A short note on the global convergence of the unmodified PRP method*

Weijun Zhou†

September 20, 2011

Abstract

It is well-known that the search direction generated by the standard (unmodified) PRP nonlinear conjugate gradient method is not necessarily a descent direction of the objective function, which brings difficulty for its global convergence for general functions. However, to our surprise, it is easily proved in this short note that the unmodified PRP method still converges globally even for nonconvex optimization by the use of an approximate descent line search.

Keywords. The PRP method, global convergence, backtracking line search.

AMS subject classification 2010. 90C30, 65K05.

1 Introduction

It is well-known that the PRP method [6, 7] is one of the most efficient nonlinear conjugate gradient methods for optimization. However, a shortcoming of the PRP method is that it can not guarantee to produce a descent direction of the objective function, which brings difficulty for its global convergence for general functions. To ensure global convergence of the PRP method, some modifications (or some line searches which force it generate descent direction such as that of [3]) are often adopted such as the PRP+ method [2] and the three-term PRP method [8]. We refer to the book [1] and the survey paper [4] for more related topics.

The purpose of this paper is to investigate the global convergence of the unmodified PRP method for general unconstrained optimization. Consider the problem:

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

where $f : R^n \to R$ is a smooth function and its gradient $g(x) \triangleq \nabla f(x)$ is available. The direction generated by the standard PRP method is given by

$$d_k = \begin{cases} -g_k, & \text{if} \quad k = 0, \\ -g_k + \beta_k^{PRP} d_{k-1}, & \text{if} \quad k \geq 1, \end{cases} \tag{1.2}$$

where $g_k = \nabla f(x_k)$ and

$$\beta_k^{PRP} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}, \quad y_{k-1} = g_k - g_{k-1}. \tag{1.3}$$

*This work was supported by the NSF foundation (10901026) of China and The Open Fund Project of Key Research Institute of Philosophies and Social Sciences in Hunan Universities.

†Department of Mathematics, Changsha University of Science and Technology, Changsha 410004, China (Email: weijunzhou@126.com).

In this short note we use another way to globalize the PRP method, that is, we use the following approximate function descent backtracking line search to compute the steplength $\alpha_k = \max\{1, \rho^1, \rho^2, \ldots\}$ satisfying

$$f(x_k + \alpha_k d_k) \le f(x_k) - \delta\|\alpha_k d_k\|^2 + \epsilon_k, \tag{1.4}$$

where $\{\epsilon_k\}$ is a positive sequence satisfying

$$\sum_{k=0}^{\infty} \epsilon_k \le \epsilon < \infty, \tag{1.5}$$

$\rho \in (0, 1)$, ϵ and δ are positive constants.

It is clear that the line search (1.4) is well-defined whether d_k is a descent direction or not, which is a variation of the approximate norm descent line search proposed by Li and Fukushima [5] for solving nonlinear equations.

For clarity, we give the steps of the PRP method for (1.1) as follows.

Algorithm 1.

Step 0. Choose $x_0 \in R^n$, $\delta > 0$, $\rho \in (0, 1)$, $\epsilon > 0$ and a positive sequence $\{\epsilon_k\}$ satisfying (1.5). Let $k := 0$.

Step 1. Compute d_k by (1.2)-(1.3).

Step 2. Compute the stepsize α_k by (1.4).

Step 3. Set $x_{k+1} = x_k + \alpha_k d_k$. Let $k := k + 1$ and go to Step 1.

In the next section, we make some assumptions and show its global convergence. Throughout the paper, we denote $s_k = x_{k+1} - x_k = \alpha_k d_k$.

2 Global convergence

To begin with, let us define the level set

$$\Omega = \{x| \ f(x) \le f(x_0) + \epsilon\}. \tag{2.1}$$

It is clear that $x_k \in \Omega$ for all $k \ge 0$. Then we make the following assumptions to ensure global convergence of Algorithm 1.

Assumption 1.

(i) The level set Ω defined by (2.1) is bounded.

(ii) In some neighborhood N of Ω, the gradient is Lipschitz continuous, namely, there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \le L\|x - y\|, \quad \forall x, y \in N. \tag{2.2}$$

Assumption 1 implies that there exists a positive constant M such that

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

Lemma 2.1. *Let Assumption 1 hold. Then we have*

$$\sum_{k=0}^{\infty} \|\alpha_k d_k\|^2 = \sum_{k=0}^{\infty} \|s_k\|^2 < \infty.$$

Proof. It follows from the line search (1.4) and (1.5) directly. □

Lemma 2.1 implies

$$\lim_{k\to\infty} \alpha_k d_k = \lim_{k\to\infty} s_k = 0. \tag{2.4}$$

Lemma 2.2. *Let Assumption 1 hold. If there is a constant $\tau > 0$ such that*

$$\|g_k\| \geq \tau, \quad \forall k \geq 0, \tag{2.5}$$

then there is a constant $M_1 > 0$ such that

$$\|d_k\| \leq M_1, \quad \forall k \geq 0. \tag{2.6}$$

Proof. The proof is similar to that of Lemma 3.1 in [8]. From (1.3), (2.3), (2.5), (2.2) and (2.4), we have

$$|\beta_k^{PRP}| \leq \frac{\|g_k\|\|y_{k-1}\|}{\|g_{k-1}\|^2} \leq \frac{ML}{\tau^2}\|s_{k-1}\| \to 0, \tag{2.7}$$

which implies that there exist an integer k_0 and a constant $r \in (0,1)$ such that

$$|\beta_k^{PRP}| \leq r, \forall k \geq k_0.$$

This and (1.2) yield

$$\begin{aligned} \|d_k\| &\leq \|g_k\| + |\beta_k|\|d_{k-1}\| \\ &\leq M + r\|d_{k-1}\| \\ &\leq M(1 + r + r^2 + \cdots + r^{k-k_0-1}) + r^{k-k_0}\|d_{k_0}\| \\ &\leq \frac{M}{1-r} + \|d_{k_0}\| \\ &\leq M_1, \end{aligned}$$

where $M_1 = \max\{\|d_0\|, \|d_1\|, \ldots, \|d_{k_0}\|, \frac{M}{1-r} + \|d_{k_0}\|\}$. This finishes the proof. □

Theorem 2.1. *Let Assumption 1 hold. Then the sequence $\{x_k\}$ generated by Algorithm 1 converges globally in the sense that*

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{2.8}$$

Proof. We prove this theorem by contradiction. Suppose (2.8) is not true, then there exists a positive constant τ such that (2.5) holds for all $k \geq 0$. Hence Lemma 2.2 holds.

(i) If $\liminf_{k\to\infty} \|d_k\| = 0$, then from (1.2), (2.7) and (2.6), we get

$$\liminf_{k\to\infty} \|g_k\| = \liminf_{k\to\infty} \|d_k - \beta_k^{PRP} d_{k-1}\| \leq \liminf_{k\to\infty} \|d_k\| + \lim_{k\to\infty} \|\beta_k^{PRP} d_{k-1}\| = 0,$$

which contradicts to (2.5).

(ii) If $\liminf_{k\to\infty} \|d_k\| > 0$, then (2.4) implies

$$\lim_{k\to\infty} \alpha_k = 0. \tag{2.9}$$

This and (1.4) show that $\alpha'_k = \alpha_k/\rho$ does not satisfy (1.4), namely,

$$f(x_k + \alpha'_k d_k) \geq f(x_k) - \delta\|\alpha'_k d_k\|^2 + \epsilon_k > f(x_k) - \delta\|\alpha'_k d_k\|^2, \tag{2.10}$$

which yields

$$\frac{f(x_k + \alpha'_k d_k) - f(x_k)}{\alpha'_k} \geq -\delta \alpha'_k \|d_k\|^2. \tag{2.11}$$

By the mean-value theorem and (1.2), there exists $\theta_k \in (0, 1)$ such that

$$\begin{aligned} & \frac{f(x_k + \alpha'_k d_k) - f(x_k)}{\alpha'_k} \\ = \; & g(x_k + \theta_k \alpha'_k d_k)^T d_k \\ = \; & -g(x_k + \theta_k \alpha'_k d_k)^T g_k + g(x_k + \theta_k \alpha'_k d_k)^T (\beta_k^{PRP} d_{k-1}). \end{aligned} \tag{2.12}$$

Note that (2.7), (2.6) and (2.3) imply

$$\lim_{k \to \infty} g(x_k + \theta_k \alpha'_k d_k)^T (\beta_k^{PRP} d_{k-1}) = 0. \tag{2.13}$$

Since $\{x_k\} \subset \Omega$ is bounded. Then without loss of generality, we assume $x_k \to x^*$. Let $k \to \infty$ in (2.11)-(2.13), we obtain

$$-g(x^*)^T g(x^*) = -\|g(x^*)\|^2 \geq 0,$$

which means

$$g(x^*) = 0.$$

This leads to a contradiction to (2.5). The proof is then completed. □

References

[1] Y. H. Dai and Y. Yuan, Nonlinear Conjugate Gradient Methods, Shanghai Science and Technology Publisher, Shanghai, 2000.

[2] J. C. Gilbert and J. Nocedal, Global convergence properties of conjugate gradient methods for optimization, SIAM Journal on Optimization, 2 (1992), pp. 21-42.

[3] L. Grippo and S. Lucidi, A globally convergent version of the Polak-Ribière conjugate gradient method, Mathematical Programming, 78 (1997), pp. 375-391.

[4] W. W. Hager and H. Zhang, A survey of nonlinear conjugate gradient methods, Pacific Journal of Optimization, 2 (2006), pp. 35-58.

[5] D. Li and M. Fukushima, A globally and superlinearly convergent Gauss-Newton-based BFGS method for symmetric nonlinear equations, SIAM Journal on Numerical Analysis, 37 (1999), pp. 152-172.

[6] E. Polak and G. Ribière, Note surla convergence de méthodes de directions conjuguées, Revue Francaise dinformatique et de Recherche Opérationnele, 16 (1969), pp. 35-43.

[7] B. T. Polyak, The conjugate gradient method in extreme problems, USSR Computational Mathematics and Mathematical Physics, 9 (1969), pp. 94-112.

[8] L. Zhang, W. Zhou and D. Li, A descent modified Polak-Ribière-Polyak conjugate gradient method and its global convergence, IMA Journal of Numerical Analysis, 26 (2006), pp. 629-640.