



Adaptive three-term PRP algorithms without gradient Lipschitz continuity condition for nonconvex functions

Gonglin Yuan¹ · Heshu Yang¹  · Mengxiang Zhang¹

Received: 31 May 2021 / Accepted: 6 January 2022 / Published online: 20 January 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

At present, many conjugate gradient methods with global convergence have been proposed in unconstrained optimization, such as MPRP algorithm proposed by Zhang et al. (*IMA J. Numer. Anal.* **26**(4):629–640, 2006). Unfortunately, almost all of these methods require gradient Lipschitz continuity condition. As far as we know, how do the current conjugate gradient methods deal with gradient non-Lipschitz continuity problems is basically blank. For gradient non-Lipschitz continuity problems, Algorithm 1 and Algorithm 2 are proposed in this paper based on MPRP algorithm. The proposed algorithms have the following characteristics: (i) Algorithm 1 retains sufficient descent property independent of line search technology in MPRP algorithm; (ii) for nonconvex and gradient non-Lipschitz continuous functions, the global convergence of Algorithm 1 is obtained in combination with the trust region property and the weak Wolfe-Powell line search technique; (iii) based on Algorithm 1, Algorithm 2 is further improved which global convergence can be obtained independently of line search technique; (iv) according to numerical experiments, the proposed algorithms perform competitively with other similar algorithms.

Keywords Conjugate gradient · Nonconvex functions · Sufficient descent property · Global convergence · Gradient Lipschitz continuity condition

Mathematics Subject Classifications (2010) 90C26

✉ Heshu Yang
yangheshuy@163.com

Gonglin Yuan
glyuan@gxu.edu.cn

Mengxiang Zhang
mxzhang@st.gxu.edu.cn

¹ College of Mathematics and Information Science, Guangxi University, Center for Applied Mathematics of Guangxi, Nanning, Guangxi, 530004, China

1 Introduction

Consider unconstrained optimization problems

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

where the objective function $f : R^n \rightarrow R$ is continuously differentiable. The conjugate gradient (CG) algorithm for (1.1) has the iteration formula:

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, 2, \dots, \quad (1.2)$$

where x_k is the current iteration point, α_k is the step size, and d_k is the search direction, which is defined by

$$d_k = \begin{cases} -g_k + \beta_{k-1} d_{k-1}, & \text{if } k \geq 1, \\ -g_k, & \text{if } k = 0, \end{cases} \quad (1.3)$$

where $g_k = \nabla f(x_k)$ is the gradient of $f(x)$ at x_k , and $\beta_k \in R$ is a CG scalar. Various well-known conjugate gradient methods (see [4–6, 9, 15–17]), including the Fletcher-Reeves (FR) method [6], the Polak-Ribière-Polyak (PRP) method [16, 17], the Hestenes-Stiefel (HS) method [9], the Dai-Yuan (DY) method [4] and the conjugate descent method [5]. Zoutendijk [42] proved the global convergence of the FR method for general functions under an exact line search. In practice, the FR method [6] may produce a continuous small step length, which will lead to poor numerical performance despite its satisfactory convergence. The global convergence of the DY method [4] and the CD method [5] has been proven, and the global convergence of the DY method [4] can be proven without descent condition. However, the performances of the DY method and the CD method are weaker than those of the PRP method [16, 17] and the HS method [9], where the performance of the HS method is similar to that of the PRP method. Furthermore, the PRP method has a built-in restart feature that can automatically adjust β_k to solve the interference problem. Therefore, we are more interested in the PRP formula with

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

where $\|\cdot\|$ denotes the Euclidean norm of vectors and $y_{k-1} = g_k - g_{k-1}$. Many scholars have studied the convergence (see [2, 3, 17–19, 35, 41] et al.) of the PRP method. Polark and Ribière show that for convex problems, the global convergence of the PRP method can be proven using an exact line search [16, 17]. However, for the nonconvex problems, Powell [19] gave a three-dimensional counterexample and pointed out that the convergence of the PRP method could not be guaranteed even if an exact line search was adopted. Based on the above work, the research results of Gilbert and Nocedal [7] show that if β_k^{PRP} is strictly non-negative and the step size α_k is determined by a line search step that satisfies sufficient descent, then the PRP method has global convergence.

It is well known that sufficient descent is a very important property for CG algorithms which has the form

$$d_k^T g_k \leq -c \|g_k\|^2, \quad (1.4)$$

where $c > 0$ is a constant and this property has been proven to play a critical role in analyzing the global convergence of CG methods (see [7, 10, 36, 39] etc.). To ensure that the sufficient descent condition is satisfied and establish the convergence of the PRP method, the following weak Wolfe-Powell (WWP) inexact line search technique is proposed in optimization algorithms to find α_k such that

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k \tag{1.5}$$

and

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k, \tag{1.6}$$

where $\delta \in (0, \frac{1}{2})$ and $\sigma \in (\delta, 1)$.

The above work is particularly important for the global convergence of algorithms, and the gradient Lipschitz continuity condition is usually used as a hypothetical condition in the convergence proof for each algorithm. Next, we will analyze the convergence of various classical CG formulas and the use of the gradient Lipschitz continuity condition in the corresponding proofs.

The FR method [1] proves global convergence under the strong Wolfe line search, while generating sufficient descent directions. In the global convergence proof of the FR method, the following three conditions are required: (i) the objective function f is continuously differentiable; (ii) the level set $S = \{x \in R^n : f(x) \leq f(x_1)\}$ is bounded; (iii) the gradient function $g(x)$ satisfies the gradient Lipschitz continuity condition. According to the above conditions, we can get $\liminf_{k \rightarrow \infty} \|g_k\| = 0$.

The DY method [4] is a classic CG formula that always generates descent directions under a standard Wolfe line search. The global convergence of the DY method is assumed under two conditions: (i) f is bounded below on R^n and is continuously differentiable in a neighborhood N of the level set $S = \{x \in R^n : f(x) \leq f(x_1)\}$; (ii) the gradient $\nabla f(x)$ is Lipschitz continuous in N , i.e., there exists a constant $L > 0$ such that $\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$ for any $x, y \in N$. In addition, the DY method satisfies the Zoutendijk condition with

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

under the assumption of the gradient Lipschitz continuity condition.

In [16, 17], the PRP method can converge at a square rate, and its finite-dimensional smooth objective function $f(x)$ needs to satisfy the requirement that $f(x)$ is a twice differentiable strongly convex function, with a bounded second derivative that satisfies gradient Lipschitz continuity condition. Then, we can obtain

$$\|x^{kN} - x^*\| \leq cq^{2k}, q < 1.$$

Moreover, the gradient Lipschitz continuity condition is used in the proof of the above inequality. This also indicates that the PRP method with an exact line search is globally convergent under the assumption that f is strongly convex.

The method that was proposed by Hager and Zhang (see [10]) in 2005 is also a well-known CG formula. This method is considered to satisfy either the Goldstein conditions [8]

$$\delta_1 \alpha_k g_k^T d_k \leq f(x_k + \alpha_k d_k) - f(x_k) \leq \delta_2 \alpha_k g_k^T d_k,$$

where $0 < \delta_2 < \frac{1}{2} < \delta_1 < 1$ and $\alpha_k > 0$, or the Wolfe conditions [25, 26]. Assuming that the objective function f is strongly convex and the gradient function satisfies gradient Lipschitz continuity condition, global convergence can be achieved. Moreover, if the line search satisfies the Goldstein conditions, then

$$\alpha_k \geq \frac{(1 - \delta_1) |g_k^T d_k|}{L \|g_k\|^2},$$

and if the line search satisfies the Wolfe conditions [25, 26], then

$$\alpha_k \geq \frac{(1 - \sigma) |g_k^T d_k|}{L \|g_k\|^2}.$$

Inspired by the above observations, many scholars have proposed modified CG methods based on the PRP formula (see [7, 10, 11, 14, 27–32, 36, 38, 40] etc.). An improved three-term Polak-Ribière-polyak (MPRP) method proposed by Zhang et al. is one of the good results. Contradiction proof is used in the MPRP algorithm for global convergence, and $\|d_k\| \leq M_d$ (M_d is a positive constant) is obtained in the proof process, which makes us naturally associate with the trust region property. It is well known that the trust region method is an effective method for optimization problems ([13, 20–24, 33, 34] etc.), which has desirable theoretical properties, i.e., it possesses very strong convergent features. The subproblem model of the trust region to find a trial step d_k is determined by

$$\begin{aligned} \min_{x \in \mathbb{N}^n} \quad & g_k^T d_k + \frac{1}{2} d^T B_k d = q_k(d), \\ \text{s.t.} \quad & \|d_k\| \leq \Delta_k. \end{aligned}$$

where Δ_k is called trust region radius. According to the subproblem model, we think that d_k has the trust region feature if it can ensure that the following relation

$$\|d_k\| \leq c_M \|g_k\|. \quad (1.7)$$

holds, where c_M is a positive constant. Inspired by MPRP algorithm, we consider to prove the global convergence of algorithms combined with the trust region properties. However, it is well known that the gradient Lipschitz continuity condition is still needed to prove the global convergence of MPRP algorithm. The following question is considered: can the global convergence proceed smoothly without gradient Lipschitz continuity condition? At present, few scholars have obtained good results [12, 37]. Based on the MPRP method, this paper gives a positive answer, and the following results are obtained:

- The sufficient descent property (1.4) of MPRP algorithm independent of line search technology is retained in Algorithm 1.
- For nonconvex and gradient non-Lipschitz continuous functions, the global convergence of Algorithm 1 is established by combining trust region properties and using the weak Wolfe-Powell line search technique.
- Algorithm 2 is further improved from Algorithm 1, and the global convergence can be obtained independently of the line search technique.
- The proposed algorithms perform competitively with other similar algorithms.

The motivation and Algorithm 1 are presented in the next section. Section 3 introduces the global convergence of Algorithm 1 under the WWP line search, as well as the sufficient descent property. In Section 4, Algorithm 2 is analyzed and the global convergence of Algorithm 2 is given without gradient Lipschitz continuity condition. In Section 5, the numerical results of image restoration are introduced by A-PRP-A algorithm, A-PRP-A-W algorithm, MPRP algorithm and Hager-Zhang algorithm.

2 Motivation and algorithm

Sufficient descent property plays an important role in the proof of the global convergence of the CG algorithm. The first three-term CG formula proposed by Zhang et al. [43] is called the MPRP method which have the sufficient descent property described above with $d_k^T g_k = -\|g_k\|^2$ independent of any line search. The search direction d_k iterative formula is as follows

$$d_k = \begin{cases} -g_k, & \text{if } k = 0, \\ -g_k + \frac{g_k^T y_{k-1} d_{k-1} - g_k^T d_{k-1} y_{k-1}}{\|g_{k-1}\|^2}, & \text{if } k \geq 1. \end{cases} \tag{2.1}$$

The global convergence of MPRP algorithm is proved under Armijo line search technique and gradient Lipschitz continuity condition. In addition, in the process of proof using the contradiction, assuming $\|g_k\| \geq \epsilon > 0$, you can obtain

$$\|d_k\| \leq M_d,$$

where M_d is a positive constant. Inspired by the form of the above formula, we consider combining trust region property in the proof of global convergence and expect to obtain the trust region property of the following form

$$\|d_k\| \leq c_M \|g_k\|$$

holds, where c_M is a positive constant. According to the above, for gradient non-Lipschitz continuous functions, we present a new formula on the basis of formula (2.1), and combine the trust region properties to prove global convergence. The search direction d_k iterative formula is as follows:

$$d_k = \begin{cases} -g_k, & \text{if } k = 0, \\ -g_k + \alpha_{k-1} \frac{g_k^T y_{k-1} d_{k-1} - g_k^T d_{k-1} y_{k-1}}{\|g_{k-1}\|^2}, & \text{if } k \geq 1, \end{cases} \tag{2.2}$$

In this formula, prefix α_{k-1} is added to the formula (2.1) and the specified formula is called an adaptive three-term CG formula. This modification is helpful for the implementation of the algorithm in this paper. The sufficient descent property, global convergence and other properties of this method will be proven in the next section. Based on the above discussions, Algorithm 1 is presented as follows.

Algorithm 1: An adaptive three-term PRP algorithm (A-T-PRP-A)

Step 0: Given any initial point $x_0 \in R^n$, $d_0 = -g_0$, constants $\epsilon \in (0, 1)$, $\delta \in (0, \frac{1}{2})$, $\sigma \in (\delta, 1)$, $\mu > 0$, and set $k = 0$.

- Step 1: Stop if $\|g_k\| \leq \epsilon$ holds.
- Step 2: Find α_k satisfying the WWP line search (1.5) and (1.6).
- Step 3: The next point $x_{k+1} = x_k + \alpha_k d_k$ is defined.
- Step 4: Stop if $\|g_{k+1}\| \leq \epsilon$ is true.
- Step 5: Compute the conjugate gradient direction by (2.2).
- Step 6: Set $k = k + 1$, and go to Step 2.

3 The global convergence of Algorithm 1

In this section, we will prove that Algorithm 1 is globally convergent. First, the sufficient descent property of the new formula will be demonstrated.

Lemma 1 *The direction is generated by Algorithm 1 where $g_k^T d_k = -\|g_k\|^2$, i.e., Algorithm 1 has sufficient descent (1.4).*

Proof Algorithm 1 generates the sequence $\{x_k, d_k, \alpha_k, g_k\}$, then, if $k = 0, d_0 = -g_0$. In this case, Algorithm 1 has sufficient descent (1.4).

If $k \geq 1$, the following relation can be obtained from the iterating formula (2.2):

$$\begin{aligned} g_k^T d_k &= g_k^T (-g_k + \alpha_{k-1} \frac{g_k^T y_{k-1} d_{k-1} - g_k^T d_{k-1} y_{k-1}}{\|g_{k-1}\|^2}) \\ &= -\|g_k\|^2 + g_k^T \frac{g_k^T y_{k-1} (x_k - x_{k-1}) - g_k^T (x_k - x_{k-1}) y_{k-1}}{\|g_{k-1}\|^2} \\ &= -\|g_k\|^2. \end{aligned}$$

Then sufficient descent property (1.4) holds. The proof is complete. □

Remark 3.1 (1) It is not difficult to observe that the direction has the property of sufficient descent and requires no additional assumptions.

(2) From the above results, the following relation can be obtained:

$$-\|d_k\| \|g_k\| \leq g_k^T d_k = -\|g_k\|^2,$$

this means that

$$\|g_k\| \leq \|d_k\|, \forall k. \tag{3.1}$$

The following assumption is needed to prove the global convergence of Algorithm 1.

Assumption (I). The level set $S = \{x \mid f(x) \leq f(x_0)\}$ is defined and bounded, and the nonconvex function $f(x)$ is continuously differentiable and bounded below.

Remark 3.2 (1) Under Assumption I, we conclude that there exists a constant $c_0 > 0$ satisfying

$$\|x_k\| \leq c_0, \forall k.$$

(2) From Assumption I, the nonconvex function $f(x)$ is continuously differentiable then the gradient function $g(x)$ is continuous. In addition, the level set S is

bounded; therefore, the function $f(x)$ is bounded in the level set S . And the gradient function $g(x)$ is continuous, that is, the gradient function $g(x)$ is bounded in the level set S . Namely there exists a constant $M_g > 0$ such that

$$\|g(x)\| \leq M_g, \forall x \in R^n. \tag{3.2}$$

- (3) Note, gradient non-Lipschitz continuous functions are common. We give a simple example where the original function $f(x) = x^{\frac{3}{2}}$, $x \in [0, +\infty]$, the gradient function $g(x) = \frac{3}{2}x^{\frac{1}{2}}$, and at $x = 0$ the gradient does not satisfy gradient Lipschitz continuity condition.

Theorem 1 *Suppose that Assumption 1 is satisfied and sequence $\{x_k, d_k, \alpha_k, g_k\}$ is generated by Algorithm 1, then, we get*

$$\liminf_{k \rightarrow \infty} \|g_k\| = 0. \tag{3.3}$$

Proof In this proof, global convergence is obtained by using contradiction. First of all, suppose

$$\|g_k\| \geq \varepsilon_0 \tag{3.4}$$

holds. The global convergence of Algorithm 1 is proved below. Using (1.5) and the relation (1.4) of Lemma 1, we have

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k \leq f(x_k) - \delta \alpha_k \|g_k\|^2,$$

and

$$\delta \alpha_k \|g_k\|^2 \leq f(x_k) - f(x_k + \alpha_k d_k).$$

Summing the above inequalities from $k = 0$ to ∞ , we obtain

$$\delta \sum_{k=0}^{\infty} \alpha_k \|g_k\|^2 < \infty,$$

then

$$\lim_{k \rightarrow \infty} \alpha_k \|g_k\|^2 = 0. \tag{3.5}$$

Therefore, the following relation can be obtained from (3.4) and (3.5)

$$\alpha_k \rightarrow 0, k \rightarrow \infty.$$

In the next place, similar to the proof of Lemma 1, if $k = 0$, $d_0 = -g_0$. Then,

$$\|d_0\| = \|g_0\|. \tag{3.6}$$

If $k \geq 1$, it can be deduced from formulas (3.4) and Remark 3.2

$$\begin{aligned} \|d_k\| &= \left\| -g_k + \alpha_{k-1} \frac{g_k^T y_{k-1} d_{k-1} - g_k^T d_{k-1} y_{k-1}}{\|g_{k-1}\|^2} \right\| \\ &= \left\| -g_k + \frac{g_k^T y_{k-1} (x_k - x_{k-1}) - g_k^T (x_k - x_{k-1}) y_{k-1}}{\|g_{k-1}\|^2} \right\| \\ &\leq \|g_k\| + \frac{2\|g_k\| \|y_{k-1}\| \|x_k - x_{k-1}\|}{\|g_{k-1}\|^2} \\ &\leq \|g_k\| + \frac{2\|g_k\| (\|g_k\| + \|g_{k-1}\|) \cdot 2c_0}{\|g_{k-1}\|^2} \\ &\leq \left(1 + \frac{8c_0 M_g}{\epsilon_0^2}\right) \|g_k\|. \end{aligned}$$

Let $M = 1 + \frac{8c_0 M_g}{\epsilon_0^2} > 0$. According to the above derivation and (3.6), that is to say

$$\|d_k\| \leq M \|g_k\|, \tag{3.7}$$

where $k \geq 0$. In addition, according to (3.1) in Remark 3.1, we can infer that

$$\|g_k\| \leq \|d_k\| \leq M \|g_k\|.$$

From the above derivation, the level set is bounded means that

$$x_{k_j} \rightarrow x^*, j \rightarrow \infty,$$

where $\{x_{k_j}\}$ is a convergent sequence in x_k . By Assumption I, the objective function $f(x)$ is continuously differentiable. Therefore, for all $\epsilon_1 > 0$, there exists an integer $N_1 > 0$ such that

$$\|g(x_{k_j}) - g(x^*)\| < \epsilon_1, \forall j > N_1. \tag{3.8}$$

Using (3.7), for all $\epsilon_2 > 0$, there exists an integer $N_2 > 0$ that satisfies

$$\|d(x_{k_j}) - d(x^*)\| < \epsilon_2, \forall j > N_2. \tag{3.9}$$

From (3.8), (3.9) and Remark 3.1(2), we obtain

$$g(x^*)^T d(x^*) \leq -\|g(x^*)\|^2 < -c\epsilon_0 < 0. \tag{3.10}$$

According to the second relation (1.6) of the WWP line search, the following relation can be obtained

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

this means that

$$g_{k+1}^T d_k - \sigma g_k^T d_k \geq 0,$$

Let $N = \max\{N_1, N_2\}$, when $k > N$, we can deduce that

$$\lim_{k \rightarrow \infty} (g_{k+1}^T d_k - \sigma g_k^T d_k) = (1 - \sigma)g(x^*)^T d(x^*) \geq 0.$$

Finally, the above relation also implies that

$$g(x^*)^T d(x^*) \geq 0$$

holds. This contradicts the relation (3.10) that was obtained under our hypothesis (3.3). The proof is completed. □

- Remark 3.3* (1) According to the above proof process, the global convergence can be proven in the absence of gradient Lipschitz continuity condition.
- (2) In the process of proving the global convergence, we use contradiction, and the proof for trust region simplifies the whole proof process.

4 Algorithm 2 and its proof of convergence

In the above section, we use the WWP line search method to determine α_k and propose a new CG formula for determining the direction of descent d_k . It is not difficult to observe that the proof of global convergence is still valid in the absence of gradient Lipschitz continuity condition. This section will continue to explore and give an algorithm without a line search or the gradient Lipschitz continuity condition, and global convergence will also hold. The algorithm is presented as follows:

Algorithm 2: An adaptive three-term PRP algorithm without gradient Lipschitz continuity condition (A-T-PRP-A-W)

- Step 0: Given any initial point $x_0 \in R^n, d_0 = -g_0$, constants $\epsilon \in (0, 1), \delta \in (0, 1), \sigma_0 \in (0, 1), \mu > 0$, and set $k=0$.
- Step 1: Stop if $\|g_k\| \leq \epsilon$ holds.
- Step 2: The next point $x_{k+1} = x_k + d_k$ is defined.
- Step 3: If $f(x_{k+1}) > f(x_k) + \delta_0 g_k^T d_k$, set $f(x_{k+1}) = f(x_k) + \sigma_0 \delta_0 g_k^T d_k$.
- Step 4: Stop if $\|g_{k+1}\| \leq \epsilon$ is true.
- Step 5: Compute the conjugate gradient direction by (2.1).
- Step 6: Set $k=k+1$, and go to Step 2.

Remark 4.1 (1) The difference between Algorithm 1 and Algorithm 2 is the step 2 and the step 3. We assume that α_k is always equal to 1, and formula (2.2) will become a classical formula (2.1).

- (2) According to the Taylor formula, we obtain

$$f(x_k + d_k) = f(x_k) + g_k^T d_k + O(\|d_k\|^2).$$

Using Lemma 1 has $g_k^T d_k = -\|g_k\|^2$. In the third step of Algorithm 2, if we choose a sufficiently small scalar δ_0 , then $f(x_{k+1}) \leq f(x_k) + g_k^T d_k + O(\|d_k\|^2)$ will be true in most cases. This also means that case $f(x_{k+1}) = f(x_k) + \sigma_0 \delta_0 g_k^T d_k$ might not hold. Therefore, it can be divided into the following two situations with

Case(i): $f(x_{k+1}) \leq f(x_k) + \delta_0 g_k^T d_k$ and **Case(ii):** $f(x_{k+1}) = f(x_k) + \sigma_0 \delta_0 g_k^T d_k$.

- (3) In the following we set $f^b(x_{k+1}) = f(x_{k+1})$ in Case(ii)

Lemma 2 Suppose that Assumption 1 holds, for Case(ii), we can deduce that $f^b(x_k)$ is bounded below.

Proof By the definition of f^b , we have

$$f^b(x_{k+1}) = f(x_k) + \sigma_0 \delta_0 g_k^T d_k = f(x_k) - \sigma_0 \delta_0 \|g_k\|^2.$$

Then

$$f(x_k) - \sigma_0 \delta_0 \beta^2 \leq f^b(x_{k+1}) \leq f(x_k).$$

Since $f(x)$ is bound below, the result of this lemma is true. The proof is complete. □

Theorem 2 *Suppose that Assumption I hold, then we get*

$$\lim_{k \rightarrow \infty} \|g_k\| = 0. \tag{4.1}$$

Proof This theorem will be proved in the following four situations.

Situation I: Case(i) always holds. According to Assumption I, (1.4), we obtain

$$f(x_{k+1}) \leq f(x_k) + \delta_0 g_k^T d_k = f(x_k) - \delta_0 \|g_k\|^2,$$

then

$$\delta_0 \|g_k\|^2 \leq f(x_k) - f(x_{k+1}).$$

Summing the above inequalities from $k = 0$ to ∞ and using Assumption I, we get

$$\delta_0 \sum_{k=0}^{\infty} \|g_k\|^2 \leq f_0 - f_{\infty} < \infty,$$

so (4.1) is true.

Situation II: Case(ii) always holds. According to the above definition and (1.4), we obtain

$$\begin{aligned} f^b(x_1) &= f(x_0) + \sigma_0 \delta_0 g_0^T d_0 = f(x_0) - \sigma_0 \delta_0 \|g_0\|^2, \\ f^b(x_2) &= f^b(x_1) + \sigma_0 \delta_0 g_1^T d_1 = f^b(x_1) - \sigma_0 \delta_0 \|g_1\|^2, \\ f^b(x_3) &= f^b(x_2) + \sigma_0 \delta_0 g_2^T d_2 = f^b(x_2) - \sigma_0 \delta_0 \|g_2\|^2, \\ f^b(x_4) &= f^b(x_3) + \sigma_0 \delta_0 g_3^T d_3 = f^b(x_3) - \sigma_0 \delta_0 \|g_3\|^2, \\ &\dots, \\ f^b(x_{k+1}) &= f^b(x_k) + \sigma_0 \delta_0 g_k^T d_k = f^b(x_k) - \sigma_0 \delta_0 \|g_k\|^2, \\ &\dots, \end{aligned}$$

summing up the above relations obtains

$$\sigma_0 \delta_0 \sum_{k=0}^{\infty} \|g_k\|^2 = f(x_0) - f^b(x_{\infty}) < \infty,$$

From Lemma 2 and Assumption I, it can also be inferred that (4.1) holds.

Situation III: Case(i) and Case(ii) both hold. It has

$$\begin{aligned}
 f^b(x_1) &= f(x_0) + \sigma_0 \delta_0 g_0^T d_0 = f(x_0) - \sigma_0 \delta_0 \|g_0\|^2, \\
 f(x_2) &\leq f^b(x_1) + \delta_0 g_1^T d_1 \leq f^b(x_1) - \sigma_0 \delta_0 \|g_1\|^2, \\
 f^b(x_3) &= f(x_2) + \sigma_0 \delta_0 g_2^T d_2 = f(x_2) - \sigma_0 \delta_0 \|g_2\|^2, \\
 f(x_4) &= f^b(x_3) + \delta_0 g_3^T d_3 = f^b(x_3) - \sigma_0 \delta_0 \|g_3\|^2, \\
 f^b(x_5) &= f(x_4) + \sigma_0 \delta_0 g_4^T d_4 = f(x_4) - \sigma_0 \delta_0 \|g_4\|^2, \\
 &\dots,
 \end{aligned}$$

similar to Situation I and II, we can also obtain

$$\sigma_0 \delta_0 \sum_{k=0}^{\infty} \|g_k\|^2 < \infty,$$

which means that (4.1) is true.

Situation IV: Case(i) and Case(ii) hold with irregular alternation. By a way similar to the above analysis, we can turn out that (4.1) is true. The proof is complete. \square

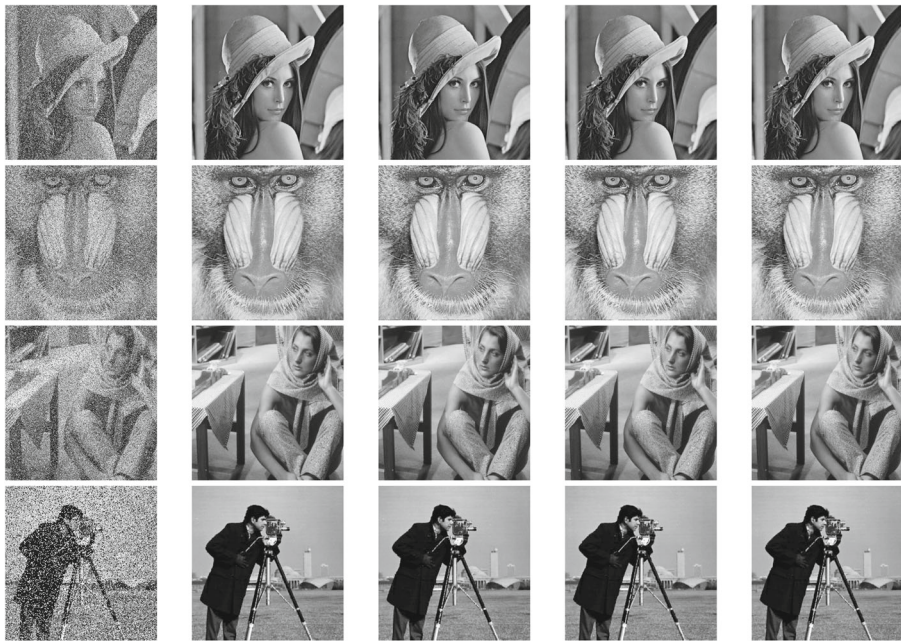


Fig. 1 From left to right, the images disturbed by 30% salt-and-pepper noise, the images restored by A-T-PRP-A, the images restored by A-T-PRP-A-W algorithm, the images restored by MPRP algorithm and the images restored by Hager-Zhang algorithm



Fig. 2 From left to right, the images disturbed by 45% salt-and-pepper noise, the images restored by A-T-PRP-A, the images restored by A-T-PRP-A-W algorithm, the images restored by MPRP algorithm and the images restored by Hager-Zhang algorithm

5 Image restoration problems

In this section, the numerical results of image restoration are introduced. This section aims at restoring the original image from an image that was damaged by impulse noise. Image restoration is of great importance in optimization experiments and is applied in many fields. All the numerical tests are run on an Intel (R) Core (TM) i5-4590 CPU @ 3.30 GHz with RAM 4.0 GB of memory on the Windows 10 operating system.

Usually, after image compression, the output image differs from the original image. In experiments, to measure the image quality after processing, we usually refer to the PSNR value to measure whether a process is satisfactory. In general, the higher the value of PSNR is, the better the quality of the output image. Therefore, PSNR is an objective standard for image evaluation, which is defined as

$$PSNR = 10 \times \log_{10} \left(\frac{(2^n - 1)^2}{MSE} \right),$$

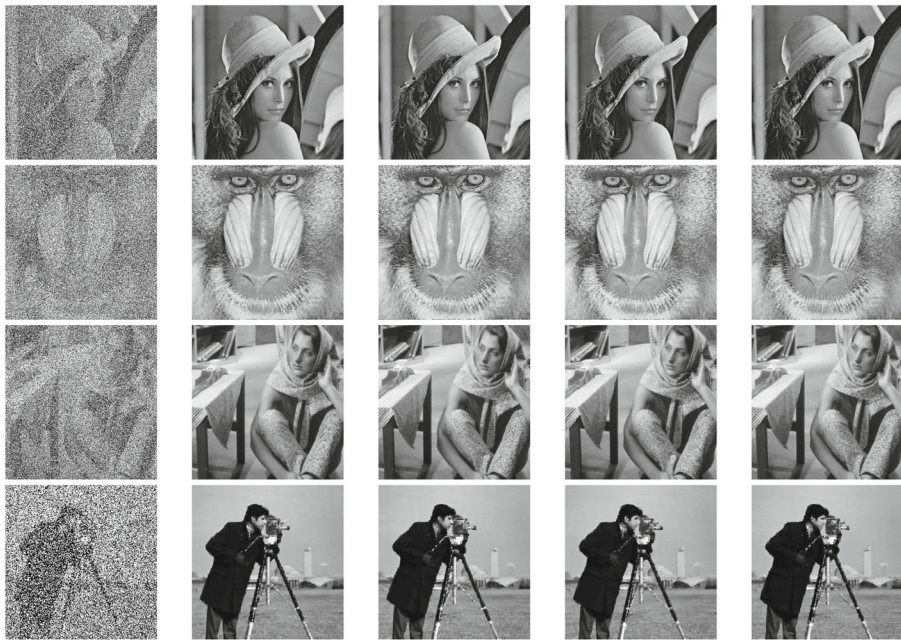


Fig. 3 From left to right, the images disturbed by 60% salt-and-pepper noise, the images restored by A-T-PRP-A, the images restored by A-T-PRP-A-W algorithm, the images restored by MPRP algorithm and the images restored by Hager-Zhang algorithm

where MSE is the mean square error between the original image (voice) and the processed image (speech) and n is the number of bits per sample value. The algorithm will stop if $\frac{\|f_{k+1} - f_k\|}{\|f_k\|} < 10^{-4}$ is satisfied. The parameters of the algorithms are: $\delta = 0.2, \sigma = 0.9, \delta_0 = 0.2, \sigma_0 = 0.9, \mu = 0.01$.

In the experiment, we choose Lena(512×512), Baboon(512×512), Barbara (512×512) and Cameraman (512×512) as the test images. We use the A-T-PRP-A algorithm and A-T-PRP-A-W algorithm to restore the images after adding noise, and compare the results with those of the MPRP algorithm and Hager-Zhang algorithm. The detailed performance results are shown in Figs. 1, 2 and 3. Figures 1, 2 and 3 describe the results of repair by the A-T-PRP-A algorithm, A-T-PRP-A-W algorithm, MPRP algorithm and Hager-Zhang algorithm under noises addition of 30%, 45% and 60% respectively. As shown in Figs. 1, 2 and 3, the A-T-PRP-A algorithm and A-T-PRP-A-W algorithm can successfully restore the image with noise. In addition, Table 1 lists the PSNR values in the experiment. According to Table 1, although the differences in PSNR values of the restored images that are obtained by the four algorithms are small, it can still be seen that the PSNR values that are obtained by the A-T-PRP-A algorithm and A-T-PRP-A-W algorithm are superior to those obtained by the MPRP algorithm and Hager-Zhang algorithm. This shows that the A-T-PRP-A algorithm and A-T-PRP-A-W algorithm are competitive with the MPRP algorithm and Hager-Zhang algorithm in performance.

Table 1 Peak signal-to-noise ratio (PSNR) of A-T-PRP-A algorithm, A-T-PRP-A-W algorithm, MPRP algorithm and Hager-Zhang algorithm

	<i>Lena</i>	<i>Baboon</i>	<i>Barbara</i>	<i>Cameraman</i>	<i>Total</i>
30% noise					
A-T-PRP-A algorithm	38.6029	27.4365	29.0481	30.6062	125.6937
A-T-PRP-A-W algorithm	38.5858	27.4209	29.0968	30.7759	125.8794
MPRP algorithm	38.4313	27.3959	28.9551	30.3775	125.1598
Hager-Zhang algorithm	38.4798	27.4198	28.9841	30.4838	125.3675
45% noise					
A-T-PRP-A algorithm	35.8123	25.2508	27.0440	28.0649	116.1720
A-T-PRP-A-W algorithm	35.8195	25.2406	26.9782	28.0817	116.1200
MPRP algorithm	35.6602	25.1841	26.9259	28.0441	115.8143
Hager-Zhang algorithm	35.5914	25.1990	26.8680	27.6658	115.3242
60% noise					
A-T-PRP-A algorithm	33.3761	23.4403	25.4296	26.2770	108.5230
A-T-PRP-A-W algorithm	33.4423	23.4871	25.4456	25.9833	108.3583
MPRP algorithm	33.1939	23.3829	25.4194	25.8205	107.8167
Hager-Zhang algorithm	33.2362	23.4011	25.4153	25.8349	107.8875

6 Conclusion

For non-convex and gradient non-Lipschitz continuous functions, two adaptive three-term PRP algorithms are presented in this paper. Algorithm 1 is proposed on the basis of MPRP algorithm and its sufficient descent property is preserved independent any line search. The global convergence of Algorithm 1 is obtained using WWP line search technique without the gradient Lipschitz continuity condition, and the proof is combined with the trust region properties by using contradiction. Algorithm 2 is further improved from Algorithm 1 to obtain global convergence independent of line search technique. In addition, numerical experiment shows that Algorithm 1 and Algorithm 2 have certain competitiveness compared with other algorithms.

Algorithm 1 and Algorithm 2 are proposed based on MPRP algorithm. In addition, it can be seen from the proof of Lemma 1 and (3.3) in this paper that if sufficient descent and trust region properties are satisfied, global convergence may be established. However, for classical CG algorithms, such as PRP algorithm, it is a pity that even sufficient descent properties cannot be satisfied.

In the future, there are several aspects can be further studied. The first one is whether there exist other proof methods for the global convergence of CG algorithms and nonconvex without gradient Lipschitz continuity condition. The second one is the convergence of some improved CG method in the absence of gradient Lipschitz continuity condition.

Funding 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), the Special Funds for Local Science and Technology Development Guided by the Central Government (No. ZY20198003), and Special Foundation for Guangxi Ba Gui Scholars.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Al-Baali, M.: Descent property and global convergence of the Fletcher-Reeves method with inexact line search[J]. *IMA J. Numer. Anal.* **5**(1), 121–124 (1985)
2. Dai, Y.: Analysis of conjugate gradient methods, Ph.D. Thesis. Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences (1997)
3. Dai, Y.: Convergence properties of the BFGS algorithm[J]. *SIAM J. Optim.* **13**(3), 693–701 (2002)
4. Dai, Y., Yuan, Y.: A nonlinear conjugate gradient method with a strong global convergence property[J]. *SIAM J. Optim.* **10**(1), 177–182 (1999)
5. Fletcher, R.: *Practical Methods of Optimization*, 2nd. Wiley, New York (1987)
6. Fletcher, R., Reeves, C.: Function minimization by conjugate gradients[J]. *Comput. J.* **7**(2), 149–154 (1964)
7. Gilbert, J., Nocedal, J.: Global convergence properties of conjugate gradient methods for optimization[J]. *SIAM J. Optim.* **2**(1), 21–42 (1992)
8. Goldstein, A.A.: On steepest descent[J]. *J. Soc. Indust. Appl. Math. Series A: Control* **3**(1), 147–151 (1965)
9. Hestenes, M., Stiefel, E.: Method of conjugate gradient for solving linear equations. *J. Res. Natl. Bur. Stand.* **49**, 409–436 (1952)
10. Hager, W., Zhang, H.: A new conjugate gradient method with guaranteed descent and an efficient line search[J]. *SIAM J. Optim.* **16**(1), 170–192 (2005)
11. Hager, W., Zhang H.: Algorithm 851: CG-DESCENT, a conjugate gradient method with guaranteed descent[J]. *ACM Trans. Math. Softw. (TOMS)* **32**(1), 113–137 (2006)
12. Liu, J., Lu, Z., Xu, J., Wu, S., Tu, Z.: An efficient projection-based algorithm without Lipschitz continuity for large-scale nonlinear pseudo-monotone equations[J]. *Journal of Computational and Applied Mathematics*, 113822. <https://doi.org/10.1016/j.cam.2021.113822> (2021)
13. Levenberg, K.: A method for the solution of certain non-linear problems in least squares[J]. *Quart. Appl. Math.* **2**(2), 164–168 (1944)
14. Li, X., Wang, S., Jin, Z., et al: A conjugate gradient algorithm under Yuan-Wei-Lu line search technique for large-scale minimization optimization models[J]. *Math. Probl. Eng.*, 2018 (2018)
15. Liu, Y., Storey, C.: Efficient generalized conjugate gradient algorithms, part 1: Theory[J]. *J. Optim. Theory Appl.* **69**(1), 129–137 (1991)
16. Polyak, B.: The conjugate gradient method in extremal problems[J]. *USSR Comput. Math. Math. Phys.* **9**(4), 94–112 (1969)
17. Polak, E., Ribiere, G.: Note sur la convergence de méthodes de directions conjuguées[J]. *ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique* **3**(R1), 35–43 (1969)
18. Powell, M.: Convergence properties of algorithms for nonlinear optimization[J]. *SIAM Rev.* **28**(4), 487–500 (1986)
19. Powell, M.: *Nonconvex Minimization Calculations and the Conjugate Gradient Method*[M]. Springer, Berlin (1984)
20. Powell, M.J.D.: Convergence Properties of a Class of Minimization Algorithms[M]/*Nonlinear Programming 2*. Academic Press, pp. 1–27 (1975)
21. Sheng, Z., Ouyang, A., Liu, L., et al.: A novel parameter estimation method for Muskingum model using new Newton-type trust region algorithm[J]. *Math. Probl. Eng.*, 2014 (2014)

22. Sheng, Z., Yuan, G.: An effective adaptive trust region algorithm for nonsmooth minimization[J]. *Comput. Optim. Appl.* **71**(1), 251–271 (2018)
23. Sheng, Z., Yuan, G., Cui, Z., et al.: An adaptive trust region algorithm for large-residual nonsmooth least squares problems[J]. *J. Industr. Manag. Optim.* **14**(2), 707 (2018)
24. Sheng, Z., Gonglin, Y., Zengru, C.U.I.: A new adaptive trust region algorithm for optimization problems[J]. *Acta Math. Sci.* **38**(2), 479–496 (2018)
25. Wolfe, P.: Convergence conditions for ascent methods[J]. *SIAM Rev.* **11**(2), 226–235 (1969)
26. Wolfe, P.: Convergence conditions for ascent methods. II Some corrections[J]. *SIAM Rev.* **13**(2), 185–188 (1971)
27. Wei, Z., Yao, S., Liu, L.: The convergence properties of some new conjugate gradient methods[J]. *Appl. Math. Comput.* **183**(2), 1341–1350 (2006)
28. Yuan, G., Li, T., Hu, W.: A conjugate gradient algorithm for large-scale nonlinear equations and image restoration problems[J]. *Appl. Numer. Math.* **147**, 129–141 (2020)
29. Yuan, G., Lu, X.: A modified PRP conjugate gradient method[J]. *Ann. Oper. Res.* **166**(1), 73–90 (2009)
30. Yuan, G., Lu, J., Wang, Z.: The PRP conjugate gradient algorithm with a modified WWP line search and its application in the image restoration problems[J]. *Appl. Numer. Math.* **152**, 1–11 (2020)
31. Yuan, G., Lu, J., Wang, Z.: The modified PRP conjugate gradient algorithm under a non-descent line search and its application in the Muskingum model and image restoration problems[J]. *Soft. Comput.* **25**(8), 5867–5879 (2021)
32. Yuan, G., Lu, X., Wei, Z.: A conjugate gradient method with descent direction for unconstrained optimization[J]. *J. Comput. Appl. Math.* **233**(2), 519–530 (2009)
33. Yuan, G., Lu, S., Wei, Z.: A new trust-region method with line search for solving symmetric nonlinear equations[J]. *Int. J. Comput. Math.* **88**(10), 2109–2123 (2011)
34. Yuan, G., Lu, X., Wei, Z.: BFGS trust-region method for symmetric nonlinear equations[J]. *J. Comput. Appl. Math.* **230**(1), 44–58 (2009)
35. Yuan, G.: Modified nonlinear conjugate gradient methods with sufficient descent property for large-scale optimization problems. *Optim. Lett.* **3**, 11–21 (2009)
36. Yuan, G., Meng, Z., Li, Y.: A modified Hestenes and Stiefel conjugate gradient algorithm for large-scale nonsmooth minimizations and nonlinear equations[J]. *J. Optim. Theory Appl.* **168**(1), 129–152 (2016)
37. Yuan, G., Zhang, M., Zhou, Y.: Adaptive scaling damped BFGS method without gradient Lipschitz continuity. [J] **124**, 107634 (2022). <https://doi.org/10.1016/j.aml.2021.107634>
38. Yuan, G., Wei, Z., Li, G.: A modified Polak-Ribière-Polyak conjugate gradient algorithm for nonsmooth convex programs[J]. *J. Comput. Appl. Math.* **255**, 86–96 (2014)
39. Yuan, G., Wei, Z., Yang, Y.: The global convergence of the Polak-Ribière-Polyak conjugate gradient algorithm under inexact line search for nonconvex functions[J]. *J. Comput. Appl. Math.* **362**, 262–275 (2019)
40. Yuan, G., Zhang, M.: A three-terms Polak-Ribière-Polyak conjugate gradient algorithm for large-scale nonlinear equations[J]. *J. Comput. Appl. Math.* **286**, 186–195 (2015)
41. Yuan, Y.: Analysis on the conjugate gradient method[J]. *Dyn. Syst.* **2**(1), 19–29 (1993)
42. Zoutendijk, G.: Nonlinear programming, computational methods[J]. *Integer & Nonlinear Programming*, 37–86 (1970)
43. Zhang, L., Zhou, W., Li, D.: A descent modified Polak-Ribière-Polyak conjugate gradient method and its global convergence[J]. *IMA J. Numer. Anal.* **26**(4), 629–640 (2006)

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.