Filomat 38:12 (2024), 4101–4115
https://doi.org/10.2298/FIL2412101L

Published by Faculty of Sciences and Mathematics,
University of Niš, Serbia
Available at: http://www.pmf.ni.ac.rs/filomat

A sufficient descent three-term conjugate gradient method and its global convergence

Mina Lotfi[a], S. Mohammad Hosseini[a]

[a] *Department of Applied Mathematics, Faculty of Mathematical Sciences, Tarbiat Modares University*

Abstract. In this paper, we presented a new three-term conjugate gradient method based on combining the conjugate gradient method proposed by Cheng et al [15] with the idea of the modified FR method [22]. In our method, search direction satisfies the sufficient descent condition independent of the line search. Under some standard assumptions, we establish the global convergence property and the r-linear convergence rate of the proposed method. Numerical results on the standard test problems, in some well-known library, illustrate computational efficiency of the new method.

1. Introduction

We consider the following unconstrained optimization problem:

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

where $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is continuously differentiable and its gradient is denoted by $g(x) = \nabla f(x)$. Conjugate gradient (CG) methods are efficient iterative methods, with some important properties such as low memory requirement and strong global convergence, making them useful tools in solving large-scale unconstrained optimization problems. A CG method generates a sequence of points $x_k \in \mathbb{R}^n$, obtained by

$$x_{k+1} = x_k + \alpha_k d_k, \quad k \geq 0, \tag{2}$$

where x_k is the current approximation to a solution, and $d_k \in \mathbb{R}^n$ is a search direction defined by

$$d_k = \begin{cases} -g_0, & k = 0, \\ -g_k + \beta_k d_{k-1}, & k \geq 1. \end{cases} \tag{3}$$

The vector g_k denotes $g(x_k)$, the β_k (called here CG parameter) is a scalar, which distinguishes any two different CG methods, and the step length $\alpha_k > 0$ is usually determined to satisfy the strong Wolfe line search conditions

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k, \tag{4}$$

2020 *Mathematics Subject Classification*. 90C30.
Keywords. Unconstrained optimization; Three-term conjugate gradient method; Sufficient descent property; Global convergence
Received: 29 May 2021; Revised: 07 March 2022; Accepted: 12 February 2024
Communicated by Predrag Stanimirović
Email addresses: minalotfi@modares.ac.ir (Mina Lotfi), hossei_m@modares.ac.ir (S. Mohammad Hosseini)

$$|g(x_k + \alpha_k d_k)^T d_k| \leq \sigma |g_k^T d_k|, \tag{5}$$

where $0 < \delta < \frac{1}{2}$, $\delta < \sigma < 1$. The most well-known CG methods include Fletcher–Reeves (FR) method [1], the Dai–Yuan (DY) method [2], the Hestenes–Stiefel (HS) method [3], Liu–Storey (LS) method [4] and the Polak–Ribière–Polyak (PRP) method [5, 6]. From theoretical and numerical performance point of view, the FR and DY methods have strong convergence properties, while the HS, LS and PRP methods have better computational performances [18]. Moreover, HS method satisfies the conjugacy condition $d_k^T y_{k-1} = 0$, for all $k \geq 0$, independent of the line search conditions and the objective function convexity. Dai and Liao (DL) [19], with a modification of conjugacy condition, introduced a class of CG methods. Although the DL method seldom generates uphill search direction in an actual computation, this search direction is not necessarily a descent one in theory. This motivated many researchers to make various modifications on the DL method, in order to achieve some descent properties, see [8, 9, 11, 13, 15].

In this paper, we propose a three-term version of the modified DL method proposed by Cheng et al. [15]. As a significant property, our method ensures sufficient descent, independent of any line search. Its global convergence and the r-linear convergence rate, are shown under standard assumptions. We have examined our method on the test problems from CUTEst collection. Numerical results show efficiency and robustness of our proposed method in practice. The rest of this paper is organized as follows. In Section 2, we present details of the new CG method and its computational algorithm. In Section 3, we establish the global convergence property of the proposed method for general functions. The r-linear convergence rate of our method is discussed in Section 4. Numerical results, obtained from applying the new method on the unconstrained optimization problems from CUTEst collection, are reported in Section 5.

2. Motivation and algorithm

Based on an extended conjugacy condition, Dai and Liao (DL) [19] presented a class of CG methods with CG parameter given by

$$\beta_k^{DL} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} - t \frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}},$$

where $t > 0$, $y_{k-1} = g_k - g_{k-1}$, and $s_{k-1} = x_k - x_{k-1}$. They have proved the convergence of those methods for convex functions; moreover, by using the truncation technique of [7, 10], they have also established the convergence for more general functions. Although the DL method is computationally efficient, it may fail to generate a descent direction (i.e. there may be some k for which $g_k^T d_k < 0$ fails to hold). In [8], Hager and Zhang (HZ) proposed a subclass of DL methods known as CG_DESCENT,

$$\beta_k^{HZ} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} - \theta \frac{\|y_{k-1}\|^2 g_k^T d_{k-1}}{(d_{k-1}^T y_{k-1})^2}, \qquad \theta > \frac{1}{4}, \tag{6}$$

where $\|.\|$ stands for Euclidean norm. They showed that the search direction in this method satisfies the sufficient descent condition $g_k^T d_k \leq (\frac{1}{4\theta} - 1)\|g_k\|^2$. In addition, in order to guarantee the global convergence, they have updated β_k^{HZ} by

$$\beta_k^{HZ+} = \max\{\eta_k, \beta_k^{HZ}\}, \quad \text{with } \eta_k = -\frac{1}{\|d_{k-1}\| \min\{\eta, \|g_{k-1}\|\}}, \tag{7}$$

and $\eta > 0$. Numerical results obtained by CG_DESCENT method outperforms many existing CG methods. Yao et al. [11, 13] proposed two DL-type conjugate gradient methods where their search directions satisfy the sufficient descent condition under the strong Wolfe line search for $\sigma < \frac{1}{4}$ and $\sigma < \frac{1}{3}$. They obtained those methods by replacing the first term in the expression of β_k^{DL}, respectively, by β_k^{MHS} (of the modified HS method [12]) and β_k^{WYL} (of the WYL method [14]). Recently, based on the same approach and by using a modified version of LS method (MLS) of [12], Cheng et al. [15] proposed the following CG parameter

$$\beta_k^{MLS-DL} = \beta_k^{MLS} - t\frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}}, \tag{8}$$

where $\beta_k^{MLS} = \frac{g_k^T(g_k - \frac{\|g_k\|}{\|g_{k-1}\|}g_{k-1})}{-d_{k-1}^T g_{k-1}}$.

In [15], it is shown that the MLS-DL method satisfies the sufficient descent condition under the Wolfe line search for $\sigma < \frac{1}{2}$, and it is also computationally superior to the CG_DESCENT method [8]. In spite of the mentioned promising features of MLS-DL method, the condition $\sigma < \frac{1}{2}$, may limit the performance of the proposed method, see [25] for more details. In [22], Narushima et al. proposed a modified FR method, in which the directions are defined by

$$d_k = -g_k + \beta_k^{FR} d_{k-1} - \beta_k^{FR}\frac{g_k^T d_{k-1}}{\|g_k\|^2} g_k. \tag{9}$$

Moreover, it has been shown [22] that the search direction (9) satisfies the sufficient descent condition

$$g_k^T d_k = -\|g_k\|^2. \tag{10}$$

independently of choices of the CG parameter β_k.

Here, to increase computational efficiency and robustness of the MLS-DL method, we employ the idea of the modified FR method and propose the following search direction

$$d_k = -g_k + \beta_k^{MLS-DL} d_{k-1} - \beta_k^{MLS-DL}\frac{g_k^T d_{k-1}}{\|g_k\|^2} g_k. \tag{11}$$

Now, we express the steps of the new proposed method in Algorithm 2.1.

Algorithm 2.1. *New Three-Term Conjugate Gradient Method*
Step 0 : Consider constants $\varepsilon > 0$, $0 < \delta < \frac{1}{2}$, $\delta < \sigma < 1$, *choose an initial point* $x_0 \in \mathbb{R}^n$ *and set* $k = 0$, $t > 0$, $d_0 = -g_0$.
Step 1 : Stop if $\|g_k\|_\infty < \varepsilon$.
Step 2 : Determine the step length α_k *by the strong Wolfe line search (4) and (5).*
Step 3 : Let $x_{k+1} = x_k + \alpha_k d_k$; *Set* $k = k + 1$.
Step 4 : Compute β_k^{MLS-DL} *by (8).*
Step 5 : Compute d_k *by (11), go to Step 1.*

3. Convergence Analysis

In this section, we establish global convergence property of Algorithm 2.1. To this end, the following assumptions are considered on the objective function.

Assumption 4.1 The level set $\mathcal{L} = \{x \in \mathbb{R}^n | f(x) \leq f(x_0)\}$ is bounded, namely, there exists a constant $B > 0$ such that

$$\|x\| \leq B, \quad \forall x \in \mathcal{L}. \tag{12}$$

Assumption 4.2 In some neighborhood $\mathcal{N}$ of $\mathcal{L}$, f is continuously differentiable and its gradient is Lipschitz continuous, namely, there exists a positive constant $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \qquad \forall x, y \in \mathcal{N}. \tag{13}$$

Remark 4.1 Assumption 4.1 implies that there exists a positive constant M such that

$$\|g(x)\| \leq M, \quad \forall x \in \mathcal{L}. \tag{14}$$

The following well-known lemma which was proved by Zoutendijk [21] will be needed for convergence analysis.

Lemma 3.1. *Suppose that the Assumptions* 4.1 *and* 4.2 *hold. Consider any CG method of the form* (2) *and* (3)*, where* d_k *is a sufficient descent direction and the steplength* α_k *satisfies the strong Wolfe line search conditions. Then*

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

which, from (10), reduces to

$$\sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty. \tag{15}$$

Lemma 3.2. *Suppose that the Assumptions* 4.1 *and* 4.2 *hold. Let* x_k *be the sequence of points generated by Algorithm 2.1. Moreover, suppose that there exists a constant* $\mu > 0$ *such that*

$$\|g_k\| \geq \mu, \quad \forall k \geq 0. \tag{16}$$

Then there exist positive constants C1 and C2 such that

$$|\beta_k^{MLS-DL}| \leq C_1 \|s_{k-1}\|, \qquad |\beta_k^{MLS-DL}| \frac{|g_k^T d_{k-1}|}{\|g_k\|^2} \leq C_2 \|s_{k-1}\|. \tag{17}$$

Proof. Using the strong Wolfe condition (5) together with (10), we obtain

$$d_{k-1}^T y_{k-1} \geq -(1-\sigma) g_{k-1}^T d_{k-1} = (1-\sigma)\|g_{k-1}\|^2. \tag{18}$$

Using the same lines of proof of Lemma 3.4 in [15], we get

$$|\beta_k^{MLS-DL}| \leq \frac{2\|g_k\|\|y_{k-1}\|}{\|g_{k-1}\|^2} + t\frac{\|g_k\|\|s_{k-1}\|}{d_{k-1}^T y_{k-1}}. \tag{19}$$

From (13), (14), (16), (18) and (19) we conclude

$$\begin{aligned} |\beta_k^{MLS-DL}| &\leq \frac{2\|g_k\|\|y_{k-1}\|}{\|g_{k-1}\|^2} + t\frac{\|g_k\|\|s_{k-1}\|}{(1-\sigma)\|g_{k-1}\|^2} \\ &\leq \frac{2ML\|s_{k-1}\|}{\mu^2} + t\frac{M\|s_{k-1}\|}{(1-\sigma)\mu^2} \\ &\leq \left(\frac{2ML}{\mu^2} + t\frac{M}{(1-\sigma)\mu^2}\right)\|s_{k-1}\|. \end{aligned}$$

Setting $C_1 = \frac{2ML}{\mu^2} + t\frac{M}{(1-\sigma)\mu^2}$, it follows that

$$|\beta_k^{MLS-DL}| \leq C_1 \|s_{k-1}\|. \tag{20}$$

Combining this upper bound for $|\beta_k^{MLS-DL}|$ with (5), (10), (14) and (16), we have

$$|\beta_k^{MLS-DL}| \frac{|g_k^T d_{k-1}|}{\|g_k\|^2} \leq C_1 \|s_{k-1}\| \frac{\sigma |g_{k-1}^T d_{k-1}|}{\|g_k\|^2} \tag{21}$$

$$\leq C_1 \|s_{k-1}\| \frac{\sigma \|g_{k-1}\|^2}{\|g_k\|^2} \tag{22}$$

$$\leq C_1 \|s_{k-1}\| \frac{\sigma M^2}{\mu^2} \tag{23}$$

$$\leq C_2 \|s_{k-1}\|. \tag{24}$$

where $C_2 = C_1\frac{\sigma M^2}{\mu^2}$. This completes the proof of the Lemma. □

Lemma 3.3. *Suppose that the Assumptions 4.1 and 4.2 hold. Let x_k be the sequence of points generated by Algorithm 2.1. If there exists a constant $\mu > 0$ such that $||g_k|| > \mu$ for all $k > 0$, then we have*

$$\sum_{k=0}^{\infty} ||u_k - u_{k-1}||^2 < \infty, \tag{25}$$

where $u_k = \frac{d_k}{||d_k||}$.

Proof. From $||g_k|| > \mu > 0$ and descent direction property (10), it follows that $d_k \neq 0$, and thus u_k is well defined. Moreover, from (15) and $||g_k|| > \mu > 0$, we conclude that

$$\sum_{k=0}^{\infty} \frac{||g_k||^4}{||d_k||^2} < \infty. \tag{26}$$

Let us define

$$r_k = \frac{w_k}{||d_k||},\ \delta_k = \beta_k^{MLS}\frac{||d_{k-1}||}{||d_k||}, \tag{27}$$

where

$$w_k = -\left(1 + \beta_k^{MLS-DL}\frac{g_k^T d_{k-1}}{||g_k||^2}\right)g_k - t\frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}}d_{k-1}. \tag{28}$$

Then, it follows from (8), (11), (27) and (28) that:

$$\begin{aligned} u_k &= \frac{d_k}{||d_k||} \\ &= \frac{-g_k + \beta_k^{MLS-DL}d_{k-1} - \beta_k^{MLS-DL}\frac{g_k^T d_{k-1}}{||g_k||^2}g_k}{||d_k||} \\ &= \frac{-g_k - \beta_k^{MLS-DL}\frac{g_k^T d_{k-1}}{||g_k||^2}g_k + (\beta_k^{MLS}d_{k-1} - t\frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}}d_{k-1})}{||d_k||} \\ &= \frac{-\left(1 + \beta_k^{MLS-DL}\frac{g_k^T d_{k-1}}{||g_k||^2}\right)g_k - t\frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}}d_{k-1}}{||d_k||} + \beta_k^{MLS}\frac{||d_{k-1}||}{||d_{k-1}||}\frac{d_{k-1}}{||d_k||} \\ &= r_k + \delta_k u_{k-1}. \end{aligned}$$

By similar arguments of the proof of Lemma 3.1 in [8], we obtain

$$||u_k - u_{k-1}|| \leq 2||r_k||. \tag{29}$$

Since $||r_k|| = \frac{||w_k||}{||d_k||}$, it is enough to obtain an upper bound for $||w_k||$. Utilizing (5), (12), (14), (17), (18) and (28) gives

$$\begin{aligned} ||w_k|| &\leq ||g_k|| + |\beta_k^{MLS-DL}|\frac{|g_k^T d_{k-1}|}{||g_k||^2}||g_k|| + t\frac{\sigma|g_{k-1}^T d_{k-1}|||s_{k-1}||}{(1-\sigma)||g_{k-1}||^2} \\ &\leq M + C_2||s_{k-1}||M + t\frac{\sigma||g_{k-1}||^2||s_{k-1}||}{(1-\sigma)\mu^2} \\ &\leq M(1 + 2C_2B + tM\frac{2\sigma B}{(1-\sigma)\mu^2}). \end{aligned} \tag{30}$$

Thus, it follows from (26), (29) and (30) that

$$\begin{aligned}\sum_{k=0}^{\infty}\|u_k-u_{k-1}\|^2 &\le 4\sum_{k=0}^{\infty}\|r_k\|^2\\ &\le 4\sum_{k=0}^{\infty}\frac{\|w_k\|^2}{\|d_k\|^2}\\ &\le \frac{4M^2(1+2C_2B+tM\frac{2\sigma B}{(1-\sigma)\mu^2})^2}{\mu^4}\sum_{k=0}^{\infty}\frac{1}{\|d_k\|^2}\\ &< \infty, \end{aligned} \tag{31}$$

which completes the proof. □

Theorem 3.4. *Suppose that Assumptions* 4.1 *and* 4.2 *hold. If the sequence* $\{x_k\}$ *is generated by Algorithm 2.1, then either* $\|g_k\| = 0$ *for some k, or*

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

Proof. We proceed by contradiction. We assume that there exists a constant $\mu > 0$ such that

$$\|g_k\| \ge \mu, \quad \forall k \ge 0.$$

By squaring both sides of (11), and using (12), (14) and (17) we obtain

$$\begin{aligned}\|d_k\|^2 &\le \left(\|g_k\| + |\beta_k^{MLS-DL}|\frac{|g_k^T d_{k-1}|}{\|g_k\|^2}\|g_k\| + |\beta_k^{MLS-DL}|\|d_{k-1}\|\right)^2\\ &\le (M + MC_2\|s_{k-1}\| + C_1\|s_{k-1}\|\|d_{k-1}\|)^2\\ &\le 2(M+2MC_2B)^2 + 2C_1^2\|s_{k-1}\|^2\|d_{k-1}\|^2.\end{aligned}$$

Following the same line of proof of Theorem 3.2 in [8], it can be seen that $\|d_k\| \le \gamma$ holds for some positive constant γ. This leads to

$$\sum_{k=0}^{\infty}\frac{\|g_k\|^4}{\|d_k\|^2} \ge \frac{\mu^4}{\gamma^2}\sum_{k=0}^{\infty}1 = \infty,$$

which contradicts (15). Therefore, the proof is complete. □

4. Convergence Rate Analysis

In this section, we provide the r-linear convergence rate of Algorithm 2.1. We assume that the objective function f is twice continuously differentiable and uniformly convex, namely, there are positive constants $m_1 \le m_2$ such that

$$m_1\|y\|^2 \le y^T\nabla^2 f(x)y \le m_2\|y\|^2, \quad \forall y \in \mathbb{R}^n, x \in \mathcal{L}.$$

This implies that the sequence $\{x_k\}$ converges to the unique minimizer x^* of the problem (1). Moreover, it implies that:

$$\frac{1}{2}m_1\|x-x^*\|^2 \le (f(x)-f(x^*)) \le \frac{1}{2}m_2\|x-x^*\|^2, \quad \forall x \in \mathbb{R}^n, \tag{32}$$

$$m_1\|x-x^*\| \le \|g(x)\| \le m_2\|x-x^*\|, \quad \forall x \in \mathbb{R}^n. \tag{33}$$

Theorem 4.1. *Suppose that the Assumptions* 4.1 *and* 4.2 *hold. Moreover, assume that the objective function f is twice continuously differentiable and uniformly convex. If the sequence $\{x_k\}$ is generated by Algorithm 2.1, there exists a constant $D > 0$ such that*

$$\alpha_k \geq D, \quad \forall k > 0. \tag{34}$$

Proof. Since f is a twice continuously differentiable and uniformly convex function, there exist positive constants ν and c such that

$$s_k^T y_k \geq \nu \|s_k\|^2, \quad -g_k^T d_k \geq c\alpha_k \|d_k\|^2, \quad \text{(see[16])}. \tag{35}$$

Following the same lines of proof of Lemma 3.4 in [15], it can be shown that

$$|\beta_k^{MLS}| \leq \frac{2\|g_k\|\|y_{k-1}\|}{|d_{k-1}^T g_{k-1}|}. \tag{36}$$

From (13), (35) and (36) it follows that

$$\begin{aligned} |\beta_k^{MLS}| &\leq \frac{2\|g_k\|\|y_{k-1}\|}{c\alpha_{k-1}\|d_{k-1}\|^2} \\ &\leq \frac{2\|g_k\|L\|s_{k-1}\|}{c\alpha_{k-1}\|d_{k-1}\|^2} \\ &\leq \frac{2L\|g_k\|}{c\|d_{k-1}\|}. \end{aligned} \tag{37}$$

Thus, by inserting (35) and (37) into (11), we conclude that

$$\begin{aligned} \|d_k\| &\leq \|g_k\| + |\beta_k^{MLS-DL}| \frac{\|g_k\|\|d_{k-1}\|}{\|g_k\|^2} \|g_k\| + |\beta_k^{MLS-DL}| \|d_{k-1}\| \\ &\leq \|g_k\| + 2\left(\frac{2L\|g_k\|}{c\|d_{k-1}\|} + t\frac{\|g_k\|\|s_{k-1}\|}{d_{k-1}^T y_{k-1}}\right)\|d_{k-1}\| \\ &\leq \|g_k\| + 2\left(\frac{2L\|g_k\|}{c} + t\frac{\|g_k\|\|s_{k-1}\|}{s_{k-1}^T y_{k-1}}\|s_{k-1}\|\right) \\ &\leq \|g_k\| + 2\left(\frac{2L\|g_k\|}{c} + t\frac{\|g_k\|\|s_{k-1}\|^2}{\nu\|s_{k-1}\|^2}\right) \\ &\leq \left(1 + 2\left(\frac{2L}{c} + t\frac{1}{\nu}\right)\right)\|g_k\| \end{aligned}$$

which implies

$$\frac{\|g_k\|^2}{\|d_k\|^2} \geq \left(1 + 2\left(\frac{2L}{c} + t\frac{1}{\nu}\right)\right)^{-2} = \lambda. \tag{38}$$

On the other hand, from (5), (10), and (13), we obtain

$$(1-\sigma)\|g_k\|^2 = (\sigma - 1)g_k^T d_k \leq (g_{k+1} - g_k)^T d_k \leq L\alpha_k \|d_k\|^2. \tag{39}$$

Thus, from (38) and (39), we get

$$\alpha_k \geq \frac{1-\sigma}{L}\frac{\|g_k\|^2}{\|d_k\|^2} \geq D,$$

where $D = \frac{1-\sigma}{L}\lambda$. This completes the proof □

The following theorem, which is similar to Theorem 2.6 in [16], establishes the r-linear convergence rate of the proposed method.

Theorem 4.2. *Suppose that the Assumptions* 4.1 *and* 4.2 *hold, and the objective function f is twice continuously differentiable and uniformly convex. Then the sequence of points $\{x_k\}$ generated by Algorithm 2.1, is r-linear convergent, i.e. there are constants $a > 0$ and $r \in]0, 1[$ such that*

$$\|x_k - x^*\| \leq ar^k, \quad \text{for all } k \geq 0.$$

Proof. The statements of the proof of Theorem 2.6 in [16], it is simply shown that

$$f(x_{k+1}) - f(x^*) \leq (1 - \frac{2\delta D m_1^2}{m_2})(f(x_k) - f(x^*)).$$

This implies that

$$f(x_k) - f(x^*) \leq r^{2k}(f(x_0) - f(x^*)), \tag{40}$$

where $r = \sqrt{(1 - \frac{2\delta D m_1^2}{m_2})}$. Then, from (32) and (40), we obtain

$$\|x_k - x^*\|^2 \leq \frac{2}{m_1}(f(x_k) - f(x^*)) \leq \frac{2}{m_1}(f(x_0) - f(x^*))r^{2k},$$

which implies $\|x_k - x^*\|^2 \leq a^2 r^{2k}$ for $a = \sqrt{\frac{2}{m_1}(f(x_0) - f(x^*))}$. □

5. Numerical Results

In this section, we present the computational results obtained from the implementation of the Algorithm 2.1, denoted by "TMLS-DL". We compare the performance of TMLS-DL with several recent well-performing methods :

- HZ + : denotes CG_DESCENT [8] by using the values $\theta = 2$ in (6), $\eta = 0.01$ in (7), resp.
- MLS : The CG method proposed by [12].
- MLS-DL : The CG method proposed by [15].
- NLS : Algorithm 1 proposed by [16].
- TTRMIL : The three-term CG method proposed by [23].
- NTPA : New three-term Perry Algorithm proposed by [24].

The codes of all the algorithms (methods) investigated in this study were written in MATLAB, and run on a PC (CPU 2.5 GHZ, RAM 3.8 GB) with Linux operating system. The test problems were taken from the CUTEst library [20]. The dimension of the problems ranges from 50 to 10,000. Table 1 shows names of those problems and their dimensions. In all the algorithms, we used the strong Wolfe line search conditions and computed the initial guess of the step length by scheme proposed in [21]. In all algorithms, we have used the following values for the parameters:

$$\varepsilon = 10^{-6}, \quad \delta = 0.01, \quad \sigma = 0.1,$$

$$\alpha_{0,0} = 1, \quad \alpha_{k,0} = \alpha_{k-1}\frac{g_{k-1}^T d_{k-1}}{g_k^T d_k}.$$

Table 1 The test problems and their dimensions.

No	Name	Dim	No	Name	Dim	No	Name	Dim
1	ARGLINA	200	2	ARGLINA	100	3	BDEXP	1000
4	BDEXP	5000	5	BIGSB1	100	6	BIGSB1	1000
7	COSINE	100	8	COSINE	1000	9	CURLY10	10000
10	CURLY20	10000	11	CURLY30	10000	12	DEGTRID	110
13	DIXMAANA	3000	14	DIXMAANA	9000	15	DIXMAANB	3000
16	DIXMAANB	9000	17	DIXMAANC	3000	18	DIXMAANC	9000
19	DIXMAAND	3000	20	DIXMAAND	9000	21	DIXMAANE	3000
22	DIXMAANE	9000	23	DIXMAANF	3000	24	DIXMAANF	9000
25	DIXMAANG	3000	26	DIXMAANG	9000	27	DIXMAANH	3000
28	DIXMAANH	9000	29	DIXMAANI	3000	30	DIXMAANI	9000
31	DIXMAANJ	3000	32	DIXMAANJ	9000	33	DIXMAANK	3000
34	DIXMAANK	9000	35	DIXMAANL	1500	36	DIXMAANL	9000
37	DIXMAANM	3000	38	DIXMAANM	9000	39	DIXMAANN	3000
40	DIXMAANN	9000	41	DIXMAANO	3000	42	DIXMAANO	9000
43	DIXMAANP	3000	44	DIXMAANP	9000	45	DQDRTIC	1000
46	DQDRTIC	5000	47	DQRTIC	1000	48	DQRTIC	5000
49	DIXON3DQ	100	50	DIXON3DQ	1000	51	DECONVU	61
52	EG2	1000	53	FLETCBV	1000	54	FLETCHCR	100
55	FLETCHCR	100	56	FMINSRF2	5625	57	FMINSRF2	10000
58	FMINSURF	5625	59	FMINSURF	10000	60	LIARWHD	5000
61	LIARWHD	10000	62	LMINSURF	5625	63	LMINSURF	10000
64	MANCINO	50	65	MANCINO	100	66	MOREBV	1000
67	MOREBV	5000	68	MSQRTALS	529	69	MSQRTALS	1024
70	MSQRTBLS	529	71	MSQRTBLS	1024	72	NLMSURF	5625
73	NLMSURF	10000	74	NONDIA	1000	75	NONDIA	5000
76	NONDQUAR	500	77	NONDQUAR	1000	78	NONSCOMP	5000
79	POWELLSG	5000	80	POWELLSG	10000	81	SPARSQUR	5000
82	SPARSQUR	10000	83	SPMSRTLS	1000	84	SPMSRTLS	4999
85	TOINTGSS	5000	86	TOINTGSS	10000	87	TRIDIA	5000
88	TRIDIA	10000	89	WOOD	4000	90	WOOD	10000

Moreover, the algorithms were stopped when the number of iteration exceeded 10000 or $\|g_k\|_\infty \leq \varepsilon$.
We adopted the performance profile of Dolan and Moré [17] (in log_2 scale) to compare numerical results of the reported algorithms. The performance comparison of the above-mentioned methods has been conducted. Figs. 1-4, illustrate the performance comparison of TMLS-DL, MLS, MLS-DL and HZ+ in terms of number of iterations n_i, number of function evaluations n_f, number of gradient evaluations n_g, and CPU time t in second, respectively. Figs. 5-8, do the same for three-term methods NLS, TTRMIL, NTPA, and TMLS-DL. In Tables 2-3, we present the percentage of the test problems that are solved by each algorithm with the lowest value of n_i, n_f, n_g and t. The Tables 2-3 and Figs. 1-8 support the conclusion that the method TMLS-DL performs better than the other considered methods, with respect to the number of iterations, the number of function evaluations, and the number of gradient evaluations.

6. Conclusions

In this paper, we presented a new three-term CG method derived through combining the CG method of Cheng et al [15] and the idea of the modified FR method [22]. As a remarkable feature, the search directions of the proposed method satisfy the sufficient descent condition, independently of line searches. Under the strong Wolfe line search with $\sigma \in]0, 1[$, theoretical results show that the new method inherits global

Table 2 Percentage of the test problems that each method solves with the lowest value of n_i, n_f, n_g and t

	$HZ+$ (%)	MLS(%)	$MLS-DL$(%)	$TMLS-DL$(%)
n_i	21	27	29	**69**
n_f	24	18	18	**72**
n_g	26	17	18	**72**
t	14	20	18	**44**

Table 3 Percentage of the test problems that each method solves with the lowest value of n_i, n_f, n_g and t

	NLS(%)	$TTRMIL$(%)	$NTPA$(%)	$TMLS-DL$(%)
n_i	38	23	9	**70**
n_f	36	22	8	**70**
n_g	38	20	8	**70**
t	20	10	6	**60**

convergence property. Moreover, when the objective function is uniformly convex, we established the r-linear convergence rate of our proposed method. Numerical comparisons on the test problems from the CUTEst library have been reported which indicate the efficiency and robustness of our proposed method in practice. As a future work, it would be interesting to improve our proposed method for nonsmooth unconstrained problems.

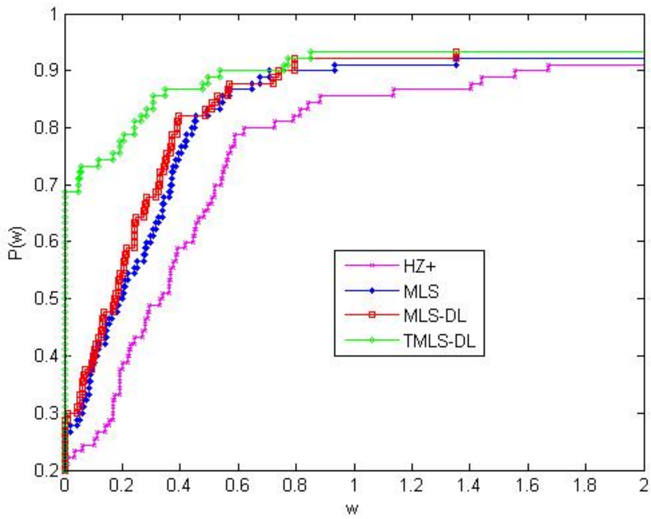

Figure 1: Performance profile of methods in terms of number of iterations.

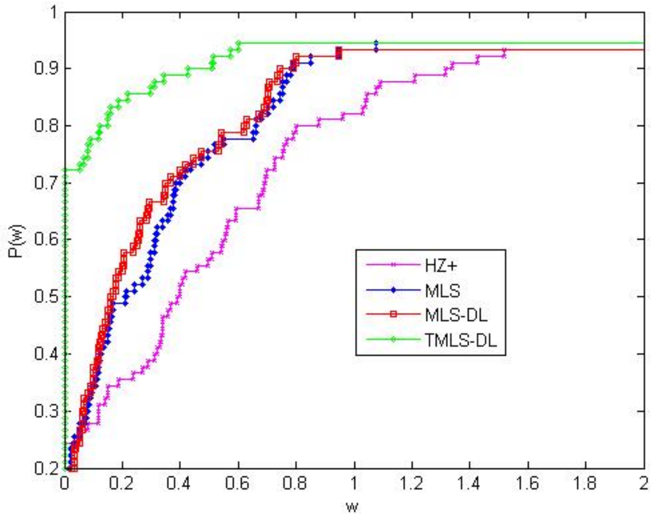

Figure 2: Performance profile of methods in terms of number of function evaluations.

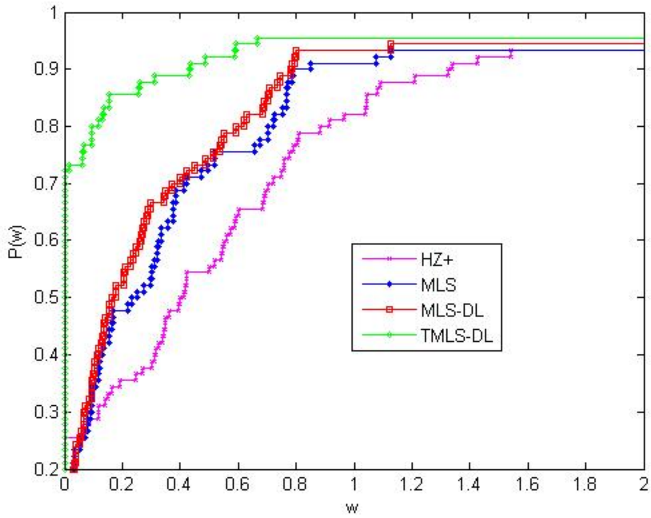

Figure 3: Performance profile of methods in terms of number of gradient evaluations.

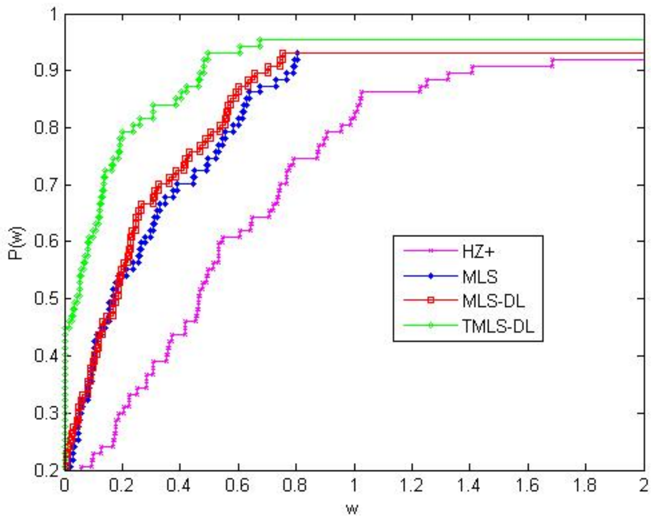

Figure 4: Performance profile of methods in terms of CPU time.

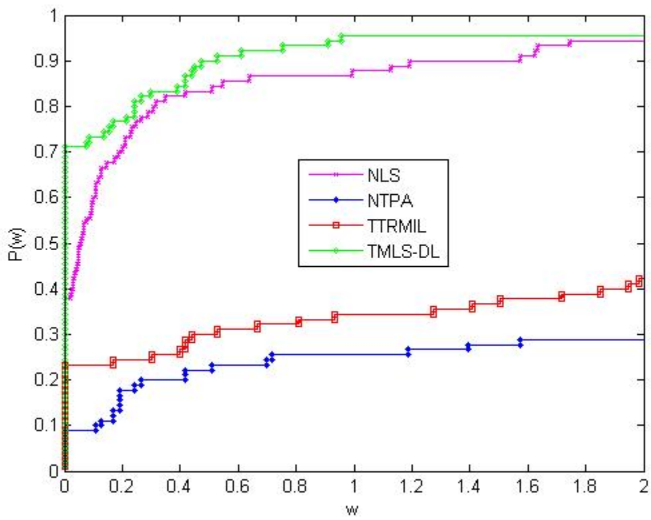

Figure 5: Performance profile of methods in terms of number of iterations.

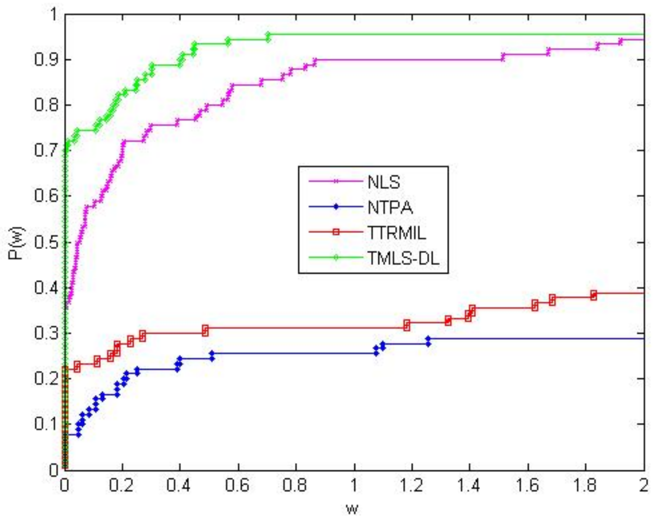

Figure 6: Performance profile of methods in terms of number of function evaluations.

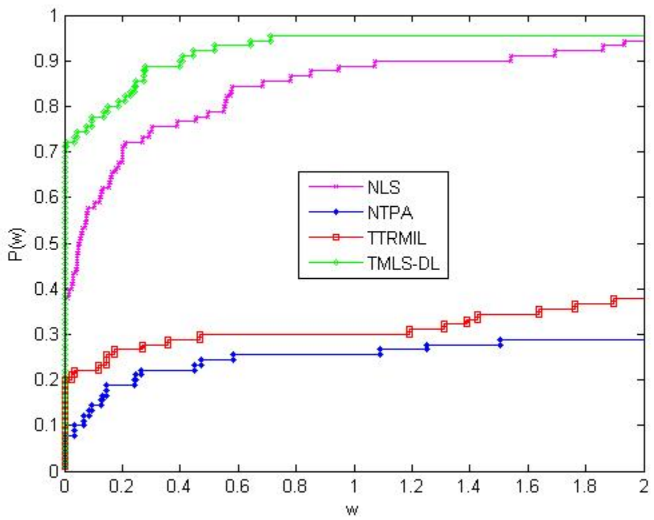

Figure 7: Performance profile of methods in terms of number of gradient evaluations.

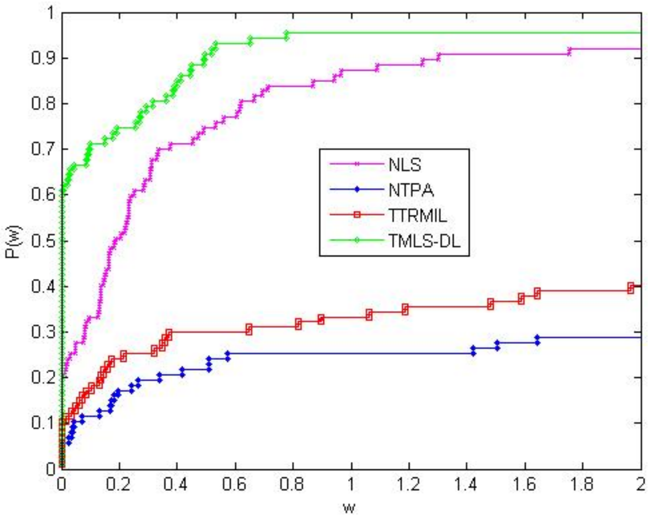

Figure 8: Performance profile of methods in terms of CPU time.

References

[1] R. Fletcher, C.M. Reeves, Function minimization by conjugate gradients. Comput. J. 7 (2) (1964), 149-154.

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

[3] M.R. Hestenes, E.L. Stiefel, Methods of conjugate gradients for solving linear systems, J. Res. Natl. Bur. Stand. (1952), 409-432.

[4] Y. Liu, C. Storey, Efficient generalized conjugate gradient algorithms. I. Theory. J. Optim. Theory Appl. 69 (1) (1991), 129-137.

[5] E. Polak, G. Ribière, Note sur la convergence des méthodes de directions conjuguées, Rev. Fr. Inform. Rech. Oper. 16 (1952), 35-43.

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

[7] J.C. Gilbert, J. Nocedal, Global convergence properties of conjugate gradient methods for optimization. SIAM J. Optim. 2 (1992), 21-42.

[8] W.W. Hager, H.A. Zhang, A new conjugate gradient method with guaranteed descent and an efficient line search. SIAM J. Optim. 16 (2005), 170-192.

[9] M. Lotfi, S. M. Hosseini, An efficient Dai-Liao type conjugate gradient method by reformulating the CG parameter in the search direction equation. J. Comput. Appl. Math. 371 (2020), 112708.

[10] M. Al-Baali, Descent property and global convergence of the Fletcher-Reeves method with inexact line search. IMA J. Numer. Anal. 5 (1985), 121-124.

[11] S. Yao, B. Qin, A hybrid of DL and WYL nonlinear conjugate gradient methods. Abstr. Appl. Anal. 1-9, Article ID 279891 (2014) .

[12] S. Yao, Z. Wei, H. Huang, A note about WYL's conjugate gradient method and its applications. Appl. Math. Comput. 191(2) (2007), 381-388.

[13] S. Yao, X. Lu, Z. Wei, A conjugate gradient method with global convergence for large-scale unconstrained optimization problems. J. Appl. Math. 1-9, Article ID 730454 (2013).

[14] Z. Wei, S. Yao, L. Liu, The convergence properties of some new conjugate gradient methods. Appl. Math. Comput. 183 (2)(2006), 1341-1350.

[15] Y. Cheng, Q. Mou, X. Pan, S. Yao, A sufficient descent conjugate gradient method and its global convergence. Optim. Methods Softw. 31(3), 577-590 (2016)

[16] M. Li, A. Qu, A sufficient descent Liu-Storey conjugate gradient method and its global convergence. Optimization. 64 (9)(2015), 1919-1934.

[17] E.D. Dolan, J.J. Moré, Benchmarking optimization software with performance profiles. Math. Progr. 91 (2) (2002), 201–213.

[18] N. Andrei, Numerical comparison of conjugate gradient algorithms for unconstrained optimization. Stud. Inform. Control. 16 (2007), 333-352.

[19] Y.H. Dai, L.Z. Liao, New conjugacy conditions and related nonlinear conjugate gradient methods. Appl. Math. Optim. 43 (2001), 87-101.

[20] N.I.M. Gould, D. Orban, ph. Toint, CUTEst : a constrained and unconstrained testing environment with safe threads for mathematical optimization. Comput. Optim. Appl. 60 (3) (2015), 545-557.

[21] G. Zoutendijk, Nonlinear programming, computational methods, in: Integer and nonlinear programming, J. Abadie (ed.), North-Holland, Amsterdam, 1970, pp. 37–86.

[22] Y. Narushima, H. Yabe, J.A. Ford, A three-term conjugate gradient method with sufficient descent property for unconstrained optimization, SIAM J. Optim. 21 (2011), 212- 230 .
[23] J.K. Liu, Y.M. Feng, L.M. Zou, Some three-term conjugate gradient methods with the inexact line search condition. Calcolo 55 (2018), 1-16.
[24] S. Yao, L. Ning, An adaptive three-term conjugate gradient method based on self-scaling memoryless BFGS matrix. J. Comput. Appl. Math. 332 (2018), 72-85.
[25] Y. Zheng, B. Zheng, Two New Dai-Liao-Type Conjugate Gradient Methods for Un- constrained Optimization Problems, J. Optim. Theory Appl. (2017) DOI:10.1007/s10957- 017-1140-1.