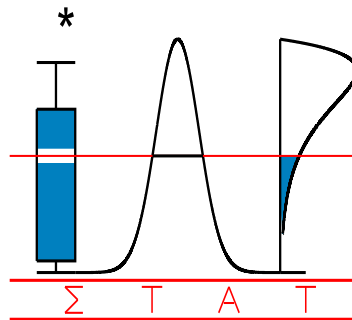


T E C H N I C A L
R E P O R T

0676

A CANONICAL DEFINITION OF SHAPE

D. PAINDAVEINE



I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

A CANONICAL DEFINITION OF SHAPE

Davy PAINDAVEINE*

Université Libre de Bruxelles

Abstract

Very general concepts of scatter, extending the traditional notion of covariance matrices, have become classical tools in robust multivariate analysis. In many problems of practical importance (principal components, canonical correlation, testing for sphericity), only homogeneous functions of the scatter matrix are of interest, and the latter thus is relevant up to a positive scalar factor only. In line with this fact, scatter functionals (such as, e.g., Tyler (1987)'s matrix) often are only defined up to a positive scalar factor, yielding a family of scatter matrices rather than a uniquely defined one. In such families, it is natural to single out one representative by imposing a *normalization constraint*: this normalized scatter is called a *shape matrix*. In the particular case of elliptical families, this constraint in turn induces a concept of *scale*; along with a location center and a *standardized radial density*, the shape and scale parameters entirely characterize an elliptical density. In this paper, we show that, among all possible normalizations, one and only one has the additional properties that (i) the resulting Fisher information matrices for shape and scale, in locally asymptotically normal (LAN) elliptical families, are block-diagonal, and (ii) that the semiparametric elliptical families indexed by location, shape, and completely unspecified radial densities are adaptive. This particular normalization, which consists in imposing that the determinant of the shape matrix be equal to one, therefore can be considered canonical.

Key words and phrases: Elliptic densities; Scatter matrix; Shape matrix; Local asymptotic normality; Semiparametric efficiency; Adaptivity

*Davy Paindaveine, Institut de Recherche en Statistique, E.C.A.R.E.S, and Département de Mathématique, Université Libre de Bruxelles, Campus de la Plaine, CP 210, B-1050 Bruxelles, Belgium

1 Introduction.

The multivariate concepts of location and scatter, extending to the multivariate context the traditional concepts of location and scale are generally characterized via their behavior under affine transformations of the observation space. More precisely, denoting by \mathbf{X} a k -variate random vector with probability distribution $P^{\mathbf{X}}$, consider a couple $(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ of functionals defined over $\{P^{\mathbf{A}\mathbf{X}+\mathbf{b}} : \mathbf{A}$ an invertible $k \times k$ real matrix, $\mathbf{b} \in \mathbb{R}^k\}$ mapping $P^{\mathbf{X}}$ onto $(\boldsymbol{\theta}^{\mathbf{X}}, \boldsymbol{\Sigma}^{\mathbf{X}}) \in \mathbb{R}^k \times \mathcal{S}_k$, where \mathcal{S}_k denotes the set of symmetric positive definite real $k \times k$ matrices—throughout, $k \geq 2$. This couple is called a location-scatter functional iff

$$(\boldsymbol{\theta}^{\mathbf{A}\mathbf{X}+\mathbf{b}}, \boldsymbol{\Sigma}^{\mathbf{A}\mathbf{X}+\mathbf{b}}) = (\mathbf{A}\boldsymbol{\theta}^{\mathbf{X}} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}^{\mathbf{X}}\mathbf{A}') \quad (1)$$

for any invertible $k \times k$ matrix \mathbf{A} and any $\mathbf{b} \in \mathbb{R}^k$. The traditional example of such a couple of course is the mean and the covariance matrix, but many other solutions exist, and the problem of defining robust counterparts to means and covariances has generated a huge literature which we do not attempt to review here—see Maronna et al. (2006) or Zuo (2006) for recent surveys.

In many problems in multivariate analysis, it is sufficient to know—or to estimate—*normalized* versions of scatter matrices to be able to perform the analysis (see below). In line with this fact, scatter matrices often are only defined up to a positive factor—see, for instance, Tyler (1983, 1987). In such families of scatter matrices, it is natural to pick out one representative by imposing a *normalization constraint*. More specifically, let $S : \mathcal{S}_k \rightarrow \mathbb{R}_0^+$ be a *homogeneous* function—i.e., satisfying $S(\lambda\boldsymbol{\Sigma}) = \lambda S(\boldsymbol{\Sigma})$ for all $\lambda > 0$ —and define $\mathcal{V}_k^S := \{\mathbf{V} \in \mathcal{S}_k : S(\mathbf{V}) = 1\}$: the elements of \mathcal{V}_k^S are called *shape matrices*, and $\mathbf{V}_S^{\mathbf{X}} := \boldsymbol{\Sigma}^{\mathbf{X}}/S(\boldsymbol{\Sigma}^{\mathbf{X}})$ is called the *shape matrix of \mathbf{X}* .

The choice of S is arbitrary and, to some extent, inessential (see the comments below). Classical choices include

- (i) $S(\boldsymbol{\Sigma}) = \Sigma_{11}$ (Hallin and Paindaveine (2006), Hallin et al. (2006), Hettmansperger and Randles (2002), and Randles (2000)),

- (ii) $S(\boldsymbol{\Sigma}) = (\text{tr } \boldsymbol{\Sigma})/k$ (Dümbgen (1998), Ollila et al. (2004), and Tyler (1987)), and
- (iii) $S(\boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{1/k}$ (Dümbgen and Tyler (2005), Salibian-Barrera et al. (2006), Taskinen et al. (2006), and Tatsuoka and Tyler (2000)).

Now consider the particular case of a k -variate *elliptical* random vector \mathbf{X} , that is, letting $d(\mathbf{x}, \boldsymbol{\theta}; \boldsymbol{\Sigma}) := ((\mathbf{x} - \boldsymbol{\theta})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\theta}))^{1/2}$ and denoting by $|\mathbf{M}|$ the determinant of the square matrix \mathbf{M} , assume that $P^{\mathbf{X}}$ admits the density

$$\mathbf{x} \mapsto \frac{c_{k,f_1}}{|\boldsymbol{\Sigma}|^{1/2}} f_1(d(\mathbf{x}, \boldsymbol{\theta}; \boldsymbol{\Sigma})), \quad (2)$$

where $\boldsymbol{\theta}$, the *center of symmetry*, is a k -dimensional real vector, $\boldsymbol{\Sigma}$ belongs to \mathcal{S}_k , and $f_1 : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$, the *standardized radial density* is such that $\mu_{k-1,f_1} := \int_0^\infty r^{k-1} f_1(r) dr < \infty$ (c_{k,f_1} is a normalization factor). To ensure identifiability of $\boldsymbol{\Sigma}$ and $c_{k,f_1} \times f_1$ without imposing any moment conditions, the pdf of $d(\mathbf{X}, \boldsymbol{\theta}; \boldsymbol{\Sigma})$ under (2) (that is, $r \mapsto \tilde{f}_{1k}(r) := (\mu_{k-1,f_1})^{-1} r^{k-1} f_1(r) I_{[r>0]}$) is assumed to have median one.

Under this elliptical setting, the only solutions of (1) are the couples $(\boldsymbol{\theta}, \lambda \boldsymbol{\Sigma})$, with arbitrary $\lambda > 0$. It follows that the shape $\mathbf{V}_S = \boldsymbol{\Sigma}/S(\boldsymbol{\Sigma})$ is uniquely defined for any homogeneous function S , and that $\sigma_S := (S(\boldsymbol{\Sigma}))^{1/2}$, as the median of $d(\mathbf{X}, \boldsymbol{\theta}; \mathbf{V}_S)$, has the interpretation of a *scale parameter*. This allows for rewriting (2) as

$$\mathbf{x} \mapsto \frac{c_{k,f_1}}{\sigma_S^k |\mathbf{V}_S|^{1/2}} f_1 \left(\frac{1}{\sigma_S} d(\mathbf{x}, \boldsymbol{\theta}; \mathbf{V}_S) \right) = \frac{c_{k,f}}{|\mathbf{V}_S|^{1/2}} f(d(\mathbf{x}, \boldsymbol{\theta}; \mathbf{V}_S)). \quad (3)$$

This latter density is indexed by $\boldsymbol{\theta}$, \mathbf{V}_S , and the *(non-standardized) radial density* f . Under finite second-order moments, of course, $\boldsymbol{\Sigma}$ reduces to a multiple of $\text{Cov}[\mathbf{X}]$, and hence $\mathbf{V}_S = \text{Cov}[\mathbf{X}]/S(\text{Cov}[\mathbf{X}])$.

Whatever the choice of S , the shape matrix \mathbf{V}_S is a parameter of primary interest in a number of standard problems in multivariate analysis. Principal component analysis (PCA), canonical correlation analysis (CCA), and the problem of testing for sphericity, among others, only depend on shape—rather than on scatter or covariance matrices; see, for instance, Croux and Haesbroeck (2000), Hallin and Paindaveine (2006), and Taskinen et al. (2006). Inference on shape is thus an essential issue in the domain.

The choice of S so far is still arbitrary. The objective of this paper is to show that decision-theoretic arguments, involving the structure of Fisher information and semiparametric efficiency, strongly suggest adopting the determinant-based normalization $S(\boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{1/k}$. This particular choice indeed is the only one for which

- (a) the Fisher information matrices for scale and shape, in locally asymptotically normal (LAN) elliptical families, are block-diagonal, and
- (b) the semiparametric elliptical families indexed by location $\boldsymbol{\theta}$, shape \mathbf{V}_S , and completely unspecified radial densities f (see (3)) are adaptive; this adaptivity result is much stronger than the one established in Bickel (1982) (see Section 2 for a discussion).

These two properties considerably simplify the structure of information, and in principle allow for parametrically efficient inference for shape, under unspecified $(\boldsymbol{\theta}, f)$ (equivalently, unspecified $(\boldsymbol{\theta}, \sigma_S, f_1)$). The determinant-based concepts of shape and scale therefore can be considered canonical.

The paper is organized as follows. In Section 2, we list and discuss the assumptions that are needed in the sequel, and state, for an arbitrary normalization S , the local and asymptotic normality (LAN) property of elliptical families. Section 3 states the main result of the paper and discusses its implications. Finally, the proofs are given in Section 4.

2 Assumptions, notation, and local asymptotic normality.

The following notation will be used. For any $k \times k$ matrix \mathbf{A} , let $\text{vec } \mathbf{A}$ be the k^2 -vector resulting from stacking the columns of \mathbf{A} on top of each other. If \mathbf{A} moreover is symmetric, write $\text{vech } \mathbf{A} := (A_{11}, (\overset{\circ}{\text{vech}} \mathbf{A})')'$ for the $(K + 1)$ -vector (throughout, $K = k(k + 1)/2 - 1$) obtained by stacking the upper-triangular elements of $\mathbf{A} = (A_{ij})$: $\overset{\circ}{\text{vech}} \mathbf{A}$ thus stands for $\text{vech } \mathbf{A}$ deprived of its first component A_{11} . On the scale functional S we make the following assumption.

ASSUMPTION (A1). *The scale functional $S : \mathcal{S}_k \rightarrow \mathbb{R}_0^+$ (i) is homogeneous (see Section 1), (ii) is differentiable, with $\frac{\partial S}{\partial \Sigma_{11}}(\boldsymbol{\Sigma}) \neq 0$ for all $\boldsymbol{\Sigma} \in \mathcal{S}_k$, and (iii) satisfies $S(\mathbf{I}_k) = 1$, where \mathbf{I}_k denotes the k -dimensional identity matrix.*

Clearly, one can also look at $\boldsymbol{\Sigma} \mapsto S(\boldsymbol{\Sigma})$ as a function of $\text{vech } \boldsymbol{\Sigma}$: with a slight abuse of notation, we indifferently write $S(\boldsymbol{\Sigma})$ or $S(\text{vech } \boldsymbol{\Sigma})$ in the sequel, and denote by $\nabla S(\text{vech } \boldsymbol{\Sigma})$ the gradient $\text{grad}_{\text{vech } \boldsymbol{\Sigma}} S(\text{vech } \boldsymbol{\Sigma})$. Under Assumption (A1), $(\mathbf{V}_S)_{11}$, for any $\mathbf{V}_S \in \mathcal{V}_k^S$, can be recovered from $(\text{vech } \mathbf{V}_S)$. Also note that the special role of Σ_{11} in Assumption (A1) could have been played by any other entry of $\boldsymbol{\Sigma}$. Assuming that some other component of ∇S is non-zero would allow, for instance, for dealing with scale functionals such as $S(\boldsymbol{\Sigma}) = \Sigma_{12}$ or $S(\boldsymbol{\Sigma}) = (\prod_{i=2}^k \Sigma_{ii})^{1/(k-1)}$ —with appropriate redefinition of the vech operator. As the extension of our results to such cases is trivial, we stick to Assumption (A1) in the sequel.

Denote by $P_{\boldsymbol{\theta}, \boldsymbol{\Sigma}, f_1}^n$ or equivalently (for any given S) $P_{\boldsymbol{\theta}, \sigma_S^2, \mathbf{V}_S, f_1}^n$ the distribution of an i.i.d. n -tuple $(\mathbf{X}_1, \dots, \mathbf{X}_n)$ with density (2) or (3). For given S satisfying Assumption (A1), the scatter parameter $\boldsymbol{\Sigma}$ (in vector form, $\text{vech } \boldsymbol{\Sigma}$) decomposes into scale and shape parameters through $\boldsymbol{\Sigma} = \sigma_S^2 \mathbf{V}_S$, where $\sigma_S^2 = S(\boldsymbol{\Sigma})$ and $\mathbf{V}_S := \boldsymbol{\Sigma}/S(\boldsymbol{\Sigma}) \in \mathcal{V}_k^S$. In vector form, dropping $(\mathbf{V}_S)_{11}$, the new parameter is thus $\boldsymbol{\vartheta}_S := (\boldsymbol{\theta}', \sigma_S^2, (\text{vech } \mathbf{V}_S)')' \in \boldsymbol{\Theta}_S (:= \mathbb{R}^k \times \mathbb{R}_0^+ \times \text{vech } \mathcal{V}_k^S)$. Theorem 2.1 below states that, under mild regularity conditions on f_1 , the families of distributions $\mathcal{P}_{S; f_1}^n := \{P_{\boldsymbol{\vartheta}_S, f_1}^n : \boldsymbol{\vartheta}_S \in \boldsymbol{\Theta}_S\}$ are locally asymptotically normal (LAN; see Le Cam (1986)). This theorem extends to an arbitrary scale functional S the result obtained for $S(\boldsymbol{\Sigma}) = \Sigma_{11}$ in Hallin and Paindaveine (2006), where minimal assumptions are given; here, for the sake of simplicity, we rather provide the following sufficient one.

ASSUMPTION (A2). *The standardized radial density f_1 belongs to the collection \mathcal{F} of absolutely continuous functions, with a.e.-derivative \dot{f}_1 , and, letting $\varphi_{f_1} := -\dot{f}_1/f_1$, the quantities $\mathcal{I}_k(f_1) := \int_0^\infty (\varphi_{f_1}(r))^2 \tilde{f}_{1k}(r) dr$ and $\mathcal{J}_k(f_1) := \int_0^\infty r^2 (\varphi_{f_1}(r))^2 \tilde{f}_{1k}(r) dr$ are finite.*

The finiteness of the radial Fisher informations for location $\mathcal{I}_k(f_1)$ and scale $\mathcal{J}_k(f_1)$ does

not imply any moment conditions. Hence, Assumption (A2) is extremely mild and turns out to be satisfied at Gaussian densities as well as at all Student and power-exponential densities (see Hallin and Paindaveine 2006).

The following notation is needed in the statement of LAN. Denoting by \mathbf{e}_ℓ the ℓ th vector of the canonical basis of \mathbb{R}^k , let $\mathbf{K}_k := \sum_{i,j=1}^k (\mathbf{e}_i \mathbf{e}_j') \otimes (\mathbf{e}_j \mathbf{e}_i')$ be the $k^2 \times k^2$ commutation matrix, and put $\mathbf{J}_k := (\text{vec } \mathbf{I}_k)(\text{vec } \mathbf{I}_k)'$. Write also $\mathbf{A}^{\otimes 2}$ for the Kronecker product $\mathbf{A} \otimes \mathbf{A}$. Finally, for any $\boldsymbol{\Sigma} \in \mathcal{S}_k$ and S satisfying (A1), let $\mathbf{M}_S^\boldsymbol{\Sigma} := \mathbf{M}_{S,k}^\boldsymbol{\Sigma}$ be the $K \times k^2$ matrix such that $(\mathbf{M}_S^\boldsymbol{\Sigma})'(\text{vech } \mathbf{v}) = \text{vec } \mathbf{v}$ for any symmetric $k \times k$ matrix \mathbf{v} satisfying $(\nabla S(\text{vech } \boldsymbol{\Sigma}))'(\text{vech } \mathbf{v}) = 0$. Note that $(\nabla S(\text{vech } \boldsymbol{\Sigma}))'(\text{vech } \mathbf{v}) = 0$ holds for $S(\boldsymbol{\Sigma}) = \Sigma_{11}$, $S(\boldsymbol{\Sigma}) = (\text{tr } \boldsymbol{\Sigma})/k$, and $S(\boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{1/k}$, iff $(\mathbf{v})_{11} = 0$, $\text{tr } \mathbf{v} = 0$, and $\text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{v}) = 0$, respectively.

Theorem 2.1 Under Assumptions (A1) and (A2), the family $\mathcal{P}_{S,f_1}^n = \{P_{\boldsymbol{\vartheta}_S, f_1}^n : \boldsymbol{\vartheta}_S \in \boldsymbol{\Theta}_S\}$ is LAN. More precisely, for any $\boldsymbol{\vartheta}_S = (\boldsymbol{\theta}', \sigma_S^2, (\text{vech } \mathbf{V}_S)')'$ and any bounded sequence $\boldsymbol{\tau}_n \in \mathbb{R}^{k+K+1}$, we have that (i) under $P_{\boldsymbol{\vartheta}_S, f_1}^n$,

$$\log \left(dP_{\boldsymbol{\vartheta}_S + n^{-1/2} \boldsymbol{\tau}_n, f_1}^n / dP_{\boldsymbol{\vartheta}_S, f_1}^n \right) = \boldsymbol{\tau}_n' \boldsymbol{\Delta}_{\boldsymbol{\vartheta}_S, f_1}^n - \frac{1}{2} \boldsymbol{\tau}_n' \boldsymbol{\Gamma}_{\boldsymbol{\vartheta}_S, f_1} \boldsymbol{\tau}_n + o_P(1),$$

where, letting $d_i := d(\mathbf{X}_i, \boldsymbol{\theta}; \mathbf{V}_S)$ and $\mathbf{U}_i := \mathbf{V}_S^{-1/2}(\mathbf{X}_i - \boldsymbol{\theta})/d_i$ (throughout, $\mathbf{V}_S^{1/2}$ is taken symmetric), $\boldsymbol{\Delta}_{\boldsymbol{\vartheta}_S, f_1}^n := ((\boldsymbol{\Delta}_{\boldsymbol{\vartheta}_S, f_1; 1}^n)', \boldsymbol{\Delta}_{\boldsymbol{\vartheta}_S, f_1; 2}^n, \boldsymbol{\Delta}_{\boldsymbol{\vartheta}_S, f_1; 3}^n)'$, with

$$\begin{aligned} \boldsymbol{\Delta}_{\boldsymbol{\vartheta}_S, f_1; 1}^n &:= \frac{1}{\sigma_S \sqrt{n}} \sum_{i=1}^n \varphi_{f_1} \left(\frac{d_i}{\sigma_S} \right) \mathbf{V}_S^{-1/2} \mathbf{U}_i, \\ \boldsymbol{\Delta}_{\boldsymbol{\vartheta}_S, f_1; 2}^n &:= \frac{1}{2\sigma_S^2 \sqrt{n}} \sum_{i=1}^n \left(\varphi_{f_1} \left(\frac{d_i}{\sigma_S} \right) \frac{d_i}{\sigma_S} - k \right), \end{aligned} \quad (4)$$

and

$$\boldsymbol{\Delta}_{\boldsymbol{\vartheta}_S, f_1; 3}^n := \frac{1}{2\sqrt{n}} \mathbf{M}_S^{\mathbf{V}_S} (\mathbf{V}_S^{\otimes 2})^{-1/2} \sum_{i=1}^n \text{vec} \left(\varphi_{f_1} \left(\frac{d_i}{\sigma_S} \right) \frac{d_i}{\sigma_S} \mathbf{U}_i \mathbf{U}_i' - \mathbf{I}_k \right), \quad (5)$$

and that (ii) the central sequence $\boldsymbol{\Delta}_{\boldsymbol{\vartheta}_S, f_1}^n$, still under $P_{\boldsymbol{\vartheta}_S, f_1}^n$, is asymptotically normal with mean zero and covariance matrix

$$\boldsymbol{\Gamma}_{\boldsymbol{\vartheta}_S, f_1} := \begin{pmatrix} \boldsymbol{\Gamma}_{\boldsymbol{\vartheta}_S, f_1; 11} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_{\boldsymbol{\vartheta}_S, f_1; 22} & \boldsymbol{\Gamma}'_{\boldsymbol{\vartheta}_S, f_1; 32} \\ \mathbf{0} & \boldsymbol{\Gamma}_{\boldsymbol{\vartheta}_S, f_1; 32} & \boldsymbol{\Gamma}_{\boldsymbol{\vartheta}_S, f_1; 33} \end{pmatrix}, \quad (6)$$

with

$$\begin{aligned}\mathbf{\Gamma}_{\boldsymbol{\vartheta}_S, f_1; 11} &:= \frac{\mathcal{I}_k(f_1)}{k\sigma_S^2} \mathbf{V}_S^{-1}, \\ \mathbf{\Gamma}_{\boldsymbol{\vartheta}_S, f_1; 22} &:= \frac{\mathcal{J}_k(f_1) - k^2}{4\sigma_S^4}, \quad \mathbf{\Gamma}_{\boldsymbol{\vartheta}_S, f_1; 32} := \frac{\mathcal{J}_k(f_1) - k^2}{4k\sigma_S^2} \mathbf{M}_S^{\mathbf{V}_S} (\text{vec } \mathbf{V}_S^{-1}),\end{aligned}$$

and

$$\mathbf{\Gamma}_{\boldsymbol{\vartheta}_S, f_1; 33} := \frac{1}{4} \mathbf{M}_S^{\mathbf{V}_S} (\mathbf{V}_S^{\otimes 2})^{-1/2} \left[\frac{\mathcal{J}_k(f_1)}{k(k+2)} (\mathbf{I}_{k^2} + \mathbf{K}_k + \mathbf{J}_k) - \mathbf{J}_k \right] (\mathbf{V}_S^{\otimes 2})^{-1/2} (\mathbf{M}_S^{\mathbf{V}_S})'. \quad (7)$$

The block-diagonal structure of the information matrix (6) implies that the non-specification of the location centre $\boldsymbol{\theta}$ does not affect optimal parametric performances when estimating \mathbf{V}_S and/or σ_S^2 , or when performing tests about the same; more precisely, when estimating \mathbf{V}_S for instance, the optimal asymptotic covariance matrix that can be achieved (at $P_{\boldsymbol{\theta}, \sigma_S^2, \mathbf{V}_S, f_1}^n$) by an estimator of \mathbf{V}_S is the same in $\mathcal{P}_{S; \sigma_S^2, f_1}^n := \{P_{\boldsymbol{\theta}, \sigma_S^2, \mathbf{V}_S, f_1}^n : \boldsymbol{\theta} \in \mathbb{R}^k, \mathbf{V}_S \in \mathcal{V}_k^S\}$ as in $\mathcal{P}_{S; \boldsymbol{\theta}, \sigma_S^2, f_1}^n := \{P_{\boldsymbol{\theta}, \sigma_S^2, \mathbf{V}_S, f_1}^n : \mathbf{V}_S \in \mathcal{V}_k^S\}$, where $\boldsymbol{\theta}$ is specified, and is actually given by $(\mathbf{\Gamma}_{\boldsymbol{\vartheta}_S, f_1; 33})^{-1}$. Since the latter does only depend on \mathbf{V}_S and f_1 , so does this optimal performance.

On the contrary, the non-zero covariance between the scale and shape parts of the central sequences implies that, when estimating \mathbf{V}_S , the non-specification of σ_S^2 affects the optimal parametric performance at f_1 . The latter actually is given by the f_1 -efficient Fisher information for shape

$$\begin{aligned}\mathbf{\Gamma}_{\boldsymbol{\vartheta}_S, f_1; 33}^* &:= \mathbf{\Gamma}_{\boldsymbol{\vartheta}_S, f_1; 33} - \mathbf{\Gamma}_{\boldsymbol{\vartheta}_S, f_1; 32} \mathbf{\Gamma}_{\boldsymbol{\vartheta}_S, f_1; 22}^{-1} \mathbf{\Gamma}_{\boldsymbol{\vartheta}_S, f_1; 32}' \\ &= \frac{\mathcal{J}_k(f_1)}{4k(k+2)} \mathbf{M}_S^{\mathbf{V}_S} (\mathbf{V}_S^{\otimes 2})^{-1/2} \left[\mathbf{I}_{k^2} + \mathbf{K}_k - \frac{2}{k} \mathbf{J}_k \right] (\mathbf{V}_S^{\otimes 2})^{-1/2} (\mathbf{M}_S^{\mathbf{V}_S})', \quad (8)\end{aligned}$$

that is, the asymptotic covariance matrix, under $P_{\boldsymbol{\vartheta}_S, f_1}^n$, of the f_1 -efficient central sequence for shape

$$\begin{aligned}\Delta_{\boldsymbol{\vartheta}_S, f_1; 3}^{*n} &:= \Delta_{\boldsymbol{\vartheta}_S, f_1; 3}^n - \mathbf{\Gamma}_{\boldsymbol{\vartheta}_S, f_1; 32} \mathbf{\Gamma}_{\boldsymbol{\vartheta}_S, f_1; 22}^{-1} \Delta_{\boldsymbol{\vartheta}_S, f_1; 2}^n \\ &= \frac{1}{2\sqrt{n}} \mathbf{M}_S^{\mathbf{V}_S} (\mathbf{V}_S^{\otimes 2})^{-1/2} \sum_{i=1}^n \varphi_{f_1} \left(\frac{d_i}{\sigma_S} \right) \frac{d_i}{\sigma_S} \text{vec} \left(\mathbf{U}_i \mathbf{U}_i' - \frac{1}{k} \mathbf{I}_k \right); \end{aligned}$$

as explained in a general parametric setup in, e.g, section 2.4 of Bickel et al. (1993) locally optimal inference on shape—under unspecified σ_S^2 —should be based on $\Delta_{\boldsymbol{\vartheta}_S, f_1; 3}^{*n}$.

Note that, unlike the original central sequence for shape in (5), $\Delta_{\boldsymbol{\vartheta}_S, f_1; 3}^{*n}$ remains centered under any $P_{\boldsymbol{\vartheta}_S, g_1}^n$, $g_1 \neq f_1$. Actually, letting $K_{f_1}(u) = \varphi_{f_1}(\tilde{F}_{1k}^{-1}(u)) \times \tilde{F}_{1k}^{-1}(u)$ ($u \in (0, 1)$), where \tilde{F}_{1k} stands for the cdf associated with the pdf \tilde{f}_{1k} , and denoting by $R_i = R_i(\boldsymbol{\theta}, \mathbf{V}_S)$ the rank of $d_i = d(X_i, \boldsymbol{\theta}; \mathbf{V}_S)$ among d_1, \dots, d_n , a trivial extension of the proof of Lemma 4.1 in Hallin and Paindaveine (2006) (which is restricted to $S(\boldsymbol{\Sigma}) = \Sigma_{11}$) yields that, under $P_{\boldsymbol{\vartheta}_S, f_1}^n$, as $n \rightarrow \infty$,

$$\Delta_{\boldsymbol{\vartheta}_S, f_1; 3}^{*n} = \frac{1}{2\sqrt{n}} \mathbf{M}_S^{\mathbf{V}_S} (\mathbf{V}_S^{\otimes 2})^{-1/2} \sum_{i=1}^n K_{f_1} \left(\frac{R_i}{n+1} \right) \text{vec} \left(\mathbf{U}_i \mathbf{U}_i' - \frac{1}{k} \mathbf{I}_k \right) + o_P(1),$$

which shows that $\Delta_{\boldsymbol{\vartheta}_S, f_1; 3}^{*n}$ admits an asymptotically equivalent version based on the ranks R_i and the multivariate signs \mathbf{U}_i . This asymptotic equivalence, along with the invariance properties of the families $\mathcal{P}_{S; \boldsymbol{\vartheta}_S}^n = \{P_{\boldsymbol{\vartheta}_S, g_1}^n : g_1 \in \mathcal{F}\}$ (see, for $S(\boldsymbol{\Sigma}) = \Sigma_{11}$, section 4.1 in Hallin and Paindaveine 2006) and a general result by Hallin and Werker (2003), shows that the (semiparametrically) optimal performance (at $P_{\boldsymbol{\vartheta}_S, f_1}^n$), when performing inference on shape in $\mathcal{P}_S^n = \{P_{\boldsymbol{\vartheta}_S, g_1}^n : \boldsymbol{\vartheta}_S \in \boldsymbol{\Theta}_S, g_1 \in \mathcal{F}\}$ coincides with the optimal performance achievable in the parametric model $\mathcal{P}_{S; f_1}^n = \{P_{\boldsymbol{\vartheta}_S, f_1}^n : \boldsymbol{\vartheta}_S \in \boldsymbol{\Theta}_S\}$ (as characterized by the efficient information matrix $\boldsymbol{\Gamma}_{\boldsymbol{\vartheta}_S, f_1; 33}^*$ in (8)). This confirms the adaptivity result first obtained in Example 4 of Bickel (1982), where it is shown that the non-specification of f_1 has no cost when estimating *the inverse shape matrix* $\mathbf{V}^{-1} := \boldsymbol{\Sigma}^{-1}/(\text{tr} \boldsymbol{\Sigma}^{-1})$; note that although this adaptivity result restricts to what is called there a “most general” normalization (the one based on the trace), its proof actually holds for an arbitrary scale functional S .

Summing up, when estimating the shape \mathbf{V}_S in $\mathcal{P}_S^n = \{P_{\boldsymbol{\vartheta}_S, g_1}^n : \boldsymbol{\vartheta}_S \in \boldsymbol{\Theta}_S, g_1 \in \mathcal{F}\}$, the non-specification of σ_S^2 alone is responsible for a loss of efficiency (as already mentioned, the non-specification of $\boldsymbol{\theta}$ does not play any role). This property—call it *Bickel adaptivity*—actually holds for an arbitrary scale functional S . In this paper, we consider a stronger adaptivity concept—*full adaptivity*, say—under which $\boldsymbol{\theta}$, σ_S^2 and f_1 (rather than f_1 alone) lie in the nuisance space of the semiparametric model. The next section shows that full adaptivity holds for one and only one scale functional S , which therefore can be considered canonical.

3 A canonical definition of shape.

We are now ready to state the main result of the paper, which shows that there exists a *unique* scale functional S (thus a unique definition of the shape \mathbf{V}_S) for which the non-specification of σ_S^2 does not cause any loss of efficiency when performing inference on \mathbf{V}_S , hence, for which inference on \mathbf{V}_S in \mathcal{P}_S^n and $\mathcal{P}_{S;\sigma_S^2,f_1}^n$ (equivalently, $\mathcal{P}_{S;\boldsymbol{\theta},\sigma_S^2,f_1}^n$) yields the same optimal performance.

Theorem 3.1 Let Assumptions (A1) and (A2) hold. Then $\boldsymbol{\Gamma}_{\boldsymbol{\vartheta}_S,f_1;32} = \mathbf{0}$ for all $\boldsymbol{\vartheta}_S \in \boldsymbol{\Theta}_S$ iff $S = S_d$, where $S_d(\boldsymbol{\Sigma}) := |\boldsymbol{\Sigma}|^{1/k}$.

The decomposition of scatter into scale and shape through the functional S_d is thus the only one that guarantees (a) mutual (asymptotic) orthogonality of the scale and shape parts of the central sequence (hence, independence in the asymptotic multinormal distribution), and, consequently, (b) *full adaptivity* (see above) in the estimation of shape. This finding strongly pleads in favor of the determinant-based definition of shape which, with its block-diagonal information matrix, is also more convenient from the point of view of statistical inference.

For this canonical parametrization, the shape part of the central sequence and Fisher information matrix take the simple form

$$\boldsymbol{\Delta}_{\boldsymbol{\vartheta}_S,f_1;3}^n = \frac{1}{2\sqrt{n}} \mathbf{M}_S^{\mathbf{V}_S} (\mathbf{V}_S^{\otimes 2})^{-1/2} \sum_{i=1}^n \varphi_{f_1} \left(\frac{d_i}{\sigma_S} \right) \frac{d_i}{\sigma_S} \text{vec}(\mathbf{U}_i \mathbf{U}_i') = \boldsymbol{\Delta}_{\boldsymbol{\vartheta}_S,f_1;3}^{*n},$$

and

$$\boldsymbol{\Gamma}_{\boldsymbol{\vartheta}_S,f_1;33} = \frac{\mathcal{J}_k(f_1)}{4k(k+2)} \mathbf{M}_S^{\mathbf{V}_S} [\mathbf{I}_{k^2} + \mathbf{K}_k] (\mathbf{V}_S^{\otimes 2})^{-1} (\mathbf{M}_S^{\mathbf{V}_S})' = \boldsymbol{\Gamma}_{\boldsymbol{\vartheta}_S,f_1;33}^* \quad (9)$$

(note indeed that the proof of Theorem 3.1 shows that $\mathbf{M}_S^{\mathbf{V}_S} (\text{vec } \mathbf{V}_S^{-1}) = \mathbf{0}$ for all $\mathbf{V}_S \in \mathcal{V}_k^S$, which entails that $\mathbf{M}_S^{\mathbf{V}_S} (\mathbf{V}_S^{\otimes 2})^{-1/2} \mathbf{J}_k = \mathbf{0}$). Theorem 3.1 shows that the determinant-based definition of scale/shape is the only one for which parametric and semiparametric efficiency bounds do coincide (hence, no other choice of S is such that $\boldsymbol{\Delta}_{\boldsymbol{\vartheta}_S,f_1;3}^{*n}$ and $\boldsymbol{\Delta}_{\boldsymbol{\vartheta}_S,f_1;3}^n$ are equal up to $o_P(1)$ terms). Also note that, since $\mathcal{J}_k(f_1) = k(k+2)$ at the multinormal, the canonical parametrization of shape is also the only one for which the information matrix

for shape (either in its original or efficient version) in (9) has *at any* f_1 the same structure as that of the parametric information matrix (7) at the multinormal.

4 Proofs.

In this final section, we prove Theorems 2.1 and 3.1.

For any S satisfying Assumption (A1), consider the mapping $V_{11}^S : \mathring{\text{vech}} \mathcal{S}^k \rightarrow \mathbb{R}$ defined by $S((V_{11}^S(\mathring{\text{vech}} \mathbf{V}), (\mathring{\text{vech}} \mathbf{V})')) = 1$, the existence of which—locally around any $\mathbf{V}_S \in \mathcal{V}_k^S$ —is guaranteed by Assumption (A1) and the implicit function Theorem. We then start with the following lemma.

Lemma 4.1 Let Assumption (A1) hold and fix $\mathbf{V}_S \in \mathcal{V}_k^S$. Let also \mathbf{P}_k be the $(K+1) \times k^2$ matrix such that $\mathbf{P}'_k(\text{vech } \mathbf{v}) = \text{vec } \mathbf{v}$ for any symmetric $k \times k$ matrix \mathbf{v} . Then, $\mathbf{M}_S^{\mathbf{V}_S} = (\nabla V_{11}^S(\mathring{\text{vech}} \mathbf{V}_S) : \mathbf{I}_K) \mathbf{P}_k$.

PROOF OF LEMMA 4.1. Differentiating (at $\mathring{\text{vech}} \mathbf{V}_S$) both sides of $S((V_{11}^S(\mathring{\text{vech}} \mathbf{V}), (\mathring{\text{vech}} \mathbf{V})')) = 1$ with respect to $\mathring{\text{vech}} \mathbf{V}$, we obtain

$$\nabla V_{11}^S(\mathring{\text{vech}} \mathbf{V}_S) = -\frac{\nabla_2 S(\text{vech } \mathbf{V}_S)}{\nabla_1 S(\text{vech } \mathbf{V}_S)}, \quad (10)$$

where $\nabla S(\text{vech } \mathbf{V}_S) = (\nabla_1 S(\text{vech } \mathbf{V}_S), (\nabla_2 S(\text{vech } \mathbf{V}_S))')' \in \mathbb{R} \times \mathbb{R}^K$. Now, $\mathbf{M}_S^{\mathbf{V}_S}$ is such that $(\mathbf{M}_S^{\mathbf{V}_S})'(\mathring{\text{vech}} \mathbf{v}) = (\text{vec } \mathbf{v})$ for any symmetric $k \times k$ matrix $\mathbf{v} = (v_{ij})$ satisfying $(\nabla S(\text{vech } \mathbf{V}_S))'(\mathring{\text{vech}} \mathbf{v}) = (\nabla_1 S(\text{vech } \mathbf{V}_S)) \times v_{11} + (\nabla_2 S(\text{vech } \mathbf{V}_S))'(\mathring{\text{vech}} \mathbf{v}) = 0$, that is, any \mathbf{v} satisfying (see (10))

$$(\nabla V_{11}^S(\mathring{\text{vech}} \mathbf{V}_S))'(\mathring{\text{vech}} \mathbf{v}) = v_{11}.$$

The result follows since, for any such \mathbf{v} , one also has $((\nabla V_{11}^S(\mathring{\text{vech}} \mathbf{V}_S) : \mathbf{I}_K) \mathbf{P}_k)'(\mathring{\text{vech}} \mathbf{v}) = \mathbf{P}'_k(v_{11}, (\mathring{\text{vech}} \mathbf{v})')' = \mathbf{P}'_k(\text{vech } \mathbf{v}) = \text{vec } \mathbf{v}$. \square

PROOF OF THEOREM 2.1. As shown in the proof of Proposition 2.1 in Hallin and Paindaveine (2006), the family $\mathcal{P}_{f_1}^n := \{P_{\boldsymbol{\xi}, f_1}^n = P_{\boldsymbol{\theta}, \boldsymbol{\Sigma}, f_1}^n : \boldsymbol{\xi} = (\boldsymbol{\theta}', (\text{vech } \boldsymbol{\Sigma})')' \in \Xi := \mathbb{R}^k \times (\text{vech } \mathcal{S}_k)\}$, under Assumption (A2), is LAN. Now, for any scale functional S satisfying

Assumption (A1), consider the function $h^S : \Theta_S \rightarrow \Xi$ that maps ϑ_S onto the corresponding value of ξ , namely onto $\xi = h^S(\vartheta_S) = (\theta', \sigma_S^2(\text{vech } \mathbf{V}_S)')' = (\theta', \sigma_S^2 V_{11}^S(\text{vech } \mathbf{V}_S), \sigma_S^2(\text{vech } \mathbf{V}_S)')'$. Since h^S is a diffeomorphism, $\mathcal{P}_{S;f_1}^n = \{P_{\vartheta_S, f_1}^n : \vartheta_S \in \Theta_S\}$ is also LAN, and the corresponding central sequence is given by

$$\Delta_{\vartheta_S, f_1}^n = (Dh^S(\vartheta_S))' \bar{\Delta}_{h(\vartheta_S), f_1}^n, \quad (11)$$

where

$$Dh^S(\vartheta_S) = \begin{pmatrix} \mathbf{I}_k & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & V_{11}^S(\text{vech } \mathbf{V}_S) & \sigma_S^2(\nabla V_{11}^S(\text{vech } \mathbf{V}_S))' \\ \mathbf{0} & \text{vech } \mathbf{V}_S & \sigma_S^2 \mathbf{I}_K \end{pmatrix}$$

is the Jacobian matrix of h^S at ϑ_S and where, letting $d_i(\xi) = d(\mathbf{X}_i, \theta; \Sigma)$ and $\mathbf{U}_i(\xi) = \Sigma^{-1/2}(\mathbf{X}_i - \theta)/d_i(\xi)$ (where $\Sigma^{1/2}$ denotes the symmetric root of Σ),

$$\bar{\Delta}_{\xi, f_1}^n := \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^n \varphi_{f_1}(d_i(\xi)) \Sigma^{-1/2} \mathbf{U}_i(\xi) \\ \frac{1}{2\sqrt{n}} \mathbf{P}_k (\Sigma^{\otimes 2})^{-1/2} \sum_{i=1}^n \text{vec}(\varphi_{f_1}(d_i(\xi)) d_i(\xi) \mathbf{U}_i(\xi) \mathbf{U}_i'(\xi) - \mathbf{I}_k) \end{pmatrix}$$

is the central sequence in the LAN family $\mathcal{P}_{f_1}^n = \{P_{\xi, f_1}^n : \xi \in \Xi\}$ (see the proof of Proposition 2.1 in Hallin and Paindaveine 2006). The result then readily follows from (11), since (i) the scale part $\Delta_{\vartheta_S, f_1; 2}^n$ of $\Delta_{\vartheta_S, f_1}^n$ can be written as in (4) by noting that $(\text{vech } \mathbf{V}_S)' \mathbf{P}_k (\mathbf{V}_S^{\otimes 2})^{-1/2} \text{vec } \mathbf{v} = (\text{tr } \mathbf{v})$ for any $k \times k$ symmetric matrix \mathbf{v} , and since (ii) the shape part $\Delta_{\vartheta_S, f_1; 3}^n$ of $\Delta_{\vartheta_S, f_1}^n$ can be directly put under the form (5) by using Lemma 4.1. \square

Now, for any $\Sigma \in \mathcal{S}_k$ and S satisfying Assumption (A1), define $\mathbf{C}_S^\Sigma := \mathbf{C}_{S, k}^\Sigma$ as the upper-triangular $k \times k$ matrix such that $\text{vech } \mathbf{C}_S^\Sigma := \nabla S(\text{vech } \Sigma)$, and let $\mathbf{D}_S^\Sigma := (\mathbf{C}_S^\Sigma + (\mathbf{C}_S^\Sigma)')/2$. Clearly, $(\text{vec } \mathbf{D}_S^\Sigma)'(\text{vec } \mathbf{v}) = (\text{vech } \mathbf{C}_S^\Sigma)'(\text{vech } \mathbf{v})$ for any symmetric $k \times k$ matrix \mathbf{v} . The matrix \mathbf{M}_S^Σ is then such that $(\mathbf{M}_S^\Sigma)'(\text{vech } \mathbf{v}) = (\text{vec } \mathbf{v})$ for any symmetric $k \times k$ matrix \mathbf{v} satisfying $(\text{vec } \mathbf{D}_S^\Sigma)'(\text{vec } \mathbf{v}) = 0$ (equivalently, satisfying $\text{tr}(\mathbf{D}_S^\Sigma \mathbf{v}) = 0$). For $S(\Sigma) = \Sigma_{11}$, $S(\Sigma) = (\text{tr } \Sigma)/k$, and $S(\Sigma) = |\Sigma|^{1/k}$, one has $\mathbf{D}_S^\Sigma = \mathbf{e}_1 \mathbf{e}_1'$, $\mathbf{D}_S^\Sigma = \frac{1}{k} \mathbf{I}_k$, and $\mathbf{D}_S^\Sigma = \frac{1}{k} |\Sigma|^{1/k} \Sigma^{-1}$, respectively. The following result states some important properties of \mathbf{M}_S^Σ and \mathbf{D}_S^Σ , which are needed in the proof of Theorem 3.1.

Lemma 4.2 Let Assumption (A1) hold and fix $\Sigma \in \mathcal{S}_k$. Then, (i) $\mathbf{D}_S^\Sigma = \mathbf{D}_S^{\lambda\Sigma}$ for all $\lambda > 0$; (ii) $S(\Sigma) = \text{tr}(\mathbf{D}_S^\Sigma \Sigma)$; (iii) letting $\Sigma_\lambda := (1 - \lambda)\mathbf{I}_k + \lambda\Sigma$, $S(\Sigma) = \exp[\int_0^1 \text{tr}((\Sigma - \mathbf{I}_k)\mathbf{D}_S^{\Sigma_\lambda}/S(\Sigma_\lambda)) d\lambda]$; (iv) \mathbf{M}_S^Σ has (maximal) rank K ; (v) denoting by $\ker \mathbf{A}$ the null space of a matrix \mathbf{A} , $(\ker \mathbf{M}_S^\Sigma) \cap (\text{vec } \mathcal{S}_k) = \{\lambda(\text{vec } \mathbf{D}_S^\Sigma) : \lambda \in \mathbb{R}\}$.

PROOF OF LEMMA 4.2. (i) The claim readily follows from the definition of \mathbf{D}_S^Σ and the homogeneity of S , which implies that the gradient of S is constant along half lines of the form $\{\lambda\Sigma : \lambda > 0\}$. (ii) By taking derivatives at $\lambda = 1$ of both sides of the equality $\lambda S(\text{vech } \Sigma) = S(\lambda(\text{vech } \Sigma))$, one obtains that $S(\Sigma) = (\text{vech } \mathbf{C}_S^\Sigma)'(\text{vech } \Sigma) = (\text{vec } \mathbf{D}_S^\Sigma)'(\text{vec } \Sigma) = \text{tr}(\mathbf{D}_S^\Sigma \Sigma)$. (iii) The result follows by integrating between 0 and 1 the mapping $\lambda \mapsto d\log S(\Sigma_\lambda)/d\lambda = (\text{vech } (\Sigma - \mathbf{I}_k))'(\text{vech } \mathbf{C}_S^{\Sigma_\lambda})/S(\Sigma_\lambda) = (\text{vec } (\Sigma - \mathbf{I}_k))'(\text{vec } \mathbf{D}_S^{\Sigma_\lambda})/S(\Sigma_\lambda) = \text{tr}((\Sigma - \mathbf{I}_k)\mathbf{D}_S^{\Sigma_\lambda}/S(\Sigma_\lambda))$. (iv) It is clear from the definition of \mathbf{M}_S^Σ that $(\mathbf{M}_S^\Sigma)'$ maps the canonical basis of \mathbb{R}^K into a collection of linearly independent vectors (in $(\text{vec } \mathcal{S}_k)$, actually). Hence, $\text{rank}(\mathbf{M}_S^\Sigma) = K$. (v) It follows from the definition of \mathbf{M}_S^Σ that, for any symmetric $k \times k$ matrix \mathbf{v} satisfying $(\text{vec } \mathbf{D}_S^\Sigma)'(\text{vec } \mathbf{v}) = 0$, $(\text{vec } \mathbf{D}_S^\Sigma)'(\mathbf{M}_S^\Sigma)'(\text{vech } \mathbf{v}) = 0$. Since Assumption (A1) guarantees that $(\mathbf{D}_S^\Sigma)_{11} = \frac{\partial S}{\partial \Sigma_{11}}(\Sigma) \neq 0$ for all $\Sigma \in \mathcal{S}_k$, this entails that $\mathbf{M}_S^\Sigma(\text{vec } \mathbf{D}_S^\Sigma) = \mathbf{0}$. Now, the proof of (ii) shows that the restriction (L , say) to $(\text{vec } \mathcal{S}_k)$ of the linear mapping from \mathbb{R}^{k^2} to \mathbb{R}^K with matrix \mathbf{M}_S^Σ has rank K . Hence, the null space of L has dimension 1, which establishes the result. \square

We are now able to prove Theorem 3.1.

PROOF OF THEOREM 3.1. Assume that $\Gamma_{\vartheta_S, f_1; 32} = \mathbf{0}$ for all $\vartheta_S \in \Theta_S$. Since Assumption (A2) guarantees that $\mathcal{J}_k(f_1) > k^2$ (see Hallin and Paindaveine 2006), we must have that $\mathbf{M}_S^{\mathbf{V}_S}(\text{vec } \mathbf{V}_S^{-1}) = \mathbf{0}$ for all $\mathbf{V}_S \in \mathcal{V}_k^S$. Lemma 4.2(v) shows that, for all $\mathbf{V}_S \in \mathcal{V}_k^S$, there exists some $\lambda(\mathbf{V}_S) \neq 0$ such that $\mathbf{V}_S^{-1} = \lambda(\mathbf{V}_S)\mathbf{D}_S^{\mathbf{V}_S}$. Lemma 4.2(ii) then yields that $1 = S(\mathbf{V}_S) = \text{tr}((\lambda(\mathbf{V}_S))^{-1}\mathbf{I}_k) = (\lambda(\mathbf{V}_S))^{-1}k$, so that $\lambda(\mathbf{V}_S) = k$ for all $\mathbf{V}_S \in \mathcal{V}_k^S$. Hence, Lemma 4.2(i) entails that, for any $\Sigma \in \mathcal{S}_k$, $\mathbf{D}_S^\Sigma/S(\Sigma) = \mathbf{D}_S^{\Sigma/S(\Sigma)}/S(\Sigma) = \frac{1}{k}(\Sigma/S(\Sigma))^{-1}/S(\Sigma) = \frac{1}{k}\Sigma^{-1}$. Since we also have that $\mathbf{D}_{S_d}^\Sigma/S_d(\Sigma) = \frac{1}{k}\Sigma^{-1}$ (see the paragraph below the proof of Theorem 2.1), Lemma 4.2(iii) yields that $S(\Sigma) = S_d(\Sigma)$ for any $\Sigma \in \mathcal{S}_k$, which establishes the result. \square

References

- [1] Bickel, P.J. (1982). On adaptive estimation, *Ann. Statist.* **10**, 647–671.
- [2] Bickel, P.J. , C.A.J. Klaassen, Y. Ritov and J.A. Wellner (1993). *Efficient and Adaptive Estimation for Semiparametric Models* (Johns Hopkins University Press, Baltimore).
- [3] Croux, C. and G. Haesbroeck (2000). Principal component analysis based on robust estimators of the covariance or correlation matrix : influence functions and efficiencies, *Biometrika* **87**, 603–618.
- [4] Dümbgen, L., and D.E. Tyler (2005). On the breakdown properties of some multivariate M-Functionals, *Scand. J. Statist.* **32**, 247–264.
- [5] Dümbgen, L. (1998). On Tyler’s M-functional of scatter in high dimension, *Ann. Inst. Statist. Math.* **50**, 471–491.
- [6] Hallin, M. and D. Paindaveine (2006). Semiparametrically efficient rank-based inference for shape. I. Optimal rank-based tests for sphericity, *Ann. Statist.* **34**, in press.
- [7] Hallin, M., H. Oja, and D. Paindaveine (2006). Semiparametrically efficient rank-based inference for shape. II. Optimal R-estimation of shape, *Ann. Statist.* **34**, in press.
- [8] Hallin, M., and B.J.M. Werker (2003). Semiparametric efficiency, distribution-freeness, and invariance, *Bernoulli* **9**, 137–165.
- [9] Hettmansperger, T.P. and R.H. Randles (2002). A practical affine equivariant multivariate median, *Biometrika* **89**, 851–860.
- [10] Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory* (Springer-Verlag, New York).

- [11] Maronna, R., D. Martin, and V. Yohai (2006). *Robust Statistics. Theory and Methods* (Wiley, New York).
- [12] Ollila, E., T.P. Hettmansperger, and H. Oja (2004). Affine equivariant multivariate sign methods. Preprint, University of Jyväskylä.
- [13] Randles, R. H. (2000). A simpler, affine-invariant, multivariate, distribution-free sign test, *J. Amer. Statist. Assoc.* **95**, 1263–1268.
- [14] Salibian-Barrera, M., S. Van Aelst, and G. Willems (2006). Principal components analysis based on multivariate MM-estimators with fast and robust bootstrap, *J. Amer. Statist. Assoc.* **101**, 1198–1211.
- [15] Taskinen, S., C. Croux, A. Kankainen, E. Ollila, and H. Oja (2006). Influence functions and efficiencies of the canonical correlation and vector estimates based on scatter and shape matrices, *J. Multivariate Anal.* **97**, 359–384.
- [16] Tatsuoka, K.S., and D.E. Tyler (2000). On the uniqueness of S-Functionals and M-functionals under nonelliptical distributions, *Ann. Statist.* **28**, 1219–1243.
- [17] Tyler, D.E. (1983). Robustness and efficiency properties of scatter matrices, *Biometrika* **70**, 411–420.
- [18] Tyler, D.E. (1987). A distribution-free M-estimator of multivariate scatter, *Ann. Statist.* **15**, 234–251.
- [19] Zuo, Y. (2006). Robust location and scatter estimators in multivariate analysis, in: J. Fan and H.L. Koul, Eds, *Frontiers of Statistics (in honor of Professor P.J. Bickel's 65th Birthday)*, Imperial College, pp. 467–490.