Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 21, No. 1, pp. 107–123, 2025

A NEW THREE-TERM CONJUGATE GRADIENT METHOD WITH ADAPTIVELY ALTERNATIVE TWO-SIDE APPROXIMATING STRATEGY THAT GENERATES DESCENT SEARCH DIRECTION*

Xiaoliang Dong† and Deren Han

Abstract: A new three–term conjugate gradient algorithm is developed, in which the involved search direction satisfies the sufficient descent condition at each iteration under Wolfe conditions. Different from the existent methods, a dynamical two–side approximating mechanism is proposed, which adaptively adjusts the relative weight between sufficient descent condition and making fully use of curvature information of objective functions. To some extent, the above strategy meaningfully exploits Hessian approximation of the objective function and therefore increases the efficiency of the algorithm in practical computation. Under mild conditions, we prove that the presented method converges globally for general objective functions. Numerical results are reported, which illustrate that the proposed algorithm is practically encouraging.

Key words: *three–term conjugate gradient method, adaptively alternative two-side approximating strategy, sufficient descent condition, conjugacy condition, Rayleigh quotient*

Mathematics Subject Classification: *65K05, 90C53*

1 Introduction

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable and nonlinear function, and its gradient abbreviated as $g(x) = \nabla f(x)$ is available. Here, we consider the following unconstrained optimization problem:

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

Conjugate gradient (CG) methods have drawn considerable attention due to their effectiveness and simplicity, especially for solving large–scale unconstrained optimization. For a given initial point $x_1 \in \mathbb{R}^n$, the method generates the iterates via the recursion as follows:

$$x_{k+1} = x_k + s_k, s_k = \alpha_k d_k, \quad \forall k \geq 1, \tag{1.2}$$

where the search direction d_k takes the following form

$$d_1 = -g_1, d_{k+1} = -g_{k+1} + \beta_{k+1} d_k, \quad \forall k \geq 1, \tag{1.3}$$

*This work is supported by the Natural Science Basic Research Plan in Shaanxi Province of China (2021JM-398), the National Natural Science Foundation of China (12131004,11601012).

†Corresponding author

DOI: https://doi.org/10.61208/pjo-2023-028

and $\alpha_k > 0$ is a steplength generally determined by the well-known Wolfe conditions with constants $0 < \rho < \sigma < 1$, that is,

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \rho \alpha_k g_k^T d_k, \tag{1.4}$$

$$g\left(x_k + \alpha_k d_k\right)^T d_k \geq \sigma g_k^T d_k. \tag{1.5}$$

In (1.3), β_k is a judiciously constructed CG parameter, which brought diversity of algorithms, and with it quite diverse computational behavior and convergence results. Among the CG methods, the Hestenes and Stiefel [17] (HS) scheme has been the focus of recent work, not only for its essential history importance and impressive computational performance, but also its theoretical property of satisfying conjugacy condition, namely,

$$d_k^T y_{k-1} = 0, \tag{1.6}$$

where $y_{k-1} = g_k - g_{k-1}$. On the other hand, in [14], Gilbert and Nocedal established the convergence of the HS+ method for the general objective functions, in which

$$\beta_{k+1}^{HS+} = \frac{(g_{k+1}^T y_k)^+}{d_k^T y_k}. \tag{1.7}$$

Here, $b^+ = \max\{b, 0\}$, where b is a constant.

Recently, various variations and great improvements of the HS method have been made by designing sophisticated techniques for guaranteeing the Dai and Liao (DL) conjugacy condition and the sufficient descent condition:

- As an extension of (1.6), Dai and Liao [10] proposed the following conjugacy condition:

$$d_k^T y_{k-1} = -t g_k^T s_{k-1}, \tag{1.8}$$

  where $t > 0$ is a constant. Based on condition number analysis of iteration matrix of search direction, Andrei [3] and Babaie–Kafaki [4, 5] gave several choices of optimal parameter t, which fuels the boom in the open problem posed by Andrei [2], i.e., what is the best conjugacy condition? For more details, we refer to the excellent survey [6].

- In some convergence analyses, the sufficient descent condition is required, namely,

$$d_k^T g_k \leq -c\|g_k\|^2, \ \forall k \in \mathbb{N}, \quad for \ \ some \ \ c > 0. \tag{1.9}$$

  It is worth mentioning that Hager and Zhang (HZ), and Dai and Kou (DK) pioneered new technique to force the presented search direction to satisfy automatically (1.9). Here, we call the involved methods the HZ and DK methods for short. Meanwhile, the resulting CG_DESCENT [15] and CGOPT [11] are public domain software packages.

Our attention will be on the three–term conjugate gradient (TTCG) method. As a natural extension of the standard CG method, it has received much study (see [1, 8, 19, 21] and references therein), which not only enhances the freedom and flexibility of the selection of parameters but also substantially embeds some favorable properties in the search direction.

Following a modified DL (MDL) method [7], we construct another search direction by combining the advantages of the CGOPT method and quasi-Newton method to compensate the loss of second-order curvature information of $f(x)$ being caused by slightly improper truncation in the MDL method. This can be viewed as the inheritance and development of

properties of the MDL method in the sense that the twin search directions are computationally comparable, while the alternatively generated search direction makes the corresponding optimization behavior close to that of sequential the current one as much as possible.

We structure the remainder of the paper as follows. In Sect.2, we present formally the modified MDL method and investigate the sufficient descent condition. In Sect.3, we establish the convergence analysis of the above method. In Sect.4, a variant of the proposed method is given. In Sect.5, numerical results are report to demonstrate the efficiency of the proposed methods. The paper is concluded in the last section.

2 A New CG Algorithm

Recently, Babaie–Kafaki and Ghanbari [8] proposed the MDL method, in which the main contribution is to combine the DL method and the TTCG algorithm framework in such a way that the search direction can automatically satisfy the standard Newton equation. However, theoretically, the above method lacks significant descent property. To circumvent this difficult, a constant $\xi > 0$ was introduced and an improved search direction was proposed:

$$d_{k+1} = -g_{k+1} + \left(\frac{g_{k+1}^T y_k}{d_k^T y_k} - \varpi_k^{MDL} \frac{g_{k+1}^T s_k}{d_k^T y_k} \right) d_k - \theta_{k+1} y_k, \tag{2.1}$$

where

$$\varpi_{k+1}^{MDL} = \max \left\{ \xi, 1 - \frac{\|y_k\|^2}{s_k^T y_k} \right\}, \tag{2.2}$$

$$\theta_{k+1} = \frac{g_{k+1}^T d_k}{d_k^T y_k}. \tag{2.3}$$

It can be seen that the above method converges for uniformly convex objective functions, under the Wolfe conditions, with a promising computational behavior. Meanwhile, the truncation scheme (2.2) is slightly inefficient in that an excessive use of the value of ξ may cause the information being close to the Newton direction to be regretfully neglected.

However, what one loses on the swings, he gets back on the roundabouts. General iterative schemes, which are usually based on a quadratic model have been successfully in solving (1.1). Consequently, to adequately utilize curvature information of $f(x)$ plays a significant role in accelerating the iterations. To compensate the loss of second-order curvature information, another search direction that most closely approximates that of the quasi-Newton method is promptly supplemented in a proper way that acceleration of the whole of iteration would be anticipated.

Formally, we obtain the leading search directions as follows

$$d_{k+1}^{main} = \begin{cases} d_{k+1}^{main1}, & if \ \ \dfrac{s_k^T y_k}{\|y_k\|^2} \geq \dfrac{||g_{k+1}||^2}{\varepsilon^2}, \\ d_{k+1}^{main2}, & otherwise, \end{cases} \tag{2.4}$$

in which $\varepsilon > 0$ is an acceptable tolerance of the norm of the unconstrained stationary point.

In (2.4), the search direction d_{k+1}^{main} consists of twin sub-directions, given by,

$$d_{k+1}^{main1} = -g_{k+1} + \beta_{k+1}^{DK+} d_k + \tau_{k+1}^{+} y_k, \tag{2.5}$$

$$d_{k+1}^{main2} = -g_{k+1} + \beta_{k+1}^{MDL+} d_k - \theta_{k+1}^{+} y_k, \tag{2.6}$$

where involved parameters are given as follows:

$$\beta_{k+1}^{DK+} = \beta_{k+1}^{HS+} - \frac{\|y_k\|^2}{(d_k^T y_k)^2}(g_{k+1}^T d_k)^+, \tag{2.7}$$

$$\tau_{k+1}^{+} = \left(1 - \frac{s_k^T y_k}{||y_k||^2}\right)\frac{(g_{k+1}^T d_k)^+}{d_k^T y_k}, \tag{2.8}$$

$$\beta_{k+1}^{MDL+} = \beta_{k+1}^{HS+} - \left(1 - \frac{\|y_k\|^2}{s_k^T y_k}\right)\frac{(g_{k+1}^T s_k)^+}{d_k^T y_k}, \tag{2.9}$$

$$\theta_{k+1}^{MDL+} = \frac{(g_{k+1}^T d_k)^+}{d_k^T y_k}. \tag{2.10}$$

Remark 2.1. The search direction of the TTCG method d_{k+1} is ususlly generated by a linear combination of $-g_{k+1}$, d_k and y_k. Based on the Broyden-Fletcher-Goldfarb-Shanno (BFGS) update in the sense that the updated matrix $H_k = I$, a search direction is derived, which can be viewed as that of a class of four–term extension of the DK method, given by

$$d_{k+1} = \underbrace{-g_{k+1} + \beta_{k+1}^{DK} d_k}_{d_{k+1}^{DK}} + \Lambda_{k+1} y_k - \Lambda_{k+1} s_k. \tag{2.11}$$

In (2.11), $\beta_{k+1}^{DK} = \dfrac{g_{k+1}^T y_k}{d_k^T y_k} - \dfrac{g_{k+1}^T d_k ||y_k||^2}{(d_k^T y_k)^2}$ and $\Lambda_{k+1} = \dfrac{g_{k+1}^T s_k}{s_k^T y_k}$.

The simple deletion of the last term in (2.11) leads to another TTCG search direction, say, $\tilde{d}_{k+1}$. As a remedy, $\tilde{d}_{k+1}$ satisfies the quasi–Newton equation $H_{k+1} y_k = s_k$, provided that Λ_{k+1} is replaced by

$$\tau_{k+1} = \left(1 - \frac{s_k^T y_k}{||y_k||^2}\right)\frac{g_{k+1}^T d_k}{d_k^T y_k}. \tag{2.12}$$

The fact above partially accounts for the motivation of the search direction d_{k+1}^{main1}.

Remark 2.2. Notice that d_{k+1}^{main1} and d_{k+1}^{main2} are essentially designed based on an adaptive switch from quasi-Newton equation $H_{k+1} y_k = s_k$ to "pure" conjugacy condition (1.6) when $g_{k+1}^T d_k \le 0$. Furthermore, we combine the most recently observed information about the objective function with the existing knowledge of second-order curvature information embedded in the alternative Hessian approximation as much as possible. Specifically, if the current Hessian matrix "incorrectly" abandons the curvature in the objective function, and if this bad estimate may slow down the iteration, then the alternative Hessian approximation will tend to correct itself within a few steps. Perhaps, it would be more appropriate to call it "adaptively alternative two-sided approximating strategy".

Then, the AMDL method can be described, where A stands for "approximating".

The search direction of the BFGS quasi–Newton method is given by $d_{k+1} = -H_{k+1} g_{k+1}$, updated by the following iterative formula from the previous approximation H_k of $\nabla^2 f(x_{k+1})^{-1}$: $H_{k+1} = H_k + \left(1 + \dfrac{y_k^T H_k y_k}{s_k^T y_k}\right)\dfrac{s_k s_k^T}{s_k^T y_k} - \dfrac{s_k y_k^T H_k + H_k y_k s_k^T}{s_k^T y_k}$.

Algorithm 2.3. (A modified DL-type TTCG algorithm with adaptively alternative two-side approximating strategy)
Step 1. Given positive constants ε, ε_1, and $\rho < \sigma < 1$. Choose an initial point $x_1 \in \mathbb{R}^n$ and set $d_1 = -g_1$ and $k = 1$.
Step 2. Determine a steplength α_k satisfying the Wolfe conditions (1.4) and (1.5).
Step 3. Let $x_{k+1} = x_k + \alpha_k d_k$ and calculate g_{k+1}. **If** $\|g_{k+1}\| < \varepsilon$, **then stop**.
Step 4. If $g_{k+1}^T y_k \leq \varepsilon_1$, **then** set $d_{k+1} = -g_{k+1}$ and $k = k + 1$, and **goto** Step 2.
Step 5. If $g_{k+1}^T y_k > \varepsilon_1$, **then** compute scalars β_{k+1}^{DK+}, τ_{k+1}^{+}, β_{k+1}^{MDL+} and θ_{k+1}^{MDL+} by (2.7), (2.8), (2.9) and (2.10). Finally, compute the search direction d_{k+1} as follows:

$$d_{k+1} = \begin{cases} -g_{k+1} + \beta_{k+1}^{HS}, & if \quad k \in K_1, \\ -g_{k+1} + \eta_{k+1} d_k, & if \quad k \in K_2, \\ d_{k+1}^{main}, & if \quad k \in K_3, \end{cases} \tag{2.13}$$

where d_{k+1}^{main} is defined by (2.4) and

$$\eta_{k+1} = \frac{-1}{||d_k|| \min\{\eta, ||g_{k+1}||\}}. \tag{2.14}$$

Also, the index sets K_1, K_2 and K_3 in (2.13) are individually presented by

$$K_1 = \{k | k \in \mathbb{N} | g_{k+1}^T y_k > \varepsilon_1, g_{k+1}^T d_k \leq 0\}. \tag{2.15}$$

$$K_2 = \{k | k \in \mathbb{N} | g_{k+1}^T y_k > \varepsilon_1, g_{k+1}^T d_k > 0, \beta_{k+1}^{DK+} \leq \eta_{k+1} \quad or \quad \beta_{k+1}^{MDL+} \leq \eta_{k+1}\}, \tag{2.16}$$

$$K_3 = \{k | k \in \mathbb{N} | g_{k+1}^T y_k > \varepsilon_1, g_{k+1}^T d_k > 0, \beta_{k+1}^{DK+} > \eta_{k+1} \quad and \quad \beta_{k+1}^{MDL+} > \eta_{k+1}\}. \tag{2.17}$$

Set $k = k + 1$ and **goto** Step 2.

Remark 2.4. As commented by Dai and Kou [11] and Kou et al. [18], the method with nonnegative β_k, firstly proposed by Powell [20], processes attractive properties of establishing global convergence for the general objective functions and preventing effectively jamming phenomenon from occurring. We employ the search direction of the HS+ method to invoke the restarting strategy. Also, we cautiously replace the restarting condition $g_{k+1}^T y_k \leq 0$ by a variant, i.e., $g_{k+1}^T y_k \leq \varepsilon_1$ to avoid possible "divide by zero" occurs in the scalar τ_{k+1}^{+}.

Remark 2.5. If $||g_{k+1}|| < \varepsilon$, then we get from Step 3 of the Algorithm 2.3 that the iterations stop, and so, the condition $\dfrac{||g_{k+1}||^2}{\varepsilon^2} > 1$ always hold provided that d_{k+1}^{main} is used. The above basic fact will fascinate the proof of the following lemma.

Lemma 2.6. *Suppose that the steplength α_k satisfies the Wolfe conditions. Then the search directions $\{d_k\}$ of Algorithm 2.3 satisfy* (1.9) *with $c = 1$.*

Proof. We provide a proof by induction. The basis of induction is verified by $k = 1$. Suppose that (1.9) holds for k, that is, $d_k^T g_k \leq -\|g_k\|^2$. What remains to do is to show the conclusion holds for $k + 1$, which concerned with the following cases.

Case (i) If $g_{k+1}^T y_k \leq \varepsilon_1$, then the conclusion holds clearly.

Case (ii) If $k \in K_1$, then the method reduces to the HS method. We get from (1.5) that

$$0 < -(1 - \sigma) g_k^T d_k < d_k^T y_k. \tag{2.18}$$

It is obviously seen that the conclusion follows immediately since $\eta_{k+1} < 0 < \beta_{k+1}^{HS}$.

Case (iii) Consider the truncation form of the search direction, i.e., $k \in K_2$. By direct computations, we have $g_{k+1}^T d_{k+1} \leq -||g_{k+1}||^2$.

Case (iv) Last but not least, if $k \in K_3$, then the rest of the proof falls naturally within two cases:

First, if $\frac{s_k^T y_k}{\|y_k\|^2} \geq \frac{||g_{k+1}||^2}{\varepsilon^2}$, then

$$\tau_{k+1}^+ = \left(1 - \frac{s_k^T y_k}{||y_k||^2}\right) \frac{(g_{k+1}^T d_k)^+}{d_k^T y_k} \leq \left(1 - \frac{||g_{k+1}||^2}{\varepsilon^2}\right) \frac{(g_{k+1}^T d_k)^+}{d_k^T y_k} \leq 0, \tag{2.19}$$

which together with (2.5) and $g_{k+1}^T y_k > \varepsilon_1$ gives

$$\begin{aligned} g_{k+1}^T d_{k+1}^{main1} &= -||g_{k+1}||^2 + \beta_{k+1}^{DK+} g_{k+1}^T d_k + \tau_{k+1}^+ g_{k+1}^T y_k \\ &\leq -||g_{k+1}||^2 + \beta_{k+1}^{DK} g_{k+1}^T d_k \\ &\leq -\frac{3}{4}||g_{k+1}||^2. \end{aligned} \tag{2.20}$$

The last inequality is seen from Lemma 2.6, please see [11].

Second, taking inner product on both sides of (2.6) with g_{k+1}^T, we obtain that

$$\begin{aligned} g_{k+1}^T d_{k+1}^{main2} &= -||g_{k+1}||^2 + \beta_{k+1}^{MDL+} g_{k+1}^T d_k - \theta_{k+1}^{MDL+} g_{k+1}^T y_k \\ &= -||g_{k+1}||^2 - \left(1 - \frac{\|y_k\|^2}{s_k^T y_k}\right) \frac{(g_{k+1}^T s_k)^+}{d_k^T y_k} g_{k+1}^T d_k \\ &\leq -||g_{k+1}||^2 - (1 - \frac{\varepsilon^2}{||g_{k+1}||^2}) \cdot \frac{\alpha_k}{d_k^T y_k} (g_{k+1}^T d_k)^2 \\ &\leq -||g_{k+1}||^2. \end{aligned} \tag{2.21}$$

Based on the discussion above, we set $c = 1$ to finish the proof. □

3 Convergence Analysis

In this section, we study the convergence analysis of the presented algorithm for the general functions in the sense that $\liminf_{k\to\infty} ||g_k|| = 0$. We assume that $g_k \neq 0$, otherwise, a stationary point has been obtained. Thus, we assume that there exists a positive constant ε such that

$$||g_k|| \geq \varepsilon, \ \forall k \in N. \tag{3.1}$$

Meanwhile, the following regular assumptions are commonly used to analyze the global convergence of the CG methods.

Assumption 3.1. (A1): The level set $\Omega = \{x \in \mathbb{R}^n | f(x) \leq f(x_1)\}$ is bounded; (A2): In some neighborhood Ω_0 of Ω, the objective function f is continuously differentiable and its gradient g is Lipschitz continuous, namely, there exists a constant $L > 0$ such that $||g(x) - g(y)|| \leq L||x - y||, \forall x, y \in \Omega_0$.

The assumptions imply there exist positive constants B and γ, such that $||x|| \leq B, \forall x \in L$, and $||g(x)|| \leq \gamma$, for all $x \in \Omega_0$.

Lemma 3.1. *Suppose that Assumptions 3.1 hold. Let $\{x_k\}$ be generated by Algorithm 2.3. If* (3.1) *holds, then there exist positive constants C_1 and M such that*

$$|\beta_k| \leq C_1||s_{k-1}||, \quad ||p_k|| \leq M, \tag{3.2}$$

where

$$p_k = \begin{cases} -g_k, & if\ k \in K_1 \quad or \quad g_k^T y_{k-1} \leq \varepsilon_1, \\ -g_k + \eta_k d_{k-1}, & if\ k \in K_2, \\ -g_k + \left(\beta_k^{MDL+}\right)^- d_{k-1} - \theta_k^{MDL+} y_{k-1}, & if\ k \in K_3 \quad \left(\beta_k = \beta_k^{MDL+}\right), \\ -g_k + \left(\beta_k^{DK+}\right)^- d_{k-1} + \tau_k^+ y_{k-1}, & if\ k \in K_3 \quad \left(\beta_k = \beta_k^{DK+}\right). \end{cases} \tag{3.3}$$

Proof. We estimate a bound for β_k. First, we should state the following fact. Both CG parameters β_k^{MDL+} and β_k^{DK+} reduce to β_k^{HS+} provided that $g_{k+1}^T y_k > \varepsilon_1$ and $g_{k+1}^T d_k \leq 0$ are satisfied simultaneously. So, for future use, we readily get the following relationships

$$\theta_k^{MDL+} = \frac{\left(g_k^T d_{k-1}\right)^+}{d_{k-1}^T (g_k - g_{k-1})} \in [0, 1). \tag{3.4}$$

Before we give the bounds for β_k^{MDL+} and β_k^{DK+}, we can drop the superscript symbol "+" in their expressions (2.7) and (2.9), and obtain that:

$$\begin{aligned} |\beta_k^{MDL+}| &= \left| \beta_k^{HS} + \left(1 - \frac{\|y_{k-1}\|^2}{s_{k-1}^T y_{k-1}}\right) \cdot \frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}} \right| \\ &\leq \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} + \left(1 + \frac{\|y_{k-1}\|^2}{s_{k-1}^T y_{k-1}}\right) \cdot \frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}} \\ &\leq \frac{(1+L)||g_k|| \cdot ||s_{k-1}||}{d_{k-1}^T y_{k-1}} + \frac{\|y_{k-1}\|^2}{d_{k-1}^T y_{k-1}} \cdot \frac{g_k^T d_{k-1}}{d_{k-1}^T y_{k-1}} \\ &\leq \frac{(1+L)||g_k|| \cdot ||s_{k-1}||}{d_{k-1}^T y_{k-1}} + L^2 \frac{\|s_{k-1}\|^2}{d_{k-1}^T y_{k-1}} \\ &\leq \frac{(L+1)\gamma + 2BL^2}{(1-\sigma)\varepsilon^2} ||s_{k-1}||. \end{aligned} \tag{3.5}$$

Analogously, we get that $|\beta_k^{DK+}| \leq \dfrac{(2BL^2 + L)\gamma}{(1-\sigma)\varepsilon^2} ||s_{k-1}||$.

Set $C_1 = \dfrac{(L+1)\gamma + 2BL^2}{(1-\sigma)\varepsilon^2}$, and the desired conclusion $|\beta_k| \leq C_1 ||s_{k-1}||$ is satisfied.

Subsequently, we estimate the upper bound for p_k. We consider the case where $g_k^T d_{k-1} > 0$ and $g_k^T y_{k-1} > \varepsilon_1$ and $\beta = \beta_k^{MDL+}$. We have from (3.3), (3.4) and (2.9) that

$$\begin{aligned} ||p_k|| &\leq ||g_k|| + |\left(\beta_k^{MDL+}\right)^-| \cdot ||d_{k-1}|| + |\theta_k^{MDL+}| \cdot ||y_{k-1}|| \\ &= ||g_k|| - \min\{\beta_k^{MDL+}, 0\} ||d_{k-1}|| + \frac{(g_k^T d_{k-1})^+}{d_{k-1}^T y_{k-1}} \cdot ||g_k - g_{k-1}|| \\ &\leq ||g_k|| - \eta_k ||d_{k-1}|| + ||g_k - g_{k-1}|| \\ &= 3\gamma + \frac{1}{||d_{k-1}|| \min\{\eta, ||g_{k-1}||\}} ||d_{k-1}|| \\ &\leq 3\gamma + \frac{1}{\min\{\eta, \varepsilon\}}. \end{aligned} \tag{3.6}$$

We now insert the Lipschitz estimate (1.5) for y_{k-1} into the expression (2.8) to get:

$$|\tau_k^+| \leq \left| 1 - \frac{s_{k-1}^T y_{k-1}}{\|y_{k-1}\|^2} \right| \leq 1 + \frac{s_{k-1}^T y_{k-1}}{\|y_{k-1}\|^2} \leq 1 + 2B\frac{\gamma}{\varepsilon_1} \triangleq C_\tau, \tag{3.7}$$

where the lase inequality comes from the fact that $\frac{1}{||y_{k-1}||} < \frac{||g_k||}{\varepsilon_1} < \frac{\gamma}{\varepsilon_1}$.

By an analogous philosophy, we can deduce that

$$|p_k| \leq \left(1 + 4B\frac{\varepsilon_1}{\epsilon}\right)\gamma + \frac{1}{\min\{\eta, \varepsilon\}} \tag{3.8}$$

for $\beta_k = \beta_k^{DK+}$.

Assertion (3.2) follows directly from inequalities (3.6) and (3.8). □

To proceed, we introduce the following lemma, called the Zoutendijk condition, which is often used to prove global convergence of the nonlinear CG method.

Lemma 3.2 ([22]). *Suppose that Assumptions 3.1 hold. Consider any iterative method of the form* (1.2)*, where d_k satisfies $g_k^T d_k < 0$ and α_k is obtained by the Wolfe conditions. Then* $\sum_{k\geq 1} \frac{(g_k^T d_k)^2}{||d_k||^2} < \infty$.

Next, we will establish the global convergence of the proposed method. To this end, similar to [14], we will establish a bound for changes of the normalized direction $u_k = d_k/||d_k||$. Clearly, if $g_k^T y_{k-1} \leq \varepsilon_1$, then $u_k = -\frac{g_k}{||d_k||}$; otherwise, it suffices to consider $g_k^T y_{k-1} > \varepsilon_1$ for which u_k is given by

$$u_k = \begin{cases} \dfrac{-g_k + \left(\beta_k^{MDL+}\right)^- d_{k-1} - \theta_k^{MDL+} y_{k-1}}{||d_k||} + \left(\beta_k^{MDL+}\right)^+ \dfrac{d_{k-1}}{||d_k||}, & if \quad \beta_k = \beta_k^{MDL+}, \\ \dfrac{-g_k + \left(\beta_k^{DK+}\right)^- d_{k-1} + \tau_k^+ y_{k-1}}{||d_k||} + \left(\beta_k^{DK+}\right)^+ \dfrac{d_{k-1}}{||d_k||}, & if \quad \beta_k = \beta_k^{DK+}, \\ -\dfrac{g_k}{||d_k||} + \eta_k \dfrac{d_{k-1}}{||d_k||}, & if \quad \beta_k = \eta_k, \\ -\dfrac{g_k}{||d_k||} + \beta_k^{HS} \dfrac{d_{k-1}}{||d_k||}, & if \quad \beta_k = \beta_k^{HS}, \end{cases} \tag{3.9}$$

We also define $r_k = \frac{p_k}{||d_k||}$ and the nonnegative parameter $\delta_k = \frac{||d_{k-1}||}{||d_k||}(\beta_k)^+$, given by

$$\delta_k = \begin{cases} \dfrac{||d_{k-1}||}{||d_k||}(\beta_k^{MDL+})^+, & if \quad \beta_k = \beta_k^{MDL+}, \\ \dfrac{||d_{k-1}||}{||d_k||}(\beta_k^{DK+})^+, & if \quad \beta_k = \beta_k^{DK+}, \\ 0, & if \quad \beta_k = 0 \ \ (g_k^T y_{k-1} \leq \varepsilon_1) \text{ or } \beta_k = \eta_k, \\ \dfrac{||d_{k-1}||}{||d_k||}\beta_k^{HS+}, & if \quad \beta_k = \beta_k^{HS}. \end{cases} \tag{3.10}$$

Now, we can reformulate u_k as follows:

$$u_k = \frac{p_k}{||d_k||} + \frac{||d_{k-1}||}{||d_k||}(\beta_k)^+ \cdot \frac{d_{k-1}}{||d_{k-1}||} = r_k + \delta_k u_{k-1}. \tag{3.11}$$

Lemma 3.3. *Suppose that Assumptions 3.1 and* (3.1) *are satisfied. Let $\{x_k\}$ and $\{d_k\}$ be generated by Algorithm 2.3. If* (3.1) *holds, then we have $d_k \neq 0$ and*

$$\sum_{k=1}^{+\infty} ||u_k - u_{k-1}||^2 < \infty. \tag{3.12}$$

Proof. From Lemma 2.6, we have $d_k \neq 0$. Note that $||u_k|| = 1$, we have

$$||r_k|| = ||u_k - \delta_k u_{k-1}|| = ||u_{k-1} - \delta_k u_k|| = \sqrt{1 + \delta_k^2 - 2\delta_k u_{k-1}^T u_k}. \tag{3.13}$$

Equality (3.13), together with the triangle inequality and $\delta_k \geq 0$ implies that

$$\begin{aligned} ||u_k - u_{k-1}|| &\leq (1 + \delta_k)||u_k - u_{k-1}|| \\ &\leq ||u_k - \delta_k u_{k-1}|| + ||u_{k-1} - \delta_k u_k|| \\ &= 2||r_k||. \end{aligned} \tag{3.14}$$

Also, from (3.2) we get

$$\begin{aligned} \sum_{k\geq 1} ||r_k||^2 &= \sum_{k\geq 1} \frac{||p_k||^2}{||d_k||^2} \\ &\leq \sum_{k\geq 1} \frac{M^2}{||g_k||^4} \frac{||g_k||^4}{||d_k||^2} \\ &\leq \frac{M^2}{\varepsilon^4} \sum_{k\geq 1} \frac{(g_k^T d_k)^2}{||d_k||^2} < +\infty, \end{aligned} \tag{3.15}$$

where the second inequality follows from (1.9). Finally, (3.15) and (3.14) ensure (3.12), and so, the proof is completed. □

Now, we deal with the global convergence of Algorithm 2.3. In this context, we present a theorem that shows our method inherits the built-in self-restarting mechanism of the HS method, which was firstly proposed by Gilbert and Nocedal in [14] and then slightly modified by Dai and Liao [10].

Subsequently, we give the convergence result of our presented method, in which the proof is analogous to that of Theorem 3.2 in [15], and we omit it here.

Theorem 3.4. *Suppose that Assumptions 3.1 hold. Let* $\{x_k\}$ *be generated by Algorithm 2.3. If* (3.1) *holds, then* $\liminf_{k\to\infty} ||g_k|| = 0$.

Proof. The proof is divided into the following two phases.

Phase 1. (Finding a bound for the steps $\{s_k\}_{k\geq 1}$) We observe that for any $l \geq k$,

$$x_l - x_k = \sum_{j=k}^{l-1} x_{j+1} - x_j = \sum_{j=k}^{l-1} ||s_j|| u_k + \sum_{j=k}^{l-1} ||s_j|| \, (u_j - u_k). \tag{3.16}$$

Using the triangle inequality and recalling that $||u_k|| = 1$, we can write

$$\sum_{j=k}^{l-1} ||s_j|| \leq ||x_l - x_k|| + \sum_{j=k}^{l-1} ||s_j||||u_j - u_k|| \leq B + \sum_{j=k}^{l-1} ||s_j||||u_j - u_k||. \tag{3.17}$$

Now, let $\triangle$ be a positive integer, chosen large enough such that $\triangle \geq 4BC_1$, where B and C_1 are defined in Assumption (A1) and (3.2), respectively. By Lemma 3.3, there exists an index k_0 such that

$$\sum_{i\geq k_0} ||u_{i+1} - u_i||^2 \leq \frac{1}{4\Delta}. \tag{3.18}$$

If $j > k > k_0$ and $j - k \leq \Delta$, then by (3.18) and Cauchy inequality we have

$$||u_j - u_k|| \leq \sum_{i=k}^{j-1} ||u_{i+1} - u_i|| \leq \sqrt{j-k} \left(\sum_{i=k}^{j-1} ||u_{i+1} - u_i||^2 \right)^{\frac{1}{2}} \leq \frac{\sqrt{j-k}}{2\sqrt{\Delta}} \leq \frac{1}{2}. \tag{3.19}$$

We obtain from the above inequality and (3.17) that

$$\sum_{j=k}^{l-1} ||s_k|| \leq 2B. \tag{3.20}$$

Phase 2. (Finding a bound for the directions $\{d_l\}$) We obtain that scalars β_k^{MDL+} and β_k^{HZ+} are simultaneously truncated by η_k. Then, from (2.4), we obtain

$$\begin{aligned} ||d_l||^2 &\leq \left(\max\{||g_l|| + |\beta_l^{DK+}|||d_{l-1}|| + ||\tau_l y_{l-1}||, ||g_l|| + ||\beta_l^{MDL+} d_{l-1}|| + ||\theta_l y_{l-1}||\}\right)^2 \\ &\leq \left((1+2C_\tau)\gamma + C_1||s_{l-1}|| \cdot ||d_{l-1}||\right)^2 \\ &\leq 2(1+2C_\tau)^2\gamma^2 + 2\left(C_1||s_{l-1}|| \cdot ||d_{l-1}||\right)^2, \end{aligned} \tag{3.21}$$

where the second inequality follows from (3.2), (3.4) and (3.7). Now, set $S_i = 2C_1^2||s_i||^2$. By induction, we have

$$||d_l||^2 \leq \begin{cases} 2(1+2C_\tau)^2\gamma^2 + S_{k0}||d_{k0}||^2, & l = k_0 + 1, \\ 2(1+2C_\tau)^2\gamma^2 \left(1 + \sum\limits_{i=k_0+1}^{l-1} \prod\limits_{j=i}^{l-1} S_j\right) + ||d_{k_0}||^2 \prod\limits_{j=k_0}^{l-1} S_j, & l > k_0 + 1. \end{cases} \tag{3.22}$$

Let us consider a product of Δ consecutive S_j for $k \geq k_0$, that is

$$\prod_{j=k}^{k+\Delta-1} S_j = \prod_{j=k}^{k+\Delta-1} 2C_1^2||s_j||^2 \leq \left(\frac{\sum\limits_{j=k}^{k+\Delta-1} \sqrt{2}C_1||s_j||}{\Delta} \right)^{2\Delta} \leq \left(\frac{2\sqrt{2}BC_1}{\Delta} \right)^{2\Delta} \leq \frac{1}{2^\Delta}, \tag{3.23}$$

where the first inequality follows form arithmetic-geometry inequality, the second one follows from (3.20) and the third one is due to $\triangle \geq 4BC_1$. The product of Δ consecutive S_j is bounded by $1/2^\Delta$, and hence, using (3.23), we obtain that the sum in (3.22) is bounded, independent of l. Furthermore, we can conclude that there exist two positive constants p and q which are independent of $l > k_0$ such that $||d_l||^2 \leq pl + q$. Finally, from Lemma 2.6, we get $\sum\limits_{k\geq 1} \dfrac{(g_k^T d_k)^2}{||d_k||^2} \geq \sum\limits_{k\geq 1} \dfrac{\varepsilon^2}{p} = +\infty$, which contradicts Lemma 3.3. Hence, the proof is finished. □

4 A Variant of the AMDL Method

In this section, we are interested in a variant of Algorithm 2.3, which stems from the following observations.

- First, closely associated to the curvature information of $f(x)$ is the expression $1 - \dfrac{s_k^T y_k}{\|s_k\|^2}$ in (2.2), in which the very parameter $\dfrac{s_k^T y_k}{\|s_k\|^2}$ contains the information of a simple

approximation of the Hessian of $f(x)$ along the line segment $[x_k, x_{k+1}]$, and therefore it is closely related to the estimated Rayleigh quotient of the Hessian of $f(x)$.

- Second, by using a singular value study, the author [13] proposed the optimal value of the parameter $\omega_{k+1} = 0$ in (2.2) in the sense that the condition number in the iteration matrix of the search direction is at minimum, which is preciously the same as the famous 3HS+ method proposed by Narushima et al. in [19].
- Third, we get from Cauchy–Schwarz inequality and (2.18) that

$$1 - \frac{\|y_k\|^2}{s_k^T y_k} \leq 1 - \frac{s_k^T y_k}{\|s_k\|^2} \leq 1. \tag{4.1}$$

If $1 - \dfrac{\|y_k\|^2}{s_k^T y_k} = 0$ and the Wolfe conditions are employed, then the MDL method reduces to the 3HS+ method. Consequently, all singular values of the iteration matrix are equal to 1, which is no other than the simple approximation expression $\dfrac{s_k^T y_k}{\|s_k\|^2} = 1$, then it will obtain an ideal distribution of the singular values.

It seems to hint us that the choice of $\xi = 0$ in (2.2) has its intrinsic advantage, which will increase opportunity of boosting the well–conditioned matrix of search direction. It should be pointed out that the new variant is different from the original AMDL method in that d_{k+1}^{main} in Algorithm 2.3, here we denote it as $\bar{d}_{k+1}$, is renewed. For simplicity, we only give the search direction in Step 5 of the algorithm, and the strong Wolfe conditions are employed, namely, (1.4) and

$$|g\left(x_k + \alpha_k d_k\right)^T d_k| \leq -\sigma g_k^T d_k, \tag{4.2}$$

while the other steps are as exactly the same as Algorithm 2.3.

The complete Hessian approximation by (2.5) and (2.6) with $(g_{k+1}^T d_k)^+$ replaced with $g_{k+1}^T d_k$, and $(g_{k+1}^T s_k)^+$ replaced with $g_{k+1}^T s_k$.

Algorithm 4.1. (A variant of the AMDL method with the strong Wolfe conditions)

$$\bar{d}_{k+1} = \begin{cases} -g_{k+1} + \eta_{k+1}\bar{d}_k, & if \quad k \in K_4, \\ \bar{d}_{k+1}^{main}, & if \quad k \in K_5, \end{cases} \tag{4.3}$$

where η_{k+1} is defined by (2.14) and index sets are given by as follows

$$K_4 = \{k | k \in \mathbb{N} | g_{k+1}^T y_k > \varepsilon_1, \bar{\beta}_{k+1}^{DK} \leq \eta_{k+1} \quad or \quad \bar{\beta}_{k+1}^{MDL} \leq \eta_{k+1}\}, \tag{4.4}$$

$$K_5 = \{k | k \in \mathbb{N} | g_{k+1}^T y_k > \varepsilon_1, \bar{\beta}_{k+1}^{DK} > \eta_{k+1} \quad and \quad \bar{\beta}_{k+1}^{MDL} > \eta_{k+1}\}. \tag{4.5}$$

In (4.3), (4.4) and (4.5), the search direction $\bar{d}_{k+1}^{main}$ and involved scalars are presented by

$$\bar{d}_{k+1}^{main} = \begin{cases} -g_{k+1} + \bar{\beta}_{k+1}^{MDL} d_k - \bar{\theta}_{k+1} y_k, & if \quad s_k^T y_k \geq \|y_k\|^2, \\ -g_{k+1} + \bar{\beta}_{k+1}^{DK} d_k + \bar{\tau}_{k+1} y_k, & if \quad s_k^T y_k < \|y_k\|^2, \end{cases} \tag{4.6}$$

in which

$$\bar{\beta}_{k+1}^{MDL} = \beta_{k+1}^{HS+} - \left(1 - \frac{\|y_k\|^2}{s_k^T y_k}\right) \frac{g_{k+1}^T s_k}{d_k^T y_k}, \tag{4.7}$$

$$\bar{\tau}_{k+1} = \left(1 - \frac{s_k^T y_k}{||y_k||^2}\right) \frac{g_{k+1}^T d_k}{d_k^T y_k}, \tag{4.8}$$

$$\bar{\theta}_{k+1}^{MDL} = \frac{g_{k+1}^T d_k}{d_k^T y_k}, \tag{4.9}$$

$$\bar{\beta}_{k+1}^{DK} = \beta_{k+1}^{HS+} - \frac{\|y_k\|^2}{(d_k^T y_k)^2} g_{k+1}^T d_k. \tag{4.10}$$

Remark 4.2. We drop the sign "+" in the superscript of terms $g_{k+1}^T d_k$ and $g_{k+1}^T s_k$ appearing in parameters β_{k+1}^{DK+}, τ_{k+1}^{+}, β_{k+1}^{MDL+} and θ_{k+1}^{MDL+} defined by (2.7), (2.8), (2.9) and (2.10), which come (4.7), (4.9), (4.10) and (4.8). In order to globalize the Algorithm 4.1, we employ exactly the same restarting mechanism of the PRP+ method and the truncation form of the CG_DESCENT method as Algorithm 2.3 does.

Theorem 4.3. *If the stepsize is determined by the strong Wolfe conditions* (1.4) *and* (4.2) *with* $0 < \rho < \sigma < 1/2$*, then the search direction* $\{d_k\}$ *of Algorithm 4.1 satisfies* $d_k^T g_k \leq -\|g_k\|^2$.

Proof. It suffices to consider the case where $\bar{d}_{k+1} = -g_{k+1} + \bar{\beta}_{k+1}^{DK} d_k + \bar{\tau}_{k+1} y_k$ without truncation form is used. In such a case, $g_{k+1}^T y_k > \varepsilon_1$ is satisfied.

First, if $g_{k+1}^T d_k \leq 0$, then it is easy to verify that $\bar{\tau}_{k+1} \leq 0$, and so,

$$g_{k+1}^T \bar{d}_{k+1} \quad = \quad -||g_{k+1}||^2 + \bar{\beta}_{k+1}^{DK} g_{k+1}^T d_k + \bar{\tau}_{k+1} g_{k+1}^T y_k \leq -\frac{3}{4}||g_{k+1}||^2. \tag{4.11}$$

Second, consider the case where $g_{k+1}^T d_k > 0$. We get from (2.18) and (4.2) that

$$\bar{\tau}_{k+1} = \left(1 - \frac{s_k^T y_k}{||y_k||^2}\right) \frac{g_{k+1}^T d_k}{d_k^T y_k} < \frac{g_{k+1}^T d_k}{d_k^T y_k} \leq \frac{\sigma}{1-\sigma} \triangleq \bar{C}_\sigma, \tag{4.12}$$

which together with $0 < \sigma < 1/2$ gives $0 < \bar{\tau}_{k+1} < \bar{C}_\sigma < 1$.

By direct computation, we obtain that

$$\begin{aligned} g_{k+1}^T \bar{d}_{k+1} &= -||g_{k+1}||^2 + \frac{g_{k+1}^T d_k}{d_k^T y_k} g_{k+1}^T y_k (1 + \bar{\tau}_{k+1}) - \frac{(g_{k+1}^T d_k)^2}{(d_k^T y_k)^2} ||y_k||^2 \\ &= -||g_{k+1}||^2 + \left(\frac{1+\bar{\tau}_{k+1}}{2}\right)^2 ||g_{k+1}||^2 - \left(\frac{1+\bar{\tau}_{k+1}}{2} g_{k+1} + \frac{g_{k+1}^T d_k}{d_k^T y_k} y_k\right)^2 \\ &\leq -\left(1 - \left(\frac{1+\bar{\tau}_{k+1}}{2}\right)^2\right) ||g_{k+1}||^2 \\ &\leq -\left(1 - \left(\frac{1+\bar{C}_\sigma}{2}\right)^2\right) ||g_{k+1}||^2. \end{aligned} \tag{4.13}$$

□

The convergence analysis of the Algorithm 4.1 can be derived using a similar argument, so we omit it here.

Theorem 4.4. *Suppose that Assumptions 3.1 hold. Let* $\{x_k\}$ *be generated by Algorithm 4.1. If* (3.1) *holds, then* $\liminf_{k\to\infty} ||g_k|| = 0$.

5 Numerical Experiments

In this section, we report numerical results in order to evaluate the numerical performance of our proposed methods with that of the existing methods CG_DESCENT, 3HS+ and MDL:

- The HZ method [15]: The CG method with the parameter $\beta_{k+1} = \max\left\{\frac{g_{k+1}^T y_k}{d_k^T y_k} - 2\frac{||y_k||^2}{(d_k^T y_k)^2} g_{k+1}^T d_k, \eta\frac{g_k^T d_k}{||d_k||^2}\right\}$, where $\eta = 0.4$.
- The MDL method [8]: The CG method with the parameters (2.1), (2.2) and (2.3), where $\varpi_{k+1}^{MDL} = \max\left\{0.66, 1 - \frac{\|y_k\|^2}{s_k^T y_k}\right\}$;
- The 3HS+ method [19]: The CG method with the parameters (2.1), (2.3) and $\beta_{k+1} = \frac{g_{k+1}^T y_k}{d_k^T y_k}$, that is to say, $\varpi_{k+1}^{MDL} = 0$;
- Algorithm 2.3: We call it the AMDL1 method, where $\varepsilon_1 = 10^{-14}$;
- Algorithm 4.1: We call it the AMDL2 method, where $\varepsilon_1 = 10^{-14}$.

The implementation code is written in C on a PC (CPU 2.9 GHz with 4GB RAM). Our experiments have been done on a set of 145 nonlinear unconstrained optimization test problems of the CUTEr collection [9], with default dimensions as presented in Hager's homepage: `http://www.math.ufl.edu/~hager/papers/CG`. Since CG_DESCENT is based on the CG method by Hager and Zhang, HZ denotes CG_DESCENT itself.

It should be first stressed that in order to enhance the numerical performance, we uniformly replace the older truncation scheme (2.14) in CG_DESCENT version 5.3, by the latest truncation form $\eta_{k+1}^* = \eta\frac{g_k^T d_k}{||d_k||^2}$, where $\eta = 0.4$. And we employ exactly the same truncation above form in Algorithms 2.3 and 4.1 in actual computation.

All attempts to solve the test problems were limited to achieving a solution which satisfies the termination condition of [16], namely, the inequality $||g_k||_\infty \leq 10^{-6}(1 + |f(x_k)|)$ is satisfied. Meanwhile, the first four algorithms use exactly the same implementation of the Wolfe line search conditions with $\rho = 0.1, \sigma = 0.9$, while the AMDL2 method utilizes the strong Wolfe conditions, in which $\rho = 0.1, \sigma = 0.4$. The performance profile of [12] proposed by Dolan and Morè are employed to get more insight on the performance of the above methods, on the ground of the number of iterations, the number of function evaluations, the number of gradient evaluations and the CPU time.

From Figure 1, it is concluded that the most efficient algorithm in terms of the number of iterations is the AMDL1 method, being the fastest for 50% of the problems, followed by AMDL2 and HZ, for nearly 40% and 36% of the problems. The other methods correspond to the 3HS+ method and the MDL method.

The second part of our comparisons was made on the performance of our methods with that of the other methods based on the total number of function and gradient evaluations being equal to $Nf + 3Ng$ [16], where Nf and Ng respectively denote the number of function and gradient evaluations. As seen in Figures 2, the curves of the AMDL1 and AMDL2 methods are very close, which are often preferable to the HZ method.

Figure 3 vividly demonstrates that the performance profile of AMDL2 is under the others, and AMDL1 is more time consuming comparatively.

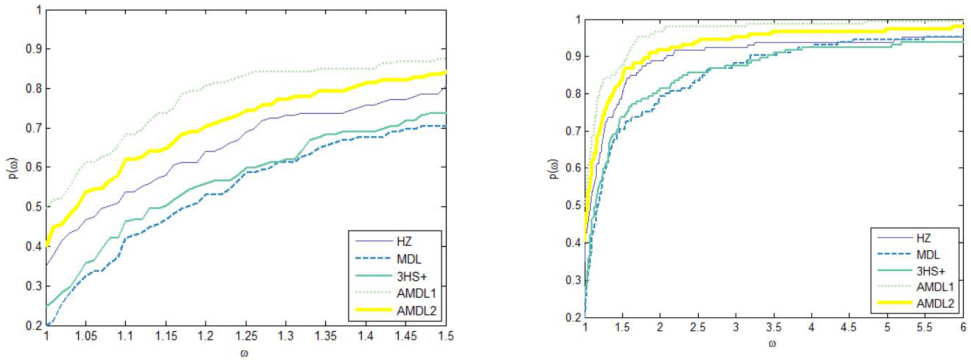

Figure 1: Performance profile of iterations

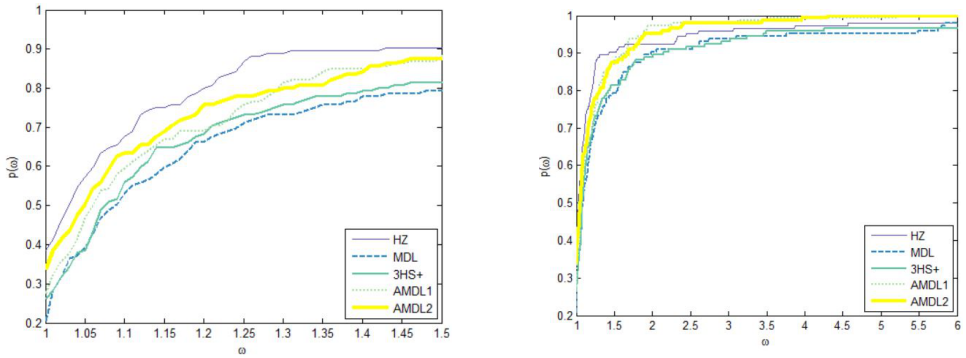

Figure 2: Performance profile of total number of function and gradient evaluations

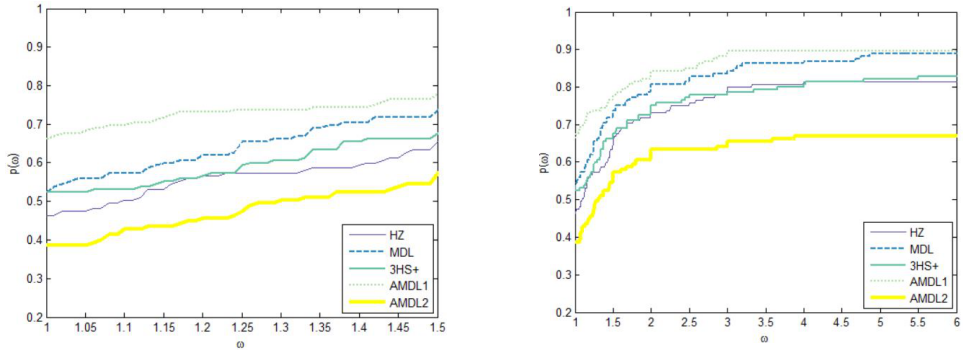

Figure 3: Performance profile of CPU time

Computationally, AMDL1 is not always superior to the other methods, but this method performed better on average. The corresponding search directions make full use of curvature information by exploiting new approximating strategy, which accelerate iteration and improve the reduction in functions values. Thus, the proposed methods seem to be practically promising, specially for solving large-scale problems.

6 Conclusions

In this paper, we propose a new method with adaptively alternative two–side approximating strategy, to compensate the loss of second–order curvature information. Specifically, if it happens that the truncation parameter fails to satisfy Newton equation/qausi–Newton equation after some iterations, the strategy mentioned above begins to work. In such situation, we employ a computationally equivalent yet effective search direction, which simultaneously preserves as much information as possible from the current approximation. By doing so, it maintains almost the same efficiency as the original method and achieves much higher accuracy. Theoretically, we proved that the new method can be globalized even if the objective function is nonconvex. To some extent, numerical experiments showed that exploitation of the reawakened techniques enhance the computational efficiency

Acknowledgements

We would like to thank Professor W.W. Hager and Professor H. Zhang for the CG_DESCENT code. The authors would like to appreciate two anonymous referees for their kind and valuable comments, which have made the paper clearer and more comprehensive than the earlier version.

References

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

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

[3] N. Andrei, A Dai–Liao conjugate gradient algorithm with clustering of eigenvalues, *Numer. Algor.* 77 (2018) 1273–1282.

[4] S. Babaie-Kafaki and R. Ghanbari, The Dai–Liao nonlinear conjugate gradient method with optimal parameter choices, *Eur. J. Oper. Res.* 234 (2014) 625–630.

[5] S. Babaie-Kafaki and R. Ghanbari, A descent family of Dai-Liao conjugate gradient methods. *Optim.Methods Softw.* 243 (2014) 583–591.

[6] S. Babaie-Kafaki, A survey on the Dai-Liao family of nonlinear conjugate gradient methods, RAIRO-Operations Research, 57 (2023) 43–58.

[7] S. Babaie-Kafaki and Z.Aminifard, Two-parameter scaled memoryless BFGS methods with a nonmonotone choice for the initial step length, *Numer.Algor.* 82 (2019) 1345–1357.

[8] S. Babaie-Kafaki and R. Ghanbari, Two modified three-term conjugate gradient method with sufficient descent property. *Optim. Lett.* 8 (2014) 2285–2297.

[9] I. Bongartz, A. R. Conn, N. Gould and P. L. Toint, Cute: constrained and unconstrained testing environment. *ACM T. Math. Software.* 124 (1993) 123–160.

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

[11] 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.

[12] E.D. Dolan and J.J. Moré, Benchmarking optimization software with performance profiles, *Math. Program.* 91(2,Ser.A) (2002) 201–213.

[13] X. Dong, Comment on "A new three–term conjugate gradient method for unconstrained problem", *Numer.Algor.* 72 (2016) 173–179.

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

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

[16] W.W. Hager and H. Zhang, Algorithm 851: CG_Descent, a conjugate gradient method with guaranteed descent, *ACM Trans. Math. Softw.* 32(1) (2006) 113–137.

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

[18] C. Kou, W. Zhang, W. Ai and Y. Liu, On the use of Powell's restart strategy to conjugate gradient methods, *Pac. J. Optim.* 10 (2014) 85–104.

[19] 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.

[20] M.J.D. Powell, Convergence properties of algorithms for nonlinear optimization, *SIAM Rev.* 28 (1986) 487–500.

[21] K. Sugiki, Y. Narushima and H. Yabe, Globally convergent three–term conjugate gradient methods that use secant conditions and generate descent search directions for unconstrained optimization, *J. Optim. Theory Appl.* 153 (2012) 733–757.

[22] P. Wolfe, Convergence conditions for ascent methods, *SIAM Rev.* 11 (1969) 226–235.

Manuscript received 7 February 2023
revised 18 August 2023
accepted for publication 25 September 2023

Xiaoliang Dong
College of Science, Xi'an Shiyou University
Xi'an, 710065, P.R. China;
School of Mathematics Science, Nanjing Normal University
Nanjing, 210023, P.R. China
E-mail address: dongxl@stu.xidian.edu.cn

Deren Han
School of Mathematics and System Science
Beijing Advanced Innovation Center for
Big Data and Brain Computing (BDBC)
Beihang University, Beijing, 100191, P.R. China
E-mail address: handr@buaa.edu.cn