Global and Suplinear Convergent VM-Algorithms for nonlinear Optimization

Abbas Y. Al-Bayati and Maha S. Al-Salih
College of Computers Sciences and Mathematics
University of Mosul

Abstract

In this paper a new class of self-scaling VM-algorithms for nonlinear optimization are investigated. Some theoretical results are given on the scaling strategies that guarantee the global and super linear convergence of the new proposed algorithms. Numerical evidence on thirty two well-known nonlinear test functions is generally encouraging.

Introduction

Consider the nonlinear optimization problem $\min_{x\in R^n} f(x)$, where f is a nonlinear differentiable function. Assume that an exact line search is used at the beginning of each iteration k, and that for an estimate vector x_k there is a symmetric and positive definite matrix B_k. The new iteration is computed by

$$d_k = -B_k^{-1} g_{k,} \qquad \ldots(1)$$

$$x_{k+1} = x_k + \lambda_k d_k, k \geq 1 \qquad \ldots(2)$$

where g_k is the gradient of the objective function at x_k. λ_k is a steplength satisfies the Wolfe conditions with exact line search strategy, i.e.

$$f(x_k + \lambda_k d_k) \leq f(x_k) + \alpha \lambda_k g_k^T d_k \qquad \ldots(3)$$

$$g(x_k + \lambda_k d_k)^T d_k \geq \beta g_k^T d_k \qquad \ldots(4)$$

for $0 < \alpha < \frac{1}{2}$ and $\alpha < \beta < 1$.

It is important for d_k to be a descent direction so that

$$f(x_k + \lambda_k d_k) < f(x_k)$$

for some $\lambda_k > 0$. Thus we most have

$$d_k^T g_k < 0$$

where $g_k = \nabla f(x_k)$

Quasi-Newton Methods

Here

$$d_{k+1} = -H_{k+1}\, g_{k+1} \qquad \ldots(5)$$

with H_{k+1} , an approximation to

$G_{k+1} = \nabla^2 f(x_{k+1})$ which satisfy the QN-condition defined by:

$$H_{k+1} y_k = \delta_k \qquad \ldots(6a)$$

where

$$\left.\begin{aligned} \delta_k &= x_{k+1} - x_k \\ y_k &= g_{k+1} - g_k \end{aligned}\right\} \qquad \ldots(6b)$$

A family of H_{k+1} satisfy (5) is Broyden family

$$H_{k+1} = H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} \qquad \ldots(7)$$

$$+ \frac{s_k y_k^T}{s_k^T y_k} + \phi_k\,(y_k^T H_k y_k) L_k L_k^T$$

where

$$L_k = \frac{S_k}{S_k^T y_k} - \frac{H_k y_k}{y_k^T H_k y_k} \qquad \ldots(8)$$

and ϕ_k is free parameter.Quasi Newton methods are quite efficient but need to store H_k and require $O(n^2)$multiplications per iteration to update H_k .

Note that this is done only for a quadratic model. But for non quadratic models, see(Al-Bayati,1993,Al-Bayati&Al-Assady,1994 and Al-Bayati,2001). for the details of standard VM steps. For the next iteration B_{k+1} is updated by Al-Bayati's VM-update i.e.

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{s_k^T B_k y_k}{(s_k^T y_k)^2} \cdot y_k^T y_k \qquad \ldots(9)$$

See (Al-Bayati,1991) for more details and properties of this algorithm.

New Suggestion

In this section we describe the prototype for the new suggested class of algorithms with self-scaling strategies:

Algorithm (1):

(1) For an starting point x_1 and non singular matrix V_1 ; set k =1.

(2) Terminate if $\|g_{k+1}\|_2 < \in, \in$is small positive real number.

(3) Compute

$$d_k = -V_k^T V_k^{-1} g_k$$
$$x_{k+1} = x_k + \lambda_k d_k$$

λ_k is computed by exact line search .

(4) 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 = diag [c_1 , c_2 , …., c_n] where

$$c_i = \begin{cases} \dfrac{\sigma_k}{\|w_i\|} & \text{if } \|w_i\| < \sigma_k \\ \dfrac{\mu_k}{\|w_i\|} & \text{if } \|w_i\| > \mu_k \\ \dfrac{\zeta_i}{\|w_i\|} \text{ where } \zeta_k = \dfrac{y_k^T V_k y_k}{y_k^T s_k} & \text{otherwise} \end{cases} \qquad (10)$$

(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

$$\left.\begin{aligned} B_1 &= V_1 V_1^T \\ B_k &= V_k V_k^T \\ &= W_{k-1}^T C_{k-1}^2 W_{k-1}^T \qquad k > 1 \end{aligned}\right\} \qquad \ldots(11)$$

and the update is performed directly on V_k .

2- It will be shown that one has considered freedom in choosing σ_k and μ_k of every iteration while still maintaining global convergence of the above algorithm .

The Global Convergence of the New Algorithm (1)

In this section, we will prove that the new algorithm suggested in section (3) with an appropriate choice of the scaling parameters is globally convergent on strictly convex objective functions.

Lemma 1: For any n×n matrices A and C, where C in diagonal matrix

$$\text{Tr}(ACA^T) = \text{tr}(AA^T) + \text{tr}[(C - I) A^T A] \qquad \ldots(12)$$

Where tr, denotes trace of any matrix.

Proof: For any two matrices A and B

$tr(AB) = tr(BA)$

$\Rightarrow tr(ACA^T) = tr(CA^TA)$

$= tr(AA^T) + tr(CA^TA) - tr(A^TA)$

Eq. (12) follows directly from the last equality #

Lemma 2: Let $h(u) = \ln u - u$ for $u > 0$

Let $\delta_1 > 0$, $\delta_2 > 0$ $\exists$ δ_3 and δ_4 $\ni$

$$x \in (0, \delta_1] \text{ and } y \in (0, x] \Rightarrow h(y) - h(x) \leq \delta_3 \quad \ldots(13)$$

And

$$x \in [\delta_2 , \infty) \text{ and } y \in [x , \infty) \Rightarrow h(y) - h(x) \leq \delta_4 \quad \ldots(14)$$

Proof: To prove eq.(13) we first note that h(u) is strictly concave and its maximum occurs at u =1. If $x \in (0, \min(\delta_1, 1))$ we conclude that for any $y \in (0, x]$.

$h(y) - h(x) \leq 0$ since h(u) is strictly increasing for $0 < u \leq 1$.

On the other hand, if $x \in [\min(\delta_1 , 1), \delta_1]$ then for any $y \in (0, x]$ we have $h(y) - h(x) \leq h[\min(\delta_1 , 1), - h(\delta_1)]$. Thus eq.(13) holds in either case with $\delta_3 = h[\min(\delta_1 , 1) - h(\delta_1)]$. We can prove eq.(14) in a similar line with $\delta_4 = h[\max(\delta_2 , 1)-h(\delta_2)]$. Details and explanations may be found in (Byrd et al, 1987).

Now let G(x) denote the Hessian matrix of f at x.

Let $D(\bar{x}) = \{x \in R^n ; f(x) \leq f(\bar{x})\}$ be the level set of f at $\bar{x}$.

Let x_1 be the starting point. Assume also

(1) f is twice continuously differentiable.

(2) $D(x_1)$ is convex.

(3) $\exists$ m > o and M $\ni$ $\forall z \in R^n$ and $x \in D(x_1)$

$$m \| z \|^2 \leq z^T G(x) z \leq M \| z \|^2$$

These three assumptions readily imply that f is strictly convex in $D(x_1)$. Also $\exists$ a unique minimizer x^* of f in $D(x_1)$ and for any positive define matrix B, we define

$$\psi(B) = tr(B) - \ln(\det(B)) \quad \ldots(15)$$

This result has been used by (Byrd & Nocedal,1989;Griewank, 1991)in their analysis of QN methods.

Let us define

$$\cos\theta_k = \frac{s_k^T B_k s_k}{\| s_k \| \| B_k s_k \|} \quad \ldots(16)$$

So that θ_k is the angle between the search direction d_k and the steepest - descent direction – g_k . Define also

$$q_k = \frac{s_k^T B_k s_k}{s_k^T s_k} \qquad \ldots(17)$$

Also assume that the scaling parameters σ_k and μ_k are bounded such that for all k.

$$\sigma_k \le \sigma_{max} \ , \mu_k \le \mu_{min} \qquad \ldots(18)$$

for some σ_{max} and μ_{min} .

The following new theorem provides the foundation for the proof of global convergence of our new suggested algorithm given in section (3). It generalizes a similar result given by (Byrd & Nocedal, 1989)for their algorithm but for the case of unscaled BFGS algorithm.

Theorem: 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 σ_k and μ_k satisfying eq.(18) and for any $\rho \in (0, 1)$ $\exists$ a constant β_1 ∍ for any k >1 the relation Cos $\theta_j \ge \beta_1$ holds for at least $[P_k]$ values of $j \in [1, k]$.

Proof: First we note that the symmetric matrices $B_k = V_k V_k^T = W_{k-1} C_{k-1}^2 W_{k-1}^T$ generated by the algorithm are positive definite, because the W_{k-1} are nonsingular as a consequence of the (Al-Bayati , 1991) update, and the C_{k-1} are nonsingular by construction.

Using the definition (15) of ψ , eq.(11) and lemma (4.1), we have

$$\psi (B_{k+1}) = tr (B_{k+1}) - \ln (\det (B_{k+1}))$$
$$= tr (W_k C_k^2 W_k^T) - \ln (\det (W_k C_k^2 W_k^T))$$
$$= tr (W_k W_k^T) - tr[(C_k^2 - I) W_k^T W_k)] - \ln \det (W_k W_k^T) - \ln \det (C_k^2)$$
$$= \psi (W_k W_k^T) + tr((C_k^2 - I) W_k^T W_k) - \ln \det (C_k^2)$$
$$= \psi (W_k W_k^T) + \sum_{i=1}^{n} \left[(C_i^2 - I) \| w_i \|^2 - \ln C_i^2 \right]$$

Where w_i is the i th Column of W_k now scaling up and down the set of indices of the column W_k as

$$I_k = \left(i \in [1, n] : \| w_i \| < \sigma_k \right) \qquad \ldots(19)$$

and

$$J_k = \left(i \in [1, n] : \| w_i \| > \mu_k \right) \qquad \ldots(20)$$

and $R_{tc} = (i \in [1, n])$ otherwise

Therefore by define of the scalar c_i in our new proposed algorithm

$$\psi(B_{K+1}) = \psi(W_k\ W_k^T) + \sum_{i\in I_k}\left[\left(\frac{\sigma^2{}_k}{\|w_i\|^2}-1\right)\|w_i\|^2 - \ln\frac{\sigma^2{}_k}{\|w_i\|^2}\right]$$

$$+\sum_{i\in J_k}\left[\left(\frac{\mu^2_k}{\|w_i\|^2}-1\right)\|w_i\|^2 - \ln\frac{\mu^2_k}{\|w_i\|^2}\right]$$

$$+\sum_{i\in R_k}\left[\left(\frac{\zeta^2{}_k}{\|w_i\|^2}-1\right)\|w_i\|^2 - \ln\frac{\zeta^2{}_k}{\|w_i\|^2}\right]$$

$$= \psi(W_k\ W_k^T) + \sum_{i\in I_k}\left[(\ln\|w_i\|^2 - \|w_i\|^2) - (\ln\sigma_k^2 - \sigma_k^2)\right]$$

$$+\sum_{i\in J_k}\left[(\ln\|w_i\|^2 - \|w_i\|^2) - (\ln\mu_k^2 - \mu_k^2)\right]$$

$$+\sum_{i\in R_k}\left[(\ln\|w_i\|^2 - \|w_i\|^2) - (\ln\zeta_k^2 - \zeta_k^2)\right]$$

We will now involve lemma (4.2) with $\delta_1 = \sigma_{max}$ and $\delta_2 = \mu_{min}$ since $\|w_i\| \le \sigma_k$ for $i\in I_k$ whereas $\|w_i\| \ge \mu_k$ for $i \in J_k$ and $\|w_i\| \ge \zeta_k$ for $i \in R_k$ we can therefore apply eq.(13) to each term of the first summation, and eq.(14) to each term of the 2nd summation to obtain

$$\psi(B_{k+1}) \le \psi(w_k\ w_k^T) + n\,\delta_3 + n\,\delta_4 \qquad \ldots(21)$$

for the constants δ_3 and δ_4 given by lemma (4. 2).

Now step (4) of our new suggested algorithm (1) indicates that the matrix $W_k\ W_k^T$ is Al-Bayati's update of B_k.Therefore by the same procedure of (Byrd & Nocedal, 1989)we can claim that $\psi\ (B_{k+1})$ is bounded and $\cos\theta_j \ge B_1$ .To ensure that the new algorithm generates a sequence of $\{x_k\}$ that converge to x^*, i.e.

$$\sum_{k=1}^{\infty}\|x_k - x^*\| < \infty \qquad \ldots(22)$$

and $\quad f_{k+1} - f^* \le r^k\ (f_1 - f^*) \qquad \ldots(23)$

for some constant $r \in [0, 1)$

To prove (23) let us start with $f_{k+1} - f^* \le (1 - \delta_6 \cos^2\theta_k)\ (f_k - f^*)$ see (Byrd et al, 1987)for the theoretical explanations.

Now since $\cos\theta_j \ge \beta_1$ then

$f_{k+1} - f^* \le (1 - \delta_6\ \beta_1^2)\ (f_k - f^*) \le r^k\ (f_k - f^*)$

with $r = (1 - \delta_6\ \beta_1^2) \in [0, 1)$ where $\delta_{6=\frac{\alpha_m(1-B_k)}{M}}, \alpha = \beta = B_1$

The assumption on f also imply that $\frac{1}{2} m \| x_k - x^* \|^2 \le f_k - f^*$ …(24)

Therefore combining (24) with (23) we obtain

$$\sum_{k=1}^{\infty} \| x_k - x^* \| \le (\frac{2}{m})^{\frac{1}{2}} \sum_{k=1}^{\infty} (f_k - f^*)^{\frac{1}{2}} \le \left[\frac{2(f_1 - f^*)}{m} \right]^{\frac{1}{2}} \sum_{k=0}^{\infty} (r^{\frac{1}{2}})^k < \infty$$

(since the series is geometric series and it converges to a finite sum)

This proves the global convergence of our new proposed algorithm (3.1)#

Super Linear Convergence

First we define the following quantities to be used in this section:

$$\overline{B}_k = G_*^{-\frac{1}{2}} B_k G_*^{-\frac{1}{2}} \quad , \quad \overline{W}_k = G_*^{-\frac{1}{2}} W_k \quad \ldots(25)$$

$$\bar{s}_k = G_*^{\frac{1}{2}} s_k \quad , \quad \bar{y}_k = G_*^{-\frac{1}{2}} y_k \quad \ldots(26)$$

$$\overline{M}_k = \frac{\bar{y}_k^T \bar{y}_k}{\bar{y}_k^T \bar{s}_k} \quad , \quad \overline{m}_k = \frac{\bar{y}_k^T \bar{s}_k}{\bar{s}_k^T \bar{s}_k} \quad \ldots(27)$$

$$\bar{q}_k = \frac{\bar{s}_k^T \overline{B}_k \bar{s}_k}{\bar{s}_k^T \bar{s}_k} \quad , \quad \mathrm{Cos}\,\bar{\theta}_k = \frac{\bar{s}_k^T \overline{B}_k \bar{s}_k}{\|\bar{s}_k\| \|\overline{B}_k \bar{s}_k\|} \quad \ldots(28)$$

where G_* is the Hessian of f at the minimizer x_* .

We have shown (see lemma 4.2) that the limiting behavior of $\bar{q}_k$ and $\mathrm{Cos}\,\bar{\theta}_k$ 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 (Dennis & More , 1977) characterization, is reproduced in the following lemma.

Lemma: Suppose that the sequence of iterates $\{x_k\}$ is generated by algorithm (1)-(2) using some positive definite sequence $\{B_k\}$, and that $\lambda_k = 1$ whenever this value satisfies Wolfe conditions (3)-(4). If $x_k \to x_*$ then the following two conditions are equivalent :

(i) The steplength $\lambda_k = 1$ satisfies conditions (3)-(4) for all larg k and the rate of convergence is superlinear.

(ii) $\lim_{k\to\infty} \mathrm{Cos}\,\bar{\theta}_k = \lim_{k\to\infty} \bar{q}_k = 1$ …(29)

Proof: Proof this lemma can be found in (Byrd &Nocedal, 1989). The next theorem specifies conditions on the scaling parameters σ_k and η_k that allow $\bar{q}_k$ and $\mathrm{Cos}\,\bar{\theta}_k$, produced by Algorithm 3.1, to exhibit the desirable limiting behavior of Lemma 5.1 . Such conditions involve following quantities:

$$\gamma_k = \sum_{i\in I_k}\left[(\ln \left\| G_*^{-\frac{1}{2}}\, wi \right\|^2 - \left\| G_*^{-\frac{1}{2}}\, wi \right\|^2) - (\ln \sigma_k^2 \frac{\left\| G_*^{-\frac{1}{2}}\, wi \right\|^2}{\|wi\|} - \sigma_k^2 \frac{\left\| G_*^{-\frac{1}{2}}\, wi \right\|^2}{\|wi\|^2}) \right] \quad \ldots(30)$$

and

$$\mu_k = \sum_{i\in J_k}\left[(\ln \left\| G_*^{-\frac{1}{2}}\, wi \right\|^2 - \left\| G_*^{-\frac{1}{2}}\, wi \right\|^2) - (\ln \eta_k^2 \frac{\left\| G_*^{-\frac{1}{2}}\, wi \right\|^2}{\|wi\|} - \eta_k^2 \frac{\left\| G_*^{-\frac{1}{2}}\, wi \right\|^2}{\|wi\|^2}) \right]$$

$$\phi_k = \sum_{i\in R_k}\left[(\ln \left\| G_*^{-\frac{1}{2}}\, wi \right\|^2 - \left\| G_*^{-\frac{1}{2}}\, wi \right\|^2) - (\ln \zeta_k^2 \frac{\left\| G_*^{-\frac{1}{2}}\, wi \right\|^2}{\|wi\|} - \zeta_k^2 \frac{\left\| G_*^{-\frac{1}{2}}\, wi \right\|^2}{\|wi\|^2}) \right] \quad \ldots(31)$$

And whether or not they sum finitely. Note that γ_k and μ_k need not be positive. Recall that the sets I_k and J_k defined by (19) and (20) contain the indices of the columns that are scaled down at iteration k. We are now ready to state the theorem.

Theorem: Let f, x_1, B_1, σ_k and η_k satisfy the assumptions in theorem 4.3 . In addition, assume that G is Lipschitz continuous at x_*. Let $\{x_k\} \to x_*$ be generated by Algorithm 3.1; then if

$$\sum_{k=1}^{\infty} \gamma_k < \infty \quad \ldots(32)$$

$$\sum_{k=1}^{\infty} \mu_k < \infty$$

$$\sum_{k=1}^{\infty} \phi_k < \infty \quad \ldots(33)$$

the iterates converge superlinearly.

Proof: From the definition (15) of ψ and from (11), (12) and (25), we have

$$\psi(\overline{B}_{k+1}) = \mathrm{tr}(G_*^{-\frac{1}{2}}\, W_k\, C_k^2\, W_k^T\, G_*^{-\frac{1}{2}}) - \ln\det(G_*^{-\frac{1}{2}}\, W_k\, C_k^2\, W_k^T\, G_*^{-\frac{1}{2}})$$

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

$$= \psi(\tilde{W}_k\ \tilde{W}_k^T) + \sum_{i=1}^{n}\left[(c_i^2 - 1)\| G_*^{-\frac{1}{2}}\ W_i \|^2 - \ln c_i^2\right]$$

Then by the definition (10) of c_i ,

$$\psi(\tilde{B}_{k+1}) = \psi(\tilde{W}_k\ W_k^T) + \sum_{i\in J_k}\left[\left(\frac{\sigma_k^2}{\| W_i \|^2} - 1\right)\| G_*^{-\frac{1}{2}}\ W_i \|^2 - \ln\frac{\sigma_k^2}{\| W_i \|^2}\right]$$

$$+ \sum_{i\in J_k}\left[\left(\frac{\eta_k^2}{\| W_i \|^2} - 1\right)\| G_*^{-\frac{1}{2}}\ W_i \|^2 - \ln\frac{\eta_k^2}{\| W_i \|^2}\right]$$

$$+ \sum_{i\in R_k}\left[\left(\frac{\zeta_k^2}{\| W_i \|^2} - 1\right)\| G_*^{-\frac{1}{2}}\ W_i \|^2 - \ln\frac{\zeta_k^2}{\| W_i \|^2}\right]$$

$$= \psi(\tilde{W}_k\ W_k^T) + \sum_{i\varepsilon I_k}\left[\sigma_k^2\ \frac{\| G_*^{-\frac{1}{2}}\ W_i \|^2}{\| W_i \|^2} - \| G_*^{-\frac{1}{2}}\ W_i \|^2\right.$$

$$\left. - \ln\sigma_k^2\ \frac{\| G_*^{-\frac{1}{2}}\ W_i \|^2}{\| W_i \|^2} + \ln\| G_*^{-\frac{1}{2}}\ W_i \|^2\right]$$

$$+ \sum_{i\in J_k}\left[\eta_k^2\ \frac{\| G_*^{-\frac{1}{2}}\ W_i \|^2}{\| W_i \|} - \| G_*^{-\frac{1}{2}}\ W_i \|^2\right.$$

$$\left. - \ln\eta_k^2\ \frac{\| G^{-\frac{1}{2}}{}_*\ W_i \|^2}{\| W_i \|^2} + \ln\| G_*^{-\frac{1}{2}}\ W_i \|^2\right]$$

$$+ \sum_{i\in J_k}\left[\zeta_k^2\ \frac{\| G_*^{-\frac{1}{2}}\ W_i \|^2}{\| W_i \|} - \| G_*^{-\frac{1}{2}}\ W_i \|^2\right.$$

$$\left. - \ln\zeta_k^2\ \frac{\| G_*^{-\frac{1}{2}}\ W_i \|^2}{\| W_i \|^2} + \ln\| G_*^{-\frac{1}{2}}\ W_i \|^2\right]$$

$$= \psi(\tilde{W}_k\ W_k^T) + \gamma_k + \mu_k + \phi_k \qquad \ldots(34)$$

Since $\tilde{W}_k\ W_k^T$ is the matrix obtained by updating B_k using the (Al-Bayati,1991)formula,which is invariant under the transformation (25)– (28), we have:

$$\psi(\tilde{W}_k \tilde{W}_k^T) = \psi(\tilde{B}_k) + (\tilde{M}_k - \ln \tilde{m}_k - 1) + (1 - \frac{\tilde{q}_k}{\cos^2\theta_k} + \ln \frac{\tilde{q}_k}{\cos^2\theta_k}) + \ln \cos^2\tilde{\theta}_k \quad \ldots(35)$$

Therefore, using (35) in (34), we have:

$$\left.\begin{aligned}\psi(\tilde{B}_{k+1}) &= \psi(\tilde{B}_k) + (\tilde{M}_k - \ln \tilde{m}_k - 1) + (1 - \frac{\tilde{q}_k}{\cos^2\tilde{\theta}_k} + \ln \frac{\tilde{q}_k}{\cos^2\tilde{\theta}_k}) \\ &\quad + \ln \cos^2\tilde{\theta}_k + \gamma_k + \mu_k \\ &= \psi(\tilde{B}_1) + \sum_{j=1}^{k}(\tilde{M}_j - \ln \tilde{m}_j - 1) + \sum_{j=1}^{k}[(1 - \frac{\tilde{q}_k}{\cos^2\tilde{\theta}_k} + \ln \frac{\tilde{q}_k}{\cos^2\tilde{\theta}_k}) \\ &\quad + \ln \cos^2\tilde{\theta}_j] + \sum_{j=1}^{k}\gamma_j + \sum_{j=1}^{k}\mu_j + \sum_{j=1}^{k}\phi_j\end{aligned}\right\} \ldots(36)$$

By Theorem 4.3, 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 (Byrd & Nocedal, 1989)that:

$$\sum_{j=1}^{k}(\tilde{M}_j - \ln \tilde{m}_j - 1) < \infty \quad \ldots(37)$$

Moreover, the hypothesis of the theorem guarantees that the last two summations in (36) 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:

$$\lim_{k\to\infty} (1 - \frac{\tilde{q}_k}{\cos^2\tilde{\theta}_k} + \ln \frac{\tilde{q}_k}{\cos^2\theta_k}) = \lim_{k\to\infty} \ln \cos^2\tilde{\theta}_k = 0$$

Which implies that both $\tilde{q}_k$ and $\cos^2\tilde{\theta}_k \to 1$. Hence, superlinear convergence follows from Lemma (5.1) #.

Now in the following section we describe a specific and modified implementation of algorithm 3.1 and make use of theory developed so far to show that it is globally and superlinearly convergent for strictly convex objective functions.

Algorithm

Automatic column scaling (Al-Bayati, 1991)VM-algorithm. This is a modified version from our first proposed algorithm (3.1).

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

(1) Terminate if a stopping criterion is satisfied.

(2) Find an orthogonal matrix Q_k such that $L_k = V_k Q_k$ is lower triangular.

Compute :

$d_k = -L_k^{-T} L_k^{-1} g_k,$

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

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

Compute:

$s_k = x_{k+1} - x_k$

$y_k = g_{k+1} - g_k$

(3) Perform the following steps to update L_k to W_k so that $W_k W_k^T$ is the Al-Bayati update of $L_k L_k^T$ defined in (9):

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

(3.2) Find an orthogonal and lower matrix Ω_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

$$w_i^k = \begin{cases} y_k / \sqrt{y_k^T s_k} & , i = 1 \\ L_k \Omega_k e_i & , i = 2, 3, \ldots, n \end{cases}$$

(4) Compute the scaling parameters:

If k = 1, $\sigma_1^2 = \eta_1^2 = \dfrac{y_1^T y_1}{s_1^T y_1} = \zeta_1^2$

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

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

And $\eta_k^2 = \dfrac{1}{n} [(n - |J_{k-1}|)\, \eta_{k-1}^2 + \sum_{i \in J_{k-1}} \| W_i^{k-1} \|^2]$,

And $\zeta_k^2 = \dfrac{1}{n} [(n - |R_{k-1}|)\, \zeta_{k-1}^2 + \sum_{i \in R_{k-1}} \| W_i^{k-1} \|^2]$,

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

Construct C_k = diagonal $(c_1, c_2, \ldots, c_n)$ where c_i givin by:

$$c_i = \begin{cases} \dfrac{\sigma_k}{\| W_i^k \|} & \text{if} \quad \| W_i^k \| < \sigma_k \\ \dfrac{\eta_k}{\| W_i^k \|} & \text{if} \quad \| W_i^k \| > \eta_k \\ \dfrac{\zeta_k}{\| w_i^k \|} \; where\, \zeta_k = \dfrac{y_k^T V_k y_k}{y_k^T s_k} & \text{otherwise} \end{cases}$$

Compute: $\gamma_{k+1} = W_k \; C_k$
(5) Set k = k + 1 and go to step (1).
Note that: at each iteration k begins with lower matrix V_k which defines $B_k = V_k \; V_k^T$. Also since $L_k = V_k \; Q_F$ we have that $B_k = L_k \; L_k^T$. This allows V to compute the search direction by two triangular solves.

Numerical Results

In order to asses the value of this new technique, numerical tests were carried out on a number of unconstraint 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 6.1 (which it has been proved to be global and superlinear convergent).

All the numerical results were presented in table (1)-(4). All the algorithm terminate whenever $g_{k+1}^T \, g_{k+1} < 1 \times 10^{-5}$ and the two algorithms use exactly the same line search strategy, namely, the cubic fitting technique directly adapted from that published by (Bunday, 1984).

Analysis of the four 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 the automatic scaling strategy.

Table (1): Comparison between standard BFGS algorithm with the new proposed algorithm n = 4 .

Test Function	New algorithm		Standard BFGS	
	NOI	NOF	NOI	NOF
Resonbrok (-1.2, 1, …)	12	41	31	93
Cubic (1.2, 1, …)	7	34	8	26
Freud (30, 3, …)	7	23	7	27
Powell (3, -1, 0, 1, …)	18	81	22	84
Wood (-3, -1, -3, -1, …)	28	100	56	159
Dixon (-1, …)	10	27	14	37
Miele (1, 2, 2, 2, …)	19	78	25	94
Cantrell (1, 2, 2, 2, …)	15	85	13	63
Total	116	469	176	583

Percentage improvement of the new algorithm compared against standard BFGS algorithm

BFGS	**100 % NOI**	**100 % NOF**
New	65.9	80.4

Table (2): Comparison between standard BFGS algorithm with the new proposed algorithm n = 40 .

Test Function	New algorithm		Standard BFGS	
	NOI	NOF	NOI	NOF
Resonbrok (-1.2, 1, …)	14	42	132	398
Cubic (1.2, 1, …)	10	42	9	29
Freud (30, 3, …)	8	25	8	29
Powell (3, -1, 0, 1, …)	37	101	35	100
Wood (-3, -1, -3, -1, …)	126	399	201	576
Dixon (-1, …)	43	90	60	123
Miele (1, 2, 2, 2, …)	24	92	30	105
Cantrell (1, 2, 2, 2, …)	16	91	13	63
Total	278	882	488	1423

Percentage improvement of the new algorithm compared against standard BFGS algorithm

BFGS	100 % NOI	100 % NOF
New	56.9	61.9

Table (3): Comparison between standard BFGS algorithm with the new proposed algorithm n = 100 .

Test Function	New algorithm		Standard BFGS	
	NOI	NOF	NOI	NOF
Resonbrok (-1.2, 1, …)	18	55	169	521
Cubic (1.2, 1, …)	10	40	13	37
Freud (30, 3, …)	8	25	8	29
Powell (3, -1, 0, 1, …)	41	128	42	129
Wood (-3, -1, -3, -1, …)	21	68	37	114
Dixon (-1, …)	93	192	129	262
Miele (1, 2, 2, 2, …)	28	104	31	107
Cantrell (1, 2, 2, 2, …)	16	91	14	69
Total	235	700	443	1268

Percentage improvement of the new algorithm compared against standard BFGS algorithm

BFGS	100 % NOI	100 % NOF
New	53	55.2

Table (4): Comparison between standard BFGS algorithm with the new proposed algorithm n = 200 .

Test Function	New algorithm		Standard BFGS	
	NOI	NOF	NOI	NOF
Resonbrok (-1.2, 1, …)	17	51	159	483
Cubic (1.2, 1, …)	9	37	13	39
Freud (30, 3, …)	8	23	10	32
Powell (3, -1, 0, 1, …)	39	117	40	120
Wood (-3, -1, -3, -1, …)	32	99	56	165
Dixon (-1, …)	89	183	123	249
Miele (1, 2, 2, 2, …)	28	104	31	107
Cantrell (1, 2, 2, 2, …)	16	91	14	69
Total	238	705	446	1264

Percentage improvement of the new algorithm compared against standard BFGS algorithm

BFGS	**100 % NOI**	**100 % NOF**
New	53.3	55.7

Final Remarks

We have described in this paper the conditions under which new automatic self-scaling algorithms based on the direct form of (Al-Bayati, 1991) VM-Update can be proved to be globally and super linearly convergent. Also some sort of numerical experiments have been done to inform the effectiveness of the new proposed algorithms. It is also possible to describe another similar algorithm based on the inverse scaled-BFGS algorithm. A column scaling algorithm which was proposed by (Siegel, 1991) may be modified and implemented with this family of algorithms.

However, values of σ_k, μ_k selected in the new algorithm may be described (in more details) in our further work. It might occasionally be better to increase σ_k and to decrease μ_k . in any case, the theory developed in this paper will prove to be useful for analyzing the global and super linear convergence of these algorithms.Finally this idea may be extended to constrained optimization problems see (Al-Bayati & Hamed, 1998) for more details.

References

- Al-Bayati A.Y.,(1991): A new Family of self-scaling variable metric algorithms for unconstrained optimization, Journal of Educ. And Sci., Mosul, Vol. 12, pp. 25-54.
- Al-Bayati A.Y.,(1993): A new non-quadratic model for unconstrained nonlinear optimization, J. of Mu'tah, Jordan, Vol. 8, pp. 131-155.
- Al-Bayati A.Y. and Al-Assady (1994): Minimization of extended quadratic functions with inexact line searches, JOTA, Vol.82, pp.139-147.
- 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.
- 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.
- Bunday, B.D.,(1984): Basic optimization problems Edward Arnold, Bedford square, London.
- Byrd, R and Nocedal, J.,(1989): A tool for analysis of quasi-Newton methods with applications to unconstrained minimization, SIAM J. Numerical Analysis,Vol. 26, pp. 727-739.
- Byrd, R. ; Nocedal, J. and Yuan, Y.,(1987): Global Convergence of a class of quasi-Newton methods on convex problems, SIAM J. Numerical analysis,Vol. 24, pp. 1171-1190.
- Dennis J.E and More J.J.,(1977): Quasi-Newton methods, motivation and theory, SIAM Review, Vol.19, pp. 46-89.
- Fletcher, R.,(1987): Practical Methods of optimization, John Wiely and sons, New York, Toronto and Singapore.
- Griewank, A.,(1991): The global convergence of partitioned BFGS on problems with convex decompositions and Lipschitzian gradients; Math. Programming, Vol. 50, pp. 141-175.
- 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.

Appendix

1. Generalization Powell Function:

$$f=\sum_{i=1}^{n/4}[x_{4i-3}+10\,x_{4i-2})^2+5\,(x_{4i-1}-x_{4i})^2+(x_{4i-2}-2\,x_{4i-1})^4$$
$$+10\,(x_{4i-3}-x_{4i})^4] \quad ; \qquad x_o=(3,-1,0,1,\ldots)^T$$

2. Generalized Wood Function:

$$f=\sum_{i=1}^{n/4}100\,[(x_{4i-2}+x_{4i-3}^2)^2]+(1-x_{4i-3})^2+90\,(x_{4i}-x_{4i-1}^2)^2$$
$$+(1-x_{4i-1})^{24}+10.1[(x_{4i-2}+1)^2+(x_{4i-1}-1)^2]+19.8\,(x_{4i-2}-1)\,(x_{4i}-1)$$
$$;\,x_o=(-3,-1,-3,-1;\ldots)^T$$

3. Generalized Cantrel Function:

$$f=\sum_{i=1}^{n/4}[\exp(x_{4i-3})-x_{4i-2})+100\,(x_{4i-2}-x_{4i-1})^6+(\arctan(x_{4i-1}-x_{4i}))^4+x_{4i-3}$$
$$;\,x_o=(1,2,2,2;\ldots)^T$$

4. Generalized Miele Function:

$$f=\sum_{i=1}^{n/4}[\exp(x_{4i-3}-x_{4i-1})^2+100\,(x_{4i-2}-x_{4i-1})^6+(\tan(x_{4i-1}-x_{4i}))^4$$
$$+x_{4i-3}^8+(x_{4i}-1)^2] \qquad ;\,x_o=(1,2,2,2;\ldots)^T$$

5. Cubic Function:

$$f=100\,(x_2-x_1^3)^2+(i-x_1)^2 \qquad ;\,x_o=(1,2,2,2;\ldots)^T$$

6. **Rosenbrock Function:**

$$f = 100(x_2 - x_1^2)^2 + (1 - x_1)^2 \qquad ; x_0 = (-1.2, 2, 1; \ldots)^T$$

7. **Dixon Function:**

$$f = (1 - x_1)^2 + (1 - x_{10})^2 + \sum_{i=2}^{9}(x_i - x_{i+1})^2 \qquad , x_0 = (-1, \ldots)^T$$

8. **Freudenstein Function:**

$$f = [-13 + x_1 - ((5 - x_2)x_2 - 2)x_2] + [-29 + x_1 + ((1 - x_2)x_2) - 14)x_2]^2$$

$$; x_0 = (30, 3)^T$$

List of symbols

Symbol	**Meaning**
n	is the dimensions of the problems
K	is the K-th step of iterations
F	is the twice differentiable real value function
x^*	is the local minimum of f(x)
x	is an approximation to x^*
g	is the n ×1 gradient vector of f(x)
d	is the n ×1 search direction vector
G	is the n ×n Hessian matrix
H	is the n ×n approximation to G^{-1} matrix
B	is the n ×n approximation to G matrix
Y	is the n ×1 difference vector between two successive gradients
v	is the n ×1 difference vector between two successive points
λ	is the positive scalar which minimizes f(x- λHg)
ELS	is the exact line search
ILS	is the inexact line search
QN	is the Quasi-Newton
VM	is the Variable metric
CG	is the Conjugate Gradient
NOF	is the number of function evaluations
NOI	is the number of iterations

Journal of Kirkuk University –Scientific Studies , vol.2, No.3,2007

خوارزميات ذوات التقارب الشامل والسرعة فوق الخطية في الأمثلية اللاخطية

عباس يونس البياتي و مها صلاح الصالح
كلية علوم الحاسبات والرياضيات
جامعة الموصل

الخلاصة

في هذا البحث تم التطرق إلى صنف جديد من خوارزميات المتري المتغير وفق تقنية خاصة بالقياس الذاتي . وتم كذلك دراسة بعض النتائج النظرية التي تؤكد التقارب الشامل والسرعة فوق الخطية للخوارزميات الجديدة المقترحة مع دراسة عملية تؤيد كفاءة الخوارزميات المقترحة. وباستعمال (٣٢) دالة غير خطيه معروفة.