Journal of Mathematical Sciences: Advances and Applications
Volume 7, Number 1, 2011, Pages 1-10

A MODIFIED CONJUGATE GRADIENT METHOD FOR UNCONSTRAINED OPTIMIZATION

SHUO WANG, ZHIBIN ZHU, QIAN QING and JUANJUAN HU

Department of Computational Science & Mathematics
Guilin University of Electronic Technology
Guilin, 541004
Guangxi
P. R. China
e-mail: shuo1987@gmail.com

Abstract

In this paper, based on a new quasi-Newton equation and the conjugacy condition, we propose a modified conjugate gradient method. Under some suitable conditions, it is proven that the method converges globally for general functions.

1. Introduction

The nonlinear conjugate gradient method is designed to solve the following unconstrained optimization problem:

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

2010 Mathematics Subject Classification: 49M37, 90C30, 65K05.

Keywords and phrases: conjugate gradient method, the strong Wolfe line search, global convergence.

This work was supported in part by NNSF(No.11061011) of China and Program for Excellent Talents in Guangxi Higher Education Institutions ([2009]156).

Received October 14, 2010; Revised December 9, 2010

where f is smooth and its gradient g is available. The nonlinear conjugate gradient method is very useful for solving (1.1), especially when n is large, and has the following form:

$$x_{k+1} = x_k + \alpha_k d_k, \tag{1.2}$$

and

$$d_k = \begin{cases} -g_k, & k = 1, \\ -g_k + \beta_k d_{k-1}, & k \geq 2, \end{cases} \tag{1.3}$$

where $\alpha_k > 0$ is step-length, β_k is a scalar, and g_k denotes $\nabla f(x_k)$.

The PRP, HS, and LS methods are three well-known conjugate gradient methods [2, 4], and other conjugate gradient methods can be found [1, 6] et al..

Recently, based on the memoryless BFGS quasi-Newton method, [6] propose a new LS type method, which converges globally for general functions with the Grippo-Lucidi line search.

In this paper, based on a new quasi-Newton equation proposed in [5], we establish a new conjugate gradient method, by modifying the Liu-Storey type nonlinear conjugate gradient method and the conjugacy condition, which derived from [1]. Under some suitable properties, it is proven that the global convergence of the method for general functions with the strong Wolfe line search.

2. A Modified Conjugate Gradient Method

In [5], a new quasi-Newton equation was proposed for the problem (1.1):

$$B_{k+1}s_k = y_k^*, \quad y_k^* = y_k + \frac{\gamma_k}{\|s_k\|^2} s_k, \tag{2.1}$$

where B_{k+1} is the Hessian approximation of $f(x)$, $s_{k+1} = x_{k+1} - x_k = \alpha_k d_k$, $y_k = g_{k+1} - g_k$,

$$\gamma_k = 2(f_k - f_{k+1}) + (g_k + g_{k+1})^\top s_k. \tag{2.2}$$

We give our new conjugacy condition by using this new quasi-Newton equation. Let H_{k+1} be the inverse Hessian approximation of $f(x)$. Since B_{k+1} satisfies the new quasi-Newton equation: $B_{k+1}s_k = y_k^*$, $y_k^* = y_k + \frac{\gamma_k}{\|s_k\|^2} s_k$, and the search direction is calculated as following:

$$d_{k+1} = -H_{k+1}g_{k+1}.$$

We obtain that

$$d_{k+1}^\top y_k^* = d_{k+1}^\top (B_{k+1}s_k) = -g_{k+1}^\top s_k, \tag{2.3}$$

where y_k^* is defined as (2.1). Finally, taking (2.3) into consideration and adding the parameter t, we achieve our new conjugacy condition:

$$d_{k+1}^\top y_k^* = -tg_{k+1}^\top s_k. \tag{2.4}$$

Using the new conjugacy condition (2.4), we are now ready to give some new nonlinear conjugate gradient methods. Multiplying the search direction with y_k in [6] and using (2.4), we obtain the following formula for computing:

$$d_{k+1}^* \triangleq -g_{k+1} + \beta_{k+1}^* d_k^* + \theta_{k+1}(y_k^* - s_k). \tag{2.5}$$

We obtain that

$$(\beta_{k+1}^0)^* \triangleq \frac{g_{k+1}^\top y_k^*}{-(d_k^*)^\top g_k} - t\frac{\|y_k^*\|^2 g_{k+1}^\top d_k^*}{|(d_k^*)^\top g_k|^2},$$

$$\theta_{k+1} \triangleq \frac{g_{k+1}^\top d_k^*}{-(d_k^*)^\top g_k}, \tag{2.6}$$

where y_k^* is defined as (2.1). It can be seen that if the objective function f is uniformly convex, then we have

$$(d_k^*)^\top y_k^* = (d_k^*)^\top \left(y_k + \frac{2(f_k - f_{k+1}) + (g_k + g_{k+1})^\top s_k}{\|s_k\|^2} s_k \right)$$

$$= (d_k^*)^\top (g_{k+1} - g_k) + 2\alpha_{k+1}^{-1}(f_k - f_{k+1}) + (g_k + g_{k+1})^\top d_k^*$$

$$= 2(d_k^*)^\top g_{k+1} + 2\alpha_{k+1}^{-1}(f_k - f_{k+1}) > 0,$$

where the last inequality is from the uniform convexity of f. However, if f is a general function, then $(d_k^*)^\top y_k^*$ may equal to 0. In order to overcome this problem, we make the following modification. Define $\tilde{y}_k^* = y_k + \frac{\max\{\gamma_k, 0\}}{\|s_k\|^2} s_k$.

$$(\beta_{k+1}^1)^* \triangleq \frac{g_{k+1}^\top \tilde{y}_k^*}{-(d_k^*)^\top g_k} - t\frac{\|\tilde{y}_k^*\|^2 g_{k+1}^\top d_k^*}{|(d_k^*)^\top g_k|^2}.$$

Using this modification, when the strong Wolfe line search is used and the descent condition holds, $(\beta_{k+1}^1)^*$ is well-defined since we always have

$$(d_k^*)^\top \tilde{y}_k^* \geq (d_k^*)^\top y_k \geq -(1-\delta) g_k^\top d_k^* > 0. \tag{2.7}$$

In addition, following Dai and Liao's approach [1], we also give the following formula:

$$(\beta_{k+1}^2)^* \triangleq \max\{\frac{g_{k+1}^\top \tilde{y}_k^*}{-(d_k^*)^\top g_k}, 0\} - t\frac{\|\tilde{y}_k^*\|^2 g_{k+1}^\top d_k^*}{|(d_k^*)^\top g_k|^2}. \tag{2.8}$$

Algorithm:

Step 0. Given constants $0 < \delta < \sigma < 1$, $t > 1$. Chosen an initial point $x_1 \in R^n$. Let $k = 1$.

Step 1. Compute d_k^* by $d_1 = -g_1$, and for $k > 1$,

$$d_k^* = \begin{cases} -g_k + (\beta_k^2)^* d_{k-1}^* + \theta_k (y_k^* - s_{k-1}), \text{ if } (\beta_k^2)^* \geq 0, \\ -g_k, \text{ otherwise,} \end{cases} \tag{2.9}$$

where $(\beta_k^2)^*$ and θ_k are defined by (2.8) and (2.6), respectively.

Step 2. Determine α_k by the strong Wolfe line search (3.4).

Step 3. Let the next iterate be $x_{k+1} = x_k + \alpha_k d_k^*$.

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

3. Convergence Analysis

In this section, we will discuss the global convergence of the method (2.8). We show that the conjugate gradient method with (2.8) is globally convergent for general function.

Throughout this section, we assume that $g_k \neq 0$, for all $k \geq 1$.

The following basic assumptions on the objective function will be used in this section:

H 3.1. (1) The level set $\Gamma = \{x \mid f(x) \leq f(x_1)\}$ is bounded, namely, there exists a constant $B > 0$ such that $\|x\| \leq B$, for all $x \in \Gamma$.

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

$$\|\nabla f(x) - \nabla f(\tilde{x})\| \leq L\|x - \tilde{x}\|, \text{ for all } x, \tilde{x} \in N. \tag{3.1}$$

Under the above assumptions on f, there exists a constant $\gamma \geq 0$ such that

$$\|\nabla f(x)\| \leq \gamma, \text{ for all } x \in \Gamma. \tag{3.2}$$

We state the following two properties, which play important roles in our later analysis. We say the descent condition holds, if for each search direction d_k^* : $g_k^\top d_k^* \leq 0$, for all $k \geq 1$.

In addition, we say the sufficient descent condition holds, if there exists a constant $c > 0$ such that for each search direction d_k^*, we have

$$g_k^\top d_k^* \leq -c\|g_k\|^2, \text{ for all } k \geq 1. \tag{3.3}$$

The step-length α_k in (1.2) is obtained by some line search scheme. In conjugate gradient methods, the strong Wolfe conditions, namely,

$$f(x_k + \alpha_k d_k^*) - f_k \leq \delta\alpha_k g_k^\top d_k^*, \quad |g(x_k + \alpha_k d_k^*)^\top d_k^*| \leq -\sigma g_k^\top d_k^*, \quad (3.4)$$

where $0 < \delta < \sigma < 1$, are often imposed on the line search (in this case, we call the line search the strong Wolfe line search).

For any conjugate gradient method with the strong Wolfe line search, we have the following Lemmas 3.1 and 3.2.

Lemma 3.1. *Let* α_k *be obtained by the strong Wolfe line search. Suppose that* H 3.1 *and the descent condition hold. Then, we have*

$$\sum_{k=1}^{\infty} -\alpha_k g_k^\top d_k^* < \infty. \quad (3.5)$$

Lemma 3.2. *Consider any conjugate gradient method in the form* (1.2)-(1.3), *where* α_k *is obtained by the strong Wolfe line search. Suppose that* H 3.1 *and the descent condition hold. Then, the following so-called Zoutendijk condition holds*:

$$\sum_{k=1}^{\infty} \frac{(g_k^\top d_k^*)^2}{\|d_k^*\|^2} < \infty. \quad (3.6)$$

Lemma 3.3. *Let* $\{x_k\}$ *and* $\{d_k^*\}$ *be generated by the new conjugate gradient method, then we have* $(d_{k+1}^*)^\top g_{k+1} \leq -(1 - \frac{1}{t})\ \|g_{k+1}\|^2$.

Proof. If $g_{k+1}^\top \tilde{y}_k^* > 0$, then $(\beta_{k+1}^2)^* \triangleq \frac{g_{k+1}^\top \tilde{y}_k^*}{-(d_k^*)^\top g_k} - t\frac{\|\tilde{y}_k^*\|^2 g_{k+1}^\top d_k^*}{|(d_k^*)^\top g_k|^2}$.

So,

$$(d_{k+1}^*)^\top g_{k+1} = -\|g_{k+1}\|^2 + 2\frac{g_{k+1}^\top \tilde{y}_k^* g_{k+1}^\top d_k^*}{-(d_k^*)^\top g_k} - t\frac{\|\tilde{y}_k^*\|^2 g_{k+1}^\top d_k^*}{|(d_k^*)^\top g_k|^2} - \frac{\alpha_k (g_{k+1}^\top d_k^*)^2}{-(d_k^*)^\top g_k}$$

$$\leq -\|g_{k+1}\|^2 + 2(\frac{1}{\sqrt{t}} g_{k+1})^\top (\frac{\sqrt{t} g_{k+1}^\top d_k^*}{-(d_k^*)^\top g_k} \tilde{y}_k^*) - t\frac{\|\tilde{y}_k^*\|^2 g_{k+1}^\top d_k^*}{|(d_k^*)^\top g_k|^2}$$

$$= -(1 - \frac{1}{t})\|g_{k+1}\|^2.$$

If $g_{k+1}^\top \tilde{y}_k^* < 0$, then $(\beta_{k+1}^2)^* \triangleq -t\frac{\|\tilde{y}_k^*\|^2 g_{k+1}^\top d_k^*}{|(d_k^*)^\top g_k|^2}$. So,

$$(d_{k+1}^*)^\top g_{k+1} = -\|g_{k+1}\|^2 + \frac{g_{k+1}^\top \tilde{y}_k^* g_{k+1}^\top d_k^*}{-(d_k^*)^\top g_k} - t\frac{\|\tilde{y}_k^*\|^2 g_{k+1}^\top d_k^*}{|(d_k^*)^\top g_k|^2} - \frac{\alpha_k (g_{k+1}^\top d_k^*)^2}{-(d_k^*)^\top g_k}$$

$$\leq -\|g_{k+1}\|^2 + \frac{1}{4t}\|g_{k+1}\|^2 + t\frac{\|\tilde{y}_k^*\|^2 (g_{k+1}^\top d_k^*)}{|(d_k^*)^\top g_k|^2} - t\frac{\|\tilde{y}_k^*\|^2 g_{k+1}^\top d_k^*}{|(d_k^*)^\top g_k|^2}$$

$$= -(1 - \frac{1}{4t})\|g_{k+1}\|^2$$

$$\leq -(1 - \frac{1}{t})\|g_{k+1}\|^2,$$

which shows that the conclusion holds by induction since $d_1 = -g_1$ and $t > 1$. This completes the proof. □

Lemma 3.4. *Consider the conjugate gradient method in the form* (1.2)-(1.3), *where* β_k *is calculated by* (2.8) *and* α_k *is obtained by the strong Wolfe line search. Suppose that* H 3.1 *and the sufficient descent condition* (3.3) *hold. Suppose that there exists a constant* $\gamma_1 > 0$ *such that* $\|g_{k+1}\| \geq \gamma_1$, *for all* $k \geq 0$. *Then, we have* $d_k \neq 0$ *and* $\sum_{k=1}^{\infty}\|\mu_{k+1} - \mu_k\|^2 < \infty$, $\mu_{k+1} = \frac{d_{k+1}}{\|d_{k+1}\|}$.

Proof. We only need to consider the case $(\beta_{k+1}^2)^* \geq 0$. We rewrite (2.9) as follows:

$$d_{k+1}^* = -g_{k+1} + (\beta_{k+1}^2)^* d_k^* + \theta_{k+1}(y_k^* - s_k) \triangleq \nu_{k+1} + (\beta_{k+1}^2)^* d_k^*. \quad (3.7)$$

Now, we define

$$\gamma_{k+1} \triangleq \frac{\nu_{k+1}}{\|d_{k+1}^*\|} \text{ and } \delta_{k+1} \triangleq \frac{(\beta_{k+1}^2)^*\|d_k^*\|}{\|d_{k+1}^*\|}. \quad (3.8)$$

Then by (3.7), we have for all $k > 1$,

$$\mu_{k+1} = \gamma_{k+1} + \delta_{k+1}\mu_k. \tag{3.9}$$

Using the identity $\|\mu_{k+1}\| = \|\mu_k\| = 1$ and (3.9), we have

$$\|\gamma_{k+1}\| = \|\mu_{k+1} - \delta_{k+1}\mu_k\| = \|\delta_{k+1}\mu_{k+1} - \mu_k\|. \tag{3.10}$$

Using the condition $\delta_{k+1} \geq 0$, the triangle inequality, and (3.10), we obtain

$$\|\mu_{k+1} - \mu_k\| \leq \|\mu_{k+1} - \delta_{k+1}\mu_k\| + \|\delta_{k+1}\mu_{k+1} - \mu_k\| = 2\|\gamma_{k+1}\|. \tag{3.11}$$

It follows from (3.8), (2.6), (3.4), and (3.2) that there exists a constant $M_1 \neq 0$ such that

$$\|\nu_{k+1}\| \leq \|g_{k+1}\| + \sigma(\|g_{k+1} - g_k\| + \|x_{k+1} - x_k\|) \leq \gamma + 2\sigma(\gamma + B) \triangleq M_1. \tag{3.12}$$

So, we have from (3.7), (3.12), (3.2), and (3.6) that

$$\sum_{k=0}^{\infty}\|\gamma_{k+1}\|^2 = \sum_{k=0}^{\infty}\frac{\|\nu_{k+1}\|^2}{\|d_{k+1}^*\|^2} \leq \sum_{k=0}^{\infty}\frac{M_1^2}{\|d_{k+1}^*\|^2} \leq \frac{M_1^2}{\gamma_1^4}\sum_{k=0}^{\infty}\frac{\|g_{k+1}\|^4}{\|d_{k+1}^*\|^2} < \infty,$$

which together with (3.9) completes our proof. □

Let N^* denote the set of positive integers. For $\lambda > 0$ and a positive integer Δ, we define the set of index: $K_{k,\Delta}^{\lambda} = \{i \in N^* | k \leq i \leq k + \Delta - 1, \|s_{i-1}\| > \lambda\}$.

Lemma 3.5. *Let $\{x_k\}$ and $\{d_k\}$ be generated by the Algorithm. Then, if $\|g_{k+1}\| \geq \gamma_1$, for all $k \geq 0$ holds, there exists $\lambda > 0$ such that, for any $\Delta \in N^*$ and any index k_0, there is an index $k > k_0$ such that $|K_{k,\Delta}^{\lambda}| > \frac{\Delta}{2}$.*

Proof. We proceed by contradiction. Suppose that for any $\lambda > 0$, there exist $\Delta \in N^*$ and k_0 such that

$$|K^{\lambda}_{k,\Delta}| \le \frac{\Delta}{2}, \quad \text{for all} \quad k \ge k_0. \tag{3.13}$$

Let $b > 1$ and $\lambda > 0$ such that for all $k, |\beta_k| \le b$ and $\|s_{k-1}\| \le \lambda \Rightarrow |\beta_k| \le \frac{1}{2b}$. For $\lambda > 0$, we choose Δ and k_0 such that (3.13) holds. For $l \ge k_0 + 1$, we have from (3.7) and (3.12) that $\|d^*_l\|^2 \le (\|\nu_1\| + |(\beta^2_l)^*|\|d^*_{l-1}\|)^2 \le 2M_1^2 + 2((\beta^2_l)^*)^2\|d^*_{l-1}\|^2$. Now, the remaining argument is standard in the same way as Lemma 4.2 in [2]. So, we omit it here. The proof is then completed. □

Theorem 3.6. *Let* $\{x_k\}$ *and* $\{d^*_k\}$ *be generated by Algorithm. Then*, $\lim_{k\to\infty} \inf\|g_k\| = 0$.

4. Numerical Experiments

We carry out some numerical experiments. This algorithm has been tested on some problems from [3]. The results are summarized in Table 1. For each test problem, No. is the number of the test problem in [3], x_0 is the initial point, x_k is the final point, and FV is the final value of the objective function.

Table 1. The detail information of numerical experiments

No.	x_0	x_k	$\|g_k\|$	FV
*s*201	(8, 9)	(5, 6)	1.907348633*r* – 006	9.094947018*e* – 013
*s*206	(–1.2, 1)	(1, 1.0004)	7.964932417*e* – 006	1.410015189*e* – 009
*s*207	(–1.2, 1)	(0.9999891, 0.999736)	8.8142009*e* – 006	1.399722747*e* – 008
*s*309	(0, 0)	(3.48268, 3.9)	9.6410967*e* – 006	– 3.9871708
*s*208	(–1.2, 1)	(0.999903, 0.999805)	9.725340738*e* – 006	9.451272023*e* – 009
*s*212	(2, 0)	(0.141906, – 0.141907)	8.311214097*e* – 006	4.898648281*e* – 012
*s*311	(1, 1)	(3, 2)	0	0

References

[1] Y. Dai and L. Z. Liao, New conjugacy conditions and related nonlinear conjugate gradient methods, Appl. Math. Optim. 43 (2001), 87-101.

[2] J. C. Gilbert and J. Nocedal, Global convergence properties of conjugate gradient methods for optimization, SIAM J. Optim. 2 (1992), 21-42.

[3] W. Hock and K. Schittkowski, Test Examples for Nonlinear Programming Codes, Lecture Notes in Economics and Mathematical Systems 187, Springer, Berlin, 1981.

[4] J. J. Moreau, Proximite et dualite dans un espace Hilbertien, Bull. Soc. Math. France 93 (1965), 273-299.

[5] Z. Wei, G. Li and L. Qi, New quasi-Newton methods for unconstrain optimization, preprint.

[6] Li Zhang, A new Liu-Storey type nonlinear conjugate gradient method for unconstrained optimization problems, J. Comput. Appl. Math. 225 (2009), 146-157.

■