RAIRO-Oper. Res. 50 (2016) 567–574
DOI: 10.1051/ro/2016009

RAIRO Operations Research
www.rairo-ro.org

A DESCENT HYBRID MODIFICATION OF THE POLAK–RIBIÈRE–POLYAK CONJUGATE GRADIENT METHOD

Saman Babaie-Kafaki[1] and Reza Ghanbari[2]

Abstract. Hybridizing self-adjusting approach of Dong *et al.* and three-term formulation of Zhang *et al.*, a nonlinear conjugate gradient method is proposed. The method reduces to the Polak–Ribière–Polyak method under the exact line search and satisfies the sufficient descent condition independent of the line search and the objective function convexity. Similar to the Polak–Ribière–Polyak method, the method possesses an automatic restart feature which avoids jamming. Global convergence analyses are conducted when the line search fulfills the popular Wolfe conditions as well as an Armijo-type condition. Numerical experiments are done on a set of CUTEr unconstrained optimization test problems. Results of comparisons show computational efficiency of the proposed method in the sense of Dolan–Moré performance profile.

Mathematics Subject Classification. 90C53, 49M37, 65K05.

Received August 10, 2015. Accepted January 26, 2016.

1. Introduction

Optimization problems occur in most disciplines like engineering, physics, mathematics, economics, administration, commerce, social sciences, and even politics. As a topic of great significance in optimization and nonlinear analysis, unconstrained optimization deals with the following problem:

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

where $f:\mathbb{R}^n \to \mathbb{R}$ is called the objective function, here assumed to be continuously differentiable. Emphasizing importance of unconstrained optimization, in a general strategy that has evolved in recent years, a constrained optimization problem is reformulated as an unconstrained optimization problem, redefining the objective function such that the constraints are simultaneously satisfied when the objective function is minimized.

The most essential approach to optimization is based on computational methods in which iterative numerical procedures are used to generate a series of progressively improved solutions to the problem, starting with an initial estimate of the solution. The process is terminated when some convergence criteria are satisfied. Iterations

Keywords. Unconstrained optimization, conjugate gradient method, sufficient descent condition, line search, global convergence.

[1] Department of Mathematics, Faculty of Mathematics, Statistics and Computer Science, Semnan University, P.O. Box: 35195–363, Semnan, Iran. sbk@semnan.ac.ir

[2] Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, P.O. Box: 9177948953, Mashhad, Iran. rghanbari@um.ac.ir

Article published by EDP Sciences

of the line search–based numerical methods for solving (1.1) are in the following form:

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

in which α_k is a step length to be determined by a line search technique and d_k is a descent search direction; that is, $d_k^T g_k < 0$ in which $g_k = \nabla f(x_k)$ [19].

Among the iterative methods for solving large-scale cases of (1.1), conjugate gradient (CG) methods have attracted especial attention since they require storage for only a few n-vectors. Search directions of the CG methods are in the following form:

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

where β_k is a scalar called the CG (update) parameter. Different CG methods mainly correspond to different choices for β_k [15].

As known, there exists a class of CG methods with an approximate restart feature which avoids jamming [15]. One of the efficient member of this class has been proposed by Polak, Ribière [17] and Polyak [18] (PRP), with the following CG parameter:

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

where $y_k = g_{k+1} - g_k$, and $||.||$ stands for the Euclidean norm. Note that if the iterations jam, then $x_{k+1} - x_k$ is small. So, the factor y_k in the numerator of β_k^{PRP} tends to zero and consequently, β_k becomes small. Therefore, the search direction d_{k+1} tends to the steepest descent direction. This automatic restart feature is considerable in the computational point of view. However, in theoretical point of view, it is remarkable that the PRP method may fail to generate descent directions.

In a recent effort to achieve a descent version of the PRP method, based on the approach of [6] Babaie–Kafaki and Ghanbari [4] proposed an extension of β_k^{PRP} as follows:

$$\beta_k^{EPRP} = \beta_k^{PRP} - t\frac{g_{k+1}^T d_k}{||g_k||^2}, \tag{1.4}$$

in which

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

with $p > \dfrac{1}{4}$ and $q \geq -1$. Note that if $q = 0$, then the method reduces to the DPRP method proposed by Yu *et al.* [21] which satisfies the sufficient descent condition, *i.e.*,

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

where c is a positive constant (see also [3,22]). In another attempt, Zhang *et al.* [23] (ZZL) proposed a three-term CG method with the following search directions:

$$d_0 = -g_0,\ d_{k+1}^{ZZL} = -g_{k+1} + \beta_k^{PRP} d_k - \frac{g_{k+1}^T d_k}{||g_k||^2} y_k,\ \forall k \geq 0, \tag{1.6}$$

satisfying the sufficient descent condition $d_k^T g_k = -||g_k||^2$, $\forall k \geq 0$, independent of the line search and the objective function convexity. Spectral PRP methods with sufficient descent property have been developed by Andrei [1], Wan *et al.* [20], and Deng *et al.* [8]. Using the project of PRP search direction, Cheng [5] dealt with a modified PRP method which satisfies (1.5).

Here, we deal with another extension of the PRP method with sufficient descent property, hybridizing self-adjusting approach of Dong *et al.* [11] and three-term formulation of Zhang *et al.* [23]. The remainder of this work is organized as follows. In Section 2, the method is proposed and its global convergence is discussed. Numerical experiments are done in Section 3 and conclusions are drawn in Section 4.

2. A modified Polak–Ribière–Polyak method

In a recent attempt to make a modification on the Hestenes–Stiefel [16] (HS) method in order to achieve the descent property, Dong *et al.* [11] proposed the following CG parameter:

$$\beta_k^{CHS} = \underbrace{\frac{g_{k+1}^T y_k}{d_k^T y_k}}_{\beta_k^{HS}} - t\frac{\max\{g_{k+1}^T d_k, 0\}}{(d_k^T y_k)^2}\left(\frac{g_{k+1}^T y_k}{||g_{k+1}||}\right)^2,$$

where t is a real parameter in $(\frac{1}{4}, +\infty)$. It can be seen that the CHS search directions satisfy the descent condition when $g_{k+1}^T y_k \geq 0$. In [11] an adaptive version of CHS has been proposed which fulfills the sufficient descent condition (see also [2,10,12] and the references therein to learn more about the recent extensions of the HS method with sufficient descent property).

Here, considering similarity between β_k^{HS} and β_k^{PRP}, we suggest the following modification on β_k^{PRP}:

$$\beta_k^{CPRP} = \beta_k^{PRP} - t\frac{\max\{g_{k+1}^T d_k, 0\}}{||g_k||^4}\left(\frac{g_{k+1}^T y_k}{||g_{k+1}||}\right)^2, \tag{2.1}$$

with $t \in (\frac{1}{4}, +\infty)$. So, defining $d_{k+1}^{CPRP} = -g_{k+1} + \beta_k^{CPRP} d_k$, we get

$$\begin{aligned} d_{k+1}^{CPRP^T} g_{k+1} = -||g_{k+1}||^2 + \frac{g_{k+1}^T y_k}{||g_k||^2} g_{k+1}^T d_k \\ - t\frac{\max\{g_{k+1}^T d_k, 0\}}{||g_k||^4}\left(\frac{g_{k+1}^T y_k}{||g_{k+1}||}\right)^2 g_{k+1}^T d_k. \end{aligned}$$

Now, assume that $g_{k+1}^T y_k \geq 0$. If $g_{k+1}^T d_k \leq 0$, then we have

$$d_{k+1}^{CPRP^T} g_{k+1} \leq -||g_{k+1}||^2. \tag{2.2}$$

On the other hand, if $g_{k+1}^T d_k > 0$, then we have

$$\begin{aligned} d_{k+1}^{CPRP^T} g_{k+1} &= -||g_{k+1}||^2 \\ &+ ||g_{k+1}||^2 \left(\frac{(g_{k+1}^T y_k)(g_{k+1}^T d_k)}{||g_k||^2 ||g_{k+1}||^2} - t\frac{(g_{k+1}^T y_k)^2 (g_{k+1}^T d_k)^2}{||g_k||^4 ||g_{k+1}||^4}\right). \end{aligned} \tag{2.3}$$

Note that for any real constant ξ we have

$$\xi - \xi^2 t \leq \frac{1}{4t}.$$

Thus, from (2.3) we get

$$d_{k+1}^{CPRP^T} g_{k+1} \leq -\left(1 - \frac{1}{4t}\right)||g_{k+1}||^2. \tag{2.4}$$

So, from (2.2) and (2.4), if $g_{k+1}^T y_k \geq 0$, then CPRP satisfies the descent condition (2.4). However, when $g_{k+1}^T y_k < 0$, CPRP fails to guarantee the descent property. In such situation, we switch from d_{k+1}^{CPRP} to d_{k+1}^{ZZL} given by (1.6) which satisfies the sufficient descent condition $d_{k+1}^{ZZL^T} g_{k+1} = -||g_{k+1}||^2$. So, based on this hybrid scheme, we propose a class of one–parameter nonlinear CG methods with the following search directions:

$$d_0 = -g_0, \; d_{k+1}^{HCPRP} = \begin{cases} d_{k+1}^{CPRP}, & g_{k+1}^T y_k \geq 0, \\ d_{k+1}^{ZZL}, & g_{k+1}^T y_k < 0, \end{cases} \tag{2.5}$$

with $t \in (\frac{1}{4}, +\infty)$ in (2.1). The following theorem is now immediate.

Theorem 2.1. *For the HCPRP method, if $t \geq \bar{t}$ in which $\bar{t} > \frac{1}{4}$ is a real constant, then the sufficient descent condition* (1.5) *holds with $c = 1 - \frac{1}{4\bar{t}}$.*

Although search directions of other extended PRP methods which satisfy the sufficient descent condition can be used instead of d_{k+1}^{ZZL} in (2.5), our numerical experiments showed that hybridizing the CPRP and ZZL methods seem to be reasonable in computational pint of view. When the line search is exact, since we have $g_{k+1}^T d_k = 0$, the HCPRP method reduces to the PRP method. Also, it can be seen that HCPRP inherits the automatic restart feature from the PRP method. So, in practical computations jamming does not occur for the HCPRP method. In what follows, we deal with global convergence of the HCPRP method under the Wolfe line search conditions, *i.e.*,

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k \nabla f(x_k)^T d_k, \tag{2.6}$$

$$\nabla f(x_k + \alpha_k d_k)^T d_k \geq \sigma \nabla f(x_k)^T d_k, \tag{2.7}$$

where $0 < \delta < \sigma < 1$ [19], being popular in convergence analysis and implementation of the CG methods [7]. The following basic assumptions are now needed [22].

Assumption 2.2.

(1) The level set $\Omega = \{x \in \mathbb{R}^n |\ f(x) \leq f(x_0)\}$ is bounded, where x_0 is a given starting point. That is, there exists a positive constant B such that

$$||x|| \leq B,\ \forall x \in \Omega. \tag{2.8}$$

(2) In an open convex set Ω_0 that contains Ω, f has a lower bound and its gradient is Lipschitz continuous; namely, there exists a positive constant L such that

$$||\nabla f(x) - \nabla f(y)|| \leq L||x - y||,\ \forall x, y \in \Omega_0. \tag{2.9}$$

The following global convergence theorem can be established similar to Theorem 3.2 of [22]. So, the proof is omitted.

Theorem 2.3. *Suppose that Assumption* 2.2 *holds. Consider the HCPRP method in which $t \geq \bar{t}$ with the constant $\bar{t} > \frac{1}{4}$, and the step length α_k is determined such that the Wolfe conditions* (2.6) *and* (2.7) *are satisfied. If there exists a positive constant α^* such that $\alpha_k \geq \alpha^*$, $\forall k \geq 0$, then $\lim_{k\to\infty} ||g_k|| = 0$.*

Next, we deal with global convergence of the HCPRP method under the following Armijo-type line search strategy proposed in [23].

Line search 2.4. Choose the constants ρ and δ in the interval $(0, 1)$. Compute the step length $\alpha_k = \max\{\rho^j,\ j = 0, 1, \ldots\}$, satisfying the following condition:

$$f(x_k + \alpha_k d_k) \leq f(x_k) - \delta \alpha_k^2 ||d_k||^2. \tag{2.10}$$

Assumption 2.2 implies that there exists a positive constant γ such that

$$||\nabla f(x)|| \leq \gamma,\ \forall x \in \Omega. \tag{2.11}$$

The line search condition (2.10) ensures that $\{x_k\}_{k\geq 0} \subset \Omega$ and the sequence $\{f(x_k)\}_{k\geq 0}$ is decreasing. Moreover, from Assumption 2.2, the sequence $\{f(x_k)\}_{k\geq 0}$ is bounded below and hence, it is convergent. Therefore, from (2.10) we get

$$\sum_{k\geq 0} \alpha_k^2 ||d_k||^2 < \infty,$$

which yields

$$\lim_{k\to\infty} \alpha_k ||d_k|| = 0. \tag{2.12}$$

In what follows, we assume that there exists a positive constant Γ such that

$$\max\{g_{k+1}^T d_k, 0\} \leq \Gamma, \ \forall k \geq 0. \tag{2.13}$$

Considering the exact line search, if α_k is near to the optimal step length, then inequality (2.13) may be satisfied for the enough large values of Γ. The following lemma plays an essential role in our analyses.

Lemma 2.5. *For the HCPRP method, suppose that Assumption 2.2 holds, $t \geq \bar{t}$ with the constant $\bar{t} > \dfrac{1}{4}$, inequality (2.13) is satisfied, and the step length α_k is determined using Line search 2.4. If there exists a positive constant ε such that*

$$||g_k|| \geq \varepsilon, \ \forall k \geq 0, \tag{2.14}$$

then there exists a positive constant M such that

$$||d_k|| \leq M, \ \forall k \geq 0.$$

Proof. From Cauchy–Schwarz inequality, (1.6), (2.9), (2.11) and (2.14) we have

$$\begin{aligned} ||d_{k+1}^{ZZL}|| &\leq ||g_{k+1}|| + \frac{|g_{k+1}^T y_k|}{||g_k||^2}||d_k|| + \frac{|g_{k+1}^T d_k|}{||g_k||^2}||y_k|| \\ &\leq ||g_{k+1}|| + 2\frac{||g_{k+1}|| \ ||y_k|| \ ||d_k||}{||g_k||^2} \leq \gamma + 2\frac{L\gamma\alpha_k||d_k||}{\varepsilon^2}||d_k||. \end{aligned} \tag{2.15}$$

In addition, from (2.8) and (2.13) we get

$$\begin{aligned} ||d_{k+1}^{CPRP}|| &\leq ||g_{k+1}|| + \frac{|g_{k+1}^T y_k|}{||g_k||^2}||d_k|| + t\frac{\max\{g_{k+1}^T d_k, 0\}}{||g_k||^4}\frac{(g_{k+1}^T y_k)^2}{||g_{k+1}||^2}||d_k|| \\ &\leq ||g_{k+1}|| + \frac{L\gamma\alpha_k||d_k||}{\varepsilon^2}||d_k|| + t\frac{2\Gamma L^2 B\alpha_k||d_k||}{\varepsilon^4}||d_k||. \end{aligned} \tag{2.16}$$

Now, from (2.5), (2.15) and (2.16) we have

$$||d_{k+1}^{HCPRP}|| \leq \gamma + \left(2\frac{L\gamma}{\varepsilon^2} + t\frac{2\Gamma L^2 B}{\varepsilon^4}\right)(\alpha_k||d_k||)||d_k||.$$

Considering (2.12), there exists a constant $r \in (0,1)$ and a large enough integer k_0 such that

$$\left(2\frac{L\gamma}{\varepsilon^2} + t\frac{2\Gamma L^2 B}{\varepsilon^4}\right)\alpha_k||d_k|| \leq r, \ \forall k \geq k_0.$$

Hence, for any $k > k_0$ we have

$$\begin{aligned} ||d_{k+1}^{HCPRP}|| &\leq \gamma + r||d_k|| \\ &\leq \gamma(1 + r + r^2 + \ldots + r^{k-k_0}) + r^{k-k_0+1}||d_{k_0}|| \\ &\leq \frac{\gamma}{1-r} + ||d_{k_0}||. \end{aligned}$$

So, if we let

$$M = \max\{||d_0||, ||d_1||, \ldots, ||d_{k_0-1}||, \frac{\gamma}{1-r} + ||d_{k_0}||\},$$

then the proof is complete. □

Now, the following global convergence theorem can be established for HCPRP, similar to the proof of Theorem 3.2 of [23].

Theorem 2.6. *For the HCPRP method, if Assumption 2.2 holds, $t \geq \bar{t}$ with the constant $\bar{t} > \frac{1}{4}$, inequality (2.13) is satisfied, and the step length α_k is determined using Line search 2.4, then*

$$\liminf_{k\to\infty} ||g_k|| = 0. \tag{2.17}$$

Proof. If $\liminf_{k\to\infty} \alpha_k > 0$, then Theorem 2.1, (2.11) and (2.12) lead to (2.17). Assume that $\liminf_{k\to\infty} \alpha_k = 0$. That is, there exists an infinite index set K such that

$$\lim_{k\in K,\ k\to\infty} \alpha_k = 0. \tag{2.18}$$

So, considering Line search 2.4, for enough large values $k \in K$, $\rho^{-1}\alpha_k$ does not satisfy (2.10), *i.e.*,

$$f(x_k + \rho^{-1}\alpha_k d_k) - f(x_k) > -\delta\rho^{-2}\alpha_k^2||d_k||^2. \tag{2.19}$$

Also, from the mean-value theorem, for some $\zeta \in (0,1)$ we can write

$$f(x_k + \rho^{-1}\alpha_k d_k) - f(x_k) = \rho^{-1}\alpha_k \nabla f(x_k + \zeta\rho^{-1}\alpha_k d_k)^T d_k,$$

which by considering Theorem 2.1 and inequalities (1.5) and (2.9) we get

$$\begin{aligned} f(x_k + \rho^{-1}\alpha_k d_k) - f(x_k) &= \rho^{-1}\alpha_k g_k^T d_k + \rho^{-1}\alpha_k(\nabla f(x_k + \zeta\rho^{-1}\alpha_k d_k) - g_k)^T d_k \\ &\leq -c\rho^{-1}\alpha_k||g_k||^2 + L\rho^{-2}\alpha_k^2||d_k||^2. \end{aligned} \tag{2.20}$$

So, from (2.19) and (2.20) we get

$$||g_k||^2 \leq \frac{L+\delta}{c}\rho^{-1}\alpha_k||d_k||^2. \tag{2.21}$$

Now, if (2.17) does not hold, then there exists a positive constant ε such that (2.14) holds. Therefore, from Lemma 2.5, $||d_k||$ is bounded above and as a result, (2.21) leads to a contradiction with (2.18). So, the proof is complete. □

3. Numerical experiments

Here, we present some numerical results obtained by applying C++ implementations of the HCPRP and ZZL methods, and also, the DPRP method proposed by Yu *et al.* [21] in which, as mentioned in Section 1, the CG parameter is computed by (1.4) with

$$t = \mu\frac{||y_k||^2}{||g_k||^2}, \tag{3.1}$$

where $\mu > \frac{1}{4}$ is a real constant. 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 [13] with default dimensions being at least equal to 1000, as specified in [4].

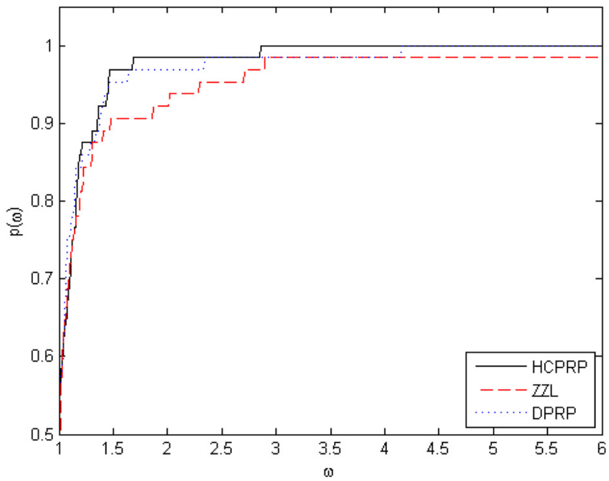

FIGURE 1. Total number of function and gradient evaluations performance profiles.

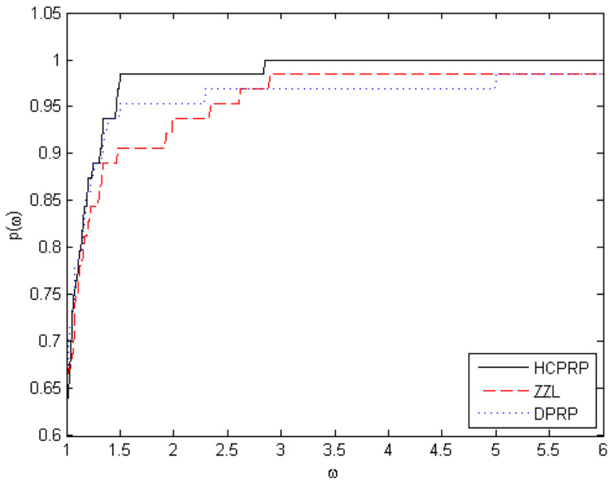

FIGURE 2. CPU time performance profiles.

For the HCPRP method, we set $t = 1.0$ because of promising numerical results obtained among the different values $t \in \{0.1k\}_{k=3}^{20}$. Also, for the DPRP method we set $\mu = 0.5$ in (3.1), as suggested in [21]. We used the effective approximate Wolfe conditions proposed by Hager and Zhang [14] in the line search procedure, with the same parameter values as specified in [14]. Moreover, 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 [9] on the running time and the total number of function and gradient evaluations being equal to $N_f + 3N_g$, 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 ω of the best. Figures 1 and 2 show the results of comparisons. As the figures show, HCPRP is preferable to ZZL and DPRP both in the perspectives of the total number of function and gradient evaluations, and the running time. Thus, our hybridization scheme seem to be practically effective. It is interesting that averagely in 83.39% of the iterations search directions of the HCPRP method is equal to search directions of the CPRP method.

The figures also show that DPRP is generally preferable to ZZL, especially with respect to the total number of function and gradient evaluations.

4. Conclusions

Based on a recent modification of the Hestenes–Stiefel conjugate gradient method made by Dong *et al.*, a modified Polak–Ribière–Polyak method, namely HCPRP, has been proposed. The method can also be considered as an adaptive version of the three-term conjugate gradient method proposed by Zhang *et al.* which retains the sufficient descent property without convexity assumption on the objective function. A brief global convergence analysis has been made when the line search fulfills the Wolfe conditions. Also, the method has been shown to be globally convergent under an Armijo-type line search condition. Preliminary numerical results showed that the method is computationally promising.

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 which improved the quality of this study.

References

[1] N. Andrei, A modified Polak–Ribière–Polyak conjugate gradient algorithm for unconstrained optimization. *Optimization* **60** (2011) 1457–1471.
[2] N. Andrei, An adaptive conjugate gradient algorithm for large–scale unconstrained optimization. *J. Comput. Appl. Math.* **292** (2016) 83–91.
[3] 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) 235–242.
[4] S. Babaie-Kafaki and R. Ghanbari, A descent extension of the Polak–Ribière–Polyak conjugate gradient method. *Comput. Math. Appl.* **68** (2014) 2005–2011.
[5] W. Cheng, A two-term PRP-based descent method. *Numer. Funct. Anal. Optim.* **28** (2007) 1217–1230.
[6] Y.H. Dai and L.Z. Liao, New conjugacy conditions and related nonlinear conjugate gradient methods. *Appl. Math. Optim.* **43** (2001) 87–101.
[7] Y.H. Dai, J.Y. Han, G.H. Liu, D.F. Sun, H.X. Yin and Y.X. Yuan, Convergence properties of nonlinear conjugate gradient methods. *SIAM J. Optim.* **10** (1999) 348–358.
[8] S. Deng, Z. Wan and X. Chen, An improved spectral conjugate gradient algorithm for nonconvex unconstrained optimization problems. *J. Optim. Theory Appl.* **157** (2013) 820–842.
[9] E.D. Dolan and J.J. Moré, Benchmarking optimization software with performance profiles. *Math. Program.* **91** (2002) 201–213.
[10] X.L. Dong, H.W. Liu and Y.B. He, New version of the three-term conjugate gradient method based on spectral scaling conjugacy condition that generates descent search direction. *Appl. Math. Comput.* **269** (2015) 606–617.
[11] 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) 225–241.
[12] 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) 239–249.
[13] N.I.M. Gould, D. Orban and Ph.L. Toint, CUTEr: a constrained and unconstrained testing environment, revisited. *ACM Trans. Math. Softw.* **29** (2003) 373–394.
[14] W.W. Hager and H. Zhang, Algorithm 851: CG_Descent, a conjugate gradient method with guaranteed descent. *ACM Trans. Math. Softw.* **32** (2006) 113–137.
[15] W.W. Hager and H. Zhang, A survey of nonlinear conjugate gradient methods. *Pac. J. Optim.* **2** (2006) 35–58.
[16] M.R. Hestenes and E. Stiefel, Methods of conjugate gradients for solving linear systems. *J. Res. Nat. Bur. Standards* **49** (1952) 409–436.
[17] E. Polak and G. Ribière, Note sur la convergence de méthodes de directions conjuguées. *Rev. Française Informat. Recherche Opérationnelle* **3** (1969) 35–43.
[18] B.T. Polyak, The conjugate gradient method in extreme problems. *USSR Comp. Math. Math. Phys.* **9** (1969) 94–112.
[19] W. Sun and Y.X. Yuan, Optimization Theory and Methods: Nonlinear Programming. Springer, New York (2006).
[20] Z. Wan, Z.L. Yang and Y.L. Wang, New spectral PRP conjugate gradient method for unconstrained optimization. *Appl. Math. Lett.* **24** (2011) 16–22.
[21] 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) 565–579.
[22] G.L. Yuan, Modified nonlinear conjugate gradient methods with sufficient descent property for large-scale optimization problems. *Optim. Lett.* **3** (2009) 11–21.
[23] 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) 629–640.