CHAPTER 10

Application to Type II Problems: No Special Group Structure, But Global Cross Section Exists

10.1. Characteristic roots of a positive definite matrix. Let $S \in PD(p)$ be a random matrix with distinct characteristic roots $\lambda_1 > \cdots > \lambda_p > 0$ and distribution $p(S)(dS)$ as in Examples 8.1 and 8.7. The results of Example 8.7 can be copied by changing n to p, with the results

$$(dS) = 2^{-p} \prod_{i<j} (\lambda_i - \lambda_j) \mu_y(dy)(d\Lambda) \quad \text{at} \quad S = \Lambda, \tag{10.1.1}$$

$$P(d\Lambda) = 2^{-p} \prod_{i<j} (\lambda_i - \lambda_j)(d\Lambda) \int p(\Gamma \Lambda \Gamma') \mu_{O(p)}(d\Gamma). \tag{10.1.2}$$

Equation (10.1.2) can also be found in Muirhead (1982), Theorem 3.2. 17, (but note that Muirhead's Haar measure on $O(p)$ is normalized). If in (10.1.2) $p(S)$ depends on S only through Λ, then formula (2) of Theorem 13.2.1 in Anderson (1984) is reproduced. In particular, if $S \sim W(n, I_p)$ (take formula (9.2.8) with $\Sigma = I_p$), then the result is

$$\begin{aligned} P(d\Lambda) = {} & (2\pi)^{-\frac{1}{2}pn} 2^{-2p} c_n c_{n-p}^{-1} c_p \\ & \cdot \prod_{i<j} (\lambda_i - \lambda_j) |\Lambda|^{\frac{1}{2}(n-p-1)} e^{-\frac{1}{2} \operatorname{tr} \Lambda} (d\Lambda) \end{aligned} \tag{10.1.3}$$

(cf. Anderson, 1984, Section 13.3, (11)). Formula (10.1.2) will also be applied to the singular values in Section 10.2 and indirectly to the characteristic roots of Q in Section 10.3.

10.2. Singular values. Let X be a random $q \times p$ matrix of rank $s = \min(p, q)$ and distribution $p(X)(dX)$. Its singular values $\ell_1 > \cdots > \ell_s > 0$ (strict inequalities with probability one) are the square roots of the characteristic roots $\lambda_1 > \cdots > \lambda_s > 0$ of $X'X$ if $q \geq p$, or of XX' if $q < p$. Let $L : q \times p$ be a matrix with $\ell_1, \ldots, \ell_s$ on, and zeros off, the diagonal. The distribution of L may be obtained by putting $S = X'X$ or XX' according as $q \geq$ or $< p$; then apply (9.2.7) (with p and q interchanged if $q < p$), followed by an application of (10.1.2). This actually produces the joint distribution of the ℓ_i^2. Then use $d\ell_i^2 = 2\ell_i d\ell_i$. The result is

$$P(dL) = 2^{-s} c_{|p-q|}^{-1} \prod_{i=1}^{s} \ell_i^{|p-q|} \prod_{i<j} (\ell_i^2 - \ell_j^2)(dL) \cdot \int p(\Gamma L \Delta') \mu_{O(q)}(d\Gamma) \mu_{O(p)}(d\Delta), \tag{10.2.1}$$

in which $s = \min(p, q)$ and c_n is defined in (7.7.9). Distribution of singular values is also treated in Farrell (1985) with group methods.

10.3. Characteristic roots of $Q = US^{-1}U'$. Let U and S be as in Section 9.4. If $q \leq p$, then the distribution of the characteristic roots of Q could be written down by using (9.4.3) followed by (10.1.2). When $q > p$ this procedure fails since Q is then singular and (10.1.2) does not apply. However, it is possible to avoid this difficulty by taking a different route. Take X of (9.3.3) and observe that the squares of the singular values ℓ_i of X are the nonzero characteristic roots λ_i of $Q = XX'$. Thus, apply (10.2.1) to X having distribution (9.3.6). This involves three integrals, in which the integration over $\prod_1^p t_{ii}^{-i}(dT) = \mu_{LT(p)}(dT)$ and over $\mu_{O(p)}(d\Delta)$ can be contracted to one integration over $\mu_{GL(p)}(dC) = |C|^{-p}(dC)$, by (7.7.10) and (7.7.1). The result is

$$\begin{aligned} P(dL) &= 2^{p-s}c_{|p-q|}^{-1}\prod_{i=1}^{s}\ell_i^{|p-q|}\prod_{i<j}(\ell_i^2-\ell_j^2)(dL) \\ &\quad\cdot\int p(\Gamma LC', CC')|C|^{q+1}(dC)\mu_{O(q)}(d\Gamma), \end{aligned} \tag{10.3.1}$$

in which $p(U,S)$ is the density of (U,S) and $s = \min(p,q)$. Formula (10.3.1) gives the joint distribution of the ℓ_i $(i = 1, \ldots, s)$. The joint distribution of the characteristic roots λ_i follows easily from $\lambda_i = \ell_i^2$. If the rows of U are iid $N(0, I_p)$ and S is an independent $W(n, I_p)$ matrix, then the integration in (10.3.1) can easily be carried out and reproduces the well-known distribution of the characteristic roots in the central MANOVA problem (Anderson, 1984, Section 13.2; Muirhead, 1982, Section 10.4). Note that our formula (10.3.1) takes care of all p, q so that it is not necessary (as is done in the conventional derivation, see, e.g., Anderson 1984, Section 13.2.4) to first demonstrate how the case $q > p$ can be reduced to $q < p$.

10.4. Characteristic roots of $S_1^{-1}S_2$. Let $S_1, S_2 \in PD(p)$ with joint density $p(S_1, S_2)$. The distribution of the characteristic roots of $S_1^{-1}S_2$ was treated by Koehn (1970) as a type II problem, starting from scratch. However, we may obtain the result also by combining the results of Section 9.5 and 10.1. Let $S_1 = T_S T_S'$, $T_S \in PD(p)$, be the Cholesky decomposition of S_1 and put $U = T_S^{-1}S_2T_2'^{-1}$. The characteristic roots λ_i of $S_1^{-1}S_2$ are the same as those of U. The distribution of the latter follows from (9.5.1) with $k = 1$ and the present U is U_1 in (9.5.1). Then apply (10.1.2). The resulting double integral over $LT(p)$ and $O(p)$ can be contracted, as in Section 10.3, to one over $GL(p)$, by (7.7.10). The result is

$$P(d\Lambda) = \prod_{i<j}(\lambda_i - \lambda_j)(d\Lambda)\int p(CC', C\Lambda C')|C|^{p+2}(dC). \tag{10.4.1}$$

This result will also find application in the next section.

10.5. Canonical correlations. Let $S \in PD(p+q)$ be partitioned $S = ((S_{ij}))$, $i, j = 1, 2$, with $S_{11} : p \times p$; assume $p \leq q$. Let the distribution of S be $P(dS) = p(S_{11}, S_{12}, S_{22})(dS_{11})(dS_{12})(dS_{22})$. The canonical correlations $r_1 > \cdots > r_p > 0$ (strict inequalities with probability one) are the square roots of the characteristic roots $\lambda_1 > \cdots > \lambda_p > 0$ of $S_{11}^{-1}S_{12}S_{22}^{-1}S_{21}$ and constitute a maximal invariant under the group $G = GL(p) \times GL(q)$ with action $S_{11} \to BS_{11}B'$, $S_{12} \to BS_{12}C'$, $S_{22} \to CS_{22}C'$, $B \in GL(p)$, $C \in GL(q)$. It is again possible to avoid deriving the joint distribution of the r_i from scratch as a Type II problem by combining the results from Sections 9.4 and 10.4. Let $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_p)$ and let $R = ((r_{ij}))$ be a $p \times q$ matrix with $r_{ij} = 0$ for $i \neq j$, and $r_{ii} = r_i = \lambda_i^{1/2}$. Then $RR' = \Lambda$. Rename S_{11} by S_1 and put $S_{12}S_{22}^{-1}S_{21} = S_2$, so that the λ_i are the characteristic roots of $S_1^{-1}S_2$. The joint density of S_1 and S_2 follows from (9.4.3) applied to $Q = S_2$. Then apply (10.4.1) (with C replaced by another symbol, say B). This produces the distribution of Λ, and that of R follows from $\Lambda = RR'$. In the first integration (9.4.3) the integrand contains the function $p(S_1, XC', CC')$, where X is any matrix such that $XX' = S_2$. Then in the second integration (10.4.1), S_1 is replaced by BB' and X by any matrix such that $XX' = B\Lambda B'$ (since in (10.4.1) the second argument of p is to be replaced by $B\Lambda B'$). We may take $X = BR$. Thus, in the double integral the arguments of p are $p(BB', BRC', CC')$. The final result is

$$
\begin{aligned}
P(dR) = 2^q c_{q-p}^{-1} \prod_{i=1}^{p} r_i^{q-p} \prod_{i<j} (r_i^2 - r_j^2)(dR) \\
\cdot \int p(BB', BRC', CC')|B|^{q+1}|C|^{p+1}(dB)(dC)
\end{aligned}
\tag{10.5.1}
$$

in which the integration may be taken over all $B : p \times p$ and $C : q \times q$. The constant c_n is defined in (7.7.9). IF S is a Wishart matrix and the two sets of variates are independent, then (10.5.1) easily reproduces the null distribution of canonical correlations, as given, for instance, in Anderson (1984), Section 13.4; Kshirsager (1972), Section 7.5; Giri (1977), Section 10.3.2; Muirhead (1982), Corollary 11.3.3.

10.6. Simultaneous orthogonal reduction of several positive definite matrices. We shall only treat here the simplest case of two matrices; the extension to more than two matrices is straightforward. Let $\mathcal{X}$ consist of all (S_1, S_2), $S_i \in PD(p)$, $i = 1, 2$, and let $G = O(p)$ with action $S_i \to \Gamma S_i \Gamma'$, $i = 1, 2$, $\Gamma \in G$. In order to construct a maximal invariant, via a cross section $\mathcal{Z}$, we may, for instance, reduce S_2 to diagonal form $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_p)$ by a transformation with $\Gamma \in G$, and let S be the resulting transformed S_1. That is, suppose $\Gamma_S \in G$ is such that $\Gamma_S S_2 \Gamma'_S = \Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_p)$, then let $S = \Gamma_S S_1 \Gamma'_S$. By removing from $\mathcal{X}$ a set of Lebesgue measure 0 we may assume $\lambda_1 > \cdots > \lambda_p > 0$. The resulting cross section $\mathcal{Z}$ consists of all pairs of matrices of the form $z = (S, \Lambda)$. The only members of G that leave Λ invariant are matrices of the form $E = \operatorname{diag}(\pm 1, \ldots, \pm 1)$. Then the isotropy subgroup of G at $z \in \mathcal{Z}$ consists of all those E that also leave S invariant. For most z there are no such E except the identity I_p. In order to guarantee that G_z be the same group G_0 for all $z \in \mathcal{Z}$ we have to require that for every $z \in \mathcal{Z}$ no E leaves S invariant except $E = I_p$. Define $PD^*(p)$ to be those $S \in PD(p)$ that do not commute with any E but $E = I_p$. Then remove from $\mathcal{X}$ the set of Lebesgue measure 0 consisting of all orbits of points $z = (S, \Lambda)$ with $S \notin PD^*(p)$. Equivalently, remove all (S_1, S_2) for which the two matrices have a common invariant subspace in their spectral decomposition of R^p. We end up with $\mathcal{X}$ such that $G_z = G_0 = \{e\}$ for every $z \in \mathcal{Z}$.

Denote $\lambda = (\lambda_1, \ldots, \lambda_p) \in R^p$ and define $\mathcal{T} = \{t = (S, \lambda) : S \in PD^*(p), \lambda_1 > \cdots > \lambda_p > 0\}$. Then define the function s of (8.6) by $s(t) = (S, \Lambda) \in \mathcal{X}$ (these are of course the points of $\mathcal{Z}$). Since $G_0 = \{e\}$ we have $\mathcal{Y} = G$. Equation (8.7), which can be written $x = gs(t)$, reads here

$$S_1 = \Gamma S \Gamma', \quad S_2 = \Gamma \Lambda \Gamma', \quad \Gamma \in G. \tag{10.6.1}$$

Differentiate (10.6.1) at $\Gamma = I_p$ and take the wedge product $(dS_1)(dS_2)$. The computation of (dS_2) was done already in Example 8.7, with the result $(dS_2) = \prod_{i<j}(\lambda_i - \lambda_j)(d\Lambda)(d\Gamma)$. The wedge product (dS_1)

also has factors that contain terms with $d\gamma_{ij}$. However, every such term is multiplied by a corresponding $d\gamma_{ij}$ of $(d\Gamma)$ in (dS_2), and since $d\gamma_{ij} \wedge d\gamma_{ij} = 0$, all such terms drop out. The result is the same as if we had $(dS_1) = (dS)$. Therefore,

$$(dS_1)(dS_2) = \prod_{i<j}(\lambda_i - \lambda_j)(dS)(d\Lambda)(d\Gamma) \quad \text{at} \quad y = [e]. \tag{10.6.2}$$

The left-hand side of (10.6.2) is $\lambda(dx)$ of (8.15), and $(d\Gamma)$ of the right-hand side is $\mu_{\mathcal{Y}}(dy)$ of (8.15). By comparison of (8.15) and (10.6.2) we have therefore

$$\mu_{\mathfrak{T}}(dt) = \prod_{i<j}(\lambda_i - \lambda_j)(dS)(d\Lambda) \tag{10.6.3}$$

and the factorization (8.14) becomes (note that $\chi = 1$ since G is orthogonal)

$$(dS_1)(dS_2) = \prod_{i<j}(\lambda_i - \lambda_j)(dS)(d\Lambda)\mu_{O(p)}(d\Gamma) \tag{10.6.4}$$

in which the various matrices are related by (10.6.1). If the distribution of (S_1, S_2) is $p(S_1, S_2)(dS_1)(dS_2)$, then by the general formula (8.12) the distribution of the maximal invariant $T = (S, \lambda)$ is

$$P(dS, d\Lambda) = \prod_{i<j}(\lambda_i - \lambda_j)(dS)(d\Lambda) \cdot \int p(\Gamma S \Gamma', \Gamma\Lambda\Gamma')\mu_{O(p)}(d\Gamma). \tag{10.6.5}$$

This problem occurs, for instance, as a final invariance reduction in GMANOVA, and also in Example 11.6.

10.7. Covariance matrix of complex structure. This section deals with the distributional aspects of a problem treated in Andersson, Brøns, and Jensen (1983), Section 2, and also in Andersson and Perlman (1984). The problem itself is to test the hypothesis that a covariance matrix has complex structure. There is a related problem, also treated in the above mentioned references, to test the hypothesis that a covariance matrix of complex structure has real structure. We shall only consider the first problem here and refer the reader to Wijsman (1986), Section 7.7(b), for an application of the method of cross sections to the second problem. Distributions of maximal invariants were derived by Andersson, Brøns, and Jensen (1983) and by Andersson and Perlman (1984) under the assumption that the covariance matrix S has a Wishart distribution. Here we shall derive an integral expression for the distribution of a maximal invariant using a global cross section, assuming about the distribution of S only that it is absolutely continuous with respect to Lebesgue measure.

A matrix $C \in GL(2p)$ is said to have **complex structure** if it is of the form

$$C = \begin{bmatrix} C_1 & -C_2 \\ C_2 & C_1 \end{bmatrix} \tag{10.7.1}$$

with some $p \times p$ matrices C_1, C_2. Let $\mathcal{X} = PD(2p)$ and let $S \in \mathcal{X}$ be partitioned as $((S_{ij}))$ with $S_{ij} : p \times p$, $i, j = 1, 2$. The statistical problem that motivates the choice of the group G that acts on $\mathcal{X}$ is as follows. It is given that $S \sim W(n, \Sigma)$, with some unknown $\Sigma \in PD(2p)$. The problem is to test that Σ has complex structure. This problem is invariant under the group G of nonsingular matrices C of the form (10.7.1), with action

$$S \to CSC', \quad C \in G. \tag{10.7.2}$$

Now we shall drop the assumption that S has a Wishart distribution and only assume that S has a distribution of the form $p(S)(dS)$.

A chart on the whole of G may be taken as the elements of C_1 and C_2. Left Haar measure on G follows from a left invariant differential form of maximum degree ($= 2p^2$). According to Section 5.3 this can be chosen as the wedge product of any $2p^2$ linearly independent elements of the $(2p) \times (2p)$ matrix of (5.3.7), which in the present notation is $C^{-1}dC$. It is obvious that

$$C^{-1} \begin{bmatrix} dC_1 \\ dC_2 \end{bmatrix} \tag{10.7.3}$$

constitutes such a set of linearly independent differentials. Take the wedge product of the elements of (10.7.3) and use (9.1.2), then we obtain

$$\mu_G(dg) = |C|^{-p}(dC_1)(dC_2). \tag{10.7.4}$$

The multiplier χ of the action (10.7.2) is, by (9.1.4),

$$\chi(g) = |C|^{2p+1}. \tag{10.7.5}$$

It follows from Andersson, Brøns, and Jensen (1983) or from Andersson and Perlman (1984) that in Assumption 8.2 we may take $\mathcal{T} = \{\ell = (\ell_1, \dots, \ell_p) : 1 > \ell_1 > \cdots > \ell_p > 0\}$ (after removing from $\mathcal{X}$ a set of Lebesgue measure 0) and $s(t) = \operatorname{diag}(I_p + L, I_p - L) \in \mathcal{X}$, where $L = \operatorname{diag}(\ell_1, \dots, \ell_p)$. Then $\{s(t) : t \in \mathcal{T}\} = \mathcal{Z}$, where $\mathcal{Z} \subset \mathcal{X}$ is the global cross section corresponding to $\mathcal{T}$. The isotropy subgroup G_0 consists of all matrices $C = \operatorname{diag}(E, E)$, with $E = \operatorname{diag}(\pm 1, \dots, \pm 1) : p \times p$. Then G_0 is a finite group with 2^p elements, and by Proposition 7.7.6 we have

$$\mu_{\mathcal{Y}}(dy) = 2^p \mu_G(dg). \tag{10.7.6}$$

Equation (8.7), i.e, $x = gs(t)$, is obtained from (10.7.2) by setting on the right-hand side of (10.7.2) $S = \operatorname{diag}(I_p + L, I_p - L)$. The result is

$$S_{11} = C_1(I + L)C_1' + C_2(I - L)C_2' \tag{10.7.7}$$

$$S_{22} = C_1(I - L)C_1' + C_2(I + L)C_2' \tag{10.7.8}$$

$$S_{12} = C_1(I + L)C_2' - C_2(I - L)C_1' \tag{10.7.9}$$

in which I is short for I_p. These equations are to be differentiated at $C = I_{2p}$, i.e., $C_1 = I_p$, $C_2 = 0$, in order to compute the wedge products of differentials for substitution into equation (8.15). For this purpose it is convenient to introduce

$$T_1 = \frac{1}{2}(S_{11} + S_{22}), \quad T_2 = \frac{1}{2}(S_{11} - S_{22}). \tag{10.7.10}$$

Temporarily write these two equations as $U = \frac{1}{2}(A+B)$, $V = \frac{1}{2}(A-B)$ and compute $(dU)(dV)$. Since A and B, therefore also U and V, are symmetric, we only consider their elements on and below the diagonal. We have, for $i \geq j$, $du_{ij} \wedge dv_{ij} = \frac{1}{4}(da_{ij} + db_{ij}) \wedge (da_{ij} - db_{ij}) = \frac{1}{2}da_{ij} \wedge db_{ij}$ (disregarding the sign). Taking the wedge product over all $i \geq j$ gives $(dU)(dV) = 2^{-\frac{1}{2}p(p+1)}(dA)(dB)$. In the original notation this becomes

$$(dS_{11})(dS_{22}) = 2^{\frac{1}{2}p(p+1)}(dT_1)(dT_2). \tag{10.7.11}$$

Now differentiate (10.7.7) and (10.7.8) at $C_1 = I_p$, $C_2 = 0$:

$$dS_{11} = dC_1(I + L) + dL + (I + L)dC_1' \tag{10.7.12}$$

$$dS_{22} = dC_1(I - L) - dL + (I - L)dC_1'. \tag{10.7.13}$$

Add and subtract (10.7.12) and (10.7.13), then we get $dT_1 = dC_1 + dC_1'$, $dT_2 = dC_1L + LdC_1' + dL$, at $g = e$. Compute the wedge product of the elements of dT_1 and dT_2:

$$(dT_1)(dT_2) = 2^p \prod_{i<j}(\ell_i - \ell_j)(dL)(dC_1), \quad \text{at} \quad g = e. \tag{10.7.14}$$

This is to be substituted for the right-hand side of (10.7.11) to give an expression for $(dS_{11})(dS_{22})$ in terms of $(dL)(dC_1)$. Next, it has to be wedge-multiplied by (dS_{12}) resulting from differentiation of (10.7.9). However, any term in the latter arising from dC_1 or dL will be annihilated by a differential in (dC_1) or (dL) in the expression for $(dS_{11})(dS_{22})$. Hence, we may pretend that

$$dS_{12} = (I + L)dC_2' - dC_2(I - L), \quad \text{at} \quad g = e, \tag{10.7.15}$$

and taking the wedge product over all the elements of dS_{12} produces

$$(dS_{12}) = 2^{\frac{1}{2}p(p+1)} \prod_{i=1}^{p} \ell_i \prod_{i<j}(\ell_i + \ell_j)(dC_2), \quad \text{at} \quad g = e. \tag{10.7.16}$$

Combine (10.7.11), (10.7.14), and (10.7.16):

$$(10.7.17)\quad \begin{aligned}&(dS_{11})(dS_{22})(dS_{12})\\ &\quad = 2^{p(p+2)}\prod_{i=1}^{p}\ell_i\prod_{i<j}(\ell_i^2-\ell_j^2)(dL)(dC_1)(dC_2), \quad \text{at } g=e.\end{aligned}$$

With help of (10.7.4) and (10.7.6) this can be written

$$(10.7.18)\quad \begin{aligned}&(dS_{11})(dS_{22})(dS_{12})\\ &\quad = 2^{p(p+2)}\prod_{i=1}^{p}\ell_i\prod_{i<j}(\ell_i^2-\ell_j^2)(dL)\mu_{\mathcal{Y}}(dy), \quad \text{at } y=[e].\end{aligned}$$

Comparison with (8.15) then shows that

$$(10.7.19)\qquad \mu_{\mathcal{T}}(dt) = 2^{p(p+1)}\prod_{i=1}^{p}\ell_i\prod_{i<j}(\ell_i^2-\ell_j^2)(dL).$$

Substitution of (10.7.19) into the general formula (8.12), taking into account (10.7.4) and (10.7.5), produces the distribution of the maximal invariant L:

$$(10.7.20)\quad \begin{aligned}P(dL) &= 2^{p(p+1)}\prod_{i=1}^{p}\ell_i\prod_{i<j}(\ell_i^2-\ell_j^2)(dL)\\ &\quad\cdot\int p\left(C\begin{bmatrix} I+L & 0\\ 0 & I-L\end{bmatrix}C'\right)|C|^{p+1}(dC_1)(dC_2),\end{aligned}$$

in which C has the form (10.7.1). If one takes $S\sim W(n,I_{2p})$, then (10.7.20) reproduces the distribution described in Andersson, Brøns, and Jensen (1983), Theorem 1, and by Andersson and Perlman (1984), Equation (7.11). If $S\sim W(n,\Sigma)$ with arbitrary $\Sigma\in PD(2p)$, then the method we have employed here, using a global cross section, provides an alternative way compared to Andersson and Perlman (1984), Section 7, for deriving the distribution of L.