ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 7, No. 1, 2012, *pp*. 067-071

Global Convergence of a New Conjugate Gradient Method with Wolfe Type Line Search+

Yuanyuan Chen[1,2]

[1]School of Management, University of Shanghai for Science and Technology, Shanghai 200093,
[2]College of Mathematics, Qingdao University, Qingdao, 266071.

(*Received August 27, 2011, accepted September 15, 2011*)

Abstract. In this paper, a new nonlinear conjugate gradient methods with wolfe type line search for solving unstrained optimization problems is proposed. The direction generated by the new methods produce sufficient descent search direction. Under some conditions, we give the global convergence results for the new nonlinear conjugate gradient methods with Wolfe type line search. Finally, the numerical results show that the new method is also efficient for general unconstrained optimizations.

Key Words: Unconstrained optimization; nonlinear conjugate gradient method; line search; global convergence

1. Introduction

In this paper, we consider the following large scale unconstrained optimization problem

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

where $f : R^n \to R$ is continuously differentiable and its gradient $g(x) = \nabla f(x)$ is available. This problem has been well studied. Iterative methods are widely used for solving (1.1) and the iterative formula is given by

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

where $x_k \in R^n$ is the k-th approximation to the solution of (1.1), α_k is a stepsize obtained by carrying out a line search and d_k is a search direction.

Due to the simplicity of its iteration and low memory requirements, the nonlinear conjugate gradient method are one of the most famous methods for solving the above unconstrained optimization problem (1.1), especially in case of the dimension n of $f(x)$ is large, which are often proposed in scientific and engineering computation. The search direction d_k is defined by

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

where β_k is a scalar and $g_k = g(x_k)$ .Different conjugate gradient methods correspond to different scalar β_k in (1.3), i.e. [1-6]

$$\beta_k^{FR} = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}}, \text{ (Fletcher-Reeves)}, \quad \beta_k^{HS} = \frac{g_k^T (g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1}}, \text{ (Hestenes-Stiefel)},$$

+ Supported by national science foundation of China(11101231, 11171221), A project of Shandong province higher educational science and technology program(J10LA05), Shanghai municipal committee of science and technology(10550500800).

$\beta_k^{PRP} = \frac{g_k^T(g_k - g_{k-1})}{g_{k-1}^T g_{k-1}}$, (Polak-Ribiere-Polyak), $\beta_k^{CD} = -\frac{g_k^T g_k}{d_{k-1}^T g_{k-1}}$, (Conjugate Descent),

$\beta_k^{LS} = -\frac{g_k^T(g_k - g_{k-1})}{d_{k-1}^T g_{k-1}}$, (Liu-Storey), $\beta_k^{DY} = \frac{g_k^T g_k}{(g_k - g_{k-1})^T d_{k-1}}$, (Dai-Yuan).

The nonlinear conjugate gradient methods and the global convergence results about Fletcher-Reeves(FR) method, Polak-Ribiere-Polyak(PRP) method, Hestenes-Stiefel(HS)method, Dai-Yuan(DY) method, Conjugate Descent(CD) method and Liu-Storey(LS) method can see [5-11]. Quite recently, a new nonlinear conjugate gradient method was proposed in [12]. The search direction β_k are given by the following way

$$\beta_k = \frac{1}{d_{k-1}^T y_{k-1}}(y_{k-1} - 2d_{k-1}\frac{\|y_{k-1}\|^2}{d_{k-1}^T y_{k-1}})^T g_k, \tag{1.4}$$

where $y_{k-1} = g_k - g_{k-1}$, $\|\cdot\|$ stands for the Euclidean norm.

In this paper, based on the above new nonlinear conjugate gradient method in [12], under some mild conditions, we give the global convergence of the new nonlinear conjugate gradient method with Wolfe type line search. The rest of the paper is organized as follows. In Section 2, we present the Wolfe type line search and the new nonlinear conjugate gradient method. In Section 3, we present the global convergence results of the new nonlinear conjugate gradient method with Wolfe type line search. Numerical results and some discussions are given in the Section 4.

2. New nonlinear conjugate gradient method

In this section, we will give the following assumptions on objective function $f(x)$ in (1.1). The two assumptions have been often used in the literature to prove the global convergence of nonlinear conjugate gradient methods with exact and inexact line searches for unconstrained optimization problems.

Assumption 2.1. (i) The level set $L_0 = \{x \in R^n \mid f(x) \le f(x_0)\}$ is bounded. (ii) In some neighborhood U of L_0, $f(x)$ is continuously differentiable and its gradient is Lipschitz continuous, namely, there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \le L\|x - y\|, \quad \forall x, y \in U.$$

And $g(x)$ also satisfies the following condition

$$(g(x) - g(y))(x - y) \ge \tau\|x - y\|^2,$$

for $\tau > 0$ and $\forall x, y \in U$.

Now we consider the modified new conjugate gradient method for (1.1) with Wolfe type line search. Firstly, we give the following Wolfe type line search (we have proposed in [14]).

The line search is to choose $\alpha_k > 0$ such that

$$f(x_k + \alpha_k d_k) - f(x_k) \le -\rho\alpha_k^2\|d_k\|^2, \tag{2.1}$$

$$g(x_k + \alpha_k d_k)^T d_k \ge -2\sigma\alpha_k\|d_k\|^2, \tag{2.2}$$

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

Now we present the new conjugate gradient methods as follows.

Algorithm 2.1.

Step 0. Given $x_0 \in R^n$, set $d_0 = -g_0$. If $g_0 = 0$, then stop.

Step 1. Find $\alpha_k > 0$ satisfying the Wolfe type line search (2.1), (2.2), by (1.2), x_{k+1} is given. If $g_{k+1} = 0$, then stop.

Step 2. Compute d_k by (1.3) and (1.4).

Step 3. Set $k := k+1$, go to Step 1.

In order to establish the global convergence of the Algorithm 2.1, we give the following lemmas.

Lemma 2.1. Suppose that Assumption 2.1 holds, then the Wolfe type line search (2.1), (2.2) is feasible.

The proof is essentially the same as Lemma 1 of [14], hence we omit it.

Lemma 2.2. Suppose direction d_k is given by (1.3) and (1.4), then we have

$$g_k^T d_k \le -\frac{7}{8}\|g_k\|^2,$$

holds for any $k \ge 0$, i.e. d_k is descent search direction.

The proof is essentially the same as Theorem 1.1 of [12], hence we omit it.

Lemma 2.3. Suppose that Assumption 2.1 holds, α_k is determined by Wolfe line search (2.1), (2.2), we have

$$\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty \tag{2.3}$$

Proof. From (2.1), (2.2), Lemma 2.2 and Assumption 2.1, we obtain

$$(2\sigma + L)\alpha_k \|d_k\|^2 \ge -g_k^T d_k .$$

Then, we know

$$\alpha_k \|d_k\| \ge \frac{1}{2\sigma + L}\left(-\frac{g_k^T d_k}{\|d_k\|}\right).$$

Squaring both sides of above formula, we get

$$\alpha_k^2 \|d_k\|^2 \ge \left(\frac{1}{2\sigma + L}\right)^2 \left(\frac{g_k^T d_k}{\|d_k\|}\right)^2 .$$

From (2.1), we have

$$\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} \le (2\sigma + L)^2 \sum_{k=1}^{\infty} \alpha_k^2 \|d_k\|^2 \le \frac{(2\sigma + L)^2}{\rho} \sum_{k=1}^{\infty} \{ f(x_k) - f(x_{k+1})\} < +\infty.$$

So we get (2.3), this completes the proof.

Lemma 2.4. Suppose that Assumption 2.1 holds, d_k is generated by (1.3) and (1.4), α_k is determined by Wolfe line search (2.1), (2.2), we have

$$\sum_{k \ge 0} \frac{\|g_k\|^4}{\|d_k\|^2} < +\infty . \tag{2.4}$$

Proof From Lemma 2.2 and Lemma 2.3, we obtain (2.4).

3. Global convergence result for the method

From the above analysis of Section 2, we give the following result of global convergence for the new conjugate gradient method. The general idea for proving the global convergence of the new conjugate gradient method with Wolfe type line search is to assume first by contradiction, then derive contradictory relations. We now give the global convergence theorem for the new conjugate gradient method in a similar way to the global convergence results in [12-13] under mild assumptions.

Theorem 3.1. Consider Algorithm 2.1, suppose that Assumptions 2.1 holds. Then we have

$$\liminf_{k \to \infty} \|g_k\| = 0$$

Proof. Firstly, we suppose that the conclusion is not true. Suppose by contradiction that there exists a positive constant $\varepsilon > 0$, such that $\|g_k\| \geq \varepsilon$ holds for all $k \geq 0$. From (1.3), (1.4) and Assumptions 2.1, we get

$$\|d_{k+1}\|^2 = \|g_{k+1}\|^2 + \beta_k^2 \|d_k\|^2 - 2\beta_k g_{k+1}^T d_k.$$

$$\beta_k \leq (\frac{L}{\tau} + \frac{2L^2}{\tau^2}) \frac{\|g_{k+1}\|}{\|d_k\|}.$$

$$\|d_{k+1}\|^2 \leq \|g_{k+1}\|^2 + \beta_k^2 \|d_k\|^2 + 2 | \beta_k | \|g_{k+1}\| \|d_k\|.$$

Dividing both sides of the above formula by $\|g_{k+1}\|^4$, we have

$$\frac{\|d_{k+1}\|^2}{\|g_{k+1}\|^4} \leq \frac{\beta_k^2 \|d_k\|^2 + 2|\beta_k| \|g_{k+1}\| \|d_k\|}{\|g_{k+1}\|^4} + \frac{1}{\|g_{k+1}\|^2} \leq \xi \frac{1}{\|g_{k+1}\|^2} \leq \frac{\xi}{\varepsilon^2},$$

where $\xi = (\frac{L}{\tau} + \frac{2L^2}{\tau^2} + 1)^2$. We obtain

$$\sum_{k \geq 0} \frac{\|g_k\|^4}{\|d_k\|^2} \geq \varepsilon^2 \sum_{k \geq 0} \frac{1}{\xi} = +\infty,$$

which contradicts (2.4). Therefor, we complete the proof of this theorem.

4. Numerical experiments and discussions

In this section, we give some numerical experiments for the above new nonlinear conjugate gradient method and some finical discussions. The problem that we tested are from [14]. Our line search subroutine computes α_k such that the line search condition (2.1) and (2.2) hold with ρ =0.001 and σ =0.01. We also use the condition $\|g_{k+1}\| \leq 10^{-6}$ as the stopping criterion in Algorithm 2.1. We use MATLAB7.0 to test the chosen problems. The numerical results of Algorithm 2.1 are listed in Table 4.1, where the items in each column have the following meanings:

Name: the name of the test problem;

Dim: the dimension of the problem;

NI: the number of iterations;

NF: the number of function evaluations;

NG: the number of gradient evaluations.

Table 4.1. Test results for the Algorithm 2.1

Name	dim	NI	NF	NG
LIN1	2	1	51	2
BAND	30	25	205	38
BAND	3	12	123	20
TRID	200	48	206	62
TRID	100	56	175	70
IE	200	10	21	21
TRIG	100	160	1326	256

Discussions. From the analysis of the global convergence of Algorithm 2.1, we can see when d_k satisfied the property of efficient descent search direction, we can get the global convergence of the corresponding conjugate gradient method with Wolfe type line search without other assumptions. Such as the following method, which have been discussed in [12]

$$d_k = -g_{k+1} + \tau d_k, \tag{4.1}$$

where $\tau \in [\beta_k^N, \max\{\beta_k^N, 0\}]$, is computed by (1.4). From the fact of $g_{k+1}^T d_{k+1} \le -\frac{7}{8}\|g_{k+1}\|^2$, we can get the global convergence result as Theorem 3.1.

Remark 4.1. Consider Algorithm 2.1, suppose that Assumptions 2.1 holds, d_k is computed by (4.1). Then we have

$$\liminf_{k\to\infty}\|g_k\| = 0.$$

5. References

[1] R.Fletcher, C.Reeves. Function Minimization by Conjugate Gradients. *J. Comput.* 1964, **7**: 149-154.

[2] B.T.Polyak. The conjugate gradient method in extreme problems. *USSR Comp.Math. Math. Phys.* 1969, **9**: 94-112.

[3] Y.H.Dai, Y.X.Yuan. A nonlinear conjugate gradient method with nice global convergence properties. *SIAM J. Optim.* 1999, **10**: 177-182.

[4] M.R.Hestenes, E.Stiefel. Method of conjugate gradient for solving linear equations. *J.Res.Nat.Bur.Stand*. 1952, **49**: 409-436.

[5] R.Fletcher. Practical methods of optimization. *unconstrained optimization*. New York: Wiley, 1987.

[6] Y. Liu, C. Storey. Efficient generalized conjugate gradient algorithms. *Part 1: Theory, J. Optim. Theory Appl.* 1991, **69**: 129-137.

[7] M. Raydan. The Barzilain and Borwein gradient method for the large unconstrained minimization problem. *SIAM J. Optim*. 1997, **7**: 26-33.

[8] E.G. Birgin, J.M. Martinez. A spectral conjugate gradient method for unconstrained optimization. *Appl. Math. Optim*. 2001, **43**: 117-128.

[9] J.C. Gilbert, J. Nocedal. Global Convergence properties of conjugate gradient methods for optimization. *SIAM J. Optim*. 1992, **2**: 21-24.

[10] J. Sun, J. Zhang. Convergence of conjugate gradient methods without line search. *Annals Operations Research*. 2001, **103**: 161-173.

[11] L. Zhang, W. Zhou, D. Li. Global convergence of a modified Fletcher-Reeves conjugate gradient method with Armijo-type line search. *Numer. Math*. 2006, **104**: 561-572.

[12] W.W. Hager, H. Zhang. A new conjugate gradient method with guaranteed descent and an efficient line search. *SIAM J. Optim*. 2005, **16**: 170-192.

[13] C. Wang, Y. Chen, S. Du. Further insight into the Shamanskii modification of Newton method. *Appl. Math. Comput*. 2006, **180**: 46-52.

[14] J.J. More, B.S. Garbow, K.E. Hillstrom. Testing unconstrained optimization software. *ACM Trans. Math. Software*. 1981, **7**: 17-41.