FACULTY OF ACTUARIAL SCIENCE & ECONOMETRICS

A & E REPORT

REPORT AE 1/85

SYMMETRY, 0-1 MATRICES, AND JACOBIANS: A REVIEW

J.R. Magnus and H. Neudecker

University of Amsterdam

Title: SYMMETRY, 0-1 MATRICES, AND JACOBIANS: A REVIEW

Authors: J.R. Magnus and H. Neudecker

Address: University of Amsterdam
Faculty of Actuarial Science and Econometrics
Jodenbreestraat 23
1011 NH Amsterdam

Date: January 1985

Series and Number: AE Report 1/85

Pages: 46

Price: No charge

JEL Subject Classification: 213

Keywords: Kronecker product; vec operator; commutation and duplication matrix

Abstract:

The purpose of the 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 vec operators is fairly exhaustive; the treatment of 0-1 matrices is (deliberately) selective.

Corresponding Author

Jan R. Magnus
London School of Economics
Department of Economics
Houghton Street
London WC2A 2AE
United Kingdom

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 vec operators 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 K_{mn} is introduced as the matrix which transforms vec A into vec A' for any m×n matrix A. Its algebraic properties are discussed in section 4, and its role in normal distribution theory in section 5. Closely related to the commutation matrix is the symmetric idempotent matrix N_n defined as $N_n = \frac{1}{2}(I_{n^2} + K_{nn})$, whose main properties are obtained in section 6. If A is a symmetric n×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 A, this defines a new vector which we denote as v(A). The matrix which transforms, for symmetric A, v(A) into vec A is the duplication matrix 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. Finally, in sections 9-11, 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. We also define the Hessian matrix. If the Jacobian matrix is square, its determinant is the Jacobian of the transformation. Some examples show that the evaluation of Jacobian matrices, Hessian matrices, and Jacobians can be short, elegant, and easy, also if the transformations involve symmetric (or L-structured) matrix arguments.

The historical references in sections 2-4 are taken from Henderson and Searle's (1981) 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; A' denotes the transpose of A, A^+ its Moore-Penrose inverse, and $r(A)$ its rank; if A is square, $\operatorname{tr} A$ denotes its trace, $|A|$ its determinant, and A^{-1} its inverse (when A is non-singular). $\mathbb{R}^{m \times n}$ 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 I_n. Mathematical expectation is denoted by E; variance (variance-covariance matrix) by var.

2. The Kronecker product

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

$$\begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{pmatrix} \qquad (2.1)$$

is called the Kronecker product of A and B and written $A \otimes B$.

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

$$A \otimes B \otimes C = (A \otimes B) \otimes C = A \otimes (B \otimes C); \qquad (2.2)$$

$$(A + B) \otimes (C + D) = A \otimes C + A \otimes D + B \otimes C + B \otimes D, \qquad (2.3)$$

if $A + B$ and $C + D$ exist; and

$$(A \otimes B)(C \otimes D) = AC \otimes BD, \qquad (2.4)$$

if AC and BD exist.

If α is a scalar, then

$$\alpha \otimes A = \alpha A = A\alpha = A \otimes \alpha. \qquad (2.5)$$

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

$$a' \otimes b = ba' = b \otimes a'. \qquad (2.6)$$

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

$$(A \otimes B)' = A' \otimes B', \qquad (A \otimes B)^+ = A^+ \otimes B^+. \tag{2.7}$$

If A and B are square matrices (not necessarily of the same order), then

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

(Of course, the trace of $A \otimes B$ may exist even when A and B are not square matrices; in that case the expression for $\operatorname{tr}(A \otimes B)$ is more complicated.[1]) If A and B are nonsingular, then

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}. \tag{2.9}$$

(The nonsingularity of A and B is not only sufficient, but also necessary for the nonsingularity of $A \otimes B$; this follows from rank considerations, see (2.11).)

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

Lemma 2.1. Let A be an $m \times m$ matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$, and let B be a $p \times p$ matrix with eigenvalues $\mu_1, \mu_2, \ldots, \mu_p$. Then the mp eigenvalues of $A \otimes B$ are $\lambda_i \mu_j$ $(i = 1,\ldots,m,\ j = 1,\ldots,p)$.

Proof. By Schur's Theorem (Bellman (1970, p.202)) there exist nonsingular (in fact, unitary) matrices S and T such that

$$S^{-1}AS = L, \qquad T^{-1}BT = M,$$

---

1) See Neudecker and Wansbeek (1983, Theorem 3.2).

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

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

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

Remark. If x is an eigenvector of A and y an eigenvector of B, then $x \otimes y$ is clearly an eigenvector of $A \otimes B$. It is not generally true, however, that every eigenvector of $A \otimes B$ is the Kronecker product of an eigenvector of A and an eigenvector of B, as the following example shows.[2] Let

$$A = B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Both eigenvalues of A (and B) are zero and the only eigenvector is e_1. The four eigenvalues of $A \otimes B$ are all zero (in concordance with Lemma 2.1), but the eigenvectors of $A \otimes B$ are not just $e_1 \otimes e_1$, but also $e_1 \otimes e_2$ and $e_2 \otimes e_1$.

Lemma 2.1 has several important corollaries. First, if A and B are positive (semi)definite, then $A \otimes B$ is positive (semi)definite. Secondly, since the determinant of $A \otimes B$ is equal to the product of its eigenvalues, we obtain

---

2) This fact is emphasised here because it is often stated incorrectly, see e.g. Bellman (1970, p.235).

$$|A \otimes B| = |A|^p \, |B|^m , \qquad (2.10)$$

where A is an m×m matrix and B is a p×p matrix. Thirdly, we can obtain the rank of $A \otimes B$ from Lemma 2.1 as follows. The rank of $A \otimes B$ is equal to the rank of $AA' \otimes BB'$. 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 2.1, the eigenvalues of $AA' \otimes BB'$ are $\lambda_i \mu_j$, where λ_i are the eigenvalues of AA' and μ_j are the eigenvalues of 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 $AA' \otimes BB'$ is the product of the number of nonzero eigenvalues of AA' and the number of nonzero eigenvalues of BB'. Thus the rank of $A \otimes B$ is

$$r(A \otimes B) = r(A)\, r(B). \qquad (2.11)$$

<u>Historical note</u>. The original interest in the Kronecker product focussed on the determinantal result (2.10), which seems to have been first studied by Zehfuss (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.

3. The vec-operator

Let A be an m×n matrix and a_j its j^{th} column, then vec A is the mn×1 vector

$$\text{vec } A = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} . \tag{3.1}$$

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

A very simple but often useful property is

$$\text{vec } a' = \text{vec } a = a \tag{3.2}$$

for any column-vector a. The basic connection between the vec-operator and the Kronecker product is

$$\text{vec } ab' = b \otimes a \tag{3.3}$$

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

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

$$(\text{vec } A)'\text{vec } B = \text{tr } A'B , \tag{3.4}$$

where A and B are matrices of the same order. This is easy to verify since both the left side and the right side of (3.4) are equal to $\sum_i \sum_j a_{ij} b_{ij}$.

Let us now generalize the basic properties (3.3) and (3.4). The generalization of (3.3) is the following well-known result.

Lemma 3.1. Let A, B, and C be three matrices such that the matrix product ABC is defined. Then,

$$\text{vec } ABC = (C' \otimes A) \text{ vec } B. \tag{3.5}$$

Proof. Assume that 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×q identity matrix I_q, so that $B = \sum_{j=1}^{q} b_j e_j'$. Then, using (3.3),

$$\begin{aligned} \text{vec } ABC &= \text{vec} \sum_{j=1}^{q} Ab_j e_j' C = \sum_{j=1}^{q} \text{vec } (Ab_j)(C'e_j)' \\ &= \sum_{j=1}^{q} (C'e_j \otimes Ab_j) = (C' \otimes A) \sum_{j=1}^{q} (e_j \otimes b_j) \\ &= (C' \otimes A) \sum_{j=1}^{q} \text{vec } b_j e_j' = (C' \otimes A) \text{ vec } B. \quad \| \end{aligned}$$

One special case of Lemma 3.1 is

$$\text{vec } AB = (B' \otimes I_m) \text{ vec } A = (B' \otimes A) \text{ vec } I_n = (I_q \otimes A) \text{ vec } B, \tag{3.6}$$

where A is an m×n matrix and B is an n×q matrix. Another special case arises when the matrix C in (3.5) is replaced by a vector. Then we obtain, using (3.2),

$$ABd = (d' \otimes A) \text{ vec } B = (A \otimes d') \text{ vec } B', \tag{3.7}$$

where d is a q×1 vector.

The equality (3.4) can be generalized as follows.

Lemma 3.2. Let A, B, C, and D be four matrices such that the matrix product ABCD is defined and square. Then,

$$\operatorname{tr} ABCD = (\operatorname{vec} D')'(C' \otimes A) \operatorname{vec} B = (\operatorname{vec} D)'(A \otimes C') \operatorname{vec} B'. \quad (3.8)$$

Proof. We have

$$\begin{aligned} \operatorname{tr} ABCD &= \operatorname{tr} D(ABC) = (\operatorname{vec} D')' \operatorname{vec} ABC && \text{(by (3.4))} \\ &= (\operatorname{vec} D')'(C' \otimes A) \operatorname{vec} B && \text{(by (3.5))}. \end{aligned}$$

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

Historical note. The idea of stacking the elements of a matrix in a vector goes back at least to Sylvester (1884a,b). The notation "vec" was introduced by Koopmans, Rubin and Leipnik (1950). Lemma 3.1 is due to Roth (1934).

4. The commutation matrix K_{mn}

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

$$K_{mn} \text{ vec } A = \text{vec } A'. \tag{4.1}$$

Since K_{mn} is a permutation matrix it is orthogonal, i.e., $K'_{mn} = K^{-1}_{mn}$. Also, premultiplying (4.1) by K_{nm} gives $K_{nm}K_{mn}$ vec A = vec A so that $K_{nm}K_{mn} = I_{mn}$. Hence,

$$K'_{mn} = K^{-1}_{mn} = K_{nm}. \tag{4.2}$$

Further, using (3.2),

$$K_{n1} = K_{1n} = I_n. \tag{4.3}$$

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 4.1. Let A be an m×n matrix and B a p×q matrix. Then

$$K_{pm}(A \otimes B) = (B \otimes A)K_{qn}. \tag{4.4}$$

Proof. Let X be an arbitrary q×n matrix. Then, by repeated application of (3.5) and (4.1),

$$K_{pm}(A \otimes B) \text{ vec } X = K_{pm} \text{ vec } BXA' = \text{vec } AX'B'$$

$$= (B \otimes A) \text{ vec } X' = (B \otimes A)K_{qn} \text{ vec } X.$$

Since X is arbitrary the result follows. ||

Immediate consequences of Lemma 4.1 are

$$K_{pm}(A \otimes B)K_{nq} = B \otimes A \tag{4.5}$$

and

$$K_{pm}(A \otimes b) = b \otimes A, \quad K_{mp}(b \otimes A) = A \otimes b, \tag{4.6}$$

where b is a p×1 vector.

All these properties follow from the implicit definition (4.1) of the commutation matrix. The following lemma gives an explicit expression for K_{mn} which is often useful.

<u>Lemma 4.2</u>. Let H_{ij} be the m×n matrix with 1 in its ij^{th} position and zeroes elsewhere. Then

$$K_{mn} = \sum_{i=1}^{m} \sum_{j=1}^{n} (H_{ij} \otimes H'_{ij}). \tag{4.7}$$

<u>Proof</u>. Let X be an arbitrary m×n matrix. Let e_i denote the i^{th} column of I_m and u_j the j^{th} column of I_n, so that $H_{ij} = e_i u'_j$. Then

$$X' = I_n X' I_m = (\sum_{j=1}^{n} u_j u'_j)\, X' (\sum_{i=1}^{m} e_i e'_i)$$

$$= \sum_{ij} u_j (u'_j X' e_i) e'_i = \sum_{ij} u_j (e'_i X u_j) e'_i$$

$$= \sum_{ij} (u_j e'_i) X (u_j e'_i) = \sum_{ij} H'_{ij} X H'_{ij} .$$

Taking vecs we obtain

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

using (3.5). The result follows. ||

Lemma 4.2 shows that K_{mn} is a square matrix of order mn, partitioned into mn submatrices each of order n×m, such that the ij[th] submatrix has unity in its ji[th] position and zeros elsewhere. For example,

$$K_{23} = \left(\begin{array}{cc|cc|cc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right) . \qquad (4.8)$$

The explicit form (4.7) of K_{mn} enables us to find the trace and the determinant of K_{mn}.

<u>Lemma 4.3.</u>[3] The trace of the commutation matrix is

$$\operatorname{tr} K_{mn} = 1 + \gcd(m-1,n-1) , \qquad (4.9)$$

where gcd(m,n) is the greatest common divisor of m and n; its determinant is

$$|K_{mn}| = (-1)^{\frac{1}{4}mn(m-1)(n-1)} . \qquad (4.10)$$

<u>Proof</u>. We shall only prove the case $m = n$. (For a proof of the more difficult case $m \neq n$, see Magnus and Neudecker (1979, Theorem 3.1.) Let e_j be the j^{th} column of I_n. Then, from (4.7),

$$\operatorname{tr} K_{nn} = \operatorname{tr} \sum_{i=1}^{n} \sum_{j=1}^{n} (e_i e_j' \otimes e_j e_i') = \sum_{ij} \operatorname{tr} (e_i e_j' \otimes e_j e_i')$$

$$= \sum_{ij} (\operatorname{tr} e_i e_j')(\operatorname{tr} e_j e_i') = \sum_{ij} \delta_{ij}^2 = n ,$$

---

3) For a discussion of the characteristic polynomial of the commutation matrix, see Hartwig and Morris (1975) and Don and van der Plas (1981).

where $\delta_{ij} = 0$ if $i \neq j$, $\delta_{ii} = 1$. Since K_{nn} is real, orthogonal, and symmetric, it has eigenvalues +1 and -1 only. (The eigenvalues of 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

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

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

$$|K_{nn}| = (-1)^p = (-1)^{\frac{1}{2}n(n-1)} = (-1)^{\frac{1}{4}n^2(n-1)^2}. \quad \|$$

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.

<u>Lemma 4.4</u>. Let A be an $m \times n$ matrix and B a $p \times q$ matrix. Then

$$\operatorname{vec}(A \otimes B) = (I_n \otimes K_{qm} \otimes I_p)(\operatorname{vec} A \otimes \operatorname{vec} B). \qquad (4.11)$$

<u>Proof</u>. Let a_i $(i = 1,\ldots,n)$ and b_j $(j = 1,\ldots,q)$ denote the columns of A and B, respectively. Also, let e_i $(i = 1,\ldots,n)$ and u_j $(j = 1,\ldots,q)$ denote the columns of I_n and I_q, respectively. Then we can write A and B as

$$A = \sum_{i=1}^{n} a_i e_i', \qquad B = \sum_{j=1}^{q} b_j u_j',$$

and we obtain

$$\begin{aligned}\operatorname{vec}(A \otimes B) &= \sum_{i=1}^{n}\sum_{j=1}^{q} \operatorname{vec}(a_i e_i' \otimes b_j u_j') \\ &= \sum_{ij} \operatorname{vec}(a_i \otimes b_j)(e_i \otimes u_j)' = \sum_{ij} (e_i \otimes u_j \otimes a_i \otimes b_j) \\ &= \sum_{ij} (I_n \otimes K_{qm} \otimes I_p)(e_i \otimes a_i \otimes u_j \otimes b_j)\end{aligned}$$

$$= (I_n \otimes K_{qm} \otimes I_p)\{(\sum_i \text{vec } a_i e_i') \otimes (\sum_j \text{vec } b_j u_j')\}$$

$$= (I_n \otimes K_{qm} \otimes I_p)(\text{vec } A \otimes \text{vec } B). \quad \|$$

In particular, by noting that

$$\text{vec } A \otimes \text{vec } B = (I_{nm} \otimes \text{vec } B)\text{vec } A = (\text{vec } A \otimes I_{qp}) \text{ vec } B,$$

using (2.5), we obtain

$$\text{vec } (A \otimes B) = (I_n \otimes G) \text{ vec } A = (H \otimes I_p) \text{ vec } B, \qquad (4.12)$$

where

$$G = (K_{qm} \otimes I_p)(I_m \otimes \text{vec } B), \quad H = (I_n \otimes K_{qm})(\text{vec } A \otimes I_q). \qquad (4.13)$$

Historical note. The original interest in the commutation matrix focussed on its role in reversing ("commuting") the order of Kronecker products (Lemma 4.1), a role which seems to have been first recognized byLedermann (1936) and Murnaghan (1938, pp.68-69), while Vartak (1955) generalized Murnaghan's result to rectangular matrices. Tracy and Dwyer (1969) rediscovered the commutation matrix and based their definition on the fact that K_{mn} is the matrix obtained by rearranging the rows of I_{mn} by taking every m^{th} row starting with the first, then every m^{th} row starting with the second, and so on. (For example, the rows of K_{23} are rows 1,3,5,2,4, and 6 of I_6.) The fruitful idea of defining K_{mn} by its transformation property (4.1) comes from Barnett (1973), and is the definition adopted in this paper.

Among the many alternative names of the commutation matrix we mention permutation matrix, permuted identity matrix, vec-permutation matrix, shuffle matrix, and universal flip matrix. Alternative notations for K_{mn} include $E_{m,n}$, $E^{n\times m}_{m\times n}$, $U_{m\times n}$, $P_{n,m}$, $I_{(n,m)}$, and $I_{n,m}$.

Lemma 4.1 goes back at least to Ledermann (1936). Concise proofs are given by Barnett (1973), Hartwig and Morris (1975), and Magnus and Neudecker (1979). Lemmas 4.2 and 4.3 are due to Magnus and Neudecker (1979), and Lemma 4.4 to Neudecker and Wansbeek (1983).

For further reading on the commutation matrix we recommend Hartwig and Morris (1975), Balestra (1976), Magnus and Neudecker (1979), Henderson and Searle (1981), and Neudecker and Wansbeek (1983).

5. 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 lemma.

Lemma 5.1. Let u be an $n\times 1$ vector of independent and standard normally distributed random variables $u_1,\ldots,u_n$, that is, $u \cong N(0,I_n)$. Then,

$$\operatorname{var}(u \otimes u) = I_{n^2} + K_{nn} . \qquad (5.1)$$

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

$$u'Au = u'Bu = u'T\Lambda T'u = v'\Lambda v = \sum_{i=1}^{n} \lambda_i v_i^2 .$$

Since $v \cong N(0,I_n)$, it follows that $v_1^2,\ldots,v_n^2$ are independently distributed with $\operatorname{var} v_i^2 = 2$, so that

$$\operatorname{var} u'Au = \operatorname{var} \sum_i \lambda_i v_i^2 = \sum_i \lambda_i^2 (\operatorname{var} v_i^2) = 2 \operatorname{tr} \Lambda^2$$

$$= 2 \operatorname{tr} B^2 = \operatorname{tr} A'A + \operatorname{tr} A^2 = (\operatorname{vec} A)'(I+K_{nn}) \operatorname{vec} A,$$

using (3.4). Also, since $u'Au = \operatorname{vec} u'Au = (u \otimes u)' \operatorname{vec} A$,

$$\operatorname{var} u'Au = \operatorname{var}((u \otimes u)' \operatorname{vec} A) = (\operatorname{vec} A)'(\operatorname{var} u \otimes u) \operatorname{vec} A.$$

Hence,

$$(\operatorname{vec} A)'(\operatorname{var} u \otimes u) \operatorname{vec} A = (\operatorname{vec} A)'(I+K_{nn}) \operatorname{vec} A$$

for every $n\times n$ matrix A. The result follows. ||

We can generalize Lemma 5.1 by considering normal random variables which are not necessarily independent or identically distributed. This leads to Lemma 5.2.

Lemma 5.2. Let $x \cong N(\mu,V)$ where V is a positive semidefinite $n\times n$ matrix. Then

$$\operatorname{var}(x \otimes x) = (I_{n^2}+K_{nn})(V \otimes V + V \otimes \mu\mu' + \mu\mu' \otimes V). \qquad (5.2)$$

Proof. We write $x = V^{\frac{1}{2}}u + \mu$ with $u \cong N(0,I_n)$, so that

$$\begin{aligned} x \otimes x &= V^{\frac{1}{2}}u \otimes V^{\frac{1}{2}}u + V^{\frac{1}{2}}u \otimes \mu + \mu \otimes V^{\frac{1}{2}}u + \mu \otimes \mu \\ &= (V^{\frac{1}{2}} \otimes V^{\frac{1}{2}})(u \otimes u) + (I + K_{nn})(V^{\frac{1}{2}}u \otimes \mu) + \mu \otimes \mu \\ &= (V^{\frac{1}{2}} \otimes V^{\frac{1}{2}})(u \otimes u) + (I + K_{nn})(V^{\frac{1}{2}} \otimes \mu)u + \mu \otimes \mu , \end{aligned}$$

using (4.6) and (2.5). Since the two vectors $u \otimes u$ and u are uncorrelated with $\operatorname{var}(u \otimes u) = I + K_{nn}$ and $\operatorname{var} u = I_n$, we obtain

$$\begin{aligned} \operatorname{var}(x \otimes x) &= \operatorname{var}\{(V^{\frac{1}{2}} \otimes V^{\frac{1}{2}})(u \otimes u)\} + \operatorname{var}\{(I + K_{nn})(V^{\frac{1}{2}} \otimes \mu)u\} \\ &= (V^{\frac{1}{2}} \otimes V^{\frac{1}{2}})(I+K_{nn})(V^{\frac{1}{2}} \otimes V^{\frac{1}{2}}) + (I+K_{nn})(V^{\frac{1}{2}} \otimes \mu)(V^{\frac{1}{2}} \otimes \mu)'(I+K_{nn}) \\ &= (I+K_{nn})(V \otimes V) + (I+K_{nn})\big(V \otimes \mu\mu' + K_{nn}(\mu\mu' \otimes V)\big) \\ &= (I+K_{nn})(V \otimes V + V \otimes \mu\mu' + \mu\mu' \otimes V) , \end{aligned}$$

using (4.4) and the fact (implied by (4.2)) that $(I + K_{nn})K_{nn} = I + K_{nn}$. ||

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

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

The joint distribution of the elements of the matrix

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

is said to be Wishart with k degrees of freedom and is denoted by $W_n(k,V,M)$, where M is the k×n matrix

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

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

Theorem 5.1. Let S be Wishart distributed $W_n(k,V,M)$, V positive semidefinite. Then,

$$ES = kV + M'M \tag{5.3}$$

and

$$\text{var}(\text{vec } S) = (I_{n^2} + K_{nn})\{k(V \otimes V) + V \otimes M'M + M'M \otimes V\}. \tag{5.4}$$

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

$$ES = E\sum_i y_i y_i' = \sum_i E y_i y_i' = \sum_i (V + \mu_i \mu_i') = kV + M'M ,$$

and

$$\begin{aligned} \text{var}(\text{vec } S) &= \text{var}(\text{vec} \sum_i y_i y_i') = \text{var}(\sum_i y_i \otimes y_i) = \sum_i \text{var}(y_i \otimes y_i) \\ &= \sum_i (I + K_{nn})(V \otimes V + V \otimes \mu_i \mu_i' + \mu_i \mu_i' \otimes V) \\ &= (I + K_{nn})\{k(V \otimes V) + V \otimes M'M + M'M \otimes V\} , \end{aligned}$$

using (5.2). ||

Historical Note. The results in this section are taken from Magnus and Neudecker (1979), but the proofs are somewhat simplified. More general results can be found in Magnus and Neudecker (1979) and Neudecker and Wansbeek (1983). In the latter paper it is shown, inter alia, that the normality assumption in Lemma 5.1 is not essential. More precisely, if u is an n×1 vector of independent random variables $u_1,\ldots,u_n$ with $Eu_i = 0$, $Eu_i^2 = \sigma^2$, $Eu_i^4 = \psi^4$, then

$$\text{var}(u \otimes u) = \sigma^4 (I_{n^2} + K_{nn} + \gamma \sum_{i=1}^{n} (E_{ii} \otimes E_{ii})), \tag{5.5}$$

where $\gamma = (\psi^4/\sigma^4) - 3$ (the kurtosis) and E_{ii} is the n×n matrix with 1 in its i^{th} diagonal position and zeros elsewhere.

6. The matrix N_n

In the previous section the $n^2 \times n^2$ matrix $I_{n^2} + K_{nn}$ played a central role. This matrix appears in many applications and we shall now study it in more detail. For reasons which will become apparent shortly it is convenient to investigate the properties not of $I + K_{nn}$, but of the $n^2 \times n^2$ matrix

$$N_n = \tfrac{1}{2}(I_{n^2} + K_{nn}) . \qquad (6.1)$$

This matrix, is symmetric and idempotent,

$$N_n = N_n' = N_n^2 , \qquad (6.2)$$

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

$$r(N_n) = \operatorname{tr} N_n = \tfrac{1}{2} n(n+1) . \qquad (6.3)$$

The matrix N_n transforms an arbitrary $n \times n$ matrix A into the symmetric matrix $\tfrac{1}{2}(A + A')$:

$$N_n \operatorname{vec} A = \operatorname{vec} \tfrac{1}{2}(A + A'). \qquad (6.4)$$

Of course, if A is symmetric to begin with, the transformation has no effect. (This shows again that N_n must be idempotent.)

Further properties of N_n include

$$N_n K_{nn} = N_n = K_{nn} N_n , \qquad (6.5)$$

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

$$N_n (A \otimes B) N_n = N_n (B \otimes A) N_n , \qquad (6.6)$$

$$N_n (A \otimes B + B \otimes A) N_n = N_n (A \otimes B + B \otimes A) = (A \otimes B + B \otimes A) N_n , \qquad (6.7)$$

$$N_n(A\otimes A)N_n = N_n(A\otimes A) = (A\otimes A)N_n\ , \tag{6.8}$$

and

$$N_n(A\otimes b) = N_n(b\otimes A) = \tfrac{1}{2}(A\otimes b + b\otimes A), \tag{6.9}$$

for any $n\times 1$ vector b.

The explicit form of N_n is easily derived from K_{nn}. For example, for $n=2$ and 3, we have

$$N_2 = \left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \hline 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{array}\right),\ N_3 = \left(\begin{array}{ccc|ccc|ccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ \hline 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \hline 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right). \tag{6.10}$$

Historical note. The matrix N_n was introduced by Magnus and Neudecker (1980, p.424).

7. Symmetry: the duplication matrix D_n

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

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

and

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

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

$$D_n v(A) = \text{vec } A \qquad (A = A') . \tag{7.1}$$

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

$$D_n^+ = (D_n'D_n)^{-1}D_n' . \tag{7.2}$$

For $n = 3$, we have

$$D_3 = \left(\begin{array}{ccc|cc|c} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right), \qquad D_3^{+'} = \left(\begin{array}{ccc|cc|c} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ \hline 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \\ \hline 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right) . \tag{7.3}$$

Some further properties of D_n are easily derived from its implicit definition 7.1. For symmetric A we have

$$K_{nn} D_n v(A) = K_{nn} \operatorname{vec} A = \operatorname{vec} A = D_n v(A)$$

and

$$N_n D_n v(A) = N_n \operatorname{vec} A = \operatorname{vec} A = D_n v(A).$$

Again, the symmetry of A does not restrict $v(A)$, so that

$$K_{nn} D_n = D_n = N_n D_n . \tag{7.4}$$

Also, from (7.2) and (7.4),

$$D_n^+ K_{nn} = D_n^+ = D_n^+ N_n . \tag{7.5}$$

Finally, we obtain

$$D_n^+ D_n = I_{\frac{1}{2}n(n+1)} , \quad D_n D_n^+ = N_n . \tag{7.6}$$

The first of the two equalities in (7.6) is an immediate consequence of (7.2), while the second follows from $N_n D_n = D_n$ (see (7.4)).[4)]

---

4) Let A be an idempotent, symmetric $n \times n$ matrix, and B an $n \times r$ matrix with full column-rank r. Then $AB = B$ if and only if $A = B(B'B)^{-1}B'$.

Much of the interest in the duplication matrix is due to the importance of the matrices $D_n^+(A \otimes A)D_n$ and $D_n'(A \otimes A)D_n$, whose properties we shall now investigate. We first prove Lemma 7.1.

<u>Lemma 7.1</u>. Let A be an n×n matrix. Then,

$$D_nD_n^+(A \otimes A)D_n = (A \otimes A)D_n, \tag{7.7}$$

$$D_n^+(A \otimes A)D_nD_n^+ = D_n^+(A \otimes A), \tag{7.8}$$

and, if A is nonsingular,

$$(D_n^+(A \otimes A)D_n)^{-1} = D_n^+(A^{-1} \otimes A^{-1})D_n . \tag{7.9}$$

<u>Proof</u>. The first two equalities follow from $D_nD_n^+ = N_n$, $N_n(A \otimes A) = (A \otimes A)N_n$, $N_nD_n = D_n$ and $D_n^+N_n = D_n^+$. (See (7.6), (6.8), (7.4), and (7.5).) The last equality follows by direct verification since

$$D_n^+(A \otimes A)D_nD_n^+(A^{-1} \otimes A^{-1})D_n = D_n^+(A \otimes A)(A^{-1} \otimes A^{-1})D_n = D_n^+D_n = I,$$

using (7.7) and (7.6). ||

In fact, the property $D_nD_n^+(A \otimes A)D_n = (A \otimes A)D_n$ for <u>arbitrary</u> square A is the Kronecker counterpart to $D_nD_n^+ \operatorname{vec} A = \operatorname{vec} A$ for symmetric A, just as the property $K_{nn}(A \otimes A) = (A \otimes A)K_{nn}$ is the Kronecker counterpart to $K_{nn} \operatorname{vec} A = A'$. We see this immediately if we let X be symmetric and substitute the symmetric matrix AXA' for X in $D_nD_n^+ \operatorname{vec} X = \operatorname{vec} X$, yielding

$$D_nD_n^+(A \otimes A)D_nv(X) = (A \otimes A)D_nv(X) .$$

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

Lemma 7.2. Let 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 $D_n^+(A \otimes A)D_n$ is also diagonal (upper triangular, lower triangular) with diagonal elements $a_{ii}a_{jj}$ $(1 \leq j \leq i \leq n)$.

Proof. Let E_{ij} be the $n \times n$ matrix with 1 in the ij^{th} position and zeros elsewhere, and define

$$T_{ij} = E_{ij} + E_{ji} - \delta_{ij}E_{ii} .$$

Then, for $i \geq j$,

$$D^+(A \otimes A)D\, v(E_{ij}) = D^+(A \otimes A)D\, v(T_{ij})$$

$$= D^+(A \otimes A) \operatorname{vec} T_{ij} = D^+ \operatorname{vec} AT_{ij}A' = v(AT_{ij}A'),$$

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

$$(v(E_{st}))'D^+(A \otimes A)D\, v(E_{ij}) = (v(E_{st}))'\, v(AT_{ij}A')$$

$$= (AT_{ij}A')_{st} = a_{si}a_{tj} + a_{sj}a_{ti} - \delta_{ij}a_{si}a_{ti} .$$

In particular, if A is upper triangular, we obtain

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

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

$$(v(E_{ij}))' D^+ (A \otimes A) D\, v(E_{ij}) = a_{ii} a_{jj} \qquad (j \leq i)$$

are its diagonal elements. The case where A is lower triangular is proved similarly. The case where A is diagonal follows as a special case. ||

Lemma 7.2 is instrumental in proving our main result concerning the matrix $D_n^+(A \otimes A)D$.

<u>Theorem 7.1</u>. Let A be an n×n matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Then the eigenvalues of the matrix $D_n^+(A \otimes A)D_n$ are $\lambda_i \lambda_j$ $(1 \leq i \leq j \leq n)$, and its trace and determinant are given by

$$\operatorname{tr}\,(D_n^+(A \otimes A)D_n) = \tfrac{1}{2}\operatorname{tr} A^2 + \tfrac{1}{2}(\operatorname{tr} A)^2 \qquad (7.10)$$

and

$$|D_n^+(A \otimes A)D_n| = |A|^{n+1} . \qquad (7.11)$$

<u>Proof</u>. By Schur's Theorem (Bellman (1970, p.202)) there exists a nonsingular matrix S such that $S^{-1}AS = M$, where M is an upper triangular matrix with the eigenvalues $\lambda_1, \ldots, \lambda_n$ of A on its diagonal. Thus

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

Since $D_n^+(S^{-1} \otimes S^{-1})D_n$ is the inverse of $D_n^+(S \otimes S)D_n$ (see (7.9)), it follows that $D_n^+(A \otimes A)D_n$ and $D_n^+(M \otimes M)D_n$ have the same set of eigenvalues. By Lemma 7.2, the latter matrix is upper triangular with eigenvalues (diagonal elements) $\lambda_i \lambda_j$ $(1 \leq j \leq i \leq n)$. These are therefore the eigenvalues of $D_n^+(A \otimes A)D_n$ too.

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

$$\operatorname{tr} D_n^+(A \otimes A)D_n = \sum_{i \geq j} \lambda_i \lambda_j = \tfrac{1}{2} \sum_i \lambda_i^2 + \tfrac{1}{2} \sum_{ij} \lambda_i \lambda_j$$

$$= \tfrac{1}{2} \operatorname{tr} A^2 + \tfrac{1}{2}(\operatorname{tr} A)^2,$$

and

$$|D_n^+(A \otimes A)D_n| = \prod_{i \geq j} \lambda_i \lambda_j = \prod_i \lambda_i^{n+1} = |A|^{n+1} . \quad \|$$

Let us now establish the nature of the nonsingular $\tfrac{1}{2}n(n+1) \times \tfrac{1}{2}n(n+1)$ matrix $D_n'D_n$. Let $B = (b_{ij})$ and $C = (c_{ij})$ be arbitrary symmetric $n \times n$ matrices, and let E_{ii} be the $n \times n$ matrix with 1 in the i^{th} diagonal position and zeros elsewhere. Then

$$(v(B))'D_n'D_n v(C) = (\operatorname{vec} B)' \operatorname{vec} C = \sum_{ij} b_{ij} c_{ij}$$

$$= 2 \sum_{i \geq j} b_{ij} c_{ij} - \sum_i b_{ii} c_{ii}$$

$$= 2(v(B))'v(C) - \sum_i \left((v(B))'v(E_{ii})\right)\left((v(E_{ii}))'v(C)\right)$$

$$= (v(B))'\left\{2I - \sum_i v(E_{ii})(v(E_{ii}))'\right\}v(C),$$

so that

$$D_n'D_n = 2\, I_{\frac{1}{2}n(n+1)} - \sum_{i=1}^{n} v(E_{ii})(v(E_{ii}))'. \quad (7.12)$$

Hence, $D_n'D_n$ is a diagonal matrix with diagonal elements 1 (n times) and 2 ($\tfrac{1}{2}n(n-1)$ times) and determinant

$$|D_n'D_n| = 2^{\frac{1}{2}n(n-1)} . \quad (7.13)$$

With the help cf (7.13) we can now prove Lemma 7.3.

Lemma 7.3. Let A be an n×n matrix. Then

$$|D_n'(A \otimes A)D_n| = 2^{\frac{1}{2}n(n-1)}|A|^{n+1} , \qquad (7.14)$$

and, if A is nonsingular,

$$(D_n'(A \otimes A)D_n)^{-1} = D_n^{+}(A^{-1} \otimes A^{-1})D_n^{+'} . \qquad (7.15)$$

Proof. Since, from (7.7),

$$D_n'(A \otimes A)D_n = (D_n'D_n)(D_n^{+}(A \otimes A)D_n) ,$$

(7.14) follows from (7.13) and (7.11), and (7.15) follows from (7.9) and (7.2). ||

Historical note. The idea of putting into a single vector just the distinct elements of a symmetric matrix goes back at least to Aitken (1949). Properties of the duplication matrix were studied, inter alia, by Tracy and Singh (1972), Browne (1974), Vetter (1975), Richard (1975), Balestra (1976), Nel (1978), and Henderson and Searle (1979). For further properties of the duplication matrix the reader should consult Magnus and Neudecker (1980).

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 $\mathcal{D}$ 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 $\mathcal{D}$. The $mn \times s$ matrix

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

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

$$L(\Delta) = \{X \mid X \in \mathbb{R}^{m\times n}, \text{ vec } X \in \mathcal{D}\} \tag{8.1}$$

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 $\mathcal{D}$, then so is ΔE for any nonsingular E. This fact suggests that it might be more appropriate to regard L as function of $\mathcal{D}$ 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 Definition (8.1) 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 D_n. Other examples of L-Structures are (strictly) triangular, skew-symmetric, diagonal, circulant, and Toeplitz matrices.

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

$$\Delta\psi(A) = \text{vec}\, A\,. \tag{8.2}$$

Since Δ has full column-rank s, we obtain

$$\Delta^{+}\Delta = I_s\,, \tag{8.3}$$

which implies that $\psi(A)$ can be solved uniquely from (8.2), the unique solution being

$$\psi(A) = \Delta^{+}\text{vec}\, A \qquad (A \in L(\Delta))\,. \tag{8.4}$$

Thus, given the choice of Δ, ψ is uniquely determined by (8.2). (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. One may verify that for <u>arbitrary</u> A, $D^{+}\text{vec}\, A = \frac{1}{2} v(A+A')$; for <u>symmetric</u> A this becomes $D^{+}\text{vec}\, A = v(A)$.

Of special interest is the symmetric idempotent $s\times s$ matrix N_Δ defined as

$$N_\Delta = \Delta\Delta^{+}\,. \tag{8.5}$$

(In the case of symmetry, this is the matrix N.) If we substitute $\Delta^{+}\text{vec}\, A$ for $\psi(A)$ in (8.2), we obtain

$$N_\Delta\,\text{vec}\, A = \Delta\Delta^{+}\text{vec}\, A = \text{vec}\, A\,. \tag{8.6}$$

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

$$(\Delta E)(\Delta E)^{+} = (\Delta E)^{+'}(\Delta E)' = (\Delta E)^{+'}E'\Delta'$$

$$= (\Delta E)^{+'}E'\Delta'\Delta\Delta^{+} = (\Delta E)^{+'}(\Delta E)'\Delta EE^{-1}\Delta^{+}$$

$$= (\Delta E)(\Delta E)^{+}(\Delta E)E^{-1}\Delta^{+} = \Delta EE^{-1}\Delta^{+} = \Delta\Delta^{+} .$$

Hence N_Δ is invariant to the choice of Δ.

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

$$BXA' \in L(\Delta) \tag{8.7}$$

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

$$\Delta\Delta^{+}(A \otimes B)\Delta = (A \otimes B)\Delta , \tag{8.8}$$

and, if A and B are nonsingular,

$$(\Delta^{+}(A \otimes B)\Delta)^{-1} = \Delta^{+}(A^{-1} \otimes B^{-1})\Delta \tag{8.9}$$

and

$$(\Delta'(A \otimes B)\Delta)^{-1} = \Delta^{+}(A^{-1} \otimes B^{-1})\Delta^{+'} . \tag{8.10}$$

To prove (8.8), let $X \in L(\Delta)$. Then

$$\Delta\Delta^{+}(A \otimes B)\Delta\psi(X) = \Delta\Delta^{+}(A \otimes B)\,\mathrm{vec}\,X = \Delta\Delta^{+}\,\mathrm{vec}\,BXA'$$

$$= \mathrm{vec}\,BXA' = (A \otimes B)\,\mathrm{vec}\,X = (A \otimes B)\Delta\psi(X) .$$

The restriction $X \in L(\Delta)$ does not restrict $\psi(X)$; hence (8.8) follows. Property (8.8) together with (8.3) implies (8.9), since

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

while (8.10) also follows from (8.8) and (8.3), 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):

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

$$\Delta_3 = \left(\begin{array}{ccc} \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot \\ \cdot & 1 & \cdot \\ \hline -1 & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 \\ \hline \cdot & -1 & \cdot \\ \cdot & \cdot & -1 \\ \cdot & \cdot & \cdot \end{array}\right) \quad \Delta_4 = \left(\begin{array}{ccc} \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot \\ \cdot & 1 & \cdot \\ \hline \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 \\ \hline \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{array}\right) \quad \Delta_5 = \left(\begin{array}{ccc} 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \hline \cdot & \cdot & \cdot \\ \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot \\ \hline \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 \end{array}\right) \quad \Delta_6 = \left(\begin{array}{ccc} 1 & \cdot & \cdot \\ \cdot & 1 & \cdot \\ \cdot & \cdot & 1 \\ \hline \cdot & \cdot & 1 \\ 1 & \cdot & \cdot \\ \cdot & 1 & \cdot \\ \hline \cdot & 1 & \cdot \\ \cdot & \cdot & 1 \\ 1 & \cdot & \cdot \end{array}\right).$$

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

$$v_s(A) = (a_{21}, a_{31}, a_{32})' . \tag{8.11}$$

We also define the n×1 vectors

$$v_d(A) = (a_{11}, a_{22}, \ldots, a_{nn})' \tag{8.12}$$

and

$$v_1(A) = (a_{11}, a_{21}, \ldots, a_{n1})' . \tag{8.13}$$

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

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

$$\psi_1(A) = \psi_2(A) = v(A), \qquad \psi_3(A) = \psi_4(A) = v_s(A),$$

$$\psi_5(A) = v_d(A), \qquad \psi_6(A) = v_1(A) .$$

<u>Historical note</u>. Patterned matrices (with only equality relationships) among their elements were studied by Tracy and Singh (1972) with the purpose of finding matrix derivatives of certain matrix transformations. Lower triangular (and symmetric) matrices were discussed by Magnus & Neudecker (1980), and skew-symmetric, strictly lower triangular and diagonal matrices by Neudecker (1983). The present section is based on Magnus (1983) who introduced the concept of an L-structure in the context of solving linear matrix equations where the solution matrix is known to be L-structured.

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 $D_j f_i(x)$ denote the partial derivative of f_i with respect to the j^{th} coordinate. Then the m×n matrix

$$\begin{pmatrix} D_1 f_1(x) & D_2 f_1(x) & \ldots & D_n f_1(x) \\ D_1 f_2(x) & D_2 f_2(x) & \ldots & D_n f_2(x) \\ \vdots & \vdots & & \vdots \\ D_1 f_m(x) & D_2 f_m(x) & \ldots & D_n f_m(x) \end{pmatrix} \qquad (9.1)$$

is called the Jacobian matrix of f at x (the gradient vector, if $m = 1$) and is denoted $\nabla f(x)$ or $\partial f(x)/\partial x'$.

The generalization to matrix functions of matrices is straightforward. Let $F : S \rightarrow \mathbb{R}^{m\times p}$ be a matrix function defined and differentiable on a set S in $\mathbb{R}^{n\times q}$. Then we define the Jacobian matrix of F at X (the gradient vector, if $m = p = 1$) as the mp×nq matrix

$$\nabla F(X) = \frac{\partial \operatorname{vec} F(X)}{\partial (\operatorname{vec} X)'}, \qquad (9.2)$$

whose ij^{th} element is the partial derivative of the i^{th} component of vec F(X) with respect to the j^{th} coordinate of vec X.

We emphasize that (9.2) is the only sensible definition of a matrix derivative. There are, of course, other ways in which the mnpq partial derivatives of F could be displayed (Balestra (1976), Rogers (1980)), 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 (1984) and Magnus and Neudecker (1984).

The computation of Jacobian matrices is made extremely simple by the use of differentials (Neudecker (1969), Magnus and Neudecker (1984)). The essential property here is that

$$\text{vec}\, dF(X) = A(X)\, \text{vec}\, dX \tag{9.3}$$

if, and only if,

$$\nabla F(X) = A(X). \tag{9.4}$$

Thus, if we can find a matrix A (depending, in general, on X, but not on dX) satisfying (9.3), 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 Y = AXB where A and B are two matrices of constants. Taking differentials we have

$$dY = A(dX)B ,$$

from which we obtain, upon vectorizing,

$$\text{vec}\, dY = \text{vec}\, A(dX)B = (B' \otimes A)\, \text{vec}\, dX. \tag{9.5}$$

Hence the Jacobian matrix is

$$\frac{\partial\, \text{vec}\, Y}{\partial (\text{vec}\, X)'} = B' \otimes A \tag{9.6}$$

If X is constrained to be symmetric, we substitute D dv(X) for vec dX in (9.5), where D is the duplication matrix. This gives

$$\text{vec}\, dY = (B' \otimes A) D\, dv(X) ,$$

so that

$$\frac{\partial \operatorname{vec} Y}{\partial (v(X))'} = (B' \otimes A)D . \qquad (9.7)$$

Of course, we can also obtain (9.7) from (9.6) using the chain rule, since for symmetric X,

$$\frac{\partial \operatorname{vec} X}{\partial (v(X))'} = D . \qquad (9.8)$$

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

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

$$dY = dX^{-1} = -X^{-1}(dX)X^{-1}$$

and vecs,

$$\operatorname{vec} dY = -((X')^{-1} \otimes X^{-1}) \operatorname{vec} dX,$$

thus leading to the Jacobian matrix

$$\frac{\partial \operatorname{vec} Y}{\partial (\operatorname{vec} X)'} = -(X')^{-1} \otimes X^{-1} . \qquad (9.9)$$

Again, if X is symmetric (L-structured), we postmultiply (9.9) by D (Δ, in general).

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

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

so that the gradient vector is

$$\frac{\partial \phi(X)}{\partial (\text{vec } X)'} = (\text{vec } A')' \qquad (9.10)$$

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

$$\frac{\partial \phi(X)}{\partial (v(X))'} = (\text{vec } A')'D = \left(\text{vec } (A + A' - dg(A))\right)', \qquad (9.11)$$

where dg(A) is the diagonal matrix with the diagonal elements of 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 $D^2_{ij}\phi(x)$ denote the second-order partial derivative of ϕ with respect to the i^{th} and j^{th} coordinates. Then the $n \times n$ matrix $(D^2_{ij}\phi(x))$ is called the Hessian matrix of ϕ at x and is denoted $H\phi(x)$ or $\partial\phi(x)/\partial x \partial x'$. Since ϕ is twice differentiable at x, $H\phi(x)$ is a symmetric matrix.

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

$$H\phi(X) = \frac{\partial \phi(X)}{\partial \operatorname{vec} X \; \partial(\operatorname{vec} X)'} , \qquad (10.1)$$

whose ij^{th} element is the second-order partial derivative of ϕ with respect to the i^{th} and j^{th} coordinates of vec X.

The computation of Hessian matrices is based on the property that

$$d^2\phi(X) = (\operatorname{vec} dX)' B(X) (\operatorname{vec} dX) \qquad (10.2)$$

if, and only if,

$$H\phi(X) = \tfrac{1}{2}(B(X) + B'(X)) , \qquad (10.3)$$

where B may depend on X, but not on dX.

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

$$d^2\phi(X) = 2 \operatorname{tr} A(dX)B(dX)' = 2(\operatorname{vec} dX)'(B' \otimes A)(\operatorname{vec} dX) . \qquad (10.4)$$

The Hessian matrix is therefore

$$\frac{\partial^2\phi(X)}{\partial \operatorname{vec} X\ \partial(\operatorname{vec} X)'} = B' \otimes A + B \otimes A' . \qquad (10.5)$$

If X is constrained to be symmetric, we have

$$\frac{\partial^2\phi(X)}{\partial v(X)\ \partial(v(X))'} = D'(B' \otimes A + B \otimes A')D . \qquad (10.6)$$

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

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

and therefore

$$\begin{aligned} d^2\phi(X) &= -\operatorname{tr}(dX^{-1})(dX)X^{-1} - \operatorname{tr} X^{-1}(dX)(dX^{-1}) \\ &= 2\operatorname{tr} X^{-1}(dX)X^{-1}(dX)X^{-1} = 2(\operatorname{vec} dX')'(X'^{-2} \otimes X^{-1})(\operatorname{vec} dX) \\ &= 2(\operatorname{vec} dX)'K(X'^{-2} \otimes X^{-1})(\operatorname{vec} dX) , \qquad (10.7) \end{aligned}$$

so that the Hessian matrix becomes

$$\frac{\partial^2\phi(X)}{\partial \operatorname{vec} X\ \partial(\operatorname{vec} X)'} = K(X'^{-2} \otimes X^{-1} + X'^{-1} \otimes X^{-2}) . \qquad (10.8)$$

For symmetrix X, we find

$$\begin{aligned} \frac{\partial^2\phi(X)}{\partial v(X)\ \partial(v(X))'} &= D'(X^{-2} \otimes X^{-1} + X^{-1} \otimes X^{-2})D \\ &= 2D'(X^{-1} \otimes X^{-2})D , \qquad (10.9) \end{aligned}$$

using (7.4) and Lemma 4.1.

11. Jacobians involving L-structures

Let $F : S \to \mathbb{R}^{m\times p}$ be a matrix function defined and differentiable on a set S in $\mathbb{R}^{n\times q}$. If $mp = nq$, the Jacobian matrix $\nabla F(X)$ defined by (9.2) is a square matrix. Its determinant is called the Jacobian (or Jacobian determinant) and is denoted by $J_F(X)$. Thus,

$$J_F(X) = |\nabla F(X)| . \tag{11.1}$$

Example (i). The linear transformation $F(X) = AXB$, where X and F(X) are m×n matrices, and A and B are nonsingular matrices of constants of orders m×m and n×n, respectively. From (9.6) we know the Jacobian matrix $\nabla F(X) = B' \otimes A$, so that the Jacobian is

$$J_F(X) = |B' \otimes A| = |A|^n |B|^m . \tag{11.2}$$

Example (ii). The nonlinear transformation $F(X) = X^{-1}$, where X is a nonsingular n×n matrix. The Jacobian matrix is given in (9.9) as $\nabla F(X) = -(X')^{-1} \otimes X^{-1}$, so that the Jacobian of the transformation is

$$J_F(X) = |-(X')^{-1} \otimes X^{-1}| = (-1)^n |X|^{-2n}. \tag{11.3}$$

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

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

$$\text{vec}\, dF(X) = (A \otimes A)\, \text{vec}\, dX.$$

Since dX and dF(X) are symmetric, we obtain

$$dv(F(X)) = D_n^+(A \otimes A)D_n\, dv(X),$$

so that,

$$J_F(X) = \left|\frac{\partial v(F(X))}{\partial (v(X))'}\right| = |D_n^+(A \otimes A)D_n| = |A|^{n+1}, \qquad (11.4)$$

using (7.11).

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

$$\text{vec } dF(X) = -((X')^{-1} \otimes X^{-1}) \text{ vec } dX,$$

so that

$$dv(F(X)) = -D_n^+((X')^{-1} \otimes X^{-1})D_n\, dv(X).$$

The Jacobian of this transformation then follows from (7.11):

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

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

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

$$dF(X) = (dX)'X + X'dX,$$

we obtain

$$\begin{aligned}\mathrm{vec}\, dF(X) &= (X' \otimes I)\,\mathrm{vec}\,(dX)' + (I \otimes X')\,\mathrm{vec}\, dX \\ &= ((X' \otimes I)K_{nn} + I \otimes X')\,\mathrm{vec}\, dX \\ &= (I + K_{nn})(I \otimes X')\,\mathrm{vec}\, dX = 2N_n(I \otimes X')\,\mathrm{vec}\, dX.\end{aligned}$$

Now let L_n' be the Δ-matrix with the property that

$$L_n' v(A) = \mathrm{vec}\, A$$

for every lower triangular $n\times n$ matrix A. Then, since dX is lower triangular and dF(X) is symmetric, we obtain

$$\begin{aligned}dv(F(X)) &= 2\, D_n^+ N_n (I \otimes X') L_n' dv(X) \\ &= 2(D_n' D_n)^{-1} D_n' (I \otimes X') L_n' dv(X),\end{aligned}$$

using (7.5) and (7.2). The Jacobian matrix is therefore

$$\frac{\partial v(F(X))}{\partial (v(X))'} = 2(D_n' D_n)^{-1} (L_n (I \otimes X) D_n)' ,$$

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

$$J_F(X) = \left|\frac{\partial v(F(X))}{\partial (v(X))'}\right| = 2^n \prod_{i=1}^{n} x_{ii}^{i} , \qquad (11.5)$$

using (7.13) and Lemma 4.1(iii) of Magnus and Neudecker (1980).

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 (1951) and Olkin (1953)), induction (Jack (1966)), and functional equations induced on the relevant spaces (Olkin and Sampson (1972)). Our approach finds its root in Tracy and Singh (1972) 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 (1980); the matrix L_n introduced in example (v) is their so-called "elimination" matrix. Neudecker (1983) obtained Jacobians of transformations with skew-symmetric, strictly lower triangular, or diagonal matrix arguments.

Aitken, A.C. (1949). On the Wishart distribution in statistics. Biometrika, 36, 59-62.

Balestra, P. (1976). La Dérivation Matricielle. Collection de l'Institut de Mathématiques Economiques, No. 12, Sirey, Paris.

Barnett, S. (1973). Matrix differential equations and Kronecker products. SIAM Journal of Applied Mathematics, 24, 1-5.

Bellman, R. (1970). Introduction to matrix analysis. McGraw-Hill, New York, 2nd edition.

Browne, M.W. (1974). Generalized least squares estimators in the analysis of covariance structures. South African Statistical Journal, 8, 1-24.

Deemer, W.L. and Olkin, I. (1951). The Jacobians of certain matrix transformations useful in multivariate analysis. Biometrika, 38, 345-367.

Don, F.J.H., and van der Plas, A.P. (1981). A note on the characteristic polynomial of the commutation matrix. Linear Algebra and Its Applications, 37, 135-142.

Hartwig, R.E., and Morris, S.B. (1975). The universal flip matrix and the generalized faro-shuffle. Pacific Journal of Mathematics, 58, 445-455.

Henderson, H.V., and Searle, S.R. (1979). Vec and vech operators for matrices, with some uses in Jacobians and multivariate statistics. The Canadian Journal of Statistics, 7, 65-81.

Henderson, H.V., and Searle, S.R. (1981). The vec-permutation matrix, the vec operator and Kronecker products: a review. Linear and Multilinear Algebra, 9, 271-288.

Jack, H. (1966). Jacobians of transformations involving orthogonal matrices. Proceedings of the Royal Society of Edinburgh, Section A, 67, 81-103.

Koopmans, T.C., Rubin, H., and Leipnik, R.B. (1950). Measuring the equation systems of dynamic economics, in: Statistical Inference in Dynamic Economic Models (T.C. Koopmans, ed.). John Wiley, New York.

Ledermann, W. (1936). On singular pencils of Zehfuss, compound, and Schläflian matrices. Proceedings of the Royal Society of Edinburgh, 56, 50-89.

Magnus, J.R. (1983). L-structured matrices and linear matrix equations. Linear and Multilinear Algebra, 14, 67-88.

Magnus, J.R., and Neudecker, H. (1979). The Commutation matrix: some properties and applications. The Annals of Statistics, 7, 381-394.

Magnus, J.R., and Neudecker, H. (1980). The elimination matrix: some lemmas and applications. SIAM Journal on Algebraic and Discrete Methods, 1, 422-449.

Magnus, J.R., and Neudecker, H. (1984). Matrix differential calculus with applications to simple, Hadamard, and Kronecker products, submitted for publication.

Murnaghan, F.D. (1938). The theory of group representations. Johns Hopkins Press, Baltimore.

Nel, D.G. (1978). On the symmetric multivariate normal distribution and the asymptotic expansion of a Wishart matrix. South African Statistical Journal, 12, 145-159.

Neudecker, H. (1968). The Kronecker matrix product and some of its applications in econometrics. Statistica Neerlandica, 22, 69-81.

Neudecker, H. (1969). Some theorems on matrix differentiation with special reference to Kronecker matrix products. Journal of the American Statistical Association, 64, 953-963.

Neudecker, H. (1983). On Jacobians of transformations with skew-symmetric, strictly (lower) triangular or diagonal matrix arguments. Linear and Multilinear Algebra, 14, 271-295.

Neudecker, H., and Wansbeek, T. (1983). Some results on commutation matrices, with statistical applications. The Canadian Journal of Statistics, 11, 221-231.

Olkin, I. (1953). Note on 'The Jacobians of certain matrix transformations useful in multivariate analysis'. Biometrika, 40, 43-46.

Olkin, I., and Sampson, A.R. (1972). Jacobians of matrix transformations and induced functional equations. Linear Algebra and Its Applications, 5, 257-276.

Pollock, D.S.G. (1984). Tensor products and matrix differential calculus (revised version). Discussion paper No. 86, Department of Economics, Queen Mary College, London.

Richard, J.F. (1975). A note on the information matrix of the multivariate normal distribution. Journal of Econometrics, 3, 57-60.

Rogers, G.S. (1980). Matrix derivatives. Marcel Dekker, New York.

Roth, W.E. (1934). On direct product matrices. Bulletin of the American Mathematical Society, 40, 461-468.

Sylvester, J. (1884a). 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, Paris, Tome 99, 117-118.

Sylvester, J. (1884b). Sur la résolution générale de l'équation linéaire en matrices d'un ordre quel-conque, Comptes Rendus des Séances de l'Académie des Sciences, Paris, Tome 99, 409-412 and 432-436.

Tracy, D.S., and Dwyer, P.S. (1969). Multivariate maxima and minima with matrix derivatives. Journal of the American Statistical Association, 64, 1576-1594.

Tracy, D.S., and Singh, R.P. (1972). Some modifications of matrix differentiation for evaluating Jacobians of symmetric matrix transformations, in: Symmetric Functions in Statistics (D.S. Tracy, ed.). University of Windsor, Ontario, Canada.

Vartak, N.N. (1955). On an application of Kronecker product of matrices to statistical designs. Annals of Mathematical Statistics, 26, 420-438.

Vetter, W.J. (1975). Vector structures and solutions of linear matrix equations. Linear Algebra and Its Applications, 10, 181-188.

Zehfuss, G. (1858). Ueber eine gewisse Determinante. Zeitschrift für Mathematik und Physik, 3, 298-301.