Matematika, 2002, Jilid 18, bil. 2, hlm. 117–127

Convergence of a Positive-Definite Scaled Symmetric Rank One Method

Malik b. Hj. Abu Hassan, Mansor b. Monsi & Leong Wah June
Department of Mathematics
University Putra Malaysia
43400 Serdang
Selangor, Malaysia

Abstract We discuss the convergence rate of a positive-definite scaled symmetric rank one (SSR1) method. This method is developed by Malik et al. (2002). In general, a restart procedure is derived and used together with the symmetric rank one (SR1) method. The restart procedure provides a replacement for the non-positive definite H_k with a positive multiple of the identity matrix. However if we choose the initial approximation for the inverse Hessian as an identity matrix, the sequences of steps produced by the SSR1 do not usually seem to have the "uniform linear independence" property that is assumed in some recent convergence analysis for SR1. Therefore, we present a new analysis that shows that the SSR1 method with a line search is $n+1$ step q-superlinearly convergent without the assumption of linearly independent iterates. This analysis only assumes that the Hessian approximations are positive-definite and bounded asymptotically, which are the actual conditions given by SSR1. Numerical experiments indicate that the SSR1 method is very competitive with the BFGS method and is easily implemented.

Keywords SSR1 method, BFGS method, q-superlinear rate of convergence.

Abstrak *Kami membincang kadar penumpuan bagi suatu kaedah terskala pangkat satu yang simetri (SSR1) dan tentu-positif. Kaedah tersebut dibangunkan oleh Malik et al. (2002). Secara umum, suatu prosidur mula-semula diterbitkan dan digunakan bersama dengan kaedah pangkat satu yang simetri (SR1). Prosidur mula-semula ini membekalkan suatu penggantian untuk* H_k *yang tidak tentu-positif dengan suatu gandaan positif bagi matriks identiti. Walau bagaimanapun, jika kita memilih hampiran awal bagi songsangan Hessian sebagai matriks identiti. jujukan langkah yang dihasilkan oleh SSR1 tidak semestinya mempunyai sifat "kemerdekaan linear seragam" yang diandaikan dalam beberapa analisis penumpuan yang terkini bagi SR1. Justeru itu, kami kemukakan suatu analisis baru yang menunjukkan kaedah SSR1 dengan gelintaran garis menumpu pada* $n+1$ *langkah* q*-superlinear tanpa andaian lelaran merdeka* *linear. Analisis tersebut hanya mengandaikan bahawa hampiran Hessian adalah tentu-positif dan terbatas secara asimptot, yang merupakan syarat sebenar yang dipaparkan oleh SSR1. Ujikaji berangka menunjukkan bahawa SSR1 setanding dengan kaedah BFGS dan mudah diimplementasikan.*

Katakunci *Kaedah SSR1, kaedah BFGS, kadar penumpuan q-superlinear.*

1 Introduction

We consider the quasi-Newton methods for finding a local minimum of the unconstrained optimization problem

$$\min_{x \in \Re^n} f(x) \tag{1}$$

with $f(x)$ assumed to be at least twice continuously differentiable.

The algorithms for solving (1) are iterative with the basic framework of an iteration of a secant method described as follows:

Given the current iteration x_c, $f(x_c)$, $\nabla f(x_c)$ or finite difference approximation, and $B_c \in \Re^{n \times n}$ symmetric (secant approximation to $\nabla^2 f(x_c)$), select the new iterate x_+ by a line search method. Update B_c to B_+ such that B_+ is symmetric and satisfies the secant equation $B_+ s_c = y_c$, where $s_c = x_+ - x_c$ and $y_c = \nabla f(x_+) - \nabla f(x_c)$.

In this paper, we consider the SR1 update for the Hessian approximation,

$$B_+ = B_c + \frac{(y_c - B_c s_c)(y_c - B_c s_c)^T}{s_c^T (y_c - B_c s_c)} \tag{2}$$

and the BFGS update

$$B_+ = B_c + \frac{y_c y_c^T}{y_c^T y_c} + \frac{B_c s_c s_c^T B_c}{s_c^T y_c}. \tag{3}$$

Throughout if $H = B^{-1}$, the inverse update respected to SR1 is given by

$$H_+ = H_c + \frac{(s_c - H_c y_c)(s_c - H_c y_c)^T}{y_c^T (s_c - H_c y_c)} \tag{4}$$

and, the inverse BFGS update is

$$H_+ = H_c + \left(1 + \frac{y_c^T H_c y_c}{s_c^T y_c}\right) \frac{s_c s_c^T}{s_c^T y_c} - \left(\frac{s_c y_c^T H_c + H_c y_c s_c^T}{s_c^T y_c}\right). \tag{5}$$

For the literature of these updates and others see Fletcher [10], Gill et al. [11], and Dennis and Schnabel [8].

The BFGS update has been the most commonly used secant update for many years. It makes a symmetric, rank two change to the previous Hessian approximation B_c, and if B_c is positive definite and $s_c^T y_c > 0$, then B_+ is positive definite.

The BFGS method has been shown by Broyden et al. [1] to be locally q-superlinearly convergent provided that the initial Hessian approximation is sufficiently accurate. Powell [14] proved a global superlinear convergence result for the BFGS method when applied to strictly convex functions and used in conjunction with line searches that satisfy the conditions of Wolfe [16] (eqs. 9-10). The BFGS update has been used successfully in many production codes for unconstrained optimization.

The SR1 formula, on the other hand, makes a symmetric rank one change to the previous Hessian approximation B_c. Compared with other secant updates, the SR1 update is simpler and may require less computation per iteration. A basic disadvantage to the SR1 update, however, is that the SR1 update may not preserve positive-definiteness even if this is positive, i.e., when B_c is positive-definite and $s_c^T y_c > 0$. A simple remedy by Malik et al. [12] to this problem (SSR1 method) is to replace the non-positive definite B_+ with a positive multiple of the identity matrix whenever this difficulty arises.

In the next section, we present the algorithm of the SSR1 method using standard line search for unconstrained optimization problems. We also report on situations that the assumption of uniform linear independence of the sequence of steps which is required by the theory of Conn et al. [5] may not be satisfied for many problems. Therefore in Section 3, we prove a new convergence result without the assumption of uniform linear independence steps. Instead, it requires the assumption of boundedness and positive-definiteness of the Hessian approximation. In Section 4, we compare the results from the SSR1 method with the widely used BFGS method.

2 SSR1 Algorithm

Algorithm 2.1 SSR1 method

Step 0. Given an initial point x_0, an initial positive matrix $H_0 = I$, set $k = 0$.

Step 1. If the convergence criterion

$$\| \nabla f(x_k) \| \leq \varepsilon \times \max(1, \| x_k \|) \tag{6}$$

is achieved, then stop.

Step 2. Compute a quasi-Newton direction

$$\begin{aligned} p_k &= -H_k \nabla f(x_k), \text{where } H_k \text{ is given by} \\ H_+ &= H_c + \frac{(s_c - H_c y_c)(s_c - H_c y_c)^T}{y_c^T (s_c - H_c y_c)}. \end{aligned} \tag{7}$$

Step 3. If $p_k^T \nabla f(x_k) > 0$, (H_k is not positive definite) or $k = 1$, set $H_k = \widetilde{\delta}_{k-1} I$,

$$\widetilde{\delta}_{k-1} = \frac{s_{k-1}^T s_{k-1}}{y_{k-1}^T s_{k-1}} - \left\{ \frac{\left(s_{k-1}^T s_{k-1}\right)^2}{\left(y_{k-1}^T s_{k-1}\right)^2} - \frac{s_{k-1}^T s_{k-1}}{y_{k-1}^T s_{k-1}} \right\}^{1/2} \tag{8}$$

and subsequently $p_k = -\widetilde{\delta}_{k-1} \nabla f(x_k)$. Else retain (7).

Step 4. Using a backtracking line search, find an acceptable steplength, λ_k such that the Wolfe's condition

$$f(x_k + \lambda_k p_k) \geq f(x_k) + \alpha \lambda_k \nabla f(x_k)^T p_k \tag{9}$$

and

$$\nabla f(x_k + \lambda_k p_k)^T p_k \leq \alpha' \nabla f(x_k)^T p_k \tag{10}$$

are satisfied. ($\lambda_k = 1$ is always tried first, $\alpha = 10^{-4}$ and $\alpha' = 0.9$)

Step 5. Set $x_{k+1} = x_k + \lambda_k p_k$.

Step 6. Compute the next inverse Hessian approximation H_{k+1} .

Step 7. Set $k = k + 1$, and go to Step 1.

Fiacco and McCormick [9] showed that if the SR1 update is applied to positive-definite quadratic function in a line search method, then, provided that the updates are all well defined, the solution is reached in at most $n + 1$ iterations. Furthermore, if $n + 1$ iterations are required, then the final Hessian approximation is the actual Hessian at the solution. This result is not true, in general, for BFGS update or other members of the Broyden family, unless exact line searches are used.

For non-quadratic functions, however, convergence of the SR1 is not as well understood as convergence analysis of the BFGS method. In fact, Broyden et al. [1] have shown that under their assumptions the SR1 update can be undefined, and thus their convergence analysis cannot be applied in this case. Also, no global convergence result similar to that for the BFGS method given by Powell [14] exists, so far, for the SR1 method when applied to a non-quadratic function.

Recent work by Conn et al. [3], [4], and [5] has sparked renewed interest in the SR1 update. Conn et al. [5] proved that the sequence of matrices generated by the SR1 formula converges to the actual Hessian at the solution $\nabla^2 f(x^*)$, provided that the steps taken are uniformly linearly independent, that the SR1 update denominator is always sufficiently different from zero, and that the iterates converge to finite limit (Using this result it is simple to prove that the rate of convergence is q-superlinear).

The condition of linear independence of the sequence $\{s_k\}$ under which Conn et al. [5] analyze the performance of the SR1 method may be too strong in practice. Therefore in this paper we consider the convergence rate of the SR1 method without this condition. We will show that if we drop the condition of uniform linear independence of $\{s_k\}$ but add instead the condition that the sequence $\{B_k\}$ remains positive definite and bounded, then the line search Algorithm 2.1 generate at least p q-superlinear steps out of every $n+p$ steps. This will enable us to prove that convergence is $2n$-step q-quadratic.

The basic idea behind our proof is that, if any step falls close enough to a subspace spanned $m \leq n$ recent steps, then the Hessian approximation must be quite accurate in this subspace. Thus, if in addition the step is the full secant step $-B_k^{-1}\nabla f(x_k)$, it should be a superlinear step. But in a line search method, for the step to be the full secant step, B_k must be positive definite, which accounts for the algorithm by Malik et al. [12]. In the following section we will give the convergence proof.

3 Convergence Rate of the SR1 without Uniform Linear Independence

Throughout this section the following assumptions will frequently be made:

Assumptions 1

A1. The function f has a local minimizer at a point x^* such that $\nabla^2 f(x^*)$ is positive definite, and its Hessian $\nabla^2 f(x)$ is Lipschitz continuous near x^*, that is, there exists a constant $\gamma > 0$ such that for all x, y in some neighborhood of x^*,

$$\| \nabla^2 f(x) - \nabla^2 f(y) \| \geq \gamma \| x - y \| .$$

A2. The sequence $\{x_k\}$ converges to local minimizer x^*.

We first state the following result, due to Conn et al. [5], which does not assume linear independence of the step directions.

Lemma 3.1 *Let $\{x_k\}$ be a sequence of iterates defined by $x_{k+1} = x_k + s_k$. Suppose that Assumptions 1 hold, that the sequence of matrices $\{B_k\}$ is generated from the SR1 updates, and that for each iteration*

$$|s_k^T (y_k - B_k s_k)| \geq r \| s_k \| \| y_k - B_k s_k \| \tag{11}$$

where $r \in (0, 1)$ is a constant. Then, for each j, $\| y_j - B_{j+1} s_j \| = 0$, and

$$\| y_j - B_i s_j \| \leq \frac{\gamma}{r} \left(\frac{2}{r} + 1 \right)^{i-j-2} \eta_{i,j} \| s_j \| \tag{12}$$

for all $i \geq j + 2$, where $\eta_{i,j} = \max\{\| x_q - x_t \| \, | j \leq t \leq q \leq i\}$, and γ is the Lipschitz constant from Assumptions 1.

Actually, it is apparent from the proof of Lemma 3.1 by Conn et al. [5] that, if the update is skipped whenever (11) is violated, then (12) still holds for all j for which (11) is true.

In the lemma below, proven by Byrd et al. [2] it is shown that if the sequence of steps generated by an iterative process using the SR1 update satisfies (11), and the sequence of matrices is bounded, then out of any set of $n + 1$ steps, at least one is very good. As in the previous lemma, condition (3.1) actually needs only hold at this set of $n + 1$ steps, as long as the update is not made when that condition fails.

Lemma 3.2 *Suppose the assumptions of Lemma 3.1 are satisfied for the sequences $\{x_k\}$ and $\{B_k\}$, and that in addition there exists M for which $\| B_k \| \leq M$ for all k. Then there exist $K \geq 0$ with $S = \{s_{k_j} : K \leq k_1 \leq \cdots \leq k_{n+1}\}$ and an index $k_m, m \in \{2, 3, \cdots, n+1\}$, such that*

$$\frac{\| (B_{k_m} - \nabla^2 f(x^*)) s_{k_m} \|}{\| s_{k_m} \|} < \bar{c} \varepsilon_s^{1/n}$$

where

$$\varepsilon_s = \max_{i \leq j \leq n+1} \{\| x_{k_j} - x^* \|\}$$

and

$$\bar{c} = 4 \left[\gamma + \sqrt{n} \frac{\gamma}{r} \left(\frac{2}{r} + 1 \right)^{k_{n+1}-k_1-2} + M + \| \nabla^2 f(x^*) \| \right].$$

In order to use this lemma to establish a rate of convergence we need the following result, which is closely related to the well-known superlinear convergence characterization of Dennis and Moré [6].

Lemma 3.3 *Suppose the function f satisfies Assumption 1. If the quantities*

$$e_k = \| x_k - x^* \| \quad \textit{and} \quad \frac{\| \left(B_k - \nabla^2 f(x^*) \right) s_k \|}{\| s_k \|}$$

are sufficiently small, and if $B_k s_k = -\nabla f(x_k)$, then

$$\| x_k + s_k - x^* \| \leq \| \left(\nabla^2 f(x^*) \right)^{-1} \| \left[2 \frac{\| \left(B_k - \nabla^2 f(x^*) \right) s_k \|}{\| s_k \|} e_k + \frac{\gamma}{2} e_k^2 \right].$$

Proof. By the definition of s_k

$$\nabla^2 f(x^*) s_k = \left(\nabla^2 f(x^*) - B_k \right) s_k - \nabla f(x_k)$$

so that

$$s_k = -(x_k - x^*) + \left(\nabla^2 f(x^*) \right)^{-1} \left[\left(\nabla^2 f(x^*) - B_k \right) s_k - \nabla f(x_k) + \nabla^2 f(x^*)(x_k - x^*) \right]. \quad (13)$$

Therefore, using Taylor's theorem and Assumptions 1,

$$\| x_k - x^* + s_k \| \leq \| \left(\nabla^2 f(x^*) \right)^{-1} \| \left[\| \left(\nabla^2 f(x^*) - B_k \right) s_k \| + \frac{\gamma}{2} e_k^2 \right]. \quad (14)$$

Now it follows from (13) that if $\| \left(\nabla^2 f(x^*) \right)^{-1} \| \dfrac{\| \left(B_k - \nabla^2 f(x^*) \right) s_k \|}{\| s_k \|} \leq \dfrac{1}{3}$, then by Taylor's theorem,

$$\| s_k \| \leq \frac{3}{2} \left[\| x_k - x^* \| + \| \left(\nabla^2 f(x^*) \right)^{-1} \| \frac{\gamma}{2} \| x_k - x^* \|^2 \right] \leq 2 \| x_k - x^* \|,$$

if e_k is sufficiently small. Using this inequality together with (13) gives the result. $\diamond$

Using these lemmas one can show that for any $p > n$, Algorithm 2.1 will generate at least $p - n$ superlinear steps for every p iterations provided that B_k is safely positive definite, which implies that H_k is not replaced by $\widetilde{\delta}_{k-1} I$ in Step 3. This results, which is contained in the following theorem, is proven and used to establish a rate of convergence for Algorithm 2.1 under the assumption that the sequence $\{B_k\}$ becomes, and stays, positive definite. In a corollary we also show that this implies that the rate of convergence for Algorithm 2.1 is $2n$-step q-quadratic. We are assuming here that if B_k is positive definite, then it is not replaced in Step 3, i.e., we are assuming that "safely positive definite" just means positive definite.

Theorem 3.1 *Consider Algorithm 2.1 and suppose that Assumptions 1 hold. Assume also that for all $k \geq 0$,*

$$\left| s_k^T \left(y_k - B_k s_k \right) \right| \geq r \| s_k \| \| y_k - B_k s_k \|,$$

for a fixed $r \in (0,1)$, and that $\exists M$ for which $\| B_k \| \geq M \; \forall k$. Then, if $\exists K_0$ such that B_k is positive definite for all $k \geq K_0$, then for any $p \geq n+1$ there exists K_1 such that for all $k \geq K_1$,

$$e_{k+p} \leq \alpha e_k^{p/n} \tag{15}$$

where α is a constant and e_j is defined as $\| x_j - x^* \|$.

Proof. Since $\nabla^2 f(x^*)$ is positive definite, there exists $K_1, \beta_1 > 0$, and $\beta_2 > 0$ such that

$$\beta_1 \left[f(x_k) - f(x^*)\right]^{1/2} \leq \| x_k - x^* \| \leq \beta_2 \left[f(x_k) - f(x^*)\right]^{1/2} \tag{16}$$

for all $k \geq K_1$. Therefore, since we have a descent method, for all $l > k > K_1$,

$$\| x_l - x^* \| \leq \frac{\beta_1}{\beta_2} \| x_k - x^* \| .$$

Now, given $k > K_1$ we apply Lemma 3.2 to the set $\{s_k, s_{k+1}, \cdots, s_{k+n}\}$. Thus, there exists $l_1 \in \{k+1, \cdots, k+n\}$ such that

$$\frac{\| \left(B_{l_1} - \nabla^2 f(x^*)\right) s_{l_1} \|}{\| s_{l_1} \|} < \overline{c} \left(\frac{\beta_2}{\beta_1} e_k\right)^{1/n} \tag{17}$$

(If there is more than one such index l_1, we choose the smallest.) Equation (17) implies that for $\| x_{l_1} - x^* \|$ sufficiently small, by Dennis and Moré's theorem [7], which states that if $\{x_k\}$ converges to a point x^* at which $\nabla^2 f(x_k)$ is positive definite and

$$\lim_{k \to \infty} \frac{\| \nabla f(x_i) + \nabla^2 f(x_k) p_k \|}{p_k} = 0,$$

then Algorithm 2.1 will choose $\lambda_{l_1} = 1$ so that $x_{l_1+1} = x_{l_1} + s_{l_1}$. This fact, together with Lemma 3.3 and (17), implies that if e_k is sufficiently small, than

$$e_{l_1+1} \leq \hat{\alpha} e_k^{1/n} e_{l_1} \tag{18}$$

for some constant $\hat{\alpha}$. Now we can apply Lemma 3.2 to the set

$$\{s_k, s_{k+1}, \cdots, s_{k+n}, s_{k+n+1}\} - \{s_{l_1}\}$$

to get l_2. Repeating this $n-p$ times we get a set of integers $l_1 < l_2 < \cdots < l_{p-n}$, with $l_1 > k$ and $l_{p-n} < k+p$ such that

$$e_{l_1+1} \leq \hat{\alpha} e_k^{1/n} e_{l_1} \tag{19}$$

for each l_1. Now letting $h_j = \left[f(x_j) - f(x^*)\right]^{1/2}$, since we have a descent method,

$$h_{j+1} \leq h_j \tag{20}$$

and using (16) we have that for the arbitrary $k \geq K_1$,

$$\begin{aligned} h_{l_i+1} &\leq \frac{1}{\beta_1} e_{l_i+1} \\ &\leq \frac{\hat{\alpha}}{\beta_1} e_k^{1/n} e_{l_i} \\ &\leq \frac{\hat{\alpha}\beta_2}{\beta_1} e_k^{1/n} h_{l_i} \end{aligned} \tag{21}$$

for $i = 1, 2, \cdots, p-n$. Therefore using (20) and (21) we have that

$$h_{k+p} \leq \left(\frac{\hat{\alpha}\beta_2}{\beta_1} e_k^{1/n} \right)^{p-n} h_k$$

which, by (16) implies that

$$e_{k+p} \leq \frac{\beta_2}{\beta_1} \left(\frac{\hat{\alpha}\beta_2}{\beta_1} e_k^{1/n} \right)^{p-n} e_k.$$

Therefore,

$$e_{k+p} \leq \hat{\alpha}^{p-n} \left(\frac{\beta_2}{\beta_1} \right)^{p-n+1} e_k^{p/n}$$

and (15) follows. ⋄

Corollary 3.1 *Under the assumptions of Theorem 3.1, the sequence $\{x_k\}$ generated by Algorithm 2.1 is $n+1$-step q-superlinear, i.e.,*

$$\frac{e_{k+n+1}}{e_k} \rightarrow 0,$$

and is $2n$-step q-quadratic, i.e.,

$$\lim_{k\rightarrow\infty} \sup \frac{e_{k+2n}}{e_k^2} \leq \infty.$$

Proof. Let $p = n+1$, and $p = 2n$ in Theorem 3.1. ⋄

Note that a $2n$-step q-quadratically convergent sequence has an r order of $\left(\sqrt{2}\right)^{1/n}$. Since the integer p in the theorem is arbitrary, an interesting, purely theoretical question is what value of p will prove the highest r-convergence order for the sequence. It is not hard to show that, by choosing p to be an integer close to en, the r order approaches $e^{1/en} \approx 1.44^{1/n}$ for n sufficiently large, and that this value is optimal for this technique of analysis.

4 Comparison of the SSR1 and BFGS Methods

Using the above outlined assumptions, we tested the SSR1 method and the BFGS method (TOMS 500 algorithm, Shanno and Phua, [15]) on a set of test problems selected from Moré et al. [13], i.e. Penalty function I, II, Rosenbrook function, Powell function, Wood function, Beale function and Trigonometric function.

First derivatives were calculated analytically. The stopping tolerance, ε used in both methods was 10^{-5}. The upper bound used on the number of functions was 999. The symbol 'EX' was given when the number of functions exceeded 999. As done in Conn et al. [4] and our convergence proofs, we skipped the SR1 update if

$$\left| s_k^T (y_k - B_k s_k) \right| < r \parallel s_k \parallel \parallel y_k - B_k s_k \parallel,$$

where $r = 10^{-2}$. All experiments were run using double precision arithmetic that has a machine epsilon of order 10^{-16}. Table 1 reports the performance of the SSR1 and BFGS methods.

Tabel 1 indicates that the SSR1 is very competitive with the widely used BFGS method.

Table 1: Comparison of the SSR1 and the BFGS methods

	SSR1		BFGS	
	n_I	n_f	n_I	n_f
Penalty I				
$n = 4$	31	44	26	65
$n = 20$	50	84	51	64
$n = 400$	61	83	58	77
Penalty II				
$n = 4$	28	34	30	35
$n = 20$	284	439	692	816
$n = 400$	EX	EX	EX	EX
Trigonometric				
$n = 4$	9	11	17	20
$n = 20$	41	55	44	50
$n = 400$	38	44	48	58
Rosenbrook				
$n = 4$	30	40	29	39
$n = 20$	37	42	33	41
$n = 400$	34	48	33	44
Powell				
$n = 4$	39	47	42	43
$n = 20$	35	50	36	37
$n = 400$	36	38	54	55
Wood				
$n = 4$	28	40	36	43
$n = 20$	27	36	31	44
$n = 400$	38	48	35	49
Beale				
$n = 4$	16	21	15	16
$n = 20$	17	27	15	17
$n = 400$	14	18	16	18

5 Conclusion

We have attempted, in this paper, to investigate theoretical and numerical aspects of the SSR1 formula for unconstrained optimization.

We tested the SSR1 method on a set of standard test problems from Moré et al. [13]. Our test results shows that on the set of problems we tried, the SSR1 method, generally requires somewhat fewer iterations than the BFGS method in a line search algorithm.

Under conditions that do not assume uniform linear independence of the generated steps, but do assume positive definiteness and boundedness of the Hessian approximations, we were able to prove $n+1$-step q-superlinear convergence, and $2n$ step quadratic convergence, of a line search SSR1 method.

References

[1] C.G. Broyden, J.E. Jr. Dennis, and J.J. Moré, *On the local and superlinear convergence of quasi-Newton methods*, Journal of the Institute of Mathematics and its Applications 12 (1973), 223-246.

[2] R. H. Byrd, H. Khalfan, and R. B. Schnabel, *A theoretical and experimental study of the symmetrix rank one update*, Technical Report CU-CS-489-90, University of Colorado at Boulder, Colorado, 1990.

[3] A. R. Conn, N.I.M. Gould, and Ph. Toint, *Global convergence of a class of trust region algorithms for optimization with simple bounds*, SIAM Journal of Numerical Analysis 25/ 2 (1988a), 433-460.

[4] A.R. Conn, N.I.M. Gould, Ph. and Toint, *Testing a class of methods for solving minimization problems with simple bounds on the variables*, Mathematics of Computation 50/182 (1988b), 399-430.

[5] A.R. Conn, N.I.M. Gould, Ph. and Toint, *Convergence of quasi-Newton matrices generated by the symmetric rank one update*, Mathematical Programming 48 (1991), 549-560.

[6] J.E. Dennis, and J.J. Moré, *A characterization of superlinear convergence and its application to quasi-Newton methods*, Mathematics of Computation 28 (1974), 549-560.

[7] J.E. Dennis, and J.J. Moré, *Quasi-Newton methods, motivation and theory*, SIAM Review 19 (1977), 46-89.

[8] J.E. Dennis, and R.B. Schnabel, *Numerical Methods for Nonlinear Equations and Unconstrained Optimization*, Prentice Hall, Englewood Cliffs, New Jersey, 1983.

[9] A.V. Fiacco, and G.P. McCormick, *Nonlinear Programming*, John Wiley & Sons, New York, 1968.

[10] R. Fletcher, *Practical Methods of Optimization*, John Wiley & Sons, New York, 1980.

[11] P.E. Gill, , W. Murray, and M.H. Wright, *Practical Optimization*, Academic Press, London, 1981.

[12] A.H. Malik, M. Mansor, and W.J. Leong, *Scaling symmetric rank one update for unconstrained optimization*, to appear in J. ANALISIS 9/1&2, 2002.

[13] J.J. Moré, B.S. Garbow, and K.E. Hillstrom, *Testing unconstrained optimization software*, TOMS 7 (1981), 17-41.

[14] M.J.D. Powell, *Convergence properties of a class of minimization algorithms*, in Nonlinear Programming 2, O. Mangasarian, R. Meyer, and S. Robinson, eds., Academic Press, New York (1975), 1-27.

[15] D.F. Shanno, and K.H. Phua, *Remark on algorithm 500: minimization of unconstrained multivariate functions*, ACM Transactions on Mathematical Software 6 (1980), 618-622.

[16] P. Wolfe, *Convergence conditions for ascent methods*, SIAM Review 11/2 (1969) 226-235.