

Memoryless Modified Symmetric Rank-One Method for Large-Scale Unconstrained Optimization

Farzin Modarres, Malik Abu Hassan and Wah June Leong
Department of Mathematics, Faculty of Science,
University Putra Malaysia, 43400 Serdang, Selangor, Malaysia

Abstract: Problem statement: Memoryless QN methods have been regarded effective techniques for solving large-scale problems that can be considered as one step limited memory QN methods. In this study, we present a scaled memoryless modified Symmetric Rank-One (SR1) algorithm and investigate the numerical performance of the proposed algorithm for solving large-scale unconstrained optimization problems. **Approach:** The basic idea is to apply the modified Quasi-Newton (QN) equations, which uses both the gradients and the function values in two successive points in the frame of the scaled memoryless SR1 update, in which the modified SR1 update is reset, at every iteration, to the positive multiple of the identity matrix. The scaling of the identity is chosen such that the positive definiteness of the memoryless modified SR1 update is preserved. **Results:** Under some suitable conditions, the global convergence and rate of convergence are established. Computational results, for a test set consisting of 73 unconstrained optimization problems, show that the proposed algorithm is very encouraging. **Conclusion/Recommendations:** In this study a memoryless QN method developed for solving large-scale unconstrained optimization problems, in which the SR1 update based on the modified QN equation have applied. An important feature of the proposed method is that it preserves positive definiteness of the updates. The presented method owns global and R-linear convergence. Numerical results showed that the proposed method is encouraging comparing with the methods MMBFGS and FRCG.

Key words: Large-scale unconstrained optimization, symmetric rank-one method, memoryless method, modified quasi-Newton equation

INTRODUCTION

We consider the following unconstrained optimization problem:

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

where, $f : \mathfrak{R}^n \rightarrow \mathfrak{R}$ is a twice continuously differentiable nonlinear function and n , the number of variables, is large. There are various iterative methods for solving problem (1); Quasi-Newton (QN) methods are one of the most exploited methods.

The following iterative method can be seen as the general QN procedure to solve the problem (1).

Calculate the search direction p_k by solving the equation $p_k = -B_k^{-1}g_k$ then set $x_{k+1} = x_k + \alpha_k p_k$ at the k th iteration, where g_k denotes the gradient vector of f at x_k , B_k is the secant approximation to the $\nabla^2 f(x_k)$ and α_k is the step length that is updated by line search. The

matrix B_k is usually required to be positive definite to ensure a descent direction for f . B_k is updated at every iteration to a new Hessian approximation B_{k+1} for which the general QN equation:

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

Where:

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

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

is satisfied.

We are interested in elaborating an algorithm for solving large-scale problems for which we could use both function and gradient values to possess more accurate information. Memoryless QN methods were firstly introduced by Perry^[10] and Shanno^[12]. They can be considered as the QN methods for which the approximation to the inverse of the Hessian is taken as the identity matrix at every iteration. Limited memory

Corresponding Author: Farzin Modarres, Department of Mathematics, Faculty of Science, University Putra Malaysia, 43400 Serdang, Selangor, Malaysia Tel: +60172496327 Fax: +60389437958

BFGS (LBFGS) method^[9] and conjugate gradient method are two important classes of methods to solve large-scale unconstrained optimization problems. Wei *et al.*^[13] proposed the modified QN equation:

$$B_{k+1}s_k = y_k^* \tag{2}$$

Where:

$$y_k^* = y_k + A_k s_k \tag{3}$$

and A_k is a simple symmetric and positive definite matrix. The modified QN equation uses not only the gradient but also function value information in order to get the higher order accuracy in approximating the curvature of the objective function.

In this study, we consider the famous SR1 method:

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{s_k^T (y_k - B_k s_k)}$$

which makes a rank-one modification to the previous Hessian approximation B_k and hence it is a simpler update to use and it requires less computational effort per iteration. Conn *et al.*^[4] and Khalfan *et al.*^[7] have investigated the computational and numerical results of the SR1 methods. The results showed that in practice when the SR1 update solves a given problem, its efficiency is at least, if not better, as good as other QN methods. With these encouragement, it seems reasonable to extend the modified QN equations of Wei *et al.*^[13] to the memoryless method and obtain the memoryless modified SR1 update.

MATERIALS AND METHODS

Modified QN equation: We begin by reviewing the modified QN equation. Li and Fukushima^[8] proposed the modified QN equation:

$$B_{k+1}s_k = y_k^* \tag{4}$$

where, $y_k^* = y_k + t_k \|g_k\| s_k$, with $t_k = 1 + \max \left\{ -\frac{s_k^T y_k}{\|s_k\|^2}, 0 \right\}$.

The modified QN equation takes advantage of not only the gradients but also the function values. The proposed equation guarantees the global convergence without using convexity assumption but this modification does not outperform the BFGS update with general QN equation $B_{k+1}s_k = y_k$. Inspired by this Wei *et al.*^[13] proposed the similar modified QN equation:

$$B_{k+1}s_k = y_k^* \tag{5}$$

Where:

$$y_k^* = y_k + \frac{\phi_k}{s_k^T u_k} u_k \tag{6}$$

$$\phi_k = 2[f(x_k) - f(x_{k+1})] + (g_{k+1} + g_k)^T s_k$$

and $u_k \in \mathbb{R}^n$ is any vector such that $s_k^T u_k \neq 0$. By substituting u_k with s_k in (6), we obtain the following modified QN equation:

$$B_{k+1}s_k = \tilde{y}_k = y_k + \mu_k s_k \tag{7}$$

Where:

$$\mu_k = \frac{\phi_k}{\|s_k\|^2}$$

With (7), the SR1 update has the form:

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

Furthermore, to avoid solving a linear system of equations on every iteration, we can update $B_k^{-1} = H_k$ directly based upon (5) as an approximation to the inverse of the Hessian. So the search direction p_k can be computed by $p_k = -H_k g_k$ and (7) is written as:

$$s_k = H_{k+1} \tilde{y}_k$$

Where:

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

This modified SR1 update efficiently exploits both gradient and function information.

Since memoryless SR1 formula updated from modified QN Eq. 7 may lose positive definiteness of updated matrix in the following we present a scaling factor to update modified SR1 formula from a scaled identity matrix.

Scaling factor: In 1993 Dennis and Wolkowicz^[5] suggested the following measure:

$$\sigma(A) = \frac{\zeta_A}{\det(A)^{1/n}} \tag{8}$$

Where:

A = An n×n positive definite matrix

ζ_A = The largest eigenvalue of A

Note that finding the optimal scaling factor for the modified SR1 update in the measure given by (8) is easier than to find it in the l₂-norm condition number. Hence in the following theorem we try to find the 'best' modified SR1 update from a positive multiple of identity matrix that satisfies the modified secant Eq. 7 and preserves positive definiteness of the update.

Theorem 1: Let:

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

Then the modified SR1 matrix updated from $\frac{1}{\lambda_k} I$:

$$B_{k+1} = \frac{1}{\lambda_k} I + \frac{[\tilde{y}_k - (1/\lambda_k)s_k][\tilde{y}_k - (1/\lambda_k)s_k]^T}{s_k^T [\tilde{y}_k - (1/\lambda_k)s_k]} \quad (10)$$

is the unique solution of:

$$\min \sigma(B_{k+1}^{-1}) \quad (11)$$

$$\text{s.t. } B_{k+1}^{-1} \tilde{y}_k = s_k$$

and B_{k+1}⁻¹ positive definite.

Proof: The proof of this theorem is a result of Theorem 1 of^[14] which H_k and y_k is replaced by the identity matrix and the modified \tilde{y}_k , respectively. Note that since the proof in^[14] is not depended to y_k, the result is true after such replacement.

Corollary 1: Let:

$$\tilde{\lambda}_k = \frac{s_k^T s_k}{s_k^T \tilde{y}_k} - \left(\frac{(s_k^T s_k)^2}{(s_k^T \tilde{y}_k)^2} - \frac{s_k^T s_k}{\tilde{y}_k^T \tilde{y}_k} \right)^{1/2} \quad (12)$$

Then the scaled modified memoryless SR1 update:

$$H_{k+1} = \tilde{\lambda}_k I + \frac{(s_k - \tilde{\lambda}_k \tilde{y}_k)(s_k - \tilde{\lambda}_k \tilde{y}_k)^T}{\tilde{y}_k^T (s_k - \tilde{\lambda}_k \tilde{y}_k)} \quad (13)$$

is the unique solution of:

$$\min \sigma(H_{k+1}^{-1}) \quad (14)$$

$$\text{s.t. } H_{k+1}^{-1} s_k = \tilde{y}_k$$

and H_{k+1}⁻¹ positive definite.

Proof: The proof is the direct result of the above theorem by interchanging the role of s and \tilde{y} .

Remark 1: The scaling factor $\tilde{\lambda}_k$ is positive by using the Cauchy-Schwartz inequality.

Description of algorithm:

MMSR1 algorithm:

- Given an initial point x₀, set k = 0
- Set the termination test, $\|\nabla f(x_k)\| \leq \epsilon$. If the termination test is achieved, then stop
- Compute the scaling modified memoryless SR1 direction, for k = 0, let p₀ = -g₀ and for k>0 compute p_k by p_k = -H_kg_k:

$$p_k = -\tilde{\lambda}_{k-1} g_k + \tilde{\lambda}_{k-1} \left(\frac{s_{k-1}^T g_k - \tilde{\lambda}_{k-1} \tilde{y}_{k-1}^T g_k}{\tilde{y}_{k-1}^T s_{k-1} - \tilde{\lambda}_{k-1} \tilde{y}_{k-1}^T \tilde{y}_{k-1}} \right) \tilde{y}_{k-1} - \left(\frac{s_{k-1}^T g_k - \tilde{\lambda}_{k-1} \tilde{y}_{k-1}^T g_k}{\tilde{y}_{k-1}^T s_{k-1} - \tilde{\lambda}_{k-1} \tilde{y}_{k-1}^T \tilde{y}_{k-1}} \right) s_{k-1} \quad (15)$$

where, $\tilde{\lambda}_{k-1}$ and H_k is given by (12) and (13) and:

$$\tilde{y}_k = y_k + \frac{\phi_k}{\|s_k\|^2} s_k$$

where, $\phi_k = 2[f(x_k) - f(x_{k+1})] + (g_{k+1} + g_k)^T s_k$.

- Find the step-length β_k such that β_k satisfy Wolfe conditions:

$$f(x_k + \beta_k p_k) \leq f(x_k) + \delta_1 \beta_k g_k^T p_k \quad (16)$$

$$g(x_k + \beta_k p_k)^T p_k \geq \delta_2 g_k^T p_k \quad (17)$$

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

- Update $x_{k+1} = x_k + \beta_k p_k$
- Set k = k+1 and go to the step 1

RESULTS

Convergence result: We establish our analysis for the convergence result of MMSR1 Algorithm. We apply

our results to obtain globally convergence and R-linearly result under some standard conditions. For this purpose, we present the following assumption:

Assumptions:

- (i) The level set $\Omega = \{x | f(x_0) \leq f(x)\}$ is contained in a convex set D.
- (ii) Let G be the matrix of second derivatives of f then there exists constant L_1 and L_2 such that:

$$L_1 \|z\|^2 \leq z^T G(x) z \leq L_2 \|z\|^2 \tag{18}$$

for all $z \in \mathbb{R}^n$ and all $x \in D$.

Theorem 2: Let f satisfy the Assumptions and the sequences $\{x_k\}$ be generated by MMSR1 Algorithm. Then the sequences $\{x_k\}$ converge globally to x^* .

Proof: The Wolfe conditions in step 3 of MMSR1 Algorithm and the positive definiteness and boundedness of the memoryless modified SR1 matrix implies that:

$$f(x_{k+1}) \leq f(x_k) - \beta_1 \alpha_k q \|g_k\|^2 \tag{19}$$

for some positive constant q. Therefore $f(x_{k+1}) \leq f(x_k)$ for all k and since f is bounded below, it follows that:

$$\lim_{k \rightarrow \infty} [f(x_k) - f(x_{k+1})] = 0$$

Since $f(x_{k+1}) - f(x_k) \leq -\beta_1 \alpha_k q \|g_k\|^2$, as a consequence $\|g_k\|$ goes to zero, i.e., $\{x_k\}$ converges to x^* .

Theorem 3: Let f satisfies in the Assumptions and the sequences $\{x_k\}$ generated by MMSR1 Algorithm. Then there exists a constant $0 \leq t < 1$ such that:

$$f(x_k) - f(x^*) \leq t^k (f(x_0) - f(x^*)) \tag{20}$$

hence we can deduce that $\{x_k\}$ converges R-linearly to x^* .

Proof: Using Theorem 3.1 and Corollary 3.1 of^[14], one can show that:

$$\frac{1}{\lambda_k} \text{ and } \frac{1}{\lambda'_k}$$

Where:

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

$$\lambda'_k = \frac{\tilde{y}_k^T \tilde{y}_k}{\tilde{y}_k^T s_k} + \left(\frac{(\tilde{y}_k^T \tilde{y}_k)^2}{(\tilde{y}_k^T s_k)^2} - \frac{\tilde{y}_k^T \tilde{y}_k}{s_k^T s_k} \right)^{1/2}$$

are two distinct eigenvalues of B_{k+1} . Using Assumption

(ii), we can show the boundedness of $\frac{\tilde{y}_k^T \tilde{y}_k}{\tilde{y}_k^T s_k}$ and $\frac{\tilde{y}_k^T \tilde{y}_k}{s_k^T s_k}$.

Therefore there exists constant m_1 and m_2 such that $m_1 \leq \mu_i \leq m_2$ for each eigenvalues μ_i of B_{k+1} . Hence by taking the trace operation on B_{k+1} and the fact that the sequence $\{\|B_k\|\}$ is bounded we can see that the trace of B_{k+1} satisfies:

$$\text{tr}(B_{k+1}) \leq nm_2 \tag{21}$$

and the determinant of B_{k+1} is bounded from below

$$\det(B_{k+1}) \geq m_1^n \tag{22}$$

Therefore from (21) and (22), we conclude that there exists a positive constant δ such that:

$$\cos \theta_k = \frac{s_k^T B_k s_k}{\|s_k\| \|B_k s_k\|} \geq \delta \tag{23}$$

By using Assumptions and the Wolfe conditions in step 3 of MMSR1 Algorithm, we can show that there exists a positive constant c such that:

$$f(x_{k+1}) - f(x^*) \leq (1 - c \cos^2 \theta_k) (f(x_k) - f(x^*)) \tag{24}$$

See for example^[11].

Applying (24) recursively together (23), we obtain (20). Finally, from (18):

$$\frac{1}{2} L_1 \|x_k - x^*\|^2 \leq f(x_k) - f(x^*)$$

which together with (20) implies $\|x_k - x^*\|^2 \leq t^{k/2} [2(f(x_0) - f(x^*)) / L_1]^{1/2}$. Hence the sequence $\{x_k\}$ is also R-linearly convergent.

Table 1: Ratio of algorithm MMSR1 cost to MMBFGS cost

Mean	MMSR1	
	Iterations	Function evaluations
Arithmetic	0.67	0.88
Geometric	0.56	0.80

Table 2: Ratio of algorithm MMSR1 cost to FRSCG cost

Mean	MMSR1	
	Iterations	Function evaluations
Arithmetic	0.51	0.58
Geometric	0.43	0.46

Numerical result: Now we present the computational performance of MMSR1 Algorithm. We use the following existing packages:

- The limited memory QN package by Liu and Nocedal^[9] was modified to fit with the modified QN equation. The package has the variable storage capability controlled by the parameter m , where m is the number of stored updates. For our case, $m = 1$ is chosen. Therefore this method is called memoryless modified BFGS (MMBFGS) method in this study
- The conjugate gradient algorithm by Birgin and Martínez^[2] (FRSCG), which is mainly a scaled variant of Perry's^[10]. The algorithm is preserving the nice geometrical properties of Perry's direction and uses Fletcher-Reeve formula. It is implemented in such a manner in which the parameter scaling the gradient defining the search direction is selected by means of a spectral formula.

All the algorithms are implemented in Fortran 77 and stopped whenever $\|g_k\| \leq \epsilon$ where $\epsilon = 10^{-5}$. Seventy-three large-scale unconstrained optimization functions has been tested from Cute collection^[3], along with other large-scale optimization problems from^[1]. The problems tested have variable dimensions and we tested our algorithms with dimensions $1000 \leq n \leq 10000$. This has resulted in a total of 730 runs.

In Table 1 and 2, we summarize our numerical experiments by using the Geometric and Arithmetic means of number of iterations and function/gradient evaluations requires to solve these problems by the MMSR1 Algorithm to the corresponding mean for the MMBFGS and FRSCG method.

Also in order to access the performances of these algorithms, the performances have been evaluated using the performance profiling of Dolan and More^[6]. In Fig. 1 and 2 we compare the performance of MMSR1, MMBFGS and FRSCG referring to the number of iteration, function/gradient calls, respectively.

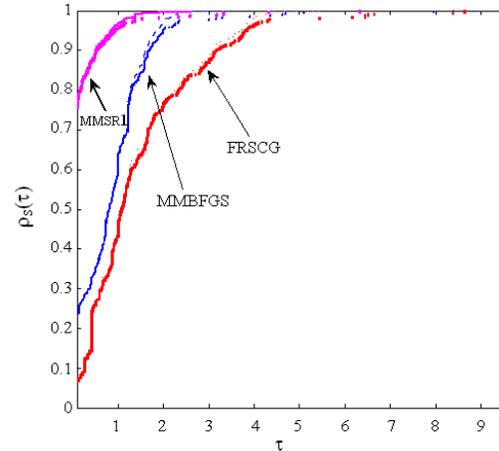


Fig. 1: Performance profile based on iterations

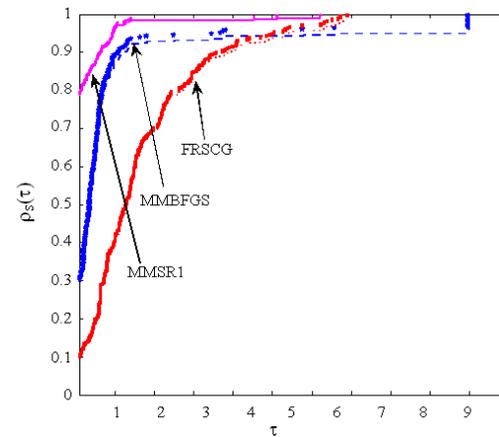


Fig. 2: Performance profile based on function/gradient calls

DISCUSSION

In this study we presented a new method, in which the well known SR1 update is computed in the framework of memoryless method through modified secant condition. Memoryless method is very powerful method for solving large-scale unconstrained optimization problems characterized by low memory requirements and strong local and global convergence properties. The method takes both the available gradient and the function values information in two successive iteration points and achieves high-order accuracy in approximating the second-order curvature of the minimizing function. Under very mild assumptions, the method is proved to be globally convergent. It is shown that the convergence of the MMSR1 algorithm is R-linear. Also the results presented in Table 1 and 2 imply

that MMSR1 Algorithm improved significantly over the performance of MMBFGS and FRSCG methods. Thus, the improvement of Algorithm MMSR1 over MMBFGS is 12-40%, in average, respectively in terms of the number of iterations and function/gradient calls.

Similarly, the improvement of algorithm MMSR1 over FRSCG is 40-50%, in average, respectively, in terms of the number of iterations and function /gradient calls. Therefore the algorithm MMSR1 is 12-50% in average faster and cheaper than MMBFGS and FRSCG methods. Moreover, the experimental results also show that MMSR1 solves 80% of all problems while MMBFGS and FRSCG can solve 70% and 72% of all problems, respectively.

From Fig. 1 and 2, we observe that, MMSR1 performs better than the MMBFGS and FRSCG methods. From the numerical results, MMSR1 Algorithm is clearly better than MMBFGS algorithm and vast superior than FRSCG algorithm.

CONCLUSION

We have presented a scaled memoryless modified SR1 algorithm, MMSR1, based on the modified QN equation by employing modified SR1 formula within the memoryless QN framework for solving large scale unconstrained optimization problems. We use the positive multiple of identity matrix to update the modified SR1 matrix. MMSR1 algorithm can be computed without the storage of matrices, namely n^2 storages. Therefore we conclude that by incorporating a simple scaling factor into the modified memoryless SR1 update we can preserve positive definiteness of the modified SR1 matrix. Moreover the numerical results show that, the MMSR1 method performs significantly better than the other two methods and the improvements exhibited by the MMSR1 algorithm over MMBFGS and FRCS are evident in both function/gradient evaluations and iterations. Also we have considered the global and local behavior of the scaled memoryless modified SR1 algorithm with some standard conditions. Specifically, we have showed R-linear convergence property of our method.

REFERENCES

1. Andrei, N., 2008. An unconstrained optimization test functions collection. *Adv. Model. Optim.*, 10: 147-161. www.ici.ro/camo/journal/vol10/v10a10.pdf
2. Birgin, E. and J.M. Martínez, 2001. A spectral conjugate gradient method for unconstrained optimization. *Applied Math. Optim.*, 43: 117-128. DOI: 10.1007/s00245-001-0003-0
3. Bongartz, I., A.R. Conn, N.I.M. Gould and Ph.L. Toint, 1995. CUTE: Constrained and unconstrained testing environment: *ACM Trans. Math. Software*, 21: 123-160. <http://portal.acm.org/citation.cfm?doid=200979.201043>
4. Conn, A.R., N.I.M. Gould and Ph.L. Toint, 1991. Convergence of quasi-Newton matrices generated by the symmetric rank one update. *Math. Program.*, 48: 549-560. DOI: 10.1007/BF01594934
5. Dennis, J.E. and H. Wolkowicz, 1993. Sizing and least change secant methods. *SIAM J. Numer. Anal.*, 30: 1291-1313. DOI: 10.1137/0730067
6. Dolan, E.D. and J.J. Moré, 2002. Benchmarking optimization software with performance profiles. *Math. Program.*, 91: 201-203, DOI: 10.1007/s101070100263
7. Khalfan, H., R.H. Byrd and R.B. Schnabel, 1993. A theoretical and experimental study of the symmetric rank one update. *SIAM J. Optim.*, 3: 1-24. DOI: 10.1137/0803001
8. Li, D. and M. Fukushima, 2001. A modified BFGS method and its global convergence in nonconvex minimization. *J. Comput. Appl. Math.*, 129: 15-35.
9. Liu, D.C. and J. Nocedal, 1989. On the limited memory method for large-scale optimization. *Math. Program.*, B45: 503-528. DOI: 10.1016/S0377-0427(00)00540-9.
10. Perry, J.M., 1977. A class of conjugate gradient algorithm with a two step variable metric memory: Discussion Paper 269, Center for Mathematical studies in Economics and Management Science, Northwestern University, Chicago. <http://www.kellogg.northwestern.edu/research/math/papers/269.pdf>
11. Powell, M.J.D., 1976. Some global convergence properties of a variable metric algorithm for minimization without exact line searches. *Proceeding of the Symposium on Applied Mathematics of the American Mathematical Society and the Society for Industrial and Applied Mathematics*, April 11-12, New York, pp: 53-72. <http://www.zentralblatt-math.org/zmath/en/search/?an=0736.90068>
12. Shanno, D.F., 1978. Conjugate gradient methods with inexact searches. *Math. Operat. Res.*, 3: 244-256, DOI: 10.1287/moor.3.3.244
13. Wei, Z., G. Li and L. Qi, 2006. New quasi-Newton methods for unconstrained optimization problems. *J. Applied Math. Comput.*, 175: 1156-1188, DOI: 10.1016/j.amc.2005.08.027
14. Wolkowicz, H., 1994. Measure for symmetric rank-one updates. *Math. Operat. Res.*, 19: 815-830, DOI: 10.1287/moor.19.4.815