J. Edu. & Sci., Vol. (25), No. (3) 2012

A Modified Globally Convergent Self-Scaling BFGS Algorithm for Unconstrained Optimization

Prof. Dr. Abbas Y. AL-Bayati **Dr. Basim A. Hassan**
Department of Mathematics
College of Computers Sciences and Mathematics
University of Mosul

&

Sawsan S. Ismael
Department of Mathematics / College of Education
University of Mosul

Received **17 / 10 / 2010** **Accepted** **06 / 04 / 2011**

الملخص

في هذا البحث تم التقصي في خوارزمية مطورة ل BFGS ذاتي القياس و ذات التقارب الشامل في حل مسائل الأمثلية المحدبة وباستعمال خطوط بحث تامة و جميع التقريبات لمعكوس مصفوفة هسي موجبة التعريف . النتائج العددية أثبتت كفاءة الخوارزمية المقترحة مقارنة بخوارزمية− BFGS القياسية.

Abstract

In this paper, a modified globally convergent self-scaling BFGS algorithm for solving convex unconstrained optimization problems was investigated in which it employs exact line search strategy and the inverse Hessian matrix approximations were positive definite. Experimental results indicate that the new proposed algorithm was more efficient than the standard BFGS- algorithm.

1. Introduction.

Consider the unconstrained optimization problem

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

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

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

where $v_k = x_{k+1} - x_k$, $y_k = g_{k+1} - g_k$.[5]
The search direction is computed by:

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

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

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

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

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

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

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

The famous update B_k is the BFGS formula for which B_{k+1} is updated as :

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

In [6] Zhang and Xu proposed a new QN-condition defined by:

$$B_{k+1}v_k = y_k^* \qquad ....(8)$$

By using (8), The modified BFGS–update may be given by:

$$B_{k+1} = B_k - \frac{B_k v_k v_k^T B_k}{v_k^T B_k v_k} + \frac{y_k^* y_k^{*T}}{v_k^T y_k^*} \qquad ....(9)$$

where

$$y_k^* = y_k + \frac{6[f_k - f_{k+1}] + 3(g_{k+1} + g_k)^T v_k}{v_k^T y_k} y_k \qquad ....(10)$$

Dai [3] proved that updates (9)-(10) may fail for convergence, especially for the non-convex functions. To overcome this drawback, let us consider the self-scaling BFGS update due to Al-Bayati [1].

1.1 Al-Bayati Self-Scaling BFGS-Method [1].

The standard BFGS update can be separated into two components, $B^{(1)}$ *and* $B^{(2)}$ so that:

$$B_{BFSG} = B^{(1)} + B^{(2)} \qquad ....(11)$$

where $B^{(1)} = B_k - \dfrac{B_k v_k v_k^T B_k}{v_k^T B_k v_k}$, $B^{(2)} = \dfrac{y_k y_k^T}{v_k^T y_k}$

AL-Bayati's [1] modifications for the BFGS formula can then be written as:

$$B_{AL-Bayati} = B^{(1)} + \rho_k B^{(2)} , \quad \rho_k = \frac{v_k^T B_k v_k}{y_k^T v_k} \qquad ....(12)$$

which will satisfy

$$B_{k+1} v_k = \rho\, y_k \qquad ....(13)$$

This relaxation of the QN-condition is of particular interest in deriving different VM-algorithms for non–quadratic objective functions.

1.2 A Modified Self-Scaling BFGS-Algorithm.

In this section we will deal with an algorithm which generates the sequence of B_{k+1} matrices which converge to the Hessian matrix G and satisfies the following modified QN-condition:

$$B_{k+1} v_k = \rho^* y_k^* \qquad ....(14)$$

According to the above strategy, $B^{(1)}$ *and* $B^{(2)}$ of Al-Bayati's update can be represented as:

$$B_{\text{MBFGS}} = B^{(1)} + B^{(2)} \qquad ....(15)$$

where $B^{(1)} = B_k - \dfrac{B_k v_k v_k^T B_k}{v_k^T B_k v_k}$, $B^{(2)} = \dfrac{y_k^* y_k^{*T}}{v_k^T y_k^*}$

By using (12), the modified AL-Bayati's BFGS- update may be written as:

$$B_{k+1}^{new} = B_k - \frac{B_k v_k v_k^T B_k}{v_k^T B_k v_k} + \rho^*{}_k \frac{y_k^* y_k^{*T}}{v_k^T y_k^*} \qquad ....(16)$$

where $\rho^*{}_k = \dfrac{v_k^T B_k v_k}{y_k^{*T} v_k}$, y_k^* is a vector defined in (10).

1.3 Outline of the Modified BFGS-Algorithm.

The outline of the modified BFGS algorithm is as follows:

Step 0 : Choose an initial point $x_1 \in R^n$ and an initial positive definite

matrix $B_1 = I_{n*n}$, $\varepsilon = 1*10^{-4}$, set $k = 1$.

Step 1 : If $\|g_k\| \le \varepsilon$, stop .

Step 2 : Solve $d_k = -H_k g_k$ to obtain a search direction d_k.

Step 3 : Find α_k by (WP) step-size rules (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 (16) with (10).

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

2. The Global Convergence Property for the Modified BFGS-Algorithm.

The important property of the line search method is the global convergence property defined by the relation :

$$\lim_{k \Rightarrow \infty} \inf \|g_k\| = 0 \qquad ....(17a)$$

To prove this property, it is equivalent to prove that the modified BFGS-updating formula (16) generates identical conjugate gradient search directions provided that the function is quadratic and exact line searches are used . Let us consider the following property :

Property (2.1).

Let f be given by

$$f(x) = \frac{1}{2} x^T G x + b^T x \qquad ....(17b)$$

where G is symmetric positive definite. Choose an initial approximation $H_1 = H$ where H is any symmetric positive definite matrix. Obtain H_{k+1} from H where $d_k = -Hg_k$ is the search direction and assuming exact lime searches then:

$$H_{k+1} g^* = H g^* \qquad ....(18)$$

Proof : see [1] .

It is well know that the inverse equivalent BFGS-update formula of the modified updating formula (16) can be written also as :

$$H_{k+1} = H_k - \frac{H_k y_k^* v_k^T + v_k y_k^{*T} H_k}{y_k^{*T} H_k y_k^*} + \frac{v_k v_k^T}{v_k^T y_k^*} \left[\rho^*{}_k + \frac{y_k^{*T} H_k y_k^*}{v_k^T y_k^*} \right] \qquad .....(19)$$

where $\rho^*{}_k = \dfrac{y_k^{*T} H_k y_k^*}{v_k^T y_k^*}$, y_k^* is a vector defined in (10)

which is the dual of the modified updating formula (16) in sense that $H_k \Rightarrow B_k$, $v_k \Rightarrow y_k$ and $\rho_k \Rightarrow \rho_k^{-1}$.

The following Lemma (2.1) which was given in [6] page 273, yields:

Lemma (2.1).

Suppose that if $(\alpha_k, x_k, y_k, d_k)$ generated by the BFGS update and that G is continuous at x^*. Then we have

$$\lim_{k \Rightarrow \infty} \left\| \frac{6[f_k - f_{k+1}] + 3(g_{k+1} + g_k)^T v_k}{v_k^T y_k} \right\| = 0$$

New Theorem (2.1).

Assume that $f(x)$ be a quadratic function defined in (17b) and that the line searches are exact: let H be any symmetric positive definite matrix for the modified self scaling BFGS-updating formula defined by:

$$H_{k+1} = H_k - \frac{H_k y_k^* v_k^T + v_k y_k^{*T} H_k}{y_k^{*T} H_k y_k^*} + \frac{v_k v_k^T}{v_k^T y_k^*}\left[\rho^*{}_k + \frac{y_k^{*T} H_k y_k^*}{v_k^T y_k^*}\right] \qquad .....(20a)$$

where $\rho^*{}_k = \dfrac{y_k^{*T} H_k y_k^*}{v_k^T y_k^*}$ $\qquad (20b)$

Then the search direction

$$d_{MBFGS} = -H_{k+1} g^* \qquad ....(21)$$

is identical to the Hestenes and Stiefel conjugate gradient direction d_{HSCG} defined by:

$$d_{HSCG} = -g^* + \frac{y^{*T} g^*}{y^{*T} d} d \quad for\ k \geq 1 \qquad ....(22)$$

Proof :

$$H_{k+1} = H_k - \frac{H_k y_k^* v_k^T}{v^T y_k^*} - \frac{v_k y_k^{*T} H_k}{v^T y_k^*} + \frac{v_k v_k^T}{v_k^T y_k^*}\left[\rho^*{}_k + \frac{y_k^{*T} H_k y_k^*}{v_k^T y_k^*}\right]$$

Now

$$d_{MBFGS} = -H_k g_k^* + \frac{H_k y_k^* v_k^T g_k^*}{v_k^T y_k^*} + \frac{y_k^{*T} H_k g_k^*}{v_k^T y_k^*} v_k - 2\frac{y_k^{*T} H_k y_k^* v_k^T g_k^*}{(v_k^T y_k^*)^2} v \qquad .....(23)$$

$$= -H_k g_k^* + \frac{y_k^{*T} H_k g_k^*}{v_k^T y_k^*} v_k \qquad ....(24)$$

using the property $v_k^T g_k^* = 0$ quoted earlier which holds for exact line searches. The vector g_k^* can be substituted for $H_k g_k^*$ by using property (2.1). Therefore :

$$d_{MBFGS} = -g_k^* + \frac{y_k^{*T} g_k^*}{v_{k\ new}^T y_k^*} v_k \qquad ....(25)$$

New by using Lemma (2.1) we have

$$d_{MBFGS} = -g_k^* + \frac{y_k^T g_k^*}{v_k^T \; y_k} v_k \qquad \text{..........(26)}$$

We also know that d_{MBFGS} and d_{HSCG} are identical [4] with exact line searches. Hence equation (25) becomes

$$d_{k+1(MBFGS)} = -g_k^* + \frac{y_k^T H_k g_k^*}{d_{k(HSCG)}^T y_k} d_{k(HSCG)} \qquad \text{..........(27)}$$

This completes the proof of the global convergence property of the modified BFGS-algorithm.

3. Numerical Results.

This section was devoted to numerical experiments. Our purpose was to check whether the modified self scaling AL-Bayati's BFGS-algorithm provide improvements on the corresponding standard BFGS-algorithm. The programs were written in Fortran 90. The test functions were commonly used for unconstrained test problems with standard starting points [6] and a summary of the results of these test functions was given in **Table (3.1)**. The same line search was employed in each algorithm, this was the cubic interpolation technique which is satisfy conditions (5) and (6) for convex optimization with $\delta_1 = 0.0001$ and $\delta_2 = 0.1$. The initial inverse approximation was $H_0 = I$. The stopping criterion was taken to be $\|g_{k+1}\| \prec 1*10^{-4}$. We tabulate for comparison of these algorithms, the number of function evaluations (NOF) and the number of iterations (NOI) .

Table (3.1) gives the comparison between the standard BFGS-algorithm and the modified self-scaling BFGS-algorithm for convex optimization, this table indicates that the modified algorithm saves 11% NOI and 14% NOF, overall against the standard BFGS-algorithm, especially for our selected test problems.

Table (3.1)

Test-functions	**N**	**Modified algorithm NOF (NOI)**	**BFGS algorithm NOF (NOI)**
Non-diagonal	10	99 (40)	108 (40)
	40	121 (53)	118 (50)
	100	116 (51)	135 (56)
	400	119 (52)	144 (59)
	1000	127 (54)	150 (62)
Wolfe	10	31 (15)	32 (15)
	40	85 (42)	87 (43)
	100	113 (56)	125 (62)

	400	125 (62)	139 (69)
	1000	143 (71)	168 (83)
Cubic	10	61 (24)	70 (26)
	40	85 (36)	83 (35)
	100	79 (33)	98 (41)
	400	80 (33)	109 (48)
	1000	101 (45)	112 (48)
Powell	10	82 (26)	82 (23)
	40	100 (38)	132 (46)
	100	98 (36)	140 (51)
	400	113 (40)	157 (54)
	1000	103 (38)	120 (39)
Cantrell	10	41 (10)	52 (11)
	40	48 (11)	52 (11)
	100	54 (12)	56 (12)
	400	57 (13)	71 (14)
	1000	71 (17)	71 (14)
Miele	10	56 (19)	65 (21)
	40	80 (26)	92 (30)
	100	76 (26)	93 (30)
	400	83 (28)	98 (31)
	1000	84 (28)	97 (32)
Total		2631 (1035)	3056 (1156)

The Percentage Performance of the improvements is given by:

Tools	BFGS	Modified algorithm
NOI	100 %	89.53 %
NOF	100 %	86.09 %

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

6. Cantrell *function*:

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

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

References.

1) Al-Bayati, A., (1991), A new family of self-scaling variable metric algorithms for unconstrained optimization, Journal of Educ. and Sci., Iraq, Vol.12, pp.25-54.

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

3) Dai Y., (2003), Convergence properties of the BFGS algorithm, SIAM J. Optimization, Vol. 13, pp. 693-701.

4) Nazareth, L., (1979), A relationship between BFGS and conjugate gradient algorithm and its implementation for new algorithms, SIAM Journal Numerical Analysis Vol.61, pp.409-436.

5) Zhang, H., Hager, W., (2004), A non-monotone line search technique and its application to unconstrained optimization. SIAM J. Optimization, Vol.14, pp.1043- 1056.

6) Zhang, J., Xu, C., (2001), Properties and numerical performance of quasi – Newton Methods with modified quasi – Newton equations Com. and Applied Mathematics, Vol. 137, pp.269-278.