Int. J. Open Problems Compt. Math., Vol. 4, No. 3, September 2011
ISSN 1998-6262; Copyright © ICSRS Publication, 2011
www.i-csrs.org

Variational Nonlinear Problems with Trace Determinant Functions in Variable Metric Algorithms

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

This paper is devoted to investigate a general variational approach for the Variable Metric (VM) updates. Al-Bayati [1] and Oren [6] algorithms can be obtained from the minimization of the trace determinant function $\psi(X)$. The positive definiteness property of the optimal solution is guaranteed by the nature of $\psi(X)$. Finally, it has been shown that the Al-Bayati [1] and Oren [6] updating formulas satisfy the least change property with respect to this new measure.

Key Words: *Variational Approach, Variable Metric (Quasi-Newton) Updates, Trace Determinant Function.*
2010 Mathematics Subject Classification: 35A15, 35J20, 47J20, 90C53, 11F72

1 Introduction

In the optimization problems, for obtaining Al-Bayati [1] and Oren [6] updates, the constraints must have the form:

$$\left\{X \in R^{n*n} : Xv = y,\ X^T = X\right\} \qquad \text{..........(1)}$$

where v and y are given vectors in R^n. The first affine equation $Xv = y$ is called the secant equation or (Quasi-Newton equation), and the constraint $X^T = X$ is included since the Hessian matrix (or its inverse) is always symmetric, and so should be any approximations to it. We endow the vector space R^{n*n} with the trace inner product:

$$\langle X,Y\rangle = tr(X^T,Y) = \sum_{i,j=1}^{n} X_{ij}Y_{ij} \ , \qquad \text{..........(2)}$$

which induces the trace inner product

$$\langle X,Y\rangle = tr(XY) \qquad \text{..........(3)}$$

in S^n. Both vector spaces become Euclidean spaces with these inner products see [7].

The function $\psi : R^{n*n} \rightarrow R$ defined by :

$$\psi(X) = trX - In\det X = \langle I, X\rangle - In\det X \qquad \text{..........(4)}$$

is used by Byrd and Nocedal [3] in the convergence analysis of the BFGS update rule. Later, Hassan [4] and Al-Bayati and Hassan [2] show that Al-Bayati [1] and Oren [6] update rules can be obtained from the minimization of function $\psi(X)$ subject to the same constraints as in the least squares minimization case.

Lemma 1.1:
For nonsingular X, the derivative of $\det(X)$ is given by $d(\det(X))/d(a_{ij}) = (X^{-1})_{ji}\det(X)$. For the proof and more details see [4,5]

Lemma 1.2:
Let $v, y \in R^n$, $v \neq 0$. Consider the affine subspace $A = \{X \in S^n : Xv = y\}$ in the vector space S^n. The linear subspace corresponding to A is $\Psi = \{X \in S^n : Xv = 0\}$. Let $\{u_k\}_1^n$ be a basis of R^n, and define the matrices $S_k = vu_k^T + u_k v^T$, $k = 1,.........n$. The matrices $\{S_k\}_1^n$ are linearly independent and Ψ is the intersection of n hyper planes in S^n, i.e.

$$\Psi = \{X \in S^n : \langle X, S_k\rangle = 0,\ k = 1,........,n\}. \qquad \text{..........(5)}$$

Moreover,

$$\Psi^{\perp} = span\{S_1,...........,S_n\} = \{v\lambda^T + \lambda v^T : \lambda \in R^n\}. \qquad \text{..........(6)}$$

For the proof and more details see [7].

2 Preliminaries

2.1 Trace determinant function for minimization problems in Quasi-Newton methods:

A variational result for the Al-Bayati [1] and Oren [6] updating formulas:

$$H_{k+1}^{\text{Al-Bayati}} = H_k - \frac{H_k y_k v_k^T + v_k y_k^T H_k}{v_k^T y_k} + \frac{v_k v_k^T}{v_k^T y_k}\left[\vartheta_k + \frac{y_k^T H_k y_k}{v_k^T y_k}\right] \qquad \text{..........(7)}$$

$$B_{k+1}^{Oren} = B_k - \frac{y_k v_k^T B_k + B_k v_k y_k^T}{v_k^T y_k} + \frac{y_k y_k^T}{v_k^T y_k}\left[\xi_k + \frac{v_k^T B_k v_k}{v_k^T y_k}\right] \qquad \text{..........(8)}$$

where

$$\vartheta_k = y_k^T H_k y_k / v_k^T y_k \quad \text{and} \quad \xi_k = v_k^T B_k v_k / v_k^T y_k \qquad ....(9)$$

Occupies a central roles in unconstrained optimization. Here v_k and y_k denoted certain difference vectors of the points and gradients respectively occurring on iteration k of Quasi-Newton, with $v_k^T y_k > 0$, [5]. B_k represents approximations of the Hessian and H_k represents approximations of the inverse Hessian[4]. The main results of this paper is to show that these two formulas also satisfy the minimum property with respect to the measure function ψ of Byrd and Nocedal defined in (4).

Theorem 2.1:

Let the affine set $\{X \in R^{n*n} : Xv = y\}$ contain a positive definite matrix. The solution $\bar{X}$ to the minimization problem in S^n

$$\min \ \psi(\mu X) = \langle I, \mu X \rangle - In \det \mu X \qquad ....(10)$$

$$\text{s.t. :} \quad Xv = y \qquad ....(11)$$

satisfies

$$\bar{X}^{-1} = I + \mu \frac{vv^T}{\langle y,v \rangle} - \frac{vY^T + Yv^T}{\langle y,v \rangle} + \frac{\langle y,y \rangle}{\langle y,v \rangle^2} vv^T \qquad ....(12)$$

Proof :

The first derivatives of $\psi(\mu X)$ is given by

$$\nabla \ \psi(\mu X) = I - \mu X^{-1} \qquad ....(13)$$

Lemma 1.2 imply that the optimal solution $\bar{X}$ satisfies the condition

$$I - \mu X^{-1} = v\lambda^T + \lambda v^T \qquad ....(14)$$

$$\frac{I}{\mu} - X^{-1} = \frac{v\lambda^T + \lambda v^T}{\mu} \qquad ....(15)$$

$$\left(\frac{I}{\mu} - X^{-1}\right) y = \frac{\langle \lambda, y \rangle v}{\mu} + \frac{\langle y, v \rangle \lambda}{\mu} \qquad ....(16)$$

$$\frac{Iy}{\mu} - v = \frac{\langle \lambda, y \rangle v}{\mu} + \frac{\langle y, v \rangle \lambda}{\mu} \qquad ....(17)$$

for $\lambda \in R^n$. Using the secant equation $\bar{X}^{-1} y = v$, we get

$$\left\langle \left(\frac{y}{\mu} - v \right), y \right\rangle = \left\langle \left(\frac{I}{\mu} - X^{-1} \right) y, y \right\rangle$$
$$= \left\langle \left(\frac{v\lambda^T + \lambda v^T}{\mu} \right) y, y \right\rangle \qquad ....(18)$$
$$= \frac{2\langle \lambda, y \rangle \langle y, v \rangle}{\mu}$$

which yields

$$\langle \lambda, y^T \rangle = \frac{\left\langle \left(\frac{y}{\mu} - v \right), y \right\rangle}{\frac{2\langle y, v \rangle}{\mu}} \qquad ....(19)$$
$$= \frac{\mu \left\langle \left(\frac{y}{\mu} - v \right), y \right\rangle}{2\ \langle y, v \rangle} .$$

Substituting this in the equation

$$\frac{Iy}{\mu} - v = \frac{\langle \lambda, y \rangle v}{\mu} + \frac{\langle y, v \rangle \lambda}{\mu} \qquad ....(20)$$

$$(y - \mu v) - \langle \lambda, y \rangle v = \langle y, v \rangle \lambda \qquad ....(21)$$

gives

$$\lambda = \frac{(y - \mu v)}{\langle y, v \rangle} - \frac{\langle \lambda, y \rangle}{\langle y, v \rangle^2} v \qquad ....(22)$$

$$\lambda = \frac{(y - \mu v)}{\langle y, v \rangle} - \frac{\langle y - \mu v, y \rangle}{2\langle y, v \rangle^2} v . \qquad ....(23)$$

Substituting this in (14) and simplifying the result givens (12) .

Corollary 2.2:
Let $\langle v_k, y_k \rangle > 0$ and $H_k \in S^n$ be positive definite, where H_k is the approximation to the inverse Hessian at iteration k. The matrix H_{k+1} in

the Al-Bayati [1] update formula (7) satisfies $H_{k+1} = \bar{B}^{-1}$ where $\bar{B}$ is the optimal solution to the minimization problem in S^n

$$\min\ \psi(H_k^{1/2} \rho B_k H_k^{1/2}) = \langle H_k, \rho B \rangle - In \det \rho B + const \qquad ....(24)$$

$$\text{s.t. :} \quad Bv_k = y_k . \qquad ....(25)$$

Proof :
The change of variables

$$X = H_k^{1/2} \rho B_k H_k^{1/2},\ y = (\rho H_k)^{1/2} y_k,\ v = (H_k/\rho)^{-1/2} v_k \quad ....(26)$$

Reduces the problem to the minimization problem $(10-11)$ in Theorem (2.1). Substituting the values of X, y, v above in equation (12) and simplifying, we obtain

$$\bar{B}^{-1} = H_k - \frac{H_k y_k v_k^T + v_k y_k^T H_k}{v_k^T y_k} + \rho \frac{v_k v_k^T}{v_k^T y_k} + \frac{y_k^T H_k y_k}{v_k^T y_k} \frac{v_k v_k^T}{v_k^T y_k} \quad ....(27)$$

The right hand side of this formula is identical to the one in (4) by using transformation of variables given in (26). Consequently, $\bar{B} = H_{k+1}^{-1}$. This completes the proof.

Corollary 2.3:
Let $\langle v_k, y_k \rangle > 0$ and $B_k \in S^n$ be positive definite, where B_k is the approximation to the inverse Hessian at iteration k. The matrix B_{k+1} in

the Oren update formula (8) satisfies $B_{k+1} = \bar{H}^{-1}$ where $\bar{H}$ is the optimal solution to the minimization problem in S^n

$$\min\ \psi(B_k^{1/2} \xi H_k B_k^{1/2}) = \langle B_k, \xi H \rangle - In \det \xi H + const \quad ....(28)$$

$$\text{s.t. :} \quad H y_k = v_k. \quad ....(29)$$

Proof :
The change of variables

$$X = B_k^{1/2} \xi H B_k^{1/2},\ y = (H_k/\xi)^{-1/2} y_k,\ v = (\xi B_k)^{1/2} v_k \quad ....(30)$$

Reduces the problem to the minimization problem $(10-11)$ in Theorem (2.1). Substituting the values of X, y, v above in equation (12) and simplifying, we obtain

$$\bar{H}^{-1} = B_k - \frac{y_k v_k^T B_k + B_k v_k y_k^T}{v_k^T y_k} + \xi \frac{y_k y_k^T}{v_k^T y_k} + \frac{v_k^T B_k v_k}{v_k^T y_k} \frac{y_k y_k^T}{v_k^T y_k} \quad ....(31)$$

The right hand side of this formula is identical to the one in (4) by using the 2nd set of transformation of variables given in (30). Consequently,

$\bar{H} = B_{k+1}^{-1}$. This completes the proof.

3 Dual of the trace determinant function for minimization problems in Quasi- Newton methods:

In this section, we have proposed two dual problems for the trace determinant function minimization problems. As we will see, the optimal solution to the dual problem is directly related to the Al-Bayati [1] update formula. Similar results also hold true for the Oren [6] update formula. The following theorem (3.1) is given in [7].

Theorem 3.1:
Let $X_0, Y_0 \in S^n$. The following minimization problem are duals of each other,

$$(P) \qquad \min \frac{1}{2}\|x - x_0\|^2 \qquad x \in y_0 + V \qquad ....(32)$$

$$(D) \qquad \min \frac{1}{2}\|y - y_0\|^2 \qquad y \in x_0 + V^{\perp} \qquad ....(33)$$

If (P) and (D) both have feasible positive definite solutions, then they optimal solutions and the strong duality theorem holds true. Furthermore, the optimal solutions of (P) and (D) are inverses of each other, that is, $\bar{Y} = (\bar{X})^{-1}$ where $\bar{X}$ and $\bar{Y}$ are the optimal solutions of (P) and (D), respectively. The following corollary is immediate, using the choices $X_0 = H_k, Y_0 \in S^n$ is any matrix satisfying the condition $Bv_k = y_k$ (such as $Y_0 = y_k y_k^T / \langle v_k, y_k \rangle$), and $\Im = \{B \in S^n : Bv_k = 0\}$ [7].

Corollary 3.2 :
Let v_k, y_k, H_k and H_{k+1} be defined as in Corollary (2.2). H_{k+1} in the Al-Bayati update formula (7) is the optimal $\bar{Y}$ to the minimization problem

$$\min \ \langle Y_0, Y \rangle - In\det Y \qquad ....(34)$$
$$Y = H_k/\rho + \lambda v_k^T/\rho + v_k \lambda^T/\rho\ ,\ \lambda \in R^n .$$

Thus, we obtain here a new result that the Al-Bayati updating formula B_{k+1} and $H_{k+1} = B_{k+1}^{-1}$ come from the primal dual problem (24) and (34), respectively. A similar result holds for the Oren updating formula.

4 Conclusions

Different measures lead to different Variable Metric or (Quasi-Newton) formulas. The idea may be extended for any positive definite symmetric matrix of Variable Metric class based on the trace determinant functions.

5 Open Problems

As we mentioned in [8], Biggs VM-update can be expressed as:

$$H_{k+1}^{Biggs} = \Phi\left\{H_k - \frac{H_k y_k s_k^T + s_k y_k^T H_k}{s_k^T y_k}\right\} + \Omega\left\{\frac{s_k s_k^T}{s_k^T y_k}\left[1 + \frac{y_k^T H_k y_k}{s_k^T y_k}\right]\right\} \qquad (35)$$

where Φ and Ω are selected in such away so that they satisfy the Quasi-Newton like condition; this is can be further analyized as:

$$H_{k+1}^{BFGS} = \left\{H_k - \frac{H_k y_k s_k^T + s_k y_k^T H_k}{s_k^T y_k}\right\} + \left\{\frac{s_k s_k^T}{s_k^T y_k}\left[1 + \frac{y_k^T H_k y_k}{s_k^T y_k}\right]\right\} \qquad ....(36)$$

In formula (36) Biggs update satisfy the least change property with respect to the new measure and it can be further expressed as a Projected-BFGS update or as a hybrid standard CG- Projected BFGS update.

References

[1] A.Y. Al-Bayati, A New Family of Self-Scaling Variable Metric Algorithm for Unconstrained Optimization. Journal of Education and Science, University of Mosul, Iraq, (12) (1991) 25-54.

[2] A.Y. Al-Bayati, and B.A. Hassan, B., Some Theoretical Results for Oren Variable Metric Method. J. of Kirkuk University, Iraq, (1) (2006) 26-31.

[3] R. Byrd and J. Nocedal, A Tool for the Analysis of Quasi-Newton Methods with Application to Unconstrained Minimization, SIAM J., (24) (1989) 727-739.

[4] B.A. Hassan, A New Theoretical Result for Quasi-Newton formulae for Unconstrained Optimization. Al-Rafidain Journal of Computer and Math. Sciences, Iraq, (2) (2005) 41-46.

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

[6] S.S. Oren, Self-Scaling Variable Metric Algorithm, Part II, Management Science, (20) (1979) 862-874.

[7] G. Osman; G. Filiz and S. Olena, Variational Problems in QN Methods Department of Mathematic. and Statistics, University of Maryland Baltimore, County, Maryland, 21250, USA, (2006) 1-22.

[8] A.Y. Al-Bayati, and B.A. Hassan, Some Variational Problems Based on the Least Square Technique, International Journal of Open Problems in Computer Science and Mathematics, IJOPCM, Vol. (4), No.(1), pp. 141-151, 2011.