RAIRO-Oper. Res. 51 (2017) 67–77
DOI: 10.1051/ro/2016028

RAIRO Operations Research
www.rairo-ro.org

SOME MODIFIED YABE–TAKANO CONJUGATE GRADIENT METHODS WITH SUFFICIENT DESCENT CONDITION

Xiao Liang Dong[1], Wei Jun Li[2] and Yu Bo He[3]

Abstract. Descent condition is a crucial factor to establish the global convergence of nonlinear conjugate gradient method. In this paper, we propose some modified Yabe–Takano conjugate gradient methods, in which the corresponding search directions always satisfy the sufficient descent property independently of the convexity of the objective function. Differently from the existent methods, a new update strategy in constructing the search direction is proposed to establish the global convergence of the presented methods for the general nonconvex objective function. Numerical results illustrate that our methods can efficiently solve the test problems and therefore is promising.

Mathematics Subject Classification. 49M37, 65K05, 90C53.

Received August 22, 2015. Accepted April 2, 2016.

1. Introduction

The problem considered in this paper is

$$\min f(x), \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is smooth and its gradient $g(x) = \nabla f(x)$ is available.

Throughout the whole paper, for a sufficiently differentiable function f at x_k, we introduce the following notations: $f_k = f(x_k)$; $g_k = g(x_k)$; $y_k = g_{k+1} - g_k$; $s_k = x_{k+1} - x_k$.

Conjugate gradient methods have attracted special attention for solving large–scale unconstrained problems, due to their low memory and simple computational scheme, that is,

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

where $\alpha_k > 0$ is commonly chosen to satisfy certain line search conditions. Among them, the Wolfe conditions have attracted special attention in the convergence analyses and the implementations of CG methods, requiring that

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \rho_1 \alpha_k g_k^T d_k, \tag{1.3}$$

$$g\left(x_k + \alpha_k d_k\right)^T d_k \geq \sigma g_k^T d_k, \tag{1.4}$$

Keywords. Yabe–Takano conjugate gradient method, global convergence, sufficient descent condition, conjugacy condition, numerical comparison.

[1] School of Mathematics and Information science, Beifang University of Nationalities, No. 204 North Wenchang Rd,Yinchuan, Ningxia 750021, P.R. China. dongxl@nun.edu.cn

[2] Network Information Technology Center, Beifang University of Nationalities, No. 204 North Wenchang Rd, Yinchuan, Ningxia 750021, P.R. China. liweijun51@163.com

[3] Department of Mathematics and Applied Mathematics, Huaihua University, No. 612 Yingfeng East Road, Huaihua, Hunan 418008, P.R. China. heyinprc@csu.edu.cn

Article published by EDP Sciences

where $0 < \rho_1 < \sigma < 1$ are constants. The search direction d_k is of the form:

$$d_k = \begin{cases} -g_k, & k = 1, \\ -g_k + \beta_k d_{k-1}, & k \geq 2. \end{cases} \tag{1.5}$$

Different CG methods correspond to different choices for the parameter β_k, which in turn lead to quite diverse computational efficiency and convergence results of the corresponding methods. Well-known formulate for β_k are called the Fletcher–Reeves [18], Polak−Ribière−Polyak [22], Conjugate−Descent [17], Dai–Yuan [8], and Liu–Store [29] and Hestenes–Stiefel [25] formulae. We refer to an excellent survey [24] for a review on recent advances in this area.

Conjugacy condition is an important factor in CG methods. The searching directions in CG methods are often selected in such a way that, when applied to minimize a strongly quadratic convex function, two successive directions are conjugate if no round-off error exists. But for general nonlinear functions,in most methods, the search directions rarely satisfy the conjugacy condition.

In 1952, Hestenes and Stiefel derive β_k^{HS} by requiring the search direction d_k to be $\nabla^2 f(x_k)$-conjugate to d_{k-1}, *i.e.*, enforcing $d_k \nabla^2 f(x_k) d_{k-1} = 0$. For the general function, this condition is replaced by the pure conjugacy condition $d_k^T y_{k-1} = 0$, which is independent of the objective function and line search used. By introducing a constant $\tau > 0$, Dai and Liao [9] replaced the HS conjugacy condition by the the Dai–Liao (DL) conjugacy condition

$$d_k^T y_{k-1} = -\tau g_k^T s_{k-1}, \tag{1.6}$$

and the Dai–Liao (DL) CG method

$$\beta_k^{DL} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} - \tau \frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}}. \tag{1.7}$$

The DL conjugacy condition has been intensively studied. More recently, much effort was directed to utilize second order information to improve the numerical efficiency of conjugate gradient methods based on the condition (1.6) with higher order accuracy in the approximation of the curvature of the objective function. For example, Babaie-Kafaki *et al.* [4] and Yabe and Takano [35] applied a revised form of the modified secant equation proposed by Zhang and Xu [39] and Zhang *et al.* [40] and the modified secant equation proposed by Yuan [37] (see also [38]). Zhou and Zhang [43] applied the modified secant equation proposed by Li and Fukushima [26, 27]. Li *et al.* [28] used the modified secant equation proposed by Wei *et al.* [32]. Ford *et al.* [21] employed the Multi-step quasi-Newton conditions proposed by Ford and Moghrabi [19, 20, 30].

Conjugate gradient methods based on the secant conditions above do not always generate a descent search direction theoretically, but they perform well in practice for solving large–scale unconstrained optimization problems. Thus, it is very indispensable to derive a conjugate gradient method which employs second order information of the objective function and generates a descent search direction.

In an effort to make a modification on the CG method in order to achieve the sufficient descent condition, that is, there exists a constant $c > 0$ such that

$$d_k^T g_k \leq -c\|g_k\|^2, \forall k \in N. \tag{1.8}$$

The papers by Hager and Zhang [23], Dai and Kou [7], Y. Narushima, H. Yabe and J.A. Ford [31], Zhang, Zhou and Li [41], Cheng [6], Andrei [2, 3], Yu [36] and Dong [11–16] proposed several conjugate gradient methods together with their global convergence, in which their search directions satisfy (1.8) independently of the line search used. Extensive numerical results showed the exploitation of new techniques greatly enhance the numerical efficiency.

My concern, also my hope for further study the Yabe–Takano (YT) method, in this paper, is that the sufficient descent property and global convergence can simultaneously shine through to efficiently enhance the numerical efficiency. Furthermore, it is indeed highly desirable to establish the global convergence for the general nonconvex objective function of interest.

The remainder of this paper is organized as follows. In Section 2, the motivation and our method are proposed. Section 3 is devoted to proving the global convergence of our proposed method under the Wolfe line search. In Section 4, we extend this new technique to the other CG methods. In Section 5, some numerical results are given to illustrate the efficiency and robustness of the algorithms. Conclusions are stated in the last section.

2. Motivation and properties

Following Dai and Liao, Yabe and Takano [35] incorporated a nonnegative parameter ρ into the modified secant condition by Zhang *et al.* [39, 40], and presented the YT method, in which

$$\beta_k^{YT} = \frac{g_k^T \varpi_{k-1}}{d_{k-1}^T \varpi_{k-1}} - \tau \frac{g_k^T s_{k-1}}{d_{k-1}^T \varpi_{k-1}}, \tag{2.1}$$

where $\tau > 0$ is a constant and

$$\varpi_{k-1} = y_{k-1} + \rho \frac{\theta_{k-1}}{s_{k-1}^T u_k} u_k. \tag{2.2}$$

In (2.2), the parameter θ_k is presented by

$$\theta_{k-1} = 6(f_{k-1} - f_k) + 3(g_{k-1} + g_k)^T s_{k-1} \tag{2.3}$$

and $u_k \in \mathbb{R}^n$ satisfying $s_{k-1}^T u_k \neq 0$.

Although the DL and YT methods had a reasonable conjugacy condition, and seldom generated uphill search direction in actual computation, its direction may not be a descent one in theory. Moreover, Andrei [1] pointed out that these methods are very dependent on the parameter τ.

Also, the original YT method is globally convergent for uniformly convex function provided that the search direction is a descent direction and the strong Wolfe line search is used.

Obviously, the above strict conditions greatly rule out the application of the YT method. To overcome this defect, Zheng *et al.* [42] presented a modified YT method, in which

$$d_k^{MYT} = -g_k + \beta_k^{YT} d_{k-1} - \frac{g_k^T d_{k-1}}{d_{k-1}^T \varpi_{k-1}} \left(\varpi_{k-1} - \tau s_{k-1} \right), \tag{2.4}$$

and the property $d_k^T g_k = -\|g_k\|^2$ always holds.

As an important point, under mild condition, they establish the global convergence of the MYT method above for the convex objective function. However, the aforementioned methods suffer from one or two of the following problems:

- 1. The MYT method rely on the convexity of the objective function, that is, the global convergence of the MYT method is established for convex objective function.
- 2. By the relationship $s_k = \alpha_k d_k$, we rewrite the search direction d_k^{MYT} as follows

$$\begin{aligned} d_k^{MYT} &= -g_k + \beta_k^{YT} d_{k-1} - \frac{g_k^T d_{k-1}}{\omega_{k-1}^T d_{k-1}} (\omega_{k-1} - \tau s_{k-1}) \\ &= -g_k + \frac{g_k^T \omega_{k-1}}{\omega_{k-1}^T d_{k-1}} d_{k-1} - \tau \frac{g_k^T s_{k-1}}{\omega_{k-1}^T d_{k-1}} d_{k-1} - \frac{g_k^T d_{k-1}}{\omega_{k-1}^T d_{k-1}} (\omega_{k-1} - \tau \alpha_{k-1} d_{k-1}) \\ &= -g_k + \frac{g_k^T \omega_{k-1}}{\omega_{k-1}^T d_{k-1}} d_{k-1} - \frac{g_k^T d_{k-1}}{\omega_{k-1}^T d_{k-1}} \omega_{k-1}. \end{aligned} \tag{2.5}$$

We obviously obtain from (2.5) that the search direction d_k^{MYT} is independent of the choice of the parameter τ, or equivalently, which corresponds to $\tau = 0$.

In this paper, we consider another approach to generate an efficient and robust conjugate gradient algorithm. Specifically, we take a little modification to the YT method, given by

$$d_k^{DYT1} = -g_k + \beta_k^{DYT1} d_{k-1} - \frac{g_k^T d_{k-1}}{d_{k-1}^T \lambda_{k-1}} \lambda_{k-1}. \tag{2.6}$$

In (2.6), the parameter β_k^{DYT1} and λ_{k-1} are defined by

$$\beta_k^{DYT1} = \frac{g_k^T \lambda_{k-1}}{d_{k-1}^T \lambda_{k-1}} - \xi \frac{g_k^T s_{k-1}}{d_{k-1}^T \lambda_{k-1}}, \tag{2.7}$$

$$\lambda_{k-1} = y_{k-1} + \rho \frac{\max\{0, \theta_{k-1}\}}{s_{k-1}^T u_k} u_k, \tag{2.8}$$

where the parameter θ_{k-1} is defined by (2.3) and ξ and ρ are two positive constants.

Lemma 2.1. *Suppose that the search direction is defined by* (2.6) *and the steplength* α_k *satisfies the Wolfe conditions* (1.3) *and* (1.4). *Then it satisfies the sufficient descent condition,* i.e.*,*

$$d_k^T g_k \le -\|g_k\|^2, \ \forall k \in N. \tag{2.9}$$

Proof. We prove this lemma by induction. For $k = 1$, it holds since $d_1 = -g_1$. Assume that the conclusion of this lemma holds for $k - 1$.

Subsequently, we prove that (2.9) holds for k. We obtain from (1.4) and the relationship $s_k = \alpha_k d_k$ that

$$d_{k-1}^T \lambda_{k-1} \ge \ y_{k-1}^T d_{k-1} > -(1-\sigma) g_{k-1}^T d_{k-1} > (1-\sigma)||g_{k-1}||^2 > 0. \tag{2.10}$$

For all $k > 1$, multiplying (2.6) by g_k^T, we have

$$g_k^T d_k = -||g_k||^2 + \beta_k^{DYT1} g_k^T d_{k-1} - \frac{g_k^T d_{k-1}}{d_{k-1}^T \lambda_{k-1}} g_k^T \lambda_{k-1} \qquad = -||g_k||^2 - \xi \frac{g_k^T s_{k-1}}{d_{k-1}^T \lambda_{k-1}} g_k^T d_{k-1} \le -||g_k||^2.$$

In order to ensure the global convergence of our method, we propose a new update strategy in constructing the search direction, presented by

$$d_k = \begin{cases} -g_k, & \Upsilon_k \|d_{k-1}\| \ge \mu \|g_k\|, \\ d_k^{DYT1}, & \Upsilon_k \|d_{k-1}\| < \mu \|g_k\|, \end{cases} \tag{2.11}$$

where

$$\Upsilon_k = \max\{\|g_k\| \cdot \|\lambda_{k-1}\|, \quad \xi |g_k^T s_{k-1}|\} \tag{2.12}$$

and μ is a positive constant.

For convenience, we call the corresponding method as DYT method (D denotes descent) here and formally state the steps of this method as follows.

Algorithm 2.2 (DYT1 method)**.**

Step 1. Given positive constants ε, μ, ξ, $\rho_1 < \sigma < 1$, and choose an initial point $x_1 \in R^n$, set $k = 1$.
Step 2. Test a criterion for stopping the iterations. If $\|g_k\| < \varepsilon$, then stop, otherwise calculate β_k^{DYT1} by (2.7) and d_k by (2.11).
Step 3. Determine the steplength α_k by the Wolfe conditions.
Step 4. Set $x_{k+1} = x_k + \alpha_k d_k$ and $k = k + 1$. Return to Step 2.

3. Global convergence of the DYT1 method

In this part, we come to show the convergence of our method. Throughout the paper, we assume, for the sake of contradiction, that there exists a constant $\epsilon > 0$ such that the relation $||g_k|| \geq \epsilon$ holds for all k, otherwise a stationary point has been found.

The following assumption and Zoutendijk condition is often needed in convergence analysis.

Assumption 3.1. The level set, defined by $\Omega = \{x \in R^n | f(x) \leq f(x_1)\}$ with x_1 to be the initial point, is bounded.

Theorem 3.2 ([34]). *Suppose that Assumption* 3.1 *hold. Consider the conjugate gradient method defined by* (1.2)*, where the search direction* d_k *satisfies* $g_k^T d_k < 0$ *and the steplength* α_k *satisfies the Wolfe conditions* (1.3) *and* (1.4)*, then* $\sum_{k\geq 1} \frac{(g_k^T d_k)^2}{||d_k||^2} < \infty$.

Theorem 3.3. *Suppose that Assumption* 3.1 *hold. Let* $\{x_k\}$ *and* $\{d_k\}$ *be generated by Algorithm* 2.2*. If there exists a constant* $\epsilon > 0$ *such that the relation* $||g_k|| \geq \epsilon$ *holds for all* $k \in N$*, then there exist a constant* $\overline{c} > 0$ *such that*

$$||d_k|| \leq \overline{c}||g_k|| \quad \forall k \in N. \tag{3.1}$$

Proof. It should be first noted that $d_k \neq 0, \forall k \in N$, otherwise, as in the proof of Lemma 2.1, we can see that $||g_k|| \leq ||d_k||$ holds.

Clearly, it suffices to consider the case where $\Upsilon_k \|d_{k-1}\| < \mu \|g_k\|$ is satisfied.

From the definition of β_k^{DYT1} in (2.7), we obtain that

$$\begin{aligned}
|\beta_k^{DYT1}| &= \left| \frac{g_k^T \lambda_{k-1}}{d_{k-1}^T \lambda_{k-1}} - \xi \frac{g_k^T s_{k-1}}{d_{k-1}^T \lambda_{k-1}} \right| \\
&\leq \frac{|g_k^T \lambda_{k-1}| + |\xi g_k^T s_{k-1}|}{d_{k-1}^T \lambda_{k-1}} \\
&\leq 2 \frac{\Upsilon_k}{d_{k-1}^T y_{k-1}} \\
&\leq 2\mu \frac{\|g_k\|}{||d_{k-1}||} \frac{1}{d_{k-1}^T y_{k-1}} \\
&\leq 2 \frac{\mu}{\epsilon^2 (1-\sigma)} \frac{\|g_k\|}{||d_{k-1}||},
\end{aligned}$$

where the third and last inequalities follows from (2.12) and(2.10), respectively.

The above inequality together with (2.11) gives

$$\begin{aligned}
||d_k|| &\leq ||g_k|| + |\beta_k^{DYT1}| \cdot ||d_{k-1}|| + \frac{||g_k||||d_{k-1}||}{d_{k-1}^T \lambda_{k-1}} ||\lambda_{k-1}|| \\
&\leq ||g_k|| + 2 \frac{\mu}{\epsilon^2(1-\sigma)} ||g_k|| + \Upsilon_k \frac{||d_{k-1}||}{d_{k-1}^T y_{k-1}} \\
&\leq ||g_k|| + 2 \frac{\mu}{\epsilon^2(1-\sigma)} ||g_k|| + \mu \frac{\|g_k\|}{||d_{k-1}||} \frac{||d_{k-1}||}{\epsilon^2(1-\sigma)} \\
&\leq ||g_k|| + 3 \frac{\mu}{\epsilon^2(1-\sigma)} ||g_k||
\end{aligned} \tag{3.2}$$

Letting $\overline{c} = 1 + 3\frac{\mu}{\epsilon^2(1-\sigma)}$, the conclusion is seen to hold.

In the end of this section, we state the convergence result of our method.

Theorem 3.4. *Suppose that Assumption* 3.1 *hold. Let* $\{x_k\}$ *be generated by the Algorithm* 2.2. *Then* $g_k = 0$ *for some* k *or* $\liminf_{k\to\infty} ||g_k|| = 0$.

Proof. We proceed by contradiction. If the conclusion does not hold, there exists a constant $\epsilon > 0$ *such that*

$$||g_k|| \geq \epsilon, \forall k \in N; \tag{3.3}$$

otherwise a stationary point has been found. We obtain from Theorem 3.2 that

$$\sum_{k\geq 1} (\overline{c})^{-2} ||g_k||^2 \leq \sum_{k\geq 1} \frac{||g_k||^4}{||d_k||^2} < \sum_{k\geq 1} \frac{(g_k^T d_k)^2}{||d_k||^2} < \infty, \tag{3.4}$$

where the first inequality follows from (3.1) *and the second one from* (2.9). *So we have* $\lim_{k\to\infty} ||g_k|| = 0$, *which contradicts* (3.3). *Therefore the conclusion holds.*

4. Generalization of the new technique

In this section, we intend to take the advantages of the results on the new versions of the YT method in Section 2 and the HZ method to construct new methods. For simplicity, we only give the search dierction d_k of these algorithms, the other steps are as same as Algorithm 2.2.

We first introduce a constant ζ and propose a new parameter for β_k:

$$\beta_k^{YTHZ} = \frac{g_k^T \lambda_{k-1}}{d_{k-1}^T \lambda_{k-1}} - \zeta \frac{\|\lambda_{k-1}\|^2}{(d_{k-1}^T \lambda_{k-1})^2} g_k^T d_{k-1}. \tag{4.1}$$

Algorithm 4.1 (DYT2 method)**.**

$$d_k = \begin{cases} -g_k, & \Gamma_k \|d_{k-1}\| \geq \mu \|g_k\|, \\ -g_k + \beta_k^{YTHZ} d_{k-1} - \dfrac{g_k^T d_{k-1}}{d_{k-1}^T \lambda_{k-1}} \lambda_{k-1}, & \Gamma_k \|d_{k-1}\| < \mu \|g_k\|, \end{cases} \tag{4.2}$$

Algorithm 4.2 (YT–HZ method)**.**

$$d_k = \begin{cases} -g_k, & \Gamma_k \|d_{k-1}\| \geq \mu \|g_k\|, \\ -g_k + \beta_k^{YTHZ} d_{k-1} & \Gamma_k \|d_{k-1}\| < \mu \|g_k\|, \end{cases} \tag{4.3}$$

In (4.2) and (4.3), where $\Gamma_k = \|g_k\| \cdot \|\lambda_{k-1}\|$, $\mu > 0$,and $\zeta > \dfrac{1}{4}$ are constant.

Analogously, we can establish the global convergence of these Algorithms.

Theorem 4.3. *Suppose that Assumption* 2.2 *hold. Let* β_k *be generated by* (4.1) *and* d_k *be generated by* (4.2) *and* (4.3). *Then the sufficient descent condition* (1.8) *holds. Furthermore, the Algorithms* 4.1 *and* 4.2 *converge in the sense that* $\liminf_{k\to\infty} ||g_k|| = 0$ *holds.*

Proof. We first consider the sufficient descent condition. For $k = 1$, it holds since $d_1 = -g_1$. Assume that the conclusion of this theorem holds for k.

If the negative gradient direction is used, then sufficient descent condition (2.9) holds. Furthermore, for all $k > 1$, multiplying (4.2) by g_{k+1}^T, we have

$$g_{k+1}^T d_{k+1} = -||g_{k+1}||^2 + \beta_{k+1}^{YTHZ} g_{k+1}^T d_k - \frac{g_{k+1}^T d_k}{d_k^T \lambda_k} g_{k+1}^T \lambda_k \quad = -||g_{k+1}||^2 - \zeta \frac{\|\lambda_k\|^2 (g_{k+1}^T d_k)^2}{(d_k^T \lambda_k)^2} \leq -||g_{k+1}||^2.$$

Also, we mainly consider the search direction in the YT–HZ method for $\mathit{\Gamma}_{k+1}\|d_k\| < \mu\|g_{k+1}\|$. We obtain form (4.3) and Cauchy inequatily that

$$\begin{aligned}
g_{k+1}^T d_{k+1} &= -||g_{k+1}||^2 + \beta_{k+1}^{YTHZ} g_{k+1}^T d_k \\
&= -||g_{k+1}||^2 - \frac{g_{k+1}^T \lambda_k}{d_k^T \lambda_k} g_{k+1}^T d_k - \zeta \frac{\|\lambda_k\|^2}{(d_k^T \lambda_k)^2}(g_{k+1}^T d_k)^2 \\
&\le -||g_{k+1}||^2 + \frac{||g_{k+1}||||\lambda_k||}{d_k^T \lambda_k} g_{k+1}^T d_k - \zeta \frac{\|\lambda_k\|^2}{(d_k^T \lambda_k)^2}(g_{k+1}^T d_k)^2 \\
&= -\left(1 - \frac{1}{4\zeta}\right) ||g_{k+1}||^2 - \left(\frac{||g_{k+1}||}{2\sqrt{\zeta}} - \sqrt{\zeta}\frac{\|\lambda_k\|}{(d_k^T \lambda_k)} g_{k+1}^T d_k\right)^2 \\
&\le -\left(1 - \frac{1}{4\zeta}\right) ||g_{k+1}||^2.
\end{aligned}$$

Subsequently, we establish the global convergence of the Algorithms 4.1 and 4.2. In the rest of this section, we assume there exists a constant $\epsilon > 0$ such that $||g_k|| > \epsilon$ for all k, otherwise a stationary point has been found. The main result can be proved in a very similar way to the proof of Algorithm 2.2. The only remarkable difference is at the point where we have to prove that there exists a positive constant M, such that

$$|\beta_k| \le M \frac{\|g_k\|}{||d_{k-1}||}. \tag{4.4}$$

From the definition of β_k^{DYHZ} in (4.1) we obtain that

$$\begin{aligned}
|\beta_k^{DYHZ}| &\le \frac{||g_k||||\lambda_{k-1}||}{d_{k-1}^T \lambda_{k-1}} + \zeta \left(\frac{||\lambda_{k-1}||}{d_{k-1}^T \lambda_{k-1}}\right)^2 ||g_k||||d_{k-1}|| \\
&\le \mu \frac{\|g_k\|}{||d_{k-1}||} \frac{1}{d_{k-1}^T y_{k-1}} + \zeta \left(\frac{\mu}{||d_{k-1}||}\right)^2 \frac{||g_k||||d_{k-1}||}{(d_{k-1}^T y_{k-1})^2} \\
&\le \left(\frac{\mu}{\epsilon^2(1-\sigma)} + \zeta \frac{\mu^2}{(1-\sigma)^2 \epsilon^4}\right) \frac{\|g_k\|}{||d_{k-1}||}.
\end{aligned}$$

Setting $M = \dfrac{\mu}{\epsilon^2(1-\sigma)} + \zeta \dfrac{\mu^2}{(1-\sigma)^2\epsilon^4}$, we finish the proof.

5. Numerical experiments

Finally, we report some numerical results on 73 nonlinear unconstrained test problems, 26 out of which are from the CUTEr collection established in [5]. For each test problem, the dimension n is set to 6000 and the Fortran expression of its function and gradient can be downloaded from Andrei's website[4]. There are 2 problems that are eliminated from the given problems above, that is "Extended Cliff" and "VARDIM", which lead an overflow error during the iteration.

Here we utilize the source code Fortran 77 on Hager's webpage[5]. In the line search procedure, the Wolfe conditions (1.3) and (1.4) were used with $\rho_1 = 10^{-4}, \sigma = 0.9$. The implementations are run on PC with 1.3 GHz CPU processor and 760 MB RAM memory.

We uniformly set $\rho = 10^{-6}$, $u = s_{k-1}$ in λ_{k-1} defined by (2.8). Furthermore, since too many restarts may reduce the speed of convergence, the parameter μ is set as large as possible such that the relation $\mathit{\Gamma}_k < \mu\|g_k\|$ holds for most of indices k. The parameter μ in our methods use exactly the same setting with $\mu = 10^{20}$. We stop the iteration if the inequality $||g_k|| < 10^{-6}$ is satisfied.

[4]http://www.ici.ro/camo/neculai/SCALCG/evalfg.for

[5]http://clas.ufl.edu/users/hager/papers/CG/Archive/

The whole numerical results, including the CPU time in seconds and the number of iterations, the number of function and gradient evaluations implementation for each method, can be found in:

http://shuxueyou.blog.sohu.com/307224137.html.

The following 6 CG methods are test, including

- 1. The HZ method [23]: the CG method with the truncation, that is, the parameter $\beta_k = \max\left\{\beta_k^{HZ}, \frac{-1}{||d_{k-1}|| \min\{\eta, ||g_{k-1}||\}}\right\}$, where $\eta = 0.1$.
- 2. The YT method: the CG method with the parameter β_k^{YT} defined by (2.1), where $u_k = s_{k-1}$, $\tau = 0.1$ and $\rho = 10^{-6}$ in (2.2).
- 3. The MYT method [42]: the three–term CG method with the search direction d_k^{MYT} defined by (2.4), where $u_k = s_{k-1}$, $\tau = 0.1$ and $\rho = 10^{-6}$.
- 4. The DYT1 method (Algorithm 2.2): the three–term CG method with the search direction d_k^{DYT1} defined by (2.11), where $\rho = 10^{-6}$, $u_k = s_{k-1}$ and $\xi = 0.1$.
- 5. The DYT2 method (Algorithm 4.1): the three–term CG method with the search direction d_k^{YTHZ} defined by (4.2), where $\rho = 10^{-6}$, $u_k = s_{k-1}$ and $\zeta = 0.1$.
- 6. The YT–HZ method (Algorithm 4.2): the CG method with the search direction d_k^{DYT2} defined by (4.3), where $\rho = 10^{-6}$, $u_k = s_{k-1}$ and $\zeta = 0.5$.

Further to demonstrate the efficiency of our methods, the use of profiles of Dolan and Moré [10] will present a wealth of information including efficiency and robustness. More analytically, the left side of the figure presents the percentage of test problem for which a method performs fastest, the right side gives the percentage of the test problems that are successfully solved. The top curve is the method that solved the most problems in a time that was within a factor ω of the best time.

Since there are too many lines collapsing at $\omega = 1$, it is difficult to see which algorithm was more efficient. As constructively suggested by a reviewer, to present a detailed numerical comparison, two different scales have been considered for ω-axis: $\omega \in [1, 2]$, to perceive what happens for the values of ω near to 1; and $\omega \in [1, 10]$, to perceive which is the trend for the large values of ω.

From Figure 1, it concludes that the most efficient algorithm in terms of number of iteration was DYT2, being the fastest for 56% of the problems, followed by DYT1, YT–HZ, HZ, MYT and YT that were the most efficient for respectively 45%, 42%, 40%, 40%, and 35% of the problems.

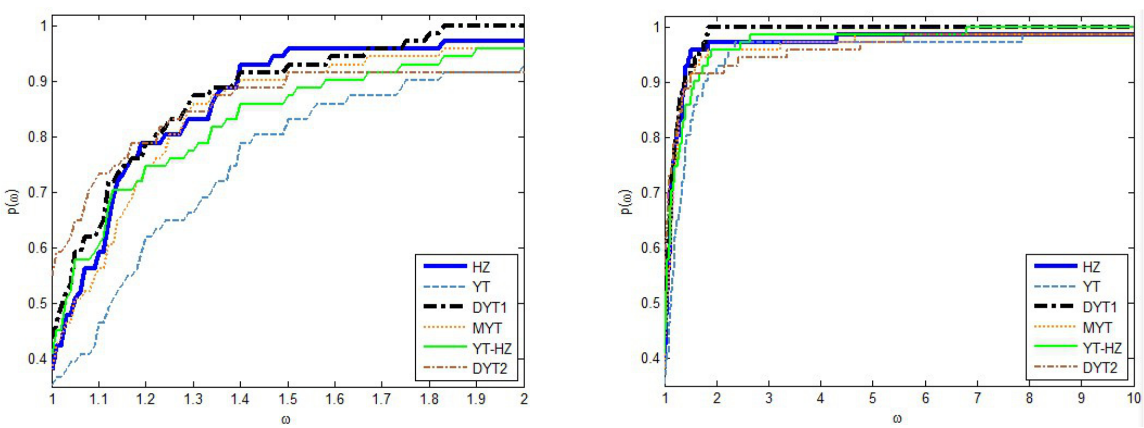

FIGURE 1. Performance profile of iterations.

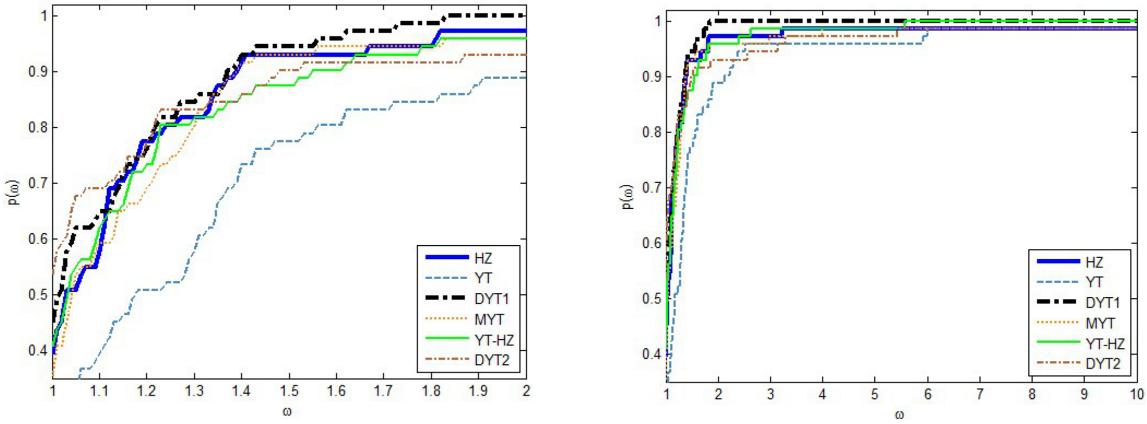

FIGURE 2. Performance profile of function evaluations.

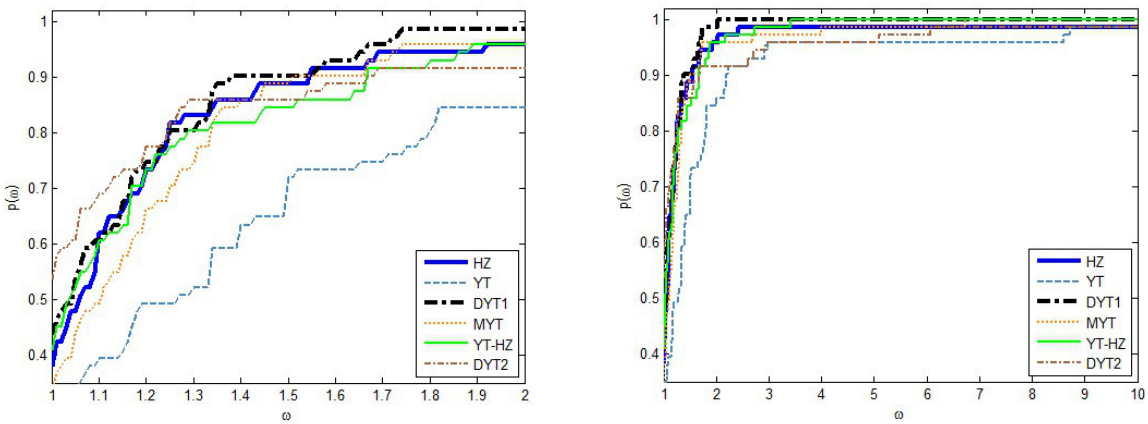

FIGURE 3. Performance profile of gradient evaluations.

In Figures 2 and 3, we compare the performance of our method with that of other methods based on the numbers of function evaluations and its gradient evaluations. As can be seen form these figures, the performance profile illustrate that the DYT2 method and the MYT method approximately solve 56% and 39% of the test problems with the least numbers of function evaluations and its gradient evaluations while the DYT1 method, the YT–HZ, and HZ method approximately solve about 44%, which almost have identical performance. Therefore, our methods outperforms slightly other method with respect to numbers of the function evaluations.

From Figures 1–3, we can observe that, for $\omega > 1.5$, our proposed and "DYT1" methods numerically outperforms with slight superiority to the other methods. There figures graphically illustrate that the curve of "DYT1" is always the top performer for almost all values of ω. However, the best performance, relative to the CPU time in Figure 4, is obtained by the "HZ" method, which solve 60% of the test problems with the least CPU time. The other methods have the second best performance in which it solves about 40% in the same situation. On the other hand, the right graphics on the right side of the above four figures illustrate that all algorithms can ultimately solve 100% of the test problems. Thus, the above methods have efficient roundness with respect to the iteration and CPU time.

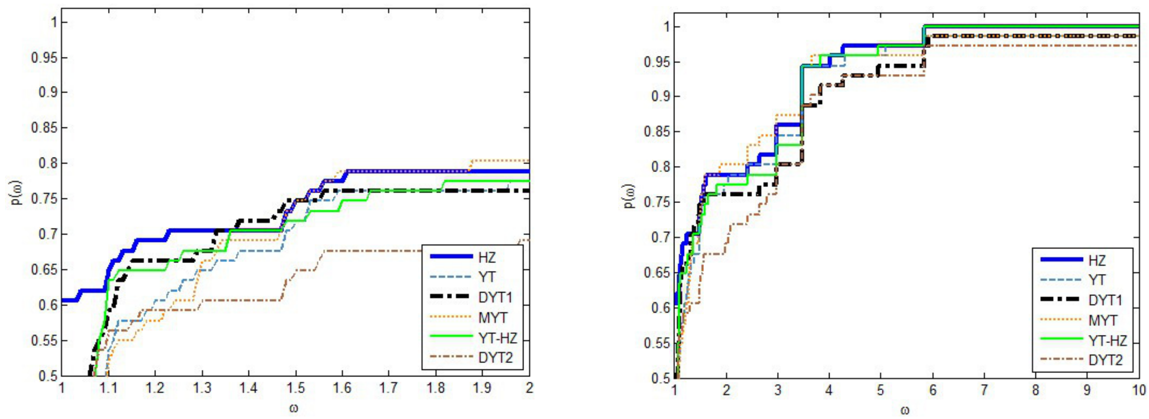

FIGURE 4. Performance profile of CPU time.

Numerical comparisons of the HZ method, the YT method, the MYT method and our proposed methods show that the DYT1 method, which requires fewer iterations, but needs slightly more CPU time during the iteration, is efficient for solving these test problems. The possible reason is that our presented search direction needs slightly more computational cost for inner-products than that of the other methods.

6. Conclusion

In this paper, we proposed some modifications to the Yabe–Takano method, in which the search direction always satisfy the sufficient descent condition. Under mild condition, we can establish the global convergence of our proposed method even if the objective function is nonconvex. Numerical results show that our proposed method is efficient for solving the optimization problems.

Acknowledgements. We are grateful to the anonymous referees and editor for their useful comments that help us improve the presentation of this paper. We would like to thank Professors W.W. Hager and H. Zhang for their CG_DESCENT code for numerical comparison. This work is supported by the Natural Science Foundation of China (11361001, 11601012), Research Foundation of Beifang University of Nationalities (2016XYZ06, 2014XBZ09, 2013XYZ028).

References

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

[2] N. Andrei, A simple three–term conjugate gradient algorithm for unconstrained optimization. *J. Comput. Appl. Math.* **241** (2013) 19–29.

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

[4] S. Babaie-Kafaki, R. Ghanbari and N. Mahdavi-Amiri, Two new conjugate gradient methods based on modified secant equations. *J. Comput. Appl. Math.* **234** (2010) 1374–1386.

[5] K.E. Bongartz, A.R. Connand and N.I.M. Gould, P.L. Toint, CUTE: constrained and unconstrained testing environments, *ACM Trans. Math. Software* **21** (1995) 123–160.

[6] W.Y. Cheng and D.H. Li, An active set modified Polak–Ribière–Polyak method for large–scale nonlinear bound constrained optimization. *J. Optim. Theory Appl.* **155** (2012) 1084–1094.

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

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

[9] Y. Dai and L. Liao. New conjugacy conditions and related nonlinear conjugate gradient methods. *Appl. Math. Optim.* **43** (2001) 87–101.
[10] E.D. Dolan and J.J. Moré, Benchmarking optimization software with performance profiles. *Math. Program. Ser. A* **91** (2002) 201–213.
[11] X. Dong, Comment on "A new three–term conjugate gradient method for unconstrained problem". *Numer. Algor.* **72** (2016) 173–179.
[12] X. Dong, H. Liu and Y. He, A self–adjusting conjugate gradient method with sufficient descent condition and conjugacy condition. *J. Optim. Theory Appl.* **165** (2015) 225–241.
[13] X. Dong, H. Liu and Y. He, New version of three–term conjugate gradient method based on spectral scaling conjugacy condition that generates descent direction. *Appl.Math. Comput.* **239** (2015) 1606–1617.
[14] X. Dong, H. Liu, Y. He, S. Babaie-Kafakid and R. Ghanbari, A new three-term conjugate gradient method with descent direction for unconstrained optimization. *Math. Model. Anal.* (to appear).
[15] X. Dong, H. Liu, Y. He, Y. Xu and X. Yang, Some nonlinear conjugate gradient methods with sufficient descent condition and global convergence. *Optim. Lett.* **9** (2015) 1421–1432.
[16] X. Dong, H. Liu, Y. He and X. Yang, A modified Hestenes–Stiefel conjugate gradient method with sufficient descent condition and conjugacy condition. *J. Comput. Appl. Math.* **281** (2015) 239–249.
[17] R. Fletcher, Practical methods of optimization, John Wiley and Sons (2013).
[18] R. Fletcher and C. Reeves, Function minimization by conjugate gradients. *Comput. J.* **7** (1964) 149–154.
[19] J.A. Ford and I.A. Moghrabi, Multi-step quasi-Newton methods for optimization. *J. Comput. Appl. Math.* **50** (1994) 305–323.
[20] J.A. Ford and I.A. Moghrabi, Using function–values in Multi-step quasi-Newton methods. *J. Comput. Appl. Math.* **66** (1996) 201–211.
[21] J.A. Ford, Y. Narushima, and H. Yabe, Multi-step nonlinear conjugate gradient methods for unconstrained minimization. *Comput. Optim. Appl.* **40** (2008) 191–216.
[22] J.C. Gilbert and J. Nocedal, Global convergence properties of conjugate gradient methods for optimization. *SIAM J. Optim.* **2** (1992) 21–42.
[23] W.W. Hager and H. Zhang, A new conjugate gradient method with guaranteed descent and an efficient line search. *SIAM J. Optim.* **16** (2005) 170–192.
[24] W.W. Hager and H. Zhang, A survey of nonlinear conjugate gradient methods. *Pac. J. Optim.* **2** (2006) 35–58.
[25] M.R. Hestenes and E. Stiefel, Methods of conjugate gradients for solving linear systems. *J. Res. Natl. Bur. Stand.* **49** (1952) 409–436.
[26] D.H. Li and M. Fukushima, A modified BFGS method and its global convergence in nonconvex minimization. *J. Comput. Appl. Math.* **129** (2001) 15–35.
[27] D.H. Li and M. Fukushima, On the global convergence of the BFGSmethod for nonconvex unconstrained optimization problems. *SIAM J. Optim.* **11** (2001) 1054–1064.
[28] G. Li, C. Tang, and Z.Wei, New conjugacy condition and related new conjugate gradient methods for unconstrained optimization. *J. Comput. Appl. Math.* **202** (2007) 523–539.
[29] Y.L. Liu and C.S. Storey, Efficient generalized conjugate gradient algorithms. *Part 1: Theory, J. Optim. Theory Appl.* **69** (1991) 129–137.
[30] Y. Narushima and H. Yabe. Conjugate gradient methods based on secant conditions that generate descent search directions for unconstrained optimization. *J. Comput. Appl. Math.* **236** (2012) 4303–4317.
[31] Y. Narushima, H. Yabe and J.A. Ford, A three-term conjugate gradient method with sufficient descent property for unconstrained optimization. *SIAM J. Optim.* **21** (2011) 212–230.
[32] Z. Wei, G. Li and L. Qi, New quasi-Newton methods for unconstrained optimization problems. *Appl. Math. Comput.* **175** (2006) 1156–1188.
[33] F. Wen, Z. He, Z. Dai and X. Yang, Characteristics of investors' risk preference for stock markets. *Econ. Comput. Econ. Cybern. Stud. Res.* **3** (2014) 235–254.
[34] P. Wolfe, Convergence conditions for ascent methods. *SIAM Rev.* **11** (1969) 226–235.
[35] H. Yabe and M. Takano, Global convergence properties of nonlinear conjugate gradient methods with modified secant condition. *Comput. Optim. Appl.* **28** (2004) 203–225.
[36] G. Yu and L. Guan, A descent sprctral conjugate gradient methods for inpulse noise removal. *Appl. Math. Lett.* **2** (2010) 555–560.
[37] Y. Yuan, A modified BFGS algorithm for unconstrained optimization. *IMA J. Numer. Anal.* **11** (1991) 325–332.
[38] Y. Yuan and R.H. Byrd, Non-quasi-Newton updates for unconstrained optimization. *J. Comput. Math.* **13** (1995) 95–107.
[39] J.Z. Zhang, N.Y. Deng and L.H. Chen, New quasi-Newton equation and related methods for unconstrained optimization, *J. Optim. Theory Appl.* **102** (1999) 147–167.
[40] J. Zhang and C. Xu, Properties and numerical performance of quasi-Newton methods with modified quasi-Newton equations. *J. Comput. Appl. Math.* **137** (2001) 269–278.
[41] L. Zhang, W. Zhou and D. Li, A descent modified Polak–Ribière–Polyak conjugate gradient method and its global convergence. *IMA J. Numer. Anal.* **6** (2006) 629–640.
[42] X. Zheng, H. Liu and A. Lu, Sufficient descent conjugate gradient methods for large-scale optimization problems. *Int. J. Comput Math.* **88** (2011) 3436–3447.
[43] W. Zhou and L. Zhang, A nonlinear conjugate gradient method based on the MBFGS secant condition. *Optim. Methods Softw.* **21** (2006) 707–714.