The Annals of Statistics
1981, Vol. 9, No. 4, 792-802

A NEW CLASS OF MULTIVARIATE TESTS BASED ON THE UNION-INTERSECTION PRINCIPLE[1]

BY INGRAM OLKIN AND JACK L. TOMSKY

Stanford University and Lockheed Aircraft Corporation

Using Roy's union-intersection principle, a unified treatment is developed for the construction of multivariate tests. These include Wilks' determinantal criteria, Hotelling-Lawley trace criterion, and Roy's largest characteristic root criterion. The key lies in the extension of an index set from vectors to matrices plus the use of elementary symmetric functions of characteristic roots to test component hypotheses.

1. Introduction and background preliminaries. In a variety of multivariate problems, the ultimate step is to choose as a test statistic a particular function of p statistics. Three functions have emerged as candidates: the product (usually the likelihood ratio test, Wilks (1932)), the sum or trace criterion (Lawley (1938), Hotelling (1947)), or the maximum (usually as a result of employing the union-intersection principle, Roy (1953). Each of these choices is reasonable in certain circumstances. However, the fact that one has to employ a different procedure to obtain each of these tests suggests that there may be a single procedure that generates all these tests.

We show how a wide spectrum of test statistics can be obtained via one procedure, namely, the union-intersection principle. The key idea is encompassed in the followng. Consider the class of elementary symmetric functions defined on the set $\{x:x_1 \geq \cdots \geq x_p\}$:

$$\begin{aligned} T_{m,k}(x) &\equiv T_{m,k}(x_1, \cdots, x_p) = E_m(x_1, \cdots, x_k) \\ &= \sum_{x_{i_1}<\cdots<x_{i_m}} x_{i_1}x_{i_2}\cdots x_{i_m}, \qquad m \leq k. \end{aligned} \tag{1.1}$$

That is, $T_{m,k}(x) = E_m(x_1, \cdots, x_k)$ is the mth elementary symmetric function of the k largest x's. Important special cases are

$$T_{1,p}(x) = \textstyle\sum_1^p x_i, \qquad T_{p,p}(x) = \prod_1^p x_i, \qquad T_{1,1}(x) = \max(x_1, \cdots, x_p).$$

In the problems considered the x's are the characteristic roots of a $p \times p$ symmetric matrix, in which case we use the notation

$$\text{tr}_m A \equiv E_m(\lambda_1(A), \cdots, \lambda_p(A)), \tag{1.2}$$

where $\lambda_1(A), \cdots, \lambda_p(A)$ are the characteristic roots of A.

The union-intersection principle involves two steps: (i) designation of component hypotheses, and (ii) choice of tests for testing each of the component hypotheses. It is exactly the flexibility permitted in (i) and (ii) that yields a wide range of tests.

The actual application of the union-intersection principle to multivariate problems generally involves the solution of an extremal problem. These extremal problems are central to obtaining a solution, but are peripheral to the statistical development. For this

Received November, 1979; revised August, 1980.

[1] Work supported in part by the National Science Foundation Grant GP-32326X and by the Lockheed Independent Research Program.

AMS 1970 *subject classifications*. Primary 62H12, 60E15.

Key words and phrases. Union-intersection principle, multivariate tests, characteristic roots, elementary symmetric functions, multivariate linear hypothesis, sphericity test, testing for the equality of covariance matrices, monotonicity of power function.

reason these results are given in Section 4 at the end of the paper. This section also contains notational preliminaries.

Section 2 deals with three standard multivariate (normal) problems (i) the general linear model, (ii) testing that the covariance matrix is the identity, (iii) testing for the equality of two covariance matrices. In each case, $T_{m,k}$ statistics are derived.

We note that the class of tests $T_{m,k}$ provides not only a unification to the construction of tests, but also yields a number of tests previously not considered. However, since no single member of the class $T_{m,k}$ is uniformly better than any other, the deciding factor in the choice of m and k depends on the alternatives of concern. Thus, it is important to compare power functions of the alternative statistics. This issue is outside the scope of the present study.

In Section 3, we show that each of the proposed tests for the general linear model satisfies a "monotonicity" property in the parameters.

The union-intersection principle was introduced by Roy (1953) as a heuristic method of test construction that can be described as follows:

Let $\{\omega_a, \omega_a^*, a \in \Gamma\}$ be a collection of sets in the parameter space. It is assumed that $\cap_{a\in\Gamma}\omega_a$ is nonnull and Γ is an arbitrary index set. Define component hypotheses H_a and alternatives K_a as

$$H_a: \theta \in \omega, \qquad K_a: \theta \in \omega*.$$

Suppose that there exists a test of size α_a in which one accepts H_a over K_a for sample points in the set A_a and rejects H_a otherwise. For simplicity, we assume that $\alpha_a = \alpha*$ is the same for all component tests.

The union-intersection testing problem is constructed from the component testing problems. Set

$$H = \cap_{a\in\Gamma} H_a, \qquad K = \cup_{a\in\Gamma} K_a$$

by which we mean that the null hypothesis H is true if and only if every component null hypothesis H_a is true. Similarly, the alternative hypothesis K is true if and only if at least one component alternative hypothesis is true. Under the union-intersection principle, H is accepted over K if and only if each component test accepts H_a over the corresponding K_a. That is, the acceptance region for a union-intersection test is given by $A = \cap_{a\in\Gamma} A_a$.

In practice, one starts with a null hypothesis H, an alternative hypothesis K, and a significance level α. H and K are then represented as an intersection of component null hypotheses and a union of component alternative hypotheses, respectively. The (common) size α^* of each component test is so determined that the size of A under H is equal to the preassigned α.

Union-intersection tests are generally dependent upon the selection of the representation for H and K. In most applications, the index set Γ is chosen to consist of all nonzero p-dimensional vectors. These union-intersection tests are then constructed from well-known univariate tests.

In the present paper, we extend the index set Γ from vectors to $k \times p$ matrices of rank $k \le p$. The choice of tests of the component hypotheses is based on elementary symmetric functions.

Before presenting the results, we note some related references. Morrison (1976) provides a general discussion of the union-intersection principle. Gabriel (1970) shows that if likelihood ratio tests at a constant level are used for testing the component hypotheses, then the induced union-intersection test (LR-UI) is contained in the likelihood ratio critical region for the global test. He also obtains conditions when the LR-UI test is equivalent to the likelihood test of the overall hypothesis. The use of an index set consisting of matrices was used by Mudholkar, Davidson, and Subbaiah (1974), and by Khatri (1978) who then generate certain classes of tests. Although their class of tests is not as broad as the $T_{m,k}$ class, these papers do contain some of the flavor of the present work.

2. Union-Intersection tests for some standard multivariate problems.

2.1 *General Remarks.* We now obtain union-intersection tests for three standard problems in multivariate analysis. In the usual application of the union-intersection principle, the unions and intersections have been performed over univariate tests. This has led in most cases to a test based on the extreme sample characteristic roots, or as in the case of the stepdown procedure, on a finite number of independent statistics.

The idea here is to perform the unions and intersections over multivariate tests, of lower dimensionality k, which are indexed by a $k \times p$ matrix A. A matrix W_A is said to be a likelihood ratio component matrix if the likelihood ratio test for the component hypothesis is equivalent to rejection if $|W_A| > c$. In general, there are many matrices satisfying this property.

The test criterion for each component hypothesis is to reject the component hypothesis H_A if $\text{tr}_m W_A > c$. The null hypothesis H is rejected if at least one H_A is rejected. Thus, for each k, the union-intersection test is based on the maximum of $\text{tr}_m W_A$ (defined by (1.1 and (1.2)) with respect to the index matrix A. For $m \le k$, the resulting union-intersection test generally depends on the rule for selecting W_A.

By denoting this test statistic as $T_{m,k}$, where $1 \le m \le k \le p$, it can be seen that several standard multivariate tests are included in this class. For example, $T_{p,p}$ coincides with the likelihood ratio test based on determinants, $T_{1,p}$ is a Hotelling-Lawley "trace" criterion, and $T_{1,1}$ is the union-intersection test of Roy based on the extreme sample characteristic roots.

Since we are not directly concerned with distribution problems, the constant c which appears in the rejection regions is a generic constant and may differ for different statistics.

Two sets of matrices continually arise: $C(k, p)$, the set of all $k \times p$ matrices A of rank $k \le p$, and $O(k, p)$, the set of all $k \times p$ matrices A satisfying $AA' = I_k$.

2.2 *Testing the general linear hypothesis.* Let X be a random $p \times n$ matrix with

$$EX' = B\theta, \tag{2.1}$$

where $B{:}n \times m$ is the design matrix of rank $r \le m \le n$ and $\theta{:}m \times p$ is a matrix of unknown parameters. It is assumed that $n > p$ and that the columns of X are independently distributed and have a multivariate normal distribution with unknown covariance matrix Σ.

The hypothesis to be tested is

$$H{:}L\theta = 0 \quad \text{versus} \quad K{:}L\theta \neq 0, \tag{2.2}$$

where $L{:}q \times m$, $q \le m \le r$, is the "hypothesis" matrix.

The component null and alternative hypotheses are chosen as

$$H_A{:}L\theta A' = 0 \quad \text{versus} \quad K_A{:}L\theta A' \neq 0, \qquad A \in \mathscr{C}(k, p).$$

Partition B, θ, and L:

$$B = (B_I, B_D), \qquad \theta' = (\theta_I', \theta_O'), \qquad L = (L_I, L_D),$$

where B_I is $n \times r$, θ_I is $r \times p$, and L_I is $q \times r$.

Next, define the matrices S_H and S_E due to hypothesis and to error, respectively,

$$\begin{aligned} S_H &= XB_I(B_I'B_I)^{-1}L_I'[L_i(B_I'B_I)^{-1}L_I']^{-1}L_I(B_I'B_I)^{-1}B_I'X'/q, \\ S_E &= X[I_n - B_I(B_I'B_I)^{-1}B_I']X'/(n - r). \end{aligned} \tag{2.3}$$

A likelihood ratio component test matrix [see Anderson (1958, page 217)] for testing H_A against K_A is

$$A(S_E + S_H)A'(AS_EA')^{-1} = I_k + (AS_HA')(AS_EA')^{-1} \equiv I_k + Q. \tag{2.4}$$

The component hypothesis H_A is rejected against K_A if

$$\text{tr}_m(I_k + (AS_hA')(AS_EA')^{-1}) = E_m(1 + \lambda_1(Q), \cdots, 1 + \lambda_k(Q)) > 0, \tag{2.5}$$

by definition (1.2).

The null hypothesis H is rejected if and only if $E_m > c$ for some $A \in \mathscr{C}(k,p)$. To find the maximum of E_m, use Theorem 4.8 with $f(x) = 1 + x$, $x_0 = 0$ and $a = \infty$. This maximum is

$$\begin{aligned} &\max_{A\in\mathscr{C}(k,p)} E_m(1 + \lambda_1(Q), \cdots, 1 + \lambda_k(Q)) \\ &\qquad = E_m(1 + \lambda_1(S_HS_E^{-1}), \cdots, 1 + \lambda_k(S_HS_E^{-1})). \end{aligned} \tag{2.6}$$

The resulting test for H against K is to reject H if

$$T_{m,k} = E_m[1 + \lambda_1(S_HS_E^{-1}), \cdots, 1 + \lambda_k(S_HS_E^{-1})] > c. \tag{2.7}$$

In Section 3.1, we show that each of the tests based on $T_{m,k}$ satisfy a monotonicity property; that is, the power function is an increasing function of each characteristic root of the matrix $\mu'\Sigma^{-1}\mu$, where $\mu = (\mu_1, \mu_2, \cdots, \mu_r)$.

Various combinations of m and k yield well-known standard multivariate tests. The choice $m = k = p$ yields

$$T_{p,p} = \textstyle\prod_{j=1}^{p} (1 + \lambda_j(S_HS_E^{-1})) = | I_p + S_HS_E^{-1} | = | S_E + S_H |/| S_E |, \tag{2.8}$$

which is equivalent to the likelihood ratio statistic. The choice $m = 1$, $k = p$ yields

$$T_{1,p} = \textstyle\sum_{j=1}^{p} (1 + \lambda_j(S_HS_E^{-1})) = \text{tr}(I_p + S_HS_E^{-1}) = p + \text{tr}(S_HS_E^{-1}),$$

which is equivalent to the Hotelling-Lawley "trace" statistic. The choice $m = k = 1$ yields

$$T_{1,1} = 1 + \lambda_1(S_HS_E^{-1}),$$

which is equivalent to the Roy "maximum root" statistic.

REMARK. The case $m = 1$ and $B' = (1, \cdots, 1)$ yields Hotelling's T^2 statistic. This model is discussed in Morrison (1976) with the choice $A \in \mathscr{C}(1, p)$. The extension to the larger class $A \in \mathscr{C}(k, p)$ does not yield a new test.

2.3 *Testing that the covariance matrix is the identity matrix.* Suppose that $S > 0$ has a Wishart distribution, $W(\Sigma, p, n)$ and we wish to test

$$H: \Sigma = I_p \quad \text{versus} \quad K: \Sigma \neq I_p.$$

Define the component null and alternative hypotheses H_A and K_A by

$$H_A: A\Sigma A' = I_k, \qquad K_A: A\Sigma A' \neq I_k, \qquad A \in O(k, p), \tag{2.9}$$

A likelihood ratio component test matrix [see Anderson (1958, page 265)] is $(ASA')^{-1}\exp(ASA')$, so that the component test for H_A against K_A is to reject H_A if

$$\text{tr}_m((ASA')^{-1}\exp(ASA')) > c. \tag{2.10}$$

The corresponding union-intersection test is to reject H against K if

$$\max_{A\in O(k,p)} \text{tr}_m((ASA')^{-1}\exp(ASA')) > c. \tag{2.11}$$

By Corollary 4.9, (2.11) is equivalent to

$$T_{m,k} = E_m(\lambda_{[1]}^{-1}e^{\lambda_{[1]}}, \cdots \lambda_{[k]}^{-1}e^{\lambda_{[k]}}), > c, \tag{2.12}$$

where $\lambda_{[j]}$ are the characteristic roots of S arranged according to decreasing values of $f(\lambda) = \lambda^{-1}\exp(\lambda)$.

The special choice $m = k = p$ yields

$$T_{p,p} = \textstyle\prod_{j=1}^{p} \lambda_j^{-1}e^{\lambda_j} = (\exp \text{tr } S)/| S |,$$

which is equivalent to the likelihood ratio statistic. A Hotelling-Lawley trace statistic is given by the choice $m = 1$, $k = p$, namely,

$$T_{1,p} = \sum_{j=1}^{p} \lambda_j^{-1} e^{\lambda_j} = \operatorname{tr}(S^{-1} \exp S).$$

The choice $m = k = 1$ yields

$$T_{1,1} = \lambda_{[1]}^{-1} e^{\lambda_{[1]}},$$

which is identical to a test based on λ_1 and/or λ_p. This test differs slightly from Roy's test, which is based on "equal tails" for λ_1 and λ_p.

2.4 *Testing for the equality of two covariance matrices.* Suppose that S_1 and S_2 have Wishart distributions $W(\Sigma_1, p, n_1)$ and $W(\Sigma_2, p, n_2)$, respectively, and we wish to test

$$H:\Sigma_1 = \Sigma_2 \quad \text{versus} \quad K:\Sigma_1 \neq \Sigma_2.$$

Define the component null hypotheses H_A and the component alternative hypotheses K_A by

$$H_A: A\Sigma_1 A' = A\Sigma_2 A', \qquad K_A: A\Sigma_1 A' \neq A\Sigma_2 A', \qquad A \in \mathscr{C}(k, p). \tag{2.13}$$

A test matrix for the likelihood ratio component test [see Anderson (1958), page 256)] is

$$(AS_1A')^{-n_1} (A(S_1 + S_2)A')^{n_1+n_2} (AS_2A')^{-n_2},$$

so that the test for H_A against K_A is to reject H_A if

$$\operatorname{tr}_m[(AS_1A')^{-n_1}(A(S_1 + S_2)A')^{n_1+n_2}(AS_2A')^{-n_2}] > c. \tag{2.14}$$

The union-intersection test is to reject H against K if

$$\max_{A\in\mathscr{C}(k,p)} \operatorname{tr}_m[(AS_1A')^{-n_1}(A(S_1 + S_2)A')^{n_1+n_2}(AS_2A')^{-n_2}] > c. \tag{2.15}$$

which by Corollary 4.10 is equivalent to

$$T_{m,k} = E_m(\lambda_{[1]}^{-n_1}(1 - \lambda_{[1]})^{-n_2}, \cdots, \lambda_{[k]}^{-n_1}(1 - \lambda_{[k]})^{-n_2}) > c, \tag{2.16}$$

where $\lambda_{[j]}$ are characteristic roots of $S_1(S_1 + S_2)^{-1}$ arranged according to decreasing values of $f(\theta) = \theta^{-n_1}(1 - \theta)^{-n_2}$.

The choice $m = k = p$ yields

$$T_{p,p} = \prod_{j=1}^{p} \lambda_j^{-n_1}(1 - \lambda_j)^{-n_2} = |S_1 + S_2|^{n_1+n_2}/(|S_1|^{n_1}|S_2|^{n_2})$$

which is the likelihood ratio test. The choice $m = 1$, $k = p$ yields

$$T_{1,p} = \sum_{j=1}^{p} \lambda_j^{-n_1}(1 - \lambda_j)^{-n_2} = \operatorname{tr}[(I_p + S_2S_1^{-1})^{n_1}(I_p + S_1S_2^{-1})^{n_2}],$$

which is similar to a Hotelling-Lawley trace test. The choice $m = k = 1$ yields

$$T_{1,1} = \lambda_{[1]}^{-n_1}(1 - \lambda_{[1]})^{-n_2},$$

which is equivalent to a test based on λ_1 and/or λ_p.

3. Monotonicity of the power functions of some proposed tests for the general linear model. In many testing problems in multivariate analysis, tests that are invariant under a group of transformations depend upon the sample characteristic roots. The power of such tests is a function of the population characteristic roots, which can be thought of as noncentrality parameters. A test is said to have the monotonicity property if its power function is a monotonically increasing function of each population characteristic root. We now show that the class of tests proposed for the general linear model problem satisfies the monotonicity property.

The canonical form for the general linear model is given by Roy (1957). Let U be a

$p \times q$ matrix and V be a $p \times n - r$ matrix, where p is the number of variates, q is the degrees of freedom for the hypothesis, and $n - r$ is the degrees of freedom for the error. The joint density function $f(U, V)$ of U and V is proportional to

$$f(U, V) \propto \exp[-\tfrac{1}{2}\{\operatorname{tr} VV' + \textstyle\sum_{i=1}^{t} (u_{ii} - \theta_i)^2 + \sum_{i=t+1}^{p} u_{ii} + \sum_{i=1}^{p} \sum_{j\neq i=1}^{s} u_{ij}^2], \tag{3.1}$$

where $\theta_1 \geq \cdots \geq \theta_t$. The hypothesis to be tested is $H{:}\theta_1 = \cdots = \theta_t = 0$, against the alternative $K{:}\theta_1 > 0$.

A sufficient condition for a test to satisfy the monotonicity property (for testing H against K) has been given by Anderson, Das Gupta, and Mudholkar (1964) and is stated below.

THEOREM 3.1 *For each i ($i = 1, \cdots, q$) and for each set of fixed values of u_j's ($j \neq i$) and V, suppose there exists an orthogonal transformation: $u_i \to Mu_i = (u_{1i}^*, \cdots, u_{pi}^*)'$ such that the region $\omega_i(u_i)$, a section of the acceptance region ω in the space of u_i for a set of fixed values of u_j's ($j \neq i$) and V, is transformed into the region $\omega_i^*(u_i^*)$ which has the following property: any section of $\omega_i^*(u_i^*)$ for fixed values of u_{ki}^* ($k \neq j$) is an interval, symmetric about $u_{ji}^* = 0$. Then the power function of the test, having acceptance region ω, monotonically increases in each θ_i.*

We use Theorem 3.1 to prove that the tests for the general linear model based on the $T_{m,k}$ criterion introduced in Section 2.2 satisfy the monotonicity property.

THEOREM 3.2. *In testing the general linear hypothesis, the test with acceptance region*

$$T_{m,k} = E_m(1 + \lambda_1, \cdots, 1 + \lambda_k) \leq c \tag{3.2}$$

has a power function monotonically increasing in each θ_i. (The λ_j are the k largest characteristic roots of $UU'(VV')^{-1}$.)

PROOF. The idea in this proof is to show that Theorem 3.1 is satisfied for $M = I_p$. Let V and all columns of U except u_i be fixed. (Without loss of generality, we can assume $i = 1$.)

Next partition U into (u_1, U_2), where u_1 is the first column of U. Then

$$I_p + UU'(VV')^{-1} = I_p + U_2U_2'(VV') + u_1u_1'(VV')^{-1} \equiv G. \tag{3.3}$$

Let

$$\mathscr{U} = (u_1 : E_m(\lambda_1(G), \cdots, \lambda_k(G)) \leq c; \qquad U_2 \text{ and } V \text{ fixed}). \tag{3.4}$$

The conditions of Theorem 3.1 are satisfied if we can show that for $u_1 \in \mathscr{U}$, it follows that $\alpha u_1 \in \mathscr{U}$ for $-1 \leq \alpha \leq 1$. Let $G(\alpha)$ be formed from G by replacing u_1 with αu_1, that is,

$$\begin{aligned} G(\alpha) &= I_p + U_2U_2'(VV')^{-1} + \alpha^2 u_1u_1'(VV')^{-1} \\ &= (VV' + U_2U_2' + \alpha u_1u_1')(VV')^{-1}. \end{aligned} \tag{3.5}$$

Since $G(\alpha)$ is the product of two positive-definite matrices, the characteristic roots of $G(\alpha)$ are positive. [See e.g., Bellman (1960, page 134)]. For $\alpha^2 \leq 1$,

$$G - G(\alpha) = G(1) - G(\alpha) = (1 - \alpha^2)u_1u_1'(VV')^{-1}. \tag{3.6}$$

Since the characteristic roots of $G - G(\alpha)$ are nonnegative for $\alpha^2 \leq 1$, it follows that

$$0 \leq \lambda_j(G(\alpha)) \leq \lambda_j(G), \qquad j = 1, \cdots, k.$$

Therefore, since E_m is a nondecreasing function in each of its arguments when they are all nonnegative,

$$E_m(\lambda_1(G(\alpha)), \cdots, \lambda_k(G(\alpha)) \leq E_m(\lambda_1(G), \cdots, \lambda_k(G)).$$

Consequently, $\alpha u_1 \in \mathscr{U}$ for $\alpha^2 \le 1$, and thus the region $\mathscr{U}$ satisfies Theorem 3.1 as was to be shown. □

4. Some results on matrix extremal problems.

4.1 *Preliminaries.* This section contains results on extremal problems that are needed in the derivation of the union-intersection tests. Recall that $\mathscr{C}(k, p) = \{A : k \times p, \text{rank } A = k\}$, $O(k, p) = \{A : k \times p, AA' = I_k\}$. In particular, $O(1, p)$ is the class of unit vectors and $O(p, p)$ is the class of orthogonal $p \times p$ matrices. Because we deal with real matrices, the results are stated for real matrices; however, with the obvious changes, the results can be readily extended to apply to complex matrices.

By $A \ge B$ $(A > B)$ we mean that for symmetric matrices A and B the difference $A - B$ is positive semidefinite (positive definite). The characteristic roots of an $n \times n$ matrix A are denoted by $\lambda(A)$, which when real, are ordered by

$$\lambda_1(A) \ge \lambda_2(A) \ge \cdots \ge \lambda_n(A).$$

4.2 *Extremal properties for a positive semi-definite hermitian matrix.* Our starting point is a result known as the Poincaré Separation Theorem. The special case $k = 1$ was obtained by Cauchy in 1829, in which case the result is also called the interlacing property.

THEOREM 4.1 (*Poincaré* (1890)). *If* $S : p \times p$, $S \ge 0$, *and* $A \in O(k, p)$ *then*

$$\text{(i) } \lambda_j(ASA') \le \lambda_j(S), \tag{4.1}$$

$$\text{(ii) } \lambda_{k-j+1}(ASA') \ge \lambda_{p-j+1}(S), \qquad j = 1, \cdots, k.$$

From Theorem 4.1, we deduce the important corollary.

COROLLARY 4.2. *If* $S : p \times p$, $S \ge 0$, *then*

(1a) $\min_{O(k,p)} \lambda_1(ASA') = \lambda_{p-k+1}(S)$, (1b) $\max_{O(k,p)} \lambda_1(ASA') = \lambda_1(S)$,

(2a) $\min_{O(k,p)} \lambda_k(ASA') = \lambda_p(S)$, (2b) $\max_{O(k,p)} \lambda_k(ASA') = \lambda_k(S)$.

PROOF. The choice $j = 1$ and $j = k$ in (4.1) yields $\lambda_{p-k+1}(S) \le \lambda_1(ASA') \le \lambda_1(S)$, $\lambda_p(S) \le \lambda_k(ASA') \le \lambda_k(S)$, for which equality is achieved by the choices $A = (I_k, 0)\Gamma'$ and $A = (0, I_k)\Gamma'$, respectively, where Γ is an orthogonal matrix satisfying $S = \Gamma D_\lambda \Gamma'$, $D_\lambda \equiv \text{diag}(\lambda_1, \cdots, \lambda_p)$. □

If we take the sum and the product over j in (4.1) we obtain

$$\textstyle\sum_{j=1}^k \lambda_{p-j+1}(S) \le \sum_{j=1}^k \lambda_{k-j+1}(ASA') = \text{tr } ASA' \le \sum_{j=1}^k \lambda_j(S),$$

$$\textstyle\prod_{j=1}^k \lambda_{p-j+1}(S) \le \prod_{j=1}^k \lambda_{k-j+1}(ASA') = |ASA'| \le \prod_{j=1}^k \lambda_j(S).$$

Equality is achieved by choosing A as in Corollary 4.2, which yields the following.

COROLLARY 4.3. *If* $S : p \times p$, $S \ge 0$, *then*

(1a) $\max_{O(k,p)} \text{tr } ASA' = \sum_{j=1}^k \lambda_j(S)$, (1b) $\min_{O(k,p)} \text{tr } ASA' = \sum_{j=1}^k \lambda_{p-j+1}(S)$,

(2a) $\max_{O(k,p)} |ASA'| = \prod_{j=1}^k \lambda_j(S)$, (2b) $\min_{O(k,p)} |ASA'| = \prod_{j=1}^k \lambda_{p-j+1}(S)$.

Part (1a) follows from von Neumann (1937) and Fan (1951); (1b) is given by Fan (1951); (2a) and (2b) are due to Fan (1949). The derivations differ from the ones presented here.

DEFINITION. Let A be a $p \times q$ matrix and let $m \le \min(p, q)$. The mth compound of a matrix A, denoted by $A^{(m)}$ is a matrix of dimension $\binom{p}{m}$ by $\binom{q}{m}$ obtained by forming all minors of A of order m and arranging them in lexicographic order.

Properties concerning the compound are discussed in Aitken (1956) and Karlin (1968). In particular $(A^{(m)})' = (A')^{(m)}$, and $(A^{(m)})^{-1} = (A^{-1})^{(m)}$. A central property of compound matrices is the Binet-Cauchy Theorem.

THEOREM 4.4. $(AB)^{(m)} = A^{(m)}B^{(m)}$.

As a consequence of Theorem 4.4, if $A:k \times p$ is row-orthogonal, i.e., $AA' = I_k$, then $A^{(m)}$ is row-orthogonal.

A second key property is that if $A:p \times p$ has characteristic roots $\lambda_1, \cdots, \lambda_p$, then the characteristic roots of $A^{(m)}$ are the $\binom{p}{m}$ products of the form $\lambda_{i_1}\lambda_{i_2} \cdots \lambda_{i_m}$. As a consequence, an extremal property involving $\operatorname{tr} A = \Sigma\lambda_i = E_1(\lambda_1, \cdots, \lambda_p)$ can be extended to $\operatorname{tr} A^{(m)} = E_m(\lambda_1, \cdots, \lambda_p)$. The notation $\operatorname{tr}_m A$ is generally used for $\operatorname{tr} A^{(m)}$.

THEOREM 4.5. *If S is a $p \times p$ matrix, $S \geq 0$, then for $m = 1, \cdots, k$,*

$$(1)\ \max_{O(k,p)} \operatorname{tr}_m(ASA') = E_m(\lambda_1(S), \cdots, \lambda_k(S)),$$

$$(2)\ \min_{O(k,p)} \operatorname{tr}_m(ASA') = E_m(\lambda_{p-k+1}(S), \cdots, \lambda_p(S)).$$

PROOF. As a consequence of the Binet-Cauchy Theorem,

$$\operatorname{tr}_m(ASA') = \operatorname{tr}(ASA')^{(m)} = \operatorname{tr} A^{(m)}S^{(m)}A'^{(m)}.$$

Since $A:k \times p$ is row orthogonal, $B \equiv A^{(m)}: \binom{k}{m} \times \binom{p}{m}$ is row orthogonal. Thus

$$\begin{aligned}
\max_{A\in O(k,p)} \operatorname{tr} A^{(m)}S^{(m)}A^{(m)\prime} &\leq \max_{B\in O(\binom{k}{m},\binom{p}{m})} \operatorname{tr} BS^{(m)}B' \\
&= \textstyle\sum_1^{\binom{k}{m}} \lambda_j(S^{(m)}) = E_m(\lambda_1(S), \cdots, \lambda_k(S)), \\
\min_{A\in O(k,p)} \operatorname{tr} A^{(m)}S^{(m)}A^{(m)} &\geq \min_{B\in O(\binom{k}{m},\binom{p}{m})} \operatorname{tr} BS^{(m)}B' \\
&= \textstyle\sum_1^{\binom{k}{m}} \lambda_{p-j+1}(S^{(m)}) = E_m(\lambda_{p-k+1}(S), \cdots, \lambda_p(S)).
\end{aligned}$$

The results follows from the fact that equality can be achieved. □

Berkowitz (1974) obtains a result for elementary symmetric functions of the form $\operatorname{tr}_m(AUA' + V)$, where $U:p \times p$, $V:k \times k$, $U \geq 0$, $V \geq 0$ and $A \in O(k,p)$, namely,

$$\max_{O(k,p)} \operatorname{tr}_m(AUA' + V) = E_m(\nu_1 + \mu_k, \cdots, \nu_k + \mu_1),$$

$$\min_{O(k,p)} \operatorname{tr}_m(AUA' + V) = E_m(\nu_1 + \mu_{p-k+1}, \cdots, \nu_k + \mu_p),$$

where $\nu = \lambda(V)$ and $\mu = \lambda(U)$. The case $V = 0$ reduces to Theorem 4.5.

4.3 *Extremal properties for two positive-definite Hermitian matrices.* The results presented thus far concern a single positive semi-definite matrix S. Since some applications involve two matrices, we require extensions of the previous results. Given two $p \times p$ matrices S_1 and S_2 with $S_1 \geq 0$, $S_2 > 0$ we need to study the characteristic roots of the matrix $(BS_1B')(BS_2B')^{-1}$, where $B \in \mathscr{C}(k,p)$.

The matrices S_1 and S_2 can be simultaneously diagonalized by a nonsingular $p \times p$ matrix W:

$$S_1 = WD_\theta W', \qquad S_2 = WW',$$

where $D_\theta = \operatorname{diag}(\theta_1, \cdots, \theta_p)$, and $\theta_j \equiv \lambda_j(S_1S_2^{-1})$ are the ordered characteristic roots of $S_1S_2^{-1}$. Then

$$\begin{aligned}
\max_{B\in\mathscr{C}(k,p)} \lambda_j((BS_1B')(BS_2B')^{-1} &= \max_{B\in\mathscr{C}(k,p)} \lambda_j((BWD_\theta W'B')(BWW'B')^{-1}) \\
&= \max_{G\in\mathscr{C}(k,p)} \lambda_j(GD_\theta G')(GG')^{-1} \\
&= \max_{G\in\mathscr{C}(k,p)} \lambda_j((GG')^{-1/2}GD_\theta G'(GG')^{-1/2}) \\
&= \max_{H\in O(k,p)} \lambda_j(HD_\theta H') = \theta_j.
\end{aligned}$$

The fact that equality can be achieved by $B = (I_k 0) W^{-1}$ yields an equality. This argument can be extended to provide a number of variations.

THEOREM 4.6. *Suppose $S_1 \geq 0$ and $S_2 > 0$. For $A \in \mathscr{C}(k, p)$, $k \leq p$, the following results hold. (The maximizations are over the set $\mathscr{C}(k, p)$.)*

(1a) $\lambda_j((AS_1A')(AS_2A')^{-1}) \leq \lambda_j(S_1S_2^{-1}), \quad (j = 1, \cdots, k)$

(1b) $\lambda_{k-j+1}((AS_1A')(AS_2A')^{-1}) \geq \lambda_{p-j+1}(S_1S_2^{-1}), \quad (j = 1, \cdots, k)$

(2a) $\min_A \lambda_1((AS_1A')(AS_2A')^{-1}) = \lambda_{p-k+1}(S_1S_2^{-1})$,

(2b) $\max_A \lambda_1((AS_1A')(AS_2A')^{-1}) = \lambda_1(S_1S_2^{-1})$,

(3a) $\min_A \lambda_k((AS_1A')(AS_2A')^{-1}) = \lambda_p(S_1S_2^{-1})$,

(3b) $\max_A \lambda_k((AS_1A')(AS_2A')^{-1}) = \lambda_k(S_1S_2^{-1})$,

(4a) $\max_A \mathrm{tr}_m((AS_1A')(AS_2A')^{-1}) = E_m(\lambda_1(S_1S_2^{-1}), \cdots, \lambda_k(S_1S_2^{-1}))$,

(4b) $\min_A \mathrm{tr}_m((AS_1A')(AS_2A')^{-1}) = E_m(\lambda_{p-k+1}(S_1S_2^{-1}), \cdots, \lambda_p(S_1S_2^{-1}))$.

In constructing union-intersection tests, we are led to maximizing elementary symmetry functions with respect to a matrix A. The following theorem is basic in solving this problem.

THEOREM 4.7. *Let $f(x)$ be a nonnegative analytic function of a real variable x over the interval $0 \leq x \leq a$. Suppose that for some $x_0 \in (0, a)$, the function $f(x)$ is decreasing for $0 \leq x \leq x_0$ and increasing for $x_0 \leq x \leq a$. Let $S \geq 0$ have characteristic roots $\theta_1 \geq \theta_2 \geq \cdots \geq \theta_p$. Let $\theta_{[1]}, \theta_{[2]}, \cdots, \theta_{[p]}$ denote the θ's arranged in decreasing order according to the values of $f(\theta)$; that is, $f(\theta_{[1]}) \geq f(\theta_{[2]}) \geq \cdots \geq f(\theta_{[p]})$. Then for $m = 1, 2, \cdots, k$,*

$$\max_{O(k,p)} E_m[f(\lambda_1(ASA')), \cdots, f(\lambda_k(ASA'))] = E_m[f(\theta_{[1]}), \cdots, f(\theta_{[k]})]. \tag{4.2}$$

PROOF. We first show that for any $A \in O(k, p)$,

$$E_m[f(\lambda_1(ASA')), \cdots, f(\lambda_k(ASA'))] \leq E_m[f(\theta_{[1]}), \cdots, f(\theta_{[k]})]. \tag{4.3}$$

Consider first $k = 1$. For any $A \in O(1, p)$, A is a row vector of unit length. The single characteristic root $\lambda_1(ASA')$ is equal to the scalar ASA'. By the characterization of θ_1 and θ_p as the extreme values of $x'Sx$ when $x'x = 1$, we have $\theta_p \leq \lambda(ASA') \leq \theta_1$. Because of the assumptions on the function $f(x)$, the maximum in every interval occurs at one of its end points. Thus, $f(ASA')$ cannot be greater than both $f(\theta_1)$ and $f(\theta_p)$, and hence

$$\max_{\mathscr{A}(k,p)} f(\lambda(ASA')) \leq f(\theta_{[1]}). \tag{4.4}$$

When $k = 1$, we must have $m = 1$ and since $E_1(x) = x$, (4.4) holds for $k = 1$.

Consider now the case for general k. Let $\theta_{[1]}, \theta_{[2]}, \cdots, \theta_{[k]}$ be the k θ's which have the largest values of $f(\theta)$. In terms of the original ordering of θ, these form sequences at either one or both ends of the θ's. Let us denote these θ's by $\{\theta_1, \cdots, \theta_m, \theta_{p-k+m+1}, \cdots, \theta_p\}$, admitting the possibility that $m = 0$ or k. By the characterization of the function f, we must have $\theta_m \leq x_0 \leq \theta_{p-k+m+1}$, where x_0 minimizes f. By Theorem 4.1,

$$\begin{aligned} \lambda_j(ASA') &\leq \theta_j, & j &= 1, \cdots, m, \\ \lambda_{k-j+1}(ASA') &\geq \theta_{p-j+1}, & j &= 1, \cdots, k-m. \end{aligned} \tag{4.5}$$

Because of the assumptions on f,

$$\begin{aligned} f(\lambda_j(ASA')) &\leq f(\theta_j), & j &= 1, \cdots, m, \\ f(\lambda_{k-j+1}(ASA')) &\leq f(\theta_{p-j+1}), & j &= 1, \cdots, k-m. \end{aligned} \tag{4.6}$$

Now $E_m(x_1, \cdots, x_k)$ is a nondecreasing function of each argument when all $x_i \geq 0$.

Therefore,

$$(4.7)\quad E_m[f(\lambda_1(ASA')), \cdots, f(\lambda_k(ASA'))] \le E_m[f(\theta_1), \cdots, f(\theta_m), f(\theta_{p-k+m+1}), \cdots, f(\theta_p)]$$
$$= E_m[f(\theta_{[1]}), \cdots, f(\theta_{[k]})].$$

All that remains to be shown is that this upper bound can be attained. Let S be factored according to $S = \Gamma D_{[\theta]}\Gamma'$, where Γ is a $p \times p$ orthogonal matrix and $D_{[\theta]} = \mathrm{diag}(\theta_{[1]}, \cdots, \theta_{[p]})$. Then the upper bound is attained for $A = (I_k, 0)\Gamma'$. □

If x_0 is equal to 0 or a then f is monotone. From the discussion leading to Theorem 4.5, we can generalize Theorem 4.7 to the case of two matrices.

THEOREM 4.8. *Let $f(x)$ be a function of a real variable x satisfying the conditions specified in Theorem 4.7. If $S_1 \ge 0$ and $S_2 > 0$, let $\theta_1 \ge \theta_2 \ge \cdots \ge \theta_p$ be the ordered characteristic roots of $S_1S_2^{-1}$. Let $\theta_{[1]}, \theta_{[2]}, \cdots, \theta_{[p]}$ denote the ordering of the θ's according to $f(\theta)$. That is, $f(\theta_{[1]}) \ge f(\theta_{[2]}) \ge \cdots \ge f(\theta_{[p]})$. Then for $m = 1, \cdots, k$,*

$$(4.8)\quad \max_{O(k,p)} E_m[f(\lambda_1((AS_1A')(AS_2A')^{-1})), \cdots, f(\lambda_k((AS_1A')(AS_2A')^{-1}))]$$
$$= E_m[f(\theta_{[1]}), \cdots, f(\theta_{[k]})].$$

4.4 *Matrix functions.* Suppose S_1 is an $n \times n$ symmetric matrix that can be factored according to $S = \Gamma D_\lambda \Gamma'$, where Γ is orthogonal and $D_\lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$. If g is an analytic function such that $g(\lambda_i)$ is defined for each i, then we define $g(S)$ as the symmetric matrix

$$g(S) = \Gamma D_{g(\lambda)}\Gamma',$$

where $D_{g(\lambda)} = \mathrm{diag}[g(\lambda_1), \cdots, g(\lambda_n)]$.

For example, if $g(x) = \exp(x)$ and $S = \Gamma D_\lambda \Gamma'$, then $\exp(S) = \Gamma\ \mathrm{diag}(\exp \lambda_1, \cdots, \exp \lambda_n)\Gamma'$. If λ is a characteristic root of S, then $g(\lambda)$ is a characteristic root of $g(S)$; the characteristic vectors of S and $g(S)$ are identical.

COROLLARY 4.9. *If $S: p \times p$, $S > 0$, then for $m = 1, \cdots, k$,*

$$(4.9)\quad \max_{A \in O(k,p)} \mathrm{tr}_m[(ASA')^{-1}\exp(ASA')] = E_m(\theta_{[1]}^{-1}e^{\theta_{[1]}}, \cdots, \theta_{[k]}^{-1}e^{\theta_{[k]}}),$$

where $\theta_{[j]}$ are the characteristic roots of S arranged according to decreasing values of $f(\theta) = \theta^{-1}\exp(\theta)$.

PROOF. If the characteristic roots of ASA'^{-1} are denoted by $\lambda = \lambda(ASA')$, then the characteristic roots of $(ASA')^{-1}\exp(ASA')$ are $\lambda^{-1}\exp(\lambda)$. Thus,

$$\mathrm{tr}_m[(ASA')^{-1}\exp(ASA')] = E_m(\lambda_1^{-1}e^{\lambda_1}, \cdots, \lambda_k^{-1}e^{\lambda_k}).$$

Let $f(x) = x^{-1}e^x$. Since $f'(x) = x^{-2}(x-1)e^x$, the conditions of Theorem 4.7 on $f(x)$ are satisfied with $a = \infty$ and $x_0 = 1$. □

COROLLARY 4.10. *If $S_1 > 0$ and $S_2 > 0$, then for $m = 1, \cdots, k$,*

$$\max_{A \in O(k,p)} \mathrm{tr}_m[(AS_1A')^{-m_1}(A(S_1+S_2)A')^{n_1+n_2}(AS_2A')^{-n_1}]$$
$$= E_m(\theta_{[1]}^{-n_1}(1-\theta_{[1]})^{-n_2}, \cdots, \theta_{[k]}^{-n_1}(1-\theta_{[k]})^{-n_2}),$$

where $\theta_{[j]}$ are the characteristic roots of $S_1(S_1+S_2)^{-1}$ arranged according to decreasing values of $f(\theta) = \theta^{-n_1}(1-\theta)^{-n_2}$.

PROOF. For any $A \in O(k,p)$, we first apply a simultaneous factorization to AS_1A' and $A(S_1+S_2)A'$.

$$(4.10)\quad A(S_1+S_2)A' = WW', \qquad AS_1A' = WD_\lambda W', \qquad AS_2A' = W(I_k - D_\lambda)W'.$$

Here, W is a $k \times k$ nonsingular matrix and D_λ is a diagonal matrix whose diagonal elements are the k characteristic roots of $(AS_1A')[A(S_1 + S_2)A']^{-1}$.

$$(4.11) \quad (AS_1A')^{-n_1}(A(S_1 + S_2)A')^{n_1+n_2}(AS_2A')^{-n_2} = (WD_\lambda W')^{-n_1}(WW')^{n_1+n_2}(W(I_k - D_\lambda)W')^{-n_2}.$$

We next make use of the fact that the characteristic roots of the product of square matrices are invariant under commutation; that is, $\lambda(UV) = \lambda(VU)$. Applying this fact repeatedly, the characteristic roots of $(WD_\lambda W')^{-n_1}(WW')^{n_1+n_2}[W(I_k - D_\lambda)W']^{-n_2}$ are the same as the characteristic roots of $D_\lambda^{-n_1}(I_k - D_\lambda)^{-n_2}$. These characteristic roots are equal to $\lambda^{-n_1}(1 - \lambda)^{-n_2}$. Therefore,

$$(4.12) \quad \operatorname{tr}_m[(AS_1A')^{-n_1}(A(S_1 + S_2)A')^{n_1+n_2}(AS_2A')^{-n_2}] = E_m[\lambda_1^{-n_1}(1 - \lambda_1)^{-n_2}, \cdots, \lambda_k^{-n_k}(1 - \lambda_k)^{-n_2}].$$

Let $f(x) = x^{-n_1}(1 - x)^{-n_2}$. Since

$$f'(x) = x^{-(n_1+1)}(1 - x)^{-(n_2+1)}[(n_1 + n_2)x - n_1],$$

the conditions of Theorem 4.8 on $f(x)$ are satisfied with $a = 1$ and $x_0 = n_1/(n_1 + n_2)$. □

REFERENCES

AITKEN, A. C. (1956). *Determinants and Matrices*, 9th ed. Oliver and Boyd, Edinburgh.

ANDERSON, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.

ANDERSON, T. W., DAS GUPTA, S., and MUDHOLKAR, G. S. (1964). Monotonicity of the power functions of some tests of the multivariate linear hypothesis. *Ann. Math. Statist.* **35** 200–205.

BELLMAN, R. (1960). *Introduction to Matrix Analysis*. McGraw-Hill, New York.

BERKOWITZ, S. (1974). Extrema of elementary symmetric polynomials of the eigenvalues of the matrix $P^*KP + L$. *Linear Algebra Appl.* **8** 273–280.

FAN, K. (1949). On a theorem of Weyl concerning the eigenvalues of linear transformations I. *Proc. Nat. Acad. Sci.* U.S.A. **35** 652–655.

FAN, K. (1951). Maximum properties and inequalities for the eigenvalues of completely continuous operators. *Proc. Nat. Acad. Sci.* U.S.A. **37** 760–766.

GABRIEL, K. R. (1970). On the relation between union intersection and likelihood ratio tests. In *Essays in Probability and Statistics*. ed, by R. C. Bose, I. M. Chakravarti, P. C. Mahalauobis, C. R. Rao, and K. J. C. Smith, 251–266. Univ. North Carolina Press, Chapel Hill.

HOTELLING, H. (1947). Multivariate quality control, illustrated by the air testing of sample bombsights. In *Techniques of Statistical Analysis*. ed, by C. Eisenhart, M. W. Kastay, W. A. Wallis 111–184. McGraw-Hill, New York.

KARLIN, S. (1968). *Total Positivity*. Stanford Univ. Press. Stanford, Calif.

KHATRI, C. G. (1978). Some optimization problems with application to canonical correlations and sphericity tests. *J. Multivariate Anal.* **8** 453–467.

LAWLEY, D. N. (1938). A generalization of Fisher's z test. *Biometrika* **30** 180–187.

MORRISON, D. (1976). *Multivariate Statistical Methods*, 2nd ed. McGraw-Hill, New York.

MUDHOLKAR, G. S., DAVIDSON, M. S., and SUBBAIAH, P. (1974). A note on the union-intersection character of some MANOVA procedures. *J. Multivariate Anal.* **4** 486–493.

VON NEUMANN, J. (1937). Some matrix-inequalities and metrization of metric-space. *Tomsk Univ. Rev.* **1** 286–296. (John von Neumann Collected Works (A. M. Taub, ed.) 205–218. Pergamon, Oxford, 1962.)

POINCARÉ, H. (1890). Sur les equations aux derivées partielles de la physique mathematique. *Amer. Journ. Math.* **12** 211–294.

ROY, S. N. (1953). On a heuristic method of test construction and its use in multivariate analysis. *Ann. Math. Statist.* **24** 220–238.

ROY, S. N. (1957). *Some Aspects of Multivariate Analysis*. Wiley, New York.

WILKS, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika*. **24** 471–494.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

LOCKHEED AIRCRAFT CORPORATION
LOCKHEED PALO ALTO RESEARCH LABORATORY
PALO ALTO, CALIFORNIA 94304