International Journal of Mathematical Archive-2(10), 2011, Page: 1764-1769
IJMA Available online through www.ijma.info ISSN 2229 – 5046

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

***Lirong Wang[1], Zhijun Luo[2], Jian Xie[3] and Jing Xiong[4]**

[1]Department of Computational Science and Technology,Hunan University of Humanities, Science and Technology, Loudi, 417000, PR China

[2, 3,4]Department of Mathematics & Applied Mathematics, Hunan University of Humanities, Science and Technology, Loudi, 417000, PR China
E-mail: ldwlr1234@163.com

(Received on: 02-09-11; Accepted on: 16-09-11)

ABSTRACT

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

Keywords: *Unconstrained optimization; DY Conjugate gradient method; Wolfe line search; Global convergence*

Mathematics Subject Classification (2000): *90C30 65k05.*

---

1. INTRODUCTION:

Conjugate gradient method is a method for solving nonlinear optimization problems. In this paper, we consider the following an unconstrained minimization problem:

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

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

As you know, conjugate gradient method is a line search method that takes the form:

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

where d^k is a descent direction of $f(x)$ at x^k is a step size. If x^k is the current iterate, we denote $f(x^k) \triangleq f_k, \nabla f(x^k) \triangleq g_k, \nabla^2 f(x^k) \triangleq G_k$ and $f(x^*) \triangleq f_*$, 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

---

Corresponding author: *Lirong Wang[1], *E-mail: ldwlr1234@163.com

$$g_k^T d^k \le 0,$$

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

$$g_k^T d^k \le -c \| g_k \|^2 \ , \tag{1.3}$$

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

$$d^k = \begin{cases} -g_k, & if \quad k = 0 \\ -g_k + \beta_k d^{k-1}, & if \quad k \ge 1 \end{cases} \tag{1.4}$$

Where

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

Or β_k is represented by other formulae [3]. The corresponding methods are called FR (Fletcher-Reeves) [4]; PRP (Polak-Ribiere-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 modified DY method. In order to make the direction d^k satisfies the sufficient descent condition (1.3), we take the following modified direction:

$$d^k = \begin{cases} -g_k, & if \quad k = 0 \\ -\theta_k g_k + \beta_k^{DY} d^{k-1}, & if \quad k \ge 1 \end{cases}, \tag{1.5}$$

Where $\theta_k = 1 + \beta_k^{DY} \frac{g_k^T d^{k-1}}{\| g_k \|^2}$. The general Wolfe line search is proposed for the modified DY conjugate gradient method. Under some mild conditions, the general Wolfe line search can guarantee the global convergence of modified 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

$L_0 = \{x \in R^n \mid f(x) \le f(x^0)\}$ with x^0 being given, i.e., there exists $L > 0$ such that

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

Algorithm A:

Step 0: Choose $x^0 \in R^n$ and $\forall \varepsilon > 0$;

Step 1: Set $d^0 = -g_0$, $k := 0$. If $\| g_0 \| \le \varepsilon$ then STOP else go to Step 2.

Step 2: The general Wolfe line search

Compute step size α_k, such that,

$$f(x^k + \alpha_k d^k) - f(x^k) \le \delta \alpha_k g_k^T d^k, \tag{2.1}$$

$$\sigma_1 g_k^T d^k \le g(x^k + \alpha_k d^k)^T d^k \le -\sigma_2 g_k^T d^k, \tag{2.2}$$

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 $\| g_k \| \le \varepsilon$, STOP. Otherwise, go to step 4.

Step 4: Computer $d^k = -\theta_k g_k + \beta_k^{DY} d^{k-1}$, 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 the direction d^k given by (1.5) satisfies the sufficient descent condition

$$g_k^T d^k = -\| g_k \|^2, \tag{2.3}$$

Proof: From (1.5), if $k = 0, d^0 = -g_o$, we have $g_0^T d^0 = -\| g_0 \|^2$, the result clearly holds.

For $k \ge 1$, we can obtain

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

Lemma: 2.2 If **H2.1** and **H2.2** hold, consider the Algorithm A, α_k satisfying (2.1) and (2.2), then

$$0 < \beta_k^{DY} \le \frac{\| g_k \|^2}{(1 - \sigma_1) \| g_{k-1} \|^2}.$$

Proof: From the line search condition (2.1), (2.2) and the sufficient descent condition (2.3), we have

$$\begin{aligned} d^{(k-1)T}(g_k - g_{k-1}) &= d^{(k-1)T} g_k - d^{(k-1)T} g_{k-1} \\ &\ge (\sigma_1 - 1) d^{(k-1)T} g_{k-1} = (1 - \sigma_1) \| g_{k-1} \|^2 > 0. \end{aligned}$$

In this case, it is easy to show that

$$0<\beta_k^{DY}=\frac{\|g_k\|^2}{d^{(k-1)T}(g_{k-1}-g_k)}\le\frac{\|g_k\|^2}{(1-\sigma_1)\|g_{k-1}\|^2}. \qquad (2.4)$$

3. GLOBAL CONVERGENCE OF ALGORITHM:

Now we analyze the global convergence of the Algorithm.

Lemma: 3.1 If **H2.1** and **H2.2** hold, α_k satisfying (2.1) and (2.2), then

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

Proof: In view of (2.2), we have

$$(\sigma_1-1)d^{kT}g_k\le(g_{k+1}-g_k)^T d^k.$$

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

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

Thus,

$$\frac{(1-\sigma_1)|d^{kT}g_k|}{L\|d^k\|^2}\le\alpha_k. \qquad (3.1)$$

From (2.1) and (3.1), we have $f_k-f_{k+1}\ge\delta\frac{(1-\sigma_1)(g_k^T d^k)^2}{L\|d^k\|^2}$,

Then,

$$\sum_{k=0}^{\infty}f_k-f_{k+1}\ge\sum_{k=0}^{\infty}\delta\frac{(1-\sigma_1)(g_k^T d^k)^2}{L\|d^k\|^2}.$$

So we have

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

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 (1.5), we have

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

Then, $\|d^k\|^2+2\theta_k g_k^T d^k+(\theta_k)^2\|g_k\|^2=(\beta_k^{DY})^2\|d^{k-1}\|^2,$

i.e.

$$\begin{aligned}\frac{\|d^k\|^2}{(g_k^T d^k)^2} &= -(\theta_k)^2 \frac{\|g_k\|^2}{(g_k^T d^k)^2} - 2\theta_k \frac{1}{g_k^T d^k} + (\beta_k^{DY})^2 \frac{\|d^{k-1}\|^2}{(g_k^T d^k)^2} \\ &= (\beta_k^{DY})^2 \frac{\|d^{k-1}\|^2}{(g_k^T d^k)^2} - (\theta_k - 1)^2 \frac{\|g_k\|^2}{(g_k^T d^k)^2} + \frac{\|g_k\|^2}{(g_k^T d^k)^2} \\ &\le (\beta_k^{DY})^2 \frac{\|d^{k-1}\|^2}{(g_k^T d^k)^2} + \frac{\|g_k\|^2}{(g_k^T d^k)^2} = (\beta_k^{DY})^2 \frac{\|d^{k-1}\|^2}{\|g_k\|^4} + \frac{1}{\|g_k\|^2}.\end{aligned} \tag{3.2}$$

In view of (2.4) we can obtain

$$0 < \beta_k^{DY} \le \frac{\|g_k\|^2}{(1-\sigma_1)\|g_{k-1}\|^2}.$$

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

$$\begin{aligned}\frac{\|d^k\|^2}{(g_k^T d^k)^2} &= -(\theta_k)^2 \frac{\|g_k\|^2}{(g_k^T d^k)^2} - 2\theta_k \frac{1}{g_k^T d^k} + (\beta_k^{DY})^2 \frac{\|d^{k-1}\|^2}{(g_k^T d^k)^2} \\ &\le [\frac{\|g_k\|^2}{(1-\sigma_1)\|g_{k-1}\|^2}]^2 \frac{\|d^{k-1}\|^2}{\|g_k\|^4} + \frac{1}{\|g_k\|^2} \\ &= \frac{\|d^{k-1}\|^2}{(1-\sigma_1)^2 \|g_{k-1}\|^4} + \frac{1}{\|g_k\|^2} \\ &= \frac{\|d^{k-1}\|^2}{(1-\sigma_1)^2 (g_{k-1}^T)^2} + \frac{1}{\|g_k\|^2}.\end{aligned}$$

Noting that

$$\frac{\|d^0\|^2}{(g_0^T d^0)^2} = \frac{1}{\|g_0\|^2},$$

from (2.4) we get

$$\frac{\|d^k\|^2}{(g_k^T d^k)^2} \le \sum_{l=0}^{\infty} \frac{1}{(1-\sigma_1)^2 \|g_l\|^2}. \tag{3.3}$$

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

$$\|g_k\|^2 > c, \forall k = 0,1,2,\ldots,$$

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

$$\frac{\|d^k\|^2}{(g_k^T d^k)^2} \le \frac{k}{(1-\sigma_1)^2 c^2},$$

i.e.

$$\frac{(g_k^T d^k)^2}{\|d^k\|^2} \ge \frac{(1-\sigma_1)^2 c^2}{k}.$$

So, it is easy to see that

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

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

ACKNOWLEDGEMENT:

This work was supported in part by 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, Inc., 1999.

[2] Y. Yuan, Numerical Methods for Nonlinear Programming, Shanghai Scientific & 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.

****

