

## Research Article

# A Modified Conjugacy Condition and Related Nonlinear Conjugate Gradient Method

Shengwei Yao,<sup>1,2</sup> Xiwen Lu,<sup>2</sup> and Bin Qin<sup>1</sup>

<sup>1</sup> School of Information and Statistics, Guangxi University of Finance and Economics, Nanning 530003, China

<sup>2</sup> School of Science, East China University of Science and Technology, Shanghai 200237, China

Correspondence should be addressed to Shengwei Yao; idhot@163.com

Received 23 November 2013; Accepted 13 February 2014; Published 18 March 2014

Academic Editor: Jein-Shan Chen

Copyright © 2014 Shengwei Yao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The conjugate gradient (CG) method has played a special role in solving large-scale nonlinear optimization problems due to the simplicity of their very low memory requirements. In this paper, we propose a new conjugacy condition which is similar to Dai-Liao (2001). Based on this condition, the related nonlinear conjugate gradient method is given. With some mild conditions, the given method is globally convergent under the strong Wolfe-Powell line search for general functions. The numerical experiments show that the proposed method is very robust and efficient.

## 1. Introduction

The conjugate gradient (CG) method has played a special role in solving large-scale nonlinear optimization problems due to the simplicity of their iterations and their very low memory requirements. In fact, the CG method is not among the fastest or most robust optimization methods for nonlinear problems available today, but it remains very popular for engineers and mathematicians who are interested in solving large-scale problems. The conjugate gradient method is designed to solve the following unconstrained optimization problem:

$$\min \{f(x) \mid x \in R^n\}, \quad (1)$$

where  $f(x) : R^n \rightarrow R$  is a smooth, nonlinear function whose gradient will be denoted by  $g(x)$ . The iterative formula of the conjugate gradient method is given by

$$x_{k+1} = x_k + s_k, \quad s_k = \alpha_k d_k, \quad (2)$$

where  $\alpha_k$  is a step-length which is computed by carrying out a line search, and  $d_k$  is the search direction defined by

$$d_k = \begin{cases} -g_k & \text{if } k = 1 \\ -g_k + \beta_k d_{k-1} & \text{if } k \geq 2, \end{cases} \quad (3)$$

where  $\beta_k$  is a scalar and  $g_k$  denotes the gradient  $\nabla f(x_k)$ . The different conjugate gradient methods correspond to different computing ways of  $\beta_k$ . If  $f$  is a strictly convex quadratic function, namely,

$$f(x) = \frac{1}{2}x^T Hx + b^T x, \quad (4)$$

where  $H$  is a positive definite matrix, and if  $\alpha_k$  is the exact one-dimensional minimizer along the direction  $d_k$ , then the method with (2) and (3) is called linear conjugate gradient method. Otherwise, it is called nonlinear conjugate gradient method. The most important feature of linear conjugate gradient method is that the search directions satisfy the following conjugacy condition:

$$d_i^T H d_j = 0, \quad i \neq j. \quad (5)$$

For nonlinear conjugate gradient methods, (5) does not hold, since the Hessian  $\nabla^2 f(x)$  changes at different iterations.

Some well-known formulae for  $\beta_k$  are the Fletcher-Reeves (FR), Polak-Ribière (PR), Hestense-Stiefel (HS), and Dai-Yuan (DY), which are given, respectively, by

$$\begin{aligned}\beta_k^{\text{FR}} &= \frac{\|g_k\|^2}{\|g_{k-1}\|^2}; \\ \beta_k^{\text{PR}} &= \frac{g_k^T(g_k - g_{k-1})}{\|g_{k-1}\|^2}; \\ \beta_k^{\text{HS}} &= \frac{g_k^T(g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1}}; \\ \beta_k^{\text{DY}} &= \frac{\|g_k\|^2}{(g_k - g_{k-1})^T d_{k-1}},\end{aligned}\quad (6)$$

where  $\|\cdot\|$  denotes the Euclidean norm. Their corresponding conjugate methods are abbreviated as FR, PR, HS, and DY methods. In the past two decades, the convergence properties of these methods have been intensively studied by many researchers (e.g., [1–9]). Although all these methods are equivalent in the linear case, namely, when  $f$  is a strictly convex quadratic function and  $\alpha_k$  are determined by exact line search, their behaviors for general objective functions may be far different.

For general functions, Zoutendijk [10] proved the global convergence of FR method with exact line search. (Here and throughout this paper, for global convergence, we mean that the sequence generated by the corresponding methods will either terminate after finite steps or contain a subsequence such that it converges to a stationary point of the objective function from a given initial point.) Although one would be satisfied with its global convergence properties, the FR method performs much worse than the PR and HS methods in real computations. Powell [11] analyzed a major numerical drawback of the FR method; namely, if a small step is generated away from the solution point, the subsequent steps may be also very short. On the other hand, in practical computation, the HS method resembles the PR method, and both methods are generally believed to be the most efficient conjugate gradient methods since these two methods essentially perform a restart if a bad direction occurs. However, Powell [12] constructed a counterexample and showed that the PR and HS methods without restarts can cycle infinitely without approaching the solution. This example suggests that these two methods have a drawback that they are not globally convergent for general functions. Therefore, over the past few years, much effort has been put to find out new formulae for conjugate methods such that they are not only globally convergent for general functions but also have robust and efficient numerical performance.

Recently, using a new conjugacy condition, Dai and Liao [13] proposed two new methods. Interestingly, one of their methods is not only globally convergent for general functions but also performs better than HS and PR methods. In this paper, similar to Dai and Liao's approach, we propose a new conjugacy condition. Based on the proposed condition, a new formula for computing  $\beta_k$  is given. And then, we analyze

the convergence properties for the given method and also carry the numerical experiment which shows that the given method is robust and efficient.

The remainder of this paper is organized as follows. In Section 2, after a short description of Dai and Liao's conjugacy condition and related methods, the motivations of this paper are represented. According to the motivations, we propose a new conjugacy condition and related method at the end of Section 2. In Section 3, convergence analysis for the given method is presented. In the last Section we perform the numerical experiments by testing a set of large-scale problems and do some numerical comparisons with some existing methods.

## 2. Motivations, New Conjugacy Condition, and Related Method

**2.1. Dai-Liao's Methods.** It is well-known that the linear conjugate gradient methods generate a sequence of search directions  $d_k$  such that conjugacy condition (5) holds. Denote  $y_{k-1}$  to be the gradient change, which means that

$$y_{k-1} = g_k - g_{k-1}. \quad (7)$$

For a general nonlinear function  $f$ , we know by the mean value theorem that there exists some  $t \in (0, 1)$  such that

$$y_{k-1}^T d_k = \alpha_{k-1} d_k^T \nabla^2 f(x_{k-1} + t\alpha_{k-1} d_{k-1}) d_{k-1}. \quad (8)$$

Therefore, it is reasonable to replace (5) with the following conjugacy condition:

$$y_{k-1}^T d_k = 0. \quad (9)$$

Recently, extension of (9) has been studied by Dai and Liao in [13]. Their approach is based on the quasi-Newton techniques. Recall that, in the quasi-Newton method, an approximation matrix  $H_{k-1}$  of the Hessian  $\nabla^2 f(x_{k-1})$  is updated such that the new matrix  $H_k$  satisfies the following quasi-Newton equation:

$$H_k s_{k-1} = y_{k-1}. \quad (10)$$

The search direction  $d_k$  in quasi-Newton method is calculated by

$$d_k = -H_k^{-1} g_k. \quad (11)$$

Combining these two equations, we obtain

$$d_k^T y_{k-1} = d_k^T (H_k s_{k-1}) = -g_k^T s_{k-1}. \quad (12)$$

The above relation implies that (9) holds if the line search is exact since in this case  $g_k^T d_{k-1} = 0$ . However, practical numerical algorithms normally adopt inexact line searches instead of exact line search. For this reason, it seems more reasonable to replace conjugacy condition (9) with the condition

$$d_k^T y_{k-1} = -t g_k^T s_{k-1} \quad t \geq 0, \quad (13)$$

where  $t \geq 0$  is a scalar.

To ensure that the search direction  $d_k$  satisfies conjugacy condition (13), one only needs to multiply (3) with  $y_{k-1}$  and use (13), yielding

$$\beta_k^{\text{DL1}} = \frac{g_k^T (y_{k-1} - t s_{k-1})}{d_{k-1}^T y_{k-1}}. \quad (14)$$

It is obvious that

$$\beta_k^{\text{DL1}} = \beta_k^{\text{HS}} - t \frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}}. \quad (15)$$

For simplicity, we call the method with (2), (3), and (14) as DL1 method. Dai and Liao also proved that the conjugate gradient method with DL1 is globally convergent for uniformly convex functions. For general functions, Powell [12] constructed an example showing that the PR method may cycle without approaching any solution point if the step-length  $\alpha_k$  is chosen to be the first local minimizer along  $d_k$ . Since the DL1 method reduces to the PR method in the case that  $g_k^T d_{k-1} = 0$  holds, this implies that the method with (14) need not converge for general functions. To get the global convergence, Dai and Liao made a restriction on  $\beta_k^{\text{DL1}}$  as follows

$$\beta_k = \max \{ \beta_k^{\text{PR}}, 0 \}. \quad (16)$$

Dai and Liao replaced (14) by

$$\begin{aligned} \beta_k^{\text{DL}} &= \max \left\{ \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, 0 \right\} - t \frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}} \\ &= \max \{ \beta_k^{\text{HS}}, 0 \} - t \frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}}. \end{aligned} \quad (17)$$

We also call the method with (2), (3), and (17) as DL method; Dai and Liao show that DL method is globally convergent for general functions under sufficient descent condition (31) and some suitable conditions. Besides, some numerical experiments in [13] indicate the efficiency of this method.

Similar to Dai and Liao's approach, Li et al. [14] proposed another conjugate condition and related conjugate gradient methods. And they also proved that the proposed methods are globally convergent under some assumptions.

Recently, based on a modified secant condition given by Zhang et al. [15], Yabe and Takano [16] derive an update parameter  $\beta_k^{\text{YT}}$  and show that the YT+ scheme is globally convergent under some conditions:

$$\beta_k^{\text{YT}} = \frac{g_{k+1}^T (z_k - t s_k)}{d_k^T z_k}, \quad (18)$$

where

$$z_k = y_k + \left( \frac{\rho \theta_k}{s_k^T u_k} \right) u_k, \quad (19)$$

$$\theta_k = 6(f_k - f_{k+1}) + 3(g_k + g_{k+1})^T s_k,$$

$\rho \geq 0$  is a constant

$$\beta_k^{\text{YT+}} = \max \left\{ \frac{g_{k+1}^T z_k}{d_k^T z_k}, 0 \right\} - t \frac{g_{k+1}^T s_k}{d_k^T z_k}. \quad (20)$$

**2.2. Motivations.** From the above discussions, Dai and Liao's approach is effective; the main reason is that the search directions  $d_k$  generated by DL1 method or DL method not only contain the gradient information but also contain some Hessian  $\nabla^2 f(x)$  information. From (15) and (17),  $\beta_k^{\text{DL1}}$  and  $\beta_k^{\text{DL}}$  are formed by two parts; the first part is  $\beta_k^{\text{HS}}$  and the second part is  $-t(g_k^T s_{k-1}/d_{k-1}^T y_{k-1})$ . So we also consider DL1 and DL methods as the modified forms of the HS method by adding some information of Hessian  $\nabla^2 f(x)$  which is contained in the second part.

From the structure of (17), we know that the parameter  $\beta_k^{\text{DL}}$  may be negative since the second part  $-t(g_k^T s_{k-1}/d_{k-1}^T y_{k-1})$  may be less than zero. In conjugate gradient methods, if the  $\beta_k < 0$  and  $|\beta_k|$  is large, then the generated directions  $d_k$  and  $d_{k-1}$  may tend to be opposite. This type of methods is susceptible to jamming.

On the other hand, in conjugate gradient methods, the following strong Wolfe-Powell line search is often used to determine the step size  $\alpha_k$ :

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k, \quad (21)$$

$$|g(x_k + \alpha_k d_k)^T d_k| \leq \sigma |g_k^T d_k|, \quad (22)$$

where  $0 < \delta < \sigma < 1$ ; a typical choice of  $\sigma$  is  $\sigma = 0.1$ . From the structure of (17), we know that  $\beta_k^{\text{DL}}$  depends on the directional derivative  $g_k^T d_{k-1}$  which is determined by the line search. For PRP+ algorithm with the strong Wolfe-Powell line search, in order to make sufficient descent condition (31) hold, people often used Lemarechal [17], Fletcher [18], or Moré and Thuente's [19] strategy to make the directional derivative  $|g_k^T d_{k-1}|$  sufficiently small. Under this strategy, the second part of  $\beta_k^{\text{DL}}$  will tend to vanish. This means that the DL method is much line-search-dependent.

The above discussions motivate us to propose a modified conjugacy condition and the related conjugate gradient method, which should possess the following properties

- (1) Nonnegative property  $\beta_k \geq 0$ .
- (2) The new formula contains not only the gradient information but also some Hessian information.
- (3) The formula should be less line-search-dependent.

**2.3. The Modified Conjugacy Condition and Related Method.** From the above discussion, it seems reasonable to replace conjugacy condition (13) with the following modified conjugacy condition:

$$d_k^T y_{k-1} = -t g_{k-1}^T s_{k-1}, \quad t > 0. \quad (23)$$

To ensure that the search direction  $d_k$  satisfies condition (23), one only needs to multiply (3) with  $y_{k-1}$  and use (23), yielding

$$\beta_k^{\text{MDL}} = \frac{g_k^T y_{k-1} - t g_{k-1}^T s_{k-1}}{d_{k-1}^T y_{k-1}}. \quad (24)$$

It is obvious that

$$\beta_k^{\text{MDL}} = \beta_k^{\text{HS}} - t \frac{g_{k-1}^T s_{k-1}}{d_{k-1}^T y_{k-1}}. \quad (25)$$

For simplicity, we call the method with (2), (3), and (25) as MDL method. Similar to Gilbert and Nocedal's [4] approach, we propose the following restricted parameter  $\beta_k^{\text{MDL}+}$ :

$$\beta_k^{\text{MDL}+} = \max\{\beta_k^{\text{HS}}, 0\} - t \frac{g_{k-1}^T s_{k-1}}{d_{k-1}^T y_{k-1}}. \quad (26)$$

And we call the method with (2), (3), and (26) as MDL+ method and give the nonlinear conjugate gradient algorithm as below.

*Algorithm 1* (MDL+). *Step 1.* Given  $x_1 \in R^n$ ,  $\varepsilon \geq 0$ , set  $d_1 = -g_1$ ,  $k = 1$  if  $\|g_1\| \leq \varepsilon$ , then stop.

*Step 2.* Compute  $\alpha_k$  by the strong Wolfe-Powell line search.

*Step 3.* Let  $x_{k+1} = x_k + \alpha_k d_k$ ,  $g_{k+1} = g(x_{k+1})$ ; if  $\|g_{k+1}\| \leq \varepsilon$ , then stop.

*Step 4.* Compute  $\beta_k$  by (26) and generate  $d_{k+1}$  by (3).

*Step 5.* Set  $k := k + 1$  and go to Step 2.

### 3. Convergence Analysis

In the convergence analysis of conjugate gradient methods, we often make the following basic assumptions on the objective functions.

*Assumption A.* (i) The level set  $\Gamma = \{x \in R^n : f(x) \leq f(x_1)\}$  is bounded; namely, there exists a constant  $B > 0$  such that

$$\|x\| \leq B, \quad \forall x \in \Gamma. \quad (27)$$

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

$$\|g(x) - g(y)\| \leq L \|x - y\|, \quad \forall x, y \in N. \quad (28)$$

Under the above assumptions of  $f$ , there exists a constant  $\bar{\gamma} \geq 0$  such that

$$\|\nabla f(x)\| \leq \bar{\gamma}, \quad \forall x \in \Gamma. \quad (29)$$

We say the descent condition holds if for each search direction  $d_k$ ,

$$g_k^T d_k < 0 \quad \forall k \geq 1. \quad (30)$$

In addition, we say the sufficient descent condition holds if there exists a constant  $c > 0$  such that, for each search direction  $d_k$ , we have

$$g_k^T d_k \leq -c \|g_k\|^2 \quad \forall k \geq 1. \quad (31)$$

Under Assumption A, based on the Zoutendijk condition in [10], for any conjugate gradient method with the strong Wolfe-Powell line search, Dai et al. in [20] proved the following general result.

**Lemma 2.** *Suppose that Assumption A holds. Consider any conjugate gradient method in the form (2)-(3), where  $d_k$  is a descent direction and  $\alpha_k$  is obtained by the strong Wolfe-Powell line search. If*

$$\sum_{k \geq 1} \frac{1}{\|d_k\|^2} = \infty, \quad (32)$$

*we have that*

$$\liminf_{k \rightarrow \infty} \|g_k\| = 0. \quad (33)$$

If the objective functions are uniformly convex, we can prove that the norm of  $d_k$  generated by Algorithm 1 (MDL+) is bounded above. Thus by Lemma 2 we immediately have the following result.

**Theorem 3.** *Suppose that Assumption A holds. Consider MDL+ method, where  $d_k$  is a descent direction and  $\alpha_k$  is obtained by the strong Wolfe-Powell line search. If the objective functions are uniformly convex, namely, there exists a constant  $\mu > 0$  such that*

$$(\nabla f(x) - \nabla f(y))^T (x - y) \geq \mu \|x - y\|^2, \quad \forall x, y \in \Gamma, \quad (34)$$

*we have that*

$$\lim_{k \rightarrow \infty} \|g_k\| = 0. \quad (35)$$

*Proof.* It follows from (34) that

$$d_{k-1}^T y_{k-1} \geq \mu \alpha_{k-1} \|d_{k-1}\|^2. \quad (36)$$

By (3), (24), (28), (29), and (36), we have that

$$\begin{aligned} \|d_k\| &\leq \|g_k\| + |\beta_k^{\text{MDL}}| \|d_{k-1}\| \\ &\leq \|g_k\| + \left| \frac{g_k^T y_{k-1} - t g_{k-1}^T s_{k-1}}{\mu \alpha_{k-1} \|d_{k-1}\|^2} \right| \|d_{k-1}\| \\ &\leq \|g_k\| + \frac{\|g_k\| \|g_k - g_{k-1}\| + t \|g_{k-1}^T s_{k-1}\|}{\mu \alpha_{k-1} \|d_{k-1}\|^2} \|d_{k-1}\| \\ &\leq \|g_k\| + \frac{\|g_k\| L \|s_{k-1}\| + t \|g_{k-1}\| \|s_{k-1}\|}{\mu \alpha_{k-1} \|d_{k-1}\|^2} \|d_{k-1}\| \\ &\leq \bar{\gamma} + \frac{\bar{\gamma} L \|s_{k-1}\| + t \bar{\gamma} \|s_{k-1}\|}{\mu \alpha_{k-1} \|d_{k-1}\|^2} \|d_{k-1}\| \\ &= \bar{\gamma} \mu^{-1} (\mu + L + t), \end{aligned} \quad (37)$$

which implies the truth of (32). Therefore, by Lemma 2, we have (33), which is equivalent to (35) for uniformly convex functions. The proof is completed.  $\square$

For the method with  $\beta_k^{\text{MDL}+}$ , if descent condition (30) holds and  $\alpha_k$  satisfies Wolfe-Powell condition (22), then the parameter  $\beta_k^{\text{MDL}+}$  is nonnegative.

**Lemma 4.** *In any conjugate gradient method, if the parameter  $\beta_k$  is computed by (26) and  $d_k$  is a descent direction, when  $\alpha_k$  is determined by strong Wolfe-Powell line search conditions (21) and (22), then*

$$\beta_k^{\text{MDL}+} \geq 0. \quad (38)$$

*Proof.* By SWP condition (22), we have  $d_{k-1}^T y_{k-1} \geq \sigma g_{k-1}^T d_{k-1} - g_{k-1}^T d_{k-1} \geq 0$  and  $g_{k-1}^T s_{k-1} < 0$ . So we have

$$\begin{aligned} \beta_k^{\text{MDL}+} &= \max \left\{ \beta_k^{\text{HS}}, 0 \right\} - t \frac{g_{k-1}^T s_{k-1}}{d_{k-1}^T y_{k-1}}, \\ &\geq 0 \end{aligned} \quad (39)$$

The proof is completed.  $\square$

**Theorem 5.** *Suppose that Assumption A holds. Consider MDL+ method, where  $d_k$  is a sufficient descent direction and  $\alpha_k$  is obtained by strong Wolfe-Powell line search. If there exists a constant  $\gamma > 0$  such that*

$$\|g_k\| \geq \gamma, \quad \forall k \geq 1, \quad (40)$$

then  $d_k \neq 0$  and

$$\sum_{k \geq 2} \|u_k - u_{k-1}\|^2 < \infty, \quad (41)$$

where  $u_k = d_k / \|d_k\|$ .

*Proof.* First, note that  $d_k \neq 0$ ; otherwise (31) is false. Therefore  $u_k$  is well defined. In addition, by relation (40) and Lemma 2, we have that

$$\sum_{k \geq 1} \frac{1}{\|d_k\|^2} < \infty. \quad (42)$$

Now, we divide formula  $\beta_k^{\text{MDL}+}$  into two parts as follows:

$$\beta_k^1 = \max \left\{ \beta_k^{\text{HS}}, 0 \right\}, \quad \beta_k^2 = -t \frac{g_{k-1}^T s_{k-1}}{d_{k-1}^T y_{k-1}}, \quad (43)$$

and define

$$r_k := \frac{\vartheta_k}{\|d_k\|}, \quad \delta_k := \beta_k^1 \frac{\|d_{k-1}\|}{\|d_k\|}, \quad (44)$$

where  $\vartheta_k = -g_k + \beta_k^2 d_{k-1}$ .

Then by (3) we have, for all  $k \geq 2$ ,

$$u_k = r_k + \delta_k u_{k-1}. \quad (45)$$

Using the identity  $\|u_k\| = \|u_{k-1}\| = 1$  and (45), we can obtain

$$\|u_k - \delta_k u_{k-1}\|^2 = 1 + \delta_k^2 - 2\delta_k u_k^T u_{k-1}, \quad (46)$$

$$\|\delta_k u_k - u_{k-1}\|^2 = 1 + \delta_k^2 - 2\delta_k u_k^T u_{k-1}, \quad (47)$$

$$\|r_k\| = \|u_k - \delta_k u_{k-1}\| = \|\delta_k u_k - u_{k-1}\|. \quad (48)$$

Using the condition  $\delta_k = \max\{\beta_k^{\text{HS}}, 0\}(\|d_{k-1}\|/\|d_k\|) \geq 0$ , the triangle inequality, and (48), we obtain

$$\begin{aligned} \|u_k - u_{k-1}\| &\leq \|(1 + \delta_k) u_k - (1 + \delta_k) u_{k-1}\| \\ &\leq \|u_k - \delta_k u_{k-1}\| + \|\delta_k u_k - u_{k-1}\| \\ &= 2\|r_k\|. \end{aligned} \quad (49)$$

On the other hand, line search condition (22) gives

$$y_{k-1}^T d_{k-1} \geq (\sigma - 1) g_{k-1}^T d_{k-1} > 0. \quad (50)$$

Equations (22), (31), and (50) imply that

$$\left| \frac{g_{k-1}^T d_{k-1}}{d_{k-1}^T y_{k-1}} \right| \leq \frac{1}{1 - \sigma}. \quad (51)$$

It follows from the definition of  $\vartheta_k$ , (27), (29), and (51) that

$$\begin{aligned} \|\vartheta_k\| &\leq \|g_k\| + t \left| \frac{g_{k-1}^T s_{k-1}}{d_{k-1}^T y_{k-1}} \right| \|d_{k-1}\| \\ &= \|g_k\| + t \left| \frac{g_{k-1}^T d_{k-1}}{d_{k-1}^T y_{k-1}} \right| \|s_{k-1}\| \\ &\leq \bar{\gamma} + t \frac{1}{1 - \sigma} 2B. \end{aligned} \quad (52)$$

So we have

$$\begin{aligned} \sum \|u_k - u_{k-1}\|^2 &\leq 4 \sum \|r_k\|^2 \leq 4 \sum \frac{\vartheta_k^2}{\|d_k\|^2} \\ &\leq 4 \left( \bar{\gamma} + t \frac{1}{1 - \sigma} 2B \right)^2 \sum \frac{1}{\|d_k\|^2} \\ &< \infty \end{aligned} \quad (53)$$

and the proof is completed.  $\square$

Gilbert and Nocedal [4] introduced property (\*) which is very important for the convergence analysis of the conjugate gradient methods. In fact, with Assumption A, (40), and (50), if (31) holds with some constant  $c > 0$ , the method with  $\beta_k^{\text{MDL}+}$  possesses such property (\*).

**Property 1** (\*). Consider a method of forms (2) and (3). Suppose that

$$0 < \gamma \leq \|g_k\| \leq \bar{\gamma} \quad \forall k \geq 1. \quad (54)$$



We say that the method has property (\*), if, for all  $k$ , there exist constants  $b > 1$ ,  $\lambda > 0$  such that  $|\beta_k| \leq b$ , and if  $\|s_{k-1}\| \leq \lambda$ , we have  $|\beta_k| \leq 1/2b$ .

In fact, by (31), (40), and (50), we have

$$\begin{aligned} d_{k-1}^T y_{k-1} &\geq (\sigma - 1) g_{k-1}^T d_{k-1} \geq c(1 - \sigma) \|g_{k-1}\|^2 \\ &\geq (1 - \sigma) c \gamma^2. \end{aligned} \quad (55)$$

Combining (55) with (27) and (28) and (29), we obtain

$$|\beta_k^{\text{MDL+}}| \leq \frac{\bar{\gamma} L \|s_{k-1}\| + t \bar{\gamma} \|s_{k-1}\|}{(1 - \sigma) c \gamma^2} \leq \frac{2B(L + t) \bar{\gamma}}{(1 - \sigma) c \gamma^2} =: b. \quad (56)$$

Note that  $b$  can be defined such that  $b > 1$ . Therefore we can say that  $b > 1$ . As a result, we define

$$\lambda := \frac{(1 - \sigma) c \gamma^2}{2b(L + t) \bar{\gamma}}, \quad (57)$$

and we get from the first inequality in (56) that if  $\|s_{k-1}\| \leq \lambda$ , then

$$\begin{aligned} |\beta_k^{\text{MDL+}}| &\leq \frac{\bar{\gamma}(L + t) \lambda}{(1 - \sigma) c \gamma^2} \\ &= \frac{\bar{\gamma}(L + t)}{(1 - \sigma) c \gamma^2} \frac{(1 - \sigma) c \gamma^2}{2b(L + t) \bar{\gamma}} = \frac{1}{2b}. \end{aligned} \quad (58)$$

Let  $N^*$  denote the set of positive integers. For  $\lambda > 0$  and a positive integer  $\Delta$ , denote

$$K_{k,\Delta}^\lambda := \{i \in N^* : k \leq i \leq k + \Delta - 1, \|s_{k-1}\| > \lambda\}. \quad (59)$$

Let  $|K_{k,\Delta}^\lambda|$  denote the number of elements in  $K_{k,\Delta}^\lambda$ . From the above property (\*), we can prove the following theorem.

**Theorem 6.** Suppose that Assumption A holds. Consider MDL+ method, where  $d_k$  satisfies condition (31) with  $c > 0$ , and  $\alpha_k$  is obtained by the strong Wolfe-Powell line search. Then if (40) holds, there exists  $\lambda > 0$  such that, for any  $\Delta \in N^*$  and any index  $k_0$ , there is an index  $k \geq k_0$  such that

$$|K_{k,\Delta}^\lambda| > \frac{\Delta}{2}. \quad (60)$$

The proof of this theorem is similar to the proof of Lemma 3.5 in [13]. So, we omit the proof.

According to the above lemmas and theorems, we can prove the following convergence result for the MDL+ method.

**Theorem 7.** Suppose that Assumption A holds. Consider MDL+ method, where  $d_k$  satisfies condition (31) with  $c > 0$ , and  $\alpha_k$  is obtained by the strong Wolfe-Powell line search. Then we have  $\liminf_{k \rightarrow \infty} \|g_k\| = 0$ .

*Proof.* We proceed by contradiction. If  $\liminf_{k \rightarrow \infty} \|g_k\| > 0$ , then (40) must hold. Then the conditions of Theorem 6 hold. Defining  $u_i = d_i / \|d_i\|$ , we have, for any indices  $l, k$ , with  $l \geq k$ ,

$$\begin{aligned} x_l - x_{k-1} &= \sum_{i=k}^l x_i - x_{i-1} \\ &= \sum_{i=k}^l \alpha_{i-1} d_{i-1} = \sum_{i=k}^l u_{i-1} \|s_{i-1}\| \\ &= \sum_{i=k}^l \|s_{i-1}\| u_{k-1} + \sum_{i=k}^l \|s_{i-1}\| (u_{i-1} - u_{k-1}). \end{aligned} \quad (61)$$

Consider  $\|u_i\| = 1$ ; (27) and (61) give that

$$\begin{aligned} \sum_{i=k}^l \|s_{i-1}\| &\leq \|x_l - x_{k-1}\| + \sum_{i=k}^l \|s_{i-1}\| \|u_{i-1} - u_{k-1}\| \\ &\leq 2B + \sum_{i=k}^l \|s_{i-1}\| \|u_{i-1} - u_{k-1}\|. \end{aligned} \quad (62)$$

Let  $\lambda > 0$  be given by Theorem 6 and define  $\Delta := \lceil 8B/\lambda \rceil$  to be the smallest integer not less than  $8B/\lambda$ . By Theorem 6, we can find an index  $k_0 \geq 1$  such that

$$\sum_{i \geq k_0} \|u_{i-1} - u_{k-1}\|^2 \leq \frac{1}{4\Delta}. \quad (63)$$

With this  $\Delta$  and  $k_0$ , Theorem 6 gives an index  $k \geq k_0$  such that

$$|K_{k,\Delta}^\lambda| > \frac{\Delta}{2}. \quad (64)$$

For any index  $i \in [k, k + \Delta - 1]$ , by Cauchy-Schwarz, the geometric inequalities, and (63),

$$\begin{aligned} \|u_i - u_{k-1}\| &\leq \sum_{j=k}^i \|u_j - u_{j-1}\| \\ &\leq (i - k + 1)^{1/2} \left( \sum_{j=k}^i \|u_j - u_{j-1}\|^2 \right)^{1/2} \\ &\leq \Delta^{1/2} \left( \frac{1}{4\Delta} \right)^{1/2} = \frac{1}{2}. \end{aligned} \quad (65)$$

From relations (64) and (65), by taking  $l = k + \Delta - 1$  in (62), we get

$$2B \geq \frac{1}{2} \sum_{i=k}^{k+\Delta-1} \|s_{i-1}\| > \frac{\lambda}{2} |K_{k,\Delta}^\lambda| > \frac{\lambda \Delta}{4}. \quad (66)$$

Thus  $\Delta < 8B/\lambda$ , which contradicts the definition of  $\Delta$ . The proof is completed.  $\square$

#### 4. Numerical Results

In this section, we report the performance of Algorithm 1 (MDL+) on a set of test problems. The codes were written in Fortran 77 and in double precision arithmetic. All the tests were performed on the same PC (Intel Core i3 CPU M370 @ 2.4 GHz, 2 GB RAM). The experiments were performed on a set of 73 nonlinear unconstrained problems collected by Neculai Andrei. Some of the problems are from CUTE [21] library. For each test problem, we have performed 10 numerical experiments with a number of variables  $n = 1000, 2000, \dots, 10000$ .

In order to assess the reliability of the MDL+ algorithm, we also tested this method against the DL method and HS method using the same problems. All these algorithms terminate when  $\|g_k\| \leq 10^{-5}$ . We also force the routines to stop if the iterations exceed 1000 or the number of function evaluations reaches 2000. The parameters  $\Delta$  and  $\sigma$  in Wolfe-Powell line search conditions (21) and (22) are set to be  $10^{-4}$  and  $10^{-1}$  respectively. For DL method,  $t = 0.1$ , which is the same with [13]. We also test MDL+ algorithm with different parameters  $t$  to see that  $t = 0.05$  is the best choice.

The comparing data contain the iterations, function, and gradient evaluations and CPU time. To approximatively assess the performance of MDL+, HS, and DL methods, we use the profile of Dolan and Moré [22] as an evaluated tool.

Dolan and Moré [22] gave a new tool to analyze the efficiency of algorithms. They introduced the notion of a performance profile as a means to evaluate and compare the performance of the set of solvers  $S$  on a test set  $P$ . Assuming that there exist  $n_s$  solvers and  $n_p$  problems, for each problem  $p$  and solver  $s$ , they defined  $t_{p,s}$  = computing cost (iterations or function and gradient evaluations or CPU time) required to solve problem  $p$  by solver  $s$ .

Requiring a baseline for comparisons, they compared the performance on problem  $p$  by solver  $s$  with the best performance by any solver on this problem; that is, using the performance ratio

$$r_{p,s} = \frac{t_{p,s}}{\min \{t_{p,s} : s \in S\}}. \quad (67)$$

Suppose that a parameter  $M \geq r_{p,s}$  for all  $p, s$ . Set  $r_{p,s} = M$  if and only if solver  $s$  does not solve problem  $p$ . Then they defined

$$\rho_s(\tau) = \frac{1}{n_p} \text{size} \{p \in P : r_{p,s} \leq \tau\}. \quad (68)$$

Thus  $\rho_s(\tau)$  is the probability for solver  $s$  that a performance ratio  $r_{p,s}$  is within factor  $\tau \geq 1$  of the best possible ratio. Then function  $\rho_s$  is the distribution function for the performance ratio. The performance profile  $\rho_s$  is a nondecreasing, piecewise constant function. That is, for subset of the methods being analyzed, we plot the fraction  $P$  of the problems for which any given method is within a factor  $\tau$  of the best.

For the testing problems, if all three methods can not terminate successfully, then we got rid of it. In case one method fails, but there is another method that terminates successfully, then the performance ratio of the failed method

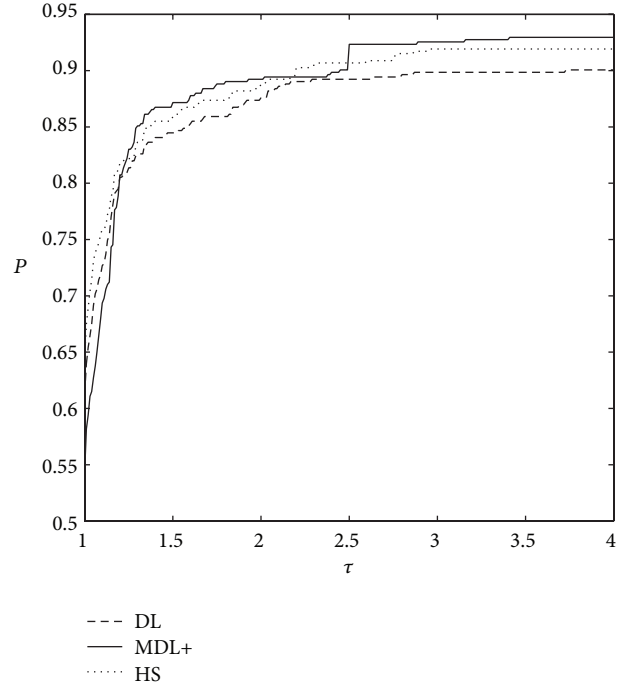


FIGURE 1: Iterations.

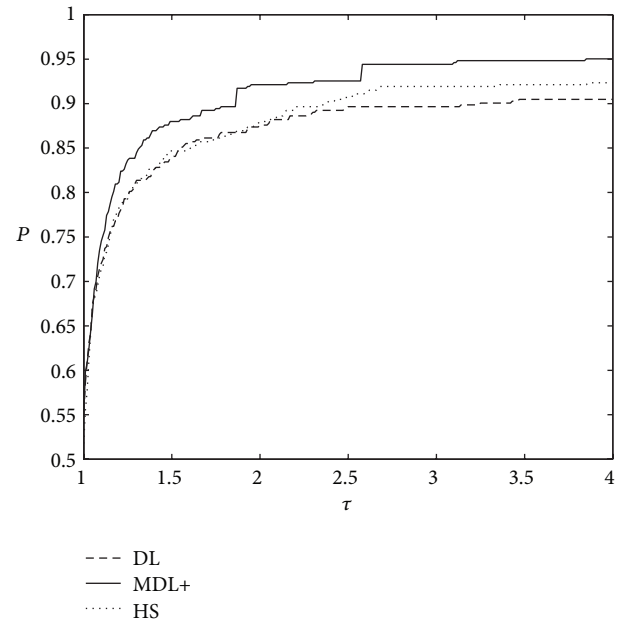


FIGURE 2: Function and gradient evaluations.

is set to be  $M$  ( $M$  is the maxima of the performance ratios). The performance profiles based on iterations, function and gradient evaluations, and CPU time of the three methods are plotted in Figures 1, 2, and 3, respectively.

From Figure 1, which plots the performance profile based on iterations, when  $\tau = 1$ , the HS method performs better than MDL+ and DL methods. With the increasing of  $\tau$ , when  $\tau \geq 1.3$ , the profile of MDL+ method outperforms HS and DL

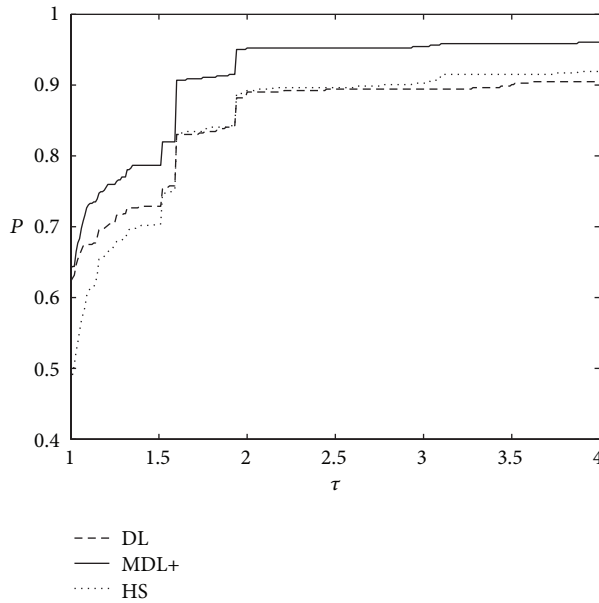


FIGURE 3: CPU time.

methods. This means that, from the iteration points of view, for a subset of problems, HS method is better than MDL+ and DL methods. But, for all the testing problems, MDL+ method is much robust than HS and DL methods.

From Figure 2, which plots the performance profile based on function and gradient evaluations, it is easy to see that, for all  $\tau \geq 1$ , MDL+ method performs much better than HS and DL methods. It is an interesting phenomenon, since, when  $\tau \leq 1.3$ , the profiles of HS based on iterations outperform MDL+ method. This means that, during process of iteration, the required function and gradient evaluations of MDL+ method are much less than HS and DL methods. From this point of view, the CPU time consumed by MDL+ method should be much less than HS and DL methods, since the CPU time is mainly dependent on function and gradient evaluations. Figure 3 validates that the CPU time consumed by MDL+ method is much less than HS and DL methods.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

This research was supported by Guangxi High School Foundation Grant (Grant no. 2013BYB210) and Guangxi University of Finance and Economics Science Foundation Grant no. 2013A015.

## References

- [1] M. Al-baali, "Descent property and global convergence of the Fletcher-Reeves method with inexact line search," *IMA Journal of Numerical Analysis*, vol. 5, no. 1, pp. 121–124, 1985.
- [2] Y. H. Dai and Y. Yuan, "Convergence properties of the Fletcher-Reeves method," *IMA Journal of Numerical Analysis*, vol. 16, no. 2, pp. 155–164, 1996.
- [3] W. W. Hager and H. Zhang, "A survey of nonlinear conjugate gradient methods," *Pacific Journal of Optimization*, vol. 2, pp. 35–58, 2006.
- [4] J. C. Gilbert and J. Nocedal, "Global convergence properties of conjugate gradient methods for optimization," *SIAM Journal on Optimization*, vol. 2, no. 1, pp. 21–42, 1992.
- [5] L. Grippo and S. Lucidi, "A globally convergent version of the Polak-Ribière conjugate gradient method," *Mathematical Programming B*, vol. 78, no. 3, pp. 375–391, 1997.
- [6] L. Grippo and S. Lucidi, "Convergence conditions, line search algorithms and trust region implementations for the Polak-Ribière conjugate gradient method," *Optimization Methods and Software*, vol. 20, no. 1, pp. 71–98, 2005.
- [7] J. Sun and J. Zhang, "Global convergence of conjugate gradient methods without line search," *Annals of Operations Research*, vol. 103, no. 1–4, pp. 161–173, 2001.
- [8] Z. X. Wei, G. Y. Li, and L. Q. Qi, "Global convergence of the Polak-Ribière-Polyak conjugate gradient method with an Armijo-type inexact line search for nonconvex unconstrained optimization problems," *Mathematics of Computation*, vol. 77, no. 264, pp. 2173–2193, 2008.
- [9] L. Zhang, W. Zhou, and D. Li, "Global convergence of a modified Fletcher-Reeves conjugate gradient method with Armijo-type line search," *Numerische Mathematik*, vol. 104, no. 4, pp. 561–572, 2006.
- [10] G. Zoutendijk, "Nonlinear programming computational methods," in *Integer and Nonlinear Programming*, J. J. Moré, Ed., pp. 37–86, North-Holland, Amsterdam, The Netherlands, 1970.
- [11] M. J. D. Powell, "Restart procedures for the conjugate gradient method," *Mathematical Programming*, vol. 12, no. 1, pp. 241–254, 1977.
- [12] M. J. D. Powell, "Nonconvex minimization Calculations and the Conjugate Gradient Method," in *Numerical Analysis*, vol. 1066 of *Lecture Notes in Mathematics*, pp. 122–141, Springer, Berlin, Germany, 1984.
- [13] Y.-H. Dai and L.-Z. Liao, "New conjugacy conditions and related nonlinear conjugate gradient methods," *Applied Mathematics and Optimization*, vol. 43, no. 1, pp. 87–101, 2001.
- [14] G. Li, C. Tang, and Z. Wei, "New conjugacy condition and related new conjugate gradient methods for unconstrained optimization," *Journal of Computational and Applied Mathematics*, vol. 202, no. 2, pp. 523–539, 2007.
- [15] J. Z. Zhang, N. Y. Deng, and L. H. Chen, "New quasi-Newton equation and related methods for unconstrained optimization," *Journal of Optimization Theory and Applications*, vol. 102, no. 1, pp. 147–167, 1999.
- [16] H. Yabe and M. Takano, "Global convergence properties of nonlinear conjugate gradient methods with modified secant condition," *Computational Optimization and Applications*, vol. 28, no. 2, pp. 203–225, 2004.
- [17] C. Lemarechal, "A view of line searches," in *Optimization and Optimal Control*, A. Auslander, W. Oettli, and J. Stoer, Eds., Lecture Notes in Control and Information Science 30, pp. 59–78, Springer, Berlin, Germany, 1981.
- [18] R. Fletcher, *Practical Methods of Optimization Vol. 1: Unconstrained Optimization*, John Wiley & Sons, New York, NY, USA, 1987.



- [19] J. J. Moré and D. J. Thuente, *On Line Search Algorithms with Guaranteed Sufficient Decrease*, Mathematics and Computer Science Division Preprint MCS-P330-1092, Argonne National Laboratory, 1990.
- [20] Y. Dai, J. Han, G. Liu, D. Sun, H. Yin, and Y.-X. Yuan, “Convergence properties of nonlinear conjugate gradient methods,” *SIAM Journal on Optimization*, vol. 10, no. 2, pp. 345–358, 2000.
- [21] I. Bongartz, A. R. Conn, N. Gould, and P. L. Toint, “CUTE: constrained and unconstrained testing environment,” *ACM Transactions on Mathematical Software*, vol. 21, no. 1, pp. 123–160, 1995.
- [22] E. D. Dolan and J. J. Moré, “Benchmarking optimization software with performance profiles,” *Mathematical Programming B*, vol. 91, no. 2, pp. 201–213, 2002.

