MINIMAX ESTIMATORS OF A COVARIANCE MATRIX

by

Francois Perron
University of Montreal

Technical Report #90-16

Department of Statistics
Purdue University

March 1990

*

Minimax Estimators of a Covariance Matrix

Francois Perron
University of Montreal

Abstract

Let $S : p \times p$ have a nonsingular Wishart distribution with unknown matrix Σ and n degrees of freedom, $n \geq p$. For estimating Σ a family of minimax estimators, with respect to the entropy loss, is presented. These estimators are of the form $\hat{\Sigma}(S) = R\Phi(L)R^t$ where R is orthogonal, L and Φ are diagonal and $RLR^t = S$. Conditions under which the components of Φ and L follow the same order relation are stated (i.e. writing $L =$ diag $((\ell_1, \ldots, \ell_p)^t)$ and $\Phi =$ diag $((\varphi_1, \ldots, \varphi_p)^t)$ we have $\varphi_1 \geq \ldots \geq \varphi_p$ if and only if $\ell_1 \geq \ldots \geq \ell_p$).

Key words and phrases: combinatoric, convexity, dominate, equivariant, isotonic regression, minimax, risk.

1. Introduction

Let $S: p \times p$ have a nonsingular Wishart distribution with unknown matrix Σ and n degrees of freedom ($S \sim W_p(\Sigma, n)$), $n \geq p$. Consider the problem of estimating Σ using a strictly convex loss function $\mathcal{L}$ invariant with respect to the general linear group of transformations ($G_\ell(p)$). The best equivariant estimator ($\hat{\Sigma}^M$), with respect to the group of lower triangular matrices with positive diagonal elements (G_T^+), is minimax and has constant risk (cf. Kiefer 1957). Moreover, $\hat{\Sigma}^M$ has the form

$$\hat{\Sigma}^M(S) = TDT^t$$

where $T \in G_T^+$, $TT^t = S$, $D = \text{diag}(d)$, $d = (d_1, \ldots, d_p)^t$ and d does not depend on S. The optimal choice for d varies with $\mathcal{L}$. Corresponding to any $G \in G_\ell(p)$ define $G^*\hat{\Sigma}^M$ as

$$(G^*\hat{\Sigma}^M)(S) = G^{-1}\hat{\Sigma}^M(GSG^t)G^{-1^t}.$$

It follows that $G^*\hat{\Sigma}^M$ and $\hat{\Sigma}^M$ have the same risk function. Suppose now that G is randomly distributed independently of S. Define Σ^* as

$$\Sigma^*(S) = E[(G^*\hat{\Sigma}^M)(S)|S].$$

The strict convexity of $\mathcal{L}$ combined with the constant risk of $G^*\hat{\Sigma}^M$ imply that $\hat{\Sigma}^*$ dominates $\hat{\Sigma}^M$ as long as $P_I[\hat{\Sigma}^* = G^*\hat{\Sigma}^M] < 1$ for almost all $G \in G_\ell(p)$. Eaton (1970) made this observation and suggested using the uniform distribution on the orthogonal group ($0(p)$) as a distribution for G. We shall denote the corresponding estimator $\hat{\Sigma}^U$. This estimator is orthogonally equivariant. It has the form

$$\hat{\Sigma}^U(S) = R\Phi(\ell)R^t \tag{1.1}$$

where $R \in 0(p)$, $RLR^t = S$, $L = \text{diag}(\ell)$ with $\ell = (\ell_1, \ldots, \ell_p)^t$, $\ell_1 \geq \ell_2 \geq \ldots \geq \ell_p > 0$ and $\Phi = \text{diag}(\varphi)$ with $\varphi = (\varphi_1, \ldots, \varphi_p)^t$. Sharma and Krishnamoorthy (1983) carried out the computation of φ for $p = 2$. When $p > 2$ no simple expressions of φ are available. Takemura (1984) derived a series expansion for $p = 3$. He also provided the decomposition

$$\varphi(\ell) = L\,W(L)d \tag{1.2}$$

and showed that W is doubly stochastic.

Let $\lambda = (\lambda_1, \ldots, \lambda_p)^t$ with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p > 0$ be the eigenvalues of Σ. Given an orthogonally equivariant estimator $\hat{\Sigma}$ of Σ, $\hat{\Sigma}(S) = R\ \mathrm{diag}(\varphi)R^t$, φ estimates λ. At first hand, since $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$ we should have $\varphi_1 \geq \varphi_2 \geq \ldots \geq \varphi_p$ which we shall call the ordering property. On the other hand, we know (Takemura 1984) that ℓ_1/n overestimate λ_1 and ℓ_p/n underestimate λ_p. This fact suggests having $\Psi_1 \leq \Psi_2 \leq \ldots \leq \Psi_p$ where $\varphi_i = \Psi_i \ell_i$. This requirement has the effect of shrinking the estimates of λ_i towards each other. Let us call it the shrinkage property. For $p = 2$ and $d_1 \leq d_2$ $\hat{\Sigma}^U$ is minimax and satisfies the ordering and the shrinkage properties. When $p > 2$ and $d_1 \leq \ldots \leq d_p$ $\hat{\Sigma}^U$ still minimax and it is conjectured that $\hat{\Sigma}^U$ satisfies the ordering and the shrinkage properties as well. The other orthogonally equivariant minimax estimators in the literature (Stein 1977, Dey and Srinivasan 1985, 1986) do not satisfy the ordering property. Modifying these estimators using isotonic regression is suggested by Stein (cf. Lin and Perlman 1985). However it is not proven that the resulting estimators still minimax. For sure, the modified estimators are inadmissible.

In this article we propose a minimax orthogonally equivariant estimator $\hat{\Sigma}^I$ of Σ which satisfies the ordering and the shrinkage properties. The estimator $\hat{\Sigma}^I$ is based on a representation of $\hat{\Sigma}^U$ derived in section 2. This representation involves expectations of ratios. Replacing expectations of ratios by ratios of expectations leads to $\hat{\Sigma}^I$. $\hat{\Sigma}^I$ has the form

$$\hat{\Sigma}^I(S) = R\ \mathrm{diag}(\varphi^I(\ell))R^t$$

where $\varphi^I(\ell) = L\, W^I(L)d$ and W^I approximates W. The minimaxity, ordering and shrinkage properties of $\hat{\Sigma}^I$ are demonstrated in section 3. Finally, in section 4, a family of estimators is generated from $\hat{\Sigma}^I$. These estimators have the form

$$\hat{\Sigma}^h(S) = R\ \mathrm{diag}(\varphi^h(\ell))R^t$$

where $\varphi^h(\ell) = L\, W^h(L)d,$ h is a function and $W^h(L) = W^I(\ \mathrm{diag}((h(\ell_1), \ldots, h(\ell_p))^t)$.

2. A representation for $\hat{\Sigma}^U$

Suppose that we want to compute $\hat{\Sigma}^U$ explicitly. Since $\hat{\Sigma}^U$ is orthogonally equivariant it is sufficient to evaluate $\hat{\Sigma}^U$ at $L = \text{diag}(\ell)$ with $\ell_1 \geq \ldots \geq \ell_p > 0$. Let $L^{1/2} = \text{diag}((\ell_1^{1/2}, \ldots, \ell_p^{1/2})^t)$. From section 1 we know that $\hat{\Sigma}^U(L) = E[G^*\hat{\Sigma}^M(L)|L]$ where G is uniformly distributed on $0(p)$. We also know that $\hat{\Sigma}^U(L) = \text{diag}(L\,W(L)d)$. By manipulating $E[G^*\hat{\Sigma}^M(L)|L]$ we shall find a representation for W. The result is presented in theorem 2.1. Before proving this result we need the following lemma.

Lemma 2.1. $(G^*\hat{\Sigma}^M)(L) = L^{1/2}\Gamma^t D\Gamma L^{1/2}$ with $GL^{1/2} = U\Gamma,\ U \in G_T^+,\ \Gamma \in 0(p)$.

Proof. $(G^*\hat{\Sigma}^M)(L) = G^{-1}\hat{\Sigma}^M(GLG^t)G^{-1^t}$

$$= L^{1/2}(GL^{1/2})^{-1}\hat{\Sigma}^M((GL^{1/2})(GL^{1/2})^t)(GL^{1/2})^{-1^t}L^{1/2}$$
$$= L^{1/2}\Gamma^t U^{-1}\hat{\Sigma}^M(UU^t)U^{-1^t}\Gamma L^{1/2},\ GL^{1/2} = U\Gamma,\ U \in G_T^+,\ \Gamma \in 0(p)$$
$$= L^{1/2}\Gamma^t D\Gamma L^{1/2}.\ (\hat{\Sigma}^M \text{ is equivariant})$$

□

We need also to introduce new notations and recall the Binet–Cauchy's theorem (cf. Gantmacher 1959). Let

$$\delta_k = \sum_{j=k}^{p} d_j,\ k = 1,\ldots,p,$$

$D_k = \text{diag}\ (1,\ldots,1,0,\ldots,0)$ where $\text{tr}(D_k) = k,\ k = 1,\ldots,p,$

G_k: the $k \times p$ matrix consisting of the k first rows of G, $k = 1,\ldots,p,$

$L_i = \text{diag}\ ((\ell_1,\ldots,\ell_{i-1},\ 0,\ \ell_{i+1},\ldots,\ell_p)^t),\ i = 1,\ldots,p,$

$L_{ij} = L_i + L_j - L,\ i \neq j,\ i,j = 1,\ldots,p.$

Theorem (Binet–Cauchy). If $C = AB$ with $A: p \times q$ and $B: q \times r$ then

$$C\begin{pmatrix} i_1 & \cdots & i_k \\ j_1 & \cdots & j_k \end{pmatrix} = \sum_{1 \leq m_1 < \ldots < m_k \leq q} A\begin{pmatrix} i_1 & \cdots & i_k \\ m_1 & \cdots & m_k \end{pmatrix} B\begin{pmatrix} m_1 & \cdots & m_k \\ j_1 & \cdots & j_k \end{pmatrix}$$

where
$$C\begin{pmatrix} i_1 & \cdots & i_k \\ j_1 & \cdots & j_k \end{pmatrix} = \det\ ((c_{i_m j_n})_{m,n=1,\ldots,k}).$$

Theorem 2.1. $W(L) = EQ(L)$ where $q_{i1} = 1 - \det^{-1}(G_1 L G_1^t) \det(G_1 L_i G_1^t)$, $q_{ik} = \det^{-1}(G_{k-1} L G_{k-1}^t) \det(G_{k-1} L_i G_{k-1}^t) - \det^{-1}(G_k L G_k^t) \det(G_k L_i G_k^t)$ for $i = 1, \ldots, p$, $k = 2, \ldots, p$ and G is uniformly distributed on $0(p)$.

Proof. We have $\hat{\Sigma}^U(L) = E[(G^* \hat{\Sigma}^M)(L)|L] = \operatorname{diag}(\varphi(\ell))$ and $\varphi(\ell) = LW(L)d$. From lemma 2.1 $(G^*\hat{\Sigma}^M)(L) = L^{1/2}\Gamma^t D \Gamma L^{1/2}$ with $GL^{1/2} = U\Gamma$, $U \in G_T^+$, $\Gamma \in 0(p)$. If we express $\Gamma^t D \Gamma$ in terms of D_k we get

$$\Gamma^t D \Gamma = \sum_k \delta_k \Gamma^t D_k \Gamma = \sum_k \delta_k P_k = \sum_k d_k (P_k - P_{k-1})$$

where the P_k matrices are idempotent and correspond to the orthogonal projections of $\mathbb{R}^p$ onto the linear spaces spanned by the rows of $G_k L^{1/2}$, $k = 1, \ldots, p$. Therefore $P_k = L^{1/2} G_k^t (G_k L G_k^t)^{-1} G_k L^{1/2}$ for $k = 1, \ldots, p$ and $P_0 = 0$. Denoting by ρ_{ik} the (i,i) element of P_k the (i,i) element of $(G^*\hat{\Sigma}^M)(L)$ becomes $\ell_i \sum_k d_k(\rho_{ik} - \rho_{ik-1})$ and $q_{ik} = \rho_{ik} - \rho_{ik-1}$. After straightforward computations we get

$$\rho_{ik} = \frac{\sum\limits_{1 \le i_2 < \ldots < i_k \le p} G^2 \begin{pmatrix} 1 & 2 & \ldots & k \\ i & i_2 & \ldots & i_k \end{pmatrix} \ell_i \prod\limits_{j=2}^{k} \ell_{i_j}}{\sum\limits_{1 \le i_1 < \ldots < i_k \le p} G^2 \begin{pmatrix} 1 & 2 & \ldots & k \\ i_1 & i_2 & \ldots & i_k \end{pmatrix} \prod\limits_{j=1}^{k} \ell_{i_j}} \quad i, k = 1, \ldots, p \tag{2.1}$$

$$= 1 - \det^{-1}(G_k L G_k^t) \det(G_k L_i G_k^t). \qquad \square$$

Note that $\rho_{i0} = 0$ and $\rho_{ip} = 1$ for $i = 1, \ldots, p$.

Example: If $p = 2$ then $E\ \rho_{i1} = (\ell_1^{1/2} + \ell_2^{1/2})^{-1}\ \ell_i^{1/2}$.

3. A minimax estimator

In section 2, the difficulty in computing $\hat{\Sigma}^U$ explicitly was due to the fact that we could not find a simple formula for the expectation of ρ_{ij}, $i, j = 1, \ldots, p$. Following expression (2.1) the ρ_{ij} are ratios. In this section we consider a crude approximation of the expectation of ρ_{ij} which consists in taking the ratio of the expectations instead of the expectation of the ratio. Let

$$w_{ik}^I(L) = \operatorname{tr}_{k-1}^{-1}(L)\operatorname{tr}_{k-1}(L_i) - \operatorname{tr}_k^{-1}(L)\operatorname{tr}_k(L_i) \tag{3.1}$$

where

$$\operatorname{tr}_k(L) = \begin{cases} 1 & \text{if } k = 0 \\ \sum\limits_{1 \le i_1 < \ldots < i_k \le p} \prod\limits_{j=1}^{k} \ell_{i_j} & \text{if } k \in \{1, \ldots, p\} \\ 0 & \text{otherwise} \end{cases} \tag{3.2}$$

and set

$$\hat{\Sigma}^I(L) = R \text{ diag } (\varphi^I(\ell))R^t$$

with $\varphi^I(\ell) = L\ W^I(L)d$. Define also Ψ^I as $\Psi^I(\ell) = W^I(L)d$. The approximation W^I turns out to be equal to W when evaluated at $L = cI$, $c > 0$. In this particular case $\hat{\Sigma}^I(cI) = \hat{\Sigma}^U(cI) = c \operatorname{tr}(D)/pI$. As W, W^I is also doubly stochastic.

In the following, the shrinkage, ordering and minimax properties of $\hat{\Sigma}^I$ will be demonstrated. The proofs use some results concerning the function tr_k. These results are technical and are reported to the appendix, section 5.

Theorem 3.1. W^I is doubly stochastic.

Proof. $\sum\limits_i w^I_{ik}(L) = \operatorname{tr}_k^{-1}(L) \sum\limits_i \{\operatorname{tr}_{k-1}(L_i) - \operatorname{tr}_k(L_i)\} = (p-k+1) - (p-k) = 1.$

$$\sum_k w^I_{ik}(L) = \sum_k \{\operatorname{tr}_{k-1}^{-1}(L)\operatorname{tr}_{k-1}(L_i) - \operatorname{tr}_k^{-1}(L)\operatorname{tr}_k(L_i)\} = \operatorname{tr}_0^{-1}(L)\operatorname{tr}_0(L_i) - \operatorname{tr}_p^{-1}(L)\operatorname{tr}_p(L_i) = 1.$$

$$\begin{aligned} w^I_{ik}(L) &= \operatorname{tr}_{k-1}^{-1}(L)\operatorname{tr}_{k-1}(L_i) - \operatorname{tr}_k^{-1}(L)\operatorname{tr}_k(L_i) \\ &= \{\operatorname{tr}_{k-1}(L)\operatorname{tr}_k(L)\}^{-1}\{\operatorname{tr}_{k-1}(L_i)\operatorname{tr}_k(L) - \operatorname{tr}_{k-1}(L)\operatorname{tr}_k(L_i)\} \\ &= \{\operatorname{tr}_{k-1}(L)\operatorname{tr}_k(L)\}^{-1}\ell_i\{\operatorname{tr}^2_{k-1}(L_i) - \operatorname{tr}_k(L_i)\operatorname{tr}_{k-2}(L_i)\} \\ &\ge 0 \text{ by lemma 5.1d).} \end{aligned}$$

□

Theorem 3.2. If $d_1 \le d_2 \le \ldots \le d_p$ then $d_1 \le \Psi^I_1 \le \Psi^I_2 \le \ldots \le \Psi^I_p \le d_p$.

Proof. $\Psi^I_i(\ell) = \sum\limits_k w^I_{ik}(L)d_k = \sum\limits_{k=1}^{p-1} (\operatorname{tr}_k^{-1}(L)\operatorname{tr}_k(L_i))(d_{k+1} - d_k) + d_1$ which is nondecreasing in i by lemma 5.1 c). Moreover, $\Psi^I_i(\ell)$ is a convex combination of $d_1, \ldots, d_p$ therefore $d_1 \le \Psi^I_i \le d_p$ for $i = 1, \ldots, p$.

Theorem 3.3 (ordering property). $\varphi_i^I(\ell) \geq \varphi_j^I(\ell)$ if and only if $\ell_i \geq \ell_j$.

Proof. In order to prove this result we must show that $\ell_i = \ell_j$ implies $\varphi_i^I(\ell) = \varphi_j^I(\ell)$ and also $\ell_i > \ell_j$ implies $\varphi_i^I(\ell) > \varphi_j^I(\ell)$. If $\ell_i = \ell_j$ then $\mathrm{tr}_k(L_i) = \mathrm{tr}_k(L_j)$ for $k = 0, \dots, p$ and $\varphi_i^I(\ell) = \varphi_j^I(\ell)$. If $\ell_i > \ell_j$ then $\ell_i\, w_{i1}^I(L) - \ell_j w_{j1}^I(L) = (\ell_i^2 - \ell_j^2)/\mathrm{tr}(L) > 0$ by lemma 5.2a) and $\ell_i w_{ik}^I(L) - \ell_j w_{jk}^I(L) \geq 0$ for $k = 2, \dots, p$ by lemma 5.2b) therefore $\varphi_i^I(\ell) - \varphi_j^I(\ell) = \sum_k (\ell_i w_{ik}^I(L) - \ell_j w_{jk}^I(L)) d_k > 0$. □

To evaluate the precision of $\hat{\Sigma}^I$ a common loss function used is the entropy loss $\mathcal{L}_0$ given by $\mathcal{L}_0(\Sigma, \hat{\Sigma}) = \mathrm{tr}(\Sigma^{-1}\hat{\Sigma}) - \log(\det(\Sigma^{-1}\hat{\Sigma})) - p$. The optimal choice of D corresponding to $\mathcal{L}_0$ is given by $d_i = (n + p + 1 - 2i)^{-1}$ and the minimax risk is the risk of $\hat{\Sigma}^M$ which is given by

$$\mathcal{R}(\Sigma, \hat{\Sigma}^M) = -\sum_i [\log(d_i) + E\log(\chi^2_{n-i+1})]$$

(cf. James and Stein (1961)). In order to show that $\hat{\Sigma}^I$ is minimax for $\mathcal{L}_0$ we shall use the unbiased estimator $\hat{\mathcal{R}}$ of an orthogonally equivariant estimator proposed by Stein (1977). We shall show that $\hat{\mathcal{R}}(\Sigma, \hat{\Sigma}^I) - \mathcal{R}(\Sigma, \hat{\Sigma}^M) < 0$. Following Stein (1977),

$$\begin{aligned} \hat{\mathcal{R}}(\Sigma, \hat{\Sigma}^I) - \mathcal{R}(\Sigma, \hat{\Sigma}^M) &= (n - p + 1)\sum_i \Psi_i^I(\ell) + \sum\sum_{i \neq j} \frac{\ell_i \Psi_i^I(\ell) - \ell_j \Psi_j^I(\ell)}{\ell_i - \ell_j} - p \\ &+ 2\sum_i \ell_i \frac{\partial}{\partial \ell_i} \Psi_i^I(\ell) - \sum_i [\log(\Psi_i^I(\ell)) - \log(d_i)]. \end{aligned} \tag{3.3}$$

Theorem 3.4. $\hat{\Sigma}^I$ is minimax for $\mathcal{L}_0$.

Proof. Substituting in expression (3.3) and using lemmas 5.1, 5.2 and theorem 3.1 we get

$$\begin{aligned} \hat{\mathcal{R}}(\Sigma, \hat{\Sigma}^I) - \mathcal{R}(\Sigma, \hat{\Sigma}^M) &= (n - p + 1)\sum_k d_k + 2\sum_k (p - k) d_k - p \\ &- 2\sum_{i=1}^{p}\sum_{k=1}^{p-1} \frac{\mathrm{tr}_k(L_i)}{\mathrm{tr}_k(L)}\left(1 - \frac{\mathrm{tr}_k(L_i)}{\mathrm{tr}_k(L)}\right)(d_{k+1} - d_k) - \sum_i \left[\log\left(\sum_k w_{ik}^I(L) d_k\right) - \log(d_i)\right] \\ &= -2\sum_{i=1}^{p}\sum_{k=1}^{p-1} \frac{\mathrm{tr}_k(L_i)}{\mathrm{tr}_k(L)}\left(1 - \frac{\mathrm{tr}_k(L_i)}{\mathrm{tr}_k(L)}\right)(d_{k+1} - d_k) - \sum_i \left[\log\left(\sum_k w_{ik}^I(L) d_k\right) - \sum_k w_{ik}^I(L)\log(d_k)\right] \end{aligned}$$

< 0 by the concavity of the function log and the Jensen inequality. □

4. A family of minimax estimators

Consider a function $h: \mathsf{R}_+ \longrightarrow \mathsf{R}_+$, let $h_i = h(\ell_i)$, $\tilde{h}(\ell) = (h_1(\ell), \ldots, h_p(\ell))^t$ and $H =$ diag$(\tilde{h})$. According to section 3 define $\Psi_i^h = \Psi_i^I \circ \tilde{h}$, $\varphi_i^h = \ell_i \Psi_i^h$, $W^h = W^I \circ \tilde{h}$ and finally

$$\hat{\Sigma}^h(S) = R \text{ diag } (\varphi^h(\ell)) R^t$$

where $R \in 0(p)$, $L =$ diag (ℓ), $\ell_1 \geq \ell_2 \geq \ldots \geq \ell_p$ and $S = RLR^t$. When $p = 2$ and h is given by $h(x) = \sqrt{x}$ we get $\hat{\Sigma}^h = \hat{\Sigma}^U$ and $\hat{\Sigma}^h(cI) = \hat{\Sigma}^U(cI)$ no matter what h and p are. In section 3, the case h equal the identity has been covered. In this section, assuming some regularity conditions, we prove the validity of theorems 3.1 to 3.4 when the superscripts I are replaced by h.

Theorem 4.1. W^h is doubly stochastic.

Theorem 4.2. The relation $d_1 \leq \ldots \leq d_p$ implies $d_1 \leq \Psi_1^h \leq \ldots \leq \Psi_p^h \leq d_p$ if and only if h is nondecreasing.

The proofs of theorems 4.1 and 4.2 are replications of the proofs of theorems 3.1 and 3.2 respectively.

Theorem 4.3. If $xh(x)$ and $x/h(x)$ are nondecreasing in x then $\varphi_i^h(\ell) \geq \varphi_j^h(\ell)$ is equivalent to $\ell_i \geq \ell_j$.

Proof. We have $\ell_i w_{i1}^h(L) - \ell_j w_{j1}^h(L) = (\ell_i h_i - \ell_j h_j)/\text{tr}(H) > 0$ if $\ell_i > \ell_j$. Using lemma 5.3 and reproducing the proof of theorem 3.3 we get the proof of this theorem. □

Theorem 4.4. If h satisfies the relation $\sum\limits_i \ell_i \frac{\partial}{\partial \ell_i} \Psi_i^h(\ell) \leq \sum\limits_{i>j}\sum \frac{(\ell_i h_j - \ell_j h_i)}{(\ell_i - \ell_j)(h_i - h_j)} \; (\Psi_i^h(\ell) - \Psi_j^h(\ell))$ for all ℓ then $\hat{\Sigma}^h$ is minimax for $\mathcal{L}_0$.

Proof. From expression (3.3) we have

$$\begin{aligned}
\hat{\mathcal{R}}(\Sigma, \hat{\Sigma}^h) - \mathcal{R}(\Sigma, \hat{\Sigma}^M) &= (n-p+1)\sum_i \Psi_i^h(\ell) + \sum\sum_{i \neq j} \frac{\ell_i \Psi_i^h(\ell) - \ell_j \Psi_j^h(\ell)}{\ell_i - \ell_j} \\
&\quad - p + 2\sum_i \ell_i \frac{\partial}{\partial \ell_i}\Psi_i^h(\ell) - \sum_i [\log(\Psi_i^h(\ell)) - \log(d_i)] \\
&= (n-p+1)\sum_i \Psi_i^I(\tilde{h}) + \sum\sum_{i \neq j} \frac{h_i \Psi_i^I(\tilde{h}) - h_j \Psi_j^I(\tilde{h})}{h_i - h_j} - p \\
&\quad + \sum\sum_{i \neq j} \frac{(h_i \ell_j - h_j \ell_i)}{(\ell_i - \ell_j)(h_i - h_j)}(\Psi_i^h(\ell) - \Psi_j^h(\ell)) + 2\sum_i \ell_i \frac{\partial}{\partial \ell_i}\Psi_i^h(\ell) \\
&\quad - \sum_i \left[\log\left(\sum_k w_{ik}^h(L) d_k\right) - \sum_k w_{ik}^h(L)\log(d_k)\right] \\
&\leq \sum\sum_{i \neq j} \frac{h_i \ell_j - h_j \ell_i}{(\ell_i - \ell_j)(h_i - h_j)}(\Psi_i^h(\ell) - \Psi_j^h(\ell)) + 2\sum_i \ell_i \frac{\partial}{\partial \ell_i}\Psi_i^h(\ell) \\
&< 0 \text{ by assumptions combined with Jensen's inequality.}
\end{aligned}$$

□

Corollary 4.1. $\hat{\Sigma}^U$ is minimax for $p = 2$.

Proof. If $h(x) = \sqrt{x}$ and $p = 2$ then $\hat{\Sigma}^U = \hat{\Sigma}^h$ and $\sum_i \ell_i \frac{\partial}{\partial \ell_i}\Psi_i^h(\ell) = \frac{(\ell_1 h_2 - \ell_2 h_1)}{(\ell_1 - \ell_2)(h_1 - h_2)}$ $(\Psi_1^h(\ell) - \Psi_2^h(\ell))$. □

Corollary 4.2. If h is differentiable and $h(x)/x$ is nondecreasing in x then $\hat{\Sigma}^h$ is minimax for $\mathcal{L}_0$.

Proof. We have $h' > 0$ by assumptions so

$$h_i \frac{\partial}{\partial \ell_i}\Psi_i^h(\ell) = h_i \frac{\partial}{\partial h_i}\Psi_i^I(\tilde{h})\frac{dh(\ell_i)}{d\ell_i} = -\sum_{k=1}^{p-1} \frac{\mathrm{tr}_k(H_i)}{\mathrm{tr}_k(H)}\left(1 - \frac{\mathrm{tr}_k(H_i)}{\mathrm{tr}_k(H)}\right) h'(\ell_i)(d_{k+1} - d_k) \leq 0.$$

Furthermore, since h is increasing $(\Psi_i^h(\ell) - \Psi_j^h(\ell))/(h_i - h_j) \leq 0$ by theorem 4.2 and $(\ell_i h_j - \ell_j h_i)/(\ell_i - \ell_j) \leq 0$ by assumptions therefore $\sum\sum_{i>j} \frac{(\ell_i h_j - \ell_j h_i)}{(\ell_i - \ell_j)(h_i - h_j)}(\Psi_i^h(\ell) - \Psi_j^h(\ell)) \geq$ $\sum_i \ell_i \frac{\partial}{\partial \ell_i}\Psi_i^h(\ell)$. □

5. Appendix

Let $\Lambda = \text{diag}(\lambda)$ with $\lambda = (\lambda_1, \ldots, \lambda_p)^t$, $\lambda_i \geq 0$, $i = 1, \ldots, p$ and $L = \text{diag}(\ell)$ with $\ell_1 \geq \ldots \geq \ell_p > 0$. Consider also the functions W^I, tr_k defined by expressions 3.1 and 3.2 and W^h defined at the beginning of section 4.

Lemma 5.1.

a) $\sum_i \text{tr}_k(\Lambda_i) = (p-k)\text{tr}_k(\Lambda)$ and $\sum\sum_{i \neq j} \text{tr}_k(\Lambda_{ij}) = (p-k)(p-k-1)\text{tr}_k(\Lambda)$.

b) $\text{tr}_k(\Lambda) = \lambda_i \text{tr}_{k-1}(\Lambda_i) + \text{tr}_k(\Lambda_i)$.

c) $\text{tr}_k(\Lambda_i) - \text{tr}_k(\Lambda_j) = (\lambda_j - \lambda_i)\text{tr}_{k-1}(\Lambda_{ij})$.

d) $\text{tr}_k^2(\Lambda) - \text{tr}_{k-1}(\Lambda)\text{tr}_{k+1}(\Lambda) \geq 0$.

e) $\lambda_i \frac{\partial}{\partial \lambda_i} \frac{\text{tr}_k(\Lambda_i)}{\text{tr}_k(\Lambda)} = -\frac{\text{tr}_k(\Lambda_i)}{\text{tr}_k(\Lambda)}\left(1 - \frac{\text{tr}_k(\Lambda_i)}{\text{tr}_k(\Lambda)}\right)$.

Proof. The proofs of a) and b) are essentially combinatoric and are omitted. The proof of c) is a direct application of b).

d) If $k \notin \{1, \ldots, p-1\}$ the proof is trivial. In order to complete the proof for $k \in \{1, \ldots, p-1\}$ define $A(m_1, m_2)$ and c as $A(m_1, m_2) = \{(\alpha_{11}, \ldots, \alpha_{1m_1}, \alpha_{21}, \ldots, \alpha_{2m_2}) : 1 \leq \alpha_{11} < \ldots < \alpha_{1m_1} \leq p,\ 1 \leq \alpha_{21}, < \ldots < \alpha_{2m_2} \leq p,\ \{\alpha_{11}, \ldots, \alpha_{1m_1}\} \cap \{\alpha_{21}, \ldots, \alpha_{2m_2}\} = \phi\}$ and $c(m) = (m+1)^{-1}\binom{2m}{m}$. By a combinatoric argument we get

$$\text{tr}_k^2(\Lambda) - \text{tr}_{k-1}(\Lambda)\text{tr}_{k+1}(\Lambda) = \sum_{m_2 = 0 \vee 2k-p}^{k} c(k - m_2) \sum_{A(2(k-m_2), m_2)} \prod_{r=1}^{2} \prod_{s=1}^{m_r} \lambda_{\alpha_{rs}}^{r} \quad (5.1)$$

which is always greater or equal to zero.

e) $$\lambda_i \frac{\partial}{\partial \lambda_i} \frac{\text{tr}_k(\Lambda_i)}{\text{tr}_k(\Lambda)} = -\frac{\lambda_i \text{tr}_k(\Lambda_i)\text{tr}_{k-1}(\Lambda_i)}{\text{tr}_k^2(\Lambda)} = -\frac{\text{tr}_k(\Lambda_i)}{\text{tr}_k(\Lambda)}\left(1 - \frac{\text{tr}_k(\Lambda_i)}{\text{tr}_k(\Lambda)}\right). \qquad \square$$

Lemma 5.2.

a) $(\ell_i w_{ik}^I(L) - \ell_j w_{jk}^I(L))/(\ell_i - \ell_j) = \mathrm{tr}_{k-1}^{-1}(L)\mathrm{tr}_{k-1}(L_{ij}) - \mathrm{tr}_k^{-1}(L)\mathrm{tr}_k(L_{ij})$.

b) $(\ell_i w_{ik}^I(L) - \ell_j w_{jk}^I(L))/(\ell_i - \ell_j) \geq 0$.

c) $\sum_{i>j}\sum(\ell_i w_{ik}^I(L) - \ell_j w_{jk}^I(L))/(\ell_i - \ell_j) = (p-k)$.

Proof.

a) $(\ell_i w_{ik}^I(L) - \ell_j w_{jk}^I(L))/(\ell_i - \ell_j)$

$$= \left[\left\{\ell_i \frac{\mathrm{tr}_{k-1}(L_i)}{\mathrm{tr}_{k-1}(L)} - \ell_i \frac{\mathrm{tr}_k(L_i)}{\mathrm{tr}_k(L)}\right\} - \left\{\ell_j \frac{\mathrm{tr}_{k-1}(L_j)}{\mathrm{tr}_{k-1}(L)} - \ell_j \frac{\mathrm{tr}_k(L_j)}{\mathrm{tr}_k(L)}\right\}\right] / (\ell_i - \ell_j)$$

$$= \left\{\frac{\ell_i \mathrm{tr}_{k-1}(L_i) - \ell_j \mathrm{tr}_{k-1}(L_j)}{(\ell_i - \ell_j)\mathrm{tr}_{k-1}(L)}\right\} - \left\{\frac{\ell_i \mathrm{tr}_k(L_i) - \ell_j \mathrm{tr}_k(L_j)}{(\ell_i - \ell_j)\mathrm{tr}_k(L)}\right\}$$

$$= \left\{\frac{\mathrm{tr}_k(L_j) - \mathrm{tr}_k(L_i)}{(\ell_i - \ell_j)\mathrm{tr}_{k-1}(L)}\right\} - \left\{\frac{\mathrm{tr}_{k+1}(L_j) - \mathrm{tr}_{k+1}(L_i)}{(\ell_i - \ell_j)\mathrm{tr}_k(L)}\right\} \text{ by lemma 5.1b)}$$

$$= \frac{\mathrm{tr}_{k-1}(L_{ij})}{\mathrm{tr}_{k-1}(L)} - \frac{\mathrm{tr}_k(L_{ij})}{\mathrm{tr}_k(L)} \text{ by lemma 5.1c).}$$

b) If $k = p$ then $\ell_i w_{ik}^I(L) = \ell_j w_{jk}^I(L)$. For $0 < k < p$

$$\frac{\mathrm{tr}_{k-1}(L_{ij})}{\mathrm{tr}_{k-1}(L)} - \frac{\mathrm{tr}_k(L_{ij})}{\mathrm{tr}_k(L)} = \frac{\mathrm{tr}_{k-1}(L_{ij})\mathrm{tr}_{k-1}(L_i)}{\mathrm{tr}_{k-1}(L_i)\mathrm{tr}_{k-1}(L)} - \frac{\mathrm{tr}_k(L_{ij})\mathrm{tr}_k(L_i)}{\mathrm{tr}_k(L_i)\mathrm{tr}_k(L)}$$

$$= w_{jk}^I(L_i)\frac{\mathrm{tr}_{k-1}(L_i)}{\mathrm{tr}_{k-1}(L)} + w_{ik}^I(L)\frac{\mathrm{tr}_k(L_{ij})}{\mathrm{tr}_k(L_i)} \geq 0.$$

c) The proof of c) is immediate from lemmas 5.1a) and 5.2a). □

Lemma 5.3. If $xh(x)$ and $x/h(x)$ are nondecreasing in x then $(\ell_i w_{ik}^h - \ell_j w_{jk}^h)/(\ell_i - \ell_j) \geq 0$ for $k = 1, \ldots, p$, $\ell_i \neq \ell_j$.

Proof. In order to prove lemma 5.1d) we introduced the sets $A(m_1, m_2)$. Consider the partition of $A(m_1, m_2)$ given by

$$A(m_1, m_2, (i,0)) = \{(\alpha_{11}, \ldots, \alpha_{1m_1}, \alpha_{21}, \ldots, \alpha_{2m_2}) \in A(m_1, m_2) : i \notin \{\alpha_{11}, \ldots, \alpha_{1m_1}, \alpha_{21}, \ldots, \alpha_{2m_2}\}\}$$

$$A(m_1, m_2, (i,1)) = \{(\alpha_{11}, \ldots, \alpha_{1m_1}, \alpha_{21}, \ldots, \alpha_{2m_2}) \in A(m_1, m_2) : i \in \{\alpha_{11}, \ldots, \alpha_{1m_1}\}\}$$

$$A(m_1, m_2, (i,2)) = \{(\alpha_{11}, \ldots, \alpha_{1m_1}, \alpha_{21}, \ldots, \alpha_{2m_2}) \in A(m_1, m_2) : i \in \{\alpha_{21}, \ldots, \alpha_{2m_2}\}\}$$

and let $A(m_1, m_2, (i,r), (j,s)) = A(m_1, m_2, (i,r)) \cap A(m_1, m_2, (j,s))$. Assume that $\ell_i \geq \ell_j$ so $\ell_i h_i^{1-r} \geq \ell_j h_j^{1-r}$ for $r = 0, 1, 2$ by assumptions. From the proof of theorem 3.1 and expression (5.1) we get

$$\{\mathrm{tr}_{k-1}(H)\mathrm{tr}_k(H)\}\ell_i w_{ik}^h(L) = \sum_{m_2=0\vee 2(k-1)-p}^{k-1} c(k-1-m_2) \sum_{A(2(k-1-m_2),m_2,(i,0))} \ell_i h_i \prod_{r=1}^{2}\prod_{s=1}^{m_r} h_{\alpha_{rs}}^r.$$

In order to reduce the notations, suppose that k and m_2 are fixed from now on and let $B(i,r) = A(2(k-1-m_2), m_2, (i,r))$ and $B((i,r),(j,s)) = B(i,r) \cap B(j,s),\ r,s = 0,1,2$. We have

$$\begin{aligned}
\sum_{B(i,0)} \ell_i h_i \prod_{r=1}^{2}\prod_{s=1}^{m_r} h_{\alpha_{rs}}^r &= \sum_{t=0}^{2} \sum_{B((i,0),(j,t))} \ell_i h_i \prod_{r=1}^{2}\prod_{s=1}^{m_r} h_{\alpha_{rs}}^r \\
&= \sum_{t=0}^{2} \sum_{B((j,0),(i,t))} \ell_i h_i^{1-t} h_j^t \prod_{r=1}^{2}\prod_{s=1}^{m_r} h_{\alpha_{rs}}^r \\
&\geq \sum_{t=0}^{2} \sum_{B((j,0),(i,t))} \ell_j h_j \prod_{r=1}^{2}\prod_{s=1}^{m_r} h_{\alpha_{rs}}^r \\
&= \sum_{B(j,0)} \ell_j h_j \prod_{r=1}^{2}\prod_{s=1}^{m_r} h_{\alpha_{rs}}^r
\end{aligned}$$

therefore $\ell_i w_{ik}^h(L) - \ell_j w_{jk}^h(L) \geq 0$. □

References

Dey, D.K. and Srinivasan, C. (1985). Estimation of a covariance matrix under Stein's loss. *Ann. Statist.* **13**, 1581–1591.

Dey, D.K. and Srinivasan, C. (1986). Trimmed minimax estimator of a covariance matrix. *Ann. Inst. Statist. Math.* **38**, 101–108.

Eaton, M.L. (1970). Some problems in covariance estimation. Tech. Report No. 49, Stanford University.

Gantmacher, F.R. (1959). The theory of matrices. Vol. 1, Chelsea, New York.

James, W. and Stein, C. (1961). Estimation with quadratic loss. Proc. Fourth Berkeley Symp. Math. Statist. Prob. I, 361–380, University of California Press.

Kiefer, J. (1957). Invariance minimax sequential estimation and continuous processes. *Ann. Math. Statist.* **25**, 40–75.

Lin, S.P. and Perlman, M.D. (1985). A Monte Carlo comparison of four estimators for a covariance matrix. *Multivariate Analysis* **6**, 411–429, ed. P.R. Krishnaiah, North Holland, Amsterdam.

Sharma, D. and Krishnamoorthy, K. (1983). Orthogonal equivariant estimators of bivariate normal covariance matrix and precision matrix. Calcutta Statistical Association Bulletin, **32**, 23–45.

Stein, C. (1977). Lectures on the theory of estimation of many parameters (in Russian). Studies in the Statistical Theory of Estimation, Part I (Ibragimov, I.A. and Nikulin, M.S. eds.) Proceeding of Scientific Seminars of the Steklov Institute, Leningrad division **74**, 4–65.

Takemura, A. (1984). An orthogonally invariant minimax estimator of the covariance matrix of a multivariate normal population. *Tsukuba J. Math.* **8**, 367–376.

Department of Mathematics and Statistics
University of Montreal
C.P. 6128, succ. A
Montreal, (Quebec) Canada
H3C 3J7