Journal of Computational Mathematics, Vol.20, No.6, 2002, ??–??.

A NOTE ON THE NONLINEAR CONJUGATE GRADIENT METHOD *1)

Y.H. Dai Y. Yuan

(*State Key Laboratory of Scientific and Engineering Computing, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, P. O. Box 2719, Beijing 100080, China*)

Abstract

The conjugate gradient method for unconstrained optimization problems varies with a scalar. In this note, a general condition concerning the scalar is given, which ensures the global convergence of the method in the case of strong Wolfe line searches. It is also discussed how to use the result to obtain the convergence of the famous Fletcher-Reeves, and Polak-Ribiére-Polyak conjugate gradient methods. That the condition cannot be relaxed in some sense is mentioned.

Key words: unconstrained optimization, conjugate gradient, line search, global convergence.

1. Introduction

The conjugate gradient method is highly useful for minimizing a smooth function of n variables,

$$\min_{x\in R^n} f(x), \tag{1.1}$$

especially when n is large. It has the following form

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

$$d_k = \begin{cases} -g_k, & \text{for } k=1; \\ -g_k + \beta_k d_{k-1}, & \text{for } k \ge 2, \end{cases} \tag{1.3}$$

where $g_k = \nabla f(x_k)$, α_k is a stepsize obtained by a one-dimensional line search and β_k is a scalar. Because α_k is not the exact one-dimensional minimizer in practice and f is not a quadratic, many formulas have been proposed to compute the scalar β_k. Two well-known formulas for β_k are called the Fletcher-Reeves (FR), and Polak-Ribiére-Polyak (PRP) formulas (see [8, 16, 17]). They are given by

$$\beta_k^{FR} = \|g_k\|^2/\|g_{k-1}\|^2 \tag{1.4}$$

and

$$\beta_k^{PRP} = g_k^T(g_k - g_{k-1})/\|g_{k-1}\|^2 \tag{1.5}$$

respectively, where $\|\cdot\|$ means the Euclidean norm. See Dai & Yuan [2], Daniel [6], Fletcher [7], Hestenes & Stiefel [11], Liu & Storey [13] *etc.* for other formulas of β_k.

In recent years, many authors studied the nonlinear conjugate gradient method especially from the angle of global convergence. Because its properties can be very different with the choice of β_k (see Powell [14]), the nonlinear conjugate gradient method was often analyzed

* Received.

1) Research partially supported by Chinese NSF grants 19525101, 19731010 and 19801033.

individually, for example, see Al-Baali [1], Dai & Yuan [3, 4], Gilbert & Nocedal [9], Grippo & Lucidi [10], Liu *et al* [12], Powell [15], Qi [19] and Touati-Ahmed & Storey [20]. Dai *et al* [5] studied the general conjugate gradient method in the absence of the sufficient descent condition and proposed a sufficient condition ensuring the global convergence (see also Lemma 2.3). Since the nonlinear conjugate gradient method varies with the choice of β_k, we wonder what condition on β_k guarantees the convergence of the method.

This paper is organized as follows. After giving some preliminaries in the next section, we will prove in Section 3 that a mild condition on β_k results in global convergence of the nonlinear conjugate gradient method in the case of strong Wolfe line searches. Section 4 discusses how to use the result to obtain the convergence of the famous FR, and PRP conjugate gradient method. In the last section, it is mentioned that the condition on β_k cannot be relaxed in some sense.

2. Preliminaries

For convenience, we assume that $g_k \neq 0$ for all k, for otherwise a stationary point has been found. We give the following basic assumption on the objective function.

Assumption 2.1. *(i) The level set $\mathcal{L} = \{x \in \Re^n : f(x) \le f(x_1)\}$ is bounded; (ii) In some neighborhood $\mathcal{N}$ of $\mathcal{L}$, f is differentiable and its gradient g is Lipschitz continuous, namely, there exists a constant $L > 0$ such that*

$$\|g(x) - g(\tilde{x})\| \le L\|x - \tilde{x}\|, \qquad \text{for any } x,\ \tilde{x} \in \mathcal{N}. \tag{2.1}$$

Assumption 2.1 implies that there exists a constant $\overline{\gamma}$ such that

$$\|g(x)\| \le \overline{\gamma}, \qquad \text{for all } x \in \mathcal{L}. \tag{2.2}$$

The stepsize α_k in () is computed by carrying out certain line searches. The Wolfe line search [21] is to find a positive stepsize α_k such that

$$f(x_k + \alpha_k d_k) - f(x_k) \le \delta \alpha_k g_k^T d_k, \tag{2.3}$$

$$g(x_k + \alpha_k d_k)^T d_k \ge \sigma g_k^T d_k, \tag{2.4}$$

where $0 < \delta < \sigma < 1$. Under Assumption 2.1 on f, we state the following result, which was essentially obtained by Zoutendijk [23] and Wolfe [21, 22].

Lemma 2.2. *Suppose that x_1 is a starting point for which Assumption 2.1 holds. Consider any iterative method (), where d_k is a descent direction and α_k is computed by the standard Wolfe line search. Then*

$$\sum_{k \ge 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty. \tag{2.5}$$

In the convergence analysis and implementation of conjugate gradient methods, the stepsize α_k is often computed by the strong Wolfe line search, namely, (2.3) and

$$|g(x_k + \alpha_k d_k)^T d_k| \le -\sigma g_k^T d_k, \tag{2.6}$$

where also $0 < \delta < \sigma < 1$. Dai *et al* [5] proved the following general convergence result for any conjugate gradient method using the strong Wolfe line search.

Lemma 2.3. *Suppose that x_1 is a starting point for which Assumption 2.1 holds. Consider any method ()–(), where d_k is a descent direction and α_k is computed by the strong Wolfe line search. If*

$$\sum_{k\geq 1} \frac{\|g_k\|^t}{\|d_k\|^2} = \infty \tag{2.7}$$

for some $t \in [0,4]$, we have that

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{2.8}$$

3. Analyses of general conjugate gradient method

In this section, we will give a general condition on the scalar β_k, which ensures the global convergence of the nonlinear conjugate gradient method in the case of strong Wolfe line searches. We need the following lemma.

Lemma 3.1. *Suppose that x_1 is a starting point for which Assumption 2.1 holds. Consider any method ()–(), where d_k is a descent direction and α_k is computed by the strong Wolfe line search. If*

$$\sum_{k\geq 1} \frac{|g_k^T d_k|^r}{\|d_k\|^2} = \infty \tag{3.1}$$

for some $r \in [0,2]$, the method converges in the sense that (2.8) is true.

Proof. For any $r \in [0,2]$, if $|g_k^T d_k| > 1$, then we have that $|g_k^T d_k|^r \leq (g_k^T d_k)^2$. It follows that

$$|g_k^T d_k|^r \leq 1 + (g_k^T d_k)^2 \tag{3.2}$$

and consequently

$$\sum_{k\geq 1} \frac{1}{\|d_k\|^2} \geq \sum_{k\geq 1} \frac{|g_k^T d_k|^r}{\|d_k\|^2} - \sum_{k\geq 1} \frac{|g_k^T d_k|^2}{\|d_k\|^2}. \tag{3.3}$$

Therefore by Lemma 2.2, (3.1) and (3.3), we know that (2.7) holds with $t = 0$. Due to Lemma 2.3, (2.8) is true. □

The above lemma provides another sufficient condition ensuring the convergence of the nonlinear conjugate gradient method. Combining Lemmas 2.3 and 3.1, we now can prove a more general result.

Corollary 3.2. *Suppose that x_1 is a starting point for which Assumption 2.1 holds. Consider any method ()–(), where d_k is a descent direction and α_k is computed by the strong Wolfe line search. If*

$$\sum_{k\geq 1} \frac{\|g_k\|^t \, |g_k^T d_k|^r}{\|d_k\|^2} = \infty \tag{3.4}$$

for some positive constants t and r satisfying $t + 2r \leq 4$, the method converges in the sense that (2.8) is true.

Proof. Assume that the statement is not true, we have by Lemma 2.3 that

$$\sum_{k\geq 1} \frac{\|g_k\|^{t+2r}}{\|d_k\|^2} < \infty. \tag{3.5}$$

For any $t, r > 0$, if $|g_k^T d_k| \le \|g_k\|^2$, we have that $\|g_k\|^t\, |g_k^T d_k|^r \le \|g_k\|^{t+2r}$; otherwise we have $\|g_k\|^t\, |g_k^T d_k|^r \le |g_k^T d_k|^{\frac{t}{2}+r}$. Therefore the following relation always holds:

$$\|g_k\|^t\, |g_k^T d_k|^r \le \|g_k\|^{t+2r} + |g_k^T d_k|^{\frac{t}{2}+r}. \tag{3.6}$$

By (3.4) and (3.5), we obtain

$$\sum_{k\ge 1} \frac{|g_k^T d_k|^{\frac{t}{2}+r}}{\|d_k\|^2} = \infty. \tag{3.7}$$

Due to Lemma 3.1 and the fact that $\frac{t}{2} + r \in [0, 2]$, we see that (2.8) holds. Thus we obtain a contradiction. This completes our proof. $\square$

In the following, we give a general condition on the scalar β_k and show that such a condition can ensure the global convergence of the conjugate gradient method in the case of strong Wolfe line searches. Note that the sufficient descent condition, namely,

$$g_k^T d_k \le -c\|g_k\|^2 \tag{3.8}$$

for all $k \ge 1$ and some constant $c > 0$, is not revoked in Theorem 3.3.

Theorem 3.3. *Suppose that x_1 is a starting point for which Assumption 2.1 holds. Consider any method ()–(), where d_k is a descent direction and α_k is computed by the strong Wolfe line search. If there exist an infinite subsequence $\{k_i\}$ and a positive constant c, such that the values of β_j satisfies*

$$\prod_{j=l}^{k_i} |\beta_j| \le c \tag{3.9}$$

for any $i \ge 1$ and $l \le k_i$, the method converges in the sense that (2.8) is true.

Proof. For $k \ge 2$, () implies that

$$d_k + g_k = \beta_k d_{k-1}. \tag{3.10}$$

Squaring both sides of the above relation, we get that

$$\|d_k\|^2 \le -2g_k^T d_k + \beta_k^2 \|d_{k-1}\|^2. \tag{3.11}$$

Using (3.11) recursively, we obtain

$$\|d_k\|^2 \le -2g_k^T d_k - 2\sum_{j=2}^{k} \prod_{i=j}^{k} \beta_i^2 g_{j-1}^T d_{j-1}, \tag{3.12}$$

which, with condition (3.9), shows that

$$\|d_{k_i}\|^2 \le (1 + c^2) \sum_{j=1}^{k_i} (-2g_j^T d_j). \tag{3.13}$$

If

$$\liminf_{i\to\infty} \|d_{k_i}\| < \infty, \tag{3.14}$$

then $\sum_{k\ge 1} \frac{1}{\|d_k\|^2} = \infty$ and hence by Lemma 2.3, (2.8) holds. Otherwise, the following relation holds

$$\lim_{i\to\infty} \|d_{k_i}\| = \infty. \tag{3.15}$$

In this case, it follows from (3.13) and (3.15) that

$$\lim_{i\to\infty}\sum_{j=1}^{k_i}(-g_j^T d_j)=\infty. \tag{3.16}$$

Using relations (3.13) and (3.16) and Lemma 6 in Pu & Yu [18], we can obtain that

$$\lim_{i\to\infty}\sum_{j=1}^{k_i}\frac{-g_j^T d_j}{\|d_j\|^2}=\infty. \tag{3.17}$$

Therefore by Lemma 3.1, we also have (2.8). □

4. Convergence of FR and PRP methods

In this section, we briefly discuss how to use Theorem 3.3 to obtain the global convergence of the FR, and PRP conjugate gradient methods. First, we state the following lemma for the FR method, whose proof can be found in Al-Baali [1].

Lemma 4.1. *Suppose that x_1 is a starting point for which Assumption 2.1 holds. Consider the FR method (), () and (1.4). If the stepsize α_k is computed by the strong Wolfe line search (2.3) and (2.6) with $\sigma \le 1/2$, then for all $k \ge 1$,*

$$-2\|g_k\|^2 \le g_k^T d_k < 0. \tag{4.1}$$

Theorem 4.2. *Suppose that x_1 is a starting point for which Assumption 2.1 holds. Consider the FR method (), () and (1.4). If the stepsize α_k is computed by the strong Wolfe line search (2.3) and (2.6) with $\sigma \le 1/2$, then the method converges in the sense that (2.8) is true.*

Proof. ¿From Lemma 4.1, we know that each d_k is a descent direction. Assume that the convergence relation (2.8) does not hold. Then there exists a positive constant γ such that

$$\|g_k\| \ge \gamma, \quad \text{for all } k. \tag{4.2}$$

By the definition (1.4) of β_k^{FR}, (2.2) and (4.2), we obtain

$$\prod_{j=l}^{k}\beta_j^{FR}=\frac{\|g_k\|^2}{\|g_{l-1}\|^2}\le\frac{\bar\gamma^2}{\gamma^2}\triangleq c. \tag{4.3}$$

Due to Theorem 3.3, (2.8) holds. Therefore we obtain a contradiction. This completes our proof. □

For the PRP method, Powell [15] constructed counter-examples showing that the method may cycle without approaching any solution point. Nevertheless, Gilbert & Nocedal [9] proved that a modification of setting $\beta_k=\max\{\beta_k^{PRP},0\}$ results in global convergence. To give another proof to the result by Theorem 3.3, we draw the following lemma from [9].

Lemma 4.3. *Suppose that x_1 is a starting point for which Assumption 2.1 holds. Consider any method ()–() with $\beta_k \ge 0$, where α_k satisfies (2.3), (2.6) and $g_{k+1}^T d_{k+1} < 0$. Denote $s_k = x_{k+1} - x_k$ and*

$$\mathcal{K}_{k,\Delta}^{\lambda}=\{i \text{ is an integer} : k\le i\le k+\Delta-1, \|s_{i-1}\|>\lambda\}. \tag{4.4}$$

Let $|\mathcal{K}^{\lambda}_{k,\Delta}|$ be the number of elements of $\mathcal{K}^{\lambda}_{k,\Delta}$. If (4.2) holds, then for any $\lambda > 0$, there exist an integer Δ and an index k_0, such that for any $k \geq k_0$,

$$|\mathcal{K}^{\lambda}_{k,\Delta}| \leq \frac{\Delta}{2}. \tag{4.5}$$

Theorem 4.4. *Suppose that x_1 is a starting point for which Assumption 2.1 holds. Consider the method ()–() with $\beta_k = \max\{\beta_k^{PRP}, 0\}$, where α_k satisfies (2.3), (2.6) and $g_{k+1}^T d_{k+1} < 0$. Then the method converges in the sense that (2.8) is true.*

Proof. Assume that (4.2) holds. For

$$\lambda = \frac{\gamma^4}{2L\bar{\gamma}^3}, \tag{4.6}$$

Lemma 4.3 gives an integer Δ and an index k_0 such that (4.5) holds for any $k \geq k_0$. It follows that for any integers m and $k \geq k_0$,

$$|\mathcal{K}^{\lambda}_{k,m\Delta}| \leq \frac{m\Delta}{2}. \tag{4.7}$$

For any $k \leq i \leq k + m\Delta - 1$, if $i \notin \mathcal{K}^{\lambda}_{k,m\Delta}$, we have that $\|s_{i-1}\| \leq \lambda$ and hence by (2.1), (2.2) and (4.2),

$$|\beta_i| \leq |\beta_i^{PRP}| \leq \frac{L\bar{\gamma}\|s_{i-1}\|}{\gamma^2} \leq \frac{\gamma^2}{2\bar{\gamma}^2}. \tag{4.8}$$

When $i \in \mathcal{K}^{\lambda}_{k,m\Delta}$, by (2.2) and (4.2), we have that

$$|\beta_i| \leq |\beta_i^{PRP}| \leq \frac{\|g_i\|(\|g_i\| + \|g_{i-1}\|)}{\|g_{i-1}\|} \leq \frac{2\bar{\gamma}^2}{\gamma^2}. \tag{4.9}$$

Therefore if we pick

$$c = \left(\frac{2\bar{\gamma}^2}{\gamma^2}\right)^{\max\{k_0, \Delta\}}, \tag{4.10}$$

for any integers m and $l \leq m\Delta$, we can obtain from (4.7), (4.8) and (4.9) that

$$\prod_{j=l}^{m\Delta} |\beta_j| \leq c. \tag{4.11}$$

Due to Theorem 3.3, (2.8) is true. Therefore we obtain a contradiction. This completes our proof. □

5. Discussion

We have further studied the general conjugate gradient method using the strong Wolfe line search and provided a condition on the scalar β_k, namely, (3.9), which guarantees the global convergence of the method in the absence of the sufficient descent condition. Using the result, intuitive proofs have been given to the convergence of the famous FR, and PRP conjugate gradient methods. We believe that our result will lead to a better understanding of already-existing convergence results in the conjugate gradient field and provide a unified tool in the convergence analysis of the conjugate gradient method.

Let us now consider the $n = 2, m = 8$ example of Powell [15]. Assuming that the line search is exact, Powell constructed a two-dimensional function such that the PRP method will cycle near eight points, none of which is a solution point. Direct calculations show that in the example, the limit of $\{\beta_{4j+i} : i = 1, 2, 3, 4\}$ with j is

$$-\frac{2}{9}, \quad 2, \quad \frac{3}{4}, \quad 6$$

respectively. It follows that

$$\lim_{j\to\infty} \prod_{i=1}^{4} |\beta_{4j+i}| = 2, \tag{5.1}$$

which implies that condition (3.9) is not satisfied. Therefore condition (3.9) cannot be relaxed in certain sense in ensuring the convergence of the nonlinear conjugate gradient method.

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), pp. 121-124.

[2] Y. H. Dai and Y. Yuan, *A Nonlinear Conjugate Gradient Method with Nice Global Convergence Properties*, Research report ICM-95-038, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 1995.

[3] Y. H. Dai and Y. Yuan, *Convergence properties of the Fletcher-Reeves method*, IMA J. Numer. Anal. Vol. **16** No. 2 (1996), 155-164.

[4] Y. H. Dai and Y. Yuan, *Convergence properties of the conjugate descent method*, Mathematical Advances Vol. **25** No. 6 (1996), 552-562.

[5] Y. H. Dai, J. Y. Han, G. H. Liu, D. F. Sun, H. X. Yin, and Y. Yuan, *Convergence properties of nonlinear conjugate gradient methods*, 1998. (accepted by SIAM Journal of Optimization)

[6] J. W. Daniel, *The conjugate gradient method for linear and nonlinear operator equations*, SIAM J. Numer. Anal., **4** (1967), 10-26.

[7] R. Fletcher, *Practical Methods of Optimization vol. 1: Unconstrained optimization,* John Wiley & Sons (New York), 1987.

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

[9] J. C. Gilbert and J. Nocedal, *Global convergence properties of conjugate gradient methods for optimization*, SIAM. J. Optimization. Vol. **2** No. 1 (1992), pp. 21-42.

[10] L. Grippo and S. Lucidi, *A globally convergent version of the Polak-Ribière conjugate gradient method,* Math. Prog. 78 (1997), pp. 375-391.

[11] M. R. Hestenes and E. L. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Res. Nat. Bur. Standards Sect. **5**, 49 (1952), 409-436.

[12] G. H. Liu, J. Y. Han and H. X. Yin, *Global convergence of the Fletcher-Reeves algorithm with an inexact line search*, Report, Institute of Applied Mathematics, Chinese Academy of Sciences, 1993.

[13] Y. Liu and C. Storey, *Efficient Generalized Conjugate Gradient Algorithms, Part 1: Theory*, Journal of Optimization Theory and Applications, Vol. **69** (1991), 129-137.

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

[15] M. J. D. Powell, *Nonconvex minimization calculations and the conjugate gradient method*, in: Lecture Notes in Mathematics vol. **1066**, Springer-Verlag (Berlin) (1984), pp. 122-141.

[16] E. Polak and G. Ribière, *Note sur la convergence de directions conjugées*, Rev. Francaise Informat Recherche Opertionelle, 3e Année **16** (1969), pp. 35-43.

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

[18] D. Pu and W. Yu, *On the convergence properties of the DFP algorithms*, Annals of Operations Research 24 (1990) 175-184.

[19] H. D. Qi, J. Y. Han and G. H. Liu, *A modified Hestenes-Stiefel conjugate gradient algorithm*, Chinese Annals of Mathematics 17A : 3 (1996), 277-284. (in Chinese)

[20] D. Touati-Ahmed and C. Storey, *Efficient hybrid conjugate gradient techniques*, J. Optimization Theory Appl. **64** (1990) 379-397.

[21] P. Wolfe, *Convergence conditions for ascent methods*, SIAM Review **11** (1969), pp. 226-235.

[22] P. Wolfe, *Convergence conditions for ascent methods. II: Some corrections*, SIAM Review **13** (1969), pp. 185-188.

[23] G. Zoutendijk, *Nonlinear Programming, Computational Methods*, in: Integer and Nonlinear Programming (J. Abadie, ed.), North-Holland (Amsterdam), 1970, pp. 37-86.