Modified the CG-Algorithm for Unconstrained Non-Linear Optimization by Using Oren's Update

Abbas Y. Al-Bayati   **Abdulghafor M. Al-Rozbayani**

abdulghafor_rozbayani@uomosul.edu.iq

College of Computer Sciences and Mathematics, University of Mosul**, Iraq**

ABSTRACT

In this paper we have modified a new extended generalized conjugate gradient steps with self-scaling variable metric updates for unconstrained optimization. The new proposed algorithm is based on the inexact line searches and it is examined by using different non-linear test functions with various dimensions.

Keywords: Conjugate Gradient, Unconstrained Optimization, Nonlinear Optimization, Variable Metric.

تطوير خوارزمية التدرج المترافق للأمثلية غير المقيدة اللاخطية باستخدام تحديث أورين

عباس يونس البياتي   **عبدالغفور محمد امين الروژبياني**

كلية علوم الحاسوب والرياضيات/جامعة الموصل/العراق

الملخص

في هذا البحث تم تطوير خوارزمية التدرج المترافق المعمم الموسعة بخوارزمية المتري المتغير ذاتي القياس في الأمثلية اللاخطية. هذه الخوارزمية تستعمل خطوط بحث غير تامة, والحسابات العددية بينت أن الخوارزمية المطورة أكفأ من عدد من الخوارزميات الأخرى في هذا المجال وباستخدام عشرة دوال لا خطية معروفة في هذا الحقل.

الكلمات المفتاحية: خوارزمية التدرج المترافق, الامثلية اللامقيدة, الامثلية غير الخطية, خوارزمية المتري المتغير , خطوط بحث تامة.

1.**Introduction:**

We try to solve the unconstrained minimization problem

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

when f is twice continuously differentiable function. This problem is usually solved iteratively. Starting with an initial estimate x_1 of the minimum point x^*, each subsequent point x_{k+1}, $k \geq 1$, will be derived by search along a descent direction d_k , such that $d_k^T g_k < 0$, where $g_k = \nabla f(x)$,

such that

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

where α_k is the step-length satisfying line search conditions such that

$$f(x_k + \alpha_k d_k) \le f(x_k) + c_1 \alpha\, g_k^T d_k \tag{3}$$

and

$$\left| g_{k+1}^T d_k \right| \le -c_2 g_k^T d_k \tag{4}$$

where $0 < c_1 < c_2 < 1$.

We consider that

$$v_k = x_{k+1} - x_k \tag{5}$$

$$y_k = g_{k+1} - g_k \tag{6}$$

Conjugate Gradient (CG) method is one of the few practical methods for solving large dimension problems because it does not require matrix storage and its iteration cost is very low.

Now we begin from the initial direction d_1 such that

$$d_1 = -g_1. \tag{7}$$

The search direction for the next iteration has the following form:

$$d_{k+1} = -g_{k+1} + \beta_k d_k \tag{8}$$

Where β_k is a constant parameter defined by either

$$\beta_k = \frac{\left\| g_{k+1} \right\|^2}{\left\| g_k \right\|^2} \tag{9a}$$

Or

$$\beta_k = \frac{g_{k+1}^T (g_{k+1} - g_k)}{\left\| g_k \right\|^2} \tag{9b}$$

The definition of β_k in (9a) is due to Fletcher and Reeves [6] and β_k in (9b) is due to Polak Ribiere [12].

2. Generalized CG-Methods:

CG-algorithm usually requires more function evaluations than the variable Metric (VM) method to solve small dimensionality problems. Therefore, many extensions and modifications have been proposed in this field .Liu and Storey [8] introduced a generalized PR algorithm. They studied the effect of the inexact line search on conjugacy in unconstrained optimization, and they showed that their algorithm has global convergence for twice continuously differentiable functions with a bounded level set.

2.1 Generalized CG-algorithm (Liu and Storey, 1991)

Step 1:let x_1 be an initial point of the minimizer x^* of f.
Step 2:set the iteration k=1 and the search direction $d_k = -g_k$

Step 3:do a line search: set $x_{k+1}=x_k+\alpha_k d_k$, set k=k+1, α is a line search parameter.

Step 4:if $\left\|g_{k+1}\right\|<$ ε, where ε=5×10^{-5}, take x* as x_{k+1} and stop;
Otherwise go to step 5.

Step 5: if k>n>2,go to step 9; otherwise go to step 6.

Step 6:let $t_k = d_k^T G_{k+1} d_k$, $s_k = g_{k+1}^T G_{k+1} g_{k+1}$, and $u_k = g_{k+1}^T G_{k+1} d_k$.

Step7: if t_k>0, s_k>0, $1 - u_k^2/(t_k s_k) \geq 1/(4r_k)$, and

$(s_k/g_{k+1}^T g_{k+1})/(t_k/d_k^T d_k) \leq r_k$, r_k>0 then go to step 8;
Otherwise go to step 9.

Step8: let

$$d_{k+1} = [(u_k g_{k+1}^T d_k - t_k g_{k+1}^T g_{k+1})g_{k+1} + (u_k g_{k+1}^T g_{k+1} - s_k g_{k+1}^T d_k)d_k]/w_k,$$

Where w_k= $t_k s_k$- u_k^2, go to step 3.

Step 9:set x_{k+1} =x_1 and go to step 2.

This algorithm was compared with currently available standard routines and their results demonstrate a general efficient GPR algorithm.

Usually CG-algorithms are implemented with restarts, in order to avoid the effects of an accumulation error. Fletcher [5] in his standard method suggested to restart his algorithm with the steepest descent direction every n or n+1 iteration, where n is a dimension of the problem, another restarting direction was suggested by Powell [13]. He developed a new procedure for starting CG-methods. Powell checked that the new search direction d_{k+1} will be sufficiently downhill if their inequalities

$\left|g_{k+1}^T g_k\right| > 0.2\left|g_{k+1}^T g_{k+1}\right|$ is satisfied.

3.Self –Scaling VM methods:

The practical and theoretical merits of the quasi Newton (QN) family of methods for unconstrained optimization have been systematically explored since the classic paper of Fletcher and Powell [7] analyzing Davidon's VM method.

In the family of VM updating, it is useful to multiply each H_k by some scale factor $\rho_k > 0$ before using the update formula. With exact line searches, this can be shown to present the conjugacy property in the quadratic case, although we may no longer have $H_{k+1} \equiv G^{-1}$. However, the focus here is to improve the single-step rather than the n-step convergence behavior of the algorithm. Methods that automatically prescript scale factor in a manner such that, if the function is quadratic then the eigenvalues of

$d_k^T H_k G_{k+1}$ tend to be spread above and below are called self-scaling methods [3].

Thus in 1970's the self-scaling VM algorithms were introduced showing significant improvement in efficiency over standard VM-methods.In particular, in a series of papers by Oren [11], Al-Bayati [1], Nocedel [10] and Al-Bayati and Al-Salih [2] algorithms for minimizing an unconstrained nonlinear function f (x) were developed.

Now we summarize the scaled BFGS algorithms due to Oren [11].

3.1 Oren's algorithm (1974)

Step 1: Start with an initial point x_1, Set k=1 and choose H_1=I, where I is the identity matrix.

Step 2:Determine the step-size α_k that minimizes f $(x_k+\alpha_k d_k)$ where d_k=-$H_k g_k$ and obtain x_{k+1}=$x_k+\alpha_k d_k$, α_k is determined by inexact line search.

Step 3: If $\left\| g_{k+1} \right\| < \varepsilon$, where ε=5×10^{-5}, take x* as x_{k+1} and stop; Otherwise continue.

Step 4:Update H_k by H_{K+1} by using Oren's update as follows

$$H_{k+1}^{Oren} = \left[H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} + w_k w_k^T \right]^T \psi_k + \frac{v_k v_k^T}{v_k^T y_k} \qquad (10)$$

where

$$\psi_k = \frac{v_k^T y_k}{y_k^T H_k y_k} \qquad (11a)$$

and w_k is vector defined by

$$w_k = (y_k^T H_k y_k)^{1/2} \left\{ \frac{v_k}{v_k^T y_k} - \frac{H_k y_k}{y_k^T H_K y_k} \right\} \qquad (11b)$$

Step 5: Set k=k+1 and go to step 1.

4. Preconditioned CG-Methods:

Any problem in an unconstrained optimization can be solved by CG or VM methods, each of these methods has particular advantages and disadvantages. In general, a CG-method normally requires more iterations than a QN or VM method to obtain an equally good minimum point but a CG-method requires less storage for implementation per iteration. CG-methods have proved to be valuable where n is large because at each step a

few n vectors have to be stored and hence the computational costs and storage requirements are affordable, even for large problems.

The new class of CG-methods has been developed, termed preconditioned conjugate gradient methods (PCG); the idea of the preconditioning has been extended directly to nonlinear problem [9].

4.1 New Modified Generalized CG-algorithm:

Step 1: Let x_1 be an initial point of the minimizer x* of f and $H_1=I$ where I is the identity matrix.

Step 2: Set k=1 and $d_k=-H_k g_k$.

Step 3: Do a line search to compute α_k that minimizes f $(x_k+\alpha_k d_k)$ then compute $x_{k+1}=x_k+\alpha_k d_k$.

Step 4: If $\left\|g_{k+1}\right\|$<ε, where ε=5×10^{-5}, take x* as x_{k+1} and stop; otherwise go to step 5.

Step 5: If k=n or $\left|g_{k+1}^T g_k\right| > 0.2\left|g_{k+1}^T g_{k+1}\right|$, go to step 2; otherwise continue.

Step 6: Let $t_k = d_k^T G_{k+1} d_k , s_k = g_{k+1}^T G_{k+1} g_{k+1}$, and $u_k = g_{k+1}^T G_{k+1} d_k$.

Step 7: If t_k>0,s_k>0, $1-u_k^2/(t_k s_k) \geq 1/(4r_k)$, and

$(s_k/g_{k+1}^T g_{k+1})/(t_k/d_k^T d_k) \leq r_k , r_k$>0 then go to step 8; otherwise go to step 10.

Step 8: Let

$$d_{k+1} = [(u_k g_{k+1}^T d_k - t_k g_{k+1}^T H_k g_{k+1})H_k g_{k+1} + (u_k g_{k+1}^T H_k g_{k+1} - s_k g_{k+1}^T d_k)d_k]/w_k$$

where $w_k= t_k s_k - u_k^2$.

Step 9: Update H_k by H_{k+1} using Oren's [11] formula as given in (10).

Step 10: Set $x_{k+1} =x_1$, k=k+1 and go to step 3.

4.2 Derivation of the Modified Generalized CG Direction:

Let H be any symmetric positive-definite preconditioned matrix, then by Nazareth [9] with Choleski Factorization H can be factorized as:

H=LL^T,   (12)

where L is a real lower triangular matrix and non-singular matrix. Let f be the strictly quadratic function

f(x)=x^{T}Gx/2+b^{T}x+c,   (13)

then the gradient is

g(x)= f $'$(x) =Gx+b^T.   (14)

Let x=Lz,   (15)

where z defines a new vector spaces and is defined as:

h(z)=f(Lz)=(Lz)^{T}G(Lz)/2+(Lz)^{T}b^T+c,   (16)

which implies that

$$h'(z) = f'(Lz) = L^T GLz + L^T b^T$$
$$= L^T(GLz + b^T)$$
$$= L^T g(x). \tag{17}$$

Eq.(17) gives a relationship between the gradients in x-space and z-space, i.e.

$$\bar{g} = L^T g, \tag{18}$$

where $\bar{g}$ is the gradient in z-space while g is the gradient in x-space, so that if h (z*)=0, then g (Lz*)=0.

Set $z_{k+1} = z_k + \alpha_k \bar{d}_k$. \hfill (19)

Multiplication of Eq. (19) by L we get,

$$Lz_{k+1} = Lz_k + \alpha_k L\bar{d}_k. \tag{20}$$

Using Eq. (15), then Eq. (20) becomes

$$x_{k+1} = x_k + \alpha_k d_k, \tag{21}$$

therefore,

$$d_k = L\bar{d}_k, \tag{22}$$

which implies that

$$\bar{d}_k = L^{-1} d_k. \tag{23}$$

Set

$$\bar{y}_k = \bar{g}_{k+1} - \bar{g}_k \tag{24}$$

where $\bar{g}_k$, $\bar{g}_{k+1}$ are the gradients of h (z) at the point z_k, z_{k+1} ,respectively.

By using Eq. (18), then Eq. (24) becomes

$$\bar{y}_k = L^T g_{k+1} - L^T g_k = L^T y_k. \tag{25}$$

Now consider applying the conjugate gradient method,

$$\bar{d}_{k+1} = [(u_k \bar{g}_{k+1}^T \bar{d}_k - t_k \bar{g}_{k+1}^T \bar{g}_{k+1})\bar{g}_{k+1} + (u_k \bar{g}_{k+1}^T \bar{g}_{k+1} - s_k \bar{g}_{k+1}^T \bar{d}_k)\bar{d}_k]/w_k$$

By using Eqs. (15), (18), (23) and (25) we get

$$L^{-1} d_{k+1} = [(u_k g_{k+1}^T LL^{-1} d_k - t_k g_{k+1}^T LL^T g_{k+1})L^T g_{k+1} + (u_k g_{k+1}^T LL^T g_{k+1} - s_k g_{k+1}^T LL^{-1} d_k)L^{-1} d_k]/w_k, \tag{26}$$

multiply (26) by L and using (12) and $LL^{-1}=I$ where I is the identity matrix we get:

$$d_{k+1} = [(u_k g_{k+1}^T d_k - t_k g_{k+1}^T Hg_{k+1})Hg_{k+1} + (u_k g_{k+1}^T Hg_{k+1} - \\ s_k g_{k+1}^T d_k)d_k]/w_k \qquad (27)$$

Equation (27) is our proposed CG Method with Metric update H, where H is any positive-definite symmetric matrix, or the preconditioned CG Method (PCG). Thus the search direction defined in Eq. (27) is a new direction for the Generalized Polak-Ribiere CG-direction.

5. Numerical Results:

In this section we are going to employ the new proposed algorithm with a certain effective CG-method; namely Liu and Storey [8].

We will show that our algorithm is better than the Generalized CG-of Liu and Storey's [8] algorithm. Using Oren [11] self-scaling VM update Oren [11] as an acceleration tool to the Generalized CG-algorithm to decrease the number of iterations (NOI) and number of function evaluations (NOF). The comparison tests involve ten well-known test functions with different dimensions. All the results were obtained using double precision on the (Pentium II Computer) using programs written in Fortran Language. The terminate criteria is (5×10^{-5}) with Powell's restarting criterion $\left| g_{k+1}^T g_k \right| > 0.2 \left\| g_{k+1} \right\|^2$ and the algorithms use the cubic line search strategy, with function and gradient values and it is an adaptation from that published by Bunday [4].

The comparative performances of the algorithms are evaluated by considering both the total numbers of iterations (NOI) and total number function evaluations (NOF).

Thus all our numerical results are presented in table (1). In this table we have compared our modified proposed algorithm with Liu and Storey [8] algorithm by using (Ten) cases and for dimension $2 \leq n \leq 1000$.

It is clear that the modified algorithm outperformance the standard Liu and Sorey's algorithm as results of this comparison.

17

Table (1)
Comparative performance of all algorithms for test functions with
$(2 \leq n \leq 1000)$.

Test Functions	N	Liu and Storey's algorithm		Modified GCG algorithm	
		NOI	NOF	NOI	NOF
Sallow	2	9	24	9	24
Cantrel	4	17	145	16	77
Rosen	10	305	781	16	44
Cubic	70	167	425	37	99
Powell	100	239	1145	156	785
Recip	120	9	24	5	17
Sum	180	18	85	11	48
Powell	200	201	1727	175	519
Recip	600	7	20	7	18
Sum	1000	20	79	13	57
Total		992	4455	445	1688

We see from the above table that taking the cubic routine as 100% NOI and NOF yields:

Table (2)

Tools	Liu and Storey's Algorithm	Modified GCG algorithm
NOI	100	44.8
NOF	100	37.8

It is clear from the above table that the modified algorithm is more efficient than Generalized Liu and Storey's CG-algorithm.

6. Conclusions:

In the field of unconstrained optimization, a new search direction is implemented for modified extended generalized CG-VM steps. The technique uses inexact line search with promising numerical results.

REFERENCES

[1]   Al-Bayati, A.Y., (1991), "A New Family of Self-Scaling Variable Metric Algorithms for Unconstrained Optimization", **Journal of Education and Science**, Mosul University, Vol. (12), pp.25-54.

[2]   Al-Bayati, A.Y.and Al-Salih, M.S. (1994), "New VM-Methods for Nonlinear Unconstrained Optimization", **Journal of Education and Science, Mosul University**, Vol. (21), pp.169-182.

[3]   Bazaraa, M.S. (1993), **Nonlinear Programming**, London : English, Universities Press.

[4]   Bunday, B. (1984), "**Basic Optimization Methods**", London: Edward Arnold.

[5]   Fletcher, R. (1987), "**Practical Methods by Optimization**", John Wiely and Sons, Chichester, New York, Brisbance, Tornoto and Singapore.

[6]   Fletcher, R.and Reeves, C.M. (1964), "Function Minimization by Conjugate Gradients", **Computer Journal**, 7,pp.149-154.

[7]   Fletcher, R.and Powell, M.J.D. (1963), "A Rapidly Convergent Descent Method for Minimization", **Computer Journal**, 6, pp.163-168.

[8]   Liu, Y. and Storey, C. (1991). "Efficient Generalized Conjugate Gradient Algorithms", Part 1,Thorey, **Journal of Optimization Theory and Applications**, 69,pp.129-137.

[9]   Nazareth, L. (1979) "A Relationship between BFGS and Conjugate Gradient Algorithms and its Implementations for New Algorithms", **SIAM Journal Numerical Analysis**, 16, pp. 794-800.

[10]   Nocedal, J. (1993) "Analysis of a Self-Scaling Quasi-Newton Method", **Mathematical Programming**, 611.pp.19-37.

[11]   Oren, S.S. (1974), "Self-Scaling Variable Metric Algorithm, Part II", **Management Science**, 20, pp. 863-874.

[12]   Polak, E. (1969), "Computational Methods in Optimization a Unified Approach", **Academics press New York**.

[13]   Powell, M.J.D. (1977), "Restart Procedures for the Conjugate Gradient Method", **Mathematical Programming**, 12,pp.241-254.