DOI: 10.13656/j.cnki.gxkx.2001.01.003
广西科学 Guangxi Sciences 2001, 8 (1): 7~ 9, 12

Convergence Conditions of General Conjugate Gradient Method with Some Types of Inexact Line Searches
几类非精确线搜索下共轭梯度法的收敛条件

Liu Yun Liang Yumei
刘 云 梁玉梅*

(Dept. of Math, Yunnan Univ., Kunming, Yunnan, 650091, China)
(云南大学数学系 云南昆明 650091)

Abstract Certain convergence conditions for conjugate gradient method with some types of inexact line search are discussed. Convergence analysis of a class of methods is given as an example by applying our results.

Key words conjugate gradient method, convergence property, inexact line search, nonlinear programming

摘要 在几类非精确线搜索下讨论一般共轭梯度法的收敛条件,运用此条件,对一类新算法的收敛性进行分析.

关键词 共轭梯度法 收敛性 非精确线搜索 非线性规划
中图法分类号 O 224; O 241

1 Introduction

We consider the unconstrained nonlinear optimization problem:

$$\min_{x\in R^n} f(x). \quad (1)$$

Where $f: R^n \to R$ is continuously differentiable and its gradient is denoted by g. g_k and f_k represent $g(x_k)$, $f(x_k)$ respectively. A general conjugate gradient algorithm is given by

$$x_{k+1} = x_k + \lambda_k d_k. \quad (2)$$

$$d_k = \begin{cases} -g_k, & k = 1 \\ -g_k + \beta_k d_{k-1}, & k \geq 2. \end{cases} \quad (3)$$

Where β_k is a scalar and λ_k is a step length obtained by a line search.

The well-known formulae for β_k are Fletcher-Reeves (FR), Polak-Ribiere-Polyak (PRP), Hestenes-Stiefel (HS) and Conjugate-Descent (CD) formulae:

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \quad (4)$$

$$\beta_k^{PRP} = \frac{g_k^T(g_k - g_{k-1})}{\|g_{k-1}\|^2}, \quad (5)$$

$$\beta_k^{HS} = \frac{g_k^T(g_k - g_{k-1})}{d_{k-1}^T(g_k - g_{k-1})}, \quad (6)$$

$$\beta_k^{CD} = -\frac{\|g_k\|^2}{d_{k-1}^T g_{k-1}}. \quad (7)$$

Zoutendijk[1] proved that the FR method with exact line search was globally convergent. Al-Baali[2] extended this result to strong Wolfe line search. Powell[3] showed that the PRP method might not converge to a stationary point, and he suggested that β_k should not be less than zero, Gilbert and Nocedal[4] proved that $\beta_k = \max(\beta_k^{PRP}, 0)$ could make the method converged globally with the Wolfe line search and the sufficient descent condition ($g_k^T d_k \leq -C\|g_k\|^2, C > 0$) holding. But Grippo and Lucidi[5] showed that choosing β_k might not be the only way. Chen and Jiao[6] presented a new formula to compute the scalar β_k:

$$\beta_k = \begin{cases} \beta_k, & |g_k^T d_{k-1}| \geq \rho_1 \text{ and } \|g_k\| \cdot \|d_{k-1}\| \leq \rho_2 \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

Where

$$-\frac{\bar\sigma}{\sigma}\beta_k^{CD} \leq \beta^{(k)} \leq \frac{\bar\sigma}{\sigma}\beta_k^{CD}, \quad (9)$$

$0 < \rho_1 < \rho_2 < +\infty$, $\sigma \in (0, 1)$, $\bar\sigma \in (0, 1/2)$

In this method $\beta_k < 0$ is permitted, and in Reference [6] the global convergence is proved with the generalized Curry line search:

(A) $\lambda_k = \min\{\lambda \mid g(x_k + \lambda d_k)^T d_k = \_ g_k^T d_k, \lambda > 0\}, \_ \in (0, \sigma)$.

We call this method New-Method.

As we all known, line search method plays an im-

2000-06-01收稿.

* Dept. of Math & Info. Sci., Guangxi Univ., Nanning, Guangxi, 530004 (广西大学数学与信息科学系,广西南宁, 530004)

portant role in optimization algorithms. There are many inexact line searches such as Wolfe line search (B), strong Wolfe line search (B$'$), ideal line search (E), Generalized Curry line search (A), etc. From above discussion we know that most of the results of the convergence properties are based on the Wolfe or strong Wolfe line searches. What's happen on the other inexact line searches is the main object of this paper. In Section 3, we discuss the global convergence properties of the general algorithms (1) ~ (3) with nine types of line searches. We will show in Section 3 that Theorems 1 and 3 are very good tools for convergence analysis. In Section 4, we analyze the convergence properties of the New-Method mentioned above. It is a good example for applying our results. In Section 5, we make a further discuss on the line search (D).

2 Basic Assumptions and Definitions

We give the following basic assumptions:

(AS1) The level set $L_1 = \{x | f(x) \leqslant f(x_1)\}$ is bounded.

(AS2) In some neighborhood N of L_1, the objective function f is continuously differentiable and its gradient is Lipchitz continuous, i. e. there exists a constant $L > 0$ such that

$$\| g(x) - g(y) \| \leqslant L \| x - y \| \quad \forall x, y \in N. \tag{10}$$

Line search (A) and the following eight line searches are considered in this paper.

$$(\mathrm{B}) \begin{cases} f_{k+1} \leqslant f_k + c_1 \lambda_k g_k^T d_k, & (\mathrm{B1}) \\ g_{k+1}^T d_k \geqslant c_2 g_k^T d_k. & (\mathrm{B2}) \end{cases}$$

$$(\mathrm{B}') \begin{cases} f_{k+1} \leqslant f_k + c_1 \lambda_k g_k^T d_k, & (\mathrm{B1}') \\ | g_{k+1}^T d_k | \leqslant - c_2 g_k^T d_k. & (\mathrm{B2}') \end{cases}$$

$$(\mathrm{C}) \begin{cases} f_{k+1} \leqslant f_k, & (\mathrm{C1}) \\ f_{k+1} \leqslant f(x_k + t\lambda_k d_k) + X_k, \forall t \in (0, 1), & (\mathrm{C2}) \\ g_{k+1}^T d_k \geqslant c_2 g_k^T d_k. & (\mathrm{C3}) \end{cases}$$

$$(\mathrm{C}') \begin{cases} f_{k+1} \leqslant f_k, & (\mathrm{C1}') \\ f_{k+1} \leqslant f(x_k + t\lambda_k d_k) + X_k, \forall t \in (0, 1), & (\mathrm{C2}') \\ | g_{k+1}^T d_k | \leqslant - c_2 g_k^T d_k. & (\mathrm{C3}') \end{cases}$$

$$(\mathrm{D}) \begin{cases} f_{k+1} \leqslant f_k, & (\mathrm{D1}) \\ f_{k+1} \leqslant \min\{f(x_k + \lambda d_k) | \lambda \geqslant 0\} + X_k, & (\mathrm{D2}) \\ g_{k+1}^T d_k \geqslant c_2 g_k^T d_k. & (\mathrm{D3}) \end{cases}$$

$$(\mathrm{D}') \begin{cases} f_{k+1} \leqslant f_k, & (\mathrm{D1}') \\ f_{k+1} \leqslant \min\{f(x_k + \lambda d_k) | \lambda \geqslant 0\} + X_k, & (\mathrm{D2}') \\ | g_{k+1}^T d_k | \leqslant - c_2 g_k^T d_k. & (\mathrm{D3}') \end{cases}$$

$$(\mathrm{E}) f_{k+1} \leqslant f(x_k + \bar{\lambda}_k d_k),$$

$$(\mathrm{F}) \begin{cases} f_{k+1} \leqslant f_k + \_{1} \lambda_k g_k^T d_k, & (\mathrm{F1}) \\ f_{k+1} \geqslant f_k + \_{2} \lambda_k g_k^T d_k. & (\mathrm{F2}) \end{cases}$$

Where $0 < c_1 < c_2 < 1$, $X_k \geqslant 0$, $\sum_{k=1}^{\infty} X_k < +\infty$, $0 < \_{1} < \_{2} < 1$. $\bar{\lambda}_k$ is the smallest positive stationary point of the function $h_k(\lambda) = f(x_k + \lambda d_k)$.

The angle between $-g_k$ and d_k is denoted by θ_k. We denote

$$V_k = \cos\theta_k \cdot \| g_k \| = - \frac{g_k^T d_k}{d_k}. \tag{11}$$

3 Main Results

Lemma 1[4] For any conjugate gradient method with Formulae (1) ~ (3), if Zoutendijk condition

$$\sum_{k=1}^{\infty} V_k^2 = \sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\| d \|^2} < +\infty \tag{12}$$

holds and $\frac{\| d_k \|}{\| g_k \|}$ is bounded, then

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

Lemma 2 Suppose the assumption (AS1) and (AS2) hold, and consider any iteration with Formulae (2) ~ (3), where d_k is a descent direction and λ_k satisfies the Wolfe line search (B) or the ideal line search (E) or the Goldstein line search (F), then Zoutendijk condition (12) holds.

Proof With the line searches (B) or (E), we can see Zoutendijk[1]; with the line search (F) we can see Xu[7].

For the line searches (C) and (D), we can obtain similar results:

Theorem 1 Suppose the conditions of Lemma 2 hold with the line searches (C) or (D), then Zoutendijk condition (12) holds.

Proof By assumption (AS2) and Formula (C3) (or (D3)) we obtain

$$(c_2 - 1) g_k^T d_k \leqslant (g_{k+1} - g_k)^T d_k \leqslant \lambda_k L \| d_k \|^2.$$

$$\therefore \quad \lambda_k \geqslant \frac{(c_2 - 1) g_k^T d_k}{L \| d \|^2} \triangleq \bar{\lambda}_k > 0. \tag{13}$$

3.1 Under the line search (C)

By Formulae (C2) and (13) $\exists\, a_k \in (x_k, x_k + \bar{\lambda}_k d_k)$,

$$f_{k+1} \leqslant f(x_k + \bar{\lambda}_k d_k) + X_k$$

$$= f_k + \bar{\lambda}_k g_k^T d_k + \bar{\lambda}_k (g(a_k) - g_k)^T d_k + X_k$$

$$\leqslant f_k + \bar{\lambda}_k g_k^T d_k + \bar{\lambda}_k^2 L \| d_k \|^2 + X_k = f_k - \frac{1 - c_2}{L} V_k^2 + \frac{(1 - c_2)^2}{L} V_k^2 + X_k = f_k - \frac{(1 - c_2) c_2}{L} V_k^2 + X_k,$$

$$f_{k+1} \leqslant f_1 - \frac{(1 - c_2) c_2}{L} \sum_{i=1}^{k} V_i^2 + \sum_{i=1}^{k} X_i.$$

From (AS1) (AS2), $\{f_k\}$ is bounded.

$\therefore \quad \sum_{k=1}^{\infty} V_k^2 < +\infty$.

3.2 Under the line search (D)

Since the proof is similar to (i), we omit it here.

Remark 1 Since the line searches (B′) (C′) (D′) are stronger than (B) (C) (D), hence Theorem 1 holds with the line searches (B′) (C′) (D′).

Now we consider the line search (A).

Theroem 2 If d_k is a descent direction then the line search (A) is stronger than (B′).

Proof (i) By (A) we have

$\forall \lambda \in [0, \lambda_k], g(x_k + \lambda_k d_k)^T d_k \leqslant \_ g_k^T d_k$.

$\exists \bar{\lambda_k} \in (0, \lambda_k), f_{k+1} - f_k = \lambda_k g(x_k + \bar{\lambda_k} d_k)^T d_k \leqslant \_ \lambda_k g_k^T d_k$.

(ii) By (A) again, $|g_{k+1}^T d_k| = -\_ g_k^T d_k \leqslant -e g_k^T d_k$.

Now what we need to do is to let $c_1 = \_, c_2 = e$.

From the above lemmas and theorems, we have the following theorem.

Theorem 3 Suppose the conditions of Lemma 2 hold with one of the nine line searches (A) to (F), and $\frac{\|d_k\|}{\|g_k\|}$ is bounded, then $\lim_{k\to\infty} \|g_k\| = 0$.

4 The Convergence Analysis of the New-Method

Now we analyse the convergence property of the New-Method we have mentioned in Section 1.

Theorem 4 Suppose the assumptions (AS1) (AS2) hold, let $\{x_k\}$ be generated by the New-Method with the line searches (B′), (C′), (D′), c_2 satisfies

$$\frac{\bar{\sigma} c_2}{\sigma} < 1, \tag{14}$$

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

Proof (i) To prove $g_k^T d_k < 0$ for all k.

When $k = 1$, $g_k^T d_k = -\|g_k\|^2 < 0$. We suppose by induction that $g_k^T d_k < 0$, then,

$$g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 + U_{k+1} g_{k+1}^T d_k$$

$$\leqslant -\|g_{k+1}\|^2 + \frac{\bar{e}\|g_{k+1}\|^2}{e(-g_k^T d_k)} |g_{k+1}^T d_k|$$

$$\leqslant -\|g_{k+1}\|^2 + \frac{\bar{e} c_2}{e} \|g_{k+1}\|^2$$

$$= -\left(1 - \frac{\bar{e} c_2}{e}\right) \|g_{k+1}\|^2 < 0.$$

Hence d_k is descent direction and Formulae (B′), (C′), (D′) can be satisfied.

(ii) To prove $\frac{\|d_k\|}{\|g_k\|}$ is bounded.

If $k = 1$ or $U_k = 0$, then $\frac{\|d_k\|}{\|g_k\|} = 1$.

If $k \neq 1, U_k \neq 0$, then

$$\|d_k\|^2 = \|g_k\|^2 - 2U_k g_k^T d_{k-1} + U_k^2 \|d_{k-1}\|^2$$

$$\leqslant \|g_k\|^2 + 2\frac{\bar{e}}{e} \frac{\|g_k\|^2}{|g_{k-1}^T d_{k-1}|} |g_k^T d_{k-1}| + \frac{\bar{e}^2}{e^2} \frac{\|g_k^T\|^4}{|g_{k-1}^T d_{k-1}|^2} \|d_{k-1}\|^2$$

$$\leqslant \|g_k\|^2 + \frac{2\bar{e} c_2}{e} \|g_k\|^2 + \left(\frac{\bar{e} c_2 d_2}{e d_1}\right)^2 \|g_k\|^2.$$

$$\therefore \frac{\|d_k\|^2}{\|g_k\|^2} = 1 + \frac{2\bar{e} c_2}{e} + \left(\frac{\bar{e} c_2 d_2}{e d_1}\right)^2.$$

Remark 2 Form Theorem 2 we obtain that Theorem 4 holds with the line search (A). Therefore the result of Chen [6] can be deduced in our method.

5 Discussion

Now we make a further discussion on the line search (D). We suppose the following assumption (AS3) holds, then we show that Formulae (D1) and (D2) are enough to ensure the Zoutendijk condition.

(AS3) $f \in C^2(N)$ and there exists a constant $M > 0$ such that

$\|\nabla^2 f(x)\| \leqslant M, \forall x \in N, i, j = 1, 2, \cdots, n.$

Where $\|\nabla^2 f(x)\|$ is some norm of $\nabla^2 f$ such that

$\|\nabla^2 f(x) \cdot y\| \leqslant \|\nabla^2 f(x)\| \cdot \|y\|, \forall x, y \in N, i, j = 1, 2, \cdots, n.$

Lemma 3 $\theta(\lambda) \in C^2[0, b], \theta'(0) < 0$ then any zero point $\_$ of $\theta'(\lambda)$, $\_ \in [0, b]$, satisfies

$$\_ \geqslant -\theta'(0)/Q. \tag{15}$$

Where $|\theta''(\lambda)| \leqslant Q, \quad \forall \lambda \in [0, b]$.

Theorem 5 Suppose the assumptions (AS1), (AS3) hold and consider any iteration with the formulae (2) and (3), where d_k is a descent direction and λ_k satisfies the following line search (D″)

$$(D'')\begin{cases} f_{k+1} \leqslant f_k, & (D1'') \\ f_{k+1} \leqslant \min\{f(x_k + \lambda d_k) | \lambda \geqslant 0\} + X_k, & (D2'') \end{cases}$$

then the Zoutendijk condition (13) holds.

Proof Let $\theta(\lambda) = f(x_k + \lambda d_k) \quad \lambda \in [0, \lambda_k^*]$, then

$\theta''(\lambda) = d_k^T \nabla^2 f(x_k + \lambda d_k) d_k \leqslant M\|d_k\|^2 \quad \forall \lambda \in [0, \lambda_k^*], \forall k$, and

$\theta'(0) = g_k^T d_k < 0.$

By Lemma 3, we obtain that the zero point λ_k^* of $\theta'(\lambda)$ satisfies

$$\lambda_k^* \geqslant \bar{\lambda_k} \triangleq -\theta'(0)/(M\|d_k\|^2).$$

(下转第 12页 Continue on page 12)

使 $E|S|^{W} < \infty$，且 $E|S_n - S|^{W} \to 0$，(当 $n \to \infty$ 时)，又 $\forall X > 0$,

$$P(|S_{2^k} - S| > X) \ll E|S_{2^k} - S|^{W} \leqslant \limsup_n E|S_{2^k} - S_n|^{W} \ll \sum_{i=2^k+1}^{\infty} i^{-(1+2W)} \ll 2^{-kW}, \quad (9)$$

$$P(\max_{2^{k-1}<n<2^k} |S_n - S_{2^{k-1}}| > X) \ll E \max_{2^{k-1}<n<2^k} |S_n - S_{2^{k-1}}|^{W} \ll E \max_{2^{k-1}<n<2^k} |\sum_{i=2^{k-1}+1}^{n} Y_i i^{-\theta}|^{W} \leqslant E \sum_{i=2^{k-1}+1}^{2^k} |Y_i i^{-\theta}|^{W} \ll \sum_{i=2^{k-1}+1}^{2^k} E|Y_i i^{-\theta}|^{W} \ll 2^{-kW}, \quad (10)$$

由 (9) 和 (10) 有 $S_n \to S$, a.s.

从而定理获证．证毕．

参考文献

1 Bradley R C. On the spectral density and asymptotic normality of weakly dependent random fields. J Theoret Probab, 1992, 5: 355~374.

2 Bradley R C. Equivalent mixing conditions for random fields. Technical Report No. 336, Center for Stochastic Processes, Univ of North Carolina, Chapel Hill, 1990.

3 Bryc W, Smolenski W. Moment conditions for almost sure convergence of weakly correlated random variables. Proceeding of American Math Society, 1993, 119 (2): 629~635.

4 杨善朝．一类随机变量部分和的矩不等式及其应用．科学通报，1998, 43 (17): 1823~1827.

5 杨善朝．混合序列加权和的强收敛性．系统科学与数学，1995, 15 (3): 254~265.

(责任编辑：黎贞崇)

(上接第 9页 Continue from page 9)

$\therefore \exists t \in (0,1)$, $f(x_k + \bar{\lambda}_k d_k) = f_k + \bar{\lambda}_k g_k^T d_k + \frac{1}{2}\bar{\lambda}_k^2 d_k^T \nabla^2 f(x_k + t\bar{\lambda}_k d_k) d_k \leqslant f_k - \frac{(g_k^T d_k)^2}{M\|d_k\|^2} + \frac{1}{2}\frac{(g_k^T d_k)^2}{M^2\|d_k\|^4} M\|d_k\|^2 \leqslant f_k - \frac{1}{2M}V_k^2$.

$f_{k+1} \leqslant f(x_k + \lambda_k^* d_k) + X_k \leqslant f(x_k + \bar{\lambda}_k d_k) + X_k \leqslant f_k - \frac{1}{2M}V_k^2 + X_k$.

$\therefore f_{k+1} \leqslant f_1 - \frac{1}{2M}\sum_{i=1}^{k} V_k^2 + \sum_{i=1}^{k} X_k$.

Let $k \to \infty$, we obtain

$$\sum_{k=1}^{\infty} V_k^2 < +\infty.$$

Acknowledgements We would like to thank professor Wei Zengxin for giving us valuable suggestions on this paper and lectures about the optimization methods. We are also grateful to him and other teachers for reading the draft of this paper and giving us constructive criticism and constant encouragement.

References

1 Zoutendijk G. Nonlinear programming computational methods. In Jabadie. Integer and Nonlinear Programming. North-Holland, Amsterdam, 1970. 37~86.

2 Al-Baali M. Descent property and global convergence of the fletcher-Reeves method with inexact line search. IMA J Numer, Anal, 1985, 5: 121~124.

3 Powell M J D. Nonconvex minimization calculations and the conjugate gradient method. In: D F Griffiths. Numerical Analysis. 1984.

4 Gilbert J C, Nocedal J. Global convergence properties of conjugate gradient methods for optimization. SIAM J. Optimization, 1992, 2: 21~42.

5 Grippo L, Lucidi S. A Global convergence properties of conjugate gradient methods for optimization. SIAM, 1995, 71: 399~405.

6 Chen L P, Jiao B C. Global convergence of modified conjugate gradient methods. Mathematics in Practice and Theory, 1999, 29 (2): 134~139.

7 Xu Z S. Global convergence properties of a class of conjugate gradient methods with goldstein search. J of Math (PRC), 2000, 20 (1): 13~16.

(责任编辑：蒋汉明)