RESEARCH ARTICLE

A Conjugate Gradient Algorithm with Function Value Information and N-Step Quadratic Convergence for Unconstrained Optimization

Xiangrong Li¹, Xupei Zhao², Xiabin Duan¹, Xiaoliang Wang^{1*}

1 Guangxi Colleges and Universities Key Laboratory of Mathematics and Its Applications, College of Mathematics and Information Science, Guangxi University, Nanning, Guangxi, P.R. China, **2** Mathematics and Physics Institute of Henan University of Urban Construction, Pingdingshan, Henan, P. R. China

* xliangwang@126.com

OPEN ACCESS

Citation: Li X, Zhao X, Duan X, Wang X (2015) A Conjugate Gradient Algorithm with Function Value Information and N-Step Quadratic Convergence for Unconstrained Optimization. PLoS ONE 10(9): e0137166. doi:10.1371/journal.pone.0137166

Editor: Hans A Kestler, Leibniz Institute for Age Research, GERMANY

Received: January 18, 2015

Accepted: August 13, 2015

Published: September 18, 2015

Copyright: © 2015 Li et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper.

Funding: Mrs. Li and Dr. Zhao are the members of the Guangxi NSF (Grant No. 2012GXNSFAA053002) and the China NSF (Grant No. 11261006 and 11161003). Moreover Dr. Zhao is one member of the Fund of Henan University of Urban Construction (Grant No. 2014JBS003).

Competing Interests: The authors have declared that no competing interests exist.

Abstract

It is generally acknowledged that the conjugate gradient (CG) method achieves global convergence—with at most a linear convergence rate—because CG formulas are generated by linear approximations of the objective functions. The quadratically convergent results are very limited. We introduce a new PRP method in which the restart strategy is also used. Moreover, the method we developed includes not only n-step quadratic convergence but also both the function value information and gradient value information. In this paper, we will show that the new PRP method (with either the Armijo line search or the Wolfe line search) is both linearly and quadratically convergent. The numerical experiments demonstrate that the new PRP algorithm is competitive with the normal CG method.

Introduction

Consider

$$\min_{x \in \mathbb{R}^n} f(x), \quad (1)$$

where $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ is a continuously differentiable function. This optimization problem can be used to model other problems (see [1–5]). The iterative processes of the conjugate gradient (CG) methods are determined by

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

where x_k is the k -th iteration; α_k is the step length, which is determined by a line search method;

and d_k is the search direction defined by

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

where $g_k = g(x_k)$ is the gradient $\nabla f(x_k)$ of $f(x)$ at the point x_k and β_k is a scalar. In addition, the parameter β_k varies between different CG formulas (see [6–17] etc.). This paper focuses on the PRP method designed by [13, 14]

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

where g_{k-1} is the gradient $\nabla f(x_{k-1})$ of $f(x)$ at the point x_{k-1} , and $\|\cdot\|$ denotes the Euclidean norm of vectors. Recently, there have been several successful research projects involving the PRP algorithm (see [7, 18–29] for details). The linear convergence of the standard PRP method is well known. If a restart strategy is used, the algorithm can be n -step quadratic convergent [30–32]; however, the quadratically convergent results are very limited. Zhang et al. [33] presented a three-term CG algorithm by modifying the PRP (MPRP) method and proved its global convergence. Moreover, Li and Tian [34] proved that the MPRP method with a restart strategy exhibited quadratic convergence under certain inexact line searches and suitable assumptions. In this paper, we focus primarily on the MPRP method. On the basis of the MPRP method [33], a new three-term CG formula is proposed in which the formula includes not only the gradient information but also the function information. By applying the conclusion of [34], we prove that the new CG method obtains global convergence for general functions and n -step quadratic convergence for uniformly convex functions. The numerical results show that the new method performs quite well. The main attributes of this method are stated below.

- The new PRP CG method possesses not only the gradient information but also information about function values.
- The sufficient descent property is satisfied without any conditions.
- The global convergence, linear convergent rate, and n -step quadratic convergence are established.
- The numerical results demonstrate that this method is competitive with the normal CG method for the given problems.

This paper is arranged as follows. In Section 2, we review our motivation, review the MPRP method of [33], and introduce our new PRP method. The r -step linear convergence of the new PRP method with the standard Armijo line search is proven in Section 3. In Section 4, we introduce a strategy for choosing the initial step length using the Armijo or Wolfe line search and provide the steps of the new PRP method. In Section 5, the n -step quadratic convergence of the given CG algorithm is established. In Section 6, various numerical experiments are performed to test the n -step quadratic convergence of the new PRP algorithm, and its performance is compared with that of the normal CG method. The conclusion is presented in the last section.

Motivation and Algorithm

In this section, we will provide our motivation based on the MPRP method [33] and the BFGS formulas.

The MPRP method of [33]

Zhang et al. [33] proposed a modified PRP method (a three-term CG method) called the MPRP method. The search direction of the MPRP method is defined by

$$d_k = \begin{cases} -g_k, & \text{if } k = 0, \\ -g_k + \beta_k^{PRP} d_{k-1} - \theta_k y_{k-1}, & \text{if } k \geq 1. \end{cases} \quad (5)$$

where $y_{k-1} = g_k - g_{k-1}$,

$$\beta_k^{PRP} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}, \quad \theta_k = \frac{g_k^T d_{k-1}}{g_k^T y_{k-1}} \cdot \beta_k^{PRP} = \frac{g_k^T d_{k-1}}{\|g_{k-1}\|^2}. \quad (6)$$

This method can guarantee that d_k provides a descent direction of f at x_k , namely, d_k satisfies $d_k^T g_k = -\|g_k\|^2$. Because this method is regarded as the PRP method when the exact line search technique is used, we call it the MPRP method. This method possesses global convergence with the Armijo line search but fails using the Wolfe line search because the search direction does not belong to a trust region. The numerical results thus show that this method is effective.

Motivation based on the BFGS formula and the MPRP method

The BFGS method is one of the most effective methods for unconstrained optimization problems. Many satisfactory results concerning this method can be found in the literature (see [35–43] etc.), wherein Wei et al. [43] presented a new BFGS update generated by Taylor's formula

$$B_k = B_{k-1} - \frac{B_{k-1} s_{k-1} s_{k-1}^T B_{k-1}}{s_{k-1}^T B_{k-1} s_{k-1}} + \frac{y_{k-1}^m y_{k-1}^{mT}}{s_{k-1}^T y_{k-1}^m},$$

with $y_{k-1}^m = y_{k-1} + \frac{\rho_{k-1}}{\|s_{k-1}\|^2} s_{k-1}$ and $\rho_{k-1} = 2[f_{k-1} - f_k] + (g_k + g_{k-1})^T s_{k-1}$; B_k is an approximation of the Hessian matrix of the objective function f , with $B_0 = I$. Clearly, the formula contains not only gradient value information but also function value information. Then, it might be argued that the resulting methods will substantially outperform the original method in theory.

Practical computations demonstrate that this method is better than the normal BFGS method [43, 44]. Based on this method, to obtain the global convergence and the superlinear convergence of generally convex functions, Yuan and Wei [45] proposed a new quasi-Newton equation with

$$B_k s_{k-1} = y_{k-1}^*,$$

where $y_{k-1}^* = y_{k-1} + \frac{\max\{\rho_{k-1}, 0\}}{\|s_{k-1}\|^2} s_{k-1}$ and

$$B_k = B_{k-1} - \frac{B_{k-1} s_{k-1} s_{k-1}^T B_{k-1}}{s_{k-1}^T B_{k-1} s_{k-1}} + \frac{y_{k-1}^* (y_{k-1}^*)^T}{s_{k-1}^T y_{k-1}^*}, \quad (7)$$

can ensure that B_k inherits the positive definiteness of B_{k-1} for generally convex functions; the numerical results show that this method is better than the normal CG method.

Motivated by the above discussions and the CG Eq (5), the modified PRP formula is used to replace y_{k-1} by y_{k-1}^* in Eq (5). Specifically, the new CG formula is defined by

$$d_k^* = \begin{cases} -g_k, & \text{if } k = 0, \\ -g_k + \beta_k^{PRP^*} d_{k-1}^* - \theta_k^* y_{k-1}^*, & \text{if } k \geq 1, \end{cases} \quad (8)$$

where

$$\begin{aligned} y_{k-1}^* &= y_{k-1} + \frac{\max\{\rho_{k-1}, 0\}}{\|s_{k-1}\|} s_{k-1}, \quad s_{k-1} = \alpha_{k-1} d_{k-1}^* \\ \beta_k^{PRP^*} &= \frac{g_k^T y_{k-1}^*}{\|g_{k-1}\|^2}, \quad \theta_k^* = \frac{g_k^T d_{k-1}^*}{g_k^T y_{k-1}^*} \cdot \beta_k^{PRP^*} = \frac{g_k^T d_{k-1}^*}{\|g_{k-1}\|^2} \end{aligned} \quad (9)$$

We call this the $MPRP^*$ formula. Using (8), we can obtain

$$d_k^{*T} g_k = -\|g_k\|^2 \quad (10)$$

and

$$\|g_k\| \leq \|d_k^*\|. \quad (11)$$

Linear convergence of the $MPRP^*$ method

The Armijo line search finds a step length $\alpha_k = \max\{\rho^i | i = 0, 1, 2, \dots\}$ that satisfies

$$f(x_k + \alpha_k d_k^*) \leq f(x_k) + \sigma_1 \alpha_k g_k^T d_k^*, \quad (12)$$

where $\rho \in (0, 1)$ and $\sigma_1 \in (0, \frac{1}{2})$. The Wolfe line search conditions are

$$\begin{cases} f(x_k + \alpha_k d_k^*) \leq f(x_k) + \sigma_1 \alpha_k g_k^T d_k^* \\ g(x_k + \alpha_k d_k^*)^T d_k^* \geq \sigma_2 g_k^T d_k^*, \end{cases} \quad (13)$$

where $0 < \sigma_1 < \frac{1}{2}$ and $\sigma_1 < \sigma_2 < 1$.

To establish the r-linear convergence of the $MPRP^*$ method, we make the following assumptions.

Assumption (i) f is a twice continuously differentiable, uniformly convex function, which means that there are positive constants $M \geq m > 0$ such that

$$m \|d\|^2 \leq d^T \nabla^2 f(x) d \leq M \|d\|^2, \quad \forall x, d \in \mathbb{R}^n,$$

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

Under Assumption (i), it is not difficult to obtain

$$\frac{1}{2} m \|x - x^*\|^2 \leq f(x) - f(x^*) \leq \frac{1}{2} M \|x - x^*\|^2, \quad \forall x \in \mathbb{R}^n \quad (14)$$

and

$$m \|x_k - x^*\| \leq \|g_k\| \leq M \|x_k - x^*\|, \quad (15)$$

where x^* denotes the unique solution of the problem (1). Based on Assumption (i), we can easily obtain the following lemma and thus omit its proof.

Lemma 0.1 Let Assumption (i) hold, and let the sequence $\{x_k\}$ be generated by the $MPRP^*$ method with the Armijo or Wolfe line search. We have

$$-\sum_{k=0}^{\infty} g_k^T s_k^* < \infty, \quad \sum_{k=0}^{\infty} \|s_k^*\| < \infty, \quad c_1 \alpha_k \|d_k^*\|^2 \leq -g_k^T d_k^*,$$

where $s_k^* = x_{k+1} - x_k$ and $c_1 = \frac{1}{2}(1 - \sigma_1)^{-1}m$. Moreover, when the Wolfe line search is used, we

obtain

$$-g_k^T d_k^* \leq (1 - \sigma_2)^{-1} M \alpha_k \|d_k^*\|^2.$$

Lemma 0.2 Let Assumption (i) hold, and let the sequence $\{x_k\}$ be generated by the MPRP* method with the Armijo line search or the Wolfe line search. Then, there is a constant $c > 0$ satisfying

$$\alpha_k > c, \quad \forall k \geq 0. \quad (16)$$

Proof. Let

$$G_{k-1} = \int_0^1 f(x_{k-1} + \tau s_{k-1}) d\tau,$$

where $s_{k-1}^* = x_k - x_{k-1} = \alpha_{k-1} d_{k-1}^*$, $y_{k-1}^* = y_{k-1} + \frac{\max\{\rho_{k-1}, 0\}}{\|s_{k-1}^*\|^2} \cdot s_{k-1}^*$. Using the definition of $\beta_k^{PRP^*}$, we obtain

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

Case 1: If $\rho_{k-1} \leq 0$, we have $y_{k-1}^* = g_k - g_{k-1}$.

Using the mean-value theorem, we have

$$y_{k-1}^* = g_k - g_{k-1} = G_{k-1} \cdot s_{k-1}^* = \alpha_{k-1} G_{k-1} d_{k-1}^*,$$

and

$$\begin{aligned} |\beta_k^{PRP^*}| &= \frac{|g_k^T (g_k - g_{k-1})|}{\|g_{k-1}\|^2} = \frac{|\alpha_{k-1} g_k^T G_{k-1} d_{k-1}^*|}{-g_{k-1}^T d_{k-1}^*} \\ &\leq \frac{|\alpha_{k-1}| \cdot \|g_k\| \cdot \|G_{k-1}\| \cdot \|d_{k-1}^*\|}{c_1 \alpha_{k-1} \|d_{k-1}^*\|^2} \\ &\leq M c_1^{-1} \frac{\|g_k\|}{\|d_{k-1}^*\|}. \end{aligned}$$

Moreover, we obtain

$$|\theta_k^*| = \frac{|g_k^T d_{k-1}^*|}{\|g_{k-1}\|^2} = \frac{|g_k^T d_{k-1}^*|}{-g_{k-1}^T d_{k-1}^*} \leq c_1^{-1} \alpha_{k-1}^{-1} \frac{\|g_k\|}{\|d_{k-1}^*\|}.$$

Thus, it follows from (8) and the Lipschitz continuity of g that

$$\begin{aligned} \|d_k^*\| &= \|-g_k + \beta_k^{PRP^*} d_{k-1}^* - \theta_k^* y_{k-1}^*\| \\ &\leq \|g_k\| + |\beta_k^{PRP^*}| \cdot \|d_{k-1}^*\| + |\theta_k^*| \cdot \|y_{k-1}^*\| \\ &\leq \|g_k\| + M c_1^{-1} \|g_k\| + c_1^{-1} \alpha_{k-1}^{-1} \cdot \frac{\|g_k\|}{\|d_{k-1}^*\|} \cdot \|y_{k-1}^*\| \\ &\leq (1 + 2M c_1^{-1}) \|g_k\|. \end{aligned}$$

Case 2: If $\rho_{k-1} > 0$, we have

$$\begin{aligned} y_{k-1}^* &= y_{k-1} + \frac{\rho_{k-1}^T s_{k-1}^*}{\|s_{k-1}^*\|^2} \\ &= (g_k - g_{k-1}) + \frac{[2(f(x_{k-1}) - f(x_k)) + (g(x_k) + g^T(x_{k-1}) \cdot s_{k-1}^*)] \cdot s_{k-1}^*}{\|s_{k-1}^*\|^2} \\ &= (g_k - g_{k-1}) + \frac{(f(x_{k-1}) - f(x_k)) \cdot s_{k-1}^*}{\|s_{k-1}^*\|^2} + (g_k + g_{k-1}) \\ &= 2g_k + \frac{(f(x_{k-1}) - f(x_k)) \cdot s_{k-1}^*}{\|s_{k-1}^*\|^2}. \end{aligned}$$

By

$$f(x_{k-1}) - f(x_k) \leq \frac{1}{2}M \|x_{k-1} - x_k\|^2 = \frac{1}{2}M \|x_k - x_{k-1}\|^2 = \frac{1}{2}M \alpha_{k-1}^2 \|d_{k-1}^*\|^2$$

and

$$s_{k-1}^* = \alpha_{k-1} \cdot d_{k-1}^*, \quad \|g_k\| \leq M \|x_k - x_{k-1}\| = M \alpha_{k-1} \|d_{k-1}^*\|,$$

we obtain

$$\begin{aligned} \|y_{k-1}^*\| &\leq 2 \|g_k\| + \frac{\|f(x_{k-1}) - f(x_k)\| \cdot \|s_{k-1}^*\|}{\|s_{k-1}^*\|^2} \\ &\leq 2 \left(M \alpha_{k-1} \|d_{k-1}^*\| + \frac{\frac{1}{2}M \alpha_{k-1}^2 \|d_{k-1}^*\|^2 \alpha_{k-1} \|d_{k-1}^*\|}{\alpha_{k-1}^2 \|d_{k-1}^*\|^2} \right) \\ &= 3M \alpha_{k-1} \|d_{k-1}^*\|. \end{aligned}$$

Then, we have

$$|\beta_k^{PRP^*}| = \frac{|g_k^T y_{k-1}^*|}{\|g_{k-1}\|^2} \leq \frac{\|g_k^T\| \cdot \|y_{k-1}^*\|}{\|g_{k-1}\|^2} \leq \frac{3M \alpha_{k-1} \cdot \|g_k^T\| \cdot \|d_{k-1}^*\|}{c_1 \alpha_{k-1} \cdot \|d_{k-1}^*\|^2} \leq 3M c_1^{-1} \frac{\|g_k\|}{\|d_{k-1}^*\|}$$

and

$$|\theta_k^*| = \frac{|g_k^T d_{k-1}^*|}{\|g_{k-1}\|^2} = \frac{|g_k^T d_{k-1}^*|}{-g_{k-1}^T d_{k-1}^*} \leq c_1^{-1} \alpha_{k-1}^{-1} \frac{\|g_k\|}{\|d_{k-1}^*\|}.$$

Therefore, it follows from (8) and the Lipschitz continuity of g that

$$\begin{aligned} \|d_k^*\| &= \|-g_k + \beta_k^{PRP^*} d_{k-1}^* - \theta_k^* y_{k-1}^*\| \\ &\leq \|g_k\| + |\beta_k^{PRP^*}| \cdot \|d_{k-1}^*\| + |\theta_k^*| \cdot \|y_{k-1}^*\| \\ &\leq \|g_k\| + \frac{3M \alpha_{k-1} \|g_k^T\| \cdot \|d_{k-1}^*\|}{c_1 \alpha_{k-1} \cdot \|d_{k-1}^*\|^2} \cdot \|d_{k-1}^*\| + c_1^{-1} \alpha_{k-1}^{-1} \frac{\|g_k\|}{\|d_{k-1}^*\|} \cdot \|y_{k-1}^*\| \quad (17) \\ &\leq \|g_k\| + 3c_1^{-1} M \|g_k\| + c_1^{-1} \alpha_{k-1}^{-1} \frac{\|g_k\|}{\|d_{k-1}^*\|} \cdot 3M \alpha_{k-1} \|d_{k-1}^*\| \\ &\leq (1 + 6c_1^{-1} M) \|g_k\|. \end{aligned}$$

If the Armijo line search is used, and $\alpha_k \neq 1$, then $\alpha'_k = \alpha_k \rho^{-1}$ such that $f(x_k + \alpha'_k d_k^*) - f(x_k) > \sigma_1 \alpha'_k g_k^T d_k^*$. By the mean-value theorem and the above relation, we can deduce that there exists a

scalar $\mu_k \in (0,1)$ satisfying

$$\begin{aligned} \sigma_1 \alpha'_k g_k^T d_k^* &< f(x_k + \alpha'_k d_k^*) - f(x_k) = \alpha'_k g(x_k + \mu_k \alpha'_k d_k^*)^T d_k^* \\ &= \alpha'_k (g(x_k + \mu_k \alpha'_k d_k^*) - g(x_k))^T d_k^* + \alpha'_k g_k^T d_k^* \\ &= \mu_k (\alpha'_k)^2 d_k^{*T} \int_0^1 \nabla^2 f(x_k + \tau \mu_k \alpha'_k d_k^*) d\tau + \alpha'_k g_k^T d_k^* \\ &\leq (\alpha'_k)^2 M \cdot \|d_k^*\|^2 + \alpha'_k g_k^T d_k^*. \end{aligned}$$

Thus, the relation

$$\alpha_k = \rho \alpha'_k \geq -\frac{(1 - \sigma_1)\rho}{M} \cdot \frac{g_k^T d_k^*}{\|d_k^*\|^2} = \frac{(1 - \sigma_1)\rho}{M} \cdot \frac{\|g_k\|^2}{\|d_k^*\|^2}$$

holds. By the relation $1 + 6Mc_1^{-1} > 1 + 2Mc_1^{-1} > 0$, we can find that

$$\frac{1}{1 + 2Mc_1^{-1}} > \frac{1}{1 + 6Mc_1^{-1}}.$$

Then, we obtain

$$\alpha_k = \rho \alpha'_k \geq -\frac{(1 - \sigma_1)\rho}{M} \cdot \frac{g_k^T d_k^*}{\|d_k^*\|^2} = \frac{(1 - \sigma_1)\rho}{M} \cdot \frac{\|g_k\|^2}{\|d_k^*\|^2} \geq M^{-1}(1 - \sigma_1)\rho(1 + 2Mc_1^{-1})^{-2} \equiv \bar{c}.$$

Setting $c = \min\{1, \bar{c}\}$, we have [Eq \(16\)](#). If the Wolfe line search is used, using [Eq \(13\)](#), we have

$$M\alpha_k \|d_k^*\|^2 \geq (g(x_k + \alpha_k d_k^*) - g_k)^T d_k^* \geq -(1 - \sigma_2)g_k^T d_k^*.$$

By analyzing the Armijo line search technique in a similar manner, we can find a lower positive bound of the step size α_k . The proof is complete.

Similar to [\[34\]](#), we can establish the following theorem. Here, we state the theorem below but omit its proof.

Theorem 0.1 *Let Assumption (i) hold, and let the sequence $\{x_k\}$ be generated by the MPRP* method with the Armijo line search technique or Wolfe line search technique. Then, there are constants $a > 0$ and $r \in (0,1)$ that satisfy*

$$\|x_k - x^*\| \leq ar^k. \quad (18)$$

The restart MPRP* method and its convergence

As with [\[34\]](#), we define an initial step length γ_k^* as follows:

$$\gamma_k^* \equiv \frac{\epsilon_k \|g_k\|^2}{d_k^{*T}(g(x_k + \epsilon_k d_k^*) - g(x_k))}, \quad (19)$$

using the integer sequence $\{\epsilon_k\} \rightarrow 0$ with $k \rightarrow \infty$. Moreover, we can obtain

$$|\bar{\alpha}_k - \gamma_k^*| \rightarrow 0.$$

Theorem 0.2 *Let sequence $\{x_k\}$ be generated by the MPRP* method, and let Assumption (i) hold. Then, for sufficiently large k , γ_k^* satisfies the Armijo line search and the Wolfe-Powell line search conditions.*

Theorem 0.2 shows that, for sufficiently large k , γ_k^* can be defined by [Eq \(19\)](#) such that the Armijo line search and Wolfe-Powell line search conditions are satisfied. In the following, $|\gamma_k|$ is used as the initial step length of the restart MPRP* method.

Algorithm 4.1 (RMPPR^{*})

Step 0: Given $x_0 \in \mathbb{R}^n$, $\sigma_1 \in (0, \frac{1}{2})$, $\rho \in [0, 1)$, $\epsilon \in [0, 1)$, and $\gamma > 0$ is an integer, let $k := 0$.

Step 1: If $\|g_k\| \leq \epsilon$, stop.

Step 2: If the inequality $f(x_k + \gamma_k d_k^*) \leq f(x_k) + \sigma_1 \gamma_k |g_k^T d_k^*|$ holds, we set $\alpha_k = \gamma_k$; otherwise, we determine $\alpha_k = \max\{\gamma_k \rho^j | j = 0, 1, 2, \dots\}$ satisfying

$$f(x_k + \alpha_k d_k^*) \leq f(x_k) + \sigma_1 \alpha_k g_k^T d_k^*. \quad (20)$$

Step 3: Let $x_{k+1} = x_k + \alpha_k d_k^*$, and $k := k+1$.

Step 4: If $\|g_k\| \leq \epsilon$, stop.

Step 5: If $k = \gamma$, we let $x_0 := x_k$. Go to step 1.

Step 6: Compute d_k^* using (8). Go to step 2.

We now establish the global convergence of Algorithm 4.1.

Theorem 0.3 Let the conditions of Theorem 0.2 hold. Then, the following relation

$$\lim_{k \rightarrow \infty} \|g_k\| = 0 \quad (21)$$

holds.

Proof. We will prove this theorem by contradiction. Suppose that Eq (21) does not hold; then, for all k , there exists a constant $\epsilon_1 > 0$ satisfying

$$\|g_k\| \geq \epsilon_1. \quad (22)$$

Using Eqs (10) and (20), if f is bounded from below, we can obtain

$$\sum_{k=0}^{\infty} \alpha_k^2 \|d_k^*\|^2 < \infty. \quad (23)$$

In particular, we have

$$\lim_{k \rightarrow \infty} \alpha_k \|d_k^*\| = 0. \quad (24)$$

If $\lim_{k \rightarrow \infty} \alpha_k > 0$, by Eqs (10) and (24), we obtain $\lim_{k \rightarrow \infty} \|g_k\| = 0$. This contradicts Eqs (22) and (21) holds. Otherwise, if $\lim_{k \rightarrow \infty} \alpha_k = 0$, then there is an infinite index set K_0 satisfying

$$\lim_{k \in K_0, k \rightarrow \infty} \alpha_k = 0. \quad (25)$$

According to Step 2 of Algorithm 4.1, when k is sufficiently large, $\alpha_k \rho^{-1}$ does not satisfy Eq (20), which implies that

$$f(x_k + \alpha_k \rho^{-1} d_k^*) - f(x_k) > -\delta \rho^{-2} \alpha_k^2 \|d_k^*\|^2. \quad (26)$$

By Eq (22), similar to the proof of Lemma 2.1 in [33], we can deduce that there is a constant $\varrho > 0$ such that

$$\|d_k^*\| \leq \varrho, \quad \forall k. \quad (27)$$

Using Eqs (27) and (10), and the mean-value theorem, we have

$$\begin{aligned} f(x_k + \alpha_k \rho^{-1} d_k^*) - f(x_k) &= \rho^{-1} \alpha_k g(x_k + \xi_0 \rho^{-1} \alpha_k d_k^*)^T d_k^* \\ &= \rho^{-1} \alpha_k g_k^T d_k^* + \rho^{-1} \alpha_k (g(x_k + \xi_0 \rho^{-1} \alpha_k d_k^*) - g_k)^T d_k^* \\ &\leq \rho^{-1} \alpha_k g_k^T d_k^* + M \rho^{-2} \alpha_k^2 \|d_k^*\|^2, \end{aligned}$$

where $\xi_0 \in (0, 1)$ and the last inequality follows Eq (15). Combining this result with Eq (26), for

all sufficiently large $k \in K_0$, we obtain

$$\|g_k\|^2 \leq \rho^{-1}(M + \delta)\alpha_k \|d_k^*\|^2.$$

By Eq (27) and $\lim_{k \rightarrow \infty} \alpha_k = 0$, the above inequality then implies that $\lim_{k \in K_0, k \rightarrow \infty} \|g_k\| = 0$. This is also a contradiction. The proof is complete.

Lemma 0.3 Let Assumption (i) hold, and let the sequence $\{x_k\}$ be generated by the RMPRP* method. Then, there are four positive numbers $c_i, i = 1, 2, 3, 4$ that satisfy

$$\|g_{k+1}\| \leq c_1 \|d_k^*\|, \quad |\beta_{k+1}^{PRP^*}| \leq c_2, \quad |\theta_{k+1}^*| \leq c_3, \quad \|d_{k+1}^*\| \leq c_4 \|d_k^*\|. \quad (28)$$

Proof. Considering the first inequality of Eq (28), we have

$$\begin{aligned} \|g_{k+1}\| &= \|g_k + (g_{k+1} - g_k)\| \leq \|g_k\| + |\gamma_k| \cdot \|\hat{A}_k d_k^*\| \\ &\leq \|d_k^*\| + \frac{M}{m} \cdot \|d_k^*\| = \left(1 + \frac{M}{m}\right) \|d_k^*\| \equiv c_1 \|d_k^*\|, \end{aligned} \quad (29)$$

where $\hat{A}_k = \int_0^1 \nabla^2 f(x_k + \tau | \gamma_k | d_k^*) d\tau$. Using $\beta_{k+1}^{PRP^*}$, we will discuss the three other inequalities of Eq (28). For $\beta_{k+1}^{PRP^*}$, we have

$$|\beta_{k+1}^{PRP^*}| = \frac{|g_{k+1}^T \gamma_k|}{\|g_k\|^2} \leq \frac{\|g_{k+1}^T\| \cdot \|\gamma_k^*\|}{\|g_k\|^2}.$$

Case 1: If $\rho_k \leq 0$, then $y_k^* = y_k = g_{k+1} - g_k$. Therefore, we have

$$\begin{aligned} |\beta_{k+1}^{PRP^*}| &\leq \frac{\|g_{k+1}\|}{\|g_k\|^2} \|g_{k+1} - g_k\| \leq \frac{c_1 |\gamma_k| \|d_k^*\|^2 M}{\|g_k\|^2} \\ &= \frac{c_1 \|d_k^*\|^2 M}{d_k^{*T} A_k d_k^*} \leq \frac{c_1 \|d_k^*\|^2 M}{d_k^{*T} A_k d_k^*} \leq c_1 m^{-1} M \equiv \bar{c}_2 \end{aligned}$$

where $A_k = \int_0^1 \nabla^2 f(x_k + \tau \epsilon_k d_k^*) d\tau$ and the second inequality follows from the mean-value theorem and Eq (29).

Case 2: If $\rho_k > 0$, then we have

$$\begin{cases} y_k^* = y_k + \frac{\rho_k^T s_k^*}{\|s_k^*\|^2} \\ \rho_k = 2[f(x_k) - f(x_{k+1})] + (g(x_{k+1}) + g(x_k))^T s_k^*. \end{cases}$$

Thus, we have

$$\begin{aligned}
 |\beta_{k+1}^{PRP^*}| &\leq \frac{\|g_{k+1}^T\| \cdot \|y_k^*\|}{\|g_k\|^2} \\
 &\leq \frac{\|g_{k+1}\|}{\|g_k\|^2} \left[\|g_{k+1} - g_k\| + \frac{2\|f(x_{k+1}) - f(x_k)\| \cdot \|s_k^*\| + \|g_{k+1} + g_k\| \cdot \|s_k^*\|^2}{\|s_k^*\|^2} \right] \\
 &\leq \frac{\|g_{k+1}^T\| \cdot \|g_{k+1} - g_k\|}{\|g_k\|^2} + \frac{\|g_{k+1}\|}{\|g_k\|^2} \cdot \frac{2\|f(x_{k+1}) - f(x_k)\| + \|g_{k+1} + g_k\| \cdot \|s_k^*\|}{\|s_k^*\|} \\
 &\leq \bar{c}_2 + \frac{\|g_{k+1}\|}{\|g_k\|^2} \left(\frac{2\|f(x_{k+1}) - f(x_k)\|}{\|s_k^*\|} + \|g_{k+1} + g_k\| \right) \\
 &\leq \frac{c_1 \|d_k^*\|}{(1 + 2Mc_1^{-1})^{-2} \|d_k^*\|^2} \left[\frac{M \cdot \|s_k^*\|^2}{\|s_k^*\|} + (1 + c_1) \|d_k^*\| \right] + \bar{c}_2 \\
 &\leq \bar{c}_2 + [c_1(1 + 2Mc_1^{-1})^2(M + c_1 + 1)] \equiv \bar{c}_2
 \end{aligned}$$

Let $c_2 = \max\{\bar{c}_2, \bar{c}_2\}$. We obtain

$$|\beta_{k+1}^{PRP^*}| \leq c_2$$

and

$$|\theta_{k+1}^*| = \left| \frac{g_{k+1}^T d_k^*}{\|g_k\|^2} \right| \leq \|g_{k+1}\| \cdot \frac{\|d_k^*\|}{\|g_k\|^2} \leq c_1 \frac{\|d_k^*\|^2}{\|g_k\|^2} \leq c_1(1 + 2Mc_1^{-1})^2 \equiv c_3.$$

Using the definition of d_{k+1}^* , we obtain

$$\begin{aligned}
 \|d_{k+1}^*\| &= \|-g_{k+1} + \beta_{k+1}^{PRP^*} d_k^* - \theta_{k+1}^* y_k^*\| \\
 &\leq \|g_{k+1}\| + |\beta_{k+1}^{PRP^*}| \cdot \|d_k^*\| + |\theta_{k+1}^*| \cdot \|y_k^*\|.
 \end{aligned}$$

Case 1: If $\rho_k \leq 0$, we have

$$\begin{aligned}
 \|d_{k+1}^*\| &\leq \|g_{k+1}\| + |\beta_{k+1}^{PRP^*}| \cdot \|d_k^*\| + |\theta_{k+1}^*| (\|g_{k+1}\| + \|g_k\|) \\
 &= (1 + |\theta_{k+1}^*|) \|g_{k+1}\| + |\beta_{k+1}^{PRP^*}| \cdot \|d_k^*\| + |\theta_{k+1}^*| \cdot \|g_k\| \\
 &\leq (1 + c_3)c_1 \|d_k^*\| + c_2 \|d_k^*\| + c_3 \|d_k^*\| \\
 &= [(1 + c_3)c_1 + c_2 + c_3] \cdot \|d_k^*\| \equiv \bar{c}_4 \|d_k^*\|.
 \end{aligned}$$

Case 2: If $\rho_k > 0$, we have

$$\|y_k^*\| \leq 3M\alpha_k \|d_k^*\| \leq 3M \|d_k^*\|, \quad (0 < \alpha_k \leq 1),$$

and

$$\begin{aligned}
 \|d_{k+1}^*\| &\leq \|g_{k+1}\| + |\beta_{k+1}^{PRP^*}| \cdot \|d_k^*\| + |\theta_{k+1}^*| \cdot \|y_k^*\| \\
 &\leq c_1 \|d_k^*\| + c_2 \|d_k^*\| + 3c_3 M \|d_k^*\| \\
 &= (c_1 + c_2 + 3Mc_3) \|d_k^*\| \equiv \bar{c}_4 \|d_k^*\|.
 \end{aligned}$$

Letting $c_4 = \max\{\bar{c}_4, \bar{c}_4\}$, we obtain $\|d_{k+1}^*\| \leq c_4 \|d_k^*\|$. The proof is complete.

Assumption (ii) In some neighborhood N of x^* , $\nabla^2 f$ is Lipschitz continuous.

Similar to [34], we can also establish the n -order quadratic convergence of the $RMPRP^*$ method. Here, we state the theorem but omit the proof.

Table 1. Definition of the benchmark problems and their features.

No.	Functions	Definition	Multimodal?	Separable?	Regular?
1	Sphere	$f_{Sph}(x) = \sum_{i=1}^p x_i^2$ $x_i \in [-5.12, 5.12], x^* = (0, 0, \dots, 0), f_{Sph}(x^*) = 0.$	no	yes	n/a
2	Schwefel's	$f_{SchDS}(x) = \sum_{i=1}^p (\sum_{j=1}^i x_j)^2$ $x_i \in [-65.536, 65.536], x^* = (0, 0, \dots, 0), f_{SchDS}(x^*) = 0.$	no	no	n/a
3	Rastrigin	$f_{Ras}(x) = 10p + \sum_{i=1}^p (x_i^2 - 10\cos(2\pi x_i))$ $x_i \in [-5.12, 5.12], x^* = (0, 0, \dots, 0), f_{Ras}(x^*) = 0.$	yes	yes	n/a
4	Schwefel	$f_{Sch}(x) = 418.9829p + \sum_{i=1}^p x_i \sin \sqrt{ x_i }$ $x_i \in [-512.03, 511.97], x^* = (-420.9678, -420.9678, \dots, -420.9678), f_{Sch}(x^*) = 0.$	yes	yes	n/a
5	Griewank	$f_{Gri}(x) = 1 + \sum_{i=1}^p \frac{x_i^2}{4000} - \prod_{i=1}^p \cos \frac{x_i}{i}$ $x_i \in [-600, 600], x^* = (0, 0, \dots, 0), f_{Gri}(x^*) = 0.$	yes	no	yes
6	Rosenbrock	$f_{Ros}(x) = \sum_{i=1}^{p-1} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2]$ $x_i \in [-2.048, 2.048], x^* = (1, \dots, 1), f_{Ros}(x^*) = 0.$	no	no	n/a
7	Ackley	$f_{Ack}(x) = 20 + e - 20e^{-0.2\sqrt{\frac{1}{p}\sum_{i=1}^p x_i^2}} - e^{\frac{1}{p}\sum_{i=1}^p \cos(2\pi x_i)}$ $x_i \in [-30, 30], x^* = (0, \dots, 0), f_{Ack}(x^*) = 0.$	yes	no	yes
8	Langerman	$f_{Lan}(x) = -\sum_{i=1}^m c_i e^{-\frac{1}{2}\sum_{j=1}^p (x_j - a_{ij})^2} \cos(\pi \sum_{j=1}^p (x_j - a_{ij}))$ $x_i \in [0, 10], m = p, x^* = random, f_{Lan}(x^*) = random.$	yes	no	no

doi:10.1371/journal.pone.0137166.t001

Theorem 0.4 Let Assumptions (i) and (ii) hold. Then, there exists a constant $c' > 0$ such that

$$\limsup_{k \rightarrow \infty} \frac{\|x_{kr+n} - x^*\|}{\|x_{kr} - x^*\|^2} \leq c' < \infty. \quad (30)$$

Specifically, the RMPRP* method is quadratically convergent.

Numerical Results

This section reports on various numerical experiments with Algorithm 4.1 and the normal algorithm to demonstrate the effectiveness of the given algorithm. We utilize the normal algorithm (called **Algorithm N**), whereby the algorithm does not use the restart technique, but the other steps are the same as those in Algorithm 4.1. We will test the following benchmark problems using these two algorithms. These problems are listed in Table 1. These benchmark problems and discussions concerning the choice of tested problems for an algorithm can be found at

<http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume24/ortizboyer05a-html/node6.html>.

We will use these benchmark problems to perform the experiments with these two algorithms. The code was written in MATLAB 7.6.0 and run on a PC with a Core 2 Duo CPU, E7500 @2.93 GHz, 2.00 GB memory and Windows XP operation system. The parameters of the algorithms are chosen to be $\rho = 0.1$, $\gamma = 10$, $\sigma_1 = 0.0001$, and $\gamma_k = e_1 = \epsilon = 10^{-4}$. The dimension can be found in Tables 2–3. Because the line search cannot always ensure that the descent condition $d_k^T g_k < 0$ holds, uphill searches will occur in the experiments. To avoid this situation, the step size α_k will be accepted if the searching number is larger than ten in every line search technique. The *Himmeblau* stop rule will be used:

If $|f(x_k)| > e_1$, set $stop1 = \frac{|f(x_k) - f(x_{k+1})|}{|f(x_k)|}$; otherwise, set $stop1 = |f(x_k) - f(x_{k+1})|$.

For each problem, the program will stop if $\|g(x)\| < e_3$ or $stop1 < e_2$ is satisfied, where $e_1 = e_2 = e_3 = 10^{-4}$.

Table 2. Test results using Algorithm 4.1.

No.	Dim x_0	NI/NFG/ $f(\bar{x})$ (-5, ...)	NI/NFG/ $f(\bar{x})$ (-3, ...)	NI/NFG/ $f(\bar{x})$ (2, ...)	NI/NFG/ $f(\bar{x})$ (4, ...)
1	30	2/6/8.355554e-023	2/6/3.274972e-024	2/6/1.455543e-024	2/6/5.822172e-024
	10000	2/6/2.788150e-020	2/6/1.091657e-021	2/6/4.851810e-022	2/6/1.940724e-021
	100000	2/6/5.970100e-019	2/6/8.813541e-020	2/6/4.871394e-021	2/6/1.948557e-020
	103000	2/6/5.962874e-019	2/6/9.164034e-020	2/6/4.997364e-021	2/6/1.998946e-020
	x_0	(-0.002, ...)	(-0.001, ...)	(0.001, ...)	(0.0002, ...)
2	30	4/12/7.338081e-006	4/12/1.834520e-006	4/12/1.834520e-006	3/9/4.070200e-007
	60	5/15/1.607584e-005	4/12/1.503191e-005	4/12/1.503191e-005	3/9/3.227252e-006
	100	6/18/2.574911e-005	5/15/1.889258e-005	5/15/1.889258e-005	4/12/2.785399e-006
	200	8/24/3.830680e-005	7/21/2.110989e-005	7/21/2.110989e-005	5/15/6.065297e-006
	x_0	(-0.02, 0, ...)	(0.01, 0, ...)	(0.001, 0, ...)	(0.006, 0, ...)
3	30	3/9/0.000000e+000	3/9/0.000000e+000	2/6/4.547474e-013	3/9/0.000000e+000
	60	3/9/0.000000e+000	3/9/0.000000e+000	2/6/1.136868e-012	3/9/0.000000e+000
	100	3/9/0.000000e+000	3/9/0.000000e+000	2/6/1.136868e-012	3/9/0.000000e+000
	200	3/9/0.000000e+000	3/9/0.000000e+000	2/6/1.136868e-012	3/9/0.000000e+000
	x_0	(-500, ...)	(-10, ...)	(-300, ...)	(50, ...)
4	30	9/28/4.588939e+002	5/15/1.233277e+004	10/30/8.751388e+003	2/16/1.469607e+004
	1000	9/28/1.529646e+004	5/15/4.110923e+005	10/30/2.917129e+005	2/16/4.898690e+005
	10000	9/28/1.529646e+005	5/15/4.110923e+006	10/30/2.917129e+006	2/16/4.898690e+006
	100000	9/28/1.529646e+006	5/15/4.110923e+007	10/30/2.917129e+007	2/16/4.898690e+007
	x_0	(-2, ...)	(-1, ...)	(1, ...)	(2, ...)
5	30	2/6/8.548717e-015	2/6/5.644248e-009	2/6/5.644248e-009	2/6/8.548717e-015
	1000	2/6/0.000000e+000	2/6/0.000000e+000	2/6/0.000000e+000	2/6/0.000000e+000
	10000	2/6/0.000000e+000	2/6/0.000000e+000	2/6/0.000000e+000	2/6/0.000000e+000
	100000	2/6/0.000000e+000	2/6/0.000000e+000	2/6/0.000000e+000	2/6/0.000000e+000
	x_0	(1.01, ...)	(0.9999, ...)	(1.003, ...)	(1.005, ...)
6	30	12/36/6.832422e-005	2/6/1.919312e-006	6/18/5.497859e-005	9/27/6.152662e-005
	1000	12/36/6.562848e-005	3/9/3.607120e-007	6/18/5.343543e-005	9/27/5.799691e-005
	10000	12/36/6.382394e-005	3/9/3.615221e-007	7/21/4.048983e-005	9/27/6.078016e-005
	100000	12/36/5.969884e-005	3/9/3.632823e-007	7/21/4.484866e-005	9/27/5.885821e-005
	x_0	(-0.1, ...)	(0.1, ...)	(0.01, ...)	(0.03, ...)
7	30	11/50/1.684545e+000	11/50/1.684545e+000	11/58/1.684467e+000	6/25/1.684388e+000
	1000	14/67/1.717428e+000	14/67/1.717428e+000	9/46/1.717374e+000	11/56/1.717448e+000
	10000	15/75/1.718189e+000	15/75/1.718189e+000	9/46/1.718252e+000	11/57/1.718327e+000
	100000	15/75/1.718273e+000	15/75/1.718273e+000	9/46/1.718340e+000	11/57/1.718410e+000
	x_0	(-5, ...)	(-1, ...)	(2, ...)	(4, ...)
8	30	1/3/-1.889333e-104	2/16/-8.226096e-004	1/3/-2.009161e-015	1/3/-1.228836e-064
	300	1/3/0.000000e+000	1/3/-3.893431e-040	1/3/-1.008253e-163	1/3/0.000000e+000
	500	1/3/0.000000e+000	1/3/-1.459270e-067	1/3/-4.297704e-274	1/3/0.000000e+000
	1000	1/3/0.000000e+000	1/3/-2.213361e-136	1/3/0.000000e+000	1/3/0.000000e+000

doi:10.1371/journal.pone.0137166.t002

Table 3. Test results using Algorithm N.

No.	Dim x_0	NI/NFG/ $f(\bar{x})$ (-5, ...)	NI/NFG/ $f(\bar{x})$ (-3, ...)	NI/NFG/ $f(\bar{x})$ (2, ...)	NI/NFG/ $f(\bar{x})$ (4, ...)
1	30	2/6/8.355554e-023	2/6/3.274972e-024	2/6/1.455543e-024	2/6/5.822172e-024
	10000	2/6/2.788150e-020	2/6/1.091657e-021	2/6/4.851810e-022	2/6/1.940724e-021
	100000	2/6/5.970100e-019	2/6/8.813541e-020	2/6/4.871394e-021	2/6/1.948557e-020
	103000	2/6/5.962874e-019	2/6/9.164034e-020	2/6/4.997364e-021	2/6/1.998946e-020
	x_0	(-0.002, ...)	(-0.001, ...)	(0.001, ...)	(0.0002, ...)
2	30	4/12/7.338081e-006	4/12/1.834520e-006	4/12/1.834520e-006	3/9/4.070200e-007
	60	5/15/1.607584e-005	4/12/1.503191e-005	4/12/1.503191e-005	3/9/3.227252e-006
	100	6/18/2.574911e-005	5/15/1.889258e-005	5/15/1.889258e-005	4/12/2.785399e-006
	200	8/24/3.830680e-005	7/21/2.110989e-005	7/21/2.110989e-005	5/15/6.065297e-006
	x_0	(-0.02, 0, ...)	(0.01, 0, ...)	(0.001, 0, ...)	(0.006, 0, ...)
3	30	3/9/0.000000e+000	3/9/0.000000e+000	2/6/4.547474e-013	3/9/0.000000e+000
	60	3/9/0.000000e+000	3/9/0.000000e+000	2/6/1.136868e-012	3/9/0.000000e+000
	100	3/9/0.000000e+000	3/9/0.000000e+000	2/6/1.136868e-012	3/9/0.000000e+000
	200	3/9/0.000000e+000	3/9/0.000000e+000	2/6/1.136868e-012	3/9/0.000000e+000
	x_0	(-500, ...)	(-10, ...)	(-300, ...)	(50, ...)
4	30	9/28/4.588939e+002	5/15/1.233277e+004	10/30/8.751388e+003	2/16/1.469607e+004
	1000	9/28/1.529646e+004	5/15/4.110923e+005	10/30/2.917129e+005	2/16/4.898690e+005
	10000	9/28/1.529646e+005	5/15/4.110923e+006	10/30/2.917129e+006	2/16/4.898690e+006
	100000	9/28/1.529646e+006	5/15/4.110923e+007	10/30/2.917129e+007	2/16/4.898690e+007
	x_0	(-2, ...)	(-1, ...)	(1, ...)	(2, ...)
5	30	2/6/8.548717e-015	2/6/5.644248e-009	2/6/5.644248e-009	2/6/8.548717e-015
	1000	2/6/0.000000e+000	2/6/0.000000e+000	2/6/0.000000e+000	2/6/0.000000e+000
	10000	2/6/0.000000e+000	2/6/0.000000e+000	2/6/0.000000e+000	2/6/0.000000e+000
	100000	2/6/0.000000e+000	2/6/0.000000e+000	2/6/0.000000e+000	2/6/0.000000e+000
	x_0	(1.01, ...)	(0.9999, ...)	(1.003, ...)	(1.005, ...)
6	30	16/48/8.156782e-005	2/6/1.919312e-006	6/18/5.497859e-005	9/27/6.152662e-005
	1000	15/45/8.236410e-005	3/9/3.607120e-007	6/18/5.343543e-005	9/27/5.799691e-005
	10000	14/42/8.568194e-005	3/9/3.615221e-007	7/21/4.048983e-005	9/27/6.078016e-005
	100000	13/39/8.458113e-005	3/9/3.632823e-007	7/21/4.484866e-005	9/27/5.885821e-005
	x_0	(-0.1, ...)	(0.1, ...)	(0.01, ...)	(0.03, ...)
7	30	11/50/1.684545e+000	11/50/1.684545e+000	11/58/1.684467e+000	6/25/1.684388e+000
	1000	14/67/1.717428e+000	14/67/1.717428e+000	9/46/1.717374e+000	11/56/1.717448e+000
	10000	15/75/1.718189e+000	15/75/1.718189e+000	9/46/1.718252e+000	11/57/1.718327e+000
	100000	15/75/1.718273e+000	15/75/1.718273e+000	9/46/1.718340e+000	11/57/1.718410e+000
	x_0	(-5, ...)	(-1, ...)	(2, ...)	(4, ...)
8	30	1/3/-1.889333e-104	2/16/-8.226096e-004	1/3/-2.009161e-015	1/3/-1.228836e-064
	300	1/3/0.000000e+000	1/3/-3.893431e-040	1/3/-1.008253e-163	1/3/0.000000e+000
	500	1/3/0.000000e+000	1/3/-1.459270e-067	1/3/-4.297704e-274	1/3/0.000000e+000
	1000	1/3/0.000000e+000	1/3/-2.213361e-136	1/3/0.000000e+000	1/3/0.000000e+000

doi:10.1371/journal.pone.0137166.t003

Table 4. The performance of Algorithm 4.1 and Algorithm N on NFG.

Algorithm 4.1	Algorithm N
0.99	1

doi:10.1371/journal.pone.0137166.t004

The program also stops if more than five thousand iterations are performed because the corresponding method for the problem is regarded as having failed. The meanings of the columns in Tables 2–3 are as follows:

x_0 : the initial point NFG: the total number of NF and NG, i.e., $NFG = NF + NG$.

NI: the total number of iterations Dim: the dimension of the problem;

$f(\bar{x})$ denotes the function value at the point $\bar{x}$ when the program stops.

The results in Tables 2–3 show that these two algorithms effectively solve the benchmark problems—except for the fourth problem. In the experiment, the results when the algorithms are applied to problems 2 and 3 are not satisfactory when the dimensions of the problem are large; thus, we use a lower dimension. The dimensions of problems 8 and 9 are less than 1,000; the reasoning is similar to that for problems 2 and 3. However, the dimensions of all the problems are larger than 30, which is fixed. For many problems, the results are similar, except for those of problem 6. Clearly, the restart algorithm is competitive with the normal algorithm without using the restart technique.

To clearly to show the performance of Algorithm 4.1 and Algorithm N on NFG, we use the tool (S1 File) in [46] to analyze the algorithms. The results are listed in Table 4. It is easy to see that Algorithm 4.1 outperforms Algorithm N by approximately 1%, and we can conclude that the proposed method is better than the normal method. Thus, we hope that the given algorithm will be utilized in the future.

Conclusion

1. This paper presented a new conjugate gradient method that exhibits global convergence, linear convergence, and quadratic convergence when suitable assumptions are made. Our proposed CG formula includes not only the gradient value information but also the function value information. For the test benchmark problems, the numerical results showed that the given algorithm is more effective than the normal method without using the restart technique.
2. We performed tests using benchmark problems using the presented algorithm and the normal algorithm without employing the restart technique. These two methods were shown to be very effective for solving the given problems. Moreover, we did not fix the dimension because $n = 30$, and the largest dimension was higher than 100,000 (103,000). Additional numerical problems (such as the problems in [47]) should be investigated in the future to examine this algorithm.
3. Recently, we solved nonsmooth optimization problems using the relative gradient methods and obtained various results; therefore, in the future, we will use the *RMPRP** method to solve nonsmooth optimization problems and hopefully obtain interesting results. Moreover, we will study the convergence of the CG method with other line search rules.

Supporting Information

S1 File. Supporting Information for Table 4.
(PDF)

Acknowledgments

The authors would like to thank the editor and the referees for their useful suggestions and comments, which greatly improved the paper.

Author Contributions

Conceived and designed the experiments: XZ XW XL. Performed the experiments: XZ XW. Analyzed the data: XW XD. Contributed reagents/materials/analysis tools: XL XZ XW. Wrote the paper: XW XZ.

References

1. Gu B and Sheng VS. Feasibility and finite convergence analysis for accurate on-line v-support vector learning. *IEEE Transactions on Neural Networks and Learning Systems*. 2013; 24:1304–1315. doi: [10.1109/TNNLS.2013.2250300](https://doi.org/10.1109/TNNLS.2013.2250300)
2. Li J, Li XL, Yang B, Sun XM. Segmentation-based Image Copy-move Forgery Detection Scheme, *IEEE Transactions on Information Forensics and Security*. 2015; 10: 507–518. doi: [10.1109/TIFS.2014.2381872](https://doi.org/10.1109/TIFS.2014.2381872)
3. Wen XZ, Shao L, Fang W, and Xue Y. Efficient Feature Selection and Classification for Vehicle Detection. *IEEE Transactions on Circuits and Systems for Video Technology*. 2015;. doi: [10.1109/TCSVT.2014.2358031](https://doi.org/10.1109/TCSVT.2014.2358031)
4. Zhang H, Wu J, Nguyen TM, Sun MX. Synthetic Aperture Radar Image Segmentation by Modified Student's t-Mixture Model. *IEEE Transaction on Geoscience and Remote Sensing*. 2014; 52: 4391–4403. doi: [10.1109/TGRS.2013.2281854](https://doi.org/10.1109/TGRS.2013.2281854)
5. Fu ZJ. Achieving Efficient Cloud Search Services: Multi-keyword Ranked Search over Encrypted Cloud Data Supporting Parallel Computing. *IEICE Transactions on Communications*. E98-B, 2015: 190–200. doi: [10.1587/transcom.E98.B.190](https://doi.org/10.1587/transcom.E98.B.190)
6. Dai Y, Yuan Y. A nonlinear conjugate gradient with a strong global convergence properties. *SIAM J. Optim.* 2000; 10: 177–182. doi: [10.1137/S1052623497318992](https://doi.org/10.1137/S1052623497318992)
7. Dai Y, Yuan Y. *Nonlinear conjugate gradient Methods*. Shanghai Scientific and Technical Publishers, 1998.
8. Fletcher R. *Practical methods of optimization, Vol I: Unconstrained Optimization*. 2nd edition, Wiley, New York, 1997.
9. Fletcher R, Reeves C. Function minimization by conjugate gradients, *Comput. J.* 1964; 7: 149–154. doi: [10.1093/comjnl/7.2.149](https://doi.org/10.1093/comjnl/7.2.149)
10. Hager WW, Zhang H. A new conjugate gradient method with guaranteed descent and an efficient line search. *SIAM J. Optim.* 2005; 16: 170–192. doi: [10.1137/030601880](https://doi.org/10.1137/030601880)
11. Hestenes MR, Stiefel E. Method of conjugate gradient for solving linear equations. *J. Res. Nat. Bur. Stand.* 1952; 49:409–436. doi: [10.6028/jres.049.044](https://doi.org/10.6028/jres.049.044)
12. Liu Y, Storey C. Efficient generalized conjugate gradient algorithms part 1: Theory, *J. Appl. Math. Comput.* 1992; 69: 17–41.
13. Polak E. The conjugate gradient method in extreme problems, *Comput. Math. Mathem. Phys.* 1969; 9: 94–112. doi: [10.1016/0041-5553\(69\)90035-4](https://doi.org/10.1016/0041-5553(69)90035-4)
14. Polak E, Ribière G. Note sur la convergence de directions conjuguées. *Rev. Fran. Inf. Rech. Opérat.* 1969; 3: 35–43.
15. Wei Z, Yao S, Liu L. The convergence properties of some new conjugate gradient methods. *Appl. Math. Comput.* 2006; 183: 1341–1350. doi: [10.1016/j.amc.2006.05.150](https://doi.org/10.1016/j.amc.2006.05.150)
16. Yuan GL, Lu XW. A modified PRP conjugate gradient method. *Anna. Operat. Res.* 2009; 166: 73–90. doi: [10.1007/s10479-008-0420-4](https://doi.org/10.1007/s10479-008-0420-4)
17. Yuan GL, Lu XW, Wei ZX. A conjugate gradient method with descent direction for unconstrained optimization. *J. Comput. Appl. Math.* 2009; 233: 519–530. doi: [10.1016/j.cam.2009.08.001](https://doi.org/10.1016/j.cam.2009.08.001)

18. Dai Y. Analysis of conjugate gradient methods, Ph.D. Thesis, Institute of Computational Mathematics and Scientific/Engineering Computing. Chese Academy of Sciences, 1997.
19. Dai ZF, Tian BS. Global convergence of some modified PRP nonlinear conjugate gradient methods. *Optim. Let.*
20. Gilbert J C, Nocedal J. Global convergence properties of conjugate gradient methods for optimization. *SIAM J. Optim.* 1992; 2: 21–42. doi: [10.1137/0802003](https://doi.org/10.1137/0802003)
21. Hager WW, Zhang H. Algorithm 851: *CG_DESCENT*, A conjugate gradient method with guaranteed descent. *ACM Trans. Mathem. Soft.* 2006; 32: 113–137. doi: [10.1145/1132973.1132979](https://doi.org/10.1145/1132973.1132979)
22. Powell MJD. Nonconvex minimization calculations and the conjugate gradient method. *Lecture Notes in Mathematics*, Vol. 1066, Springer-Verlag, Berlin, 1984, pp. 122–141.
23. Powell MJD. Convergence properties of algorithm for nonlinear optimization. *SIAM Rev.* 1986; 28: 487–500. doi: [10.1137/1028154](https://doi.org/10.1137/1028154)
24. Yu GH. Nonlinear self-scaling conjugate gradient methods for large-scale optimization problems. thesis of Doctor's Degree, Sun Yat-Sen University, 2007.
25. Yuan GL. Modified nonlinear conjugate gradient methods with sufficient descent property for large-scale optimization problems, *Optim. Let.* 2009; 3: 11–21. doi: [10.1007/s11590-008-0086-5](https://doi.org/10.1007/s11590-008-0086-5)
26. Yuan YX. Analysis on the conjugate gradient method. *Optim. Meth. Soft.* 1993; 2: 19–29. doi: [10.1080/10556789308805532](https://doi.org/10.1080/10556789308805532)
27. Yuan GL, Wei ZX and Li GY. A modified Polak-Ribière-Polyak conjugate gradient algorithm for non-smooth convex programs. *Journal of Computational and Applied Mathematics.* 2014; 255: 86–96. doi: [10.1016/j.cam.2013.04.032](https://doi.org/10.1016/j.cam.2013.04.032)
28. Yuan GL, Wei ZX, and Zhao QM. A modified Polak-Ribière-Polyak conjugate gradient algorithm for large-scale optimization problems. *IIE Transactions*, 2014; 46: 397–413. doi: [10.1080/0740817X.2012.726757](https://doi.org/10.1080/0740817X.2012.726757)
29. Yuan GL, Zhang M J. A modified Hestenes-Stiefel conjugate gradient algorithm for large-scale optimization. *Numerical Functional Analysis and Optimization.* 2013; 34:914–937. doi: [10.1080/01630563.2013.777350](https://doi.org/10.1080/01630563.2013.777350)
30. Burmeister W. Die Konvergenzordnung des Fletcher-Powell Algorithmus, *Z. Angew. Math. Mech.* 1973; 53: 693–699. doi: [10.1002/zamm.19730531007](https://doi.org/10.1002/zamm.19730531007)
31. Cohen A. Rate of convergence of several conjugate gradient algorithms. *SIAM J. Numer. Anal.* 1972; 9: 248–259. doi: [10.1137/0709024](https://doi.org/10.1137/0709024)
32. Ritter K. On the rate of superlinear convergence of a class of variable metric methods, *Numer. Math.* 1980; 35: 293–313. doi: [10.1007/BF01396414](https://doi.org/10.1007/BF01396414)
33. Zhang L, Zhou W, Li DH. A descent modified Polak-Ribière-Polyak conjugate gradient method and its global convergence, *IMA J. Numer. Anal.* 2006; 26: 629–640. doi: [10.1093/imanum/drl016](https://doi.org/10.1093/imanum/drl016)
34. Li DH, Tian BS. N-step quadratic convergence of the MPRP method with a restart strategy, *J. comput. Appl. Math.* 2011; 235: 4978–4990. doi: [10.1016/j.cam.2011.04.026](https://doi.org/10.1016/j.cam.2011.04.026)
35. Broyden CG, Dennis JE, Moré JJ. On the local and superlinear convergence of quasi-Newton methods, *J. Ins. Math. Appl.*, 12 (1973), pp. 223–246. doi: [10.1093/imanamat/12.3.223](https://doi.org/10.1093/imanamat/12.3.223)
36. Byrd R, Nocedal J. A tool for the analysis of quasi-Newton methods with application to unconstrained minimization. *SIAM J. Numer. Anal.* 1989; 26: 727–739. doi: [10.1137/0726042](https://doi.org/10.1137/0726042)
37. Byrd R, Nocedal J, Yuan Y. Global convergence of a class of quasi-Newton methods on convex problems. *SIAM J. Numer. Anal.* 1987; 24: 1171–1189. doi: [10.1137/0724077](https://doi.org/10.1137/0724077)
38. Dai Y. Convergence properties of the BFGS algorithm. *SIAM J. Optim.* 2003; 13: 693–701. doi: [10.1137/S1052623401383455](https://doi.org/10.1137/S1052623401383455)
39. Dennis JE, Moré JJ. A characterization of superlinear convergence and its application to quasi-Newton methods. *Math. Comput.* 1974; 28:549–560. doi: [10.1090/S0025-5718-1974-0343581-1](https://doi.org/10.1090/S0025-5718-1974-0343581-1)
40. Li D, Fukushima M A modified BFGS method and its global convergence in nonconvex minimization. *J. Comput. Appl. Math.* 2001; 129: 15–35. doi: [10.1016/S0377-0427\(00\)00540-9](https://doi.org/10.1016/S0377-0427(00)00540-9)
41. Li D, Fukushima M. On the global convergence of the BFGS method for nonconvex unconstrained optimization problems. *SIAM J. Optim.* 2001; 11: 1054–1064. doi: [10.1137/S1052623499354242](https://doi.org/10.1137/S1052623499354242)
42. Liu GH, Han JY, Sun DF. Global convergence Analysis of the BFGS Algorithm with Nonmonotone line-search. *Optim.* 1995; 34: 147–159. doi: [10.1080/02331939508844101](https://doi.org/10.1080/02331939508844101)
43. Wei Z, Yu G, Yuan G, Lian Z. The superlinear convergence of a modified BFGS-type method for unconstrained optimization. *Comput. Optim. Appl.* 2004; 29: 315–332. doi: [10.1023/B:COAP.0000044184.25410.39](https://doi.org/10.1023/B:COAP.0000044184.25410.39)

44. Wei Z, Li G, Qi L. New quasi-Newton methods for unconstrained optimization problems. *Appl. Math. Comput.* 2006; 175: 1156–1188. doi: [10.1016/j.amc.2005.08.027](https://doi.org/10.1016/j.amc.2005.08.027)
45. Yuan GL, Wei ZX. Convergence analysis of a modified BFGS method on convex minimizations. *Comput. Optim. Appl.* 2010; 47: 237–255. doi: [10.1007/s10589-008-9219-0](https://doi.org/10.1007/s10589-008-9219-0)
46. Yuan GL, Lu XW. A modified PRP conjugate gradient method. *Annals of Operations Research*. 2009; 166: 73–90. doi: [10.1007/s10479-008-0420-4](https://doi.org/10.1007/s10479-008-0420-4)
47. Nicholas I, Gould M, Dominique Orban and Philippe L. Toint, CUTer(and SifDec): A Constrained and Unconstrained Testing Environment, revisited. *ACM Trans. Mathem. Soft.* 2003; 29: 373–394.