The Annals of Statistics
1974, Vol. 2, No. 2, 337–346

MATRIX DERIVATIVES WITH AN APPLICATION TO AN ADAPTIVE LINEAR DECISION PROBLEM

BY ELIZABETH CHASE MACRAE

University of Maryland

A theory of matrix differentiation is presented which uses the concept of a matrix of derivative operators. This theory allows matrix techniques to be used in both the derivation and the description of results. Several new operations and identities are presented which facilitate the process of matrix differentiation. The derivative theorems and new operations are then applied to the problem of determining optimal policies in a linear decision model with unknown coefficients, a problem which would be cumbersome if not impossible to solve without these theorems and operations.

1. Introduction. The problem of differentiating the elements of one matrix with respect to the elements of another matrix is an important one in multivariate analysis. Ideally, matrix differentiation should retain the advantages of matrix notation by producing an array of derivatives which can be expressed in terms of the original matrices. In addition, there should be some basic underlying concept which permits the systematic derivation of derivatives for general matrix expressions without recourse to *ad hoc* rules.

Previous work on matrix derivatives has produced several formulas which describe an array of matrix derivatives in terms of original arrays. The derivation of these formulas has generally followed one of two procedures: either a typical element of the matrix array is examined in hopes of inferring a matrix expression for the entire array [1]; or matrices of total differentials are calculated and the resulting expressions then transformed into arrays of derivatives, using special theorems [4].

Both these approaches are somewhat unsatisfactory because there is no explicit underlying concept which suggests what to do if existing rules and theorems are inapplicable to a particular expression. Furthermore, because of inadequate notation for dealing with special arrangements of matrix elements, arrays of matrix derivatives are often described not in terms of the matrices involved, but in terms of rows or columns or even single elements.

The purpose of this paper is to present a consistent general approach to the problem of matrix differentiation, and to introduce several new matrix operations and identities to handle the particular notational problems that matrix differentiation creates. The basic notation for matrix derivatives is described in Section 2 in terms of a matrix of partial derivative operators. In addition, several special matrix operations are presented along with identities, some of which are new,

Received October 1971; revised May 1973.
AMS 1970 *subject classifications.* Primary 26A60, 47F05; Secondary 93E10, 93E20.
Key words and phrases. Matrix differentiation, derivative operators, adaptive decision problem.

describing their properties. Section 3 lists the most useful of the theorems regarding matrix differentiation and sketches proofs which follow from the basic notation discussed in Section 2. In Section 4, the theorems and notation developed in earlier sections are applied to an adaptive decision problem which has not been solved prior to the introduction of the theorems and identities presented in this paper.

2. Basic notation.

2.1. *Matrix differentiation.* Let X be an m by n matrix, and let Y be a p by q matrix whose elements are functions of the elements of X. Let d/dX be a matrix of derivative operators, $[\partial/\partial x_{ij}]$. Then the derivative of matrix Y with respect to X is defined to be an mp by nq matrix of partial derivatives, dY/dX, given by

$$dY/dX = Y \otimes d/dX\,, \tag{2.1}$$

where $\otimes$ indicates a Kronecker product. Multiplication of a matrix element by a derivative operator corresponds to the operation of differentiation. Although Vetter [7] and Neudecker [4] arrange matrix derivatives in a pattern which corresponds to the reversed Kronecker product, $d/dX \otimes Y$, they do not employ the concept of derivative operator, and consequently must obtain their results through *ad hoc* examination of individual elements.

2.2. *Pack operator.* The pack of an m by n matrix X, $\mathcal{P}(X)$, is defined to be an mn-dimensional column vector formed by packing the columns of X below one another, i.e., the ijth element of X becomes the $(i + (j - 1)m)$th element of $\mathcal{P}(X)$. The usefulness of the pack operator is enhanced by the following identities. Let X, Y and Z be conformable to the indicated multiplication. Then as shown by Nissen [5],

$$\mathcal{P}(XYZ) = (Z' \otimes X)\mathcal{P}(Y)\,. \tag{2.2}$$

In addition, when W is r by s, Z is s by t, X is m by n, and Y is n by p, then

$$WZ \otimes XY = [I_{rm} \otimes \mathcal{P}'(Y')][I_r \otimes \mathcal{P}(X')\mathcal{P}'(Z') \otimes I_p][\mathcal{P}(W') \otimes I_{pt}]\,, \tag{2.3}$$

where the subscripts on I indicate the dimension of the identity matrix. The most useful special cases of (2.3) occur when one or more of the matrices W, Z, X, Y are identity matrices, or when either WZ or XY is a scalar.

2.3. *Permuted identity matrix.* Define $I_{(m,n)}$ to be a square mn-dimensional matrix partitioned into m by n submatrices such that the ijth submatrix has a 1 in its jith position and zeroes elsewhere. E.g.,

$$I_{(2,2)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.4}$$

This matrix is identical to the matrix $I_{(n)}$ defined by Tracy and Dwyer [6]. The new notation, however, has the advantage of indicating the total dimension of the matrix. The following identities can be verified by direct examination.

$$I_{(m,1)} = I_{(1,m)} = I_m , \tag{2.5}$$

$$I_{(m,n)} = I'_{(n,m)} , \tag{2.6}$$

$$I_{(m,n)} I_{(n,m)} = I_{mn} . \tag{2.7}$$

In addition, the permuted identity matrix may be used to reverse the order of a Kronecker product:

$$B \otimes A = I_{(m,p)}(A \otimes B) I_{(q,n)} , \tag{2.8}$$

where A is m by n and B is p by q. These identities are particularly useful in the calculation of matrix derivatives. One further identity, analogous to (2.3) is:

$$\begin{aligned}(WZ \otimes XY) I_{(p,t)} \\ = [I_{rm} \otimes \mathscr{V}'(Z')][I_r \otimes (X \otimes I_s) I_{(s,n)} (I_s \otimes Y) \otimes I_t][\mathscr{V}(W') \otimes I_{pt}] ,\end{aligned} \tag{2.9}$$

where the dimensions of W, X, Y and Z are the same as for identity (2.3).

2.4. *Star product*. Many problems involving matrices call for special operations which cannot be readily described in terms of standard matrix operations. The star product is a new operation introduced here to fill the gap at least partially.

Let A be an m by n matrix, and let B be a mp by nq matrix. Then the star product of A and B is a p by q matrix C defined by

$$C = A * B = \textstyle\sum_{ij} a_{ij} B_{ij} , \tag{2.10}$$

where a_{ij} is the ijth element of A, and B_{ij} is the ijth submatrix of B when B is partitioned into submatrices of dimension p by q. If A and B have the same dimension, the star product is equivalent to the trace of $A'B$:

$$A * B = \operatorname{tr}\{A'B\} , \qquad (A, B \text{ same dimension}) . \tag{2.11}$$

The following identities can be established by direct computation, assuming the matrices are conformable to the indicated operations:

$$(X * Y) \otimes Z = X * (Y \otimes Z) , \tag{2.12}$$

$$ABC = B * \mathscr{V}(A) \mathscr{V}'(C') , \tag{2.13}$$

$$ABC = B' * (C \otimes I_m) I_{(m,n)} (A \otimes I_n) , \tag{2.14}$$

and

$$A * D = D * (A \otimes \mathscr{V}(I_s) \mathscr{V}'(I_t)) , \tag{2.15}$$

where A, B, C and D are m by p, p by q, q by n and sm by tp matrices respectively.

Two additional identities deal with the star product in the context of quadratic expressions. If A and B are m by n and mp by nq matrices, and x and y are p and

q dimensional vectors, then

$$A * (I_m \otimes x')B(I_n \otimes y) = x'(A * B)y = (I_m \otimes x)A(I_n \otimes y') * B \,. \tag{2.16}$$

Furthermore, if the vectors x and y and each submatrix of B are partitioned as follows,

$$x = \begin{bmatrix} u \\ v \end{bmatrix}, \quad y = \begin{bmatrix} w \\ z \end{bmatrix}, \quad B_{ij} = \begin{bmatrix} B_{ij}^{uw} & B_{ij}^{uz} \\ B_{ij}^{vw} & B_{ij}^{vz} \end{bmatrix} \tag{2.17}$$

then

$$x'(A * B)y = u'(A * B^{uw})w + u'(A * B^{uz})z + v'(A * B^{vw})w + v'(A * B^{vz})z \,, \tag{2.18}$$

where B^{uw}, etc. are matrices formed of all the noncontiguous submatrices B_{ij}^{uw} for all i, j.

3. General derivative theorems. In this section, theorems are derived for matrix differentiation of simple expressions involving the usual matrix operations and the new operations presented in Section 2. By repeated application of these rules, some of which are new, derivatives may be calculated for compound expressions.

THEOREM 1. *Let X be an m by n matrix. Then*

$$dX/dX = \mathscr{P}(I_m)\mathscr{P}'(I_n) \tag{3.1}$$

and

$$dX'/dX = I_{(m,n)} \,. \tag{3.2}$$

These two results may be verified by direct application of the definition of a matrix derivative. An alternative statement of Theorem 1 with different notation is given by Vetter [7].

The following theorem shows how a compound matrix expression can be differentiated piece by piece.

THEOREM 2. (Decomposition Rule.) *Let $W = F(Y, Z)$ be a matrix function of matrices Y and Z which are in turn functions of a matrix X. Then*

$$dW/dX = dF/dX|_{Z \text{ constant}} + dF/dX|_{Y \text{ constant}} \,. \tag{3.3}$$

This is simply a multidimensional analog of the scalar chain rule, where $w = f(y, z)$ and,

$$\frac{dw}{dx} = \frac{\partial f}{\partial y} \cdot \frac{dy}{dx} + \frac{\partial f}{\partial z} \cdot \frac{dz}{dx} = \left.\frac{df}{dx}\right|_{z \text{ constant}} + \left.\frac{df}{dx}\right|_{y \text{ constant}} \,. \tag{3.4}$$

Note, however, that unlike the scalar case, the derivative dF/dX with, for example, Z constant, cannot in general be written as the product of dF/dY and dY/dX.

THEOREM 3. (Sum Rule.) *Let Y and Z be matrix functions of X, and let their* *sum be defined. Then*

$$d(Y + Z)/dX = dY/dX + dZ/dX . \tag{3.5}$$

THEOREM 4. (Product Rule.) *Let Y and Z be matrix functions of an m by n matrix X, and let the product YZ be defined. Then*

$$d(YZ)/dX = (dY/dX)(Z \otimes I_n) + (Y \otimes I_m)(dZ/dX) . \tag{3.6}$$

By the Decomposition Rule, each term on the right-hand side of (3.6) is the derivative of the product YZ when Z or Y is treated as a matrix of constants. Using the definition of matrix derivative, the first term is

$$d(YZ)/dX|_{Z \text{ constant}} = (YZ \otimes d/dX)|_{Z \text{ constant}} , \tag{3.7}$$

and using the rules of Kronecker products, (3.7) becomes

$$\begin{aligned}(YZ \otimes (d/dX)I_n)|_{Z \text{ constant}} &= (Y \otimes d/dX)(Z \otimes I_n) \\ &= (dY/dX)(Z \otimes I_n) .\end{aligned} \tag{3.8}$$

The second term of (3.6) is derived similarly.

THEOREM 5. (Inverse Rule). *Let Y be a nonsingular matrix function of an m by n matrix X. Then*

$$d(Y^{-1})/dX = -(Y^{-1} \otimes I_m)(dY/dX)(Y^{-1} \otimes I_n) . \tag{3.9}$$

The proof follows from the Product Rule after writing Y^{-1} as $Y^{-1}YY^{-1}$.

THEOREM 6. (Kronecker Product Rule). *Let the s by t matrix Y and the p by q matrix Z be functions of an m by n matrix X. Then*

$$d(Y \otimes Z)/dX = (Y \otimes dZ/dX) + (I_{(p,s)} \otimes I_m)(Z \otimes dY/dX)(I_{(t,q)} \otimes I_n) . \tag{3.10}$$

Under the Decomposition Rule, the first term on the right-hand side of (3.10) is the derivative with Y held constant. Using the definition of derivative and the associativity of the Kronecker product, the first term is

$$\begin{aligned}d(Y \otimes Z)/dx|_{Y \text{ constant}} &= (Y \otimes Z) \otimes d/dX \\ &= Y \otimes (Z \otimes d/dX) \\ &= Y \otimes dZ/dX .\end{aligned} \tag{3.11}$$

The second term of (3.10), the derivative when Z is constant, is derived by first using (2.8) to reverse the order of the Kronecker product, obtaining

$$Y \otimes Z = I_{(p,s)}(Z \otimes Y)I_{(t,q)} , \tag{3.12}$$

then using the Product Rule. Vetter [7] and Neudecker [4] give results similar to Theorem 6, but they must define *ad hoc* matrices for the second term on the right-hand side of (3.10) since they do not have the permuted identity matrix and identity (2.8).

In many cases it is convenient to take derivatives of the pack of a matrix or with respect to the pack of a matrix. Pack derivatives are especially useful in calculating the Hessian which involves differentiating first with respect to $\mathcal{P}(X)$ and then with respect to $\mathcal{P}'(X)$. In general, derivatives involving packs may be expressed in terms of ordinary matrix derivatives through the use of identities (2.2), (2.3), and (2.9). As an example, consider the differentiation of the p by q matrix Y by the pack of an m by n matrix X. Since the vector of derivative operators $d/d\mathcal{P}(X)$ is equivalent to $\mathcal{P}(d/dX)$, the derivative may be written, using (2.2) as

$$(3.13) \qquad dY/d\mathcal{P}(X) = Y \otimes \mathcal{P}(d/dX) = Y \otimes (d/dX' \otimes I_m)\mathcal{P}(I_m) ,$$

which may be factored to give

$$(3.14) \qquad dY/d\mathcal{P}(X) = (dY/dX' \otimes I_m)(I_q \otimes \mathcal{P}(I_m)) .$$

It is possible to calculate many other derivatives which involve packs in various ways, but since they are not of general interest they will not be given here.

Derivatives of star products may be calculated from the definition of matrix derivatives using identities (2.12) and (2.15) along with the Decomposition Theorem.

THEOREM 7. (Star Product Rule.) *Let Y and Z be p by q and ps by qt matrix functions of an m by n matrix X.*

$$(3.15) \qquad d(Y * Z)/dX = Y * dZ/dX + Z * (I_p \otimes \mathcal{P}(I_s) \otimes I_m)(dY/dX)(I_q \otimes \mathcal{P}'(I_t) \otimes I_n) .$$

COROLLARY. *If X and Z are the same dimension, then*

$$(3.16) \qquad d(Y * Z)/dX = Y * dZ/dX + Z * dY/dX .$$

The Star Product Rule may also be used to find derivatives of Y' in terms of derivatives of Y, since from the star product identity (2.14) the transpose of a p by q matrix Y may be written as $Y' = Y * I_{(q,p)}$.

The final derivative theorem involves a scalar function of a matrix. While it is possible to derive a general rule for differentiating a matrix function of a matrix (see Vetter [7]), the expressions in general are of a form which can be handled by repeated application of the foregoing theorems, and a special theorem is not necessary.

THEOREM 8. (Chain Rule.) *Let $w = f(Y)$ be a scalar-valued function of a matrix Y which is in turn a function of a matrix X. Then the derivative of w with respect to X may be written*

$$(3.17) \qquad dw/dX = dw/dY * dY/dX .$$

The elements of dw/dX may be evaluated using the ordinary chain rule;

$$(3.18) \qquad dw/dX = \textstyle\sum_{ij} \partial w/\partial y_{ij} \cdot dy_{ij}/dX .$$

But dy_{ij}/dX is simply the ijth submatrix of dY/dX, so that (3.18) is the star product given in (3.17).

Thus far, relations among the elements of the matrix X, if any, have been ignored; derivatives with respect to each element have been calculated with all other elements held constant. If matrix elements are functionally related to one another, however, it may be desirable to calculate "total partial derivatives" where not all of the other matrix elements are held constant while differentiating with respect to one element. In the two most common instances, where the matrix X is either symmetric or diagonal, it is intuitively clear how the appropriate constrained matrix derivatives should be constructed. Unfortunately, when the relationships among the elements of X are more general, the nature of the constrained matrix derivative is ambiguous inasmuch as it may be unclear which matrix elements are to be held constant or not during differentiation.

When matrix constraints appear in the context of an optimization problem, they are most easily handled with Lagrangean multipliers, rather than by calculation of constrained matrix derivatives for the first order conditions. For example, suppose the problem is to maximize a scalar function $f(X)$ of an m by n matrix X subject to a system of linear constraints,

$$X * B = C\,, \tag{3.19}$$

where B is mp by nq and C is p by q. Introduce a p by q matrix of Lagrangean multipliers, Λ, and form the augmented objective function, $g(X, \Lambda)$,

$$g(X, \Lambda) = f(X) + \Lambda * (X * B - C)\,. \tag{3.20}$$

The first-order conditions, obtained by differentiating with respect to X and Λ while holding the other matrix constant, and setting the results equal to zero, are:

$$\partial g/\partial X = df/dX + \Lambda * B^+ = 0 \tag{3.21}$$

and

$$\partial g/\partial \Lambda = X * B - C = 0\,, \tag{3.22}$$

where B^+ is a permuted version of B:

$$B^+ = I_{(p,m)} B I_{(n,q)}\,. \tag{3.23}$$

4. Application. The theorems, identities and operations developed in this paper can be applied in many areas of multivariate analysis. This section deals with one such application, the determination of an optimal strategy in a dynamic model characterized by a system of linear stochastic equations with unknown coefficients and a quadratic objective function. An approximate solution for the scalar or single-equation case has been developed by MacRae [3], but without appropriate notation for matrix differentiation and special operations the extension to the multi-equation case is almost impossibly cumbersome.

In the multi-equation decision problem, the vectors of policy variables, $u_0, u_1, \cdots, u_{N-1}$ and the vectors of state variables, $x_0, x_1, \cdots, x_N$ are related through a system of linear stochastic equations with unknown coefficients. The changing level of information about the coefficients (gained from observations on the state and policy variables in each period) is captured by modeling the unknown coefficients as random variables, deterministically related between periods to reflect the changes in the level of uncertainty. Mathematically, the system of equations is

$$x_{k+1} = \mathbf{B}_k u_k + \mathbf{C} z_k + \varepsilon_k\,, \qquad k = 0, \cdots, N-1\,. \tag{4.1}$$

The vectors $z_0, z_1, \cdots, z_{N-1}$ are known exogenous variables, and ε_k is a normally distributed random vector with mean zero, with covariance matrix Ω (same for all k), and independent of ε_j if $j \neq k$. For conciseness, (4.1) may be written

$$x_{k+1} = \mathbf{D}_k w_k + \varepsilon_k\,, \qquad k = 0, \cdots, N-1\,, \tag{4.2}$$

where $\mathbf{D}_k \equiv [\mathbf{B}_k, \mathbf{C}_k]$ and $w_k' \equiv [u_k', z_k']$. The n by p matrices $\mathbf{D}_k$, $k = 0, \cdots, N-1$, are random with mean D (same for all k) and covariance matrix Γ_k (covariances among the elements of $\mathbf{D}_k$ arranged by rows), and are statistically independent of one another and ε's. The np by np-matrix Γ_k is defined by a dynamic equation:

$$\Gamma_{k+1}^{-1} = \Gamma_k^{-1} + E\{(I_n \otimes w_k)\Omega^{-1}(I_n \otimes w_k') \mid x_0\} \qquad k = 0, \cdots, N-1 \tag{4.3}$$

where the expectation is with respect to the $\mathbf{D}$'s and ε's, and Γ_0 and x_0 are given. The problem of the decision maker is to choose vectors of policy variables, $u_0, u_1, \cdots, u_{N-1}$, subject to constraints (4.2) and (4.3) so as to minimize the expected value of a quadratic objective function,

$$J = E\{\textstyle\sum_{k=1}^{N} \frac{1}{2} x_k' Q_k x_k + \frac{1}{2} u_{k-1}' R_k u_{k-1} + x_k' s_k + u_{k-1}' t_k \mid x_0\}\,, \tag{4.4}$$

where the expectation is with respect to the $\mathbf{D}$'s and ε's.

To solve the problem, matrices of Lagrangean multipliers are introduced to deal with the covariance constraint (4.3), yielding an augmented objective function:

$$V = J + \textstyle\sum_{k=0}^{N-1} M_{k+1} * (\Gamma_{k+1}^{-1} - \Gamma_k^{-1} - E\{(I_n \otimes w_k)\Omega^{-1}(I_n \otimes w_k') \mid x_0\}\,. \tag{4.5}$$

The algorithm, then, for deriving optimal strategies is recursive. First of all, the Lagrangean matrices, $M_0, \cdots, M_N$, are assumed to be arbitrarily chosen. Then in each period k it is assumed that policy variables for all previous periods (through u_{k-2}) are given and that all future variables (from u_k^*, x_{k+1}^* on) have been optimally chosen as functions of earlier variables. Next, expressions are determined for the current variables, u_{k-1}, x_k, Γ_{k-1} to minimize V subject to the state constraint (4.2). Finally, the matrices M_k are chosen to satisfy the variance constraint (4.3) (which depends on M through the optimal expressions for Γ, u, and x) so that V and J are equal.

Let V_k be the value of V in period k when the optimal expressions for future policy and state variables are substituted. The form of V_k will be the same for all k, namely:

$$\begin{aligned} V_k = E\{&\tfrac{1}{2} x_k' K_k x_k + \tfrac{1}{2} u_{k-1}' R_k u_{k-1} + x_k' g_k + u_{k-1}' t_k \\ &- \tfrac{1}{2} M_k^*(I_n \otimes w_{k-1})\Omega^{-1}(I_n \otimes w_{k-1}') \mid x_0\} + \tfrac{1}{2}(M_{k-1} - M_k) * \Gamma_{k-1}^{-1} \\ &+ (\text{terms not involving } u_{k-1}, \Gamma_{k-1}, \text{ or } x_k)\,. \end{aligned} \tag{4.6}$$

This form is clearly valid for $k = N$ with $g_N = s_N$. By induction it will be shown valid for all k, $k = 1, \cdots, N$.

Substitute the state constraint into (4.5), rewrite the second line as a quadratic form using the star product identity (2.16), and combine terms to obtain:

$$
\begin{aligned}
V_k = {} & E\{\tfrac{1}{2}w'_{k-1}(\mathbf{D}'_{k-1}Q_k\mathbf{D}_{k-1} - \Omega^{-1} * M_k)w_{k-1} + \tfrac{1}{2}u'_{k-1}R_k u_k \\
& + w'_{k-1}\mathbf{D}'_{k-1}g_k + u'_{k-1}t_k \mid x_0\} + \tfrac{1}{2}(M_{k-1} - M_k) * \Gamma_{k-1}^{-1} \\
& + (\text{terms not involving } u_{k-1}, \Gamma_{k-1}, x_k)\,.
\end{aligned} \tag{4.7}
$$

Next, evaluate the expectation of the **D** expressions, using the star product identity (2.13) to separate out the nonrandom Q_k in the first expression, and recalling that $\mathbf{D}_{k-1}$ is independent of w_{k-1}, yielding:

$$
\begin{aligned}
V_k = {} & E\{\tfrac{1}{2}w'_{k-1}(D'Q_kD + Q_k * \Gamma_{k-1} - \Omega^{-1} * M_k)w_{k-1} + \tfrac{1}{2}u'_{k-1}R_k u_{k-1} \\
& + w'_{k-1}D'g_k + u'_{k-1}t_k \mid x_0\} + \tfrac{1}{2}(M_{k-1} - M_k) * \Gamma_{k-1}^{-1} \\
& + (\text{terms not involving } u_{k-1}, \Gamma_{k-1}, x_k)\,.
\end{aligned} \tag{4.8}
$$

The expression V_k will be minimized with respect to u_{k-1} if the expression within the expectation is minimized. Use (2.17) to expand the first quadratic term of (4.8) in terms of x_{k-1}, u_{k-1}, z_{k-1} and submatrices of Γ_{k-1} and M_k, then differentiate with respect to u_{k-1}, using the Product Rule and the pack identity (2.2):

$$
\begin{aligned}
\partial V_k/\partial u_{k-1} = {} & [B'Q_kB + Q_k * \Gamma_{k-1}^{BB} - \Omega^{-1} * M_k^{BB} + R_k]u_{k-1} \\
& + [B'Q_kC + Q_k * \Gamma_{k-1}^{BC} - \Omega^{-1} * M_k^{BC}]z_{k-1} + B'g_k + t_k\,,
\end{aligned} \tag{4.9}
$$

where the superscripts on Γ and M refer to particular submatrices. The optimal expression for u_{k-1} is therefore given by setting $\partial V_k/\partial u_{k-1}$ equal to zero, yielding:

$$
u_{k-1} = -H_{k-1}^{-1}f_{k-1} \tag{4.10}
$$

where H_{k-1} is the expression in the first pair of brackets in (4.9) and f_{k-1} is the expression on the last line. Minimization with respect to Γ_{k-1} is done by differentiating with respect to Γ_{k-1}^{-1}, using the identity (2.16) again, the Inverse Rule, the Star Product Rule, and the fact that Γ_{k-1} is nonrandom, yielding a first order condition:

$$
M_{k-1} = M_k + \Gamma_{k-1}[E\{(I_n \otimes w_{k-1})Q_k(I_n \otimes w'_{k-1})x_0\}]\Gamma_{k-1}\,. \tag{4.11}
$$

To show that V_{k-1} has the same form as V_k, substitute the optimal expression for u_{k-1} into V_k:

$$
\begin{aligned}
V_{k-1} & = V_k(u_0, \cdots, u_{k-1}^*, x_0, \cdots, x_k{}^*) \\
& = E\{\tfrac{1}{2}x'_{k-1}Q_{k-1}x_{k-1} + \tfrac{1}{2}u'_{k-2}R_{k-1}u_{k-2} + x'_{k-1}g_{k-1} + u'_{k-2}t_{k-1} \\
& \quad - \tfrac{1}{2}M_{k-1} * (I_n \otimes w_{k-2})\Omega^{-1}(I_n \otimes w'_{k-2}) \mid x_0\} \\
& \quad + \tfrac{1}{2}(M_{k-2} - M_{k-1}) * \Gamma_{k-2} \\
& \quad + (\text{terms not involving } u_{k-2}, x_{k-1} \text{ or } \Gamma_{k-2})\,,
\end{aligned} \tag{4.12}
$$

and

$$
g_{k-1} = s_{k-1} - F'_{k-1}H_{k-1}^{-1}f_{k-1}\,. \tag{4.13}
$$

Since V_k has the same form for all k, the optimal strategies for this problem are defined for all k by equations (4.10), the state constraints (4.2), the first-order conditions for variance (4.11), and the variance constraints (4.3) which implicitly determine the Lagrangean matrices, M. The equation (4.13) is definitional. For a more extensive discussion and interpretation of this problem and its solution, see MacRae [2].

5. Conclusion. This paper has presented a number of identities and developed a number of theorems for matrix calculus which facilitate operations on multivariate problems. The derivative theorems, which are based on the concept of derivative operators, give general rules for dealing with the more common matrix expressions, while the new special operations and identities provide a means of rewriting cumbersome matrix expressions in more concise and useful forms. The utility of these identities and theorems in demonstrated by applying them to an adaptive decision problem, the solution to which would have been tedious if not impossible without them.

REFERENCES

[1] ANDERSON, T. W. (1958). *Introduction to Multivariate Analysis*. Wiley, New York.

[2] MACRAE, ELIZABETH CHASE (1972). Optimal estimation and control: A structural approximation. Special Studies Paper 27, Federal Reserve Board.

[3] MACRAE, ELIZABETH CHASE (1972). Linear decision with experimentation. *Ann. Econ. Soc. Measurement* **1** 437–447.

[4] NEUDECKER, H. (1969). Some theorems on matrix differentiation with special references to Kronecker matrix products. *J. Amer. Statist. Assoc.* **64** 953–963.

[5] NISSEN, DAVID H. (1968). A note on the variance of a matrix. *Econometrica* **36** 603–604.

[6] TRACY, DERRICK S. and DWYER, PAUL S. (1969). Multivariate maxima and minima with matrix derivatives. *J. Amer. Statist. Assoc.* **64** 1576–1594.

[7] VETTER, W. J. (1970). Derivatives operations on matrices. *IEEE Trans. Automatic Control* **AC-15** 241–244.

DEPARTMENT OF ECONOMICS
UNIVERSITY OF MARYLAND
COLLEGE PARK, MARYLAND 20742