Journal of Mathematical Research & Exposition
Mar., 2011, Vol. 31, No. 2, pp. 261–268
DOI:10.3770/j.issn:1000-341X.2011.02.009
Http://jmre.dlut.edu.cn

Global Convergence of a Modified Spectral CD Conjugate Gradient Method

Wei CAO*, **Kai Rong WANG**, **Yi Li WANG**
College of Mathematics and Statistics, Chongqing University, Chongqing 401331*, P. R. China*

Abstract In this paper, we present a new nonlinear modified spectral CD conjugate gradient method for solving large scale unconstrained optimization problems. The direction generated by the method is a descent direction for the objective function, and this property depends neither on the line search rule, nor on the convexity of the objective function. Moreover, the modified method reduces to the standard CD method if line search is exact. Under some mild conditions, we prove that the modified method with line search is globally convergent even if the objective function is nonconvex. Preliminary numerical results show that the proposed method is very promising.

Keywords unconstrained optimization; conjugate gradient method; armijo-type line search; global convergence.

Document code A
MR(2010) Subject Classification 90C30
Chinese Library Classification O221.1

1. Introduction

Consider the unconstrained optimization problem

$$\min_{x\in R^n} f(x), \quad x\in R^n \tag{1.0}$$

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

Conjugate gradient methods are very efficient for solving large-scale unconstrained optimization problems (1.0). Its iterative formula is as follows:

$$x_{k+1}=x_k+\alpha_k d_k, \tag{1.1}$$

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

where step-size α_k is positive, $g_k=\nabla f(x_k)$ and $\beta_k$is a scalar. In addition, α_k is a step length which is computed by carrying out line search. There are several line search rules for choosing step-size α_k (see [1]). In this paper, we analyze the general results on convergence of line search methods with the following line search rules:

Received May 12, 2009; Accepted December 8, 2009
Supported by the Key Project of 2010 Chongqing Higher Education Teaching Reform (Grant No. 102104).
* Corresponding author
E-mail address: caowei220@sina.com (W. CAO)

Armijo-type line search rule:

Determine a step-size $\alpha_k = \max\{\rho^j, j = 0, 1, 2, \ldots\}$ by the following constrained condition

$$f(x_k + \alpha_k d_k) \le f(x_k) + \delta_1 \alpha_k g_k^{\mathrm{T}} d_k - \delta_2 \alpha_k^2 \|d_k\|^2 \tag{1.3}$$

where $\rho \in (0, 1)$, $\delta_1 \in (0, 1)$, $\delta_2 > 0$. The Wolfe type line search [15]

$$f(x_k) - f(x_k + \alpha_k d_k) \ge \delta \alpha_k^2 \|d_k\|^2, \tag{1.4}$$

$$g(x_k + \alpha_k d_k)^{\mathrm{T}} d_k \ge -2\sigma \alpha_k \|d_k\|^2. \tag{1.5}$$

The strong Wolfe line search:

$$\begin{aligned} &f(x_k + \alpha_k d_k) \le f(x_k) + \delta \alpha_k g_k^{\mathrm{T}} d_k, \\ &|d_k^{\mathrm{T}} g(x_k + \alpha_k d_k)| \le |\sigma d_k^{\mathrm{T}} g_k|, \end{aligned} \tag{1.6}$$

where $0 < \delta < \sigma < 1$.

Since 1952, there have been many formulas to construct the scalar, for example:

$$\beta_k^{\mathrm{FR}} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \ \beta_k^{\mathrm{PRP}} = \frac{g_k^{\mathrm{T}} y_{k-1}}{\|g_{k-1}\|^2}, \ \beta_k^{\mathrm{DY}} = \frac{\|g_k\|^2}{d_{k-1}^{\mathrm{T}} y_{k-1}}, \ \beta_k^{\mathrm{CD}} = -\frac{\|g_k\|^2}{d_{k-1}^{\mathrm{T}} g_{k-1}},$$

where $y_{k-1} = g_k - g_{k-1}$, and $\|\cdot\|$ stands for the Euclidean norm. They are successively called FR, PRP, DY and CD methods, while correlative conjugate gradient methods can be found in [2–6].

The PRP, FR and DY methods are thereof the well-known conjugate gradient methods. Although the CD nonlinear conjugate gradient method has a similar structure to the FR and DY methods, research about this method is very rare. If the line search is exact, i.e., $g_k^{\mathrm{T}} d_{k-1} = 0$, then DY and CD methods reduce to FR method. Hence, DY and CD methods have the same disadvantage. The CD method and DY method are proved to have global convergence under strong Wolfe line search [5, 6, 12, 13]. The conjugate gradient methods have extensive applications in many fields [6, 14]. Recently, Birgin and Martinez [8] proposed a spectral conjugate gradient method by combining conjugate gradient method and spectral gradient method [7]. The direction d_k is given by

$$d_k = -\theta_k g_k + \beta_k d_{k-1}, \ \beta_k = \frac{(\theta_k y_{k-1} - s_{k-1})^{\mathrm{T}} g_k}{d_{k-1}^{\mathrm{T}} y_{k-1}},$$

where $y_{k-1} = g_k - g_{k-1}$, $s_{k-1} = x_k - x_{k-1}$, $\theta_k = \frac{s_{k-1}^{\mathrm{T}} s_{k-1}}{s_{k-1}^{\mathrm{T}} y_{k-1}}$ is a parameter, and θ_k is taken to be the spectral gradient. Unfortunately, the spectral conjugate gradient method [8] cannot guarantee to generate descent directions.

The purpose of this paper is to describe the modified CD method whose form is similar to that of [9] but with different parameters θ_k and β_k. Under some mild conditions, we give the global convergence of the modified spectral CD method with the Armijo-type line search rules.

2. The algorithm of the modified CD method

We first show that if exact line search is used, the new method is identical to the CD method. The iterates of the new method are obtained by (1.1) with

$$d_k = \begin{cases} -g_k, & \text{if } k = 1; \\ -\theta_k g_k + \beta_k d_{k-1}, & \text{if } k > 1, \end{cases} \tag{2.1}$$

where $\theta_k = 1 - \frac{g_k^{\mathrm{T}} d_{k-1}}{d_{k-1}^{\mathrm{T}} g_{k-1}}$ is a parameter, $\beta_k = \beta_k^{\mathrm{CD}}$. We call the methods (1.1) and (2.1) as the MCD method. From β_k^{CD} and (2.1), it is easy to see that for any $k \geq 1$,

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

This is a very important property for conjugate gradient methods.

Now we give the concrete algorithm and assumptions as follows:

Algorithm

Step 0. Given $x_1 \in R^n$, $\epsilon \geq 0$; set $d_1 = -g_1 = -\nabla f(x_1)$, $k = 1$, if $\|g_1\| \leq \epsilon$, then stop.

Step 1. Find $\alpha_k > 0$ satisfying line search rules.

Step 2. Let $x_{k+1} = x_k + \alpha_k d_k$ and $g_{k+1} = g(x_{k+1})$; if $\|g_{k+1}\| \leq \epsilon$, then stop.

Step 3. Compute β_{k+1} by the formulae $\beta_{k+1}^{\mathrm{CD}}$, then generate d_{k+1} by (2.1).

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

Assumptions

Throughout this paper, we make the following assumptions:

(i) The level set $\Omega = \{x \in R^n | f(x) \leq f(x_1)\}$ is bounded;

(ii) In some neighborhood N of Ω, $f(x)$ is continuously differentiable and its gradient is Lipschitz continuous, namely, there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \quad \forall x, y \in N. \tag{2.3}$$

It follows directly from above assumptions that there exist two positive constants β and γ such that

$$\|x\| \leq B, \|g(x)\| \leq \gamma, \quad \forall x \in \Omega. \tag{2.4}$$

3. Convergence analysis

From above, the MCD conjugate gradient method provides a descent direction for the objective function and this property depends neither on the line search rule, nor on the convexity of the objective function. We will give the global convergence of algorithm as follows.

Lemma 3.1 *Suppose that assumptions (i) and (ii) hold. Let $\{g_k\}$ and $\{d_k\}$ be generated by the proposed algorithm with Armijo- type line search rules, then there exists a constant $\tau > 0$, such that the following inequality holds for all k,*

$$\alpha_k \geq \tau \frac{\|g_k\|^2}{\|d_k\|^2}. \tag{3.1}$$

Proof When $\alpha_k = 1$, we have $\|g_k\| \le \|d_k\|$ by (2.2), which implies that (3.1) holds for $\tau = 1$. When $\alpha_k < 1$, by the definition of Armijo- type line search rule, $\rho^{-1}\alpha_k$ does not satisfy inequality (1.3). This means

$$f(x_k + \rho^{-1}\alpha_k d_k) - f(x_k) > \delta_1 \alpha_k \rho^{-1} g_k^{\mathrm{T}} d_k - \delta_2 \rho^{-2} \alpha_k^2 \|d_k\|^2. \tag{3.2}$$

By the mean-value theorem and inequality (2.3), there is a $t_k \in (0,1)$ such that $x_k + t_k \rho^{-1}\alpha_k d_k \in N$, and

$$\begin{aligned} & f(x_k + \rho^{-1}\alpha_k d_k) - f(x_k) = \alpha_k \rho^{-1} g(x_k + t_k \rho^{-1}\alpha_k d_k)^{\mathrm{T}} d_k \\ & \quad = \alpha_k \rho^{-1} g_k^{\mathrm{T}} d_k + \alpha_k \rho^{-1}(g(x_k + t_k\rho^{-1}\alpha_k d_k) - g_k)^{\mathrm{T}} d_k \le \alpha_k \rho^{-1} g_k^{\mathrm{T}} d_k + L\rho^{-2}\alpha_k^2 \|d_k\|^2. \end{aligned}$$

From the last inequality and (3.2), we get

$$\alpha_k > \frac{(1-\delta_1)\rho \|g_k\|^2}{(L+\delta_2)\|d_k\|^2}.$$

Letting $\tau = \min\{1, \frac{(1-\delta_1)\rho}{(L+\delta_2)}\}$, we get (3.1). □

Lemma 3.2 *Suppose that assumptions (i) and (ii) hold. Let $\{g_k\}$ and $\{d_k\}$ be generated by the proposed algorithm with Armijo- type line search rules. Then, we have the Zoutendijk condition [11]*

$$\sum_{k\ge 0} \frac{\|g_k\|^4}{\|d_k\|^2} < +\infty. \tag{3.3}$$

Proof We have from (1.3) and assumption (i) that

$$\sum_{k\ge 0} (-\delta_1 \alpha_k g_k^{\mathrm{T}} d_k + \delta_2 \alpha_k^2 \|d_k\|^2) < \infty,$$

which together with equality (2.2) yields

$$\sum_{k\ge 0} \alpha_k^2 \|d_k\|^2 < \infty, \quad \sum_{k\ge 0} \alpha_k \|g_k\|^2 = -\sum_{k\ge 0} \alpha_k g_k^{\mathrm{T}} d_k < \infty. \tag{3.4}$$

From (3.1) and (3.4), we can easily obtain the above Zoutendijk condition. □

Lemma 3.3 *Suppose that assumptions (i) and (ii) hold. Let $\{g_k\}$ and $\{d_k\}$ be generated by the proposed algorithm with Wolfe type line search rule. Then*

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

Proof We have from (1.4), (1.5) and assumption (i) that $(2\sigma + L)\alpha_k \|d_k\|^2 \ge -g_k^{\mathrm{T}} d_k$. So we have $\alpha_k \|d_k\| \ge \frac{1}{2\sigma+L}(\frac{-g_k^{\mathrm{T}} d_k}{\|d_k\|})$. Combining with inequality (1.4), we know that

$$\begin{aligned} \sum_{k=1}^{\infty} \frac{(g_k^{\mathrm{T}} d_k)^2}{\|d_k\|^2} &= \sum_{k=1}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} \le (2\sigma + L)^2 \sum_{k=1}^{\infty} \alpha_k^2 \|d_k\|^2 \\ &\le \frac{(2\sigma+L)^2}{\delta} \sum_{k=1}^{\infty} \{f(x_k) - f(x_{k+1})\} < +\infty. \end{aligned}$$

From (2.2), we can easily obtain the Zoutendijk condition. □

The following is our main theorem of this paper.

Therorem 3.1 *Suppose that assumptions (i) and (ii) hold. Let $\{g_k\}$ and $\{d_k\}$ be generated by the proposed algorithm. Then we have*

$$\lim_{k\to\infty} \inf \|g_k\| = 0. \tag{3.5}$$

Proof Suppose by contradiction that there exists a positive constant $\bar{\gamma} > 0$, such that

$$\|g_k\| \geq \bar{\gamma} \tag{3.6}$$

holds for all $k \geq 1$. From (2.1), we have

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

Dividing both sides of this equality by $(g_k^{\mathrm{T}} d_k)^2$, we get from (2.2), (3.6) and β_k^{CD} that

$$\begin{aligned}
\frac{\|d_k\|^2}{\|g_k\|^4} = \frac{\|d_k\|^2}{(g_k^{\mathrm{T}} d_k)^2} &= \frac{-2\theta_k d_k^{\mathrm{T}} g_k}{(g_k^{\mathrm{T}} d_k)^2} + \frac{(\beta_k^{\mathrm{CD}})^2 \|d_{k-1}\|^2}{(g_k^{\mathrm{T}} d_k)^2} - \frac{\theta_k^2 \|g_k\|^2}{(g_k^{\mathrm{T}} d_k)^2} \\
&= \Big(-\frac{\|g_k\|^2}{d_{k-1}^{\mathrm{T}} g_{k-1}} \Big)^2 \frac{\|d_{k-1}\|^2}{(g_k^{\mathrm{T}} d_k)^2} + \frac{2\theta_k}{\|g_k\|^2} - \frac{\theta_k^2}{\|g_k\|^2} \\
&= \frac{\|d_{k-1}\|^2}{(g_{k-1}^{\mathrm{T}} d_{k-1})^2} - \frac{1}{\|g_k\|^2} (\theta_k^2 - 2\theta_k + 1 - 1) \\
&= \frac{\|d_{k-1}\|^2}{(g_{k-1}^{\mathrm{T}} d_{k-1})^2} - \frac{(\theta_k - 1)^2}{\|g_k\|^2} + \frac{1}{\|g_k\|^2} \\
&\leq \frac{\|d_{k-1}\|^2}{(g_{k-1}^{\mathrm{T}} d_{k-1})^2} + \frac{1}{\|g_k\|^2} \leq \sum_{i=1}^{k} \frac{1}{\|g_i\|^2} \leq \frac{k}{\bar{\gamma}^2}.
\end{aligned}$$

Hence, we obtain

$$\sum_{k\geq 1} \frac{\|g_k\|^4}{\|d_k\|^2} \geq \bar{\gamma}^2 \sum_{k\geq 1} \frac{1}{k} = \infty,$$

which contradicts (3.3). Therefore, we can conclude that (3.5) holds. □

4. Numerical results

In this part, we have tested the algorithms suggested by this paper. In order to weigh the numerical effects of the different methods, we also test the following CG methods:

MCD1: the (2.1) formula with (1.3) conditions, where $\delta = 0.001$, $\sigma = 0.01$, $\rho = 0.49$, the termination condition is $\|g_k\| \leq 10^{-6}$ or It-limit > 9999;

MCD2: the (2.1) formula with (1.4), (1.5) conditions, where $\delta = 0.001$, $\sigma = 0.01$, the termination condition is $\|g_k\| \leq 10^{-6}$or It-limit > 9999;

CD: the β_k^{CD} formula with (1.6) conditions, where $\delta = 0.001$, $\sigma = 0.01$, the termination condition is $\|g_k\| \leq 10^{-6}$ or It-limit > 9999;

MFR [9]: the Algorithm with (1.3) conditions, where $\delta = 0.001$, $\sigma = 0.01$, $\rho = 0.49$, the termination condition is $\|g_k\| \leq 10^{-6}$ or It-limit > 9999.

Where It-limit denotes the iterative limit.

N	NAME	DIM	NFR: NI/NF/NG	MCD1: NI/NF/NG	MCD2: NI/NF/NG
1	ROSE	2	105/1000/106	78/716/79*	87/1967/1286
2	FROTH	2	70/776/71	67/735/68*	···
4	BADSCB	2	167/7437/167*	184/8104/184	···
5	BEALE	2	53/265/54*	57/285/58	103/268/152
6	JENSAM	2 ($m=8$)	40/1579/40	35/1510/35*	···
7	HELIX	3	260/2628/261	114/1003/115	38/326/206*
9	GAUSS	3	25/76/26	25/76/26	9/19/12*
10	MEYER	3	584/25916/584	511/21579/511/*	···
16	BD	4	283/5987/283	139/3289/139*	···
20	WATSON	5	781/6718/782	303/2338/304	85/343/204*
21	ROSEX	8	113/1082/114	79/726/80*	88/1969/1287
		100	126/1243/127	82/747/83*	93/1979/1292
		500	165/1648/166	96/918/97	93/1979/1292*
24	PEN2	100	191/2455/191	160/2160/160*	···
25	VARDIM	2	20/89/21	20/89/21	2/32/21*
		5	27/203/28	27/203/28	2/61/41*
		10	19/221/20	19/221/20	2/87/58*
31	BAND	5	2/1010/2	2/1010/2	···
		10	42/2285/42	40/2197/40*	···

Table 1 Test results for the MCD1, MCD2 methods and the MFR method

In Tables 1 and 2, “NAME”,“DIM”,“NI”,“NF” and “NG” stand for the name of the test problem in MATLAB, the dimension of the problem, the number of iterations, the number of function evaluations and the number of gradient evaluations, respectively while the star * denotes that this result is the best one among these methods and "$\cdots$" means the iteration failed.

The experiments are carried out on some famous test problems which can be found in [9]. We compare the MCD1 and MCD2 methods with MFR in Table 1. It can be seen that the MCD1 method performes better than the other methods, and becomes more effective when the problem is a large scale one. We see from Table 2 that "MCD1" has the best performance since it solves more test problems than the "CD" with the least number of iterations. Although the method "CD" has good performance in some test problems, it fails to terminate successfully for many problems. The method "MCD1" can solve all test problems successfully in Table 2.

NAME	DIM	CD: NI/NF/NG	MCD1: NI/NF/NG
ROSE	2	109/381/329	78/716/79*
FROTH	2	17/100/80*	67/735/68
BADSCB	2	$\cdots$	184/8104/184*
BEALE	2	70/202/171	57/285/58*
JENSAM	$2\,(m=8)$	13/67/40*	35/1510/35
HELIX	3	125/363/314	114/1003/115*
BD	4	$\cdots$	139/3289/139*
WATSON	5	96/303/257*	303/2338/304
ROSEX	8	115/402/347	79/726/80*
	100	116/404/348	82/747/83*
	500	122/425/367	96/918/97*
PEN2	100	60/236/193*	160/2160/160
BAND	5	12/62/40	2/1010/2*
	10	$\cdots$	40/2197/40*

Table 2 Test results for the MCD1 and the CD method

In the above tables, all codes are written in MATLAB 7.0.1 and run on a PC with a 1.73 GHz Genuine Intel(R)CPU processor and 1.00 GB RAM memory and the Windows VistaTM Home Basic system.

As a conclusion, we remark that further research on the new algorithm is still needed to be done. For example, we should analyze the convergence of the new algorithm with some other line search rules, employ our algorithm to solve some large practical problems and study the effect of the parameter on the algorithm.

References

[1] SHI Zhenjun. *Convergence of line search methods for unconstrained optimization* [J]. Appl. Math. Comput., 2004, **157**(2): 393–405.

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

[3] POLAK E, RIBIÈRE G. *Note sur la convergence de méthodes de directions conjuguées* [J]. Rev. Française Informat. Recherche Opérationnelle, 1969, **3**(16): 35–43. (in French)

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

[5] DAI Yuhong, YUAN Yaxiang. *A nonlinear conjugate gradient method with a strong global convergence property* [J]. SIAM J. Optim., 1999, **10**(1): 177–182.

[6] FLETCHER R. *Practical Methods of Optimization* [M]. John Wiley & Sons, Ltd., Chichester, 1987.

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

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

[9] ZHANG Li, ZHOU Weijun, LI Donghui. *Global convergence of a modified Fletcher-Reeves conjugate gradient method with Armijo-type line search* [J]. Numer. Math., 2006, **104**(4): 561–572.

[10] MORÉ J J, GARBOW B S, HILLSTROM K E. *Testing unconstrained optimization software* [J]. ACM Trans. Math. Software, 1981, **7**(1): 17–41.

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

[12] DAI Yuhong, YUAN Yaxiang. *Convergence properties of the Fletcher-Reeves method* [J]. IMA J. Numer. Anal., 1996, **16**(2): 155–164.

[13] DAI Yuhong, YUAN Yaxiang. *Convergence properties of the conjugate descent method* [J]. Adv. in Math. (China), 1996, **25**(6): 552–562.

[14] YUAN Yaxiang. *Numerical Methods for Nonlinear Programming* [M]. Shanghai: Shanghai Science and Technology Publisher, 1993. (in Chinese)

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