SIAM J. OPTIM.
Vol. 10, No. 1, pp. 177–182
© 1999 Society for Industrial and Applied Mathematics

Downloaded 01/02/13 to 129.173.72.87. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

# A NONLINEAR CONJUGATE GRADIENT METHOD WITH A STRONG GLOBAL CONVERGENCE PROPERTY*

Y. H. DAI† AND Y. YUAN†

**Abstract.** Conjugate gradient methods are widely used for unconstrained optimization, especially large scale problems. The strong Wolfe conditions are usually used in the analyses and implementations of conjugate gradient methods. This paper presents a new version of the conjugate gradient method, which converges globally, provided the line search satisfies the standard Wolfe conditions. The conditions on the objective function are also weak, being similar to those required by the Zoutendijk condition.

**Key words.** unconstrained optimization, new conjugate gradient method, Wolfe conditions, global convergence

**AMS subject classifications.** 65K, 90C

**PII.** S1052623497318992

## 1. Introduction.

Our problem is to minimize a function of $n$ variables

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

where $f$ is smooth and its gradient $g(x)$ is available. Conjugate gradient methods for solving (1.1) are iterative methods of the form

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

where $\alpha_k > 0$ is a steplength and $d_k$ is a search direction. Normally the search direction at the first iteration is the steepest descent direction, namely, $d_1 = -g_1$. The other search directions can be defined recursively:

$$d_{k+1} = -g_{k+1} + \beta_k d_k. \tag{1.3}$$

$\beta_k \in \Re$ is so chosen that (1.2)–(1.3) reduces to the linear conjugate gradient method if $f(x)$ is a strictly convex quadratic function and if $\alpha_k$ is the exact one-dimensional minimizer. Well-known formulas for $\beta_k$ are the Fletcher–Reeves (FR), Polak–Ribière–Polyak (PRP), and Hestenes–Stiefel (HS) formulas (see [6]; [10], [11]; and [7], respectively) and are given by

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

$$\beta_k^{PRP} = g_{k+1}^T y_k/\|g_k\|^2, \tag{1.5}$$

$$\beta_k^{HS} = g_{k+1}^T y_k/d_k^T y_k, \tag{1.6}$$

where $y_k = g_{k+1} - g_k$ and $\|\cdot\|$ denotes the Euclidean norm.

*Received by the editors March 26, 1997; accepted for publication (in revised form) March 26, 1999; published electronically November 29, 1999. This work was supported by Chinese National Natural Science Foundation grants 19525101 and 19731010.

http://www.siam.org/journals/siopt/10-1/31899.html

†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 (dyh@lsec.cc.ac.cn, yyx@lsec.cc.ac.cn).

Downloaded 01/02/13 to 129.173.72.87. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

The global convergence properties of the FR, PRP, and HS methods without regular restarts have been studied by many authors, including Zoutendijk [15], Al-Baali [1], Liu, Han, and Yin [9], Dai and Yuan [2], Powell [12], Gilbert and Nocedal [8], and Dai and Yuan [4]. To establish the convergence results of these methods, it is normally required that the steplength $\alpha_k$ satisfy the following strong Wolfe conditions:

$$f(x_k) - f(x_k + \alpha_k d_k) \geq -\delta\alpha_k g_k^T d_k, \tag{1.7}$$

$$|g(x_k + \alpha_k d_k)^T d_k| \leq -\sigma g_k^T d_k, \tag{1.8}$$

where $0 < \delta < \sigma < 1$. Some convergence analyses even require the $\alpha_k$ be computed by the exact line search, namely,

$$f(x_k + \alpha_k d_k) = \min_{\alpha > 0} f(x_k + \alpha d_k). \tag{1.9}$$

On the other hand, many other numerical methods for unconstrained optimization are proved to be convergent under the standard Wolfe conditions (1.7) and

$$g(x_k + \alpha_k d_k)^T d_k > \sigma g_k^T d_k. \tag{1.10}$$

For example, see Fletcher [5]. Hence it is interesting to investigate whether there exists a conjugate gradient method that converges under the standard Wolfe conditions.

In this paper, we give a new formula for $\beta_k$. It is shown that this new conjugate gradient method is globally convergent as long as the standard Wolfe conditions (1.7) and (1.10) are satisfied. Moreover, the conditions on the objective function are also weaker than the usual ones.

**2. New formula for $\boldsymbol{\beta_k}$.** One motivation for our new formula for $\beta_k$ is the descent property of the conjugate descent method (see Fletcher [5]), which uses

$$\beta_k^{CD} = \|g_{k+1}\|^2 / (-d_k^T g_k). \tag{2.1}$$

It can be shown that the conjugate descent method always produces a descent direction if the strong Wolfe conditions are satisfied. We try to find a conjugate gradient method which generates descent directions provided the standard Wolfe conditions are satisfied. Suppose the current search direction $d_k$ is a descent direction, namely, $d_k^T g_k < 0$. Now we need to find a $\beta_k$ that defines a descent direction $d_{k+1}$. This requires that

$$-\|g_{k+1}\|^2 + \beta_k g_{k+1}^T d_k < 0. \tag{2.2}$$

We assume that $\beta_k > 0$. Denote $\tau_k = \|g_{k+1}\|^2/\beta_k$. The above inequality is equivalent to

$$\tau_k > g_{k+1}^T d_k. \tag{2.3}$$

Therefore, we can let $\tau_k = d_k^T y_k$, giving our new formula

$$\beta_k = \|g_{k+1}\|^2 / d_k^T y_k. \tag{2.4}$$

This formula is well defined because line search condition (1.10) implies $d_k^T y_k > 0$. If line searches are exact, the above formula is the same as the FR formula (1.4). Therefore we see that (2.4) corresponds to a nonlinear conjugate gradient method. It is interesting to note that (2.4) has the same numerator as the FR formula (1.4) and has the same denominator as the HS formula (1.6). Now we can define the new method, as follows.

Downloaded 01/02/13 to 129.173.72.87. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

ALGORITHM 2.1 (A new CG method).

*Step* 1. *Given* $x_1 \in \Re^n$, $d_1 = -g_1$, $k := 1$, *if* $g_1 = 0$, *then stop.*

*Step* 2. *Compute an* $\alpha_k > 0$ *satisfying* (1.7) *and* (1.10).

*Step* 3. *Let* $x_{k+1} = x_k + \alpha_k d_k$. *If* $g_{k+1} = 0$, *then stop.*

*Step* 4. *Compute* $\beta_k$ *by* (2.4) *and generate* $d_{k+1}$ *by* (1.3), $k := k+1$, *go to Step* 2.

It follows from (1.3) and (2.4) that

$$g_{k+1}^T d_{k+1} = \frac{\|g_{k+1}\|^2}{d_k^T y_k} g_k^T d_k = \beta_k g_k^T d_k. \tag{2.5}$$

The above relation can be rewritten as

$$\beta_k = \frac{g_{k+1}^T d_{k+1}}{g_k^T d_k}. \tag{2.6}$$

This formula is very important in our convergence analysis.

## 3. Convergence of the new method.

In this section, we establish a convergence theorem for Algorithm 2.1. We assume that the objective function satisfies the following conditions.

*Assumption* 3.1. (1) $f$ is bounded below on $\Re^n$ and is continuously differentiable in a neighborhood $\mathcal{N}$ of the level set $\mathcal{L} = \{x \in \Re^n : f(x) \le f(x_1)\}$; (2) the gradient $\nabla f(x)$ is Lipschitz continuous in $\mathcal{N}$, i.e., there exists a constant $L > 0$ such that

$$\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\| \text{ for any } x,\ y \in \mathcal{N}. \tag{3.1}$$

Under Assumption 3.1, we give a useful lemma which was essentially proved by Zoutendijk [15] and Wolfe [13, 14].

LEMMA 3.2. *Suppose that* $x_1$ *is a starting point for which Assumption* 3.1 *is satisfied. Consider any method of the form* (1.2), *where* $d_k$ *is a descent direction and* $\alpha_k$ *satisfies the standard Wolfe conditions* (1.7) *and* (1.10). *Then we have that*

$$\sum_{k \ge 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty. \tag{3.2}$$

*Proof.* It follows from (1.10) that

$$d_k^T y_k = d_k^T (g_{k+1} - g_k) \ge (\sigma - 1) g_k^T d_k. \tag{3.3}$$

On the other hand, the Lipschitz condition (3.1) implies

$$(g_{k+1} - g_k)^T d_k \le \alpha_k L \|d_k\|^2. \tag{3.4}$$

The above two inequalities give

$$\alpha_k \ge \frac{\sigma - 1}{L} \cdot \frac{g_k^T d_k}{\|d_k\|^2}, \tag{3.5}$$

which with (1.7) implies that

$$f_k - f_{k+1} \ge c \frac{(g_k^T d_k)^2}{\|d_k\|^2}, \tag{3.6}$$

where $c = \delta(1-\sigma)/L$. Summing (3.6) and noting that $f$ is bounded below, we see that (3.2) holds, which concludes the proof. □

THEOREM 3.3. *Suppose that $x_1$ is a starting point for which Assumption 3.1 holds. Let $\{x_k,\, k=1,2\ldots\}$ be generated by Algorithm 2.1. Then the algorithm either terminates at a stationary point or converges in the sense that*

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{3.7}$$

*Proof.* If the algorithm does not terminate after finite many iterations, we have that

$$\|g_k\| > 0 \quad \text{for all } k. \tag{3.8}$$

First we show all search directions are descent, namely,

$$g_k^T d_k < 0 \tag{3.9}$$

for all $k$. The above inequality is obvious for $k = 1$. Now we prove it for all $k \geq 1$ by induction. Assume (3.9) holds for $k$. It follows from the line search conditions that

$$d_k^T y_k \geq (\sigma - 1) d_k^T g_k > 0. \tag{3.10}$$

The above inequality and (2.5) imply that (3.9) holds for $k+1$. This shows that (3.9) is true for all $k \geq 1$.

We now rewrite (1.3) as

$$d_{k+1} + g_{k+1} = \beta_k d_k. \tag{3.11}$$

Squaring both sides of the above equation, we get

$$\|d_{k+1}\|^2 = \beta_k^2 \|d_k\|^2 - 2 g_{k+1}^T d_{k+1} - \|g_{k+1}\|^2. \tag{3.12}$$

Dividing both sides by $(g_{k+1}^T d_{k+1})^2$ and applying (2.6), we obtain that

$$\begin{aligned}
\frac{\|d_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2} &= \frac{\|d_k\|^2}{(g_k^T d_k)^2} - \frac{2}{g_{k+1}^T d_{k+1}} - \frac{\|g_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2} \\
&= \frac{\|d_k\|^2}{(g_k^T d_k)^2} - \left( \frac{1}{\|g_{k+1}\|} + \frac{\|g_{k+1}\|}{g_{k+1}^T d_{k+1}} \right)^2 + \frac{1}{\|g_{k+1}\|^2} \\
&\leq \frac{\|d_k\|^2}{(g_k^T d_k)^2} + \frac{1}{\|g_{k+1}\|^2}.
\end{aligned} \tag{3.13}$$

Because $\|d_1\|^2/(g_1^T d_1)^2 = 1/\|g_1\|^2$, (3.13) shows that

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} \leq \sum_{i=1}^{k} \frac{1}{\|g_i\|^2} \tag{3.14}$$

for all $k$. If the theorem is not true, there exists a constant $c > 0$ such that

$$\|g_k\| \geq c \quad \text{for all } k. \tag{3.15}$$

Downloaded 01/02/13 to 129.173.72.87. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 01/02/13 to 129.173.72.87. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Therefore it follows from (3.14) and (3.15) that

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} \leq \frac{k}{c^2}, \tag{3.16}$$

which implies that

$$\sum_{k\geq 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} = \infty. \tag{3.17}$$

Relation (3.17) contradicts the Zoutendijk condition (3.2). This contradiction shows that the theorem is true. $\square$

**4. Discussion.** It is shown in the previous section that the new conjugate gradient method converges under the standard Wolfe line search conditions. It should be noted that our assumption that the objective function is bounded below is weaker than the usual assumption that the level set

$$\{x \in \Re^n : f(x) \leq f(x_1)\} \tag{4.1}$$

is bounded.

From the proof of Theorem 3.3, we can see that the equivalent form (2.6) of the formula (2.4) plays an important role in the convergence analysis. Relation (2.6) enables us to establish the recurrence relation (3.13), which is about the sequence of the reciprocal $\{(g_k^T d_k)^2/\|d_k\|^2\}$. The term $(g_k^T d_k)^2/\|d_k\|^2$ is exactly the one that appears in the Zoutendijk condition (3.2). This makes our convergence analysis very simple. It is known that to obtain the convergence of the FR, PRP, and HS methods, one normally has to consider two sequences. For example, Al-Baali [1] considered the sequences $\{\|d_k\|^2\}$ and $\{g_k^T d_k/\|g_k\|^2\}$ for the FR method, and Gilbert and Nocedal [8] considered $\{\|d_k\|^2\}$ and $\{g_k^T d_k\}$ for the PRP and HS methods.

It is also worth noting that Al-Baali [1] and Gilbert and Nocedal [8] proved or required the sufficient descent condition, namely,

$$g_k^T d_k \leq -c\|g_k\|^2 \quad \text{for some } c > 0 \text{ and for all } k \geq 1. \tag{4.2}$$

However, our method does not guarantee this inequality. But if the strong Wolfe line search conditions are satisfied at every iteration, we have that

$$l_k = \frac{g_{k+1}^T d_k}{g_k^T d_k} \in [-\sigma, \sigma]. \tag{4.3}$$

Formula (2.5) can be rewritten as

$$g_{k+1}^T d_{k+1} = \frac{1}{l_k - 1}\|g_{k+1}\|^2. \tag{4.4}$$

The above two relations show that (4.2) holds with $c = 1/(1+\sigma)$. This indicates that our method also has the sufficient descent property (4.2) if the strong Wolfe line search conditions are used.

Dai and Yuan [3] considered a class of methods that use

$$\beta_k \in [(\sigma - 1)/(1+\sigma), 1]\bar{\beta}_k, \tag{4.5}$$

where $\bar{\beta}_k$ is given by (2.4). It is shown in [3] that Algorithm 2.1 is still convergent if, in Step 4, $\beta_k$ computed by (2.4) is replaced by any $\beta_k$, satisfying (4.5).

Downloaded 01/02/13 to 129.173.72.87. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

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, *Convergence properties of the Fletcher-Reeves method*, IMA J. Numer. Anal., 16 (1996), pp. 155–164.

[3] Y. H. Dai and Y. Yuan, *A nonlinear conjugate gradient method with a nice global convergence property*, Research report ICM-95-038, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, Beijing, 1995.

[4] Y. H. Dai and Y. Yuan, *Further studies on the Polak-Ribière-Polyak method*, Research report ICM-95-040, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, Beijing, 1995.

[5] R. Fletcher, *Practical Methods of Optimization*, 2nd ed., John Wiley, New York, 1989.

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

[7] M. R. Hestenes and E. L. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Research Nat. Bur. Standards, 49 (1952), pp. 409–436.

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

[9] 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, Beijing, 1993.

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

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

[12] M. J. D. Powell, *Nonconvex Minimization Calculations and the Conjugate Gradient Method*, Lecture Notes in Math. 1066, Springer-Verlag, Berlin, 1984, pp. 122–141.

[13] P. Wolfe, *Convergence conditions for ascent methods*, SIAM Rev., 11 (1969), pp. 226–235.

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

[15] G. Zoutendijk, *Nonlinear programming, computational methods*, in Integer and Nonlinear Programming, J. Abadie, ed., North-Holland, Amsterdam, 1970, pp. 37–86.