International Journal of Applied Mathematics and Statistics,
Int. J. Appl. Math. Stat.; Vol. 31; Issue No. 1; Year 2013,
ISSN 0973-1377 (Print), ISSN 0973-7545 (Online)

Analysis on an Improved Global Convergence for a Spectral Conjugate Gradient Method

Songhai Deng[1], **Xiaohong Chen**[2] **and Zhong Wan**[1]

[1]School of Mathematics and Statistics
Central South University
410083, Hunan Changsha, China
Correspondence author: Zhong Wan, wanmath@163.com

[2]School of Bussiness
Central South University
410083, Hunan Changsha, China
csu_cxh@163.com

ABSTRACT

In this paper, we establish the theory of global convergence for a spectral conjugate gradient algorithm recently developed by Z. Wan etc. An assumption, that the inequalities $0 < g_k^T g_{k-1} \leq 2\|g_k\|^2$ *are satisfied for any* k*, is first investigated by numerical experiments. It is shown that such assumption holds only for* k *large enough in solving some benchmark problems, not for all ones. Another contribution of this paper is to obtain the same convergence result under some weaker assumptions.*

Keywords: unconstrained optimization, conjugate gradient, global convergence

2010 Mathematics Subject Classification: 90C25, 90C30

1 Introduction

Consider the following unconstrained optimization problem:

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

where $f : R^n \to R$ is continuously differentiable such that its gradient is available. Let $g(x)$ denote the gradient of f at x.

Due to the features of low memory requirement and simple computation, the conjugate gradient method is popular to be employed for solving problem (1.1). By this method, starting from an initial point $x_0 \in R^n$, a sequence of solutions x_k, $k \geq 1$, is generated by

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

where α_k is a stepsize obtained by some line search rule and d_k is a search direction given by

$$d_k = \begin{cases} -g_k, & \text{if } \ k = 0, \\ -g_k + \beta_k d_{k-1}, & \text{if } \ k > 0. \end{cases} \tag{1.2}$$

In (1.2), β_k is a fundamental algorithmic parameter. With different value of β_k, the corresponding algorithm has distinct numerical behavior (see, for example, [5], [8], [10], [11], [12], [16], [22] and the references therein). One of the most popular choices is the formula given by the Polak-Ribière-Polyak conjugate gradient method, it reads

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

Very recently, in [15], a new spectral PRP conjugate gradient algorithm is developed. The search direction d_k is determined by

$$d_k = \begin{cases} -g_k, & \text{if } k = 0, \\ -\theta_k g_k + \beta_k^{PRP} d_{k-1}, & \text{if } k > 0, \end{cases} \tag{1.4}$$

where

$$\theta_k = \frac{d_{k-1}^T y_{k-1}}{\|g_{k-1}\|^2} - \frac{d_{k-1}^T g_k g_k^T g_{k-1}}{\|g_k\|^2 \|g_{k-1}\|^2}. \tag{1.5}$$

The stepsize in the algorithm is obtained by the following modified Wolf-type line search rule:

$$\begin{cases} f(x_k) - f(x_k + \alpha_k d_k) \geq \rho \alpha_k^2 \|d_k\|^2, \\ g(x_k + \alpha_k d_k)^T d_k \geq -2\sigma\alpha_k \|d_k\|^2. \end{cases} \tag{1.6}$$

Thus, the framework of the algorithm is described as follows.

Algorithm 1. **(Modified Spectral PRP Conjugate Gradient Algorithm)**

Step 0 Given constants $0 < \rho < \sigma < 1, \epsilon > 0$. Choose an initial point $x_0 \in R^n$. Set $k := 0$.

Step 1 If $\|g_k\| \leq \epsilon$, then the algorithm stops. Otherwise, compute d_k by (1.4) and (1.5), and go to Step 2.

Step 2 Determine a steplength $\alpha_k > 0$ such that (1.6) is satisfied.

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

In [15], it has been demonstrated by numerical experiment that Algorithm 1 is powerful to solve the benchmark test problems from [9]. However, in [15], the establishment of global convergence needs the following assumption

$$0 < g_k^T g_{k-1} \leq 2 g_k^T g_k. \tag{1.7}$$

To observe whether the inequalities (1.7) hold or not, we implement Algorithm 1 to solve some test problems from [9], with choices of $\rho = 0.5$, $\sigma = 0.75$ and $\epsilon = 10^{-7}$. In Table 1, the results are reported, which indicate that (1.7) must not hold in some cases.

Table 1: Results of testing assumption

Function	x_0	total times	$0 < g_k^T g_{k-1} \leq 2g_k^T g_k$
Rosenbrock	(-1.2,1)	90	None
Helical valley	(-1,0,0)	103	None
Box three-dimensional	(0,10,20)	6867	$k \geq 6866$
Biggs EXP6	(1,2,1,1,1,1)	57560	$k \geq 57552$

The numerical experiments show that the probability that (1.7) is satisfied reduces in the case that the number of iterations is small, but for a large number of iterations, it is often that the inequalities (1.7) hold.

Motivated by the above observation, we are going to establish the global convergence of Algorithm 1 without Assumption (1.7), but the same convergence result as that in [15] is to be obtained.
This paper is organized as follows. In the next section, the main results will be presented. Section 3 will be devoted to prove the main results. Some final remarks are given in the last section.

2 Main results

Assumption 1. The level set $\Omega = \{x \in R^n \mid f(x) \le f(x_0)\}$ is bounded.

Assumption 2. In some neighborhood N of Ω, f is continuously differentiable and its gradient is Lipschitz continuous, namely, there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \le L\|x - y\|, \quad \forall x, y \in N. \tag{2.1}$$

Lemma 2.1. *Suppose that d_k is given by (1.3)-(1.5). Then, the following result*

$$g_k^T d_k = -\|g_k\|^2 \tag{2.2}$$

holds for any $k \ge 0$.

Lemma 2.1 shows that the direction d_k in Algorithm 1 is always sufficiently descent at each iteration.

Lemma 2.2. *With Assumption 2, there exists $\alpha_k > 0$ satisfying (1.6).*

In the following lemma, it is stated that the Zoutendijk condition holds, which is useful to prove the global convergence of conjugate gradient method (see [24] and [18]).

Lemma 2.3. *Under Assumptions 1 and 2, it holds that*

$$\sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty.$$

The following property of the algorithm plays an important role in the proof of the main theorem.

Lemma 2.4. *Let $\{ \alpha_k \}$ and $\{ d_k \}$ be the step size and the search direction sequences generated by Algorithm 1, respectively. Then,*

$$\lim_{k\to\infty} \alpha_k^2 \|d_k\|^2 = 0.$$

With Lemmas 2.1, 2.2, 2.3 and 2.4, we can prove the global convergence.

Theorem 2.5. *Let $\{ g_k \}$ be the gradient sequence generated by Algorithm 1. Under Assumptions 1 and 2, the following result*

$$\lim_{k\to\infty} inf \|g_k\| = 0. \tag{2.3}$$

holds.

3 Proof of main results

In this section, we present the proofs for the main results in this paper.
Since Lemmas 2.1, 2.2 and 2.3 are directly from the corresponding conclusions in [15] under the same conditions, we here needs to prove Lemma 2.4 and Theorem 2.5.
The proof of Lemma 2.4:
From the line search strategy (1.6), we know

$$f(x_k) - f(x_{k+1}) \geq \rho\alpha_k^2\|d_k\|^2.$$

Thus, it is obtained that

$$\sum_{k=1}^{n}(f(x_k) - f(x_{k+1})) \geq \rho\sum_{k=1}^{n}\alpha_k^2\|d_k\|^2.$$

Since f is bounded in the level set Ω, there exists a positive constant scalar M such that

$$2M \geq f(x_1) - f(x_{n+1}) \geq \rho\sum_{k=1}^{n}\alpha_k^2\|d_k\|^2.$$

It follows that the series

$$\sum_{k=1}^{\infty}\alpha_k^2\|d_k\|^2$$

is convergent. Thus,

$$\lim_{k\to\infty}\alpha_k^2\|d_k\|^2 = 0. \tag{3.1}$$

The desired result is proved. □
The proof of Theorem 1:
If the result (2.3) does not hold, then there exists a positive $\epsilon > 0$ such that for all k, $\|g_k\| \geq \epsilon$.
From (1.4), it follows that

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

Dividing by $\|g_k\|^4$ on the both sides of this equality, we obtain

$$\begin{aligned}
\frac{\|d_k\|^2}{\|g_k\|^4} &= \frac{(\beta_k^{PRP})^2\|d_{k-1}\|^2 - 2\theta_k d_k^T g_k - \theta_k^2\|g_k\|^2}{\|g_k\|^4} \\
&= \frac{(g_k^T(g_k - g_{k-1}))^2}{\|g_{k-1}\|^4}\frac{\|d_{k-1}\|^2}{\|g_k\|^4} - \frac{(\theta_k - 1)^2}{\|g_k\|^2} + \frac{1}{\|g_k\|^2} \\
&= \frac{(g_k^T(g_k - g_{k-1}))^2}{\|g_k\|^4}\frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^4} - \frac{(\theta_k - 1)^2}{\|g_k\|^2} + \frac{1}{\|g_k\|^2} \\
&\le \frac{\|g_k - g_{k-1}\|^2}{\|g_k\|^2}\frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^4} + \frac{1}{\|g_k\|^2} \\
&\le \frac{L^2\alpha_{k-1}^2\|d_{k-1}\|^2}{\epsilon^2}\frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^4} + \frac{1}{\|g_k\|^2}.
\end{aligned}$$

From Lemma 2.4, it is clear that there exists an integer number $k_0 \ge 0$ such that for $k \ge k_0$, we have

$$0 \le \alpha_{k-1}^2\|d_{k-1}\|^2 < \frac{\epsilon^2}{L^2}.$$

Thus,

$$\begin{aligned}
\frac{\|d_k\|^2}{\|g_k\|^4} &< \frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^4} + \frac{1}{\|g_k\|^2} \\
&\le \ldots \le \frac{\|d_{k_0}\|^2}{\|g_{k_0}\|^4} + \sum_{i=k_0+1}^{k}\frac{1}{\|g_i\|^2} \\
&\le \frac{C_0}{\epsilon^2} + \frac{k-k_0}{\epsilon^2} \\
&= \frac{C_0 + (k-k_0)}{\epsilon^2},
\end{aligned}$$

where

$$C_0 = \epsilon^2\frac{\|d_{k_0}\|^2}{\|g_{k_0}\|^4} > 0$$

is a constant. Therefore, it is obtained that

$$\sum_{k\ge 1}\frac{\|g_k\|^4}{\|d_k\|^2} \ge \sum_{k>k_0}\frac{\|g_k\|^4}{\|d_k\|^2} > \epsilon^2\sum_{k>k_0}\frac{1}{C_0 + (k-k_0)} = \infty,$$

which contradicts the result of Lemma 2.4. This shows that the desired conclusion holds. □

4 Final Remarks

In this paper, we have established the theory of global convergence for a class of spectral conjugate gradient methods. Different from the existing results available in the literature, we have obtained the same conclusion under more weaker assumption conditions.

Acknowledgment

This work is supported by National Natural Science Fund of China (grant number: 71071162, 70921001).

References

[1] N. Andrei, *Acceleration of conjugate gradient algorithms for unconstrained optimization*, Appl. Math. Comput., **213**(2009), no. 3, 361–369.

[2] N. Andrei, *Open problems in nonlinear conjugate gradient algorithms for unconstrained optimization*, Bulletin of the Malaysian Mathematical Sciences Society, **34**(2011), no. 2, 319–330.

[3] E. Birgin, J.M. Martínez, *A spectral conjugate gradient method for unconstrained optimiztion*. Appl. Math. Optim. **43**(2001), 117–128.

[4] S. Q. Du, Y. Y. Chen, *Global convergence of a modified spectral FR conjugate gradient method*, App. Math. and Comput. **202**(2008), 766–770.

[5] J. C. Gilbert, J. Nocedal, *Global convergence properties of conjugate gradient methods for optimization*, SIAM J. Optim. **2**(1992), no. 1, 21–42.

[6] W.W. Hager, H. Zhang, *A survey of nonlinear conjugate gradient methods*, Pacific J. Optim. **2**(2006), no. 1, 35–58.

[7] J. Nocedal, J. S. Wright, *Numerical Optimization*, Springer-Verlag, New York. 1999.

[8] E. Polak, G. Ribiére, *Note sur la convergence de directions conjuguées*, Rev. Francaise Infomat Recherche Operatonelle, 3e Année. **16**(1969), 35–43.

[9] J. J. Moré , B.S. Garbow and K.E. Hillstrome, *Testing unconstrained optimization software*, ACM Trans. Math. Software, **7**(1981), 17–41.

[10] B. T. Polyak, *The conjugate gradient method in extreme problems*, USSR Comput. Math. Math. Phys. **9**(1969), no. 1, 94–112.

[11] M. J. D. Powell, *Restart procedures of the conjugate gradient method*, Math. Program. **2**(1977), 241–254.

[12] M. J. D. Powell, *Nonconvex minimization calculations and the conjugate gradient method*, Lecture Notes in Mathematics, vol. 1066, Springer, Berlin, 1984, pp. 122–141.

[13] Z. J. Shi, *Restricted PR conjugate gradient method and its global convergence*, Adv. Math. **31**(2002), no. 1, 47–55.

[14] Z. Wan, C.M. Hu and Z.L. Yang, *A spectral PRP conjugate gradient methods for nonconvex optimization problem based on modified line search*, Discrete and Continuous Dynamical Systems: Series B, **16(4)**(2011), no. 1, 1157-1169.

[15] Z. Wan, Z. L. Yang and Y. L. Wang, *New spectral PRP conjugate gradient method for unconstrained optimization*, Appl. Math. Letter, **24**(2011), no. 1, 16–22.

[16] C. Wang, Y. Chen, S. Du, *Further insight into the Shamanskii modification of Newton method*, Appl. Math. Comput. **180**(2006), 46–52.

[17] Z. Wei, G. Li, L. Qi, *Global convergence of the Polak-Ribiere-Polyak conjugate gradient method with an Armijo-type inexact line search for nonconvex unconstrained optimization problems*, Math. Comput., **77**(2008), 2173–2193.

[18] P. Wolfe, *Convergence conditions for ascent methods*, SIAM Rev., **11**(1969), 226–235.

[19] G.H. Yu, L.T. Guan and Z.X. Wei, *Globally convergent Polak-Ribiére-Polyak conjugate gradient methods under a modified Wolfe line search*, Appl. Math. Comput., **215**(2009), no. 8, 3082–3090.

[20] G.Yuan, *Modified nonlinear conjugate gradient methods with sufficient descent property for large-scale optimization problems*, Optimization Letters, **3**(2009), 11–21.

[21] G. Yuan, X.Lu and Z. Wei, *A conjugate gradient method with descent direction for unconstrained optimization*, J. Comput. Appl. Math., **233**(2009), 519–530.

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

[23] L. Zhang, W. Zhou, D. Li., *Global convergence of a modified Fletcher-Reeves conjugate gradient method with Armijo-type line search*, Numer. Math. **104**(2006), 561–572.

[24] G. Zoutendijk, *Nonlinear Programming, Computational methods, Integer and Nonlinear Programming.* J.Abadie(ed.), North-Holland, Amsterdam(1970), 37–86.