Applied Mathematics, 2011, 2, 1119-1123
doi:10.4236/am.2011.29154 Published Online September 2011 (http://www.SciRP.org/journal/am)

A New Descent Nonlinear Conjugate Gradient Method for Unconstrained Optimization

Hao Fan, Zhibin Zhu, Anwa Zhou
School of Mathematics and Computing Science, Guilin University of Electronic Technology, Guilin, China
E-mail: 540103003@qq.com
Received June 15, 2011; *revised July* 10, 2011; *accepted July* 17, 2011

Abstract

In this paper, a new nonlinear conjugate gradient method is proposed for large-scale unconstrained optimization. The sufficient descent property holds without any line searches. We use some steplength technique which ensures the Zoutendijk condition to be held, this method is proved to be globally convergent. Finally, we improve it, and do further analysis.

Keywords: Large Scale Unconstrained Optimization, Conjugate Gradient Method, Sufficient Descent Property, Globally Convergent

1. Introduction

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

$$\min f(x), \quad x \in \Re^n$$

where $f: \Re^n \to \Re$ is a smooth nonlinear function, and the gradient of f at x is denoted by $g(x)$. The iterative formula of the conjugate gradient methods is given by

$$x_{k+1} = x_k + t_k d_k, \quad k = 0,1,..., \tag{1.1}$$

where t_k is a steplength which is computed by carrying out some line search, d_k is the search direction defined by

$$d_k = \begin{cases} -g_k, & \text{if } k = 1, \\ -g_k + \beta_k d_{k-1}, & \text{if } k \geq 2, \end{cases} \tag{1.2}$$

where β_k is a scalar, g_k denotes $g(x_k)$.

There are some well-known formulas for β_k, which are given as follows:

$$\beta_k^{PRP} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}, \text{ (Polak-Ribiere-Polyak [1])}, \tag{1.3}$$

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \text{ (Fletcher-Reeves [2])}, \tag{1.4}$$

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

In addition, the sufficient descent condition is defined as follows:

$$g_k^T d_k \leq -c\|g_k\|^2, \tag{1.5}$$

where $c > 0$, has often been used in the literature to analyze the global convergence of conjugate gradient methods with inexact line searches.

Generally, the PRP method was much better than the FR method judging from the numerical calculation. When the objective function was convex, Polak and Ribie`re proved that the PRP method with the exact line search was globally convergent. But Powell showed that there existed nonconvex functions on which the PRP method did not converge globally. He suggested that β_k should not be less than zero. Under the sufficient descent condition, Gilbert and Nocedal proved that the modified PRP method $\beta_k = \max\{0, \beta_k^{PRP}\}$ was globally convergent with the Wolfe-Powell line search.

Recently, G. Yu [3] proposed a modified FR (MFR) formula such as

$$\beta_k^{MFR}(\mu) = \frac{\mu_1\|g_k\|^2}{\mu_2\left|g_k^T d_{k-1}\right| + \mu_3\|g_{k-1}\|^2}, \tag{1.6}$$

where $\mu_1 \in (0,+\infty)$, $\mu_2 \in [\mu_1 + \varepsilon_1, +\infty)$, $\mu_3 \in (0,+\infty)$ and ε_1 was an any given positive constant. They proved that for any line search, (1.6) satisfied the condition (1.5), in which $c = 1 - \frac{\mu_1}{\mu_2}$. In fact, the term $\mu_2\left|g_k^T d_{k-1}\right|$ in the denominator of (1.6) played an important role in enhancing descent. It essentially controled the relative weight placed on conjugacy versus descent. Along this way, G. Yu [3] proposed a new nonlinear conjugate gradient formula such as

$$\beta_k^N(\mu)=\begin{cases}\dfrac{\|g_k\|^2-|g_k^T g_{k-1}|}{\mu|g_k^T d_{k-1}|+\|g_{k-1}\|^2} & if\ \|g_k\|^2\ge |g_k^T g_{k-1}|\\ 0 & \text{otherwise,}\end{cases} \tag{1.7}$$

in which $\mu>1$. It possessed the sufficient descent property for any line search, and had an advancement that the directions would approach to the steepest descent directions while the steplength was small. They also proved the algorithm which possed the global convergence property with the weak Wolfe-Powell.

Z. Wei [4] proposed a new nonlinear conjugate gradient formula such as

$$\beta_k^*=\frac{g_k^T\left(g_k-\dfrac{\|g_k\|}{\|g_{k-1}\|}g_{k-1}\right)}{\|g_{k-1}\|^2}, \tag{1.8}$$

In [4], a new conjugate gradient formula β_k^* was given to compute the search directions for unconstrained optimization problems. It was discussed some general convergence results for the proposed formula with some line searches such as the exact line search, the Wolfe-Powell line search and the Grippo-Lucidi line search. The given formula $\beta_k^*\ge 0$, and had the similar form with β_k^{PRP}.

Combining the algorithms above, in this paper, we propose a new modified scalar formula $\beta_k^N(\mu)$, denoted $\beta_k^{MN}(\mu)$, the new algorithm calls *MN* algorithm, where

$$\beta_k^{MN}(\mu)=\frac{\|g_k\|^2-\dfrac{\|g_k\|}{\|g_{k-1}\|}g_k^T g_{k-1}}{\mu|g_k^T d_{k-1}|+\|g_{k-1}\|^2} \tag{1.9}$$

in which $\mu>1$. We add some parameters for $\beta_k^{MN}(\mu)$ so that it is generalization, then we have *VMN* algorithm

$$\beta_k^{VMN}(\mu)=\frac{\mu_1\left(\|g_k\|^2-\dfrac{\|g_k\|}{\|g_{k-1}\|}g_k^T g_{k-1}\right)}{\mu_2|g_k^T d_{k-1}|+\mu_3\|g_{k-1}\|^2}, \tag{1.10}$$

in which $\mu_1\in(0,+\infty)$, $\mu_2\in[\mu_1+\varepsilon_1,+\infty)$, $\mu_3\in(0,+\infty)$, and ε_1 is any given positive number, calls VMN algorithm.

In the next section, we present the global convergence of MN algorithm and establish some good properties for which. The global convergence results of VMN algorithm are given in Section 3. Finally, we have a conclusion section.

2. The Global Convergence of MN Algorithm

Firstly, we can prove $\beta_k^{MN}(\mu)$ in (1.9) is non-negative,

$$\beta_k^{MN}(\mu)=\frac{\|g_k\|^2-\dfrac{\|g_k\|}{\|g_{k-1}\|}g_k^T g_{k-1}}{\mu|g_k^T d_{k-1}|+\|g_{k-1}\|^2}$$

$$\ge\frac{\|g_k\|^2-\dfrac{\|g_k\|}{\|g_{k-1}\|}\|g_k\|\|g_{k-1}\|}{\mu|g_k^T d_{k-1}|+\|g_{k-1}\|^2}=0,$$

The following theorem shows that MN algorithm possesses the sufficient descent property for any line search.

Theorem 2.1. Consider any method (1.1) and (1.2), where $\beta_k=\beta_k^{MN}(\mu)$ .Then for all $k\ge 1$

$$g_k^T d_k\le-\left(1-\frac{1}{\mu}\right)\|g_k\|^2. \tag{2.1}$$

Proof. If $g_{k+1}^T d_k=0$, for $\mu>1$, then we have

$$g_{k+1}^T d_{k+1}=-\|g_{k+1}\|^2\le-\left(1-\frac{1}{\mu}\right)\|g_{k+1}\|^2.$$

Otherwise, from the definition of $\beta_k^{MN}(\mu)$, we can Obtain

$$\beta_k^{MN}(\mu)\le\frac{\|g_{k+1}\|^2}{\mu|g_{k+1}^T d_k|},$$

and then we have

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

From $g_1^T d_1=-\|g_1\|^2<0$, we can deduce that (2.1) holds for all $k\ge 1$.

In order to establish the global convergence result for MN algorithm, we will impose the following assumptions.

Assumption A.

1) The level set

$$\Omega=\{x\in R^n\,|\,f(x)\le f(x_0)\}\ \text{ is bounded.}$$

2) In some neighborhood N of Ω, f is differentiable and its gradient g is Lipschitz continuous, that is to say, there exists a constant $L>0$ such that

$$\|g(x)\le g(y)\|\le L\|x-y\|,\ \forall x,y\in N. \tag{2.2}$$

By using the Assumption A, we can deduce that there exists B and $M>0$ such that

$$\|x\|\le B,\ \ \|g(x)\|\le M,\ \ \forall x\in\Omega. \tag{2.3}$$

The following important result is obtained by Zoutendijk [5].

Lemma 2.2. Suppose that Assumption A holds. Consider any iteration method of the form (1.1) and (1.2), and t_k is obtained by the Wolfe line search (1.4). Then

$$\sum_{k=0}^{\infty} \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} < +\infty. \quad (2.4)$$

Then we will analyse the global convergence property of MN algorithm.

Gilbert and Nocedal [6] introduced the following Property A which pertains to the PRP method under the sufficient descent condition. Now we will show that this Property A pertains to the new method.

Property A. Consider a method of form (1.1) and (1.2). Suppose that

$$0 < \gamma \le \|g_k\| \le \bar{\gamma}. \quad (2.5)$$

We say that the method has Property A, if for all k, there exist constants $b > 1$, $\lambda > 0$ such that $|\beta_k| \le b$ and we have

$$|\beta_k| \le \frac{1}{2b} \text{ if } \|s_{k-1}\| \le \lambda .$$

The following lemma shows that the new method has the Property A.

Lemma 2.3. Consider the method of form (1.1) and (1.2) in which $\beta_k = \beta_k^{MN}(\mu)$. Suppose that Assumption A holds, then the method has Property A.

Proof. Set

$$b = \frac{\bar{\gamma}^2\left(\gamma + \bar{\gamma}\right)}{\gamma^3} > 1, \ \lambda = \frac{\gamma^2}{4L\bar{\gamma}b}.$$

By (1.9) and (2.5) we have

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

$$\le \frac{\|g_k\| \left(\|g_k\| + \frac{\bar{\gamma}}{\gamma} \|g_{k-1}\| \right)}{\|g_{k-1}\|^2}$$

$$\le \frac{\bar{\gamma}\left(\bar{\gamma} + \frac{\bar{\gamma}^2}{\gamma} \right)}{\gamma^2} = \frac{\bar{\gamma}^2\left(\gamma + \bar{\gamma}\right)}{\gamma^3} = b.$$

By the Assumption A (2) and (2.2) hold, if $\|s_{k-1}\| \le \lambda$, then

$$\left|\beta_k^{MN}(\mu)\right| \le \frac{\|g_k\| \left(\|g_k - g_{k-1}\| + \left\| g_{k-1} - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1} \right\| \right)}{\|g_{k-1}\|^2}$$

$$\le \frac{\|g_k\| \left(L\lambda + \left| \|g_{k-1}\| - \|g_k\| \right| \right)}{\|g_{k-1}\|^2}$$

$$\le \frac{\|g_k\| \left(L\lambda + \|g_{k-1} - g_k\| \right)}{\|g_{k-1}\|^2}$$

$$\le \frac{2L\lambda \|g_k\|}{\|g_{k-1}\|^2} = \frac{2L\bar{\gamma}\lambda}{\gamma^2} = \frac{1}{2b}.$$

The proof is finished.

If (2.5) holds and the methods have Property A, then the small steplength should not be too many. The following lemma shows this property.

Lemma 2.4. Suppose that Assumption A and (1.5) hold. Let $\{x_k\}$ and $\{d_k\}$ be generated by (1.1) and (1.2) in which t_k satisfies the Wolfe-Powell line search, β_k has Property A. If (2.5) holds, then, for any $\lambda > 0$, there exist $\Delta \in N^+$ and $k_0 \in N^+$, for all $k \ge k_0$, such that

$$\left|\kappa_{k,\Delta}^{\lambda}\right| \ge \frac{\Delta}{2},$$

where

$$\kappa_{k,\Delta}^{\lambda} = \left\{ i \in Z^+ : k \le i \le k + \Delta - 1, \|s_{i-1}\| \ge \lambda \right\}, \left|\kappa_{k,\Delta}^{\lambda}\right|$$

denotes the number of the $\kappa_{k,\Delta}^{\lambda}$.

Lemma 2.5. Suppose that Assumption A and (1.5) hold. Let $\{x_k\}$ be generated by (1.1) and (1.2), t_k satisfies the Wolfe-Powell line search, and $\beta_k \ge 0$ has Property A. Then,

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

The proofs of Lemmas 2.4 and 2.5 had been given in [7]. By the above three lemmas, it is easy to obtain the following convergence result.

Theorem 2.6. Suppose that Assumption A holds. Let $\{x_k\}$ be generated by (1.1) and (1.2), t_k satisfies the Wolfe-Powell line search, β_k is computed by (1.9), then

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

3. The Global Convergence of VMN Algorithm

In this section we will add some parameters of $\beta_k^{MN}(\mu)$ so that it is generalization, then we have VMN algorithm

$$\beta_k^{VMN}(\mu) = \frac{\mu_1 \left(\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1} \right)}{\mu_2 \left|g_k^T d_{k-1}\right| + \mu_3 \|g_{k-1}\|^2}, \quad (3.1)$$

Table 1. The detail information of numerical experiments for MN algorithm.

No.	x_0	x_k	$\|g_k\|$	k
S201	(8, 9)	(5.00000000000007, 6.00000000000002)	5.338654227543967e−013	2
S202	(15, −2)	(11.41277974501077, −0.89680520867268)	7.343412874359107e−007	30
S205	(0,0)	(3.00000000072742, 0.49999999510645)	2.411787046907220e−007	12
S206	(−1.2, 1)	(1.00000000400789, 1.00000020307759)	3.907063255896894e−007	5
S311	(1, 1)	(−3.77931025686871, −3.28318599743028)	5.006611497285485e−007	6
S314	(2, 2)	(1.79540285273286, 1.37785978124895)	7.487443339742852e−008	6

in which

$$\mu_1 \in (0,+\infty), \quad \mu_2 \in [\mu_1+\varepsilon_1,+\infty), \quad \mu_3 \in (\mu_1,+\infty),$$

ε_1 is any given positive number.

Theorem 3.1. Suppose that Assumption A holds. Let Let $\{x_k\}$ be generated by (1.1) and (1.2), t_k satisfies the Wolfe-Powell line search, and $\beta_k \geq 0$ has Property A. Then,

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

Similar to the second part of the discussion for the above general algorithm $\beta_k^{VMN}(\mu)$, we can get $\beta_k^{MN}(\mu) \geq 0$. And the algorithm possesses the sufficient descent property, in which

$$c = 1 - \frac{\mu_1}{\mu_2}$$

The proof is similar as the one of Theorem 2.1 in Section 2.

4. Numerical Experiments

In this section, we carry out some numerical experiments. The *MN* algorithm has been tested on some problems from [8]. The results are summarized in **Table 1**. For the test problem, No. is the number of the test problem in [8], x_0 is the initial point, x_k is the final point, k is the number of times of iteration for the problem.

Table 1 shows the performance of the *MN* algorithm relative to the iteration. It is easily to see that, for all the problems, the algorithm is very efficient. The results for each problem are accurate, and with less number of times of iteration.

5. Conclusions

In this paper, we have proposed a new nonlinear conjugate gradient method-MN algorithm. The sufficient descent property holds without any line searches, and the algorithm satisfys Property A. We also have proved, employing some steplength technique which ensures the Zoutendijk condition to be held, this method is globally convergent. Judging from the numerical experiments in **Table 1**, compared to most other algorithms, *MN* algorithm has higher precision and less number of times of iteration. Finally, we have proposed *VMN* algorithm, it also have the sufficient descent property and Property A, and it is global convergence under weak Wolfe-Powell line search.

6. References

[1] M. Al-Baali, "Descent Property and Global Convergence of the Fletcher-Reeves Method with Inexact Line Search," *IMA Journal of Numerical Analysis*, Vol. 5, No. 1, 1985, pp. 121-124. doi:10.1093/imanum/5.1.121

[2] Y. F. Hu and C. Storey, "Global Convergence Result for Conjugate Gradient Method," *Journal of Optimization Theory and Applications*, Vol. 71, No. 2, 1991, pp. 399-405. doi:10.1007/BF00939927

[3] G. Yu, Y. Zhao and Z. Wei, "A Descent Nonlinear Conjugate Gradient Method for Large-Scale Unconstrained Optimization," *Applied Mathematics and Computation*, Vol. 187, No. 2, 2007, pp. 636-643. doi:10.1016/j.amc.2006.08.087

[4] Z. Wei, S. Yao and L. Liu, "The Convergence Properties of Some New Conjugate Gradient Methods," *Applied Mathematics and Computation*, Vol. 183, No. 2, 2006, pp. 1341-1350. doi:10.1016/j.amc.2006.05.150

[5] G. Zoutendijk, "Nonlinear Programming, Computational Methods," In: J. Abadie, Ed., *Integer and Nonlinear Programming*, North-Holland Publisher Co., Amsterdam, 1970, pp. 37-86.

[6] J. C. Gilbert and J. Nocedal, "Global Convergence Properties of Conjugate Gradient Methods for Optimization," *SIAM Journal Optimization*, Vol. 2, No. 1, 1992, pp. 21-42. doi:10.1137/0802003

[7] Y. H. Dai and Y. Yuan, "Nonlinear Conjugate Gradient

Methods," Shanghai Scientific and Technical Publishers, Shanghai, 1998, pp. 37-48.

[8] W. Hock and K. Schittkowski, "Test Examples for Nonlinear Programming Codes," *Journal of Optimization Theory and Applications*, Vol. 30, No. 1, 1981, pp. 127-129. doi:10.1007/BF00934594