Journal of Modern Mathematics and Statistics
MEDWELL PUBLICATIONS SINCE 2000

Modified BFGS Update Based on Determinant Property of Hessian Matrix

Saad Shakir Mahmood and Jaafer Hmood Eidi
Department of Mathematics, College of Education, University of Al-Mustansiriya, Baghdad, Iraq

Key words: Quasi-Newton condition, Zhang-Xu condition, unconstrained optimization, iteration, approximation, minimizer

Corresponding Author:
Saad Shakir Mahmood
Department of Mathematics, College of Education, University of Al-Mustansiriya, Baghdad, Iraq

Page No.: 22-26
Volume: 14, Issue 2, 2020
ISSN: 1994-5388
Journal of Modern Mathematics and Statistics

Abstract: The aim of this study is to modified the BFGS update based on the determinant property of Hessian matrix by multiply the vector y (difference between the next gradient and the current gradient) with a real number say such that the determinant of the next Hessian matrix equal to one at every iteration and because of the choice of the initial Hessian approximation can be identity matrix, so, the determinant of initial Hessian matrix is also equal one and hence, the sequence of Hessian matrix produced by the method never go to a near singular matrix numerically which make the program never break before get the minimizer of the objective function.

INTRODUCTION

Consider the unconstrained optimization problem Eq. 1:

$$\min_{x\in R^n} f: R^n \to R \tag{1}$$

BFGS update considered as a popular update to solve the unconstrained optimization problem[1], proposed a modified BFGS update by updating the vector y_k which represent the difference between the next gradient and the current gradient by multiply with a real number to satisfy any property where we needed in this case the Quasi-Newton condition must be extended to Zhang-Xu condition and we have the extended Quasi-Newton condition $B_{k+1}s_k = \beta_k y_k$ where, B_{k+1} is the next approximation of Hessian matrix, $s_k = x_{k+1}-x_k$, x_k is the current solution, x_{k+1} is the next solution, $y_k = \nabla f(x_{k+1})-\nabla f(x_k)$ and ∇f is the gradient of the objective function f.

The problem is to solve (Eq. 1) by produce a sequence of symmetric and positive definite Hessian matrix which never convergent to a near singular matrix which make the numerical computation break before getting the minimizer because of the singularity of the Hessian matrix numerically. The best solution of this problem is to fixed the value of the determinant of Hessian matrix far away from zero at every iteration, so, the program never break before get the minimizer.

MATERIALS AND METHODS

β BFGS update: Consider the BFGS update[2] Eq. 2:

$$B_{k+1} = B_k + \frac{y_k y_k^T}{s_k^T s_k} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} \tag{2}$$

By Zhang and Xu[1] condition Eq. 3:

$$y_k^{\wedge} = \beta_k y_k,\ \beta_k \in R \tag{3}$$

Based on Eq. 3, the Quasi-Newton condition becomes as follows Eq. 4:

$$B_{k-1}s_k = y_k^{\wedge} = \beta_k y_k \quad (4)$$

The solution of (Eq. 4) is Eq. 5:

$$B_{k+1} = B_k + \beta_k \frac{y_k y_k^T}{y_k^T s_k} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} \quad (5)$$

This is called the β-BFGS update, to determine β_k, the following lemma be needed.

Lemma 1: For the β-BFGS update the determinant of the next approximation of Hessian matrix is given by:

$$|B_{k+1}| = |B_k| \beta_k \frac{y_k^T s_k}{s_k^T B_k s_k}$$

Proof:

$$|B_{k+1}| = \left| B_k + \beta_k \frac{y_k y_k^T}{y_k^T s_k} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} \right|$$

Since, the current Hessian matrix is symmetric and positive definite then, exist a triangular matrix $L_k \varepsilon R^{nxn}$ such that $B_k = L_k L_k^T$ and then:

$$|B_{k+1}| = \left| L_k L_k^T + \beta_k \frac{y_k}{y_k^T s_k} y_k^T - \frac{L_k L_k^T s_k}{s_k^T B_k s_k} s_k^T L_k L_k^T \right| =$$
$$|L_k| \left| I + \beta_k \frac{L_k^{-1} y_k}{y_k^T s_k} \left(L_k^{-1} y_k\right)^T - \frac{L_k^T s_k}{s_k^T B_k s_k} \left(L_k^T s_k\right)^T \right| |L_k^T|$$

By Sherman-Morrison-Woodburg Eq. 4, we have:

$$|B_{k+1}| = |B_k| \left[1 + \beta_k \frac{\left(L_k^{-1} y_k\right)^T}{y_k^T s_k} L_k^{-1} y_k \right] \left[1 + \frac{\left(L_k^T s_k\right)^T L_k^T s_k}{s_k^T B_k s_k} \right] -$$
$$|B_k| \left[\beta_k \frac{\left(L_k^{-1} y_k\right)^T}{y_k^T s_k} L_k^T s_k \right] \left[\left(L_k^{-1} y_k\right)^T \frac{L_k^T s_k}{s_k^T B_k s_k} \right] = |B_k| \beta_k \frac{y_k^T s_k}{s_k^T B_k s_k}$$

Hence, if we set $|B_{k+1}| = |B_k| = 1$ then,

$$\beta_k = \frac{s_k^T B_k s_k}{s_k^T y_k}$$

and the β-BFGS update becomes as follows Eq. 6:

$$B_{k+1} = B_k + \frac{s_k^T B_k s_k}{\left(s_k^T y_k\right)^2} y_k y_k^T - \frac{B_k s_k s_k^T}{s_k^T B_k s_k} \quad (6)$$

Lemma 2: β-BBFGS update produced a symmetric Hessian matrix if the current Hessian matrix is symmetric.

Proof: Since, $B_k^T = B_k \left(y_k y_k^T\right)^T = y_k y_k^T$ and $\left(B_k S_k S_k^T B_k\right)^T = B_k S_k S_k^T B_k$ then the proof is complete. The next lemma is show that the β-BFGS update is preserve the positive definiteness of the Hessian matrix more better than BFGS done because the condition $y^T s > 0$ is a sufficient condition in BFGS update but in β-BFGS update this condition will be delete it.

Lemma 3: Given B_k symmetric and positive definite matrix then, β-BFGS update produced a positive definite Hessian matrix.

Proof: For $0 \neq z \varepsilon R^n$, we have:

$$z^T B_{k+1} z = z^T B_k z - \frac{z^T B_k s_k s_k^T B_k z}{s_k^T B_k s_k} + \frac{s_k^T B_k s_k}{\left(y_k^T s_k\right)^2} \left(z^T y_k\right)^2$$

The third term is positive, so that, clear now we must prove that the first term is greater than or equal the second term by using Cauchy-Shwarz inequality and since, B_k is symmetric positive definite then exist a lower triangular matrix $L_k \varepsilon R^{nxn} \ni B_k = L_k L_k^T$ and hence:

$$z^T B_k z - \frac{z^T B_k s_k s_k^T B_k z}{s_k^T B_k s_k} = \left(L_k^T z\right)^T L_k^T z - \frac{\left(L_k^T Z\right)^T L_k^T s_k \left(L_k^T s_k\right)^T L_k^T z}{\left(L_k^T s_k\right)^T L_k^T s_k}$$

If we set $a = L_k^T z$ and $b = L_k^T s_k$ then, we have:

$$z^T B_k z - \frac{z^T B_k s_k s_k^T B_k z}{s_k^T B_k s_k} = a^T a - \frac{a^T b b^T a}{b^T b} =$$
$$\frac{a^T a b^T b - a^T b b^T a}{b^T b} = \frac{\|a\|^2 \|b\|^2 - \|a^T b\|^2}{\|b\|^2} \geq 0$$

And the proof is complete.

Lemma 4: The inverse formula of β-BFGS update is given by Eq. 7:

$$H_{k+1} = \left[I - \frac{s_k y_k^T}{y_k^T s_k} \right] H_k \left[I - \frac{y_k s_k^T}{y_k^T s_k} \right] + \frac{s_k s_k^T}{s_k^T B_k s_k y_k^T s_k} \quad (7)$$

Where: $H_{k+1} = B_{k+1}^{-1}$ and $H_k = B_k^{-1}$

Proof: Let we denote by $y_k = y$, $s_k = s$, $B_k = B$ and $H_k = H$, then by Sherman-Morreson Eq. 4, we have Eq. 8:

$$\left[B+\frac{s^TBs}{(y^Ts)^2}yy^T-\frac{Bss^TB}{s^TBs}\right]^{-1}=\left[B+\frac{s^TBs}{(y^Ts)^2}yy^T\right]^{-1}-$$

$$+\frac{\left[B+\frac{s^TBs}{(y^Ts)^2}yy^T\right]^{-1}\frac{Bss^TB}{s^TBs}\left[B+\frac{s^TBs}{(y^Ts)}yy^T\right]^{-1}}{1-s^TB\left[B+\frac{s^TBs}{(y^Ts)^2}yy^T\right]^{-1}\frac{Bs}{s^TBs}} \quad (8)$$

And again by using Sherman-Morreson formula, we have Eq. 9:

$$\left[B+\frac{s^TBs}{(y^Ts)^2}yy^T\right]^{-1}=H-\frac{s^TBsHyy^T}{(y^Ts)^2+s^TBsy^THy} \quad (9)$$

By substituting (Eq. 9 in 8), we get:

$$H_{k+1}=H+\frac{ss^T}{s^TBss^Ty}-\frac{sy^TH}{s^Ty}-\frac{Hys^T}{s^Ty}+\frac{ss^Ty^THy}{(s^Ty)^2}=$$

$$\left[I-\frac{s_ky_k^T}{y_k^Ts_k}\right]H_k\left[I-\frac{y_ks_k^T}{y_k^Ts_k}\right]+\frac{s_ks_k^T}{s_k^TB_ks_ky_k^Ts_k}$$

Algorithm 1 β-BFGS update:
1. Choose the starting point x^0 and the initial approximation $B_0 = I$, error, set $k = 0$
2. Compute $\nabla f(x^k)$
3. Solve the system $B_kp_k = -\nabla f(x^k)$ for p_k
4. Do line search to find $\alpha_k\in R$, $\ni f(x^k+\alpha_k p_k)<f(x^k)$
5. Set $x^{k+1} = x^k+\alpha_kp_k$
6. Set $s_k = x^{k+1}-x^k$, $y_k = \nabla f(x^{k+1})-\nabla f(x^k)$
7. Compute B_{k+1} from (6)
8. If $\|\nabla f(x^{k+1})\|<$ error then stop and x^{k+1} is the solution, else $k = k+1$ and go to 3

The convergence of β-BFGS update: In this study, we introduce the global convergence for β-BFGS update under exact line search. The following assumption be needed:

Assumption 1[3]: $f:R^n\to R$ is twice continuously differentiable on convex set $D\subseteq R^n$. f(x) is uniformly convex, i.e., there exist a positive constants m and M such that for all $x\in L(x) = \{x:f(x)\le f(x^0)\}$ which is convex, we have $m\|u\|^2\le u^T\nabla^2f(x)u\le M\|u\|^2$, $\forall u\in R^n$ and x^0 is the starting point.

Lemma 5[4]: Let f: $R^n\to R$ satisfy assumption 1, then $\|s_k\|/\|y_k\|\|y_k\|/\|s_k\|, s_k^Ty_k/\|s_k\|^2, s_k^Ty_k/\|y_k\|^2$ and $\|y_k\|^2/s_k^ty_k$ are bounded. As a result from the lemma 5, we have $s_k^Ty_k/s_k^TB_ky_k$ and $s_k^TB_ky_k/s_k^Ty_k$ are bounded.

Lemma 6[5]: Under exact line search $\Sigma\|s_k\|^2$ and $\Sigma\|y_k\|^2$ are convergent.

Theorem (convergence of β-BFGS update): Suppose that f(x) satisfy assumption 1, then under inexact line search the sequence $\{x^k\}$ generated by β-BFGS update convergence to the minimizer x* of f.

Proof:

$$B_{k+1}=B_k+\beta_k\frac{y_ky_k^T}{y_k^Ts_k}-\frac{B_ks_ks_k^TB_k}{s_k^TB_ks_k}$$

$$\text{Trace}(B_{k+1})=\text{Trace}(B_k)-\frac{\|B_ks_k\|^2}{s_k^TB_ks_k}+\beta_k\frac{\|y_k\|^2}{y_k^Ts_k} \quad (10)$$

$$\text{Define } m_k=\frac{y_k^Ts_k}{s_k^Ts_k} \text{ and } M_k=\frac{y_k^Ty_k}{y_k^Ts_k} \quad (11)$$

m_k and M_k are bounded, define:

$$\text{Cos}\theta_k=\frac{s_k^TB_ks_k}{\|s_k\|\|B_ks_k\|} \text{ And } q_k=\frac{s_k^TB_ks_k}{s_k^Ts_k} \quad (12)$$

where, θ_k is the angle between s_k and B_ks_k and define Eq. 13:

$$\Phi(B)=\text{Trace}(B)-\text{In}[\det(B)] \quad (13)$$

Clear that $\Phi(B)>0$ and:

$$\Phi(B_{k+1})=\text{Trace}(B_{k+1})-\text{In}[\det(B_{k+1})]=\text{Trace}(B_k)-$$

$$\frac{\|B_ks_k\|^2}{s_k^TB_ks_k}+B_k\frac{\|y_k\|^2}{s_k^Ty_k}-\text{In}\left(|B_k|\beta_k\frac{s_k^Ty_k}{s_k^TB_ks_k}\right)$$

$$\Phi(B_{k+1})=\Phi(B_k)+\beta_kM_k-\text{In}(\beta_k)\text{In}(m_k)+$$

$$\text{In}(\cos^2\theta_k)-1+\left(1-\frac{q_k}{\cos^2\theta_k}+\text{In}-\frac{q_k}{\cos^2\theta_k}\right)$$

Since, the last term is not positive and by Lemma 5 and 6 we have:

$$\Phi(B_{k+1})\le\Phi(B_k)+C+\text{In}(\cos^2\theta_k)$$

Where:

$$C=\beta_kM_k-\text{In}(\beta)-\text{In}(m_k)-1\in R$$

By summing the last inequality up to k (Eq. 14):

$$\sum_{j=0}^{k}\Phi\left(B_{j+1}\right)\le\sum_{j=0}^{k}\Phi\left(B_j\right)+\sum_{j=0}^{k}C+\sum_{j=0}^{k}\text{In}\left(\cos^2\theta_j\right)$$
$$\Phi\left(B_{k+1}\right)\le\Phi\left(B_0\right)+\left(k+1\right)c+\sum_{j=0}^{k}\text{In}\left(\cos^2\theta_j\right) \quad (14)$$

By Zoutendijk condition[2]:

$$\sum_{k=0}^{\infty}\text{In}\left(\cos^2\theta_k\right)\left\|\nabla f\left(x_k\right)\right\|<\infty$$

And hence:

$$\lim_{k\to\infty}\text{In}\left(\cos^2\theta_k\right)\left\|\nabla f\left(x_k\right)\right\|=0$$

Case 1: If θ_k is bounded away from $\pi/2$, $\exists\delta>0 \ni \cos\theta_k>\delta>0$ for k sufficiently large and then $\lim_{k\to\infty}\|\nabla f(x_k)\| = 0$ and $\{x_k\}\to x^*$ the proof is complete.

Case 2: If $\cos\theta_k\to 0$, then $\exists k_1>0 \ni \forall j>k_1$, we have $\text{In}\cos^2\theta_k<-2c$, therefore, for a sufficient large k:

$$0<\Phi\left(B_{k+1}\right)<\Phi\left(B_0\right)+C\left(k+1\right)+$$
$$\sum_{j=0}^{k_1}\text{In}\left(\cos^2\theta_j\right)-2C\left(k-k_1\right)<0$$

Which contradiction and the proof is complete.

RESULTS AND DISCUSSION

Numerical experiments: This study is devoted to numerical experiments. Our purpose was to check whether the β-BFGS update algorithm provide improvements on the corresponding standard BFGS update algorithm. The program are written in MATLAB with single precision. The test functions are commonly used unconstrained test problems with same starting point and a summary of which is given in Table 1. The test functions are chosen as follows[4]:

$$F\left(x\right)=\left(1-x_1\right)^2+\left(1-x_2\right)^2$$

Brown's badly scaled function:

$$F\left(x\right)=\left(x_1-10^6\right)^2+\left(x_2-2\times10^{-6}\right)^2+\left(x_1x_2-2\right)^2$$

$$F\left(x\right)=\left(1-x_1\right)^2+\left(x_2-x_1\right)^2$$

Rosenbroc'k Cliff function:

$$F\left(x\right)=10^{-4}\left(x_1-3\right)^2-\left(x_1-x_2\right)+e^{20\left(x_1-x_2\right)}$$

Generalized edeger function:

$$F\left(x\right)=\sum_{i=1}^{n/2}\left[\left(x_{2i-1}-2\right)^4+\left(x_{2i-1}-2\right)^2 * x_{2i}^2+\left(x_{2i}+1\right)^2\right]$$

Extended Himmelbla function:

$$F\left(x\right)=\sum_{i=1}^{n/2}\left(x_{2i-1}^2+x_{2i}-11\right)^2+\left(x_{2i-1}+x_{2i}^2-7\right)^2$$

Rosen Rock's function:

$$F\left(x\right)=\sum_{i=1}^{n/2}\left[100\left(x_i-x_i^3\right)^2+\left(1-x_i\right)^2\right]$$

Trigonometric function:

$$f\left(x\right)=\sum_{i=1}^{n}\left[n-\sum_{j=1}^{n}\cos x_j+i\left(1-\cos x_i\right)-\sin x_i+e^{xi}-1\right]^2$$

Extended Rosen rock function:

$$F\left(x\right)=\sum_{i=1}^{n/2}c\left(x_{2i}-x_{2i-1}^2\right)^2+\left(1-x_{2i-1}\right)^2 . C=100$$

Brown's badly scaled function:

$$F\left(x\right)=\left(x_1-10^6\right)^2+\left(x_2-2*10^{-6}\right)^2+\left(x_1x_2-2\right)^2$$

Watson function:

$$F\left(x\right)=\sum f_i^2\left(x\right)$$

Where:

$$f_i\left(x\right)=\sum_{j=2}^{3}\left(j-1\right)x_j t_j^{j-2}-\left(\sum_{j=1}^{3}x_j t_j^{j-1}\right)^2-1 \text{ and } t_j=\frac{i}{29}$$

Table 1, clear that β-BFGS update tends to the minimum of the function in all test problems and with all starting point, BFGS update also tends to the minimum of the function but if we compare the value of the objective function (Feval) between the two methods we can see that the β-BFGS update continue to the minimum of the objective function but the BFGS update stopped because the singularity of the Hessian matrix.

Table 1: Test problems

			BFGS		β-BFGS	
Problems	Starting points	Dim	FevalIter.	Values	FevalIter	Value
1	[-1; -1]	2	2.7900e-020	2	2.7900e-020	2
2	[0;...]	6	2.4954e-005	5	1.2415e-005	11
2	[1;-1;…]	4	2.3487e-005	14	4.9866e-009	28
3	[0; 0]	2	1.9471e-018	11	1.0951e-018	9
3	[-5; -5]	2	2.2838e-016	32	1.0216e-016	13
4	[-1; 0;-1; 0]	4	0.2011	16	0.1998	19
4	[0.5; …]	12	0.2004	3	0.2005	2
5	[1; 1…]	18	3.2184e-010	6	2.6306e-007	5
5	[1; 0]	2	2.6653e-010	5	9.8618e-009	5
6	[1;1]	2	9.4582e-011	6	1.6978e-011	7
6	[0; 0]	2	2.8607e-009	8	2.6041e-013	9
7	[-1; 1…]	8	7.6554e-011	10	1.1818-010	8
7	[0.2;…]	4	0.9901	3	0.9901	3
8	[-0.5;…]	12	7.1047-006	12	3.5263e-006	28
8	[0.5;…]	12	4.8273e-006	13	2.9004e-006	28
9	[-1.2;…]	3	2.03565e-006	17	1.5507e-009	24
9	[0; 0]	2	1.1462e-007	14	1.1276e-010	18
10	[0; 0]	2	2.4954e-005	5	1.2415e-005	11
10	[1;1]	2	2.3326e-005	17	8.4894e-006	15
11	[1; 1; …]	4	1.8054e-010	4	1.1649e-017	4
11	[0; 0; …]	10	1.1288e-009	4	1.7427e-013	5

CONCLUSION

In this study, The BFGS update is modified to preserve the determinant value of Hessian matrix at each iteration equal one and guarantee the strong positive definite property that the Hessian matrix never near singular at each iteration which make the computation continue until the objective function terminate at the minimum.

REFERENCES

01. Zhang, J. and C. Xu, 2001. Properties and numerical performance of Quasi-Newton methods with modified Quasi-Newton equations. J. Comput. Appl. Math., 137: 269-278.
02. Nocedal, J. and S.J. Wright, 2006. Numerical Optimization. 2nd Edn., Springer, Berlin, Germany, ISBN-13:978-0387-30303-1, Pages: 664.
03. Mahmood, S.S. and S.H. Shnywer, 2017. On modified DFP update for unconstrained optimization. Am. J. Appl. Math., 5: 19-30.
04. Mahmood, S.S. and H. Farqad, 2017. On extended symmetric rank one update for unconstrained optimization: University of Almustansiriah. J. Educ., 1: 206-220.
05. Sun, W. and Y.X. Yuan, 2006. Optimization Theory and Methods: Nonlinear Programming. Vol. 1, Springer, Berlin, Germany, ISBN-13: 978-0-387-24975-9, Pages: 687.