TAMKANG JOURNAL OF MATHEMATICS
Volume 54, Number 2, 139-154, 2023
DOI:10.5556/j.tkjm.54.2023.4185

Solving unconstrained optimization problems with some three-term conjugate gradient methods

Ladan Arman, Yuanming Xu, Mohammad Reza Bayat, Long Liping

Abstract. In this paper, based on the efficient Conjugate Descent (CD) method, two generalized CD algorithms are proposed to solve the unconstrained optimization problems. These methods are three-term conjugate gradient methods which the generated directions by using the conjugate gradient parameters and independent of the line search satisfy in the sufficient descent condition. Furthermore, under the strong Wolfe line search, the global convergence of the proposed methods are proved. Also, the preliminary numerical results on the CUTEst collection are presented to show effectiveness of our methods.

Keywords. Conjugate gradient method, unconstrained optimization, global convergence, strong Wolfe line search

1 Introduction

Consider the following unconstrained optimization problem

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

where $f:\mathbb{R}^n\to\mathbb{R}$ is a continuously differentiable function and its gradient $g := \bigtriangledown f$ is available. Conjugate Gradient (CG) methods are effective iterative methods for solving (1.1), especially for large-scale problems. The important properties of these methods are the use only first-order derivatives, little storage and computation requirements, and strong local and global convergence properties [1, 9, 18, 24]. Starting from an initial guess $x_0 \in \mathbb{R}^n$, the CG methods generate a sequence $\{x_k\}_{k\geq 0}$ as

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

where $\alpha_k > 0$ is step-length and usually obtained using some inexact line search. Furthermore, d_k is the search direction calculated by

$$d_k = \begin{cases} -g_k, & k=0, \\ -g_k + \beta_k d_{k-1}, & k\geq 1, \end{cases} \tag{1.3}$$

Received date: May 3, 2021; Published online: September 16, 2022.
2010 *Mathematics Subject Classification*. 90C30, 65K05.
Corresponding author: Long Liping.

in which $g_k = g(x_k)$ and β_k is a scalar. There are many variants of CG methods, which are obtained with different choices for the parameter β_k. The most important CG methods proposed by Fletcher–Reeves (FR) [16], Hestenes–Stiefel (HS) [19], Conjugate Descent (CD) by Fletcher [15], Polak–Ribiére–Polyak (PRP) [24, 25], Dai-Yuan (DY) [10] and Hager-Zhang (HZ) [17] are defined by

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \qquad \beta_k^{HS} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, \qquad \beta_k^{CD} = -\frac{\|g_k\|^2}{g_{k-1}^T d_{k-1}}, \tag{1.4}$$

$$\beta_k^{PRP} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}, \qquad \beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}}, \qquad \beta_k^{HZ} = \left(y_{k-1} - 2d_{k-1}\frac{\|y_{k-1}\|^2}{d_{k-1}^T y_{k-1}}\right)^T \frac{g_k}{d_{k-1}^T y_{k-1}}, \tag{1.5}$$

in which $\|\cdot\|$ is the Euclidean norm and $y_{k-1} = g_k - g_{k-1}$. These methods are the identical where the objective function f is quadratic and exact line search is used [23], but for general objective functions the behavior of these methods is different. CG methods are used in many applications problems such as image denoising and image deblurring in image processing, see [20].

Generally, in the iterative methods, we need the search direction d_k satisfy the descent condition

$$g_k^T d_k < 0, \qquad \forall k \geq 0. \tag{1.6}$$

In order to guarantee the global convergence of CG methods, the direction d_k must satisfy the sufficient descent condition

$$g_k^T d_k \leq -c\|g_k\|^2, \qquad \forall k \geq 0, \tag{1.7}$$

in which c is a positive constant. There are many CG methods which satisfy (1.7), see [3, 17, 22].

In practical the step-length α_k is determined by inexact line search. Some inexact line search techniques have been provided in [23]. The standard Wolfe conditions are [26]

$$f(x_k + \alpha_k d_k) - f(x_k) \leq c_1 \alpha_k g_k^T d_k, \tag{1.8}$$

$$g_{k+1}^T d_k \geq c_2 g_k^T d_k, \tag{1.9}$$

where $0 < c_1 < c_2 < 1$. To convergence analysis and numerical implementations of the CG methods, the step-length α_k is often obtained from the strong Wolfe line search [27] by

$$f(x_k + \alpha_k d_k) - f(x_k) \leq c_1 \alpha_k g_k^T d_k, \tag{1.10}$$

$$\left|g_{k+1}^T d_k\right| \leq -c_2 g_k^T d_k. \tag{1.11}$$

Furthermore, the generalized Wolfe conditions for $0 < c_1 < c_3 < 1$ and $c_4 \geq 0$ are as follows:

$$f(x_k + \alpha_k d_k) - f(x_k) \leq c_1 \alpha_k g_k^T d_k, \tag{1.12}$$

$$c_3 g_k^T d_k \leq g_{k+1}^T d_k \leq -c_4 g_k^T d_k. \tag{1.13}$$

For the first time, the general three-term conjugate gradient (TTCG) methods were proposed by Beale [7] to solve the unconstrained optimization problems. In this approaches, the search direction d_k is

$$d_k = -g_k + \beta_k d_k + \gamma_k d_t, \tag{1.14}$$

where $\beta_k = \beta_k^{FR}, \beta_k^{HS}, \beta_k^{DY}$. Furthermore, d_t is a restart direction and

$$\gamma_k = \begin{cases} 0, & k = t+1, \\ \dfrac{g_k^T y_t}{d_t^T y_t}, & k > t+1. \end{cases}$$

However, TTCG methods are obtained to improve traditional conjugate gradient methods and different choices for three-term conjugate gradient parameters lead to different TTCG methods. Further efforts have been made to develop the TTCG methods with the sufficient descent property [2, 6, 28], the descent and conjugacy properties [4, 11] and the sufficient descent and conjugacy properties [13, 14]. A comparison between some TTCG methods is reported for solving unconstrained optimization problems, see [5].

In this paper, we introduce two three-term conjugate gradient methods based on CD algorithm. Also, the generated search directions satisfy the sufficient descent property, independent of line search. The global convergence of the new methods are proven for general functions under mild assumptions. Also, numerical experiments confirm that our methods are efficient to solve unconstrained optimization problems in compared to some conjugate gradient method.

The structure of this paper is as follows. In Section 2, we propose two generalize of CD algorithm which are TTCG methods. The sufficient descent property of generated directions and the global convergence of the proposed algorithms are established in Section 3. In Section 4, we provide some numerical experiments to demonstrate the efficiency of our methods. Finally, some conclusions are given in Section 5.

2 Motivation and the new algorithms

In this section, we introduce two three-term conjugate gradient algorithms to solve unconstrained optimization problem (1.1) based on CD method. Fletcher in [15] proposed the CD conjugate gradient method which is closely related to the FR method. Note that to obtain the step-length α_k, we should solve the following one-dimensional optimization problem

$$\alpha_k = \underset{\alpha>0}{\operatorname{argmin}} f(x_k + \alpha d_k). \tag{2.1}$$

The CD conjugate gradient method is equal to FR conjugate gradient method when the exact line search is uesd. The exact line search implies $g_{k+1}^T d_k = 0$. Therefore, from (1.3), we get

$$g_{k-1}^T d_{k-1} = g_{k-1}^T\big(-g_{k-1} + \beta_{k-1} d_{k-2}\big) = -\|g_{k-1}\|^2 + \beta_{k-1} g_{k-1}^T d_{k-2} = -\|g_{k-1}\|^2.$$

Hence

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2} = -\frac{\|g_k\|^2}{g_{k-1}^T d_{k-1}} = \beta_k^{CD}.$$

On the other hand, the generated directions by CD method satisfy the sufficient descent condition with strong Wolfe line search [18]. Also, from the generalized Wolfe condition with $c_3 < 1$ and $c_4 = 0$, we obtain $0 \le \beta_k^{CD} \le \beta_k^{FR}$. Hence, the global convergence of CD method will be obtained by Theorem 2.2 in [1]. Now, we generalize the CD method to obtain a new three-term conjugate gradient method (NTTCD) where the direction d_k is calculated by

$$d_k = \begin{cases} -g_k, & k = 0, \\ -g_k + \beta_k^{CD} d_{k-1} + \theta_k g_k, & k \ge 1, \end{cases} \tag{2.2}$$

where the parameter θ_k is to grantee the sufficient descent condition and defined by

$$\theta_k = \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}}. \tag{2.3}$$

We will show that the search direction (2.2) satisfy $g_k^T d_k = -\|g_k\|^2$, independent of the line search and the objective function convexity. Furthermore, using the exact line search NTTCD method is reduced to CD method . To augment the efficiency of NTTCD method, we consider the following modification of this method. Hence, we get MNTTCD method while the search direction is generated by

$$d_k = \begin{cases} -g_k, & k = 0, \\ -g_k + \beta_k^{CD} d_{k-1} + t_k \theta_k g_k, & k \geq 1, \end{cases} \tag{2.4}$$

in which

$$t_k = \begin{cases} \max\left\{1, \min\left\{\eta_1, \dfrac{g_k^T d_{k-1}}{\max\{\zeta_1, \|y_{k-1}\|\|d_{k-1}\|\}}\right\}\right\}, & g_k^T d_{k-1} > 0, \\ \max\left\{\eta_2, \dfrac{g_k^T d_{k-1}}{\max\{\zeta_2, \|y_{k-1}\|\|d_{k-1}\|\}}\right\}, & g_k^T d_{k-1} \leq 0, \end{cases} \tag{2.5}$$

where $\eta_2 < 0 < \eta_1$ and $\zeta_1, \zeta_2 > 0$ are constant and are selected to increase the numerical efficiency and guarantee global convergence of the new algorithm. Note that for $t_k = 0$ and $t_k = 1$ the MNTTCD method reduces to CD and NTTCD methods, respectively. Now, we present the structure of new three-term conjugate gradient algorithms as follows:

Algorithm 1: The new three-term conjugate gradient method (NTTCD)

Step 0: Choose positive constant ϵ, $0 < c_1 < c_2 < 1$ and an initial point $x_0 \in \mathbb{R}^n$. Set $k = 0$, $d_0 = -g_0$.
Step 1: Terminate the algorithm once $\|g_k\| \leq \epsilon$ holds.
Step 2: Find the step-length α_k satisfying the strong Wolfe condition (1.10)-(1.11).
Step 3: Generate the new iterate by $x_{k+1} = x_k + \alpha_k d_k$.
Step 4: Calculate g_{k+1} and the conjugate parameter β_{k+1}^{CD} by (1.4).
Step 5: Obtain the parameter θ_{k+1} with (2.3) and the new search direction d_{k+1} by (2.2).
Step 6: Set $k = k + 1$ and go to *Step 1.*

Algorithm 2: The modification of new three-term conjugate gradient method (MNTTCD)

Step 0: Choose positive constants ϵ, ζ_1, ζ_2, $\eta_2 < 0 < \eta_1$, $0 < c_1 < c_2 < 1$ and an initial point $x_0 \in \mathbb{R}^n$. Set $k = 0$, $d_0 = -g_0$.
Step 1: Terminate the algorithm once $\|g_k\| \leq \epsilon$ holds.
Step 2: Find the step-length α_k satisfying the strong Wolfe condition (1.10)-(1.11).
Step 3: Generate the new iterate by $x_{k+1} = x_k + \alpha_k d_k$.
Step 4: Calculate g_{k+1} and the conjugate parameter β_{k+1}^{CD} by (1.4).
Step 5: Obtain the parameters θ_{k+1} with (2.3), t_k by (2.5) and the new direction d_{k+1} by (2.4).
Step 6: Set $k = k + 1$ and go to *Step 1.*

---

3 Convergence analysis

In this section, the sufficient descent property and the global convergence of the new algorithms are established. To this aim, we make some assumptions on the objective function as follows:

Assumption 3.1 The level set $L(x_0) = \{x \in \mathbb{R}^n | f(x) \leq f(x_0)\}$ is bounded, i.e., there exists a constant $M > 0$ such that

$$\|x\| \leq M, \qquad \forall x \in L(x_0). \tag{3.1}$$

Assumption 3.2 In some neighborhood $\Omega \subseteq L(x_0)$, the gradient of the objective function f is Lipschitz continuous, i.e., there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \qquad \forall x, y \in \Omega. \tag{3.2}$$

Lemma 3.1. *Suppose that $\{d_k\}_{k\geq 0}$ is generated by* NTTCD *algorithm. Then, we have*

$$g_k^T d_k = -\|g_k\|^2. \tag{3.3}$$

Proof. By multiplying (2.2) in g_k^T, using (1.4) and (2.3), we obtain

$$\begin{aligned} g_k^T d_k &= -\|g_k\|^2 + \beta_k^{CD} g_k^T d_{k-1} + \theta_k \|g_k\|^2 \\ &= -\|g_k\|^2 - \frac{\|g_k\|^2}{g_{k-1}^T d_{k-1}} g_k^T d_{k-1} + \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}} \|g_k\|^2 \\ &= -\|g_k\|^2 < 0. \end{aligned}$$

Therefore, the proof is complete. □

Lemma 3.2. *Let $\{d_k\}_{k\geq 0}$ be generated direction by* MNTTCD *algorithm. Then, $\{d_k\}_{k\geq 0}$ satisfy the sufficient descent condition* (1.7) *with $c = 1$, i.e.*

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

Proof. We prove this lemma in two following cases.

Case(1): Let $g_k^T d_{k-1} > 0$. From (1.4), (2.3) and (2.4), we get

$$g_k^T d_k = -\|g_k\|^2 - \frac{\|g_k\|^2}{g_{k-1}^T d_{k-1}} g_k^T d_{k-1} + t_k \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}} \|g_k\|^2. \tag{3.5}$$

Using (2.5), there are two choices for parameter t_k.

(i) For $t_k = 1$, we have

$$\begin{aligned} g_k^T d_k &= -\|g_k\|^2 - \frac{\|g_k\|^2}{g_{k-1}^T d_{k-1}} g_k^T d_{k-1} + \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}} \|g_k\|^2 \\ &= -\|g_k\|^2 < 0. \end{aligned}$$

(ii) If $t_k = \min\left\{\eta_1, \dfrac{g_k^T d_{k-1}}{\max\{\zeta_1, \|y_{k-1}\|\|d_{k-1}\|\}}\right\} > 1$, then we use induction over k to prove this item. Now, induction hypothesis implies $g_{k-1}^T d_{k-1} \leq -\|g_{k-1}\|^2 < 0$. Therefore, we have

$$\frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}} \|g_k\|^2 < 0.$$

Hence

$$t_k \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}} \|g_k\|^2 < \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}} \|g_k\|^2. \tag{3.6}$$

So, (3.5) and (3.6) give us

$$\begin{aligned} g_k^T d_k &\leq -\|g_k\|^2 - \frac{\|g_k\|^2}{g_{k-1}^T d_{k-1}} g_k^T d_{k-1} + \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}} \|g_k\|^2 \\ &= -\|g_k\|^2 < 0. \end{aligned}$$

Therefore, for this case d_k satisfy the sufficient descent condition.

CASE(2): If $g_k^T d_{k-1} \leq 0$, then

$$t_k = \max\left\{\eta_2, \frac{g_k^T d_{k-1}}{\max\{\zeta_2, \|y_{k-1}\|\|d_{k-1}\|\}}\right\} \leq 0.$$

Similar to CASE (1), using induction over k, we have $g_{k-1}^T d_{k-1} \leq -\|g_{k-1}\|^2 < 0$. Hence

$$\frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}} \|g_k\|^2 \geq 0, \tag{3.7}$$

yielding

$$t_k \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}} \|g_k\|^2 \leq 0. \tag{3.8}$$

Finally, from (3.5), (3.7) and (3.8), we obtain

$$g_k^T d_k \leq -\|g_k\|^2 < 0.$$

So, we obtain desired result. □

Lemma 3.3. *Let $\{d_k\}_{k\geq 0}$ be a sufficient descent direction and the step-length α_k satisfies the strong Wolfe line search (1.10)-(1.11). Then, based on Assumptions 3.1 and 3.2, we have*

$$\sum_{k=0}^{+\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty. \tag{3.9}$$

Proof. Since d_k is the sufficient descent direction, the proof is similar to [30]. □

Lemma 3.4. *Under strong Wolfe line search (1.10)-(1.11), the parameter θ_k satisfies*

$$-1 \leq \theta_k \leq 1. \tag{3.10}$$

Proof. From (1.11), it is clear that

$$c_2 g_{k-1}^T d_{k-1} \leq g_k^T d_{k-1} \leq -c_2 g_{k-1}^T d_{k-1}. \tag{3.11}$$

Since $g_{k-1}^T d_{k-1} \leq -\|g_{k-1}\|^2 < 0$, we get

$$\theta_k = \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}} \leq \frac{c_2 g_{k-1}^T d_{k-1}}{g_{k-1}^T d_{k-1}} = c_2 < 1,$$

and

$$\theta_k = \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}} \geq -\frac{c_2 g_{k-1}^T d_{k-1}}{g_{k-1}^T d_{k-1}} = -c_2 > -1.$$

Hence

$$-1 \leq \theta_k \leq 1.$$

□

Theorem 3.1. *Let $\{d_k\}_{k\geq 0}$ be a sufficient descent direction and $\{x_k\}_{k\geq 0}$ be the generated sequence by* `NTTCD` *algorithm. Moreover, suppose that the Assumptions 3.1 and 3.2 hold. Then*

$$\lim_{k\to\infty} \inf \|g_k\| = 0. \tag{3.12}$$

Proof. By contradiction there exists $\epsilon_1 > 0$ such that $\|g_k\| > \epsilon_1$ for any k. So

$$\frac{1}{\|g_k\|^2} \leq \frac{1}{\epsilon_1^2}. \tag{3.13}$$

From (2.2), we get

$$d_k = (\theta_k - 1)g_k + \beta_k^{CD} d_{k-1}.$$

Now, (1.4), (2.3) and (3.3) imply

$$\begin{aligned}
\|d_k\|^2 &= (\theta_k - 1)^2\|g_k\|^2 + (\beta_k^{CD})^2\|d_{k-1}\|^2 + 2(\theta_k - 1)\beta_k^{CD} g_k^T d_{k-1} \\
&= (\theta_k - 1)^2\|g_k\|^2 + \frac{\|g_k\|^4}{(g_{k-1}^T d_{k-1})^2}\|d_{k-1}\|^2 - 2(\theta_k - 1)\frac{\|g_k\|^2}{g_{k-1}^T d_{k-1}} g_k^T d_{k-1} \\
&= (\theta_k - 1)^2\|g_k\|^2 + \frac{\|g_k\|^4}{\|g_{k-1}\|^4}\|d_{k-1}\|^2 - 2\theta_k \frac{\|g_k\|^2}{g_{k-1}^T d_{k-1}} g_k^T d_{k-1} + 2\frac{\|g_k\|^2}{g_{k-1}^T d_{k-1}} g_k^T d_{k-1} \\
&= (\theta_k - 1)^2\|g_k\|^2 + \frac{\|g_k\|^4}{\|g_{k-1}\|^4}\|d_{k-1}\|^2 - 2\frac{\|g_k\|^2}{(g_{k-1}^T d_{k-1})^2}(g_k^T d_{k-1})^2 + 2\frac{\|g_k\|^2}{g_{k-1}^T d_{k-1}} g_k^T d_{k-1} \\
&\leq (\theta_k - 1)^2\|g_k\|^2 + \frac{\|g_k\|^4}{\|g_{k-1}\|^4}\|d_{k-1}\|^2 + 2\frac{\|g_k\|^2}{g_{k-1}^T d_{k-1}} g_k^T d_{k-1}.
\end{aligned}$$

The above inequality along with (3.11) result

$$\begin{aligned}
\|d_k\|^2 &\leq (\theta_k - 1)^2\|g_k\|^2 + \frac{\|g_k\|^4}{\|g_{k-1}\|^4}\|d_{k-1}\|^2 + 2c_2\frac{\|g_k\|^2}{g_{k-1}^T d_{k-1}} g_{k-1}^T d_{k-1} \\
&= (\theta_k - 1)^2\|g_k\|^2 + \frac{\|g_k\|^4}{\|g_{k-1}\|^4}\|d_{k-1}\|^2 + 2c_2\|g_k\|^2.
\end{aligned} \tag{3.14}$$

By dividing both sides of this inequality in $(g_k^T d_k)^2$ and using (3.3), we have

$$\begin{aligned}
\frac{\|d_k\|^2}{(g_k^T d_k)^2} &\leq \frac{(\theta_k - 1)^2\|g_k\|^2}{(g_k^T d_k)^2} + \frac{\|g_k\|^4\|d_{k-1}\|^2}{\|g_{k-1}\|^4(g_k^T d_k)^2} + \frac{2c_2\|g_k\|^2}{(g_k^T d_k)^2} \\
&= \frac{(\theta_k - 1)^2\|g_k\|^2}{\|g_k\|^4} + \frac{\|g_k\|^4\|d_{k-1}\|^2}{\|g_{k-1}\|^4\|g_k\|^4} + \frac{2c_2\|g_k\|^2}{\|g_k\|^4} \\
&= \frac{(\theta_k - 1)^2}{\|g_k\|^2} + \frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^4} + \frac{2c_2}{\|g_k\|^2}.
\end{aligned}$$

By Lemma 3.4, $-2 \leq \theta_k - 1 \leq 0$ and $0 \leq (\theta_k - 1)^2 \leq 4$. Hence

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} \leq \frac{\|d_{k-1}\|^2}{(g_{k-1}^T d_{k-1})^2} + \frac{\omega_1}{\|g_k\|^2}, \tag{3.15}$$

in which $\omega_1 := 2(c_2 + 2)$. By applying (3.13) and (3.15), we can result

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} \leq \frac{\|d_{k-1}\|^2}{(g_{k-1}^T d_{k-1})^2} + \frac{\omega_1}{\|g_k\|^2} \leq \frac{\|d_{k-2}\|^2}{(g_{k-2}^T d_{k-2})^2} + \frac{\omega_1}{\|g_{k-1}\|^2} + \frac{\omega_1}{\|g_k\|^2} \leq \cdots \leq \sum_{i=0}^{k} \frac{\omega_1}{\|g_i\|^2} \leq \frac{k\omega_1}{\epsilon_1^2}.$$

Therefore

$$\frac{(g_k^T d_k)^2}{\|d_k\|^2} \geq \frac{\epsilon_1^2}{\omega_1} \frac{1}{k}.$$

Finally

$$\sum_{k=0}^{+\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} \geq \frac{\epsilon_1^2}{\omega_1} \sum_{k=0}^{+\infty} \frac{1}{k} = +\infty,$$

which contradicts with Lemma 3.3. □

Now, we investigate the convergence of `MNTTCD` algorithm in three cases. For $t_k = 1$, this method reduces to `NTTCD` algorithm which its convergence established in Theorem 1. Therefore, we prove other cases in the following theorem.

Theorem 3.2. *Let $\{d_k\}_{k\geq 0}$ be a sufficient descent direction and $\{x_k\}_{k\geq 0}$ be the generated sequence by* `MNTTCD` *algorithm. Then*

$$\lim_{k\to\infty} \inf \|g_k\| = 0. \tag{3.16}$$

Proof. We use contradiction to proof this theorem. Hence, there exists a constant $\epsilon_2 > 0$ such that $\|g_k\| > \epsilon_2$ for any k and

$$\frac{1}{\|g_k\|^2} \leq \frac{1}{\epsilon_2^2}. \tag{3.17}$$

Now, (2.4) implies

$$d_k = (t_k\theta_k - 1)g_k + \beta_k^{CD} d_{k-1}.$$

By substituting (1.4) and (2.3) in above equality, we get

$$\begin{aligned}
\|d_k\|^2 &= (t_k\theta_k - 1)^2\|g_k\|^2 + (\beta_k^{CD})^2\|d_{k-1}\|^2 + 2(t_k\theta_k - 1)\beta_k^{CD} g_k^T d_{k-1} \\
&= (t_k\theta_k - 1)^2\|g_k\|^2 + \frac{\|g_k\|^4}{(g_{k-1}^T d_{k-1})^2}\|d_{k-1}\|^2 - 2(t_k\theta_k - 1)\frac{\|g_k\|^2}{g_{k-1}^T d_{k-1}} g_k^T d_{k-1} \\
&= (t_k\theta_k - 1)^2\|g_k\|^2 + \frac{\|g_k\|^4}{(g_{k-1}^T d_{k-1})^2}\|d_{k-1}\|^2 - 2t_k\theta_k \frac{\|g_k\|^2}{g_{k-1}^T d_{k-1}} g_k^T d_{k-1} + 2\frac{\|g_k\|^2}{g_{k-1}^T d_{k-1}} g_k^T d_{k-1}.
\end{aligned} \tag{3.18}$$

We consider two following cases:

Case (i): If $g_k^T d_{k-1} > 0$, then $1 < t_k \leq \eta_1$. Also, Lemma 3.3 implies

$$\frac{1}{(g_k^T d_k)^2} \leq \frac{1}{\|g_k\|^4}. \tag{3.19}$$

Now, Lemma 3.4 along with (2.3) give us $-1 \leq \theta_k < 0$. From (3.18), we have

$$\|d_k\|^2 \leq (t_k\theta_k - 1)^2\|g_k\|^2 + \frac{\|g_k\|^4}{(g_{k-1}^T d_{k-1})^2}\|d_{k-1}\|^2.$$

We divide both sides of this inequality in $(g_k^T d_k)^2$ and use (3.19). Hence

$$\begin{aligned}\frac{\|d_k\|^2}{(g_k^T d_k)^2} &\leq \frac{(t_k\theta_k - 1)^2\|g_k\|^2}{(g_k^T d_k)^2} + \frac{\|g_k\|^4}{(g_{k-1}^T d_{k-1})^2(g_k^T d_k)^2}\|d_{k-1}\|^2 \\ &\leq \frac{(t_k\theta_k - 1)^2\|g_k\|^2}{\|g_k\|^4} + \frac{\|g_k\|^4}{(g_{k-1}^T d_{k-1})^2\|g_k\|^4}\|d_{k-1}\|^2 \\ &= \frac{(t_k\theta_k - 1)^2}{\|g_k\|^2} + \frac{\|d_{k-1}\|^2}{(g_{k-1}^T d_{k-1})^2}. \end{aligned} \tag{3.20}$$

Since $1 < t_k \leq \eta_1$, we have

$$\begin{aligned} 1 < t_k \leq \eta_1 &\Longrightarrow \theta_k\eta_1 \leq t_k\theta_k < \theta_k < 0 \\ &\Longrightarrow \theta_k\eta_1 - 1 \leq t_k\theta_k - 1 < -1 \\ &\Longrightarrow -\eta_1 - 1 \leq t_k\theta_k - 1 < -1 \\ &\Longrightarrow (t_k\theta_k - 1)^2 \leq (\eta_1 + 1)^2 := \omega_2. \end{aligned}$$

This inequality and (3.20) result

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} \leq \frac{\|d_{k-1}\|^2}{(g_{k-1}^T d_{k-1})^2} + \frac{\omega_2}{\|g_k\|^2}.$$

Case (ii): If $g_k^T d_{k-1} \leq 0$, then $\eta_2 \leq t_k \leq 0$. Also, Lemma 3.4 give us

$$0 \leq \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}} \leq 1.$$

Now, from (3.18), we have

$$\|d_k\|^2 \leq (t_k\theta_k - 1)^2\|g_k\|^2 + \frac{\|g_k\|^4}{(g_{k-1}^T d_{k-1})^2}\|d_{k-1}\|^2 - 2t_k\theta_k^2\|g_k\|^2 + 2\|g_k\|^2.$$

By dividing both sides of this inequality in $(g_k^T d_k)^2$ and using (3.19)

$$\begin{aligned}\frac{\|d_k\|^2}{(g_k^T d_k)^2} &\leq \frac{(t_k\theta_k - 1)^2\|g_k\|^2}{(g_k^T d_k)^2} + \frac{\|g_k\|^4}{(g_{k-1}^T d_{k-1})^2(g_k^T d_k)^2}\|d_{k-1}\|^2 + \frac{2(1-t_k)}{(g_k^T d_k)^2}\|g_k\|^2 \\ &\leq \frac{(t_k\theta_k - 1)^2 + 2(1-t_k)}{\|g_k\|^2} + \frac{\|d_{k-1}\|^2}{(g_{k-1}^T d_{k-1})^2}. \end{aligned} \tag{3.21}$$

Since $0 \leq \theta_k \leq 1$, we get

$$(t_k\theta_k - 1)^2 + 2(1-t_k) = t_k^2\theta_k^2 - 2t_k\theta_k - 2t_k + 3 \leq t_k^2 - 4t_k + 3 = (t_k - 2)^2 - 1,$$

and

$$\eta_2 \leq t_k \leq 0 \Longrightarrow \eta_2 - 2 \leq t_k - 2 \leq -2$$

$$\Longrightarrow (t_k-2)^2 \le (\eta_2-2)^2$$
$$\Longrightarrow (t_k-2)^2-1 \le (\eta_2-2)^2-1 := \omega_3.$$

By subsuiting this inequality to (3.21), we obtain

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} \le \frac{\|d_{k-1}\|^2}{(g_{k-1}^T d_{k-1})^2} + \frac{\omega_3}{\|g_k\|^2}.$$

Hence, in both cases similar to Theorem 3.1, we have

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} \le \frac{\|d_{k-1}\|^2}{(g_{k-1}^T d_{k-1})^2} + \frac{\omega_j}{\|g_k\|^2} \le \frac{\|d_{k-2}\|^2}{(g_{k-2}^T d_{k-2})^2} + \frac{\omega_j}{\|g_{k-1}\|^2} + \frac{\omega_j}{\|g_k\|^2}$$
$$\le \cdots \le \sum_{i=0}^{k} \frac{\omega_j}{\|g_i\|^2} \le \frac{k\omega_j}{\epsilon_2^2}, \qquad j=2,3.$$

Hence

$$\frac{(g_k^T d_k)^2}{\|d_k\|^2} \ge \frac{\epsilon_2^2}{\omega_j}\frac{1}{k} \qquad j=2,3.$$

Finally

$$\sum_{k=0}^{+\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} \ge \frac{\epsilon_2^2}{\omega_j}\sum_{k=0}^{+\infty}\frac{1}{k} = +\infty, \qquad j=2,3.$$

Therefore, by this contradicts, the proof is complete. □

4 Numerical experiments

In this section, we express numerical results on a set of some nonlinear unconstrained optimization test functions on the `CUTEst` collection [8] which are given in Table 1. The dimensions of test functions are from 2 to 12005 while the initial points are standard ones proposed in `CUTEst`. We apply the following algorithms to solve these test functions:

- `FR`: Fletcher-Reeves conjugate gradient method [16],
- `HS`: Hestenes-Stiefel conjugate gradient method [19],
- `DY`: Dai-Yuan conjugate gradient method [10],
- `CD`: Conjugate Descent conjugate gradient method [15],
- `NTTCD`: New three-term conjugate gradient method,
- `MNTTCD`: Modification of the new three-term conjugate gradient method.

All algorithms are implemented in Matlab 2011 programming environment on a 2.3Hz Intel core i3 processor laptop and 4GB of RAM with the double precision data type in Linux operations system. The iterations stop whenever the inequality

$$\|g_k\| \le 10^{-6},$$

be satisfied or the total number of iterates exceeds 10000. Furthermore, we choose the parameters $\zeta_1 = 100$, $\zeta_2 = 50$, $\eta_1 = 15$, $\eta_2 = -10$, $c_1 = 10^{-3}$ and $c_2 = 0.95$.

Here, we use the performance profiles of Dolan and Moré [12] to compare the performance of the algorithms on the test functions. We consider P as designates the percentage of problems which are solved within a factor τ of the best solver. The horizontal axis of the figure gives the percentage of the test functions for which a method is the fastest (efficiency), while the vertical axis gives the percentage of the test functions that were successfully solved by each method (robustness).

Figures 1-3 show the performance of all algorithms to solve the unconstrained optimization problems. In these figures, $P(\tau)$ is designates the percentage of problems which are solved within a factor τ of the best solver. Figure 1 shows that the `MNTTCD` method wins about 32% of test problems with the smallest number of iterations. We conclude from Figure 2 that the `NTTCD` method is the most effective for most test functions in total number of function evaluations about 39%. From figure 3, we can see that `NTTCD` method is better than other methods about 26% of the most wins in terms of CPU times.

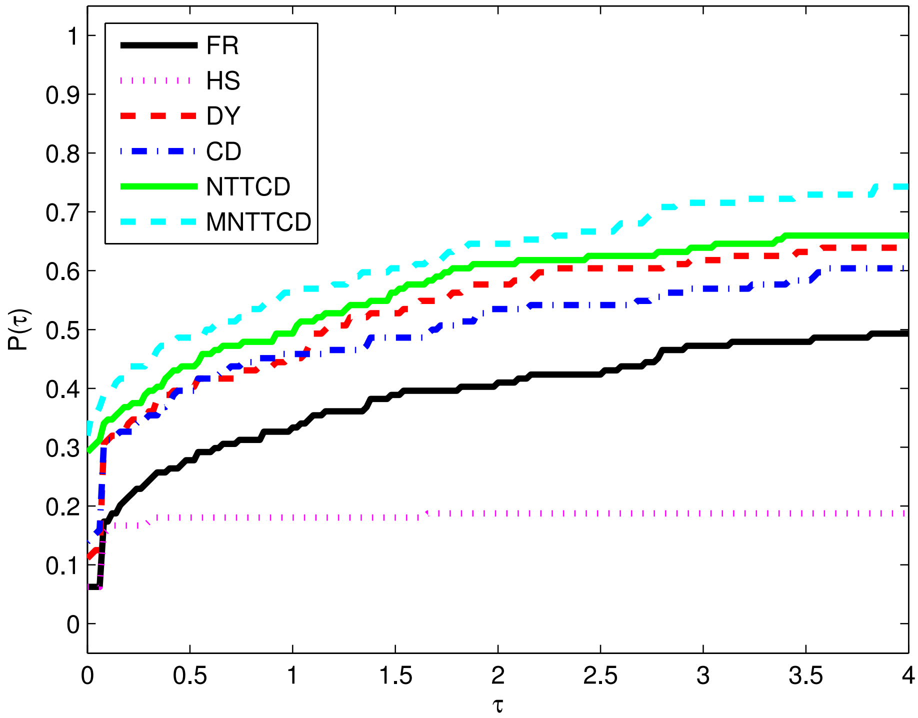

Figure 1: The Dolan-Moré performance profiles for the total number of iterations.

5 Conclusion

In this work, we propose two three-term conjugate gradient directions based on `CD` conjugate gradient method. It is shown that the proposed directions always fulfills the sufficient descent property, independent of the line search. Under standard assumptions, we prove the conver-

Table 1: Test functions taken from CUTEst collection

No.	Test function	Dim	No.	Test function	Dim	No.	Test function	Dim
1	3PK	30	49	DQDRTIC	10000	97	NONDIA	5000
2	AIRCRFTB	8	50	DQRTIC	5000	98	NONDQUAR	5000
3	ALLINIT	4	51	EDENSCH	100	99	OSCIPANE	5000
4	ALLINITU	4	52	EG2	1000	100	OSCIPATH	10
5	ARGLINA	500	53	EG3	10000	101	OSLBQP	8
6	ARGLINB	200	54	EIGENA	2000	102	PALMER1C	8
7	ARWHEAD	5000	55	ENGVAL1	100	103	PALMER1D	7
8	BARD	3	56	ENGVAL2	3	104	PALMER2C	8
9	BDQRTIC	100	57	ERRINROS	50	105	PALMER3C	8
10	BEALE	2	58	EXPFIT	2	106	PALMER4C	8
11	BIGGS6	6	59	EXTROSNB	1000	107	PALMER5C	6
12	BIGGSB1	5000	60	FLETCBV2	10000	108	PALMER6C	9
13	BOX2	3	61	FLETCHCR	500	109	PALMER7C	8
14	BOX3	3	62	FMINSRF2	5625	110	PALMER8A	6
15	BRKMCC	2	63	FMINSURF	5625	111	PALMER8C	8
16	BROWNDEN	4	64	FREUROTH	2	112	PENALTY1	100
17	BROYDN3D	5000	65	GENHUMPS	5000	113	PENALTY2	50
18	BROYDN7D	500	66	GENROSE	500	114	POWELLBC	1000
19	BROYDNBD	5000	67	GROWTHLS	3	115	POWELLSG	5000
20	BRYBND	500	68	GULF	3	116	QR3DLS	610
21	CHAINWOO	1000	69	HAIRY	2	117	QUARTC	25
22	CHNROSNB	50	70	HATFLDD	3	118	ROSENBR	2
23	CLIFF	2	71	HATFLDF	3	119	S308	2
24	COSINE	1000	72	HATFLDFL	3	120	SCHMVETT	100
25	CRAGGLVY	1000	73	HEART6LS	6	121	SENSORS	100
26	CUBE	2	74	HEART8LS	3	122	SINEVAL	2
27	CUBENE	2	75	HELIX	3	123	SINVALNE	2
28	DALLASM	196	76	HILBERTA	10	124	SISSER	2
29	DALLASS	46	77	HILBERTB	10	125	SNAIL	2
30	DECONVU	63	78	HIMMELBA	2	126	SPARSINE	1000
31	DENSCHNA	2	79	HIMMELBC	2	127	SPARSQUR	10000
32	DENSCHNB	2	80	HIMMELBF	4	128	SPMSRTLS	4999
33	DENSCHNC	2	81	HIMMELBG	2	129	SROSENBR	1000
34	DENSCHNF	2	82	HIMMELBH	2	130	TAME	2
35	DIXMAANA	9000	83	HUMPS	2	131	TESTQUAD	100
36	DIXMAANB	3000	84	JENSMP	2	132	TOINTGOR	50
37	DIXMAANC	3000	85	KOWOSB	4	133	TOINTGSS	10000
38	DIXMAAND	3000	86	LIARWHD	5000	134	TOINTPSP	50
39	DIXMAANE	3000	87	LOGHAIRY	2	135	TOINTQOR	50
40	DIXMAANF	3000	88	MANCINO	100	136	TQUARTIC	500
41	DIXMAANG	3000	89	MATRIX2	6	137	TRIDIA	5000
42	DIXMAANH	3000	90	METHANOL	12005	138	VAREIGVL	500
43	DIXMAANI	3000	91	MODBEALE	2	139	VIBRBEAM	8
44	DIXMAANJ	3000	92	MOREBV	5000	140	WATSON	12
45	DIXMAANK	3000	93	MSQRTALS	1024	141	WEEDS	3
46	DIXMAANL	3000	94	MSQRTBLS	1024	142	WOODS	100
47	DIXON3DQ	1000	95	MINE5D	10733	143	YFITU	3
48	DJTL	2	96	NONCVXU2	1000	144	ZANGWIL2	2

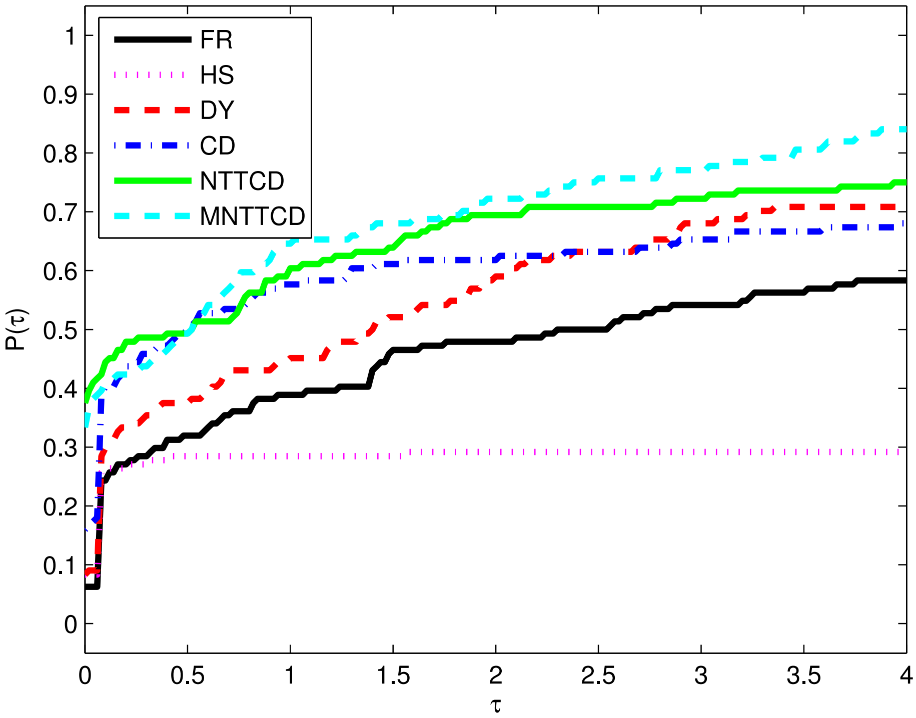

Figure 2: The Dolan-Moré performance profiles for the total number of function evaluations.

gence properties of the new schemes. The preliminary numerical experiment on a set of the test functions collection indicates that the new algorithms are effective.

References

[1] M. Al-Baali, Descent property and global convergence of the Fletcher-Reeves method with inexact line search, IMA Journal of Numerical Analysis. **5**, 121–124 (1985).

[2] M. Al-Baali, Y. Narushima and H. Yabe, A family of three-term conjugate gradient methods with sufficient descent property for unconstrained optimization, Computational Optimization and Applications. **60**, 89–110 (2015).

[3] Z. Aminifard and S. Babaie-Kafaki, An optimal parameter choice for the Dai-Liao family of conjugate gradientmethods by avoiding a direction of the maximum magnification by the search direction matrix, 4OR. **17**, 317–330 (2019).

[4] N. Andrei, A new three-term conjugate gradient algorithm for unconstrained optimization, Numerical Algorithm. **68**, 305–321 (2015).

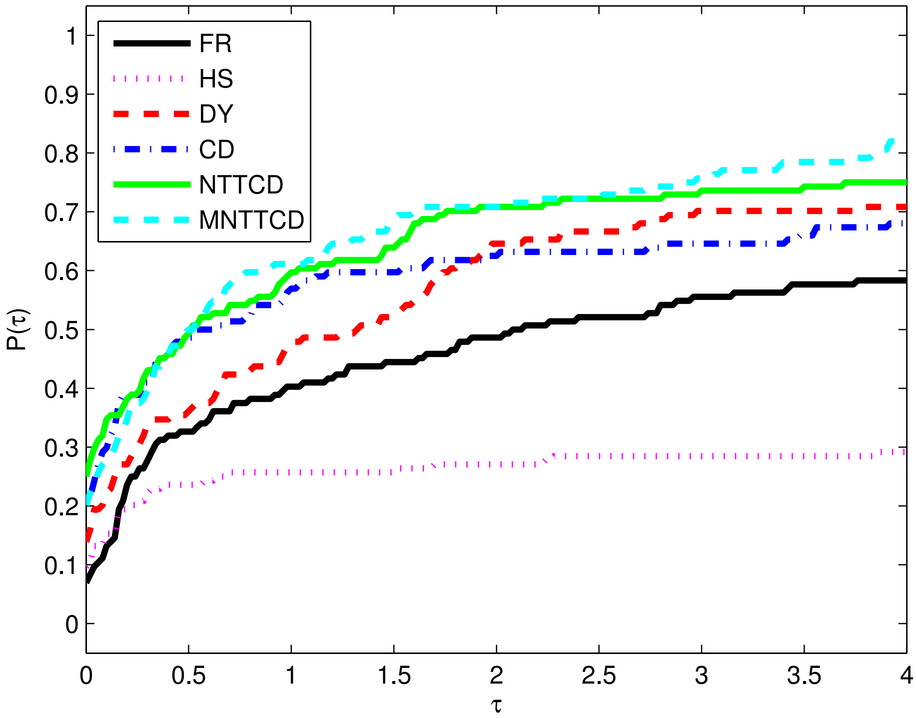

Figure 3: The Dolan-Moré performance profiles for the CPU times.

[5] L. Arman, Y. Xu, M. Rostami and F. Rahpeymaii, Some three-term conjugate gradient methods for solving unconstrained optimization problems, Pacific Journal of Optimization. (2020).

[6] S. Babaie–Kafaki, A modified three-term conjugate gradient method with sufficient descent property, Applied Mathematics-A Journal of Chinese Universities. **30(3)**, 263–272 (2015).

[7] E. M. L. Beale, A derivative of conjugate gradients In: Lootsma, F.A (ed.) Numerical Methods for Nonlinear Optimization, Academic, London. 39–43 (1972).

[8] A. R. Conn, N. I .M. Gould and Ph. L. Toint, CUTE: constrained and unconstrained testing environment, ACM Transactions on Mathematical Software. **21**, 123–160 (1995).

[9] Y. H. Dai, J. Y. Han, G. H. Liu, D. F. Sun, H. X. Yin and Y. Yuan, Convergence properties of nonlinear conjugate gradient methods, SIAM Journal on Optimization. **10**, 345–358 (1999).

[10] Y. H. Dai and Y. Yuan, A nonlinear conjugate gradient method with a strong global convergence property, SIAM Journal on Optimization. **10**, 177–182 (1999).

[11] S. Deng and Z. Wan, A three-term conjugate gradient algorithm for large-scale unconstrained optimization problems, Applied Numerical Mathematics. **92**, 70–81 (2015).

[12] E. D. Dolan and J. J. Moré, Benchmarking optimization software with performance profiles, Mathematical Programming. **91**, 201–213 (2002).

[13] X. L. Dong, D. R. Han, R. Ghanbari, X. L. Li and Z. F. Dai, Some new three-term HestenesStiefel conjugate gradient methods with affine combination, Optimization. **66(5)**, 759–776 (2017).

[14] X. L. Dong, H. W. Liu and Y. B. He, New version of the three-term conjugate gradient method based on spectral scaling conjugacy condition that generates descent search direction, Applied Mathematics and Computation. **269**, 606–617 (2015).

[15] R. Fletcher, Practical Methods of Optimization vol. 1: Unconstrained Optimization, John Wiley & Sons, New York, 1987.

[16] R. Fletcher and C. Reeves, Function minimization by conjugate gradients, Computer Journal. **7(2)**, 149–154 (1964).

[17] W. W. Hager and H. Zhang, A new conjugate gradient method with guaranteed descent and an efficient line search, SIAM Journal on Optimization. **16**, 170–192 (2005).

[18] W. W. Hager and H. Zhang, A survey of nonlinear conjugate gradient methods, Pacific journal of Optimization. **2(1)**, 35–58 (2006).

[19] M. R. Hestenes and E. L. Stiefel, Methods of conjugate gradients for solving linear systems, Journal of research of the National Bureau of Standards. **49**, 409–436 (1952).

[20] M. Kimiaei and M. Rostami, Impulse noise removal based on new hybrid conjugate gradient approach, Kybernetika. **52(5)**, 791–823 (2016).

[21] A. B. Muhammad, C. Tammer, A. M. Awwal and R. Elster, A Dai-Liao-like projection method for solving convex constrained nonlinear monotone equations and minimizing the l_1-regularized problem, Appl. Set-Valued Anal. Optim, **3**, 259–279, (2021).

[22] Y. Narushima, H. Yabe and J. A. Ford, A three-term conjugate gradient method with sufficient descent property for unconstrained optimization, SIAM Journal on Optimization. **21**, 212–230 (2011).

[23] J. Nocedal and S. J. Wright, Numerical Optimization, Springer series in Operation Research, Springer, New York, 1999.

[24] E. Polak and G. Ribière, Note sur la convergence de directions conjugées, Rev. Francaise Informat Recherche Opertionelle, 3e Année. **16**, 35–43 (1969).

[25] B. T. Polyak, The conjugate gradient method in extreme problems, USSR computational mathematics and mathematical physics. **9**, 94–112 (1969).

[26] P. Wolfe, Convergence conditions for ascent methods, SIAM Review. **11**, 226–235 (1968).

[27] P. Wolfe, Convergence conditions for ascent methods. II: some corrections, SIAM Review. **13(2)**, 185–188 (1971).

[28] G. Yuan and M. Zhang, A three-terms Polak–Ribière–Polyak conjugate gradient algorithm for large-scale nonlinear equations, Journal of Computational and Applied Mathematics. **286**, 186–195 (2015).

[29] X. Zhao, A projected subgradient method for nondifferentiable quasiconvex multiobjective optimization problems, J. Optim. Theory Appl, **190**, 82107, (2021).

[30] G. Zoutendijk, Nonlinear programming, computational methods. In: Abadie J (ed) Integer and nonlinear programming. Amsterdam, North-holland. 37–86 (1970).

Ladan Arman School of Aeronautic Science and Engineering, Beihang University, Beijing, China

E-mail: ladan.arman@gmail.com

Yuanming Xu School of Aeronautic Science and Engineering, Beihang University, Beijing, China

E-mail: xuymg@sina.com

Mohammad Reza Bayat Ph.D student, School of Electronics and Information Engineering, Beihang University, Beijing, China

E-mail: bayat_mrz@yahoo.com

Long Liping Institute of Mechaincs, Chinese Academy of Sciences, Beijing 100190, China

E-mail: lplong05@163.com