On Self-Scaling Variable-Metric Algorithms

Abbas Y.AL-Bayati **Mardin Sh. Taher**
profabbasalbayati@yahoo.com
College of Computer sciences and Mathematics
University of Mosul/Iraq

Received on: 25/01/2005 **Accepted on: 09/05/2005**

ABSTRACT

In this paper, we have developed a new self-scaling VM-method for solving unconstrained nonlinear optimization problems. The numerical and theoretical results demonstrate the general effectiveness of the new self-scaling VM-method when compared with PHUA & ZENG algorithm ; we have tested these algorithms on several high-dimension test functions with promising numerical results.

Keywords: unconstrained optimization, nonlinear problems, self-scaling VM-method.

حول خوارزميات المتري المتغير ذاتي القياس

عباس يونس البياتي **ماردين شوكت طاهر**
كلية علوم الحاسوب والرياضيات/جامعة الموصل/العراق

تاريخ استلام البحث: 2005/01/25 **تاريخ قبول البحث: 2005/05/09**

الملخص

في هذا البحث تم استحداث خوارزمية جديدة للمتري المتغير ذاتي القياس لحل المسائل اللاخطية في ألا مثلية غير المقيدة. النتائج النظرية والعملية أثبتت كفاءة الخوارزمية الجديدة بمقارنتها مع خوارزمية PHUA & ZENG وباستخدام عدد من الدوال غير الخطية ذات الإبعاد الكبيرة .

الكلمات المفتاحية: الامثلية غير المقيدة، المسائل اللاخطية، المتري المتغير ذاتي القياس.

1-Introduction

In this section we begin with a numerical method for solving unconstrained nonlinear minimization problem

$$\min f(x) \qquad (1)$$

where the function f(x) is a twice continuously differentiable real valued function in n-dimensional space.

A popular class of method for solving (1) is the Quasi-Newton (QN) or Variable Metric (VM) methods. Given an initial point x_1 ,an nxn matrix H_1,(H_1 is non-singular), VM-methods proceed to generate the following iterative sequence at the k^{th} iteration: -

1- Compute $g_k=\nabla f(x_k)$,the gradient of the function f(x) at the current point x_k.

2- Compute the search direction

$$d_k = -H_k g_k \qquad (2)$$

3- Apply an appropriate line search strategy along the search direction d_k to find a step-size $\alpha_k > 0$, such that the following Wolf's condition are satisfied:

$$f(x_k + \alpha_k d_k) \le f(x_k) + \beta_1 \alpha_k g^T_k d_k \qquad (3)$$

$$g(x_k + \alpha_k d_k)^T d_k \ge \beta_2 g^T_k d_k \qquad (4)$$

where $0 < \beta_1 < 0.5$ and $\beta_1 < \beta_2 < 1$.

4- Set $x_{k+1} = x_k + \alpha_k d_k$. (5)

5- Update the H_k by

$$H_{k+1} = H_k + \Delta H_k . \qquad (6)$$

where ΔH_k is a correction matrix .Usually ,$H_1=I$ is chosen ,and the updating matrix H_{k+1} is chosen to satisfy the following QN- like condition

$$H_{k+1} y_k = \lambda_k v_k \qquad (7)$$

where $v_k = x_{k+1} - x_k$,

$y_k = g_{k+1} - g_k$,and λ_k is scalar

A general class of QN updates was proposed by Broyden (1967).

$$H_{k+1}=H_k-(H_k y_k y^T_k H_k)/(y^T_k H_k y_k)+v_k v^T_k/ v^T_k y_k+\phi_k(y^T_k H_k y_k)R_k R^T_k \qquad (8)$$

where

$$R_k=v_k/v_k^T y_k-H_k y_k/ y^T_k H_k y_k \qquad (9)$$

$$\delta_k= y^T_k H_k y_k /v_k^T y_k \qquad (10)$$

$$\mu_k=v^T_k H_k v_k/ v_k^T y_k \qquad (11)$$

$$\phi_k = \phi_k (\theta_k)=(1- \theta_k)/(1+ \theta_k (\delta_k \ \mu_k)) \qquad (12)$$

where $\theta_k \in R$ is parameter;

The most popular QN-methods are defined from (8) by choosing different values of θ_k,

For $\theta_k = v^T_k y_k/ (v^T_k y_k - v^T_k H_k v_k)$, we get the symmetric rank-one formula, Broyden (1967).

For $\theta_k = 1$, we get the DFP formula, due to Davidon, Fletcher and Powll(1963).

For $\theta_k = 0$, we get the BFGS formula ,due to Broyden, Fletcher,Goldfarb and Shanno (1970).

2-The modification of QN-methods

These methods can be expressed as follows: -

$$H_{k+1}=(H_k-(H_k y_k y^T_k H_k)/(y^T_k H_k y_k)+\phi_k(y^T_k H_k y_k) R_k R^T_k) st_1+v_k v^T_k/t_k v^T_k y_k \qquad (13)$$

such that chosen H_{k+1} to satisfy the following modified QN- Condition

$$H_{k+1} y_k= \lambda_k v_k \qquad (14)$$

where $\lambda_k > 0$ is scaling parameter ;

$$\mathbf{t_k=1/\lambda_k=(6(f(x_k)-f(x_{k+1})+v^T{}_k g_{k+1})/v^T{}_k y_k)-2} \quad \textbf{(15)}$$

$$\boldsymbol{\gamma_k=t_k g^T{}_k v_k/g^T{}_k H_k y_k+(1-t_k)v^T{}_k y_k/\ y^T{}_k H_k\ y_k} \quad \textbf{(16)}$$

and ϕ_k is defined by (12)

now we have special families for (13) ,like

1- H_{k+1} (ϕ_k,1,1) = the Broyden family of QN-update(1970).

2- H_{k+1} (ϕ_k, 1, t_k)= the Biggs family of modified BFGS update(1973).

3- H_{k+1} (ϕ_k, st_1,1)= the Oren family of self –scaling QN-update(1974).

4-H_{k+1}(ϕ_k,1,σ_k)=AL-Bayati's(1991)self-scaling method with $\sigma_k=y^T{}_k H_k y_k / v_k{}^T y_k$

5- H_{k+1} (ϕ_k, st_1, t_k)= the PHUA&ZENG family of self –scaling QN-update.

where the parameter (st_1) can be chosen as:

$$st_1=\begin{cases} \varepsilon_1 & \text{if} \quad \gamma_k \le \varepsilon_1 \\ \gamma_k & \text{if} \quad \varepsilon_1 \le \gamma_k \le \varepsilon_2 \\ \varepsilon_2 & \text{if} \quad \gamma_k > \varepsilon_2 \end{cases} \quad \textbf{(17)}$$

where ε_1 is a small positive constant , ε_2 is bigger constant and γ_k is defined by(16)

3-New self-scaling VM-algorithm

However, we can modify the above algorithm by using our new self-scaling QN-update which is given by

$$\mathbf{H_{k+1}=H_k-(H_k y_k y^T{}_k H_k)/(\ y^T{}_k H_k\ y_k)+\ \phi_k(y^T{}_k H_k\ y_k)R_k R^T{}_k\ )\ cs_1\ +v_k v^T{}_k/\ t_k\ v^T{}_k y_k} \quad \textbf{(18)}$$

where the parameter (cs_1) can be chosen as:

$$cs_1=\begin{cases} st_1 \quad \text{if} \quad st_1 \ge st_2\ \&\ st_2 > t_k \\ st_2 \qquad\qquad \text{otherwise} \end{cases} \quad (19a)$$

where $st_2 = v_k{}^T y_k / y^T{}_k H_k y_k$ (19b)

where st_1 is defined by (17),t_k is defined by(15)

we can see that the new algorithm satisfies the quasi-Newton like condition $H_{k+1}y_k$= (cs_1)v_k ,and it generates conjugate search directions.

Theorem (3.1): -

Assume that f(x) be a quadratic function defined by $f(x)=1/2x^TGx+b^Tx$, where G is symmetric positive definite matrix ,b is constant . If the new $H_{k+1}$is any symmetric positive definite matrix then we can define a new updating formula (18) and (19), and the search direction

$d_{k+1}^{new}= -H_{k+1}g_{k+1}$ is identical to the CG-direction defined by

$$d_{k+1}^{CG}= \begin{cases} -g_k & \text{for } k=1 \\ -g_{k+1} + \dfrac{y_k^T g_{k+1}}{y_k^T d_k} d_k & \text{for } k>1 \end{cases} \qquad (20)$$

proof:

The updating formula (18)-(19) can be written as

$$\mathbf{H_{k+1}^{new}=(H_k-v_ky^T_kH_k/v_k^Ty_k- H_ky_kv_k^T/v_k^Ty_k)cs_1+(1+y^T_kH_ky_k/v_k^Ty_k)v_kv^T_k/t_kv^T_ky_k}$$ **Now**

Let $d_{k+1}=-H_{k+1}g_{k+1}$

$$\mathbf{d_{k+1}=-H_kg_{k+1}(cs_1)+y^T_kH_kg_{k+1}(cs_1)v_k/v_k^Ty_k+v_k^Tg_{k+1}H_ky_k(cs_1)/v_k^Ty_k - (y^T_kH_k y_k)\, v_k^Tg_{k+1}v_k/t_k\, v_k^Ty_k)- v_k^Tg_{k+1}v_k/\, t_kv^T_ky_k}$$

since $v_k^Tg_{k+1}=0$ for exact line search, then

$$\mathbf{d_{k+1}=-(cs_1)H_kg_{k+1}+y^T_kH_kg_{k+1}(cs_1)d_k/\, v^T_ky_k} \qquad \mathbf{(21)}$$

for constant (cs1), it is clear that equation (21)is identical search direction for cs1 =1 ,or a parallel search directions with a multiple of (cs1) i.e.$H_{k+1}y_k=(cs_1)\, v_k$ hence the proof .#

It is like standard CG-method in its global convergence property. For more details see (WANG and Li, 2005).

4-Numerical results

Our computation results involved a number of well-known nonlinear functions with different dimensions. The results are performed in double precision using programs written in Fortran.

All the algorithms use exactly the same line search strategy which is the cubic fitting technique, directly adapted from Bundy (1984), these algorithms are assumed to have convergence when each element of the gradient vector is less than 1.E-5 ,that is $\|gk+1\|<1.E-5$.

Our numerical results are presented in two tables (1) and (2), the comparative performance of all the algorithm is evaluated by considering the total number of function evaluations (NOF) and the total number of iterations (NOI).

Table (1) contains the results for problems of dimensionality between 4and 20 i.e. ($4\leq n\leq 20$) and table (2) contains the results for problems of dimensionality between 100and 500 i.e. ($100\leq n\leq 500$).

In table (1) we can see that the new modified algorithm (new) improves PHUA &ZENG algorithm about 45% NOI and 25% NOF, and in table (2) we can see that the new modified algorithm (new) improves PHUA&ZENG algorithm about53% NOI and 31% NOF .As a general conclusion we can say that a proper self-scaling technique is very effective for large scale problems if it is well-defined theoretically, experimentally and for this group of selected functions .

Table (1.a)

Comparison between new and PHUA and ZENG method for 4≤n≤20

Test Function	N	**PHUA&ZENG**		**New**	
		NOI	**NOF**	**NOI**	**NOF**
Cubic	4	20	58	14	55
Powll	4	22	71	27	92
Miele	4	25	74	25	91
Wood	4	38	106	20	71
Wolfe	4	7	16	9	24
Rosen	4	33	92	17	72
Sum	4	4	24	5	28
Non-Diagonal	4	26	66	19	64
Cubic	20	34	87	14	55
Powll	20	37	98	37	123
Miele	20	35	98	26	93
Wood	20	89	227	20	71
Wolfe	20	24	50	26	75
Rosen	20	96	255	18	78
Non-Diagonal	20	52	117	21	86
Total		542	1439	298	1078

Table (1.b)
Performance percentage of improving

Tools	PHUA&ZENG	New
NOI	**100%**	**55%**
NOF	**100%**	**75%**

Table (2.a)
Comparison between new and PHUA&ZENG method for $100 \le n \le 500$

Test function	N	PHUA&ZENG		New	
		NOI	**NOF**	**NOI**	**NOF**
Cubic	100	43	106	14	55
Powll	100	52	131	42	138
Miele	100	35	101	26	93
Wood	100	72	145	42	127
Wolfe	100	92	212	21	87
Non-Diagonal	500	48	115	14	55
Powll	500	47	116	42	131
Miele	500	39	106	27	96
Wolfe	500	82	165	45	136
Non-Diagonal	500	111	276	22	92
Total		**621**	**1473**	**295**	**1012**

Table (2.b)
Performance percentage of improving

Tools	PHUA&ZENG	New
NOI	**100%**	**47%**
NOF	**100%**	**69%**

5-Appendix

All the test functions are from general literature.

1-Cubic function: -

$f=100(x_2 - x_1^3)^2+(1-x_1)^2$ $\qquad$ $x_0=(-1.2,1)^T$

2-Generalized Powll function: -

$$f=\sum_{i=1}^{n/4} [(x_{4i-9}-10x4_{4i-2})^2 +5(x_{4i-1}-x_{4i})^2+(x_{4i-2}-2x_{4i-1})^2+10(x_{4i-9}-10x_{4i})^2] ,\quad x_0=(3,-1,0,1,\dots)^T$$

3-Generalized Miele function: -

$$f=\sum_{i=1}^{n/4} \exp(x_{4i-3}-x_{4i-1})^2 +100(x_{4i-2}-x_{4i-1})^6+(\tan(x_{4i-1}-2x_{4i}))^2+ (x_{4i-3})^8(x_{4i-9})^2] ,\quad x_0=(-1,2,2,2,\dots)^T$$

4- Wolfe function: -

$$f= [-x_1(3- x_1/2)+2x_2 -1)]^2+\sum_{i=1}^{n-1}[x_{i-1}-x_1(3- x_1/2)+2 x_2 -1)]^2+[-x_1(3- x_1/2)+2 x_2 -1)]^2,\quad x_0=(-1,\dots)^T$$

5-Non-diagonal variant of Rosenbrock function: -

$$f=\sum_{i=1}^{n/2} [100(x_1- x_i^2)^2 +(1-x_i)^2] \qquad x_0=(-1,\dots)^T$$

6-Rosenbrock function: -

$$f=\sum_{i=1}^{n/2} [100(x_1- x_i^2)^2 +(1-x_i)^2] \qquad x_0=(-1.2,1,\dots)^T$$

REFERENCES

[1] Al-Bayati, A.Y. (1991) "A New Family Of Self– Scaling Variable– Metric Algorithms for Unconstrained Optimization", **Journal of education and Science**, (12), PP.25-54.

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

[3] Broyden, C.G. (1970) "The Convergence of a Class of Double–Rank Minimization Algorithms", **Journal of Institute of Mathematics , and its Application**, (6), PP.221-231.

[4] Broyden, C.G. (1967) "Quasi-Newton Method and their Application to function Minimization " ,**Math. Comp.**, 21, PP.368-381.

[5] Biggs, M.C. (1973) "A Note on Minimization Algorithm Which Make Use of Non-Quadratic Properties of the Objective Function ", **Journal of Institute of Mathematics ,and its Application**,(12), PP.337-338.

[6] Fletcher, R. and M.J.D. Powll (1963) "A Rapidly Convergent Descent Method For minimization", **Computer Journal**,(6),PP.163-168.

[7] Goldfarb, D. (1970) " A Family of Variable Metric Updates Derived Variational Means " **Mathematics of computations**,(24), PP.23-26

[8] Oren, S.S. (1974) "On Selection of Parameters in Self Scaling Variable " **mathematical programming**, 3,1974.

[9] Shanno D.F. (1970) "Conditioning of Quasi-Newton Method for Unconstrained optimization"**, Mathematics of computations** ,24, PP.647-656.

[10] WANG,C. and M. Li (2005) "Convergence Property of the FRCG with Error ", **Journal of industrial and management optimization**, Vol. (1),(2), PP.193-200.