The Global Convergence of a New Spectral Conjugate Gradient Method

W.Q. Zeng
The School of Computer Science of Guangdong Polytechnic Normal University
Guangzhou, Guangdong, China
zwq6381508@163.com

H.L. Liu
The School of Computer Science of Guangdong Polytechnic Normal University
Guangzhou, Guangdong, China
lklhl@sina.com

Abstract—**In this pager, we propose a new spectral Conjugate Gradient method to solve unconstrained optimization problems. It has the following properties: (1) the new method satisfies the sufficient descent condition with any line searches condition; (2) this method possesses inherent properties when $\theta_k \geq 0$ (3) under the strong Wolfe line search the method is globally convergent. Preliminary numerical results show that method is very efficient.**

Keywords-sufficient decent; conjugate gradient method; line search; global convergence

I. INTRODUCTION

For solving the nonlinear unconstrained optimization problem

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

Where $f: R^n \to R$ is a continuously differentiable and its gradient is available, bound from below. Generally, we use the iterative method to solve (1), and the iterative formulas is given by

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

Where the x_k is current iterative, α_k is a positive scalar and named the step-size, determined by some line search, d_k is the search direction defined by

$$d_k = \begin{cases} -g_k, k = 1 \\ -g_k + \beta_k d_{k-1}, k > 1 \end{cases} \tag{3}$$

Where $g_k = \nabla f(x_k)$ is the gradient of f(x) at the x_k and β_k is a scalar which determined different conjugate gradient methods [1, 2].The line search in the conjugate gradient algorithms often is based on the general Wolfe condition [3,4]:

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k \tag{4}$$

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k \tag{5}$$

Where $0<\delta<\sigma<1$. There are many well known formulas for β_k , such as the Fletcher-Reeves(FR) [5], Polar-Ribiere-Polyak(PRP) [6, 7] .Hestenes-Stiefe(HS) [8],and Conjugate Descent(CD)[9], Dai-Yuan(DY)[10]Quite recently, Birgin and Martinez [11] propose a spectral conjugate gradient method by combining conjugate gradient and spectral gradient method .Unfortunately, the spectral conjugate gradient method cannot guarantee to generate descent direction. So base on the FR formula, Zhang et al. [12] modified the FR conjugate gradient method.

$$d_k = \begin{cases} -g_k, k = 1 \\ -\theta_k g_k + \beta_k^{FR} d_{k-1}, k > 1 \end{cases} \tag{6}$$

Where $\beta_k = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}$, $\theta_k = \frac{d_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2}$

And they proved the global convergence of the modified spectral FR method with the mild conditions. And then Huabin Jiang and Songhai Deng [13] propose a spectral conjugate gradient method about PRP method, they proof the global convergence under the Armijo line search, they proof when the exact line search is used, when $\theta_k = 1$.

II. NEW SPECTRAL CONJUGATE GRADIENT METHOD

Consider the general spectral conjugate gradient method (2) and

$$d_k = \begin{cases} -g_k, k = 1 \\ -\theta_k g_k + \beta_k d_{k-1}, k > 1 \end{cases} \tag{7}$$

Where α_k (2) is a step-size satisfy the Wolfe condition(4),(5)and θ_k is a coefficient, β_k is a scalar which determined different spectral conjugate gradient methods .We assume that the search direction is descent direction for k-1 steps, and we hope the conclusion also for k steps, so we have

$$g_k^T d_k < 0 \tag{8}$$

From the (7) and multiplying by g_k^T , and from the (8) we have

$$-\theta_k \|g_k\|^2 + \beta_k g_k^T d_{k-1} < 0 \tag{9}$$

We observe the formula (9), let the form of θ_k to be $\theta_K = 1 + A_K$, and we assume $\beta_k > 0$, we have

$$\frac{\theta_k \|g_k\|^2}{\beta_k} > g_k^T d_{k-1} \quad (10)$$

From (5) we have

$$\frac{\|g_k\|^2}{\beta_k} + \frac{A_k \|g_k\|^2}{\beta_k} = -g_{k-1}^T d_{k-1} + g_k^T d_{k-1} \quad (11)$$

From our assume we known that $g_{k-1}^T d_{k-1} < 0$ and $\beta_k > 0$, and from (11) we let

$$\frac{\|g_k\|^2}{\beta_k} = -g_{k-1}^T d_{k-1} \quad (12)$$

$$\frac{A_k \|g_k\|^2}{\beta_k} = g_k^T d_{k-1} \quad (13)$$

From the (12), (13) we know that

$$\theta_k = 1 - \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}} \quad (14)$$

$$\beta_k = -\frac{\|g_k\|^2}{g_{k-1}^T d_{k-1}} \quad (15)$$

III. ALGORITHM AND LEMMAS

In order to establish the global convergence of the proposed method, we need the assumption on objective function, which have been used often in the literature to analyze the global convergence of nonlinear conjugate gradient with inexact line search.

Assumption 3.1

(1) The level set $\Omega = \{x \mid f(x) \le f(x_1)\}$ is bounded, where x_1 is the starting point.

(2) In some neighborhood N of Ω ,the objective function is continuously differentiable , and its gradient is Lipchitz continuous ,so there exists a constant L>0 such that

$$\|g(x) - g(y)\| \le L\|x - y\|, \forall x, y \in N$$

Algorithm 3.2

Step 1. Data: $x_1 \in R^n, \varepsilon \ge 0$ .Set $d_1 = -g_1$, if $\|g_1\| \le \varepsilon$,then stop, otherwise go to Step 2;

Step 2.Computer α_k by some line search

Step3.Let $x_{k+1} = x_k + \alpha_k d_k, g_{k+1} = g(k+1)$,if $\|g_{k+1}\| \le \varepsilon$, then stop, otherwise go to the Step 4.

Step 4.Computer β_{k+1} by (15), and generate d_{k+1} by (7)

Step 5.Set k=k+1, go to Step 2

Lemma 3.3. Consider any method (2), (7), where(14),(15) and the step-size α_k be determined by the any line search, then

$$g_k^T d_k < 0 \quad (16)$$

Proof: From the (7) and multi plying by g_k^T ,the parameter determined by (71), (72), we have

$$g_k^T d_k = -(1 - \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}})\|g_k\|^2 - \frac{\|g_k\|^2}{g_{k-1}^T d_{k-1}} g_k^T d_{k-1} = -\|g_k\|^2 \quad (17)$$

From the above, we obtain that the conclusion

And from the the equation (15) and (17) we have

$$\beta_k = \frac{g_k^T d_k}{g_{k-1}^T d_{k-1}} \quad (18)$$

Lemma 3.4 Support that assumption (16) holds, Consider any method (2), (7), where(14),(15) and the step-size α_k be determined by the Wolfe line search, then

$$\sum_{k \ge 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty \quad (19)$$

The above lemma (19) often called Zoutendijk condition, is used to prove the global convergence properties of nonlinear conjugate gradient methods. It was originally given by Zoutendijk.

IV. GLOBAL CONVERGENCE PROPERTY

Theorem 20 .Suppose that Assumption (16) holds, Consider any method (2), (7), where(14),(15) and the step-size α_k be determined by the Strong Wolfe line search, then

$$\lim_{k \to +\infty} \inf \|g_k\| = 0 \quad (20)$$

Proof：Support by contradiction that there exists a positive constant $\gamma > 0$, such that

$$\|g_k\| \ge \gamma \quad (21)$$

From (7), we have $d_k + \theta_k g_k = \beta_k d_{k-1}$, and by squaring it, we get

$$\|d_k\|^2 = \beta_k^2 \|d_{k-1}\|^2 - 2\theta_k g_k^T d_k - \theta_k^2 \|g_k\|^2 \quad (22)$$

From the above equation (18)and dividing by $(g_k^T d_k)^2$ we have

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} = \frac{\|d_{k-1}\|^2}{(g_{k-1}^T d_{k-1})^2} - 2\frac{\theta_k}{g_k^T d_k} - \theta_k^2 \frac{\|g_k\|^2}{(g_k^T d_k)^2} \quad (23)$$

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

$$\le \frac{\|d_{k-1}\|^2}{(g_{k-1}^T d_{k-1})^2} + \frac{1}{\|g_k\|^2} \quad (24)$$

Using recursively and $\|d_1\|^2 = -g_1^T d_1 = \|g_1\|^2$, we have

$$\frac{\|d_k\|^2}{(g_k^T d_k)} \le \sum_{i=1}^{k} \frac{1}{\|g_i\|^2} \quad (25)$$

From (21) and (24), we have

$$\frac{(g_k^T d_k)^2}{\|d_k\|^2} \ge \frac{\gamma^2}{k} \quad (26)$$

Which indicates

$$\sum_{k\ge 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} = +\infty \quad (27)$$

This contradicts the Zoutendijk condition (19).Therefore the conclusion (20) holds.

V. THE INHERENT PROPERTY OF THE METHOD AND ITS APPLICATIONS

We will introduce the inherent nature of the method under the any line search condition and we define

$$q_k = \frac{\|d_k\|^2}{(g_k^T d_k)^2} \quad (28)$$

And

$$r_k = -\frac{g_k^T d_k}{\theta_k \|g_k\|^2} \quad (29)$$

We use (28), (29) change the equation (23), we have

$$q_k = q_{k-1} + \frac{1}{\|g_k\|^2}\cdot\frac{2}{r_k} - \frac{1}{\|g_k\|^2}\cdot\frac{1}{r_k^2} \quad (30)$$

We assume exit positive scalar γ and $\bar{\gamma}$ satisfy

$$\gamma \le \|g_k\| \le \bar{\gamma}, \forall k \ge 1 \quad (31)$$

Lemma5.1 Consider the method (2) ,(7),(14),(15), $\theta_k > 0$ and(16) if(31) holds, there exit positive constants $\delta_1, \delta_2, \delta_3$,such that relations

$$-g_k^T d_k \ge \frac{\delta_1}{\sqrt{k}} \quad (32)$$

$$\|d_k\|^2 \ge \frac{\delta_2}{\sqrt{k}} \quad (33)$$

$$r_k \ge \frac{\delta_3}{\sqrt{k}} \quad (34)$$

Proof: Using (30) recursively, and noting that $d_1 = -g_1$, and $q_k \ge 0$,we have

$$\frac{1}{\|g_k\|^2}\left(-\frac{2}{r_k} + \frac{1}{r_k^2}\right) \le \sum_{i=1}^{k-1} \frac{1}{\|g_i\|^2}\left(\frac{2}{r_i} - \frac{1}{r_i^2}\right) \quad (35)$$

From (35) and (31) we have

$$\frac{1}{r_k^2} - \frac{2}{r_k} - \frac{\bar{\gamma}}{\gamma}(k-1) \le 0 \quad (36)$$

From (36) and $r_k > 0$, we have

$$\frac{1}{r_k} \le 1 + \sqrt{1 + \frac{\bar{\gamma}^2}{\gamma^2}(k-1)} \le 1 + \frac{\bar{\gamma}}{\gamma}\sqrt{k} \le \frac{2\bar{\gamma}}{\gamma}\sqrt{k} \quad (37)$$

From (37), we have

$$r_k \ge \frac{\delta_3}{\sqrt{k}} \qquad \delta_3 = \frac{\gamma}{2\bar{\gamma}} \quad (38)$$

From (29), we have

$$-g_k^T d_k = \|g_k\|^2 r_k \theta_k \quad (39)$$

$$\|d_k\| \ge \|g_k\| r_k \theta_k \quad (40)$$

From (39), (40), (38) and assume $\theta_k \ge \xi$,we have

$$-g_k^T d_k \ge \frac{\delta_1}{\sqrt{k}} \quad (41)$$

$$\|d_k\|^2 \ge \frac{\delta_2}{\sqrt{k}} \quad (42)$$

And $\delta_1 = \delta_3 \gamma^2 \xi$, $\delta_2 = \delta_3^2 \gamma^2 \xi$

Now we will construct a new line search condition:

$$f_k - f_{k-1} \ge c\min\left\{-g_k^T d_k, \|d_k\|^2, q_k^{-1}\right\} \quad (43)$$

Which c is a scalar and c>0, q_k is from (5.1) and the assumption (16) holds. We can easy to proof

$$f_k - f_{k-1} \ge c q_k^{-1} \quad (44)$$

By the Wolfe line search condition and we can also proof

$$f_k - f_{k+1} \ge c\min\left\{-g_k^T d_k, q_k^{-1}\right\} \quad (45)$$

By the Armijo line search condition. So from the (44), we can know that it satisfy Wolfe line condition and Armijo line condition.

VI. NUMERICAL EXPERIMENTS

From the derivation of the two parameter β_k and θ_k we known that is very similar with the DY methods, and D.Y prove it global convergence under the Wolfe line search condition. The parameter β_k in the new spectral conjugate gradient method is the same to the CD conjugate gradient method. FR is the famous conjugate gradient methods. So in this section, we compare the Spectral conjugate gradient method, donated the SP method, to FR method, CD method under the strong Wolfe line search with the given initial points and dimensions. The parameters are chosen as follows: $\delta = 0.01, \sigma = 0.1$ .The problems we test are from [14], If $\|g_k\| \le 10^{-6}$ is satisfied, we will stop the program. We will be also stopped the program if the number of iteration is more then 9999. All codes were written in FORTRAN90 and run on a PC with 2.00GHz CPU processor and 2.00 G Memory and Windows XP operation system.

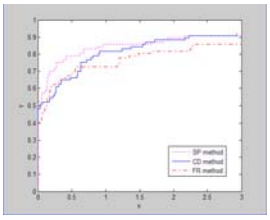

FIGURE I. PERFORMANCE PROFILES FOR THE NUMBER OF ITERATIONS.

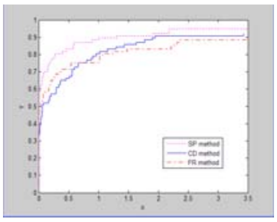

FIGURE II. PERFORMANCE PROFILES FOR THE NUMBER OF FUNCTION EVALUATIONS.

We adopt the performance profile by Dolan and More [15] to compare the SP method with FR method, and CD method. Figure 1-4 show the performance of the three methods relative to the number of iterations, the number of function evaluations, the number of the gradient evaluations and the CPU time.

Obviously, from the four figures show that SP method outperforms FR method, CD method for the given test problems in CPU time, the number of function evaluations, and the number of gradient evaluations, so the SP method is computationally efficient.

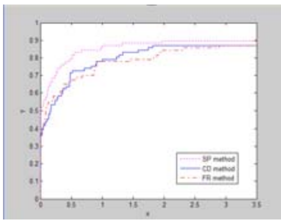

FIGURE III. PERFORMANCE PROFILES FOR THE NUMBER OF GRADIENT EVALUATIONS.

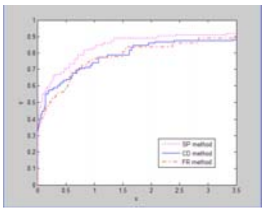

FIGURE IV. PERFORMANCE PROFILES FOR THE CPU TIME.

REFERENCES

[1] G.H.YU, Nonlinear self-scaling conjugate gradient methods for large-scale optimization [Ph.D. thesis], Sun Yat-Sen University, Guangzhou, China 2007

[2] G. Yuan and Z. Wei, New line search methods for unconstrained optimization .Journal of the Korean Statistical Society, vol.38, no.1, pp.29-39, 2009

[3] Wolfe, P., Convergence conditions for ascent methods, SIAM Rev., Vol. 11, pp.226-235, 1968

[4] Wolfe, P., Convergence conditions for ascent methods, (Ⅱ): some corrections. SIAM Review 13 (1971) 185-188

[5] Fletcher, R. and Reeves, C. M. “Function minimization by conjugate gradients,” The Computer Journal, vol. 7. pp. 149-154, 1964

[6] E. Polak and G. Ribiere, “Note sur la convergence de methods de direction conjugate,” vol. 3, no. 16, pp.35-43,1969.

[7] Polyak, B.T., The conjugate gradient method in extreme problems. USSR Comp. Meth. Meth. Phys. 9, 94-112(1969)

[8] M. R. Hestenes and E. stiefel, “Methods of conjugate gradients for solving linear systems,” Journal of Research of the National of Standards, vol. 49, pp.409=436, 1952

[9] R, Fletcher, Practical methods of Optimization: Unconstrained Optimization, vol. John Wiley＆Sons, New York., NY, USA, 1987

[10] Dai, Y.H., Yuan, Y., A nonlinear conjugate gradient method with a strong global convergence property, SIAM. Math. Optim. 43, 87-101(2001)

[11] E. G. Birgin and J. M. Martinez, “A Spectral conjugate gradient method for unconstrained optimization,” Applied Mathematics and Computation, vol. 180, pp. 46-52, 2006

[12] L. Zhang, W. Zhou, and D. Li, “Global convergence of a modified Fletcher-Reeves conjugate gradient method with Armijo-type line search,” Numerische Mathematic, vol. 104, no.64, pp. 561-572, 2006

[13] Huabin Jiang, Songhai Deng, Xiaodong Zhang, and Zhang Wan, “Global Convergence of a Modified Spectral Conjugate Gradient Method,” Journal of Applied Mathematics, 2012

[14] J.J. More, B.S. Garbow and K.E. Hillstrom, Testing unconstrained optimization software, ACM Transactions on Mathematical Software 7(1981) 75-82

[15] E.D. Dolan and J.J. More, “Benchmarking optimization software with performances profiles,” Mathematical Programming, vol. 91, no. 2, pp.201-213, 2002