14p

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF HOUSTON HOUSTON, TEXAS

CR 134216

PARTIAL DERIVATIVES FOR VARIOUS
SCALER FUNCTIONS OF MATRICES
BY JOHN QUIREIN JULY 1973

N74-20177

(NASA-CR-134216) PARTIAL DERIVATIVES FOR VARIOUS SCALER FUNCTIONS OF MATRICES (Houston Univ.) 12 p HC $4.00 CSCL 12A

Unclas
G3/19 35028

APR 1974
RECEIVED
NASA STI FACILITY
INPUT BRANCH

PREPARED FOR
EARTH OBSERVATION DIVISION , JSC
UNDER
CONTRACT NAS-9-12777

3801 CULLEN BLVD.
HOUSTON, TEXAS 77004

Report #29

Partial Derivatives for Various Scaler Functions of Matrices

by

J. A. Quirein

Mathematics Department

University of Houston

July 1973

NAS-9-12777 MOD 1S

INTRODUCTION

This paper computes the partial derivative of the scaler function $\Phi = \mathrm{tr}(B\Lambda B^T) - \mathrm{tr}\{M^T[(B\Lambda B^T)^{-1} - (BDB^T)^{-1} - I_k]\}$ with respect to the matrices B and D where:

- tr ; the trace of a matrix
- B ; a k by n real matrix of rank $k \le n$
- Λ ; an n by n positive definite, real symmetric matrix
- M ; a real k by k symmetric matrix (assumed to be constant)
- D ; an n by n positive definite diagonal matrix
- I_k ; the k by k identity matrix

DISCUSSION - Define the scaler

$$\Phi_1 = \frac{1}{2}\,\mathrm{tr}(B\Lambda B^T)$$

We use the notation $\frac{\partial\Phi_1}{\partial B}$ to denote the matrix $(\frac{\partial\Phi_1}{\partial b_{ij}})$ where $B = (b_{ij})$

Lemma 1 $\left(\frac{\partial\Phi_1}{\partial B}\right)^T = \Lambda B^T$, so that $\Phi_1 = \frac{1}{2}\,\mathrm{tr}[B\left(\frac{\partial\Phi_1}{\partial B}\right)^T]$

Proof: $\frac{\partial\Phi_1}{\partial b_{ij}} = \frac{1}{2}\,\mathrm{tr}[\ \frac{\partial B}{\partial b_{ij}}\Lambda B^T + B\Lambda\frac{\partial B^T}{\partial b_{ij}}\ ] = \frac{1}{2}\,\mathrm{tr}[\ (\frac{\partial B}{\partial b_{ij}}\Lambda B^T) + \left(\frac{\partial B}{\partial b_{ij}}\Lambda B^T\right)^T]$

But for any square matrices A_1 and A_2, $\mathrm{tr}(A_i) = \mathrm{tr}(A_i^T)$ and $\mathrm{tr}(A_1 + A_2) = \mathrm{tr}(A_1) + \mathrm{tr}(A_2)$ so that

$$\frac{\partial\Phi_1}{\partial b_{ij}} = \mathrm{tr}\ [\left(\frac{\partial B}{\partial b_{ij}}\Lambda B^T\right)] = \alpha_{ji}$$

where $\Lambda B^T = (\alpha_{ij})$

Thus it follows

$$\frac{\partial \Phi_1}{\partial b_{ij}} = (\alpha_{ji}), \text{ and thus}$$

$$\left(\frac{\partial \Phi_1}{\partial B}\right)^T = \left(\frac{\partial \Phi_1}{\partial b_{ij}}\right)^T = (\alpha_{ji})^T = (\alpha_{ij}) = \Lambda B^T$$

proving the Lemma.

<u>Lemma 2:</u> $\frac{\partial}{\partial b_{ij}} (B\Lambda B^T)^{-1} = -(B\Lambda B^T)^{-1}[\frac{\partial}{\partial b_{ij}}(B\Lambda B^T)](B\Lambda B^T)^{-1}$ where $B = (b_{ij})$

Proof: Since $(B\Lambda B^T)(B\Lambda B^T)^{-1} = I_k$, it follows

$$[\frac{\partial}{\partial b_{ij}} (B\Lambda B^T)](B\Lambda B^T)^{-1} + (B\Lambda B^T)[\frac{\partial}{\partial b_{ij}} (B\Lambda B^T)^{-1}] = (0)$$

so that

$$\frac{\partial}{\partial b_{ij}} (B\Lambda B^T)^{-1} = -(B\Lambda B^T)^{-1}[\frac{\partial}{\partial b_{ij}} (B\Lambda B^T)](B\Lambda B^T)^{-1}$$

as desired.

<u>Lemma 3</u> - Let $\Phi_2 = \frac{1}{2} \text{tr}[M^T(B\Lambda B^T)^{-1}]$. Then $\left(\frac{\partial \Phi_2}{\partial B}\right)^T = -\Lambda B^T(B\Lambda B^T)^{-1}M^T(B\Lambda B^T)^{-1}$ and $\Phi_2 = -\frac{1}{2} \text{tr}[B\left(\frac{\partial \Phi_2}{\partial B}\right)^T]$

Proof: $$\frac{\partial \Phi_2}{\partial b_{ij}} = \frac{\partial}{\partial b_{ij}} \frac{1}{2} \text{tr}[M^T(B\Lambda B^T)^{-1}]$$

$$= \frac{1}{2} \text{tr}[M^T \frac{\partial}{\partial b_{ij}} (B\Lambda B^T)^{-1}]$$

$$= -\frac{1}{2}\,\mathrm{tr}\{M^T(B\Lambda B^T)^{-1}\,[\frac{\partial}{\partial b_{ij}}\,(B\Lambda B^T)](B\Lambda B^T)^{-1}\} \qquad \text{(Lemma 2)}$$

$$= -\frac{1}{2}\,\mathrm{tr}\{M^T(B\Lambda B^T)^{-1}[\frac{\partial B}{\partial b_{ij}}\,\Lambda B^T + B\Lambda\frac{\partial B^T}{\partial b_{ij}}](B\Lambda B^T)^{-1}\}$$

$$= -\frac{1}{2}\,\mathrm{tr}\{\frac{\partial B}{\partial b_{ij}}\,\Lambda B^T(B\Lambda B^T)^{-1}M^T(B\Lambda B^T)^{-1}\}$$

$$-\frac{1}{2}\,\mathrm{tr}\{(B\Lambda B^T)^{-1}M^T(B\Lambda B^T)^{-1}B\Lambda\,\frac{\partial B^T}{\partial b_{ij}}\}$$

$$= -\mathrm{tr}\{\frac{\partial B}{\partial b_{ij}}\,\Lambda B^T(B\Lambda B^T)^{-1}M^T(B\Lambda B^T)^{-1}\}$$

$$= -\gamma_{ji}$$

where $(\gamma_{ij}) = \Lambda B^T(B\Lambda B^T)^{-1}M^T(B\Lambda B^T)^{-1}$ so that the Lemma follows since

$$\left(\frac{\partial\Phi_2}{\partial B}\right)^T = \left(\frac{\partial\Phi_2}{\partial b_{ij}}\right)^T = -(\gamma_{ji})^T = -(\gamma_{ij}) = -\Lambda B^T(B\Lambda B^T)^{-1}M^T(B\Lambda B^T)^{-1}$$

Now define the scaler

$$\Phi_3 = \mathrm{tr}[M^T(BDB^T)^{-1}]$$

and we use the notation $\frac{\partial\Phi_3}{\partial D}$ to denote the matrix of partial derivatives $\left(\frac{\partial\Phi_3}{\partial d_{ij}}\right)$ where $D = (d_{ij})$

<u>Lemma 4</u> - If $\Phi_3 = \mathrm{tr}[M^T(BDB^T)^{-1}]$, then

$$\frac{\partial\Phi_3}{\partial D} = -B^T(BDB^T)^{-1}M^T(BDB^T)^{-1}B \quad \text{so that whenever}$$

$$BB^T = I_k, \quad \Phi_3 = -\mathrm{tr}[BD \frac{\partial\Phi_3}{\partial D} B^T]$$

Proof: $\frac{\partial\Phi_3}{\partial d_{ij}} = \mathrm{tr}[M^T \frac{\partial}{\partial d_{ij}} (BDB^T)^{-1}]$

$$= -\mathrm{tr}\{M^T(BDB^T)^{-1}[\frac{\partial}{\partial d_{ij}} (BDB^T)](BDB^T)^{-1}\}$$

$$= -\mathrm{tr}\{M^T(BDB^T)^{-1}(B \frac{\partial D}{\partial d_{ij}} B^T)(BDB^T)^{-1}\}$$

But for any two matrices A_1, A_2,

$\mathrm{tr}(A_1A_2) = \mathrm{tr}(A_2A_1)$ whenever both matrix products are defined; thus letting

$$A_1 = M^T(BDB^T)^{-1}B \quad \text{and} \quad A_2 = \frac{\partial D}{\partial d_{ij}} B^T(BDB^T)^{-1},$$

it follows

$$\frac{\partial\Phi_3}{\partial d_{ij}} = -\mathrm{tr}\{\frac{\partial D}{\partial d_{ij}} B^T(BDB^T)^{-1}M^T(BDB^T)^{-1}B\}$$

and as in Lemmas 1 and 3 it follows

$$\frac{\partial\Phi_3}{\partial D} = \left(\frac{\partial\Phi_3}{\partial D}\right)^T = -B^T(BDB^T)^{-1}M^T(BDB^T)^{-1}B$$

<u>Lemma 5</u> - If $\Phi_4 = \ln|B\Lambda B^T|$ corresponds to the natural logarithm of the determinant of $B\Lambda B^T$, then

$$\left(\frac{\partial\Phi_4}{\partial B}\right)^T = \Lambda B^T(B\Lambda B^T)^{-1}$$

and thus $B\left(\frac{\partial\Phi_4}{\partial B}\right)^T = I_k$

Proof: Let $\lambda_1,\ldots,\lambda_k$ be the strictly positive eigenvalues of $B\Lambda B^T$, so that

$$\Phi_4 = \frac{1}{2}\,\ln|B\Lambda B^T| = \frac{1}{2}\,\ln(\lambda_1,\ldots,\lambda_n)$$

and thus

$$\frac{\partial\Phi_4}{\partial b_{ij}} = \frac{1}{2}\,\frac{\frac{\partial\lambda_1}{\partial b_{ij}}\,\lambda_2\cdot\ldots\cdot\lambda_k}{\lambda_1\cdots\lambda_k} + \ldots\ldots\ldots + \frac{\lambda_1\cdots\lambda_{k-1}\,\frac{\partial\lambda_k}{\partial b_{ij}}}{\lambda_1\cdots\lambda_k}$$

$$= \frac{1}{2}\,\frac{\frac{\partial\lambda_1}{\partial b_{ij}}}{\lambda_1} + \frac{\frac{\partial\lambda_2}{\partial b_{ij}}}{\lambda_2} + \ldots + \frac{\frac{\partial\lambda_k}{\partial b_{ij}}}{\lambda_k}$$

$$= \frac{1}{2}\,\mathrm{tr}\{W^{-1}\,\frac{\partial W}{\partial b_{ij}}\} \quad \text{where}$$

$$W = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_k \end{pmatrix} \quad \text{is a diagonal matrix}$$

of eigenvalues of $B\Lambda B^T$

But since $(B\Lambda B^T)$ is a real symmetric matrix, there exists an orthogonal matrix U satisfying

$$U(B\Lambda B^T)U^T = W \quad \text{and} \quad UU^T = I_k$$

(Note $(B\Lambda B^T)^{-1} = U^T W^{-1} U$). Thus

$$\frac{\partial W}{\partial b_{ij}} = \frac{\partial U}{\partial b_{ij}} (B\Lambda B^T)U^T + U[\frac{\partial}{\partial b_{ij}}(B\Lambda B^T)]U^T$$

$$+ U(B\Lambda B^T)\frac{\partial U^T}{\partial b_{ij}}$$

and thus

$$U^T \frac{\partial W}{\partial b_{ij}} U = U^T \frac{\partial U}{\partial b_{ij}} (B\Lambda B^T) + \frac{\partial}{\partial b_{ij}}(B\Lambda B^T)$$

$$+ (B\Lambda B^T)\frac{\partial U^T}{\partial b_{ij}} U$$

But

$$\frac{\partial \Phi_4}{\partial b_{ij}} = \frac{1}{2}\operatorname{tr}\{W^{-1}\frac{\partial W}{\partial b_{ij}}\}$$

$$= \frac{1}{2}\operatorname{tr}\{U^T W^{-1} U U^T \frac{\partial W}{\partial b_{ij}} U\}$$

$$= \frac{1}{2}\operatorname{tr}\{(B\Lambda B^T)^{-1}[U^T \frac{\partial U}{\partial b_{ij}}(B\Lambda B^T) + \frac{\partial}{\partial b_{ij}}(B\Lambda B^T) + (B\Lambda B^T)\frac{\partial U^T}{\partial b_{ij}}U]\}$$

$$= \frac{1}{2}\operatorname{tr}(U^T \frac{\partial U}{\partial b_{ij}} + \frac{\partial U^T}{\partial b_{ij}}U) + \frac{1}{2}\operatorname{tr}[(B\Lambda B^T)^{-1}\frac{\partial}{\partial b_{ij}}(B\Lambda B^T)]$$

But since $U^T U = I_k$, it follows

$$U^T \frac{\partial U}{\partial b_{ij}} + \frac{\partial U^T}{\partial b_{ij}}U = 0, \text{ so that}$$

$$\frac{\partial\Phi_4}{\partial b_{ij}} = \frac{1}{2}\,\mathrm{tr}[(B\Lambda B^T)^{-1}\frac{\partial}{\partial b_{ij}}(B\Lambda B^T)]$$

$$= \mathrm{tr}[\frac{\partial B}{\partial b_{ij}}\Lambda B^T(B\Lambda B^T)^{-1}] \quad \text{so that}$$

$$\left(\frac{\partial\Phi_4}{\partial B}\right)^T = \Lambda B^T(B\Lambda B^T)^{-1}, \quad \text{completing the proof.}$$

Now, recall the definition of the function

$$\Phi = \mathrm{tr}(B\Lambda B^T) - \mathrm{tr}\{M^T[(B\Lambda B^T)^{-1} - (BDB^T)^{-1} - I_k\} = 2\Phi_1 - 2\Phi_2 + \Phi_3 + \mathrm{tr}\{M^T\}$$

But

$$\left(\frac{\partial\Phi_1}{\partial B}\right)^T = \Lambda B^T \qquad \text{(Lemma 1)}$$

$$\left(\frac{\partial\Phi_2}{\partial B}\right)^T = -\Lambda B^T(B\Lambda B^T)^{-1}M^T(B\Lambda B^T)^{-1} \qquad \text{(Lemma 3)}$$

$$\left(\frac{\partial\Phi_3}{\partial B}\right)^T = -2DB^T(BDB^T)^{-1}M^T(BDB^T)^{-1} \qquad \text{(Lemma 3)}$$

$$\frac{\partial\Phi_3}{\partial D} = -B^T(BDB^T)^{-1}M^T(BDB^T)^{-1}B \qquad \text{(Lemma 4)}$$

$$\frac{\partial\Phi_1}{\partial D} = \frac{\partial\Phi_2}{\partial D} = 0$$

so that

$$\left(\frac{\partial\Phi}{\partial B}\right)^T = 2\left(\frac{\partial\Phi_1}{\partial B}\right)^T - 2\left(\frac{\partial\Phi_2}{\partial B}\right)^T + \left(\frac{\partial\Phi_3}{\partial B}\right)^T$$

$$= 2\Lambda B^T[I_k + (B\ B^T)^{-1}M^T(B\ B^T)^{-1}]$$

$$- 2DB^T(BDB^T)^{-1}M^T(BDB^T)^{-1}$$

and

$$\left(\frac{\partial\Phi}{\partial D}\right) = \left(\frac{\partial\Phi}{\partial D}\right)^T = -\ B^T(BDB^T)^{-1}M^T(BDB^T)^{-1}B$$

where we have assumed $\qquad \frac{\partial M^T}{\partial B} = \frac{\partial M^T}{\partial D} = 0$

ADDITIONAL CONSIDERATIONS

Consider the problem of maximizing

$$X = tr(B\Lambda B^T)$$

subject to the constraint $I_k - B\Lambda B^T$ is positive definite.

Since $I_k - B\Lambda B^T$ is symmetric, there exists an orthogonal matrix Q satisfying

$$Q(I_k - B\Lambda B^T)Q^T = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_k \end{pmatrix} = S$$

and $QQ^T = I_k$, and S is a diagonal matrix of eigenvalues of $I_k - B\Lambda B^T$. Thus the constraint is equivalent to $\lambda_i > 0$ for all i. Also,

$$
\begin{aligned}
X &= \mathrm{tr}(B\Lambda B^T) \\
&= \mathrm{tr}(QB\Lambda B^T Q^T) \qquad \text{(for any orthogonal matrix } Q) \\
&= \mathrm{tr}(\hat{B}\Lambda\hat{B}^T)
\end{aligned}
$$

where $\hat{B} = QB$ is a k by n matrix

It follows $\hat{B}$ satisfies

$$I_k - \hat{B}\Lambda\hat{B}^T = S$$

so that the row vectors of $\hat{B}$ may always be chosen Λ orthogonal. Alternately, let

$$\hat{B} = \begin{pmatrix} b_1^T \\ \\ b_k^T \end{pmatrix}$$

and the constraints are equivalent to

$$b_i^T \Lambda b_j = 0 \qquad \begin{array}{l} i = 1,\ldots,k-1 \\ j = i+1,\ldots,k \end{array}$$

$$1 - b_i^T \Lambda b_i > 0 \qquad i = 1,\ldots,k$$

Thus the problem is to maximize

$$X = \mathrm{tr}(\hat{B}\Lambda\hat{B}^T)$$

subject to the above constraints.

Thus it appears the solution to the problem

$$\max X = tr(B\Lambda B^T)$$

subject to the constraint $I_k - B\Lambda B^T$ is positive definite is given by any k eigenvectors $e_1, e_2, \ldots, e_k$ of Λ, appropriately "scaled" with scaler α so that

$$1 - \alpha^2 e_i^T \Lambda e_i > 0 \qquad \forall_i$$

where we assume $e_i^T \Lambda e_i = 1 \qquad \forall_i$