J. Appl. Math. & Informatics Vol. **29***(2011), No. 5 - 6, pp. 1303 - 1309*
Website: http://www.kcam.biz

GLOBAL CONVERGENCE OF A NEW SPECTRAL PRP CONJUGATE GRADIENT METHOD

LIU JINKUI

Abstract. Based on the PRP method, a new spectral PRP conjugate gradient method has been proposed to solve general unconstrained optimization problems which produce sufficient descent search direction at every iteration without any line search. Under the Wolfe line search, we prove the global convergence of the new method for general nonconvex functions. The numerical results show that the new method is efficient for the given test problems.

AMS Mathematics Subject Classification : 65H05, 65F10.
Key words and phrases : Unconstrained optimization, Conjugate gradient method, PRP conjugate gradient method, Wolfe line search, Global convergence.

1. Introduction

The object of this paper is to study the global convergence properties and practical computational performance of a new spectral PRP conjugate gradient method for unconstrained optimization problems with inexact line searches. Consider the unconstrained optimization problem

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

where$f : R^n \to R$is continuously differentiable and its gradient is denoted by g. Conjugate gradient methods for solving (1) are iterative formulas of the form:

$$x_{k+1} = x_k + \alpha_k d_k, \tag{2}$$

$$d_k = \begin{cases} -g_k, & for\ k = 1, \\ -g_k + \beta_k d_{k-1}, & for\ k \geq 2. \end{cases} \tag{3}$$

where $g_k = \bigtriangledown f(x_k)$, $\alpha_k > 0$ is a step length which is determined by some line search,d_k is the search direction and β_k is a scalar.In a standard PRP conjugate

Received March 28, 2011. Revised April 20, 2011. Accepted April 26, 2011.

gradient method, the scalar $\beta_k$is determined by

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

where $\|\cdot\|$ is the Euclidean norm. In the past two decades, the convergence property of the famous PRP method has been intensively studied by many researchers (e.g.,[1-6]).

In practical computation, the famous PRP method is generally believed to be the most efficient conjugate gradient method, and has got meticulous in recent years. One remarkable property of the method is that it essentially performs a restart if a bad direction occurs (see [7]). But, Powell [8] constructed an example showed that the method can cycle infinitely without approaching any stationary point even if an exact line search is used. This counter-example also indicates that the method has a drawback that they may not globally be convergent when the objective function is non-convex. Therefore, during the past few years, much effort has been investigated to create new formula for β_k and research direction for d_k, which not only possess global convergence for general functions but are also superior to original method from the computation point of view. For example, Bergin and Martine [9] proposed another kind of conjugate gradient method, called spectral conjugate gradient method. Let $y_{k-1} = g_k - g_{k-1}$and $s_{k-1} = x_k - x_{k-1}$. Then, the search direction in this method was defined by

$$d_k = -\theta_k g_k + \beta_k d_{k-1} \tag{5}$$

where $\beta_k = \frac{(\theta_k y_{k-1} - s_{k-1})^T g_k}{d_{k-1}^T y_{k-1}}$ and $\theta_k = \frac{s_{k-1}^T s_{k-1}}{s_{k-1}^T y_{k-1}}$. But, in the above method, the direction must not always be a descent direction. Zhong Wan, ZhanLu Yang, YaLin Wang[11] proposed a spectral PRP conjugate gradient method (called SPRP method). The search direction d_k in this method was defined by (5), where β_k is computed by (4), and $\theta_k = \frac{d_{k-1}^T y_{k-1}}{\|g_{k-1}\|^2} - \frac{d_{k-1}^T g_k g_k^T g_{k-1}}{\|g_k\|^2 \cdot \|g_{k-1}\|^2}$. Under the modified Wolfe-type line search, the authors proved the global convergence of the above spectral conjugate gradient method.

Based on the above methods and the famous PRP method, we develop a new PRP spectral conjugate gradient method (called MSPRP method). The direction is given by the following way

$$d_k = \begin{cases} -g_k, & for k = 1, \\ -\theta_k g_k + \beta_k d_{k-1}, & for k \geq 2. \end{cases} \tag{6}$$

where β_k is specified by (4) and

$$\theta_k = 1 + \beta_k^{PRP} \cdot \frac{g_k^T d_{k-1}}{\|g_k\|^2} \tag{7}$$

And under the Wolfe line search, we are going to establish the global convergence of the new PRP spectral conjugate gradient method.

This paper is organized as follows. In section 2, we propose our algorithm.

In section 3, global convergence analysis is provided with suitable conditions. Preliminary numerical results are presented in section 4.

2. New Algorithm and sufficient descent property

In this section, we give the specific form of the developed spectral conjugate gradient method as follows.
Algorithm 2.1(New Spectral PRP Conjugate Gradient Method):
Step 1: Data: $x_1 \in R^n$, $\varepsilon \geq 0$. Set $d_1 = -g_1$, if $||g_1|| \leq \varepsilon$, then stop.
Step 2: Compute α_k by the Wolfe line search ($0 < \delta < \sigma < 1$):

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k, \tag{8}$$

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k, \tag{9}$$

Step 3: Let $x_{k+1} = x_k + \alpha_k d_k$, $g_{k+1} = g(x_{k+1})$, if $||x_{k+1} - x_k|| \leq \varepsilon$, then stop.
Step 4: Compute β_{k+1} by (4), and generate d_{k+1} by (6).
Step 5: Set $k = k + 1$, go to step 2.
$d_k$is said sufficient descent if it satisfies

$$g_k^T d_k \leq -c\|g_k\|^2, c > 0, \tag{10}$$

where c is a positive constant. (10) plays a vital role in guaranteeing the global convergence of conjugate gradient methods. In most references, we can see that (10) was always ensured. Now we have the following theorem, which illustrate that the new spectral conjugate gradient method can guarantee (10) hold without any line search.

Theorem 2.1. *Suppose that d_k is given by (2),(6) and (7). Then for all $k \geq 1$*

$$g_k^T d_k = -\|g_k\|^2. \tag{11}$$

Proof. Firstly, for $k = 1$, it is easy to see that (11) holds since $d_1 = -g_1$. Secondly, multiplying (6) by g_k^T, we get

$$g_k^T d_k = -\theta_k \|g_k\|^2 + \beta_k^{PRP} \cdot g_k^T d_{k-1}. \tag{12}$$

Since (7), we get

$$g_k^T d_k = -(1 + \beta_k^{PRP} \cdot \frac{g_k^T d_{k-1}}{\|g_k\|^2}) \cdot \|g_k\|^2 + \beta_k^{PRP} g_k^T d_{k-1} = -\|g_k\|^2.$$

Therefore, the (11) holds. □

3. Global Convergence

In order to establish the global convergence of the proposed method, we need the following assumption on objective function, which have been used often in the literature to analyze the global convergence of nonlinear conjugate gradient methods with inexact line searches.
Assumption (3.1) :
(i) The level set $\Omega = \{x | f(x) \leq f(x_1)\}$ is bounded, where x_1 is the starting point.

(ii) In some neighborhood N of Ω, the objective function is continuously differentiable, and its gradient is Lipschitz continuous, i.e., there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \ for \ \forall \ x, y \in N.$$

In addition, we need another assumption for our paper in the following.
Assumption (3.2) : For k large enough, the inequalities

$$0 < g_k^T g_{k-1} \leq 2 g_k^T g_k. \tag{13}$$

hold.
This assumption (13) is reasonable (see [11]).
The conclusion of the following lemma, often called the Zoutendijk condition, is used to prove the global convergence of nonlinear conjugate gradient methods. It was originally given by Zoutendijk [10].

Lemma 3.1. *Suppose Assumption (3.1) holds. Consider any iteration of (2) and (6), where* d_k *satisfies* $g_k^T d_k < 0$ *for* $k \in N$ *and* α_k *satisfies the Wolfe line search. Then*

$$\sum_{k \geq 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty. \tag{14}$$

Remark 3.1. From (11) and (14), we have

$$\sum_{k \geq 1} \frac{\|g^T\|^4}{\|d_k\|^2} < +\infty. \tag{15}$$

Theorem 3.2. *Suppose that Assumption (3.1-3.2) holds. Let* $\{x_k\}$ *and* $\{d_k\}$ *be generated by (2) and (6), in which* α_k *satisfies the Wolfe line searches. Then*

$$\liminf_{k \to +\infty} \|g_k\| = 0. \tag{16}$$

Proof. To obtain this result, we proceed by contradiction. Suppose that (16) does not hold, which means that there exists $r > 0$ such that

$$\|g_k\| \geq r, fork \geq 1. \tag{17}$$

From (6), we have

$$\begin{aligned} \|d_k\|^2 &= (-\theta_k g_k + \beta_k^{PRP} d_{k-1})^T (-\theta_k g_k + \beta_k^{PRP} d_{k-1}) \\ &= \theta_k^2 \|g_k\|^2 - 2\theta_k \beta_k^{PRP} g_k^T d_{k-1} + (\beta_k^{PRP})^2 \|d_{k-1}\|^2. \end{aligned} \tag{18}$$

From (6), we have

$$\beta_k^{PRP} d_{k-1} = d_k + \theta_k g_k.$$

Then, we get

$$\begin{aligned} \|d_k\|^2 &= \theta_k^2 \|g_k\|^2 - 2\theta_k g_k^T (d_k + \theta_k g_k) + (\beta_k^{PRP})^2 \|d_{k-1}\|^2 \\ &= -\theta_k^2 \|g_k\|^2 - 2\theta_k g_k^T d_k + (\beta_k^{PRP})^2 \|d_{k-1}\|^2. \end{aligned}$$

Dividing by $\|g_k\|^4$ in the both sides of above equality, then from (4), (11), (13) and (17), we have

$$\frac{\|d_k\|^2}{\|g_k\|^4} = \frac{-\theta_k^2\|g_k\|^2 - 2\theta_k g_k^T d_k + (\beta_k^{PRP})^2\|d_{k-1}\|^2}{\|g_k\|^4}$$

$$= \frac{\|d_{k-1}\|^2}{\|g_k\|^4}\cdot(\beta_k^{PRP})^2 - \frac{\theta_k^2\|g_k\|^2 + 2\theta_k g_k^T d_k}{\|g_k\|^4} = \frac{\|d_{k-1}\|^2}{\|g_k\|^4}\cdot(\beta_k^{PRP})^2 - \frac{\theta_k^2\|g_k\|^2 - 2\theta_k\|g_k^T\|^2}{\|g_k\|^4}$$

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

$$= \frac{\|d_{k-1}\|^2}{\|g_k\|^4}\cdot(\frac{\|g_k\|^4}{\|g_{k-1}\|^4} - \frac{(2\|g_k\|^2 - g_k^T g_{k-1})g_k^T g_{k-1}}{\|g_{k-1}\|^4}) - \frac{(\theta_k-1)^2-1}{\|g_k\|^2}$$

$$= \frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^4} - \frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^4}\cdot\frac{(2\|g_k\|^2 - g_k^T g_{k-1})g_k^T g_{k-1}}{\|g_k\|^4} - \frac{(\theta_k-1)^2}{\|g_k\|^2} + \frac{1}{\|g_k\|^2}$$

$$\leq \frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^4} + \frac{1}{\|g_k\|^2} \leq \sum_{i=1}^{k}\frac{1}{\|g_i\|^2} < \frac{k}{r^2}. \quad (19)$$

From (19), we know

$$\sum_{k\geqslant 1}\frac{\|g_k\|^4}{\|d_k\|^2} \geq r^2\sum_{k\geq 1}\frac{1}{k} = +\infty,$$

which contradicts (16), so $\liminf_{k\to+\infty}\|g_k\| = 0$. □

4. Numerical results

In this section, we report some numerical experiments. We test the Algorithm3.1 by solving the ten problems from [12], and compare its performance to that of the PRP method and the SPRP method under the Wolfe line search conditions. All codes were written in Mat lab 7.0 and run on a PC with 2.0GHz CPU processor and 512 MB memory and Windows XP operation system.
In Algorithm 3.1, we have $\delta = 0.01$ and $\sigma = 0.1$, and the termination condition is $\|g_k\| \leq 1.0 \times 10_{-6}$, or It-max ¿9999. It-max denotes the maximal number of iterations.

The numerical results of our tests are reported in the Table 1. The first column "Problem " represents the name of the tested problem. Dim denotes the dimension of the test problems. NI, NF, NG, CPU denote the number of iterations, function evaluations, gradient evaluations and the CPU time in seconds, respectively. From the numerical results in Table1, it is shown that the proposed spectral conjugate gradient method is promising.

Table 1: The performances of the PRP, SPRP and MSPRP methods

Problem	Dim	Algorithm	NI	NF	NG	CUP
FROTH	2	PRP	12	78	59	0.0698
	2	SPRP	9	56	40	0.0276
	2	MSPRP	9	56	40	0.0301
BADSCP	2	PRP	99	394	336	0.4000
	2	SPRP	110	419	361	1.0000
	2	MSPRP	109	411	352	0.4000
BADSCB	2	PRP	13	30	19	0.0448
	2	SPRP	13	39	26	0.1220
	2	MSPRP	13	39	26	0.1241
BEALE	2	PRP	12	48	35	0.0402
	2	SPRP	13	43	30	0.0394
	2	MSPRP	13	43	30	0.1239
BRAD	3	PRP	30	100	80	0.2027
	3	SPRP	14	44	34	0.0430
	3	MSPRP	14	44	34	0.0441
KOWOSB	4	PRP	71	234	203	0.3000
	4	SPRP	65	205	181	0.2000
	4	MSPRP	65	205	181	0.2000
WATSON	5	PRP	193	535	473	1.4000
	5	SPRP	123	370	320	0.4000
	5	MSPRP	127	358	309	0.5000
PEN1	50	PRP	64	344	280	0.5
	50	SPRP	71	345	286	0.4000
	50	MSPRP	71	343	285	0.3000
PEN1	100	PRP	23	160	122	0.2212
	100	SPRP	27	170	131	0.2272
	100	MSPRP	27	170	131	0.3187
TRIG	50	PRP	45	93	85	0.3426
	50	SPRP	47	97	92	0.5114
	50	MSPRP	47	97	92	0.5109
TRIG	100	PRP	58	120	113	0.4573
	100	SPRP	54	117	108	0.5788
	100	MSPRP	54	117	108	0.5954
TRIG	200	PRP	64	135	128	2.0248
	200	SPRP	62	136	131	1.9746
	200	MSPRP	62	136	131	1.9860
TRID	500	PRP	35	78	74	0.5188
	500	SPRP	35	78	73	0.4961
	500	MSPRP	35	78	73	0.4966
TRID	1000	PRP	34	76	72	1.6832
	1000	SPRP	34	76	72	1.6623
	1000	MSPRP	34	76	72	1.6431

References

1. E. Polak, G. Ribire, *Note sur la xonvergence de directions conjugees*, Rev Francaise informat Recherche Operatinelle 3e Annee 16(1969) 35-43.
2. B.T.Polak, *The conjugate gradient method in extreme problems*, MUSSR Comput. Math. Math. Phys. 9(1969) 94-112.
3. J. C. Gilbert and J. Nocedal, *Global convergence properties of conjugate gradient methods for optimization*, SIAM Journal on Optimization, 2(1992) 21-42.
4. L. Grippo and S. Lucidi, *A globally convergent version of the Polak-Ribire conjugate gradient method*, Mathematical Programming, 78(1997) 375-391.
5. L. Grippo and S. Lucidi, *Convergence conditions, line search algorithms and trust region implementations for the Polak-Ribire conjugate gradient method*, Optimization Methods and Software, 20(2005) 71-98.
6. L.Zhang, W.J.Zhou, D.H.Li, *A descent modified Polak-Ribire-Polyak conjugate gradient method and its global convergence*, IMA J. Numer. Anal. 26(2006)629-640.
7. W. W. Hager and H. Zhang, *A survey of nonlinear conjugate gradient methods*, Pacific Journal of Optimization, 2(2006) 35-58.
8. M. J. D. Powell, *Nonconvex minimization calculations and the conjugate gradient method*, in Numerical Analysis (Dundee, 1983), vol. 1066 of Lecture Notes in Mathematics, pp. 122-141, Springer, Berlin, Germany, 1984.
9. E.Birgin, J.M.Martinez, *A spectral conjugate gradient method for unconstrained optimization*, Appl. Math. Optim. 43(2001)117-128.
10. G.Zoutendijk, *Nonlinear Programming, Computational Methods*, In: J.Abadie(eds0, Integer and Nonlinear Programming, North-Holland, 1970, pp. 37-86.
11. Zhong Wan, ZhanLu Yang, YaLin Wang, *New spectral PRP conjugate gradient method for constrained optimization*, Applied Mathematics Letters, 24(2011)16-22.
12. J.J.More, B.S.Garbow, K.E.Hillstrome, *Testing unconstrained optimization software*, ACM Trains. Math. Software 7(1981) 17-41.

Liu Jinkui received Master's degree at Chongqing University. Since 2009, he has been at Chongqing Three Gorges University. His research interests include optimization and numerical computation.

College of Mathematics and Statistics, Chongqing Three Gorges University, Wanzhou, Chongqing 400040, China.
e-mail: liujinkui2006@126.com