Filomat 38:24 (2024), 8433–8448
https://doi.org/10.2298/FIL2424433B

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

A family of scaled conjugate gradient methods under a new modified weak-Wolfe-Powell line search for large-scale optimization

S. Bojari[a,*], M. R. Eslahchi[b]

[a]*Department of Basic Science, Hamedan University of Technology, Hamedan, Iran*
[b]*Department of Computer Science, Faculty of Mathematical Sciences, Tarbiat Modares University, P.O. Box 14115-134, Tehran, Iran*

Abstract. In this paper, a family of three-term conjugate gradient methods is proposed to solve a large-scale unconstrained optimization problem. With the help of suitable features of the new family (like sufficient descent directions) a strong global convergence theorem for uniformly convex functions under weak Wolfe-Powell line search technique is established. Furthermore, a new well-defined modification of weak Wolfe-Powell line search technique is presented and a strong global convergence theorem for general smooth functions is obtained. In two competitions contained two line search techniques, five well behaved conjugate gradient methods and 200 standard problems the efficiency of these new methods in numerical experience is indicated.

1. Introduction

In a large-scale unconstrained optimization problem

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

where $f(x) : \mathbb{R}^n \longrightarrow \mathbb{R}$ is smooth and bounded from below, finding the second-order information of f is a costly process. Therefore, using first-order (gradient-type) methods to solve problem (1) numerically is more efficient than using Newton or quasi-Newton methods. The steepest descent method is a simple gradient-type algorithm which use only the first-order information (the gradient) of f. But this method has a slow convergence rate and may generate zigzagging directions [32]. So, the conjugate gradient (CG) method is a natural choice for solving the large-scale optimization problem (1).

In a CG algorithm, in iteration x_{k+1}, first a descent direction d_{k+1} in the form of

$$d_0 = -g_0, \quad d_{k+1} = -g_{k+1} + \beta_{k+1} d_k + \gamma_{k+1} p_k, \quad \text{for } k = 0, 1, 2, 3, \ldots, \tag{2}$$

is calculated. Note that in (2), g_{k+1} is the gradient of f at x_{k+1}, β_{k+1} and γ_{k+1} are CG parameters and p_k is an arbitrary vector related to previous iterations. Then by an exact line search

$$\min_{\alpha\in\mathbb{R}>0} f(x_{k+1} + \alpha d_{k+1}), \tag{3}$$

2020 *Mathematics Subject Classification*. 90C30, 90C06, 90C26.
Keywords. Large-scale optimization problem, Nonconvex optimization, Conjugate gradient method, Global convergence, Weak Wolfe-Powell line search technique.
Received: 22 September 2020; Revised: 29 April 2021; Accepted: 09 June 2024
Communicated by Predrag Stanimirović
* Corresponding author: S. Bojari
Email addresses: sanaz.bojari@gmail.com (S. Bojari), eslahchi@modares.ac.ir (M. R. Eslahchi)

or an inexact line search technique like the weak-Wolf-Powell (WWP) technique

$$\begin{cases} f(x_{k+1}+\alpha d_{k+1}) \le f_{k+1} + \sigma_1 \alpha g_{k+1}^T d_{k+1}, \\ g(x_{k+1}+\alpha d_{k+1})^T d_{k+1} \ge \sigma_2 g_{k+1}^T d_{k+1}, \end{cases} \tag{4}$$

where $f_{k+1} = f(x_{k+1})$, $0 < \sigma_1 < (1/2)$ and $\sigma_1 < \sigma_2 < 1$, the step length α_{k+1} is obtained [32]. Finally, the next iteration x_{k+2} is considered as

$$x_{k+2} = x_{k+1} + \alpha_{k+1} d_{k+1}. \tag{5}$$

Some well-known CG methods with $\gamma_{k+1} = 0$ (two-term CG methods) are Hestenes-Stiefel (*HS*) [26], Fletcher-Reeves (*FR*) [23], Polak-Ribiére-Polyak (*PRP*) [33, 34]. It is worth to notice that, these methods are equivalent for convex quadratic functions.

To the best knowledge of authors, the introduction of the new conjugacy condition

$$d_{k+1}^T y_k = -t g_{k+1}^T s_k, \tag{6}$$

where $t \in [0, \infty)$ and $s_k = x_{k+1} - x_k = \alpha_k d_k$ by Dai and Liao [16] has been a source of inspiration for many like Hager and Zhang [25], Babaie-Kafaki and Ghanbari [9], Dong et al. [21], Zheng and Zheng [46] and Liu and Liu [30] to create new nonlinear CG methods.

By choosing $\beta_{k+1} = \beta_{k+1}^{HS}$, $\gamma_{k+1} = g_{k+1}^T y_l / d_l^T y_l$ and $p_k = d_l$ where $1 \le l < k$ for (2), Beale introduced the first three-term CG method in [10]. The directions of this method was not always descent but the idea of three-term CG method was interesting. Therefore, many researchers like Zhang et al. [44], Andrei [5], Narushima et al. [31], Al-Baali et al. [2], Wu [38] and Liu et al. [28] developed this idea and proposed more efficient three-term CG methods.

One way to have a better numerical behavior in CG methods is to multiply the gradient part of CG directions in a positive parameter τ_k or actually using the scaled CG direction

$$d_0 = -g_0, \quad d_{k+1} = -\tau_k g_{k+1} + \beta_{k+1} d_k + \gamma_{k+1} p_k, \quad \text{for } k = 0, 1, 2, \dots. \tag{7}$$

In [11], Birgin and Martínez introduced two parameters for τ_k and showed that their scaled methods are more effective than *PRP* and *FR* methods. Interested readers can find other examples of scaled CG directions in [4, 22, 29, 36, 45]. The authors of these papers indicated that scaled CG methods usually produce more efficient directions than non-scaled ones, in both analytical and numerical points of view.

On the other hand, by carefully examining the structure of the limited memory BFGS (L-BFGS) direction

$$d_{k+1} = -g_{k+1} + \left[\frac{g_{k+1}^T y_k}{s_k^T y_k} - \left(\tau_k + \frac{\|y_k\|^2}{s_k^T y_k} \right) \frac{g_{k+1}^T s_k}{s_k^T y_k} \right] s_k + \frac{g_{k+1}^T s_k}{s_k^T y_k} y_k, \quad \text{for some } \tau_k \ge 0, \tag{8}$$

it turns out that the L-BFGS directions and the three-term CG directions have similar structure. For example, the L-BFGS direction (8) can be considered as a modification of *HS* method because they are identical under exact line search. These similarities enable the authors of [12, 14] to add some parameters to (8) and create some families of three-term CG methods. Interested readers can find other research articles in this field in [7, 17, 20, 40].

An important subject in nonlinear CG methods is the global convergence of algorithms. Based on [1] Gilbert and Nocedal [24] presented a process to prove the global convergence of *HS*, *FR* and *PRP* methods. An important key in their process is that the directions of the method satisfy the sufficient descent condition

$$g_{k+1}^T d_{k+1} \le -c \|g_{k+1}\|^2, \quad \text{for some } 0 < c \le 1 \text{ and all } k \ge 0. \tag{9}$$

For general smooth functions, under WWP line search technique and conditions (9), some researchers used Powell's restart procedures [35] and set $\beta_{k+1}^+ = max\{\beta_{k+1}, 0\}$ or obtained a weaker global convergence theorem.

Another main element in proving the convergence of a CG algorithm is the property

$$y_k^T s_k \geq \mu \|s_k\|^2, \tag{10}$$

of uniformly convex functions. Therefore, in the case of non-convex functions, Yuan et al. [42] proposed a modified WWP line search technique (MWWP) and showed that under a few more conditions, the BFGS and *PRP* methods are globally convergent. Recently, the authors of [13] introduced another modified WWP line search technique. In some numerical experiences, they showed that this technique behave better than both WWP and MWWP techniques. Furthermore, they indicated that under this line search technique, property (10) and the global convergence of BFGS method are true for general smooth functions .Some other modified line search techniques are presented in [15, 19, 27, 41].

Due to the vast applications of large-scale unconstrained optimization problems and the efficiency of CG methods to solve them, in this paper, we introduce a family of scaled three-term CG methods. Furthermore, to achieve an appropriate global convergence theorem for the new family, we propose a modified WWP line search technique. Our main contributions in this paper are as follows:

- Showing that the directions of our family are sufficiently descent.
- Proving a strong global convergence theorem for the new family for uniformly convex and general smooth functions.
- Indicating that the new modified WWP line search technique is well-defined.
- Observing the efficiency and effectiveness of the new presented methods in comparison with similar methods in some numerical experiments.

The rest of this paper is organized as follows. In the next section, we will introduce the new family of three-term CG methods. Then, in Section 3, we will prove the global convergence of the new family for uniformly convex and general smooth functions under WWP line search techniques. In Section 4, we will propose a new modified WWP line search technique and obtain the strong global convergence theorem of the presented family for general smooth functions. In the last section, we will present the results of numerical experiments.

2. The new family of three-term CG methods and its properties

As we mentioned in introduction, two directions (7) and (8) are the source of inspiration of many researchers to introduce efficient CG methods. In both of these directions, a free parameter is added to control the balance between the parts of directions. Plus, the free parameters in a family of CG directions enable the researchers to choose an appropriate member of the family for solving their special problems. Many numerical experiences also indicated that using the free parameters in directions (7) and (8) yielded better numerical results. Given this fact, we decided to use some free parameters in a three-term CG direction which created by combining directions (7) and (8). Therefore, in this section, we first introduce the new family of three-term CG direction as

$$d_{k+1} = -\tau_1 g_{k+1} + \frac{1}{d_k^T y_k}\left(\tau_1 g_{k+1}^T y_k - \tau_2 c_k \|y_k\|^2 - \tau_3 g_{k+1}^T s_k\right) d_k - \tau_1 c_k y_k, \tag{11}$$

where $c_k = g_{k+1}^T s_k / y_k^T s_k$, $0 < \tau_1 \leq 1$ and $\tau_2, \tau_3 \geq 0$. Then, in the following of this section, we show that conditions (6) and (9) are true for (11).

Proposition 2.1. *The directions of the family (11) satisfy the conjugacy condition (6), whenever the line search algorithm guaranty* $y_k^T s_k > 0$.

Proof. For directions (11) we have

$$\begin{aligned} d_{k+1}^T y_k &= -\tau_1 g_{k+1}^T y_k + \tau_1 g_{k+1}^T y_k - \tau_2 c_k \|y_k\|^2 - \tau_3 g_{k+1}^T s_k - \tau_1 c_k \|y_k\|^2 \\ &= -\left(\frac{(\tau_1 + \tau_2)\|y_k\|^2}{y_k^T s_k} + \tau_3 \right) g_{k+1}^T s_k := -t_k g_{k+1}^T s_k. \end{aligned}$$

So, if we use any line search technique which can guaranty $y_k^T s_k > 0$, then t_k is non-negative. Therefore, direction (11) satisfy conjugacy condition (6). □

Throughout this article, for simplicity, we will call the new family of scaled three-term CG directions (11) as *STTCGF*. Since under exact line search we have $g_{k+1}^T s_k = 0$, so *STTCGF* family can be considered as a modification of *HS* method.

Proposition 2.2. *Under any line search technique which can guaranty the positiveness of $y_k^T s_k$, the directions of STTCGF family are sufficient descent.*

Proof. For *STTCGF* we have

$$d_{k+1}^T g_{k+1} = -\tau_1 \|g_{k+1}\|^2 + \tau_1 c_k g_{k+1}^T y_k - \tau_2 c_k^2 \|y_k\|^2 - \tau_3 \frac{(g_{k+1}^T s_k)^2}{y_k^T s_k} - \tau_1 c_k g_{k+1}^T y_k \le -\tau_1 \|g_{k+1}\|^2. \tag{12}$$

So, the sufficient descent condition (9) is satisfied. □

Algorithm 2.3. *Scaled three-term CG family STTCGF*
Initialization: *Choose $\epsilon > 0$ and $x_0 \in \mathbb{R}^n$. Set $k = 0$.*

```
while ||g_k||_∞ > ε do
  if k = 0 then
    Set d_k = −g_k,
  else
    Calculate the search direction d_k by (11).
  end if
  Calculate α_k by an appropriate line search technique.
  Set x_{k+1} = x_k + α_k d_k.
  Set k = k + 1.
end while
```

Remark 2.4. *If we set τ_2 or τ_3 equal to zero in STTCGF family, we can create two new families of scaled three-term CG directions:*

- $\tau_2 = 0$:

$$d_{k+1} = -\tau_1 g_{k+1} + \frac{1}{d_k^T y_k} \left(\tau_1 g_{k+1}^T y_k - \tau_3 g_{k+1}^T s_k \right) d_k - \tau_1 c_k y_k,$$

where, $0 < \tau_1 \le 1$ *and* $\tau_3 \ge 0$.

- $\tau_3 = 0$:

$$d_{k+1} = -\tau_1 g_{k+1} + \frac{1}{d_k^T y_k} \left(\tau_1 g_{k+1}^T y_k - \tau_2 c_k \|y_k\|^2 \right) d_k - \tau_1 c_k y_k,$$

where, $0 < \tau_1 \le 1$ *and* $\tau_2 \ge 0$.

Note that, Propositions 2.1 and 2.2 are true for these families.

In Algorithm 2.3, we propose the pseudo code of the *STTCGF* family. In the next section, we prove the global convergence of Algorithm 2.3 for uniformly convex and general smooth functions under WWP line search technique.

3. Global convergence analysis under WWP line search technique

In this section, we first consider an assumption and Zoutendijk lemma [32]. Then, we prove the global convergence property of Algorithm 2.3 under WWP line search technique for uniformly convex and general smooth functions. Note that, since WWP line search technique can guaranty the positiveness of $y_k^T s_k$, so Propositions 2.1 and 2.2 of *STTCGF* family are true under it.

Assumption 3.1.

i) The level set $\Omega = \{x \in \mathbb{R}^n | f(x) < f(x_0)\}$ *is bounded.*

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

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

Remark 3.2. *From Assumption 3.1, we know that*

- *There exists a constant* $B > 0$ *such that*

$$\|s_k\| \leq B, \quad \forall k \geq 0. \tag{13}$$

- g *is bounded or actually there exists a constant* $\overline{\theta} > 0$ *such that*

$$\|g(x)\| \leq \overline{\theta}, \quad \forall x \in N. \tag{14}$$

- *For all* $k \geq 1$

$$\|y_k\| \leq L\|s_k\|. \tag{15}$$

Lemma 3.3. *(**Zoutendijk lemma**) Under Assumption 3.1 and WWP line search technique, for any iteration of the form (5) with a descent direction* d_k *we have*

$$\sum_{k \geq 0} \frac{(g_{k+1}^T d_{k+1})^2}{\|d_{k+1}\|^2} < \infty. \tag{16}$$

Proof. See Theorem 3.2 of [32]. □

From Propositions 2.2, we are sure that the directions of *STTCGF* family are descent, therefore, we can apply Zoutendijk lemma 3.3 to d_k and g_k generated by Algorithm 2.3. In the next theorem, we establish the global convergence of Algorithm 2.3 for uniformly convex functions.

Theorem 3.4. *Let Assumption 3.1 be true for a uniformly convex function* f *and an initial point* x_0. *Under WWP line search technique, for any sequence* $\{x_k\}$ *generated by Algorithm 2.3 we have*

$$\lim_{k \to +\infty} \|g_{k+1}\| = 0.$$

Proof. From Theorem 1.3.16 of [37], for a uniformly convex function f there exists a constant $\mu > 0$ such that for all $k \geq 0$

$$y_k^T s_k \geq \mu \|s_k\|^2. \tag{17}$$

Now, by considering (11), (14), (15), (17) and Cauchy–Schwarz inequality, we have

$$\begin{aligned}
\|d_{k+1}\| &= \left\| -\tau_1 g_{k+1} + \frac{1}{d_k^T y_k}(\tau_1 g_{k+1}^T y_k - \tau_2 c_k \|y_k\|^2 - \tau_3 g_{k+1}^T s_k) d_k - \tau_1 c_k y_k \right\| \\
&\leq |\tau_1| \|g_{k+1}\| + |\tau_1| \frac{|g_{k+1}^T y_k| \|d_k\|}{|d_k^T y_k|} + |\tau_2| \frac{|g_{k+1}^T s_k| \|y_k\|^2 \|d_k\|}{|y_k^T s_k| |d_k^T y_k|} \\
&+ |\tau_3| \frac{|g_{k+1}^T s_k| \|d_k\|}{|d_k^T y_k|} + |\tau_1| \frac{|g_{k+1}^T s_k| \|y_k\|}{|y_k^T s_k|} \\
&\leq \tau_1 \|g_{k+1}\| + 2\tau_1 \frac{\|g_{k+1}\| \|y_k\| \|s_k\|}{\mu \|s_k\|^2} + \tau_2 \frac{\|g_{k+1}\| \|y_k\|^2 \|s_k\|^2}{\mu^2 \|s_k\|^4} \\
&+ \tau_3 \frac{\|g_{k+1}\| \|s_k\|^2}{\mu \|s_k\|^2} \\
&\leq \left(\tau_1 + 2\tau_1 \frac{L}{\mu} + \tau_2 \frac{L^2}{\mu^2} + \frac{\tau_3}{\mu} \right) \|g_{k+1}\| \\
&\leq \left(\tau_1 + 2\tau_1 \frac{L}{\mu} + \tau_2 \frac{L^2}{\mu^2} + \frac{\tau_3}{\mu} \right) \overline{\theta} := M. \qquad (18)
\end{aligned}$$

Therefore, inequalities (12), (16) and (18) yield

$$\infty > \sum_{k \geq 0} \frac{(g_{k+1}^T d_{k+1})^2}{\|d_{k+1}\|^2} \geq \frac{\tau_1^2}{M^2} \sum_{k \geq 0} \|g_{k+1}\|^4.$$

So, the proof is complete. □

At the end of this section, we prove a weaker global convergence theorem for general smooth functions under WWP line search technique, Assumption 3.1 and one of the following conditions:

$A1$: There exists $\mu' > 0$ such that $s_k^T y_k > \mu'$ for all $k \geq 0$.

$A2$: $\alpha := \min_{k \geq 0}\{\alpha_k\}$ is positive.

Theorem 3.5. *Consider Assumption 3.1 and one of the conditions $A1$ and $A2$ for a smooth function f and an initial point x_0. Under WWP line search technique for sequence $\{x_{k+1}\}$ generated by Algorithm 2.3 we have*

$$\liminf_{k \to +\infty} \|g_{k+1}\| = 0. \qquad (19)$$

Proof. The proof is similar to Lemma 2 and Theorem 2 of [12]. □

In the next section, we obtain a strong global convergence theorem for Algorithm 2.3 for general smooth functions under a new modification of WWP line search technique.

4. The new modified WWP line search technique

As we mentioned before, inequality (10) play the main role of proving the strong global convergence of many CG algorithms. The obvious way to attain inequality (10) is using the uniformly convexity property of f (similar to Theorem 3.4). On the other hand, it was shown that by improving the WWP line search technique, it may be possible to achieve inequality (10) for general smooth functions [13]. Motivated by this idea, in this section, we first introduce the modified WWP line search

$$\begin{cases} f(x_k + \alpha_k d_k) \leq f_k + \sigma_1 \alpha_k g_k^T d_k + \delta h(\alpha_k, d_k), \\ g(x_k + \alpha_k d_k)^T d_k \geq \sigma_2 g_k^T d_k - \delta \alpha_k \|d_k\|^2 h(\alpha_k, d_k), \end{cases} \qquad (20)$$

where

$$h(\alpha, d) = \begin{cases} 0, & \text{if } \alpha = 0, \\ -e^{-\left(\frac{\alpha^2\|d\|^2}{2}\right)}, & \text{if } \alpha > 0, \end{cases} \tag{21}$$

and

$$0 < \sigma_1 < \frac{1}{2}, \quad \sigma_1 < \sigma_2 < 1, \quad 0 < \delta < 1,$$

which we will call as M-WWP line search technique. Then, we prove that M-WWP line search technique is well-defined. At the end of this section, we indicate that under M-WWP line search technique, Algorithm 2.3 is globally convergence for general smooth functions.

Theorem 4.1. *Let Assumption 3.1 be true for a function f and $g_k^T d_k \le 0$. For any $0 < \sigma_1 < (1/2)$, $\sigma_1 < \sigma_2 < 1$ and $0 < \delta < 1$, there exists $\alpha_k \in [\alpha_l, \alpha_u]$ where $0 < \alpha_l < \alpha_u$, which satisfies condition (20).*

Proof. Consider function $\varphi_k(\alpha)$ as

$$\varphi_k(\alpha) = f(x_k + \alpha d_k) - f(x_k) - \sigma_1 \alpha g_k^T d_k - \delta h(\alpha, d_k).$$

So, obviously we have

$$\varphi_k(0) = 0, \quad \lim_{\alpha \to +\infty} \varphi_k(\alpha) = +\infty. \tag{22}$$

Furthermore,

$$\varphi_k^{'+}(0) = (1 - \sigma_1) g_k^T d_k < 0. \tag{23}$$

Therefor, from (22) and (23) there exists an $\alpha^{'} > 0$, such that

$$\varphi_k(\alpha^{'}) = 0, \tag{24}$$

and

$$\varphi_k(\alpha) < 0, \quad \text{for all } \alpha \in (0, \alpha^{'}), \tag{25}$$

which means that

$$f(x_k + \alpha d_k) \le f(x_k) + \sigma_1 \alpha g_k^T d_k + \delta h(\alpha, d_k), \quad \text{for all } \alpha \in [0, \alpha^{'}].$$

Now, from (23), (24), (25) and the fact that φ_k is a differentiable function for all $\alpha > 0$, there exists an interval $[\alpha_l, \alpha_u] \subset (0, \alpha^{'}]$ as which

$$\varphi_k^{'}(\alpha) \ge 0, \quad \text{for all } \alpha \in [\alpha_l, \alpha_u],$$

or namely for all $\alpha \in [\alpha_l, \alpha_u]$, we have

$$g(x_k + \alpha d_k)^T d_k \ge \sigma_1 g_k^T d_k - \delta\alpha \|d_k\|^2 h(\alpha, d_k) \ge \sigma_2 g_k^T d_k - \delta\alpha \|d_k\|^2 h(\alpha, d_k).$$

Note that the last inequality is true because $\sigma_1 < \sigma_2$ and $g_k^T d_k < 0$. The proof is complete. □

In the next theorem, we present some appropriate properties of M-WWP line search technique.

Theorem 4.2. *Under M-WWP line search technique, the following three properties are true for any descent direction d_k ($g_k^T d_k \le 0$):*

1) Consider Assumption 3.1. We have

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

2) For all $k \geq 0$ we have

$$s_k^T y_k > 0. \tag{26}$$

3) There exists a constant $\mu^{'} > 0$ such that for all $k \geq 0$ we have

$$s_k^T y_k > \mu^{'} ||s_k||^2.$$

Proof.

1) From part (*ii*) of Assumption 3.1 and the second part of conditions (20), we have

$$\alpha_k L||d_k||^2 \geq (g_{k+1} - g_k)^T d_k \geq (\sigma_2 - 1) g_k^T d_k - \delta \alpha_k ||d_k||^2 h(\alpha_k, d_k) \geq (\sigma_2 - 1) g_k^T d_k,$$

or

$$-\alpha_k \leq \frac{(1 - \sigma_2)}{L||d_k||^2} g_k^T d_k. \tag{27}$$

Now, by using the fact that $h(\alpha, d) \leq 0$ for all $\alpha > 0$ and vector d and by substituting inequality (27) into the first condition of (20), we get

$$\begin{aligned} f(x_k + \alpha_k d_k) \leq f_k + \sigma_1 \alpha_k g_k^T d_k + \delta h(\alpha_k, d_k) &\leq f_k + \sigma_1 \alpha_k g_k^T d_k \\ &\leq f_k - \frac{\sigma_1 (1 - \sigma_2)}{L||d_k||^2} (g_k^T d_k)^2 = f_k - \xi \frac{(g_k^T d_k)^2}{||d_k||^2}, \end{aligned} \tag{28}$$

where $\xi = \sigma_1 (1 - \sigma_2)/L$ is a positive constant. If we organize inequality (28) as

$$f_k - f(x_k + \alpha_k d_k) \geq \xi \frac{(g_k^T d_k)^2}{||d_k||^2},$$

and sum it over index k, then from part (*i*) of Assumption 3.1 the proof is complete.

2) From the second part of conditions (20), we have

$$s_k^T y_k \geq (\sigma_2 - 1) g_k^T s_k - \delta ||s_k||^2 h(\alpha_k, d_k). \tag{29}$$

Since function $h(\alpha, d)$ is always non-positive and $\sigma_2 < 1$, so the proof is complete.

3) From inequality (29) we have

$$s_k^T y_k \geq \delta ||s_k||^2 \, e^{-\left(\frac{||s_k||^2}{2}\right)}.$$

Now, by considering inequality (13) the proof is complete with $0 < \mu^{'} := \delta e^{-\left(\frac{B^2}{2}\right)} < \infty$.

□

Table 1: The test problems.

N.	Name	N.	Name
1	Extended Trigonometric function	21	Quadratic QF2 function
2	Extended Rosenbrock function	22	Extended quadratic exponential EP1 function
3	Extended Beale function	23	Extended Tridiagonal 2 function
4	Extended Penalty function	24	DQDRTIC function (CUTE)
5	Perturbed Quadratic function	25	Broyden Tridiagonal function
6	Raydan 2 function	26	Almost Perturbed Quadratic function
7	Hager function	27	Perturbed Tridiagonal Quadratic function
8	Generalized Tridiagonal 1 function	28	ENGVAL1 function (CUTE)
9	Extended Tridiagonal 1 function	29	EDENSCH function (CUTE)
10	Extended TET function	30	BDEXP function (CUTE)
11	Diagonal 4 function	31	QUARTC function (CUTE)
12	Diagonal 5 function	32	Extended DENSCHNB function (CUTE)
13	Extended Himmelblau function	33	Extended DENSCHNF function (CUTE)
14	Extended PSC1 function	34	COSINE function (CUTE)
15	Extended BD1 function	35	Generalized Quartic function
16	Extended Maratos function	36	Diagonal 7 function
17	Extended Wood function	37	Diagonal 8 function
18	Quadratic QF1 function	38	Full Hessian FH3 function
19	Extended quadratic penalty QP1 function	39	SINCOS function
20	Extended quadratic penalty QP2 function	40	HIMMELBG function (CUTE)

Here, we are ready to prove a strong global convergence theorem for Algorithm 2.3 for general smooth functions. Note that from (26) we know that under M-WWP line search technique, Propositions 2.1 and 2.2 of *STTCGF* family are true.

Theorem 4.3. *Consider Assumption 3.1 for a smooth function f and an initial point x_0. Under M-WWP line search technique for any sequence $\{x_k\}$ generated by Algorithm 2.3 we have*

$$\lim_{k\to+\infty} \|g_{k+1}\| = 0.$$

Proof. By considering Theorem 4.2, the proof is similar to Theorem 3.4. □

At the end of this section, we present two remarks to explain some other appropriate advantages of our proposed methods.

Remark 4.4. *By using M-WWP instead of WWP line search technique the global convergence theorems of many line search methods (like the Broyden family in quasi-Newton and DL family in CG classes) can be obtained for general smooth functions.*

Remark 4.5. *In recent decades, many researchers such as [39, 43, 47] have developed conjugate gradient methods for solving systems of nonlinear equations. They showed that the roots of the equation $F(x) = 0$ where $F : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is Lipschitz continuous and the solutions of the optimization problem*

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|F(x)\|^2,$$

are equivalent. Consequently, STTCGF family can be considered as a family of derivative-free methods for solving systems of nonlinear equations. Note that, if we use M-WWP line search technique, the resulting algorithm will be convergence for non-monotone problems.

5. Numerical results

In this section, we are going to prove the numerically efficiency of M-WWP line search technique and *STTCGF* family. To this aim, we design two competitions including forty standard test problems and their initial points from [3]. The first competition is between WWP and M-WWP line search techniques and the second one is between *STTCGF* family and five different CG methods. We compare all of these methods in four terms:

Table 2: The methods which participate in the second competition.

Name	Direction	Reference
CGLFZ	$d_{k+1} = -g_{k+1} + \frac{g_{k+1}^T y_k}{\|d_k\|^2} d_k - \frac{g_{k+1}^T d_k}{\|d_k\|^2} y_k,$	[28]
CGYN	$d_{k+1} = -g_{k+1} + max\left(\frac{t_k\, g_{k+1}^T y_k - g_{k+1}^T s_k}{d_k^T y_k}, 0\right) d_k + t_k\, \frac{g_{k+1}^T s_k}{s_k^T y_k} y_k,$	[40]
	$t_k = min\left\{\frac{(s_k^T y_k)^2}{(s_k^T y_k)^2 + \|s_k\|^2\|y_k\|^2}, \frac{s_k^T y_k}{\|y_k\|^2}\right\}$	
CGDW	$d_{k+1} = -g_{k+1} - \left(\left(1 - \min\left\{1, \frac{\|y_k\|^2}{s_k^T y_k}\right\}\right)\frac{g_{k+1}^T s_k}{s_k^T y_k} - \frac{g_{k+1}^T y_k}{s_k^T y_k}\right) s_k - \frac{g_{k+1}^T s_k}{s_k^T y_k} y_k,$	[17]
CGBKG	$d_{k+1} = -g_{k+1} + \left(\frac{g_{k+1}^T y_k}{d_k^T y_k} - \left(\frac{s_k^T y_k}{\|s_k\|^2} + \frac{\|y_k\|}{\|s_k\|}\right)\frac{g_{k+1}^T s_k}{d_k^T y_k}\right) d_k,$	[8]
CGHZ	$d_{k+1} = -g_{k+1} + \left(\frac{g_{k+1}^T y_k}{d_k^T y_k} - \left(2\frac{\|y_k\|^2}{s_k^T y_k}\right)\frac{g_{k+1}^T s_k}{d_k^T y_k}\right) d_k,$	[25]
STTCGFs	$d_{k+1} = -\tau_1 g_{k+1} + \frac{1}{d_k^T y_k}\left(\tau_1 g_{k+1}^T y_k - \tau_2 \frac{g_{k+1}^T s_k}{s_k^T y_k}\|y_k\|^2 - \tau_3 g_{k+1}^T s_k\right) d_k$	presented
	$-\tau_1 \frac{g_{k+1}^T s_k}{s_k^T y_k} y_k,$ $(\tau_1, \tau_2, \tau_3) = (0.7, 0.2, 0.75)$	method

- k: The number of iterations.
- kf: The number of function evaluations.
- kg: The number of gradient evaluations.
- t: The CPU time in seconds.

The names of these chosen test problems are shown in Table 1. The methods which participate in the second competition are presented in Table 2. We consider all these problems in five dimensions [100, 500, 1000, 5000, 10000] for the first competition and in five dimensions [1000, 5000, 10000, 15000, 20000] for the second one and run all the codes in MATLAB 8.4.1 and a LAP's (Intel Core i7-7500U, up to 3.5 GHz, 8GB Memory) with Windows 10 operating system.

It is important to notice that, *CGLFZ*, *CGYN* and *CGDW* methods are three well-behaved newly developed three-term CG methods with global convergence property under WWP line search technique. On the other hand, *CGBKG* is a member of DL family with optimal parameter which shows interesting results in numerical experiments. Likewise, *CGHZ* is a well-known member of DL family with sufficiently descent directions and superior numerical results.

Remark 5.1. *In Table 2, we consider a member of STTCGF family and called it STTCGFs. Since defining the best CG direction and also finding an optimal parameter for CG classes are open problems [6], the parameters of STTCGFs method are selected based on some numerical experiments and they are not optimal. Therefore, better values for parameters* (τ_1, τ_2, τ_3) *or actually more efficient members of STTCGF family can be find to solve the chosen set of test problems in Table 1 or for any set of test problems.*

For both WWP and M-WWP line search techniques, we use a bisection algorithm similar to Algorithm 2.5.1 of [37] and set the initial step lengths as

$$\alpha_0^0 = 1, \quad \alpha_{k+1}^0 = \alpha_k \frac{\|d_k\|}{\|d_{k+1}\|}, \quad \text{for } k = 0, 1, 2, \ldots.$$

Furthermore, in all experiments, we set $\sigma_1 = 10^{-4}$, $\sigma_2 = 0.8$ for WWP and M-WWP line search techniques and $\epsilon = 10^{-5}$ for termination condition in CG algorithms. Also, we terminate the CG algorithms when the number of iterations or the number of function evaluations are greater than 4000 or 20000, respectively. Furthermore, to avoid an uphill search direction in numerical experiments, we stop the loop of line search

Table 3: Some numerical results of Comparing three line search techniques.

No.	Dim. ($\times 10^2$)	WWP $k/kf/kg/t$	M-WWPs1 $k/kf/kg/t$	M-WWPs2 $k/kf/kg/t$
2	1	248/731/474/0.0227	241/740/479/0.0222	248/731/474/0.0213
	5	2371/2955/2648/0.2214	324/932/613/0.0511	2315/2899/2592/0.2328
	10	2510/3092/2786/0.3724	2510/3092/2786/0.3931	2481/3063/2757/0.3942
	50	2434/3066/2736/1.594	799/1480/1124/0.6636	2434/3066/2736/1.632
3	10	110/268/178/0.0901	87/204/133/0.0678	110/268/178/0.0901
5	1	276/436/341/0.0264	213/339/261/0.0132	276/436/341/0.0184
	5	793/1427/1092/0.0741	781/1402/1074/0.0789	763/1374/1050/0.0778
	10	1044/1860/1432/0.1346	1014/1827/1401/0.1453	1015/1804/1391/0.1425
11	10	175/399/274/0.0298	175/399/274/0.0317	174/401/274/0.0318
16	1	469/1097/768/0.0437	407/993/686/0.0328	469/1097/768/0.0377
	5	441/906/658/0.0578	447/1014/714/0.0646	441/906/658/0.0604
	10	359/801/565/0.0798	355/800/564/0.0827	359/801/565/0.0831
	100	291/675/469/0.5803	275/666/454/0.6099	291/675/469/0.6614
17	1	939/1891/1396/0.2861	965/1901/1414/0.310	939/1891/1396/0.288
	5	1740/2571/2138/1.909	1993/2723/2339/2.079	1740/2571/2138/1.920
	10	2913/3461/3170/5.424	1807/2506/2138/3.792	2911/3459/3168/5.679
	50	1923/2635/2260/22.50	1911/2663/2252/23.71	1923/2635/2260/23.25
	100	1886/2743/2297/44.26	1950/2814/2363/45.19	1886/2743/2297/44.61
18	5	791/1396/1076/0.0714	819/1444/1116/0.0860	722/1254/970/0.0669
	10	1420/1878/1631/0.1799	1451/1997/1705/0.1967	1420/1878/1631/0.1900
20	1	147/408/267/0.0180	125/356/227/0.0145	147/408/267/0.0164
	10	236/689/444/0.0909	208/630/402/0.0830	236/689/444/0.0948
	100	448/840/623/1.154	456/909/663/1.233	448/840/623/1.191
21	1	331/531/416/0.0288	367/609/472/0.0271	325/529/411/0.0250
24	1	343/631/469/0.0276	310/614/444/0.0260	326/608/449/0.0273
	5	404/751/559/0.0500	347/659/484/0.0487	394/724/540/0.0550
	10	197/383/271/0.0385	263/501/364/0.0595	199/378/270/0.0427
	50	370/704/518/0.3163	275/506/372/0.2385	370/704/518/0.3321
	100	359/714/518/0.6190	352/689/501/0.6384	379/737/539/0.6483
25	10	305/527/403/0.0743	279/482/366/0.0711	305/527/403/0.0782
26	5	503/918/693/0.0451	669/1054/844/0.0636	503/918/693/0.0501
	10	2084/2573/2310/0.2686	916/1679/1278/0.1254	1524/2013/1750/0.1991
27	1	279/426/338/0.0214	305/478/377/0.0222	262/413/322/0.0191
	5	712/1119/898/0.0848	919/1492/1186/0.1281	712/1119/898/0.0975
	10	1279/1845/1544/0.2459	1279/1845/1544/0.2512	1149/1715/1414/0.2324

algorithm after 15 tries. In these case we does not stop the CG algorithm because some CG methods may find better direction after a very small step.

For comparing the methods, we use the technique of Dolan and Moré [18]. They supposed that there are n_q solvers $\{q_1, q_2, \dots q_{n_q}\}$ in set $\mathcal{Q}$ and n_p problems $\{p_1, p_2, \dots p_{n_p}\}$ in set $\mathcal{P}$. To comparing all solvers in $\mathcal{Q}$ in term of property a, they defined a probability function $P_{q_i}(\tau)$ with a threshold $\tau \geq 1$. For example, in term of t, $P_{q_1}(\tau) = 0.7$ means that solver q_1 solved 70% of all the problems in $\mathcal{P}$ within a factor τ of the best possible t. We name the graph of $P_{q_i}(\tau)$ for $i = 1, 2, \dots n_q$, the performance profiles.

In the first competition, we set $\delta = 10^{-8}$ and $\delta = 10^{-13}$ in M-WWP line search technique and called them M-WWPs1 and M-WWPs2, respectively. For finding the directions, we consider a member of *STTCGF* family which is presented in Table 2 as *STTCGs* method.

Since showing all the results of this competition needed a huge space, in Table 3 we just present the problems which different methods solve them with different k. In addition, the $P_q(1)$ of comparing these three algorithms is shown in Table 4 in three parts:

- **Part 1:** $P_q(1)$ of comparing three algorithms WWP, M-WWPs1 and M-WWPs2 with each other.
- **Part 2:** $P_q(1)$ of comparing two algorithms WWP and M-WWPs1 with each other.
- **Part 3:** $P_q(1)$ of comparing two algorithms WWP and M-WWPs2 with each other.

Note that, for example, the number 0.3650 in the second row and last column of Table 4 means that M-WWPs1 solved 36.50% of test problems with best t.

The results of our first competition in Tables 3 and 4 show that:

1) Even with a small coefficient δ (nearly zero in M-WWPs2), WWP and M-WWP techniques can eventuate different results (Table 3).

Table 4: $P_q(1)$ of the first competition.

Part 1:	k	kf	kg	t
WWP	0.8100	0.8150	0.8150	0.2550
M-WWPs1	0.8600	0.8500	0.8550	0.3650
M-WWPs2	0.8400	0.8550	0.8550	0.3200
Part 2:	k	kf	kg	t
WWP	0.8400	0.8450	0.8450	0.3350
M-WWPs1	0.8800	0.8750	0.8750	0.6050
Part 3:	k	kf	kg	t
WWP	0.8750	0.8750	0.8750	0.3850
M-WWPs2	0.9300	0.9300	0.9350	0.5550

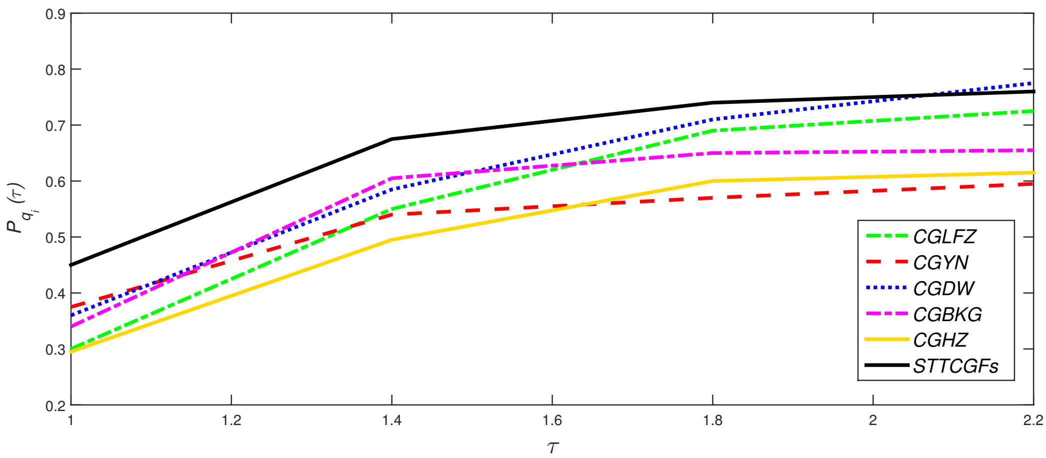

Figure 1: The performance profiles of six methods of second competition in term of k.

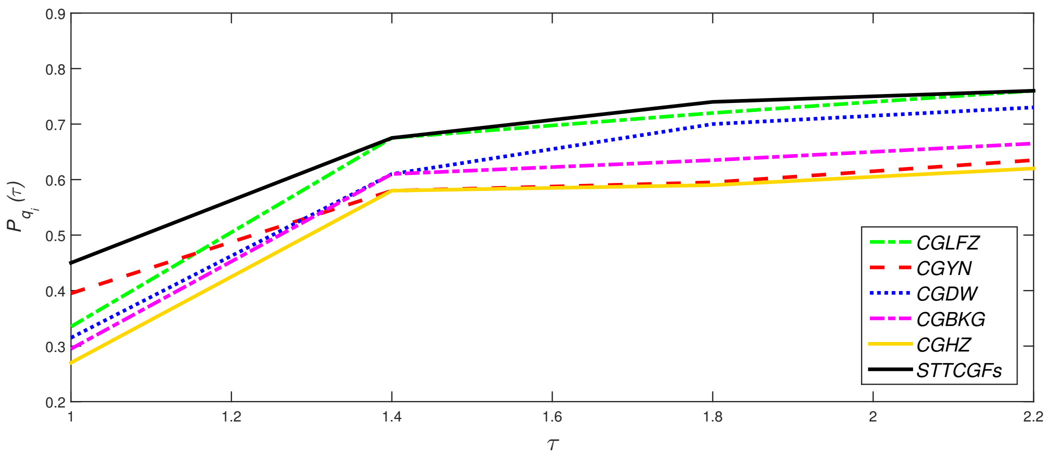

Figure 2: The performance profiles of six methods of second competition in term of kf.

2) As the coefficient δ is increased, the number of iterations (k) and the CPU time (t) have decreased and the numbers of function and gradient evaluations (kf and kg) have increased (**Part 1** of Table 4).

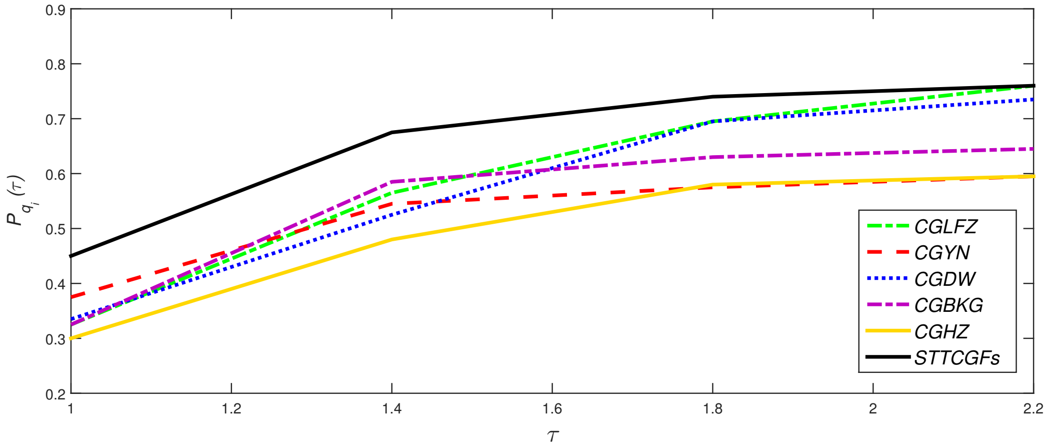

Figure 3: The performance profiles of six methods of second competition in term of kg.

Table 5: $P_q(1)$ of all methods in second competition.

	CGLFZ	*CGYN*	*CGDW*	*CGBKG*	*CGHZ*	*STTCGFs*
k	0.3000	0.3750	0.3600	0.3400	0.2950	0.3950
kf	0.3350	0.3950	0.3150	0.2950	0.2700	0.4500
kg	0.3250	0.3750	0.3350	0.3250	0.3000	0.4000
t	0.1300	0.0750	0.1450	0.1800	0.1000	0.2450

3) M-WWP technique outstrips WWP technique in numerical experiments (**Parts 2** and **3** of Table 4).

4) Since the result of Theorem 4.3 is true as long as $\delta > 0$, we can enjoy the advantage of using M-WWP technique at no extra numerical cost over using WWP technique (**Part 3** of Table 4).

The second result of our first competition can be confirm from theoretical point of view. Because as the coefficient δ has increased, the conditions of M-WWP technique become stronger and thus the resulted value of α_k is closer to its optimal value (the solution of exact line search problem (3)). Therefore, it is likely that the number of iterations which are required for the CG algorithm to achieve the optimal solution (k), will be reduced. On the other hand, as the conditions of M-WWP technique are stronger, so the number of iterations of the line search algorithm and naturally kf and kg have increased. In addition, the behavior of t depends on the tested problem. Here, since in most of the problems in Table 1, the cost of computing the values of the objective functions and their gradients are low, so the cost of the CG algorithm depends on k. As a result, the value of t has also decreased with decreasing k.

In the second competition, we compare *STTCGFs* method (which is introduced in Table 2 as a member of *STTCGF* family) with five other well-behaved CG methods. Since we do not prove the global convergence of the first five methods of Table 2 under M-WWP line search technique, in all the CG algorithms of this competition we use WWP line search technique.

We propose the performance profiles of the second competition in Figs. 1, 2, 3 and 4. In order to have more clearance, we present the $P_q(1)$ of all the methods of the second competition in Table 5.

As we can see, the selected member of *STTCGF* family is the best one in all terms. Therefore, this competition indicate the efficiency of our proposed family of scaled three-term CG methods in numerical experiments.

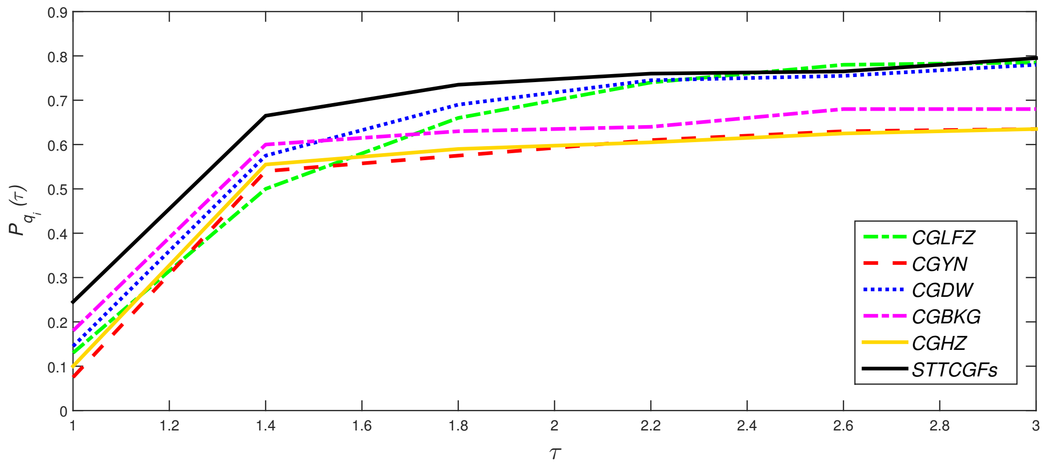

Figure 4: The performance profiles of six methods of second competition in term of t.

6. Conclusion

In this paper,

- We proposed a family of scaled three-term CG methods (*STTCGF*) to solve a large-scale smooth unconstrained optimization problem.
- We indicated that, the directions of *STTCGF* family are sufficiently descent and fulfill the Dai-Liao conjugacy conditions.
- We proved the strong global convergence theorem of *STTCGF* family for uniformly convex functions and a weaker one for general smooth functions under WWP line search technique.
- We introduced a modified WWP line search technique (M-WWP) and established the strong global convergence theorem of *STTCGF* family for general smooth functions.
- We observed the numerical efficiency and effectiveness of the proposed new methods in comparison with similar methods, in two sets of numerical experiments.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgment

The authors would like to thank the anonymous reviewers for their precious comments which were greatly helpful to improve the quality of this paper.

References

[1] 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.

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

[3] N. Andrei, *An unconstrained optimization test functions collection*, Adv. Model. Optim. **10** (2008) 147–161.

[4] N. Andrei, *Accelerated scaled memoryless BFGS preconditioned conjugate gradient algorithm for unconstrained optimization*, Eur. J. Oper. Res. **204** (2010) 410–420.

[5] N. Andrei, *A modified Polak-Ribiére-Polyak conjugate gradient algorithm for unconstrained optimization*, Optimization **60** (2011) 1457–1471.

[6] N. Andrei, *Open problems in nonlinear conjugate gradient algorithms for unconstrained optimization*, B. Malays. Math. Sci. Soc. **34** (2011) 319–330.

[7] N. Andrei, *On three-term conjugate gradient algorithms for unconstrained optimization*, Appl. Math. Comput. **219** (2013) 6316–6327.

[8] S. Babaie-Kafaki, *On optimality of two adaptive choices for the parameter of Dai-Liao method*, Optim. Lett. **10** (2016) 1789–1797.

[9] S. Babaie-Kafaki, R. Ghanbari ,*A descent extension of the Polak-Ribiére-Polyak conjugate gradient method*, Comput. Math. Appl. **68** (2014) 2005–2011.

[10] E. M. L. Beale, *A derivation of conjugate gradients*, Numer. Methods for Nonlinear Optim. (Academic Press) (1972) 39–43.

[11] E. G. Birgin, J. M. Martínez, *A spectral conjugate gradient method for unconstrained optimization*, Appl. Math. Opt. **43** (2001) 117–128.

[12] S. Bojari, M.R. Eslahchi, *Two families of scaled three-term conjugate gradient methods with sufficient descent property for nonconvex optimization*, Numer. Algorithms **83** (2020) 901–933.

[13] S. Bojari, M.R. Eslahchi, *Global convergence of a family of modified BFGS methods under a modified weak-Wolfe-Powell line search for nonconvex functions*, 4OR-Q. J. Oper. Res. **18** (2020) 219–244.

[14] S. Bojari, M.R. Eslahchi, *Global convergence of a new sufficient descent spectral three-term conjugate gradient class for large-scale optimization*, Optim. Methods Softw. **37** (2022) 830–843.

[15] Y. H. Dai, C. X. Kou, *A nonlinear conjugate gradient algorithm with an optimal property and an improved Wolfe line search*, SIAM J. Optim. **23** (2013) 296–320.

[16] Y. Dai, L. Liao, *New conjugacy conditions and related nonlinear conjugate gradient methods*, Appl. Math. Opt. **43** (2001) 87–101.

[17] S. Deng, Z. Wan, *A three-term conjugate gradient algorithm for large-scale unconstrained optimization problems*, Appl. Numer. Math. **92** (2015) 70–81.

[18] E. D. Dolan, J.J. Moré, *Benchmarking optimization software with performance profiles*, Math. Program. **91** (2002) 201–213.

[19] Y. Dong, *New step lengths in conjugate gradient methods*, Comput. Math. Appl. **60** (2010) 563–571.

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

[21] X. L. Dong, H. W. Liu, Y. B. He, *A self-adjusting conjugate gradient method with sufficient descent condition and conjugacy condition*, J. Optimiz. Theory App. **165** (2015) 225–241.

[22] M. Fatemi, *A scaled conjugate gradient method for nonlinear unconstrained optimization*, Optim. Method. Softw. **32** (2016) 1095–1112.

[23] R. Fletcher, C. Reeves, *Function minimization by conjugate gradients*, Comput. J. **7** (1962) 149–154.

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

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

[26] M. R. Hestenes, E. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Res. Nat. Bur. Stand. **49** (1952) 409–436.

[27] Y. Huang, C. Liu, *Dai-Kou type conjugate gradient methods with a line search only using gradient*, J. Inequal. Appl. **66** (2017) 1–17.

[28] 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.

[29] J. K. Liu, Y. M. Feng, L. M. Zou, *A spectral conjugate gradient method for solving large-scale unconstrained optimization*, Comput. Math. Appl. **77** (2019) 731–739.

[30] Z. Liu, H. Liu, *Several efficient gradient methods with approximate optimal step sizes for large-scale unconstrained optimization*, J. Comput. Appl. Math. **328** (2018) 400–413.

[31] 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.

[32] J. Nocedal, S.J. Wright, *Numerical optimization*, (2rd edition), Springer Series in Operations Research, Springer, 2006.

[33] E. Polak, G. Ribiére, *Note sur la convergence de directions conjuguées*, Francaise Infomat Recherche Operatonelle **3** (1969) 35–43.

[34] B. T. Polyak, *The conjugate gradient method in extreme problems*, Ussr. Comp. Math. Math+ **9** (1969) 94–112.

[35] M. J. D. Powell, *Restart procedures for the conjugate gradient method*, Math. Program. **12** (1977) 241–254.

[36] F. Rahpeymaii, K. Amini, T. Allahviranloo, M. R. Malkhalifeh, *A new class of conjugate gradient methods for unconstrained smooth optimization and absolute value equations*, Calcolo **56** (2019) 1–28.

[37] W. Sun, Y.X. Yuan, *Optimization theory and methods: nonlinear programming*, Springer Optimization and Its Applications, Springer, 2006.

[38] Y. Wu, *A modified three-term PRP conjugate gradient algorithm for optimization models*, J. Inequal. Appl. **97** (2017) 1–14.

[39] Y. J. Xie, C. F. Ma, *The scaling conjugate gradient iterative method for two types of linear matrix equations*, Comput. Math. Appl. (2015) 1098–1113.

[40] 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.

[41] G. Yu, L. Guan, Z. Wei, *Globally convergent Polak-Ribiére-Polyak conjugate gradient methods under a modified Wolfe line search*, Appl. Math. Comput. **215** (2009) 3082–3090.

[42] G. Yuan, Z. Wei, X. Lu, *Global convergence of BFGS and PRP methods under a modified weak Wolfe-Powell line search*, Appl. Math. Model. **47** (2017) 811-825.

[43] L. Zhang, *A derivative-free conjugate residual method using secant condition for general large-scale nonlinear equations*, Numer. Algorithms **83** (2020) 1277–1293.

[44] L. Zhang, W. Zhou, D. H. Li, *A descent modified Polak-Ribiére-Polyak conjugate gradient method and its global convergence*, IMA J. Numer. Anal. **26** (2006) 629–640.

[45] L. Zhang, W. Zhou, D. H. Li, *Global convergence of a modified FR conjugate gradient method with Armijo-type line search*, Numer. Math. **104** (2006) 561–572.

[46] Y. Zheng, B. Zheng, *Two new Dai-Liao-type conjugate gradient methods for unconstrained optimization problems*, J. Optimiz. Theory App. **175** (2017) 502–509.

[47] W. Zhou, D.H. Li, *On the Q-linear convergence rate of a class of methods for monotone nonlinear equations*, Pac. J. Optim. **14** (2018) 723-737.