Institutional Repository - Research Portal
Dépôt Institutionnel - Portail de la Recherche
researchportal.unamur.be
University of Namur

THESIS / THÈSE

MASTER EN SCIENCES MATHÉMATIQUES

Analyse de convergence et étude comparative de trois méthodes en programmation convexe. La méthode de Han et les méthodes primale et primale-duale de Iusem

GEURY, Fabienne

Award date:
1996

Awarding institution:
Universite de Namur

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 02. May. 2026

FACULTES UNIVERSITAIRES NOTRE DAME DE LA PAIX
NAMUR
FACULTES DES SCIENCES

Analyse de la convergence et étude comparative de trois méthodes en programmation convexe. La méthode de Han et les méthodes primale et primale-duale de Iusem.

Promoteur: Mr J.J. Strodiot

GEURY Fabienne

Année académique : 1995-1996

Je remercie spécialement Monsieur le Professeur J.J. Strodiot pour l'aide et les précieux conseils qu'il m'a apportés tout au long de l'élaboration de ce mémoire.
Je souhaite également remercier Maman, Yves et Nathalie de m'avoir soutenue et supportée tout au long de mes études.

Résumé

Dans ce mémoire nous prouvons la convergence de la méthode de Han, méthode de projections successives qui permet de minimiser des fonctions quadratiques strictement convexes sous des contraintes convexes.

Pour résoudre les problèmes généraux de programmation convexe nous étudions ensuite les méthodes primale et primale-duale de Iusem. A chaque itération un sous-problème de minimisation sous une ou deux contraintes linéaires est considéré. Nous établissons la convergence de la méthode primale et la comparons avec celle de Han. La méthode duale génère deux suites, l'une primale qui converge vers la solution du problème, l'autre duale qui approxime un vecteur optimal des multiplicateurs de Karush-Kuhn-Tucker. Au cours de cette analyse, nous comparons les deux méthodes de Iusem entre elles.

De toutes ces comparaisons, nous concluons que la méthode primale-duale de Iusem a le meilleur rapport vitesse de convergence/place mémoire.

Abstract

In this essay, we prove first the convergence of Han's method, a successive projection method which allows to minimize strictly convex quadratic functions subject to convex constraints.

To resolve general problems of convex programming, we study in the second place Iusem's primal and primal-dual methods. At each iteration a subproblem of minimization subject to one or two linear equations is considered. We establish the convergence of the primal method and compare it with Han's. The primal-dual method generates two sequences, the first one, primal, which converges to the solution of the problem, the other one, dual, which approximates a vector of the optimal Karush-Kuhn-Tucker multipliers. Throughout this analysis, we compare the two Iusem methods with each other.

From all these comparisons, we conclude that the primal-dual Iusem method has the best proportion between rate of convergence and memory place.

Geury Fabienne

Analyse de la convergence et étude comparative de trois méthodes en programmation convexe.
La méthode de Han et les méthodes primale et primale-duale de Iusem.

Errata

p23 : avant dernière ligne
$= \langle z_1^k - P_1(z_1^k), x_1^{k+1} - P_1(z_1^k)\rangle + \langle z_1^{k+1} - P_1(z_1^{k+1}), x_1^k - P_1(z_1^{k+1})\rangle$
au lieu de $=\langle z_1^k - P_1(z_1^k), x_1^{k+1} - P_1(z_1^k)\rangle + \langle z_1^{k+1} - P_1(z_1^{k+1}), x_1^k - P_1(x_1^{k+1})\rangle$

p24 : preuve de la proposition 1.5.8
$a_k = \|x_2^k - x_1^{k+1}\|^2$
au lieu de $a_k = \|x_2^k - x_1^{k+1}\|$

p25 : fin de la preuve
donc, $lim_{k\to\infty} kb_k = -lim_{k\to\infty} 2kc_k$
au lieu de donc, $lim_{k\to\infty} kb_k = lim_{k\to\infty} 2kc_k$

p35 : introduction dernier paragraphe
Nous donnons également deux stratégies pour le choix de la contrainte ...
au lieu de Nous également deux stratégies pour le choix de la contrainte ...

p37 : 2ième dessin
C au lieu de T
cette erreur revient aussi sur les dessins des pages 38, 43, 44 et 47.

p60 : ligne 11 1ère remarque
... hypothèses (3.13)-(3.15) sont satisfaites.
au lieu de ... hypothèses (3.13)-(3.14) sont satisfaites.

p64 : avant avant dernière ligne
...si $x_k \in C$ à partir d'un certain k, alors...
au lieu de ...si $x_k \in C$ pour tout k, alors...

p65 : 6ième ligne en bas
$g_i(x^{l_k}) \leq 0$
au lieu de $g_i(x^{l_k}) < 0$

p76 : preuve du corollaire 3.6.1
...ceci découle directement de la proposition 3.6.7
au lieu de ...ceci découle directement de la proposition 3.6.6

Table des matières

Introduction **3**

1 La méthode de Han:une méthode de projections successives **4**
1.1 Introduction 4
1.2 La méthode de Han 5
1.2.1 Position du problème 5
1.2.2 La méthode 5
1.2.3 Exemple 6
1.3 Etude de deux problèmes duaux de (P) 9
1.4 Analyse de la convergence 12
1.5 Analyse de la convergence lorsque $m = 2$ 17
1.5.1 Préliminaires 17
1.5.2 Caractère borné des suites $\{x_i^k\}$ $(i = 1, 2)$ 20
1.5.3 Convergence des suites et optimalité des limites 22
1.5.4 Nécessité de l'hypothèse $D_1 \neq \emptyset$ 26
1.6 Version simultanée de l'algorithme de Han 27
1.6.1 La méthode 27
1.6.2 Exemple 28
1.7 Analyse de la convergence de la version simultanée 29
1.8 Conditions suffisantes pour la convergence 33

2 La méthode primale de Iusem:une méthode primale de type "row-action" **35**
2.1 Introduction 35
2.2 La méthode 36
2.2.1 Définition de la méthode de type "row-action" 36
2.2.2 Position du problème 36
2.2.3 Description géométrique de l'algorithme primal 37
2.3 L'algorithme de base 38
2.3.1 Le problème (PP') 38
2.3.2 L'algorithme de base pour le problème (PP') 38
2.3.3 Propriétés élémentaires de convergence 39
2.3.4 Reformulation du sous-problème 42
2.3.5 Une variante de l'algorithme de base 44
2.3.6 Un choix spécifique de S_k 45
2.4 L'algorithme pour le problème (PP) 45
2.4.1 Une version séquentielle de l'algorithme pour le problème (PP) . . . 46
2.4.2 Stratégies pour le choix de la contrainte à utiliser à chaque itération 48

2.4.3 Une version simultanée de l'algorithme pour le problème (PP) . . . 49
2.5 Comparaison avec la méthode de Han 51
2.6 Conclusion 53

3 La méthode primale-duale de Iusem:une méthode primale-duale de type "row-action" 54
3.1 Introduction 54
3.2 La méthode 55
3.2.1 Position du problème 55
3.2.2 Description de l'algorithme primal-dual 55
3.3 L'algorithme de base 56
3.4 Version séquentielle de l'algorithme de base 57
3.4.1 Préliminaires 57
3.4.2 Algorithme et commentaires 57
3.4.3 Analyse de la convergence 61
3.5 Reformulation de l'algorithme primal-dual S 67
3.6 Taux de convergence de l'algorithme primal-dual S 70
3.7 Convergence de la suite duale 79
3.8 Conclusion 84

Conclusion Générale 85

Bibliographie 86

A Preuves Supplémentaires 87

Introduction

Nous commençons par étudier une méthode de projections successives appelée méthode de Han. Celle-ci permet de minimiser des fonctions quadratiques strictement convexes sous des contraintes convexes générales.

Nous établissons la convergence de cette méthode et présentons ensuite une version simultanée de celle-ci, pour laquelle une analyse de convergence sera également réalisée.

Dans le second chapitre, une nouvelle méthode pour minimiser une fonction strictement convexe sous des contraintes convexes générales est présentée. Cette méthode de type "row-action" est proposée par Iusem. Elle utilise une contrainte à la fois et n'effectue aucun changement sur les fonctions contraintes ni sur leur matrice jacobienne (d'où sa nature "row-action"). A chaque itération elle résout un sous-problème consistant en la minimisation de la fonction objectif sous une ou deux contraintes linéaires.

Nous établissons la convergence de la méthode primale et donnons également une version simultanée de celle-ci avant de la comparer avec la méthode de Han.

Nous terminons en présentant la seconde méthode de type "row-action" proposée par Iusem, c'est à dire la méthode primale-duale, laquelle permet encore de minimiser une fonction strictement convexe sous des contraintes convexes générales.

Celle-ci se différencie de la méthode primale de par le nombre de suites engendrées. En effet, la méthode primale-duale génère deux suites, l'une primale qui converge vers la solution du problème, l'autre duale qui approxime un vecteur optimal des mutiplicateurs de Karush-Kuhn-Tucker pour le problème considéré.

Afin d'approximer la solution duale, une matrice $m \times n$ doit être mémorisée au cours des itérations, et à chacune de celles-ci une de ses lignes est mise à jour. Ceci est le prix à payer pour avoir un algorithme primal-dual. Nous verrons, cependant, que cette matrice possède deux avantages significatifs par rapport à celle considérée dans la méthode de Han.

Nous prouvons que dans le cas de contraintes linéaires, cette méthode est toujours linéairement convergente, et nous en déduisons la convergence de la suite duale.

Chapitre 1

La méthode de Han: une méthode de projections successives

1.1 Introduction

Dans ce chapitre, nous allons étudier une méthode de projections successives appelée méthode de Han. Celle-ci est utilisée pour minimiser des fonctions quadratiques strictement convexes sur l'intersection d'un nombre fini d'ensembles convexes. De plus, elle possède une caractéristique importante : elle permet de décomposer le problème initial en plusieurs petits sous-problèmes.

La méthode de Han consiste à projecter un point sur l'intersection d'un nombre fini d'ensembles fermés et convexes de $\mathbb{R}^n$. Pour ce faire nous utilisons une suite de projections successives sur chacun de ceux-ci.

Nous analyserons la convergence de cette méthode par l'intermédiaire de problèmes duaux et étendrons cette analyse au cas d'une fonction soumise à deux contraintes.

Ensuite nous présenterons une seconde méthode qui découle de celle de Han. Celle-ci consiste encore à projeter le point courant simultanément sur chaque ensemble mais cette fois le nouvel itéré est une combinaison convexe de ces projections.

Dans ce cas-ci, nous prouverons aussi que la méthode converge (y compris dans le cas où l'intersection des ensembles est vide). Pour terminer nous formulerons des conditions suffisantes pour la convergence.

Mais avant de continuer, rappelons que la fonction indicatrice de l'ensemble C, $\delta(. \mid C)$, est définie comme suit

$$\delta(x \mid C) = \begin{cases} 0 & \text{si } x \in C, \\ \infty & \text{sinon.} \end{cases}$$

Tandis que la fonction d'appui de l'ensemble C, $\delta^*(. \mid C)$, est définie par

$$\delta^*(y \mid C) = \sup \{\langle y, x\rangle \mid x \in C\}$$

Notons par $\partial f(x)$, le sous-différentiel d'une fonction convexe f au point x, c'est à dire l'ensemble des sous-gradients de f en x, ou encore

$$\partial f(x) = \{v \in \mathbb{R}^n \mid \forall y \in \mathbb{R}^n : f(y) \geq f(x) + \langle v, y - x\rangle\}.$$

Nous définissons aussi la norme associée à la matrice Q, $\|.\|_Q$ de la manière suivante

$$\forall x \in \mathbb{R}^n \quad \|x\|_Q^2 = \langle x, Qx \rangle,$$

où Q est une matrice $n \times n$, symétrique et définie positive, et nous indiquons par riC, l'intérieur relatif de C .

1.2 La méthode de Han

1.2.1 Position du problème

Considérons le problème suivant:

$$(P) \begin{cases} \min & f(x) = \frac{1}{2}\langle x - x^0, Q(x - x^0) \rangle \\ \text{s.c.} & x \in C = \bigcap_{i=1}^m C_i \end{cases}$$

où x^0 est un point donné de $\mathbb{R}^n$,
C_i, $1 \leq i \leq m$, sont des ensembles convexes, fermés $\subseteq \mathbb{R}^n$,
Q est une matrice $n \times n$, symétrique, définie positive.

Notons que, grâce à la définition de la fonction indicatrice de l'ensemble C, le problème peut encore s'écrire comme

$$(\mathrm{P}) \quad \min_{x \in \mathbf{R}^n} f(x) + \delta(x \mid C),$$

où $f(x) = \frac{1}{2}\|x - x^0\|_Q^2$ et avec x^0, C_i, Q sont comme ci-dessus.

1.2.2 La méthode

Soit $a \in \mathbb{R}^n$. Notons $P_i(a)$ la projection de a sur l'ensemble C_i par rapport à la norme associée à la matrice Q.

Ce point $P_i(a)$ est l'unique solution du problème

$$\begin{cases} \min & f(x) = \frac{1}{2}\langle x - a, Q(x - a) \rangle \\ \text{s.c.} & x \in C_i \end{cases}$$

et peut aussi être caractérisé comme le seul point $\bar{x} \in C_i$ vérifiant l'inégalité

$$\langle z - \bar{x}, Q(a - \bar{x}) \rangle \leq 0 \ , \ \forall z \in C_i. \tag{1.1}$$

Pour résoudre le problème (P), nous pourrions considérer une méthode de projections successives directes c'est-à-dire x^0 serait projeté sur C_1, le point obtenu serait à son tour projeté sur C_2 et ainsi de suite. Un exemple simple (cfr section suivante) montre que cette méthode ne fournit pas toujours la solution de (P). C'est pourquoi, dans la méthode de Han, nous modifions le point avant de le projeter.

ALGORITHME DE HAN

Initialisation

- $x_m^0 = x^0$ un point donné,
- $y_i^0 = 0, \quad 1 \leq i \leq m.$

Itération

Pour $k = 1, 2, ...$

- $x_0^k = x_m^{k-1}$,
- Pour $i = 1, ..., m$

$$z_i^k = x_{i-1}^k + y_i^{k-1}, \tag{1.2}$$

$$x_i^k = P_i(z_i^k), \tag{1.3}$$

$$y_i^k = z_i^k - P_i(z_i^k). \tag{1.4}$$

où l'itération majeure est notée par un indice supérieur k tandis que l'itération mineure est notée par un indice inférieur i. (1.2)-(1.4) peut aussi se réécrire comme $y_i^k = x_{i-1}^k + y_i^{k-1} - x_i^k$.

Dans cette méthode, nous calculons le vecteur y_i^k normal externe à l'ensemble C_i au point x_i^k, et nous ajoutons le vecteur normal externe y_i^{k-1} au vecteur x_{i-1}^k avant de le projeter sur l'ensemble C_i. Par conséquent, à chaque itération nous remplaçons un ancien vecteur normal externe par un nouveau et ainsi cette suite convergera vers une solution d'un problème dual de (P). Il faudra donc étudier dans la suite quelques problèmes duaux de (P). Mais commençons d'abord par regarder, sur un exemple simple, ce que fait cet algorithme.

1.2.3 Exemple

Considérons le problème suivant :

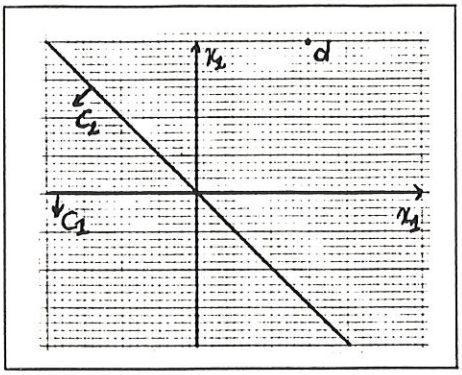
$$(\text{P1}) \begin{cases} \min & \|x - d\|^2 \\ \text{s.c.} & x \in C_i, \quad i = 1, 2 \end{cases}$$

où $C_1 = \{(x_1, x_2) \mid x_2 \leq 0\}$,
$C_2 = \{(x_1, x_2) \mid x_1 + x_2 \leq 0\}$,
d un point donné de $\mathbb{R}^2$.

Appliquons l'algorithme de Han à ce problème pour seulement deux itérations afin d'avoir une idée géométrique de cet algorithme :

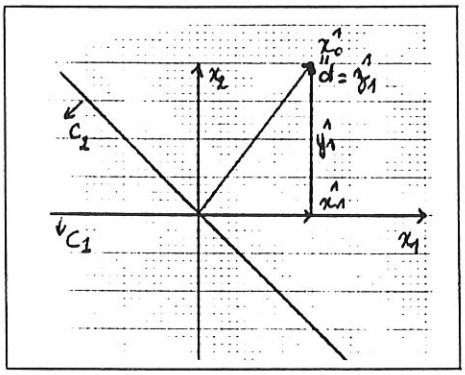
1) Situation initiale

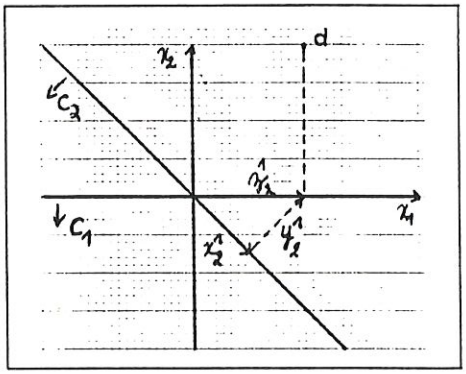
$$\begin{cases} y_1^0 = y_2^0 = 0 \\ x_2^0 = d \end{cases}$$



2) Première itération : $k = 1$

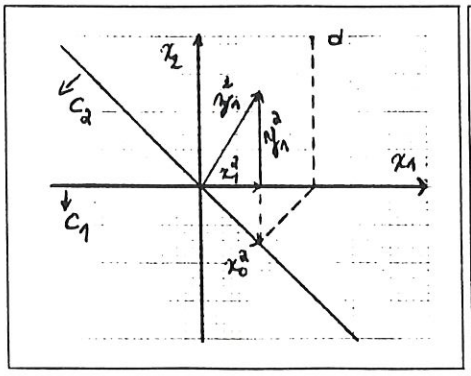
$$\begin{cases} x_0^1 = x_2^0 \\ i = 1 \begin{cases} z_1^1 = x_0^1 + y_1^0 \\ x_1^1 = P_1(z_1^1) \\ y_1^1 = z_1^1 - P_1(z_1^1) \end{cases} \\ i = 2 \begin{cases} z_2^1 = x_1^1 + y_2^0 \\ x_2^1 = P_2(z_2^1) \\ y_2^1 = z_2^1 - P_2(z_2^1) \end{cases} \end{cases}$$

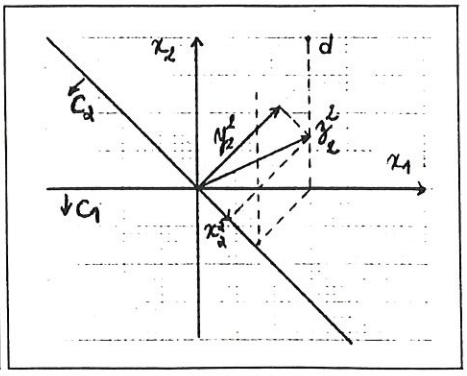




3) Deuxième itération : $k = 2$

$$\begin{cases} x_0^2 = x_2^1 \\ i = 1 \begin{cases} z_1^2 = x_0^2 + y_1^1 \\ x_1^2 = P_1(z_1^2) \\ y_1^2 = z_1^2 - P_1(z_1^2) \end{cases} \\ i = 2 \begin{cases} z_2^2 = x_1^2 + y_2^1 \\ x_2^2 = P_2(z_2^2) \\ y_2^2 = z_2^2 - P_2(z_2^2) \end{cases} \end{cases}$$



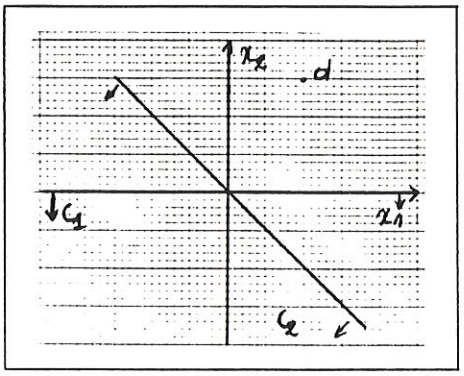


Si nous utilisons la méthode de projections successives directes pour le même problème (P1) avec un point de départ d dont les coordonnées sont telles que $d_1 \neq 0$ et $d_1 \neq d_2$, alors cette méthode ne converge pas vers la solution du problème considéré.

Nous le constatons en effectuant les deux premières itérations :

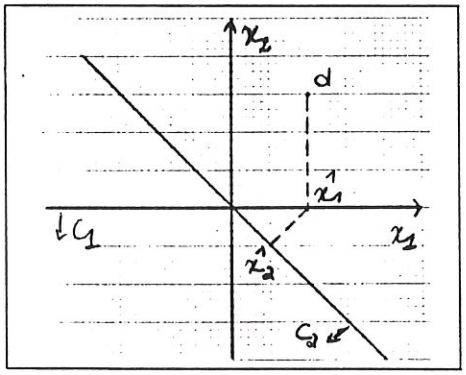
1) Situation initiale

$x_2^0 = d$



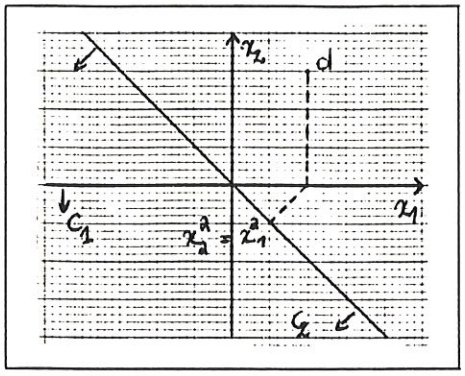
2) Première itération :

$$\begin{cases} x_0^1 = x_2^0 = d \\ x_1^1 = P_1(x_0^1) \\ x_2^1 = P_2(x_1^1) \end{cases}$$



3) Deuxième itération :

$$\begin{cases} x_0^2 = x_2^1 \\ x_1^2 = P_1(x_0^2) = P_1(x_2^1) = x_2^1 \\ x_2^2 = P_2(x_1^2) = P_2(x_2^1) = x_2^1 \end{cases}$$



Nous resterons donc au fur et à mesure des itérations au point x_2^1 (car $x_2^1 \in C_1 \bigcap C_2$) qui n'est pas la solution du problème (P1).

1.3 Etude de deux problèmes duaux de (P)

Afin d'analyser la convergence de la méthode de Han, nous devons étudier des problèmes duaux de (P). Ici, considérons les deux problèmes duaux de (P) suivants :

$$(\mathrm{D}) \quad \min_{y \in \mathbb{R}^n} g(y) = f(y) + \delta^*(y \mid QC)$$

et

$$(\mathrm{D}') \quad \min_{y_i \in \mathbb{R}^n} h(y_1, \ldots, y_m) = f(y_1 + \ldots + y_m) + \sum_{i=1}^{m} \delta^*(y_i \mid QC_i)$$

où $QC = \{y \in \mathbb{R}^n \mid y = Qx,\ x \in C\}$ et $QC_i = \{y_i \in \mathbb{R}^n \mid y_i = Qx_i,\ x_i \in C_i\}$.

Pour commencer, nous donnerons une description de la relation entre les problèmes (P) et (D).

Théorème 1.3.1 *Soient $\bar{x}, \bar{y} \in \mathbb{R}^n$ tels que $\bar{x} + \bar{y} = x^0$ et soit $C \neq \emptyset, C \neq \mathbb{R}^n$.*
Alors les assertions suivantes sont équivalentes,
(a) $\bar{x}$ *résout* (P),
(b) $\bar{y}$ *résout* (D),
(c) $\bar{x} \in C$ *et* $\delta^*(\bar{y} \mid QC) = \langle \bar{x}, Q\bar{y} \rangle$.
De plus, les solutions de (P) *et* (D) *existent et sont uniques.*

Preuve

Comme $\partial f(\bar{x}) = \{-Q\bar{y}\}$ et $\partial f(\bar{y}) = \{-Q\bar{x}\}$, nous avons successivement

$$\begin{aligned}
(a) &\Leftrightarrow 0 \in [\partial f(\bar{x}) + \partial \delta(\bar{x} \mid C)] \\
&\Leftrightarrow Q\bar{y} \in \partial \delta(\bar{x} \mid C) \\
&\Leftrightarrow \delta(x \mid C) - \langle Q\bar{y}, x \rangle \geq \delta(\bar{x} \mid C) - \langle Q\bar{y}, \bar{x} \rangle \quad \forall x \in \mathbb{R}^n \\
&\Leftrightarrow \delta(\bar{x} \mid C) + \delta^*(Q\bar{y} \mid C) = \langle \bar{x}, Q\bar{y} \rangle \\
&\Leftrightarrow (c)
\end{aligned}$$

où l'avant dernière équivalence est obtenue grâce aux définitions du sous-gradient et de la fonction d'appui δ^* sur l'ensemble C.

De manière analogue,

$$\begin{aligned}
(b) &\Leftrightarrow 0 \in [\partial f(\bar{y}) + \partial\delta^*(\bar{y} \mid QC)] \\
&\Leftrightarrow Q\bar{x} \in \partial\delta^*(\bar{y} \mid QC) \\
&\Leftrightarrow \delta(Q\bar{x} \mid QC) + \delta^*(\bar{y} \mid QC) = \langle \bar{x}, Q\bar{y} \rangle \\
&\Leftrightarrow \delta(\bar{x} \mid C) + \delta^*(\bar{y} \mid QC) = \langle \bar{x}, Q\bar{y} \rangle \\
&\Leftrightarrow (c).
\end{aligned}$$

De plus, comme une projection sur un ensemble convexe fermé non vide existe et est unique, il suit que les problèmes (P) et (D) admettent une et une seule solution.

□

A présent, examinons les liens existant entre les problèmes (P), (D) et (D'). Commençons par donner une condition nécessaire et suffisante pour que (D') ait une solution.

Lemme 1.3.1 $(\bar{y}_1, \ldots, \bar{y}_m) \in \mathbb{R}^{m \times n}$ *résout* (D') *si et seulement si les deux conditions suivantes sont satisfaites,*
(a) $\bar{x} = x^0 - \bar{y}_1 - \bar{y}_2 - \ldots - \bar{y}_m \in C$,
(b) $\forall i, 1 \leq i \leq m,\ \langle \bar{x}, Q\bar{y}_i \rangle = \delta^*(\bar{y}_i \mid QC_i)$.

Preuve

Remarquons que $\partial h(y_1, \ldots, y_m) = \begin{pmatrix} -Qx + \partial\delta^*(y_1 \mid QC_1) \\ \ldots \\ -Qx + \partial\delta^*(y_m \mid QC_m) \end{pmatrix}$

où $x = x^0 - y_1 - \ldots - y_m$.
Dès lors,

$$\begin{aligned}
&(\bar{y}_1, \ldots, \bar{y}_m) \text{ résout (D')} \\
&\Leftrightarrow 0 \in \partial h(\bar{y}_1, \ldots, \bar{y}_m) \\
&\Leftrightarrow Q\bar{x} \in \partial\delta^*(\bar{y}_i \mid QC_i), \quad 1 \leq i \leq m \\
&\Leftrightarrow \langle \bar{x}, Q\bar{y}_i \rangle = \delta^*(\bar{y}_i \mid QC_i) + \delta(\bar{x} \mid C_i), \quad 1 \leq i \leq m \\
&\Leftrightarrow \langle \bar{x}, Q\bar{y}_i \rangle = \delta^*(\bar{y}_i \mid QC_i),\ \bar{x} \in C_i, \quad 1 \leq i \leq m \\
&\Leftrightarrow \bar{x} \in C \text{ et } \langle \bar{x}, Q\bar{y}_i \rangle = \delta^*(\bar{y}_i \mid QC_i), \quad 1 \leq i \leq m \\
&\Leftrightarrow (a) \text{ et } (b).
\end{aligned}$$

□

Le théorème que nous introduisons maintenant est essentiel pour l'analyse de la convergence de la méthode de Han.

Théorème 1.3.2 *Soient* C_u, $1 \leq u \leq k$ *des polyèdres et* C_v, $k+1 \leq v \leq m$ *des ensembles convexes fermés tels que*

$$\left(\bigcap_{u=1}^{k} C_u \right) \bigcap \left(\bigcap_{v=k+1}^{m} riC_v \right) \neq \emptyset.$$

Alors $\bar{x}$ *résout* (P) *si et seulement si* $\begin{cases} \text{il existe } \bar{y}_1, \ldots, \bar{y}_m \text{ tels que } \bar{x} + \sum_{i=1}^{m} \bar{y}_i = x^0 \\ \text{et } (\bar{y}_1, \ldots, \bar{y}_m) \text{ résout (D')}. \end{cases}$

De plus, dans ce cas-là, le vecteur $\bar{y} = \sum_{i=1}^{m} \bar{y}_i$ *résout* (D) *et*

$$\inf_{\bar{y}_i \in \mathbf{R}^n} h(\bar{y}_1, \ldots, \bar{y}_m) = \inf_{\bar{y} \in \mathbf{R}^n} g(\bar{y})$$

Preuve

Supposons d'abord que $\bar{x}$ résout (P). Alors $Q\bar{y} \in \partial\delta(\bar{x} \mid C)$ où $\bar{y} = x^0 - \bar{x}$. Comme par hypothèse, $\partial\delta(\bar{x} \mid C) = \partial\delta(\bar{x} \mid C_1) + \ldots + \partial\delta(\bar{x} \mid C_m)$, il existe $\bar{y}_1, \ldots, \bar{y}_m$ tels que

$$\bar{y} = \bar{y}_1 + \ldots + \bar{y}_m \quad \text{et} \quad Q\bar{y}_i \in \partial\delta(\bar{x} \mid C_i).$$

D'où,

$$\langle \bar{x}, Q\bar{y}_i \rangle = \delta^*(Q\bar{y}_i \mid C_i) = \delta^*(\bar{y}_i \mid QC_i),$$

et par le lemme 1.3.1, $(\bar{y}_1, \ldots, \bar{y}_m)$ résout (D').

Réciproquement, supposons que $(\bar{y}_1, \ldots, \bar{y}_m)$ résout (D'), $\bar{y} = \bar{y}_1 + \ldots + \bar{y}_m$ et $\bar{x} + \bar{y} = x^0$. Du lemme 1.3.1, nous déduisons

$$\bar{x} \in C, \quad \text{et} \quad \langle \bar{x}, Q\bar{y} \rangle = \sum_{i=1}^{m} \langle \bar{x}, Q\bar{y}_i \rangle = \sum_{i=1}^{m} \delta^*(\bar{y}_i \mid QC_i).$$

Par hypothèse et en utilisant le corollaire 16.4.1 de [1], nous avons

$$\langle \bar{x}, Q\bar{y} \rangle = \sum_{i=1}^{m} \delta^*(\bar{y}_i \mid QC_i) \geq \delta^*(\bar{y} \mid QC).$$

D'autre part, comme $\bar{x} \in C$ implique que $\delta^*(\bar{y} \mid QC) \geq \langle \bar{x}, Q\bar{y} \rangle$, nous obtenons

$$\langle \bar{x}, Q\bar{y} \rangle = \delta^*(\bar{y} \mid QC).$$

Et par le théorème 1.3.1, $\bar{x}$ résout (P) et $\bar{y}$ résout (D).

De plus,

$$\begin{aligned} \inf h(\bar{y}_1, \ldots, \bar{y}_m) &= \inf \left(f(\bar{y}) + \sum_{i=1}^{m} \delta^*(\bar{y}_i \mid QC_i) \right) \\ &= \inf \left(f(\bar{y}) + \delta^*(\sum_{i=1}^{m} \bar{y}_i \mid Q \bigcap_{i=1}^{m} C_i) \right) \\ &= \inf g(\bar{y}). \end{aligned}$$

□

1.4 Analyse de la convergence

La méthode de projections successives présentée fournit trois suites $\{(y_1^k, \ldots, y_m^k)\}$, $\{y^k\}$ et $\{x^k\}$, où $y^k = \sum_{i=1}^m y_i^k$ et $x^k = x_m^k$. Afin de montrer que ces trois suites convergent respectivement vers les solutions de (D'), (D) et (P), établissons la décroissance des valeurs $\{h(y_1^k, \ldots, y_m^k)\}$.

Pour simplifier, supposons que $C \neq \mathbb{R}^n$ et $C \neq \emptyset$, et notons $\bar{x}$ et $\bar{y}$ les solutions uniques respectives de (P) et (D).

Lemme 1.4.1 *Pour tout $i = 1, \ldots, m$ et pour tout $k = 1, 2, \ldots$ les assertions suivantes ont lieu,*
(a) $x_i^k \in C_i$,
(b) $x^k + y^k = x^0$ *où* $x^k = x_m^k$ *et* $y^k = \sum_{i=1}^m y_i^k$,
(c) $\langle x_i^k, Q y_i^k \rangle = \delta^*(y_i^k \mid QC_i)$.

Preuve

(a) Par définition x_i^k est la projection d'un point sur C_i.

(b) En sommant les équations $y_i^k = x_{i-1}^k + y_i^{k-1} - x_i^k$ pour $i = 1, \ldots, m$, nous obtenons l'égalité suivante

$$\sum_{i=1}^m y_i^k = x_0^k - x_m^k + \sum_{i=1}^m y_i^{k-1},$$

ou encore, $y^k = x_0^k - x^k + y^{k-1}$. Comme $x_0^k = x_m^{k-1} = x^{k-1}$, nous avons ensuite

$$x^k + y^k = x^{k-1} + y^{k-1},$$

et par récurrence nous obtenons finalement $x^k + y^k = x^0 + y^0$. Or $y^0 = \sum_{i=1}^m y_i^0 = 0$, par conséquent l'assertion (b) est vérifiée.

(c) Puisque x_i^k est la projection de $x_{i-1}^k + y_i^{k-1}$ sur l'ensemble fermé convexe C_i, par (1.1), nous avons

$$\langle z - x_i^k, Q(x_{i-1}^k + y_i^{k-1} - x_i^k) \rangle \leq 0 \quad \forall z \in C_i.$$

Or, $y_i^k = x_{i-1}^k + y_i^{k-1} - x_i^k$, donc

$$\langle z, Q y_i^k \rangle \leq \langle x_i^k, Q y_i^k \rangle, \quad \forall z \in C_i,$$

et comme $x_i^k \in C_i$, nous obtenons

$$\sup_{z \in C_i} \langle Qz, y_i^k \rangle = \langle Q x_i^k, y_i^k \rangle,$$

c'est à dire (c)

□

Lemme 1.4.2 *L'inégalité suivante est vérifiée:*

$$h(y_1^{k-1}, \ldots, y_m^{k-1}) \geq h(y_1^k, \ldots, y_m^k) + \frac{1}{2} \sum_{i=1}^m \|y_i^k - y_i^{k-1}\|_Q^2.$$

Preuve

De la définition de y_i^k, nous avons $x_{i-1}^k = x_i^k + y_i^k - y_i^{k-1}$. D'où,

$$\|x_{i-1}^k\|_Q^2 = \|x_i^k\|_Q^2 + \|y_i^k - y_i^{k-1}\|_Q^2 + 2\langle x_i^k, Q(y_i^k - y_i^{k-1})\rangle.$$

Il suit alors des assertions (a) et (c) du lemme 1.4.1,

$$\begin{aligned}
&\frac{1}{2}\|x_{i-1}^k\|_Q^2 + \delta^*(y_i^{k-1} \mid QC_i) \\
&\quad \geq \frac{1}{2}\|x_{i-1}^k\|_Q^2 + \langle x_i^k, Qy_i^{k-1}\rangle \\
&\quad = \frac{1}{2}\|x_i^k\|_Q^2 + \langle x_i^k, Qy_i^k\rangle + \frac{1}{2}\|y_i^k - y_i^{k-1}\|_Q^2 \\
&\quad = \frac{1}{2}\|x_i^k\|_Q^2 + \delta^*(y_i^k \mid QC_i) + \frac{1}{2}\|y_i^k - y_i^{k-1}\|_Q^2.
\end{aligned}$$

En sommant les inégalités extrêmes pour $i = 1, \ldots, m$, nous obtenons après simplification des termes semblables,

$$\begin{aligned}
&\frac{1}{2}\|x_0^k\|_Q^2 + \sum_{i=1}^m \delta^*(y_i^{k-1} \mid QC_i) \geq \\
&\qquad \frac{1}{2}\|x_m^k\|_Q^2 + \sum_{i=1}^m \delta^*(y_i^k \mid QC_i) + \frac{1}{2}\sum_{i=1}^m \|y_i^k - y_i^{k-1}\|_Q^2.
\end{aligned}$$

La thèse suit alors de la définition de f et du fait que $x_0^k = x_m^{k-1} = x^0 - y_1^{k-1} - \ldots - y_m^{k-1}$ et $x_m^k = x^0 - y_1^k - \ldots - y_m^k$.

□

Nous venons d'établir la décroissance des valeurs $\{h(y_1^k, \ldots, y_m^k)\}$. Montrons maintenant que les trois suites $\{(y_1^k, \ldots, y_m^k)\}$, $\{y^k\}$ et $\{x^k\}$ fournies par la méthode de Han convergent.

Corollaire 1.4.1 *Pour tout $i = 1, \ldots, m$ les assertions suivantes ont lieu*
(a) $\lim_{k\to\infty} \|y_i^k - y_i^{k-1}\|_Q = 0$,
(b) $\lim_{k\to\infty} \|x_{i-1}^k - x_i^k\|_Q = 0$.
Par conséquent, $\lim_{k\to\infty} \|y^k - y^{k-1}\|_Q = 0$ *et* $\lim_{k\to\infty} \|x^k - x^{k-1}\|_Q = 0$.

Preuve

Une preuve de ce corollaire peut être trouvée en annexe.

□

Proposition 1.4.1 *Tout point d'accumulation de $\{(y_1^k, \ldots, y_m^k)\}$ résout* (D').

Preuve

Sans perte de généralité, nous pouvons supposer que $\{(y_1^k, \ldots, y_m^k)\}$ converge vers $(\tilde{y}_1, \ldots, \tilde{y}_m)$. Définissons $\tilde{x} = x^0 - \tilde{y}_1 - \ldots - \tilde{y}_m$, alors il suit du corollaire 1.4.1 que pour tout i, $\lim_{k\to\infty} x_i^k = \tilde{x}$. Comme, $x_i^k \in C_i$, pour tout i, nous avons

$$\tilde{x} \in C.$$

D'où, par la semi-continuité inférieure de δ^* en $\tilde{y}_i$, et par le lemme 1.4.1(c) nous avons aussi pour tout i,

$$\begin{aligned} \langle \tilde{x}, Q\tilde{y}_i \rangle &\leq \delta^*(\tilde{y}_i \mid QC_i) \\ &\leq \liminf \delta^*(y_i^k \mid QC_i) \\ &= \liminf \langle x_i^k, Qy_i^k \rangle \\ &= \langle \tilde{x}, Q\tilde{y}_i \rangle, \end{aligned}$$

et donc

$$\langle \tilde{x}, Q\tilde{y}_i \rangle = \delta^*(\tilde{y}_i \mid QC_i).$$

Par le lemme 1.3.1, nous déduisons que $(\tilde{y}_1, \ldots, \tilde{y}_m)$ résout (D').

□

Dans les théorèmes suivants nous donnons des conditions suffisantes pour que les suites $\{x^k\}$ et $\{y^k\}$ convergent vers les solutions respectives de (P) et (D).

Théorème 1.4.1 *Soient $C_u, 1 \leq u \leq k$ des polyhèdres et $C_v, k+1 \leq v \leq m$ des ensembles convexes fermés tels que*

$$\left(\bigcap_{u=1}^{k} C_u\right) \bigcap \left(\bigcap_{v=k+1}^{m} riC_v\right) \neq \emptyset.$$

Si la suite $\{(y_1^k, \ldots, y_m^k)\}$ a un point d'accumulation, alors les suites $\{x^k\}$ et $\{y^k\}$ convergent respectivement vers les solutions de (P) *et* (D).

Preuve

Par hypothèse et par le théorème 1.3.2, nous avons $\inf_{y_i \in \mathbf{R}^n} h(y_1, \ldots, y_m) = \inf_{y \in \mathbf{R}^n} g(y)$. D'autre part, de la proposition 1.4.1 et de la décroissance de $\{h(y_1^k, \ldots, y_m^k)\}$, il suit que

$$\lim_{k\to\infty} h(y_1^k, \ldots, y_m^k) = \inf_{y_i \in \mathbf{R}^n} h(y_1, \ldots, y_m) = \inf_{y \in \mathbf{R}^n} g(y).$$

Et comme $h(y_1^k, \ldots, y_m^k) \geq g(y^k)$, nous avons

$$\lim_{k\to\infty} h(y_1^k, \ldots, y_m^k) \geq \lim_{k\to\infty} g(y^k) \geq \inf_{y \in \mathbf{R}^n} g(y).$$

D'où,

$$\lim_{k\to\infty} g(y^k) = \inf_{y \in \mathbf{R}^n} g(y) \stackrel{def}{=} g(\bar{y}),$$

où $\bar{y}$ est solution de (D). Dès lors, comme g est convexe, propre, s.c.i.

$$y^k \to \bar{y}.$$

De plus, comme $x^k + y^k = \bar{x} + \bar{y} = x^0$, où $\bar{x}$ est la solution de (P), nous avons

$$x^k \to \bar{x}.$$

□

Lemme 1.4.3 *Si* int$C \neq \emptyset$, *alors la suite* $\{(y_1^k, \ldots, y_m^k)\}$ *est bornée.*

Preuve

Soit $z \in \text{int}C$, nous avons

$$\begin{aligned}
h(y_1, \ldots, y_m) &= \frac{1}{2}\|y_1 + \ldots + y_m - x^0\|_Q^2 + \sum_{i=1}^m \delta^*(y_i \mid QC_i) \\
&= \frac{1}{2}\|y_1 + \ldots + y_m - x^0 + z\|_Q^2 - \frac{1}{2}\|z\|_Q^2 \\
&\quad -\langle y_1 + \ldots + y_m - x^0, Qz\rangle + \sum_{i=1}^m \delta^*(y_i \mid QC_i) \\
&= \frac{1}{2}\|y_1 + \ldots + y_m - x^0 + z\|_Q^2 + \sum_{i=1}^m \delta^*(y_i \mid QC_i) \\
&\quad -\sum_{i=1}^m \delta^*(y_i \mid Qz) + \langle x^0 - \frac{1}{2}z, Qz\rangle \\
&= \frac{1}{2}\|y_1 + \ldots + y_m - x^0 + z\|_Q^2 + \sum_{i=1}^m \delta^*(y_i \mid Q(C_i - z)) \\
&\quad +\langle x^0 - \frac{1}{2}z, Qz\rangle.
\end{aligned}$$

Du lemme 1.4.2 et de l'égalité ci-dessus, nous avons

$$\begin{aligned}
\frac{1}{2}\|x^0 - z\|_Q^2 &= h(y_1^0, \ldots, y_m^0) + \frac{1}{2}\langle z, Qz\rangle - \langle x^0, Qz\rangle \\
&\geq h(y_1^k, \ldots, y_m^k) + \frac{1}{2}\langle z, Qz\rangle - \langle x^0, Qz\rangle \\
&\geq \sum_{i=1}^m \delta^*(y_i^k \mid Q(C_i - z)).
\end{aligned}$$

Puisque $0 \in Q(C_i - z)$, nous avons $\delta^*(y_i^k \mid Q(C_i - z)) \geq 0$. D'où pour tout i,

$$y_i^k \in S_i \stackrel{def}{=} \{y \mid \delta^*(y \mid Q(C_i - z)) \leq \frac{1}{2}\|x^0 - z\|_Q^2\}.$$

Et par le corollaire 14.2.2 de [1], les ensembles S_i sont bornés car

$$0 \in \text{int}(Q(C_i - z)) = \text{int}(\text{dom}(\delta(. \mid Q(C_i - z)))).$$

□

Théorème 1.4.2 *Si* int$C \neq \emptyset$, *alors les suites* $\{x^k\}$ *et* $\{y^k\}$ *convergent respectivement vers les solutions de* (P) *et* (D).

Preuve

Elle découle directement du lemme 1.4.3 et du théorème 1.4.1

□

Théorème 1.4.3 *Soient* C_u, $1 \leq u \leq p$ *des polyhèdres et* C_v, $p+1 \leq v \leq m$ *des ensembles convexes fermés tels que*

$$\left(\bigcap_{u=1}^{p} C_u\right) \bigcap \left(\bigcap_{v=p+1}^{m} \text{int}C_v\right) \neq \emptyset,$$

alors les suites $\{x^k\}$ *et* $\{y^k\}$ *convergent respectivement vers les solutions de* (P) *et* (D).

Preuve

Une preuve de ce théorème peut être trouvée dans [5] (théorème 4.8 p8-9).

□

D'un point de vue théorique, il est intéressant d'établir la convergence des suites $\{x^k\}$ et $\{y^k\}$ vers les solutions respectives de (P) et (D).

Seulement, ce n'est pas parce que ces suites convergent, que nécessairement la suite $\{y_1^k, \ldots, y_m^k\}$ convergera. Or d'un point de vue pratique, ces vecteurs sont réellement calculés, c'est pourquoi la convergence de cette suite est plus importante.

Par conséquent nous nous intéressons maintenant à des situations où la convergence de $\{y_1^k, \ldots, y_m^k\}$ est assurée. De telles conditions sont considérées dans le lemme et le théorème qui suivent.

Lemme 1.4.4 *Si* (D') *a une solution unique alors la suite* $\{(y_1^k, \ldots, y_m^k)\}$ *converge vers la solution de* (D').

Preuve

Par hypothèse, les ensembles de niveaux de h sont bornés, donc, la suite $\{(y_1^k, \ldots, y_m^k)\}$ est bornée, et tout point d'accumulation de cette suite résout (D'), par la proposition 1.4.1. Par conséquent le résultat découle directement de l'unicité de la solution.

□

Théorème 1.4.4 *Soient* C_u, $1 \leq u \leq p$ *des polyhèdres et* C_v, $p+1 \leq v \leq m$ *des ensembles convexes fermés tels que*

$$\left(\bigcap_{u=1}^{p} C_u\right) \bigcap \left(\bigcap_{v=p+1}^{m} riC_v\right) \neq \emptyset$$

Si les sous-espaces affines $\text{aff}(\partial\delta(\bar{x} \mid C_i))$ *sont linéairement indépendants dans le sens que*

$$\sum z_i = 0 \quad et \quad z_i \in \text{aff}(\partial\delta(\bar{x} \mid C_i)) \quad \Rightarrow \quad z_i = 0 \quad \forall i,$$

alors les suites $\{x^k\}$, $\{y^k\}$ *et* $\{(y_1^k, \ldots, y_m^k)\}$ *convergent respectivement vers les solutions de* (P), (D) *et* (D').

Preuve

Par le théorème 1.3.2, le problème (D') possède des solutions. Si $\{(\tilde{y}_1, \ldots, \tilde{y}_m)\}$ et $\{(\bar{y}_1, \ldots, \bar{y}_m)\}$ sont deux solutions de (D') alors, toujours par le théorème 1.3.2, $\tilde{y} = \tilde{y}_1 + \ldots + \tilde{y}_m = \bar{y}_1 + \ldots + \bar{y}_m$ est l'unique solution de (D). Par conséquent, avec $z_i \stackrel{def}{=} \tilde{y}_i - \bar{y}_i$, nous avons $\sum z_i = 0$ et $z_i \in \text{aff}(\partial\delta(\bar{x} \mid C_i))$. Il suit de la condition d'indépendance linéaire que $(\tilde{y}_1, \ldots, \tilde{y}_m) = (\bar{y}_1, \ldots, \bar{y}_m)$. Le théorème découle alors immédiatement du lemme 1.4.4 et du théorème 1.3.2

□

1.5 Analyse de la convergence lorsque $m = 2$

1.5.1 Préliminaires

Nous étendons l'analyse de la convergence de la méthode de Han pour $m = 2$ (i.e. deux contraintes). Pour cela, nous substituons l'hypothèse $intC \neq \emptyset$ par l'existence de points $a_1 \in C_1$ et $a_2 \in C_2$ tels que $d(a_1, a_2) = d =$ la distance entre les deux ensembles convexes C_1 et C_2.

Commençons par réécrire l'algorithme de Han pour $m = 2$ sous une forme un peu différente.

Initialisation

- $x^0 \in \mathbb{R}^n$ un point quelconque,
- $z_1^1 = x_0$,
- $z_2^1 = P_1(x^0)$.

Itération

Pour $k = 1, 2, \ldots$

- $x_i^k = P_i(z_i^k), \quad 1 \leq i \leq 2,$ (1.5)
- $z_1^{k+1} = x_2^k + z_1^k - x_1^k,$ (1.6)
- $z_2^{k+1} = x_1^{k+1} + z_2^k - x_2^k.$ (1.7)

La proposition suivante est une conséquence de la relation de récurrence (1.5)-(1.7).

Proposition 1.5.1 *Pour $k = 1, 2, \ldots$, les égalités suivantes ont lieu,*

$$z_1^k = x^0 + \sum_{j=1}^{k-1}(x_2^j - x_1^j), \tag{1.8}$$

$$z_2^k = x_1^k + \sum_{j=1}^{k-1}(x_1^j - x_2^j) = x_2^k + \sum_{j=1}^{k}(x_1^j - x_2^j), \tag{1.9}$$

$$x^0 = z_1^{k+1} + z_2^k - x_2^k. \tag{1.10}$$

Preuve

La preuve est immédiate par récurrence sur (1.5)-(1.7).

□

L'analyse de la convergence de la méthode de Han donnée ci-avant est réalisée pour $intC \neq \emptyset$ (avec $C = C_1 \bigcap C_2$), et nous y prouvons que les deux suites $\{x_1^k\}$ et $\{x_2^k\}$ convergent vers $x^* = \text{argmin}_{x \in C} \|x^0 - x\|$.

Dans la prochaine analyse l'intérieur de C peut être vide. Dans ce cas le problème primal peut être inadmissible et le problème dual peut ne pas être borné de même que la suite $\{z_i^k\}$, comme nous le montrerons. C'est pourquoi nous allons travailler dans une optique assez différente.

Soit $d = d(C_1, C_2) = \inf\{\|x - y\| : x \in C_1,\ y \in C_2\}$. Considérons $D_1 = \{x \in C_1 : d(x, C_2) = d\}$ et $D_2 = \{x \in C_2 : d(x, C_1) = d\}$. Il est clair que $D_1 \neq \emptyset$ si et seulement si $D_2 \neq \emptyset$. Sous l'hypothèse de convergence de base $D_1 \neq \emptyset$, nous allons montrer que $\{x_i^k\}$ converge vers $x_i^* = \text{argmin}_{x \in D_i} \|x^0 - x\|$, c'est à dire vers la projection orthogonale de x^0 sur D_i.

Nous présentons d'abord quelques rappels à propos des projections orthogonales.

Soient C un ensemble convexe fermé et P l'opérateur de projection sur C. P est bien défini et continu, de plus il suit du théorème de séparation des convexes que

$$\langle x - P(x), y - P(x) \rangle \leq 0 \quad \forall y \in C, \quad \forall x \in \mathbb{R}^n. \tag{1.11}$$

Une conséquence de (1.11) est donnée dans la proposition suivante.

Proposition 1.5.2 *Les assertions suivantes ont lieu,*
(i) $\|P(x) - P(y)\| \leq \|x - y\| \quad \forall x, y \in \mathbb{R}^n$,
(ii) *si* $\|P(x) - P(y)\| = \|x - y\|$ *alors*
(a) $P(x) - P(y) = x - y$,
(b) $\langle P(x) - x, y - x \rangle = 0$.

Preuve

Considérons l'égalité suivante,

$$\begin{aligned} \|P(x) - P(y) - x + y\|^2 &= \|x - y\|^2 - \|P(x) - P(y)\|^2 \\ &+2[\langle x - P(x), P(y) - P(x) \rangle + \langle y - P(y), P(x) - P(y) \rangle] \end{aligned} \tag{1.12}$$

et remarquons par (1.11) que les produits internes du membre de droite de (1.12) sont négatifs. Et comme le membre de gauche est toujours positif, l'assertion (i) est vraie. De l'hypothèse de l'assertion (ii), le membre de droite se réduit aux produits internes négatifs et est donc négatif. Or cette même expression est positive, et par conséquent nulle ce qui prouve (a) et (b).

□

Nous utiliserons dans la suite l'implication suivante

$$\langle x-w, y-w\rangle \leq 0 \Rightarrow \|x-y\| \geq \|x-w\| \tag{1.13}$$

qui découle de l'égalité $\|x-y\|^2 = \|x-w\|^2 + \|w-y\|^2 - 2\langle x-w, y-w\rangle$.

Les résultats requis sur les ensembles D_i sont exprimés dans la proposition suivante.

Proposition 1.5.3 *Les assertions suivantes ont lieu.*
(i) D_i *est fermé et convexe* $(i = 1, 2)$,
(ii) $u \in D_1 \Rightarrow P_2(u) \in D_2$, $P_1(P_2(u)) = u$ *et* $\|u - P_2(u)\| = d$,
(iii) *si* $u \in C_1$ *et* $P_1(P_2(u)) = u$, *alors* $u \in D_1$,
(iv) *si* $u = \operatorname{argmin}_{x\in D_1}\|x-w\|$ *où* $w \in \mathbb{R}^n$, *alors* $P_2(u) = \operatorname{argmin}_{y\in D_2}\|y-w\|$.

Preuve

(i) Les fonctions $g_i : \mathbb{R}^n \to \mathbb{R}$ définies par $g_i(x) = d(x, C_i)$, $i = 1, 2$ sont convexes et continues. Ainsi, D_1 peut se réécrire $D_1 = \{x \in C_1 : g_2(x) \leq d\}$; c'est à dire un ensemble de niveau d'une fonction convexe continue. Il est donc fermé et convexe. Un raisonnement analogue implique que D_2 est aussi fermé et convexe.

(ii) Si $u \in D_1$, alors $\|u - P_2(u)\| = d(u, C_2) = d$. Or, comme $d(P_2(u), C_1) = d$ alors $P_2(u) \in D_2$. Par conséquent $\|P_1(P_2(u)) - P_2(u)\| = d$ et ainsi $P_1(P_2(u)) = u$ par unicité de la projection orthogonale.

(iii) Soient $x \in C_1, y \in C_2$ arbitraires. De (1.11),

$$\langle P_2(u) - u, x - u\rangle \leq 0$$

et

$$\langle u - P_2(u), y - P_2(u)\rangle = \langle P_1(P_2(u)) - P_2(u), y - P_2(u)\rangle \leq 0.$$

Additionnant les deux inégalités, nous avons

$$\|u - P_2(u)\|^2 \leq \langle u - P_2(u), x - y\rangle \leq \|u - P_2(u)\|\|x - y\|,$$

ou

$$\|u - P_2(u)\| \leq \|x - y\|.$$

Ainsi $\|u - P_2(u)\| = d$ et $u \in D_1$.

(iv) Soit $y \in D_2$. On définit $x = P_1(y)$. Du point (ii) de cette proposition, nous avons $P_2(x) = y$ et $x \in D_1$. Du point (i) de la proposition 1.5.2 appliqué deux fois, nous obtenons

$$\|u - x\| = \|P_1(P_2(u)) - P_1(P_2(x))\| \leq \|P_2(u) - P_2(x)\| \leq \|u - x\|.$$

Et grâce à la proposition 1.5.2 (ii),

$$u - x = P_2(u) - P_2(x) = P_2(u) - y,$$

ainsi,

$$\begin{aligned}\langle w - P_2(u), y - P_2(u)\rangle &= \langle w - P_2(u), x - u\rangle \\ &= \langle w - u, x - u\rangle - \langle P_2(u) - u, x - u\rangle.\end{aligned}$$

Puisque par hypothèse u est la projection orthogonale de w sur D_1 et $x \in D_1$, $\langle w - u, x - u\rangle$ est négatif par (1.11). Et $\langle P_2(u) - u, x - u\rangle$ est nul par le point (ii) de la proposition 1.5.2. Ainsi,

$$\langle w - P_2(u), y - P_2(u)\rangle \leq 0.$$

Par (1.13),

$$\|w - P_2(u)\| \leq \|w - y\|,$$

d'où la thèse.

□

1.5.2 Caractère borné des suites $\{x_i^k\}$ $(i = 1, 2)$

Dans la suite les termes dont le signe de sommation est indéfini (c'est à dire lorsque la limite supérieure est plus petite que la limite inférieure) seront considérés comme nuls.

Proposition 1.5.4 *Soit $k \geq 1$, $s_j \in \mathbb{R}^n$ $(1 \leq j \leq k+1)$, $t_j \in \mathbb{R}^n$ $(1 \leq j \leq k)$, $w \in \mathbb{R}^n$. Alors l'identité suivante a lieu*

$$\begin{aligned}2\Bigg[\sum_{l=1}^{k}\left\langle w - s_1 + \sum_{j=1}^{l-1}(t_j - s_{j+1}), s_{l+1} - s_l\right\rangle + \sum_{l=1}^{k-1}\left\langle \sum_{j=1}^{l}(s_j - t_j), t_{l+1} - t_l\right\rangle \\ + \sum_{l=1}^{k-1}\langle t_k - s_{k+1}, s_{l+1} - s_{k+1}\rangle + \sum_{l=1}^{k-1}\langle s_{k+1} - t_k, t_l - t_k\rangle\Bigg] \\ = \|s_k - s_{k+1}\|^2 + \|w - s_1\|^2 - \|w - s_{k+1}\|^2 \\ + \sum_{l=1}^{k-1}\left(\|s_l - t_l + t_k - s_{k+1}\|^2 + \|t_l - s_{l+1} + s_{k+1} - t_k\|^2\right). \qquad (1.14)\end{aligned}$$

Preuve

Il suffit de développer toutes les sommes et produits internes du membre de gauche et tous les carrés du membre de droite, pour constater l'égalité.

□

Proposition 1.5.5 *Supposons $D_1 \neq \emptyset$. Soit $u \in D_1$. Définissons $v = P_2(u)$ et $\gamma = \|x^0 - u\|^2 - \|x^0 - x_1^1\|^2$. Alors,*

(i) $\|x_1^k - u\|^2 \leq \gamma$ *pour* $k \geq 2$,
(ii) $\sum_{l=1}^{\infty}\|x_1^l - x_2^l + v - u\|^2 \leq \gamma$,
(iii) $\sum_{l=1}^{\infty}\|x_2^l - x_1^{l+1} + u - v\|^2 \leq \gamma$,
(iv) $\sum_{l=1}^{\infty}\langle u - v, x_1^{l+1} - u\rangle \leq \gamma$,
(v) $\sum_{l=1}^{\infty}\langle u - v, v - x_2^l\rangle \leq \gamma$.

Preuve

Nous spécialisons (1.14) pour la suite définie par (1.5)-(1.6), dans le cas où $D_1 \neq \emptyset$. De la proposition 1.5.3(ii), $P_1(v) = u$. Nous appliquons la proposition 1.5.4 avec $k \geq 2$, $w = x^0$, $s_j = x_1^j$ $(1 \leq j \leq k)$, $t_j = x_2^j$ $(1 \leq j \leq k-1)$, $s_{k+1} = u$, $t_k = v$. Utilisant (1.5) et (1.8)-(1.9) et la proposition 1.5.4, (1.14) devient

$$\begin{aligned}
2\Big[&\sum_{l=1}^{k-1}\langle z_1^l - P_1(z_1^l), x_1^{l+1} - P_1(z_1^l)\rangle + \sum_{l=1}^{k-2}\langle z_2^l - P_2(z_2^l), x_2^{l+1} - P_2(z_2^l)\rangle \\
&+ \sum_{l=1}^{k-1}\langle v - P_1(v), x_1^{l+1} - P_1(v)\rangle + \sum_{l=1}^{k-1}\langle u - P_2(u), x_2^l - P_2(u)\rangle \\
&+\langle z_1^k - P_1(z_1^k), u - P_1(z_1^k)\rangle + \langle z_2^{k-1} - P_2(z_2^{k-1}), v - P_2(z_2^{k-1})\rangle\Big] \\
= \;& \|x_1^k - u\|^2 + \|x^0 - x_1^1\|^2 - \|x^0 - u\|^2 \\
&+ \sum_{l=1}^{k-1}(\|x_1^l - x_2^l + v - u\|^2 + \|x_2^l - x_1^{l+1} + u - v\|^2). \qquad (1.15)
\end{aligned}$$

Grâce à la définition de la projection orthogonale, nous réécrivons maintenant (1.15) comme suit

$$\begin{aligned}
&\|x_1^k - u\|^2 + \sum_{l=1}^{k-1}\left(\|x_1^l - x_2^l + v - u\|^2 + \|x_2^l - x_1^{l+1} + u - v\|^2\right) \\
&+2\sum_{l=1}^{k-1}\langle x_1^l - z_1^l, x_1^{l+1} - x_1^l\rangle + 2\sum_{l=1}^{k-2}\langle x_2^l - z_2^l, x_2^{l+1} - x_2^l\rangle \\
&+2\sum_{l=1}^{k-1}\langle u - v, x_1^{l+1} - u\rangle + 2\sum_{l=1}^{k-1}\langle u - v, v - x_2^l\rangle \\
&+2\langle x_1^k - z_1^k, u - x_1^k\rangle + 2\langle x_2^{k-1} - z_2^{k-1}, v - x_2^{k-1})\rangle \\
&= \|x^0 - u\|^2 - \|x^0 - x_1^1\|^2 = \gamma. \qquad (1.16)
\end{aligned}$$

Par (1.11) tous les produits internes du membre de gauche de (1.15) sont négatifs. Ainsi, tous les termes du membre de gauche de (1.16) sont positifs. Cela étant vrai pour tout $k \geq 2$, nous pouvons passer à la limite sur k et les différentes assertions sont alors vérifiées.

□

Corollaire 1.5.1 *Si $D_1 \neq \emptyset$, la suite $\{x_1^k\}$ est bornée.*

Preuve

Le corollaire suit de la proposition 1.5.5(i).

□

Les deux corollaires suivants seront utilisés afin d'étudier la convergence des suites $\{x_i^k\}$ et l'optimalité de leurs limites dans la sous-section suivante.

Corollaire 1.5.2 *Si $D_1 \neq \emptyset$ et x_1^* est la limite d'une sous-suite $\{x_1^{j_k}\}$ de $\{x_1^k\}$, alors $x_1^* \in D_1$ et la sous-suite $\{x_2^{j_k}\}$ converge vers un point $x_2^* \in D_2$ tel que $\|x_1^* - x_2^*\| = d$.*

Preuve

Soit $u \in D_1$. Soit $v = P_2(u)$. De la proposition 1.5.5(ii), $\lim_{k\to\infty} \|x_1^k - x_2^k + v - u\| = 0$, et par suite $\lim_{k\to\infty}(x_1^k - x_2^k) = u - v$ ainsi que $\lim_{k\to\infty}(x_1^{j_k} - x_2^{j_k}) = u - v$. Par conséquent $x_2^{j_k}$ converge. Si x_2^* est sa limite, $x_2^* = x_1^* - u + v$ et $\|x_1^* - x_2^*\| = \|u - v\| = d$. De plus $x_i^* \in C_i$ $(i = 1, 2)$, puique C_i est fermé. Et donc $x_i^* \in D_i$ $(i = 1, 2)$.

□

Corollaire 1.5.3 *Supposons $D_1 \neq \emptyset$. Soit $u \in D_1$. Définissons $v = P_2(u), \gamma = \|x^0 - u\|^2 - \|x^0 - x_1^1\|^2$, $c_k = \langle u - v, x_1^{k+1} - x_2^k - u + v\rangle$. Alors $c_k \geq 0$ et $\sum_{k=1}^{\infty} c_k < \infty$.*

Preuve

Nous réécrivons c_k :

$$\begin{aligned} c_k &= -\langle v - u, x_1^{k+1} - u\rangle - \langle u - v, x_2^k - v\rangle \\ &= -\langle v - P_1(v), x_1^{k+1} - P_1(v)\rangle - \langle u - P_2(u), x_2^k - P_2(u)\rangle \\ &\geq 0, \quad \text{(par (1.11))}. \end{aligned}$$

La convergence de la série est une conséquence des points (iv) et (v) de la proposition 1.5.5.

□

1.5.3 Convergence des suites et optimalité des limites

Dans cette section nous montrons que la suite $\{x_1^k\}$ converge et que sa limite est solution du problème

$$\min_{x \in D_1} \|x - x^0\|. \tag{1.17}$$

Dans le cas linéaire (c'est à dire lorsque les C_i sont des demi-espaces), l'approche que nous allons suivre consiste à prouver une propriété d'optimalité intermédiaire. Cette propriété considère des ensembles D_1^k convergeant vers l'ensemble D_1 et affirme que $\{x_1^k\}$ résout $\min\{\|x - x^0\| : x \in D_1^k\}$. Nous montrerons aussi que la limite de toute sous-suite de $\{x_1^k\}$ résout (1.17). Et comme la solution de (1.17) est unique, la suite $\{x_1^k\}$ converge.

Soient $A \subset \mathbb{R}^n$ et $a \in \mathbb{R}^n$, pour lesquels nous définissons $A + a = \{x + a : x \in A\}$. Avec cette notation, la propriété intermédiaire d'optimalité est la suivante.

Proposition 1.5.6 *Soit $D_1^k = C_1 \bigcap [C_2 + (x_1^k - x_2^{k-1})]$. Alors x_1^k est la solution de* $\min \|x^0 - x\|$ s.c. $x \in D_1^k$.

Preuve

Comme $x_2^{k-1} \in C_2$ (i.e. $0 \in C_2 - x_2^{k-1}$), il suit que $x_1^k \in C_2 + (x_1^k - x_2^{k-1})$, et comme x_1^k est la projection de z_1^k sur C_1, $x_1^k \in D_1^k$. Si nous prenons $x \in D_1^k$ alors, $x \in C_1$ et il existe $y \in C_2$ tel que $x = y + x_1^k - x_2^{k-1}$. Utilisant (1.10), nous trouvons

$$\begin{aligned} \langle x^0 - x_1^k, x - x_1^k \rangle &= \langle z_1^k - x_1^k, x - x_1^k \rangle + \langle z_2^{k-1} - x_2^{k-1}, x - x_1^k \rangle \\ &= \langle z_1^k - x_1^k, x - x_1^k \rangle + \langle z_2^{k-1} - x_2^{k-1}, y - x_2^{k-1} \rangle \\ &\leq 0, \end{aligned}$$

l'inégalité étant obtenue par (1.11). Par (1.13), $\|x - x^0\| \geq \|x_1^k - x^0\| \ \forall x \in D_1^k$, d'où la thèse.

□

Remarquons que si $D_1 \neq \emptyset$, $x_1^k - x_2^{k-1} \to u - v$ lorsque $k \to \infty$, avec $u \in D_1$, $v = P_2(u)$ par la proposition 1.5.5(iii) et il est immédiat que $C_1 \bigcap [C_2 + (u - v)] = D_1$. Cependant dans le cas non linéaire la propriété intermédiaire d'optimalité n'est pas valable à la limite. Ainsi nous devons calculer $\langle x^0 - x_1^k, u - x_1^k \rangle$, pour $u \in D_1$, plus explicitement. Observons que si $u \in D_1$ et $v = P_2(u)$, en utilisant (1.10), nous obtenons l'égalité suivante,

$$\begin{aligned} &\langle x^0 - x_1^{k+1}, u - x_1^{k+1} \rangle \\ &= \langle z_1^{k+1} - x_1^{k+1}, u - x_1^{k+1} \rangle + \langle z_2^k - x_2^k, v - x_2^k \rangle \\ &\quad + \langle z_2^k - x_2^k, (x_2^k - x_1^{k+1}) - (v - u) \rangle. \end{aligned} \tag{1.18}$$

Par (1.11), les deux premiers termes du membre de droite de (1.18) sont négatifs et par la proposition 1.5.5(iii), $x_2^k - x_1^{k+1} \to v - u$ lorsque $k \to \infty$. Si la suite $\{z_2^k\}$ était bornée nous aurions immédiatement que tout point limite du membre de droite de (1.18) serait négatif et alors par la caractérisation de la projection, $\|x^0 - x_1^*\| \leq \|u - x_1^*\|$ pour tout $u \in D_1$ et pour tout point limite x_1^* de $\{x_1^k\}$. Mais lorque $C_1 \bigcap C_2 = \emptyset, \{z_2^k\}$ n'est pas bornée. Ainsi nous devons étudier de plus près le comportement de la suite.

Proposition 1.5.7 *Pour tout $k \geq 2$, nous avons*

$$\|x_1^{k+1} - x_2^k\| \leq \|x_1^k - x_2^k\| \leq \|x_1^k - x_2^{k-1}\|.$$

Preuve

Montrons que l'inégalité gauche est vérifiée.

$$\begin{aligned} &\langle x_2^k - x_1^{k+1}, x_1^k - x_1^{k+1} \rangle \\ &= \langle x_2^k - z_1^{k+1}, x_1^k - x_1^{k+1} \rangle + \langle z_1^{k+1} - x_1^{k+1}, x_1^k - x_1^{k+1} \rangle \\ &= \langle z_1^k - x_1^k, x_1^{k+1} - x_1^k \rangle + \langle z_1^{k+1} - x_1^{k+1}, x_1^k - x_1^{k+1} \rangle \\ &= \langle z_1^k - P_1(z_1^k), x_1^{k+1} - P_1(z_1^k) \rangle + \langle z_1^{k+1} - P_1(z_1^{k+1}), x_1^k - P_1(x_1^{k+1}) \rangle \\ &\leq 0, \end{aligned} \tag{1.19}$$

où la seconde égalité découle de (1.6), la suivante de (1.5) et la dernière inégalité de (1.11). Utilisant ensuite (1.13), nous obtenons finalement,

$$\|x_2^k - x_1^k\| \geq \|x_1^{k+1} - x_2^k\|.$$

L'inégalité droite se vérifie de manière similaire. Nous avons $\langle x_1^k - x_2^k, x_2^{k-1} - x_2^k \rangle \leq 0$ grâce à (1.7), (1.5) et (1.11), et nous obtenons la thèse par (1.13).

□

Proposition 1.5.8 *Supposons que $D_1 \neq \emptyset$. Soit $u \in D_1$. Définissons $v = P_2(u)$, $a_k = \|x_2^k - x_1^{k+1} + u - v\|^2 + 2\langle u - v, x_1^{k+1} - x_2^k - u + v \rangle$.*
Alors la suite $\{a_k\}$ est décroissante.

Preuve

Nous avons,

$$\begin{aligned} a_k &= \|x_2^k - x_1^{k+1}\| + \|u - v\|^2 + 2\langle x_2^k - x_1^{k+1}, u - v \rangle \\ &\quad +2\langle u - v, x_1^{k+1} - x_2^k \rangle - 2\langle u - v, u - v \rangle \\ &= \|x_1^{k+1} - x_2^k\|^2 - \|u - v\|^2. \end{aligned}$$

La thèse suit de la proposition 1.5.7.

□

Maintenant nous avons besoin de deux résultats simples d'analyse. Nous les présentons dans la proposition suivante.

Proposition 1.5.9 *Soit une suite $\{\alpha_k\} \subset \mathbb{R}^+$ telle que $\sum_{k=1}^{\infty} \alpha_k < \infty$. Alors,*
(i) la suite $\{(1/\sqrt{k}) \sum_{j=1}^{k} \sqrt{\alpha_j}\}$ est bornée,
(ii) si $\{\alpha_k\}$ n'est pas croissante alors $\lim_{k \to \infty} k\alpha_k = 0$.

Preuve

Une preuve de cette proposition peut être trouvée en annexe.

□

Proposition 1.5.10 *Supposons que $D_1 \neq \emptyset$. Soient $u \in D_1$ et $v = P_2(u)$. Alors,*

$$\lim_{k \to \infty} \sum_{j=1}^{k} \langle x_1^j - x_2^j, x_1^{k+1} - x_2^k - u + v \rangle = 0.$$

Preuve

Posons

$$b_k = \|x_2^k - x_1^{k+1} + u - v\|^2, \quad c_k = \langle u - v, x_1^{k+1} - x_2^k - u + v\rangle,$$
$$h_k = \|x_1^k - x_2^k - u + v\|^2, \quad \text{et} \quad a_k = b_k + 2c_k.$$

Alors,

$$\begin{aligned}
&\sum_{j=1}^{k}\langle x_1^j - x_2^j, x_1^{k+1} - x_2^k - u + v\rangle \\
&= \sum_{j=1}^{k}\langle x_1^j - x_2^j - u + v, x_1^{k+1} - x_2^k - u + v\rangle \\
&\quad + k\langle u - v, x_1^{k+1} - x_2^k - u + v\rangle \\
&= \sum_{j=1}^{k}\langle x_1^j - x_2^j - u + v, x_1^{k+1} - x_2^k - u + v\rangle + kc_k. \qquad (1.20)
\end{aligned}$$

En prenant les valeurs absolues dans (1.20) et en se souvenant que $c_k \geq 0$ (corollaire 1.5.3), nous arrivons aux inégalités suivantes,

$$\begin{aligned}
&\left|\sum_{j=1}^{k}\langle x_1^j - x_2^j, x_1^{k+1} - x_2^k - u + v\rangle\right| \\
&\leq \sum_{j=1}^{k}|\langle x_1^j - x_2^j - u + v, x_1^{k+1} - x_2^k - u + v\rangle| + kc_k \\
&\leq \|x_1^{k+1} - x_2^k - u + v\| \sum_{j=1}^{k}\|x_1^j - x_2^j - u + v\| + kc_k \\
&= \sqrt{b_k}\sum_{j=1}^{k}\sqrt{h_j} + kc_k \\
&= \sqrt{kb_k}\left(\frac{1}{\sqrt{k}}\sum_{j=1}^{k}\sqrt{h_j}\right) + kc_k. \qquad (1.21)
\end{aligned}$$

Si nous considérons la suite $\{a_k\}$, du corollaire 1.5.3 et de la proposition 1.5.5(iii), nous avons $\sum_{k=1}^{\infty} a_k < \infty$. De la proposition 1.5.8, la suite $\{a_k\}$ n'est pas croissante. Ainsi par la proposition 1.5.9(ii), $0 = \lim_{k\to\infty} ka_k = \lim_{k\to\infty}(kb_k + 2kc_k)$, donc $\lim_{k\to\infty} kb_k = \lim_{k\to\infty} 2kc_k$ Comme $c_k \geq 0$, nous avons $\lim_{k\to\infty} kc_k \geq 0$, de même comme $b_k \geq 0$, nous avons $\lim_{k\to\infty} kb_k \geq 0$, il suit donc que $\lim_{k\to\infty} kc_k = \lim_{k\to\infty} kb_k = 0$. D'où, $\lim_{k\to\infty}\sqrt{kb_k} = 0$. De la proposition 1.5.5(ii), $\sum_{k=1}^{\infty} h_k < \infty$. Ainsi en appliquant la proposition 1.5.9(i), $(1/\sqrt{k})\sum_{j=1}^{k}\sqrt{h_j}$ est bornée. Grâce à tous ces résultats le membre de droite de (1.21) tend vers zéro, la thèse est par conséquent établie.

□

Théorème 1.5.1 *Si $D_1 \neq \emptyset$, alors la suite $\{x_1^k\}$ converge vers x_1^*, la projection orthogonale de x^0 sur D_1 et la suite $\{x_2^k\}$ converge vers $x_2^* = P_2(x_1^*)$, qui est la projection orthogonale de x^0 sur D_2.*

Preuve

Soient $u \in D_1$ et $v = P_2(u)$. Comme dans (1.18),

$$\begin{aligned}
&\langle x^0 - x_1^{k+1}, u - x_1^{k+1}\rangle \\
&= \langle z_1^{k+1} - x_1^{k+1}, u - x_1^{k+1}\rangle + \langle z_2^k - x_2^k, v - x_2^k\rangle \\
&\quad -\langle z_2^k - x_2^k, x_1^{k+1} - x_2^k - u + v\rangle \\
&= \langle z_1^{k+1} - x_1^{k+1}, u - x_1^{k+1}\rangle + \langle z_2^k - x_2^k, v - x_2^k\rangle \\
&\quad -\sum_{j=1}^{k}\langle x_1^j - x_2^j, x_1^{k+1} - x_2^k - u + v\rangle, \qquad (1.22)
\end{aligned}$$

où (1.9) est utilisé dans la dernière égalité. Par (1.11) les deux premiers termes du membre de droite de (1.22) sont négatifs, et par la proposition 1.5.10 le dernier terme tend vers zéro. Ainsi, un point limite de $\{\langle x^0 - x_1^{k+1}, u - x_1^{k+1}\rangle\}$ est négatif. Comme par le corollaire 1.5.1 la suite $\{x_1^k\}$ est bornée, il existe donc une sous-suite $\{x_1^{j_k}\}$ convergente. Soit $u^* = \lim_{k\to\infty} x_1^{j_k}$. Par le corollaire 1.5.2, $u^* \in D_1$. D'où

$$\langle x^0 - u^*, u - u^*\rangle = \lim_{k\to\infty}\langle x^0 - x_1^{j_k}, u - x_1^{j_k}\rangle \le 0, \quad \forall u \in D_1. \qquad (1.23)$$

De (1.23) et de (1.13), pour tout $u \in D_1$, $\|x^0 - u\| \ge \|x^0 - u^*\|$. Ainsi, $u^* = \operatorname{argmin}_{u\in D_1}\|x^0 - u\| = x_1^*$. Puisqu'une telle projection est unique, nous concluons que tous les points limites de $\{x_1^k\}$ sont les mêmes, et la suite converge. L'assertion concernant $\{x_2^k\}$ découle du corollaire 1.5.2 et de la proposition 1.5.3(iv).

□

1.5.4 Nécessité de l'hypothèse $D_1 \neq \emptyset$

A présent, nous montrons que l'hypothèse $D_1 \neq \emptyset$ est non seulement suffisante mais aussi nécessaire pour la convergence des suites $\{x_i^k\}$ $(i = 1, 2)$.

Théorème 1.5.2 *Si $x_i^k \to x_i^*$ lorsque $k \to \infty$, alors $x_i^* \in D_i$ $(i = 1, 2)$.*
En particulier, D_1 et D_2 sont non vides.

Preuve

Soit $\hat{x}_i^k = (1/k)\sum_{j=1}^{k} x_i^j$ $(i = 1, 2)$. Par la convexité de C_i, $\hat{x}_i^k \in C_i$ et $\lim_{k\to\infty}\hat{x}_i^k = \lim_{k\to\infty} x_i^k = x_i^*$. Considérons les points

$$\tilde{x}_1^k = x_1^{k+1} + (1/k)(z_1^{k+1} - x_1^{k+1}), \quad \tilde{x}_2^k = x_2^k + (1/k)(z_2^k - x_2^k).$$

Puisque $\tilde{x}_2^k$ appartient à la demi-droite partant de x_2^k et passant par z_2^k, $P_2(\tilde{x}_2^k) = x_2^k$. D'une manière similaire, $P_1(\tilde{x}_1^k) = x_1^{k+1}$. Utilisant (1.8) et (1.9), nous avons

$$\begin{aligned}
\tilde{x}_1^k &= x_1^{k+1} + \frac{1}{k}\left[x^0 + \sum_{j=1}^{k}(x_2^j - x_1^j) - x_1^{k+1}\right] \\
&= x_1^{k+1} + \frac{1}{k}x^0 + (\hat{x}_2^k - \hat{x}_1^k) - \frac{1}{k}x_1^{k+1}
\end{aligned}$$

$$
\begin{aligned}
&\rightarrow \quad x_1^* + x_2^* - x_1^* = x_2^* \quad \text{quand } k \rightarrow \infty, \\
\tilde{x}_2^k &= x_2^k + \frac{1}{k}\left[x_2^k + \sum_{j=1}^{k}(x_1^j - x_2^j) - x_2^k\right] \\
&= x_2^k + \hat{x}_1^k - \hat{x}_2^k \\
&\rightarrow \quad x_2^* + x_1^* - x_2^* = x_1^* \quad \text{quand } k \rightarrow \infty.
\end{aligned}
$$

Par la continuité de P_i, nous obtenons

$$P_1(x_2^*) = P_1\left(\lim_{k\to\infty} \tilde{x}_1^k\right) = \lim_{k\to\infty} P_1(\tilde{x}_1^k) = \lim_{k\to\infty} x_1^{k+1} = x_1^*,$$

$$P_2(x_1^*) = P_2\left(\lim_{k\to\infty} \tilde{x}_2^k\right) = \lim_{k\to\infty} P_2(\tilde{x}_2^k) = \lim_{k\to\infty} x_2^k = x_2^*.$$

Dès lors, $P_1(P_2(x_1^*)) = x_1^*$. Par la proposition 1.5.3(iii), $x_1^* \in D_1$, et par le point (ii) de cette même proposition, $x_2^* \in D_2$.

□

Remarque: cette analyse de la convergence (pour laquelle l'intérieur de C peut être vide) peut être étendue, moyennant quelques complications techniques, au cas de m ensembles convexes fermés $C_1, \ldots, C_m$. Il suffit de définir $D_i = \{x \in C_i : d(x, C_j) = d(C_i, C_j), (1 \leq j \leq m)\}$ et la condition suffisante pour la convergence de la suite $\{x_i^k\}$ est $D_i \neq \emptyset$ $(1 \leq i \leq m)$, sous laquelle elle converge vers la projection orthogonale de x^0 sur D_i.

Cependant pour la version simultanée de l'algorithme (donnée à la section suivante), le cas $m = 2$ est suffisant, nous le montrons dans la suite.

1.6 Version simultanée de l'algorithme de Han

1.6.1 La méthode

L'algorithme que l'on propose ici est une application de la méthode de Han. Il est appellé algorithme parallèle ou simultané et consiste à calculer le même type de projection simultanément sur chaque ensemble, mais à prendre comme nouvel itéré une combinaison convexe de ces projections.

ALGORITHME PARALLELE

Initialisation

- $\lambda_i \in \mathbb{R}^+$, $1 \leq i \leq m$ tels que
$$\sum_{i=1}^{m} \lambda_i = 1, \tag{1.24}$$
- x^0 un point donné,
- $z_i^1 = x^0$, $1 \leq i \leq m$.

Itération

Pour $k = 1, 2, \ldots$

- $x^k = \sum_{i=1}^{m} \lambda_i P_i(z_i^k)$, (1.25)
- $z_i^{k+1} = x^k + z_i^k - P_i(z_i^k),\ 1 \leq i \leq m.$ (1.26)

L'avantage principal de cet algorithme parallèle par rapport à celui de Han est qu'il peut être facilement implémenté en utilisant des processeurs parallèles qui calculent les projections orthogonales $P_i(z_i^k)$ indépendemment.

1.6.2 Exemple

Nous rappellons le problème (P1) considéré pour l'algorithme de Han :

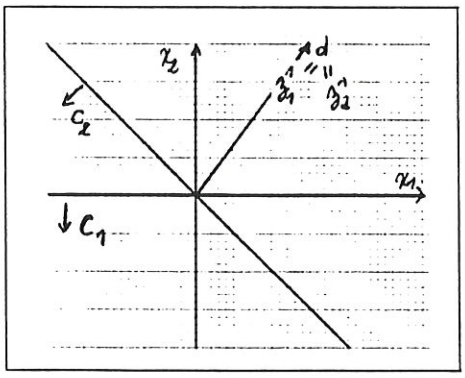
$$(\text{P1}) \begin{cases} \min & \|x - d\|^2 \\ \text{s.c.} & x \in C_i, \quad i = 1, 2 \end{cases}$$

où $C_1 = \{(x_1, x_2) \mid x_2 \leq 0\}$,
$C_2 = \{(x_1, x_2) \mid x_1 + x_2 \leq 0\}$,
d un point donné.

Pour la version simultanée, nous obtenons l'interprétation géométrique suivante pour les deux premières itérations.

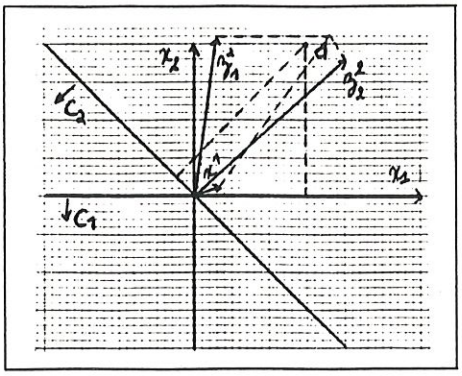
1) Situation initiale

$$\begin{cases} d \text{ donné} \\ z_1^1 = z_2^1 = d \\ \lambda_1 = \frac{1}{4}, \lambda_2 = \frac{3}{4} \end{cases}$$



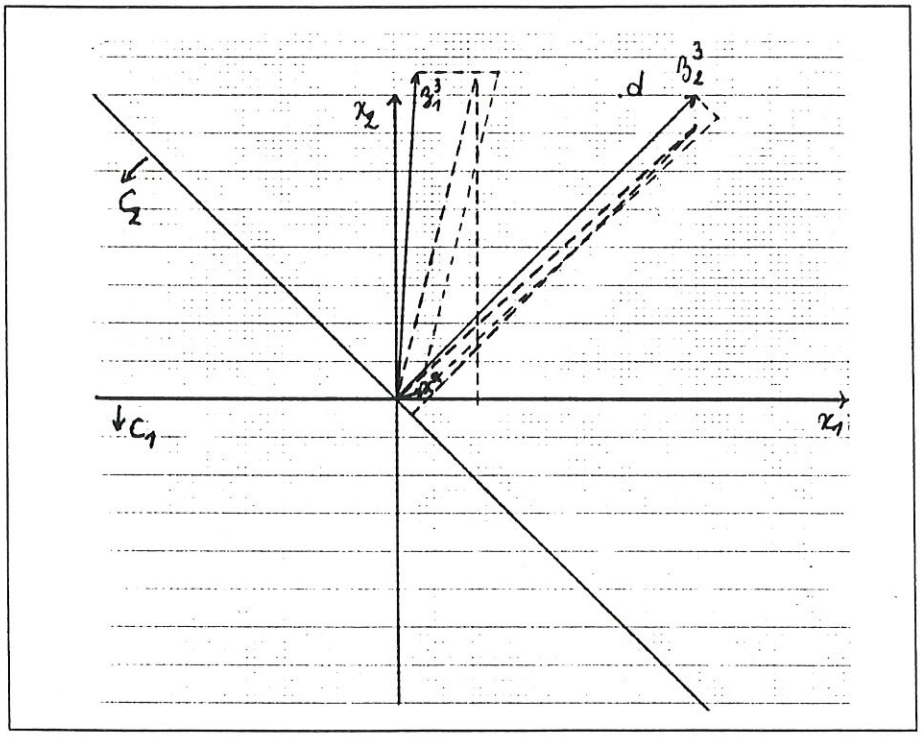
2) Première itération : $k = 1$

$$\begin{cases} x^1 = \frac{1}{4} P_1(z_1^1) + \frac{3}{4} P_2(z_2^1) \\ z_1^2 = x^1 + z_1^1 - P_1(z_1^1) \\ z_2^2 = x^1 + z_2^1 - P_2(z_2^1) \end{cases}$$



3) Deuxième itération : $k = 2$

$$\begin{cases} x^2 = \frac{1}{4}P_1(z_1^2) + \frac{3}{4}P_2(z_2^2) \\ z_1^3 = x^2 + z_1^2 - P_1(z_1^2) \\ z_2^3 = x^2 + z_2^2 - P_2(z_2^2) \end{cases}$$



Nous avons doublé l'échelle du dernier schéma pour le rendre plus clair.

1.7 Analyse de la convergence de la version simultanée

Nous prouvons que la convergence de la version simultanée peut être réduite à la convergence de la méthode de Han pour deux ensembles.

Les suites $\{x^k\}$, $\{z_i^k\}$ se réfèrent à l'algorithme défini par (1.25)-(1.26), et P_i est défini comme précédemment. Définissons $f : \mathbb{R}^n \to \mathbb{R}$ comme $f(x) = \sum_{i=1}^m \lambda_i \|x - P_i(x)\|^2$. Soit G l'ensemble des minima globaux de f. Nous allons montrer que la suite $\{x^k\}$ converge

vers $x^* = \text{argmin}_{x\in G}\|x^0 - x\|$. Nous verrons que si $C = \bigcap_{i=1}^m C_i \neq \emptyset$, alors $G = C$ et x^* est une solution de (P).

Nous travaillons dans $\mathbb{R}^{n\times m}$ et représentons $x \in \mathbb{R}^{n\times m}$ par $x = (x_1, \ldots, x_m), x_i \in \mathbb{R}^n$ $(1 \leq i \leq m)$. Nous considérons dans $\mathbb{R}^{n\times m}$ le produit interne défini par

$$\langle x, y\rangle_{\bar{Q}} = \sum_{i=1}^m \lambda_i \langle x_i, y_i\rangle, \tag{1.27}$$

et les ensembles $\bar{C}_1, \bar{C}_2$ définis par

$$\bar{C}_1 = \{x \in \mathbb{R}^{n\times m} : x_i = x_{i+1} (1 \leq i \leq m-1)\}, \tag{1.28}$$

$$\bar{C}_2 = C_1 \times \ldots \times C_m. \tag{1.29}$$

Soit

$$\bar{x}^0 = (x^0, \ldots, x^0) \in \mathbb{R}^{n\times m}. \tag{1.30}$$

Nous montrons que la suite générée par (1.25)-(1.26) est équivalente à celle générée par (1.5)-(1.7) utilisant pour cela $\bar{x}^0, \bar{C}_1, \bar{C}_2$ et le produit interne $\langle ., .\rangle_{\bar{Q}}$, ainsi le théorème de convergence pour (1.25)-(1.26) sera une conséquence du théorème 1.5.2.

Nous commençons par donner quelques propriétés élémentaires de $\bar{P}_1, \bar{P}_2$ définies respectivement comme les projections orthogonales relativement à (1.27), de $\mathbb{R}^{n\times m}$ sur $\bar{C}_1, \bar{C}_2$.

Proposition 1.7.1 *Pour tout $x \in \mathbb{R}^{n\times m}$, nous avons*
(i) $[\bar{P}_1(x)]_i = \sum_{l=1}^m \lambda_l x_l$ $(1 \leq i \leq m)$,
(ii) $[\bar{P}_2(x)]_i = P_i(x_i)$ $(1 \leq i \leq m)$.

Preuve

(i) Si $\bar{y} \in \bar{C}_1$, alors $\bar{y} = (y, \ldots, y)$, $y \in \mathbb{R}^n$. Soit $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_m)$ défini comme $\bar{x}_i = \sum_{l=1}^m \lambda_l x_l$ $(1 \leq i \leq m)$. Alors $\bar{x} \in \bar{C}_1$ et

$$\begin{aligned}
\langle x - \bar{x}, \bar{y} - \bar{x}\rangle_{\bar{Q}} &= \sum_{i=1}^m \lambda_i \langle x_i - \bar{x}_i, y - \bar{x}_i\rangle \\
&= \sum_{i=1}^m \lambda_i \langle x_i - \sum_{l=1}^m \lambda_l x_l, y - \sum_{l=1}^m \lambda_l x_l\rangle \\
&= \langle \sum_{i=1}^m \lambda_i x_i - \sum_{l=1}^m \lambda_l x_l, y - \sum_{l=1}^m \lambda_l x_l\rangle = 0,
\end{aligned}$$

où (1.24) est utilisée dans la troisième égalité. Ainsi, par (1.13), $\|x - \bar{y}\|_{\bar{Q}} \geq \|x - \bar{x}\|_{\bar{Q}}$, ce qui implique que

$$\bar{x} = \text{argmin}_{\bar{y}\in\bar{C}_1}\|x - \bar{y}\|_{\bar{Q}} \stackrel{def}{=} \bar{P}_1(x).$$

L'assertion (i) est par conséquent établie.

(ii) Soit $\bar{x} \in \bar{C}_2$ tel que $\bar{x}_i = P_i(x_i)$, soit $y \in \bar{C}_2$, nous obtenons par (1.11),

$$\langle x - \bar{x}, y - \bar{x}\rangle_{\bar{Q}} = \sum_{i=1}^m \lambda_i \langle x_i - P_i(x_i), y_i - P_i(x_i)\rangle \leq 0,$$

Par conséquent, par (1.13), $\|x-y\|_{\bar{Q}} \geq \|x-\bar{x}\|_{\bar{Q}}$, ce qui implique

$$\bar{x} = \text{argmin}_{y\in\bar{C}_2}\|x-y\|_{\bar{Q}} \stackrel{def}{=} \bar{P}_2(x).$$

L'assertion (ii) est donc établie.

□

Nous considérons maintenant les suites $\{\bar{x}_i^k\}$, $\{\bar{z}_i^k\}$ $(i=1,2)$ générées par (1.5)-(1.7) utilisant $\bar{x}^0, \bar{C}_1, \bar{C}_2$ et $\langle .,. \rangle_{\bar{Q}}$ comme définis en (1.27)-(1.30).

Les propositions suivantes établissent les relations entre les suites $\{x^k\}$, $\{z_i^k\}$ générées par (1.25)-(1.26) et les suites $\{\bar{x}_i^k\}$, $\{\bar{z}_i^k\}$. Nous utilisons encore la convention selon laquelle une somme est nulle si l'indice supérieur de sommation est plus petit que l'indice inférieur.

Proposition 1.7.2 *Pour $k \geq 1$ et $1 \leq i \leq m$, les assertions suivantes ont lieu*
(i) $(\bar{z}_2^k)_i = z_i^k$,
(ii) $(\bar{x}_2^k)_i = P_i(z_i^k)$,
(iii) $(\bar{z}_1^{k+1})_i = P_i(z_i^k) + \sum_{j=1}^{k-1}[P_i(z_i^j) - x^j]$,
(iv) $(\bar{x}_1^{k+1})_i = x^k$.

Preuve

Une preuve détaillée de cette proposition peut être trouvée en annexe.

□

Soit $\bar{D}_1 = \{x \in \bar{C}_1 : \bar{d}(x,\bar{C}_2) = \bar{d}(\bar{C}_1,\bar{C}_2)\}$, où $\bar{d}$ est la distance définie à partir du produit interne relatif à $\bar{Q}$ $\langle .,. \rangle_{\bar{Q}}$. La proposition 1.7.2 signifie que la suite $\{x^k\}$ converge si et seulement si la suite $\{\bar{x}_1^k\}$ converge et cela se produit, en accord avec les théorèmes 1.5.1 et 1.5.2, si et seulement si $\bar{D}_1 \neq \emptyset$. Afin de présenter les théorèmes de convergence pour l'algorithme simultané nous devons ramener l'hypothèse $\bar{D}_1 \neq \emptyset$ à l'espace original.

Soit $P : \mathbb{R}^n \to \mathbb{R}$ défini par $P(x) = \sum_{i=1}^m \lambda_i P_i(x)$. Soit F l'ensemble des points fixes de P, c'est à dire $F = \{x \in \mathbb{R}^n : P(x) = x\}$.

Proposition 1.7.3 $\bar{D}_1 = \{\bar{u} \in \mathbb{R}^{n\times m} : \bar{u} = (u,\ldots,u) \text{ avec } u \in F\}$.

Preuve

De la proposition 1.5.3(ii) et (iii), $\bar{u} \in \bar{D}_1$ si et seulement si $\bar{P}_1(\bar{P}_2(\bar{u})) = \bar{u}$. De la proposition 1.7.1,

$$\bar{u}_i = [\bar{P}_1(\bar{P}_2(\bar{u}))]_i = \sum_{l=1}^m \lambda_l [\bar{P}_2(\bar{u})]_l = \sum_{l=1}^m \lambda_l P_l(\bar{u}_l).$$

Ainsi,

$$\bar{u} \in \bar{D}_1 \quad \Leftrightarrow \quad \sum_{l=1}^m \lambda_l P_l(\bar{u}_l) = \bar{u}_i. \tag{1.31}$$

Comme le membre de gauche de l'égalité de (1.31) ne dépend pas de i, nous concluons que

$$\bar{u} \in \bar{D}_1 \quad \Leftrightarrow \quad \bar{u} = (u,\ldots,u) \text{ et } \sum_{l=1}^m \lambda_l P_l(u) = u.$$

D'où la thèse.

□

Théorème 1.7.1 *Si $F \neq \emptyset$, la suite $\{x^k\}$ générée par (1.25)- (1.26) converge vers la solution de* $\min \|x^0 - x\|$ *s.c.* $x \in F$. *De plus, si la suite $\{x^k\}$ converge, alors $F \neq \emptyset$.*

Preuve

Si $F \neq \emptyset$, par la proposition 1.7.3, $\bar{D}_1 \neq \emptyset$. Par le théorème 1.5.1, la suite $\{\bar{x}_1^k\}$ converge. Et par la proposition 1.7.2(iv) la suite $\{x^k\}$ converge. Si $x^k \to (x^*, \ldots, x^*) = \bar{x}^*$ lorsque $k \to \infty$. Par le théorème 1.5.1, $\bar{x}^*$ minimise $\|\bar{x}^0 - \bar{x}\|_{\bar{Q}}^2$ *s.c.* $\bar{x} \in \bar{D}_1$, c'est à dire

$$\begin{aligned} \bar{x}^* &= \text{argmin}_{\bar{x} \in \bar{D}_1} \sum_{i=1}^{m} \lambda_i \|\bar{x}_i^0 - \bar{x}_i\|^2 \\ &= \text{argmin}_{x \in F} \sum_{i=1}^{m} \lambda_i \|x^0 - x\|^2 \\ &= \text{argmin}_{x \in F} \|x^0 - x\|^2, \end{aligned}$$

où nous avons utilisé respectivement (1.30), la proposition 1.7.3 et (1.24) dans chaque égalité.

Pour la seconde assertion, si x^k converge vers $\bar{x}^*$, par la proposition 1.7.2(iv) $(\bar{x}_1^{k+1})_i$ converge vers $\bar{x}^*$, par le théorème 1.5.2 $\bar{x}^* \in \bar{D}_1$ et par la proposition 1.7.3 $\bar{x}^* \in \mathbb{R}^{n \times m}$ tel que $\bar{x}^* = (x^*, \ldots, x^*)$ et $x^* \in F$, et donc $F \neq \emptyset$.

□

Maintenant, nous allons donner une autre caractéristique de l'ensemble F. Définissons $f : \mathbb{R}^n \to \mathbb{R}$ comme

$$f(x) = \sum_{i=1}^{m} \lambda_i \|P_i(x) - x\|^2.$$

Soit G l'ensemble des minima globaux de f.

Proposition 1.7.4 *Nous avons, $F = G$.*

Preuve

Une preuve de cette proposition peut être trouvée dans [2] (théorème 1).

□

Observons que si $C = \bigcap_{i=1}^{m} C_i \neq \emptyset$, alors $C = G$ parce que $f(x) \geq 0$ et $f(x) = 0$ si et seulement si $P_i(x) = x$ $(1 \leq i \leq m)$ si et seulement si $x \in C_i$ $(1 \leq i \leq m)$ si et seulement si $x \in C$

Remarquons aussi que lorsque $C = \emptyset$, G peut être interprété comme l'ensemble des solutions des moindres carrés pondérées (par les λ_i) du problème convexe admissible. Ainsi, dans ce cas-là ($G \neq \emptyset$) la suite converge vers les solutions des moindres carrés pondérées les plus proches de x^0 en norme Q. Nous résumons cette assertion dans le théorème suivant:

Théorème 1.7.2 *Si* $G \neq \emptyset$, *alors* $\lim_{k\to\infty} x^k = x^*$, *lequel résout* $\min \|x - x^0\|^2$ *s.c.* $x \in \operatorname{argmin}_{y\in\mathbb{R}^n}\{\sum_{i=1}^m \lambda_i \|P_i(y)-y\|^2\}$. *Si* $C = \bigcap_{i=1}^m C_i \neq \emptyset$, *alors* x^* *résout* $\min \|x-x^0\|^2$ *s.c.* $x \in C$.

Preuve

Immédiate par le théorème 1.7.1 et la proposition 1.7.4.

□

1.8 Conditions suffisantes pour la convergence

Comme la condition $G \neq \emptyset$ n'est pas facile à vérifier, il est intéressant de fournir quelques conditions suffisantes qui l'impliquent. Nous donnons ici un théorème regroupant ces conditions.

Théorème 1.8.1 *Si une des conditions suivantes à lieu, la suite* $\{x^k\}$ *générée par (1.25)-(1.26) converge.*

(i) $C = \bigcap_{i=1}^m C_i \neq \emptyset$,

(ii) *il existe* C_i *borné,*

(iii) *les* C_i *sont des polyèdres,*

(iv) $\exists v_i \in C_i$ *s.c.* $\|v_i - v_j\| = d(C_i, C_j)$ $(1 \leq i, j \leq m)$.

Preuve

(i) Découle du théorème 1.7.2 et de la remarque qui le précède.

(ii) Si C_i est borné pour un certain i, alors l'image de P_i est bornée, et nous obtenons

$$\lim_{\|x\|\to\infty} \lambda_i \|P_i(x) - x\| = \infty,$$

et donc $\lim_{\|x\|\to\infty} f(x) = \infty$. Il suit que pour $\alpha \in \mathbb{R}$, $N_\alpha = \{x : f(x) \leq \alpha\}$ est compact (s'il est non vide). Soit $\alpha \in \mathbb{R}$ tel que $N_\alpha \neq \emptyset$, alors

$$G = \{\operatorname{argmin}_{x\in\mathbb{R}^n} f(x)\} = \{\operatorname{argmin}_{x\in N_\alpha} f(x)\} \neq \emptyset$$

par continuité de f et compacité de N_α. La thèse suit du théorème 1.7.2.

(iii) Nous pouvons voir que la projection orthogonale sur un polyèdre est une "pièce" affine sur le polyèdre, c'est à dire qu'il existe un recouvrement de $\mathbb{R}^n$ par un nombre fini de polyèdres fermés tel qu'il coïncide dans chacun d'eux avec une fonction affine. Les fonctions affines dans ce cas-ci sont les projections orthogonales sur le contour affine des faces de dimension j du polyèdre $(1 \leq j \leq m-1)$ et l'identité. Ainsi $\|P_i(x) - x\|^2$ est une pièce quadratique sur le polyèdre et est donc $f(x)$. Par le théorème de Frank-Wolfe (Corollaire 27.3.1 de [1]) f atteint son minimum sur chaque polyèdre du recouvrement. Comme leur nombre est fini, f atteint son minimum global. Par le théorème 1.7.2 le résultat est établi.

(iv) Nous avons, $\|v_i - v_j\| = d(C_i, C_j)$ implique $P_i(v_j) = v_i$ $(1 \leq i, j \leq m)$. Soit $x \in C_i$, $\{y \in \mathbb{R}^n : P_i(y) = x\}$ est convexe. Soit $v = \sum_{i=1}^{m} \lambda_i v_i$. Alors,

$$\begin{aligned} P_i(v) = v_i &\Rightarrow \sum_{i=1}^{m} \lambda_i P_i(v) = \sum_{i=1}^{m} \lambda_i v_i = v \\ &\Rightarrow P(v) = v \\ &\Rightarrow v \in F \\ &\Rightarrow F \neq \emptyset \end{aligned}$$

et $G \neq \emptyset$ par la proposition 1.7.4. la thèse suit du théorème 1.7.2.

□

Chapitre 2

La méthode primale de Iusem: une méthode primale de type "row-action"

2.1 Introduction

Nous présentons dans ce chapitre une méthode de type "row-action" pour minimiser une fonction strictement convexe sous des contraintes convexes générales. Celle-ci procède de la manière suivante: les contraintes sont utilisées une à là fois, aucun changement n'est effectué sur les fonctions contraintes, ni sur leur matrice jacobienne, et à chaque itération un sous-problème est résolu consistant en la minimisation de la fonction objectif sous contrainte d'une ou deux équations linéaires.

Nous établissons la convergence de l'algorithme et nous comparons la nouvelle méthode avec celle de Han.

Nous commençons par définir la nouvelle méthode, après quoi nous posons le problème et donnons une description géométrique de l'algorithme primal. Nous présentons ensuite la méthode sous une forme assez générale, et donnons quelques propriétés élémentaires de convergence de celle-ci. Une présentation plus explicite et une variante du sous-problème à résoudre à chaque itération sont fournies respectivement dans les sections 2.3.4 et 2.3.5. Ensuite nous discutons à propos d'un choix particulier de S_k et établissons des preuves de convergence pour un tel choix. Dans la section 2.4.1 nous présentons une version calculable (implémentable) de la méthode. Nous également deux stratégies pour le choix de la contrainte utilisée à chaque itération. Dans la section 2.4.3 nous proposons une version simultanée de l'agorithme de base pour laquelle toutes les contraintes sont combinées à chaque itération. Et pour terminer nous comparons l'algorithme S avec la méthode de Han.

2.2 La méthode

2.2.1 Définition de la méthode de type "row-action"

Une méthode itérative de type "row-action" est une méthode ayant les propriétés suivantes:

1) A chaque itération une seule contrainte est utilisée.

2) Chaque itération calcule un n-vecteur x^k en utilisant seulement x^{k-1} (et peut-être quelques n-vecteurs auxiliaires).

3) Aucune opération n'est effectuée sur la matrice jacobienne des contraintes (on garde sa structure).

2.2.2 Position du problème

Considèrons le problème suivant

$$(\text{PP}) \begin{cases} \min f(x) \\ \text{s.c. } g_i(x) \leq 0 \ (1 \leq i \leq m), \end{cases}$$

où $f, g_i : \mathbb{R}^n \to \mathbb{R}$, $1 \leq i \leq m$. Soient

$$C_i = \{x \in \mathbb{R}^n : g_i(x) \leq 0\} \quad \text{et} \quad C = \bigcap_{i=1}^{m} C_i.$$

Sur les contraintes, nous faisons les hypothèses suivantes,

(A1) g_i est convexe $(1 \leq i \leq m)$.
(A2) g_i est continue $(1 \leq i \leq m)$.
(A3) $C \neq \emptyset$.
(A4) L'ensemble des sous gradients de g_i $(1 \leq i \leq m)$ est uniformément borné sur des sous-ensembles compacts de $\mathbb{R}^n$.

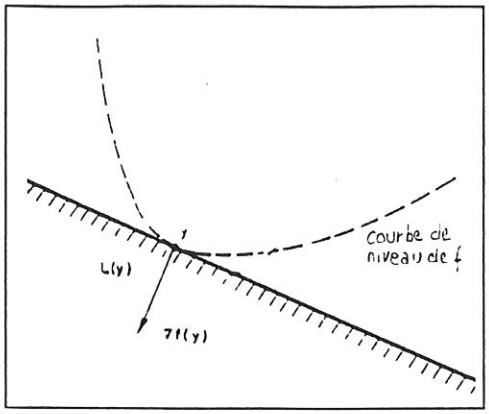
Nous faisons aussi quelques hypothèses générales sur f afin de simplifier l'analyse de convergence,

(B1) f est strictement convexe.
(B2) f est différentiable.
(B3) Les ensembles de niveau de f sont bornés.

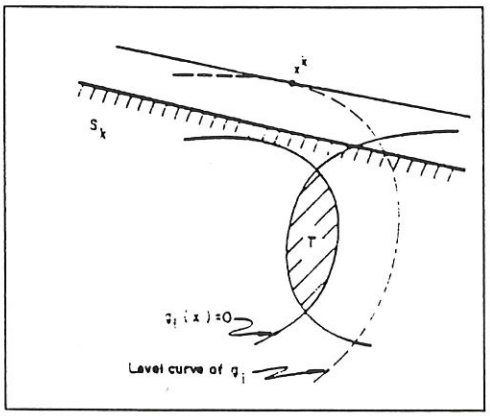
L'hypothèse (B1) est essentielle pour l'analyse de la méthode, tandis que (B2) n'est pas nécessaire et est ajoutée seulement pour simplifier les notations. En effet si f est continue mais non différentiable, il suffit de remplacer le gradient de f en x par un élément du sous gradient de f en x. Cela ne modifiera pas les résultats de convergence. L'hypothèse (B3) est utile car elle assure que f atteint un minimum (unique grâce à (B1)) sur tout sous-ensemble convexe et fermé de $\mathbb{R}^n$.

2.2.3 Description géométrique de l'algorithme primal

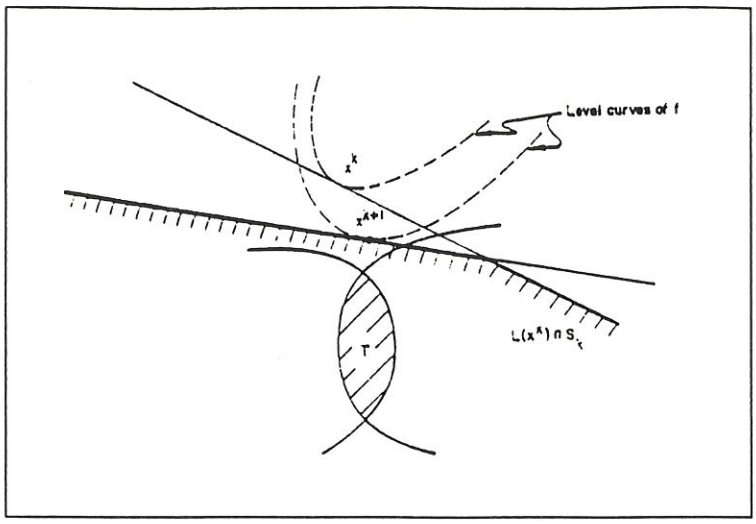
Soit $y \in \mathbb{R}^n$. Définissons $L(y) = \{x \in \mathbb{R}^n : \nabla f(y)^t(x - y) \geq 0\}$. Si nous prenons l'hyperplan tangent à l'ensemble de niveau de f en y, alors $L(y)$ est le demi-espace dont la frontière est cet hyperplan et $L(y)$ ne contient pas l'ensemble de niveaux (sauf si y est le minimum de f sans contrainte, dans quel cas $L(y) = \mathbb{R}^n$).



L'algorithme procède comme suit. A l'itération k, le $k$ème itéré x^k est donné et nous construisons un demi-espace $S_k = \{x \in \mathbb{R}^n : a^t x \leq \beta\}$ où $a \in \mathbb{R}^n$, $\beta \in \mathbb{R}$. Cet espace contient l'ensemble admissible mais pas x^k, c'est à dire que la frontière de S_k est un hyperplan (appelé H_k) séparant x^k de l'ensemble admissible.



L'itéré suivant x^{k+1} sera le minimum de f sur $L(x^k) \bigcap S_k$. Il peut être prouvé par récurrence que si la solution du problème appartient à $L(x^0)$, alors elle appartient à $L(x^k)$ pour tout k.



Et si x^k est admissible alors c'est la solution du problème (notée x^*).

2.3 L'algorithme de base

2.3.1 Le problème (PP')

Si nous définissons $g(x) = \max_{1 \leq i \leq m} g_i(x)$, le problème (PP) se réduit au problème (PP') suivant

$$(\mathrm{PP}') \begin{cases} \min f(x) \\ \text{s.c. } g(x) \leq 0 \end{cases}$$

où g satisfait les hypothèses (A1)-(A4), tant que les g_i les satisfont. Le problème (PP') n'est en général pas approprié pour les calculs réels mais il est pratique pour présenter une formulation générale de la méthode et pour prouver certains résultats élémentaires de convergence.

2.3.2 L'algorithme de base pour le problème (PP')

Soit x^* l'unique solution du problème (PP') (l'existence suit des hypothèses (A2), (A3), (B3) et la continuité de f via (B2), tandis que l'unicité découle de (A1) et (B1)). Pour chaque $x \in \mathbb{R}^n$ nous déterminons

$$L(x) \stackrel{def}{=} \{y \in \mathbb{R}^n : \nabla f(x)^t (y - x) \geq 0\}. \tag{2.1}$$

Remarquons que $L(x)$ est le plus grand ensemble convexe fermé tel que x est le minimum de f sur cet ensemble.

Définissons aussi

$$U \stackrel{def}{=} \{x \in \mathbb{R}^n : x^* \in L(x)\}. \tag{2.2}$$

Notons que x appartient à U si et seulement si, il existe un ensemble convexe fermé contenant x^* tel que x est le minimum de f sur cet ensemble.

Considérons aussi une suite d'ensembles $S_k \subset \mathbb{R}^n$ tel que S_k est soit $\mathbb{R}^n$, soit un demi-espace satisfaisant, pour tout $k \geq 0$,

$$C \subset S_k. \tag{2.3}$$

L'algorithme de base généré par une suite $\{x^k\} \subset \mathbb{R}^n$ est

Initialisation

$$x^0 \in U \tag{2.4}$$

Itération

$$x^{k+1} = \text{argmin } f(x) \quad \text{s.c. } x \in L(x^k) \bigcap S_k. \tag{2.5}$$

Nous allons montrer dans la proposition 2.3.2 que $L(x^k) \bigcap S_k$ est non vide pour tout k. Dans un tel cas cet ensemble est convexe et fermé, et (B1)-(B3) garantissent l'existence et l'unicité de x^{k+1}. En regardant l'initialisation (2.4), nous pouvons toujours choisir x^0 comme le minimum sans contrainte de f dont l'existence et l'unicité découlent des hypothèses (B1)-(B3). Pour un tel choix nous avons $\nabla f(x^0) = 0$ ainsi $L(x^0) = \mathbb{R}^n$ et (2.4) a évidemment lieu.

Bien sûr nous ne pouvons espérer prouver la convergence de la suite définie par (2.4), (2.5) vers x^*, à moins que d'imposer certaines conditions sur S_k en plus de (2.3). Cependant dans la suite nous montrerons que (2.3)-(2.5) sont suffisantes pour établir plusieurs propriétés importantes de convergence.

2.3.3 Propriétés élémentaires de convergence

Nous définissons tout d'abord la convergence faible.

Définition 2.3.1 Une suite $\{x^k\}$ est *faiblement convergente* vers la solution d'un certain problème si

(i) la suite est bornée,
(ii) $\lim_{k\to\infty}(x^{k+1} - x^k) = 0$,
(iii) tout point limite de $\{x^k\}$ est une solution du problème.

Nous prouvons ci-après que (i) et (ii) ont lieu pour la suite $\{x^k\}$ définie par (2.4), (2.5) et que si un point limite de $\{x^k\}$ appartient à C alors il est égal à la solution x^*.

A partir d'ici, à moins que ce ne soit spécifié, la suite $\{x^k\}$ désigne la suite générée par (2.4), (2.5). Commençons avec un résultat élémentaire concernant cette suite.

Proposition 2.3.1 *Pour tout $x \in \mathbb{R}^n$, nous avons*

$$x = \text{argmin}_{y \in L(x)} f(y).$$

Preuve

Soit $y \in L(x)$, par la convexité de f nous avons,

$$0 \leq \nabla f(x)^t(y-x) \leq f(y) - f(x). \tag{2.6}$$

Il découle donc que $f(x) \leq f(y)\ \forall y \in L(x)$.

□

Proposition 2.3.2 *Pour tout* $k \geq 0$,
(i) $x^* \in L(x^k)$,
(ii) $L(x^k) \bigcap S_k \neq \emptyset$.

Preuve

(i) Nous procédons par récurrence. L'assertion (i) est vraie pour $k = 0$ par (2.4). Supposons que ce soit vrai pour k, i.e. $x^* \in L(x^k)$. Montrons que cela reste vrai pour x^{k+1}.

Comme x^{k+1} minimise f sur l'ensemble convexe $L(x^k) \bigcap S_k$, nous avons la condition d'optimalité équivalente,

$$\nabla f(x^{k+1})^t(y - x^{k+1}) \geq 0$$

pour tout $y \in L(x^k) \bigcap S_k$. Or $x^* \in C \subset S_k$ par définition et $x^* \in L(x^k)$ par hypothèse de récurrence, nous pouvons donc remplacer y par x^* dans l'inégalité ci-dessus pour obtenir

$$\nabla f(x^{k+1})^t(x^* - x^{k+1}) \geq 0$$

et par conséquent $x^* \in L(x^{k+1})$.

(ii) Par définition $x^* \in C \subset S_k$, de plus par (i) $x^* \in L(x^k)$. Donc $L(x^k) \bigcap S_k \neq \emptyset$.

□

Corollaire 2.3.1 *Pour tout* $k \geq 0$,
(i) $f(x^k) \leq f(x^{k+1}) \leq f(x^*)$,
(ii) *si* $x^k \neq x^{k+1}$ *alors* $f(x^k) < f(x^{k+1})$.

Preuve

(i) Comme par définition $x^{k+1} \in L(x^k)$, l'inégalité $f(x^k) \leq f(x^{k+1})$ découle de la proposition 2.3.1. Par la preuve de la proposition 2.3.2(ii) $x^* \in L(x^{k+1}) \bigcap S_{k+1}$, et donc en utilisant (2.5) nous obtenons $f(x^{k+1}) \leq f(x^*)$. La première assertion est par conséquent établie.

(ii) Comme $x^{k+1} \in L(x^k)$, (2.6) a lieu avec $x = x^k$, $y = x^{k+1}$. La stricte convexité de f implique que l'inégalité droite dans (2.6) est stricte sauf si $x = y$. La seconde assertion est donc vérifiée.

□

Corollaire 2.3.2 *La suite* $\{x^k\}$ *est bornée.*

Preuve

Par le corollaire 2.3.1(i), $\{x^k\} \subset \{x \in \mathbb{R}^n : f(x) \leq f(x^*)\}$, or cet ensemble est borné par (B3), donc la suite $\{x^k\}$ est bornée.

□

Corollaire 2.3.3 *Nous avons,*
(i) soit $\tilde{x}$ *un point limite de* $\{x^k\}$, *si* $\tilde{x} \in C$ *alors* $\tilde{x} = x^*$,
(ii) si pour certains $k \geq 0$, $x^k \in C$ *alors* $x^k = x^*$.

Preuve

(i) Du corollaire 2.3.1(i) et de la continuité de f, $f(\tilde{x}) \leq f(x^*)$. Si $\tilde{x} \in C$, alors par définition de $\tilde{x}$, $f(x^*) \leq f(\tilde{x})$ impliquant $f(\tilde{x}) = f(x^*) \leq f(x)$ pour tout $x \in C$. Ainsi $\tilde{x}$ est une solution de (PP') et par conséquent $\tilde{x} = x^*$ par unicité de x^*.

(ii) Raisonnement analogue.

□

Le corollaire 2.3.3(ii) indique que si $x^k \in C$ pour k quelconque alors les calculs devraient s'arrêter. Dans un tel cas l'algorithme produit $x^j = x^k = x^*$ pour tout $j \geq k$.

Proposition 2.3.3 $\lim_{k\to\infty}(x^{k+1} - x^k) = 0$.

Preuve

Par l'absurde. Supposons que le résultat n'ait pas lieu, alors il existe $\delta > 0$ et une sous-suite $\{x^{l_k}\}$ de $\{x^k\}$ telle que $\|x^{l_k+1} - x^{l_k}\| \geq \delta$. Utilisant le corollaire 2.3.2 choisissons une sous-suite $\{x^{j_k}\}$ de $\{x^{l_k}\}$ telle que $\{x^{j_k}\}$ et $\{x^{j_k+1}\}$ convergent vers $\tilde{x}$ et $\bar{x}$. Alors $\|\tilde{x} - \bar{x}\| \geq \delta$ implique que $\tilde{x} \neq \bar{x}$. Soient

$$z^k = \frac{1}{2}(x^{j_k+1} + x^{j_k})$$

et

$$z = \lim_{k\to\infty} z^k = \frac{1}{2}(\bar{x} + \tilde{x}).$$

Du corollaire 2.3.1(i) la suite $\{f(x^k)\}$ est croissante et bornée et donc convergente. Notons sa limite γ. Par conséquent, $f(\bar{x}) = \lim_{k\to\infty} f(x^{j_k+1}) = \lim_{k\to\infty} f(x^{j_k}) = f(\tilde{x}) = \gamma$. Par la stricte convexité de f, $f(z) < \gamma$. D'où pour k assez grand,

$$f(z^k) < f(x^{j_k}), \tag{2.7}$$

$x^{j_k} \in L(x^{j_k})$ par définition de $L(x)$ et $x^{j_k+1} \in L(x^{j_k})$ par (2.5). Ainsi $z^k \in L(x^{j_k})$ par la convexité de $L(x^{j_k})$. D'autre part, grâce à la définition de $L(x)$ et à la convexité de f nous avons que

$$0 \leq \nabla f(x^{j_k})^t(z^k - x^{j_k}) \leq f(z^k) - f(x^{j_k}), \tag{2.8}$$

ce qui contredit (2.7). Le résultat est donc établi.

□

Corollaire 2.3.4 *Soit* $d(x^k, S_k) = \inf_{x \in S_k} \|x^k - x\|$, *nous avons*

$$\lim_{k \to \infty} d(x^k, S_k) = 0.$$

Preuve

Comme $x^{k+1} \in S_k$ par définition, nous avons $0 \leq d(x^k, S_k) \leq \|x^k - x^{k+1}\|$. La thèse suit de la proposition 2.3.3.

□

2.3.4 Reformulation du sous-problème

$L(x^k)$ et S_k sont tous deux soit des demi-espaces soit l'espace tout entier ($L(x^k) = \mathbb{R}^n$ si et seulement si x^k est le minimum sans contraintes de f). Montrons que le problème min $f(x)$ s.c. $x \in L(x^k) \bigcap S_k$ se réduit à la minimisation de f sur un hyperplan ou sur l'intersection de deux hyperplans. Soit

$$G_k = \{x \in \mathbb{R}^n : \nabla f(x^k)^t (x - x^k) = 0\} \tag{2.9}$$

et soit H_k défini de la manière suivante,

$$H_k = \begin{cases} \mathbb{R}^n & \text{si } S_k = \mathbb{R}^n \\ \{x \in \mathbb{R}^n : a^t x = \beta\} & \text{si } S_k = \{x \in \mathbb{R}^n : a^t x \leq \beta\}. \end{cases} \tag{2.10}$$

où $a \in \mathbb{R}^n, \beta \in \mathbb{R}$. Définissons aussi,

$$z^1 = \text{argmin}_{x \in H_k} f(x), \tag{2.11}$$

$$z^2 = \text{argmin}_{x \in G_k \bigcap H_k} f(x). \tag{2.12}$$

Proposition 2.3.4 *Soit* $\{x^k\}$ *la suite définie par (2.4)-(2.5),*

$$\begin{array}{ll} (i) & \textit{si } x^k \in S_k \\ & \textit{alors } x^{k+1} = x^k, \\ (ii) & \textit{si } x^k \notin S_k \\ & \textit{et si } z^1 \in L(x^k) \\ & \textit{alors } x^{k+1} = z^1, \\ (iii) & \textit{si } x^k \notin S_k \\ & \textit{et si } z^1 \notin L(x^k) \\ & \textit{alors } x^{k+1} = z^2. \end{array} \tag{2.13}$$

Preuve

(i) Comme $x^k \in L(x^k)$, si $x^k \in S_k$ nous avons $x^k \in L(x^k) \bigcap S_k$ ainsi $f(x^{k+1}) \leq f(x^k)$ par (2.5). Il suit du corollaire 2.3.1(i) que $f(x^{k+1}) = f(x^k)$ et du corollaire 2.3.1(ii) que $x^{k+1} = x^k$.

(ii) $x^k \notin S_k$ implique que $x^{k+1} \in H_k$.

Sinon x^{k+1} est à l'intérieur de S_k impliquant que la contrainte $x \in S_k$ dans (2.5) est redondante, i.e., par la convexité du problème

$$x^{k+1} = \text{argmin}_{x \in L(x^k) \bigcap S_k} f(x) = \text{argmin}_{x \in L(x^k)} f(x) = x^k$$

où la dernière égalité découle de la proposition 2.3.1. Donc $x^k = x^{k+1} \in S_k$, ce qui contredit l'hypothèse, et l'implication est par conséquent correcte.

Comme $x^{k+1} \in H_k$, par définition de z^1,

$$f(z^1) \leq f(x^{k+1}).$$

D'autre part, par hypothèse $z^1 \in L(x^k)$ impliquant $z^1 \in L(x^k) \bigcap H_k \subset L(x^k) \bigcap S_k$ ainsi nous obtenons

$$f(z^1) \geq f(x^{k+1})$$

par (2.5). Dès lors, $f(z^1) = f(x^{k+1})$ et $z^1 = x^{k+1}$ par unicité du minimum dans (2.5).

(iii) Dans la preuve de (ii) nous avons vu que $x^{k+1} \in H_k$ parce que $x^k \notin S_k$. De (2.5), (2.12), il suffit alors de montrer que $x^{k+1} \in G_k$. Supposons que $x^{k+1} \notin G_k$ alors x^{k+1} est dans l'intérieur de $L(x^k)$ et la contrainte $x \in L(x^k)$ dans (2.5) est redondante i.e.,

$$x^{k+1} = \text{argmin}_{x \in S_k} f(x),$$

or $x^{k+1} \in H_k \subset S_k$, donc

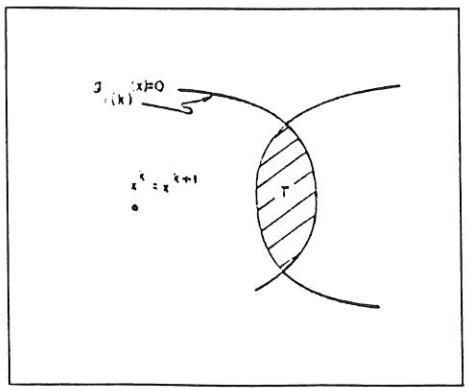
$$x^{k+1} = \text{argmin}_{x \in H_k} f(x) = z^1.$$

Comme $x^{k+1} \in L(x^k)$, il suit que $z^1 \in L(x^k)$, ce qui contredit l'hypothèse.

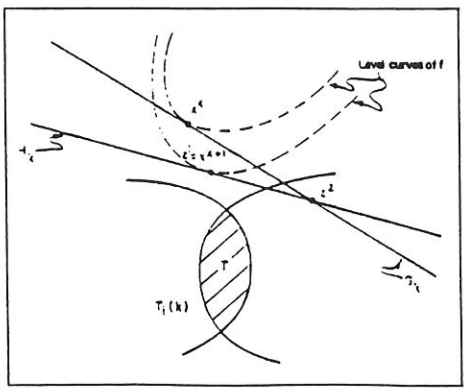
□

Les cas (i), (ii) et (iii) de la proposition 2.3.4 sont illustrés dans les figures suivantes.

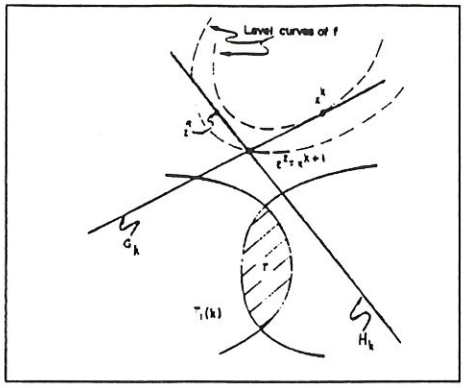
(i)



(ii)



(iii)



2.3.5 Une variante de l'algorithme de base

Comme nous l'avons vu dans la proposition 2.3.4, (2.5) peut être réécrit comme

$$x^{k+1} = \begin{cases} x^k \text{ si } x^k \in S_k, \\ z^1 \text{ si } x^k \notin S_k,\ z^1 \in L(x^k), \\ z^2 \text{ si } x^k \notin S_k,\ z^1 \notin L(x^k), \end{cases} \tag{2.14}$$

avec z^1, z^2 comme défini par (2.11), (2.12). Nous proposons de remplacer (2.13) par

$$f(z^1) \geq f(x^k), \tag{2.15}$$

i.e. de remplacer (2.14) par

$$x^{k+1} = \begin{cases} x^k \text{ si } x^k \in S_k, \\ z^1 \text{ si } x^k \notin S_k,\ f(z^1) \geq f(x^k), \\ z^2 \text{ si } x^k \notin S_k,\ f(z^1) < f(x^k). \end{cases} \tag{2.16}$$

Comme (2.13) est équivalent à $0 \leq \nabla f(x^k)^t(z^1 - x^k)$, il suit de la convexité de f que (2.13) implique (2.15), mais la réciproque n'est pas vraie. Dans des cas où (2.14) nécessite la minimisation de f sur l'intersection de deux hyperplans, l'avantage de (2.16) est que la minimisation de f s'effectuera sur un seul hyperplan. Les calculs seront donc plus simples.

Les résultats de la section 2.3.3 restent valables pour (2.16). A l'itération k, appellons $\hat{x}^{k+1}$ le point produit par (2.16) et x^{k+1} le résultat de (2.14). Si $x^{k+1} \neq \hat{x}^{k+1}$ alors $\hat{x}^{k+1} = z^1$ et $x^{k+1} = z^2$. Comme $f(z^1) \leq f(z^2)$ par définition de z^1 et z^2, nous obtenons de (2.15) et du corollaire 2.3.1(i),

$$f(x^k) \leq f(\hat{x}^{k+1}) \leq f(x^{k+1}) \leq f(x^*). \tag{2.17}$$

Il suit de (2.17) que le corollaire 2.3.1 est satisfait pour la suite générée avec (2.16) au lieu de (2.5), et tous les autres résultats après le corollaire 2.3.1 sont également satisfaits pour cette variante de l'algorithme.

Cependant (2.16) a une faiblesse qui se voit dans (2.17): si nous mesurons la progression de l'algorithme avec une suite croissante $\{f(x^k)\}$, (2.17) montre que, quand (2.5) et (2.16) diffèrent, (2.16) fournit un itéré moins bon que (2.5).

Seules des expériences numériques peuvent décider si la minimisation plus simple de (2.16) compense la croissance plus grande des valeurs fonctionnelles.

2.3.6 Un choix spécifique de S_k

Nous proposons ici un choix particulier de S_k pour lequel la convergence de la suite $\{x^k\}$ est établie. Soit $\epsilon > 0$ et une suite de paramètres de relaxation $\{\alpha_k\} \subset \mathbb{R}$ tels que

$$0 < \epsilon \leq \alpha_k \leq 1. \tag{2.18}$$

Choisissons $0 \neq v^k \in \partial g(x^k)$ et définissons

$$S_k = \{x \in \mathbb{R}^n : (v^k)^t(x - x^k) + \alpha_k g(x^k) \leq 0\}. \tag{2.19}$$

Nous pouvons supposer $v^k \neq 0$, car si $\partial g(x^k) = \{0\}$ alors x^k est un minimum de g impliquant $x^k \in C$ et par conséquent grâce au corollaire 2.3.3(ii), $x^k = x^*$. Supposons donc $\partial g(x^k) > 0$ pour tout k, dans quel cas par (2.18), (2.19) et la définition du sous-gradient, nous avons

$$(v^k)^t(x - x^k) + \alpha_k g(x^k) \leq (v^k)^t(x - x^k) + g(x^k) \leq g(x);$$

ainsi, si $x \in C$ alors $x \in S_k$, ce qui signifie que (2.3) est satisfait.

Proposition 2.3.5 *Si S_k est donné par (2.19) alors la suite $\{x^k\}$ définie par (2.4), (2.5) (ou par (2.4), (2.16)) est convergente vers x^*.*

Preuve

Il suffit, au vu des corollaires 2.3.2 et 2.3.3, de montrer que tout point limite de $\{x^k\}$ appartient à C, dans quel cas tous sont égaux à x^* et $x^* = \lim_{k\to\infty} x^k$.

Si $x^k \in S_k$ alors $d(x^k, S_k) = 0$ et sinon par (A4),

$$d(x^k, S_k) = \frac{\alpha_k}{\|v^k\|} g(x^k) \geq (\epsilon/\eta) g(x^k) \geq 0,$$

où η est la borne du sous-différentiel sur l'ensemble compact $\{x \in \mathbb{R}^n : f(x) \leq f(x^*)\}$, utilisant le corollaire 2.3.1(i) et (B3). Ainsi,

$$d(x^k, S_k) \geq (\epsilon/\eta) \max\{0, g(x^k)\}.$$

Du corollaire 2.3.4, $\lim_{k\to\infty} \max\{0, g(x^k)\} = 0$. Si $\tilde{x}$ est un point limite de $\{x^k\}$, nous concluons de (A2) que

$$\max\{0, g(\tilde{x})\} = 0, \quad \text{i.e. } g(\tilde{x}) \leq 0.$$

Par conséquent, $\tilde{x} \in C$.

□

2.4 L'algorithme pour le problème (PP)

Dans cette section nous revenons au problème

$$(\text{PP}) \begin{cases} \min & f(x) \\ \text{s.c.} & g_i(x) \leq 0 \ (1 \leq i \leq m). \end{cases}$$

Nous présentons un choix possible d'ensembles S_k tenant compte de chaque contrainte séparément (à la manière des méthodes de type "row-action").

2.4.1 Une version séquentielle de l'algorithme pour le problème (PP)

Nous présentons une formulation adaptée au problème (PP) plutôt qu'au problème (PP').

Soit $\{i(k)\}$ une suite de nombres entre 1 et m. $i(k)$ indique l'indice de la contrainte utilisée à la $k$ème itération pour définir S_k. Soit $\epsilon > 0$ et une suite de paramètres de relaxation $\{\alpha_k\} \subset \mathbb{R}$ tels que

$$0 < \epsilon \leq \alpha_k \leq 1. \tag{2.20}$$

Définissons alors S_k comme

$$S_k = \begin{cases} \mathbb{R}^n & \text{si } g_{i(k)}(x^k) \leq 0, \\ \{x \in \mathbb{R}^n : (v^k)^t(x - x^k) + \alpha_k g_{i(k)}(x^k) \leq 0\} & \text{sinon,} \end{cases} \tag{2.21}$$

où $0 \neq v^k \in \partial g_{i(k)}(x^k)$. Comme dans la section 2.3.6, il suit de la définition du sous-gradient que $C_{i(k)} \subset S_k$. Puisque $C \subset C_i$ pour tout i, $C \subset S_k$ est toujours satisfait et les résultats de la section 2.3.3 ont lieu pour ce choix-ci de S_k. Il découle de (2.21) que $x^k \in S_k$ si et seulement si $x^k \in C_{i(k)}$. Ainsi lorsque $x^k \notin C_{i(k)}$,

$$H_k = \{x \in \mathbb{R}^n : (v^k)^t(x - x^k) + \alpha_k g_{i(k)}(x^k) = 0\}.$$

Réécrivons l'algorithme en utilisant (2.14) (ou (2.16)) au lieu de (2.5) et en substituant (2.11) et (2.12) par les conditions d'optimalité du premier ordre, qui sont suffisantes par convexité. Cette reformulation-ci de l'algorithme de base sera appellée Algorithme (Primal) S.

ALGORITHME PRIMAL S

Initialisation

- Soit $x^0 \in \mathbb{R}^n$ tel que

$$\nabla f(x^0)^t(x^* - x^0) \geq 0, \tag{2.22}$$

(pour l'instant, $x^0 = \operatorname{argmin}_{x \in \mathbb{R}^n} f(x)$).

Itération

- Soit donné x^k, $i(k)$, α_k.
- Si $g_{i(k)}(x^k) \leq 0$,
  alors $x^{k+1} = x^k$,
  sinon (i.e. $g_{i(k)}(x^k) > 0$),
  prendre v^k tel que $v^k \neq 0$, $v^k \in \partial g_{i(k)}(x^k)$ et
  trouver $x \in \mathbb{R}^n$ et $\lambda \in \mathbb{R}$ tels que

$$\begin{cases} \nabla f(x) = \lambda v^k, \\ (v^k)^t(x - x^k) + \alpha_k g_{i(k)}(x^k) = 0, \end{cases} \tag{2.23}$$

  si $\nabla f(x^k)^t(x - x^k) \geq 0$ (ou si $f(x) \geq f(x^k)$),

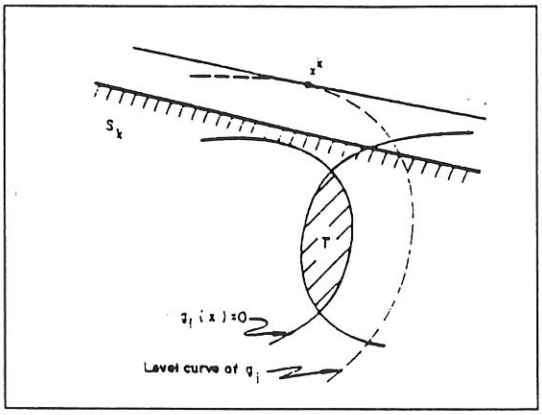
alors $x^{k+1} = x$,
sinon,
trouver $x \in \mathbb{R}^n$ et $\lambda, \mu \in \mathbb{R}$ tels que

$$\begin{cases} \nabla f(x) = \lambda v^k + \mu \nabla f(x^k), \\ (v^k)^t(x - x^k) + \alpha_k g_{i(k)}(x^k) = 0, \\ \nabla f(x^k)^t(x - x^k) = 0, \end{cases} \tag{2.24}$$

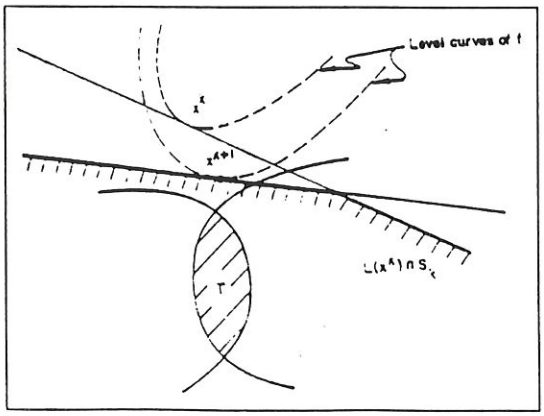
et prendre $x^{k+1} = x$.

Remarquons que (2.23) est un système de $n+1$ équations à $n+1$ inconnues et que (2.24) est un système de $n+2$ équations à $n+2$ inconnues.

Afin de donner une interprétation géométrique de l'algorithme, nous choisissons de prendre S_k comme décrit ci-dessus mais sans paramètre de relaxation, et nous supposons que les g_i sont différentiables. A l'itération k, nous choisissons une contrainte, la $i$ème. Si x^k satisfait cette contrainte, alors nous prenons $x^{k+1} = x^k$, sinon nous considérons $H_k = \{x \in \mathbb{R}^n : \nabla g_i(x^k)^t(x^k - x) = g_i(x^k)\}$. H_k est parallèle à l'hyperplan tangent à la courbe de niveau de g_i en x^k et se situe entre x^k et l'ensemble admissible de la $i$ème contrainte. S_k est donc le demi-espace qui ne contient pas x^k et dont la frontière est H_k.



C'est à dire $S_k = \{x \in \mathbb{R}^n : \nabla g_i(x^k)^t(x^k - x) \geq g_i(x^k)\}$ et x^{k+1} est le minimum de f sur $L(x^k) \bigcap S_k$.



Afin de montrer que cet algorithme génère une suite convergeant vers x^* il est nécessaire de spécifier la suite dite de contrôle $\{i(k)\}$. En effet, une suite arbitraire ne fonctionnera pas car si on prend $i(k) = l$ pour tout k, $\{x^k\}$ convergera en général vers la solution de $\min f(x)$ s.c. $g_l(x) \leq 0$ et non vers x^*.

2.4.2 Stratégies pour le choix de la contrainte à utiliser à chaque itération

Le contrôle de contrainte la plus violée

Considérons l'Algorithme S où $i(k)$ est défini comme

$$i(k) = \text{argmax}_{1 \leq l \leq m}\{g_l(x^k)\}. \tag{2.25}$$

Théorème 2.4.1 *La suite $\{x^k\}$ générée par l'Algorithme S avec $i(k)$ comme dans (2.25) converge vers la solution x^* du problème* (PP).

Preuve

Appellons $g(x) = \max_{1 \leq i \leq m} g_i(x)$, alors cet algorithme-ci est précisément la méthode de la section 2.3.6 pour le problème (PP'). Au vu de (2.19), (2.21), il suffit de montrer que

$$\partial g_{i(k)}(x^k) \subset \partial g(x^k), \tag{2.26}$$

ainsi le résultat suivra directement de la proposition 2.3.5. Montrons donc que l'inclusion (2.26) est satisfaite. Soit $v \in \partial g_{i(k)}(x^k)$. Pour tout $y \in \mathbb{R}^n$, nous avons

$$v^t(y - x^k) \leq g_{i(k)}(y) - g_{i(k)}(x^k) = g_{i(k)}(y) - g(x^k) \leq g(y) - g(x^k), \tag{2.27}$$

où la première inégalité découle de la définition de $\partial g_{i(k)}(x^k)$, l'égalité suit de (2.25) et la seconde inégalité de la définition de g. Finalement des extrémités de (2.27), nous obtenons $v \in \partial g(x^k)$.

□

La suite de contrôle presque cyclique

Définition 2.4.1 Soit une suite $\{i(k)\}$ telle que il existe $r \geq m$ ayant la propriété suivante

$$\forall l\ 1 \leq l \leq m \text{ et } \forall k\ \exists j\ k \leq j \leq k + r \text{ satisfaisant } i(j) = l.$$

Une telle suite est appelée *suite presque cyclique* et r est *la constante de presque cyclicité.*

Cette définition signifie que chaque contrainte est employée au moins une fois dans un bloc de r itérations consécutives.

Définition 2.4.2 Lorsque $r = m$ la suite est *cyclique.*

Théorème 2.4.2 *La suite $\{x^k\}$ générée par l'Algorithme S avec une suite de contrôle presque cyclique $\{i(k)\}$ converge vers la solution x^* du problème* (PP).

Preuve

Au vu des corollaires 2.3.2 et 2.3.3, si nous montrons que chaque point limite de $\{x^k\}$ appartient à C nous concluons que x^* est le seul point limite et $\lim_{k\to\infty} x^k = x^*$.

Soit une sous-suite convergente $\{x^{j_k}\}$ de $\{x^k\}$. Soit $\tilde{x} = \lim_{k\to\infty} x^{j_k}$. Soit p entre 1 et m. Soit $l_k = \min\{l > j_k : i(l) = p\}$. Du fait que $\{i(k)\}$ est presque cyclique, nous avons $0 \leq l_k - j_k \leq r$. Par conséquent,

$$\|x^{l_k} - x^{j_k}\| \leq \sum_{j=j_k}^{l_k-1} \|x^{j+1} - x^j\| \leq \sum_{j=j_k}^{j_k+r} \|x^{j+1} - x^j\|. \tag{2.28}$$

De (2.28) et de la proposition 2.3.3, $\lim_{k\to\infty} \|x^{l_k} - x^{j_k}\| = 0$, impliquant $\lim_{k\to\infty} x^{l_k} = \tilde{x}$. De la définition de S_k, si $g_{i(l_k)}(x^{l_k}) \leq 0$ alors $d(x^{l_k}, S_{l_k}) = 0$ et sinon de (2.20) et (A4),

$$d(x^{l_k}, S_{l_k}) = \frac{\alpha_{l_k}}{\|v^{l_k}\|} g_{i(l_k)}(x^{l_k}) = \frac{\alpha_{l_k}}{\|v^{l_k}\|} g_p(x^{l_k}) \geq \frac{\epsilon}{\eta} g_p(x^{l_k}) > 0,$$

où η est la borne du sous-différentiel sur l'ensemble compact $\{x \in \mathbb{R}^n : f(x) \leq f(x^*)\}$. Il suit que

$$d(x^{l_k}, S_{l_k}) \geq (\epsilon/\eta) \max\{0, g_p(x^{l_k})\}.$$

Du corollaire 2.3.4, $0 = \lim_{k\to\infty} \max\{0, g_p(x^{l_k})\} = \max\{0, g_p(\tilde{x})\}$ impliquant $g_p(\tilde{x}) \leq 0$. Comme p est arbitraire, on obtient $g_i(\tilde{x}) \leq 0$ pour tout i, c'est à dire $\tilde{x} \in C$.

□

2.4.3 Une version simultanée de l'algorithme pour le problème (PP)

Dans cette section nous présentons une implémentation simultanée de l'algorithme de base dans laquelle toutes les contraintes sont combinées à chaque itération. Nous considérons donc un choix de S_k tenant compte simultanément de toutes les contraintes du problème (PP).

Soient des nombres réels $\epsilon, \bar{\epsilon} > 0$, α_k satisfaisant $0 < \epsilon \leq \alpha_k \leq 1$ et ξ_{ik} $(1 \leq i \leq m,\ k \geq 0)$ tels que, pour tout i, k,

$$0 < \bar{\epsilon} < \xi_{ik}, \tag{2.29}$$

et pour tout k,

$$\sum_{i=1}^{m} \xi_{ik} = 1. \tag{2.30}$$

Soit $I_k = \{i : g_i(x^k) > 0\}$, c'est à dire l'ensemble des indices des contraintes non vérifiées en x^k. Par le corollaire 2.3.3(ii), nous pouvons supposer $I_k \neq \emptyset$ pour tout k, car lorsque $I_k = \emptyset$ nous obtenons $x^k = x^*$. Pour $i \in I_k$, prenons $0 \leq v^{ik} \in \partial g_i(x^k)$ et définissons

$$h^k = \sum_{i\in I_k} \xi_{ik} v^{ik}, \tag{2.31}$$

$$\theta_k = \sum_{i\in I_k} \xi_{ik} g_i(x^k). \tag{2.32}$$

Alors les ensembles S_k sont donnés par

$$S_k = \{x \in \mathbb{R}^n : (h^k)^t (x - x^k) + \alpha_k \theta_k \leq 0\}. \tag{2.33}$$

Théorème 2.4.3 *La suite $\{x^k\}$ générée par (2.4), (2.5) (ou par (2.4), (2.16)) avec S_k comme dans (2.33) converge vers la solution x^* du problème* (PP).

Preuve

D'abord vérifions que (2.33) satisfait (2.3).
Soit $x \in C$. Par la définition de I_k, (2.20), (2.29) et la définition du sous-gradient, nous avons

$$\begin{aligned}(h^k)^t(x-x^k)+\alpha_k\theta_k &= \sum_{i\in I_k}\xi_{ik}[(v^{ik})^t(x-x^k)+\alpha_k g_i(x^k)] \\ &\leq \sum_{i\in I_k}\xi_{ik}[(v^{ik})^t(x-x^k)+g_i(x^k)] \\ &\leq \sum_{i\in I_k}\xi_{ik}g_i(x)\leq 0, \end{aligned} \tag{2.34}$$

donc $x \in S_k$.

Grâce aux corollaires 2.3.2 et 2.3.3 il suffit de vérifier que tout point limite $\tilde{x}$ de $\{x^k\}$ appartient à C. Il découle de (2.34) et de la définition de S_k que $d(x^k, S_k) = \alpha_k\theta_k/\|h^k\|$, et

$$\begin{aligned} d(x^k,S_k) &= \frac{\alpha_k\theta_k}{\|\sum_{i\in I_k}\xi_{ik}v^{ik}\|} \\ &\geq \frac{\alpha_k\theta_k}{\eta\sum_{i\in I_k}\xi_{ik}} \\ &\geq \frac{\alpha_k\theta_k}{\eta} \\ &\geq \frac{\epsilon}{\eta}\sum_{i\in I_k}\xi_{ik}g_i(x^k) \\ &\geq \frac{\epsilon\bar{\epsilon}}{\eta}\sum_{i\in I_k}g_i(x^k) \\ &\geq \frac{\epsilon\bar{\epsilon}}{\eta}\sum_{i=1}^{m}\max\{0,g_i(x^k)\}\geq 0 \end{aligned} \tag{2.35}$$

en utilisant (A4) (η est la borne du sous-différentiel sur l'ensemble compact $\{x \in \mathbb{R}^n : f(x) \leq f(x^*)\}$) et la convexité de la norme dans la première inégalité, (2.30) dans la seconde, (2.32) dans la troisième, (2.29) dans la quatrième.

Par le corollaire 2.3.4 et (2.35), nous avons $\lim_{k\to\infty}\max\{0, g_i(x^k)\} = 0$ pour tout i, ainsi que $g_i(\tilde{x}) \leq 0$ pour tout i et tout point limite $\tilde{x}$. Par conséquent, tout point limite $\tilde{x}$ de $\{x^k\}$ appartient à C.

□

Substituant v^k, $g_{i(k)}(x^k)$ par h^k, θ_k respectivement dans (2.23), (2.24) nous obtenons une version simultanée de l'algorithme de base pour le problème (PP).

L'avantage principal de la version simultanée est la plus grande liberté dans la sélection des poids ξ_{ik} attribués aux contraintes. Ce qui contraste avec la version séquentielle où seulement le paramètre de relaxation α_k peut être arbitrairement choisi.

2.5 Comparaison avec la méthode de Han

Rappelons d'abord que la méthode de Han considère une fonction objectif quadratique et des contraintes convexes générales.

Plaçons-nous donc dans ces conditions.

Supposons

$$f(x) = \frac{1}{2}(x-u)^t M(x-u) \tag{2.36}$$

où $M \in \mathbb{R}^{n\times n}$ est définie positive, symétrique et $u \in \mathbb{R}^n$. Soit $v^k \in \partial g_{i(k)}(x^k)$ et définissons

$$\sigma = (v^k)^t M^{-1} v^k, \tag{2.37}$$

$$\rho = (v^k)^t (x^k - u), \tag{2.38}$$

$$\pi = (x^k - u)^t M (x^k - u) = 2f(x^k). \tag{2.39}$$

Alors l'algorithme S se réécrit comme suit

Initialisation

$x^0 = u$ (le minimum sans contrainte de f).

Itération

La solution de (2.23) est

$$x = u + \frac{1}{\sigma}(\rho - \alpha_k g_{i(k)}(x^k)) M^{-1} v^k, \tag{2.40}$$

et la solution de (2.24) est

$$x = u + \lambda M^{-1} v^k + \mu(x^k - u), \tag{2.41}$$

où λ, μ résolvent le système linéaire suivant

$$\sigma\lambda + \rho\mu = \rho - \alpha_k g_{i(k)}(x^k), \tag{2.42}$$

$$\rho\lambda + \pi\mu = \pi, \tag{2.43}$$

dont la solution est facile.

Ensuite réécrivons l'algorithme de Han. Celui-ci génère $m+1$ suites de n-vecteurs $\{x^k\}$ et $\{y_i^k\}$ $(1 \le i \le m)$ de la manière suivante

Initialisation

$$\begin{aligned} &x^0 = u, \\ &y_i^0 = 0 \ (1 \le i \le m). \end{aligned} \tag{2.44}$$

Itération

Soit donné x^k, y_i^k $(1 \leq i \leq m)$ et $i(k)$,

$$x^{k+1} = \begin{cases} \min & \frac{1}{2}(x^k + y_{i(k)}^k - x)^t M(x^k + y_{i(k)}^k - x) \\ \text{s.c.} & g_{i(k)}(x) \leq 0 \end{cases} \tag{2.45}$$

et

$$y_i^{k+1} = \begin{cases} y_i^k & \text{si } i \neq i(k) \\ y_i^k + x^{k+1} - x^k & \text{si } i = i(k) \end{cases} \tag{2.46}$$

Comparons maintenant l'algorithme S (2.40)-(2.43) avec celui de Han (2.44)-(2.46). Deux observations sont immédiates.

D'abord (2.44)-(2.46) garde en mémoire d'une itération à l'autre non seulement l'itéré courant x^k mais aussi une matrice $m \times n$ dont les lignes sont les vecteurs duaux associés aux contraintes, tandis que l'algorithme S mémorise seulement le n-vecteur x^k. Comme la matrice dont les lignes sont y_i^k sera généralement dense, la méthode de Han deviendra impraticable pour des grandes valeurs de m et n.

Deuxièmement, le sous-problème (2.45) est un problème assez difficile en soi, il n'a pas de solution explicite, tandis que le sous-problème (2.40)-(2.43) demande seulement la minimisation de f sur un hyperplan ou l'intersection de deux hyperplans, et donne une formule facilement calculable et proche de la solution de $My = v^k$.

Afin de mieux cerner cette différence, considérons le cas le plus simple des contraintes non linéaires, par exemple quadratiques. Supposons que,

$$g_i(x) = \frac{1}{2}(x - w^i)^t Q_i (x - w^i) \quad (1 \leq i \leq m), \tag{2.47}$$

avec Q_i symétrique, définie positive, $w^i \in \mathbb{R}^n$.

Dans le cas de l'algorithme S, nous remplaçons simplement $v^k = Q_{i(k)}(x^k - w^{i(k)})$ dans (2.37)-(2.43).

Ensuite regardons la méthode de Han. Soit $z^k = x^k + y_{i(k)}^k$, comme donné par (2.44)-(2.46) et prenons $i(k) = l$. Les conditions d'optimalité donnent $x^{k+1} = x^k$ si $g_l(x^k) \leq 0$ et sinon x^{k+1} est la solution de

$$\begin{cases} M(x - z^k) = \lambda Q_l (x - w^l), \\ (x - w^l)^t Q_l (x - w^l) = 0. \end{cases} \tag{2.48}$$

Ainsi $x^{k+1} = (M - \lambda Q_l)^{-1}(Mz^k - \lambda Q_l w^l)$ où λ est la solution de

$$\begin{aligned} &[(M - \lambda Q_l)^{-1}(Mz^k - \lambda Q_l w^l) - w^l]^t Q_l \\ &\quad [(M - \lambda Q_l)^{-1}(Mz^k - \lambda Q_l w^l) - w^l] = 0 \end{aligned} \tag{2.49}$$

Et (2.49) ne peut être résolu explicitement. Si nous le résolvons, avec la méthode de Newton, générant une suite $\{\lambda_j\}$ qui converge vers la solution λ^* de (2.49), à chaque itération de la méthode de Newton la fonction d'évaluation nécessite la solution en y du système linéaire

$$(M - \lambda_j Q_l)y = Mz^j - \lambda_j Q_l w^l, \tag{2.50}$$

autrement dit, même en s'arrêtant après quelques itérations de Newton, chaque itération de la méthode de Han demande la solution d'une suite de systèmes linéaires $n \times n$ avec des matrices différentes. Tandis que pour l'algorithme S chaque itération requiert la solution d'un seul système linéaire $n \times n$.

De plus pour la convergence de (2.44)-(2.46), une condition de Slater doit être vérifiée dans la méthode de Han. Soit I l'ensemble des i tels que C_i est un polyèdre. La condition est

$$\left(\bigcap_{i\in I} C_i\right) \bigcap \left(\bigcap_{i\notin I} riC_i\right) \neq \emptyset, \tag{2.51}$$

où ri dénote l'intérieur relatif. Par contre pour la convergence de l'algorithme S, la seule condition requise est

$$C = \bigcap_{i=1}^{m} C_i \neq \emptyset.$$

2.6 Conclusion

Nous venons de montrer que la méthode proposée par Iusem (à travers l'algorithme S) est plus intéressante d'un point de vue place mémoire (pour une éventuelle implémentation) que la méthode de Han.

Seulement la nature de cette méthode est purement primale, de ce fait son taux de convergence peut être sous-linéaire. De plus, dans beaucoup d'applications la détermination d'une solution optimale duale est aussi importante que celle d'une solution optimale primale x^*. Puisque l'algorithme présenté dans ce chapitre ne fournit pas une telle solution duale, il est intéressant de regarder une variante qui génère aussi une suite $\{z^k\} \subset \mathbb{R}^n$ convergeant vers une solution optimale duale z^*. Ceci fera l'objet du chapitre suivant.

Chapitre 3

La méthode primale-duale de Iusem: une méthode primale-duale de type "row-action"

3.1 Introduction

Nous présentons une méthode primale-duale de type "row-action" pour la minimisation d'une fonction strictement convexe sous des contraintes convexes générales. Ici, comme dans l'algorithme primal, les contraintes ne sont utilisées qu'une par une et aucun changement n'est effectué sur celles-ci, ni sur leur matrice jacobienne. A chaque itération un sous-problème est résolu consistant en la minimisation d'une fonction objectif sous contrainte d'une ou deux équations linéaires.

La différence avec la méthode primale est que l'algorithme considéré ici génère deux suites : l'une, appelée primale, converge vers la solution du problème, et l'autre, appelée duale, approxime un vecteur optimal des multiplicateurs de Karush-Kuhn-Tucker, pour le problème considéré.

Afin d'approximer la solution duale, une matrice $m \times n$ doit être mémorisée au cours des itérations, et à chacune de celles-ci une de ses lignes est mise à jour. Ceci est le prix à payer afin d'avoir un algorithme primal-dual.

Nous prouvons la convergence de la suite primale pour des contraintes convexes générales. Dans le cas de contraintes linéaires, nous prouvons également que la suite primale converge au moins linéairement et nous en déduisons la convergence de la suite duale.

Pour commencer nous posons le problème et donnons une description de l'algorithme primal-dual. Nous présentons ensuite l'algorithme sous une forme générale et établissons ses propriétés de convergence. Dans la section 3.4, nous obtenons une version séquentielle de cet algorithme. Dans la suivante, l'algorithme est reformulé et présenté sous une forme plus adéquate pour les calculs. Dans les deux dernières sections, nous regardons le cas de contraintes d'inégalités linéaires. Nous y prouvons que la suite primale a un taux de convergence linéaire et nous en déduisons que la suite duale est convergente.

3.2 La méthode

3.2.1 Position du problème

Considérons le problème (PPD) suivant

$$(\text{PPD}) \begin{cases} \min & f(x) \\ \text{s.c.} & g_i(x) \leq 0 \ (1 \leq i \leq m), \end{cases}$$

où $f, g_i : \mathbb{R}^n \to \mathbb{R}$. Soit

$$C_i = \{x \in \mathbb{R}^n : g_i(x) \leq 0\}, \ C = \bigcap_{i=1}^{m} C_i.$$

Sur les contraintes, nous faisons les hypothèses suivantes,

(A'1) g_i est convexe $(1 \leq i \leq m)$.
(A'2) g_i est continument différentiable $(1 \leq i \leq m)$.
(A'3) $C \neq \emptyset$.

Nous faisons aussi quelques hypothèses générales sur f qui simplifie l'analyse de la convergence,

(B'1) f est strictement convexe.
(B'2) f est différentiable.
(B'3) Les ensembles de niveau de f sont bornés.

Les hypothèses (A'2) et (B'2) ne sont pas essentielles et peuvent être remplacées par la continuité. Dans ce cas il suffit d'utiliser le sous-gradient de f ou de g_i au point x au lieu du gradient de f ou de g_i en x dans tout le développement suivant. Par souci de clarté dans la présentation, nous supposerons les fonctions f et g_i différentiables. L'hypothèse (B'3) quant à elle assure l'existence du minimum de f sur tout ensemble fermé convexe contenu dans $\mathbb{R}^n$; d'autre part l'unicité de ce minimum suit de (B'1).

3.2.2 Description de l'algorithme primal-dual

L'algorithme primal décrit dans le chapitre précédent ne fournit pas de solution duale, par conséquent son taux de convergence peut être sous-linéaire. L'algorithme présenté ici, génère deux suites, l'une d'elles approxime la solution primale et l'autre approxime une solution duale de (PPD). Chaque itération consiste encore à minimiser f sous une ou deux contraintes linéaires. L'une est un hyperplan séparant l'itéré courant de l'ensemble admissible d'une des contraintes, l'autre garantit la propriété de monotonie de base de l'algorithme. Afin d'approximer la solution duale une matrice $m \times n$ doit être mémorisée le long de la procédure, et à chaque itération on utilise ou on met à jour une de ses lignes. C'est le coût nécessaire pour avoir un algorithme primal-dual. Cependant, cette matrice a deux avantages significatifs par rapport à celle considérée à chaque itération dans la méthode de Han:

- elle a la même structure creuse que la matrice jacobienne des contraintes,

- seules des entrées correspondant à des entrées non constantes de la matrice jacobienne doivent être stockées et mises à jour.

Dans ce sens, la méthode primale-duale est compatible avec les exigences des méthodes de type "row-action".

3.3 L'algorithme de base

Considérons le problème décrit ci-avant

$$(\text{PPD}) \begin{cases} \min & f(x) \\ \text{s.c.} & g_i(x) \leq 0 \ (1 \leq i \leq m) \end{cases}$$

Soit x^* la solution de (PPD). Cette solution existe par (B'3) et est unique par (B'1). Nous considérons une suite $\{W_k\}$ de sous-ensembles fermés convexes de $\mathbb{R}^n$ et une suite $\{x^k\} \subset \mathbb{R}^n$ définie par

$$x^0 \in \mathbb{R}^n, \tag{3.1}$$

$$x^{k+1} = \text{argmin}_{x \in W_k} f(x). \tag{3.2}$$

Les hypothèses (B'1), (B'3) garantissent l'existence et l'unicité de x^{k+1}. Soit

$$U = \{x \in \mathbb{R}^n : \langle \nabla f(x), x^* - x \rangle \geq 0\}, \tag{3.3}$$

et pour $x \in \mathbb{R}^n$, définissons $L(x) \subset \mathbb{R}^n$ comme

$$L(x) = \{y \in \mathbb{R}^n : \langle \nabla f(x), y - x \rangle \geq 0\}. \tag{3.4}$$

Pour $x \in \mathbb{R}^n$ et $E \subset \mathbb{R}^n$, soit

$$d(x, E) = \inf_{y \in E} \|x - y\|. \tag{3.5}$$

Proposition 3.3.1 *Soient les ensembles W_k et la suite $\{x^k\}$ définie par (3.1)-(3.2), satisfaisant pour tout $k \geq 0$,*

$$x^* \in W_k \subset L(x^k). \tag{3.6}$$

Alors,

(i) $f(x^k) \leq f(x^{k+1}) \leq f(x^*)$, *pour tout* $k \geq 0$,
$f(x^k) = f(x^{k+1})$ *implique* $x^k = x^{k+1}$,

(ii) $\{x^k\}$ *est bornée,*

(iii) $\lim_{k \to \infty}(x^{k+1} - x^k) = 0$,

(iv) $\lim_{k \to \infty} d(x^k, W_k) = 0$.

Preuve

Une preuve de cette proposition peut être trouvée en annexe.

□

Corollaire 3.3.1 *Soit $\{x^k\}$ une suite définie par (3.1)-(3.2) avec W_k satisfaisant (3.6). Alors,*

(i) si un point limite $\tilde{x}$ de $\{x^k\}$ appartient à C, alors $\tilde{x} = x^*$,

(ii) si $x^k \in C$ pour certains $k \geq 0$, alors $x^k = x^*$.

Preuve

Une preuve de ce corollaire peut être trouvée en annexe.

□

La proposition 3.3.1 et le corollaire 3.3.1 indiquent que si les ensembles W_k sont choisis tels que (3.6) est satisfait et si tous les points limites appartiennent à C, alors la suite $\{x^k\}$ converge vers x^*. Ces résultats sont une généralisation de ceux de la section 2.3.3.

Dans la section suivante, nous donnons un algorithme qui est un cas particulier de (3.1)-(3.2) pour lequel (3.6) a lieu.

3.4 Version séquentielle de l'algorithme de base

3.4.1 Préliminaires

L'idée de base de l'algorithme est qu'à chaque itération, la fonction f doit être minimisée sur un ensemble convexe ayant les caractéristiques suivantes: l'ensemble doit contenir la solution du problème (PPD), et être le plus simple possible. En plus, nous voulons que la suite des valeurs fonctionnelles choisies parmi ces minima au cours des différentes itérations soit une suite croissante.

Dans la suite, nous présentons la version séquentielle de l'algorithme et nous analysons sa convergence.

3.4.2 Algorithme et commentaires

Nous générons quatre suites,

$$\{x^k\} \subset \mathbb{R}^n, \quad \{z^k\} \subset \mathbb{R}^n, \quad \{\bar{A}^k\} \subset \mathbb{R}^{m \times n}, \quad \{\bar{b}^k\} \subset \mathbb{R}^m.$$

De plus à chaque itération nous utilisons aussi une matrice $A^k \in \mathbb{R}^{m \times n}$ et un vecteur $b^k \in \mathbb{R}^n$. Ceux-ci ne sont pas mis à jour mais calculés directement à partir des données du problème. Nous verrons que

- x^k est une approximation de x^*;
- z^k approxime une solution duale (i.e. un vecteur des multiplicateurs de Karush-Kuhn-Tucker pour le problème (PPD));
- A^k et b^k fournissent une approximation linéaire des contraintes (i.e. le système $A^k x \leq b^k$ approxime $g_i(x) \leq 0$ $(1 \leq i \leq m)$);
- $\bar{A}^k, \bar{b}^k$ définissent un système "substitut" de contraintes, $\bar{A}^k x \leq \bar{b}^k$ qui stocke l'information requise pour la mise à jour des variables primales et duales.

Nous employons encore deux autres suites : une suite $\{\alpha_k\} \subset \mathbb{R}$, de paramètres de relaxation satisfaisant,

$$0 \leq \alpha_k \leq 1 \tag{3.7}$$

et une suite $\{i(k)\} \subset \{1, \ldots, m\}$ appelée suite de contrôle qui détermine quelle contrainte est utilisée à chaque itération. $i(k)$ est choisie pour être presque cyclique (cfr définition 2.4.1).

Soit $e^i \in \mathbb{R}^m$, $1 \leq i \leq m$ définie comme $e^i_j = \delta_{ij}$, le symbole de Kronecker, et définissons $E^i \in \mathbb{R}^{m \times m}$ comme

$$E^i = e^i (e^i)^t; \tag{3.8}$$

c'est à dire, la $i$ème entrée diagonale de E^i est 1, et toutes les autres entrées sont nulles. Comme l'algorithme que nous présentons maintenant est une version primale-duale de l'algorithme donné dans le chapitre précédent, nous le noterons algorithme primal-dual S.

L'algorithme primal-dual S est défini de la façon suivante:

ALGORITHME PRIMAL - DUAL S

Initialisation

Soient $x^0 \in \mathbb{R}^n, z^0 \in \mathbb{R}^m, b^0, \bar{b}^0 \in \mathbb{R}^m, A^0, \bar{A}^0 \in \mathbb{R}^{m \times n}$ où $a^{0,i}, \bar{a}^{0,i}, 1 \leq i \leq m$, satisfont respectivement

$$x^0 \in U, \tag{3.9}$$

$$a^{0,i} = \bar{a}^{0,i} = \nabla g_i(x^0), \quad 1 \leq i \leq m, \tag{3.10}$$

$$b^0_i = \langle a^{0,i}, x^0 \rangle - \alpha_0 g_i(x^0), \quad 1 \leq i \leq m, \tag{3.11}$$

$$\bar{b}^0_i = \langle a^{0,i}, x^0 \rangle - g_i(x^0), \quad 1 \leq i \leq m, \tag{3.12}$$

$$z^0 \geq 0, \tag{3.13}$$

$$\nabla f(x^0) = -(A^0)^t z^0, \tag{3.14}$$

$$(z^0)^t (\bar{A}^0 x^0 - \bar{b}^0) = 0. \tag{3.15}$$

Itération

Soient $x^k, z^k, \bar{A}^k, \bar{b}^k$, effectuer les opérations suivantes.

(a) Définir $A^k \in \mathbb{R}^{m \times n}$ et $b^k \in \mathbb{R}^m$ comme suit,

$$a^{k,i} = \nabla g_i(x^k), \quad 1 \leq i \leq m, \tag{3.16}$$

$$b^k_i = \langle a^{k,i}, x^k \rangle - \alpha_k g_i(x^k), \quad 1 \leq i \leq m. \tag{3.17}$$

(*b*) Si $g_{i(k)}(x^k) > 0$, effectuer une étape externe.
Mettre à jour de la manière suivante,

$$x^{k+1} = \quad \text{argmin} f(x), \tag{3.18}$$

$$\text{s.c.} \quad \langle \nabla f(x^k), x \rangle \geq \langle \nabla f(x^k), x^k \rangle, \tag{3.19}$$

$$-\langle a^{k,i(k)}, x \rangle \geq -b^k_{i(k)}. \tag{3.20}$$

Soient

$$\mu_k \geq 0 \quad \text{et} \quad \lambda_k \geq 0 \tag{3.21}$$

les multiplicateurs associés respectivement aux contraintes (3.19), (3.20) dans la solution du problème (3.18)-(3.20).
Mettre à jour z^k comme suit,

$$z^{k+1} = \mu_k z^k + \lambda_k e^{i(k)}. \tag{3.22}$$

Définir

$$\gamma_k = \begin{cases} 0, & \text{si } z^{k+1}_{i(k)} = 0, \\ \lambda_k / z^{k+1}_{i(k)} & \text{sinon.} \end{cases} \tag{3.23}$$

Mettre à jour $\bar{A}^k, \bar{b}^k$ comme suit,

$$\bar{A}^{k+1} = \bar{A}^k + \gamma_k E^{i(k)}(A^k - \bar{A}^k), \tag{3.24}$$

$$\bar{b}^{k+1} = \bar{b}^k + \gamma_k E^{i(k)}(b^k - \bar{b}^k). \tag{3.25}$$

(*c*) Si $g_{i(k)}(x^k) \leq 0$, effectuer une étape interne.
Mettre à jour de la manière suivante,

$$x^{k+1} = \quad \text{argmin} f(x), \tag{3.26}$$

$$\text{s.c.} \quad \langle \nabla f(x^k) + z^k_{i(k)} \bar{a}^{k,i(k)}, x \rangle \geq$$

$$\langle \nabla f(x^k), x^k \rangle + z^k_{i(k)} \bar{b}^k_{i(k)}, \tag{3.27}$$

$$-\langle \bar{a}^{k,i(k)}, x \rangle \geq -\bar{b}^k_{i(k)}. \tag{3.28}$$

Soient

$$\mu_k \geq 0 \quad \text{et} \quad \lambda_k \geq 0 \tag{3.29}$$

les multiplicateurs associés respectivement aux contraintes (3.27), (3.28) dans la solution du problème (3.26)-(3.28).
Mettre à jour $z^k, \bar{A}^k, \bar{b}^k$ comme suit,

$$z^{k+1} = \mu_k z^k + (\lambda_k - \mu_k z^k_{i(k)}) e^{i(k)}, \tag{3.30}$$

$$\bar{A}^{k+1} = \bar{A}^k, \tag{3.31}$$

$$\bar{b}^{k+1} = \bar{b}^k. \tag{3.32}$$

Nous montrerons que cet algorithme est un cas particulier de (3.1)-(3.2) satisfaisant (3.6), après avoir prouvé dans la proposition 3.4.2 quelques propriétés de la mise à jour (3.9)-(3.32).
Mais avant cela, faisons quelques remarques à propos de l'algorithme.

Remarques

1) Si nous prenons x^0 le minimum sans contraintes de f, (qui existe par (B'1)-(B'3)), et $z^0 = 0$, alors les conditions d'initialisation (3.9) et (3.13)-(3.15) sont vérifiées. Le minimum sans contrainte de f est calculable, (nous supposons implicitement que des problèmes tels que (3.18)-(3.20) et (3.26)-(3.28) sont résolvables).

Nous remarquons que pour un $x^0 \in U$ quelconque (pas nécessairement le minimum de f), il est toujours possible de trouver un z^0 approprié de la façon suivante. Nous ajoutons la contrainte linéaire $x \in L(x^0)$ au système des contraintes, i.e. nous considérons

$$g_{m+1}(x) = \langle \nabla f(x^0), x^0 - x \rangle.$$

Comme $x^0 \in U$, $x^* \in L(x^0)$, nous avons $g_{m+1}(x^*) \leq 0$, ainsi x^* est aussi la solution du problème modifié. Si nous prenons ensuite $z^0 \in \mathbb{R}^{m+1}$ comme $z^0 = (0, \ldots, 0, 1)^t$, les hypothèses (3.13)-(3.14) sont satisfaites. Cette astuce peut être utilisée dans des applications pour des fonctions qui ne satisfont pas (B'3) (et qui par conséquent peuvent ne pas avoir de minimum sans contrainte), car pour de telles fonctions un point dans U peut être facilement trouvé.

2) Les matrices A^k et les vecteurs b^k sont introduits pour simplifier les notations. A l'itération k, nous calculons x^{k+1}, z^{k+1} en utilisant seulement x^k, z^k et les $i(k)$èmes lignes et composantes de $A^k, \bar{A}^k$ et $b^k, \bar{b}^k$.

De plus, de (3.8), (3.24), (3.25), (3.31) et (3.32), seulement les $i(k)$ème composante et ligne respective de $\bar{A}^k$ et $\bar{b}^k$ sont mises à jour.

Finallement, notons que durant l'itération externe, $\bar{A}^k$ et $\bar{b}^k$ sont mis à jour mais non utilisés, tandis que durant l'itération interne, ils sont utilisés mais inchangés.

3) Conserver $\bar{A}^k$ une matrice $m \times n$ au cours des itérations semble être un grand prix à payer pour la suite duale $\{z^k\}$, mais ce n'est pas le cas comme nous le montrons ci-après.

Considérons des paires (i, j) telles que $\partial g_i(x)/\partial x_j$ est une constante, appelée a_{ij}. Il découle de (3.10), (3.16) que

$$a_j^{k,i} = a_{ij}, \quad \forall k,$$

et ensuite, de (3.8), (3.10), (3.24), que

$$\bar{a}_j^{k,i} = a_{ij}, \quad \forall k.$$

Ainsi les seules entrées de $\bar{A}^k$ qui doivent être mises à jour et gardées sont celles qui correspondent à des entrées non constantes de la matrice jacobienne du système des contraintes. En particulier, si ce système est linéaire, noté $Ax \leq b$ où $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$, nous obtenons

$$\bar{A}^k = A^k = A, \quad \forall k,$$

et nous pouvons complètement éliminer les matrices $\bar{A}^k, A^k$ de la procédure. En plus, si pour (i, j) nous avons

$$\partial g_i(x)/\partial x_j = 0, \quad \forall x,$$

alors

$$a_j^{k,i} = \bar{a}_j^{k,i} = 0, \quad \forall k,$$

c'est à dire que les matrices $\bar{A}^k, A^k$ ont la même structure creuse que la matrice jacobienne. Ceci est une autre caractéristique des méthodes de type "row-action". La structure creuse de $\bar{A}^k$ contraste avec la matrice considérée dans la méthode de Han (dont les lignes sont les y_i^k), qui est plus dense.

3.4.3 Analyse de la convergence

A partir d'ici, $x^k, z^k, A^k, \bar{A}^k, b^k$ et $\bar{b}^k$ représentent les variables de l'algorithme primal-dual S.

Dans la proposition suivante nous exposons les conditions de Karush-Kuhn-Tucker pour (3.18)-(3.20) et (3.26)-(3.28).

Proposition 3.4.1 *Pour tout $k \geq 0$, les relations suivantes ont lieu,*

(i) Si une itération externe est effectuée à l'itération k, alors

$$\nabla f(x^{k+1}) = \mu_k \nabla f(x^k) - \lambda_k a^{k,i(k)}, \tag{3.33}$$

$$\langle \nabla f(x^k), x^{k+1} - x^k \rangle \geq 0, \tag{3.34}$$

$$\langle a^{k,i(k)}, x^{k+1} \rangle \leq b^k_{i(k)}, \tag{3.35}$$

$$\lambda_k (b^k_{i(k)} - \langle a^{k,i(k)}, x^{k+1} \rangle) = 0, \tag{3.36}$$

$$\mu_k \langle \nabla f(x^k), x^{k+1} - x^k \rangle = 0. \tag{3.37}$$

(ii) Si une itération interne est effectuée à l'itération k, alors

$$\nabla f(x^{k+1}) = \mu_k (\nabla f(x^k) + z^k_{i(k)} \bar{a}^{k,i(k)}) - \lambda_k \bar{a}^{k,i(k)}, \tag{3.38}$$

$$\langle \nabla f(x^k), x^{k+1} - x^k \rangle \geq z^k_{i(k)} (\bar{b}^k_{i(k)} - \langle \bar{a}^{k,i(k)}, x^{k+1} \rangle), \tag{3.39}$$

$$\langle \bar{a}^{k,i(k)}, x^{k+1} \rangle \leq \bar{b}^k_{i(k)}, \tag{3.40}$$

$$\lambda_k (\bar{b}^k_{i(k)} - \langle \bar{a}^{k,i(k)}, x^{k+1} \rangle) = 0, \tag{3.41}$$

$$\mu_k [\langle \nabla f(x^k), x^{k+1} - x^k \rangle - z^k_{i(k)} (\bar{b}^k_{i(k)} - \langle \bar{a}^{k,i(k)}, x^{k+1} \rangle)] = 0. \tag{3.42}$$

Preuve

(3.33)-(3.37) et (3.38)-(3.42) sont les conditions de Karush-Kuhn-Tucker respectives des problèmes (3.18)-(3.20) et (3.26)-(3.28).

□

Soit $g(x) \in \mathbb{R}^m$ le vecteur de composantes $g_i(x), 1 \leq i \leq m$, pour tout $x \in \mathbb{R}^n$.

Proposition 3.4.2 *Pour tout $k \geq 0$, nous avons,*

(i) $z^k \geq 0$,

(ii) $\nabla f(x^k) = -(\bar{A}^k)^t z^k$,

(iii) $(z^k)^t (\bar{b}^k - \bar{A}^k x^k) = 0$,

(iv) $A^k x - b^k \leq g(x) - (1 - \alpha_k) g(x^k)$, *pour tout* $x \in \mathbb{R}^n$,

(v) $\bar{A}^k x - \bar{b}^k \leq g(x)$, *pour tout* $x \in \mathbb{R}^n$.

Preuve

Une preuve de cette proposition peut être trouvée en annexe.

□

Nous pouvons maintenant donner une motivation heuristique pour les mises à jour de l'algorithme primal-dual S. L'idée est qu'à chaque itération, f soit minimisée sur un ensemble convexe ayant les propriétés suivantes: il contient la solution optimale du problème (PPD) et il est aussi simple que possible. Nous voulons aussi que la suite des valeurs fonctionnelles en ces minima soit croissante. Dans un tel cas en accord avec la proposition 3.3.1 et le corollaire 3.3.1, nous avons de bonnes chances de converger vers la solution x^* du problème (PPD).

La simplicité ici est de prendre l'ensemble convexe comme l'intersection de deux demi-espaces, donnés par deux contraintes d'inégalité linéaire. L'un de ceux-ci (donné par (3.20) ou (3.28)) garantit que x^* est contenu dans l'ensemble admissible du sous-problème: pour (3.20), ceci découle de la proposition 3.4.2(iv) et de $g_i(x^k) > 0$; pour (3.28), cela suit de la proposition 3.4.2(v): si $g(x) \leq 0$ alors $\bar{A}^k x - \bar{b}^k \leq 0$. L'autre demi-espace (donné par (3.19) ou (3.27)) garantit que la région admissible du sous-problème est contenue dans $L(x^k)$, ce qui assure (cfr proposition 3.3.1) que la suite $\{f(x^k)\}$ est croissante. Il suit donc que ces mises à jour satisfont (3.6), et ainsi les résultats de la proposition 3.3.1 et du corollaire 3.3.1 ont lieu. Nous établissons ces résultats plus formellement dans la proposition 3.4.3 donnée ci-dessous.

Nous pouvons aussi justifier les propriétés de la suite $\{z^k\}$ à l'aide de la proposition 3.4.2. Comme $\bar{A}^k$ approxime la matrice jacobienne de g en x^k et $\bar{A}^k x^k - \bar{b}^k$ approxime $g(x^k)$, (i)-(iii) approximent les conditions de Karush-Kuhn-Tucker du problème (PPD) incluant les variables duales, i.e.

$$z \geq 0, \quad \nabla f(x) + \sum_{i=1}^{m} z_i \nabla g_i(x) = 0, \quad z^t g(x) = 0,$$

qui ont lieu à chaque itération. Seulement la condition d'admissibilité primale $g(x) \leq 0$ est perdue; nous verrons que celle-ci dépend seulement de la suite $\{x^k\}$. L'itération interne de l'algorithme primal-dual S est requise pour générer une suite $\{z^k\}$ satisfaisant les points (i)-(iii) de la proposition 3.4.2. La mise à jour de la suite primale $\{x^k\}$ durant les itérations externes est par conséquent exactement la même pour l'algorithme primal-dual S que pour l'algorithme S présenté dans la section 2.4.1.

Il est aussi possible d'obtenir une version simultanée de l'algorithme primal-dual S, il suffit de remplacer $a^{k,i(k)}, b^k_{i(k)}$ dans (3.20) (respectivement $\bar{a}^{k,i(k)}, \bar{b}^k_{i(k)}$ dans (3.27)-(3.28)) par une combinaison convexe des lignes et composantes de A^k, b^k (respectivement de $\bar{A}^k, \bar{b}^k$), exactement comme dans la version simultanée de l'algorithme primal de la section 2.4.3.

Notons que l'ensemble admissible de la contrainte (3.19) est $L(x^k)$ comme défini dans (3.4). Soient $S_k, \bar{L}_k, \bar{S}_k$ les ensembles admissibles des contraintes (3.20), (3.27), (3.28) respectivement. Définissons

$$W_k = \begin{cases} L(x^k) \bigcap S_k, & \text{si la } k\text{ème itération est externe,} \\ \bar{L}_k \bigcap \bar{S}_k & \text{sinon.} \end{cases} \tag{3.43}$$

Les problèmes (3.18)-(3.20) et (3.26)-(3.28) peuvent alors se réécrire comme,

$$\min f(x), \quad \text{s.c. } x \in W_k.$$

Proposition 3.4.3 *Si W_k est défini comme dans (3.43), alors pour tout $k \geq 0$,*

$$x^* \in W_k \subset L(x^k). \tag{3.44}$$

Preuve

Nous divisons la preuve en cinq parties.

(i) Pour tout $k \geq 0$, $W_k \subset L(x^k)$. En effet, si la $k$ème itération est une étape externe, ceci suit directement de (3.43). Si c'est une étape interne, on prend $x \in W_k$, i.e. tel que (3.27), (3.28) ont lieu. Multipliant les deux membres de (3.28) par $z^k_{i(k)}$, (positif par la proposition 3.4.2(i)), et additionant (3.27), nous obtenons

$$\langle \nabla f(x^k), x - x^k \rangle \geq 0, \quad \text{i.e. } x \in L(x^k).$$

(ii) Si la $k$ème itération est une étape externe, alors $x^* \in S_k$. En effet, par la proposition 3.4.2(iv) avec $x = x^*$, et considérant la $i(k)$ème ligne,

$$\langle a^{k,i(k)}, x^* \rangle - b^k_{i(k)} \leq g_{i(k)}(x^*) - (1 - \alpha_k) g_{i(k)}(x^k).$$

Comme $\alpha_k \leq 1$ (par (3.7)), $g_{i(k)}(x^k) > 0$ (par la définition du pas externe) et $g_{i(k)}(x^*) \leq 0$, il suit que

$$\langle a^{k,i(k)}, x^* \rangle \leq b^k_{i(k)}, \quad \text{i.e., } x^* \in S_k.$$

(iii) Si la $k$ème itération est une étape interne, alors $x^* \in \bar{S}_k$. En effet, par la proposition 3.4.2(v) avec $x = x^*$ et considérant la $i(k)$ème ligne,

$$\langle \bar{a}^{k,i(k)}, x^* \rangle - \bar{b}^k_{i(k)} \leq g_{i(k)}(x^*) \leq 0, \quad \text{i.e. } x^* \in \bar{S}_k.$$

(iv) Si la $k$ème itération est une étape interne, alors $x^* \in W_k$. En effet, par la proposition 3.4.2(ii) et (iii), nous obtenons après multiplication par $x^* - x^k$,

$$\begin{aligned} \langle \nabla f(x^k), x^* - x^k \rangle &= -(z^k)^t \bar{A}^k (x^* - x^k) \\ &= (z^k)^t (\bar{b}^k - \bar{A}^k x^*) \\ &= \sum_{i=1}^{m} z^k_i (\bar{b}^k_i - \langle \bar{a}^{k,i}, x^* \rangle). \end{aligned} \tag{3.45}$$

Utilisons maintenant la proposition 3.4.2(v) dans (3.45) nous obtenons

$$\begin{aligned} &\langle \nabla f(x^k), x^* - x^k \rangle - z^k_{i(k)} (\bar{b}^k_{i(k)} - \langle \bar{a}^{k,i(k)}, x^* \rangle) \\ &= \sum_{\substack{i=1 \\ i \neq i(k)}}^{m} z^k_i (\bar{b}^k_i - \langle \bar{a}^{k,i}, x^* \rangle) \\ &\geq - \sum_{\substack{i=1 \\ i \neq i(k)}}^{m} z^k_i g_i(x^*) \geq 0, \end{aligned} \tag{3.46}$$

où la dernière inégalité découle de la proposition 3.4.2(i) et du fait que $x^* \in C$. (3.46) implique immédiatement que (3.27) a lieu avec $x = x^*$, i.e. que $x^* \in \bar{L}_k$. Le résultat suit en utilisant (iii) et (3.43).

(v) $x^* \in W_k$, pour tout $k \geq 0$. En effet par récurrence, le résultat tient pour $k = 0$ par (3.3), (3.4), (3.9). Supposons que ce soit vrai pour k. Comme x^{k+1} est le minimum de f sur W_k, la dérivée directionnelle de f en x^{k+1} dans toute direction contenue dans W_k doit être positive, et par la convexité de W_k, $x^* - x^{k+1}$ est une telle direction. La positivité de cette dérivée directionnelle signifie que

$$\langle \nabla f(x^{k+1}), x^* - x^{k+1} \rangle \geq 0, \quad \text{i.e. } x^* \in L(x^{k+1}). \tag{3.47}$$

Maintenant considérons deux cas. Si la $(k+1)$ème itération est une étape interne, alors $x^* \in W_{k+1}$ par (iv). Si c'est une étape externe, alors $x^* \in W_{k+1}$ par (ii) et (3.47). La thèse est donc établie.

□

La preuve de la proposition 3.4.3 nous permet de donner une motivation heuristique pour les mises à jour spécifiques de l'algorithme primal-dual S, particulièrement pour les étapes internes, qui constituent la différence principale entre les algorithmes primal et primal-dual S. A l'itération interne, comme montré dans (3.45) avec x au lieu de x^*, $\langle \nabla f(x^k), x - x^k \rangle$ est une combinaison positive des $\bar{b}_i^k - \langle \bar{a}^{k,i}, x \rangle$, lesquelles sont toutes positives dans la région admissible C (par la proposition 3.4.2(v)). Nous prenons l'une d'entre elles pour créer l'hyperplan séparant donné par (3.28), et l'éliminons de la combinaison positive, générant une nouvelle fonctionnelle qui, étant une combinaison positive des fonctionnelles restantes, est aussi positive sur C, comme montré dans (3.46). Cette fonctionnelle-ci définit le demi-espace donné par (3.27), produisant une coupure de sorte que la suite $\{x^k\}$ reste dans C durant les itérations internes aussi. L'effet radical de ces étapes internes sur le taux de convergence sera discuté dans la section 3.6.

Nous résumons dans le corollaire suivant les résultats de convergence pour l'algorithme primal-dual S obtenus jusqu'ici.

Corollaire 3.4.1 *Soit $\{x^k\}$ une suite quelconque générée par l'algorithme primal-dual S. Alors,*

(i) $\{x^k\}$ *est bornée,*
(ii) $\lim_{k\to\infty}(x^{k+1} - x^k) = 0$,
(iii) $\lim_{k\to\infty} d(x^k, W_k) = 0$,
(iv) si un point limite $\tilde{x}$ de $\{x^k\}$ appartient à C, alors $\tilde{x} = x^*$,
(v) si x^k appartient à C pour certains k, alors $x^k = x^*$.

Preuve

Elle découle de ce qui précède.

□

Le point (iv) du corollaire 3.4.1 indique que, si x^k appartient à C pour tout k, alors les calculs devraient s'arrêter. Supposons par convention que dans ce cas, $x^l = x^k$ pour tout $l > k$.

Afin de complèter la preuve de convergence, il suffit de montrer que $\lim_{k\to\infty} d(x^k, W_k) = 0$ implique que tout point limite de $\{x^k\}$ appartient à C. Afin d'avoir cette implication il faut renforcer l'hypothèse (3.7) sur la suite $\{\alpha_k\}$ des paramètres de relaxation. Celle-ci doit être limitée autour de 0.

Théorème 3.4.1 *S'il existe $\bar{\alpha} > 0$ tel que, pour tout $k \geq 0$,*

$$\bar{\alpha} \leq \alpha_k \leq 1, \tag{3.48}$$

alors toute suite générée par l'algorithme primal-dual S converge vers la solution x^* du problème (PPD).

Preuve

Prenons un point limite $\tilde{x}$ de $\{x^k\}$, qui existe par le corollaire 3.4.1(i). Considérons une sous-suite $\{x^{j_k}\}$ de $\{x^k\}$ telle que $\lim_{k\to\infty} x^{j_k} = \tilde{x}$. Fixons une contrainte, la $i$ème. Soit $l_k = \min\{l > j_k : i(l) = i\}$. Par le caractère presque cyclique de $\{i(k)\}$, nous avons $0 \leq l_k - j_k < r$. Par conséquent,

$$\|x^{l_k} - x^{j_k}\| \leq \sum_{j=j_k}^{l_k-1} \|x^{j+1} - x^j\| \leq \sum_{j=j_k}^{j_k+r} \|x^{j+1} - x^j\|. \tag{3.49}$$

Par le corollaire 3.4.1(ii), nous déduisons que $\lim_{k\to\infty}(x^{l_k} - x^{j_k}) = 0$, impliquant que $\lim_{k\to\infty} x^{l_k} = \tilde{x}$. Estimons $d(x^{l_k}, W_{l_k})$.

Supposons que la $l_k$ème itération soit une étape externe. Comme $S_{l_k} \supset W_{l_k}$,

$$d(x^{l_k}, W_{l_k}) \geq d(x^{l_k}, S_{l_k}). \tag{3.50}$$

De (3.16) et (3.17),

$$\langle a^{l_k,i}, x^{l_k}\rangle - b_i^{l_k} = \alpha_{l_k} g_i(x^{l_k}) \geq \bar{\alpha} g_i(x^{l_k}) > 0, \tag{3.51}$$

ainsi $x^{l_k} \notin S_{l_k}$. Il suit que $d(x^{l_k}, S_{l_k})$ est la distance de x^{l_k} à l'hyperplan limitant de S_{l_k}, notamment $\alpha_{l_k} g_i(x^{l_k})/\|\nabla g_i(x^{l_k})\|$. Ainsi, de (3.50) et de la proposition 3.3.1(i),

$$d(x^{l_k}, W_{l_k}) \geq d(x^{l_k}, S_{l_k}) = \alpha_{l_k} g_i(x^{l_k})/\|\nabla g_i(x^{l_k})\| \geq \bar{\alpha} g_i(x^{l_k})/w > 0, \tag{3.52}$$

où

$$w = \sup_{x\in E}\|\nabla g_i(x)\|, \quad E = \{x \in \mathbb{R}^n : f(x) \leq f(x^*)\}.$$

De (A'2) et (B'3), w est fini.

Supposons maintenant que la $l_k$ème itération est une étape interne. Dans un tel cas $g_i(x^{l_k}) < 0$, ainsi $\max\{0, g_i(x^{l_k})\} = 0$. Il suit de (3.52) que, pour les itérations internes et externes,

$$d(x^{l_k}, W_{l_k}) \geq (\bar{\alpha}/w)\max\{0, g_i(x^{l_k})\} \geq 0. \tag{3.53}$$

Prenant la limite dans (3.53) lorsque $k \to \infty$ et utilisant $\lim_{k\to\infty} x^{l_k} = \tilde{x}$ et le corollaire 3.4.1(iii), nous avons

$$0 = (\bar{\alpha}/w)\max\{0, g_i(\tilde{x})\}, \tag{3.54}$$

i.e., $g_i(\tilde{x}) \leq 0$. Comme i est une contrainte arbitraire, nous concluons que $g_i(\tilde{x}) \leq 0$ pour tout i, ou de manière équivallente que $\tilde{x} \in C$. Par le corollaire 3.4.1(iv), $\tilde{x} = x^*$. Comme $\tilde{x}$ est un point limite arbitraire de $\{x^k\}$, il suit que x^* est le seul point limite possible de la suite bornée $\{x^k\}$, c'est à dire $\lim_{k\to\infty} x^k = x^*$. D'où la thèse.

□

Nous terminons cette section avec une propriété de la suite duale dont nous aurons besoin dans la section 3.6 afin de prouver la convergence linéaire de la suite primale. Cette propriété est d'ailleurs un intérêt en elle-même.

Proposition 3.4.4 *Si pour certains i, il existe k_0 tel que $g_i(x^k) < 0$ pour tout $k \geq k_0$, alors il existe k_1 tel que $z_i^k = 0$ pour tout $k > k_1$.*

Preuve

Prenons $k_1 = \min\{k > k_0 : i(k) = i\}$. Nous procédons par récurrence. Commençons en $k_1 + 1$. Comme $g_i(x^{k_1}) < 0$ par hypothèse, la $k_1$ème itération est une étape interne. Comme $i(k_1) = i$, il suit de (3.30) que $z_i^{k_1+1} = \lambda_{k_1}$, le multiplicateur associé à la contrainte (3.28), qui peut être écrite comme $\langle \bar{a}^{k_1,i}, x\rangle - \bar{b}_i^{k_1} \leq 0$. Evaluant en x^{k_1+1} et utlisant la proposition 3.4.2(v), nous avons

$$\langle \bar{a}^{k_1,i}, x^{k_1+1}\rangle - \bar{b}_i^{k_1} \leq g_i(x^{k_1+1}) < 0, \tag{3.55}$$

parce que $k_1 + 1 \geq k_0$. Par (3.41) concluons que $0 = \lambda_{k_1} = z_i^{k_1+1}$.

Supposons à présent que $z_i^k = 0$ pour un certain $k > k_1$. Si $i(k) \neq i$, alors de (3.22) ou (3.30), $z_i^{k+1} = \mu_k z_i^k = 0$. Si $i(k) = i$, alors $z_i^{k+1} = \lambda_k$, et comme dans (3.55),

$$\langle \bar{a}^{k,i}, x^{k+1}\rangle - \bar{b}_i^k \leq g_i(x^k) < 0, \tag{3.56}$$

impliquant par (3.41) que $0 = \lambda_k = z_i^{k+1}$.

□

Corollaire 3.4.2 *Si $g_i(x^*) < 0$, alors il existe k_1 tel que $z_i^k = 0$ pour tout $k > k_1$.*

Preuve

Par le théorème 3.4.1, $x^* = \lim_{k\to\infty} x^k$, il existe donc k_0 tel que $g_i(x^k) < 0$ pour tout $k \geq k_0$. Le résultat suit de la proposition 3.4.4.

□

Le corollaire 3.4.2 signifie que les variables duales, associées aux contraintes qui sont vérifiées en tant qu'inégalité stricte à la solution x^*, deviennent nulles à une certaine itération et restent nulles par la suite.

3.5 Reformulation de l'algorithme primal-dual S

La version de l'algorithme primal-dual S, définie par (3.9)-(3.32), est appropriée pour l'analyse de la convergence, mais pas pour les calculs. C'est pourquoi, dans cette section nous réécrivons l'algorithme sous une forme assez compacte, mais surtout plus pratique pour les calculs. Comme nous avons besoin de garder les matrices $A^k, \bar{A}^k$, il est plus simple de remplacer les sous-problèmes (3.18)-(3.20) et (3.26)-(3.28) par la minimisation sur des hyperplans, plutôt que sur des demi-espaces.

Nous supposons que (3.48) est satisfait. Soient $G_k, F_k, \bar{G}_k, \bar{F}_k$ les hyperplans limites des demi-espaces définis par respectivement (3.19), (3.20), (3.27), (3.28) i.e.

$$
\begin{aligned}
G_k &= \{x \in \mathbb{R}^n : \langle \nabla f(x^k), x - x^k \rangle = 0\}, \\
F_k &= \{x \in \mathbb{R}^n : \langle a^{k,i(k)}, x \rangle = b^k_{i(k)}\}, \\
\bar{G}_k &= \{x \in \mathbb{R}^n : \langle \nabla f(x^k), x - x^k \rangle = z^k_{i(k)}(\bar{b}^k_{i(k)} - \langle a^{k,i(k)}, x \rangle)\}, \\
\bar{F}_k &= \{x \in \mathbb{R}^n : \langle \bar{a}^{k,i(k)}, x \rangle = \bar{b}^k_{i(k)}\}.
\end{aligned}
$$

Rappellons que $S_k, \bar{S}_k$ sont respectivement les ensembles admissibles des contraintes (3.20) et (3.28) .

Proposition 3.5.1 *Nous avons,*

(i) *Si la kème itération est une étape externe, soient* $\hat{x} = \text{argmin} f(x), \text{s.c. } x \in F_k$, *et* $\bar{x} = \text{argmin} f(x), \text{s.c. } x \in F_k \bigcap G_k$. *Alors,* $x^{k+1} = \hat{x}$, *si* $\hat{x} \in L(x^k)$ *et* $x^{k+1} = \bar{x}$, *sinon.*

(ii) *Si la kème itération est une étape interne, soient* $\hat{x} = \text{argmin} f(x), \text{s.c. } x \in \bar{G}_k$, *et* $\bar{x} = \text{argmin} f(x), \text{s.c. } x \in \bar{F}_k \bigcap \bar{G}_k$. *Alors,* $x^{k+1} = x^k$, *si* $z^k_{i(k)} = 0$; $x^{k+1} = \hat{x}$, *si* $\hat{x} \in \bar{S}_k$, $z^k_{i(k)} \neq 0$; *et* $x^{k+1} = \bar{x}$, *si* $\hat{x} \notin \bar{S}_k$, $z^k_{i(k)} \neq 0$.

Preuve

(i) Nous affirmons que $x^{k+1} \in F_k$, i.e. (3.20) est vérifiée en tant qu'égalité à la solution de (3.18)-(3.20).

> Sinon, (3.18)-(3.20) est équivalent à la minimisation de f sous la seule contrainte (3.19), i.e. min $f(x)$ s.c. $x \in L(x^k)$. Il suit de la convexité de f que la solution de ce problème-ci est précisément x^k, ainsi $x^{k+1} = x^k$, dans quel cas x^k satisfait (3.20) impliquant, par (3.17), $0 \geq \alpha_k g_{i(k)}(x^k)$, ce qui ne peut avoir lieu lors d'une itération externe.

Comme (3.20) est vérifiée en tant qu'égalité à la solution de (3.18)-(3.20), nous pouvons remplacer (3.20) par $x \in F_k$. Nous minimisons f sur F_k, et obtenons $\hat{x}$ avec un multiplicateur $\hat{\lambda}$ associé à la contrainte $x \in F_k$. Si $\hat{x}$ satisfait (3.19), alors $\hat{x}$ est la solution de (3.18)-(3.20), i.e. $x^{k+1} = \hat{x}$ et nous pouvons prendre $\lambda_k = \hat{\lambda}, \mu_k = 0$. Sinon, (3.19) doit être vérifiée en tant qu'égalité en x^{k+1}, i.e.

$$x^{k+1} = \text{argmin} f(x), \quad \text{s.c. } x \in F_k \bigcap G_k,$$

dans quel cas $x^{k+1} = \bar{x}$, et nous pouvons prendre λ_k, μ_k comme les multiplicateurs associés aux contraintes $x \in F_k$, $x \in G_k$ respectivement.

(ii) Observons d'abord que la condition d'une itération interne $g_{i(k)}(x^k) \leq 0$ et la proposition 3.4.2(v) impliquent que x^k satisfait (3.28). Notons aussi que, quand $z^k_{i(k)} = 0$, alors x^k satisfait (3.27). Ainsi, sous l'hypothèse $z^k_{i(k)} = 0$, x^k est admissible pour le problème (3.26)-(3.28), impliquant que $f(x^k) \geq f(x^{k+1})$. D'un autre coté, des propositions 3.4.3 et 3.3.1(i), nous avons $f(x^k) \leq f(x^{k+1})$, ainsi $f(x^k) = f(x^{k+1})$; donc par la proposition 3.3.1(i), $x^k = x^{k+1}$ et nous pouvons prendre, au vu de (3.38)-(3.42), $\lambda_k = 0, \mu_k = 1$. Maintenant, nous affirmons que, quand $z^k_{i(k)} \neq 0$, (3.27) doit être vérifiée en tant qu'égalité en x^{k+1}.

Sinon, (3.27) est superflue dans le problème (3.26)-(3.28), qui se réduit à min $f(x)$, s.c. (3.28). Comme x^k satisfait (3.28), il est admissible pour ce problème-ci, impliquant comme avant que $x^{k+1} = x^k$ dans quel cas x^k satisfait (3.27) qui devient

$$z^k_{i(k)}(\bar{b}^k_{i(k)} - \langle \bar{a}^{k,i(k)}, x^k \rangle) = 0.$$

Comme $z^k_{i(k)} \neq 0$, nous obtenons

$$\langle \bar{a}^{k,i(k)}, x^k \rangle = \bar{b}^k_{i(k)},$$

et puisque $x^k = x^{k+1}$, nous obtenons que (3.27) est vérifiée en tant qu'égalité en x^{k+1} ce qui contredit l'hypothèse d'absurdité.

Nous pouvons alors substituer (3.27) par $x \in \bar{G}_k$ dans le problème (3.26)-(3.28). Dans un tel cas, nous résolvons min $f(x)$ s.c. $x \in \bar{G}_k$ et obtenons $\hat{x}$ avec un multiplicateur $\hat{\mu}$ associé à la contrainte $x \in \bar{G}_k$. Si $\hat{x}$ satisfait (3.28), alors $x^{k+1} = \hat{x}$ et nous pouvons prendre $\lambda_k = 0$ et $\mu_k = \hat{\mu}$. Sinon, (3.28) doit aussi être vérifiée en tant qu'égalité et le problème (3.26)-(3.28) se réduit à min $f(x)$, s.c. $x \in \bar{F}_k \bigcap \bar{G}_k$ avec la solution $x^{k+1} = \bar{x}$ et λ_k, μ_k égaux aux multiplicateurs associés respectivement aux contraintes $x \in \bar{F}_k, x \in \bar{G}_k$.

□

La reformulation de l'algorithme primal-dual S utilise les résultats de la proposition 3.5.1 sous forme des conditions de Karush-Kuhn-Tucker des problèmes de minimisation correspondants et tient compte des remarques faites dans la section précédente.

Soit K l'ensemble des paires (i,j), $1 \leq i \leq m$, $1 \leq j \leq n$ telle que $\partial g_i / \partial x_j$ est constante sur $\mathbb{R}^n$.

La nouvelle version de l'algorithme primal-dual S est la suivante.

ALGORITHME PRIMAL-DUAL S

Initialisation

Soient $x^0 \in \mathbb{R}^n, z^0 \in \mathbb{R}^m, \bar{b}^0 \in \mathbb{R}^m, \bar{A}^0 \in \mathbb{R}^{m \times n}$ où $\bar{a}^{0,i}, 1 \leq i \leq m$, satisfont respectivement,

$$x^0 \in U, \tag{3.57}$$

$$\bar{a}^{0,i} = \nabla g_i(x^0), \quad 1 \leq i \leq m, \tag{3.58}$$

$$\bar{b}^0_i = \langle \bar{a}^{0,i}, x^0 \rangle - g_i(x^0), \quad 1 \leq i \leq m, \tag{3.59}$$

$$z^0 \geq 0, \tag{3.60}$$

$$\nabla f(x^0) = -(\bar{A}^0)^t z^0, \tag{3.61}$$

$$(z^0)^t(\bar{A}^0 x^0 - \bar{b}^0) = 0. \tag{3.62}$$

Itération
Soient $x^k, z^k, \bar{A}^k, \bar{b}^k$, effectuer les opérations suivantes.

(a) Si $g_{i(k)}(x^k) > 0 (étapeexterne)$, trouver $\hat{x}, \hat{\lambda}$ satisfaisant

$$\nabla f(\hat{x}) = -\hat{\lambda}\nabla g_{i(k)}(x^k), \tag{3.63}$$

$$\langle \nabla g_{i(k)}(x^k), \hat{x} - x^k \rangle = -\alpha_k g_{i(k)}(x^k). \tag{3.64}$$

Si $\langle \nabla f(x^k), \hat{x} - x^k \rangle \geq 0$, alors $x^{k+1} = \hat{x}, \lambda_k = \hat{\lambda}, \mu_k = 0$.
Si $\langle \nabla f(x^k), \hat{x} - x^k \rangle < 0$, trouver $x^{k+1}, \lambda_k, \mu_k$ résolvant le système suivant,

$$\nabla f(x^{k+1}) = \mu_k \nabla f(x^k) - \lambda_k \nabla g_{i(k)}(x^k), \tag{3.65}$$

$$\langle \nabla g_{i(k)}(x^k), x^{k+1} - x^k \rangle = -\alpha_k g_{i(k)}(x^k), \tag{3.66}$$

$$\langle \nabla f(x^k), x^{k+1} - x^k \rangle = 0. \tag{3.67}$$

Mettre à jour z^k comme suit,

$$z_i^{k+1} = \begin{cases} \mu_k z_i^k, & \text{si } i \neq i(k), \\ \mu_k z_i^k + \lambda_k, & \text{si } i = i(k). \end{cases} \tag{3.68}$$

Définir

$$\gamma_k = \begin{cases} 0, & \text{si } z_{i(k)}^{k+1} = 0, \\ \lambda_k / z_{i(k)}^{k+1}, & \text{sinon.} \end{cases} \tag{3.69}$$

Mettre à jour $\bar{A}^k, \bar{b}^k$ comme suit,

$$\bar{a}_j^{k+1,i} = \begin{cases} \bar{a}_j^{k,i}, & \text{si } i \neq i(k) \text{ ou } (i,j) \in K, \\ (1-\gamma_k)\bar{a}_j^{k,i} + \gamma_k \nabla g_i(x^k)_j, & \\ & \text{si } i = i(k) \text{ et } (i,j) \notin K, \end{cases} \tag{3.70}$$

$$\bar{b}_i^{k+1} = \begin{cases} \bar{b}_i^k, & \text{si } i \neq i(k), \\ (1-\gamma_k)\bar{b}_i^k & \\ +\gamma_k(\langle \nabla g_i(x^k), x^k \rangle - \alpha_k g_i(x^k)), & \text{sinon.} \end{cases} \tag{3.71}$$

(b) Si $g_{i(k)}(x^k) \leq 0$ (étape interne), alors procéder comme suit.
Si $z_{i(k)}^k = 0$, alors $x^{k+1} = x^k$, $\lambda_k = 0$, $\mu_k = 1$.
Si $z_{i(k)}^k \neq 0$, alors trouver $\hat{x}$, $\hat{\mu}$ satisfaisant

$$\nabla f(\hat{x}) = \hat{\mu}(\nabla f(x^k) + z_{i(k)}^k \bar{a}^{k,i(k)}), \tag{3.72}$$

$$\langle \nabla f(x^k), \hat{x} - x^k \rangle = z_{i(k)}^k (\bar{b}_{i(k)}^k - \langle \bar{a}^{k,i(k)}, \hat{x} \rangle). \tag{3.73}$$

Si $\langle \bar{a}^{k,i(k)}, \hat{x} \rangle \leq \bar{b}^k_{i(k)}$, alors $x^{k+1} = \hat{x}$, $\lambda_k = 0$, $\mu_k = \hat{\mu}$.
Si $\langle \bar{a}^{k,i(k)}, \hat{x} \rangle > \bar{b}^k_{i(k)}$, alors trouver x^{k+1}, λ_k, μ_k résolvant le système suivant,

$$\nabla f(x^{k+1}) = \mu_k(\nabla f(x^k) + z^k_{i(k)} \bar{a}^{k,i(k)}) - \lambda_k \bar{a}^{k,i(k)}, \tag{3.74}$$

$$\langle \nabla f(x^k), x^{k+1} - x^k \rangle = z^k_{i(k)}(\bar{b}^k_{i(k)} - \langle \bar{a}^{k,i(k)}, x^{k+1} \rangle), \tag{3.75}$$

$$\langle \bar{a}^{k,i(k)}, x^{k+1} \rangle = \bar{b}^k_{i(k)}. \tag{3.76}$$

Mettre à jour z^k et $\bar{A}^k, \bar{b}^k$ comme suit,

$$z^{k+1}_i = \begin{cases} \mu_k z^k_i, & \text{si } i \neq i(k), \\ \lambda_k, & \text{sinon.} \end{cases} \tag{3.77}$$

$$\bar{A}^{k+1} = \bar{A}^k, \tag{3.78}$$

$$\bar{b}^{k+1} = \bar{b}^k. \tag{3.79}$$

Observons qu'à chaque itération, la tâche calculatoire la plus élevée est soit la résolution d'un système non linéaire de $n+1$ équations à $n+1$ inconnues, soit la résolution d'un système non linéaire de $n+2$ équations à $n+2$ inconnues. Dans le premier cas, lors d'une itération externe (respectivement interne) le système correspondant à résoudre est celui défini par (3.63)-(3.64) avec $\hat{x} \in \mathbb{R}^n, \hat{\lambda} \in \mathbb{R}$ (respectivement (3.72)-(3.73) avec $\hat{x} \in \mathbb{R}^n, \hat{\mu} \in \mathbb{R}$). Dans le second cas, lors d'une itération externe (respectivement interne) le système à résoudre est défini par (3.65)-(3.67) (respectivement (3.74)-(3.76)) avec les inconnues $x^{k+1} \in \mathbb{R}^n, \lambda_k \in \mathbb{R}, \mu_k \in \mathbb{R}$. Donc la tâche calculatoire de l'algorithme primal-dual S est approximativement la même que dans l'algorithme S du chapitre précédent; la différence se marque dans les mises à jour duales (3.68)-(3.71) et (3.77) qui ne sont pas présentes dans l'algorithme S, lequel est purement primal. Donc la comparaison avec la méthode de Han donnée dans la section 2.5 tient aussi pour l'algorithme primal-dual S.

Dans les deux sections suivantes nous donnons des résultats de convergence dans le cas de **contraintes linéaires**. Nous considérons donc à partir de maintenant que les contraintes définies dans le problème (PPD) sont des inégalités linéaires.

3.6 Taux de convergence de l'algorithme primal-dual S

Nous prouvons dans cette section que, lorsque les contraintes sont des inégalités linéaires, lorsque f est deux fois continument différentiable à la solution x^* du problème (PPD), et que le hessien de f en x^* est non singulier, la suite $\{x^k\}$ converge au moins linéairement.

Supposons que (PPD) soit de la forme

$$\begin{cases} \min & f(x) \\ \text{s.c.} & Ax \leq b, \end{cases}$$

avec $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$. Soient a^i, $1 \leq i \leq m$ les lignes de A. Ainsi,

$$C = \{x \in \mathbb{R}^n : Ax \leq b\}, \quad C_i = \{x \in \mathbb{R}^n : \langle a^i, x \rangle \leq b_i\}.$$

Comme nous l'avons vu dans la troisième remarque de la section 3.4,

$$\bar{A}^k = A^k = A. \tag{3.80}$$

Nous faisons aussi l'hypothèse simplificatrice que $\alpha_k = 1$ pour tout k, i.e. nous travaillons avec la version non relaxée de l'algorithme primal-dual S. Remplaçant (3.80) et $g_i(x) = \langle a^i, x \rangle - b_i$ dans (3.11), (3.17), nous obtenons $b^k = b$ pour tout k et, de (3.11), (3.12), $\bar{b}^0 = b^0 = b$. Il découle de (3.25) que, pour tout $k \geq 0$,

$$\bar{b}^k = b^k = b. \tag{3.81}$$

Par conséquent, l'itération externe se réduit à (3.18)-(3.22) et l'itération interne à (3.26)-(3.30), avec A (respectivement b) au lieu de A^k, $\bar{A}^k$ (respectivement $b^k, \bar{b}^k$).

Le résultat de convergence linéaire est considérable parce que dans ce cas-ci l'algorithme primal S peut converger de manière sous-linéaire. Donc, l'ajout de la suite duale implique un meilleur comportement de convergence. Rappellons que l'algorithme primal S consiste en la mise à jour (3.18)-(3.20) dans le cas d'une itération externe, et $x^{k+1} = x^k$ dans le cas d'une itération interne.

Exemple: Soient $m = 3, \quad n = 2, \quad i(k) = (1, 2, 3, 1, 2, 3, \ldots),$

$$f(x) = (1/2)\|x - u\|^2, \quad \text{où } u^t = (3, 0).$$

Soit le système $Ax \leq b$ défini par

$$A = \begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad b = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

et avec $(x^0)^t = (1/2, 0)$. Il suit que $x^* = (0, 0)^t$. Effectuons les trois premières itérations avec l'algorithme primal S:

Pour l'initialisation, nous connaissons x^0 par l'énoncé. Nous avons aussi A et b.

La première itération est réalisée avec $i(k) = 1$. Comme $g_1(x^0) = 1/2 > 0$, nous sommes dans une étape externe. Il faut donc résoudre le système suivant

$$\begin{cases} \nabla f(\hat{x}) = -\hat{\lambda} \nabla g_1(x^0), \\ \langle \nabla g_1(x^0), \hat{x} - x^0 \rangle = -g_1(x^0). \end{cases}$$

et nous obtenons

$$\hat{x} = (3/2, 3/2), \quad \hat{\lambda} = 3/2.$$

Comme $\langle \nabla f(x^0), \hat{x} - x^0 \rangle < 0$ nous devons résoudre le système suivant

$$\begin{cases} \nabla f(x^1) = \mu_0 \nabla f(x^0) - \lambda_0 \nabla g_1(x^0), \\ \langle \nabla g_1(x^0), x^1 - x^0 \rangle = -g_1(x^0), \\ \langle \nabla f(x^0), x^1 - x^0 \rangle = 0, \end{cases}$$

et nous obtenons

$$\hat{x} = (1/2, 1/2), \quad \mu_0 = 4/5, \quad \lambda_0 = 1/2.$$

Passons ensuite à la seconde itération avec $i(k) = 2$.

Comme $g_2(x^1) = 1/2 > 0$, nous sommes encore dans une étape externe. Il faut donc résoudre le système suivant

$$\begin{cases} \nabla f(\hat{x}) = -\hat{\lambda}\nabla g_2(x^1), \\ \langle \nabla g_2(x^1), \hat{x} - x^1 \rangle = -g_2(x^1), \end{cases}$$

et nous obtenons

$$\hat{x} = (3, 0), \quad \hat{\lambda} = 0$$

Comme $\langle \nabla f(x^1), \hat{x} - x^1 \rangle < 0$ nous devons résoudre le système suivant

$$\begin{cases} \nabla f(x^2) = \mu_1 \nabla f(x^1) - \lambda_1 \nabla g_2(x^1), \\ \langle \nabla g_2(x^1), x^2 - x^1 \rangle = -g_2(x^1), \\ \langle \nabla f(x^1), x^2 - x^1 \rangle = 0, \end{cases}$$

et nous obtenons

$$\hat{x} = (2/5, 0), \quad \mu_1 = 26/25, \quad \lambda_1 = 13/25.$$

La troisième itération avec $i(k) = 3$ quant à elle, est une étape interne cette fois et donc

$$x^3 = x^2.$$

Par récurrence il peut être prouvé que x^{3k} est de la forme $(c, 0)^t$, avec $0 \leq c < 1$. Il suit alors que $x^{3k+1} = (c, c)^t$, $x^{3k+2} = (1 - c/(3 - c))x^{3k}$, et $x^{3k+3} = x^{3k+2}$, ainsi

$$\|x^{3k+3} - x^*\| = (1 - x_1^{3k}/(3 - x_1^{3k}))\|x^{3k} - x^*\|.$$

Comme $\lim_{k\to\infty} x_1^{3k} = 0$ par le thèorème 3.4.1, nous obtenons

$$\lim_{k\to\infty} \|x^{3k+3} - x^*\|/\|x^{3k} - x^*\| = 1.$$

et la convergence est sous-linéaire.

Si nous prenons comme points de départ de l'algorithme primal-dual S le même x^0 et $z^0 = (5/4, 5/4, 5/4)^t$, les deux points vérifient (3.60)-(3.62). Les itérations de la forme $3k + 3$ sont des itérations internes, et nous n'avons plus toujours que $x^{3k+3} = x^{3k+2}$. En fait dans le cas de l'exemple nous obtenons $x^{3k+3} = x^*$ et la convergence est finie.

La convergence finie de cet exemple résulte du fait que les dimensions choisies ici sont petites, mais les résultats de cette section montrent que pour avoir la convergence linéaire, les itérations internes (principale différence entre les algorithmes primal S et primal-dual S pour la suite primale) sont essentielles.

Cet exemple montre que si nous laissons l'itéré courant inchangé lorsqu'il satisfait la contrainte courante, alors la convergence peut être sous-linéaire.

Nous commençons la preuve de convergence linéaire avec deux conséquences des formules de mises à jour de l'algorithme primal-dual S.

Proposition 3.6.1 *Si $z_i^k = 0$ et $l > k$ est tel que $i(j) \neq i$ pour j entre k et $l - 1$, alors $z_i^l = 0$.*

Preuve

$i(j) \neq i$ implique que $e_i^{i(j)} = 0$. De (3.22) ou (3.30), $z_i^{j+1} = \mu_j z_i^j$ pour j entre k et $l-1$. Par conséquent,

$$z_i^l = \left(\prod_{j=k}^{l-1} \mu_j \right) z_i^k = 0.$$

□

Proposition 3.6.2 *Si la k ème itération est une étape externe, alors* $\langle a^{i(k)}, x^{k+1} \rangle = b_{i(k)}$.

Preuve

Comme il a été vu dans la preuve de la proposition 3.5.1(i), pour des itérations externes (3.20) est vérifiée en tant qu'égalité en x^{k+1}. Il suffit alors d'utiliser (3.80) et (3.81), et le résultat suit.

□

Ensuite, nous approximons l'algorithme dans un voisinage de $x^* = \lim_{k \to \infty} x^k$. L'existence et l'optimalité de x^* pour le problème min $f(x)$, s.c. $x \in C$ sont garanties par le théorème 3.4.1.

Soit

$$Q = \nabla^2 f(x^*). \tag{3.82}$$

Nous supposons que Q est définie positive.

Proposition 3.6.3 *Pour tout* $x, y \in \mathbb{R}^n$, *il existe* $\eta(x,y) \in \mathbb{R}$ *tel que*

$$f(x) = f(y) + \langle \nabla f(y), x - y \rangle + (1/2)(1 + \eta(x,y)) \|x - y\|_Q^2, \tag{3.83}$$

$$|\eta(x,y)| \leq \|\nabla^2 f(\hat{y}) - \nabla^2 f(x^*)\|_Q, \tag{3.84}$$

pour $\hat{y}$ *dans le segment entre* x *et* y.

Preuve

Le développement de Taylor du premier ordre de f autour de y s'écrit

$$f(x) = f(y) + \langle \nabla f(y), x - y \rangle + (1/2)(x - y)^t \nabla^2 f(\hat{y})(x - y), \tag{3.85}$$

pour $\hat{y}$ dans le segment entre x et y. De (3.85) et (3.82),

$$\begin{aligned} f(x) &= f(y) + \langle \nabla f(y), x - y \rangle + (1/2)(x - y)^t Q (x - y) \\ &+ (1/2)(x - y)^t (\nabla^2 f(\hat{y}) - \nabla^2 f(x^*))(x - y) \\ &= f(y) + \langle \nabla f(y), x - y \rangle + (1/2)(1 + \eta(x,y)) \|x - y\|_Q^2. \end{aligned} \tag{3.86}$$

où

$$\eta(x,y) = (x - y)^t (\nabla^2 f(\hat{y}) - \nabla^2 f(x^*))(x - y) / \|x - y\|_Q^2. \tag{3.87}$$

Prenant la norme Q sur les deux membres de (3.87), nous avons

$$|\eta(x,y)| \leq \|\nabla^2 f(\hat{y}) - \nabla^2 f(x^*)\|_Q. \tag{3.88}$$

□

Soit

$$\eta_k = \eta(x^*, x^k), \tag{3.89}$$

$$\tilde{\eta}_k = \eta(x^{k+1}, x^k). \tag{3.90}$$

Proposition 3.6.4 *Les résultats suivants ont lieu,*
(i) pour tout $k \geq 0$, $f(x^{k+1}) \geq f(x^k) + (1/2)(1+\tilde{\eta}_k)\|x^{k+1} - x^k\|_Q^2$,
(ii) pour k *assez grand,* $f(x^*) = f(x^k) + (1/2)(1+\eta_k)\|x^* - x^k\|_Q^2$.

Preuve

(i) Il suffit d'appliquer la proposition 3.6.3 avec $x = x^{k+1}, y = x^k$, puis d'utiliser pour l'itération externe (3.19), et pour l'itération interne la proposition 3.4.2(i), (3.27), et (3.28).

(ii) Il suffit d'appliquer la proposition 3.6.3 avec $x = x^*, y = x^k$. Puis de vérifier que $\langle \nabla f(x^k), x^* - x^k \rangle = 0$. Par la proposition 3.4.2(ii), nous obtenons après multiplication par $x^* - x^k$,

$$\begin{aligned} \langle \nabla f(x^k), x^* - x^k \rangle &= (z^k)^t A(x^k - x^*) \\ &= (z^k)^t (Ax^k - b) + (z^k)^t (b - Ax^*) \\ &= \sum_{i=1}^{m} z_i^k (b_i - \langle a^i, x^* \rangle), \end{aligned}$$

en utilisant la proposition 3.4.2(iii) dans la dernière égalité. Par le corollaire 3.4.2 pour k assez grand, $z_i^k = 0$ quand $\langle a^i, x^* \rangle \neq b_i$. Le résultat est donc établi.

□

Soit

$$\bar{\eta}_k = \max_{k \leq j \leq k+r-1} \tilde{\eta}_j \tag{3.91}$$

où r est la constante de presque cyclicité.

La proposition suivante établit la propriété de décroissance de la suite $\{x^k\}$.

Proposition 3.6.5 *Pour* k *assez grand,*

$$\begin{aligned} &(1+\eta_{k+r})\|x^* - x^{k+r}\|_Q^2 \\ &\leq (1+\eta_k)\|x^* - x^k\|_Q^2 - (1/r)(1+\bar{\eta}_k)\left(\sum_{j=k}^{k+r-1} \|x^{j+1} - x^j\|_Q\right)^2. \end{aligned} \tag{3.92}$$

Preuve

De la proposition 3.6.4(i), il suit que

$$f(x^*) - f(x^{k+1}) \leq f(x^*) - f(x^k) - (1/2)(1+\tilde{\eta}_k)\|x^{k+1} - x^k\|_Q^2.$$

Itérant cette inégalité, nous obtenons

$$\begin{aligned} f(x^*) - f(x^{k+r}) &\leq f(x^*) - f(x^k) - (1/2)\sum_{j=k}^{k+r-1}(1+\tilde{\eta}_j)\|x^{j+1} - x^j\|_Q^2 \\ &\leq f(x^*) - f(x^k) - (1/2r)(1+\bar{\eta}_k)\left(\sum_{j=k}^{k+r-1}\|x^{j+1} - x^j\|_Q\right)^2. \end{aligned}$$

Le résultat suit en appliquant la proposition 3.6.4(ii) aux deux extrêmités de l'inégalité ci-dessus.

□

Soit

$$d_Q(x,S) = \inf_{y\in S}\|x-y\|_Q, \quad S \subset \mathbb{R}^n.$$

Définissons encore

$$H_i = \{x \in \mathbb{R}^n : \langle a^i, x\rangle = b_i\}, \quad I_k = \{i : z_i^k = 0\}, \quad J_k = \{i : z_i^k > 0\},$$

et

$$V_k = \left(\bigcap_{i\in I_k} C_i\right) \bigcap \left(\bigcap_{i\in J_k} H_i\right).$$

Proposition 3.6.6 *Pour k assez grand, $x^* \in V_k$.*

Preuve

Comme $x^* \in C$, x^* appartient à C_i pour tout i. Considérons i tel que $x^* \notin H_i$, i.e., $\langle a^i, x^*\rangle < b_i$. Par le corollaire 3.4.2, $z_i^k = 0$, i.e., $i \in I_k$. Il suit que $x^* \in \bigcap_{i\in J_k} H_i$.

□

Proposition 3.6.7 *Il existe $\beta_k > 0$ tel que, pour tout $x \notin V_k$,*

$$\max\left\{\max_{i\in I_k} d_Q(x,C_i), \max_{i\in J_k} d_Q(x,H_i)\right\} \geq \beta_k d_Q(x,V_k). \tag{3.93}$$

Preuve

Une preuve de cette proposition peut être trouvée dans [3] (Proposition 5.7 p100-101).

□

Nous remarquons que β_k peut prendre seulement un ensemble fini de valeurs, car le nombre de paires (I_k, J_k), et par conséquent le nombre d'ensembles V_k, est borné par 2^n. Ainsi, si nous définissons β comme

$$\beta = \min_{k \geq 0} \beta_k, \tag{3.94}$$

nous obtenons que $\beta > 0$ car $\beta_k > 0$, et nous utilisons β dans (3.93). Nous énonçons ce fait dans le corollaire suivant.

Corollaire 3.6.1 *Pour tout $k > 0$ et pour tout $x \notin V_k$,*

$$\max \left\{ \max_{i \in I_k} d_Q(x, C_i), \max_{i \in J_k} d_Q(x, H_i) \right\} \geq \beta d_Q(x, V_k), \tag{3.95}$$

avec β comme dans (3.94)

Preuve

Ceci découle directement de la proposition 3.6.6 et de la discussion précédente.

□

Nous utilisons le corollaire 3.6.1 pour trouver une borne inférieure appropriée pour le second terme du membre de droite de (3.92). Nous le faisons en deux étapes, dans les deux propositions suivantes. Définissons

$$\hat{x}^k = \text{argmin}_{x \in V_{k+r}} \|x - x^{k+r}\|_Q, \tag{3.96}$$

$$\hat{\eta}_k = \eta(\hat{x}^k, x^{k+r}), \tag{3.97}$$

avec η comme dans (3.84), où r est la constante de presque cyclicité.

Proposition 3.6.8 *Pour k assez grand,*

$$\|\hat{x}^k - x^{k+r}\|_Q \geq ((1 + \eta_{k+r})/(1 + \hat{\eta}_k))^{1/2} \|x^* - x^{k+r}\|_Q. \tag{3.98}$$

Preuve

Observons que

$$V_{k+r} \subset \bigcap_{i=1}^{m} C_i = C,$$

car $H_i \subset C_i$ pour tout i. Et comme $\hat{x}^k \in V_{k+r}$, nous obtenons que $\hat{x}^k \in C$, impliquant que

$$f(x^*) \leq f(\hat{x}^k). \tag{3.99}$$

De la proposition 3.6.4(ii), nous avons

$$f(x^*) = f(x^{k+r}) + (1/2)(1 + \eta_{k+r}) \|x^* - x^{k+r}\|_Q^2. \tag{3.100}$$

Appliquons la proposition 3.6.3, avec $y = x^{k+r}, x = \hat{x}^k$, pour trouver

$$f(\hat{x}^k) = f(x^{k+r}) + \langle \nabla f(x^{k+r}), \hat{x}^k - x^{k+r} \rangle + (1/2)(1 + \hat{\eta}_k) \|\hat{x}^k - x^{k+r}\|_Q^2. \tag{3.101}$$

Le second terme de (3.101) peut se réécrire comme:

$$\begin{aligned}\langle \nabla f(x^{k+r}), \hat{x}^k - x^{k+r} \rangle &= (z^{k+r})^t A(x^{k+r} - \hat{x}^k) \\ &= (z^{k+r})^t (Ax^{k+r} - b) + (z^{k+r})^t (b - A\hat{x}^k) \\ &= \sum_{i \in J_{k+r}} z_i^{k+r} (b_i - \langle a^i, \hat{x}^k \rangle), \end{aligned} \tag{3.102}$$

en utilisant la proposition 3.4.2(iii) dans la dernière égalité de (3.102). Par (3.96) et la définition de V_k, nous avons, pour tout $i \in J_{k+r}$,

$$\hat{x}^k \in H_i, \quad \text{i.e. } \langle a^i, \hat{x}^k \rangle = b_i.$$

Ainsi, le membre de droite de (3.102) est nul, impliquant par (3.101) que

$$f(\hat{x}^k) = f(x^{k+r}) + (1/2)(1 + \hat{\eta}_k)\|\hat{x}^k - x^{k+r}\|_Q^2. \tag{3.103}$$

Le résultat suit de (3.100), (3.103) placés dans (3.99).

□

Proposition 3.6.9 *Pour k assez grand,*

$$\beta \|\hat{x}^k - x^{k+r}\|_Q \leq \sum_{j=k}^{k+r-1} \|x^{j+1} - x^j\|_Q, \tag{3.104}$$

avec β comme dans (3.94).

Preuve

Soit l l'index où le maximum du membre de gauche de (3.95) est atteint. Définissons q comme

$$q = \max\{j : k \leq j \leq k + r - 1, i(j) = l\}. \tag{3.105}$$

Nous considérons deux cas.

(i) $l \in I_{k+r}$. Du corollaire 3.6.1 et de (3.96),

$$d_Q(x^{k+r}, C_l) \geq \beta d_Q(x^{k+r}, V_{k+r}) = \beta \|\hat{x}^k - x^{k+r}\|_Q, \tag{3.106}$$

avec β comme dans (3.94). De (3.35) ou (3.40), $x^{q+1} \in C_l$, impliquant que

$$\|x^{k+r} - x^{q+1}\|_Q \geq d_Q(x^{k+r}, C_l). \tag{3.107}$$

De (3.106), (3.107), nous avons

$$\begin{aligned}\beta \|\hat{x}^k - x^{k+r}\|_Q &\leq \|x^{k+r} - x^{q+1}\|_Q \\ &\leq \sum_{j=q+1}^{k+r-1} \|x^{j+1} - x^j\|_Q \\ &\leq \sum_{j=k}^{k+r-1} \|x^{j+1} - x^j\|_Q. \end{aligned} \tag{3.108}$$

(ii) $l \in J_{k+r}$. Par (3.35) ou (3.40), $x^{q+1} \in C_l$. Nous affirmons que $x^{q+1} \in H_l$.

Sinon, i.e. $\langle a^l, x^{q+1}\rangle < b_l$. Par la proposition 3.6.2, une étape interne était prise à l'itération q, signifiant par (3.30) que

$$z_l^{q+1} = \mu_q z_l^q + \lambda_q e_l^l - \mu_q z_l^q = \lambda_q.$$

Comme la contrainte (3.20) n'est pas vérifiée en tant qu'égalité, nous avons $\lambda_q = 0$ et par conséquent $z_l^{q+1} = 0$. Utilisant la proposition 3.6.1 et (3.105), nous concluons que

$$z_l^{k+r} = 0, \quad \text{i.e.}, l \in I_{k+r},$$

contredisant le fait que $l \in J_{k+r}$.

Comme le maximum du membre de gauche de (3.95) est atteint à l'itération l, utilisant $x^{q+1} \in H_l$, le corollaire 3.6.1, et (3.96), nous avons

$$\begin{aligned} \|x^{k+r} - x^{q+1}\|_Q &\geq d_Q(x^{k+r}, H_l) \\ &\geq \beta d_Q(x^{k+r}, V_{k+r}) \\ &= \beta \|\hat{x}^k - x^{k+r}\|_Q, \end{aligned} \tag{3.109}$$

et nous obtenons le résultat exactement comme dans (3.108).

□

Finalement, nous présentons le résultat du taux de convergence.

Théorème 3.6.1 *Si f est deux fois continument différentiable au point limite x^* de la suite $\{x^k\}$ et si le Hessien de f en x^* est non singulier, alors la suite $\{x^k\}$ définie par le primal-dual S converge linéairement vers la solution x^* de min $f(x)$, s.c. $Ax \leq b$, avec une erreur asymptotique constante égale à $(1+\beta^2/r)^{-1/2} < 1$, où β est comme dans (3.94) et r est la constante de presque cyclicité, i.e.,*

$$\text{limsup}_{k\to\infty} \|x^{k+r} - x^*\|_Q / \|x^k - x^*\|_Q \leq (1 + \beta^2/r)^{-1/2}. \tag{3.110}$$

Preuve

Par la proposition 3.6.5, et par la proposition 3.6.9, pour k assez grand, nous avons

$$\begin{aligned} &(1 + \eta_{k+r})\|x^* - x^{k+r}\|_Q^2 \\ &\quad \leq (1 + \eta_k)\|x^* - x^k\|_Q^2 - (\beta^2/r)(1 + \bar{\eta}_k)\|\hat{x}^k - x^{k+r}\|_Q^2. \end{aligned}$$

Par la proposition 3.6.8 il suit que,

$$\begin{aligned} &(1 + \eta_{k+r})\|x^* - x^{k+r}\|_Q^2 \\ &\quad \leq (1 + \eta_k)\|x^* - x^k\|_Q^2 - (\beta^2/r)\left[\frac{(1 + \bar{\eta}_k)(1 + \eta_{k+r})}{(1 + \hat{\eta}_k)}\right]\|x^* - x^{k+r}\|_Q^2. \end{aligned}$$

Réarrangeant les termes, nous obtenons

$$\begin{aligned} &\|x^* - x^{k+r}\|_Q / \|x^* - x^k\|_Q \\ &\quad \leq [(1 + \eta_k)/(1 + \eta_{k+r})]^{1/2} \times \{1 + (\beta^2/r)[(1 + \bar{\eta}_k)/(1 + \hat{\eta}_k)]\}^{-1/2}. \end{aligned} \tag{3.111}$$

Nous évaluons ensuite les limites de η_k, $\bar{\eta}_k$, $\hat{\eta}_k$ lorsque $k \to \infty$. De (3.88), (3.89),

$$|\eta_k| \leq \|\nabla^2 f(y^k) - \nabla^2 f(x^*)\|_Q$$

avec y^k entre x^k et x^*. Puisque $\lim_{k\to\infty} x^k = x^*$, nous obtenons $\lim_{k\to\infty} y^k = x^*$, impliquant que $\lim_{k\to\infty} \|\nabla^2 f(y^k) - \nabla^2 f(x^*)\|_Q = 0$, et par conséquent

$$\lim_{k\to\infty} \eta_k = 0.$$

De (3.88), (3.90),

$$|\tilde{\eta}_k| \leq \|\nabla^2 f(y^k) - \nabla^2 f(x^*)\|_Q,$$

avec y^k entre x^k et x^{k+1}. Dans ce cas-ci aussi nous avons $\lim_{k\to\infty} y^k = x^*$, et par conséquent $\lim_{k\to\infty} \tilde{\eta}_k = 0$. De (3.91), comme r est fixé,

$$\lim_{k\to\infty} \bar{\eta}_k = 0.$$

De (3.88), (3.97),

$$|\hat{\eta}_k| \leq \|\nabla^2 f(y^k) - \nabla^2 f(x^*)\|_Q,$$

avec y^k entre x^{k+r} et $\hat{x}^k$. Par la proposition 3.6.9, comme $\lim_{k\to\infty}(x^{k+1} - x^k) = 0$, nous obtenons que $\lim_{k\to\infty} \hat{x}^k = \lim_{k\to\infty} x^{k+r} = x^*$. Il suit que $\lim_{k\to\infty} y^k = x^*$ et par conséquent,

$$\lim_{k\to\infty} \hat{\eta}_k = 0.$$

Le résultat suit en prenant la limite lorsque $k \to \infty$ dans (3.111).

□

3.7 Convergence de la suite duale

Dans cette section, nous prouvons la convergence de la suite duale $\{z^k\}$ vers un vecteur optimal z^* des multiplicateurs de Karush-Kuhn-Tucker pour le problème min $f(x)$, s.c. $Ax \leq b$. Si les contraintes actives en x^* sont linéairement indépendantes, c'est immédiat. En effet, soit

$$J = \{i : \langle a^i, x^* \rangle = b_i\},$$

et soit A_J la matrice formée des lignes $i \in J$ de A. Par la proposition 3.4.2(ii) et le corollaire 3.4.2, pour k assez grand,

$$\nabla f(x^k) = -A_J^t z_J^k,$$

où z_J^k est le vecteur formé des composantes $i \in J$ de z^k. Comme A_J est de rang plein par rapport à ses lignes,

$$z_J^k = -(A_J A_J^t)^{-1} A_J \nabla f(x^k),$$

impliquant que

$$\lim_{k\to\infty} z_J^k = -(A_J A_J^t)^{-1} A_J \nabla f(x^*),$$

et par conséquent la suite $\{z^k\}$ converge, parce que $z_i^k = 0$ pour $i \notin J$ et pour k assez grand. Dans ce cas-ci, $\{z^k\}$ a aussi un taux de convergence linéaire.

Le cas compliqué survient lorsque A_J n'est pas de rang plein par rapport à ses lignes. L'idée est alors de borner λ_k et μ_k par un multiple de $\|x^k - x^*\|$, pour conclure par le théorème 3.6.1 que

$$\sum_{k=0}^{\infty} \lambda_k < \infty, \quad \sum_{k=0}^{\infty} \mu_k < \infty,$$

et ensuite à partir de (3.22) et (3.30) obtenir la convergence de la suite $\{z^k\}$.

Afin d'obtenir la borne de λ_k et μ_k, nous avons besoin de quatre résultats préparatoires.

Proposition 3.7.1 *S'il existe $\xi \in \mathbb{R}$, $\xi \leq 0$, et l tel que $\langle a^l, x^* \rangle = b_l$ et $\nabla f(x^*) = \xi a^l$, alors la convergence est finie.*

Preuve

Soit $j = \min\{k : i(k) = l\}$.
Alors $x^{j+1} = x^*$, avec $\mu_j = 0, \lambda_j = -\xi$. Ce choix satisfait automatiquement soit (3.33) soit (3.38); (3.35) et (3.40) sont satisfaits car $x^* \in C_l$. Comme $\langle a^l, x^* \rangle = b_l$, (3.36) et (3.41) ont lieu et le membre de droite de (3.39) disparaît; (3.34) et (3.39) ont lieu parce que par la proposition 3.4.3, $\langle \nabla f(x^k), x^* - x^k \rangle \geq 0$, pour tout k. Finallement, $\mu_j = 0$ implique que (3.37) et (3.42) sont satisfaits. Nous concluons que x^* satisfait les conditions d'optimalité de (3.18)-(3.20) et (3.26)-(3.28), et par conséquent $x^{j+1} = x^*$ quelque soit l'étape (interne ou externe) prise à l'itération j.

□

Proposition 3.7.2 *S'il existe $\xi > 0$ et $l \in \{1, \ldots, m\}$ tels que $\nabla f(x^*) = \xi a^l$, alors $z_l^k = 0$ pour k assez grand.*

Preuve

Une preuve de cette proposition peut être trouvée dans [3] (proposition 6.2 pages 106-107).

□

Proposition 3.7.3 *Si à l'itération k une étape interne est prise et si $z_{i(k)}^k = 0$, alors $z^{k+1} = z^k$.*

Preuve

(3.38)-(3.42) sont satisfaits avec $x^{k+1} = x^k, \mu_k = 1, \lambda_k = 0$ impliquant que $z^{k+1} = z^k$ par (3.30).

□

Au vu de la proposition 3.7.1 nous pouvons supposer que

$$\nabla f(x^*) \neq \xi a^l, \quad \forall l \in \{1, \ldots, m\} \text{ et } \quad \forall \xi \leq 0,$$

dans l'étude des propriétés de convergence de la suite $\{z^k\}$. Nous pouvons aussi omettre de considérer les contraintes l telles que

$$\nabla f(x^*) = \xi a^l, \quad \text{avec } \xi > 0 \text{ ou } \langle a^l, x^* \rangle < b_l,$$

signifiant aussi que les itérations k telle que $i(k) = l$ pour un tel l peuvent être omises, parce que de telles itérations n'ont pas d'effet sur la suite $\{z^k\}$ pour k assez grand. Définissons $\tilde{\mu}_k, \tilde{\lambda}_k$ comme suit,

$$\tilde{\mu}_k = 1 - \mu_k \tag{3.112}$$

$$\tilde{\lambda}_k = \begin{cases} \lambda_k, & \text{si la } k\text{ème itération est externe,} \\ \lambda_k - \mu_k z^k_{i(k)}, & \text{sinon.} \end{cases} \tag{3.113}$$

Il suit de (3.33), (3.38) que, pour tout k,

$$\nabla f(x^k) - \nabla f(x^{k+1}) = \tilde{\mu}_k \nabla f(x^k) + \tilde{\lambda}_k a^{i(k)}, \tag{3.114}$$

et de (3.22), (3.30)

$$z^{k+1} - z^k = -\tilde{\mu}_k z^k + \tilde{\lambda}_k e^{i(k)}. \tag{3.115}$$

Proposition 3.7.4 *Pour k assez grand,*

$$\tilde{\mu}_k = \frac{(\nabla f(x^k) - \nabla f(x^{k+1}))^t (\|a^{i(k)}\|_2^2 \nabla f(x^k) - \langle a^{i(k)}, \nabla f(x^k) \rangle a^{i(k)})}{\|\nabla f(x^k)\|_2^2 \|a^{i(k)}\|_2^2 - \langle \nabla f(x^k), a^{i(k)} \rangle^2}, \tag{3.116}$$

$$\tilde{\lambda}_k = \frac{(\nabla f(x^k) - \nabla f(x^{k+1}))^t (\|\nabla f(x^k)\|_2^2 a^{i(k)} - \langle a^{i(k)}, \nabla f(x^k) \rangle \nabla f(x^k))}{\|\nabla f(x^k)\|_2^2 \|a^{i(k)}\|_2^2 - \langle \nabla f(x^k), a^{i(k)} \rangle^2}. \tag{3.117}$$

Preuve

Multiplier (3.114) une première fois par $\nabla f(x^k)$ et ensuite par $a^{i(k)}$, pour obtenir un système de deux équations linéaires en $\tilde{\lambda}_k, \tilde{\mu}_k$. Si le dénominateur de (3.116) et (3.117) n'est pas nul, les membres de droite de (3.116), (3.117) sont les solutions d'un tel système. Observons que, si nous considérons une sous-suite $\{x^{j_k}\}$ de $\{x^k\}$ avec $i(j_k) = l$ pour un l quelconque alors si nous prenons la limite lorsque $k \to \infty$ du dénominateur nous obtenons $\|\nabla f(x^*)\|_2^2 \|a^l\|_2^2 - \langle \nabla f(x^*), a^l \rangle^2$, qui est plus grand que zéro à moins que

$$\nabla f(x^*) = \xi a^l, \quad \xi \in \mathbb{R},$$

comme discuté avant, nous pouvons supposer qu'un tel cas ne se produit pas. Ainsi, le dénominateur est positif pour k assez grand et le résultat a lieu.

□

Proposition 3.7.5 *Soient*

$$\sigma = 2 \max\{2\|\nabla f(x^*)\|_2^2 \max_{1 \le i \le m} \|a^i\|_2, 2\|\nabla f(x^*)\|_2 \max_{1 \le i \le m} \|a^i\|_2^2\},$$

$$\nu = (1/2) \min_{1 \le i \le m} \{\|\nabla f(x^*)\|_2^2 \|a^i\|_2^2 - \langle \nabla f(x^*), a^i \rangle^2\}.$$

Alors, pour k assez grand,

$$\max\{|\tilde{\lambda}_k|, |\tilde{\mu}_k|\} \le (4\sigma/\nu)\|Q\|_2 \|x^k - x^*\|_2. \tag{3.118}$$

Preuve

Nous pouvons supposer $\nu > 0$, car nous omettons les contraintes l telles que $\nabla f(x^*) = \xi a^l$, $\forall \xi \in \mathbb{R}$. Comme $\lim_{k\to\infty} \nabla f(x^k) = \nabla f(x^*)$, la norme du dénominateur dans (3.116), (3.117) est plus grande que ν pour k assez grand, et la norme des seconds facteurs dans les numérateurs est plus petite que σ. Ainsi,

$$\begin{aligned}\max\{|\tilde{\lambda}_k|, |\tilde{\mu}_k|\} &\leq (\sigma/\nu)\|\nabla f(x^k) - \nabla f(x^{k+1})\|_2 \\ &\leq (\sigma/\nu)(\|\nabla f(x^k) - \nabla f(x^*)\|_2 + \|\nabla f(x^{k+1}) - \nabla f(x^*)\|_2). \end{aligned} \tag{3.119}$$

Etendant $\nabla f(x)$ autour de x^*, nous avons

$$\nabla f(x^k) = \nabla f(x^*) + \langle \nabla^2 f(y^k), x^k - x^* \rangle,$$

avec y^k entre x^k et x^*; comme $\lim_{k\to\infty} x^k = x^*$ implique $\lim_{k\to\infty} \nabla^2 f(y^k) = \nabla^2 f(x^*) = Q$, nous obtenons, pour k assez grand,

$$\|\nabla f(x^k) - \nabla f(x^*)\|_2 \leq 2\|Q\|_2 \|x^k - x^*\|_2. \tag{3.120}$$

Le résultat suit en utilisant (3.120) dans (3.119), et en observant que $\|x^{k+1} - x^*\|_2$ est plus petit que $\|x^k - x^*\|_2$ pour k assez grand par le théorème 3.6.1.

□

Corollaire 3.7.1 *Nous avons,*

$$\sum_{k=0}^{\infty} \tilde{\mu}_k < \infty, \quad \sum_{k=0}^{\infty} \tilde{\lambda}_k < \infty.$$

Preuve

Par le théorème 3.6.1, $\{\|x^k - x^*\|_Q\}$ converge vers zéro avec un taux linéaire de convergence. Comme

$$\|x^k - x^*\|_2 \leq \tau \|x^k - x^*\|_Q, \quad \forall \tau > 0,$$

il découle que $\{\|x^k - x^*\|_2\}$ a un taux linéaire de convergence, i.e.,

$$\limsup_{k\to\infty} (\|x^k - x^*\|_2)^{1/k} \leq \limsup_{k\to\infty} (\tau \|x^k - x^*\|_Q)^{1/k} < 1,$$

impliquant que

$$\sum_{k=0}^{\infty} \|x^k - x^*\|_2 < \infty,$$

et le résultat suit de la proposition 3.7.5.

□

Afin de prouver la convergence de la suite duale nous avons encore besoin d'un résultat qui est énoncé dans la proposition suivante.

Proposition 3.7.6 *Considérons une suite* $\{\pi_k\} \subset \mathbb{R}^+$ *qui satisfait la récurrence*

$$\pi_{k+1} = (1 - \delta_k)\pi_k + \epsilon_k,$$

où

$$\delta_k \leq 1, \quad \sum_{k=0}^{\infty} |\delta_k| < \infty, \quad \sum_{k=0}^{\infty} |\epsilon_k| < \infty,$$

alors, la suite $\{\pi_k\}$ *converge.*

Preuve

Premièrement, π_k est bornée. En effet, par la définition de π_k, nous avons

$$\begin{aligned} \pi_k &= \pi_0 \prod_{j=0}^{k-1}(1 - \delta_j) + \sum_{j=0}^{k-1} \epsilon_j \prod_{l=j+1}^{k-1} (1 - \delta_l) \\ &\leq \pi_0 \prod_{j=0}^{k-1}(1 + |\delta_j|) + \sum_{j=0}^{k-1} |\epsilon_j| \prod_{l=j+1}^{k-1} (1 + |\delta_l|) \\ &\leq \left[\prod_{k=0}^{\infty}(1 + |\delta_k|)\right] \left(\pi_0 + \sum_{k=0}^{\infty} |\epsilon_k|\right) \end{aligned} \tag{3.121}$$

On sait aussi que

$$\sum_{k=0}^{\infty} |\delta_k| < \infty \Rightarrow \prod_{k=0}^{\infty}(1 + |\delta_k|) < \infty.$$

Ainsi, le membre de droite de (3.121) est fini, notons le $\bar{\pi}$. Nous obtenons alors, par la définition de π_{k+1}

$$|\pi_{k+1} - \pi_k| \leq |\delta_k|\pi_k + |\epsilon_k| \leq |\delta_k|\bar{\pi} + |\epsilon_k|. \tag{3.122}$$

Comme la série, ayant pour terme général le membre de droite de (3.122), est convergente, la série $\sum_{k=0}^{\infty}(\pi_{k+1} - \pi_k)$ est absolument convergente, et par conséquent convergente. Il découle donc que la suite $\{\pi_k\}$ converge car,

$$\lim_{k \to \infty} \pi_k = \sum_{k=0}^{\infty}(\pi_{k+1} - \pi_k) < \infty.$$

□

Finalement, nous présentons le résultat de convergence.

Théorème 3.7.1 *Si* f *est deux fois continument différentiable en* x^* *et si le Hessien de* f *en* x^*, $\nabla^2 f(x^*)$, *n'est pas singulier, alors la suite duale* $\{z^k\}$ *définie par (3.22), (3.30) converge vers un vecteur* z^* *des multiplicateurs de Karush-Kuhn-Tucker pour le problème min* $f(x)$, *s.c.* $Ax \leq b$.

Preuve

De (3.115), nous avons

$$z_i^{k+1} = (1 - \tilde{\mu}_k) z_i^k + \tilde{\lambda}_k e_i^{i(k)}. \tag{3.123}$$

Appliquons la proposition 3.7.6 avec $\pi_k = z_i^k, \epsilon_k = \tilde{\lambda}_k e_i^{i(k)}, \delta_k = \tilde{\mu}_k$. Puisque $e_i^{i(k)}$ vaut soit 1 soit 0, le corollaire 3.7.1 implique que les hypothèses de la proposition 3.7.6 sont satisfaites; par conséquent, les suites $\{z_i^k\}$ convergent pour tout i, i.e., $\{z^k\}$ converge (vers z^*). Prenant les limites dans la proposition 3.4.2, nous avons

$$z^* \geq 0, \quad \nabla f(x^*) = -A^t z^*, \quad (z^*)^t (b - Ax^*) = 0, \quad Ax^* \leq b,$$

qui sont les conditions de Karush-Kuhn-Tucker pour le problème min $f(x)$, s.c. $Ax \leq b$.

□

3.8 Conclusion

Nous venons de prouver que l'introduction de la suite duale implique un meilleur comportement de convergence dans le cas des contraintes linéaires. Pour l'algorithme primal-dual la suite primale converge au moins linéairement, contairement à celle de l'algorithme primal, qui peut converger de manière sous-linéaire.

Comme conséquence, nous en avons déduit la convergence de la suite duale.

Conclusion générale

Nous venons d'analyser la convergence des différentes méthodes de Han et Iusem et nous avons comparé ces méthodes entre elles.

D'une première comparaison entre la méthode primale de type "row-action" de Iusem et la méthode de projections succesives de Han, nous avons conclu que la méthode de Iusem était la plus intéressante des deux d'un point de vue place mémoire pour une même convergence.

Seulement, cette méthode était purement primale et de ce fait pouvait converger sous-linéairement.

C'est pourquoi, nous avons ensuite introduit une seconde méthode proposée par Iusem, toujours de type "row-action", mais de nature primale-duale. La convergence de cette nouvelle méthode a également été analysée.

Ensuite nous avons comparé ces deux méthodes de Iusem entre elles, et avons constaté que dans le cas de contraintes linéaires, la suite primale engendrée par la seconde avait un taux de convergence supérieur à celui de la suite générée par la première méthode.

De ces différentes comparaisons, nous concluons finalement que dans le cas de contraintes linéaires, parmi les trois méthodes analysées dans ce mémoire, c'est la méthode primale-duale de Iusem qui a le meilleur rapport vitesse de convergence/place mémoire. Le cas de contraintes convexes générales reste une question ouverte.

Bibliographie

[**1**] R.T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, New Jersey, 1970).

[**2**] S.P. Han and G. Lou, "A parallel algorithm for a class of convex programs", *SIAM Journal on Control and Optimisation* 26(2) (1988) 345-355.

[**3**] A.N. Iusem and B.F. Svaiter, " Primal-Dual Row-Action Method for Convex Programming", *Journal of optimization theory and applications* Vol 86 No 1 (July 1995) 73-112.

[**4**] A.N. Iusem and B.F. Svaiter, "A row-action method for convex programming", *Mathematical Programming* 64 (1994) 149-171.

[**5**] S.P. Han, "A successive projection method", *Mathematical Programming* 40 (1988) 1-14.

[**6**] A.N. Iusem and A.R. De Pierro, "On the convergence of Han's method for convex programming with quadratic objective", *Mathematical Programming* 52 (1991) 265-284.

Annexe A

Preuves Supplémentaires

Preuve du corollaire 1.4.1

(a) En sommant les inégalités du lemme 1.4.2 de 1 à k, nous trouvons

$$h(y_1^0,\ldots,y_m^0) - h(y_1^k,\ldots,y_m^k) \geq \frac{1}{2}\sum_{j=1}^{k}(\sum_{i=1}^{m}\|y_i^j - y_i^{j-1}\|_Q^2).$$

D'autre part pour tout $(y_1,\ldots,y_m)$, nous avons

$$\begin{aligned} h(y_1,\ldots,y_m) &= f(y_1+\ldots+y_m) + \sum_{i=1}^{m}\delta^*(y_i \mid QC_i) \\ &\geq g(y) \\ &\geq g(\bar{y}) \end{aligned}$$

où $y = y_1 + \ldots + y_m$ et $\bar{y}$ est la solution de (D). Par conséquent, h est bornée inférieurement et pour tout k nous avons,

$$h(y_1^0,\ldots,y_m^0) - g(\bar{y}) \geq \frac{1}{2}\sum_{j=1}^{k}(\sum_{i=1}^{m}\|y_i^j - y_i^{j-1}\|_Q^2).$$

Lorsque $k \to \infty$, la série est bornée, et donc son terme général converge vers 0. Or ce terme général est une série qui est bornée (puisque convergente), donc son terme général converge aussi vers 0. Ainsi, l'assertion (a) est prouvée.

(b) Par la définition de y_i^k, nous avons

$$x_{i-1}^k - x_i^k = y_i^k - y_i^{k-1},$$

et la thèse découle de l'assertion (a).

- La dernière assertion est bien une conséquence de (a) et (b), grâce à la preuve de l'assertion (b) du lemme 1.4.1.

Preuve de la proposition 1.5.9

(i) Nous avons,

$$\left(\frac{1}{k}\sum_{j=1}^{k}\sqrt{\alpha_j}\right)^2 \leq \frac{1}{k}\sum_{j=1}^{k}\alpha_j \leq \frac{1}{k}\sum_{j=1}^{\infty}\alpha_j.$$

En prenant la racine carrée et en multipliant ensuite par $\sqrt{k}$, nous obtenons,

$$\frac{1}{\sqrt{k}}\sum_{j=1}^{k}\sqrt{\alpha_j} \leq \sqrt{\sum_{j=1}^{\infty}\alpha_j}.$$

(ii) Nous avons,

$$0 \leq 2k\alpha_{2k} \leq 2\sum_{j=k+1}^{2k}\alpha_j \leq 2\sum_{j=k+1}^{\infty}\alpha_j,$$

cette dernière série tend vers 0 lorsque $k \to \infty$. Ainsi, $2k\alpha_{2k} \to 0$ quand $k \to \infty$, et $k\alpha_k \to 0$ quand $k \to \infty$.

Preuve de la proposition 1.7.2

Nous procédons par récurrence. Commençons par vérifier les cinq assertions pour $k = 1$.

(i) Il suffit d'utiliser respectivement l'initialisation de l'algorithme (1.5)-(1.7), la proposition 1.7.1(i), (1.30), (1.24) et l'initialisation de l'algorithme parallèle, i.e.,

$$(\bar{z}_2^1)_i = [\bar{P}_1(\bar{x}^0)]_i = \sum_{l=1}^{m}\lambda_l\bar{x}_l^0 = \sum_{l=1}^{m}\lambda_l x^0 = x^0 = z_i^1. \tag{A.1}$$

(ii) Nous avons,

$$(\bar{x}_2^1)_i = [\bar{P}_2(\bar{z}_2^1)]_i = P_i[(\bar{z}_2^1)_i] = P_i(z_i^1), \tag{A.2}$$

en employant respectivement (1.5), la proposition 1.7.1(ii) et (A.1).

(iii) Nous trouvons,

$$\begin{aligned}(\bar{z}_1^2)_i &= (\bar{x}_2^1)_i + (\bar{z}_1^1)_i - (\bar{x}_1^1)_i \\ &= P_i(z_i^1) + \bar{x}_i^0 - [\bar{P}_1(\bar{x}^0)]_i \\ &= P_i(z_i^1) + x^0 - \sum_{l=1}^{m}\lambda_l\bar{x}_l^0 \\ &= P_i(z_i^1) + x^0 - x^0 = P_i(z_i^1),\end{aligned} \tag{A.3}$$

où la première égalité découle de (1.6), la seconde de (A.2), l'initialisation et (1.5), la troisième de (A.1), la proposition 1.7.1(i) et la quatrième de (1.24).

(iv) Nous avons,

$$(\bar{x}_1^2)_i = [\bar{P}_1(\bar{z}_1^2)]_i = \sum_{l=1}^{m} \lambda_l (z_1^2)_l = \sum_{l=1}^{m} \lambda_l P_l(z_l^1) = x^1,$$

en utilisant respectivement (1.5), la proposition 1.7.1(i), (A.3) et (1.25).

Supposons maintenant que ce soit vrai pour k, et montrons que ça reste vrai pour $k+1$.

(i) Il suffit d'utiliser respectivement (1.7), l'hypothèse de récurrence et (1.26)

$$(\bar{z}_2^{k+1})_i = (\bar{x}_1^{k+1})_i + (\bar{z}_2^k)_i - (\bar{x}_2^k)_i = x^k + z_i^k - P_i(z_i^k) = z_i^{k+1}, \tag{A.4}$$

(ii) Nous avons,

$$(\bar{x}_2^{k+1})_i = [\bar{P}_2(\bar{z}_2^{k+1})]_i = P_i[(\bar{z}_2^{k+1})_i] = P_i(z_i^{k+1}), \tag{A.5}$$

en employant (1.5), la proposition 1.7.1(ii) et (A.4).

(iii) Nous avons,

$$\begin{aligned}
(\bar{z}_1^{k+2})_i &= (\bar{x}_2^{k+1})_i + (\bar{z}_1^{k+1})_i - (\bar{x}_1^{k+1})_i \\
&= P_i(z_i^{k+1}) + P_i(z_i^k) + \sum_{j=1}^{k-1} [P_i(z_i^j) - x^j] - x^k \\
&= P_i(z_i^{k+1}) + \sum_{j=1}^{k} [P_i(z_i^j) - x^j],
\end{aligned} \tag{A.6}$$

en utilisant (1.6), (A.5) et l'hypothèse de récurrence.

(iv) Nous avons,

$$\begin{aligned}
(\bar{x}_1^{k+2})_i &= [\bar{P}_1(\bar{z}_1^{k+2})]_i \\
&= \sum_{l=1}^{m} \lambda_l (\bar{z}_1^{k+2})_l \\
&= \sum_{l=1}^{m} \lambda_l P_l(z_l^{k+1}) + \sum_{j=1}^{k} \left[\sum_{l=1}^{m} \lambda_l P_l(z_l^j) - \sum_{l=1}^{m} \lambda_l x^j \right] \\
&= x^{k+1} + \sum_{j=1}^{k} \left[\sum_{l=1}^{m} \lambda_l P_l(z_l^j) - x^j \right] \\
&= x^{k+1},
\end{aligned}$$

où la première égalité découle de (1.5), la seconde de la proposition 1.7.1(i), la troisième de (A.6), la quatrième de (1.25), (1.24) et la dernière de (1.25).

Preuve de la proposition 3.3.1

(i) Comme $x^{k+1} \in W_k \subset L(x^k)$, nous avons

$$\langle \nabla f(x^k), x^{k+1} - x^k \rangle \geq 0. \tag{A.7}$$

D'autre part, de la convexité de f, nous avons aussi

$$f(x^{k+1}) \geq f(x^k) + \langle \nabla f(x^k), x^{k+1} - x^k \rangle. \tag{A.8}$$

D'où $f(x^k) \leq f(x^{k+1})$. De $x^* \in W_k$ et (3.2) nous avons, $f(x^{k+1}) \leq f(x^*)$. Et donc la première assertion de (i) est vérifiée.

Pour la seconde assertion, il suffit de remarquer que (A.8) est une inégalité stricte à moins que $x^k = x^{k+1}$, par (B'1).

(ii) Par (i), $\{x^k\} \subset \{x : f(x) \leq f(x^*)\}$, borné par (B'3).

(iii) Par l'absurde, supposons qu'il existe une sous-suite $\{x^{l_k}\}$ de $\{x^k\}$ telle que la suite $\{x^{l_k+1} - x^{l_k}\}$ ne converge pas vers zéro. Par (ii), nous pouvons extraire une sous-suite $\{x^{j_k}\}$ de $\{x^{l_k}\}$ telle que $\{x^{j_k}\}$ et $\{x^{j_k+1}\}$ convergent respectivement vers des points nommés $\bar{x}, \tilde{x}$. L'hypothèse sur $\{x^{l_k}\}$ implique $\bar{x} \neq \tilde{x}$. Soient

$$y^k = (1/2)(x^{j_k} + x^{j_k+1}), \quad y = \lim_{k\to\infty} y^k = (1/2)(\bar{x} + \tilde{x}).$$

Comme par (i), $\{f(x^k)\}$ est croissante et bornée, $\{f(x^k)\}$ converge.Notons sa limite ζ. Il suit que $\{f(x^{j_k})\}$ et $\{f(x^{j_k+1})\}$ convergent aussi vers ζ, i.e., $f(\bar{x}) = f(\tilde{x}) = \zeta$; or f est strictement convexe, donc $f(y) < \zeta$. Ainsi, pour k assez grand,

$$f(y^k) < f(x^{j_k}). \tag{A.9}$$

En plus $x^{j_k} \in L(x^{j_k})$ par la définition de $L(x)$, et $x^{j_k+1} \in L(x^{j_k})$ car $x^{j_k+1} \in W_{j_k} \subset L(x^{j_k})$. Donc, $y^k \in L(x^{j_k})$ par convexité de $L(x^{j_k})$. Comme f est convexe et comme $y^k \in L(x^{j_k})$, il suit que

$$f(y^k) \geq f(x^{j_k}) + \langle \nabla f(x^{j_k}), y^k - x^{j_k} \rangle \geq f(x^{j_k}). \tag{A.10}$$

(A.9) et (A.10) se contredisent, par conséquent, le résultat est vérifié.

(iv) Comme $x^{k+1} \in W_k$, nous avons $d(x^k, W_k) \leq \|x^{k+1} - x^k\|$, et le résultat découle de (iii).

Preuve du corollaire 3.3.1

(i) Par la proposition 3.3.1(i), $f(\tilde{x}) \leq f(x^*)$. Par la définition de x^*, $f(x^*) \leq f(x)$, pour tout $x \in C$, impliquant $f(\tilde{x}) \geq f(x^*)$. Ainsi $f(\tilde{x}) = f(x^*)$, et $\tilde{x}$ est par conséquent un minimum de f sur C. Par (B'1), $\tilde{x} = x^*$.

(ii) Le même argument est valable pour (ii) avec x^k au lieu de $\tilde{x}$.

Preuve de la proposition 3.4.2

(i) Nous procédons par récurrence. (i) est vrai pour $k = 0$ par (3.13). Supposons que ce soit vrai pour k. Montrons que ça reste vrai pour $k + 1$.

1) Si l'itération k est externe, $z^{k+1} \geq 0$ par (3.21), (3.22) et par l'hypothèse de récurrence.

2) Si l'itération k est interne, $z_i^{k+1} = \mu_k z_i^k$ pour $i \neq i(k)$, par (3.30) ainsi $z_i^{k+1} \geq 0$ par (3.29) et par l'hypothèse de récurrence. Si $i = i(k)$, nous obtenons de (3.30) $z_{i(k)}^{k+1} = \lambda_k \geq 0$ par (3.29).

(ii) Par récurrence. (ii) est vrai pour $k = 0$ par (3.14). Supposons que ce soit vrai pour k. Nous montrons que ça reste vrai pour $k + 1$.

1) Si l'itération k est externe, alors

$$\begin{aligned}
(\bar{A}^{k+1})^t z^{k+1} &= (\bar{A}^k + \gamma_k E^{i(k)}(A^k - \bar{A}^k))^t(\mu_k z^k + \lambda_k e^{i(k)}) \\
&= \mu_k(\bar{A}^k)^t z^k + \lambda_k(\bar{A}^k)^t e^{i(k)} \\
&+ \gamma_k(A^k - \bar{A}^k)^t e^{i(k)}(e^{i(k)})^t(\mu_k z^k + \lambda_k e^{i(k)}) \\
&= -\mu_k \nabla f(x^k) + \lambda_k \bar{a}^{k,i(k)} + \gamma_k(\mu_k z_{i(k)}^k + \lambda_k)(a^{k,i(k)} - \bar{a}^{k,i(k)}) \\
&= -\mu_k \nabla f(x^k) + \lambda_k \bar{a}^{k,i(k)} + \lambda_k(a^{k,i(k)} - \bar{a}^{k,i(k)}) \\
&= -\mu_k \nabla f(x^k) + \lambda_k a^{k,i(k)} \qquad \text{(A.11)} \\
&= -\nabla f(x^{k+1}),
\end{aligned}$$

en utilisant (3.22) et (3.24) dans la première égalité, (3.8) dans la seconde, l'hypothèse de récurrence dans la troisième, (3.22), (3.23) dans la quatrième, et (3.33) dans la dernière. Notons que, si $\mu_k z_{i(k)}^k + \lambda_k$ disparait si $\lambda_k = 0$.

2) Si l'itération k est interne, alors

$$\begin{aligned}
(\bar{A}^{k+1})^t z^{k+1} &= (\bar{A}^k)^t[\mu_k z^k + (\lambda_k - \mu_k z_{i(k)}^k)e^{i(k)}] \\
&= \mu_k(\bar{A}^k)^t z^k + (\lambda_k - \mu_k z_{i(k)}^k)\bar{a}^{k,i(k)} \\
&= -\mu_k(\nabla f(x^k) + z_{i(k)}^k \bar{a}^{k,i(k)}) + \lambda_k \bar{a}^{k,i(k)} \\
&= -\nabla f(x^{k+1}), \qquad \text{(A.12)}
\end{aligned}$$

en utilisant (3.30) et (3.31) dans la première égalité, l'hypothèse de récurrence dans la troisième, et (3.38) dans la dernière.

(iii) Nous procédons par récurrence. (iii) est vrai pour $k = 0$ par (3.15). Supposons que ce soit vrai pour k. Montrons que ça reste vrai pour $k + 1$.

1) Si l'itération k est externe,alors

$$\begin{aligned}
&(z^{k+1})^t(\bar{b}^{k+1} - \bar{A}^{k+1}x^{k+1}) \qquad \text{(A.13)} \\
&= (\mu_k z^k + \lambda_k e^{i(k)})^t\{\bar{b}^k - \bar{A}^k x^{k+1} + \gamma_k E^{i(k)}[b^k - \bar{b}^k - (A^k - \bar{A}^k)x^{k+1}]\} \\
&= \mu_k(z^k)^t(\bar{b}^k - \bar{A}^k x^{k+1}) + \lambda_k(\bar{b}_{i(k)}^k - \langle \bar{a}^{k,i(k)}, x^{k+1}\rangle)
\end{aligned}$$

$$
\begin{aligned}
&+ \gamma_k(\mu_k z^k_{i(k)} + \lambda_k)(b^k_{i(k)} - \bar{b}^k_{i(k)} - \langle a^{k,i(k)} - \bar{a}^{k,i(k)}, x^{k+1}\rangle)\\
&= \mu_k(z^k)^t(\bar{b}^k - \bar{A}^k x^{k+1}) + \lambda_k(b^k_{i(k)} - \langle a^{k,i(k)}, x^{k+1}\rangle)\\
&= \mu_k(z^k)^t(\bar{b}^k - \bar{A}^k x^{k+1})\\
&= \mu_k((z^k)^t\bar{b}^k + \langle \nabla f(x^k), x^{k+1}\rangle)\\
&= \mu_k((z^k)^t\bar{b}^k + \langle \nabla f(x^k), x^k\rangle)\\
&= \mu_k(z^k)^t(\bar{b}^k - \bar{A}^k x^k) = 0, \qquad \text{(A.14)}
\end{aligned}
$$

en utilisant (3.22),(3.24),(3.25) dans la première égalité, (3.8) et un peu d'algèbre dans la seconde, (3.23) dans la troisième (avec la même observation que dans (ii)), (3.36) dans la quatrième, (ii) dans la cinquième, (3.37) dans la sixième, (ii) encore dans la septième, et l'hypothèse de récurrence dans la dernière.

2) Si l'itération k est interne, alors

$$
\begin{aligned}
&(z^{k+1})^t(\bar{b}^{k+1} - \bar{A}^{k+1}x^{k+1}) \qquad \text{(A.15)}\\
&= [\mu_k z^k + (\lambda_k - \mu_k z^k_{i(k)})e^{i(k)}]^t(\bar{b}^k - \bar{A}^k x^{k+1})\\
&= \mu_k(z^k)^t(\bar{b}^k - \bar{A}^k x^{k+1}) + (\lambda_k - \mu_k z^k_{i(k)})(\bar{b}^k_{i(k)} - \langle \bar{a}^{k,i(k)}, x^{k+1}\rangle)\\
&= \mu_k((z^k)^t\bar{b}^k + \langle \nabla f(x^k) + z^k_{i(k)}\bar{a}^{k,i(k)}, x^{k+1}\rangle - z^k_{i(k)}\bar{b}^k_{i(k)})\\
&= \mu_k((z^k)^t\bar{b}^k + \langle \nabla f(x^k), x^k\rangle)\\
&= \mu_k(z^k)^t(\bar{b}^k - \bar{A}^k x^k) = 0, \qquad \text{(A.16)}
\end{aligned}
$$

en utilisant (3.30), (3.31), (3.32) dans la première égalité, (3.41) dans la troisième, (3.42) dans la quatrième, (ii) dans la cinquième, et l'hypothèse de récurrence dans la dernière.

(iv) De (3.10), (3.11), (3.16), (3.17), nous avons

$$
(A^k x - b^k)_i = \langle \nabla g_i(x^k), x - x^k\rangle + \alpha_k g_i(x^k) \leq g_i(x) - (1-\alpha_k)g_i(x^k), \qquad \text{(A.17)}
$$

en utilisant la convexité de g_i dans l'inégalité.

(v) Encore par récurrence. Pour $k = 0$,

$$
\begin{aligned}
(\bar{A}^0 x - \bar{b}^0)_i &= (A^0 x)_i - \langle a^{0,i}, x^0\rangle + g_i(x^0)\\
&= \langle \nabla g_i(x^0), x - x^0\rangle + g_i(x^0)\\
&\leq g_i(x), \qquad \text{(A.18)}
\end{aligned}
$$

en utilisant (3.12) dans la première égalité, (3.10) dans la seconde, et la convexité de g_i dans l'inégalité. Supposons que (v) est vrai pour k, montrons que ça reste vrai pour $k+1$.

1) Si l'itération k est externe et $i \neq i(k)$, alors

$$
\begin{aligned}
(\bar{A}^{k+1}x - \bar{b}^{k+1})_i &= \langle \bar{a}^{k+1,i}, x\rangle - \bar{b}^{k+1}_i\\
&= \langle \bar{a}^{k,i}, x\rangle - \bar{b}^k_i\\
&\leq g_i(x), \qquad \text{(A.19)}
\end{aligned}
$$

par (3.24), qui implique $\bar{a}^{k+1,i} = \bar{a}^{k,i}, \bar{b}^{k+1}_i = \bar{b}^k_i$, et l'hypothèse de récurrence.
Si $i = i(k)$, de (3.8) et (3.24), nous avons

$$
\bar{a}^{k+1,i(k)} = \gamma_k a^{k,i(k)} + (1-\gamma_k)\bar{a}^{k,i(k)}, \qquad \text{(A.20)}
$$

$$\bar{b}_{i(k)}^{k+1} = \gamma_k b_{i(k)}^k + (1-\gamma_k)\bar{b}_{i(k)}^k. \tag{A.21}$$

Utilisant (i), (3.21), (3.23), nous avons

$$0 \leq \gamma_k \leq 1. \tag{A.22}$$

De (A.20), (A.21), nous avons

$$\begin{aligned}
(\bar{A}^{k+1}x - \bar{b}^{k+1})_{i(k)} &= \langle \bar{a}^{k+1,i(k)}, x\rangle - \bar{b}_{i(k)}^{k+1} \\
&= \gamma_k(\langle a^{k,i(k)}, x\rangle - b_{i(k)}^k) + (1-\gamma_k)(\langle \bar{a}^{k,i(k)}, x\rangle - \bar{b}_{i(k)}^k) \\
&\leq \gamma_k(\langle a^{k,i(k)}, x\rangle - b_{i(k)}^k) + (1-\gamma_k) g_{i(k)}(x) \\
&\leq \gamma_k[g_{i(k)}(x) - (1-\alpha_k) g_{i(k)}(x^k)] + (1-\gamma_k) g_{i(k)}(x) \\
&\leq \gamma_k g_{i(k)}(x) + (1-\gamma_k) g_{i(k)}(x) \\
&= g_{i(k)}(x)
\end{aligned} \tag{A.23}$$

en utilisant l'hypothèse de récurrence et (A.22) dans la première inégalité, (iv) et (A.22) dans la seconde, et (A.22), (3.7), et $g_{i(k)}(x^k) > 0$ (résultant du fait que la $k$ème itération est externe), dans la troisième. De (A.19), (A.23),

$$(\bar{A}^{k+1}x - \bar{b}^{k+1})_i \leq g_i(x), \quad \forall x$$

i.e.

$$\bar{A}^{k+1}x - \bar{b}^{k+1} \leq g(x).$$

2) Si l'itération k est interne, alors le résultat découle de (3.31), (3.32) et de l'hypothèse de récurrence.
