On Some Pattern-Reduction Matrices Which Appear in Statistics

Douglas P. Wiens*
Department of Mathematics, Statistics, and Computing Science
Dalhousie University
Halifax, Nova Scotia, Canada B3H 4H8

Submitted by George P. H. Styan

ABSTRACT

We review and extend some recent results concerning the structure of pattern-reduction matrices, which effect the reduction of the vec of a patterned matrix to the vector consisting only of the functionally independent elements of the matrix. The results are applied to the calculation of certain Jacobians, and to the construction of ellipsoidal confidence regions for covariance matrices, on the basis of maximum likelihood or robust M-estimators.

1. INTRODUCTION

In this paper, we review and extend some recent results concerning the structure of *pattern-reduction* matrices, which effect the reduction of the vec of a patterned matrix to the vector consisting only of the functionally independent elements of the matrix. Such matrices occur in the calculation of Jacobians, when the domain or range of the transformation is a patterned matrix, and in the elimination of singularities in the covariance matrices of random patterned matrices. Examples of such applications are given in Section 5. In Sections 2 and 3, we investigate the calculus of matrix differentiation, obtaining structural properties of pattern-reduction matrices from those of the (generally well-studied) Jacobian matrices in which they appear. These properties are then exploited in Section 4, where we exhibit some general results on the determinants and inverses of combinations of pattern-reduction matrices with arbitrary Kronecker products.

*Research supported by the Natural Sciences and Engineering Research Council under Grant No. A8603.

52 Vanderbilt Ave., New York, NY 10017 0024-3795/85/$3.30

By vec X, we understand the column vector formed from the columns of X, in their natural order. Of central importance is the vec permutation matrix $I_{(p,q)}$, defined by its action

$$\operatorname{vec} A_{p\times q} = I_{(p,q)} \operatorname{vec} A'. \tag{1.1}$$

This is as in Henderson and Searle [4], who also [5] give a number of properties of vec X and $I_{(p,q)}$, together with an interesting history of their derivations. Further properties, with applications to the determination of moments in distributions related to the multivariate normal, may be found in Magnus and Neudecker [7]. We mention in particular the properties

$$\operatorname{vec} ABC = (C'\otimes A)\operatorname{vec} B \qquad \text{(Roth [17])}, \tag{1.2}$$

$$I_{(p,m)}(A_{p\times q}\otimes B_{m\times n})I_{(n,q)} = B\otimes A \qquad \text{(MacRae [6])}. \tag{1.3}$$

See Graham [3] for properties of the Kronecker product $\otimes$ and its role in matrix calculus.

2. THE CALCULUS OF MATRIX DIFFERENTIATION

In this section, we give some basic definitions and rules for calculating the derivatives of matrix-valued functions $Y(X)$, with matrix-valued arguments X. It is assumed that the elements of X are functionally independent and variable.

Our definition of a matrix derivative combines the common "vector rearrangement" definition with the "derivative operator" approach of MacRae [6], McCulloch [9], and Rogers [16]:

$$\frac{\partial Y}{\partial X} := (\operatorname{vec} Y)\left(\operatorname{vec}\frac{\partial}{\partial X}\right)', \tag{2.1}$$

where $\partial/\partial X = (\partial/\partial x_{ij})$ is a matrix of derivative operators, with multiplication corresponding to partial differentiation. See Nel [12] for a comparison of several methods of defining matrix derivatives. The use of $\partial/\partial X$ allows (2.1) to extend naturally to higher derivatives:

$$\frac{\partial^k Y}{\partial X^k} := \frac{\partial}{\partial X}\left(\frac{\partial^{k-1} Y}{\partial X^{k-1}}\right) = \left(\operatorname{vec}\frac{\partial^{k-1} Y}{\partial X^{k-1}}\right)\left(\operatorname{vec}\frac{\partial}{\partial X}\right)' \tag{2.2}$$

It follows that for $k \geqslant 2$,

$$\frac{\partial^k Y}{\partial X^k} = \left(\text{vec}\,\frac{\partial}{\partial X}\right) \otimes \frac{\partial^{k-1} Y}{\partial X^{k-1}}. \tag{2.3}$$

EXAMPLE 2.1. Taylor's theorem can be conveniently formulated in the above notation. Let $\mathbf{F}: D \subseteq \mathbb{R}^q \to \mathbb{R}^p$ be $n+1$ times differentiable on an open convex set $D_0 \subseteq D$. Let $\mathbf{x}, \mathbf{y} \in D_0$, set $\mathbf{z} = \mathbf{y} - \mathbf{x}$, let $\mathbf{z}^{[k]}$ be the k-fold iterated Kronecker product of $\mathbf{z}$ with itself, and let $\times$ represent the Schur product $\mathbf{t} \times \mathbf{z} = (t_1 z_1, \dots, t_q z_q)'$. Writing out the standard proof of Taylor's theorem in matrix terms gives the following expansion, for some $\mathbf{t}$, with $0 < t_i < 1$, $1 \leqslant i \leqslant q$:

$$\mathbf{F}(\mathbf{y}) - \mathbf{F}(\mathbf{x}) = \sum_{k=1}^{n} \frac{1}{k!}\left(\mathbf{z}^{[k-1]\prime} \otimes I_p\right) \frac{\partial^k \mathbf{F}}{\partial \mathbf{u}^k} \mathbf{z}\bigg|_{\mathbf{u}=\mathbf{x}}$$

$$+ \frac{1}{(n+1)!}\left(\mathbf{z}^{[n]\prime} \otimes I_p\right) \frac{\partial^{n+1} \mathbf{F}}{\partial \mathbf{u}^{n+1}} \mathbf{z}\bigg|_{\mathbf{u}=\mathbf{x}+\mathbf{t}\times\mathbf{z}}.$$

Let $Y = Y(X)$, and denote by dY and dX the matrices of differentials. Then (Deemer and Olkin [2])

(R1) $\text{vec}\, dY = A\, \text{vec}\, dX$ iff $\partial Y / \partial X = A$.

A chain rule is immediate from (2.1):

(R2) If $Z = Z(Y_1, \dots, Y_m)$ and $Y_i = Y_i(X)$, then

$$\frac{\partial Z}{\partial X} = \sum_{i=1}^{m} \frac{\partial Z}{\partial Y_i} \frac{\partial Y_i}{\partial X}.$$

The next rule follows from (1.1), but may also be taken, as in McDonald and Swaminathan [10], as defining $I_{(p,q)}$:

(R3) $\partial (Y_{p\times q})' / \partial Y = I_{(q,p)}$, so that

$$\frac{\partial Y'}{\partial X} = I_{(q,p)} \frac{\partial Y}{\partial X}.$$

An application of (1.2) gives

(R4) $\partial Y_1 Y_2 \cdots Y_m / \partial Y_i = Y_m' \ldots Y_{i+1}' \otimes Y_1 \ldots Y_{i-1}$ $(Y_0 = I,\ \ Y_{m+1} = I)$, so that

$$\frac{\partial Y_1 \ldots Y_m}{\partial X} = \sum_{i=1}^{m} \left(Y_m' \ldots Y_{i+1}' \otimes Y_1 \ldots Y_{i-1} \right) \left(\frac{\partial Y_i}{\partial X} \right).$$

In particular,

$$\frac{\partial \underset{p \times q}{Y_1} \underset{q \times r}{Y_2}}{\partial X} = \left(Y_2' \otimes I_p \right) \frac{\partial Y_1}{\partial X} + \left(I_r \otimes Y_1 \right) \frac{\partial Y_2}{\partial X}.$$

The calculation of second and higher derivatives requires the differentiation of Kronecker products. For this, let $Y_i : p_i \times q_i$ be functions of X $(1 \leqslant i \leqslant m)$. Put

$$p = \prod_1^m p_j, \qquad q = \prod_1^m q_j;$$

$$\alpha_i = \prod_1^{i-1} p_j, \quad \alpha_1 = 1; \qquad \beta_i = \sum_{i+1}^{m} p_j, \quad \beta_m = 1;$$

$$a_i = \prod_1^{i-1} q_j, \quad a_1 = 1; \qquad b_i = \prod_{i+1}^{m} q_j, \quad b_m = 1.$$

Let $A_i = Y_1 \otimes \cdots \otimes Y_{i-1} : \alpha_i \times a_i$; $A_1 = 1$; $B_i = Y_{i+1} \otimes \cdots \otimes Y_m : \beta_i \times b_i$; $B_m = 1$. Define $Q_i : pq \times pq$ by

$$Q_i = I_{a_i} \otimes \left[I_{(p_i \alpha_i,\, q_i b_i)} \left(I_{\alpha_i} \otimes I_{(q_i,\, p_i)} \otimes I_{b_i} \right) \right] \otimes I_{\beta_i}.$$

Then

$$\begin{aligned} \operatorname{vec}(A_i \otimes Y_i \otimes B_i) &= Q_i (\operatorname{vec} A_i \otimes \operatorname{vec} Y_i \otimes \operatorname{vec} B_i) \\ &= Q_i \left(\operatorname{vec} A_i \otimes I_{p_i q_i} \otimes \operatorname{vec} B_i \right) \operatorname{vec} Y_i . \end{aligned} \qquad (2.4)$$

The first equality above follows easily, once it is checked for matrices

A_i, Y_i, B_i of rank one; this in turn is straightforward. We then get

(R5) With notation as above,

$$\frac{\partial Y_1\otimes\cdots\otimes Y_m}{\partial Y_i}=Q_i\left(\operatorname{vec}A_i\otimes I_{p_iq_i}\otimes\operatorname{vec}B_i\right),$$

so that

$$\frac{\partial Y_1\otimes\cdots\otimes Y_m}{\partial X}=\sum_{i=1}^{m}Q_i\left(\operatorname{vec}A_i\otimes\frac{\partial Y_i}{\partial X}\otimes\operatorname{vec}B_i\right).$$

A particular case of (2.4) is

$$\operatorname{vec}(Y_1\otimes Y_2)=\left(I_{q_1}\otimes I_{(p_1,q_2)}\otimes I_{p_2}\right)(\operatorname{vec}Y_1\otimes\operatorname{vec}Y_2)$$

(Neudecker and Wansbeek [14]), implying (Bentler and Lee [13])

$$\frac{\partial Y_1\otimes Y_2}{\partial X}=\left(I_{q_1}\otimes I_{(p_1,q_2)}\otimes I_{p_2}\right)\left(\operatorname{vec}Y_1\otimes\frac{\partial Y_2}{\partial X}+\frac{\partial Y_1}{\partial X}\otimes\operatorname{vec}Y_2\right).$$

EXAMPLE 2.2. To illustrate the use of these rules, and some useful techniques, we calculate $\partial^2X^{-1}/\partial X^2$, where X is $p\times p$. Taking differentials in $XX^{-1}=I$ and applying (R1) gives

$$\frac{\partial X^{-1}}{\partial X}=-(X^{-T}\otimes X^{-1}),\qquad\text{where}\quad X^{-T}=(X^{-1})'. \tag{2.5}$$

Put $Q=I_p\otimes I_{(p,p)}\otimes I_p$ and apply (R5), then (1.1) and (R3), then (1.3) to get

$$\frac{\partial^2X^{-1}}{\partial X^2}=-Q\left[I_{(p,p)}\otimes I_{p^2}\right]\left[I_{p^4}+I_{(p^2,p^2)}\right]\left(\frac{\partial X^{-1}}{\partial X}\otimes\operatorname{vec}X^{-1}\right).$$

A basis of p^4-dimensional Euclidean space is provided by vectors of the form $\mathbf{a}\otimes\mathbf{b}\otimes\mathbf{c}\otimes\mathbf{d}$. Since the action of $Q[I_{(p,p)}\otimes I_{p^2}]$ on such vectors is that of $I_{(p,p^2)}\otimes I_p$, these matrices are equal. This gives

$$\frac{\partial^2X^{-1}}{\partial X^2}=\left[I_{(p,p^2)}\otimes I_p\right]\left[I_{p^4}+I_{(p^2,p^2)}\right]\left[X^{-T}\otimes X^{-1}\otimes\operatorname{vec}X^{-1}\right].$$

Inserting this into the expression in Example 2.1 and continuing gives the familiar expansion of the inverse of a matrix Y in terms of a "nearby" matrix X:

$$Y^{-1}=X^{-1}\left[\sum_{i=0}^{n}(-ZX^{-1})^{i}+(-UX^{-1})^{n+1}\right],$$

where $Z=Y-X$, $U=X+(T\times Z)$, $0<t_{ij}<1$.

3. PATTERNED MATRICES AND THEIR DERIVATIVES

DEFINITION 3.1. A variable matrix $X: r\times s$ has (linear) pattern p ($X\in\mathcal{M}_p^{(r,s)}$) if

(i) X has p^* functionally independent elements, arranged in some order as

$$\operatorname*{vec}_p X: p^*\times 1;$$

(ii) The remaining $rs-p^*$ elements are linear combinations of the elements of $\operatorname*{vec}_p X$.

We shall write $\mathcal{M}_p^{(r,s)}$ as $\mathcal{M}_p$. An equivalent definition is that $X\in\mathcal{M}_p$ iff there exists a permutation matrix $Q_p: rs\times rs$ and a constant matrix $A_p: rs-p^*\times p^*$ such that

$$\operatorname{vec} X=Q_p\begin{pmatrix} I_{p^*}\\ A_p\end{pmatrix}\operatorname*{vec}_p X, \tag{3.1}$$

where the elements of $\operatorname*{vec}_p X$ are functionally independent. Note that this definition may be broadened to include matrices X with nonzero constant elements by applying it to $X-C$, where C contains only those nonzero constants in X. Nel [12] seems to overlook this distinction—his relation 6.1.2, corresponding to (3.2ii) below, can hold only if the constant elements of X are all zeros.

DEFINITION 3.2.

(i) $G_p = Q_p \begin{pmatrix} I_{p^*} \\ A_p \end{pmatrix} : rs \times p^*$.
(ii) $G_p^+ = (G_p' G_p)^{-1} G_p' : p^* \times rs$.
(iii) $M_p = G_p G_p^+ : rs \times rs$.
(iv) $\mathscr{H}_p = \{ H_p : p^* \times rs \mid H_p G_p = I_{p^*} \} = \{ H_p \mid H_p = G_p^+ + C(I_{rs} - M_p)$ for some $C: p^* \times rs \}$.

From this point onwards, any statement involving H_p is to be understood as holding for any $H_p \in \mathscr{H}_p$, in particular for G_p^+. In the following immediate consequences of Definition 3.2, X is any member of $\mathscr{M}_p$:

$$H_p \operatorname{vec} X = \underset{p}{\operatorname{vec}} X, \tag{3.2i}$$

$$G_p \underset{p}{\operatorname{vec}} X = \operatorname{vec} X, \tag{3.2ii}$$

$$M_p \operatorname{vec} X = \operatorname{vec} X, \tag{3.2iii}$$

$$G_p H_p \operatorname{vec} X = \operatorname{vec} X, \tag{3.2iv}$$

$$M_p G_p = G_p, \tag{3.2v}$$

$$H_p M_p = G_p^+ . \tag{3.2vi}$$

Following Henderson and Searle [4], Tracy and Singh [19], and Nel [12], we define derivatives of patterned matrices by restricting the vector rearrangement method to the functionally independent elements.

DEFINITION 3.3. If $Y \in \mathscr{M}_p$, $X \in \mathscr{M}_q$, and $Y = Y(X)$, then

$$\frac{\partial_p Y}{\partial_q X} = \frac{\partial \underset{p}{\operatorname{vec}} Y}{\partial \underset{q}{\operatorname{vec}} X} : p^* \times q^*, \qquad \frac{\partial_p^k Y}{\partial_q X^k} = \frac{\partial}{\partial \underset{q}{\operatorname{vec}} X} \operatorname{vec}\left(\frac{\partial_p^{k-1} Y}{\partial_q X^{k-1}} \right) : p^* q^{*k-1} \times q^*.$$

Using (R1) and (3.2i, ii), we have

$$\underset{p}{\operatorname{vec}} dY = H_p \left(\frac{\partial Y}{\partial X} \right) G_q \underset{q}{\operatorname{vec}} dX,$$

so that

$$\frac{\partial_p Y}{\partial_q X} = H_p\left(\frac{\partial Y}{\partial X}\right)G_q, \tag{3.3}$$

with $\partial Y/\partial X$ calculated ignoring functional relationships. Then by induction,

$$\frac{\partial_p^k Y}{\partial_q X^k} = \left(G_q'^{[k-1]}\otimes H_p\right)\left(\frac{\partial^k Y}{\partial X^k}\right)G_q, \tag{3.4}$$

where the exponent in square brackets denotes an iterated Kronecker product.

LEMMA 3.1. *For all* $Y = Y(X) \in \mathcal{M}_p$, $X \in \mathcal{M}_q$, $H_p, H_{p,1} \in \mathcal{H}_p$, $H_q \in \mathcal{H}_q$:

$$H_p\left(\frac{\partial Y}{\partial X}\right)G_q = H_{p,1}\left(\frac{\partial Y}{\partial X}\right)G_q, \tag{3.5}$$

$$G_p H_p\left(\frac{\partial Y}{\partial X}\right)G_q = \left(\frac{\partial Y}{\partial X}\right)G_q, \tag{3.6}$$

$$M_p\left(\frac{\partial Y}{\partial X}\right)G_q = \left(\frac{\partial Y}{\partial X}\right)G_q, \tag{3.7}$$

and if $\partial Y/\partial X$ *is invertible, then*

$$\left[H_p\left(\frac{\partial Y}{\partial X}\right)G_q\right]^{-1} = H_q\left(\frac{\partial Y}{\partial X}\right)^{-1}G_p. \tag{3.8}$$

Proof. Equation (3.5) is merely the observation that (3.3) does not depend upon the choice of H_p. Equation (3.8) follows from (3.6), and (3.7) follows from (3.5) upon expressing $H_p - H_{p,1}$ as $C(I_{rs} - M_p)$ for arbitrary $C\colon p^* \times rs$. Now use $H_{p,1} = G_p^+$ in (3.5), premultiply by G_p, and apply (3.7) to obtain (3.6). ■

From (3.6), we get the analogue of (R2): If $Z \in \mathcal{M}_p$ is a function of matrices $Y_i \in \mathcal{M}_{q_i}$, all of which are functions of $X \in \mathcal{M}_r$, then

$$\frac{\partial_p Z}{\partial_r X} = \sum_i \left(\frac{\partial_p Z}{\partial_{q_i} Y_i}\right)\left(\frac{\partial_{q_i} Y_i}{\partial_r X}\right).$$

We now consider some particular patterns. In the following examples, all matrices X, Y, A, B are $n \times n$, and $q = n(n+1)/2$.

DEFINITION 3.4. For any square X, X_L is the lower triangle of X, X_U the upper triangle, and X_D the diagonal, all augmented by zeros. Also, $X_{LD} = X_L - X_D$, $X_{UD} = X_U - X_D$. Thus,

$$X = X_{LD} + X_D + X_{UD} = X_L + X_{UD} = X_U + X_{LD}. \tag{3.9}$$

The matrices M_p, $p \in \{l, u, d, ld, ud\}$, defined below effect the decomposition of $\operatorname{vec} X$ into orthogonal components $\operatorname{vec} X_L$, $\operatorname{vec} X_{UD}$, etc. This is useful in isolating those elements of a matrix which are functionally, but not linearly, dependent upon the others—see Examples 5.3 and 5.4 below. For $p \in \{l, ld, d\}$, the elements of $\underset{p}{\operatorname{vec}} X$ are ordered antilexicographically, for $p \in \{u, ud\}$ they are ordered lexicographically.

EXAMPLE 3.1. Let $\mathscr{M}_l$ be the class of lower triangular matrices. Then

$$G_l = \bigoplus_{i=1}^{n} \begin{pmatrix} 0 \\ \hline I \end{pmatrix}_{n-i+1}^{i-1} : n^2 \times q, \qquad G_l^+ = G_l', \qquad M_l = \bigoplus_{i=1}^{n} \begin{pmatrix} 0_{i-1} & 0 \\ 0 & I_{n-i+1} \end{pmatrix}. \tag{3.10}$$

For any X, triangular or not,

$$G_l^+ \operatorname{vec} X = \underset{l}{\operatorname{vec}} X_L = G_l^+ \operatorname{vec} X_L, \qquad M_l \operatorname{vec} X = \operatorname{vec} X_L. \tag{3.11}$$

Consider the transformation $Y = TX$, with $X, Y, T \in \mathscr{M}_l$, for which $(\partial_l Y / \partial_l X) = H_l(I_n \otimes T) G_l$. Then (3.6) yields

$$G_l H_l (I_n \otimes T) G_l = (I_n \otimes T) G_l. \tag{3.12}$$

EXAMPLE 3.2. Let $\mathscr{M}_u$ be the class of upper triangular matrices. If $X \in \mathscr{M}_u$, then

$$\underset{u}{\operatorname{vec}} X = \underset{l}{\operatorname{vec}} X',$$

so that

$$I_{(n,n)} G_l \underset{u}{\operatorname{vec}} X = I_{(n,n)} \operatorname{vec} X' = \operatorname{vec} X = G_u \underset{u}{\operatorname{vec}} X. \tag{3.13}$$

Since $\underset{u}{\operatorname{vec}} X$ varies freely,

$$G_u = I_{(n,n)} G_l, \qquad G_u^+ = G_u' = G_l^+ I_{(n,n)}, \qquad M_u = I_{(n,n)} M_l I_{(n,n)}. \tag{3.14}$$

For arbitrary X, A, B,

$$G_u^+ \operatorname{vec} X = \operatorname{vec}_u X_U = G_u^+ \operatorname{vec} X_U, \qquad M_u \operatorname{vec} X = \operatorname{vec} X_U,$$

$$G_l^+ (A \otimes B) G_u = G_u^+ (B \otimes A) G_l. \tag{3.15}$$

Applying (3.6) to the transformation $Y = XS'$, $X, Y \in \mathcal{M}_l$, $S \in \mathcal{M}_u$ gives

$$G_l H_l (S \otimes I_n) G_l = (S \otimes I_n) G_l. \tag{3.16}$$

EXAMPLE 3.3. Let $\mathcal{M}_d$ be the class of diagonal matrices. Then $G_d = (E_1' \ \cdots \ E_n')' : n^2 \times n$, where

$$E_i = \operatorname{diag}(0 \ \cdots \ 0 \ \underset{\substack{\uparrow \\ i}}{1} \ 0 \ \cdots \ 0);$$

and $M_d = \oplus_{i=1}^n E_i$. For arbitrary X,

$$G_d^+ \operatorname{vec} X = \operatorname{vec}_d X_D = G_d^+ \operatorname{vec} X_D, \qquad M_d \operatorname{vec} X = \operatorname{vec} X_D, \tag{3.17}$$

implying

$$G_d^+ I_{(n,n)} = G_d^+, \qquad M_d I_{(n,n)} = M_d. \tag{3.18}$$

Then for arbitrary A and B,

$$G_d^+ (A \otimes B) G_d = G_d^+ (B \otimes A) G_d. \tag{3.19}$$

EXAMPLE 3.4. Let $\mathcal{M}_{ld}$ be the class of lower triangular matrices with null diagonals, $\mathcal{M}_{ud}$ the class of upper triangular matrices with null diagonals. Then the relations (3.10)–(3.15) hold, with the notational changes $i \to i+1$, $l \to ld$, $u \to ud$, $L \to LD$, $U \to UD$.

LEMMA 3.2. *In each of the following identities, the matrices on the left are mutually orthogonal:*

(i) $M_{ud} + M_d = M_u$,
(ii) $M_{ld} + M_d = M_l$,
(iii) $M_u + M_{ld} = I$,
(iv) $M_l + M_{ud} = I$,
(v) $M_{ud} + M_d + M_{ld} = I$.

Proof. To establish the identities, consider the actions of the matrices on $\operatorname{vec} X$ for arbitrary X. Then since all are idempotent, they are mutually orthogonal. ■

EXAMPLE 3.5. Let $\mathscr{M}_s$ be the class of symmetric matrices, with $\operatorname{vec}_s X$ ordered antilexicographically. Recall (3.1). We may take $A_s = \oplus_{i=1}^{n}(0 \,\vdots\, I_{n-i})$, yielding

$$G_s'G_s = \operatorname{diag}(1,2,\dots,2;1,2,\dots,2;\dots;1,2;1) := D_s : q \times q. \quad (3.20)$$

As in Henderson and Searle [4], we have the identities

$$I_{(n,n)}G_s = G_s, \qquad G_s^+ I_{(n,n)} = G_s^+,$$

$$M_s = M_s I_{(n,n)} = I_{(n,n)} M_s = \tfrac{1}{2}(I + I_{(n,n)}), \quad (3.21)$$

implying that for arbitrary A and B,

$$G_s^+(A \otimes B)G_s = G_s^+(B \otimes A)G_s, \qquad M_s(A \otimes A) = (A \otimes A)M_s. \quad (3.22)$$

By (3.21) and (3.2vi),

$$H_s(I + I_{(n,n)}) = 2G_s^+. \quad (3.23)$$

Consider the transformation $Y = AXA'$ with $X, Y \in \mathscr{M}_s$, for which $\partial_s Y/\partial_s X = H_s(A \otimes A)G_s$. Then (3.5), (3.6) yield

$$H_s(A \otimes A)G_s = G_s^+(A \otimes A)G_s, \qquad G_s H_s(A \otimes A)G_s = (A \otimes A)G_s,$$

$$(H_s(A \otimes A)G_s)^+ = H_s(A^+ \otimes A^+)G_s. \quad (3.24)$$

By Theorem 3.7 of Deemer and Olkin [2],

$$|H_s(A \otimes A)G_s| = |A|^{n+1}. \quad (3.25)$$

The relations (3.24), (3.25) were also obtained by Henderson and Searle [4], without reference to matrix derivatives. If $X \in \mathscr{M}_s$, then

$$G_l^+ G_s \operatorname{vec}_s X = G_l^+ \operatorname{vec} X = \operatorname{vec}_l X_L = \operatorname{vec}_s X,$$

by (3.11). Hence, similarly,

$$G_l^+ G_s = G_u^+ G_s = G_s' G_l = G_s' G_u = I_q, \qquad \text{i.e.} \quad \{G_l, G_u\} \subset \mathscr{H}_s,$$

$$G_s' \in \mathscr{H}_l \cap \mathscr{H}_u; \quad (3.26)$$

$$G_s^+ = \tfrac{1}{2}(G_l^+ + G_u^+) = \tfrac{1}{2} G_l^+ (I + I_{(n,n)}) = G_l^+ M_s. \qquad (3.27)$$

For arbitrary X,

$$G_s^+ \operatorname{vec} X = \operatorname{vec}_s \left(\frac{X + X'}{2} \right), \qquad M_s \operatorname{vec} X = \operatorname{vec}\left(\frac{X + X'}{2} \right). \qquad (3.28)$$

EXAMPLE 3.6. Let $\mathscr{M}_{ss}$ be the class of skew-symmetric matrices, with $\operatorname{vec}_{ss} X$ ordered antilexicographically. We may take $A_{ss} = (-I_{n(n-1)/2} \vdots 0_{n(n-1)/2 \times n})'$, yielding

$$G_{ss}' G_{ss} = 2I_{n(n-1)/2}, \qquad G_{ss}^+ = \tfrac{1}{2} G_{ss}'. \qquad (3.29)$$

Proceeding as in the previous example, we find

$$I_{(n,n)} G_{ss} = -G_{ss}, \qquad G_{ss}^+ I_{(n,n)} = -G_{ss}^+, \qquad (3.30)$$

correcting Nel [12, p. 163, line 2]; and

$$M_{ss} = \tfrac{1}{2}(I - I_{(n,n)}) = -M_{ss} I_{(n,n)} = -I_{(n,n)} M_{ss}, \qquad G_{ss} G_{ss}' = I - I_{(n,n)};$$

$$(3.31)$$

$$G_{ss}^+ (A \otimes B) G_{ss} = G_{ss}^+ (B \otimes A) G_{ss}, \qquad M_{ss}(A \otimes A) = (A \otimes A) M_{ss};$$

$$H_{ss}(I - I_{(n,n)}) = 2G_{ss}^+, \qquad G_{ld}^+ G_{ss} = -G_{ud}^+ G_{ss} = I, \qquad (3.32)$$

$$G_{ss}^+ = \tfrac{1}{2}(G_{ld}^+ - G_{ud}^+) = G_{ld}^+ M_{ss}; \qquad (3.33)$$

$$G_{ss} \operatorname{vec} X = \operatorname{vec}_{ss} \left(\frac{X - X'}{2} \right), \quad M_{ss} \operatorname{vec} X = \operatorname{vec}\left(\frac{X - X'}{2} \right) \qquad \text{for all } X.$$

$$(3.34)$$

As in Lemma 3.2, (3.34), together with (3.28), implies that M_s and M_{ss} are mutually orthogonal. Thus in Nel [12, p. 163], the second product in line 2 and both products in line 6 should be zero, rather than as stated there. Considering the transformation $Y = AXA'$, $X, Y \in \mathscr{M}_{ss}$, and employing Theorem 3.10 of Deemer and Olkin [2] gives the analogues of (3.24) and (3.25):

$$H_{ss}(A\otimes A)G_{ss} = G_{ss}^{+}(A\otimes A)G_{ss}, \qquad G_{ss}H_{ss}(A\otimes A)G_{ss} = (A\otimes A)G_{ss},$$
$$|H_{ss}(A\otimes A)G_{ss}| = |A|^{n-1}. \tag{3.35}$$

LEMMA 3.3. *Let $S \in \mathscr{M}_u$, $T \in \mathscr{M}_l$ be arbitrary. In the following, the matrices in the second column are the inverses of those in the first, for any H_l, $H_{l,1} \in \mathscr{H}_l$. The expressions in the third column are the determinants of the matrices in the first column.*

	Inverse	Determinant	
$H_l(S\otimes I_n)G_l$	$H_{l,1}(S^{-1}\otimes I_n)G_l$	$\lvert S\rvert^{n+1}\prod_1^n s_{ii}^{-i}$ (= 0 if $\lvert S\rvert = 0$)	(3.36)
$G_l^{+}(I_n\otimes S)H_l'$	$G_l^{+}(I_n\otimes S^{-1})H_l'$	$\prod_1^n s_{ii}^{i}$	(3.37)
$G_l^{+}(S\otimes I_n)G_s$	$G_s^{+}(S^{-1}\otimes S^{-1})G_sG_l^{+}(I_n\otimes S)H_l'$	$\lvert S\rvert^{n+1}\prod_1^n s_{ii}^{-i}$ (= 0, if $\lvert S\rvert = 0$)	(3.38)
$G_l^{+}(I_n\otimes T)G_s$	$G_s^{+}(T^{-1}\otimes T^{-1})G_sG_l^{+}(T\otimes I_n)H_l'$	$\prod_1^n t_{ii}^{i}$	(3.39)

Proof. For (3.36) the invariance, under different choices of H_l, follows from (3.5), with $Y = XS'$. Apply (3.8) to obtain the inverse. For the determinant, take transposes in Exercise 1.27(ii) of Srivastava and Khatri [18]. Recall that $G_s' \in \mathscr{H}_l$, and note that if $|S| = 0$, then there exists $\mathbf{x} \neq \mathbf{0}$ with $\mathbf{x}'S = \mathbf{0}'$. Then

$$G_s(\mathbf{x}\otimes\mathbf{x}) = \underset{s}{\mathrm{vec}}\,\mathbf{x}\mathbf{x}' \neq \mathbf{0},$$

and $[G_s^{+}(\mathbf{x}\otimes\mathbf{x})]'G_s'(S\otimes I_n)G_l = (\mathbf{x}\otimes\mathbf{x})'(S\otimes I_n)G_l = (\mathbf{x}'S\otimes\mathbf{x}')G_l = \mathbf{0}'$, using (3.2iii). Thus $|G_s'(S\otimes I_n)G_l| = 0$ if $|S| = 0$.

The development of (3.37) is similar—consider $Y = TX$ with $X, Y, T \in \mathcal{M}_l$, apply Theorem 3.8 of Deemer and Olkin [2], then take transposes.

By (3.24), $G_l^+(S \otimes I_n)G_s = G_l^+(I_n \otimes S^{-1})G_s G_s^+(S \otimes S)G_s$. Now (3.38) follows from (3.37) and a further application of (3.24). The proof of (3.39) is similar. ∎

4. SOME PROPERTIES OF $G_p^+(A \otimes B)G_q$ FOR ARBITRARY A AND B

In this section, we evaluate determinants and inverses for some arrangements $G_p^+(A \otimes B)G_q$, where $A, B\colon n \times n$ are arbitrary. Some of our results generalize those of Neudecker [13], who established Corollary 4.1 and Theorem 4.3 for $A, B \in \mathcal{M}_l$, and Theorem 4.4 for simultaneously diagonalizable $A, B \in \mathcal{M}_s$.

DEFINITION 4.1. For $C\colon n \times n$,

$$|C|_i = \det(c_{ab})_{1 \leqslant a, b \leqslant i},$$

$$_i|C| = \det(c_{ab})_{i \leqslant a, b \leqslant n},$$

$$C^{(i)} = (c_{ab})_{1 \leqslant a \leqslant n, 1 \leqslant b \leqslant i},$$

$$C_{(n-i)} = (c_{ab})_{1 \leqslant a \leqslant n, i+1 \leqslant b \leqslant n}.$$

THEOREM 4.1. *For $A, B\colon n \times n$, $|G_l^+(A \otimes B)G_s| = 0$ if $|A||B| = 0$. If $|A||B| \neq 0$, then*

$$|G_l^+(A \otimes B)G_s| = |A||B|^n \prod_{i=1}^{n-1} |AB^{-1}|_i = \prod_{i=0}^{n} |A'^{(i)} \vdots B'_{(n-i)}|. \tag{4.1}$$

Proof. The proof of the first statement is similar to that used in establishing (3.36). Assume now that $|A||B| \neq 0$. By taking a perturbation if necessary, we may assume that $|AB^{-1}|_i \neq 0$ for all i. We can then write AB^{-1} as $T^{-1}S^{-1}$, where $S \in \mathcal{M}_u$, $T \in \mathcal{M}_l$, $s_{ii} = 1$, $t_{ii} = |AB^{-1}|_{i-1}/|AB^{-1}|_i$. (See, e.g., Exercise 1.20 of Srivastava and Khatri [18].) Applying first (3.12), then (3.24) gives

$$G_l^+(A \otimes B)G_s = G_l^+(I \otimes S)G_l \cdot G_l^+(I \otimes T)G_s \cdot G_s^+(A \otimes A)G_s. \tag{4.2}$$

The first equality in (4.1) then follows from (3.37) with $H_l = G_l'$, (3.39), and (3.25). Standard results on inverses and determinants of partitioned matrices give

$$|AB^{-1}|_i = \frac{|A'^{(i)} \vdots B'_{(n-i)}|}{|B|},$$

yielding the second equality in (4.1). ∎

COROLLARY 4.1. *For $A, B\colon n \times n$,*

$$|G_s^+(A \otimes B)G_l| = 2^{-n(n-1)/2}|A||B|^n \prod_{i=1}^{n-1} |B^{-1}A|_i$$

$$= 2^{-n(n-1)/2} \prod_{i=0}^{n} |A^{(i)} \vdots B_{(n-i)}|. \qquad (4.3)$$

Proof. Take transposes, recall (3.20), then use $G_s^{+\prime} = G_s D_s^{-1}$. ∎

Note that applications of (1.3) and (3.14) give analogous expressions for $|G_u^+(A \otimes B)G_s|$ and $|G_s^+(A \otimes B)G_u|$.

Assume now that $|G_l^+(A \otimes B)G_s| \neq 0$, so that $|AB^{-1}|_i \neq 0$ for all i and the decomposition $BA^{-1} = ST$ holds.

THEOREM 4.2. *With notation as above,*

$$[G_l^+(A \otimes B)G_s]^{-1} = G_s^+(A^{-1} \otimes A^{-1})[(T^{-1} \otimes T^{-1})G_l D_s G_l'(T \otimes T)]$$

$$\cdot (I \otimes AB^{-1})G_l,$$

where

$$G_l D_s G_l' = \bigoplus_{i=1}^{n} \operatorname{diag}(\underbrace{0,\ldots,0}_{i-1}, 1, \underbrace{2,\ldots,2}_{n-i}).$$

Proof. Applying (3.39), (3.24), (3.37) to (4.2), then using (3.26) and (3.12), then (3.24) gives

$$\begin{aligned}
[G_l^+(A \otimes B)G_s]^{-1} &= G_s^+(A^{-1} \otimes A^{-1})G_s \cdot G_s^+(T^{-1} \otimes T^{-1})G_s \\
&\quad \cdot G_l^+(T \otimes I)G_s \cdot G_l^+(I \otimes S^{-1})G_l \\
&= G_s^+(A^{-1} \otimes A^{-1})G_s \cdot G_s^+(T^{-1} \otimes T^{-1})G_s \\
&\quad \cdot G_l^+(T \otimes I)(I \otimes S^{-1})G_l \\
&= G_s^+(A^{-1} \otimes A^{-1})(T^{-1} \otimes T^{-1})G_s \\
&\quad \cdot G_l^+(T \otimes T)(I \otimes AB^{-1})G_l.
\end{aligned}$$

But by (3.27) and (3.10), $G_s G_l^+ = G_s^{+\prime} D_s G_l^+ = M_s G_l D_s G_l'$, whence the result follows by (3.22). ∎

COROLLARY 4.2. *Under the decomposition* $A^{-1}B = ST$, $S \in \mathcal{M}_u$, $T \in \mathcal{M}_l$, $t_{ii} = 1$,

$$[G_s^+(A\otimes B)G_l]^{-1} = G_l(I\otimes B^{-1}A)[(S\otimes S)G_l D_s G_l'(S^{-1}\otimes S^{-1})] \cdot (A^{-1}\otimes A^{-1})G_s.$$

Note from (3.10) that if B_{ij} is the submatrix consisting of the last $n-i+1$ rows and $n-j+1$ columns of B, then $G_l^+(A\otimes B)G_l = (a_{ij}B_{ij})_{1\le i,j\le n}$. The determinant of this block matrix has a surprisingly simple form.

THEOREM 4.3. *For* $A, B: n\times n$,

$$|G_l^+(A\otimes B)G_l| = |G_u^+(B\otimes A)G_u| = \prod_{i=1}^n |A|_i \cdot {}_i|B|. \tag{4.4}$$

Proof. The first equality follows from (3.14). By taking a perturbation, if necessary, we may represent A, B as $A = T_1 S_1$, $B = S_2 T_2$, where $T_1, T_2 \in \mathcal{M}_l$, $S_1, S_2 \in \mathcal{M}_u$, $S_{1,ii} = S_{2,ii} = 1$. Then by (3.16) and (3.12),

$$G_l^+(A\otimes B)G_l = G_l^+(I\otimes S_2)G_l\cdot G_l^+(T_1\otimes I)G_l\cdot G_l^+(S_1\otimes I)G_l\cdot G_l^+(I\otimes T_2)G_l,$$

so that by (3.12), (3.36), (3.10), and the construction of T_1,

$$\begin{aligned}|G_l^+(A\otimes B)G_l| &= |G_l^+(I\otimes S_2)G_l G_l^+(I\otimes T_2)G_l| \\ &\quad\cdot |G_l^+(T_1\otimes I)G_l||G_l^+(S_1\otimes I)G_l| \\ &= |G_l^+(I\otimes B)G_l|\cdot|T_1|^{n+1}\prod_{i=1}^n t_{1,ii}^{-i}\cdot 1 \\ &= \prod_{i=1}^n {}_i|B|\cdot|A|^{n+1}\cdot\prod_{i=1}^n\left(\frac{|A|_i}{|A|_{i-1}}\right)^{-i} \\ &= \prod_{i=1}^n |A|_i\cdot{}_i|B|.\end{aligned}$$

∎

THEOREM 4.4. *For* $A, B: n \times n$, $|G_s^+(A\otimes B)G_s| = 0$ *if* $|A||B| = 0$. *If* $|A||B| \neq 0$, *then*

$$|G_s^+(A\otimes B)G_s| = |A||B|^n \prod_{i<j} \frac{\lambda_i + \lambda_j}{2}, \tag{4.5}$$

where $\lambda_1, \dots, \lambda_n$ *are the characteristic roots of* AB^{-1}.

Proof. The proof of the first statement is similar to that used in establishing (3.36). Assume now that $|A||B| \neq 0$, and put $C = AB^{-1}$. By (3.24), $G_s^+(A\otimes B)G_s = G_s^+(C\otimes I)G_s \cdot G_s^+(B\otimes B)G_s$, so that by (3.25), $|G_s^+(A\otimes B)G_s| = |B|^{n+1}|G_s^+(C\otimes I)G_s|$. Now let $\lambda_1, \dots, \lambda_n$ be the roots of C, with characteristic vectors $\mathbf{x}_1, \dots, \mathbf{x}_n$. Using (3.2iii) and (3.28),

$$\begin{aligned} G_s^+(C\otimes I)G_s \cdot G_s^+(\mathbf{x}_i\otimes\mathbf{x}_j + \mathbf{x}_j\otimes\mathbf{x}_i) &= \lambda_i G_s^+(\mathbf{x}_i\otimes\mathbf{x}_j) + \lambda_j G_s^+(\mathbf{x}_j\otimes\mathbf{x}_i) \\ &= \frac{\lambda_i + \lambda_j}{2} \operatorname{vec}_s\left(\mathbf{x}_j\mathbf{x}_i' + \mathbf{x}_i\mathbf{x}_j'\right) \\ &= \frac{\lambda_i + \lambda_j}{2} G_s^+(\mathbf{x}_i\otimes\mathbf{x}_j + \mathbf{x}_j\otimes\mathbf{x}_i). \end{aligned}$$

Thus if the $(\lambda_i + \lambda_j)/2$ $(i \leqslant j)$ are all distinct, they are the roots of $G_s^+(C\otimes I)G_s$, whence $|G_s^+(C\otimes I)G_s| = |AB^{-1}|\prod_{i<j}(\lambda_i + \lambda_j)/2$. Taking a perturbation gives the general result. ■

COROLLARY 4.3. *For* $C: n \times n$,

$$\operatorname{tr} G_s^+(C\otimes I)G_s = \operatorname{tr} G_s^+(I\otimes C)G_s = \frac{n+1}{2}\operatorname{tr} C.$$

5. EXAMPLES

The results of Sections 2–4 allow one to easily manipulate many of the Jacobian matrices arising in multivariate statistical analysis. The three evaluations of Jacobian determinants given in Examples 5.1–5.3 below appear to be new. Magnus and Neudecker [8] discuss special cases of Example 5.1.

EXAMPLE 5.1. Let the transformation be $Y_{n\times n} = AX'B' + BXA'$, where $X \in \mathscr{M}_l$, $Y \in \mathscr{M}_s$, and A, B are constant and arbitrary. Since $\operatorname{vec} Y = [(B\otimes A)I_{(n,n)} + (A\otimes B)]\operatorname{vec} X = (I + I_{(n,n)})(A\otimes B)\operatorname{vec} X$, the Jacobian matrix is $\partial_s Y/\partial_l X = H_s(I + I_{(n,n)})(A\otimes B)G_l = 2G_s^+(A\otimes B)G_l$, by (3.23). Then by Corollary 4.1,

$$\left|\frac{\partial_s Y}{\partial_l X}\right| = 2^n|A||B|^n \prod_{i=1}^{n} |B^{-1}A|_i.$$

EXAMPLE 5.2. If the transformation is $Y = AXB' + BXA'$ with $X, Y \in \mathscr{M}_s$, then $\partial_s Y/\partial_s X = 2G_s^+(A\otimes B)G_s$ with, in the notation of Theorem 4.4,

$$\left|\frac{\partial_s Y}{\partial_s X}\right| = |B|^{n+1} \prod_{i\leqslant j} (\lambda_i + \lambda_j).$$

EXAMPLE 5.3. Let $Y_{n\times n}$ have all leading principal minors nonzero, so that we may represent it uniquely as $Y = TS$, $T \in \mathscr{M}_l$, $S \in \mathscr{M}_u$, $s_{ii} = 1$. Here, we show that the Jacobian J for this transformation has

$$|J|_+ = \prod_{i=1}^{n-1} |t_{ii}|^{n-i}. \tag{5.1}$$

Arrange the partial derivatives as

$$J = \begin{pmatrix} \dfrac{\partial_l Y_L}{\partial_l T} & \dfrac{\partial_l Y_L}{\partial_{ud} S} \\[2ex] \dfrac{\partial_{ud} Y_{UD}}{\partial_l T} & \dfrac{\partial_{ud} Y_{UD}}{\partial_{ud} S} \end{pmatrix}.$$

From $\operatorname{vec} Y = (S'\otimes I_n)\operatorname{vec} T$ we get, using (3.11), that $\operatorname{vec} Y_L = M_l \operatorname{vec} Y = M_l(S'\otimes I_n)\operatorname{vec} T$, so that $\partial_l Y_L/\partial_l T = G_l^+ M_l(S'\otimes I_n)G_l = G_l^+(S'\otimes I_n)G_l$. The other terms are calculated in a similar fashion, yielding $J = (G_l^{+\prime} \vdots G_{ud}^{+\prime})'(S'\otimes I \vdots I\otimes T)(G_l \oplus G_{ud})$. Noting that products ABC and BCA have the same nonzero characteristic roots, we have that J is the product of the roots of $(S'\otimes I)M_l + (I\otimes T)M_{ud}$. Using (3.10) and Lemma 3.2, we see that this matrix is lower triangular, with diagonal $\oplus_{i=1}^{n} \operatorname{diag}(t_{11}, \ldots, t_{i-1,i-1}, s_{ii}, \ldots, s_{ii})$, from which (5.1) follows.

The next example uses the following result, due to S. N. Roy, from Olkin and Roy [15].

THEOREM 5.1. *Let* $\mathbf{x} \in \mathbb{R}^{m+n}$ *be subject to* n *constraints* $\mathbf{g}(\mathbf{x}) = \mathbf{0}_{n \times 1}$. *Partition* $\mathbf{x}$ *as*

$$\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} \begin{matrix} m \\ n \end{matrix},$$

and let $\mathbf{y} \in \mathbb{R}^m$ *be a function of* $\mathbf{x}$. *Then, under the conditions of the implicit-function theorem,*

$$\left| \frac{\partial \mathbf{y}}{\partial \mathbf{x}_1} \right|_+ = \left| \frac{\partial \mathbf{y}, \mathbf{g}(\mathbf{x})}{\partial \mathbf{x}} \right|_+ \Big/ \left| \frac{\partial \mathbf{g}(\mathbf{x})}{\mathbf{x}_2} \right|_+ .$$

EXAMPLE 5.4. If $Y: p \times n$ has rank $p \leqslant n$, then we may express it as $Y_{p \times n} = T_{p \times p} \Gamma_{p \times n}$, where $T \in \mathcal{M}_l$, $t_{ii} > 0$, and $\Gamma\Gamma' = I_p$. This transformation is central to the derivation of the Wishart distribution, and its Jacobian was evaluated by Olkin and Roy [15], using Theorem 5.1 above. We believe that the following derivation, also based on Theorem 5.1, is somewhat simpler. In the theorem, put $\mathbf{y} = \operatorname{vec} Y: pn \times 1$. Partition Γ as $\Gamma = (\Lambda_{p \times p} \,\vdots\, \Omega_{p \times (n-p)})$. Put

$$\mathbf{x}_1 = \begin{pmatrix} \underset{l}{\operatorname{vec}} T \\ \underset{ud}{\operatorname{vec}} \Lambda \\ \operatorname{vec} \Omega \end{pmatrix} \begin{matrix} p(p+1)/2 \\ p(p-1)/2 \\ p(n-p) \end{matrix} : pn \times 1, \qquad \mathbf{x}_2 = \underset{l}{\operatorname{vec}} \Lambda,$$

$$\mathbf{g}(\mathbf{x}) = \operatorname{vec}\left(\Gamma\Gamma' - I_p\right), \qquad \mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}.$$

Then $\mathbf{x}$ is a permutation of

$$\begin{pmatrix} \operatorname{vec} & \Gamma \\ \underset{l}{\operatorname{vec}} & T \end{pmatrix},$$

and $\Gamma\Gamma' = I_p$ are the $p(p+1)/2$ constraints of the theorem, regarded as constraining $\underset{l}{\operatorname{vec}} \Lambda$. The absolute value of the Jacobian is then

$$|J|_+ = \left| \frac{\partial \mathbf{y}}{\partial \mathbf{x}_1} \right|_+ = \left| \frac{\partial \operatorname{vec} Y, \underset{s}{\operatorname{vec}} \Gamma\Gamma'}{\partial \operatorname{vec} \Gamma, \underset{l}{\operatorname{vec}} T} \right|_+ \Big/ \left| \frac{\partial \underset{s}{\operatorname{vec}} \Gamma\Gamma'}{\partial \underset{l}{\operatorname{vec}} \Lambda} \right|_+ .$$

To determine

$$\frac{\partial \underset{s}{\operatorname{vec}} \Gamma\Gamma'}{\partial \underset{l}{\operatorname{vec}} \Lambda},$$

we take differentials in $\Gamma\Gamma' = \Lambda\Lambda' + \Omega\Omega'$, holding Ω fixed, and obtain

$$\underset{s}{\operatorname{vec}}\, d\Gamma\Gamma' = 2G_s^+\left(\Lambda \otimes I_p\right)G_l \underset{l}{\operatorname{vec}}\, d\Lambda.$$

By Corollary 4.1,

$$\left|\frac{\partial \underset{s}{\operatorname{vec}} \Gamma\Gamma'}{\partial \operatorname{vec} \Lambda}\right|_+ = 2^p \prod_{i=1}^{p} \left(|\Lambda|_i\right)_+.$$

We easily find that

$$\left(\frac{\partial \operatorname{vec} Y, \underset{s}{\operatorname{vec}} \Gamma\Gamma'}{\partial \operatorname{vec} \Gamma, \underset{l}{\operatorname{vec}} T}\right) = \begin{pmatrix} I_n \otimes T & \left(\Gamma' \otimes I_p\right)G_l \\ 2G_s\left(\Gamma \otimes I_p\right) & 0 \end{pmatrix}$$

with determinant $|T|^n| - 2G_s^+(\Gamma \otimes I_p)(I_n \otimes T^{-1})(\Gamma' \otimes I_p)G_l| = (-2)^{p(p+1)/2}|T|^n|G_s^+(I_p \otimes T^{-1})G_l|$. This is evaluated by writing G_s^+ as $D_s^{-1}G_s'$, and then taking transposes in (3.37). The result is $(-1)^{p(p+1)/2}2^p|T|^n\prod_{i=1}^p t_{ii}^{-i}$, whence

$$|J|_+ = \prod_{i=1}^{p} \frac{t_{ii}^{n-i}}{\left(|\Lambda|_i\right)_+}.$$

EXAMPLE 5.5. We first recall some recent results of Tyler [20]. Suppose that $y_1, \ldots, y_n$ is a random sample from a p-dimensional elliptically symmetric population with density $f(y) = |\Sigma|^{-1/2} g((y - \mu)'\Sigma^{-1}(y - \mu))$, where $\Sigma > 0$. Let $\{V_n\}$ be a consistent sequence of affinely invariant estimators of Σ—e.g. sample covariances, maximum-likelihood estimators, robust M-estimators. Put $Z_n = \Sigma^{-1/2}V_n\Sigma^{-1/2\prime}$, where $\Sigma = \Sigma^{1/2}\Sigma^{1/2\prime}$, $V_n = V_n^{1/2}V_n^{1/2\prime}$. Then Z_n is a consistent estimator of I_p, based upon a sample with density $f(z) = g(z'z)$. Suppose that

$$\sqrt{n}\left(Z_n - I_p\right) \xrightarrow{\mathscr{L}} Z, \tag{1}$$

where Z has a radial distribution, i.e. $QZQ' \sim Z$ for all orthogonal Q. Tyler shows that the covariance matrix of any such Z is of the form

$$\operatorname{cov}[\operatorname{vec} Z] := C = \sigma_1(I + I_{(p,p)}) + \sigma_2(\operatorname{vec} I)(\operatorname{vec} I)' \tag{5.2}$$

for some constants $\sigma_1 \geqslant 0$, $\sigma_2 \geqslant -2\sigma_1/p$ defined by $\operatorname{VAR}[z_{12}] = \sigma_1$, $\operatorname{COV}[z_{11}, z_{22}] = \sigma_2$, $\operatorname{VAR}[z_{11}] = 2\sigma_1 + \sigma_2$. Under quite general conditions,

$$\operatorname{vec} Z \sim N_{p^2}(\mathbf{0}, C). \tag{2}$$

If (1) and (2) hold, then

$$n \operatorname*{vec}_s (Z_n - I_p) \xrightarrow{\mathscr{L}} N_q(\mathbf{0}, H_s C H_s') \qquad [q = p(p+1)/2],$$

with a nonsingular covariance matrix if $\sigma_2 > -2\sigma_1/p$; hence

$$n\left(\operatorname*{vec}_s (Z_n - I_p)\right)'(H_s C H_s')^{-1} \operatorname*{vec}_s (Z_n - I_p) \xrightarrow{\mathscr{L}} \chi_q^2.$$

With

$$C_\Sigma : q \times q := H_s(\Sigma^{1/2} \otimes \Sigma^{1/2}) G_s \cdot H_s C H_s' \cdot [H_s(\Sigma^{1/2} \otimes \Sigma^{1/2}) G_s]',$$

we then have

$$n\left(\operatorname*{vec}_s (V_n - \Sigma)\right)' C_\Sigma^{-1}\left(\operatorname*{vec}_s (V_n - \Sigma)\right) \xrightarrow{\mathscr{L}} \chi_q^2. \tag{5.3}$$

A Taylor series expansion of V_n^{-1} about Σ^{-1}, using (2.5) and (3.24), shows that, with

$$D_\Sigma : q \times q := [H_s(\Sigma^{-1/2\prime} \otimes \Sigma^{-1/2\prime}) G_s] H_s C H_s' [H_s(\Sigma^{-1/2\prime} \otimes \Sigma^{-1/2\prime}) G_s]',$$

we have

$$n\left(\operatorname*{vec}_s (V_n^{-1} - \Sigma^{-1})\right)' D_\Sigma^{-1}\left(\operatorname*{vec}_s (V_n^{-1} - \Sigma^{-1})\right) \xrightarrow{\mathscr{L}} \chi_q^2. \tag{5.4}$$

Suppose that σ_1, σ_2 either are known, or have consistent estimators $\hat{\sigma}_1, \hat{\sigma}_2$. Since V_n is consistent for Σ, one can obtain asymptotic, ellipsoidal confidence regions for Σ and Σ^{-1} by replacing C_Σ with C_{V_n} and D_Σ with D_{V_n} in (5.3) and (5.4) respectively. In fact, (5.3) already describes an ellipsoidal confidence region for Σ^{-1}, and (5.4) an ellipsoidal region for Σ, without further estimation of C_Σ or D_Σ.

LEMMA 5.1. *With Z_n as above, put $k = \sigma_2/(2\sigma_1 + p\sigma_2)$,*

$$X_n = \frac{n}{2\sigma_1}\left[\operatorname{vec}(Z_n - I_p)\right]'\left[I - k(\operatorname{vec} I)(\operatorname{vec} I)'\right]\operatorname{vec}(Z_n - I_p).$$

Then X_n and the following three quantities are equal:

$$\frac{n}{2\sigma_1}\left[\operatorname{tr}(Z_n - I_p)^2 - k\operatorname{tr}^2(Z_n - I_p)\right], \tag{i}$$

$$n\left(\operatorname{vec}_s(V_n - \Sigma)\right)' C_\Sigma^{-1} \operatorname{vec}_s(V_n - \Sigma), \tag{ii}$$

$$n\left(\operatorname{vec}_s(V_n^{-1} - \Sigma^{-1})\right)' D_{V_n}^{-1} \operatorname{vec}_s(V_n^{-1} - \Sigma^{-1}). \tag{iii}$$

Put

$$Y_n = \frac{n}{2\sigma_1}\left[\operatorname{vec}(Z_n^{-1} - I_p)\right]'\left[I - k(\operatorname{vec} I)(\operatorname{vec} I)'\right]\operatorname{vec}(Z_n^{-1} - I_p).$$

Then Y_n and the following three quantities are equal:

$$\frac{n}{2\sigma_1}\left[\operatorname{tr}(Z_n^{-1} - I_p)^2 - k\operatorname{tr}^2(Z_n^{-1} - I_p)\right], \tag{iv}$$

$$n\left(\operatorname{vec}_s(V_n^{-1} - \Sigma^{-1})\right)' D_\Sigma^{-1} \operatorname{vec}_s(V_n^{-1} - \Sigma^{-1}), \tag{v}$$

$$n\left(\operatorname{vec}_s(V_n - \Sigma)\right)' C_{V_n}^{-1} \operatorname{vec}_s(V_n - \Sigma). \tag{vi}$$

Proof. We prove the first part; the second is similar. That X_n equals the quantity at (i) follows from the identity $(\operatorname{vec} A)'\operatorname{vec} B = \operatorname{tr} A'B$. Consider the expression (iii). Using (3.23), (3.20) we find

$$H_s C H_s' = 2\sigma_1 D_s^{-1} + \sigma_2 \Big(\underset{s}{\operatorname{vec}} I \Big) \Big(\underset{s}{\operatorname{vec}} I \Big)',$$

so that

$$\left(H_s C H_s' \right)^{-1} = \frac{1}{2\sigma_1} G_s' \left[I - k(\operatorname{vec} I)(\operatorname{vec} I)' \right] G_s.$$

Then using (2.2), (3.24), and the identity $\operatorname{tr}(ABC) = \operatorname{tr}(BCA)$, the quantity at (iii) is

$$\frac{1}{2\sigma_1} \left[\operatorname{vec}\left(V_n^{1/2\prime} \Sigma^{-1} V_n^{1/2} - I \right) \right]' \left[I - k(\operatorname{vec} I)(\operatorname{vec} I)' \right]$$

$$\cdot \operatorname{vec}\left(V_n^{1/2\prime} \Sigma^{-1} V_n^{1/2} - I \right)$$

$$= \frac{1}{2\sigma_1} \left[\operatorname{tr}\left(V_n^{1/2\prime} \Sigma^{-1} V_n^{1/2} - I \right)^2 - k \operatorname{tr}^2\left(V_n^{1/2\prime} \Sigma^{-1} V_n^{1/2} - I \right) \right]$$

$$= \frac{1}{2\sigma_1} \left[\operatorname{tr}(Z_n - I)^2 - k \operatorname{tr}^2(Z_n - I) \right] = \frac{X_n}{n}$$

The reduction of (ii) to X_n is similar. ■

The following is now immediate.

THEOREM 5.2. *Suppose that* σ_1, σ_2 *are known. If conditions* (1) *and* (2) *in Example* 5.5 *hold, then*

$$X_n, Y_n \overset{\mathscr{L}}{\to} \chi_q^2.$$

Whether (1) *and* (2) *hold or not, let* F_n, G_n *be the d.f.'s of* X_n, Y_n *respectively. Then a* $100(1-\alpha)\%$ *confidence ellipsoid for* $\underset{s}{\operatorname{vec}} \Sigma^{-1}$ *is*

$$n\Big(\underset{s}{\operatorname{vec}} \left(\Sigma^{-1} - V_n^{-1} \right) \Big)' D_{V_n}^{-1} \underset{s}{\operatorname{vec}} \left(\Sigma^{-1} - V_n^{-1} \right) \leqslant F_n^{-1}(1-\alpha). \tag{5.5}$$

A 100(1 − α)% confidence ellipsoid for $\underset{s}{\operatorname{vec}}\Sigma$ *is*

$$n\left(\underset{s}{\operatorname{vec}}(\Sigma - V_n)\right)' C_{V_n}^{-1} \underset{s}{\operatorname{vec}}(\Sigma - V_n) \leqslant G_n^{-1}(1-\alpha). \tag{5.6}$$

Consider now the region described by (5.5), and assume that we sample from a $N_p(\mu, \Sigma)$ population. Let V_n be the unbiased sample covariance matrix $\Sigma_1^n(y_i - \bar{y})(y_i - \bar{y})'/(n-1)$. For normal populations, $\sigma_1 = 1$, $\sigma_2 = 0$, $k = 0$, and $(n-1)Z_n \sim W_p(I, n-1)$, the Wishart distribution. Put $N = n - 1$; let $U_N \sim W_p(I, N)$. In order that the statistic have, in finite samples, its asymptotic expectation, we set $\tilde{X}_n = [(n-1)/n]X_n$, with d.f. $\tilde{F}_n$. Then

$$E[\tilde{X}_n] = q, \quad \tilde{X}_n \overset{\mathscr{L}}{\to} \chi_q^2, \quad E[U_N] = NI_p, \text{ and } \tilde{X}_n \sim \frac{1}{2N}\operatorname{tr}(U_N - NI_p)^2.$$

The confidence region becomes

$$\tilde{X}_n = N\left[\operatorname{vec}(\Sigma^{-1} - V_n^{-1})\right]'(V_n \otimes V_n)\operatorname{vec}(\Sigma^{-1} - V_n^{-1}) \leqslant \tilde{F}_n(1-\alpha).$$

We note that Nagao [11] has proposed $\tilde{X}_n$ (his "T_1") as a test statistic for the hypothesis $H_0: \Sigma = \Sigma_0$ against $H_A: \Sigma \neq \Sigma_0$, with $U_N = NV_n\Sigma_0^{-1}$. Nagao obtains an expansion of the form $F_n^{-1}(1-\alpha) = \chi_q^2(1-\alpha) + c_\alpha/N + O(N^{-2})$, with c_α given explicitly.

For the region (5.6) we proceed similarly. Let

$$V_n = \sum_1^n \frac{(y_i - \bar{y})(y_i - \bar{y})'}{N - p - 1},$$

set

$$c_{P,N} = \frac{(p+1)(N-1)(N-p+3)}{(N+1)\left[N(p+1) - (p+1)^2 + 2\right]},$$

and let $\tilde{Y}_n = c_{P,N}Y_n$, with d.f. $\tilde{G}_n$. Then

$$\tilde{Y}_n \sim \frac{(N+1)(N-p-1)^2}{2} c_{P,N} \operatorname{tr}\left(U_N^{-1} - \frac{I_p}{N-p-1}\right)^2,$$

with

$$E[\tilde{Y}_n] = q, \qquad \tilde{Y}_n \overset{\mathscr{L}}{\to} \chi_q^2, \qquad E[U_N^{-1}] = \frac{I_p}{N-p-1}.$$

The confidence region becomes

$$\tilde{Y}_n = nc_{P,N}\left[\operatorname{vec}(\Sigma - V_n)\right]'\left(V_n^{-1}\otimes V_n^{-1}\right)\operatorname{vec}(\Sigma - V_n) \leqslant \tilde{G}_n^{-1}(1-\alpha).$$

See Wiens [21] for asymptotic expansions of $\tilde{G}_n$ and $\tilde{G}_n^{-1}$.

REFERENCES

1 P. M. Bentler and S. Y. Lee, Matrix derivatives with chain rule and rules for simple, Hadamard and Kronecker products. *J. Math. Psychol.* 17:225–262 (1978).

2 W. L. Deemer and I. Olkin, Jacobians of matrix transformations useful in multivariate analysis, *Biometrika* 38:345–367 (1951).

3 A. Graham, *Kronecker Products and Matrix Calculus with Applications*, Halsted (Wiley), New York, 1981.

4 H. V. Henderson and S. R. Searle, vec and vech operators for matrices, with some uses in Jacobians and multivariate statistics, *Canad. J. Statist.* 7:65–81 (1979).

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

6 E. C. MacRae, Matrix derivatives with an application to an adaptive linear decision problem, *Ann. Statist.* 2:337–346 (1974).

7 J. R. Magnus and H. Neudecker, The commutation matrix: Some properties and applications, *Ann. Statist.* 7:381–394 (1979).

8 J. R. Magnus and H. Neudecker, The elimination matrix: Some lemmas and applications, *SIAM J. Algebraic Discrete Methods* 1:422–449 (1980).

9 C. E. McCulloch, Symmetric matrix derivatives with applications, *J. Amer. Statist. Assoc.* 77:679–682 (1982).

10 R. P. McDonald and H. Swaminathan, A simple matrix calculus with applications to multivariate analysis, *Gen. Systems* 18:37–54 (1973).

11 H. Nagao, On some test criteria for covariance matrix, *Ann. Statist.* 1(4): 700–709 (1973).

12 D. G. Nel, On matrix differentiation in statistics, *South African Statist. J.* 14:137–193 (1980).

13 H. Neudecker, Some results on Jacobians, Report AE1/78, Faculty of Actuarial Science and Econometrics, Univ. of Amsterdam, 1978.

14 H. Neudecker and T. Wansbeek, Some results on commutation matrices, with statistical applications, *Canad. J. Statist.* 11:221–231 (1983).

15 I. Olkin and S. N. Roy, On multivariate distribution theory, *Ann. Math. Statist.* 25:329–339 (1954).

16 G. S. Rogers, *Matrix Derivatives*, Marcel Dekker, New York, 1980.

17 W. E. Roth, On direct product matrices, *Bull. Amer. Math. Soc.* 40:461–468 (1934).

18 M. S. Srivastava and C. G. Khatri, *An Introduction to Multivariate Analysis*, North Holland, New York, 1979.

19 D. S. Tracy and R. P. Singh, Some modifications of matrix differentiation for evaluating Jacobians of symmetric matrix transformations, in *Symmetric Functions in Statistics, Proceedings of a Symposium in Honor of Paul S. Dwyer* (Derrick S. Tracy, Ed.), Univ. Windsor, Windsor, Ontario, 1972, pp. 203–224.

20 D. E. Tyler, Radial estimates and the test for sphericity, *Biometrika*, 69, 2, 429–436 (1982).

21 D. Wiens, Ellipsoidal confidence regions for a normal covariance matrix, *Comm. Statist. A*, to appear.

Received 7 December 1983; revised 19 November 1984