Research Article

A Conjugate Gradient Method with Global Convergence for Large-Scale Unconstrained Optimization Problems

Shengwei Yao,[1,2] Xiwen Lu,[1] and Zengxin Wei[3]

[1] *School of Science, East China University of Science and Technology, Shanghai 200237, China*
[2] *School of Information and Statistics, Guangxi University of Finance and Economics, Nanning 530003, China*
[3] *College of Mathematics and Information Science, Guangxi University, Nanning 530004, China*

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

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. This paper proposes a conjugate gradient method which is similar to Dai-Liao conjugate gradient method (Dai and Liao, 2001) but has stronger convergence properties. The given method possesses the sufficient descent condition, and is globally convergent under strong Wolfe-Powell (SWP) line search for general function. Our numerical results show that the proposed method is very efficient for the test problems.

1. Introduction

The conjugate gradient (CG) method has played a special role in solving large-scale nonlinear optimization 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 algorithms for nonlinear problems available today, but it remains very popular for engineers and mathematicians who are interested in solving large problems. The conjugate gradient method is designed to solve the following unconstrained optimization problem:

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

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

$$x_{k+1} = x_k + s_k, \qquad s_k = \alpha_k d_k, \tag{2}$$

where α_k is a step length which is computed by carrying out a line search, and d_k is the search direction defined by

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

where β_k is a scalar and g_k denotes the gradient $\nabla f(x_k)$. If f is a strictly convex quadratic function, namely,

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

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

$$d_i^T H d_j = 0, \quad i \neq j. \tag{5}$$

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

Some well-known formulas for β_k are the Fletcher-Reeves (FR), Polak-Ribière (PR), Hestense-Stiefel (HS), and Dai-Yuan (DY) methods which are given, respectively, by

$$\beta_k^{\mathrm{FR}} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \tag{6}$$

$$\beta_k^{\mathrm{PR}} = \frac{g_k^T \left(g_k - g_{k-1}\right)}{\|g_{k-1}\|^2}, \tag{7}$$

$$\beta_k^{\mathrm{HS}} = \frac{g_k^T \left(g_k - g_{k-1}\right)}{\left(g_k - g_{k-1}\right)^T d_{k-1}}, \tag{8}$$

$$\beta_k^{\mathrm{DY}} = \frac{\|g_k\|^2}{\left(g_k - g_{k-1}\right)^T d_{k-1}}, \tag{9}$$

where $\| \ \|$ denotes the Euclidean norm. Their corresponding conjugate methods are abbreviated as FR, PR, HS, and DY methods. Although all these method are equivalent in the linear case, namely, when f is a strictly convex quadratic function and α_k are determined by exact line search, their behaviors for general objective functions may be far different.

For general functions, Zoutendijk [1] proved the global convergence of FR methods 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 (HS) method in real computations. Powell [2] 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 [3] constructed a counterexample and showed that the PR method and HS method 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, in the past two decades, much effort has been exceterd to find out new formulas for conjugate methods such that not only they are globally convergent for general functions but also they have good numerical performance.

Recently, using a new conjugacy condition, Dai and Liao [4] 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 another formula for β_k, 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, we firstly state the corresponding formula which is proposed by Dai and Liao [4] and the motivations of this paper, and then we propose the new nonlinear conjugate gradient method. In Section 3, convergence analysis for the given method is presented. Numerical results are reported in Section 4. Finally, some conclusions are given in Section 5.

2. Motivations and New Nonlinear Conjugate Gradient 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 the conjugacy condition (5) holds. Denote y_{k-1} to be the gradient change, which means that

$$y_{k-1} = g_k - g_{k-1}. \tag{10}$$

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

$$y_{k-1}^T d_k = \alpha_{k-1} d_k^T \nabla^2 f \left(x_{k-1} + t\alpha_{k-1} d_{k-1}\right) d_{k-1}. \tag{11}$$

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

$$y_{k-1}^T d_k = 0. \tag{12}$$

Recently, extension of (12) has been studied by Dai and Liao in [4]. 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:

$$H_k s_{k-1} = y_{k-1}. \tag{13}$$

The search direction d_k in Quasi-Newton method is calculated by

$$d_k = -H_k^{-1} g_k. \tag{14}$$

Combining these two equations, we obtain

$$d_k^T y_{k-1} = d_k^T \left(H_k s_{k-1}\right) = -g_k^T s_{k-1}. \tag{15}$$

The previous relation implies that (12) 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 searches. For this reason, it seems more reasonable to replace the conjugacy condition (12) with the condition

$$d_k^T y_{k-1} = -t g_k^T s_{k-1}, \quad t \geq 0, \tag{16}$$

where $t \geq 0$ is a scalar.

To ensure that the search direction d_k satisfies the conjugate condition (16), one only needs to multiply (3) with y_{k-1} and use (16), yielding

$$\beta_k^{\mathrm{DL1}} = \frac{g_k^T \left(y_{k-1} - t s_{k-1}\right)}{d_{k-1}^T y_{k-1}}. \tag{17}$$

It is obvious that

$$\beta_k^{\text{DL1}} = \beta_k^{\text{HS}} - t \frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}}. \tag{18}$$

For simplicity, we call the method with (2), (3), and (17) as DL1 method. Dai and Liao also prove that the conjugate gradient method with DL1 is globally convergent for uniformly convex functions. For general functions, Powell [3] constructed an example showing that the PR method may cycle without approaching any solution point if the step length α_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 (17) need not converge for general functions. To get the global convergence, like Gilbert and Nocedal [5], who have proved the global convergence of the PR method with the restriction that $\beta_k^{\text{PR}} \geq 0$, Dai and Liao replaced (17) by

$$\begin{aligned}
\beta_k^{\text{DL}} &= \max\left\{ \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, 0 \right\} - t \frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}} \\
&= \max\left\{ \beta_k^{\text{HS}}, 0 \right\} - t \frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}}.
\end{aligned} \tag{19}$$

We also call the method with (2), (3), and (19) as DL method, Dai and Liao show that DL method is globally convergent for general functions under the sufficient descent condition (21) and some suitable conditions. Besides, some numerical experiments in [4] indicate the efficiency of this method.

Similar to Dai and Liao's approach, Li et al. [6] proposed another conjugate condition and related conjugate gradient methods, and they also prove that the proposed methods are globally convergent under some assumptions.

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 (18) and (19), β_k^{DL1} and β_k^{DL} are formed by two parts; the first part is β_k^{HS}, and the second part is $-t(g_k^T s_{k-1}/d_{k-1}^T y_{k-1})$. So, we also can consider DL1 and DL methods as some modified forms of the HS method by adding some information of Hessian $\nabla^2 f(x)$ which is contained in the second part. The convergence properties of the HS method are similar to PR method; it does not converge for general functions even if the line search is exact. In order to get the convergence, one also needs the nonnegative restriction $\beta_k = \max\{\beta_k^{\text{HS}}, 0\}$ and the sufficient descent assumption (21). From the above discussion, the descent condition or sufficient descent condition and nonnegative property of β_k play important roles in the convergence analysis. We say that the descent condition holds if for each search directions d_k

$$g_k^T d_k < 0, \quad \forall k \geq 1. \tag{20}$$

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

$$g_k^T d_k \leq -c \|g_k\|^2, \quad \forall k \geq 1. \tag{21}$$

Motivated by the above ideal, in this paper, we focus on finding the new conjugate gradient method which possesses 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 search directions d_k generated by the proposed method satisfy the sufficient descent conditions (21).

2.3. The New Conjugate Gradient Method.

From the structure of (6), (7), (8), and (9), the PR and HS methods have the common numerator $g_k^T y_{k-1}$, and the FR and DY methods have the common numerator $\|g_k\|^2$; and this different choice makes them have different properties. Generally speaking, FR and DY methods have better convergence properties, and PR and HS methods have better numerical experiments. Powell [3] pointed out that the FR method, with exact line search, was susceptible to jamming. That is, the algorithm could take many short steps without making significant progress to the minimum. If the line search is exact, that means $g_k^T d_{k-1} = 0$, in this case, DY method will turn out to be FR method. So, these two methods have the same disadvantage. The PR and HS methods which share the common numerator $g_k^T y_{k-1}$ possess a built-in restart feature to avoid the jamming problem: when the step $x_k - x_{k-1}$ is small, the factor y_{k-1} in the numerator of β_k tends to zero. Hence, the next search direction d_k is essentially the steepest descent direction $-g_k$. So, the numerical performance of these methods is better than the performance of the methods with $\|g_k\|^2$ in numerator of β_k.

Just as above discussions, great attentions were given to find the methods which not only have global convergent properties but also have nice numerical experiments.

Recently, Wei et al. [7] proposed a new formula

$$\beta_k^{\text{WYL}} = \frac{g_k^T y_{k-1}^*}{\|g_{k-1}\|^2}, \qquad y_{k-1}^* = g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1}. \tag{22}$$

The method with formula β_k^{WYL} not only has nice numerical results but also possesses the sufficient descent condition and global convergence properties under the strong Wolfe-Powell line search. From the structure of β_k^{WYL}, we know that the method with β_k^{WYL} can also avoid jamming: when the step $x_k - x_{k-1}$ is small, $\|g_k\|/\|g_{k-1}\|$ tends to 1 and the next search direction tends to the steepest descent direction which is similar to PR method. But WYL method has some advantages, such as under strong Wolfe-Powell line search, $\beta_k^{\text{WYL}} \geq 0$, and if the parameter $\sigma \leq 1/4$ in SWP, WYL method possesses the sufficient descent condition which deduces the global convergence of the WYL method.

In [8, 9], Shengwei et al. extended such modification to HS method as follows:

$$\beta_k^{\mathrm{MHS}} = \frac{g_k^T y_{k-1}^*}{d_{k-1}^T y_{k-1}}, \qquad y_{k-1}^* = g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1}. \qquad (23)$$

The previous formulae β_k^{WYL} and β_k^{MHS} can be considered as the modification forms of β_k^{PR} and β_k^{HS} by using y_{k-1}^* to replace y_{k-1}, respectively. In [8, 9], the corresponding methods are proved to be globally convergent for general functions under the strong Wolfe-Powell line search and Grippo-Lucidi line search. Based on the same approach, some authors give other discussions and modifications in [10–12]. In fact, y_{k-1}^* is not our point at the beginning, our purpose is involving the information of the angle between g_k and g_{k-1}. From this point of view, β_k^{WYL} has the following form:

$$\beta_k^{\mathrm{WYL}} = \beta_k^{\mathrm{FR}} \left(1 - \cos \left(\overline{\theta}_k \right) \right), \qquad (24)$$

where $\overline{\theta}_k$ is the angle between g_k and g_{k-1}. By multiplying β_k^{FR} with $1 - \cos \overline{\theta}_k$, the method not only has similar convergence properties with FR method, but also avoids jamming which is similar to PR method.

The above analysis motivates us to propose the following formula to compute β_k:

$$\beta_k^{\mathrm{MDL1}} = \max \left\{ \frac{g_k^T y_{k-1}^*}{d_{k-1}^T y_{k-1}}, 0 \right\} - t \frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}}, \qquad (25)$$

where $y_{k-1}^* = g_k - (\|g_k\|/\|g_{k-1}\|) g_{k-1}$. Since the β_k^{MHS} are nonnegative under the strong Wolfe-Powell line search, we omit the nonnegative restriction and propose the following formula:

$$\beta_k^{\mathrm{MDL}} = \frac{g_k^T y_{k-1}^*}{d_{k-1}^T y_{k-1}} - t \frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}} = \beta_k^{\mathrm{MHS}} - t \frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}}. \quad (26)$$

From (25) and (26), we know that we only substitute y_{k-1} in the first part of the numerator of β_k^{DL1} or β_k^{DL} by y_k^*. The reason is that we hope the formulae (25) and (26) contain the angle information between g_k and g_{k-1}. In fact, β_k^{MDL} can be expressed as

$$\begin{aligned} \beta_k^{\mathrm{MDL}} &= \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}} \left(1 - \cos \overline{\theta}_k \right) - t \frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}} \\ &= \beta_k^{\mathrm{DY}} \left(1 - \cos \overline{\theta}_k \right) - t \frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}}. \end{aligned} \qquad (27)$$

For simplicity, we call the method generated by (2), (3), and (26) as MDL method and give the algorithm as follows.

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 t_k by some line searches.

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

Step 4. Compute β_k by (26) and generate d_{k+1} by (3).

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

We 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

$$\|x\| \leq B, \quad \forall x \in \Gamma. \qquad (28)$$

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

$$\|g(x) - g(y)\| \leq L \|x - y\|, \quad \forall x, y \in N. \qquad (29)$$

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

$$\|\nabla f(x)\| \leq \overline{\gamma}, \quad \forall x \in \Gamma. \qquad (30)$$

The step length α_k in Algorithm 1 (MDL) is obtained by some line search scheme. In conjugate gradient methods, the strong Wolfe-Powell conditions; namely,

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k, \qquad (31)$$

$$\left| g(x_k + \alpha_k d_k)^T d_k \right| \leq -\sigma g_k^T d_k, \qquad (32)$$

where $0 < \delta < \sigma < 1$, are often imposed on the line search (SWP).

3. Convergence Analysis

Under Assumption A, based on the Zoutendijk condition in [1], for any conjugate gradient method with the strong Wolfe-Powell line search, Dai et al. in [13] 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 α_k is obtained by the strong Wolfe-Powell line search. If*

$$\sum_{k \geq 1} \frac{1}{\|d_k\|^2} = \infty. \qquad (33)$$

One has that

$$\liminf_{k \to \infty} \|g_k\| = 0. \qquad (34)$$

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

Theorem 3. *Suppose that Assumption A holds. Consider MDL method, where d_k is a descent direction and α_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*

$$\left(\nabla f(x) - \nabla f(y)\right)^T (x - y) \geq \mu \|x - y\|^2, \quad \forall x, y \in \Gamma. \quad (35)$$

One has that

$$\lim_{k \to \infty} \|g_k\| = 0. \quad (36)$$

Proof. It follows from (35) that

$$d_{k-1}^T y_{k-1} \geq \mu \alpha_{k-1} \|d_{k-1}\|^2. \quad (37)$$

By (3), (26), (29), (30), and (37), we have

$$\|d_k\|$$

$$\leq \|g_k\| + \left|\beta_k^{\mathrm{MDL}}\right| \|d_{k-1}\|$$

$$\leq \|g_k\| + \frac{g_k^T \left(g_k - \left(\|g_k\| / \|g_{k-1}\|\right) g_{k-1} - t s_{k-1}\right)}{\mu \alpha_{k-1} \|d_{k-1}\|^2} \|d_{k-1}\|$$

$$= \|g_k\|$$

$$\quad + \frac{g_k^T \left(g_k - g_{k-1} + g_{k-1} - \left(\|g_k\| / \|g_{k-1}\|\right) g_{k-1} - t s_{k-1}\right)}{\mu \alpha_{k-1} \|d_{k-1}\|^2}$$

$$\quad \times \|d_{k-1}\|$$

$$\leq \|g_k\|$$

$$\quad + \frac{\|g_k\| \left(\|g_k - g_{k-1}\| + \|g_{k-1} \left(1 - \|g_k\| / \|g_{k-1}\|\right)\| + t \|s_{k-1}\|\right)}{\mu \alpha_{k-1} \|d_{k-1}\|^2}$$

$$\quad \times \|d_{k-1}\|$$

$$\leq \|g_k\|$$

$$\quad + \frac{\|g_k\| \left(\|g_k - g_{k-1}\| + \|\|g_{k-1}\| - \|g_k\|\| + t \|s_{k-1}\|\right)}{\mu \alpha_{k-1} \|d_{k-1}\|^2} \|d_{k-1}\|$$

$$\leq \|g_k\| + \frac{\|g_k\| \left(2L \|s_{k-1}\| + t \|s_{k-1}\|\right)}{\mu \alpha_{k-1} \|d_{k-1}\|^2} \|d_{k-1}\|$$

$$\leq \overline{\gamma} \left(1 + \frac{2L + t}{\mu}\right) = \overline{\gamma} \mu^{-1} \left(\mu + 2L + t\right), \quad (38)$$

which implies the truth of (33). Therefore, by Lemma 2 we have (34), which is equivalent to (36) for uniformly convex functions. The proof is completed. $\qquad \square$

In order to prove the convergence of the MDL method, we need to state some properties of β_k^{MHS}.

Lemma 4. *In any conjugate gradient methods, if the parameter β_k is computed by (23), namely, $\beta_k = \beta_k^{\mathrm{MHS}}$, and α_k is determined by strong Wolfe-Powell line search of (31) and (32), then*

$$\beta_k^{\mathrm{MHS}} \geq 0. \quad (39)$$

Proof. By SWP condition (32), we have $d_{k-1}^T y_{k-1} \geq \sigma g_{k-1}^T d_{k-1} - g_{k-1}^T d_{k-1} \geq 0$, since $\sigma < 1$ and $g_1^T d_1 = -\|g_1\|^2 < 0$. So we have

$$\beta_k^{\mathrm{MHS}} = \frac{g_k^T \left(g_k - \left(\|g_k\| / \|g_{k-1}\|\right) g_{k-1}\right)}{d_{k-1}^T y_{k-1}}$$

$$= \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}} \left(1 - \cos \overline{\theta}_k\right) \geq 0. \quad (40)$$

The proof is completed. $\qquad \square$

In addition, we can also prove that, in conjugate gradient method of forms (2)-(3), if β_k is computed by β_k^{MDL} (26) and α_k is determined by strong Wolfe-Powell line search, then the search direction d_k satisfies the sufficient descent condition (21).

Theorem 5. *In any conjugate gradient methods, in which the parameter β_k is computed by (26), namely, $\beta_k = \beta_k^{\mathrm{MDL}}$, and α_k is determined by strong Wolfe-Powell line search of (31) and (32), if $\sigma < 1/3$, then the search direction d_k satisfied the sufficient descent condition (21).*

Proof. We prove this theorem by induction. Firstly, we prove the descent condition $d_k^T g_k < 0$ as follow.

Since $g_1^T d_1 = -\|g_1\|^2 < 0$, supposing that $g_i^T d_i < 0$ holds for $i \leq k - 1$, we deduce that the descent condition holds by proving that $g_i^T d_i < 0$ holds for $i = k$ as follow.

By SWP condition (32), we have $d_{k-1}^T y_{k-1} \geq (\sigma - 1) g_{k-1}^T d_{k-1} > 0$. Combining (3) and (26), we have

$$\frac{g_k^T d_k}{\|g_k\|^2} = -1 + \frac{\left(1 - \cos \overline{\theta}_k\right)}{d_{k-1}^T y_{k-1}} g_k^T d_{k-1}$$

$$\quad - t * \frac{\left(g_k^T d_{k-1}\right)^2 \alpha_{k-1}}{\left(d_{k-1}^T y_{k-1}\right) \|g_k\|^2}$$

$$\leq -1 + 2 \frac{-\sigma g_{k-1}^T d_{k-1}}{d_{k-1}^T y_{k-1}} \quad (41)$$

$$\leq -1 + 2 \frac{-\sigma g_{k-1}^T d_{k-1}}{(\sigma - 1) g_{k-1}^T d_{k-1}}$$

$$\leq \frac{-1 + 3\sigma}{1 - \sigma} < 0.$$

Equation (41) means that descent condition holds.

Secondly, we prove the following sufficient descent condition.

Set $c = 1 - 2\sigma/(1 - \sigma)$; since the restriction $\sigma < 1/3$, we have $0 < c < 1$. Combining $g_1^T d_1 = -\|g_1\|^2$ and (41), the sufficient descent condition (21) holds immediately. □

By Theorem 5, we can prove the following Lemma 6.

Lemma 6. *Suppose that Assumption A holds. Consider MDL method, where α_k is obtained by strong Wolfe-Powell lien search with $\sigma < 1/3$. If there exists a constant $\gamma > 0$ such that*

$$\|g_k\| \geq \gamma, \quad \forall k \geq 1, \tag{42}$$

then $d_k \neq 0$ and

$$\sum_{k \geq 2} \|u_k - u_{k-1}\|^2 < \infty, \tag{43}$$

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

Proof. Firstly, note that $d_k \neq 0$; otherwise, (21) is false. Therefore, u_k is well defined. In addition, by relation (42) and Lemma 2, we have

$$\sum_{k \geq 1} \frac{1}{\|d_k\|^2} < \infty. \tag{44}$$

Now, we divide formula β_k^{MDL} into two parts as follows:

$$\beta_k^1 = \frac{g_k^T y_{k-1}^*}{d_{k-1}^T y_{k-1}}, \qquad \beta_k^2 = -t\frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}}, \tag{45}$$

and define

$$r_k := \frac{\vartheta_k}{\|d_k\|}, \qquad \delta_k := \beta_k^1 \frac{\|d_{k-1}\|}{\|d_k\|}, \tag{46}$$

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

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

$$u_k = r_k + \delta_k u_{k-1}. \tag{47}$$

Using the identity $\|u_k\| = \|u_{k-1}\| = 1$ and (47) we can obtain

$$\|r_k\| = \|u_k - \delta_k u_{k-1}\| = \|\delta_k u_k - u_{k-1}\|, \tag{48}$$

using the condition $\delta_k = \beta_k^{\text{MHS}}(\|d_{k-1}\|/\|d_k\|) \geq 0$, the triangle inequality, and (48), it follows that

$$\|u_k - u_{k-1}\| \leq \|(1 + \delta) u_k - (1 + \delta) u_{k-1}\|$$
$$\leq \|u_k - \delta_k u_{k-1}\| + \|\delta_k u_k - u_{k-1}\| \tag{49}$$
$$= 2\|r_k\|.$$

On the other hand, the line search condition (32) gives

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

Equations (50), (32), and (21) imply that

$$\left| \frac{g_k^T d_{k-1}}{d_{k-1}^T y_{k-1}} \right| \leq \frac{\sigma}{1 - \sigma}. \tag{51}$$

It follows from the definition of ϑ_k, (51), (28), and (30) that

$$\|\vartheta_k\| \leq \|\delta_k\| + t \left| \frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}} \right| \|d_{k-1}\|$$
$$= \|\delta_k\| + t \left| \frac{g_k^T d_{k-1}}{d_{k-1}^T y_{k-1}} \right| \|s_{k-1}\| \tag{52}$$
$$\leq \bar{\gamma} + t\frac{\sigma}{1 - \sigma}2B.$$

So, we have

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

The proof is completed. □

Gilbert and Nocedal [5] introduced property ($*$) which is very important for the convergence properties of the conjugate gradient methods. We are going to show that method with β_k^{MDL} possesses such property ($*$).

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

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

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 (50), (21), and (42), we have

$$d_{k-1}^T y_{k-1} \geq (\sigma - 1) g_{k-1}^T d_{k-1} \geq c(1 - \sigma) \|g_{k-1}\|^2$$
$$\geq (1 - \sigma) c\gamma^2. \tag{55}$$

Using this, (28), (29), and (30) we obtain

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

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

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

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

$$\left|\beta_k^{\text{MDL}}\right| \leq \frac{(2L + t)\bar{\lambda}}{(1 - \sigma) c\lambda^2} = \frac{1}{2b}. \tag{58}$$

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

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

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

Lemma 7. *Suppose that Assumption A holds. Consider MDL method, where α_k is obtained by the strong Wolfe-Powell line search in which $\sigma < 1/3$. Then if (42) 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*

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

The proof of this lemma is similar to the proof of Lemma 3.5 in [4]. In [4], authors proved that method with (19) has this property, if the search direction d_k satisfies the sufficient descent condition (21). In our paper, we do not need this assumption, since the directions generated by MDL method with strong Wolfe-Powell line search always possess the sufficient descent condition (21). So, we omit the proof of this lemma.

According to the previous lemmas and theorems, we can prove the following convergence theorem for the MDL.

Theorem 8. *Suppose that Assumption A holds. Consider MDL method, if α_k is obtained by strong Wolfe-Powell line search with $\sigma < 1/3$. Then we have $\lim \inf_{k \to \infty} \|g_k\| = 0$.*

Proof. We proceed by contradiction. If $\lim \inf_{k \to \infty} \|g_k\| > 0$, then (42) must hold. Then the conditions of Lemmas 6 and 7 hold. Defining $u_i = d_i / \|d_i\|$, we have for any indices l, k, with $l \geq k$,

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

Equation (61), $\|u_i\| = 1$, and (28) give

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

Let $\lambda > 0$ be given by Lemma 7, and define $\Delta := \lceil 8B/\lambda \rceil$ to be the smallest integer not less than $8B/\lambda$. By Lemma 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 Δ and k_0, Lemma 7 gives an index $k \geq k_0$ such that

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

For any index $i \in [k, k+\Delta-1]$, by Cauchy-Schwartz inequality 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 these relations (65) and (64) and taking $l = k + \Delta - 1$ in (62), we get

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

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

4. Numerical Results

From (26) and (27), the MDL method can be considered as

(i) form 1: a modification form of DL method;

(ii) form 2: a modification form of MHS method;

(iii) form 3: a modification form of DY method.

In form 1, the β_k^{HS} in β_k^{DL} is replaced by β_k^{MHS}. By this modification, we can guarantee the nonnegativity restrictions in DL method. In form 2, β_k^{MDL} is obtained by β_k^{MHS} adding an adjusting term $-t(g_k^T s_{k-1}/y_{k-1}^T d_{k-1})$ which contains some Hessian information of the objective function. In form 3, $\beta_k^{MDL} = \beta_k^{DY}(1 - \cos \overline{\theta}_k) - t(g_k^T s_{k-1}/d_{k-1}^T y_{k-1})$ shows that β_k^{MDL} is obtained by multiplying β_k^{DY} with $(1 - \cos \overline{\theta}_k)$ and adding the second term $-t(g_k^T s_{k-1}/y_{k-1}^T d_{k-1})$.

From the above convergence analysis, we know that MDL method has stronger convergent properties than DL method, and similar convergent properties with MHS method and DY method. So, in this section, we test the following four CG methods:

(i) MDL method: method of the forms (2) and (3), in which β_k is computed by β_k^{MDL} (26);

(ii) DL method: method of the forms (2) and (3), in which β_k is computed by β_k^{DL} (19);

(iii) MHS method: method of the forms (2) and (3), in which β_k is computed by β_k^{MHS} (23);

(iv) DY method: method of the forms (2) and (3), in which β_k is computed by β_k^{DY} (9).

TABLE 1: Numerical results.

Problem	Dim	MDL	DL	MHS	DY
ROSE	2	35/349/83	F	38/267/91	63/800/106
FROTH	2	18/88/29	9/25/18	15/84/26	16/38/26
BADSCP	2	28/275/64	36/510/96	42/362/96	F
BADSCB	2	26/446/48	F	28/452/50	F
BEALE	3	16/87/27	11/81/22	14/83/25	47/193/74
JENSAM	2	11/31/21	F	11/31/21	11/31/21
HELIX	3	49/347/81	28/164/54	47/390/73	80/406/126
BARD	3	18/38/26	24/145/37	18/86/26	48/148/77
GAUSS	3	4/9/5	3/7/4	4/9/5	4/9/5
SING	4	134/501/209	78/396/124	111/411/172	650/3254/1104
WOOD	4	102/613/182	179/865/306	207/1352/365	F
KOWOSB	4	39/178/66	46/383/72	53/259/88	462/1760/796
BIGGS	6	18/279/25	85/564/14	20/286/31	210/644/342
OSB2	11	268/1001/445	185/888/293	186/701/310	F
WATSON	20	1455/3587/2274	1426/4240/2255	1922/4843/3018	548/1480/864
ROSEX	8	36/446/90	26/421/62	38/362/93	63/764/100
	50	46/548/101	32/469/84	44/412/101	86/707/146
	100	45/459/99	23/445/57	46/414/102	71/856/112
SINGX	4	134/501/209	78/396/124	111/411/172	650/3254/1104
PEN1	2	5/18/12	12/182/34	5/18/12	5/18/12
PEN2	4	10/82/26	12/89/27	11/133/29	32/167/57
	50	131/764/254	405/1453/683	136/1056/256	121/724/242
VARDIM	2	3/9/7	3/9/7	3/9/7	3/9/7
	50	10/52/36	10/52/36	10/52/36	10/52/36
TRIG	3	13/129/27	11/82/25	15/225/27	162/974/267
	50	38/320/70	38/222/68	38/225/71	206/1662/290
	100	48/340/90	43/425/76	48/294/90	225/3077/286
BV	3	9/17/11	12/25/16	11/20/13	13/27/18
	10	64/171/97	50/148/81	64/172/99	59/163/93
IE	200	5/59/7	6/13/8	5/59/7	6/61/8
	500	5/11/7	6/13/8	6/13/8	6/13/8
TRID	3	14/33/18	10/26/17	14/33/18	15/84/21
	200	31/68/39	30/66/37	31/68/39	36/78/42
BAND	3	7/64/12	9/20/13	7/64/12	7/64/12
	50	19/670/26	15/278/23	19/670/26	F
	100	18/712/27	16/373/26	18/712/27	F
	500	18/677/26	16/339/27	18/677/26	F
LIN	1000	1/3/3	1/3/3	1/3/3	1/3/3
LIN1	10	1/3/3	1/3/3	1/3/3	1/3/3

The step length α_k in all methods is determined such that the strong Wolfe-Powell conditions (31) and (32) hold with $\delta = 0.01$ and $\sigma = 0.1$.

The test problems are drawn from [14]. The numerical results of our tests are reported in Table 1.

The column problem represents the problem name in [14], Dim represents the dimension of the problems. The numerical results are given in the form of $I/F/G$, where I, F, and G denote the numbers of iterations, function evaluations and gradient evaluations, respectively. The stopping condition is $\|g_k\| \leq 10^{-6}$. Since we want to compare the

performance of the different methods, in the numerical results, we omit the problems if all the four methods perform equally. The notation F means that, for this problem, the corresponding method fails.

5. Conclusions

In this paper, based on β_k^{DY} and β_k^{DL}, a new formula is proposed to compute the parameter β_k of the conjugate gradient methods. The main motivations are to improve both

the convergence properties and numerical behavior of the conjugate gradient method. For general conjugate gradient methods, in order to get the global convergence results, the methods are required to possess the following major properties:

(1) the generated directions d_k are descent directions;

(2) the parameters β_k are nonnegative.

In addition, to ensure that the methods have robust and efficient numerical behavior, the parameter β_k needs to approach zero, when the small step s_k occurs.

From the convergence analysis of this paper, we known that the directions d_k generated by MDL method are descent directions, which is not true for DY or DL methods, and the proposed MDL method is globally convergent for general functions. In the previous section, we compare the numerical performance of the MDL method with the DY, MHS, and DL methods. From the convergence analysis and numerical results, comparing with the DL, DY, and MHS method, we can have the following.

(a) MDL method versus DL method: from the computational point of view, for most of the test problems, MDL method performs quite similarly with DL method. There are 15 problems in which MDL method outperforms the DL method and 18 problems in which DL method outperforms the MDL method. But, from the convergent point of view, the MDL method outperforms the DL method.

(b) MDL method versus DY method: the convergence properties of MDL method are similar to DY method. By comparing the numerical results of MDL method with DY method, there are 27 test problems in which MDL method outperforms the DY method and only 4 test problems in which DY method outperforms the MDL method. Therefore, we could say that MDL method is much better than the DY method in numerical behavior.

(c) MDL method versus MHS method: they possess similar convergence properties; the numerical results show that MDL method performs little better than the MHS method.

Acknowledgments

References

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

[2] M. J. D. Powell, "Restart procedures for the conjugate gradient method," *Mathematical Programming*, vol. 12, no. 2, pp. 241–254, 1977.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] Z. Wei, S. Yao, and L. Liu, "The convergence properties of some new conjugate gradient methods," *Applied Mathematics and Computation*, vol. 183, no. 2, pp. 1341–1350, 2006.

[8] Y. Shengwei, Z. Wei, and H. Huang, "A note about WYL's conjugate gradient method and its applications," *Applied Mathematics and Computation*, vol. 191, no. 2, pp. 381–388, 2007.

[9] H. Huang, S. Yao, and H. Lin, "A new conjugate gradient method based on HS-DY methods," *Journal of Guangxi University of Technology*, no. 4, pp. 63–66, 2008.

[10] L. Zhang, "An improved Wei-Yao-Liu nonlinear conjugate gradient method for optimization computation," *Applied Mathematics and Computation*, vol. 215, no. 6, pp. 2269–2274, 2009.

[11] L. Zhang, "Further studies on the Wei-Yao-Liu nonlinear conjugate gradient method," *Applied Mathematics and Computation*, vol. 219, no. 14, pp. 7616–7621, 2013.

[12] Z. Dai and F. Wen, "Another improved Wei-Yao-Liu nonlinear conjugate gradient method with sufficient descent property," *Applied Mathematics and Computation*, vol. 218, no. 14, pp. 7421–7430, 2012.

[13] 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, 1999.

[14] J. J. Moré, B. S. Garbow, and K. E. Hillstrom, "Testing unconstrained optimization software," *ACM Transactions on Mathematical Software*, vol. 7, no. 1, pp. 17–41, 1981.