A New Nonlinear Conjugate Gradient Method Based on the Scaled Matrix

Basim A. Hassan ✉, **Haneen A. Alashoor**
Department of Mathematics, College of Computers Sciences and Mathematics

Article Info
Received 7/6/2015
Accepted 7/9/2015

ABSTRACT
In this paper, a new type nonlinear conjugate gradient method based on the Scale Matrix is derived. The new method has the decent and globally convergent properties under some assumptions. Numerical results indicate the efficiency of this method to solve the given test problems.

Keywords: Conjugate gradient, Descent condition, global convergent, Numerical results.

الخلاصة
تم في هذا البحث اشتقاق نوع جديد من طريقة التدرج المترافق غير خطية المعتمدة على المصفوفة القياسية. الطريقة الجديدة تمتلك خاصيتي الانحدار والتقارب الشامل تحت بعض الفرضيات. وأشارت النتائج العددية إلى كفاءة هذه الطريقة في حل دوال الاختبار المعطاة.

INTRODUCTION

In this paper, we consider the unconstrained optimization problem :

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

where f is smooth and its gradient g is available. For solving this problem, starting from an initial guess , $x_0 \in R^n$.

Conjugate gradient methods are very efficient for solving large-scale unconstrained optimization problems (1). The iterates of conjugate gradient methods are obtained by:

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

with

$$d_{k+1} = \begin{cases} -g_{k+1} & \text{if } k = 0 \\ -g_{k+1} + \beta_k d_k & \text{if } k > 0 \end{cases} \tag{3}$$

where step size α_k is positive, $g_{k+1} = \nabla f(x_{k+1})$ and β_k is generated by the definition of conjugated directions :

$$d_{k+1}^T G d_k = 0 \tag{4}$$

where G is the Hessian of f at the point x_{k+1} In addition, α_k is a step length which is computed by carrying out some line search. In this paper we analyze the general results on convergence of line search methods with the following line search rules. We consider line searches that satisfy the Wolfe (WP) conditions:

$$f(x_k + \alpha_k d_k) \le f(x_k) + \delta_1 \alpha_k d_k^T g_k \tag{5}$$

$$g(x_k + \alpha_k d_k)^T d_k \ge \delta_2 d_k^T g_k \tag{6}$$

where d_k is descent direction, i.e. :

$$g_k^T d_k < 0 \tag{7}$$

and $0 < \delta_1 \le \delta_2 < 1$. More performance profile, is given in [3].

Different conjugate gradient algorithms correspond to different choices for the scalar parameter β_k . We list some of these methods as Fletcher and Reeves (FR) [7], Dai and Yuan (DY) [5] and Conjugate Descent (CD) [6] and their scalar parameters:

$$\beta_k^{FR}=\frac{g_{k+1}^T g_{k+1}}{g_k^T g_k},\beta_k^{DY}=\frac{g_{k+1}^T g_{k+1}}{y_k^T d_k},\beta_k^{CD}=-\frac{g_{k+1}^T g_{k+1}}{g_k^T d_k} \qquad (8a)$$

These methods have strong convergence properties, but they may have modest practical performance due to jamming. On the other hand, the methods of Polak and Ribiere (PR) [11], Hestenes and Stiefel (HS) [8], and Liu and Storey, (LS) [9] with the parameters:

$$\beta_{k+1}^{PRP}=\frac{g_{k+1}^T y_k}{\|g_k\|^2},\beta_k^{HS}=\frac{g_{k+1}^T y_k}{y_k^T d_k},\beta_k^{LS}=-\frac{g_{k+1}^T y_k}{g_k^T d_k}, \qquad (8b)$$

in general, may not be convergent, but they often have better computational performances. Dai and Liao [4] used the secant condition of quasi-Newton methods that is

$$v_k = B_{k+1} y_k \qquad (9)$$

where B_{k+1} is an approximation of the inverse Hessian. It is symmetric positive definite. For quasi-Newton methods, the search direction d_{k+1} can be calculated in the form:

$$d_{k+1} = -B_{k+1} g_{k+1} \ . \qquad (10)$$

It is necessary to modify and extend these methods to make them suiTable for large problems [10]. For this reason, in the present article we prefer the scalar matrix approach. By taking the scalar approximation to the Hessian inverse:

$$B_{k+1} = \gamma_{k+1} I \cong G^{-1}, \gamma_{k+1} > 0 \ . \qquad (11)$$

If we apply the secant Equation to the chosen approximations x_k and x_{k+1}, our task is to find a scalar γ_{k+1} satisfying:

$$v_k = \gamma_{k+1} y_k \ . \qquad (12)$$

The corresponding scalar matrix $B_{k+1} = \gamma_{k+1} I$. Then the direction:

$$d_{k+1} = -\gamma_{k+1}\ g_{k+1} \ . \qquad (13)$$

The structure of the paper is as follows. In section 2, we present the new conjugate gradient method and new Algorithm. Section 3 global Convergence is studied. In Section 4, numerical results are presented and in Section 5 we give brief conclusions and discussions.

Derivation a new formulae β_k and Algorithm

In this section, we derive a new conjugate gradient method based expression of the denominator $d_k^T G v_k$. From (3) and (4) we get:

$$\begin{aligned} d_{k+1}^T G d_k &= (-g_{k+1} + \beta_k d_k)^T G d_k \\ &= -g_{k+1}^T G d_k + \beta_k d_k^T G d_k = 0 \end{aligned} \qquad (14)$$

as a result,

$$\beta_k = \frac{g_{k+1}^T G v_k}{d_k^T G v_k} \ . \qquad (15)$$

On the other hand, a class of the quasi-Newton condition where the approximation of the Hessian inverse is presented by appropriate scalar matrix. This scheme is also analyzed in [2], where the step length γ_{k+1} is computed by:

$$\gamma_{k+1} = \|v_k\| / \|y_k\| \ . \qquad (16)$$

The corresponding scalar matrix:

$$G = \frac{1}{\gamma_{k+1}} I_{n*n} \ . \qquad (17)$$

Therefore, we can define a new denominator nonlinear conjugate gradient method:

$$\begin{aligned} d_k^T G v_k &= d_k^T \left(\gamma_{k+1}^{-1} I_{n*n}\right) v_k \\ &= \gamma_{k+1}^{-1} d_k^T \left(I_{n*n}\right) v_k \\ &= \gamma_{k+1}^{-1}\ d_k^T v_k \end{aligned} \qquad (18)$$

from (15) and (18) we get a new formula denoted by β_k^{BSI} :

$$\beta_k^{BSI} = \frac{g_{k+1}^T y_k}{\delta_{k+1} d_k^T v_k} \tag{19}$$

where $\delta_{k+1} = \gamma_{k+1}^{-1}$. If the line search is exact, then :

$$\beta_k^{BSI} = \frac{g_{k+1}^T g_{k+1}}{\delta_{k+1} d_k^T v_k} \tag{20}$$

This completes the derivation.

Now we can obtain the new conjugate gradient algorithm, as follows :

New Algorithm (BSI Algorithm) :

Step 1. Initialization. Select $x_1 \in R^n$ and the parameters $0 < \delta_1 < \delta_2 < 1$. Compute $f(x_1)$ and g_1. Consider $d_1 = -g_1$ and set the initial guess $\alpha_1 = 1/\|g_1\|$.

Step 2. Test for continuation of iterations. If $\|g_{k+1}\| \le 10^{-6}$, then stop.

Step 3. Line search. Compute $\alpha_{k+1} > 0$ satisfying the Wolfe line search condition (4) and (5) and update the variables $x_{k+1} = x_k + \alpha_k d_k$.

Step 4. Compute β_k conjugate gradient parameter which defined in (20).

Step 5. Compute the search direction d_{k+1} as in (3). If the restart criterion of Powell $|g_{k+1}^T g_k| \ge 0.2\|g_{k+1}\|^2$, is satisfied, then set $d_{k+1} = -g_{k+1}$ otherwise set $k = k+1$ and continue with step 2.

Convergence analysis

In [3] and throughout this section we assume that f is strongly convex and Lipschitz continuous on the level set

$$L_0 = \{x \in R^n : f(x) \le f(x_0)\}. \tag{21}$$

Assumption A : There exist constants $\mu > 0$ and L such that :

$$(\nabla f(x) - \nabla f(y))^T (x - y) \ge \mu \|x - y\|^2 \tag{22}$$

and

$$(\nabla f(x) - \nabla f(y)) \le L\|x - y\| \tag{23}$$

for all x and y from L_0.

Now we have the following theorem, which illustrates that the new conjugate gradient method can guarantee the descent property with the Wolfe line searches.

Theorem (3.1) :

Suppose that α_k in (2) satisfies the Wolfe conditions (4) and (5), then the direction d_{k+1} given by (3) with (20) is a descent direction.

Proof :

Since $d_0 = -g_0$, we have $g_0^T d_0 = -\|g_0\|^2 \le 0$. Suppose that $g_k^T d_k < 0$ for all $k \in n$. Since δ_{k+1} and $d_k^T v_k$ are positive then $\delta_{k+1} d_k^T v_k > 0$. Thus, we have $\beta_k = \|g_{k+1}\|^2 / \delta_{k+1} d_k^T v_k > 0$. If $g_{k+1}^T d_k \le 0$, we get :

$$\begin{aligned} g_{k+1}^T d_{k+1} &= g_{k+1}^T(-g_{k+1} + \beta_k g_{k+1}^T d_k) \\ &\le -\|g_{k+1}\|^2 < 0 \ . \end{aligned} \tag{24}$$

If $g_{k+1}^T d_k > 0$, by $\delta_{k+1} d_k^T v_k \ge d_k^T y_k = g_{k+1}^T d_k - g_k^T d_k$ and $g_k^T d_k < 0$, we find that $\delta_{k+1} d_k^T v_k \ge g_{k+1}^T d_k$, and hence, $g_{k+1}^T d_k / \delta_{k+1} d_k^T v_k < 1$. So, we have :

$$\begin{aligned} g_{k+1}^T d_{k+1} &= g_{k+1}^T(-g_{k+1} + \beta_k g_{k+1}^T d_k) \\ &\le -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{\delta_{k+1} d_k^T v_k} g_{k+1}^T d_k \\ &\le \left[-1 + \frac{g_{k+1}^T d_k}{\delta_{k+1} d_k^T v_k}\right] \|g_{k+1}\|^2 \\ &< [-1 + 1]\|g_{k+1}\|^2 = 0 . \end{aligned} \tag{25}$$

The following Lemma is the result for general iterative methods :

Lemma (3.1) [12,13]

Suppose that **Assumption A** is satisfied and consider any method with Equation (2) where α_k satisfies Equation s (4) and (5). Then,

$$\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty \ . \tag{26}$$

Assumption A and **Lemma (3.1)** lead to the following lemma. This lemma is immediately obtained by using the result.

Lemma(3.2) [5,13]
Suppose that Assumption A hold and that β_k satisfies

$$0 \le \beta_{k-1} \le \frac{g_k^T d_k}{g_{k-1}^T d_{k-1}} \quad , \; k \ge 2 \,. \tag{27}$$

Then, the (x_k) generated by Equation s (2) and (3) either terminates at a stationary point or converges in the sense that :

$$\lim_{k\to\infty} \inf \| g_k \| = 0 \;. \tag{28}$$

Now, we shall present the convergence analysis of our Algorithm **BSI** :

Theorem (3.2) :
Suppose that **Assumption A** holds. Then, Algorithm BSI either terminates at a stationary point or converges in the sense that $\lim_{k\to\infty} \inf \| g_k \| = 0$.

Proof : In equality mean that, for all $k \in n$.

$$\delta_{k+1} d_k^T v_k > 0 \tag{29}$$

Hence, we get $\beta_k > 0$ for all $k \in n$. Moreover, by Theorem (3.1), we have that, for all $k \in n$.

$$\begin{aligned} \delta_{k+1} d_k^T v_k - g_{k+1}^T d_k &\ge d_k^T y_k - g_{k+1}^T d_k \\ &= g_{k+1}^T d_k - g_k^T d_k - g_{k+1}^T d_k \\ &= -g_k^T d_k > 0 \end{aligned} \tag{30}$$

The definition of search direction and Formula (20) ensure that :

$$\begin{aligned} g_{k+1}^T d_{k+1} &= -g_{k+1}^T g_{k+1} + \beta_k g_{k+1}^T d_k \\ &= (-\delta_{k+1} d_k^T v_k + g_{k+1}^T d_k) \beta_k , \end{aligned} \tag{31}$$

and hence, for all $k \in n$,

$$0 < \beta_k = \frac{-g_{k+1}^T d_{k+1}}{\delta_{k+1} d_k^T v_k - g_{k+1}^T d_k} \le \frac{-g_{k+1}^T d_{k+1}}{-g_k^T d_k} \tag{32}$$

These facts guarantee that β_k satisfies the condition in Lemma (3.2). Therefore, Algorithm **BSI** is globally convergent.

Numerical Results

In this section, we report some results of the numerical experiments. We test the new algorithm and compare its performance with those of FR method whose results be given by [7] by solving the 15 benchmark problems from [1].
All codes of the computer procedures are written in Fortran. The parameters are chosen as follows :
$\varepsilon = 10^{-6}$, $\xi = 0.5$, $\delta_1 = 0.001$, $\delta_2 = 0.9$
In the following Tables, the numerical results are written in the form NOI/IRS, where NOI, IRS denote the number of iteration, the number of restart respectively, n denotes the dimension of the test problems.
From the numerical results, we can see that the proposed method performs better than the FR method for some problems.

Table 6: Comparison of the algorithms for $n = 100$.

Test Problems	**FR - algorithm**		**New algorithm**	
	NOI	**IRS**	**NOI**	**IRS**
Hager	61	45	55	38
Extended Three Expo Terms	15	6	13	8
Generalized Tridiagonal 2	37	8	44	17
Extended PSC1	15	9	9	6
EDENSCH (CUTE)	69	50	52	32
ENGVAL1 (CUTE)	34	16	30	12
DENSCHNA (CUTE)	20	11	16	10
DENSCHNC (CUTE)	49	22	14	9
DENSCHNB (CUTE)	12	7	8	5
Extended Block-Diagonal	122	62	15	9
Generalized quartic GQ1	11	6	9	6
SINCOS	15	9	9	6
Generalized quartic GQ2	112	55	45	18
HIMMELBH (CUTE)	11	5	7	3
Trigonometric	19	12	20	12

Total	602	323	346	191

Table 7: Comparison of the algorithms for $n = 1000$.

Test Problems	FR algorithm		New algorithm	
	NOI	IRS	NOI	IRS
Hager	535	509	396	357
Extended Three Expo Terms	127	117	8	5
Generalized Tridiagonal 2	73	27	53	23
Extended PSC1	8	6	7	5
EDENSCH (CUTE)	98	82	63	49
ENGVAL1 (CUTE)	142	126	134	118
DENSCHNA (CUTE)	19	11	15	9
DENSCHNC (CUTE)	129	67	17	11
DENSCHNB (CUTE)	12	7	8	5
Extended Block-Diagonal	130	66	16	10
Generalized quartic GQ1	9	5	9	4
SINCOS	8	6	7	5
Generalized quartic GQ2	110	54	50	21
HIMMELBH (CUTE)	11	5	7	3
Trigonometric	38	23	37	20
Total	**1449**	**1111**	**827**	**645**

Conclusions and Discussions

In this paper, we have proposed new a nonlinear CG- algorithms based on the Scaled Matrix defined by (17) under some assumptions. The new algorithm has been shown to be globally convergent and satisfies the descent property. The computational experiments show that the new kinds given in this paper are successful. Table 6 and Table 7 give comparisons between the new-algorithm and FR algorithms whose results be given by [7] for convex optimization. Table 8 saves the new algorithm of $(42)\%$ NOI and $(41)\%$ IRS of overall against the standard FR algorithm, especially for our selected group of test problems.

Table 8: Relative efficiency of the new Algorithm.

Tools	NOI	IRS
Algorithm with β_k^{FR}	100 %	100 %
New Algorithm	57.19 %	58.29 %

References

[1] Andrie N. (2008) ' An Unconstrained Optimization Test functions collection ' Advanced Modeling and optimization. 10, pp.147-161.

[2] Barzilai. J. Borwein. J. M. (1988). Two point step size gradient method IMA J.Numer Numer. Anal. 8, pp. 141–148.

[3] Cao W. and Wang K. (2010) ' Global convergence of a new conjugate gradient method for modified Liu-Storey formula' Journal of East China Normal University (Natural Science), 1, pp.44-51.

[4] Dai Y. and Liao Z. (2001). New conjugate conditions and related nonlinear conjugate gradient methods. Appl. Math. Optim. 43, pp. 87-101.

[5] Dai, Y. and Yuan Y. (1999) ' A non-linear conjugate gradients method with a strong global convergence property' SIAM J. Optimization,10, pp.177-182.

[6] Fletcher, R. (1987) ' Practical Methods of Optimization (second edition). John Wiley and Sons, New York.

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

[8] Hestenes, M. R. and Stiefel E. L. (1952) ' Method of conjugate gradients for solving linear systems' J. Research Nat. Standards 49, pp.409-436.

[9] Liu Y. and Storey C. (1991) ' Efficient generalized conjugate gradients algorithms ' Part 1 : Theory. J. Optimization Theory and Applications 69, pp.129-137.

[10] Nocedal J. and Wright S. J. (1999). Numerical Optimization. Springer. New York .

[11] Polak, E. and Ribiere, G. (1969) ' Note for Convergence Direction Conjugate, Revue Francaise Informant, Research. Operational, 16, 35-43.

[12] Zoutendijk G., 1970. Nonlinear Porgramming, computational methods, in: Integr and Nonlinear Programming, North-Holland, Amsterdam, pp. 37-86.

[13] Yasushi N. and Hideaki I., (2011). Conjugate gradient methods using value of objective function for unconstrained optimization Optimization Letters, V.6, Issue 5, 941-955.