Accepted Manuscript

A new efficient conjugate gradient method for unconstrained optimization

Masoud Fatemi

PII: S0377-0427(15)00647-0
DOI: http://dx.doi.org/10.1016/j.cam.2015.12.035
Reference: CAM 10433

To appear in: *Journal of Computational and Applied Mathematics*

Received date: 2 July 2015
Revised date: 14 November 2015

Please cite this article as: M. Fatemi, A new efficient conjugate gradient method for unconstrained optimization, *Journal of Computational and Applied Mathematics* (2016), http://dx.doi.org/10.1016/j.cam.2015.12.035

A new efficient conjugate gradient method for unconstrained optimization

Masoud Fatemi

Department of Mathematics, K. N. Toosi University of Technology, Tehran, Iran
Scientific Computations in Optimization and System Engineering (SCOPE), K. N. Toosi University of Technology, Tehran, Iran

Abstract

We propose a nonlinear conjugate gradient method for unconstrained optimization based on solving a new optimization problem. Our optimization problem combines the good features of the linear conjugate gradient method using some penalty parameters. We show that the new method is a subclass of Dai-Liao family, the fact that enables us to analyze the family, closely. As a consequence, we obtain an optimal bound for Dai-Liao parameter. The global convergence of the new method is investigated under mild assumptions. Numerical results show that the new method is efficient and robust, and outperforms CG-DESCENT.

Keywords: Conjugate gradient method, Dai-Liao family, Unconstrained optimization, Line search

1. Introduction

Here, we consider the following unconstrained optimization problem,

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

where f is a smooth function. Conjugate gradient algorithms are a class of efficient methods for solving (1), specially, when n is large [1, 2, 3, 4, 5, 6, 7, 8, 9]. This class were originally invented by Hestenes and Stiefel [1] for solving a symmetric and positive definite linear system of equations, and

Email address: smfatemi@kntu.ac.ir (Masoud Fatemi)

then was extended by many authors to handle general optimization problems [10, 11].

In a conjugate gradient algorithm, a sequence of iterates, x_{k+1}, are generated by the following scheme:

$$x_{k+1} = x_k + \alpha_k d_k, \tag{2}$$

where the search direction d_k is computed by

$$d_{k+1} = -g_{k+1} + \beta_k d_k, \quad d_0 = -g_0. \tag{3}$$

The step length $\alpha_k > 0$ usually satisfies the Wolfe conditions,

$$f(x_{k+1}) - f(x_k) \leq c_1 \alpha_k g_k^T d_k, \tag{4}$$

$$g_{k+1}^T d_k \geq c_2 g_k^T d_k, \tag{5}$$

where, $0 < c_1 < c_2 < 1$ are some arbitrary constants and $g_k := \nabla f(x_k)$.

The linear conjugate gradient method uses (2) and (3) with the exact line search to solve a strongly convex quadratic function. The method has some remarkable properties:

(i) The sufficient descent condition, namely, there exists a scalar $c > 0$ such that

$$g_{k+1}^T d_{k+1} \leq -c \parallel g_{k+1} \parallel^2 . \tag{6}$$

(ii) The conjugacy condition, namely,

$$d_{k+1}^T y_k = 0. \tag{7}$$

(iii) The orthogonality property:

$$g_{k+1}^T d_i = 0, \tag{8}$$

for $i = 0 \ldots k$.

It is known that a linear conjugate gradient algorithm always terminates in finite iteration, and it is actually a remarkable property. Unfortunately, this property can not be guaranteed for the nonlinear conjugate gradient algorithms.

The good features of the linear conjugate gradient method persuade the authors to follow this idea in nonlinear optimization. We refer the interested readers to the nice surveys of Hager and Zhang [10] and Dai [11] about nonlinear conjugate gradient methods. Hager and Zhang [7] have recently introduced an efficient nonlinear conjugate gradient method. Their method is a subclass of Dai-Liao family corresponding to the choice

$$\beta_k^{DL} = \frac{g_{k+1}^T y_k - \tau g_{k+1}^T s_k}{y_k^T d_k}, \tag{9}$$

with $\tau = \lambda_k \frac{\|y_k\|^2}{s_k^T y_k}$. They proved the global convergence of a truncated version of the method similar to PRP$^+$ of Gilbert and Nocedal [12] under mild assumptions. Numerical results showed that the method outperforms many existing conjugate gradient methods. Nowadays, it is known as a most efficient conjugate gradient method. An implementation of the method called CG-DESCENT is now available from the Hager's homepage.

In order to design an efficient nonlinear conjugate gradient method, we combine (i)-(iii), and introduce the following optimization problem:

$$\min_{\beta_k} \left[g_{k+1}^T d_{k+1} + M \left((g_{k+2}^T s_k)^2 + (d_{k+1}^T y_k)^2 \right) \right], \tag{10}$$

where M is a penalty parameter. The first term in (10) contains the information about (i), whereas, the second one contains the information about (ii) and (iii). A large value of M clearly increases the chance of satisfying (ii) and (iii), and a small one increases the effect of the sufficient descent property (i).

More recently in [13], we used the same idea, and proposed an optimal parameter for Dai-Liao family of conjugate gradient methods. In this work, we pay attention to M and introduce an efficient penalty parameter, having some useful properties. A new expression for β_k is then obtained by solving (10). We show that the resulting method is a subclass of Dai-Liao family. This observation enables us to closely analyze Dai-Liao family. As a consequence, we show that the optimal Dai-Liao parameter should be somewhere in interval $(0, \frac{1}{2})$. We also investigate the global convergence of the method under suitable assumptions. Finally, we show that the new method is efficient, and outperforms CG-DESCENT.

The paper is organized as follow: In Section 2, we present some motivations and background of the method. Introducing penalty parameter M is the subject of Section 3. The global convergence of the new method is investigated in Section 4, and numerical results are reported in Section 5. Finally, conclusions and discussions are made in the last section.

2. Motivations and backgrounds

In this section, we introduce a new family of conjugate gradient methods by solving (10).

To solve (10), we should replace g_{k+2} by some of its appropriate estimation, because it is not available in the current iteration.

Here, we consider the quadratic approximation of the objective function,

$$\Phi(d) = f_{k+1} + g_{k+1}^T d + \frac{1}{2} d^T B_{k+1} d,$$

and, take $\nabla\Phi(\alpha_{k+1} d_{k+1})$ as an estimation of g_{k+2}. It is easy to see that

$$\nabla\Phi(\alpha_{k+1} d_{k+1}) = \alpha_{k+1} B_{k+1} d_{k+1} + g_{k+1}. \tag{11}$$

Now, we modify (11) and set

$$g_{k+2} := t B_{k+1} d_{k+1} + g_{k+1}, \tag{12}$$

where, $t > 0$ is a suitable approximation of α_{k+1}. Substituting (3) and (12) in (10), we obtain

$$\begin{aligned} \beta_k = \frac{1}{X} \Big[& - g_{k+1}^T d_k + 2Mt^2 (s_k^T B_{k+1} g_{k+1})(s_k^T B_{k+1} d_k) \\ & - 2Mt(s_k^T g_{k+1})(s_k^T B_{k+1} d_k) \\ & + 2M(y_k^T g_{k+1})(y_k^T d_k) \Big], \end{aligned} \tag{13}$$

where

$$X = 2Mt^2 (s_k^T B_{k+1} d_k)^2 + 2M(y_k^T d_k)^2.$$

To simplify (13), we use the secant condition $B_{k+1} s_k = y_k$, and obtain the new formula

$$\beta_k = \frac{-1}{2M(1+t^2)} \frac{g_{k+1}^T d_k}{(y_k^T d_k)^2} + \frac{y_k^T g_{k+1}}{y_k^T d_k} - \frac{t}{(1+t^2)} \frac{s_k^T g_{k+1}}{y_k^T d_k}. \tag{14}$$

2.1. An optimal bound for Dai-Liao parameter

Consider the method (2) and (3) with β_k in (14), let M approaches infinity, we have

$$\beta_k = \frac{y_k^T g_{k+1}}{y_k^T d_k} - \frac{t}{(1+t^2)} \frac{s_k^T g_{k+1}}{y_k^T d_k}. \tag{15}$$

This β_k would seem to be the most efficient conjugate gradient parameter, because it increases the probability of satisfying (ii) and (iii), making the method to reflect the good features of the linear conjugate gradient algorithm. It is easy to see that (15) is a member of Dai-Liao family with $\tau = \frac{t}{1+t^2}$. Since $\tau \in (0, \frac{1}{2}]$, it is reasonable to claim that the optimal Dai-Liao parameter τ should be somewhere between 0 and $\frac{1}{2}$.

Unfortunately, we have some difficulties to ensure the sufficient descent condition of the search direction d_{k+1} when it is computed using (15). Obviously, the sufficient descent information are lost when M approaches infinity.

3. The sufficient descent direction

In this section, we intended to overcome the difficulties addressed in Section 2.1 by introducing a suitable penalty parameter M.

As we mentioned earlier, a suitable penalty parameter M should have the two important properties. Firstly, the search direction d_{k+1} equipped with β_k in (14) must satisfy (i), and secondly, M should approach infinity during the iterations. In the following lemma, we propose such a suitable penalty parameter.

Lemma 1. *Assume the method (2) and (3) with the standard Wolfe line search, where β_k is defined in (14), then, for some positive scalars γ_1 and γ_2 satisfying $\gamma_1 + \gamma_2 < 1$, we have*

$$g_{k+1}^T d_{k+1} \leq -(1 - \gamma_1 - \gamma_2) \parallel g_{k+1} \parallel^2, \tag{16}$$

whenever

$$|t - 1| \leq \left(\frac{2\gamma_2 (y_k^T s_k)}{\parallel s_k \parallel^2} \right)^{\frac{1}{2}}, \tag{17}$$

and

$$M = \frac{2\gamma_1}{(1 + t^2) \parallel y_k - \frac{\lambda_k}{2} s_k \parallel^2}, \tag{18}$$

where $\lambda_k \leq 1$ is an arbitrary scalar.

PROOF. We prove lemma by mathematical induction. It is easy to see using (3) that $g_0^T d_0 = - \parallel g_0 \parallel^2 \leq -(1 - \gamma_1 - \gamma_2) \parallel g_0 \parallel^2$. Now, assume

$$g_k^T d_k \leq -(1 - \gamma_1 - \gamma_2) \parallel g_k \parallel^2, \tag{19}$$

we show that (16) holds. Note that using (5) and (19), we have

$$y_k^T d_k = g_{k+1}^T d_k - g_k^T d_k \geq (c_2 - 1) g_k^T d_k > 0. \tag{20}$$

Using (3) and (14), we have

$$g_{k+1}^T d_{k+1} = - \parallel g_{k+1} \parallel^2 + \frac{(y_k^T g_{k+1})(g_{k+1}^T s_k)}{y_k^T s_k} - \frac{1}{2M(1+t^2)} \frac{(g_{k+1}^T s_k)^2}{(y_k^T s_k)^2} - \frac{t}{(1+t^2)} \frac{(s_k^T g_{k+1})^2}{y_k^T s_k}.$$

Since $\lambda_k \leq 1$, we also have $\lambda_k (s_k^T g_{k+1})^2 \leq (s_k^T g_{k+1})^2$, and so, using (20)

$$g_{k+1}^T d_{k+1} \pm \frac{(s_k^T g_{k+1})^2}{2 y_k^T s_k} \leq - \parallel g_{k+1} \parallel^2 + \frac{(y_k - \frac{\lambda_k}{2} s_k)^T g_{k+1} (g_{k+1}^T s_k)}{y_k^T s_k} - \frac{1}{2M(1+t^2)} \frac{(g_{k+1}^T s_k)^2}{(y_k^T s_k)^2} + \left(\frac{1}{2} - \frac{t}{1+t^2}\right) \frac{(s_k^T g_{k+1})^2}{y_k^T s_k}.$$

Now, using the following inequality:

$$ab \leq \frac{t'}{4} a^2 + \frac{1}{t'} b^2,$$

where a, b and t' are positive scalars, we have

$$g_{k+1}^T d_{k+1} \leq - \parallel g_{k+1} \parallel^2 + \frac{t'}{4} \left((y_k - \frac{\lambda_k}{2} s_k)^T g_{k+1} \right)^2 + \frac{1}{t'} \frac{(g_{k+1}^T s_k)^2}{(y_k^T s_k)^2} - \frac{1}{2M(1+t^2)} \frac{(g_{k+1}^T s_k)^2}{(y_k^T s_k)^2} + \frac{(t-1)^2}{2(1+t^2)} \frac{(s_k^T g_{k+1})^2}{y_k^T s_k}.$$

Let $t' = 2M(1+t^2)$,

$$g_{k+1}^T d_{k+1} \leq - \parallel g_{k+1} \parallel^2 + \frac{M(1+t^2)}{2} \left((y_k - \frac{\lambda_k}{2} s_k)^T g_{k+1} \right)^2 + \frac{(t-1)^2}{2(1+t^2)} \frac{(s_k^T g_{k+1})^2}{y_k^T s_k}.$$

Finally, the Cauchy-Schwarz inequality implies

$$g_{k+1}^T d_{k+1} \leq - \left[1 - \frac{M(1+t^2)}{2} \parallel y_k - \frac{\lambda_k}{2} s_k \parallel^2 - \frac{(t-1)^2}{2(1+t^2)} \frac{\parallel s_k \parallel^2}{y_k^T s_k} \right] \parallel g_{k+1} \parallel^2 . \tag{21}$$

Now, the proof is completed using (17), (18) and (21). □

In follow, we present some remarks concerning to Lemma 1

Remark 1. Interestingly, we can use λ_k to accelerate the growth-rate of M. More exactly, we set

$$\lambda_k = \min(1, \frac{2s_k^T y_k}{\| s_k \|^2}), \tag{22}$$

which minimizes the denominator of (18).

Remark 2. In Newton type algorithms, it is common to set $\alpha_k = 1$, initially, because it finally leads to quadratic convergence. However, the step size variations are not predictable in conjugate gradient algorithms. Thus, we use the following heuristic to update t.

$$t = \begin{cases} \alpha_k, & |\alpha_k - 1| \leq \left(\frac{2\gamma_2(y_k^T s_k)}{\|s_k\|^2}\right)^{\frac{1}{2}}, \\ 1 + \left(\frac{2\gamma_2(y_k^T s_k)}{\|s_k\|^2}\right)^{\frac{1}{2}}, & \text{O.W.} \end{cases} \tag{23}$$

Note that substituting (18) in (14), we have

$$\beta_k = \frac{y_k^T g_{k+1}}{y_k^T d_k} - \frac{t}{1+t^2} \frac{g_{k+1}^T s_k}{y_k^T d_k} - \frac{\| y_k - \frac{\lambda_k}{2} s_k \|^2}{4\gamma_1} \frac{(g_{k+1}^T d_k)}{(y_k^T d_k)^2}, \tag{24}$$

which converts to the most desirable case of (15), when $\| y_k - \frac{\lambda_k}{2} s_k \| = 0$.

4. Global convergence

We now analyze the global convergence of the method (2), (3) and (24) with λ_k and t defined in (22) and (23), respectively.

Here, we assume that the step length α_k satisfies the standard Wolfe conditions (4) and (5). The following standard assumptions are considered in this section:

(A1) The gradient vector g is Lipschitz continuous; namely, there exists a constant $L > 0$ such that

$$\| \nabla f(x) - \nabla f(y) \| \leq L \| x - y \|, \quad x, y \in \mathbb{R}^n.$$

(A2) $f(x)$ is a differentiable and bounded below function on the level set

$$\mathcal{L} = \{x \in \mathbb{R}^n \mid f(x) \leq f(x_0)\}.$$

(A3) The generated sequence of iterates, x_k, is bounded.

The global convergence of descent methods with standard Wolfe line search essentially relies on the following Zoutendijk condition.

Lemma 2. *suppose that A1-A3 holds. consider any descent method of the form (2) where α_k is determined by standard Wolfe line search. Then we have that*

$$\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\| d_k \|^2} < \infty. \tag{25}$$

The global convergence analysis of this section is basically similar to that of Hager and Zhang in [7].

4.1. Global convergence for strongly convex functions

First, we have the following definition:

Definition 1. A differentiable function f is called strongly convex, if there exists a constant $\mu > 0$ such that

$$\mu \| x - y \|^2 \leq (\nabla f(x) - \nabla f(y))^T (x - y),$$

for all $x, y \in \mathcal{L}$.

Fortunately, we can prove the global convergence of the method without any modification of β_k when f is strongly convex.

Theorem 3. *Assume the method (2) and (3) with β_k defined in (24) is implemented on a strongly convex function f, then, we have* $\lim_{k\to\infty} g_k = 0$.

PROOF. We first show that the search direction d_{k+1} is bounded above.

Using (A1), Definition 1, Cauchy-Schwarz inequality and the fact that $\mu\alpha_k \parallel d_k \parallel^2 \leq y_k^T d_k$, we have

$$\begin{aligned}
|\beta_k| &\leq \frac{\parallel y_k \parallel\parallel g_{k+1} \parallel}{\mu\alpha_k \parallel d_k \parallel^2} + \frac{t}{1+t^2}\frac{\parallel s_k \parallel\parallel g_{k+1} \parallel}{\mu\alpha_k \parallel d_k \parallel^2} + \frac{\parallel y_k - \frac{\lambda_k}{2}s_k \parallel^2\parallel d_k \parallel\parallel g_{k+1} \parallel}{4\gamma_1\mu^2 \parallel d_k \parallel^4 \alpha_k^2} \\
&\leq \frac{(L+1) \parallel s_k \parallel\parallel g_{k+1} \parallel}{\mu\alpha_k \parallel d_k \parallel^2} + \frac{(L+1)^2 \parallel s_k \parallel^2\parallel d_k \parallel\parallel g_{k+1} \parallel}{4\gamma_1\mu^2 \parallel d_k \parallel^4 \alpha_k^2} \\
&\leq \frac{(L+1) \parallel g_{k+1} \parallel}{\mu \parallel d_k \parallel} + \frac{(L+1)^2 \parallel g_{k+1} \parallel}{4\gamma_1\mu^2 \parallel d_k \parallel} \\
&\leq \Big(\frac{L+1}{\mu} + \frac{(L+1)^2}{4\gamma_1\mu^2}\Big)\frac{\parallel g_{k+1} \parallel}{\parallel d_k \parallel}.
\end{aligned}$$

Now,

$$\parallel d_{k+1} \parallel\leq\parallel g_{k+1} \parallel + |\beta_k| \parallel d_k \parallel\leq \Big(\frac{L+1}{\mu} + \frac{(L+1)^2}{4\gamma_1\mu^2}\Big) \parallel g_{k+1} \parallel . \tag{26}$$

As a consequence, (16), (26) and the Zoutendijk condition imply that

$$\sum_{k=1}^{\infty} \parallel g_k \parallel^2 < \infty,$$

which completes the proof. □

4.2. Global convergence for general functions

Here, we consider the following modification of β_k:

$$\beta_k^T = \max(\beta_k, \chi_k), \tag{27}$$

where χ_k is a real valued function having the following properties:

(p1) $|\chi_k| \parallel d_k \parallel$ is bounded above.

((p2)) For some $\epsilon < 1$,

$$\chi_k \leq \frac{\epsilon \parallel g_{k+1} \parallel^2}{g_{k+1}^T d_k},$$

whenever $g_{k+1}^T d_k > 0$.

Note that, (p1) and (p2) ensure that the search direction d_k is bounded, and the sufficient descent property (6) holds, respectively.

Lemma 4. *Suppose the method (2) and (3) with β_k^T defined in (27). Moreover, assume that the standard Wolfe line search conditions (4) and (5) are used, then*

$$g_{k+1}^T d_{k+1} \leq -\max(1-\epsilon, 1-\gamma_1-\gamma_2) \parallel g_{k+1} \parallel^2, \tag{28}$$

Proof. If $\beta_k^T = \beta_k$, then (16) implies that

$$g_{k+1}^T d_{k+1} \leq -(1-\gamma_1-\gamma_2) \parallel g_{k+1} \parallel^2 .$$

If $\beta_k^T = \chi_k$ and $g_{k+1}^T d_k < 0$, then our previous analysis and the fact that $\beta_k \leq \chi_k$ imply that

$$\begin{aligned} g_{k+1}^T d_{k+1} = - \parallel g_{k+1} \parallel^2 + \chi_k g_{k+1}^T d_k &\leq - \parallel g_{k+1} \parallel^2 + \beta_k g_{k+1}^T d_k \\ &\leq -(1-\gamma_1-\gamma_2) \parallel g_{k+1} \parallel^2 . \end{aligned}$$

If $\beta_k^T = \chi_k$ and $g_{k+1}^T d_k > 0$, then (p2) implies

$$\begin{aligned} g_{k+1}^T d_{k+1} &= - \parallel g_{k+1} \parallel^2 + \chi_k g_{k+1}^T d_k \\ &\leq -(1-\epsilon) \parallel g_{k+1} \parallel^2 . \end{aligned}$$

Now, the proof is completed. □

The following lemma is analogue of Lemma 4.3 in [14]

Lemma 5. *Suppose A1-A3 holds, then for method (2) and (3) with β_k^T defined in (27), and a line search satisfying standard Wolfe conditions, then, we have*

$$\sum_{k=1}^{\infty} \parallel u_k - u_{k-1} \parallel^2 < \infty,$$

where, $u_k = \frac{d_k}{\|d_k\|}$, *whenever* $\inf \parallel g_k \parallel \neq 0$.

Proof. The proof is basically similar to Lemma 4.3 in [14]. We let

$$z_k^{(1)} = \max(\beta_k - \chi_k, 0),$$

and

$$z_k^{(2)} = \chi_k.$$

It is easy to see that $\beta_k^T = z_k^{(1)} + z_k^{(2)}$. Let

$$w_k = \frac{-g_k + z_{k-1}^{(2)} d_{k-1}}{\| d_k \|},$$

and

$$\delta_k = \frac{z_{k-1}^{(1)} \| d_{k-1} \|}{\| d_k \|}.$$

Following the statements of the proof of Lemma 4.3 in [14], we reach to

$$\| u_k - u_{k-1} \| \leq 2 \| w_k \|,$$

Since we assumed that $|\chi_k| \| d_k \|$ is bounded, there exists a constant $\epsilon > 0$ such that

$$\| -g_k + z_{k-1}^{(2)} d_{k-1} \| \leq \| g_k \| + |\chi_{k-1}| \| d_{k-1} \| \leq \epsilon.$$

Thus,

$$\| u_k - u_{k-1} \| \leq \frac{2\epsilon}{\| d_k \|}.$$

Now, the proof is completed using (28) and the Zoutendijk condition. □

There are some special choices of χ_k in literatures. For example, the Hager and Zhang choice of

$$\chi_k = \frac{-1}{\| d_k \| \min(\eta, \| g_k \|)}, \tag{29}$$

and the Dai and Kou choice of

$$\chi_k = \eta \frac{g_{k+1}^T d_k}{\| d_k \|^2}.$$

We now state the main result.

Theorem 6. *Suppose that A1-A3 holds. If the method (2) and (3) with β_k^T defined in (27) is implemented on f and the standard Wolfe line search conditions (4) and (5) are used, then*

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

Proof. Assume that there exists η_1 such that $\| g_k \| > \eta_1$ for all k.

If there exists a subsequence k_j such that $\beta_{k_j}^T = \chi_{k_j}$, then, using (p1), we have for some $\epsilon > 0$,

$$\| d_{k_j+1} \| = \| -g_{k_j+1} + \beta_{k_j}^T d_{k_j} \| \leq \| g_{k_j+1} \| + |\chi_{k_j}| \, \| d_{k_j} \| \leq \epsilon.$$

This bound for d_{k_j+1} and (28) yield a contradiction using Zoutendijk condition.

We now assume that $\beta_k^T = \beta_k$ for sufficiently large k. Following the statements of the proof of Theorem 3.2 in [7], we only address the changes in the parts of the proof.

Part I. (A bound for β_k) Using inequalities

$$y_k^T d_k \geq (1 - c_2) \max(1 - \epsilon, 1 - \gamma_1 - \gamma_2)\eta_1^2, \tag{30}$$

and

$$\frac{|g_{k+1}^T d_k|}{|y_k^T d_k|} \leq \max(\frac{c_2}{1 - c_2}, 1), \tag{31}$$

see [7], we show that there exists a constant $C > 0$ such that

$$|\beta_k| \leq C \, \| s_k \| \, . \tag{32}$$

It is easy to see using (24) that

$$|\beta_k| \leq \frac{|y_k^T g_{k+1}|}{y_k^T d_k} + \frac{\| y_k - \frac{\lambda_k}{2} s_k \|^2 \, |g_{k+1}^T d_k|}{4\gamma_1 (y_k^T d_k)^2} + \frac{t}{(1 + t^2)} \frac{|s_k^T g_{k+1}|}{y_k^T d_k}.$$

Now, $\frac{t}{1+t^2} < 1$ implies that

$$|\beta_k| \leq \frac{1}{y_k^T d_k} \left[|y_k^T g_{k+1}| + \| y_k - \frac{\lambda_k}{2} s_k \|^2 \, \frac{|g_{k+1}^T d_k|}{y_k^T d_k} + |s_k^T g_{k+1}| \right].$$

Using Cauchy-Schwarz inequality,

$$|\beta_k| \leq \frac{1}{y_k^T d_k} \left[\eta_2 L \, \| s_k \| + \frac{(L+1)^2}{4\gamma_1} \, \| s_k \|^2 \, \frac{|g_{k+1}^T d_k|}{y_k^T d_k} + \eta_2 \, \| s_k \| \right]. \tag{33}$$

where η_2 is an upper bound on $\| g_k \|$.

Now, it is easy to see from (33) that using (A1), (A3), (30) and (31), there exist a constant $C > 0$ such that (32) holds. The rest of the proof is essentially similar to Theorem 3.2 in [7]. $\square$

5. Numerical results

We now investigate the numerical comparisons of our algorithms, that is, the method (2), (3) and (24) with λ_k and t defined in (22) and (23), respectively, with a modification version of original CG-DESCENT [7] based on the work by Dai and Kou [14]. The two algorithm were coded in MATLAB 2007 and tested on a 2.4 Intel Core 2Duo processor computer with 2GB of RAM. We compare the two algorithms on 154 unconstrained problems of CUTEr collection [15]. The performance profile of Dolan and Moré [16] is used to compare the efficiency of the algorithms. Furthermore, we used the CG-DESCENT line search procedure with the initial parameters reported in [7]. As in the CG-DESCENT, algorithms terminate if either

$$\| \nabla f(x_k) \|_\infty \leq \max(10^{-6}, 10^{-12} \| \nabla f(x_1) \|_\infty),$$

or the number of iterations exceed 50000. The problems CLPLATE(A-C), NONMSQRT, NONCVXUN, RAYBEND(L-S), SBRYBND, SCOSINE and SCURLY(10-30) were removed, because the iteration count limit was reached, for the two algorithms, before having the chance to detect convergence . The MATLAB codes and raw data of numerical tests are available from website:

WWW.wp.kntu.ac.ir/smfatemi/publications.htm

We choose $\gamma_1 = 0.98$ and $\gamma_2 = 0.01$ in our numerical test. These values clearly ensure that the penalty parameter M in (18) is as large as possible.
Figures 1-4 show the performance profile for the number of iteration, the number of function and gradient evaluations and CPU time. As these figures indicate, the new algorithm outperforms CG-DESCENT in every respect. This clearly confirms the effectiveness of the new method.

6. Conclusions

We have presented a nonlinear conjugate gradient method for unconstrained optimization based on solving a new optimization problem. We showed that the new method is a subclass of Dai-Liao family, the fact that enabled us to analyze the family, closely. As a consequence, an optimal bound for Dai-Liao parameter is presented. The global convergence of the method was investigated under mild assumptions. The numerical comparing results indicated that the new method is efficient and outperforms CG-DESCENT.

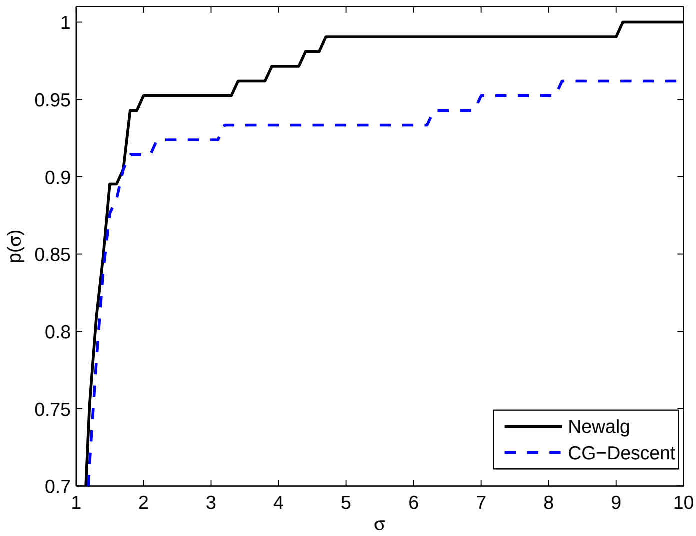

Figure 1: Iteration performance profile.

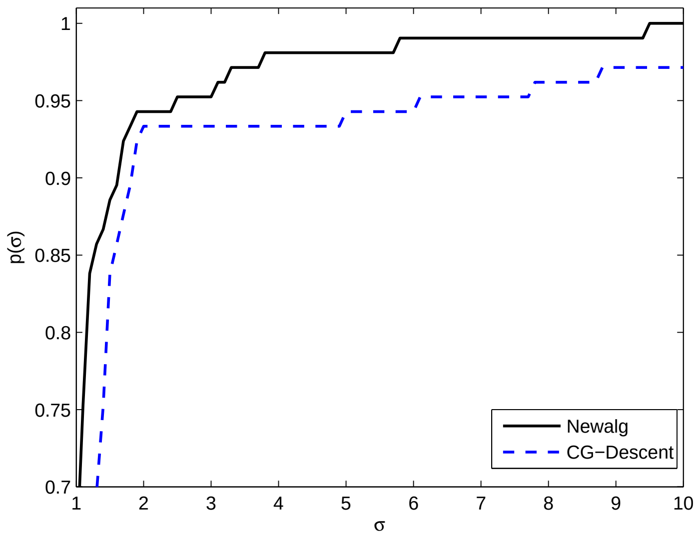

Figure 2: Number of function evaluation performance profile.

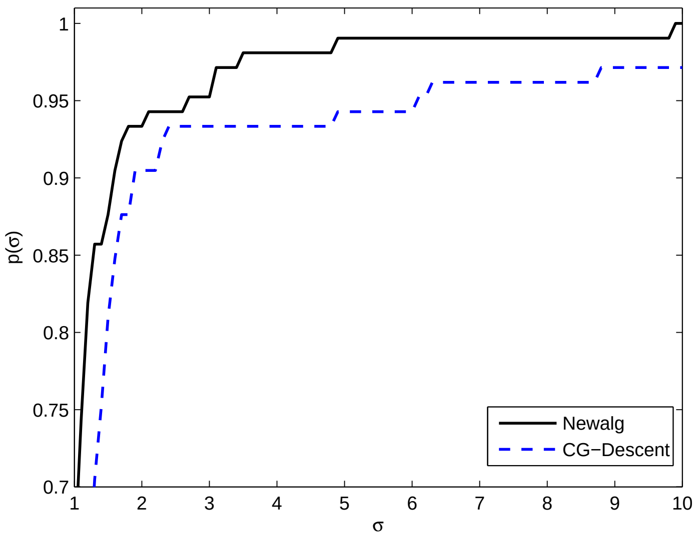

Figure 3: Number of gradient evaluation performance profile.

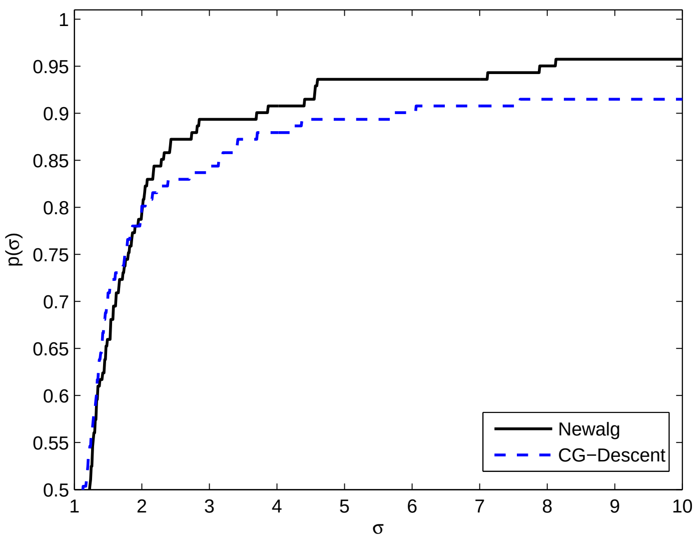

Figure 4: CPU time performance profile.

Acknowledgments

The author thank the Research Council of K. N. Toosi University of Technology for supporting this work.

References

References

[1] M.R. Hestenes, E. Stiefel, Method of conjugate gradient for solving linear system, J. Res. Nat. Bur. Stand. 49 (1952) 409-436.

[2] R. Fletcher, C.M. Reeves, Function minimization by conjugate gradients, Comput. J. 7 (1964) 149-154.

[3] E. Polak, G. Ribière, Not sur la convergence de méthods directions conjugées, Revue Franccaise d'Informatique et de Recherche opiérationnelle. 16 (1966) 35-43.

[4] B.T. Polyak, The conjugate gradient method in extreme problems, USSR Comp Math and Math. Phys. 9 (1969) 94-112.

[5] Y.H. Dai, Y. Yuan, A nonlinear conjugate gradient method with a strong global convergence properties, Siam Journal on Optimization. 10 (1999) 177-182.

[6] Y.H. Dai, L.Z. Liao, New conjugacy conditions and related nonlinear conjugate gradient methods, Applied Mathematics and Optimization. 43 (2001) 87-101.

[7] W.W. Hager, H. Zhang, A new conjugate gradient method with guaranteed descent and an efficient line search, Siam Journal on Optimization, 16 (2005) 170-192.

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

[9] S. Babaie-Kafaki, R. Ghanbari, Two optimal Dai-Liao conjugate gradient methods, Optimization, DOI:10.1080/02331934.2014.938072, (2014).

[10] W.W. Hager, H. Zhang, A survey of nonlinear conjugate gradient methods, Pacific Journal of Optimization. 2:1 (2006) 335-358.

[11] Y.H. Dai, Nonlinear Conjugate Gradient Methods, Wiley Encyclopedia of Operations Research and Management Science, 2011.

[12] J.C. Gilbert, J. Nocedal, Global convergence properties of conjugate gradient methods for optimization. Siam Journal on Optimization. 2 (1992) 21-42.

[13] M. Fatemi, An optimal parameter for Dai-Liao family of conjugate gradient methods, J. Optim Theory Appl, DOI 10.1007/s10957-015-0786-9. 2015.

[14] Y.H. Dai, C.X. Kou, A nonlinear conjugate gradient algorithm with an optimal property and an improved wolfe line search, Siam Journal on Optimization. 23 (2013) 296-320.

[15] N.I.M. Gould, D. Orban, Ph.L. Toint, CUTEr (and SifDec), a constrained and unconstrained testing environment, revisited, Technical Report TR/PA/01/04, CERFACS, Toulouse, France, 2001.

[16] E.D. Dolan, J.J. Moré, Benchmarking optimization software with performance profile, Math. Programming, 91 (2002) 201-213.