ISSN: 1017-060X (Print)

ISSN: 1735-8515 (Online)

Bulletin of the Iranian Mathematical Society

Vol. 43 (2017), No. 7, pp. 2437–2448

Title:

Extensions of the Hestenes–Stiefel and Polak–Ribière–Polyak conjugate gradient methods with sufficient descent property

Author(s):

S. Babaie–Kafaki and R. Ghanbari

Published by the Iranian Mathematical Society
http://bims.ims.ir

Bull. Iranian Math. Soc.
Vol. 43 (2017), No. 7, pp. 2437–2448
Online ISSN: 1735-8515

EXTENSIONS OF THE HESTENES–STIEFEL AND POLAK–RIBIÈRE–POLYAK CONJUGATE GRADIENT METHODS WITH SUFFICIENT DESCENT PROPERTY

S. BABAIE–KAFAKI* AND R. GHANBARI

(Communicated by Maziar Salahi)

Abstract. Using search directions of a recent class of three–term conjugate gradient methods, modified versions of the Hestenes–Stiefel and Polak–Ribière–Polyak methods are proposed which satisfy the sufficient descent condition. The methods are shown to be globally convergent when the line search fulfills the (strong) Wolfe conditions. Numerical experiments are done on a set of CUTEr unconstrained optimization test problems. They demonstrate efficiency of the proposed methods in the sense of the Dolan–Moré performance profile.
Keywords: Unconstrained optimization, conjugate gradient method, sufficient descent property, line search, global convergence.
MSC(2010): Primary: 90C53; Secondary: 49M37, 65K05.

1. **Introduction**

Consider the unconstrained optimization problem

$$\min_{x\in\mathbb{R}^n} f(x), \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable and analytic expression of its gradient is available. As practical tools for solving (1.1), iterative methods define a sequence of approximations that are expected to be closer and closer to the exact solution in a given norm, stopping the iterations using some predefined criterion, and obtaining a vector which is only an approximation of the solution. When the dimension n is large, iterative methods which require low memory storage are more encouraging. Among them there are the conjugate gradient (CG) methods with the following iterative formula:

$$x_0 \in \mathbb{R}^n,\ x_{k+1} = x_k + s_k,\ s_k = \alpha_k d_k,\ k = 0, 1, \ldots, \tag{1.2}$$

Article electronically published on December 30, 2017.
Received: 23 August 2016, Accepted: 15 May 2017.
*Corresponding author.

where α_k is a step length to be computed by a line search technique along the search direction d_k defined by

$$d_0 = -g_0,\ d_{k+1} = -g_{k+1} + \beta_k d_k,\ k = 0, 1, \ldots, \tag{1.3}$$

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

Different CG methods correspond to different choices for β_k with dissimilar computational behaviors [1, 13, 23]. Two well–known and numerically effective CG parameters have been proposed by Hestenes and Stiefel [24] (HS), and Polak, Ribière and Polyak [26, 27] (PRP) with

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k}, \text{ and } \beta_k^{PRP} = \frac{g_{k+1}^T y_k}{||g_k||^2},$$

where $y_k = g_{k+1} - g_k$, and $||.||$ stands for the ℓ_2 norm. However, in theoretical point of view, the methods fail to guarantee the descent property which is defined as follows:

$$d_k^T g_k < 0,\ \forall k \geq 0. \tag{1.4}$$

To overcome this defect, efforts have been made to suggest extensions of the HS and PRP methods with sufficient descent property, i.e.,

$$d_k^T g_k \leq -c||g_k||^2,\ \forall k \geq 0, \tag{1.5}$$

where c is a positive constant, being stronger than the descent condition (1.4). A brief review of the literature reveals an abundance of such attempts. As examples, interested readers can study the references [2, 5, 10, 12, 14, 17, 18, 21, 29, 32–38].

In a general scheme, recently Narushima et al. [25] proposed a class of three–term CG methods with the search directions $d_0 = -g_0$ and

$$d_{k+1} = -g_{k+1} + \beta_k (g_{k+1}^T p_{k+1})^{\dagger} \left((g_{k+1}^T p_{k+1}) d_k - (g_{k+1}^T d_k) p_{k+1} \right), \tag{1.6}$$

for all $k \geq 0$, where β_k is an arbitrary CG parameter, $p_{k+1} \in \mathbb{R}^n$ is any vector, and

$$a^{\dagger} = \begin{cases} \dfrac{1}{a}, & a \neq 0, \\ 0, & a = 0. \end{cases}$$

It is worth noting that search directions of the method satisfy the sufficient descent condition $g_k^T d_k = -||g_k||^2$, for all $k \geq 0$, independent to the choices of β_k and p_{k+1}, the objective function convexity and the line search. Also, if $g_{k+1}^T p_{k+1} \neq 0$, then (1.6) reduces to

$$d_{k+1} = -g_{k+1} + \beta_k d_k - \beta_k \frac{g_{k+1}^T d_k}{g_{k+1}^T p_{k+1}} p_{k+1}, \tag{1.7}$$

or equivalently,

$$d_{k+1} = -g_{k+1} + \beta_k Q_{k+1} d_k,$$

with

$$Q_{k+1} = I - \frac{p_{k+1}g_{k+1}^T}{g_{k+1}^T p_{k+1}}.$$

As seen, the matrix Q_{k+1} is a projection matrix onto the orthogonal complement of Span$\{g_{k+1}\}$ along Span$\{p_{k+1}\}$. If $p_{k+1} = g_{k+1}$, then Q_{k+1} is an orthogonal projection matrix and, in such situation, (1.7) can be written as follows:

$$d_{k+1} = -g_{k+1} + \beta_k d_k - \beta_k \frac{g_{k+1}^T d_k}{||g_{k+1}||^2} g_{k+1}, \tag{1.8}$$

playing an important role in our approach of modifying the HS and PRP methods. This work is organized as follows: in Sections 2 and 3, we respectively deal with two descent extensions of the HS and PRP methods, together with their global convergence analyses. We present comparative numerical experiments in Section 4. Finally, we make conclusions in Section 5.

2. A descent adaptive Dai–Liao method

Here, at first we suggest an extension of the HS method which can be considered as an adaptive version of the CG method proposed in [15]. Then, we conduct a brief global convergence analysis for uniformly convex objective functions.

Employing the features of quasi–Newton methods [31], Dai and Liao [15] (DL) proposed an extended conjugacy condition which yields one of the well–known extensions of β_k^{HS} as follows:

$$\beta_k^{DL} = \beta_k^{HS} - t\frac{g_{k+1}^T s_k}{d_k^T y_k}, \tag{2.1}$$

where t is a nonnegative parameter. It can be seen that the efficient CG methods proposed by Hager and Zhang [21,22], and Dai and Kou [14] can be regarded as adaptive versions of the DL method with sufficient descent property. However, generally the DL method fails to guarantee even the descent condition (1.4). Also, numerical behavior of the DL method is very dependent to the parameter t for which there is no any optimal choice [3]. Recently, Babaie–Kafaki and Ghanbari [6, 7, 9, 11], and Fatemi [19] proposed several adaptive choices for the parameter t in (2.1). Among them, the choices suggested in [9] and [19] may guarantee the sufficient descent condition (1.5). Here, using (1.8), we suggest another adaptive choice for the DL parameter which ensures (1.5).

Note that from (1.3), search direction of the DL method can be written as follows:

$$d_{k+1}^{DL} = -g_{k+1} + \beta_k^{HS} d_k - t\frac{g_{k+1}^T s_k}{d_k^T y_k} d_k. \tag{2.2}$$

Also, if we let $\beta_k = \beta_k^{HS}$ in (1.8), then we achieve a three–term extension of the HS method, here called TTHS, with the following search direction:

$$d_{k+1}^{TTHS} = -g_{k+1} + \beta_k^{HS} d_k - \beta_k^{HS} \frac{g_{k+1}^T d_k}{||g_{k+1}||^2} g_{k+1}. \tag{2.3}$$

As mentioned in Section 1, TTHS satisfies the sufficient descent condition (1.5). Also, as shown in [25], a truncated version of (2.3) turns out to be numerically effective. So, motivated by these desirable features, it is reasonable to compute the DL parameter t in a way to tend d_{k+1}^{DL} to d_{k+1}^{TTHS} as closer as possible. That is, similar to the approach of [14], a reasonable choice for t can be achieved by solving the following least–squares problem:

$$\min_t ||d_{k+1}^{DL} - d_{k+1}^{TTHS}||^2. \tag{2.4}$$

After some algebraic manipulations, the solution of (2.4) can be obtained as follows:

$$\bar{t}_{k_1}^* = \frac{(g_{k+1}^T y_k)(g_{k+1}^T s_k)}{||g_{k+1}||^2 ||s_k||^2}. \tag{2.5}$$

However, although its denominator is positive, $\bar{t}_{k_1}^*$ is not necessarily nonnegative. Moreover, considering the eigenvalue analysis carried out in [9], the DL method with $t = \bar{t}_{k_1}^*$ may not possesses the descent property. To overcome these defects, we suggest the following modified adaptive version of (2.5):

$$t_{k_1}^* = \max\left\{ \frac{(g_{k+1}^T y_k)(g_{k+1}^T s_k)}{||g_{k+1}||^2 ||s_k||^2}, \theta \frac{||y_k||^2}{s_k^T y_k} \right\}, \tag{2.6}$$

where $\theta > \dfrac{1}{4}$ is a real constant. The following theorem is now immediate.

Theorem 2.1. *For the DL method with $t = t_{k_1}^*$ given by (2.6) in which $\theta > \dfrac{1}{4}$, if the line search procedure guarantees that $d_k^T y_k > 0$, for all $k \geq 0$, then the sufficient descent condition (1.5) holds with*

$$c = 1 - \frac{1}{4\theta}. \tag{2.7}$$

Proof. Note that from (2.6) we have

$$d_{k+1}^{DL^T} g_{k+1} \leq d_{k+1}^{\theta^T} g_{k+1},$$

where

$$d_{k+1}^{\theta} = -g_{k+1} + \beta_k^{\theta} d_k, \ \beta_k^{\theta} = \beta_k^{HS} - \theta \frac{||y_k||^2}{s_k^T y_k} \frac{g_{k+1}^T s_k}{d_k^T y_k}.$$

So, since [5, Theorem 1] ensures that d_{k+1}^{θ} satisfies the sufficient descent condition (1.5) with the constant c given by (2.7), the proof is complete. □

As seen in (2.6), $t^*_{k_1}$ is computed based on an adaptive switch from $\bar{t}^*_{k_1}$ given by (2.5) to $\theta\frac{||y_k||^2}{s_k^T y_k}$ with $\theta > \frac{1}{4}$ when the optimal choice $t = \bar{t}^*_{k_1}$ fails to guarantee the sufficient descent property. Note that the popular Wolfe line search conditions, i.e.

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k d_k^T g_k, \tag{2.8}$$

$$d_k^T g_{k+1} \geq \sigma d_k^T g_k, \tag{2.9}$$

with $0 < \delta < \sigma < 1$, ensure that $d_k^T y_k > 0$, as supposed in Theorem 2.1. In what follows, we discuss global convergence of the DL method with $t = t^*_{k_1}$ when the line search fulfills the strong Wolfe conditions consisting of (2.8) and the following strengthened version of (2.9):

$$|d_k^T g_{k+1}| \leq -\sigma d_k^T g_k. \tag{2.10}$$

So, we need to consider the following standard assumptions.

Assumption 2.1. The level set $\mathcal{L} = \{x| \ f(x) \leq f(x_0)\}$ 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

$$||\nabla f(x) - \nabla f(y)|| \leq L||x - y||, \ \forall x, y \in \mathcal{N}. \tag{2.11}$$

Now, we can establish the following global convergence theorem, using [15, Theorem 3.3].

Theorem 2.2. *Suppose that Assumption 2.1 holds. Consider a CG method in the form of (1.2)–(1.3) with the CG parameter β_k^{DL} defined by (2.1) in which $t = t^*_{k_1}$ given by (2.6) with $\theta > \frac{1}{4}$. If the objective function f is uniformly convex on $\mathcal{N}$ and the step length α_k is determined such that the strong Wolfe conditions (2.8) and (2.10) are satisfied, then the method converges in the sense that $\lim_{k\to\infty} ||g_k|| = 0$.*

Proof. At first, note that from Theorem 2.1 and the line search condition (2.8), the sequence $\{x_k\}_{k\geq 0}$ is a subset of the level set $\mathcal{L}$. Also, uniform convexity of the differentiable function f ensures that there exists a positive constant μ such that

$$s_k^T y_k \geq \mu ||s_k||^2. \tag{2.12}$$

(See [30, Theorem 1.3.16].) So, from Cauchy–Schwarz inequality, (2.11) and (2.12) we get

$$t^*_{k_1} \leq \max\left\{\frac{||g_{k+1}||^2 L||s_k||^2}{||g_{k+1}||^2||s_k||^2}, \theta\frac{L^2||s_k||^2}{\mu||s_k||^2}\right\} = \max\left\{L, \theta\frac{L^2}{\mu}\right\},$$

ensuring boundedness of $t^*_{k_1}$. Remainder of the proof is similar to the proof of [15, Theorem 3.3] and here is omitted. □

In order to achieve the global convergence without convexity assumption on the objective function, we can employ Powell's nonnegative restriction of the CG parameters [15,28] as follows:

$$\beta_k^{DL+} = \max\left\{\frac{g_{k+1}^T y_k}{d_k^T y_k}, 0\right\} - t_{k_1}^* \frac{g_{k+1}^T s_k}{d_k^T y_k}, \tag{2.13}$$

with $t_{k_1}^*$ given by (2.6). Using [15, Theorem 3.6], global convergence of the above adaptive version of the DL+ method can be established independent to the objective function convexity.

3. A descent extension of the Polak–Ribière–Polyak method

Here, we propose an extension of the PRP method which can be regarded as an adaptive version of the CG method proposed in [8]. Also, we show that the method is globally convergent under the Wolfe conditions.

In a recent effort to make a modification on the PRP method in order to achieve the descent property, similar to extension of the HS method proposed by Dai and Liao [15], Babaie–Kafaki and Ghanbari [8] dealt with an extension of β_k^{PRP} as follows:

$$\beta_k^{EPRP} = \beta_k^{PRP} - t\frac{g_{k+1}^T d_k}{||g_k||^2}, \tag{3.1}$$

where t is a nonnegative parameter. The eigenvalue analysis carried out in [8] showed that if

$$t = p\frac{||y_k||^2}{||g_k||^2} + q\left(\frac{1}{2}\frac{d_k^T y_k}{||d_k||||g_k||} - \frac{||g_k||}{||d_k||}\right)^2, \tag{3.2}$$

with $p > \frac{1}{4}$ and $q \geq -1$, then EPRP satisfies the descent condition (1.4). Moreover, for $q = 0$ the method reduces to the DPRP method proposed by Yu et al. [32] which satisfies the sufficient descent condition (1.5) (see also [4]). Nevertheless, finding the optimal value of t in (3.1) can be considered as an open problem. Next, we propose an adaptive choice for the EPRP parameter t, similar to the approach of Section 2.

Note that from (1.3) and (3.1) we have

$$d_{k+1}^{EPRP} = -g_{k+1} + \beta_k^{PRP} d_k - t\frac{g_{k+1}^T d_k}{||g_k||^2} d_k. \tag{3.3}$$

Furthermore, the choice $\beta_k = \beta_k^{PRP}$ in (1.8) yields a three–term extension of the PRP method, namely TTPRP, that its search direction is given by

$$d_{k+1}^{TTPRP} = -g_{k+1} + \beta_k^{PRP} d_k - \beta_k^{PRP} \frac{g_{k+1}^T d_k}{||g_{k+1}||^2} g_{k+1}, \tag{3.4}$$

being theoretically effective, especially because of satisfying the sufficient descent condition (1.5). In addition, a truncated version of (3.4) turns out to be computationally promising [25]. Hence, it is reasonable to approach d_{k+1}^{EPRP} to d_{k+1}^{TTPRP} in the sense of computing the parameter t as a solution of the following minimization problem:

$$\min_t ||d_{k+1}^{EPRP} - d_{k+1}^{TTPRP}||^2,$$

which yields

$$\bar{t}_{k_2}^* = \frac{(g_{k+1}^T y_k)(g_{k+1}^T d_k)}{||g_{k+1}||^2 ||d_k||^2}, \tag{3.5}$$

where is not generally nonnegative and also, considering the inequality (3.2), may not guarantee the descent condition (1.4). To overcome these defects, we propose the following modified adaptive version of (3.5):

$$t_{k_2}^* = \max\left\{\frac{(g_{k+1}^T y_k)(g_{k+1}^T d_k)}{||g_{k+1}||^2 ||d_k||^2}, \xi \frac{||y_k||^2}{||g_k||^2}\right\}, \tag{3.6}$$

where $\xi > \frac{1}{4}$ is a real constant. The following theorem is now immediate.

Theorem 3.1. *For the EPRP method with $t = t_{k_2}^*$ given by (3.6) in which $\xi > \frac{1}{4}$, if $g_k \neq 0$, then the sufficient descent condition (1.5) holds with*

$$c = 1 - \frac{1}{4\xi}. \tag{3.7}$$

Proof. From (3.6) we have

$$d_{k+1}^{EPRP^T} g_{k+1} \leq d_{k+1}^{DPRP^T} g_{k+1},$$

where, as mentioned before, DPRP is an adaptive version of EPRP with $p = \xi > \frac{1}{4}$ and $q = 0$ in (3.2) [32], satisfying the sufficient descent condition (1.5) with the constant c given by (3.7). (See [4, Theorem 2.2] or [32, Theorem 2.1].) So, the proof is complete. □

As seen in (3.6), by the choice $t = t_{k_2}^*$ the EPRP method adaptively reduces to the DPRP method of [32] to guarantee the sufficient descent property. Now, we deal with global convergence of the EPRP method with $t = t_{k_2}^*$ when the line search fulfills the Wolfe conditions. Similar to the convergence analysis conducted in Section 2, here we need to consider Assumption 2.1 in order to establish a convergence result.

Theorem 3.2. *Consider a CG method in the form of (1.2)–(1.3) with the CG parameter β_k^{EPRP} defined by (3.1) in which $t = t_{k_2}^*$ given by (3.6) with $\xi > \frac{1}{4}$,* *and the step length α_k is determined such that the Wolfe conditions* (2.8) *and* (2.9) *are satisfied. If there exists a positive constant α^* such that $\alpha_k \geq \alpha^*$, $\forall k \geq 0$, then $\lim_{k\to\infty} ||g_k|| = 0$.*

Proof. Considering Theorem 3.1, since the sufficient descent condition (1.5) is satisfied, the proof is similar to the proof of [33, Theorem 3.2] and here is omitted. □

4. **Numerical experiments**

Here, we present some numerical results obtained by applying C++ implementations of the following nonlinear CG methods:

- ADL: the method with the search direction (2.2) in which the parameter t is computed by (2.6), being an adaptive version of the DL method;
- AEPRP: the method with the search direction (3.3) in which the parameter t is computed by (3.6), being an adaptive version of the EPRP method;

comparing with the three–term CG methods TTHS and TTPRP, respectively with the search directions (2.3) and (3.4). The codes were run on a PC with 3.6 GHz Intel I7–4790 of CPU, 4 GB of RAM and Centos 6.2 server Linux operation system. Since CG methods are appropriate for solving large–scale problems, the experiments were performed on a set of 64 unconstrained optimization test problems of the CUTEr collection [20] with default dimensions being at least equal to 1000, as specified in [8]. (See also Hager's home page: 'http://www.math.ufl.edu/∼hager/'.)

For all the four methods, we used the approximate Wolfe conditions proposed by Hager and Zhang [21] in the line search procedure, with the same parameter values as specified in [22]. Also, we used the steepest descent direction as the initial search direction of the methods. That is, for all the four methods we set $d_0 = -g_0$. For the ADL and AEPRP methods, we respectively set $\theta = 1.1$ in (2.6) and $\xi = 1.6$ in (3.6) because of their promising computational results among the different choices of the set $\{0.1k\}_{k=3}^{20}$. In addition, 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 Dolan–Moré performance profile [16] which for every $\omega \geq 1$ yields 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 results of comparisons. Note that in Figure 1 total number of function and gradient evaluations is equal to $N_f + 3N_g$ where N_f and N_g respectively stand for the number of function and gradient evaluations [22].

As shown by the figures, although ADL and AEPRP are approximately competitive, AEPRP is slightly preferable to ADL, especially with respect to the CPU time, and both of the methods outperform the TTHS method. Moreover,

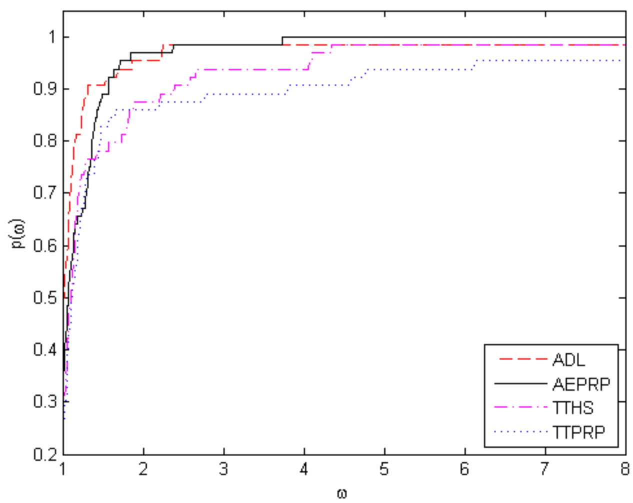

Figure 1. Total number of function and gradient evaluations performance profiles

the figures show that TTHS outperforms TTPRP. Thus, our adaptive choices for the parameters of the DL and EPRP methods turn out to be practically effective.

5. Conclusions

We have dealt with the open problem of finding an optimal value for parameter of the Dai–Liao method by approaching its search directions to the search directions of a descent three–term extension of the Hestenes–Stiefel method. Our analysis has led to a choice for the Dai–Liao parameter given by (2.6) which guarantees the sufficient descent property as well as the global convergence for uniformly convex objective functions when the line search fulfills the strong Wolfe conditions. We have conducted a similar analysis on a recent extended Polak–Ribière–Polyak method and achieved another globally convergent one–parameter nonlinear conjugate gradient method with sufficient descent property in which the parameter is computed by (3.6). Although the choices (2.6) and (3.6) are not optimal, they are adaptive hybridizations of the optimal choice and another term which plays an important role to achieve the effective sufficient descent property as well as to ensure nonnegativity of the parameter. Preliminary numerical results showed that the proposed methods are computationally promising.

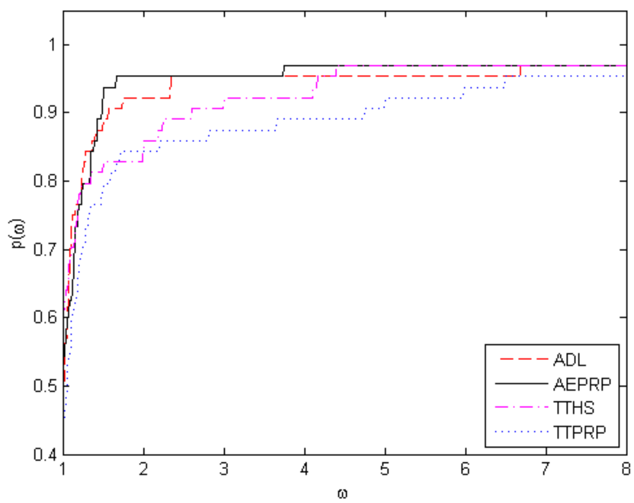

FIGURE 2. CPU time performance profiles

Acknowledgements

This research was supported by Research Councils of Semnan University and Ferdowsi University of Mashhad. The authors are grateful to Professor William W. Hager for providing the line search code. They also thank the anonymous reviewers for their valuable comments and suggestions helped to improve the presentation.

References

[1] N. Andrei, Numerical comparison of conjugate gradient algorithms for unconstrained optimization, *Stud. Inform. Control* **16** (2007), no. 4, 333–352.

[2] N. Andrei, A modified Polak–Ribière–Polyak conjugate gradient algorithm for unconstrained optimization, *Optimization* **60** (2011), no. 12, 1457–1471.

[3] N. Andrei, Open problems in conjugate gradient algorithms for unconstrained optimization, *Bull. Malays. Math. Sci. Soc.* **34** (2011), no. 2, 319–330.

[4] S. Babaie–Kafaki, An eigenvalue study on the sufficient descent property of a modified Polak–Ribière–Polyak conjugate gradient method, *Bull. Iranian Math. Soc.* **40** (2014), no. 1, 235–242.

[5] S. Babaie–Kafaki, On the sufficient descent condition of the Hager–Zhang conjugate gradient methods, *4OR* **12** (2014), no. 3, 285–292.

[6] S. Babaie–Kafaki, On optimality of two adaptive choices for the parameter of Dai–Liao method, *Optim. Lett.* **10** (2016), no. 8, 1789–1797.

[7] S. Babaie–Kafaki and R. Ghanbari, The Dai–Liao nonlinear conjugate gradient method with optimal parameter choices, *European J. Oper. Res.* **234** (2014), no. 3, 625–630.

[8] S. Babaie–Kafaki and R. Ghanbari, A descent extension of the Polak–Ribière–Polyak conjugate gradient method, *Comput. Math. Appl.* **68** (2014), no. 12, 2005–2011.
[9] S. Babaie–Kafaki and R. Ghanbari, A descent family of Dai–Liao conjugate gradient methods, *Optim. Methods Softw.* **29** (2014), no. 3, 583–591.
[10] S. Babaie–Kafaki and R. Ghanbari, Two modified three–term conjugate gradient methods with sufficient descent property, *Optim. Lett.* **8** (2014), no. 8, 2285–2297.
[11] S. Babaie–Kafaki and R. Ghanbari, Two optimal Dai–Liao conjugate gradient methods, *Optimization* **64** (2015), no. 11, 2277–2287.
[12] W. Cheng, A two–term PRP–based descent method, *Numer. Funct. Anal. Optim.* **28** (2007), no. 11-12, 1217–1230.
[13] Z. Dai, X. Chen and F. Wen, A modified Perry's conjugate gradient method–based derivative–free method for solving large–scale nonlinear monotone equations, *Appl. Math. Comput.* **270** (2015), no. 7, 378–386.
[14] Y.H. Dai and C.X. Kou, A nonlinear conjugate gradient algorithm with an optimal property and an improved Wolfe line search, *SIAM J. Optim.* **23** (2013), no. 1, 296–320.
[15] Y.H. Dai and L.Z. Liao, New conjugacy conditions and related nonlinear conjugate gradient methods, *Appl. Math. Optim.* **43** (2001), no. 1, 87–101.
[16] E.D. Dolan and J.J. Moré, Benchmarking optimization software with performance profiles, *Math. Program.* **91** (2002), no. 2, Ser. A, 201–213.
[17] X.L. Dong, H.W. Liu and Y.B. He, A self–adjusting conjugate gradient method with sufficient descent condition and conjugacy condition, *J. Optim. Theory Appl.* **165** (2015), no. 1, 225–241.
[18] X.L. Dong, H.W. Liu, Y.B. He and X.M. Yang, A modified Hestenes–Stiefel conjugate gradient method with sufficient descent condition and conjugacy condition, *J. Comput. Appl. Math.* **281** (2015), no. 1, 239–249.
[19] M. Fatemi, An optimal parameter for Dai–Liao family of conjugate gradient methods, *J. Optim. Theory Appl.* **169** (2016), no. 2, 587–605.
[20] N.I.M. Gould, D. Orban and Ph.L. Toint, CUTEr: a constrained and unconstrained testing environment, revisited, *ACM Trans. Math. Software* **29** (2003), no. 4, 373–394.
[21] W.W. Hager and H. Zhang, A new conjugate gradient method with guaranteed descent and an efficient line search, *SIAM J. Optim.* **16** (2005), no. 1, 170–192.
[22] W.W. Hager and H. Zhang, Algorithm 851: CG_Descent, a conjugate gradient method with guaranteed descent, *ACM Trans. Math. Software* **32** (2006), no. 1, 113–137.
[23] W.W. Hager and H. Zhang, A survey of nonlinear conjugate gradient methods, *Pac. J. Optim.* **2** (2006), no. 1, 35–58.
[24] M.R. Hestenes and E. Stiefel, Methods of conjugate gradients for solving linear systems, *J. Research Nat. Bur. Standards* **49** (1952), no. 6, 409–436.
[25] Y. Narushima, H. Yabe and J.A. Ford, A three–term conjugate gradient method with sufficient descent property for unconstrained optimization, *SIAM J. Optim.* **21** (2011), no. 1, 212–230.
[26] E. Polak and G. Ribière, Note sur la convergence de méthodes de directions conjuguées, *Rev. Française Informat. Recherche Opérationnelle* **3** (1969), no. 16, 35–43.
[27] B.T. Polyak, The conjugate gradient method in extreme problems, *USSR Comp. Math. Math. Phys.* **9** (1969), no. 4, 94–112.
[28] M.J.D. Powell, Convergence properties of algorithms for nonlinear optimization, *SIAM Rev.* **28** (1986), no. 4, 487–500.
[29] 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** (2012), no. 3, 733–757.

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

[31] C. Xu and J.Z. Zhang, A survey of quasi–Newton equations and quasi–Newton methods for optimization, *Ann. Oper. Res.* **103** (2001), no. 1-4, 213–234.

[32] G. Yu, L. Guan and G. Li, Global convergence of modified Polak–Ribière–Polyak conjugate gradient methods with sufficient descent property, *J. Ind. Manag. Optim.* **4** (2008), no. 3, 565–579.

[33] G.L. Yuan, Modified nonlinear conjugate gradient methods with sufficient descent property for large–scale optimization problems, *Optim. Lett.* **3** (2009), no. 1, 11–21.

[34] G.L. Yuan, X.W. Lu and Z.X. Wei, A conjugate gradient method with descent direction for unconstrained optimization, *J. Comput. Appl. Math.* **233** (2009), no. 2, 519–530.

[35] G.L. Yuan, Z. Wei and Q. Zhao, A modified Polak–Ribière–Polyak conjugate gradient algorithm for large–scale optimization problems, *IIE Trans.* **46** (2014), no. 4, 397–413.

[36] G.L. Yuan and M. Zhang, A modified Hestenes–Stiefel conjugate gradient algorithm for large–scale optimization, *Numer. Funct. Anal. Optim.* **34** (2013), no. 8, 914–937.

[37] L. Zhang, W. Zhou and D.H. Li, A descent modified Polak–Ribière–Polyak conjugate gradient method and its global convergence, *IMA J. Numer. Anal.* **26** (2006), no. 4, 629–640.

[38] L. Zhang, W. Zhou and D.H. Li, Some descent three–term conjugate gradient methods and their global convergence, *Optim. Methods Softw.* **22** (2007), no. 4, 697–711.

(Saman Babaie–Kafaki) Department of Mathematics, Faculty of Mathematics, Statistics and Computer Science, Semnan University, P.O. Box 35195–363, Semnan, Iran.

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

(Reza Ghanbari) Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, P.O. Box: 9177948953, Mashhad, Iran.

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