Sun and Liu *Journal of Inequalities and Applications* (2015) 2015:125
DOI 10.1186/s13660-015-0649-9

Journal of Inequalities and Applications
a SpringerOpen Journal

RESEARCH

Open Access

# Three modified Polak-Ribière-Polyak conjugate gradient methods with sufficient descent property

Min Sun[1*] and Jing Liu[2]

*Correspondence:
ziyouxiaodou@163.com
[1]School of Mathematics and Statistics, Zaozhuang University, Beian Street, Zaozhuang, 277160, P.R. China
Full list of author information is available at the end of the article

**Abstract**
In this paper, three modified Polak-Ribière-Polyak (PRP) conjugate gradient methods for unconstrained optimization are proposed. They are based on the two-term PRP method proposed by Cheng (Numer. Funct. Anal. Optim. 28:1217-1230, 2007), the three-term PRP method proposed by Zhang *et al.* (IMA J. Numer. Anal. 26:629-640, 2006), and the descent PRP method proposed by Yu *et al.* (Optim. Methods Softw. 23:275-293, 2008). These modified methods possess the sufficient descent property without any line searches. Moreover, if the exact line search is used, they reduce to the classical PRP method. Under standard assumptions, we show that these three methods converge globally with a Wolfe line search. We also report some numerical results to show the efficiency of the proposed methods.

**Keywords:** conjugate gradient method; sufficient descent property; global convergence

## 1 Introduction

Consider the unconstrained optimization problem:

$$\min f(x), x \in \mathcal{R}^n, \tag{1}$$

where $f : \mathcal{R}^n \to \mathcal{R}$ is continuously differentiable, and its gradient $g(x)$ is available. Conjugate gradient methods are efficient for solving (1), especially for large-scale problems. A conjugate gradient method generates an iterate sequence $\{x_k\}$ by

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, \ldots, \tag{2}$$

where $x_k$ is the current iterate, $\alpha_k > 0$ is the step size and computed by certain line search, and $d_k$ is the search direction defined by

$$d_k = \begin{cases} -g_k, & \text{if } k = 0, \\ -g_k + \beta_k d_{k-1}, & \text{if } k \geq 1, \end{cases} \tag{3}$$

in which $\beta_k$ is an important parameter. Generally, different conjugate gradient methods correspond to different choices of the parameter $\beta_k$. Some well-known formulas for $\beta_k$

include the Fletcher-Reeves (FR) [1], the Polak-Ribière-Polyak (PRP) [2, 3], the Liu-Storey (LS) [4], the Dai-Yuan (DY) [5], the Hestenes-Stiefel (HS) [6] and the conjugate descent (CD) [7] formulas. In this paper, we focus our attention on the PRP method, in which the parameter $\beta_k$ is given by

$$\beta_k^{\mathrm{PRP}} = \frac{g_k^\top (g_k - g_{k-1})}{\|g_{k-1}\|^2}, \tag{4}$$

where $\|\cdot\|$ is the 2-norm. In the convergence analysis and implementations of conjugate gradient methods, one often requires the line search to be an inexact line search such as a Wolfe line search, a strong Wolfe line search or an Armijo line search. The Wolfe line search is finding a step size $\alpha_k$ satisfying

$$\begin{cases} f(x_k + \alpha_k d_k) \le f(x_k) + \rho \alpha_k g_k^\top d_k, \\ g(x_k + \alpha_k d_k)^\top d_k \ge \sigma g_k^\top d_k, \end{cases} \tag{5}$$

where $0 < \rho < \sigma < 1$. The strong Wolfe line search is computing $\alpha_k$ such that

$$\begin{cases} f(x_k + \alpha_k d_k) \le f(x_k) + \rho \alpha_k g_k^\top d_k, \\ |g(x_k + \alpha_k d_k)^\top d_k| \le \sigma |g_k^\top d_k|, \end{cases} \tag{6}$$

where $0 < \rho < 1/2$ and $\sigma \in (\rho, 1)$. The Armijo line search is finding a step size $\alpha_k = \max\{\rho^j | j = 0, 1, \ldots\}$ satisfying

$$f(x_k + \alpha_k d_k) \le f(x_k) + \delta \alpha_k g_k^\top d_k, \tag{7}$$

where $\delta \in (0,1)$ and $\rho \in (0,1)$ are two constants.

The PRP method is generally regarded to be one of the most efficient conjugate gradient methods and has been studied by many researchers [2, 3, 8]. Polak and Ribière [2] proved that the PRP method with the exact line search is globally convergent under a strong convexity assumption for the objective function $f$. Gilbert and Nocedal [3] conducted an elegant analysis and showed that the PRP method is globally convergent if $\beta_k^{\mathrm{PRP}}$ is restricted to be non-negative (denoted $\beta_k^{\mathrm{PRP+}}$) and $\alpha_k$ is determined by a line search step satisfying the sufficient descent condition

$$g_k^\top d_k \le -c\|g_k\|^2, \quad c > 0, \tag{8}$$

in addition to the Wolfe line search condition (5). Grippo and Lucidi [8] proposed new line search conditions, which can ensure that the PRP method is globally convergent for nonconvex problems. However, the method given by Grippo and Lucidi [8] does not perform better than the PRP method, which employs $\beta_k^{\mathrm{PRP+}}$ and the Wolfe line search in the numerical computations. Therefore, great attention is given to the problem of finding the methods which not only have global convergence but also have nice numerical performance [9–16].

Recently, two new conjugate gradient methods, obtained by modifying the PRP method, called two-term PRP method (denoted CTPRP) and three-term PRP method (denoted

ZTPRP), have been proposed by Cheng [9] and Zhang *et al.* [10], respectively, in which the direction $d_k$ is given by

$$d_k^{\mathrm{CTPRP}} = -\left(1 + \beta_k^{\mathrm{PRP}} \frac{g_k^\top d_{k-1}}{\|g_k\|^2}\right) g_k + \beta_k^{\mathrm{PRP}} d_{k-1}, \quad \forall k \geq 1$$

or

$$d_k^{\mathrm{ZTPRP}} = -g_k + \beta_k^{\mathrm{PRP}} d_{k-1} - \theta_k y_{k-1}, \quad \forall k \geq 1,$$

where

$$y_{k-1} = g_k - g_{k-1}, \qquad \theta_k = \frac{g_k^\top d_{k-1}}{\|g_{k-1}\|^2}.$$

An attractive feature of the CTPRP method and the ZTPRP method is that they satisfy $g_k^\top d_k = -\|g_k\|^2$, which is independent of line search used. Moreover, these two methods are globally convergent if some kind of modified Armijo type line search or strong Wolfe line search is used, and the presented numerical results in [9, 10] show some potential advantages of the proposed methods. Moreover, Yu *et al.* [11] proposed another type variation of PRP method, denoted YTPRP, whose direction is defined by

$$d_k^{\mathrm{YTPRP}} = -g_k + \beta_k^{\mathrm{YPRP}} d_{k-1}, \quad \forall k \geq 1,$$

where

$$\beta_k^{\mathrm{YPRP}} = \beta_k^{\mathrm{PRP}} - C \frac{\|y_{k-1}\|^2 g_k^\top d_{k-1}}{\|g_{k-1}\|^4} \quad \text{and} \quad C > \frac{1}{4}.$$

An attractive feature of the $d_k^{\mathrm{YTPRP}}$ is that it satisfies $g_k^\top d_k \leq -(1 - 1/4C)\|g_k\|^2$, which is also independent of line search used.

Note that the global convergence of the above three methods is established under some Armijo type line search or strong Wolfe line search. It is well known that the step size generated by the Armijo line search maybe approaches zero, and thus the reduction of the objective function is very little. This slows down the optimization process. Obviously, the strong Wolfe line search can avoid this phenomenon when the parameter $\sigma \to 0^+$, and in this case, the strong Wolfe line search is close to the exact line search. Thus, the computational load of the strong Wolfe line search increases heavily. In fact, the Wolfe line search can also avoid the above phenomenon. However, compared with the strong Wolfe line search, the Wolfe line search needs less computation to get a suitable step size at each iteration. Therefore, the Wolfe line search can enhance the efficiency of the conjugate gradient method.

In this paper, we shall investigate some variations of PRP method under a Wolfe line search. In fact, we take a little modification to the $\beta_k^{\mathrm{PRP}}$ and propose three modified PRP methods based on the iterate directions $d_k^{\mathrm{CTPRP}}$, $d_k^{\mathrm{ZTPRP}}$, and $d_k^{\mathrm{YTPRP}}$, which possess not only the sufficient descent property for any line search but also global convergence with a Wolfe line search. In order to do so, the remainder of the paper is organized as follows:

In Section 2, we propose the modified PRP methods and prove their convergence. In Section 3, we present some numerical results by using the test problems in [17]. Section 4 concludes the paper with final remarks.

## 2 Three modified PRP methods

First, we give the following basic assumption as regards the objection function $f(x)$.

**Assumptions**

(H1) The level set $R_0 = \{x | f(x) \leq f(x_0)\}$ is bounded.

(H2) In some neighborhood $N$ of $R_0$, the gradient $g(x)$ is Lipschitz continuous on an open convex set $B$ that contains $R_0$, *i.e.*, there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \quad \text{for any } x, y \in B.$$

Assumptions (H1) and (H2) imply that there exist positive constants $\gamma$ and $B$ such that

$$\|g(x)\| \leq \gamma, \quad \forall x \in R_0 \tag{9}$$

and

$$\|x - y\| \leq B, \quad \forall x, y \in R_0. \tag{10}$$

Recently, Wei *et al.* [18] proposed a variation of the FR method which we call the VFR method, in which the parameter $\beta_k$ is defined by

$$\beta_k^{\mathrm{VFR}} = \frac{\mu_1 \|g_k\|^2}{\mu_2 |g_k^\top d_{k-1}| + \mu_3 \|g_{k-1}\|^2},$$

where $\mu_1 \in (0, +\infty)$, $\mu_2 \in (\mu_1 + \epsilon_1, +\infty)$, $\mu_3 \in (0, +\infty)$, and $\epsilon_1$ is any given positive constant. An attractive feature of the VFR method is that the sufficient descent condition $g_k^\top d_k \leq -(1 - \frac{\mu_1}{\mu_2})\|g_k\|^2$ always holds which is independent of the line search used. The idea of Wei *et al.* [18] was further extended to the Wei-Yao-Liu method by Dai and Wen [19]. Here, motivated by the ideas of Wei *et al.* [18] and Dai and Wen [19], we construct two modified PRP methods, in which the parameter $\beta_k$ is specified as follows:

$$\beta_k^{\mathrm{MPRP}} = \frac{g_k^\top (g_k - g_{k-1})}{\mu |g_k^\top d_{k-1}| + \|g_{k-1}\|^2} \tag{11}$$

or

$$\beta_k^{\mathrm{MPRP+}} = \max\left\{\frac{g_k^\top (g_k - g_{k-1})}{\mu |g_k^\top d_{k-1}| + \|g_{k-1}\|^2}, 0\right\}, \tag{12}$$

where $\mu \geq 0$ is a constant. Obviously, if $\mu = 0$ or the line search is exact, the new parameter $\beta_k^{\mathrm{MPRP}}$ or $\beta_k^{\mathrm{MPRP+}}$ reduces to the classical parameter $\beta_k^{\mathrm{PRP}}$ in [2] or $\beta_k^{\mathrm{PRP+}}$ in [3].

First, using the parameter $\beta_k^{\mathrm{MPRP}}$ and the direction $d_k^{\mathrm{CTPRP}}$, we present the following conjugate gradient method (denoted the TMPRP1 method).

**TMPRP1 method** (Two-term modified PRP method)

Step 0. Give an initial point $x_0 \in \mathcal{R}^n$, $\mu \geq 0$, $0 < \rho < \sigma < 1$, and set $d_0 = -g_0$, $k := 0$.
Step 1. If $\|g_k\| = 0$ then stop; otherwise go to Step 2.
Step 2. Compute $d_k$ by

$$d_k = \begin{cases} -g_k, & \text{if } k = 0, \\ -(1 + \beta_k^{\mathrm{MPRP}} \frac{g_k^\top d_{k-1}}{\|g_k\|^2}) g_k + \beta_k^{\mathrm{MPRP}} d_{k-1}, & \text{if } k \geq 1. \end{cases} \tag{13}$$

Determine the step size $\alpha_k$ by Wolfe line search (5).
Step 3. Set $x_{k+1} = x_k + \alpha_k d_k$, and $k := k + 1$; go to Step 1.

Similarly, using the parameter $\beta_k^{\mathrm{MPRP}}$ and the direction $d_k^{\mathrm{ZTPRP}}$, we present the following conjugate gradient method (denoted the TMPRP2 method).

**TMPRP2 method** (Three-term modified PRP method)

Step 0. Give an initial point $x_0 \in \mathcal{R}^n$, $\mu \geq 0$, $0 < \rho < \sigma < 1$, and set $d_0 = -g_0$, $k := 0$.
Step 1. If $\|g_k\| = 0$ then stop; otherwise go to Step 2.
Step 2. Compute $d_k$ by

$$d_k = \begin{cases} -g_k, & \text{if } k = 0, \\ -g_k + \beta_k^{\mathrm{MPRP}} d_{k-1} - \vartheta_k y_{k-1}, & \text{if } k \geq 1, \end{cases} \tag{14}$$

where $\vartheta_k = g_k^\top d_{k-1} / (\|g_{k-1}\|^2 + \mu |g_k^\top d_{k-1}|)$. Determine the step size $\alpha_k$ by Wolfe line search (5).
Step 3. Set $x_{k+1} = x_k + \alpha_k d_k$, and $k := k + 1$; go to Step 1.

Using a parameter similar to $\beta_k^{\mathrm{YPRP}}$, we present the following conjugate gradient method (denoted the TMPRP3 method).

**TMPRP3 method** (Three-term descent PRP method)

Step 0. Give an initial point $x_0 \in \mathcal{R}^n$, $\mu \geq 0$, $t > 1$, $0 < \rho < \sigma < 1$, and set $d_0 = -g_0$, $k := 0$.
Step 1. If $\|g_k\| = 0$ then stop; otherwise go to Step 2.
Step 2. Compute $d_k$ by

$$d_k = \begin{cases} -g_k, & \text{if } k = 0, \\ -g_k + \beta_k^{\mathrm{VPRP}} d_{k-1} + \nu_k (y_{k-1} - s_{k-1}), & \text{if } k \geq 1, \end{cases} \tag{15}$$

where

$$\begin{aligned} \beta_k^{\mathrm{VPRP}} &= \frac{g_k^\top (g_k - g_{k-1})}{\mu |g_k^\top d_{k-1}| + \|g_{k-1}\|^2} - t \frac{\|y_{k-1}\|^2 g_k^\top d_{k-1}}{(\mu |g_k^\top d_{k-1}| + \|g_{k-1}\|^2)^2}, \\ \nu_k &= \frac{g_k^\top d_{k-1}}{\mu |g_k^\top d_{k-1}| + \|g_{k-1}\|^2}. \end{aligned} \tag{16}$$

Determine the step size $\alpha_k$ by Wolfe line search (5).
Step 3. Set $x_{k+1} = x_k + \alpha_k d_k$, and $k := k + 1$; go to Step 1.

**Remark 2.1** If the constant $\mu = 0$, then the TMPRP1 method and TMPRP2 method reduce to the methods proposed by Cheng [9] and Zhang *et al.* [10], respectively, and the TMPRP3 method reduces to a method similar to that proposed by Yu *et al.* [20].

**Remark 2.2** Obviously, if the line search is exact, then the direction generated by (13) or (14) or (15) reduces to (3) with $\beta_k = \beta_k^{\mathrm{PRP}}$. Therefore, in the following, we assume that $\mu > 0$.

**Remark 2.3** From (13) and (14), we can easily obtain

$$g_k^\top d_k = -\|g_k\|^2 \quad \text{and} \quad \|g_k\| \le \|d_k\|. \tag{17}$$

This indicates that the TMPRP1 method and the TMPRP2 method satisfy the sufficient descent property. In addition, from the following lemma, we can see that the TMPRP3 method also satisfies this property.

**Lemma 2.1** *Let* $\{x_k\}$ *and* $\{d_k\}$ *be generated by the* TMPRP3 method, *then we have*

$$g_k^\top d_k \le -\left(1 - \frac{1}{t}\right)\|g_k\|^2. \tag{18}$$

*Proof* We have from (15) and (16)

$$\begin{aligned}
g_k^\top d_k &= -\|g_k\|^2 + \left(\frac{g_k^\top(g_k - g_{k-1})}{\mu|g_k^\top d_{k-1}| + \|g_{k-1}\|^2} - t\frac{\|y_{k-1}\|^2 g_k^\top d_{k-1}}{(\mu|g_k^\top d_{k-1}| + \|g_{k-1}\|^2)^2}\right) g_k^\top d_{k-1} \\
&\quad + \frac{g_k^\top d_{k-1}}{\mu|g_k^\top d_{k-1}| + \|g_{k-1}\|^2}\left(g_k^\top y_{k-1} - g_k^\top s_{k-1}\right) \\
&\le -\|g_k\|^2 + 2\frac{g_k^\top y_{k-1} g_k^\top d_{k-1}}{\mu|g_k^\top d_{k-1}| + \|g_{k-1}\|^2} - t\frac{\|y_{k-1}\|^2 (g_k^\top d_{k-1})^2}{(\mu|g_k^\top d_{k-1}| + \|g_{k-1}\|^2)^2} \\
&\quad - \frac{\alpha_{k-1}(g_k^\top d_{k-1})^2}{\mu|g_k^\top d_{k-1}| + \|g_{k-1}\|^2} \\
&\le -\|g_k\|^2 + 2\left(\frac{1}{\sqrt{t}} g_k\right)^\top \left(\frac{\sqrt{t} g_k^\top d_{k-1}}{\mu|g_k^\top d_{k-1}| + \|g_{k-1}\|^2} y_{k-1}\right) - t\frac{\|y_{k-1}\|^2 (g_k^\top d_{k-1})^2}{(\mu|g_k^\top d_{k-1}| + \|g_{k-1}\|^2)^2} \\
&\le -\|g_k\|^2 + \frac{1}{t}\|g_k\|^2 + t\frac{\|y_{k-1}\|^2 (g_k^\top d_{k-1})^2}{(\mu|g_k^\top d_{k-1}| + \|g_{k-1}\|^2)^2} - t\frac{\|y_{k-1}\|^2 (g_k^\top d_{k-1})^2}{(\mu|g_k^\top d_{k-1}| + \|g_{k-1}\|^2)^2} \\
&= -\left(1 - \frac{1}{t}\right)\|g_k\|^2,
\end{aligned}$$

which indicates that (18) holds by induction since $d_0 = -g_0$ and $t > 1$. This completes the proof. □

**Remark 2.4** From the proof of Lemma 2.1, we can see that if the term $s_{k-1}$ in $d_k$ is deleted, then the above sufficient descent property still holds.

The global convergence proof of the above three methods is similar, here, we only prove the global convergence of the TMPRP1 method. In the case of the other two methods, the argument is similar.

The following lemma, called the Zoutendijk condition, is often used to prove global convergence of conjugate gradient method. It was originally given by Zoutendijk in [21].

**Lemma 2.2** *Suppose that $x_0$ is a starting point for which assumptions* (H1) *and* (H2) *hold. Consider any method in the form of* (2), *where $d_k$ is a descent direction and $\alpha_k$ satisfies the Wolfe condition* (5) *or the strong Wolfe condition* (6). *Then we have*

$$\sum_{k=0}^{\infty} \frac{(g_k^\top d_k)^2}{\|d_k\|^2} < +\infty.$$

*This together with* (17) *shows that*

$$\sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < +\infty. \tag{19}$$

**Definition 2.1** The function $f(x)$ is said to be uniformly convex on $\mathcal{R}^n$, if there is a positive constant $m$ such that

$$d^\top \nabla^2 f(x) d \geq m\|d\|^2, \quad \forall x, d \in \mathcal{R}^n,$$

where $\nabla^2 f(x)$ is the Hessian matrix of the function $f(x)$.

Now we prove the strongly global convergence of TMPRP1 method for uniformly convex functions.

**Lemma 2.3** *Let the sequences $\{x_k\}$ and $\{d_k\}$ be generated by* TMPRP1 method, *and the function $f(x)$ be uniformly convex, then we have*

$$c_1 \alpha_k \|d_k\|^2 \leq -g_k^\top d_k, \tag{20}$$

*where $c_1 = (1-\rho)^{-1} m/2$.*

*Proof* See Lemma 2.1 in [22]. □

The proof of the following theorem is similar to that of Theorem 2.1 in [22]. For completeness, we give the proof.

**Theorem 2.1** *Suppose that the assumptions* (H1) *and* (H2) *hold, and $f(x)$ is uniformly convex, then we have*

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

*Proof* From (11), (20), and (H2), we have

$$\left|\beta_k^{\mathrm{MPRP}}\right| \leq \left|\frac{g_k^\top (g_k - g_{k-1})}{\|g_{k-1}\|^2}\right| \leq \frac{L\alpha_{k-1}\|g_k\|\|d_{k-1}\|}{-g_{k-1}^\top d_{k-1}} \leq \frac{L}{c_1}\frac{\|g_k\|}{\|d_{k-1}\|}.$$

This together with (13) shows that

$$\begin{aligned} \|d_k\| &\le \|g_k\| + \left|\beta_k^{\mathrm{MPRP}}\right| \frac{\|g_k\|\|d_{k-1}\|}{\|g_k\|^2}\|g_k\| + \left|\beta_k^{\mathrm{MPRP}}\right|\|d_{k-1}\| \\ &\le \|g_k\| + \frac{2L}{c_1}\|g_k\| \\ &= \left(1 + \frac{2L}{c_1}\right)\|g_k\|. \end{aligned}$$

Then, letting $\sqrt{A} = 1 + \frac{2L}{c_1}$, we get $\|d_k\|^2 \le A\|g_k\|^2$. So, by (19), we get

$$\lim_{k\to\infty} \|g_k\|^2 = \lim_{k\to\infty} \frac{\|g_k\|^4}{\|g_k\|^2} \le A \lim_{k\to\infty} \frac{\|g_k\|^4}{\|d_k\|^2} = 0.$$

This completes the proof. □

We are going to investigate the global convergence of the TMPRP1 method with Wolfe line search (5) for nonconvex function. In the last part of this subsection, we use $\beta_k^{\mathrm{MPRP}+}$ to replace $\beta_k^{\mathrm{MPRP}}$ in (13).

The next lemma corresponds to Lemma 4.3 in [23] and Theorem 3.2 in [24].

**Lemma 2.4** *Suppose that assumptions* (H1) *and* (H2) *hold. Let* $\{x_k\}$ *be the sequence generated by* TMPRP1 method. *If there exists a constant* $\varepsilon > 0$ *such that* $\|g_k\| \ge \varepsilon$ *for all* $k \ge 0$, *then we have*

$$\sum_{k=0}^{\infty} \|u_{k+1} - u_k\|^2 < +\infty, \tag{21}$$

*where* $u_k = d_k/\|d_k\|$.

*Proof* From (17) and $\|g_k\| \ge \varepsilon$ for all $k$, we have $\|d_k\| > 0$ for all $k$. Therefore, $u_k$ is well defined. Define

$$r_k = -\frac{(1 + \beta_k^{\mathrm{MPRP}+}\frac{g_k^\top d_{k-1}}{\|g_k\|^2})}{\|d_k\|} g_k \quad \text{and} \quad \delta_k = \beta_k^{\mathrm{MPRP}+}\frac{\|d_{k-1}\|}{\|d_k\|}.$$

Then we have

$$u_k = r_k + \delta u_{k-1}.$$

Since $u_{k-1}$ and $u_k$ are unit vectors, we can write

$$\|r_k\| = \|u_k - \delta u_{k-1}\| = \|\delta u_k - u_{k-1}\|.$$

Noting that $\delta_k \ge 0$, we get

$$\|u_k - u_{k-1}\| \le \left\|(1 + \delta_k)(u_k - u_{k-1})\right\| \le \|u_k - \delta u_{k-1}\| + \|\delta u_k - u_{k-1}\| = 2\|r_k\|. \tag{22}$$

From (10), (11), and (H2), we have

$$\left|\beta_k^{\mathrm{MPRP}+}\right| \frac{|g_k^\top d_{k-1}|}{\|g_k\|^2} \leq \frac{\|g_k\| LB}{\mu |g_k^\top d_{k-1}|} \frac{|g_k^\top d_{k-1}|}{\|g_k\|^2} \leq \frac{LB}{\varepsilon\mu}. \tag{23}$$

From (9), (10), and (23), it follows that there exists a constant $M_1 \geq 0$ such that

$$\left\| -\left(1 + \beta_k^{\mathrm{MPRP}+} \frac{g_k^\top d_{k-1}}{\|g_k\|^2}\right) g_k \right\| \leq \|g_k\| + \frac{LB}{\varepsilon\mu}\gamma \leq \gamma + \frac{LB}{\varepsilon\mu}\gamma \doteq M_1. \tag{24}$$

Thus, from (19) and (24), we get

$$\sum_{k=0}^{\infty} \|r_k\|^2 \leq \sum_{k=0}^{\infty} \frac{M_1^2}{\|d_k\|^2} = \sum_{k=0}^{\infty} \frac{M_1^2}{\|g_k\|^4} \frac{\|g_k\|^4}{\|d_k\|^2} \leq \frac{M_1^2}{\varepsilon^4} \sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < +\infty,$$

which together with (22) completes the proof. □

The following theorem establishes the global convergence of the TMPRP1 method with Wolfe line search (5) for general nonconvex functions. The proof is analogous to that of Theorem 3.2 in [24].

**Theorem 2.2** *Let the assumptions* (H1) *and* (H2) *hold. Then the sequence* $\{x_k\}$ *generated by* TMPRP1 method *satisfies*

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{25}$$

*Proof* Assume that the conclusion (25) is not true. Then there exists a constant $\varepsilon > 0$ such that for all

$$\|g_k\| \geq \varepsilon, \quad \forall k \geq 0.$$

The proof is divided into the following two steps.

Step I. A bound on the steps $s_k$. We observe that for any $l \geq k$,

$$x_l - x_k = \sum_{j=k}^{l-1} (x_{j+1} - x_j) = \sum_{j=k}^{l-1} \|s_j\| u_j = \sum_{j=k}^{l-1} \|s_j\| u_k + \sum_{j=k}^{l-1} \|s_j\| (u_j - u_k), \tag{26}$$

where $s_j = x_{j+1} - x_j$ and $u_k$ is defined in Lemma 2.4. Using the triangle inequality and $\|u_k\| = 1$, we can write (26) as

$$\sum_{j=k}^{l-1} \|s_j\| \leq \|x_l - x_k\| + \sum_{j=k}^{l-1} \|s_j\| \|u_j - u_k\| \leq B + \sum_{j=k}^{l-1} \|s_j\| \|u_j - u_k\|. \tag{27}$$

Let $\Delta$ be an arbitrary but fixed positive integer. It follows from Lemma 2.4 that there is an index $k_\Delta$ such that

$$\sum_{i \geq k_\Delta} \|u_{i+1} - u_i\|^2 \leq \frac{1}{4\Delta}. \tag{28}$$

If $j > k \geq k_\Delta$ with $j - k \leq \Delta$, then by (28) and Cauchy-Schwarz inequality, we have

$$\begin{aligned}\|u_j - u_k\| &\leq \sum_{i=k}^{j-1} \|u_{i+1} - u_i\| \\ &\leq \left( (j-k) \sum_{i=k}^{j-1} \|u_{i+1} - u_i\|^2 \right)^{\frac{1}{2}} \\ &\leq \left( \Delta \frac{1}{4\Delta} \right)^{\frac{1}{2}} = \frac{1}{2}.\end{aligned}$$

Combining this with (27) yields

$$\sum_{j=k}^{l-1} \|s_j\| \leq 2B, \tag{29}$$

where $l > k \geq k_\Delta$ with $l - k \leq \Delta$.

Step II. A bound on the direction $d_k$. From (13) and (24), we have

$$\begin{aligned}\|d_k\|^2 &\leq \left( \left\| -\left(1 + \beta_k^{\mathrm{MPRP+}} \frac{g_k^\top d_{k-1}}{\|g_k\|^2}\right) g_k \right\| + \left|\beta_k^{\mathrm{MPRP+}}\right| \|d_{k-1}\| \right)^2 \\ &\leq \left(M_1 + \left|\beta_k^{\mathrm{MPRP+}}\right| \|d_{k-1}\|\right)^2 \\ &\leq 2M_1^2 + 2\left(\beta_k^{\mathrm{MPRP+}}\right)^2 \|d_{k-1}\|^2 \\ &\leq 2M_1^2 + \frac{2L^2\gamma^2\|s_{k-1}\|^2}{\varepsilon^2} \|d_{k-1}\|^2.\end{aligned}$$

By the use of the same argument of the Case III of Theorem 3.2 in [3], we can get the conclusion (25). This completes the proof. □

**Remark 2.5** From Theorem 2.2, we can see that the TMPRP1 method possesses better convergence properties than CTPRP method in [2]. Since the TMPRP1 method converges globally for nonconvex minimization problems with a Wolfe line search, while the CTPPR method converges globally for nonconvex minimization problems with a strong Wolfe line search. We also note that the term $\mu|g_k^\top d_{k-1}|$ in the denominator of (11) plays an important role in the proof of Lemma 2.4.

## 3 Numerical results

In this section, we present some numerical results to compare the performance of the TMPRP1 method, the CG_DESCENT method in [24] and the DTPRP method in [19].

- TMPRP1: the TMPRP1 method with Wolfe line search (5), with $\mu = 10^{-4}$, $\rho = 0.1$, $\sigma = 0.5$;
- CG_DESCENT: the CG_DESCENT method with Wolfe line search (5), with $\rho = 0.1$, $\sigma = 0.5$;
- DTPRP: the DTPRP method with Wolfe line search (5), with $\mu = 1.2$, $\rho = 0.1$, $\sigma = 0.5$.

All codes were written in Matlab 7.1 and run on a portable computer. We stopped the iteration if the number of iterations exceeded 1,000 or $\|g_k\| < 10^{-5}$. Here, we use some

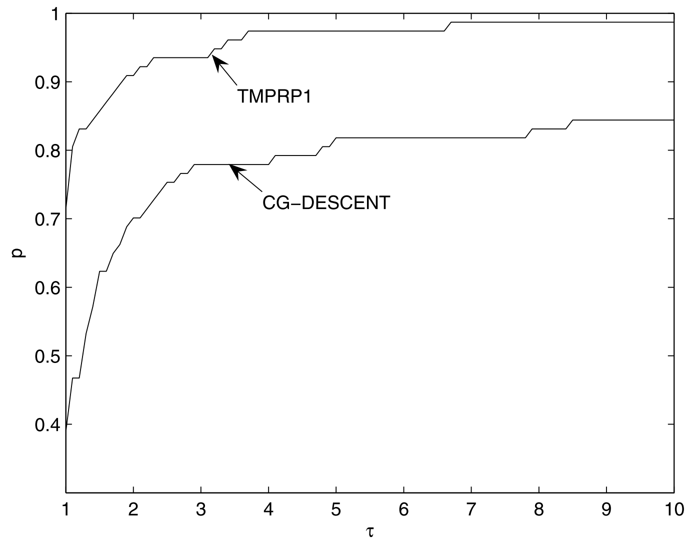


**Figure 1 Performance profiles of TMPRP1 and CG_DESCENT about CPU time.**

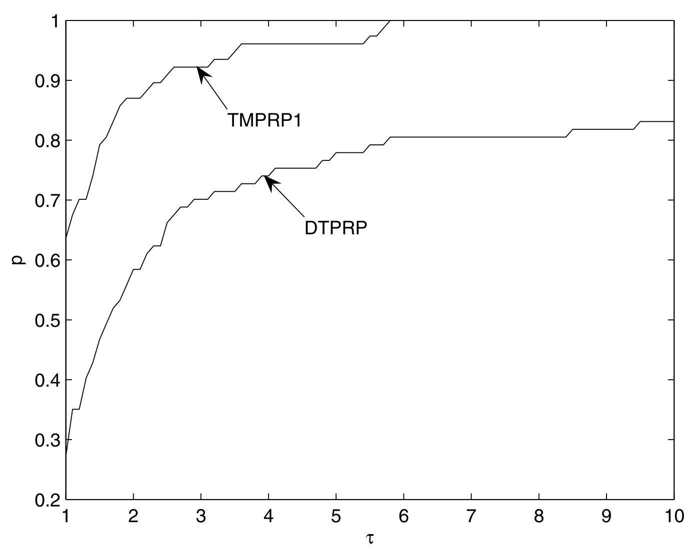


**Figure 2 Performance profiles of TMPRP1 and DTPRP about CPU time.**

test problems in [17] with different dimensions. Our numerical results are listed in the form NI/NF/CPU, where the symbols NI, NF, and CPU mean the number of iterations, the number of function evaluations and the CPU time in seconds, respectively. 'F' means the method failed. Here, the code of Wolfe line search (5) is adapted from [25]. In Figures 1 and 2, we adopt the performance profiles by Dolan and Moré [12] to compare the performance based on the CPU time between the TMPRP1 method, the CG_DESCENT method and the DTPRP method. That is, for each method, we plot the fraction P of problems for which the method is within a factor $\tau$ of the best time. The left side of the figure gives the percentage of the test problems for which a method is fastest; while the right side gives the percentage of thee test problems that are successfully solved by each of the methods.

The top curve is the method that solved the most problems in a time that was within a factor $\tau$ of the best time. From Table 1 and Figures 1 and 2, we can see that the TMPRP1 method performs better than the CG_DESCENT method and the DTPRP method, thus the proposed TMPRP1 method is computationally efficient.

**Table 1 The results for the methods on the tested problems**

| P | *n* | TMPRP1 | CG_DESCENT | DTPRP |
|---|---|---|---|---|
| Freudenstein and Roth | 100 | 52/1,017/0.4,688 | 53/1,030/0.4219 | 94/2,037/0.8125 |
| Trigonometric | 5,000 | 118/539/5.6094 | 75/603/5.6250 | 57/170/2.2813 |
| Extended Rosenbrock | 5,000 | 44/868/1.7344 | 119/2,195/3.6875 | 54/956/1.8281 |
| Generalized Rosenbrock | 10 | 223/5,114/1.2500 | 567/13,632/3.5156 | 305/6,522/1.6719 |
| White | 1,000 | 48/874/1.3594 | 101/2,321/3.3438 | 71/1,474/2.0625 |
| Beale | 5,000 | 45/933/3.7344 | 98/2,182/8.3906 | 43/555/2.2031 |
| Penalty | 5,000 | 30/593/1.2969 | 26/516/1.1094 | F |
| Perturbed quadratic | 100 | 92/1,674/0.5469 | 114/1,974/0.6563 | 99/2,302/0.6875 |
| Raydan 1 | 500 | 171/3,083/1.9219 | 231/3,882/2.3125 | 150/2,333/1.4688 |
| Raydan 2 | 5,000 | 5/6/0.3750 | 6/60/0.6250 | 6/7/0.3438 |
| Diagonal 1 | 100 | 88/1,462/0.6250 | 74/880/0.4063 | 83/1,608/0.6563 |
| Diagonal 2 | 100 | 780/781/0.5938 | 104/341/0.1875 | 780/781/0.5313 |
| Diagonal 3 | 100 | 101/1,492/0.7188 | 154/2,321/1.0781 | 77/767/0.4219 |
| Hager | 100 | 44/640/0.3281 | 32/251/0.1563 | 34/403/0.2188 |
| Generalized tridiagonal 1 | 1,000 | 41/578/1.4844 | 31/403/1.0469 | 58/1,078/2.8594 |
| Extended tridiagonal 1 | 1,000 | 41/432/1.1250 | 40/497/1.2188 | 46/724/1.8281 |
| Extended three expo terms | 5,000 | 45/759/5.5156 | 31/246/2.0781 | 21/174/1.5469 |
| Generalized tridiagonal 2 | 1,000 | 56/785/1.3594 | 404/11,638/19.5938 | 61/1031/1.7813 |
| Diagonal 4 | 5,000 | 48/815/1.3906 | 128/2,383/3.6406 | 55/673/1.1719 |
| Diagonal 5 | 5,000 | 4/5/0.2969 | 4/8/0.3906 | 4/5/0.3281 |
| Extended Himmelblau | 5,000 | 30/438/1.0625 | 23/214/0.7500 | 20/178/0.7344 |
| Generalized PSC1 | 5,000 | 222/1,100/5.7344 | 672/5,554/27.0156 | F |
| Extended PSC1 | 5,000 | 55/916/5.0156 | 24/173/1.5156 | 24/187/1.4375 |
| Extended Powell | 5,000 | 193/2,649/17.4844 | F | 536/7,005/42.2031 |
| Extended BD1 | 5,000 | 35/431/1.5156 | 49/856/2.8281 | 33/452/1.6250 |
| Extended Maratos | 1,000 | 66/1,121/0.6563 | F | 136/2,206/1.2344 |
| Extended Cliff | 5,000 | 48/262/1.6094 | 123/1,275/6.5000 | F |
| Quadratic diagonal perturbed | 5,000 | 433/6,793/3.6875 | F | 247/3,834/2.1094 |
| Extended Wood | 5,000 | 199/2,976/5.7188 | F | 131/2,075/4.1094 |
| Extended Hiebert | 5,000 | 2/32/0.4844 | 2/33/0.5313 | 3/62/0.5469 |
| Quadratic QF1 | 5,000 | 731/12,453/19.8438 | 790/13,180/20.1875 | 882/14,508/22.3906 |
| Extended QP1 | 1,000 | 65/1,662/1.0156 | 25/361/0.2813 | 16/157/0.1875 |
| Extended QP2 | 5,000 | 64/988/5.1719 | 143/2,919/13.9219 | 78/1,170/6.4063 |
| Quadratic QF2 | 5,000 | 777/14,146/24.6406 | 968/17,331/31.9219 | 814/14,358/24.7344 |
| Extended EP1 | 5,000 | 101/2,391/6.6094 | 12/195/1.0000 | 136/3,136/8.0781 |
| Extended tridiagonal 2 | 5,000 | 46/615/2.0156 | 63/1,270/3.4844 | 32/170/0.9219 |
| BDQRTIC | 100 | 159/2,473/0.8438 | F | 185/3,133/1.0156 |
| TRIDIA | 100 | 310/4,816/1.5938 | 440/7,143/2.2344 | 364/6,190/1.8281 |
| ARWHEAD | 5,000 | 35/702/2.2813 | F | F |
| NONDIA | 5,000 | 30/626/1.8906 | F | 209/4,327/10.7344 |
| NONDQUAR | 5 | 713/779/0.4531 | 97/809/0.2813 | F |
| DQDRTIC | 5,000 | 80/1,234/2.7031 | 117/2,386/4.8750 | 81/1,108/2.4688 |
| EG2 | 100 | 165/2,715/1.5,000 | 85/1,136/0.7969 | F |
| DIXMAANA | 5,001 | 21/191/7.0938 | 13/177/6.4688 | 10/70/2.8281 |
| DIXMAANB | 5,001 | 22/45/2.0000 | 13/127/4.7344 | 7/14/0.8750 |
| DIXMAANC | 5,001 | 17/136/5.1719 | 15/231/8.7500 | 6/17/0.9219 |
| DIXMAANE | 102 | 346/451/0.7031 | 186/5,359/5.4844 | 321/325/0.5313 |
| Partial perturbed quadratic | 100 | 87/1,905/1.6094 | 116/2,180/1.6719 | 77/1,326/0.9844 |
| Broyden tridiagonal | 5,000 | 114/1,927/5.1719 | 101/1,884/4.9531 | 119/2,029/5.2656 |
| Almost perturbed quadratic | 5,000 | 854/19,329/30.2969 | F | 866/19,516/32.0938 |
| Tridiagonal perturbed quadratic | 5,000 | 760/16,744/42.7813 | 959/22,230/52.9063 | 774/17,831/43.9063 |
| EDENSCH | 1,000 | 40/615/1.7188 | 35/450/1.1875 | 49/1,273/3.0938 |
| HIMMELBHA | 5,000 | 15/69/1.4063 | F | 17/18/0.6406 |
| STAIRCASE S1 | 100 | 341/5,058/1.5781 | F | 510/7,591/2.4844 |

**Table 1** (***Continued***)

| P | *n* | TMPRP1 | CG_DESCENT | DTPRP |
|---|---|---|---|---|
| LIARWHD | 5,000 | 39/727/1.9688 | 165/3,873/9.7500 | 262/6,799/16.2500 |
| DIAGONAL 6 | 5,000 | 5/6/0.3594 | 6/60/0.5313 | 6/7/0.3594 |
| DIXON3DQ | 100 | 578/9,208/3.0625 | F | 499/7,241/2.1563 |
| ENGVAL1 | 5,000 | 36/611/1.7344 | 52/1,264/3.3906 | F |
| DENSCHNA | 5,000 | 23/249/1.9063 | 27/318/2.5938 | 19/93/1.0469 |
| DENSCHNB | 5,000 | 21/54/0.4844 | 10/79/0.4531 | 20/327/0.9063 |
| DENSCHNC | 5,000 | 23/191/2.6094 | 34/357/4.5938 | F |
| DENSCHNF | 5,000 | 25/343/1.1094 | 24/348/1.1250 | 23/363/1.2656 |
| SINQUAD | 100 | 505/10,201/4.1250 | F | F |
| BIGGSB1 | 100 | 489/5,248/1.7031 | F | 533/5,660/1.6406 |
| Extended block-diagonal | 1,000 | 30/506/0.8906 | 36/508/0.8750 | 26/374/0.5938 |
| Generalized quartic 1 | 5,000 | 21/159/0.7500 | 18/342/1.0469 | 36/777/1.8594 |
| DIAGONAL 7 | 5,000 | 53/2,509/14.1563 | 54/2,477/13.8125 | F |
| DIAGONAL 8 | 5,000 | 57/2,710/18.3906 | 56/2,622/17.5781 | F |
| Full Hessian | 5,000 | 17/239/1.6094 | 18/305/1.9688 | 46/1,643/8.8750 |
| SINCOS | 5,000 | 26/250/1.5313 | 22/132/1.1719 | F |
| Generalized quartic 2 | 5,000 | 48/996/2.4688 | 35/606/1.4375 | 39/704/1.6250 |
| EXTROSNB | 5,000 | 39/741/1.7500 | 159/7,756/14.8750 | 43/1,010/2.3281 |
| ARGLINB | 100 | 101/5,024/1.7969 | 111/5,498/1.7813 | 23/691/0.3125 |
| FLETCHCR | 5,000 | 61/1,662/4.0469 | 36/661/1.7813 | 61/1,976/4.7969 |
| HIMMELBG | 2 | F | 2/4/0.0313 | F |
| HIMMELBH | 5,000 | 18/103/0.7969 | 23/224/1.1406 | 16/91/0.6719 |
| DIAGONAL 9 | 5,000 | 1/3/0.3594 | 1/3/0.3906 | 1/3/0.3594 |

## 4 Conclusion

This paper proposed three modified PRP conjugate gradient methods, which are some improvements of recently proposed PRP conjugate gradient methods. The global convergence of the proposed methods are established under the Wolfe line search. The effectiveness of the proposed methods have been shown by some numerical examples. We find that the performance of the TMPRP1 method is related to the parameter $\mu$ in $\beta_k^{\mathrm{MPRP}}$; therefore, how to choose a suitable parameter $\tau$ deserves further investigation.

**Competing interests**
The authors declare that they have no competing interests.

**Authors' contributions**
The first author has designed the three methods and the second author has refined them. Both authors have equally contributed in the numerical results. All authors read and approved the final manuscript.

**Author details**
[1]School of Mathematics and Statistics, Zaozhuang University, Beian Street, Zaozhuang, 277160, P.R. China. [2]School of Mathematics and Statistics, Zhejiang University of Finance and Economics, Xueyuan Street, Hangzhou, 310018, P.R. China.

**Acknowledgements**
The authors gratefully acknowledge the helpful comments and suggestions of the anonymous reviewers. This work was partially supported by the domestic visiting scholar project funding of Shandong Province outstanding young teachers in higher schools, the foundation of Scientific Research Project of Shandong Universities (No. J13LI03), and the Shandong Province Statistical Research Project (No. 20143038).

Received: 3 December 2014 Accepted: 30 March 2015 Published online: 08 April 2015

**References**

1. Fletcher, R, Reeves, C: Function minimization by conjugate gradients. J. Comput. **7**, 149-154 (1964)
2. Polak, B, Ribière, G: Note sur la convergence des méthodes de directions conjuguées. Rev. Fr. Inform. Rech. Oper. **16**, 35-43 (1969)
3. Gilbert, JC, Nocedal, J: Global convergence properties of conjugate gradient methods for optimization. SIAM J. Optim. **2**, 21-42 (1992)
4. Liu, YL, Storey, CS: Efficient generalized conjugate gradient algorithms, part 1: theory. J. Optim. Theory Appl. **69**, 129-137 (1991)

5. Dai, YH, Yuan, YX: A nonlinear conjugate gradient method with a strong global convergence property. SIAM J. Optim. **10**, 177-182 (2000)
6. Hestenes, MR, Stiefel, EL: Method of conjugate gradient for solving linear systems. J. Res. Natl. Bur. Stand. **49**, 409-432 (1952)
7. Fletcher, R: Practical Methods of Optimization. Volume 1: Unconstrained Optimization. Wiley, New York (1987)
8. Grippo, L, Luidi, S: A globally convergent version of the Polak-Ribière gradient method. Math. Program. **78**, 375-391 (1997)
9. Cheng, WY: A two-term PRP-based descent method. Numer. Funct. Anal. Optim. **28**, 1217-1230 (2007)
10. Zhang, L, Zhou, WJ, Li, DH: A descent modified Polak-Ribière-Polyak conjugate gradient method and its global convergence. IMA J. Numer. Anal. **26**, 629-640 (2006)
11. Yu, GH, Guan, LT, Chen, WF: Spectral conjugate gradient methods with sufficient descent property for large-scale unconstrained optimization. Optim. Methods Softw. **23**, 275-293 (2008)
12. Dolan, ED, Moré, JJ: Benchmarking optimization software with performance profiles. Math. Program. **91**, 201-213 (2002)
13. Wei, ZX, Li, G, Qi, LQ: Global convergence of the Polak-Ribière-Polyak conjugate gradient method with inexact line searches for non-convex unconstrained optimization problems. Math. Comput. **77**, 2173-2193 (2008)
14. Li, G, Tang, CM, Wei, ZX: New conjugacy condition and related new conjugate gradient methods for unconstrained optimization. J. Comput. Appl. Math. **202**, 523-539 (2007)
15. Yu, G, Guan, L, Li, G: Global convergence of modified Polak-Ribière-Polyak conjugate gradient methods with sufficient descent property. J. Ind. Manag. Optim. **3**, 565-579 (2008)
16. Dai, YH, Kou, CX: A nonlinear conjugate gradient algorithm with an optimal property and an improved Wolfe line search. SIAM J. Optim. **23**, 296-320 (2013)
17. Neculai, A: Unconstrained optimization by direct searching (2007). http://camo.ici.ro/neculai/UNO/UNO.FOR
18. Wei, ZX, Yao, SW, Liu, LY: The convergence properties of some new conjugate gradient methods. Appl. Math. Comput. **183**, 1341-1350 (2006)
19. Dai, ZF, Wen, FH: Another improved Wei-Yao-Liu nonlinear conjugate gradient method with sufficient descent property. Appl. Math. Comput. **218**, 7421-7430 (2012)
20. Yu, GH, Zhao, YL, Wei, ZX: A descent nonlinear conjugate gradient method for large-scale unconstrained optimization. Appl. Math. Comput. **187**, 636-643 (2007)
21. Zoutendijk, G: Nonlinear programming, computational methods. In: Abadie, J (ed.) Integer and Nonlinear Programming, pp. 37-86. North-Holland, Amsterdam (1970)
22. Dai, ZF, Tian, BS: Global convergence of some modified PRP nonlinear conjugate gradient methods. Optim. Lett. **5**(4), 615-630 (2011)
23. Dai, YH, Liao, LZ: New conjugacy conditions and related nonlinear conjugate gradient methods. Appl. Math. Optim. **43**, 97-101 (2001)
24. Hager, WW, Zhang, HC: A survey of nonlinear conjugate gradient methods. Pac. J. Optim. **2**, 35-58 (2006)
25. Wu, QZ, Zheng, ZY, Deng, W: Operations Research and Optimization, MATLAB Programming, pp. 66-69. China Machine Press, Beijing (2010)