

# On the convergence properties of the modified Polak–Ribière–Polyak method with the standard Armijo line search

Lijun Li<sup>1</sup>      Weijun Zhou<sup>2</sup>

(Received 21 May 2013; revised 14 April 2014)

## Abstract

Zhang et al. [IMA J. Numer. Anal., 26 (2006) 629–640] proposed a modified Polak–Ribière–Polyak method for non-convex optimization and proved its global convergence with some backtracking type line search. We further study its convergence properties. Under the standard Armijo line search condition, we show that the modified Polak–Ribière–Polyak method has better global convergence property and locally R-linear convergence rate for non-convex minimization. Some preliminary numerical results are also reported to show its efficiency.

*Subject class:* 90C30, 65K05

*Keywords:* conjugate gradients; PRP method; strong convergence; linear convergence

---

<http://journal.austms.org.au/ojs/index.php/ANZIAMJ/article/view/7098>

gives this article, © Austral. Mathematical Soc. 2014. Published April 16, 2014. ISSN 1446-8735. (Print two pages per sheet of paper.) Copies of this article must not be made otherwise available on the internet; instead link directly to this URL for this article.

# Contents

<b>1 Introduction</b>	<b>E80</b>
<b>2 Convergence properties</b>	<b>E82</b>
<b>3 Numerical experiments</b>	<b>E84</b>
<b>4 Conclusions</b>	<b>E88</b>
<b>References</b>	<b>E88</b>

## 1 Introduction

Consider the problem

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

where  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  is a continuously differentiable function and its gradient  $g(x) \triangleq \nabla f(x)$  is available.

Zhang et al. [6] presented the following modified Polak–Ribière–Polyak (MPRP) method for solving the general unconstrained optimization problem (1). We denote  $g_k = \nabla f(x_k)$ , then the scheme of the MPRP method is

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

where  $\alpha_k$  is a step size and the search direction  $d_k$  is generated by the MPRP formula

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

$$\beta_k^{\text{PRP}} = \frac{g_k^\top y_{k-1}}{\|g_{k-1}\|^2}, \quad \theta_k = \frac{g_k^\top d_{k-1}}{\|g_{k-1}\|^2}, \quad y_{k-1} = g_k - g_{k-1}. \quad (4)$$

An important property of the MPRP method is that

$$\mathbf{g}_k^T \mathbf{d}_k = -\|\mathbf{g}_k\|^2, \quad (5)$$

which is independent of line searches used and the convexity of the objective function  $f$ . Moreover, the MPRP method reduces to the standard PRP method [4, 5] when the exact line search is used. Zhang et al. [6] proved that the MPRP method converges globally for non-convex optimization in the sense that

$$\liminf_{k \rightarrow \infty} \|\mathbf{g}_k\| = 0$$

by the use of the backtracking type line search, that is,  $\alpha_k = \max\{\rho^j \mid j = 0, 1, 2, \dots\}$  satisfies

$$f(\mathbf{x}_k + \alpha_k \mathbf{d}_k) \leq f(\mathbf{x}_k) - \delta \alpha_k^2 \|\mathbf{d}_k\|^2, \quad (6)$$

where  $\rho \in (0, 1)$  and  $\delta > 0$  are constants.

Since the MPRP method was proposed, it has attracted much attention. Li and Tian [2] showed  $n$ -step quadratic convergence of the MPRP method with a restart strategy for strongly convex objective functions using the Armijo or Wolfe line search. Zhu, Xiao and Wu [7] applied the MPRP method to solving large sparse signal recovery problems and obtained good numerical results. Li [1] extended the MPRP method to solving non-negatively constrained optimization problems.

We further study the convergence properties of the MPRP method for non-convex minimization with the following standard Armijo line search. Let  $\rho, \delta \in (0, 1)$ ,  $\tau > 0$  and  $\sigma_k = \tau \|\mathbf{g}_k^T \mathbf{d}_k\| / \|\mathbf{d}_k\|^2$ , compute  $\alpha_k = \max\{\sigma_k \rho^j \mid j = 0, 1, 2, \dots\}$  such that

$$f(\mathbf{x}_k + \alpha_k \mathbf{d}_k) \leq f(\mathbf{x}_k) + \delta \alpha_k \mathbf{g}_k^T \mathbf{d}_k. \quad (7)$$

Section 2 establishes the strongly global convergence property of the MPRP method and show its locally  $\mathbf{R}$ -linear convergence rate when the line search (7) is used. Section 3 reports some numerical experiments to show the performance of the MPRP method with the line searches (6) and (7) for several test problems.

## 2 Convergence properties

To begin with, let us define the level set

$$\Omega = \{x \mid f(x) \leq f(x_0)\}. \quad (8)$$

We know that  $x_k \in \Omega$  for all  $k \geq 0$  since  $g_k^T d_k = -\|g_k\|^2$ . We make the following standard assumptions to ensure strongly global convergence of the method.

**Assumption 1.** 1. The level set  $\Omega$  defined by (8) is bounded.

2. In some neighbourhood  $N$  of  $\Omega$ , the gradient is Lipschitz continuous; that is, there exists a constant  $L > 0$  such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \quad \text{for all } x, y \in N. \quad (9)$$

Under Assumption 1, the following result shows that the step size  $\alpha_k$  is bounded below.

**Lemma 2.** Let the sequence  $\{x_k\}$  be generated by the MPRP method with the line search (7). Then there exist two positive constants  $c_1$  and  $c_2$  such that

$$\alpha_k \geq c_1 |g_k^T d_k| / \|d_k\|^2 \geq c_2, \quad \text{for all } k \geq 0. \quad (10)$$

**Proof:** If  $\alpha_k \neq \sigma_k = \tau |g_k^T d_k| / \|d_k\|^2$ , then  $\alpha'_k = \alpha_k / \rho$  does not satisfy the line search (7), namely,

$$f(x_k + \alpha'_k d_k) > f(x_k) + \delta \alpha'_k g_k^T d_k. \quad (11)$$

By the mean value theorem and (9),


$$\begin{aligned} f(x_k + \alpha'_k d_k) - f(x_k) &= \alpha'_k g_k^T d_k + \alpha'_k (g(x_k + \tau_k \alpha'_k d_k) - g_k)^T d_k \\ &\leq \alpha'_k g_k^T d_k + L \alpha_k'^2 \|d_k\|^2, \end{aligned}$$

where  $\tau_k \in (0, 1)$ . This and (11) mean that the first inequality in (10) holds with  $c_1 = \min\{\tau, \rho(1 - \delta)/L\}$ . By the line search (7), we know that  $\alpha_k \leq \sigma_k = \tau |g_k^T d_k| / \|d_k\|^2$ , which together with (5) implies that

$$\alpha_k \|d_k\|^2 \leq \tau |g_k^T d_k| = \tau \|g_k\|^2. \quad (12)$$

Hence, from (3), (4), (9) and (12),


$$\|d_k\| \leq \|g_k\| + \frac{2L \|g_k\| \alpha_{k-1} \|d_{k-1}\|^2}{\|g_{k-1}\|^2} \leq (1 + 2L\tau) \|g_k\|. \quad (13)$$

It follows from (13) and (5) that the second inequality in (10) holds with  $c_2 \triangleq c_1 / (1 + 2L\tau)$ . 

Then we obtain the following strongly global convergence result for the MPRP method.

**Theorem 3.** *Consider the MPRP method with the line search (7):*

$$\lim_{k \rightarrow \infty} \|g_k\| = 0. \quad (14)$$

**Proof:** Since  $\{f(x_k)\}$  is decreasing and bounded from below, from (7) and (5), we know  $\lim_{k \rightarrow \infty} \alpha_k \|g_k\|^2 = 0$ , which together with (10) yields (14). 

Theorem 3 shows that every limit point of the sequence  $\{x_k\}$  is a stationary point of  $f$ . Moreover, if the Hessian matrix at one limit point  $x^*$  is positive definite, which means that  $x^*$  is a strict local optimal solution of the problem (1), then the whole sequence  $\{x_k\}$  converges to  $x^*$  since (5) implies that  $\|x_{k+1} - x_k\| = \alpha_k \|d_k\| \leq \tau \|g_k\|^2 / \|d_k\| \leq \tau \|g_k\| \rightarrow 0$ . Hence, in the local convergence analysis, we assume that the whole sequence  $\{x_k\}$  converges.

The following result shows that the MPRP method with the line search (7) converges  $\mathbf{R}$ -linearly.

**Theorem 4.** *Let  $f$  be a twice continuously differentiable. Consider the MPRP method with the line search (7). Suppose that  $\{\mathbf{x}_k\}$  converges to  $\mathbf{x}^*$ , where  $\mathbf{g}(\mathbf{x}^*) = 0$  and  $\nabla^2 f(\mathbf{x}^*)$  is positive definite. Then there exist constants  $\mathbf{c}_3 > 0$  and  $r \in (0, 1)$  such that*

$$\|\mathbf{x}_k - \mathbf{x}^*\|^2 \leq \mathbf{c}_3 r^k. \quad (15)$$

**Proof:** Since  $\nabla^2 f(\mathbf{x}^*)$  is positive definite, then  $f$  is uniformly convex in some neighbourhood  $\mathbf{N}_1$  of  $\mathbf{x}^*$  and there exist two positive constants  $m$  and  $M$  such that

$$m\|\mathbf{x} - \mathbf{x}^*\|^2 \leq \|\mathbf{g}(\mathbf{x})\|^2 \leq M\|\mathbf{x} - \mathbf{x}^*\|^2, \quad (16)$$


$$m\|\mathbf{x} - \mathbf{x}^*\|^2 \leq f(\mathbf{x}) - f(\mathbf{x}^*) \leq M\|\mathbf{x} - \mathbf{x}^*\|^2. \quad (17)$$

Without loss of generality, we assume that  $\{\mathbf{x}_k\} \subseteq \mathbf{N}_1$ . By (7), (5) and (10),

$$f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k) + \delta \alpha_k \mathbf{g}_k^T \mathbf{d}_k = f(\mathbf{x}_k) - \delta \alpha_k \|\mathbf{g}_k\|^2 \leq f(\mathbf{x}_k) - \delta \mathbf{c}_2 \|\mathbf{g}_k\|^2.$$

This together with (16) and (17) shows that

$$f(\mathbf{x}_{k+1}) - f(\mathbf{x}^*) \leq (1 - \delta m \mathbf{c}_2 / M) [f(\mathbf{x}_k) - f(\mathbf{x}^*)] \leq \dots \leq r^{k+1} [f(\mathbf{x}_0) - f(\mathbf{x}^*)],$$

where  $r = 1 - \delta m \mathbf{c}_2 / M < 1$ . This and (17) yield (15) with constant  $\mathbf{c}_3 = (f(\mathbf{x}_0) - f(\mathbf{x}^*)) / m$ . 

### 3 Numerical experiments

This section reports some numerical experiments on the MPRP method with the line searches (6) and (7), that is,

- MPRP1: the MPRP method with the line search (6). We set parameters  $\rho = 0.12$  and  $\delta = 10^{-3}$ ;

- MPRP2: the MPRP method with the line search (7). We set parameters  $\rho = 0.12$ ,  $\delta = 10^{-3}$  and  $\tau = 3$ .

All codes were written in Matlab 7.4. We stopped the iteration if the total number of iterations exceeds 500 or  $\|g_k\| < 10^{-4}$ . We tested the following three problems [3] with the form  $f(x) = \frac{1}{2} \sum_{i=1}^m f_i^2(x)$ .

1. The extended Rosenbrock function:  $m = n$ ,  $n$  is an even integer and

$$f_{2i-1}(x) = 10(x_{2i} - x_{2i-1}^2), \quad f_{2i}(x) = 1 - x_{2i-1}, \quad i = 1, 2, \dots, n/2.$$

2. The trigonometric function:  $m = n$  and

$$f_i(x) = n - \sum_{j=1}^n \cos x_j + i(1 - \cos x_i) - \sin x_i, \quad i = 1, 2, \dots, n.$$

3. The Broyden tridiagonal function:  $m = n$  and

$$f_i(x) = (3 - 2x_i)x_i - x_{i-1} - 2x_{i+1} + 1, \quad i = 1, 2, \dots, n,$$

where  $x_0 = x_{n+1} = 0$ .

Tables 1–2 lists the numerical results for the MPRP1 and MPRP2 methods on the three test problems with different initial points and sizes, where “P” and “Init” stand for the problem and the initial point, respectively; “Iter” is the total number of iterations; “fcnt” indicates the total number of function evaluations; “Time” is the CPU time in seconds and “—” means that the method failed to find the solution within 500 iterations. In the last row of Table 2, “Best” means the number of test problems that one method wins over the other method on the number of iterations and function evaluations, and the CPU time.

Tables 1–2 show that both methods performed well and the MPRP2 method seems more efficient since it requires less computations on function evaluations for most problems.

Table 1: First group of test results of the MPRP1 and MPRP2 methods, where  $10^{-2}$  means that the initial point  $x_0 = 10^{-2} * \text{ones}(n,1)$ .

P	Init	n	MPRP1				MPRP2			
			Iter	fcnt	Time	$\ g_k\ $	Iter	fcnt	Time	$\ g_k\ $
1	$10^{-2}$	10	78	268	1.4	$9 \times 10^{-5}$	72	223	1.5	$6 \times 10^{-5}$
		50	78	266	1.4	$8 \times 10^{-5}$	62	183	1.3	$4 \times 10^{-5}$
		100	71	243	2.0	$5 \times 10^{-5}$	73	219	1.5	$9 \times 10^{-5}$
		200	95	328	2.0	$3 \times 10^{-5}$	81	232	1.8	$7 \times 10^{-5}$
		500	78	262	3.0	$8 \times 10^{-5}$	86	259	3.2	$4 \times 10^{-5}$
		1000	86	297	8.1	$3 \times 10^{-5}$	87	257	7.6	$1 \times 10^{-5}$
		2000	97	336	28.0	$7 \times 10^{-5}$	78	223	20.3	$6 \times 10^{-5}$
1	$10^{-4}$	10	74	249	1.3	$8 \times 10^{-5}$	71	188	1.6	$8 \times 10^{-5}$
		50	74	248	7.1	$9 \times 10^{-5}$	64	173	1.6	$9 \times 10^{-5}$
		100	74	251	2.8	$5 \times 10^{-5}$	81	223	1.9	$6 \times 10^{-5}$
		200	84	282	3.7	$5 \times 10^{-5}$	83	250	2.1	$6 \times 10^{-5}$
		500	66	225	3.0	$4 \times 10^{-5}$	72	190	3.1	$7 \times 10^{-5}$
		1000	76	255	7.1	$9 \times 10^{-5}$	87	238	9.0	$7 \times 10^{-5}$
		2000	94	316	27.1	$6 \times 10^{-5}$	86	243	23.8	$6 \times 10^{-5}$
2	0.1	10	134	135	1.3	$9 \times 10^{-5}$	57	74	1.4	$8 \times 10^{-5}$
		50	27	32	2.0	$4 \times 10^{-5}$	34	49	1.6	$7 \times 10^{-5}$
		100	14	22	1.6	$9 \times 10^{-5}$	36	58	1.6	$7 \times 10^{-5}$
		200	19	26	1.8	$8 \times 10^{-5}$	44	82	2.5	$6 \times 10^{-5}$
		500	13	23	2.2	$7 \times 10^{-5}$	—	—	—	$9 \times 10^6$
2	1	10	31	37	1.2	$9 \times 10^{-5}$	39	72	1.4	$9 \times 10^{-5}$
		20	42	49	1.4	$9 \times 10^{-5}$	32	67	1.5	$7 \times 10^{-5}$
		50	23	39	1.7	$5 \times 10^{-5}$	40	62	1.8	$8 \times 10^{-5}$
		100	—	—	—	$1 \times 10^5$	91	275	3.1	$8 \times 10^{-5}$
2	10	10	135	147	2.8	$9 \times 10^{-5}$	42	75	1.9	$7 \times 10^{-5}$
		20	61	83	1.5	$9 \times 10^{-5}$	71	110	1.5	$8 \times 10^{-5}$
		50	29	68	1.6	$8 \times 10^{-5}$	48	95	1.7	$7 \times 10^{-5}$
		100	—	—	—	$1 \times 10^5$	117	410	2.5	$9 \times 10^{-5}$



Table 2: Second group of test results of the MPRP1 and MPRP2 methods, where  $10^{-2}$  means that the initial point  $x_0 = 10^{-2} * \text{ones}(n,1)$ .

P	Init	n	MPRP1				MPRP2			
			Iter	fcnt	Time	$\ g_k\ $	Iter	fcnt	Time	$\ g_k\ $
3	-10	10	149	416	1.4	$7 \times 10^{-5}$	84	215	1.4	$9 \times 10^{-5}$
		50	184	519	6.9	$8 \times 10^{-5}$	80	212	1.7	$8 \times 10^{-5}$
		100	114	324	1.7	$9 \times 10^{-5}$	75	200	2.0	$9 \times 10^{-5}$
		200	209	583	1.9	$9 \times 10^{-5}$	67	186	1.7	$8 \times 10^{-5}$
		500	233	656	5.6	$8 \times 10^{-5}$	61	166	2.9	$8 \times 10^{-5}$
		1000	130	369	9.4	$7 \times 10^{-5}$	61	169	6.5	$7 \times 10^{-5}$
		2000	121	347	29.7	$9 \times 10^{-5}$	61	169	29.6	$7 \times 10^{-5}$
3	-1	10	53	152	1.3	$9 \times 10^{-5}$	49	166	1.6	$8 \times 10^{-5}$
		50	100	281	1.5	$7 \times 10^{-5}$	52	172	1.7	$6 \times 10^{-5}$
		100	91	256	1.7	$9 \times 10^{-5}$	52	174	1.7	$9 \times 10^{-5}$
		200	104	295	1.9	$9 \times 10^{-5}$	52	177	1.7	$7 \times 10^{-5}$
		500	99	280	3.2	$9 \times 10^{-5}$	50	169	3.8	$7 \times 10^{-5}$
		1000	146	407	10.6	$8 \times 10^{-5}$	56	194	6.7	$9 \times 10^{-5}$
3	0	10	79	224	2.3	$7 \times 10^{-5}$	62	202	1.6	$7 \times 10^{-5}$
		50	153	428	2.3	$9 \times 10^{-5}$	181	545	2.0	$9 \times 10^{-5}$
		100	423	1168	2.1	$9 \times 10^{-5}$	229	692	1.9	$8 \times 10^{-5}$
		200	393	1088	2.7	$9 \times 10^{-5}$	323	953	2.4	$9 \times 10^{-5}$
		300	—	—	—	$1 \times 10^{-3}$	441	1314	5.3	$7 \times 10^{-5}$
Best			15	11	18		30	34	27	

## 4 Conclusions

We proved that the MPRP method has the strongly global convergence and locally  $R$ -linear convergence rate for non-convex optimization, which improves the existing results on the MPRP method. Numerical results showed that the MPRP method with the standard Armijo line search is efficient.

**Acknowledgments** We thank the editor and the referee whose helpful suggestions led to much improvement of the paper. This work was supported by the NSF (11371073) of China, the Project of the Scientific Research Fund (11C0049 and 12A004) of the Hunan Provincial Education Department, and the NSF (13JJ4062) of Hunan Province.

## References

- [1] C. Li, A conjugate gradient type method for the nonnegative constraints optimization problems, *J. Appl. Math.*, Volume 2013, Article ID 986317, 6 pages. [E81](#)
- [2] D. Li and B. Tian,  $n$ -step quadratic convergence of the MPRP method with a restart strategy, *J. Comput. Appl. Math.*, 235 (2011) 4978–4990. doi:[10.1016/j.cam.2011.04.026](https://doi.org/10.1016/j.cam.2011.04.026) [E81](#)
- [3] J. J. More, B. S. Garbow and K. H. Hillstom, Testing unconstrained optimization software, *ACM Trans. Math. Softw.*, 7 (1981) 17–41. [E85](#)
- [4] E. Polak and G. Ribière, Note sur la convergence de méthodes de directions conjuguées, *Rev. Fr. Inform. Rech. Oper.*, 16 (1969) 35–43. [E81](#)
- [5] B. T. Polyak, The conjugate gradient method in extreme problems, *USSR Comput. Math. Math. Phys.*, 9 (1969) 94–112. [E81](#)

- [6] 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. doi:[10.1093/imanum/dr1016](https://doi.org/10.1093/imanum/dr1016) E80, E81
- [7] H. Zhu, Y. Xiao and S. Wu, Large sparse signal recovery by conjugate gradient algorithm based on smoothing technique, *Comput. Math. Appl.*, 66 (2013) 24–32. E81

## Author addresses

1. **Lijun Li**, Department of Mathematics, Changsha University of Science and Technology, Changsha 410004, China.
2. **Weijun Zhou**, Department of Mathematics, Changsha University of Science and Technology, Changsha 410004, China.  
<mailto:weijunzhou@126.com>