Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 17, No. 1, pp. 47–55, 2021

A PROJECTED PRP METHOD FOR OPTIMIZATION WITH CONVEX CONSTRAINT*

Weijun Zhou

Abstract: In this paper, we introduce a projected PRP method for optimization with convex constraint and establish its global convergence, which is a direct generalization of the classical PRP method for unconstrained optimization. We also did some numerical experiments to show its efficiency.

Key words: *PRP, projected, line search, global convergence*

Mathematics Subject Classification: *90C30*

1 Introduction

The classical Polak-Ribière-Polyak (PRP) method is very popular for solving smooth unconstrained optimization, which has been regarded as one of the most efficient nonlinear conjugate gradient methods [11, 12]. Similar to the BFGS method [18, 20, 25], the PRP method also has been greatly extended to solve nonlinear equations [16, 17, 23, 24].

In general, the direction generated by the PRP formula with the standard Wolfe or Armijo line search may be not a descent direction of the objective function. To guarantee global convergence of the PRP method for nonconvex problems, two different ways are often used. One way is to modify the PRP formula such as the PRP+ method and the three-term PRP method [1, 2, 4, 5, 8, 10, 19]. Another way is to adopt special search rules such as descent type line searches [7, 9, 14, 15] and non-descent type line searches [21, 22].

We note that the above described PRP type methods aim to solve unconstrained optimization. However, they are not suitable for solve constrained optimization problems directly. The objective of the paper is to generalize the unmodified PRP method to solve optimization with convex constraint by utilizing the projection and some non-descent line search.

The paper is organized as follows. In Sect. 2, we propose a projected PRP method for optimization with convex constraint in detail. In Sect. 3, we investigate global convergence property of the proposed method. In Sect. 4, we do some numerical experiments.

*This work was supported by Hunan Provincial Key Laboratory of Mathematical Modeling and Analysis in Engineering.

2 The Projected PRP Method

In this section, we first simply recall iterative form of the unmodified PRP method for unconstrained optimization. Then we present a projected PRP method for solving optimization with convex constraint.

The general scheme of the unmodified PRP method for solving the smooth unconstrained optimization problem

$$\min_{x \in R^n} f(x), \tag{2.1}$$

is given by $x_{k+1} = x_k + \alpha_k d_k, k = 0, 1, \ldots$. Here the stepsize α_k is computed by some line search, and the search direction d_k is generated by the PRP formula

$$d_k = \begin{cases} -g_k, & \text{if} \quad k = 0, \\ -g_k + \beta_k^{PRP} d_{k-1}, & \text{if} \quad k \geq 1, \end{cases} \tag{2.2}$$

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

$$\beta_k^{PRP} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}, \quad y_{k-1} = g_k - g_{k-1}. \tag{2.3}$$

Throughout the paper, we denote 2-norm by $\|\cdot\|$.

The aim of this paper is to generalize the PRP method for the problem (2.1) to solve the following constrained optimization problem

$$\min_{x \in \Omega} f(x), \tag{2.4}$$

where $\Omega \subseteq R^n$ is a closed convex set, $f : R^n \to R$ is a smooth function and its gradient $g(x) = \nabla f(x)$ is available. It is clear that if x^* is a local minimizer of the problem (2.4), then it is a stationary point which satisfies the following definition.

Definition 2.1. We say that $x^* \in \Omega$ is a stationary point of the problem (2.4) if it satisfies

$$g(x^*)^T (x - x^*) \geq 0, \quad \forall x \in \Omega.$$

Let $P_\Omega : R^n \to \Omega$ be the projected operator, that is,

$$P_\Omega(x) = \operatorname{Arg}\min_{y \in \Omega} \|y - x\|.$$

Set

$$r_k = P_\Omega(x_k - g_k) - x_k. \tag{2.5}$$

It is well-known that $r_k = 0$ if and only if x_k is a stationary point of the problem (2.4). Using the projection, we introduce the following projected PRP method for the problem (2.4).

Algorithm 1. (The projected PRP method)

Step 0. Choose $x_0 \in \Omega$, $\delta > 0$, $\rho \in (0,1)$, $0 < \lambda_{\min} < \lambda_{\max} < \infty$. Select a positive sequence $\{\eta_k\}$ satisfying

$$\sum_{k=0}^{\infty} \eta_k \leq \eta < \infty. \tag{2.6}$$

Set $d_0 = -g_0$. Let $k := 0$.

Step 1. If $r_k = 0$, then stop. Otherwise, go to Step 2.

Step 2. Compute d_k by (2.2).

Step 3. Compute the stepsize $\alpha_k = \max\{\sigma_k \rho^i |\ i = 0, 1, \ldots\}$ satisfying

$$f(P_\Omega(x_k + \alpha_k d_k)) \leq f(x_k) - \delta\|\alpha_k d_k\|^2 + \eta_k, \tag{2.7}$$

where $\sigma_k \in [\lambda_{\min}, \lambda_{\max}]$.

Step 4. Set $x_{k+1} := P_\Omega(x_k + \alpha_k d_k)$. Let $k := k + 1$ and go to Step 1.

Remark 2.2. By (2.5), if g_k=0, then $r_k = 0$, which means that x_k is a stationary point of the problem (2.4). Moreover, from the continuity of the projected operator and $\eta_k > 0$, then the line search (2.7) is satisfied for all sufficiently small $\alpha > 0$. The non-descent line search (2.7) is a modification of that of [21].

Now we give some important properties of the projected operator, which are very useful for global convergence analysis of Algorithm 1. The following two lemmas come from [3].

Lemma 2.3. *If $z \in \Omega$, then*

$$\big(P_\Omega(x) - x\big)^T \big(z - P_\Omega(x)\big) \geq 0, \quad \forall x \in R^n, \tag{2.8}$$

$$\|P_\Omega(x) - P_\Omega(y)\| \leq \|x - y\|, \quad \forall x, y \in R^n. \tag{2.9}$$

Lemma 2.4. *For any fixed $x \in \Omega$, $\frac{\|P_\Omega(x-\alpha g(x))-x\|}{\alpha}$ is nonincreasing in $\alpha > 0$.*

Lemma 2.5. *Let $x_k \in \Omega$. Then*

$$g_k^T \big(x_k - P_\Omega(x_k - \alpha g_k)\big) \geq \frac{\|P_\Omega(x_k - \alpha g_k) - x_k\|^2}{\alpha}, \quad \forall \alpha > 0. \tag{2.10}$$

Proof. In fact, from (2.8) and $x_k \in \Omega$, we have

$$\begin{aligned}
& g_k^T \big(x_k - P_\Omega(x_k - \alpha g_k)\big) \\
= \ & \frac{1}{\alpha}\Big(x_k - P_\Omega(x_k - \alpha g_k) + P_\Omega(x_k - \alpha g_k) - (x_k - \alpha g_k)\Big)^T \big(x_k - P_\Omega(x_k - \alpha g_k)\big) \\
= \ & \frac{\|P_\Omega(x_k - \alpha g_k) - x_k\|^2}{\alpha} + \frac{1}{\alpha}\big(P_\Omega(x_k - \alpha g_k) - (x_k - \alpha g_k)\big)^T \big(x_k - P_\Omega(x_k - \alpha g_k)\big) \\
\geq \ & \frac{\|P_\Omega(x_k - \alpha g_k) - x_k\|^2}{\alpha}.
\end{aligned}$$

The proof is completed. □

3 Global Convergence

In this secton, we discuss global convergence property of Algorithm 1 under the following assumptions. To begin with, let us define the level set

$$\Omega_1 = \{x|\ f(x) \le f(x_0) + \eta\} \cap \Omega, \tag{3.1}$$

where η satisfies (2.6). It is clear that $x_k \in \Omega_1$ for all $k \ge 0$.

Assumption 1.

(i) The level set Ω_1 defined by (3.1) is bounded.

(ii) There exists some convex neighborhood N of Ω_1 such that the gradient $g(x)$ is Lipschitz continuous in $N \cap \Omega$, namely, there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \le L\|x - y\|, \quad \forall x, y \in N \cap \Omega. \tag{3.2}$$

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

$$\|g(x)\| \le M, \quad \forall x \in N \cap \Omega. \tag{3.3}$$

Clearly, by the line search (2.7) and (2.6), we have

$$\lim_{k\to\infty} \alpha_k d_k = 0. \tag{3.4}$$

The following result shows that Algorithm 1 converges globally.

Theorem 3.1. *Let Assumption 1 hold and the sequence $\{x_k\}$ be generated by Algorithm 1. Then*

$$\liminf_{k\to\infty} \|r_k\| = 0. \tag{3.5}$$

Proof. We prove this theorem by contradiction. If it is not true, then there exists a constant $\tau > 0$ such that

$$\|r_k\| \ge \tau, \quad \forall k \ge 0. \tag{3.6}$$

This implies that for some positive constant τ_1,

$$\|g_k\| \ge \tau_1, \quad \forall k \ge 0. \tag{3.7}$$

Otherwise, there exists some infinite subset $K \subseteq \{0, 1, 2, \dots\}$ such that

$$\lim_{k\in K, k\to\infty} \|r_k\| = \lim_{k\in K, k\to\infty} \|P_\Omega(x_k - g_k) - x_k\| \le \lim_{k\in K, k\to\infty} \|g_k\| = 0,$$

where the last inequality uses (2.9) and the fact $P_\Omega(x_k) = x_k$. This contradicts with (3.6).

Therefore, by (2.3), (3.2), (2.9), (3.3), (3.4) and (3.7), we obtain

$$|\beta_k^{PRP}| \le \frac{L\|g_k\|\alpha_{k-1}\|d_{k-1}\|}{\tau_1^2} \to 0. \tag{3.8}$$

This together with (3.3) and (2.2) means that there exists a constant $M_1 > 0$ such that

$$\|d_k\| \le M_1. \tag{3.9}$$

By (3.8)-(3.9) and (2.2), we have

$$\beta_k^{PRP} d_{k-1} \to 0, \quad d_k \to -g_k. \tag{3.10}$$

(i) If $\limsup_{k\to\infty} \alpha_k > 0$, then (3.4) yields

$$\liminf_{k\to\infty} \|d_k\| = 0,$$

which leads a contradiction with (3.10) and (3.7).
(ii) If $\limsup_{k\to\infty} \alpha_k = 0$, then $\alpha'_k = \frac{\alpha_k}{\rho}$ can not satisfy the line search (2.7). Thus,

$$f(P_\Omega(x_k + \alpha'_k d_k)) - f(x_k) > -\delta\|\alpha'_k d_k\|^2 + \eta_k > -\delta\|\alpha'_k d_k\|^2. \tag{3.11}$$

By the mean value theorem, we have

$$\begin{aligned}
&\frac{f(P_\Omega(x_k + \alpha'_k d_k)) - f(x_k)}{\alpha'_k} = \frac{g(\xi_k)^T(P_\Omega(x_k + \alpha'_k d_k) - x_k)}{\alpha'_k} \\
= \;& \frac{g_k^T(P_\Omega(x_k - \alpha'_k g_k) - x_k)}{\alpha'_k} + \frac{(g(\xi_k) - g_k)^T(P_\Omega(x_k + \alpha'_k d_k) - x_k)}{\alpha'_k} \\
&+ \frac{g_k^T(P_\Omega(x_k + \alpha'_k d_k) - P_\Omega(x_k - \alpha'_k g_k))}{\alpha'_k} \\
= \;& \frac{g_k^T(P_\Omega(x_k - \alpha'_k g_k) - x_k)}{\alpha'_k} + \Delta_k \\
\le \;& -\frac{\|P_\Omega(x_k - \alpha'_k g_k) - x_k\|^2}{\alpha'^2_k} + \Delta_k,
\end{aligned}$$

where ξ_k lies in the segment between x_k and $P_\Omega(x_k + \alpha'_k d_k)$ and the last inequality follows from Lemma 2.5. This together with (3.11) yields

$$\frac{\|P_\Omega(x_k - \alpha'_k g_k) - x_k\|^2}{\alpha'^2_k} \le |\Delta_k| + \delta\alpha'_k\|d_k\|^2. \tag{3.12}$$

Moreover, we have

$$\begin{aligned}
&|\Delta_k| \\
\le \;& \|g(\xi_k) - g_k\| \left\| \frac{(P_\Omega(x_k + \alpha'_k d_k) - x_k)}{\alpha'_k} \right\| + \|g_k\| \left\| \frac{(P_\Omega(x_k + \alpha'_k d_k) - P_\Omega(x_k - \alpha'_k g_k))}{\alpha'_k} \right\| \\
\le \;& M_1\|g(\xi_k) - g_k\| + \|g_k\|\|d_k + g_k\| \\
= \;& M_1\|g(\xi_k) - g_k\| + M\|\beta_k^{PRP} d_{k-1}\|,
\end{aligned}$$

where the second inequality uses (2.9) and (3.9). By the continuity of g and $\alpha'_k \to 0$ and (3.10), we get that $\Delta_k \to 0$. By (2.5), Lemma 2.4 and (3.12), we obtain

$$\|r_k\|^2 = \|P_\Omega(x_k - g_k) - x_k\|^2 \le \frac{\|P_\Omega(x_k - \alpha'_k g_k) - x_k\|^2}{\alpha'^2_k} \le |\Delta_k| + \delta\alpha'_k\|d_k\|^2 \to 0.$$

This contradicts with (3.6). The proof is then completed. □

Table 1: Test results on the problem (4.2) with $\gamma = (1, 2, \ldots, n-1)^T$.

	Algorithm 1			The PG method		
n	Iter	$\|r_k\|_\infty$	Time	Iter	$\|r_k\|_\infty$	Time
100	74	9.1256e-06	0.049779	70	9.9467e-06	0.027479
500	60	8.2435e-06	0.092343	91	8.7241e-06	0.10784
1000	59	8.3732e-06	0.16285	104	9.289e-06	0.22721
1500	55	9.0037e-06	0.21005	116	8.5336e-06	0.38664
2000	62	8.9135e-06	0.30178	122	8.1618e-06	0.53083
2500	78	9.2436e-06	0.55164	128	9.8512e-06	0.71693
3000	71	9.451e-06	0.5897	131	9.2855e-06	0.91072
3500	60	9.1161e-06	0.5529	129	9.8574e-06	1.0503

Table 2: Test results on the problem (4.2) with $\gamma = \frac{1}{n}(1, 2^2, \ldots, (n-1)^2)^T$.

	Algorithm 1			The PG method		
n	Iter	$\|r_k\|_\infty$	Time	Iter	$\|r_k\|_\infty$	Time
100	72	9.7616e-06	0.044399	67	9.1474e-06	0.025313
500	60	9.2637e-06	0.092921	98	9.2032e-06	0.12744
1000	71	9.2677e-06	0.19903	102	9.5268e-06	0.2381
1500	62	8.0182e-06	0.2427	115	8.5563e-06	0.40841
2000	66	8.5835e-06	0.35081	121	8.1772e-06	0.57016
2500	75	8.7576e-06	0.54326	125	8.4656e-06	0.74873
3000	82	9.4057e-06	0.78966	130	9.3253e-06	0.98772
3500	63	8.9983e-06	0.59293	128	8.3004e-06	1.137

4 Numerical Results

In this section, we compare the performance of the following two methods for solving the problem (2.4).

• Algorithm 1. We set the parameters $\delta = 0.1$, $\rho = 0.1$, $\lambda_{min} = \lambda_{max} = 1$ and $\eta_k = 0.5^k$.

• The following classical projected gradient method (PG) [13] with the same parameters as Algorithm 1 in our numerical experiments:

Step 0. Choose $x_0 \in \Omega$, $\delta > 0$, $\rho \in (0, 1)$. Set $d_0 = -g_0$. Let $k := 0$.

Step 1. If $r_k = 0$, then stop. Otherwise, go to Step 2.

Step 2. Compute $d_k = -g_k$.

Step 3. Compute the stepsize $\alpha_k = \max\{\sigma_k \rho^i |\ i = 0, 1, \ldots\}$ satisfying

$$f(P_\Omega(x_k + \alpha_k d_k)) \leq f(x_k) + \delta g_k^T \big(P_\Omega(x_k + \alpha_k d_k) - x_k\big), \tag{4.1}$$

where $\sigma_k \in [\lambda_{\min}, \lambda_{\max}]$.

Step 4. Set $x_{k+1} := P_\Omega(x_k + \alpha_k d_k)$. Let $k := k + 1$ and go to Step 1.

We test both methods on the following problem with different γ_i.

$$f(x) = \frac{1}{2}\sum_{i=1}^{n-1}(x_i - x_{i+1})^2 + \frac{1}{12}\sum_{i=1}^{n-1}\gamma_i(x_i - x_{i+1})^4 + \frac{1}{2}x^T x \tag{4.2}$$

with the constrained set $\Omega = \{x| - 10 \leq x_i \leq 10,\ i = 1, 2, ..., n\}$, where $\gamma_i \geq 0, i = 1, 2, \ldots, n-1$, are constants. We denote $\gamma = [\gamma_1, \gamma_2, \ldots, \gamma_{n-1}]^T$.

The codes were written in Matlab R2015a. We stopped the iteration if $k \geq 500$ or $\|r_k\|_\infty \leq 10^{-5}$. We chose the initial point $x_0 = (1, 1, \ldots, 1)^T$. Numerical results are listed in Table 1 and Table 2, where Iter, $\|r_k\|_\infty$ and Time stand for the total number of iterations, the infinite norm of r_k at the stopping point and the CPU times in second respectively.

From both Tables, we can see that Algorithm 2.1 performs better than the PG method since it requires less iterations and less CPU times especially for relatively large n.

5 Conclusions

Using the projection and some non-descent line search, we present a projected PRP method for optimization with convex constraint. It is a natural extension of the classical PRP method for unconstrained optimization. We show that the proposed method converges globally. Numerical results show that the proposed method performs better than the classical projected gradient method for the given test problem. Our further study is to apply the proposed method to some financial optimization problems [6].

References

[1] N. Andrei, Another conjugate gradient algorithm with guaranteed descent and the conjugacy conditions for large-scaled unconstrained optimization, *J. Optim. Theory Appl.* 159 (2013) 159–182.

[2] S. Babaie-Kafaki, A quadratic hybridization of Polak-Ribière-Polyak and Fletcher-Reeves conjugate gradient methods, *J. Optim. Theory Appl.* 154 (2012) 916–932.

[3] P.H. Calamai and J.J. Moré, Projected gradient methods for linearly constrained problems, *Math. Program.* 39 (1987) 93–116.

[4] W. Cheng, A two term PRP-based descent method, *Numer. Funct. Anal. Optim.* 28 (2008) 1217–1230.

[5] Z. Dai and B. Tian, Global convergence of some modified PRP nonlinear conjugate gradient methods, *Optim. Lett.* 5 (2011) 615–630.

[6] Z. Dai, H. Zhu and F. Wen, *Two nonparametric approaches to mean absolute deviation portfolio selection model*, *J. Ind. Manag. Optim.* 16 (2020) 2283–2303.

[7] Y. Dai, Conjugate gradient methods with Armijo-type line searches, *Acta Math. Appl. Sin.-Engl. Ser.* 18 (2002) 123–130.

[8] J.C. Gilbert and J. Nocedal, Global convergence properties of conjugate gradient methods for optimization, *SIAM J. Optim.* 2 (1992) 21–42.

[9] L. Grippo and S. Lucidi, A globally convergent version of the Polak-Ribière conjugate gradient method, *Math. Program.* 78 (1997) 375–391.

[10] M. Li, A family of three-term nonlinear conjugate gradient methods close to the memoryless BFGS method, *Optim. Lett.* 12 (2018) 1911–1927.

[11] B.T. Polyak, The conjugate gradient method in extreme problems, *USSR Comput. Math. Math. Phys.* 9 (1969) 94–112.

[12] E. Polak and G. Ribière, Note sur la convergence de méthodes de directions conjuguées, *Rev. Fr. Inform. Rech. Oper.* 16 (1969) 35–43.

[13] W. Sun and Y. Yuan, *Optimization Theory and Methods*, Springer Science and Business Media, LLC, New York, 2006.

[14] G. Yuan, Z. Wei and X. Lu, Global convergence of BFGS and PRP methods under a modified weak Wolfe-Powell line search, *Appl. Math. Model.* 47 (2017) 811–825.

[15] G. Yuan, Z. Wei and Y. Yang, The global convergence of the Polak-Ribière-Polyak conjugate gradient algorithm under inexact line search for nonconvex functions, *J. Comput. Appl. Math.* 362 (2019) 262–275.

[16] G. Yuan and M. Zhang, A three-terms Polak-Ribière-Polyak conjugate gradient algorithm for large-scale nonlinear equations, *J. Comput. Appl. Math.* 286 (2015) 186–195.

[17] L. Zhang, A derivative-free conjugate residual method using secant con- dition for general large-scale nonlinear equations, *Numer. Algor.* 83 (2020) 1277–1293.

[18] L. Zhang and H. Tang, A hybrid MBFGS and CBFGS method for nonconvex minimization with a global complexity bound, *Pacific J. Optim.* 14 (2018) 693– 702.

[19] L. Zhang, W. Zhou and D. Li, A descent modified Polak-Ribière-Polyak conjugate gradient method and its global convergence, *IMA J. Numer. Anal.* 26 (2006) 629–640.

[20] W. Zhou, A modified BFGS type quasi-Newton method with line search for symmetric nonlinear equations problems, *J. Comput. Appl. Math.* 367 (2020) 112454.

[21] W. Zhou, A short note on the global convergence of the unmodified PRP method, *Optim. Lett.* 7 (2013) 1367–1372.

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

[23] W. Zhou and D. Li, On the Q-linear convergence rate of a class of methods for monotone nonlinear equations, *Pacific J. Optim.* 14 (2018) 723–737.

[24] W. Zhou and D. Shen, Convergence properties of an iterative method for solving symmetric nonlinear equations, *J. Optim. Theory Appl.* 164 (2015) 277–289.

[25] W. Zhou and L. Zhang, A BFGS method using inexact gradient for general nonlinear equations, to appear in Pacific J. Optim.

Manuscript received 11 September 2019
revised 22 October 2020
accepted for publication 6 November 2020

Weijun Zhou
Department of Mathematics and Statistics
Changsha University of Science and Technology
Changsha 410114, China
E-mail address: mathwjzhou@csust.edu.cn