BULLETIN OF THE BRAZILIAN MATHEMATICAL SOCIETY

Bull Braz Math Soc, New Series 36(2), 197-204

CG Global Convergence Properties with Goldstein Linesearch*

Jian Wang and Xuebin Chi

Abstract. This paper explores the convergence of nonlinear conjugate gradient methods with Goldstein line search without regular restarts. Under this line search, global convergence for a subsequence is given for the famous conjugate gradient methods, Fletcher-Reeves method. The same result can be obtained for Polak-Ribiére-Polyak method and others.

Keywords: unconstrained optimization, conjugate gradient method, Goldstein conditions, line search, global convergence.

Mathematical subject classification: 65K10, 74P99, 78M50, 80M50.

1 Introduction

We consider the global convergence of nonlinear conjugate gradient methods for a smooth, nonlinear, and unconstrained function of n variables

$$\min f(x) \tag{1.1}$$

where $f : R^n \to R^1$ is continuously differentiable and its gradient is denoted by g. We consider only the case where the methods are implemented without regular restarts. The iterative formula is given by

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

where α_k is a step-length and d_k is the search direction defined by

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

Received 14 September 2004.

*This work was partially supported by National Hitech Program (863,2002AA104540) and National Natural Science Foundation of China (No.60373060).

where β_k is a scalar and g_k denotes $g(x_k)$.

The best-known formulas for β_k are called the Fletcher-Reeves (FR), the Polak-Ribiére(PR) and Hestenes-Stiefel(HS) formulas and are given by:

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

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

$$\beta_k^{HS} = \frac{g_k^T(g_k - g_{k-1})}{d_{k-1}^T(g_k - g_{k-1})} \tag{1.6}$$

where $\|\cdot\|$ means the Euclidean norm. The conjugate gradient methods (CG) are available for large-scale unconstrained optimization because their storage are relatively small. Numerical results showed that if f is easy to be computed and if its dimension n is very large, the CG is still the best choice for solving (1.1).

In the already-existing convergence analysis and implementations of the CG, the strong Wolfe conditions, namely,

$$f(x_k + \alpha_k d_k) \le f(x_k) + c_1 \alpha_k g_k^T d_k \tag{1.7a}$$

$$|g(x_k + \alpha_k d_k)^T d_k| \le c_2 |g_k^T d_k| \tag{1.7b}$$

where $0 < c_1 < c_2 < 1$, are often imposed on the line search. Since Al-Baali [1] first extended the globally convergent property of nonlinear CG to inexact line search using the strong Wolfe conditions, some important global convergence results for CG have been given. But there is little result about the CG using the Goldstein conditions, namely,

$$f(x_k) + (1-c)\alpha_k g_k^T d_k \le f(x_k + \alpha_k d_k) \tag{1.8a}$$

$$f(x_k + \alpha_k d_k) \le f(x_k) + c\alpha_k g_k^T d_k \tag{1.8b}$$

where $0 < c < \frac{1}{2}$. Gilbert and Nocedal's analysis [2] on the CG was greatly different from that used by Al-Baali [1]. Dai and Yuan [3] showed that the FR method is globally convergent if the line search conditions (1.7) are satisfied. [7] presents a new CG with (1.7) and gives the proof. [4] investigates the convergence property by using different choices for β_k. [5] establishes the convergence results in the absence of the sufficient condition. [6] propose a new line search algorithm that ensures global convergence of CG.

In this paper, we will propose the convergence properties of the CG using the Goldstein conditions, (1.8a) and (1.8b). Especially, we take example for Fletcher-Reeves to give the global convergence property.

2 Results for general conjugate gradient methods

In this section, we always assume that $\|g_k\| \neq 0$ for all k, or else a stationary point has been obtained. And from (1.3) we have that

$$g_k^T d_k = -\|g_k\|^2 + \beta_k g_k^T d_{k-1} \tag{2.1}$$

If $g_k^T d_{k-1} > 0$ and $\beta_k < 0$, we have that $g_k^T d_k < 0$. If $g_k^T d_{k-1} < 0$ and $\beta_k > 0$, we still have that $g_k^T d_k < 0$. In other words, we can select the right β_k to enable (1.3) to satisfy the descent condition $g_k^T d_k < 0$ at every search direction d_k.

The condition (1.8b) indicates the sufficient decrease, whereas the (1.8a) means to control the step length from below.

Assumption 2.1.

(i) *f is bounded below on the level set $L = \{x \mid f(x) < f(x_0)\}$, where x_0 is the starting point.*

(ii) *In some neighborhood N of L, f is continuously differentiable, and its gradient is Lipschitz continuous; namely, there is a constant $K > 0$ such that*

$$\|g(x) - g(y)\| \leq K\|x - y\| \quad \text{for all } x, y \in N. \tag{2.2}$$

From **Assumption 2.1** and if the level set L is bounded, we can know that there exists a positive constant η such that

$$\|g(x)\| \leq \eta. \tag{2.3}$$

The following important result was obtained by Zoutendijk [9]

Lemma 2.2. *Suppose that* **Assumption 2.1** *holds. Consider any iteration method of the form (1.2)-(1.3) with d_k satisfying (2.1) and with the Goldstein line search (1.8). Then*

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

Proof. From (1.8a) we have that

$$(1-c)\alpha_k g_k^T d_k \leq f(x_k + \alpha_k d_k) - f(x_k) \tag{2.4}$$

And from the mean value theorem we have that

$$f(x_k + \alpha_k d_k) - f(x_k) = \alpha_k g^T(x_k + \theta_k d_k) d_k \tag{2.5}$$

where

$$\theta_k \in (0, \alpha_k). \tag{2.6}$$

By combining (2.4), (2.5) and (2.2), we obtain

$$-c g_k^T d_k \leq K \theta_k \|d_k\|^2 \tag{2.7}$$

By (2.6) and (2.7), we obtain

$$\alpha_k \geq \frac{-c g_k^T d_k}{K \|d_k\|^2} \tag{2.8}$$

By this inequality into (1.8b) and (2.1), we have

$$f_{k+1} \leq f_k - \frac{c^2 (g_k^T d_k)^2}{K \|d_k\|^2} \tag{2.9}$$

By summing this expression over all indices less than or equal to k, we obtain

$$f_{k+1} \leq f_0 - \sum_{i=0}^{k} \frac{c^2 (g_i^T d_i)^2}{K \|d_i\|^2} \tag{2.10}$$

Since f is bounded below, we have that $f_0 - f_{k+1}$ is less than some positive constant, for all k. Hence by taking limits in (2.10), we obtain

$$\sum_{i=0}^{\infty} \frac{(g_i^T d_i)^2}{\|d_i\|^2} < \infty \tag{2.11}$$

which concludes the proof. □

From (2.11), we can see

$$\lim_{k \to \infty} \frac{g_k^T d_k}{\|d_k\|} = 0 \tag{2.12}$$

Theorem 2.3. *From (1.8b), we can have that*

$$\lim_{k \to \infty} |g_k^T d_k| = 0 \tag{2.13}$$

This result is easy obtained.

A disadvantage of the Goldstein conditions vs. the Wolfe conditions is that the (1.8a) may exclude all minimizers of $f(x_k + \alpha d_k)$. However, the Goldstein and Wolfe conditions have much in common. The Goldstein conditions are often used in Newton-type methods but are not well suited for quasi-Newton methods that maintain a positive definite Hessian approximation.

3 Global Convergence

The CG with Goldstein line search is as the following:

- Step 1 Given $x_1 \in R^n$,$d_1 = -g_1$,$k := 1$, if $g_1 = 0$ then stop; otherwise continue.
- Step 2 Compute an α_k satisfying (1.8).
- Step 3 Generate x_{k+1} by (1.2). If $g_{k+1} = 0$ then stop.
- Step 4 Compute β_k, and generate d_{k+1} by (1.3).
- $k := k + 1$, go to Step2.

Lemma 3.1. ([8].) *Suppose that $m(> 0)$ and c are constant, $\{a_i\}$ is a positive series, if the following for all k holds*

$$\sum_{i=1}^{k} a_i \geq mk + c. \tag{3.1}$$

We have that

$$\sum_{i=1}^{\infty} \frac{a_i^2}{i} = \infty. \tag{3.2}$$

$$\sum_{k=1}^{\infty} \frac{a_i^2}{\sum_{i=1}^{k} a_i} = \infty. \tag{3.3}$$

Take example for Fletcher-Reeves method, i.e. $\beta_k = \beta_k^{FR}$, we give the proof of the global convergence property.

Theorem 3.2. *Suppose that x_0 is a starting point for which **Assumption 2.1** holds. Let $\{x_k, k = 1, 2, \cdots\}$ be generated by (1.2) and (1.3) with a line search (1.8). Then (1.2) and (1.3) terminate at a stationary point or converge in the sense that*

$$\liminf_{k\to\infty} \|g_k\| = 0 \tag{3.4}$$

Proof. When $k \geq 2$ we now rewrite (1.3) as

$$d_k + g_k = \beta_k^{FR} d_{k-1} \tag{3.5}$$

Squaring both sides of the above equation, we get

$$\|d_k\|^2 = -\|g_k\|^2 - 2g_k^T d_k + \beta_k \|d_{k-1}\|^2 \tag{3.6}$$

Set

$$t_k = \frac{\|d_k\|^2}{\|g_k\|^4}, \quad r_k = -\frac{g_k^T d_k}{\|g_k\|^2} \tag{3.7}$$

Note that $t_1 = \frac{1}{\|g_k\|^2}$, $r_1 = 1$, and $r_k > 0$.

From (3.6), (3.7) and (1.4), we have that

$$t_n = -\sum_{k=1}^{n} \frac{1}{\|g_k\|^2} + 2\sum_{k=1}^{n} \frac{r_k}{\|g_k\|^2} \tag{3.8}$$

The proof is by contradiction. If (1.2) and (1.3) do not terminate after many iterations, we have that there is positive constant $\mu > 0$ such that

$$\|g_k\| \geq \mu \quad \text{for all} \quad k. \tag{3.9}$$

Therefore from (2.3) and (3.9), it is easy to obtain that

$$\frac{1}{\eta} \leq \frac{1}{\|g_k\|} \leq \frac{1}{\mu} \tag{3.10}$$

From (3.7) and (3.10), we obtain that

$$t_n \leq -\frac{n}{\eta^2} + \frac{2}{\mu^2} \sum_{k=1}^{n} r_k \tag{3.11}$$

Hence

$$t_n \leq \frac{2}{\mu^2} \sum_{k=1}^{n} r_k \tag{3.12}$$

For $t_k \geq 0$, from (3.11) we can get

$$\sum_{k=1}^{n} r_k \geq \frac{\mu^2 n}{2\eta^2} \tag{3.13}$$

From (3.11), (3.12) and Lemma 3.1, we obtain that

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} = \sum_{k=1}^{\infty} \frac{r_k^2}{t_k} = \infty. \tag{3.14}$$

The relation (3.14) contradicts (2.11). This contraction shows that the theorem is true. □

Theorem 3.3. *Suppose that x_0 is a starting point for which* ***Assumption 2.1*** *holds. Let $\{x_k, k = 1, 2, \cdots\}$ be generated by (1.2) and (1.3), where β_k satisfies (1.5) and (1.6), with a line search (1.8). Then*

$$\liminf_{k\to\infty} \|g_k\| = 0 \tag{3.15}$$

We can prove theorem 2.1 as theorem 3.2.

4 Discussion

In this paper we have presented the global convergence property for nonlinear CG, where the step-length is computed by the Goldstein conditions under the assumption that all the search directions are descent directions. It is shown that in the previous section that the CG converges globally under the Goldstein line search conditions. The assumption that the objective function is bounded below is weaker than the usual assumption that the level set

$$\{x \mid f(x) < f(x_0)\} \tag{4.1}$$

is bounded.

References

[1] M. AL-Baali, Descent property and global convergence of the Fletcher-Reeves method with inexact line search, *IMA Journal of Numerical Analysis.* **5** (1985), 121–124.

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

[3] Y. H. Dai and Y. X. Yuan, Convergence properties of the Fletcher-Reeves methods, *Journal of Numerical Analysis.* **16** (1996), 155–164.

[4] Han Jiye, Liu Guanghui and Yin Hongxia, Convergence properties of conjugate gradient methods with strong Wolfe line search, *Systems Science and Mathematical Sciences*, Vol.11 No. 2, pp.112–116 Apr. 1998.

[5] Yuhong Dai et al. Convergence Properties of Nonlinear Conjugate Gradient Methods, *SIAM Journal of Optimization*, **10**(2): (1999), 345–358.

[6] L. Grippo and S. Lucidi, A Globally Convergent Version of the Polak-Ribière Conjugate Gradient Method, *Mathematical Programming*, **78** (1997), 375–391.

[7] Li Rongsheng and Liu Guanghui, A class of new conjugate gradient methods with inexact line search, *Advances in Mathematics*, Vol.26, No,1 Feb., 1997.

[8] Yuhong Dai and Yaxiang Yuan, Convergence of the Fletcher-Reeves Method under A Generalized Wolfe Search, Numerical Mathematics A Journal of Chinese Universities.

[9] G. Zoutendijk, *Nonlinear Programming, Computational Methods*, in: Integer and Nonlinear Programming (J. Abadie, ed.), (1970), 37–86.

Jian Wang
Supercomputing Center
Computer Network Information Center
and
Institute of Software
CAS, Beijing 100080
and
Graduate School of CAS
Beijing 100039

E-mail: jwang@sccas.cn

Xuebin Chi
Supercomputing Center
Computer Network Information Center
CAS, Beijing 100080

E-mail: chi@sccas.cn