Automatic Self-Scaling Strategies for VM Updates

Abbas Y. Al-Bayati **Maha S. Al-Salih**
profabbasalbayati@yahoo.com
Collage of Computers and Mathematics Science,
University of Mosul/Iraq

Received on: 12/08/2002 **Accepted on: 07/09/2002**

Abstract

In this paper a class of self-scaling VM-algorithms for unconstrained optimization is investigated. Some theoretical results are given on the scaling strategies that guarantee the global convergence of the new proposed algorithm.

Keywords: unconstrained optimization, VM-algorithms, self-scaling, global convergence.

وسيلة ذاتية القياس لتحديثات المتري المتغير

مهى صلاح الصالح **عباس يونس البياتي**
كلية علوم الحاسوب والرياضيات/جامعة الموصل/العراق

تاريخ استلام البحث: 2002/08/12 **تاريخ قبول البحث: 2002/09/07**

الملخص

في هذا البحث تم التطرق إلى صنف جديد من خوارزميات المتري المتغير على وفق تقنية خاصة بالقياس الذاتي. وتمت كذلك دراسة بعض النتائج النظرية التي تؤكد التقارب الشامل للخوارزمية الجديدة المقترحة.

الكلمات المفتاحية: الأمثلية غير المقيدة، خوارزميات المتري المتغير، القياس الذاتي، التقارب الشامل.

1. Introduction

Consider the unconstrained 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

$$d_k - B_k^{-1} g_k \tag{1}$$

$$x_{k+1} = x_k + \lambda_k d_k, k \geq 1 \tag{2}$$

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

$$f(x_k + \lambda_k d_k) \leq f(x_k) + \alpha\lambda_k g_k^T d_k \tag{3}$$

$$g(x_k + \lambda_k d_k)^T d_k \geq \beta g_k^T d_k \tag{4}$$

for $0 \leq \alpha \leq \frac{1}{2}$ and $\alpha \leq \beta \leq 1$. See Fletcher [6] for the details of standard VM step. For the next iteration B_{k+}, is updated by Al-Bayati's VM-update i.e.

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{s_k^T B_k y_k}{(s_k^T y_k)^2} . y_k^T y_k \tag{5a}$$

Where

$$\left.\begin{aligned} s_k &= x_{k+1} - x_k \\ y_k &= g_{k+1} - g_k \end{aligned}\right\} \tag{5b}$$

See Al-Bayati [1 ,2] for more details and properties of this algorithm.

2. New Suggestion

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

Algorithm:

1-For a starting point x, and non singular matrix V_1; set k =1.

2- Terminate if $\|g_{k+1}\| < \varepsilon$, ε is a small positive real number.

3-Compute

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

λ_k is computed ;by exact line search .

4- Update

$$W_k = V_k - \frac{V_k s_k s_k^T V_k}{s_k^T V_k s_k} + \frac{s_k^T V_k y_k}{(y_k^T s_k)^2} .y_k y_k^T$$

5- Compute the scaling parameter $\sigma_k \geq 0$ and , $\mu_k \geq 0$ such that $\sigma_k \leq \mu_k$. If w; represents the column of W_k Put $C_k = diag[c_1, c_2, ...c_n]$ where σ_k = 0.5 and μ_k =1

$$c_i = \begin{cases} \dfrac{\sigma_k}{\|w_i\|} & if \quad \|w_i\| \leq \sigma_k \\ \dfrac{\mu_k}{\|w_i\|} & if \quad \|w_i\| > \mu_k \\ \dfrac{s_k^T V_k y_k}{y_k^T} & otherwise \end{cases}$$

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

$$\left.\begin{aligned} B_1 &= V_1V_1^T \\ B_k &= V_kV_k^T \\ &= W_{k-1}^T C_{k-1}^2 W_{k-1}^T, k \geq 1 \end{aligned}\right\} \quad (6)$$

and the update is performed directly on V_k .

2-It will be shown that one has considered freedom in choosing δ_k and μ_k of every iteration, while still maintaining global convergence of the above algorithm. It is necessary that the choice of these values be made carefully.

3. Global Convergence of the New Algorithm

In this section, we prove that the new algorithm suggested in section (2) with an appropriate choice of the scaling parameters is globally convergent on strictly convex objective functions.

Lemma 3.1: For any nxn matrices A and C, where C diagonal matrix

$$\text{Tr}(ACA^T)=\text{tr}(AA^T)+\text{tr}[(C-1)AA^T] \quad (7)$$

Where tr, denotes trace of any matrix see [7].

Proof: For any two matrices A and B

tr (AB) = tr (BA)

$$\Rightarrow \begin{aligned} & tr(ACA^T) = tr(CA^T A) \\ & = tr(AA^T) + tr(CA^T A) - tr(A^T A) \end{aligned}$$

Eq. (7) follows directly from the last equality #

Lemma3.2: Let h(u) = In u-u for u>0

Let $\delta_1 > 0, \delta_2 > 0 \exists \delta_3$ *and* $\delta_4 \ni$

$$x \in (0, \delta_1] and \quad y \in (0, x] \Rightarrow h(y) - h(x) \leq \delta_3 \quad (8)$$

And

$$x \in [\delta_2 and \ \ y \in [x, \infty) \Rightarrow h(y) - h(x) \leq \delta_4 \quad (9)$$

Proof: To prove eq.(8) 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∈(o, X].

h(y)-h(x)≤0 since h(u) is strictly increasing for 0 < u<- 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) _< h [min (δ_1 , 1), - h(δ_1)]. Thus eq.(8) holds in either case with δ_3 = h [min (δ_1, 1) - h(δ_1)]. We can prove eq.(9) in a similar line with δ_4 = h [max (b_2 , 1) - h(δ_2)]. Details and explanations can be found in [3].

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

Let D(x)-{x E R *n ;* f(x) <- f(X)} be the level set of f at x . Let x, be the starting point. Assume also

(1) f is twice continuously differentiable. (2) D(x_l) is convex.

(3) $\exists m > 0$ and $M \ni \forall z \in R^n$ and x$\in D(x_1)$

$m\|z\|^2 \leq z^T G(x) z \leq M\|z\|^2$ Where the norm is norm two.

These three assumptions readily imply that f is strictly convex in D(x_l). Also 3 is a unique minimizer x^* of f in D(x_l) and for any positive defined matrix B, we define

y (B) = tr (B) - In (det (B)) (10)

This result has been used by Byrd and Nocedal [4] and Griewank [5] in their analysis of QN methods.

Let us defne

$$\cos\theta_k = \frac{s_k^T B_k s_k}{\|s_k\|\|B_k s_k\|} \tag{11}$$

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

$$q_k = \frac{s_k^T B_k s_k}{s_k^T s_k} \tag{12}$$

Also assume that the scaling parameters 6_k and ~tk are bounded such that for all k.

$$\sigma_k \le \sigma_{\max}, \mu_k \le \mu_{\min} \text{ for some } \sigma_{\max} \le \mu_{\min} \qquad .(13)$$

The following new theorem provides the foundation for the proof of global convergence of our new suggested algorithm given in section 2. It generalizes a similar result given by Byrd and Nocedal [4] for their algorithm but for the case of unscaled BFGS algorithm.

Theorem 3.3: Let x, be a starting point for which f satisfies eq.(7) and let B_I be a positive definite starting Hessian approximation. Let {x_k} be generated by the new proposed algorithm with 6_k and ~tk satisfying eq.(13) and for any p E (0, 1) 1 a constant (3 , 3 for any k > 1 the relation Cos 0_j >_ (3, holds for at least [P_k] values of $j \in [1, k]$

Proof: We note that the symmetric matrices B_k = V_k Vk = Wk_ I C k_~ Wk 1 generated by the algorithm are positive definite, because $W_{k\_}$ I are nonsingular as a consequence of the Al-Bayati [2] update, and the $C_{k\_}$, are nonsingular by construction.

Using the definition (10) of yr , eq.(6) and lemma (3.1), we have

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

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

$$I_k = (i \in [1, n] : \|w_i\| \leq \sigma_k) \qquad (14)$$

And

$$J_k = (i \in [1, n] : \|w_i\| \geq \mu_k) \qquad (15)$$

Therefore by defining the scalar c in our new proposed algorithm

$$\psi(B_{k+1}) = \psi(W_k W_k^T) + \sum_{i\in I_k}\left[\left(\frac{\partial^2 k}{\|w_i\|^2} - 1\right)\|w_i\|^2 - \ln\frac{\sigma_k}{\|w_i\|^2}\right] + \sum\left[\left(\frac{\mu}{\|w_i\|^2} - 1\right)\|w_i\|^2 - \ln\frac{\mu_k}{\|w_i\|^2}\right]$$

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

We will now involve lemma (3.2) 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 $i \in J_k$ we can therefore apply eq.(8) to each term of the first summation, and eq.(9) to each term of the 2nd summation to obtain

$$\psi(B_{k+1}) \le \psi(w_k w_k^T) + n\delta_3 + n\delta_4 \qquad (16)$$

for the constants δ_3 and δ_4 given by lemma (3.2).

Now step (4) of our new suggested algorithm indicates that the matrix is $w_k w_k^T$ Al-Bayati's update of B_k. Therefore by the same procedure of [4] 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.

$$\sum_{k=1}^{\infty}\|x_k - x^*\| \le \infty \qquad (17)$$

and $f_{k+1} - f^* \le r^k(f_1 - f^*)$.(18)

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

To prove (18) let us start with

$f_{k+1} - f^* \le (1-\delta_5 \cos^2\theta_k)(f_k - f^*)$ (see [3] for the theoretical explanations).

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

$$f_{k+1} - f^* \le (1-\delta_5\beta_1^2)(f_k - f^*) \le r^k (f_k - f^*)$$

With $r = (1-\delta_5\beta_1^2) \in [0,1)$

The assumption on f also implies that $\frac{1}{2} m \|x_k - x^*\|^2 \le f_k - f^*$

(19) Therefore combining (19) with (18) we obtain

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

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

This proves the global convergence of our new proposed algorithm. #

4. Final Remarks

We have described in this paper the conditions under which a new automatic self-scaling algorithm based on the direct form of Al-Bayati's VM-Update [2] can be proven to be globally convergent. It should be noted that using extra theoretical results the super linear convergence of this new algorithm may be found. Also some sort of numerical experiments needs to inform the effectiveness of the new proposed algorithm. This will be certainly done in our next research paper.

It is also possible to describe another similar algorithm based on the inverse scaled-BFGS algorithm. Also a column-scaling algorithm which was proposed by Siegel [9] may be modified and implemented with this family of algorithms by Nocedal [8].

However, optimal values of 6k, ~Ik selected in the new algorithm may be described in our further work, but for this proposed algorithm let 6_k=0.5 and ~t,,=1. It might occasionally be better to increase 6_k and to decrease ~t_k . in any case, the theory developed in this paper will prove to be useful for analyzing the global convergence of the algorithm and it may be useful to prove the super linear convergent of the new proposed algorithm in the following research paper.

REFERENCES

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

[2] Al-Bayati, A.Y. (2001) " New generalized CG-methods fot the nonquadratic model in unconstrained optimizations " J. of Al-Yarmok, JORDAN, vol (10) PP. 1-17.

[3] Byrd, R. ; Nocedal, J. and Yuan, Y. (1987). "Global Convergence of a class of quasi-Newton methods on convex problems, SIAM J. Numerical analysis, 24, pp. 1171-1190.

[4] Byrd, R and Nocedal, J. (1989). " A tool for analysis of quasi-Newton methods with applications to unconstrained minimization, SIAM J. Numerical Analysis, 26, pp. 727-739.

[5] Griewank, A. (1991). " The global convergence of partitioned BFGS on problems with convex decompositions and Lipschitzian gradients " ; Math. Programming, 50, pp. 141-175.

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

[7] Nocedal, J and Morales,L (1997) "Auto matrix Preconditioning by limited memory QN-updates " TR OTC 97/08, Optimization Technology Center, Northwestren University.

[8] Nocedal, J. (1999). " An interior point algorithm for large scale nonlinear programming ", SAM J. optimization Vol (9), PP 877-900.

[9] 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.