Filomat 30:14 (2016), 3715–3723
DOI 10.2298/FIL1614715B

Published by Faculty of Sciences and Mathematics, University of Niš, Serbia
Available at: http://www.pmf.ni.ac.rs/filomat

# An Adaptive Hager-Zhang Conjugate Gradient Method

**Saman Babaie-Kafaki[a], Reza Ghanbari[b]**

[a]*Department of Mathematics, Faculty of Mathematics, Statistics and Computer Science, Semnan University, P.O. Box: 35195-363, Semnan, Iran*
[b]*Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, P.O. Box: 9177948953, Mashhad, Iran*

**Abstract.** Based on a singular value study, lower and upper bounds for the condition number of the matrix which generates search directions of the Hager-Zhang conjugate gradient method are obtained. Then, based on the insight gained by our analysis, a modified version of the Hager-Zhang method is proposed, using an adaptive switch form the Hager-Zhang method to the Hestenes-Stiefel method when the mentioned condition number is large. A brief global convergence analysis is made for the uniformly convex objective functions. Numerical comparisons between the implementations of the proposed method and the Hager-Zhang method are made on a set of unconstrained optimization test problems of the CUTEr collection, using the performance profile introduced by Dolan and Moré. Comparative testing results are reported.

## 1. Introduction

Conjugate gradient (CG) methods comprise a class of unconstrained optimization algorithms characterized by low memory requirements and strong global convergence properties [4, 12] which made them popular for engineers and mathematicians engaged in solving large-scale problems in the following form:

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

where $f : \mathbb{R}^n \to \mathbb{R}$ is a smooth nonlinear function and its gradient is available. The iterative formula of a CG method is given by

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

where $\alpha_k$ is a steplength to be computed by a line search procedure [20], and $d_k$ is the search direction defined by

$$\begin{aligned} &d_0 = -g_0, \\ &d_{k+1} = -g_{k+1} + \beta_k d_k,\ k = 0, 1, ..., \end{aligned} \tag{2}$$

2010 *Mathematics Subject Classification*. Primary 65K05; Secondary 90C30, 65F35, 15A18
*Keywords*. Unconstrained optimization, Large-scale optimization, Singular value, Condition number, Global convergence
Received: 14 October 2014; Accepted: 15 June 2015
Communicated by Predrag Stanimirović
*Email addresses:* sbk@semnan.ac.ir (Saman Babaie-Kafaki), rghanbari@um.ac.ir (Reza Ghanbari)

where $g_k = \nabla f(x_k)$, and $\beta_k$ is a scalar called the CG (update) parameter. The steplength $\alpha_k$ in the CG methods is often determined to satisfy some versions of Wolfe conditions [22]. The (standard) Wolfe conditions are

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k \nabla f(x_k)^T d_k, \tag{3}$$

$$\nabla f(x_k + \alpha_k d_k)^T d_k \geq \sigma \nabla f(x_k)^T d_k, \tag{4}$$

where $0 < \delta < \sigma < 1$. Also, the strong Wolfe conditions consist of (3) and the following strengthened version of (4):

$$|\nabla f(x_k + \alpha_k d_k)^T d_k| \leq -\sigma \nabla f(x_k)^T d_k. \tag{5}$$

Different choices for the CG parameter lead to different CG methods (see [12] and the references therein). One of the recent essential CG methods has been proposed by Hager and Zhang [10], with the following CG parameter:

$$\beta_k^{HZ} = \underbrace{\frac{y_k^T g_{k+1}}{d_k^T y_k}}_{\beta_k^{HS}} - 2\frac{\|y_k\|^2}{d_k^T y_k}\frac{d_k^T g_{k+1}}{d_k^T y_k} = \beta_k^{HS} - 2\frac{\|y_k\|^2}{d_k^T y_k}\frac{d_k^T g_{k+1}}{d_k^T y_k}, \tag{6}$$

where $y_k = g_{k+1} - g_k$, and $\beta_k^{HS}$ is the CG parameter proposed by Hestenes and Stiefel [13]. Also, $\|.\|$ stands for the Euclidean norm. The CG parameter $\beta_k^{HZ}$ is well-defined if the line search ensures that $d_k^T y_k \neq 0$, as guaranteed by the (strong) Wolfe conditions.

As an important property, the HZ method satisfies the sufficient descent condition in the sense that

$$d_k^T g_k \leq -\frac{7}{8}\|g_k\|^2, \; k = 0, 1, ...,$$

independent of the line search and the objective function convexity [1], and also, the method is globally convergent when the line search fulfills the Wolfe or the Goldstien conditions [20]. In the perspective of the numerical performance, the HZ method with an approximate Wolfe conditions outperforms the two efficient unconstrained optimization algorithms of the limited memory BFGS (L-BFGS) method proposed by Liu and Nocedal [14], and the PRP+ method, i.e. a CG method with the nonnegative restriction of the CG parameter proposed by Polak, Ribière and Polyak [17, 18], which was firstly proposed in [19] and then studied in [8]. Moreover, the HZ method can be considered as an adaptive version of the nonlinear CG method proposed by Dai and Liao [6] as well as a member of the two-parameter family of descent CG methods proposed by Babaie-Kafaki and Ghanbari [2].

It is remarkable that, from (2) and (6), search directions of the HZ method can be written as:

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

where

$$Q_{k+1} = I - \frac{s_k y_k^T}{s_k^T y_k} + 2\frac{\|y_k\|^2}{s_k^T y_k}\frac{s_k s_k^T}{s_k^T y_k}. \tag{7}$$

So, the HZ method can be considered as a quasi-Newton method [20] in which the inverse Hessian in each iteration is approximated by the nonsymmetric matrix $Q_{k+1}$. Since $Q_{k+1}$ is determined based on a rank-two update, its determinant can be computed by

$$\det(Q_{k+1}) = 2\frac{\|s_k\|^2\|y_k\|^2}{(s_k^T y_k)^2}, \tag{8}$$

(see equality (1.2.70) of [20]) and so, from Cauchy-Schwarz inequality we have

$$\det(Q_{k+1}) \geq 2,$$

which ensures that $Q_{k+1}$ is nonsingular.

Here, based on a singular value study on the matrix $Q_{k+1}$, we make a modification on the HZ method. The remainder of this work is organized as follows. In Section 2, we propose our adaptive nonlinear CG method and discuss its global convergence. In Section 3, we numerically compare our method with the HZ method and report comparative testing results. Finally, we make conclusions in Section 4.

## 2. An Adaptive Nonlinear Conjugate Gradient Method

In this section, after computing the singular values of the matrix $Q_{k+1}$ defined by (7), we study its condition number and consequently, we propose an adaptive CG method as a modified version of the HZ method.

One essential factor which plays an important role in the sensitivity analysis of a numerical problem related to a matrix is the matrix condition number. For an arbitrary nonsingular matrix $A$, the scalar $\kappa(A)$ defined by

$$\kappa(A) = ||A||||A^{-1}||,$$

is called the condition number of $A$. The matrix $A$ with a large condition number is called an ill-conditioned matrix since instability may occur in the computations related to this matrix. Here, the following theorem is needed.

**Theorem 2.1.** [21] *Let $A \in \mathbb{R}^{n\times n}$ be a nonsingular matrix with the singular values $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_n > 0$. Then*

$$\kappa(A) = \frac{\sigma_1}{\sigma_n}. \tag{9}$$

To compute the singular values of $Q_{k+1}$, firstly note that since $s_k^T y_k \neq 0$, there exists a set of mutually orthonormal vectors $\{u_k^i\}_{i=1}^{n-2}$ such that

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

which leads to

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

That is, $Q_{k+1}$ has $n-2$ singular values being equal to 1. Now, we find the two remaining singular values of $Q_{k+1}$ namely $\sigma_k^-$ and $\sigma_k^+$.

Since $||Q_{k+1}||_F^2$ is equal to $\mathrm{tr}(Q_{k+1}^T Q_{k+1})$, and also, is equal to sum of squares of the singular values of $Q_{k+1}$, we can write

$$\begin{aligned} \mathrm{tr}(Q_{k+1}^T Q_{k+1}) &= n - 2 + \frac{||s_k||^2||y_k||^2}{(s_k^T y_k)^2} + 4\frac{||s_k||^4||y_k||^4}{(s_k^T y_k)^4} \\ &= \underbrace{1^2 + ... + 1^2}_{(n-2) \text{ times}} + {\sigma_k^-}^2 + {\sigma_k^+}^2. \end{aligned}$$

Hence,

$${\sigma_k^-}^2 + {\sigma_k^+}^2 = \frac{||s_k||^2||y_k||^2}{(s_k^T y_k)^2} + 4\frac{||s_k||^4||y_k||^4}{(s_k^T y_k)^4}. \tag{10}$$

Also, from (8) and since $|\det(A)| = \prod_{i=1}^{n} \sigma_i$, we have

$$\underbrace{1 \times ... \times 1}_{(n-2) \text{ times}} \times \sigma_k^- \times \sigma_k^+ = 2\frac{||s_k||^2||y_k||^2}{(s_k^T y_k)^2}.$$

Hence,

$$\sigma_k^- \sigma_k^+ = 2\frac{||s_k||^2||y_k||^2}{(s_k^T y_k)^2}. \tag{11}$$

Now, from (10) and (11), after some algebraic manipulations we obtain the singular values $\sigma_k^-$ and $\sigma_k^+$ as follows:

$$\sigma_k^{\pm} = \frac{1}{2}\frac{||s_k||||y_k||}{|s_k^T y_k|}\left\{\sqrt{4\frac{||s_k||^2||y_k||^2}{(s_k^T y_k)^2}+5} \pm \sqrt{4\frac{||s_k||^2||y_k||^2}{(s_k^T y_k)^2}-3}\right\}. \tag{12}$$

In what follows, we obtain lower and upper bounds for the singular values $\sigma_k^-$ and $\sigma_k^+$. In this context, note that the singular value $\sigma_k^-$ can be written as:

$$\begin{aligned}\sigma_k^- &= \frac{1}{2}\frac{||s_k||||y_k||}{|s_k^T y_k|}\frac{8}{\sqrt{4\frac{||s_k||^2||y_k||^2}{(s_k^T y_k)^2}+5} + \sqrt{4\frac{||s_k||^2||y_k||^2}{(s_k^T y_k)^2}-3}} \\ &= \frac{4||s_k||||y_k||}{\sqrt{4||s_k||^2||y_k||^2+5(s_k^T y_k)^2} + \sqrt{4||s_k||^2||y_k||^2-3(s_k^T y_k)^2}}.\end{aligned}$$

Therefore, from Cauchy-Schwarz inequality we can write

$$\sigma_k^- \leq \frac{4||s_k||||y_k||}{\sqrt{4||s_k||^2||y_k||^2} + \sqrt{||s_k||^2||y_k||^2}} = \frac{4}{3}. \tag{13}$$

Also, from (10), (11), and Cauchy-Schwarz inequality we have

$$\begin{aligned}\sigma_k^- = 2\frac{||s_k||^2||y_k||^2}{(s_k^T y_k)^2}\frac{1}{\sigma_k^+} &\geq 2\frac{||s_k||^2||y_k||^2}{(s_k^T y_k)^2}\frac{1}{\sqrt{\sigma_k^{-^2}+\sigma_k^{+^2}}} \\ &= \frac{2||s_k||^2||y_k||^2}{\sqrt{||s_k||^2||y_k||^2(s_k^T y_k)^2 + 4||s_k||^4||y_k||^4}} \\ &\geq \frac{2}{\sqrt{5}} \approx 0.8944.\end{aligned} \tag{14}$$

Hence, from (13) and (14) we have

$$\frac{2}{\sqrt{5}} \leq \sigma_k^- \leq \frac{4}{3}. \tag{15}$$

For the singular value $\sigma_k^+$, although for general functions from Cauchy-Schwarz inequality we have the following lower bound:

$$\sigma_k^+ \geq 2, \tag{16}$$

here it is possible to obtain an upper bound when the function is uniformly convex. So, the following theorem is now necessary.

**Theorem 2.2.** [20] *Let $\mathcal{S} \subset \mathbb{R}^n$ be a nonempty open convex set and consider the differentiable function $f : \mathcal{S} \to \mathbb{R}$. The function $f$ is uniformly convex on $\mathcal{S}$ if and only if its gradient is uniformly monotone on $\mathcal{S}$, i.e. there exists a positive constant $\mu$ such that*

$$(\nabla f(x) - \nabla f(y))^T (x-y) \geq \mu||x-y||^2, \ \forall x, y \in \mathcal{S}. \tag{17}$$

In our analysis, we need to make the following standard assumptions on the objective function.

**Assumptions 2.1.** *The level set $\mathcal{L} = \{x | f(x) \leq f(x_0)\}$, with $x_0$ to be the starting point of the iterative method (1)-(2), is bounded. Also, in a 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*

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \ \forall x, y \in \mathcal{N}. \tag{18}$$

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

$\|\nabla f(x)\| \leq \gamma, \ \forall x \in \mathcal{L}$ (see Proposition 3.1 of [3]).

Since search directions of the HZ method are descent directions, inequality (3) ensures that the sequence $\{x_k\}_{k\geq 0}$ generated by the HZ method is a subset of the level set $\mathcal{L}$. Hence, if $f$ is a uniformly convex function on the neighborhood $\mathcal{N}$ of $\mathcal{L}$, then from (17) we have

$$s_k^T y_k \geq \mu \|s_k\|^2. \tag{19}$$

Also, from (18) we have

$$\|y_k\| \leq L\|s_k\|. \tag{20}$$

Now, from (12), (19) and (20), we get

$$\begin{aligned}
\sigma_k^+ &= \frac{1}{2}\frac{\|s_k\|\|y_k\|}{|s_k^T y_k|}\left\{\sqrt{4\frac{\|s_k\|^2\|y_k\|^2}{(s_k^T y_k)^2} + 5} + \sqrt{4\frac{\|s_k\|^2\|y_k\|^2}{(s_k^T y_k)^2} - 3}\right\} \\
&\leq \frac{1}{2}\frac{L}{\mu}\left\{\sqrt{4\frac{L^2}{\mu^2} + 5} + \sqrt{4\frac{L^2}{\mu^2} - 3}\right\} \\
&\leq \frac{5}{2}\frac{L^2}{\mu^2}.
\end{aligned} \tag{21}$$

So, from (16) and (21) we have

$$2 \leq \sigma_k^+ \leq \frac{5}{2}\frac{L^2}{\mu^2}. \tag{22}$$

As mentioned in Section 1, the HZ method can be considered as a quasi-Newton method in which the nonsingular matrix $Q_{k+1}$ is used to approximate the inverse Hessian in each iteration. Numerical performance of the quasi-Newton methods is much influenced by condition number of the successive approximations of the (inverse) Hessian [20]. As an important result, from (9), (15) and (22), we conclude that the condition number of $Q_{k+1}$ is bounded in the sense that

$$\frac{3}{2} \leq \kappa(Q_{k+1}) \leq \frac{5\sqrt{5}}{4}\frac{L^2}{\mu^2} \approx 2.7951\frac{L^2}{\mu^2}, \ \forall k \geq 0. \tag{23}$$

Here, by the following approximations for the uniformly convex function $f$:

$$\begin{aligned}
L &\approx \max_{x\in\mathcal{L}} \|\nabla^2 f(x)\|, \\
\mu &\approx \min_{x\in\mathcal{L}} \|\nabla^2 f(x)\|,
\end{aligned}$$

which are obtained based on the mean-value theorem, i.e.,

$$y_k = \nabla^2 f(\xi_k x_k + (1 - \xi_k)x_{k+1})s_k,$$

for some $\xi_k \in (0, 1)$, it can be stated that if $\mu \approx L$, or equivalently, the Hessian $\nabla^2 f$ is not ill-conditioned on the level set $\mathcal{L}$, then from (23) the matrix $Q_{k+1}$ is a well-conditioned matrix and as a result, the HZ method have a nice numerical behavior.

Now, based on the insight gained by the above singular value analysis, we make a modification on the HZ method and propose a nonlinear CG method in which the CG parameter is adaptively switched from $\beta_k^{HZ}$ to $\beta_k^{HS}$ when encountering to a point $x_k$ with an ill-conditioned Hessian $\nabla^2 f(x_k)$.

Since inequality (15) shows that $\sigma_k^-$ is bounded below, $\kappa(Q_{k+1})$ is large when $\sigma_k^+$ is large. Also, note that from (21) it can be seen that a large value of $\frac{||s_k||^2||y_k||^2}{(s_k^T y_k)^2}$, arisen from the third term of $Q_{k+1}$ in (7), makes the value of $\sigma_k^+$ to be a large number, or equivalently, makes $Q_{k+1}$ to be an ill-conditioned matrix. Motivated by this, we suggest a modified form of $Q_{k+1}$ as follows:

$$Q_{k+1}^M = I - \frac{s_k y_k^T}{s_k^T y_k} + 2t_k \frac{||y_k||^2}{s_k^T y_k} \frac{s_k s_k^T}{s_k^T y_k},$$

where $t_k \in [0, 1]$ is a scalar embedded to control the values of $\kappa(Q_{k+1}^M)$, for all $k \geq 0$. (Recently, similar study on the Perry's CG method [16] has been made by Liu et al. [15].) Here, $t_k$ is computed by

$$t_k = \begin{cases} 1, & \frac{||s_k||^2||y_k||^2}{(s_k^T y_k)^2} < \tau, \\ 0, & \text{otherwise}, \end{cases} \tag{24}$$

with a positive constant $\tau$. In other words, we propose an adaptive version of the HZ method, namely AHZ, with the following CG parameter:

$$\beta_k^{AHZ} = \beta_k^{HS} - 2t_k \frac{||y_k||^2}{d_k^T y_k} \frac{d_k^T g_{k+1}}{d_k^T y_k},$$

in which $t_k$ is computed by (24). Note that if $t_k = 1$, then $\beta_k^{AHZ} = \beta_k^{HZ}$, and if $t_k = 0$, then $\beta_k^{AHZ} = \beta_k^{HS}$. Hence, $\beta_k^{AHZ}$ can be considered as an extension of $\beta_k^{HZ}$ and $\beta_k^{HS}$. Also, since the AHZ method can be regarded as an adaptive version of the Dai-Liao method [6], if the search directions are descent directions and the steplengthes are determined to satisfy the strong Wolfe conditions (3) and (5), then Theorem 3.3 of [6] ensures global convergence of the method for uniformly convex objective functions.

**Remark 2.1.** *Recently, Dai and Kou [5] proposed an efficient family of CG methods with the following parameter:*

$$\beta_k(\tau_k) = \beta_k^{HS} - \left(\tau_k + \frac{||y_k||^2}{s_k^T y_k} - \frac{s_k^T y_k}{||s_k||^2}\right) \frac{g_{k+1}^T s_k}{d_k^T y_k},$$

*where $\tau_k$ is a parameter corresponding to scaling factor in the scaled memoryless BFGS method [20] for which an effective choice has been given as follows:*

$$\tau_k = \frac{s_k^T y_k}{||s_k||^2}. \tag{25}$$

*Considering the similarity between the CG parameter $\beta_k(\tau_k)$ with the choice (25) for $\tau_k$ and the CG parameter $\beta_k^{HZ}$, a similar singular value analysis can be made for the Dai-Kou method and consequently, an adaptive version of the Dai-Kou can be proposed.*

## 3. Numerical experiments

Here, we present some numerical results obtained by applying C++ implementations of the CG methods of AHZ and HZ. 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. Furthermore, the experiments were performed on a set of 145 unconstrained optimization test problems of the CUTEr collection [9], with default dimensions as given in Hager's home page: 'http://www.math.ufl.edu/~hager/'.

In our experiments, we used the approximate Wolfe conditions proposed by Hager and Zhang [10, 11] in the line search procedure, with the same parameter values as suggested in [11]. For the AHZ method, among the different values $\tau \in \{5k\}_{k=1}^{30}$ in (24), we set $\tau = 70$ because of its promising numerical results. Although the descent condition may not always hold for the AHZ method, uphill search direction seldom occurred in our experiments; when encountering, we restarted the algorithm with $d_k = -g_k$ [6]. 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 Moré [7], on the CPU time and the total number of function and gradient evaluations [11], here denoted by $N_T$ and defined by $N_T = N_f + 3N_g$, where $N_f$ and $N_g$ respectively denote the number of function and gradient evaluations. Performance profile gives, for every $\omega \geq 1$, the proportion $p(\omega)$ of the test problems that each considered algorithmic variant has a performance within a factor of $\omega$ of the best. Comparisons results have been demonstrated by Figures 1 and 2. As the figures show, the AHZ method outperforms the HZ method both in the perspectives of the total number of function and gradient evaluations, and the CPU time. Thus, our modified version of the HZ method turns out to be practically efficient.

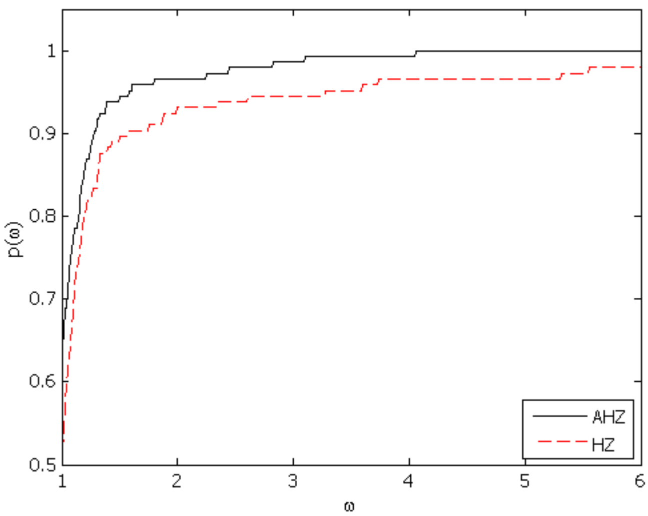


Figure 1: Total number of function and gradient evaluations performance profiles for AHZ and HZ

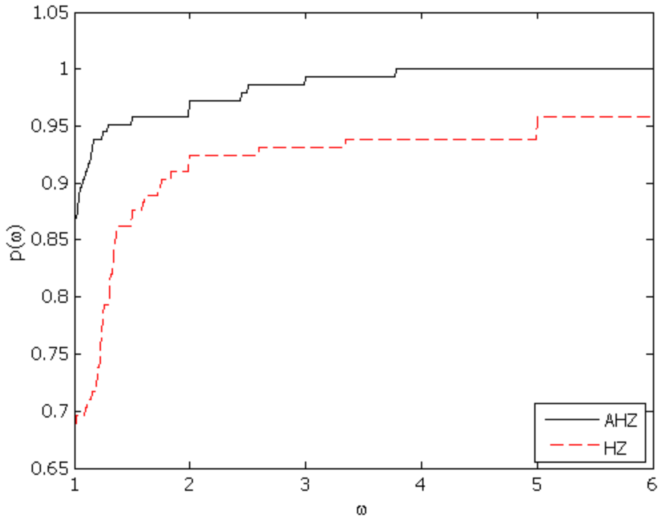


Figure 2: CPU time performance profiles for AHZ and HZ

## 4. Conclusions

In order to control the condition number of the matrix which generates search directions of the nonlinear conjugate gradient method proposed by Hager and Zhang [10], we suggested a modified version of the Hager-Zhang method based on an adaptive switch form the Hager-Zhang method to the Hestenes-Stiefel method [13] when the mentioned condition number is large. Under proper conditions, we briefly showed that our method is globally convergent for uniformly convex objective functions. Using a set of unconstrained optimization test problems of the CUTEr collection [9], a numerical comparison of the implementations of our adaptive method (AHZ) and the Hager-Zhang method (HZ) has been made. The results showed that our adaptive approach seems to be practically effective.

## Acknowledgements

This research was in part supported by Research Council of Semnan University, and in part by Research Council of Ferdowsi University of Mashhad. The authors are grateful to Professor William W. Hager for providing the line search code. They also thank the anonymous referee for his/her valuable comments helped to improve the quality of this work.

## References

[1] S. Babaie-Kafaki, On the sufficient descent condition of the Hager-Zhang conjugate gradient methods, 4OR 11 (2013) 361–374.
[2] S. Babaie-Kafaki, R. Ghanbari, A descent family of Dai-Liao conjugate gradient methods, Optimization Methods and Software 29 (2014) 583–591.
[3] S. Babaie-Kafaki, R. Ghanbari, N. Mahdavi-Amiri, Two new conjugate gradient methods based on modified secant equations, Journal of Computational and Applied Mathematics 234 (2010) 1374–1386.
[4] Y.H. Dai, J.Y. Han, G.H. Liu, D.F. Sun, H.X. Yin, Y.X. Yuan, Convergence properties of nonlinear conjugate gradient methods, SIAM Journal on Optimization 10 (1999) 348–358.
[5] 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.
[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] E.D. Dolan, J.J. Moré, Benchmarking optimization software with performance profiles, Mathematical Programming 91 (2002) 201–213.

[8] J.C. Gilbert, J. Nocedal, Global convergence properties of conjugate gradient methods for optimization, SIAM Journal on Optimization 2 (1992) 21–42.
[9] N.I.M. Gould, D. Orban, Ph.L. Toint, CUTEr: a constrained and unconstrained testing environment, revisited, ACM Transactions on Mathematical Software 29 (2003) 373–394.
[10] 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.
[11] W.W. Hager, H. Zhang, Algorithm 851: CG_Descent, a conjugate gradient method with guaranteed descent, ACM Transactions on Mathematical Software 32 (2006) 113–137.
[12] W.W. Hager, H. Zhang, A survey of nonlinear conjugate gradient methods, Pacific Journal of Optimization 2 (2006) 35–58.
[13] M.R. Hestenes, E. Stiefel, Methods of conjugate gradients for solving linear systems, Journal of Research of the National Bureau of Standards 49 (1952) 409–436.
[14] D.C. Liu, J. Nocedal, On the limited memory BFGS method for large-scale optimization, Mathematical Programming 45 (1989) 503–528.
[15] D.Y. Liu, G.Q. Xu, Applying Powell's symmetrical technique to conjugate gradient methods, Computational Optimization and Applications 49 (2011) 319–334.
[16] A. Perry, A modified conjugate gradient algorithm, Operations Research 26 (1976) 1073–1078.
[17] E. Polak, G. Ribière, Note sur la convergence de méthodes de directions conjuguées, Revue Française d'Informatique et de Recherche Opérationelle 3 (1969) 35–43.
[18] B.T. Polyak, The conjugate gradient method in extreme problems, USSR Computational Mathematics and Mathematical Physics 9 (1969) 94–112.
[19] M.J.D. Powell, Restart procedures for the conjugate gradient method, Mathematical Programming 12 (1977) 241–254.
[20] W. Sun, Y.X. Yuan, Optimization Theory and Methods: Nonlinear Programming, Springer, New York, 2006.
[21] D.S. Watkins, Fundamentals of Matrix Computations, John Wiley and Sons, New York, 2002.
[22] P. Wolfe, Convergence conditions for ascent methods, SIAM Review 11 (1969) 226–235.