Soft Computing (2021) 25:5867–5879
https://doi.org/10.1007/s00500-021-05580-0

METHODOLOGIES AND APPLICATION

The modified PRP conjugate gradient algorithm under a non-descent line search and its application in the Muskingum model and image restoration problems

Gonglin Yuan[1] · Junyu Lu[1] · Zhan Wang[1]

Published online: 27 January 2021

Abstract

In this paper, a modified Polak–Ribière–Polyak (PRP) method, which possesses the following desired properties for unconstrained optimization problems, is presented. (i) The search direction of the given method has the gradient value and the function value. (ii) A non-descent backtracking-type line search technique is proposed to obtain the step size α_k and construct a point. (iii) The method inherits an important property of the classical PRP method: the tendency to turn towards the steepest descent direction if a small step is generated away from the solution, preventing a sequence of tiny steps from happening. (iv) The strongly global convergence and R-linear convergence of the modified PRP method for nonconvex optimization are established under some suitable assumptions. (v) The numerical results show that the modified PRP method not only is interesting in practical computation but also has better performance than the normal PRP method in estimating the parameters of the nonlinear Muskingum model and performing image restoration.

Keywords PRP method · Strongly global convergence · Non-descent line search · R-linear convergence

1 Introduction

Consider the following unconstrained optimization problem:

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

where $f:\Re^n \to \Re$, $f \in C^2$ is continuously differentiable. Conjugate gradient (CG) methods (Dai 2001; Grippo and Lucidi 1997; Khoda et al. 1992; Nocedal and Wright 2006; Shi 2002) are particularly powerful for solving large-scale problems due to their simplicity and lower storage (Birgin and Martínez 2001; Cohen 1972; Shanno 1978; Yuan 1993); thus, they are especially popular for solving unconstrained optimization problems. CG methods generate an iterative sequence $\{x_k\}$ by:

$$x_{k+1} = x_k + \alpha_k d_k,\ \ k = 0,\ 1,\ 2, \ldots, \tag{1.2}$$

where x_k is the k-th iteration point, the step size $\alpha_k > 0$ can be computed by certain line search techniques, and the search direction d_k is defined by the following formula:

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

where β_k is a scalar and can be defined by the following six formulas (or other formulas):

This work was supported by the National Natural Science Foundation of China (Grant No. 11661009), the High Level Innovation Teams and Excellent Scholars Program in Guangxi institutions of higher education (Grant No. [2019]52), the Guangxi Natural Science Foundation (2020GXNSFAA159069) and the Guangxi Natural Science Key Fund (No. 2017GXNSFDA198046).

✉ Gonglin Yuan
glyuan@gxu.edu.cn

Junyu Lu
ljy2412095857@163.com

Zhan Wang
13283799361@163.com

1 College of Mathematics and Information Science, Guangxi University, Nanning, Guangxi, People's Republic of China

$$\beta_k^{PRP} = \frac{g_k^T(g_k - g_{k-1})}{\|g_{k-1}\|^2}, \ \beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2},$$

$$\beta_k^{HS} = \frac{g_k^T(g_k - g_{k-1})}{d_{k-1}^T(g_k - g_{k-1})},$$

$$\beta_k^{CD} = \frac{\|g_k\|^2}{-d_{k-1}^T g_{k-1}}, \ \beta_k^{LS} = \frac{g_k^T(g_k - g_{k-1})}{-d_{k-1}^T g_{k-1}},$$

$$\beta_k^{DY} = \frac{\|g_k\|^2}{(g_k - g_{k-1})^T d_{k-1}},$$

where g_{k-1} is the gradient $\nabla f(x_{k-1})$ of $f(x)$ at the point x_{k-1} and $\|\cdot\|$ is the Euclidean norm. The corresponding methods are called the Polak–Ribière–Polyak (PRP) (Polak and Ribière 1969; Polyak 1969), Fletcher–Reeves (FR) (Fletcher and Reeves 1964), Hestenses–Stiefel (HS) (Hestenes and Stiefel 1952), conjugate descent (CD) (Fletcher 1997), Liu-Storrey (LS) (Liu and Storey 2000), and Dai-Yuan (DY) (Dai and Yuan 1999) CG methods, respectively. The convergence of the CD method, DY method, and FR method are relatively easy to establish, but their numerical results are not ideal in real computations. Powell (1986) presented an explanation of the numerical disadvantages of the FR method, such as subsequent steps being very short if a small step is originated away from the solution point. However, if a poor direction occurs in practical computation, the PRP, HS, or LS method will perform a restart, so these three methods perform much better than the above three methods. They are generally regarded as the most efficient conjugate gradient methods.

In this paper, we specifically study the modified PRP method. There has been extensive study regarding the global convergence of the PRP method. Polak and Ribière (1969) proved that the PRP method with an exact line search is globally convergent for strongly convex functions but that it fails to satisfy the property of global convergence for the general functions under the Wolfe line search technique, and this is a still an open problem. Yuan (1993) further established global convergence with the modified Wolfe line search under the condition that the search direction is descending. All convergence discussion of the PRP algorithm hinted that the key issue for the study of the PRP method is the sufficient descent condition. However, there are several limitations of the PRP method such that it may not provide a descent direction of the objective function under the exact line search, which creates serious consequences for the global convergence of the algorithm for general functions. For example, Powell (1984) provided a counterexample to show that the PRP method might circle infinitely without approaching the solution even if α_k is chosen as the least positive minimizer of the line search. Through Powell's analysis Powell (1986), the PRP method did not satisfy global convergence, probably due to β_k being negative, so inspired by his study, Gilbert and Nocedal (1992) found a convergent consequence in the PRP method when $\beta_k^{PRP+} = \max\{0, \beta^{PRP}\}$ for general nonconvex functions with a suitable line search. Hence, other researchers (Cheng 2007; Yuan et al. 2020, ?) modified β_k such that the algorithm satisfies global convergence for general functions. Inspired by Li et al. (2015), Yuan and Wei (2010), in this paper we, present a modified PRP method as follows:

$$d_k = \begin{cases} -g_k, & \text{if } k = 0, \\ -g_k + \beta_k^{*PRP} d_{k-1}, & \text{if } k \geq 1, \end{cases} \tag{1.3}$$

and

$$\beta_k^{*PRP} = \frac{g_k^T \tilde{y}_{k-1}}{\|g_{k-1}\|^2}, \tag{1.4}$$

where $\tilde{y}_{k-1} = y_{k-1} + \frac{\rho_{k-1} s_{k-1}}{\|s_{k-1}\|^2}$, $\rho_{k-1} = 2[f(x_{k-1}) - f(x_k)] + [g(x_k) + g(x_{k-1})] s_{k-1}$ and $s_{k-1} = x_k - x_{k-1}$. In addition to modifying β_k, some researchers choose to modify the line search to obtain the global convergence of the PRP method for general functions. Grippo and Lucidi (1997) presented a new descent line search technique as follows. There exist constants $\mu > 0$, $\delta > 0$ and $0 < t < 1$, $\alpha_k = \max\{t^j \left(\frac{\mu |g_k^T d_k|}{\|d_k\|^2}\right); \ j = 0, 1, \ldots\}$ such that

$$f(x_{k+1}) \leq f(x_k) + \delta \alpha_k^2 \|d_k\|^2, \tag{1.5}$$

and

$$-t_2 \|g_{k+1}\|^2 \leq g_{k+1}^T d_{k+1} \leq -t_1 \|g_{k+1}\|^2, \tag{1.6}$$

where $0 < t_1 < 1 < t_2$ are constants. Dai (2002) proposed another descent line search as follows:

$$f(x_{k+1}) \leq f(x_k) + \delta \alpha g_k^T d_k, \tag{1.7}$$

and

$$g_{k+1}^T d_{k+1} \leq -\sigma_2 \|d_{k+1}\|^2, \tag{1.8}$$

where $t \in (0, 1)$, $\delta > 0$, $\sigma_2 \in (0, 1)$ and $\alpha_k = t^m$, $m > 0$. When the above two line searches are used, the PRP method satisfies the property of convergence for general functions. (1.5), (1.6) or (1.7), (1.8) will require more time to compute gradient evaluations compared with the Aromijo line search, so Zhou and Li (2014) introduced a non-descent backtracking-type line search in a different way from that in Grippo and Lucidi (1997), Dai (2002). For given constants $\tau > 0$, $\mu > 0$, $\delta \in (0, 1)$, $t \in (0, 1)$, and $\sigma_k = \min\{\tau, \frac{\mu \|g_k\|^2}{\|d_k\|^2}\}$, let $\alpha_k = \max\{\sigma_k t^j; \ j = 0, 1, \ldots\}$ satisfy

$$f(x_{k+1}) \leq f(x_k) - \delta \|\alpha_k g_k\|^2 + \eta_k \min\{1, \|g_k\|^2\}, \tag{1.9}$$

where η_k is a positive sequence satisfying

$$\sum_{k=0}^{\infty} \eta_k \le \eta < \infty, \tag{1.10}$$

where η is a positive constant. It is easy to see that the line search technique (1.9) is well defined and does not compute other gradient evaluations except g_k regardless of whether d_k is a descent direction or not. Based on the above discussion, the modified PRP method that we present has the following attributes.

- The modified PRP method possesses the information about function values.
- Strongly global convergence and R-linear convergence are established.
- The numerical results demonstrate that the modified PRP method is competitive with the normal PRP method for the given problems.
- The modified PRP method is applied to the engineering Muskingum model and image restoration problems.

The present paper is organized as follows. In Sect. 2, we show that the modified PRP method has strong convergence and a locally R-linear convergence rate for general functions under the non-descent backtracking line search. In Sect. 3, we perform some numerical experiments to compare the performance of the classical PRP method and modified PRP method with some of the line search techniques mentioned above.

2 Convergence properties

Based on the above line search technique and the modified PRP formula, we present a modified PRP algorithm, which is listed as follows.

Algorithm (Modified PRP Algorithm)

Step 1 Choose an initial point $x_0 \in \Re^n$, $\tau > 0$, $\mu > 0$, $\varepsilon \in (0, 1)$, and $\delta \in (0, 1)$, $t \in (0, 1)$. Set $d_0 = -g_0 = -\nabla f(x_0)$ and a positive sequence η_k satisfying (1.10), k:=0.
Step 2 Stop if $\|g_k\| \le \varepsilon$.
Step 3 Compute the step size α_k using the non-descent line search rule (1.9).
Step 4 Let $x_{k+1} = x_k + \alpha_k d_k$.
Step 5 If $\|g_{k+1}\| \le \varepsilon$, then the modified PRP algorithm stops.
Step 6 Calculate the search direction

$$d_{k+1} = -g_{k+1} + \beta_k^{*PRP} d_k. \tag{2.1}$$

Step 7 Set k:=k+1 and go to step 3.

In the following paper, there are some indispensable assumptions for the global convergence of the algorithm on objective functions.

Assumption B (i) The level set $T_0 = \{x \mid f(x) \le f(x_0)+\eta\}$ is bounded.

(ii) In some neighbourhood N of T_0, f is differentiable, and its gradient function g is Lipschitz continuous, namely,

$$\|g(x) - g(y)\| \le L\|x - y\|, \tag{2.2}$$

where $L > 0$ is a constant and any $x, y \in N$.

Remark (i) Assumption B implies that there exists a constant $A > 0$ satisfying

$$\|g(x)\| \le A, \quad \forall x \in N. \tag{2.3}$$

Moreover, from (1.9) and (1.10) such that

$$\begin{aligned} f(x_k + \alpha_k d_k) &\le f(x_k) - \delta\|\alpha_k g_k\|^2 + \eta_k \min\{1, \|g_k\|^2\} \\ &\le f(x_k) + \eta_k \\ &\le f(x_0) + \sum_{j=0}^{\infty} \eta_j \\ &< f(x_0) + \eta, \end{aligned}$$

where $x_k \in T_0$ for all $k \ge 0$. The following useful lemmas are presented to conveniently show the strongly global convergence of the modified PRP method with the line search (1.9).

Lemma 2.1 *Let $\{\xi_k\}$ and $\{\beta_k\}$ be positive sequences satisfying $\xi_k \le (1 + \beta_k)\xi_k + \beta_k$ and $\sum_{k=0}^{\infty} \beta_k < \infty$. $\{\xi_k\}$ is able to converge.*

Proof Omitted. The proof of the above lemma is the same as the proof of Lemma 3.3 Dennis and Moré (1974). □

Lemma 2.2 *Based on Assumption B, the sequence $\{f(x_k)\}$ that satisfies the line search technique (1.9) in the algorithm converges.*

Proof We can find a constant κ such that $f(x) > \kappa$ for all $x \in T_0$ from Assumption B (i), so we have $f(x_{k+1}) - \kappa \le f(x_k) - \kappa + \eta_k$ under the line search technique (1.9), which shows that the sequence $\{f(x_k) - \kappa\}$ converges through Lemma 2.1. Then, $\{f(x_k)\}$ converges. □

Lemma 2.3 *d_k is defined by (1.3); then, we have*

$$\|d_k\| \le \lambda\|g_k\|. \tag{2.4}$$

Proof First, by (1.9), such that

$$\alpha_k\|d_k\|^2 \le \sigma_k\|d_k\|^2 \le \frac{\mu\|g_k\|^2}{\|d_k\|^2}\|d_k\|^2 = \mu\|g_k\|^2. \tag{2.5}$$

On the other hand, from (2.2), using the mean-value theorem we have

$$\begin{aligned}\rho_{k-1} &= (g_{k-1} + g_k - 2g(x_{k-1} + as_{k-1}))^T s_{k-1}\\ &\le (\|g_{k-1} - g(x_{k-1} + as_{k-1})\|\\ &\quad +\|g_k - g(x_{k-1} + as_{k-1})\|)\|s_{k-1}\|\\ &\le (La\|s_{k-1}\| + L(1-a)\|s_{k-1}\|)\|s_{k-1}\|\\ &= L\|s_{k-1}\|^2,\end{aligned} \tag{2.6}$$

where $a \in (0, 1)$. Combining (1.3), (1.4), (2.2), (2.6), (2.5), then,

$$\begin{aligned}\|d_k\| &\le \|g_k\| + \|\beta_k^{*PRP} d_{k-1}\|\\ &\le \|g_k\| + \|\beta_k^{*PRP}\|\|d_{k-1}\|\\ &\le \|g_k\| + \frac{\|g_k^T y_{k-1}\|\|d_{k-1}\|}{\|g_{k-1}\|^2} + \frac{\|g_k^T \rho_{k-1}s_{k-1}\|\|d_{k-1}\|}{\|g_{k-1}\|^2\|s_{k-1}\|^2}\\ &\le \|g_k\| + \frac{L\alpha_{k-1}\|g_k\|\|d_{k-1}\|^2}{\|g_{k-1}\|^2} + \frac{L\alpha_{k-1}\|g_k\|\|d_{k-1}\|^2}{\|g_{k-1}\|^2}\\ &\le \|g_k\| + L\mu\|g_k\| + L\mu\|g_k\| = (1+2L)\mu\|g_k\|,\end{aligned}$$

where $\lambda = (1+2L)\mu$. Hence, the proof is complete.

It is obvious that $\lim_{k\to\infty} \alpha_k\|g_k\| = 0$ holds from the line search technique (1.9) and Lemma 2.2; when combined with (2.4), we obtain

$$\lim_{k\to\infty}\|s_k\| = \lim_{k\to\infty}\alpha_k\|d_k\| \le \lim_{k\to\infty}\lambda\alpha_k\|g_k\| = 0. \tag{2.7}$$

□

Lemma 2.4 *Let Assumption B hold and the sequence $\{x_k\}$ be generated by the algorithm; then, there exists a constant $A_1 > 0$ satisfying*

$$\alpha_k \ge A_1 \min\left\{1, \frac{-g_k^T d_k}{L\|d_k\|^2 + \delta\|g_k\|^2}\right\}. \tag{2.8}$$

Proof *Case i* If $\alpha_k = \sigma_k = \min\{\tau, \frac{\mu\|g_k\|^2}{\|d_k\|^2}\}$, from (2.4), we obtain

$$\alpha_k \ge \min\left\{\tau, \frac{\mu}{\lambda^2}\right\} = A_1.$$

Hence, together with $\min\{1, \frac{-g_k^T d_k}{L\|d_k\|^2+\delta\|g_k\|^2}\} \le 1$, we obtain (2.8).

Case ii If $\alpha_k \ne \sigma_k$, let $\alpha_k^* \doteq \frac{\alpha_k}{t}$ not satisfy (1.9), namely,

$$\begin{aligned}f(x_k + \alpha_k^* d_k) &> f(x_k) - \delta\|\alpha_k^* g_k\|^2 + \eta_k \min\{1, \|g_k\|^2\}\\ &> f(x_k) - \delta\|\alpha_k^* g_k\|^2.\end{aligned} \tag{2.9}$$

By the mean value theorem and (2.2), we have

$$\begin{aligned}f(x_k + \alpha_k^* d_k) - f(x_k) &= g(x_k + \nu_k\alpha_k^* d_k)^T \alpha_k^* d_k\\ &= \alpha_k^* g_k^T d_k\\ &\quad + (g(x_k + \nu_k\alpha_k^* d_k) - g_k)^T \alpha_k^* d_k\\ &\le \alpha_k^* g_k^T d_k + L\|\alpha_k^* d_k\|^2,\end{aligned}$$

where $\nu_k \in (0, 1)$, combining with (2.9), shows that (2.8) holds with $A_1 = t$.

The above lemma gives an estimation to the step size α_k. Now, we can prove the global convergence property of the modified PRP method under the line search (1.9). □

Theorem 2.1 *Supposing that Assumption B holds, consider the modified PRP method where d_k is satisfied* (1.3). *Then,*

$$\lim_{k\to\infty}\|g_k\| = 0. \tag{2.10}$$

Proof We will obtain this result by contradiction. We suppose that (2.10) does not hold; then, there exists a constant $n_1 > 0$ and an infinite index G such that

$$\|g_k\| \ge n_1, \quad \forall k \in G. \tag{2.11}$$

By (2.2), (2.4) and $\alpha_k \le \sigma_k \le \tau$ such that

$$\begin{aligned}\|g_k\| &\le \|g_k - g_{k-1}\| + \|g_{k-1}\| \le L\alpha_{k-1}\|d_{k-1}\| + \|g_{k-1}\|\\ &\le A_2\|g_{k-1}\|,\end{aligned} \tag{2.12}$$

where $A_2 = 1 + L\tau\lambda$. from (1.3), (1.4), (2.4), (2.6), (2.7) and (2.12), we obtain

$$\begin{aligned}\|g_k + d_k\| = |\beta_k^{*PRP}|\|d_{k-1}\| &\le \frac{L\|g_k\|\|s_{k-1}\|\|d_{k-1}\|}{\|g_{k-1}\|^2}\\ &\quad + \frac{\|g_k\|\|\rho_{k-1}s_{k-1}\|\|d_{k-1}\|}{\|g_{k-1}\|^2\|s_{k-1}\|^2}\\ &\le 2LA_2\lambda\|s_{k-1}\| \to 0.\end{aligned}$$

From (2.11), we know that there exists a constant $n_2 > 0$ such that

$$\|d_k\| \geq n_2. \tag{2.13}$$

Combining with (2.7) we obtain

$$\lim_{k\to\infty} \alpha_k = 0, \quad \forall k \in G. \tag{2.14}$$

By (1.3), (2.4), and (2.8), for all $k \in G$, we have

$$\begin{aligned}\alpha_k &\geq A_1 \frac{\|g_k\|^2 - \beta^{*PRP} g_k^T d_{k-1}}{L\|d_k\|^2 + \delta\|g_k\|^2} \\ &= A_1 \frac{\|g_k\|^2}{L\|d_k\|^2 + \delta\|g_k\|^2} - A_1 \frac{\beta^{*PRP} g_k^T d_{k-1}}{L\|d_k\|^2 + \delta\|g_k\|^2} \\ &\geq \frac{A_1}{L\lambda^2 + \delta} - A_1 \frac{\beta^{*PRP} g_k^T d_{k-1}}{L\|d_k\|^2 + \delta\|g_k\|^2}. \end{aligned} \tag{2.15}$$

For the last inequality, from (2.2), (2.3), (2.4), (2.6), (2.11), and (2.13), such that

$$\begin{aligned} &A_1 \frac{|\beta^{*PRP} g_k^T d_{k-1}|}{L\|d_k\|^2 + \delta\|g_k\|^2} \leq \frac{A_1}{L n_2^2 + \delta n_1^2} \\ &\left(\frac{L\|g_k\|^2 \|s_{k-1}\| \|d_{k-1}\|}{\|g_{k-1}\|^2} + \frac{\|g_k^T \rho_{k-1} s_{k-1}\| \|g_k\| \|d_{k-1}\|}{\|g_{k-1}\|^2 \|s_{k-1}\|^2} \right) \\ &\leq \frac{2LAA_1A_2\lambda}{L n_2^2 + \delta n_1^2} \|s_{k-1}\| \to 0. \end{aligned}$$

This with (2.15) implies that (2.14) does not hold, thus contradicting (2.14). The proof is completed. □

Theorem 2.2 *Supposing that Assumption B holds, consider the modified PRP method and (1.9); then, for large k, we have a positive constant n_3 such that*

$$\alpha_i \geq n_3 \tag{2.16}$$

holds for at least half of the indices $i \in \{0, 1, 2 \ldots, k\}$.

Proof Without loss of generality, by (2.8), we have

$$\alpha_k \geq A_1 \frac{-g_k^T d_k}{L\|d_k\|^2 + \delta\|g_k\|^2},$$

using (1.3), (1.4), (1.10), and (2.4), (2.6); then,

$$\begin{aligned}\alpha_k &\geq A_1 \frac{\|g_k\|^2 - g_k^T \beta_k^{*PRP} d_{k-1}}{L\|d_k\|^2 + \delta\|g_k\|^2} \\ &\geq A_1 \frac{\|g_k\|^2}{L\|d_k\|^2 + \delta\|g_k\|^2} (1 - 2L\lambda^2 \alpha_{k-1}).\end{aligned}$$

If $\alpha_{k-1} < \frac{1}{2L\lambda^2}$, we obtain

$$\alpha_k \geq \frac{A_1}{L\lambda^2 + \delta}(1 - 2L\lambda^2\alpha_{k-1}) = n_4 - n_5\alpha_{k-1},$$

where

$$n_4 = \frac{A_1}{L\lambda^2 + \delta}, \quad n_5 = \frac{2LA_1\lambda^2}{L\lambda^2 + \delta}.$$

Then, we obtain $n_4 \leq \alpha_k + n_5\alpha_{k-1} \leq 2\max\{1, n_5\}$ $\max\{\alpha_k, \alpha_{k-1}\}$, so

$$\max\{\alpha_k, \alpha_{k-1}\} \geq \frac{n_4}{2\max\{1, n_5\}},$$

Otherwise, from $\alpha_{k-1} \geq \frac{1}{2L\lambda^2}$, we can obtain

$$\max\{\alpha_k, \alpha_{k-1}\} \geq \alpha_{k-1} \geq \frac{1}{2L\lambda^2}.$$

Thus, when $n_3 = \min\{\frac{n_4}{2\max\{1,n_5\}}, \frac{1}{2L\lambda^2}\}$, the theorem holds. □

Theorem 2.1 shows that every limit point of the sequence $\{x_k\}$ is a stationary point of f. Moreover, if the Hessian matrix at one limit point x^* is positive definite, which means that x^* is a strict local optimal solution of the problem (1.1), then the whole sequence $\{x_k\}$ converges to x^*. Hence, in the local convergence analysis, we assume that the whole sequence $\{x_k\}$ converges.

Lemma 2.5 *Assume that f is twice continuously differentiable and uniformly convex on R^n and that Assumption B holds; then, $f(x_k)$ has a unique minimal point x^*, and there exist constants $0 < m < M$ satisfying*

$$m\|x - x^*\|^2 \leq \|g(x)\|^2 \leq M\|x - x^*\|^2 \tag{2.17}$$

and

$$m\|x - x^*\|^2 \leq f(x) - f(x^*) \leq M\|x - x^*\|^2. \tag{2.18}$$

Proof Omitted. For the proof, see (Ortega and Rheinboldt 1970; Rockafellar 1970). □

Theorem 2.3 *Let f be twice continuously differentiable. Consider the modified PRP method, where d_k satisfies (1.3), and suppose that $\{x_k\}$ converging to x^* satisfies $g(x^*) = 0$ and that $\nabla^2 f(x^*)$ is positive definite. Then, there exists a constant $A_3 > 0$ and $r \in (0, 1)$ such that*

$$\|x_k - x^*\| \leq A_3 r^k. \tag{2.19}$$

Proof Since $\nabla^2 f(x^*)$ is positive definite, f is uniformly convex in some neighbourhood N_1 of x^* if $\{x_k\} \subset N_1$ and $\{x_k\}$ satisfies (2.17) and (2.18). Denote the index as follows:

$$I_1 = \{i | i \leq k, \quad \alpha_i \geq n_3\}. \tag{2.20}$$

Table 1 Test problems

Nr.	Test problem	Nr.	Test problem
1	Extended Freudenstein and Roth Function	38	ARWHEAD Function (CUTE)
2	Extended Trigonometric Function	39	ARWHEAD Function (CUTE)
3	Extended Rosenbrock Function	40	NONDQUAR Function (CUTE)
4	Extended White and Holst Function	41	DQDRTIC Function (CUTE)
5	Extended Beale Function	42	EG2 Function (CUTE)
6	Extended Penalty Function	43	DIXMAANA Function (CUTE)
7	Perturbed Quadratic Function	44	DIXMAANB Function (CUTE)
8	Raydan 1 Function	45	DIXMAANC Function (CUTE)
9	Raydan 2 Function	46	DIXMAANE Function (CUTE)
10	Diagonal 1 Function	47	Partial Perturbed Quadratic Function
11	Diagonal 2 Function	48	Broyden Tridiagonal Function
12	Diagonal 3 Function	49	Almost Perturbed Quadratic Function
13	Hager Function	50	Tridiagonal Perturbed Quadratic Function
14	Generalized Tridiagonal 1 Function	51	EDENSCH Function (CUTE)
15	Extended Tridiagonal 1 Function	52	VARDIM Function (CUTE)
16	Extended Three Exponential Terms Function	53	STAIRCASE S1 Function
17	Generalized Tridiagonal 2 Function	54	LIARWHD Function (CUTE)
18	Diagonal 4 Function	55	DIAGONAL 6 Function
19	Diagonal 5 Function	56	DIXON3DQ Function (CUTE)
20	Extended Himmelblau Function	57	DIXMAANF Function (CUTE)
21	Generalized PSC1 Function	58	DIXMAANG Function (CUTE)
22	Extended PSC1 Function	59	DIXMAANH Function (CUTE)
23	Extended Powell Function	60	DIXMAANI Function (CUTE)
24	Extended Block Diagonal BD1 Function	61	DIXMAANJ Function (CUTE)
25	Extended Maratos Function	62	DIXMAANK Function (CUTE)
26	Extended Cliff Function	63	DIXMAANL Function (CUTE)
27	Quadratic Diagonal Perturbed Function	64	DIXMAAND Function (CUTE)
28	Extended Wood Function	65	ENGVAL1 Function (CUTE)
29	Extended Hiebert Function	66	FLETCHCR Function (CUTE)
30	Quadratic Function QF1 Function	67	COSINE Function (CUTE)
31	Extended Quadratic Penalty QP1 Function	68	Extended DENSCHNB Function (CUTE)
32	Extended Quadratic Penalty QP2 Function	69	DENSCHNF Function (CUTE)
33	A Quadratic Function QF2 Function	70	SINQUAD Function (CUTE)
34	Extended EP1 Function	71	BIGGSB1 Function (CUTE)
35	Extended Tridiagonal-2 Function	72	Partial Perturbed Quadratic PPQ2 Function
36	BDQRTIC Function (CUTE)	73	Scaled Quadratic SQ1 Function
37	TRIDIA Function (CUTE)	74	Scaled Quadratic SQ2 Function

Case i If $k \in I_1$, using (1.9) and (2.16),

$$f(x_{k+1}) \leq f(x_k) - \delta\|\alpha_k g_k\|^2 + \eta_k\|g_k\|^2$$
$$\leq f(x_k) - \delta n_3^2\|g_k\|^2 + \eta_k\|g_k\|^2,$$

and $\eta_k \to 0$ since (1.10) holds, so we have $f(x_{k+1}) \leq f(x_k) - \frac{\delta n_3^2}{2}\|g_k\|^2$, following from (2.17) and (2.18), such that

$$f(x_{k+1}) - f(x^*) \leq r_0(f(x_k) - f(x^*)),$$

where $0 < r_0 = 1 - \frac{\delta n_3^2 m}{2M} < 1$.

Case ii If $k \in I/I_1$, using the line search (1.9), we obtain $f(x_{k+1}) \leq f(x_k) + \eta_k\|g_k\|^2$ from (2.17) and (2.18), such

Fig. 1 Performance profiles of these methods (CPU)

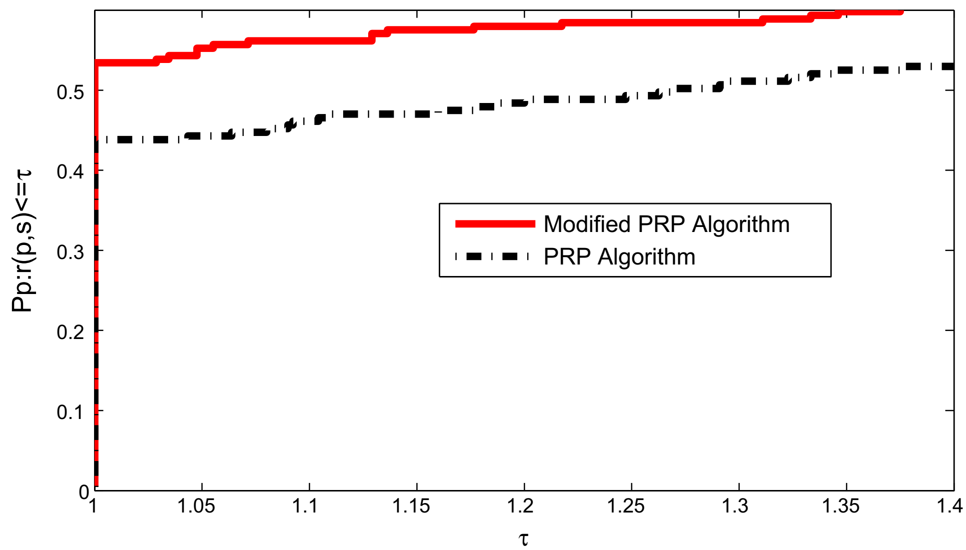

Fig. 2 Performance profiles of these methods (NI)

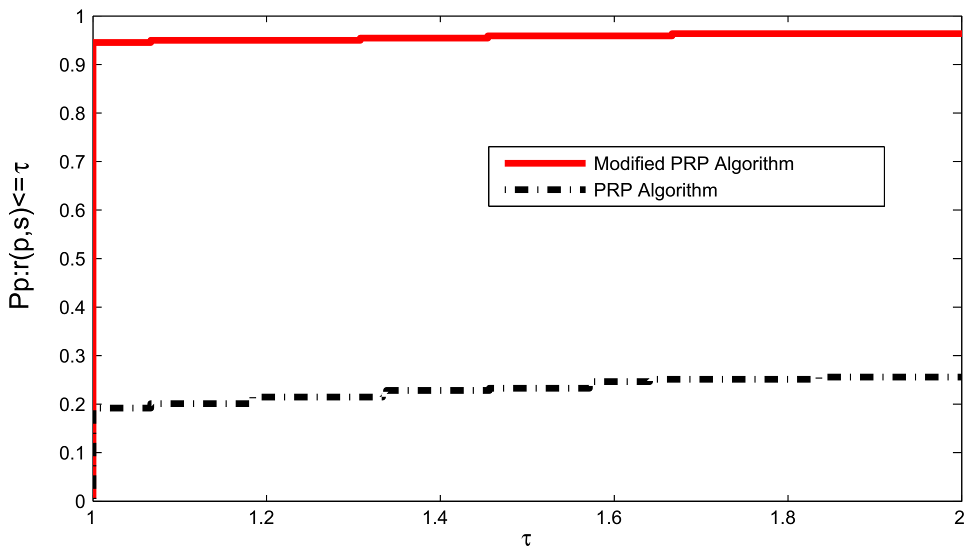

Fig. 3 Performance profiles of these methods (NFG)

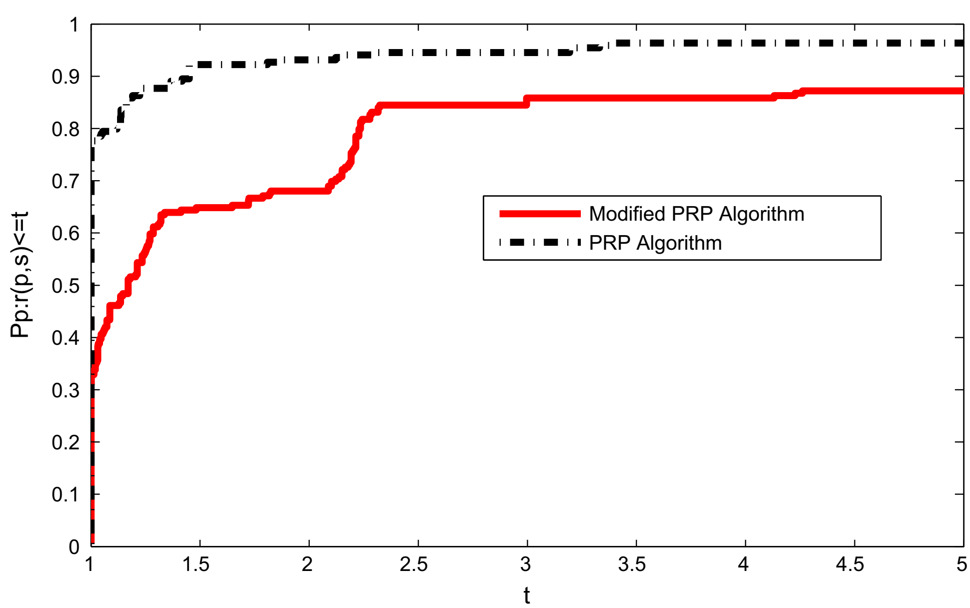

Table 2 Results of the algorithms

Algorithm	x_1	x_2	x_3
BFGS Geem (2006)	10.8156	0.9826	1.0219
HIWO Ouyang et al. (2015)	13.2813	0.8001	0.9933
Algorithm 1	4.29997	0.0000	0.0000

that

$$f(x_{k+1}) - f(x^*) \leq \left(1 + \frac{M}{m}\eta_k\right)(f(x_k) - f(x^*)),$$

we can know that there exists a constant $A_4 > 0$ satisfying the following inequality from (1.10):

$$\prod_{k=0}^{\infty}\left(1 + \frac{M}{m}\eta_k\right) \leq A_4,$$

for all large k, combining Case i and Case ii, we obtain

$$\begin{aligned} f(x_{k+1}) - f(x^*) &\leq \left(\prod_{i \in I/I_1}\left(1 + \frac{M}{m}\eta_i\right)\right)\left(r_0^{\sum_{k \in I_1}}\right) \\ &\quad \times (f(x_0) - f(x^*)) \\ &\leq A_4 r_0^{k/2}(f(x_0) - f(x^*)), \end{aligned}$$

by (2.18), such that

$$\|x_{k+1} - x^*\| \leq \left(\frac{A_4(f(x_0) - f(x^*))}{m r_0^{\frac{1}{2}}}\right)^{\frac{1}{2}} (r_0^{\frac{1}{4}})^{k+1}.$$

Then, (2.19) holds, where $A_3 = (\frac{A_4(f(x_0)-f(x^*))}{m r_0^{\frac{1}{2}}})^{\frac{1}{2}}$ and $r = r_0^{\frac{1}{4}}$. The proof is complete. □

3 Numerical experiments

In this section, we report some different numerical results for the PRP algorithm and modified PRP algorithm. Normal unconstrained optimization problems and engineering problems are included. All codes are written in MATLAB and run on a 2.30 GHz CPU with 8.00 GB of memory on the Windows 10 operating system.

3.1 Normal unconstrained optimization problems

The test problems are listed in Table 1. We will report on various numerical experiments with the modified PRP algorithm and the normal PRP algorithm with the same non-descent line search technique to demonstrate the effectiveness for the given problems. We introduce the stop rules, dimension and some parameters in the numerical experiments as follows:
Stop rules (the Himmeblau stop rule Yuan and Sun 1999): If $|f(x_k)| > e_1$, let $stop1 = \frac{|f(x_k)-f(x_{k+1})|}{|f(x_k)|}$ or $stop1 = |f(x_k) - f(x_{k+1})|$. If the conditions $\|g(x)\| < \epsilon$ or $stop1 < e_2$ are satisfied, where $e_1 = e_2 = 10^{-5}$ and $\epsilon = 10^{-6}$, the algorithm will stop if the number of iterations is greater than 1000.
Dimension: 3000, 6000, and 9000 variables.
Parameters: τ=1, μ=5, δ=0.2, and $t = 0.8$.
The columns of Table 1 have the following meanings:
Nr.: The number of tested problems.
Test problem: The name of the problem.

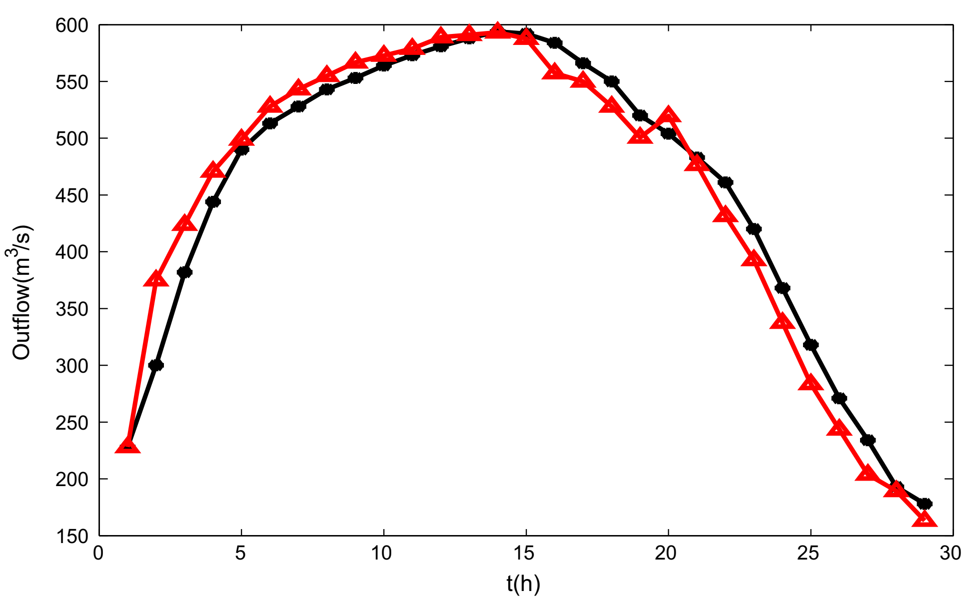

Fig. 4 Performance of Data in 19

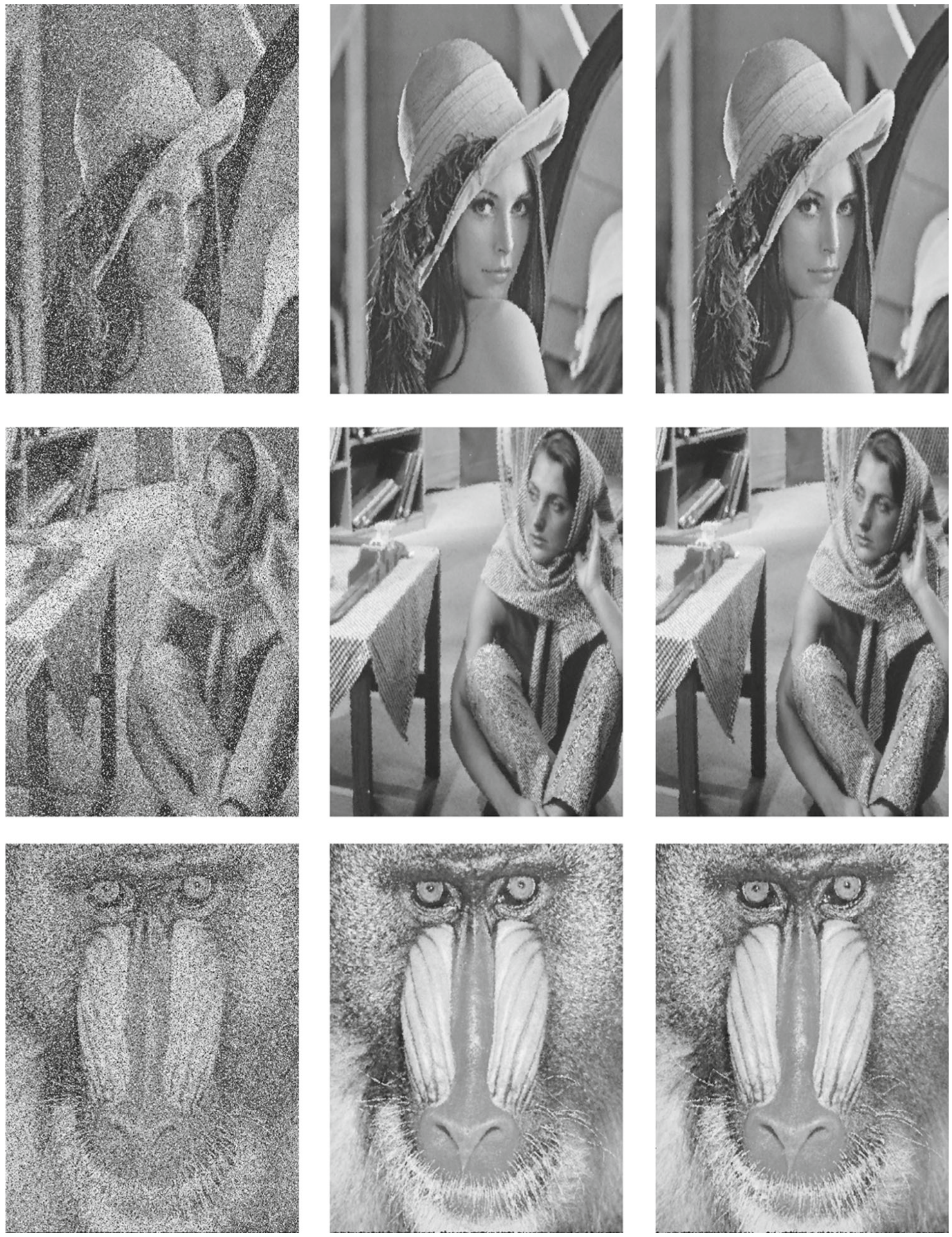

Fig. 5 Restoration of the Lena image, Baboon image, and Barbara image by the modified PRP method and the normal PRP method. From left to right: a noisy image with 30% salt-and-pepper noise and restorations obtained by minimizing z with the normal PRP method and modified PRP method

Dolan and Moré (2002) presented a new tool to show their performance in order to analyse the efficiency of these two methods, and Figs. 1, 2 and 3 show the profiles relative to the CPU time, NI, and NFG, respectively.

From Figs. 1, 2 and 3, we can see that modified PRP algorithm is more competitive than the normal method since its performance curves corresponding to the number of iterations, the total of the function and gradient evaluations, and the CPU time are best in the three figures. The modified PRP algorithm can successfully solve most of the test problems. Altogether, it is clear that the modified method is efficient based on the experimental results. The modified PRP algorithm is slightly more robust than the normal PRP algorithm in terms of the CPU time, as shown in Fig. 1. In further work, from the number of iterations in Fig. 2 and , it is not difficult to see that the modified PRP algorithm performs best between the two methods. However, from the total of the function and gradient values of the methods in Fig. 3, The modified PRP algorithm is not very good di?erent from the performances of Figs. 1 and 2. But that does not change the fact that we think the modified method is better than the normal method. Overall, we think that the modified method provides one of the most efficient approaches for solving unconstrained optimization problems.

Fig. 6 Restoration of the Lena image, Baboon image, and Barbara image by the modified PRP method and the normal PRP method. From left to right: a noisy image with 55% salt-and-pepper noise and restorations obtained by minimizing z with the normal PRP method and modified PRP method

3.2 The Muskingum model in engineering problems

As is known, some optimization algorithms are considered a significant challenge in engineering problems. Many authors endeavour to design effective algorithms for solving these engineering problems. Parameter estimation is one of the important means for the study of a well-known hydrologic engineering application problem called the nonlinear Muskingum model. This subsection discusses the nonlinear Muskingum model, a common example of such an application.

Muskingum Model Ouyang et al. (2015): is defined by:

$$\min \quad f(x_1, x_2, x_3)$$

$$= \sum_{i=1}^{n-1} \Bigg((1 - \frac{\Delta t}{6}) x_1 \big(x_2 I_{i+1} + (1 - x_2) Q_{i+1}\big)^{x_3} - (1 - \frac{\Delta t}{6}) x_1 \big(x_2 I_i + (1 - x_2) Q_i\big)^{x_3} - \frac{\Delta t}{2}(I_i - Q_i) + \frac{\Delta t}{2}(1 - \frac{\Delta t}{3})(I_{i+1} - Q_{i+1}) \Bigg)^2,$$

where n denotes the total time, x_1 denotes the storage time constant, x_2 denotes the weighting factor, x_3 denotes an additional parameter, Δt is the time step at time t_i ($i = 1, 2, \ldots, n$), I_i is the observed inflow discharge and Q_i is the observed outflow discharge. In the experiment, observed data of the flood run-off process between Chenggouwan and

Fig. 7 Restoration of the Lena image, Baboon image, and Barbara image by the modified PRP method and normal PRP method. From left to right: a noisy image with 70% salt-and-pepper noise, restorations obtained by minimizing z with the normal PRP method and the modified PRP method

Linqing of Nanyunhe in the Haihe Basin, Tianjin, China, is used. In the experiment, $\Delta t = 12(h)$ and the initial point $x = [0, 1, 1]^T$ are chosen. The detailed data for I_i and Q_i in 1961 can be found in Ouyang et al. (2014). The tested results of the final points are listed in Table 2.

From Fig. 4 and Table 2, we make the following conclusions: (i) Similar to the BFGS method and the HIWO method, the modified algorithm provides a good approximation for these data, and the given algorithm is effective for the nonlinear Muskingum model. (ii) The final points are competitive with the final points of similar algorithms. (iii) The Muskingum model may have more optimum approximation points since the final points (x_1, x_2, and x_3) of the modified PRP algorithm are different from those of the BFGS method and the HIWO method.

3.3 Image restoration problems

This subsection deals with image restoration problems to recover an original image from an image corrupted by impulse noise. These problems are proven to be difficult, and they have many applications in many fields. The code

Table 3 The CPU time of algorithm 1 and the normal PRP algorithm in seconds

30% noise	Lena	Barbara	Baboon	Total
Modified PRP Algorithm	2.984375	4.281250	4.17875	11.44438
PRP algorithm	3.218750	4.359375	4.396025	11.97415
55% noise	Lena	Barbara	Baboon	Total
Modified PRP Algorithm	7.218750	11.12500	11.06250	29.40625
PRP algorithm	8.031250	11.28125	11.54688	30.85938
70% noise	Man	Barbara	Baboon	Total
Modified PRP Algorithm	9.718750	14.48438	13.18750	37.39063
PRP algorithm	9.75000	14.90625	13.64063	38.29688

will be stopped if the condition $\frac{|\|f_{k+1}\|-\|f_k\||}{\|f_k\|} < 10^{-3}$ or $\frac{\|x_{k+1}-x_k\|}{\|x_k\|} < 10^{-3}$ holds. The experiments choose Lena (256×256), Baboon(512×512) and Barbara (512×512) as the test images. The detailed performance results are given in Figs. 5, 6 and 7. It is easy to see that both of these algorithms are successful for restoring these three images. The CPU time spent is listed in Table 3 to compare these two algorithms. The proposed algorithm is competitive with other approaches in terms of CPU time.

The results from Table 3, Figs. 5, 6 and 7, indicate that the modified PRP method with a non-descent line search is effective for image restoration. This method requires approximately one minute to restore the image from a noisy image with 55% salt-and-pepper noise, but the cost is higher to restore the image from a noisy image with 70% salt-and-pepper noise, so as the salt-and-pepper noise increases, the cost to restore the image increases.

4 Conclusion

In this paper, we present a non-descent line search and prove the global convergence and R-linear convergence of the modified PRP method with this technique for nonconvex functions. The numerical results show that the modified PRP method is competitive with the normal PRP method for nonconvex optimization. For further research, we can study the proposed algorithm with other line searches such as the Yuan-Wei-Lu line search technique Yuan et al. (2017), which guarantees the convergence property; additionally, the non-descent line search likely can be applied in nonlinear equations and nonsmooth convex minimization. Moreover, more numerical experiments for large practical problems should be performed in the future.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant No. 11661009), the High Level Innovation Teams and Excellent Scholars Program in Guangxi institutions of higher education (Grant No. [2019]52)), the Guangxi Natural Science Key Fund (No. 2017GXNSFDA198046), and the Special Funds for Local Science and Technology Development Guided by the Central Government (No. ZY20198003).

Conflict of interest The authors declare that there no conflicts of interest regarding the publication of this paper.

References

Birgin EG, Martínez JM (2001) A spectral conjugate gradient method for unconstrained optimization. Appl Math Optim 43:117–128
Cohen A (1972) Rate of convergence of several conjugate gradient algorithms. SIAM J Numer Anal 9:248–259
Cheng W (2007) A two-term PRP-based descent method. Numer Funct Anal Optim 28:1217–1230
Dai Y (2001) New properties of nonlinear conjugate gradient algorithms. J Numer Math 89:83–98
Dai Y (2002) Conjugate gradient methods with Armijo-type line searches. Acta Math Appl Sin (English Series) 18:123–130
Dolan ED, Moré JJ (2002) Benchmarking optimization software with performance profiles. Math Program 91:201–213
Dennis JE, Moré JJ (1974) A characterization of superlinear convergence and its applications to quasi-Newton methods. Math Comput 28:549–560
Dai Y, Yuan Y (1999) A nonlinear conjugate gradient with a strong global convergence property. SIAM J Optim 10(1):177–182
Fletcher R (1997) Practical method of optimization, vol I: unconstrained optimization. Wiley, New York
Fletcher R, Reeves CM (1964) Function minimization by conjugate gradients. J Comput 7:149–154
Gilbert JC, Nocedal J (1992) Global convergence properties of conjugate gradient methods for optimization. SIAM J Optim 2:21–42
Grippo L, Lucidi S (1997) A globally convergent version of the Polak-Ribière-Polyak conjugate gradient method. Math Program 78:375–391
Geem ZW (2006) Parameter estimation for the nonlinear Muskingum model using the BFGS technique. J Irrig Drain Eng 132:474–478
Hestenes MR, Stiefel E (1952) Method of conjugate gradient for solving linear equations. J Res Natl Bureau Stand 49:409–436
Khoda KM, Liu Y, Storey C (1992) Generalized Polak–Ribière algorithm. J Optim Theory Appl 75:345–354
Li X, Zhao X, Duan X (2015) A conjugate gradient algorithm with function value information and N-step quadratic convergence for unconstrained optimization. PLoS ONE 10(9):e137166
Liu Y, Storey C (2000) Effcient generalized conjugate gradient algorithms part 1: theory. J Optim Theory Appl Math 10:177–182
Nocedal J, Wright SJ (2006) Numberical optimization, 2nd edn. Springer Series in Operations Research. Springer, New York

Ouyang A, Liu L, Sheng Z, Wu F (2015) A class of parameter estimation methods for nonlinear Muskingum model using hybrid invasive weed optimization algorithm. Math Probl Eng 2015:15 (**Article ID 573894**)

Ouyang A, Tang Z, Li K, Sallam A, Sha E (2014) Estimating parameters of Muskingum model using an adaptive hybrid PSO algorithm. Int J Pattern Recogn Artif Intell 28:29 (**Article ID 1459003**)

Ortega JM, Rheinboldt WC (1970) Iterative solution of nonlinear equation in seveal variables. Academic Press, Cambridge

Polak E, Ribière G (1969) Note Sur la convergence de directions conjugèes. Rev. Fr Inf Rech Operationelle, 3e Annèe. 16:35–43

Polyak BT (1969) The conjugate gradient method in extreme problems. USSR Comput Math Math Phys 9:94–112

Powell MJD (1984) Nonconvex minimization calculations and the conjugate gradient method, vol 1066. Lecture Notes in Mathematics. Spinger, Berlin

Powell MJD (1986) Convergence properties of algorithm for nonlinear optimization. SIAM Rev 28:487–500

Rockafellar RT (1970) Convex analysis. Princeton University Press, Princeton

Shanno DF (1978) Conjugate gradient methods with inexact line searches. Math Oper Res 3:244–256

Shi Z (2002) Restricted PR conjugate gradient method and its global convergence. Adv Math 31(1):47–55

Yuan G, Li T, Hu W (2020) A conjugate gradient algorithm for large-scale nonlinear equations and image restoration problems. Appl Numer Math 147:129–141

Yuan G, Lu J, Wang Z (2020) The PRP conjugate gradient algorithm with a modified WWP line search and its application in the image restoration problems. Appl Numer Math 152:1–11

Yuan G, Wei Z (2010) Convergence analysis of a modified BFGS method on convex minimizations. Comput Optim Appl 47(2):237–255

Yuan G, Wei Z, Lu X (2017) Global convergence of the BFGS method and the PRP method for general functions under a modified weak Wolfe-Powell line search. Appl Math Modell 47:811–825

Yuan Y (1993) Analysis on the conjugate gradient method. Optim Methematics Softw 2:19–29

Yuan Y, Sun W (1999) Theory and methods of optimization. Science Press of China, Beijing

Zhou W, Li D (2014) On the convergence properties of the unmodified PRP method with a non-descent line search. Optim Methods Softw 29:484–496