

## An Alternative Scaling Factor In Broyden's Class Methods for Unconstrained Optimization

<sup>1</sup>Muhammad Fauzi bin Embong, <sup>2</sup>Mustafa bin Mamat, <sup>1</sup>Mohd Rivaie and <sup>2</sup>Ismail bin Mohd

<sup>1</sup>Department of Computer Science and Mathematics, University Technology MARA,  
 Kuala Terengganu Campus, 21080 Kuala Terengganu, Malaysia

<sup>2</sup>Department of Mathematics, Faculty of Science and Technology,  
 University Malaysia Terengganu, 21030 Kuala Terengganu, Malaysia

**Abstract: Problem statement:** In order to calculate step size, a suitable line search method can be employed. As the step size usually not exact, the error is unavoidable, thus radically affect quasi-Newton method by as little as 0.1 percent of the step size error. **Approach:** A suitable scaling factor has to be introduced to overcome this inferiority. Self-scaling Variable Metric algorithms (SSVM's) are commonly used method, where a parameter is introduced, altering Broyden's single parameter class of approximations to the inverse Hessian to a double parameter class. This study proposes an alternative scaling factor for the algorithms. **Results:** The alternative scaling factor had been tried on several commonly test functions and the numerical results shows that the new scaled algorithm shows significant improvement over the standard Broyden's class methods. **Conclusion:** The new algorithm performance is comparable to the algorithm with initial scaling on inverse Hessian approximation by step size. An improvement over unscaled BFGS is achieved, as for most of the cases, the number of iterations are reduced.

**Key words:** Broyden's class, eigenvalues, Hessian, scaling factor, self-scaling variable matrices

### INTRODUCTION

The quasi-Newton methods are very popular and efficient methods for solving unconstrained optimization problem:

$$\min_{x \in \mathbb{R}^n} f(x) \quad (1)$$

where,  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  is a twice continuously differentiable function. There are a large number of quasi-Newton methods but the Broyden's class of update is more popular. As other quasi-Newton method, the Broyden's method are iterative, whereby at (k+1)-th iteration:

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

where,  $d_k$  denotes the search direction and  $\alpha_k$  is its step size. The search direction,  $d_k$ , is calculated by using:

$$d_k = -H_k g_k \quad (3)$$

Where:

$g_k$  = The gradient of  $f$  evaluated at the current iterate  $x_k$

$H_k$  = The inverse Hessian approximation.

The step size  $\alpha_k$  is a positive step length chosen by a line search so that at each iteration either:

$$f(x_k + \alpha_k d_k) \leq f(x_k) - \eta_1 \alpha_k \frac{g_k^T d_k}{s_k^T y_k} \quad (4)$$

Or:

$$f(x_k + \alpha_k d_k) \leq f(x_k) - \eta_2 g_k^T d_k \quad (5)$$

where,  $\eta_1$  and  $\eta_2$  are positive constants.

Note that the conditions (4) and (5) are assumes in Byrd and Nocedal (1989). They cover a large class of line search strategies under suitable conditions. If the gradient of  $f$  is Lipschitz continuous, then for several well known line search satisfy the Wolfe conditions:

$$f(x_k + \alpha_k d_k) \leq f(x_k) - \delta_1 \alpha_k d_k^T g_k \quad (6)$$

$$(g(x_k + \alpha_k d_k))^T d_k \geq \delta_2 d_k^T g_k \quad (7)$$

where,  $1 < \delta < \frac{1}{2}$  and  $\delta_1 < \delta_2 < 1$  (4).

**Corresponding Author:** Muhammad Fauzi bin Embong, Department of Computer Science and Mathematics,  
 University Technology MARA, Kuala Terengganu Campus, 21080 Kuala Terengganu, Malaysia

Byrd and Nocedal (1989) prove that if the ratio between successive trial values of  $\alpha$  is bounded away from zero, the new iteration produced by a backtracking line search satisfies (4) and (5). The inverse Hessian approximation is then updated by:

$$H_{k+1}^\phi = H_k + \frac{s_k s_k^T}{s_k^T y_k} - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} \phi v_k v_k^T \quad (8)$$

Where:

$$s_k = x_{k+1} - x_k \quad (9)$$

$$y_k = g_{k+1} - g_k \quad (10)$$

$$v_k = \left( y_k^T H_k y_k \right)^{\frac{1}{2}} \left( \frac{s_k}{s_k^T y_k} - \frac{H_k y_k}{y_k^T H_k y_k} \right) \quad (11)$$

and  $\phi$  is a parameter that may take any real value.

According to Dennis and More (1977), there are two updated formulae that are contained in the Broyden's class method, namely BFGS update if the parameter  $\phi_k = 0$  and DFP update if  $\phi_k = 1$ . Consequently, (1.8) may be written as:

$$H^\phi = (1 - \phi)H^{DFP} + \phi H^{BFGS} \quad (12)$$

If we let  $\phi_1 \in [0,1]$  (1.8) is called Broyden convex family. Meanwhile, if  $\phi_1 \in [0,1-\sigma]$  for  $\sigma \in [0,1]$ . then (1.8) is called the restricted Broyden's class method (Byrd *et al.*, 1987).

## MATERIALS AND METHODS

### Scaling the Broyden's class method:

**Self-scaling Variable Metric (SSVM) method:** Many modifications have been applied on quasi-Newton methods in attempt to improve its efficiency. Now, the discussion will be on the self-scaling variable metric algorithms developed by Oren (1973) and Oren and Luenberger (1974). Multiplying  $H_k$  by  $\gamma_k$  and then replacing  $\gamma_k H_k$  in (8), the Broyden's class formula can be written as:

$$H_{k+1}^\phi = \left( H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} + \phi v_k v_k^T \right) \gamma_k + \frac{y_k y_k^T}{y_k^T s_k} \quad (13)$$

where  $\gamma_k$  is a self-scaling parameter. The formula (13) is known as self-scaling variable metric (SSVM) formula.

Clearly, when  $\gamma_k = 1$ , the formula (13) is reduced to Broyden's class update (8).

**Choices of the scaling factor:** The choice of a suitable scaling factor can be determined by the following theorem.

Theorem (Oren and Luenberger, 1974)

Let  $\phi \in [0,1]$  and  $\gamma_k > 0$ . Let  $H_k$  be the inverse Hessian approximation and  $H_{k+1}^\phi$  be defined by (13). Let  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$  and  $\mu_1^\phi \geq \mu_2^\phi \geq \dots \geq \mu_n^\phi$  be eigenvalues of  $H_k$  and  $H_{k+1}$  respectively. Then the following statements hold:

- If  $\gamma_k \lambda_n \geq 1$ , then  $\mu_n^\phi = 1$  and  $1 \leq \gamma_k \lambda_{i+1} \leq \mu_n^\phi \leq \gamma_k \lambda_i$ ,  $i = 1, 2, \dots, n-1$
- If  $\gamma_k \lambda_i \leq 1$  then  $\mu_n^\phi = 1$  and  $\gamma_k \lambda_i \leq \mu_i^\phi \leq \gamma_k \lambda_{i-1} \leq 1$ ,  $i = 2, 3, \dots, n$
- If  $\gamma_k \lambda_1 \leq 1 \leq \gamma_k \lambda_i$  and  $i_0$  is an index with  $\gamma_k \lambda_{i_0} \leq 1 \leq \gamma_k \lambda_{i_0-1}$ ,  $\gamma_k \lambda_1 \geq \mu_1^\phi \geq \gamma_k \lambda_2 \geq 1 \mu_2^\phi \geq \dots \geq \gamma_k \lambda_{i_0}$  then  $\mu_{i_0}^\phi \geq 1 \geq \mu_{i_0+1}^\phi \geq \gamma_k \lambda_{i_0+1} \geq \dots \geq \gamma_k \lambda_k$  and there is at least one eigenvalue in  $\mu_{i_0}^\phi$  and  $\mu_{i_0+1}^\phi$  which equals 1.

Readers who are interested in the proof for the above theorem may refer Oren and Luenberger (1974), or Sun and Yuan (2006). From the above theorem, it can be shown that:

$$\gamma_k = \frac{s_k^T y_k}{s_k^T H_k s_k} \quad (14)$$

is a suitable scaling factor (Sun and Yuan, 2006).

Shanno and Phua (1978) suggested a simple initial scaling which require no additional information about the object function than that routinely required by variable metric algorithm. Initially  $H_0 = I$  may be used to determine  $x_1$ , where  $\alpha_0$  is chosen according to some step length or linear search criterion to assure sufficient reduction in the function  $f$ . Once  $x_1$  has been chosen, but before  $H_1$  is calculated,  $H_0$  now being scaled by:

$$\hat{H}_0 = \alpha_0 H_0 \quad (15)$$

and:

$$H_1^\phi = \hat{H}_0 + \frac{s_k s_k^T}{s_k^T y_k} - \frac{\hat{H}_0 y_k y_k^T \hat{H}_0}{y_k^T \hat{H}_0 y_k} \phi v_k v_k^T \quad (16)$$

$$v_k = \left( y_k^T \hat{H}_0 y_k \right)^{\frac{1}{2}} \left( \frac{s_k}{s_k^T y_k} - \frac{\hat{H}_0 y_k}{y_k^T \hat{H}_0 y_k} \right) \quad (17)$$

Substitution of (15) into (16) yields:

$$H_1^\phi \left( H_0 - \frac{H_0 y_k y_k^T H_0}{y_k^T H_0 y_k} \phi v_k v_k^T \right) \alpha + \frac{y_k y_k^T}{y_k^T s_k} \quad (18)$$

After the initial scaling of  $H_0$  by an appropriate step size  $\alpha$ , the inverse Hessian approximation is never rescaled. Numerical experiments show that the initial scaling is simple and effective for a lot of problems in which the curvature changes smoothly (Sun and Yuan, 2006).

**An alternative scale factor:** In this article, the smallest eigenvalue of inverse Hessian approximation was proposed as an alternative scaling factor of initial scaling on  $H$  as in (15). Replacing step size,  $\alpha$  with the smallest eigen value of  $H$ ,  $\lambda$  into (18) yields:

$$H_1^\phi \left( H_0 - \frac{H_0 y_k y_k^T H_0}{y_k^T H_0 y_k} \phi v_k v_k^T \right) \lambda + \frac{y_k y_k^T}{y_k^T s_k} \quad (19)$$

As proposed by Shanno and Phua (1978), this update is also an initial scaling on inverse Hessian approximation. After the initial iteration, the inverse Hessian approximation is never rescaled. The following algorithm is proposed with the smallest eigenvalue of  $H_1$ ,  $\lambda$  as the scaling factor.

**Eigenvalue scaled algorithm:** For simplicity, we let  $\phi = 1$ , thus a modification of the BFGS method is obtained. For other members of Broyden's class methods, this algorithm is also applicable.

- Step 1: Initialization Given  $x_0$ , set  $k = 0$  and  $H_0 = I$ .
- Step 2: Computing search direction  $d_k = -H_k g_k$  If  $g_k = 0$ , then stop.
- Step 3: Computing step size,  $\alpha$ .
- Step 4: Updating new point,  $x_{k+1} = x_k + \alpha_k d_k$
- Step 5: Updating approximation of inverse Hessian approximation,  $H_{k+1}$ . For  $k = 1$ , use (19), else, use (8).
- Step 6: Convergent test and stopping criteria If  $(x_{k+1} < x_k)$  and  $\|g_k\| \leq \epsilon$ , then stop. Otherwise go to Step 1 with  $k = k+1$ .

### RESULTS

A MAPLE subroutine was programmed to test three algorithms, BFGS algorithm without scaling

(BFGS), with eigenvalue scaling (19) (denoted as ES-BFGS) and with the initial scaling (18) (denoted as S-BFGS). The three algorithms was applied on eight commonly tested functions, consist of two variables ( $n = 2$ ) (functions and four variable) ( $n = 4$ ) functions:

- Rosenbrock function with  $n = 2$

$$f(x) = 100(x_2 - x_1^2)^2 + (x_1 - 1)^2$$

- Cube function with  $n = 2$

$$f(x) = 100(x_2 - x_1^3)^2 + (x_1 - 1)^2$$

- Shalow function with  $n = 2$

$$f(x) = (x_2 - x_1^2)^2 + (1 - x_1)^2$$

- Strait function with  $n = 2$

$$f(x) = (x_2 - x_1^2)^2 + 100(x_1 - 1)^2$$

- Rosenbrock function with  $n = 4$

$$f(x) = 100(x_2 - x_1^2)^2 + (x_1 - 1)^2 + 100(x_4 - x_3)^2 + (x - 1_3)^2$$

- Cube function with  $n = 4$

$$f(x) = 100(x_2 - x_1^3)^2 + (x_1 - 1)^2 + 100(x_4 - x_3^3)^2 + (x_3 - 1)^2$$

- Shalow function with  $n = 4$

$$f(x) = (x_2 - x_1^2)^2 + (1 - x_1)^2 + (x_4 - x_3^2)^2 + (1 - x_3)^2$$

- Wood function with  $n = 4$

$$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2 + 90(x_4 - x_3^2)^2 + (1 - x_3)^2 + 10(x_2 + x_4 - 2)^2 + 0.1(x_2 - x_4)^2$$

The numerical results produced by implementing the three algorithms to the test functions are presented in the Table 1. The efficiency of the algorithm are based on the number of iterations needed to reach the minimum value of the functions. Algorithm with less iteration is considered more efficient. A tolerance of  $\epsilon = 10^{-6}$  is set as the stopping criteria.

Table 1: Numerical results produced by the three tested algorithms, BFGS without scaling (BFGS), with Eigenvalue Scaling (ES-BFGS) and with step size scaling (S-BFGS)

Test function	Initial point	Number of iterations		
		BFGS	ES-BFGS	S-BFGS
Rosenbrock (n = 2)	(-2,-2)	17	15	16
	(-100,100)	30	28	28
	(10 000, 10 000)	133	11	131
Cube (n = 2)	(-1.2, 1.6)	17	8	8
	(1.5,-150)	62	54	63
	(100, 50)	44	44	35
Shalow (n = 2)	(5, 5)	9	9	9
	(-100,100)	15	13	13
	(1000,-5000)	11	9	9
Strait (n = 2).	(2, 2)	5	5	5
	(100,100)	9	13	9
	(1000, 1000)	12	13	13
Rosenbrock (n = 4).	(-2,-2,-2,-2)	16	18	17
	(-100,100,100,100)	30	29	28
	(100, 100, 100, 1.5)	156	102	137
Cube (n = 4).	(1.5,-1.5, 1.5,-1.5)	42	40	39
	(10,-10, 10,-10)	18	17	18
	(15,-15, 15,-15)	45	28	24
Shalow (n = 4)	(2, 4, 2, 4)	7	8	7
	(-200,400,200,400)	326	23	23
	(2000, 2000, 2000, 2000)	78	74	99
Wood (n = 4)	(2,-2, 2,-2)	29	24	22
	(200,-5, 200,-5)	43	29	30
	(2000, 2000, 2000, 2000)	80	56	58

### DISCUSSION

From the calculation using MAPLE software, an interesting relation between step size and eigenvalues is observed. For initial iteration, the value of step size and the smallest eigen value of Hessian approximation is almost identical, but after a number of iterations, the difference increased accordingly. This explains why the ES-BFGS is effective on initial scaling only, a similarity with the scaled algorithm proposed by Shanno and Phua (1978). When the scaling on Hessian approximation was done on every iteration, the performance of the scaled BFGS deteriorated, or even failed at all.

The numerical results show that the new algorithm (ES-BFGS) performance is comparable to the algorithm with initial scaling (S-BFGS).

### CONCLUSION

An improvement over unscaled BFGS is achieved, as for most of the cases (18 out of 24), the number of iterations are reduced. Further investigation will be carried out using the alternative scaling factor,  $\lambda$  on the other types of quasi-Newton methods. The relationship between the smallest eigenvalue of Hessian approximation and the optimal step size is also of the interest in future research, triggering the possibility of

using eigenvalue as a new step size in the quasi-Newton methods.

For every test function, the minimum point is (0,0) for the two variables functions and (0,0,0,0) for the four variables functions . The minimum value is equal to zero for all the test functions.

### REFERENCES

- Byrd, R.H., J. Nocedal and J. Y.X Yuan, 1987. Global convergence of a class of quasi-newton methods on Convex problems. *SIAM J. Numer. Anal.*, 24: 1171-1189. DOI: 10.1137/0724077
- Byrd, R.H. and J. Nocedal, 1989. A tool for the analysis of quasi-Newton methods with application to unconstrained minimization. *SIAM J. Numer. Anal.*, 26: 727-739. DOI: 10.1137/0726042
- Dennis, J.E. and J.J. More, 1977. Quasi-Newton methods, motivation and theory. *SIAM Rev.*, 19: 46-89. DOI: 10.1137/1019005
- Oren, S.S. and D.G. Luenberger, 1974. Self-scaling Variable Metric (SSVM) algorithms. *Manage. Sci.*, 20: 845-862. URL: <http://www.jstor.org/stable/2630094>
- Oren, S.S., 1973. Self-scaling variable metric algorithms without line search for unconstrained minimization. *Math. Comput. Am. Math. Soc.*, 27: 863-874. URL: <http://www.jstor.org/stable/2005523>

Sun, W. and Y.X. Yuan, 2006. Optimization Theory and Method. Springer Optimization and Its Applications. 1st Edn., New York USA., ISBN: 10:0-387-24975-3, pp: 273-282.

Shanno, D.F. and K.H. Phua, 1978. Matrix conditioning and nonlinear optimization. Math. Programm., 14: 149-160. DOI:10.1007/BF01588962