A modified scaled conjugate gradient method with global convergence for nonconvex functions

Saman Babaie-Kafaki*   Reza Ghanbari

Abstract

Following Andrei's approach, a modified scaled memoryless BFGS preconditioned conjugate gradient method is proposed based on the modified secant equation suggested by Li and Fukushima. It is shown that the method is globally convergent without convexity assumption on the objective function. Furthermore, for uniformly convex objective functions, sufficient descent property of the method is established based on an eigenvalue analysis. Numerical experiments are employed to demonstrate the efficiency of the method.

1 Introduction

Conjugate gradient (CG) methods have attracted special attention for solving large-scale unconstrained optimization problems in the form of $\min_{x \in \mathbb{R}^n} f(x)$, with the smooth nonlinear function $f : \mathbb{R}^n \to \mathbb{R}$, because of low memory requirement, simple computation and strong global convergence. Iterations of the CG methods are given by

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

where α_k is a steplength to be computed by a line search procedure and d_k is the search direction defined by

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

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

Different CG methods mainly correspond to different choices for the CG parameter [21]. Furthermore, in convergence analysis and implementation of the CG methods Wolfe line search conditions [32], i.e.,

$$
\begin{aligned}
f(x_k + \alpha_k d_k) - f(x_k) &\leq \delta \alpha_k \nabla f(x_k)^T d_k, &\quad (1.2) \\
\nabla f(x_k + \alpha_k d_k)^T d_k &\geq \sigma \nabla f(x_k)^T d_k, &\quad (1.3)
\end{aligned}
$$

with $0 < \delta < \sigma < 1$, have been widely paid attention to.

To enhance efficiency of the CG methods using information of the Hessian of the objective function more straightly, Andrei [1, 2, 3, 4, 5] proposed the family of scaled CG algorithms of SCALCG, hybridizing the memoryless BFGS preconditioned CG method suggested by Shanno [29] and the spectral CG method suggested by Birgin and Martínez [15], based on the standard secant equation [32], i.e.,

$$
\nabla^2 f(x_{k+1}) s_k = y_k, \quad (1.4)
$$

where $y_k = g_{k+1} - g_k$. In order to employ the objective function information in addition to the gradient information, recently Babie-Kafaki [9] made a modification on the SCALCG, using a revised form of the modified secant equation suggested by Zhang et al. [33] which was proposed in [13], that is,

$$
\nabla^2 f(x_{k+1}) s_k = z_k, \ z_k = y_k + \rho_k \frac{\max\{\vartheta_k, 0\}}{s_k^T u_k} u_k, \quad (1.5)
$$

where ρ_k is a nonnegative parameter, $u_k \in \mathbb{R}^n$ is a vector parameter satisfying $s_k^T u_k \neq 0$, and $\vartheta_k = 6(f_k - f_{k+1}) + 3(g_k + g_{k+1})^T s_k$.

Search directions of the scaled CG methods suggested in [1, 2, 3, 4, 5, 9] are in the following form:

$$
d_0 = -g_0, \ d_{k+1} = -Q_{k+1} g_{k+1}, \ k = 0, 1, ..., \quad (1.6)
$$

with the matrix $Q_{k+1} \in \mathbb{R}^{n \times n}$ defined by

$$
Q_{k+1} = \theta_{k+1} I - \theta_{k+1} \frac{v_k s_k^T + s_k v_k^T}{s_k^T v_k} + \left(1 + \theta_{k+1} \frac{v_k^T v_k}{s_k^T v_k}\right) \frac{s_k s_k^T}{s_k^T v_k}, \quad (1.7)
$$

where, depending to the secant equations (1.4) and (1.5), $v_k = y_k$ in [1, 2, 3, 4, 5] and $v_k = z_k$ in [9], and also, θ_{k+1} is a scaling parameter determined based on a two-point approximation of the corresponding secant equation [14, 27], i.e.,

$$
\theta_{k+1} = \frac{s_k^T s_k}{s_k^T v_k}. \quad (1.8)
$$

Note that the matrix Q_{k+1} is not formed; only its product with g_{k+1} is applied instead. Also, Q_{k+1} is precisely the BFGS update in which approximation of the inverse Hessian is restarted as $\theta_{k+1} I$. Since from (1.3) we have

$$
s_k^T z_k \geq s_k^T y_k = s_k^T g_{k+1} - s_k^T g_k \geq -(1 - \sigma) s_k^T g_k > 0, \quad (1.9)
$$

the matrix Q_{k+1} is positive definite [26, 32] and consequently, the search directions (1.6) are descent directions. In addition, for uniformly convex objective functions [32], Babaie-Kafaki [7, 10, 9] showed that the search directions (1.6) with $v_k = y_k$ and $v_k = z_k$ satisfy the sufficient descent condition, that is,

$$g_k^T d_k \leq -\tau ||g_k||^2, \ k = 0, 1, ..., \tag{1.10}$$

where τ is a positive constant and $||.||$ stands for the Euclidean norm.

Here, we suggest a modified scaled CG method based on the modified secant equation proposed by Li and Fukushima [23]. The remainder of this work is organized as follows. In Section 2, we propose our method and discuss its global convergence and sufficient descent property. In Section 3, we present comparative numerical results. Finally, we make conclusions in Section 4.

2 A modified scaled conjugate gradient method

In this section, at first we briefly discuss the modified secant equation proposed by Li and Fukushima [23], being necessary in explanation of our modified scaled CG method.

Convexity assumption on the objective function plays an important role in convergence analysis of the quasi-Newton methods (see [6] and the references therein). Nevertheless, Li and Fukushima [23] proposed a modified BFGS method which is globally and superlinearly convergent for nonconvex objective functions (see also [19, 24, 34]). The method has been designed based on the following modified secant equation:

$$\nabla^2 f(x_{k+1})s_k = w_k, \ w_k = y_k + h_k||g_k||^r s_k, \tag{2.1}$$

where $r > 0$, and $h_k > 0$ 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. As an interesting property, for the modified secant equation (2.1) we have $s_k^T w_k > 0$, independent of the line search and the objective function convexity, which guarantees heredity of positive definiteness for the related BFGS updates. Note that from (1.9), Wolfe conditions ensure that $h_k = C, \forall k \geq 0$. In such situation, if $||g_k|| < 1, \forall k \geq K$, with some positive integer K large enough, and $r \geq 1$, then we have $h_k||g_k||^r = O(||g_k||)$, which leads to nice theoretical properties as given by Theorem 4.1 of [23].

Recently, efforts have been made to suggest unconstrained optimization algorithms based on the modified secant equation (2.1). Examples include the nonlinear CG method proposed by Zhou and Zhang [34], the hybrid CG method proposed by Babaie-Kafaki et al. [12], the descent CG method proposed by Narushima and Yabe [25], the three-term CG method proposed by Sugiki et al. [31], and the two-point stepsize gradient algorithm proposed by Babaie-Kafaki and Fatemi [11].

Effectiveness of the modified secant equation (2.1) motivated us to apply it in order to suggest another modified version of SCALCG. More precisely, here we deal with a scaled CG method in which the search directions are computed by (1.6)-(1.8) with $v_k = w_k$ defined in (2.1). Thus, letting $d_0 = -g_0$, search directions of our method are given by

$$
\begin{aligned}
d_{k+1} \;=\; & -\theta_{k+1}g_{k+1} + \theta_{k+1}\frac{s_k^T g_{k+1}}{s_k^T w_k}w_k + \theta_{k+1}\frac{w_k^T g_{k+1}}{s_k^T w_k}s_k \\[1mm]
& - \left(1 + \theta_{k+1}\frac{w_k^T w_k}{s_k^T w_k}\right)\frac{s_k^T g_{k+1}}{s_k^T w_k}s_k, \; k \geq 1,
\end{aligned}
\tag{2.2}
$$

in which

$$
\theta_{k+1} = \frac{s_k^T s_k}{s_k^T w_k}.
\tag{2.3}
$$

As mentioned in Section 1, since $s_k^T w_k > 0$, the matrix Q_{k+1} defined by (1.7) with $v_k = w_k$ is positive definite and consequently, the search directions (2.2) are descent directions. Next, we show that our method is globally convergent without convexity assumption on the objective function. In this context, we need to make the following standard assumptions on the objective function.

Assumption A1: The level set $\mathcal{L} = \{x \mid f(x) \leq f(x_0)\}$, 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

$$
||\nabla f(x) - \nabla f(y)|| \leq L||x - y||, \; \forall x, y \in \mathcal{N}.
\tag{2.4}
$$

Note that these assumptions imply that there exists a positive constant γ such that

$$
||\nabla f(x)|| \leq \gamma, \; \forall x \in \mathcal{L} \text{ (see Proposition 3.1 of [13]).}
\tag{2.5}
$$

The following important lemma plays an essential role in proving the global convergence theorem of our method.

Lemma 2.1. [31] Suppose that Assumptions A1 and A2 hold. Consider any iterative method in the form of (1.1), where d_k and α_k satisfy the sufficient descent condition (1.10) and the Wolfe conditions (1.2) and (1.3), respectively. If

$$
\sum_{k \geq 0}\frac{1}{||d_k||^2} = \infty,
\tag{2.6}
$$

then the method converges in the sense that

$$
\liminf_{k \to \infty}||g_k|| = 0.
\tag{2.7}
$$

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

Theorem 2.1. *Suppose that Assumptions A1 and A2 hold. Consider any iterative method in the form of (1.1) in which the search direction d_k is computed by (2.2) with θ_{k+1} defined by (2.3), and the steplength α_k is determined to fulfill the Wolfe conditions (1.2) and (1.3). If the sufficient descent condition (1.10) is satisfied, then the method converges in the sense that (2.7) holds.*

Proof. At first, note that search directions of the method are descent directions. So, from (1.2) the sequence $\{x_k\}_{k \geq 0}$ is a subset of the level set $\mathcal{L}$. Now, to prove the theorem by contradiction, we suppose that there exists a positive constant ε such that

$$||g_k|| \geq \varepsilon, \ \forall k \geq 0. \tag{2.8}$$

Hence, since (1.9) ensures that from the Wolfe conditions we have $s_k^T y_k > 0$, from (2.8) we get

$$s_k^T w_k = s_k^T y_k + C||g_k||^r ||s_k||^2 \geq C\varepsilon^r ||s_k||^2, \tag{2.9}$$

which leads to

$$\theta_{k+1} \leq \frac{1}{C\varepsilon^r}. \tag{2.10}$$

Also, from (2.4) and (2.5) we can write

$$||w_k|| \leq ||y_k|| + C\gamma^r ||s_k|| \leq (L + C\gamma^r)||s_k||. \tag{2.11}$$

Now, from Cauchy-Schwarz inequality, (2.5), (2.9), (2.10) and (2.11) we have

$$
\begin{aligned}
||d_{k+1}|| \ &\leq \ \theta_{k+1}||g_{k+1}|| + 2\theta_{k+1}\frac{||s_k||\,||w_k||\,||g_{k+1}||}{s_k^T w_k} \\
&\quad + \left(1 + \theta_{k+1}\frac{||w_k||^2}{s_k^T w_k}\right)\frac{||s_k||\,||g_{k+1}||}{s_k^T w_k}||s_k|| \\
&\leq \ \frac{1}{C\varepsilon^r}\gamma + 2\frac{1}{C\varepsilon^r}\frac{(L + C\gamma^r)\gamma||s_k||^2}{C\varepsilon^r||s_k||^2} \\
&\quad + \left(1 + \frac{1}{C\varepsilon^r}\frac{(L + C\gamma^r)^2||s_k||^2}{C\varepsilon^r||s_k||^2}\right)\frac{\gamma||s_k||}{C\varepsilon^r||s_k||^2}||s_k|| \\
&= \ \frac{\gamma}{C\varepsilon^r} + 2\frac{(L + C\gamma^r)\gamma}{C^2\varepsilon^{2r}} + \left(1 + \frac{(L + C\gamma^r)^2}{C^2\varepsilon^{2r}}\right)\frac{\gamma}{C\varepsilon^r},
\end{aligned}
$$

which leads to (2.6). So, from Lemma 2.1, (2.7) holds, contradicting (2.8). ∎

Although search directions of our method are descent directions, the sufficient descent condition in the assumptions of Theorem 2.1 may seem to be to some extent strong. In what follows, based on an eigenvalue analysis we show that search directions of our method possess the sufficient descent property for uniformly convex objective functions.

Lemma 2.2. For the search directions (1.6) with the matrix Q_{k+1} defined by (1.7), if for all $k \geq 0$,

$$||v_k|| \leq \Lambda ||s_k||, \tag{2.12}$$

and

$$s_k^T v_k \geq \xi ||s_k||^2, \tag{2.13}$$

with some positive constants Λ and ξ, then the sufficient descent condition (1.10) holds.

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

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

which leads to

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

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

Since the trace of a square matrix is equal to the sum of its eigenvalues, we have

$$
\begin{aligned}
\text{tr}(Q_{k+1}) &= (n - 2)\theta_{k+1} + \left(1 + \theta_{k+1} \frac{||v_k||^2}{s_k^T v_k}\right)\theta_{k+1} \\
&= \underbrace{\theta_{k+1} + ... + \theta_{k+1}}_{(n-2) \text{ times}} + \lambda_k^- + \lambda_k^+,
\end{aligned}
$$

which leads to

$$\lambda_k^- + \lambda_k^+ = \left(1 + \theta_{k+1}\frac{||v_k||^2}{s_k^T v_k}\right)\theta_{k+1}. \tag{2.14}$$

Furthermore, from the properties of the Frobenius norm we have

$$
\begin{aligned}
||Q_{k+1}||_F^2 &= \text{tr}(Q_{k+1}^T Q_{k+1}) = \text{tr}(Q_{k+1}^2) \\
&= (n - 2)\theta_{k+1}^2 - 2\theta_{k+1}^2 + \left(1 + \theta_{k+1}\frac{||v_k||^2}{s_k^T v_k}\right)^2 \theta_{k+1}^2 \\
&= \underbrace{\theta_{k+1}^2 + ... + \theta_{k+1}^2}_{(n-2) \text{ times}} + \lambda_k^{-2} + \lambda_k^{+2},
\end{aligned}
$$

which yields

$$\lambda_k^{-2} + \lambda_k^{+2} = -\theta_{k+1}^2 + 2\theta_{k+1}^3\frac{||v_k||^2}{s_k^T v_k} + \theta_{k+1}^4\frac{||v_k||^4}{(s_k^T v_k)^2}. \tag{2.15}$$

Now, from (2.14) and (2.15), after some algebraic manipulations we get

$$\lambda_k^- \lambda_k^+ = \theta_{k+1}^2, \tag{2.16}$$

and consequently, from (2.3), (2.14) and (2.16), λ_k^- and λ_k^+ can be computed by

$$\lambda_k^{\pm} = \frac{1}{2}\theta_{k+1}\left(1 + \frac{||s_k||^2||v_k||^2}{(s_k^Tv_k)^2} \pm \sqrt{\left(1 + \frac{||s_k||^2||v_k||^2}{(s_k^Tv_k)^2}\right)^2 - 4}\right). \qquad (2.17)$$

From Cauchy-Schwarz inequality we get

$$\lambda_k^+ \geq \frac{1}{2}\theta_{k+1}\left(1 + \frac{||s_k||^2||v_k||^2}{(s_k^Tv_k)^2}\right) \geq \theta_{k+1},$$

and

$$\lambda_k^- \leq \frac{1}{2}\theta_{k+1}\left(1 + \frac{||s_k||^2||v_k||^2}{(s_k^Tv_k)^2} - \right.$$
$$\left.\sqrt{\left(1 + \frac{||s_k||^2||v_k||^2}{(s_k^Tv_k)^2}\right)^2 - 4\frac{||s_k||^2||v_k||^2}{(s_k^Tv_k)^2}}\right)$$
$$= \frac{1}{2}\theta_{k+1}\left(1 + \frac{||s_k||^2||v_k||^2}{(s_k^Tv_k)^2} - \left(\frac{||s_k||^2||v_k||^2}{(s_k^Tv_k)^2} - 1\right)\right) = \theta_{k+1}.$$

Thus, λ_k^- is the smallest eigenvalue of the positive definite matrix Q_{k+1}. Now, from Cauchy-Schwarz inequality, (2.12) and (2.13) we have

$$\lambda_k^- = \frac{1}{2}\theta_{k+1}\frac{4}{1 + \frac{||s_k||^2||v_k||^2}{(s_k^Tv_k)^2} + \sqrt{\left(1 + \frac{||s_k||^2||v_k||^2}{(s_k^Tv_k)^2}\right)^2 - 4}}$$

$$\geq \frac{1}{2}\theta_{k+1}\frac{4}{2\left(1 + \frac{||s_k||^2||v_k||^2}{(s_k^Tv_k)^2}\right)} = \frac{||s_k||^2(s_k^Tv_k)}{(s_k^Tv_k)^2 + ||s_k||^2||v_k||^2}$$

$$\geq \frac{||s_k||^2(s_k^Tv_k)}{2||s_k||^2||v_k||^2} \geq \frac{\xi}{2\Lambda^2}. \qquad (2.18)$$

Therefore, from (2.18) we can write

$$d_{k+1}^Tg_{k+1} = -g_{k+1}^TQ_{k+1}g_{k+1} \leq -\frac{\xi}{2\Lambda^2}||g_{k+1}||^2, \; \forall k \geq 0. \qquad (2.19)$$

So, from (2.19) and since $d_0^Tg_0 = -||g_0||^2$, to complete the proof it is enough to let $\tau = \min\{1, \frac{\xi}{2\Lambda^2}\}$ in the sufficient descent condition (1.10). $\blacksquare$

Note that in our method we have $v_k = w_k$. So, if Assumptions A1 and A2 hold and the objective function is uniformly convex on $\mathcal{N}$, then there exists a positive constant μ such that

$$s_k^Ty_k \geq \mu||s_k||^2,$$

which leads to

$$s_k^T w_k = s_k^T y_k + C||g_k||^r||s_k||^2 \geq s_k^T y_k \geq \mu||s_k||^2. \tag{2.20}$$

Thus, from (2.11) and (2.20), Lemma 2.2 guarantees the sufficient descent property for an iterative method in the form of (1.1) with the search directions (2.2).

3 Numerical experiments

Here, we present some numerical results obtained by applying a MATLAB 7.7.0. 471 (R2008b) implementation of the iterative method (1.1) in which the search directions are computed by (1.6)-(1.8) with the following three choices for the vector parameter v_k:

- The method M1: $v_k = w_k$ defined by (2.1) which yields our modified scaled CG method;

- The method M2: $v_k = z_k$ defined by (1.5) which yields the modified scaled CG method suggested in [9];

- The SCALCG method: $v_k = y_k$;

comparing with the HS+ method, i.e., a CG method with the parameter $\beta_k = \max\left\{\dfrac{g_{k+1}^T y_k}{d_k^T y_k}, 0\right\}$ obtained by nonnegative restriction of the CG parameter suggested by Hestenes and Stiefel [22], firstly proposed in [28] and then studied in [17]. The implementations were performed on a computer, Intel(R) Core (TM)2 Duo CPU 2.00 GHz, with 1 GB of RAM. Since CG methods have been mainly designed for solving large-scale problems, our experiments have been done on a set of 105 unconstrained optimization test problems of the CUTEr collection [18] with various dimensions $n \in [100, 15625]$, as specified in [8].

For M1, we set $C = 10^{-6}$ in (2.1) because of the promising numerical results obtained among the different values $C \in \{10^{-k}\}_{k=0}^6$. Also, we adopt the suggestion of [34] for the computation of r in the sense that we set $r = 3$, if $||g_k|| < 1$, and $r = 1$, otherwise. For M2, we adopt the suggestions of [9] and set $u_k = s_k$ in (1.5), for all $k \geq 0$, and also, $\rho_k = 1$, if $||s_k|| < 1$, and $\rho_k = 0$, otherwise.

In our implementations, if

$$g_k^T d_k > -10^{-10}||g_k||||d_k||, \tag{3.1}$$

then we considered d_k as an improper search direction. Although (3.1) seldom occurred, when encountering we restarted the algorithm with $d_k = -g_k$. We used the Wolfe line search conditions (1.2) and (1.3) with $\delta = 0.0001$ and $\sigma = 0.9$, and computed the steplength α_k using the interpolation scheme proposed in [26] with the initial trial value being equal to $||s_{k-1}||/||d_k||$, for all $k \geq 1$, and $||g_k||_\infty^{-1}$, for $k = 0$ [3, 30]. Also, all attempts to solve the test problems were limited to reaching a maximum of 10000 iterations or achieving a solution with $||g_k||_\infty < 10^{-6}(1 + |f(x_k)|)$.

Efficiency comparisons were made using the performance profile introduced by Dolan and Moré [16], on the running time and the total number of function and gradient evaluations being equal to $N_f + 3N_g$ [20], 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 ω of the best. Figures 1 and 2 show the comparisons results. As shown by Figure 1, although M2 outperforms M1, and SCALCG is often preferable to M1, at times M1 outperforms SCALCG with respect to the total number of function and gradient evaluations. Also, Figure 2 shows that although the methods M1, M2 and SCALCG are approximately competitive, M1 slightly outperforms SCALCG, and M2 slightly outperforms M1 with respect to the running time. Moreover, the figures show that all the three methods M1, M2 and SCALCG outperform the HS+ method.
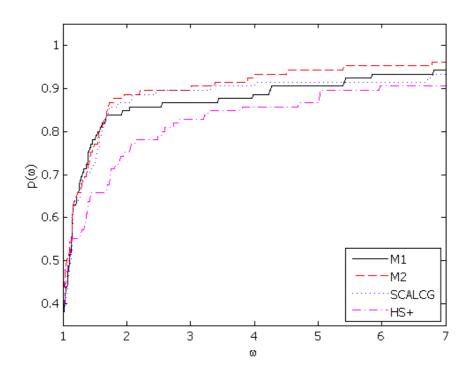

Figure 1: Total number of function and gradient evaluations performance profiles for M1, M2, SCALCG and HS+

4 Conclusions and future work

A modified scaled conjugate gradient method is proposed following Andrei's approach of hybridizing the memoryless BFGS preconditioned conjugate gradient method suggested by Shanno [29] and the spectral conjugate gradient method suggested by Birgin and Martínez [15], based on the modified secant equation suggested by Li and Fukushima [23]. When the line search fulfills the Wolfe conditions, it has been shown that the method is globally convergent without convexity assumption on the objective function. Also, based on an eigenvalue analysis, it has been shown that search directions of the method satisfy the sufficient
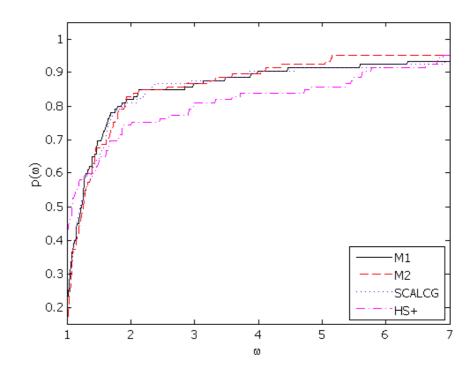
Figure 2: CPU time performance profiles for M1, M2, SCALCG and HS+

descent property for uniformly convex objective functions. The method (M1) has been numerically compared with the modified scaled conjugate gradient method suggested in [9] (M2), and the methods of SCALCG and HS+. Numerical results showed that the method is practically promising.

As a future work, because of the relationship between the scaled conjugate gradient methods and the BFGS method, it would be interesting to study the superlinear convergence of the scaled conjugate gradient methods.

Acknowledgements

References

[1] N. Andrei. A scaled BFGS preconditioned conjugate gradient algorithm for unconstrained optimization. *Appl. Math. Lett.*, 20(6):645–650, 2007.

[2] N. Andrei. Scaled conjugate gradient algorithms for unconstrained optimization. *Comput. Optim. Appl.*, 38(3):401–416, 2007.

[3] N. Andrei. Scaled memoryless BFGS preconditioned conjugate gradient algorithm for unconstrained optimization. *Optim. Methods Softw.*, 22(4):561–571, 2007.

[4] N. Andrei. A scaled nonlinear conjugate gradient algorithm for unconstrained optimization. *Optimization*, 57(4):549–570, 2008.

[5] N. Andrei. Accelerated scaled memoryless BFGS preconditioned conjugate gradient algorithm for unconstrained optimization. *European J. Oper. Res.*, 204(3):410–420, 2010.

[6] S. Babaie-Kafaki. A modified BFGS algorithm based on a hybrid secant equation. *Sci. China Math.*, 54(9):2019–2036, 2011.

[7] S. Babaie-Kafaki. A note on the global convergence theorem of the scaled conjugate gradient algorithms proposed by Andrei. *Comput. Optim. Appl.*, 52(2):409–414, 2012.

[8] S. Babaie-Kafaki. A quadratic hybridization of Polak-Ribière-Polyak and Fletcher-Reeves conjugate gradient methods. *J. Optim. Theory Appl.*, 154(3):916–932, 2012.

[9] S. Babaie-Kafaki. A modified scaled memoryless BFGS preconditioned conjugate gradient method for unconstrained optimization. *4OR*, 11(4):361-374, 2013.

[10] S. Babaie-Kafaki. A new proof for the sufficient descent condition of Andrei's scaled conjugate gradient algorithms. *Pac. J. Optim.*, 9(1):23–28, 2013.

[11] S. Babaie-Kafaki and M. Fatemi. A modified two-point stepsize gradient algorithm for unconstrained minimization. *Optim. Methods Softw.*, 28(5): 1040-1050, 2013.

[12] S. Babaie-Kafaki, M. Fatemi, and N. Mahdavi-Amiri. Two effective hybrid conjugate gradient algorithms based on modified BFGS updates. *Numer. Algorithms*, 58(3):315–331, 2011.

[13] S. Babaie-Kafaki, R. Ghanbari, and N. Mahdavi-Amiri. Two new conjugate gradient methods based on modified secant equations. *J. Comput. Appl. Math.*, 234(5):1374–1386, 2010.

[14] J. Barzilai and J.M. Borwein. Two-point stepsize gradient methods. *IMA J. Numer. Anal.*, 8(1):141–148, 1988.

[15] E. Birgin and J.M. Martínez. A spectral conjugate gradient method for unconstrained optimization. *Appl. Math. Optim.*, 43(2):117–128, 2001.

[16] E.D. Dolan and J.J. Moré. Benchmarking optimization software with performance profiles. *Math. Program.*, 91(2, Ser. A):201–213, 2002.

[17] J.C. Gilbert and J. Nocedal. Global convergence properties of conjugate gradient methods for optimization. *SIAM J. Optim.*, 2(1):21–42, 1992.

[18] N.I.M. Gould, D. Orban, and Ph.L. Toint. CUTEr: a constrained and unconstrained testing environment, revisited. *ACM Trans. Math. Softw.*, 29(4):373–394, 2003.

[19] 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(1-2):435–446, 2008.

[20] W.W. Hager and H. Zhang. Algorithm 851: CG_Descent, a conjugate gradient method with guaranteed descent. *ACM Trans. Math. Softw.*, 32(1):113–137, 2006.

[21] W.W. Hager and H. Zhang. A survey of nonlinear conjugate gradient methods. *Pac. J. Optim.*, 2(1):35–58, 2006.

[22] M.R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *J. Research Nat. Bur. Standards*, 49(6):409–436, 1952.

[23] D.H. Li and M. Fukushima. A modified BFGS method and its global convergence in nonconvex minimization. *J. Comput. Appl. Math.*, 129(1-2):15–35, 2001.

[24] D.H. Li and M. Fukushima. On the global convergence of the BFGS method for nonconvex unconstrained optimization problems. *SIAM J. Optim.*, 11(4):1054–1064, 2001.

[25] Y. Narushima and H. Yabe. Conjugate gradient methods based on secant conditions that generate descent search directions for unconstrained optimization. *J. Comput. Appl. Math.*, 236(17):4303–4317, 2012.

[26] J. Nocedal and S.J. Wright. *Numerical Optimization*. Springer, New York, 2006.

[27] S.S. Oren and E. Spedicato. Optimal conditioning of self-scaling variable metric algorithms. *Math. Program.*, 10(1):70–90, 1976.

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

[29] D.F. Shanno. Conjugate gradient methods with inexact searches. *Math. Oper. Res.*, 3(3):244–256, 1978.

[30] D.F. Shanno and K.H. Phua. Algorithm 500: Minimization of unconstrained multivariate functions. *ACM Trans. Math. Softw.*, 2(1):87–94, 1976.

[31] 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(3):733–757, 2012.

[32] W. Sun and Y.X. Yuan. *Optimization Theory and Methods: Nonlinear Programming*. Springer, New York, 2006.

[33] 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(1):147–167, 1999.

[34] W. Zhou and L. Zhang. A nonlinear conjugate gradient method based on the MBFGS secant condition. *Optim. Methods Softw.*, 21(5):707–714, 2006.

Department of Mathematics, Faculty of Mathematics,
Statistics and Computer Sciences, Semnan University,
P.O. Box: 35195-363, Semnan, Iran,
E-mail address: `sbk@semnan.ac.ir`

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