Hindawi Publishing Corporation
Abstract and Applied Analysis
Volume 2013, Article ID 742815, 5 pages
http://dx.doi.org/10.1155/2013/742815

Research Article

Nonlinear Conjugate Gradient Methods with Wolfe Type Line Search

Yuan-Yuan Chen and Shou-Qiang Du

College of Mathematics, Qingdao University, Qingdao 266071, China

Correspondence should be addressed to Yuan-Yuan Chen; usstchenyuanyuan@163.com

Received 20 January 2013; Accepted 6 February 2013

Academic Editor: Yisheng Song

Nonlinear conjugate gradient method is one of the useful methods for unconstrained optimization problems. In this paper, we consider three kinds of nonlinear conjugate gradient methods with Wolfe type line search for unstrained optimization problems. Under some mild assumptions, the global convergence results of the given methods are proposed. The numerical results show that the nonlinear conjugate gradient methods with Wolfe type line search are efficient for some unconstrained optimization problems.

1. Introduction

In this paper, we focus our attention on the global convergence of nonlinear conjugate gradient method with Wolfe type line search. We consider the following unconstrained optimization problem:

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

In (1), f is continuously differentiable function, and its gradient is denoted by $g(x) = \nabla f(x)$. Of course, the iterative methods are often used for (1). The iterative formula is given by

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

where x_k, $x_{k+1} \in R^n$ is the kth and (k+1)th iterative step, α_k is a step size, and d_k is a search direction. Here, in the following, we define the search direction by

$$d_k = \begin{cases} -g_k, & \text{if } k = 1, \\ -g_k + \beta_k d_{k-1}, & \text{if } k \geq 2. \end{cases} \tag{3}$$

In (3), β_k is a conjugate gradient scalar, and the well-known useful formulas are β_k^{FR}, β_k^{PRP}, β_k^{HS}, and β_k^{DY} (see [1–6]). Recently, some kinds of new nonlinear conjugate gradient methods are given in [7–11]. Based on the new method, we give some new kinds of nonlinear conjugate gradient methods and analyze the global convergence of the methods with Wolfe type line search.

The rest of the paper is organized as follows. In Section 2, we give the methods and the global convergence results for them. In the last section, numerical results and some discussions are given.

2. The Methods and Their Global Convergence Results

Firstly, we give the Wolfe type line search, which will be used in our new nonlinear conjugate gradient methods. In the following section of this paper, $\|\cdot\|$ stands for the 2-norm.

We have used the Wolfe type line search in [12].

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

$$f(x_k + \alpha_k d_k) \leq f(x_k) - \rho \alpha_k^2 \|d_k\|^2, \tag{4}$$

$$g(x_k + \alpha_k d_k)^T d_k \geq -2\sigma \alpha_k \|d_k\|^2, \tag{5}$$

where $\rho, \sigma \in (0, 1)$, $\rho < \sigma$.

Now, we present the nonlinear conjugate gradient methods as follows.

Algorithm 1. We have the following steps.

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

Step 1. Find $\alpha_k > 0$ satisfying (4) and (5), and by (2), x_{k+1} is given. If $\|g_{k+1}\| = 0$, then stop.

Step 2. Compute d_k by the following equation:

$$d_k = \beta_k d_{k-1} - \left(1 + \beta_k \frac{g_k^T d_{k-1}}{\|g_k\|^2}\right) g_k, \tag{6}$$

in which $\beta_k = -\|g_k\|^2 / d_{k-1}^T g_{k-1}$. Set $k := k + 1$, and go to Step 1.

Before giving the global convergence theorem, we need the following assumptions.

Assumption 1. (A1) The set $L_0 = \{x \in R^n \mid f(x) \leq f(x_0)\}$ is bounded.

(A2) In the neighborhood of L_0, denoted as U, f is continuously differentiable. Its gradient is Lipschitz continuous; namely, for $x, y \in U$, there exists $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\| \ . \tag{7}$$

In order to establish the global convergence of Algorithm 1, we also need the following lemmas.

Lemma 2. *Suppose that Assumption 1 holds; then,* (4) *and* (5) *are well defined.*

The proof is essentially the same as Lemma 1 of [12]; hence, we do not rewrite it again.

Lemma 3. *Suppose that direction d_k is given by* (6)*; then, one has*

$$d_k^T g_k = -\|g_k\|^2 \leq 0 \tag{8}$$

holds for all $k \geq 0$. So, one knows that d_k is descent search direction.

Proof. From the definitions of d_k and β_k, we can get it. □

Lemma 4. *Suppose that Assumption 1 holds, and α_k is determined by* (4) *and* (5)*; one has*

$$\sum_{k=1}^{\infty} \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} < +\infty. \tag{9}$$

Proof. By (4), (5), Lemma 3, and Assumption 1, we can get

$$-(2\sigma + L)\alpha_k \|d_k\|^2 \leq g_k^T d_k. \tag{10}$$

Then, we know that

$$(2\sigma + L)\alpha_k \|d_k\| \geq -\frac{g_k^T d_k}{\|d_k\|}. \tag{11}$$

By squaring both sides of the previous inequation, we get

$$(2\sigma + L)^2 \alpha_k^2 \|d_k\|^2 \geq \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2}. \tag{12}$$

By (4), we know that

$$\begin{aligned} \sum_{k=1}^{\infty} \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} &\leq \sum_{k=1}^{\infty} (2\sigma + L)^2 \alpha_k^2 \|d_k\|^2 \\ &\leq \frac{(2\sigma + L)^2}{\rho} \sum_{k=1}^{\infty} \{f(x_k) - f(x_{k+1})\} \\ &< +\infty. \end{aligned} \tag{13}$$

So, we get (9), and this completes the proof of the lemma. □

Lemma 5. *Suppose that Assumption 1 holds, d_k is computed by* (6)*, and α_k is determined by* (4) *and* (5)*; one has*

$$\sum_{k \geq 0} \frac{\|g_k\|^4}{\|d_k\|^2} < +\infty. \tag{14}$$

Proof. From Lemmas 3 and 4, we can obtain (14). □

Theorem 6. *Consider Algorithm 1, and suppose that Assumption 1 holds. Then, one has*

$$\liminf_{k \to \infty} \|g_k\| = 0. \tag{15}$$

Proof. We suppose that the theorem is not true. Suppose by contradiction that there exists $\epsilon > 0$ such that

$$\|g_k\| \geq \epsilon \tag{16}$$

holds for $k \geq 0$.

From (6) and Lemma 3, we get

$$\begin{aligned} \|d_k\|^2 = {} & (\beta_k)^2 \|d_{k-1}\|^2 - \left(1 + \beta_k \frac{g_k^T d_{k-1}}{\|g_k\|^2}\right)^2 \|g_k\|^2 \\ & - 2\left(1 + \beta_k \frac{g_k^T d_{k-1}}{\|g_k\|^2}\right) d_k^T g_k. \end{aligned} \tag{17}$$

Dividing the previous inequation by $(g_k^T d_k)^2$, we get

$$\begin{aligned} \frac{\|d_k\|^2}{\|g_k\|^4} &= \frac{\|d_k\|^2}{\left(g_k^T d_k\right)^2} \\ &= \frac{\beta_k^2 \|d_{k-1}\|^2}{\left(g_k^T d_k\right)^2} - \frac{\left(1 + \beta_k g_k^T d_{k-1} / \|g_k\|^2\right)^2 \|g_k\|^2}{\left(d_k^T g_k\right)^2} \\ &\quad - \frac{2\left(1 + \beta_k g_k^T d_{k-1} / \|g_k\|^2\right)}{d_k^T g_k} \end{aligned}$$

$$
\begin{aligned}
&= \frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^4} + \frac{1}{\|g_k\|^2} \\
&\quad - \frac{\beta_k^2 \left(g_k^T d_{k-1}\right)^2 / \|g_k\|^4}{\|g_k\|^2} \\
&\le \frac{1}{\|g_k\|^2} + \frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^4} \\
&\le \sum_{i=0}^{k-1} \frac{1}{\|g_i\|^2} \\
&\le \frac{k}{\epsilon^2}.
\end{aligned} \tag{18}
$$

So, we obtain

$$\sum_{k\ge 1} \frac{\|g_k\|^4}{\|d_k\|^2} \ge \sum_{k\ge 1} \epsilon^2 \frac{1}{k} = +\infty, \tag{19}$$

which contradicts (14). Hence we get this theorem. □

Remark 7. In Algorithm 1, we also can use the following equations to compute d_k:

$$d_k = \beta_k d_{k-1} - \beta_k \frac{g_k^T d_{k-1}}{\|g_k\|^2} - g_k, \tag{20}$$

where $\beta_k = \max\{\min\{\beta_k^{\mathrm{FR}}, \beta_k^{\mathrm{PRP}}\}, 0\}$;

$$d_k = \beta_k d_{k-1} - \beta_k \frac{g_k^T d_{k-1}}{\|g_k\|^2} - g_k, \tag{21}$$

where $\beta_k = \max\{\min\{\beta_k^{\mathrm{LS}}, \beta_k^{\mathrm{CD}}\}, 0\}$.

Algorithm 8. We have the following steps.

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

Step 1. Find $\alpha_k > 0$ satisfying (4) and (5), and by (2), x_{k+1} is given. If $\|g_{k+1}\| = 0$, then stop.

Step 2. Compute β_k by formula

$$\beta_k = \begin{cases} \beta_k^{\mathrm{DY}} = \dfrac{g_k^T g_k}{\left(g_k - g_{k-1}\right)^T d_{k-1}}, & \left\|g_{k-1}\right\|^2 \le g_k^T d_{k-1}, \\[2ex] \beta_k^{\mathrm{FR}} = \dfrac{\|g_k\|^2}{\|g_{k-1}\|^2}, & \left\|g_{k-1}\right\|^2 > g_k^T d_{k-1}, \end{cases} \tag{22}$$

and compute d_{k+1} by (3). Set $k := k+1$, and go to Step 1.

Lemma 9. *Suppose that Assumption 1 holds, and β_k is computed by* (22)*; if $\|g_k\| \ne 0$, then one gets $g_k^T d_k < 0$ for all $k \ge 2$ and $\beta_k^{\mathrm{FR}} \ge |\beta_k|$ (see [9]).*

Lemma 10. *Suppose that $l > 0$, and ν is a constant. If positive series ζ_k satisfied*

$$\sum_{i=1}^{k} \zeta_i \ge lk + \nu, \tag{23}$$

one has

$$
\begin{gathered}
\sum_{i\ge 1} \frac{\zeta_i^2}{i} = +\infty, \\
\sum_{k\ge 1} \frac{\zeta_k^2}{\sum_{i=1}^{k} \zeta_i} = +\infty.
\end{gathered} \tag{24}
$$

From the previous analysis, we can get the following global convergence result for Algorithm 8.

Theorem 11. *Suppose that Assumption 1 holds, and $\|g(x)\|^2 \le \bar{c}$, where $\bar{c}$ is a constant. Then, one has*

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{25}$$

Proof. Suppose by contradiction that there exists $\varepsilon > 0$ such that

$$\left\|g_k\right\|^2 \ge \varepsilon \tag{26}$$

holds for all k. From (3), we have

$$d_k = \beta_k d_{k-1} - g_k. \tag{27}$$

Squaring both sides of the previous equation, we get

$$\left\|d_k\right\|^2 = \left(\beta_k \left\|d_{k-1}\right\|\right)^2 - \left\|g_k\right\|^2 - 2 g_k^T d_k. \tag{28}$$

Let $\vartheta_k = \|d_k\|^2/\|g_k\|^4$ and $r_k = -g_k^T d_k/\|g_k\|^2$; from (22), we have

$$\vartheta_k \le \vartheta_{k-1} - 2\frac{r_k}{\|g_k\|^2} - \frac{1}{\|g_k\|^2}. \tag{29}$$

By $\vartheta_1 = 1/\|g_1\|^2$, $r_1 = 1$, we know that

$$\vartheta_k \le \sum_{i=1}^{k} \frac{2}{\varepsilon}\left|r_i\right| - \frac{k}{\bar{c}}. \tag{30}$$

By (30), we get

$$
\begin{gathered}
\vartheta_k \le \sum_{i=1}^{k} \frac{2}{\varepsilon}\left|r_i\right|, \\
\sum_{i=1}^{k} \left|r_i\right| \ge 2\frac{\varepsilon k}{\bar{c}}.
\end{gathered} \tag{31}
$$

From (31) and Lemma 10, we have

$$\sum_{k\ge 1} \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} = +\infty, \tag{32}$$

which contradicts Lemma 4. Therefore, we get this theorem. □

Algorithm 12. We have the following steps.

Step 0. Given $x_0 \in R^n$, $\mu > 1/4$, set $d_0 = -g_0$, $k := 0$. If $\|g_0\| = 0$, then stop.

Step 1. Find $\alpha_k > 0$ satisfying (4) and (5), and by (2), x_{k+1} is given. If $\|g_{k+1}\| = 0$, then stop.

Step 2. Compute d_k by

$$d_k = \begin{cases} -g_k, & k = 0, \\ \beta_k d_{k-1} - \left(1 + \beta_k \dfrac{g_k^T d_{k-1}}{\|g_k\|^2}\right) g_k, & k \geq 1, \end{cases} \tag{33}$$

where

$$\beta_k = \frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2} - \mu \frac{\|g_k - g_{k-1}\|^2 g_k^T d_{k-1}}{\|g_{k-1}\|^4}. \tag{34}$$

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

Lemma 13. *Suppose that direction d_k is given by* (33) *and* (34)*; then, one has*

$$d_k^T g_k = -\|g_k\|^2 \leq 0 \tag{35}$$

holds for any $k \geq 0$.

Lemma 14. *Suppose that Assumption 1 holds, d_k is generated by* (33) *and* (34)*, and α_k is determined by* (4) *and* (5)*; one has*

$$\sum_{k\geq 0} \frac{\|g_k\|^4}{\|d_k\|^2} < +\infty. \tag{36}$$

Proof. From Lemma 4 and Lemma 13, we obtain (36). □

Lemma 15. *Suppose that f is convex. That is, $d^T \nabla^2 f(x) d \geq 0$, for all $d \in R^n$, where $\nabla^2 f(x)$ is the Hessian matrix of f. Let $\{x_k\}$ and $\{d_k\}$ be generated by Algorithm 12; one has*

$$\rho \alpha_k \|d_k\|^2 \leq -g_k^T d_k. \tag{37}$$

Proof. By Taylor's theorem, we can get

$$f(x_{k+1}) = f(x_k) + g_k^T s_k + \frac{1}{2} s_k^T G_k s_k, \tag{38}$$

where $s_k = x_{k+1} - x_k$, and $G_k = \int_0^1 \nabla^2 f(x_k + \tau s_k) d\tau s_k$.

By Assumption 1, (4), and (38), we get

$$-\rho \alpha_k^2 \|d_k\|^2 \geq f(x_{k+1}) - f(x_k) \geq g_k^T s_k = \alpha_k g_k^T d_k. \tag{39}$$

So, we get (37). □

Theorem 16. *Consider Algorithm 12, and suppose that Assumption 1 and the assumption of Lemma 15 hold. Then, one has*

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{40}$$

Proof. We suppose that the conclusion is not true. Suppose by contradiction that there exists $\epsilon > 0$ such that

$$\|g_k\| \geq \epsilon \tag{41}$$

holds for all $k \geq 0$.

By Lemma 13, we have

$$\beta_k = -\frac{g_k^T (g_k - g_{k-1})}{g_{k-1}^T d_{k-1}} - \mu \left(\frac{\|g_k - g_{k-1}\|}{-g_{k-1}^T d_{k-1}}\right)^2 g_k^T d_{k-1}. \tag{42}$$

From Assumption 1, Lemma 15, and (42), we know that

$$|\beta_k| \leq \left(\frac{\mu L^2 + \rho L}{\rho^2}\right) \frac{\|g_k\|}{\|d_{k-1}\|}. \tag{43}$$

Therefore, by (33), we get

$$\begin{aligned} \|d_k\| &\leq \|g_k\| + 2\left(\frac{\mu L^2 + \rho L}{\rho^2}\right)\|g_k\| \\ &= \left(1 + 2\frac{L}{\rho} + 2\frac{\mu L^2}{\rho^2}\right)\|g_k\|. \end{aligned} \tag{44}$$

We obtain

$$\sum_{k\geq 1} \frac{\|g_k\|^4}{\|d_k\|^2} \geq +\infty, \tag{45}$$

which contradicts (36). Therefore, we have

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{46}$$

So, we complete the proof of this theorem. □

Remark 17. In Algorithm 12, β_k can also be computed by the following formula:

$$\begin{aligned} \beta_k = {} & \frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2} \\ & - \min\left\{\frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2}, \mu \frac{\|g_k - g_{k-1}\|^2 g_k^T d_{k-1}}{\|g_{k-1}\|^4}\right\}, \end{aligned} \tag{47}$$

where $\mu > 1/4$.

3. Numerical Experiments and Discussions

In this section, we give some numerical experiments for the previous new nonlinear conjugate gradient methods with Wolfe type line search and some discussions. The problems that we tested are from [13]. We use the condition $\|g_{k+1}\| \leq 10^{-6}$ as the stopping criterion. We use MATLAB 7.0 to test the chosen problems. We give the numerical results of Algorithms 1 and 12 to show that the method is efficient for unconstrained optimization problems. The numerical results of Algorithms 1 and 12 are listed in Tables 1 and 2.

Table 1: Test results for Algorithm 1.

Name	Dim	NI	NF	NG
GULF	3	2	52	3
VARDIM	2	3	54	6
LIN	50	1	3	3
LIN	2	1	3	3
LIN	1000	1	3	3
LIN1	2	1	51	2
LIN1	10	1	3	3
LIN0	4	1	3	3

Name: the test problem name; Dim: the problem dimension; NI: the iterations number; NF: the function evaluations number; NG: the gradient evaluations number.

Table 2: Test results for Algorithm 12.

Name	Dim	NI	NF	NG
GULF	3	2	52	3
BIGGS	6	1000	1149	1095
IE	50	1000	1053	1003
IE	3	1000	1101	1052
TRIG	1000	1103	1052	3
BV	10	5	203	55
BV	3	6	109	9

Name: the test problem name; Dim: the problem dimension; NI: the iterations number; NF: the function evaluations number; NG: the gradient evaluations number.

Discussion 1. From the analysis of the global convergence of Algorithm 1, we can see that if d_k satisfies the property of efficient descent search direction, we can get the global convergence of the corresponding nonlinear conjugate gradient method with Wolfe type line search without other assumptions.

Discussion 2. In Algorithm 8, we use a Wolfe type line search. Overall, we also feel that nonmonotone line search (see [14]) also can be used in our algorithms.

Discussion 3. From the analysis of the global convergence of Algorithm 12, we can see that when d_k is an efficient descent search direction, we can get the global convergence of the corresponding conjugate gradient method with Wolfe type line search without requiring uniformly convex function.

Acknowledgments

This work is supported by the National Science Foundation of China (11101231 and 10971118), Project of Shandong Province Higher Educational Science and Technology Program (J10LA05), and the International Cooperation Program for Excellent Lecturers of 2011 by Shandong Provincial Education Department.

References

[1] B. T. Polyak, "The conjugate gradient method in extremal problems," *USSR Computational Mathematics and Mathematical Physics*, vol. 9, no. 4, pp. 94–112, 1969.

[2] R. Fletcher and C. M. Reeves, "Function minimization by conjugate gradients," *The Computer Journal*, vol. 7, pp. 149–154, 1964.

[3] M. R. Hestenes, "E.Stiefel. Method of conjugate gradient for solving linear equations," *Journal of Research of the National Bureau of Standards*, vol. 49, pp. 409–436, 1952.

[4] Y. H. Dai and Y. Yuan, "A nonlinear conjugate gradient method with a strong global convergence property," *SIAM Journal on Optimization*, vol. 10, no. 1, pp. 177–182, 1999.

[5] R. Fletcher, *Practical Methods of Optimization, Unconstrained Optimization*, Wiley, New York, NY, USA, 2nd edition, 1987.

[6] M. Raydan, "The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem," *SIAM Journal on Optimization*, vol. 7, no. 1, pp. 26–33, 1997.

[7] L. Zhang and W. Zhou, "Two descent hybrid conjugate gradient methods for optimization," *Journal of Computational and Applied Mathematics*, vol. 216, no. 1, pp. 251–264, 2008.

[8] A. Zhou, Z. Zhu, H. Fan, and Q. Qing, "Three new hybrid conjugate gradient methods for optimization," *Applied Mathematics*, vol. 2, no. 3, pp. 303–308, 2011.

[9] B. C. Jiao, L. P. Chen, and C. Y. Pan, "Global convergence of a hybrid conjugate gradient method with Goldstein line search," *Mathematica Numerica Sinica*, vol. 29, no. 2, pp. 137–146, 2007.

[10] G. Yuan, "Modified nonlinear conjugate gradient methods with sufficient descent property for large-scale optimization problems," *Optimization Letters*, vol. 3, no. 1, pp. 11–21, 2009.

[11] Z. Dai and B. Tian, "Global convergence of some modified PRP nonlinear conjugate gradient methods," *Optimization Letters*, vol. 5, no. 4, pp. 615–630, 2011.

[12] C. Y. Wang, Y. Y. Chen, and S. Q. Du, "Further insight into the Shamanskii modification of Newton method," *Applied Mathematics and Computation*, vol. 180, no. 1, pp. 46–52, 2006.

[13] J. J. Moré, B. S. Garbow, and K. E. Hillstrom, "Testing unconstrained optimization software," *ACM Transactions on Mathematical Software*, vol. 7, no. 1, pp. 17–41, 1981.

[14] L. Grippo, F. Lampariello, and S. Lucidi, "A nonmonotone line search technique for Newton's method," *SIAM Journal on Numerical Analysis*, vol. 23, no. 4, pp. 707–716, 1986.