A new self-scaling VM-algorithm for non-convex optimization, part 1

Prof Dr. Abbas Y. AL-Bayati
Department of mathematics,
College of Computers Sciences and Mathematics
University of Mosul, IRAQ
&
Dr. Maha S.Y. AL-Salih
Department of Computers Science / College of Education
University of Mosul, IRAQ

Received	Accepted
06 / 05 / 2010	02 / 03 / 2011

الملخص

خوارزمية المتري المتغير ذاتي القياس تحل مسائل الامثلية غير الخطية وغير المقيدة بواسطة مصفوفة هسي التقريبية وقبل هذا تحّ دث عند كل تكرار ، ولتجنب مقدار الخطوة الكبيرة لمصفوفات هسي التقريبية لدالة الهدف . قمنا ببرهان هذه الخوارزميات في حالة التقارب الشام ل والسرعة فوق الخطية عندما تكون الدالة غير محدبة.

في هذا البحث تم استحداث خوارزمية جديدة في الأمثلية غير الخطية وغير المقيدة في مجال المتري المتغير ذاتي القياس بأستخدام خط بحث خاص بالدوال غير المتزايدة (non-monotone line search) وتم دراسة خاصية التقارب فوق الخطي والتقارب الشامل للخوارزمية المقترحة في الامثلية غير المحدبة.

Abstract

The self-scaling VM-algorithms solves an unconstrained non-linear optimization problems by scaling the Hessian approximation matrix before it is updated at each iteration to avoid the possible large eigen-values in the Hessian approximation matrices of the objective function f(x).It has been proved that these algorithms have a global and super-linear convergences when f(x)is non- convex.

In this paper we are going to propose a new self-scaling VM-algorithm with a new non-monotone line search procedure with a detailed study of the global and super-linear convergence property for the new proposed algorithm in non-convex optimization.

Keywords: VM-methods, non-monotone line searches, self-scaling AL-Bayati VM- method, global converge, super-linear convergence.

1. Introduction

We study the global convergence of a self –scaling Al-Bayati [4], VM- method with non-monotone line searches for solving the unconstrained optimization problem (see[3]and[12])

$$\min_{x \in R^n} f(x) \qquad \ldots\ldots (1.1)$$

where f is a continuously differentiable function of n variables. At the k^{th} iteration of the self-scaling method, asymmetric and positive definite matrix B_k is given, and a search direction is computed by

$$d_k = -B_k^{-1} g_k \qquad \ldots\ldots (1.2)$$

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

$$x_{k+1} = x_k + \alpha_k d_k \qquad \ldots\ldots(1.3)$$

where the step size α_k satisfies the wolfe conditions

$$f(x_k + \alpha_k d_k) \le f(x_k) + \delta_1 \alpha_k d_k^T g_k \qquad \ldots\ldots(1.4)$$

and

$$g(x_k + \alpha_k d_k)^T d_k \ge \delta_2 d_k^T g_k \qquad \ldots\ldots(1.5)$$

Where $0 < \delta_1 < \frac{1}{2}$ and $\delta_1 < \delta_2 < 1$.

and satisfies another conditions as we explain it later.

Variable Metric Method

The theory of Variable Metric methods is beautiful. We now have a fairly good understanding of their properties. Much of this knowledge has been obtained recently, and we will discuss it in this paper.

The BFGS method is a line search method. At the k-th iteration, a symmetric and positive B_k is given and a search direction is computed by (1.2) and (1.3). It has been found that it is best to implement BFGS formula (Broyden (1970), Fletcher (1970), Goldfarb (1970) and Shanno (1970)), (see [1], [2]).

$$B_{k+1} = B_k - \frac{B_k v_k v_k^T B_k}{v_k^T B_k v_k} + \frac{y_k y_k^T}{y_k^T v_k} \qquad \ldots\ldots (1.6)$$

Where

$$y_k = g_{k+1} - g_k \text{ and } v_k = x_{k+1} - x_k \qquad \ldots\ldots (1.7)$$

The analysis has been extended by Byrd, et. al. (see [11]) to restricted Broyden class of Quasi-Newton method in (1.6) which is replaced by

$$B_{k+1} = B_k - \frac{B_k v_k v_k^T B_k}{v_k^T B_k v_k} + \frac{y_k y_k^T}{y_k^T v_k} + \phi(v_k^T B_k v_k) v_k v_k^T \qquad \ldots\ldots (1.8)$$

where $\phi \in [0, 1]$, and

$$v_k = \left[\frac{y_k}{y_k^T v_k} - \frac{B_k v_k}{v_k^T B_k v_k} \right]$$

The choice $\phi = 0$ gives rise to the BFGS update where as $\phi = 1$ defines the DFP method, the first Variable Metric method proposed by Davidon, Fletcher and Powell (see Fletcher, [13]). And then the Hessian approximation is then updated by Al-Bayati [4].

$$B_{k+1} = B_k + \frac{B_k v_k v_k^T B_k}{v_k^T B_k v_k} + \rho_k \frac{y_k y_k^T}{v_k^T y_k^T}$$

where

$\rho_k = \dfrac{v_k^T B_k v_k}{y_k^T v_k}$ (for more details see Al-Bayati [4])

Note that this is done only for a quadratic model. But for non quadratic models, see (Al-Bayati [5], Al-Bayati & Al-Assady [6] and Al-Bayati [7]). For the constraint optimization problems and scaled sequential BFGS algorithm see (AL-Bayati and Hammed [8]).

2. A new Non-monotones self-scaling Al-Bayati (1991) VM-algorithm

First we give the outline of the above algorithm as follows:

Algorithm (2.1)

At the k[th] iteration denote $f_k = f(x_k)$ and $g_k = g(x_k)$

Step (0) For given x_o and initial symmetric positive definite matrix H_o let k=0

Step (1) If $\|g_k\| = 0$ stop

Step (2) Determine the search direction

$$d_k = -B_k^{-1} g_k \qquad \text{……(2.1)}$$

Step (3) Find step length α_k by a new line search approach (NLS) below

Step (4) Generate a new iteration point by

$$x_{k+1} = x_k + \alpha_k d_k \qquad \text{……(2.2)}$$

Step (5) Update the Hessian matrix G_k by the following Al-Bayati [4] self-scaling

VM- update

$$B_{k+1} = B_k - \frac{B_k v_k v_k^T B_k}{v_k^T B_k v_k} + \rho_k \frac{y_k y_k^T}{v_k^T y_k} \qquad \text{……(2.3)}$$

$$v_k = x_{k+1} - x_k \qquad \text{…… (2.4)}$$

$$y_k = g_{k+1} - g_k \qquad \text{…… (2.5)}$$

$$\rho_k = \frac{v_k^T B_k v_k}{v_k^T y_k} \qquad \text{…… (2.6)}$$

Step (6) Let k=k+1 and go to step (1).

Now we produce a new non-monotone line search approach to determine the step- length in step (3). For this we introduce the concept of forcing function.

Definition 2.1: The mapping σ: [0, +∞]→[0, +∞] is called a forcing function, if for any sequence $\{t_i\}$ with $t_i \geq 0$ we have

$$\lim_{i\to\infty} \sigma(t_i) = 0 \Rightarrow \lim_{i\to\infty} t_i = 0 \quad \text{...... (2.7)}$$

2.3 A new non-monotone line search (NLS) approach

let σ_1, σ_2 be two forcing functions
let m be a positive integer
Given parameters $0<\delta_1<\beta<1$ and $M\geq 0$, $\delta_2>0$
At the iteration k, the step length α_k (for more details see[9]).
satisfies that:

$$f(x_k + \alpha_k d_k) \leq C_k - \delta_1 \min\{\sigma_1(\mu_k), \sigma_2(\zeta_k)\} - \delta_2 \alpha_k^2 \|d_k\|^2 \quad \text{......(2.8)}$$

$$g(x_k + \alpha_k d_k)^T d_k \geq \beta_k g_k^T d_k \quad \text{......(2.9)}$$

$$\mu_k = -\frac{g_k^T d_k}{\|d_k\|} \quad \zeta_k = -\alpha_k g_k^T d_k \quad \text{......(2.10)}$$

$C_o=f_o$, $Q_o=1$ and $C_{k+1} = \dfrac{\gamma_k Q_k C_k + f_{k+1}}{Q_{k+1}}$ ......(2.11)

with $\gamma_k \in [0, 1]$ and

$$Q_{k+1}=\gamma_k Q_k+1 \quad \text{......(2.12)}$$

The updating of C_k in (2.11) was recently given by Zhang and Hager [15]. Clearly from (2.11) and (2.12) C_k is a convex combination of the function values $f_o, f_1, \ldots, f_k$.

The choice of γ_k controls the degree of the non-monotonicity If $\gamma_k= 0$ and $\min\{\sigma_1, \sigma_2\}=\zeta_k$ for each k then (2.8) reduces to the monotone line search. If $\gamma_k=1$ for each k then C_k is the average of $f_o, f_1, \ldots, f_k$ and (2.8) reduces to the monotone non-increasing function.

Theorem 2.4:

Consider $x_{k+1}=x_k+\alpha_k d_k$, α_k satisfying (2.8)-(2.10) if $\{f_k\}$ in infinite and bounded below then

$$\sum_{k=0}^{\infty} \frac{\min\{\sigma_1(\mu_k), \sigma_2(\zeta)\} + \alpha^2 \|d_k\|^2}{Q_{k+1}} < \infty \quad \text{......(2.13)}$$

Proof:

From (2.8) it is clear that

$\alpha_k^2 \|d_k\|^2 \geq 0$ and since d is descent

then $\mu_k \geq 0$ and $\zeta_k \geq 0$

$$\Rightarrow \min\{\sigma_1, \sigma_2\} \geq 0 \quad \text{......(2.14)}$$

$$\Rightarrow f_{k+1} \leq C_k \text{ from (2.8)} \quad \text{......(2.15)}$$

and

$$C_{k+1} = \frac{\gamma_k Q_k C_k + f_{k+1}}{Q_{k+1}} \quad \ldots\ldots(2.16)$$

$$\leq \frac{\gamma_k Q_k C_k + \{C_k - \min\{\sigma_1, \sigma_2\} - \alpha_k^2 \|d_k\|^2\}}{Q_{k+1}}$$

$$\leq \frac{C_k\{\gamma_k Q_k + 1\}}{Q_{k+1}} - \frac{\min\{\sigma_1, \sigma_2\} + \alpha_k^2 \|d_k\|^2}{Q_{k+1}}$$

from (2.12) we get

$$\leq C_k - \frac{\min\{\sigma_1, \sigma_2\} + \alpha_k^2 \|d_k\|^2}{Q_{k+1}}$$

$\therefore \{C_k\}$ is monotone non-increasing since $\{f_k\}$ is bounded from (2.15) $\{f_k\}$ is non- increasing (see [10]).This implies

$$\sum \frac{\min\{\sigma_1, \sigma_2\} + \alpha_k^2 \|d_k\|^2}{Q_{k+1}} < \infty \quad \ldots\ldots(2.17)$$

3. The global convergence property of the new proposed algorithm (2.1)

Let us consider the following

Assumption 3.1

(1) for a given $x_1 \in R^n$

let the level set $S=\{x \in R^n: f(x) \leq f(x_1)\}$ be a bounded ……(3.1)

(2) In some neighborhood N(S) of S. g(x) satisfies Lipschitz continuous condition, i.e.

for all $x, \bar{x} \in N(S)$

$$\|g(x) - g(\bar{x})\| \leq L\|x - \bar{x}\| \quad \ldots\ldots(3.2)$$

Lemma 3.2

Under assumption 3.1 if α_k satisfies

$$g_{k+1}^T d_k \geq \beta g_k^T d_k \qquad 0<\beta<1 \quad \ldots\ldots(3.3)$$

then $\alpha_k \geq \frac{1-\beta}{L} \frac{\mu_k}{\|d_k\|}$.

Proof:

Clearly

$(g_{k+1}-g_k)^T d_k \leq \|g_{k+1}-g_k\| \|d_k\|$ ……(3.4)

from (3.2) $\leq L\|x - \bar{x}\|\|d_k\|$

$\leq L \alpha_k \|d_k\|^2$ ……(3.5)

Also from (3.3) we can write

$g_{k+1}^T d_k - g_k^T d_k \geq \beta g_k^T d_k - g_k^T d_k$.

$\Rightarrow (g_{k+1} - g_k)^T d_k \geq (\beta - 1) g_k^T d_k \geq 0$ ......(3.6)

since 0<β<1

from (3.5) and (3.6) we have

$$L\alpha_k \|d_k\|^2 \geq (\beta - 1) g_k^T d_k \qquad ......(3.7)$$

$$\Rightarrow \alpha_k \geq \frac{\beta - 1 \, g_k^T d_k}{L\|d_k\|^2} \qquad ......(3.8)$$

from (2.10) we have

$$\alpha_k \geq \frac{1-\beta}{L} \frac{\mu_k}{\|d_k\|} \qquad ......(3.9)$$

Now before starting the proof of the global convergence of the proposed algorithm, we define an auxiliary matrix sequence $\{\bar{B}_k\}$ as Nocedal and Yuan did in [14],

assume that $\overline{B_1} = B_1$ and

$$\bar{B}_{k+1} = \frac{v_k^T y_k}{v_k^T \bar{B} \, v_k} B_{k+1}, \ k \geq 1 \qquad ......(3.10)$$

from (2.1) we have

$v_k = -\overline{\alpha_k} \ \overline{B_k} \ g_k$

where

$$\overline{\alpha_k} = \frac{v_{k-1}^T y_{k-1}}{v_{k-1}^T \bar{B}_{k-1}^{-1} v_{k-1}} \alpha_k \qquad ......(3.11)$$

since ρ_k >0 then if $v_k^T y_k > 0 \Rightarrow \bar{B}_{k+1}$ *is a positive definite*

Lemma 3.3 : From [12]

For $\{\bar{B}_k\}$ defined in (3.10) we have

$$tr(\bar{B}_{k+1}) = tr(\bar{B}_k) - \frac{\|\bar{B}_k v_k\|^2}{v_k^T \bar{B}_k v_k} + \frac{y_k^T v_k}{v_k^T \bar{B}_k v_k} + \frac{y_k^T v_k}{v_k^T \bar{B}_k v_k} \frac{\|y_k\|^2}{y_k^T v_k}$$

$$\bar{B}_{k+1} = \bar{B}_k - \frac{\bar{B}_k v_k v_k^T \bar{B}_k}{v_k^T \bar{B}_k v_k} + \frac{y_k^T v_k}{v_k^T \bar{B}_k v_k} \cdot \frac{y_k y_k^T}{v_k^T y} \qquad ......(3.12)$$

Then for all k≥1

⇒

$$tr(\bar{B}_{k+1}) = tr(\bar{B}_k) - \frac{\|\bar{B}_k v_k\|^2}{v_k^T \bar{B}_k v_k} + \frac{y_k^T v_k}{v_k^T \bar{B}_k v_k} \left(1 + \frac{\|y_k\|^2}{y_k^T v_k}\right) \qquad (3.13)$$

Theorem 3.3

Let x_1 be a starting point for which assumption 3.1 holds, let $\{x_k\}$ be the sequence generated by the new proposed algorithm 2.1. If there exists a positive number k≥1 for which

$$\|y_k\| \le (1-\beta)\,\|g_k\| \qquad \text{……(3.14)}$$

for all k ≥ K then

$$\lim_{k\to\infty} \inf \|g_k\| = 0 \qquad \text{……(3.15)}$$

Proof:

By Contradiction

Assume that there exists a positive $c_1>0$ for which

$$\|g_k\| \ge c_1 \qquad \text{…… (3.16)}$$

For all k ≥ 1. Now since

$$v_k = -\bar{\alpha}_k \bar{B}_k^{-1} g_k \qquad \text{…… (3.17)}$$

$$\Rightarrow \bar{B}_k v_k = -\bar{\alpha}_k g_k \qquad \text{…… (3.18)}$$

Let us start with

$$v_k^T \bar{B}_k v_k = (-\bar{\alpha}_k \bar{B}_k^{-1} g_k)^T (-\bar{\alpha}_k g_k)$$

from (3.11), (3.17) and (3.18) we get

$$= (-\bar{\alpha}_k g_k)^T \left(-\frac{v_{k-1}^T y_{k-1}}{v_{k-1}^T \bar{B}_{k-1}^{-1} v_{k-1}} \alpha_k \bar{B}_k^{-1} g_k\right) \qquad \text{…… (3.19)}$$

From (3.10) take k+1=k and then (2.1) we get.

$$\text{…… (3.20)}$$

$$\Rightarrow v_k^T \bar{B}_k v_k = -\bar{\alpha}_k \alpha_k g_k^T d_k$$

Now from (3.6) we have

$$y_k^T d_k \ge (\beta-1) g_k^T d_k$$

$$\Rightarrow y_k^T v_k \ge (\beta-1)\alpha_k g_k^T d_k \qquad \text{…… (3.21)}$$

substitute (3.18), (3.20) and (3.21) into (3.13)

$$tr(\bar{B}_{k+1}) \le tr(\bar{B}_k) - \frac{(-\bar{\alpha}_k g_k)^2}{-\bar{\alpha}_k \alpha_k g_k^T d_k} + \frac{(\beta-1)\alpha_k g_k^T d_k}{-\bar{\alpha}_k \alpha_k g_k^T d_k} + \frac{(\beta-1)\alpha_k g_k^T d_k}{-\bar{\alpha}_k \alpha_k g_k^T d_k}\,\frac{\|y_k\|^2}{(\beta-1)\alpha_k g_k^T d_k}$$

$$tr(\bar{B}_{k+1}) \le tr(\bar{B}_k) + \frac{\bar{\alpha}_k^2 g_k^2}{\bar{\alpha}_k \alpha_k g_k^T d_k} + \frac{(\beta-1) g_k^T d_k}{\bar{\alpha}_k g_k^T d_k} + \frac{\|y_k\|^2}{-\bar{\alpha}_k \alpha_k g_k^T d_k}$$

from (3.14) we have

$$tr(\bar{B}_{k+1}) \le tr(B_k) + \frac{\bar{\alpha}_k \|g_k\|^2}{\alpha_k g_k^T d_k} + \frac{(1-\beta) g_k^T d_k}{\bar{\alpha}_k g_k^T d_k} + \frac{(1-\beta)^2 \|g_k\|^2}{\bar{\alpha}_k \alpha_k g_k^T d_k}$$

$$tr(\bar{B}_{k+1}) \le tr(\bar{B}_k) + \frac{\bar{\alpha}_k \|g_k\|^2}{\alpha_k g_k^T d_k} + \frac{(1-\beta)}{\bar{\alpha}_k}\left[1 - \frac{(1-\beta)\|g_k\|^2}{\alpha_k g_k^T d_k}\right] \qquad \text{……(3.22)}$$

since β<1 and $d_k^T g_k < 0$ clearly

$tr(\bar{B}_{k+1}) \le tr(\bar{B}_k)$ for all $k \ge K$

Thus ∃ a constant $C_2 > 0$ for which

$$tr(\bar{B}_k) \le C_2 \qquad \text{for all } k \qquad \text{…… (3.23)}$$

let $0 < \lambda_1 < \lambda_2 < \dots < \lambda_n$ be the n eigen values of $\bar{B}_k$ then by (3.23)

$$\lambda_n(\bar{B}_k) \le C_2 \qquad \text{…… (3.24)}$$

$$C_3 = \prod_{i=1}^{n} \lambda_i(\bar{B}_k) \le \lambda_1(\bar{B}_k) C_2^{n-1}$$

∴ ∃ a constant $C_4 > 0$ ∍

$$\lambda_1(\bar{B}_k) \ge C_4 \qquad \text{…… (3.25)}$$

i.e. $\lambda_1(\bar{B}_k)$ is the largest eigen value of the matrix $(\bar{B}_k)$

Now

$$\text{Let } \bar{d}_k = -\bar{B}_k^{-1} g_k \qquad \text{…… (3.26)}$$

take the norm

$$\|\bar{d}_k\| \le \frac{1}{\lambda_1 \bar{B}_k} \|g_k\| \qquad \text{by (3.25)}$$

$$\le \frac{1}{c_4} \|g_k\|, \qquad \text{…… (3.27)}$$

Now

$$-\bar{g}_k^T d_k = g_k^T \bar{B}_k^{-1} g_k \qquad \text{by (3.26)}$$

$$\ge \frac{1}{\lambda_n(\bar{B}_k)} \|g_k\|^2$$

$$\ge C_2 \|g_k\|^2 \qquad \text{…… (3.28)}$$

From the def. of μ_k and (2.10) we have

$$\mu_k = \frac{-g_k^T d_k}{\|d_k\|} = \frac{-g_k^T \bar{d}_k}{\|\bar{d}_k\|} \ge \frac{C_4}{C_2} \|g_k\| \ge \frac{C_1 C_4}{C_2} \ge 0$$

from (3.16)

Also

$$\zeta_k = -\alpha_k g_k^T d_k \text{ from (3.9)} \qquad \text{…… (3.29)}$$

$$\ge \frac{-g_k^T d_k}{\|d_k\|} \left(\frac{1-\beta}{L} \mu_k \right) \qquad \text{…… (3.30)}$$

from (2.10)

$$= \left(\frac{1-\beta}{L} \right) \mu_k^2 \qquad \text{…… (3.31)}$$

$$\ge \left(\frac{1-\beta}{L} \right) \left(\frac{C_1^2 C_4^2}{C_2^2} \right) \succ 0 \qquad \text{…… (3.32)}$$

Holds for any k …… (3.33)

since $\beta<1$ $L>0 \Rightarrow \zeta_k > 0$ $\forall_k$

$\Rightarrow \min\{\sigma_1(\mu_k), \sigma_2(\zeta_k)\} > C_5 > 1 \ \forall_k$ …… (3.34)

From $Q_{k+1} = 1+ \gamma_k Q_k$ from (2.12)

$$= 1 + \sum_{i=0}^{k} \prod_{m=0}^{i} \gamma_{k-m} \le k+2 \qquad \text{…… (3.35)}$$

$$\rightarrow \sum \frac{\min(\sigma_1, \sigma_2) + \alpha_k^2 \|d_k\|^2}{Q_{k+1}} \ge \sum_{k=0}^{\infty} \frac{C_5}{k+2} \rangle \infty \qquad \text{…… (3.36)}$$

which contradicts (2.17)
so we complete the proof

Remark: we can say that the above theorem will be true for the following case: depending on determinant of the matrix B_k
since

$$\det\left(B_{k+1}^{Al-Bayati}\right) = \rho_k \ \det\ (B_k^{BFGS}) \qquad \text{…… (3.37)}$$

$$\text{with } \rho_k = \min\left\{\frac{y_k^T B_k y_k}{y_k^T v_k}, 1\right\} \qquad \text{…… (3.48)}$$

similarly, be taking the auxiliary matrix $\overline{B}_k$ as

$$\overline{B}_1 = B_1,\ \overline{B}_{k+1} = \max\left\{\frac{y_k^T v_k}{v_k^T \overline{B}_k^T v_k}, 1\right\} B_{k+1} \quad ; \quad k \ge 1 \qquad \text{…… (3.49)}$$

we claim that (3.38) and (3.39) are superior on the actual BFGS update.

References

1) Al-Baali, M., (1998): Global and super-linear convergence of a restricted class of self-scaling methods with in exact line searches for convex functions. Comput. Optim. Appl.,Vol.9, pp.191-203.

2) Al-Baali, M., (1998): Numerical experience with a class of self-scaling quasi- Newton algorithms. JOTA, Vol.96, pp.533-553.

3) Al-Baali, M., (2001): On the behavior of a combined extra-updating/self-scaling BFGS method. J. Comput. Appl. Math., Vol.134, pp.269-281.

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

5) Al-Bayati, A. Y., (1993): Anew non-quadratic model for unconstrained nonlinear optimization, J. of Mu'tah, Jordan, Vol.8, pp.131-155.

6) Al-Bayati, A. Y. and Al-Assady, (1994): Minimization of extended quadratic functions with inexact line searches, JOTA, Vol.82, pp.139-147.

7) 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.

8) 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.

9) Al-Bayati, A. Y. and Al-Salih, M. S., (2007): Global and super-linear convergent VM-algorithms for nonlinear optimization, J of Kirkuk University-Scientific Studies, Vol.2, pp.88-106.

10) Birgin, E. G., Martinez, J. M. Raydan, M., (2000): Non-monotone spectral projected gradient methods on convex set. SIAM J.Optim.,Vol.10, pp.1196-1211.

11) 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.

12) Dai, Y. H., (2002): On the non-monotone line search. JOTA, Vol.112, pp.123-160.

13) Fletcher, R., (1987): Practical methods of optimization, John Wiley and sons, New York, Toronto and Singapore.

14) Nocedal, J., Yuan, Y., (1993): Analysis of a self-scaling quasi-Newton method. Math. Program, Vol.61, pp.19-37.

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

16) Yin & Du., (2007): The Global Convergence Of Self-Scaling BFGS Algorithm with Nonmontone Line Search for Unconstrained Non convex Optimization problems. ACTA Mathematica Sinica English Series pp 1232- 1240.