

# Convergence of Symmetric Rank-one Method Based on Modified Quasi-Newton Equation

Farzin Modarres (Corresponding author)

Department of Mathematics, Faculty of Science, Universiti Putra Malaysia

43400 Serdang, Selangor, Malaysia

Tel: 60-1-7249-6327 E-mail: farzin.modarres@gmail.com

Malik Abu Hassan

Department of Mathematics, Faculty of Science Universiti Putra Malaysia

43400 Serdang, Selangor, Malaysia

Tel: 60-1-3390-6116 E-mail: malik@fsas.upm.edu.my

Wah June Leong

Department of Mathematics, Faculty of Science, Universiti Putra Malaysia

43400 Serdang, Selangor, Malaysia

Tel: 60-1-2226-6562 E-mail: leong@math.upm.edu.my

## Abstract

In this paper we investigate on convergence rate of a modified symmetric rank-one (SR1) method for unconstrained optimization problems. In general, the modified SR1 method incorporates a modified secant equation into the standard SR1 method. Also a restart procedure is applied to avoid the loss of positive definiteness and zero denominator. A remarkable feature of the modified SR1 method is that it possesses at most  $n + 1$ -step  $q$ -superlinearly convergent and  $2n$ -step quadratic convergent without uniformly independent assumptions of steps.

**Keywords:** Unconstrained minimization, Convergence, Symmetric rank-one update, Secant equation, Hessian approximation.

## 1. Introduction

We consider the unconstrained optimization problem

$$\min_{x \in \mathbb{R}^n} f(x) \quad (1)$$

where  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is a smooth function whose gradient at point  $x_k$  is  $\nabla f(x_k)$  or  $g_k$ , for the sake of simplicity. We assume that  $f$  is continuous and at least twice differentiable. Among various iterative methods for solving eq.(1), quasi-Newton methods constitute an important class. These methods have been developed based on Newton's method, in which the Hessian matrix of  $f$  at  $x_k$ ,  $\nabla^2 f(x_k)$  is substituted by some matrix  $B_k$  to avoid the calculation of a Hessian matrix. The QN method for solving eq.(1) takes the following iterative process.

Given the  $k^{\text{th}}$  iterate  $x_k$  and the gradient of the function at  $x_k$ ,  $\nabla f(x_k)$ , we determine the QN direction  $d_k$  by  $B_k d_k + \nabla f(x_k) = 0$  where  $B_k$  is an secant approximation to  $\nabla^2 f(x)$ . Once  $d_k$  is obtained, the next iterate  $x_{k+1}$  is generated by  $x_{k+1} = x_k + \alpha_k d_k$ . Update  $B_k$  to  $B_{k+1}$  such that  $B_{k+1}$  satisfies in the following secant equation

$$B_{k+1} s_k = y_k$$

where  $s_k = x_{k+1} - x_k$  and  $y_k = g_{k+1} - g_k$ . However the general secant equation employ only the gradient information and ignores function information. Therefore Wei et al. (Wei *et al.*, (2006)) proposed the modified secant equation

$$B_{k+1} s_k = y_k^*, \quad (2)$$

where

$$y_k^* = y_k + A_k s_k, \quad (3)$$

and  $A_k$  is a simple symmetric and positive definite matrix.

In order to use both function and gradient information in the secant equation, they proposed a modified BFGS-type method for the solution of eq.(1) and the superlinear convergence of the BFGS-type algorithm was proved. On the other hand

Conn et al. (Conn *et al.*, (1991)) and Khalfan et al. (Khalfan *et al.*, (1993)) have analyzed the computational and numerical results of the SR1 methods and the numerical results showed that SR1 is a competitive formula among the QN method. Motivated by this Modarres et al. (Farzin *et al.*, (2009)) presented a modification to the secant equation of Wei et al. and employed this modification to the SR1 update. They used a restart procedure to preserve positive definiteness and to avoid the unbounded updates in the modified SR1 update. The global convergence for this method has been proved. Convergence rate of SR1 method has been studied by Conn et al. and proved that the rate of convergence is  $q$ -superlinear by the assumption of the uniformly linearly independent steps. This condition may be too strong in practice. Therefore Khalfan et al. made the weaker assumptions, in which the matrices are positive definite and uniformly bounded. They showed that the convergence rate for the standard SR1 update is at most  $n + 1$ -step  $q$ -superlinear and  $2n$ -step  $q$ -quadratic. Hence, it seems possible to extend the similar results of Khalfan et al. to the modified SR1 method. In the next section, we present the algorithm of the modified SR1 method. Finally we concentrate our attention on the proof of the superlinear convergence of the modified SR1 method.

## 2. Description of algorithm

### Modified SR1 Algorithm (+MSR1)

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

*Step 1.* If the convergence criterion  $\|\nabla f(x_k)\| \leq \varepsilon \times \max(1, \|x_k\|)$  is achieved, then stop.

*Step 2.* Compute a quasi-Newton direction by  $d_k = -H_k \nabla f(x_k)$ .

*Step 3.* Find an acceptable steplength such that the Wolfe conditions

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta_1 \alpha_k \nabla f(x_k)^T d_k, \quad (4)$$

$$\nabla f(x_k + \alpha_k d_k)^T d_k \geq \delta_2 \nabla f(x_k)^T d_k, \quad (5)$$

where  $0 < \delta_1 < \delta_2 < 1$ ,  $\delta_1 < \frac{1}{2}$ , are satisfied.

( $\alpha_k = 1$  is always tried first,  $\delta_1 = 10^{-4}$ ,  $\delta_2 = 0.9$ ).

*Step 4.* Set  $x_{k+1} = x_k + \alpha_k d_k$ .

*Step 5.* If

$$s_k^T y_k - y_k^T H_k y_k < 0, \quad (H_k \text{ might not be positive definite}) \quad (6)$$

or

$$|\tilde{y}_k^T (s_k - H_k \tilde{y}_k)| < r \|\tilde{y}_k\| \|s_k - H_k \tilde{y}_k\|, \quad (7)$$

where  $r \in (0, 1)$ , (denominator in  $H_k$  is sufficiently close to zero)

or

$$\|H_k\|_\infty > L, \quad (\text{where } L \text{ is a preset constant}), \quad (8)$$

set  $H_{k+1} = \tilde{\lambda}_k I$ , where  $\tilde{\lambda}_k$  is given by

$$\tilde{\lambda}_k = \frac{s_k^T s_k}{\tilde{y}_k^T s_k} - \left\{ \frac{(s_k^T s_k)^2}{(\tilde{y}_k^T s_k)^2} - \frac{s_k^T s_k}{\tilde{y}_k^T \tilde{y}_k} \right\}^{1/2}, \quad (9)$$

and subsequently  $d_{k+1} = -\tilde{\lambda}_k \nabla f(x_k)$ .

*Step 6.* Calculate  $\tilde{y}_k$  by using the following equation

$$\tilde{y}_k = y_k + \text{sgn}(\psi_k) \frac{\psi_k}{s_k^T s_k} s_k,$$

where  $\psi_k = 2(f(x_k) - f(x_{k+1})) + (\nabla f(x_{k+1}) + \nabla f(x_k))^T s_k$ .

*Step 7.* Compute the next inverse Hessian approximation  $H_{k+1}$  as follows

$$H_{k+1} = H_k + \frac{(s_k - H_k \tilde{y}_k)(s_k - H_k \tilde{y}_k)^T}{(s_k - H_k \tilde{y}_k)^T \tilde{y}_k}. \quad (10)$$

Step 8. Set  $k = k + 1$ , and go to step 1.

**Remark 1** Note that the +MSR1 method preserves positive definiteness, however in the case  $B_k$  is non-positive definite, by replacing  $\tilde{\lambda}_k I$  we can preserve positive definiteness of Hessian matrix.

**Remark 2** The scaling factor  $\tilde{\lambda}_{k-1}$  is derived in such a way to improve condition of modified SR1 method while preserving positive definiteness.

### 3. Convergence Rate of +MSR1 Algorithm

In this section we show that the modified SR1 update generated by +MSR1 Algorithm is at most  $n + 1$ -step  $q$ -superlinear convergence, and  $2n$ -quadratic convergence.

To give the convergence results, the following assumptions are given:

- (i) The sequence of iterates  $\{x_k\}$  remains in a closed bounded convex set  $D$ .
- (ii) The function  $f$  has an unique minimizer at a point  $x_*$  such that its Hessian  $\nabla^2 f(x_*)$  is positive definite, and  $\nabla^2 f(x)$  is Lipschitz continuous near  $x_*$ , that is, there exists a constant  $\tau > 0$  such that for all  $x, y$  in some neighborhood of  $x_*$ ,  $\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq \tau \|x - y\|$ .
- (iii) The sequence  $\{x_k\}$  converges to  $x_*$ .

Since modified SR1 method always generates positive definite updates, then for a strongly convex objective function, a line search implementation with Wolfe conditions will ensure that Assumption (iii) holds.

We first extend Lemma 1 from (Conn *et al.*, (1991)) to the modified SR1 update, which does not assume linear independence of the step directions.

**Lemma 1** Let  $\{x_k\}$  be a sequence of iterates generated by the +MSR1 Algorithm. Suppose that Assumptions (i)-(iii) holds, Assume, furthermore, for each iteration eq.(7) holds. Then, for each  $j$ ,

$$\|\tilde{y}_j - B_{j+1} s_j\| = 0, \quad (11)$$

and

$$\|\tilde{y}_j - B_i s_j\| \leq \left(\frac{\tau}{r} \left(\frac{2}{r} + 1\right)^{i-j-2} \eta_{i,j}\right) \|s_j\|, \quad (12)$$

for all  $i \geq j + 2$ , where

$$\eta_{i,j} = \max \left\{ \|x_q - x_r\| : j \leq t \leq q \leq i \right\}, \quad (13)$$

and  $\tau$  is the Lipschitz constant from Assumption (ii).

**Proof 1** It is obvious that eq.(11) and eq.(12) with  $i = j + 1$  are immediate result of  $B_{k+1} s_k = \tilde{y}_k$ . Now we proof eq.(12) by induction. First of all we choose  $k \geq j + 1$  and suppose that eq.(12) holds for every  $i = j + 1, \dots, k$ . Now consider

$$\left| (\tilde{y}_k - B_k s_k)^T s_j \right| = \left| \tilde{y}_k^T s_j - s_k^T B_k s_j \right| + \left( \frac{\tau}{r} \left( \frac{2}{r} + 1 \right)^{k-j-2} \eta_{k,j} \right) \|s_j\| \|s_k\| \quad (14)$$

In the last inequality, we used inductive assumption and Cauchy-Schwarz inequality.

Using the mean value theorem, we obtain that, for all  $l$ ,

$$H_l s_l = \tilde{y}_l, \quad (15)$$

where  $H_l = \int_0^1 H(x_l + t s_l) dt$ . Then from eq.(14) we have,

$$\begin{aligned} \left| (\tilde{y}_k - B_k s_k)^T s_j \right| &\leq \left| s_k^T (H_k - H_j) s_j \right| + \left( \frac{\tau}{r} \left( \frac{2}{r} + 1 \right)^{k-j-2} \eta_{k,j} \right) \|s_j\| \|s_k\| \\ &\leq (\tau \eta_{k+1,j} + \frac{\tau}{r} \left( \frac{2}{r} + 1 \right)^{k-j-2} \eta_{k,j}) \|s_j\| \|s_k\|. \end{aligned} \quad (16)$$

Furthermore we know from the triangle inequality and eq.(7) that

$$\|\tilde{y}_j - B_{k+1} s_j\| = \left\| \tilde{y}_j - B_k s_j - \frac{(\tilde{y}_k - B_k s_k)(\tilde{y}_k - B_k s_k)^T s_j}{(\tilde{y}_k - B_k s_k)^T s_k} \right\|$$

$$\leq \|\tilde{y}_j - B_k s_j\| + \frac{\|\tilde{y}_k - B_k s_k\| \left| (\tilde{y}_k - B_k s_k)^T s_j \right|}{r \|\tilde{y}_k - B_k s_k\| \|s_k\|}, \tag{17}$$

Hence if we use induction assumption and eq.(16) in eq.(17) we have

$$\begin{aligned} \|\tilde{y}_j - B_{k+1} s_j\| &\leq \left(\frac{\tau}{r} \left(\frac{2}{r} + 1\right)^{k-j-2} \eta_{k,j}\right) \|s_j\| + \frac{\tau}{r} \eta_{k+1,j} \|s_j\| \\ &+ \left(\frac{\tau}{r^2} \left(\frac{2}{r} + 1\right)^{k-j-2} \eta_{k,j}\right) \|s_j\|, \end{aligned} \tag{18}$$

If we put  $i = k + 1$  in eq.(18), when one takes into account the fact that  $r \in (0, 1)$  and use the inequality  $\eta_{k,j} \leq \eta_{k+1,j}$  we can deduce eq.(12).  $\square$

We then have the following useful two Lemmas by using the Lemma 1, which state that by using +MSR1 Algorithm, there is at least  $p - n$  superlinear steps for each  $p > n$  iterations.

**Lemma 2** Let  $\{x_k\}$  be the sequences of iterates generated by +MSR1 Algorithm and  $B_k$  be the corresponding Hessian approximation, also suppose the assumptions of Lemma 1 are satisfied for the sequences and that in addition there exists  $M$  for which  $\|B_k\| \leq M$ , for all  $k$ . Then there exist  $K \geq 0$  with  $S = \{s_{k_i} : K \leq k_1 \leq \dots \leq k_{n+1}\}$  and an index  $k_m, m \in \{2, 3, \dots, n + 1\}$  such that

$$\frac{\|(B_{k_m} - f(x_*)s_{k_m})\|}{\|s_{k_m}\|} < \bar{\rho} \varepsilon_M^{1/n} \tag{19}$$

where

$$\varepsilon_M = \max_{1 \leq j \leq n+1} \{\|x_j - x_*\|\} \tag{20}$$

and

$$\bar{\rho} = 4 \left[ \tau + \sqrt{n} \frac{\tau}{r} \left(\frac{2}{r} + 1\right)^{k_{n+1}+k_1-2} + M + \|\nabla^2 f(x_*)\| \right] \tag{21}$$

**Proof 2** The proof is similar to the proof in the Lemma 1 of (Khalfan et al., (1993)).  $\square$

For our final result, Theorem 1 below, which establish convergence rate of the modified SR1 method (+MSR1), we first establish the following Lemma which is closely related to the well-known superlinear convergence characterization of Dennis and Moré (Dennis and Moré, (1974)).

**Lemma 3** Suppose the Assumptions ((i)-(iii)) hold for objective function  $f$ . If the following quantities in the +MSR1 method  $e_k = \|x_k - x_*\|$  and  $\frac{\|(B_k - f(x_*)s_k)\|}{\|s_k\|}$  are sufficiently small, and if  $B_k s_k = -\nabla f(x_k)$ , then

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

**Proof 3** See Khalfan et al. (Khalfan et al., (1993)).  $\square$

We are now ready to prove our final result.

**Theorem 1** Consider the sequence  $\{x_k\}$  generated by the +MSR1 Algorithm and suppose that Assumptions (i)-(iii) holds. If there 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  $k \geq K_1$ ,

$$\|x_{k+p} - x_*\| \leq \alpha \|x_k - x_*\|^{p/n}, \tag{22}$$

where  $\alpha$  is a constant.

**Proof 4** Since  $\nabla^2 f(x_*)$  is positive definite, there exists  $K_1, \mathfrak{h}_1 > 0$  and  $\mathfrak{h}_2 > 0$  such that

$$\mathfrak{h}_1 (f(x_k) - f(x_*))^{1/2} \leq \|x_k - x_*\| \leq \mathfrak{h}_2 (f(x_k) - f(x_*))^{1/2} \tag{23}$$

for all  $k \geq K_1$ . Therefore as +MSR1 method is a descent method, for all  $t > k > K_1$ , we have  $\|x_t - x_*\| \leq \frac{1}{\mathfrak{h}_1} \|x_k - x_*\|$ .

Applying Lemma 2 to the set  $\{s_k, s_{k+1}, \dots, s_{k+n}\}$ , there exist  $t_1 \in \{k + 1, \dots, k + n\}$  such that

$$\frac{\|(B_{t_1} - \nabla^2 f(x_*)s_t)\|}{\|s_{t_1}\|} < \bar{\rho} \left(\frac{t_2}{t_1}\right)^{1/n} \quad (24)$$

Eq.(24) implies that for sufficiently small  $e_1$  and by Lemma 3, we can choose a steplength  $\alpha_{t_1}$  in +MSR1 Algorithm so that  $x_{t_1+1} = x_{t_1} + s_{t_1}$ .

On the other hand using this fact, together with Lemma 3 and eq.(24) implies that if  $e_k$  is sufficiently small then

$$e_{t_1+1} \leq \hat{\alpha} e_k^{1/n} e_{t_1}, \quad (25)$$

for some constants  $\hat{\alpha}$ .

We may also apply Lemma 2 to the set  $\{s_k, s_{k+1}, \dots, s_{k+n}, s_{k+n+1}\} - \{s_{t_1}\}$  to get  $t_2$ . Hence, by repeating this step for  $n - p$  times we get a set of integers  $t_1 < t_2 < \dots < t_{p-n}$  with  $t_1 > k$  and  $t_{p-n} < k + p$  such that

$$e_{t_i+1} \leq \hat{\alpha} e_k^{1/n} e_{t_i}, \quad (26)$$

for each  $t_i$ . Since we have a descent method, it follows that

$$(f(x_{j+1}) - f(x_*))^{1/2} < (f(x_j) - f(x_*))^{1/2}. \quad (27)$$

Using eq.(23) we have that for an arbitrary  $k \geq K_1$ ,

$$(f(x_{t_i+1}) - f(x_*))^{1/2} \leq \frac{1}{t_1} e_{t_i+1} \leq \frac{\hat{\alpha}}{t_1} e_k^{1/n} e_{t_i} \leq \frac{\hat{\alpha} t_2}{t_1} e_k^{1/n} (f(x_{t_i+1}) - f(x_*))^{1/2}, \quad (28)$$

for  $i = 1, 2, \dots, p - n$ . Therefore using eq.(26) and eq.(27) we obtain

$$(f(x_{k+p+1}) - f(x_*))^{1/2} \leq \left(\frac{\hat{\alpha} t_2}{t_1} e_k^{1/n}\right)^{p-n} (f(x_{k+1}) - f(x_*))^{1/2}$$

which by eq.(23) implies that

$$e_{k+p} \leq \left(\frac{t_2}{t_1} e_k^{1/n}\right)^{p-n} e_k.$$

Thus,

$$e_{k+p} \leq \hat{\alpha}^{p-n} \left(\frac{t_2}{t_1}\right)^{p-n} e_k^{p/n}.$$

and the inequality eq.(22) holds.  $\square$

Note that Theorem 1 only requires positive definiteness at the  $p - n$  out of  $p$  "good iterations" (which is, steps where  $f$  is reduced).

Finally, we give the rate of convergence for the +MSR1 algorithm:

**Corollary 1** Under the assumptions of Theorem 1, the sequence  $\{x_k\}$  generated by the +SSR1 algorithm is at most  $n+1$ -step  $q$ -superlinear, i.e.,

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

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

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

**Proof 5** The results follows by setting  $p = n + 1$ , and  $p = 2n$ , respectively in Theorem 1.

**Remark 3** By choosing  $p = n + 1$  in Corollary 1, we need just the positive definiteness condition just at only 1 step out of  $n + 1$  "good steps". Hence, for every  $n + 1$  steps greater than  $k$ , we will have at least 1 good step (which is, where  $B_k$  (or  $H_k$ ) is positive definite and bounded).

#### 4. Conclusion

In summary, we have considered the convergence properties of the modified SR1 method based on modified secant equation of Wei et al. An important feature of the proposed method is that it preserves positive definiteness of the updates.

Specifically by assuming positive definiteness and boundedness of Hessian approximations for modified SR1, we have proved that the modified SR1 is at most  $n + 1$ -step  $q$ -superlinear, without requiring an extra assumption such as uniformly linearly independent steps.

## References

- Conn A. R., Gould, N. I. M., & Toint, Ph. L. (1991). Convergence of quasi-Newton matrices generated by the symmetric rank one update. *Math. Program*, **48**, 549-560.
- Dennis, J. E., & Moré, J. J. (1974). A characterization of superlinear convergence and its application to quasi-Newton methods. *Math. Comp.*, **28** (5) 549-560.
- Khalfan, H., Byrd, R.H., & Schnabel, R.B. (1993). A theoretical and experimental study of the symmetric rank one update. *SIAM J. Optim*, **3**, 1-24.
- Leong, W.J., & Hassan, M.A. (2009). A restarting approach for the symmetric rank one update for unconstrained optimization. *Comp. optim. Appl.*, **42** (3), 327-334.
- Modarres, F., Hassan, M. A., & Leong, W. J. (2009). Improved Hessian approximation with modified quasi-Newton equations for symmetric rank-one method. *submitted to J. Comput. Appl. Math.*,
- Osborne, M.R., & Sun, L. (1999). A new approach to symmetric rank-one updating. *IMA J. Numer. Anal.*, **19**, 497-507.
- Spellucci, P. (2001). A modified rank one update which converges Q-superlinearly. *Comput. Optim. Appl.*, **19**, 273-296.
- Sun, L.P. (2004). Updating the self-scaling symmetric rank one algorithm with limited memory for large-scale unconstrained optimization. *Comp. Optim. Appl.*, **27** (1), 23-29.
- Wei, Z., Li, G., & Qi, L. (2006). New quasi-Newton methods for unconstrained optimization problems. *J. Appl. Math Comput.*, **175** Theory Appl 1156-1188.
- Wolkowicz, H. (1994). Measure for symmetric rank-one updates. *Math. Oper. Res.*, **19** (4), 815-830.
- Yabe, H., Ogasawara, H., & Yoshino, M. (2007). Local and superlinear convergence of quasi-Newton methods based on modified secant conditions. *J. Comput. Appl. Math*, **205** 617-632.