Some Theoretical Results for Oren Variable-Metric Method

Basim A.Hassan & Abbas Y.Al-Bayati
College of Computers Science and Mathmetics-Mosul University

Abstract

The main result of this paper is to show that Oren updating formula for unconstraind optimization is also satisfy a minimum property with respect to the measure function Ψ of Byrd and Nocedal. The Oren formula gives the unique solution of the variational problem.This idea may be extended for the inverse of any positive difinite matrix.

Introduction.

(Byrd &Nocedal,1989) introduced the measure function $\Psi : R^{n*n} \rightarrow R$ defined by

$$\Psi(A) = trace(A) - f(A) \qquad \text{............................(1)}$$

where $f(A)$ denotes the function

$$f(A) = In(\det A) \qquad \text{..(2)}$$

Byrd and Nocedal use this function to unify and extend certain convergence results for Quasi-Newton methods. We examine a self scaling Quasi-Newton (Hu&Storey,1994) updats as described above. In particular,we look at many possible updating formulae that have been suggested .Mostly the context is that of a line search method, but these ideas are also relevant to use in the trust region methods. The current preformance for the BFGS formula was suggested by (Fletcher,1987) ,and is known as the BFGS formula

$$H_{BFGS}^{K+1} = H - \frac{\delta\gamma^T H + H\gamma\delta^T}{\delta^T\gamma} + \left(1 + \frac{\gamma^T H\gamma}{\delta^T\gamma}\right)\frac{\delta\delta^T}{\delta^T\gamma} \qquad \text{..........(3)}$$

It can be motivated in the following way . If H^{-1} is denoted by B (so that $B^K = H^{(K)^{-1}}$ approximates the Hessian matrix G), then it can be verified (by establishing that $B_{BFGS}^{K+1} H_{BFGS}^{K+1} = I$) that

$$B_{BFGS}^{K+1} = B - \frac{B\delta\delta^T B}{\delta^T B\delta} + \frac{\gamma\gamma^T}{\gamma^T\delta}$$

This resembles the DFP formula

$$H_{DFP}^{K+1} = H - \frac{H\gamma\gamma^T H}{\gamma^T H\gamma} + \frac{\delta\delta^T}{\delta^T\gamma}$$

but with the interchanges $B \leftrightarrow H$ and $\gamma \leftrightarrow \delta$ having been made .Formulae related in this way are siad to be dual .Similarly it follows from (3) that

$$B_{DFP}^{K+1} = B - \frac{\gamma\delta^T B + B\delta\gamma^T}{\gamma^T\delta} + \left(1 + \frac{\delta^T B\delta}{\gamma^T\delta}\right)\frac{\gamma\gamma^T}{\gamma^T\delta}$$

In fact (given any formula for H^{K+1} the dual formula for B^{K+1} follows by inter changing as above) and then a dual formula for H^{K+1} can be obtained using the Sherman-Morrison formula. Also the Quasi-Newton condition $\delta = H\gamma$ is preserved by the duality operation .It can be established that rank one formula is self-dual and another self-dual formula which preserves positive definite H matrix is given by (Fletcher,1987). In this paper, simple proofs of some of the properties of these functions are given. These properties are form a new variational result for the Oren updating formula (Oren,1974) .

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

Proof .Let $A = diag(a_i)$.Then $\nabla^2 f = diag(-1/a_i^2)$ and is negative definite since $a_i > 0$ for all i. Hence f is strictly concave (Hu&Storey,1994).

LEMMA 1. 2. $f(A)$ is strictly concave function on the set of positive definite symmetric $n \times n$ matrices(Fletcher,1991).

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

$X^T CX = (1-\theta)X^T AX + \theta X^T BX = (1-\theta)\Lambda + \theta I$........................(4)

Also

$f(X^T AX) = In\det(X^T AX) = In(\det^2 X \det A) = f(A) + In\det^2 X$,.......(5)

And likewise

$f(X^T BX) = f(B) + In\det^2 X$..(6)

$f(X^T CX) = f(C) + In\det^2 X$..(7)

Now $A \neq B \Leftrightarrow \Lambda \neq I$, so by Lemma 1.1 and Eq.(4) 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 $(5)-(7)$,

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

and so the lemma is established (Fletcher,1991).

LEMMA 1.3. $\Psi(A)$ 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)(Fletcher,1991).

LEMMA1.4. for nonsingular A the derivative of $\det(A)$ is given by $d(\det A)/da_{ij} = [A^{-T}]_{il} \det A.$. (Byrd &Nocedal,1989)

Proof. From the the well-known identity $\det(I + uv^T) = 1 + v^T u$ it follows that $\det(\xi A + \varepsilon e_i e_j^T) = \det(I + \varepsilon \xi e_i e_j^T A^{-1}) \det \xi A = (1 + \varepsilon \xi (A^{-1})_{ji}) \det \xi A$.

Hence

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

THEOREM 1.1. $\psi(A)$ is globally and uniquely minimized by $A = I$ over the set of positive definite symmetric $n \times n$ matrices .

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

$$\frac{d\psi}{da_{ij}} = I_{ij} - \frac{1}{\det \xi A} \frac{d}{da_{ij}} \det \xi A = (I - \xi A^{-T})_{ij} \text{ ,..........................(8)}$$

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 (Byrd &Nocedal,1989) that $A = I$ is a global minimizer of $\psi(A)$.

A new variational result

The Oren updating formula

$$H^{k+1} = \left\{ H - \frac{H\gamma\gamma^T H}{\gamma^T H \gamma} \right\} \frac{1}{\xi} + \frac{\delta\delta^T}{\delta^T \gamma} \text{,.........................(9)}$$

where $\xi = \delta^T \gamma / \gamma^T H \gamma$

occupy 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 , (Fletcher,1987)) A significant result due to (Goldfarb,1970) 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 (Fletcher,1987)) .

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) .

NEW THEOREM 2.1: if $B^{(k)}$ is positive definite and $\delta^T \gamma > 0$, the variation problem

$$\underset{H>0}{\text{minimize}}\ \Psi(B^{(K)1/2}\, \xi H\, B^{(K)1/2}) \text{……………………..(10)}$$

subject to $H^T = H$..(11)

$H\gamma = \delta$..(12)

is solved uniquely by the matrix $B^{(k+1)}$ given by the formula (9).
proof: the matrix product that forms the argument of Ψ can be cyclically permuted so that

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

$$= \Psi(B^{(K)}\,\xi H) = \Psi(\xi H B^{(K)}) \text{.......................(13)}$$

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

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

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

where ∧ and λ are lagrange multipliers for (11) and (12), respectively. To solve the first order conditions, it is necessary to find B, ∧ and λ to satisfy (11), (12), and the equations $\partial L/\partial H_{ij} = 0$. Using the identity $\partial H/\partial H_{ij} = e_i e_j^T$ and lemma (1.4), it follows that

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

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

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

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

or

$$\xi H^{-1} = B^{(K)} + \lambda\gamma^T + \gamma\lambda^T = 0\ \text{,.............................(14)}$$

$$H^{-1} = B/\xi + \lambda\gamma^T/\xi + \gamma\lambda^T/\xi$$

which shows that the optimum matrix inverse involves a rank-2 correction of $H^{(k)}$. to determine λ, (14) is post-multiplied by δ. It then follows, using the equation $H^{-1}\delta = \gamma$ derived from (12), that

$$\gamma = B\delta/\xi + \lambda\gamma^T\delta/\xi + \gamma\lambda^T\delta/\xi$$

and hence

$$\delta^T\gamma = \delta^T B\delta/\xi + \delta^T\lambda\gamma^T\delta/\xi + \delta^T\gamma\lambda^T\delta/\xi\,.$$

$$\delta^T\gamma = \delta^T B\delta/\xi + 2\delta^T\lambda\gamma^T\delta/\xi$$

$$\xi\delta^T\gamma = \delta^T B\delta + 2\delta^T\lambda\gamma^T\delta$$

$$\xi\delta^T\gamma - \delta^T B\delta = 2\delta^T\lambda\gamma^T\delta$$

$$\xi - \delta^T B\delta/\gamma^T\delta = 2\delta^T\lambda$$

Rearranging this gives $\delta^T \lambda = \frac{1}{2}(\xi - \delta^T B\delta / \gamma^T \delta)$

and so

$$\gamma = B\delta / \xi + \lambda\gamma^T \delta / \xi + \gamma\lambda^T \delta / \xi$$
$$\gamma = B\delta / \xi + \lambda\gamma^T \delta / \xi + \gamma\delta^T \lambda / \xi$$
$$\lambda\gamma^T \delta / \xi = \gamma - B\delta / \xi - \gamma\delta^T \lambda / \xi$$
$$\lambda\gamma^T \delta = \xi\gamma - B\delta - \gamma\delta^T \lambda$$
$$\lambda\gamma^T \delta = \xi\gamma - B\delta - \frac{\gamma}{2}\left[\xi - \delta^T B\delta / \gamma^T \delta\right]$$
$$\lambda = (\xi\gamma - B\delta - \frac{\gamma}{2}\left[\xi - \delta^T B\delta / \gamma^T \delta\right]) / \gamma^T \delta \text{ ,.......................(15)}$$

from (15) we have

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

substituting this expression into (2.6) gives the equation

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

Further more, the dual is

$$\xi B^{-1} = H - \frac{H\gamma\delta^T + \delta\gamma^T H}{\delta^T \gamma} + \frac{\delta\delta^T}{\delta^T \gamma}\left[\xi + \gamma^T H\gamma / \delta^T \gamma\right]$$
$$= H - \frac{H\gamma\delta^T}{\delta^T \gamma} - \frac{\gamma^T \delta H}{\delta^T \gamma} + \xi\frac{\delta\delta^T}{\delta^T \gamma} + \frac{\delta\delta^T(\gamma^T H\gamma)}{(\delta^T \gamma)^2}$$
$$= H - \frac{H\gamma\gamma^T H}{\delta^T \gamma} - \frac{\delta\delta^T}{\delta^T \gamma} + \xi\frac{\delta\delta^T}{\delta^T \gamma} + \frac{\delta\delta^T(\gamma^T H\gamma)}{(\delta^T \gamma)^2}$$
$$= H - \frac{H\gamma\gamma^T H}{\delta^T \gamma} + \xi\frac{\delta\delta^T}{\delta^T \gamma}$$
$$B^{-1} = \left\{H - \frac{H\gamma\gamma^T H}{\delta^T \gamma}\right\} * \frac{1}{\xi} + \frac{\delta\delta^T}{\delta^T \gamma}$$

and hence the proof .

Conclusions

It is a well-known consequence of the sherman-Morrison formula (Fletcher,1987) that there exists a corresponding rank-2 update for H, which is given by the right – hand side of (9). Moreover the conditions of the theorem (9) ensure that the resulting updated matrix H is positive definite (as in (Fletcher,1987)). This establishes that the Oren formula satisfies first order conditions (including feasibility) for the variational problem. Finally, $\Psi(B^{(K)1/2}\xi H\, B^{(K)1/2})$ is seen to be a strictly convex function on $H \succ 0$ by virtue of (13) and Lemma (1.2), so it follows that the Oren formula gives the unique solution of the variational problem..

References

- 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.
- FLETCHER R.,(1987) Practical Methods of Optimization, Second Edition, Jonhn Wiley, Chichester,.
- FLETCHER R.,(1991) , A New variational Result for Quasi-Newton formula , SIAM Journal on Optimization , Vol. 1,pp. 18-21.
- GOLDFARB D., (1970) ,A family of variable metric methods derived by variational means, Math. Comp., 24, pp. 23-26.
- 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 .
- Oren , S.S. , (1974) , On the selection of parameters in self-scaling variable metric algorithm , Math. Prog. , 7,351-367 .

Journal of Kirkuk University –Scientific Studies , vol.1, No.1 ,2006

بعض النتائج النظرية الى طريقة Oren للمتري المتغير

باسم حسن وعباس البياتي
كلية علوم الحاسبات والرياضيات-جامعة الموصل

الخلاصة

تم في هذا البحث اثبات ان صيغة Oren في المتري المتغير للامثلية غير الخطية تحقق خاصية النهاية الصغرى لدالة ψ Byrd &Nocedal مع بعض الصفات الخاصة بالدالة ψ،كذلك صيغة Oren تعطي الحل الوحيد للمفاهيم النظرية الجديدة.الصيغة الجديدة يمكن توسيعها الى معكوس ايه مصفوفة موجبة التعريف .