New Variable Metric algorithms for Uncnostrained optimization

Hussein O. Dahwi , Basim A. Hassan , Thaair M. Y. Ameen
Department of mathematics, College of Computers Sciences and Mathematics , University of Mosul, Mosul , IRAQ
(**Received:** 9 / 4 / 2009 ---- **Accepted:** 21 / 2 / 2010)

Abstract

In unconstrained optimization, the original quasi-Newton condition $B_{k+1}v_k = y_k$ where y_k is the difference of the gradients at the least two iterations. In this paper, we propose VM-algorithms based on a new Quasi–Newton equation $B_{k+1}v_k = y_k^*$ where y_k^* defined in eq. (24).

Experimental results indicate that the new proposed algorithms was more efficient than the standard BFGS & DFP algorithms.

Introduction

Consider the unconstrained optimization problem

$$\min\left\{ f(x) \;\middle|\; x \in R^n \right\} \qquad \text{..........(1)}$$

where f is a continuously differentiable function of n variables .Quasi-Newton methods for solving (1) often needed to update the iterate matrix B_k .Traditionally , $\{B_k\}$ satisfies the following quasi-Newton equation :

$$B_{k+1}v_k = y_k \qquad \text{..........(2)}$$

where $v_k = x_{k+1} - x_k$, $y_k = g_{k+1} - g_k$. [3,5]

The search direction is computed by

$$d_k = -B_k^{-1} g_k \qquad \text{..........(3)}$$

where g_k is the gradient of f evaluated at the current iteration x_k .One then computes the next iteration by

$$x_{k+1} = x_k + \alpha_k d_k \qquad \text{.........(4)}$$

where the step size α_k satisfies the Wolfe – Powell (WP) conditions

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

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

where $\delta_1 < 1/2$ and $\delta_1 < \delta_2 < 1$. **[1]**

In **[4,6]** the famous update B_k is the Broyden-Fletcher-Goldfarb-Shanno (BFGS) formula for which B_{k+1} is updated as :

$$B_{k+1} = B_k - \frac{B_k v_k v_k^T B_k}{v_k^T B_k v_k} + \frac{y_k y_k^T}{v_k^T y_k} \qquad \text{..........(7)}$$

& the Davidon-Fletcher-Powell (DFP) formula defined by

$$B_{k+1} = B_k - \left(1 + \frac{v_k^T B_k v_k}{v_k^T y_k}\right)\frac{y_k y_k^T}{v_k^T y_k} - \frac{y_k v_k^T B_k + B_k v_k y_k}{v_k^T y_k} \qquad \text{...(8)}$$

Modified Quasi-Newton Equation :

In this section we derive the modified quasi-Newton equation which exploit not only the gradients but also the function values. Now we assume that the function f is smooth enough. From

$$f_k = f_{k+1} - g_{k+1}^T v_k + \frac{1}{2} v_k^T G_{k+1} v_k - \frac{1}{6} v_k^T (T_{k+1} v_k) v_k + O(\|v_k\|^4) \qquad \text{.........(9)}$$

$$v_k^T g_k = g_{k+1}^T v_k - v_k^T G_{k+1} v_k + \frac{1}{2} v_k^T (T_{k+1} v_k) v_k + O(\|v_k\|^4) \qquad \text{.........(10)}$$

Where T^{n*n*n} is the tensor of f at x_{k+1} & $v_k^T (T_{k+1} v_k) v_k = \sum \frac{\partial^3 f(x_{k+1})}{\partial x^i \partial x^j \partial x^l} v_k^i v_k^j v_k^l$

We obtain, by canceling the terms which include the tensor,

$$v_k^T G_{k+1} v_k = (g_{k+1} - g_k)^T v_k + 6(f_k - f_{k+1}) + 3(g_{k+1} + g_k)^T v_k \qquad \text{.........(11)}$$

$$\frac{1}{6} v_k^T G_{k+1} v_k = \frac{1}{6}(g_{k+1} - g_k)^T v_k + (f_k - f_{k+1}) + \frac{1}{2}(g_{k+1} + g_k)^T \qquad \text{..........(12)}$$

where $v_k = \alpha_k d_k = -\alpha_k g_k$. Since $B_{k+1}v_k$ is required to approximate $G_{k+1}v_k$

$$B_{k+1} = \frac{1}{\alpha_K^2 g_k^T g_k} \left\{ 6(f_k - f_{k+1}) - \alpha_k (g_{k+1} - g_k)^T g_k - 3\alpha_k (g_{k+1} + g_k)^T g_k \right\} \qquad \text{...(13)}$$

Then we take B_{k+1} as an approximation of G_{k+1}. It is clear that if $B_{k+1} > 0$ this approximation will be positive definite hence to complete the method .

We must consider the situation when $B_{k+1} < 0$, i.e. if

$$6(f_k - f_{k+1}) - \alpha_k (g_{k+1} - g_k)^T g_k - 3\alpha_k (g_{k+1} + g_k)^T g_k < 0 \qquad \text{..........(14)}$$

In this case we can change the step size α_k as $\alpha_k + \rho_k$ s.t.

$$6(f_k - f_{k+1}) - \alpha_k (g_{k+1} - g_k)^T g_k - 3\alpha_k (g_{k+1} + g_k)^T g_k > 0 \qquad \text{..........(15)}$$

$$6(f_k - f_{k+1}) - \alpha_k (g_{k+1} - g_k)^T g_k - 3\alpha_k (g_{k+1} + g_k)^T g_k - \delta = 0 \qquad \text{.......(16)}$$

$$6(f_k - f_{k+1}) - [\alpha_k + \rho_k](g_{k+1} - g_k)^T g_k \quad \text{.......(17)}$$
$$-3[\alpha_k + \rho_k](g_{k+1} + g_k)^T g_k - \delta = 0$$
$$\therefore \rho_k = \frac{1}{(g_{k+1} - g_k)^T g_k + 3(g_{k+1} + g_k)^T g_k} Q \quad \text{.........(18)}$$

Wher $Q =$

$$6(f_k - f_{k+1}) - \alpha_k (g_{k+1} - g_k)^T g_k - 3\alpha_k (g_{k+1} + g_k)^T g_k - \delta \quad \text{....(19)}$$

where $\delta \geq 0$, slack variable.

$$v_k^T B_{k+1} v_k = (g_{k+1} - g_k)^T v_k + 6(f_k - f_{k+1}) + 3(g_{k+1} + g_k)^T v_k \quad \text{.........(20)}$$

$$v_k^T B_{k+1} v_k = (g_{k+1} - g_k)^T v_k + \theta, \quad \text{.........(21)}$$

where $\theta = 6(f_k - f_{k+1}) + 3(g_{k+1} + g_k)^T v_k$

From (20) we get

$$B_{k+1} v_k = y_k + \frac{\theta}{v_k^T v_k} v_k = y_k - \frac{\theta}{[\alpha_k + \rho_k]^2 g_k^T g_k} [\alpha_k + \rho_k] g_k \quad \text{.......(22)}$$

$$B_{k+1} v_k = y_k - \frac{\theta}{[\alpha_k + \rho_k]^2 g_k^T g_k} [\alpha_k + \rho_k] g_k \quad \text{.........(23)}$$

Where $B_{k+1} v_k = y_k^*$

$$y_k^* = y_k - \frac{\theta}{[\alpha_k + \rho_k]^2 g_k^T g_k} [\alpha_k + \rho_k] g_k \quad \text{.........(24)}$$

It is clear that this Quasi-Newton equation yields the following two new formulas:

- Formula 1: Using the new quasi-Newton equation for BFGS updates we can obtain a BFGS-Type formula as follows:

$$B_{k+1} = B_k - \frac{B_k v_k v_k^T B_k}{v_k^T B_k v_k} + \frac{y_k^* y_k^{*T}}{v_k^T y_k^*} \quad \text{.........(25)}$$

- Formula 2 : Using the new quasi-Newton equation for DFP updates we can obtain a DFP-Type formula as follows:

$$B_{k+1} = B_k - \left(1 + \frac{v_k^T B_k v_k}{v_k^T y_k^*}\right) \frac{y_k^* y_k^T}{v_k^T y_k^*} - \frac{y_k^* v_k^T B_k + B_k v_k y_k^{*T}}{v_k^T y_k^*} \quad \text{.........(26)}$$

New Algorithms (2.1)

The outline of the New algorithms as follows :

Step 0 : Choose an initial point $x_1 \in R^n$ and an initial positive definite matrix B_1, set $k = 1$.

Step 1 : If $g_k = 0$, stop .

Step 2 : Solve $B_k d_k + g_k = 0$ to obtain a search direction d_k .

Step 3 : Find α_k by (WP) step size rule (5) and (6) .

Step 4 : Generate a new iteration point by $x_{k+1} = x_k + \alpha_k d_k$ and calculate

The new updating formula (25) and (26) .

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

Numerical Results :

In this section, we have been reported the numerical results for the new formula (25) & (26).The test functions are commonly used unconstrained test problems with standard points. We have used the dimension of the problem (N), N=10,50,100,500,1000. For each problem, we choose the initial matrix $H_{k+1} = I$. These algorithms use the cubic fitting technique which satisfy Wolfe-Powel (WP) condition as a line search technique in which $\delta_1 = 0.0001$ $\delta_2 = 0.1$,**[2]**. The convergence criterion employed is a test on the norm of the gradient, $\|g_k\| < 0.0001$

We will test the following VM-algorithms.

1. The standard BFGS algorithm .
2. The standard DFP algorithm.
3. Modified (1) algorithm (BFGS-type formula defied in (25).
4. Modified (2) algorithm (DFP-type formula defied (26).

Shows the computation results , where the columns have the following meanings :

Problem : the name of the test problem .

NOI : number of iterations .

NOF : number of function evaluations .

From Table **(3-1)** ,we observe that the average performances of the formula (25) are better than the standard BFGS and for unconstrained minimization problems. We claim that the formula (25) will be more efficient and better than the standard BFGS formula. Namely, there are about 7 % improvement in NOF and 8 % improvement in NOI **&** from Table **(3-2)** ,we observe that the average performances of the formula (26) are better than the standard DFP and for unconstrained minimization problems. We claim that the formula (26) will be more efficient and better than the standard DFP formula. Namely, there are about 16 % improvement in NOF and 2 % improvement in NOI .

Table (3-1): Comparison results of BFGS algorithm & formula (25).

Test-function	N	Modified (1) algorithm NOF (NOI)	BFGS-algorithm NOF (NOI)
Non-diagnal	10	115 (39)	108 (40)
	50	119 (50)	117 (50)
	100	113 (56)	135 (56)
	500	139 (58)	140 (59)
	1000	152 (63)	150 (62)
Wolfe	10	32 (15)	32 (15)
	50	101 (50)	101 (50)
	100	125 (62)	125 (62)
	500	141 (70)	141 (70)
	1000	180 (89)	168 (83)
Cubic	10	71 (26)	70 (26)
	50	95 (40)	81 (33)
	100	92 (40)	98 (41)
	500	105 (45)	114 (48)
	1000	109 (465)	112 (48)
Powell	10	57 (18)	82 (23)
	50	72 (25)	145 (52)
	100	112 (41)	140 (51)
	500	102 (38)	102 (34)
	1000	127 (44)	120 (39)
Miele	10	66 (21)	65 (21)
	50	78 (24)	89 (29)
	100	84 (27)	93 (30)
	500	85 (28)	97 (31)
	1000	87 (29)	93 (30)
total		2549 (1044)	2718 (1131)

The percentages of improvements given by :

	BFGS	Modified (1)
NOI	100	92.30
NOF	100	93.78

Table (3-2): Comparison results of DFP algorithm & formula (26).

Test-function	N	Modified(2) algorithm NOF (NOI)	DFP-algorithm NOF (NOI)
Non-diagnal	10	315 (52)	136 (41)
	50	174 (64)	182 (67)
	100	142 (51)	140 (50)
	500	159 (57)	154 (56)
	1000	162 (59)	164 (60)
Wolfe	10	33 (15)	32 (15)
	50	101 (50)	101 (50)
	100	125 (62)	125 (62)
	500	141 (70)	141 (70)
	1000	169 (84)	165 (82)
Cubic	10	88 (27)	92 (28)
	50	110 (38)	91 (33)
	100	105 (37)	123 (45)
	500	115 (42)	130 (47)
	1000	136 (49)	142 (52)
Powell	10	69 (18)	283 (49)
	50	78 (24)	436 (70)
	100	146 (42)	447 (75)
	500	145 (39)	244 (45)
	1000	160 (40)	242 (45)
Miele	10	83 (21)	79 (22)
	50	90 (23)	125 (28)
	100	112 (27)	125 (28)
	500	173 (34)	145 (32)
	1000	178 (36)	144 (32)
total		3309 (1061)	3903 (1074)

The percentages of improvements given by :

	DFP	Modified(2)
NOI	100	98.78
NOF	100	84.78

Conclasion :

In this paper, we have proposed a VM-type for unconstrained minimization based on a modified quasi-Newton condition. The computational experiments show that the modified approach given in this paper is very successful.

Appendix

1.*Cubic function* :

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

$$Starting\ point:(-1.2,1,-1.2,1,......)^T$$

2.*Non – diagonal function* :

$$f(x)=\sum_{i=1}^{n/2}(100(x_i-x_i^3)^2+(1-x_i)^2)$$

$$Starting\ point:(-1,.......................)^T$$

3.*Generalized powell function* :

$$f(x)=\sum_{i=1}^{n/4}(x_{4i-3}-10x_{4i-2})^2+5(x_{4i-1}-x_{4i})^2+(x_{4i-1}-2x_{4i})^2+10(x_{4i-9}-x_{4i})^4+(x_{4i-2}-2x_{4i-1}-x_{4i})^2)$$

$$Starting\ point:(3,1,0,1,.......................)^T$$

4.*Miele function* :

$$f(x)=\sum_{i=1}^{n/4}[\exp(x_{4i-3})-x_{4i-2}]^2+100(x_{4i-2}-x_{4i-1})^6+[\tan(x_{4i-1}-x_{4i})]^4+x_{4i-3}^8+(x_{4i}-1)^2$$

$$Starting\ point:(1,2,2,2,................)^T$$

5.*Welfe function* :

$$f(x)=(-x_1(3-x_1/2)+2x_2-1)^2+\sum_{i=1}^{n-1}(x_{i-1}-x_i(3-x_i(3-x_i/2)+2x_{i+1}-1)^2+(x_{n+1}-x_n(3x_n/2-1)^2$$

$$Starting\ point:(-1,.......................)^T$$

References

1- Al-Bayati, A.Y. and Al-Salih, M.S.,(2008):New self-scaling VM-algorithm for non – convex optimization, to be appear in the 2nd confers Aleppo-Syria.

2- Bunday B., (1984):Basic Optimization Methods, Edward Arold, Bedford Square, London, U. K.

3- Fletcher R. (1987): Practical Methods of Optimization, Wiley , Chi Chester .

4- Kinsella J.,(2006):Course Notes for MS4327 Optimization.

5- Storey C. and HU .Y. F., (1993): Preconditioned Low-Order Newton methods , J. of optimization theory and application ,Vol. 79, No. 2, pp.311-331.

6- Onwubiko C., (2000): Introduction to Engineering Design Optimization, chapter 4, Prentice Hall.

خوارزميات جديدة للمتري المتغير في الامثلية غير المقيدة

حسين عبيد ضحوي[١] ، باسم عباس حسن [٢] ، ثائر محمد يونس[١]

[١] *رئاسة جامعة الموصل ، الموصل ، العراق*

[٢] *كلية علوم الحاسبات والرياضيات ، جامعة الموصل ، الموصل ، العراق*

(تاريخ الاستلام: ٩ / ٤ / ٢٠٠٩ ---- **تاريخ القبول:** ٢١ / ٢ / ٢٠١٠)

الملخص

في الامثلية غير المقيدة الشرط الأساسي لخوارزمية شبيه نيوتن هو $B_{k+1}v_k = y_k$ عندما y_k تعرف بأنها فرق المشتقات المتعاقبة. في هذا البحث نتطرق للتقصي عن خوارزميات جديدة للمتري المتغير عندما تستخدم الشرط التالي $B_{k+1}v_k = y_k^*$ المعرفة في (٢٤). النتائج العددية أشارت إلى ان الخوارزميات الجديدة أكثر كفاءة من خوارزميات BFGS و DFP القياسية .