SOME THEOREMS ON MATRIX DIFFERENTIATION WITH SPECIAL REFERENCE TO KRONECKER MATRIX PRODUCTS

H. NEUDECKER*
University of Birmingham, U. K.
Middle East Technical University, Ankara

Matrix differentiation, the procedure of finding partial derivatives of the elements of a matrix function with respect to the elements of the argument matrix, is the subject matter of this paper. The method followed here amounts to linking the differentials of the matrix function and the argument matrix, and then identifying matrices of partial derivatives. The basic assumption made is mathematical independence of the elements of the argument matrix. Matrix functions are divided into two categories: Kronecker matrix products and non-Kronecker (=ordinary) matrix products. Three definitions of partial derivatives matrices are used, to be denoted by D_1, D_2 and D_3 in this abstract. D_2 and D_3 are applied to Kronecker products, D_1 is applied to ordinary products. This is a matter of efficiency only. D_1 is developed first. Transforming matrices into column vectors turns out to be very convenient. D_2 and D_3 are developed then. D_3 is a generalisation of D_1.

1. INTRODUCTION

TURNBULL, Dwyer and MacPhail, Deemer and Olkin, Olkin, Bodewig, Roy, Dwyer and others have presented a series of theorems for solving problems in the field of matrix differentiation. However, there is still a need for developing others, primarily since the theorems available at the moment are not adequate for all the matrix derivatives currently needed. In particular, matrix differentiation for Kronecker matrix products is needed in many applications. This category comprises matrix functions whose differentials can be written as $A\otimes B(dX)C$ or $B(dX)C\otimes A$ (or their transposes). The present matrix calculus permits such calculations in a relatively simple manner.

2. SOME MATRIX ALGEBRA THEOREMS

(2.1) Let $A=[a_{ij}]$ be an (m, n) matrix and B an (s, t) matrix, then the Kronecker product $A\otimes B$ is defined as the (ms, nt) matrix

$$A \otimes B = [a_{ij}B].$$

We state the following properties of Kronecker matrix products, all of which may be proved in an elementary fashion.

The matrices involved are assumed to be conformable. In (2.7) it is further assumed that A and B are *square* of order m and s respectively.

(2.2) $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$

(2.3) $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$

(2.4) $(A \otimes B)' = A' \otimes B'$

(2.5) $(A + B) \otimes (C + D) = (A \otimes C) + (A \otimes D) + (B \otimes C) + (B \otimes D)$

* The author is indebted to editors and referees of J.A.S.A. for their most constructive criticism.

(2.6) $A \otimes (B \otimes C) = (A \otimes B) \otimes C$

(2.7) $|A \otimes B| = |A|^s |B|^m$

(2.8) $\operatorname{tr}(A \otimes B) = (\operatorname{tr} A)(\operatorname{tr} B)$.

In the following a_{ij} will denote the ijth element of A. $A_{i.}$ and $A_{.j}$ will denote the ith row and jth column of A, respectively. If A is of order (m, n), we define the (mn) column vector vec A:

(2.9) $$\operatorname{vec} A = \begin{bmatrix} A_{.1} \\ \cdot \\ \cdot \\ \cdot \\ A_{.n} \end{bmatrix}$$

In particular, if y is a column vector, then vec $y = \operatorname{vec} y' = y$.

We now give five theorems connecting vec A and the Kronecker product. The proofs are straightforward. It is assumed that the matrices involved are conformable. The unit matrices have appropriate orders.

(2.10) $$\operatorname{vec} ABC = (C' \otimes A) \operatorname{vec} B.$$

Proof: The jth subvector of vec ABC equals

$$ABC_{.j} = \sum_i c_{ij} AB_{.i} = (C'_{j.} \otimes A) \operatorname{vec} B.$$

The conclusion follows.

(2.10.1) $$\operatorname{vec} AB = (I \otimes A) \operatorname{vec} B = (B' \otimes I) \operatorname{vec} A = (B' \otimes A) \operatorname{vec} I = \sum_i B'_{.i} \otimes A_{.i}.$$

Proof: By substituting $C = I$; $A = I$, $B = A$, $C = B$; $B = I$, $C = B$ respectively, in (2.10) the first three results follow easily. The fourth result follows from the third.

(2.11) $$\operatorname{tr} ABC = (\operatorname{vec} A')'(I \otimes B) \operatorname{vec} C$$

(2.11.1) $$\operatorname{tr} AB = (\operatorname{vec} A')' \operatorname{vec} B$$

(2.12) $$\operatorname{tr} AZ'BZC = (\operatorname{vec} Z)'(A'C' \otimes B) \operatorname{vec} Z = (\operatorname{vec} Z)'(CA \otimes B') \operatorname{vec} Z.$$

Proof: Application of 2.10 and 2.11.1 gives $\operatorname{tr} AZ'BZC = \operatorname{tr} Z'BZCA = (\operatorname{vec} Z)'$ vec $BZCA = (\operatorname{vec} Z)' (A'C' \otimes B) \operatorname{vec} Z$. Transposition obviously does not affect this result; it leads to $(\operatorname{vec} Z)' (CA \otimes B') \operatorname{vec} Z$ by (2.4.)

3. DEFINITIONS ON MATRIX DIFFERENTIAL CALCULUS

(3.1) Let

$$dX = [dx_{ij}], \quad \frac{dY}{d\xi} = \left[\frac{dy_{kl}}{d\xi}\right], \quad \frac{\partial f}{\partial X} = \left[\frac{\partial f}{\partial x_{ij}}\right],$$

$$\frac{\partial y}{\partial x} = \left[\frac{\partial y_j}{\partial x_i}\right], \quad \frac{\partial y_j}{\partial x_i} \text{ being the } ij\text{th element of } \frac{\partial y}{\partial x};$$

where Y is a matrix function of the scalar ξ, f is a scalar function of the matrix X, and y is a column vector function of the column vector x. The third and the fourth definitions agree if X is a column vector and y is a scalar.

(3.2) If Y is a matrix function of the matrix X, $\partial \text{ vec } Y/\partial \text{ vec } X$ is a matrix of partial derivatives $\partial y_{kl}/\partial x_{ij}$, *uniquely* ordered. Especially for the treatment of Kronecker matrix products, we define two more partial derivative matrices:

$$\left[\frac{\partial Y}{\partial x_{ij}}\right] = \begin{bmatrix} \dfrac{\partial Y}{\partial x_{11}} & \cdots & \dfrac{\partial Y}{\partial x_{1m}} \\ \vdots & & \vdots \\ \dfrac{\partial Y}{\partial x_{n1}} & \cdots & \dfrac{\partial Y}{\partial x_{nm}} \end{bmatrix}, \quad X \text{ being of order } (n, m). \tag{3.3}$$

$$\left[\frac{\partial \text{ vec } Y_{qr}}{\partial \text{ vec } X}\right] = \begin{bmatrix} \dfrac{\partial \text{ vec } Y_{11}}{\partial \text{ vec } X} & \cdots & \dfrac{\partial \text{ vec } Y_{1v}}{\partial \text{ vec } X} \\ \vdots & & \vdots \\ \dfrac{\partial \text{ vec } Y_{u1}}{\partial \text{ vec } X} & \cdots & \dfrac{\partial \text{ vec } Y_{uv}}{\partial \text{ vec } X} \end{bmatrix}, \tag{3.4}$$

where Y has the following partitioning:

$$Y = [Y_{qr}] \qquad q = 1 \cdots u; \quad r = 1 \cdots v.$$

The matrices $[\partial Y/\partial x_{ij}]$ and $[\partial \text{ vec } Y_{qr}/\partial \text{ vec } X]$ were referred to as D_2 and D_3 in the abstract.

4. MATRIX DIFFERENTIATION THEOREMS. FIRST PART

In this section we shall list theorems on differentiation of matrix functions *not* being Kronecker matrix products. Kronecker matrix products will be considered in section 5.

Argument matrices (and vectors) can have any order except (1, 1). Matrix (and vector) functions can have any order. We shall make the *assumption*: $\partial x_{kl}/\partial x_{ij}=0$ for $i\neq k$ or $j\neq l$, both x_{kl} and x_{ij} being elements of the argument matrix X. With this assumption we can easily identify partial derivatives from relations between differentials. The main reasoning will, therefore, be done in terms of differentials. The three following theorems will enable us to work with matrix differentials:

(4.1) For conformable matrices X and Y: $d(XY)=(dX)Y+XdY$.

(4.2) If l is a linear scalar function of the matrix X then $dl(X)=l(dX)$.

(4.3) $d \text{ vec } X = \text{vec } dX$.

The following five theorems are straightforward. They involve: f, a scalar function of the matrix X; y, a column vector function of the column vector x; g, a scalar function of the column vector x; Y, a matrix function of the matrix X; the matrices P, M, N, H and the column vector m being of appropriate orders.

(4.4) $\partial f/\partial X=P$ implies $\partial f/\partial \text{ vec } X=\text{vec } P$, and conversely.

(4.5) $dy = M dx$ implies $\partial y/\partial x = M'$, and conversely.
(4.5.1) vec $dY = N$ vec dX implies ∂ vec Y/∂ vec $X = N'$, and conversely.
(4.5.2) $dg = m'dx$ implies $\partial g/\partial x = m$, and conversely.
(4.5.3) $df = \operatorname{tr} H'dX$ implies $\partial f/\partial X = H$, and conversely.

With the help of the previous theorems we can now establish the main theorem of section 4:

(4.6) If the matrix Y is a function of the matrix X, X having t columns, the matrices P_i, Q_i, R_j and S_j ($i = 1 \cdots f; j = 1 \cdots g$) being of appropriate orders, then

$$dY = \sum_i P_i(dX)Q_i + \sum_j R_j(dX)'S_j$$

implies

$$\frac{\partial \text{ vec } Y}{\partial \text{ vec } X} = \sum_i Q_i \otimes P_i' + \sum_j \begin{bmatrix} S_j \otimes (R_j')_{1.} \\ \cdot \\ \cdot \\ S_j \otimes (R_j')_{t.} \end{bmatrix},$$

and conversely.

Proof: We can without loss of generality confine ourselves to $dY = P(dX)Q + R(dX)'S$. From this we derive

$$\begin{aligned} \text{vec } dY &= \text{vec } P(dX)Q + \text{vec } R(dX)'S \\ &= (Q' \otimes P) \text{ vec } dX + (S' \otimes R) \text{ vec } dX', \end{aligned}$$

by means of (2.10). According to (4.5.1) and (2.4) the first term contributes towards ∂ vec Y/∂ vec X: $Q \otimes P'$; the second term contributes towards ∂ vec Y/∂ vec X': $S \otimes R'$. Given the structural relationship between ∂ vec Y/∂ vec X' and ∂ vec Y/∂ vec X the second term contributes towards ∂ vec Y/∂ vec X:

$$\begin{bmatrix} S \otimes (R')_{1.} \\ \cdot \\ \cdot \\ S \otimes (R')_{t.} \end{bmatrix}.$$

This concludes the proof.

We shall now quote some theorems about *linking* derivative matrices. They involve: scalar f; column vectors x, y and z; matrices X, Y and Z; constant matrix A:

(4.7) $$\frac{\partial z}{\partial x} = \frac{\partial y}{\partial x} \cdot \frac{\partial z}{\partial y}$$

(4.7.1) $$\frac{\partial \text{ vec } Z}{\partial \text{ vec } X} = \frac{\partial \text{ vec } Y}{\partial \text{ vec } X} \cdot \frac{\partial \text{ vec } Z}{\partial \text{ vec } Y}$$

(4.7.2) $$\frac{\partial f}{\partial x} = \frac{\partial y}{\partial x} \frac{\partial f}{\partial y}$$

(4.8) $$\frac{\partial y}{\partial x} = A' \frac{\partial y}{\partial Ax}$$

Proof: According to (4.7): $\partial y/\partial x = \partial Ax/\partial x \; \partial y/\partial Ax$.
According to (4.5): $\partial Ax/\partial x = A'$. From this the theorem follows.

$$\frac{\partial \text{ vec } Y}{\partial \text{ vec } X} = (I \otimes A') \frac{\partial \text{ vec } Y}{\partial \text{ vec } AX} \tag{4.8.1}$$

$$\frac{\partial f}{\partial x} = A' \frac{\partial f}{\partial Ax} \, . \tag{4.8.2}$$

It goes without saying that the matrix of second-order partial derivatives of a *scalar* function can be derived by applying (4.6) to the matrix of first-order partial derivatives. Let us use the notation

$$\hat{\nabla}^2_{\text{vec } X} f = \frac{\partial \dfrac{\partial f}{\partial \text{ vec } X}}{\partial \text{ vec } X} \, .$$

We can then establish the two useful theorems for scalar f:

(4.9) $\quad d^2 f = \text{tr } A(dX)'B(dX)C \quad$ implies

$$\hat{\nabla}^2_{\text{vec } X} f = \tfrac{1}{2}(A'C' \otimes B + CA \otimes B'), \quad \text{and conversely.}$$

Proof: By definition $d^2 f = (\text{vec } dX)' \, \nabla^2_{\text{vec } X} f \, (\text{vec } dX)$. We rewrite:

$$\text{tr } A(dX)'B(dX)C = \tfrac{1}{2}(\text{vec } dX)'(A'C' \otimes B + CA \otimes B') \text{ vec } dX,$$

using (2.12). From this the theorem follows.

$$d^2 f = (dx)'Adx \quad \text{implies} \quad \nabla^2_x f = \tfrac{1}{2}(A + A'). \tag{4.9.1}$$

5. MATRIX DIFFERENTIATION THEOREMS. SECOND PART

Obviously the theorems of the previous section, in particular theorem 4.6, cannot be applied to Kronecker matrix products. There is a need, however, for treatment of such products. Three obvious cases are: First-order partial derivatives of a Kronecker product involving a matrix X, second-order partial derivatives of a matrix Y with respect to a matrix X, and third-order partial derivatives of a scalar function of a matrix X with respect to X. We shall employ the two types of derivatives defined in section 3 under (3.2) and (3.3). Five theorems will be developed.

(5.1) $dY = A \otimes B(dX)C$ implies

$$\left[\frac{\partial Y}{\partial x_{ij}}\right] = [A \otimes B_{.i}C_{j.}], \quad \text{and conversely.}$$

Proof: Using elementary matrices E_{ij}, having zeros everywhere except in the ijth position where they have unity, we rewrite:

$$\begin{aligned} dY &= A \otimes B(dX)C = A \otimes B(\sum_{ij} E_{ij} dx_{ij})C \\ &= A \otimes \sum_{ij} (BE_{ij}C) dx_{ij} = A \otimes \sum_{ij} B_{.i}C_{j.} dx_{ij} \\ &= \sum_{ij} (A \otimes B_{.i}C_{j.}) dx_{ij}. \end{aligned}$$

Thus

$$\frac{\partial Y}{\partial x_{ij}} = A \otimes B_{.i}C_{j.}$$

and

$$\left[\frac{\partial Y}{\partial x_{ij}}\right] = [A \otimes B_{.i}C_{j.}].$$

This proves the theorem.

(5.2) $dZ = F(dX)G \otimes H$ implies

$$\left[\frac{\partial Z}{\partial x_{ij}}\right] = (\text{vec } F)(\text{vec } G')' \otimes H, \quad \text{and conversely.}$$

Proof: Obviously

$$dZ = F(dX)G \otimes H = \sum_{ij} (F_{.i}G_{j.} \otimes H)dx_{ij}.$$

Thus

$$\frac{\partial Z}{\partial x_{ij}} = F_{.i}G_{j.} \otimes H$$

and

$$\left[\frac{\partial Z}{\partial x_{ij}}\right] = [F_{.i}G_{j.} \otimes H] = (\text{vec } F)(\text{vec } G')' \otimes H.$$

(5.3) If $dY = A \otimes B(dX)C$, where $dY = [dY_{qr}]$, $dY_{qr} = a_{qr}B(dX)\text{C}$, then

$$\left[\frac{\partial \text{ vec } Y_{qr}}{\partial \text{ vec } X}\right] = A \otimes C \otimes B'$$

and conversely.

Proof: $dY_{qr} = a_{qr}B(dX)C$, so

$$\frac{\partial \text{ vec } Y_{qr}}{\partial \text{ vec } X} = C \otimes (a_{qr}B)' = a_{qr}C \otimes B',$$

according to (4.6). Therefore

$$\left[\frac{\partial \text{ vec } Y_{qr}}{\partial \text{ vec } X}\right] = A \otimes C \otimes B'.$$

(5.4) If $dZ = F(dX)G \otimes H$, where $dZ = [dZ_{qr}]$, $dZ_{qr} = \{F_{q.}(dX)G_{.r}\}H$, $q = 1 \cdots w$, then

$$\left[\frac{\partial \text{ vec } Z_{qr}}{\partial \text{ vec } X}\right] = \begin{Bmatrix} G \otimes (F_{1.})' (\text{vec } H)' \\ \vdots \\ G \otimes (F_{w.})' (\text{vec } H)' \end{Bmatrix},$$

and conversely.

Proof: $dZ_{qr}=\{F_{q.}(dX)G_{.r}\}H$. So

$$d \text{ vec } Z_{qr} = \{F_{q.}(dX)G_{.r}\} \text{ vec } H = (\text{vec } H)F_{q.}(dX)G_{.r},$$

and consequently

$$\begin{aligned}\frac{\partial \text{ vec } Z_{qr}}{\partial \text{ vec } X} &= G_{.r} \otimes \{(\text{vec } H)F_{q.}\}' \\ &= G_{.r} \otimes (F_{q.})'(\text{vec } H)' \quad \text{by 4.6.}\end{aligned}$$

This leads to

$$\left[\frac{\partial \text{ vec } Z_{qr}}{\partial \text{ vec } X}\right] = \begin{bmatrix} G \otimes (F_{1.})' (\text{vec } H)' \\ \cdot \\ \cdot \\ G \otimes (F_{w.})' (\text{vec } H)' \end{bmatrix}.$$

We can conclude from theorems (5.1–5.4) that the derivative $[\partial \text{ vec } Y_{qr}/\partial \text{ vec } X]$ notationally best suits differentials like $dY=A\otimes B(dX)C$, and that the derivative $[\partial Y/\partial x_{ij}]$ best suits differentials like $dY=F(dX)G\otimes H$. In case of a *mixed* differential like $dY=A\otimes B(dX)C+F(dX)G\times H$, we obviously have to choose one of the two derivatives. We cannot have a mixture of definitions. $[\partial Y/\partial x_{ij}]$ definitely has the edge on $[\partial \text{ vec } Y_{qr}/\partial \text{ vec } X]$ in general. The reason is that application of the second derivative presupposes one unique partition of dY. As can be seen from (5.3) and (5.4) this requirement cannot be met generally. In (5.3) it is the order of A that determines the partition; in (5.4) it is the order of $F(dX)G$ that does so, and there is no reason why the two orders should be identical. A comparable situation arises when Y depends on X'. It is obvious that $[\partial Y/\partial x_{ji}]$ can easily be derived from $[\partial Y/\partial x_{ij}]$ by *blockwise* transposition. $[\partial \text{ vec } Y_{qr}/\partial \text{ vec } X']$ cannot so easily be derived from $[\partial \text{ vec } Y_{qr}/\partial \text{ vec } X]$.

6. JACOBIANS

If Y is a new set of $m\times n$ variables and X is the original set of also $m\times n$ variables, the Jacobian will be equal to (the absolute value of) the determinant of $[\partial \text{ vec } Y/\partial \text{ vec } X]^{-1}$. With the help of theorem (4.6) this expression can be evaluated, provided $\partial \text{ vec } Y/\partial \text{ vec } X$ is nonsingular. Complications arise if columns of $\partial \text{ vec } Y/\partial \text{ vec } X$ are linearly dependent. This may happen, if e.g., elements of Y are multiple, proportional, or if they have relationships of the type: $y_{ij}=y_{kt}+c$, c constant. This may happen even if the elements of X are still mathematically independent. If this assumption basic to our analysis has to be dropped, the position becomes even more complicated. Theorem (4.6) cannot be applied then any longer. A case of this type is the transformation $Y=TXT'$ where $X'=X$, T being a constant matrix. Clearly $Y'=Y$, so $\partial \text{ vec } Y/\partial \text{ vec } X$ contains multiple columns. It also contains multiple rows because $X'=X$. Therefore, $T'\otimes T'\neq\partial \text{ vec } Y/\partial \text{ vec } X$.

There are several approaches to this kind of problem but we shall not discuss them here.

7. APPLICATIONS

(7.1) *Jacobians*. In the following we shall evaluate the Jacobians of several transformations. All matrices will be assumed to be square of order (n, n), X and Y variable, A, B etc. constant.

(7.1.1) $Y=AXB$ leads to $dY=A(dX)B$,

$$\frac{\partial \text{ vec } Y}{\partial \text{ vec } X} = B \otimes A',$$

$$\left|\frac{\partial \text{ vec } Y}{\partial \text{ vec } X}\right|^{-1} = |B \otimes A'|^{-1} = |A|^{-n}|B|^{-n} \quad \text{by (2.7) and (4.6)}$$

(7.1.2) $Y=X^2$ leads to $dY=XdX+(dX)X$,

$$\frac{\partial \text{ vec } Y}{\partial \text{ vec } X} = I \otimes X' + X \otimes I,$$

$$\left|\frac{\partial \text{ vec } Y}{\partial \text{ vec } X}\right|^{-1} = |I \otimes X' + X \otimes I|^{-1} = \left(\prod_k \xi_k\right)^{-n} \prod_{ij}\left(1 + \frac{\xi_i}{\xi_j}\right)^{-1},$$

the ξ_i being the characteristic roots of X.

For $n = 3$ the Jacobian becomes $\dfrac{9}{8[(\text{tr } X)^3 - \text{tr } X^3]^2 |X|}$.

For $n = 2$ the Jacobian reduces to $\dfrac{1}{4(\text{tr } X)^2 |X|}$.

(7.2) *Miscellaneous*

(7.2.1) $Y=X^2$ has $\partial \text{ vec } Y/\partial \text{ vec } X=I\otimes X'+X\otimes I$. Further, $dZ=d(\partial \text{ vec } Y/\partial \text{ vec } X)=I\otimes(dX)'+(dX)\otimes I$. From this we derive the matrix of second-order partial derivatives:

$$\left[\frac{\partial Z}{\partial x_{ij}}\right] = [I \otimes E_{ij}] + (\text{vec } I)(\text{vec } I)' \otimes I.$$

(7.2.2) $Y=X\otimes X$, X being a matrix of order (m, n), has first-order differential $dY=(dX)\otimes X+X\otimes dX$. This implies

$$\left[\frac{\partial Y}{\partial x_{ij}}\right] = (\text{vec } I_m)(\text{vec } I_n)' \otimes X + [X \otimes I_{.i}I_{j.}]$$

and

$$\left[\frac{\partial \text{ vec } Y_{ij}}{\partial \text{ vec } X}\right] = \begin{Bmatrix} I_n \otimes (I_{1.})' (\text{vec } X)' \\ \cdot \\ \cdot \\ \cdot \\ I_n \otimes (I_{m.})' (\text{vec } X)' \end{Bmatrix} + X \otimes I_n \otimes I_m,$$

where

$$dY_{ij} = dx_{ij}X + x_{ij}dX. \qquad (i = 1 \cdots m;\ j = 1 \cdots n)$$

(7.3) *Maximum Likelihood Estimation.* Let $U=[u_{ij}]$ be an (N, T) matrix of normally distributed random variables with $Eu_{ij}=0$ and

$$Eu_{ij}u_{k\ell} = \begin{cases} 0 & \text{if } j \neq \ell \\ \sigma_{ik} & \text{if } j = \ell \end{cases}.$$

Consider the model

$$BY + \Gamma Z = U,$$

where B and Γ are (N, N) and (N, Λ) matrices of parameters. The problem is to obtain maximum likelihood estimators of B and Γ. Define $\Sigma=[\sigma_{ik}]$ $(i, k=1 \cdots N)$ so that $EU=0$ and $E(1/T)UU'=\Sigma$. The log likelihood is given by

$$\log p(U) = \text{constant} + T \log | B | + \frac{T}{2} \log | \Sigma^{-1} | - \frac{T}{2} \operatorname{tr} \Sigma^{-1} AMA',$$

where

$$M = \frac{1}{T}\begin{bmatrix} Y \\ Z \end{bmatrix}(Y' \,\vdots\, Z'), \quad \text{and} \quad A = (B \,\vdots\, \Gamma).^1$$

Differentiation of $\log p(U)$ gives:

$$0 = d \log p(U) = T \operatorname{tr} B^{-1}dB + \frac{T}{2} \operatorname{tr} \Sigma d\Sigma^{-1} - \frac{T}{2} \operatorname{tr} AMA'd\Sigma^{-1}$$

$$2 \qquad 2$$

$$- T \operatorname{tr} MA'\Sigma^{-1}dA = T \operatorname{tr}\left\{\begin{pmatrix} B^{-1} \\ 0 \end{pmatrix} - MA'\Sigma^{-1}\right\} dA$$

$$+ \frac{T}{2} \operatorname{tr}(\Sigma - AMA')d\Sigma^{-1},$$

From which follow the likelihood equations:

$$\begin{pmatrix} B^{-1} \\ 0 \end{pmatrix} = MA'\Sigma^{-1} \quad \text{and} \tag{1}$$

$$\Sigma = AMA'. \tag{2}$$

After substitution we get:

$$\begin{pmatrix} B^{-1} \\ 0 \end{pmatrix} = MA'(AMA')^{-1}. \tag{3}$$

This equation is nonlinear in the unknown parameters. For dealing with the nonlinearity several procedures have been devised. They involve $\nabla^2_{\text{vec } A}\lambda^*$, where λ^* is the concentrated likelihood function:

$$\lambda^* = \text{constant} + T \log | B | - \frac{T}{2} \log | AMA' | .$$

---

[1] See Fisk (pp. 6 et seq., 22 et seq.)

The Newton-Raphson method is one of these methods. (If all equations are just-identified solving equation (3) is straightforward.)

We shall now derive $\nabla^2_{\text{vec } A}\lambda^*$. Obviously

$$\partial\lambda^*/\partial A = T[(B')^{-1} \vdots\, 0] - T(AMA')^{-1}AM.$$

(By (4.5.3).) Then

$$\begin{aligned} d\frac{\partial\lambda^*}{\partial A} = &- T[(B')^{-1}\ \vdots 0](dA)'[(B')^{-1}\ \vdots 0] \\ &+ T(AMA')^{-1}(dA)MA'(AMA')^{-1}AM \\ &+ T(AMA')^{-1}AM(dA)'(AMA')^{-1}AM - T(AMA')^{-1}(dA)M, \end{aligned}$$

from which follows (by virtue of (4.6)):

$$\begin{aligned} \nabla^2_{\text{vec } A}\lambda^* = &- T\begin{Bmatrix} [(B')^{-1}\ \vdots 0] \otimes B_{1.}^{-1} \\ \vdots \\ [(B')^{-1}\ \vdots 0] \otimes B_{p.}^{-1} \end{Bmatrix} + TMA'(AMA')^{-1}AM \otimes (AMA')^{-1} \\ &+ T\begin{bmatrix} (AMA')^{-1}AM \otimes \{MA'(AMA')^{-1}\}_{1.} \\ \vdots \\ (AMA')^{-1}AM \otimes \{MA'(AMA')^{-1}\}_{q.} \end{bmatrix} - TM \otimes (AMA')^{-1}. \end{aligned}$$

(We put: $q = N + \Lambda,\quad p = N$.)

REFERENCES

[1] Bellman, R. *Introduction to Matrix Analysis.* New York: McGraw-Hill Book Company, 1960.

[2] Bodewig, E. *Matrix Calculus.* Amsterdam: North Holland Publishing Company, 1956.

[3] Deemer, W. L. and Olkin, I., "Jacobians of matrix transformations useful in multivariate analysis," *Biometrika,* 38 (1951), 345–67.

[4] Dwyer, P. S., "Some applications of matrix derivatives in multivariate analysis," *Journal of the American Statistical Association,* 62 (1967), 607–25.

[5] Dwyer, P. S. and MacPhail, M. S., "Symbolic matrix derivatives," *Annals of Mathematical Statistics,* 19 (1948), 517–34.

[6] Fisk, P. R. *Stochastically Dependent Equations.* London: Charles Griffin and Company, 1967.

[7] Jack, H., "Jacobians of transformations involving orthogonal matrices," *Proceedings of the Royal Society of Edinburgh,* Section A, LXVII (1964–5), 81–103.

[8] Olkin, I., "Note on 'Jacobians of matrix transformations useful in multivariate analysis'," *Biometrika,* 40(1953), 43–46.

[9] Rothenberg, T. J. and Leenders, C. T., "Efficient estimation of simultaneous equation systems," *Econometrica,* 32 (1964) 57–76.

[10] Roy, S. N. *Some Aspects of Multivariate Analysis.* New York: John Wiley and Sons, 1957.

[11] Turnbull, H. W., "On differentiating a matrix," *Proceedings of the Edinburgh Mathematical Society,* Second Series, I (1927–29), 111–28.

[12] Turnbull, H. W., "A matrix form of Taylor's theorem," *Proceedings of the Edinburgh Mathematical Society,* Second Series, II (1930–31), 33–54.

[13] Turnbull, H. W., "Matrix differentiation of the characteristic function," *Proceedings of the Edinburgh Mathematical Society,* Second Series, II (1930–31), 256–64.

POSTSCRIPT

When reading the proofs, I noticed a paper by Conlisk. This provides some interesting applications from which I take the following: To find

$$\frac{\partial \operatorname{vec} S}{\partial \operatorname{vec} A}$$

for $S = ASA' + V$, where $S' = S$ is an $m \times m$ matrix. Differentiation leads to:

$$dS = (dA)SA' + A(dS)A' + AS(dA)' + dV$$

or

$$dT = (dA)SA' + AS(dA)' + dV,$$

where

$$dT = dS - A(dS)A'.$$

Using 4.6 and 4.7.1 we get:

$$\frac{\partial \operatorname{vec} S}{\partial \operatorname{vec} A} = \frac{\partial \operatorname{vec} T}{\partial \operatorname{vec} A}\left(\frac{\partial \operatorname{vec} T}{\partial \operatorname{vec} S}\right)^{-1}$$
$$= [SA' \otimes I + (I \otimes SA')_+](I \otimes I - A' \otimes A')^{-1},$$

where

$$(I \otimes SA')_+ \text{ is shorthand for } \begin{bmatrix} I \otimes (SA')_{1.} \\ \vdots \\ I \otimes (SA')_{m.} \end{bmatrix}$$

REFERENCE

[1] Conlisk, J., "The equilibrium covariance matrix of dynamic econometric models," *Journal of the American Statistical Association,* 64 (1969) 277–279.