# AN EXTENDED DAI–LIAO CONJUGATE GRADIENT METHOD WITH GLOBAL CONVERGENCE FOR NONCONVEX FUNCTIONS

Mohammad Reza Arazm, Saman Babaie–Kafaki* and Reza Ghanbari

Semnan University, Ferdowsi University of Mashhad, Iran

Abstract. Using an extension of some previously proposed modified secant equations in the Dai–Liao approach, a modified nonlinear conjugate gradient method is proposed. As interesting features, the method employs the objective function values in addition to the gradient information and satisfies the sufficient descent property with proper choices for its parameter. Global convergence of the method is established without convexity assumption on the objective function. Results of numerical comparisons are reported. They demonstrate efficiency of the proposed method in the sense of the Dolan–Moré performance profile.

## 1. Introduction

We consider the minimization of a smooth nonlinear function $f : \mathbb{R}^n \to \mathbb{R}$, that is,

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

in the case where the number of variables $n$ is large and where analytic expressions for the function $f$ and its gradient $\nabla f$ are available. Among the most useful tools for solving the problem are the limited memory quasi–Newton methods ([24]) and the conjugate gradient (CG) methods ([20]), because the amount of memory storage required by the methods is low. This study is devoted to a class of CG methods constructed based on the secant (quasi–Newton) equations ([3, 31]).

In addition to the low memory requirement, CG methods possess the attractive feature of simple iterative formula, that is,

(1.1)          $x_0 \in \mathbb{R}^n, \ x_{k+1} = x_k + s_k, \ s_k = \alpha_k d_k, \ k = 0, 1, ...,$

where $\alpha_k$ is a steplength to be computed by a line search procedure along the search direction $d_k$ defined by

(1.2)          $d_0 = -g_0, \ d_{k+1} = -g_{k+1} + \beta_k d_k, \ k = 0, 1, ...,$

in which $g_k = \nabla f(x_k)$ and $\beta_k$ is a scalar called the CG (update) parameter.

Different CG methods mainly correspond to different choices for the CG parameter for which a nice review has been presented in [20]. Although CG methods are equivalent in the linear case, that is, when $f$ is a strictly convex quadratic function and $\alpha_k$ is computed by an exact line search, their numerical behavior for general functions may be quite different ([1, 20, 25, 26]).

The interesting feature of linear CG methods leading to the important at most $n$–step termination property is the generation of search directions $d_k$, $k \geq 0$, satisfying the conjugacy condition, i.e.,

(1.3)          $d_i^T G d_j = 0, \ \forall i \neq j,$

where $G$ is the Hessian of the objective function ([28]). For a general nonlinear function $f$, from the mean–value theorem we know that there exists some $\tau \in (0, 1)$ such that

$$d_{k+1}^T(g_{k+1} - g_k) = \alpha_k d_{k+1}^T \nabla^2 f(x_k + \tau s_k) d_k.$$

Hence, in order to compute the CG parameter $\beta_k$ in a nonlinear CG method in the form of (1.1)–(1.2), it is reasonable to replace (1.3) with the following conjugacy condition:

(1.4)          $d_{k+1}^T y_k = 0,$

where $y_k = g_{k+1} - g_k$. The conjugacy condition (1.4) is effective since it leads to the efficient CG method proposed by Hestenes and Stiefel ([21]) (HS) in which, considering (1.2) and (1.4), the CG parameter is given by

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k}.$$

In order to employ quasi–Newton aspects in the conjugacy condition (1.4), Dai and Liao ([12]) (DL) noted that if we consider $B_{k+1} \in \mathbb{R}^{n \times n}$ as an approximation of $\nabla^2 f(x_{k+1})$ given by a quasi–Newton method, since $B_{k+1}$ satisfies the standard secant equation, i.e.,

(1.5)          $B_{k+1} s_k = y_k,$

and the search direction is computed by solving the following system of linear equations:

$$B_{k+1} d_{k+1} = -g_{k+1},$$

then we can write

$$(1.6) \qquad d_{k+1}^T y_k = d_{k+1}^T (B_{k+1} s_k) = -g_{k+1}^T s_k.$$

Now, if the line search is exact, then $g_{k+1}^T s_k = 0$, and consequently, (1.6) reduces to (1.4). However, practical numerical algorithms normally adopt inexact line searches. Hence, the following extension of the conjugacy condition (1.4) has been proposed in [12]:

$$(1.7) \qquad d_{k+1}^T y_k = -t g_{k+1}^T s_k,$$

where $t$ is a nonnegative parameter. If $t = 0$ or the line search is exact, then (1.7) reduces to (1.4), and if $t = 1$, then (1.7) reduces to (1.6) which implicitly contains the effective standard secant equation (1.5). Also, for small values of $t$, the conjugacy condition (1.7) tends to the conjugacy condition (1.4). Thus, the conjugacy condition (1.7) can be regarded as a hybridization of the conjugacy conditions (1.4) and (1.6).

Taking inner product of (1.2) with $y_k$ and using (1.7), Dai and Liao ([12]) obtained the following formula for the CG parameter:

$$\beta_k^{DL} = \frac{g_{k+1}^T y_k}{d_k^T y_k} - t \frac{g_{k+1}^T s_k}{d_k^T y_k}.$$

The DL method is very dependent on the parameter $t$ for which there is no any optimal choice ([2]). It is worth noting that if

$$(1.8) \qquad t = 2 \frac{||y_k||^2}{s_k^T y_k},$$

where $||.||$ denotes the Euclidean norm, then the CG parameter proposed by Hager and Zhang ([18]) is achieved. Also, the choice

$$(1.9) \qquad t = \frac{||y_k||^2}{s_k^T y_k},$$

yields another CG parameter suggested by Dai and Kou ([11]). The choices (1.8) and (1.9) are effective since they guarantee the sufficient descent condition, i.e.,

$$(1.10) \qquad g_k^T d_k \leq -\varrho ||g_k||^2, \ k = 0, 1, ...,$$

where $\varrho$ is a positive constant, independent of the line search and the objective function convexity, and lead to numerically efficient CG methods ([4, 11, 19]). Recently, Babaie–Kafaki and Ghanbari ([5,6,9]) proposed several other choices for the parameter $t$ in the DL method.

The Dai–Liao approach has been attracted special attention. Note that in the DL method only the gradient information is used. Motivated by this, in several efforts employing the objective function values has been focused on, using modified secant equations. For example, Yabe and Takano ([32]) used the modified secant equation proposed by Zhang et al. ([36]). Li et

al. ([23]) applied the modified secant equation proposed by Wei et al. ([29]). Ford et al. ([15]) used the multi–step quasi–Newton equations proposed by Ford et al. ([14]). Babaie–Kafaki et al. ([10]) employed a revised form of the modified secant equation proposed by Zhang et al. ([36]) and the modified secant equation proposed by Yuan ([33]). Since the DL method is globally convergent for uniformly convex objective functions ([12]), in another effort Zhou and Zhang ([37]) applied the modified secant equation proposed by Li and Fukushima ([22]) and suggested a modified DL method which is globally convergent without convexity assumption on the objective function.

Here, based on an extension of the above–mentioned modified secant equations, we deal with another modification of the DL method which simultaneously employs the objective function values and is globally convergent without convexity assumption. This work is organised as follows. In Section 2, the method and its descent property are discussed in details. In Section 3, global convergence analysis of the method is studied. Comparative numerical comparisons to demonstrate efficiency of the method are made in Section 4. Finally, conclusions are drawn in Section 5.

## 2. A modified Dai–Liao conjugate gradient method

To the best of our knowledge, convexity assumption on the objective function plays an essential role in convergence analysis of the quasi–Newton methods (see [3] and the references therein). Nevertheless, Li and Fukushima ([22]) proposed a modified quasi–Newton method which is globally and locally superlinearly convergent without convexity assumption (see also [17, 37]). In the method, the following modified version of the standard secant equation (1.5) is applied:

$$(2.1) \qquad B_{k+1}s_k = \bar{y}_k, \ \bar{y}_k = y_k + h_k ||g_k||^r s_k,$$

in which $r$ is a positive constant and $h_k$ is defined by

$$h_k = C + \max\{-\frac{s_k^T y_k}{||s_k||^2}, 0\}||g_k||^{-r},$$

with some positive constant $C$. It is worth noting that independent of the line search and the objective function convexity, we have $s_k^T \bar{y}_k > 0$, being necessary to guarantee positive definiteness of successive approximations of the Hessian generated by the quasi–Newton methods. Also, if the popular Wolfe conditions ([30]) are employed in the line search procedure, that is,

$$(2.2) \qquad f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k \nabla f(x_k)^T d_k,$$
$$(2.3) \qquad \nabla f(x_k + \alpha_k d_k)^T d_k \geq \sigma \nabla f(x_k)^T d_k,$$

with $0 < \delta < \sigma < 1$, since $s_k^T y_k > 0$, then we have $h_k = C$. Hence, the modified secant equation (2.1) reduces to

$$(2.4) \qquad B_{k+1} s_k = y_k + C \|g_k\|^r s_k.$$

In another point of view, researchers paid attention to employ objective function values in the standard secant equation (1.5) which only applies the gradient information. For example, Zhang et al. ([36]) proposed the following modified secant equation:

$$(2.5) \qquad B_{k+1} s_k = \tilde{y}_k, \ \tilde{y}_k = y_k + \frac{\vartheta_k}{s_k^T u_k} u_k,$$

in which

$$\vartheta_k = 6(f_k - f_{k+1}) + 3 s_k^T (g_k + g_{k+1}),$$

and $u_k \in \mathbb{R}^n$ is a vector parameter satisfying $s_k^T u_k \neq 0$ (see also [34, 35]). Also, Wei et al. ([29]) proposed the following modified secant equation:

$$(2.6) \qquad B_{k+1} s_k = \hat{y}_k, \ \hat{y}_k = y_k + \frac{\theta_k}{s_k^T u_k} u_k,$$

where

$$(2.7) \qquad \theta_k = 2(f_k - f_{k+1}) + s_k^T (g_k + g_{k+1}),$$

and $u_k \in \mathbb{R}^n$ is a vector parameter satisfying $s_k^T u_k \neq 0$ (see also [33]). Note that in the modified secant equations (2.5) and (2.6) we can let $u_k = s_k$, or in the case of using the Wolfe line search conditions, since $s_k^T y_k > 0$, we can let $u_k = y_k$. It is important that in the modified secant equations (2.5) and (2.6) the curvature $s_k^T y_k$ is locally approximated more accurate than the standard secant equation (1.5).

Here, we suggest an extension of the secant equations (1.5), (2.4), (2.5) and (2.6) as follows:

$$(2.8) \qquad B_{k+1} s_k = w_k, \ w_k = y_k + \xi \frac{\theta_k}{s_k^T u_k} u_k + C \|g_k\|^r s_k,$$

in which $\xi$, $C$ and $r$ are nonnegative constants, $\theta_k$ is defined by (2.7), and $u_k \in \mathbb{R}^n$ is a vector parameter satisfying $s_k^T u_k \neq 0$. As seen, if $\xi = C = 0$, then (2.8) reduces to (1.5), and if $\xi = 0$ and $C > 0$, then (2.8) reduces to (2.4). Also, setting $C = 0$, the choices $\xi = 3$ and $\xi = 1$ respectively yield the modified secant equations (2.5) and (2.6).

Now, following Dai–Liao approach ([12]), based on the extended secant equation (2.8) we suggest the following CG parameter:

$$(2.9) \qquad \bar{\beta}_k^{EDL} = \frac{g_{k+1}^T w_k}{d_k^T w_k} - t \frac{g_{k+1}^T s_k}{d_k^T w_k},$$

where $t$ is a nonnegative parameter. Note that proper choices for the nonnegative parameters $\xi$ and $C$ in (2.9) yield the CG parameters proposed in [37],

[32] and [23], respectively obtained from the modified secant equations (2.4), (2.5) and (2.6).

Although the Wolfe conditions ensure that $d_k^T y_k > 0$, generally the denominator of $\bar{\beta}_k^{EDL}$, i.e.,

$$d_k^T w_k = d_k^T y_k + \xi \frac{\theta_k}{\alpha_k} + C\alpha_k ||g_k||^r ||d_k||^2,$$

may not be nonzero (positive). That is, the CG parameter $\bar{\beta}_k^{EDL}$ is not well–defined. Motivated by this, we suggest the following revised form of (2.8):

$$(2.10) \qquad B_{k+1}s_k = z_k, \ z_k = y_k + \xi \frac{\max\{\theta_k, 0\}}{s_k^T u_k} u_k + C||g_k||^r s_k,$$

with the vector parameter $u_k$ satisfying $s_k^T u_k \neq 0$, and consequently, based on the Dai–Liao approach we propose the following CG parameter:

$$(2.11) \qquad \beta_k^{EDL} = \frac{g_{k+1}^T z_k}{d_k^T z_k} - t\frac{g_{k+1}^T s_k}{d_k^T z_k}, \ t \geq 0,$$

being a revised version of (2.9). Note that for $C = 0$, the choices $\xi = 3$ and $\xi = 1$ in (2.11) respectively yield two CG parameters proposed in [10] and [23].

Next, we establish a descent property for the EDL method using the eigenvalue analysis conducted in [7]. Hereafter, we assume that the line search fulfills the Wolfe conditions (2.2) and (2.3). Also, in (2.10) we set $u_k = s_k$, for all $k \geq 0$.

THEOREM 2.1. *For a CG method in the form of (1.1)–(1.2) with the CG parameter $\beta_k^{EDL}$ defined by (2.11) in which*

$$(2.12) \qquad t = \rho_k \frac{||z_k||^2}{s_k^T z_k},$$

*with the real parameter $\rho_k > \dfrac{1}{4}$, we have*

$$d_{k+1}^T g_{k+1} \leq - \left(1 - \frac{1}{4\rho_k}\right) ||g_{k+1}||^2, \ \forall k \geq 0.$$

PROOF. Firstly, note that search directions of the method can be written as:

$$d_{k+1} = -Q_{k+1}g_{k+1}, \ k = 0, 1, ...,$$

where

$$Q_{k+1} = I - \frac{s_k z_k^T}{s_k^T z_k} + \rho_k \frac{||z_k||^2}{s_k^T z_k} \frac{s_k s_k^T}{s_k^T z_k},$$

and consequently,

$$(2.13) \qquad d_{k+1}^T g_{k+1} = -g_{k+1}^T Q_{k+1}^T g_{k+1} = -g_{k+1}^T \frac{Q_{k+1}^T + Q_{k+1}}{2} g_{k+1}.$$

Hence, we focus on finding eigenvalues of the following symmetric matrix:

$$A_{k+1} = \frac{Q_{k+1}^T + Q_{k+1}}{2} = I + \rho_k \frac{||z_k||^2}{s_k^T z_k} \frac{s_k s_k^T}{s_k^T z_k} - \frac{1}{2} \frac{s_k z_k^T + z_k s_k^T}{s_k^T z_k}.$$

Since $\alpha_k > 0$ and $d_k^T z_k > 0$, we have $s_k^T z_k \neq 0$ and consequently, $s_k \neq 0$ and $z_k \neq 0$. So, there exists a set of mutually orthogonal vectors $\{u_k^i\}_{i=1}^{n-2}$ such that

$$s_k^T u_k^i = z_k^T u_k^i = 0, \ ||u_k^i|| = 1, \ i = 1, ..., n-2,$$

which leads to

$$A_{k+1} u_k^i = u_k^i, \ i = 1, ..., n-2.$$

That is, the vectors $u_k^i$, $i = 1, ..., n-2$, are the eigenvectors of $A_{k+1}$ corresponding to the eigenvalue 1. Now, we find the two remaining eigenvalues of $A_{k+1}$, namely $\lambda_k^-$ and $\lambda_k^+$.

Since the trace of a square matrix is equal to the sum of its eigenvalues, by defining $\varsigma_k$ as:

$$\varsigma_k = \frac{||s_k||^2 ||z_k||^2}{(s_k^T z_k)^2},$$

we have

$$\text{tr}(A_{k+1}) = n - 1 + \rho_k \varsigma_k$$
$$= \underbrace{1 + ... + 1}_{(n-2) \text{ times}} + \lambda_k^- + \lambda_k^+,$$

which leads to

$$(2.14) \qquad \lambda_k^- + \lambda_k^+ = 1 + \rho_k \varsigma_k.$$

On the other hand, from properties of the Frobenius norm we have

$$||A_{k+1}||_F^2 = \text{tr}(A_{k+1}^T A_{k+1}) = \text{tr}(A_{k+1}^2)$$
$$= n - \frac{3}{2} + \frac{1}{2}\varsigma_k + \rho_k^2 \varsigma_k^2$$
$$= \underbrace{1 + ... + 1}_{(n-2) \text{ times}} + \lambda_k^{-2} + \lambda_k^{+2},$$

and so,

$$(2.15) \qquad \lambda_k^{-2} + \lambda_k^{+2} = \frac{1}{2} + \frac{1}{2}\varsigma_k + \rho_k^2 \varsigma_k^2.$$

Now, from (2.14) and (2.15) we get

$$(2.16) \qquad \lambda_k^- \lambda_k^+ = \frac{1}{4} + \left(\rho_k - \frac{1}{4}\right)\varsigma_k,$$

and consequently, from (2.14) and (2.16), $\lambda_k^-$ and $\lambda_k^+$ can be computed as the solutions of the following quadratic equation:

$$\lambda^2 - (1 + \rho_k \varsigma_k)\lambda + \frac{1}{4} + \left(\rho_k - \frac{1}{4}\right)\varsigma_k = 0.$$

More precisely,

$$\lambda_k^\pm = \frac{1 + \rho_k \varsigma_k \pm \sqrt{\rho_k^2 \varsigma_k^2 + (1 - 2\rho_k)\varsigma_k}}{2}.$$

Since Cauchy–Schwarz inequality ensures that $\varsigma_k \geq 1$, we have

$$\lambda_k^+ = \frac{1 + \rho_k \varsigma_k + \sqrt{(\rho_k \varsigma_k - 1)^2 + \varsigma_k - 1}}{2} \geq 1.$$

Now, in order to show that $\lambda_k^- > 0$, we define the following function:

$$q(\varphi) = \frac{1 + \rho_k \varphi - \sqrt{\rho_k^2 \varphi^2 + (1 - 2\rho_k)\varphi}}{2},$$

for which $q(\varsigma_k) = \lambda_k^-$. It can be seen that if $\rho_k \neq 1/2$, then $q(\varphi)$ is a strictly decreasing function on $[1, +\infty)$, and consequently, since $\rho_k > 1/4$, we have

$$\lambda_k^- > \lim_{\varphi \to +\infty} q(\varphi) = 1 - \frac{1}{4\rho_k} > 0.$$

Furthermore, if $\rho_k = \dfrac{1}{2}$, then

$$\lambda_k^- = \frac{1}{2} = 1 - \frac{1}{4\rho_k} > 0.$$

Let $\bar{\lambda}_k$ be the smallest eigenvalue of $A_{k+1}$. From the above discussion, we have $\bar{\lambda}_k \geq 1 - \dfrac{1}{4\rho_k} > 0$. Thus, the matrix $A_{k+1}$ is positive definite and from (2.13) we have

$$d_{k+1}^T g_{k+1} = -g_{k+1}^T A_{k+1} g_{k+1} \leq -\bar{\lambda}_k \|g_{k+1}\|^2 \leq -\left(1 - \frac{1}{4\rho_k}\right)\|g_{k+1}\|^2,$$

which completes the proof.                                                              □

REMARK 2.2. From Theorem 2.1, if $\rho_k \geq \rho > 1/4$, then the sufficient descent condition (1.10) holds for the EDL method with $\varrho = 1 - \dfrac{1}{4\rho}$.

## 3. Global Convergence

Here, we show that in addition to using the objective function values and satisfying the sufficient descent condition, the EDL method is globally convergent without convexity assumption on the objective function. In this context, we need to consider the following standard assumptions.

Assumption A1: The level set $\mathcal{L} = \{x| \ f(x) \leq f(x_0)\}$, with $x_0$ to be the starting point of the iterative method (1.1), is bounded.

Assumption A2: In some open convex neighborhood $\mathcal{N}$ of $\mathcal{L}$, $f$ is continuously differentiable and its gradient is Lipschitz continuous; that is, there exists a positive constant $L$ such that

$$(3.1) \qquad \qquad ||\nabla f(x) - \nabla f(y)|| \leq L||x - y||, \ \forall x, y \in \mathcal{N}.$$

Note that these assumptions imply that there exists a positive constant $\gamma$ such that

$$(3.2) \qquad \qquad ||\nabla f(x)|| \leq \gamma, \ \forall x \in \mathcal{L}.$$

The following important lemma plays an essential role in our global convergence analysis.

Lemma 3.1 ([27]). *Suppose that Assumptions A1 and A2 hold. Consider any iterative method in the form of (1.1), where $d_k$ and $\alpha_k$ satisfy the sufficient descent condition (1.10) and the Wolfe conditions (2.2) and (2.3), respectively. If*

$$(3.3) \qquad \qquad \sum_{k \geq 0} \frac{1}{||d_k||^2} = \infty,$$

*then the method converges in the sense that*

$$(3.4) \qquad \qquad \liminf_{k \to \infty} ||g_k|| = 0.$$

Now, we can establish the following global convergence theorem for our method.

Theorem 3.2. *Suppose that Assumptions A1 and A2 hold. Consider a CG method in the form of (1.1)–(1.2) with the parameter $\beta_k^{EDL}$ given by (2.11) in which $C > 0$, there exists a positive constant $M$ such that $t \leq M$, and the steplength $\alpha_k$ is determined to fulfill the Wolfe conditions (2.2) and (2.3). If the sufficient descent condition (1.10) is satisfied, then the method converges in the sense that (3.4) holds.*

Proof. At first, note that sufficient descent condition (1.10) and the Wolfe condition (2.2) ensure that $\{x_k\}_{k \geq 0} \subseteq \mathcal{L}$. Now, to prove the theorem by contradiction, we suppose that there exists a positive constant $\varepsilon$ such that

$$(3.5) \qquad \qquad ||g_k|| \geq \varepsilon, \ \forall k \geq 0.$$

Hence, since from the Wolfe condition (2.3) we have $s_k^T y_k > 0$, considering (3.5) we get

$$(3.6) \qquad s_k^T z_k = s_k^T y_k + \xi \max\{\theta_k, 0\} + C\|g_k\|^r\|s_k\|^2 > C\varepsilon^r\|s_k\|^2.$$

On the other hand, from the mean–value theorem we can write

$$|\theta_k| = |2(f_k - f_{k+1}) + s_k^T(g_k + g_{k+1})|$$
$$= |(-2\nabla f(x_\mu) + \nabla f(x_k) + \nabla f(x_{k+1}))^T s_k|,$$

where $x_\mu = \mu x_k + (1 - \mu)x_{k+1}$, for some $\mu \in (0, 1)$. Therefore, from (3.1) we have

$$(3.7) \qquad |\theta_k| \leq (\|\nabla f(x_k) - \nabla f(x_\mu)\| + \|\nabla f(x_{k+1}) - \nabla f(x_\mu)\|) \|s_k\|$$
$$\leq (L(1 - \mu)\|s_k\| + L\mu\|s_k\|) \|s_k\| = L\|s_k\|^2.$$

Thus, from (3.1), (3.2) and (3.7), and since $u_k = s_k$, we can write

$$(3.8) \qquad \|z_k\| \leq \|y_k\| + \xi \frac{|\theta_k|}{\|s_k\|^2}\|s_k\| + C\|g_k\|^r\|s_k\|$$
$$\leq L\|s_k\| + L\xi\|s_k\| + C\gamma^r\|s_k\| = (L + L\xi + C\gamma^r)\|s_k\|.$$

Now, from (1.2) and (2.11) we have

$$d_{k+1} = -g_{k+1} + \frac{g_{k+1}^T z_k}{s_k^T z_k}s_k - t\frac{g_{k+1}^T s_k}{s_k^T z_k}s_k,$$

which together with Cauchy–Schwarz inequality, (3.2), (3.6) and (3.8) yields

$$\|d_{k+1}\| \leq \| - g_{k+1}\| + \frac{\|g_{k+1}\|\|z_k\|}{s_k^T z_k}\|s_k\| + t\frac{\|g_{k+1}\|\|s_k\|}{s_k^T z_k}\|s_k\|$$
$$\leq \gamma + \frac{(L + L\xi + C\gamma^r)\gamma}{C\varepsilon^r} + M\frac{\gamma}{C\varepsilon^r},$$

leading to (3.3). So, from Lemma 3.1, (3.4) holds, contradicting (3.5). $\qquad \square$

REMARK 3.3. From Theorem 2.1, the EDL method with the parameter $t$ given by (2.12) possesses the sufficient descent property for general functions. Also, if the objective function is uniformly convex, then $s_k^T y_k \geq \varpi\|s_k\|^2$. So, form (3.8) we have

$$t \leq \rho_k \frac{(L + L\xi + C\gamma^r)^2\|s_k\|^2}{\varpi\|s_k\|^2} = \rho_k \frac{(L + L\xi + C\gamma^r)^2}{\varpi},$$

and in this situation, if $\rho_k$ is bounded above, then there exists a positive constant $M$ such that $t \leq M$. Hence, assumptions of Theorem 3.2 seem to be reasonable.

## 4. Numerical Experiments

Here, we present some numerical results obtained by applying C++ implementations of four versions of a CG method with the parameter (2.11) in which $t$ is computed by (2.12) with $\rho_k = 1$, guaranteeing the sufficient descent property (1.10) and being an extended form of (1.9) which is an effective choice as stated in [11]. The methods are abbreviated by

- EDL: in which $\xi = 1.1$ and $C = 10^{-4}$ because of the promising numerical results obtained respectively among the different values $\xi \in \{0.1k\}_{k=1}^{30}$ and $C \in \{10^{-k}\}_{k=0}^{6}$;
- DK: in which $\xi = C = 0$, being equivalent to the CG method proposed by Dai and Kou ([11]);
- YT: in which $\xi = 3$ and $C = 0$, being equivalent to the CG method proposed by Yabe and Takano ([32]);
- ZZ: in which $\xi = 0$ and $C = 10^{-4}$, being equivalent to the CG method proposed by Zhou and Zhang ([37]).

For the parameter $r$ of the EDL and ZZ methods, we adopted the suggestion of [37] and set $r = 1$, when $\|g_k\| \geq 1$, and $r = 3$, otherwise. For all the four methods, we used the effective approximate Wolfe conditions proposed by Hager and Zhang ([19]) in the line search procedure, with the same parameter values as suggested in [19]. Moreover, all attempts to solve the test problems were terminated when $\|g_k\|_\infty < 10^{-5} \max(1, \|g_0\|_\infty)$.

The codes were run on a PC with 3.2 GHz Intel I3 of CPU, 4 GB of RAM and Centos 6.2 server Linux operation system. Since CG methods have been mainly designed to solve large–scale problems, the experiments were performed on a set of 64 unconstrained optimization test problems of the CUTEr collection ([16]) with default dimensions being at least equal to 1000, as given in Hager's home page: `http://www.math.ufl.edu/~hager/`. The test problems data has been also clarified in [8].

Efficiency comparisons were made using the Dolan–Moré performance profile ([13]) on the running time and the total number of function and gradient evaluations being equal to $N_f + 3N_g$, where $N_f$ and $N_g$ respectively denote the number of function and gradient evaluations. Performance profile gives, for every $\omega \geq 1$, the proportion $p(\omega)$ of the test problems that each considered algorithmic variant has a performance within a factor of $\omega$ of the best. Figures 1 and 2 show the results of comparisons.

As seen in the figures, EDL is preferable to the other methods both in the perspectives of the total number of function and gradient evaluations and the running time. Hence, using the extended secant equation (2.10) in the Dai–Liao approach turns out to be practically effective. Moreover, exactness of the modified secant equation (2.5) leads to reasonable numerical behavior of the YT method. Since EDL, YT and ZZ are preferable to DK (especially with respect to the running time), it can be stated that the choice (2.12) for
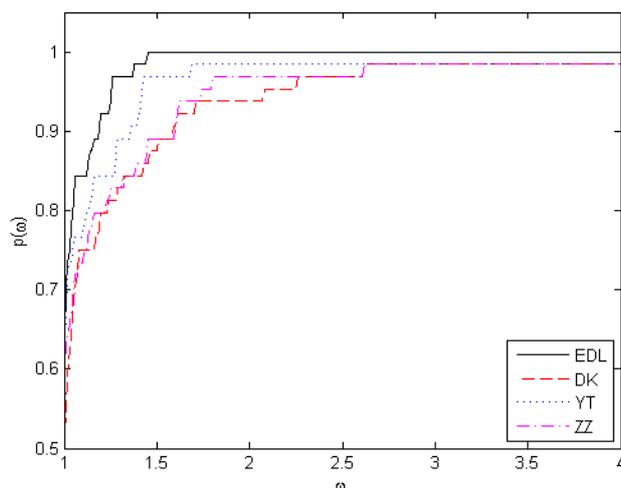
FIGURE 1. Total number of function and gradient evaluations performance profiles

the parameter $t$ with nonzero values of $\xi$ or $C$ (which is given based on the the extended secant equation (2.10)) seem to be more effective than the choice (1.9).

## 5. Conclusions

Following Dai–Liao approach, a nonlinear conjugate gradient method has been proposed based on the extended secant equation (2.10). In addition to the gradient information, the method employs the objective function values. With the choice (2.12) for the parameter $t$ in (2.11), a sufficient descent property has been established for the method. It has been shown that the method is globally convergent for general functions. Numerical comparisons have been made between the implementations of the proposed method and three other nonlinear conjugate gradient methods proposed in [11, 32, 37], being especial cases of our method, using the Dolan–Moré performance profile. In the numerical experiments, a set of 64 unconstrained optimization test problems of the CUTEr collection has been chosen with default dimensions being at least equal to 1000. Results of comparisons showed efficiency of the proposed method, implying effectiveness of the extended secant equation (2.10).
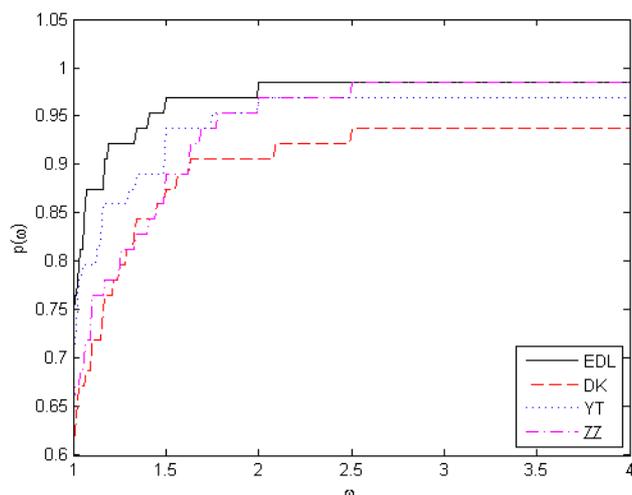
FIGURE 2. CPU time performance profiles

William W. Hager for providing the line search code. They also thank the anonymous reviewer for his/her valuable comments and suggestions helped to improve the quality of this work.

## REFERENCES

[1] N. Andrei, *Numerical comparison of conjugate gradient algorithms for unconstrained optimization*, Stud. Inform. Control **16** (2007), 333–352.

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

[3] S. Babaie–Kafaki, *A modified BFGS algorithm based on a hybrid secant equation*, Sci. China Math. **54** (2011), 2019–2036.

[4] S. Babaie–Kafaki, *On the sufficient descent condition of the Hager–Zhang conjugate gradient methods*, 4OR **12** (2014), 285–292.

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

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

[7] S. Babaie–Kafaki and R. Ghanbari, *An extended three–term conjugate gradient method with sufficient descent property*, Miskolc Math. Notes **16** (2015), 45–55.

[8] 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. **30** (2015), 673–681.

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

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

[11] 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. **23** (2013), 296–320.

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

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

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

[15] J.A. Ford, Y. Narushima and H. Yabe, *Multi–step nonlinear conjugate gradient methods for unconstrained minimization*, Comput. Optim. Appl. **40** (2008), 191–216.

[16] N.I.M. Gould, D. Orban and Ph.L. Toint, *CUTEr: a constrained and unconstrained testing environment, revisited*, ACM Trans. Math. Softw. **29** (2003), 373–394.

[17] Q. Guo, J.G. Liu and D.H. Wang, *A modified BFGS method and its superlinear convergence in nonconvex minimization with general line search rule*, J. Appl. Math. Comput. **28** (2008), 435–446.

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

[19] W.W. Hager and H. Zhang, *Algorithm 851: CG–Descent, a conjugate gradient method with guaranteed descent*, ACM Trans. Math. Software **32** (2006), 113–137.

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

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

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

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

[24] D.C. Liu and J. Nocedal, *On the limited memory BFGS method for large–scale optimization*, Math. Programming **45** (1989), 503–528.

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

[26] M.J.D. Powell, *Nonconvex minimization calculations and the conjugate gradient method*, in *D.F. Griffiths (Ed.),* Numerical analysis (Dundee, 1983), Lecture Notes in Math. 1066, Springer, Berlin, 1984, 122–141.

[27] 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.

[28] W. Sun and Y.X. Yuan, Optimization theory and methods. Nonlinear programming, Springer, New York, 2006.

[29] Z. Wei, G. Li and L. Qi, *New quasi–Newton methods for unconstrained optimization problems*, Appl. Math. Comput. **175** (2006), 1156–1188.

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

[31] C. Xu and J.Z. Zhang, *A survey of quasi–Newton equations and quasi–Newton methods for optimization*, Ann. Oper. Res. **103** (2001), 213–234.

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

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

[34] Y.X. Yuan and R.H. Byrd, *Non–quasi–Newton updates for unconstrained optimization*, J. Comput. Math. **13** (1995), 95–107.

[35] J. Zhang and C. Xu, *Properties and numerical performance of quasi–Newton methods with modified quasi–Newton equations*, J. Comput. Appl. Math. **137** (2001), 269–278.

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

M. R. Arazm
Department of Mathematics
Faculty of Mathematics, Statistics and Computer Science
Semnan University
P.O. Box: 35195–363, Semnan
Iran
*E-mail*: mohamadreza.arazm@semnan.ac.ir

S. Babaie–Kafaki
Department of Mathematics
Faculty of Mathematics, Statistics and Computer Science
Semnan University
P.O. Box: 35195–363, Semnan
Iran
*E-mail*: sbk@semnan.ac.ir

R. Ghanbari
Faculty of Mathematical Sciences
Ferdowsi University of Mashhad
P.O. Box: 9177948953, Mashhad
Iran
*E-mail*: rghanbari@um.ac.ir