synthèse sur les différents résultats des méthodes du gradient conjugué pour la minimisation des fonctions

M.Belloufi,* R.Benzine,† Y.Laskri
Analyse Numérique, optimisation et statistique – Annaba-
Université Ibn khaldoun -TIARET -
email : mbelloufi@ yahoo.fr

octobre 2010

Abstract

Dans ce travaille on va essayer de présenter une synthèse sur les différents résultats de convergence des méthodes du gradient conjugué pour la minimisation des fonctions sans contraintes. Ces méthodes seront utilisées avec une recherche linéaire inexacte . L'analyse couvre plusieurs classes de méthodes qui sont globalement convergentes pour des fonctions régulières non nécessairement convexes. Dans la première famille, ce sont certaines propriétés de la méthode de Fletcher-Reeves qui jouent un rôle crucial, tandis que la seconde famille partage avec la méthode de Polak-Ribière-Polyak une propriété importante.

Keywords : Gradient conjugué, Algorithme, Convergence globale, Recherche linéaire inexact

*Tel.: 0661426259;.
E-mail address: mbelloufi@yahoo.fr
†Tel.: 0661426259;.
E-mail address: rabenzine@yahoo.fr

1 Introduction

Soit $f : \mathbb{R}^n \to \mathbb{R}$ et (P) le problème d'optimisation sans contraintes suivant:

$$\min\{f(x) : x \in \mathbb{R}^n\} \tag{1.1}$$

où f est régulière (continûment différentiable) et g est son gradient. Notons par g_k le gradient de f au point x_k .

Rappelons que les différentes méthodes du gradient conjugué génèrent des suites $\{x_k\}$ de la forme suivante:

$$x_{k+1} = x_k + \alpha_k d_k \tag{1.2}$$

où la direction recherchée est définie par la formule de récurrence suivante:

$$d_k = \begin{cases} -g_0 & \text{si } k = 0 \\ -g_k + \beta_k d_{k-1} & \text{si } k \geq 1 \end{cases} \tag{1.3}$$

Le coéfficient β_k détermine la méthode du gradient conjugué, le pas $\alpha_k \in \mathbb{R}^+$ étant déterminé par une recherche linéaire

2 Recherches linéaires inexactes

2.1 La règle d'Armijo

La règle d'Armijo [2, 1966] impose une contrainte sur le choix de α_k suffisante pour minimiser localement $f(\alpha_k)$.

Une condition naturelle est de demander que f décroisse autant qu'une portion $\rho \in]0,1[$ de ce que ferait le modèle linéaire de f en x_k. Cela conduit à l'inégalité suivante, parfois appelée *condition d'Armijo* ou *condition de décroissance linéaire* :

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \rho\alpha_k \nabla^T f(x_k) d_k \tag{1.4}$$

2.2 La règle de Goldstein&Price

Les conditions de Goldstein & Price ([23, 1969]) suivant sont, comme on va le prouver,suffisante pour assurer la convergence sous certaines conditions et indépendamment de l'algorithme qui calcule le paramètre .

Etant données deux réels ρ et σ tels que $0 < \rho < \delta < 1$; ces conditions sont :

$$f(x_k + \alpha d_k) \leq f(x_k) + \rho\alpha \nabla^T f(x_k) d_k \tag{1.5}$$

$$f(x_k + \alpha d_k) \geq f(x_k) + \delta\alpha \nabla^T f(x_k) d_k \tag{1.6}$$

2.3 La règle de Wolfe

La conditions (2.8)-(2.9) "règle de Goldstein&Price" peuvent exclure un minimum ce qui est peut être un inconvénient.

Les conditions de wolfe ([42, 1969]) n'ont pas cet inconvénient.

Etant donnés deux réels ρ et σ tel que $0 < \rho < \sigma < 1$, ces conditions sont :

$$f(x_k + \alpha d_k) \leq f(x_k) + \rho\alpha \nabla^T f(x_k) d_k \tag{1.7}$$

$$\nabla^T f(x_k + \alpha d_k) d_k \geq \sigma \nabla^T f(x_k) d_k \tag{1.8}$$

2.4 Conditions de Wolfe fortes

Pour certains algorithmes (par exemple le gradient conjugué non linéaire, voir chapitre4) il est parfois nécéssaire d'avoir une condition plus restrictive que (2.11).

Pour cela la deuxième condition (2.11) est remplacée par
$\left|\nabla^T f(x_k + \alpha d_k) d_k\right| \leq \sigma \left|\nabla^T f(x_k) d_k\right| = -\nabla^T f(x_k) d_k$.
On aurra donc les conditions de Wolfe fortes ([43,1971]) :

$$\begin{aligned} f(x_k + \alpha d_k) &\leq f(x_k) + \rho\alpha\nabla^T f(x_k) d_k \quad (1.9) \\ \left|\nabla^T f(x_k + \alpha d_k) d_k\right| &\leq -\sigma\nabla^T f(x_k) d_k \quad (1.10) \end{aligned}$$

2.5 La règle de Wolfe relaxée

Proposée par Dai et Yuan ([11, 1996]), cette règle consiste à choisir le pas satisfaisant aux conditions :

$$\begin{aligned} f(x_k + \alpha d_k) &\leq f(x_k) + \rho\alpha\nabla^T f(x_k) d_k \quad (1.11) \\ \sigma_1\nabla^T f(x_k) d_k &\leq \nabla^T f(x_k + \alpha_k d_k) d_k \leq -\sigma_2\nabla^T f(x_k) d_k \quad (1.12) \end{aligned}$$

2.6 Recherches linéaires non monotone

2.7 la recherche lineaire nonmonotone d'Armijo

Soit $0 < \lambda_1 < \lambda_2$, $\sigma \in (0,1)$, $\delta \in (0,1)$, et que M est un entier positif; supposer que, lors de la ième itération, un procès avant le pas $\overline{\alpha_k} \in (\lambda_1, \lambda_2)$ est donné. Le nonmonotone de recherche en ligne est de choisir la première nonnegative h_k entier tel que le pas

(1.13) $$\alpha_k = \overline{\alpha}_k \sigma^{h_k}$$

satisfait

(1.14) $$f(x_k + \alpha_k d_k) \leq \underset{0\leq j\leq m(k)}{max} f(x_{k-j}) + \delta\alpha_k g_k^T d_k,$$

où

(1.15) $$m(0) = 0 \text{ et } 0 \leq m(k) \leq min[m(k-1)+1, M-1], \quad k \geq 1.$$

2.8 Recherche lineaire non-monotone de wolfe

Déterminer λ_k tels que:

(1.16) $$ff(x_k + \alpha_k d_k) \leq \underset{0\leq j\leq m(k)}{max} f(x_{k-j}) + \rho\alpha_k g_k^T d_k,$$

(1.17) $$\mid \langle g\left(x_{k+}\lambda_k d_k\right), d_k\rangle \mid \leq -\delta g_k^T d_k$$

là où 0‹ρ‹δ‹1⁄2et où l'ordre de nombre entier $\{m(k)\}$ est produit par la formule

(1.18) $$m(O) = O, m(k) \leq \min\{m(k-1)+1, M_O\},$$

à où $M_{O\rangle O}$ est un nombre entier.

3 Les différents formules pour β_k :

Si on note $y_{k-1} = g_k - g_{k-1}$, on obtient les variantes suivantes:

(1.19) $$\beta_k^{HS} = \frac{g_k^T(g_k - g_{k-1})}{d_{k-1}^T(g_k - g_{k-1})},$$ Gradient conjugué variante Hestenes-Stiefel

(1.20) $$\beta_k^{PRP} = \frac{g_k^T y_k}{\|g_{k-1}\|^2}$$ Gradient conjugué variante Polak-Ribière-Polyak

(1.21) $$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}$$ Gradient conjugué variante Fletcher Reeves

(1.22) $$\beta_k^{CD} = \frac{\|g_k\|^2}{-d_{k-1}^T g_{k-1}}$$ Gradient conjugué variante descente conjugué

(1.23) $$\beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}}$$ Gradient conjugué variante de *Dai-Yuan*

Méthodes hybrides

Les méthodes hybrides sont des combinaisons de la méthode de Dai-Yuan et de la méthode de Hestenes-Stiefel.

Ces méthodes ont été construites par Dai et Yuan ([15, 2001]). On obtient à partir des formules (3.12) et (4.6) les méthodes hybride suivantes:

1^{ere} méthode hybride

Le coefficient β_k est déterminé par la formule hybride suivante

(1.24) $$\beta_k = max\left\{-c\beta_k^{DY}, \min\left\{\beta_k^{HS}, \beta_k^{DY}\right\}\right\};$$

Où $c = \frac{1-\sigma}{1+\sigma}$

$2^{ème}$ méthode hybride

Les coefficients β_k sont déterminés par la formule suivante:

(1.25) $$\beta_k = max\left\{0, \min\left\{\beta_k^{HS}, \beta_k^{DY}\right\}\right\};$$

Où $c = \frac{1-\sigma}{1+\sigma}$

4 Synthèse des résultats de convergence des méthodes du gradient conjugué

On expose dans ce chapitre une synthèse concernant la convergence des différentes variantes du gradient conjugué avec la recherche linéaire inexacte

4.1 Résultats de Convergence du Gradient Conjugué, version Fletcher-Reves

Le premier résultat de convergence de la méthode du gradient conjugué non linéaire (version *Flecher Reeves*) avec la recherche linéaire inexacte de *Wolfe* forte a été démontré par *Al-Baali* ([1985]).

•*Touati Ahmed* et *Story* ([41, 1990]) ont généralisé ce résultat pour

$$0 \leq \beta_k \leq \beta_k^{FR}. \tag{1.26}$$

• *Gilbert* et *Nocedal* ([21, 1992]) ont généralisé ce résultat pour

$$|\beta_k| \leq \beta_k^{FR}. \tag{1.27}$$

• Récemment, *Dai et Yuan* ([10, 1996]) a montré que la méthode *FR* ets globalement convergente avec la recherche linéaire inexacte de *Wolfe relaxée.*

On donne dans ce paragraphe les principaux résultats de convergence de la méthode du gradient conjugué non linéaire avec la recherche linéaire inexacte de *Wolfe* forte [*Gilbertet et Nocedal*], aussi avec la recherche linéaire inexacte de *Wolfe relaxée* [*Dai et Yuan*].

4.2 Résultats de Convergence du Gradient Conjugué, version Polak Ribière

Polak et *Ribière* ([1969]). Il établit la convergence de cette méthode pour une fonction fortement convexe avec une recherche linéaire exacte.

Remarque 3.4. Si f n'est pas convexe, la méthode de *Polak-Ribière-Polyak* peut ne pas converger.

Remarque 3.5. *Powell* [(38, 1984)] a donné un exemple de fonction pour laquelle l'algorithme génère une suite $\{x_k\}$ dont aucun des points d'adhérence n'est stationnaire. On peut trouver des remèdes simples a ce comportement inattendu [21, 1992]. Le plus simple et apparemment le plus efficace est de prendre

$$\beta_k = (\beta_k^{PRP})^+ = \max(0, \beta_k^{PRP}), \tag{1.28}$$

ce qui revient a redémarrer l'algorithme chaque fois que $\beta_k^{PRP} < 0$.On peut montrer la convergence globale pour cette valeur de β_k et une recherche linéaire adaptée.

Les performances numériques de cette méthode sont trés semblables a celles de la méthode de *Polak-Ribière.*

4.3 Résultats de Convergence du Gradient Conjugué, version(Dai-Yuan)

Récemment *Dai* et *yuan* ([9, 1998]) ont introduit une formule pour β_k

Cette méthode possède plusieurs propriétés, par exemple elle possède la propriété de descente à chaque itération et la convergence, si le pas est déterminé par la règle de (*Wolfe faible, Armijo et Goldstein),* si f est strictement convexe ou régulière.

4.4 Résultats de Convergence du Gradient Conjugué avec la recherche lineaire non-monotone

La méthode **PR** converge si en utilisant des recherches linéaires non-monotone (**GLL)**

La méthode **HS** converge si en utilisant des recherches linéaires non-monotone (**GLL)**

4.5 Résultats de Convergence du 1^ere^ méthode hybride

Cette méthode possède plusieurs propriétés, par exemple elle possède la propriété de descente à chaque itération, la convergence au sens

$$\lim_{k\to\infty} \inf \|g_k\| = 0 \tag{1.29}$$

si le pas est déterminé par la règle de Wolfe faible.

4.6 Résultats de Convergence du $2^{\text{ème}}$ méthode hybride

Cette méthode possède plusieurs propriétés, par exemple elle possède la propriété de descente à chaque itération, la convergence au sens

$$\lim_{k\to\infty} \inf \|g_k\| = 0$$

si le pas est déterminé par la règle de Wolfe faible.

References

[1] M. Al-Baali, Descent property and global convergence of the Fletcher–Reeves method with inexact line search, IMA J. Numer. Anal. 5 (1985) 121–124.

[2] L. Armijo, Minimization of functions having Lipschits continuous partial derivatives, Pacific J. Math. 16 (1966) 1–3.

[3] Y.H. Dai,Y.Yuan, An eficient hybrid conjugate gradient method for unconstrained optimization, Ann. Oper. Res. 103 (2001) 33–47.

[4] Y.H. Dai,Y.Yuan, A nonlinear conjugate gradient method with a strong global convergence property, SIAM J. Optim. 10 (1999) 177–182.

[5] R. Fletcher, C. Reeves, Function minimization by conjugate gradients, Comput. J. 7 (1964) 149–154.

[6] A.A. Goldstein, On steepest descent, SIAM J. Control 3 (1965) 147–151.

[7] J.C. Gilbert, J. Nocedal, Global convergence properties of conjugate gradient methods for optimization, SIAM. J. Optim. 2 (1992) 21–42.

[8] M.R. Hestenes, E. Stiefel, Method of conjugate gradient for solving linear systems, J. Res. Nat. Bur. Stand. 49 (1952) 409–436.

[9] W.W. Hager, H. Zhang, A new conjugate gradient method with guaranteed descent and an efficient line search, SIAM J. Optim. 16 (2005) 170–192.

[10] E. Polak, G. Ribiére, Note sur la convergence de directions conjuguées, Rev. Francaise Infomat Recherche Operatonelle, 3e Année 16 (1969) 35–43.