World Applied Sciences Journal 7 (5): 610-615, 2009
ISSN 1818-4952

Multi-steps Symmetric Rank-one Update for 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: In this paper, we present a generalized Symmetric Rank-one (SR1) method by employing interpolatory polynomials in order to possess a more accurate information from more than one previous step. The basic idea is to incorporate the SR1 update within the framework of multi-step methods. Hence iterates could be interpolated by a curve in such a way that the consecutive points define the curves. However to preserve the positive definiteness of the SR1 updates a restart procedure is applied, in which we restart the SR1 update by a scale of the identity. Comparison to multi-steps BFGS method, the proposed algorithm shows significant improvements in numerical results.

Mathematics subject classification: 65K10 . 90C53

Key words: Unconstrained optimization . symmetric rank-one update . multi-step methods . Hessian approximation

INTRODUCTION

In this paper we are concerned with the numerical methods for solving the following unconstrained nonlinear optimization problem

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

where the objective function f: $\mathbb{R}^n \to \mathbb{R}$ is a twice continuously differentiable defined in n-dimensional space.

Among numerous iterative methods for solving (1), quasi-Newton (QN) methods constitute a popular class of methods. These methods are variants of Newton's method, in which the Hessian matrix $\nabla^2 f(x_k)$ is substituted by an approximation to the Hessian. QN methods, require only the gradient of the objective function to be supplied at each iterate. The improvement over other methods like steepest descent is dramatic, epecially on difficult problems. Moreover, since second derivatives are not required, QN methods are sometimes more efficient than Newton's method. For backgound of QN Updates [8].

Among various QN methods, symmetric rank-one (SR1) method is regarded as very competitive formula compared with the widely used BFGS formula. In this method, Hessian approximation matrix B_k is updated by the following formula:

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

and inverse of the Hessian approximation by

$$H_{k+1} = H_k + \frac{(s_k - H_k y_k)(s_k - H_k y_k)^T}{y_k^T (s_k - H_k y_k)} \tag{3}$$

where $s_k = x_{k+1} - x_k$, $y_k = g_{k+1} - g_k$ and $g_k = \nabla f(x_k)$ denotes the gradient vector of f at current iteration point x_k. Several studies in minimization algorithms using SR1 formula in both line search and trust region context by Conn *et al*. [3], Khalfan *et al*. [9], Leong and Hassan [10] have sparked renewed interest in this formula. Conn *et al*. [3] showed that the SR1 update generate more accurate Hessian approximation compared with BFGS and DFP.

Ford and Moghrabi [6] introduced multi-steps method, their two-steps QN method are very similar to the QN method with standard secant equation

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

in which (4) is replaced with the modified secant equation

$$B_{k+1} r_k = w_k \tag{5}$$

Corresponding Author: Farzin Modarres, Department of Mathematics, Faculty of Science, University Putra Malaysia, 43400 Serdang, Selangor, Malaysia

where $r_k = s_k - \chi_k s_{k-1}$ and $w_k = y_k - \chi_k y_{k-1}$ with χ_k as a scalar. Equation (5) is approximated by the interpolating quadratic curves $\{x(\theta)\}$ and $\{h(\theta)\}$ where $\{x(\theta)\}$ interpolates the three latest iterates x_{i-1}, x_i and x_{i+1} while $\{h(\theta)\}$ interpolates the corresponding gradient values g_{i-1}, g_i and g_{i+1}. They applied multi-step method by using the well known BFGS formula, in which the vectors s_k and y_k substitute with r_k and w_k, respectively in BFGS formula so that the new Hessian approximation B_{k+1} can be obtained as follows:

$$B_{k+1} = B_k - \frac{B_k r_k r_k^T B_k}{r_k^T B_k r_k} + \frac{w_k w_k^T}{w_k^T r_k} \tag{6}$$

The numerical results show that the new methods offer a significant improvement in performance when compared with the standard BFGS method.

Hence it seems reasonable if we employ the SR1 formula within the framework of the multi-step method. Therefore we expect to have higher order accuracy in approximating the Hessian of the objective function.

However a setback with SR1 update is the fact that its denominator may be zero or it may loss the positive definiteness even when B_{k+1} updated from positive definite update. To overcome these difficulties, we can set the updated matrix with the scale of initial approximation, mostly the positive multiple of identity. But the scaling factor should be derived in such a way that it improves the conditions of the SR1 update while preserving as much information from the previous iterate.

We organize this paper as follows. In the next section we briefly describe multi-step method. In section 3 we will construct SR1 formula obeying modified secant equation (5) and introduce the optimal scaling factor to maintain the positive definiteness of the Hessian approximation matrix. Then we will describe our algorithm in section 4. Finally, computational experiments are reported in section 5 which shows that the multi-step SR1 algorithm is encouraging comparing with the multi-step BFGS method.

MULTI-STEP METHOD

The concept of multi-step methods, developed by Ford and Moghrabi [7], consider data from the m recent steps is employed in the construction of an interpolating path x (θ)x (θ) is any differentiable curve, denoted by χ in $\mathbb{R}^n$), in which the standard secant equation corresponds to m = 1. In multi-step methods the polynomial forms for χ have been considered to interpolate (for a prescribed set of values $\{\theta_k\}_{k=0}^{m}$) the m most recent iteration (m>1):

$x(\theta_k) = x_{i-m+k+1}$ for k = 0, 1,…,m. Thus the curve and the gradient (when restricted to χ), can be approximated by the interpolatory polynomials,

$$x(\theta) = \sum_{k=0}^{m} L_k(\theta) x_{i-m+k+1} \tag{7}$$

$$g(x(\theta)) \approx \sum_{k=0}^{m} L_k(\theta) g(x_{i-m+k+1}) \tag{8}$$

where $L_k(\theta)$ is k^{th} the Lagrangian polynomial form

$$L_k(\theta) \equiv \prod_{\substack{j=0\\ j\neq k}}^{m} \frac{(\theta - \theta_j)}{(\theta_k - \theta_j)} \tag{9}$$

Thus, one can obtain the condition that satisfies in (5) by determining the derivatives of (7) and (8) where,

$$r_i \stackrel{\Delta}{=} x'(\theta_m) = \sum_{k=0}^{m} L'_k(\theta_m) x_{i-m+k+1} \tag{10}$$

$$w_i \stackrel{\Delta}{=} g'(x(\theta_m)) \approx \sum_{k=0}^{m} L'_k(\theta_m) g(x_{i-m+k+1}) \tag{11}$$

The coefficients in (10) and (11) are given explicitly by

$$L'_k(\theta_m) = (\theta_k - \theta_m)^{-1} \prod_{\substack{j=0\\ j\neq k}}^{m-1} \frac{(\theta_m - \theta_j)}{(\theta_k - \theta_j)} \text{ for } k \neq m$$

and

$$L'_m(\theta_m) = \sum_{j=0}^{m-1} (\theta_m - \theta_j)^{-1}$$

It was shown by Ford and Moghrabi that r_i and w_i can represent in terms of the most recent "step-vectors" $\{s_{i-j}\}_{j=0}^{m-1}$ and $\{y_{i-j}\}_{j=0}^{m-1}$

$$r_i = \sum_{j=0}^{m-1} s_{i-j} \{ \sum_{k=m-j}^{m} L'_k(\theta_m) \} \tag{12}$$

$$w_i = \sum_{j=0}^{m-1} y_{i-j} \{ \sum_{k=m-j}^{m} L'_k(\theta_m) \} \tag{13}$$

Also in order to determine the values of $\{\theta_k\}_{k=0}^{m}$, first we introduce a metric ϕ_M defined on $\mathbb{R}^n$ which is generated by any appropriate positive-definite matrix (M, say)

$$\phi_M(l_1, l_2) \stackrel{\text{def}}{=} \left\{ (l_1 - l_2)^T M (l_1 - l_2) \right\}^{1/2} \tag{14}$$

The choices of matrix M considered in [6] were: $M = I$, $M = B_k$ and $M = B_{k+1}$. In this paper we will consider the case $M = B_k$. Now by fixing one of the points, say j, we calculate the distance (measured by metric ϕ_M) between each consecutive pair of points in the sequence by:

$$\theta_k = -\sum_{p=k+1}^{j} \phi_M(x_{i-m+p+1}, x_{i-m+p}), \quad \text{for} \quad k < j \tag{15}$$

$$\theta_j = \theta \tag{16}$$

$$\theta_k = \sum_{p=j+1}^{k} \phi_M(x_{i-m+p+1}, x_{i-m+p}), \quad \text{for} \quad k > j \tag{17}$$

THE SR1 FORMULA BASED ON MULTI-STEP SECANT EQUATION

In this section SR1 formula based on multi-step equation will be proposed. Moreover a scaling factor will be present to avoid the loss of positive definiteness in the modified formula.

SR1 formula with new secant equation: In this subsection we propose modified SR1 formula based on secant equation (5). By substituting m = 2 in (10) and (11), we obtain (after removal of a common scaling factor)

$$r_i = s_i - \left(\frac{\delta^2}{2\delta+1}\right)s_{i-1} \tag{18}$$

$$w_i = y_i - \left(\frac{\delta^2}{2\delta+1}\right)y_{i-1} \tag{19}$$

where

$$\delta = \frac{(\theta_2 - \theta_1)}{(\theta_1 - \theta_0)} \tag{20}$$

In the case of two-step method, to specify the values of $\{\theta_k\}_{k=0}^{2}$, the matrix in (14) is taken to be B_k, the current Hessian approximation. Therefore we determine the measurement of the relevant distances (using (15-17)) by

$$\begin{aligned} \theta_0 &= -\left\{(s_k + s_{k-1})^T B_k (s_k + s_{k-1})\right\}^{1/2} \\ \theta_1 &= -\left\{s_k^T B_k s_k\right\}^{1/2} \\ \theta_2 &= 0 \end{aligned} \tag{21}$$

Therefore the modified SR1 formula based on secant equation (5) where r_i and w_i are expressed as above can be written as follows

$$B_{k+1} = B_k + \frac{(w_k - B_k r_k)(w_k - B_k r_k)^T}{r_k^T (w_k - B_k r_k)} \tag{22}$$

Furthermore, by using updates of B_k^{-1} in modified SR1 formula (22), we can avoid solving linear system of equations to obtain the search direction p_k. By denoting the inverse Hessian approximation of B_k with H_k, the modified secant equation for H_k is written as

$$r_k = H_{k+1} w_k \tag{23}$$

and the SR1 update that approximates the inverse Hessian can be obtained as follows:

$$H_{k+1} = H_k + \frac{(r_k - H_k w_k)(r_k - H_k w_k)^T}{w_k^T (r_k - H_k w_k)} \tag{24}$$

Possible instability of new SR1 update: The modified SR1 update clearly possesses desirable features (like SR1 update) but it has some major disadvantages:

if B_k is well-defined and positive definite, B_{k+1} may fail to inherit these qualities. Therefore in order to avoid zero denominator and non-positive definite updates in modified SR1 update we should have some stabilizing options to overcome these difficulties.

A simple remedy for misbehaved update would be to simply skip the update (set $B_{k+1} = B_k$) or take a gradient step (set $B_{k+1} = I$). These procedures are aesthetically crude and may run the risk of losing valuable information obtained during the descent process and also by looking to the numerical results of Leong and Malik [10], we can observe the lack of convergence to the minimizer whenever we skip the update with identity matrix.

On the other hand Osborne and Sun [11] presented a SR1 update of the form $B_{k+1} = \theta B_k + k_s z_k z_k$, where $z_k = y_k - \theta_s B_k s_k$, $k_s = 1/s_k^T z_k$ and θ_k is a scalar scaling factor that can be chosen to preserve the positive definiteness of update. However scaling the update may prevent fast convergence and it will be redundant if we scale the modified update each iteration. Our motivation in here is that B_{k+1} preserve built up information from B_k. To reach to this aim we scale the identity matrix with a scaling factor, in which the scaling factor is derived in such a way that we may use the valuable information of previous iteration while improving the condition of update. Therefore to find such a scaling factor for the modified SR1 update (24), we present the σ-measure suggested by Dennis and Wolkwicz [4]

$$\sigma(A) = \frac{\lambda_{max}}{\det(A)^{1/n}} \tag{25}$$

where A is an positive definite matrix and λ_{max} is the largest eigenvalue of A. Dennis and Wolkwicz [4] had shown that any σ-optimal will also be κ-optimal as well (κ(A) denotes the l_2-condition number of A) and have a common spectral property.

The following result will be used in the multi-step iterates.

Theorem 3.1: Let

$$\mu_k = \frac{w_k^T w_k}{w_k^T r_k} - \left(\frac{\left(w_k^T w_k\right)^2}{\left(w_k^T r_k\right)^2} - \frac{w_k^T w_k}{r_k^T r_k} \right)^{1/2} \tag{26}$$

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

$$B_{k+1} = \frac{1}{\mu_k} I + \frac{[w_k - (1/\mu_k) r_k][w_k - (1/\mu_k) r_k]^T}{r_k^T [w_k - (1/\mu_k) r_k]} \tag{27}$$

is the unique solution of

$$\min \quad \sigma(B_{k+1}^{-1}) \tag{28}$$

$$\text{suchthat } B_{k+1}^{-1} w_k = r_k$$

and B_{k+1}^{-1}ispositivedefinite.

Proof: See [10].

Remark 3.2: The scaling factor μ_k is positive by using the Cauchy-Schwarz inequality.

Remark 3.3: Since $r_k^T[w_k - (1/\mu_k) r_k] \neq 0$, then the sequences of matrices B_{k+1} generated by (27) are bounded.

Corollary 1 Let

$$\tilde{\mu}_k = \frac{r_k^T r_k}{r_k^T w_k} - \left(\frac{\left(r_k^T r_k\right)^2}{\left(r_k^T w_k\right)^2} - \frac{r_k^T r_k}{w_k^T w_k} \right)^{1/2} \tag{29}$$

Then the inverse SR1 matrix updated from $\tilde{\mu}_k I$:

$$H_{k+1} = \tilde{\mu}_k I + \frac{(r_k - \tilde{\mu}_k w_k)(r_k - \tilde{\mu}_k w_k)^T}{w_k^T (r_k - \tilde{\mu}_k w_k)} \tag{30}$$

is the unique solution of

$$\min \quad \sigma(H_{k+1}^{-1}), \tag{31}$$

$$\text{suchthat } H_{k+1}^{-1} r_k = w_k$$

and H_{k+1}^{-1} ispositivedefinite.

Proof: The proof is the direct result of the theorem 3.1 by interchanging the role of s and y.

DESCRIPTION OF ALGORITHM

Multi-steps symmetric rank-one Algorithm (MSSR1)

- Given an initial point $x_0 \in$ í^n, an initial positive matrix $H_0 = I$, compute f (x_0) and $g_0 = \nabla f(x_0)$. Set k=0.
- Termination test. If the convergence criterion $\|g_k\| \leq \varepsilon$ is achieved, then stop.
- Compute a QN direction, d_k, by $d_k = -H_k g_k$.
- Find an acceptable steplength, α_k, such that the Wolfe conditions

$$f(x_k + \alpha_k p_k) \leq f(x_k) + \delta_1 \alpha_k g_k^T p_k \tag{32}$$

$$\nabla f(x_k + \alpha_k p_k)^T d_k \geq \delta_2 g_k^T p_k \tag{33}$$

where $0<\delta_1<\delta_2<1$, $\delta_1 < \frac{1}{2}$, are satisfied . ($\delta_1 = 10^{-4}$, $\delta_2 = 0.9$).

- Set $x_{k+1} = x_k + \alpha_k d_k$.
- If single-step iteration is being executed, set $r_k = s_k$ and $w_k = y_k$.
- Stabilizing. If

$$r_k^T w_k - w_k^T H_k w_k < 0; \text{(Hessianmatrixisnotpositivedefinite)}$$

or $$\left| w_k^T (r_k - H_k w_k) \right| < t \|w_k\| \|r_k - H_k w_k\|$$

where $t \in (0,1)$; (denominator in Hessian matrix is sufficiently close to zero),

or $\|H_k\|_\infty > L$; (whereLisapresetconstant)

set $H_{k+1} = \mu_k I$, where

$$\mu_k = \frac{r_k^T r_k}{w_k^T r_k} - \left\{ \frac{\left(r_k^T r_k\right)^2}{\left(w_k^T r_k\right)^2} - \frac{r_k^T r_k}{w_k^T w_k} \right\}^{1/2}$$

- Compute the next inverse Hessian approximation H_{k+1} by (24) (or (3) in case of single-step).
- Set k = k+1 and go to step 1.

Table 1: Test problems

Andrei [1] test problems:	
	Ex. Freudenstein & Roth, Ex. Trigonometric, Ex. Rosenbrock, Ex. White & Holst, Ex. Beale, Ex. Penalty, Perturbed Quadratic, Raydan 1&2, Hager, Ge. Tridiagonal, Ex. Tridiagonal, Ex. Three Expo Terms, Ge. Tridiagonal 2, Ex. Himmelblau, Ge. PSC1, Ex. PSC1, Ex. Powell, Ex. BD1, Ex. Maratos, Ex. Cliff, Quadratic Diagonal Perturbed, Ex. Wood, Ex. Hiebert, QF1, QF2, QP1, QP2, Ex. EP1, Ex. Tridiagonal 2, PPQ1, Broyden Tridiagonal, Almost Perturbed Quadratic, Tridiagonal Perturbed Quadratic, STAIRCASE S1, PPQ2, SQ1, SQ2, Diagonal (1-6)
Cute [2] test problems:	
	BDQRTIC, TRIDIA, ARWHEAD, NONDIA, NONDQUAR, DQDRTIC, EG2, ENGVAL1, FLETCHCR, COSINE, Ex. DENSCHNB, Ex. DENSCHNF, SINQUAD, BIGGSB1, EDENSCH, VARDIM, LIARWHD, DIXMAANA, DIXMAANB, DIXMAANC, DIXMAANE, DIXON3DQ, DIXMAANF, DIXMAANG, DIXMAANH, DIXMAANI, DIXMAANJ, DIXMAANK, DIXMAANL

NUMERICAL RESULTS

In this section we present the performance of the MSSR1 on a set of 73 unconstrained optimization problems. At the same time, we compare the performance of MSSR1 with multi-step algorithm for standard BFGS update (MSBFGS).

All the algorithms are implemented in Fortran 77. In all cases, convergence is assumed if $\|g_k\|\leq\varepsilon$ for $\varepsilon = 10^{-4}$. The algorithm also stopped whenever the number of iterations or function evaluations exceeds 999. We selected 73 unconstrained optimization test problems from the Cute [2] library, along with other optimization test problems from [1]. Each function is tested with variable dimensions $2\leq n\leq 1000$. This has resulted in a total of 1300 runs. The names of the problems are listed in Table 1.

In Table 2, we present the Geometric and Arithmetic means of number of iterations and function/gradient evaluations requires to solve these problems by the MSSR1 Algorithm to the corresponding mean for the MSBFGS method corresponding to these 730 test problems, referring to the total number of iterations, the total number of function/gradient evaluations. We presented the summary of our results in Table 2.

The results presented in Table 2 imply that MSSR1 Algorithm improved significantly over the performance of MSBFGS method. The improvement of Algorithm MSSR1 over MSBFGS is 17% to 24%, in average, in terms of the number of iterations and 13% to 18%, in average, in terms of the number of function/gradient calls.

Also in order to compare the performance of our algorithms to access their performance and represent graphically, we use the performance profiling proposed by Dolan and More [5].

Table 2: Ratio of algorithm MSSR1 cost to MSBFGS cost

	MMSR1	
Mean	Iterations	Function evaluations
Arithmetic	0.83	0.87
Geometric	0.76	0.82

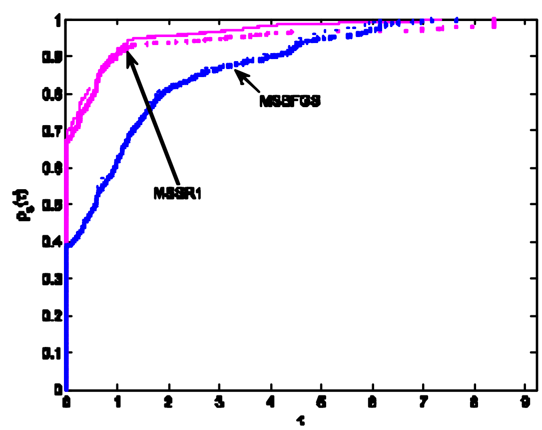

Fig. 1: Performance profile based on iterations

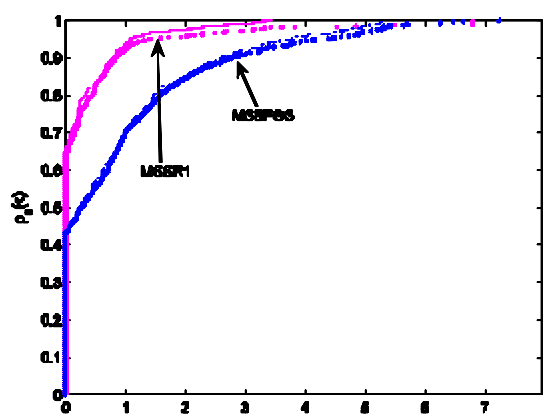

Fig. 2: Performance profile based on function/gradient calls

As it is represented in Fig. 1 and 2, MSSR1 performs better than the MSBFGS method. Therefore the efficiency of our proposed algorithm over MSBFGS is clearly observed.

CONCLUSION

We have presented a new method called the multi-step SR1 algorithm (MSSR1) which involves alternating two-step QN methods with the standard SR1 method that rely on one of two approaches to define the distribution of recent iterates on an interpolating curve. We use the positive multiple of identity matrix to update the presented SR1 formula. The scaling factor is derived in such a way that the multi-step SR1 update is optimally conditioned. Moreover by this consideration, we preserve positive definiteness of updated matrix. The numerical results for a broad class of test problems show that the new method are efficient and robust in solving unconstrained optimization problems. Finally, we may note that the new method outperform multi-step BFGS method. Specifically encouraging improvements could be realized by our new method as the size of the problem increases.

REFERENCES

1. Andrei, N., 2008. An unconstrained optimization test functions collection. Adv. Model. Optim., 10: 147-161.
2. Bongartz, I., A.R. Conn and N.I.M. Gould and Ph.L. Toint, 1995. CUTE: Constrained and unconstrained testing environment. ACM Trans. Math. Software, 21: 123-160.
3. 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.
4. Dennis, J.E. and H. Wolkowicz, 1993. Sizing and least change secant methods. SIAM J. Numer. Anal., 30 (5): 1291-1313.
5. Dolan, E.D. and J.J. More', 2002. Benchmarking optimization software with performance profiles. Mathematical Programming, 91 (2): 201-203.
6. Ford, J.A. and I.A. Moghrabi, 1994. Multi-step quasi-Newton methods for optimization. J. Comput. Appl. Math., 50: 305-323.
7. Ford, J.A. and I.A. Moghrabi, 1997. Alternating multistep quasiNewton methods for unconstrained optimization. J. Comput. Appl. Math., 82: 105-116.
8. Gill, P.E., W. Murray and M.H. Wright, 1981. Practical Optimization. Academic, London.
9. 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.
10. Leong, W.J. and M.A. Hassan, 2009. A restarting approach for the symmetric rank one update for unconstrained optimization. Comp. Optim. Appl., 42 (3): 327-334.
11. Osborne, M.R. and L. Sun, 1999. A new approach to symmetric rank-one updating. IMA J. Numer. Anal., 19: 497-507.
12. Wolkowicz, H., 1994. Measure for symmetric rank-one updates. Math. Oper. Res., 19 (4): 815-830.