A modified quasi-Newton methods for unconstrained optimization

Basim A. Hassan
Department of Mathematics
College of Computers Sciences and Mathematics
University of Mosul
Iraq
basimabas39@gmail.com

Abstract. In this paper, we present a new modified SR1 method based on the new quasi-Newton equation. Under some suitable condition, we prove the global convergence of method with Wolfe condition. The Numerical results show that the proposed method effective for the given some test problems.
Keywords: quasi-Newton equations, symmetric rank-one update, convergence.

1. Introduction

In this paper, the proposed method is designed to solve the following problem of unconstrained optimization:

$$Minf(x), x \in \mathbb{R}^n, \tag{1}$$

where is a continuously differentiable objective function. [7] This problem can be solved by using several numerical methods including quasi-Newton methods, which are well-known to be the effective ones. The iterative formula of the Quasi-Newton methods is given by

$$x_{k+1} = x_k + \alpha_k d_k, \tag{2}$$

where α_k is the parameter determined by exact line-search as:

$$\alpha_k = \frac{\nabla f(x_x)^T d_k}{d_k^T Q d_k} \tag{3}$$

and calculate the search direction by:

$$B_k d_k + \nabla f(x_k) = 0. \tag{4}$$

The updated matrix B_k is generated by:

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

where $s_k = x_{k+1} - x_k$ and $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$, (5) is very famous and named quasi-Newton equation. Based on the above equation (5), the very famous update B_k is the SR1 formula:

$$B_{k+1}^{SRI} = B_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{(y_k - B_k s_k)^T s_k}. \tag{6}$$

For more details can be found in [11]. The SR1 method of updates"that approximates the inverse Hessian can be obtained from(5):

$$H_{k+1}^{SRI} = H_k + \frac{(s_k - H_k y_k)(s_k - H_k y_k)^T}{(s_k - H_k y_k)^T y_k}. \tag{7}$$

For standard SR1 method, the steplength α_k in (3) is determined by Wolfe-Powell line search:

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k \nabla f(x_k)^T d_k, \tag{8}$$

$$d_k^T g(x_k + \alpha_k d_k) \geq \sigma f(x_k)^T d_k, \tag{9}$$

where $0 \prec \delta \prec \sigma$. For more details can be found in (Wolfe, [8,9]). These methods can be modified using other alternatives of QN algorithms for upgrading their efficiency whilst maintaining important properties (see [2]). Furthermore, the idea of modified quasi-Newton equation had been studied by many researchers for example, see (Powell, [6]); (Li and Fukushima, [3]); (Wei, Li, and Qi, [10]) ; and (Zhang, Deng, and Chen, [14]).

Moreover, among all QN updating formulas, the SR1 formula has very effective properties of correction; therefore, it is currently regarded as the most effective one (see [5]).

Now we will derive new quasi-Newton equations and analyze its convergence.

2. Motivation based on quasi-Newton equation

In 1999, Zhang et al. [14] presented a modified secant condition exploits not only the gradients but also the function values. Using the Taylor expansion to third-order terms, f and $\nabla f(x_k)^T s_k$ can be written as respectively:

$$f_k = f_{k+1} - \nabla f(x_{k+1})^T + \frac{1}{2} s_k^T G_{k+1} + \frac{1}{6} s_k^T T_{k+1} s_k + O(\|s_k\|^4), \tag{10}$$

$$\nabla f(x_k)^T s_k = \nabla f(x_{k+1}^T s_k - s_k^T G_{k+1} s_k + \frac{1}{2} s_k^T (T_{k+1} s_k) s_k + O(\|s_k\|^4), \tag{11}$$

where $T_{k+1} \in R^{n \times n \times n}$ is the tensor of at x_{k+1} . Canceling the terms which include the tensor yields:

$$s_k^T G_{k+1} s_k = y_k^T s_k = 6(f_k - f_{k+1}) + 3(\nabla f(x_{k+1} + \nabla f(x_k))^T s_k. \tag{12}$$

For more details can be found in [12,14].

Now, the motivation quasi-Newton equation based on the above relation:

$$\frac{1}{6}s_k^T G_{k+1}s_k = (f_k - f_{k+1}) + \frac{2}{3}s_{k+1}^T + \frac{1}{3}s_k^T \nabla f(x_k). \tag{13}$$

It follows from the meaning of α_k:

$$s_k^T G_{k+1}s_k = (f_k - f_{k+1}) + \frac{2}{3}s_k^T \nabla f(x_{k+1}) - \frac{1}{2}s_k^T \nabla f(x_k). \tag{14}$$

It follows from the meaning ofy_k we get:

$$y_k^T s_k = \nabla f(x_{k+1})^T s_k - \nabla f(x_k)^T s_k. \tag{15}$$

By putting (15)in (14) we get:

$$s_k^T G_{k+1}s_k = \frac{1}{2}s_k^T y_k + (f_k - f_{k+1}) + \frac{1}{6}s_k^T \nabla f(x_{k+1}). \tag{16}$$

To obtain a higher truth in estimating the Hessian matrix by B_{k+1} it is reasonable to let B_{k+1} satisfy:

$$s_k^T B_{k+1}s_k = \frac{1}{2}s_k^T y_k + (f_k - f_{k+1}) + \frac{1}{6}s_k^T \nabla f(x_{k+1}). \tag{17}$$

good choice to estimate $B_{k+1}s_k$ is known by:

$$B_{k+1}s_k = y_k^*, y_k^* = \frac{1}{2}y_k + \frac{(f_k - f_{k+1}) + 1/6(\nabla f(x_{k+1})^T s_k)}{s_k^T u_k} u_k, \tag{18}$$

where u_k is any vector such that $s_k^T u_k \neq 0$.

Here we will give our two choices of vector u_k in above equation.

1. Putting $u_k = \nabla f(x_{k+1})$ in (18) we have:

$$y_k^* = \frac{1}{2}y_k + \frac{(f_k - f_{k+1}) + 1/6(\nabla f(x_{k+1})^T s_k)}{s_k^T \nabla f(x_{k+1})} \nabla f(x_{k+1}). \tag{19}$$

2. Putting $u_k = y_k$ in (18) we have:

$$y_k^* = \frac{1}{2}y_k + \frac{(f_k - f_{k+1}) + 1/6(\nabla f(x_{k+1})^T s_k)}{s_k^T y_k} y_k. \tag{20}$$

Applying any new quasi-Newton equation to the SR1 formula, this yields the new formula.

The new algorithm method of Modified SR1-type is expressed formally as follows.

2.1 New algorithm

1. Given an initial pointx_0 an initial positive definite matrix $H_0 = I$.
2. If the convergence criterion is achieved, the stop.
3. Compute a quasi-Newton direction by $d_k = -H_k \nabla f(x_k)$.
4. Find an acceptable steplength such that the Wolfe condition.
5. Set $x_{k+1} = x_k + \alpha_k d_k$.
6. Calculate y_k^* by using equation.
7. Compute the next inverse Hessian approximation H_{k+1} by (7).
8. Set $k = k + 1$ and go to Step 1.

The following theorem shows good property of the new secant equation.

Theorem 2.1. *Let* $(\alpha_k,\ x_{k+1},\ g_{k+1},\ d_{k+1})$ *be generated through the new algorithm. Then* H_{k+1} *is non negative definite for* $\forall k$ *provided that* $s_k^T y_k^* \succ 0$.

Proof of Theorem 2.1. And by evaluating the amount, we have by the Wolfe conditions (4) that:

$$s_k^T y_k^* = \frac{1}{2} s_k^T y_k + (f_k - f_{k+1}) + \frac{1}{6} s_k^T \nabla f(x_{k+1}),$$

$$s_k^T y_k^* = (f_k - f_{k+1}) + \frac{2}{3} s_k^T \nabla f(x_{k+1}) - \frac{1}{2} s_k^T \nabla f(x_k). \tag{21}$$

By Wolfe condition (5), we get:

$$\begin{aligned} s_k^T y_k^* &\geq -\delta \nabla f(x_k)^T s_k + \frac{2}{3} \sigma \nabla f(x_k)^T s_k - \frac{1}{2} s_k^T \nabla f(x_k) \\ &\geq s_k^T \nabla f(x_k)(\delta + \frac{1}{2} - \frac{2}{3}\sigma) \geq s_k^T \nabla f(x_k)(\delta + \frac{1}{2} - \frac{2}{3}\delta) \geq s_k^T \nabla f(x_k)(\frac{1}{2} + \frac{1}{3}\delta), \end{aligned}$$

$$s_k^T y_k^* \geq s_k^T \nabla f(x_k)(\frac{1}{2} + \frac{1}{3}\delta). \tag{22}$$

Noting the $s_k^T \nabla f(x_k) = \alpha_k d_k^T \nabla f(x_k) \prec 0$, we know that there exists a constant $m \succ 0$ such that:

$$s_k^T y_k \geq -m d_k^T \nabla f(x_k) \succ 0 \tag{23}$$

The proof is complete.

3. Global convergence

Throughout this section, the following assumptions will be considered about the objective function f.

Assumption:

Let G be the matrix of second derivative of f.

i. The objective function f is continuously differentiable in a neighbor-hood Q of the level set $D = \{x \in R^n : f(x) \leq f(x_0)\}$ and bounded below R^nin.

ii. The “gradient is Lipschitz continuous, i.e., there exists a constant $L \succ 0$ such that:

$$||\nabla f(x_2) - \nabla f(x_1)|| \leq ||x_2 - x_1||, \forall x_1, x_2 \in Q. \tag{24}$$

Let us define:

$$\cos(\theta_k) = \frac{\nabla f(x_k)^T d_k}{||\nabla f(x_k)||||d_k||}. \tag{25}$$

For more details can be found in [1,11].

Theorem 3.1. *Let x_0 be a starting point for which f satisfies Assumption. Consider $\{x_k\}$ the sequence of points generated by the updating scheme $x_{k+1} = x_k + \alpha_k d_k$ where the $\{B_k\}$ sequence is generated by BFGS and α_k satisfies the Wolfe conditions. Then:*

$$\sum \cos^2(\theta_k) = ||\nabla f(x_k)||^2 \prec \infty. \tag{26}$$

Proof of Theorem 3.1. *From Wolfe conditions we have that:*

$$(\nabla f(x_{k+1}) - \nabla f(x_k))^T d_k \geq (\sigma - 1)\nabla f(x_k)^T d_k. \tag{27}$$

Moreover, from Lipschitz condition we obtain:

$$(\nabla f(x_{k+1}) - \nabla f(x_k))^T d_k \leq \alpha_k L||d_k||^2. \tag{28}$$

Combining the above inequalities gives:

$$(\nabla f(x_k)^T d_k)/||d_k||^2 \frac{\sigma - 1}{L} \leq \alpha_k. \tag{29}$$

Therefore, by using the first Wolfe condition and above inequality, we have:

$$f_{k+1} \leq f_k + \delta(\frac{\sigma - 1}{L})(\nabla f(x_k)^T d_k)^2/||d_k||^2. \tag{30}$$

We now use definition $\cos(\theta_k)$ *to write this relation as:*

$$f_{k+1} \leq f_k + c\cos^2\theta_k||\nabla f(x_k)||^2, \tag{31}$$

where $c = \delta(\sigma - 1)/LSumming$ this expression and recalling that f is bounded below we obtain:

$$\sum \cos^2\theta_k||\nabla f(x_k)||^2 \prec \infty, \tag{32}$$

which concludes the proof.

Table 1: Comparison of different SR1-algorithms with different test functions and different dimensions

P.No.	n	SR1 algorithm *NI*	*NF*	BSR1 with $u_k = y_k$ *NI*	*NF*	BSR1 with $u_k = \nabla f(x_{k+1})$ *NI*	*NF*
1	2	4	34	10	28	4	11
2	2	13	53	12	49	3	30
3	2	15	73	13	68	8	44
4	2	2	27	2	27	2	27
5	3	34	162	15	82	7	20
6	3	16	51	17	75	5	15
7	3	2	4	2	-4	2	4
8	3	2	27	2	27	2	27
9	3	2	27	2	27	2	27
10	4	16	52	17	99	6	18
11	4	19	86	19	65	4	13
12	4	19	101	10	74	5	12
13	4	15	70	14	44	5	16
14	5	2	27	2	27	2	27
15	6	6	21	6	21	4	11
16	11	3	31	3	31	3	31
17	20	22	118	13	88	4	12
18	400	17	76	9	53	6	18
19	400	2	27	2	27	2	27
20	200	2	5	2	5	2	5
21	100	2	27	2	27	2	27
22	500	8	49	11	41	6	43
23	500	2	4	2	4	2	4
24	500	11	82	11	82	6	15
25	500	10	69	8	52	6	41
26	500	2	4	2	4	2	4
27	500	3	7	3	7	3	7
28	500	3	7	3	7	3	7
Total		235	1321	229	1145	103	543

4. Numerical result

Table 1 lists numerical results. In Table 1, “problem” and “n” stand for the test problem name and the“dimension”of the test problem, respectively. NI , NF are the total number of the“iterations, the function evaluations, respectively.

In this section, we report the detailed numerical result of a number of problems by Algorithm 2.1. Furthermore, the original SRI methods are given. All cods were written in Matlab code. we used unconstrained test problems form More Garbow and Hillstrom (see 4). We test the following three BEGS methods: $\sigma_1 = 0.1$, $\sigma_2 = 0.9$, $\epsilon = 10^{-5}$. We stop the iteration: if $|f(x_k)| \succ 10^{-5}$ let $stop1 = \frac{|f(x_k)-f(x_{k+1}|}{|f(x_k)|}$ Otherwise, let $stop1 = |f(x_k) - f(x_{k+1}|$. For every problem, if $||\nabla f(x_k)|| \leq \epsilon$ $stop1 \prec 10^{-5}$ or is satisfied, the program will be stopped. Himmeblau (see 13).

Problems numbers indicant for: "1. is the Froth, 2. is the Badscb, 3. is the Beale, 4. is the Jensam, 5. is the Helix, 6. is Bard, 7. is the Gauss, 8. is the Gulf, 9. is the Box, 10. is the Sing, 11. is the Wood, 12. is the Kowosb, 13. is the Bd. 14. is the Osb1, 15. is the Biggs, 16. is the Osb2, 17. is the Watson, 18. is the Singx, 19. is the Pen1, 20. is the Pen2, 21. is Vardim, 22. is the Trig, 23. is the Bv, 24. is the Trid, 25. is the Band, 26. is the Lin, 27. is the Lin1, 28. is the Lino".

5. Conclusions

We have presented the BSR1 method with a new quasi-Newton equation. Under suitable conditions, we proved that the"proposed method" is globally convergent. Some limited"numerical results" are also reported, which show that the BSR1 method is more efficient than the $SR1$ method. Commonly, we can compute the percentage performance of the new proposed algorithms $BSR1$ compared in opposition to the standard SR1 algorithm for the general tools NI and NF as follows:

Table 2: Relative efficiency of the new Algorithms

	SR1 algorithm	BSR1 with $u_k = y_k$	BSRI with $u_k = \nabla f(x_{k+1})$
NI	100	97.44	43.82
NF	100	86.67	41.10

References

[1] M. Farzin, H.M. Abu and J.L. Wah, *Improved Hessian approximation with modified secant equations for symmetric rank-one method*, Journal of Computational and Applied Mathematics, 235 (2011) 2423-2431.

[2] B. Fahimeh, A.H. Malik and J.L. Wah, *New quasi-Newton methods via higher order tensor models*, Journal of Computational and Applied Mathematics, 235 (2011), 2412-2422.

[3] D. Li, M. Fukushima, *A modified BFGS method and its global convergence in non convex minimization*, J. Comput. Appl. Math., 129 (2001), 15-35.

[4] J. More, B. Garbow, K. Hillstrome, *Testing unconstrained optimization software*, ACM Trans. Math. Software, 7 (1981), 17-41.

[5] J. Nocedal, *Theory of algorithms for unconstrained optimization*, Acta Numerica, 1 (1992), 199242.

[6] D. Powell, *Algorithms for Nonlinear Constraints that Use Lagrange Functions,*, Math. Programming., 14 (1978), 224-248.

[7] Y. Shengwei, N. Liangshuo, *An adaptive three-term conjugate gradient method based on self-scaling memoryless BFGS matrix*, Journal of Computational and Applied Mathematics, 2017, 1-27.

[8] P. Wolfe, *Convergence conditions for ascent methods*, SIAM Rev., 11 (1969), 226-235.

[9] P. Wolfe, P. Wolfe, *Convergence conditions for ascent methods*, II:Some corrections SIAM Review, 13 (1970), 185-188.

[10] Z. Wei, G. Li, L. Qi, *New quasi-Newton methods for unconstrained optimization problems*, Math. Program. Applied Mathematics and Computation., 175 (2006): 1156-1188.

[11] J.L. Wah, A.H. Malik, *Scaled memoryless symmetric rank one method for large-scale optimization*, Applied Mathematics and Computation, 218 (2011), 413-418.

[12] X. Yunhai, W. Zengxin, *A modified BFGS method without line searches for nonconvex unconstrained optimization*, Advances in theoretical and applied mathematics, 2006, 149-162.

[13] Y. Yuan and W. Sun, *Theory and Methods of Optimization*, Science Press of China, 1999.

[14] J.Z. Zhang, N.Y. Deng, L.H. Chen, *New quasi-Newton equation and related methods for unconstrained optimization,* Journal of Optimization Theory and Applications, 102 (1999), 147-167.

.

Accepted: 13.12.2018