

HOCINE MOKHTAR-KHARROUBI

**Sur quelques méthodes de gradient réduit
sous contraintes linéaires**

RAIRO. Analyse numérique, tome 13, n° 2 (1979), p. 167-180

http://www.numdam.org/item?id=M2AN_1979__13_2_167_0

© AFCET, 1979, tous droits réservés.

L'accès aux archives de la revue « RAIRO. Analyse numérique » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). 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/>

SUR QUELQUES MÉTHODES DE GRADIENT RÉDUIT SOUS CONTRAINTES LINÉAIRES (*)

par Hocine MOKHTAR-KHARROUBI (1)

Communiqué par J CEA

Résumé — On montre que les méthodes de gradient réduit sous contraintes linéaires sont aussi des méthodes de directions réalisables. Ceci nous permettra à partir de deux lemmes dus à Zoutendijk d'établir la convergence de deux variantes assez générales de la méthode de Wolfe, et ce sous des hypothèses plus faibles que celles généralement formulées dans la littérature sur le sujet.

Abstract — In the case of linear constraints, we show that reduced gradient methods are also methods of feasible directions. Using two lemmas due to Zoutendijk, this allows us to prove the convergence of two fairly general variants of the method of Wolfe, under hypotheses which are weaker than those generally assumed in the literature.

INTRODUCTION

La première méthode de gradient réduit fut proposée en 1962 par Wolfe [6] en vue de résoudre des programmes à contraintes linéaires et fonction économique non linéaire de classe C^1 .

Cette méthode fut par la suite reprise et complétée par Faure et Huard [1]. Mais bien qu'elle fut utilisée avec succès pour résoudre des problèmes réels de dimensions importantes, sa « convergence » théorique ne put être établie; bien plus, en 1966 l'auteur de la méthode publia un exemple numérique simple sur lequel l'algorithme convergait vers un point non optimal [7]. Ce cas de non-convergence est en fait dû, d'une part à un phénomène dit de « zigzag », lors duquel les points engendrés par l'algorithme oscillent entre deux ou plusieurs contraintes en tendant vers un point non optimal, et d'autre part à un phénomène dit de « blocage » sur les contraintes au point limite.

Ceci sera explicité plus clairement lors de l'étude.

Par ailleurs le gradient réduit comportant des opérations analogues à celles de la méthode simpliciale (en particulier l'utilisation des bases et des changements de base), peut comme cette méthode, être sujet à des phénomènes de dégénérescence.

Le problème de la convergence théorique du gradient réduit fut étudié par Huard [2], qui sous une hypothèse (H) de non-dégénérescence, établit cette convergence pour des fonctions économiques concaves de classe C^2 et à dérivées secondes bornées.

(*) Reçu septembre 1977

(1) I N A , boulevard Jourdan, Paris

Une variante particulière est donnée par la méthode « convex-simplex » de Zangwill [8] qui en avait établi la convergence pour des fonctions économiques concaves de classe C^1 , sous la même hypothèse (H) et l'une des deux hypothèses (à notre sens trop fortes) suivantes :

($\overline{H1}$) l'algorithme ne cycle pas [8];

($\overline{H2}$) à chaque itération les plus grandes variables sont en base [9].

Enfin comme autre variante connue, citons la méthode de Luenberger [3], dont la convergence fut établie par son auteur pour des fonctions économiques concaves de classe C^1 sous les hypothèses (H) et ($\overline{H2}$).

Dans ce qui suit, sont proposées deux variantes assez générales constituant des extensions des méthodes de Zangwill et de Luenberger.

La convergence de ces variantes sera établie sous l'hypothèse (H) de non-dégénérescence, pour des fonctions économiques ⁽²⁾ de classe C^1 . Nous nous affranchirons ainsi des hypothèses ($\overline{H1}$) et ($\overline{H2}$) et de celles portant sur les dérivées secondes.

Deux grandes parties composeront ce travail : dans la première, après un bref rappel sur des directions réalisables, nous montrons que les méthodes de gradient réduit sous contraintes linéaires, sont aussi des méthodes de directions réalisables.

Ceci nous permettra dans l'autre partie d'établir à l'aide de deux lemmes dus à Zoutendijk, la convergence des méthodes proposées.

Note : Dans tout ce qui suit, par « algorithme convergent », il faut entendre tout algorithme engendrant des suites qui sont :

1° soit finies;

2° soit infinies et vérifiant la propriété qu'en chacun de leurs points d'accumulation les conditions d'optimalité de Kuhn et Tucker sont satisfaites.

1. NOTATIONS : RAPPELS SUR LES DIRECTIONS RÉALISABLES ET PRÉSENTATION DES MÉTHODES DE GRADIENT RÉDUIT

1.1. Notation

R^n désigne l'espace euclidien usuel, muni du produit scalaire $x \cdot y = \sum_{i=1}^n x_i y_i$ et de la norme induite $\|x\|^2 = x \cdot x$,
 si $x, y \in R^n$, $[x, y] = \{z \in R^n / z = x + t(y - x); t \in [0, 1]\}$;
 ∇f désigne le gradient de $f : R^n \rightarrow R$ différentiable;

(²) Concaves

N sera l'ensemble des entiers naturels et si $\{a_k/k \in N\}$ est une suite de nombre réels et si $K \subset N$, on écrira :

$$\sum_K a_k \quad \text{pour} \quad \sum_{k \in K} a_k.$$

1.2. Rappels sur les directions réalisables

Soit le programme général

$$(P) \quad \left\{ \begin{array}{l} \max f(x), \\ x \in D, \end{array} \right. \quad \text{où} \quad \left\{ \begin{array}{l} D \subset R^n, \\ f: R^n \rightarrow R \text{ de classe } C^1. \end{array} \right.$$

DÉFINITION (Zoutendijk [10]) : Soit $x \in D$; $s \in R^n$ est dit direction réalisable en x si :

$$\exists \theta > 0 / [x, x + \theta s] \subset D, \quad (1)$$

$$\nabla f(x) \cdot s > 0, \quad (2)$$

notons

$$S(x) = \{s \in R^n / s \text{ direction réalisable en } x\}.$$

Procédure générale de direction réalisable

On dispose de $\hat{x} \in D$; $k=0$.

Itération k

(★) si $S(\hat{x}) = \emptyset$ stop sinon choisir $s^k \in S(\hat{x})$ et calculer :

$$\theta_M^k = \sup \{ \theta \mid [x, x + \theta s^k] \subset D \},$$

$$\theta_m^k = \sup \{ \theta / \nabla f(x + t s^k) \cdot s^k > 0, \forall t \in [0, \theta] \},$$

$$x^{k+1} = x + \theta_k s^k \quad \text{où} \quad \theta_k = \min \{ \theta_m^k, \theta_M^k \},$$

aller en (★) avec $k+1$ au lieu de k .

REMARQUE : (i) θ_m fournit le premier maximum local de f sur la demi-droite

$$D_k = \{z / z = x^k + t s^k; t \geq 0\};$$

(ii) la procédure précédente ne fournit pas nécessairement des points stationnaires au sens $S(x) = \emptyset$, car s^k est pris quelconque dans $S(\hat{x})$; par contre ce choix arbitraire donne une grande généralité aux deux lemmes suivants :

LEMME 1 2 1 (Zoutendijk [10]) Soit $\{(x^k, s^k)/k \in K\}$ une sous-suite d'une suite infinie bornée engendrée par la procédure précédente

si $\exists \alpha > 0/\forall k \in K, \nabla f(x^k) \cdot s^k \geq \alpha$ alors

$$f \text{ bornée} \Rightarrow \sum_K \theta_k < +\infty$$

LEMME 1 2 2 (Zoutendijk [10]) Sous les hypothèses du lemme précédent

$$\exists \bar{k}/\forall k \geq \bar{k}, k \in K, \Rightarrow \theta_k = \theta_M^k$$

1.3. Présentation des méthodes de gradient réduit

Nous en donnons un exposé différent des exposés originaux afin de montrer que ce sont des méthodes de directions réalisables

Le programme considéré est de la forme

$$(P) \quad \begin{cases} \max f(x), \\ Ax = a, \\ x \geq 0, \end{cases} \quad \text{où} \quad \begin{cases} A \text{ est une matrice d'ordre } (m, n), \\ a \in R^m, \\ f: R^n \rightarrow R \text{ de classe } C^1 \end{cases}$$

NOTATIONS 1 3 On note D le domaine de (P) soit $Q = \{1, \dots, n\}$, pour $L \subset Q$, on note

$$\bar{L} = Q - L,$$

A^L est la sous-matrice de A dont les colonnes sont indexées dans L ,

$\nabla_L f(x)$ est le vecteur formé de composantes de $\nabla f(x)$ indexées dans L ,

$I \subset Q$ est une base si $|I| = m$ et $\text{rang } A^I = m$

On note alors

$$J(I) = (A^I)^{-1},$$

$T(I) = J(I)A$ le tableau simplicial,

$t(I) = J(I)a$ le second membre

REMARQUES 1 3 1 (i) Soient $x \in D, s \in R^n$, alors

$$s \in S(x) \Leftrightarrow \begin{cases} A s = 0 & \nabla f(x) \cdot s > 0 \\ \exists \theta > 0/x + \theta s \geq 0, \end{cases} \quad \begin{matrix} (1 \ 3 \ a) \\ (1 \ 3 \ b) \end{matrix}$$

$$(ii) \quad (1 \ 3 \ b) \Leftrightarrow \begin{cases} s_j \geq 0, & \forall j/x_j = 0 \end{cases} \quad (1 \ 3 \ c)$$

HYPOTHESE (HR) $\text{rang } A = m$

Par (HR), il existe une base I , induisant sur s la partition $[s_I, s_{\bar{I}}]$ telle que

$$A^I s_I + A^{\bar{I}} s_{\bar{I}} = 0,$$

d'où

$$s_I = -T^I(I) s_{\bar{I}} \quad (1.3d)$$

Par ailleurs :

$$\nabla f(x) \cdot s = \nabla_I f(x) \cdot s_I + \nabla_{\bar{I}} f(x) \cdot s_{\bar{I}}$$

et avec (1.3b) :

$$\Rightarrow \nabla f(x) \cdot s = d^I(x, I) \cdot s_I$$

où

$$d(x, I) = \nabla f(x) - \nabla_I f(x) T(I);$$

Les conditions (1.3a) et (1.3c) deviennent alors :

$$\left. \begin{array}{l} s_I = -T^I(I) s_{\bar{I}}, \quad d^I(x, I) \cdot s_I > 0, \\ s_j \geq 0, \quad \forall j/x_j = 0. \end{array} \right\} \quad (1.3e)$$

REMARQUE 1.3.2 : Toutes les méthodes de gradient réduit diffèrent pour l'essentiel par la façon de définir s à partir de (1.3e).

Dans ce qui suit nous discutons quelques définitions de s_I . Pour simplifier l'écriture nous noterons (lorsqu'aucune confusion n'est à craindre) t , T , d pour $t(I)$, $T(I)$ et $d(x, I)$.

1.3.1. DÉFINITION D 1 (Wolfe [6]):

$$\forall j \in \bar{I}, \quad s_j = \begin{cases} d_j & \text{si } d_j \geq 0 \text{ ou } x_j > 0, \\ 0 & \text{sinon.} \end{cases}$$

1.3.2. DÉFINITIONS PROPOSÉES.

DÉFINITION D 2 : Soit $\rho \in]0, 1]$ quelconque fixé une fois pour toutes, soient

$$\left\{ \begin{array}{l} d_l = \max \{ d_j ; j \in \bar{I} \}, \quad x_p d_p = \{ \min x_j d_j ; j \in \bar{I} \}, \\ \alpha = \max \{ d_l, |x_p d_p| \}, \\ \bar{I}_\rho = \{ j \in \bar{I} / |x_j d_j| \geq \rho \alpha \}, \end{array} \right.$$

ceci étant, on définit s_I comme suit

$$s_j = \begin{cases} d_j & \text{si } d_j \geq 0 \text{ ou } j \in \bar{I}_\rho, \\ 0 & \text{sinon.} \end{cases}$$

REMARQUES 1.3.3 : (i) Dans la méthode « convexe-simplex », Zangwill définit s_I comme suit :

a) si $d_l \geq |x_p d_p|$, s_I est défini par

$$s_j = 0, \quad \forall j \neq l \text{ et } s_l = d_l;$$

b) *si* $s_{\bar{I}}$ *est* *donnée* *par*

$$s_j = 0, \quad \forall j \neq p \quad \text{et} \quad s_p = d_p,$$

on a donc une méthode de relaxation, on notera D 4 cette définition

(ii) La prise en compte d'un plus grand nombre de composantes, tout en ne changeant rien à la preuve de la convergence, fournit une méthode « plus proche » de la méthode originale

(iii) On n'est pas tenu de prendre en compte, tous les éléments de \bar{I}_p . On peut ne considérer qu'un nombre quelconque d'entre eux, ce nombre pouvant même changer d'une itération à l'autre

DÉFINITION D 3 C'est une extension d'une définition proposée par Luenberger [3] Soient n fonctions $\varphi_i : R \rightarrow R$ (éventuellement égales) vérifiant

- (H1) φ_i est continue,
- (H2) $\varphi_i(t) \geq 0, \quad \forall t \in R_+,$
- (H3) $\varphi_i(t) = 0 \Leftrightarrow t = 0,$
- (H4) $d\varphi_i/dt(0)$ existe et est finie

On définit alors $s_{\bar{I}}$ comme suit

$$s_j = \begin{cases} d_j & \text{si } d_j \geq 0, \\ \varphi_j(x_j) d_j & \text{sinon} \end{cases}$$

REMARQUES 1 3 4 Si $\forall i, \forall t, \varphi_i(t) = t$, on retrouve la méthode de Luenberger

PROPRIÉTÉ 1 3 2 (1) *Pour les définitions D 1 et D 3*

$s_{\bar{I}} = 0 \Rightarrow x$ *vérifie les conditions de Kuhn et Tucker*

(2) *Pour les définitions D 2 et D 4*

$\alpha = 0 \Rightarrow x$ *vérifie les conditions de Kuhn et Tucker*

(3) *Pour les quatre définitions*

si $s_i \neq 0$ *et* *si* $\exists \theta > 0 / x + \theta s \geq 0$ *ou*

alors $s \in S(x)$ $s = [s_I, s_{\bar{I}}]$ *et* $s_I = -T^{\bar{I}} s_{\bar{I}}$,

Preuve 1° Pour les définitions D 1, il est facile de vérifier que $s_{\bar{I}} = 0 \Leftrightarrow \{d_{\bar{I}} \leq 0 \text{ et } x_{\bar{I}} d_{\bar{I}} = 0\}$ On voit alors aisément qu'il existe $u = \nabla_I f(x) J$ tel que $d = \nabla f(x) + u A \leq 0$ et $x d = 0$ ce qui exprime les conditions de Kuhn et Tucker

2° Concernant D 2 et D 4, on a

$$\alpha = 0 \Leftrightarrow \begin{cases} d_I = 0, \\ x_p d_p = 0, \end{cases} \tag{1}$$

$$(1) \Leftrightarrow \{d_{\bar{I}} \leq 0, \tag{2}$$

ce qui avec (2) donne

$$\{x_j d_j = 0, \quad \forall j \in I \quad (4)$$

(3) et (4) expriment les conditions $(K - T)$.

3° Pour ce qui est de D 3, remarquons que :

$$s_r = 0 \Leftrightarrow \begin{cases} d_r \leq 0, \\ d_j < 0 \Rightarrow \varphi_j(x_j) = 0, \end{cases} \quad (5)$$

$$(6) \text{ et (H3)} \Leftrightarrow \{d_j < 0 \Rightarrow x_j = 0, \quad (6)$$

$$(7) \Leftrightarrow \{x_i \cdot d_r = 0, \quad (7)$$

(5) et (7) expriment les conditions $(K - T)$.

Pour toutes les définitions, s a été construit de façon à satisfaire :

$$As = 0 \quad \text{et} \quad \nabla f(x) \cdot s = d_r \cdot s_r > 0,$$

si de plus $\exists \theta/x + \theta s \geq 0$, alors $s \in S(x)$.

REMARQUE 1.3.5 : Si $s_r \neq 0$ et si (ii) n'est pas vérifiée :

$$\exists r \in I/x_r = 0 \quad \text{et} \quad s_r < 0.$$

On dira alors que le couple (A^I, x) est dégénéré; dans ce qui suit, on fera l'hypothèse suivante :

HYPOTHÈSES DE NON-DÉGÉNÉRESCENCE (Huard [2]) :

$$(H) \quad \forall x \in D, \quad \text{rang } A^M = m \quad \text{où } M = \{j/x_j > 0\}.$$

Conséquences : (i) sous l'hypothèse (H), $\forall x \in D, \exists$ une base $I/x_i > 0$;

(ii) dans ces conditions $\theta_M = \sup \{ \theta/[x, x + \theta s] \in D \} > 0$, le couple (A^I, x) est alors dit non dégénéré.

2. ALGORITHME GÉNÉRAL DE GRADIENT RÉDUIT ET ÉTUDE DE LA CONVERGENCE

On dispose de $x \in D$, d'un couple (A^I, x) non dégénéré et du tableau simplicial $T = JA$.

Itération courante :

$$(1) \text{ calculer } d_r = \nabla_T f(x) - \nabla_I f(x) T^I;$$

(2) définir s_r selon 1.3.2;

(2a) si $\left\{ \begin{array}{l} s_T=0 \text{ pour D 1; D 3} \\ \alpha=0 \text{ pour D 2; D 4} \end{array} \right\}$ stop;

(2b) sinon aller en (3).

(3) calculer :

$$\left\{ \begin{array}{l} s_j = -T^T s_T \quad \text{et} \quad s = [s_i, s_j], \\ \theta_M = \min \{ x_i / -s_i; s_i < 0 \}, \\ \theta_m = \sup \{ \theta / \nabla f(x + ts) \cdot s > 0, \forall t \in [0, \theta] \}, \\ \bar{x} = x + \theta s \quad \text{où} \quad \theta = \min \{ \theta_m, \theta_M \}, \end{array} \right.$$

définir $Z = \{j/\bar{x}_j = 0\}$ aller en (4);

(4) (a) si $\theta < \theta_M$ ou $Z \cap I = \emptyset$ reprendre en (1) avec \bar{x} et le même tableau T ,

(b) sinon changer de base afin d'éliminer les variables indicées dans $Z \cap I$, puis reprendre en (1) avec \bar{x} et la dernière base obtenue.

2.2. Élimination des variables basiques nulles

Soit $[A^I, A^{\bar{I}}]$ une partition de A , où I est une base.

Soit $[x_I, x_{\bar{I}}]$ la partition induite sur $x \in D$; notons

$$\begin{aligned} M &= \{j/x_j > 0\}, & \bar{M} &= M \cap \bar{I}, \\ Z &= \{j/x_j = 0\}, & \bar{Z} &= Z \cap I. \end{aligned}$$

LEMME 2.2 (Huard [2]) : Si rang $A^M = m$ et si $\bar{M} \neq \emptyset$ alors :

$$\forall r \in Z, \exists p \in \bar{M}/T_r^p \neq 0 \quad \text{où} \quad T = (A^I)^{-1} A.$$

REMARQUE : Le lemme 2.2 assure la possibilité d'élimination des variables basiques nulles.

CRITÈRE DE CHOIX DE p : Soit

$$\begin{aligned} \bar{I}_r &= \{j \in \bar{I}/T_r^j \neq 0\}, \\ \bar{x}_q &= \max \{ \bar{x}_j / j \in \bar{I}_r \}. \end{aligned}$$

Le candidat p à la base peut être n'importe quel indice $j \in \bar{I}_r$,⁽³⁾ vérifiant :

$$\bar{x}_j \geq \beta \bar{x}_q \quad \text{où} \quad \beta \in]0, 1].$$

⁽³⁾ Wolfe [6] propose de prendre $p = q$. Un choix totalement différent de ceux qui précèdent est proposé par Huard [2] dans le cadre de la variante qu'il propose.

Dans ce qui suit on supposera pour simplifier que l'algorithme utilise le critère de Wolfe, les démonstrations étant les mêmes pour le cas général.

2.3. Sur la convergence des algorithmes

PROPRIETE 2 3 1 Soit $\{x^k/k \in N\}$ une suite infinie engendree par l'algorithme general, si $\exists K \subset N$ infini tel que $\{\|s_l^k\|/k \in K\} \rightarrow 0$ alors tout point d'accumulation de la sous-suite $\{x^k/k \in K\}$ verifie les conditions (K - T)

REMARQUE Par un abus de notation, nous ecrivons s_l^k pour $s_{I_k}^k$ (I_k etant la base en x^k)

Preuve de la propriete Soient $\bar{x} \in \bar{K} \subset K/\{x^k/k \in \bar{K}\} \rightarrow \bar{x}$

Les bases etant en nombre fini, $\exists \bar{K} \subset \bar{K}$, une base $I/k \in \bar{K} \Rightarrow I_k = I$, par suite

d'où $d(x) = \nabla f(x) - \nabla_I f(x) T$ est une fonction continue, (2 3 a)

$$\{d(x^k)/k \in \bar{K}\} \rightarrow \bar{d} = \nabla f(\bar{x}) - \nabla_I f(\bar{x}) T$$

A ce niveau on affirme que

$$\bar{d} \leq 0 \quad \text{et} \quad \bar{x} \bar{d} = 0, \tag{2 3 b}$$

sinon s'il existait un indice j tel que (1) ou (2)

(1) $\bar{d}_j > 0$, ceci avec (2 3 a) \Rightarrow

$$\exists \bar{\alpha} > 0, \quad \bar{k}/k \geq \bar{k}, \quad k \in \bar{K} \Rightarrow d_j^k \geq \bar{\alpha}, \tag{2 3 c}$$

(2) $\bar{d}_j < 0$ et $\bar{x}_j > 0$ ce qui toujours avec (2 3 a) \Rightarrow

$$\exists \bar{\alpha} > 0, \quad \bar{k}/k \geq k, \quad k \in K \Rightarrow \left\{ \begin{array}{l} |d_j^k| \geq \bar{\alpha} \\ x_j^k \geq \bar{\alpha} \end{array} \right\} \tag{2 3 d}$$

(2 3 d) et (H3) (hypothese de la definition D 3) donnent

$$\exists \dot{\alpha} > 0/k \geq \bar{k}, \quad k \in \bar{K} \Rightarrow \varphi_j(x_j^k) \geq \dot{\alpha} \tag{2 3 e}$$

Conclusion (i) On voit alors aisement a partir de (2 3 c, d, e) que pour l'algorithme general

$$\exists \alpha_0 > 0, \quad k_0/k \geq k_0, \quad k \in \bar{K} \Rightarrow \|s_l^k\| \geq \alpha_0$$

et ceci contredit l'hypothese de la propriete,

(ii) (2 3 b) est donc etabli, ce qui exprime precisement les conditions (K - T) pour \bar{x}

PROPRIETE 2 3 2 [2] Si la suite engendree par l'algorithme converge, le nombre de changements de base est fini

PROPRIÉTÉ 2.3.3 : S'il existe $K \subset N$ tel que $\{\|s_i^k\|/k \in K\} \rightarrow 0$ et si f est concave, alors tout point d'accumulation de la suite totale $\{\bar{x}/k \in N\}$ est optimal.

Preuve : Par la propriété 2.3.1, on sait que tout point d'accumulation de $\{\bar{x}/k \in K\}$ est stationnaire.

Soit \bar{x} un de ces points : f étant concave, \bar{x} est optimal et $\bar{f} = f(\bar{x})$ est la valeur optimale du programme.

La suite $\{f(\bar{x}^k)/k \in N\}$ étant croissante, f est constante sur l'adhérence de $\{\bar{x}^k/k \in N\}$; par conséquent tout point d'accumulation x^* vérifie $f(x^*) = \bar{f}$; ce qui établit la propriété.

2.4. Analyse de la convergence

HYPOTHÈSES : (i) f est de classe C^1 et concave;
(ii) D est borné.

Soit $\{(x^k, s^k)/k \in N\}$ une suite engendrée par l'algorithme général.

Mis à part le cas évident d'une suite finie (le dernier point \bar{x} engendré étant optimal) nous avons à envisager les deux cas suivants :

1^{er} cas : $\inf\{\|s_i^k\|/k \in N\} = 0$.

Alors $\exists K \subset N/\{\|s_i^k\|/k \in K\} \rightarrow 0$.

On peut alors (sous les hypothèses ci-dessus) utiliser la propriété 2.3.3 et donc tout point d'accumulation de la suite totale $\{\bar{x}/k \in N\}$ est optimal.

2^e cas : $\inf\{\|s_i^k\|/k \in N\} = \alpha > 0$.

Il est alors facile de vérifier (*) à l'aide des définitions 1.3.2 que

$$\exists \alpha^* > 0 / \forall k \in N, \quad \nabla f(x^k) \cdot s^k \geq \alpha^*. \quad (1)$$

Par application des lemmes 1.2.1 et 1.2.2, on sait que

$$\sum_N \theta_k < +\infty \quad \text{et} \quad \exists k_0/k \geq k_0 \Rightarrow \theta_k = \theta_M^k \quad (2)$$

(*) Cette vérification est essentiellement technique, nous n'avons pas jugé utile de la reproduire

(2) donne

$$\{\theta_M^k/k \in N\} \rightarrow 0, \quad (3)$$

le nombre de bases étant fini

$$\exists \bar{\rho} > 0 / \|T(I_k)\| \leq \bar{\rho}, \quad \forall k,$$

D étant compact, ∇f continu ainsi que les fonctions φ_i :

$$\left. \begin{aligned} &\exists \bar{\rho} > 0 / \|s_I^k\| \leq \bar{\rho}, \quad \forall k, \\ &\|s^k\| \leq \|s_I^k\| + \|s_J^k\| \leq (1 + \bar{\rho}) \bar{\rho} = \rho > 0. \end{aligned} \right\} \quad (4)$$

A ce niveau, on affirme que la suite totale converge; en effet

$$x^{k+1} - x^0 = \sum_{l=0}^k x^{l+1} - x^l = \sum_{l=0}^k \theta_l s^l.$$

La série de terme général $\theta_l s^l$ est convergente car elle est normalement convergente; en effet

$$\sum_{l \in N} \|\theta_l s^l\| = \sum_{l \in N} \theta_l \|s^l\| \leq \rho \sum_{l \in N} \theta_l < +\infty \quad \text{par (2).}$$

Comme la convergence de la série $\sum \theta_l s^l$ équivaut à la convergence de la suite $\{x^k/k \in N\}$, on a le résultat annoncé.

La suite étant convergente, le nombre de changements de base est fini (propriété 2.3.2), c'est-à-dire :

$$\exists \bar{k}, \text{ une base } I/k \geq \bar{k} \Rightarrow I_k = I. \quad (5)$$

Par suite

$$\{d(x^k, I_k)/k \in N\} \rightarrow \bar{d} = \nabla f(\bar{x}) - \nabla_I f(\bar{x}) T. \quad (6)$$

Soit $k^* > \max\{k_0, \bar{k}\}$: de (3) et (5) on déduit que

$$\left. \begin{aligned} &\exists K \subset N, \quad q \in \bar{I}/k \geq k^*, \quad k \in K \\ &\Rightarrow \theta_M^k = \bar{x}_q^k - s_q^k \quad \text{et} \quad x_q^{k+1} = 0 \\ &\text{donc } \{x_q^k/k \in K\} \rightarrow \bar{x}_q = 0. \end{aligned} \right\} \quad (7)$$

A ce niveau considérons séparément les différentes définitions :

a) Cas de la définition D 1

(5) exprime qu'à partir de l'itération \bar{k} , l'algorithme coïncide avec l'algorithme du gradient projeté dont on sait qu'il ne converge pas (cf [7])

(3) et (7) expriment qu'à partir de l'itération k^s on peut avoir un phénomène de zig-zag, suivi d'un phénomène de blocage sur la contrainte $x_q=0$ (cf [7])

b) Cas de la définition D 3

(2) et (7) donnent

$$\theta_M^k = x_q^k / -s_q^k \quad \text{où} \quad s_q^k = \varphi_q(x_q^k) d_q^k \quad \text{et} \quad x_q^{k+1} = 0,$$

$$\bar{x}_q = 0 \quad \text{donc} \quad \varphi_q(\bar{x}_q) = 0,$$

ceci permet d'écrire

$$\theta_M^k = \frac{1}{d_q^k} \left[\frac{x_q^k - \bar{x}_q}{\varphi_q(x_q^k) - \varphi_q(\bar{x}_q)} \right],$$

par suite

$$\{\theta_M^k / k \in N\} \rightarrow -\frac{1}{\bar{d}_q} \left[\frac{1}{d\varphi_q(0)/dt} \right] \neq 0$$

Ceci contredit (3), le cas (b) est donc exclu pour D 3 ce qui nous ramène au cas (a) pour lequel on a la convergence

CONCLUSION Dans le cas de la définition D 3, l'algorithme est convergent

c) Cas des définitions D 2 et D 4

(7) donne

$$\text{où} \quad \left. \begin{array}{l} q \in \bar{I}, \\ \forall k \in K, \quad |x_q^k d_q^k| \geq \rho \alpha^k, \\ \alpha^k = \max \{ d_i^k, |x_p^k d_p^k| \} \end{array} \right\} \quad (8)$$

REMARQUE S'il s'agit de D 4, on a $\rho = 1$ et $q = p$

A partir de l'itération \bar{k} , la base est constante, par suite, $\forall J \in \bar{I}$, la fonction

$$x \rightarrow d_J(x) = \nabla_J f(x) - \nabla_I f(x) T^J$$

est continue

Il en est alors de même des fonctions

$$\begin{aligned} x &\rightarrow \max \{ d_J(x), J \in \bar{I} \} = d_I(x), \\ x &\rightarrow \min \{ x_J d_J(x), J \in \bar{I} \} = x_p d_p(x) \end{aligned}$$

De cela, on déduit la continuité de la fonction

$$\alpha \quad x \rightarrow \alpha(x) = \max \{ d_l(x), |x_p d_p(x)| \},$$

de (7) et (8) on tire

$$\{ \alpha^k = \alpha(x^k) / k \in N \} \rightarrow 0,$$

ce qui avec la continuité de α , donne

$$\alpha(\bar{x}) = \bar{\alpha} = 0$$

où

$$\bar{\alpha} = \max \{ \bar{d}_{l_*}, | \bar{x}_{p_*} \bar{d}_{p_*} | \} \quad \text{et} \quad \bar{d} = d(\bar{x})$$

et l_* et p_* tels que

$$\bar{d}_{l_*} = \max \{ \bar{d}_j, j \in \bar{I} \} \quad \text{et} \quad \bar{x}_{p_*} \bar{d}_{p_*} = \min \{ \bar{x}_j \bar{d}_j, j \in \bar{I} \}$$

$\bar{\alpha} = 0$ exprime donc que \bar{x} vérifie les conditions de Kuhn et Tucker, \bar{x} est donc optimal dans ce second cas

Nous avons finalement établi le résultat suivant

THEOREME DE CONVERGENCE *On suppose*

- (i) *f concave, de classe C^1 ,*
- (ii) *le domaine D borné,*
- (iii) *L'hypothèse (H) de non-dégénérescence vérifiée*

Alors dans ces conditions, l'algorithme général est convergent pour les définitions D2, D3 et D4

Commentaires 1° L'étude des méthodes de gradient réduit par les directions réalisables a l'avantage de montrer que sous l'hypothèse (H), les cas de non-convergence ne peuvent provenir que de phénomènes de zigzag et de blocage, situation correspondant aux relations (2), (3) et (5) de 2 4

2° Nous avons établi la convergence des variantes proposées sous l'hypothèse (H) de non-dégénérescence pour des fonctions C^1 concaves. Nous nous sommes donc affranchis des hypothèses $(\bar{H}1)$ et $(\bar{H}2)$ (Zangwill-Luenberger) et des hypothèses d'existence et de finitudes des dérivées secondes (Huard)

3° Remarquons que concernant la définition D3, un point de recherche intéressant sur le plan pratique, pourrait être la détermination de systèmes de fonction φ_i permettant « d'améliorer » la convergence. Des fonctions φ_i satis-

faisant les hypothèses H1, 2, 3, 4) existent en grand nombre; par exemple :

$$\begin{aligned}\varphi(t) &= qt; & q > 0, \\ \varphi(t) &= qt e^t, \\ \varphi(t) &= \sum_{j=1}^p a_j t^j & \text{où } a_j \geq 0; & j=1, p.\end{aligned}$$

BIBLIOGRAPHIE

1. P. FAURE et P. HUARD, *Résolution de programmes mathématiques à fonction non linéaire par la méthode du Gradient réduit*, R.F.R.O., n° 36, 1965, p. 167-206.
2. P. HUARD, *Convergence de la méthode du Gradient réduit dans Nonlinear Programming Symposium*, Madison, avril 1974, O. L. MANGASARIAN, R. R. MEYER; S. M. ROBINSON, éd., Academic Press, 1975, p. 29-54.
3. D. LUENBERGER, *Introduction to Linear and Nonlinear Programming*, Addison Wesley, 1973.
4. G. McCORMICK, *Antizigzagging by Bending*, Management Science, vol. 15, 1969, p. 315-320.
5. H. MOKHTAR-KHARROUBI, *Thèse 3^e cycle*, Paris VI, juin 1976
6. P. WOLFE, *The Reduced Gradient Method Rand Document*, juin 1962.
7. P. WOLFE, *On the Convergence of Gradient Method under Constraints*, I.B.M. J., juillet 1972, p. 407-411.
8. W. ZANGWILL, *The Convex-Simplex Method*, Management Science, vol. 14, n° 3, 1967, p. 221-238.
9. W. ZANGWILL, *Nonlinear Programming*, Prentice Hall, 1969.
10. G. ZOUTENDIJK, *Methods of Feasible Directions*, Elsevier Publishing Co., Amsterdam 1960.