TELKOMNIKA, Vol.11, No.11, November 2013, pp. 6373~6380
e-ISSN: 2087-278X

A Modified Conjugate Gradient Method for Unconstrained Optimization

Can Li
College of Mathematics, Honghe University, Mengzi, China
Corresponding author, e-mail: lican841014@163.com

Abstract

Conjugate gradient methods are an important class of methods for solving unconstrained optimization problems, especially for large-scale problems. Recently, they have been studied in depth. In this paper, we further study the conjugate gradient method for unconstrained optimization. We focus our attention to the descent conjugate gradient method. This paper presents a modified conjugate gradient method. An interesting feature of the presented method is that the direction is always a descent direction for the objective function. Moreover, the property is independent of the line search used. Under mild conditions, we prove that the modified conjugate gradient method with Armijo-type line search is globally convergent. We also present some numerical results to show the efficiency of the proposed method.

Keywords: *unconstrained optimization problem, conjugate gradient method, armijo-type line search, global convergence*

1. Introduction

Let us consider the following unconstrained optimization problem:

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

Where $f: R^n \to R$ is a continuously differentiable function and its gradient is denoted by $g(x) = \nabla f(x)$. Optimization problems exist in many areas [1-2], such as engineering, production management, economy etc. Generally, Optimization problem is solved by converted to unconstrained optimization problem.

Conjugate gradient methods constitute an excellent choice for efficiently solving the optimization problem (1), especially when the dimension n is large due to the simplicity of their iteration, convergence properties and their low memory requirements. In fact, conjugate gradient methods have played special roles in solving large scale nonlinear optimization problems. Although conjugate gradient methods are not the fastest or most robust optimization algorithms for nonlinear problems available today, they remain very popular for engineers and mathematicians engaged with solving large problems.

Conjugate gradient methods generate a sequence of points $\{x_k\}$, starting from an initial guess $x_0 \in R^n$, using the iterative formula

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

Where $\alpha_k > 0$ is obtained by line search, and the direction d_k is generated by:

$$d_k = \begin{cases} -g_k & k = 0, \\ -g_k + \beta_k d_{k-1} & k \geq 1, \end{cases} \tag{3}$$

Where $g_k = g(x_k)$ and β_k is a scalar. Some well-known formulae for β_k are given in [3-8].

Received April 9, 2013; Revised May 31, 2013; Accepted Jun 27, 2013

For a conjugate gradient method in the form (2)-(3), we say that the descent condition holds if:

$$g_k^{\mathrm{T}} d_k < 0, \qquad \forall k \geq 0. \tag{4}$$

In addition, we say that the sufficient descent condition holds if there exists a constant $c_1 > 0$ such that:

$$g_k^{\mathrm{T}} d_k < -c_1 \| g_k \|^2, \qquad \forall k \geq 0, \tag{5}$$

Where $\|\cdot\|$ stand for the Euclidean norm of vectors. The descent condition or the sufficient descent condition is often used in the literature to analyze the global convergence of conjugate gradient method with inexact line search, and may be crucial for conjugate gradient methods. But for classical conjugate gradient methods, the descent condition or the sufficient descent condition holds depending on the line search used. During the last decade, much effort has been devoted to generate descent conjugate gradient methods independent of the line search used. Similar to the spectral conjugate gradient method [9], Zhang, Zhou and Li [10] proposed a descent modified FR conjugate gradient method as:

$$d_k = -\frac{d_{k-1}^{\mathrm{T}} y_{k-1}}{\| g_{k-1} \|^2} g_k + \frac{\| g_k \|^2}{\| g_{k-1} \|^2} d_{k-1},$$

Where $d_0 = -g_0$ and $y_{k-1} = g_k - g_{k-1}$. A remarkable property of the method is that it produces descent direction,

$$g_k^{\mathrm{T}} d_k = -\| g_k \|^2, \qquad \forall k \geq 0.$$

and this property is independent of the line search used. Motivated by this nice descent property, Zhang, Xiao and Wei [11] introduced a descent three-term conjugate gradient method based on the Dai-Liao method [12] as:

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

Where $d_0 = -g_0$, $s_{k-1} = x_k - x_{k-1}$ and $t \geq 0$. Again, it is easy to see that the sufficient descent condition also holds independent by the line search, i.e. for this method $g_k^{\mathrm{T}} d_k = -\| g_k \|^2$ for all k. Other descent conjugate gradient methods and their global convergence can be found in [13-17] etc.

In this paper, we propose a modified conjugate gradient method. The direction generated by the proposed method is always a descent direction of the objective function. This property is independent of the line search used. Under mild conditions, we prove that the modified conjugate gradient method with Armijo-type line search is globally convergent.

In the next section, we propose the method. Section 3 is devoted to the global convergence of the method. At last, we present some numerical results in section 4.

2. Algorithm

Recently, Li and Feng [18] proposed a modified Liu-Storey (MLS) method by letting:

$$\beta_k = -\frac{g_k^{\mathrm{T}} y_{k-1}}{d_{k-1}^T g_{k-1}} - t \frac{\|y_{k-1}\|^2 g_k^{\mathrm{T}} d_{k-1}}{(d_{k-1}^T g_{k-1})^2},$$

Where $t > \frac{1}{4}$. In [18], Li and Feng proved that the MLS method can always generate descent directions which satisfy the sufficient descent condition.

$$g_k^{\mathrm{T}} d_k \le -(1-\frac{1}{4t}) \| g_k \|^2 .$$

Li and Feng also establish the global convergence of the MLS method with strong Wolfe line search. Based on the MLS method, we propose our algorithm as follows.

Algorithm 1:

Step 1. Given constants $\rho > 0$, $\frac{1}{4} < \mu < 1$, $0 < \delta < 1$, $\varepsilon > 0$. Choose an initial point

$x_0 \in R^n$ Set $d_0 = -g_0$, $k =: 0$.

Step 2. Calculate the search direction d_k by (3), where β_k is defined by:

$$\beta_k = \frac{1}{d_{k-1}^T g_{k-1}} (g_k - \mu \frac{\|g_k\|^2}{d_{k-1}^T g_{k-1}} d_{k-1})^T g_k . \tag{6}$$

Where $\frac{1}{4} < \mu < 1$.

Step 3. Determine $\alpha_k = \max\{\rho^j, j = 0,1,2,...\}$ satisfying

$$f(x_k + \alpha_k d_k) \le f(x_k) - \delta \alpha_k^2 \|d_k\|^4 . \tag{7}$$

Step 4. Let $x_{k+1} = x_k + \alpha_k d_k$. If $\|g_{k+1}\|^2 < \varepsilon$, then stop. Otherwise go to step 5.

Step 5. Let $k =: k+1$ and go to step 2.

The following theorem shows that the Algorithm 1 possesses the sufficient condition (4).

Theorem 1. Let the sequences $\{g_k\}$ and $\{d_k\}$ be generated by the Algorithm 1, then:

$$g_k^T d_k \le -(1-\frac{1}{4\mu})\|g_k\|^2, \qquad \forall k \ge 0. \tag{8}$$

Proof: Since $d_0 = -g_0$, we have $g_0^T d_0 \le -\|g_0\|^2$, which satisfies (8). Multiplying the direction d_k by g_k^{T}, we have from (3) and (6) that for all $k \ge 1$

$$\begin{aligned} g_k^T d_k &= g_k^T \left[-g_k + \frac{1}{d_{k-1}^T g_{k-1}} (g_k - \mu \frac{\|g_k\|^2}{d_{k-1}^T g_{k-1}} d_{k-1})^T g_k . d_{k-1} \right] \\ &= -\|g_k\|^2 + \frac{\|g_k\|^2 (g_k^T \mathrm{d}_{k-1})}{d_{k-1}^T g_{k-1}} - \mu \frac{\|g_k\|^2 (g_k^T \mathrm{d}_{k-1})^2}{(d_{k-1}^T g_{k-1})^2} . \end{aligned} \tag{9}$$

Applying the inequality $u^{\mathrm{T}} v \le \frac{1}{2}(\| u \|^2 + \| v \|^2)$ into $\frac{\|g_k\|^2 (g_k^T \mathrm{d}_{k-1})}{d_{k-1}^T g_{k-1}}$, we obtain:

$$\frac{\|g_k\|^2 (g_k^T \mathrm{d}_{k-1})}{d_{k-1}^T g_{k-1}} = \frac{\left[\frac{1}{\sqrt{2\mu}} (g_{k-1}^T \mathrm{d}_{k-1}) g_k \right]^T \left[\sqrt{2\mu} (g_k^T \mathrm{d}_{k-1}) g_k \right]}{(d_{k-1}^T g_{k-1})^2}$$

$$\leq \frac{\frac{1}{2}\left[\frac{1}{2\mu}(d_{k-1}^T g_{k-1})^2 \|g_k\|^2 + 2\mu(g_k^T \mathrm{d}_{k-1})^2 \|g_k\|^2\right]}{(d_{k-1}^T g_{k-1})^2}$$

$$= \frac{1}{4\mu}\|g_k\|^2 + \mu \frac{\|g_k\|^2 (g_k^T \mathrm{d}_{k-1})^2}{(d_{k-1}^T g_{k-1})^2}$$

Substituting the inequality into (9), we get:

$$g_k{}^T d_k \leq -\|g_k\|^2 + \frac{1}{4\mu}\|g_k\|^2 = -(1 - \frac{1}{4\mu})\|g_k\|^2 .$$

From the proof of Theorem 1, we can see that d_k provides a descent direction of f at x_k, and this property is independent of the line search used.

3. The Properties and the Global Convergence

In the following, we assume that $g_k \neq 0$ for all k, otherwise a stationary point has been found. The following assumptions are often used to prove the convergence of the conjugate gradient method.

Assumption A:

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

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

$$\|g(x) - g(y)\| \leq L\|x - y\|, \forall x, y \in N. \tag{10}$$

Since d_k is a descent direction of f at x_k, Algorithm 1 is well defined. Moreover, it follows from (7) that the function value sequence $\{f(x_k)\}$ is decreasing. We also have from (7) that $\sum_{k=0}^{\infty} \alpha_k^2 \|d_k\|^4 < \infty$, if f is bounded from below. In particular, we have

$$\lim_{k\to\infty} \alpha_k \|d_k\|^2 = 0 \tag{11}$$

In addition, we can get from Assumption A that there exists a constant $\gamma > 0$ such that:

$$\|g_k\| \leq \gamma, \quad \forall x \in \Omega. \tag{12}$$

Lemma 1. Let the sequences $\{x_k\}$ be generated by Algorithm 1. If there exists a constant $\varepsilon > 0$ such that:

$$\|g_k\| \geq \varepsilon, \forall k, \tag{13}$$

Then there exists a constant $Q > 0$ such that:

$$\|d_k\| \leq Q. \tag{14}$$

Proof: By the definition of β_k, we can simplify its expression as:

$$\beta_k = \frac{1}{g_{k-1}^T d_{k-1}}(g_k - \mu \frac{\|g_k\|^2}{g_{k-1}^T d_{k-1}} d_{k-1})^T g_k = \frac{\|g_k\|^2}{g_{k-1}^T d_{k-1}}\left(1-\mu \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}}\right)$$
$$= \frac{\|g_k\|^2}{(g_{k-1}^T d_{k-1})^2}\left(g_{k-1}^T d_{k-1} - \mu g_k^T d_{k-1}\right) < \frac{\|g_k\|^2}{\left(g_{k-1}^T d_{k-1}\right)^2} \cdot \mu \left(g_{k-1}^T d_{k-1} - g_k^T d_{k-1}\right)$$
$$= \frac{\mu \|g_k\|^2}{\left(g_{k-1}^T d_{k-1}\right)^2}\left(g_{k-1} - g_k\right)^T d_{k-1} \tag{15}$$

By (15), (8), (12), (13) and (10), we obtain:

$$|\beta_k| \le \frac{\mu\gamma^2}{\omega^2\varepsilon^2}\|g_{k-1} - g_k\| \cdot \|d_{k-1}\| \le \frac{L\mu\gamma^2}{\omega^2\varepsilon^2}\|\alpha_{k-1} d_{k-1}\| \cdot \|d_{k-1}\| = \frac{L\mu\gamma^2}{\omega^2\varepsilon^2}\alpha_{k-1}\|d_{k-1}\|^2 \tag{16}$$

Where $\omega = (1 - \frac{1}{4\mu})$.

By the definition of d_k, we get from (16) and (12) that:

$$\|d_k\| = \|g_k\| + |\beta_k| \cdot \|d_{k-1}\| \le \gamma + \frac{L\mu\gamma^2\alpha_{k-1}\|d_{k-1}\|^2}{\omega^2\varepsilon^2}\|d_{k-1}\|.$$

Since $\lim_{k\to\infty}\alpha_k\|d_k\|^2 = 0$, there exists a constant $c \in (0,1)$ and an integer k_0, such that the following inequality holds for all $k \ge k_0$.

$$\frac{L\mu\gamma^2}{a^2\varepsilon^2}\alpha_{k-1}\|d_{k-1}\|^2 \le c$$

Hence, we have for all $k \ge k_0$.

$$\|d_k\| \le \gamma + c\|d_{k-1}\| \le \gamma(1 + c + c^2 + \cdots + c^{k-k_0-1}) + c^{k-k_0}\|d_{k_0}\| \le \frac{\gamma}{1-c} + \|d_{k_0}\|$$

Let $Q = \max\left\{\|d_1\|, \|d_2\|, \cdots, \|d_{k_0}\|, \frac{\gamma}{1-c} + \|d_{k_0}\|\right\}$, we have $\|d_k\| \le Q,\ \forall k \ge k_0$.

Theorem 2. Let the sequences $\{x_k\}$ be generated by Algorithm 1, then:

$$\lim_{k\to\infty}\inf\|g_k\| = 0 \tag{17}$$

Proof. We prove the result of this theorem by contradiction. Assume that the theorem is not true, there exists a constant $\varepsilon > 0$ such that:

$$\|g_k\| \ge \varepsilon,\ \forall k = 0,1,2,\ldots. \tag{18}$$

If $\lim_{k\to\infty}\inf \alpha_k > 0$, we get from (8) and (11) that $\lim_{k\to\infty}\inf \|g_k\| = 0$. This contradicts assumption (18).

If $\lim_{k\to\infty}\inf \alpha_k = 0$, there exists an infinite index set K such that:

$$\lim_{k\in K, k\to\infty} \alpha_k = 0 \cdot \quad (19)$$

By the step 3 of Algorithm 1, we have:

$$f(x_k + \rho^{-1}\alpha_k d_k) > f(x_k) - \delta(\rho^{-1}\alpha_k)^2 \|d_k\|^4 \quad (20)$$

By the mean-value theorem, (10) and (8), there is a constant $\tau \in (0,1)$ such that:

$$\begin{aligned} & f(x_k + \rho^{-1}\alpha_k d_k) - f(x_k) = \rho^{-1} g(x_k + \tau\rho^{-1}\alpha_k d_k)^T d_k \\ & = \rho^{-1}\alpha_k g_k^T d_k + \rho^{-1}\alpha_k [g(x_k + \tau\rho^{-1}\alpha_k d_k) - g(x_k)]^T d_k \\ & \le \rho^{-1}\alpha_k g_k^T d_k + L\rho^{-2}\alpha_k^2 \|d_k\|^2 \le -(1-\frac{1}{4\mu})\rho^{-1}\alpha_k \|g_k\|^2 + L\rho^{-2}\alpha_k^2 \|d_k\|^2 \end{aligned}$$

Substituting the inequality into (20), we get for all $k \in K$ sufficiently large,

$$(1-\frac{1}{4\mu})\|g_k\|^2 \le L\rho^{-1}\alpha_k \|d_k\|^2 + \delta\rho^{-1}\alpha_k \|d_k\|^4$$

With (14), we get:

$$(1-\frac{1}{4\mu})\|g_k\|^2 \le L\rho^{-1}\alpha_k \|d_k\|^2 + \delta\rho^{-1}Q^2\alpha_k \|d_k\|^2$$

Dividing both sides of this inequality by $(1-\frac{1}{4\mu}) > 0$, we get:

$$\|g_k\|^2 \le \frac{\rho^{-1}(L+\delta Q^2)}{1-\frac{1}{4\mu}}\alpha_k \|d_k\|^2 \cdot$$

Since $\lim_{k\to\infty} \alpha_k \|d_k\|^2 = 0$, the last inequality implies that $\lim_{k\in K, k\to\infty}\inf \|g_k\| = 0$. This also contradicts assumption (18). The contradiction shows that (17) is true.

4. Numerical Experiments

In this section, we report some results of the numerical experiments. We test Algorithm 1 and compare its performance with those of PRP method whose results be given by [19]. The problem that we tested are from [20]. Our line search subroutine computes α_k such that the Armojo type line search condition (7) holds with $\mu = \rho = 0.5$ and $\delta = 0.01$.

The termination condition is $\| g_k \| \le 10^{-5}$ or the iteration number exceeds 2×10^4 or the function evaluation number exceeds 3×10^5. In the following tables, each column has the following meanings:

Problem: the name of the problem;
Dim: the dimension of the problem;

NI: the number of iterations;

NF: the total number of function evaluations;

From the numerical results, we can see that the proposed method performs better than the PRP method for some problems. For some test problems, although the number of iterations of Algorithm 1 is more than the PRP method, the total number of function evaluations of Algorithm 1 is less than the PRP method. In summary, the numerical results show that Algorithm 1 is more efficient than the PRP method and provides an efficient method for solving uncontrained optimizaiton problems.

Table 1. Numerical Results

Problem	Dim	Algorithm 1 NI/NF	PRPSWP NI/NF
rose	2	36/118	29/502
helix	3	41/207	49/255
bard	3	26/89	23/98
Gulf	3	1/2	1/2
kowosb	4	77/251	62/361
biggs	6	125/350	121/495
osb2	11	141/528	293/1372
watson	20	616/1307	990/2773
vardim	50	8/51	10/52
Trig	100	68/241	46/342
Ie	500	4/9	6/13
Lin	1000	1/3	1/3

5. Conclusion

Different conjugate gradient algorithms correspond to different choices for the scalar parameter β_k. For classic conjugate gradient methods, the parameter β_k is selected so that when applied to minimize a strongly quadratic convex function, the directions d_k and d_{k+1} are conjugate subject to the Hessian of the quadratic function. Therefore, to minimize a convex quadratic function in a subspace spanned by a set of mutually conjugate directions is equivalent to minimize this function along each conjugate direction in turn. This is a very good idea. However, the directions in which the parameter β_k is computed by (6) does not satisfy the conjugacy condition in this pape.

References

[1] Asady B, Mansouri P, Gupta N. The modify version of artificial bee colony algorithm to solve real optimization problems. *International Journal of Electrical and Computer Engineering.* 2012; 2(4): 473-480.

[2] Seman K, Puspita FM, Taib BM, Shafii Z. An improved optimization model of internet charging scheme in muti service networks. *TELKOMNIKA.* 2012; 10(3): 592-598.

[3] Polak E, Ribière G. Note sur la convergence de methodes de directions conjugées. *Rev. Francaise Informat Recherche Opertionelle, 3e Année.* 1969; 16(1): 35- 43.

[4] Polyak BT. The conjugate gradient method in extremal problems. *USSR Comp. Math. and Math. Phys*. 1969; 9(4): 94-112.

[5] Hestenes MR, Stiefel EL. Method of conjugate gradients for solving linear systems. *J. Res. Nat. Bur. Stand.* 1952; 49(6): 409-432.

[6] Fletcher R, Reeves C. Function minimization by conjugate gradients. *J. Comput.* 1964; 7(2): 149-154.

[7] Dai YH, Yuan Y. A nonlinear conjugate gradient method with a strong global convergence property. *SIAM J. Optim.* 1999; 10(1): 177-182.

[8] Fletcher R. Practical Methods of Optimization. New York: John Wiley & Sons, 1987: 1. Unconstrained Optimization.

[9] Birgin E, Martinez JM. A spectral conjugate gradient method for unconstrained optimization. *App. Math. Optim.* 2001; 43(2): 117-128.

[10] Zhang L, Zhou WJ, Li DH. Global convergence of a modified Fletcheer-Reeves conjugate gradient method with Arnijo-type line search. *Numerische Mathematik*. 2006; 104(4): 561-572.

[11] Zhang J, Xiao Y, Wei Z. Nonlinear conjugate gradient methods with sufficient descent condition for large-scale unconstrained optimization. *Mathematical Problems in Engineering*. 2009, Article ID 243290. http://dx.doi.org/10.1155/2009/243290.
[12] Dai YH, Liao LZ. New conjugacy conditions and related nonlinear conjugate gradient methods. *Appl. Math. Optim.* 2001; 43(1): 87-101.
[13] Zhang L, Zhou W, Li DH. A descent modified Polak-Ribiere-Polyak conjugate gradient method and its global convergence. *IMA Journal of Numerical Analysis*. 2006; 26(4): 629-640.
[14] Cheng W. A two-term PRP-based descent method. *Numer. Funct. Anal. Optim.* 2007; 28(11-12): 1217-1230.
[15] Al-Bayati AY, Sharif WH. A new three-term conjugate gradient method for unconstrained optimization. *Can. J. Sci. Eng. Math*. 2010; 1: 108-124.
[16] Zhang L, Zhou W, Li DH. Some descent three-term conjugate gradient methods and their global convergence. *Optim. Methods software*. 2007; 22(4): 697-711.
[17] Li C, Fang L, Cao XL. Global convergence of a kind of conjugate gradient method. *TELKOMNIKA.* 2013; 11(1): 544-549.
[18] Li M, Feng H. A sufficient descent LS conjugate gradient method for unconstrained optimization problems. *Applied Mathematics and Computation.* 2011; 218(5): 1577-1586.
[19] Wei ZX, Yao SW, Liu LY. The convergence properties of some new conjugate gradient methods. *Applied Mathematics and Computation.* 2006; 183(2): 1341-1350.
[20] More JJ, Garbow BS, Hillstrom KE. Testing unconstrained optimization software. *ACM Trans. Math. Softw*. 1981; 7(1): 17-41.