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

Basim A. Hassan
College of Computer Sciences and Mathematics
University of Mosul

Received on: 15/3/2005 **Accepted on: 9/5/2005**

الملخص

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

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.

1.Introduction.

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

$\Psi(A) = trace(A) - f(A)$..........................(1.1)

denotes the function $f(A)$ where

$f(A) = In(\det A)$......................................(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].

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

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

Denote $C = (1-q)A + qB, q \in (0,1)$.

Then

$X^T CX = (1-q)X^T AX + qX^T BX = (1-q)\Lambda + qI$..........................(1.3)

Also

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

and likewise

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

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

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

$f(X^T CX) = f((1-q)\Lambda + qI \;\mathbf{f}\; (1-q)f(A) + qf(I) = (1-q)f(X^T AX) + qf(X^T BX)$.

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

$$f(C) \;\mathbf{f}\; (1-q)f(A) + qf(B),$$

and so the Lemma is established [5].

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

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

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

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

$\det(rA + ee_i e_j^T) = \det(I + ere_i e_j^T A^{-1}) \det rA = (1 + er(A^{-1})_{ji}) \det rA$.

Hence

$$\frac{d \det A}{da_{ij}} = \lim_{e \to 0} \frac{\det(rA + ee_i e_j^T) - \det rA}{e} = (rA^{-1})_{ji} \det A \quad .$$

<u>Theorem1.1.</u> $y(A)$ is globally and uniquely minimized by $A = I$ over the set of positive definite symmetric $n*n$ matrices .

<u>Proof.</u> Because A is nonsingular , y is continuously differentiable and so

$$\frac{dy}{da_{ij}} = I_{ij} - \frac{1}{\det rA}\frac{d}{da_{ij}}\det rA = (I - rA^{-T})_{ij}\ ,\ldots\ldots\ldots\ldots(1.7)$$

using Lemma 1.4. Hence y 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 $y\,(A)$.

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

$$H^{k+1} = H^k + \left[\frac{2g^T H^k g}{(d^T g\ )^2}\right] dd^T - \frac{H^k g\ d^T + d\ g^T H}{d^T g}\ ,\ldots\ldots\ldots\ldots(2.1)$$

Occupies a central role in unconstrained optimization . (Here d and g denoted certain difference vectors occurring on iteration k of a Quasi-Newton method , with $d^T g \mathbf{f} 0$. $B^{(k)}$ 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 y of Byrd and Nocedal defined in (1.1) .

Theorem2.1: if $H^{(k)}$ is positive definite and $d^T g \mathbf{f} 0$, the variation problem

$$\underset{B>0}{\text{minimize}}\ \Psi(H^{(K)1/2}\ rB\,H^{(K)1/2})\ldots\ldots\ldots\ldots(2.2)$$

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

$$Bd = g\ldots\ldots\ldots\ldots(2.4)$$

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

$$\Psi(H^{(K)1/2}\ rB\,\mathrm{H}^{(K)1/2}) = trace(H^{(K)} rB) - \ln(\det H^{(K)} \det rB)$$
$$= \Psi(H^{(K)}\ rB) = \Psi(rBH^{(K)})\ldots\ldots\ldots\ldots(2.5)$$

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

A suitable lagrangian function is

$$L(B,\wedge,l)=\frac{1}{2}y\,(H^{(K)1/2}\,rBH^{(K)1/2}+trace(\wedge^{T}(B^{T}-B))+l^{T}(Bd-g)$$
$$=\frac{1}{2}(trace(H^{(K)}\,rB)-\ln\det H^{(K)}-\ln\det rB)+trace(\Lambda^{T}(B^{T}-B))+l^{T}(Bd-g)$$

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

$$\partial L/\partial B_{ij}=0=\frac{1}{2}(trace(H^{(K)}\,re_i e_j^T)-(rB^{-1})_{ji})+trace(\Lambda^{T}(e_j e_i^T-e_i e_j^T))+l^{T}e_i e_j^T d$$
$$=\frac{1}{2}((rH^{(K)})_{ji}-(rB^{-1})_{ji})+\Lambda_{ji}-\Lambda_{ij}+(l\,d^{T})_{ij}.$$

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

$$H^{(K)}-rB^{-1}+l\,d^{T}+d\,l^{T}=0$$

or

$$rB^{-1}=H^{(K)}+l\,d^{T}+d\,l^{T}=0\ ,\text{..........................}(2.6)$$
$$B^{-1}=H/r+l\,d^{T}/r+d\,l^{T}/r$$

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 $B^{-1}g=d$ derived from (2.4), that

$$d=Hg/r+l\,d^{T}g/r+d\,l^{T}g/r$$

and hence

$$g^{T}d=g^{T}Hg/r+g^{T}l\,d^{T}g/r+g^{T}d\,l^{T}g/r\,.$$
$$g^{T}d=g^{T}Hg/r+2g^{T}l\,d^{T}g/r$$
$$rg^{T}d=g^{T}Hg+2g^{T}l\,d^{T}g$$
$$rg^{T}d-g^{T}Hg=2g^{T}l\,d^{T}g$$
$$r-g^{T}Hg/d^{T}g=2g^{T}l$$

Rearranging this gives $\quad g^{T}l=\frac{1}{2}(r-g^{T}Hg/d^{T}g)$

and so

$$d=Hg/r+l\,d^{T}g/r+d\,l^{T}g/r$$
$$d=Hg/r+l\,d^{T}g/r+dg^{T}l/r$$

$$l\,d^Tg\,/\,r = d - Hg\,/\,r - dg^T l\,/\,r$$

$$l\,d^Tg = rd - Hg - dg^T l$$

$$l\,d^Tg = rd - Hg - \frac{d}{2}\left[r - g^T Hg\,/\,d^Tg\right]$$

$$l = (rd - Hg - \frac{d}{2}\left[r - g^T Hg\,/\,d^Tg\right])/d^Tg \quad ,\dots\dots\dots\dots\dots(2.7)$$

from (2.7) we have

$$l\,d^T = -\frac{Hgd^T}{d^Tg} + \frac{dd^T}{2d^Tg}\left[r + g^T Hg\,/\,d^Tg\right]$$

$$l^T = -\frac{g^TH}{d^Tg} + \frac{d^T}{2d^Tg}\left[r + g^T Hg\,/\,d^Tg\right]$$

$$d\,l^T = -\frac{dg^TH}{d^Tg} + \frac{dd^T}{2d^Tg}\left[r + g^T Hg\,/\,d^Tg\right]$$

substituting this expression into (2.6) gives the equation

$$rB^{-1} = H - \frac{Hgd^T + dg^TH}{d^Tg} + \frac{dd^T}{d^Tg}\left[r + g^T Hg\,/\,d^Tg\right]$$

where

$$r = g^T Hg\,/d^{\,T}g$$

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, $\Psi(H^{(K)1/2}\; rB\,H^{\,(K)1/2})$ 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 .