Journal of Mathematical Modeling
Vol. 8, No. 1, 2020, pp. 21-39

JMM

New nonlinear conjugate gradient methods based on optimal Dai-Liao parameters

Saeed Nezhadhosein*

Department of Mathematics, Payame Noor University, Tehran 193953697, Iran
Email: s_nejhadhosein@pnu.ac.ir

Abstract. Here, three new nonlinear conjugate gradient (NCG) methods are proposed, based on a modified secant equation introduced in (IMA. J. Num. Anal. 11 (1991) 325-332) and optimal Dai-Liao (DL) parameters (Appl. Math. Optim. 43 (2001) 87-101). Firstly, an extended conjugacy condition is obtained, which leads to a new DL parameter. Next, to set this parameter, we use three approaches such that the search directions be close to some descent or quasi-newton directions. Global convergence of the proposed methods for uniformly convex functions and general functions is proved. Numerical experiments are done on a set of test functions of the CUTEr collection and the results of these NCGs are compared with some well-known methods.

Keywords: Unconstrained optimization, Modified secant equations, Dai-Liao conjugate gradient method.
AMS Subject Classification : 90C30, 65K05.

1 Introduction

Unconstrained optimization problems arised in many applications in science and engineering [23]. The general form of these problems are as follows :

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

*Corresponding author.

Received: 15 October 2019 / Revised: 7 December 2019 / Accepted: 8 December 2019.
DOI: 10.22124/jmm.2019.14737.1338

where $f : \mathbb{R}^n \to \mathbb{R}$ is a smooth nonlinear function and its gradient is available. The iterative algorithms for solving the unconstrained optimization problems construct a sequence of solutions as $\{x_k\}$, with an initial point $x_0 \in \mathbb{R}^n$, by following recursive formula :

$$x_{k+1} = x_k + s_k, \qquad k = 0, 1, 2, \ldots, \tag{2}$$

where s_k is called step at k-th iteration. For line search (LS) methods the step at the $k-$th iteration is computed as $s_k = \alpha_k d_k$, where d_k is the search direction and α_k is the step length along this direction. General form of the search direction is $d_k = -H_k g_k$, where $g_k = \nabla f(x_k)$ and H_k is the matrix of search direction with dimension $n \times n$. The structure of this matrix is vital for the appropriate search direction [23], which should be descent or sufficient descent. The step length α_k usually is chosen to satisfy certain line search conditions [30], as inexact line search. Among them, the so-called Wolfe conditions [30], have attracted special attention in the convergence analyse and the implementations of conjugate gradient (CG) methods, requiring:

$$f(x_k + \alpha_k d_k) - f(x_k) \le \delta \alpha_k g_k^T d_k, \tag{3}$$

$$g(x_k + \alpha_k d_k)^T d_k \ge \sigma g_k^T d_k, \tag{4}$$

where $0 < \delta < \sigma < 1$. These conditions guaranteed that $s_k^T y_k > 0$, where $y_k = g_{k+1} - g_k$.

Among LS methods the CG methods are characterized by low memory requirements and strong global convergence properties [9], which makes them popular for engineers and mathematicians engaged in solving large-scale problems. These methods, using eqn (2), lead to a sequence of the approximate solutions as $\{x_k\}$ with the following recursive formula for their search directions:

$$d_0 = -g_0, \qquad d_{k+1} = -g_{k+1} + \beta_k d_k, \quad k = 0, 1, 2, \ldots, \tag{5}$$

where β_k is a scalar called the CG (update) parameter.

Different choices for the CG parameters in Eq. (5), lead to different CG methods based on different conjugacy conditions. In early CG methods, the conjugacy condition is based on the convex quadratic function and the exact line search, which is $d_i^T H d_j = 0,\ \forall i \neq j$, where H is the Hessian of the objective function. These methods lead to the classic or linear CG methods such as Fletcher-Reeves (FR) [13], Hestenes-Stiefel (HS) [19], Polak-Ribiere-Polyak (PRP) [25, 26] and Dai-Yuan (DY) [11]. All of them have the same performance for classic CG methods, but have different global convergence properties and numerical performances for general nonlinear objective functions or inexact line search. For a complete overview of classic methods, their categorization and global convergence properties, see [18]. Furthermore, for general nonlinear functions under exact line search all of them satisfy the conjugacy condition as $y_k^T d_{k+1} = 0$.

Along with classic CG methods, there are three approaches in defining CG parameter, β_k, which lead to new nonlinear CG (NCG) methods in literature. In the first approach, which is called descent approach, the directions in classic CGs are approached or converted to a descent or sufficient descent directions. For example Zhang et al. [39], construct some descent classic CG direction as three term CG methods with sufficient descent directions. As special case, they proposed a three term HS, TTHS, with following search direction [39]:

$$d_{k+1}^{TTHS} = -g_{k+1} + \beta_k^{HS} d_k - \theta_{k+1} y_k, \tag{6}$$

where

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k}, \qquad \theta_{k+1} = \frac{g_{k+1}^T d_k}{d_k^T y_k}. \tag{7}$$

It is also obvious that if exact line search is applied, then $\theta_{k+1} = 0$, and the above method is reduced to the classic HS method. By replacing the HS method with other linear CG methods, some new descent methods, such as TTPR and TTFR can be achieved (see [39]). An attractive feature of these methods is that the search directions satisfy the sufficient descent condition, i.e. $d_k^T g_k = -\|g_k\|^2$, which is independent of line search procedure.
Also, Babaie-Kafaki and Ghanbari [39] used a descent approach for a hybridization of HS and DY methods, named HCG, based on solving a least-square problem of minimizing the distance between the search directions of the HCG method and a three term CG method proposed by Zhang et al. [39], which possesses the sufficient descent property.

In the second approach, the search directions in NCGs are achieved by extended conjugacy conditions. For example, Perry [24] incorporated the secant equation [23,29] to conjugacy condition and proposed a method with new extended conjugacy condition, which is $y_k^T s_{k+1} = -s_k^T g_{k+1}$. Next, the search direction defined as follows:

$$d_{k+1}^P = -g_{k+1} + \beta_k^P d_k = -Q_{k+1}^P g_{k+1}, \tag{8}$$

where Q_{k+1}^P is the direction matrix, as a nonsymmetric matrix which approximate the inverse Hessian of the objective function at current iteration, and β_k^P is the Perry CG parameter, which are defined as follows:

$$\beta_k^P = \frac{g_{k+1}^T y_k}{d_k^T y_k} - \frac{g_{k+1}^T s_k}{d_k^T y_k}, \qquad Q_{k+1}^P = I - \frac{s_k y_k^T}{y_k^T s_k} + \frac{s_k s_k^T}{y_k^T s_k}. \tag{9}$$

Note that using Wolfe conditions, Eqs. (3)-(4), we have $s_k^T y_k > 0$, so the matrix Q_{k+1}^P in (9) is well-defined. In Perry approach, the direction matrix, Q_{k+1}^P, is not symmetric and also doesn't satisfied the secant equation [28]. To overcome these defects, Shanno [28] combined the Perry method and memoryless BFGS method [23], as the most famous QN method, to introduce a new matrix direction as follows :

$$Q_{k+1}^S = I - \frac{s_k y_k^T + y_k s_k^T}{s_k^T y_k} + (1 + \frac{y_k^T y_k}{s_k^T y_k}) \frac{s_k s_k^T}{s_k^T y_k}. \tag{10}$$

In 2001, Dai and Liao [10] using an extended Perry conjugacy condition by a parameter, which is $y_k^T s_{k+1} = -t s_k^T g_{k+1}$, introduced the new NCG method by following direction:

$$d_{k+1}^{DL} = -g_{k+1} + \beta_k^{DL} d_k = -Q_{k+1}^{DL} g_{k+1}, \tag{11}$$

where

$$\beta_k^{DL} = \frac{g_{k+1}^T y_k}{d_k^T y_k} - t \frac{g_{k+1}^T s_k}{d_k^T y_k}, \qquad Q_{k+1}^{DL} = I - \frac{s_k y_k^T}{s_k^T y_k} + t \frac{s_k s_k^T}{s_k^T y_k}, \tag{12}$$

where t is a nonnegative scalar and called the DL parameter. Note that if $t = 0$, then β_k^{DL} reduces to β_k^{HS}, Eq. (7), and if $t = 1$, then β_k^{DL} reduces to β_k^P, Eq. (9). Since the setting of the DL parameter is an open problem in NCG methods [22], many efforts have been made by researchers to adjust it. For example, in descent approach based on an eigenvalue study, the authors in [3] proposed a descent family of DL method, namely, DDL. The search direction in DDL method is as follows [3]:

$$d_{k+1}^{DDL} = -(I + \frac{s_k y_k^T}{s_k^T y_k} - t_k^{p,q} \frac{s_k s_k^T}{s_k^T y_k}) g_{k+1}, \tag{13}$$

where $t_k^{p,q}$ is DL parameter as follows:

$$t_k^{p,q} = p \frac{\|y_k\|^2}{s_k^T y_k} - q \frac{s_k^T y_k}{\|s_k\|^2}, \tag{14}$$

where p and q are nonnegative parameters, which $p < \frac{1}{4}$ and $q \geq \frac{1}{4}$. For more information about setting of the DL parameter, see [1, 2, 5, 6, 37].

For global convergence properties of general objective functions, the authors in [10] considered the a truncated form of the DL method, with an extended CG parameter, namely β_k^{DL+}, and following direction:

$$\begin{aligned} d_{k+1}^{DL+} &= -g_{k+1} + \beta_k^{DL+} d_k \\ &= -g_{k+1} + \left(\max\{\frac{g_{k+1}^T y_k}{d_k^T y_k}, 0\} - t\frac{g_{k+1}^T s_k}{d_k^T s_k} \right) d_k, \end{aligned} \tag{15}$$

In the third approach of determining the NCG parameter, the second order information is applied in the sense of the QN aspects. It means that the QN or Newton search directions used to approximate of search direction. For example as a descent CG method, independent to type of line search, Hager and Zhang (HZ) [17] introduced the following CG parameter:

$$\beta_k^{HZ} = \frac{g_{k+1}^T y_k}{d_k^T y_k} - 2\frac{\|y_k\|^2}{d_k^T y_k}\frac{g_{k+1}^T d_k}{d_k^T y_k}, \tag{16}$$

This method can be viewed as an adaptive version of the DL parameter corresponding to $t = 2\frac{\|y_k\|^2}{s_k^T y_k}$ in Eq. (12), where $\|.\|$ denotes the Euclidean norm. Another adaptive DL parameter is based on scaled memoryless BFGS, suggested by Dai and Kou (DK) [8], as follows:

$$\beta_k^{DK}(\tau_k) = \frac{g_{k+1}^T y_k}{d_k^T y_k} - (\tau_k + \frac{\|y_k\|^2}{s_k^T y_k} - \frac{s_k^T y_k}{\|s_k\|^2})\frac{g_{k+1}^T s_k}{d_k^T y_k}, \tag{17}$$

in which τ_k is a parameter corresponding to the scaling factor in the scaled memoryless BFGS method.

All three above approaches for determining NCG parameter can be applied by variation of conjugacy conditions with modified secant equations. For example, based on modified BFGS, MBFGS, method in [20], Zhang and Zhou [38], introduced the modified TTHS method (see Eq. (6)), called MTTHS, as follows :

$$d_{k+1}^{MTTHS} = -g_{k+1} + \beta_k^{MHS} d_k - \theta_k^M z_k, \tag{18}$$

where

$$\beta_k^{MHS} = \frac{g_{k+1}^T z_k}{d_k^T z_k}, \qquad \theta_k^M = \frac{g_{k+1}^T d_k}{d_k^T z_k}, \tag{19}$$

and

$$z_k \triangleq y_k + c\|g_k\|^r s_k, \tag{20}$$

where $r \geq 0$ and $c > 0$ are some constant. In Eq. (20), z_k plays an important role in the global convergence of the MBFGS method for nonconvex function [38]. This idea motivated us to apply a modified secant equation for combining the three mentioned approaches in setting the new NCG parameters based on some optimal DL parameters. In other word, firstly, by introducing new conjugacy condition, we parametrize the search direction and next to specify the DL parameter two descent and QN approaches are used.

The remainder of this paper is organized as follows. In Section 2, we propose a new extended conjugacy condition based on a new modified secant equation, proposed in [31], and DL approach, Eq. (11). Then we discuss three methods to setting the DL parameter by descent and QN search direction approaches. In the first method, we use the MTTHS descent method, Eqs. (18)-(19). In second method, we try to match the direction matrix of the CG method to the Shanno QN direction matrix, Q^S_{k+1} given by (10). In third method, we use the condition number of the search matrix direction, similar to [3] for DL method, to descent the search direction. Afterwards, in Section 3, we discuss the global convergence of the proposed methods. In Section 4, we numerically compare our methods with the DK and DDL methods and report comparative testing results. Finally, we make conclusions in Section 5.

2 New nonlinear conjugate gradient methods

Using modified secant equations are common in CG methods to define new conjugacy conditions by DL approach, for example see [14, 15, 20, 31–33, 35, 36, 38]. However, as far as we know, modified secant equations have not been manipulated in the DL parameter. Thus for the purpose of setting the DL parameter in a more accurate fashion, we introduce new NCG methods, based on a modified secant equation, proposed by Yuan [31]. Therefore, in this section, we first define a new extended modified conjugacy condition for NCG methods, and then propose three methods for approximating the optimal NCG parameters.

2.1 New conjugate condition

As a disadvantage, only the gradient information is used in the standard secant equation, and the value of the function does not play any role in it. To resolve this issue, based on Taylor expansion and quadratic interpretation of the objective function, Yuan [31] proposed a modified secant equation, which then studied by Wei et al. [32, 33] and Li et al. [34]. This equation is as follows :

$$B_{k+1}s_k = \overline{y}_k, \qquad \overline{y}_k = y_k + A_k u_k, \tag{21}$$

where B_{k+1} is an approximation of the Hessian matrix of the objective function, $u_k \in \mathbb{R}^n$ is a vector that satisfy $s_k^T u_k \neq 0$ and $A_k = \frac{\theta_k}{s_k^T u_k}$ where θ_k is as follows:

$$\theta_k = 2(f_k - f_{k+1}) + (g_k - g_{k+1})^T s_k. \tag{22}$$

It is remarkable that for a quadratic objective function f, we have $\theta_k = 0$ for every $k \geq 0$, and so the modified secant equation in Eq. (21), convert to the standard secant equation [23]. According to (21), similar to DL conjugate condition [10], the new extended conjugacy condition is presented as follows:

$$d_{k+1}^T \overline{y}_k = -t^{\overline{DL}} g_{k+1}^T s_k, \tag{23}$$

where $t^{\overline{DL}}$ is called $\overline{DL}$ (update) parameter and the Eq. (23) is named $\overline{DL}$ conjugate condition. Using CG direction in (5) and the conjugacy condition in (23), we have the following CG parameter:

$$\beta_k^{\overline{DL}} = \frac{g_{k+1}^T \overline{y}_k}{d_k^T \overline{y}_k} - t^{\overline{DL}} \frac{g_{k+1}^T s_k}{d_k^T \overline{y}_k}, \tag{24}$$

By replacing the (24) in (5) and rearranging the vectors, we have the following new search direction :

$$d_{k+1}^{\overline{DL}} = -Q_{k+1}^{\overline{DL}} g_{k+1} = -(I + \frac{s_k \overline{y}_k^T}{s_k^T \overline{y}_k} - t^{\overline{DL}} \frac{s_k s_k^T}{s_k^T \overline{y}_k}) g_{k+1}. \tag{25}$$

The associate CG method is called $\overline{DL}$. Now, similar to DL parameter, the setting of the $\overline{DL}$ parameter is an important factor. In following, we use two approaches with three methods to set it.

2.2 Setting the $\overline{DL}$ parameter

To set the $\overline{DL}$ parameter, we apply three methods based on two approaches, descent and QN directions. In the first and third methods the descent approach and in second method, the QN approach are applied .

2.2.1 Descent method

Using descent directions in LS approaches is an important issue to convergence analysis. Since the $\overline{DL}$ search directions, Eq. (25), may not satisfy the descent condition, so with two approaches, we try to satisfy the descent condition. In the first approach, similar to [7] for DL method, to close the $\overline{DL}$ search direction to the MTTHS direction in (18)-(19), consider the following subproblem:

$$t_k^{\overline{DL}*} = \texttt{argmin}\|d_{k+1}^{\overline{DL}} - d_{k+1}^{MTTHS}\|, \tag{26}$$

where $\|.\|$ is the Euclidean norm. By taking simple algebraic calculations, we get the $\overline{DL}$ parameter as follows:

$$t_k^{\overline{DL}} = \frac{1}{a_2}(a_1 - a_3 + \frac{a_4}{\|d_k\|^2} d_k^T z_k), \tag{27}$$

where

$$a_1 = \frac{g_{k+1}^T \overline{y}_k}{d_k^T \overline{y}_k}, \qquad a_2 = \frac{g_{k+1}^T s_k}{d_k^T \overline{y}_k},$$
$$a_3 = \frac{g_{k+1}^T z_k}{d_k^T z_k}, \qquad a_4 = \frac{g_{k+1}^T d_k}{d_k^T z_k},$$

where z_k and $\overline{y}_k$ are defined in Eqs. (20) and (21), respectively. Note that from Wolfe conditions, Eqs. (3)-(4), we have $s_k^T y_k > 0$, and so $d_k^T z_k > 0$ and $d_k^T \overline{y}_k > 0$. Therefore, the fractional expressions in Eq. (27) are well-defined. After some simplification, the Eq. (27) can be written as follows:

$$t_k^{\overline{DL}} = -\frac{\overline{y}_k^T g_{k+1}}{s_k^T g_{k+1}}. \tag{28}$$

To ensure the well-definiteness of the DL parameter in Eq. (28), we set an adaptive version of it as follows :

$$t_k^{\overline{DL}*} = \begin{cases} t_k^{\overline{DL}}, & |s_k^T g_{k+1}| > \varepsilon, \\ 1, & \text{otherwise}, \end{cases} \tag{29}$$

with a enough small positive constant ε. So, by replacing (29) in (25), we get a new nonlinear $\overline{DL}$ direction as follows :

$$d_{k+1}^{NDL-1} = -g_{k+1} + (\frac{g_{k+1}^T \overline{y}_k}{d_k^T \overline{y}_k} - t_k^{\overline{DL}*} \frac{g_{k+1}^T s_k}{d_k^T \overline{y}_k}) d_k, \tag{30}$$

The NCG method based on the search direction d_{k+1}^{NDL-1} in Eq. (30), called "$NDL - 1$" method.

2.2.2 QN method

To access the CG direction matrices to approximate the inverse Hessian matrix, similar to [24, 28] for the quasi-newton methods, we improve the efficiency of CG method. For this purpose, we try to approach the matrix direction of the $\overline{DL}$ method, $Q_{k+1}^{\overline{DL}}$, in Eq. (25), to the Shanno quasi-Newton search direction matrix, Q_{k+1}^{S}, Eq. (10). Therefore, consider the following subproblem:

$$t_k^{\overline{DL}*} = \texttt{argmin}\|Q_{k+1}^{\overline{DL}} - Q_{k+1}^{S}\|_F, \tag{31}$$

where $\|.\|_F$ is Frobenius norm. Using the property $tr(AA^T) = \|A\|_F^2$ and after some algebraic calculations, we have

$$t_k^{\overline{DL}*} = 1 + \frac{\overline{y}_k^T \overline{y}_k}{s_k^T \overline{y}_k} - \frac{s_k^T \overline{y}_k}{\|s_k\|^2}. \tag{32}$$

Note that using Wolfe conditions, we have $s_k^T \overline{y}_k > 0$ and the $\overline{DL}$ parameter in Eq. (32), is well-defined. By replacing (32) in (25), we get another new $\overline{DL}$ direction as follows:

$$d_{k+1}^{NDL-2} = -g_{k+1} + \left(\frac{g_{k+1}^T \overline{y}_k}{d_k^T \overline{y}_k} - t_k^{\overline{DL}*} \frac{g_{k+1}^T s_k}{d_k^T \overline{y}_k}\right) d_k, \tag{33}$$

where $t_k^{\overline{DL}*}$ is defined in Eq. (32). The NCG method based on d_{k+1}^{NDL-2} in Eq. (33), called "$NDL-2$" method.

2.2.3 Eigenvalue approach

In the third method, similar to DL method in [3], we apply a descent approach in Section 2.2.1, using an eigenvalue study of the $\overline{DL}$ direction matrix. Therefore, using Eqs. (13)-(14), with replacing the vector y_k with $\overline{y}_k$ in Eq. (21), we have the following search direction:

$$d_{k+1}^{NDL-3} = -\left(I + \frac{d_k^T \overline{y}_k}{d_k^T \overline{y}_k} - t_k^{\overline{DL}*(p,q)} \frac{d_k^T s_k}{d_k^T \overline{y}_k}\right) g_{k+1} = -Q_{k+1}^{D\overline{DL}} g_{k+1}, \tag{34}$$

where p and q are nonnegative parameters, which $p < \frac{1}{4}$ and $q \geq \frac{1}{4}$ (see [3]), and $t_k^{\overline{DL}*(p,q)}$ is as follows:

$$t_k^{\overline{DL}*(p,q)} = p\frac{\|\overline{y}_k\|^2}{s_k^T \overline{y}_k} - q\frac{s_k^T \overline{y}_k}{\|s_k\|^2}. \tag{35}$$

By Wolfe conditions we have $s_k^T \overline{y}_k > 0$, and so the Eq. (35), is well-defined. The NCG method with the $\overline{DL}$ parameter in (35) called "$NDL-3$".

3 Convergence analysis

In this section, we discuss the global convergence of our NCG methods, consist of "$NDL - i$", $i = 1, 2, 3$. In our analysis, we need to make the following basic assumptions on the objective function, commonly used in the convergence analysis of the CG methods [9].

Assumption (A):
We assume that the objective function f is strongly convex and ∇f is Lipschitz continuous on the level set

$$S = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\} \tag{36}$$

That is, there exists constants $\mu > 0$ and L such that

$$(\nabla f(x) - \nabla f(y))^T (x - y) \geq \mu \|x - y\|^2 \quad , \quad \forall x, y \in S \tag{37}$$

and

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \quad , \quad \forall x, y \in S \tag{38}$$

Note that these assumptions implies that there exists a positive constant Γ such that for all $x \in S$; $\|\nabla f(x)\| \leq \Gamma$.

Lemma 1. *[29] Suppose that the Assumption (A) holds. Consider any CG method in the form of (2) and (5) in which for all $k \geq 0$, the search direction d_k is a descent direction and the step length α_k is determined to satisfy the Wolfe conditions, (3)-(4). If*

$$\sum_{k \geq 0} \frac{1}{\|d_k\|^2} = \infty, \tag{39}$$

then the method converges in the sense that

$$\liminf_{k \to \infty} \|g_k\| = 0. \tag{40}$$

Theorem 1. *Suppose that the Assumption (A) holds for the objective function f in (1). Consider a CG method in the form of (2) and (5) with the CG direction defined by (30), "$NDL - 1$" method, in which the step length α_k is computed such that the Wolfe conditions (3)-(4) are satisfied. If the objective function f is uniformly convex on $\mathcal{S}$, then the method converges in the sense that (40) holds.*

Proof. The uniform convexity of the differentiable function f ensures that there exists a positive constant μ such that (see Theorem 1.3.16 of [29])

$$y_k^T s_k \geq \mu \|s_k\|^2. \tag{41}$$

Note that similar inequality can be proved by replacing y_k with $\overline{y_k}$. Therefore, there exists a positive constant $\overline{\mu}$, such that:

$$\overline{y}_k^T s_k \geq \overline{\mu} \|s_k\|^2. \tag{42}$$

Also from the second equation of the Wolf conditions, Eq. (3), we have:

$$g_{k+1}^T d_k \geq \sigma g_k^T d_k. \tag{43}$$

On the other hand, from (21) we have :

$$\|\overline{y_k}\| \leq \|y_k\| + \|A_k u_k\| = \|y_k\| + \|w_k\|, \tag{44}$$

where $w_k = A_k u_k$ and A_k is defined in (21). Now we show that $\|\overline{y}_k\| \leq L_1 \|s_k\|$, where $L_1 > 0$ is a constant. For this purpose, first of all, using Taylor expansion of θ_k in Eq. (22), we have:

$$|\theta_k| < M \|s_k\|^2. \tag{45}$$

Then, considering the Eq. (21), we have two cases for w_k [20]: $w_k = \frac{\theta_k s_k}{\|s_k\|^2}$ or $w_k = \frac{\theta_k y_k}{s_k^T y_k}$, which θ_k is defined in (22). In the first case, from Eqs. (38), (44) and (45), we get:

$$\|\overline{y}_k\| \leq \|y_k\| + \frac{|\theta_k| \|s_k\|}{\|s_k\|^2} = (L + M)\|s_k\| = M_1 \|s_k\|, \tag{46}$$

where $M_1 = L + M$. In the second case, from (38), (41) and (45) we have:

$$\|\overline{y}_k\| \leq \|y_k\| + \frac{ML\|s_k\|^3}{\mu \|s_k\|^2} \leq L(1 + \frac{M}{\mu})\|s_k\| = M_2 \|s_k\|, \tag{47}$$

where $M_2 = L(1 + \frac{M}{\mu})$. Now, by defining $L_1 = \max\{M_1, M_2\}$, we have:

$$\|\overline{y}_k\| \leq L_1 \|s_k\|. \tag{48}$$

Next we can show that $\|z_k\| \leq L_2 \|s_k\|$, where z_k is defined in (20). From the Eqs. (20) and (38), we have :

$$\begin{aligned} \|z_k\| &= \|y_k + c\|g_k\|^r s_k\| \leq \|y_k\| + c\|g_k\|^r \|s_k\| \leq L\|s_k\| + c\|g_k\|^r \|s_k\| \\ &\leq (L + c\Gamma^r)\|S_k\| = L_2 \|s_k\|, \end{aligned} \tag{49}$$

where $L_2 = L + c\Gamma^r$. Moreover, from (41) and (20) we have:

$$\begin{aligned} s_k^T z_k &= s_k^T (y_k + c\|g_k\|^r s_k) = s_k^T y_k + c\|g_k\|^r \|s_k\|^2 \\ &\geq (\mu + c\|g_k\|^r)\|s_k\|^2 \geq \mu\|s_k\|^2, \end{aligned} \tag{50}$$

Which implies that $s_k^T z_k \geq \mu\|s_k\|^2$. Hence from this inequality and Eqs. (42), (48), (49), (50), (3) and Cauchy-Shwartz inequality we have :

$$\begin{aligned} |t_k^{\overline{DL}*}| &= \frac{d_k^T \overline{y}_k}{g_{k+1}^T s_k} \left(\frac{g_{k+1}^T \overline{y}_k}{d_k^T \overline{y}_k} + \frac{g_{k+1}^T z_k}{s_k^T z_k} + \frac{g_{k+1}^T d_k}{\|d_k\|^2} \right) \\ &\leq \frac{\|s_k\|\|\overline{y}_k\|}{\sigma\|g_{k+1}\|\|s_k\|} \left(\frac{\|g_{k+1}\|\|\overline{y}_k\|}{\overline{\mu}\|s_k\|^2} + \frac{\|g_{k+1}\|\|z_k\|}{\mu\|s_k\|^2} + \frac{\|g_{k+1}\|\|s_k\|}{\|s_k\|^2} \right) \\ &\leq \frac{L_1}{\sigma} \left(\frac{L_1}{\overline{\mu}} + \frac{L_2}{\mu} + 1 \right). \end{aligned} \tag{51}$$

That is $t_k^{\overline{DL}*}$ is bounded for uniformly convex objective function. So, if we use the Wolfe conditions, (3)-(4), similar to Theorem (2.1) in [3], the search directions are bounded away, which with Lemma 1 complete the proof. □

Theorem 2. *Suppose that the Assumption (A) holds for the objective function f in (1). Consider a CG method in the form of (2) and (5) with the CG direction defined by (33), "$NDL-2$" method, in which the step length αk is computed such that the Wolfe conditions (3)-(4) are satisfied. If the objective function f is uniformly convex on $\mathcal{S}$, then the method converges in the sense that (40) holds.*

Proof. Considering the Assumption (A) and the assumptions of Theorem 1, from Eqs. (37), (42), (46), (48) and definition of $t_k^{\overline{DL}*}$ in Eq. (32), we have:

$$\begin{aligned} |t_k^{\overline{DL}*}| &= |1 + \frac{\overline{y}_k^T \overline{y}_k}{s_k^T \overline{y}_k} - \frac{s_k^T \overline{y}_k}{\|s_k\|^2}| \leq 1 + \frac{\|\overline{y}_k\|^2}{|s_k^T \overline{y}_k|} + \frac{|s_k^T \overline{y}_k|}{\|s_k\|^2} \\ &\leq 1 + \frac{L_1^2\|s_k\|^2}{\|s_k\|^2} + \frac{\|s_k\|\|\overline{y}_k\|}{\|s_k\|^2} = 1 + L_1 + L_1^2. \end{aligned} \tag{52}$$

So, similar to Theorem 1, the search directions are bounded away and the proof is complete. □

Theorem 3. *Suppose that the Assumption (A) holds for the objective function f in (1). Consider a CG method in the form of (2) and (5) with the CG direction defined by (30), "$NDL-3$" method, in which the step length α_k is computed such that the Wolfe conditions (3)-(4) are satisfied. If the objective function f is uniformly convex on $\mathcal{S}$, then the method converges in the sense that (40) holds .*

Proof. Considering the Assumption (A) and the assumptions of Theorem 1, from Eqs. (37), (42), (46), (48) and definition of $t_k^{\overline{DL}*(p,q)}$, Eq. (35), where $p < \frac{1}{4}$ and $q \geq \frac{1}{4}$, we have :

$$|t_k^{\overline{DL}*(p,q)}| = |p\frac{\|\overline{y}_k^2\|}{s_k^T\overline{y}_k} - q\frac{s_k^T\overline{y}_k}{\|s_k^2\|}| \leq p\frac{\|\overline{y}_k^2\|}{|s_k^T\overline{y}_k|} + q\frac{|s_k^T\overline{y}_k|}{\|s_k^2\|} \leq p\frac{L_1^2}{\bar{\mu}} + q\bar{\mu},$$

So, similar to Theorem 3, the "$NDL-3$" method is convergent. □

Remark 1. *For general objective functions, similar to the truncated CG parameter in [10], for DL method, which employed the Powell's nonnegative restriction on the CG parameter [27], "DL+" method, we can use new truncated parameters, $t_k^{\overline{DL}+}$. Now, using the Theorem 3.6 in [3], for descent search directions, the global convergence properties is satisfied.*

4 Numerical Experiments

Here, we present some numerical results, obtained by applying a MATLAB 8.8.0.1 (R2013a) implementation of the proposed NCG methods, "$NDL-i$", $i = 1, 2, 3$. The numerical results are compared with the DDL [3] with parameters $p = 0.2$ and $q = 0.9$, and DK [8] with parameter $\tau_k = \frac{s_k^T y_k}{\|s_k\|^2}$. The implementations were performed on a computer, Intel(R) Core (TM) A10-8700P CPU 3.20 Gigahertz 64-bit desktop with 8 Gigabyte RAM. Our experiments have been done on the test problems of unconstrained optimization problems of CUTEr collection [16]. The names and dimensions of these problems are presented in Table 1. The dimensions of the test problems range from 2 to 1000 variables.

For ensuring the descent property, we restarted the algorithm with Powell Restart [29], which is $|g_k^T g_{k+1}| < 0.2\|g_{k+1}\|$. In all the methods, we used the effective approximate Wolfe conditions described in (3)-(4) with parameters $\sigma = 0.9$ and $\rho = 10^{-4}$. The same stop condition is considered for all methods which is $\|g_k\|_\infty \leq 10^{-6}$ and the maximum number of iterations is limited to 1000.

The comparing data contain the number of iterations, CPU time and the number of evaluations for function, n_f, and gradient, n_g as $n_f + 3n_g$ [3]. To approximately assess the performance of different algorithms, we use the performance profile introduced by Dolan and More [12]. As shown in Figure 1, with respect to the number of iteration, the "$NDL-2$" method is the best among all methods and "$NDL-3$" method is competitive with the DDL method. In addition, Figure 2 shows that with respect to the number of

Table 1: The test problems.

Name	Dim	Name	Dim	Name	Dim
AKIVA	2	DIXMAANH	3000	LIARWHD	5000
ALLINITU	4	DIXMAANI	3000	LOGHAIRY	2
ARGLINA	200	DIXMAANJ	3000	MANCINO	100
ARGLINB	200	DIXMAANK	3000	MARATOSB	2
ARGLINC	200	DIXMAANL	3000	MEXHAT	2
ARWHEAD	5000	DIXMAANM	15	NONCVXU2	5000
BARD	3	DIXMAANN	15	NONCVXUN	5000
BDQRTIC	5000	DIXMAANO	15	NONDIA	5000
BEALE	2	DIXMAANP	15	NONDQUAR	5000
BEALE	2	DIXON3DQ	10000	OSBORNEA	5
BIGGS6	6	DJTL	2	PALMER1C	8
BOX	10000	DQDRTIC	5000	PALMER2C	8
BOX3	3	DQRTIC	5000	PALMER3C	8
BRKMCC	2	EDENSCH	2000	PALMER4C	8
BROWNAL	200	EG2	1000	PALMER5C	6
BROWNDEN	4	ENGVAL1	5000	PALMER6C	8
BROYDN7D	5000	ENGVAL2	3	PALMER6C	8
BRYBND	5000	ERRINROS	50	PALMER8C	8
CHAINWOO	4000	ERRINRSM	50	HIMMELBF	4
CHNROSNB	50	EXPFIT	2	HIMMELBG	2
CHNRSNBM	50	EXTROSNB	1000	HIMMELBH	2
CLIFF	2	FLETBV3M	5000	POWELLSG	5000
CLIFF	2	FLETCHBV	5000	POWER	10000
CUBE	2	FLETCHCR	1000	QUARTC	5000
CURLY10	10000	FMINSRF2	5625	ROSENBR	2
CURLY20	10000	FMINSURF	5625	SINEVAL	2
CURLY30	10000	FREUROTH	5000	SISSER	2
DECONVU	63	GENHUMPS	5000	SPARSINE	5000
DENSCHNA	2	GENROSE	500	SPARSQUR	10000
DENSCHNB	2	GULF	3	SPMSRTLS	4999
DENSCHNC	2	HAIRY	2	SROSENBR	5000
DENSCHND	3	HATFLDD	3	SSCOSINE	10
DENSCHNE	3	HATFLDE	3	TESTQUAD	5000
DENSCHNF	2	HATFLDFL	3	TOINTGOR	50
DIXMAANC	3000	HEART6LS	6	TOINTPSP	50
DIXMAANA	3000	HEART8LS	8	TOINTQOR	50
DIXMAANB	3000	HELIX	3	TQUARTIC	5000
DIXMAANC	3000	HILBERTA	2	TRIDIA	5000
DIXMAAND	3000	HILBERTB	10	VARDIM	20
DIXMAANE	3000	HIMMELBB	2	WOODS	4000
DIXMAANF	3000	JENSMP	2	YFITU	3
DIXMAANG	3000	KOWOSB	4	ZANGWIL2	2

iteration, the "$NDL-2$" is the best methods. Also, the "$NDL-1$" method is competitive with DK method. Moreover, from Figure 3, "$NDL-2$" and "$NDL-3$" methods, are the best methods.

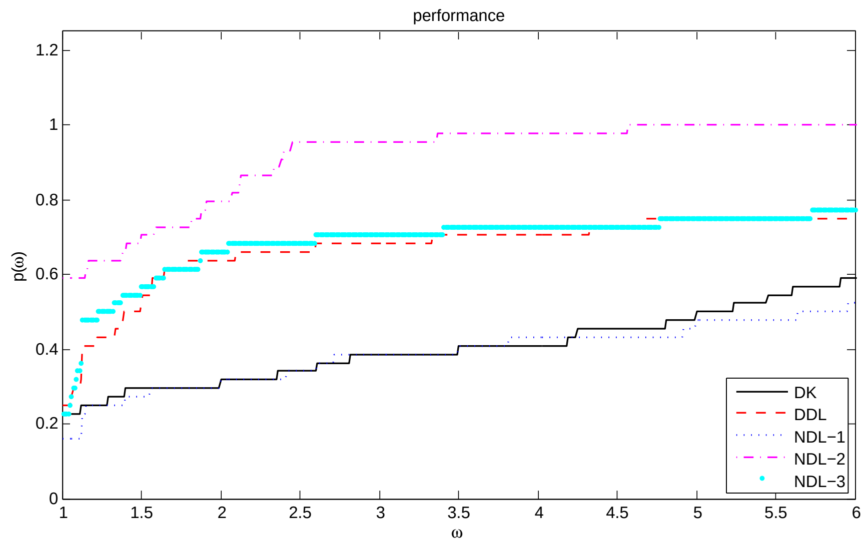

Figure 1: Performance profiles based on the number of iteration for "$NDL-i$", $i = 1, 2, 3$, DDL and DK methods.

On the whole, Figures 1-3, show that the "$NDL-2$" method outperforms all methods, contains "$NDL-1$", "$NDL-3$", DDL and DK, with respect to all criterias.

5 Conclusion

The DL approach has been used to provide a new conjugacy condition using the modified secant equation [31]. To adjust the parameter of the new conjugacy condition, $\overline{DL}$ parameter, two approaches are used. The convergence analysis is presented for uniformly convex and general nonlinear functions. The comparison of the new NCGs with some well-known methods, shows that "$NDL-2$" method is the best in the all iteration criterias.

References

[1] Z. Aminifard and S. Babaie-Kafaki, *An optimal parameter choice for the Dai Liao family of conjugate gradient methods by avoiding a direction of the maximum magnification by the search direction matrix*, 4OR. (2018) 1-14.

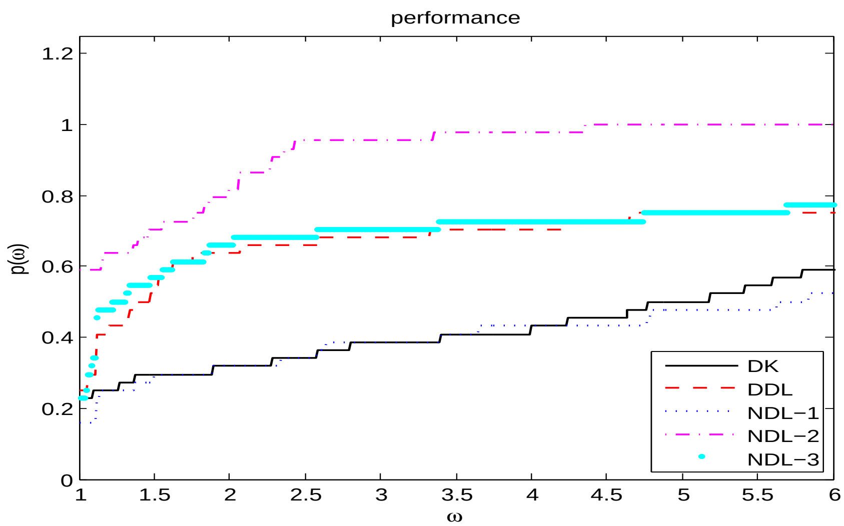

Figure 2: Performance profiles based on the $n_f + 3n_g$ for "$NDL - i,\ i = 1, 2, 3$", DDL and DK methods.

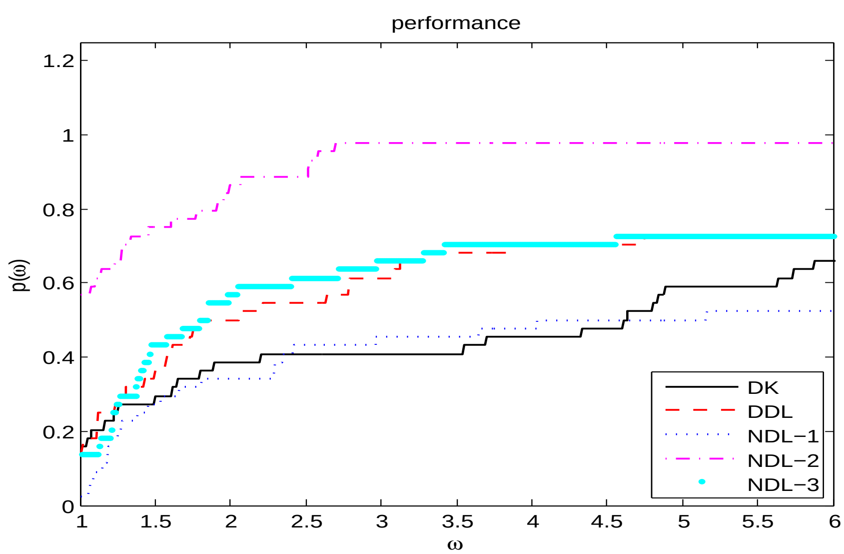

Figure 3: Performance profiles based on CPU time for "$NDL - i,\ i = 1, 2, 3$", DDL and DK methods.

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

[3] S. Babaie-Kafaki and R. Ghanbari, *A descent family of Dai Liao conjugate gradient methods*, Optim. Methods Softw. **3**(29) (2014) 583-591.

[4] S. Babaie-Kafaki and R. Ghanbari, *A hybridization of the Hestenes Stiefel and Dai Yuan conjugate gradient methods based on a least-squares approach*, Optim. Methods Softw. **4**(30) (2015) 673-681.

[5] S. Babaie-Kafaki and R. Ghanbari, *The Dai Liao nonlinear conjugate gradient method with optimal parameter choices*, European J. Oper. Res. **3**(234) (2014) 625-630.

[6] S. Babaie-Kafaki and R. Ghanbari, *Two optimal Dai Liao conjugate gradient methods*, Optimization **11**(64) (2015) 2277-2287.

[7] S. Babaie-Kafaki, R. Ghanbari and N. Mahdavi-Amiri, *Two new conjugate gradient methods based on modified secant equations*, J. Comput. Appl. Math. **5**(234) (2010) 1374-1386.

[8] Y.H. Dai and C.X. Kou, *A nonlinear conjugate gradient algorithm with an optimal property and an improved Wolfe line search*, SIAM J. Optim. **1**(23) (2013) 296 320.

[9] Y.H. Dai, J. Han, G. Liu, D. Sun, H. Yin and Y.X. Yuan, *Convergence properties of nonlinear conjugate gradient methods*, SIAM J. Optim. **2**(10) (2000) 345-358.

[10] Y.H. Dai and L.Z. Liao, *New conjugacy conditions and related nonlinear conjugate gradient methods*, Appl. Math. Optim. **1**(43) (2001) 87-101.

[11] Y.-H. Dai and Y. Yuan, P, *A nonlinear conjugate gradient method with a strong global convergence property*, SIAM J. Optim. **1**(10) (1999) 177-182.

[12] E.D. Dolan and J.J. Mor, *Benchmarking optimization software with performance profiles*, Math. Prog. **2**(91) (2002) 201-213.

[13] R. Fletcher and C.M. Reeves, *Function minimization by conjugate gradients*, The comput. J. **2**(7) (1964) 149-154.

[14] J.A. Ford and I.A. Moghrabi, *Alternative parameter choices for multistep Quasi-Newton methods*, Optim. Methods Softw. **2**(3) (1993) 357-370.

[15] J.A. Ford and I.A. Moghrabi, *Multi-step quasi-Newton methods for optimization*, J. Comput. Appl. Math. **1**(50) (1994) 305-323.

[16] N.I.M. Gould, D. Orban and P.L. Toint, *CUTER : a constrained and unconstrained testing environment*, ACM Trans. Math. Software. **4**(29) (2003) 373-394.

[17] W.W. Hager and H. Zhang, *A new conjugate gradient method with guaranteed descent and an efficient line search*, SIAM J. Optim. **1**(16) (2005) 170-192.

[18] W.W. Hager and H. Zhang, *A survey of nonlinear conjugate gradient methods*, Pac. J. Optim. **1**(2) (2006) 35-58.

[19] M.R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, NBS, Washington DC, 1952.

[20] D.H. Li and M. Fukushima, *A modified BFGS method and its global convergence in nonconvex minimization*, J. Comput. Appl. Math. **129**(1) (2001) 15-35.

[21] D.H. Li and M. Fukushima, *On the global convergence of the BFGS method for nonconvex unconstrained optimization problems*, SIAM J. Optim. **4**(11) (2001) 1054-1064.

[22] A. Neculai, *Open problem in conjugate gradient algorithms for unconstrained optimization*, Bull. Malays. Math. Sci. Soc. **2**(34) (2011) 319-330.

[23] J. Nocedal and S.J. Wright, *Numerical Optimization*, Springer, New York, NY, USA, Second edition, 2006.

[24] A. Perry, *A modified conjugate gradient algorithm*, Oper. Res. **6**(26) (1978) 1073-1078.

[25] B.T. Polyak, *The conjugate gradient method in extremal problems*, USSR Comput. Math. & Math. Phys. **4**(9) (1969) 94-112.

[26] E. Polak and G. Ribiere, *Note sur la convergence de mthodes de directions conjugues*, ESAIM: Mathematical Modelling and Numerical Analysis-Modlisation Mathmatique et Analyse Numrique **3** (1969) 35-43.

[27] M.J.D. Powell, *Nonconvex minimization calculations and the conjugate gradient method*, Griffiths DF (ed) Numerical analysis (Dundee, 1983)

volume 1066 of Lecture Notes in Math. Springer, Berlin. (1984) 122-141.

[28] D.F. Shanno, *Conjugate gradient methods with inexact searches*, Math. Oper. Res. **3(3)** (1978) 244-256.

[29] W. Sun and Y.X. Yuan, *Optimization theory and methods nonlinear programming*, Springer, New-York, 2006.

[30] P. Wolfe, *Convergence conditions for ascent methods*, SIAM review **2**(11) (1969) 226-235.

[31] Y.-X. Yuan, *A modified BFGS algorithm for unconstrained optimization*, IMA J. Numer. Anal. **3**(11) (1991) 325-332.

[32] W. Zengxin, G. Li and L. Qi, *New quasi-Newton methods for unconstrained optimization problems*, Appl. Math. Comput. **175** (2006) 156-1188.

[33] Z. Wei, G. Yu, G. Yuan and Z. Lian, *The Superlinear convergence of a modified BFGS-type method for unconstrained optimization*, Comput. Optim. Appl. **3**(29) (2004) 315-332.

[34] G. Li, C. Tang and Z. Wei, *New conjugacy condition and related new conjugate gradient methods for unconstrained optimization*, J. Comput. Appl. Math. **2**(202) (2007) 523-539.

[35] H. Yabe and M. Takano, *Global convergence properties of nonlinear conjugate gradient methods with modified secant condition*, Comput. Optim. Appl. **2**(28) (2004) 203-225.

[36] 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.

[37] K. Zhang, H. Liu and Z. Liu, *A new Dai-Liao conjugate gradient method with optimal parameter choice*, Numer. Funct. Anal. Optim. **40**(2) (2019) 194-215.

[38] L. Zhang and W. Zhou, *A nonlinear conjugate gradient method based on the MBFGS secant condition*, Optim. Methods Softw. **5**(21) (2006) 707-714.

[39] L. Zhang, W. Zhou and D. Li, *Some descent three-term conjugate gradient methods and their global convergence*, Optim. Methods Softw. **4**(22) (2007) 697-711.