Int. J. Open Problems Compt. Math., Vol. 4, No. 1, March 2011
ISSN 1998-6262; Copyright © ICSRS Publication, 2011
www.i-csrs.org

Some Variational Problems Implemented By The Least Square Method

Abbas Y. AL-Bayati and Basim A. Hassan

Department of Mathematics, College of Computers Sciences and Mathematics
University of Mosul, IRAQ
e-mail: profabbasalbayati@yahoo.com

Abstract

In this paper, we have constructed some new variational problems whose optimal solutions coincide with the quasi-Newton update matrices. Different measures lead to different formulas. The new updating problems are based on the quasi-Newton condition. These new variational problems may be useful in suggesting new quasi-Newton updates in the future.

Keyword: *Quasi Newton, Least Square, Variational Problems, Quasi Newton Condition.*
2010 Mathematics Subject Classification: 90C53, 93E24.

1. Introduction

The main point of this paper is some variational problems in Quasi-Newton (QN) methods. By working directly in the space of symmetric matrices, we handle the symmetry constraint implicitly [2]. This approach simplifies both the formulations and the proofs given in the literature, such as the ones given in [3] and [4]. In Quasi-Newton methods for unconstrained minimization, an approximation of the true Hessian is used in the Newton step. Then, at each iteration, the approximation is updated to reflect the information coming form the new point or the gradient at the new point. There are two popular and successful updates, Biggs [1] and Oren [5] scaled Variable Metric (VM) updates. In these updates, the matrices B_k

represent an approximation for the Hessian matrix while the matrices H_k represent an approximation for the inverse Hessian matrix. Superscripts Biggs and Oren will be used to refer to the corresponding updates. For easy reference, these formulas are presented below in two equivalent formulas:

$$B_{k+1}^{Oren} = B_k - \frac{y_k s_k^T B_k + B_k s_k y_k^T}{s_k^T y_k} + \frac{y_k y_k^T}{s_k^T y_k}\left[\xi + \frac{s_k^T B_k s_k}{s_k^T y_k}\right] \qquad \ldots\ldots\ldots(1a)$$

$$B_{k+1}^{Biggs} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{1}{\vartheta}\frac{y_k y_k^T}{s_k^T y_k} \qquad \ldots\ldots\ldots(1b)$$

$$H_{k+1}^{Oren} = \left\{H_k - \frac{H_k s_k s_k^T H_k}{y_k^T H_k y_k}\right\}\frac{1}{\xi} + \frac{s_k s_k^T}{s_k^T y_k} \qquad \ldots\ldots\ldots(2a)$$

$$H_{k+1}^{Biggs} = H_k - \frac{H_k y_k s_k^T + s_k y_k^T H_k}{s_k^T y_k} + \frac{s_k s_k^T}{s_k^T y_k}\left[\vartheta + \frac{y_k^T H_k y_k}{s_k^T y_k}\right] \qquad \ldots\ldots\ldots(2b)$$

where

$$\xi = s_k^T y_k / y_k^T H_k y_k \;,\quad s_k = x_{k+1} - x_k \text{ and } y_k = g_{k+1} - g_k \qquad \ldots\ldots\ldots(3a)$$

and

$$\vartheta = \frac{1}{t_k} \;,\quad t_k = \frac{6}{s_k y_k}\left[f(x_k) - f(x_{k+1}) + s_k^T g_{k+1}\right] - 2 \;. \qquad \ldots\ldots\ldots(3b)$$

This paper is organized as follows: in Section 2, we formulate various norm minimization problems in quasi-Newton methods as least squares problems in S^n, the vector space $n*n$ of symmetric matrices. In Section 3, we formulate the Dual of these least squares problems showing that the optimal solutions to the Dual problems are also produce quasi-Newton update formulas. In Section 4, a general conclusion has been given while an open problem is listed in Section 5.

2. Preliminaries

2.1 QN-methods with least squares

It is well-known that the approximation to the Hessian matrices in various QN methods are updated using the solution of some optimization problems. In the optimization problems, for obtaining the Oren and Biggs updates, the constraints have the form

$$\left\{X \in R^{n*n} : Xs = y,\; X^T = X\right\} \qquad \ldots\ldots\ldots(4)$$

where, s and y are given vectors in R^n. The first affine equation $Xs = y$ is called the secant equation (or Quasi-Newton equation), and the constraint $X^T = X$ is included since a Hessian matrix (or its inverse) is always symmetric, and so should be any approximations to it. Now, we solve such problems in a simpler way, we first noted that the symmetry constraints can be eliminated entirely simply by working in the vector space S^n of $n*n$ symmetric matrices instead of the vector space R^{n*n}. We endow the vector space R^{n*n} with the trace inner product.

$$\langle X,Y\rangle = tr(X^T,Y) = \sum_{i,\,j=1}^{n} X_{ij}Y_{ij} \ , \qquad \text{.........}(5a)$$

which induces the trace inner product

$$\langle X,Y\rangle = tr(XY) \qquad \text{.........}(5b)$$

in S^n. Both vector spaces become Euclidean spaces with these inner products see [6].

For the least squares method, the objective function f has the following special form:

$$f(x) = \frac{1}{2}\sum_{k=1}^{m} r_k^2(x) \qquad \text{.........}(6a)$$

where each r_k is a smooth function from R^n to R. Least squares problems arise in many areas of applications, and may be the largest source of unconstrained optimization problems. To see why the special form of f often makes least square problems easier to solve than the general unconstrained minimization problems, we assemble the individual components r_k from $(6a)$ into a residual vectors $r : R^n \to R^m$ as follows

$$X(x) = (r_1(x), r_2(x). \ . \ . \ . \ . \ ., r_m(x))^T \qquad \text{.........}(6b)$$

Using this notation, we can rewrite f as

$$f(x) = \frac{1}{2}\|X\|^2 . \qquad \text{.........}(6c)$$

For more details see [7].

Lemma 2.2: Let $s, y \in R^n$, $s \neq 0$. Consider the affine subspace $\mathrm{A} = \{X \in S^n : Xs = y\}$ in the vector space S^n. The linear subspace corresponding to A is $\Psi = \{X \in S^n : Xs = 0\}$. Let $\{u_k\}_1^n$ be a basis of R^n, and define the matrices $S_k = su_k^T + u_k s^T$, $k = 1, \ldots\ldots\, n$. The matrices $\{S_k\}_1^n$ are linearly independent and Ψ is the intersection of n hyper planes in S^n, i.e.

$$\Psi = \{X \in S^n : \langle X, S_k \rangle = 0,\ k = 1, \dots\dots\dots, n\}. \qquad \dots\dots\dots(7)$$

Moreover,

$$\Psi^\perp = span\{S_1, \dots\dots\dots \dots, S_n\} = \{s\lambda^T + \lambda s^T : \lambda \in R^n\}. \qquad \dots\dots\dots(8)$$

Proof : The formula for Ψ is obvious. Notice that the equation $Xs = 0$ in R^{n*n} is equivalent to the component equations.

$$\langle X, su_k^T \rangle = 0,\ k = 1, \dots\dots\dots \dots\dots, n \qquad \dots\dots\dots(9)$$

Since

$$0 = \langle u_k, Xs \rangle = tr(u_k^T Xs) = tr(Xsu_k^T) = \langle X, su_k^T \rangle,\ k = 1, \dots\dots\dots \dots\dots, n \qquad \dots.(10)$$

Since X is symmetric , we also have

$$\langle X, u_k s^T \rangle = tr(Xu_k s^T) = tr(su_k^T X) = tr(u_k^T Xs) = tr(Xsu_k^T) = \langle X, su_k^T \rangle, \qquad \dots.(11)$$

So that the equation $Xs = 0$ is also equivalent to be the component equations

$$\langle X, u_k s^T \rangle = 0, \quad k = 1, \dots\dots\dots \dots, n. \qquad \dots\dots\dots(12)$$

Consequently, $\Psi \subseteq S^n$ can be written as an intersection of n hyper planes

$$\langle X, S_k \rangle = 0, \quad k = 1, \dots\dots\dots \dots\dots, n. \qquad \dots\dots\dots(13)$$

The formula $\Psi^\perp = span\{S_1, \dots\dots\dots \dots, S_n\}$ follows immediately; and the linear combination $\sum_{k=1}^{n} \delta_k S_k \in \Psi^\perp$ can be written as:

$$\sum_{k=1}^{n} \delta_k (su_k^T + u_k s^T) = s\lambda^T + \lambda s^T \qquad \dots\dots\dots(14)$$

where

$$\lambda = \sum_{k=1}^{n} \delta_k u_k \qquad \dots\dots\dots(15)$$

The matrices $\{S_k\}_1^n$ are linearly independent: the equation

$$0 = \sum_{k=1}^{n} \delta_k S_k = s\lambda^T + \lambda s^T \qquad \dots\dots\dots(16)$$

gives

$$0 = (s^T \lambda)\lambda + \|\lambda\|^2 s, \qquad \dots\dots\dots(17)$$

and taking the inner product of both sides with s yields

$$(s^T \lambda)^2 + \|\lambda\|^2 . \|s\|^2 = 0 . \qquad \dots\dots\dots(18)$$

Thus

$$\|\lambda\|^2 . \|s\|^2 = 0 , \qquad \text{.........(19)}$$

and since $s \neq 0$, we have $\lambda = 0$. Since $\lambda = \sum_{k=1}^{n} \delta_k u_k = 0$ and $\{u_k\}_1^n$ is a basis of R^n, we have

$$\delta_k = 0,\ k-1, \text{.........} \text{......}, n . \qquad \text{.........(20)}$$

Most of the variational problems encountered in VM-methods are closely related to the following generic least squares problem see [6].

Theorem 2.3: The solution to the minimization problem in S^n

$$\min \ \frac{1}{2}\|X\|^2 \qquad \text{.........(21)}$$

$$S.t. \quad Xs = \rho y \ , \ \rho \text{ is any arbitrary scalar} \qquad \text{.........(22)}$$

is given by

$$\bar{X} = \frac{sy^T + ys^T}{\langle s,s\rangle} - \frac{\langle y,s\rangle}{\langle s,s\rangle^2} ss^T \ . \qquad \text{.........(23)}$$

Proof : Define

$$f(X) = \|X\|^2 = \langle X, X\rangle / 2 \ . \qquad \text{.........(24)}$$

We have $\nabla f(X) = X$ and

$$\nabla^2 f(X) = I, \qquad \text{.........(25)}$$

and the function f is convex. Consequently, Lemma 2.2 implies that $\bar{X}$ is characterized by the equation

$$\bar{X} = \lambda s^T + s\lambda^T \ , \qquad \text{.........(26)}$$

for some $\lambda \in R^n$. We have

$$\langle \rho y, s\rangle = \langle \rho \bar{X} s, s\rangle = \langle \rho[\lambda s^T + s\lambda^T]s, s\rangle = 2\rho\langle \lambda, s\rangle\|s\|^2 \ , \qquad \text{.........(27)}$$

and

$$\langle \lambda, s\rangle = \langle \rho y, s\rangle \big/ (2\rho\|s\|^2) = \langle y, s\rangle \big/ (2\|s\|^2) . \qquad \text{.........(28)}$$

Substituting this in the equation

$$\rho y \quad = \rho \bar{X} s$$
$$\rho y \quad = \rho\langle \lambda, s\rangle s + \rho\|s\|^2 \lambda$$
$$\rho\|s\|^2 \lambda = \rho y - \rho\langle \lambda, s\rangle s$$
$$\rho\|s\|^2 \lambda = \rho y - \rho \frac{\langle \rho y, s\rangle}{2\rho\|s\|^2} s$$

gives

$$\lambda = \frac{1}{\|s\|^2} y - \frac{\langle y, s\rangle}{2\|s\|^4} s \ . \qquad \text{………(29)}$$

Finally, substituting this in

$$\bar{X} = \lambda s^T + s\lambda^T \qquad \text{………(30)}$$

gives equation (23).

Corollary 2.4: Let $X_0 \in S^n$ be given and Q be symmetric positive definite weight matrix. The optimal solution $\bar{X}$ to the minimization problem in S^n

$$\min \ \frac{1}{2}\left\|Q^{1/2}(X - X_0)Q^{1/2}\right\|^2 \qquad \text{………(31)}$$

$$S.t. \quad Xs = \rho y \ \ , \ \rho \text{ is any arbitrary scalar} \qquad \text{………(32)}$$

is given by

$$\bar{X} = X_0 + \frac{Q^{-1}s(\rho y - X_0 s)^T + (\rho y - X_0 s)s^T Q^{-1}}{\langle s, Q^{-1}s\rangle} - \frac{\langle \rho y - X_0 s, s\rangle}{\langle s, Q^{-1}s\rangle^2} Q^{-1}ss^T Q^{-1} \quad \text{…(33)}$$

Proof : With the following changes of the variables

$$\hat{X} = Q^{1/2}(X - X_0)Q^{1/2} \ , \quad \hat{y} = Q^{1/2}(\rho y - X_0 s) \ , \quad \hat{s} = Q^{-1/2} s \ , \qquad \text{………(34)}$$

The problem (31) reduces to problem (21). After substituting the expressions for $\hat{X}, \hat{y},$ *and* $\hat{s}$ into (23), we multiply the resulting equality by $Q^{-1/2}$ from both sides to get the desired expression for $\bar{X}$.

Corollary 2.5: (**New result**): The Oren update matrix B_{k+1} in $(1a)$ is the optimal solution to the minimization problem in S^n

$$\min \quad \frac{1}{2}\left\| Q^{1/2}(B - B_k)Q^{1/2} \right\|^2 \qquad \text{.........(35)}$$

$$S.t. \quad Bs_k = \xi\, y_k\,,\ \xi \text{ is defined in } (3a) \qquad \text{.........(36)}$$

where $B_k \in S^n$ is an approximation to the Hessian of f at iteration k,

$$y = \nabla f(x_{k+1}) - \nabla f(x_k)\,,\ s_k = x_{k+1} - x_k \qquad \text{.........(37)}$$

and Q is any symmetric positive definite matrix satisfying $Qy_k = s_k$.

Proof : Using Corollary 2.4 , we have

$$B_{k+1} = B_k + \frac{Q^{-1}s(\rho y - B_k s)^T + (\rho y - B_k s)s^T Q^{-1}}{\langle s, Q^{-1}s\rangle} - \frac{\langle \rho y - B_k s, s\rangle}{\langle s, Q^{-1}s\rangle^2} Q^{-1}ss^T Q^{-1} \quad \text{...(38)}$$

The requirement

$$Qy_k = s_k \text{ or } y_k = Q^{-1}s_k \qquad \text{.........(39)}$$

simplifies the above expression and makes B_{k+1} independent of Q. It is easy to verify that the resulting formula for B_{k+1} is the same as the one obtained by expanding $(1a)$.

Corollary 2.6: (**New result**): The Biggs update matrix H_{k+1} in $(2b)$ is the optimal solution to the minimization problem in S^n

$$\min \quad \frac{1}{2}\left\| Q^{1/2}(H - H_k)Q^{1/2} \right\|^2 \qquad \text{.........(40)}$$

$$S.t. \quad Hy_k = \vartheta s_k\,,\ \vartheta \text{ is defined in } (3b) \qquad \text{.........(41)}$$

where s_k and y_k are defined as in Corollary 2.5, $H_k \in S^n$ is a matrix approximating the inverse Hessian of f at iteration k, and Q is any symmetric positive definite matrix satisfying $Qs_k = y_k$.

Proof: The proof is similar to the proof of Corollary 2.5.

3. QN-methods with dual least squares problems

In this section, we have proposed two dual problems for the least squares minimization problem. The dual problem turns out to be minimization problem whose solution coincides with the Oren and Biggs update formulas.

The following, theorem 3.1 and lemma 3.2, are given in [2,6].

Theorem 3.1 Let $x_0, y_0 \in E$ be given point in a Euclidean space E, and let $V \subseteq E$ be a linear subspace of E. The following least squares problems are duals of each other. Furthermore, they have the same optimal solution.

$$(P) \quad \min \frac{1}{2}\|x - x_0\|^2 \quad \text{......}(41a) \qquad (D) \quad \min \frac{1}{2}\|y - y_0\|^2 \quad \text{......}(42a)$$

$$x \in y_0 + V \quad \text{......}(41b) \qquad y \in x_0 + V^{\perp} \quad \text{......}(42b)$$

Let $Q \in S^n$ be a positive definite matrix. Consider W-norm on S^n given by

$$\|X\|_W^2 = tr(Q^{1/2} X Q^{1/2})^2 = tr(QXQX), \qquad \text{.........}(43)$$

and the corresponding inner-product

$$\langle X, Y\rangle = tr(QXQY) = tr((Q \otimes Q)XY) = \langle (Q \otimes Q)X, Y\rangle, \qquad \text{.........}(44)$$

Where $(Q \otimes Q)X = QXQ$ .In the Euclidean space $(S^n, \| \, . \, \|_W)$, the problem (35) becomes :

$$\min \quad \frac{1}{2}\|B - B_k\|_W^2 \qquad \text{.........}(45)$$

$$S.t. \quad Bs_k = \xi\, y_k \qquad \text{.........}(46)$$

to which Theorem (3.1) applies. Let $\bar{B}$ be any matrix in the affine constraint set $\mu = \{B \in S^n : Bs_k = \xi\, y_k\}$. Then $\mu = \bar{B} + \Psi$ where $\Psi = \{B : Bs_k = 0\}$. In order to determine the dual problem in this setting, we need to compute the orthogonal complement of Ψ. This is done in the lemma below, which is an analogue of Lemma 2.2.

Lemma 3.2 Let $Q \in S^n$ be a positive definite matrix and $0 \neq s \in R^n$. The orthogonal complement of the linear subspace $\Psi = \{B : Bs = 0\}$ in the Euclidean space $(S^n, \| \, . \, \|_W)$ is

$$\Psi^{\perp} = \{\lambda (Q^{-1}s)^T + (Q^{-1}s)\lambda^T : \lambda \in R^n\}. \qquad \text{.........}(47)$$

Lemma 3.3: (**New result**): The matrix B_{k+1} in the Oren update formula $(1.a)$ is the optimal solution to the least squares problem

$$\min \quad \left\{ \frac{1}{2}\left\| B_k - \xi \bar{B} + \lambda y_k^T + y_k \lambda^T \right\|_W^2 : \lambda \in R^n \right\}, \qquad \text{.........}(48)$$

where y_k, s_k, B_k, *and* Q satisfy the conditions in Corollary 2.5, and $\bar{B} \in S^n$ is any matrix satisfying the secant equation $\bar{B} s_k = \xi y_k$. In particular, we may choose

$$\bar{B} = \xi \frac{y_k y_k^T}{\langle s_k, y_k \rangle} \;, \text{ where } \xi = \frac{s_k^T y_k}{y_k^T H_k y_k} \qquad \text{.........(49)}$$

Proof : The affine constraint set in (45) is $\bar{B} + \Psi$ where $\Psi = \{B : Bs_k = 0\}$, and we have $Q^{-1} s_k = y_k$. The proof is follows immediately from Theorem 3.1 and Lemma 3.2.

Similarly, we also have

Lemma 3.4: (**New result**): The matrix H_{k+1} in the Biggs update formula (2*b*) is the optimal solution to the least squares problem

$$\min \left\{ \frac{1}{2} \left\| H_k - \vartheta \bar{H} + \lambda s_k^T + s_k \lambda^T \right\|_W^2 : \lambda \in R^n \right\} , \qquad \text{.........(50)}$$

where y_k, s_k, H_k, *and* Q satisfy the conditions in Corollary 2.6, and $\bar{H} \in S^n$ is any matrix satisfying the secant equation $\bar{H} s_k = \vartheta y_k$. For more details of these parameters see [8, 9]. In particular, we may choose

$$\bar{H} = \vartheta \frac{s_k s_k^T}{\langle s_k, y_k \rangle} \quad \text{where} \quad \vartheta = \frac{1}{t_k} \;, \qquad \text{.....(51)}$$

$$t_k = \frac{6}{s_k y_k} \left[f(x_k) - f(x_{k+1}) + s_k^T g_{k+1} \right] - 2 .$$

Proof: is similar to the proof of lemma 3.3.

4. Conclusions

In this paper, we have scaled the Hessian by a scalar ρ, for which the QN-like condition ($H_{k+1} y_k = \rho_k s_k$) is satisfied for two different values of the parameter ρ_k .The main result of this paper is to show that these formulas gives the optimal solution to the least squares problem defined in (21) and the Dual least squares problem defined in (46) and (48). Moreover, the condition of the theorem 2.2 ensures that the resulting updated matrices B_{k+1} and H_{k+1} are positive definite.

5. Open Problem

The two VM updates for Oren and Biggs can be further generalized as:

$$H_{k+1}^{Oren} = \Psi\left\{H_k - \frac{H_k s_k s_k^T H_k}{y_k^T H_k y_k}\right\} + \Gamma\left[\frac{s_k s_k^T}{s_k^T y_k}\right] \quad \text{………(52)}$$

$$H_{k+1}^{Biggs} = \Phi\left\{H_k - \frac{H_k y_k s_k^T + s_k y_k^T H_k}{s_k^T y_k}\right\} + \Omega\left\{\frac{s_k s_k^T}{s_k^T y_k}\left[1 + \frac{y_k^T H_k y_k}{s_k^T y_k}\right]\right\} \quad \text{………(53)}$$

where Ψ, Γ, Φ and Ω are selected in such away so that they satisfy the Quasi Newton like condition. These parameters will generalized both Oren and Biggs to improve the performance of the standard variable metric update. These search directions under exact line searches will produce mutually conjugate directions which are equivalent with the standard conjugate directions using the standard quadratic functions [10].

References

[1] M.C. Biggs, A Note Of Minimization Algorithm Which Make Use Of Non-Quadratic Properties Of The Objective Function, Journal of Institute of Mathematics and its application. (1973) 337-338.

[2] R. Courant and D. Hilbert, Methods Of Mathematical Physics, Inter science publishers, Inc., New York, Vol. I. (1953).

[3] B.A. Hassan, A New Theoretical Result For Quasi-Newton Formulae For Unconstrained Optimization, Raf. of comp. and Math. 2(2) (2005) 41-46.

[4] R. Fletcher, A New Variation Result For Quasi-Newton Formulae, SIAM J. Optimization., 1(1) (1991) 18-21.

[5] S.S. Oren, On The Selection of Parameter In Self-Scaling Variable Metric Algorithm, Mathematical Programming., Vol. 7 (1974) 351-367.

[6] G. Osman, G. Filiz and S. Olena, Variational Problems In Quasi-Newton Methods Department of Mathematics and Statistics, University of Maryland Baltimore, County, Maryland 21250, USA, (2006) 1-22.

[7] J. Nocedal and J. Wright, Numerical Optimization, Springer Series In Operations Research, Springer Verlag, New York, USA, (2006).

[8] X. Wei, and Z. Wang, "A limited memory BFGS-type method for large-scale unconstrained optimization," *Computers & Mathematics with Applications*, Vol. 56(4), (2008), 1001–1009.

[9] L. Liu, S. Yao, and Z. Wei, The global and super-linear convergence of a new non-monotone MBFGS algorithm on convex objective functions, Journal of Computational and Applied Mathematics, Vol. 220(1-2), (2008), 422-438.

[10] N. Andrei, "Another nonlinear conjugate gradient algorithm for unconstrained optimization," *Optimization Methods & Software*, Vol. 24(1), (2009), 89–104.