J. Appl. Math. & Informatics Vol. **34***(2016), No. 1 - 2, pp. 157 - 165*
http://dx.doi.org/10.14317/jami.2016.157

A NONLINEAR CONJUGATE GRADIENT METHOD AND ITS GLOBAL CONVERGENCE ANALYSIS†

AJIE CHU, YIXIAO SU AND SHOUQIANG DU*

Abstract. In this paper, we develop a new hybridization conjugate gradient method for solving the unconstrained optimization problem. Under mild assumptions, we get the sufficient descent property of the given method. The global convergence of the given method is also presented under the Wolfe-type line search and the general Wolfe line search. The numerical results show that the method is also efficient.

AMS Mathematics Subject Classification : 65K05, 90C30.
Key words and phrases : Nonlinear conjugate gradient method, global convergence, line search.

1. Introduction

We consider the unconstrained optimization problem

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

where $f : R^n \to R$ is a continuously differentiable function. The nonlinear conjugate gradient method is very useful for solving (1.1), especially when n is large. For any given $x \in R^n$, the nonlinear conjugate gradient method generates x_k, $k = 1, 2, \dots, n$, by the following recursive relation

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

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

Received November 22, 2014. Revised September 5, 2015. Accepted September 8, 2015. *Corresponding author. †This work was supported by National Natural Science Foundation of China (11101231, 11401331), Natural Science Foundation of Shandong(ZR2015AQ013) and Key issues of statistical research in Shandong Province(KT15173).

where $g_k = \nabla f(x_k)$ is the gradient of f at x_k and β_k is typically given by some formulas (such as [1-5]).

$$\beta_k^{FR} = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}} \qquad \beta_k^{PRP} = \frac{g_k^T (g_k - g_{k-1})}{g_{k-1}^T g_{k-1}}$$

$$\beta_k^{HS} = \frac{g_k^T (g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1}} \qquad \beta_k^{DY} = \frac{g_k^T g_k}{d_{k-1}^T (g_k - g_{k-1})}$$

To achieve good computational performance and maintain the attractive feature of strong global convergence, in the past years, there exist many hybridizations of the basic conjugate gradient methods (see [6-10]). Based on the above papers, in this paper, we present a new hybridization nonlinear conjugate gradient method, where β_k is given as

$$\beta_k = \begin{cases} \dfrac{-g_k^T g_k}{|g_{k-1}^T d_{k-1}| + \dfrac{1}{u}|g_k^T d_{k-1}|}, & 0 < u < 1,\ \|g_k\|^2 < |g_k^T d_k|, \\ 0, & \text{otherwise.} \end{cases} \tag{1.4}$$

α_k is computed by the Wolfe-type line search, which is proposed in [11]

$$f(x_k + \alpha_k d_k) - f(x_k) \le \max\{\delta \alpha_k g_k^T d_k, -\gamma \alpha_k^2 \|d_k\|^2\}, \tag{1.5}$$

$$g(x_k + \alpha_k d_k)^T d_k \ge \max\{\sigma g_k^T d_k, -2\sigma \alpha_k \|d_k\|^2\}, \tag{1.6}$$

where $0 < \delta < \sigma < 1, 0 < \gamma < 1$. Based on the hybridization of β_k as given by (1.4) we give the nonlinear conjugate gradient methods under the Wolfe-type line search and the general Wolfe line search.

In Section 2, we give the Method 2.1 and prove the global convergence of the proposed method with Wolfe-type line search. In Section 3, some discussions and the numerical results of the Method 2.1 are also given.

2. Method 2.1 and its global convergence analysis

Now, we give the Method 2.1 for solving (1.1).

Method 2.1

Step 1. Choose initial point $x_0 \in R^n$, $\varepsilon \ge 0$, $0 < \delta < \sigma < 1$, $u, \gamma \in (0, 1)$.
Step 2. Set $d_1 = -g_1$, $k = 1$, if $\|g_1\| = 0$, then stop.
Step 3. Let $x_{k+1} = x_k + \alpha_k d_k$, compute α_k by (1.5) and (1.6).
Step 4. Compute g_{k+1}, if $\|g_{k+1}\| \le \varepsilon$, then stop. Otherwise, go to next step.
Step 5. Compute β_{k+1} by (1.4) and generate d_{k+1} by (1.3).
Step 6. Set $k = k + 1$, go to step 3.

In order to establish the global convergence of the Method 2.1, we need the following assumption, which are often used in the literature to analyze the global convergence of nonlinear conjugate gradient methods.

Assumption 2.2

(i) The level set $\Omega = \{x \in R^n | f(x) \le f(x_1)\}$ is bounded, i.e., there exists a positive constant C such that $\|x\| \le C$, for all $x \in \Omega$.

(ii) In some neighborhood Ω of L, f is continuously differentiable and its gradient is Lipchitz continuous, i.e., there exists a constant $L > 0$, such that

$$\|g(x) - g(y)\| \leq L\|x - y\|,$$

for all x,$y \in \Omega$.

Theorem 2.1. *Let the sequences* $\{x_k\}$ *and* $\{d_k\}$ *be generated by the method* (1.2), (1.3)*, and* β_k *is computed by* (1.4)*. Then, we have*

$$g_k^T d_k \leq -(1-u)\|g_k\|^2, \tag{2.1}$$

for all $k \geq 1$*, where* $u \in (0,1)$*.*

Proof. If $k = 1$, from (1.3), we get

$$g_k d_k = -\|g_k\|^2.$$

Then, we can easily conclude (2.1). If $k \geq 2$, multiplying (1.3) by g_k^T, from (1.4), we get

$$\begin{aligned} g_k^T d_k &= -\|g_k\|^2 + \beta_k g_k^T d_{k-1} \\ &\leq -\|g_k\|^2 + |\beta_k| \cdot |g_k^T d_{k-1}| \\ &\leq -\|g_k\|^2 + \frac{\|g_k\|^2}{|g_{k-1}^T d_{k-1}| + \frac{1}{u}|g_k^T d_{k-1}|} \cdot |g_k^T d_{k-1}| \\ &\leq -\|g_k\|^2 + \frac{\|g_k\|^2}{\frac{1}{u}|g_k^T d_{k-1}|} \cdot |g_k^T d_{k-1}| \\ &\leq -\|g_k\|^2 + u\|g_k\|^2 \\ &= -(1-u)\|g_k\|^2. \end{aligned}$$

□

Theorem 2.2. *Suppose that Assumption 2.2 holds. By the Method 2.1, we have*

$$\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty. \tag{2.2}$$

Proof. From Theorem 2.1 and Assumption (i), we can know that $\{f(x_k)\}$ is bounded and monotonically decreasing, i.e., $\{f(x_k)\}$, $k = 1, 2, \ldots, n$, is convergent series. By (1.6), we have that

$$d_k^T (g_{k+1} - g_k) \geq (\sigma - 1) g_k^T d_k. \tag{2.3}$$

From Assumption 2.2, we get

$$d_k^T (g_{k+1} - g_k) \leq L\alpha_k \|d_k\|^2. \tag{2.4}$$

So, from (2.3) and (2.4), we have

$$\alpha_k \|d_k\| \geq \frac{1-\sigma}{L} \frac{(-g_k^T d_k)}{\|d_k\|}. \tag{2.5}$$

Square both sides of (2.5), we have

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

Therefore, by $f(x_k) - f(x_{k+1}) \geq \gamma \alpha_k^2 \|d_k\|^2$, we get

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

According to the convergence of $\{f(x_k)\}$, we can conclude that

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

□

Remark 2.3. Suppose that Assumption 2.2 holds. By the Method 2.1, we know that

$$\sum_{k=1}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < +\infty. \tag{2.6}$$

Proof. From Theorem 2.1, we know that

$$g_k^T d_k \leq -(1-u)\|g_k\|^2, \tag{2.7}$$

where $u \in (0,1)$.
Square both sides of (2.7), we have

$$(g_k^T d_k)^2 \geq (1-u)^2 \|g_k\|^4.$$

Divided both sides of the above inequation by $\|d_k\|^2$, we get

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

From Theorem 2.2, we can conclude that

$$\sum_{k=1}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < +\infty.$$

□

Theorem 2.4. *Suppose that Assumption 2.2 holds. If $\{x_k\}$ $(k = 1, 2, \ldots, n)$ is generated by Method 2.1, we have*

$$\lim_{k\to\infty} inf \|g_k\| = 0. \tag{2.8}$$

Proof. If (2.8) does not hold, there exists $\widetilde{r} > 0$, such that

$$\|g_k\| \geq \widetilde{r}, \tag{2.9}$$

holds for all $k \geq 1$. From (1.4), if $\|g_k\|^2 < |g_k^T d_k|$, we have

$$\beta_k = \frac{-g_k^T g_k}{|g_{k-1}^T d_{k-1}| + \frac{1}{u}|g_k^T d_{k-1}|}.$$

From (1.3) and (1.4), we know that

$$d_{k+1} + g_{k+1} = \beta_{k+1} d_k. \tag{2.10}$$

Square both sides of (2.10) , we get

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

Divided both sides of the above equation by $(g_{k+1}^T d_{k+1})^2$, we get

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

By

$$\frac{\|d_1\|^2}{(g_1^T d_1)^2} = \frac{1}{\|g_1\|^2},$$

we have

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} \leq \sum_{i=1}^{k} \frac{1}{\|g_i\|^2}. \tag{2.11}$$

By (2.9) and (2.11), we know that

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} \leq \frac{k}{\widetilde{r}^2}.$$

Therefore, by $\frac{(g_k^T d_k)^2}{\|d_k\|^2} \geq \frac{\widetilde{r}^2}{k}$, we have

$$\sum_{k \geq 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} = +\infty. \tag{2.12}$$

If $\|g_k\|^2 \geq |g_k^T d_k|$, we get

$$\beta_k = 0,\ d_k = -g_k.$$

We can easily conclude that

$$\sum_{k\geq 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} = +\infty,$$

which leads to a contradiction with (2.2). This shows (2.8) holds. We finish the proof of the theorem. □

3. Discussions of the Method 2.1 and Numerical Results

The line search in Method 2.1 can also given by the general Wolfe line search

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k, \tag{3.1}$$

$$\sigma_1 g_k^T d_k \leq g_{k+1}^T d_k \leq -\sigma_2 g_k^T d_k, \tag{3.2}$$

where $0 < \sigma_1 \leq \sigma_2 < 1$.

Theorem 3.1. *Suppose that Assumption 2.2 holds. Consider the Method 2.1, where α_k satisfies* (3.1), (3.2). *Then, we have*

$$\sum_{k\geq 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty.$$

Proof. From (3.2), we get

$$(\sigma_1 - 1) g_k^T d_k \leq (g_{k+1} - g_k)^T d_k,$$

$$(1-\sigma_1)|g_k^T d_k| \leq \|g_{k+1} - g_k\| \cdot \|d_k\| \leq L\alpha_k \|d_k\|^2.$$

So

$$\alpha_k \geq \frac{(1-\sigma_1)|g_k^T d_k|}{L\|d_k\|^2}. \tag{3.3}$$

From (3.1), we have

$$f(x_k) - f(x_{k+1}) \geq -\delta \alpha_k g_k^T d_k.$$

By (3.3), we get

$$f(x_k) - f(x_{k+1}) \geq \delta \frac{1-\sigma_1}{L} \cdot \frac{(g_k^T d_k)^2}{\|d_k\|^2}.$$

That is

$$\sum_{k\geq 1} [f(x_k) - f(x_{k+1})] \geq \sum_{k\geq 1} \delta \frac{1-\sigma_1}{L} \frac{(g_k^T d_k)^2}{\|d_k\|^2}.$$

We have

$$\frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty.$$

□

Discussion 3.1 By Theorem 3.1, we also can get the global convergence of the Method 2.1 with (3.1), (3.2).

Discussion 3.2 If the line search in the Method 2.1 is given by the other Wolfe-type line search, which is given in [12]

$$f(x_k + \alpha_k d_k) - f(x_k) \le -\gamma \alpha_k^2 \|d_k\|^2,$$
$$g(x_k + \alpha_k d_k)^T d_k \ge -2\sigma \alpha_k \|d_k\|^2,$$

the method is also globally convergent.

Discussion 3.3 If in the Method 2.1, β_k is given as

$$\beta_k = \begin{cases} \dfrac{-\theta_1 g_k^T g_k}{\dfrac{1}{\theta_2}|g_{k-1}^T d_{k-1}| + \dfrac{1}{\theta_3}|g_k^T d_{k-1}|}, & 0 < \theta_1,\ \theta_2,\ \theta_3 < 1,\ \theta_1 + \theta_2 + \theta_3 = 1,\ \|g_k\|^2 < |g_k^T d_k|, \\ 0, & \text{otherwise}, \end{cases}$$

α_k satisfies (1.5), (1.6) or (3.1), (3.2), we also can get the global convergence of the Method 2.1.

Numerical Results 3.4

Now, we test the Method 2.1, where α_k satisfing (1.5), (1.6) or (3.1), (3.2) by using double precision versions of the unconstrained optimization problems in the CUTE library [13].

For the Method 2.1, α_k is computed by (1.5) and (1.6) with $\delta = 0.4$ and $\sigma = 0.7$ in the Table 3.1. α_k is computed by (3.1) and (3.2) with $\sigma_1 = 0.5$ and $\sigma_2 = 0.6$ in the Table 3.2.

The numerical results are given in the form of NI/NF/NG/g, where NI, NF, NG denote the numbers of iterations, function evaluations, and gradient evaluations and g denotes the finally gradient norm. Finally, all attempts to solve the test problems were limited to reaching maximum of achieving a solution with $\|g_k\| \le 10^{-3}$.

Table 3.1

PROBLEM	DIM	NI/NF/NG/g
JENSAM	2	2/164/54/0
GAUSS	3	1/4/2/7.806334e-04
GULF	3	42/115/94/3.236456e-04
BOX	3	133/278/134/9.589702e-04
OSB2	11	4/155/56/4.218750e-04
PEN1	4	4/8/5/1.110815e-05
TRIG	3	36/7/37/7.273758e-04
TRIG	50	35/14/36/9.274977e-04
TRIG	100	26/10/27/8.383886e-04

KOWOSB	4	109/4545/2042/9.897984e-04
IE	3	8/211/102/6.736650e-04
IE	50	9/213/101/5.091395e-04
IE	100	9/213/101/7.386004e-04
IE	200	9/213/101/9.148532e-04
IE	500	10/215/102/3.810671e-04
TRID	50	70/741/258/9.296880e-04
TRID	200	188/1446/305/9.335612e-04
LIN	2	1/2/2/5.117875e-16
LIN	50	1/2/2/3.140185e-15
LIN	500	1/2/2/4.468561e-14
LIN	1000	1/2/2/6.366875e-13

Table 3.2

PROBLEM	DIM	NI/NF/NG/g
ROSE	2	130/3061/959/9.428937e-04
GAUSS	3	3/107/48/5.070183e-04
GULF	3	21/1136/513/8.595130e-04
KOWOSB	4	109/5047/2178/9.868359e-04
ROSEX	8	165/2421/681/9.110545e-04
PEN1	2	3/111/55/9.909755e-04
PEN2	4	35/801/355/9.206275e-04
VARDIM	2	8/135/52/7.889590e-04
VARDIM	50	22/496/23/6.552673e-04
IE	3	9/217/109/5.867130e-04
IE	50	9/214/105/6.771816e-04
IE	100	9/213/105/5.703362e-04
IE	200	10/215/106/2.006254e-04
IE	500	10/215/106/3.500399e-04
TRID	50	126/1663/540/6.969663e-04
BAND	3	12/169/60/3.566964e-04
BAND	50	53/1102/404/8.178660e-04
BAND	100	24/346/112/7.547444e-04
BAND	200	24/346/113/8.209428e-04
LIN	2	1/2/2/5.117875e-16
LIN	50	1/2/2/3.140185e-15
LIN	500	1/2/2/4.468561e-14
LIN	1000	1/2/2/6.366875e-13
LIN1	10	26/576/55/9.240066e-04
LIN0	4	17/245/59/4.271362e-04

References

1. R. Fletcher and C.M. Reeves, *Function minimization by conjugate gradients*, Computer Journal, **7** (1964), 149-154.
2. E. Polak and G. Ribiere, *Note sur la convergence de directions conjugees*, Rev Francaise Informat Recherche Operatinelle, 3e Annee. **16** (1969), 35-43.
3. B.T. Polyak, *The conjugate gradient method in extreme problems*, USSR Computational Mathematics and Mathematical Physics, **9** (1969), 94-112.
4. M.R. Hestenes and E.L. Stiefel, *Methods of conjugate gradients for solving linear systems*, Journal of Research of the National Bureau of Standards, **49** (1952), 409-436.
5. Y.H. Dai and Y. Yuan, *A nonlinear conjugate gradient method with a strong global convergence property*, SIAM Journal on Optimization, **10** (1999), 177-182.
6. J.C. Gilbert and J. Nocedal, *Global convergence properties of conjugate gradient methods for optimization*, SIAM Journal on Optimization, **2** (1992), 21-42.
7. N. Andrei, *Accelerated hybrid conjugate gradient algorithm with modified secant condition for unconstrained optimization*, Numerical Algorithms, **54** (2010), 23-46.
8. S. Babaie-Kafaki, R. Ghanbari and N. Mahdavi-Amiri, *Two new conjugate gradient methods based on modified secant equations*, Journal of Computational and Applied Mathematics, **234** (2010), 1374-1386.
9. Y.H. Dai and Y. Yuan, *An efficient hybrid conjugate gradient method for unconstrained optimization*, Annals of Operations Research, **103** (2001), 33 -47.
10. G. Li, C. Tang and Z. Wei, *New conjugacy condition and related new conjugate gradient methods for unconstrained optimization*, Journal of Computational and Applied Mathematics, **202** (2007), 523-539.
11. X.J. Zhang and A.N. Xu, *Global convergence properties of a new class of nonlinear conjugate gradient methods*, Guangxi Sciences, **12** (2005), 282-283.
12. C.Y. Wang, S.Q. Du and Y.Y. Chen, *Global covergence properties of three-term conjugate gradient method with new-type line search*, Journal of Systems Science and Complexity, **17** (2004), 412-420.
13. I. Bongartz, A.R. Conn, N. Gould and P.L. Toint, *CUTE: constrained and unconstrained testing environment*, Acm Transactions on Mathematical Software, **50** (1995), 123-160.

Ajie Chu received her BS from Xingtan CollegeQufu Normal University, China in 2013. Since 2013 she has been at Qingdao University, China. Her research interests include optimization, nonsmooth equations.

College of Mathematics, Qingdao University, Qingdao 266071, China.
e-mail: 13658689660@163.com

Yixiao Su received his BS from Dezhou College, China in 2013. Since 2013 he has been at Qingdao University, China. His research interests include optimization, complementarity problems.

College of Mathematics, Qingdao University, Qingdao 266071, China.
e-mail: 124890405@qq.com

Shouqiang Du PhD, Vice Professor. Since 2003 he has been at College of Mathematics, Qingdao University. His research interests include numerical optimization and nonsmooth optimization.

College of Mathematics, Qingdao University, Qingdao, 266071, China.
e-mail: dsq8933@163.com; sqdu@qdu.edu.cn