Optim Lett (2018) 12:1911–1927
https://doi.org/10.1007/s11590-017-1205-y

ORIGINAL PAPER

A family of three-term nonlinear conjugate gradient methods close to the memoryless BFGS method

Min Li[1]

Received: 17 February 2017 / Accepted: 3 October 2017 / Published online: 6 October 2017

Abstract Based on the memoryless BFGS quasi-Newton method, a family of three-term nonlinear conjugate gradient methods are proposed. For any line search, the directions generated by the new methods are sufficient descent. Using some efficient techniques, global convergence results are established when the line search fulfills the Wolfe or the Armijo conditions. Moreover, the r-linear convergence rate of the methods are analyzed as well. Numerical comparisons show that the proposed methods are efficient for the unconstrained optimization problems in the CUTEr library.

Keywords Nonlinear conjugate gradient method · Memoryless BFGS method · Sufficient descent property · Global convergence

1 Introduction

In this paper, we consider the unconstrained optimization problem

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

where $f: \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable and its gradient $g(x)$ is available. Among different kinds of numerical methods for solving problem (1.1), nonlinear conjugate gradient (CG) methods comprise a class of efficient approaches, especially for large-scale problems, due to their low memory requirements and good global convergence properties. A CG method generates a sequence of iterates $x_0, x_1, x_2, \ldots$ by using the recurrence

✉ Min Li
liminmath@aliyun.com

1 Department of Mathematics and Computational Science, Huaihua University, Huaihua 418008, Hunan, People's Republic of China

$$x_{k+1} = x_k + \alpha_k d_k, \qquad k = 0, 1, \ldots \tag{1.2}$$

where α_k is a positive steplength computed by a line search and d_k is the search direction generated by the rule

$$d_k = -g_k + \beta_k d_{k-1}, \qquad d_0 = -g_0, \tag{1.3}$$

where $g_k = g(x_k)$ and β_k is a CG parameter. Different choices for the CG parameter β_k correspond to different CG methods. So far, the researches on the CG methods have made great progress. There have been many famous CG methods, for example, see [5,6,9–12,14,19,22,23,33], etc. In the survey paper [13], Hager and Zhang gave a comprehensive review of the development of different versions of CG methods, with special attention given to their global convergence properties. We refer to the survey paper for more details.

In this paper, we are particularly interested in the Hestenes–Stiefel (HS) [14] method, the Polak–Ribière–Polyak (PRP) [22,23] method and the Liu–Storey (LS) [19] method, in which the CG parameters are specified by

$$\beta_k^{\mathrm{HS}} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, \qquad \beta_k^{\mathrm{PRP}} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}, \qquad \beta_k^{\mathrm{LS}} = \frac{g_k^T y_{k-1}}{-d_{k-1}^T g_{k-1}}, \tag{1.4}$$

where $y_{k-1} = g_k - g_{k-1}$ and $\|.\|$ is the Euclidean norm. These methods have been regarded as the most efficient CG methods and studied extensively, see [7,11,12,14,17, 22–25,29,30,32,33] etc. However, in summary, the convergences of these methods for general nonlinear functions are still uncertain. In [24], Powell designed a 3 dimensional example and showed that, when the function is not strongly convex, the PRP method may not converge, even with an exact line search. Moreover, by Powell's example [24], the HS method may not converge for a general nonlinear function as well, since $\beta_k^{\mathrm{HS}} = \beta_k^{\mathrm{PRP}}$ holds with an exact line search. In [24], for the PRP method, Powell also suggested to restrict the CG parameter to be non-negative, namely, $\beta_k^{\mathrm{PRP+}} = \max\{\beta_k^{\mathrm{PRP}}, 0\}$. Inspired by Powell's work, Gilbert and Nocedal [11] gave an elegant analysis and proved that the PRP+ method is globally convergent when the search direction satisfies the sufficient descent condition $g_k^T d_k \le -c\|g_k\|^2$ and the stepsize α_k is determined by the standard Wolfe line search. This technique can also be used to analyze the global convergence of the HS+ method in which $\beta_k^{\mathrm{HS+}} = \max\{\beta_k^{\mathrm{HS}}, 0\}$.

Recently, based on the standard HS, PRP and LS method, some descent CG methods have been developed, see [4,12,16,29,32,33], etc. The first one is the well-known CG_DESCENT method proposed by Hager and Zhang [12,13], in which the parameter β_k is defined by

$$\beta_k^{\mathrm{HZ}} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} - \theta_k \frac{\|y_{k-1}\|^2 g_k^T d_{k-1}}{(d_{k-1}^T y_{k-1})^2}, \quad \theta_k > \frac{1}{4}. \tag{1.5}$$

An attractive property of this method is that, for any line search, the direction d_k satisfies the sufficient descent condition

$$g_k^T d_k \le -\left(1 - \frac{1}{4\theta_k}\right) \|g_k\|^2. \tag{1.6}$$

Hager and Zhang dynamically adjusted the lower bound on β_k^{HZ} by letting

$$\overline{\beta}_k^{\mathrm{HZ}} = \max\left\{\beta_k^{\mathrm{HZ}}, \eta_k\right\}, \quad \eta_k = \frac{-1}{\|d_{k-1}\| \min\{\eta, \|g_{k-1}\|\}}, \quad \eta > 0,$$

and established the global convergence of their method when the line search fulfills the Wolfe conditions

$$f(x_k + \alpha_k d_k) \le f(x_k) + \delta \alpha_k g_k^T d_k, \tag{1.7}$$

$$g(x_k + \alpha_k d_k)^T d_k \ge \sigma g_k^T d_k, \tag{1.8}$$

where $0 < \delta \le \sigma < 1$.

Similar to the CG_DESCENT method, there are some modified PRP and LS methods, such as the modified LS (MLS) method in [16] and the descent PRP (DPRP) method in [29]. In these two methods, the CG parameters are defined by

$$\beta_k^{\mathrm{MLS}} = -\frac{g_k^T y_{k-1}}{g_{k-1}^T d_{k-1}} - \theta_k \frac{\|y_{k-1}\|^2 g_k^T d_{k-1}}{\left(g_{k-1}^T d_{k-1}\right)^2} \tag{1.9}$$

and

$$\beta_k^{\mathrm{DPRP}} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2} - \theta_k \frac{\|y_{k-1}\|^2 g_k^T d_{k-1}}{\|g_{k-1}\|^4}, \tag{1.10}$$

where $\theta_k > \frac{1}{4}$ is a constant. It is clear that the formulas (1.5), (1.9) and (1.10) have similar structures. More interestedly, both MLS method and DPRP method satisfy the sufficient descent condition (1.6). Based on some conditions, the global convergent results were established.

Based on the secant condition often satisfied by quasi-Newton method, Zhang, Zhou and Li developed sufficient descent three-term PRP and HS methods [32,33]. In the three-term PRP method [33], the search direction is defined by

$$d_k = -g_k + \beta_k^{\mathrm{PRP}} d_{k-1} - \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2} y_{k-1}, \quad d_0 = -g_0. \tag{1.11}$$

Similarly, in the modified HS method [32], they set

$$d_k = -g_k + \beta_k^{\mathrm{HS}} d_{k-1} - \frac{g_k^T d_{k-1}}{d_{k-1}^T y_{k-1}} y_{k-1}, \quad d_0 = -g_0. \tag{1.12}$$

A remarkable feature of these methods is that the sufficient descent condition $g_k^T d_k = -\|g_k\|^2$ will always hold. Zhang et al. [32,33] analyzed the global convergence of their methods with suitable line search.

Among the descent CG methods above, the CG_DESCENT method is the most famous one. When the exact line search is used, the method will reduce to the standard HS method. Interestedly, it can also be seemed as a member of the Dai–Liao [5] CG methods, in which

$$\beta_k^{\mathrm{DL}} = \frac{g_k^T y_{k-1}}{y_{k-1}^T d_{k-1}} - t\frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}}, \quad t \geq 0,$$

where $s_{k-1} = \alpha_{k-1} d_{k-1}$. The CG_DESCENT method can be viewed as an adaptive version of DL method corresponding to

$$t = \theta_k \frac{\|y_{k-1}\|^2}{s_{k-1}^T y_{k-1}}, \quad \theta_k > \frac{1}{4}.$$

Due to the existence of the parameter t, the DL method has been seemed as a family of CG methods and studied extensively, see [1,2,15,28,31], etc. Very recently, to seek an optimal choice of the parameter t in the DL method, Dai and Kou [4] provided the following family of CG methods for unconstrained optimization

$$\beta^{\mathrm{DK}}(\tau_k) = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} - \left(\tau_k + \frac{\|y_{k-1}\|^2}{s_{k-1}^T y_{k-1}} - \frac{s_{k-1}^T y_{k-1}}{\|s_{k-1}\|^2}\right) \frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}},$$

where the parameter τ_k is corresponding to the scaling parameter in the scaled memoryless BFGS method proposed by Perry [21] and Shanno [27]. Dai and Kou [4] established the global convergence of their method with a improved Wolfe line search.

Inspired by the Dai–Kou method and the descent CG methods introduced above, in this paper, we attempt to structure a family of three-term CG methods, in which the search directions are close to the directions of the memoryless BFGS method in [20,26]. The rest of the paper is organized as follows: in the next section, we propose the algorithm. In Sects. 3 and 4, we analyze the convergence of the proposed methods. In the last section, we present some numerical results to test the efficiency of the algorithms.

2 The algorithm

The limited memory BFGS method [18,20] is an adaptation of the BFGS method for large-scale problems. It requires minimal storage and often provides a fast rate of linear convergence. The direction of the self-scaling memoryless BFGS method [21,27] is defined by

$$d_k^{\mathrm{PS}} = -H_k g_k,$$

where

$$H_k = \frac{1}{\tau_k}\left(I - \frac{s_{k-1}y_{k-1}^T + y_{k-1}s_{k-1}^T}{s_{k-1}^T y_{k-1}}\right) + \left(1 + \frac{1}{\tau_k}\frac{\|y_{k-1}\|^2}{s_{k-1}^T y_{k-1}}\right)\frac{s_{k-1}s_{k-1}^T}{s_{k-1}^T y_{k-1}},$$

where I denotes the identity matrix and τ_k is a scaling parameter. Therefore, the direction can be rewritten as

$$d_k^{\mathrm{PS}} = -g_k + \left[\frac{g_k^T y_{k-1}}{s_{k-1}^T y_{k-1}} - \left(\tau_k + \frac{\|y_{k-1}\|^2}{s_{k-1}^T y_{k-1}}\right)\frac{g_k^T d_{k-1}}{d_{k-1}^T y_{k-1}}\right] s_{k-1} + \frac{g_k^T d_{k-1}}{d_{k-1}^T y_{k-1}} y_{k-1}.$$

Dai and Kou [4] derived the new formula by seeking the conjugate gradient direction which is closest to d_k^{PS}.

In d_k^{PS}, if set $\tau_k = 1$, we will get the memoryless BFGS method proposed by Nocedal [20] and Shanno [26], in which

$$d_k^{\mathrm{LBFGS}} = -g_k + \left(\frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} - \frac{\|y_{k-1}\|^2 g_k^T d_{k-1}}{(d_{k-1}^T y_{k-1})^2}\right) d_{k-1} + \frac{g_k^T d_{k-1}}{d_{k-1}^T y_{k-1}}(y_{k-1} - s_{k-1}).$$

We note that the direction d_k^{LBFGS} is in fact a linear combination of g_k, d_{k-1} and y_{k-1}, since $s_{k-1} = \alpha_{k-1} d_{k-1}$. Based one the strategies used to design the the Dai–Kou method [4], we think it is reasonable to replace the term $(y_{k-1} - s_{k-1})$ in d_k^{LBFGS} by $t_k y_{k-1}$ to construct a new search direction. Here, the scalar t_k is a suitable parameter. Combining the similar structure of formulas (1.5), (1.9) and (1.10), we give the following scheme

$$d_k = -g_k + \beta_k^{\mathrm{NEW}} d_{k-1} + \lambda_k y_{k-1}, \qquad d_0 = -g_0, \tag{2.1}$$

$$\beta_k^{\mathrm{NEW}} = \frac{g_k^T y_{k-1}}{z_k} - \frac{\|y_{k-1}\|^2 g_k^T d_{k-1}}{z_k^2}, \tag{2.2}$$

$$\lambda_k = t_k \frac{g_k^T d_{k-1}}{z_k}, \qquad 0 \le t_k \le \bar{t} < 1, \tag{2.3}$$

where z_k is a scalar to be specific, and $\bar{t}$ is a constant to guarantee the sufficient descent property of the new search direction. In practical computation, we set $\bar{t} = 0.3$ and calculate t_k by

$$t_k = \min\left\{\bar{t},\ \max\left\{0,\ 1 - \frac{y_{k-1}^T s_{k-1}}{\|y_{k-1}\|^2}\right\}\right\}. \tag{2.4}$$

In the formula above, the scalar $(1 - \frac{y_{k-1}^T s_{k-1}}{\|y_{k-1}\|^2})$ is the solution of the univariate minimal problem

$$\min \|(y_{k-1} - s_{k-1}) - t y_{k-1}\|^2, \quad t \in \mathbb{R}.$$

We calculate t_k by (2.4) to make the direction defined by (2.1) close to the direction of the memoryless BFGS method.

Lemma 1 *If the sequences* $\{x_k\}$ *and* $\{d_k\}$ *are generated by (1.2) and (2.1), then*

$$g_k^T d_k \le -\left(1 - \frac{(1+\bar{t})^2}{4}\right)\|g_k\|^2. \tag{2.5}$$

Proof Since $d_0 = -g_0$, we have $g_0^T d_0 = -\|g_0\|^2$, which satisfies (2.5) since $0 \le t_k \le \bar{t} < 1$. Multiplying (2.1) by g_k^T, we have

$$\begin{aligned}
g_k^T d_k &= -\|g_k\|^2 + \beta_k^{\text{NEW}} g_k^T d_{k-1} + \lambda_k g_k^T y_{k-1} \\
&= -\|g_k\|^2 + \left(\frac{g_k^T y_{k-1}}{z_k} - \frac{\|y_{k-1}\|^2 g_k^T d_{k-1}}{z_k^2}\right) g_k^T d_{k-1} + t_k \frac{g_k^T d_{k-1}}{z_k} g_k^T y_{k-1} \\
&= -\|g_k\|^2 + (1+t_k)\frac{g_k^T y_{k-1} g_k^T d_{k-1}}{z_k} - \frac{\|y_{k-1}\|^2 \left(g_k^T d_{k-1}\right)^2}{z_k^2} \\
&= -\|g_k\|^2 + 2\left(\frac{1+t_k}{2} g_k^T\right)\left(\frac{y_{k-1} g_k^T d_{k-1}}{z_k}\right) - \frac{\|y_{k-1}\|^2 \left(g_k^T d_{k-1}\right)^2}{z_k^2} \\
&\le -\|g_k\|^2 + \frac{(1+t_k)^2}{4}\|g_k\|^2 + \frac{\|y_{k-1}\|^2 \left(g_k^T d_{k-1}\right)^2}{z_k^2} - \frac{\|y_{k-1}\|^2 \left(g_k^T d_{k-1}\right)^2}{z_k^2} \\
&= -\|g_k\|^2 + \frac{(1+t_k)^2}{4}\|g_k\|^2 \\
&\le -\left(1 - \frac{(1+\bar{t})^2}{4}\right)\|g_k\|^2,
\end{aligned}$$

which completes the proof. □

In the scheme (2.1)–(2.3), different choices for the scalar z_k in β_k^{NEW} will correspond to different modified CG methods. In this paper, we are mainly interested in the following three cases.

- For $z_k = d_{k-1}^T y_{k-1}$, we get a modified HS method and the parameter β_k^{NEW} equal to β^{HZ} with $\theta_k = 1$. If an exact line search is used, this method will reduce to the standard HS method since $g_k^T d_{k-1} = 0$ holds for $k > 0$.
- For $z_k = \|g_{k-1}\|^2$, we get a descent PRP method.
- For $z_k = -d_{k-1}^T g_{k-1}$, we get a modified LS method.

It follows from Lemma 1 that the new methods are sufficient descent. However, how to prove the global convergence is still a problem, especially when the Armijo or the standard Wolfe line search is used. To establish the global convergences of CG methods, a common technique is to use the truncated technique in [11] to restrict the the parameter β_k to be nonnegative. Differently, in this paper, we will use a new technique to set the lower bound for z_k to guarantee the global convergence of the proposed methods.

- We set $z_k = \max\{\tau\|d_{k-1}\|, d_{k-1}^T y_{k-1}\}$ with some constant $\tau > 0$ to get a modified HS method, which we call the THS method.
- We let $z_k = \max\{\tau\|d_{k-1}\|, \|g_{k-1}\|^2\}$ to obtain a modified PRP method which we call the TPRP method.
- We choose $z_k = \max\{\tau\|d_{k-1}\|, -d_{k-1}^T g_{k-1}\}$ to get a LS type method which we call the TLS method.

In these methods, we always have $z_k \geq \tau\|d_{k-1}\|$. This inequality is very helpful to establish the global convergence of the proposed methods. In the rest of this paper, for simplicity, we will only analyze the global convergence of the TPRP method. The conclusion can be extended to other methods in a similar way. Based on the precess above, we present concrete steps of the TPRP method as follows:

Algorithm 2.1 (The TPRP method)

Step 0. Initiate $x_0 \in \mathbb{R}^n$, $\varepsilon > 0$, $\tau > 0$ *and* $0 < \delta \leq \sigma < 1$. *Set* $k := 0$;
Step 1. Stop if $\|g_k\|_\infty \leq \varepsilon$; *Otherwise go to the next step;*
Step 2. Compute d_k *by* (2.1) *with* $z_k = \max\{\tau\|d_{k-1}\|, \|g_{k-1}\|^2\}$;
Step 3. Determine the steplength α_k *by a line search.*
Step 4. Let $x_{k+1} = x_k + \alpha_k d_k$;
Step 5. Set $k := k + 1$ *and go to Step 1.*

It is not difficult to establish the global convergence of the TPRP method when the Wolfe or Armijo line search is used. However, the numerical performance of the TPRP method in practical computation is not as good as we expect. In the new method, we still can not guarantee the scalar β_k^{NEW} to be nonnegative. When $\beta_k^{\text{NEW}} < 0$, we think that the term $\beta_k^{\text{NEW}} d_{k-1}$ in (2.1) will reduce the efficiency of d_k, since d_{k-1} is a sufficient descent direction of f at x_{k-1} which is close to x_k. Therefore, we give a truncated TPRP (TPRP+) method by setting

$$d_k = \begin{cases} -g_k, & \text{if} \quad \beta_k^{\text{NEW}} \leq 0, \\ d_k^{\text{TPRP}}, & \text{else,} \end{cases}$$

where d_k^{TPRP} is the direction defined in the TPRP method. It is clear that the TPRP+ method is sufficient descent and satisfies (2.5). Similarly, we can define THS+ and TLS+ methods by using this truncated technique.

In the next sections, we will establish the global convergence of the TPRP method when the Wolfe or the Armijo line search is used. All the convergence results of the TPRP method can be extended to the TPRP+ method in a similar way. From now on, throughout the paper, we always suppose the following assumption holds.

Assumption A (I) The level set $\Omega = \{x \in \mathbb{R}^n: f(x) \leq f(x_0)$ is bounded.
(II) In some neighborhood N of Ω, function f is continuously differentiable and its gradient is Lipschitz continuous, namely, there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \qquad \forall x, y \in N. \tag{2.6}$$

The assumption implies that there are positive constants B and γ_1 such that

$$\|x\| \le B \qquad \text{and} \qquad \|g(x)\| \le \gamma_1, \qquad \forall x \in \Omega. \tag{2.7}$$

3 Convergence analysis of the TPRP method with Wofle line search

In this section, we devote to the global convergence of the TPRP method when the Wolfe line search is used. The following useful lemma was essentially proved by Zoutendijk [34].

Lemma 2 *Suppose that the conditions in Assumption A hold,* $\{g_k\}$ *and* $\{d_k\}$ *are generated by the TPRP method with the Wolfe line search* (1.7)–(1.8)*, then*

$$\sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < +\infty. \tag{3.1}$$

Theorem 1 *Suppose that the conditions in Assumption A hold and* $\{g_k\}$ *is generated by the TPRP method with the Wolfe line search, then*

$$\lim_{k\to\infty} \|g_k\| = 0. \tag{3.2}$$

Proof It follows from the descent condition (2.5) that $\|d_{k-1}\| \neq 0$ holds for $k > 1$. Since $z_k = \max\{\tau\|d_{k-1}\|, \|g_{k-1}\|^2\}$, we have

$$z_k \ge \tau\|d_{k-1}\| > 0.$$

Combining this with (2.2), (2.6) and (2.7) gives

$$\begin{aligned}
\|d_k\| &\le \|g_k\| + \left|\beta_k^{\text{NEW}}\right| \|d_{k-1}\| + |\lambda_k| \|y_{k-1}\| \\
&\le \|g_k\| + \left| \frac{g_k^T y_{k-1}}{z_k} - \frac{\|y_{k-1}\|^2 g_k^T d_{k-1}}{z_k^2} \right| \|d_{k-1}\| + \left| t_k \frac{g_k^T d_{k-1}}{z_k} \right| \|y_{k-1}\| \\
&\le \|g_k\| + \left(\frac{\|g_k\|\|y_{k-1}\|}{\tau\|d_{k-1}\|} + \frac{\|y_{k-1}\|^2\|g_k\|\|d_{k-1}\|}{\tau^2\|d_{k-1}\|^2} \right) \|d_{k-1}\| \\
&\quad + \frac{t_k\|g_k\|\|d_{k-1}\|}{\tau\|d_{k-1}\|} \|y_{k-1}\| \\
&\le \|g_k\| + \left(\frac{\|g_k\| 2\gamma_1}{\tau\|d_{k-1}\|} + \frac{4\gamma_1^2\|g_k\|}{\tau^2\|d_{k-1}\|} \right) \|d_{k-1}\| + \frac{t_k\|g_k\|}{\tau} 2\gamma_1 \\
&= \left(1 + \frac{2\gamma_1}{\tau} + \frac{4\gamma_1^2}{\tau^2} + \frac{2t_k\gamma_1}{\tau} \right) \|g_k\|
\end{aligned}$$

$$\leq \left(1 + \frac{4\gamma_1}{\tau} + \frac{4\gamma_1^2}{\tau^2}\right) \|g_k\|$$
$$\triangleq C\|g_k\|.$$

Namely,

$$\|d_k\| \leq C\|g_k\|. \tag{3.3}$$

Combining this with (3.1) gives

$$\sum_{k=0}^{\infty} \|g_k\|^2 \leq \sum_{k=0}^{\infty} \frac{C^2\|g_k\|^4}{\|d_k\|^2} < \infty.$$

This implies $\lim_{k\to\infty} \|g_k\| = 0$. The proof is completed. □

4 Convergence analysis of the TPRP method with Armijo line search

In this section, we analyze the global convergence of the TPRP method with the Armijo line search, that is, the steplength satisfies

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta\alpha_k g_k^T d_k, \tag{4.1}$$

where $\alpha_k = \max\{\alpha_0\rho^i, i = 0, 1, 2, \ldots\}$ with $0 < \rho, \delta < 1$, $\alpha_0 \in (0, 1]$ is an initial guess of the steplength. If the conditions in Assumption A hold, it follows directly from (3.3) and (2.7) that

$$\|d_k\| \leq C\|g_k\| \leq C\gamma_1. \tag{4.2}$$

Theorem 2 *Suppose that the conditions in Assumption A hold, $\{g_k\}$ is generated by the TPRP method with the Armijo line search (4.1). Then either $\|g_k\| = 0$ for some k or*

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{4.3}$$

Proof Suppose for contradiction that $\liminf_{k\to\infty} \|g_k\| > 0$ and $\|g_k\| \neq 0$ for all $k \geq 0$. Define

$$\gamma = \inf\{\|g_k\|: k \geq 0\}.$$

Then

$$\|g_k\| \geq \gamma > 0, \quad \forall k \geq 0. \tag{4.4}$$

From (2.5), (4.1) and the conditions in Assumption A, we have

$$\lim_{k\to\infty} \alpha_k \|g_k\|^2 = 0. \tag{4.5}$$

On one hand, if $\liminf_{k\to\infty} \alpha_k > 0$, (4.5) gives $\liminf_{k\to\infty} \|g_k\| = 0$, which contradicts (4.4). On the other hand, if $\liminf_{k\to\infty} \alpha_k = 0$, then there exists a infinite index set $\mathbb{K}$ such that

$$\lim_{k\in\mathbb{K}, k\to\infty} \alpha_k = 0 \tag{4.6}$$

When $k \in \mathbb{K}$ is large enough, the Armijo line search rule implies that $\rho^{-1}\alpha_k$ does not satisfy (4.1), namely

$$f(x_k + \rho^{-1}\alpha_k d_k) - f(x_k) > \delta\rho^{-1}\alpha_k g_k^T d_k \tag{4.7}$$

We get from the mean value theorem and (2.6) that, there is a $\xi_k \in [0, 1]$, such that

$$\begin{aligned} f(x_k + \rho^{-1}\alpha_k d_k) - f(x_k) &= \rho^{-1}\alpha_k g(x_k + \xi_k\rho^{-1}\alpha_k d_k)^T d_k \\ &= \rho^{-1}\alpha_k g_k^T d_k + \rho^{-1}\alpha_k (g(x_k + \xi_k\rho^{-1}\alpha_k d_k) - g_k)^T d_k \\ &\le \rho^{-1}\alpha_k g_k^T d_k + L\rho^{-2}\alpha_k^2 \|d_k\|^2. \end{aligned}$$

This together with (4.7), (4.2) and (2.5) gives

$$\begin{aligned} (1-\delta)\left(1 - \frac{(1+\bar{t})^2}{4}\right)\|g_k\|^2 &\le (\delta - 1) g_k^T d_k \\ &\le L\rho^{-1}\alpha_k \|d_k\|^2 \\ &\le \alpha_k L\rho^{-1} C^2 \gamma_1^2. \end{aligned}$$

Combining this with (4.6) gives $\liminf_{k\in\mathbb{K}, k\to\infty} \|g_k\| = 0$. This also yields contradiction and the proof is completed. □

5 Linear convergence rate

In this section, we analyze the r-linear convergence rate of the TPRP method when the objective function f is twice continuously differentiable and uniformly convex, that is, there are positive constants $m \le M$ such that

$$m\|y\|^2 \le y^T \nabla^2 f(x) y \le M\|y\|^2, \quad \forall x, y \in \mathbb{R}^n, \tag{5.1}$$

where $\nabla^2 f(x)$ denotes the Hessian of f at x.

Lemma 3 *Suppose that f is twice continuously differentiable and uniformly convex. Then the problem (1.1) has a unique solution x^* and*

$$\frac{1}{2}m\|x - x^*\|^2 \le f(x) - f(x^*) \le \frac{1}{2}M\|x - x^*\|^2, \quad \forall x \in \mathbb{R}^n, \tag{5.2}$$

$$m\|x - x^*\| \le \|g(x)\| \le M\|x - x^*\|, \quad \forall x \in \mathbb{R}^n, \tag{5.3}$$

$$m\|x - y\|^2 \le (g(x) - g(y))^T (x - y) \le M\|x - y\|^2, \quad \forall x, y \in \mathbb{R}^n, \tag{5.4}$$

$$\|g(x) - g(y)\| \le M\|x - y\|, \quad \forall x, y \in \mathbb{R}^n. \tag{5.5}$$

Moreover, for any $x_0 \in \mathbb{R}^n$*, the level set* $\Omega_0 \triangleq \{x \in \mathbb{R}^n : f(x) \le f(x_0)\}$ *is bounded and there is a constant* $B_0 > 0$ *such that*

$$\|x\| \le B_0, \quad \forall x \in \Omega_0. \tag{5.6}$$

Based on the assumptions above, it is no difficult to prove the following convergence theorem.

Theorem 3 *Suppose that* f *is twice continuously differentiable and uniformly convex. If* $\{x_k\}$ *is generated by the TPRP method with the Wolfe or the Armijo line search, then this sequence converges to the unique solution of problem* (1.1).

In order to prove the r-linear convergence of the TPRP method, we first give the following lemma, which gives a lower bound of the stepsize α_k.

Lemma 4 *Suppose that* f *is twice continuously differentiable and uniformly convex, the sequence* $\{x_k\}$ *is generated by the TPRP method with the Wolfe or the Armijo line search. Then there is a constant* $C_1 > 0$ *such that*

$$\alpha_k \ge C_1, \quad \forall k > 0. \tag{5.7}$$

Proof Since $z_k = \max\{\tau\|d_{k-1}\|, \|g_{k-1}\|^2\}$, we get from (2.1), (2.2), (2.3), (5.5), (5.6) that

$$\begin{aligned}
\|d_k\| &\le \|g_k\| + \left|\beta_k^{\text{NEW}}\right| \|d_{k-1}\| + |\lambda_k|\|y_{k-1}\| \\
&\le \|g_k\| + \left|\frac{g_k^T y_{k-1}}{z_k} - \frac{\|y_{k-1}\|^2 g_k^T d_{k-1}}{z_k^2}\right| \|d_{k-1}\| + \left|t_k \frac{g_k^T d_{k-1}}{z_k}\right| \|y_{k-1}\| \\
&\le \|g_k\| + \left(\frac{\|g_k\|\|y_{k-1}\|}{\tau\|d_{k-1}\|} + \frac{\|y_{k-1}\|^2\|g_k\|\|d_{k-1}\|}{\tau^2\|d_{k-1}\|^2}\right)\|d_{k-1}\| \\
&\quad + \frac{t_k\|g_k\|\|d_{k-1}\|}{\tau\|d_{k-1}\|}\|y_{k-1}\| \\
&\le \|g_k\| + \left(\frac{\|g_k\| 2MB_0}{\tau\|d_{k-1}\|} + \frac{4M^2B_0^2\|g_k\|}{\tau^2\|d_{k-1}\|}\right)\|d_{k-1}\| + \frac{t_k\|g_k\|}{\tau} 2MB_0 \\
&\le \left(1 + \frac{2MB_0}{\tau} + \frac{4M^2B_0^2}{\tau^2} + \frac{t_k 2MB_0}{\tau}\right)\|g_k\| \\
&\le \left(1 + \frac{4MB_0}{\tau} + \frac{4M^2B_0^2}{\tau^2}\right)\|g_k\|.
\end{aligned}$$

Therefore

$$\frac{\|g_k\|^2}{\|d_k\|^2} \ge \left(1 + \frac{4MB_0}{\tau} + \frac{4M^2B_0^2}{\tau^2}\right)^{-2} \triangleq C_2. \tag{5.8}$$

When the Wolfe line search is used, we get from (2.5), (1.8) and (5.5) that

$$\frac{(1-\sigma)\left(4-(1+\bar{t})^2\right)}{4}\|g_k\|^2 \le (\sigma-1)g_k^T d_k \le (g_{k+1}-g_k)^T d_k \le M\alpha_k\|d_k\|^2.$$

By (5.8),

$$\alpha_k \ge \frac{(1-\sigma)\left(4-(1+\bar{t})^2\right)\|g_k\|^2}{4M\|d_k\|^2} \ge \frac{(1-\sigma)\left(4-(1+\bar{t})^2\right)C_2}{4M}$$

When the Armijo line search is used. By the line search rule, $\alpha_k \ne \alpha_0$ implies $\rho^{-1}\alpha_k$ dose not satisfy (4.1). Similar to the proof of Theorem 2, we can prove that

$$\alpha_k \ge \frac{\rho(1-\delta)(4-(1+\bar{t})^2)}{4M}\frac{\|g_k\|^2}{\|d_k\|^2} \ge \frac{\rho(1-\delta)(4-(1+\bar{t})^2)C_2}{4M}$$

So we can get (5.7) by setting

$$C_1 = \min\left\{\frac{(1-\sigma)(4-(1+\bar{t})^2)C_2}{4M},\quad \frac{\rho(1-\delta)(4-(1+\bar{t})^2)C_2}{4M}\right\}.$$

The proof is completed. □

Theorem 4 *Suppose that f is twice continuously differentiable and uniformly convex, x^* is the unique solution of problem (1.1) and the sequence $\{x_k\}$ is generated by the TPRP method with the Wolfe or the Armijo line search. Then there are constants $a > 0$ and $r \in (0, 1)$ such that*

$$\|x_k - x^*\| \le ar^k. \tag{5.9}$$

Proof From the Wolf condition (1.7) or the Armijo condition (4.1), we have

$$\begin{aligned} f(x_{k+1}) - f(x^*) &\le f(x_k) - f(x^*) + \delta\alpha_k g_k^T d_k \\ &\le f(x_k) - f(x^*) - \delta\alpha_k\left(1 - \frac{(1+\bar{t})^2}{4}\right)\|g_k\|^2 \\ &\le f(x_k) - f(x^*) - \delta C_1\left(1 - \frac{(1+\bar{t})^2}{4}\right)m^2\|x_k - x^*\|^2 \\ &\le f(x_k) - f(x^*) - \frac{2\delta C_1 m^2}{M}\left(1 - \frac{(1+\bar{t})^2}{4}\right)(f(x_k) - f(x^*)) \\ &= \left[1 - \frac{2\delta C_1 m^2}{M}\left(1 - \frac{(1+\bar{t})^2}{4}\right)\right](f(x_k) - f(x^*)). \end{aligned}$$

Then we get,

$$f(x_k) - f(x^*) \le r^2\left(f(x_{k-1}) - f(x^*)\right) \le \cdots \le r^{2k}\left(f(x_0) - f(x^*)\right),$$

where

$$r = \left[1 - \frac{2\delta C_1 m^2}{M}\left(1 - \frac{(1+\bar{t})^2}{4}\right)\right]^{1/2} \in (0, 1).$$

Combining this with (5.2) gives

$$\|x_k - x^*\|^2 \leq \frac{2}{m}\left(f(x_k) - f(x^*)\right) \leq \frac{2}{m}\left(f(x_0) - f(x^*)\right) r^{2k}.$$

Hence we can obtain (5.9) by letting $a = \sqrt{2\left(f(x_0) - f(x^*)\right)/m}$. The proof is completed. □

6 Numerical results

In this section, we report some numerical results of the proposed methods and compare their numerical performances with that of the CG_DESCENT method [12]. The test problems are the unconstrained problems in the CUTEr library [3] with dimensions varying from 50 to 1000. We excluded the problems for which different solvers converge to different local minimizers. We often ran two versions of the test problem for which the dimension could be chosen arbitrarily. Table 1 lists the names (Prob) and dimensions (Dim) of the 152 valid test problems.

All the methods were coded in Fortran and ran on a PC with 3.7 GHz CPU processor and 4 GB RAM. The code for our methods are modifications of the subroutine of CG_DESCENT method. We terminated the iteration when $\|g_k\|_\infty \leq 10^{-6}$. Detailed numerical results are omitted here since the data is too much.

Figures 1, 2, 3 and 4 plot the performances of the methods relative to the total number of iterations and the CPU time by using the profiles of Dolan and Moré [8]. The curves in the figures have the following meanings:

- "CG_DESCENT" stands for the CG_DESCENT method with the approximate Wolfe line search [12]. We use the Fortran code (Version 1.4, November 14, 2005) from Prof. Hager's web page: http://www.math.ufl.edu/~hager/ and the default parameters there.
- "TPRP" stands for the TPRP method with the same line search as "CG_DESCENT". We set $\tau = 10^{-6}$ for the scalar z_k in β_k^{NEW} and the parameter t_k in (2.3) is calculated (2.4) with $\bar{t} = 0.3$.
- "TPRP+", "THS","THS+","TLS","TLS+" denote the TPRP+, THS, THS+, TLS and TLS+ methods with the same line search and parameters as "TPRP", respectively.

We observe from Figs. 1, 2, 3 and 4 that the performances of the TPRP+, THS+ and TLS+ methods are close and obviously better than that of the CG_DESCENT method. Since all the methods are implemented with the same line search, we conclude that, on average, the TPRP+, THS+ and TLS+ methods seem to generate more efficient search directions. We also observe that, the performance of the TPRP+, THS+ and TLS+

Table 1 The problems and their dimensions

No.	Prob	Dim	No.	Prob	Dim	No.	Prob	Dim
1	ARGLINA	100	52	DIXMAANK	300	103	NONDQUAR	500
2	ARGLINA	200	53	DIXMAANK	1500	104	NONDQUAR	5000
3	ARGLINB	100	54	DIXMAANL	300	105	NONMSQRT	100
4	ARGLINB	200	55	DIXMAANL	1500	106	OSCIPATH	100
5	ARGLINC	100	56	DIXON3DQ	1000	107	OSCIPATH	500
6	ARGLINC	200	57	DIXON3DQ	10000	108	PENALTY1	500
7	ARWHEAD	100	58	DQDRTIC	500	109	PENALTY1	1000
8	ARWHEAD	1000	59	DQDRTIC	1000	110	PENALTY2	50
9	BDQRTIC	100	60	DQRTIC	100	111	PENALTY2	100
10	BDQRTIC	1000	61	EDENSCH	2000	112	PENALTY3	50
11	BDQRTIC	5000	62	EG2	1000	113	PENALTY3	100
12	BOX	100	63	ENGVAL1	100	114	POWELLSG	1000
13	BOX	1000	64	ENGVAL1	5000	115	POWELLSG	10000
14	BROWNAL	100	65	ERRINROS	50	116	POWER	500
15	BROWNAL	200	66	EXTROSNB	100	117	POWER	1000
16	BROYDN7D	5000	67	EXTROSNB	1000	118	QUARTC	500
17	BROYDN7D	10000	68	FLETCBV2	5000	119	QUARTC	10000
18	BRYBND	1000	69	FLETCBV2	10000	120	SCHMVETT	500
19	BRYBND	10000	70	FLETCBV3	100	121	SCOSINE	100
20	CHNROSNB	50	71	FLETCHBV	100	122	SCURLY10	100
21	COSINE	1000	72	FLETCHCR	100	123	SCURLY20	100
22	COSINE	10000	73	FLETCHCR	1000	124	SCURLY30	100
23	CRAGGLVY	1000	74	FMINSRF2	1024	125	SENSORS	100
24	CRAGGLVY	5000	75	FMINSRF2	5625	126	SINQUAD	1000
25	CURLY10	100	76	FMINSURF	5625	127	SINQUAD	10000
26	CURLY10	1000	77	FREUROTH	100	128	SPARSINE	50
27	CURLY20	100	78	FREUROTH	5000	129	SPARSINE	100
28	CURLY20	1000	79	GENHUMPS	100	130	SPARSQUR	1000
29	CURLY30	100	80	GENHUMPS	5000	131	SPARSQUR	5000
30	CURLY30	1000	81	GENROSE	100	132	SPMSRTLS	4999
31	DECONVU	61	82	GENROSE	500	133	SPMSRTLS	10000
32	DIXMAANA	3000	83	HILBERTB	50	134	SROSENBR	5000
33	DIXMAANA	9000	84	HYDC20LS	99	135	SROSENBR	10000
34	DIXMAANB	3000	85	INDEF	1000	136	TESTQUAD	1000
35	DIXMAANB	9000	86	INDEF	5000	137	TESTQUAD	5000
36	DIXMAANC	90	87	LIARWHD	5000	138	TOINTGOR	50
37	DIXMAANC	9000	88	LIARWHD	10000	139	TOINTGSS	1000
38	DIXMAAND	3000	89	MANCINO	50	140	TOINTGSS	5000
39	DIXMAAND	9000	90	MANCINO	100	141	TOINTPSP	50
40	DIXMAANE	300	91	MODBEALE	200	142	TOINTQOR	50

Table 1 continued

No.	Prob	Dim	No.	Prob	Dim	No.	Prob	Dim
41	DIXMAANE	1500	92	MOREBV	100	143	TQUARTIC	1000
42	DIXMAANF	3000	93	MOREBV	500	144	TQUARTIC	5000
43	DIXMAANF	9000	94	MSQRTALS	100	145	TRIDIA	100
44	DIXMAANG	1500	95	MSQRTALS	1024	146	TRIDIA	10000
45	DIXMAANG	9000	96	MSQRTBLS	100	147	VARDIM	100
46	DIXMAANH	3000	97	MSQRTBLS	1024	148	VARDIM	200
47	DIXMAANH	9000	98	NONCVXU2	5000	149	VAREIGVL	1000
48	DIXMAANI	300	99	NONCVXU2	10000	150	VAREIGVL	5000
49	DIXMAANI	1500	100	NONCVXUN	100	151	WOODS	4000
50	DIXMAANJ	300	101	NONDIA	5000	152	WOODS	10000
51	DIXMAANJ	1500	102	NONDIA	10000			

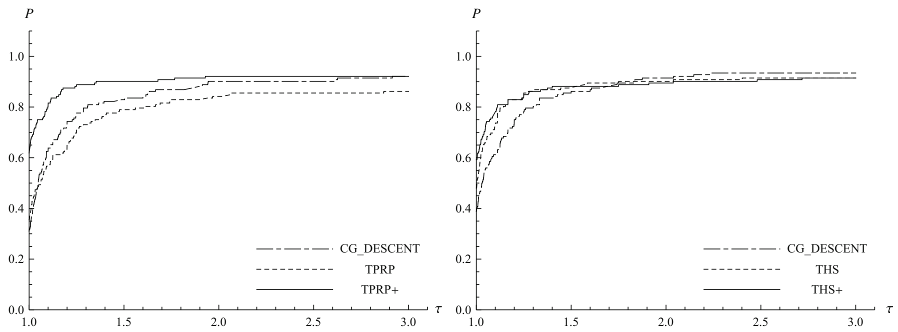

Fig. 1 Performance profiles relative to the total number of iterations (a)

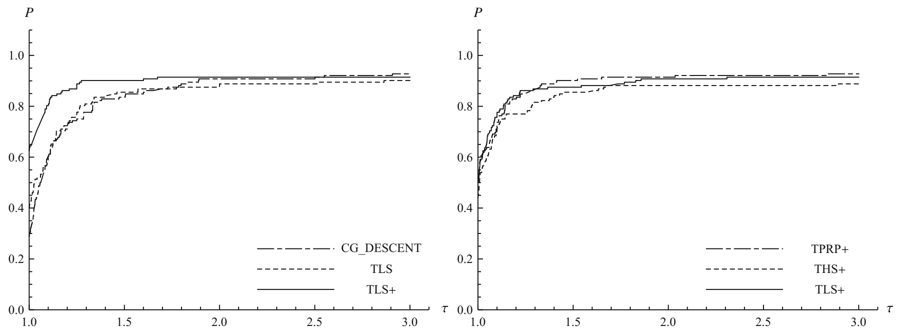

Fig. 2 Performance profiles relative to the total number of iterations (b)

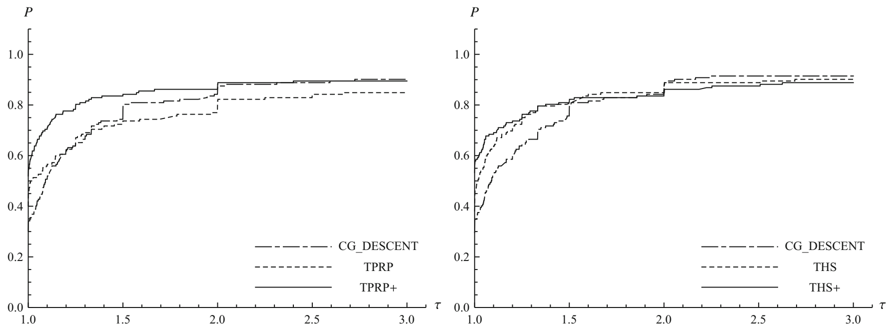

Fig. 3 Performance profiles relative to the CPU time (a)

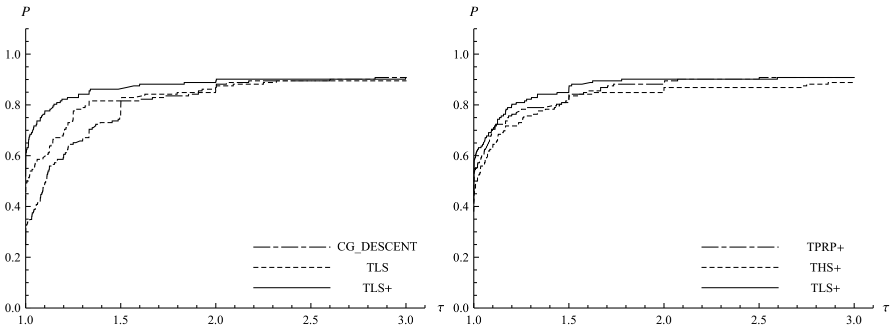

Fig. 4 Performance profiles relative to the CPU time (b)

methods are better than that of the TPRP, THS and TLS methods, correspondingly. This means that the nonnegative restriction on the CG parameter β_k^{NEW} is benefit to improve the efficiency of the methods.

Acknowledgements This work is supported by the Chinese NSF Grant (No. 11401242) and the SFED Grant (No. 14B139) of Hunan Province.

References

1. Andrei, N.: Open problems in nonlinear conjugate gradient algorithms for unconstrained optimization. Bull. Malays. Math. Sci. Soc. **34**, 319–330 (2011)
2. Babaie-Kafaki, S., Reza, G.: A descent family of Dai–Liao conjugate gradient methods. Optim. Methods Softw. **21**, 1–9 (2013)
3. Bongartz, I., Conn, A., Gould, N., Toint, P.: CUTE: constrained and unconstrained testing environments. ACM Trans. Math. Softw. **21**, 123–160 (1995)
4. Dai, Y., Kou, C.: A nonlinear conjugate gradient algorithm with an optimal property and an improved Wolfe line search. SIAM J. Optim. **23**, 296–320 (2013)
5. Dai, Y., Liao, L.: New conjugate conditions and related nonlinear conjugate gradient methods. Appl. Math. Optim. **43**, 87–101 (2001)
6. Dai, Y., Yuan, Y.: A nonlinear conjugate gradient method with a strong global convergence property. SIAM J. Optim. **10**, 177–182 (2000)
7. Dai, Z.: Two modified HS type conjugate gradient methods for unconstrained optimization problems. Nonlinear Anal. **74**, 927–936 (2011)

8. Dolan, E., Moré, J.: Benchmarking optimization software with performance profiles. Math. Program. **91**, 201–213 (2002)
9. Fletcher, R.: Practical Method of Optimization, Vol. 1: Unconstrained Optimization. Wiley, New York (1987)
10. Fletcher, R., Reeves, C.: Function minimization by conjugate gradients. Comput. J. **7**, 149–154 (1964)
11. Gilbert, J., Nocedal, J.: Global convergence properties of conjugate gradient methods for optimization. SIAM J. Optim. **2**, 21–42 (1992)
12. Hager, W., Zhang, H.: A new conjugate gradient method with guaranteed descent and an efficient line search. SIAM J. Optim. **16**, 170–192 (2005)
13. Hager, W., Zhang, H.: A survey of nonlinear conjugate gradient methods. Pac. J. Optim. **2**, 35–58 (2006)
14. Hestenes, M., Stiefel, E.: Methods of conjugate gradients for solving linear systems. J. Res. Natl. Bur. Stand. **49**, 409–436 (1952)
15. Li, G., Tang, C., Wei, Z.: New conjugacy condition and related new conjugate gradient methods for unconstrained optimization. J. Comput. Appl. Math. **202**, 523–539 (2007)
16. Li, M., Feng, H.: A sufficient descent LS conjugate gradient method for unconstrained optimization problems. Appl. Math. Comput. **218**, 1577–1586 (2011)
17. Li, M., Liu, J., Feng, H.: The global convergence of a descent PRP conjugate gradient method. Comput. Appl. Math. **31**(1), 59–83 (2012)
18. Liu, D., Nocedal, J.: On the limited memory BFGS method for large-scale optimization. Math. Program. **45**, 503–528 (1989)
19. Liu, Y., Storey, C.: Efficient generalized conjugate gradient algorithms, part 1: theory. J. Optim. Theory Appl. **69**, 177–182 (1991)
20. Nocedal, J.: Updating quasi-Newton matrices with limited storage. Math. Comput. **35**, 773–782 (1980)
21. Perry, J.M.: A class of conjugate gradient algorithms with a two-step variable-metric memory. Discussion Paper 269, Center for Mathematical Studies in Economics and Management Sciences, Northwestern University, Evanston, Illinois (1977)
22. Polak, B., Ribière, G.: Note sur la convergence de méthodes de directions conjuguées. Rev. Francaise Inform. Recherche. Opérat. **16**, 35–43 (1969)
23. Polyak, B.: The conjugate gradient method in extreme problems. USSR Comput. Math. Math. Phys. **9**, 94–112 (1969)
24. Powell, M.: Nonvonvex minimization calculations and the conjugate gradient method. In: Lecture Notes in Mathematics, vol. 1066. Springer, Berlin (1984)
25. Powell, M.: Convergence properties of algorithms for nonlinear optimization. SIAM Rev. **28**, 487–500 (1986)
26. Shanno, D.: Conjugate gradient methods with inexact searches. Math. Oper. Res. **3**, 244–256 (1978)
27. Shanno, D.F.: On the convergence of a new conjugate gradient algorithm. SIAM J. Numer. Anal. **15**, 1247–1257 (1978)
28. Yabe, H., Takano, M.: Global convergence properties of nonlinear conjugate gradient methods with modified secant condition. Comput. Optim. Appl. **28**, 203–225 (2004)
29. Yu, G., Guan, L.: Modified PRP methods with sufficient descent property and their convergence properties. Acta Scientiarum Naturalium Universitatis Sunyatseni **45**(4), 11–14 (2006). **(Chinese)**
30. Yuan, G.: Modified nonlinear conjugate gradient methods with sufficient descent property for large-scale optimization problems. Optim. Lett. **3**, 11–21 (2009)
31. Zhang, J., Deng, N., Chen, L.: New quasi-Newton equation and related methods for unconstrained optimization. J. Optim. Theory Appl. **102**, 147–167 (1999)
32. Zhang, L.: New versions of the Hestenes–Stiefel nonlinear conjugate gradient method based on the secant condition for optimization. Comput. Appl. Math. **28**, 1–23 (2009)
33. Zhang, L., Zhou, W., Li, D.: A descent modified Polak–Ribière–Polyak conjugate gradient method and its global convergence. IMA J. Numer. Anal. **26**, 629–640 (2006)
34. Zoutendijk, G.: Nonlinear programming, computational methods. In: Abadie, J. (ed.) Integer and Nonlinear Programming, pp. 37–86. North-Holland, Amsterdam (1970)