

A. EL GHALI

**Méthode du sous-gradient réduit généralisé comme
extension du GRG d'Abadie au cas non différentiable**

*Revue française d'automatique, d'informatique et de recherche
opérationnelle. Recherche opérationnelle*, tome 26, n° 3 (1992),
p. 237-267.

http://www.numdam.org/item?id=RO_1992__26_3_237_0

© AFCET, 1992, tous droits réservés.

L'accès aux archives de la revue « Revue française d'automatique, d'informatique et de recherche opérationnelle. Recherche opérationnelle » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/legal.php>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

MÉTHODE DU SOUS-GRADIENT RÉDUIT GÉNÉRALISÉ COMME EXTENSION DU GRG D'ABADIE AU CAS NON DIFFÉRENTIABLE (*)

par A. EL GHALI ⁽¹⁾

Communiqué par Y. SMEERS

Résumé. — Depuis l'apparition de la méthode du gradient réduit introduite par Wolfe, beaucoup d'auteurs ont étudié la convergence de cette méthode ou des méthodes modifiées. Comme le cas de Luenberger, elle a pour objet de résoudre les programmes non linéaires avec contraintes d'égalité linéaires et des contraintes de non-négativité des variables, et utilise le principe de décomposition des variables en variables de base et hors base et le calcul de pivot comme extension naturelle de la méthode simpliciale.

Cette méthode a connu une généralisation au cas des contraintes d'égalité non linéaires en appliquant le théorème des fonctions implicites qu'on appellera méthode du gradient réduit généralisé : GRG. Cette généralisation est due à Abadie et Carpentier.

L'objet de notre article est d'étudier l'adaptation du GRG au cas non différentiable en utilisant les techniques des faisceaux introduites par Lemaréchal et utilisées dans le cas de contraintes linéaires par Bihain, Nguyen et Strodriot.

Mots clés : Programmation non différentiable; problème à contraintes non linéaires; algorithme de sous-gradient réduit généralisé; techniques des faisceaux.

Abstract. — Since the appearing of the reduced gradient method given by Wolfe, many authors have tried to obtain the convergence of this method or of a modified form if it (see e.g. Luenberger).

This method is designed to solve a non-linear problem subject to linear equality and bound constraints, use the principle of the variables decomposition into basic and non-basic variables and the computation of the pivot as an extension of the simplex method.

It has been generalized to non-linear equality constraints by application of the implicit function theorem, this generalization is called the generalized reduced gradient method: GRG due to Abadie and Carpentier.

In this paper, we study the adaptation of the GRG to the non-differentiable case by using bundle techniques introduced by Lemaréchal and applied to linear constraints by Bihain, Nguyen and Strodriot.

Keywords : Non-differentiable programming; non-linearly constrained minimization; generalized reduced subgradient algorithm; bundle techniques.

(*) Reçu en mars 1991.

⁽¹⁾ Département de Mathématiques, Facultés Universitaires de Namur, Namur, Belgique, et Département de Mathématique, Faculté des Sciences, Université My Ismail, Meknes, Maroc.

1. INTRODUCTION

L'objet de cet article est de résoudre les programmes non linéaires à fonction objectif non différentiable avec des contraintes d'égalité différentiables non linéaires et des contraintes de non négativité des variables.

Toutes les méthodes de sous-gradient utilisent le principe de perturbation des sous-différentiels et des dérivées directionnelles pour garantir la convergence à des solutions ε -optimales comme le cas de Lemaréchal [12], Wolfe [26] et Bihain, Nguyen et Strodiot [2].

Aussi les techniques des faisceaux qui sont des approximations de l' ε -sous-différentiel s'imposent afin que ces méthodes deviennent implémentables.

Dans le cas de contraintes linéaires et à fonction objectif de classe C^1 , Wolfe [23] a introduit la méthode du gradient réduit. Comme cette méthode ne converge pas [24], beaucoup d'auteurs ont proposé des modifications algorithmiques de celle-ci. On cite par exemple Huard [9], Zangwill [27], Luenberger [15] et Mokhtar-Kharroubi [20].

Dans le cas où la fonction économique n'est plus différentiable, on peut citer l'algorithme de sous-gradient réduit de Bihain, Nguyen et Strodiot [2].

La méthode de Wolfe a été généralisée au cas de contraintes d'égalité non linéaires par Abadie et Carpentier [1] dans la méthode appelée gradient réduit généralisé: GRG, tout en appliquant la décomposition des variables en variables de base et des variables hors base, en résolvant le système d'équations non linéaires par l'application du théorème des fonctions implicites et en éliminant par conséquent l'ensemble des variables de base.

Une fois que la direction réalisable réduite pour les variables hors base est obtenue, on la complète par linéarisation du système non linéaire.

Cette généralisation dans le cas différentiable connaît le problème de convergence. Ainsi, Smeers [21] et Mokhtar [19] ont étudié sa convergence en la modifiant sous l'hypothèse de non dégénérescence et sous la propriété de base introduite par Smeers dans la procédure de calcul de nouvelles bases non dégénérées.

Dans ce papier on utilise la variante de Luenberger [15] du gradient réduit adaptée au cas non différentiable mais aussi à contraintes non linéaires par généralisation du GRG d'Abadie. On applique le théorème des fonctions implicites pour éliminer les variables de base et on obtient localement un programme équivalent réduit en variables et en contraintes.

Pour la recherche d'une direction réalisable réduite en variables hors base, tout d'abord l' ε -sous-différentiel est approché par le faisceau ne contenant

qu'un nombre fini de sous-gradients. Chaque vecteur de ce faisceau est réduit de la même manière que dans le GRG puis modifié par la transformation proposée par Luenberger et à la fin, on résout un sous-problème quadratique approprié.

La recherche linéaire économique s'effectue le long de la direction réduite pour les variables hors base.

Une fois que l'itéré suivant est obtenu, on utilise la procédure algorithmique de Smeers [21] pour calculer la nouvelle base.

Notre étude se base sur la caractérisation des solutions ε -optimales pour des problèmes convexes non différentiables démontrée par Strodiot, Nguyen et Heukemes [22].

Dans la Section 2, on étudie la résolution du système non linéaire par l'application du théorème des fonctions implicites et on obtient un problème réduit.

Dans la Section 3, c'est la recherche de la direction réalisable qui est entamée en résolvant un sous-problème approprié. La Section 4 sera consacrée à la recherche linéaire de type économique, le calcul de la fonction implicite avec précision des contraintes de tolérance de ce calcul et on rappelle la propriété fondamentale dans la procédure de calcul de la nouvelle base.

Dans la Section 5, c'est l'algorithme qui est énoncé, sa convergence est étudiée en Section 6.

Notations générales

\mathbb{R}^n désigne l'espace euclidien usuel muni du produit scalaire $\langle x, y \rangle = \sum x_i y_i$ et la norme associée $\|x\|^2 = \langle x, x \rangle$.

∂f désigne le sous-différentiel de $f: \mathbb{R}^n \rightarrow \mathbb{R}$ où f est une fonction convexe.

Soient $Q = \{1, \dots, n\}$ et $L \subset Q$.

Soit g un sous-gradient appartenant à $\partial f(x)$, g_L désigne le vecteur ayant pour composantes celles de g indicées dans L .

On note aussi l' ε -sous-différentiel de f en x par $\partial_\varepsilon f(x)$. C'est un ensemble convexe compact de \mathbb{R}^n . De plus, pour tout ensemble borné S de \mathbb{R}^n , $\partial_\varepsilon f(S)$ reste borné

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

Dans le cas $\varepsilon = 0$, on retrouve le sous-différentiel ordinaire $\partial f(x)$.

Rappelons la définition de l' ε -dérivée directionnelle donnée par

$$f'_\varepsilon(x, d) = \inf_{t > 0} \frac{f(x + td) - f(x) + \varepsilon}{t}$$

liée à l' ε -sous-différentiel par la formule suivante :

$$f'_\varepsilon(x, d) = \max_{g \in \partial_\varepsilon f(x)} \langle d, g \rangle.$$

Notons qu'une généralisation de la notion de sous-différentiel a été donnée par Clarke dans le cas localement lipschitzien appelée *gradient généralisé* et définie par

$$\partial f(x) = Co \{ g : g = \lim_{y \rightarrow x} \nabla f(y), \nabla f(y) \text{ existe} \}$$

où Co désigne l'enveloppe convexe. C'est un ensemble convexe compact de \mathbb{R}^n .

Aussi, la dérivée directionnelle généralisée de f dans la direction d est définie par

$$\begin{aligned} f^0(x, d) &= \lim_{\substack{h \rightarrow 0 \\ t \rightarrow 0 \\ >}} \sup \frac{f(x + td + h) - f(x + h)}{t} \\ &= \max_{g \in \partial f(x)} \langle d, g \rangle. \end{aligned}$$

$f^0(x, \cdot)$ est une fonction convexe positivement homogène.

Dans le cas f convexe, le gradient généralisé de f et le sous-différentiel coïncident.

Remarquons aussi, lorsque f est localement lipschitzienne et que la dérivée directionnelle existe, on a

$$f'(x, d) \leq f^0(x, d).$$

Le programme mathématique considéré dans ce travail est de la forme :

$$\begin{aligned} &\min f(x), \\ &h_i(x) = 0, \quad i = 1, \dots, m, \\ &x \geq 0, \end{aligned} \tag{P}$$

où f est une fonction convexe définie sur \mathbb{R}^n , non nécessairement différentiable, h_i , pour tout i sont des fonctions numériques continûment différentiables sur \mathbb{R}^n .

On note D l'ensemble des solutions réalisables de (P).

2. RECHERCHE DE LA DIRECTION

2.1. Le problème réduit

Soient $x \in D$ et $\nabla h(x)$ la matrice jacobienne de la fonction h en x , ainsi $\nabla h(x) = [\nabla h_1(x), \dots, \nabla h_m(x)]^t$.

On fera l'hypothèse de non-dégénérescence (H) suivante :

$$\forall x \in D, \quad \text{si } M(x) = \{j \mid x_j > 0\} \text{ on a : } \left. \begin{array}{l} \text{rang } \nabla h(x)^{M(x)} = m; \end{array} \right\} \quad (\text{H})$$

par suite, il est possible de représenter le système $h(x) = 0$ autrement. En effet, le théorème des fonctions implicites nous permet d'exprimer au voisinage de x , m variables en fonction des $n - m$ restantes et d'extraire de la matrice $\nabla h(x)$ une sous-matrice inversible d'ordre $(m \times m)$ que l'on appellera la matrice de base.

Par suite, il existe une base $I: I \subset Q$ et $|I| = m$ telle que :

$$x = [x_I, x_{\bar{I}}] \quad (2.1)$$

où $\bar{I} = Q - I$,

$$\nabla h(x) = [\nabla_I h(x), \nabla_{\bar{I}} h(x)] \quad (2.2)$$

où $\nabla_I h(x)$ est une matrice carrée d'ordre m non singulière, et

$$x_I > 0. \quad (2.3)$$

On note la matrice de base $\nabla_I h(x)$ par $B(x)$ et son complémentaire dans $\nabla h(x)$ par $N(x)$.

x_I représente le m -vecteur dont les indices sont exactement les indices des vecteurs colonnes de la matrice jacobienne formant $B(x)$, il est appelé vecteur de base.

$x_{\bar{I}}$ est le $(n - m)$ vecteur correspondant à la sous-matrice $N(x)$ appelé vecteur hors base.

Pour simplifier, on écrit comme variables de base les m premières variables, mais ceci n'est pas chaque fois le cas.

$B(y)$ signifie aussi que la matrice de base est calculée en (y, I) , c'est-à-dire que

$$B(y) = \nabla_I h(y).$$

Pour le système $h(x) = 0$, le théorème des fonctions implicites implique l'existence de deux voisinages $V(x_{\bar{I}}, r)$ et $V(x_I, c)$ et une fonction φ continûment différentiable telle que :

$$\left. \begin{array}{l} \varphi: V(x_{\bar{I}}, r) \rightarrow V(x_I, C), \\ y \rightarrow \varphi(y) = z \Leftrightarrow h(\varphi(y), y) = 0, \end{array} \right\} \quad (2.4)$$

$$\nabla \varphi(y) = -B^{-1}(\varphi(y), y) N(\varphi(y), y). \quad (2.5)$$

L'existence et la construction de la fonction φ seront analysées dans la section suivante.

Au voisinage de x , on définit aussi la nouvelle fonction \hat{f} par :

$$\hat{f}(u) = f[\varphi(u), u], \quad \forall u \in V(x_{\bar{I}}, r)$$

et le problème (P) peut alors s'écrire localement comme :

$$\begin{array}{ll} \min \hat{f}(u), \\ u \in \mathbb{R}^{n-m}, \\ \text{sous les contraintes} \\ u \geq 0, \\ \varphi(u) \geq 0. \end{array} \quad (\hat{P})$$

Comme $\varphi(x_{\bar{I}}) = x_I > 0$ et que φ est une fonction continue, alors φ sera non négative au voisinage de $x_{\bar{I}}$ et la contrainte $\varphi(u) \geq 0$ sera vérifiée dans le voisinage.

Localement donc, le problème (P) sera équivalent au problème réduit en variables et en contraintes (P_R) :

$$\left. \begin{array}{l} \min \hat{f}(u), \\ u \geq 0. \end{array} \right\} \quad (P_R)$$

2.2. Calcul d'une direction réalisable en $x_{\bar{I}}$

Soit un sous-gradient $g \in \partial f(x)$, on définit le sous-gradient réduit \hat{g} en $x_{\bar{I}}$ par

$$\hat{g} = g_{\bar{I}} - N^t (B^{-1})^t g_I \quad (2.6)$$

où B et N sont les matrices de base et hors-base respectivement calculées au point x .

Il s'agit d'une adaptation de la définition du gradient réduit généralisé (GRG) d'Abadie [1] au cas non différentiable et trouve sa justification, d'après le théorème de composition de sous gradients, dans le fait que

$$\partial \hat{f}(x_{\bar{I}}) \subset Co \{ [-N^t (B^{-1})^t, Id] g : g \in \partial f(x) \} \quad (2.7)$$

où Id est la matrice identité et Co est l'enveloppe convexe. Par suite, pour tout sous-gradient, on définit tout d'abord à partir de la définition de Luenberger donnée dans le cas différentiable [15] la transformation suivante, notée $L(x_{\bar{I}}, \hat{g})$:

$$\forall j \in \bar{I}, \quad L_j = \begin{cases} \hat{g}_j & \text{si } \hat{g}_j \leq 0, \\ x_j \hat{g}_j & \text{sinon.} \end{cases} \quad (2.8)$$

THÉORÈME 1 D'OPTIMALITÉ: *S'il existe, en x réalisable, un sous-gradient $g \in \partial f(x)$ tel que :*

$$L(x_{\bar{I}}, \hat{g}) = 0,$$

alors x est un point stationnaire de (P) .

Preuve: On démontre que x vérifie les conditions nécessaires de type Kuhn-Tucker [22], ce qui revient à prouver l'existence des constantes μ_i , $i = 1, \dots, m$ et $u \geq 0$, $u \in \mathbb{R}^n$ telles que

$$g + \sum \mu_i \nabla h_i(x) - u = 0 \quad \text{et} \quad \langle u, x \rangle = 0$$

si et seulement si

$$g_I + \mu B - u_I = 0, \quad (2.9)$$

$$g_{\bar{I}} + \mu N - u_{\bar{I}} = 0, \quad (2.10)$$

$$\langle u_I, x_I \rangle = 0 \quad \text{et} \quad \langle u_{\bar{I}}, x_{\bar{I}} \rangle = 0; \quad (2.11)$$

En choisissant $u_I = 0$, (2.9) devient

$$\mu = -(B^{-1})^t g_I$$

et (2.10) devient

$$\begin{aligned} u_{\bar{I}} &= g_{\bar{I}} - N^t (B^{-1})^t g_I \\ &= \hat{g}. \end{aligned}$$

Par hypothèse, on a $L(x_{\bar{I}}, \hat{g}) = 0$ d'où

$$u_{\bar{I}} = \hat{g} \geq 0.$$

De plus, pour tout indice $j \in \bar{I}$ tel que $\hat{g}_j > 0$, on a $x_j = 0$, c'est-à-dire

$$0 = \langle x_{\bar{I}}, \hat{g} \rangle = \langle x_{\bar{I}}, u_{\bar{I}} \rangle,$$

ce qui démontre que x est un point stationnaire de (P). ■

2.3. Sélection d'une direction réalisable et de descente

2.3.1. Cas du problème de type réduit

Notre problème de type réduit (P_R) est de type

$$\left. \begin{array}{l} \min k(x) \\ x \geq 0, \end{array} \right\} \quad (Q)$$

où k est une fonction supposée convexe ou même localement lipschitzienne.

En x réalisable pour (Q), on se donne un sous-gradient $\delta \in \partial k(x)$ appelé *gradient généralisé* au sens de Clarke si k est localement lipschitzienne et nous définissons la direction réalisable S comme suit :

$$S = -L(x, \delta)$$

telle que

$$L_j = \begin{cases} \delta_j & \text{si } \delta_j \leq 0, \\ x_j \delta_j & \text{sinon.} \end{cases}$$

De même que dans le théorème 1 d'optimalité, s'il existe un sous-gradient δ tel que

$$L(x, \delta) = 0 \Leftrightarrow \begin{cases} \delta \geq 0, \\ \langle x, \delta \rangle = 0, \end{cases}$$

alors x est une solution optimale de (Q) si k est convexe ou stationnaire si k est localement lipschitzienne. Dans le cas contraire, on résout le sous-problème quadratique associé à x :

$$\left. \begin{array}{l} \min \frac{1}{2} \|L(x, \delta)\|^2, \\ \delta \in \partial k(x). \end{array} \right\} \quad (D_x)$$

La fonction économique de ce sous-problème que l'on note par

$$\theta(\delta) = \frac{1}{2} \|L(x, \delta)\|^2$$

admet des dérivées partielles continues sur le sous-différentiel $\partial k(x)$:

$$\nabla_j \theta(\delta) = \begin{cases} \delta_j & \text{si } \delta_j \leq 0, \\ x_j^2 \delta_j & \text{sinon.} \end{cases} \quad (2.12)$$

Soit δ^* une solution optimale de (D_x) , alors δ^* vérifie l'inégalité variationnelle suivante :

$$\langle \nabla \theta(\delta^*), \delta - \delta^* \rangle \geq 0, \quad \forall \delta \in \partial k(x)$$

ou encore

$$\langle \nabla \theta(\delta^*), \delta \rangle \geq \langle \nabla \theta(\delta^*), \delta^* \rangle, \quad \forall \delta \in \partial k(x).$$

De (2.8) et (2.12), on remarque que

$$\langle \nabla \theta(\delta^*), \delta^* \rangle = \|L(x, \delta^*)\|^2 > 0;$$

Par conséquent,

$$\langle \nabla \theta(\delta^*), \delta \rangle \geq \|L(x, \delta^*)\|^2 > 0, \quad \forall \delta \in \partial k(x).$$

Si on prend la direction réalisable $d = -\nabla \theta(\delta^*)$, on aura alors

$$\langle d, \delta \rangle \leq -\|L(x, \delta^*)\|^2 < 0, \quad \forall \delta \in \partial k(x).$$

Dès lors, la dérivée directionnelle, quand elle existe, vérifie pour le cas lipschitzien :

$$k'(x, d) \leq -\|L(x, \delta^*)\|^2 < 0$$

ce qui veut dire que d est une direction réalisable et de descente pour le problème (Q) en x .

2.3.2. Cas du problème (P)

La recherche d'une direction de descente nécessite la résolution d'un sous-problème de type (D_x) avec le sous-différentiel tout entier, alors que, analytiquement et numériquement, il est difficile de définir une expression de cet ensemble. Pour cette raison, on préfère utiliser les techniques des faisceaux introduites par Lemaréchal [11], ce qui, en d'autres termes, veut dire le calcul d'une approximation du sous-différentiel.

Pour cela, si on dispose de sous-gradients g^1, \dots, g^k de f en x^1, \dots, x^k respectivement, on définit α_i^k par :

$$\alpha_i^k = \alpha(x^k, x^i, g^i) = f(x^k) - f(x^i) - \langle g^i, x^k - x^i \rangle, \quad \forall i = 1, \dots, k \quad (2.13)$$

où $g^i \in \partial f(x^i)$.

α_i^k représente l'erreur faite lorsque la fonction convexe est linéarisée en x^i avec le sous-gradient g^i .

Alors, pour $\varepsilon \geq 0$, une approximation de $\partial_\varepsilon f(x^k)$ est l'ensemble $G_\varepsilon(x^k)$ défini par

$$G_\varepsilon(x^k) = \left\{ \sum_{i=1}^k \lambda_i g^i : \sum_{i=1}^k \lambda_i = 1, \lambda_i \geq 0, i = 1, \dots, k \text{ et } \sum_{i=1}^k \lambda_i \alpha_i^k \leq \varepsilon \right\}. \quad (2.14)$$

Ceci est justifié par le fait que

$$G_\varepsilon(x^k) \subseteq \partial_\varepsilon f(x^k). \quad (2.15)$$

Pour la recherche d'une direction réalisable réduite en x_F^k , on résout un sous-problème similaire à (D_x) en remplaçant le sous-différentiel par une approximation de type (2.14) et les sous-gradients par les sous-gradients réduits \hat{g}^i définis dans (2.6). Ceci nous amène à résoudre le sous-problème

quadratique suivant :

$$\left. \begin{array}{l} \min \frac{1}{2} \left\| L(x_{\bar{T}}^k, \sum_{i=1}^k \lambda_i \hat{g}^i) \right\|^2, \\ \text{sous les contraintes} \\ \sum_{i=1}^k \lambda_i = 1, \\ \lambda_i \geq 0, \quad i = 1, \dots, k, \\ \sum_{i=1}^k \lambda_i \alpha_i^k \leq \varepsilon. \end{array} \right\} \quad (D^k)$$

Soit u_k le multiplicateur de Lagrange correspondant à la contrainte additionnelle $\sum_{i=1}^k \lambda_i \alpha_i^k \leq \varepsilon$ dans (D^k) , alors le sous-problème (D^k) est équivalent au sous-problème suivant :

$$\left. \begin{array}{l} \min \frac{1}{2} \left\| L(x_{\bar{T}}^k, \sum_{i=1}^k \lambda_i \hat{g}^i) \right\|^2 + u_k \sum_{i=1}^k \lambda_i \alpha_i^k, \\ \text{sous les contraintes} \\ \sum_{i=1}^k \lambda_i = 1, \\ \lambda_i \geq 0, \quad i = 1, \dots, k. \end{array} \right\} \quad (\bar{D}^k)$$

Ainsi, pour λ^* , une solution optimale de (D^k) , on pose

$$\hat{p} = \sum_{i=1}^k \lambda_i^* \hat{g}^i$$

et on définit tout d'abord le gradient de la fonction $\theta(g) = 1/2 \|L(x_{\bar{T}}^k, g)\|^2$ en \hat{p} par :

$$j \in \bar{I}, \quad \nabla_j \theta(\hat{p}) = \begin{cases} \hat{p}_j & \text{si } \hat{p}_j \leq 0, \\ (x_j^k)^2 \hat{p}_j & \text{sinon.} \end{cases} \quad (2.16)$$

La direction réduite en $x_{\bar{T}}^k$ sera donc

$$d_{\bar{T}}^k = -\nabla \theta(\hat{p}). \quad (2.17)$$

Cette direction vérifie les propriétés suivantes :

THÉORÈME 2 : Soit d_I^k définie par (2.16) et (2.17).

Alors :

- (i) d_I^k est une direction réalisable en x_I^k ,
- (ii) On a $\langle d_I^k, \hat{g}^i \rangle - u_k \alpha_i^k \leq -\|L(x_I^k, \hat{p})\|^2 - u_k \varepsilon$, $i = 1, \dots, k$.

Preuve : (i) Évident à partir de la définition de d_I^k .

(ii) Si ω et ω_i ($1 \leq i \leq k$) sont les multiplicateurs de Lagrange correspondant aux contraintes $\sum \lambda_i = 1$ et $\lambda_i \geq 0$, $i = 1, \dots, k$ respectivement dans (\bar{D}^k) , alors les conditions d'optimalité pour (\bar{D}^k) peuvent s'écrire comme

$$\langle \nabla \theta(\hat{p}), \hat{g}^i \rangle + u_k \alpha_i^k + \omega - \omega_i = 0, \quad \forall i = 1, \dots, k,$$

$$\lambda_i^* \omega_i = 0,$$

$$\omega_i \geq 0,$$

$$u_k \left[\sum_{i=1}^k \lambda_i^* \alpha_i^k - \varepsilon \right] = 0.$$

Par conséquent,

$$\begin{aligned} \langle -d_I^k, \hat{g}^i \rangle + u_k \alpha_i^k &\geq \langle -d_I^k, \hat{g}^i \rangle + u_k \alpha_i^k - \omega_i = -\omega = -\sum \lambda_i^* \omega \\ &\geq \langle -d_I^k, \sum \lambda_i^* \hat{g}^i \rangle + u_k \sum \lambda_i^* \alpha_i^k. \end{aligned}$$

Or, de la définition de d_I^k , on a

$$\langle -d_I^k, \sum \lambda_i^* \hat{g}^i \rangle = \langle -d_I^k, \hat{p} \rangle = \|L(x_I^k, \hat{p})\|^2$$

et de la dernière condition d'optimalité, on a

$$u_k \sum \lambda_i^* \alpha_i^k = u_k \varepsilon$$

d'où on peut déduire

$$\langle -d_I^k, \hat{g}^i \rangle + u_k \alpha_i^k \geq \|L(x_I^k, \hat{p})\|^2 + u_k \varepsilon. \quad \blacksquare$$

3. RECHERCHE LINÉAIRE

I étant la base choisie en x , le système

$$h(x) = h(x_I, x_{\bar{I}}) = 0$$

implique l'existence de la fonction implicite φ du voisinage $V(x_{\bar{I}}, r)$ de $x_{\bar{I}}$ dans $V(x_I, c)$, voisinage de x_I , telle que

- $h[\varphi(u), u] = 0, \forall u \in V(x_{\bar{I}}, r),$
- $x_I = \varphi(x_{\bar{I}}).$

Rappelons aussi que le problème (P) est localement équivalent au problème (\hat{P}) si la contrainte $\varphi(u) \geq 0$ reste vérifiée pour u appartenant à un certain voisinage de $x_{\bar{I}}$, ce qui veut dire que le successeur x^+ sera de la forme :

$$x^+ = [\varphi(x_{\bar{I}} + td_{\bar{I}}), x_{\bar{I}} + td_{\bar{I}}]$$

et doit appartenir au domaine réalisable en vérifiant :

$$(i) \quad x_{\bar{I}} + td_{\bar{I}} \geq 0, \quad (3.1)$$

$$(ii) \quad \varphi(x_{\bar{I}} + td_{\bar{I}}) \geq 0. \quad (3.2)$$

La première condition (i) est réalisable pour

$$t \leq r \leq \min_{i=1, \dots, n-m} \left\{ -\frac{(x_{\bar{I}})_i}{(d_{\bar{I}})_i} : (d_{\bar{I}})_i < 0 \right\}. \quad (3.3)$$

La seconde est vérifiée en choisissant la base I telle que $x_I = \varphi(x_{\bar{I}}) > 0$ et $\varphi(x_{\bar{I}} + td_{\bar{I}}) \in V(x_I, c)$ où

$$c \leq \min_{i=1, \dots, m} (x_I)_i. \quad (3.4)$$

Notons la fonction $\psi(t)$ telle que

$$\begin{aligned} \psi: [0, r] &\rightarrow V(x_I, c) \times V(x_{\bar{I}}, r), \\ t &\rightarrow [\varphi(x_{\bar{I}} + td_{\bar{I}}), x_{\bar{I}} + td_{\bar{I}}]. \end{aligned}$$

Cette fonction est différentiable sur $[0, r]$ et

$$\nabla \psi(t) = [-N^t(\psi(t)) B^{-1}(\psi(t))^t d_{\bar{I}}, d_{\bar{I}}]. \quad (3.5)$$

Remarquons aussi que

$$\nabla \psi(0) = [-N^t(x) B^{-1}(x)^t d_{\bar{I}}, d_{\bar{I}}]. \quad (3.6)$$

$\nabla \psi(0)$ est la direction en x obtenue par Abadie [1] en linéarisant $h(x)$. Mais dans notre recherche linéaire économique de type Lemaréchal [14], on prendra plutôt $\nabla \psi(t)$ à la place de la direction d .

Cette recherche consiste à trouver un scalaire $t \in]0, r]$ et un sous-gradient $g(t) \in \partial f(\psi(t))$ tels que au moins une des conditions suivantes soit vérifiée :

1. $t = r$ et

$$\begin{aligned} f[\psi(t)] &\leq f[\psi(0)] - m_1 t S \\ &= f(x) - m_1 t S; \end{aligned} \quad (3.7)$$

2. $t \in]0, r[$ et

$$\begin{aligned} f[\psi(t)] &\leq f(x) - m_1 t S, \\ \langle \nabla \psi(t), g(t) \rangle &\geq -m_2 S; \end{aligned} \quad (3.8)$$

3. $t \in]0, r[$ et

$$\langle \nabla \psi(t), g(t) \rangle \geq -m_2 S, \quad (3.9)$$

$$\alpha(x, \psi(t), g(t)) \leq \varepsilon^1,$$

$$\|x - \psi(t)\| \leq \varepsilon^2, \quad (3.10)$$

où $0 < m_1 < m_2 < 1$, $0 < \varepsilon^1$, $0 < \varepsilon^2$.

Et du Théorème 2, on définit S :

$$S = \|L(x_{\bar{f}}, \hat{p})\|^2 + u\varepsilon > 0. \quad (3.11)$$

La condition (3.10) signifie que la distance entre le point courant x et le point $\psi(t)$ n'est pas très grande.

Les conditions (3.7) et $t = r$ caractérisent l'itération maximale.

Les conditions (3.7) et (3.8) caractérisent l'itération sérieuse.

Les conditions (3.8), (3.9) et (3.10) caractérisent l'itération nulle.

Remarquons que $[N'(\psi(t)) B^{-1}(\psi(t))^t, \text{Id}_{n-m}] g(t)$ n'est pas forcément dans l'ensemble $\partial \hat{f}(x_{\bar{f}} + t d_{\bar{f}})$ où Id_{n-m} est la matrice identité. Lorsque la fonction implicite peut être calculée, on peut avoir la recherche linéaire en un nombre fini d'itérations par bisection de l'intervalle $[0, r]$ et en appliquant l'algorithme rappelé ci-dessous :

3.1. Algorithme de recherche linéaire

Soient $S > 0$, $0 < m_1 < m_2 < 1$, $\varepsilon^1 > 0$, $\varepsilon^2 > 0$, r et c vérifiant (3.3) et (3.4) respectivement.

1. Poser $t^0 = r$, $t_L^0 = r$, $n = 0$ et calculer $g(t^0) \in \partial f(\psi(t^0))$.

Si t^0 vérifie (3.7), aller en 7;

2. Poser $t_L^0 = 0$, $t_R^0 = r$;

3. Calculer $g(t^n) \in \partial f(\psi(t^n))$.

Si t^n vérifie (3.7), poser $t_R^{n+1} = t_R^n$, $t_L^{n+1} = t^n$ et aller en 4.

Sinon, poser $t_R^{n+1} = t^n$, $t_L^{n+1} = t_L^n$ et aller en 5;

4. Si t^n satisfait (3.8), aller en 7;

5. Si $t_L^{n+1} \neq 0$, aller en 6.

Sinon, si t^n vérifie (3.8), (3.9) et (3.10), aller en 7.

6. Poser $t^{n+1} = (t_L^{n+1} + t_R^{n+1})/2$, $n = n+1$ et aller en 3;

7. Poser $g^+ = g(t^n)$, $t^+ = t^n$, $x^+ = \psi(t^+)$;

on a : l'itération maximale si $t^n = r$,

l'itération nulle si $t_L^{n+1} = 0$,

l'itération sérieuse si $t_L^{n+1} \in]0, r[$.

THÉORÈME 3 : *Si la fonction f est convexe ou localement lipschitzienne semi-régulière (« semi-smooth » au sens de Mifflin), c'est-à-dire si la dérivée directionnelle existe, alors l'algorithme se termine en un nombre fini d'itérations.*

Preuve : Similaire à Lemaréchal [14] ou Bihain [3].

Remarque : La recherche linéaire suppose la résolution du système $h(x) = 0$ à chaque itération; ceci exige une procédure algorithmique infinie, ce qui fait que cette recherche linéaire reste théorique. Cependant, elle est complètement implémentable dans le cas où les contraintes sont linéaires et où on a une forme explicite simple de la fonction implicite.

Numériquement, on se contente d'une forme approchée de cette fonction dans le cas non linéaire donné par la méthode rappelée ci-dessous. ■

3.2. Calcul de la fonction implicite

Le calcul de la fonction implicite φ est évidemment équivalent à la résolution du système $h(\cdot, y) = 0$ où y appartient à un voisinage de x_f . Différentes méthodes sont proposées pour résoudre ce problème. Parmi celles-ci, on peut citer la méthode pseudo-Newton, développée et analysée par Blum [4] et reprise par Bihain [3].

Cette méthode se présente de la manière suivante :

Comme h est une fonction de classe C^1 , il existe en x des constantes $\theta > 0$, $L_\theta^d > 0$ et $L_\theta > 0$ telles que, pour tout $y, z \in V(x, \theta)$,

$$\|B(y) - B(z)\| \leq L_\theta^d \|y - z\|,$$

$$\|h(y) - h(z)\| \leq L_\theta \|y - z\|.$$

En supposant que $h(x)=0$, on pose $b=\|B^{-1}(x)\|$ et $y=x_{\bar{I}}+v$. On considère la méthode pseudo-Newton telle que

$$k_n(v)=k_{n-1}(v)+B^{-1}(x_I, x_{\bar{I}}+v)h(x_I+k_{n-1}(v), x_{\bar{I}}+v) \quad (3.12)$$

sous les conditions que les rayons r et c des voisinages $x_{\bar{I}}$ et x_I vérifient respectivement

$$c \leq \min\left(\theta, \frac{1}{4L_\theta^d b}\right), \quad (3.13)$$

$$r \leq \min\left(\frac{c}{4bL_\theta}, \sqrt{\frac{1}{4(L_\theta^d)^2 b^2} - c^2}, \theta\right). \quad (3.14)$$

Alors, pour tout $v \in V(0, r)$, on a

1. La suite $k_n(v)$ bien définie et $\|k_n(v)\| \leq c, \forall n \in \mathbb{N}$;
2. La suite $k_n(v)$ converge vers le point $k^*(v)$ pour tout point initial $k_0(v) \in V(0, c)$ et le point z tel que $z = x_I + k^*(v)$ est l'unique point dans $V(x_I, c)$ solution du système $h(z, x_{\bar{I}}+v) = h(z, y) = 0$;
3. La fonction implicite φ est définie par

$$\begin{aligned} \varphi: V(x_{\bar{I}}, r) &\rightarrow V(x_I, c), \\ y &\rightarrow \varphi(y) = x_I + k^*(y - x_{\bar{I}}), \end{aligned}$$

et est continûment différentiable sur le voisinage $V(x_{\bar{I}}, r)$ telle que

$$\nabla \varphi(y) = -B^{-1}(\varphi(y), y)N(\varphi(y), y).$$

Commentaire : Les constantes θ , L_θ et L_θ^d dépendent de x . Seulement pour assurer la convergence, il faut que les constantes θ soient bornées inférieurement et que L_θ et L_θ^d soient bornées supérieurement. Afin d'éviter cette difficulté, on supposera qu'on a l'hypothèse que h est uniforme localement lipschitzienne, c'est-à-dire que ces constantes ne dépendent pas des points courants.

Cependant, cette hypothèse est vraie pour une suite convergente et, à partir d'un certain rang assez grand si on prend ces constantes au point limite. ■

3.3. Renouvellement de la procédure

Par la recherche linéaire, on obtient t^+ et le nouvel itéré $x^+ = \psi(t^+)$. Mais rien n'assure que $x_I^+ > 0$ et que la matrice de base associée reste non singulière. D'où le calcul d'une nouvelle base I_+ en x^+ .

Remarquons aussi que suivant un exemple dû à Smeers [21], on peut avoir un phénomène de blocage sur les contraintes, ce qui se traduit par le fait que pour toute sous-suite convergente, la base à la limite risque de ne pas être non dégénérée. Ainsi, Smeers a proposé une procédure algorithmique pour le choix de la base assurant la propriété de base (PB) suivante :

Sous l'hypothèse de non dégénérescence (H),

si $\{x^k, k \in K\} \rightarrow \bar{x}$,

si $B(x^k) = B_k \rightarrow \bar{B}$, où $\forall k \in K, I_k = I$,

et si \bar{x} n'est pas stationnaire,

Alors

(i) \bar{B} est non singulière,

(ii) $\bar{x}_I > 0$,

c'est-à-dire que \bar{B} est non dégénérée en \bar{x} .

Remarque : D'un point de vue strictement numérique, la procédure coûte cher du fait qu'elle exige le calcul de la matrice inverse avec un pivot sophistiqué, mais, d'un autre côté, elle assure la convergence.

Par ailleurs, dans le cas de contraintes linéaires, l'algorithme ne nécessite pas la propriété (PB); le choix de la base selon le critère de Wolfe est suffisant pour assurer la convergence.

Dans ce qui suit, on énonce l'algorithme SGRG dont nous établirons la convergence en supposant la base choisie par le procédé de Smeers.

4. ALGORITHME DE SOUS-GRADIENT RÉDUIT GÉNÉRALISÉ

On suppose que l'hypothèse de non dégénérescence (H) est satisfaite. On se donne les paramètres m_1 et m_2 tels que $0 < m_1 < m_2 < 1$ et ε_1 une constante strictement positive. Les constantes θ , L_0^d et L_0 sont fixées par l'hypothèse d'uniforme localement lipschitzienne de h .

4.1. Étape d'initialisation

Soit x^1 un point réalisable de (P), calculer $g^1 \in \partial f(x^1)$. Poser $y^1 = x^1$, $\alpha_1^1 = \alpha(x^1, y^1, g^1) = 0$, $k = 1$.

4.2. Choix de base

Par la procédure de Smeers, calculer la base I_k , la matrice associée $B_k = B(x^k)$, et son inverse B_k^{-1} en x^k .

Poser $b_k = \|B_k^{-1}\|$.

4.3. Calcul de la direction réduite

3. a. Calculer les sous-gradients réduits généralisés

$$\hat{g}^i = g_{I_k}^i - N_k^t (B_k^{-1})^t g_{I_k}^i, \quad \forall i = 1, \dots, k.$$

3. b. Résoudre le sous-problème suivant :

$$\left. \begin{aligned} \min \frac{1}{2} \left\| L \left(x_{I_k}^k, \sum_{i=1}^k \lambda_i \hat{g}^i \right) \right\|^2, \\ \text{sous les contraintes} \\ \sum_{i=1}^k \lambda_i = 1, \\ \lambda_i \geq 0, \quad i = 1, \dots, k, \\ \sum_{i=1}^k \lambda_i \alpha_i^k \leq \varepsilon_k. \end{aligned} \right\} \quad (D^k)$$

Soit λ^k une solution optimale de (D^k) et u_k le multiplicateur de Lagrange associé à la contrainte additionnelle $\sum_{i=1}^k \lambda_i \alpha_i^k \leq \varepsilon_k$.

Poser $p^k = \sum_{i=1}^k \lambda_i^k \hat{g}^i$ et $d_{I_k}^k$ la direction réalisable réduite en $x_{I_k}^k$ telle que

$$j \in \bar{I}_k, \quad d_j^k = \begin{cases} -p_j^k & \text{si } p_j^k \leq 0, \\ -(x_j^k)^2 p_j^k & \text{sinon.} \end{cases}$$

4.4. Test de convergence

Si $d_{I_k}^k \neq 0$, aller en 5.

Sinon, poser $\varepsilon_k = \varepsilon_k/2$ et aller en 3.

4.5. Recherche linéaire

Calculer

$$\bar{c}_k = \min_{i=1, \dots, m} (x_{I_k}^k)_i,$$

$$\bar{r}_k = \min_{i=1, \dots, n-m} \left\{ -\frac{x_i^k}{d_i^k} : d_i^k < 0 \quad \text{et} \quad i \in \bar{I}_k \right\},$$

et poser les valeurs de tolérance de la fonction implicite :

$$c_k = \min \left\{ \theta, \frac{1}{4 L_\theta^d b_k}, \bar{c}_k \right\},$$

$$r_k = \min \left\{ \frac{c_k}{4 b_k L_\theta}, \sqrt{\frac{1}{4 (L_\theta^d b_k)^2} - c_k^2}, \bar{r}_k, \theta \right\}.$$

Appliquer la recherche linéaire en x^k avec les valeurs :

$$S_k = \|L(x_{I_k}^k, p^k)\|^2 + u_k \varepsilon_k,$$

$$\varepsilon^1 = \frac{\varepsilon_k}{k},$$

$$\varepsilon^2 = \frac{1}{k};$$

on obtient le pas t^+ , le point x^+ et $g^+ \in \partial f(x^+)$.

4.6. Itération suivante

Soit $\varepsilon_{k+1} = \min \{ \|d_{I_k}^k\|, \varepsilon_k \}$.

Dans le cas de l'itération sérieuse ou maximale, poser

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

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

$$g^{k+1} = g^+,$$

$$t^k = t^+.$$

Calculer

$$\alpha_i^{k+1} = \alpha(x^{k+1}, y^i, g^i), \quad \forall i = 1, \dots, k,$$

$$\alpha_{k+1}^{k+1} = 0.$$

Poser $k = k + 1$ et aller en 2.

Dans le cas de l'itération nulle, poser

$$\begin{aligned}x^{k+1} &= x^k, \\y^{k+1} &= x^+, \\g^{k+1} &= g^+, \\t^k &= t^+, \\ \alpha_i^{k+1} &= \alpha_i^k, \quad \forall i = 1, \dots, k.\end{aligned}$$

Calculer $\alpha_{k+1}^{k+1} = \alpha(x^{k+1}, y^{k+1}, g^{k+1})$.

Poser $k + 1 = k$ et aller en 3.

Remarque: Numériquement, on s'arrête lorsque ε_k et $\|d_{I_k}^k\|$ sont suffisamment petits, mais, dans le cas non convexe ou si les contraintes sont non linéaires, on ne peut caractériser le point final, ce qui nous amène à considérer l'algorithme infini avec la propriété d'optimalité asymptotique. ■

5. ÉTUDE DE LA CONVERGENCE

On fait les hypothèses suivantes :

(H₁) le domaine réalisable de (P) est borné;

(H) l'hypothèse de non dégénérescence est vérifiée;

(H₂) l'hypothèse d'uniforme localement lipschitzienne pour h et son gradient;

(H₃) f est convexe.

Dans notre étude, l'objectif sera de prouver que, si on n'a pas le cas où $d_{I_k}^k$ restera toujours nulle à partir d'une itération k , ce qui marque le point d'arrêt de l'algorithme, alors on obtient une suite infinie dont tout point d'accumulation est stationnaire.

THÉORÈME 4 : Sous l'hypothèse (H₁), les suites $(\|B_k^{-1}\|)_{k \in \mathbb{N}}$, $(g^k)_{k \in \mathbb{N}}$ et $(d_{I_k}^k)_{k \in \mathbb{N}}$ sont bornées.

Preuve : La suite $(x^k)_{k \in \mathbb{N}}$ étant bornée, soit M une constante telle que

$$\|x^k\| \leq M, \quad \forall k \in \mathbb{N}$$

et soit \bar{x} un point d'accumulation de cette suite :

$$x^k \xrightarrow[k \in K]{} \bar{x}$$

où K infini $\subset \mathbb{N}$.

(i) Puisque le nombre de bases est fini, on peut supposer que, sur K , on a la même base $I_k = I$, $\forall k \in K$.

Dès lors, par le choix des bases par Smeers, on a

$$\lim_K B^{-1}(x^k) = B^{-1}(\bar{x})$$

et donc la suite $(\|B_k^{-1}\|)_{k \in K}$ est bornée.

(ii) Par la définition de g^k dans la recherche linéaire, nous avons

$$g^k \in \partial f(y^k)$$

où $\|x^k - y^k\| \leq (1/k - 1)$.

La suite $(x^k)_k$ étant bornée, la suite $(y^k)_{k \in \mathbb{N}}$ appartient à un compact de \mathbb{R}^n et, par conséquent, la suite $(g^k)_{k \in \mathbb{N}}$ est bornée.

(iii) h étant une fonction de classe C^1 , sous l'hypothèse (H_1) , on peut affirmer que la suite $(N_k)_{k \in K}$ est bornée de même que $(B_k^{-1})_{k \in K}$.

D'où le vecteur $\hat{g}^i = g_{I_k}^i - N_k^t (B_k^{-1})^t g_{I_k}^i$ est borné par une constante qui ne dépend pas de i ni de k .

Par suite et par définition de $d_{I_k}^k$, on peut conclure que la suite $(d_{I_k}^k)_{k \in K}$ est bornée. ■

Dans la suite de cette étude, en se basant sur quatre lemmes, on démontre le théorème suivant :

THÉOREME 5 : *La suite de directions réduites $(d_{I_k}^k)_{k \in K}$ admet 0 comme point d'accumulation.*

Preuve : Tout d'abord, on suppose par l'absurde qu'on a le contraire, c'est-à-dire qu'il existe une constante strictement positive δ telle que

$$(5.1) \quad \|d_{I_k}^k\| \geq \delta, \quad \forall k \in K,$$

ce qui revient aussi à supposer l'existence d'une constante $\eta > 0$ telle que

$$\|L_{I_k}\| = \|L(x_{I_k}^k, p^k)\| \geq \eta \quad (5.2)$$

par le fait que, si l'une tend vers 0, l'autre aussi.

Comme dans toutes les méthodes des faisceaux, on a le résultat suivant :

LEMME 6 : Pour toute sous-suite convergente de $(x^k)_{k \in \mathbb{N}}$, on a

1. Pour tout $b \in \mathbb{R}^+$, il existe un rang n_0 tel que :

$\forall k, i \in K$ avec $k \geq i \geq n_0$, on a

$$\alpha_i^k \leq \alpha_i^i + b.$$

2. La suite $(u_k)_{k \in \mathbb{N}}$ des multiplicateurs de Lagrange est bornée.

Preuve : En remarquant que la suite $(L_{I_k})_{k \in \mathbb{N}}$ est bornée de la même façon que la suite des directions réduites, la preuve est similaire à [13]. ■

LEMME 7 : Les suites $(c_n)_{n \in \mathbb{N}}$ et $(r_n)_{n \in \mathbb{N}}$ sont bornées inférieurement par des constantes $c > 0$ et $r > 0$.

Preuve :

(i) Rappelons que

$$c_k = \min \left\{ \theta, \frac{1}{4 L_0^d b_k}, \bar{c}_k \right\}$$

où

$$\bar{c}_k = \min_{i=1, \dots, m} (x_{I_k})_i.$$

Tout d'abord, on a vu que les $b_k = \|B_k^{-1}\|$ sont bornés supérieurement.

Il reste donc à démontrer que les \bar{c}_k sont bornés inférieurement.

Sinon, il existe une sous-unité $K' \subset K$ vérifiant

$$\begin{aligned} B_k &= B(x^k), & \forall k \in K', \\ \lim x^k &= \bar{x} & \text{et} & \quad I_k = I, \\ \lim \bar{c}_k &= 0. \end{aligned}$$

Par définition des \bar{c}_k , on peut écrire

$$\begin{aligned} 0 &= \lim_k \bar{c}_k = \lim_k \min_i (x_{I_k}^k)_i \\ &= \min_i (\bar{x}_I)_i = 0 \end{aligned}$$

ce qui contredit le fait que $\bar{x}_I > 0$.

(ii) De même pour la suite $(r_k)_{k \in \mathbb{N}}$, on rappelle que

$$r_k = \min \left\{ \frac{c_k}{4 b_k L_\theta}, \sqrt{\frac{1}{4 (L_\theta^d b_k)} - c_k^2}, \bar{r}_k, \theta \right\}$$

où

$$\bar{r}_k = \min_{i=1, \dots, n-m} \left\{ -\frac{x_i^k}{d_i^k} : d_i^k < 0 \quad \text{et} \quad i \in \bar{I}_k \right\}.$$

On suppose que sur une sous-suite infinie $\bar{K} \subset K$, on a

$$\bar{r}_k = \frac{x_i^k}{-(d_{\bar{I}_k}^k)_i}$$

et, dans ce cas $-(d_{\bar{I}_k}^k)_i = (x_i^k)^2 p_i^k$ avec $p_i^k > 0$, d'où

$$\bar{r}_k = \frac{1}{x_i^k p_i^k} \geq r > 0$$

par la simple raison que les suites (x^k) et (p^k) sont bornées. ■

Nous sommes maintenant en mesure d'énoncer la propriété fondamentale que doit avoir l'algorithme pour converger.

LEMME 8 : *Si l'hypothèse (5.1) est vérifiée, le nombre d'itérations maximales est fini.*

Preuve : Par contradiction, supposons qu'il existe une sous-suite infinie \bar{K} telle que $t^k = r_k$ et que

$$f(x^{k+1}) \leq f(x^k) - m_1 r_k S_k, \quad \forall k \in \bar{K}.$$

Mais, d'après la condition (5.2), on a

$$S_k = \|L_{\bar{I}_k}\|^2 + u_k \varepsilon_k \geq \|L_{\bar{I}_k}\|^2 \geq \eta^2 > 0$$

et

$$r_k \geq r > 0,$$

d'où finalement,

$$f(x^{k+1}) \leq f(x^k) - m_1 r \eta^2, \quad \forall k \in \bar{K},$$

ce qui contredit le fait que f est bornée inférieurement. ■

Enfin, comme dans toutes les méthodes utilisant les techniques des faisceaux, une fois que le nombre d'itérations maximales est fini, on se ramène au cas similaire de situation de problème d'optimisation sans contraintes [13]. Les contradictions s'obtiennent par conséquent facilement.

LEMME 9 : Soit la suite des x^k engendrée par l'Algorithme 4 et \bar{x} un point d'accumulation de cette suite tel que

$$x^k \xrightarrow[k \in K]{} \bar{x}, \quad K \subset \mathbb{N}.$$

Alors la suite $(d_{\bar{I}_k}^k)_{k \in K}$ des directions réalisables réduites converge nécessairement vers 0, c'est-à-dire que l'hypothèse (5.1) ne peut avoir lieu.

Preuve : Par l'absurde, on suppose que l'hypothèse (5.1) est vérifiée.

\bar{x} étant un point d'accumulation de la suite $(x^k)_{k \in \mathbb{N}}$,

$$x^k \xrightarrow[k \in K]{} \bar{x}, \quad K \subset \mathbb{N}.$$

Comme le nombre de bases possibles est fini, on peut supposer que, sur une sous-suite infinie de K , on a

$$I_k = I$$

et

$$x_{I_k}^k = x_I^k, \quad x_{\bar{I}_k}^k = x_{\bar{I}}^k.$$

On peut prendre cette sous-suite égale à K .

D'après le théorème 4, on peut écrire que

$$\begin{aligned} \lim_k g^k &= g, \\ \lim_k d_{\bar{I}_k}^k &= d_{\bar{I}} \quad \text{et} \quad \|d_{\bar{I}}\| \geq \delta > 0, \\ \lim_k L_{\bar{I}_k} &= L_{\bar{I}} \quad \text{avec} \quad \|L_{\bar{I}}\| \geq \eta > 0. \end{aligned}$$

De même, soient $\lim_k \varepsilon_k = \varepsilon$, $0 \leq \varepsilon \leq \varepsilon_1$, $\lim_k u_k = u$ et $u \geq 0$.

Comme $S_k = \|L_{\bar{I}_k}\|^2 + u_k \varepsilon_k$, alors la suite $(S_k)_{k \in K}$ converge vers

$$S = \|L_{\bar{I}}\|^2 + u \varepsilon \geq \|L_{\bar{I}}\|^2 \geq \eta^2 > 0. \quad (5.3)$$

D'après le Lemme 8, on peut supposer que $t^k \neq r_k$ pour tout $k \in K$. Aussi, de la recherche linéaire à l'itération k , on a

$$\langle \nabla \psi^k(t^k), g^{k+1} \rangle \geq -m_2 S_k. \quad (5.4)$$

De la condition (ii) du Théorème 2, pour tout $k, p \in K$, avec $p > k$, on a

$$\langle d_{I_p}^p, \tilde{g}^{k+1} \rangle - u_p \alpha_{k+1}^p \leq -S_p. \quad (5.5)$$

Mais, par définition du sous-gradient réduit \tilde{g}^{k+1} à l'itération p , on peut écrire

$$\langle \tilde{g}^{k+1}, d_{I_p}^p \rangle = g^{k+1} \left(\begin{array}{c} -B^{-1}(x^p) N(x^p) \\ \text{Id}_{n-m} \end{array} \right) d_{I_p}^p = \langle g^{k+1}, \nabla \psi^p(0) \rangle.$$

Soustrayons maintenant (5.4) de (5.5):

$$\langle \nabla \psi^p(0) - \nabla \psi^k(t^k), g^{k+1} \rangle - u_p \alpha_{k+1}^p \leq -S_p + m_2 S_k. \quad (5.6)$$

Rappelons que

$$\nabla \psi^k(0) = [-N^t(x^k) B^{-1}(x^k)^t, \text{Id}_{n-m}] d_{I_k}^k$$

et

$$\nabla \psi^k(t^k) = [-N^t(\psi^k(t^k)) \cdot B^{-1}(\psi^k(t^k)) d_{I_k}^k, d_{I_k}^k],$$

$$\text{avec } \psi^k(t) = [\varphi_k(x^k + t d_{I_k}^k), x_{I_k}^k + t x_{I_k}^k],$$

$$B(x^k) = \nabla_I h(x^k) \quad \text{et} \quad N(x^k) = \nabla_{\bar{I}} h(x^k).$$

Comme $d_{I_k}^k$ et $d_{I_k}^p$ convergent vers $d_{\bar{I}}$, h de classe C^1 et g^{k+1} borné, alors

$$\nabla \psi^p(0) \xrightarrow{p \rightarrow \infty} [-N^t(\bar{x}) B^{-1}(\bar{x})^t, \text{Id}_{n-m}] d_{\bar{I}}. \quad (5.7)$$

De même, en démontrant que $\psi^k(t^k) \xrightarrow{k \rightarrow \infty} \bar{x}$, on obtient

$$\nabla \psi^k(t^k) \xrightarrow{k \rightarrow \infty} [-N^t(\bar{x}) B^{-1}(\bar{x})^t, \text{Id}_{n-m}] d_{\bar{I}}. \quad (5.8)$$

En effet, tout d'abord on remarque que les fonctions implicites φ_k , à partir d'un certain rang, deviennent égales à une même fonction φ . Ensuite on a

deux situations à considérer :

1^{re} situation :

Sur une sous-suite infinie, on a des itérations sérieuses telles que

$$\psi^k(t^k) = x^{k+1},$$

vérifiant

$$f(x^{k+1}) \leq f(x^k) - m_2 t^k \|L_{\bar{I}_k}\|^2 \leq f(x^k) - m_2 t^k \eta^2,$$

d'où

$$0 \leq t^k \leq \frac{f(x^k) - f(x^{k+1})}{m_2 \eta^2}.$$

puisque la suite $(f(x^k))_k$ est décroissante bornée, alors $t^k \xrightarrow[k \rightarrow \infty]{} 0$ et par conséquent $\psi^k(t^k) \xrightarrow[k \rightarrow \infty]{} \bar{x}$.

2^e situation :

Sur une sous-suite infinie, on a des itérations nulles telles que

$$\psi^k(t^k) = y^{k+1}, \quad x^{k+1} = x^k \quad \text{et} \quad \|y^{k+1} - x^{k+1}\| \leq \frac{1}{k}.$$

Comme $x^k \xrightarrow[k \in K]{} \bar{x}$, alors nécessairement $\psi^k(t^k) = y^{k+1} \xrightarrow[k \rightarrow \infty]{} \bar{x}$.

Maintenant de (5.7) et (5.8), il en découle

$$\langle \nabla \psi^p(0) - \nabla \psi^k(t^k), g^{k+1} \rangle \xrightarrow[k, p \rightarrow \infty]{} 0. \quad (5.9)$$

Par ailleurs, du Lemme 6, on a

$$\alpha_{k+1}^p \leq \alpha_{k+1}^{k+1} + b$$

et de la recherche linéaire, nous avons

$$\alpha_{k+1}^{k+1} \leq \frac{\varepsilon_k}{k};$$

Donc, de (5.6), on peut écrire

$$\langle \nabla \psi^p(0) - \nabla \psi^k(t^k), g^{k+1} \rangle - u_p \left(\frac{\varepsilon_k}{k} + b \right) \leq m_2 S_k - S_p.$$

Par passage à la limite pour $k, p \in K$ et $p > k$, en utilisant (5.9), on obtient

$$-ub \leq (m_2 - 1)S. \quad (5.10)$$

Dans le Lemme 6, on choisit b tel que

$$b = \begin{cases} 1 & \text{si } u = 0, \\ \frac{(1 - m_2)S}{2u} & \text{si } u \neq 0. \end{cases}$$

A ce niveau, on peut distinguer deux cas :

1^{er} cas : Si $u = 0$.

Alors l'inégalité (5.10) devient

$$(m_2 - 1)S \geq 0 \quad \text{avec } m_2 - 1 < 0 \quad \text{et } S > 0,$$

ce qui est impossible.

2^e cas : Si $u \neq 0$.

Par le choix de b dans ce cas, on obtient de (5.10) :

$$(m_2 - 1)\frac{S}{2} \geq 0,$$

Pour les mêmes raisons que dans le premier cas, là aussi, on a l'impossibilité, d'où la contradiction. ■

Remarque : Comme la suite $(d_{I_k}^k)_{k \in K}$ converge nécessairement vers 0, alors, d'après le choix des $\varepsilon_k = \min \{ \|d_{I_{k-1}}^{k-1}\|, \varepsilon_{k-1} \}$ dans l'algorithme, cette suite des ε_k converge aussi vers 0. ■

Ceci nous permet de caractériser le point final dans le cas où la fonction f est convexe.

THÉORÈME 10 : Soit la suite $(x^k)_{k \in \mathbb{N}}$ engendrée par l'algorithme.

Si on suppose que f est convexe, alors tout point d'accumulation de la suite est un point stationnaire de (P).

Preuve : Soit \bar{x} un tel point d'accumulation.

D'après le Lemme 9, on a

$$x^k \xrightarrow[k \in K]{} \bar{x} \quad \text{et} \quad d_{I_k}^k \xrightarrow[k \in K]{} 0$$

où K est une sous-suite infinie de \mathbb{N} .

Rappelons qu'à l'itération k , en résolvant le sous-problème (D_k) , on obtenait une solution optimale λ^k .

Notons

$$p^k = \sum \lambda_i^k \hat{g}^i$$

où \hat{g}^i est le sous-gradient réduit de g^i :

$$\hat{g}^i = g_{I_k}^i - g_{I_k}^i N_k^t (B_k^{-1})^t$$

avec $B_k = B(x^k)$ la matrice de base en x^k .

p^k s'écrit alors

$$p^k = (\sum \lambda_i^k g_{I_k}^i) - (\sum \lambda_i^k g_{I_k}^i) N_k^t (B_k^{-1})^t.$$

Du fait que $\sum \lambda_i^k \alpha_i^k \leq \varepsilon_k$ et que f est convexe, on a :

$$g^{*k} = \sum \lambda_i^k g^i \in \partial_{\varepsilon_k} f(x^k).$$

Par conséquent, p^k est le sous-gradient réduit de g^{*k} .

Comme $\varepsilon_k \rightarrow 0$ et la suite $(g^{*k})_{k \in K}$ est bornée, on peut déduire que sur une sous-suite $K' \subset K$ on a

$$g^{*k} \rightarrow g^* \in \partial f(\bar{x}).$$

Par ailleurs, comme dans le Lemme 9, on peut supposer qu'on a la même base $I_k = I$, pour tout $k \in K$. Ainsi, on a la suite $(p^k)_{k \in K'}$ qui converge vers p :

$$p = g_I^* - g_I^* N^t(\bar{x}) (B^{-1}(\bar{x}))^t.$$

Enfin, par définition des directions réduites $d_{I_k}^k$:

$$(d_{I_k}^k)_j = \begin{cases} -p_j^k & \text{si } p_j^k \leq 0, \\ -(x_j^k)^2 p_j^k & \text{sinon,} \end{cases}$$

et le fait que $d_{I_k}^k \rightarrow 0$,

on a nécessairement $p \geq 0$ et $\langle \bar{x}_I, p \rangle = 0$, ce qui constitue, comme dans le théorème 1, les conditions de stationnarité de \bar{x} . ■

Remarque: Dans le cas où la fonction est localement lipschitzienne, Mifflin [18] a proposé une modification des fonctions linéarisantes de f :

$$\alpha(x^k, y^i, g^i) = \max \{ f(x^k) - f(y^i) - \langle g^i, x^k - y^i \rangle, r \|x^k - y^i\| \}$$

où $g^i \in \partial f(y^i)$ et g^i est appelé gradient généralisé au sens de Clarke.

Si f est convexe, on prend $r=0$.

Ces fonctions $\alpha: \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ vérifient les propriétés suivantes : elles sont continues et

- (i) $\forall x \in \mathbb{R}^n, \forall g \in \mathbb{R}^n, \alpha(x, x, g) = 0$;
- (ii) $\forall x \in \mathbb{R}^n, \forall g \in \mathbb{R}^n, \forall g \in \partial f(y)$:

$$\alpha(x, y, g) = 0 \Rightarrow g \in \partial f(x).$$

On définit maintenant $G_\varepsilon(x)$:

$$G_\varepsilon(x) = \left\{ \sum \lambda_i g^i : \sum \lambda_i = 0, \lambda_i \geq 0 \quad \text{et} \quad \sum \lambda_i \alpha(x, x^i, g^i) \leq \varepsilon \right\}. \quad (5.11)$$

Si f est convexe, on a $G_\varepsilon(x) \subseteq \partial_\varepsilon f(x)$; par contre, dans le cas localement lipschitzien, on dispose de la propriété suivante:

Soient $(x^k)_{k \in K}$ une suite convergente, $(y^k)_{k \in K}$ une suite bornée, $(g^k)_{k \in K}$ une suite telle que $g^k \in \partial f(y^k)$ et $(\varepsilon_k)_{k \in K}$ une suite de scalaires positifs convergente vers 0.

Si on considère $G_{\varepsilon_k}(x^k)$ définie dans (5.11), alors

$$\overline{\lim} G_{\varepsilon_k}(x^k) \subseteq \partial f(\bar{x})$$

où

$$\lim_k x^k = \bar{x}.$$

Ainsi, dans la preuve du théorème 10, il suffit de remarquer que

$$g^{**k} = \sum_{i=1}^k \lambda_i g^i \in G_{\varepsilon_k}(x^k)$$

et qu'à la limite $g^{**k} \rightarrow g^* \in \partial f(\bar{x})$.

Pour le reste, on arrive aux mêmes conditions:

$$p \geq 0 \quad \text{et} \quad \langle \bar{x}, p \rangle = 0$$

où p est le gradient généralisé réduit au sens de Clarke de g^* , ce qui constitue les conditions de type Kuhn-Tucker [22] pour que \bar{x} soit un point stationnaire de (P). ■

REMERCIEMENTS

L'auteur tient à remercier vivement M. Nguyen Van Hien pour ses précieux conseils, ainsi que M. J.-J. Strodiot pour ses encouragements tout au long de son travail.

BIBLIOGRAPHIE

1. J. ABADIE et J. CARPENTIER, Generalization of the Wolfe Reduced Gradient Method to the Case of Nonlinear Constraints, In *Optimization*, R. FLETCHER ed., London Academic, 1969.
2. A. BIHAIN, V. H. NGUYEN et J. J. STRODIOT, A Reduced Subgradient Algorithm, *Math. Programming Study*, 1987, 30, pp. 127-149.
3. A. BIHAIN, Numerical and Algorithmic Contributions to the Constrained Optimization of Some Classes of Non-Differentiable Functions, *Ph. D. Thesis*, F.U.N.D.P., Namur, Belgium, 1984.
4. E. K. BLUM, Numerical Analysis and Computation, *Theory and Practice*, Addison-Wesley, New York, 1972.
5. J. A. CHATELON, D. W. HEARN et J. J. LOWE, A Subgradient Algorithm for Certain Minimax and Minisum Problems, *S.I.A.M., J. Control Optim.*, 1982, 20, pp. 455-469.
6. M. GAUDIOSO et M. F. MONACO, A Bundle Type Approach to the Unconstrained Minimization of Convex Nonsmooth Functions, *Math. Programming*, 1982, 23, pp. 216-226.
7. W. GOCHET et Y. SMEERS, A Modified Reduced Gradient Method for a Class of Nondifferentiable Problems, *Math. Programming*, 1980, 19, pp. 137-154.
8. P. HUARD, Convergence of the Reduced Gradient Method, In *Nonlinear Programming*, O. L. MANGASARIAN, R. R. MEYER et S. M. ROBINSON, éd., Academic Press, New York, 1975, 2, pp. 29-54.
9. P. HUARD, Un algorithme général de gradient réduit, *Bulletin de la Direction des Études et Recherches*, E.D.F., 1980, Série C, 2, pp. 91-109.
10. C. LEMARÉCHAL et R. MIFFLIN, *Nonsmooth optimization*, Pergamon Press, New York, 1977.
11. C. LEMARÉCHAL, Extensions diverses des méthodes de gradients et applications, *Thèse d'État*, Paris-IX Dauphine, Paris, 1980.
12. C. LEMARÉCHAL, An Extension of Davidon Methods to Non-Differentiable Problems, *Math. Programming Study*, 1975, 3, pp. 95-109.
13. C. LEMARÉCHAL, J. J. STRODIOT et A. BIHAIN, On a Bundle Algorithm for Nonsmooth Optimization, In *Nonlinear Programming*, O. L. MANGASARIAN, R. R. MEYER et S. M. ROBINSON éd., Academic Press, New York, 1981, 4, pp. 245-282.
14. C. LEMARÉCHAL A View of Line Search, In *Lecture Notes in Control and Information Science* A. AUSLENDER, W. OETTLI et J. STOER, éd., Springer, Berlin, 1981.
15. D. G. LUENBERGER, Introduction to Linear and Nonlinear Programming, Academic Press, New York, 1973.
16. R. MIFFLIN, A Stable Method for Solving Certain Constrained Least Squares Problems, *Math. Programming*, 1979, 16, pp. 141-158.
17. R. MIFFLIN, An Algorithm for Constrained Optimization with Semi Smooth Functions, *Math. Oper. Res.*, 1977, 2, pp. 191-207.

18. R. MIFFLIN, Convergence of a Modification of Lemaréchal's Algorithm for non Smooth Optimization, In *Progress in non differentiable Optimization*, E. A. Nurminski ed., I.I.A.S.A., Laxenburg, Austria, 1982.
19. H. MOKHTAR-KHARROUBI, Sur la convergence théorique de la méthode du gradient réduit généralisé, *Numer. Math.*, 1980, 34, pp. 73-85.
20. H. MOKHTAR-KHARROUBI, Sur quelques méthodes de gradient réduit sous contraintes linéaires, *R.A.I.R.O., Anal. Numér.*, 1979, 13, n° 2, pp. 167-180.
21. Y. SMEERS, Generalized Reduced Gradient Method as an Extension of Feasible Directions Methods, *J. Optim. Theory Appl.*, 1977, 22, n° 2, pp. 209-226.
22. J. J. STRODIOT, V. H. NGUYEN et N. HEUKEMES, ϵ -Optimal Solutions in non Differentiable Convex Programming and some Related Questions, *Math. Programming*, 1983, 25, pp. 307-328.
23. P. WOLFE, Reduced Gradient Method, Rand Document, June 1962.
24. P. WOLFE, On the Convergence of Gradient Methods under Constraints, *I.B.M. Journal*, 1972, pp. 407-411.
25. P. WOLFE, Convergence Conditions for Ascent Methods, *S.I.A.M. Review*, 1969, 11, pp. 226-234.
26. P. WOLFE, A Method for Conjugate Subgradients for Minimizing non Differentiable Functions, *Math. Programming Study*, 1975, 3, pp. 145-173.
27. W. ZANGWILL, The Convex-Simplex Methods, *Management Sci.*, 1967, 14, pp. 221-238.