Journal of Inequalities and Applications
a SpringerOpen Journal

RESEARCH Open Access

A scaled three-term conjugate gradient method for unconstrained optimization

Ibrahim Arzuka*, Mohd R Abu Bakar and Wah June Leong

*Correspondence:
arzukaibrahim@yahoo.com
Institute for Mathematical Research, Universiti Putra Malaysia, Serdang, Selangor 43400, Malaysia

Abstract

Conjugate gradient methods play an important role in many fields of application due to their simplicity, low memory requirements, and global convergence properties. In this paper, we propose an efficient three-term conjugate gradient method by utilizing the DFP update for the inverse Hessian approximation which satisfies both the sufficient descent and the conjugacy conditions. The basic philosophy is that the DFP update is restarted with a multiple of the identity matrix in every iteration. An acceleration scheme is incorporated in the proposed method to enhance the reduction in function value. Numerical results from an implementation of the proposed method on some standard unconstrained optimization problem show that the proposed method is promising and exhibits a superior numerical performance in comparison with other well-known conjugate gradient methods.

Keywords: unconstrained optimization; nonlinear conjugate gradient method; quasi-Newton methods

1 Introduction

In this paper, we are interested in solving nonlinear large scale unconstrained optimization problems of the form

$$\min f(x), \quad x \in \Re^n, \tag{1}$$

where $f : \Re^n \to \Re$ is an at least twice continuously differentiable function. A nonlinear conjugate gradient method is an iterative scheme that generates a sequence $\{x_k\}$ of an approximation to the solution of (1), using the recurrence

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, 2, 3, \dots, \tag{2}$$

where $\alpha_k > 0$ is the steplength determined by a line search strategy which either minimizes the function or reduces it sufficiently along the search direction and d_k is the search direction defined by

$$d_k = \begin{cases} -g_k; & k = 0, \\ -g_k + \beta_k d_{k-1}; & k \ge 1, \end{cases}$$

where g_k is the gradient of f at a point x_k and β_k is a scalar known as the conjugate gradient parameter. For example, Fletcher and Reeves (FR) [1], Polak-Ribiere-Polyak (PRP) [2], Liu and Storey (LS) [3], Hestenes and Stiefel (HS) [4], Dai and Yuan (DY) [5] and Fletcher (CD) [6] used an update parameter, respectively, given by

$$\beta_k^{\mathrm{FR}} = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}}, \qquad \beta_k^{\mathrm{PRP}} = \frac{g_k^T y_{k-1}}{g_{k-1}^T g_{k-1}}, \qquad \beta_k^{\mathrm{LS}} = \frac{-g_k^T y_{k-1}}{d_{k-1}^T g_{k-1}},$$
$$\beta_k^{\mathrm{HS}} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, \qquad \beta_k^{\mathrm{DY}} = \frac{g_k^T g_k}{d_{k-1}^T y_{k-1}}, \qquad \beta_k^{\mathrm{CD}} = -\frac{g_k^T g_k}{d_{k-1}^T y_{k-1}},$$

where $y_{k-1} = g_k - g_{k-1}$. If the objective function is quadratic, with an exact line search the performances of these methods are equivalent. For a nonlinear objective function different β_k lead to a different performance in practice. Over the years, after the practical convergence result of Al-Baali [7] and later of Gilbert and Nocedal [8] attention of researchers has been on developing on conjugate gradient methods that possesses the sufficient descent condition

$$g_k^T d_k \leq -c\|g_k\|^2, \tag{3}$$

for some constant $c > 0$. For instance the CG-DESCENT of Hager and Zhang [9]

$$\beta_k^{\mathrm{HZ}} = \max\left\{\beta_k^N, \eta_k\right\}, \tag{4}$$

where

$$\beta_k^N = \frac{1}{d_{k-1}^T y_{k-1}}\left(y_{k-1} - 2d_{k-1}\frac{\|y_{k-1}\|^2}{d_{k-1}^T y_{k-1}}\right)^T g_k$$

and

$$\eta_k = \frac{-1}{\|d_{k-1}\| \min\{\|g_{k-1}\|, \eta\}},$$

which is based on the modification of HS method. Another important class of conjugate gradient methods is the so-called three-term conjugate gradient method in which the search direction is determined as a linear combination of g_k, s_k, and y_k as

$$d_k = -g_k - \tau_1 s_k + \tau_2 y_k, \tag{5}$$

where τ_1 and τ_2 are scalar. Among the generated three-term conjugate gradient methods in the literature we have the three-term conjugate methods proposed by Zhang *et al.* [10, 11] by considering a descent modified PRP and also a descent modified HS conjugate gradient method as

$$d_{k+1} = -g_{k+1} + \left(\frac{g_{k+1}^T y_k}{g_k^T g_k}\right)d_k - \left(\frac{g_{k+1}^T d_k}{g_k^T g_k}\right)y_k,$$

and

$$d_{k+1} = -g_{k+1} + \left(\frac{g_{k+1}^T y_k}{s_k^T y_k}\right)s_k - \left(\frac{g_{k+1}^T s_k}{s_k^T y_k}\right)y_k,$$

where $s_k = x_{k+1} - x_k$. An attractive property of these methods is that at each iteration, the search direction satisfies the descent condition, namely $g_k^T d_k = -c\|g_k\|^2$ for some constant $c > 0$. In the same manner, Andrei [12] considers the development of a three-term conjugate gradient method from the BFGS updating scheme of the inverse Hessian approximation restarted as an identity matrix at every iteration where the search direction is given by

$$d_{k+1} = -g_{k+1} + \frac{y_k^T g_{k+1}}{y_k^T s_k} - \left(y - 2\frac{\|y_k\|^2}{y_k^T s_k}\right)^T \frac{s_k^T g_{k+1}}{y_k^T s_k} s_k - \left(\frac{s_k^T g_{k+1}}{y_k^T s_k}\right)y_k.$$

An interesting feature of this method is that both the sufficient and the conjugacy conditions are satisfied and we have global convergence for a uniformly convex function. Motivated by the good performance of the three-term conjugate gradient method, we are interested in developing a three-term conjugate gradient method which satisfies both the sufficient descent condition, the conjugacy condition, and global convergence. The remaining part of this paper is structured as follows: Section 2 deals with the derivation of the proposed method. In Section 3, we present the global convergence properties. The numerical results and discussion are reported in Section 4. Finally, a concluding remark is given in the last section.

2 Conjugate gradient method via memoryless quasi-Newton method

In this section, we describe the proposed method which would satisfied both the sufficient descent and the conjugacy conditions. Let us consider the DFP method, which is a quasi-Newton method belonging to the Broyden class [13]. The search direction in the quasi-Newton methods is given by

$$d_k = -H_k g_k, \tag{6}$$

where H_k is the inverse Hessian approximation updated by the Broyden class. This class consists of several updating schemes, the most famous being the BFGS and the DFP; if H_k is updated by the DFP then

$$H_{k+1} = H_k + \frac{s_k s_k^T}{s_k^T y_k} - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k}, \tag{7}$$

such that the secant equation

$$H_{k+1} y_k = s_k \tag{8}$$

is satisfied. This method is also known as a variable metric method, developed by Davidon [14], Fletcher and Powell [15]. A remarkable property of this method is that it is a conjugate direction method and one of the best quasi-Newton methods that encompassed the advantage of both the Newton method and the steepest descent method, while avoiding their shortcomings [16]. Memoryless quasi-Newton methods are other techniques for solving (1), where at every step the inverse Hessian approximation is updated as an identity matrix. Thus, the search direction can be determined without requiring the storage of any matrix. It was proposed by Shanno [17] and Perry [18]. The classical conjugate gradient methods PRP [2] and FR [1] can be seen as memoryless BFGS (see Shanno [17]). We proposed our three-term conjugate gradient method by incorporating the DFP updating scheme of the inverse Hessian approximation (7), within the frame of a memoryless quasi-Newton method where at each iteration the inverse Hessian approximation is restarted as a multiple of the identity matrix with a positive scaling parameter as

$$Q_{k+1} = \mu_k I + \frac{s_k s_k^T}{s_k^T y_k} - \mu_k \frac{y_k y_k^T}{y_k^T y_k}, \tag{9}$$

and thus, the search direction is given by

$$d_{k+1} = -Q_{k+1} g_{k+1} = -\mu_k g_{k+1} - \frac{s_k^T g_{k+1}}{s_k^T y_k} s_k + \mu_k \frac{y_k^T g_{k+1}}{y_k^T y_k} y_k. \tag{10}$$

Various strategies can be considered in deriving the scaling parameter μ_k; we prefer the following which is due to Wolkowicz [19]:

$$\mu_k = \frac{s_k^T s_k}{y_k^T s_k} - \sqrt{\left(\frac{s_k^T s_k}{y_k^T s_k}\right)^2 - \frac{s_k^T s_k}{y_k^T y_k}}. \tag{11}$$

The new search direction is then given by

$$d_{k+1} = -\mu_k g_{k+1} - \varphi_1 s_k + \varphi_2 y_k, \tag{12}$$

where

$$\varphi_1 = \frac{s_k^T g_{k+1}}{s_k^T y_k} \tag{13}$$

and

$$\varphi_2 = \mu_k \frac{y_k^T g_{k+1}}{y_k^T y_k}. \tag{14}$$

We present the algorithm of the proposed method as follows.

2.1 Algorithm (STCG)

In this section, we present the algorithm of the proposed method. It has been reported that the line search in conjugate gradient method performs more function evaluations so as to obtain a desirable steplength α_k due to poor scaling of the search direction (see Nocedal [20]). As a consequence, we incorporate the acceleration scheme proposed by Andrei [21], so as to have some reduction in the function evaluations. The new approximation to the minimum instead of (2) is determined by

$$x_{k+1} = x_k + \alpha_k \vartheta_k d_k, \tag{15}$$

where $\vartheta_k = \frac{-r_k}{q_k}$, $r_k = \alpha_k g_k^T d_k$, $q_k = -\alpha_k (g_k - g_z) d_k = -\alpha_k y_k d_k$, $g_z = \nabla f(z)$ and $z = x_k + \alpha_k d_k$.

Algorithm 1

Step 1. Select an initial point x_o and determine $f(x_o)$ and $g(x_o)$. Set $d_o = -g_o$ and $k = 0$.
Step 2. Test the stopping criterion $\|g_k\| \leq \epsilon$, if satisfied stop. Else go to Step 3.
Step 3. Determine the steplength α_k as follows:
Given $\delta \in (0,1)$ and p_1, p_2, with $0 < p_1 < p_2 < 1$.
(i) Set $\alpha = 1$.
(ii) Test the relation

$$f(x + \alpha d_k) - f(x_k) \leq \alpha \delta g_k^T d_k. \tag{16}$$

(iii) If (16) is satisfied, then $\alpha_k = \alpha$ and go to Step 4 else choose a new $\alpha \in [p_1\alpha, p_2\alpha]$ and go to (ii).
Step 4. Determine $z = x_k + \alpha_k d_k$, compute $g_z = \nabla f(z)$ and $y_k = g_k - g_z$.
Step 5. Determine $r_k = \alpha_k g_k^T d_k$ and $q_k = -\alpha_k y_k^T d_k$.
Step 6. If $q_k \neq 0$, then $\vartheta_k = \frac{r_k}{q_k}$, $x_{k+1} = x_k + \vartheta_k \alpha_k d_k$ else $x_{k+1} = x_k + \alpha_k d_k$.
Step 7. Determine the search direction d_{k+1} by (12) where μ_k, φ_1, and φ_2 are computed by (11), (13), and (14), respectively.
Step 8. Set $k := k + 1$ and go to Step 2.

3 Convergence analysis

In this section, we analyze the global convergence of the propose method, where we assume that $g_k \neq 0$ for all $k \geq 0$ else a stationary point is obtained. First of all, we show that the search direction satisfies the sufficient descent and the conjugacy conditions. In order to present the results, the following assumptions are needed.

Assumption 1 The objective function f is convex and the gradient g is Lipschitz continuous on the level set

$$K = \{x \in \Re^n | f(x) \leq f(x_0)\}. \tag{17}$$

Then there exist some positive constants ψ_1, ψ_2, and L such that

$$\|g(x) - g(y)\| \leq L\|x - y\| \tag{18}$$

and

$$\psi_1 \|z\|^2 \leq z^T G(x) z \leq \psi_2 \|z\|^2, \tag{19}$$

for all $z \in R^n$ and $x, y \in K$ where $G(x)$ is the Hessian matrix of f.

Under Assumption 1, we can easily deduce that

$$\psi_1 \|s_k\|^2 \leq s_k^T y_k \leq \psi_2 \|s_k\|^2, \tag{20}$$

where $s_k^T y_k = s_k^T \bar{G} s_k$ and $\bar{G} = \int_0^1 G(x_k + \lambda s_k) s_k \, d\lambda$. We begin by showing that the updating matrix (9) is positive definite.

Lemma 3.1 *Suppose that Assumption* 1 *holds*; *then the matrix* (9) *is positive definite.*

Proof In order to show that the matrix (9) is positive definite we need to show that μ_k is well defined and bounded. First, by the Cauchy-Schwarz inequality we have

$$\begin{aligned}\left(\frac{s_k^T s_k}{y_k^T s_k}\right)^2 - \frac{s_k^T s_k}{y_k^T y_k} &= \frac{(s_k^T s_k)((s_k^T s_k)(y_k^T y_k) - (y_k^T s_k)^2)}{(y_k^T s_k)^2(y_k^T y_k)} \\ &\geq 0,\end{aligned}$$

and this implies that the scaling parameter μ_k is well defined. It follows that

$$\begin{aligned}0 < \mu_k &= \frac{s_k^T s_k}{y_k^T s_k} - \left(\left(\frac{s_k^T s_k}{y_k^T s_k}\right)^2 - \frac{s_k^T s_k}{y_k^T y_k}\right)^{\frac{1}{2}} \\ &\leq \frac{s_k^T s_k}{y_k^T s_k} \leq \frac{\|s_k\|^2}{\psi_1^2\|s_k\|^2} = \frac{1}{\psi_1^2}.\end{aligned}$$

When the scaling parameter is positive and bounded above, then for any non-zero vector $p \in \Re^n$ we obtain

$$\begin{aligned}p^T Q_{k+1} p &= \mu_k p^T p I + \frac{p^T s_k s_k^T p}{s_k^T y_k} - \mu_k \frac{p^T y_k y_k^T p}{y_k^T y_k} \\ &= \mu_k \left[\frac{(p^T p)(y_k^T y_k) - p^T y_k y_k^T p}{y_k^T y_k}\right] + \frac{(p^T s_k)^2}{s_k^T y_k}.\end{aligned}$$

By the Cauchy-Schwarz inequality and (20), we have $(p^T p)(y_k^T y_k) - (p^T y_k)(y_k^T p) \geq 0$ and $y_k^T s_k > 0$, which implies that the matrix (9) is positive definite $\forall k \geq 0$.

Observe also that

$$\begin{aligned}\operatorname{tr}(Q_{k+1}) &= \operatorname{tr}(\mu_k I) + \frac{s_k^T s_k}{s_k^T y_k} - \mu_k \frac{y_k^T y_k}{y_k^T y_k} \\ &= (n-1)\mu_k + \frac{s_k^T s_k}{s_k^T y_k} \\ &\leq \frac{n-1}{\psi_1^2} + \frac{\|s_k\|^2}{\psi_1\|s_k\|^2} \\ &= \frac{\psi_1 + n - 1}{\psi_1^2}.\end{aligned} \tag{21}$$

Now,

$$0 < \frac{1}{\psi_2} \leq \left(\frac{s_k^T s_k}{y_k^T s_k}\right) \leq \operatorname{tr}(Q_{k+1}) \leq \frac{\psi_1 + n - 1}{\psi_1^2}. \tag{22}$$

Thus, $\operatorname{tr}(Q_{k+1})$ is bounded. On the other hand, by the Sherman-Morrison House-Holder formula (Q_{k+1}^{-1} is actually the memoryless updating matrix updated from $\frac{1}{\mu_k}I$ using the direct DFP formula), we can obtain

$$Q_{k+1}^{-1} = \frac{1}{\mu_k} I - \frac{1}{\mu_k}\frac{y_k s_k^T + s_k y_k^T}{s_k^T y_k} + \left(1 + \frac{1}{\mu_k}\frac{s_k^T s_k}{s_k^T y_k}\right)\frac{y_k y_k^T}{s_k^T y_k}. \tag{23}$$

We can also establish the boundedness of $\mathrm{tr}(Q_{k+1}^{-1})$ as

$$\begin{aligned}
\mathrm{tr}\bigl(Q_{k+1}^{-1}\bigr) &= \mathrm{tr}\biggl(\frac{1}{\mu_k}I\biggr) - \frac{2}{\mu_k}\frac{s_k^T y_k}{s_k^T y_k} + \frac{\|y_k\|^2}{s_k^T y_k} + \frac{1}{\mu_k}\frac{\|s_k\|^2\|y_k\|^2}{(s_k^T y_k)^2} \\
&\le \frac{n}{\mu_k} - \frac{2}{\mu_k} + \frac{L^2\|s_k\|^2}{\psi_1\|s_k\|^2} + \frac{1}{\mu_k}\frac{L^2\|s_k\|^4}{\psi_1^2\|s_k\|^4} \\
&\le \frac{(n-2)}{\psi_1^2} + \frac{L^2}{\psi_1} + \frac{L^2}{\psi_1^4} \\
&= \omega,
\end{aligned} \tag{24}$$

where $\omega = \frac{(n-2)}{\psi_1^2} + \frac{L^2}{\psi_1} + \frac{L^2}{\psi_1^4} > 0$, for $n \ge 2$. □

Now, we shall state the sufficient descent property of the proposed search direction in the following lemma.

Lemma 3.2 *Suppose that Assumption* 1 *holds on the objective function f then the search direction* (12) *satisfies the sufficient descent condition* $g_{k+1}^T d_{k+1} \le -c\|g_{k+1}\|^2$.

Proof Since $-g_{k+1}^T d_{k+1} \ge \frac{1}{\mathrm{tr}(Q_{k+1}^{-1})}\|g_{k+1}\|^2$ (see for example Leong [22] and Babaie-Kafaki [23]), then by using (24) we have

$$-g_{k+1}^T d_{k+1} \ge c\|g_{k+1}\|^2, \tag{25}$$

where $c = \min\{1, \frac{1}{\omega}\}$. Thus,

$$g_{k+1}^T d_{k+1} \le -c\|g_{k+1}\|^2. \tag{26}$$

Dai-Liao [24] extended the classical conjugacy condition from $y_k^T d_{k+1} = 0$ to

$$y_k^T d_{k+1} = -t\bigl(s_k^T g_{k+1}\bigr), \tag{27}$$

where $t \ge 0$. Thus, we can also show that our proposed method satisfies the above conjugacy condition. □

Lemma 3.3 *Suppose that Assumption* 1 *holds, then the search direction* (12) *satisfies the conjugacy condition* (27).

Proof By (12), we obtain

$$\begin{aligned}
y_k^T d_{k+1} &= -\mu y_k^T g_{k+1} - \frac{s_k^T g_{k+1}}{s_k^T y_k} y_k^T s_k + \mu\frac{y_k^T g_{k+1}}{y_k^T y_k} y_k^T y_k \\
&= -\mu y_k^T g_{k+1} - \frac{s_k^T g_{k+1}}{s_k^T y_k} s_k^T y_k + \mu\frac{y_k^T g_{k+1}}{y_k^T y_k} y_k^T y_k \\
&= -\mu y_k^T g_{k+1} - s_k^T g_{k+1} + \mu y_k^T g_{k+1} \\
&= -s_k^T g_{k+1},
\end{aligned}$$

where the result holds for $t = 1$. The following lemma gives the boundedness of the search direction. □

Lemma 3.4 *Suppose that Assumption* 1 *holds then there exists a constant* $p > 0$ *such that* $\|d_{k+1}\| \leq P\|g_{k+1}\|$, *where* d_{k+1} *is defined by* (12).

Proof A direct result of (10) and the boundedness of $\operatorname{tr}(Q_{k+1})$ gives

$$\begin{aligned} \|d_{k+1}\| &= \|Q_{k+1}g_{k+1}\| \\ &\leq \operatorname{tr}(Q_{k+1})\|g_{k+1}\| \\ &\leq P\|g_{k+1}\|, \end{aligned} \tag{28}$$

where $P = (\frac{\psi_1+n-1}{\psi_1^2})$. □

In order to establish the convergence result, we give the following lemma.

Lemma 3.5 *Suppose that Assumption* 1 *holds. Then there exist some positive constants* γ_1 *and* γ_2 *such that for any steplength* α_k *generated by Step* 3 *of Algorithm* 1 *will satisfy either of the following*:

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \frac{-\gamma_1(g_k^T d_k)^2}{\|d_k\|^2}, \tag{29}$$

or

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \gamma_2 g_k^T d_k. \tag{30}$$

Proof Suppose that (16) is satisfied with $\alpha_k = 1$, then

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta g_k^T d_k, \tag{31}$$

implies that (30) is satisfied with $\gamma_2 = \delta$.

Suppose $\alpha_k < 1$, and that (16) is not satisfied. Then for a steplength $\alpha \leq \frac{\alpha_k}{p_1}$ we have

$$f(x_k + \alpha d_k) - f(x_k) > \delta\alpha g_k^T d_k. \tag{32}$$

Now, by the mean-value theorem there exists a scalar $\tau_k \in (0,1)$ such that

$$f(x_k + \alpha d_k) - f(x_k) = \alpha g(x_k + \tau\alpha d_k)^T d_k. \tag{33}$$

From (32) we have

$$\begin{aligned} (\delta - 1)\alpha g_k^T d_k &< \alpha\big(g(x_k + \tau_k\alpha d_k) - g_k\big)^T d_k \\ &= \alpha y_k^T d_k \\ &< L\big(\alpha\|d_k\|\big)^2, \end{aligned}$$

which implies

$$\alpha \geq -\frac{(1-\delta)(g_k^T d_k)}{L\|d_k\|^2}. \tag{34}$$

Now,

$$\alpha_k \geq p_1 \alpha \geq -\frac{(1-\delta)(g_k^T d_k)}{L\|d_k\|^2}. \tag{35}$$

Substituting (34) in (16) we have the following:

$$\begin{aligned} f(x_k+\alpha_k d_k) - f(x_k) &\leq -\frac{\delta(1-\delta)(g_k^T d_k)}{L\|d_k\|^2}\left(g_k^T d_k\right) \\ &= \frac{-\gamma_1 (g_k^T d_k)^2}{\|d_k\|^2}, \end{aligned}$$

where

$$\gamma_1 = \frac{\delta(1-\delta)}{L}.$$

Therefore

$$f(x_k+\alpha_k d_k) - f(x_k) \leq \frac{-\gamma_1 (g_k^T d_k)^2}{\|d_k\|^2}. \tag{36}$$

□

Theorem 3.6 *Suppose that Assumption* 1 *holds. Then Algorithm* 1 *generates a sequence of approximation* $\{x_k\}$ *such that*

$$\lim_{k\to\infty} \|g_k\| = 0. \tag{37}$$

Proof As a direct consequence of Lemma 3.4, the sufficient descent property (26), and the boundedness of the search direction (28) we have

$$\begin{aligned} f(x_k+\alpha_k d_k) - f(x_k) &\leq \frac{-\gamma_1 (g_k^T d_k)^2}{\|d_k\|^2} \\ &\leq \frac{-\gamma_1 c^2 \|g_k\|^4}{P^2 \|g_k\|^2} \\ &= \frac{-\gamma_1 c^2}{P^2} \|g_k\|^2 \end{aligned} \tag{38}$$

or

$$\begin{aligned} f(x_k+\alpha_k d_k) - f(x_k) &\leq \gamma_1 g_k^T d_k \\ &\leq -\gamma_1 c^2 \|g_k\|^2. \end{aligned} \tag{39}$$

Hence, in either case, there exists a positive constant γ_3 such that

$$f(x_k + \alpha_k d_k) - f(x_k) \leq -\gamma_3 \|g_k\|^2. \tag{40}$$

Since the steplength α_k generated by Algorithm 1 is bounded away from zero, (38) and (39) imply that $f(x_k)$ is a non-increasing sequence. Thus, by the boundedness of $f(x_k)$ we have

$$0 = \lim_{k\to\infty} \big(f(x_{k+1}) - f(x_k)\big) \leq -\gamma_3 \lim_{k\to\infty} \|g_k\|^2,$$

and as a result

$$\lim_{k\to\infty} \|g_k\| = 0. \tag{41}$$

□

4 Numerical results

In this section, we present the results obtained from the numerical experiment of our proposed method in comparison with the CG-DESCENT (CG-DESC) [9], three-term Hestenes-Stiefel (TTHS) [11], three-term Polak-Ribiere-Polyak (TTPRP) [10], and TTCG [12] methods. We evaluate the performance of these methods based on iterations and function evaluations. By considering some standard unconstrained optimization test problems obtained from Andrei [25], we conducted ten numerical experiments for each test function with the size of the variable ranging from $70 \leq n \leq 45{,}000$. The algorithms were implemented using Matlab subroutine programming on a PC (Intel(R) core(TM)2 Duo E8400 3.00 GHz 3 GB) 32-bit Operating system. The program terminates whenever $\|g_k\| < \epsilon$ where $\epsilon = 10^{-6}$ or a method failed to converges within 2,000 iterations. The latter requirement is represented by the symbol '-'. An Armijo-type line search suggested by Byrd and Nocedal [26] was used for all the methods under consideration. Table 1 in the appendices gives the performance of the algorithms in terms of iterations and function evaluations. TTPRP solves 81% of the test problems, TTHS solves 88% of the test problems, CG-DESCENT solves 85% of the test problems, and STCG solves 90% of the test problems, whereas TTCG solves 85% of the test problems. The performance of STCG over TTPRP is that TTPRP needs 16% and 60% more, on average, in terms of the number of iterations and function evaluations, respectively, than STCG. The improvement of STCG over TTHS is that STCG needs 2% and 57% less, on average, in terms of number of iterations and function evaluations, respectively, than TTHS. The improvement of STCG over CG-DESCENT algorithms is that CG-DESCENT needs 10% and 70% more, on average, in terms of the number of iterations and function evaluations, respectively, than STCG. Similarly, the improvement of STCG over TTCG is that STCG needs 21% and 79% less, on average, in terms of the number of iterations and function evaluations, respectively, than TTCG. In order to further examine the performance of these methods, we employ the performance profile of Dolan and Moré [27]. Figures 1-2 give the performance profile plots of these methods in terms of iterations and function evaluations and the top curve corresponds to the method with the highest win which indicates that the performance of the proposed method is highly encouraging and substantially outperforms any of the other methods considered.

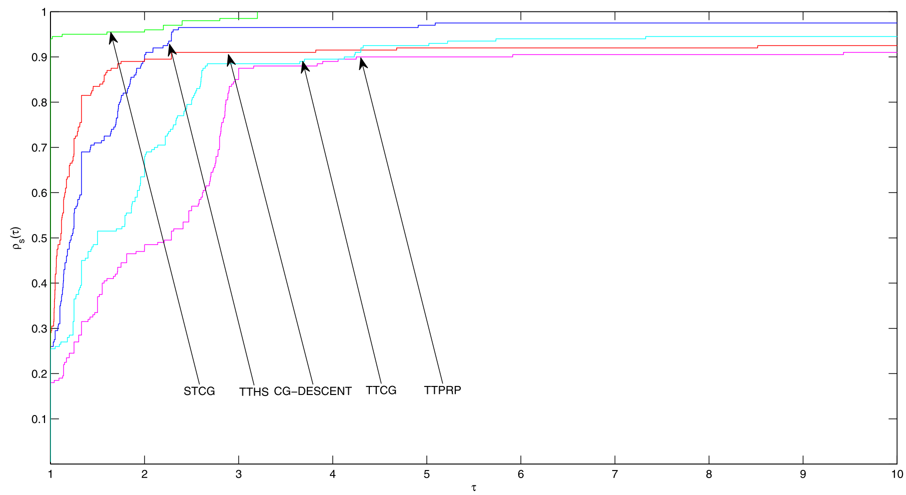

Figure 1 Performance profiles based on iterations.

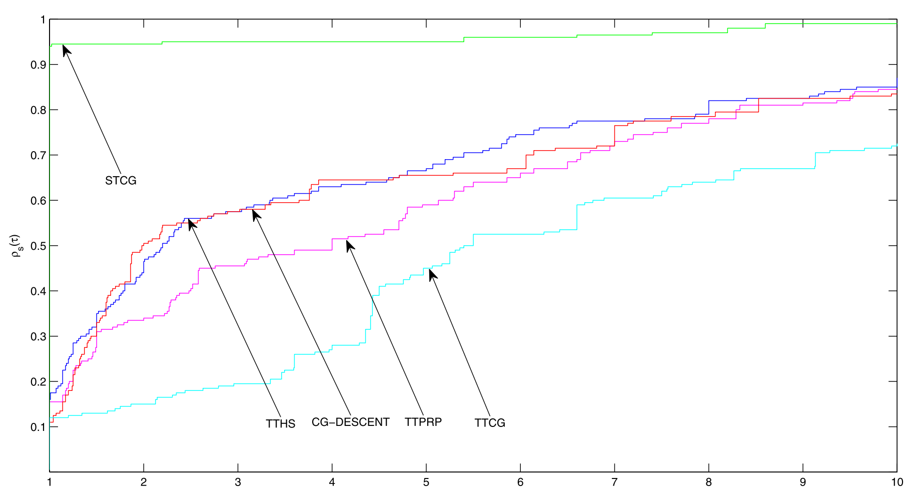

Figure 2 Performance profiles based on function evaluations.

5 Conclusion

We have presented a new three-term conjugate gradient method for solving nonlinear large scale unconstrained optimization problems by considering a modification of the quasi-Newton memoryless DFP update of the inverse Hessian approximation. A remarkable property of the proposed method is that both the sufficient and the conjugacy conditions are satisfied and the global convergence is established under some mild assumption. The numerical results show that the proposed method is promising and more efficient than any of the other methods considered.

Appendix

Table 1 Numerical results of TTPRP, TTHS, CG-DESCENT, STCG, and TTCG

Test functions	Dimension	TTPRP		TTHS		CG-DESC.		STCG		TTCG	
		NI	NF	NI	NF	NI	NF	NI	NF	NI	NF
Extended BD1	70	27	73	39	142	28	102	19	31	25	133
	180	28	77	50	207	28	102	19	31	26	157
	863	31	85	51	194	31	124	20	33	26	157
	1,362	31	85	65	259	31	124	20	33	26	157
	6,500	31	85	37	144	31	124	20	33	28	164
	11,400	31	85	52	216	31	124	20	33	28	164
	17,000	31	85	55	215	31	124	20	33	28	164
	33,200	32	88	59	249	31	124	21	34	28	164
	42,250	32	88	56	205	31	124	-	-	28	164
	45,000	32	88	58	220	31	124	22	37	28	164
Extended Rosenbrock	70	44	227	23	156	55	590	125	156	87	828
	180	52	272	40	264	42	349	119	186	129	1,243
	863	55	285	41	290	33	269	100	136	115	1,098
	1,362	60	315	46	323	28	230	91	142	-	-
	6,500	62	326	22	143	27	220	103	147	-	-
	11,400	74	401	23	159	27	209	116	191	-	-
	17,000	82	436	22	143	30	236	111	141	-	-
	33,200	62	322	39	240	28	213	83	125	-	-
	42,250	67	355	21	155	29	223	133	157	-	-
	45,000	75	393	22	158	22	174	134	157	-	-
Diagonal 7	70	11	22	4	40	4	52	3	4	6	18
	180	11	22	4	40	4	52	3	4	6	18
	863	12	24	4	40	4	52	3	4	6	18
	1,362	12	24	4	40	4	52	3	4	6	18
	6,500	12	24	4	40	4	52	4	5	6	18
	11,400	12	24	4	40	4	52	4	5	6	18
	17,000	34	53	4	40	4	52	4	5	6	18
	33,200	-	-	4	40	4	52	4	5	6	18
	42,250	-	-	4	40	4	52	4	5	6	18
	45,000	-	-	4	40	4	52	4	5	6	18
DENSCHNF	70	25	126	47	403	20	171	6	17	15	126
	180	25	126	49	420	20	171	6	18	16	136
	863	27	136	50	429	21	179	7	18	16	135
	1,362	27	136	52	446	22	188	7	18	16	135
	6,500	28	141	53	455	22	188	19	31	16	135
	11,400	28	141	53	455	22	188	19	31	16	135
	17,000	29	146	53	455	22	188	19	31	16	135
	33,200	29	146	54	463	22	188	19	31	16	135
	42,250	29	146	54	463	22	188	19	31	16	135
	45,000	29	146	55	472	22	188	19	31	16	135
Extended Himmelblau	70	34	135	20	126	19	114	9	15	18	124
	180	36	143	16	85	19	114	9	15	18	124
	863	36	143	15	76	18	121	9	15	18	124
	1,362	37	147	12	75	18	121	9	15	18	124
	6,500	38	151	9	54	19	128	9	15	20	137
	11,400	39	155	11	78	19	128	9	15	20	137
	17,000	39	155	12	76	19	128	9	15	20	137
	33,200	40	159	24	152	19	128	9	15	20	137
	42,250	40	159	16	136	19	128	9	15	20	137
	45,000	40	159	13	69	19	128	9	15	20	137

Table 1 (*Continued*)

Test functions	Dimension	TTPRP		TTHS		CG-DESC.		STCG		TTCG	
		NI	NF	NI	NF	NI	NF	NI	NF	NI	NF
DQDRTIC	70	5	5	5	24	5	36	-	-	43	362
	180	5	5	5	27	5	35	-	-	46	391
	863	5	5	5	31	5	35	-	-	44	371
	1,362	5	5	5	29	5	35	-	-	46	395
	6,500	5	5	5	28	5	35	-	-	111	961
	11,400	5	5	5	24	5	35	-	-	60	516
	17,000	5	5	5	33	5	35	-	-	37	320
	33,200	5	5	5	26	5	35	-	-	59	516
	42,250	5	5	5	29	5	35	-	-	64	548
	45,000	5	5	5	27	5	35	-	-	55	462
HIMMELH	70	-	-	-	-	-	-	7	7	23	80
	180	-	-	-	-	-	-	7	7	18	54
	863	-	-	-	-	-	-	7	7	23	70
	1,362	-	-	-	-	-	-	7	7	22	77
	6,500	-	-	-	-	-	-	7	7	23	83
	11,400	-	-	-	-	-	-	7	7	28	71
	17,000	-	-	-	-	-	-	7	7	22	65
	33,200	-	-	-	-	-	-	7	7	33	90
	42,250	-	-	-	-	-	-	7	7	29	89
	45,000	-	-	-	-	-	-	7	7	26	91
Extended BD2	70	30	96	19	73	9	37	13	23	36	237
	180	31	100	22	87	9	37	13	23	39	254
	863	34	110	11	50	10	43	13	23	41	264
	1,362	34	110	11	44	10	43	13	23	43	276
	6,500	35	114	12	40	10	40	13	23	42	272
	11,400	36	118	25	66	10	43	13	23	43	273
	17,000	36	118	18	70	10	43	13	23	36	231
	33,200	37	122	10	45	10	48	13	23	39	246
	42,250	37	122	17	71	10	38	13	23	37	237
	45,000	37	122	9	39	10	38	13	23	37	236
Extended Maratos	70	25	135	23	60	26	86	37	103	109	934
	180	25	143	24	63	26	86	37	103	107	895
	863	27	143	24	63	26	86	38	104	102	871
	1,362	27	147	33	82	-	-	38	104	112	887
	6,500	27	151	94	209	-	-	38	104	121	1,034
	11,400	120	308	110	246	-	-	38	104	118	949
	17,000	278	681	102	229	-	-	38	104	119	1,004
	33,200	-	-	140	302	-	-	38	104	96	814
	42,250	-	-	140	302	-	-	38	104	105	867
	45,000	-	-	159	350	-	-	38	104	92	787
NONDIA	70	8	52	-	-	12	141	12	37	34	441
	180	11	83	-	-	14	170	18	27	-	-
	863	14	119	-	-	23	337	68	77	-	-
	1,362	11	96	-	-	26	389	32	41	-	-
	6,500	16	155	-	-	1,029	1,555	77	90	-	-
	11,400	16	167	-	-	-	-	161	185	-	-
	17,000	16	168	-	-	-	-	-	-	-	-
	33,200	16	168	-	-	-	-	-	-	-	-
	42,250	21	234	-	-	-	-	-	-	-	-
	45,000	23	247	-	-	-	-	-	-	-	-

Table 1 (***Continued***)

Test functions	Dimension	TTPRP		TTHS		CG-DESC.		STCG		TTCG	
		NI	NF	NI	NF	NI	NF	NI	NF	NI	NF
DENSCHNB	70	25	50	30	82	6	13	5	6	13	29
	180	25	50	27	76	6	13	5	6	14	31
	863	27	54	28	67	6	13	6	7	14	31
	1,362	27	54	30	73	6	13	6	7	14	31
	6,500	28	56	23	55	6	13	6	7	14	31
	11,400	28	56	34	86	6	13	6	7	14	31
	17,000	29	58	31	76	6	13	6	7	14	31
	33,200	29	58	31	83	6	13	6	7	14	31
	42,250	29	58	38	93	6	13	6	7	14	31
	45,000	29	58	29	79	6	13	6	7	14	31
EG2	70	31	102	12	37	35	180	19	59	-	-
	180	87	321	18	27	31	172	68	96	-	-
	863	-	-	68	77	-	-	25	100	-	-
	1,362	-	-	32	41	-	-	-	-	-	-
	6,500	-	-	77	90	-	-	25	107	-	-
	11,400	-	-	-	-	-	-	-	-	-	-
	17,000	-	-	-	-	-	-	89	361	-	-
	33,200	-	-	-	-	-	-	-	-	-	-
	42,250	-	-	-	-	-	-	-	-	-	-
	45,000	-	-	92	158	-	-	33	138	-	-
Raydan 2	70	5	5	5	5	5	5	4	4	5	41
	180	5	5	5	5	5	5	4	4	5	41
	863	5	5	5	5	5	5	4	4	5	41
	1,362	5	5	5	5	5	5	4	4	5	41
	6,500	6	6	6	6	6	6	4	4	5	41
	11,400	6	6	6	6	6	6	4	4	5	41
	17,000	6	6	6	6	6	6	4	4	5	41
	33,200	6	6	6	6	6	6	4	4	5	41
	42,250	6	6	6	6	6	6	4	4	5	41
	45,000	6	6	6	6	6	6	4	4	5	41
ENGVAL1	70	49	137	30	139	29	181	53	60	54	375
	180	48	154	28	137	30	157	52	61	50	413
	863	67	273	29	145	31	135	49	88	55	402
	1,362	81	325	33	159	30	134	49	98	57	525
	6,500	283	1,486	32	139	23	156	50	100	39	263
	11,400	100	461	29	173	28	192	51	97	40	271
	17,000	-	-	23	132	27	175	52	131	43	302
	33,200	-	-	30	164	29	244	52	94	54	373
	42,250	-	-	25	160	29	244	52	91	44	318
	45,000	-	-	27	119	29	244	51	104	39	223
HIMMELBG	70	5	5	5	5	5	5	4	4	4	4
	180	5	5	5	5	5	5	5	5	5	5
	863	6	6	6	6	6	6	5	5	5	5
	1,362	6	6	6	6	6	6	5	5	6	6
	6,500	7	7	7	7	7	7	6	6	6	6
	11,400	7	7	7	7	7	7	6	6	6	6
	17,000	8	8	8	8	8	8	6	6	6	6
	33,200	8	8	8	8	8	8	7	7	7	7
	42,250	8	8	8	8	8	8	7	7	7	7
	45,000	8	8	8	8	8	8	7	7	7	7

Table 1 (***Continued***)

Test functions	Dimension	TTPRP		TTHS		CG-DESC.		STCG		TTCG	
		NI	NF	NI	NF	NI	NF	NI	NF	NI	NF
Diagonal 5	70	4	4	4	4	4	4	4	4	3	21
	180	4	4	4	4	4	4	4	4	3	21
	863	4	4	4	4	4	4	4	4	3	21
	1,362	4	4	4	4	4	4	4	4	3	21
	6,500	4	4	4	4	4	4	4	4	3	21
	11,400	4	4	4	4	4	4	4	4	4	22
	17,000	4	4	4	4	4	4	4	4	4	22
	33,200	4	4	4	4	4	4	4	4	4	22
	42,250	4	4	4	4	4	4	4	4	4	22
	45,000	4	4	4	4	4	4	4	4	4	22
Extended Tridigonal 1	70	343	465	19	24	19	24	22	40	17	51
	180	-	-	21	24	20	24	22	40	20	51
	863	-	-	21	25	21	26	28	46	20	61
	1,362	-	-	22	29	21	26	28	46	20	61
	6,500	-	-	23	30	23	28	31	55	21	97
	11,400	-	-	23	30	23	28	31	56	21	97
	17,000	-	-	20	34	23	28	32	50	21	97
	33,200	-	-	23	44	24	29	31	49	21	97
	42,250	-	-	22	27	24	29	42	63	21	97
	45,000	-	-	22	38	24	29	46	64	21	97
Extended Quadratic Penalty QP1	70	8	25	10	53	7	33	7	15	15	99
	180	9	36	10	53	6	21	9	18	18	124
	863	12	44	11	58	6	25	12	24	21	154
	1,362	8	32	8	48	8	49	13	25	16	135
	6,500	13	48	10	51	12	121	14	32	69	796
	11,400	11	43	15	107	30	328	15	32	188	976
	17,000	7	26	11	52	58	702	15	30	381	1,616
	33,200	12	55	13	85	231	2,500	16	43	-	-
	42,250	13	52	10	61	381	2,950	16	43	-	-
	45,000	8	39	10	61	433	3,584	15	33	-	-
Diagonal 8	70	9	18	4	9	3	7	3	5	4	33
	180	9	18	4	9	3	7	3	5	4	33
	863	10	20	4	10	4	9	3	5	4	33
	1,362	10	20	4	12	4	10	3	5	4	33
	6,500	10	20	4	10	4	8	3	5	4	33
	11,400	10	20	4	10	4	8	3	5	4	33
	17,000	10	20	4	12	4	8	3	5	4	33
	33,200	13	24	4	11	4	8	3	5	4	33
	42,250	23	36	4	10	4	8	3	5	4	33
	45,000	23	36	4	10	4	8	3	5	4	33
Extended Tridigonal 2	70	44	155	13	25	19	24	18	23	17	68
	180	45	157	20	22	20	25	18	23	20	84
	863	42	146	21	43	21	26	17	21	20	84
	1,362	42	146	20	27	21	26	17	21	20	84
	6,500	42	146	22	26	23	28	17	22	21	97
	11,400	42	146	20	27	23	28	17	22	21	97
	17,000	42	146	23	27	23	28	17	22	21	97
	33,200	41	143	23	25	24	29	17	22	21	97
	42,250	41	143	24	51	24	29	17	22	21	97
	45,000	41	143	25	50	24	29	17	22	21	97

Competing interests

We hereby declare that there are no competing interests with regard to the manuscript.

Authors' contributions

We all participated in the establishment of the basic concepts, the convergence properties of the proposed method and in the experimental result in comparison of the proposed method with order existing methods.

Received: 30 May 2016 Accepted: 9 November 2016 Published online: 13 December 2016

References

1. Fletcher, R, Reeves, CM: Function minimization by conjugate gradients. Comput. J. **7**(2), 149-154 (1964)
2. Polak, E, Ribiere, G: Note sur la convergence de méthodes de directions conjuguées. ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique **3**(R1), 35-43 (1969)
3. Liu, Y, Storey, C: Efficient generalized conjugate gradient algorithms, part 1: theory. J. Optim. Theory Appl. **69**(1), 129-137 (1991)
4. Hestenes, MR: The conjugate gradient method for solving linear systems. In: Proc. Symp. Appl. Math VI, American Mathematical Society, pp. 83-102 (1956)
5. Dai, Y-H, Yuan, Y: A nonlinear conjugate gradient method with a strong global convergence property. SIAM J. Optim. **10**(1), 177-182 (1999)
6. Fletcher, R: Practical Methods of Optimization. John Wiley & Sons, New York (2013)
7. Al-Baali, M: Descent property and global convergence of the Fletcher-Reeves method with inexact line search. IMA J. Numer. Anal. **5**(1), 121-124 (1985)
8. Gilbert, JC, Nocedal, J: Global convergence properties of conjugate gradient methods for optimization. SIAM J. Optim. **2**(1), 21-42 (1992)
9. Hager, WW, Zhang, H: A new conjugate gradient method with guaranteed descent and an efficient line search. SIAM J. Optim. **16**(1), 170-192 (2005)
10. Zhang, L, Zhou, W, Li, D-H: A descent modified Polak-Ribière-Polyak conjugate gradient method and its global convergence. IMA J. Numer. Anal. **26**(4), 629-640 (2006)
11. Zhang, L, Zhou, W, Li, D: Some descent three-term conjugate gradient methods and their global convergence. Optim. Methods Softw. **22**(4), 697-711 (2007)
12. Andrei, N: On three-term conjugate gradient algorithms for unconstrained optimization. Appl. Math. Comput. **219**(11), 6316-6327 (2013)
13. Broyden, C: Quasi-Newton methods and their application to function minimisation. Mathematics of Computation **21**, 368-381 (1967)
14. Davidon, WC: Variable metric method for minimization. SIAM J. Optim. **1**(1), 1-17 (1991)
15. Fletcher, R, Powell, MJ: A rapidly convergent descent method for minimization. Comput. J. **6**(2), 163-168 (1963)
16. Goldfarb, D: Extension of Davidon's variable metric method to maximization under linear inequality and equality constraints. SIAM J. Appl. Math. **17**(4), 739-764 (1969)
17. Shanno, DF: Conjugate gradient methods with inexact searches. Math. Oper. Res. **3**(3), 244-256 (1978)
18. Perry, JM: A class of conjugate gradient algorithms with a two step variable metric memory. Center for Mathematical Studies in Economies and Management Science. Northwestern University Press, Evanston (1977)
19. Wolkowicz, H: Measures for symmetric rank-one updates. Math. Oper. Res. **19**(4), 815-830 (1994)
20. Nocedal, J: Conjugate gradient methods and nonlinear optimization. In: Linear and Nonlinear Conjugate Gradient-Related Methods, pp. 9-23 (1996)
21. Andrei, N: Acceleration of conjugate gradient algorithms for unconstrained optimization. Appl. Math. Comput. **213**(2), 361-369 (2009)
22. Leong, WJ, San Goh, B: Convergence and stability of line search methods for unconstrained optimization. Acta Appl. Math. **127**(1), 155-167 (2013)
23. Babaie-Kafaki, S: A modified scaled memoryless BFGS preconditioned conjugate gradient method for unconstrained optimization. 4OR **11**(4), 361-374 (2013)
24. Dai, Y-H, Liao, L-Z: New conjugacy conditions and related nonlinear conjugate gradient methods. Appl. Math. Optim. **43**(1), 87-101 (2001)
25. Andrei, N: An unconstrained optimization test functions collection. Adv. Model. Optim. **10**(1), 147-161 (2008)
26. Byrd, RH, Nocedal, J: A tool for the analysis of quasi-Newton methods with application to unconstrained minimization. SIAM J. Numer. Anal. **26**(3), 727-739 (1989)
27. Dolan, ED, More, JJ: Benchmarking optimization software with performance profiles. Math. Program. **91**(2), 201-213 (2002)