A PCG ALGORITHM FOR NON-LINEAR OPTIMIZATION BASED ON A QUADRATIC MODEL

BY

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

Professor | **Postgraduate**

Department of Mathematics,
College of Computers and Mathematical Sciences,
University of Mosul, Mosul, Iraq.

Received 1/7/ 2001 | **Accepted** 1/11/ 2000

الملخص :-

في هذا البحث تم استحداث خوارزمية تداخلية جديدة تعمل على تعجيل خوارزمية التدرج المترافق الموسع بواسطة المتري المتغير ذاتي القياس في الامثلية اللاخطية. هذه الخوارزمية الجديدة تستعمل خطوط بحث غير تامة والحسابات العددية بينت كفاءة الخوارزمية الجديدة مقارنةً بشبيهاتها من الخوارزميات في هذا المجال.

Abstract :-

In this paper a new preconditioned Extended Conjugate Gradient with self-scaling Variable Metric method is proposed for unconstrained optimization. This method is based on the inexact line searches and it is examined by using different nonlinear test functions with various dimensions.

1. Introduction

conjugate Gradient (CG) methods are some of the most useful algorithms to solve large problems for unconstrained optimization problem:

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

when f is twice continuously differentiable function and its domain contains a bounded level set L

$$L=\{x\in R: f(x)\le f(x_1)\}; \tag{1.2}$$

where x_1 is an initial point. Then the minimum point of f is to be found at each subsequent point x_{k+1} ; k>1 and derived by searching along a decent direction d_k such that $d_k^T g_k < 0$; $g_k=\nabla f(x_k)$ so that

$$x_{k+1}=x_k+\lambda_k d_k \quad , k\ge 1 \tag{1.3}$$

$$\text{where } \lambda_k=\arg\min_{\lambda} f(x_k+\lambda_k d_k) \tag{1.4a}$$

and λ_k satisfies line search conditions such that

$$|g_{k+1}^T d_k| < c_1 |g_k^T d_k|; \qquad 0 < c_1 < 1/2. \tag{1.4b}$$

and

$$f(x_{k+1}) \le f(x_k) + c_2 \lambda_k g_k d_k; \qquad 0< c_2 <1/2. \tag{1.4c}$$

We consider that

$$v_k=x_{k+1}-x_k. \tag{1.5}$$

$$y_k=g_{k+1}-g_k. \tag{1.6}$$

We know that (CG) method is one of the few practical methods for solving large dimensionality problems because it does not require matrix storage and its iteration cost is very low. Usually the initial direction is given by

$$d_1=-g_1. \tag{1.7}$$

Then the search direction for the next iteration has the form

$$d_{k+1}=-g_{k+1}+\beta_k d_k, \tag{1.8}$$

Where β_k is a constant parameter determined by the conjugacy condition

$$d_{k+1}^T G_{k+1} d_k = 0; k \ge 1, \tag{1.9}$$

Any CG-method is a conjugate direction (CD) method where the vectors d_k is also satisfy

$d_k \in [d_0, d_1, ..., d_{k-1}, g_k]$ for k=0,1,2,…,n-1; (Al-Bayati, 1996).
that has the properties:

(a) $d_1,...,d_k$ are conjugate with respect to f, (1.10a)

(b) $g_k^T d_i = 0$ i=1,2,..., k, (1.l0b)

(c) $g_k^T g_i = 0$ i=0,1,..., k-1, (1.l0c)

Thus Dixon, (1975) seeks a direction d_{k+1} that is conjugate to the member of the set d_k and is a linear combination of k+l directions described in (1.8).

Then is β_k defined either by

$$\beta_k = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k}, k \geq 1 \text{ or} \quad (1.11\ a)$$

$$\beta_k = \frac{g_{k+1}^T (g_{k+1} - g_k)}{g_k^T g_k}; k \geq 1 \quad (1.11\ b)$$

Then definition of β_k in (1.11 a) is due to (Fletcher-Reevrs, 1964)(FR) and P_k in (1.ll b)is due to (Polka-Ribiere 1969)(PR). Extensive numerical experience has shown that the PR algorithm is more efficient than the original FR algorithm.

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 (1991) introduced a generalized PR algorithm. They studied the effect of the inexact line search on conjugacy in unconstrained optimization, and they show that their algorithm has global convergence for twice continuously differentiable functions with a bounded level set.

Before we give the algorithm by Liu and Storey (1991), we first established the general convergence theorem.

2.1 Theorem

Let the line search direction be

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

where t_k, u_k, s_k, w_k satisfy

$t_k > 0$, $s_k > 0$,

$1-u_k^2/(t_k s_k) \geq 1/(4r_k)$, $0<r_k<\infty$,

$w_k = t_k s_k - u_k^2$

$(s_k/g_{k+1}^T g_{k+1})/(t_k/d_k^T d_k) \leq r_k$, $0<r_k<\infty$,

or $t_k=1$, $u_k=s_k=0$, $w_k = g_{k+1}^T g_{k+1} \neq 0, r_k = 0$.

$g_k^T d_k < 0$,

Then, the following descent property holds:

if in addition there exists $p_k > 0$ such that

$$f(x_{k+1}) - f(x_k) \leq -p_k (g_k^T d_k)^2 / d_k^T d_k.$$

Then if

$$\sum_{k=1}^{\infty} p_k/(1+r_k^2) = \infty,$$

it follows that

$$\lim_{k\to\infty} \inf g_{k+1}^T g_{k+1} = 0;$$

and if

$$\lim_{k\to\infty} \inf p_k/(1+r_k^2) > 0.$$

Then we have the convergence property

$$\lim_{k\to\infty} g_{k+1}^T g_{k+1} = 0.$$

Proofs: see (Liu and Storey, 1991)

2.2 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 k=1 and $d_k=-g_k$

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

Step 4:if $\|g_{k+1}\| < \varepsilon$, where $\varepsilon=5\times10^{-5}$, take x^* as x_{k+1} and stop; Otherwise go to step 5.

Step 5:if k+1>n>2,go to step 9; otherwise go to step 6.

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

In practice, 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 (1987) in his standard method suggested to restart his algorithm with the steepest descent direction every n or n+l iteration, where n is the dimension of the problem, Another restarting direction was suggested by (Powell, 1977). They developed a new

procedure for starting CG-methods. Powell checked that the new search direction d_{k+1} will be sufficiently downhill if there inequalities

$$d_{k+1}^T g_{k+1} \le -0.8\|g_{k+1}\|^2$$

is satisfies.

3. Quasi-Newton Methods

Let H_{k+1} be an approximation to G_{k+1}^{-1}. Satisfying the Quasi-Newton condition

$$H_{k+1} y_k = v_k \tag{3.1}$$

a family of H_{k+1} satisfying (3. I) is

$$H_{k+1} = H_k - \frac{H_k y_k y_k^T H_K}{y_k^T H_K y_k} + \frac{v_k v_k^T}{v_k^T y_k} + \varphi(y_k^T H_K y_k) w_k w_k^T \tag{3.2}$$

where

$$w_k = \frac{v_k}{v_k^T y_k} - \frac{H_k y_k}{y_k^T H_K y_k} \tag{3.3}$$

and φ_k is a free parameter. Quasi-Newton methods are quite efficient but need to store H_k and require O (n^2) multiplication per iteration to update H_k (Fletcher and Powell, 1963).

It is generally agreed that the best member of Broyden family is the BFGS algorithm obtained by taking φ_k (Hu, Y.F. Khoda, K.M. Liu, Y. Storey, C and Touati-Ahmed, D. (1995)).

3.1 Self-Scaling VM Methods

To alleviate the family of VM updating, it 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 $H_{k+1}G^{-1}$. 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 mamaner such that, if the function is quadratic then the eigenvalues of d_k H_kG_{k+1} tend to be spread above and below are called self-scaling methods (Bazaraa, 1993).

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, 1979), (Al-Bayati, 1991), (Nocedal, 1993) and (Al-Bayati and Al-Salih, 1994). Algorithms for minimizing an unconstrained nonlinear function f (x) were developed.

Now we summarize the scaled BFGS algorithm due to (Al-Bayati, 1991).

3.2 Self-Scalin algorithm (Al-Bayati, 1991)

Start with an initial point x_1,

Step 1:Set k=1 and choose $H_1=I$, where I is the identity matrix.

Step 2:determine the step-size λ_k that minimizes $f(x_k+\lambda d_k)$ where

$d_k=-H_kg_k$ and obtain $x_{k+1}=x_k+\lambda_k d_k$.

Step 3:Update H_k by H_{k+1} by using Al-Bayati's update as follows

$$H_{k+1}^{Al-Bayati} = H_k - \left\{I - \frac{v_k y_k^T}{v_k^T y_k}\right\} H_k \left\{I - \frac{y_k v_k^T}{v_k^T y_k}\right\} + \varphi_k \frac{v_k v_k^T}{v_k^T y_k} + w_k w_k^T \tag{3.2.1}$$

where $\varphi_k = \dfrac{y_k^T H_k y_k}{v_k^T y_k}$

and w_k is vector defined by

$$w_k = (y_k^T H_k y_k)^{1/2} \left\{ \frac{v_k}{v_k^T y_k} - \frac{H_k y_k}{y_k^T H_K y_k} \right\} \qquad (3.2.2)$$

Step 4: Set k=k+1 and go to step 1.

4. Preconditioned CG-Methods

In applications of the general functions, the CG and QN methods each have 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 when 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.

Therefore new class of CG-methods has been developed, termed preconditioned conjugate gradient methods (PCG); the idea of the preconditioning is to transform the problem so that the Hessian of the transformed problem has clustered eigenvalues and is well conditioned.

The aim of PCG Method is to keep the storage requirements of order n while improving the convergence properties, (see Nazareth 1979).

It will be seen that Quasi-Newton inverse Hessian approximation has desirable properties as preconditions.

The idea of preconditioning has been extended directly to nonlinear problems. The standard CG method is not always effective, but preconditioning using an appropriate matrix can accelerate convergence of the CG method by a transformation of variables while keeping the basic properties of the method.

The preconditioning matrices described in this paper are based upon the inverse Hessian approximations generated in a Quasi-Newton method. Therefore, the main idea behind this method is the use of the Quasi-Newton or Variable Metric updates to accelerate the CG methods.

4.1 New Preconditioned 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 set $x_{k+1}=x_k+\lambda_k d_k$.

Step 4: if $\|g_{k+1}\|<\varepsilon$, where $\varepsilon=5\times10^{-5}$, take x^* as x_{k+1} and stop; otherwise go to step 5.

Step 5: if $k+1>n>2$, go to step 9; otherwise go to step 6 .

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

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

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

Step8: 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 Al-Bayati (1991) formula as given in (3.2).

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

4.2 The Derivation of the New PCG Search Direction

Let H be any symmetric positive-definite preconditioned matrix, then (By Nazareth, 1979) with Choleski Factorization H can be factorized as:

$$H=LL^T \tag{5.1}$$

where L is a real lower triangular matrix and non-singular matrix.

Let f be the strictly quadratic function

$$f(x)=x^TGx/2+b^Tx+c \tag{5.2}$$

then the gradient is

$$g(x)= f'(x) =Gx+b \tag{5.3}$$

$$\text{Let } x=Lz \tag{5.4}$$

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

$$h(z)=f(Lz)=(Lz)^TG(Lz)/2+(Lz)^Tb+c \tag{5.5}$$

which implies that

$$\begin{aligned} h'(z) &= f'(Lz)=L^TGLz+L^Tb \\ &= L^T(GLz+b)= L^Tg(x). \end{aligned} \tag{5.6}$$

Eq. (5.6) gives a relationship between the gradients in x-space and z-pace, i.e

$$\bar{g}=L^Tg \tag{5.7}$$

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 $z_{k+1}=z_k+\alpha_k \bar{d}_k$ (5.8)

multiplication of Eq. (5.8) by L we get

$Lz_{k+1}=Lz_k+\alpha_k L\bar{d}_k$. (5.9)

using Eq. (5.4), then Eq. (5.9) becomes

$x_{k+1}=x_k+\alpha_k d_k$ (5.10)

herefore,

$d_k=L\bar{d}_k$ (5.11)

which implies that

$\bar{d}_k=L^{-1}d_k$. (5.12)

Set $\bar{y}_k=\bar{g}_{k+1}-\bar{g}_k$ (5.13)

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

by using Eq. (5.7), Eq. (5.13) becomes

$\bar{y}_k L^T g_{k+1}-L^T g_k=L^T y_k$ (5.14)

now consider applying the conjugate gradient method,

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

By using Eqs. (5.4), (5.7), (5.12) and (5.14) we get

$$L^{-1}d_{k+1}=[(u_k g^T_{k+1}LL^{-1}d_k-t_k g^T_{k+1}LL^T g_{k+1})L^T g_{k+1}+(u_k g^T_{k+1}LL^T g_{k+1}-s_k g^T_{k+1}LL^{-1}d_k)d_k]/w_k ,\quad (5.15)$$

multiply (2.23) by L and using (2.9) and $LL^{-1}=I$ where I is the identity matrix we get:

$$d_{k+1}=[(u_k g^T_{k+1}d_k-t_k g^T_{k+1}Hg_{k+1})Hg_{k+1}+(u_k g^T_{k+1}Hg_{k+1}-s_k \dot{g}^T_{k+1}d_k)d_k]/w_k \quad (5.16)$$

equation (5.16) is called 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. (5.16)) is a new preconditioned Hestens and Stiefel PCG-method.

The relative advantages over current algorithms of this type are they require less storage and computation time, they are not so sensitive to the exactness of the line searches, and the extension to minimize the general function in a finite number of steps, i.e. this PCG algorithm, has a quadratic termination property.

5. Numerical Results

In this section we will show that our algorithm is better than GCGStore's algorithm. Using self-scaling VM update (Al-bayati, 1991) as an acceleration tool to the Generalized CG-Storey's algorithm to decrease the number of iterations (NOI) and number of function evaluations (NOF). The comparison tests involve twenty-five well-known test functions with different dimension (see Appendix). All the results were obtained using double precision on the (Pentium II Computer). Using programs written in Fortran. The algorithms terminate whenever $\|g_{k+1}\|<5\times10^{-5}$ with Powell's restart $|g_{k+1}^T g_k| > 0.2\|g_{k+1}\|$ and the algorithms are using the cubic line search strategy, which uses function and gradient values and it is adaptation that published by Bunday (Bunday, 1984).

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

Indeed, all our numerical results, the GCG-algorithm and new algorithm are presented in table (1).

In this table we have compared our new proposed algorithm with General CG algorithm by using (25) cases and for dimension ($2\leq n \leq 1000$). It is clear that the new algorithm outperform the standard Storey's algorithm as a result of this comparison.

9.Generalized Penalty (1) Function:

$$f(x)=\sum_{i=1}^{n}\left[(x_i-1)^2+\exp(x_i^2-0.25)^2\right], \qquad x_0=(1,2,\ldots,n)^T.$$

10.Generalized Penalty (2) Function:

$$f(x)=\sum_{i=1}^{n}\left[\exp(x_i-1)^2+(x_i^2-0.25)^2\right], \qquad x_0=(1,2,\ldots,n)^T.$$

11.Generalized Powell Function:

$$f(x)=\sum_{i=1}^{n/4}\begin{bmatrix}(x_{4i-3}+10x_{4i-2})^2+5(x_{4i-1}-x_{4i})^2+\\(x_{4i-2}-2x_{4i-1})^4+10(x_{4i-3}-x_{4i})^4\end{bmatrix}, \qquad x_0=(3,-1,0,3;\ldots)^T.$$

12.Generalized Powell 3 Function:

$$f(x)=\sum_{i=1}^{n/3}\left\{3-\left[\frac{1}{1+(x_i-x_{2i})^2}\right]-\sin\left(\frac{\pi x_{2i}x_{3i}}{2}\right)-\exp\left[-\left(\frac{x_i+x_{3i}}{x_{2i}}-2\right)^2\right]\right\},$$

$x_0=(0,1,2;\ldots)^T$.

7. Reference:

Al-Bayati, A.Y., (1996), "A New PCG Method for Unconstrained Non-Linear Optimization", Journal of Abhath Al-Yarmouk, Jordan, Vol. S,No.I, pp. 71-92.

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.

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.

Bazaraa, M.S. (1993), "Nonlinear Programming", English, Universities Press, London.

Bunday, B. (1984), "Basic Optimization Methods", Edward Arnold, London.

Dixon, L. C. W. (1975), "Conjugate Gradient Algorithms: Quadratic Termination without Line Searches", journal of the Institute of Mathematics and its Applications Vol. 15,pp.9-18.

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

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

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

Hu, Y.F., Khoda, K.M., Liu, Y. Storey, C.and Touati-Ahmed, D. (1995), "Some Research on Conjugate Gradient Methods for Unconstrained Optimization" J.Fac.sci.U.A..E.Uni.Vol (7), No. 10, (1995), pp.123-131.

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

Nazareth, L. and Nocedal, J. (1979), "Conjugate Direction Methods with Variable Storage"Jornal of Mathematical Programming 23, pp. 326-340.

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

Oren, SS. (1979), "Self- Scaling Variable Metric Algorithm without Line Search for Unconstrained Optimization", Mathematics of Commutation 27.

Polak, E. and Ribiere (1969), "Computational Methods in Optimization a Unified Approach", Academic press, New York.

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