MASTER EN SCIENCES MATHÉMATIQUES

Application de la méthode affine à un problème d'optimisation convexe non différentiable

Donfut, Delphine

Award date:
1997

Facultés Universitaires Notre-Dame de la Paix
Namur
Facultés des Sciences - Département de Mathématiques

APPLICATION DE LA MÉTHODE AFFINE A UN PROBLÈME D'OPTIMISATION CONVEXE NON DIFFÉRENTIABLE

Mémoire présenté pour l'obtention du grade
de Licencié en Sciences
Mathématiques
par

Promoteur : J.-J. Strodiot

DONFUT Delphine

Année académique 1996-1997

Je tiens à exprimer toute ma gratitude envers Monsieur J.-J. Strodiot pour sa disponibilité et les conseils qu'il m'a prodigués tout au long de l'élaboration et de la rédaction de ce mémoire.

Table des matières

Résumé

L'objet de ce mémoire est l'étude d'un algorithme de résolution de problèmes d'optimisation convexe non différentiable. L'algorithme étudié est une combinaison de la méthode affine et de la méthode du sous-gradient. Il est également une généralisation de la méthode du point intérieur. A chaque itération, on associe au point courant une transformation affine et on calcule un sous-gradient de la fonction objectif augmentée d'une fonction barrière logarithmique. La direction de descente est obtenue en projetant suivant la nouvelle métrique ce sous-gradient dans le sous-espace associé aux contraintes. Le choix des pas est analogue à celui de la méthode classique du sous-gradient. On peut montrer, sous la condition d'angle, la convergence de cette méthode.

Abstract

We propose an algorithm for solving nondifferentiable convex programming problems. This algorithm combines the idea of the affine scaling method with the subgradient method. It is a generalization of the interior point method. At each iteration, we associate a scaling and we calculate a subgradient of the objective function with a logarithmic barrier term. The search direction is obtained by projecting on the scaled space this subgradient. The stepsize choice is analogous to the stepsize choice of the classic subgradient method. Convergence of the method is established if the angle condition is assumed.

1.1 Définitions

Soit $f : I\!R^n \to I\!R$ une fonction convexe non différentiable et soit $x \in I\!R^n$.

Le *sous-différentiel* de la fonction f au point x, noté $\partial f(x)$, est défini de la manière suivante :

$$\partial f(x) = \{ \; x^* \in I\!R^n \mid f(z) \; \geq \; f(x) + \; < x^*, z - x > \; \forall z \in I\!R^n \; \} .$$

Les éléments de $\partial f(x)$ sont appelés *sous-gradients* de f au point x.

Pour les fonctions convexes, le sous-gradient est donc bien une généralisation du gradient.

Voici deux propriétés du sous-différentiel que nous utiliserons par la suite :

1. Si f est différentiable en x,
   alors $\partial f(x) = \{\nabla f(x)\}$.

2. Si f et g sont deux fonctions convexes de $I\!R^n$ dans $I\!R$,
   alors $\partial(f + g)(x) = \partial f(x) + \partial g(x)$, pour tout $x \in I\!R^n$.

1.2 Les algorithmes

Nous allons commencer par traiter le cas des problèmes sans contrainte c'est-à-dire les problèmes qui peuvent s'écrire sous la forme suivante :

$$\min_{x \in I\!R^n} f(x),$$

où $f : I\!R^n \to I\!R$ est une fonction convexe non différentiable. On supposera qu'on peut calculer un sous-gradient γ de f en tout $x \in I\!R^n$. En s'inspirant de la méthode du gradient, nous obtenons l'algorithme suivant :

(a) $x^{(0)} \in I\!R^n$ et k = 0.

(b) Trouver $\gamma_k \in \partial f(x^{(k)})$;
si $\gamma_k = 0 \longrightarrow$ fin ; $x^{(k)}$ est un point optimal,
sinon $\longrightarrow x^{(k+1)} = x^{(k)} - \lambda_k \dfrac{\gamma_k}{\|\gamma_k\|}$.

(c) Test d'arrêt;
s'il est satisfait $\longrightarrow$ fin,
sinon $\longrightarrow$ k = k+1 ; retour en (b).

Nous pouvons généraliser cette méthode sans trop de difficultés au cas d'un problème avec contraintes de la forme suivante :

$$\min_{x \in X} f(x),$$

où $X \subseteq I\!R^n$ est un ensemble convexe et fermé et $f : I\!R^n \to I\!R$ une fonction convexe, non différentiable et pour laquelle on peut calculer un sous-gradient γ en tout $x \in I\!R^n$. Dans ces conditions, l'algorithme de la méthode du sous-gradient est le suivant :

(a) $x^{(0)} \in X$ et k = 0.

(b) Trouver $\gamma_k \in \partial f(x^{(k)})$;
si $\gamma_k = 0 \longrightarrow$ fin; $x^{(k)}$ est un point optimal,
sinon $\longrightarrow x^{(k+1)} = Proj_X \left[x^{(k)} - \lambda_k \frac{\gamma_k}{\|\gamma_k\|} \right]$.

(c) Test d'arrêt;
s'il est satisfait $\longrightarrow$ fin,
sinon $\longrightarrow$ k = k+1 ; retour en (b).

A l'étape (b), $Proj_X(y)$ représente la projection orthogonale du vecteur y sur X.

1.3 Convergence de la méthode du sous-gradient

Avant de démontrer la convergence de la méthode vers une solution optimale x^*, nous devons nous assurer de l'existence d'une telle solution. Dans ce but, introduisons les deux hypothèses suivantes :

1) $f(x) \to \infty$ lorsque $\|x\| \to \infty$,

2) X est borné.

Si au moins une de ces deux hypothèses est vérifiée, l'existence d'une solution optimale $x^* \in X$ est assurée en vertu du théorème de Weierstrass.

Dans la suite de cette section, nous allons donc supposer qu'au moins une de ces deux hypothèses est vérifiée et chercher les conditions à imposer sur les pas λ_k pour avoir la convergence de la méthode. C'est l'objet du théorème suivant.

Théorème 1.1
Soient $f : I\!R^n \to I\!R$ une fonction convexe et continue et $X \subseteq I\!R^n$ un ensemble convexe et borné. Supposons que les pas λ^k vérifient les deux conditions suivantes :

$$\lambda_k \to 0 \text{ lorsque } k \to \infty,$$
$$\sum_{k=1}^{\infty} \lambda_k = +\infty.$$

Alors la suite $(x^{(k)})$ engendrée par le second algorithme est telle que

$$\liminf_{k\to\infty} f(x^{(k)}) = \min_{x\in X} f(x).$$

Dans ce cas, la suite des pas λ_k se comporte comme une progression géométrique de raison $\sqrt{1-\chi^2}$, ce qui revient à poser

$$\lambda_k = \lambda_0 \alpha^k \qquad \text{avec} \left\{ \begin{array}{l} \alpha = \sqrt{1-\chi^2} \\ \\ \lambda_0 = \chi d(x^{(0)}). \end{array} \right.$$

Le meilleur taux de convergence est par conséquent $\sqrt{1-\chi^2}$ et il est obtenu lorsque l'on fait le meilleur choix pour t à chaque pas. Cependant, en pratique, ni χ ni $d(x^{(0)})$ ne sont connus. Cela implique qu'en général, on a la convergence linéaire avec un taux :

$$\alpha > \sqrt{1-\chi^2}.$$

Pour être un peu plus précis, introduisons quelques notations fort utiles :

$$\begin{aligned} C &= \max \left\{ \frac{1}{\alpha} \, ; \, \frac{\chi - \sqrt{\chi^2 - (1-\alpha^2)}}{1-\alpha^2} \right\} , \\ D &= \frac{\chi + \sqrt{\chi^2 - (1-\alpha^2)}}{1-\alpha^2} , \\ z &= \left\{ \begin{array}{ll} \sqrt{1-\chi^2} & si\ \chi \leq \frac{\sqrt{2}}{2} \\ \dfrac{1}{2\chi} & si\ \chi > \frac{\sqrt{2}}{2}. \end{array} \right. \end{aligned}$$

Nous pouvons maintenant formuler le théorème de convergence pour la méthode de la série convergente.

Théorème 1.2

Si $z \leq \alpha < 1$ et si $d(x^{(0)}) \in [\ C\lambda_0\ ,\ D\lambda_0\ ]$, alors, pour tout k, $d(x^{(k)}) \leq \alpha^k d(x^{(0)})$ (convergence linéaire vers un point optimal).

Si $z \leq \alpha < 1$ et si $d(x^{(0)}) < C\lambda_0$, alors, pour tout k, $d(x^{(k)}) \leq \alpha^k C \lambda_0$ (convergence linéaire vers un point optimal).

Si $\alpha < z$ ou si $d(x^{(0)}) > D\lambda_0$, alors l'algorithme peut converger vers un point non optimal.

<u>*Preuve*</u>

Commençons par majorer $d^2(x^{(k+1)})$. Nous avons successivement

$$\begin{aligned}
d^2(x^{(k+1)}) &= \| \, x^{(k+1)} - s(x^{(k+1)}) \, \|^2 \\
&\leq \| \, x^{(k+1)} - s(x^{(k)}) \, \|^2 \\
&= \left\| \, x^{(k)} - \lambda_k \, \frac{\gamma_k^T}{\|\gamma_k\|} - s(x^{(k)}) \, \right\|^2 \\
&= \| \, x^{(k)} - s(x^{(k)}) \, \|^2 + \lambda_k^2 - 2\lambda_k \, \frac{\gamma_k^T}{\|\gamma_k\|} \, (\, x^{(k)} - s(x^{(k)}) \,) \\
&\leq d^2(x^{(k)}) + \lambda_k^2 - 2\lambda_k \, \chi \, \| \, x^{(k)} - s(x^{(k)}) \, \| \\
&= d^2(x^{(k)}) + \lambda_k^2 - 2\lambda_k \, \chi \, d(x^{(k)}) \\
&= d^2(x^{(k)}) + \lambda_0^2 \, \alpha^{2k} - 2\lambda_0 \, \alpha^k \, \chi \, d(x^{(k)})
\end{aligned}$$

puisque nous travaillons avec la méthode de la série convergente.

Montrons ensuite que

$$\exists \, M \qquad \text{t.q.} \qquad d(x^{(k)}) \leq M \, \alpha^k \qquad \Longrightarrow \qquad d(x^{(k+1)}) \leq M \, \alpha^{k+1}.$$

Supposons que $d(x^{(k)}) = K \, \alpha^k$ avec $K \in [\, 0 \, , \, M \,]$. Dès lors,

$$\begin{aligned}
d^2(x^{(k+1)}) &\leq d^2(x^{(k)}) + \lambda_0^2 \, \alpha^{2k} - 2\lambda_0 \, \alpha^k \, \chi \, d(x^{(k)}) \\
&= K^2 \, \alpha^{2k} + \lambda_0^2 \, \alpha^{2k} - 2\lambda_0 \, \chi \, \alpha^{2k} \, K.
\end{aligned}$$

Le théorème sera donc vérifié si on parvient à montrer que

$$\exists \, M \text{ t.q } d(x^{(0)}) \leq M \Longrightarrow K^2 \, \alpha^{2k} + \lambda_0^2 \, \alpha^{2k} - 2\lambda_0 \, \alpha^{2k} \, \chi \, K \leq M^2 \, \alpha^{2k+2}.$$

Supposons $d(x^{(0)}) \leq M$ sans préciser ce que vaut M et regardons quelles conditions doivent être vérifiées pour avoir l'implication désirée.

Considérons l'inégalité suivante :

$$K^2 \, \alpha^{2k} + \lambda_0^2 \, \alpha^{2k} - 2\lambda_0 \, \chi \, \alpha^{2k} K - M^2 \, \alpha^{2k+2} \leq 0.$$

Le membre de gauche étant une fonction convexe de K, cette inégalité sera satisfaite si elle est vérifiée pour $K = 0$ et $K = M$, c-à-d si :

$$\lambda_0^2 \; \alpha^{2k} - M^2 \; \alpha^{2k+2} \leq 0 \; , \tag{1.1}$$

$$M^2 \; \alpha^{2k} + \lambda_0^2 \; \alpha^{2k} - 2\lambda_0 \; \chi \; \alpha^{2k} M - M^2 \; \alpha^{2k+2} \leq 0. \tag{1.2}$$

Pour simplifier ces deux relations, posons

$$\frac{M}{\lambda_0} = \omega.$$

Alors (1.1) et (1.2) deviennent :

$$1 - \omega\alpha \leq 0 \; , \tag{1.3}$$

$$\omega^2(1-\alpha^2) + 1 - 2\chi\omega \leq 0. \tag{1.4}$$

Regardons à présent ce que nous obtenons pour les différentes valeurs possibles de α.

1) $\alpha < \sqrt{1-\chi^2}$

Dans ce cas, nous avons

$$\begin{aligned} \omega^2 \; (\; 1-\alpha^2 \;) - 2\chi\omega + 1 \quad &> \quad \omega^2(1 - 1 + \chi^2) - 2\omega\chi + 1 \\ &= \quad \omega^2\chi^2 - 2\omega\chi + 1 \\ &= \quad (\; \omega\chi - 1 \;)^2 \geq 0 \end{aligned}$$

et par conséquent l'equation (1.4) n'est jamais vérifiée.

2) $\alpha \geq \sqrt{1-\chi^2}$

Nous avons cette fois

$$\omega^2 \; (\; 1-\alpha^2 \;) - 2\omega\chi + 1 \leq 0.$$

En calculant le réalisant du membre de gauche, nous trouvons deux racines positives :

$$\left\{ \frac{\chi - \sqrt{\chi^2 - (1-\alpha^2)}}{(\; 1-\alpha^2 \;)} \right\} \quad \text{et} \quad \left\{ \frac{\chi + \sqrt{\chi^2 - (1-\alpha^2)}}{(1-\alpha^2)} \right\};$$

l'inéquation est dès lors vérifiée si nous nous trouvons entre ces deux racines.

Par conséquent, les équations (1.3) et (1.4) sont vérifiées en même temps si et seulememt si :

$$\alpha \geq \sqrt{1-\chi^2} \qquad \text{et} \qquad \frac{\chi + \sqrt{\chi^2 - (1-\alpha^2)}}{(1-\alpha^2)} \geq \frac{1}{\alpha}.$$

Réécrivons la deuxième inégalité :

$$\begin{aligned}
\frac{\chi + \sqrt{\chi^2 - (1-\alpha^2)}}{(1-\alpha^2)} \geq \frac{1}{\alpha} \quad & \Longleftrightarrow \quad \alpha\sqrt{\chi^2 - (1-\alpha^2)} \geq 1 - \alpha^2 - \chi\alpha \\
& \Longleftrightarrow \quad 1 - \alpha^2 - \alpha\chi \leq 0 \\
& \qquad \text{ou } \alpha^2[\ \chi^2 - (1-\alpha^2)\ ] \geq (1 - \alpha^2 - \alpha\chi)^2 \\
& \Longleftrightarrow \quad \alpha \geq \frac{-\chi + \sqrt{\chi^2+4}}{2} \\
& \qquad \text{ou } \alpha \geq \frac{1}{2\chi}.
\end{aligned}$$

En résumé, les équations (1.1) et (1.2) sont vérifiées simultanément si et seulement si

$$\left\{\ \alpha \geq \sqrt{1-\chi^2}\ \right\} \qquad \text{et} \qquad \left\{\ [\ \alpha \geq \frac{-\chi + \sqrt{\chi^2+4}}{2}\ ] \text{ ou } [\ \alpha \geq \frac{1}{2\chi}\ ]\right\}.$$

ce qui est équivalent à la condition $z \leq \alpha$.

Revenons à l'hypothèse faite au début de la démonstration, à savoir

$$\begin{aligned}
d(x^{(0)}) &\leq \lambda_0\omega \qquad \omega \in [\ C\ ,\ D\ ] \\
d(x^{(0)}) &\leq M \qquad M \in [\ C\lambda_0\ ,\ D\lambda_0\ ].
\end{aligned}$$

Le plus petit M possible donnera la meilleure borne de $d(x^{(0)})$.

Si $d(x^{(0)}) \in [\ C\lambda_0\ \ D\lambda_0\ ]$,	alors $M = d(x^{(0)})$ convient.
Si $d(x^{(0)}) < C\lambda_0$,	alors $M = C\lambda_0$ convient.
Si $d(x^{(0)}) > D\lambda_0$,	alors il est impossible de trouver un M qui convient.

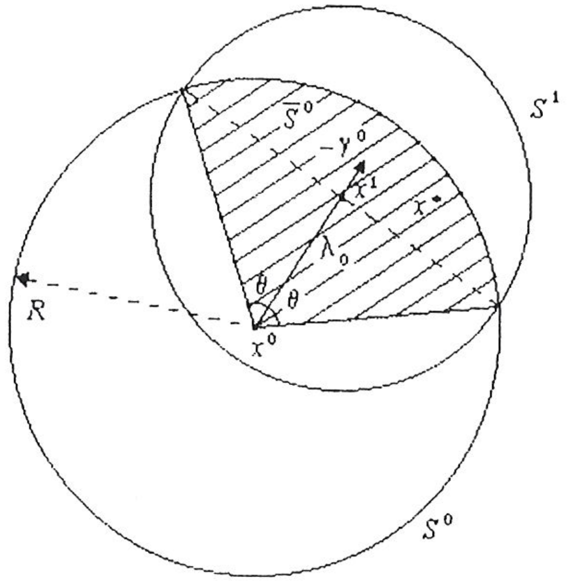
□

La méthode de la série convergente peut être interprétée géométriquement de la manière suivante.

Travaillons dans le cas où l'ensemble des points optimaux Ω est un singleton, ce qui signifie que l'optimum x^* est unique. Supposons que l'on connaisse un majorant R_0 de la distance entre le point de départ $x^{(0)}$ et la solution x^*, c'est-à-dire que

$$R_0 \geq \| x^{(0)} - x^* \|.$$

Le point optimal est donc situé dans la sphère S_0 de centre $x^{(0)}$ et de rayon R_0.
Soit γ_0 un sous-gradient quelconque de f au point $x^{(0)}$ et $\chi = \cos\theta$ la condition de f.



interprétation géométrique de la méthode de la série convergente.

En regardant le dessin, on constate que le point x^* est contenu dans l'intersection de la sphère S_0 et du cône de sommet $x^{(0)}$ et d'angle au sommet 2θ. Notons $\bar{S}_0$ l'ensemble ainsi défini.

Remarquons que dans le cas où $\theta \geq \pi/3$, c-à-d $\chi = \cos\theta \leq 1/2$, l'ensemble $\bar{S}_0$ est totalement contenu dans la sphère de centre $x^{(1)}$ et de rayon R_1 calculés de la manière suivante :

$$x^{(1)} = x^{(0)} - \lambda_0 \frac{\gamma_0}{\|\gamma_0\|}, \qquad \text{avec } \lambda_0 = R_0 \cos\theta = R_0 \chi,$$
$$R_1 = R_0 \sin\theta = R_0 \sqrt{1-\chi^2}.$$

Nous avons donc localisé l'optimum dans une sphère S_1 de rayon strictement inférieur à R_0. Nous pouvons répéter le processus en partant du point $x^{(1)}$. Soit γ_1 un sous-gradient de f au point $x^{(1)}$. On constate que le point x^* est contenu dans la sphère S_2 de centre $x^{(2)}$ et de rayon R_2 calculés de la manière suivante :

$$x^{(2)} = x^{(1)} - \lambda_1 \frac{\gamma_1}{\|\gamma_1\|}, \qquad \text{avec } \lambda_1 = R_1 \cos\theta = R_0 \chi \sqrt{1-\chi^2},$$
$$R_2 = R_1 \sin\theta = R_0 (\sqrt{1-\chi^2})^2.$$

En généralisant à l'étape k, la procédure revient à choisir les pas λ_k de la manière suivante :

$$\lambda_k = \lambda_0 \alpha^k \qquad \text{avec} \left\{ \begin{array}{l} \lambda_0 = R_0 \chi, \\ \\ \alpha = \sqrt{1-\chi^2}. \end{array} \right.$$

Nous venons donc de montrer à nouveau, géométriquement cette fois, que la méthode converge linéairement vers le point x^* avec un taux de convergence égal à $\sqrt{1-\chi^2}$.

Chapitre 2

Méthode affine

Dans ce chapitre, nous allons étudier la minimisation d'une fonction convexe et différentiable sur un ensemble défini par des contraintes linéaires. Nous utiliserons pour cela la méthode affine dans sa forme la plus simple. A chaque itération, une transformation affine est associée au point courant et l'itéré suivant est alors obtenu après recherche linéaire le long de la direction opposée à celle du gradient projeté dans l'espace transformé.

La plupart des méthodes affines nécessitent une fonction objectif deux fois continûment différentiable dont les dérivées secondes sont Lipschitziennes. Pour notre part, nous allons travailler avec l'hypothèse de non dégénerescence primale et il suffira que la fonction objectif soit continûment différentiable.

Nous allons commencer par expliciter le problème à résoudre et définir la non dégénerescence primale. Ensuite, nous présenterons l'idée de la méthode ainsi que l'algorithme. La deuxième section sera consacrée aux propriétés des différentes suites engendrées par l'algorithme. Nous introduirons alors la notion d'estimateur des multiplicateurs en un point x et terminerons ce chapitre par une analyse de la convergence de la suite engendrée par cette méthode.

Notons que l'algorithme et l'analyse de la convergence peuvent être modifiés pour s'accorder avec une recherche linéaire inexacte telle que les méthodes d'Armijo ou de Golstein.

2.1 Algorithme

2.1.1 Position du problème

Dans ce chapitre, nous allons nous intéresser au problème suivant :

$$(P_a) \begin{cases} \min \ f(x) \\ s.c. \quad Ax = b \\ \qquad\quad x \geq 0 \end{cases}$$

avec $A \in I\!R^{m \times n}$ de rang m, $b \in I\!R^m$, $x \in I\!R^n$ et $f : I\!R^n \to I\!R$ une fonction convexe et continûment différentiable.

Notons $F = \{ \ x \in I\!R^n \mid Ax = b \text{ et } x \geq 0 \ \}$, l'*ensemble admissible* du problème (P_a).

Supposons qu'il existe un point $x^{(0)} \in F$ dont les composantes sont strictement positives et posons

$$S = \{ \ x \in F \mid f(x) \leq f(x^{(0)}) \ \} \qquad \text{et} \qquad S^0 = \{ \ x \in S \mid x > 0 \ \}.$$

Ces ensembles sont appelés respectivement *ensemble admissible effectif* et *ensemble des points intérieurs admissibles.*

Nous allons maintenant nous intéresser aux conditions à imposer à ces différents ensembles pour pouvoir définir la méthode.

2.1.2 Hypothèses

Le cadre défini par le problème (P_a) est un peu trop vaste. Les trois hypothèses suivantes vont nous permettre de le restreindre, de définir la méthode affine et d'établir sa convergence.

1. L'intérieur relatif de F est non vide,
   c'est-à-dire il existe $x^{(0)} \in F$ t.q. $x^{(0)} > 0$.

2. L'ensemble S est borné.
   Dans le cas où f est convexe, cela revient à exiger que l'ensemble des solutions optimales de (P_a) soit non vide et borné.

3. Tout point de S est primal non dégénéré dans le sens décrit ci-après.

2.1.3 Non dégénérescence primale

Considérons un point x de S.
Posons $N = \{ \ i = 1, ..., n \mid x_i = 0 \ \}$ et $B = \{ \ 1, ..., n \ \} \backslash N$.
On notera A_B la sous matrice de A correspondant aux composantes dont les indices appartiennent à B.
On définit alors la *non dégénérescence primale* de la manière suivante :

$$x \text{ est primal non dégénéré } \iff A_B \text{ est de rang } m.$$

La non dégéneresccence primale étant peu utilisée, nous allons faire le lien avec une notion plus courante, l'hypothèse de qualification de contraintes.

Lemme 2.1
Le point $x \in S$ satisfait l'hypothèse de qualification de contraintes si et seulement si x est primal non dégénéré.

<u>*Preuve*</u>

Commençons par traduire, dans le cadre du problème qui nous occupe, les qualifications de contraintes. Elles seront satisfaites si les gradients des contraintes d'inégalités projetés sur le noyau de la matrice A sont linéairement indépendants. Les contraintes d'inégalités pouvant se réécrire $I_i^T x \geq 0$ où I_i représente la i-ème colonne de la matrice identité, nous obtenons

$$\begin{array}{l}\text{les qualifications de contraintes sont satisfaites}\\ \text{si } \{P_A I_i\}_{i=1}^n \text{ sont linéairement indépendants,}\end{array}$$

où P_A représente la projection sur le noyau de la matrice A.

Passons à la démonstration de l'équivalence. Utilisons pour cela les contraposées.

Supposons x primal dégénéré et montrons que x ne satisfait pas l'hypothèse de qualification de contraintes.
Par définition, x est primal dégénéré si la matrice A_B n'est pas de rang m, c'est-à-dire s'il existe $y \in I\!R^n$ non nul tel que $A_B^T y = 0$.
Posons $\lambda_N = A_N^T y$ et $\lambda_B = 0$. Nous obtenons

$$\begin{array}{l}\lambda = A^T y\\ y \neq 0\\ A \text{ de rang plein.}\end{array}$$

Donc λ est non nul et $P_A \lambda = 0$, ce qui signifie que x ne satisfait pas les qualifications de contraintes.

Réciproquement, supposons que x ne satisfait pas les qualifications de contraintes et montrons que x est primal dégénéré.
Si x ne satisfait pas les qualifications de contraintes, il existe $\lambda \in I\!R^n$, non nul, tel que $\lambda_N \neq 0$, $\lambda_B = 0$ et $P_A \lambda = 0$. Cela implique qu'il existe $y \in I\!R^n$ tel que $\lambda = A^T y$; en particulier, $\lambda_B = A_B^T y = 0$.
Or $\lambda_N = A_N^T y \neq 0$, donc il existe $y \in I\!R^n$ non nul tel que $A_B^T y = 0$, ce qui signifie que A_B n'est pas de rang m.

$\square$

2.1.4 La méthode

L'idée de l'algorithme est de projeter, à chaque itération k, le gradient de f en $x^{(k)}$ sur le noyau de la matrice $AX^{(k)}$ où $X^{(k)}$ est la matrice diagonale

$$X^{(k)} = diag\ (x_1^{(k)}, x_2^{(k)}, ..., x_n^{(k)}).$$

L'opérateur de projection sur ce noyau, noté $P_{AX^{(k)}}$, est donné par la formule explicite suivante :

$$P_{AX^{(k)}} = I - X^{(k)} A^T \left(A\ (X^{(k)})^2\ A^T\right)^{-1} AX^{(k)}.$$

Remarquons que cette projection est bien définie dès que la matrice AX^2A^T est non singulière. C'est l'objet du théorème suivant.

Théorème 2.1
Soit $x \in S$ primal non dégénéré.
Alors la matrice AX^2A^T est non singulière.

Preuve

Comme X est diagonale, nous avons immédiatement

$$AX^2A^T = AXXA^T = (AX)(AX)^T.$$

Dès lors, AX^2A^T est non singulière si et seulement si AX est de rang m.

D'autre part, par hypothèse, x est non dégénéré et donc A_B est de rang m.
Comme $X_B > 0$, la matrice A_BX_B est également de rang m et donc AX est de rang m.
La thèse est alors immédiate.

□

2.1.5 L'algorithme

Soit donné $\theta \in]0,1[$ et $x^{(0)} \in S^0$, $k = 0$

Répéter $\quad \bar{z}^{(k)} = P_{AX^{(k)}} X^{(k)} \nabla f(x^{(k)}),$

$$\bar{d}^{(k)} = -\frac{\bar{z}^{(k)}}{\|\bar{z}^{(k)}\|},$$

$$d^{(k)} = X^{(k)} \bar{d}^{(k)},$$

$$z^{(k)} = (X^{(k)})^{-1}\ \bar{z}^{(k)},$$

$$\lambda_m = \sup \{ \lambda \mid x^{(k)} + \lambda d^{(k)} \geq 0 \}.$$

Choisir $\bar{\lambda} \in [\theta, \lambda_m[$,

$$\lambda_k \in \text{ argmin}\{ f(x^{(k)} + \lambda d^{(k)}) \mid \lambda \in [0, \bar{\lambda}] \},$$

$$h^{(k)} = \lambda_k\ d^{(k)},$$

$$x^{(k+1)} = x^{(k)} + h^{(k)},$$

$$k = k + 1,$$

jusqu'à la convergence.

Avant d'aller plus loin, montrons que la recherche linéaire est bien définie. Notons $e = (1, 1, ..., 1)^T$. Comme $e^T = (1, ..., 0)$ et $\|\bar{d}^{(k)}\| = 1$, le vecteur $e + \bar{d}^{(k)}$ a toutes ses composantes positives ou nulles et il en va de même pour le vecteur $x^{(k)} + d^{(k)} = X^{(k)}(e + \bar{d}^{(k)})$. Dès lors, la direction $d^{(k)}$ est admissible, λ_m est supérieur à 1 et l'intervalle $[\theta, \lambda_m[$ est non vide. D'autre part, comme la fonction f est continue, la longueur de pas λ_k est bien définie.

2.2 Propriétés des suites engendrées par l'algorithme

Avant d'étudier la convergence de la suite $(x^{(k)})_{k \in I\!N}$ engendrée par cette méthode, intéressons nous aux propriétés des différentes suites produites par l'algorithme précédent.

Lemme 2.2
La suite engendrée par l'algorithme est admissible.

Preuve

Nous procéderons par récurrence. Comme le point de départ $x^{(0)}$ est admissible, supposons que $x^{(k)}$ est admissible et montrons que $x^{(k+1)}$ le reste.

Par définition de $x^{(k+1)}$, nous avons

$$x^{(k+1)} = x^{(k)} + h^{(k)} = x^{(k)} + \lambda_k d^{(k)}$$

avec $\lambda_k \in [0, \lambda_m[$. Dès lors, en vertu de la définition de λ_m, nous obtenons $x^{(k+1)} \geq 0$.

D'autre part, nous avons successivement

$$\begin{aligned} Ax^{(k+1)} &= A\,[\; x^{(k)} + h^{(k)} \;] \\ &= A\,[\; x^{(k)} + \lambda_k d^{(k)} \;] \\ &= A\left[\; x^{(k)} - \lambda_k X^{(k)} \frac{P_{AX^{(k)}} X^{(k)} \nabla f(x^{(k)})}{\| \; P_{AX^{(k)}} X^{(k)} \nabla f(x^{(k)}) \; \|} \;\right]. \end{aligned}$$

Or, par définition de $P_{AX^{(k)}}$, $P_{AX^{(k)}} X^{(k)} \nabla f(x^{(k)}) \in Ker AX^{(k)}$ et en particulier, $d^{(k)} \in Ker A$. Mais puisque $Ax^{(k)} = b$ par l'hypothèse de récurrence, nous en déduisons immédiatement que $Ax^{(k+1)} = b$.

$\square$

Lemme 2.3
A chaque itération de l'algorithme, on a

$$\nabla f(x^{(k)})^T d^k = -\| \ X^{(k)} z^{(k)} \ \|.$$

En particulier, $d^{(k)}$ *est une direction de descente.*

<u>Preuve</u>

Par construction de la suite $(d^{(k)})$, nous avons immédiatement que

$$\nabla f(x^{(k)})^T d^{(k)} = \nabla f(x^{(k)})^T X^{(k)} \bar{d}^{(k)} = - \frac{\nabla f(x^{(k)})^T X^{(k)} \bar{z}^{(k)}}{\|\bar{z}^{(k)}\|}.$$

Comme $\bar{z}^{(k)} \in Ker AX^{(k)}$, nous obtenons, par caractérisation de la projection orthogonale sur un sous-espace vectoriel, la relation suivante

$$(\bar{z}^{(k)})^T (u - P_{AX^{(k)}} u) = 0 \qquad \forall \ u \in I\!R^n.$$

En particulier, pour $u = X^{(k)} \nabla f(x^{(k)})$, nous avons

$$(\bar{z}^{(k)})^T X^{(k)} \nabla f(x^{(k)}) = (\bar{z}^{(k)})^T P_{AX^{(k)}} X^{(k)} \nabla f(x^{(k)}),$$

et par conséquent, en utilisant l'algorithme, $(\bar{z}^{(k)})^T X^{(k)} \nabla f(x^{(k)}) = \| \ \bar{z}^{(k)} \ \|^2$.

Finalement, $\nabla f(x^{(k)})^T d^{(k)} = -\| \ \bar{z}^{(k)} \ \| = -\| \ X^{(k)} z^{(k)} \ \|$.

En particulier, $\nabla f(x^{(k)})^T d^{(k)} < 0$, si $\bar{z}^{(k)} \neq 0$.

□

Remarquons que par l'algorithme, $h^{(k)} = \lambda_k d^{(k)}$ avec $\lambda_k \geq 0$. C'est donc aussi une direction de descente et on peut lui appliquer les différents résultats obtenus pour $d^{(k)}$ dans la démonstration précédente, à savoir

$$h^{(k)} \in Ker A \qquad \text{car } Ker A \text{ est un sous-espace vectoriel}$$

$$\nabla f(x^{(k)})^T h^{(k)} = (z^{(k)})^T h^{(k)}. \tag{2.1}$$

Lemme 2.4

La suite $(f(x^{(k)}))$ est décroissante et bornée inférieurement.

<u>Preuve</u>

La suite $(f(x^{(k)}))$ est décroissante par construction de λ_k. Par conséquent, la suite $(x^{(k)})$ reste dans S. Comme S est borné et f continue, nous en déduisons immédiatement que $f(x^{(k)})$ est bornée inférieurement.

□

Lemme 2.5

Les suites $(z^{(k)})$ et $(h^{(k)})$ sont telles que la suite $(z^{(k)})^T h^{(k)}$ converge vers 0 lorsque k tend vers ∞.

<u>Preuve</u>

Supposons, par l'absurde, qu'il existe une sous-suite d'indices $K^0 \subset I\!N$ et $\delta > 0$ tels que $(z^{(k)})^T h^{(k)} \leq -\delta < 0$, pour tout $k \in K^0$.

Comme S est borné, la suite $(x^{(k)})_{k \in K^0}$ est bornée et par le théorème de Weiertrass, il existe une sous-suite de $(x^{(k)})_{k \in K^0}$ qui converge c'est-à-dire

$$\exists\ K \subset K^0 \qquad \exists\ \tilde{x} \in S \qquad \text{t.q.} \qquad x^{(k)} \longrightarrow \tilde{x} \qquad \text{pour } k \in K. \tag{2.2}$$

Par construction, $h^{(k)} = x^{(k+1)} - x^{(k)}$. Dès lors, en restreignant K, si nécessaire,

$$\exists\ \tilde{h} \in KerA \text{ t.q. } h^{(k)} \longrightarrow \tilde{h} \qquad \text{pour } k \in K.$$

Soit $\theta \in]0,1[$ fixé à l'initialisation et $\lambda \in]0,\theta]$ choisi arbitrairement. Par construction (recherche linéaire) et par différentiabilité de f, nous avons, pour tout $k \in K$,

$$\begin{aligned} f(x^{(k+1)}) &\leq f(x^{(k)} + \lambda h^{(k)}) \\ &= f(x^{(k)}) + \lambda \nabla f(x^{(k)})^T h^{(k)} + o(x^{(k)}, \lambda h^{(k)}) \\ &= f(x^{(k)}) + \lambda (z^{(k)})^T h^{(k)} + o(x^{(k)}, \lambda h^{(k)}) \end{aligned}$$

où la dernière égalité est obtenue par (2.1).

En appliquant l'hypothèse absurde, nous obtenons, pour tout $k \in K$,

$$f(x^{(k+1)}) - f(x^{(k)}) \leq -\lambda\delta + o(x^{(k)}, \lambda h^{(k)}). \tag{2.3}$$

Mais, par le lemme 2.4, la suite $(f(x^{(k)}))$ est décroissante et bornée inférieurement, ce qui signifie que $f(x^{(k+1)}) - f(x^{(k)}) \longrightarrow 0$ pour tout $k \in I\!N$, et en particulier pour tout $k \in K$.

Dès lors, si nous passons à la limite dans (2.4) en utilisant (2.2) et (2.3), nous obtenons

$$0 \leq -\lambda\delta + o(\tilde{x}, \lambda\tilde{h}).$$

Or λ ayant été choisi arbitrairement dans l'intervalle $]0, \theta]$, nous pouvons le faire tendre vers 0 et par conséquent

$$\lim_{\lambda \to 0} \frac{o(\tilde{x}, \lambda\tilde{h})}{\lambda} \geq \delta > 0,$$

ce qui est impossible car f est une fonction différentiable.

$\square$

2.3 Estimation des multiplicateurs

Le problème (P_a) est défini par m contraintes d'égalités ($-Ax+b=0$) et n contraintes d'inégalités ($-x \leq 0$). Nous allons utiliser les conditions de Kuhn-Tucker pour trouver la solution optimale de ce problème. Comme l'hypothèse de qualification de contraintes est satisfaite, elles se traduisent de la manière suivante :

$$x^* \text{ est solution optimale} \Longleftrightarrow \begin{cases} \exists\ w^* \in I\!R^m\ \exists\ z^* \in I\!R^n\ t.q. \\ \nabla f(x^*) - A^T w^* - z^* = 0 \\ z_i^* x_i^* = 0 \qquad i = 1, ..., n \\ z^* \geq 0\ ;\ Ax^* = b. \end{cases}$$

Essayons de simplifier ces conditions en projetant la première sur le noyau de A.
Nous obtenons

$$\begin{aligned} P_A \nabla f(x^*) &= P_A(A^T w^* + z^*) \\ &= P_A A^T w^* + P_A z^* \\ &= P_A z^* \qquad \text{car } A^T w^* \perp KerA. \end{aligned}$$

Ces conditions correspondent à la linéarisation du problème autour du point x^*.

Convenons d'appeler *estimateur des multiplicateurs au point* $x \in S$ tout vecteur $z \in I\!R^n$ vérifiant

$$\boxed{P_A z = P_A \nabla f(x).}$$

Nous obtenons alors les deux résultats suivants :

- $z^{(k)}$ est un estimateur des multiplicateurs au point $x^{(k)}$.
- $\bar{z}^{(k)}$ est un estimateur des multiplicateurs au point $x^{(k)}$ pour le problème dans l'espace de projection c'est-à-dire $P_{AX^{(k)}} \bar{z}^{(k)} = P_{AX^{(k)}} X^{(k)} \nabla f(x^{(k)})$.

En effet, nous avons par l'algorithme

$$\bar{z}^{(k)} = P_{AX^{(k)}} X^{(k)} \nabla f(x^{(k)}).$$

Dès lors, $X^{(k)} \nabla f(x^{(k)}) = \bar{z}^{(k)} + X^{(k)} A^T w$, avec $\bar{z}^{(k)} \in Ker AX^{(k)}$ et $w \in I\!R^m$.
Nous en déduisons :

$$\begin{aligned} \nabla f(x^{(k)}) &= (X^{(k)})^{-1} \bar{z}^{(k)} + (X^{(k)})^{-1} X^{(k)} A^T w \\ &= z^{(k)} + A^T w \end{aligned}$$

et en projetant sur $Ker A$:

$$\begin{aligned} P_A \nabla f(x^{(k)}) &= P_A (z^{(k)} + A^T w) \\ &= P_A z^{(k)}. \end{aligned}$$

Pour établir le deuxième résultat, il suffit d'utiliser l'algorithme et le fait que $\bar{z}^{(k)} \in Ker AX^{(k)}$ et donc $\bar{z}^{(k)} = P_{AX^{(k)}} \bar{z}^{(k)}$.

Grâce à ces estimateurs, nous allons pouvoir démontrer la convergence de la suite.

2.4 Convergence de la méthode affine

Avec les hypothèses retenues, nous ne pourrons établir la convergence de la suite $(x^{(k)})$ engendrée par l'algorithme de la méthode affine. Nous montrerons seulement que tout point d'accumulation de cette suite est solution optimale du problème (P_a). Pratiquement, nous allons choisir un point d'accumulation de la suite $(x^{(k)})$ et montrer que l'estimateur des multiplicateurs en ce point est le vecteur des multiplicateurs de Lagrange.

Soit x^* un point d'accumulation de la suite $(x^{(k)})_{k \in I\!N}$. Notons z^* l'estimateur des multiplicateurs au point x^* :

$$P_A z^* = P_A \nabla f(x^*).$$

Pour montrer que le vecteur z^* est le vecteur des multiplicateurs de Lagrange associé au point x^*, nous devons montrer que les quatre équations suivantes sont vérifiées :

$$\begin{aligned}
&\nabla f(x^*) = A^T w + z^* \\
&Ax^* = b \\
&z_i^* x_i^* = 0 \\
&z^* \geq 0.
\end{aligned}$$

La première de ces équations est trivialement satisfaite puisque z^* est un estimateur des multiplicateurs au point x^*; la seconde l'est aussi car x^* est un point d'accumulation de la suite et donc un point admissible. Le théorème 2.2 va nous permettre de montrer que la troisième est également vérifiée. Ensuite nous introduirons deux ensembles ainsi que quelques-unes de leurs propriétés. Cela nous permettra de vérifier la quatrième équation.

Théorème 2.2
Les suites $(x^{(k)})$ et $(z^{(k)})$ sont telles que $X^{(k)}z^{(k)}$ converge vers 0 lorsque k tend vers ∞.

<u>*Preuve*</u>

Supposons, par l'absurde, qu'il existe une sous-suite d'indices $K^0 \subset I\!N$ et $\delta > 0$ tels que $\| X^{(k)}z^{(k)} \| \geq \delta > 0$, pour tout $k \in K^0$.

Comme S est borné, la suite $(x^{(k)})_{k\in K^0}$ est bornée et par le théorème de Weierstrass, il existe une sous-suite de $(x^{(k)})_{k\in K^0}$ qui converge, c'est-à-dire

$$\exists\, K \subset K^0 \quad \exists\, \tilde{x} \in S \qquad \text{t.q.} \qquad x^{(k)} \longrightarrow \tilde{x} \qquad \text{pour } k \in K.$$

Par construction, $z^{(k)} = (X^{(k)})^{-1} P_{AX^{(k)}} X^{(k)} \nabla f(x^{(k)})$ et $d^{(k)} = -X^{(k)} \frac{x^{(k)}z^{(k)}}{\| X^{(k)}z^{(k)} \|}$.
Dès lors, en restreignant K, si nécessaire,

$$\begin{array}{llll} \exists\, \tilde{z} \in I\!R^n \text{ t.q.} & z^{(k)} \longrightarrow \tilde{z} & \text{pour } k \in K, \\ \exists\, \tilde{d} = \dfrac{-\tilde{X}^2\tilde{z}}{\| \tilde{X}\tilde{z} \|} \text{ t.q.} & d^{(k)} \longrightarrow \tilde{d} & \text{pour } k \in K. \end{array}$$

Soit $\theta \in]0,1[$ fixé à l'initialisation et $\lambda \in]0,\theta]$ choisi arbitrairement. Par construction (recherche linéaire) et par différentiabilité de f, nous avons, pour tout $k \in K$:

$$\begin{aligned} f(x^{(k+1)}) &= f(x^{(k)} + \lambda_k d^{(k)}) \\ &\leq f(x^{(k)} + \lambda d^{(k)}) \\ &= f(x^{(k)}) + \lambda \nabla f(x^{(k)})^T d^{(k)} + o(x^{(k)}, \lambda d^{(k)}) \\ &= f(x^{(k)}) - \lambda \| X^{(k)}z^{(k)} \| + o(x^{(k)}, \lambda d^{(k)}), \end{aligned}$$

où la dernière égalité est obtenue par le lemme 2.3.

En introduisant l'hypothèse absurde, nous obtenons pour $k \in K$

$$f(x^{(k+1)}) - f(x^{(k)}) \leq -\lambda\delta + o(x^{(k)}, \lambda d^{(k)}). \tag{2.4}$$

Par le lemme 2.4, la suite $(f(x^{(k)}))$ est décroissante et bornée inférieurement, ce qui signifie que $f(x^{(k+1)}) - f(x^{(k)}) \longrightarrow 0$ pour tout $k \in I\!N$, et en particulier pour tout $k \in K$.

Passons à la limite dans (2.5), cela nous donne :

$$0 \leq -\lambda\delta + o(\tilde{x}, \lambda\tilde{d}).$$

Or, λ ayant été choisi arbitrairement dans $]0, \theta]$, nous pouvons le faire tendre vers 0 :

$$\lim_{\lambda \to 0} \frac{o(\tilde{x}, \lambda\tilde{d})}{\lambda} \geq \delta > 0,$$

ce qui est impossible car f est une fonction différentiable.

□

Intéressons nous maintenant à la quatrième de ces équations.
Définissons, dans ce but, $M = \{\ i = 1, 2, ..., n \mid z_i^* \neq 0\ \}$ et $\bar{M} = \{1, 2, ..., n\} \backslash M$. Notons par x_M les composantes du vecteur x dont les indices appartiennent à M. Par le théorème 2.2, $x_i^* z_i^* = 0$, et donc, par définition de M, $x_M^* = 0$. Le point d'accumulation x^* appartient par conséquent à la face de l'ensemble admissible caractérisée par $x_M = 0$.

Sur cette face, définissons l'*ensemble terminal* par

$$\Gamma = \{\ x \in S \mid x_M = 0 \text{ et } f(x) = f(x^*)\ \}$$

et un *voisinage* de Γ par

$$\Gamma_\delta = \{\ x \in S \mid \|x - y\|_\infty \leq \delta \text{ et } y \in \Gamma\ \},$$

où δ est un nombre strictement positif.

Commençons par étudier les propriétes de l'ensemble Γ.

Lemme 2.6

L'ensemble Γ est convexe.

<u>*Preuve*</u>

Notons $co\Gamma$ l'enveloppe convexe de Γ et montrons que $\Gamma = co\Gamma$.

Par définition de $co\Gamma$, il suffit de montrer que $co\Gamma \subset \Gamma$.

Soit $x \in co\Gamma$. Comme x est une combinaison convexe de points de Γ, $x_M = 0$.

Mais, par convexité de la fonction f, $f(x) \leq f(x^*)$; il reste donc à montrer

$$f(x) \geq f(x^*).$$

Considérons la direction $h = x - x^*$. Par convexité de f,

$$f(x) \geq f(x^*) + \nabla f(x^*)^T h.$$

Or, $h \in KerA$ car x et x^* sont des points admissibles et z^* est un estimateur des multiplicateurs au point x^*,d'où

$$\nabla f(x^*)^T h = (z^*)^T h.$$

De plus, $h_M = 0$ car x et x^* appartiennent à Γ et les composantes de z^* correspondant aux indices n'appartenant pas à M sont nulles.

Finalement, nous obtenons $f(x) \geq f(x^*)$.

$\square$

Lemme 2.7
Soit $f : I\!R^n \longrightarrow I\!R$ une fonction convexe et continûment différentiable.
Supposons que f est constante sur un ensemble convexe Γ de $I\!R^n$.
Alors ∇f est constant sur Γ.

<u>*Preuve*</u>

Soit $epif = \{ \ u = (y,t) \in I\!R^n \times I\!R \mid t \geq f(y) \ \}$, l'épigraphe de f.
Nous utiliserons le résultat suivant (cf. [8]) : pour tout x, nous avons l'équivalence

$$g = \nabla f(x) \qquad \Longleftrightarrow \qquad \begin{array}{l} \beta = (g,-1) \text{ définit un hyperplan } \beta u = \alpha \\ \text{qui sépare } (x, f(x)) \text{ de } epif. \end{array}$$

Notons par $ri\Gamma$ l'intérieur relatif de Γ. Soit $\bar{x} \in ri\Gamma$.
Posons $\Psi = \{ \ (x, f(\bar{x})) \mid x \in \Gamma \ \}$.
Comme f est constante sur Γ, cet ensemble est contenu dans $epif$.
Soit $\bar{u} = (\bar{x}, f(\bar{x})) \in ri\Psi$.
Nous obtenons que $\bar{\beta} = (\nabla f(\bar{x}), -1)$ définit un hyperplan $\bar{\beta}\bar{u} = \bar{\alpha}$ qui sépare $(\bar{x}, f(\bar{x}))$ de $epif$. Notons π cet hyperplan.

Comme Ψ est contenu dans $epif$, π sépare $(\bar{x}, f(\bar{x}))$ de Ψ, ce qui signifie que

$$\bar{\beta}u \leq \bar{\alpha} \qquad \text{pour tout } u \in \Psi.$$

Montrons que $\Psi \subset \pi$.
Supposons par l'absurde qu'il existe $u_0 \in \Psi$ tel que $u_0 \notin \pi$, c'est-à-dire $\bar{\beta}u_0 < \bar{\alpha}$.
Comme $\bar{u} \in ri\Psi$, il existe $\delta > 0$ tel que $u = \bar{u} + \delta(\bar{u} - u_0) \in \Psi$. Dès lors nous avons

$$\begin{aligned} \bar{\beta}u &= \bar{\beta}\bar{u} + \delta\bar{\beta}\bar{u} - \delta\bar{\beta}u_0 \\ &= (1+\delta)\bar{\alpha} - \delta\bar{\beta}u_0 > \bar{\alpha} \end{aligned}$$

car $\bar{\beta}u_0 < \bar{\alpha}$. Ceci est impossible puisque $u \in \Psi$. Donc $\Psi \subset \pi$.

En particulier, pour tout $u \in \Psi$, nous avons $\bar{\beta}u = \bar{\alpha}$. Cela qui signifie que π sépare $u \in \Psi$ de Ψ ou encore que $\bar{\beta} = (\nabla f(\bar{x}), -1)$ définit un séparateur en chaque point u de Ψ.

Montrons que $\bar{\beta}$ définit un hyperplan séparant $(x, f(x))$ de $epif$ pour tout $x \in \Gamma$, c'est-à-dire

$$\bar{\beta}(y,t) \leq \bar{\alpha} \qquad \text{pour tout } (y,t) \in epif.$$

Procédons en trois étapes.

1) Le résultat est vrai pour les couples $(y,t) \in \Gamma \times I\!R$ tels que $t \geq f(y)$.
En effet pour de tels couples nous avons

$$\bar{\beta}(y,t) \leq \bar{\beta}(y, f(y)) = \bar{\beta}(y, f(\bar{x})) = \bar{\alpha},$$

car $\bar{\beta} = (\nabla f(\bar{x}), -1)$ et $(y, f(\bar{x})) \in \Psi$.

2) Vérifions ensuite le résultat pour les couples $(y, f(y)) \in I\!R^n \times I\!R$.
Pour obtenir une contradiction, supposons qu'il existe $\bar{y} \in I\!R^n$ tel que $\bar{\beta}(\bar{y}, f(\bar{y})) > \bar{\alpha}$.
Dans ce cas, $\bar{y} \notin \Gamma$.
Comme $\bar{x} \in ri\Gamma$, par convexité de Γ, il existe $\lambda \in]0,1[$ tel que $y = \lambda\bar{x} + (1-\lambda)\bar{y} \in \Gamma$.
Comme f est une fonction convexe, nous obtenons :

$$\begin{aligned}
\bar{\beta}(y, f(y)) &\geq \bar{\beta}\ (\ \lambda\bar{x} + (1-\lambda)\bar{y}\ ,\ \lambda f(\bar{x}) + (1-\lambda) f(\bar{y})\ ) \\
&= \bar{\beta}\ [\ \lambda(\bar{x}, f(\bar{x})) + (1-\lambda)(\bar{y}, f(\bar{y}))\ ] \\
&= \lambda\ \bar{\beta}\ (\bar{x}, f(\bar{x})) + (1-\lambda)\ \bar{\beta}\ (\bar{y}, f(\bar{y})) \\
&> \lambda\ \bar{\alpha} + (1-\lambda)\ \bar{\alpha} = \bar{\alpha},
\end{aligned}$$

ce qui est impossible car $y \in \Gamma$.

3) Enfin le résultat est valable pour les couples $(y,t) \in epif$ car:

$$\bar{\beta}(y,t) \leq \bar{\beta}(y, f(y)) \leq \bar{\alpha}.$$

En conséquence, $\bar{\beta} = (\nabla f(x), -1)$.

Comme par définition $\bar{\beta} = (\nabla f(\bar{x}), -1)$, nous en déduisons que, pour tout $x \in \Gamma$

$$\nabla f(x) = \nabla f(\bar{x}).$$

$\square$

Lemme 2.8
Si on note $z(x) = X^{-1}P_{AX}X\nabla f(x)$, pour tout $x \in I\!R^n$, alors

$$z(x) = z^*,$$

pour tout $x \in \Gamma$.

Preuve

Par définition de l'estimateur des multiplicateurs, nous avons

$$z^* = \nabla f(x^*) + A^T w \qquad \text{avec } w \in I\!R^m.$$

Par le lemme 2.7, ∇f est constant sur Γ et donc, pour tout $x \in \Gamma$,

$$\nabla f(x) = \nabla f(x^*) = z^* - A^T w.$$

Introduisons cette égalité dans la formule de $z(x)$. Nous obtenons

$$\begin{aligned} z(x) &= X^{-1}[\; I - XA^T(AX^2A^T)^{-1}AX \;]\; X\nabla f(x) \\ &= [\; I - A^T(AX^2A^T)^{-1}AX^2 \;]\; [\; z^* - A^T w \;]. \end{aligned}$$

Remarquons que $Xz^* = 0$.
En effet, pour les indices $i \in M$, $x_i = 0$ car $x \in \Gamma$, et pour les autres indices, $z^* = 0$ par définition de M. Donc

$$(Xz^*)_i = x_i z_i^* = 0 \qquad \text{pour } i = 1, 2, .., n.$$

On obtient finalement

$$z(x) = z^* - A^T w + A^T w = z^*.$$

□

Etudions maintenant les propriétés du voisinage Γ_δ de Γ.

Lemme 2.9
Si on note $d(x) = -X\frac{z(x)}{\|z(x)\|}$, *pour tout* $x \in I\!R^n$,
alors il existe $\delta > 0$ *tel que les deux relations suivantes sont vérifiées pour tout* $x \in \Gamma_\delta$:

$$(1)\ \|z(x) - z^*\| \leq \varepsilon \text{ où } \varepsilon = \frac{1}{2} \min_{i \in M} |z_i^*|,$$
$$(2)\ \text{si } x > 0, d_i(x) z_i^* < 0 \text{ pour tout } i \in M.$$

Preuve

(1) $z(.)$ est uniformément continue sur S puisque $z(.)$ est continue en chaque point de S et que S est borné. Il existe donc $\delta > 0$ tel que pour tout $x, y \in S$,

$$\|x - y\|_\infty \leq \delta \quad \Longrightarrow \quad \| z(x) - z(y) \| \leq \varepsilon.$$

Par définition de Γ_δ, pour chaque $x \in \Gamma_\delta$, il existe $y \in \Gamma$ tel que

$$\|x - y\|_\infty \leq \delta.$$

Par le lemme 2.8, $z(y) = z^*$ pour tout $y \in \Gamma$, et donc

$$\| z(x) - z^* \| \leq \varepsilon.$$

(2) Soit $x > 0$. Par définition de ε, pour tout $i \in M$,

$$| z_i(x) - z_i^* | \leq \| z(x) - z^* \| \leq \varepsilon \leq \frac{1}{2} |z_i^*|.$$

En utilisant cette relation, on peut montrer que pour tout $i \in M$, $z_i(x)$ et z_i^* sont de même signe. Supposons le contraire, deux cas sont alors possibles :

a) $z_i(x) > 0 > z_i^*$.
$\frac{1}{2}z_i^* \leq z_i(x) - z_i^* \leq -\frac{1}{2}z_i^*$, et donc $z_i(x) \leq \frac{1}{2}z_i^*$, ce qui contredit l'hypothèse.

b) $z_i(x) < 0 < z_i^*$.
$-\frac{1}{2}z_i^* \leq z_i(x) - z_i^* \leq \frac{1}{2}z_i^*$, et donc $\frac{1}{2}z_i^* \leq z_i(x)$, ce qui contredit l'hypothèse.

Par définition de $d(x)$, pour tout $i \in M$, $d_i(x)$ et $z_i(x)$ sont de signes opposés, ce qui termine la démonstration.

□

Le résultat suivant va nous permettre d'établir un lien entre les points d'accumulation de la suite et les ensembles Γ et Γ_δ.

Lemme 2.10
Soit $\tilde{x}$ un point d'accumulation de la suite $(x^{(k)})$.
Alors ou $\tilde{x} \in \Gamma$ ou $\tilde{x} \notin \Gamma_\delta$.

Preuve

Supposons, par l'absurde, que $\tilde{x} \in \Gamma_\delta$ et $\tilde{x} \notin \Gamma$.

Comme $\tilde{x}$ et x^* sont deux points d'accumulation, il existe deux sous-suites de $(x^{(k)})$ qui convergent, l'une vers $\tilde{x}$ et l'autre vers x^*.

Puisque la fonction f est convexe et que la suite $(f(x^{(k)}))$ est décroissante par le lemme 2.4, cette suite converge. Comme f est continue et par unicité de la limite,

$$f(\tilde{x}) = f(x^*).$$

Il existe donc un indice $i \in M$ tel que $\tilde{x}_i > 0$.

Appliquant le lemme 2.9, nous obtenons

$$d_i(\tilde{x}) z_i(\tilde{x}) < 0 \qquad \text{pour tout } i \in M.$$

Nous en concluons que $z_i(\tilde{x}) \neq 0$ pour tout $i \in M$, et donc $\tilde{x}_i z_i(\tilde{x}) \neq 0$ pour au moins un indice $i \in M$, ce qui contredit le théorème 2.2.

□

Voici le résultat important de cette section, la vérification de la quatrième équation citée page 32.

Théorème 2.3
Soit x^* un point d'accumulation et z^* un estimateur des multiplicateurs au point x^*.
Posons $M = \{i = 1, 2, ..., n \mid z_i^* \neq 0\}$.
Alors $z_M^* > 0$.

Preuve

Supposons, par l'absurde, que z_M^* a au moins une composante négative.
Supposons de plus que les deux assertions suivantes sont vérifiées :

(1) Il existe $K \subset I\!N$ tel que $x^{(k)} \longrightarrow \bar{x}^1$ et $x^{(k+1)} \longrightarrow \bar{x}^2$ où $\bar{x}^1 \in \Gamma$ et $\bar{x}^2 \notin \Gamma_\delta$.

(2) Il existe $j \in M$ tel que $z_j^* < 0$ et $\bar{x}_j^2 > \bar{x}_j^1 = 0$.

Considérons le pas $h^{(k)} = x^{(k+1)} - x^{(k)}$ et étudions la suite $(z^{(k)})^T h^{(k)}$.
Par le lemme 2.5, cette suite converge vers 0. Passons à la limite dans K :

$$(z^{(k)})^T h^{(k)} = (z^{(k)})^T (x^{(k+1)} - x^{(k)}) \longrightarrow (z^*)^T (\bar{x}^2 - \bar{x}^1) = (z_M^*)^T (\bar{x}_M^2 - \bar{x}_M^1)$$

puisque les autres composantes de z^* sont nulles.

Par unicité de la limite, nous avons

$$(z_M^*)^T (\bar{x}_M^2 - \bar{x}_M^1) = 0. \tag{2.5}$$

Or, pour $k \in K$ suffisament grand, $x^{(k)} \in \Gamma_\delta$ car, par le lemme 2.10, $\bar{x}^1 \in \Gamma_\delta$.

Par le lemme 2.9, $z_i^{(k)} h_i^{(k)} = z_i^{(k)} \frac{d_i^{(k)}}{\lambda_k} < 0$, pour $i \in M$.
Passons à la limite dans K :

$$z_i^* (\bar{x}_i^2 - \bar{x}_i^1) \leq 0 \qquad \text{pour } i \in M.$$

Mais une de ces inégalités est stricte, en vertu de la deuxième assertion, ce qui contredit l'égalité (2.6).

Pour que la preuve soit complète, il reste à montrer que les deux assertions sont vérifiées.

(1) Par le lemme 2.10, il existe des points d'accumulation dans Γ et à l'extérieur de Γ_δ. Par conséquent, il existe $K^0 \subset I\!N$ tel que, pour tout $k \in K^0$, $x^{(k)} \in \Gamma$ et $x^{(k+1)} \notin \Gamma_\delta$. Comme S est borné, la suite $(x^{(k)})_{k\in K^0}$ est bornée et par le théorème de Weiertrass, il existe une sous-suite de $(x^{(k)})_{k\in K^0}$ qui converge c'est-à-dire

$$\exists\ K \subset K^0 \qquad \exists\ \bar{x}^1, \bar{x}^2 \text{ t.q.} \qquad x^{(k)} \longrightarrow \bar{x}^1 \qquad \text{et} \qquad x^{(k+1)} \longrightarrow \bar{x}^2 \qquad \text{pour } k \in K.$$

Par définition, Γ_δ est fermé. Il contient donc les limites de ses suites convergentes et en particulier $\bar{x}^1$. Par le lemme 2.10, $\bar{x}^1 \in \Gamma$.

Supposons $\bar{x}^2 \in \Gamma_\delta$ et donc $\bar{x}^2 \in \Gamma$ par le lemme 2.10.
Pour $k \in K$ suffisament grand, $\| x^{(k+1)} - \bar{x}^2 \|_\infty \leq \delta$ car $x^{(k+1)} \longrightarrow \bar{x}^2$.
Cela signifie $x^{(k+1)} \in \Gamma_\delta$, ce qui est impossible.
Par conséquent, $\bar{x}^2 \notin \Gamma_\delta$.
(2) Comme $\bar{x}^1 \in \Gamma$, $\bar{x}^1_M = 0$.
Puisque la fonction f est convexe et que la suite $(f(x^{(k)}))$ est décroissante par le lemme 2.4, elle converge. Comme f est continue et par unicité de la limite,

$$f(\bar{x}^1) = f(\bar{x}^2).$$

Or, $\bar{x}^2 \notin \Gamma$, ce qui signifie qu'il existe un indice $j \in M$ tel que $\bar{x}^2_j > 0$.

Il reste à montrer $z^*_j < 0$.
Considérons le pas $h^{(k)} = x^{(k+1)} - x^{(k)}$.
Pour $k \in K$ suffisament grand, $h^{(k)}_j > 0$ et $x^{(k)} \in \Gamma_\delta$. Par conséquent,

$$d^{(k)}_j = \frac{h^{(k)}_j}{\lambda_k} > 0.$$

Or, par le lemme 2.9, $d^{(k)}_j z^*_j < 0$, et dès lors $z^*_j < 0$.

□

Chapitre 3

Méthode du point intérieur

Une des difficultés liées à la méthode affine présentée au chapitre précédent est de vérifier l'hypothèse de non dégénérescence primale. Une façon de contourner cet obstacle est de résoudre un problème dont tous les points admissibles sont primals non dégénérés. Nous allons, dans cette optique, définir les fonctions barrières. Nous nous intéresserons plus particulièrement aux fonctions barrières logarithmiques dont nous étudierons les propriétés plus en détails. Nous terminerons ce chapitre par une analyse de l'erreur commise lorsque l'on résout un problème de minimisation avec contraintes où l'on a ajouté une fonction barrière logarithmique à la fonction objectif.

3.1 Fonctions barrières

Dans ce chapitre , nous allons nous intéresser au problème (P)

$$(P)\begin{cases} \min \ f(x) \\ s.c. \quad Ax = b \\ \qquad\quad x \geq 0 \end{cases}$$

avec $A \in I\!R^{m\times n}$ de rang m, $b \in I\!R^m$, $x \in I\!R^n$ et $f : I\!R^n \to I\!R$ une fonction convexe, non différentiable et Lipschitz continue de constante L.

Notons $F = \{ \ x \in I\!R^n \mid Ax = b \text{ et } x \geq 0 \ \}$, l'*ensemble admissible* du problème (P), et $F^0 = \{ \ x \in X \mid Ax = b \text{ et } x > 0 \ \}$.

Notons aussi par Ω, l'*ensemble des solutions optimales.*

Supposons que le problème (P) admet un minimum global, que F est borné et que F^0 est non vide.

Nous dirons que la fonction B est une *fonction barrière* si elle vérifie les trois propriétés suivantes :

1) B est définie sur F^0 et est continue.

2) $B(x) \geq 0$ pour tout $x \in F^0$.

3) Si $(x^{(k)})$ est une suite dans F^0 qui converge vers un point frontière de F, alors $B(x^{(k)})$ converge vers $+\infty$.

Au lieu de résoudre le problème (P), nous allons résoudre

$$\begin{cases} \min \ f(x) + \mu \ B(x) \\ s.c. \quad x \in F^0 \end{cases}$$

où le paramètre μ est un nombre réel strictement positif.

3.2 Barrière logarithmique

Nous allons travailler plus particulièrement avec la *fonction barrière logarithmique* définie de la manière suivante

$$B(x) = -\sum_{i=1}^{n} \ln x_i \qquad \text{pour tout } x \in F^0.$$

Comme F est borné, nous pouvons considérer $B(x)$ comme une fonction barrière.

Posons, pour tout $x \in F^0$,

$$f_\mu(x) = f(x) - \mu \sum_{i=1}^{n} \ln x_i.$$

Puisque la fonction f est convexe et que le logarithme est stritement concave, la fonction f_μ est strictement convexe.

Le problème à résoudre est alors

$$(P_\mu) \begin{cases} \min \ f_\mu(x) \\ s.c. \quad x \in F^0. \end{cases}$$

Comme la fonction f_μ est seulement définie pour les vecteurs $x \in I\!R^n$ dont toutes les composantes sont strictement positives, la matrice A_B, définie au chapitre précédent, correspond à la matrice A, pour tout $x \in F^0$. Par conséquent, tous les points admissibles du problème (P_μ) sont primals non dégénérés dans le sens décrit à la section 2.1.3.

Calculons maitenant le sous-différentiel de la fonction f_μ.
Comme la fonction f_μ est la somme de fonctions convexes, nous obtenons :

$$\partial f_\mu(x) = \partial f(x) + \partial(-\mu \sum_{i=1}^{n} \ln x_i) = \partial f(x) + \nabla(-\mu \sum_{i=1}^{n} \ln x_i)$$

car la fonction logarithme est différentiable.
Si nous notons $X = diag(x_1, x_2, ..., x_n)$ et $e = (1, 1, ..., 1)^T \in I\!R^n$, nous obtenons finalement :

$$\partial f_\mu(x) = \partial f(x) - \mu X^{-1} e. \tag{3.1}$$

Caractérisons l'erreur commise si l'on résout le problème (P_μ) à la place du problème (P).

Théorème 3.1
Si $x^* \in F$ est une solution optimale du problème (P_μ), alors

$$f(x^*) \leq \min_{x \in F} f(x) + n\mu.$$

<u>*Preuve*</u>

Comme x^* est une solution optimale de (P_μ), 0 est un sous-gradient de la fonction f_μ au point x^*.

De plus, par (3.1),

$$\partial f_\mu(x^*) = \partial f(x^*) - \mu\ (X^*)^{-1}e,$$

et par conséquent, il existe $\bar{\gamma} \in \partial f(x^*)$ tel que

$$\bar{\gamma}^T - \mu\ e^T\ (X^*)^{-1} = 0$$

car (X^*) est une matrice diagonale.

Par définition du sous-différentiel, nous avons, pour tout $x \in F$,

$$f(x) \geq f(x^*) + \bar{\gamma}^T(x - x^*),$$

et en particulier, pour $x^{**} \in \Omega$,

$$\begin{aligned}
\min_{x \in F} f(x) = f(x^{**}) &\geq f(x^*) + \bar{\gamma}^T(x^{**} - x^*) \\
&= f(x^*) + \mu e^T(X^*)^{-1}(x^{**} - x^*) \\
&= f(x^*) + \mu\ [\ e^T(X^*)^{-1}x^{**} - e^T(X^*)^{-1}x^*\ ].
\end{aligned}$$

Chapitre 4

Résolution du problème

Nous allons maintenant nous intéresser à la résolution d'un problème de programmation linéaire, dans le cas où la fonction objectif est convexe, non différentiable et Lipschitz continue. Nous allons pour cela généraliser au cas non différentiable la méthode exposée au chapitre 2 et nous l'appliquerons au problème (P_μ) défini au chapitre précédent. Le choix du pas sera analogue à celui de la méthode du sous-gradient.

Ce chapitre est structuré de la manière suivante. Nous commencerons par resituer le problème et par présenter l'algorithme de la méthode. Nous établirons ensuite quelques résultats importants pour la convergence. Nous analyserons alors la convergence d'une sous-suite. Nous introduirons la condition d'angle et montrerons que, sous cette condition, toute la suite converge.

4.1 Algorithme

4.1.1 Position du problème

Dans ce chapitre, nous allons nous intéresser au problème suivant

$$(P)\begin{cases} \min \ f(x) \\ s.c. \quad Ax = b \\ \qquad\quad x \geq 0 \end{cases}$$

avec $A \in I\!R^{m\times n}$ de rang m, $b \in I\!R^m$, $x \in I\!R^n$, $f : I\!R^n \to I\!R$ une fonction convexe, non différentiable, et pour laquelle on peut calculer un sous-gradient en tout $x \in I\!R^n$. Nous supposerons en plus que f est Lipschitz continue de constante L.

Conservons les mêmes notations qu'au chapitre 3.

Le problème à résoudre est alors

$$(P_\mu)\begin{cases} \min \ f_\mu(x) \\ s.c. \quad x \in F^0. \end{cases}$$

Nous supposerons, comme dans le chapitre précédent, que F est borné et que F^0 est non vide.

Comme nous l'avons vu précédemment, le problème (P_μ) a une solution unique que nous noterons x^*.

4.1.2 L'itération

Nous allons nous baser sur la méthode affine pour construire l'algorithme. Nous savons déjà que tous les points admissibles pour le problème (P_μ) sont primals non dégénérés. Comme la fonction f n'est pas différentiable, nous allons généraliser la méthode présentée au chapitre 2, en remplaçant le gradient par un sous-gradient.

Supposons connu le point $x^{(k)} \in F^0$ et cherchons l'itéré suivant. Pour cela, nous commençerons par chercher une direction de descente $d^{(k)}$, solution du problème suivant :

$$(A.S.) \begin{cases} \min & \gamma_k^T d^{(k)} \\ s.c. & Ad^{(k)} = 0 \\ & \| (X^{(k)})^{-1} d^{(k)} \| \leq 1 \end{cases}$$

où $\gamma_k \in \partial f_\mu(x^{(k)})$.

La solution de ce problème est la direction affine, donnée par la formule ci-dessous :

$$d^{(k)} = -X^{(k)} \frac{P_{AX^{(k)}} X^{(k)} \gamma_k}{\| P_{AX^{(k)}} X^{(k)} \gamma_k \|}$$

où $P_{AX^{(k)}}$ représente la projection sur $Ker AX^{(k)}$.

Si nous choisissons une longueur de pas λ_k, l'itéré suivant $x^{(k+1)}$ sera donné par la formule

$$x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}.$$

Comme dans la méthode du sous-gradient, la suite (λ_k) doit satisfaire les deux hypothèses suivantes :

1) $\lim\limits_{k \to \infty} \lambda_k = 0$,

2) $\sum\limits_{k=1}^{\infty} \lambda_k = \infty$.

De plus, si $\lambda_k \in]0, 1[$, $x^{(k+1)}$ est admissible pour le problème (P_μ).

4.2 Résultats préliminaires

Nous allons présenter ici quelques propriétés utiles pour démontrer la convergence de la suite vers la solution optimale du problème (P_μ).

Commençons par caractériser les sous-gradients de f.

Lemme 4.1
Si $\bar{\gamma} \in \partial f(x)$, alors

$$\| \bar{\gamma} \| \leq L$$

où L est la constante de Lipschitz de f.

Preuve

Par la définition de sous-gradient, $\bar{\gamma} \in \partial f(x)$ si et seulememt si, pour tout $z \in F$,

$$f(z) - f(x) \geq \bar{\gamma}^T (z - x).$$

Dès lors, par définition de la norme, nous obtenons

$$\begin{aligned}
\| \bar{\gamma} \| &= \sup_{\|z-x\| \leq 1} |\bar{\gamma}^T (z - x)| \\
&= \sup_{\|z-x\| \leq 1} \bar{\gamma}^T (z - x) \\
&\leq \sup_{\|z-x\| \leq 1} (f(z) - f(x)) \\
&\leq \sup_{\|z-x\| \leq 1} L \|z - x\| \\
&\leq L,
\end{aligned}$$

où l'avant dernière inégalité résulte du caractère Lipschitz continu de f.

□

Lemme 4.2
Il existe une constante c_1 strictement positive telle que, pour tout $x \in F^0$,

$$\| X^{-1} P_{AX} X\gamma \| \leq c_1 \| \gamma \|$$

où $\gamma \in \partial f_\mu(x)$.

Preuve

Nous savons[1] qu'il existe une constante $\sigma > 0$ telle que

$$\| (AX^2A^T)^{-1}AX^2 \| \leq \sigma^{-1} \| (AA^T)^{-1}A \|.$$

En utilisant ce résultat et l'expression analytique de P_{AX}, nous obtenons

$$\begin{aligned}
\| X^{-1} P_{AX} X\gamma \| &= \| X^{-1} [I - XA^T(AX^2A^T)^{-1}AX] X \gamma \| \\
&= \| [I - A^T(AX^2A^T)^{-1}AX^2] \gamma \| \\
&= \| \gamma - A^T(AX^2A^T)^{-1}AX^2\gamma \| \\
&\leq (1 + \|A^T\| \| (AX^2A^T)^{-1}AX^2 \|) \| \gamma \| \\
&\leq (1 + \|A^T\| \| (AA^T)^{-1}A \| \sigma^{-1}) \| \gamma \|.
\end{aligned}$$

Il suffit alors de poser

$$c_1 = 1 + \|A^T\| \| (AA^T)^{-1}A \| \sigma^{-1}.$$

□

1. voir annexe A

Lemme 4.3

Il existe une constante c_2 strictement positive telle que, pour tout $x \in F^0$,

$$\| P_{AX} X\gamma \| \leq c_2$$

où $\gamma \in \partial f_\mu(x)$.

Preuve

Comme, par (3.1), pour tout $x \in F^0$, nous avons

$$\partial f_\mu(x) = \partial f(x) - \mu X^{-1} e,$$

le vecteur $\gamma \in \partial f_\mu(x)$ si et seulement si il existe $\bar{\gamma} \in \partial f(x)$ tel que

$$\gamma = \bar{\gamma} - \mu X^{-1} e.$$

Puisque la projection est une application contractante, nous avons

$$\begin{aligned} \| P_{AX} X\gamma \| &\leq \| X\gamma \| \\ &= \| X \, (\, \bar{\gamma} - \mu X^{-1} e) \| \\ &\leq \| X\bar{\gamma} \| + \mu \, \|e\| \\ &\leq \| X \| \, \| \bar{\gamma} \| + \mu\sqrt{n} \end{aligned}$$

par définition de la norme Euclidienne.

Or, nous avons supposé F borné, ce qui signifie qu'il existe $c > 0$ tel que, pour tout $x \in F$, $\|x\| \leq c$ et donc $\|X\| \leq c$.

De plus, par le lemme 4.1, $\| \bar{\gamma} \| \leq L$.

Par conséquent, si nous posons

$$c_2 = cL + \mu\sqrt{n},$$

nous obtenons le résultat cherché.

$\square$

Lemme 4.4

Soit $(x^{(k)})$ une suite dans F^0 convergeant vers $\bar{x}$ lorsque k tend vers ∞.
Alors $\lim_{k \to \infty} P_{AX^{(k)}} e = P_{A\bar{X}} e$.

Preuve

Comme F est fermé borné, il contient $\bar{x}$.

Dans le cas où $\bar{x} \in F^0$, en utilisant l'expression analytique de $P_{A\bar{X}}$, le résultat est immédiat.

Supposons donc $\bar{x} \in F$ et $\bar{x} \notin F^0$.

Notons $I = \{ \ i = 1, 2, ..., n \mid \bar{x}_i = 0 \ \}$ et $J = \{ \ i = 1, 2, ..., n \mid \bar{x}_i > 0 \ \}$.
L'ensemble I est non vide.

Avec cette notation, tout vecteur $x \in I\!R^n$ peut sécrire $x = (x_I, x_J)$.

Remarquons ensuite que $\bar{y} = P_{A\bar{X}} e$ et que $y^{(k)} = P_{AX^{(k)}} e$ sont respectivement solutions des problèmes

$$(\bar{P}) \begin{cases} \min & \|z - e\|^2 \\ s.c. & A\bar{X} z = 0 \end{cases}$$

et

$$(P^{(k)}) \begin{cases} \min & \|z - e\|^2 \\ s.c. & AX^{(k)} z = 0. \end{cases}$$

Le lemme sera démontré si on parvient à montrer que $\lim_{k \to \infty} y^{(k)} = \bar{y}$.

Nous allons procéder en deux étapes : nous établirons d'abord la convergence de la suite $(y_J^{(k)})$ vers $\bar{y}_J$ et ensuite celle de $(y_I^{(k)})$ vers $\bar{y}_I$.

1) Commençons par montrer que $\lim_{k\to\infty} y_J^{(k)} = \bar{y}_J$.

Dans ce but, définissons la multi-application $\theta : F^0 \to I\!R^{|J|}$ de la manière suivante :

$$\theta(x) = \{ \ z \in I\!R^{|J|} \mid A_J X_J z + A_I X_I y_I = 0 \ \}.$$

où $y = (y_I, y_J)$ désigne la projection de e sur $KerAX$.
Notons, pour tout $x \in F^0$,

$$\Theta(x) = argmin \ \{ \ \|z - e_J\| \mid z \in \theta(x) \ \}.$$

L'application $\Theta(x)$ est continue sur F^0 puisque la multi- application $\theta(x)$ l'est [2].

Or, par hypothèse, $x^{(k)}$ converge vers $\bar{x}$, $y_J^{(k)} = \Theta(x^{(k)})$ et $\bar{y} = \Theta(\bar{x})$.
Par conséquent, par continuité, $y_J^{(k)}$ converge vers $\bar{y}_J$ lorsque k tend vers ∞.

2) Montrons maintenant que $\lim_{k\to\infty} y_I^{(k)} = \bar{y}_I$.

Par définition de I, $\bar{x}_I = 0$. Nous en déduisons $\bar{y}_I = e_I$, et donc

$$\begin{aligned} \|y^{(k)} - e\|^2 &= \|y_I^{(k)} - e_I\|^2 + \|y_J^{(k)} - e_J\|^2 \\ &= \|y_I^{(k)} - \bar{y}_I\|^2 + \|y_J^{(k)} - e_J\|^2. \end{aligned}$$

Si nous passons à la limite supérieure, nous obtenons

$$\limsup_{k\to\infty} \|y_I^{(k)} - \bar{y}_I\|^2 = \limsup_{k\to\infty} \|y^{(k)} - e\|^2 - \|\bar{y}_J - e_J\|^2 \tag{4.1}$$

car $\lim_{k\to\infty} y_J^{(k)} = \bar{y}_J$.

Définissons la suite $(\hat{y}^{(k)})$ de la manière suivante :

$$\begin{aligned} &\hat{y}_I^{(k)} = e_I, \\ &\hat{y}_J^{(k)} = (X_J^{(k)})^{-1}(x_J^{(k)} - \bar{x}_J) + (X_J^{(k)})^{-1}\bar{X}_J\bar{y}_J. \end{aligned}$$

2. voir annexe B

Cette suite vérifie les deux propriétés ci-dessous

1) $\lim_{k\to\infty} \hat{y}^{(k)} = \bar{y}$ car $x_J^{(k)}$ converge vers $\bar{x}_J$.

2) $AX^{(k)}\hat{y}^{(k)} = 0$ car $\bar{y}$ est solution de $(\bar{P})$ et $x^{(k)},\ \bar{x} \in F$.

Comme $\hat{y}^{(k)}$ satisfait les contraintes du problème $(P^{(k)})$ et que $y^{(k)}$ est solution optimale de ce problème, nous obtenons par définition de $(\hat{y}^{(k)})$,

$$\|y^{(k)} - e\| \leq \|\hat{y}^{(k)} - e\| = \|\hat{y}_J^{(k)} - e_J\| \leq \|\hat{y}_J^{(k)} - \bar{y}_J\| + \|\bar{y}_J - e_J\|. \tag{4.2}$$

Il résulte des inégalités (4.1) et (4.2) que

$$\begin{aligned}
0 \quad &\leq \quad \liminf_{k\to\infty} \|y_I^{(k)} - \bar{y}_I\|^2 \leq \limsup_{k\to\infty} \|y_I^{(k)} - \bar{y}_I\|^2 \\
&= \quad \limsup_{k\to\infty} \|y^{(k)} - e\|^2 - \|\bar{y}_J - e_J\|^2 \\
&\leq \quad \limsup_{k\to\infty} \|\hat{y}_J^{(k)} - \bar{y}_J\|^2 + \|\bar{y}_J - e_J\|^2 \\
&\quad - \ 2\limsup_{k\to\infty} \|\hat{y}_J^{(k)} - \bar{y}_J\|\ \|\bar{y}_J - e_J\| - \|\bar{y}_J - e_J\|^2 \\
&= \quad 0
\end{aligned}$$

car la suite $(\hat{y}^{(k)})$ converge vers $\bar{y}$ quand k tend vers ∞.

Par conséquent,

$$\lim_{k\to\infty} y_I^{(k)} = \bar{y}_I.$$

Ceci termine la démonstration.

$\square$

Remarquons, par le choix de la longueur du pas λ_k, l'itéré $x^{(k+1)}$ reste dans F^0 lorsque $x^{(k)} \in F^0$. Pour suivre, nous allons montrer que les itérés vérifient une propriété encore plus stricte. C'est l'objet du lemme suivant.

Lemme 4.5

Il existe une constante c_3 strictement positive telle que

$$\inf_{k\geq 0} \min_{1\leq i\leq n} x_i^{(k)} \geq c_3.$$

<u>*Preuve*</u>

Nous allons démontrer ce lemme par récurrence.

Comme F^0 est non vide, nous pouvons choisir le point initiale $x^{(0)} \in F^{(0)}$.

Supposons que, pour tout $i \in \{1, 2, ..., n\}$,

$$x_i^{(k)} > 0$$

et montrons que c'est encore vrai pour l'itéré suivant.

Posons, pour simplifier l'écriture,

$$p^{(k)} = \frac{P_{AX^{(k)}} X^{(k)} \gamma_k}{\| P_{AX^{(k)}} X^{(k)} \gamma_k \|} \tag{4.3}$$

où $\gamma_k \in \partial f_\mu(x^{(k)})$.

Dans ce cas, la formule donnant $x^{(k+1)}$ devient

$$x^{(k+1)} = x^{(k)} - \lambda_k X^{(k)} p^{(k)}.$$

Or, par (3.1), $\partial f_\mu(x^{(k)}) = \partial f(x^{(k)}) - \mu(X^{(k)})^{-1} e$.

Par conséquent, si $\bar{\gamma}_k \in \partial f(x^{(k)})$, nous obtenons

$$p^{(k)} = \frac{P_{AX^{(k)}} \ (\ X^{(k)} \bar{\gamma}_k - \mu e\ )}{\| P_{AX^{(k)}} X^{(k)} \gamma_k \|}.$$

Supposons que $(x^{(k)})$ converge vers $\bar{x} = 0$. Dans ce cas,

$$P_{A\bar{X}} = I - \bar{X}A^T(A\bar{X}^2A^T)^{-1}A\bar{X} = I,$$

et pour tout $i \in \{1, 2, ..., n\}$,

$$p_i^{(k)} \longrightarrow \frac{-\mu}{\| P_{AX^{(k)}} X^{(k)} \gamma_k \|} e_i^T P_{AX^{(k)}} e$$

où e_i représente la i-ème colonne de la matrice identité.

Si nous appliquons le lemme 4.4, nous obtenons

$$\lim_{k \to \infty} P_{AX^{(k)}} e = e$$

et donc, pour tout $i \in \{1, 2, ..., n\}$ tel que $x_i^{(k)}$ converge vers 0, nous avons

$$p_i^{(k)} \longrightarrow \frac{-\mu}{\| P_{AX^{(k)}} X^{(k)} \gamma_k \|} e_i^T e = \frac{-\mu}{\| P_{AX^{(k)}} X^{(k)} \gamma_k \|}.$$

Or, comme l'ensemble admissible de (P_μ) est compact, il existe une constante c telle que, pour tout $i \in \{1, 2, ..., n\}$,

$$x_i^{(k)} \leq c$$

et donc

$$x_i^{(k+1)} = x_i^{(k)} - \lambda_k x_i^{(k)} p_i^{(k)} > x_i^{(k)}$$

car λ_k, $x_i^{(k)}$ et μ sont tous strictement positifs.

La thèse est alors immédiate.

$\square$

4.3 Convergence d'une sous-suite

Nous allons montrer qu'au moins une sous-suite de $(x^{(k)})$ converge vers la solution optimale du problème (P_μ).

Notons, pour tout $k \in I\!N$,

$$L_k = \{ \ x \in F^0 \mid f_\mu(x) \leq f_\mu(x^{(k)}) \ \}.$$

Pour $\bar{x} \in riL_K$, définissons les deux quantités $\delta^{(k)}$ et $\rho^{(k)}$ de la manière suivante

$$\delta^{(k)} = (p^{(k)})^T (X^{(k)})^{-1} (x^{(k)} - \bar{x}), \tag{4.4}$$

$$\rho^{(k)} = \| \ \bar{X}^{-1} (x^{(k)} - \bar{x}) \ \|. \tag{4.5}$$

Commençons par caractériser les points de riL_k en fonction de $\rho^{(k)}$.

Lemme 4.6

Si $\bar{x} \in riL_k$, alors

$$\| \ \bar{X}^{-1} X^{(k)} p^{(k)} \ \|^2 \leq (\ \rho^{(k)} + 1 \ )^2.$$

<u>*Preuve*</u>

Par définition de la norme Euclidienne et des matrices $\bar{X}$ et $X^{(k)}$, nous avons

$$\begin{aligned}
\| \ \bar{X}^{-1} X^{(k)} p^{(k)} \ \|^2 &= \sum_{i=1}^{n} \left(\frac{x_i^{(k)}}{\bar{x}_i} \ p_i^{(k)} \right)^2 \\
&\leq \left(\max_{1 \leq j \leq n} \frac{x_j^{(k)}}{\bar{x}_j} \right)^2 \sum_{i=1}^{n} \left(p_i^{(k)} \right)^2 \\
&= \| \ \bar{X}^{-1} x^{(k)} \ \|_\infty^2 \ \| \ p^{(k)} \ \|^2 \\
&= \| \ \bar{X}^{-1} x^{(k)} \ \|_\infty^2 \\
&= \| \ \bar{X}^{-1} (x^{(k)} - \bar{x}) + e \ \|_\infty^2 \\
&\leq (\ \| \ \bar{X}^{-1} (x^{(k)} - \bar{x}) \ \|_\infty + \| \ e \ \|_\infty \ )^2 \\
&\leq (\ \| \ \bar{X}^{-1} (x^{(k)} - \bar{x}) \ \| + 1 \ )^2
\end{aligned}$$

et, par définition de $\rho^{(k)}$, nous obtenons le résultat cherché.

□

Intéressons-nous maintenant aux propriétés de $\delta^{(k)}$ et de $\rho^{(k)}$.

Lemme 4.7

Il existe une constante c_4 telle que, pour tout $\bar{x} \in riL_k$,

$$\delta^{(k)} \geq c_4 \; [\; f_\mu(x^{(k)}) - f_\mu(\bar{x}) \;].$$

<u>*Preuve*</u>

Commençons par écrire $\delta^{(k)}$ plus explicitement.
Comme la matrice $X^{(k)}$ est diagonale, en utilisant (4.3.), nous obtenons

$$(p^{(k)})^T = \frac{\gamma_k^T X^{(k)} P_{AX^{(k)}}}{\| \; P_{AX^{(k)}} X^{(k)} \gamma_k \; \|} \tag{4.6}$$

Puisque $x^{(k)}$ et $\bar{x}$ sont admissibles,

$$AX^{(k)} \; [\; (X^{(k)})^{-1}(x^{(k)} - \bar{x}) \;] = A(x^{(k)} - \bar{x}) = 0$$

et donc, $(X^{(k)})^{-1}(x^{(k)} - \bar{x}) \in KerAX^{(k)}$.

Si nous introduisons (4.6) dans la définition de $\delta^{(k)}$, nous obtenons successivement

$$\begin{aligned}
\delta^{(k)} &= (p^{(k)})^T (X^{(k)})^{-1}(x^{(k)} - \bar{x}) \\
&= \frac{\gamma_k^T X^{(k)} P_{AX^{(k)}} (X^{(k)})^{-1}(x^{(k)} - \bar{x})}{\| \; P_{AX^{(k)}} X^{(k)} \gamma_k \; \|} \\
&= \frac{\gamma_k^T X^{(k)} (X^{(k)})^{-1}(x^{(k)} - \bar{x})}{\| \; P_{AX^{(k)}} X^{(k)} \gamma_k \; \|} \\
&\geq \frac{f_\mu(x^{(k)}) - f_\mu(\bar{x})}{\| \; P_{AX^{(k)}} X^{(k)} \gamma_k \; \|}
\end{aligned}$$

car $\gamma_k \in \partial f_\mu(x^{(k)})$.

En vertu du lemme 4.3, il suffit alors de poser

$$c_4 = \frac{1}{c_2}.$$

□

Lemme 4.8
S'il existe un $x^{(k)}$ *tel que* $\rho^{(k)} < 1$, *alors*

$$(\rho^{(k+1)})^2 \leq (\rho^{(k)})^2 - 2\ \lambda_k\ [\ \delta^{(k)} - (1 + \frac{1}{1-\rho^{(k)}})(\rho^{(k)})^2 - \frac{\lambda_k}{2}\ (1+\rho^{(k)})^2\ ].$$

<u>*Preuve*</u>

Par définitions de $\delta^{(k)}$ et $\rho^{(k)}$, nous avons successivement

$$\begin{aligned}
(\rho^{(k+1)})^2 &= \ \| \bar{X}^{-1}\ (x^{(k+1)} - \bar{x})\ \|^2 \\
&= \ \| \bar{X}^{-1}\ [\ x^{(k)} - \lambda_k X^{(k)} p^{(k)} - \bar{x}\ ]\ \|^2 \\
&= \ \| \bar{X}^{-1}(x^{(k)} - \bar{x})\ \|^2 + \lambda_k^2\ \| \bar{X}^{-1} X^{(k)} p^{(k)}\ \|^2 \\
&- \ 2\ \lambda_k (x^{(k)} - \bar{x})^T \bar{X}^{-2} X^{(k)} p^{(k)} \\
\\
&= \ (\rho^{(k)})^2 + \lambda_k^2\ \| \bar{X}^{-1} X^{(k)} p^{(k)}\ \|^2 - 2\ \lambda_k (x^{(k)} - \bar{x})^T \bar{X}^{-2} X^{(k)} p^{(k)} \\
&- \ 2\ \lambda_k \delta^{(k)} + 2\lambda_k (p^{(k)})^T (X^{(k)})^{-1} (x^{(k)} - \bar{x}) \\
\\
&= \ (\rho^{(k)})^2 - 2\ \lambda_k \delta^{(k)} + \lambda_k^2\ \| \bar{X}^{-1} X^{(k)} p^{(k)}\ \|^2 \\
&+ \ 2\ \lambda_k (x^{(k)} - \bar{x})^T\ [\ (X^{(k)})^{-1} - \bar{X}^{-2} X^{(k)}\ ]\ p^{(k)} \\
\\
&= \ (\rho^{(k)})^2 - 2\ \lambda_k \delta^{(k)} + \lambda_k^2\ \| \bar{X}^{-1} X^{(k)} p^{(k)}\ \|^2 \\
&+ \ 2\ \lambda_k (x^{(k)} - \bar{x})^T \bar{X}^{-1}\ [\ \bar{X}(X^{(k)})^{-1} - \bar{X}^{-1} X^{(k)}\ ]\ p^{(k)} \\
\\
&\leq \ (\rho^{(k)})^2 - 2\ \lambda_k \delta^{(k)} + \lambda_k^2\ \| \bar{X}^{-1} X^{(k)} p^{(k)}\ \|^2 \\
&+ \ 2\ \lambda_k\ [\ |(x^{(k)} - \bar{x})^T \bar{X}^{-1}\ (\ I - \bar{X}^{-1} X^{(k)}\ ) p^{(k)}| \\
&+ \ |(x^{(k)} - \bar{x})^T \bar{X}^{-1} (\bar{X}(X^{(k)})^{-1} - I) p^{(k)}|\ ] \\
\\
&= \ (\rho^{(k)})^2 - 2\ \lambda_k\ [\ \delta^{(k)} - \frac{\lambda_k}{2}\ \| \bar{X}^{-1} X^{(k)} p^{(k)}\ \|^2 \\
&- \ |(x^{(k)} - \bar{x})^T \bar{X}^{-1} (I - \bar{X}^{-1} X^{(k)}) p^{(k)}| \\
&- \ |(x^{(k)} - \bar{x})^T \bar{X}^{-1} (\bar{X}(X^{(k)})^{-1} - I) p^{(k)}|\ ].
\end{aligned} \tag{4.7}$$

Intéressons-nous plus particulièrement aux deux valeurs absolues contenues dans (4.7). Nous allons utiliser les définitions de normes Euclidienne et infinie pour majorer ces deux valeurs absolues.

Premièrement,

$$
\begin{aligned}
|(x^{(k)}-\bar{x})^T\bar{X}^{-1}(I-\bar{X}^{-1}X^{(k)})p^{(k)}| &\leq \| (I-\bar{X}^{-1}X^{(k)})\ \bar{X}^{-1}(x^{(k)}-\bar{x}) \|\ \| p^{(k)} \| \\
&= \| (I-\bar{X}^{-1}X^{(k)})\ \bar{X}^{-1}(x^{(k)}-\bar{x}) \| \\
&\leq \| (I-\bar{X}^{-1}X^{(k)})e \|_\infty \ \| \bar{X}^{-1}(x^{(k)}-\bar{x}) \| \\
&= \rho^{(k)} \ \| \bar{X}^{-1}x^{(k)}-e\|_\infty \\
&\leq \rho^{(k)} \ \| \bar{X}^{-1}x^{(k)}-e \| \\
&= \rho^{(k)} \ \| \bar{X}^{-1}x^{(k)}-\bar{X}^{-1}\bar{x} \| \\
&= \rho^{(k)} \ \| \bar{X}^{-1}(x^{(k)}-\bar{x}) \| \\
&= (\rho^{(k)})^2 \qquad (4.8)
\end{aligned}
$$

où nous avons utilisé (4.3) et les définitions de X et de $\rho^{(k)}$.

Deuxièmement,

$$
\begin{aligned}
|(x^{(k)}-\bar{x})^T\bar{X}^{-1}(\bar{X}(X^{(k)})^{-1}-I)p^{(k)}| &\leq \| (\bar{X}(X^{(k)})^{-1}-I)\ \bar{X}^{-1}(x^{(k)}-\bar{x}) \|\ \| p^{(k)} \| \\
&= \| (\bar{X}(X^{(k)})^{-1}-I)\ \bar{X}^{-1}(x^{(k)}-\bar{x}) \| \\
&\leq \| (\bar{X}(X^{(k)})^{-1}-I)\ e \|_\infty \ \| \bar{X}^{-1}(x^{(k)}-\bar{x}) \| \\
&= \rho^{(k)} \ \| (X^{(k)})^{-1}\bar{x}-e \|_\infty \\
&\leq \frac{(\rho^{(k)})^2}{1-\rho^{(k)}}. \qquad (4.9)
\end{aligned}
$$

En introduisant (4.8), (4.9) et le lemme 4.6 dans l'inégalité (4.7), nous obtenons le résultat annoncé.

□

Comme nous l'avons déjà annoncé, nous allons maintenant montrer qu'au moins une sous-suite de $(x^{(k)})$ converge vers x^*.

Théorème 4.1
Si $(x^{(k)})$ est la suite engendrée par l'algorithme et x^* la solution optimale du problème (P_μ), alors

$$\liminf_{k\to\infty} x^{(k)} = x^*.$$

<u>*Preuve*</u>

Supposons, par l'absurde, que x^* n'est pas une valeur d'adhérence de $(x^{(k)})$.

Nous allons montrer que pour un $\bar{x} \in riL_k$ bien choisi, la quantité $\rho^{(k)}$ est strictement négative, ce qui est impossible puisque c'est une norme.

Vu que x^* est l'unique point qui minimise f_μ sur F^0, il existe $\tilde{f}$ tel que, pour tout $x \in F^0$,

$$f_\mu(x^*) < \tilde{f} < f_\mu(x). \tag{4.10}$$

Soit $\tilde{x}$ une valeur d'adhérence de la suite $(x^{(k)})$, c'est-à-dire

$$\exists\ K \subset I\!N \quad \text{t.q.} \quad x^{(k)} \longrightarrow \tilde{x} \quad \text{pour } k \in K.$$

Supposons que $\tilde{x}$ minimise f_μ sur l'adhérence de la suite $(x^{(k)})$.
Alors, pour tout $\alpha \in]0, 1[$, il existe k_α tel que

$$f_\mu(x^{(k)}) \geq (1-\alpha) f_\mu(\tilde{x}) + \alpha \tilde{f} \qquad \text{pour tout } k \geq k_\alpha. \tag{4.11}$$

Posons $\bar{x}_\alpha = (1-\alpha)\tilde{x} + \alpha x^*$ où $\alpha \in]0,1[$.

Le point $\bar{x}_\alpha \in riL_k$ car $\tilde{x}$ minimise $(x^{(k)})$ sur son adhérence.

Par analogie avec (4.4) et (4.5), posons

$$\delta_\alpha^{(k)} = (p^{(k)})^T (X^{(k)})^{-1} (x^{(k)} - \bar{x}_\alpha),$$
$$\rho_\alpha^{(k)} = \parallel \bar{X}^{-1}(x^{(k)} - \bar{x}_\alpha) \parallel.$$

En utilisant le lemme 4.7, nous obtenons, pour tout $k \in I\!N$,

$$\begin{aligned} \delta_\alpha^{(k)} &\geq c_4 \; [\; f_\mu(x^{(k)}) - f_\mu(\bar{x}_\alpha) \;] \\ &= c_4 \; [\; f_\mu(x^{(k)}) - f_\mu((1-\alpha)\tilde{x} + \alpha x^*) \;] \\ &\geq c_4 \; [\; f_\mu(x^{(k)}) - (1-\alpha) f_\mu(\tilde{x}) - \alpha f_\mu(x^*) \;] \end{aligned}$$

car f_μ est convexe.

Et donc, en utilisant (4.11), nous obtenons, pour tout $k \geq k_\alpha$,

$$\begin{aligned} \delta_\alpha^{(k)} &\geq c_4 \; [\; (1-\alpha) f_\mu(\tilde{x}) + \alpha \tilde{f} - (1-\alpha) f_\mu(\tilde{x}) - \alpha f_\mu(x^*) \;] \\ &= c_4 \alpha \; (\tilde{f} - f_\mu(x^*)). \end{aligned} \tag{4.12}$$

Supposons que $\rho_\alpha^{(k)}$ vérifie les deux propriétés suivantes :

1) Il existe $K(\alpha) \subset K$ tel que, pour tout $k \in K(\alpha)$,

$$\rho_\alpha^{(k)} = O(\alpha). \tag{4.13}$$

2) Il existe $\alpha_0 > 0$ tel que, pour tout $k \in K(\alpha_0)$,

$$(\; 1 + \frac{1}{1 - \rho_{\alpha_0}^{(k)}} \;) \; (\rho_{\alpha_0}^{(k)})^2 < \min\{\; 1, \frac{1}{3} c_4 \alpha_0 (\tilde{f} - f_\mu(x^*)) \;\}. \tag{4.14}$$

Notons par k_0 le rang correspondant à α_0 dans (4.11), ce qui signifie que, pour tout $k \geq k_0$,

$$f_\mu(x^{(k)}) \geq (1 - \alpha_0) f_\mu(\tilde{x}) + \alpha_0 \tilde{f}. \tag{4.15}$$

Comme la suite (λ_k) converge vers 0, il existe $k_1 \in K(\alpha_0)$, $k_1 \geq k_0$ tel que, pour tout $k \geq k_1$,

$$2\lambda_k < \frac{1}{3} c_4 \alpha_0 (\tilde{f} - f_\mu(x^*)) \tag{4.16}$$

puisque, par (4.10), le membre de droite est stictement positif.

De plus, pour tout $k \geq k_1$, la relation suivante est vérifiée :

$$(\rho_{\alpha_0}^{(k+1)})^2 \leq \frac{-2}{3} \; c_4 \; \alpha_0 \; (\tilde{f} - f_\mu(x^*)) + (\rho_{\alpha_0}^{(k)})^2.$$

En effet, par (4.13), nous obtenons, pour tout $k \geq k_1$,

$$\rho_{\alpha_0}^{(k)} < 1. \tag{4.17}$$

Les hypothèses du lemme 4.8 étant satisfaites, nous avons

$$\begin{aligned}
(\rho_{\alpha_0}^{(k+1)})^2 &\leq (\rho_{\alpha_0}^{(k)})^2 - 2\lambda_k \ [\ \delta_{\alpha_0}^{(k)} - (1 + \frac{1}{1-\rho_{\alpha_0}^{(k)}})(\rho_{\alpha_0}^{(k)})^2 - \frac{\lambda_k}{2}\ (1 + \rho_{\alpha_0}^{(k)})^2\ ] \\
&\leq (\rho_{\alpha_0}^{(k)})^2 - 2\lambda_k \ \{\ c_4\alpha_0(\tilde{f} - f_\mu(x^*)) - (1 + \frac{1}{1-\rho_{\alpha_0}^{(k)}})(\rho_{\alpha_0}^{(k)})^2 - 2\ \lambda_k\ \} \\
&\leq (\rho_{\alpha_0}^{(k)})^2 - 2\lambda_k \ \{\ c_4\alpha_0(\tilde{f} - f_\mu(x^*)) - (1 + \frac{1}{1-\rho_{\alpha_0}^{(k)}})(\rho_{\alpha_0}^{(k)})^2 - \frac{1}{3}\ c_4\alpha_0(\tilde{f} - f_\mu(x^*))\ \} \\
&\leq (\rho_{\alpha_0}^{(k)})^2\ \frac{-2}{3}\ \lambda_k c_4\alpha_0(\tilde{f} - f_\mu(x^*)),
\end{aligned} \tag{4.18}$$

où nous avons appliqué successivement les résultats (4.12), (4.17), (4.16) et (4.14).

Or, $\sum_{k=1}^{\infty}\lambda_k = \infty$ par hypothèse et $(\tilde{f} - f_\mu(x^*)) > 0$ par (4.10).

Finalement, le membre de droite de (4.18) est majoré par $-\infty$, ce qui est bien entendu impossible.

Pour que la démonstration soit complète, il faut encore montrer que les relations (4.13) et (4.14) sont bien vérifiées.

1) Par définition de $\bar{x}_\alpha$ et de $\rho_\alpha^{(k)}$, nous obtenons

$$\begin{aligned}
\rho_\alpha^{(k)} &= \|\ \bar{X}_\alpha^{-1}(x^{(k)} - \bar{x}_\alpha)\ \| \\
&\leq \|\ \bar{X}_\alpha^{-1}(x^{(k)} - \tilde{x})\ \| + \|\ \bar{X}_\alpha^{-1}(\tilde{x} - \bar{x}_\alpha)\ \| \\
&= \|\ \bar{X}_\alpha^{-1}(x^{(k)} - \tilde{x})\ \| + \alpha\|\ \bar{X}_\alpha^{-1}(\tilde{x} - x^*)\ \|.
\end{aligned}$$

Par la définition de la norme Euclidienne, nous avons

$$\|\ \bar{X}_\alpha^{-1}(\tilde{x} - x^*)\ \| = \sqrt{\sum_{i=1}^{n}\left(\frac{\tilde{x}_i - x_i^*}{(1-\alpha)\tilde{x}_i + \alpha x_i^*}\right)^2}.$$

Comme $\tilde{x}$ et x^* sont des points contenus dans le borné F, il existe une constante $c \geq 0$ telle que

$$\|\ \bar{X}_\alpha^{-1}(\tilde{x} - x^*)\ \| \leq \frac{c}{2}.$$

4.4 Convergence de toute la suite

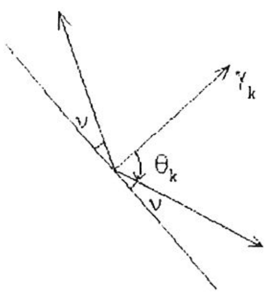
Nous allons maintenant montrer que, moyennant une condition supplémentaire, appelée condition d'angle, toute la suite $(x^{(k)})$ converge vers la solution optimale du problème (P_μ).

4.4.1 Condition d'angle

Soit (γ_k) avec $\gamma_k \in \partial f_\mu(x^{(k)})$, la suite des sous-gradients engendré.
Nous dirons que la *condition d'angle* est vérifiée si l'angle θ_k formé par les vecteurs γ_k et $(x^{(k)} - x^*)$ est uniformément borné par $\frac{\pi}{2}$, c'est-à-dire s'il existe $\nu > 0$ tel que, pour tout k,

$$|\theta_k| < \frac{\pi}{2} - \nu.$$

La condition d'angle peut s'interpréter géométriquement de la manière suivante:



$|\theta_k| < \frac{\pi}{2} - \nu.$

$\cos\ \theta_k \geq c > 0.$

De façon équivalente, la condition d'angle est satisfaite s'il existe une constante c strictement positive telle, pour tout k

$$c \leq \frac{(x^{(k)} - x^*)^T \gamma_k}{\| x^{(k)} - x^* \| \| \gamma_k \|} = \cos\ \theta_k$$

où $\gamma_k \in \partial f_\mu(x^{(k)})$ et θ_k est l'angle formé par les vecteurs γ_k et $(x^{(k)} - x^*)$.

Bibliographie

[1] J.B.G. Frenk, J.F. Sturm and S. Zhang, *An interior point subgradient method for linearly constrained non differentiable programming*, Report 9612/A, Erasmus University Rotterdam, 1996.

[2] J.L. Goffin, *On convergence rates of subradient optimization methods*, Mathematical Programming, 13, 329-347.

[3] C.C. Gonzaga and L.A. Carlos, *A primal affine-scaling algorithm for linearly constrained convex programs*, Technical Report ES-238/90, Departement of Systems Engineering and Computer Science, COPPE Federal University of Rio de Janeiro, Brazil, 1990.

[4] C.C. Gonzaga and R.A. Tapia, *On the convergence of the Mizuno-Todd-Ye algorithm to the analytic center of the solution set*, Technical Report 92-36, Departement of Mathematical Sciences, Rice University, Houston, TX, USA, 1992.

[5] N. Karmarkar, *A new polynomial-time algorithm for linear programming*, Combinatorica 4 (1984) 373-395.

[6] M. Minoux, *Mathematical Programming, Theory and Algorithms*, John Wiley and Sons, 1986.

[7] G.W. Stewart, *On scaled projections and pseudoinverses*, Linear Algebra and Its Applications 112 (1989) 189-193.

[8] J. Stoer and C. Witzgall, *Convexity and optimization in finite dimensions*, I. Springer Verlag, Berlin, 1970.