SYMBOLIC MATRIX DERIVATIVES

By Paul S. Dwyer and M. S. Macphail

University of Michigan and Queen's University

Summary. Let X be the matrix $[x_{mn}]$, t a scalar, and let $\partial X/\partial t$, $\partial t/\partial X$ denote the matrices $[\partial x_{mn}/\partial t]$, $[\partial t/\partial x_{mn}]$ respectively. Let $Y = [y_{pq}]$ be any matrix product involving X, X' and independent matrices, for example $Y = AXBX'C$. Consider the matrix derivatives $\partial Y/\partial x_{mn}$, $\partial y_{pq}/\partial X$. Our purpose is to devise a systematic method for calculating these derivatives. Thus if $Y = AX$, we find that $\partial Y/\partial x_{mn} = AJ_{mn}$, $\partial y_{pq}/\partial X = A'K_{pq}$, where J_{mn} is a matrix of the same dimensions as X, with all elements zero except for a unit in the m-th row and n-th column, and K_{pq} is similarly defined with respect to Y. We consider also the derivatives of sums, differences, powers, the inverse matrix and the function of a function, thus setting up a matrix analogue of elementary differential calculus. This is designed for application to statistics, and gives a concise and suggestive method for treating such topics as multiple regression and canonical correlation.

1. Introduction. The derivative of a matrix with respect to a scalar

$$\frac{\partial Y}{\partial x} = \frac{\partial}{\partial x}[y_{pq}] = \left[\frac{\partial y_{pq}}{\partial x}\right] \tag{1}$$

is well known and commonly used. The symbolic derivative obtained by applying a matrix of differential operators to a scalar

$$\frac{\partial y}{\partial X} = \left[\frac{\partial}{\partial x_{mn}}\right] y = \left[\frac{\partial y}{\partial x_{mn}}\right] \tag{2}$$

is not in such general use though some authors give special cases. For example, if A is a symmetric matrix and X a column matrix, so that $y = X'AX$ is a quadratic form, Fraser, Duncan and Collar [1, p. 48] write

$$\begin{bmatrix} \partial/\partial x_1 \\ \partial/\partial x_2 \\ \vdots \\ \partial/\partial x_n \end{bmatrix} y = 2AX \tag{3}$$

to indicate concisely the result of differentiating y with respect to the elements x_i of X.

It is to be noted that the matrix in (1) has the same dimensions (numbers of rows and columns) as the matrix Y, while the matrix in (2) has the dimensions of the matrix X.

We present an illustration of each of these types of symbolic matrix derivatives in order to clarify the concepts. Thus if

$$Y = \begin{bmatrix} x & 2x^3 & 3x^{-4} \\ e^x & \sin x & \log_e x \end{bmatrix},$$

we have

$$\frac{\partial Y}{\partial x} = \begin{bmatrix} 1 & 6x^2 & -12x^{-5} \\ e^x & \cos x & x^{-1} \end{bmatrix},$$

while if $y = x_{11}x_{32} - x_{31}x_{12}$ and

$$X = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{bmatrix},$$

we have

$$\frac{\partial y}{\partial X} = \begin{bmatrix} x_{32} & -x_{31} \\ 0 & 0 \\ -x_{12} & x_{11} \end{bmatrix}.$$

Suppose Y is any matrix product involving X, X' and independent matrices, for example, $Y = AXBX'C$. We may fix an element x_{mn} of X and form the matrix

$$\frac{\partial Y}{\partial x_{mn}}, \tag{4}$$

or we may fix an element y_{pq} of Y and form the matrix

$$\frac{\partial y_{pq}}{\partial X}. \tag{5}$$

The purpose of this paper is to devise a systematic method for calculating these matrices, and to give various applications in the general field of statistics.

By way of introduction we take the matrix product $Y = AX$ where

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \quad \text{and } X = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{bmatrix},$$

so that

$$Y = \begin{bmatrix} a_{11}\,x_{11} + a_{12}\,x_{21} + a_{13}\,x_{31} & a_{11}\,x_{12} + a_{12}\,x_{22} + a_{13}\,x_{32} \\ a_{21}\,x_{11} + a_{22}\,x_{21} + a_{23}\,x_{31} & a_{21}\,x_{12} + a_{22}\,x_{22} + a_{23}\,x_{32} \end{bmatrix}.$$

We have then

$$\frac{\partial Y}{\partial x_{11}} = \begin{bmatrix} a_{11} & 0 \\ a_{21} & 0 \end{bmatrix}, \qquad \frac{\partial Y}{\partial x_{12}} = \begin{bmatrix} 0 & a_{11} \\ 0 & a_{21} \end{bmatrix},$$

$$\frac{\partial Y}{\partial x_{21}} = \begin{bmatrix} a_{12} & 0 \\ a_{22} & 0 \end{bmatrix}, \qquad \frac{\partial Y}{\partial x_{22}} = \begin{bmatrix} 0 & a_{12} \\ 0 & a_{22} \end{bmatrix},$$

$$\frac{\partial Y}{\partial x_{31}} = \begin{bmatrix} a_{13} & 0 \\ a_{23} & 0 \end{bmatrix}, \qquad \frac{\partial Y}{\partial x_{32}} = \begin{bmatrix} 0 & a_{13} \\ 0 & a_{23} \end{bmatrix}.$$

These six equations can be combined in the single one

$$(6) \qquad \frac{\partial Y}{\partial x_{mn}} = AJ_{mn}$$

where J_{mn} is a matrix having dimensions of X, with all elements zero except for a unit element in the m-th row and n-th column. Similarly we find

$$\frac{\partial y_{11}}{\partial X} = \begin{bmatrix} a_{11} & 0 \\ a_{12} & 0 \\ a_{13} & 0 \end{bmatrix}, \qquad \frac{\partial y_{12}}{\partial X} = \begin{bmatrix} 0 & a_{11} \\ 0 & a_{12} \\ 0 & a_{13} \end{bmatrix},$$

$$\frac{\partial y_{21}}{\partial X} = \begin{bmatrix} a_{21} & 0 \\ a_{22} & 0 \\ a_{23} & 0 \end{bmatrix}, \qquad \frac{\partial y_{22}}{\partial X} = \begin{bmatrix} 0 & a_{21} \\ 0 & a_{22} \\ 0 & a_{23} \end{bmatrix}.$$

These four equations can be combined in the single one

$$(7) \qquad \frac{\partial y_{pq}}{\partial X} = A'K_{pq},$$

where K_{pq} is the matrix having the dimensions of Y with all elements zero except for a unit element in the p-th row and q-th column.

It should be noted that the matrices on the left of (6) and (7) are matrices composed of the basic elements $\dfrac{\partial y_{pq}}{\partial x_{mn}}$.

Other types of symbolic matrix derivatives could be defined and studied. We have selected these two main types because of their application to regression and correlation theory. The second type is more specifically indicated in the applications but the relations between the types are such that a simultaneous treatment seems appropriate.

2. Notation. Capital letters are used for matrices and small letters for scalars. It is understood that Y, U, V, $\cdots$ are matrices whose elements are functions of the elements x_{mn} of X and that A, B, $\cdots$ (unless otherwise stated) are matrices whose elements are not functions of x_{mn}. In the development of the formulas it is understood that the differentiation is carried out with respect to x_{mn} or X. The matrix function differentiated is called Y.

We have already defined J_{mn} as the matrix having the dimensions of X with all elements zero except for a unit element in the m-th row and the n-th column, and we define K_{pq} similarly with respect to Y. We now define J'_{nm} as the matrix having the dimensions of X' with all elements zero except for a unit element in the n-th row and the m-th column, and we define K'_{qp} similarly with respect to Y'. All the formulas we obtain for $\frac{\partial Y}{\partial x_{mn}}$ involve J_{mn} or J'_{nm} while all those for $\frac{\partial y_{pq}}{\partial X}$ involve K_{pq} or K'_{qp}.

3. Differentiation of a constant. If $Y = A = [a_{pq}]$ we have at once

$$\frac{\partial y_{pq}}{\partial x_{mn}} = 0.$$

It follows that

$$\frac{\partial Y}{\partial x_{mn}} = \frac{\partial}{\partial x_{mn}}[y_{pq}] = 0; \tag{8}$$

$$\frac{\partial y_{pq}}{\partial X} = \left[\frac{\partial}{\partial x_{mn}}\right] y_{pq} = 0, \tag{9}$$

where the zero matrix of (8) has the dimensions of A, while that of (9) has the dimensions of X.

4. Differentiation of a matrix with respect to itself. If $Y = X = [x_{pq}]$ we note that

$$\frac{\partial y_{pq}}{\partial x_{mn}} = \frac{\partial x_{pq}}{\partial x_{mn}} = \begin{Bmatrix} 1 & (p = m, q = n) \\ 0 & (\text{otherwise}) \end{Bmatrix}.$$

It follows that

$$\begin{aligned} \frac{\partial Y}{\partial x_{mn}} &= \frac{\partial}{\partial x_{mn}}[y_{pq}] = J_{mn}, \\ \frac{\partial y_{pq}}{\partial X} &= \left[\frac{\partial}{\partial x_{mn}}\right] y_{pq} = K_{pq}. \end{aligned} \tag{10}$$

5. Differentiation of the transpose of a matrix with respect to the matrix. Let $Y = X'$, so that

$$y_{pq} = x_{qp}.$$

Then

$$\frac{\partial y_{pq}}{\partial x_{mn}} = \frac{\partial x_{qp}}{\partial x_{mn}} = \begin{Bmatrix} 1 & (q = m,\ p = n); \\ 0 & (\text{otherwise}), \end{Bmatrix}$$

and we have

$$\frac{\partial Y}{\partial x_{mn}} = \frac{\partial}{\partial x_{mn}} [y_{pq}] = J'_{nm}, \tag{12}$$

$$\frac{\partial y_{pq}}{\partial X} = \left[\frac{\partial}{\partial x_{mn}}\right] y_{pq} = K'_{qp}, \tag{13}$$

where J'_{nm}, K'_{qp} are defined as in section **2.**

6. Differentiation of sums and differences of matrices. If

$$Y = U + V - W = [u_{pq} + v_{pq} - w_{pq}],$$

we have

$$\frac{\partial y_{pq}}{\partial x_{mn}} = \frac{\partial u_{pq}}{\partial x_{mn}} + \frac{\partial v_{pq}}{\partial x_{mn}} - \frac{\partial w_{pq}}{\partial x_{mn}},$$

then

$$\begin{aligned} \frac{\partial Y}{\partial x_{mn}} &= \frac{\partial}{\partial x_{mn}} [y_{pq}] = \frac{\partial}{\partial x_{mn}} [u_{pq} + v_{pq} - w_{pq}] \\ &= \frac{\partial}{\partial x_{mn}} [u_{pq}] + \frac{\partial}{\partial x_{mn}} [v_{pq}] - \frac{\partial}{\partial x_{mn}} [w_{pq}] \qquad (14) \\ &= \frac{\partial U}{\partial x_{mn}} + \frac{\partial V}{\partial x_{mn}} - \frac{\partial W}{\partial x_{mn}}, \end{aligned}$$

and similarly

$$\frac{\partial y_{pq}}{\partial X} = \frac{\partial u_{pq}}{\partial X} + \frac{\partial v_{pq}}{\partial X} - \frac{\partial w_{pq}}{\partial X}. \tag{15}$$

7. General formulas for the differentiation of a two factor matrix product. Suppose U is a matrix with c rows and d columns and V is a matrix with d rows and e columns, then

$$Y = UV = [y_{pq}] = \sum_{s=1}^{d} u_{ps} v_{sq}. \tag{16}$$

We have at once

$$\frac{\partial y_{pq}}{\partial x_{mn}} = \sum_{s=1}^{d} \frac{\partial u_{ps}}{\partial x_{mn}} v_{sq} + \sum_{s=1}^{d} u_{ps} \frac{\partial v_{sq}}{\partial x_{mn}}. \tag{17}$$

Now considering any fixed x_{mn} it is clear that the first term on the right of (17) is the same as the right hand term of (16) with $\frac{\partial u_{ps}}{\partial x_{mn}}$ in place of u_{ps}. The second term on the right of (17) is likewise the same as the right hand term of (16) with $\frac{\partial v_{sq}}{\partial x_{mn}}$ in place of v_{sq}. We may then write

$$\frac{\partial Y}{\partial x_{mn}} = \frac{\partial U}{\partial x_{mn}} V + U \frac{\partial V}{\partial x_{mn}}. \tag{18}$$

Also considering a fixed y_{pq} we have

$$\frac{\partial y_{pq}}{\partial X} = \sum_{s=1}^{d} \frac{\partial u_{ps}}{\partial X} v_{sq} + \sum_{s=1}^{d} u_{ps} \frac{\partial v_{sq}}{\partial X}. \tag{19}$$

It is to be noted that this formula yields matrices of the proper dimensions (those of X) since $\frac{\partial u_{ps}}{\partial X}$ and $\frac{\partial v_{sq}}{\partial X}$ have the dimensions of X. These matrices, when multiplied by the scalar values v_{sq} and u_{ps} and summed, yield matrices of the desired dimensions.

8. Some properties of matrix products involving J's and K's. Before deriving formulas for the differentiation of products of specific factors, it seems wise to derive some formulas exhibiting certain relations involving the J's and K's. Consider the matrix A having c rows and d columns and the matrix X having d rows and e columns. Then $Y = AX$ is a matrix with c rows and e columns, J_{mn} one with d rows and e columns, J'_{nm} one with e rows and d columns, K_{pq} one with c rows and e columns and K'_{qp} one with e rows and c columns.

It is easily seen by actual multiplication that

(20) *AJ_{mn} is a $c \times e$ matrix with all its elements zero except those of its n-th column which are those of the m-th column of A.* We omit further discussion of the dimensions of the matrices and assume that whenever a matrix product is written, the factors are comformable. Then we can show similarly that

(21) *$J_{mn}B$ is a matrix with all its elements zero except those of its m-th row, which are those of the n-th row of B.* Similar statements hold if J_{mn} is replaced by J'_{nm} or K_{pq} or K'_{qp}. The rules are

(a) When J_{mn} (or J'_{nm} or K_{pq} or K'_{qp}) is the postmultiplier, the *first* subscript indicates the *column* of the other matrix which is placed in the column indicated by the *second* subscript.

(b) When J_{mn} (or J'_{nm} or K_{pq} or K'_{qp}) is the premultiplier, the *second* subscript indicates the *row* of the other matrix which is placed in the row indicated by the *first* subscript.

Notice also that

(22) *$A'K_{pq}$ is a matrix with all elements zero except those of its q-th column, which are those of the p-th column of A', or the p-th row of A.* A similar result holds if K_{pq} is replaced by K'_{qp} or J_{mn} or J'_{nm}.

9. Differentiation of specific two factor products. Let us start with $Y = AX$ where the various matrices involved have the dimensions indicated in the last section. Application of (18), (8), (10) gives

$$\frac{\partial Y}{\partial x_{mn}} = \frac{\partial A}{\partial x_{mn}} X + A \frac{\partial X}{\partial x_{mn}} = 0 + AJ_{mn} = AJ_{mn}, \tag{23}$$

while application of (19), (11) yields

$$\frac{\partial y_{pq}}{\partial X} = \sum_{s=1}^{d} \frac{\partial a_{ps}}{\partial X} x_{sq} + \sum_{s=1}^{d} a_{ps} \frac{\partial x_{sq}}{\partial X} \tag{24}$$

$$= \sum_{s=1}^{d} a_{ps} K_{sq}$$

$$= a_{p1}K_{1q} + a_{p2}K_{2q} + \cdots + a_{pd}K_{dq}$$

= a $c \times e$ matrix with all elements zero except those of its q-th column which are those of the p-th row of A

$= A'K_{pq}$ by (22).

Similar treatment of $Y = XB$ yields

$$\frac{\partial Y}{\partial x_{mn}} = \frac{\partial X}{\partial x_{mn}} B + X \frac{\partial B}{\partial x_{mn}} = J_{mn} B, \tag{25}$$

$$\frac{\partial y_{pq}}{\partial X} = \sum_{s} \frac{\partial x_{ps}}{\partial X} b_{sq} = \sum_{s} K_{ps} b_{sq} = K_{pq} B'. \tag{26}$$

If we treat $Y = AX'$ in a similar fashion, we get

$$\frac{\partial Y}{\partial x_{mn}} = AJ'_{nm}, \tag{27}$$

$$\frac{\partial y_{pq}}{\partial X} = K'_{qp} A, \tag{28}$$

while $Y = X'B$ yields

$$\frac{\partial Y}{\partial x_{mn}} = J'_{nm} B, \tag{29}$$

$$\frac{\partial y_{pq}}{\partial X} = BK'_{qp}. \tag{30}$$

It is to be noted that J always has the subscripts mn, and similarly we find always J'_{nm}, K_{pq}, K'_{qp}. We may therefore omit the subscripts on these letters. When we do so we shall also write

$$\frac{\partial Y}{\partial \langle X \rangle} \text{ for } \frac{\partial X}{\partial x_{mn}}, \quad \frac{\partial \langle Y \rangle}{\partial X} \text{ for } \frac{\partial y_{pq}}{\partial X},$$

placing brackets $\langle\ \rangle$ around the matrix from which a fixed element is to be chosen. Thus if $Y = AX$, we write instead of (23) and (24)

$$\frac{\partial Y}{\partial \langle X \rangle} = AJ; \tag{23a}$$

$$\frac{\partial \langle Y \rangle}{\partial X} = A'K. \tag{24a}$$

The other results are summarized in lines 1–5 of Table I.

Examination of (18) and (19) shows that the derivatives of products with two variable factors are obtained by adding the results obtained by holding each factor constant while differentiating the other. With this in mind, (23)–(30) can be used to obtain the derivatives of double products involving X and X'. Thus if $Y = XX$, we get

$$(31) \qquad \frac{\partial Y}{\partial\langle X\rangle} = JX + XJ, \qquad \frac{\partial\langle Y\rangle}{\partial X} = KX' + X'K.$$

Other double product formulas involving X and X' are given in Table I.

TABLE I

Formula	Y	$\frac{\partial Y}{\partial\langle X\rangle}$	$\frac{\partial\langle Y\rangle}{\partial X}$
1	AB	0	0
2	AX	AJ	$A'K$
3	XB	JB	KB'
4	AX'	AJ'	$K'A$
5	$X'B$	$J'B$	BK'
6	XX	$JX + XJ$	$KX' + X'K$
7	$X'X$	$J'X + X'J$	$XK' + XK$
8	XX'	$JX' + XJ'$	$KX + K'X$
9	$X'X'$	$J'X' + X'J'$	$X'K' + K'X'$

The formulas for $\frac{\partial Y}{\partial\langle X\rangle}$ are written down very easily, but those for $\frac{\partial\langle Y\rangle}{\partial X}$ are not so easy to write. However the values of $\frac{\partial Y}{\partial\langle X\rangle}$ and $\frac{\partial\langle Y\rangle}{\partial X}$ in formulas 2–5 of Table I are such that the results for $\frac{\partial\langle Y\rangle}{\partial X}$ may be obtained from those for $\frac{\partial Y}{\partial\langle X\rangle}$ with the use of a few simple rules. They are

(a) Each J becomes K and each J' becomes K'.

(b) The pre (or post) multiplier of J becomes its transpose.

(c) The pre (or post) multiplier of J' becomes a post (or pre) multiplier of K'.

These rules are immediately applicable to the double products. Thus when $Y = X'X$ we have

$$\frac{\partial Y}{\partial\langle X\rangle} = J'X + X'J,$$

and so

$$\frac{\partial\langle Y\rangle}{\partial X} = XK' + XK.$$

10. Differentiation of three (or more) factor products. Products with three factors can be differentiated by the formulas of the last section if two adjacent factors are constant. Thus if $Y = ABX$, we have

$$\frac{\partial Y}{\partial\langle X\rangle} = ABJ, \qquad \frac{\partial\langle Y\rangle}{\partial X} = B'A'K.$$

It is not yet demonstrated that these rules are applicable to the products AXB and $AX'B$. However it can be shown by the general methods indicated earlier that if $Y = AXB$, we obtain

$$\frac{\partial Y}{\partial\langle X\rangle} = AJB, \qquad \frac{\partial\langle Y\rangle}{\partial X} = A'KB', \tag{33}$$

while if $Y = AX'B$ we have

$$\frac{\partial Y}{\partial\langle X\rangle} = AJ'B, \qquad \frac{\partial\langle Y\rangle}{\partial X} = BK'A. \tag{34}$$

It is now apparent that the rules of the last section apply to situations in which there are both pre and post multipliers.

The general theory for two-factor products is immediately extendable. Thus if $Y = UVW$ with $y_{pq} = \sum_s \sum_r u_{ps}v_{sr}w_{rq}$ then the basic element is

$$\frac{\partial y_{pq}}{\partial x_{mn}} = \sum_s \sum_r \frac{\partial u_{ps}}{\partial x_{mn}} v_{sr} w_{rq} + \sum_s \sum_r u_{ps} \frac{\partial v_{sr}}{\partial x_{mn}} w_{rq} + \sum_s \sum_r u_{ps} v_{sr} \frac{\partial w_{rq}}{\partial x_{mn}}, \tag{35}$$

and the formulas result from treating each factor in turn as the only variable. For example if $Y = XX'X$, we have

$$\frac{\partial Y}{\partial\langle X\rangle} = JX'X + XJ'X + XX'J, \tag{36}$$

and

$$\begin{aligned}\frac{\partial\langle Y\rangle}{\partial X} &= K(X'X)' + XK'X + (XX')'K \\ &= KX'X + XK'X + XX'K.\end{aligned} \tag{37}$$

The symbolic derivatives of certain triple product matrices are presented in Table II.

The rules are sufficiently general to take care of matrices with more than three factors. Thus if $Y = A'X'XB$, we have

$$\frac{\partial Y}{\partial\langle X\rangle} = A'J'XB + A'X'JB \tag{38}$$

and

$$\frac{\partial\langle Y\rangle}{\partial X} = XBK'A' + XAKB', \tag{39}$$

and in the special case $B = A$, we get

$$\frac{\partial Y}{\partial\langle X\rangle} = A'(J'X + X'J)A, \tag{40}$$

$$\frac{\partial\langle Y\rangle}{\partial X} = XA(K' + K)A'. \tag{41}$$

Similarly if $Y = X'A'AX$, we get

$$\frac{\partial Y}{\partial \langle X \rangle} = J'A'AX + X'A'AJ, \tag{42}$$

and

$$\frac{\partial \langle Y \rangle}{\partial X} = A'AXK' + A'AXK. \tag{43}$$

TABLE II

Formula	Y	$\frac{\partial Y}{\partial \langle X \rangle}$	$\frac{\partial \langle Y \rangle}{\partial X}$
1	ABC	0	0
2	ABX	ABJ	$B'A'K$
3	AXC	AJC	$A'KC'$
4	XBC	JBC	$KC'B'$
5	ABX'	ABJ'	$K'AB$
6	$AX'C$	$AJ'C$	$CK'A$
7	$X'BC$	$J'BC$	BCK'
8	AXX	$AJX + AXJ$	$A'KX' + X'A'K$
9	XBX	$JBX + XBJ$	$KX'B' + B'X'K$
10	XXC	$JXC + XJC$	$KC'X' + X'KC'$
11	$AX'X'$	$AJ'X' + AX'J'$	$X'K'A + K'AX'$
12	$X'BX'$	$J'BX' + X'BJ'$	$BX'K' + K'X'B$
13	$X'X'C$	$J'X'C + X'J'C$	$X'CK' + CK'X'$
14	$AX'X$	$AJ'X + AX'J$	$XK'A + XA'K$
15	$X'BX$	$J'BX + X'BJ$	$BXK' + B'XK$
16	$X'XC$	$J'XC + X'JC$	$XCK' + XKC'$
17	AXX'	$AJX' + AXJ'$	$A'KX + K'AX$
18	XBX'	$JBX' + XBJ'$	$KXB' + K'XB$
19	$XX'C$	$JX'C + XJ'C$	$KC'X + CK'X$
20	XXX	$JXX + XJX + XXJ$	$KX'X' + X'KX' + X'X'K$
21	XXX'	$JXX' + XJX' + XXJ'$	$KXX' + X'KX + K'XX$
22	$XX'X$	$JX'X + XJ'X + XX'J$	$KX'X + XK'X + XX'K$
23	$X'XX$	$J'XX + X'JX + X'XJ$	$XXK' + XKX' + X'XK$
24	$XX'X'$	$JX'X' + XJ'X' + XX'J'$	$KXX + X'K'X + K'XX'$
25	$X'XX'$	$J'XX' + X'JX' + X'XJ'$	$XX'K' + XKX + K'X'X$
26	$X'X'X$	$J'X'X + X'J'X + X'X'J$	$X'XK' + XK'X' + XXK$
27	$X'X'X'$	$J'X'X' + X'J'X' + X'X'J'$	$X'X'K' + X'K'X' + K'X'X'$

Finally if $Y = XAX'AX$, we get

$$\frac{\partial Y}{\partial \langle X \rangle} = JAX'AX + XAJ'AX + XAX'AJ, \tag{44}$$

$$\frac{\partial Y}{\partial \langle X \rangle} = KX'A'XA' + AXK'XA + A'XA'X'K. \tag{45}$$

11. Vector results. It should be emphasized that each of the above results is a general result. More specific results may be obtained in case one (or more) of the matrices is a vector. For example if X_c is a column matrix and $Y = X_c'BX_c$, then Y is a scalar, so K and K' are both unity and we have from Table II (15)

$$\frac{\partial\langle Y\rangle}{\partial X} = BX_c + B'X_c = (B + B')X_c. \tag{46}$$

If in addition B is symmetric, $B' = B$ and we have

$$\frac{\partial\langle Y\rangle}{\partial X} = 2BX_c,$$

which is the result indicated in (3).

12. Differentiation of the inverse of X. It is possible to use implicit differentiation to derive formulas for $\frac{\partial X^{-1}}{\partial\langle X\rangle}$ and $\frac{\partial\langle X^{-1}\rangle}{\partial X}$. We write $I = XX^{-1}$ and get

$$\frac{\partial I}{\partial\langle X\rangle} = 0 = JX^{-1} + X\frac{\partial X^{-1}}{\partial\langle X\rangle},$$

so that

$$\frac{\partial X^{-1}}{\partial\langle X\rangle} = -X^{-1}JX^{-1}, \tag{47}$$

whence

$$\frac{\partial\langle X^{-1}\rangle}{\partial X} = -(X^{-1})'K(X^{-1})'. \tag{48}$$

The formula (47) is a generalization of a known matrix differential formula [3:3.4].

In a similar way we derive

$$\frac{\partial(X')^{-1}}{\partial\langle X\rangle} = -(X')^{-1}J'(X')^{-1}, \tag{49}$$

$$\frac{\partial\langle(X')^{-1}\rangle}{\partial X} = -(X')^{-1}K'(X')^{-1}. \tag{50}$$

13. Differentiation of a function of a function. The theory developed in the earlier sections is sufficiently general to be useful in differentiating a function of a function if the functions involve addition, subtraction, premultiplication, postmultiplication, and inverse. For example if

$$Y = Z'Z \quad \text{with} \quad Z = AX \tag{51}$$

we have

$$\frac{\partial Y}{\langle\partial X\rangle} = \frac{\partial Z'}{\partial\langle X\rangle}Z + Z'\frac{\partial Z}{\partial\langle X\rangle},$$

and since

$$\frac{\partial Z'}{\partial \langle X \rangle} = J'A' \quad \text{and} \quad \frac{\partial Z}{\partial \langle X \rangle} = AJ,$$

(52)
$$\frac{\partial \langle Y \rangle}{\partial X} = J'A'Z + Z'AJ,$$

and thence

(53)
$$\frac{\partial \langle Y \rangle}{\partial X} = A'ZK' + A'ZK.$$

These results are equivalent to those of (42) and (43).

14. Differentiation of a power of a square matrix. The values of the symbolic derivatives of X^2, X^3 with respect to X are given in Tables I and II. It can be shown similarly that if n is a positive integer

(54)
$$\frac{\partial X^n}{\partial \langle X \rangle} = JX^{n-1} + \sum_{s=1}^{n-2} X^s J X^{n-s-1} + X^{n-1}J,$$

and this can be written as

(55)
$$\frac{\partial X^n}{\partial \langle X \rangle} = \sum_{s=0}^{n-1} X^s J X^{n-s-1},$$

if we adopt the convention that X^0 is I. It follows at once that

(56)
$$\frac{\partial \langle X^n \rangle}{\partial X} = \sum_{s=0}^{n-1} X'^s K (X')^{n-s-1}.$$

It is thence possible to derive formulas for the symbolic derivatives of X^{-n}. Since $X^{-n}X^n = \text{I}$, we have

(57)
$$\frac{\partial X^{-n}}{\partial \langle X \rangle} X^n + X^{-n} \left[\sum_{s=0}^{n-1} X^s J X^{n-s-1} \right] = 0,$$

so

(58)
$$\frac{\partial X^{-n}}{\partial \langle X \rangle} = -X^{-n} \left[\sum_{s=0}^{n-1} X^s J X^{n-s-1} \right] X^{-n},$$

and

(59)
$$\frac{\partial \langle X^{-n} \rangle}{\partial X} = -X^{-n} \left[\sum_{s=0}^{n-1} (X')^s K (X')^{n-s-1} \right] X^{-n}.$$

15. Applications. We consider the classical theory of least squares, a matrix presentation of which is available in [2]. Suppose that y and x_i are measured from their means and that y is to be estimated from the n variables x_i. Form the values of y into a column matrix Y and the values of x_i into an N by n matrix X. Introduce the column matrix B of n parameters b_i and define

(60)
$$E = Y - XB.$$

Note that the matrix $E'E$ is in this case the single element matrix which is the sum of the squares of the residuals. Following the least squares method we minimize this by differentiating with respect to the elements of B. We first note that

$$E'E = (Y' - B'X')(Y - XB) \tag{61}$$
$$= Y'Y - Y'XB - B'X'Y + B'X'XB$$

Then we write down first

$$\frac{\partial(E'E)}{\partial\langle B\rangle} = -Y'XJ - J'X'Y + J'X'XB + B'X'XJ, \tag{62}$$

from which we get

$$\frac{\partial\langle E'E\rangle}{\partial B} = -X'YK - X'YK' + X'XBK' + X'XBK \tag{63}$$
$$= -X'(Y - XB)(K + K') = -X'E(K + K').$$

The J's and K's are associated with B and $E'E$ respectively. Here $E'E$ is scalar so that $K = K' = 1$ and we have

$$\frac{\partial\langle E'E\rangle}{\partial B} = -2X'E. \tag{64}$$

The equation $X'E = 0$, obtained by equating the right hand side of (64) to zero, is a statement of the normal equations in matrix form.

Equation (64) may also be obtained with the use of the methods of section 13. In this case

$$\frac{\partial E}{\partial\langle B\rangle} = -XJ, \qquad \frac{\partial E'}{\partial\langle B\rangle} = -J'X',$$

and we have

$$\frac{\partial(E'E)}{\partial\langle B\rangle} = \frac{\partial E'}{\partial\langle B\rangle} E + E' \frac{\partial E}{\partial\langle B\rangle} = -J'X'E - E'XJ; \tag{65}$$

so

$$\frac{\partial\langle E'E\rangle}{\partial B} = -X'EK' - X'EK = -X'E(K' + K). \tag{66}$$

The equation (64) is also applicable to the more general problem in which y_1 and y_2 are estimated from the same set of variables x_i. The only change needed is to regard Y, B, E as two-column matrices so that $E'E$ is a matrix with two rows and columns which we denote by

$$\begin{bmatrix} \epsilon_{11} & \epsilon_{12} \\ \epsilon_{21} & \epsilon_{22} \end{bmatrix}.$$

We require $\frac{\partial \epsilon_{11}}{\partial B} = 0$ and $\frac{\partial \epsilon_{22}}{\partial B} = 0$. From equation (63), inserting subscripts, we get

$$\frac{\partial \epsilon_{11}}{\partial B} = -X'E(K_{11} + K'_{11})$$
$$= -2X'EK_{11};$$
$$\frac{\partial \epsilon_{22}}{\partial B} = -2X'EK_{22}.$$

It is easily seen that $\frac{\partial \epsilon_{11}}{\partial B} = \frac{\partial \epsilon_{22}}{\partial B} = 0$ is equivalent to $X'E = 0$, the same equation as we obtained in the last paragraph. We also arrive at the incidental result that in minimizing $\Sigma \epsilon_1^2$, and $\Sigma \epsilon_2^2$ separately we find at the same time a stationary value of $\Sigma \epsilon_1 \epsilon_2$.

In this way we can treat two or more simultaneous regression problems with this general notation as easily as we can treat one.

As a second application of the theory we outline the initial steps in the direction of the formulas for canonical correlation [4], [5]. In this case A and B are unknown column vectors with X and Y known rectangular matrices. Then XA is a column matrix:

$$XA = L = \begin{bmatrix} l_1 \\ l_2 \\ \cdot \\ \cdot \\ \cdot \\ l_N \end{bmatrix},$$

whose elements l_i may be regarded as observed values of a linear form l. Similarly $YB = \Lambda$, a column matrix whose elements may be replaced as observed values of a linear form λ. It is desired to find A and B such that l and λ may have the largest correlation coefficient, and to find the size of this coefficient. Then $A'X'XA$, $B'Y'YB$, and $B'Y'XA = A'X'YB$ are scalars, and

$$\rho = \frac{B'Y'XA}{\sqrt{(A'X'XA)(B'Y'YB)}}. \tag{67}$$

If the scales of X and Y are chosen so that $A'X'XA = 1$ and $B'Y'YB = 1$, we have

$$\rho = B'Y'XA = A'X'YB. \tag{68}$$

Using Lagrange multipliers we set

$$\phi = B'Y'XA + \frac{c}{2}(1 - A'X'XA) + \frac{d}{2}(1 - B'Y'YB), \tag{69}$$

and differentiate with respect to the elements of A and B. We first differentiate ϕ with respect to A after replacing $B'Y'XA$ by $A'X'YB$:

(70) $$\frac{\partial\phi}{\partial\langle A\rangle} = J'X'YB - \frac{c}{2}(J'X'XA + A'X'XJ);$$

(71) $$\frac{\partial\langle\phi\rangle}{\partial A} = X'YBK' - \frac{c}{2}(X'XAK' + X'XAK).$$

(The J's and K's are associated with A and ϕ respectively). We set $\frac{\partial\langle\phi\rangle}{\partial A} = \mathbf{0}$ with $K = K' = 1$ to get

(72) $$X'YB = cX'XA,$$

whence by (57)

(73) $$\rho = A'X'YB = cA'X'XA = c,$$

and

(74) $$X'YB = \rho X'XA.$$

Similar differentiation with respect to B gives $\rho = d$ and

(75) $$Y'XA = \rho Y'YB.$$

The further steps in the development of canonical correlation theory are based on (74) and (75).

A third application is to orthogonal regression. The situation is very similar to that of the first illustration, but the errors are measured orthogonal to the plane of best fit. As before we take the variates as measured from their means and so have the basic equation

(76) $$D = \frac{b_1x_1 + b_2x_2 + \cdots + b_kx_k}{\sqrt{b_1^2 + b_2^2 + \cdots + b_k^2}}.$$

This can be written as

(77) $$D = l_1x_1 + l_2x_2 + \cdots + l_kx_k = XL \text{ with } L'L = 1.$$

It follows that the quantity to be minimized is

(78) $$D'D = L'X'XL.$$

With the use of Lagrange multipliers we have

(79) $$\Phi = L'X'XL + \lambda(1 - L'L)$$

so that

(80) $$\frac{\partial\phi}{\partial\langle L\rangle} = J'X'XL + L'X'XJ - \lambda(J'L + L'J),$$

(81) $$\frac{\partial\langle\phi\rangle}{\partial L} = X'XLK' + X'XLK - \lambda(LK' + LK)$$

from which

(82) $$2X'XL - 2\lambda L = 0$$

and the values can be determined from the equation

$$(X'X - \lambda)L = 0. \tag{83}$$

The solution continues with the use of the characteristic equation.

It is to be noted from (79) and (82) that

$$D'D = L'X'XL = \lambda L'L = \lambda$$

so that (83) becomes

$$(X'X - D'D)L = 0. \tag{84}$$

A fourth illustration uses symbolic derivatives in obtaining the principal components of a total variance [5,252]. The variable portion of the exponent of the multivariate normal can be written $Y'AY$ where Y is the column vector $[y_1, \cdots, y_k]$ and A is a k by k matrix. We set this equal to a constant, say C, and get the equation of the k dimensional ellipsoid. It is desired to locate the extrema of this ellipsoid. To do this we find the extrema of $Y'Y$. Using the Lagrange multiplier we have

$$\phi = Y'Y + \lambda(C - Y'AY) \tag{85}$$

so that

$$\frac{\partial \phi}{\partial \langle Y \rangle} = J'Y + Y'J - \lambda(J'AY + Y'AJ), \tag{86}$$

$$\frac{\partial \langle \phi \rangle}{\partial Y} = YK' + YK - \lambda(AYK' + AYK), \tag{87}$$

so that there results

$$Y - \lambda AY = 0. \tag{88}$$

Pre-multiplying by A^{-1} we get

$$(A^{-1} - \lambda)Y = 0 \tag{89}$$

and pre-multiplying by Y' gives the important relation

$$Y'Y = \lambda C. \tag{90}$$

A fifth illustration utilizes symbolic differentiation in developing the theory of the linear discriminant function [6, 341] [8, 124]. As in the other illustrations, the variates are measured about their means. The unknown multipliers are indicated by the vector L. Then

$$Z = XL \tag{91}$$

is the general matrix equation while

$$\begin{aligned} Z_1 &= X_1L \\ Z_2 &= X_2L \end{aligned} \tag{92}$$

are the corresponding equations for the two groups. Then

$$\bar{Z}_1 = \bar{X}_1 L,\ \bar{Z}_2 = \bar{X}_2 L, \text{ and } \bar{Z}_1 - \bar{Z}_2 = (\bar{X}_1 - \bar{X}_2)L = DL, \tag{93}$$

$$\begin{aligned} Z_1 - \bar{Z}_1 &= (X_1 - \bar{X}_1)L = Y_1 L, \\ Z_2 - \bar{Z}_2 &= (X_2 - \bar{X}_2)L = Y_2 L. \end{aligned} \tag{94}$$

The within group variation, $L'Y_1'Y_1L + L'Y_2'Y_2L$, is then divided into the between group variation, $L'D'DL$, to get

$$G = \frac{L'D'DL}{L'Y_1'Y_1'L + L'Y_2'Y_2L} = \frac{A}{B}. \tag{95}$$

We wish to maximize G. Since A and B are scalars $\frac{\partial\langle G\rangle}{\partial L} = 0$ reduces to

$$\frac{\partial\langle B\rangle}{\partial L} = \frac{1}{G}\frac{\partial\langle A\rangle}{\partial L} \tag{96}$$

which becomes, with further differentiation

$$(Y_1'Y_1 + Y_2'Y_2)L = D'\left(\frac{DL}{G}\right). \tag{97}$$

Since $\frac{DL}{G}$ is a scalar, we have

$$(Y_1'Y_1 + Y_2'Y_2)L = cD. \tag{98}$$

Any convenient value of c can be used for purposes of discrimination. It is customary to take $c = 1$ and then to adjust (98) so that some l_i is unity.

A final illustration applies symbolic matrix differentiation to a theorem of multiple factor analysis. This presentation parallels that given by Thurstone [7,473–477] for transforming any factorial matrix into a principal axes matrix. The matrix

$$F = [a_{ij}] \tag{99}$$

has p rows and r columns, $r \leq p$, such that

$$FF' = R \tag{100}$$

where R is a $p \times p$ correlation matrix.

It is desired to apply the unitary orthogonal transformation L to F in such a way as to produce a matrix, called F_p, which has the sums of the squares in respective columns a maximum. This can be done by maximizing simultaneously the diagonal terms of $F_p'F_p$ where

$$F_p = FL. \tag{101}$$

Again using Lagrange multipliers, we have

$$\phi = L'F'FL + \lambda(I - L'L). \tag{102}$$

This equation has the same analytical form as (79). Differentiation leads to the result

$$(F'F - \lambda)L = 0. \tag{103}$$

The solution of (103) gives the value L which can be substituted in (101) to obtain F_p.

14. Conclusion. Two types of symbolic matrix derivatives have been defined. Laws have been developed for the basic operations of addition, subtraction, multiplication, inverse, and powers. Laws for more extended functions can be worked out on the basis of principles enunciated.

Applications are given to certain multivariate problems. It is our thesis that with these differentiation formulas available, much work in multivariate analysis can be carried on with a simple matrix notation.

REFERENCES

[1] R. A. Frazer, W. J. Duncan, and A. R. Collar, *Elementary Matrices and Some Applications to Dynamics and Differential Equations*, Cambridge University Press, 1936.

[2] P. S. Dwyer, "A matrix presentation of least squares and correlation theory," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 82–89.

[3] A. D. Michal, *Matrix and Tensor Calculus*, New York, John Wiley and Sons, 1947.

[4] H. Hotelling, "Relations between two sets of variates," *Biometrika*, Vol. 28 (1936), pp. 321–377.

[5] S. S. Wilks, *Mathematical Statistics*, Princeton University Press, 1946.

[6] M. G. Kendall, *Advanced Theory of Statistics*, Vol. II, London, Charles Griffin and Co. Ltd., 1946.

[7] L. L. Thurstone, *Multiple Factor Analysis*, Chicago, University of Chicago Press, 1947.

[8] P. G. Hoel, *Introduction to Mathematical Statistics*, New York, John Wiley and Sons, 1947.