Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2009, Article ID 243290, 16 pages
doi:10.1155/2009/243290

Research Article

Nonlinear Conjugate Gradient Methods with Sufficient Descent Condition for Large-Scale Unconstrained Optimization

Jianguo Zhang,[1] Yunhai Xiao,[1,2] and Zengxin Wei[3]

[1] *Institute of Applied Mathematics, College of Mathematics and Information Science, Henan University, Kaifeng 475000, China*
[2] *Department of Mathematics, Nanjing University, Nanjing 210093, China*
[3] *College of Mathematics and Information Science, Guangxi University, Nanning 530004, China*

Correspondence should be addressed to Jianguo Zhang, zjg@henu.edu.cn

Received 3 January 2009; Revised 21 February 2009; Accepted 1 May 2009

Recommended by Joaquim J. Júdice

Two nonlinear conjugate gradient-type methods for solving unconstrained optimization problems are proposed. An attractive property of the methods, is that, without any line search, the generated directions always descend. Under some mild conditions, global convergence results for both methods are established. Preliminary numerical results show that these proposed methods are promising, and competitive with the well-known PRP method.

1. Introduction

In this paper, we consider the unconstrained optimization problem

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

where $f : \mathbf{R}^n \to \mathbf{R}$ is a continuously differentiable function, and its gradient at point x_k is denoted by $g(x_k)$, or g_k for the sake of simplicity. n is the number of variables, which is automatically assumed to be large. The iterative formula of nonlinear conjugate gradient method is given by

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

where α_k is a steplength, and d_k is a search direction which is determined by

$$d_k = \begin{cases} -g_0, & \text{if } k = 0, \\ -g_k + \beta_k d_{k-1}, & \text{if } k \geq 1, \end{cases} \quad (1.3)$$

where β_k is a scalar. Since 1952, there have been many well-known formulas for the scalar β_k, for example, Fletcher-Reeves (FR), Ploak-Ribiére-Polyak (PRP), Hestenes-Stiefel (HS), and Dai-Yuan (DY) (see [1–4]),

$$\beta_k^{\text{FR}} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \qquad \beta_k^{\text{PRP}} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}, \qquad \beta_k^{\text{HS}} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, \qquad \beta_k^{\text{DY}} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}}, \quad (1.4)$$

where $y_{k-1} = g_k - g_{k-1}$, symbol $\|\cdot\|$ denotes the Euclidean norm of vectors. Their corresponding methods generally specified as FR, PRP, HS, and DY conjugate gradient methods. If f is a strictly convex quadratic function, all these methods are equivalent in the case that an exact line search is used. If the objective function is nonconvex, their behaviors may be distinctly different. In the past two decades, the convergence properties of FR, PRP, HS, and DY methods have been intensively studied by many researchers (e.g., [5–13]).

In practical computation, the HS and PRP methods, which share the common numerator $g_k^T y_{k-1}$, are generally believed to be the most efficient conjugate gradient methods, and have got meticulous in recent years. One remarkable property of both methods is that they essentially perform a restart if a bad direction occurs (see [7]). However, Powell [14] constructed an example showed that both methods can cycle infinitely without approaching any stationary point even if an exact line search is used. This counter-example also indicates that both methods have a drawback that they may not globally be convergent when the objective function is non-convex. Therefore, during the past few years, much effort has been investigated to create new formulae for β_k, which not only possess global convergence for general functions but are also superior to original method from the computation point of view (see [15–22]). An excellent survey of nonlinear conjugate gradient methods with special attention to global convergence properties is made by Hager and Zhang [7].

Recently, Dai and Liao [18] proposed two new formulae (called DL and DL+) for β_k based on the secant condition from quasi-Newton method. More lately, Li, Tang, and Wei (see [20]) also presented another two formulae (called LTW and LTW+) based on a modified secant condition in [21]. In addition, the corresponding conjugate gradient method for β_k with DL+ (or LTW+) converges globally for non-convex minimization problems, the reported numerical results showed that it excels the standard PRP method. However, the convergence result of the method for β_k with formula DL (or LTW) has not been totally explored yet. In this paper, we further study conjugate gradient method for the solution of unconstrained optimization problems. Meanwhile, we focus our attention on the scalar for β_k with DL (or LTW). Our motivation mainly comes from the recent work of Zhang et al. [22]. We introduce two versions of modified DL and LTW conjugate gradient-type methods. An attractive property of both proposed methods are that the generated directions are always descending. Besides, this property is independent of line search used and the convexity of objective function. Under some favorable conditions, we establish the global convergence of the proposed methods. We also do some numerical experiments by using a large set of unconstrained optimization problems, which indicate the proposed methods possess better performances when compared with the classic PRP method.

We organize this paper as follows. In the next Section, we briefly review the conjugate gradient methods are proposed in [18, 20]. We present two conjugate gradient methods in Sections 3 and 4, respectively. Global convergence properties are also discussed simultaneously. In the last Section we perform the numerical experiments by using a set of large problems, and do some numerical comparisons with PRP method.

2. Conjugate Gradient Methods with Secant Condition

In this section we give a short description of the new conjugate gradient method of Dai and Liao in [18]. In the following, we also briefly review another effective conjugate gradient method of Li, Tang, and Wei in [20]. Motivated by the these methods, we introduce our new versions of conjugate gradient-type methods in the following sections.

The following two assumptions are often utilized in convergence analysis for conjugate gradient algorithms.

Assumption 2.1. The objective function f is bounded below, and the level set $\mathcal{F} = \{x \in \mathbf{R}^n \mid f(x) \leq f(x_0)\}$ is bounded.

Assumption 2.2. In some neighborhood $\mathcal{N}$ of $\mathcal{F}$, f is differentiable and its gradient is Lipschitz continuous, namely, there exists a positive constant L such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \quad \forall x, y \in \mathcal{N}. \tag{2.1}$$

The above assumption implies that there exists a positive constant $\overline{\gamma}$ such that

$$\|g(x)\| \leq \overline{\gamma}, \quad \forall x \in \mathcal{F}. \tag{2.2}$$

2.1. Dai-Liao Method

Note that, in quasi-Newton method, standard BFGS method, and limited memory BFGS method, the serch direction d_k always have the common form

$$d_k = -B_k^{-1} g_k, \tag{2.3}$$

where B_k is some $n \times n$ symmetric and positive definite matrix satisfing the secant condition (or quasi-Newton equation)

$$B_k s_{k-1} = y_{k-1}. \tag{2.4}$$

Combing the above two equations, we obtain

$$d_k^T y_{k-1} = d_k^T (B_k s_{k-1}) = -g_k^T s_{k-1}. \tag{2.5}$$

Keeping these relations in mind, Dai and Liao introduced the following conjugacy condition:

$$d_k^T y_{k-1} = -t g_k^T s_{k-1}, \tag{2.6}$$

where $t \geq 0$ is a parameter. Multiplying (1.3) with y_{k-1} and making use of the new conjugacy condition (2.6), Dai and Liao obtained the following new formula for computing β_k:

$$\beta_k^{\mathrm{DL}} = \frac{g_k^T (y_{k-1} - t s_{k-1})}{d_{k-1}^T y_{k-1}}, \quad t \geq 0. \tag{2.7}$$

In order to ensure the global convergence for general functions, Dai and Liao restrict β_k to be positive, that is,

$$\beta_k^{\mathrm{DL+}} = \max\left\{ \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, 0 \right\} - t \frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}}, \quad t \geq 0. \tag{2.8}$$

The reported numerical experiments showed that the corresponding conjugate gradient method is efficient.

2.2. Li-Tang-Wei Method

Recently, Wei et al. [21] proposed a modified secant condition

$$B_k s_{k-1} = y_{k-1}^* = y_{k-1} + \lambda_{k-1} s_{k-1}, \tag{2.9}$$

where

$$\lambda_{k-1} = \frac{2\left[f(x_{k-1}) - f(x_k)\right] + (g_k + g_{k-1})^T s_{k-1}}{\|s_{k-1}\|^2}. \tag{2.10}$$

Notice this new secant condition contains not only gradient value information, but also function value information at the present and the previous step. Additionally, this modified secant condition has inspired many further studies on optimization problems (e.g., [23–25]).

Based on the modified secant condition (2.9), Li, Tang and Wei (see [20]) presented the new conjugacy condition:

$$d_k^T y_{k-1}^* = -t g_k^T s_{k-1}, \quad t \geq 0. \tag{2.11}$$

Similar to the Dai-Liao formulas in (2.7) and (2.8), Li, Tang, and Wei also constructed the following two conjugate gradient formulas for β_k:

$$\beta_k^{\mathrm{LTW}} = \frac{g_k^T\left(\tilde{y}_{k-1}^* - ts_{k-1}\right)}{d_{k-1}^T\tilde{y}_{k-1}^*}, \quad t \geq 0,$$
$$\beta_k^{\mathrm{LTW+}} = \max\left\{\frac{g_k^T\tilde{y}_{k-1}^*}{d_{k-1}^T\tilde{y}_{k-1}^*}, 0\right\} - t\frac{g_k^T s_{k-1}}{d_{k-1}^T\tilde{y}_{k-1}^*}, \quad t \geq 0, \tag{2.12}$$

where $\tilde{y}_{k-1}^* = y_{k-1} + \max\{\lambda_{k-1}, 0\}s_{k-1}$.

Obviously, the new secant condition (2.11) gives a more accurate approximation for the Hessian of the objective function (see [21]). Hence the formulas (2.12) should outperform the Dai-Liao's methods from theoretically, and the numerical results in [20] confirmed this claim. In addition, based on another modified quasi-Newton equation of Zhang et al. [26], Yabe and Takano [27] also proposed some similar conjugate gradient methods for unconstrained optimization.

Combing with strong Wolfe-Powell line search, the conjugate gradient methods with β_k from DL+ or LTW+ were proved convergent globally for non-convex minimization problems. But for β_k from DL or LTW, there are no similar results. The major contribution of our following work is to circumvent this difficulty. However, our attention does not focus on the general iterative style (1.3), our idea mainly originate from the very recently three-term conjugate gradient method of Zhang et al. [22].

3. Modified Dai-Liao Method

As we have stated in the previous section, the standard conjugate gradient method with (1.2)-(1.3) and (2.7) cannot guarantee the sequence $\{x_k\}$ approaches to any stationary point of the problem. In this section, we will appeal to a three-term form to take the place of (1.3).

The first three-term nonlinear conjugate gradient algorithm was presented by Nazareth [28], in which the search direction is determined by

$$d_{k+1} = -y_k + \frac{y_k^T y_k}{y_k^T d_k}d_k + \frac{y_{k-1}^T y_k}{y_{k-1}^T d_{k-1}}d_{k-1} \tag{3.1}$$

with $d_{-1} = 0$, $d_0 = -g_0$. The main property of d_k is that, for a quadratic function, it remains conjugate even without exact line searches. Recently, Zhang et al. [22] proposed a descent modified PRP conjugate gradient method with three terms as follows:

$$d_k = \begin{cases} -g_0, & \text{if } k = 0, \\ -g_k + \beta_k^{\mathrm{PRP}}d_{k-1} - \theta_k y_{k-1}, & \text{if } k \geq 1, \end{cases} \tag{3.2}$$

where $\theta_k = g_k^T d_{k-1}/\|g_{k-1}\|^2$. A remarkable property of the method is that it produces a descent direction at each iteration. Motivated by the nice descent property, we also give a three-term conjugate gradient method based on the DL formula for β_k in (2.7), that is,

$$d_k = \begin{cases} -g_0, & \text{if } k = 0, \\ -g_k + \beta_k^{\mathrm{DL}} d_{k-1} - \xi_k (y_{k-1} - t s_{k-1}), & \text{if } k \geq 1, \end{cases} \tag{3.3}$$

where $t \geq 0$ and $\xi_k = g_k^T d_{k-1}/d_{k-1}^T y_{k-1}$. It is easy to see that the sufficient descent condition also holds true if no line search is used, that is,

$$g_k^T d_k = -\|g_k\|^2. \tag{3.4}$$

In order to achieve the global convergence result of the PRP method, Grippo and Lucidi [9] proposed a new line search below. For given constants $\tau > 0$, $\delta > 0$, and $\lambda \in (0,1)$, let

$$\alpha_k = \max\left\{ \lambda^j \frac{\tau \left| g_k^T d_k \right|}{\|d_k\|^2};\ j = 0, 1, \ldots \right\} \tag{3.5}$$

satisfy

$$\begin{gathered} f(x_{k+1}) \leq f(x_k) - \delta \alpha_k^2 \|d_k\|^2, \\ -c_1 \|g(x_{k+1})\|^2 \leq g(x_{k+1})^T d_{k+1} \leq -c_2 \|g(x_{k+1})\|^2, \end{gathered} \tag{3.6}$$

where $0 < c_2 < 1 < c_1$ are constants. Here we prefer this new line search to the classical Armijo one for the sake of a greater reduction of objective function and wider tolerance of $\alpha\|d_k\|$ (see [9]).

Introducing the line search rule, we are now ready to state the steps of the modified Dai-Liao (MDL) conjugate gradient-type algorithm as follows.

Algorithm 3.1 (MDL).

Step 1. Given $x_0 \in \mathbf{R}^n$. Let $0 < \delta < \sigma < 1$, $t \geq 0$ and $d_0 = -g_0$. Set $k := 0$.

Step 2. If $\|g_k\| = 0$ then stop.

Step 3. Compute d_k by (3.3).

Step 4. Find the steplength α_k satisfying

$$f(x_k + \alpha_k d_k) - f(x_k) \leq -\delta \alpha_k^2 \|d_k\|^2, \tag{3.7}$$

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k. \tag{3.8}$$

Let $x_{k+1} = x_k + \alpha_k d_k$.

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

From now on, we use f_k to denote $f(x_k)$. For MDL algorithm, we have the following two important results. The proof of the following first lemma was established by Zoutendijk [29], where it is stated for slightly different circumstances. For convenience, we give the detailed proof here.

Lemma 3.2. *Consider the conjugate gradient-type method in the form* (1.2) *and* (3.3)*, and let the steplength* α_k *be obtained by the line search* (3.7)-(3.8)*. Suppose that Assumptions 2.1-2.2 hold. Then one has*

$$\sum_{k=0}^{\infty} \alpha_k^2 \|d_k\|^2 < \infty. \tag{3.9}$$

Proof. Since α_k is obtained by the line search (3.7)-(3.8). Then, by (3.4) and (3.7) we have

$$f_{k+1} - f_k \le -\delta \alpha_k^2 \|d_k\|^2 \le 0. \tag{3.10}$$

Hence, $\{f_k\}$ is a decreasing sequence and the sequence $\{x_k\}$ is contained in $\mathcal{F}$. Hence, Assumptions 2.1-2.2 imply that there exists a constant f^* such that

$$\lim_{k \to \infty} f_k = f^*. \tag{3.11}$$

From (3.11), we have

$$\sum_{k=0}^{\infty} (f_k - f_{k+1}) < +\infty. \tag{3.12}$$

This together with (3.10) implies that (3.9) holds. □

Lemma 3.3. *If there exists a constant* $\epsilon > 0$ *such that*

$$\|g_k\| \ge \epsilon, \quad \forall k \ge 0, \tag{3.13}$$

then there exists a constant $M > 0$ *such that*

$$\|d_k\| \le M, \quad \forall k \ge 0. \tag{3.14}$$

Proof. From the line search (3.7)-(3.8) and (3.4), we have

$$y_{k-1}^T d_{k-1} = g_k^T d_{k-1} - g_{k-1}^T d_{k-1} \ge -(1-\sigma) g_{k-1}^T d_{k-1} = (1-\sigma)\|g_{k-1}\|^2. \tag{3.15}$$

By the definition of d_k in (3.3) , we get from (2.1), (2.2), (3.4), and (3.13) that

$$\|d_k\| \leq \|g_k\| + 2\frac{\|g_k\| \ \ \|y_{k-1} - ts_{k-1}\|\|d_{k-1}\|}{(1-\sigma)\|g_{k-1}\|^2} \leq \overline{\gamma} + \frac{2(L+t)\overline{\gamma}\alpha_{k-1}\|d_{k-1}\|}{(1-\sigma)\epsilon^2}\|d_{k-1}\|. \tag{3.16}$$

Lemma 3.2 indicates that $\alpha_k d_k \to 0$ as $k \to \infty$, then there exists a constant $\gamma \in (0,1)$ and an integer k_0, such that the following inequality holds for all $k \geq k_0$:

$$\frac{2(L+t)\overline{\gamma}}{(1-\sigma)\epsilon^2}\alpha_{k-1}\|d_{k-1}\| \leq \gamma. \tag{3.17}$$

Hence, we have for any $k > k_0$

$$\|d_k\| \leq \overline{\gamma} + \gamma\|d_{k-1}\| \leq \overline{\gamma}\left(1 + \gamma + \gamma^2 + \cdots + \gamma^{k-k_0-1}\right) + \gamma^{k-k_0}\|d_{k_0}\| \leq \frac{\overline{\gamma}}{1-\gamma} + \|d_{k_0}\|. \tag{3.18}$$

Setting $M = \max\{\|d_1\|, \|d_2\|, \ldots, \|d_{k_0}\|, \overline{\gamma}/(1-\gamma) + \|d_{k_0}\|\}$, we deduce that $\|d_k\| \leq M$ for all k. □

Using the preceding lemmas, we are now ready to give the promised convergence results.

Theorem 3.4. *Suppose that Assumptions 2.1-2.2 hold. Let $\{x_k\}$ be generated by Algorithm MDL. Then one has*

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{3.19}$$

Proof. We proceed by contradiction. Assume that the conclusion is not true. Then there exists a positive constant ϵ such that

$$\|g_k\| \geq \epsilon, \quad \forall k \geq 0. \tag{3.20}$$

If $\liminf_{k\to\infty}\alpha_k > 0$, we have from (3.9) that $\liminf_{k\to\infty}\|g_k\| = 0$. This contradicts assumption (3.20).

Suppose that $\liminf_{k\to\infty}\alpha_k = 0$. Using Assumptions 2.1-2.2 and (3.8), we obtain

$$-(1-\sigma)g_k^T d_k \leq \left(g_{k+1} - g_k\right)^T d_k \leq L\alpha_k\|d_k\|^2. \tag{3.21}$$

Combining with (3.4) yields

$$(1-\sigma)\|g_k\|^2 \leq L\alpha_k\|d_k\|^2. \tag{3.22}$$

The above inequality and Lemma 3.3 imply $\liminf_{k\to\infty}\|g_k\| = 0$, which contradicts (3.20). This completes the proof. □

4. Modified Li-Tang-Wei Method

In a similar manner, we provide a modified Li-Tang-Wei method with three terms in the form:

$$d_k = \begin{cases} -g_0, & \text{if } k = 0, \\ -g_k + \beta_k^{\text{LTW}} d_{k-1} - \zeta_k \left(\tilde{y}_{k-1}^* - t s_{k-1}\right), & \text{if } k \geq 1, \end{cases} \tag{4.1}$$

where $\zeta_k = g_k^T d_{k-1} / d_{k-1}^T \tilde{y}_{k-1}^*$. It is not difficult to see that the sufficient descent property (3.4) also holds.

Combing with the line search (3.7)-(3.8), we state the steps of the modified Li-Tang-Wei (MLTW) conjugate gradient-type algorithm as follows.

Algorithm 4.1 (MLTW).

Step 1. Given $x_0 \in \mathbf{R}^n$. Let $0 < \delta < \sigma < 1$, $t \geq 0$ and $d_0 = -g_0$. Set $k := 0$.

Step 2. If $\|g_k\| = 0$ then stop.

Step 3. Compute d_k by (4.1).

Step 4. Find the steplength α_k satisfying (3.7)-(3.8). Let $x_{k+1} = x_k + \alpha_k d_k$.

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

Lemma 4.2. *Consider the conjugate gradient-type method in the forms* (1.2) *and* (4.1)*, let the steplength α_k be obtained by the line search* (3.7)-(3.8)*. Suppose that Assumptions 2.1-2.2 hold. Then one has*

$$\sum_{k=0}^{\infty} \alpha_k^2 \|d_k\|^2 < \infty. \tag{4.2}$$

Proof. See Lemma 3.2. □

Lemma 4.3. *Consider the conjugate gradient-type method in the form* (1.2) *and* (4.1)*, let the steplength α_k be obtained by line search* (3.7)-(3.8)*. Suppose that Assumptions 2.1-2.2 hold. Then one has*

$$\|\tilde{y}_k^*\| \leq 2L \|s_k\|, \tag{4.3}$$

where L was defined as in Assumption 2.2.

Proof. Since α_k is obtained by (3.7)-(3.8), from (3.10) we know that

$$x_k \in \mathcal{F}, \quad \forall k \geq 1. \tag{4.4}$$

By mean value theorem, we know that there exists $\eta_k \in [0,1]$ such that

$$f_{k+1} - f_k = g\left(x_k + \eta_k(x_{k+1} - x_k)\right)^T (x_{k+1} - x_k) = g\left(x_k + \eta_k s_k\right)^T s_k. \tag{4.5}$$

Using (4.4) we get

$$x_k + \eta_k s_k = x_k + \eta_k(x_{k+1} - x_k) \in \overline{\text{co}}\mathcal{F}, \tag{4.6}$$

where $\overline{\text{co}}\mathcal{F}$ denotes the closed convex hull of $\mathcal{F}$. It follows from the definition of λ_k and (4.5) that

$$\begin{aligned}
\lambda_k &= \frac{2\left[f(x_k) - f(x_{k+1})\right] + (g_{k+1} + g_k)^T s_k}{\|s_k\|^2} \\
&= \frac{\left[g_k + g_{k+1} - 2g(x_k + \eta_k s_k)\right]^T s_k}{\|s_k\|^2} \\
&\leq \frac{\left(\|g_k - g(x_k + \eta_k s_k)\| + \|g_{k+1} - g(x_k + \eta_k s_k)\|\right)}{\|s_k\|^2}\|s_k\| \\
&\leq \frac{\left(L\eta_k\|s_k\| + L(1 - \eta_k)\|s_k\|\right)}{\|s_k\|^2}\|s_k\| \\
&= L.
\end{aligned} \tag{4.7}$$

From the definition of $\tilde{y}_k^*$ and Assumption 2.2, we know that

$$\|\tilde{y}_k^*\| = \|y_k + \max\{\lambda_k, 0\}s_k\| \leq \|y_k + Ls_k\| \leq \|y_k\| + L\|s_k\| \leq 2L\|s_k\|. \tag{4.8}$$

This verifies our claims. □

Lemma 4.4. *If there exists a constant $\epsilon > 0$ such that*

$$\|g_k\| \geq \epsilon, \quad \forall k \geq 0, \tag{4.9}$$

then there exists a constant $M > 0$ such that

$$\|d_k\| \leq M, \quad \forall k \geq 0. \tag{4.10}$$

Proof. From the line search (3.7)-(3.8) and (3.4), we have

$$\begin{aligned} d_{k-1}^T \widetilde{y}_{k-1}^* &= d_{k-1}^T \left(y_{k-1} + \max\{\lambda_{k-1}, 0\} s_{k-1} \right) \\ &\geq d_{k-1}^T y_{k-1} \\ &\geq -(1-\sigma) g_{k-1}^T d_{k-1} \\ &= (1-\sigma) \|g_{k-1}\|^2. \end{aligned} \tag{4.11}$$

According to the definition of d_k in (4.1), we get from (2.1), (2.2), (4.9), and (4.11) that

$$\begin{aligned} \|d_k\| &\leq \|g_k\| + 2 \frac{\|g_k\| \|\widetilde{y}_{k-1}^* - t s_{k-1}\| \|d_{k-1}\|}{(1-\sigma)\|g_{k-1}\|^2} \\ &\leq \overline{\gamma} + \frac{2\overline{\gamma}(2L+t)\alpha_{k-1}\|d_{k-1}\|}{(1-\sigma)\epsilon^2} \|d_{k-1}\|. \end{aligned} \tag{4.12}$$

Lemma 4.3 shows that $\alpha_k d_k \to 0$ as $k \to \infty$. Hence there exists a constant $\gamma \in (0,1)$ and an integer k_0, such that the following inequality holds for all $k \geq k_0$:

$$\frac{2(2L+t)\overline{\gamma}}{(1-\sigma)\epsilon^2} \alpha_{k-1} \|d_{k-1}\| \leq \gamma. \tag{4.13}$$

Hence, we have for any $k > k_0$

$$\|d_k\| \leq \overline{\gamma} + \gamma \|d_{k-1}\| \leq \gamma_1 \left(1 + \gamma + \gamma^2 + \cdots + \gamma^{k-k_0-1} \right) + \gamma^{k-k_0} \|d_{k_0}\| \leq \frac{\overline{\gamma}}{1-\gamma} + \|d_{k_0}\|. \tag{4.14}$$

Let $M = \max\{\|d_1\|, \|d_2\|, \ldots, \|d_{k_0}\|, \overline{\gamma}/(1-\gamma) + \|d_{k_0}\|\}$, we get (4.10). □

Now we can establish the following global convergence theorem for MLTW method. Since its proof is essentially similar to Theorem 3.4, we omit it.

Theorem 4.5. *Suppose that Assumptions 2.1-2.2 hold. Let $\{x_k\}$ be generated by Algorithm MLTW. Then one has*

$$\liminf_{k \to \infty} \|g_k\| = 0. \tag{4.15}$$

Proof. Omitted. □

To end of this section, we show that MLWT method is equivalent to all the general method (1.4) if an exact line search is used. In deriving this equivalence, we work with an exact line search rule, that is, we compute α_k such that

$$f(x_k + \alpha_k d_k) = \min_{\alpha \geq 0} f(x_k + \alpha d_k) \tag{4.16}$$

is satisfied. Hence,

$$\nabla f(x_k + \alpha_k d_k)^T d_k = 0. \tag{4.17}$$

Subsequently,

$$\zeta_k = \frac{g_k^T d_{k-1}}{d_{k-1}^T \tilde{y}_{k-1}^*} = 0. \tag{4.18}$$

Moreover, let

$$f(x) = \frac{1}{2} x^T G x + b^T x + c, \tag{4.19}$$

where G is an $n \times n$ symmetric positive definite matrix, $b \in R^n$, and c is a real number. In this case, it is not difficult to see that $\lambda_{k-1} = 0$. Note that by the definition of β_k^{LWT} in (2.12), we have

$$\begin{aligned} \beta_k^{\mathrm{MLWT}} &= \frac{g_k^T (\tilde{y}_{k-1}^* - t s_{k-1})}{d_{k-1}^T \tilde{y}_{k-1}^*} \\ &= \frac{g_k^T \tilde{y}_{k-1}^*}{d_{k-1}^T \tilde{y}_{k-1}^*} \\ &= \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} \\ &= \beta_k^{\mathrm{HS}}. \end{aligned} \tag{4.20}$$

Then we have the main properties of a conjugate gradient method. The following theorem shows that MLWT method have quadratic termination property, which means that the method terminates at most n steps when it is applied to a positive definite quadratic. The proof can be found in Theorem 4.2.1 in [30] and is omitted.

Theorem 4.6. *For a positive definite quadratic function* (4.19)*, the conjugate gradient method* (1.2)–(4.1) *with exact line search terminates after* $m \le n$ *steps, and the following properties hold for all* i, *(*$0 \le i \le m$*),*

$$\begin{gathered} d_i^T G d_j = 0, \quad j = 0, 1, \ldots, i-1, \\ g_i^T g_j = 0, \quad j = 0, 1, \ldots, i-1, \\ d_i^T g_i = -g_i^T g_i, \end{gathered} \tag{4.21}$$

where m *is the number of distinct eigenvalues of* G.

The theorem also shows that conjugate gradients (1.2)–(4.1) represent conjugacy of directions, orthogonality of gradients, and descent condition. This also indicates that methods (1.2)–(4.1) preserve the property of being equivalent to the general conjugate gradient method (1.4) for strict convex quadratics with exact line search. The cases of DL, LWT, and MDL can be proved in a similar way.

5. Numerical Experiments

The main work of this section is to report the performance of the algorithms MDL and MLTW on a set of test problems. The codes were written in Fortran77 and in double precision arithmetic. All the tests were performed on a PC (`Intel Pentium Dual E2140 1.6 GHz`, `256 MB SDRAM`). Our experiments were performed on a set of 73 nonlinear unconstrained problems that have second derivatives available. These test problems are contributed by N. Andrei, and the Fortran expression of their functions and gradients are available at http://www.ici.ro/camo/neculai/SCALCG/evalfg.for. 26 out of these problems are from CUTE [31] library. For each test function we have considered 10 numerical experiments with number of variables $n = 1000, 2000, \ldots, 10000$.

In order to assess the reliability of our algorithms, we also tested these methods against the well-known routine PRP using the same problems. The PRP code is coauthored by Liu, Nocedal, and Waltz, it can be obtained from Nocedal's web page at http://www.ece.northwestern.edu/~nocedal/software.html/. While running of the PRP code, default values were used for all parameters. All these algorithms terminate when the following stopping criterion is met:

$$\|g_k\| \le 10^{-5}. \tag{5.1}$$

We also force these routines stopped if the iterations exceed 1000 or the number of function evaluations reach 2000 without achieving convergence. In MDL and MLTW, we use $\delta = 10^{-4}$, $\sigma = 0.1$. Moreover, we also test our proposed methods MDL and MLTW with different parameters t to see that $t = 1.0$ is the best choice. Since a large set of problems is used, we describe the results fully on the first author's web page at the web site: http://maths.henu.edu.cn/szdw/teachers/xyh.htm. The tables contain the number of the problem (Problem), the dimension of the problem (Dim), the Number of iterations (Iter), the number of function and gradient evaluations (Nfcnt), the CPU time required in seconds (Time), the final function value (Fv), and norm of the final gradient (Norm).

There are 30 problems that were excluded from the first two tables because they lead an "overflow error" when evaluated at some point by MDL and MLWT methods. However, the same error was occurred on 43 problems when evaluated by PRP method. From these tables, we also see that MDL and MLWT failed to satisfy the termination condition (5.1) on other 66 and 81 problems, respectively. But PRP method cannot achieve convergence on 89 problems. So only 634 problems remain where at least one method runs successfully. Now, we change our attention to consider the function values of the remaining problems founded by all three methods. We note that, on 592 problems, the differences of these functional values obtained by each method is less than the pretty small tolerance 10^{-7}. Therefore, it is reasonable to think that all the three methods obtained the same optimal solution on these problems.

To approximatively assess the performance of MDL, MLWT, and PRP methods on the remaining 592 problems, we use the profile of Dolan and Moré [32] as an evaluated tool. That

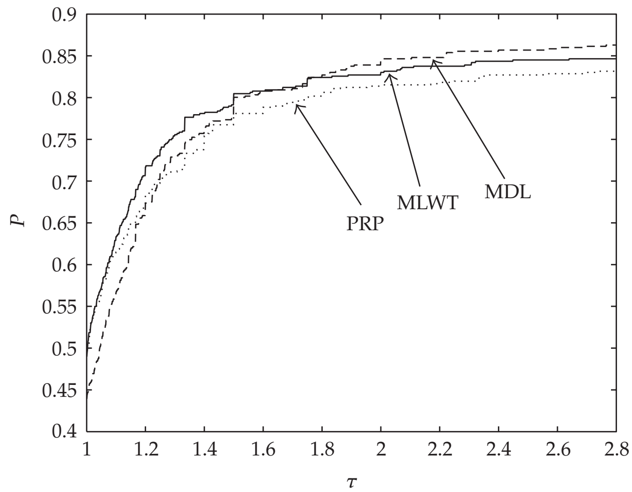

Figure 1: Performance profiles based on iterations.

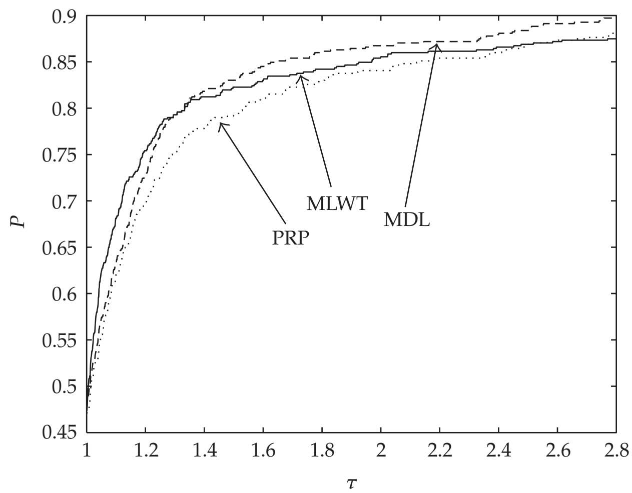

Figure 2: Performance profiles based on function and gradient evaluations.

is, for subset of the methods being analyzed, we plot the fraction P of problems for which any given method is within a factor τ of the best. Meanwhile, we use the iterations, function and gradient evaluations, and CPU time consuming as performance measure, since they reflect the main computational cost and the efficiency for each method. The performance profiles of all three methods are plotted in Figures 1, 2, and 3.

Observing Figures 1 and 2, respectively, it concludes that MDL and MLWT are always the top performer for almost all values of τ, which shows that they perform better than PRP method regarding iterations, function, and gradient evaluations. Figure 3 shows the implementation of the these methods using the total CPU time as a measure. This figure shows that PRP method is faster than the others. Why do our methods need more computing time though requiring less iterations? We think that it is highly possible that our new version of formula is a somewhat more complicated than the standard PRP method.

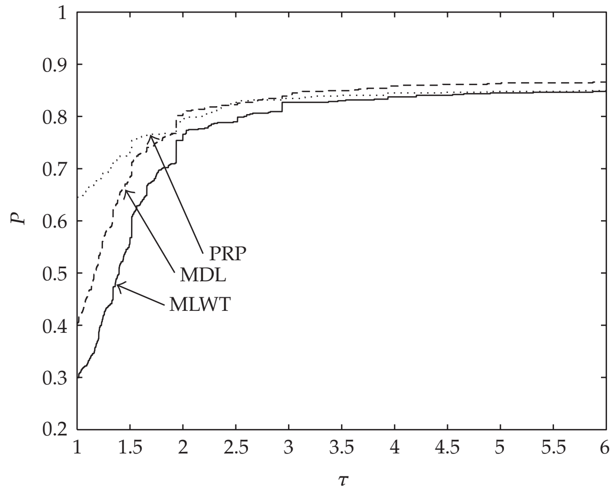

Figure 3: Performance profiles based on function and gradient evaluations.

Taking everything into consideration and albeit both proposed conjugate gradient methods did not obtain significant development as we have expected, we think that, for some specific problems, the enhancement of the proposed methods are still noticeable. Hence, we believe that each one of the new algorithm is a valid approach for the problems and has its own potential.

Acknowledgments

The authors are very grateful to two anonymous referees for their useful suggestions and comments on the previous version of this paper. This work was supported by Chinese NSF Grant 10761001, and Henan University Science Foundation Grant 07YBZR002.

References

[1] Y. H. Dai and Y. Yuan, "A nonlinear conjugate gradient method with a strong global convergence property," *SIAM Journal on Optimization*, vol. 10, no. 1, pp. 177–182, 2000.

[2] R. Fletcher and C. M. Reeves, "Function minimization by conjugate gradients," *The Computer Journal*, vol. 7, pp. 149–154, 1964.

[3] M. R. Hestenes and E. Stiefel, "Methods of conjugate gradient for solving linear systems," *Journal of Research of the National Bureau of Standards*, vol. 49, pp. 409–436, 1952.

[4] E. Polak and G. Ribière, "Note sur la convergence de directions conjugées," *Revue Francaise d'Informatique et de Recherche Operationnelle*, vol. 16, pp. 35–43, 1969.

[5] 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.

[6] 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.

[7] W. W. Hager and H. Zhang, "A survey of nonlinear conjugate gradient methods," *Pacific Journal of Optimization*, vol. 2, pp. 35–58, 2006.

[8] 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.

[9] L. Grippo and S. Lucidi, "A globally convergent version of the Polak-Ribière conjugate gradient method," *Mathematical Programming*, vol. 78, no. 3, pp. 375–391, 1997.

[10] L. Grippo and S. Lucidi, "Convergence conditions, line search algorithms and trust region implementations for the Polak-Ribière conjugate gradient method," *Optimization Methods & Software*, vol. 20, no. 1, pp. 71–98, 2005.
[11] J. Sun and J. Zhang, "Global convergence of conjugate gradient methods without line search," *Annals of Operations Research*, vol. 103, pp. 161–173, 2001.
[12] Z. Wei, G. Y. Li, and L. 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.
[13] 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.
[14] M. J. D. Powell, "Nonconvex minimization calculations and the conjugate gradient method," in *Numerical Analysis (Dundee, 1983)*, vol. 1066 of *Lecture Notes in Mathematics*, pp. 122–141, Springer, Berlin, Germany, 1984.
[15] N. Andrei, "Scaled conjugate gradient algorithms for unconstrained optimization," *Computational Optimization and Applications*, vol. 38, no. 3, pp. 401–416, 2007.
[16] N. Andrei, "Another nonlinear conjugate gradient algorithm for unconstrained optimization," *Optimization Methods & Software*, vol. 24, no. 1, pp. 89–104, 2009.
[17] E. G. Birgin and J. M. Martínez, "A spectral conjugate gradient method for unconstrained optimization," *Applied Mathematics and Optimization*, vol. 43, no. 2, pp. 117–128, 2001.
[18] 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.
[19] W. W. Hager and H. Zhang, "A new conjugate gradient method with guaranteed descent and an efficient line search," *SIAM Journal on Optimization*, vol. 16, no. 1, pp. 170–192, 2005.
[20] 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.
[21] Z. Wei, G. Li, and L. Qi, "New quasi-Newton methods for unconstrained optimization problems," *Applied Mathematics and Computation*, vol. 175, no. 2, pp. 1156–1188, 2006.
[22] L. Zhang, W. Zhou, and D.-H. Li, "A descent modified Polak-Ribière-Polyak conjugate gradient method and its global convergence," *IMA Journal of Numerical Analysis*, vol. 26, no. 4, pp. 629–640, 2006.
[23] L. Liu, S. Yao, and Z. Wei, "The global and superlinear convergence of a new nonmonotone MBFGS algorithm on convex objective functions," *Journal of Computational and Applied Mathematics*, vol. 220, no. 1-2, pp. 422–438, 2008.
[24] Z. Wei, G. Yu, G. Yuan, and Z. Lian, "The superlinear convergence of a modified BFGS-type method for unconstrained optimization," *Computational Optimization and Applications*, vol. 29, no. 3, pp. 315–332, 2004.
[25] Y. Xiao, Z. Wei, and Z. Wang, "A limited memory BFGS-type method for large-scale unconstrained optimization," *Computers & Mathematics with Applications*, vol. 56, no. 4, pp. 1001–1009, 2008.
[26] J. 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.
[27] 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.
[28] L. Nazareth, "A conjugate direction algorithm without line searches," *Journal of Optimization Theory and Applications*, vol. 23, no. 3, pp. 373–387, 1977.
[29] G. Zoutendijk, "Nonlinear programming, computational methods," in *Integer and Nonlinear Programming*, J. Jabadie, Ed., pp. 37–86, North-Holland, Amsterdam, The Netherlands, 1970.
[30] W. Sun and Y. Yuan, *Optimization Theory and Methods: Nonlinear Programming*, vol. 1 of *Springer Optimization and Its Applications*, Springer, New York, NY, USA, 2006.
[31] 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.
[32] E. D. Dolan and J. J. Moré, "Benchmarking optimization software with performance profiles," *Mathematical Programming*, vol. 91, no. 2, pp. 201–213, 2002.