International Journal of Pure and Applied Mathematics

Volume 13 No. 4 2004, 417-426

GLOBAL CONVERGENCE PROPERTIES OF THREE-TERM CONJUGATE GRADIENT METHOD WITH NEW WOLFE-TYPE LINE SEARCH

Du Shou-qiang[1 §], Chen Yuan-yuan[2]

[1,2]Department of Mathematics
Qingdao University
Qingdao, Shandong, 266071, P.R. CHINA
[1]e-mail: dsq8933@sina.com.cn

Abstract: In this paper, a new Wolfe-type line search is proposed, and we also give the global convergence properties of three-term conjugate gradient method.

AMS Subject Classification: 65K90
Key Words: unconstrained optimization, line search, conjugate gradient method

1. Introduction

We consider the unconstrained optimization problem (P)

$$\min_{x \in R^n} f(x),$$

where $f : R^n \to R$ is continuously differentiable and its gradient at x is denoted by $g(x)$. The iterative formula for (P) is given by

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

where x_1 is given, α_k is a step-length and d_k is the search direction.

In the standard conjugate gradient method, d_k is defined by

$$d_k = \begin{cases} -g_k, & k = 1; \\ -g_k + \beta_k d_{k-1}, & k \geq 2, \end{cases} \tag{2}$$

Received: November 26, 2003

§Correspondence author

in which β_k is a scalar. Well-known formulas for β_k are called Fletcher-Reeves (FR) and Polak-Ribière-Polyak (PRP) (see [7], [14], [15]).

They are given by

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

respectively, where$\|\dot{\|}$ means the Euclidean norm. There are also many other formulas of β_k, see Dai-Yuan [4], Hestenes-Stiefel [10], Liu-Storey [12] and etc. Many authors have studied the global convergence of nonlinear conjugate gradient method for unconstrained optimization problem (see [8], [5], [16], [6], [3], [2], [11], [18], [9], [17]), nice properties of standard nonlinear conjugate gradient method are given.

In order to obtain n-step quadratic convergence rate of nonlinear conjugate gradient method and also consider the second derivative information found by the search along d_{k-1}. Beale [1], gave a three-term restart conjugate gradient method, where d_k is defined by

$$d_k = -g_k + \beta_k d_{k-1} + \gamma_k d_t\,, \tag{3}$$

where d_t is a restart direction.

McGuire and Wolfe [13] and Powell [16] made further research into the Beale three-term conjugate gradient method and good numerical results were obtained. In recent years, Deng-Li [9] and Dai-Yuan [3] made further insight into this method. In [3], the general search driection was defined as follows:

$$d_k = -g_k + \beta_k d_{k-1} + \gamma_k d_{t(p)}, \tag{4}$$

where $t(p)$ stands for the $p-th$ restart iteration, and $t(p) < k \leq t(p+1)$. This work extended Beale-Powell three-term conjugate gradient method.

But the global convergence of three-term conjugate gradient method was obtained under the strong Wolfe line search. Can we propose a new weaker Wolfe-type line search for this method, and make this algorithm more efficient?

In this paper, we propose a new Wolfe-type line search in Section 2, and prove the global convergence of the general three-term conjugate gradient method (4) in Section 3. We also can see some discussions in Section 4.

2. Preliminaries

For convenience, we assume $g_k \neq 0$ for all $k \geq 1$. We also make the following basic assumptions on the objective function.

(a) $f(x)$ is bounded below on the level set $L_0 = \{x \in R^n | f(x) \leq f(x_0)\}$.

(b) In some neighborhood U of L_0, f is continuously differentiable, and its gradient is *Lipschitz* continuous, namely, there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\|\,, \quad \forall x, y \in U.$$

The step-length α_k is computed by the following new Wolfe-type line search.

New Wolfe-type line search. Choose $\alpha_k > 0$ such that

$$f(x_k + \alpha_k d_k) - f(x_k) \leq -\rho\alpha_k^2\|d_k\|^2, \tag{5}$$

$$g(x_k + \alpha_k d_k)^T d_k \geq -2\sigma\alpha_k\|d_k\|^2, \tag{6}$$

where $0 < \rho < \sigma < 1$.

Lemma 1. *Suppose that assumption (a) holds and d_k satisfies the descent condition $g_k^T d_k < 0$. Then the new Wolfe-type line search (5), (6) is feasible.*

Proof. For any $\alpha > 0$, define $\varphi(\alpha) = f(x_k + \alpha d_k) + \rho\alpha^2\|d_k\|^2$. We have

$$\lim_{\alpha\to 0^+} \frac{\varphi(\alpha) - \varphi(0)}{\alpha} = g_k^T d_k < 0.$$

So there exists a $\alpha_k' > 0$, and when $\alpha \in (0, \alpha_k']$, we have

$$\frac{\varphi(\alpha) - \varphi(0)}{\alpha} < 0. \tag{7}$$

From assumption (a), it follows that

$$\lim_{\alpha\to +\infty} \frac{\varphi(\alpha) - \varphi(0)}{\alpha} = +\infty. \tag{8}$$

Let $\hat{\alpha}_k = \inf\{\alpha > 0 | \frac{\varphi(\alpha)-\varphi(0)}{\alpha} = 0\}$. By Intermediate Value Theorem, (7) and (8), we know that $\hat{\alpha}_k$ exists and $\hat{\alpha}_k > 0$. From the continuous of φ, we have

$$\frac{\varphi(\hat{\alpha}_k) - \varphi(0)}{\hat{\alpha}_k} = 0. \tag{9}$$

Moreover, for every $\alpha \in (0, \hat{\alpha}_k]$, we have

$$\frac{\varphi(\alpha) - \varphi(0)}{\alpha} \leq 0.$$

By Mean Value Theorem and the above inequality, we have

$$\varphi'(\theta_k\hat{\alpha}_k) = \frac{\varphi'(\theta_k\hat{\alpha}_k)\hat{\alpha}_k}{\hat{\alpha}_k} = 0,$$

where $0 < \theta_k < 1$. Therefore

$$\varphi'(\theta_k\hat{\alpha}_k) = g(x_k + \theta_k\hat{\alpha}_k d_k)^T d_k + 2\rho\theta_k\hat{\alpha}_k\|d_k\|^2 = 0\,,$$

$$g(x_k + \theta_k\hat{\alpha}_k d_k)^T d_k = -2\rho\theta_k\hat{\alpha}_k\|d_k\|^2 \geq -2\sigma\theta_k\hat{\alpha}_k\|d_k\|^2.$$

The above implications imply that $\alpha_k = \theta_k\hat{\alpha}_k$ is the step-length we want to find. □

Lemma 2. *Suppose that assumptions (a), (b) hold and $\{x_k\}$ is produced by (1). d_k is a descent direction and α_k satisfies line search (5), (6). Then*

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

Proof. From line search (6) and assumption (b), we know:

$$(g_{k+1} - g_k)^T d_k \geq -g_k^T d_k - 2\sigma\alpha_k\|d_k\|^2, \quad (g_{k+1} - g_k)^T d_k \leq L\alpha_k\|d_k\|^2.$$

By the above two inequalities, we get

$$(2\sigma + L)\alpha_k\|d_k\|^2 \geq -g_k^T d_k\,, \quad \alpha_k\|d_k\| \geq \frac{1}{2\sigma + L}(\frac{-g_k^T d_k}{\|d_k\|}).$$

Squaring both sides of the above inequality, we have

$$\alpha_k^2\|d_k\|^2 \geq (\frac{1}{2\sigma + L})^2 \frac{(g_k^T d_k)^2}{\|d_k\|^2}.$$

From (5) and assumption (a), we have

$$\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} \leq (2\sigma + L)^2 \sum_{k=1}^{\infty} \alpha_k^2\|d_k\|^2$$
$$\leq \frac{(2\sigma + L)^2}{\rho} \sum_{k=1}^{\infty} \{f(x_k) - f(x_{k+1})\} < +\infty.$$

This completes the proof. □

Lemma 3. *Suppose that $\{a_i\}$, $\{b_i\}$ are two positive sequences, if $\sum_{k\geq 1} a_k = \infty$, and*

$$b_k \leq c_1 + c_2 \sum_{i=1}^{k} a_i$$

holds for $k \geq 1$, where c_1, c_2 are two positive constants. Then

$$\sum_{k\geq 1} \frac{a_k}{b_k} = \infty.$$

3. Main Results

Theorem 1. *Suppose that assumption (a), (b) hold. Consider the general three-term conjugate gradient method (4), where α_k is obtained by line search (5), (6) and d_k is a descent direction. Then if there exist positive constants σ_1, σ_2 satisfying*

$$\|\gamma_k d_{t(p)}\| \leq \sigma_1 \|g_k\|, \tag{10}$$

$$\gamma_k g_k^T d_{t(p)} \leq \sigma_2 |\beta_k g_k^T d_{k-1}| \tag{11}$$

and

$$\sum_{k\geq 1} \frac{1}{\|d_k\|^2} = \infty, \tag{12}$$

we have

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

Proof. By contradiction, we assume that there exists a constant $\gamma > 0$ such that

$$\|g_k\| \geq \gamma \tag{13}$$

holds for all k. Combining

$$d_k = -g_k + \beta_k d_{k-1} + \gamma_k d_{t(p)}\,,$$

with (10), we have

$$\|d_k\| \geq \|\beta_k d_{k-1}\| - \| - g_k + \gamma_k d_{t(p)}\| \geq |\beta_k| \|d_{k-1}\| - (1+\sigma_1)\|g_k\|\,.$$

Hence

$$|\beta_k| \|d_{k-1}\| \leq \|d_k\| + (1+\sigma_1)\|g_k\|. \tag{14}$$

Define $\tau_k = \frac{|g_k^T d_k|}{\|d_k\|}$ and $\tau'_k = \alpha_k \|d_k\|$. From Lemma 2, assumption (a) and (5), we get

$$\tau_k \to 0\,, \quad \tau'_k \to 0 \quad (k \to +\infty). \tag{15}$$

By (4), (11) and assumption (b), we obtain

$$\begin{aligned}
\|g_k\|^2 &= -g_k^T d_k + \beta_k g_k^T d_{k-1} + \gamma_k g_k^T d_{t(p)} \\
&\leq |g_k^T d_k| + (1+\sigma_2)|\beta_k||g_k^T d_{k-1}| \\
&\leq |g_k^T d_k| + (1+\sigma_2)|\beta_k|[|(g_k - g_{k-1})^T d_{k-1}| + |g_{k-1}^T d_{k-1}|] \\
&\leq |g_k^T d_k| + (1+\sigma_2)|\beta_k|[L\alpha_{k-1}\|d_{k-1}\|^2 + |g_{k-1}^T d_{k-1}|].
\end{aligned}$$

Thus

$$\begin{aligned}
\frac{\|g_k\|^2}{\|d_k\|} &\leq \tau_k + (1+\sigma_2)|\beta_k|[\frac{L\alpha_{k-1}\|d_{k-1}\|^2}{\|d_k\|} + \tau_{k-1}\frac{\|d_{k-1}\|}{\|d_k\|}] \\
&= \tau_k + (1+\sigma_2)|\beta_k|\frac{\|d_{k-1}\|}{\|d_k\|}[L\tau'_{k-1} + \tau_{k-1}].
\end{aligned}$$

From (14), we get

$$\begin{aligned}
\frac{\|g_k\|^2}{\|d_k\|} &\leq \tau_k + (1+\sigma_2)[L\tau'_{k-1} + \tau_{k-1}][1 + (1+\sigma_1)\frac{\|g_k\|}{\|d_k\|}] \\
&\leq \tau_k + (1+\sigma_2)[L\tau'_{k-1} + \tau_{k-1}] + \frac{(1+\sigma_1)(1+\sigma_2)}{\gamma}[L\tau'_{k-1} + \tau_{k-1}]\frac{\|g_k\|^2}{\|d_k\|},
\end{aligned}$$

i.e,

$$\begin{aligned}
&(1 - \frac{(1+\sigma_1)(1+\sigma_2)}{\gamma}[L\tau'_{k-1} + \tau_{k-1}])\frac{\|g_k\|^2}{\|d_k\|} \\
&\qquad \leq \tau_k + (1+\sigma_2)[L\tau'_{k-1} + \tau_{k-1}],
\end{aligned} \tag{16}$$

which, together with (15), implies that there exists a index k_0 , such that for $k \geq k_0$

$$\begin{aligned}
\frac{1}{2}\frac{\|g_k\|^2}{\|d_k\|} &\leq \tau_k + (1+\sigma_2)[L\tau'_{k-1} + \tau_{k-1}] \\
&= \tau_k + (1+\sigma_2)\tau_{k-1} + L(1+\sigma_2)\tau'_{k-1}.
\end{aligned}$$

By Lemma 2, assumption (a) and (5), we have

$$\sum_{k\geq 1} \tau_k^2 < +\infty, \quad \sum_{k\geq 1} \tau_{k-1}^2 < +\infty, \quad \sum_{k\geq 1} \tau_{k-1}'^2 < +\infty.$$

From (13) and the above inequalities, we know

$$\gamma^4 \sum_{k\geq 1} \frac{1}{\|d_k\|^2} \leq \sum_{k\geq 1} \frac{\|g_k\|^4}{\|d_k\|^2} < +\infty,$$

which contradicts (12). We conclude that

$$\liminf_{k\to\infty} \|g_k\| = 0. \qquad \square$$

Theorem 2. *In Theorem 1, if substitute (12) into the following condition*

$$\sum_{k\geq 1} \frac{|g_k^T d_k|^r}{\|d_k\|^2} = +\infty\,, \tag{17}$$

for every $r \in [0,2]$ holds, we have

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

Proof. For every $r \in [0,2]$, if $|g_k^T d_k| > 1$, we have $|g_k^T d_k|^r \leq (g_k^T d_k)^2$, thus

$$|g_k^T d_k|^r \leq 1 + (g_k^T d_k)^2.$$

So we get

$$\sum_{k\geq 1} \frac{1}{\|d_k\|^2} \geq \sum_{k\geq 1} \frac{|g_k^T d_k|^r}{\|d_k\|^2} - \sum_{k\geq 1} \frac{|g_k^T d_k|^2}{\|d_k\|^2}.$$

It follows from Lemma 2, (17) that

$$\sum_{k\geq 1} \frac{1}{\|d_k\|^2} = +\infty.$$

By theorem1, we obtain that

$$\liminf_{k\to\infty} \|g_k\| = 0. \qquad \square$$

Theorem 3. *In Theorem 1, if substitute (12) into the following condition*

$$\beta_k \gamma_k d_k^T d_{t(p)} \leq 0 \tag{18}$$

holds for $p \geq 1$ and $t(p) < k \leq t(p+1)$, and exists infinite sequence $\{k_i\}$, such that

$$\prod_{j=l}^{k_i} |\beta_i| \leq c_1 \tag{19}$$

holds for $j \geq 1$ and $j \leq k_i$, where c_1 is positive constant. Then

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

Proof. Take from of

$$d_k + g_k = \beta_k d_{k-1} + \gamma_k d_{t(p)},$$

and (10), (18), we have

$$\begin{aligned}\|d_k\|^2 = \beta_k^2\|d_{k-1}\|^2 - \|g_k\|^2 - 2g_k^T d_k + \gamma_k^2\|d_{t(p)}\|^2 + 2\gamma_k\beta_k d_{k-1}^T d_{t(p)} \\ \leq (\sigma_1^2 - 1)\|g_k\|^2 - 2g_k^T d_k + \beta_k^2\|d_{k-1}\|^2.\end{aligned}$$

Deduce the above inequality, we know

$$\|d_k\|^2 \leq (\sigma_1^2 - 1)\|g_k\|^2 - 2g_k^T d_k + \sum_{j=2}^{k} \Pi_{i=j}^k \beta_i^2[(\sigma_1^2 - 1)\|g_j\|^2 - 2g_j^T d_j].$$

By (19) and the above inequality, we obtain that

$$\|d_{k_i}\|^2 \leq (1 + c_1^2)\sum_{j=1}^{k_i} |(\sigma_1^2 - 1)\|g_j\|^2 - 2g_j^T d_j|. \tag{20}$$

We split the proof into two cases.

Case 1. If $\liminf_{i\to\infty} \|d_{k_i}\| < \infty$, then

$$\sum_{k\geq 1} \frac{1}{\|d_k\|^2} = +\infty.$$

From Theorem 1, we get

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

Case 2. When $\lim_{i\to\infty} \|d_{k_i}\| = +\infty$, (20) and Lemma 3 imply that

$$\lim_{i\to\infty} \sum_{j=1}^{k_i} \frac{|\sigma_1^2 - 1|\|g_j\|^2 - 2|g_j^T d_j|}{\|d_j\|^2} = +\infty.$$

So one of the following two equality holds.

$$\lim_{i\to\infty} \sum_{j=1}^{k_i} \frac{\|g_j\|^2}{\|d_j\|^2} = +\infty, \tag{21}$$

$$\lim_{i\to\infty}\sum_{j=1}^{k_i}\frac{|g_j^T d_j|}{\|d_j\|^2}=+\infty. \tag{22}$$

When (21) holds, by (13), we have

$$\lim_{i\to\infty}\gamma^2\sum_{j=1}^{k_i}\frac{1}{\|d_j\|^2}=+\infty.$$

We arrive at the truth of the theorem by Theorem 1. When (22) holds, we also arrive at the truth of the theorem by Theorem 2. So, the proof is completed. □

4. Discussion

We have further studied the general three-term conjugate gradient method using the new Wolfe-type line search in the absence of the sufficient descent condition. We believe that our result will lead to a better understanding of the global convergence of the general three-term conjugate gradient method and provide a unified tool in the practical analysis of this method for a large scale unconstrained optimization problem.

References

[1] E.M.L. Beale, *A Derivative of Conjugate Gradients Lootsma F A*, Numerical Methods for Nonlinear Optimization, London, Academic Press (1972).

[2] Y.H. Dai, Y.X. Yuan, Convergence properties of the Beale-Powell restart algorithm, *Sciences in China (Series A)*, **41** (1998), 1142-1150.

[3] Y.H. Dai, Y.X. Yuan, Convergence of three term conjugate gradient methods, *Mathematical Numerica Sinica*, **3** (1999), 355-362.

[4] Y.H. Dai, Y.X. Yuan, A nonlinear conjugate gradient method with nice global convergence properties, *SIAM J. Optim.*, **10** (1999), 177-182.

[5] Y.H. Dai, J.Y. Han, G.H. Liu, D.F. Sun, H.X. Yin, Y.X. Yuan, Convergence properties of nonlinear conjugate gradient methods, *SIAM J. Optim.*, **10** (2000), 345-358.

[6] N.Y. Deng, Z. Li, Global convergence of three terms conjugate gradient methods, *Optim. Method and Sofeware*, **4** (1995), 273-282.

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

[8] J.C. Gilbert, J. Nocedal, Global convergence properties of conjugate gradient methods for optimization, *SIAM J. Optim.*, **2** (1992), 21-24.

[9] L. Grippo, S. Lucidi, A global convergent version of the Polak-Ribiere conjugate gradient method, *Math. Prog.*, **78** (1997), 375-391.

[10] M.R. Hestenes, E.L. Stiefel, Methods of conjugate gradient for solving linear systems, *J. Res. Nat. Bur. Standards sect.*, **49** (1952), 409-436.

[11] Y.F. Hu, C. Storey, Global convergence result for conjugate gradient methods, *J. Optim. Theory Appl.*, **49** (1991), 409-436.

[12] Y. Liu, C. Storey, Efficient generalized conjugate gradient algorithms, Part 1: Theory, *J. Optim. theory and Appl.*, **69** (1991), 129-137.

[13] M.F. Mcguire, P. Wolfe, *Evaluating a Restart Procedure for Conjugate Gradients*, Report RC-4382, IBM Research Center, Yorktown Heights (1973).

[14] E. Polak, G. Ribiere, Note sur la convergence de directions conjugees, *Rev Francaise Informat Recher che Opertionelle*, **16** (1969), 35-43.

[15] B.T. Polyak, The conjugate gradient method in extreme problems, *USSR Comp. Math. and Math. Phys.*, **9** (1969), 94-112.

[16] M.J.D. Powell, Restart procedures of the conjugate gradient method, *Math. Prog.*, **2** (1977), 241-254.

[17] C.Y. Wang, Y.Z. Zhang, Global convergence properties of s-related conjugate gradient methods, *Chinese Science Bulletin*, **43** (1998), 1959-1965.

[18] G. Zoutendijk, Nonlinear programming, computational methods, In: *Integer and Nonlinear Programming* (Ed. J. Abodie), North-Holland, Amsterdam (1970).