Modified BFGS Update (H-Version) Based on the Determinant Property of Inverse of Hessian Matrix for Unconstrained Optimization

Saad Shakir Mahmood*

Department of Mathematics, College of Education, University of Mustansiriyah
saadshakirmahmood@uomustansiriyah.edu.iq
*Corresponding author: saadshakirmahmood@uomustansiriyah.edu.iq
*ORCID ID: 0000-0001-5433-9710

Received 10/7/2019, Accepted 13/1/2020, Published 8/9/2020

This work is licensed under a Creative Commons Attribution 4.0 International License.

Abstract:

The study presents the modification of the Broyden-Flecher-Goldfarb-Shanno (BFGS) update (H-Version) based on the determinant property of inverse of Hessian matrix (second derivative of the objective function), via updating of the vector $s$ (the difference between the next solution and the current solution), such that the determinant of the next inverse of Hessian matrix is equal to the determinant of the current inverse of Hessian matrix at every iteration. Moreover, the sequence of inverse of Hessian matrix generated by the method would never approach a near-singular matrix, such that the program would never break before the minimum value of the objective function is obtained. Moreover, the new modification of BFGS update (H-version) preserves the symmetric property and the positive definite property without any condition.

Key words: BFGS update, Hessian matrix, Positive definite matrix, Unconstrained Optimization.

Introduction:

Optimization is the great importance in the various sciences, which have included different fields and aspects. It is an important part of the sciences of mathematics and physics, as well as their importance in engineering, especially mechanical engineering, electricity, management, economy, population growth, weather and other natural phenomena.

Several attempts were made to other quasi-Newton equations to get a better approximation of the inverse of Hessian matrix. (1) proposed a modified quasi-Newton equation which uses both gradient and function value information in order to yield a higher order accuracy for approximating the second curvature of an objective function. (2) considered a modified Broyden family which includes the BFGS update. (3) modified the BFGS update based on the new quasi-Newton equation, $B_{k+1}s_k = y_k + A_k s_k$, where $A_k$ is a matrix. (4) modified SR1 update based on Zhang-Xu’s condition and provided that the update preserves the symmetric and positive definite property and they also provided the global and superlinear convergence of the proposed method. (5) proposed the modified DFP update based on Zhang-Xu’s condition and provided the global and superlinear convergence of the proposed method. (6) proposed the modified BFGS method for solving the system of non-linear equations by using Taylor theorem, this proposed method is derivative-free, so the gradient information is not needed at each iteration. (7) proposed a modified quasi-Newton (secant) equation to get a more accurate approximation of the second curvature of the objective function by using Chain rule. Then,
based on this modified secant equation, they present a new BFGS method for solving unconstrained optimization problems. The proposed method makes use of both gradient and function values, and utilizes information from two most recent steps, while the usual secant relation uses only the latest step information. (8) proposed a modified Broyden update based on the positive definite property Hessian matrix (second derivative of the objective function), via updating the vector \( \gamma \) (the difference between the next gradient of the objective function and the current gradient of the objective function) and providing the global and superlinear convergence of the proposed method. In this work, the BFGS (H-version) update was modified to get a better approximation for the inversion of Hessian matrix such that the determinant value of inverse of Hessian matrix was controlled far from zero, and also guarantee the strong positive definiteness property of the inverse of Hessian matrix at every iteration.

For the unconstrained optimization problem,

\[
\min_{x \in \mathbb{R}^n} f(x) \to R
\]  

(1)

The most efficient quasi-Newton method is the BFGS method, the matrix \( H_k \) by the BFGS method can be updating as

\[
H_{k+1} = \left[ I - \frac{s_k y_k^T}{y_k^T s_k} \right] H_k \left[ I - \frac{y_k s_k^T}{y_k^T s_k} \right] + \frac{s_k s_k^T}{y_k^T s_k}, \quad \text{and the matrix } B_k \quad \text{by the BFGS update can be updating as}
\]

\[
B_{k+1} = B_k + \frac{y_k y_k^T}{y_k^T y_k} \frac{y_k s_k^T s_k}{s_k^T y_k} B_k \quad \text{where } B_k \quad \text{be the next approximation of inverse of Hessian matrix and } B_{k+1} \text{ be the next approximation of Hessian matrix.}
\]

The BFGS update is considered as a popular update to solve Eq. (1), in which the BFGS update (H-version) preserves the positive definite property, only if \( y_k^T s_k > 0 \) at every iteration. (1), proposed a modified BFGS update by multiplying the vector \( y_k \) with a real number, in this case the quasi-Newton equation must be extended to Zhang-Xu equation and it extended quasi-Newton equation \( B_{k+1} s_k = \beta_k y_k \), where \( s_k = x_{k+1} - x_k \), \( x_k \) is the current solution, \( x_{k+1} \) is the next solution, \( y_k = \nabla f(x_{k+1}) - \nabla f(x_k) \), \( \nabla f \) is the gradient of the objective function \( f \), and \( \beta_k \in \mathbb{R} \).

The problem is to solve equation (1) by producing a sequence of symmetric and positive definite (without condition) inverse of Hessian matrix, which never converges to a near-singular matrix that makes the numerical computation break before the minimizer is obtained due to the singularity of the inverse of Hessian matrix numerically. The best solution of this problem is fixing the value of the determinant of inverse of Hessian matrix to be considerably far from zero at every iteration, so that the program does not break prior to obtaining the minimizer.

**Modified BFGS Update (H-version):** Consider the BFGS (H-version) update (9),

\[
H_{k+1} = \left[ I - \frac{s_k y_k^T}{y_k^T s_k} \right] H_k \left[ I - \frac{y_k s_k^T}{y_k^T s_k} \right] + \frac{s_k s_k^T}{y_k^T s_k} \quad (2)
\]

By extended quasi-Newton equation

\[
y_k = \beta_k y_k, \quad \beta_k \in \mathbb{R}
\]

(3)

Based on equation (3), the Quasi-Newton equation becomes as follows:

\[
H_{k+1} \beta_k y_k = s_k \quad \text{or}
\]

\[
H_{k+1} y_k = \mu_k s_k \quad (4)
\]

Where \( \mu_k = \frac{1}{\beta_k} \)

The solution of equation. (4) is

\[
H_{k+1} = \left[ I - \frac{s_k y_k^T}{y_k^T s_k} \right] H_k \left[ I - \frac{y_k s_k^T}{y_k^T s_k} \right] + \frac{s_k s_k^T}{y_k^T s_k} \quad (5)
\]

This is called Modified BFGS update (H-version), to determine \( \mu_k \), the following lemmas are necessary:

**Lemma 1**

The inverse formula of equation (5) is given by

\[
B_{k+1} = B_k + \frac{y_k y_k^T}{\mu_k y_k^T y_k} - \frac{b_k s_k s_k^T y_k}{s_k^T B_k s_k} \quad (6)
\]

where \( B_{k+1} = H_{k+1}^{-1} \), and \( B_k = H_k^{-1} \).

Proof:

By Sherman – Morrison formula (10)

\[
\left[ B_k + \frac{y_k y_k^T}{\mu_k y_k^T y_k} - \frac{b_k s_k s_k^T y_k}{s_k^T B_k s_k} \right]^{-1} = \left[ B_k + \frac{y_k y_k^T}{\mu_k y_k^T y_k} \right]^{-1} + \frac{b_k s_k s_k^T y_k}{s_k^T B_k s_k} \left[ B_k + \frac{y_k y_k^T}{\mu_k y_k^T y_k} \right]^{-1} \left( \frac{y_k y_k^T}{\mu_k y_k^T y_k} - \frac{b_k s_k s_k^T y_k}{s_k^T B_k s_k} \right) \left[ B_k + \frac{y_k y_k^T}{\mu_k y_k^T y_k} \right]^{-1} \quad (7)
\]
and again by using Sherman – Morrison formula, equation (7) becomes
\[ \frac{y y^T}{\mu y} + H = H - \frac{H y y^T H}{\mu y y^T + y H y} \]  
(8)
By substituting equation (8) in equation (7) the result is
\[ H_{k+1} = H_k + \frac{\mu_k s_k s_k^T}{s_k^T y_k} s_k y_k^T H_k - \frac{H_k s_k y_k^T}{s_k^T y_k} H k + \frac{s_k s_k^T y_k^T H_k}{s_k^T y_k} \]
\[ + \frac{(s_k y_k^T)^2}{y_k^T s_k} . \]

Lemma 2.
For the modified BFGS update (H-version), the determinant of the next approximation of inverse of Hessian matrix is given by
\[ |H_{k+1}| = |H_k| \frac{s_k^T B_k s_k}{y_k^T s_k} \]
Proof:
\[ |B_{k+1}| = |B_k + \frac{y_k y_k^T}{\mu y y_k^T + y_k s_k s_k^T B_k}|, \] since the current Hessian matrix is symmetric and positive definite. Thus, there exists a triangular matrix \( L_k \in R^{n \times n} \) such that \( B_k = L_k L_k^T \). Consequently,
\[ |B_{k+1}| = |L_k L_k^T + \frac{y_k y_k^T}{\mu y y_k^T} y_k^T - \frac{L_k L_k^T s_k}{s_k^T B_k s_k} s_k^T L_k L_k^T| \]
\[ = |L_k^T| I + \frac{L_k^T y_k}{\mu y y_k^T} (L_k y_k^T) - \frac{L_k^T s_k}{s_k^T B_k s_k} (L_k s_k^T) \]
By rank two update determinant (10), the last equation becomes
\[ |B_{k+1}| = |B_k| \frac{1 + (L_k^T y_k)^T}{\mu y y_k^T} L_k^T y_k |1 \]
\[ - \frac{L_k^T s_k}{s_k^T B_k s_k} \]
\[ + |B_k| \frac{L_k^T s_k}{s_k^T B_k s_k} \]
and because \( |H_{k+1}| = \frac{1}{|B_{k+1}|} = \frac{\mu_k s_k^T B_k s_k}{s_k^T y_k} \) the proof is complete. If the condition \(|H_{k+1}| = |H_k| \) is set, \( \mu_k = \frac{s_k^T y_k}{s_k^T B_k s_k} \) and the Modified BFGS update (H-version) becomes
\[ H_{k+1} = \left[ I - \frac{s_k^T y_k}{y_k s_k} \right] H_k \left[ I - \frac{y_k s_k^T}{y_k s_k} + \frac{s_k^T s_k}{s_k^T B_k s_k} \right] \]
(9)

Lemma 3
Modifying BFGS update (H-version) produces a symmetric inverse of Hessian matrix if the current inverse of Hessian matrix is symmetric. Proof:
Since \( H_k^T = H_k \), \( (s_k s_k^T)^T = s_k s_k^T \), and \( (I - \frac{s_k^T y_k}{y_k s_k})^T = I - \frac{y_k s_k^T}{y_k s_k} \), then the proof is complete.

The next lemma shows that the modified BFGS update (H-version) preserves the positive definiteness of the inverse of Hessian matrix better than the BFGS does. More so, the condition \( y^T s > 0 \) is sufficient in the BFGS update, but is deleted in the modified BFGS update (H-version).

Lemma (4)
Given \( H_k \) symmetric and positive definite matrix then, the modified BFGS update (H-version) produces a positive definite inverse of Hessian matrix.
Proof:
For \( 0 \neq z \in R^n \), and by definition of positive definite property (10)
\[ z^T H_{k+1} z = z^T \left[ I - \frac{y_k s_k^T}{y_k s_k} \right] H_k \left[ I - \frac{y_k s_k^T}{y_k s_k} \right] z \]
\[ + z^T \frac{s_k^T s_k}{s_k^T B_k s_k} z \]
\[ = w_k^T H_k w_k + \|s_k^T s_k\|^2 \frac{1}{s_k^T B_k s_k} , \] where \( w_k = \left[ I - \frac{y_k s_k^T}{y_k s_k} \right] z \)
since \( H_k , B_k \) are positive definite, so the proof is complete.

Modified BFGS update (H-version)
Algorithm:
1. Choose the starting point \( x^0 \), and the initial approximation \( H_0 = I \), \( \epsilon > 0 \), set \( k=0 \).
2. Compute \( \nabla f(x^k) \)
3. Solve the system \( p_k = -H_k \nabla f(x^k) \) for \( p_k \).
4. Do line search to find $\alpha_k \in R, \exists f(x^k + \alpha_k p_k) < f(x^k)$.
5. Set $x^{k+1} = x^k + \alpha_k p_k$
6. Set $s_k = x^{k+1} - x^k$, $\gamma_k = \nabla f(x^{k+1}) - \nabla f(x^k)$.
7. Compute $H_{k+1}$ from equation (9)
8. If $\|\nabla f(x^{k+1})\| < \epsilon$, then stop and $x^{k+1}$ is the solution, else k=k+1 and go back to 3

The Convergence of modified BFGS update (H-version):

In this section, the global convergence for modified BFGS update under exact line search was introduced. The following assumption is needed:

Assumption 1 (5) 

a. $f: R^n \rightarrow R$ is twice continuously differentiable on convex set $D \subseteq R^n$.

b. $f(x)$ is uniformly convex, i.e., there exist a positive constants $m$ and $M$ such that for all $x \in L(x) = \{x: f(x) \leq f(x^0)\}$, which is convex, and $m\|u\|^2 \leq u^T \nabla^2 f(x) u \leq M\|u\|^2$, $\forall u \in R^n$ and $x^0$ is the starting point.

Lemma 5 (4)

Let $f: R^n \rightarrow R$ satisfy assumption then $\|s_k\| \|y_k\|$, $\|s_k^T y_k\|$, $\|s_k\|^2$, $\|y_k\|^2$, and $\|s_k^T y_k\|^2$ are bounded. As a result from Lemma 5… $s_k^T B_k s_k$, $y_k^T H_k y_k$, amd $\|B_k s_k\|$ are bounded.

Lemma 6 (10)

Under inexact line search $\sum_k \|s_k\|^2$ and $\sum_k \|y_k\|^2$ are convergent.

Theorem (convergence of the method)

Suppose $f(x)$ satisfies Assumption 1. Then, under exact line search the sequence $\{x^k\}$ generated by modified BFGS update (H-version) converges to the minimizer $x^*$ of $f$.

Proof:

Given $H_{k+1} = H_k + \frac{y_k^T H_k y_k s_k^T}{y_k^T y_k} s_k - \frac{s_k y_k^T H_k}{y_k^T y_k}$

Trace$(H_{k+1}) = \text{Trace}(H_k) + \text{Trace}(\frac{y_k^T H_k y_k s_k^T}{y_k^T y_k}) - \text{Trace}(\frac{s_k y_k^T H_k}{y_k^T y_k}) + \text{Trace}(\frac{y_k^T H_k y_k s_k^T}{y_k^T y_k})$ + \text{Trace}(\frac{s_k y_k^T H_k}{y_k^T y_k})$
Case II. If $\cos \theta_k \to 0$, then $\exists k_1 > 0 \forall j > k_1$
then, $\ln \cos^2 \theta_j < -2C$, therefore for a
sufficient large $k$
0 < $\psi(H_{k+1}) < \psi(H_0) + kC$

$$+ \sum_{j=0}^{k_1} \ln(\cos^2 \theta_j) - 2C(k - k_1) < 0$$

which contradiction, and the proof is complete

**Numerical Experiments:**

This section is devoted to the numerical
experiments aimed at assessing whether the
modified BFGS update (H-version) algorithm
provides improvements on the corresponding
Standard BFGS update algorithm. The program
was written in MATLAB with single precision.
The test functions were commonly used
unconstrained test problems with the same
starting point, a summary of which is given in
Table 1.
The test functions are chosen as follows: (5,10).
1- $f(x) = (1 - x_1)^2 + (1 - x_2)^2$, $min. f = 0$
2- $f(x) = (x_1 - 10^6)^2 + (x_2 - 2 \times 10^{-6})^2 +$
   $(x_1 x_2 - 2)^2$, $min. f = 0$
3- $f(x) = (1 - x_1)^2 + (x_2 - x_1)^2$, $min. f = 0$
4- $f(x) = \sum_{i=1}^{n/2} [(x_{2i-1} - 2)^2 + (x_{2i} x_{2i-1} - 2)^2 x_{2i}^2 + (x_{2i} + 1)^2]$, $min. f = 0$
5- Extended Himmelbla function

$$f(x) = \sum_{i=1}^{n/2} (x_{2i-1}^2 + x_{2i}^2 - 11)^2 +$$
$(x_{2i-1} + x_{2i}^2 - 7)^2$, $min. f = 0$
6- Rosenbrock function

$$f(x) = \sum_{i=1}^{n/2} [100(x_i - x_i^3)^2 + (1 - x_i)^2],$$
$min. f = 0$
7- Trigonometric function
It is apparent in Table 1 that the modified BFGS update (H-version) tended to the minimum of the objective function \( f(x) \) in all test problems. However, a comparison of the minimum value of \( f(x) \) between BFGS update (H-version) and modified BFGS update (H-version) revealed the modified BFGS update (H-version) to continue to the minimum of \( f(x) \) in all problems except problems 1, and 6b, while the BFGS update (H-version) had to stop because of the near-singularity of inverse of Hessian matrix. All the objective functions are chosen as quadratic programming problems with an exact minimum value equal to zero (\( \text{min. } f = 0 \)), for example problem 8 is a quadratic programming problem where the exact minimum value of the objective function is 0, for the first starting point \([-2;1;…-2;1]\) the minimum value of the objective function by using BFGS update is 0.00000203565 and the minimum value of the objective function by using modified BFGS update is 0.0000000015507 and clear that the error of the modified BFGS update is less than the error of the BFGS update, and this is true for all other problems, than previously can be conclude that the modified BFGS update is the best of the BFGS update.

**Conclusion:**

In this paper, the BFGS update (H-version) was modified to preserve the determinant value of the next inverse of Hessian matrix at each iteration equal to the determinant of the current inverse of Hessian matrix 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.

**Author's declaration:**

- Conflicts of Interest: None.
- I hereby confirm that all the Figures and Tables in the manuscript are mine. Besides, the Figures and images, which are not mine, have been given the permission for republication attached with the manuscript.
- The author has signed an animal welfare statement.
- Ethical Clearance: The project was approved by the local ethical committee in University of Mustansiriyah.
References:
1. Zhang J, Xu Ch. Properties and Numerical Performance of Modified quasi–Newton Methods Equations. J. Comp. Appl. Math. 2001; 137: 269-278.
2. Yabe H, Ogasawara H, Yoshino M. Local and superlinear convergence of quasi-Newton methods based on modified secant conditions. J. Comput. Appl. Math. 2007 Aug 1;205(1):617-32.
3. Guo Q, Liu J. Global convergence properties of two modified BFGS-type methods. J. Appl. Math.Comp. 2007; 23: 311–319.
4. Mahmood S S, Farqad H. On Extended Symmetric Rank One Update for Unconstrained Optimization. J. of Edu. 2017; 1: 206-220.
5. Mahmood S S., Shnywer Samira H. On Modified DFP Update for Unconstrained Optimization. American J. Appl. Math. 2017; 5(1) : 19-30.
6. Razieh D, Narges B, Mohammad M H. A new modified BFGS method for solving system of nonlinear equations. JIM. 2019; 22 : 75-89.
7. Razieh D, Narges B, Mohammad M H. A new modified BFGS method for unconstrained optimization problems. J. Comput. Appl. Math. 2018; 37: 5113-5125.
8. Mahmood SS, Muhanah N S. Symmetric and Positive Definite Broyden Update for Unconstrained Optimization. Baghdad Sci J. 2019; 16(3): 661-666.
9. Goldfarb D. A family of variable metric methods derived by variation means. J. of Math. Comp. 1970; 24: 23–26.
10. Wenyu S, Ya-Xiang Y. Optimization Theory and Method: Nonlinear Programming. Springer, USA. 2006.
الخلاصة:

الهدف من هذا البحث هو تحوير التحديث BFGS BFGS وتصحيح النسخة H H بالاعتماد على صفة محدد معكوس المصفوفة هيسين للمثلية غير المقيدة H H.

والذي بالاعتماد على صفات المحدد لمتعدد معكوس المصفوفة هيسين ( المشتقة الثانية لدالة الهدف ) والذي يتحديث المتاجة حالي ( الفرق بين الحل القادم الحل الائي ) بحيث تكون قيمة المحدد لمتعدد معكوس المصفوفة هيسين القادماً مساوي لقيمة المحدد لمتعدد معكوس المصفوفة هيسين الائي في كل تكرار ، وذلك لأن متتابعة التحديثات للمتعدد معكوس المصفوفة هيسين المتولدة من هذه الطريقة وبناءً على معكوس المصفوفة هيسين سوف تكون قيمة المحدد لها ثابت في كل تكرار ولا تقترب من قيمة المفرد المفرد ( المحدد = صفر ) مما يؤدي إلى أن البرنامج المستخدم للحاسات العددية سوف لن يتوقف بسبب اقتراب محدد المصفوفة المتولدة من الصفر وإن البرنامج المذكور سوف يتوقف فقط عندما نحصل على الحل الامثل لدالة الهدف. إضافة إلى ذلك فان التحوير الجديد سوف يحافظ على خاصيتي التناظريّة الموجبة للمتعدد المصفوفة المتموردة، ويكون بدون شروط وفي كل تكرار.

الكلمات المفتاحية : شرط كواسي – نيوتن، التحديث BFGS BFGS، شرط زانغ – كسي، المثلية غير المقيّدة، المصفوفة هيسين.