The Annals of Statistics
1979, Vol. 7, No. 5, 1003–1018

CONSTRUCTING ALL SMALLEST SIMULTANEOUS CONFIDENCE SETS IN A GIVEN CLASS, WITH APPLICATIONS TO MANOVA[1]

BY ROBERT A. WIJSMAN

University of Illinois at Urbana-Champaign

A method is presented for the construction of all families of smallest simultaneous confidence sets (SCS) in a given class, for a family $\{\psi_i(\gamma)\}$ of parametric functions of the parameter of interest $\gamma = \gamma(\theta)$. The method is applied to the MANOVA problem (in its canonical form) of inference about $M = EX$, where X is $q \times p$ and has rows that are independently multivariate normal with common covariance matrix Σ. Let S be the usual estimate of Σ and put $W = (M - X)S^{-\frac{1}{2}}$. It is shown that smallest equivariant SCS for all $a'M$, $a \in R^q$, are necessarily those that are exact with respect to the confidence set for M determined by $\lambda_1(WW') < \text{const}$ (λ_1 = maximum characteristic root), i.e., derived from the acceptance region of Roy's maximum root test (this is strictly true for $p < q$, and true for $p \geq q$ under a weak additional restriction). It is also shown that smallest equivariant SCS for all tr NM, with rank $(N) \leq r$, are necessarily those that are exact with respect to $\|W\|_{\varphi_r} \leq 1$, where φ_r is a symmetric gauge function that, on the ordered positive cone, depends only on the first r arguments. Taking $r = 1$, the simultaneous confidence intervals for all $a'Mb$ of Roy and Bose emerge, and $r = \min(p, q)$ results in the simultaneous confidence intervals for all tr NM of Mudholkar.

1. Introduction and basic result. In this paper a method will be exhibited that is useful in several statistical situations to find for a specified family of parametric functions all families of smallest simultaneous confidence sets of a given kind. The method will be applied to the multivariate analysis of variance (MANOVA) problem, and the restriction will be made to equivariant estimators. Some of the families of parametric functions have not been treated before. Others have appeared in the literature ([4], [6], [11], [16]), but it will be shown now for the first time that the proposed simultaneous confidence sets are the only smallest equivariant ones.

Simultaneous confidence sets (abbreviated SCS hereafter) often arise in connection with a testing problem. Suppose X is a random variable with values in a measureable space $\mathcal{X}$ and having distribution P_θ, $\theta \in \Theta$. Suppose that inference is desired about some function of θ, say $\gamma = \gamma(\theta)$, with values in a space Γ. In the following the symbol γ is used to denote the function $\gamma : \Theta \to \Gamma$, or a point of Γ, depending on the context.

Received June 1975; revised May 1977.

[1]Research supported by the National Science Foundation under grant MCS 75-07978.

AMS 1970 *subject classifications.* Primary 62F25; secondary 62J10, 15A45.

Key words and phrases. Simultaneous confidence sets, simultaneous confidence intervals, smallest, exact, equivariant, MANOVA, symmetric gauge functions, extremum lemmas involving trace.

For given $\gamma_0 \in \Gamma$ let $H : \gamma(\theta) = \gamma_0$ be a hypothesis, and let T be a test for H with acceptance region $A(\gamma_0) \subset \mathfrak{X}$. If H is rejected, then the statistician would like to know which of the many different possible causes is or are responsible for $\gamma(\theta) \neq \gamma_0$. More precisely, let there be given a family $\{\psi_i : i \in I\}$ of parametric functions $\psi_i : \Gamma \to \Psi$, with I an arbitrary index set and Ψ an arbitrary space. Furthermore, let $\gamma = \gamma_0$ if and only if $\psi_i(\gamma) = \psi_i(\gamma_0)$ for every $i \in I$. Then each $i \in I$ for which $\psi_i(\gamma) \neq \psi_i(\gamma_0)$ can be called a cause for the falsity of H. Now suppose for each $i \in I$ there is given a random subset $B_i = B_i(X)$ of Ψ and suppose B_i is regarded as a confidence set for $\psi_i = \psi_i(\gamma)$. Then the simultaneous confidence statement $\{\psi_i(\gamma) \in B_i \forall i \in I\}$ is made. For those $i \in I$ for which $\psi_i(\gamma) \notin B_i$, the corresponding $\psi_i(\gamma)$ is then judged to be different from $\psi_i(\gamma_0)$, pointing to a cause, or causes, for the rejection of H. Gabriel (1969) terms the family $\{B_i\}$ *coherent* (relative to the test T) if acceptance of H implies $\psi_i(\gamma_0) \in B_i$ for every $i \in I$, and *consonant* if rejection of H implies $\psi_i(\gamma_0) \notin B_i$ for some $i \in I$. Thus, $\{B_i\}$ is coherent and consonant if the event $[\psi_i(\gamma_0) \in B_i \forall i \in I]$ is the same as the event that T accepts H. This is the case, for instance, with the Scheffé-type simultaneous confidence intervals for all contrasts in a one-way analysis of variance relative to the usual F-test (Scheffé (1959)).

Next, consider a family of hypotheses and tests, one for each $\gamma_0 \in \Gamma : \{(H(\gamma_0), T(\gamma_0)) : \gamma_0 \in \Gamma\}$, and let $A(\gamma_0)$ be the acceptance region of $T(\gamma_0)$. The family $\{A(\gamma_0) : \gamma_0 \in \Gamma\}$ can be converted in the usual way into a confidence set $F = F(X) \in \Gamma$ for γ, such that $H(\gamma_0)$ is accepted if and only if $\gamma_0 \in F$. Then $\{B_i\}$ is coherent and consonant with respect to each $(H(\gamma_0), T(\gamma_0))$ if and only if

$$[\gamma(\theta) \in F(X)] = [\psi_i(\gamma(\theta)) \in B_i(X) \forall i \in I], \qquad \forall \theta \in \Theta, \tag{1.1}$$

i.e., the two events on both sides of (1.1) are identical, no matter what the true value of $\theta \in \Theta$ is. If (1.1) is satisfied we shall say that $\{B_i\}$ is *exact* with respect to F. If $\{A_i\}$ is exact with respect to F and $A_i \subset B_i \forall i \in I$ for every $\{B_i\}$ that is exact with respect to F, then $\{A_i\}$ will be termed a family of *smallest* exact SCS with respect to F. As an example, the Scheffé-type simultaneous confidence intervals for all linear parametric functions that are 0 under a given linear hypothesis are smallest exact SCS with respect to the confidence set derived from the usual F-test. This follows as a special case of Theorem 4.1 (i).

In Theorem 1.1 it will be shown how to construct a family $\{A_i\}$ of smallest SCS given a family $\{B_i\}$, as well as the confidence set F with respect to which both are exact. In most interesting statistical problems the class of all SCS $\{B_i\}$ for $\{\psi_i\}$ may be unworkably large and it will be convenient to restrict oneself to a smaller class with "nice" properties. In estimation problems that are invariant under a group G a natural restriction is to *equivariant* estimators. Invariance reduction is especially effective in multivariate normal problems which usually possess a large amount of symmetry. Roughly speaking, an equivariant estimator is one that transforms under

G in the same way as the function to be estimated. A precise definition is given in Section 3.

A word is in order with regard to the notion of a *random subset* of a certain space, e.g. of Γ. Let F^* be a subset of $\mathfrak{X} \times \Gamma$ such that its γ-section $\{x \in \mathfrak{X} : (x, \gamma) \in F^*\}$ is measureable for each $\gamma \in \Gamma$. Then its X-section $\{\gamma \in \Gamma : (X, \gamma) \in F^*\}$ is a subset of Γ depending on X. It will be called a *random subset of* Γ and denoted $F(X)$ or simply F. We shall also say that F is measureable. Analogous considerations apply to the random subsets $B_i = B_i(X)$ of Ψ.

Let F be a random subset of Γ and $\{B_i\}$ a family of random subsets of Ψ. The definition (1.1) of $\{B_i\}$ being exact with respect to F can be rewritten more conveniently

$$F = \cap \{\psi_i^{-1}B_i : i \in I\}. \tag{1.2}$$

(Here and elsewhere we shall often omit parentheses and write $fA, f^{-1}B$ instead of $f(A), f^{-1}(B)$ if f is a function and A, B are sets). Equation (1.2) and similar ones to follow are tacitly understood to hold for every $\theta \in \Theta$. The main tool in this paper is forged in the next theorem.

1.1 THEOREM. *Let $\{B_i : i \in I\}$ be a given family of random subsets of Ψ. Define F by* (1.2) *and suppose F measureable. Also define*

$$A_i = \psi_i F, \qquad i \in I. \tag{1.3}$$

Then

$$F = \cap \{\psi_i^{-1}A_i : i \in I\} \tag{1.4}$$

and $\{A_i : i \in I\}$ is the family of smallest SCS *for $\{\psi_i\}$ that is exact with respect to F.*

PROOF. Define $F_1 = \cap \{\psi_i^{-1}A_i : i \in I\}$. It will first be shown that $A_i \subset B_i \forall i \in I$ and that $F_1 = F$, thereby proving (1.4). It follows from (1.2) and (1.3) that $A_i \subset \psi_i\psi_i^{-1}B_i = B_i$ so that $F_1 \subset F$. On the other hand, $F \subset \psi_i^{-1}\psi_i F = \psi_i^{-1}A_i \forall i \in I$, so that $F \subset \cap \{\psi_i^{-1}A_i : i \in I\} = F_1$.

By (1.4), $\{A_i\}$ is exact with respect to F. Let $\{C_i\}$ be any family of SCS exact with respect to F, then (1.2) holds with the B_i replaced by the C_i. By the first part of the proof, $A_i \subset C_i$. $\square$

For given $\{\psi_i\}$ let $\mathcal{C}$ be a class of families $\{B_i\}$ of SCS for $\{\psi_i\}$. Suppose that for each $\{B_i\} \in \mathcal{C}$ the family $\{A_i\}$ constructed by (1.2) and (1.3) is also in $\mathcal{C}$. It follows then from Theorem 1.1 that these families $\{A_i\}$ are the families of smallest SCS for $\{\psi_i\}$. This will be applied in Section 4 to various choices for $\{\psi_i\}$ in the MANOVA problem. For each $\{\psi_i\}$ the class $\mathcal{C}$ will be taken as all equivariant $\{B_i\}$ and for each such $\mathcal{C}$ it will indeed be true that $\{B_i\} \in \mathcal{C}$ implies $\{A_i\} \in \mathcal{C}$.

There are other families $\{\psi_i\}$ in the MANOVA problem, as well as other multivariate problems, where Theorem 1.1 can be used. Some of these problems are discussed in Wijsman [18], including the simultaneous confidence intervals in MANOVA based on a step-down procedure treated by Mudholkar and Subbaiah

(1975), and those associated with nonindependence between sets of variates treated by Mudholkar (1966). The point of view in [18] is slightly different from the one in the present paper in that the starting point there is a confidence set F for γ rather than a family $\{B_i\}$ for $\{\psi_i\}$.

2. Simultaneous confidence sets in the MANOVA problem. In its canonical form the MANOVA testing problem can be stated as follows (see e.g. Lehmann (1959)). The rows of the random matrices X, Y, and Z are independent p-variate normal with common, unknown, nonsingular covariance matrix Σ; $EX = M$: $q \times p$, $EY = M_Y$, $EZ = 0$ (M and M_Y unknown), and $S = Z'Z$ is nonsingular with probability one. The task is to test the hypothesis H that $M = 0$. If H is rejected, various families of parametric functions can be proposed in order to look at M in more detail. The family that is the most natural generalization of the univariate ANOVA problem is the family of all $a'M$ as a runs through R^q. More detail is provided by the "double linear compounds" $a'Mb$, $a \in R^q$, $b \in R^p$, treated by Roy and Bose (1953). These are special cases of the functions $A'MB$ (A and B matrices) treated by Gabriel (1968). The most detail results from looking at all linear combinations of the elements of M. This leads to all tr NM, treated by Mudholkar (1966), where N runs through all $p \times q$ matrices. All SCS are derived from the pivotal quantity for M

$$(X - M)S^{-1}(X - M)', \tag{2.1}$$

and use the distribution of the ordered characteristic roots $\lambda_1 > \cdots > \lambda_s > 0$ (strict inequalities with probability 1) of the matrix (2.1), where $s = \min(p, q)$. The distribution of the maximum characteristic root λ_1 plays an especially important role. Let λ_α be its upper α-point:

$$P(\lambda_1 \leqslant \lambda_\alpha) = 1 - \alpha. \tag{2.2}$$

Related to (2.2) is Roy's maximum root test which rejects the hypothesis $M = 0$ if $\lambda_1(XS^{-1}X') > \lambda_\alpha$. The confidence set for M derived from the maximum root test is

$$\{M : \lambda_1((M - X)S^{-1}(M - X)') \leqslant \lambda_\alpha\}. \tag{2.3}$$

The following SCS for all $a'M$ are implicit in Gabriel (1968) and explicit in Jensen and Mayer (1977):

$$P_\theta\{(a'M - a'X)S^{-1}(a'M - a'X)' \leqslant \|a\|^2\lambda_\alpha \forall a \in R^q\} = 1 - \alpha, \quad \forall\theta \in \Theta. \tag{2.4}$$

Roy and Bose (1953) derived for all $a'Mb$ the following simultaneous intervals:

$$P_\theta\{|a'Mb - a'Xb| \leqslant \|a\|(b'Sb)^{\frac{1}{2}}\lambda_\alpha^{\frac{1}{2}} \forall a \in R^q, b \in R^p\} = 1 - \alpha, \quad \forall\theta \in \Theta. \tag{2.5}$$

Optimality of the intervals (2.5) was shown by Gabriel (1968) within a class defined by so-called increasing root functions. Similar optimality of (2.4) is implicit in Gabriel (1968) and also shown by Jensen and Mayer (1977) within that same class. In Section 4 we shall prove the stronger result that within the larger class of all equivariant SCS those of (2.5) are smallest for all $a'Mb$ and those of (2.4) are smallest for all $a'M$ if $p < q$. If $p \geqslant q$ the sets of (2.4) are smallest provided an additional mild condition is imposed on their shape (see Theorem 4.1 (ii)). Moreover, these SCS are exact with respect to the confidence set (2.3) for M. In fact, the latter is the only $(1 - \alpha)$-confidence set for M with respect to which there exist equivariant SCS for all $a'Mb$ and for all $a'M$ if $p < q$.

The result mentioned in the previous sentence has the following somewhat unpleasant consequence for the practicing statistician. Suppose this statistician has a favored test for $H : M = 0$, e.g. Wilks' lambda test (the likelihood ratio test). Suppose also that SCS $A_{a,b}$ for all $a'Mb$ are desired in such a way that H is rejected if and only if some $A_{a,b}$ does not contain 0. However, it turns out now that this is impossible with Wilks' lambda test (or Hotelling-Lawley's trace test, or Pillai's trace test, etc.), and the only test capable of furnishing exact equivariant SCS for all $a'Mb$ is Roy's maximum root test. The same is true for all $a'M$ if $p < q$, and if $p \geqslant q$ under a mild additional condition.

Before stating the confidence intervals for all tr NM a little more notation is needed. Denote by $\mathfrak{M}_{m,n}$ the set of all $m \times n$ real matrices. Every $A \in \mathfrak{M}_{m,n}$ has a *singular value decomposition* (see e.g. [14])

$$A = UDV', \tag{2.6}$$

in which (denoting $s = \min(m, n)$) $U'U = V'V = I_s$, $D = \operatorname{diag}(d_1, \cdots, d_s)$ with $d_1 \geqslant \cdots \geqslant d_s \geqslant 0$. (If $m = n$ both U and V are orthogonal; otherwise one of U and V is incomplete orthogonal.) If $\operatorname{rank}(A) = r$, then exactly r of the d_i will be > 0. It will be convenient to denote

$$d(A) = (d_1, \cdots, d_s)', \tag{2.7}$$

i.e., $d(A)$ is the column vector of ordered singular values of A.

Next consider Euclidean n-space R^n and its *ordered positive cone*

$$R^n_{o+} = \{x \in R^n : x_1 > \cdots > x_n > 0\}. \tag{2.8}$$

Let G_s be the group of sign changes and permutations of the coordinates x_i of $x \in R^n$, i.e., G_s acts on R^n and the action gx of G_s on $x = (x_1, \cdots, x_n)'$ is a composition of transformations of the type $x_i \to -x_i$ and $x_i \leftrightarrow x_j$. Any transformation with $g \in G_s$ will be called a *symmetry* operation, and if a function f on R^n is invariant under G_s, then f will be called symmetric.

Lastly, a function $\varphi : R^n \to R$ is called a *gauge function*, or simply a *gauge*, if φ is convex and positively homogeneous (i.e., $\varphi(cx) = c\varphi(x)$ if $c > 0$) (cf. [1], [15]). Following von Neumann (1937), φ will be called a *symmetric gauge function* (sgf) if it is a gauge and symmetric. The functions $\varphi \equiv 0$ and $\varphi \equiv \infty$ will be excluded. Denote by Φ_n all remaining sgf's on R^n. Then for each $\varphi \in \Phi_n$, $\varphi(0) = 0$ and

$\varphi(x) > 0$ if $x \neq 0$. For $\varphi \in \Phi_n$ the *polar* of φ, denoted φ^o, is defined by

$$\varphi^o(x) = \max\{y'x : y \in R^n, \varphi(y) \leqslant 1\}, \qquad x \in R^n. \tag{2.9}$$

The maximum in (2.9) can also be taken over all $y \in R^n$ with $\varphi(y) = 1$.

Note on nomenclature. Von Neumann (1937) used the term "conjugate" (denoted ψ) instead of "polar." This has been followed by Mudholkar (1965, 1966). However, in the literature on convexity the term "polar" seems now well established ([1], [15]), reserving "conjugate" for the Fenchel conjugate ([3], [15]). The notation φ^o for the polar of φ follows the notation in [15]. For guiding me into the literature on convexity and duality I am much indebted to Lynn McLinden.

It is true for any gauge φ that φ^o is also a gauge and that $\varphi^{oo} = \varphi$. Moreover, if φ is symmetric, then obviously the same is true for φ^o. Therefore, $\varphi \in \Phi_n$ implies $\varphi^o \in \Phi_n$. Now let $A \in \mathfrak{M}_{m,n}$ and $\varphi \in \Phi_s$, where $s = \min(m, n)$, and denote

$$\|A\|_\varphi = \varphi(d(A)), \tag{2.10}$$

where $d(A)$ was defined in (2.7). For square matrices A von Neumann (1937) showed that (2.10) defines a norm, and Mudholkar (1965) extended this to arbitrary matrices.

The following random matrix plays an important role:

$$W = (M - X)S^{-\frac{1}{2}}. \tag{2.11}$$

(Note that the matrix (2.1) is WW'.) Its singular values are the square roots of the characteristic roots λ_i of (2.1), and the distribution of $d(W)$ (cf. (2.7)) is free of the parameters. Mudholkar (1965) recognized the importance of the symmetric gauge functions and in [11] derived the following simultaneous confidence intervals for all tr NM:

$$P_\theta\left\{|\operatorname{tr} NM - \operatorname{tr} NX| \leqslant \|S^{\frac{1}{2}}N\|_{\varphi^o} \forall N \in \mathfrak{M}_{p,q}\right\} = 1 - \alpha, \qquad \forall \theta \in \Theta, \tag{2.12}$$

in which φ is any function in Φ_s ($s = \min(p, q)$) for which

$$P\{\|W\|_\varphi \leqslant 1\} = 1 - \alpha. \tag{2.13}$$

It will be shown in Section 4 (Theorem 4.2) that for every $\varphi \in \Phi_s$ and $0 < \alpha < 1$, (2.12) is a family of smallest equivariant SCS for all tr NM and that those are the only ones if N is unrestricted.

There is a relationship between the confidence intervals of (2.12) and those of (2.5). In the latter, $a'Mb$ can be written as tr NM with $N = ba'$, a matrix of rank $\leqslant 1$. Thus (2.5) provides simultaneous confidence intervals for all tr NM as N ranges through all matrices of rank $\leqslant 1$, and it will be shown later in this section that (2.12) with N thus restricted reduces to (2.5). The conjecture arises immediately that confidence intervals analogous to (2.12) can be derived for N restricted to any intermediate rank.

Before executing the idea expressed in the preceding sentence it will be necessary to define a particular extension of a sgf to one on a space of higher dimension. The extension φ_s of φ_r (with $s > r$) defined below has the property that on R^s_{o+} it depends only on the first r coordinates and coincides there with φ_r.

2.1. LEMMA. *Let* $r < s$ *and* $\varphi_r \in \Phi_r$. *If* $x = (x_1, \cdots, x_s)'$, *denote* $x^r = (x_1, \cdots, x_r)'$. *Define, for* $x \in R^s$,

$$\varphi_s(x) = \max\{\varphi_r((gx)^r) : g \in G_s\}. \tag{2.14}$$

Then $\varphi_s \in \Phi_s$ *and*

$$\varphi_s(x) = \varphi_r(x^r) \quad \text{if} \quad x \in R^s_{o+}. \tag{2.15}$$

PROOF. Use (2.9) with $n = r$ and the roles of φ and φ^o interchanged. Then

$$\varphi_r((gx)^r) = \max\{y'(gx)^r : y \in R^r, \varphi^o_r(y) = 1\}. \tag{2.16}$$

Substituting (2.16) into (2.14) and interchanging the order in which the maxima are taken, it is seen that $\varphi_s = \max\{f_y : y \in R^r, \varphi^o_r(y) = 1\}$, where $f_y(x) = \max\{y'(gx)^r : g \in G_s\}$. It is immediate that f_y is symmetric for every y. Furthermore, it is a gauge since it is a maximum of linear functions. Therefore, $f_y \in \Phi_s$ for every y, and the same is then true of φ_s as a maximum of such functions.

To show (2.15), first observe that when φ_s is written as an iterated maximum (over y and g, with help of (2.16)), it is obviously permissible to restrict y to R^r_{o+} and its limit points, or, alternatively, to restrict y to R^r_{o+} and replace maximum over y by supremum. Then for $x \in R^s_{o+}$ the maximum over $g \in G_s$ is attained if $g = e$, and the maximum value of $y'(gx)^r$ is $\Sigma^r_1 y_i x_i = y'x^r$. Then $\varphi_s(x) = \sup\{y'x^r : y \in R^r_{o+}, \varphi^o(y) = 1\}$, and this also equals $\varphi_r(x^r)$ if $x \in R^s_{o+}$ so that $x^r \in R^r_{o+}$. □

It will be very convenient in the sequel not to introduce a notation for the extension of a function $\varphi \in \Phi_r$. Thus, if $\varphi \in \Phi_r$, then φ will also be considered a function in Φ_s for any $s > r$ via the extension defined in Lemma 2.1. In particular, this will be used in the definition (2.10) of the φ-norm of a matrix.

Define $\mathfrak{M}^r_{m,n}$ to be all $m \times n$ matrices of rank $\leqslant r$. The following lemma generalizes Lemma 3.9 in [11], which, in turn, generalizes a result of von Neumann (1937).

2.2 LEMMA. *Let* $1 \leqslant r \leqslant \min(p, q)$ *and* $\varphi \in \Phi_r$ *be given.* (i) *If* $N \in \mathfrak{M}^r_{p,q}$, *then* $\max\{\operatorname{tr} NW : W \in \mathfrak{M}_{q,p}, \|W\|_\varphi \leqslant 1\} = \|N\|_{\varphi^o}$. (ii) *If* $W \in \mathfrak{M}_{q,p}$, *then* $\max\{\operatorname{tr} NW : N \in \mathfrak{M}^r_{p,q}, \|N\|_{\varphi^o} \leqslant 1\} = \|W\|_\varphi$.

This lemma will be proved in Section 5. If desired, in (i) of the lemma the maximum may be taken over $\|W\|_\varphi = 1$, and in (ii) over $\|N\|_{\varphi^o} = 1$. Note that from each of the two equalities in the lemma there follows one with max replaced by min and the right hand side replaced by its negative. Lemma 2.2 leads immediately to simultaneous confidence intervals for all tr NM with N running through all $p \times q$ matrices of rank $\leqslant r$, generalizing Mudholkar's (1966) simultaneous confidence intervals (2.12).

2.3. THEOREM. *Let* $0 < \alpha < 1$ *and integer* r *with* $1 \leqslant r \leqslant s = \min(p, q)$ *be given, and let* $\varphi \in \Phi_r$ *be such that*

$$P\{\|W\|_\varphi \leqslant 1\} = 1 - \alpha, \tag{2.17}$$

then

$$P\{|\mathrm{tr}\, NM - \mathrm{tr}\, NX| \leqslant \|S^{\frac{1}{2}}N\|_{\varphi^o} \forall N \in \mathfrak{M}^r_{p,q}\} = 1 - \alpha. \tag{2.18}$$

In fact, the events on the left hand sides of (2.17) *and* (2.18) *are identical.*

PROOF. Using (2.11), the event on the left hand side in (2.18) is the same as the event $\{|\mathrm{tr}\, S^{\frac{1}{2}}NW| \leqslant \|S^{\frac{1}{2}}N\|_{\varphi^o} \forall N \in \mathfrak{M}^r_{p,q}\} = \{|\mathrm{tr}\, NW| \leqslant \|N\|_{\varphi^o} \forall N \in \mathfrak{M}^r_{p,q}\}$ (using the fact that $S^{\frac{1}{2}}N \in \mathfrak{M}^r_{p,q}$ if and only if $N \in \mathfrak{M}^r_{p,q}$) $= \{|\mathrm{tr}\, NW| \leqslant 1 \forall N \in \mathfrak{M}^r_{p,q}, \|N\|_{\varphi^o} = 1\} = \{-1 \leqslant \min \mathrm{tr}\, NW \leqslant \max \mathrm{tr}\, NW \leqslant 1 \forall N \in \mathfrak{M}^r_{p,q}, \|N\|_{\varphi^o} = 1\} = \{\|W\|_\varphi \leqslant 1\}$ by Lemma 2.2 (ii). □

Specializing Theorem 2.3 to $r = s$ reproduces, of course, Mudholkar's intervals (2.12). Specializing Theorem 2.1 to $r = 1$, the Roy-Bose intervals (2.5) result. Indeed, the only type of functions φ in Φ_1 is $\varphi(x) = c|x|$, $c > 0$, and its polar is $\varphi^o(x) = c^{-1}|x|$. Then $\|W\|_\varphi = c\lambda_1^{\frac{1}{2}}$ and in order that (2.17) be satisfied we must have $c^{-1} = \lambda_\alpha^{\frac{1}{2}}$ (see (2.3)). Hence, $\varphi^o(x) = \lambda_\alpha^{\frac{1}{2}}|x|$. Putting $N = ba'$, the only possible nonzero singular value of $S^{\frac{1}{2}}N$ is $(\mathrm{tr}\, N'SN)^{\frac{1}{2}} = \|a\|\,(b'Sb)^{\frac{1}{2}}$ so that $\|S^{\frac{1}{2}}N\|_{\varphi^o} = \lambda_\alpha^{\frac{1}{2}}\|a\|\,(b'Sb)^{\frac{1}{2}}$. Thus, (2.18) becomes (2.5).

In the remainder of this paper it will be sufficient to deal only with the SCS (2.4) for all $a'M$, and with the confidence intervals (2.18) for all $\mathrm{tr}\, NM$, $N \in \mathfrak{M}^r_{p,q}$, since (2.5) and (2.12) are special cases of (2.18).

3. Equivariant simultaneous set estimators. Reverting temporarily to the general notation of Section 1, it is convenient to regard the family of functions $\{\psi_i : i \in I\}$ as a single function $\psi : \Gamma \times I \to \Psi$. In the same spirit, SCS for $\{\psi_i\}$ will be regarded as a function Q on $\mathfrak{X} \times I$ whose values are subsets of Ψ. The function Q will sometimes be called a *simultaneous set estimator*.

Suppose there is a group G of invariance transformations, i.e., G acts on $\mathfrak{X}$ (measureably) and on Θ, and $P_{g\theta}(gA) = P_\theta(A)$ for every measureable $A \subset \mathfrak{X}$ and all $g \in G$, $\theta \in \Theta$. Suppose, furthermore, that an action of G on Γ can be defined such that $\gamma(g\theta) = g\gamma(\theta) \forall g \in G, \theta \in \Theta$ (this is true iff $\gamma(\theta_1) = \gamma(\theta_2) \Rightarrow \gamma(g\theta_1) = \gamma(g\theta_2) \forall g \in G, \theta_1, \theta_2 \in \Theta$). Finally, suppose that an action of G on I and, for each $i \in I$, an action of G on Ψ (denoted $\psi \to g_i\psi$) can be defined such that $\psi(g\gamma, gi) = g_i\psi(\gamma, i) \forall \gamma \in \Gamma, i \in I, g \in G$. Then the problem of estimating $\gamma(\theta)$ and estimating simultaneously all $\psi_i(\gamma)$ is invariant under G.

Writing temporarily $t = (x, i)$, $x \in \mathfrak{X}$, $i \in I$, a simultaneous set estimator Q for ψ is called *equivariant* if $Q(gt) = g_iQ(t) \forall t \in \mathfrak{X} \times I, g \in G$ (cf. Lehmann (1959), page 243, (27), for the definition of a single equivariant (called "invariant" there) confidence set). Replacing t by (x, i), and x by the random variable X, we write the definition of equivariance of Q as

$$Q(gX, gi) = g_i Q(X, i) \qquad \forall g \in G, i \in I. \tag{3.1}$$

For the application to MANOVA the random variable X in (3.1) will be replaced by (X, Y, Z), and the index i by $a \in R^q$ or by $N \in \mathfrak{M}^r_{p,q}$. The problem of obtaining SCS for all $a'M$, $a \in R^q$ [all tr NM, $N \in \mathfrak{M}^r_{p,q}$] is invariant under a group G_1 [G_2] defined below by the action of its various subgroups (labeled (a), (b), etc.). Any action not indicated is supposed to be trivial. *Notation*: $O(n)$ = all $n \times n$ orthogonal matrices; $GL(n)$ = all $n \times n$ real nonsingular matrices; $q_2[q_3]$ = number of rows of $Y[Z]$. Points of Ψ will be denoted ψ (not to be confused with the function ψ). The nontrivial actions are:

G_1(a): $\forall A_1 \in \mathfrak{M}_{q,p}$, $X \to X + A_1$, $M \to M + A_1$, $\psi \to \psi + a'A_1$;
(b): $\forall A_2 \in \mathfrak{M}_{q_2,p}$, $Y \to Y + A_2$, $M_Y \to M_Y + A_2$;
(c): $\forall U_1 \in 0(q)$, $X \to U_1 X$, $M \to U_1 M$, $a \to U_1 a$;
(d): $\forall U_2 \in 0(q_3)$, $Z \to U_2 Z$;
(e): $\forall C \in GL(p)$, $[X', Y', Z', M', M_Y'] \to C[X', Y', Z', M', M_Y']$, $\Sigma \to C\Sigma C'$, $\psi \to \psi C'$;
(f): $\forall c \neq 0$, $a \to ca$, $\psi \to c\psi$.
G_2(a): $\forall A_1 \in \mathfrak{M}_{q,p}$, $X \to X + A_1$, $M \to M + A_1$, $\psi \to \psi + \mathrm{tr}\, NA_1$;
(b): same as $G_1(b)$;
(c): $\forall U_1 \in 0(q)$, $X \to U_1 X$, $M \to U_1 M$, $N \to NU_1'$;
(d): same as $G_1(d)$;
(e): in $G_1(e)$ replace $\psi \to \psi C'$ by $N \to C'^{-1}N$;
(f): $\forall c \neq 0$, $N \to cN$, $\psi \to c\psi$.

3.1. LEMMA. *A simultaneous set estimator Q for all $a'M$, $a \in R^q$, is equivariant if and only if*

$$Q(X, Y, Z, a) = \{a'M \in R^p : a'(M - X)S^{-1}(M - X)'a \in \|a\|^2 J\}, \qquad a \in R^q, \tag{3.2}$$

in which J is any Borel subset of $[0, \infty)$.

PROOF. Set

$$Q(X, Y, Z, a) = a'X + F(X, Y, Z, a) \tag{3.3}$$

in which F is some subset of R^p, regarded as a space of row vectors. In (3.1) replace X by (X, Y, Z), i by a, and take g belonging to the subgroup defined in $G_1(a)$. The left hand side of (3.1) is $Q(X + A_1, Y, Z, a) = a'(X + A_1) + F(X + A_1, Y, Z, a)$ by (3.3). The right hand side of (3.1) is $Q(X, Y, Z, a) + a'A_1 = a'X + F(X, Y, Z, a) + a'A_1$ by (3.3). Equating the two sides of (3.1) gives $F(X + A_1, Y, Z, a) = F(X, Y, Z, a)$ for every $A_1 \in \mathfrak{M}_{q,p}$, showing that F does not depend on X. Taking g of $G_1(b)$ similarly shows that F does not depend on Y. Taking g of $G_1(d)$ shows $F(U_2 Z, a) = F(Z, a) \forall U_2 \in 0(q_3)$ from which it follows that F depends on Z only through $Z'Z = S$. Write $F = F(S, a)$. Applying g of $G_1(c)$

(which leaves $a'X$ and Q invariant) shows $F(S, U_1 a) = F(S, a) \forall U_1 \in 0(q)$ so that F depends on a only through $\|a\|$. Combining this with the transformations of $G_1(f)$ with $c > 0$ shows $F(S, \|a\|) = \|a\| F(S, 1) = \|a\| F(S)$, say. Finally, taking g of $G_1(e)$ leads to $F(CSC') = F(S)C' \forall C \in GL(p)$, $S : p \times p$ positive definite. First take $S = I_p$ and $C = U \in 0(p)$, then from $F(I) = F(I)U'$ it follows that $F(I) = \{v \in R^P : \|v\|^2 \in J\}$ for some $J \subset [0, \infty)$. Second, take $C = S^{-\frac{1}{2}}$, then $F(I) = F(S)S^{-\frac{1}{2}}$ so that $F(S) = F(I)S^{\frac{1}{2}} = \{vS^{\frac{1}{2}} \in R^p : \|v\|^2 \in J\} = \{v \in R^p : vS^{-1}v' \in J\}$. Hence, by (3.3), $Q(X, Y, Z, a) = a'X + \|a\|\{v \in R^p : vS^{-1}v' \in J\}$. Replacing v by $a'M$ yields (3.2). □

3.2. Lemma. *A simultaneous set estimator Q for all* tr NM, $N \in \mathfrak{M}^r_{p,q}$, *is equivariant if and only if*

$$Q(X, Y, Z, N) = \left\{\operatorname{tr} NM : |\operatorname{tr} N(M - X)| \in J\left(d\left(S^{\frac{1}{2}}N\right)\right)\right\}, \quad N \in \mathfrak{M}^r_{p,q}, \tag{3.4}$$

in which $d(N)$ is the vector of ordered singular values of N, and J is a Borel subset of $[0, \infty)$ *depending positively homogeneously on the vector* $d(\cdot)$.

Proof. Write $Q(X, Y, Z, N) = \operatorname{tr} NX + F(X, Y, Z, N)$. Then expressing the definition (3.1) of equivariance under g of $G_2(a)$, (b), and (d), shows as in the proof of Lemma 3.1 that $F = F(S, N)$. The transformations of $G_2(c)$ show that $F = F(S, NN')$. Applying the transformations $G_2(e)$ leads to the equation $F(CSC', C'^{-1}NN'C^{-1}) = F(S, NN')$ and this shows that F depends on its arguments only through the characteristic roots of $NN'S$, or, equivalently, through $d(S^{\frac{1}{2}}N)$ (cf. (2.7)). Applying the transformations $G_2(f)$ gives the equation $F(d(cS^{\frac{1}{2}}N)) = cF(d(S^{\frac{1}{2}}N))$, which can be written $F(|c|d(\cdot)) = cF(d(\cdot))$. Taking $c > 0$ shows F to depend on $d(\cdot)$ positively homogeneously. Taking $c = -1$ shows $F = -F$. Put $F = J \cup (-J)$, $J \subset [0, \infty)$, then $Q(X, Y, Z, N) = \operatorname{tr} NX + F(d(S^{\frac{1}{2}}N))$ can be put in the form (3.4). □

4. Smallest equivariant SCS. Denote by $\lambda_1(A) \geqslant \lambda_2(A) \geqslant \cdots$ the ordered characteristic roots of a symmetric matrix A.

4.1. Theorem. (i) *If $p < q$, then $(1 - \alpha)$-exact equivariant* SCS *for all $a'M$, $a \in R^q$, exist only with respect to the confidence set* (2.3) *for M, derived from Roy's maximum root test. The smallest such sets are given by* (2.4).

(ii) *If $p \geqslant q$, then $(1 - \alpha)$-equivariant* SCS *are smallest if and only if they are of the form* (3.2) *with J having the property*

$$P\left\{\left[\lambda_q(WW'), \lambda_1(WW')\right] \subset J\right\} = 1 - \alpha \tag{4.1}$$

(W defined in (2.11)), *and these* SCS *are exact with respect to the confidence set for M*

$$\left\{M \in \mathfrak{M}_{q,p} : \left[\lambda_q(WW'), \lambda_1(WW')\right] \subset J\right\}. \tag{4.2}$$

If the confidence set for each $a'M$ is required to be connected and contain its center $a'X$, then the same conclusion prevails as in part (i) *of the theorem. The same is true if the confidence sets are required to be convex.*

REMARK. It will appear in the proof of part (ii) that J in (4.1) may be taken to be a countable union of disjoint nondegenerate intervals, say $J = \cup I_n$. Then the event on the left hand side of (4.1) happens iff both the smallest and the largest characteristic root of WW' are in the same I_n for some $n = 1, 2, \cdots$.

PROOF. The proofs of parts (i) and (ii) run together for a while. By (2.11), if X and S are given, then to any set F of matrices M corresponds a set F^* of matrices W, and to any set B_a of vectors $a'M$, $a \in R^q$, corresponds a set B_a^* of vectors $a'W$. It is more convenient to carry out the program laid down in Theorem 1.1 in terms of the starred sets. For simplicity of notation, however, the asterisk will be dropped. The SCS (3.2) are

$$B_a = \{a'W : a'WW'a \in \|a\|^2 J\}, \qquad a \in R^q. \tag{4.3}$$

Then by (1.2) (with i replaced by $a \in R^q$), F consists of all W such that $a'WW'a \in \|a\|^2 J \forall a \in R^q$. This condition on W is clearly only a condition on $d(W)$ (defined in (2.7)). Thus, there is a set E such that

$$F = \{W \in \mathfrak{M}_{q,p} : d(W) \in E\}. \tag{4.4}$$

The set E may be taken to be a subset of R_{o+}^s, since $P\{d(W) \in R_{o+}^s\} = 1$ ($s = \min(p, q)$), but it will be convenient to allow $0 \in E$. By (1.3),

$$A_a = \{a'W : W \in F\}, \qquad a \in R^q, \tag{4.5}$$

with F given by (4.4). In order to determine the nature of these sets, singular value decompose W for every $W \in F$, $W \neq 0$ (see (2.6)):

$$W = UDV' \tag{4.6}$$

with $U \in \mathfrak{M}_{q,s}$, $V \in \mathfrak{M}_{p,s}$, $U'U = V'V = I_s$, and $D = \operatorname{diag}(d_1(W), \cdots, d_s(W))$, with $d_1(W) > \cdots > d_s(W) > 0$ being the s ordered singular values of W. Then

$$A_a = \cup_{d(W) \in E} \cup_{U,V} a'UDV'. \tag{4.7}$$

From this point on the proofs of parts (i) and (ii) diverge.

PROOF OF (i). Since $p < q$, $s = p$ and in (4.7) V runs through $O(p)$, U through all $q \times p$ matrices with orthonormal columns. It is found that for fixed $d(W)$ in (4.7) the union over U and V produces the closed ball with radius $\|a\| d_1(W)$. Taking the union over all $d(W) \in E$ yields

$$A_a = \{v \in R^p : \|v\| \leqslant \|a\| c_E\}, \qquad a \in R^q, \tag{4.8}$$

in which

$$c_E = \sup\{d_1(W) : d(W) \in E\}, \tag{4.9}$$

or with $\leqslant$ in (4.8) replaced by $<$. For probability statements it will make no difference whether $\leqslant$ or $<$ prevails, and the former will be assumed. From (4.8) the exact nature of F can be found, by (1.4):

$$F = \{W \in \mathfrak{M}_{q,p} : a'W \in A_a \forall a \in R^q\}$$
$$= \{W \in \mathfrak{M}_{q,p} : \|a'W\| \leqslant \|a\|c_E \forall a \in R^q\},$$

which can be written

$$F = \{W \in \mathfrak{M}_{q,p} : d_1(W) \leqslant c_E\}. \tag{4.10}$$

In order that the acceptance region determined by F be of level α, $c_E = \lambda_\alpha^{\frac{1}{2}}$, and (4.10) coincides with (2.3). The confidence sets (4.8) can be written as $a'WW'a \leqslant \|a\|^2\lambda_\alpha$ and coincide with those of (2.4).

PROOF OF (ii). Now $p \geqslant q$, so $s = q$. The case $q = 1$ is rather trivial since there is now essentially only one confidence set. It is seen that J in (3.2) can be any Borel subset of $[0, \infty)$ such that $P(WW' \in J) = 1 - \alpha$. Now suppose $q > 1$. In (4.7) U runs through $O(q)$ and V through all $p \times q$ matrices with orthonormal columns. Then it is seen that for fixed $d(W) \in E$ the vectors $a'UDV'$ run through all vectors $v \in R^p$ such that

$$\|a\|d_q(W) \leqslant \|v\| \leqslant \|a\|d_1(W) \tag{4.11}$$

(this is the essential difference with case (i)). For $d(W) \in R^s_{o+}$, $[d_q(W), d_1(W)]$ is nondegenerate, closed interval. The union over all $d(W) \in E$ of these intervals is a set $K \subset [0, \infty)$ consisting of a countable union of disjoint nondegenerate intervals, with possibly $\{0\}$ adjoined. Hence, after observing (4.11),

$$A_a = \{v \in R^p : \|v\| \in \|a\|K\}, \qquad a \in R^q. \tag{4.12}$$

Then (see (1.4)) $F = \{W \in \mathfrak{M}_{q,p} : a'W \in A_a \forall a \in R^q\}$ turns out to be

$$F = \{W \in \mathfrak{M}_{q,p} : [d_q(W), d_1(W)] \subset K\}. \tag{4.13}$$

Observing that the singular values $d_i(W)$ are the square roots of the characteristic roots $\lambda_i(WW')$, and taking $J = \{x^2 : x \in K\}$, the set F of (4.13) corresponds to (4.2). The sets (4.12) for $a'W$ can be written as $a'WW'a \in \|a\|^2J$, which is (3.2) with J as announced in the theorem. These confidence sets are not connected, nor contain their center $a'X$, unless $J = [0, \lambda_\alpha]$, and then the confidence sets reduce to (2.4). The same is true if the sets are required to be convex. □

4.2 THEOREM. *Let* $1 \leqslant r \leqslant s = \min(p, q)$. *Then smallest equivariant* $(1 - \alpha)$-SCS *for all* tr NM, $N \in \mathfrak{M}^r_{p,q}$, *are necessarily those given by* (2.18), *with any* $\varphi \in \Phi_r$ *satisfying* (2.17), *and the confidence intervals* (2.18) *are exact with respect to the confidence set for* M *defined by* $\|W\|_\varphi \leqslant 1$.

PROOF. As in the proof of Theorem 4.1 F will stand for a set of matrices W rather than M, and B_N for a set of values of tr NW rather than tr NM. Then (3.4) takes the form

$$B_N = \{\operatorname{tr} NW : |\operatorname{tr} NW| \in J(d(N))\}, \qquad N \in \mathfrak{M}^r_{p,q}, \tag{4.14}$$

in which $J(d(\cdot)) \subset [0, \infty)$ and depends on $d(\cdot)$ positively homogeneously. The set F of (1.2) (with i replaced by N) is the set of all W satisfying (4.14) for every $N \in \mathfrak{M}^r_{p,q}$ and is therefore determined by conditions on $d(W)$. Hence, as in the proof of Theorem 4.1, there is a set $E \subset R^s_{o+} \cup \{0\}$ such that (4.4) holds. The corresponding sets A_N of (1.3) are

$$A_N = \{\operatorname{tr} NW : W \in F\}, \qquad N \in \mathfrak{M}^r_{p,q}, \tag{4.15}$$

and (1.4) reads

$$F = \{W \in \mathfrak{M}_{q,p} : \operatorname{tr} NW \in A_N \forall N \in \mathfrak{M}^r_{p,q}\}. \tag{4.16}$$

The theorem will be proved if it is shown that

$$F = \{W \in \mathfrak{M}_{q,p} : \|W\|_\varphi \leqslant 1\} \quad \text{for some} \quad \varphi \in \Phi_r, \tag{4.17}$$

and

$$A_N = [-\|N\|_{\varphi^o}, \|N\|_{\varphi^o}], \tag{4.18}$$

because then $\operatorname{tr} NW \in A_N \forall N \in \mathfrak{M}^r_{p,q}$ is equivalent to the simultaneous confidence intervals (2.18). Employing the singular value decomposition (2.6) for W and utilizing the representation (4.4) of F, (4.15) can be written

$$A_N = \{\operatorname{tr} NUDV' : U'U = V'V = I_s, d \in E\}, \qquad N \in \mathfrak{M}^r_{p,q}, \tag{4.19}$$

in which $D = \operatorname{diag}(d_1, \cdots, d_s)$ and $d = (d_1, \cdots, d_s)'$. Denote (see (2.7)) $d(N) = n = (n_1, \cdots, n_s)'$, in which $n_i = 0$ for all $i > r$ (since N is of rank at most r). It is proved in Lemma 5.2 that for $d \in E$ fixed, with $d_1 > \cdots > d_s > 0$,

$$\{\operatorname{tr} NUDV' : U'U = V'V = I_s\} = [-n'd, n'd]. \tag{4.20}$$

Then taking the union of (4.20) over all $d \in E$ shows A_N to be a symmetric and nondegenerate interval, unless $N = 0$ in which case $A_N = \{0\}$. For probability statements it will make no difference whether the interval is open or closed, and it will be convenient to assume A_N closed for every N. Therefore,

$$A_N = [-a_N, a_N], \qquad N \in \mathfrak{M}^r_{p,q}, \tag{4.21}$$

where $a_N > 0$ for every $N \in \mathfrak{M}^r_{p,q} - \{0\}$. Then (4.16) is

$$F = \{W \in \mathfrak{M}_{q,p} : \max\{a_N^{-1} \operatorname{tr} NW : N \in \mathfrak{M}^r_{p,q} - \{0\}\} \leqslant 1\}. \tag{4.22}$$

(Note that the condition on W in (4.22) is equivalent to $\min\{a_N^{-1} \operatorname{tr} NW : N \in \mathfrak{M}^r_{p,q} - \{0\}\} \geqslant -1$ since $N \in \mathfrak{M}^r_{p,q}$ if and only if $-N \in \mathfrak{M}^r_{p,q}$.) From (4.15) and the representation (4.4) of F it follows that A_N, and therefore a_N, depends positively homogeneously on $d(N) = n$. Write a_n instead of a_N, then $a_{cn} = ca_n$ for every $c > 0$. Take $W \in F$, $W \neq 0$, fixed and denote $d(W) = w = (w_1, \cdots, w_s)'$. Take $n = (n_1, \cdots, n_s)'$ fixed, with $n_1 \geqslant \cdots \geqslant n_r \geqslant 0 = n_{r+1} = \cdots = n_s$, then it follows from Lemma 5.1 that $\max\{\operatorname{tr} NW : d(N) = n\} = n'w = \Sigma_1^r n_i w_i$. Furthermore, utilizing the group G_s of symmetry operations introduced in Section 2,

$n'w = \max\{(gn)'w : g \in G_s\}$. Now define, for $x \in R^s$, $h_n(x) = a_n^{-1}\max\{(gn)'x : g \in G_s\}$ provided $n \neq 0$, and $h_n(x) \equiv 0$ if $n = 0$, then h_n is a sgf on R^s which (for $n \neq 0$) on R^s_{o+} equals $a_n^{-1}n'x = a_n^{-1}\Sigma_1^r n_i x_i$ and therefore on R^s_{o+} does not depend on $x_{r+1}, \cdots, x_s$. Put $\varphi = \max_n h_n$, the maximum taken over all n with $n_1 \geq \cdots \geq n_r$ and $n_i = 0$ for $i > r$ (this maximum exists as a result of the positive homogeneity of a_n). Then φ is also a sgf on R^s which on R^s_{o+} depends only on the first r coordinates. Hence, $\varphi \in \Phi_r$, and the condition in (4.22) reads: $\varphi(d(W)) \leq 1$, or, equivalently (see (2.10)), $\|W\|_\varphi \leq 1$, so that (4.22) becomes (4.17). Then from (4.15), (4.21), and (4.17) one has $a_N = \max\{\text{tr } NW : \|W\|_\varphi \leq 1\} = \|N\|_{\varphi^o}$ by Lemma 2.2 (i), proving (4.18). □

5. Matrix lemmas involving traces. The next lemma is basic for various extremal problems, including several lemmas in this paper. For square (and complex valued) matrices it originated with von Neumann (1937) who gave a rather long proof. Since then the proof has been simplified by various authors. In particular, Mirsky (1975) gives a much simpler proof, using singular value decomposition. Lemma 5.1 below generalizes von Neumann's result to matrices that are not necessarily square. The proof is close to the one in [9]. We have stated the lemma for real valued matrices, since those are the only ones occurring in this paper. However, the conclusion is equally valid for complex valued matrices if "orthogonal" is replaced by "unitary."

5.1. LEMMA. *Let $A, B \in \mathfrak{M}_{m,n}$, and $a = d(A)$, $b = d(B)$ (see (2.7)) their vectors of ordered singular values ($a_1 \geq a_2 \geq \cdots, b_1 \geq b_2 \geq \cdots$). Then*

$$\max\{\text{tr } AUB'V' : U \in 0(n), V \in 0(m)\} = a'b. \tag{5.1}$$

PROOF. Without loss of generality it may be assumed that $m \geq n$. Writing A and B in their singular value decomposition it may be further assumed that the elements a_{ij} of A and b_{ij} of B are 0 except $a_{ii} = a_i$, $b_{ii} = b_i$. Redefining A and B by deleting their last $m - n$ rows, and redefining V by deleting its last $m - n$ rows and columns, it will be shown that if

$$A = \text{diag}(a_1, \cdots, a_n),\ a_1 \geq \cdots \geq a_n \geq 0, \tag{5.2}$$

$$B = \text{diag}(b_1, \cdots, b_n),\ b_1 \geq \cdots \geq b_n \geq 0, \tag{5.3}$$

and H consists of all $n \times n$ matrices $U = ((u_{ij}))$ with

$$\Sigma_i u_{ij}^2 \leq 1 \forall j, \qquad \Sigma_j u_{ij}^2 \leq 1 \quad \forall i, \tag{5.4}$$

then

$$\max\{\text{tr } AUB'V' : U, V \in H\} = \Sigma_1^n a_i b_i. \tag{5.5}$$

Obviously, the right hand side of (5.5) can be attained by taking $U = V = I_n$. Since $\text{tr } AUB'V' = \Sigma_{ij} a_i u_{ij} b_j v_{ij}$, since the a_i and b_j are ≥ 0, and since $u_{ij}v_{ij} \leq \frac{1}{2}(u_{ij}^2 + v_{ij}^2)$, it suffices to show that

$$\Sigma_{ij} a_i b_j u_{ij}^2 \leq \Sigma_1^n a_i b_i \tag{5.6}$$

if the u_{ij} satisfy (5.4). The inequality (5.6) is proved as Lemma 1A by Ky Fan in [2]. $\square$

REMARK. Another very instructive proof of (5.6) can be given by using the notions related to Schur convexity. The u_{ij}^2 may be regarded as the elements of a doubly substochastic matrix. By increasing the elements, if necessary, the matrix can be made into a doubly stochastic matrix, say P, and during this process the left hand side of (5.6) is not decreased. It is sufficient to show that $a'Pb \leqslant a'b$ if P is doubly stochastic. Theorem 2 (credited to Birkhoff) given by L. Mirsky (1963) states that P is a convex mixture of permutation matrices: $P = \Sigma\alpha_k P_k$, $\alpha_k \geqslant 0$, $\Sigma\alpha_k = 1$, and for each permutation matrix P_k it is obvious that $a'P_k b \leqslant a'b$. An independent proof of (5.6) for $((u_{ij}^2))$ doubly stochastic, is given by Mirsky (1975).

5.2. LEMMA. *Under the same conditions as in Lemma* 5.1 *we have*

$$\{\operatorname{tr} AUB'V' : U \in 0(n),\ V \in 0(m)\} = [-a'b, a'b]. \tag{5.7}$$

PROOF. From Lemma 5.1 it follows that no point outside the interval on the right hand side of (5.15) is in the set on the left hand side of (5.7). As in the proof of Lemma 5.1 it may be assumed that $m \geqslant n$ and that A and B have the form (5.2) and (5.3). It will suffice to put $V = I_n$ and to prove that

$$\{\operatorname{tr} AUB : U \in 0(n)\} = [-a'b, a'b], \tag{5.8}$$

with A, B given by (5.2), (5.3). By taking $U = I_n(-I_n)$ the right endpoint $a'b$ (left endpoint $-a'b$) is attained. It remains to be shown that every point between $-a'b$ and $a'b$ is a possible value of tr AUB.

Consider, for $1 \leqslant k \leqslant n-1$ and $0 \leqslant \theta \leqslant \pi$, the orthogonal matrix $U_k(\theta)$ of elements $u_{kk} = u_{k+1,k+1} = \cos\theta$, $u_{k,k+1} = -u_{k+1,k} = \sin\theta$, $u_{ii} = 1$ for $i \neq k$, $i \neq k+1$. Suppose first that n is even. Take $U = U_1(\theta)$ and let θ run from 0 to π. This changes tr AUB continuously from $\Sigma_1^n a_i b_i$ to $-\Sigma_1^2 a_i b_i + \Sigma_3^n a_i b_i$, so that all intermediate values are attained. If $n = 2$, the sum Σ_3^n is empty and we are done. If not, repeat the procedure with $U_3(\theta)$, etc. Eventually, all values between $\Sigma_1^n a_i b_i$ and $-\Sigma_1^n a_i b_i$ will have been reached. Now suppose n is odd, then with the above method all values between $\Sigma_1^n a_i b_i$ and $-\Sigma_1^{n-1} a_i b_i + a_n b_n$ can be reached. Then repeat the procedure, but now starting with $-U_{n-1}(\theta)$ and working backwards. This lets tr AUB attain all values between $-\Sigma_1^n a_i b_i$ and $-a_1 b_1 + \Sigma_2^n a_i b_i$. Since $-a_1 b_1 + \Sigma_2^n a_i b_i \geqslant -\Sigma_1^{n-1} a_i b_i + a_n b_n$, the two intervals of possible values overlap so that their union is the right hand side of (5.16). $\square$

PROOF OF LEMMA 2.2. (i) Write both N and W in their singular value decomposition (2.6) and in the expression of tr NW take the maximum over the orthogonal matrices involved in that decomposition. Using Lemma 5.1, that maximum is seen to equal $n'w$, where $n = d(N)$, $w = d(W)$ (see (2.7)), and both n and w are in the closure $Cl(R_{o+}^s)$ of R_{o+}^s. Since $N \in \mathfrak{M}_{p,q}^r$, $n'w = \Sigma_1^r n_i w_i$. Since $\varphi \in \Phi_r$, $\varphi(w)$ depends only on the first r coordinates of w. Therefore, n and w can be replaced by their restrictions n^r, w^r to $Cl(R_{o+}^r)$ (the notation x^r was introduced in Lemma 2.1).

Write n^r, w^r again as n, w. Then $\max\{n'w : w \in C1(R^r_{o+}), \varphi(w) \leqslant 1\} = \max\{n'w : w \in R^r, \varphi(w) \leqslant 1\} = \varphi^o(n) = \|N\|_{\varphi^o}$, by the definition (2.9) of φ^o and the definition (2.10) of the φ-norm of a matrix.

(ii) As in the proof of part (i) it is permissible to replace n and w by their restrictions n^r, w^r. Then $\max\{n'w : n \in C1(R^r_{o+}), \varphi^o(n) \leqslant 1\} = \max\{n'w : n \in R^r, \varphi^o(n) \leqslant 1\} = \varphi(w) = \|W\|_{\varphi}$, by (2.9) and (2.10). □

Acknowledgments. I would like to thank the Associate Editor and one of the referees for their penetrating and constructive comments, which have contributed to the complete rewriting of an earlier version of this paper. Special thanks also goes to K. Ruben Gabriel who introduced me to the singular value decomposition, which is one of the most useful tools when working with matrices.

REFERENCES

[1] EGGLESTON, H. G. (1958). *Convexity*. Cambridge Univ. Press, London and New York.

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

[3] FENCHEL, W. (1953). Convex cones, sets, and functions. Mimeographed notes by D. W. Blackett. Princeton Univ. Press.

[4] GABRIEL, K. R. (1968). Simultaneous test procedures in multivariate analysis of variance. *Biometrika* **55** 489–504.

[5] GABRIEL, K. R. (1969). Simultaneous test procedures—some theory of multiple comparisons. *Ann. Math. Statist.* **40** 224–250.

[6] JENSEN, D. R. and MAYER, L. S. (1977). Some variational results and their applications in multiple inference. *Ann. Statist.* **5** 922–931.

[7] LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.

[8] MIRSKY, L. (1963). Results and problems in the theory of doubly-stochastic matrices. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **1** 319–334.

[9] MIRSKY, L. (1975). A trace inequality of John von Neumann. *Monatsh. Math.* **79** 303–306.

[10] MUDHOLKAR, GOVIND S. (1965). A class of tests with monotone power functions for two problems in multivariate statistical analysis. *Ann. Math. Statist.* **36** 1794–1801.

[11] MUDHOLKAR, GOVIND S. (1966). On confidence bounds associated with multivariate analysis of variance and nonindependence between two sets of variates. *Ann. Math. Statist.* **37** 1736–1746.

[12] MUDHOLKAR, GOVIND S. and SUBBAIAH, PERLA (1975). A note on MANOVA multiple comparisons based upon step-down procedure. *Sankhyā Ser. B* **37** 300–307.

[13] NEUMANN, J. VON (1937). Some matrix inequalities and metrization of matric spaces. *Tomsk Univ. Rev.* **1** 286–300. Reproduced in *Collected Works*. (1962). Pergamon Press, Oxford, 205–219.

[14] RAO, C. R. and MITRA, S. K. (1971). *Generalized Inverse of Matrices and its Applications*. Wiley, New York.

[15] ROCKAFELLAR, R. TYRRELL (1970). *Convex Analysis*. Princeton Univ. Press.

[16] ROY, S. N. and BOSE, R. C. (1953). Simultaneous confidence estimation. *Ann. Math. Statist.* **24** 513–536.

[17] SCHEFFÉ, HENRY (1959). *The Analysis of Variance*. Wiley, New York.

[18] WIJSMAN, ROBERT A. (1979). Smallest simultaneous confidence sets, with applications in multivariate analysis. To appear in *Multivariate Analysis V*. (P. R. Krishnaiah, ed.), North-Holland, Amsterdam.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
URBANA, ILLINOIS 61801