A CLASS OF DESCENT FOUR–TERM EXTENSION OF THE DAI–LIAO CONJUGATE GRADIENT METHOD BASED ON THE SCALED MEMORYLESS BFGS UPDATE

SAMAN BABAIE–KAFAKI*
Department of Mathematics, Faculty of Mathematics, Statistics and Computer Science
Semnan University, P.O. Box: 35195–363, Semnan, Iran

REZA GHANBARI
Faculty of Mathematical Sciences, Ferdowsi University of Mashhad
P.O. Box: 9177948953, Mashhad, Iran

(Communicated by Regina Burachik)

Abstract. Hybridizing the three–term conjugate gradient method proposed by Zhang et al. and the nonlinear conjugate gradient method proposed by Dai and Liao based on the scaled memoryless BFGS update, a one–parameter class of four–term conjugate gradient methods is proposed. It is shown that the suggested class of conjugate gradient methods possesses the sufficient descent property, without convexity assumption on the objective function. A brief global convergence analysis is made for uniformly convex objective functions. Results of numerical comparisons are reported. They demonstrate efficiency of a method of the proposed class in the sense of the Dolan–Moré performance profile.

1. Introduction. Consider the following unconstrained minimization problem:

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

in which \( f : \mathbb{R}^n \to \mathbb{R} \) is a smooth nonlinear function and the analytic expression of its gradient is available. Among the most useful tools for solving large–scale cases of the problem \([1]\) are the limited memory BFGS (Broyden–Fletcher–Goldfarb–Shanno) methods \([23]\) and the conjugate gradient (CG) methods \([19]\), because the amount of memory storage required by the methods is low. This study is devoted to a class of four–term CG methods, constructed based on the scaled memoryless BFGS update \([29]\).

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

\[ x_{k+1} = x_k + s_k, \quad s_k = \alpha_k d_k, \quad k = 0, 1, \ldots, \]

2010 Mathematics Subject Classification. Primary: 65K05, 90C53; Secondary: 49M37.

Key words and phrases. Unconstrained optimization, large–scale optimization, conjugate gradient method, sufficient descent property, BFGS update, global convergence.

This research was supported by Research Councils of Semnan University and Ferdowsi University of Mashhad.

* Corresponding author: Saman Babaie–Kafaki.
where \( \alpha_k \) is a steplength to be computed by a line search procedure along the search direction \( d_k \) defined by

\[
d_0 = -g_0, \quad d_{k+1} = -g_{k+1} + \beta_k d_k, \quad k = 0, 1, \ldots, \tag{3}
\]
in which \( g_k = \nabla f(x_k) \) and \( \beta_k \) is a scalar called the CG (update) parameter. The steplength \( \alpha_k \) is usually determined to fulfill the well–known Wolfe conditions [29], i.e.,

\[
\begin{align*}
    f(x_k + \alpha_k d_k) - f(x_k) & \leq \delta \alpha_k \nabla f(x_k)^T d_k, \\
    \nabla f(x_k + \alpha_k d_k)^T d_k & \geq \sigma \nabla f(x_k)^T d_k,
\end{align*}
\]

where \( 0 < \delta < \sigma < 1 \).

Different CG methods mainly correspond to different choices for the CG parameter [19]. One of the essential CG parameters has been suggested by Hestenes and Stiefel [20] (HS), that is,

\[
\beta_{HS}^k = \frac{g_{k+1}^T y_k}{d_k^T y_k},
\]

where \( y_k = g_{k+1} - g_k \). As an important property, search directions of the HS method satisfy the conjugacy condition

\[
d_k^T y_k = 0, \quad \forall k \geq 0,
\]

independent of the line search and the objective function convexity.

In an attempt to suggest an extension of \( \beta_{HS}^k \) employing the quasi–Newton aspects [29], Dai and Liao [10] (DL) dealt with the extended conjugacy condition

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

with the nonnegative parameter \( t \), which considering (3) leads to the following CG parameter:

\[
\beta_{DL}^k = \beta_{HS}^k - t \frac{g_{k+1}^T s_k}{d_k^T y_k}.
\]

It is worth noting that if

\[
t = 2 \frac{||y_k||^2}{s_k^T y_k}, \tag{6}
\]

where \( ||.|| \) stands for the Euclidean norm, then the CG parameter proposed by Hager and Zhang [17] is achieved. Also, the choice

\[
t = \tau_k + \frac{||y_k||^2}{s_k^T y_k} - \frac{s_k^T y_k}{||y_k||^2}, \tag{7}
\]

in which \( \tau_k \) is a parameter corresponding to the scaling factor in the scaled memoryless BFGS method [29], yields another CG parameter suggested by Dai and Kou [9]. The choices (6) and (7) are theoretically effective since they can guarantee the sufficient descent condition, i.e.,

\[
g_k^T d_k \leq -c ||g_k||^2, \quad k = 0, 1, \ldots, \tag{8}
\]

where \( c \) is a positive constant, independent of the line search and the objective function convexity [2].

The Dai–Liao approach has been attracted special attention. In several efforts, modified secant equations have been applied to make modifications on the DL method. For example, Yabe and Takano [31] used the modified secant equation proposed by Zhang et al. [33]. Also, Zhou and Zhang [35] applied the modified secant equation proposed by Li and Fukushima [21]. Li et al. [22] used the modified secant equation proposed by Wei et al. [30]. Ford et al. [14] employed the multi–step quasi–Newton equations proposed by Ford et al. [13]. Babaie–Kafaki et
al. [6] applied a revised form of the modified secant equation proposed by Zhang et al. [33], and the modified secant equation proposed by Yuan [32]. In another direction, descent property of the DL method has been focused on. Recall that we say the search directions \( \{d_k\}_{k \geq 0} \) possess the descent property (or equivalently, satisfy the descent condition) if and only if
\[
g_k^T d_k < 0, \quad k = 0, 1, ...,\tag{9}
\]
being weaker than the sufficient descent condition [8]. Recently, based on an eigenvalue analysis Babaie–Kafaki and Ghanbari [3] proposed the following family of two–parameter choices for \( t \):
\[
t_{p,q}^k = p \frac{||y_k||^2}{s_k^T y_k} - q \frac{s_k^T y_k}{||s_k||^2},
\]
with \( p > \frac{1}{4} \) and \( q \leq \frac{1}{4} \), guaranteeing the descent property [9]. It is interesting that if we let \((p,q) = (2, 0)\), then \( t_{p,q}^k \) reduces to the formula (6) proposed by Hager and Zhang [17], and if we let \((p,q) = (1, 0)\), then \( t_{p,q}^k \) reduces to the formula (7) proposed by Dai and Kou [9] with the optimal choice \( \tau_k = \frac{s_k^T y_k}{||s_k||^2} \). More recently, Babaie–Kafaki and Ghanbari [4] dealt with a hybridization of the DL method and a three–term CG method proposed by Zhang et al. [34] (ZZL) with the following search directions:
\[
d_0 = -g_0, \quad d_{k+1} = -g_{k+1} + \beta^H_{k} d_k - t \frac{g_{k+1}^T s_k}{d_k^T y_k} d_k - \frac{g_{k+1}^T d_k}{d_k^T y_k} g_k, \quad \forall k \geq 0,\tag{10}
\]
where \( t \) is a nonnegative parameter. As seen, search directions (10) can be viewed as a four–term extension of the DL method. If \( t = 0 \), then the method reduces to the three–term CG method ZZL which satisfies the sufficient descent condition \( g_k^T d_k = -||g_k||^2 \), for all \( k \geq 0 \). Inheriting from the ZZL method, it can be seen that search directions (10) satisfy the sufficient descent condition [8] with \( c = 1 \), independent of the line search and the objective function convexity. Based on the standard secant equation [29], an effective choice for the parameter \( t \) in (10) has been suggested as follows:
\[
t_k = \max\{\xi, 1 - \frac{||y_k||^2}{s_k^T y_k}\},\tag{11}
\]
where \( \xi \) is a nonnegative constant.

Here, we deal with other proper choices for the parameter \( t \) in (10). This work is organised as follows. In Section 2 we briefly study the scaled memoryless BFGS update. In Section 3 we propose a family of one–parameter choices for \( t \) in (10), using the scaled memoryless BFGS update [29]. Then, we conduct a brief global convergence analysis. We present comparative numerical comparisons in Section 4. Finally, conclusions are drawn in Section 5.

2. On the scaled memoryless BFGS update. As known, quasi–Newton methods are of particular performance for solving (1) since they do not require explicit expressions of the second derivatives and are often globally and locally superlinearly convergent [29]. Iterative formula of the quasi–Newton methods is in the form of (2) in which the search direction \( d_k \) is computed by
\[
d_k = -H_kg_k,\tag{12}
\]
where \( H_k \in \mathbb{R}^{n \times n} \) is an approximation of the inverse Hessian; more precisely, \( H_k \approx \nabla^2 f(x_k)^{-1} \). The methods are characterized by the fact that \( H_k \) is effectively updated to achieve a new matrix \( H_{k+1} \) as an approximation of \( \nabla^2 f(x_{k+1})^{-1} \), satisfying a version of secant (quasi–Newton) equation which implicitly includes the second order information \([29]\). The most popular equation is the standard secant equation; that is,

\[
H_{k+1} y_k = s_k.
\] (13)

Having favorable numerical performance and strong theoretical properties in contrast to the other quasi–Newton updating formulas, the BFGS update that is separately proposed by Broyden \([8]\), Fletcher \([12]\), Goldfarb \([15]\) and Shanno \([27]\) is given by

\[
H_{k+1} = H_k - \frac{s_k y_k^T H_k + H_k y_k s_k^T}{s_k^T y_k} + \left( 1 + \frac{y_k^T H_k y_k}{s_k^T y_k} \right) \frac{s_k s_k^T}{s_k^T y_k}.
\] (14)

It can be seen that if \( H_k \) is a positive definite matrix and the line search ensures that \( s_k^T y_k > 0 \), then \( H_{k+1} \) is also a positive definite matrix \([29]\) and consequently, the descent condition (9) holds.

In order to achieve an ideal distribution of the eigenvalues of quasi–Newton updates, improving the condition number of successive approximations of the inverse Hessian and consequently, increasing the numerical stability in the iterative method (2)–(12), the scaled quasi–Newton updates have been developed \([29]\). In this context, replacing \( H_k \) by \( \theta_k H_k \) in (14) in which \( \theta_k > 0 \) is called the scaling parameter, the scaled BFGS update can be achieved as follows:

\[
H_{k+1}^{\theta_k} = \theta_k H_k - \theta_k \frac{s_k y_k^T H_k + H_k y_k s_k^T}{s_k^T y_k} + \left( 1 + \theta_k \frac{y_k^T H_k y_k}{s_k^T y_k} \right) \frac{s_k s_k^T}{s_k^T y_k}.
\] (15)

The most effective choices for \( \theta_k \) in (15) have been proposed by Oren and Luenberger \([25]\),

\[
\theta_k = \frac{s_k^T H_k^{-1} s_k}{s_k^T y_k},
\] (16)

and, Oren and Spedicato \([26]\),

\[
\theta_k = \frac{s_k^T y_k}{y_k^T H_k y_k},
\] (17)

(see also \([24]\)). Note that the scaled BFGS update with one of the parameters \([16]\) or \([17]\) is called a self–scaling BFGS update.

Although self–scaling BFGS methods are numerically efficient \([26]\), as an important defect the methods need to save the matrix \( H_k \in \mathbb{R}^{n \times n} \) in each iteration, being improper for solving large–scale problems. Hence, replacing \( H_k \) by the identity matrix in \([15]\), self–scaling memoryless BFGS update has been proposed as follows:

\[
\bar{H}^{\theta_k}_{k+1} = \theta_k I - \theta_k \frac{s_k y_k^T + y_k s_k^T}{s_k^T y_k} + \left( 1 + \theta_k \frac{y_k^T y_k}{s_k^T y_k} \right) \frac{s_k s_k^T}{s_k^T y_k},
\] (18)

(see also \([1]\)). Similarly, memoryless version of the scaling parameters \([16]\) and \([17]\) can be respectively written as:

\[
\theta_k = \frac{||s_k||^2}{s_k^T y_k},
\] (19)
and
\[ \theta_k = \frac{s_k^T y_k}{\| y_k \|^2}. \] (20)

Note that the scaling parameter (19) can also be determined based on a two–point approximation of the standard secant equation (13) [7].

3. A class of modified Dai–Liao conjugate gradient methods. Here, we deal with a new approach for computing the parameter \( t \) in the four–term extension of the DL method given by (10). In this context, we need to assume that \( s_k^T y_k > 0 \), as guaranteed by the Wolfe condition (5).

It is worth noting that the search directions (10) can be written as:
\[ d_0 = -g_0, \quad d_{k+1} = -Q_{k+1} g_{k+1}, \quad \forall k \geq 0, \] (21)
in which the matrix \( Q_{k+1} \), called the search direction matrix, is given by
\[ Q_{k+1} = I - \frac{s_k y_k}{s_k^T y_k} + \frac{y_k s_k^T}{s_k^T y_k}. \] (22)

Hence, structure of a CG method with the search directions (10) is similar to a quasi–Newton method. This motivated us to compute the parameter \( t \) in (22) in a way to make \( Q_{k+1} \) as closer as possible to an effective quasi–Newton update which is appropriate for large–scale problems. Thus, we employ the scaled memoryless BFGS update (18), due to its numerical and theoretical advantages.

Define the matrix \( E \in \mathbb{R}^{n \times n} \) as follows:
\[ E = Q_{k+1} - \bar{H}^{\theta_k}_{k+1} = (1 - \theta_k)I + (\theta_k + 1) \frac{y_k s_k^T}{s_k^T y_k} + (\theta_k - 1) \frac{s_k y_k^T}{s_k^T y_k}, \]
where the matrices \( Q_{k+1} \) and \( \bar{H}^{\theta_k}_{k+1} \) are respectively defined by (22) and (18). In order to decrease the distance between \( Q_{k+1} \) and \( \bar{H}^{\theta_k}_{k+1} \), here we determine the parameter \( t \) as a minimizer of \( \| E \|_{F} \) in which \( \| . \|_{F} \) stands for the Frobenius matrix norm. Since \( \| E \|_{F}^2 = \text{tr}(E^T E) \), we can deal with the following minimization problem:
\[ \min_{t \geq 0} \psi(t) = \left( t - 1 - \theta_k \frac{\| y_k \|^2}{s_k^T y_k} \right) \| s_k \|^4 \left( \frac{s_k^T y_k}{\| y_k \|^2} \right)^2 + 2(\theta_k + 1) \left( t - 1 - \theta_k \frac{\| y_k \|^2}{s_k^T y_k} \right) \frac{\| s_k \|^2}{s_k^T y_k} + \varpi, \] (23)
where \( \varpi \) is a real constant, being independent of \( t \). After some algebraic manipulations it can be seen that
\[ t^*_k = 1 + \theta_k \frac{\| y_k \|^2}{s_k^T y_k} - (\theta_k + 1) \frac{s_k y_k}{\| s_k \|^2}, \]
is the unique solution of the minimization problem (23). However, considering (21), since
\[ g_{k+1}^T d_{k+1} = -\| g_{k+1} \|^2 - \alpha_k \frac{(g_{k+1}^T d_k)^2}{d_k^T y_k}, \] (24)
in order to ensure the sufficient descent condition we need to have \( t \geq 0 \). So, here we suggest the following truncation of \( t^*_k \):

\[
\tilde{t}^*_k = \max \left\{ \xi, 1 + \theta_k \frac{||y_k||^2}{s_k^T y_k} - (\theta_k + 1) \frac{s_k^T y_k}{||s_k||^2} \right\},
\]

where \( \xi \) is a nonnegative constant.

The formula (25) indicates a class of one–parameter choices for the parameter \( t \) in (10) or equivalently, in (22). Different choices for the scaling parameter \( \theta_k \) lead to different choices for \( \tilde{t}^*_k \). The choice \( \theta_k = 1 \) yields

\[
\tilde{t}^*_{k_1} = \max \left\{ \xi, 1 + \frac{||y_k||^2}{s_k^T y_k} - 2 \frac{s_k^T y_k}{||s_k||^2} \right\},
\]

while the choices (19) and (20) for \( \theta_k \) respectively yield

\[
\tilde{t}^*_{k_2} = \max \left\{ \xi, \frac{||s_k||^2 ||y_k||^2}{(s_k^T y_k)^2} - \frac{s_k^T y_k}{||s_k||^2} \right\},
\]

and

\[
\tilde{t}^*_{k_3} = \max \left\{ \xi, 2 - \frac{(s_k^T y_k)^2}{||s_k||^2 ||y_k||^2} - \frac{s_k^T y_k}{||s_k||^2} \right\}.
\]

In what follows, based on the convergence analysis of [28], we discuss global convergence of a CG method with the search directions (21) in which the parameter \( t \) is computed by (26), (27) or (28). In this context, the following standard assumptions are needed.

**Assumption A1.** The level set \( L = \{ x | f(x) \leq f(x_0) \} \), with \( x_0 \) to be the starting point of the iterative method (2), is bounded.

**Assumption A2.** In some open convex neighborhood \( N \) of \( L \), \( f \) is continuously differentiable and its gradient is Lipschitz continuous; that is, there exists a positive constant \( L \) such that

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

The following lemma is now immediate, ensuring the boundedness of \( \tilde{t}^*_k, \ i = 1, 2, 3 \).

**Lemma 3.1.** Suppose that Assumptions A1 and A2 hold. Consider the iterative method (2) with the search directions \( \{d_k\}_{k \geq 0} \) given by (10) in which the parameter \( t \) is computed by (26), (27) or (28). If the objective function \( f \) is uniformly convex on \( N \) and the steplength \( \alpha_k \) is determined to fulfill the Wolfe conditions (4) and (5), then there exists a positive constant \( M \) such that

\[
\tilde{t}^*_k \leq M, \ i = 1, 2, 3, \forall k \geq 0.
\]

**Proof.** At first, note that since \( \tilde{t}^*_k \geq 0, \ i = 1, 2, 3 \), from (24) we have

\[
g_k^T d_k \leq -||y_k||^2, \ \forall k \geq 0,
\]

which together with the line search condition (4) yields \( \{x_k\}_{k \geq 0} \subseteq L \). Also, uniform convexity of the differentiable function \( f \) ensures that there exists a positive constant \( \mu \) such that

\[
s_k^T y_k \geq \mu ||s_k||^2, \ k = 0, 1, ...,
\]
(see Theorem 1.3.16 of [29]). Therefore, from (29) and (30) we get
\[
\bar{t}_{k1} \leq \max \left\{ \xi, 1 + \frac{||s_k||^2}{s_k^T y_k} + 2 \frac{s_k^T y_k}{||s_k||^2} \right\} \\
\leq \max \left\{ \xi, 1 + \frac{L^2}{\mu} + 2L \right\} = M_1, \quad (31)
\]
and similarly,
\[
\bar{t}_{k2} \leq \max \left\{ \xi, \frac{||s_k||^2}{s_k^T y_k} + \frac{s_k^T y_k}{||s_k||^2} \right\} \\
\leq \max \left\{ \xi, \frac{L^2}{\mu^2} + L \right\} = M_2. \quad (32)
\]
Moreover, from Cauchy–Schwarz inequality and (29) we have
\[
\bar{t}_{k3} \leq \max \left\{ \xi, 2 + \frac{s_k^T y_k}{||s_k||^2} \right\} \\
\leq \max \left\{ \xi, 2 + L \right\} = M_3. \quad (33)
\]
So, considering (31), (32) and (33), to complete the proof it is enough to let
\[
M = \max \{M_1, M_2, M_3\}.
\]

Lemma 3.1 plays an essential role in proving the following global convergence theorem. The proof is similar to the proof of Theorem 2.1 of [4] and here is omitted.

**Theorem 3.2.** Suppose that Assumptions A1 and A2 hold. Consider the iterative method (2) with the search directions \(\{d_k\}_{k \geq 0}\) given by (10) in which the parameter \(t\) is computed by (26), (27) or (28). If the objective function \(f\) is uniformly convex on \(N\) and the steplength \(\alpha_k\) is determined to fulfill the Wolfe conditions (4) and (5), then the method converges in the sense that \(\lim \inf_{k \to \infty} ||g_k|| = 0\).

4. Numerical experiments. Here, we present some numerical results obtained by applying C++ implementations of the following four CG methods in the form of (2) with the search directions (10) in which
- The NMDL1 method: \(t = \bar{t}_{k1}^*\) given by (26);
- The NMDL2 method: \(t = \bar{t}_{k2}^*\) given by (27);
- The NMDL3 method: \(t = \bar{t}_{k3}^*\) given by (28);
- The MDL method: \(t = t_k\) given by (11).

To justify comparison of the new methods with MDL, note that in [4] it has been shown that MDL is numerically preferable to the CG–Descent and ZZL methods.

The codes were run on a computer with 3.2 GHz of CPU, 1 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/](http://www.math.ufl.edu/~hager/) (The test problems data has been also clarified in [5]).

For the MDL method, we set \(\xi = 0.66\) in (11) as suggested in [4]. Also, for the methods of NMDL1, NMDL2 and NMDL3 respectively we set \(\xi = 0.65\) in (26), \(\xi = 0.55\) in (27) and \(\xi = 0.40\) in (28), because of the promising numerical results obtained among the different values \(\xi \in \{0.05k\}_{k=0}^{20}\). In all the four methods, we
used the effective approximate Wolfe conditions proposed by Hager and Zhang in the line search procedure, with the same parameter values as suggested in [18]. Also, all attempts to solve the test problems were terminated when $||g_k||_\infty < 10^{-6}(1 + |f(x_k)|)$.

Efficiency comparisons were made using the performance profile introduced by Dolan and More [11], on the running time and the total number of function and gradient evaluations being equal to $N_f + 3N_g$ [18], 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 in the perspectives of the total number of function and gradient evaluations and the CPU time. As the figures show, NMDL3 outperforms the other methods. Moreover, MDL is preferable to NMDL2 while NMDL2 outperforms NMDL1. Thus, proper choices for the parameter $\theta_k$ in (25) may lead to efficient CG methods in the form of (2)–(10). Also, effectiveness of the scaling parameters (27) and (28) has been demonstrated by the promising numerical behaviors of NMDL2 and NMDL3.

5. Conclusions. Taking advantage of the scaled memoryless BFGS update, a class of one–parameter choices for the parameter of a recently proposed four–term extension of the Dai–Liao method has been suggested, leading to a class of nonlinear conjugate gradient methods. Under mild assumptions, it has been shown that the methods of the suggested class possess the sufficient descent property. A global convergence property has been established for three methods of the suggested class with the assumption of uniform convexity on the objective function. Preliminary numerical results showed that practical efficiency of the methods of the suggested
class may be promising, specially when the scaling parameter is chosen appropriately.

Acknowledgments. This research was supported by Research Councils of Semnan University and Ferdowsi University of Mashhad. Also, the first author was supported in part by the grant 95813776 from Iran National Science Foundation (INSF). The authors are grateful to Professor William W. Hager for providing the line search code. They also thank the anonymous referees for their valuable comments helped to improve the presentation.

REFERENCES

[1] N. Andrei, Accelerated scaled memoryless BFGS preconditioned conjugate gradient algorithm for unconstrained optimization, European J. Oper. Res., 204 (2010), 410–420.

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

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

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

[5] 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.

[6] 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.

[7] J. Barzilai and J. Borwein, Two–point stepsize gradient methods, IMA J. Numer. Anal., 8 (1988), 141–148.

[8] C. Broyden, The convergence of a class of double–rank minimization algorithms. II. The new algorithm, J. Inst. Math. Appl., 6 (1970), 222–231.

[9] Y. Dai and C. Kou, A nonlinear conjugate gradient algorithm with an optimal property and an improved Wolfe line search, SIAM J. Optim., 23 (2013), 296–320.
[10] Y. Dai and L. Liao, New conjugacy conditions and related nonlinear conjugate gradient methods, *Appl. Math. Optim.*, 43 (2001), 87–101.

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

[12] R. Fletcher, A new approach to variable metric algorithms, *Comput. J.*, 13 (1970), 317–322.

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

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

[15] D. Goldfarb, A family of variable metric methods derived by variational means, *Math. Comp.*, 24 (1970), 23–26.

[16] N. Gould, D. Orban and P. Toint, CUTEr: a constrained and unconstrained testing environment, revisited, *ACM Trans. Math. Softw.*, 29 (2003), 353–372.

[17] 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.

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

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

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

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

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

[23] Y. Yuan, Optimization Theory and Methods: Nonlinear Programming, Springer, New York, 2006.

Received November 2014; 1st revision October 2015; 2nd revision December 2015.

E-mail address: sbk@semnan.ac.ir
E-mail address: rghanbari@um.ac.ir