REVUE FRANÇAISE D'INFORMATIQUE ET DE RECHERCHE OPÉRATIONNELLE. SÉRIE ROUGE

E. POLAK
G. RIBIERE

Note sur la convergence de méthodes de directions conjuguées

Revue française d'informatique et de recherche opérationnelle. Série rouge, tome 3, n° R1 (1969), p. 35-43

<http://www.numdam.org/item?id=M2AN_1969__3_1_35_0>

R.I.R.O.
(3e année, N° 16, 1969, p. 35-43)

NOTE SUR LA CONVERGENCE DE METHODES DE DIRECTIONS CONJUGUEES

par E. POLAK (1) et G. RIBIERE (2)

Résumé. — *Ce papier utilise un théorème général qui donne des conditions suffisantes de convergence d'une classe d'algorithmes à direction de déplacement, pour construire un algorithme convergent de directions conjuguées, destiné à la minimisation non contrainte de fonctions réelles dans R^n. Cet algorithme est dérivé, par une modification simple, de la méthode de Fletcher-Reeves. On donne quelques résultats numériques pour illustrer le comportement de ce nouvel algorithme.*

On montre également que la convergence de la méthode de Newton à pas variable, aussi bien que celle de la méthode de la plus grande pente, peut être obtenue à l'aide de ce théorème de convergence.

I. INTRODUCTION

On va s'attacher à démontrer la convergence de plusieurs algorithmes destinés à trouver le minimum d'une fonction. Parmi ceux-ci on retrouve la méthode de plus grande pente et la méthode de Newton à pas variable.

Tout d'abord on donnera un théorème de convergence assez général, puis on l'appliquera à divers algorithmes. On démontrera en particulier la convergence d'une nouvelle méthode de gradient conjugué qui est une modification de la méthode de Fletcher-Reeves (1).

2. THEOREME GENERAL DE CONVERGENCE

On définit tout d'abord un algorithme assez général destiné à minimiser une fonction continue $f(x)$, $x \in \mathbf{R}^n$. On part de x_0 arbitraire et on forme successivement :

$$x_{i+1} = a(x_i) \quad ; \quad f(x_{i+1}) < f(x_i), \tag{2.1}$$

(le procédé s'arrête à un point x_k si $f(x_{k+1}) \geqslant f(x_k)$).

(1) Department of electrical engineering and computer sciences. University of California Berkeley, California.
(2) C.N.R.S. Institut Blaise-Pascal, Service de Développement, Paris.

2.1. Définition : — *On dit que x est un point « désirable » du problème de minimisation si*

$$f(a(x)) \geqslant f(x). \tag{2.2}$$

2.2. Lemme (E. Polak [2]).

Si, pour tout x non « désirable », il existe des constantes dépendant de x, $\varepsilon(x) > 0$ et $\delta(x) > 0$, telles que

$$\|x - x_i\| \leqslant \varepsilon(x) \Rightarrow f(x_i) - f[a(x_i)] \geqslant \delta(x), \tag{2.3}$$

alors toute sous-suite convergente de la suite définie par l'algorithme (2.1) a pour point d'accumulation un point « désirable ».

Démonstration :

Voir l'article de E. Polak « On the convergence of optimization algorithms » dans ce même numéro.

Considérons maintenant une classe d'algorithmes pour minimiser f, supposée continûment différentiable ($f \in C^1(\mathbf{R})$), qui sont caractérisés par les propriétés suivantes :

On part de x_0 arbitraire, puis on calcule les itérés successifs par

$$x_{i+1} = x_i + \mu_i p_i \tag{2.4}$$

où p_i, qui ne dépend pas uniquement de x_i, est la direction de déplacement astreinte à la condition

$$-f'(x_i) \cdot p_i \geqslant \alpha \|f'(x_i)\| \, \|p_i\| \tag{2.5}$$

où $\alpha > 0$ est indépendant de x_i et μ_i est calculé de telle sorte que p_i soit orthogonal au gradient de f en x_{i+1}, plus précisément

$$\mu_i = \inf \{ \mu \mid \mu \geqslant 0 \quad , \quad f'(x_i + \mu p_i) \cdot p_i = 0 \}. \tag{2.6}$$

On peut dire aussi que x_{i+1} réalise le premier minimum de f sur la demi-droite d'origine x_i et de direction p_i. Bien entendu $f(x_{i+1}) \leqslant f(x_i)$.

2.3. Définition : — *On dit que x est stationnaire si*

$$f'(x) = 0. \tag{2.7}$$

2.4. Théorème. — *Toute sous-suite convergente de la suite définie par un algorithme du type* (2.4), (2.5) et (2.6) *a pour point d'accumulation un point stationnaire.*

Démonstration

On peut toujours supposer que les directions p sont normalisées de telle sorte que $\|p\| = 1$ et on définit :

$$S = \{ p \mid \|p\| = 1 \}$$

Soit x_0 un point non stationnaire. Alors

$$\|f'(x_0)\| = \nu > 0. \tag{2.8}$$

Du fait de la continuité de la différentielle de f, il existe une boule $B_\varepsilon(x_0)$ de centre x_0 et rayon ε telle que

$$x \in B_\varepsilon(x_0) \Rightarrow \|f'(x)\| \geqslant \nu/2. \tag{2.9}$$

On définit

$$g(x, p) = -f'(x) \cdot p.$$

Alors d'après (2.5) et (2.9) on a

$$x \in B_\varepsilon(x_0) \Rightarrow g(x, p) \geqslant \alpha \frac{\nu}{2} \|p\| = \beta \tag{2.10}$$

L'uniforme continuité de $g(x, p)$ dans le produit $B_\varepsilon(x_0) \times S$ entraîne qu'il existe λ_m tel que

$$\lambda \in [0, \lambda_m] \Rightarrow |g(x, p) - g(x + \lambda p, p)| \leqslant \beta/2. \tag{2.11}$$

Utilisant le développement de Taylor on a

$$f(x + \lambda_m p) = f(x) - \lambda_m g(x + \lambda p, p)$$

où

$$\lambda \in [0, \lambda_m].$$

D'après (2.10) et (2.11) on a aussi

$$x \in B_\varepsilon(x_0) \Rightarrow g(x + \lambda p, p) \geqslant \beta/2$$

d'où

$$f(x) - f(x + \lambda_m p) \geqslant \lambda_m \beta/2$$

mais d'après la définition (2.6)

$$f(x + \mu p) \leqslant f(x + \lambda_m p)$$

d'où

$$f(x) - f(x + \mu p) \geqslant \lambda_m \frac{\alpha}{4} \|f'(x_0)\| > 0. \tag{2.12}$$

Avec cela, on va montrer tout d'abord que x « désirable » est équivalent à x stationnaire dans le cas d'un algorithme du type (2.4), (2.5) et (2.6).

En effet, supposons que x_0 soit « désirable »; alors si x_0 n'était pas stationnaire on pourrait appliquer (2.12) avec $x = x_0$ d'où

$$0 \geqslant f(x_0) - f(x_0 + \mu p) \geqslant \lambda_m \frac{\alpha}{4} \|f'(x_0)\| > 0$$

ce qui est absurde, donc x_0 est stationnaire.

Maintenant supposons que x_0 soit stationnaire; alors d'après (2.6) $\mu = 0$ et $x_0 + \mu p \equiv x_0$ d'où $f(x_0 + \mu p) = f(x_0)$ et x_0 est « désirable ».

Pour finir la démonstration du théorème il suffit maintenant d'utiliser le lemme 2.2. avec $\delta = \lambda_m \frac{\alpha}{4} \|f'(x_0)\|$.

3. APPLICATIONS

3.1. Méthodes quasi-Newton

Ce type de méthodes, dont fait partie celle de Davidon modifiée par Fletcher-Powell [3] est définie ainsi:

On dispose d'une suite de matrices carrées d'ordre n, H_0, H_1, ..., H_i, ... toutes symétriques définies positives. La direction de déplacement est donnée, en posant $r_i = -[f'(x_i)]^T$, par

$$p_i = H_i r_i \tag{3.1}$$

et on a

$$x_{i+1} = x_i + \mu_i H_i r_i. \tag{3.2}$$

Supposant toujours $f \in C^1(\mathbf{R}^n)$ on va chercher dans quelles conditions l'hypothèse (2.5) est satisfaite, de façon à pouvoir appliquer le théorème 2.4. On a:

$$\frac{-f'(x_i) \cdot p_i}{\|f'(x_i)\| \, \|p_i\|} = \frac{r_i^T p_i}{\|r_i\| \, \|p_i\|} = \frac{r_i^T H_i r_i}{\|r_i\| \, \|H_i r_i\|} \tag{3.3}$$

Appelant $\beta_1^{(i)}$ et $\beta_n^{(i)}$ respectivement la plus grande et la plus petite valeur propre de H_i on a la minoration

$$\frac{r_i^T H_i r_i}{\|r_i\| \, \|H_i r_i\|} \geqslant \frac{\beta_n^{(i)}}{\beta_1^{(i)}}, \tag{3.4}$$

où le rapport $\frac{\beta_n^{(i)}}{\beta_1^{(i)}}$ n'est autre que l'inverse du conditionnement de H_i. Par conséquent l'hypothèse (2.5) sera vérifiée si ce conditionnement est borné pour tout i. Il ne semble pas encore qu'on ait montré que cela était vrai de la méthode de Fletcher-Powell.

Par contre on peut considérer un cas particulier de méthode quasi-Newton. On doit, pour cela, faire les hypothèses:

(3.5) i) $f \in C^2(\mathbf{R}^n)$

ii) la matrice J des dérivées partielles secondes de f vérifie

$$\lambda_n \|y\|^2 \leqslant y^T J(x) y \leqslant \lambda_1 \|y\|^2 \qquad \forall\, x, y \in \mathbf{R}^n \tag{3.6}$$

$$\lambda_n > 0, \qquad \lambda_1 > 0$$

ce qui entraîne que f est au moins strictement convexe.

On choisit pour matrices

$$H_i = J^{-1}(x_i). \tag{3.7}$$

Avec (3.1) et (3.2) la méthode ainsi définie diffère de la méthode bien connue de Newton du fait que dans (3.2) μ_i est généralement différent de 1.

Les inégalités (3.6) indiquent que le conditionnement des matrices H_i est borné et on peut appliquer le théorème 2.4.

Topkis et Veinott [4] ont obtenu des résultats analogues par un procédé différent.

3.2. Méthode du gradient conjugué

A partir de maintenant on suppose que f vérifie (3.5) et (3.6).

On va montrer que l'algorithme de Daniel [5]. dont il a déjà démontré la convergence, vérifie (2.5). Cet algorithme n'est autre qu'une méthode de directions conjuguées [6] lorsque f est quadratique.

On part de x_0 arbitraire, $p_0 = r_0 = -[f'(x_0)]^T$ et on définit les itérés :

$$x_{i+1} = x_i + \mu_i p_i \tag{3.8}$$

$$p_{i+1} = r_{i+1} + \gamma_i p_i\,; \qquad r_{i+1} = -[f'(x_{i+1})]^T \tag{3.9}$$

$$\gamma_i = -\frac{r_{i+1}^T J(x_{i+1}) p_i}{p_i^T J(x_{i+1}) p_i} \tag{3.10}$$

On voit facilement qu'en utilisant (2.6)

$$p_{i+1}^T r_{i+1} = r_{i+1}^T r_{i+1} + \gamma_i p_i^T r_{i+1} = \|r_{i+1}\|^2. \tag{3.11}$$

En outre d'après (3.9)

$$\|p_{i+1}\| \leqslant \|r_{i+1}\| + |\gamma_i| \, \|p_i\|$$

et d'après (3.10) et (3.6)

$$|\gamma_i| \leqslant \frac{\|r_{i+1}\| \, \|p_i\|}{\|p_i\|^2} \frac{\lambda_1}{\lambda_n}$$

d'où

$$\|p_{i+1}\| \leqslant \|r_{i+1}\| \left(1 + \frac{\lambda_1}{\lambda_n}\right)$$

et finalement

$$\frac{p_{i+1}^T r_{i+1}}{\|p_{i+1}\| \, \|r_{i+1}\|} \geqslant \frac{1}{1 + \lambda_1/\lambda_n}$$

C.Q.F.D.

3.3. Une modification de la méthode de Fletcher-Reeves

Dans l'algorithme de Daniel, le calcul de γ_i serait prohibitif puisqu'il faudrait connaître la matrice des dérivées secondes en chacun des itérés. C'est

pourquoi Fletcher-Reeves avaient proposé une expression plus simple; soit

$$\gamma_i = \frac{\|r_{i+1}\|^2}{\|r_i\|^2} \tag{3.12}$$

On va montrer tout d'abord que les expressions (3.10) et (3.12) sont identiques lorsque f est quadratique (1).

En effet, on a le développement de Taylor

$$-r_{i+1} = -r_i + \mu_i J p_i \tag{3.13}$$

où cette fois J est constante. Multipliant à gauche par p_{i-1} on a :

$$-p_{i-1}{}^T r_{i+1} = -p_{i-1}{}^T r_i + \mu_i p_{i-1}{}^T J p_i. \tag{3.14}$$

Tout d'abord $p_{i-1}{}^T r_i$ est nul d'après (2.9).

Ensuite prenant (3.9) et (3.10) on a

$$p_i{}^T J p_{i+1} = p_i{}^T J r_{i+1} + \gamma_i p_i{}^T J p_i, \tag{3.15}$$

dont le second membre est nul d'après la définition (3.10) de γ_i. En remplaçant i par $i-1$ dans (3.15) on a également $p_{i-1}{}^T J p_i = 0$, ce qui, porté dans (3.14) entraîne que $p_{i-1}{}^T r_{i+1} = 0$. Le vecteur r_{i+1} orthogonal à p_{i-1} et p_i sera aussi orthogonal à r_i d'après (3.9) où on remplace i par $i-1$. Pour finir, utilisant (3.13) et (3.10) on a

$$\gamma_i = -\frac{r_{i+1}{}^T J p_i}{p_i{}^T J p_i} = -\frac{r_{i+1}{}^T (r_i - r_{i+1})}{p_i{}^T (r_i - r_{i+1})} \tag{3.16}$$

ce qui, du fait que $r_{i+1}{}^T r_i = 0$ ainsi que $p_i{}^T r_{i+1} = 0$, donne (3.12).

Revenons maintenant au cas où f n'est plus quadratique. On va écrire (3.12) d'une autre façon. Le développement de Taylor

$$-r_{i+1} = -r_i + \mu_i J_{i'} p_i, \tag{3.17}$$

où

$$J_{i'} = \int_0^1 J(x_i + \theta \mu_i p_i)\, d\theta;$$

donne

$$-p_i{}^T r_{i+1} = -p_i{}^T r_i + \mu_i p_i{}^T J_{i'} p_i$$

et

$$\mu_i = \frac{p_i{}^T r_i}{p_i{}^T J_{i'} p_i}$$

ou bien, d'après (3.11)

$$\mu_i = \frac{\|r_i\|^2}{p_i{}^T J_{i'} p_i} \tag{3.18}$$

---

(1) Ceci est bien connu. Par exemple on peut en voir aussi la démonstration dans DURAND, *Résolution des équations algébriques*, t. II, Masson éd. (1962).

D'autre part en utilisant (3.17) on a

$$-r_{i+1}^T r_{i+1} = -r_{i+1}^T r_i + \mu_i r_{i+1}^T J_{i'} p_i,$$

d'où la nouvelle forme de (3.12)

$$\gamma_i = \frac{r_{i+1}^T r_i - \mu_i r_{i+1}^T J_{i'} p_i}{\|r_i\|^2}$$

qui, en utilisant (3.18), donne finalement

$$\gamma_i = \frac{r_{i+1}^T r_i}{\|r_i\|^2} - \frac{r_{i+1}^T J_{i'} p_i}{p_i^T J_{i'} p_i} \tag{3.19}$$

On voit que (3.19) diffère assez peu de l'expression (3.10) utilisée par Daniel. Le deuxième terme du second membre n'en diffère que par la matrice J dont cette fois on prend la moyenne sur le segment $[x_i, x_{i+1}]$ alors que Daniel la calcule en x_{i+1}. En outre, comme on l'a démontré au début de ce paragraphe, le terme $r_{i+1}{}^T r_i$ est nul lorsque la fonction f est quadratique.

Par conséquent si, au lieu de définir γ_i comme Fletcher-Reeves, on choisissait

$$\gamma_i = \frac{\|r_{i+1}\|^2 - r_{i+1}^T r_i}{\|r_i\|^2} = -\frac{r_{i+1}^T J_{i'} p_i}{p_i^T J_{i'} p_i} \tag{3.20}$$

on pourrait facilement, comme on l'a fait de l'algorithme de Daniel, vérifier que (2.5) est satisfait.

Pour conclure remarquons que le minimum d'une fonction f quadratique pourra être obtenu en n itérations exactement avec le nouvel algorithme proposé [avec le choix (3.20) de γ_i] puisque dans ce cas $r_{i+1}^T r_i = 0$.

3.4. Convergence globale des algorithmes de directions conjuguées

En fait avec les hypothèses (3.5) et (3.6) qu'on a dû introduire pour parler des méthodes de directions conjuguées, on peut obtenir un résultat plus fort que le théorème (2.4).

En effet, on a le développement de Taylor

$$f(x) = f(x_0) + f'(x_0) \cdot (x - x_0) + \frac{1}{2}(x - x_0)^T J(x_0 + \theta(x - x_0))(x - x_0)$$

où

$$0 \leqslant \theta \leqslant 1.$$

d'où la minoration, en utilisant (3.6)

$$f(x) \geqslant f(x_0) - \|f'(x_0)\| \, \|x - x_0\| + \frac{1}{2}\lambda_n \|x - x_0\|^2.$$

Cela entraîne que l'ensemble des x tels que $f(x) \leqslant f(x_0)$ est borné et fermé dans $\mathbf{R}^n$ donc compact. On peut donc effectivement extraire de la suite x_i de l'algorithme une sous-suite convergente. D'après le théorème 2.4, au point d'accumulation x^* on a $f'(x^*) = 0$ et puisqu'il n'existe qu'un seul point ayant cette propriété, toute la suite x_i converge en fait vers x^* le minimum unique de f (ce résultat est aussi donné par Daniel).

4. ESSAIS NUMERIQUES

On a comparé numériquement la méthode de Fletcher-Reeves et la méthode modifiée que nous proposons.

Il convient tout d'abord de faire une remarque. Fletcher et Reeves dans leur papier insistent sur la nécessité de briser la suite des directions conjuguées p_i à intervalles réguliers afin d'assurer la convergence de l'algorithme. Cela consiste à ce moment à choisir pour direction de déplacement $p_i = r_i$, donc à repartir suivant le gradient. Ils proposent, comme optimum, de briser toutes les $n + 1$ itérations. Effectivement en procédant ainsi on obtient la convergence dans la plupart des cas. Toutefois, lorsque f est convexe, on a pu observer que cela n'était pas nécessaire et c'est bien ce que montre la démonstration de convergence de notre méthode modifiée. En outre, couper la suite des directions conjuguées trop tôt peut être désastreux pour la convergence si la matrice J est très mal conditionnée au voisinage de la solution. Dans ce cas il est nécessaire de poursuivre les itérations beaucoup plus loin comme on peut le voir dans un article de Ginsburg [7] où f est quadratique.

Lorsque f n'est plus convexe, il est pratiquement nécessaire de briser la suite des directions conjuguées avec les deux méthodes. Toutefois on a vu des cas où, en utilisant la méthode modifiée, on obtenait une convergence raisonnable sans briser la suite.

En particulier avec le test proposé par A. R. Colville (New York Scientific Center I.B.M.):

minimiser

$$f(x) = 100(\xi_2 - \xi_1^2)^2 + (1 - \xi_1)^2 + 90(\xi_4 - \xi_3^2)^2 + (1 - \xi_3)^2 + 10 \cdot 1[(\xi_2 - 1)^2 + (\xi_4 - 1)^2] + 19 \cdot 8(\xi_2 - 1)(\xi_4 - 1),$$

où

$$x = \{ \xi_1, \xi_2, \xi_3, \xi_4 \}^T,$$

le minimum $f(x^*) = 0$ est obtenu pour

$$\xi_1^* = \xi_2^* = \xi_3^* = \xi_4^* = 1.$$

Le point de départ choisi était

$$\xi_1 = \xi_3 = -3; \qquad \xi_2 = \xi_4 = -1.$$

Avec les deux méthodes, en utilisant une recherche très précise du minimum d'une fonction sur une droite, on a obtenu les résultats suivants:

MÉTHODE	FLETCHER-REEVES brisé à $n + 2$	FLETCHER-REEVES sans brisure	FLETCHER-REEVES modifié brisé à $3n$	FLETCHER-REEVES modifié sans brisure
Nombres d'itérations	32	> 400	42	112

Dans tous les cas la valeur finale du coût était inférieure à 10^{-10}.

5. CONCLUSION

En somme, on voit que, en utilisant un résultat général de convergence, comme le lemme 2.2., on peut non seulement démontrer la convergence de nombreux algorithmes bien connus mais aussi voir comment modifier certains algorithmes heuristiques pour obtenir des méthodes convergentes.

ACKNOWLEDGMENT

This research was supported in part by the national Aeronotics and Space Administration under Grant No Ns G 354 (supp. 5).

REFERENCES

[1] R. FLETCHER et C. M. REEVES, « Function minimization by conjugate gradients », *The Computer Journal*, pp. 149-154 (1964).

[2] E. POLAK, « On Primal and Dual Methods for solving discrete optimal control Problems », Second Int. Conf. on Computing methods in optimization problems, San Remo, Italy, Sept. 1968.

[3] R. FLETCHER et M. J. D. POWELL, « A rapidly convergent descent method for minimization », *The Computer Journal*, vol. 6, p. 163 (1963).

[4] D. M. TOPKIS et A. F. VEINOTT (Jr), « On the convergence of some feasible direction algorithms for non-linear programming », *Siam. J. on Control*, vol. 5, n° 2, p. 268 (1967).

[5] J. W. DANIEL, « The conjugate gradient method for linear and non linear operator equations », *Siam J. Num. Anal.*, vol. 4, n° 1 (1967).

[6] M. R. HESTENES et E. STIEFEL, « Methods of conjugate gradients for solving linear systems », *J. Res. N.B.S.*, vol. 49, p. 409 (1952).

[7] T. GINSBURG, « The conjugate gradient method », *Numerische Mathematik Band* 5, Heft 2, p. 191 (1963).