New Variable Metric Algorithm by The Mean of 2nd Order Quasi-Newton Condition

Abbas Y. Al Bayati
profabbasalbayati@yahoo.com
College of Computer Sciences and Mathematics
University of Mosul, Iraq

Runak M. Abdullah
College of Science
University of Suleimani

Received on: 17/11/2006 **Accepted on: 23/01/2007**

ABSTRACT

In this paper a new class of Quasi-Newton update for solving unconstrained nonlinear optimization problem is proposed. In this work we suggested a new formula for the variable metric update with a new quasi-Newton condition used for the symmetric rank two formula.

Finally, a numerical study is reported in which the performance of this new algorithm is compared to that of various members of the unmodified family. Numerical experiments indicate that this new algorithm is effective and superior to the standard BFGS and DFP algorithms, with respect to the number of functions evaluations (NOF) and number of iterations (NOI).

Keywords: Unconstrained Optimization, Quasi-Newton Condition, Inexact Line Search.

خوارزمية المتري المتغير الجديد التي تحقق شرط QN الجديد لصيغة متماثلة من الرتبة الثانية

عباس يونس البياتي
كلية علوم الحاسوب والرياضيات، جامعة الموصل

روناك محمد عبدالله
كلية العلوم، جامعة السليمانية

تاريخ قبول البحث : 2007/1/23 **تاريخ استلام البحث : 2006/11/17**

الملخص

في هذا البحث تم اقتراح نوع جديد من خوارزميات المتري المتغير لحل المسائل غير الخطية في الأمثلية غير المقيدة. في هذا العمل تم اقتراح صيغة جديدة للمتري المتغير التي تحقق شرط جديد بصيغة QN لصيغة متماثلة من الرتبة الثانية.

وأخيرا تم حساب النتائج العددية مع التأثير على أن الخوارزمية الجديدة كفوءة مقارنة مع الخوارزميات المتماثلة في العائلة. التجارب العددية أثبتت أن الخوارزمية الجديدة كفوءة مقارنة مع خوارزميتي BFGS و DFP بالاعتماد على مقاييس حساب الدوال والتكرارات.

الكلمات المفتاحية: الأمثلية غير المقيدة، شرط شبيه–نيوتن، خط بحث غير مضبوط.

1. Introduction

The Quasi-Newton family of variable metric formula introduced by Broyden [2]. It is the most efficient technique for minimizing a non –linear function $f(x)$. Generating a sequence of points x_k and matrices H_k by the following procedure we have:

$$d_k = -H_k g_k, \quad k = 0,1,2,... \qquad ...(1)$$

where g_k is the gradient of f at x_k.

$$x_{k+1} = x_k + \alpha_k d_k \qquad ...(2)$$

The scalar α_k is chosen to ensure, at least that

$$f(x_{k+1}) < f(x_k) - \gamma\alpha_k d_k^T g_k \qquad ...(3)$$

For some predetermined γ, or more likely in theoretical analysis it is chosen to

min $f(x_k + \alpha_k d_k)$ …(4)

so that

$g_{k+1}^T d_k = 0$, …(5)

this is an exact line search (ELS).

Having determined the point x_{k+1} an improved inverse Hessian matrix H_{k+1} is obtained by incorporating the information generated in the last iteration.

The new matrix H_{k+1} is given by

$$H_{k+1} = H_k + \frac{v_k v_k^T}{v_k^T y_k} - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} + \phi R_k R_k^T \quad \ldots(6)$$

where

$y_k = g_{k+1} - g_k$

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

$$R_k = \frac{v_k}{v_k^T y_k} - \frac{H_k^T y_k}{y_k^T H_k y_k}$$

$H_{(0)} = I$

ϕ is parameter $\in (0,1)$,

Different values of the scalar ϕ in equation (6) correspond to different member of Broyden's Quasi-Newton family. It will be noted that $\phi = 0$ corresponds to the original variable metric algorithm introduced by Davidon [4] and Fletcher and Powell (DFP) [5]. In studying the theoretical behavior of these techniques it was shown by Fletcher and Powell that, on quadratic function with the accurate line search defined in (5), the original ($\phi = 0$) formula generates conjugate directions and hence minimizes a quadratic function in at most k iterations. It has frequently been postulated that fact accounts for the region near the minimum has been found.

One of the best known VM method is the BFGS method that was proposed independently by Broyden, Fletcher, Goldfarb and Shanno. The BFGS update is defined by formulae (6) where $\phi = 1$

Broyden (1970) [2] has shown that if search along all k conjugate directions is necessary then, with analysis based on the error matrix

$$K = G^{-\frac{1}{2}} H G^{-\frac{1}{2}}$$

and determined a value of ϕ for which the sequence H_k converges steadily to G^{-1}.

2. New Algorithm

In this section a new formula of preconditioned conjugate gradient (PCG) method is presented with inexact line search.

If H_{k+1} is to be viewed as an approximation to G^{-1}, it is natural to require that

$H_{k+1} y_k = v_k$ …(7)

Which is called the Quasi-Newton condition.

For the new algorithm we derive a new expression for the QN condition as follows [2]:

Let

$H_k y_k = \alpha_k v_k$ …(8)

Where $\alpha_k > 0$, and

$H_{k+1} y_k = v_k$ …(9)

Then we get the following relationship

$$H_{k+1} = \frac{1}{\alpha_k} H_k$$

and we obtain the following results

$$\frac{H_{k+1}}{H_k} = \frac{1}{\alpha_k}. \quad \text{…(10)}$$

These implies that the condition number

$$K(H) = \frac{1}{\alpha_k}, \quad \text{…(11)}$$

therefore, to compute a new formula, α_k will be chosen as follows

$$\left.\begin{array}{l} \alpha_k = \dfrac{y_k^T H_k y_k}{v_k^T y_k} \\ \text{or } \ \alpha_k \geq 2 \end{array}\right) \begin{array}{l} \text{Al-Bayati [1]} \\ \text{New suggestion} \end{array} \quad \text{…(12)}$$

then we have

$H_{k+1} y_k = v_k (1 + \alpha_k) - H_k y_k$ …(13)

Now dividing (13) by y_k and multiplying and dividing the first and the second terms of right hand side of equation(13) by v_k we get

$$H_{k+1} = \frac{(1+\alpha_k) v_k v_k^T}{y_k^T v_k} - \frac{H_k y_k v_k^T}{y_k^T v_k} \quad \text{…(14)}$$

To obtain the correctly approximation of G^{-1} we can rewrite equation (14) as follows

$$H_{k+1} = H_k - \frac{H_k y_k v_k^T}{y_k^T v_k} + (1+\alpha_k) \frac{v_k v_k^T}{y_k^T v_k} \quad \text{…(15)}$$

where α_k defined by (12).

This is a new update H_{k+1} which satisfies a new QN-like condition.

3. Outlines of The New Suggested Algorithm

Step(1): set x_0, ε, $H_0 = I$

Step(2): For $k = 1$ set $d_1 = -H_1 g_1$

Step(3): Set $x_{k+1} = x_k + \alpha_k d_k$ where α_k is optimal step- size.

Step(4): If $\|g_{k+1}\| \leq \varepsilon$ stop , otherwise

Step(5): Compute $v_k = x_{k+1} - x_k$, $y_k = g_{k+1} - g_k$

Step(6): Compute H_{k+1} by equation (15).

Step(7): $d_{k+1} = -H_{k+1} g_{k+1} + \beta_k d_k$ where $\beta_k = \dfrac{y_k^T H_{k+1} g_{k+1}}{d_k^T y_k}$ and H_{k+1} is defined by (15).

Step(8): If restart criterion is satisfied , i.e. $g_{k+1}^T d_{k+1} > 0$ and $g_k^T g_k > g_{k+1}^T g_{k+1}$ go to step(2). Else $k = k+1$ and go to step(3).

Theorem (3.1):

Let f be given by

$$f(x) = \frac{1}{2}x^T Gx + b^T x$$

Where G is symmetric positive definite. Choose an initial approximation $H_1 = H$, where H is any symmetric positive definite matrix of appropriate order. Obtain H^*_{new} from H where $d = -Hg$ is the search direction and assuming exact line searches then $H_{i+1}g^* = Hg^*$, for $0 \le i < k \le n$., $g_{k+1} = g^*$

Proof:

Apply induction on i. Let $H_0 = H$, on the above assumptions so that

$H_0 g^* = Hg^*$

using formula(16) we have

$$H_{i+1}g^* = H_k g^* - \frac{H_k y_k v_k^T}{y_k^T v_k} g^* + (1+\alpha)\frac{v_k v_k^T g^*}{y_k^T v_k} \quad \ldots(16)$$

since $v_k^T g^* = 0$ (ELS)., equation (16) become

$H_{i+1}g^* = Hg^*$

we prove it is true for *i*+1 by using the following two standard properties (not proved here) which are satisfied in the case of a quadratic function and exact line searches).

(a): $v_j^T g^* = 0$, for j = 1, 2, … , k

(b): $g_j^T H_j g^* = 0$, for j = 1, 2, … , k

The orthogonality property satisfies when f is quadratic function and exact line searches is used. Hence $H_{i+1}g^* = Hg^*$ #

4. Numerical Results:

Seven well-known test functions (given in the appendix) were tested with different dimensions ($4 \le k \le 1000$).

All programs are written in FORTRAN 90 language and for all cases the stopping criterion is taken to be $\|g_{k+1}\| < 1*10^{-5}$.

The line search routine used was cubic interpolation which uses function and gradient values and it is an adaptation of the routine published by Bundy [3].

The results are given in the Table (1) and specifically quoting the number of functions (NOF) and the number of iterations (NOI).

Experimental results in table (1A) confirm that the new algorithm is superior to standard DFP and BFGS methods. Namely there are an improvement of the new suggested about % 34.87 NOI and % 23.82 NOF by using the new suggested VM-algorithm in this.

Table (1A). Comparison of DFP, BFGS and new algorithm

Test Function	N	DFP		BFGS		New Method	
		NOI	NOF	NOI	NOF	NOI	NOF
Powel	4	22	79	18	62	28	59
	100	82	487	38	98	30	64
	500	54	284	40	100	30	64
	1000	56	294	50	128	31	66
Wood	4	40	233	36	101	29	63
	100	243	861	235	670	29	63
Cubic	4	19	68	19	56	11	32
	100	27	93	64	150	11	32
	500	48	158	81	183	11	32
	1000	76	333	82	188	11	32
Rosen	4	23	62	34	93	20	56
	100	68	187	88	256	20	56
	500	155	389	238	671	20	56
Dixon	4	10	31	9	24	9	20
	10	31	90	19	41	24	50
Shallow	4	8	26	6	18	6	17
	100	8	26	6	18	6	17
	500	8	26	6	18	6	17
	1000	8	26	6	18	6	17
Non-diagonal	4	24	72	18	55	16	41
	100	48	115	76	170	25	62
	500	74	175	86	194	21	54
	1000	78	188	86	200	22	55
Total		**1210**	**4303**	**1341**	**3512**	**422**	**1025**

From Table (1A)

Percentage performance of the new algorithm against both DFP and BFGS algorithms for 100% DFP. We have:

Table (1B)

Tools	DFP	BFGS	New
NOI	100%	65.13	34.87
NOF	100%	76.18	23.82

5. Appendix :

All the test functions used in this paper are from general literature: See [1] for the details of all these test functions.

1. Powell function (Generalized form)

$$f = \sum_{i=1}^{n/4} [(x_{4i-3} + 10x_{4i-2})^2 + 5(x_{4i-1} - x_{4i})^2 + (x_{4i-2} - 2x_{4i-1})^4 + 10(x_{4i-3} - x_{4i})^4] \ ,$$

$$x_0 = (3,-1,0,1;...)^T$$

2. Rosen function

$$f = 100(x_2 - x_1^2)^2 + (1 - x_1)^2 \,,\; x_0 = (-1.2, 1.0)^T$$

3. Cubic function

$$f = 100(x_2 - x_1^3)^2 + (1 - x_1)^2 \,,\; x_0 = (-1.2, 1.0)^T$$

4. Shallow function

$$f = \sum_{i=1}^{n/2} (x_{2i-1}^2 - x_{2i})^2 + (1 - x_{2i-1})^2 \;,\; x_0 = (-2;...)^T$$

5. Dixon function

$$f = (1 - x_1)^2 + (1 - x_{10})^2 + \sum_{i=2}^{n} (x_i^2 - x_{i+1})^2 \,,\; x_0 = (-1;...)^T$$

6. Non-diagonal function

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

7. Wood function

$$f = \sum_{i=1}^{n/4} [100(x_{4i-2} - x_{4i-3}^2)^2 + (1 - x_{4i-3})^2 + 90(x_{4i} - x_{4i-1}^2)^2 + (1 - x_{4i-1})^2 + 10.1(x_{4i-2} - 1)^2$$

$$+ (x_{4i} - 1)^2 + 19.8(x_{4i-2} - 1)(x_{4i} - 1) \quad ,\; x_0 = (-3,-1;-3,-1;...)^T$$

REFERENCES

[1] Al-Bayati, A. Y., ”A New Family of Self-Scaling Variable Metric Algorithm for Unconstrained Optimization”, Research Report series 89.23, November (1989).

[2] Broyden, C.G., "The Convergence of a Class of Double Rank Minimization Algorithms II. The New Algorithm", Journal of The Institute of Mathematics And it's Applications 6 (1970) 221-231.

[3] Bunday, B., "A Basic Optimization Methods", Edward Arnold, Bedford Square, London, (1984).

[4] Davidon, W.C., "Variable Metric Methods for minimization", Argonn, National Laboratory, Pept-ANL-5990(Rev)., (1959).

[5] Fletcher, R., and M.J.D. Powel, “A Rapidly Convergent Descent Method For Minimization”, Computer Journal 6 (1963), 163-168.

[6] Lei-Hong Zhang, Ping-Qi Pan "A Note on Pan's Second-order Quasi-Newton Updates". National Science Foundation of China (No. 1037017), (2004).