Journal of Mathematical Research & Exposition
Jan., 2010, Vol. 30, No. 1, pp. 141–148
DOI:10.3770/j.issn:1000-341X.2010.01.013
Http://jmre.dlut.edu.cn

Global Convergence of a Modified PRP Conjugate Gradient Method

Hai Dong HUANG*, **Yan Jun LI**, **Zeng Xin WEI**
College of Mathematics and Information Science, Guangxi University, Guangxi 530004*, P. R. China*

Abstract In this paper, a modified formula for $\beta_k^{\rm PRP}$ is proposed for the conjugate gradient method of solving unconstrained optimization problems. The value of $\beta_k^{\rm PRP}$ keeps nonnegative independent of the line search. Under mild conditions, the global convergence of modified PRP method with the strong Wolfe-Powell line search is established. Preliminary numerical results show that the modified method is efficient.

Keywords unconstrained optimization; conjugate gradient method; global convergence.

Document code A
MR(2000) Subject Classification 90C30
Chinese Library Classification O221.1

1. Introduction

The conjugate gradient (CG) methods are greatly efficient methods for unconstrained optimization problems, especially large-scale problems. Consider the following unconstrained optimization problem

$$\min\{f(x)|x \in \Re^n\}, \tag{1.1}$$

where $f : \Re^n \to \Re$ is a continuously differentiable function whose gradient is denoted by g. The conjugate gradient method can be described by the iterative scheme:

$$x_{k+1} = x_k + t_k d_k, \tag{1.2}$$

where the positive step-size t_k is obtained by some line search, and the direction d_k is generated by the rule:

$$d_k = \begin{cases} -g_k, & \text{if } k = 1, \\ -g_k + \beta_k d_{k-1}, & \text{if } k \geq 2, \end{cases} \tag{1.3}$$

where β_k is a scalar, and g_k denotes $g(x_k)$. Some well-known formulas for β_k are given by

$$\beta_k^{\rm FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \quad \text{(see [4])}; \tag{1.4}$$

$$\beta_k^{\rm HS} = \frac{g_k^{\rm T}(g_k - g_{k-1})}{(g_k - g_{k-1})^{\rm T} d_{k-1}} \quad \text{(see [6])}; \tag{1.5}$$

Received September 30, 2007; Accepted April 16, 2008-04-16
Supported by the National Natural Science Foundation of China (Grant No. 10761001).
* Corresponding author
E-mail address: haidongh@126.com (H. D. HUANG); lyj6677@sina.com (Y. J. LI); zxwei@gxu.edu.cn (Z. X. WEI)

$$\beta_k^{\mathrm{PRP}} = \frac{g_k^{\mathrm{T}}(g_k - g_{k-1})}{\|g_{k-1}\|^2} \quad \text{(see [8, 9])}, \tag{1.6}$$

where $\|\cdot\|$ denotes the l_2-norm. The corresponding conjugate gradient methods are abbreviated as FR, HS and PRP methods. Number experience indicates that the HS method is similar to the PRP method; both two methods tend to be more robust and efficient than the FR method.

In the FR method, if a bad direction and a tinny step from x_{k-1} to x_k are generated, the next direction d_k and the next step s_k are also likely to be poor unless a restart along the gradient direction is performed. In spite of such defeat, the FR method with the exact line search is proved to be globally convergent on the general function by Zoutendijk [16]. However, the PRP and HS methods are not globally convergent even for the exact line search since Powell [7] showed that both two methods can cycle infinitely without approaching a solution. Gibert and Nocedal [5] firstly proved that the PRP conjugate gradient method converges globally when the sufficiently decreasing condition and the so-called Property $(*)$ are satisfied. Dai and Yuan [2, 3] gave further study of the convergence of the PRP method when β_k is defined by $\beta_k = \max\{0, \beta_k^{\mathrm{PRP}}\}$. Global convergence studies of related PRP methods are also made by Wei [13–15]. Other modified PRP methods via adding some strong assumptions or using complicated line searches are proved to be globally convergent.

Birgin and Martínez [1] proposed a spectral conjugate gradient method by combing conjugate gradient method and spectral gradient method [10] in the following way:

$$d_k = -\theta_k g_k + \beta_k d_{k-1}, \tag{1.7}$$

where θ_k is a parameter and

$$\beta_k = \frac{(\theta_k y_{k-1} - s_{k-1})^{\mathrm{T}} g_k}{d_{k-1}^{\mathrm{T}} y_{k-1}}. \tag{1.8}$$

The reported numerical results show that the above method performs very well if θ_k is taken to be the spectral gradient:

$$\theta_k = \frac{s_{k-1}^{\mathrm{T}} s_{k-1}}{s_{k-1}^{\mathrm{T}} y_{k-1}}, \tag{1.9}$$

where $s_{k-1} = x_k - x_{k-1}$.

In this paper, a modified formula for β_k^{PRP} is used to calculate the research directions. Under some mild conditions, the global convergence of modified PRP method with the strong Wolfe-Powell line search is established.

The paper is organized as follows. In next section, we will provide the modified PRP method and some basic results for the conjugate gradient methods. In Section 3, we will establish global convergence results with the strong Wolfe-Powell line search. The preliminary numerical results are reported in the last section.

2. The modified PRP method and some basic results

Motivated by Wei [15] and the convergence analysis for the PRP method by Gilbert and

Noceda [5], we propose a modified formula for β_k, which is defined by

$$\beta_k^{new} = \frac{g_k^{\mathrm{T}}(g_k - \frac{g_k^{\mathrm{T}} g_{k-1}}{\|g_{k-1}\|^2} g_{k-1})}{\|g_{k-1}\|^2}. \tag{2.1}$$

It is easy to see that β_k^{new} keeps nonnegative. This property is independent of the line search. Throughout this section, we assume that every research direction d_k satisfies the descent condition

$$g_k^{\mathrm{T}} d_k < 0, \tag{2.2}$$

for all $k \geq 1$.

Now, we state another property which will play important roles in our later analysis. We say that d_k satisfies sufficient descent condition if

$$g_k^{\mathrm{T}} d_k < -c\|g_k\|^2, \tag{2.3}$$

where c is a positive constant.

In order to prove the global convergence, we use the strong Wolfe-Powell line search, which requires t_k to satisfy

$$f(x_k + t_k d_k) - f(x_k) \leq \delta t_k g_k^{\mathrm{T}} d_k, \tag{2.4}$$

and

$$|g(x_k + t_k d_k)^{\mathrm{T}} d_k| \leq \sigma |g_k^{\mathrm{T}} d_k|, \tag{2.5}$$

where $0 < \delta < \sigma < 1$. Another important line search that is often used to prove convergence of nonlinear conjugate gradient methods is strong Wolfe-Powell line search (SWP): Finding t_k to satisfy (2.4) and

$$g(x_k + t_k d_k)^{\mathrm{T}} d_k \geq \sigma g_k^{\mathrm{T}} d_k. \tag{2.6}$$

Now we can state the algorithm of the modified PRP method.

Algorithm 2.1 (A modified PRP method under SWP)

Step 0. Given $x_1 \in \Re^n$, set $d_1 = -g_1, k = 1$. If $g_1 = 0$, then stop.

Step 1. Find a $t_k > 0$ satisfying SWP.

Step 2. Let $x_{k+1} = x_k + t_k d_k$ and $g_{k+1} = g(x_{k+1})$. If $g_{k+1} = 0$, then stop.

Step 3. Compute β_k by the formula (2.1) and generate d_{k+1} by (1.3).

Step 4. Set k:=k+1, go to Step 1.

Generally, the following basic assumptions for the objective function are always used to analyze the global convergence.

Assumption A The level set

$$\Omega = \{x \in R^n \mid f(x) \leq f(x_1)\}$$

is bounded, where x_1 is given by Algorithm 2.1.

Assumption B In some neighborhood N of Ω, f is continuously differentiable, and its gradient is Lipschitz continuously differentiable, that is, there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \tag{2.7}$$

for all $x, y \in \Omega$.

Lemma 2.1 *Suppose that Assumptions A and B hold. Consider the method of the form (1.2)–(1.3), where d_k satisfies (2.2) for all k, and t_k satisfies the Wolfe-Powell line search (2.4) and (2.6). Then,*

$$\sum_{k\geq 1} \frac{(g_k^{\mathrm{T}} d_k)^2}{\|d_k\|^2} < +\infty. \tag{2.8}$$

This result was essentially proved by Zoutendijk [16] and Wolfe [11, 12]. We shall call (2.8) the Zoutendijk condition. The following Lemma is a general and positive result for conjugate gradient methods with the strong Wolfe-Powell line search.

Lemma 2.2 *Suppose that Assumptions A and B hold. Consider any method of the form (1.2)–(1.3), where d_k is a descent direction (2.2) for all k, and t_k satisfies the Wolfe-Powell line search (2.4) and (2.5). Then either*

$$\liminf_{k\to\infty} \|g_k\| = 0, \tag{2.9}$$

or

$$\sum_{k\geq 1} \frac{\|g_k\|^4}{\|d_k\|^2} < +\infty. \tag{2.10}$$

3. Global convergence results

In this section we discuss convergence properties of the MPRP method under strong Wolfe-Powell line search. The following result gives conditions on the line search that guarantee that all search directions are all descent directions.

Lemma 3.1 *Suppose that Assumptions A and B hold. Consider any method of the form (1.2)–(1.3), where β_k satisfies (2.1), and where the step-size t_k satisfies (2.4) and (2.5) with $0 < \sigma < 1/2$. Then, the method generates descent direction d_k satisfying*

$$-\frac{1}{1-\sigma} \leq \frac{g_k^{\mathrm{T}} d_k}{\|g_k\|^2} \leq \frac{2\sigma - 1}{1-\sigma}, \quad k = 1, 2, \ldots\ . \tag{3.1}$$

Proof The proof is by induction. The result clearly holds for $k = 1$ since the middle term equals -1 and $0 < \sigma < 1/2$. Assume that (3.1) holds for some $k \geq 1$. This implies that (2.2) holds, since $\frac{2\sigma-1}{1-\sigma} < 0$, by the condition $0 < \sigma < 1/2$. From (1.3) and (2.1), we have

$$\frac{g_{k+1}^{\mathrm{T}} d_{k+1}}{\|g_{k+1}\|^2} = -1 + \beta_{k+1}^{\mathrm{PRP}*} \frac{(g_{k+1}^{\mathrm{T}} d_k)}{\|g_{k+1}\|^2} = -1 + \frac{g_{k+1}^{\mathrm{T}}(g_{k+1} - \frac{g_{k+1}^{\mathrm{T}} g_k}{\|g_k\|^2} g_k)}{\|g_k\|^2} \frac{(g_{k+1}^{\mathrm{T}} d_k)}{\|g_{k+1}\|^2} \tag{3.2}$$

$$= -1 + \frac{g_{k+1}^{\mathrm{T}} d_k}{\|g_k\|^2} - \frac{(g_{k+1}^{\mathrm{T}} g_k)^2 (g_{k+1}^{\mathrm{T}} d_k)}{(\|g_k\|^2)^2 (\|g_{k+1}\|^2)} \tag{3.3}$$

$$= -1 + \frac{g_{k+1}^{\mathrm{T}} d_k}{\|g_k\|^2} - \cos^2 \vartheta_{k+1} \frac{g_{k+1}^{\mathrm{T}} d_k}{\|g_k\|^2} \tag{3.4}$$

$$= -1 + \sin^2 \vartheta_{k+1} \frac{g_{k+1}^{\mathrm{T}} d_k}{\|g_k\|^2}, \tag{3.5}$$

where ϑ_{k+1} is the angle between g_{k+1} and g_k. Using the line search conditions (2.5), we have

$$-\sigma \mid g_{k+1}^{\mathrm{T}} d_k \mid < g_{k+1}^{\mathrm{T}} d_k < \sigma \mid g_{k+1}^{\mathrm{T}} d_k \mid,$$

which, together with (3.5), gives

$$-1 - \sigma \sin^2 \vartheta_{k+1} \frac{\mid g_k^{\mathrm{T}} d_k \mid}{\|g_k\|^2} \leq \frac{g_{k+1}^{\mathrm{T}} d_{k+1}}{\|g_{k+1}\|^2} \leq -1 + \sigma \sin^2 \vartheta_{k+1} \frac{\mid g_k^{\mathrm{T}} d_k \mid}{\|g_k\|^2}. \tag{3.6}$$

Since $0 \leq \sin^2 \vartheta_{k+1} \leq 1$, we obtain from (2.2),

$$-1 + \sigma \frac{g_k^{\mathrm{T}} d_k}{\|g_k\|^2} \leq \frac{g_{k+1}^{\mathrm{T}} d_{k+1}}{\|g_{k+1}\|^2} \leq -1 - \sigma \frac{g_k^{\mathrm{T}} d_k}{\|g_k\|^2}. \tag{3.7}$$

By the induction hypothesis (3.1), we have

$$-\frac{1}{1-\sigma} \leq \frac{g_{k+1}^{\mathrm{T}} d_{k+1}}{\|g_{k+1}\|^2} \leq \frac{2\sigma - 1}{1-\sigma}. \tag{3.8}$$

We conclude that (3.1) holds for $k+1$. □

Lemma 3.1 achieves the objective: it shows that all search directions are descent directions, and the upper bound in (3.1) shows that the sufficient condition (2.3) holds.

Theorem 3.1 *Suppose that Assumptions A and B hold. Consider the method of the form (1.2)–(1.3), where β_k is obtained by (2.1) and the direction d_k satisfies (2.3) with $c = \frac{1-2\sigma}{1-\sigma}$. Then*

$$\liminf_{k \to +\infty} \|g_k\| = 0. \tag{3.9}$$

Proof Squaring the both sides of the definition of d_k, we obtain

$$\|d_k\|^2 = -\| \|g_k\|^2 - 2\beta_k (g_k^{\mathrm{T}} d_{k-1}) + \beta_k^2 \|d_{k-1}\|^2. \tag{3.10}$$

From (2.1), (3.1) and the above inequality, we have

$$\begin{aligned}
\frac{\|d_k\|^2}{\|g_k\|^4} &= \frac{1}{\|g_k\|^2} - 2\frac{\beta_k^{\text{new}} g_k^{\mathrm{T}} d_{k-1}}{\|g_k\|^4} + (\beta_k^{\text{new}})^2 \frac{\|d_{k-1}\|^2}{\|g_k\|^4} \\
&\leq \frac{1}{\|g_k\|^2} + 2\sigma \frac{\beta_k^{\text{new}} \mid g_{k-1}^{\mathrm{T}} d_{k-1} \mid}{\|g_k\|^4} + (\beta_k^{\text{new}})^2 \frac{\|d_{k-1}\|^2}{\|g_k\|^4} \\
&\leq \frac{1}{\|g_k\|^2} + 2\sigma M (1 - \cos^2 \vartheta_k) \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \frac{\|g_{k-1}\|^2}{\|g_k\|^4} + \Big[(1 - \cos^2 \vartheta_k) \frac{\|g_k\|^2}{\|g_{k-1}\|^2}\Big]^2 \frac{\|d_{k-1}\|^2}{\|g_k\|^4} \\
&\leq \frac{1 + 2\sigma M}{\|g_k\|^2} + \frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^4},
\end{aligned}$$

where $M = \max\{\frac{1}{1-\sigma}, \frac{1-2\sigma}{1-\sigma}\}$.

Denoting $t_k = \frac{\|d_k\|^2}{\|g_k\|^4}$, we have from the above inequality,

$$t_k \leq t_{k-1} + \frac{1 + 2\sigma M}{\|g_k\|^2}.$$

Note that $t_1 = \frac{1}{\|g_1\|^2}$. We obtain

$$t_k \leq (1 + 2\sigma M) \sum_{1}^{k} \frac{1}{\|g_i\|^2}. \tag{3.11}$$

Suppose that (3.9) does not hold, which means that the gradients remain bounded away from zero: there exists $\gamma > 0$ such that

$$\|g_k\| \geq \gamma, \tag{3.12}$$

for all $k \geq 1$. By (3.11) and (3.12), we have

$$t_k \leq \frac{(1 + 2\sigma M)k}{\gamma^2},$$

which implies that t_k at most increases linearly, therefore we have

$$\sum_{k \geq 1} t_k^{-1} = +\infty. \tag{3.13}$$

On the other hand, by (2.8) and (3.1), we obtain

$$\sum_{k \geq 1} \frac{\|g_k\|^4}{\|d_k\|^2} < +\infty, \tag{3.14}$$

which contradicts (3.13), hence (3.9) follows. □

4. Numerical experiments

In this section, numerical results for the modified Polak-Ribière-Polyak (MPRP) and the classical one (PRP) are reported. The problems that we tested are from Neculai Andrei Home Page: http://www.ici.ro/camo/neculai/ansoft.htm. For these tests, the parameters in the strong Wolfe-Powell line search (2.4)–(2.5) were chosen to be $\delta = 10^{-4}$ and $\sigma = 0.1$ and the initial trial value for the line search was set to $1/\|g_1\|$ for the first iteration. For the given test problems, the termination criterion was following:

$$\|g(x_k)\| \leq 10^{-5}.$$

All codes were written in Fortran and run on PC with 1.7 GHZ CPU processor and 256 M RAM memory and Linux operation system. In these runs the methods were implemented without restarting. The PRP code is co-authored by Guanghui Liu, Jorge Nocedal, and Richard Waltz and is obtained from Jorge Nocedal's web page:

http://www.eecs.northwestern.edu/∼nocedal/software.html

Number results are summarized in Table 4.1. In the Table, problem denotes the name of the test problem in Fortran, Dim denotes the number of the variable in the problem, NI denotes the number of iterations, F-G denotes the number of functions and gradient evaluations, CPU time in seconds, and -1 denotes that the line search procedure described above failed to find a step-length.

		PRP	MPRP
Problem	Dim	NI/F-G/CPU	NI/F-G/CPU
Diagonal 2	500	-1	-1
	1000	-1	502/1017 /0.6719E+00
Extended Quadratic Penalty QP1	500	7/28/0.2000E-02	8/26/0.9999E-03
	1000	8/27/0.5999E-02	7/26/0.3000E-02
Extended Powell	1000	212/473/0.6899E-01	238/541/0.6999E-01
	2000	222/500/0.1120E+00	101/237/0.5499E-01
Singuad	1000	309/702/0.3609E+00	126/417/0.2170E+00
	2000	414/967/0.9499E+00	146/486/0.4969E+00
Extended PSC1	1000	7/29/0.1800E-01	7/22/0.1500E-01
	2000	11/37/0.4299E-01	12/33/0.3999E-01
Extended Trigonometric	1000	54 /132/0.1600E+00	46/115/0.1410E+00
	2000	44/123/0.2840E+00	26/75/0.1700E+00
	5000	52/136/0.8089E+00	32/86/0.5109E+00
Extended Penalty	1000	12/58/0.6999E-02	12/53/0.5999E-02
	2000	15/64/0.1200E-01	11/56/0.8999E-02
	5000	15/70/0.4299E-01	14/62/0.3300E-01
Perturbed Quadratic	1000	161/324/0.6099E-01	161/324/0.6001E-01
	2000	228/458/0.1460E+00	228/458/0.1459E+00
	5000	361/725/0.5449E+00	361/725/0.5449E+00
Extended Block-Diagonal BD1	1000	9/84/0.3000E-01	12/73/0.2200E-01
	2000	-1	19/77/0.4399E-01
	5000	7/48/0.7599E-01	18/66/0.1130E+00

Table 4.1 Test results for the PRP method and the MPRP method

From the Table 4.1, we see that for some problem the MPRP method performs much better than the PRP method, for example the Extended Block-Diagonal BD1; whereas for the Extended Powell problem with dim=1000, the MPRP method performs worse than the PRP method. On the whole, the MPRP method performs better than the PRP method for the given test problems.

References

[1] BIRGIN E G, MARTÍNEZ J M. *A spectral conjugate gradient method for unconstrained optimization* [J]. Appl. Math. Optim., 2001, **43**(2): 117–128.

[2] DAI Yuhong, YUAN Yaxiang. *Nonlinear Conjugate Gradient Methods* [M]. Shanghai: Shanghai Scientific and Technical Publishers, 2000.

[3] DAI Yuhong, YUAN Yaxiang. *Further studies on the Polak-Ribière-Polyak method* [R]. Research Report ICM-95-040, Institute of Computational Mathematics and Scientific/ Engineering Computing, Chinese Academy of Sciences, 1995.

[4] FLETCHER R, REEVES C M. *Function minimization by conjugate gradients* [J]. Comput. J., 1964, **7**: 149–154.

[5] GILBERT J C, NOCEDAL J. *Global convergence properties of conjugate gradient methods for optimization* [J]. SIAM J. Optim., 1992, **2**(1): 21–42.

[6] HESTENES M R, STIEFEL E. *Methods of conjugate gradients for solving linear systems* [J]. J. Research Nat. Bur. Standards, 1952, **49**: 409–436.

[7] POWELL M J D. *Nonconvex Minimization Calculations and the Conjugate Gradient Method* [M]. Springer, Berlin, 1984.

[8] POLAK E, RIBIÈRE G. *Note sur la convergence de méthodes de directions conjuguées* [J]. Revue Française d'Informatique et de Recherche Opérationnelle, 1969, (16): 35–43. (in France)

[9] POLYAK B T. *The conjugate gradient method in extreme problems* [J]. USSR Comp. Math. Math. Phys., 1969, **9**: 94–112.

[10] RAYDAN M. *The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem* [J]. SIAM J. Optim., 1997, **7**(1): 26–33.

[11] WOLFE P. *Convergence conditions for ascent methods* [J]. SIAM Rev., 1969, **11**: 226–235.

[12] WOLFE P. *Convergence conditions for ascent methods. II. Some corrections* [J]. SIAM Rev., 1971, **13**: 185–188.

[13] WEI Zengxin, LI Guoyin, QI Liqun. *New nonlinear conjugate gradient formulas for large-scale unconstrained optimization problems* [J]. Appl. Math. Comput., 2006, **179**(2): 407–430.

[14] WEI Zengxin, LI Guoyin, QI Liqun. *Global convergence of the Polak-Ribière-Polyak conjugate gradient method with an Armijo-type inexact line search for nonconvex unconstrained optimization problems* [J]. Math. Comp., 2008, **77**(264): 2173–2193.

[15] WEI Zengxin, YAO Shengwei, LIU Liying. *The convergence properties of some new conjugate gradient methods* [J]. Appl. Math. Comput., 2006, **183**(2): 1341–1350.

[16] ZOUTENDIJK G. *Nonlinear Programming, Computational Methods* [M]. North-Holland, Amsterdam, 1970.