A New Theoretical Result for Quasi-Newton Formulae for Unconstrained Optimization

Basim A. Hassan
basimabas39@gmial.com
College of Computer Sciences and Mathematics
University of Mosul, Iraq

Received on: 15/03/2005 **Accepted on: 09/05/2005**

ABSTRACT

The recent measure function of Byrd and Nocedal [3] is considered and simple proofs of some its properties are given. It is then shown that the AL-Bayati (1991) formulae satisfy a least change property with respect to this new measure .The new formula has any extended positive definite matrix of Brouden Type-Updates.

Keywords: Quasi-Newton method, Some theoretical result for quasi-Newton formulae.

البراهين النظرية الجديدة لصيغ شبيه نيوتن في الأمثلية غير المقيدة

باسم عباس حسن
كلية علوم الحاسبات والرياضيات، جامعة الموصل

تاريخ استلام البحث: 2005/03/15 **تاريخ قبول البحث: 2005/05/09**

الملخص

في هذا البحث تم استحداث عدد من البراهين النظرية لبعض الصفات الخاصة بالدالة المعرفة حسب (1.1) لتكوين عدد من المفاهيم الجديدة المعطاة في الفقرة(2) من هذا البحث لصيغة Al-Bayati (1991) [1] الصيغة الجديدة فيها توسيع مصفوفة موجبة التعريف من صنف Brouden.

الكلمات المفتاحية : طريقة شبيه نيوتن , بعض البراهين النظرية لصيغ شبيه نيوتن.

1.Introduction.

Recently Byrd and Nocedal [3] introduced the measure function $\Psi : R^{n*n} \to R$ defined by

$$\Psi(A) = trace(A) - f(A) \qquad \text{.......... ....(1.1)}$$

where $f(A)$ denotes the function

$$f(A) = In(\det A) \qquad \text{.......... .... (1.2)}$$

Byrd and Nocedal use this function to unify and extend certain convergence results for Quasi-Newton methods. In this paper, simple proofs of some of the properties of these functions are given. These properties give a new variaional result for the AL-Bayati updating formulae [1] .

Lemma1.1. $f(A)$ is a strictly concave function on the set of positive definite diagonal $n*n$ matrices.

Proof. Let $A = diag(a_i)$. Then $\nabla^2 f = diag(-1/a_i^2)$ and is negative definite since $a_i \succ 0$ for all i. Hence f is strictly concave [7].

Lemma1.2. $f(A)$ is a strictly concave function on the set of positive definite symmetric $n*n$ matrices.

Proof. Let $A \neq B$ be any two such matrices. Then there exist $n*n$ matrices X and Λ (X is nonsingular, $\Lambda = diag(\lambda_i)$) such that $X^T AX = \Lambda$ and $X^T BX = I$.

Denote $C = (1-\theta)A + \theta B, \theta \in (0,1)$.

Then

$$X^T CX = (1-\theta)X^T AX + \theta X^T BX = (1-\theta)\Lambda + \theta I \quad \text{....................(1.3)}$$

Also

$$f(X^T AX) = In\det(X^T AX) = In(\det^2 X \det A) = f(A) + In\det^2 X, \text{.......(1.4)}$$

and likewise

$$f(X^T BX) = f(B) + In\det^2 X \quad \text{...(1.5)}$$

$$f(X^T CX) = f(C) + In\det^2 X \quad \text{...(1.6)}$$

Now $A \neq B \Leftrightarrow \Lambda \neq I$, so by Lemma 1.1 and Eq.(1.3) it follows for $\theta \in (0,1)$ that

$$f(X^T CX) = f((1-\theta)\Lambda + \theta I \succ (1-\theta) f(A) + \theta f(I) = (1-\theta) f(X^T AX) + \theta f(X^T BX).$$

Hence form $(1.4)-(1.6)$,

$$f(C) \succ (1-\theta) f(A) + \theta f(B),$$

and so the Lemma is established [5].

Lemma.3. $\Psi(A)$ is a strictly convex function on the set positive definite symmetric $n*n$ matrices.

Proof. This follows from Lemma 1.2 and linearity of trace(A) [5].

Lemma1.4. For nonsingular A the derivative of $\det(A)$ is given by $d(\det A)/da_{ij} = [A^{-T}]_{il} \det A.$

Proof. From the the well-known identity $\det(I + uv^T) = 1 + v^T u$ it follows that

$$\det(\rho A + \varepsilon e_i e_j^T) = \det(I + \varepsilon\rho e_i e_j^T A^{-1}) \det \rho A = (1 + \varepsilon\rho(A^{-1})_{ji}) \det \rho A \ .$$

Hence

$$\frac{d\det A}{da_{ij}} = \lim_{\varepsilon\to 0}\frac{\det(\rho A+\varepsilon e_i e_j^T)-\det\rho A}{\varepsilon} = (\rho A^{-1})_{ji}\det A \ .$$

Theorem1.1. $\psi(A)$ is globally and uniquely minimized by $A = I$ over the set of positive definite symmetric $n*n$ matrices .

Proof. Because A is nonsingular , ψ is continuously differentiable and so

$$\frac{d\psi}{da_{ij}} = I_{ij} - \frac{1}{\det\rho A}\frac{d}{da_{ij}}\det\rho A = (I-\rho A^{-T})_{ij} \ , \qquad \text{...................... ......(1.7)}$$

using Lemma 1.4. Hence ψ is stationary when $A = I$ and the theorem follows by virtue of Lemma 1.3.

Remark. It is also shown in [3] that $A = I$ is a global minimizer of $\psi(A)$.

2.A variational result . The Al-Bayati updating formula

$$H^{k+1} = H^k + \left[\frac{2\gamma^T H^k\gamma}{(\delta^T\gamma)^2}\right]\delta\delta^T - \frac{H^k\gamma\,\delta^T + \delta\,\gamma^T H}{\delta^T\gamma} \ , \qquad \text{..........................(2.1)}$$

Occupies a central role in unconstrained optimization . (Here δ and γ denoted certain difference vectors occurring on iteration k of a Quasi-Newton method , with $\delta^T\gamma \succ 0$. $B^{(k)}$ denotes the current Hessian approximation , and $H^{(k)}$ its inverse : see , for example , [4]) A significant result due to Goldfarb [6] is that the correction in the Al-Bayati formula satisfies a minimum property with respect to a function of the form $\|E\|_w^2 = trace(EWEW)$ (its corollary in [4]) .

The main result of this paper is to show that these formulae also satisfy a minimum property with respect to the measure function ψ of Byrd and Nocedal defined in (1.1) .

Theorem2.1: if $H^{(k)}$ is positive definite and $\delta^T\gamma \succ 0$, the variation problem

$$\underset{B>0}{\text{minimize}}\ \Psi(H^{(K)1/2}\rho B H^{(K)1/2}) \qquad \text{………………..(2.2)}$$

$$\text{subject to} \quad B^T = B \qquad \text{.......................................(2.3)}$$

$$B\delta = \gamma \qquad \text{.......................................(2.4)}$$

is solved uniquely by the matrix $B^{(k+1)}$ given by the formula (2.1).

proof: the matrix product that forms the argument of Ψ can be cyclically permuted so that

$$\Psi(H^{(K)1/2}\rho B H^{(K)1/2}) = trace(H^{(K)}\rho B) - \ln(\det H^{(K)}\det\rho B)$$

$$= \Psi(H^{(K)}\ \rho B) = \Psi(\rho B H^{(K)}) \qquad \text{.......... ....(2.5)}$$

A constrained stationary point of the variational problem can be obtained by the method of lagrange multipliers.

A suitable lagrangian function is

$$L(B,\wedge,\lambda) = \frac{1}{2}\psi(H^{(K)1/2}\rho B H^{(K)1/2} + trace(\wedge^T(B^T - B)) + \lambda^T(B\delta - \gamma)$$

$$= \frac{1}{2}(trace(H^{(K)}\rho B) - \ln\det H^{(K)} - \ln\det \rho B) + trace(\Lambda^T(B^T - B)) + \lambda^T(B\delta - \gamma)$$

Where $\wedge$ and λ are lagrange multipliers for (2.3) and (2.4), respectively. To solve the first order conditions, it is necessary to find B, $\wedge$ and λ to satisfy (2.3), (2.4), and the equations $\partial L/\partial B_{ij} = 0$. Using the identity $\partial B/\partial B_{ij} = e_i e_j^T$ and Lemma (1.4), it follows that

$$\partial L/\partial B_{ij} = 0 = \frac{1}{2}(trace(H^{(K)}\rho e_i e_j^T) - (\rho B^{-1})_{ji}) + trace(\Lambda^T(e_j e_i^T - e_i e_j^T)) + \lambda^T e_i e_j^T \delta$$

$$= \frac{1}{2}((\rho H^{(K)})_{ji} - (\rho B^{-1})_{ji}) + \Lambda_{ji} - \Lambda_{ij} + (\lambda\delta^T)_{ij}\,.$$

Transposing and adding, using the symmetry of $H^{(k)}$ and B, gives

$$H^{(K)} - \rho B^{-1} + \lambda\delta^T + \delta\lambda^T = 0$$

or

$$\rho B^{-1} = H^{(K)} + \lambda\delta^T + \delta\lambda^T = 0\,, \qquad \text{.......... ....(2.6)}$$

$$B^{-1} = H/\rho + \lambda\delta^T/\rho + \delta\lambda^T/\rho$$

which shows that the optimum matrix inverse involves a rank-2 correction of $H^{(k)}$. to determine λ, (2.6) is post-multiplied by γ. It then follows, using the equation $B^{-1}\gamma = \delta$ derived from (2.4), that

$$\delta = H\gamma/\rho + \lambda\delta^T\gamma/\rho + \delta\lambda^T\gamma/\rho$$

and hence

$$\gamma^T\delta = \gamma^T H\gamma/\rho + \gamma^T\lambda\delta^T\gamma/\rho + \gamma^T\delta\lambda^T\gamma/\rho\,.$$

$$\gamma^T\delta = \gamma^T H\gamma/\rho + 2\gamma^T\lambda\delta^T\gamma/\rho$$

$$\rho\gamma^T\delta = \gamma^T H\gamma + 2\gamma^T\lambda\delta^T\gamma$$

$$\rho\gamma^T\delta - \gamma^T H\gamma = 2\gamma^T\lambda\delta^T\gamma$$

$$\rho - \gamma^T H\gamma/\delta^T\gamma = 2\gamma^T\lambda$$

Rearranging this gives $\quad \gamma^T\lambda = \dfrac{1}{2}(\rho - \gamma^T H\gamma/\delta^T\gamma)$

and so

$$\delta = H\gamma / \rho + \lambda\delta^T \gamma / \rho + \delta\lambda^T \gamma / \rho$$

$$\delta = H\gamma / \rho + \lambda\delta^T \gamma / \rho + \delta\gamma^T \lambda / \rho$$

$$\lambda\delta^T \gamma / \rho = \delta - H\gamma / \rho - \delta\gamma^T \lambda / \rho$$

$$\lambda\delta^T \gamma = \rho\delta - H\gamma - \delta\gamma^T \lambda$$

$$\lambda\delta^T \gamma = \rho\delta - H\gamma - \frac{\delta}{2}\left[\rho - \gamma^T H\gamma / \delta^T \gamma\right]$$

$$\lambda = (\rho\delta - H\gamma - \frac{\delta}{2}\left[\rho - \gamma^T H\gamma / \delta^T \gamma\right]) / \delta^T \gamma \;, \qquad \text{.......... ....(2.7)}$$

from (2.7) we have

$$\lambda\delta^T = -\frac{H\gamma\delta^T}{\delta^T \gamma} + \frac{\delta\delta^T}{2\delta^T \gamma}\left[\rho + \gamma^T H\gamma / \delta^T \gamma\right]$$

$$\lambda^T = -\frac{\gamma^T H}{\delta^T \gamma} + \frac{\delta^T}{2\delta^T \gamma}\left[\rho + \gamma^T H\gamma / \delta^T \gamma\right]$$

$$\delta\lambda^T = -\frac{\delta\gamma^T H}{\delta^T \gamma} + \frac{\delta\delta^T}{2\delta^T \gamma}\left[\rho + \gamma^T H\gamma / \delta^T \gamma\right]$$

substituting this expression into (2.6) gives the equation

$$\rho B^{-1} = H - \frac{H\gamma\delta^T + \delta\gamma^T H}{\delta^T \gamma} + \frac{\delta\delta^T}{\delta^T \gamma}\left[\rho + \gamma^T H\gamma / \delta^T \gamma\right]$$

where

$$\rho = \gamma^T H\gamma / \delta^{\,T} \gamma$$

and hence the proof .

3.<u>Conclusions</u>:

It is a well-known consequence of the sherman-Morrison formula [4] that there exists a corresponding rank-2 update for B, which is given by the right – hand side of (2.1). Moreover the conditions of the theorem (2.1) ensure that the resulting updated matrix B is positive definite (as in [4]).

This establishes that the AL-Bayati formula satisfies first order conditions (including feasibility) for the variational problem. Finally, $\Psi(H^{(K)1/2}\ \rho B\, H^{\ (K)1/2})$ is seen to be a strictly convex function on $B \succ 0$ by virtue of (2.5) and Lemma (1.2), so it follows that the AL-Bayati formula gives the unique solution of the variational problem. This idea may be extended for any positive definite matrices of Broyden class.

FERENRECES

[1] Al-Bayati, A. (1991),” A new family of self Scaling Variable Metric Algorithm for unconstrained Optimization”. **Journal of Euc. And Sci.**,25,12.

[2] Bellman, R. (1960) **Introduction to Matrix Analysis** , New York: McGraw-Hill .

[3] Byrd, R.H. and Nocedal, J. (1989), “A tool for the analysis of quasi-Newton methods with application to unconstrained minimization”, **SIAM J. Numer. Anal**., 26, pp. 727-739.

[4] Fletcher, R. (1987) **Practical Methods of Optimization**, Second Edition. Chichester: Jonhn Wiley

[5] Fletcher R. (1991) **A New variational Result for Quasi-Newton formula** , **SIAM Journal on Optimization** , Vol. 1,pp. 18-21.

[6] Goldfarb D. (1970) , “A family of variable metric methods derived by variational means”, **Math. Comp**., 24, pp. 23-26.

[7] Hu Y.F. and Storey C. ,(1994) , “family of Optimally conditioned QN Updates for Unconstrained Optimization” ,**Journal of Optimization theory and applications** , Vol. 83 , No.2 ,pp. 421-431 .