Menemui Matematik (Discovering Mathematics)
Vol. 44, No. 1: 14-22 (2022)

Diagonal Quasi-Newton Updating Strategy with Cholesky Factor via Variational Principle

Tijjani Bukar[1,3], Wah June Leong[1,2], Mahani Marjugi[2], Chuei Yee Chen[2] and Hong Seng Sim[4]

[1]*Institute for Mathematical Research, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor*
[2]*Department of Mathematics, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor*
[3]*Department of Mathematics and Statistics Federal University Gashua Yobe State Nigeria,*
[4]*Department of mathematics and Actuarial Sciences, Universiti Tunku Abdulrahman Sungai Long Campus, Jalan Sungai Long 9, Bandar Sungai Long,4300 Kajang Selangor Malaysia*
[1]tijjanibukar1@gmail.com, [2]leongwj@upm.edu.my

ABSTRACT

The quasi-Newton method was popular due to the fact that only the gradient of the objective is required at each iterate and, since the second derivatives (Hessian) were not necessary, the quasi-Newton approach is often more efficient than the Newton method, especially when Hessian computation is costly. However, the method needed a full matrix storage that approximated the (inverse) Hessian. As a result, they might not be appropriate for dealing with large-scale problems. In this paper a diagonal quasi-Newton updating strategy is presented. The elements of the diagonal matrix approximating the Hessian were determined using the log-determinant norm satisfying weaker secant equation. To ensure the positive definiteness of the proposed diagonal updating matrices, their Cholesky factor will be considered within the variational problem. The corresponding variational problems are solved with the application of Lagrange multipliers approximated using Newton-Raphson method. Executable codes were developed to test the effectiveness and efficiency of the methods compared with some standard conjugate-gradient methods. Numerical results show that the proposed methods preforms better.

Keywords: Quasi-Newton methods, diagonal-updating strategy, trace and log-determinant norm, Cholesky factor, weak secant equation.

INTRODUCTION

In this paper we propose a new diagonal quasi-Newton (QN) updating strategy using Bard and Nocedal (1989), traces and log-determinant norm subject to Dennis and Wolkowicz (1993) weak secant equation. Therefore, we consider the problem of the form:

$$\min f(x), \tag{1}$$

where $f :\to R^n \to R$ is continuously differentiable of *n* variables which is assumed to be large. The first variable metric known as QN method was due to Davidon (1959), and later improved by (Fletcher & Powell, 1963). The method serves as an alternative to Newton's method. The Newton's method requires the determination of the Hessian matrix at every iteration which is known to be computationally expensive, while the QN- method requires only the computation of the gradient of the objective function at each iteration. On the other hand, the standard QN-method also has its peculiar problems. For instance, the method sometimes is affected by ill conditioning, and also may require full matrix storage which is not suitable for large scale problems. Starting from an initial point $x_0 \in \mathrm{R}^n$ and an initial approximation $B_0 \in \mathrm{R}^{n\times n}$ to the Hessian of the function f at x_0 symmetric and positive definite, the QN is an iterative of the form:

$$x_{k+1} = x_k + \alpha_k d_k, \quad \forall i = 1, \mathrm{L}\ , n \tag{2}$$

where $g_k = \nabla f(x_k)$ is the gradient vector of f at x_k and B_k^{-1} is the quasi-Newton approximation to the inverse Hessian $(\nabla^2 f(x_k))^{-1}$ at x_k . The step-size parameter $\alpha_k > 0$ is chosen by inexact line-search satisfying the Armijo, (1966) condition given by:

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k\ , \tag{3}$$

Where $\delta_k \in (0,1)$. To guarantee that the methodology includes correct curvature information then , B_{k+1} the update of B_k should satisfy the quasi-Newton equation,

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

Where $s_k = x_{k+1} - x_k$ and $y_k = g_{k+1} - g_k$. Many authors have tried to let B_k satisfy equation (4), but since it is normally very difficult to the quasi-Newton relation with a matrix of diagonal form to satisfy the relation (4), moreover, since our updating matrix is diagonal in order to maintain $O(n)$ floating point we let our B_k satisfy the Dennis and Wolkowictz, (1993) weak secant equation v which is given by:

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

Most popular among the quasi-Newton methods is the (BFGS) method which is developed (Broyden, 1970; Goldfarb, 1970; R. Fletcher, 1970; Shanno, 1970). According to Nocedal and Wright (2006), the method can be derived by solving the following system of unconstrained minimization problem

$$\begin{gathered} \min_H \| H_{k+1} - H_k \|_W \\ \text{s.t } H_{k+1} = H_{k+1}^T, \text{ and } H_{k+1} y_k = s_k \end{gathered} \tag{6}$$

Therefore, to obtain the update to the variational problem in (6), we impose additional condition that is the inverse matrix H_{k+1} is a symmetric positive definite satisfying the secant equation (4). Additionally, H_{k+1} supposed to be closer to the current matrix H_k under some Weighted Frobenius norm define by

$$\| P \|_W = \| W^{\frac{1}{2}} P W^{\frac{1}{2}} \|_F\ , \tag{7}$$

where $P = H_{k+1} - H_k$, and $\| . \|_F$ is define by $\| C \|_F^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} C_{i,j}^2$. The weighted norm W is any matrix satisfying $W y_k = s_k$. Then, the unique solution H_{k+1} give rise to the inverse Hessian update

$$H_{k+1} = (I - \rho_k y_k s_k^T) H_k (I - \rho_k s_k y_k^T) + \rho_k y_k y_k^T\ , \tag{8}$$

where $\rho_k = \dfrac{1}{y_k^T s_k}$.

Hence, working with B_k instead of H_k , the BFGS update for the Hessian approximation can be obtained by applying the Sherman Morrison formula, giving the (BFGS) method as

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k} \tag{9}$$

There are already been a list of achievements in the global convergence of the algorithm. Powell (1976) and Werner (1978) proved the global convergence properties of the method for unconstrained convex programme with large scale class of line search methods. However, there analysis did not cover the backtracking line searches. Byrd and Nocedal, (1989) introduced some mechanisms that makes the global convergence of the (BFGS) covered the backtracking strategies. They simplified the method by using the trace and log-determinant of the update matrix B_k and hence defined that for any positive definite matrix B_k the function,

$$\phi(B_k) = tr(B_k) - ln(det(B_k)), \tag{10}$$

Therefore, besides the weighted Frobenius norm, Byrd and Nocedal uses simultaneously the properties of trace and determinant of the matrix B_k to simplified the prove for the convergence of BFGS method. For this purpose they define for any positive matrix B_k on a strictly convex function, a measure of closeness between the matrix B_k and the identity matrix I such that $\phi I = 0$, where ln denotes the natural logarithm, and tr is the trace operator. Sim et al., (2018) proposed a variant of diagonal spectral gradient method using the trace and log-determinant norm while satisfying the weak secant equation of (5) . They proved the global convergence of the method under the backtracking line search with Armijo condition and it shows the method performs better than some conjugate gradient methods.

Motivated by these advances, we propose another least change secant updating strategy via the Byrd and Nocedal (1989) traces and log-determinant norm, satisfying the weak secant equation of Dennis and Wolkowicz (1993). To ensure positive definiteness of the proposed method, the Cholesky factor will be considered within the variational problem. The corresponding variational problems are solved with the application of Lagrange multipliers, approximated using the Newton-Raphson method. In Section 2, we present the derivation and mathematical formulation of the updating formulae while section 3 is the numerical results and comparison with some standard conjugate gradient methods. Finally, section is the conclusion and further research.

DIAGONAL UPDATING STRATEGY WITH CHOLESKY FACTORIZATION VIA VARIATIONAL PRINCIPLE.

Suppose that the update formula is given by B_k is diagonal and positive definite. Our main aim is to construct and define an updating formula B_{k+1} in such a way that it satisfies the weaker secant equation of (5). Taking into account the Cholesky factor of trace and log-determinant norm in (10), and letting $B_{k+1}^{\frac{1}{2}} = (b_{k+1,i}^{\frac{1}{2}}) \quad \forall i = 1, \ldots, n$.

Then, the problem is equivalent to the variational in component form given by

$$\min\Big(\sum_{i=1}^{n} b_{k+1,i}^{\frac{1}{2}}\Big) - \ln\Big(\prod_{i=1}^{n} b_{k+1,i}^{\frac{1}{2}}\Big) \tag{11}$$
$$\text{s.t} \quad \Big(\sum_{i=1}^{n} s_{k,i}^{2} (b_{k+1,i}^{\frac{1}{2}})^{2}\Big) - s_k^T y_k = 0$$

The Lagrangian function of the problem becomes

$$L = (b_{k+1,1}^{\frac{1}{2}} + b_{k+1,2}^{\frac{1}{2}} + \ldots + b_{k+1,n}^{\frac{1}{2}}) - ln(b_{k+1,1}^{\frac{1}{2}}.b_{k+1,2}^{\frac{1}{2}}.\ldots.b_{k+1,n}^{\frac{1}{2}}) - \mu\Big[\sum_{i=1}^{n} s_{k,i}^{2} (b_{k+1,i}^{\frac{1}{2}})^{2} - s_k^T y_k\Big] \tag{12}$$

where μ is the Lagrange multiplier associated with the constraints of (12). Thus in order to get the minimizer we differentiate the Lagrangian with respect to each $b_{k+1,i}^{\frac{1}{2}} \quad \forall i = 1,\ldots,n$ and setting the result equal to zero, leads to,

$$\frac{\partial L}{\partial b_{k+1,i}^{\frac{1}{2}}} = 1 - \frac{1}{b_{k+1,i}^{\frac{1}{2}}} - 2\mu b_{k+1,i}^{\frac{1}{2}}(s_{k,i}^2) = 0 \tag{13}$$

$$\Rightarrow -2\mu b_{k+1,i}^{\frac{1}{2}} s_{k,i}^2 - \frac{1}{b_{k+1,i}^{\frac{1}{2}}} + 1 = 0 \tag{14}$$

$$\Rightarrow 2\mu (b_{k+1,i}^{\frac{1}{2}})^2 s_{k,i}^2 - b_{k+1,i}^{\frac{1}{2}} + 1 = 0 \tag{15}$$

assuming $b_k^{\frac{1}{2}} \neq 0$. The quadratic in (15) is then solved to obtain

$$b_{k+1,i}^{\frac{1}{2}} = \frac{1+\sqrt{1-8\mu s_{k,i}^2}}{4\mu s_{k,i}^2}, \tag{16}$$

or

$$b_{k+1,i}^{\frac{1}{2}} = \frac{1-\sqrt{1-8\mu s_{k,i}^2}}{4\mu s_{k,i}^2}, \tag{17}$$

As suggested, the Lagrange multiplier μ, would be approximated by using the Newton-Raphson method where $\mu_0 = 0$ is the chosen initial guess but, in this place the situation changes because the initial trial guess at $(\mu_0 = 0)$ failed. Since $F(\mu)$ and $F'(\mu)$ are undefined at $\mu = 0$, we opt to do the approximation by letting $\mu = s_k^T y_k$ for the sake of simplicity and to ensure boundedness of $B_{k,i}$. Now let

$$\mu = s_k^T y_k \tag{18}$$

Subsequently, substituting (18) in (16) leads to

$$b_{k+1,i}^{\frac{1}{2}} = \Big(\frac{1+(1-8 s_k^T y_k s_{k,i}^2)^{\frac{1}{2}}}{4(s_k^T y_k) s_{k,i}^2}\Big). \tag{19}$$

Given our first updating formula in the first case as

$$B_{k+1}^I = \Big(\frac{1+(1-8 s_k^T y_k s_{k,i}^2)^{\frac{1}{2}}}{4(s_k^T y_k) s_{k,i}^2}\Big)^2, \forall i. \tag{20}$$

Following the same procedure, that is substituting (18) in (17) gives our second updating formula as

$$B_{k+1}^{II} = \Big(\frac{1-(1-8s_k^T y_k s_{k,i}^2)^{\frac{1}{2}}}{4(s_k^T y_k)s_{k,i}^2}\Big)^2, \forall i. \tag{21}$$

Now, the description of our Algorithm together with the line search is given as follows:

Algorithm 1. LDNCF 1 and LDNCF 2.

Step 1: Select an initial point $x_0 \in R^n$, $B_0 = I$ be an identity matrix, set $k = 0$ and choose the stopping criterion $\grave{o} > 0$.

Step 2: If $\| g_k \| < \grave{o}$stop. Else go to step 3.

Step 3: Compute $d_k = -B_k^{-1} g_k$.

Step 4: Find an acceptable step-length $\alpha_k > 0$ such that the Armijo backtracking line-search given by (3) is satisfied, with $\alpha = 1$as the first trial.

Step 5: Set $x_{k+1} = x_k + \alpha_k d_k$ and update B_{k+1} using (20) or (21).

Step 6: Set $k := k+1$ and go to step 2.

CONVERGENCE ANALYSIS

The convergence analysis of our methodologies is presented in this section and, before proceeding we make the following assumptions.

Assumption 1.

1. The level set $\Omega = \{x \in R^n : f(x) \le f(x_0)\}$ is convex.
2. The objective function $f \in C^2$ that is (at least twice continuously differentiable) bounded below and has a unique minimizer x^* in the level set Ω.
3. There exists positive constants $0 < M_1 < M_2$ such that :

$$M_1 \| z \|^2 \le z^T G(x) z \le M_2 \| z \|^2,$$

for all $z \in R^n$ and $x \in \Omega$.

Safeguarding strategies:

To ensure that our updating formulae are bounded, we propose the following safeguarding strategies.

Let $M_1 = 10^{-4}$ and $M_2 = 10^4$ then, B_k is updated by (20) or (21) , if:

$$M_1 \le B_{k+1} \le M_2 \quad \forall i = 1,\ldots,n$$

Otherwise,

$$B_{k+1,i} = \theta_{k,i} \text{ where, } \theta_{k,i} = \frac{s_k^T y_k}{s_k^T s_k}.$$

In general,

$$B_{k+1} = \left\{ \begin{array}{c} \text{(20 or 21) if } M_1 \le B_{k+1} \le M_2, \\ \text{Otherwise,} \\ \theta_{k,i} \text{where} \theta_{k,i} = \dfrac{s_k^T y_k}{s_k^T s_k} \end{array} \right\}$$

NUMERICAL EXPERIMENTS AND COMPARISONS

In this section, we present some numerical results from an implementation of our new proposed Ln (determinant norms) with Cholesky factor (LDNCF1) and (LDNCF2) diagonal quasi-newton methods for solving large scale unconstrained optimization problems. We therefore, evaluate the performance of our proposed (LDNCF1 and LDNCF2)-methods with some recent conjugate gradient (CG) methods. The methods used for the comparison including:

1. Fletcher and Reeves, FR (1964)
2. Polak ,Rebierre and Polyak, B.T PRP (1969)
3. Dai and Yuan, DY (1999)

All the experiments are implemented on a PC using Matlab 7.9.0 (R2015 a) with double precision Arithmetic. A total of 60 test functions are selected from Andrei (2008) test problems. For each test problem, we perform five numerical experiments with variables dimension ranging from 100 to 50,000 with Armijo (1966) line search condition from (3) used with the constant $\delta = 0.1.$ As regards to the stopping criteria used in our experiments for the algorithms, convergence is assumed when

$$\| g_k \| \le \epsilon, \tag{22}$$

is satisfied where $\epsilon = 10^{-4}$ is accepted as the standard convergence criteria. Hence we forced the algorithm to stop whenever the number of iteration exceeds 10,000 and declare the runs as failure. To compare the performance of our proposed (LDNCF1) and(LDNCF2) methods with the other CG-methods, we use the performance profile proposed by Dolan and Moré, (2002).

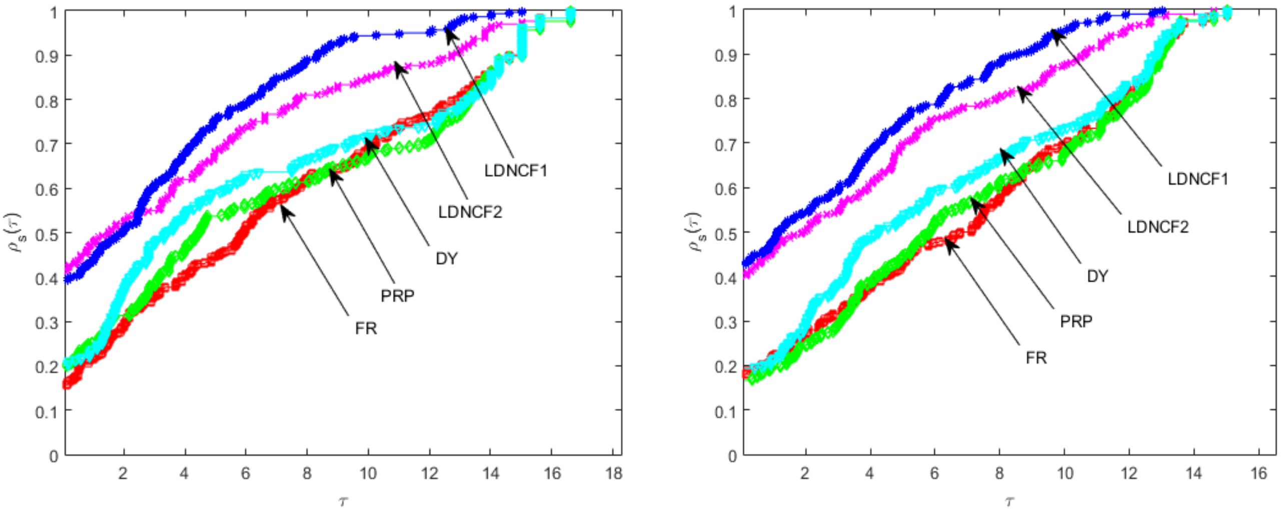

Fig.1: (a) Performance profile based on iterations for LDNCF1, LDNCF2 with DY, PRP and FR and (b) the performance profile based on function evaluation count for LDNCF1, LDNCF2 with DY, PRP and FR.

Figure 1 (a) shows the profiling graph of our proposed methods and the list of conjugate gradient methods in terms of the number of iterations. From Figure 1 (a), corresponding to the top curve in the profiling graph, it is obvious that our proposed LDNCF1 and LDNCF2 methods perform better than the conjugate gradient methods. In other words, our proposed methods require fewer iterations to get the desired minimum points. Figure 1(b) shows the profiling graph of our proposed methods and the list of conjugate gradient methods in terms of the number of function calls. Therefore, our proposed LDNCF1 and LDNCF2 methods confirm that our number of function calls is lower than the compared methods.

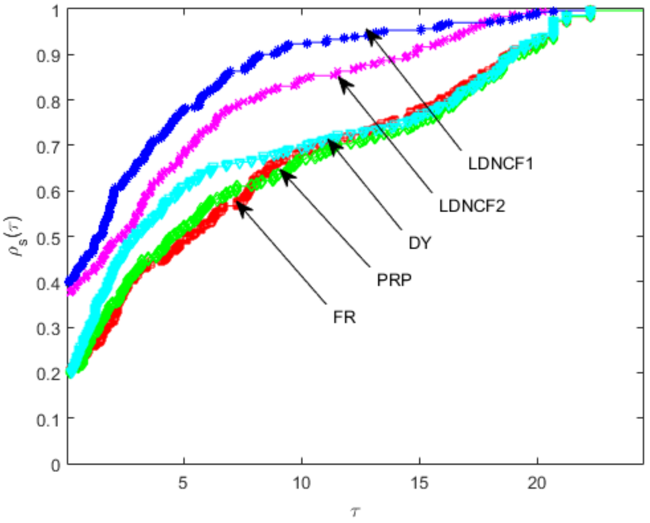

Fig. 2: Performance profile based on CPU-time for LDNCF1, LDNCF2 with DY, PRP and FR.

Figure 2 shows the profiling graph of our proposed method and the list of conjugate gradient methods in terms of CPU time, and from this figure, the graphs of the LDNCF1 and LDNCF2 methods appear to be the top curves of the profiling graph rather than the list of conjugate gradient methods, which indicates that our proposed algorithms are faster than the CG methods.

CONCLUSION

We present a new class of diagonal quasi-Newton (DQN) method in the form of Traces and log(determinant) norm. in this method, we consider the Cholesky factor or the square-root of $B_k^{\frac{1}{2}}$ to the updating diagonal Hessian matrix $B_{k+1}^{\frac{1}{2}}$. Our purpose is to determine that the updating diagonal Hessian matrix B_{k+1}, preserved positive definite. The outcome from the numerical testing confirm our claims that our proposed strategies can improve the performance of gradient based algorithms. Our safe guarding strategies are simple, inexpensive, and robust than the other compared CG-Methods. Therefore, we conclude that our numerical updating strategy provides good alternative to some of the other existing methods.

ACKNOWLEDGEMENT

This article was supported by the Fundamental Research Grant Scheme (FRGS) under Ministry of Higher Education with project number FGRS/1/2019/STG06/UPM/02/14.

REFERENCES

Andrei, N. (2008). An unconstrained optimization test functions collection. *Advanced Modelling and Optimization*, *10*(1), 147–161.

Armijo, L. (1966). Minimization of functions having lipschitz continuous first partial derivatives. *Pacific Journal of Mathematics*, *16*(1), 1–3.

Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms 1. General considerations. *J . Inst. Maths Appliications*, *76*(6), 76–90.

Byrd, R. H. and, & Nocedal, J. (1989). A tool for the analysis of quasi-Newton methods with application to unconstrained minimization. *SIAM Journal on Numerical Analysis*, *26*(3), 727–739.

Dai, Y. H. and, & Yuan, Y. (1999). A nonlinear conjugate gradient method with a strong global convergence property. *SIAM Journal on Optimization*, *10*(1), 177–182.

Davidon, C. W. (1959). Variable metric method for minimization. *Argonne National Labora atory*, *5990*, 1–32.

Dennis, J. J. E. and, & Wolkowictz, H. (1993). Sizing and least-change secant methods. *SIAM Journal of Numerical Analysis*, *30*(5), 1291–1314.

Dolan, E. D. and, & Moré, J. J. (2002). Benchmarking optimization softwire performance profile. *SIAM Journal on Optimization.*, *10*(2), 201–213.

Fletcher, R. and, & Reeves, C. M. (1964). Function minimization by conjugate gradients. *The Computer Journal*, *7*(2), 149–154.

Fletcher, R., & Powell, M. . J. D. (1963). A rapidly convergent descent method for minimization. *The Computer Journal*, *6*(2), 163–168.

Goldfarb, D. (1970). A Family of Variable-Metric Methods Derived by Variational Means. *Mathematics of Computation*, *24*(109), 23–26.

Nocedal, J. and, & Wright, S. (2006). *Numerical Optimization* (2nd Edition). Springer Science+Business Media, LLC New York.

Polyak, B. T. (1969). The conjugate gradient method in extremal pproblems. *Researchgate*, *5553*(4), 92–112.

Powell, M. J. D. (1976). “Some global convergence properties of a variable metric algorithm for minimization without exact line searches”, *Nonlinear Programming, SIAM-AMS Proceedings, American Mathematical Society, 4*(9), 53–72.

R. Fletcher. (1970). A new approach to variable metric algorithms. *The Computer Journal*, *13*(3), 317–322.

Seng Sim, H., Leong, W. J., Chen, Y. C. and, & Siti, N. I. I. (2018). Multi-Step Spectral Gradient Methods with Modified Weak Secant Relation for Large-scale Unconstrained Optimization. *Numerical Algebra Control and Optimization.*, *8*(3), 377–387.

Shanno, D. F. (1970). Conditioning of Quasi-Newton Methods for Function Minimization. *Mathematics of Computation*, *24*(111), 647.