ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 7, No. 3, 2012, *pp*. 212-217

A New Sufficient Descent Conjugate Gradient Method for Unconstrained Optimization

Sun Min [1+] and Liu Jing [2]

[1] School of Mathematics and Statistics, Zaozhuang University, Shandong 277160, China
[2] School of Mathematics and Statistics, Zhejiang University of Finance and Economics, Hangzhou, 310018, China

(*Received March 9, 2012, accepted July 10, 2012*)

Abstract. In this paper, a new conjugate conjugate method with sufficient descent property is proposed for the unconstrained optimization problem. An attractive property of the new method is that the descent direction generated by the method always possess the sufficient descent property, and this property is independent of the line search used and the choice of β_{ki}. Under mild conditions, the global convergence of the new method is proved.

Keywords: conjugate gradient method, sufficient descent property, global convergence

1. Introduction

In this paper, we consider the following unconstrained nonlinear optimization problem

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

where $f(x): R^n \to R^1$ is a continuously differentiable function whose gradient is denoted by $g(x)$.

The conjugate gradient methods are welcome methods for solving optimization problem (1). They are particularly efficient for solving large scale problems due to their simplicity and low storage [1,2]. A nonlinear conjugate gradient method generates a sequence $\{x_k\}$, starting from an initial point $x_0 \in R^n$, using the recurrence

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

where α_k is called the step size which is determined by some line search and d_k is the search direction defined by the rule

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

Here g_k is an abbreviation of $g(x_k)$ and β_k is a scalar which results in distinct conjugate gradient methods. Well known conjugate gradient methods include the Fletcher-Reeves(FR) method[1], the Polak-Ribiere-Polyak(PRP) method[2,3], the Dai-Yuan(DY) method[4] and the Conjugate Descent(CD) method[5] in which β_k are specified by

$$\beta_k^{FR} = \frac{\| g_k\|^2}{\| g_{k-1}\|^2}, \beta_k^{PRP} = \frac{g_k^T (g_k - g_{k-1})}{d_{k-1}^T (g_k - g_{k-1})} \quad \beta_k^{DY} = \frac{\| g_k\|^2}{d_{k-1}^T (g_k - g_{k-1})}, \beta_k^{CD} = -\frac{\| g_k\|^2}{g_{k-1} d_{k-1}}$$

The global convergence of above conjugate gradient methods have bee studied by many researchers, see, for instance [6] and references therein.

Descent property, that is $g_k^T d_k < 0$, is very important for an iterative method to be globally convergent. If

---

[+] Corresponding author. Tel.: +86-0632-3786733.
E-mail address: ziyouxiaodou@163.com.

the step size α_k is determined by exact line search, we have $g_k^{\bullet} d_{k-1} = 0$. Then from (3), we get

$$g_k^{\bullet} d_k = -\| g_k \|^2, \tag{4}$$

which indicates that d_k satisfies the sufficient descent condition:

$$g_k^{\bullet} d_k \le -c\| g_k \|^2,$$

for a positive constant c, independently of line search. However, it is difficult or time consuming to implement an exact line search for seeking step length in practical computation. Therefore, in most existing conjugate gradient methods, the direction d_k determined by (3) may not be descent direction of f at x_k if inexact line search such as the Armijo line search is used.

Recently, Hager and Zhang[11] proposed a new conjugate gradient method which was obtained by modifying the HS method and called CG-DESCENT method. A nice property of the CG-DESCENT method is that it generates sufficient descent directions which is independent of the line search used. Then, Zhang Li et al.[12-13] also proposed some modified conjugate gradient methods which also possess the above property. Specially, at each iteration, the generated directions satisfy $g_k^{\bullet} d_k = -\| g_k \|^2$.More recently, using the Gram-Schmidt orthogonalization to d_{k-1} and g_k, Cheng and Liu [9] proposed a sufficient descent nonlinear conjugate gradient method defined by

$$d_k = \begin{cases} -g_k, & k = 0, \\ -g_k + \beta_k \left(d_{k-1} - \dfrac{g_k^{\bullet} d_{k-1}}{\| g_k \|^2} g_k \right), & k \ge 1. \end{cases} \tag{5}$$

It is clear that the direction in (5) satisfies (4). The sufficient descent property is independent of the line search used and the choice of β_k. Moreover, the scheme (5) reduces to the standard conjugate gradient method if exact line search is used.

Similar to conjugate gradient methods, memory gradient methods can also avoid storing and remembering matrices. The main difference between them is that the latter can use the information of the previous multi-step iterations more sufficiently and hence it is helpful to design algorithms with quick convergence rate. See Cragg and Lery[7], Miele and Cantrell[8].

In this paper, we continue Cheng and Liu's research and extend their method to memory gradient type method, which uses the previous m step information sufficiently. Under mild conditions, the global convergence of the new method is proved.

The paper is organized as follows: In Sect.2, we propose the new conjugate gradient method and show the global convergence. In Sect.3, we give the numerical tests and the conclusion is given in Sect.4..

2. Algorithm and Global Convergence

Let m denotes the number of the past iterations remembered. By using the Gram-Schmidt orthogonalization to d_{k-i} and g_k $(i = 1, 2, \cdots, m)$, we define the search direction as follows:

$$d_k = -g_k + \frac{1}{m} \sum_{i=1}^{m} \beta_{ki} \left(d_{k-i} - \frac{g_k^{\bullet} d_{k-i}}{\| g_k \|^2} g_k \right), \tag{6}$$

It is clear that the direction in (6) also satisfies the sufficient descent property (4), which is independent of the line search used and the choice of β_{ki}. If β_k in (6) is specified by some existing conjugate gradient method, we get a corresponding modified conjugate gradient method. In the following, we propose a specific conjugate gradient method whose β_{ki} is defined by

$$\beta_{ki} = \frac{g_k^{\top}\left(g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1}\right)}{|g_k^{\top} d_{k-i}| + \|g_k\| \cdot \|d_{k-i}\|}. \tag{7}$$

Lemma 2.1. Let d_k be determined by (6) with β_{ki} defined by (7), then we have

$$\|d_k\| \le 3\|g_k\|. \tag{8}$$

Proof. We have from (6) and (7) that

$$\begin{aligned}
&\|d_k\| \\
&\le \|g_k\| + \frac{1}{m}\sum_{i=1}^{m} \beta_{ki}\left(\|d_{k-i}\| + \frac{|g_k^{\top} d_{k-i}|}{\|g_k\|^2}\|g_k\|\right) \\
&= \|g_k\| + \frac{1}{m}\sum_{i=1}^{m} \beta_{ki}\left(\frac{\|d_{k-i}\| \cdot \|g_k\| + |g_k^{\top} d_{k-i}|}{\|g_k\|}\right) \\
&= \|g_k\| + \frac{1}{m}\sum_{i=1}^{m} 2\|g_k\| \\
&= 3\|g_k\|
\end{aligned}$$

This complete the proof.

Now we propose our sufficient descent conjugate gradient method:

Algorithm 2.1

Step 1: Choose $x_0 \in R^n, \gamma \in (0,1), \rho \in (0,1), \epsilon > 0$, and set $k = 1$.

Step 2: If $\|g_k\| < \epsilon$, then stop.

Step 3: Compute d_k by (6) and α_k by the following Armijo line search, that is, the step length $\alpha_k = \max\{\rho^j, j = 0,1,2,\cdots\}$ satisfying

$$f(x_k + \alpha_k d_k) - f(x_k) \le \gamma \alpha_k g_k^{\top} d_k. \tag{9}$$

Step 4: Set $x_{k+1} = x_k + \alpha_k d_k$, $k := k+1$, and go to Step 2.

Throughout the paper, we make the following assumptions.

(H1): The objective function f has lower bound on the level set $L_0 = \{x \in R^n \mid f(x) \le f(x_1)\}$.

(H2): In some neighborhood N of L_0, the gradient g is Lipschitz continuous, i.e., there exists an

$L > 0$ such that

$$\| g(x) - g(y) \| \le L \| x - y \|. \tag{10}$$

The conclusion of the following lemma, often called Zoutendijk condition, is used to prove the global convergence of nonlinear conjugate gradient method.

Lemma 2.2. If (H1), (H2) hold and Algorithm 2.1 generates an infinite sequence $\{x_k\}$, then

$$\sum_{k=1}^{\infty} \frac{\| g_k \|^4}{\| d_k \|^2} < +\infty. \tag{11}$$

Proof. First we consider the following two cases: $\alpha_k = 1$ and $\alpha_k < 1$. In the first case, from (4), we have $\| g_k \| \le \| d_k \|$. Then, we get

$$\alpha_k \ge \frac{\| g_k \|^2}{\| d_k \|^2}. \tag{12}$$

If $\alpha_k < 1$, this implies that α_k / ρ violates (9). Thenwe have

$$f(x_k + \alpha_k d_k / \rho) > f(x_k) + \gamma \alpha_k g_k^{\bullet} d_k / \rho.$$

Using the mean value theorem in the above inequality, we obtain $\theta_k \in (0,1)$, such that

$$[g(x_k + \theta_k \alpha_k d_k / \rho) - g_k]^{\bullet} d_k > (1-\gamma) \| g_k \|^2.$$

By (H2), we have $L\alpha_k \| d_k \|^2 / \rho > (1-\gamma) \| g_k \|^2$. Therefore

$$\alpha_k \ge \frac{\rho(1-\gamma) \| g_k \|^2}{L \| d_k \|^2}. \tag{13}$$

Letting $\nu = \min\{1, \rho(1-\gamma)/L\}$, from (12), (13) we obtain

$$\alpha_k \ge \nu \frac{\| g_k \|^2}{\| d_k \|^2}, \forall k \tag{14}$$

From (9) and (14), we have $\gamma \nu \frac{\| g_k \|^4}{\| d_k \|^2} \le f_k - f_{k+1}$, which, together with (H1), we get (11). The proof is completed.

Theorem 2.1. If (H1)(H2) hold and Algorithm generates an infinite sequence $\{x_k\}$, then

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

Proof. From (11), we have

$$\lim_{k \to \infty} \frac{\| g_k \|^2}{\| d_k \|} = 0. \tag{15}$$

From (8), we get $\frac{\|g_k\|^2}{\|d_k\|} \geq \frac{\|g_k\|}{2}$, which, together with (15), we get the desired result. This completes the proof.

3. Numerical results

In this section, we report the detailed numerical results of a number of problems by Algorithm 2.1. We test the following two CG methods:

PRP: the PRP formula with the Armijo conditions, where $\gamma = 0.1, \rho = 0.5$, the termination condition is $\|g_k\| \leq 10^{-6}$.

NM: Algorithm 2.1 with the with the Armijo conditions, where $\gamma = 0.06, \rho = 0.4$, the termination condition is $\|g_k\| \leq 10^{-6}$.

Problem 1

$$f(x) = 10(x_1^2 - x_2)^2 + (1 - x_1)^2 + 9(x_4 - x_3^2)^2$$
$$+(1 - x_3)^2 + 10.1[(x_2 - 1)^2 + (x_4 - 1)^2] + 19.8(x_2 - 1)(x_4 - 1).$$

The initial point: $x_0 = (-3, -1, -3, -1)$.

Problem 2

$$f(x) = (x_1 + 10x_2)^4 + 5(x_3 - x_4)^4 + (x_2 - 2x_3)^4 + 10(x_1 - x_4)^4.$$

The initial point: $x_0 = (2, 2, -2, -2)$.

Problem 3

$$f(x) = (1 - x_1)^2 + \sum_{i=1}^{9} (x_i^2 - x_{i+1}^2) + (1 - x_{10})^2.$$

The initial point: $x_i = 0, i = 1, \cdots, 10$.

In the following table, the numerical results are written in the form NI/CPU, where NI, CPU denote the number of iterations, the CPU time respectively. Dim denotes the dimension of the test problems.

TABLE 1: Numerical results for Problems 1-3

P	Dim	NM	PRP
1	4	244/0.0300	309/0.0300
2	4	161/0.0200	161/0.0200
3	10	203/0.0200	135/0.0200

From the numerical results, we can see that the two methods are similar.

4. Conclusion

In this paper, we propose a new sufficient descent conjugate gradient method for unconstrained optimization. Under mild conditions, we prove its global convergence, and give some numerical results to illustrate its efficiency.

Acknowledgment

This work is supported by the Foundation of Zhejiang Provincial Education Department Under Grant Y201225096.

5. References

[1] R, Fletcher, and C, Reeves, *Function minimization by conjugate gradients*, Computer Journal, 7(1964), pp. 149-154.

[2] B, Polak, and G,Ribiere, *Note sur la convergence des methods de directions conjuguees*, Reue Francaise de

Recherche Operationnele, 16(1969), pp. 35-43.

[3] B, Polayk, *The conjugate gradient method in extreme problems*, USSR Computational Mathematics and Mathematical Physics, 9(1969), pp. 94-112.

[4] Dai Y. H., and Yuan Y. X., *A nonlinear conjugate gradient method with a strong global convergence property*, SIAM Journal on optimization, 10(2000), pp. 177-182.

[5] R, Fletcher, *Practical Methods of Optimization*, Vol. 1: Unconstrained Optimization(New York: Wiley and Sons), 1987.

[6] W, Hager, Zhang H.C., *A survey of nonlinear conjugate gradient methods*, Pacific Journal of Optimization, 2(2006), pp. 35-58.

[7] E, Cragg, and A, Levy, *Study on supermemory gradient method for the minimization of function*, Journal of Optimization Theory and Application, 4(1969), pp. 191-205.

[8] A, Miele, and J, Cantrell, *Study on a memory gradient method for the minimization of function*, Journal of Optimization Theory and Application, 3(1969), pp. 459-470.

[9] Cheng W.Y., and Liu Q.F., *Sufficient descent nonlinear conjugate gradient methods with conjugacy condition*, Numerical Algorithms, 53(2010), pp. 113-131.

[10] Shi Z.J., and Shen J., *Convergence of Liu-Storey conjugate gradient method*, European Journal of Operational Research, 182(2007), pp. 552-560.

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

[12] Zhang L., Zhou W.J., Li D.H., *A descent modified Polak-Ribiere-Polyak conjugate gradient method and its global convergence*, IMA Journal of Numerical Analysis, 26(2006), pp. 629-640.

[13] Zhang L., Zhou W.J., Li D.H., Global convergence of a modified Fletcher-Reeves conjugate method with Armijo-type line search, Numerische Mathematik, 104(2006), pp. 561-572.