

RAIRO

ANALYSE NUMÉRIQUE

B. MARTINET

**Perturbation des méthodes d'optimisation.
Applications**

RAIRO – Analyse numérique, tome 12, n° 2 (1978), p. 153-171.

http://www.numdam.org/item?id=M2AN_1978__12_2_153_0

© AFCET, 1978, 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/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/>

PERTURBATION DES MÉTHODES D'OPTIMISATION. APPLICATIONS (*)

par B. MARTINET ⁽¹⁾

Communiqué par J. CÉA

Résumé. — On se propose d'étudier dans un cadre déterministe les propriétés de quelques algorithmes d'optimisation lorsqu'à chaque itération les informations sont brouillées par des erreurs. On considère essentiellement la perturbation de la méthode du gradient projeté. On retrouve sous des hypothèses convenables les propriétés de l'algorithme classique.

On aborde ensuite de façon succincte la généralisation de ces résultats à une classe d'algorithmes de montée définis par des multiapplications. Enfin on signale quelques exemples d'applications: algorithme de régularisation diagonal, méthode de Galerkin diagonale, extension de l'algorithme dual de Hestenes-Rockafellar pour la résolution de problèmes d'optimisation convexe avec contraintes.

1. INTRODUCTION

1.1. La plupart des algorithmes pour résoudre un problème d'optimisation : $\max (f(y) \mid y \in Y)$ consistent à construire une suite (y^n) où à l'itération n on passe de y^n à y^{n+1} par une stratégie qui utilise des informations sur la fonction f , sa dérivée f' , l'ensemble Y , etc.

On se propose dans cette note d'étudier les propriétés de quelques algorithmes lorsque ces informations sont brouillées par des erreurs. L'idée d'introduire des erreurs dans les algorithmes d'optimisation a déjà été étudiée par de nombreux auteurs : par exemple Polak [10], Huard [5], Daniel [3], etc. Nous proposons ici un cadre pour son étude systématique et nous en déduisons des applications qui n'avaient pas toujours été clairement précisées.

Nous allons signaler quelques points d'intérêt de cette étude. Tout d'abord, elle constitue un modèle simple pour de nombreux algorithmes; citons à titre d'exemple :

- la modification des procédés utilisant des dérivées lorsque celles-ci sont obtenues de façon approchée (par différences finies);
- la notion d'algorithme diagonal. Si on considère les suites (f^n) et (Y^n) qui convergent en un sens à préciser respectivement vers les éléments f et Y , à chaque itération n on effectue sur f^n et Y^n les opérations prévues dans

(*) Manuscrit reçu le 20 juin 1975.

(1) U.E.R. de Mathématiques, Université de Provence, Marseille.

l'algorithme initial sur f et Y respectivement. C'est le cas par exemple si $f^n = f + \varphi^n$ où la fonction φ^n a un rôle de régularisation. Cette idée peut aussi être utilisée pour les problèmes fonctionnels discrétisés (calcul des variations, approximation, contrôle optimal); on peut en effet (*cf.* Boyer [2]) envisager un schéma de discrétisation qui varie d'une itération à l'autre;

– la résolution d'un problème max-min. Dans ce cas on a

$$f(y) = \inf_x F(x, y)$$

et il est clair que la détermination de la fonction f et de sa dérivée f' ne peut être faite que de façon approchée (*cf.* Klessig-Polak [6] et Rockafellar [11 b]).

Ce formalisme nous permet aussi de savoir que si les données à chaque moment de décision nous sont fournies par un dispositif « expérimental » (ordinateur, appareil de mesure, etc.) certains procédés restent valables et sinon comment on peut les modifier pour qu'ils soient opérationnels.

Il apparaît alors essentiel de définir au départ un modèle de perturbation. Ceci peut être fait de façon variée en fonction du problème posé et de l'algorithme de base envisagé pour le résoudre. L'aspect concret de ces questions ne peut donc souvent être abordé que sur des exemples précis. On distingue de façon générale les modèles déterministes et les modèles aléatoires. Ces derniers malgré leur intérêt ne sont pas étudiés dans cette note (*cf.* Bertrand [1]). On peut aussi considérer des perturbations « constantes » (c'est-à-dire qui ne dépendent pas de l'ordre d'itération, c'est le cas des erreurs d'arrondi) et des perturbations « variables » : c'est le cas par exemple de perturbations qui tendent vers zéro lorsque l'ordre d'itération tend vers l'infini.

Pour des perturbations de type donné, on peut se proposer d'étudier leur influence sur les algorithmes d'optimisation classiques. L'étude directe séparée de ces divers cas risque d'être bien fastidieuse car elle comporte la répétition d'arguments analogues. Il paraît donc de grand intérêt de considérer des classes d'algorithmes définies par des propriétés abstraites.

Remarquons enfin qu'on peut étendre certains de ces résultats à d'autres problèmes que l'optimisation; c'est ainsi que Sonntag [12] a étudié de façon analogue le cas des inéquations variationnelles.

1.2. Résumé

Dans le cadre limité de cette note, nous étudions d'abord la perturbation d'un procédé simple et typique : la méthode du gradient projeté dans le cas où la dérivé f' existe et satisfait à une condition de Lipchitz. On montre

les propriétés de l'algorithme perturbé qui sous certaines hypothèses sont les mêmes que celles de l'algorithme initial. Pour ne pas allonger ce texte, certains résultats sont énoncés sans démonstration.

On aborde ensuite de façon succincte la généralisation de ces résultats à une classe d'algorithmes de montée définis par des multi-applications.

Enfin, on signale quelques exemples d'applications qui nous paraissent dignes d'intérêt :

- algorithme de régularisation diagonal;
- méthode de Galerkin diagonale;
- extension de l'algorithme dual de Hestenes-Rockafellar pour la résolution de problèmes d'optimisation convexes avec contraintes.

Ce papier n'est qu'un exposé sommaire de ces questions. On pourrait le prolonger dans les directions suivantes :

- autres modèles de perturbations;
- classes plus générales d'algorithmes qui peuvent être perturbés;
- autres exemples d'applications en particulier en ce qui concerne l'approximation des problèmes d'optimisation fonctionnels.

2. MAXIMISATION D'UNE FONCTION PAR UN PROCÉDÉ DE GRADIENT PERTURBÉ

2.1. Notations. Hypothèses

Soit H un espace de Hilbert, Y une partie convexe fermée non vide de H et une fonction $f: H \rightarrow R$ concave qui possède une dérivée f' (au sens de Frechet) laquelle satisfait à une condition de Lipchitz sur H c'est-à-dire qu'il existe un nombre $A > 0$ tel que pour tout $x, y \in H$, on ait

$$\|f'(x) - f'(y)\| \leq A \|x - y\|.$$

On note \langle , \rangle et $\| \cdot \|$ respectivement le produit scalaire et la norme dans H . On suppose aussi qu'on a

$$\begin{cases} m = \sup \{ f(y) \mid y \in Y \} < +\infty, \\ M = \{ y \in Y \mid f(y) = m \} \neq \emptyset. \end{cases}$$

On étudie dans ce paragraphe un algorithme permettant de résoudre de façon approchée le problème suivant :

« trouver $y \in M$ ».

Ce procédé consiste à perturber la méthode classique du gradient projeté (cf. Levitin-Polyak [8] ou Céa [4]). On identifie ici l'espace H avec son dual.

2.2. Algorithme

On suppose ici la constante A connue ⁽²⁾. On construit la suite (y^n) d'éléments de H générés à partir de l'élément initial donné $y^0 \in H$ par les itérations suivantes :

$$y^{n+1} = P_{Y_{n+1}}(y^n + \alpha_n f'(y^n)) + \varepsilon^n. \quad (1)$$

où :

- $P_Z(z)$ est la projection de l'élément z sur l'ensemble Z ;
- (Y_n) est une suite d'ensembles convexes fermés non vides de H tels que :
 - a) $Y \supseteq Y_{n+1} \supseteq Y_n$ pour tout n ,
 - b) $Y \subseteq \liminf_{n \rightarrow \infty} Y_n$ ⁽³⁾,

(ces relations entraînent que : $Y = \limsup Y_n = \liminf Y_n$);

- $\alpha_n \in [a_1, 2/(A + 2a_2)]$ où sont fixés les nombres a_1 et a_2 tels que

$$a_2 > 0 \quad \text{et} \quad \frac{2}{A + 2a_2} \geq a_1 > 0;$$

- (ε^n) est une suite d'éléments de H tels que

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

Si on a : $\varepsilon^n = 0$ et $Y^n = Y$ on retrouve le procédé habituel du gradient projeté. Mais on remarque que les itérés successifs y^n n'appartiennent pas forcément aux ensembles Y_n et que le procédé défini par (1) n'est pas un procédé de montée. La perturbation introduite est caractérisée par le fait qu'à l'itération n on considère l'ensemble Y_{n+1} au lieu de l'ensemble Y et qu'on introduit une erreur à la fois sur $f'(y^n)$ et dans la détermination de la projection sur Y_{n+1} . Dans la plupart des cas le terme ε^n est mal connu. L'hypothèse faite sur les termes ε^n sera modifiée au paragraphe 2.4.4 où on étudiera des perturbations constantes.

2.3. Convergence de l'algorithme

Nous montrons que les hypothèses ci-dessus entraînent que l'algorithme 2.2 possède des propriétés analogues à celles du procédé de gradient habituel.

⁽²⁾ Si celle-ci est inconnue on peut modifier en conséquence le procédé proposé.

⁽³⁾ On note : $\liminf (Y_n) = \{y \in H \mid \text{Il existe une suite } (v^n) \text{ telle que } v^n \in Y_n \text{ et } \lim_{n \rightarrow \infty} v^n = y\}$;

$\limsup (Y_n) = \{y \in H \mid \text{Il existe une sous-suite infinie } v^{n'} \text{ telle que } v^{n'} \in Y_{n'} \text{ et } \lim_{n' \rightarrow \infty} v^{n'} = y\}$.

THÉORÈME 1 : Si la suite (y^n) générée par l'algorithme 2.2 est bornée ⁽⁴⁾, elle possède grâce aux hypothèses du 2.1 et du 2.2 les propriétés suivantes :

$$\lim_{n \rightarrow \infty} f(y^n) = m, \quad w.\text{adh}(y^n) \subset M \quad (5).$$

Démonstration : Posons

$$z^n = y^n - \varepsilon^{n-1} \in Y_n.$$

On a

$$f(z^{n+1}) - f(z^n) = \int_0^1 \langle f'(z^n + t(z^{n+1} - z^n)), z^{n+1} - z^n \rangle dt$$

et comme f' satisfait à une condition de Lipchitz, il vient

$$\begin{aligned} f(z^{n+1}) - f(z^n) &\geq \langle f'(z^n), z^{n+1} - z^n \rangle - \frac{A}{2} \|z^{n+1} - z^n\|^2 \\ &= \left(\frac{1}{\alpha_n} - \frac{A}{2} \right) \|z^{n+1} - z^n\|^2 + \frac{1}{\alpha_n} \langle z^n + \alpha_n f'(z^n) - z^{n+1}, z^{n+1} - z^n \rangle \\ &= \left(\frac{1}{\alpha_n} - \frac{A}{2} \right) \|z^{n+1} - z^n\|^2 \\ &\quad + \frac{1}{\alpha_n} \langle y^n + \alpha_n f'(y^n) - z^{n+1} - \varepsilon^{n-1} + \alpha_n (f'(z^n) - f'(y^n)), z^{n+1} - z^n \rangle. \quad (2) \end{aligned}$$

Or par définition de y^{n+1} on a puisque $z^n \in Y_n \subset Y_{n+1}$:

$$\langle y^n + \alpha_n f'(y^n) - z^{n+1}, z^{n+1} - z^n \rangle \geq 0.$$

D'où on tire de (2) que

$$f(z^{n+1}) - f(z^n) \geq a_2 \|z^{n+1} - z^n\|^2 - \left(A + \frac{1}{a_1} \right) \|\varepsilon^{n-1}\| \cdot \|z^{n+1} - z^n\|.$$

Et comme la suite (z^n) est bornée et que $\lim_{n \rightarrow \infty} \|\varepsilon^n\| = 0$, il vient

$$f(z^{n+1}) - f(z^n) \geq a_2 \|z^{n+1} - z^n\|^2 - \rho_n, \quad (3)$$

où les nombres ρ_n sont tels que : $\lim_{n \rightarrow \infty} \rho_n = 0$.

⁽⁴⁾ L'hypothèse qui impose *a priori* que la suite (y^n) est bornée est très forte; nous la levons en 2.4.1.

⁽⁵⁾ $w.\text{adh}(y^n)$ est l'ensemble des valeurs d'adhérence faible de la suite (y^n) .

On remarque que la suite $(f(z^n))$ n'est pas *monotone* ce qui n'était pas le cas de la méthode non perturbée. C'est la difficulté essentielle de cette étude.

Posons

$$S = \{n \mid f(z^{n+1}) < f(z^n)\}.$$

Nous allons distinguer deux cas.

a) *le cardinal de S est infini*

Montrons d'abord que pour tout $z \in Y$ on a

$$\limsup_{\substack{n \rightarrow \infty \\ n \in S}} \langle f'(z^n), z - z^n \rangle \leq 0. \quad (4)$$

Pour tout $n \in S$, on a grâce à (3) :

$$0 > a_2 \|z^{n+1} - z^n\|^2 - \rho_n.$$

Donc

$$\lim_{\substack{n \rightarrow \infty \\ n \in S}} \|z^{n+1} - z^n\| = 0.$$

Par définition de y^{n+1} on a pour tout n :

$$\langle y^n + \alpha_n f'(y^n) - z^{n+1}, u^n - y^{n+1} \rangle \leq 0, \quad \forall u^n \in Y_{n+1}. \quad (5)$$

Grâce aux propriétés des ensembles Y_n , à tout $z \in Y$ on peut faire correspondre une suite d'éléments (u^n) tels que $u^n \in Y_{n+1}$ et $\lim_{n \rightarrow \infty} \|z - u^n\| = 0$.

En posant

$$\mu^n = z - u^n + z^{n+1} - z^n,$$

l'inégalité (5) devient

$$\langle \alpha_n f'(z^n) + \alpha_n (f'(y^n) - f'(z^n)) + y^n - z^{n+1}, z - z^n - \mu^n \rangle \leq 0, \quad \forall z \in Y.$$

D'où on déduit (4) puisque

$$\lim_{n \rightarrow \infty} \|\mu^n\| = \lim_{n \rightarrow \infty} \|f'(y^n) - f'(z^n)\| = \lim_{n \rightarrow \infty} \|y^n - z^{n+1}\| = 0.$$

Soit $\tilde{y} \in M$. Il vient grâce à la concavité de f :

$$0 \geq f(z^n) - f(\tilde{y}) \geq \langle f'(z^n), z^n - \tilde{y} \rangle,$$

d'où on tire de (4) que

$$\lim_{\substack{n \rightarrow \infty \\ n \in S}} f(z^n) = f(\tilde{y}) = m. \quad (6)$$

Si le complémentaire de S est vide ou fini on a trivialement : $\lim_{n \rightarrow \infty} f(z^n) = m$.
 Supposons qu'il n'en soit pas ainsi. Alors à tout $n \notin S$, on peut faire correspondre les entiers $i(n)$ et $j(n)$ tels que

$$i(n) \in S : i(n) = \max\{l \mid l < n, l \in S\},$$

$$j(n) \in S : j(n) = \min\{l \mid l > n, l \in S\}.$$

On a par construction et grâce à (3) :

$$f(z^{i(n)}) + a_2 \|z^{i(n)+1} - z^{i(n)}\|^2 - \rho_{i(n)} \leq f(z^n) \leq f(z^{j(n)}). \tag{7}$$

D'où on tire grâce à (6) que

$$\lim_{n \rightarrow \infty} f(z^n) = m. \tag{8}$$

Soit $\hat{z} \in w.\text{adh}(z^n)$. Comme la suite (z^n) est incluse dans l'ensemble Y qui est faiblement fermé on a $\hat{z} \in Y$ donc

$$f(\hat{z}) \leq m.$$

Mais comme la fonctionnelle f est faiblement semi-continue supérieurement on a

$$f(\hat{z}) \geq \lim_{n \rightarrow \infty} f(z^n) = m.$$

Donc $\hat{z} \in Y$ et $f(\hat{z}) = m$ et par conséquent : $\hat{z} \in M$.

Par définition des éléments z^n et du fait que la suite (z^n) est bornée et que f' satisfait à une condition de Lipschitz, on voit que le théorème est démontré dans le cas où la suite S est infinie (l'énoncé de ce théorème fait intervenir les éléments y^n au lieu de z^n).

b) *le cardinal de S est nul ou fini*

A partir d'un certain rang N on a

$$m \geq f(z^{n+1}) \geq f(z^n).$$

Par conséquent : $\lim_{n \rightarrow \infty} f(z^n)$ existe et $\lim_{n \rightarrow \infty} [f(z^{n+1}) - f(z^n)] = 0$.

On tire de (3) que

$$\lim_{n \rightarrow \infty} \|z^{n+1} - z^n\| = 0.$$

et on termine la démonstration comme dans le cas précédent.

C. Q. F. D.

2.4. Résultats complémentaires concernant l'algorithme 2.2

2.4.1. Conditions a priori qui assurent que la suite (y^n) est bornée

Nous avons supposé dans le théorème précédent que la suite (y^n) est bornée. Il paraît utile de trouver des hypothèses *a priori* qui assurent cette propriété. C'est l'objet du théorème suivant qui est énoncé sans démonstration.

THÉORÈME 2 : *La suite (y^n) générée par l'algorithme 2.2 est bornée si en plus des hypothèses du 2.1 et du 2.2 on suppose que la fonction f est sup-bornée sur Y c'est-à-dire que pour tout $a \in \mathbf{R}$ l'ensemble $\{y \in Y \mid f(y) \geq a\}$ est borné.*

2.4.2. Nous examinons dans ce paragraphe le cas où la suite (ε^n) est telle que la série $\sum_{n=1}^{\infty} \|\varepsilon^n\|$ soit convergente. Nous montrons que dans le cas où H est un espace de dimension finie, l'algorithme 2.2 génère une suite convergente. Cette propriété a été signalée dans un contexte particulier par Rockafellar [11 a]

THÉORÈME 3 : *Si on suppose en plus des hypothèses du 2.1 que :*

- a) *l'espace H est de dimension finie et $Y = H$;*
- b) *la fonction f est sup-bornée sur H ;*
- c) *les nombres α_n appartiennent à l'intervalle $[a_1, 2/(A + 2 a_2)]$ où sont fixés les nombres a_1 et a_2 tels que*

$$a_2 > 0 \quad \text{et} \quad \frac{2}{A + 2 a_2} \geq a_1 > 0,$$

- d) $\sum_{n=1}^{\infty} \|\varepsilon^n\| < +\infty$, *alors la suite (y^n) générée à partir d'un élément initial $y^0 \in H$ par les itérations*

$$y^{n+1} = y^n + \alpha_n f'(y^n) + \varepsilon^n,$$

converge vers un élément de M .

Démonstration : Les théorèmes 1 et 2 permettent d'affirmer qu'on a

$$\lim_{n \rightarrow \infty} \|y^{n+1} - y^n\| = 0, \quad \lim_{n \rightarrow \infty} f(y^n) = m \quad \text{et} \quad \text{adh}(y^n) \subset M.$$

Posons

$$D = \{n \mid f(y^{n+1}) < f(y^n)\}$$

Nous allons distinguer deux cas :

a) *le cardinal de D est infini*

Comme dans le théorème 1, on montre que pour tout $n \in D$ on a

$$0 > f(y^{n+1}) - f(y^n) \geq a_2 \|y^{n+1} - y^n\|^2 - \left(A + \frac{1}{a_1}\right) \|\varepsilon^{n-1}\| \|y^{n+1} - y^n\|,$$

on en déduit que

$$\sum_{n \in D} \|y^{n+1} - y^n\| < +\infty \tag{9}$$

et

$$\sum_{n \in D} |f(y^{n+1}) - f(y^n)| < +\infty. \tag{10}$$

Comme

$$\sum_{n=1}^{\infty} (f(y^{n+1}) - f(y^n))$$

est fini, on tire de (10) que

$$\sum_{n \notin D} |f(y^{n+1}) - f(y^n)| < +\infty$$

et par conséquent

$$\sum_{n \notin D} \|y^{n+1} - y^n\|^2 < +\infty$$

ce qui compte tenu de (9) permet d'affirmer que

$$\sum_{n=1}^{\infty} \|y^{n+1} - y^n\|^2 < +\infty. \tag{11}$$

Soit $\tilde{y} \in \text{adh}(y^n)$. On a pour tout n puisque $f'(\tilde{y}) = 0$:

$$\begin{aligned} \|y^{n+1} - \tilde{y}\|^2 &= \|y^n + \alpha_n f'(y^n) + \varepsilon^n - \tilde{y}\|^2 \\ &= \|y^n - \tilde{y}\|^2 + \|\alpha_n f'(y^n) + \varepsilon^n\|^2 + 2 \langle \alpha_n (f'(y^n) - f'(\tilde{y})) + \varepsilon^n, y^n - \tilde{y} \rangle. \end{aligned}$$

D'où on déduit en utilisant la concavité de f que pour tout n on a

$$\|y^{n+1} - \tilde{y}\|^2 \leq \|y^n - \tilde{y}\|^2 + \|y^{n+1} - y^n\|^2 + 2 \|\varepsilon^n\| \cdot \|y^n - \tilde{y}\|. \tag{12}$$

Du fait que les séries $\sum_n \|\varepsilon^n\|$ et $\sum_n \|y^{n+1} - y^n\|^2$ sont convergentes et que la suite (y^n) est bornée (th. 2) on déduit de l'inégalité précédente que la suite (y^n) ne peut avoir qu'un seul point d'adhérence ce qui montre le théorème dans ce cas.

b) *le cardinal de D est nul ou fini*

A partir d'un certain rang on a

$$f(y^{n+1}) \geq f(y^n).$$

On en déduit que

$$\sum_{n=1}^{\infty} |f(y^{(n+1)}) - f(y^n)| < +\infty$$

et que

$$\sum_{n=1}^{\infty} \|y^{n+1} - y^n\|^2 < +\infty.$$

La démonstration se termine comme ci-dessus.

C. Q. F. D.

2.4.3. La fonction n'est pas concave

Il est bien connu que dans le cas où la fonction f n'est pas concave la méthode classique du gradient permet sous des hypothèses convenables d'atteindre des points stationnaires. C'est cette propriété que nous allons généraliser ici en introduisant des perturbations. Pour simplifier nous supposons que $Y_n = Y = H$ pour tout n .

THÉORÈME 4 : *Si on suppose que :*

- $Y_n = Y = H$ pour tout n ,
- f est une fonction : $H \rightarrow \mathbf{R}$ dont la dérivée de Fréchet f' existe et satisfait à une condition de Lipchitz;
- l'ensemble \mathcal{A} des points stationnaires de f est une partie fermée, bornée non vide de H . De plus il existe $\eta > 0$ tel que l'ensemble $\{y \in H \mid \|f'(y)\| \leq \eta\}$ soit borné;
- l'ensemble $f(\mathcal{A})$ ne peut prendre que l valeurs distinctes

$$f(\mathcal{A}) = \{\mathcal{V}_1, \mathcal{V}_2, \dots, \mathcal{V}_l\} \quad \text{où } \mathcal{V}_i \neq \mathcal{V}_j \text{ si } i \neq j,$$
- f est sup-borné sur H ;

alors la suite (y^n) générée par l'algorithme 2.2 est telle que

$$\lim_{n \rightarrow \infty} \|f'(y^n)\| = \lim_{n \rightarrow \infty} \|y^{n+1} - y^n\| = 0,$$

$\lim_{n \rightarrow \infty} f(y^n)$ existe et est un élément de $f(\mathcal{A})$.

Démonstration : Posons

$$S = \{n \mid f(y^{n+1}) < f(y^n)\}.$$

Si l'ensemble S est vide ou fini le résultat se montre sans difficulté comme dans le cas de la méthode non perturbée. Supposons donc le contraire. Appelons $P_1, P_2, \dots, P_k, \dots$ les parties de S constituées d'indices successifs.

Si l'ensemble des parties P_i est fini, cela implique que la dernière de ces parties est de cardinal infini. On en déduit aisément que

$$\lim_{n \rightarrow \infty} \|y^{n+1} - y^n\| = \lim_{n \rightarrow \infty} \|f'(y^n)\| = 0.$$

Cette dernière relation entraîne grâce aux hypothèses faites sur f' que la suite (y^n) est bornée.

Donc il existe un réel b tel que pour tout n assez grand on ait :

$$f(y^n) > f(y^{n+1}) \geq b.$$

On en déduit que $\lim_{n \rightarrow \infty} f(y^n)$ existe ce qui démontre le théorème dans ce cas.

Voyons maintenant le cas où l'ensemble des parties P_i est infini. Alors chacune de ces parties P_i contient un nombre fini d'éléments. Comme dans le théorème 1 on montre que

$$\lim_{\substack{k \rightarrow \infty \\ n \in P_k}} \|f'(y^n)\| = \lim_{\substack{k \rightarrow \infty \\ n \in P_k}} \|y^{n+1} - y^n\| = 0.$$

On en déduit que la suite $(y^n)_{n \in S}$ est bornée et que $\lim_{\substack{n \rightarrow \infty \\ n \in S}} d(y^n, \mathcal{A}) = 0$.

On déduit de ces relations que pour tout $\varepsilon > 0$, il existe un entier N tel que pour tout $k > 1$ il existe $j_k \in \{1, 2, \dots, p\}$ tel que pour tout $n \in p_k$ on ait

$$|f(y^n) - \mathcal{V}_{j_k}| < \frac{\varepsilon}{2}. \tag{13}$$

Nous allons montrer maintenant que $\lim_{\substack{n \rightarrow \infty \\ n \in S}} f(y^n)$ existe. Soit d_k l'indice du dernier élément de P_k et p_{k+1} l'indice du premier élément de P_{k+1} . On a par définition :

$$f(y^{p_{k+1}}) \geq f(y^{d_k}). \tag{14}$$

Soit le nombre $\varepsilon > 0$ défini par

$$\varepsilon = \min_{\substack{i, j \in \{1, 2, \dots, p\} \\ i \neq j}} |\mathcal{V}_i - \mathcal{V}_j|$$

On choisit ce nombre dans la relation (13). On déduit alors de (13) et de (14) qu'il existe N tel que pour tout $k \geq n$ on ait

$$\mathcal{V}_{j_{k+1}} \geq \mathcal{V}_{j_k}.$$

Et comme $m \geq \mathcal{V}_{j_k}$ pour tout k , on a $\lim_{k \rightarrow \infty} \mathcal{V}_{j_k}$ existe et appartient à $f(\mathcal{A})$. Ceci entraîne grâce à l'affirmation (13) que $\lim_{\substack{n \rightarrow \infty \\ n \in S}} f(y^n)$ existe. On termine alors la démonstration comme dans le théorème 1 en montrant que $\lim_{n \rightarrow \infty} f(y^n)$ existe et appartient à $f(\mathcal{A})$.

C. Q. F. D.

2.4.4. Les erreurs ε^n ne tendent pas zéro

Ce cas est intéressant car il correspond à des situations concrètes fréquentes. Une difficulté nouvelle se présente par rapport aux cas précédents : les hypothèses du théorème 2 ne suffisent plus à assurer que la suite (y^n) est bornée. Nous considérons dans ce paragraphe le cas simplifié où en plus des hypothèses du 2.1 et du 2.2 (mis à part le fait qu'on n'a plus $\lim_{n \rightarrow \infty} \|\varepsilon^n\| = 0$), on suppose que la fonction f est fortement concave c'est-à-dire qu'il existe $\lambda > 0$ tel que pour tout x et $y \in H$ on a

$$\langle f'(x) - f'(y), x - y \rangle \leq -\lambda \|x - y\|^2,$$

et que pour tout n on a

$$Y_n = Y = H \quad \text{et} \quad \|\varepsilon^n\| \leq b \quad (b > 0).$$

On note \bar{y} la solution unique du problème d'optimisation étudié.

THÉORÈME 5 : *Les hypothèses ci-dessus entraînent que la suite (y^n) générée par l'algorithme 2.2 est telle qu'il existe une constante k telle que*

$$\limsup_{n \rightarrow \infty} \|y^n - \bar{y}\| \leq kb.$$

Démonstration : Celle-ci est technique et analogue à celle du théorème 1; nous n'en donnons que les grandes lignes. Soit :

$$S = \{n \mid f(y^{n+1}) < f(y^n)\}.$$

Considérons le cas où l'ensemble S est vide ou fini. On déduit facilement que

$$\lim_{n \rightarrow \infty} f(y^n) \text{ existe et que } \lim_{n \rightarrow \infty} (f(y^{n+1}) - f(y^n)) = 0. \tag{15}$$

Comme dans le théorème 1 on peut montrer que pour tout n on a

$$f(y^{n+1}) - f(y^n) \geq a_2 \|y^{n+1} - y^n\|^2 - \frac{1}{a_1} \|\varepsilon^n\| \cdot \|y^{n+1} - y^n\|,$$

donc

$$f(y^{n+1}) - f(y^n) \geq a_2 \|y^{n+1} - y^n\|^2 - \frac{b}{a_1} \|y^{n+1} - y^n\|. \quad (16)$$

D'où on déduit grâce à (15) qu'il existe $C_1 > 0$ telle que

$$\limsup_{n \rightarrow \infty} \|y^{n+1} - y^n\| \leq C_1 b,$$

et qu'il existe $C_2 > 0$ telle que

$$\limsup_{n \rightarrow \infty} \|f'(y^n)\| \leq C_2 b. \quad (17)$$

D'où on déduit le résultat sachant que f est fortement concave et que $f'(\bar{y}) = 0$.

On peut montrer aussi ce résultat dans le cas où l'ensemble S est infini par un raisonnement analogue à celui qu'on a utilisé pour le théorème 1.

C. Q. F. D.

3. PERTURBATION DE MÉTHODES GÉNÉRALES DE MONTÉE

Il est intéressant d'étendre les résultats obtenus pour la méthode du gradient projeté à d'autres algorithmes. Nous ne le ferons ici que très partiellement en considérant une classe d'algorithmes de montée définis par un formalisme abstrait dû à Polak [10].

Cette généralisation n'est évidemment pas la seule possible, elle pourrait se faire en utilisant d'autres formalismes (*cf.* par exemple Huard [5] ou Zangwill [13]).

Nous supposons ici que l'espace H est de dimension finie, que l'ensemble Y est une partie convexe et fermée de H , que la fonction $f : H \rightarrow \mathbf{R}$ est concave et que

$$\begin{cases} m = \sup(f(y) | y \in Y) < +\infty, \\ M = \{y \in Y | f(y) = m\} \neq \emptyset. \end{cases}$$

3.1. Algorithme initial

Comme Polak [10], on considère une classe \mathcal{A} de méthodes de montée définies par une multiapplication $\Gamma : H \rightarrow \mathcal{P}(H)$ qui caractérise les itérations où on passe de z^n à z^{n+1} par $z^{n+1} \in \Gamma(z^n)$ (cette multiapplication dépend de la méthode en question et du problème posé). On suppose que :

- a) $\Gamma(x) \neq \emptyset$ pour tout $x \in H$;
- b) $\Gamma(x) = \{x\}$ pour tout $x \in M$;

c) Pour tout $x \notin M$, il existe un nombre $\eta > 0$ et un voisinage \mathcal{V}_x de x tels que pour tout $y \in \mathcal{V}_x$ et tout $z \in \Gamma(y)$ on ait

$$f(z) \geq f(y) + \eta.$$

Ce formalisme englobe de nombreux procédés classiques.

3.2. Algorithme perturbé

A cette classe \mathcal{A} d'algorithmes, on fait correspondre la classe \mathcal{A}^* des algorithmes perturbés ou à l'itération n on passe de y^n à y^{n+1} par

$$y^{n+1} \in \Gamma(y^n) + \varepsilon^n.$$

On suppose ici que les éléments ε^n de H sont tels que

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

THÉORÈME 6 : *Les hypothèses ci-dessus entraînent que pour toute suite (y^n) générée par un algorithme \mathcal{A}^* si elle est bornée a les propriétés suivantes :*

$$\begin{cases} \lim_{n \rightarrow \infty} f(y^n) = m. \\ \text{adh}(y^n) \subset M \end{cases}$$

Démonstration : Soit $u^n = y^{n+1} - \varepsilon^n$.

Par définition, on a $u^n \in \Gamma(y^n)$. Comme la fonction f est continue, que la suite (y^n) est bornée et que : $\lim_{n \rightarrow \infty} \|\varepsilon^n\| = 0$ il vient

$$\lim_{n \rightarrow \infty} (f(u^n) - f(y^{n+1})) = 0. \quad (18)$$

Or $f(u^n) \geq f(y^n)$. On déduit des deux relations précédentes que pour tout n on a

$$f(y^{n+1}) \geq f(y^n) - \rho_n \quad \text{où} \quad \lim_{n \rightarrow \infty} \rho_n = 0. \quad (19)$$

Posons

$$S = \{n \mid f(y^{n+1}) < f(y^n)\}.$$

Nous distinguerons deux cas suivant que l'ensemble S est fini ou non.

a) *L'ensemble S est infini*

On tire de (19) que

$$\lim_{\substack{n \rightarrow \infty \\ n \in S}} (f(y^{n+1}) - f(y^n)) = 0. \quad (20)$$

Soit $\tilde{y} \in \text{adh}(y^n)$. Un tel élément existe car la suite (y^n) est bornée. Il existe donc une sous-suite $(y^n)_{n \in L}$ où $L \subset S$ telle que

$$\lim_{\substack{n \rightarrow \infty \\ n \in L}} y^n = \tilde{y}.$$

Nous allons montrer que $\tilde{y} \in M$.

Supposons (en raisonnant par l'absurde), que ceci ne soit pas vrai. Alors en utilisant les propriétés de Γ on peut affirmer qu'il existe $\eta > 0$ et un entier k tel que pour tout $n \in L$ tel que $n \geq k$ on ait

$$f(u^n) - f(y^n) \geq \eta.$$

Cette affirmation contredit les relations (18) et (20) ce qui prouve que $\tilde{y} \in M$.
Donc $\text{adh}(y^n)_{n \in S} \subset M$.

La continuité de f entraîne alors que

$$\lim_{\substack{n \rightarrow \infty \\ n \in S}} f(y^n) = m.$$

On termine la démonstration comme dans le théorème 1.

b) *L'ensemble S est vide ou fini*

La démonstration se fait sans difficulté.

C. Q. F. D.

On remarque que cette étude ne généralise pas les résultats du paragraphe 2. L'hypothèse faite *a priori* que la suite (y^n) est bornée nous semble indispensable dans le cadre de ce formalisme. De même il ne paraît pas possible sans autres hypothèses d'étendre le résultat dans le cas où H est un espace de Hilbert séparable.

4. APPLICATIONS

4.1. Un procédé de régularisation diagonal

Soit le problème du 2.1 où $Y = H$ et la fonction f est sup-bornée sur H . Considérons le procédé où à l'itération n , on détermine y^{n+1} à partir de y^n par

$$y^{n+1} = y^n + \alpha_n (f'(y^n) - \eta_n y^n),$$

où le nombre α_n est choisi comme en 2.2 et où les nombres η_n sont tels que

$$\begin{cases} 0 \leq \eta_{n+1} \leq \eta_n, \\ \lim_{n \rightarrow \infty} \eta_n = 0. \end{cases}$$

Ce procédé peut être considéré comme une méthode de gradient où à chaque itération n on considère au lieu de f la fonction f_n :

$$y \rightarrow f_n(y) = f(y) - \frac{\eta_n}{2} \|y^n\|^2.$$

On peut l'interpréter comme une méthode de régularisation diagonale (dans le procédé de régularisation habituel, on a $\eta_n = \text{Cte}$). Pour nous placer dans le cadre des hypothèses du théorème 1, il nous faut vérifier que la suite (y^n) ainsi générée est bornée. Ceci est vrai car la fonction f est sup-bornée et on a pour tout n :

$$f(y^n) \geq f(y^1) - \frac{\eta_0}{2} \|y^1\|^2.$$

4.2. Méthode de Galerkin diagonale

La méthode de Galerkin est bien connue (*cf.* Levitin-Polyak [8]) pour résoudre des problèmes d'optimisation de nature fonctionnelle. On peut alors envisager de résoudre les problèmes intermédiaires posés par la méthode du gradient. Le formalisme introduit au paragraphe 2 permet de considérer un nouveau procédé qui serait une méthode de gradient où on peut éventuellement modifier si on le désire à chaque itération la base choisie. L'algorithme du paragraphe 2 où on considère une suite (Y_n) d'espaces permet de retrouver des procédés qui ont été étudiés directement par Boyer [2] dans le cadre des éléments finis. Cette étude est trop longue pour être développée ici. Elle sera présentée dans une publication ultérieure.

4.3. Résolution d'un problème min-max. Algorithme dual de Hestenes-Rockafellar

4.3.1. Énoncé du problème. Notations. Hypothèses

Soit le problème de minimisation convexe avec contraintes suivant :

$$(P) \quad \min(F(x) \mid x \in X, h_i(x) \leq 0, i = 1, 2, \dots, p),$$

où

$\left\{ \begin{array}{l} X \text{ est une partie convexe fermée bornée d'un espace de Banach réflexif } B; \\ F \text{ et } h_i \text{ sont des fonctions : } X \rightarrow \mathbf{R} \text{ convexes et semi-continues inférieurement.} \end{array} \right.$

On suppose que le problème P possède au moins une solution et des multiplicateurs de Kuhn et Tucker.

Comme Rockafellar (cf. bibliographie), on associe au problème P le lagrangien $L_r : X \times \mathbf{R}^p \rightarrow \mathbf{R}$ qui à $(x, y) \in X \times \mathbf{R}^p$ fait correspondre

$$L_r(x, y) = F(x) + \frac{1}{4r} \sum_{i=1}^p ((y_i + 2rh_i(x))^+)^2 - y_i^2,$$

où le nombre $r > 0$ est fixé.

On peut montrer (cf. Rockafellar) que :

- L_r est convexe en x et concave en y ;
- (\bar{x}, \bar{y}) est un col de L_r si et seulement si \bar{x} est solution de P et \bar{y} est multiplicateur de Kuhn et Tucker (on remarquera que les hypothèses faites n'assurant pas l'unicité de la solution de P);

- on définit $f_r : \mathbf{R}^p \rightarrow \mathbf{R} \cup \{-\infty\}$ qui à $y \in \mathbf{R}^p$ fait correspondre $f_r(y) = \inf_{x \in X} L_r(x, y)$.

On peut montrer que pour tout $y \in \mathbf{R}^p$, f_r est concave et à valeurs finies et aussi qu'elle possède une dérivée f'_r qui satisfait à une constante de Lipchitz égale à $1/2r$. On a aussi (cf. Rockafellar) le résultat suivant :

THÉORÈME : *Pour toute suite (y^n) d'éléments de \mathbf{R}^p tels que*

$$\lim_{n \rightarrow \infty} f(y^n) = m,$$

et toute suite (x^n) d'éléments de X tels que pour tout n on ait

$$L_r(x^n, y^n) - f_r(y^n) \leq \rho_n \quad \text{avec} \quad \lim_{n \rightarrow \infty} \rho_n = 0,$$

on peut affirmer que les valeurs d'adhérence faible de la suite (x^n) sont des solutions de P et que les valeurs d'adhérence de la suite (y^n) sont des multiplicateurs du problème P.

Ce théorème et les propriétés ci-dessus nous permettent d'affirmer que pour résoudre le problème P on peut utiliser une méthode duale qui nous conduit à résoudre le problème sans contrainte suivant :

$$\max(f_r(y) \mid y \in \mathbf{R}^p) \quad (r > 0 \text{ quelconque}). \tag{21}$$

Mais comme on sait que toute solution de ce problème appartient à \mathbf{R}^p_+ où $\mathbf{R}^p_+ = \{y \in \mathbf{R}^p \mid y_i \geq 0, i = 1, \dots, p\}$ on peut aussi considérer le problème

$$\max(f_r(y) \mid y \in \mathbf{R}^p_+). \tag{22}$$

Pour résoudre ces problèmes on peut utiliser l'algorithme décrit au paragraphe 2. C'est ce que nous allons voir.

4.3.2. Algorithmes

A la n -ième itération on connaît $y^n \in \mathbf{R}^p$, on détermine $x^n \in X$ tel que

$$L_r(x^n, y^n) \leq \inf_{x \in X} L_r(x, y^n) + \rho_n \quad \text{où} \quad \lim_{n \rightarrow \infty} \rho_n = 0.$$

Ensuite on détermine $y^{n+1} \in \mathbf{R}^p$ tel que

$$y^{n+1} = y^n + \alpha_n \frac{\partial}{\partial y} L_r(x^n, y^n) \quad \text{où} \quad \alpha_n \in [\mu_1, 4r - \mu_2]. \quad (23)$$

(Les constantes μ_1 et μ_2 sont telles que $\mu_1 > 0$, $\mu_2 > 0$ et $4r \geq \mu_1 + \mu_2$ et ne dépendent pas de n .) On vérifie que la formule (23) correspond bien à l'algorithme du paragraphe 2 appliqué au problème (21) car on a

$$\left\| \frac{\partial}{\partial y} L_r(x^n, y^n) - f'_r(y^n) \right\| \leq \sqrt{\frac{\rho_n}{r}}.$$

En explicitant (23) il vient

$$y_i^{n+1} = y_i^n + \alpha_n \max\left(h_i(x^n), -\frac{y_i^n}{2r}\right), \quad i = 1, 2, \dots, p.$$

Si $\alpha_n = 2r$ on retrouve l'algorithme de Hestenes-Rockafellar. Dans ce cas on trouve que $y^n \in \mathbf{R}_+^p$ pour tout n . Si $\alpha_n \neq 2r$ on a en général $y^n \notin \mathbf{R}_+^p$. Pour éviter cet inconvénient (on sait *a priori* que les multiplicateurs sont positifs) on peut envisager de résoudre le problème (22) par l'algorithme du gradient projeté

$$y_i^{n+1} = \left(y_i^n + \alpha_n \max\left(h_i(x^n), -\frac{y_i^n}{2r}\right) \right)^+, \quad i = 1, 2, \dots, p.$$

REMARQUE : Dans la note de Rockafellar citée dans la bibliographie, on justifie la méthode où $\alpha_n = 2r$ par des considérations de point fixe et ce dans le cas où $\sum_n \rho_n < +\infty$ ce qui permet de déduire la convergence de la suite (y^n) (cf. 2.4.2).

BIBLIOGRAPHIE

1. J. P. BERTRAND, *Optimisation stochastique dans un espace de Hilbert par la méthode du gradient*, C.R. Acad. Sc. Paris, t. 276, série A, 1973, p. 613-616.
2. R. BOYER, *Méthodes diagonales en optimisation convexe*, Thèse, Université de Provence, septembre 1974.
3. J. W. DANIEL, *The Approximate Minimization of Functionals*, Prentice Hall, 1971.

4. J. CEA, *Optimisation. Théorie et Algorithmes*, Dunod Paris, 1971.
5. P. HUARD, *Cours d'optimisation*, D.E.A., Université de Lille, 1972.
6. KLESSIG et POLAK, *A Method of Feasible Directions Using Function Approximation with Application to Min-Max Problems*, J. Math. Anal. Appl., vol. 41, 1973, p. 583-602.
7. B. LEMAIRE, Thèse, Paris, 1970.
8. LEVITIN et POLYAK, *Constrained Minimization Problems*, U.S.S.R. Comp. Math. Phys. Math., vol. 5, 1966, p. 1-50.
9. B. MARTINET, (a) Thèse, Grenoble, 1972; (b) Exposé fait au colloque d'analyse numérique de La Colle-sur-Loup, juin 1973.
10. E. POLAK, *Computational Methods in Optimization. A Unified Approach*, Academic press, 1971.
11. R. T. ROCKAFELLAR, (a) *The Multiplier Method of Hestenes and Powell Applied to Non Linear Programming*, University of Washington, 1972; (b) *A Dual Approach to Solving Non Linear Programming Problems by Unconstrained Optimization* (à paraître).
12. Y. SONNATG, Séminaire d'analyse numérique, Université de Provence, janvier 1974.
13. W. I. ZANGWILL, *Non Linear Programming. A Unified Approach*, Prentice Hall, International Series in Management, 1969.