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

Saeed Nezhadhosein*

Department of Mathematics, Payame Noor University, Tehran 193953697, Iran
Email: s_nezhadhosein@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:

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

(1)
where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ 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:

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

(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 $\alpha_k$ 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 $\alpha_k$ 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:

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k,$$

(3)

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k,$$

(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:

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

(5)

where $\beta_k$ 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, $\beta_k$, 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]:

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

where

$$\beta_{HS}^k = \frac{g_k^T y_k}{d_k^T y_k}, \quad \theta_{k+1} = \frac{g_{k+1}^T d_k}{d_k^T y_k}.$$  

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:

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

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 $\beta^P_k$ is the Perry CG parameter, which are defined as follows:

$$
\beta^P_k = \frac{g^T_{k+1} y_k}{d^T_k y_k} \frac{g^T_{k+1} s_k}{d^T_k y_k}, \quad Q^P_{k+1} = I - \frac{s_k y_k^T}{y_k^T s_k} + \frac{s_k s_k^T}{y_k^T s_k}.
$$

(9)

Note that using Wolfe conditions, Eqs. (3)-(4), we have $s^T_k y_k > 0$, so the matrix $Q^P_{k+1}$ in (9) is well-defined. In Perry approach, the direction matrix, $Q^P_{k+1}$, 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:

$$
Q^S_{k+1} = I - \frac{s_k y_k^T}{s_k^T y_k} + \left(1 + \frac{y_k^T y_k}{s_k^T y_k}\right) \frac{s_k s_k^T}{s_k^T y_k}.
$$

(10)

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

$$
d^{DL}_{k+1} = -g_{k+1} + \beta^{DL}_k d_k = -Q^{DL}_{k+1} g_{k+1},
$$

(11)

where

$$
\beta^{DL}_k = \frac{g^T_{k+1} y_k}{d^T_k y_k} - t \frac{g^T_{k+1} s_k}{d^T_k y_k}, \quad Q^{DL}_{k+1} = I - \frac{s_k y_k^T}{s_k^T y_k} + t \frac{s_k s_k^T}{s_k^T y_k},
$$

(12)

where $t$ is a nonnegative scalar and called the DL parameter. Note that if $t = 0$, then $\beta^{DL}_k$ reduces to $\beta^{HS}_k$, Eq. (7), and if $t = 1$, then $\beta^{DL}_k$ reduces to $\beta^P_k$, 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]:

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

(13)

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

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

(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 $\beta_k^{DL+}$, and following direction:

$$d_{k+1}^{DL+} = -g_{k+1} + \beta_k^{DL+} d_k$$

$$= -g_{k+1} + \left( \max \left\{ \frac{d_{k+1}^T y_k}{d_k^T y_k}, 0 \right\} - t \frac{g_{k+1}^T s_k}{d_k^T s_k} \right) d_k, \quad (15)$$

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:

$$\beta_k^{HZ} = \frac{g_{k+1}^T y_k}{d_k^T y_k} - 2 \left\| \frac{y_k}{d_k^T y_k} \right\|_2$$

$$\left( \frac{g_{k+1}^T}{d_k^T} d_k \right), \quad (16)$$

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

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

in which $\tau_k$ 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:

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

where

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

and

$$z_k \triangleq y_k + c \|g_k\|^r s_k, \quad (20)$$
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, $Q_{k+1}^S$ 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:

\[ B_{k+1}s_k = \bar{y}_k, \quad \bar{y}_k = y_k + A_ku_k, \]  

(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 \( s_k^Tu_k \neq 0 \) and \( A_k = \frac{\theta_k}{s_k^Tu_k} \) where \( \theta_k \) is as follows:

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

(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:

\[ d_k^T y_k = -tDL g_{k+1}s_k, \]  

(23)

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

\[ \beta_k^{DL} = \frac{g_{k+1}^T \bar{y}_k}{d_k^T \bar{y}_k} - tDL \frac{g_{k+1}^T s_k}{d_k^T s_k}, \]  

(24)

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

\[ d_{k+1}^{DL} = -Q_k^{DL}g_{k+1} = -(I + \frac{s_k \bar{y}_k^T}{s_k^T \bar{y}_k} - tDL \frac{s_k s_k^T}{s_k^T \bar{y}_k})g_{k+1}. \]  

(25)

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

### 2.2 Setting the \( DL \) parameter

To set the \( DL \) 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 \( DL \) 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 \( DL \) search direction to the MTTHS direction in (18)-(19), consider the following subproblem:

\[
\ell_{DL}^* = \arg\min \| d_{k+1}^{DL} - d_{k+1}^{MTTHS} \|, \tag{26}
\]

where \( \| . \| \) is the Euclidean norm. By taking simple algebraic calculations, we get the \( DL \) parameter as follows:

\[
t_{DL}^k = \frac{1}{a_2} (a_1 - a_3 + \frac{a_4}{\|d_k\|^2} d_k^T z_k), \tag{27}
\]

where

\[
\begin{align*}
a_1 &= \frac{g_{k+1}^T \bar{y}_k}{d_k^T \bar{y}_k}, \\
a_2 &= \frac{g_{k+1}^T s_k}{d_k^T \bar{y}_k}, \\
a_3 &= \frac{g_{k+1}^T z_k}{d_k^T z_k}, \\
a_4 &= \frac{g_{k+1}^T d_k}{d_k^T z_k},
\end{align*}
\]

where \( z_k \) and \( \bar{y}_k \) 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 \( d_k^T \bar{y}_k > 0 \). Therefore, the fractional expressions in Eq. (27) are well-defined. After some simplification, the Eq. (27) can be written as follows:

\[
t_{DL}^k = - \frac{\bar{y}_k^T g_{k+1}}{s_k^T g_{k+1}}. \tag{28}
\]

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

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

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

\[
d_{k+1}^{NDL-1} = -g_{k+1} + (\frac{g_{k+1}^T \bar{y}_k}{d_k^T \bar{y}_k} - t_{DL}^* \frac{g_{k+1}^T s_k}{d_k^T \bar{y}_k}) d_k, \tag{30}
\]

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, $Q_{\overline{DL}}^{k+1}$, in Eq. (25), to the Shanno quasi-Newton search direction matrix, $Q^S_{k+1}$, Eq. (10). Therefore, consider the following subproblem:

$$t_{DL}^* = \arg\min \|Q_{\overline{DL}}^{k+1} - Q^S_{k+1}\|_F,$$

(31)

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

$$t_{DL}^* = 1 + \frac{\bar{y}_k^T \bar{y}_k}{s_k^T \bar{y}_k} - \frac{s_k^T \bar{y}_k}{\|s_k\|^2}.$$  

(32)

Note that using Wolfe conditions, we have $s_k^T \bar{y}_k > 0$ 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:

$$d^{NDL-2}_{k+1} = -g_{k+1} + (\frac{g_{k+1}^T \bar{y}_k}{d_k^T \bar{y}_k} - t_{DL}^* \bar{y}_k^T \bar{y}_k) d_k,$$

(33)

where $t_{DL}^*$ is defined in Eq. (32). The NCG method based on $d^{NDL-2}_{k+1}$ 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 $\bar{y}_k$ in Eq. (21), we have the following search direction:

$$d^{NDL-3}_{k+1} = -(I + \frac{d_k^T \bar{y}_k}{d_k^T \bar{y}_k} - t_k^{DL*(p,q)} \frac{d_k^T s_k}{d_k^T \bar{y}_k}) g_{k+1} = -Q_{k+1}^{DL} g_{k+1},$$

(34)

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

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

(35)

By Wolfe conditions we have $s_k^T \bar{y}_k > 0$, 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 \(\nabla f\) is Lipschitz continuous on the level set

\[
S = \{ x \in \mathbb{R}^n : f(x) \leq f(x_0) \}
\]

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

\[
(\nabla f(x) - \nabla f(y))^T(x - y) \geq \mu \|x - y\|^2, \quad \forall x, y \in S
\]

and

\[
\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \quad \forall x, y \in S
\]

Note that these assumptions implies that there exists a positive constant \(\Gamma\) 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 \(\alpha_k\) is determined to satisfy the Wolfe conditions, (3)-(4). If

\[
\sum_{k \geq 0} \frac{1}{\|d_k\|^2} = \infty,
\]

then the method converges in the sense that

\[
\liminf_{k \to \infty} \|g_k\| = 0.
\]

**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 \(\alpha_k\) is computed such that the Wolfe conditions (3)-(4) are satisfied. If the objective function \(f\) is uniformly convex on \(S\), 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 $\mu$ such that (see Theorem 1.3.16 of [29])

$$y_k^Ts_k \geq \mu\|s_k\|^2.$$ \hfill (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:

$$\overline{y}_k^Ts_k \geq \overline{\mu}\|s_k\|^2.$$ \hfill (42)

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

$$g_{k+1}^Td_k \geq \sigma g_k^Td_k.$$ \hfill (43)

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

$$\|y_k\| \leq \|y_k\| + \|A_ku_k\| = \|y_k\| + \|w_k\|,$$ \hfill (44)

where $w_k = A_ku_k$ 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 $\theta_k$ in Eq. (22), we have:

$$|\theta_k| < M\|s_k\|^2.$$ \hfill (45)

Then, considering the Eq. (21), we have two cases for $w_k$ [20]: $w_k = \frac{\theta_ky_k}{\|s_k\|^2}$ or $w_k = \frac{\theta_ks_k}{s_k^Ty_k}$, which $\theta_k$ is defined in (22). In the first case, from Eqs. (38), (41) and (45), we get:

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

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

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

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

$$\|\overline{y}_k\| \leq L_1\|s_k\|.$$ \hfill (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:

$$\|z_k\| = \|y_k + c\|g_k\|^Ts_k\| \leq \|y_k\| + c\|g_k\|^T\|s_k\| \leq L\|s_k\| + c\|g_k\|^T\|s_k\|$$

$$\leq (L + c\Gamma)\|S_k\| = L_2\|s_k\|,$$ \hfill (49)
where $L_2 = L + c\Gamma'$. Moreover, from (41) and (20) we have:

$$s^T_k z_k = s^T_k (y_k + c\|g_k\|s_k) = s^T_k y_k + c\|g_k\|s_k^2 \geq (\mu + c\|g_k\|)\|s_k\|^2 \geq \mu\|s_k\|^2,$$

(50)

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

$$|t^D_{DL}*| = \frac{d_k^T \bar{y}_k}{g_{k+1}^T s_k} \left( \frac{g_{k+1}^T \bar{y}_k}{d_k^T \bar{y}_k} + \frac{z_k}{s_k^T d_k} \right) \leq \frac{\|s_k\|\|\bar{y}_k\|}{\sigma\|g_{k+1}\|\|s_k\|} \left( \frac{\|g_{k+1}\|\|\bar{y}_k\|}{\|g_{k+1}\|\|s_k\|} + \frac{\|g_{k+1}\|\|z_k\|}{\mu\|s_k\|^2} + \frac{\|g_{k+1}\|\|s_k\|}{\|s_k\|^2} \right) \leq \frac{L_1}{\sigma} \left( \frac{L_1}{\beta} + \frac{L_2}{\mu} + 1 \right).$$

(51)

That is $t^D_{DL}* \leq L_1 \cdot \frac{L_1}{\beta} + \frac{L_2}{\mu} + 1$. 

**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 $\alpha_k$ is computed such that the Wolfe conditions (3)-(4) are satisfied. If the objective function $f$ is uniformly convex on $S$, 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 $t^D_{DL}* \leq L_1 \cdot \frac{L_1}{\beta} + \frac{L_2}{\mu} + 1$. 

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 $\alpha_k$ is computed such that the Wolfe conditions (3)-(4) are satisfied. If the objective function $f$ is uniformly convex on $S$, 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 \( t_{k}^{DL^{*}(p,q)} \), Eq. (35), where \( p < \frac{1}{4} \) and \( q \geq \frac{1}{4} \), we have:

\[
|s_{k}^{DL^{*}(p,q)}| = |p\frac{\|s_{k}^{T}\|}{s_{k}^{T}s_{k}^{T}} - q\frac{s_{k}^{T}y_{k}}{s_{k}^{T}s_{k}^{T}}| \leq p\frac{\|s_{k}^{T}\|}{s_{k}^{T}s_{k}^{T}} + q\frac{|s_{k}^{T}y_{k}|}{s_{k}^{T}s_{k}^{T}} \leq p\frac{L_{1}^{2}}{\bar{\mu}} + q\bar{\mu},
\]

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, \( t_{k}^{DL^{+}} \). 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.

<table>
<thead>
<tr>
<th>Name</th>
<th>Dim</th>
<th>Name</th>
<th>Dim</th>
<th>Name</th>
<th>Dim</th>
</tr>
</thead>
<tbody>
<tr>
<td>AKIVA</td>
<td>2</td>
<td>DIXMAANH</td>
<td>3000</td>
<td>LIARWHD</td>
<td>5000</td>
</tr>
<tr>
<td>ALLINITU</td>
<td>4</td>
<td>DIXMAANI</td>
<td>3000</td>
<td>LOGHAIRY</td>
<td>2</td>
</tr>
<tr>
<td>ARGWINA</td>
<td>200</td>
<td>DIXMAANJ</td>
<td>3000</td>
<td>MANCINO</td>
<td>100</td>
</tr>
<tr>
<td>ARGWINB</td>
<td>200</td>
<td>DIXMAANK</td>
<td>3000</td>
<td>MARATOSB</td>
<td>2</td>
</tr>
<tr>
<td>ARGWINC</td>
<td>200</td>
<td>DIXMAANL</td>
<td>3000</td>
<td>MEXHAT</td>
<td>2</td>
</tr>
<tr>
<td>ARWHEAD</td>
<td>5000</td>
<td>DIXMAANM</td>
<td>15</td>
<td>NONCVXU2</td>
<td>5000</td>
</tr>
<tr>
<td>BARD</td>
<td>3</td>
<td>DIXMAANN</td>
<td>15</td>
<td>NONCVXUN</td>
<td>5000</td>
</tr>
<tr>
<td>BDQRTIC</td>
<td>5000</td>
<td>DIXMAANO</td>
<td>15</td>
<td>NONDIA</td>
<td>5000</td>
</tr>
<tr>
<td>BEALE</td>
<td>2</td>
<td>DIXMAANP</td>
<td>15</td>
<td>NONDQUAR</td>
<td>5000</td>
</tr>
<tr>
<td>BEALE</td>
<td>2</td>
<td>DIXON3DQ</td>
<td>10000</td>
<td>OSBORNEA</td>
<td>5</td>
</tr>
<tr>
<td>BIGGS6</td>
<td>6</td>
<td>DJTL</td>
<td>2</td>
<td>PALMER1C</td>
<td>8</td>
</tr>
<tr>
<td>BOX</td>
<td>10000</td>
<td>DQDRTIC</td>
<td>5000</td>
<td>PALMER2C</td>
<td>8</td>
</tr>
<tr>
<td>BOX4</td>
<td>3</td>
<td>DQRTIC</td>
<td>5000</td>
<td>PALMER3C</td>
<td>8</td>
</tr>
<tr>
<td>BRKMKCC</td>
<td>2</td>
<td>EDENSCH</td>
<td>2000</td>
<td>PALMER4C</td>
<td>8</td>
</tr>
<tr>
<td>BROWNAL</td>
<td>200</td>
<td>EG2</td>
<td>1000</td>
<td>PALMER5C</td>
<td>6</td>
</tr>
<tr>
<td>BROWNDEN</td>
<td>4</td>
<td>ENGVAL1</td>
<td>5000</td>
<td>PALMER6C</td>
<td>8</td>
</tr>
<tr>
<td>BROYDN7D</td>
<td>5000</td>
<td>ENGVAL2</td>
<td>3</td>
<td>PALMER6C</td>
<td>8</td>
</tr>
<tr>
<td>BRYBND</td>
<td>5000</td>
<td>ERRINROS</td>
<td>50</td>
<td>PALMER8C</td>
<td>8</td>
</tr>
<tr>
<td>CHAINWOO</td>
<td>4000</td>
<td>ERRINRSM</td>
<td>50</td>
<td>HIMMELBF</td>
<td>4</td>
</tr>
<tr>
<td>CHNROSNB</td>
<td>50</td>
<td>EXPFIT</td>
<td>2</td>
<td>HIMMELBG</td>
<td>2</td>
</tr>
<tr>
<td>CHNRSNB</td>
<td>50</td>
<td>EXTROSNB</td>
<td>10000</td>
<td>HIMMELBH</td>
<td>2</td>
</tr>
<tr>
<td>CLIFF</td>
<td>2</td>
<td>FLETBV3M</td>
<td>5000</td>
<td>POWELLSG</td>
<td>5000</td>
</tr>
<tr>
<td>CLIFF</td>
<td>2</td>
<td>FLETCBV</td>
<td>5000</td>
<td>POWER</td>
<td>10000</td>
</tr>
<tr>
<td>CUBE</td>
<td>2</td>
<td>FLETCCHR</td>
<td>1000</td>
<td>QUARTC</td>
<td>5000</td>
</tr>
<tr>
<td>CURLY10</td>
<td>10000</td>
<td>FMINSRF2</td>
<td>5625</td>
<td>ROSENBR</td>
<td>2</td>
</tr>
<tr>
<td>CURLY20</td>
<td>10000</td>
<td>FMINSURF</td>
<td>5625</td>
<td>SINEVAL</td>
<td>2</td>
</tr>
<tr>
<td>CURLY30</td>
<td>10000</td>
<td>FREUROTH</td>
<td>5000</td>
<td>SISSER</td>
<td>2</td>
</tr>
<tr>
<td>DECONVU</td>
<td>63</td>
<td>GENHUMPS</td>
<td>5000</td>
<td>SPARISNE</td>
<td>5000</td>
</tr>
<tr>
<td>DENSCHNA</td>
<td>2</td>
<td>GENROSE</td>
<td>5000</td>
<td>SPARSQUR</td>
<td>10000</td>
</tr>
<tr>
<td>DENSCHNB</td>
<td>2</td>
<td>GULF</td>
<td>3</td>
<td>SPMSRTLS</td>
<td>4999</td>
</tr>
<tr>
<td>DENSCHNC</td>
<td>2</td>
<td>Hairy</td>
<td>2</td>
<td>SROSENBR</td>
<td>5000</td>
</tr>
<tr>
<td>DENSCHND</td>
<td>3</td>
<td>HATFLDD</td>
<td>3</td>
<td>SSCOSINE</td>
<td>10</td>
</tr>
<tr>
<td>DENSCHNE</td>
<td>3</td>
<td>HATFLDE</td>
<td>3</td>
<td>TESTQUAD</td>
<td>5000</td>
</tr>
<tr>
<td>DENSCHNF</td>
<td>2</td>
<td>HATFLDFL</td>
<td>3</td>
<td>TOINTGOR</td>
<td>50</td>
</tr>
<tr>
<td>DIXMAANC</td>
<td>3000</td>
<td>HEART6LS</td>
<td>6</td>
<td>TOINTPSP</td>
<td>50</td>
</tr>
<tr>
<td>DIXMAANA</td>
<td>3000</td>
<td>HEART8LS</td>
<td>8</td>
<td>TOINTQOR</td>
<td>50</td>
</tr>
<tr>
<td>DIXMAANB</td>
<td>3000</td>
<td>HELIX</td>
<td>3</td>
<td>TQUARTIC</td>
<td>5000</td>
</tr>
<tr>
<td>DIXMAANC</td>
<td>3000</td>
<td>HILBERTA</td>
<td>2</td>
<td>TRIDIA</td>
<td>5000</td>
</tr>
<tr>
<td>DIXMAAND</td>
<td>3000</td>
<td>HILBERTB</td>
<td>10</td>
<td>VARDIM</td>
<td>20</td>
</tr>
<tr>
<td>DIXMAANE</td>
<td>3000</td>
<td>HIMMELBB</td>
<td>2</td>
<td>WOODS</td>
<td>4000</td>
</tr>
<tr>
<td>DIXMAANF</td>
<td>3000</td>
<td>JENSMP</td>
<td>2</td>
<td>YFITU</td>
<td>3</td>
</tr>
<tr>
<td>DIXMAANG</td>
<td>3000</td>
<td>KOWOSB</td>
<td>4</td>
<td>ZANGWIL2</td>
<td>2</td>
</tr>
</tbody>
</table>

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.
New nonlinear conjugate gradient methods based on ...

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, 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

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.


New nonlinear conjugate gradient methods based on ...


