ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 9, No. 2, 2014, *pp*.091-096

A new conjugate gradient method based on the modified secant equations

Weijuan Shi [1], Chunyan Hu [2 +], Benxin Zhang[1]
[1] School of Mathematics and Computing Science,
Guilin University of Electronic Technology, Guilin, 541004, P. R. China
[2] School of Electronic Engineering and Automation,
Guilin University of Electronic Technology, Guilin, 541004, P. R. China
(*Received Frbruary 28, 2013, accepted October 28, 2013*)

Abstract. Based on the secant equations proposed by Zhang, Deng and Chen, we propose a new nonlinear conjugate gradient method for unconstrained optimization problems. Global convergence of this method is established under some proper conditions.

Keywords: Unconstrained optimization problem, modified secant equations, conjugate gradient method, global convergence.

1. Introduction

Conjugate gradient method plays a specific role in solving large-scale nonlinear minimization problems. Some good references of the conjugate gradient methods can be found in many research. In the past few years, many efforts have been made to research new conjugate gradient methods which process not only the global convergence property for general functions but also good numerical performances. Many of these new conjugate gradient methods are based on the secant equations ([10,11]).

In this paper, we present a conjugate gradient method to solve the following unconstrained optimization problem

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

where $f: R^n \to R$ is a smooth nonlinear function.

The iterative formula is

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

where α_k is the step size which is computed by some line search([1]). The search direction d_k is defined by

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

where β_k is a scalar and $g_k = \nabla f(x)$ is the gradient of $f(x)$.

Remark Here, we use the strong Wolf line search condition (see[1]), that is, the step size α_k satisfies

$$\begin{aligned} & f(x_k + \alpha d_k) - f(x_k) \le \delta \alpha_k g_k^T d_k, \\ & \left| g(x_k + \alpha d_k)^T d_k \right| \le -\sigma g_k^T d_k, \end{aligned} \tag{4}$$

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

The gradient method proposed in this paper is based on a modified secant equations. Therefore, let us introduce shortly the secant equations firstly.

Note $G_k = \nabla^2 f(x_k)$, the matrix B_k is the approximation of G_k.

The secant equations is defined as follows:

$$B_{k+1} s_k = y_k, \tag{5}$$

where $s_k = x_{k+1} - x_k$, $y_k = g_{k+1} - g_k$. (5)sometimes is said to be the standard secant equations.

+ Corresponding author. Tel.: +86-0773-2291522; fax: +86-0773-2291522.
E-mail address: huchyel@hotmail.com.

Definition 1 A twice continuously differentiable function f is said to be uniformly convex on the nonempty open convex set S if and only if there exists some constant $M > 0$ such that

$$(g(x) - g(y))^T (x - y) \geq M\|x - y\|^2, \ \forall x, y \in S.$$

2. The new conjugate gradient method and the global convergence

In this paper we assume $g_k \neq 0,$ otherwise the current point x_k is the stationary point of the problem (1).

Definition 2 It is said that the descent condition holds for the conjugate gradient method, if

$$g_k^T d_k < 0, \ \forall k \geq 1,$$

and sufficient descent condition holds for the conjugate gradient method, if

$$g_k^T d_k < -c\|g_k\|^2, \ \forall k \geq 1,$$

To discuss the validity and the global convergence of the new conjugate gradient method, we make the following basic assumptions for the objective function.

H 1 The level set $L = \{x \mid f(x) < f(x_1)\}$ is bounded, that is, there exists a constant B such that

$$\|x\| \leq B, \ \forall x \in L. \tag{6}$$

H 2 In some neighborhood N of L, $L \subseteq N$, the function f is continuously differentiable, its gradient is Lipschitz continuous, that is there exists a constant L such that

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \ \forall x \in N. \tag{7}$$

The following proposition is obtained by [5] directly.

Proposition 1 If f the satisfies assumptions H 1 and H 2, then there exists a constant $\gamma > 0$ such that

$$\|\nabla f(x)\| \leq \gamma, \ \forall x \in L. \tag{8}$$

For any conjugate gradient method which satisfies the strong Wolf line search conditions, we have the following general conclusions, see[3].

Lemma 1 Assume that the assumptions H 1 and H 2 hold. For the conjugate gradient method (2)-(3) , where d_k is the descent direction, α_k is computed by the strong Wolf line search conditions, if

$$\sum_{K \geq 1} \frac{1}{\|d_k\|^2} = \infty, \tag{9}$$

then

$$\lim_{k \to \infty} \|g_k\| = 0. \tag{10}$$

Firstly, let us describe the modified secant equations in [6,7], it has better theory properties than the standard secant equations ([8]). Based on the modified secant equations, the relevant conjugate gradient method is proposed, and then the global convergence is discussed.

Supposed that the objective function f is smooth enough. We make its Taylor expansion at point $x_{k-1} = x_k - s_{k-1},$

$$f_{k-1} = f_k - s_{k-1}^T g_k + \frac{1}{2} s_{k-1}^T G_k s_{k-1} - \frac{1}{6} s_{k-1}^T (T_k s_{k-1}) s_{k-1} + O(\|s_{k-1}\|^4),$$

$$s_{k-1}^T g_{k-1} = s_{k-1}^T g_k - s_{k-1}^T G_k s_{k-1} + \frac{1}{2} s_{k-1}^T (T_k s_{k-1}) s_{k-1} + O(\|s_{k-1}\|^4),$$

where

$$s_{k-1}^T (T_k s_{k-1}) s_{k-1} = \sum_{i,j,l=1}^{n} \frac{\partial^3 f(x_k)}{\partial x^i \partial x^j \partial x^l} s_{k-1}^i s_{k-1}^j s_{k-1}^l, \tag{11}$$

The formula can be written as (see Zhang and Xu [7])

$$s_{k-1}^T G s_{k-1} = s_{k-1}^T y_{k-1} + \theta_{k-1}, \tag{12}$$

where

$$\theta_{k-1} = 6(f_{k-1} - f_k) + 3(g_{k-1} - g_k)^T s_{k-1}. \tag{13}$$

Recently, Yabe and Takano([5]) considered the following modified secant equations by embedding a parameter ρ,

$$B_k s_{k-1} = z_{k-1},\quad z_{k-1} = y_{k-1} + \rho \frac{\theta_{k-1}}{s_{k-1}^T u} u, \tag{14}$$

where $u \in R^n$ is any vector which satisfies $s_{k-1}^T u \neq 0$. And the following conjugacy condition was represented:

$$d_k^T z_{k-1} = -t g_k^T s_{k-1} (t \geq 0). \tag{15}$$

They obtained a new formula about β_k as:

$$\beta_k^{YT}(t) = \frac{g_k^T z_{k-1} - t g_k^T s_{k-1}}{d_{k-1}^T z_{k-1}} (t \geq 0). \tag{16}$$

Based on the formulae (14)-(16), we get the extension about the equation (14):

$$B_k s_{k-1} = \overline{y}_{k-1}^*,\quad \overline{y}_{k-1}^* = y_{k-1} + \rho_{k-1} \frac{|\theta_{k-1}|}{s_{k-1}^T y_{k-1}} y_{k-1} + (1-\rho_{k-1}) \frac{|\theta_{k-1}|}{3 s_{k-1}^T s_{k-1}} s_{k-1}, \tag{17}$$

where $\rho_{k-1} \in [0,1]$.

The new conjugacy condition is established as follows:

$$d_k^T \overline{y}_{k-1}^* = -t g_k^T s_{k-1}, (t \geq 0). \tag{18}$$

A new β_k is obtained by (3):

$$\overline{\beta}_k(t) = \frac{g_k^T \overline{y}_{k-1}^* - t g_k^T s_{k-1}}{d_{k-1}^T \overline{y}_{k-1}^*} (t \geq 0). \tag{19}$$

Obviously, we have the following result.

$$d_{k-1}^T \overline{y}_{k-1}^* \geq d_{k-1}^T y_{k-1} \geq -(1-\sigma) d_{k-1}^T g_{k-1}. \tag{20}$$

Theorem 1 If $d_{k-1}^T \overline{y}_{k-1}^* \neq 0$, $d_k = -g_k + \overline{\beta}_k(t) d_{k-1}$, $\overline{\beta}_k(t)$ is defined by (19), then $|g_k^T d_k| \geq c_1 \|g_k\|^2$, $c_1 = \min\left\{1, \left|1 - \frac{Q}{M}\right|\right\}$, where $Q = 2L + \frac{3L^2}{M} + t > 0$.

Proof. When $k = 1$, $|g_k^T d_k| = \|g_k\|^2$.

When $k \geq 2$,

$$\begin{aligned} |g_k^T d_k| &= \left|g_k^T\left(-g_k + \overline{\beta}_k(t) d_{k-1}\right)\right| \\ &\geq \left|\|g_k\|^2 - \left|\overline{\beta}_k(t)\right| \left|g_k^T d_{k-1}\right|\right| \\ &= \left|\|g_k\|^2 - \left|\frac{g_k^T \overline{y}_{k-1}^* - t g_k^T s_{k-1}}{d_{k-1}^T \overline{y}_{k-1}^*}\right| \left|g_k^T d_{k-1}\right|\right| \\ &\geq \left|\|g_k\|^2 - \left|\frac{\left(2L + \frac{3L^2}{M} + t\right)\|g_k\|\|s_{k-1}\|}{M \alpha_{k-1}^{-1} \|s_{k-1}\|^2}\right| \left|g_k^T d_{k-1}\right|\right| \\ &\geq \left|\|g_k\|^2 - \left|\frac{Q\|g_k\|^2}{M}\right|\right| \\ &= \left|1 - \frac{Q}{M}\right| \|g_k\|^2. \end{aligned}$$

Thus $|g_k^T d_k| \geq \min\left\{\|g_k\|^2, \left|1 - \frac{Q}{M}\right| \|g_k\|^2\right\} = c_1 \|g_k\|^2$.

Theorem 2 If the assumptions $H1$ and $H2$ hold, then there exists $\alpha_k \geq \frac{c_1(1-\sigma)}{L}\frac{\|g_k\|^2}{\|d_k\|^2} > 0$ satisfies (4).

Proof. By the assumption $H2$ and the strong wolf condition, we have that $g_{k+1}^T d_k \geq \sigma g_k^T d_k$.
So

$$(1-\sigma)\left|g_k^T d_k\right| = (\sigma - 1) g_k^T d_k \leq (g_{k+1} - g_k)^T d_k \leq \|g_{k+1} - g_k\| \cdot \|d_k\| \leq L\alpha_k \|d_k\|^2,$$

that is

$$\alpha_k \geq \frac{(1-\sigma)\left|g_k^T d_k\right|}{L\|d_k\|^2} \geq \frac{(1-\sigma)c_1}{L}\frac{\|g_k\|^2}{\|d_k\|^2} > 0.$$

The proof is completed.

Algorithm

Step 0 Given a starting point x_0 and the precision $\varepsilon > 0$.

Step 1 Choose $d_0 = -\nabla f(x_0)$, if $\|\nabla f(x_0)\| \leq \varepsilon$, stop. Otherwise, set $k = 0$, go to Step 2.

Step 2 Solve α_k, such that $f(x_k + \alpha_k d_k) - f(x_k) \geq \delta \alpha_k g_k^T d_k$, and $\left|g(x_k + \alpha_k d_k)^T d_k\right| \leq -\sigma g_k^T d_k$, set $x_{k+1} = x_k + \alpha_k d_k$ go to Step 3.

Step 3 If $\|\nabla f(x_{k+1})\| \leq \varepsilon$, stop, the minimal point is x_{k+1}. Otherwise go to Step 4.

Step 4 Set $d_{k+1} = -\nabla f(x_{k+1}) + \overline{\beta}_{k+1}(t) d_k$, $\overline{\beta}_{k+1}(t)$ is defined by (19). Set $k = k+1$, go to Step 2.

The following conclusion is obtained by the above algorithm.

Lemma 2 If the assumptions $H1$ and $H2$ hold, θ_{k-1} is defined by (13). Then

$$|\theta_{k-1}| \leq 3L\|s_{k-1}\|^2, \tag{21}$$

where L is defined by (7).

Corollary 1 If the assumptions $H1$ and $H2$ hold, $\overline{y}_{k-1}^*$ is defined by (17). Then

$$\left\|\overline{y}_{k-1}^*\right\| \leq \left(2 + \frac{3L}{M}\right) L\|s_{k-1}\|. \tag{22}$$

Proof. By Definition 1 we have

$$d_{k-1}^T \overline{y}_{k-1}^* \geq d_{k-1}^T y_{k-1} \geq M\alpha_{k-1}^{-1}\|s_{k-1}\|^2. \tag{23}$$

Considering Lemma 2, If the assumptions $H1$ and $H2$ hold, $\rho_{k-1} \in [0,1]$, we have

$$\begin{aligned}\left\|\overline{y}_{k-1}^*\right\| &\leq \|y_{k-1}\| + \rho_{k-1}\frac{|\theta_{k-1}|}{\left|s_{k-1}^T y_{k-1}\right|}\|y_{k-1}\| + (1-\rho_{k-1})\frac{|\theta_{k-1}|}{3\left|s_{k-1}^T s_{k-1}\right|}\|s_{k-1}\| \\ &\leq L\|s_{k-1}\| + \frac{3L\|s_{k-1}\|^2}{M\|s_{k-1}\|^2}\|y_{k-1}\| + \frac{L\|s_{k-1}\|^2}{\|s_{k-1}\|^2}\|s_{k-1}\| \\ &\leq \left(2 + \frac{3L}{M}\right) L\|s_{k-1}\|.\end{aligned}$$

The proof is completed.

Now, for the uniformly convex function, we consider the global convergence of the new conjugate gradient method with $\beta_k = \overline{\beta}_k(t)$.

Theorem 3 If the assumptions $H1$, $H2$ and the descent condition hold, choose $\beta_k = \overline{\beta}_k(t)$ in the conjugate gradient method which is defined by (2), (3), where α_k is computed by the strong Wolfe line search condition that is defined by (4). If the objective function is uniformly convex on L (L is introduced by (6), then $\lim_{k\to\infty}\|g_k\| = 0.$

Proof. Due to the descent condition, it hold that $d_k \neq 0$. By Lemma 1, it is enough to prove that $\|d_k\|$ is bounded. By Corollary 1 and Proposition 1, we have

$$\begin{aligned} \left|g_k^T \overline{y}^*_{k-1} - t g_k^T s_{k-1}\right| &\le \left|g_k^T \overline{y}^*_{k-1}\right| + \left|t g_k^T s_{k-1}\right| \\ &\le \|g_k\|\left\|\overline{y}^*_{k-1}\right\| + t\|g_k\|\|s_{k-1}\| \\ &\le \left(2 + \frac{3L}{M}\right)L\gamma\|s_{k-1}\| + t\gamma\|s_{k-1}\| \\ &\le \left(2L + \frac{3L^2}{M} + t\right)\gamma\|s_{k-1}\|. \end{aligned} \tag{24}$$

By (3), (23) and (24) we have

$$\begin{aligned} \|d_k\| &= \left\|- g_k + \beta_k(t)d_{k-1}\right\| \le \|g_k\| + \left|\beta_k(t)\right|\|d_{k-1}\| \\ &= \|g_k\| + \frac{\left|g_k^T \overline{y}^*_{k-1} - t g_k^T s_{k-1}\right|}{\left|d_{k-1}^T \overline{y}^*_{k-1}\right|}\|d_{k-1}\| \\ &\le \gamma + \frac{\left(2L + \frac{3L^2}{M} + t\right)\gamma\|s_{k-1}\|}{M\alpha_{k-1}^T\|s_{k-1}\|^2}\|d_{k-1}\| \\ &\le \gamma + \frac{\left(2L + \frac{3L^2}{M} + t\right)}{M}\gamma. \end{aligned}$$

The proof is completed.

3. Conclusions

We have introduced a nonlinear conjugate gradient methods for unconstrained optimization problems which is based on a modified secant equation proposed by Zhang, Deng and Chen. We have proved that, under some proper conditions, the proposed method is globally convergent for general functions.

Acknowledgements

This work was supported in part by NNSF (No.11061011, 11361018) of China and Guangxi Fund for Distinguished Young Scholars(2012GXSFFA060003).

4. References

[1] J.Nocedal, S.J.Wright, *Numerical Optimization*, Springer, New York, 1999.

[2] W. Sun, Y.X. Yuan, *Optimization Theory and Methods:* Nonlinear Programming, Springer, 2006.

[3] Y.H. Dai, J.Y. Han, G.H. Liu, D.F. Sun, H.X. Yin, Y.X. Yuan, *Convergence properties of nonlinear conjugate gradient methods*, SIAM J. Optim. 10 (2) (1999)348-358.

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

[5] H. Yabe, M. Takano, *Global convergence properties of nonlinear conjugate gradient methods with modified secant condition*, Comput. Optim. Appl. 28 (2004) 203-225.

[6] J.Z. Zhang, N.Y. Deng, L.H. Chen, *New quasi-Newton equation and related methods for unconstrained optimization*, J. Optim. Theory Appl. 102 (1999)147-167.

[7] J.Z. Zhang, C.X. Xu, *Properties and numerical performance of quasi-Newton methods with modified quasi-Newton equations*, J. Comput. Appl. Math.137 (2001) 269-278.

[8] Z. Wei, G. Li, L. Qi, *New quasi-Newton methods for unconstrained optimization problems*, Appl. Math. Comput. 175 (2) (2006) 1156-1188.

[9] B.K.Saman, G.Reza , M.A.Nezam ,*Two new conjugate gradient methods based on modified secant equations*, Journal of Computational and Applied Mathematics 234 (2010) 1374-1386.

[10] N. Andrei, *A scaled BFGS preconditioned conjugate gradient algorithm for unconstrained optimization,* Appl. Math. Lett. 20 (2007) 645-650

[11] W. Zhou, L. Zhang, *A nonlinear conjugate gradient method based on the MBFGS secant condition,* Optim. Methods Softw. 21 (5) (2006) 707-714.