International Journal of Pure and Applied Mathematics
Volume 71 No. 3 2011, 343-349

CONVERGENCE OF DAI-YUAN CONJUGATE METHOD WITH GENERAL WOLFE LINE SEARCH

Zhijun Luo[1 §], Jian Xie[2], Rong Xiong[3]
Department of Mathematics and Applied Mathematics
Humanities and Science and Technology Institute of Hunan
Loudi, 417000, P.R. CHINA

Abstract: Conjugate gradient method is a method for solving nonlinear optimization problems. In this paper, we develop a new general Wolfe line search for DY(Dai-Yuan) conjugate gradient method. Under some mild conditions, the general Wolfe line search can guarantee the global convergence of original DY method.

AMS Subject Classification: 90C30, 65K05
Key Words: unconstrained optimization, DY conjugate gradient method, Wolfe line search, global convergence

1. Introduction

Conjugate gradient method is a method for solving nonlinear optimization problems. It is not only one of the most useful methods to solve large-scale linear equations, but also one of the most effective algorithms to solve large nonlinear optimization problems. In this paper, we consider the following an unconstrained minimization problem:

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

where R^n denotes an n-dimensional Euclidean space and $f : R^n \longrightarrow R$ is a smooth and nonlinear function.

It is well known, conjugate gradient method is a line search method that takes the form

$$x^{k+1} = x^k + \alpha_k d^k, \quad k = 0, 1, 2, \cdots, \tag{1.2}$$

where d^k is a descent direction of $f(x)$ at x^k is a step size. If x_k is the current

Received: October 10, 2010
§Correspondence author

iterate, we denote $f(x^k)$ by $f_k, \nabla f(x^k)$ by g_k, $\bigtriangledown f^2(x^k)$ by G_k and $f(x^*)$ by f^*, respectively. If G_k is available and inverse, then $d^k = -G_k^{-1} g_k$ leads to the Newton method and $d^k = -g_k$ results in the steepest descent method [1]. The search direction d^k is generally required to satisfy

$$g_k^T d^k < 0,$$

which guarantees that d^k is a descent direction of $f(x)$ at x^k [2]. In order to guarantee the global convergence, we sometimes require d^k to satisfy a sufficient descent condition

$$g_k^T d^k \leq -c\|g_k\|^2,$$

where $c > 0$ is a constant. In line search methods, the well-known conjugate gradient method has the form (1.2) in which

$$d^k = \begin{cases} -g_k, & if\ k = 0, \\ -g_k + \beta_k d^{k-1}, & \text{if } k \geq 1, \end{cases} \tag{1.3}$$

where

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

or β_k is represented by other formulae [3]. The corresponding methods are called FR (Fletcher- Reeves) [4], PRP (Polak-Ribiére-Polyak) [5]-[6] and DY (Dai-Yuan) [7] conjugate gradient method, respectively.

Although the above mentioned conjugate gradient algorithms are equivalent to each other for minimizing strong convex quadratic functions under exact line search, they have different performance when using them to minimize non-quadratic functions or using inexact line searches. For non-quadratic objective functions, the FR method has global convergence when exact line search or strong Wolfe line search [8]-[9] is used. In this paper, we devote to the global convergence of original DY method. A new general Wolfe line search is proposed for the original DY conjugate gradient method. Under some mild conditions,the general Wolfe line search can guarantee the global convergence of original DY method.

The rest of this paper is organized as follows. The algorithm is presented in Section 2. In Sections 3 the global convergence is analyzed.

2. Description of Algorithm

We first assume that:

H 2.1. The objective function $f(x)$ is continuously differentiable on R^n and has a lower bound.

H 2.2. The gradient $g(x)$ of $f(x)$ is Lipschitz continuous on an open convex set U that contains the level set $L_0 = \{x \in R^n \mid f(x) \leq f(x^0)\}$ with x^0 being given, i.e., there exists $L > 0$ such that

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

Algorithm.
Step 0: Choose $x^0 \in R^n$ and $\varepsilon > 0$;
Step 1: Set $d^0 = -g_0$, $k := 0$. If $\|g_0\| \leq \varepsilon$ then stop else go to Step 2.
Step 2 (General Wolfe Line Search) Compute step size α_k, such that,

$$f(x^k + \alpha_k d^k) - f(x^k) \leq \delta \alpha_k g_k^T d^k \tag{2.4}$$

$$\sigma_1 g_k^T d^k \leq g(x_k + \alpha_k d^k)^T d^k \leq -\sigma_2 g_k^T d^k \tag{2.5}$$

where $\delta, \sigma_1, \sigma_2 \in (0, 1)$. Let $x^{k+1} = x^k + \alpha_k d^k$, $k := k + 1$
Step 3: Computer g_k, if $\|g_k\| \leq \varepsilon$, stop. Otherwise, go to step step 4.
Step 4: Computer $d^k = -g_k + \beta_k^{DY} d^{k-1}$, go to step 2.

Lemma 2.1. *Assume that H2.1 and H2.2 hold. DY conjugate gradient method with the general Wolfe line search generates an infinite sequence $\{x^k\}$. Then*

$$g_k^T d^k < 0.$$

Proof. If $k = 0$, $d^0 = -g_0$, we have

$$g_0^T d^0 = -\|g_0\|^2 < 0.$$

For $k \geq 1$, by the assumption H2.1 and H2.2, it is easy know that function $f(x)$ is strictly convex function. We can obtain

$$(x^k - x^{k-1})^T (g_k - g_{k-1}) > 0.$$

i.e.

$$\alpha_{k-1} d^{(k-1)T} (g_k - g_{k-1}) > 0.$$

Then, we have

$$\begin{aligned} g_k^T d^k &= g_k^T(-g_k + \beta_k^{DY} d^{k-1}) \\ &= -\|g_k\|^2 + \frac{\|g_k\|^2}{d^{(k-1)T}(g_k - g_{k-1})} g_k^T d^{k-1} \\ &= \frac{\|g_k\|^2}{d^{(k-1)T}(g_k - g_{k-1})} g_{k-1}^T d^{k-1} \\ &= \beta_k^{DY} g_{k-1}^T d^{k-1}. \end{aligned} \tag{2.6}$$

Which implies $g_k^T d^k < 0$.

Lemma 2.2. *If H2.1 and H2.2 hold, DY conjugate gradient method with the general Wolfe line search generates an infinite sequence* $\{x^k\}$, α_k *satisfying (2.4) and (2.5), then*

$$\frac{1}{1-\sigma_1} \geq -\frac{g_k^T d^k}{\|g_k\|^2} \geq \frac{1}{1+\sigma_2} \tag{2.7}$$

Proof. If $k = 0$, $d^0 = -g_0$, (2.7) holds. For $k \geq 1$, from lemma 2.1, we have

$$g_k^T d^k + \|g_k\|^2 = \beta_k^{DY} g_k^T d^{k-1}. \tag{2.8}$$

Thus, in view of (2.5), we obtain

$$\sigma_1 \beta_k^{DY} g_{k-1}^T d^{k-1} \leq \beta_k^{DY} g_k^T d^{k-1} \leq -\sigma_2 \beta_k^{DY} g_{k-1}^T d^{k-1}$$

Taking into account (2.6) and (2.8), it is easy to get

$$\sigma_1 g_k^T d^k \leq g_k^T d^k + \|g_k\|^2 \leq -\sigma_2 g_k^T d^k.$$

i.e.

$$-(1-\sigma_1) g_k^T d^k \leq \|g_k\|^2 \leq -(1+\sigma_2) g_k^T d^k.$$

So, we have

$$\frac{1}{1-\sigma_1} \geq -\frac{g_k^T d^k}{\|g_k\|^2} \geq \frac{1}{1+\sigma_2}.$$

This result shows that the descent direction d^k satisfy a sufficient descent condition,i.e.

$$g_k^T d^k \leq -c\|g_k\|^2,$$

where $c > 0$ is a constant.

3. Global Convergence of Algorithm

In this section, we analyze the global convergence of the Algorithm.

Lemma 3.1. *If H2.1 and H2.2 hold, α_k satisfying (2.4) and (2.5), then*

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

Proof. In view of (2.5), we have

$$(\sigma_1 - 1) g_k^T d^k \leq (g(x_{k+1}) - g_k)^T d^k.$$

Using Cauchy-Schwartz inequality and H2.2, it is easy to obtain

$$(1 - \sigma_1)|g_k^T d^k| \leq \|g_{k+1} - g_k\| \|d^k\| \leq L\alpha_k \|d^k\|^2.$$

Thus,

$$\alpha_k \geq \frac{(1-\sigma_1)|g_k^T d^k|}{L\|d^k\|^2}. \tag{3.1}$$

From (2.4) and (3.1), we have

$$f(x^k) - f(x^{k+1}) \geq \delta \frac{(1-\sigma_1)}{L} \frac{(g_k^T d^k)^2}{\|d^k\|^2},$$

then,

$$\sum_{k=1}^{\infty} [f(x^k) - f(x^{k+1})] \geq \sum_{k=1}^{\infty} \delta \frac{(1-\sigma_1)}{L} \frac{(g_k^T d^k)^2}{\|d^k\|^2}.$$

So, we have

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

Theorem 3.1. *Assume that H2.1 and H2.2 hold. DY conjugate gradient method with the general Wolfe line search generates an infinite sequence $\{x^k\}$. Then*

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

Proof. From (1.3), we have

$$d^k + g_k = \beta_k d^{k-1}.$$

Then,

$$\|d^k\|^2 + 2g_k^T d^k + \|g_k\|^2 = \beta_k^2 \|d^{k-1}\|^2,$$

i.e.

$$\frac{\|d^k\|^2}{(g_k^T d^k)^2} = -\frac{\|g_k\|^2}{(g_k^T d^k)^2} - \frac{2}{(g_k^T d^k)} + \beta_k^2 \frac{\|d^{k-1}\|^2}{(g_k^T d^k)^2}. \tag{3.2}$$

In view of (2.6) we can obtain

$$\beta_k^{DY} = \frac{\|g_k\|^2}{d^{(k-1)T}(g_k - g_{k-1})} = \frac{g_k^T d^k}{g_{k-1}^T d^{k-1}}.$$

So, taking into account (3.2), it is easy to get

$$\begin{aligned}\frac{\|d^k\|^2}{(g_k^T d^k)^2} &= -\frac{\|g_k\|^2}{(g_k^T d^k)^2} - \frac{2}{(g_k^T d^k)} + \frac{\|d^{k-1}\|^2}{(g_{k-1}^T d^{k-1})^2} \\ &= \frac{\|d^{k-1}\|^2}{(g_{k-1}^T d^{k-1})^2} - \left|\frac{\|g_k\|}{(g_k^T d^k)} + \frac{1}{\|g_k\|}\right|^2 + \frac{1}{\|g_k\|^2},\end{aligned}$$

i.e.

$$\frac{\|d^k\|^2}{(g_k^T d^k)^2} \le \frac{\|d^{k-1}\|^2}{(g_{k-1}^T d^{k-1})^2} + \frac{1}{\|g_k\|^2}.$$

Then, we have

$$\frac{\|d^k\|^2}{(g_k^T d^k)^2} \le \sum_{l=0}^{k} \frac{1}{\|g_l\|^2}. \tag{3.3}$$

Now, we prove the results by contradiction. Suppose that the following inequality

$$\|g_k\|^2 > c, \forall k = 1, 2, \ldots,$$

where $c > 0$ is a constant. Hence, it follows from (3.3)that

$$\frac{\|d^k\|^2}{(g_k^T d^k)^2} \le \frac{k}{c},$$

i.e.

$$\frac{(g_k^T d^k)^2}{\|d^k\|^2} \ge \frac{c}{k}.$$

So, it is easy to see that

$$\sum_{k=0}^{\infty} \frac{(g_k^T d^k)^2}{\|d^k\|^2} = +\infty,$$

which contradicts lemma 3.1, i.e. $\lim_{k\to\infty} inf\|g_k\| = 0$.

Acknowledgments

This work was supported in part by Scientific Research Fund of Hunan Institute of Humanities, Science and Technology of P.R. China (No. 2009QN07) and Educational Reform Research Fund of Hunan Institute of Humanities, Science and Technology (No. RKJGY1030).

References

[1] J. Nocedal, J.S. Wright, *Numerical Optimization*, Springer-Verlag, New York (1999).

[2] Y. Yuan, *Numerical Methods for Nonlinear Programming*, Shanghai Scientific and Technical Publishers (1993).

[3] Y. Liu, C. Storey, Efficient generalized conjugate gradient algorithms. Part 1: Theory, *Journal of Optimization Theory and Applications*, **69** (1991), 129-137.

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

[5] E. Polak, G. Ribie′re, Note sur la convergence de directions conjugue′es, *Rev. Francaise Infomat Recherche Operatonelle*, 3e Anne′e, **16** (1969), 35-43.

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

[7] Y.H. Dai, Y. Yuan, A nonlinear conjugate gradient method with a strong global convergence property, *SIAM J. Optim.*, **10** (1999), 177-182.

[8] M. Al-Baali, New property and global convergence of the Fletcher-Reeves method with inexact line searches, *IMA Journal of Numerical Analysis*, **5** (1985), 122-124.

[9] Y.H. Dai, Y. Yuan, Convergence properties of the Fletcher-Reeves method, *IMA Journal of Numerical Analysis*, **16** (1996), 155-164.