Kronecker Products and Matrix Calculus in System Theory

JOHN W. BREWER

Abstract—The paper begins with a review of the algebras related to Kronecker products. These algebras have several applications in system theory including the analysis of stochastic steady state. The calculus of matrix valued functions of matrices is reviewed in the second part of the paper. This calculus is then used to develop an interesting new method for the identification of parameters of linear time-invariant system models.

I. INTRODUCTION

THE ART of differentiation has many practical applications in system analysis. Differentiation is used to obtain necessary and sufficient conditions for optimization in analytical studies or it is used to obtain gradients and Hessians in numerical optimization studies. Sensitivity analysis is yet another application area for differentiation formulas. In turn, sensitivity analysis is applied to the design of feedback and adaptive feedback control systems.

The conventional calculus could, in principle, be applied to the elements of a matrix valued function of a matrix to achieve these ends. "Matrix calculus" is a set of differentiation formulas which is used by the analyst to preserve the matrix notation during the operation of differentiation. In this way, the relationships between the various element derivatives are more easily perceived and simplifications are more easily obtained. In short, matrix calculus provides the same benefits to differentiation that matrix algebra provides to the manipulation of systems of algebraic equations.

The first purpose of this paper is to review matrix calculus [5], [16], [22], [24], [28], with special emphasis on applications to system theory [1], [2], [6]–[10], [15], [19], [20], [26], [29], [30]. After some notation is introduced in Section II, the algebraic basis for the calculus is developed in Section III. Here the treatment is more complete than is required for the sections on calculus since the algebras related to the Kronecker product have other applications in system theory [3], [12], [13], [17], [18]. Matrix calculus is reviewed in Section IV and the application to the sensitivity analysis of linear time-invariant dynamic systems [7], [9], [10] is discussed.

The second purpose of this paper is to provide a new numerical method for solving parameter identification problems. These new results are presented in Section V

and are based on the developments of Section IV. Also, a novel and interesting operator notation will be introduced in Section VI.

Concluding comments are placed in Section VII.

II. NOTATION

Matrices will be denoted by upper case boldface (e.g., A) and column matrices (vectors) will be denoted by lower case boldface (e.g., x). The kth row of a matrix such as A will be denoted A_k. and the kth column will be denoted $A_{\cdot k}$. The ik element of A will be denoted a_{ik}.

The $n \times n$ unit matrix is denoted I_n. The q-dimensional vector which is "1" in the kth and zero elsewhere is called the *unit vector* and is denoted

$$e_k \cdot$$
$$(q)$$

The parenthetical underscore is omitted if the dimension can be inferred from the context. The *elementary matrix*

$$E_{ik}^{(p \times q)} \triangleq e_i e_k' \qquad (1)$$
$$(p)(q)$$

has dimensions $(p \times q)$, is "1" in the $i - k$ element, and is zero elsewhere. The parenthetical superscript notation will be omitted if the dimensions can be inferred from context.

The *Kronecker product* [4] of A $(p \times q)$ and B $(m \times n)$ is denoted $A \otimes B$ and is a $pm \times qn$ matrix defined by

$$A \otimes B \triangleq \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1q}B \\ a_{21}B & & & \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ a_{p1}B & & & a_{pq}B \end{bmatrix}. \qquad (2)$$

The *Kronecker sum* [4] of N $(n \times n)$ and M $(m \times m)$ is defined by

$$N \oplus M = N \otimes I_m + I_n \otimes M. \qquad (3)$$

Define the *permutation matrix* by

$$U_{p \times q} \triangleq \sum_{i}^{p} \sum_{k}^{q} E_{ik}^{(p \times q)} \otimes E_{ki}^{(q \times p)}. \qquad (4)$$

This matrix is square $(pq \times pq)$ and has precisely a single "1" in each row and in each column. Define a related

<div style="column: left">

TABLE I
THE BASIC RELATIONSHIPS

δ_{ik} Denotes the Kronecker Delta

T1.1 $\quad \underset{(p)}{e_i'} \; \underset{(p)}{e_k} = \delta_{ik}$

T1.2 $\quad E_{ik}^{(p \times q)} E_{mn}^{(q \times r)} = \delta_{km} E_{in}^{(p \times r)}$

T1.3 $\quad A = \sum_i^p \sum_k^q A_{ik} E_{ik}^{(p \times q)}$

T1.4 $\quad E_{ik}^{(s \times p)} A E_{mn}^{(q \times r)} = A_{km} E_{in}^{(s \times r)}$

T1.5 $\quad (U_{p \times q})' = U_{q \times p}$

T1.6 $\quad U_{p \times q}^{-1} = U_{q \times p}$

T1.7 $\quad (E_{ik}^{(p \times q)})' = E_{ki}^{(q \times p)}$

T1.8 $\quad U_{p \times 1} = U_{1 \times p}$

$\qquad = I_p$

T1.9 $\quad U_{n \times n} = U_{n \times n}'$

$\qquad = U_{n \times n}^{-1}$

Symmetric and Orthogonal

T1.10 $\quad U_{n \times n} \bar{U}_{n \times n} = \bar{U}_{n \times n}$

Dimensions of Matrices Used in Text and in Tables II–VI

$$\begin{array}{ll}
A(p \times q) & R(s \times t) \\
B(s \times t) & Q(q \times q) \\
C(r \times l) & P(q \times q) \\
D(q \times s) & W(s \times p) \\
F(q \times u) & X(n \times n) \\
G(t \times u) & Y(n \times n) \\
H(p \times q) & x(n \times 1) \\
L(n \times n) & y(q \times 1) \\
M(m \times m) & z(n \times 1) \\
N(n \times n) &
\end{array}$$

matrix by

$$\bar{U}_{p \times q} \triangleq \sum_i^p \sum_k^q E_{ik}^{(p \times q)} \otimes E_{ik}^{(p \times q)}. \tag{5}$$

This matrix is rectangular $(p^2 \times q^2)$.

An important vector valued function of a matrix was defined by Neudecker [16]:

$$\underset{(pq \times 1)}{\text{vec } (A)} \triangleq \begin{bmatrix} A_{.1} \\ \hline A_{.2} \\ \hline \cdot \\ \cdot \\ \cdot \\ \hline A_{.q} \end{bmatrix}. \tag{6}$$

Barnett [3] and Vetter [28] introduced other vector valued functions which can be simply related to vec $(\cdot)$. These functions can simplify matrix calculus formulas somewhat, but will not be discussed further here. Let M be $(m \times m)$ and define the vector valued function [19]

$$\underset{m \times 1}{\text{vecd } (M)} \triangleq \begin{bmatrix} m_{11} \\ m_{22} \\ \cdot \\ \cdot \\ \cdot \\ m_{mm} \end{bmatrix} \tag{7}$$

</div>

<div style="column: right">

TABLE II
THE ALGEBRA OF KRONECKER PRODUCTS AND KRONECKER SUMS

Refer to Table I (bottom) for a list of dimensions. α_k is an eigenvector of M with eigenvalue λ_k; β_k is a eigenvector of N with eigenvalue μ_k; $f(\cdot)$ is an analytic function.

Theorems		References
T2.1	$(A \otimes B) \otimes C = A \otimes (B \otimes C)$	[4]
T2.2	$(A + H) \otimes (B + R) = A \otimes B + A \otimes R$	[4]
	$\quad + H \otimes B + H \otimes R$	
T2.3	$(A \otimes B)' = A' \otimes B'$	[4]
T2.4	$(A \otimes B)(D \otimes G) = AD \otimes BG$	[4]

The "Mixed Product Rule"

T2.5	$B \otimes A = U_{s \times p}(A \otimes B)U_{q \times t}$	[3], [28]
T2.6	$(N \otimes M)^{-1} = N^{-1} \otimes M^{-1}$	[4]
T2.7	$\det (N \otimes M) = (\det N)^m (\det M)^n$	[4]
T2.8	$\text{trace } (N \otimes M) = \text{trace } N \text{ trace } M$	[4]
T2.9	$(I_m \otimes N)(M \otimes I_n) = (M \otimes I_n)(I_m \otimes N)$	[17]
T2.10	$f(I_m \otimes N) = I_m \otimes f(N)$	[3], [7], [10]
T2.11	$f(N \otimes I_m) = f(N) \otimes I_m$	[3]
T2.12	$\exp (N \otimes M) = \exp (N) \otimes \exp (M)$	[17]
T2.13	$\text{vec } (ADB) = (B' \otimes A) \text{ vec } D$	[16]
T2.14	$\beta_k \otimes \alpha_i$ is an eigenvector of $N \otimes M$ with eigenvalue $\lambda_i \mu_k$ and is also an eigenvector of $N \oplus M$ with eigenvalue $\lambda_i + \mu_k$.	[4]
T2.15	$N \otimes M$ is positive definite if N, M are symmetric and sign definite of the same sign. $N \otimes M$ is negative definite if N and M are sign definite of opposite sign.	
T2.16.	If N and M are symmetric and sign definite of the same sign then $N \oplus M$ is also sign definite of that sign.	
T2.17	$(I_p \otimes z)A = A \otimes z$	
T2.18	$A(I_q \otimes z') = A \otimes z'$	

that is, the vector formed from the diagonal elements of M.

III. THE ALGEBRAIC BASIS FOR MATRIX CALCULUS

A. Basic Theorems

The implications of the above definitions are listed in Tables I and II. The dimensions of the matrices in Table II are given in Table I (bottom). The proofs of T1.1, T1.3, T1.7, T1.8, T2.1, T2.2, T2.3, and T2.8 are obvious. T1.2 follows from (1) and T1.1. Entry T1.4 follows from T1.2 and T1.3. T1.5 follows from T2.3.

Use the rules of multiplication of partitioned matrices to show that the $i - k$ partition of $(A \otimes B)(D \otimes G)$ is

$$\sum_r a_{ir} d_{rk} BG$$

</div>

which is also the $i - k$ partition of $AD \otimes BG$ so that very important T2.4 is established. This "mixed product rule" (T2.4) is used to establish many of the other theorems listed in this review.

The mixed product rule appeared in a paper published by Stephanos in 1900.

T1.6 follows from (4), T2.4, and T1.2. T1.10 is derived in a similar manner. Entries T2.6 and T2.9 are immediate consequences of T2.4. An analytic function, $f(\cdot)$, can be expressed as a power series so that T2.10 and T2.11 are also derived from T2.4. Theorem T1.9 is merely a special case of T1.5 and T1.6.

T2.14 is also a consequence of the mixed product rule. For instance, notice that

$$(N \otimes M)(\beta_k \otimes \alpha_i) = N\beta_k \otimes M\alpha_i$$

$$= \mu_k \lambda_i \beta_k \otimes \alpha_i$$

which is a proof of the first part of T2.14. The proof of the second part proceeds in a similar manner.

T2.5 is proved by substituting (4) for both permutation matrices and then applying T2.4, and then applying T1.4, and then applying T1.3.

T2.7 follows from T2.14 and the fact that the determinant of any matrix is equal to the product of its eigenvalues.

T2.12 follows from (3), T2.9, T2.10, T2.11, and the fact that the exponential of the sum of commuting matrices is the product of exponentials. Notice however, that it is not required that M, N themselves commute (indeed M and N need not even be conformable). It is remarkable that the exponential algebra with Kronecker products and sums is more analogous to the scalar case than is the exponential algebra with normal matrix products and sums.

Neudecker [16] proved T2.13 in the following way: partition AD into columns and partition the kth column of B into single elements b_{ik}. It follows from the rule for the multiplication of partitioned matrices that the partition

$$(ADB)_{\cdot k} = \sum_i (AD)_{\cdot i} b_{ik}$$

$$= \sum_i (b_{ik} A) D_{\cdot i}$$

$$= [(B_{\cdot k})' \otimes A] \text{ vec } D$$

$$= [(B')_{k \cdot} \otimes A] \text{ vec } D$$

which establishes T2.13 partition by partition.

T2.15 follows directly from T2.14. Note that many permutations and refinements of theorem T2.15 are possible. For instance, if M is positive semidefinite and N is positive definite, then $M \otimes N$ is positive semidefinite.

T2.16 is an immediate consequence of T2.14.

Substitute the obvious identity $A = A \otimes 1$ into the left-hand side of T2.17 and use T2.4 to prove T2.17. Entry T2.18 is established in a similar manner.

Example 1: The matrix equation

$$LX + XN = Y \tag{8}$$

where X is unknown, appears often in system theory [3], [14], [17], [29]. It is assumed that all matrices in (8) are $(n \times n)$. Stability theory and steady-state analysis of stochastic dynamic systems are examples of such occurrences. It follows from T2.13 that

$$[(I_n \otimes L) + (N' \otimes L_n)] \text{ vec } X = \text{vec } Y$$

so that

$$\text{vec } X = (N' \oplus L)^{-1} \text{ vec } Y. \tag{9}$$

if $N' \oplus L$ is nonsingular. Since the determinant of a matrix is equal to the product of its eigenvalues, it follows from T2.14 that solution (9) exists if

$$\mu_k + \gamma_i \neq 0 \tag{10}$$

for any i, k where μ_k is an eigenvalue of N and γ_i is an eigenvalue of L. This existence theorem is well known [4].

It is commonly thought that (9) is not useful for computations because $N' \oplus L$ is $n^2 \times n^2$. Vetter has done much to reduce the severity of this dimensionality problem [29]. An alternate computational technique based on (9) will now be presented.

A matrix and its inverse share eigenvectors and the eigenvalues of the inverse are reciprocals of those of the matrix. It follows from the principle of spectral representation [32] and from T2.14 that

$$(N' \oplus L)^{-1} = \sum_i^n \sum_k^n \frac{(b_k \otimes w_i)(\beta_k \otimes v_i)'}{\mu_k + \gamma_i} \tag{11}$$

where b_k, w_i, β_k, and v_i are, respectively, eigenvectors of N', L, N, and L'. The "reciprocal basis vectors" b_k and v_i must be normalized so that [32]

$$b_k' \beta_k = 1 = v_i' w_i.$$

Combine (9) and (11) with T2.3, T2.4, and T2.13 to obtain the solution:

$$X = \sum_i^n \sum_k^n \frac{w_i v_i' Y \beta_k b_k'}{\mu_k + \gamma_i}. \tag{12}$$

This solution is restricted to the case where $N' \oplus L$ is "nondefective" (has a full set of linearly independent eigenvectors). The above derivation is original in detail but the same result could have been obtained by using the more general theory attributed by Lancaster to Krein [14]. A reviewer of this paper also succeeded in deriving (12) by substituting the similarity transformations to diagonal form for L and N into (8).

B. Auxiliary Results

Additional algebraic relations are displayed in Table III.

T3.1 is obtained immediately from (6).

T3.4 are immediate consequences of T2.13.

Notice that $(D')_{\cdot k}$ is the kth row of D as a column

TABLE III
AUXILIARY THEOREMS

For the vec- and vecd-functions, Kronecker powers and the Khatri–Rao product. The matrices are dimensioned as in Table I (bottom).

Theorems		References
T3.1	$\text{vec }(A + H) = \text{vec }(A) + \text{vec }(H)$	
T3.2	$\text{vec }(A') = U_{p \times q}\, \text{vec }(A)$	[28]
T3.3	$\text{vec }(A) = \sum\limits_{k}^{q} \underset{(q)}{(e_k} \otimes I_p) A\, \underset{(q)}{e_k}$	[16], [28]
T3.4	$\text{vec }(AD) = (I_s \otimes A)\, \text{vec }(D)$	[16]
	$= (D' \otimes I_p)\, \text{vec }(A)$	[16]
	$= (D' \otimes A)\, \text{vec }(I_q)$	[16]
T3.5	$\text{vec }(AD) = \sum\limits_{k}^{q} (D')_{\cdot k} \otimes A_{\cdot k}$	[16]
T3.6	$A^{[k+1]} = A \otimes A^{[k]}$	[4]
T3.7	$(AD)^{[k]} = A^{[k]} D^{[k]}$	[4]
T3.8	$\text{trace }(ADW) = (\text{vec }(A'))'(I_p \otimes D)\, \text{vec }(W)$	[16]
T3.9	$\text{trace }(A'H) = (\text{vec }(A))'\, \text{vec }(H)$	[16]
T3.10	$A \odot (D' \odot F') = (A \odot D') \odot F'$	[13]
T3.11	$L \odot N = U_{n \times n}(N \odot L)$	
T3.12	$(A \otimes B)(F \odot G) = AF \odot BG$	[13]
T3.13	$\text{vec }(AVD) = (D' \odot A)\, \text{vecd }(V)$ if V is diagonal and $(q \times q)$	[19]
T3.14	$A_{\cdot k} = A\, \underset{(q)}{e_k}$	

vector and obtain T3.5 from the rule of multiplication of partitioned matrices.

It follows from (4), T1.7, and T2.13 that

$$U_{p \times q}\, \text{vec }(A) = \text{vec }\left(\sum_i^p \sum_k^q E_{ki}^{(q \times p)} A E_{ki}^{(q \times p)} \right).$$

Use T1.4 to show that the right-hand side becomes

$$\text{vec }\left(\sum_i^p \sum_k^q A_{ik} E_{ki}^{(q \times p)} \right) = \text{vec }(A')$$

so that T3.2 is established.

To establish T3.3, merely notice that $Ae_k = A_{\cdot k}$ (hence, T3.14) and that $\underset{(q)}{e_k} \otimes I_p$ is $qp \times p$ with I_p the pth partition and all other elements zero. T3.3 then follows from the rule of multiplication for partitioned matrices.

Notice that, by T2.13 and T3.3

$$(I_p \otimes D)\, \text{vec }W = \text{vec }DW$$

$$= \sum_k^p \underset{(p)}{(e_k} \otimes I_q) DW \underset{(p)}{e_k}.$$

Use this relation with the second of T3.4 to show that the

right-hand side of T3.8 becomes

$$\sum_k^p (\text{vec }(A'e_k))'\, DW\, \underset{(p)}{e_k} = \sum_k^p \underset{(p)}{e_k'}\, ADW\, \underset{(p)}{e_k}$$
$$= \text{trace }(ADW)$$

so that T3.8 is established.

T3.9 is a consequence of T3.8.

Define the *Kronecker square* of a matrix by

$$A^{[2]} \triangleq A \otimes A \tag{13}$$

with the Kronecker power, $A^{[k]}$, defined in a similar manner. T3.6 is an obvious result of this definition. T3.7 follows from T3.6 and T2.4 and is remarkable since it is not required that A and D commute.

If G is $(t \times u)$ and F is $(q \times u)$ (that is, they have the same number of columns) the Khatri–Rao product is denoted $F \odot G$ and is defined by [13]:

$$F \odot G \triangleq \left[\, F_{\cdot 1} \otimes G_{\cdot 1} \;\vdots\; F_{\cdot 2} \otimes G_{\cdot 2} \;\vdots\; \cdots \;\vdots\; F_{\cdot u} \otimes G_{\cdot u} \,\right]. \tag{14}$$

T3.10 is easily obtained and T3.11 is a consequence of (14), T2.5, the rule for multiplication of partitioned matrices and the fact that $U_{1 \times 1} = 1$.

T3.12 is obtained from the rule for multiplication of partitioned matrices, T2.4, and the facts that $AF_{\cdot k} = (AF)_{\cdot k}$ and $BG_{\cdot k} = (BG)_{\cdot k}$.

It follows from T3.5 that

$$\text{vec }(AVD) = \sum_k^q (D')_{\cdot k} \otimes (AV)_{\cdot k}.$$

However, if V is diagonal

$$(AV)_{\cdot k} = A_{\cdot k} v_{kk}$$

so

$$\text{vec }(AVD) = \sum_k^q (D')_{\cdot k} \otimes A_{\cdot k} v_{kk}. \tag{15}$$

But

$$(D' \odot A)\, \text{vecd }(V) = \sum_k^q (D' \odot A)_{\cdot k} v_{kk}$$
$$= \sum_k^q (D')_{\cdot k} \otimes A_{\cdot k} v_{kk}. \tag{16}$$

Equations (15) and (16) establish T3.13.

IV. THE MATRIX CALCULUS

Take the derivative of a matrix A with respect to a scalar b to be

$$\frac{\partial A}{\partial b} = \left(\frac{\partial a_{ik}}{\partial b} \right). \tag{17}$$

Vetter's definition of the matrix derivative is used here [24], [28]. $\partial A / \partial B$ is taken to be a partitioned matrix whose $i - k$ partition is

$$\frac{\partial A}{\partial b_{ik}}. \tag{18}$$

Vetter's calculus was built upon the foundation laid by Dwyer [11] and Neudecker [16]. Turnbull developed a calculus along alternative lines [22], [23].

Basic theorems are listed in Table IV.

T4.1 follows from definition (18).

T4.2 follows from T4.1, T2.3, and T1.7.

To establish T4.3, note that

$$\frac{\partial(AF)}{\partial b_{ik}} = \frac{\partial A}{\partial b_{ik}} F + A \frac{\partial F}{\partial b_{ik}} \qquad (19)$$

and combine this fact with T4.1, the fact that $E_{ik}^{(s\times t)} = I_s E_{ik}^{(s\times t)} = E_{ik}^{(s\times t)} I_t$, and T2.4 to show

$$\frac{\partial(AF)}{\partial B} = \sum_{i,k}\left[\left(E_{ik}^{(s\times t)}\otimes\frac{\partial A}{\partial b_{ik}}\right)(I_t\otimes F)\right.$$

$$\left. + (I_s\otimes A)\left(E_{ik}^{(s\times t)}\otimes\frac{\partial F}{\partial b_{ik}}\right)\right]$$

which is T4.3.

To prove T4.4 notice that

$$\frac{\partial(A\otimes C)}{\partial b_{ik}} = \frac{\partial A}{\partial b_{ik}}\otimes C + A\otimes\frac{\partial C}{\partial b_{ik}}.$$

Combine this fact with T4.1 and T2.1 to show

$$\frac{\partial(A\otimes C)}{\partial B} = \frac{\partial A}{\partial B}\otimes C + \sum_{i,k}E_{ik}^{(s\times t)}\otimes A\otimes\frac{\partial C}{\partial b_{ik}}. \qquad (20)$$

It follows from T2.5 that

$$E_{ik}^{(s\times t)}\otimes A\otimes\frac{\partial C}{\partial b_{ik}} = \left(I_s E_{ik}^{(s\times t)}I_t\right)\otimes\left[U_{p\times r}\frac{(\partial C\otimes A)}{\partial b_{ik}}U_{l\times q}\right]$$

$$= (I_s\otimes U_{p\times r})\left(E_{ik}^{(s\times t)}\otimes\left[\frac{\partial C}{\partial b_{ik}}\otimes A\right]\right)(I_t\otimes U_{l\times q}) \qquad (21)$$

where the last equality follows from repeated use of T2.4. T4.4 is the result of substituting (21) into (20).

T4.5 is derived in a similar manner.

Notice that $aA = a\otimes A$ for any scalar a and any matrix A. Combine this fact with T4.4 and T1.8 to derive T4.7.

Begin the derivation of T4.6 by combining T4.1 with T1.3 and the chain rule to obtain

$$\frac{\partial A}{\partial B} = \sum_{i,k}E_{ik}^{(s\times t)}\otimes\sum_{\alpha,\beta}E_{\alpha\beta}^{(p\times q)}\frac{\partial a_{\alpha\beta}}{\partial b_{ik}}$$

$$= \sum_{i,k}E_{ik}^{(s\times t)}\otimes\sum_{\alpha,\beta}E_{\alpha\beta}^{(p\times q)}\sum_{j,l}\frac{\partial a_{\alpha\beta}}{\partial c_{jl}}\frac{\partial c_{jl}}{\partial b_{ik}}$$

$$= \sum_{j,l}\sum_{i,k}E_{ik}^{(s\times t)}\frac{\partial c_{jl}}{\partial b_{ik}}\otimes\sum_{\alpha,\beta}E_{\alpha\beta}^{(p\times q)}\frac{\partial a_{\alpha\beta}}{\partial c_{jl}}$$

$$= \sum_{j,l}\frac{\partial c_{jl}}{\partial B}\otimes\frac{\partial A}{\partial c_{jl}}.$$

TABLE IV
THE BASIC DIFFERENTIATION THEOREMS

All matrices are dimensioned in Table I (bottom).

Theorem		References
T4.1	$\dfrac{\partial A}{\partial B} = \sum\limits_{i,k}E_{ik}^{(s\times t)}\otimes\dfrac{\partial A}{\partial b_{ik}}$	
T4.2 a)	$\left(\dfrac{\partial A}{\partial B}\right)' = \dfrac{\partial A'}{\partial B'}$	[28]
T4.3	$\dfrac{\partial(AF)}{\partial B} = \dfrac{\partial A}{\partial B}(I_t\otimes F)+(I_s\otimes A)\dfrac{\partial F}{\partial B}$	[28]
T4.4 b)	$\dfrac{\partial(A\otimes C)}{\partial B} = \dfrac{\partial A}{\partial B}\otimes C+(I_s\otimes U_{p\times r})$	[28]
	$\qquad\cdot\dfrac{(\partial C\otimes A)}{\partial B}(I_t\otimes U_{l\times q})$	
T4.5	$\dfrac{\partial(L\odot N)}{\partial X} = \dfrac{(\partial L\otimes I_n)(I_n\otimes[I_n\odot N])}{\partial X}$	
	$\qquad + (I_n\otimes U_{n\times n})\dfrac{(\partial N\otimes I_n)(I_n\otimes[I_n\odot L])}{\partial X}$	
T4.6	$\dfrac{\partial A(C(B))}{\partial B} = (I_s\otimes\dfrac{\partial A}{\partial\text{ vec } C})\dfrac{(\partial[\text{vec }(C')]'\otimes I_q)}{\partial B}$	[28]
	$\qquad = \dfrac{(\partial[\text{vec } C]'\otimes I_p)}{\partial B}(I_t\otimes\dfrac{\partial A}{\partial\text{ vec } C})$	

the "chain rule."

T4.7   The scalar product rule is

$$\frac{\partial(aA)}{\partial B} = \frac{\partial a}{\partial B}\otimes A + a\frac{\partial A}{\partial B}$$

a) This is a very useful result that can be used to extend any of the results in Table V. Notice also the theoretical use of this formula in Section V.

b) This formula is at variance with that published by Vetter, wherein the subscript on the last permutation matrix is $q\times l$.

Now use T2.4 with the above equation and the facts that

$$\frac{\partial c_{jl}}{\partial B} = \frac{\partial c_{jl}}{\partial B}I_t$$

and

$$\frac{\partial A}{\partial c_{jl}} = I_p\frac{\partial A}{\partial c_{jl}}$$

to obtain

$$\frac{\partial A}{\partial B} = \sum_{j,l}\left(\frac{\partial c_{jl}}{\partial B}\otimes I_p\right)\left(I_t\otimes\frac{\partial A}{\partial c_{jl}}\right)$$

which is the same as the second equality in T4.6. The first equality is obtained in an analogous manner.

Table V lists a set of theorems easily obtained from the first four tables. The readers can prove these results themselves or study the references. The only point worthy of note is that "partitioned matrix rule" T5.13, which is easily obtained from T4.4 and T2.5, is used to obtain the derivative of a partition from the derivative of the matrix. For instance consider

$$M = \begin{bmatrix} M_{11} & M_{12} \\ \hline M_{21} & M_{22} \end{bmatrix}$$

TABLE V
APPLICATIONS OF THE BASIC DERIVATIVE FORMULAS

The symmetric matrix $Q = Q'$ and the skew-symmetric matrix $P = -P'$ are $(q \times q)$. The elements of A are mathematically independent. All other matrices are dimensioned in Table I (bottom).

Theorem		References
T5.1	$\dfrac{\partial A}{\partial A} = U_{p \times q}$	[28]
T5.2	$\dfrac{\partial A'}{\partial A} = U_{p \times q}$	[28]
T5.3	$\dfrac{\partial P}{\partial P} = U_{q \times q} - U_{q \times q}$	[28]
T5.4	$\dfrac{\partial Q}{\partial Q} = U_{q \times q} + U_{q \times q} - \sum_{i}^{q} E_{ii}^{(q \times q)} \otimes E_{ii}^{(q \times q)}$	[6], [9]
T5.5	$\dfrac{\partial M^{-1}}{\partial A} = -(I_p \otimes M^{-1}) \dfrac{\partial M}{\partial A} (I_q \otimes M^{-1})$	[28]
T5.6	$\dfrac{\partial y}{\partial y} = \text{vec}(I_q)$	[28]
T5.7	$\dfrac{\partial y'}{\partial y} = I_q$	[28]
T5.8	$\dfrac{\partial (Ay)}{\partial y} = (I_q \otimes A)\, \text{vec}(I_q)$	[28]
	$= \text{vec}\, A$	
T5.9	$\dfrac{\partial (Ay)}{\partial y'} = A$	[28]
T5.10	$\dfrac{\partial (y \otimes y)}{\partial y'} = \dfrac{\partial (y \odot y)}{\partial y'}$	
	$= I_q \otimes y + y \otimes I_q$	[28]
T5.11	$\dfrac{\partial}{\partial y}(y'Qy) = Qy + \text{vec}(y'Q)$	[28]
	$= Qy + Q'y$	
T5.12	$\dfrac{\partial (y'\odot A)}{\partial y} = \dfrac{\partial (A \odot y')}{\partial y}$	
	$= I_q \odot A$	
T5.13	$U_{r \times sp}(I_p \otimes U_{s \times r}) \dfrac{\partial}{\partial A}(E_{ik}^{(r \times l)} \otimes B)$	
	$\cdot (I_q \otimes U_{l \times t}) U_{qt \times l} = E_{ik}^{(r \times l)} \otimes \dfrac{\partial B}{\partial A}$	[9]

The partitioned matrix rule.

T5.14	$\dfrac{\partial y}{\partial x} = \text{vec}\left[\dfrac{\partial y}{\partial x'}\right]$	
T5.15	$\dfrac{\partial A_{\cdot k}}{\partial B} = \dfrac{\partial A}{\partial B}(I_t \otimes e_k)$	
	$\qquad\qquad\qquad (q)$	

where all partitions are $n \times n$. Notice that

$$M = \sum_{i,k}^{2} E_{i,k}^{(2 \times 2)} \otimes M_{ik}.$$

Apply T5.13 to this equation to find

$$U_{2 \times np}(I_p \otimes U_{n \times 2}) \dfrac{\partial M}{\partial A} (I_q \otimes U_{2 \times n}) U_{qn \times 2} = \sum_{i,k}^{2} E_{i,k}^{(2 \times 2)} \otimes \dfrac{\partial M_{ik}}{\partial A}.$$

Thus if the matrix derivative is transformed in the manner indicated, the partitions of the result are the derivatives of the corresponding partitions.

Theorem T5.13 is useful only when all partitions of the matrix have the same dimensions.

Example 2 [7], [10]: The sensitivity of the fundamental solution of a set of unforced linear, time invariant differential equations to parameter variation is studied here. All matrices, vectors, and eigenvalues are defined in example 1.

Denote the fundamental solution

$$\Phi(t) = \exp(Nt). \tag{22}$$

As is well known [32]

$$\frac{d}{dt}\Phi = N\Phi \tag{23}$$

and

$$\Phi(0) = I_n. \tag{24}$$

Let $G(t \times u)$ denote a matrix of parameters and use T4.3 to show that

$$\frac{d}{dt}\frac{\partial \Phi}{\partial G} = (I_t \otimes N)\frac{\partial \Phi}{\partial G} + \frac{\partial N}{\partial G}(I_u \otimes \Phi). \tag{25}$$

Also, from (24)

$$\frac{\partial \Phi}{\partial G}(0) = 0. \tag{26}$$

The solution to (25), (26) is

$$\frac{\partial \Phi}{\partial G} = \int_0^t \exp\left\{(I_t \otimes N)[t - \tau]\right\}\frac{\partial N}{\partial G}\left\{I_u \otimes \exp(N\tau)\right\} d\tau. \tag{27}$$

Assume that N is nondefective and use T2.10 and spectral form to show that (27) becomes

$$\frac{\partial \Phi}{\partial G}(t) = \int_0^t \left[I_t \otimes \exp(N[t - \tau])\right]\frac{\partial N}{\partial G}\left[I_u \otimes \exp(N\tau)\right] d\tau$$

$$= \sum_i^n \sum_\alpha^n (I_t \otimes \beta_i b_i')\frac{\partial N}{\partial G}(I_u \otimes \beta_\alpha b_\alpha')g_{ik}(t) \tag{28}$$

where $g_{i\alpha}(t)$ is the inverse Laplace Transform of

$$\frac{1}{(s - \mu_i)(s - \mu_\alpha)}. \tag{29}$$

These results have been greatly generalized [10]. The sensitivity of Φ has also been determined with the aid of the conventional scalar calculus [21].

V. APPLICATION TO PARAMETER IDENTIFICATION BY NEWTON-RAPHSON ITERATION FOR LINEAR TIME-INVARIANT SYSTEMS

Consider a system modeled by a set of homogeneous time-invariant linear equations:

$$\frac{d}{dt}x = Nx. \tag{30}$$

Suppose that

$$N = N(p) \tag{31}$$

where p represents a set of unknown parameters and is taken to be r-dimensional.

Suppose further, that an estimate $\hat{p}$ is to be deduced from a set of discrete measurements of the state

$$x_*(t_0), x_*(t_1), \cdots, x_*(t_m). \tag{32}$$

For convenience, the arbitrary decision to take

$$\hat{x}(t_0) = x_*(t_0) \tag{33}$$

is made here.

In what follows, the best estimate of p is taken to be that vector which minimizes the ordinary least squares criterion

$$I = \sum_{k=1}^{m} \left[\hat{x}(t_k) - x_*(t_k) \right]' \left[\hat{x}(t_k) - x_*(t_k) \right] \tag{34}$$

where $\hat{x}(t)$ is the solution of (30) with $p = \hat{p}$ and subject to assumption (33). This is a nonlinear estimation problem not amenable to closed form solution. A possible numerical procedure is Newton–Raphson iteration. The gradient and Hessian (matrix of second partial derivatives) of I will be derived in order to implement this procedure.

It follows from T4.3, T5.14, T3.4, and T3.2

$$\frac{\partial I}{\partial p} = \sum_{k=1}^{m} 2 \frac{\partial x'(t_k)}{\partial p} \left[x(t_k) - x_*(t_k) \right]. \tag{35}$$

Use T4.3 and T4.2 to show that the Hessian matrix

$$\frac{\partial^2 I}{\partial p' \partial p} = 2 \sum_{k=1}^{m} \left[\left(\frac{\partial x(t_k)}{\partial p} \right)' \frac{\partial x(t_k)}{\partial p} + \frac{\partial^2 x'(t_k)}{\partial p' \partial p} \right.$$
$$\left. \cdot \left(I_r \otimes \left[x(t_k) - x_*(t_k) \right] \right) \right]. \tag{36}$$

It also follows from T4.2 that

$$\frac{\partial^2 x}{\partial p' \partial p} = \left\{ \frac{\partial}{\partial p} \left(\frac{\partial x}{\partial p'} \right) \right\}'. \tag{37}$$

As is well known, the solution to (30) and (33) is

$$x(t_k) = \Phi(t_k) x_*(t_0) \tag{38}$$

so that

$$\frac{\partial x(t_k)}{\partial p'} = \frac{\partial \Phi(t_k)}{\partial p'} \left\{ I_r \otimes x_*(t_0) \right\} \tag{39}$$

and

$$\frac{\partial}{\partial p} \left(\frac{\partial x}{\partial p'} \right) = \frac{\partial}{\partial p} \left(\frac{\partial \Phi(t_k)}{\partial p'} \right) \left\{ I_r \otimes x_*(t_0) \right\}. \tag{40}$$

If N is nondefective, it follows from (28) that

$$\frac{\partial \Phi}{\partial p'} = \sum_i \sum_\alpha \beta_i b_i' \frac{\partial N}{\partial p'} (I_r \otimes \beta_\alpha b_\alpha') g_{i\alpha}(t). \tag{42}$$

Combine (39) and (42):

$$\frac{\partial x(t_k)}{\partial p'} = \sum_i \sum_\alpha \beta_i b_i' \frac{\partial N}{\partial p'} \left\{ I_r \otimes \beta_\alpha b_\alpha' x_*(t_0) \right\} g_{i\alpha}(t_k). \tag{43}$$

The gradient of (35) is obtained by the use of (43) and T4.2.

It is clear from (36) and (40) that the *second* derivative of the fundamental solution is required. Combine (25)

with T4.3 and T4.4 to show

$$\frac{d}{dt} \left(\frac{\partial}{\partial p} \left(\frac{\partial \Phi}{\partial p'} \right) \right) = (I_r \otimes N) \frac{\partial}{\partial p} \left(\frac{\partial \Phi}{\partial p'} \right) + \frac{\partial N}{\partial p} \frac{\partial \Phi}{\partial p'}$$
$$+ \left[\frac{\partial}{\partial p} \left(\frac{\partial N}{\partial p'} \right) \right] (I_r \otimes \Phi)$$
$$+ \left(I_r \otimes \frac{\partial N}{\partial p'} U_{r \times n} \right) \left(\frac{\partial \Phi}{\partial p} \otimes I_r \right) U_{n \times r}. \tag{44}$$

Also

$$\frac{\partial}{\partial p} \left(\frac{\partial \Phi}{\partial p'} \right) \Big|_{t=0} = 0. \tag{45}$$

Once again, (44) is a linear matrix differential equation for which the solution is well known [32]. If $I_r \otimes N$ is nondefective, the principle of spectral representation leads to

$$\frac{\partial}{\partial p} \left(\frac{\partial \Phi(t)}{\partial p'} \right) = \sum_i^n \sum_\alpha^n (I_r \otimes \beta_i b_i') \frac{\partial}{\partial p} \left(\frac{\partial N}{\partial p'} \right) (I_r \otimes \beta_\alpha b_\alpha') g_{i\alpha}(t)$$
$$+ \sum_i^n \sum_\alpha^n \sum_\gamma^n (I_r \otimes \beta_\gamma b_\gamma') \frac{\partial N}{\partial p} \beta_i b_i' \frac{\partial N}{\partial p'}$$
$$\cdot (I_r \otimes \beta_\alpha b_\alpha') h_{i\alpha\gamma}(t)$$
$$+ \sum_i^n \sum_\alpha^n \sum_\gamma^n \left(I_r \otimes \beta_\gamma b_\gamma' \frac{\partial N}{\partial p'} U_{r \times n} \right)$$
$$\cdot \left\{ \left[(I_r \otimes \beta_i b_i') \frac{\partial N}{\partial p} \beta_\alpha b_\alpha' \right] \otimes I_r \right\} U_{n \times r} h_{i\alpha\gamma}(t) \tag{46}$$

where $h_{i\alpha\gamma}(t)$ is the inverse Laplace transform of

$$\frac{1}{(s - \mu_i)(s - \mu_\alpha)(s - \mu_\gamma)}. \tag{47}$$

Equations (36), (37), (40), (43), and (46) completely specify the Hessian matrix.

The parameter identification problem can be solved by the following iteration method: let $p_{(l)}$ denote the lth iterated value of p, choose $p_{(0)}$ arbitrarily, and use the familiar Newton–Raphson formula

$$p_{(l+1)} = p_{(l)} - \left\{ \frac{\partial^2 I(p_{(l)})}{\partial p' \partial p} \right\}^{-1} \left\{ \frac{\partial I(p_{(l)})}{\partial p} \right\}. \tag{48}$$

Before the method is further expounded for a particular N, some remarks about the general method will be made.

First, the method is restricted to the case of a nondefective N *at each iteration*, this restriction can be removed by using the more general form of the spectral representation [10], [32].

Second, an eigenproblem must be re-solved *at each iteration*. This fact will limit the usefulness of the method to moderately high order systems at best.

Third, iteration scheme (48) need not converge. One would expect the following to be true, however: Gradient search with (43) will bring $p_{(l)}$ close enough to a final value to allow the Newton–Raphson scheme to converge.

As is well known, switching from gradient search to Newton–Raphson iteration greatly increases the rate of convergence.

Fourth, assumption (33) greatly overweighs the value of the initial data point $x_*(t_0)$. It may be better to change the lower summation limit in (34) to zero and treat $\hat{x}(t_0)$ as unknown. Matrix calculus can then be used to obtain gradients and Hessians for the unknown vector

$$\left[\frac{p}{\hat{x}(t_0)} \right]. \tag{49}$$

Also, it may be useful to include a positive definite weighting matrix in (34) to weight the value of particular components of the data vector.

Fifth, the above procedure is easily extended to the case when data is taken on the vector

$$y = Cx \tag{50}$$

rather than on the state vector itself.

These refinements are currently being studied by the author, as are the conditions for nonsingularity of the hessian matrix.

Example 3: Suppose that (31) is in phase variable form so that

$$N = \begin{bmatrix} 0 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 1 & \cdot & \cdot & \cdot & 0 \\ \vdots & & & & & & \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & 1 \\ -p_1 & -p_2 & \cdot & \cdot & \cdot & \cdot & -p_n \end{bmatrix}. \tag{51}$$

First, it is assumed that the eigenvalues of N are distinct which is the only case for which this matrix is nondefective. Second, it is assumed that the p_k are mathematically independent. The latter assumption is sometimes not the case; but the assumption greatly simplifies this exploratory discussion. First notice that

$$N = \sum_{k}^{n-1} E_{k,k+1} - e_n \otimes p' \tag{52}$$

so

$$\begin{aligned} \frac{\partial N}{\partial p'} &= -U_{n \times 1} \left(\frac{\partial p'}{\partial p'} \otimes e_n \right) (I_n \otimes U_{n \times 1}) \\ &= -I_n ((\text{vec } (I_n))' \otimes e_n)(I_n \otimes I_n) \\ &= -([\text{vec } (I_n)]' \otimes e_n) \end{aligned} \tag{53}$$

and

$$\begin{aligned} \frac{\partial N}{\partial p} &= -(I_n \otimes U_{1 \times n}) \left(\frac{\partial p'}{\partial p} \otimes e_n \right) U_{n \times 1} \\ &= -(I_n \otimes I_n)(I_n \otimes e_n) I_n \\ &= -(I_n \otimes e_n). \end{aligned} \tag{54}$$

Further

$$\frac{\partial}{\partial p} \left(\frac{\partial N}{\partial p'} \right) = 0. \tag{55}$$

Equations (53)–(55) complete the formulation for the gradient and Hessian. For instance, (35), (43), and (53) and T4.2 lead to

$$\frac{\partial I}{\partial p} = -2 \sum_i^r \sum_\alpha^n \sum_k^m \left\{ I_n \otimes x_*'(t_0) b_\alpha \beta_\alpha' \right\}$$
$$\cdot \left\{ \text{vec } (I_n) \otimes e_n' \right\} b_i \beta_i' \left\{ \Phi(t_k) x_*(t_0) - x_*(t_k) \right\} g_{i\alpha}(t_k). \tag{56}$$

VI. A New Operator Notation

Denote the matrix of differential operators

$$\mathcal{D}_B = \left(\frac{\partial}{\partial b_{ik}} \right). \tag{57}$$

Define the *Vetter derivative* of A by

$$\mathcal{D}_B \otimes A \tag{58}$$

because the result of this operation is the same as that defined by (18) or T4.1.

Define the *Turnbull derivative* of G by

$$\mathcal{D}_B G \tag{59}$$

because of the similarity of the result of this operation to the definition of the matrix derivative proposed by Turnbull [22]. Notice that the Turnbull derivative is only defined for conformable matrices. Following Turnbull, this definition is extended to a scalar G by using the usual rule for multiplication of a matrix by a scalar.

Thus the operator notation provides a means for combining two apparently divergent theories of matrix calculus. The results obtained by Turnbull are elegant, if somewhat ignored, and it may be well to reinstitute this theory into the mainstream of the matrix calculus. A few of Turnbull's results are listed in Table VI.

The operator notation also promises to lead to new mathematical results. For instance, it follows from T2.5 that

$$(\mathcal{D}_B \otimes \mathcal{D}_C) = U_{s \times r} (\mathcal{D}_C \otimes \mathcal{D}_B) U_{l \times t}. \tag{60}$$

This is a surprising result of mathematical significance because it shows that interchanging the order of partial Vetter differentiation is not as simple as is the case in the calculus of scalar variables. In this context, it should be noted that if C is a single row, c', and B is a single column, b, it follows from (60) and T1.8 that

$$\mathcal{D}_b \otimes \mathcal{D}_{c'} = \mathcal{D}_{c'} \otimes \mathcal{D}_b \tag{61}$$

which is the result that might have been expected.

Finally, it is noted that, as is always the case with operator notation, some care must be exerted. For instance, T2.5 can be used to show

$$\mathcal{D}_B \otimes A = U_{s \times p} (A \otimes \mathcal{D}_B) U_{q \times t} \tag{62}$$

only if the elements of A are mathematically independent of the elements of B.

TABLE VI
TURNBULL DERIVATIVES [22]

All matrices dimensioned as in Table I (bottom). μ_k is an eigenvalue of N. The scalar $s_r = \sum_k^n \mu_k^r$. The subscript c means that a matrix is held constant when operated upon by the matrix of differential operators. L is a constant matrix. The p_k are the coefficients of the characteristic equation of N. The elements of N are mathematically independent.

T6.1 $\mathfrak{D}_N(X + Y) = \mathfrak{D}_N X + \mathfrak{D}_N Y$

T6.2 $\mathfrak{D}_N(XY) = \mathfrak{D}(XY_c) + \mathfrak{D}(X_c Y)$

T6.3 $\mathfrak{D}_{N'}(LN) = \text{trace } (L)I_n$

T6.4 $\mathfrak{D}_N(LN) = L'$

T6.5 $\mathfrak{D}_{N'}(NL) = nL$

T6.6 $\mathfrak{D}_N(NL) = L$

T6.7 $\mathfrak{D}_{N'}(N')^r = N'^{r-1} + NN'^{r-2} + \cdots$
$$+ N^i N'^{r-i-1} + \cdots + N^{r-1}$$

T6.8 $\mathfrak{D}_{N'} N^r = \sum_{i=1}^{n} (N^r - \lambda_i^r I_n)(N - \lambda_i I_n)^{-1}$
$$= \sum_{i=1}^{n} N^{r-1} + \lambda_i N^{r-2} + \cdots + \lambda_i^{r-1} I_n$$
$$= nN^{r-1} + s_1 N^{r-2} + \cdots + s_{j-1} N^{r-j}$$
$$+ \cdots + s_{r-1} I_n$$

T6.9 $\mathfrak{D}_{N'} s_r = rN^{r-1}$

T6.10 $\mathfrak{D}_{N'} p_1 = -I_n$

T6.11 $\mathfrak{D}_{N'} p_r = -p_r I_n - p_{r-1} N - p_{r-2} N^2$
$$- \cdots - N^{r-1}$$

VII. CONCLUDING COMMENTS

Equation (4) and Theorem T4.1 are quite transparent but, to the author's knowledge, they have not appeared in work previous to his own. This form of the definition of permutation and this theorem enable one to more quickly and more concisely derive several of the well known results listed in Tables II–V.

At least two other tabulations should be brought to the reader's attention. Athans and Schweppe [1] provide an extensive tabulation of differentials of important trace and determinant functions. Vetter [27] provides a table of differentials of other scalar valued functions of matrices: eigenvalues, specific elements, and matrix norms of polynomial type and of transcendental type.

In this paper, applications to analysis of stochastic steady-state, to dynamic sensitivity analysis, and to parameter identification theory were studied. Discussion of other applications of Kronecker algebra [3], [8], [12], [13] were omitted but may be of interest to the reader. Other applications of matrix calculus [9], [11], [15], [19], [20], [30] were not touched upon here. In addition, space limitations did not allow for discussion of many important fundamental topics in matrix calculus: matrix variational calculus [2], matrix integrals [28], and matrix Taylor series [23], [28].

Equation (46) is a generalization of the author's previous results [7], [10] to the second derivative of the ex-

ponential matrix. The generalization to still higher derivatives is thereby made apparent. It is interesting that all matrix derivatives of exp (Nt) can be obtained once the eigenproblem for N has been solved. It is noted that the appearance of the spectral matrices in the derivative formulas should come as no surprise since the relationship of these matrices to eigenproblem sensitivity is well known [31].

Finally, some comment about Kronecker products and computer memory should be made. $A \otimes B$ is $ps \times qt$ when A is $p \times q$ and B is $s \times t$ so that the Kronecker product requires $psqt$ storage locations. Actually, one has a choice: A and B can be stored individually in $qp + st$ locations and $A \otimes B$ can be achieved with a little more programming. This is yet another example of the ever present trade-off between program complexity and computer storage requirements.

REFERENCES

[1] M. Athans and F. C. Schweppe "Gradient matrices and matrix calculations," Lincoln Lab., M.I.T., Lexington, MA, Tech. Note TN 1965–63, 1965.

[2] M. Athans, "The matrix minimum principle," Inform. Contr., vol. 11, pp. 592–606, 1968.

[3] S. Barnett, "Matrix differential equations and Kronecker products," SIAM J. Appl. Math., vol. 24, no. 1, Jan. 1973.

[4] R. Bellman, Introduction To Matrix Analysis. New York: McGraw-Hill, 1960, chap. 20.

[5] E. Bodewig, Matrix Calculus. Amsterdam, The Netherlands: North Holland, 1959.

[6] J. W. Brewer, "The gradient with respect to a symmetric matrix," IEEE Trans. Automat. Contr., vol. AC-22, pp. 265–267, Apr. 1977.

[7] ——, "The derivative of the exponential matrix with respect to a matrix," IEEE Trans. Automat. Contr., vol. AC-22, pp. 656–657, Aug. 1977.

[8] ——, "Analytical gradients for optimal environmental monitoring studies," Proc. Int. Conf. Cybernetics and Society, Washington, D.C., Sept. 19–21, 1977.

[9] ——, "The derivative of the Riccati matrix with respect to a matrix," IEEE Trans. Automat. Contr., vol. AC-22, pp. 980–983, Dec. 1977.

[10] ——, "Matrix calculus and the sensitivity analysis of linear dynamic systems, IEEE Trans. Automat. Contr., vol. AC-23, Aug. 1978.

[11] P. S. Dwyer, "Some applications of matrix derivatives in multivariate analysis," J. Amer. Statist. Assoc., vol. 62, June 1967.

[12] T. R. Harper and W. J. Rugh, "Structural features of factorable Volterra systems," IEEE Trans. Automat. Contr., vol. AC-21, pp. 822–832, Dec. 1976.

[13] C. G. Khatri and C. R. Rao, "Solutions to some functional equations and their applications to characterization of probability distributions," Sankhya: The Indian J. Stat., series A, 30, pp. 167–180, 1968.

[14] P. Lancaster, "Explicit solutions of linear matrix equations," SIAM Rev., vol. 12, no. 4, Oct. 1970.

[15] P. C. Muller and H. I. Weber, "Analysis and optimization of certain qualities of controllability and observability for linear dynamical systems," Automatica, vol. 8, pp. 237–246, 1972.

[16] H. Neudecker, "Some theorems on matrix differentiation with special reference to Kronecker matrix products," J. Amer. Stat. Assoc., vol. 64, pp. 953–963, 1969.

[17] H. Neudecker, "A note on Kronecker matrix products and matrix equation systems," SIAM J. Appl. Math., vol. 17, no. 3, May 1969.

[18] P. N. Paraskevopoulos and R. E. King, "A Kronecker product approach to pole assignment by output feedback," Int. J. Contr., vol. 24, no. 3, pp. 325–334, 1976.

[19] M. N. Ransom and R. Saeks, "The connection-function theory and application," Circuit Theory and Appl., vol. 3, pp. 5–21, 1975.

[20] M. N. Ransom and R. Saeks, "A functional approach to fault analysis of linear systems," private communication.

[21] J. C. Reid and P. S. Maybeck, R. B. Asher, and J. D. Dillow "An algebraic representation of parameter sensitivity in linear time invariant systems," J. Franklin Inst., vol. 301, no. 1, Jan.–Feb. 1976.

[22] H. W. Turnbull, "On differentiating a matrix, Proc. Edinburgh Math. Soc., second series, vol. I, pp. 111–128, 1927–29.

[23] ——, "A matrix form of Taylor's theorem," Proc. Edinburgh Math. Soc., second series, vol. II, pp. 33–54, 1930–31.

[24] W. J. Vetter, "Derivative operations on matrices," *IEEE Trans. Automat. Contr.*, vol. AC-15, pp. 241–244, Apr. 1970.
[25] ——, "Correction to 'Derivative operations on matrices,'," *IEEE Trans. Automat. Contr.*, vol. AC-16, p. 113, Feb. 1971.
[26] ——, "On linear estimates, minimum variance, and least squares weighting matrices," *IEEE Trans. Automat. Contr.*, vol. AC-16, pp. 265–266, June 1971.
[27] ——, "An extension to gradient matrices," *IEEE Trans. Syst., Man., Cybernetics*, vol. SMC-1, pp. 184–186, Apr. 1971.
[28] ——, "Matrix calculus operations and Taylor expansions," *SIAM Rev.*, vol. 15, Apr. 1973.
[29] ——, "Vector structures and solutions of linear matrix equations," *Linear Algebra and Its Applications*, vol. 10, pp. 181–188, 1975.
[30] W. J. Vetter, J. G. Balchen, and F. Pohner, "Near optimal control of parameter dependent plants," *Int. J. Contr.*, vol. 15, 1972.
[31] J. Wilkinson, *The Algebraic Eigenvalue Problem.* Oxford, England: Oxford Univ. Press, 1965.
[32] L. A. Zadeh and C. A. Desoer, *Linear System Theory: The State Space Approach.* New York: McGraw-Hill, 1963.
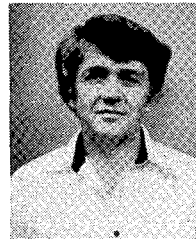
John W. Brewer received the Ph.D. degree in engineering science from the University of California at Berkeley, in 1966.

He joined the faculty of the University of California, Davis, in 1966, where he is currently Professor of Mechanical Engineering. His teaching and research activities concentrate in dynamic system theory, optimal estimation theory, and the theory of feedback control. He may have been one of the first system scientists to specialize in environmental and resource-economic systems when he published his first paper on these topics in 1962. Recent publications describe a bond graph analysis of microeconomic systems, a bond graph analysis of solar-regulated buildings, optimal impulsive control of agro-ecosystems, and optimal environmental estimation systems. His interest in matrix calculus is a result of his theoretical studies of the optimal placement of environmental monitors.

The Quotient Signal Flowgraph for Large-Scale Systems

MICHAEL K. SAIN, FELLOW, IEEE

Abstract—A system may have reached the large-scale condition either by displaying highly detailed element dynamics, or by exhibiting complicated connection patterns, or both. The quotient signal flowgraph (QSFG) is an approach to simplification of element dynamics. The principal feature of the QSFG concept is its stress upon making the simplification compatible with the connection structure. A context for the discussion is established on the generalized linear signal flowgraph (GLSFG), having node variables in an abelian group and flows determined by the morphisms of the group. A major class of examples is established, and an illustration is given from the applications literature.

I. INTRODUCTION

OUR PURPOSE here is to make a number of rather fundamental observations about large-scale interconnected systems. Intuitively speaking, a system can be considered large scale whenever a total accounting of all the connections and all the dynamics of all the elements is uneconomical for the purpose at hand. Such systems may have reached the large-scale condition either by displaying highly detailed element dynamics, or by exhibiting complicated connection patterns, or both. One may, accordingly, consider simplifications based upon either the element dynamics or the interconnections, depending upon the physical constraints of the application under consideration. For example, in large interconnected power grids, it may be an economical necessity in many cases to consider the connections as fixed; attention then turns naturally to element dynamics. A recent national workshop on power systems has made this point clear:

> A key factor in the consideration of reduced order models for power systems is the concept of structure of the system.... All of the dynamics are contained in individual subsystems.... It is important that this general structure be retained in the reduced order model so that the location of various parts of the system can be identified [1].

There are, of course, many ways to simplify dynamical elements. Surprisingly, however, little attention has been paid to the development of element simplifications that are compatible with system interconnections. This is the basic subject which we explore in the sequel.

In order to proceed precisely on this issue, we have to select a notion of an interconnected system upon which to develop the resulting concepts. For this purpose, we have chosen the signal flowgraph (SFG) of Mason [2], [3].

Briefly, an SFG is a weighted directed graph in which the nodes are variables and the edge weights are functions relating them. The flow across an edge is determined by applying the edge weighting function to the variable at the