Automatic Super-linear Self-Scaling VM-algorithm

Abbas Y. Al-Bayati Maha S. Al-Salih
profabbasalbayati@yahoo.com

College of Computers Sciences & Mathematics
University of Mosul/Iraq

Received on: 12/11/2003 Accepted on: 09/06/2004

Abstract

In this paper, a new self-scaling VM-algorithm for unconstrained non-linear optimization is investigated. Some theoretical and experimental results are given on the scaling technique, which guarantee the Super-linear of the new proposed algorithm.

Keywords: unconstrained non-linear optimization, self-scaling VM-algorithm.

الملخص

في هذا البحث تم التطرق إلى خوارزمية جديدة للمتغير ذاتي القياس الذي يضمن الفائدة من خلال تقنية خاصة بالقياس الذاتي. كما تم دراسة الجوانب النظرية والعملية للمسألة والتي تؤكد السرعة فوق الخطية للخوارزمية الجديدة المقترحة.

الكلمات المفتاحية: الأمثلية اللاحقة، المتغير ذاتي القياس.
1. Introduction

Conjugate Gradient (CG) methods were first used to solve the general unconstrained problem by Fletcher and Reeves [14]. Their algorithm (or simple variants) is still frequently used, especially for problems with a large number of variables since they require only a few vectors of length $n$ to be stored.

Given a symmetric positive definite matrix $G$, the finite set of non-null vectors $\{d_1, d_2, \ldots, d_k\}$ are said to form a conjugate set if

$$d_i^T G d_j = 0 \quad \text{for all } i \neq j$$

An important class of quasi-Newton methods for solving the unconstrained optimization problem, [13]

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

was proposed by [7]. It consists $f$ iterations of the form

$$x_{k+1} = x_k + \lambda_k d_k \quad k \geq 1,$$

where

$$d_k = -B_k^{-1} g_k$$

Here $\lambda_k$ is a step length parameter satisfies the Wolfe conditions with exact line search strategy, i.e.

$$f(x_k + \lambda_k d_k) \leq f(x_k) + \alpha \lambda_k g_k^T d_k$$

$$g(x_k + \lambda_k d_k)^T d_k \geq \beta g_k^T d_k$$

for $0 < \alpha < \frac{1}{2}$ and $\alpha < \beta < 1$, and $g_k$ denotes the gradient of $f$ at $x_k$.

The Hessian approximation $B_k$ is updated by means of the formula

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k} + \phi_k (s_k^T B_k s_k) v_k v_k^T,$$

where $\phi$ is a scalar, $y_k = g_{k+1} - g_k$, $s_k = x_{k+1} - x_k$ and

$$v_k = \left[ \begin{array}{c} y_k - B_k s_k \\ y_k^T s_k \\ s_k^T B_k s_k \end{array} \right]$$

The choice of the parameter $\phi_k$ is important, since it can greatly affect the performance of the methods. The BFGS method corresponds to $\phi_k = 0$. 

Variable metric (VM) methods were originally proposed by Davidon [11]. Subsequently, many authors have extended the theory and practice,[12] for a survey. The search direction in a VM-method is the solution of the system of equations:

$$d_k = -H_k g_k$$

where the matrix $H_k$ is an approximation to $G_k^{-1}$, the inverse Hessian of the function $f(x)$. and:

$$v_k^T G_k v_k \cong v_k^T y_k$$

This relationship is exact if the non-linear function $f$ is exactly equal $q$. The new approximation for the inverse Hessian $H_{k+1}$ is chosen to ensure that

$$H_{k+1} y_k = \xi_k v_k$$

where $\xi_k$ is a scalar; generally for the quasi-Newton (QN) method $\xi_k = 1$ and hence (3.15) reduces to

$$H_{k+1} y_k = v_k \quad \text{(called the QN-condition)}$$

And

$$H_{k+1} = H_k + C_k$$

The matrix $C_k$ is therefore, the update to $H_k$.

For the next iteration $B_{k+1}$ is updated by Al-Bayati’s VM-update, i.e.

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

(8)

See [1] for more details and properties of this algorithm.

**Algorithm 1.1:** [6]

1. For a starting point $x_1$ and non singular matrix $V_1$; set $k = 1$.
2. Terminate if $\|g_{k+1}\| < \varepsilon$, $\varepsilon$ is a small positive real number.
3. Compute
\[ d_k = -V_k^T V_k^{-1} g_k \]
\[ x_{k+1} = x_k + \lambda_k d_k \]
\( \lambda_k \) is computed by exact line search.

(4) Update \( V_k \) using Al-Bayati, 1991 VM-update.
\[ W_k = V_k - \frac{V_k s_k s_k^T V_k}{s_k^T V_k s_k} + \frac{s_k^T V_k y_k}{(y_k^T s_k)^2} \cdot y_k y_k^T \]

(5) Compute the scaling parameter \( \sigma_k \geq 0 \) and \( \mu_k > 0 \) such that \( \sigma_k \leq \mu_k \). If \( w_i \) represents the column of \( W_k \) put \( C_k = \text{diag} [c_1, c_2, \ldots, c_n] \) where
\[
\begin{cases}
\sigma_k & \text{if } \|w_i\| < \sigma_k \\
\frac{\mu_k}{\|w_i\|} & \text{if } \|w_i\| > \mu_k \\
\zeta_k & \text{otherwise}
\end{cases}
\]
\[
\zeta_k = \frac{y_k^T V_k y_k}{y_k^T s_k}
\]
Where \( \zeta_k \).

(6) Set \( V_{k+1} = W_k C_k \)

(7) set \( k = k +1 \) and go to step (1)

**Note that:**
1- In the above algorithm
\[
\begin{align*}
B_1 &= V_1 V_1^T \\
B_k &= V_k V_k^T \\
&= W_{k-1} C_{k-1}^2 W_{k-1}^T \quad k > 1
\end{align*}
\]
and the update is performed directly on \( V_k \).

2. Basic Results for Super Linear Convergence

First we define the following quantities to be used in this section:
\[
\overline{B}_k = G^2_x B_k G^2_x, \quad \overline{W}_k = G^2_x W_k
\]
\[
\overline{s}_k = G^2_x s_k, \quad \overline{y}_k = G^2_x y_k
\]
\[ \overline{M}_k = \frac{\bar{y}_k^T \bar{s}_k}{\bar{y}_k^T \bar{s}_k}, \quad \bar{m}_k = \frac{\bar{y}_k^T \bar{s}_k}{\bar{s}_k^T \bar{s}_k} \]  
\[ \bar{q}_k = \frac{\bar{s}_k^T \bar{B}_k \bar{s}_k}{\bar{s}_k^T \bar{s}_k}, \quad \cos \bar{\theta}_k = \frac{\bar{s}_k^T \bar{B}_k \bar{s}_k}{\| \bar{s}_k \| \| \bar{B}_k \bar{s}_k \|} \]  

where \( \bar{G}^* \) is the Hessian of \( f \) at the minimizer \( x^* \).

The limiting behavior of \( \bar{q}_k \) and \( \cos \bar{\theta}_k \) is enough to characterize the asymptotic rate of convergence of a sequence of iterates \( \{x_k\} \) generated by a quasi-Newton algorithm. Their result which can be seen as a restatement of the,[12] characterization, is reproduced in the following lemma.

**Lemma (2.1):**

Suppose that the sequence of iterates \( \{x_k\} \) is generated by algorithm (1.1) using some positive definite sequence \( \{B_k\} \), and that \( \lambda_k = 1 \) whenever this value satisfies Wolfe conditions (4)-(5). If \( x_k \to x^* \) then the following two conditions are equivalent:

(i) The steplength \( \lambda_k = 1 \) satisfies conditions (4)-(5) for all large \( k \) and the rate of convergence is superlinear.

(ii) \( \lim_{k \to \infty} \cos \bar{\theta}_k = \lim_{k \to \infty} \bar{q}_k = 1 \)  

**Proof:** Proof of this lemma can be found in [9]. The next theorem specifies conditions on the scaling parameters \( \sigma_k \) and \( \eta_k \) that allow \( \bar{q}_k \) and \( \cos \bar{\theta}_k \), produced by Algorithm 1.1, to exhibit the desirable limiting behavior of Lemma 2.1. Such conditions involve the following quantities:

\[ \gamma_k = \sum_{i=1}^{n} \left( \ln \left( \frac{1}{G_i \cdot \bar{w}_i} \right)^2 \right) - \left( \ln \sigma_k^2 \right) - \left( \ln \eta_k^2 \right) \]  
\[ \mu_k = \sum_{i=1}^{n} \left( \ln \left( \frac{1}{G_i \cdot \bar{w}_i} \right)^2 \right) - \left( \ln \sigma_k^2 \right) - \left( \ln \eta_k^2 \right) \]  

and

\[ \lim_{k \to \infty} \cos \bar{\theta}_k = \lim_{k \to \infty} \bar{q}_k = 1 \]  

(15)

(16) and

(17)
\[ \phi_k = \sum_{i=1}^{\infty} \left( \ln \left\| \frac{1}{G_{i,2}} w_i \right\|^2 - \ln \left\| \frac{1}{G_{i,2}} w_i \right\|^2 \right) - \left( \ln \zeta_{i,k}^2 \left\| w_i \right\|^2 - \zeta_{i,k}^2 \left\| w_i \right\|^2 \right) \] ... (17a)

and whether they sum finitely or not. Note that \( \gamma_k \) and \( \mu_k \) need not be positive. Recall that the sets \( I_k \) and \( J_k \) defined by:

\[ I_k = \{ i \in [1, n] : \| w_i \| < \sigma_k \} \]

and

\[ J_k = \{ i \in [1, n] : \| w_i \| > \mu_k \} \]

contain the indices of the columns that are scaled down at iteration \( k \).

We are now ready to state the theorem.

**Theorem 2.1:**

For the quadratic function \( f, x_1, B_1, \sigma_k \) and \( \eta_k \) satisfy the assumptions in theorem 1.1. In addition, assume that \( G \) is Lipschitz continuous at \( x^* \). Let \( \{ x_k \} \rightarrow x^* \) be generated by Algorithm 1.1; then if

\[ \sum_{k=1}^{\infty} \gamma_k < \infty \] (18)

\[ \sum_{k=1}^{\infty} \mu_k < \infty \] (19)

\[ \sum_{k=1}^{\infty} \phi_k < \infty \] (19a)

the iterates converge superlinearly (for the case of non-quadratic functions, see [2] and [3].

**Proof:** Let \( \psi (B) = \text{tr} (B) - \ln (\det (B)) \) then,

\[ \text{Tr} (ACA^T) = \text{tr} (AA^T) + \text{tr} [(C - I) A^T A] \] and from (11), we have

\[ \psi (B_{k+1}) = \text{tr} (G_{k,2}^{-2} W_k C_k^2 W_k^T G_{k,2}^{-2}) - \ln \det (G_{k,2}^{-2} W_k C_k^2 W_k^T G_{k,2}^{-2}) \]

\[ = \text{tr} (\tilde{W}_k C_k^2 \tilde{W}_k^T) - \ln \det (\tilde{W}_k \tilde{W}_k^T) - \ln \det (C_k^2) \]

\[ = \psi (\tilde{W}_k \tilde{W}_k^T) + \sum_{i=1}^{\infty} (c_i^2 - 1) \| G_{i,2} w_i \|^2 - \ln c_i^2 \]

Then by the definition (9) of \( c_i \),
$$\psi(\tilde{B}_{k+1}) = \psi(\tilde{W}_k W_k^T) + \sum_{i \in \mathbb{J}_k} \left[ \left( \frac{\sigma_k^2}{\|W_i\|^2} - 1 \right) \|G_{s_k}^{-2} W_i\|^2 \right]$$

$$+ \sum_{i \in \mathbb{J}_k} \left[ \left( \frac{\eta_k^2}{\|W_i\|^2} - 1 \right) \|G_{\eta_k}^{-2} W_i\|^2 \right]$$

$$+ \sum_{i \in \mathbb{J}_k} \left[ \left( \frac{\zeta_k^2}{\|W_i\|^2} - 1 \right) \|G_{\zeta_k}^{-2} W_i\|^2 \right]$$

$$= \psi(\tilde{W}_k W_k^T) + \sum_{i \in \mathbb{J}_k} \left[ \sigma_k^2 \|G_{s_k}^{-2} W_i\|^2 - \|G_{s_k}^{-2} W_i\|^2 \right]$$

$$- \ln \sigma_k^2 \|G_{s_k}^{-2} W_i\|^2 + \ln \|G_{s_k}^{-2} W_i\|^2$$

$$+ \sum_{i \in \mathbb{J}_k} \left[ \eta_k^2 \|G_{\eta_k}^{-2} W_i\|^2 - \|G_{\eta_k}^{-2} W_i\|^2 \right]$$

$$- \ln \eta_k^2 \|G_{\eta_k}^{-2} W_i\|^2 + \ln \|G_{\eta_k}^{-2} W_i\|^2$$

$$+ \sum_{i \in \mathbb{J}_k} \left[ \zeta_k^2 \|G_{\zeta_k}^{-2} W_i\|^2 - \|G_{\zeta_k}^{-2} W_i\|^2 \right]$$

$$- \ln \zeta_k^2 \|G_{\zeta_k}^{-2} W_i\|^2 + \ln \|G_{\zeta_k}^{-2} W_i\|^2$$

$$= \psi(\tilde{W}_k W_k^T) + \gamma_k + \mu_k + \phi_k \quad (20)$$

Since $\tilde{W}_k W_k^T$ is the matrix obtained by updating $B_k$ using the, [1] formula, which is invariant under the transformation (11) – (14), we have:

$$\psi(\tilde{W}_k W_k^T) = \psi(\tilde{B}_k) + (\tilde{M}_k - \ln \tilde{m}_k - 1)$$

$$+ (1 - \frac{\tilde{q}_k}{\cos^2 \tilde{\theta}_k} + \ln \frac{\tilde{q}_k}{\cos^2 \tilde{\theta}_k}) + \ln \cos^2 \tilde{\theta}_k \quad (21)$$

45
Therefore, using (21) in (20), we have:
\[
\psi (\tilde{B}_{k+1}) = \psi (\tilde{B}_k) + (\tilde{M}_k - \ln \tilde{m}_k - 1) + \left( 1 - \frac{\tilde{q}_k}{\cos^2 \tilde{\theta}_k} + \ln \frac{\tilde{q}_k}{\cos^2 \tilde{\theta}_k} \right) \\
+ \ln \cos^2 \tilde{\theta}_k + \gamma_k + \mu_k + \phi_j \\
= \psi (\tilde{B}_k) + \sum_{j=1}^{k} (\tilde{M}_j - \ln \tilde{m}_j - 1) + \sum_{j=1}^{k} \left( 1 - \frac{\tilde{q}_j}{\cos^2 \tilde{\theta}_k} + \ln \frac{\tilde{q}_j}{\cos^2 \tilde{\theta}_k} \right) \\
+ \ln \cos^2 \tilde{\theta}_k + \sum_{j=1}^{k} \gamma_j + \sum_{j=1}^{k} \mu_j + \sum_{j=1}^{k} \phi_j \\
\text{(22)}
\]

Now by the following theorem due to [10].

**Theorem 1.1:**

Let \( x_1 \) be a starting point for which \( f \) satisfies eq.(12) and let \( B_1 \) be a positive definite starting Hessian approximation. Let \( \{x_k\} \) be generated by the new proposed algorithm with \( \sigma_k \) and \( \mu_k \) satisfying eq.(18) and for any \( \rho \in (0, 1) \) \( \exists \) a constant \( \beta_1 \) so for any \( k > 1 \) the relation \( \cos \theta_j \geq \beta_1 \) holds for at least \( \left\lceil P_k \right\rceil \) values of \( j \in [1, k] \).

We know that the iterates converge to \( x^* \) r-linearly. Using this and the Lipschitz continuity of \( G \) at \( x^* \), it is not difficult to show, see [9] that:

\[
\sum_{j=1}^{k} (\tilde{M}_j - \ln \tilde{m}_j - 1) < \infty \\
\text{(23)}
\]

Moreover, the hypothesis of the theorem guarantees that the last two summations in (22) are bounded above. Therefore, in order for \( \psi (\tilde{B}_{k+1}) \) to remain positive as \( k \to \infty \), the sum of the nonpositive terms in the square brackets must also be bounded. This can only be true if:

\[
\lim_{k \to \infty} \left( 1 - \frac{\tilde{q}_k}{\cos^2 \tilde{\theta}_k} + \ln \frac{\tilde{q}_k}{\cos^2 \tilde{\theta}_k} \right) = \lim_{k \to \infty} \ln \cos^2 \tilde{\theta}_k = 0
\]

Which implies that both \( \tilde{q}_k \) and \( \cos^2 \tilde{\theta}_k \to 1 \). Hence, superlinear convergence follows from Lemma (2.1) #.

Now we describe a specific and modified implementation of algorithm 1.1 and make use of the theory developed so far to show that it is superlinearly convergent for strictly convex objective functions.

**New Algorithm:**

Step (0) Choose \( x_1 \) and a nonsingular and lower matrix \( V_1 \); set \( k = 1 \).

Step (1) Terminate if a stopping criterion is satisfied.
Step (2) Find an orthogonal matrix $Q_k$ such that $L_k = V_k Q_k$ is a lower triangular.

Compute:
\[
d_k = -L_k^T L_k^{-1} g_k,
\]
\[
x_{k+1} = x_k + \lambda_k d_k
\]

Where $\lambda_k$ is a steplength that satisfies the Wolfe conditions (The stepsize $\lambda_k = 1$ is always tried first and is accepted if admissible).

Compute:
\[
s_k = x_{k+1} - x_k
\]
\[
y_k = g_{k+1} - g_k
\]

Step (3) Perform the following steps to update $L_k$ to $W_k$ so that $W_k W_k^T$ become Al-Bayati update of $L_k L_k^T$ as defined in (8):

(3.1) Compute $r_k = L_k^T s_k$

(3.2) Find an orthogonal and lower matrix $\Omega_k$ such that
\[
\Omega_k e_1 = r_k / \| r_k \|.
\]

(3.3) Construct $W_k = \{ w_1^k, w_2^k, \ldots, w_n^k \}$, where $w_i^k$ is given by
\[
w_i^k = \begin{cases} y_k / \sqrt{y_k^T s_k}, & i = 1 \\ L_k \Omega_k e_i, & i = 2, 3, \ldots, n \end{cases}
\]

Step (4) Compute the scaling parameters:

If $k = 1,$
\[
\sigma_i^2 = \eta_i^2 = \frac{y_i^T y_i}{s_i^2} = \gamma_i^2
\]

Otherwise,
\[
\sigma_k^2 = \frac{1}{n} \left( (n - |I_{k-1}|) \sigma_k^{2} + \sum_{i \in I_{k-1}} \| W_i^{k-1} \|^{2} \right)
\]

where $I_{k-1} = \{ i \in [1, n] : \| W_i^{k-1} \| < \sigma_{k-1} \}$,

And
\[
\eta_k^2 = \frac{1}{n} \left( (n - |J_{k-1}|) \eta_k^{2} + \sum_{i \in J_{k-1}} \| W_i^{k-1} \|^{2} \right)
\]

where $J_{k-1} = \{ i \in [1, n] : \| W_i^{k-1} \| > \eta_{k-1} \}$

Construct $C_k = \text{diagonal} (c_1, c_2, \ldots, c_n)$ where $c_i$ is given by:
where

\[
\mathbf{c}_i = \begin{cases} 
\frac{\sigma_k}{\|W^k_i\|} & \text{if } \|W^k_i\| < \sigma_k \\
\frac{\eta_k}{\|W^k_i\|} & \text{if } \|W^k_i\| > \eta_k \\
\frac{\zeta_k}{\|W^k_i\|}, \text{ Where } \frac{y^T_k V_k y_k}{y^T_k s_k} & \text{otherwise}
\end{cases}
\]

Compute: \( \gamma_{k+1} = \mathbf{W}_k \mathbf{C}_k \)

Step (5) Set \( k = k + 1 \) and go to step (1).

3. **Numerical Results**

In order to assess the value of this new technique, numerical tests on twenty test functions were carried out for unconstrained optimization problems. As a standard for the purpose of comparison, the test functions, (from general literature) were solved using two different VM-algorithms.

(i) The standard BFGS algorithm.

(ii) The new proposed algorithm (which it has been proved to be superlinear convergent algorithm).

All the numerical results were presented in tables (1)-(2). All the algorithms terminate whenever \( g^T_{k+1} g_{k+1} < 1 \times 10^{-5} \) and the two algorithms use exactly the same line search strategy, namely, the cubic fitting technique directly adapted from that published by [8].

Analysis of the two tables shows that the new proposed VM-algorithm is superior to the standard BFGS algorithm. The superiority of the new algorithm is clear for high dimensionality test problems because of the automatic scaling strategy.
### Table (1): Comparison of the new algorithm with the standard BFGS for $2 < n < 10$.

| Test Function | N | New algorithm | Standard BFGS |
|---------------|---|---------------|---------------|
|               |   | NOI | NOF | NOI | NOF |
| OSP           | 2 | 4   | 24  | 8   | 44  |
| Helical       | 3 | 22  | 60  | 19  | 59  |
| Cubic         | 4 | 7   | 34  | 8   | 26  |
| Rosen         | 4 | 12  | 41  | 35  | 106 |
| Powell        | 4 | 16  | 72  | 19  | 79  |
| Wood          | 4 | 20  | 72  | 30  | 84  |
| NON           | 4 | 26  | 87  | 21  | 66  |
| Miele         | 4 | 23  | 79  | 25  | 94  |
| OSP           | 10| 17  | 79  | 20  | 105 |
| Full          | 10| 9   | 19  | 9   | 19  |
| **Total**     |   | 156 | 567 | 194 | 682 |

Percentage improvement of the new algorithm compared with the standard BFGS algorithm:

|        | New 100 % NOI | BFGS 124.358 |
|--------|---------------|--------------|
| **Total** | 156           | 567          |

### Table (2): Comparison of the new algorithm with the standard BFGS for $100 < n < 900$.

| Test Function | N | New algorithm | Standard BFGS |
|---------------|---|---------------|---------------|
|               |   | NOI | NOF | NOI | NOF |
| Powell        | 100| 29  | 89  | 34  | 107 |
| Wood          | 100| 122 | 340 | 232 | 747 |
| Rosen         | 100| 18  | 55  | 244 | 767 |
| Miele         | 100| 29  | 91  | 31  | 107 |
| Dixon         | 300| 231 | 644 | 244 | 644 |
| Cubic         | 700| 10  | 39  | 13  | 39  |
| Wolfe         | 800| 78  | 169 | 84  | 169 |
| Powell        | 800| 32  | 100 | 39  | 119 |
| Cantrel       | 900| 15  | 95  | 12  | 61  |
| Miele         | 900| 31  | 93  | 33  | 109 |
| **Total**     | 595| 1715| 966 | 2869|
Percentage improvement of the new algorithm compared to standard BFGS algorithm

|       | 100 % NOI | 100 % NOF |
|-------|-----------|-----------|
| New   | 100%      | 100%      |
| BFGS  | 162.352   | 167.288   |

4. Final Remarks and Conclusions

We have described in this paper the conditions under which new automatic self-scaling algorithms based on the direct form of [1] VM-Update can be proved to be superlinearly convergent. Also some sort of numerical experiments have been done to know the effectiveness of the new proposed algorithm.

It is also possible to describe another similar algorithm based on the inverse scaled-BFGS algorithm. A column scaling algorithm which was proposed by [15] may be modified and implemented with this family of algorithms.

However, values of $\sigma_k$, $\mu_k$ selected in the new algorithm are arbitrary. It might occasionally be better to increase $\sigma_k$ and to decrease $\mu_k$. In any case, the theory developed in this paper will prove to be useful for analyzing the super linear convergence of this algorithm.

Finally this idea may be extended to constrained optimization problems, see [5] for more details and for non-quadratic models see [4].
REFERENCES

[1] Al-Bayati A.Y. (1991). “A new Family of self-scaling variable metric algorithms for unconstrained optimization”, *Journal of Educ. And Sci., Mosul*, 12, pp. 25-54.

[2] Al-Bayati A.Y. (1993). “A new non-quadratic model for unconstrained nonlinear optimization”, *J. of Mu’tah, Jordan, Vol (8)* pp 131-155.

[3] Al-Bayati A.Y. and Al-Assady (1994). “Minimization of extended quadratic functions with inexact line searches”, *JOTA, Vol (82)*, pp 139-147.

[4] Al-Bayati, A.Y. (2001), “new generalized CG-methods for the non-quadratic model in unconstrained optimization”, *J. of Al-Yarmouk, Jordan, Vol.(10)*, pp. 1-17.

[5] Al-Bayati, A.Y. and Hamed, E.T. (1998). “New self-scaling sequential algorithms for the minimization of constrained non linear functions”, *J. of Dirasat, Jordan, Vol.(25)*, pp. 339-351.

[6] Al-Bayati, A.Y. and Salih M. S. (2003), “Automatic Self-Scaling Strategies For VM-Updates”, *Al-Rafidain J. of Comp. And Math. Science, Vol (1)*, pp. 93-99.

[7] Broydon C.G. (1967), “Quasi-Newton methods and their application tp function minimization”, *Math. Comp.*, 21, pp. 368-381.

[8] Bunday, B.D. (1984), “Basic optimization problems” Edward Arnold, *Bedford square, London.*

[9] Byrd, R and Nocedal, J. (1989). “A tool for analysis of quasi-Newton methods with applications to unconstrained minimization”, *SIAM J. Numerical Analysis, Vol2, No.4*, pp. 533-557.

[10] Byrd, R and Nocedal, J. (1992). “On the Behavior of Broydon’s Class of quasi-Newton Methods” *SIAM J. Numerical Analysis, 26*, pp. 727-739.

[11] Davidon, W. C. "Variable Metric Methods for minimization", *A.E.C. Research and Development, Report*
ANL-5990 Argonne National Laboratory, Argonne, Illinois, (1959).

[12] Dennis J.E and More J.J (1977), “Quasi-Newton methods, motivation and theory”, SIAM Review, Vol(19), pp. 46-89.

[13] Fletcher, R. (1987). “Practical Methods of optimization”, John Wiely and sons, New York, Toronto and Singapore.

[14] Fletcher, R. and Reeves, C. M., "Function Minimization by Conjugate Gradients", Computer Journal, Vol. 7, pp. 163-168, (1964).

[15] Siegel, D. (1991). “Modifying the BFGS Update by a new column scaling technique”, Tech. Rep. DAMTP 1991/NAS, Dept. of Applied Mathematics and Theoretical physics, University of Cambridge, U.K.