Journal of Multivariate Analysis **65**, 19–35 (1998)
Article No. MV971724

Perturbation Inequalities and Confidence Sets for Functions of a Scatter Matrix

Lutz Dümbgen

Institut fur Mathematik, Medizinische Universität zu Lübeck, Germany
E-mail: duembgen@math.mu-luebeck.de

Received October 6, 1994; revised October 13, 1995

Let Σ be an unknown covariance matrix. Perturbation (in)equalities are derived for various scale-invariant functionals of Σ such as correlations (including partial, multiple and canonical correlations) or angles between eigenspaces. These results show that a particular confidence set for Σ is canonical if one is interested in simultaneous confidence bounds for these functionals. The confidence set is based on the ratio of the extreme eigenvalues of $\Sigma^{-1}S$, where S is an estimator for Σ. Asymptotic considerations for the classical Wishart model show that the resulting confidence bounds are substantially smaller than those obtained by inverting likelihood ratio tests.

AMS 1991 subject classifications: 62H15, 62H20, 62H25.

Key words and phrases: correlation (partial, multiple, canonical), eigenspace, eigenvalue, extreme roots, Fisher's Z-transformation, nonlinear, perturbation inequality, prediction error, scatter matrix, simultaneous confidence bounds.

1. INTRODUCTION

Let Σ be an unknown parameter in the set $\mathbf{M}^+$ of all symmetric, positive definite matrices in $\mathbf{R}^{p\times p}$, and let $S\in\mathbf{M}^+$ be an estimator for Σ such that

$$\mathscr{L}(nS) \text{ is a Wishart distribution } \mathscr{W}(\Sigma, n) \tag{1.1}$$

for some fixed $n\geqslant p$. The goal of the present paper is to find a confidence set $C(S)$ for Σ, whose image $\phi(C(S))$ under various functions ϕ on $\mathbf{M}^+$ yields a reasonable confidence region for $\phi(\Sigma)$. We restrict our attention to *scale-invariant* functions ϕ; that means,

$$\phi(rM)=\phi(M) \quad \forall M\in\mathbf{M}^+ \quad \forall r>0. \tag{1.2}$$

Examples include correlations, regression coefficients, and eigenspaces.

0047-259X/98 $25.00

For $A \in \mathbf{R}^{p \times p}$ let $\lambda(A) \in \mathbf{R}^p$ denote the vector of its ordered eigenvalues $\lambda_1(A) \geqslant \lambda_2(A) \geqslant \cdots \geqslant \lambda_p(A)$, provided that they are real. Since $\lambda(\Sigma^{-1}S) = \lambda(\Sigma^{-1/2}S\Sigma^{-1/2})$ is a pivotal quantity, any Borel set $B \subset \mathbf{R}^p$ defines an equivariant confidence set

$$C(S) := \{\Gamma \in \mathbf{M}^+ : \lambda(\Gamma^{-1}S) \in B\}$$

for Σ, whose coverage probability $\mathbb{P}\{\Sigma \in C(S)\}$ does not depend on Σ. Here "equivariant" means that $\Gamma \in C(S)$ if, and only if, $A'\Gamma A \in C(A'SA)$ for any nonsingular matrix $A \in \mathbf{R}^{p \times p}$. For instance, inverting the likelihood ratio test of the hypotheses $\Sigma \in \{r\Gamma : r > 0\}$, $\Gamma \in \mathbf{M}^+$, leads to a confidence set of the form

$$C_{\mathrm{LR}}(S) := \left\{\Gamma \in \mathbf{M}^+ : -\sum_{i=1}^{p} \log\left(\frac{p\lambda_i}{\mathrm{trace}}(\Gamma^{-1}S)\right) \leqslant \beta_{\mathrm{LR}}\right\}$$

(cf. [1, Section 10.7]). Here β_{LR} is chosen such that $\mathbb{P}\{\Sigma \in C_{\mathrm{LR}}(S)\}$ equals $\alpha \in]0, 1[$. This set can be approximated by an ellipsoid if β_{LR} is small and yields simultaneous confidence bounds for linear functionals of Σ analogously to Scheffé's method for linear models. But many functionals of interest in multivariate analysis are nonlinear or even nondifferentiable so that one cannot rely on linear approximations. Some implications of this problem are discussed in Dümbgen [3].

A different confidence set for Σ, proposed by Roy [16, Chap. 14], is the set of all $\Gamma \in \mathbf{M}^+$ such that $\lambda_1(\Gamma^{-1}S) \leqslant \beta_1$ and $\lambda_p(\Gamma^{-1}S) \geqslant \beta_p$ with suitable numbers $\beta_1, \beta_p > 0$. If one is only interested in scale-invariant functions of Σ a possible modification of Roy's set is

$$\hat{C}(S) := \{\Gamma \in \mathbf{M}^+ : \gamma(\Gamma^{-1}S) \leqslant \beta\} = \left\{\Gamma \in \mathbf{M}^+ : \frac{\lambda_1}{\lambda_p}(\Gamma^{-1}S) \leqslant \frac{1+\beta}{1-\beta}\right\},$$

where

$$\gamma := \frac{\lambda_1 - \lambda_p}{\lambda_1 + \lambda_p},$$

and $\beta \in]0, 1[$ is a critical value satisfying

$$\mathbb{P}\{\gamma(\Sigma^{-1}S) > \beta\} = \alpha.$$

Note that $\gamma(M) = \gamma(rM) = \gamma(M^{-1})$ for $M \in \mathbf{M}^+$ and $r > 0$.

It is shown in Section 2 that this set $\hat{C}$ is a *canonical* candidate for C if one is interested in simultaneous confidence bounds for correlations (including partial, multiple, and canonical correlations). This approach sheds new light on Fisher's [9] Z-transformation. As a by-product one also obtains simultaneous confidence sets for regression coefficients similar to those of Roy [16].

Section 3 contains various results for eigenvalues and principal component vectors. In particular, new perturbation (in)equalities for eigenspaces of matrices in $\mathbf{M}^+$ are presented. The proofs for Sections 2 and 3 are deferred to Section 5.

Section 4 comments on the practical computation and the size of $\hat{C}$. The critical value β and the corresponding confidence bounds of Sections 2 and 3 are of order $O((p/n)^{1/2})$. This seems to be remarkable, because the parameter space $\mathbf{M}^+$ has dimension $p(p+1)/2$, so that in a linear model one would expect confidence bounds of order $O(p/n^{1/2})$. In fact, the set C_{LR} yields confidence bounds of that size.

The parametric assumption (1.1) is made here only for convenience. In order to compute $\hat{C}$ it suffices to know the distribution of $\gamma(\Sigma^{-1}S)$, at least approximately. Another example for this condition to hold is Tyler's [19] distribution-free M-estimator of scatter for elliptically symmetric distributions (see also Kent and Tyler [12] and Dümbgen [6]). Alternatively, let S be the sample covariance matrix of i.i.d. random vectors $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n \in \mathbf{R}^p$ with mean μ and covariance $\Sigma \in \mathbf{M}^+$. Under mild regularity conditions on the distribution of the standardized vectors $\Sigma^{-1/2}(\mathbf{y}_i - \mu)$, the distribution of $\Sigma^{-1/2} S \Sigma^{-1/2}$ can be estimated consistently as n tends to infinity by bootstrapping (cf. Beran and Srivastava [2] and Dümbgen [4]).

2. CORRELATIONS

Throughout this paper let $\mathbf{y}$ be a random vector in $\mathbf{R}^p$ with mean zero and covariance matrix Σ. For $v, w \in \mathbf{R}^p$, the covariance of the random variables $v'\mathbf{y}$ and $w'\mathbf{y}$ equals $v'\Sigma w$, and their correlation is given by

$$\rho(v, w \mid \Sigma) := \frac{v'\Sigma w}{\sqrt{v'\Sigma v w'\Sigma w}}$$

(where $\rho(0, \cdot \mid \Sigma) := 0$). An important function is

$$m(r, s) := \frac{r+s}{1+rs} = \tanh(\operatorname{arctanh}(r) + \operatorname{arctanh}(s)),$$

where $r \in [-1, 1]$ and $s \in]-1, 1[$. For fixed s, the Möbius transform $m(\cdot, s)$ is an increasing bijection of $[-1, 1]$ with inverse function

$m(\cdot, -s)$. It follows from Fisher's [9] results on the so-called Z-transformation $\rho \mapsto \operatorname{arctanh}(\rho)$ that for any pair (v, w) of linearly independent vectors,

$$m(\rho(v, w \mid S), -c) \leqslant \rho(v, w \mid \Sigma) \leqslant m(\rho(v, w \mid S), c),$$

with asymptotic probability $2\Phi(n^{1/2}c) - 1$ as $n \to \infty$, where Φ stands for the standard normal distribution function. For extensions of this result see, for instance, Hayakawa [10] and Jeyaratnam [11]. An interesting fact is that looking at many correlations simultaneously leads automatically to the Z-transformation without any asymptotic arguments. For notational convenience the unit sphere in $\mathbf{R}^p$ is denoted by $\mathbf{S}^{p-1}$.

LEMMA 1. *For arbitrary $M \in \mathbf{M}^+$ and any $\rho_o \in [-1, 1]$,*

$$\{\rho(v, w \mid M): v, w \in \mathbf{S}^{p-1}, v'w = \rho_o\} = [m(\rho_o, -\gamma(M)), m(\rho_o, \gamma(M))].$$

Lemma 1 extends Theorem 1 of Eaton [7], who considered the special case $\rho_o = 0$. If applied to $M = \Sigma^{-1/2} S \Sigma^{-1/2}$ or $M = S^{-1/2} \Sigma S^{-1/2}$, it shows that the quantity $\gamma(\Sigma^{-1}S) = \gamma(S^{-1}\Sigma)$ is of special interest.

COROLLARY 1. *For arbitrary fixed $\rho_o \in [-1, 1]$,*

$$\begin{aligned}\{\rho(v, w \mid S): v, w \in \mathbf{R}^p \setminus \{0\}, \rho(v, w \mid \Sigma) = \rho_o\} \\ = [m(\rho_o, -\gamma(\Sigma^{-1}S)), m(\rho_o, \gamma(\Sigma^{-1}S))].\end{aligned}$$

Further, let $C(S) := \{\Gamma \in \mathbf{M}^+: \lambda(\Gamma^{-1}S) \in B\}$ for some Borel set $B \subset \mathbf{R}^p$. Then for arbitrary $v, w \in \mathbf{R}^p \setminus \{0\}$ the set $\{\rho(v, w \mid \Gamma): \Gamma \in C(S)\}$ is an interval with endpoints

$$m\left(\rho(v, w \mid S), \pm \sup_{M \in C(I)} \gamma(M)\right).$$

Note that $\sup_{M \in C(I)} \gamma(M) \geqslant \beta$ for any confidence set $C(S) = \{\Gamma: \lambda(\Gamma^{-1}S) \in B\}$ with coverage probability $1 - \alpha$. Therefore, within this class of confidence sets, $\hat{C}$ yields the smallest possible confidence intervals for correlations. For instance, elementary calculations using Lagrange multipliers show that

$$\max_{M \in C_{\mathrm{LR}}(I)} \gamma(M) = \sqrt{1 - \exp(-\beta_{\mathrm{LR}})},$$

and it is shown in Section 4 that this can be substantially larger than β.

In addition to *simple correlations* $\rho(v, w \mid \Sigma)$ let us consider other correlation functionals. For a subspace $\mathbf{W}$ of $\mathbf{R}^p$ and $v \in \mathbf{R}^p$ let $v_{\mathbf{W}}$ be the usual orthogonal projection of v onto $\mathbf{W}$, and let $v_{\mathbf{W}\Sigma}$ be the unique minimizer of $\mathbf{W} \ni w \mapsto (v-w)' \Sigma (v-w)$. In other words, $v'_{\mathbf{W}\Sigma}\mathbf{y}_{\mathbf{W}} = v'_{\mathbf{W}\Sigma}\mathbf{y}$ is the best linear predictor of $v'\mathbf{y}$ given $\mathbf{y}_{\mathbf{W}}$ with respect to quadratic loss. For $u, v \in \mathbf{R}^p$, the *partial* correlation of $u'\mathbf{y}$ and $v'\mathbf{y}$ given $\mathbf{y}_{\mathbf{W}}$ equals

$$\rho(u, v, \mathbf{W} \mid \Sigma) := \rho(u - u_{\mathbf{W}\Sigma}, v - v_{\mathbf{W}\Sigma} \mid \Sigma)$$

and the *multiple correlation* of $v'\mathbf{y}$ and $\mathbf{y}_{\mathbf{W}}$ is given by

$$\rho(v, \mathbf{W} \mid \Sigma) := \max_{w \in \mathbf{W}} \rho(v, w \mid \Sigma) = \rho(v, v_{\mathbf{W}\Sigma} \mid \Sigma).$$

Finally, for a second subspace $\mathbf{V}$ of $\mathbf{R}^p$ the *first canonical correlation* of $\mathbf{y}_{\mathbf{V}}$ and $\mathbf{y}_{\mathbf{W}}$ equals

$$\rho(\mathbf{V}, \mathbf{W} \mid \Sigma) := \max_{v \in \mathbf{V} \setminus \{0\},\, w \in \mathbf{W} \setminus \{0\}} \rho(v, w \mid \Sigma).$$

For $1 \leqslant i \leqslant \min\{\dim(\mathbf{V}), \dim(\mathbf{W})\}$, the *ith canonical correlation* of $\mathbf{y}_{\mathbf{V}}$ and $\mathbf{y}_{\mathbf{W}}$ is given by

$$\rho_i(\mathbf{V}, \mathbf{W} \mid \Sigma) := \min_{\mathbf{V}_i} \rho(\mathbf{V}_i, \mathbf{W} \mid \Sigma),$$

where the minimum is taken over all linear subspaces $\mathbf{V}_i$ of $\mathbf{V}$ such that $\dim(\mathbf{V}_i) = \dim(\mathbf{V}) + 1 - i$. This formula is somewhat different from the usual definition of canonical correlations and follows from Rao [15, Theorem 2.2]. Here is our main result for correlation functionals.

THEOREM 1. *Let $R(\Sigma)$ stand for any correlation functional $\rho(* \mid \Sigma)$ defined above, where the first arguments "$*$" are arbitrary and fixed. Then,*

$$\max_{\Gamma \in \hat{C}(S)} R(\Gamma) = m(R(S), \beta)$$

and

$$\min_{\Gamma \in \hat{C}(S)} R(\Gamma) = \begin{cases} m(R(S), -\beta) & \text{for simple and partial correlations,} \\ m(R(S), -\beta)^+ & \text{for multiple and canonical correlations.} \end{cases}$$

Correlations are not the only class of functionals that lead automatically to the critical quantity $\gamma(\Sigma^{-1}S)$. In Lemma 1 one could also consider ratios $v'Mv/w'Mw$ or "regression coefficients" $v'Mw/w'Mw$. Instead of carrying through this program we give a corollary to Theorem 1 about regression vectors. Recall that for any linear subspace $\mathbf{W}$ of $\mathbf{R}^p$ and $v \in \mathbf{R}^p$, the vector $v_{\mathbf{W}\Sigma} \in \mathbf{W}$ minimizes $\mathbb{E}((v'\mathbf{y} - w'\mathbf{y}_{\mathbf{W}})^2)$ over all $w \in \mathbf{R}^p$. Roy [16] constructed confidence ellipsoids for $v_{\mathbf{W}\Sigma}$ for a fixed pair $(v, \mathbf{W})$; see also

Wijsman [20] for extensions and references to related work. The set $\{v_{\mathbf{W}H} \colon H \in \hat{C}(S)\}$ has the same shape as Roy's confidence set. It is larger, because one can treat arbitrary pairs $(v, \mathbf{W})$ simultaneously.

COROLLARY 2. *For any linear subspace* $\mathbf{W}$ *of* $\mathbf{R}^p$ *and* $v \in \mathbf{R}^p \setminus \mathbf{W}$,

$$\{v_{\mathbf{W}\Gamma} \colon \Gamma \in \hat{C}(S)\}$$
$$= \left\{ w \in \mathbf{W} \colon (w - v_{\mathbf{W}S})' \, S(w - v_{\mathbf{W}S}) \leqslant \frac{\beta^2}{1-\beta^2} (v - v_{\mathbf{W}S})' \, S(v - v_{\mathbf{W}S}) \right\}.$$

3. PRINCIPAL COMPONENTS

For a subspace $\mathbf{W}$ of $\mathbf{R}^p$ with orthonormal basis $\{w_1, w_2, ..., w_{\dim(\mathbf{W})}\}$ define

$$\pi(\mathbf{W} \mid \Sigma) := \mathbb{E} \, \|\mathbf{y}_{\mathbf{W}}\|^2 / \mathbb{E} \, \|\mathbf{y}\|^2 = \sum_{i=1}^{\dim(\mathbf{W})} w_i' \Sigma w_i / \mathrm{trace}(\Sigma),$$

where $\mathbf{y}$ is the random vector introduced in Section 2. Thus $\pi(\mathbf{W} \mid \Sigma)$ is the percentage of variability of $\mathbf{y}$ explained by $\mathbf{y}_{\mathbf{W}}$. Throughout we consider spectral representations

$$\Sigma = \sum_{i=1}^{p} \lambda_i(\Sigma) \, \tau_i \tau_i', \qquad S = \sum_{i=1}^{p} \lambda_i(S) \, t_i t_i'$$

with orthonormal bases $\{\tau_1, \tau_2, ..., \tau_p\}$ and $\{t_1, t_2, ..., t_p\}$ of $\mathbf{R}^p$. Then quantities such as

$$\pi_I(\Sigma) := \sum_{i \in I} \lambda_i(\Sigma) / \mathrm{trace}(\Sigma) = \pi(\mathrm{span}\{\tau_i \colon i \in I\} \mid \Sigma), \quad I \subset \{1, 2, ..., p\},$$

are of special interest; see Eaton [8, Proposition 1.44]. One can interpret $\pi_I(S)$ as an estimator for $\pi_I(\Sigma)$ as well as $\pi(\mathrm{span}\{t_i \colon i \in I\} \mid \Sigma)$. In the latter case one takes into account that the principal component vectors τ_i are unknown, too, and $1 - \pi(\mathrm{span}\{t_i \colon i \in I\} \mid \Sigma)$ can be viewed as a relative prediction error conditional on S. The following lemma provides confidence bounds for both points of view.

LEMMA 2. *For arbitrary integers* $1 \leqslant k < \ell \leqslant p$,

$$\max_{\Gamma \in \hat{C}(S)} \frac{\lambda_k}{\lambda_\ell}(\Gamma) = \frac{1+\beta}{1-\beta} \frac{\lambda_k}{\lambda_\ell}(S), \quad \min_{\Gamma \in \hat{C}(S)} \frac{\lambda_k}{\lambda_\ell}(\Gamma) = \max\left\{ \frac{1-\beta}{1+\beta} \frac{\lambda_k}{\lambda_\ell}(S), 1 \right\}.$$

Moreover,

$$m(2\Pi(S)-1, -\beta) \leqslant 2\Pi(\Gamma)-1 \leqslant m(2\Pi(S)-1, \beta) \quad \forall \Gamma \in \hat{C}(S),$$

where $\Pi(\cdot)$ *stands for* $\pi_I(\cdot)$ *or* $\pi(\mathbf{W} \mid \cdot)$. *In the latter case, these bounds are sharp if* $\mathbf{W}$ *is an eigenspace of* S.

Now we investigate special eigenspaces of Σ. For integers $1 \leqslant k \leqslant \ell \leqslant p$ let

$$\mathbf{E}_{k\ell}(\Sigma) := \operatorname{span}\{v \in \mathbf{R}^p \colon \Sigma v = \mu v \text{ for some } \mu \in [\lambda_\ell(\Sigma), \lambda_k(\Sigma)]\}.$$

A natural measure of "distance" from a subspace $\mathbf{V}$ of $\mathbf{R}^p$ to another subspace $\mathbf{W}$ is

$$\max_{v \in \mathbf{V} \cap \mathbf{S}^{p-1}} \|v - v_{\mathbf{W}}\| = \max_{v \in \mathbf{V} \setminus \{0\}} \rho(v, \mathbf{W}^\perp) = \rho(\mathbf{V}, \mathbf{W}^\perp),$$

where $\rho(*) := \rho(* \mid I)$. Therefore, it is of interest to know upper confidence bounds for the numbers $\rho(\mathbf{E}_{1k}(\Sigma), \mathbf{E}_{\ell p}(S))$ and $\rho(\mathbf{E}_{1k}(S), \mathbf{E}_{\ell p}(\Sigma))$, where $1 \leqslant k < \ell \leqslant p$. We define

$$\gamma_{k\ell} := \frac{\lambda_k - \lambda_\ell}{\lambda_k + \lambda_\ell},$$

so that $\gamma = \gamma_{1p}$.

THEOREM 2. *For* $1 \leqslant k < \ell \leqslant p$,

$$\max\{\rho(\mathbf{E}_{1k}(\Gamma), \mathbf{E}_{\ell p}(S)), \rho(\mathbf{E}_{1k}(S), \mathbf{E}_{\ell p}(\Gamma))\} \leqslant f_1(\gamma_{k\ell}(S), \beta) \quad \forall \Gamma \in \hat{C}(S),$$

where

$$f_1(r, \beta) := \frac{1}{\eta + \sqrt{1+\eta^2}}, \quad \eta := \frac{(r/\beta - 1)^+}{\sqrt{1-\beta^2}\sqrt{1-r^2}}.$$

In particular,

$$\max_{\Gamma \in \hat{C}(S)} \rho(\mathbf{E}_{1k}(\Gamma), \mathbf{E}_{k+1,\, p}(S)) = \max_{\Gamma \in \hat{C}(S)} \rho(\mathbf{E}_{1k}(S), \mathbf{E}_{k+1,\, p}(\Gamma)) = f_2(\gamma_{k,\, k+1}(S), \beta),$$

where

$$f_2(r, \beta) := \begin{cases} 2^{-1/2} \sqrt{1 - \sqrt{\dfrac{1-\beta^2/r^2}{1-\beta^2}}}, & \textit{if} \quad r > \beta, \\ 1, & \textit{if} \quad r \leqslant \beta. \end{cases}$$

Both functions f_1, f_2 satisfy

$$\frac{f_i(r, \beta)}{(\beta/2)\sqrt{1/r^2 - 1}} \to 1 \qquad \text{as} \quad r/\beta \to \infty$$

and

$$(\beta/2)\sqrt{1/\gamma_{k\ell}(S)^2 - 1} = \beta \frac{\sqrt{\lambda_k \lambda_\ell}}{\lambda_k - \lambda_\ell}(S),$$

a quantity familiar from asymptotic distributions of eigenvectors. The functions $f_{1,2}(\cdot, 0.3)$ are depicted in Fig. 1.

A possible application of Theorem 2 is testing the hypothesis "$\Sigma v = \lambda_1(\Sigma) v$" for any unit vector v. This hypothesis is to be rejected unless

$$\rho(v, \mathbf{E}_{\ell p}(S))^2 = \sum_{i=\ell}^{p} (t_i' v)^2 \leqslant \min\{f_1(\gamma_{1\ell}(S), \beta)^2, f_2(\gamma_{k,k+1}(S), \beta)^2 \colon 1 \leqslant k < \ell\}$$

for all $1 < \ell \leqslant p$.

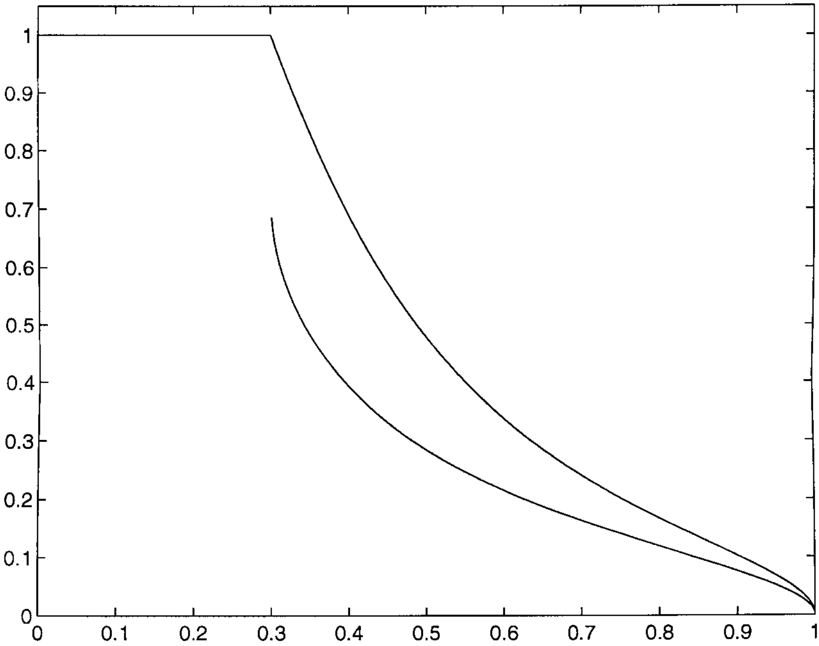

Fig. 1. The functions $f_{1,2}(\cdot, 0.3)$.

4. COMPUTATION AND SIZE OF $\hat{C}$

There is an enormous amount of literature on the distribution of $\lambda(\Sigma^{-1}S)$, and a good starting point is Muirhead [13]. These results can be used to compute the critical value β via numerical integration. Alternatively we computed Monte Carlo estimates of β based on 100,000 simulations; see Table I. We utilized Silverstein's [17] observation that the eigenvalues of $n\Sigma^{-1}S$ are distributed as the eigenvalues of the random tridiagonal matrix

$$W := \begin{pmatrix} Y_1^2 & Y_1 Z_2 & & 0 \\ Y_1 Z_2 & Y_2^2 + Z_2^2 & \ddots & \\ & \ddots & \ddots & Y_{p-1} Z_p \\ 0 & & Y_{p-1} Z_p & Y_p^2 + Z_p^2 \end{pmatrix},$$

where $Y_1, Y_2, ..., Y_p, Z_2, Z_3, ..., Z_p \geqslant 0$ are stochastically independent with $Y_i^2 \sim \chi^2_{n+1-i}$, $Z_j^2 \sim \chi^2_{p+1-j}$. Note, further, that $n^{-1/2}(W - nI)$ converges in distribution to the random matrix

$$B := \begin{pmatrix} X_1 & Z_2 & & 0 \\ Z_2 & X_2 & \ddots & \\ & \ddots & \ddots & Z_p \\ 0 & & Z_p & X_p \end{pmatrix}$$

as n tends to infinity, where $X_1, X_2, ..., X_p, Z_2, Z_3, ..., Z_p$ are independent with $X_i \sim \mathcal{N}(0, 2)$. In particular, $n^{1/2} \log(\lambda(\Sigma^{-1}S))$ converges in distribution to $\lambda(B)$, where $\log(\cdot)$ is defined componentwise. Thus $\tanh(n^{-1/2}q)$ should be a reasonable approximation for β, where q denotes the $(1-\alpha)$-quantile of $(\lambda_1 - \lambda_p)(B)/2$. The last column of Table I ("$n = \infty$") contains Monte Carlo estimates of q, and the resulting approximations $\tanh(n^{-1/2}q)$ for β are given in brackets.

As for the influence of the dimension p on the size of $\hat{C}$, we state a result without proof, which can be obtained by modifying Silverstein's [17] and Trotter's [18] techniques (see also Dümbgen [5]).

LEMMA 3. *As $p \to \infty$ and $p/n \to 0$,*

$$\beta = 2\sqrt{p/n}\,(1 + o(1)), \quad \beta_{\mathrm{LR}} = p^2/(2n)\,(1 + o(1)).$$

Thus

$$1 \leqslant \frac{\max_{M \in C_{\mathrm{LR}}(I)} \gamma(M)}{\max_{M \in \hat{C}(I)} \gamma(M)} = \frac{\sqrt{1 - \exp(-\beta_{\mathrm{LR}})}}{\beta} = \sqrt{p/8}\,(1 + o(1))$$

Table I

Monte-Carlo Estimates of β ($\tanh(n^{-1/2}q)$) and q for $\alpha = 0.1, 0.05$

p	$n=99$	$n=199$	$n=499$	$n=\infty$
2	0.213 (0.212)	0.151 (0.151)	0.096 (0.096)	2.146
	0.242 (0.241)	0.172 (0.172)	0.109 (0.109)	2.448
3	0.360 (0.355)	0.255 (0.256)	0.165 (0.164)	3.695
	0.430 (0.425)	0.310 (0.309)	0.200 (0.199)	4.510
4	0.455 (0.452)	0.330 (0.331)	0.215 (0.214)	4.850
	0.525 (0.522)	0.385 (0.387)	0.250 (0.252)	5.760
5	0.550 (0.543)	0.405 (0.404)	0.265 (0.264)	6.050
	0.610 (0.607)	0.460 (0.459)	0.305 (0.304)	7.005
6	0.605 (0.600)	0.455 (0.454)	0.300 (0.299)	6.900
	0.660 (0.660)	0.505 (0.507)	0.340 (0.339)	7.880
7	0.660 (0.650)	0.500 (0.498)	0.335 (0.332)	7.720
	0.710 (0.703)	0.550 (0.548)	0.370 (0.371)	8.690
8	0.700 (0.688)	0.540 (0.534)	0.360 (0.359)	8.395
	0.745 (0.735)	0.585 (0.580)	0.395 (0.395)	9.340
9	0.730 (0.721)	0.570 (0.566)	0.385 (0.384)	9.045
	0.775 (0.763)	0.615 (0.610)	0.420 (0.420)	9.990
10	0.760 (0.748)	0.600 (0.593)	0.410 (0.406)	9.625
	0.800 (0.786)	0.640 (0.634)	0.440 (0.440)	10.560

as $p \to \infty$ and $p^2/n \to 0$, showing that for high dimension p, the set C_{LR} is substantially "larger" than $\hat{C}$ (cf. Corollary 0).

5. PROOFS

For later reference we recall the minimax representation of eigenvalues of symmetric matrices (cf. [14, Section 1f.2]).

LEMMA 4. (Courant and Fischer). *For any symmetric matrix $M \in \mathbf{R}^{p \times p}$ and $1 \leqslant k \leqslant p$,*

$$\lambda_k(M) = \min_{\dim(\mathbf{V}) = p+1-k} \max_{v \in \mathbf{V} \cap \mathbf{S}^{p-1}} v'Mv,$$

where $\mathbf{V}$ stands for a linear subspace of $\mathbf{R}^p$. ∎

Proof of Lemma 1. Let $\mathbf{V}$ be any two-dimensional linear subspace of $\mathbf{R}^p$. There exists an orthonornmal basis $\{x, y\}$ of $\mathbf{V}$ such that

$$\tilde{M} := \begin{pmatrix} x' \\ y' \end{pmatrix} M(xy) = \operatorname{diag}(\lambda(\tilde{M})) = a \begin{pmatrix} 1+\tilde{\gamma} & 0 \\ 0 & 1-\tilde{\gamma} \end{pmatrix},$$

where $a>0$ and $\tilde{\gamma}:=\gamma(\widetilde{M})$. If v, w are unit vectors in $\mathbf{V}$ such that $v'w=\rho_o\in\,]-1,1[$, then

$$v=\cos(\theta)x+\sin(\theta)y, \quad w=\cos(\theta+\omega)x+\sin(\theta+\omega)y$$

for some $\theta\in[0,2\pi]$ and $\omega:=\arccos(\rho_o)\in\,]0,\pi[$. Repeated application of the addition rule for cosines yields

$$\begin{aligned} v'Mw/a &= \rho_o+\tilde{\gamma}\cos(2\theta+\omega), \\ v'Mv/a &= 1+\rho_o\tilde{\gamma}\cos(2\theta+\omega)+(1-\rho_o^2)^{1/2}\,\tilde{\gamma}\sin(2\theta+\omega), \\ w'Mw/a &= 1+\rho_o\tilde{\gamma}\cos(2\theta+\omega)-(1-\rho_o^2)^{1/2}\,\tilde{\gamma}\sin(2\theta+\omega) \end{aligned}$$

and, after some algebraic manipulations, one obtains

$$\begin{aligned} &\rho(v,w\,|\,M) \\ &\quad=(\rho_o+\tilde{\gamma}\cos(2\theta+\omega))/((\rho_o+\tilde{\gamma}\cos(2\theta+\omega))^2+(1-\rho_o^2)(1-\tilde{\gamma}^2))^{1/2}. \end{aligned}$$

This is a continuous, strictly increasing function of $\cos(2\theta+\omega)$ with extremal values

$$(\rho_o\pm\tilde{\gamma})/((\rho_o\pm\tilde{\gamma})^2+(1-\rho_o^2)(1-\tilde{\gamma}^2))^{1/2}=m(\rho_o,\,\pm\tilde{\gamma}).$$

But Lemma 4 implies that $\tilde{\gamma}\leqslant\gamma(M)$ with equality if $Mx=\lambda_1(M)x$ and $My=\lambda_p(M)y$. ∎

Proof of Corollary 1. First note that $\rho(\cdot)$, $C(\cdot)$ are equivariant in that

$$\begin{aligned} \rho(v,w\,|\,M)&=\rho(rv,sw\,|\,M)=\rho(Av,Aw\,|\,A^{-1\prime}MA^{-1}), \\ C(M)&=A'C(A^{-1\prime}MA^{-1})A \end{aligned}$$

for any $v,w\in\mathbf{R}^p$, $M\in\mathbf{M}^+$, $r,s>0$, and nonsingular $A\in\mathbf{R}^{p\times p}$. Hence, the first half of Corollary 1 follows straightforwardly from Lemma 1. Moreover,

$$\begin{aligned} &\{\rho(v,w\,|\,\Gamma)\colon \Gamma\in C(S)\} \\ &\quad=\{\rho(S^{1/2}v,S^{1/2}w\,|\,M)\colon M\in C(I)\} \\ &\quad=\{\rho(TS^{1/2}v,TS^{1/2}w\,|\,M)\colon M\in C(I),\ T\in\boldsymbol{R}^{p\times p}\text{ orthonormal}\} \\ &\quad=\{\rho(t,u\,|\,M)\colon M\in C(I),\ t,u\in\boldsymbol{S}^{p-1},\ t'u=\rho(v,w\,|\,S)\} \\ &\quad=\bigcup_{M\in C(I)}[m(\rho(v,w\,|\,S),-\gamma(M)),m(\rho(v,w\,|\,S),\gamma(M))]. \end{aligned}$$ ∎

Proof of Theorem 1. At first it is shown that

$$m(R(S), -\beta) \leqslant R(\Gamma) \leqslant m(R(S), \beta)$$

for any correlation functional $R(\cdot)$ and arbitrary fixed $\Gamma \in \hat{C}(S)$. For simple, multiple, and canonical correlations this follows straightforwardly from their definition, Corollary 1, and the monotonicity properties of $m(\cdot, \cdot)$. As for partial correlations, note that

$$Q_\Gamma(u, v) := (u - u_{\mathbf{W}\Gamma})' \, \Gamma(v - v_{\mathbf{W}\Gamma})$$

defines a symmetric bilinear functional on $\mathbf{R}^p$, whose restriction to any subspace $\mathbf{V}$ of $\mathbf{R}^p$ with $\mathbf{V} \cap \mathbf{W} = \{0\}$ is positive definite. Moreover, one can easily deduce from

$$Q_\Gamma(v, v) = \min_{w \in \mathbf{W}} (v - w)' \, \Gamma(v - w)$$

that

$$\lambda_p(S^{-1}\Gamma) \leqslant Q_\Gamma(v, v)/Q_S(v, v) \leqslant \lambda_1(S^{-1}\Gamma) \quad \forall v \in \mathbf{V} \backslash \{0\}.$$

Since $\rho(u, v, \mathbf{W} \mid \Gamma)$ equals $Q_\Gamma(u, w)/(Q_\Gamma(u, u) \, Q_\Gamma(v, v))^{1/2}$, one can apply Corollary 1 to $(Q_\Gamma, Q_S, \mathbf{V})$ in place of $(\Sigma, S, \mathbf{R}^p)$ in order to prove the asserted inequalities for partial correlations.

It remains to be shown that these bounds are sharp. When considering partial correlations $\rho(u, v, \mathbf{W} \mid \cdot)$, equivariance considerations show that one may assume without loss of generality that $S = I$. Further, note that $\rho(u, v, \mathbf{W} \mid \cdot) = \rho(ru - w_1, sv - w_2, \mathbf{W} \mid \cdot)$ for arbitrary $w_1, w_2 \in \mathbf{W}$, $r, s > 0$. Thus we assume that $u, v \in \mathbf{S}^{p-1} \cap \mathbf{W}^\perp$ and $-1 < \rho_o := u'v < 1$. Then there exist orthonormal vectors x, y in $\mathbf{W}^\perp$ such that

$$u = ((1 + \rho_o)/2)^{1/2} \, x + ((1 - \rho_o)/2)^{1/2} \, y,$$

$$v = ((1 + \rho_o)/2)^{1/2} \, x - ((1 - \rho_o)/2)^{1/2} \, y.$$

The matrix $\Gamma := I \pm \beta(xx' - yy')$ belongs to $\hat{C}(I)$, because $\lambda(\Gamma) = (1 + \beta, 1, ..., 1, 1 - \beta)'$. Further, $u'\Gamma w = v'\Gamma w = 0$ for all $w \in \mathbf{W}$, whence

$$\rho(u, v, \mathbf{W} \mid \Gamma) = \rho(u, v \mid \Gamma) = m(\rho_o, \pm\beta).$$

This proves the assertion for partial and simple correlations, where in the latter case $\mathbf{W} = \{0\}$.

Since multiple correlations are a special case of (first) canonical correlations, it suffices to consider $\rho_i(\mathbf{V}, \mathbf{W} \mid \cdot)$. For notational convenience we assume that $\mathbf{V} \cap \mathbf{W} = \{0\}$, the only practically relevant case. Let

$k := \dim(\mathbf{V}) \leqslant \dim(\mathbf{W})$ and $\rho_i := \rho_i(\mathbf{V}, \mathbf{W} \mid S)$. It is well known that there exists a nonsingular matrix $X = (x_1, x_2, ..., x_p) \in \mathbf{R}^{p \times p}$ such that

$$V = \operatorname{span}\{x_1, x_2, ..., x_k\},$$

$$W = \operatorname{span}\{x_{k+1}, x_{k+2}, ..., x_{k+\dim(W)}\},$$

$$X'SX = \begin{pmatrix} I_k & \operatorname{diag}(\rho_1, \rho_2, ..., \rho_k) & 0 \\ \operatorname{diag}(\rho_1, \rho_2, .., \rho_k) & I_k & 0 \\ 0 & 0 & I_{p-2k} \end{pmatrix}.$$

Now we define

$$\Gamma := X^{-1\prime} \begin{pmatrix} \operatorname{diag}((1+\rho_i\beta_i)_{1 \leqslant i \leqslant k}) & \operatorname{diag}((\rho_i+\beta_i)_{1 \leqslant i \leqslant k}) & 0 \\ \operatorname{diag}((\rho_i+\beta_i)_{1 \leqslant i \leqslant k}) & \operatorname{diag}((1+\rho_i\beta_i)_{1 \leqslant i \leqslant k}) & 0 \\ 0 & 0 & I_{p-2k} \end{pmatrix} X^{-1},$$

where $(\beta_i)_{1 \leqslant i \leqslant k}$ equals $(\beta)_{1 \leqslant i \leqslant k}$ or $(-\min\{\rho_i, \beta\})_{1 \leqslant i \leqslant k}$. Then routine calculations show that $\Gamma \in \mathbf{M}^+$ with $(\rho_i(\mathbf{V}, \mathbf{W} \mid \Gamma))_{1 \leqslant i \leqslant k}$ equal to $(m(\rho_i, \beta))_{1 \leqslant i \leqslant k}$ or $(m(\rho_i, -\beta)^+)_{1 \leqslant i \leqslant k}$, respectively. Moreover, $\Gamma \in \hat{C}(S)$, because any eigenvalue of $S^{-1}\Gamma$ equals one or

$$\lambda_{1,2}\left(\begin{pmatrix} 1 & \rho_i \\ \rho_i & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1+\rho_i\beta_i & \rho_i+\beta_i \\ \rho_i+\beta_i & 1+\rho_i\beta_i \end{pmatrix}\right) = 1 \pm \beta_i$$

for some $i \in \{1, 2, ..., k\}$. ∎

Proof of Corollary 2. For $\Gamma \in \hat{C}(S)$ a vector $w \in \mathbf{W}$ equals $v_{\mathbf{W}\Gamma}$ if, and only if, $\rho(w - v, \mathbf{W} \mid \Gamma) = 0$. Together with Theorem 1 it follows that a vector $w \in \mathbf{W}$ belongs to $\{v_{\mathbf{W}\Gamma} \colon \Gamma \in \hat{C}(S)\}$ if, and only if,

$$\rho(w - v, \mathbf{W} \mid S)^2 \leqslant \beta^2.$$

Now the assertion follows from

$$\begin{aligned} &\rho(w - v, \mathbf{W} \mid S)^2 \\ &\quad = \rho((w - v_{\mathbf{W}S}) - (v - v_{\mathbf{W}S}), \mathbf{W} \mid S)^2 \\ &\quad = \max_{x \in \mathbf{W}} \frac{((w - v_{\mathbf{W}S})' Sx)^2 / x'Sx}{(w - v_{\mathbf{W}S})' S(w - v_{\mathbf{W}S}) + (v - v_{\mathbf{W}S})' S(v - v_{\mathbf{W}S})} \\ &\quad = \frac{(w - v_{\mathbf{W}S})' S(w - v_{\mathbf{W}S})}{(w - v_{\mathbf{W}S})' S(w - v_{\mathbf{W}S}) + (v - v_{\mathbf{W}S})' S(v - v_{\mathbf{W}S})}. \end{aligned}$$ ∎

Proof of Lemma 2. By scale-invariance it suffices to consider matrices $\Gamma \in \hat{C}(S)$ such that

$$(1-\beta)\, w'Sw \leqslant w'\Gamma w \leqslant (1+\beta)\, w'Sw \quad \forall w \in \mathbf{R}^p, \tag{5.1}$$

because every point in $\hat{C}(S)$ is a positive multiple of such a matrix. Then it follows directly from Lemma 4 that

$$(1-\beta)\, \lambda_i(S) \leqslant \lambda_i(\Gamma) \leqslant (1+\beta)\, \lambda_i(S) \quad \forall i.$$

This clearly implies the asserted bounds for $(\lambda_k/\lambda_\ell)(\Gamma)$. Moreover,

$$\begin{aligned} 2\pi_I(\Gamma) - 1 &= 2\sum_{i \in I} \lambda_i(\Gamma) \Big/ \Big(\sum_{i \in I} \lambda_i(\Gamma) + \sum_{i \notin I} \lambda_i(\Gamma) \Big)^{-1} \\ &\leqslant 2(1+\beta) \sum_{i \in I} \lambda_i(S) \Big/ \Big((1+\beta) \sum_{i \in I} \lambda_i(S) + (1-\beta) \sum_{i \notin I} \lambda_i(S) \Big) - 1 \\ &= 2(1+\beta)\, \pi_I(S)/((1+\beta)\, \pi_I(S) + (1-\beta)(1-\pi_I)(S)) - 1 \\ &= m(2\pi_I(S) - 1, \beta). \end{aligned}$$

Analogously one can show that $2\pi_I(\Gamma) - 1 \geqslant m(2\pi_I(S) - 1, -\beta)$ and

$$m(2\pi(\mathbf{W} \mid S) - 1, -\beta) \leqslant 2\pi(\mathbf{W} \mid \Gamma) - 1 \leqslant m(2\pi(\mathbf{W} \mid S) - 1, \beta)$$

for any subspace $\mathbf{W}$ of $\mathbf{R}^p$.

It can be easily shown that these bounds for $(\lambda_k/\lambda_\ell)(\Gamma)$ and $\pi(\mathbf{W} \mid \Gamma)$ are sharp (if $\mathbf{W}$ is an eigenspace of S) by considering $\Gamma = \sum_{i=1}^p \mu_i t_i t_i'$ with suitable numbers $(1-\beta)\, \lambda_i(S) \leqslant \mu_i \leqslant (1+\beta)\, \lambda_i(S)$. ∎

Proof of Theorem 2. Suppose that $\gamma_o := \gamma_{k\ell}(S) \leqslant \beta$. Then

$$\Gamma := \lambda_k(S) \sum_{i \in \{k, k+1, ..., \ell\}} t_i t_i' + \sum_{i \notin \{k, k+1, ..., \ell\}} \lambda_i(S)\, t_i t_i'$$

defines a matrix $\Gamma \in \hat{C}(S)$ such that span$\{t_k, t_{k+1}, ..., t_\ell\}$ is contained in $\mathbf{E}_{1k}(\Gamma) \cap \mathbf{E}_{\ell p}(\Gamma)$. Hence $\rho(\mathbf{E}_{1k}(\Gamma), \mathbf{E}_{\ell p}(S)) = \rho(\mathbf{E}_{1k}(S), \mathbf{E}_{\ell p}(\Gamma)) = 1$.

Now suppose that $\gamma_o > \beta$, and let Γ be any fixed point in $\hat{C}(S)$. We derive upper bounds only for $\rho_o := \rho(\mathbf{E}_{1k}(\Gamma), \mathbf{E}_{\ell p}(S))$, because $\rho(\mathbf{E}_{1k}(S), \mathbf{E}_{\ell p}(\Gamma))$ can be treated analogously. Let $\mathbf{V} := \mathbf{E}_{1k}(\Gamma)$, $\mathbf{W} := \mathbf{E}_{\ell p}(S)$, and let $v \in \mathbf{V} \cap \mathbf{S}^{p-1}$, $w \in \mathbf{W} \cap \mathbf{S}^{p-1}$ such that $v'w = \rho_o$. In particular, $w - \rho_o v \in \mathbf{V}^\perp$ and $v - \rho_o w \in \mathbf{W}^\perp$. Since $\rho(\mathbf{V}, \mathbf{V}^\perp \mid \Gamma) = \rho(\mathbf{W}, \mathbf{W}^\perp \mid S) = 0$, this implies that $v'\Gamma w = \rho_o v'\Gamma v$ and $v'Sw = \rho_o w'Sw$. Consequently,

$$\rho(v, w \mid \Gamma) = \rho_o (v'\Gamma v/w'\Gamma w)^{1/2}, \quad \rho(v, w \mid S) = \rho_o\, (w'Sw/v'Sv)^{1/2}.$$

But,

$$\begin{aligned} v'\Gamma v/w'\Gamma w &\geqslant \lambda_k(\Gamma)/(w'Sw\lambda_1(S^{-1}\Gamma)) \\ &\geqslant (\lambda_k(S)\,\lambda_p(S^{-1}\Gamma))/(\lambda_\ell(S)\,\lambda_1(S^{-1}\Gamma)) \\ &\geqslant \kappa := ((1+\gamma_o)(1-\beta))/((1-\gamma_o)(1+\beta)) > 1; \end{aligned}$$

see also the proof of Lemma 2. Analogously, $v'Sv/w'Sw \geqslant \kappa$. Together with Theorem 1, it follows that

$$\kappa^{1/2}\rho_o \leqslant \rho(v, w \mid \Gamma) \leqslant m(\rho(v, w \mid S), \beta) \leqslant m(\kappa^{-1/2}\rho_o, \beta).$$

This leads to the inequality $\beta\rho_o^2 + (\kappa^{1/2} - \kappa^{-1/2})\rho_o \leqslant \beta$, whence

$$\rho_o \leqslant (1+\eta^2)^{1/2} - \eta = \Big((1+\eta^2)^{1/2} + \eta\Big)^{-1},$$

where

$$\eta := (\kappa^{1/2} - \kappa^{-1/2})/(2\beta) = (1-\beta^2)^{-1/2}\,(1-\gamma_o^2)^{-1/2}\,(\gamma_o/\beta - 1).$$

In the special case $\ell = k+1$ this bound can be refined as follows: Let $\bar{\rho}_o := (1-\rho_o^2)^{1/2}$ and

$$u := \bar{\rho}_o^{-1}(v - \rho_o w) \in \mathbf{W}^\perp \cap \mathbf{S}^{p-1} \subset \mathbf{E}_{1k}(S).$$

Then

$$\tilde{S} := \begin{pmatrix} u' \\ w' \end{pmatrix} S(uw) = a\begin{pmatrix} 1+\tilde{\gamma} & 0 \\ 0 & 1-\tilde{\gamma} \end{pmatrix},$$

where $a := (u'Su + w'Sw)/2$ and $\tilde{\gamma} = (u'Su - w'Sw)/(u'Su + w'Sw) \in [\gamma_o, 1[$. With

$$x := \bar{\rho}_o^{-1}(w - \rho_o v) \in \mathbf{V}^\perp \cap \mathbf{S}^{p-1} \subset \mathbf{E}_{k+1,\,p}(\Gamma)$$

one can show that $u = \bar{\rho}_o v - \rho_o x$, $w = \rho_o v + \bar{\rho}_o x$, whence

$$\begin{aligned} \tilde{\Gamma} &:= \begin{pmatrix} u' \\ w' \end{pmatrix} \Gamma(uw) \\ &= v'\Gamma v \begin{pmatrix} \bar{\rho}_o \\ \rho_o \end{pmatrix} (\bar{\rho}_o, \rho_o) + x'\Gamma x \begin{pmatrix} -\bar{\rho}_o \\ \bar{\rho}_o \end{pmatrix} (-\rho_o, \bar{\rho}_o) \\ &= v\begin{pmatrix} 1+rb & rc \\ rc & 1-rb \end{pmatrix}, \end{aligned}$$

where $v := \mathrm{trace}(\tilde{\Gamma})/2$, $r := \gamma(\tilde{\Gamma}) \in \;]0, 1[$, $b := 1 - 2\rho_o^2$ and $c := 2\rho_o \bar{\rho}_o$. Now we seek to minimize b under the side condition

$$\beta \geqslant \gamma(\tilde{S}^{-1}\tilde{\Gamma}) = (1 - (1 - r^2)(1 - \tilde{\gamma}^2)/(1 - \tilde{\gamma} r b)^2)^{1/2}.$$

The smallest b satisfying this inequality is

$$b(r, \tilde{\gamma}) := (1 - (1 - \beta^2)^{-1/2} (1 - \tilde{\gamma}^2)^{1/2} (1 - r^2)^{1/2})/(r\tilde{\gamma}),$$

and elementary calculations show that

$$b(r, \tilde{\gamma}) \geqslant b_o := b(r_o, \gamma_o) = (1 - \beta^2)^{-1/2} (1 - \beta^2/\gamma_o^2)^{1/2},$$

where $r_o := (1 - \beta^2)^{-1/2} (\gamma_o^2 - \beta^2)^{1/2}$. Consequently $\rho_o^2 \leqslant (1 - b_o)/2 = f_2(\gamma_o, \beta)^2$.

This bound is attained if $u = t_k$, $w = t_{k+1}$, and Γ equals

$$\sum_{i \notin \{k, k+1\}} \lambda_i(S)\, t_i t_i' + v((1 + r_o b_o)\, t_k t_k' + r_o (1 - b_o^2)^{1/2} (t_k t_{k+1}' + t_{k+1} t_k') + (1 - r_o b_o)\, t_{k+1} t_{k+1}'),$$

where $v := a(1 - \rho_o^2)(1 - r_o^2)^{-1}$. Verification of this claim is elementary and, therefore, omitted. ∎

REFERENCES

1. Anderson, T. W. (1984). *An Introduction to Multivarite Statistical Analysis, 2nd ed.* Wiley, New York.
2. Beran, R., and Srivastava, M. S. (1985). Bootstrap tests and confidence regions for functions of a covariance matrix. *Ann. Statist.* **13** 95–115.
3. Dümbgen, L. (1993). On nondifferentiable functions and the bootstrap. *Probab. Theory Related Fields* **95** 125–140.
4. Dümbgen, L. (1995). Likelihood ratio tests for principal components. *J. Multivariate Anal.* **52** 245–258.
5. Dümbgen, L. (1996a). *On the Shape and Size of Confidence Sets for High-Dimensional Parameters*. Habilitationsschrift, Universität Heidelberg.
6. Dümbgen, L. (1996b). On Tyler's M-functional of scatter in high dimension. *Ann. Inst. Statist. Math.*, to appear.
7. Eaton, M. L. (1976). A maximization problem and its application to canonical correlation analysis. *J. Multivariate Anal.* **6** 422–425.
8. Eaton, M. L. (1983). *Multivariate Statistics: A Vector Space Approach.* Wiley, New York.
9. Fisher, R. A. (1921). On the "probable error" of a coefficient of correlation deduced from a small sample. *Metron* **1** 3–32.
10. Hayakawa, T. (1987). Normalizing and variance stabilizing transformations of multivariate statistics under an elliptical population. *Ann. Inst. Statist. Math.* **39** 299–306.
11. Jeyaratnam, S. (1992). Confidence intervals for the correlation coefficient. *Statist. Probab. Lett.* **15** 389–393.

12. Kent, J. T., and Tyler, D. E. (1988). Maximum likelihood estimation for the wrapped Cauchy distribution. *J. Appl. Statist.* **15** 247–254.
13. Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York.
14. Rao, C. R. (1973). *Linear Statistical Inference and Its Applications, 2nd ed.* Wiley, New York.
15. Rao, C. R. (1979). Separation theorems for singular values of matrices and their application in multivariate analysis. *J. Multivariate Anal.* **9** 362–377.
16. Roy, S. N. (1957). *Some Aspects of Multivariate Analysis.* Wiley, New York.
17. Silverstein, J. W. (1985). The smallest eigenvalue of a large dimensional Wishart matrix. *Ann. Prob.* **13** 1364–1368.
18. Trotter, H. F. (1984). Eigenvalue distributions of large Hermitian matrices; Wigner's Semi-Circle Law and a theorem of Kac, Murdock and Szegö. *Adv. Math.* **54** 67–82.
19. Tyler, D. E. (1987). A distribution-free *M*-estimator of multivariate scatter. *Ann. Statist.* **15** 234–251.
20. Wijsman, R. A. (1980). Smallest simultaneous confidence sets with applications in multivariate analysis. *Multivariate Analysis-V* (P. R. Krishnaiah, Ed.), North-Holland, Amsterdam.