Econometric Theory, **2**, 1986, 157–190. Printed in the United States of America.

SYMMETRY, 0-1 MATRICES AND JACOBIANS

A REVIEW

JAN R. MAGNUS

London School of Economics

H. NEUDECKER

University of Amsterdam

In this paper we bring together those properties of the Kronecker product, the vec operator, and 0-1 matrices which in our view are of interest to researchers and students in econometrics and statistics. The treatment of Kronecker products and the vec operator is fairly exhaustive; the treatment of 0-1 matrices is selective. In particular we study the "commutation" matrix $\mathbf{K}$ (defined implicitly by $\mathbf{K} \operatorname{vec} \mathbf{A} = \operatorname{vec} \mathbf{A}'$ for any matrix $\mathbf{A}$ of the appropriate order), the idempotent matrix $\mathbf{N} = \frac{1}{2}(\mathbf{I} + \mathbf{K})$, which plays a central role in normal distribution theory, and the "duplication" matrix $\mathbf{D}$, which arises in the context of symmetry. We present an easy and elegant way (via differentials) to evaluate Jacobian matrices (first derivatives), Hessian matrices (second derivatives), and Jacobian determinants, even if symmetric matrix arguments are involved. Finally we deal with the computation of information matrices in situations where positive definite matrices are arguments of the likelihood function.

1. INTRODUCTION

The purpose of this paper is to bring together those properties of the (simple) Kronecker product, the vec operator, and 0-1 matrices (commutation matrix, duplication matrix) that are thought to be of interest to researchers and students in econometrics and statistics. The treatment of Kronecker products and the vec operator is fairly exhaustive; the treatment of 0-1 matrices is (deliberately) selective.

The organization of the paper is as follows. In Sections 2 and 3 we review (and prove) the main results concerning the Kronecker product and the vec operator. The *commutation matrix* $\mathbf{K}_{mn}$ is introduced as the matrix which transforms $\operatorname{vec} \mathbf{A}$ into $\operatorname{vec} \mathbf{A}'$ for any $m \times n$ matrix $\mathbf{A}$. Its algebraic properties are discussed in Section 4. Closely related to the commutation matrix is the symmetric idempotent matrix $\mathbf{N}_n$ defined as $\mathbf{N}_n = \frac{1}{2}(\mathbf{I}_{n^2} + \mathbf{K}_{nn})$, whose main properties are obtained in Section 5 and whose role in normal distribution theory is discussed in Section 6. If $\mathbf{A}$ is a symmetric $n \times n$ matrix, its $\frac{1}{2}n(n-1)$ supradiagonal elements are redundant in the sense that they can be deduced from the symmetry. If we eliminate these redundant elements from vec $\mathbf{A}$, this defines a new vector which we denote as $v(\mathbf{A})$. The matrix which transforms, for symmetric $\mathbf{A}$, $v(\mathbf{A})$ into vec $\mathbf{A}$ is the *duplication matrix* $\mathbf{D_n}$. The duplication matrix plays an essential role in matrix differentiation involving symmetric matrices, and also in solving matrix equations where the solution matrix is known to be symmetric. Its most useful properties are given in Section 7. The class of symmetric matrices is the most important example of a much wider class of matrices: *L-structures*. An **L**-structure is the totality of real matrices of a specified order that satisfy a given set of linear restrictions. Other examples of **L**-structures are (strictly) triangular, skew-symmetric, diagonal, circulant, and Toeplitz matrices. In Section 8 the concept of an **L**-structure is defined and some of its properties discussed. In Section 9 we give what we claim to be the only viable definition of a matrix derivative (Jacobian matrix)—one which preserves the rank of the transformation and allows a useful chain rule. The Hessian matrix is defined in Section 10. Some examples show that the evaluation of Jacobian matrices and Hessian matrices can be short, elegant, and easy, even if the transformations involve symmetric (or **L**-structured) matrix arguments. Section 11 deals with the evaluation of information matrices in situations where positive definite matrices are arguments of the likelihood function, while Section 12 shows how 0-1 matrices can be used to evaluate certain Jacobian determinants.

The historical references in Sections 3 and 4 are taken from Henderson and Searle's [9] interesting survey.

The following *notation* is used. Matrices are denoted by capital letters, vectors and scalars by lower case letters. An $m \times n$ matrix is one having m rows and n columns; $\mathbf{A}'$ denotes the transpose of $\mathbf{A}$, $\mathbf{A}^+$ its Moore-Penrose inverse, and $r(\mathbf{A})$ its rank; if $\mathbf{A}$ is square, tr $\mathbf{A}$ denotes its trace, $|\mathbf{A}|$ its determinant, and $\mathbf{A}^{-1}$ its inverse (when $\mathbf{A}$ is nonsingular). $\mathbb{R}^{m \times n}$ is the class of real $m \times n$ matrices and $\mathbb{R}^n$ the class of real $n \times 1$ vectors, so that $\mathbb{R}^n \equiv \mathbb{R}^{n \times 1}$. The $n \times n$ identity matrix is denoted $\mathbf{I}_n$. Mathematical expectation is denoted by $\mathscr{E}$; variance (variance-covariance matrix) by $\mathscr{V}$.

2. THE KRONECKER PRODUCT

Let $\mathbf{A}$ be an $m \times n$ matrix and $\mathbf{B}$ a $p \times q$ matrix. The $mp \times nq$ matrix defined by

$$\begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{pmatrix} \tag{1}$$

is called the *Kronecker product* of $\mathbf{A}$ and $\mathbf{B}$ and written $\mathbf{A} \otimes \mathbf{B}$.

Observe that, while the matrix product $\mathbf{AB}$ only exists if the number of columns in $\mathbf{A}$ equals the number of rows in $\mathbf{B}$ or if either $\mathbf{A}$ or $\mathbf{B}$ is a scalar, the Kronecker product $\mathbf{A} \otimes \mathbf{B}$ is defined for any pair of matrices $\mathbf{A}$ and $\mathbf{B}$. The following three properties justify the name Kronecker *product:*

$$\mathbf{A} \otimes \mathbf{B} \otimes \mathbf{C} = (\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C}); \tag{2}$$

$$(\mathbf{A} + \mathbf{B}) \otimes (\mathbf{C} + \mathbf{D}) = \mathbf{A} \otimes \mathbf{C} + \mathbf{A} \otimes \mathbf{D} + \mathbf{B} \otimes \mathbf{C} + \mathbf{B} \otimes \mathbf{D}, \tag{3}$$

if $\mathbf{A} + \mathbf{B}$ and $\mathbf{C} + \mathbf{D}$ exist; and

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}, \tag{4}$$

if $\mathbf{AC}$ and $\mathbf{BD}$ exist.

If α is a scalar, then

$$\alpha \otimes \mathbf{A} = \alpha\mathbf{A} = \mathbf{A}\alpha = \mathbf{A} \otimes \alpha. \tag{5}$$

(This property can be used, for example, to prove that $(\mathbf{A} \otimes b)\mathbf{B} = (\mathbf{AB}) \otimes b$, by writing $\mathbf{B} = \mathbf{B} \otimes 1$.) Another useful property concerns two column-vectors $\mathbf{a}$ and $\mathbf{b}$ (not necessarily of the same order):

$$\mathbf{a}' \otimes \mathbf{b} = \mathbf{ba}' = \mathbf{b} \otimes \mathbf{a}'. \tag{6}$$

The transpose and the Moore-Penrose inverse of a Kronecker product are given by

$$(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}', \qquad (\mathbf{A} \otimes \mathbf{B})^+ = \mathbf{A}^+ \otimes \mathbf{B}^+. \tag{7}$$

If $\mathbf{A}$ and $\mathbf{B}$ are square matrices (not necessarily of the same order), then

$$\operatorname{tr}(\mathbf{A} \otimes \mathbf{B}) = (\operatorname{tr} \mathbf{A})(\operatorname{tr} \mathbf{B}). \tag{8}$$

Of course, the trace of $\mathbf{A} \otimes \mathbf{B}$ may exist even when $\mathbf{A}$ and $\mathbf{B}$ are not square matrices; in that case the expression for $\operatorname{tr}(\mathbf{A} \otimes \mathbf{B})$ is more complicated, see [26, Theorem 3.2]. If $\mathbf{A}$ and $\mathbf{B}$ are nonsingular, then

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}. \tag{9}$$

(The nonsingularity of $\mathbf{A}$ and $\mathbf{B}$ is not only sufficient, but also necessary for the nonsingularity of $\mathbf{A} \otimes \mathbf{B}$; this follows from rank considerations, see equation (11).)

All these properties are easy to prove. Let us now demonstrate the following result.

LEMMA 1. *Let* $\mathbf{A}$ *be an* $m \times m$ *matrix with eigenvalues* $\lambda_1, \lambda_2, \ldots, \lambda_m$, *and let* $\mathbf{B}$ *be a* $p \times p$ *matrix with eigenvalues* $\mu_1, \mu_2, \ldots, \mu_p$. *Then the mp eigenvalues of* $\mathbf{A} \otimes \mathbf{B}$ *are* $\lambda_i\mu_j$ $(i = 1, \ldots, m, j = 1, \ldots, p)$.

Proof. By Schur's Theorem (Bellman [4, p. 202]) there exist nonsingular (in fact, unitary) matrices $\mathbf{S}$ and $\mathbf{T}$ such that

$$\mathbf{S}^{-1}\mathbf{A}\mathbf{S} = \mathbf{L}, \qquad \mathbf{T}^{-1}\mathbf{B}\mathbf{T} = \mathbf{M},$$

where $\mathbf{L}$ and $\mathbf{M}$ are upper triangular matrices whose diagonal elements are the eigenvalues of $\mathbf{A}$ and $\mathbf{B}$ respectively. Thus,

$$(\mathbf{S}^{-1} \otimes \mathbf{T}^{-1})(\mathbf{A} \otimes \mathbf{B})(\mathbf{S} \otimes \mathbf{T}) = \mathbf{L} \otimes \mathbf{M}.$$

Since $\mathbf{S}^{-1} \otimes \mathbf{T}^{-1}$ is the inverse of $\mathbf{S} \otimes \mathbf{T}$, it follows that $\mathbf{A} \otimes \mathbf{B}$ and $(\mathbf{S}^{-1} \otimes \mathbf{T}^{-1})(\mathbf{A} \otimes \mathbf{B})(\mathbf{S} \otimes \mathbf{T})$ have the same set of eigenvalues, and hence that $\mathbf{A} \otimes \mathbf{B}$ and $\mathbf{L} \otimes \mathbf{M}$ have the same set of eigenvalues. But $\mathbf{L} \otimes \mathbf{M}$ is an upper triangular matrix by virtue of the fact that $\mathbf{L}$ and $\mathbf{M}$ are upper triangular; its eigenvalues are therefore its diagonal elements $\lambda_i\mu_j$. This concludes the proof. ∎

Remark. If $\mathbf{x}$ is an eigenvector of $\mathbf{A}$ and $\mathbf{y}$ an eigenvector of $\mathbf{B}$, then $\mathbf{x} \otimes \mathbf{y}$ is clearly an eigenvector of $\mathbf{A} \otimes \mathbf{B}$. It is *not* generally true, however, that every eigenvector of $\mathbf{A} \otimes \mathbf{B}$ is the Kronecker product of an eigenvector of $\mathbf{A}$ and an eigenvector of $\mathbf{B}$. We emphasize this fact because it is often stated incorrectly, see e.g., [4, p. 235]. For example, let

$$\mathbf{A} = \mathbf{B} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad \mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Both eigenvalues of $\mathbf{A}$ (and $\mathbf{B}$) are zero and the only eigenvector is $\mathbf{e}_1$. The four eigenvalues of $\mathbf{A} \otimes \mathbf{B}$ are all zero (in concordance with Lemma 1), but the eigenvectors of $\mathbf{A} \otimes \mathbf{B}$ are not just $\mathbf{e}_1 \otimes \mathbf{e}_1$, but also $\mathbf{e}_1 \otimes \mathbf{e}_2$ and $\mathbf{e}_2 \otimes \mathbf{e}_1$.

Lemma 1 has several important corollaries. First, if $\mathbf{A}$ and $\mathbf{B}$ are positive (semi)definite, then $\mathbf{A} \otimes \mathbf{B}$ is positive (semi)definite. Secondly, since the determinant of $\mathbf{A} \otimes \mathbf{B}$ is equal to the product of its eigenvalues, we obtain

$$|\mathbf{A} \otimes \mathbf{B}| = |\mathbf{A}|^p|\mathbf{B}|^m, \tag{10}$$

where $\mathbf{A}$ is an $m \times m$ matrix and $\mathbf{B}$ is a $p \times p$ matrix. Thirdly, we can obtain the rank of $\mathbf{A} \otimes \mathbf{B}$ from Lemma 1 as follows. The rank of $\mathbf{A} \otimes \mathbf{B}$ is equal to the rank of $\mathbf{A}\mathbf{A}' \otimes \mathbf{B}\mathbf{B}'$. The rank of the latter (symmetric, in fact positive semidefinite) matrix equals the number of nonzero (in this case positive) eigenvalues it possesses. According to Lemma 1, the eigenvalues of $\mathbf{A}\mathbf{A}' \otimes \mathbf{B}\mathbf{B}'$ are $\lambda_i\mu_j$, where λ_i are the eigenvalues of $\mathbf{A}\mathbf{A}'$ and μ_j are the eigenvalues of

$\mathbf{BB}'$. Now, $\lambda_i\mu_j$ is nonzero if and only if both λ_i and μ_j are nonzero. Hence, the number of nonzero eigenvalues of $\mathbf{AA}' \otimes \mathbf{BB}'$ is the product of the number of nonzero eigenvalues of $\mathbf{AA}'$ and the number of nonzero eigenvalues of $\mathbf{BB}'$. Thus the rank of $\mathbf{A} \otimes \mathbf{B}$ is

$$r(\mathbf{A} \otimes \mathbf{B}) = r(\mathbf{A})r(\mathbf{B}). \tag{11}$$

Historical note. The original interest in the Kronecker product focussed on the determinantal result (10), which seems to have been first studied by Zehfuss [43] in 1858. The result was known to Kronecker who passed it on to his students in Berlin, where he began lecturing in 1861 at the age of 37. The exact origin of the association of Kronecker's name with the $\otimes$ operation is still obscure. See MacDuffee [15, p. 81–84] for some early references.

3. THE VEC-OPERATOR

Let $\mathbf{A}$ be an $m \times n$ matrix and a_j its jth column, then vec $\mathbf{A}$ is the $mn \times 1$ vector

$$\operatorname{vec} \mathbf{A} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}. \tag{12}$$

Thus the vec-operator transforms a matrix into a vector by stacking the columns of the matrix one underneath the other. Notice that vec $\mathbf{A}$ is defined for *any* matrix $\mathbf{A}$, not just for square matrices. Also notice that vec $\mathbf{A}$ = vec $\mathbf{B}$ does not imply $\mathbf{A} = \mathbf{B}$, unless $\mathbf{A}$ and $\mathbf{B}$ are matrices of the same order.

A very simple but often useful property is

$$\operatorname{vec} \mathbf{a}' = \operatorname{vec} \mathbf{a} = \mathbf{a} \tag{13}$$

for any column-vector $\mathbf{a}$. The basic connection between the vec-operator and the Kronecker product is

$$\operatorname{vec} \mathbf{ab}' = \mathbf{b} \otimes \mathbf{a} \tag{14}$$

for any two column-vectors $\mathbf{a}$ and $\mathbf{b}$ (not necessarily of the same order). This follows because the jth column of $\mathbf{ab}'$ is $b_j\mathbf{a}$. Stacking the columns of $\mathbf{ab}'$ thus yields $\mathbf{b} \otimes \mathbf{a}$.

The basic connection between the vec-operator and the trace is

$$(\operatorname{vec} \mathbf{A})' \operatorname{vec} \mathbf{B} = \operatorname{tr} \mathbf{A}'\mathbf{B}, \tag{15}$$

where $\mathbf{A}$ and $\mathbf{B}$ are matrices of the same order. This is easy to verify since both the left side and the right side of equation (15) are equal to $\sum_i \sum_j a_{ij} b_{ij}$.

Let us now generalize the basic properties of equations (14) and (15). The generalization of (14) is the following well-known result.

LEMMA 2. *Let* $\mathbf{A}$, $\mathbf{B}$, *and* $\mathbf{C}$ *be three matrices such that the matrix product* $\mathbf{ABC}$ *is defined. Then,*

$$\operatorname{vec} \mathbf{ABC} = (\mathbf{C}' \otimes \mathbf{A}) \operatorname{vec} \mathbf{B}. \tag{16}$$

Proof. Assume that $\mathbf{B}$ has q columns denoted $b_1, b_2, \ldots, b_q$. Similarly let $e_1, e_2, \ldots, e_q$ denote the columns of the $q \times q$ identity matrix $\mathbf{I}_q$, so that $\mathbf{B} = \sum_{j=1}^{q} b_j e_j'$. Then, using equation (14),

$$\begin{aligned} \operatorname{vec} \mathbf{ABC} &= \operatorname{vec} \sum_{j=1}^{q} \mathbf{A} b_j e_j' \mathbf{C} = \sum_{j=1}^{q} \operatorname{vec} (\mathbf{A} b_j)(\mathbf{C}' e_j)' \\ &= \sum_{j=1}^{q} (\mathbf{C}' e_j \otimes \mathbf{A} b_j) = (\mathbf{C}' \otimes \mathbf{A}) \sum_{j=1}^{q} (e_j \otimes b_j) \\ &= (\mathbf{C}' \otimes \mathbf{A}) \sum_{j=1}^{q} \operatorname{vec} b_j e_j' = (\mathbf{C}' \otimes \mathbf{A}) \operatorname{vec} \mathbf{B}. \qquad \blacksquare \end{aligned}$$

One special case of Lemma 2 is

$$\operatorname{vec} \mathbf{AB} = (\mathbf{B}' \otimes \mathbf{I}_m) \operatorname{vec} \mathbf{A} = (\mathbf{B}' \otimes \mathbf{A}) \operatorname{vec} \mathbf{I}_n = (\mathbf{I}_q \otimes \mathbf{A}) \operatorname{vec} \mathbf{B}, \tag{17}$$

where $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{B}$ is an $n \times q$ matrix. Another special case arises when the matrix $\mathbf{C}$ in equation (16) is replaced by a vector. Then we obtain, using equation (13),

$$\mathbf{ABd} = (\mathbf{d}' \otimes \mathbf{A}) \operatorname{vec} \mathbf{B} = (\mathbf{A} \otimes \mathbf{d}') \operatorname{vec} \mathbf{B}', \tag{18}$$

where $\mathbf{d}$ is a $q \times 1$ vector.

The equality (15) can be generalized as follows.

LEMMA 3. *Let* $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, *and* $\mathbf{D}$ *be four matrices such that the matrix product* $\mathbf{ABCD}$ *is defined and square. Then,*

$$\operatorname{tr} \mathbf{ABCD} = (\operatorname{vec} \mathbf{D}')'(\mathbf{C}' \otimes \mathbf{A}) \operatorname{vec} \mathbf{B} = (\operatorname{vec} \mathbf{D})'(\mathbf{A} \otimes \mathbf{C}') \operatorname{vec} \mathbf{B}'. \tag{19}$$

Proof. We have

$$\begin{aligned} \operatorname{tr} \mathbf{ABCD} &= \operatorname{tr} \mathbf{D}(\mathbf{ABC}) = (\operatorname{vec} \mathbf{D}')' \operatorname{vec} \mathbf{ABC} && \text{(by (15))} \\ &= (\operatorname{vec} \mathbf{D}')'(\mathbf{C}' \otimes \mathbf{A}) \operatorname{vec} \mathbf{B} && \text{(by 16)).} \end{aligned}$$

The second equality is proved in precisely the same way starting from $\operatorname{tr} \mathbf{ABCD} = \operatorname{tr} \mathbf{D}'(\mathbf{C}'\mathbf{B}'\mathbf{A}')$. ∎

Historical note. The idea of stacking the elements of a matrix in a vector goes back at least to Sylvester [36, 37]. The notation "vec" was introduced by Koopmans, Rubin, and Leipnik [13]. Lemma 2 is due to Roth [35].

4. THE COMMUTATION MATRIX $\mathbf{K}_{mn}$

Let $\mathbf{A}$ be an $m \times n$ matrix. The vectors $\operatorname{vec} \mathbf{A}$ and $\operatorname{vec} \mathbf{A}'$ clearly contain the same mn components, but in a different order. Hence there exists a unique $mn \times mn$ permutation matrix which transforms $\operatorname{vec} \mathbf{A}$ into $\operatorname{vec} \mathbf{A}'$. This matrix is called the *commutation matrix* and is denoted $\mathbf{K}_{mn}$. Thus

$$\mathbf{K}_{mn} \operatorname{vec} \mathbf{A} = \operatorname{vec} \mathbf{A}'. \tag{20}$$

Since $\mathbf{K}_{mn}$ is a permutation matrix it is orthogonal, i.e., $\mathbf{K}'_{mn} = \mathbf{K}_{mn}^{-1}$. Also, premultiplying equation (20) by $\mathbf{K}_{nm}$ gives $\mathbf{K}_{nm}\mathbf{K}_{mn} \operatorname{vec} \mathbf{A} = \operatorname{vec} \mathbf{A}$ so that $\mathbf{K}_{nm}\mathbf{K}_{mn} = \mathbf{I}_{mn}$. Hence,

$$\mathbf{K}'_{mn} = \mathbf{K}_{mn}^{-1} = \mathbf{K}_{nm}. \tag{21}$$

Further, using equation (13),

$$\mathbf{K}_{n1} = \mathbf{K}_{1n} = \mathbf{I}_n. \tag{22}$$

The key property of the commutation matrix (and the one from which it derives its name) enables us to interchange ("commute") the two matrices of a Kronecker product.

LEMMA 4. *Let* $\mathbf{A}$ *be an* m × n *matrix and* $\mathbf{B}$ *a* p × q *matrix. Then*

$$\mathbf{K}_{pm}(\mathbf{A} \otimes \mathbf{B}) = (\mathbf{B} \otimes \mathbf{A})\mathbf{K}_{qn}. \tag{23}$$

Proof. Let $\mathbf{X}$ be an arbitrary $q \times n$ matrix. Then, by repeated application of equations (16) and (20),

$$\begin{aligned}\mathbf{K}_{pm}(\mathbf{A} \otimes \mathbf{B}) \operatorname{vec} \mathbf{X} &= \mathbf{K}_{pm} \operatorname{vec} \mathbf{BXA}' = \operatorname{vec} \mathbf{AX}'\mathbf{B}' \\ &= (\mathbf{B} \otimes \mathbf{A}) \operatorname{vec} \mathbf{X}' = (\mathbf{B} \otimes \mathbf{A})\mathbf{K}_{qn} \operatorname{vec} \mathbf{X}.\end{aligned}$$

Since $\mathbf{X}$ is arbitrary the result follows. ∎

Immediate consequences of Lemma 4 are

$$\mathbf{K}_{pm}(\mathbf{A} \otimes \mathbf{B})\mathbf{K}_{nq} = \mathbf{B} \otimes \mathbf{A} \tag{24}$$

and

$$\mathbf{K}_{pm}(\mathbf{A} \otimes \mathbf{b}) = \mathbf{b} \otimes \mathbf{A}, \qquad \mathbf{K}_{mp}(\mathbf{b} \otimes \mathbf{A}) = \mathbf{A} \otimes \mathbf{b}, \tag{25}$$

where $\mathbf{b}$ is a $p \times 1$ vector.

All these properties follow from the implicit definition (20) of the commutation matrix. The following lemma gives an *explicit* expression for $\mathbf{K}_{mn}$ which is often useful.

LEMMA 5. *Let* $\mathbf{H}_{ij}$ *be the* $m \times n$ *matrix with* 1 *in its* ij*th position and zeroes elsewhere. Then*

$$\mathbf{K}_{mn} = \sum_{i=1}^{m} \sum_{j=1}^{n} (\mathbf{H}_{ij} \otimes \mathbf{H}'_{ij}). \tag{26}$$

Proof. Let $\mathbf{X}$ be an arbitrary $m \times n$ matrix. Let e_i denote the ith column of $\mathbf{I}_m$ and u_j the jth column of $\mathbf{I}_n$, so that $\mathbf{H}_{ij} = e_i u'_j$. Then

$$\begin{aligned}
\mathbf{X}' = \mathbf{I}_n \mathbf{X}' \mathbf{I}_m &= \left(\sum_{j=1}^{n} u_j u'_j \right) \mathbf{X}' \left(\sum_{i=1}^{m} e_i e'_i \right) \\
&= \sum_{ij} u_j (u'_j \mathbf{X}' e_i) e'_i = \sum_{ij} u_j (e'_i \mathbf{X} u_j) e'_i \\
&= \sum_{ij} (u_j e'_i) \mathbf{X} (u_j e'_i) = \sum_{ij} \mathbf{H}'_{ij} \mathbf{X} \mathbf{H}'_{ij}.
\end{aligned}$$

Taking vecs we obtain

$$\operatorname{vec} \mathbf{X}' = \sum_{ij} \operatorname{vec} \mathbf{H}'_{ij} \mathbf{X} \mathbf{H}'_{ij} = \sum_{ij} (\mathbf{H}_{ij} \otimes \mathbf{H}'_{ij}) \operatorname{vec} \mathbf{X},$$

using equation (16). The result follows. ∎

Lemma 5 shows that $\mathbf{K}_{mn}$ is a square matrix of order mn, partitioned into mn submatrices each of order $n \times m$, such that the ijth submatrix has unity in its jith position and zeros elsewhere. For example,

$$\mathbf{K}_{23} = \left(\begin{array}{cc|cc|cc}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
\hline
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{array} \right). \tag{27}$$

The explicit form (26) of $\mathbf{K}_{mn}$ enables us to find the trace and the determinant of $\mathbf{K}_{mn}$.

LEMMA 6. *The trace of the commutation matrix is*

$$\operatorname{tr} \mathbf{K}_{mn} = 1 + \gcd(\mathrm{m} - 1, \mathrm{n} - 1), \tag{28}$$

where $\gcd(\mathrm{m}, \mathrm{n})$ *is the greatest common divisor of* m *and* n*; its determinant is*

$$|\mathbf{K}_{mn}| = (-1)^{(1/4)mn(m-1)(n-1)}. \tag{29}$$

Proof. We shall only prove the case $m = n$. (For a proof of the more difficult case $m \neq n$, see Magnus and Neudecker [17, Theorem 3.1].) Let e_j be the jth column of I_n. Then, from equation (26).

$$\begin{aligned}\operatorname{tr} \mathbf{K}_{nn} &= \operatorname{tr} \sum_{i=1}^{n} \sum_{j=1}^{n} (e_i e_j' \otimes e_j e_i') = \sum_{ij} \operatorname{tr}(e_i e_j' \otimes e_j e_i') \\ &= \sum_{ij} (\operatorname{tr} e_i e_j')(\operatorname{tr} e_j e_i') = \sum_{ij} \delta_{ij}^2 = n,\end{aligned}$$

where $\delta_{ij} = 0$ if $i \neq j$, $\delta_{ii} = 1$. Since $\mathbf{K}_{nn}$ is real, orthogonal and symmetric, it has eigenvalues $+1$ and -1 only. (The eigenvalues of $\mathbf{K}_{mn}$, $m \neq n$, are, in general, complex.) Suppose the multiplicity of -1 is p. Then the multiplicity of $+1$ is $(n^2 - p)$, and

$$n = \operatorname{tr} \mathbf{K}_{nn} = \text{sum of eigenvalues of } \mathbf{K}_{nn} = -p + n^2 - p = n^2 - 2p,$$

so that $p = \frac{1}{2}n(n-1)$. Hence

$$|\mathbf{K}_{nn}| = (-1)^p = (-1)^{(1/2)n(n-1)} = (-1)^{(1/4)n^2(n-1)^2}. \qquad \blacksquare$$

An important application of the commutation matrix is that it allows us to transform the vec of a Kronecker product into the Kronecker product of the vecs, a crucial property in the differentiation of Kronecker products.

LEMMA 7. *Let* $\mathbf{A}$ *be an* $\mathrm{m} \times \mathrm{n}$ *matrix and* $\mathbf{B}$ *a* $\mathrm{p} \times \mathrm{q}$ *matrix. Then*

$$\operatorname{vec}(\mathbf{A} \otimes \mathbf{B}) = (\mathbf{I}_n \otimes \mathbf{K}_{qm} \otimes \mathbf{I}_p)(\operatorname{vec} \mathbf{A} \otimes \operatorname{vec} \mathbf{B}). \tag{30}$$

Proof. Let a_i $(i = 1, \ldots, n)$ and b_j $(j = 1, \ldots, q)$ denote the columns of $\mathbf{A}$ and $\mathbf{B}$, respectively. Also, let e_i $(i = 1, \ldots, n)$ and u_j $(j = 1, \ldots, q)$ denote the columns of $\mathbf{I}_n$ and $\mathbf{I}_q$, respectively. Then we can write $\mathbf{A}$ and $\mathbf{B}$ as

$$\mathbf{A} = \sum_{i=1}^{n} a_i e_{i'}', \qquad \mathbf{B} = \sum_{j=1}^{q} b_j u_{j'}',$$

and we obtain

$$\begin{aligned}
\operatorname{vec}(\mathbf{A} \otimes \mathbf{B}) &= \sum_{i=1}^{n} \sum_{j=1}^{q} \operatorname{vec}(a_i e_i' \otimes b_j u_j') \\
&= \sum_{ij} \operatorname{vec}(a_i \otimes b_j)(e_i \otimes u_j)' = \sum_{ij} (e_i \otimes u_j \otimes a_i \otimes b_j) \\
&= \sum_{ij} (\mathbf{I}_n \otimes \mathbf{K}_{qm} \otimes \mathbf{I}_p)(e_i \otimes a_i \otimes u_j \otimes b_j) \\
&= (\mathbf{I}_n \otimes \mathbf{K}_{qm} \otimes \mathbf{I}_p)\left\{\left(\sum_i \operatorname{vec} a_i e_i'\right) \otimes \left(\sum_j \operatorname{vec} b_j u_j'\right)\right\} \\
&= (\mathbf{I}_n \otimes \mathbf{K}_{qm} \otimes \mathbf{I}_p)(\operatorname{vec} \mathbf{A} \otimes \operatorname{vec} \mathbf{B}). \qquad \blacksquare
\end{aligned}$$

In particular, by noting that

$$\operatorname{vec} \mathbf{A} \otimes \operatorname{vec} \mathbf{B} = (\mathbf{I}_{nm} \otimes \operatorname{vec} \mathbf{B}) \operatorname{vec} \mathbf{A} = (\operatorname{vec} \mathbf{A} \otimes \mathbf{I}_{qp}) \operatorname{vec} \mathbf{B},$$

using equation (5), we obtain

$$\operatorname{vec}(\mathbf{A} \otimes \mathbf{B}) = (\mathbf{I}_n \otimes \mathbf{G}) \operatorname{vec} \mathbf{A} = (\mathbf{H} \otimes \mathbf{I}_p) \operatorname{vec} \mathbf{B}, \tag{31}$$

where

$$\mathbf{G} = (\mathbf{K}_{qm} \otimes \mathbf{I}_p)(\mathbf{I}_m \otimes \operatorname{vec} \mathbf{B}), \qquad \mathbf{H} = (\mathbf{I}_n \otimes \mathbf{K}_{qm})(\operatorname{vec} \mathbf{A} \otimes \mathbf{I}_q). \tag{32}$$

Historical note. The original interest in the commutation matrix focussed on its role in reversing ("commuting") the order of Kronecker products (Lemma 4), a role which seems to have been first recognized by Ledermann [14] and Murnaghan [21, pp. 68–69] while Vartak [40] generalized Murnaghan's result to rectangular matrices. Tracy and Dwyer [38] rediscovered the commutation matrix and based their definition on the fact that $\mathbf{K}_{mn}$ is the matrix obtained by rearranging the rows of $\mathbf{I}_{mn}$ by taking every mth row starting with the first, then every mth row starting with the second, and so on. (For example, the rows of $\mathbf{K}_{23}$ are rows 1, 3, 5, 2, 4, and 6 of I_6.) The fruitful idea of defining $\mathbf{K}_{mn}$ by its transformation property (20) comes from Barnett [3], and is the definition adopted in this paper. See also Pollock [31, p. 72–73].

Among the many alternative names of the commutation matrix we mention permutation matrix, permuted identity matrix, vec-permutation matrix, shuffle matrix, tensor commutator, and universal flip matrix. Alternative notations for $\mathbf{K}_{mn}$ include $\mathbf{E}_{m,n}$, $\mathbf{E}_{m\times n}^{n\times m}$, $\mathbf{U}_{m\times n}$, $\mathbf{P}_{n,m}$, $\mathbf{I}_{(n,m)}$, $\mathbf{I}_{n,m}$, and Ⓣ.

Lemma 4 goes back at least to Ledermann [14]. Concise proofs are given by Barnett [3], Hartwig and Morris [8], and Magnus and Neudecker [17]. Lemmas 5 and 6 are due to Magnus and Neudecker [17], and Lemma 7 to Neudecker and Wansbeek [26]. For a discussion of the characteristic polynomial of the commutation matrix, see [8] and [7].

For further reading on the commutation matrix we recommend Hartwig and Morris [8] , Balestra [2], Magnus and Neudecker [17], Henderson and Searle [10], and Neudecker and Wansbeek [26].

5. THE MATRIX $\mathbf{N}_n$

Closely related to the commutation matrix is the $n^2 \times n^2$ matrix

$$\mathbf{N}_n = \tfrac{1}{2}(\mathbf{I}_{n^2} + \mathbf{K}_{nn}). \tag{33}$$

This matrix is symmetric and idempotent,

$$\mathbf{N}_n = \mathbf{N}_n' = \mathbf{N}_n^2, \tag{34}$$

and, since tr $\mathbf{K}_{nn} = n$, its trace (and hence its rank) is easily shown to be

$$r(\mathbf{N}_n) = \operatorname{tr} \mathbf{N}_n = \tfrac{1}{2}n(n+1). \tag{35}$$

The matrix $\mathbf{N}_n$ transforms an arbitrary $n \times n$ matrix $\mathbf{A}$ into the *symmetric* matrix $\frac{1}{2}(\mathbf{A} + \mathbf{A}')$:

$$\mathbf{N}_n \operatorname{vec} \mathbf{A} = \operatorname{vec} \tfrac{1}{2}(\mathbf{A} + \mathbf{A}'). \tag{36}$$

Of course, if $\mathbf{A}$ is symmetric to begin with, the transformation has no effect. (This shows again that $\mathbf{N}_n$ must be idempotent.)

Further properties of $\mathbf{N}_n$ include

$$\mathbf{N}_n\mathbf{K}_{nn} = \mathbf{N}_n = \mathbf{K}_{nn}\mathbf{N}_n, \tag{37}$$

and, for any two $n \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$,

$$\mathbf{N}_n(\mathbf{A} \otimes \mathbf{B})\mathbf{N}_n = \mathbf{N}_n(\mathbf{B} \otimes \mathbf{A})\mathbf{N}_n, \tag{38}$$

$$\mathbf{N}_n(\mathbf{A} \otimes \mathbf{B} + \mathbf{B} \otimes \mathbf{A})\mathbf{N}_n = \mathbf{N}_n(\mathbf{A} \otimes \mathbf{B} + \mathbf{B} \otimes \mathbf{A}) = (\mathbf{A} \otimes \mathbf{B} + \mathbf{B} \otimes \mathbf{A})\mathbf{N}_n, \tag{39}$$

$$\mathbf{N}_n(\mathbf{A} \otimes \mathbf{A})\mathbf{N}_n = \mathbf{N}_n(\mathbf{A} \otimes \mathbf{A}) = (\mathbf{A} \otimes \mathbf{A})\mathbf{N}_n, \tag{40}$$

and

$$\mathbf{N}_n(\mathbf{A} \otimes \mathbf{b}) = \mathbf{N}_n(\mathbf{b} \otimes \mathbf{A}) = \tfrac{1}{2}(\mathbf{A} \otimes \mathbf{b} + \mathbf{b} \otimes \mathbf{A}), \tag{41}$$

for any $n \times 1$ vector $\mathbf{b}$.

The explicit form of $\mathbf{N}_n$ is easily derived from $\mathbf{K}_{nn}$. For example, for $n = 2$ and 3, we have

$$\mathbf{N}_2 = \left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \hline 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{array}\right), \qquad \mathbf{N}_3 = \left(\begin{array}{ccc|ccc|ccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ \hline 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \hline 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right). \tag{42}$$

Historical note. The matrix $\mathbf{N}_n$ was introduced by Magnus and Neudecker [18, p. 424].

6. THE COMMUTATION MATRIX AND THE WISHART DISTRIBUTION

Somewhat unexpectedly, the commutation matrix also plays a role in distribution theory, especially in *normal* distribution theory. This role is based on the following result.

LEMMA 8. *Let u be an $n \times 1$ vector of independent and standard normally distributed random variables $u_1, \ldots, u_n$, that is, $u \cong \mathcal{N}(\mathbf{O}, \mathbf{I}_n)$. Then*

$$\mathcal{V}(u \otimes u) = 2\mathbf{N}_n, \tag{43}$$

Note. In the scalar case $n = 1$ we find $\mathcal{V}u^2 = 2$. Formula (43) gives a natural generalization of this result.

Proof. Let $\mathbf{A}$ be an arbitrary $n \times n$ matrix and let $\mathbf{B} = (\mathbf{A} + \mathbf{A}')/2$. Let $\mathbf{T}$ be an orthogonal $n \times n$ matrix such that $\mathbf{T}'\mathbf{B}\mathbf{T} = \mathbf{\Lambda}$, where $\mathbf{\Lambda}$ is the diagonal matrix whose diagonal elements $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $\mathbf{B}$. Let $v = \mathbf{T}'u$ with components $v_1, \ldots, v_n$. Then,

$$u'\mathbf{A}u = u'\mathbf{B}u = u'\mathbf{T}\mathbf{\Lambda}\mathbf{T}'u = v'\mathbf{\Lambda}v = \sum_{i=1}^{n} \lambda_i v_i^2.$$

Since $v \cong \mathcal{N}(\mathbf{O}, \mathbf{I}_n)$, it follows that $v_1^2, \ldots, v_n^2$ are independently distributed with $\mathscr{V}(v_i^2) = 2$, so that

$$\begin{aligned}\mathscr{V}(u'\mathbf{A}u) &= \mathscr{V}\left(\sum_i \lambda_i v_i^2\right) = \sum_i \lambda_i^2(\mathscr{V} v_i^2) = 2 \operatorname{tr} \mathbf{\Lambda}^2 \\ &= 2 \operatorname{tr} \mathbf{B}^2 = \operatorname{tr} \mathbf{A}'\mathbf{A} + \operatorname{tr} \mathbf{A}^2 = (\operatorname{vec} \mathbf{A})'(\mathbf{I} + \mathbf{K}_{nn}) \operatorname{vec} \mathbf{A},\end{aligned}$$

using equation (15). Also, since $u'\mathbf{A}u = \operatorname{vec} u'\mathbf{A}u = (u \otimes u)' \operatorname{vec} \mathbf{A}$,

$$\mathscr{V}(u'\mathbf{A}u) = \mathscr{V}((u \otimes u)' \operatorname{vec} \mathbf{A}) = (\operatorname{vec} \mathbf{A})'(\mathscr{V} u \otimes u) \operatorname{vec} \mathbf{A}.$$

Hence,

$$(\operatorname{vec} \mathbf{A})'(\mathscr{V} u \otimes u) \operatorname{vec} \mathbf{A} = (\operatorname{vec} \mathbf{A})'(\mathbf{I} + \mathbf{K}_{nn}) \operatorname{vec} \mathbf{A}$$

for every $n \times n$ matrix $\mathbf{A}$. The result follows. ∎

We can generalize Lemma 8 by considering normal random variables which are not necessarily independent or identically distributed. This leads to Lemma 9.

LEMMA 9. *Let $x \cong \mathcal{N}(\mu, \mathbf{V})$ where $\mathbf{V}$ is a positive semidefinite* $n \times n$ *matrix. Then*

$$\mathscr{V}(x \otimes x) = 2\mathbf{N}_n(\mathbf{V} \otimes \mathbf{V} + \mathbf{V} \otimes \mu\mu' + \mu\mu' \otimes \mathbf{V}). \tag{44}$$

Proof. We write $x = \mathbf{V}^{1/2}u + \mu$ with $u \cong \mathcal{N}(\mathbf{O}, \mathbf{I}_n)$, so that

$$\begin{aligned}x \otimes x &= \mathbf{V}^{1/2}u \otimes \mathbf{V}^{1/2}u + \mathbf{V}^{1/2}u \otimes \mu + \mu \otimes \mathbf{V}^{1/2}u + \mu \otimes \mu \\ &= (\mathbf{V}^{1/2} \otimes \mathbf{V}^{1/2})(u \otimes u) + (\mathbf{I} + \mathbf{K}_{nn})(\mathbf{V}^{1/2}u \otimes \mu) + \mu \otimes \mu \\ &= (\mathbf{V}^{1/2} \otimes \mathbf{V}^{1/2})(u \otimes u) + (\mathbf{I} + \mathbf{K}_{nn})(\mathbf{V}^{1/2} \otimes \mu)u + \mu \otimes \mu,\end{aligned}$$

using equations (25) and (5). Since the two vectors $u \otimes u$ and u are uncorrelated with $\mathscr{V}(u \otimes u) = \mathbf{I} + \mathbf{K}_{nn}$ and $\mathscr{V}(u) = \mathbf{I}_n$, we obtain

$$\begin{aligned}\mathscr{V}(x \otimes x) &= \mathscr{V}\{(\mathbf{V}^{1/2} \otimes \mathbf{V}^{1/2})(u \otimes u)\} + \mathscr{V}\{(\mathbf{I} + \mathbf{K}_{nn})(\mathbf{V}^{1/2} \otimes \mu)u\} \\ &= (\mathbf{V}^{1/2} \otimes \mathbf{V}^{1/2})(\mathbf{I} + \mathbf{K}_{nn})(\mathbf{V}^{1/2} \otimes \mathbf{V}^{1/2}) \\ &\quad + (\mathbf{I} + \mathbf{K}_{nn})(\mathbf{V}^{1/2} \otimes \mu)(\mathbf{V}^{1/2} \otimes \mu)'(\mathbf{I} + \mathbf{K}_{nn}) \\ &= (\mathbf{I} + \mathbf{K}_{nn})(\mathbf{V} \otimes \mathbf{V}) + (\mathbf{I} + \mathbf{K}_{nn})\{\mathbf{V} \otimes \mu\mu' + \mathbf{K}_{nn}(\mu\mu' \otimes \mathbf{V})\} \\ &= (\mathbf{I} + \mathbf{K}_{nn})(\mathbf{V} \otimes \mathbf{V} + \mathbf{V} \otimes \mu\mu' + \mu\mu' \otimes \mathbf{V}),\end{aligned}$$

using equation (23) and the fact (implied by (37)) that $(\mathbf{I} + \mathbf{K}_{nn})\mathbf{K}_{nn} = \mathbf{I} + \mathbf{K}_{nn}$. ∎

Let us now consider k random $n \times 1$ vectors $y_1, \ldots, y_k$, distributed independently as

$$y_i \cong \mathcal{N}(\mu_i, \mathbf{V}), \qquad (i = 1, \ldots, k).$$

The joint distribution of the elements of the matrix

$$\mathbf{S} = \sum_{i=1}^{k} y_i y_i'$$

is said to be Wishart with k degrees of freedom and is denoted by $\mathbf{W}_n(k, \mathbf{V}, \mathbf{M})$, where $\mathbf{M}$ is the $k \times n$ matrix

$$\mathbf{M} = (\mu_1, \mu_2, \ldots, \mu_k)'.$$

(If $\mathbf{M} = 0$ the distribution is said to be central.) The following lemma gives the mean and variance of the (noncentral) Wishart distribution in a compact and readily usable form.

LEMMA 10. *Let* $\mathbf{S}$ *be Wishart distributed* $\mathbf{W}_n(k, \mathbf{V}, \mathbf{M})$, $\mathbf{V}$ *positive semidefinite. Then,*

$$\mathscr{E}\mathbf{S} = k\mathbf{V} + \mathbf{M}'\mathbf{M} \tag{45}$$

and

$$\mathscr{V} \operatorname{vec} \mathbf{S} = 2\mathbf{N}_n\{k(\mathbf{V} \otimes \mathbf{V}) + \mathbf{V} \otimes \mathbf{M}'\mathbf{M} + \mathbf{M}'\mathbf{M} \otimes \mathbf{V}\}. \tag{46}$$

Proof. We first note that $\sum_{i=1}^{k} \mu_i \mu_i' = \mathbf{M}'\mathbf{M}$. Then

$$\mathscr{E}\mathbf{S} = \mathscr{E} \sum_i y_i y_i' = \sum_i \mathscr{E} y_i y_i' = \sum_i (\mathbf{V} + \mu_i \mu_i') = k\mathbf{V} + \mathbf{M}'\mathbf{M},$$

and

$$\begin{aligned}
\mathscr{V} \operatorname{vec} \mathbf{S} &= \mathscr{V}\left(\operatorname{vec} \sum_i y_i y_i'\right) = \mathscr{V}\left(\sum_i y_i \otimes y_i\right) = \sum_i \mathscr{V}(y_i \otimes y_i) \\
&= \sum_i (\mathbf{I} + \mathbf{K}_{nn})(\mathbf{V} \otimes \mathbf{V} + \mathbf{V} \otimes \mu_i \mu_i' + \mu_i \mu_i' \otimes \mathbf{V}) \\
&= (\mathbf{I} + \mathbf{K}_{nn})\{k(\mathbf{V} \otimes \mathbf{V}) + \mathbf{V} \otimes \mathbf{M}'\mathbf{M} + \mathbf{M}'\mathbf{M} \otimes \mathbf{V}\},
\end{aligned}$$

using equation (44). ∎

We note in passing the following relationship between the covariance matrices of a noncentral Wishart matrix and its central counterpart. Let $\mathbf{S}_1$ be Wishart distributed $\mathbf{W}_n(k, \mathbf{V}, \mathbf{M})$ and let $\mathbf{S}_0$ be Wishart distributed $\mathbf{W}_n(k, \mathbf{V}, 0)$. Then, using equation (46),

$$\mathscr{V} \operatorname{vec} \mathbf{S}_0 \leqslant \mathscr{V} \operatorname{vec} \mathbf{S}_1 \tag{47}$$

in the sense that the difference between the two covariance matrices of $\operatorname{vec} \mathbf{S}_0$ and $\operatorname{vec} \mathbf{S}_1$ is negative semidefinite. For a generalization of this result, see Problem 86.2.4 in this issue.

Historical note. The results in this section are taken from Magnus and Neudecker [17], but the proofs are somewhat simplified. More general results can be found in [17] and [26]. In the latter paper it is shown, inter alia, that the normality assumption in Lemma 8 is not essential. More precisely, if u is an $n \times 1$ vector of independent random variables $u_1, \ldots, u_n$ with $\mathscr{E} u_i = 0$, $\mathscr{E} u_i^2 = \sigma^2$, $\mathscr{E} u_i^4 = \psi^4$, then

$$\mathscr{V}(u \otimes u) = \sigma^4\left(2\mathbf{N}_n + \gamma \sum_{i=1}^{n} (\mathbf{E}_{ii} \otimes \mathbf{E}_{ii})\right) \tag{48}$$

where $\gamma = (\psi^4/\sigma^4) - 3$ (the *kurtosis*) and $\mathbf{E}_{ii}$ is the $n \times n$ matrix with 1 in its ith diagonal position and zeros elsewhere.

7. SYMMETRY: THE DUPLICATION MATRIX $\mathbf{D}_n$

Let $\mathbf{A}$ be a square $n \times n$ matrix. Then $v(\mathbf{A})$ will denote the $\frac{1}{2}n(n+1) \times 1$ vector that is obtained from $\operatorname{vec} \mathbf{A}$ by eliminating all supradiagonal elements of $\mathbf{A}$. For example, if $n = 3$,

$$\operatorname{vec} \mathbf{A} = (a_{11} a_{21} a_{31} a_{12} a_{22} a_{32} a_{13} a_{23} a_{33})',$$

and

$$v(\mathbf{A}) = (a_{11} a_{21} a_{31} a_{22} a_{32} a_{33})'.$$

In this way, for symmetric $\mathbf{A}$, $v(\mathbf{A})$ contains only the distinct elements of $\mathbf{A}$. Since the elements of $\operatorname{vec} \mathbf{A}$ are those of $v(\mathbf{A})$ with some repetitions, there exists a unique $n^2 \times \frac{1}{2}n(n+1)$ matrix which transforms, for symmetric $\mathbf{A}$, $v(\mathbf{A})$ into $\operatorname{vec} \mathbf{A}$. This matrix is called the *duplication matrix* and is denoted $\mathbf{D}_n$. Thus,

$$\mathbf{D}_n v(\mathbf{A}) = \operatorname{vec} \mathbf{A} \qquad (\mathbf{A} = \mathbf{A}'). \tag{49}$$

Let $\mathbf{A} = \mathbf{A}'$ and $\mathbf{D}_n v(\mathbf{A}) = 0$. Then vec $\mathbf{A} = 0$, and so $v(\mathbf{A}) = 0$. Since the symmetry of $\mathbf{A}$ does not restrict $v(\mathbf{A})$, it follows that the columns of $\mathbf{D}_n$ are linearly independent. Hence $\mathbf{D}_n$ has full column-rank $\frac{1}{2}n(n+1)$, $\mathbf{D}_n'\mathbf{D}_n$ is nonsingular, and $\mathbf{D}_n^+$, the Moore-Penrose inverse of $\mathbf{D}_n$, equals

$$\mathbf{D}_n^+ = (\mathbf{D}_n'\mathbf{D}_n)^{-1}\mathbf{D}_n'. \tag{50}$$

For $n = 3$, we have

$$\mathbf{D}_3 = \left(\begin{array}{ccc|cc|c} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right), \qquad \mathbf{D}_3^{+\prime} = \left(\begin{array}{ccc|cc|c} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ \hline 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \\ \hline 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right). \tag{51}$$

Some further properties of $\mathbf{D}_n$ are easily derived from its implicit definition (49). For symmetric $\mathbf{A}$ we have

$$\mathbf{K}_{nn}\mathbf{D}_n v(\mathbf{A}) = \mathbf{K}_{nn} \operatorname{vec} \mathbf{A} = \operatorname{vec} \mathbf{A} = \mathbf{D}_n v(\mathbf{A})$$

and

$$\mathbf{N}_n\mathbf{D}_n v(\mathbf{A}) = \mathbf{N}_n \operatorname{vec} \mathbf{A} = \operatorname{vec} \mathbf{A} = \mathbf{D}_n v(\mathbf{A}).$$

Again, the symmetry of $\mathbf{A}$ does not restrict $v(\mathbf{A})$, so that

$$\mathbf{K}_{nn}\mathbf{D}_n = \mathbf{D}_n = \mathbf{N}_n\mathbf{D}_n. \tag{52}$$

Also, from equations (50) and (52),

$$\mathbf{D}_n^+\mathbf{K}_{nn} = \mathbf{D}_n^+ = \mathbf{D}_n^+\mathbf{N}_n, \tag{53}$$

and

$$\mathbf{D}_n^+\mathbf{D}_n = \mathbf{I}_{(1/2)n(n+1)}, \qquad \mathbf{D}_n\mathbf{D}_n^+ = \mathbf{N}_n. \tag{54}$$

The first of the two equalities in equation (54) is an immediate consequence of (50), while the second follows from $\mathbf{N}_n\mathbf{D}_n = \mathbf{D}_n$ (see (52)).[1] We see that

$\mathbf{N}_n$ is just the projection matrix $\mathbf{D}_n(\mathbf{D}_n'\mathbf{D}_n)^{-1}\mathbf{D}_n'$; its action is to project onto the column space of $\mathbf{D}_n$, i.e., the linear subspace in $\mathbb{R}^{n^2}$ representing the set of all symmetric $n \times n$ matrices.

Finally we obtain

$$\mathbf{D}_n^+ \operatorname{vec} \mathbf{A} = \tfrac{1}{2}v(\mathbf{A} + \mathbf{A}') \tag{55}$$

for any $n \times n$ matrix $\mathbf{A}$. This follows by letting $\mathbf{B} = \frac{1}{2}(\mathbf{A} + \mathbf{A}')$ and observing that

$$\mathbf{D}_n^+ \operatorname{vec} \mathbf{A} = \mathbf{D}_n^+\mathbf{N}_n \operatorname{vec} \mathbf{A} = \mathbf{D}_n^+ \operatorname{vec} \mathbf{B} = \mathbf{D}_n^+\mathbf{D}_n v(\mathbf{B}) = v(\mathbf{B})$$

using equations (53), (36), (49), and (54).

Much of the interest in the duplication matrix is due to the importance of the matrices $\mathbf{D}_n^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n$ and $\mathbf{D}_n'(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n$, whose properties we shall now investigate. We first prove Lemma 11.

LEMMA 11. *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix. Then,*

$$\mathbf{D}_n\mathbf{D}_n^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n = (\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n, \tag{56}$$

$$\mathbf{D}_n^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n\mathbf{D}_n^+ = \mathbf{D}_n^+(\mathbf{A} \otimes \mathbf{A}), \tag{57}$$

and, if $\mathbf{A}$ *is nonsingular,*

$$(\mathbf{D}_n^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n)^{-1} = \mathbf{D}_n^+(\mathbf{A}^{-1} \otimes \mathbf{A}^{-1})\mathbf{D}_n. \tag{58}$$

Proof. The first two equalities follow from $\mathbf{D}_n\mathbf{D}_n^+ = \mathbf{N}_n$, $\mathbf{N}_n(\mathbf{A} \otimes \mathbf{A}) = (\mathbf{A} \otimes \mathbf{A})\mathbf{N}_n$, $\mathbf{N}_n\mathbf{D}_n = \mathbf{D}_n$ and $\mathbf{D}_n^+\mathbf{N}_n = \mathbf{D}_n^+$. (See equations (54), (40), (52) and (53).) The last equality follows by direct verification since

$$\begin{aligned}\mathbf{D}_n^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n\mathbf{D}_n^+(\mathbf{A}^{-1} \otimes \mathbf{A}^{-1})\mathbf{D}_n &= \mathbf{D}_n^+(\mathbf{A} \otimes \mathbf{A})(\mathbf{A}^{-1} \otimes \mathbf{A}^{-1})\mathbf{D}_n \\ &= \mathbf{D}_n^+\mathbf{D}_n = \mathbf{I},\end{aligned}$$

using equations (56) and (54). ■

In fact, the property $\mathbf{D}_n\mathbf{D}_n^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n = (\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n$ for *arbitrary* square $\mathbf{A}$ is the Kronecker counterpart to $\mathbf{D}_n\mathbf{D}_n^+ \operatorname{vec} \mathbf{A} = \operatorname{vec} \mathbf{A}$ for symmetric $\mathbf{A}$, just as the property $\mathbf{K}_{nn}(\mathbf{A} \otimes \mathbf{A}) = (\mathbf{A} \otimes \mathbf{A})\mathbf{K}_{nn}$ is the Kronecker counterpart to $\mathbf{K}_{nn} \operatorname{vec} \mathbf{A} = \operatorname{vec} \mathbf{A}'$. We see this immediately if we let $\mathbf{X}$ be symmetric and substitute the symmetric matrix $\mathbf{AXA}'$ for $\mathbf{X}$ in $\mathbf{D}_n\mathbf{D}_n^+ \operatorname{vec} \mathbf{X} = \operatorname{vec} \mathbf{X}$, yielding

$$\mathbf{D}_n\mathbf{D}_n^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n v(\mathbf{X}) = (\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n v(\mathbf{X}).$$

Next we show that if $\mathbf{A}$ has a certain structure (diagonal, triangular), then $\mathbf{D}_n^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n$ often possesses the same structure.

LEMMA 12. *Let* $\mathbf{A}$ *be a diagonal (upper triangular, lower triangular)* $n \times n$ *matrix with diagonal elements* $a_{11}, a_{22}, \ldots, a_{nn}$. *Then the* $\frac{1}{2}n(n+1) \times \frac{1}{2}n(n+1)$ *matrix* $\mathbf{D}_n^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n$ *is also diagonal (upper triangular, lower triangular) with diagonal elements* $a_{ii}a_{jj}$ $(1 \leqslant j \leqslant i \leqslant n)$.

Proof. Let $\mathbf{E}_{ij}$ be the $n \times n$ matrix with 1 in the ijth position and zeros elsewhere, and define

$$\mathbf{T}_{ij} = \mathbf{E}_{ij} + \mathbf{E}_{ji} - \delta_{ij}\mathbf{E}_{ii}.$$

Then, for $i \geqslant j$,

$$\begin{aligned}\mathbf{D}_n^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n v(\mathbf{E}_{ij}) &= \mathbf{D}_n^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n v(\mathbf{T}_{ij}) \\ &= \mathbf{D}_n^+(\mathbf{A} \otimes \mathbf{A}) \operatorname{vec} \mathbf{T}_{ij} = \mathbf{D}_n^+ \operatorname{vec} \mathbf{A}\mathbf{T}_{ij}\mathbf{A}' = v(\mathbf{A}\mathbf{T}_{ij}\mathbf{A}'),\end{aligned}$$

and therefore, for $i \geqslant j$ and $s \geqslant t$,

$$\begin{aligned}(v(\mathbf{E}_{st}))'\mathbf{D}_n^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n v(\mathbf{E}_{ij}) &= (v(\mathbf{E}_{st}))'v(\mathbf{A}\mathbf{T}_{ij}\mathbf{A}') \\ &= (\mathbf{A}\mathbf{T}_{ij}\mathbf{A}')_{st} = a_{si}a_{tj} + a_{sj}a_{ti} - \delta_{ij}a_{si}a_{ti}.\end{aligned}$$

In particular, if $\mathbf{A}$ is upper triangular, we obtain[2]

$$(v(\mathbf{E}_{st}))'\mathbf{D}_n^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n v(\mathbf{E}_{ij}) = \begin{cases} a_{si}a_{tj} & (t \leqslant s \leqslant j = i \text{ or } t \leqslant j < s \leqslant i), \\ a_{si}a_{tj} + a_{sj}a_{ti} & (t \leqslant s \leqslant j < i), \\ 0 & (\text{otherwise}), \end{cases}$$

so that $\mathbf{D}_n^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n$ is upper triangular if $\mathbf{A}$ is, and

$$(v(\mathbf{E}_{ij}))'\mathbf{D}_n^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n v(\mathbf{E}_{ij}) = a_{ii}a_{jj} \qquad (j \leqslant i)$$

are its diagonal elements. The case where $\mathbf{A}$ is lower triangular is proved similarly. The case where $\mathbf{A}$ is diagonal follows as a special case. ■

Lemma 12 is instrumental in proving our main result concerning the matrix $\mathbf{D}_n^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}$.

LEMMA 13. *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix with eigenvalues* $\lambda_1, \lambda_2, \ldots, \lambda_n$. *Then the eigenvalues of the matrix* $\mathbf{D}_n^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n$ *are* $\lambda_i\lambda_j (1 \leqslant i \leqslant j \leqslant n)$, *and its trace and determinant are given by*

$$\operatorname{tr}(\mathbf{D}_n^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n) = \tfrac{1}{2} \operatorname{tr} \mathbf{A}^2 + \tfrac{1}{2}(\operatorname{tr} \mathbf{A})^2 \tag{59}$$

and

$$|\mathbf{D}_n^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n| = |\mathbf{A}|^{n+1}. \tag{60}$$

Proof. By Schur's Theorem [4, p. 202] there exists a nonsingular matrix $\mathbf{S}$ such that $\mathbf{S}^{-1}\mathbf{AS} = \mathbf{M}$, where $\mathbf{M}$ is an upper triangular matrix with the eigenvalues $\lambda_1, \ldots, \lambda_n$ of $\mathbf{A}$ on its diagonal. Thus

$$\mathbf{D}_n^+(\mathbf{S}^{-1} \otimes \mathbf{S}^{-1})\mathbf{D}_n\mathbf{D}_n^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n\mathbf{D}_n^+(\mathbf{S} \otimes \mathbf{S})\mathbf{D}_n = \mathbf{D}_n^+(\mathbf{M} \otimes \mathbf{M})\mathbf{D}_n.$$

Since $\mathbf{D}_n^+(\mathbf{S}^{-1} \otimes \mathbf{S}^{-1})\mathbf{D}_n$ is the inverse of $\mathbf{D}_n^+(\mathbf{S} \otimes \mathbf{S})\mathbf{D}_n$ (see equation (58)), it follows that $\mathbf{D}_n^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n$ and $\mathbf{D}_n^+(\mathbf{M} \otimes \mathbf{M})\mathbf{D}_n$ have the same set of eigenvalues. By Lemma 12, the latter matrix is upper triangular with eigenvalues (diagonal elements) $\lambda_i\lambda_j$ $(1 \leqslant j \leqslant i \leqslant n)$. These are therefore the eigenvalues of $\mathbf{D}_n^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n$ too.

The trace and determinant, being the sum and the product of the eigenvalues, respectively, are

$$\begin{aligned} \operatorname{tr} \mathbf{D}_n^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n &= \sum_{i \geqslant j} \lambda_i\lambda_j = \tfrac{1}{2}\sum_i \lambda_i^2 + \tfrac{1}{2}\sum_{ij} \lambda_i\lambda_j \\ &= \tfrac{1}{2} \operatorname{tr} \mathbf{A}^2 + \tfrac{1}{2}(\operatorname{tr} \mathbf{A})^2, \end{aligned}$$

and

$$|\mathbf{D}_n^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n| = \prod_{i \geqslant j} \lambda_i\lambda_j = \prod_i \lambda_i^{n+1} = |\mathbf{A}|^{n+1}. \qquad \blacksquare$$

Let us now establish the nature of the nonsingular $\frac{1}{2}n(n+1) \times \frac{1}{2}n(n+1)$ matrix $\mathbf{D}_n'\mathbf{D}_n$. Let $\mathbf{B} = (b_{ij})$ and $\mathbf{C} = (c_{ij})$ be arbitrary symmetric $n \times n$ matrices, and let $\mathbf{E}_{ii}$ be the $n \times n$ matrix with 1 in the ith diagonal position and zeros elsewhere. Then

$$\begin{aligned} (v(\mathbf{B}))'\mathbf{D}_n'\mathbf{D}_n v(\mathbf{C}) &= (\operatorname{vec} \mathbf{B})' \operatorname{vec} \mathbf{C} = \sum_{ij} b_{ij}c_{ij} \\ &= 2\sum_{i \geqslant j} b_{ij}c_{ij} - \sum_i b_{ii}c_{ii} \\ &= 2(v(\mathbf{B}))'v(\mathbf{C}) - \sum_i \big((v(\mathbf{B}))'v(\mathbf{E}_{ii})\big)\big((v(\mathbf{E}_{ii}))'v(\mathbf{C})\big) \\ &= (v(\mathbf{B}))'\left(2\mathbf{I} - \sum_i v(\mathbf{E}_{ii})(v(\mathbf{E}_{ii}))'\right)v(\mathbf{C}), \end{aligned}$$

so that

$$\mathbf{D}_n'\mathbf{D}_n = 2\mathbf{I}_{(1/2)n(n+1)} - \sum_{i=1}^{n} v(\mathbf{E}_{ii})(v(\mathbf{E}_{ii}))'. \tag{61}$$

Hence, $\mathbf{D}_n'\mathbf{D}_n$ is a diagonal matrix with diagonal elements 1 (n times) and 2 ($\frac{1}{2}n(n-1)$ times) and determinant

$$|\mathbf{D}_n'\mathbf{D}_n| = 2^{(1/2)n(n-1)}. \tag{62}$$

As a consequence of (61) we have for any $n \times n$ matrix $\mathbf{A}$,

$$\mathbf{D}_n'\mathbf{D}_n v(\mathbf{A}) = v(2\mathbf{A} - dg\ \mathbf{A}). \tag{63}$$

Also

$$\mathbf{D}_n' \operatorname{vec} \mathbf{A} = v(\mathbf{A} + \mathbf{A}' - dg\ \mathbf{A}). \tag{64}$$

Here $dg\ \mathbf{A}$ denotes the diagonal matrix with the diagonal elements of $\mathbf{A}$ on its diagonal. The proof of (64) follows by premultiplying both sides of (55) by $\mathbf{D}_n'\mathbf{D}_n$ and using (63).

With the help of (62) we can now prove Lemma 14.

LEMMA 14. *Let* $\mathbf{A}$ *be an* n × n *matrix. Then*

$$|\mathbf{D}_n'(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n| = 2^{(1/2)n(n-1)}|\mathbf{A}|^{n+1}, \tag{65}$$

and, if $\mathbf{A}$ *is nonsingular,*

$$(\mathbf{D}_n'(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n)^{-1} = \mathbf{D}_n^+(\mathbf{A}^{-1} \otimes \mathbf{A}^{-1})\mathbf{D}_n^{+\prime}. \tag{66}$$

Proof. Since, from (56),

$$\mathbf{D}_n'(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n = (\mathbf{D}_n'\mathbf{D}_n)(\mathbf{D}_n^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n),$$

(65) follows from (62) and (60), and (66) follows from (58) and (50). ■

Historical Note. The idea of putting into a single vector just the distinct elements of a symmetric matrix goes back at least to Aitken [1] in 1949. Properties of the duplication matrix were studied, inter alia, by Tracy and Singh [39], Browne [5], Vetter [41], Richard [33], Balestra [2], Nel [22], and Henderson and Searle [9]. For further properties of the duplication matrix the reader should consult Magnus and Neudecker [18].

8. A GENERALIZATION: L-STRUCTURED MATRICES

The class of symmetric matrices is just one example of a much wider class of matrices: **L**-structures. An **L**-structure (**L** stands for linear) is the totality of real matrices of a specified order that satisfy a given set of linear restrictions. To define the concept of an **L**-structure more formally, let $\mathscr{D}$ be an s-dimensional subspace (or linear manifold) of the real vector space $\mathbb{R}^{mn}$, and let $d_1, d_2, \ldots, d_s$ be a set of basis vectors for $\mathscr{D}$. The $mn \times s$ matrix

$$\Delta = (d_1, d_2, \ldots, d_s)$$

is called a *basis matrix* for $\mathscr{D}$, and the collection of real $m \times n$ matrices

$$\mathbf{L}(\Delta) = \{\mathbf{X} \,|\, \mathbf{X} \in \mathbf{R}^{m \times n}, \qquad \operatorname{vec} \mathbf{X} \in \mathscr{D}\} \tag{67}$$

is called an **L**-structure; s is called the *dimension* of the **L**-structure. A basis matrix is, of course, not unique: if Δ is a basis matrix for $\mathscr{D}$, then so is $\Delta\mathbf{E}$ for any nonsingular $\mathbf{E}$. This fact suggests that it might be more appropriate to regard **L** as function of $\mathscr{D}$ rather than Δ. It is, however, the basis matrix Δ which is relevant in applications such as matrix equations and Jacobians, so we find it convenient to retain the definition (67) as it stands.

The class of real symmetric $n \times n$ matrices is clearly an **L**-structure, the linear restrictions being the $\frac{1}{2}n(n-1)$ equalities $x_{ij} = x_{ji}$, so that the dimension of the **L**-structure is $\frac{1}{2}n(n+1)$. One choice for Δ would be the duplication matrix $\mathbf{D}_n$. Other examples of **L**-structures are (strictly) triangular, skew-symmetric, diagonal, circulant, and Toeplitz matrices.

Now consider a member $\mathbf{A}$ of the class of real $m \times n$ matrices defined by the **L**-structure $\mathbf{L}(\Delta)$ of dimension s. Since $\mathbf{A} \in \mathbf{L}(\Delta)$, the vector $\operatorname{vec} \mathbf{A}$ lies in the space $\mathscr{D}$ spanned by the columns of Δ, and hence there exists an $s \times 1$ vector, say $\psi(\mathbf{A})$, such that

$$\Delta\psi(\mathbf{A}) = \operatorname{vec} \mathbf{A}. \tag{68}$$

Since Δ has full column-rank s, we obtain

$$\Delta^+\Delta = \mathbf{I}_s, \tag{69}$$

which implies that $\psi(\mathbf{A})$ can be solved uniquely from (68), the unique solution being

$$\psi(\mathbf{A}) = \Delta^+ \operatorname{vec} \mathbf{A} \qquad (\mathbf{A} \in \mathbf{L}(\Delta)). \tag{70}$$

Thus, given the choice of Δ, ψ is uniquely determined by (68). (Of course, a different choice of Δ leads to a different ψ.) In the case of symmetry, the choice of the duplication matrix for Δ determines the choice of $v(.)$ for ψ. From (55) we have, for *arbitrary* $\mathbf{A}$, $\mathbf{D}_n^+ \operatorname{vec} \mathbf{A} = \frac{1}{2}v(\mathbf{A} + \mathbf{A}')$; for *symmetric* $\mathbf{A}$ this becomes $\mathbf{D}_n^+ \operatorname{vec} \mathbf{A} = v(\mathbf{A})$.

Of special interest is the symmetric idempotent $s \times s$ matrix $\mathbf{N}_\Delta$ defined as

$$\mathbf{N}_\Delta = \Delta\Delta^+. \tag{71}$$

(In the case of symmetry, this is the matrix $\mathbf{N}$.) If we substitute $\Delta^+ \operatorname{vec} \mathbf{A}$ for $\psi(\mathbf{A})$ in (68) we obtain

$$\mathbf{N}_\Delta \operatorname{vec} \mathbf{A} = \Delta\Delta^+ \operatorname{vec} \mathbf{A} = \operatorname{vec} \mathbf{A} \tag{72}$$

for every $\mathbf{A} \in \mathbf{L}(\Delta)$. We shall show that the matrix $\mathbf{N}_\Delta$ is *invariant* to the choice of Δ. Let Δ and $\bar{\Delta}$ be two basis matrices for $\mathscr{D}$. Since Δ and $\bar{\Delta}$ span the same subspace, there exists a nonsingular $s \times s$ matrix $\mathbf{E}$ such that $\bar{\Delta} = \Delta\mathbf{E}$. Also,

$$\begin{aligned}(\Delta\mathbf{E})(\Delta\mathbf{E})^+ &= (\Delta\mathbf{E})^{+\prime}(\Delta\mathbf{E})' = (\Delta\mathbf{E})^{+\prime}\mathbf{E}'\Delta' \\ &= (\Delta\mathbf{E})^{+\prime}\mathbf{E}'\Delta'\Delta\Delta^+ = (\Delta\mathbf{E})^{+\prime}(\Delta\mathbf{E})'\Delta\mathbf{E}\mathbf{E}^{-1}\Delta^+ \\ &= (\Delta\mathbf{E})(\Delta\mathbf{E})^+(\Delta\mathbf{E})\mathbf{E}^{-1}\Delta^+ = \Delta\mathbf{E}\mathbf{E}^{-1}\Delta^+ = \Delta\Delta^+.\end{aligned}$$

Hence $\mathbf{N}_\Delta$ is invariant to the choice of Δ.

Now suppose that $\mathbf{A}$ and $\mathbf{B}$ are square matrices of orders $n \times n$ and $m \times m$, respectively, possessing the property

$$\mathbf{B}\mathbf{X}\mathbf{A}' \in \mathbf{L}(\Delta) \tag{73}$$

for every $\mathbf{X} \in \mathbf{L}(\Delta)$. (For example, in the case of (skew-)symmetry, $\mathbf{A}\mathbf{X}\mathbf{A}'$ is (skew-)symmetric for every (skew-)symmetric $\mathbf{X}$; in the case of (strict) lower triangularity, if $\mathbf{P}$ and $\mathbf{Q}$ are lower triangular, $\mathbf{P}\mathbf{X}\mathbf{Q}$ is (strictly) lower triangular for every (strictly) lower triangular $\mathbf{X}$.) Then,

$$\Delta\Delta^+(\mathbf{A} \otimes \mathbf{B})\Delta = (\mathbf{A} \otimes \mathbf{B})\Delta, \tag{74}$$

and, if $\mathbf{A}$ and $\mathbf{B}$ are nonsingular,

$$(\Delta^+(\mathbf{A} \otimes \mathbf{B})\Delta)^{-1} = \Delta^+(\mathbf{A}^{-1} \otimes \mathbf{B}^{-1})\Delta \tag{75}$$

and

$$(\Delta'(\mathbf{A} \otimes \mathbf{B})\Delta)^{-1} = \Delta^+(\mathbf{A}^{-1} \otimes \mathbf{B}^{-1})\Delta^{+\prime}. \tag{76}$$

To prove (74) let $\mathbf{X} \in \mathbf{L}(\Delta)$. Then

$$\begin{aligned}\Delta\Delta^+(\mathbf{A} \otimes \mathbf{B})\Delta\psi(\mathbf{X}) &= \Delta\Delta^+(\mathbf{A} \otimes \mathbf{B}) \operatorname{vec} \mathbf{X} = \Delta\Delta^+ \operatorname{vec} \mathbf{B}\mathbf{X}\mathbf{A}' \\ &= \operatorname{vec} \mathbf{B}\mathbf{X}\mathbf{A}' = (\mathbf{A} \otimes \mathbf{B}) \operatorname{vec} \mathbf{X} = (\mathbf{A} \otimes \mathbf{B})\Delta\psi(\mathbf{X}).\end{aligned}$$

The restriction $\mathbf{X} \in \mathbf{L}(\Delta)$ does not restrict $\psi(\mathbf{X})$; hence (74) follows. Property (74) together with (69) implies (75), since

$$\Delta^+(\mathbf{A}^{-1} \otimes \mathbf{B}^{-1})\Delta\Delta^+(\mathbf{A} \otimes \mathbf{B})\Delta = \Delta^+(\mathbf{A}^{-1} \otimes \mathbf{B}^{-1})(\mathbf{A} \otimes \mathbf{B})\Delta = \Delta^+\Delta = \mathbf{I},$$

while (76) also follows from (74) and (69), using the symmetry of $\Delta\Delta^+$.

Examples of **L**-*structures*. The following six **L**-structures are most likely to appear in practical situations. Each defines a class of square matrices, say of order $n \times n$. The **L**-structures are (with their dimensions in brackets): (1) symmetric $[n(n+1)/2]$, (2) lower triangular $[n(n+1)/2]$, (3) skew-symmetric $[n(n-1)/2]$, (4) strictly lower triangular $[n(n-1)/2]$, (5) diagonal $[n]$, and (6) circulant $[n]$. For $n = 3$ sensible choices for Δ are (with dots representing zeros):

$$\Delta_1 = \left(\begin{array}{ccc|cc|c} 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ \hline \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ \hline \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{array}\right) \qquad \Delta_2 = \left(\begin{array}{ccc|cc|c} 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ \hline \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ \hline \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{array}\right)$$

$$\Delta_3 = \left(\begin{array}{ccc} \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot \\ \cdot & 1 & \cdot \\ \hline -1 & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 \\ \hline \cdot & -1 & \cdot \\ \cdot & \cdot & -1 \\ \cdot & \cdot & \cdot \end{array}\right) \qquad \Delta_4 = \left(\begin{array}{ccc} \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot \\ \cdot & 1 & \cdot \\ \hline \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 \\ \hline \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{array}\right)$$

$$\Delta_5 = \left(\begin{array}{ccc} 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \hline \cdot & \cdot & \cdot \\ \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot \\ \hline \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 \end{array}\right) \qquad \Delta_6 = \left(\begin{array}{ccc} 1 & \cdot & \cdot \\ \cdot & 1 & \cdot \\ \cdot & \cdot & 1 \\ \hline \cdot & \cdot & 1 \\ 1 & \cdot & \cdot \\ \cdot & 1 & \cdot \\ \hline \cdot & 1 & \cdot \\ \cdot & \cdot & 1 \\ 1 & \cdot & \cdot \end{array}\right).$$

Let $\mathbf{A} = (a_{ij})$ be an arbitrary $n \times n$ matrix. We have already encountered the $n^2 \times 1$ vector vec $\mathbf{A}$ and the $\frac{1}{2}n(n+1) \times 1$ vector $v(\mathbf{A})$, which is obtained from vec $\mathbf{A}$ by eliminating all supradiagonal elements of $\mathbf{A}$. We now define the $\frac{1}{2}n(n-1) \times 1$ vector $v_s(\mathbf{A})$, which is obtained from vec $\mathbf{A}$ by eliminating the supradiagonal *and the diagonal* elements of $\mathbf{A}$. For example, if $n = 3$, then

$$v_s(\mathbf{A}) = (a_{21}, a_{31}, a_{32})'. \tag{77}$$

We also define the $n \times 1$ vectors

$$v_d(\mathbf{A}) = (a_{11}, a_{22}, \ldots, a_{nn})' \tag{78}$$

and

$$v_1(\mathbf{A}) = (a_{11}, a_{21}, \ldots, a_{n1})'. \tag{79}$$

The vector $v_d(\mathbf{A})$ thus contains the diagonal elements of $\mathbf{A}$; the vector $v_1(\mathbf{A})$ contains the first column of $\mathbf{A}$.

The ψ-vectors associated with the six above Δ-matrices are then

$$\psi_1(\mathbf{A}) = \psi_2(\mathbf{A}) = v(\mathbf{A}), \qquad \psi_3(\mathbf{A}) = \psi_4(\mathbf{A}) = v_s(\mathbf{A}),$$
$$\psi_5(\mathbf{A}) = v_d(\mathbf{A}), \qquad \psi_6(\mathbf{A}) = v_1(\mathbf{A}).$$

Historical note. Patterned matrices (with only equality relationships among their elements) were studied by Tracy and Singh [39] with the purpose of finding matrix derivatives of certain matrix transformations. Lower triangular (and symmetric) matrices were discussed by Magnus and Neudecker [18], and skew-symmetric, strictly lower triangular and diagonal matrices by Neudecker [25]. The present section is based on Magnus [16], who introduced the concept of an $\mathbf{L}$-structure in the context of solving linear matrix equations where the solution matrix is known to be $\mathbf{L}$-structured. See also Wiens [42].

9. MATRIX DIFFERENTIATION: FIRST DERIVATIVES

Let $f = (f_1, f_2, \ldots, f_m)'$ be a *vector* function with values in $\mathbb{R}^m$ which is differentiable on a set S in $\mathbb{R}^n$. Let $\partial f_i(x)/\partial x_j$ denote the partial derivative of f_i with respect to the jth coordinate. Then the $m \times n$ matrix

$$\begin{pmatrix} \partial f_1(x)/\partial x_1 & \cdots & \partial f_1(x)/\partial x_n \\ \vdots & & \vdots \\ \partial f_m(x)/\partial x_1 & \cdots & \partial f_m(x)/\partial x_n \end{pmatrix} \tag{80}$$

is called the *derivative* or *Jacobian matrix* of f at x and is denoted $\partial f(x)/\partial x'$.

The generalization to matrix functions of matrices is straightforward. Let $\mathbf{F}: S \to \mathbb{R}^{m \times p}$ be a *matrix* function defined and differentiable on a set S in $\mathbb{R}^{n \times q}$. Then we define the *Jacobian matrix* of $\mathbf{F}$ at $\mathbf{X}$ as the $mp \times nq$ matrix

$$\frac{\partial \operatorname{vec} \mathbf{F}(\mathbf{X})}{\partial (\operatorname{vec} \mathbf{X})'} \tag{81}$$

whose ijth element is the partial derivative of the ith component of vec $\mathbf{F}(\mathbf{X})$ with respect to the jth coordinate of vec $\mathbf{X}$.

We emphasize that (81) is *the only sensible definition of a matrix derivative.*[3] There are, of course, other ways in which the $mnpq$ partial derivatives of $\mathbf{F}$ could be displayed [2, 34], but these other definitions typically do not preserve the rank of the transformation (so that the determinant of the matrix of partial derivatives is not the Jacobian), and do not allow a useful chain rule. These points are discussed in more detail by Pollock [32] and Magnus and Neudecker [19].

The computation of Jacobian matrices is made extremely simple by the use of *differentials* [24, 19]. The essential property here is that

$$\operatorname{vec} d\mathbf{F}(\mathbf{X}) = \mathbf{A}(\mathbf{X}) \operatorname{vec} d\mathbf{X} \tag{82}$$

if, and only if,

$$\frac{\partial \operatorname{vec} \mathbf{F}(\mathbf{X})}{\partial (\operatorname{vec} \mathbf{X})'} = \mathbf{A}(\mathbf{X}). \tag{83}$$

Thus, if we can find a matrix $\mathbf{A}$ (which may depend on $\mathbf{X}$, but not on $d\mathbf{X}$) satisfying (82), then this matrix is the Jacobian matrix. Some examples will show that the approach via differentials is short, elegant, and easy.

Example (i). The linear matrix function $\mathbf{Y} = \mathbf{AXB}$ where $\mathbf{A}$ and $\mathbf{B}$ are two matrices of constants. Taking differentials we have

$$d\mathbf{Y} = \mathbf{A}(d\mathbf{X})\mathbf{B},$$

from which we obtain, upon vectorizing,

$$\operatorname{vec} d\mathbf{Y} = \operatorname{vec} \mathbf{A}(d\mathbf{X})\mathbf{B} = (\mathbf{B}' \otimes \mathbf{A}) \operatorname{vec} d\mathbf{X}. \tag{84}$$

Hence the Jacobian matrix is

$$\frac{\partial \operatorname{vec} \mathbf{Y}}{\partial (\operatorname{vec} \mathbf{X})'} = \mathbf{B}' \otimes \mathbf{A}. \tag{85}$$

If $\mathbf{X}$ is constrained to be symmetric, we substitute $\mathbf{D}\, dv(\mathbf{X})$ for vec $d\mathbf{X}$ in (84), where $\mathbf{D}$ is the duplication matrix. This gives

$$\operatorname{vec} d\mathbf{Y} = (\mathbf{B}' \otimes \mathbf{A})\mathbf{D}\, dv(\mathbf{X}),$$

so that

$$\frac{\partial \operatorname{vec} \mathbf{Y}}{\partial (v(\mathbf{X}))'} = (\mathbf{B}' \otimes \mathbf{A})\mathbf{D}. \tag{86}$$

Of course, we can also obtain (86) from (85) using the chain rule, since for symmetric $\mathbf{X}$,

$$\frac{\partial \operatorname{vec} \mathbf{X}}{\partial (v(\mathbf{X}))'} = \mathbf{D}. \tag{87}$$

More generally, if $\mathbf{X}$ is $\mathbf{L}$-structured, $\mathbf{X} \in \mathbf{L}(\Delta)$, then the Jacobian matrix is $(\mathbf{B}' \otimes \mathbf{A})\Delta$.

Example (ii). The nonlinear matrix function $\mathbf{Y} = \mathbf{X}^{-1}$. We take differentials,

$$d\mathbf{Y} = d\mathbf{X}^{-1} = -\mathbf{X}^{-1}(d\mathbf{X})\mathbf{X}^{-1}$$

and vecs,

$$\operatorname{vec} d\mathbf{Y} = -((\mathbf{X}')^{-1} \otimes \mathbf{X}^{-1}) \operatorname{vec} d\mathbf{X},$$

thus leading to the Jacobian matrix

$$\frac{\partial \operatorname{vec} \mathbf{Y}}{\partial (\operatorname{vec} \mathbf{X})'} = -(\mathbf{X}')^{-1} \otimes \mathbf{X}^{-1}. \tag{88}$$

Again, if $\mathbf{X}$ is symmetric ($\mathbf{L}$-structured), we postmultiply (88) by $\mathbf{D}$ (Δ, in general).

Example (iii). The real-valued function $\phi(\mathbf{X}) = \operatorname{tr} \mathbf{AX}$, where $\mathbf{A}$ is a matrix of constants. We have

$$d\phi(\mathbf{X}) = \operatorname{tr} \mathbf{A}\, d\mathbf{X} = (\operatorname{vec} \mathbf{A}')' \operatorname{vec} d\mathbf{X},$$

so that the derivative is

$$\frac{\partial \phi(\mathbf{X})}{\partial (\operatorname{vec} \mathbf{X})'} = (\operatorname{vec} \mathbf{A}')'. \tag{89}$$

(This is usually written as $\partial\phi(\mathbf{X})/\partial\mathbf{X} = \mathbf{A}'$, which, in spite of its attractiveness, is not always commendable.) For symmetric $\mathbf{X}$, we proceed as before and find, using (64),

$$\frac{\partial\phi(\mathbf{X})}{\partial(v(\mathbf{X}))'} = (\text{vec } \mathbf{A}')'\mathbf{D} = \{v(\mathbf{A} + \mathbf{A}' - dg\ \mathbf{A})\}', \tag{90}$$

where $dg\ \mathbf{A}$ is the diagonal matrix with the diagonal elements of $\mathbf{A}$ on its diagonal.

10. MATRIX DIFFERENTIATION: SECOND DERIVATIVES

Let $\phi: S \to \mathbb{R}$ be a real-valued function defined and twice differentiable on a set S in $\mathbb{R}^n$. Let $\partial^2\phi(x)/\partial x_i \partial x_j$ denote the second-order partial derivative of ϕ with respect to the ith and jth coordinates. Then the $n \times n$ matrix $(\partial^2\phi(x)/\partial x_i \partial x_j)$ is called the *Hessian matrix* of ϕ at x and is denoted $\mathbf{H}\phi(x)$ or $\partial^2\phi(x)/\partial x \partial x'$. Since ϕ is twice differentiable at x, $\mathbf{H}\phi(x)$ is a symmetric matrix.

Next, let us consider a real-valued function $\phi: S \to \mathbb{R}$ defined and twice differentiable on $\mathbf{S} \subset \mathbb{R}^{n \times q}$. The Hessian matrix of ϕ at $\mathbf{X}$ is then the $nq \times nq$ (symmetric) matrix

$$\mathbf{H}\phi(\mathbf{X}) = \frac{\partial^2\phi(\mathbf{X})}{\partial \text{ vec } \mathbf{X}\ \partial(\text{vec } \mathbf{X})'}, \tag{91}$$

whose ijth element is the second-order partial derivative of ϕ with respect to the ith and jth coordinates of vec $\mathbf{X}$.

The computation of Hessian matrices is based on the property that

$$d^2\phi(\mathbf{X}) = (\text{vec } d\mathbf{X})'\mathbf{B}(\mathbf{X})(\text{vec } d\mathbf{X}) \tag{92}$$

if, and only if,

$$\mathbf{H}\phi(\mathbf{X}) = \tfrac{1}{2}(\mathbf{B}(\mathbf{X}) + \mathbf{B}'(\mathbf{X})), \tag{93}$$

where $\mathbf{B}$ may depend on $\mathbf{X}$, but not on $d\mathbf{X}$.

Example (i). The quadratic function $\phi(\mathbf{X}) = \text{tr}\,\mathbf{AXBX}'$, where $\mathbf{A}$ and $\mathbf{B}$ are square matrices (not necessarily of the same order) of constants. Twice taking differentials, we obtain

$$d^2\phi(\mathbf{X}) = 2 \text{ tr } \mathbf{A}(d\mathbf{X})\mathbf{B}(d\mathbf{X})' = 2(\text{vec } d\mathbf{X})'(\mathbf{B}' \otimes \mathbf{A})(\text{vec } d\mathbf{X}). \tag{94}$$

The Hessian matrix is therefore

$$\frac{\partial^2 \phi(\mathbf{X})}{\partial \operatorname{vec} \mathbf{X}\, \partial(\operatorname{vec} \mathbf{X})'} = \mathbf{B}' \otimes \mathbf{A} + \mathbf{B} \otimes \mathbf{A}'. \tag{95}$$

If $\mathbf{X}$ is constrained to be symmetric, we have

$$\frac{\partial^2 \phi(\mathbf{X})}{\partial v(\mathbf{X})\, \partial(v(\mathbf{X}))'} = \mathbf{D}'(\mathbf{B}' \otimes \mathbf{A} + \mathbf{B} \otimes \mathbf{A}')\mathbf{D}. \tag{96}$$

Example (ii). The real-valued function $\phi(\mathbf{X}) = \operatorname{tr} \mathbf{X}^{-1}$. We have

$$d\phi(\mathbf{X}) = -\operatorname{tr} \mathbf{X}^{-1}(d\mathbf{X})\mathbf{X}^{-1}$$

and therefore

$$\begin{aligned} d^2\phi(\mathbf{X}) &= -\operatorname{tr}(d\mathbf{X}^{-1})(d\mathbf{X})\mathbf{X}^{-1} - \operatorname{tr} \mathbf{X}^{-1}(d\mathbf{X})(d\mathbf{X}^{-1}) \\ &= 2 \operatorname{tr} \mathbf{X}^{-1}(d\mathbf{X})\mathbf{X}^{-1}(d\mathbf{X})\mathbf{X}^{-1} = 2(\operatorname{vec} d\mathbf{X}')'(\mathbf{X}'^{-2} \otimes \mathbf{X}^{-1})(\operatorname{vec} d\mathbf{X}) \\ &= 2(\operatorname{vec} d\mathbf{X})'\mathbf{K}(\mathbf{X}'^{-2} \otimes \mathbf{X}^{-1})(\operatorname{vec} d\mathbf{X}), \end{aligned} \tag{97}$$

so that the Hessian matrix becomes

$$\frac{\partial^2 \phi(\mathbf{X})}{\partial \operatorname{vec} \mathbf{X}\, \partial(\operatorname{vec} \mathbf{X})'} = \mathbf{K}(\mathbf{X}'^{-2} \otimes \mathbf{X}^{-1} + \mathbf{X}'^{-1} \otimes \mathbf{X}^{-2}). \tag{98}$$

For symmetric $\mathbf{X}$, we find

$$\begin{aligned} \frac{\partial^2 \phi(\mathbf{X})}{\partial v(\mathbf{X})\, \partial(v(\mathbf{X}))'} &= \mathbf{D}'(\mathbf{X}^{-2} \otimes \mathbf{X}^{-1} + \mathbf{X}^{-1} \otimes \mathbf{X}^{-2})\mathbf{D} \\ &= 2\mathbf{D}'(\mathbf{X}^{-1} \otimes \mathbf{X}^{-2})\mathbf{D}, \end{aligned} \tag{99}$$

using (52) and Lemma 4.

11. THE EVALUATION OF INFORMATION MATRICES

Of particular importance to econometricians is the evaluation of information matrices (and their inverse) in situations where positive definite matrices are arguments of the likelihood function and where, consequently, a proper treatment of symmetry is needed. Let $\mathbf{\Phi}$ be a positive definite matrix of parameters to be estimated. Then two approaches are available. The first is to take account of the symmetry by imposing the linear constraint $(\mathbf{I} - \mathbf{K}) \operatorname{vec} \mathbf{\Phi} = 0$; the second is to insert the relationship $\operatorname{vec} \mathbf{\Phi} = \mathbf{D}v(\mathbf{\Phi})$ into the likelihood function. (Recall that $v(\mathbf{\Phi})$ contains the distinct elements of $\mathbf{\Phi}$, see Section 7.) Clearly the two approaches lead to the same result, but the latter treatment is, in our view, easier to apply. Let us give one example.

Consider a sample of size m from the n-dimensional normal distribution of y with zero mean and positive definite covariance matrix $\mathbf{\Phi}$. The loglikelihood function for the sample is

$$\Lambda_m(v(\mathbf{\Phi})) = -\tfrac{1}{2}nm \log 2\pi - \tfrac{1}{2}m \log|\mathbf{\Phi}| - \tfrac{1}{2} \operatorname{tr} \mathbf{\Phi}^{-1}\mathbf{Z},$$

where

$$\mathbf{Z} = \sum_{i=1}^{m} y_i y_i'.$$

Thus Λ_m is a function of $n(n+1)/2$ parameters. The first differential of Λ_m is

$$\begin{aligned} d\Lambda_m &= -\tfrac{1}{2}m\, d \log|\mathbf{\Phi}| - \tfrac{1}{2} \operatorname{tr}(d\mathbf{\Phi}^{-1})\mathbf{Z} - \tfrac{1}{2} \operatorname{tr} \mathbf{\Phi}^{-1}\, d\mathbf{Z} \\ &= -\tfrac{1}{2}m \operatorname{tr} \mathbf{\Phi}^{-1}\, d\mathbf{\Phi} + \tfrac{1}{2} \operatorname{tr} \mathbf{\Phi}^{-1}(d\mathbf{\Phi})\mathbf{\Phi}^{-1}\mathbf{Z} \\ &= \tfrac{1}{2} \operatorname{tr}(d\mathbf{\Phi})\mathbf{\Phi}^{-1}(\mathbf{Z} - m\mathbf{\Phi})\mathbf{\Phi}^{-1}. \end{aligned}$$

Now, since $\mathbf{\Phi}$ is a *linear* function of $v(\mathbf{\Phi})$, we have $d^2\mathbf{\Phi} = 0$ and hence the second differential of Λ_m is

$$\begin{aligned} d^2\Lambda_m = \tfrac{1}{2} &\operatorname{tr}(d\mathbf{\Phi})(d\mathbf{\Phi}^{-1})(\mathbf{Z} - m\mathbf{\Phi})\mathbf{\Phi}^{-1} \\ &+ \tfrac{1}{2} \operatorname{tr}(d\mathbf{\Phi})\mathbf{\Phi}^{-1}(d\mathbf{Z} - m\, d\mathbf{\Phi})\mathbf{\Phi}^{-1} \\ &+ \tfrac{1}{2} \operatorname{tr}(d\mathbf{\Phi})\mathbf{\Phi}^{-1}(\mathbf{Z} - m\mathbf{\Phi})(d\mathbf{\Phi}^{-1}). \end{aligned}$$

Taking expectations, and observing that $\mathscr{E}Z = m\mathbf{\Phi}$ and $d\mathbf{Z} = 0$, we find

$$-\mathscr{E}d^2\Lambda_m = \frac{m}{2} \operatorname{tr}(d\mathbf{\Phi})\mathbf{\Phi}^{-1}(d\mathbf{\Phi})\mathbf{\Phi}^{-1}.$$

It is only at this stage of the computation that we need to introduce the duplication matrix $\mathbf{D}$. We have, using (19) and (49),

$$\begin{aligned} \operatorname{tr}(d\mathbf{\Phi})\mathbf{\Phi}^{-1}(d\mathbf{\Phi})\mathbf{\Phi}^{-1} &= (\operatorname{vec} d\mathbf{\Phi})'(\mathbf{\Phi}^{-1} \otimes \mathbf{\Phi}^{-1})(\operatorname{vec} d\mathbf{\Phi}) \\ &= (dv(\mathbf{\Phi}))'\mathbf{D}'(\mathbf{\Phi}^{-1} \otimes \mathbf{\Phi}^{-1})\mathbf{D}\, dv(\mathbf{\Phi}), \end{aligned}$$

so that

$$-\mathscr{E}d^2\Lambda_m = \frac{m}{2} (dv(\mathbf{\Phi}))'\mathbf{D}'(\mathbf{\Phi}^{-1} \otimes \mathbf{\Phi}^{-1})\mathbf{D}\, dv(\mathbf{\Phi}). \tag{100}$$

Hence the information matrix for $v(\mathbf{\Phi})$ is $\mathbf{\Psi}_m = m\mathbf{\Psi}$, with $\mathbf{\Psi} = \frac{1}{2}\mathbf{D}'(\mathbf{\Phi}^{-1} \otimes \mathbf{\Phi}^{-1})\mathbf{D}$, and the asymptotic covariance matrix of the ML estimator for $v(\mathbf{\Phi})$ is $\mathbf{\Psi}^{-1} = 2\mathbf{D}^{+}(\mathbf{\Phi} \otimes \mathbf{\Phi})\mathbf{D}^{+\prime}$, using (66).

Historical note. That symmetry conditions should be properly taken into account was emphasized by Richard [33] and Balestra [2]. The treatment in this section follows Magnus and Neudecker [18, Section 5] with some minor modifications.

12. JACOBIANS INVOLVING L-STRUCTURES

Let $\mathbf{F}: S \to \mathbf{R}^{m \times p}$ be a matrix function defined and differentiable on a set S in $\mathbf{R}^{n \times q}$. If $mp = nq$, the Jacobian matrix defined in (81) is a square matrix. Its determinant is called the *Jacobian* (or Jacobian determinant) and is denoted by $\mathbf{J}_{\mathbf{F}}(\mathbf{X})$. Thus,

$$\mathbf{J}_{\mathbf{F}}(\mathbf{X}) = \left| \frac{\partial \operatorname{vec} \mathbf{F}(\mathbf{X})}{\partial (\operatorname{vec} \mathbf{X})'} \right|. \tag{101}$$

Example (i). The linear transformation $\mathbf{F}(\mathbf{X}) = \mathbf{AXB}$, where $\mathbf{X}$ and $\mathbf{F}(\mathbf{X})$ are $m \times n$ matrices, and $\mathbf{A}$ and $\mathbf{B}$ are nonsingular matrices of constants of orders $m \times m$ and $n \times n$, respectively. From (85) we know that the Jacobian matrix is $\mathbf{B}' \otimes \mathbf{A}$, so that the Jacobian is

$$\mathbf{J}_{\mathbf{F}}(\mathbf{X}) = |\mathbf{B}' \otimes \mathbf{A}| = |\mathbf{A}|^{n}|\mathbf{B}|^{m}. \tag{102}$$

Example (ii). The nonlinear transformation $\mathbf{F}(\mathbf{X}) = \mathbf{X}^{-1}$, where $\mathbf{X}$ is a nonsingular $n \times n$ matrix. The Jacobian matrix is given in (88) as $-(\mathbf{X}')^{-1} \otimes \mathbf{X}^{-1}$, so that the Jacobian of the transformation is

$$\mathbf{J}_{\mathbf{F}}(\mathbf{X}) = |-(\mathbf{X}')^{-1} \otimes \mathbf{X}^{-1}| = (-1)^{n}|\mathbf{X}|^{-2n}. \tag{103}$$

The evaluation of Jacobians of transformations involving a *symmetric* $n \times n$ matrix argument $\mathbf{X}$ proceeds along the same lines, except that we must now take into account the fact that $\mathbf{X}$ contains only $\frac{1}{2}n(n+1)$ "essential" variables.

Example (iii). The linear transformation $\mathbf{F}(\mathbf{X}) = \mathbf{AXA}'$. where $\mathbf{X}$ (and hence $\mathbf{F}(\mathbf{X})$) are symmetric $n \times n$ matrices. Taking differentials and vecs, we have

$$\operatorname{vec} d\mathbf{F}(\mathbf{X}) = (\mathbf{A} \otimes \mathbf{A}) \operatorname{vec} d\mathbf{X}.$$

Since $d\mathbf{X}$ and $d\mathbf{F}(\mathbf{X})$ are symmetric, we obtain

$$dv(\mathbf{F}(\mathbf{X})) = \mathbf{D}_n^{+}(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n \, dv(\mathbf{X}),$$

so that,

$$\mathbf{J_F}(\mathbf{X}) = \left|\frac{\partial v(\mathbf{F}(\mathbf{X}))}{\partial(v(\mathbf{X}))'}\right| = |\mathbf{D}_n^+(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n| = |\mathbf{A}|^{n+1}, \tag{104}$$

using (60).

Example (iv). The inverse transformation $\mathbf{F}(\mathbf{X}) = \mathbf{X}^{-1}$ for symmetric non-singular $\mathbf{X}$ of order $n \times n$. Again taking differentials and vecs, we obtain

$$\operatorname{vec} d\mathbf{F}(\mathbf{X}) = - \quad (\mathbf{X}^{-1} \otimes \mathbf{X}^{-1}) \operatorname{vec} d\mathbf{X},$$

so that

$$dv(\mathbf{F}(\mathbf{X})) = -\mathbf{D}_n^+(\mathbf{X}^{-1} \otimes \mathbf{X}^{-1})\mathbf{D}_n\, dv(\mathbf{X}).$$

The Jacobian of this transformation then follows from (60):

$$\mathbf{J_F}(\mathbf{X}) = \left|\frac{\partial v(\mathbf{F}(\mathbf{X}))}{\partial(v(\mathbf{X}))'}\right| = |-\mathbf{D}_n^+(\mathbf{X}^{-1} \otimes \mathbf{X}^{-1})\mathbf{D}_n| = (-1)^{(1/2)n(n+1)}|\mathbf{X}|^{-(n+1)}.$$

To evaluate the Jacobian matrix (and the Jacobian) of a transformation involving more general **L**-structures is straightforward.

Example (v). The transformation $\mathbf{F}(\mathbf{X}) = \mathbf{X}'\mathbf{X}$, where $\mathbf{X} = (x_{ij})$ is a lower triangular $n \times n$ matrix. From

$$d\mathbf{F}(\mathbf{X}) = (d\mathbf{X})'\mathbf{X} + \mathbf{X}'d\mathbf{X},$$

we obtain

$$\begin{aligned}\operatorname{vec} d\mathbf{F}(\mathbf{X}) &= (\mathbf{X}' \otimes \mathbf{I}) \operatorname{vec}(d\mathbf{X})' + (\mathbf{I} \otimes \mathbf{X}') \operatorname{vec} d\mathbf{X} \\ &= ((\mathbf{X}' \otimes \mathbf{I})\mathbf{K}_{nn} + \mathbf{I} \otimes \mathbf{X}') \operatorname{vec} d\mathbf{X} \\ &= (\mathbf{I} + \mathbf{K}_{nn})(\mathbf{I} \otimes \mathbf{X}') \operatorname{vec} d\mathbf{X} = 2\mathbf{N}_n(\mathbf{I} \otimes \mathbf{X}') \operatorname{vec} d\mathbf{X}.\end{aligned}$$

Now let $\mathbf{L}_n'$ be the Δ-matrix with the property that

$$\mathbf{L}_n' v(A) = \operatorname{vec} \mathbf{A}$$

for every lower triangular $n \times n$ matrix $\mathbf{A}$. Then, since $d\mathbf{X}$ is lower triangular and $d\mathbf{F}(\mathbf{X})$ is symmetric, we obtain

$$\begin{aligned}dv(\mathbf{F}(\mathbf{X})) &= 2\mathbf{D}_n^+\mathbf{N}_n(\mathbf{I} \otimes \mathbf{X}')\mathbf{L}_n'\, dv(\mathbf{X}) \\ &= 2(\mathbf{D}_n'\mathbf{D}_n)^{-1}\mathbf{D}_n'(\mathbf{I} \otimes \mathbf{X}')\mathbf{L}_n'\, dv(\mathbf{X}),\end{aligned}$$

using (53) and (50). The Jacobian matrix is therefore

$$\frac{\partial v(\mathbf{F}(\mathbf{X}))}{\partial(v(\mathbf{X}))'} = 2(\mathbf{D}_n'\mathbf{D}_n)^{-1}(\mathbf{L}_n(\mathbf{I}\otimes\mathbf{X})\mathbf{D}_n)',$$

and its determinant is the Jacobian of the transformation. The determinant is

$$\mathbf{J}_{\mathbf{F}}(\mathbf{X}) = \left|\frac{\partial v(\mathbf{F}(\mathbf{X}))}{\partial(v(\mathbf{X}))'}\right| = 2^n \prod_{i=1}^{n} x_{ii}^{i}, \tag{105}$$

using (62) and Lemma 4.1(iii) of Magnus and Neudecker [18].

Historical note. A variety of methods has been used to account for the symmetry in the evaluation of Jacobians of transformations involving symmetric matrix arguments, notably differential techniques (Deemer and Olkin [6] and Olkin [27]; induction (Jack [12]), and functional equations induced on the relevant spaces (Olkin and Sampson [28]). Our approach finds its root in Tracy and Singh [39] who used modified matrix differentiation results to obtain Jacobians in a simple fashion. Many further Jacobians of transformations with symmetric or lower triangular matrix arguments can be found in Magnus and Neudecker [18]; the matrix $\mathbf{L}_n$ introduced in example (v) is their so-called elimination matrix. Neudecker [25] obtained Jacobians of transformations with skew-symmetric, strictly lower triangular, or diagonal matrix arguments, using the approach described here. We emphasize however that ours is by no means the only approach to evaluate Jacobians. In particular when the matrix argument is *not* L-structured, but say orthogonal, other methods are called for. Tensor and exterior algebra (and calculus) deal well with such and other cases (see Muirhead [20, p. 50–72] for an introduction to the subject), and have recently been extensively used by Phillips [29, 30] and Hillier [11].

ACKNOWLEDGMENTS

We are grateful to the editor and referee for their positive and constructive remarks, and to the Netherlands Organization for the Advancement of Pure Research (Z.W.O.) for financial support.

NOTES

1. Let $\mathbf{A}$ be an idempotent symmetric $n \times n$ matrix, and $\mathbf{B}$ an $n \times r$ matrix such that $\mathbf{AB} = \mathbf{B}$. Then $\mathbf{A} = \mathbf{BB}^+$ if, and only if, $\mathbf{A}$ and $\mathbf{B}$ have the same rank. [*Proof*: Let $\mathbf{C} = \mathbf{A} - \mathbf{BB}^+$. Then $\mathbf{CB} = 0$ and C is symmetric and idempotent. Hence $r(\mathbf{C}) = r(\mathbf{A}) - r(\mathbf{B})$ and the result follows.]

2. Since $j \leqslant i$ and $t \leqslant s$, we have to consider the following six cases: $t \leqslant s \leqslant j \leqslant i$, $t \leqslant j \leqslant s \leqslant i$, $t \leqslant j \leqslant i \leqslant s$, $j \leqslant t \leqslant s \leqslant i$, $j \leqslant t \leqslant i \leqslant s$, $j \leqslant i \leqslant t \leqslant s$. But since $\mathbf{A}$ is upper triangular, the last four of these yield zero or become special cases of the first two. Hence the only cases not yielding zero are $t \leqslant s \leqslant j \leqslant i$ and $t \leqslant j \leqslant s \leqslant i$, or equivalently $t \leqslant s \leqslant j = i$, $t \leqslant s \leqslant j < i$, and $t \leqslant j < s \leqslant i$.

3. The statement that (81) is the *only* sensible definition is perhaps too strong. Of course, if $\mathbf{F}(\mathbf{X})$ is a differentiable matrix function, then $\mathbf{P}\{\partial \operatorname{vec} \mathbf{F}(\mathbf{X})/\partial(\operatorname{vec} \mathbf{X})'\}\mathbf{Q}$, where $\mathbf{P}$ and $\mathbf{Q}$ are arbitrary permutation matrices, can also be taken as the definition of the Jacobian matrix of $\mathbf{F}$ at $\mathbf{X}$. This simply corresponds to a different ordering of the functions f_{ij} ($\mathbf{P} \operatorname{vec} \mathbf{F}$ instead of $\operatorname{vec} \mathbf{F}$) and the variables x_{st} ($\mathbf{Q}' \operatorname{vec} \mathbf{X}$ instead of $\operatorname{vec} \mathbf{X}$).

REFERENCES

1. Aitken, A. C. On the Wishart distribution in statistics. *Biometrika* 36 (1949): 59–62.
2. Balestra, P. *La dérivation matricielle.* Collection de l'Institut de Mathématiques Economiques, No. 12. Paris: Sirey, 1976.
3. Barnett, S. Matrix differential equations and Kronecker products. *SIAM Journal of Applied Mathematics* 24 (1973): 1–5.
4. Bellman, R. *Introduction to matrix analysis*, 2nd edition. New York: McGraw-Hill, 1970.
5. Browne, M. W. Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal* 8 (1974): 1–24.
6. Deemer, W. L. & I. Olkin. The Jacobians of certain matrix transformations useful in multivariate analysis. *Biometrika* 38 (1951): 345–367.
7. Don, F.J.H. & A. P. van der Plas. A note on the characteristic polynomial of the commutation matrix. *Linear Algebra and Its Applications* 37 (1981): 135–142.
8. Hartwig, R. E. & S. B. Morris. The universal flip matrix and the generalized faro-shuffle. *Pacific Journal of Mathematics* 58 (1975): 445–455.
9. Henderson, H. V. & S. R. Searle. Vec and vech operators for matrices, with some uses in Jacobians and multivariate statistics. *The Canadian Journal of Statistics* 7 (1979): 65–81.
10. Henderson, H. V. & S. R. Searle. The vec-permutation matrix, the vec operator and Kronecker products: a review. *Linear and Multilinear Algebra* 9 (1981): 271–288.
11. Hillier, G. H. On the joint and marginal densities of instrumental variable estimators in a general structural equation. *Econometric Theory* 1 (1985): 53–72.
12. Jack, H. Jacobians of transformations involving orthogonal matrices. *Proceedings of the Royal Society of Edinburgh* (Section A) 67 (1966): 81–103.
13. Koopmans, T. C., H. Rubin, & R. B. Leipnik. Measuring the equation systems of dynamic economics. In T. C. Koopmans (ed.), *Statistical inference in dynamic economic models.* New York: John Wiley, 1950.
14. Ledermann, W. On singular pencils of Zehfuss, compound, and Schläflian matrices. *Proceedings of the Royal Society of Edinburgh* 56 (1936): 50–89.
15. MacDuffee, C. C. *The theory of matrices.* Berlin, 1933, Reprinted by Chelsea, New York.
16. Magnus, J. R. L-structured matrices and linear matrix equations. *Linear and Multilinear Algebra* 14 (1983): 67–88.
17. Magnus, J. R. & H. Neudecker. The commutation matrix: some properties and applications. *The Annals of Statistics* 7 (1979): 381–394.
18. Magnus, J. R. & H. Neudecker. The elimination matrix: some lemmas and applications. *SIAM Journal on Algebraic and Discrete Methods* 1 (1980): 422–449.
19. Magnus, J. R. & H. Neudecker. Matrix differential calculus with applications to simple, Hadamard, and Kronecker products. *Journal of Mathematical Psychology* 29 (1985): 474–492.
20. Muirhead, R. J. *Aspects of multivariate statistical theory.* New York: John Wiley, 1982.
21. Murnaghan, F. D. *The theory of group representations.* Baltimore: Johns Hopkins Press, 1938.
22. Nel, D. G. On the symmetric multivariate normal distribution and the asymptotic expansion of a Wishart matrix. *South African Statistical Journal* 12 (1978): 145–159.
23. Neudecker, H. The Kronecker matrix product and some of its applications in econometrics. *Statistica Neerlandica* 22 (1968): 69–81.

24. Neudecker, H. Some theorems on matrix differentiation with special reference to Kronecker matrix products. *Journal of the American Statistical Association* 64 (1969): 953–963.
25. Neudecker, H. On Jacobians of transformations with skew-symmetric, strictly (lower) triangular or diagonal matrix arguments. *Linear and Multilinear Algebra* 14 (1983): 271–295.
26. Neudecker, H. & T. Wansbeek. Some results on commutation matrices, with statistical applications. *The Canadian Journal of Statistics* 11 (1983): 221–231.
27. Olkin, I. Note on 'The Jacobians of certain matrix transformations useful in multivariate analysis'. *Biometrika* 40 (1953): 43–46.
28. Olkin, I. & A. P. Sampson. Jacobians of matrix transformations and induced functional equations. *Linear Algebra and Its Applications* 5 (1972): 257–276.
29. Phillips, P.C.B. The exact distribution of LIML: I. *International Economic Review* 25 (1984): 249–261.
30. Phillips, P.C.B. The exact distribution of LIML: II. *International Economic Review* 26 (1985): 21–36.
31. Pollock, D.S.G. *The algebra of econometrics.* New York: John Wiley, 1979.
32. Pollock, D.S.G. Tensor products and matrix differential calculus. *Linear Algebra and Its Applications* 67 (1985): 169–193.
33. Richard, J. F. A note on the information matrix of the multivariate normal distribution. *Journal of Econometrics* 3 (1975): 57–60.
34. Rogers, G. S. *Matrix derivatives.* New York: Marcel Dekker, 1980.
35. Roth, W. E. On direct product matrices. *Bulletin of the American Mathematical Society* 40 (1934): 461–468.
36. Sylvester, J. Sur la solution du cas le plus général des équations linéaires en quantités binaires, c'est-à-dire en quaternions ou en matrices du second ordre. *Comptes Rendus des Séances de l'Académie des Sciences* 99 (1884): 117–118.
37. Sylvester, J. Sur la resolution générale de l'équation linéaire en matrices d'un ordre quelconque. *Comptes Rendus des Séances de l'Académie des Sciences* 99 (1884): 409–412 and 432–436.
38. Tracy, D. S. & P. S. Dwyer. Multivariate maxima and minima with matrix derivatives. *Journal of the Americal Statistical Association* 64 (1969): 1576–1594.
39. Tracy, D. S. & R. P. Singh. Some modifications of matrix differentiation for evaluating Jacobians of symmetric matrix transformations. In D.S. Tracy (ed.), *Symmetric functions in statistics.* University of Windsor, Ontario, Canada, 1972.
40. Vartak, N. N. On an application of Kronecker product of matrices to statistical designs. *Annals of Mathematical Statistics* 26 (1955): 420–438.
41. Vetter, W. J. Vector structures and solutions of linear matrix equations. *Linear Algebra and Its Applications* 10 (1975): 181–188.
42. Wiens, D. P. On some pattern-reduction matrices which appear in statistics. *Linear Algebra and Its Applications* 67 (1985): 233–258.
43. Zehfuss, G. Ueber eine gewisse determinante. *Zeitschrift für Mathematik und Physik* 3 (1858): 298–301.