



Statistical theory of shape under elliptical models and singular value decompositions

José A. Díaz-García^{a,*}, Francisco J. Caro-Lopera^b

^a Universidad Autónoma Agraria Antonio Narro, Department of Statistics and Computation, 25315 Buenavista, Saltillo, Coahuila, Mexico

^b Universidad de Medellín, Department of Basic Sciences, Carrera 87 No.30-65, of. 5-103, Medellín, Colombia

ARTICLE INFO

Article history:

Received 19 November 2010

Available online 16 June 2011

AMS 2000 subject classifications:

60E05

15A52

62E15

Keywords:

Shape theory

Non-central and non-isotropic shape density

Zonal polynomials

Singular value decomposition

Maximum likelihood estimators

ABSTRACT

The size-and-shape and shape distributions based on non-central and non-isotropic elliptical distributions are derived in this paper by using the singular value decomposition (SVD). The general densities require the computation of new integrals involving zonal polynomials. The invariance of the central shape distribution is also proved. Finally, some particular densities are applied in a classical data of Biology, and the inference based on exact distributions is performed after choosing the best model by using a modified *BIC** criterion.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

In the past twenty years, the concept and the statistical and mathematical techniques known as *multivariate analysis* have changed notoriously. New statistical and mathematical tools for the analysis of multivariate data have been developed by diverse areas of knowledge, and they have propitiated new disciplines such as pattern recognition, nonlinear multivariate analysis, data mining, manifold learning, generalized multivariate analysis, latent variable analysis and shape theory, among others. These and other fields constitute the so-called *modern multivariate analysis*.

As a particular case, the implementation of the elliptical distribution theory in the classical multivariate analysis of shape, usually based on the Gaussian distribution, introduces important new applications where the Gaussian model does not hold.

The general mathematical aspects of shape via singular value decomposition have been stated by Goodall [7] and Le and Kendall [13]. Almost simultaneously, Goodall and Mardia [8] set the statistical shape theory in the context of multivariate analysis, and derived the associated distributions with that transformation, under a Gaussian law, and for the planar and central cases. In this paper, we are interested in some statistical aspects of shape theory via singular value decompositions, but in a general setting, avoiding the usual assumptions of centrality, isotropy and normality. In particular, this work will derive the shape distributions via SVD, under a non-central and non-isotropic elliptical model. Then the densities are applied in a real problem of the literature. However, we must note that the general approach inherits the open problems of the Gaussian case about the statistical inference in shape theory, and these problems deserve a detailed study in a future work.

* Corresponding author.

E-mail addresses: jadiaz@uaan.mx, jadiazg@ugr.es (J.A. Díaz-García), fjcaro@udem.edu.co (F.J. Caro-Lopera).

Now, for the general approach, consider the following definition: we say that $\mathbf{X} : N \times K$ has a matrix multivariate elliptically contoured distribution if its density with respect to the Lebesgue measure is given by:

$$f_{\mathbf{X}}(\mathbf{X}) = \frac{1}{|\Sigma|^{K/2} |\Theta|^{N/2}} h \left\{ \text{tr} \left[(\mathbf{X} - \mu)' \Sigma^{-1} (\mathbf{X} - \mu) \Theta^{-1} \right] \right\},$$

where $\mu : N \times K$, $\Sigma : N \times N$, $\Theta : K \times K$, Σ positive definite ($\Sigma > \mathbf{0}$), $\Theta > \mathbf{0}$. The function $h : \Re \rightarrow [0, \infty)$ is termed the generator function, and it is such that $\int_0^\infty u^{NK-1} h(u^2) du < \infty$. Such a distribution is denoted by $\mathbf{X} \sim \mathcal{E}_{N \times K}(\mu, \Sigma, \Theta, h)$, see [6] and [9]. Observe that this class of matrix multivariate distributions includes Gaussian, Contaminated Normal, Pearson type II and VI, Kotz, Jensen-Logistic, Power Exponential, among other distributions; whose distributions have tails that are weighted more or less, and/or distributions with greater or smaller degree of kurtosis than the Gaussian model.

Next, we list some facts about the shape theory via SVD, taken from Dryden and Mardia [5]. It is known that the shape of an object is all geometrical information that remains after filtering out translation, rotation and scale information of an original figure (represented by an $N \times K$ matrix $\mathbf{X}$) which is comprised in N labeled points in K dimensions. Such data arise often in biological or medical settings, where the corresponding labeled points are termed *landmarks*. Landmark data also appear naturally in several fields, such as archeology, astronomy, cartography, manufacturing, geology and geophysics. In some instances landmarks may refer to the same physical markers identifiable in more than one map, satellite image, X-ray, etc. Hence, we say that two figures, $\mathbf{X}_1 : N \times K$ and $\mathbf{X}_2 : N \times K$ have the same shape if they are related with a special similarity transformation $\mathbf{X}_2 = \beta \mathbf{X}_1 \mathbf{H} + \mathbf{1}_N \gamma'$, where $\mathbf{H} : K \times K \in SO(K) = \{\mathbf{H} \in \Re^{K \times K} | \mathbf{H}\mathbf{H}' = \mathbf{H}'\mathbf{H} = \mathbf{I}_K \text{ and } |\mathbf{H}| = +1\}$ (the rotation), $\gamma : K \times 1$ (the translation), $\mathbf{1}_N : N \times 1$, $\mathbf{1}_N = (1, 1, \dots, 1)'$, and $\beta > 0$ (the scale). Thus, in this context, the shape of a matrix $\mathbf{X}$ is all the geometrical information about $\mathbf{X}$ that is invariant under Euclidean similarity transformations. Then, the shape space is the set of all possible shapes, it is the orbit space of the non-coincident N landmarks in $\Re^K$ under the action of the Euclidean similarity transformations. The dimension of this space is $NK - K - 1 - K(K - 1)/2$, it is, the original dimension NK is reduced by K for location, by 1 for uniform scale and by $K(K - 1)/2$ for rotation.

Now, in the classical shape theory is assumed that $\mathbf{X}$ has the isotropic matrix multivariate Gaussian distribution with mean $\mu_{\mathbf{X}}$, see [8]

$$\mathbf{X} \sim \mathcal{N}_{N \times K}(\mu_{\mathbf{X}}, \sigma^2 \mathbf{I}_N \otimes \mathbf{I}_K).$$

This paper propose a non-isotropic and non-central elliptical model, i.e.

$$\mathbf{X} \sim \mathcal{E}_{N \times K}(\mu_{\mathbf{X}}, \Sigma_{\mathbf{X}} \otimes \Theta, h).$$

Thus, two fundamental extensions of classical shape theory are provided, namely:

- The new technique assumes a matrix multivariate elliptical distribution for the landmark data instead of considering a matrix multivariate Gaussian distribution.
- Also, the usual isotropic Gaussian condition is replaced by assuming a non-isotropic elliptical model. Two important advantages are obtained: first, the errors are correlated among landmarks, this is considered with the introduction of $\Sigma_{\mathbf{X}}$, a $N \times N$ definite positive matrix; and second, the errors are correlated among the coordinates of landmarks, this condition is noticed with the introduction of Θ , a $K \times K$ definite positive matrix, which in this case is fixed by the expert (estimation of the complete covariance structure $\Sigma_{\mathbf{X}} \otimes \Theta$, demands the construction of a new general theory of shape written in terms of Kronecker products, which will be part of a future work).

The statistical theory of shape associated to this approach studies the effect of randomness and assume a probabilistic model for the original matrix in order to obtain the density of the pre-shape (cone) and shape (disk). The complete procedure for obtaining the shape coordinates $\mathbf{u}$ of $\mathbf{X}$ can be summarized in the following steps:

$$\mathbf{L}\mathbf{X}\Theta^{-1/2} = \mathbf{Y} = \mathbf{V}'\mathbf{D}\mathbf{H} = r\mathbf{V}'\mathbf{W}\mathbf{H} = r\mathbf{V}'\mathbf{W}(\mathbf{u})\mathbf{H}, \quad (1)$$

where $\Theta^{1/2}$ is the positive definite square root of the matrix Θ , i.e. $\Theta = (\Theta^{1/2})^2$, with $\Theta^{1/2} : K \times K$, the matrix $\mathbf{L} : (N - 1) \times N$ has orthonormal rows to $\mathbf{1} = (1, \dots, 1)'$. $\mathbf{L}$ can be a submatrix of the Helmert¹ matrix, for example [8]. Here $\mathbf{Y} = \mathbf{V}'\mathbf{D}\mathbf{H}$ is the SVD of matrix $\mathbf{Y}$, with $\mathbf{V} : n \times (N - 1)$ and $\mathbf{H} : n \times K$ semiorthogonal matrices and $\mathbf{D} : n \times n$, $\mathbf{D} = \text{diag}(D_1, \dots, D_n)$, $n = \min[(N - 1), K]$; $\mathbf{W} = \mathbf{D}/r$, $r = \|\mathbf{D}\| = (\sum_{i=1}^n D_i^2)^{1/2} = \|\mathbf{Y}\|$. Since $\|\mathbf{W}\| = 1$, the elements of $\mathbf{W}$ are a direction vector for shape, and $\mathbf{u}$ comprises $m = n - 1$ generalized polar coordinates.

Then the standard problem considers a model for $\mathbf{X}$ and finds the so termed cone and disk densities, which are the densities of $\mathbf{D}$ and $\mathbf{W}(\mathbf{u})$, respectively.

Alternatively, the shape coordinates $\mathbf{u}$ of $\mathbf{X}$ can be constructed by several ways in terms of the QR decomposition, see [8] and singular value decomposition, see [7].

It is clear that the Gaussian case do not support all the applications and the statistical theory of shape could enriched if complete families of cone and disk densities are available for a particular experiment and the researcher can model the

¹ The Helmert matrix is a $k \times k$ orthogonal matrix, with its first row of elements equal to $1/\sqrt{k}$, and the remaining rows (which constitute the Helmert submatrix $\mathbf{L}$) are orthogonal to the first row. The i th row of the $(k - 1) \times k$ matrix $\mathbf{L}$ consists of $L_i = -1/\sqrt{i(i + 1)}$ repeated i times, followed by $-iL_i$ and then $k - i - 1$ zeros, $i = 1, \dots, k - 1$.

situation by applying a model selection criteria, see for example Rissanen [17], Kass and Raftery [11], Raftery [16] and Yang and Yang [18], among many others.

Therefore, in this paper we propose the statistical theory of shape of Le and Kendall's approach under a general non-central elliptical model. Section 2 obtains the general densities, the cone and disk densities, with some corollaries. Then, the central case and its invariance is studied in Section 3. At the end, we illustrate some results with a simple example: the shape analysis of a mouse vertebra data based on two models: the classical Gaussian and a non-normal Kotz type I. Then the best model is chosen by a modified BIC^* criterion and the corresponding test for equality in disk mean (shape mean) is obtained.

2. Shape theory via SVD

We start with the jacobian of the corresponding decomposition.

Lemma 1. Let $\mathcal{V}_{m,n}$ denote the Stiefel manifold, defined by $\mathcal{V}_{m,n} = \{\mathbf{H} : m \times n : \mathbf{H}\mathbf{H}' = \mathbf{I}_m\}$. If $\mathbf{Y} : N - 1 \times K$, then there exist $\mathbf{V} \in \mathcal{V}_{n,N-1}$, $\mathbf{H} \in \mathcal{V}_{n,K}$ and $\mathbf{D} : n \times n$, $\mathbf{D} = \text{diag}(D_1, \dots, D_n)$, $n = \min[(N - 1), K]$; $D_1 > D_2 > \dots > D_n > 0$, such that $\mathbf{Y} = \mathbf{V}\mathbf{D}\mathbf{H}$; this factorization is termed the non-singular part of the SVD. Therefore

$$(d\mathbf{Y}) = 2^{-n} |\mathbf{D}|^{N-1+K-2n} \prod_{i < j}^n (D_i^2 - D_j^2) (d\mathbf{D}) (\mathbf{V} d\mathbf{V}') (\mathbf{H} d\mathbf{H}').$$

Proof. See [4]. $\square$

Remark 1. In Lemma 1, $(d\mathbf{Y})$ denotes the Lebesgue measure on the space of $N - 1 \times K$ matrices of rank $n = \min[(N - 1), K]$, and it is defined in terms of its mathematically independent elements y_{ij} as

$$(d\mathbf{Y}) = \bigwedge_{i=1}^{N-1} \bigwedge_{j=1}^K dy_{ij},$$

where $\bigwedge$ denotes the exterior product, see [15].

The invariant measure on a Stiefel manifold is given by the differential form

$$(\mathbf{H} d\mathbf{H}') \equiv \bigwedge_{i=1}^{N-1} \bigwedge_{j=i+1}^n \mathbf{h}_j' d\mathbf{h}_i, \quad (2)$$

where $d\mathbf{h}$ is an $N - 1 \times 1$ vector of differentials, see [10] and [15]. For $N - 1 = n$, the Stiefel manifold is the group of orthogonal matrices and $(\mathbf{H} d\mathbf{H}')$ is termed the Haar measure, [15]. Finally,

$$(d\mathbf{D}) = \bigwedge_{i=1}^n dD_i.$$

In order to obtain the joint density function of $(\mathbf{V}, \mathbf{D})$ we need the following generalization of [10, Eq. (22)].

Lemma 2. Let $\mathbf{X} : K \times n$, $\mathbf{Y} : K \times K$ and $\mathbf{H} \in \mathcal{V}_{n,K}$. Then

1.

$$\int_{\mathbf{H} \in \mathcal{V}_{n,K}} [\text{tr}(\mathbf{Y} + \mathbf{X}\mathbf{H})]^p (\mathbf{H} d\mathbf{H}') = \frac{2^n \pi^{Kn/2}}{\Gamma_n[K/2]} \sum_{f=0}^{\infty} \sum_{\lambda} \frac{(p)_{2f} (\text{tr}\mathbf{Y})^{p-2f}}{(\frac{1}{2}K)_{\lambda}} \frac{C_{\lambda}(\frac{1}{4}\mathbf{X}\mathbf{X}')}{f!},$$

where $|(\text{tr}\mathbf{Y})^{-1} \text{tr}\mathbf{X}\mathbf{H}| < 1$ and $\text{tr}\mathbf{Y} \neq 0$.

2.

$$\begin{aligned} & \int_{\mathbf{H} \in \mathcal{V}_{n,K}} \text{tr}(\mathbf{Y} + \mathbf{X}\mathbf{H}) \text{etr}(r(\mathbf{Y} + \mathbf{X}\mathbf{H})) (\mathbf{H} d\mathbf{H}') \\ &= \frac{2^n \pi^{Kn/2}}{\Gamma_n[K/2]} \text{etr}(r\mathbf{Y}) \left\{ \text{tr} {}_0F_1 \left(\frac{1}{2}K; \frac{r^2}{4} \mathbf{X}\mathbf{X}' \right) + \sum_{f=0}^{\infty} \sum_{\lambda} \frac{(f + \frac{1}{2})}{(\frac{1}{2}K)_{\lambda}} \frac{C_{\lambda}(\frac{1}{4}\mathbf{X}\mathbf{X}')}{f!} \right\}, \end{aligned}$$

where $p \in \mathfrak{N}$, $r \in \mathfrak{R}$, $C_{\kappa}(\mathbf{B})$ are the zonal polynomials of $\mathbf{B}$ corresponding to the partition $\kappa = (f_1, \dots, f_p)$ of f , with $\sum_{i=1}^p f_i = f$; and $(a)_{\kappa} = \prod_{i=1}^p (a - (j - 1)/2)_{f_j}$, $(a)_f = a(a + 1) \cdots (a + f - 1)$, are the generalized hypergeometric coefficients and ${}_0F_1$ is the Bessel function, [10].

Proof. This demonstration and the proofs of the remaining lemmas and theorems are given in the [Appendix](#). $\square$

Thus, considering the procedure described in (1), we can obtain:

Theorem 1. The joint density of $(\mathbf{V}, \mathbf{D})$ is

$$f_{\mathbf{V}, \mathbf{D}}(\mathbf{V}, \mathbf{D}) = \frac{\pi^{nK/2} |\mathbf{D}|^{N-1+K-2n} \prod_{i < j} (D_i^2 - D_j^2)}{\Gamma_n[K/2] |\Sigma|^{K/2}} \sum_{t=0}^{\infty} \sum_{\kappa} \frac{h^{(2t)} [\text{tr}(\Sigma^{-1} \mathbf{V}' \mathbf{D}^2 \mathbf{V} + \Omega)]}{t! \left(\frac{1}{2}K\right)_{\kappa}} C_{\kappa}(\Omega \Sigma^{-1} \mathbf{V}' \mathbf{D}^2 \mathbf{V}),$$

where $\mathbf{V} \in \mathcal{V}_{n, N-1}$, $D_1 > \dots > D_n > 0$, $\boldsymbol{\mu} = \mathbf{L}\boldsymbol{\mu}_x$, $\Sigma = \mathbf{L}\Sigma_x \mathbf{L}'$ and $\Omega = \Sigma^{-1} \boldsymbol{\mu} \boldsymbol{\mu}'$.

In the literature, the joint density of the n coordinates in $\mathbf{D}$ is usually termed the cone density or size-and-shape density. Then, we have arrived at the first main result of this section.

Theorem 2. The cone density is given by

$$f_{\mathbf{D}}(\mathbf{D}) = \frac{2^n \pi^{n(N-1+K)/2} |\mathbf{D}|^{N-1+K-2n} \prod_{i < j} (D_i^2 - D_j^2)}{\Gamma_n[K/2] \Gamma_n[(N-1)/2] |\Sigma|^{K/2}} \sum_{\theta, \kappa} \sum_{\phi \in \theta, \kappa} \frac{h^{(2t+l)} (\text{tr} \Omega) \Delta_{\phi}^{\theta, \kappa} C_{\phi}^{\theta, \kappa}(\mathbf{D}^2) C_{\phi}^{\theta, \kappa}(\Sigma^{-1}, \Omega \Sigma^{-1})}{t!! \left(\frac{1}{2}K\right)_{\kappa} C_{\kappa}(\mathbf{I}_{N-1})}, \quad (3)$$

where $D_1 > \dots > D_n > 0$. The notation of the sum operators, $C_{\phi}^{\theta, \kappa}$ and $\Delta_{\phi}^{\theta, \kappa}$ are given in [2], in particular $\Delta_{\phi}^{\theta, \kappa} = \frac{C_{\phi}^{\theta, \kappa}(\mathbf{I}, \mathbf{I})}{C_{\phi}(\mathbf{I})}$.

The following result is a particular case of [Theorem 2](#) and it is usually referred as the isotropic version of the cone density; it is obtained by considering $\Sigma = \sigma^2 \mathbf{I}$. It seems trivial, but in the context of shape analysis, which is an entirely applied discipline, has an important meaning; the isotropy collects the ideal conditions of measuring each landmark with the same error (same instrument, scale and experimenter, etc.), however it does not consider strong correlation among the landmarks.

Corollary 1. Let be $\Sigma = \sigma^2 \mathbf{I}$, then the isotropic cone density has the form

$$f_{\mathbf{D}}(\mathbf{D}) = \frac{2^n \pi^{n(N-1+K)/2} \prod_{i=1}^n D_i^{N-1+K-2n} \prod_{i < j} (D_i^2 - D_j^2)}{\Gamma_n[K/2] \Gamma_n[(N-1)/2] \sigma^{(N-1)K}} \sum_{t=0}^{\infty} \sum_{\kappa} \frac{h^{(2t)} \left(\text{tr} \Omega + \frac{1}{\sigma^2} \text{tr} \mathbf{D}^2 \right) C_{\kappa} \left(\frac{1}{\sigma^2} \mathbf{D}^2 \right) C_{\kappa}(\Omega)}{t! \left(\frac{1}{2}K\right)_{\kappa} C_{\kappa}(\mathbf{I})}, \quad (4)$$

where $D_1 > \dots > D_n > 0$.

Now let be $\mathbf{W} = \mathbf{D}/r$, $r = \|\mathbf{D}\| = \|\mathbf{V}' \mathbf{D} \mathbf{H}\| = \|\mathbf{Y}\|$ and noting that if $\mathbf{D} = \text{diag}(D_1, \dots, D_n)$ we define $\text{vecp}(\mathbf{D}) = (D_1, \dots, D_n)$, then

$$\text{vecp}(\mathbf{D}) = \begin{pmatrix} D_1 \\ \vdots \\ D_n \end{pmatrix}, \text{ implies that } \text{vecp}(\mathbf{W}) = \begin{pmatrix} D_1/r \\ \vdots \\ D_n/r \end{pmatrix} = \text{vecp} \frac{\mathbf{D}}{r},$$

thus

$$\begin{aligned} (d\mathbf{W}(\mathbf{u})) &= r^m \prod_{i=1}^m \sin^{m-i} \theta_i (d\mathbf{u}) \wedge dr \\ &= r^m J(\mathbf{u})(d\mathbf{u}) \wedge dr, \end{aligned}$$

with $m = n-1$, $\mathbf{u} = (\theta_1, \dots, \theta_m)'$ and $J(\mathbf{u}) = \prod_{i=1}^m \sin^{m-i} \theta_i$.

The shape density under Le and Kendall's approach it is known as disk density.

Theorem 3. The disk density is given by

$$\begin{aligned} f_{\mathbf{W}}(\mathbf{W}) &= \frac{\pi^{nK/2} \prod_{i=1}^n l_i^{*N-1+K-2n} \prod_{i < j} (l_i^{*2} - l_j^{*2}) J(\mathbf{u})}{\Gamma_n[K/2] |\Sigma|^{K/2}} \sum_{t=0}^{\infty} \sum_{\kappa} \frac{1}{t! \left(\frac{1}{2}K\right)_{\kappa}} \int_{\mathbf{V} \in \mathcal{V}_{n, N-1}} \frac{C_{\kappa}(\Omega \Sigma^{-1} \mathbf{V}' \mathbf{W}^2 \mathbf{V})}{\sum_{\theta} C_{\theta}(\Sigma^{-1} \mathbf{V}' \mathbf{W}^2 \mathbf{V})} (d\mathbf{V}) \\ &\quad \times \int_0^{\infty} s^{n(N+K-n-1)+2t-1} h^{(2t)}(s^2 + \text{tr} \Omega) (ds), \end{aligned} \quad (5)$$

where $l_1^* > \dots > l_n^* > 0$, and the number of landmarks N are selected in such way that $n(N+K-n-1)/2 + t$ is a positive integer, then $\theta = (l_1, \dots, l_{\alpha})$ is a partition of $n(N+K-n-1)/2 + t$, and $\sum_{i=1}^{\alpha} l_i = n(N+K-n-1)/2 + t$.

The isotropic case of the disk distribution follows:

Theorem 4. The isotropic disk density is given by

$$f_{\mathbf{W}}(\mathbf{W}) = \frac{2^n \pi^{n(N+K-1)/2} \prod_{i=1}^n l_i^{*N-1+K-2n} \prod_{i < j} (l_i^{*2} - l_j^{*2}) J(\mathbf{u})}{\Gamma_n[K/2] \Gamma_n[(N-1)/2] (\sigma^2)^{(N-1)K/2}} \sum_{\kappa} \frac{C_{\kappa} \left(\frac{1}{\sigma^2} \mathbf{W}^2 \right) C_{\kappa}(\mathbf{\Omega})}{t! \left(\frac{1}{2} K \right)_{\kappa} C_{\kappa}(\mathbf{I}_{N-1})} \\ \times \int_0^{\infty} r^{n(N+K-n-1)+2t-1} h(2t) \left(\text{tr} \mathbf{\Omega} + \frac{r^2}{\sigma^2} \right) dr, \quad (6)$$

where $l_1^* > \dots > l_n^* > 0$.

3. Central case

In this section we study the central cone and central disk distributions, which are obtained by taking $\mu = \mathbf{0}$ in Theorems 2 and 3, respectively:

Corollary 2. The central cone density $f_{\mathbf{D}}(\mathbf{D})$ is given by

$$\frac{2^n |\mathbf{D}|^{N-1+K-2n} \prod_{i < j} (D_i^2 - D_j^2)}{\pi^{-n(N+K-1)/2} \Gamma_n[K/2] \Gamma_n[(N-1)/2] |\Sigma|^{K/2}} \sum_{l=0}^{\infty} \sum_{\theta} \frac{h^{(l)}(0) C_{\theta}(\Sigma^{-1}) C_{\theta}(\mathbf{D}^2)}{l! C_{\theta}(\mathbf{I}_{N-1})}, \quad (7)$$

where $D_1 > \dots > D_n > 0$.

And finally, we obtain the central disk density, which does not depend on the elliptical model under consideration. This unusual property in matrix variate analysis asserts, in other words, that the disk density is invariant under the elliptical distributions.

Corollary 3. The central disk density is given by

$$f_{\mathbf{W}}(\mathbf{W}) = \frac{\Gamma[n(N+K-n-1)/2] \prod_{i=1}^n l_i^{*N-1+K-2n} \prod_{i < j} (l_i^{*2} - l_j^{*2}) J(\mathbf{u})}{2 \pi^{n(N-n-1)/2} \Gamma_n[K/2] |\Sigma|^{K/2}} \int_{\mathbf{V} \in \mathcal{V}_{n,N-1}} \frac{(\mathbf{V} d\mathbf{V})}{\sum_{\theta} C_{\theta}(\Sigma^{-1} \mathbf{V} \mathbf{W}^2 \mathbf{V})}, \quad (8)$$

where $l_1^* > \dots > l_n^* > 0$; the number of landmarks N are selected in such way that $n(N+K-n-1)/2$ is a positive integer, then $\theta = (l_1, \dots, l_{\alpha})$ is a partition of $n(N+K-n-1)/2$, and $\sum_{i=1}^{\alpha} l_i = n(N+K-n-1)/2$ where $\theta = (l_1, \dots, l_{\alpha})$ is a partition of the positive integer $n(N+K-n-1)/2$, $\sum_{i=1}^{\alpha} l_i = n(N+K-n-1)/2$.

Proof. From Theorem 3, taking $t = 0$, $\mathbf{\Omega} = \mathbf{0}$, $h^{(0)}(\cdot) \equiv h(\cdot)$ and recalling that

$$\int_0^{\infty} s^{n(N+K-n-1)-1} h(s^2) (ds) = \frac{\Gamma[n(N+K-n-1)/2]}{\pi^{n(N+K-n-1)/2}},$$

the result is obtained. $\square$

4. Example

In this section we apply the above results in a classical example of the literature of shape. As we mentioned before, the inference in statistical shape theory is problematic. In the simplest cases, the classical shape densities based on Gaussian laws, arise open problems about inference, convergence, estimability, computation of invariant polynomials for large degrees, etc., and these and new complications appear when the shape densities are based on general elliptical models, for example in the Kotz type I family, which includes the Gaussian model. Some of these open problems deserve a detail study and they are currently studied by the authors.

4.1. Technical considerations

As the reader can check the general cone and disk densities are given in terms of invariant polynomials, and at this time no inference can be performed with this densities, see [3], except if the series are truncated in the first few terms. However, there is a way to work with an exact density, it is, when we assume an isotropic model.

Just for illustration, consider the isotropic elliptical disk density of [Theorem 4](#)

$$f_{\mathbf{W}}(\mathbf{W}) = \frac{2^n \pi^{n(N+K-1)/2} \prod_{i=1}^n l_i^{*N-1+K-2n} \prod_{i < j} (l_i^{*2} - l_j^{*2}) J(\mathbf{u})}{\Gamma_n[K/2] \Gamma_n[(N-1)/2] (\sigma^2)^{(N-1)K/2}} \sum_{\kappa} \frac{C_{\kappa} \left(\frac{1}{\sigma^2} \mathbf{W}^2 \right) C_{\kappa}(\Omega)}{t! \left(\frac{1}{2} K \right)_{\kappa} C_{\kappa}(\mathbf{I}_{N-1})} \\ \times \int_0^{\infty} r^{n(N+K-n-1)+2t-1} h^{(2t)}(\text{tr} \Omega + r^2/\sigma^2) dr,$$

and the generator for the subfamily Kotz type I:

$$h(y) = \frac{R^{T-1+K(N-1)/2} \Gamma(K(N-1)/2)}{\pi^{K(N-1)/2} \Gamma(T-1+K(N-1)/2)} y^{T-1} \exp\{-Ry\},$$

with derivative

$$\frac{d^k}{dy^k} y^{T-1} \exp\{-Ry\} = (-R)^k y^{T-1} \exp\{-Ry\} \left\{ 1 + \sum_{m=1}^k \binom{k}{m} \left[\prod_{i=0}^{m-1} (T-1-i) \right] (-Ry)^{-m} \right\},$$

see [\[1\]](#) for other families (Pearson VII, Bessel, general Kotz, Jensen-Logistic) and their derivatives.

For example, the disk density for the Kotz model $T = 3$ and $R = 1/2$, is obtained after some simplification as:

$$f_{\mathbf{W}}(\mathbf{W}) = \frac{2^{(-2-M+n+Kn-n^2+nN)/2} \prod_{i=1}^n l_i^{*N-1+K-2n} \prod_{i < j} (l_i^{*2} - l_j^{*2})}{\pi^{(M+n-Kn-nN)/2} \sigma^{M-n(-1+K-n+N)} \Gamma_n[K/2] \Gamma_n[(N-1)/2]} \\ \times \frac{J(\mathbf{u})}{M(M+2)} \text{etr} \left(-\frac{\boldsymbol{\mu}' \boldsymbol{\mu}}{2\sigma^2} \right) \sum_{t=0}^{\infty} \frac{1}{t!} \left\{ \left(-8t + 16t^2 - 16t \text{tr} \left(\frac{\boldsymbol{\mu}' \boldsymbol{\mu}}{2\sigma^2} \right) + 4 \text{tr}^2 \left(\frac{\boldsymbol{\mu}' \boldsymbol{\mu}}{2\sigma^2} \right) \right) \right. \\ \left. + 4 \left(\text{tr} \left(\frac{\boldsymbol{\mu}' \boldsymbol{\mu}}{2\sigma^2} \right) - 4t \right) \Gamma[n(-1+K-n+N)/2 + t + 1] \right. \\ \left. + 4 \Gamma[n(-1+K-n+N)/2 + t + 2] \right\} \sum_{\kappa} \frac{C_{\kappa}(\mathbf{W}^2) C_{\kappa} \left(\frac{\boldsymbol{\mu}' \boldsymbol{\mu}}{2\sigma^2} \right)}{\left(\frac{1}{2} K \right)_{\kappa} C_{\kappa}(\mathbf{I}_{N-1})}.$$

The isotropic Gaussian disk density is given by

$$f_{\mathbf{W}}(\mathbf{W}) = \frac{2^{(-2-M+n+Kn-n^2+nN)/2} \prod_{i=1}^n l_i^{*N-1+K-2n} \prod_{i < j} (l_i^{*2} - l_j^{*2}) J(\mathbf{u})}{\pi^{(M+n-Kn-nN)/2} \sigma^{M-n(-1+K-n+N)} \Gamma_n[K/2] \Gamma_n[(N-1)/2]} \\ \times \text{etr} \left(-\frac{\boldsymbol{\mu}' \boldsymbol{\mu}}{2\sigma^2} \right) \sum_{t=0}^{\infty} \frac{\Gamma[n(-1+K-n+N)/2 + t]}{t!} \sum_{\kappa} \frac{C_{\kappa}(\mathbf{W}^2) C_{\kappa} \left(\frac{\boldsymbol{\mu}' \boldsymbol{\mu}}{2\sigma^2} \right)}{\left(\frac{1}{2} K \right)_{\kappa} C_{\kappa}(\mathbf{I}_{N-1})},$$

where $M = K(N-1)$ and $n = \min\{(N-1), K\}$.

And the corresponding Kotz disk density when $T = 2$ and $R = 1/2$ is given by

$$f_{\mathbf{W}}(\mathbf{W}) = \frac{2^{(-M+n+Kn-n^2+nN)/2} \prod_{i=1}^n l_i^{*N-1+K-2n} \prod_{i < j} (l_i^{*2} - l_j^{*2}) J(\mathbf{u})}{\pi^{(M+n-Kn-nN)/2} \sigma^{M-n(-1+K-n+N)} \Gamma_n \left[\frac{1}{2} K \right] \Gamma_n[(N-1)/2] M} \\ \times \text{etr} \left(-\frac{\boldsymbol{\mu}' \boldsymbol{\mu}}{2\sigma^2} \right) \sum_{t=0}^{\infty} \frac{1}{t!} \left\{ \left(\text{tr} \left(\frac{\boldsymbol{\mu}' \boldsymbol{\mu}}{2\sigma^2} \right) - 2t \right) \Gamma[n(-1+K-n+N)/2 + t] \right. \\ \left. + \Gamma[n(-1+K-n+N)/2 + t + 1] \right\} \sum_{\kappa} \frac{C_{\kappa}(\mathbf{W}^2) C_{\kappa} \left(\frac{\boldsymbol{\mu}' \boldsymbol{\mu}}{2\sigma^2} \right)}{\left(\frac{1}{2} K \right)_{\kappa} C_{\kappa}(\mathbf{I}_{N-1})}.$$

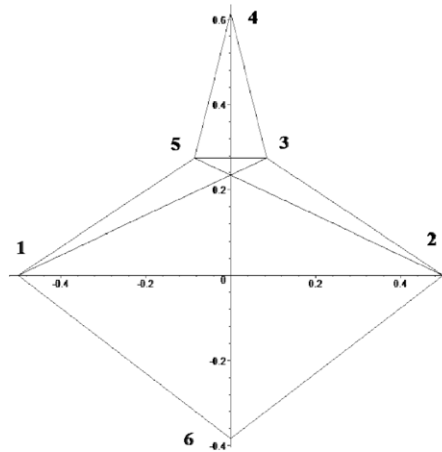
The likelihood based on the above exact densities requires the computation of some special series; a carefully comparison with the known hypergeometric series of two matrix argument indicates that these distributions can be obtained by a suitable modification of the algorithms of Koev and Edelman [\[12\]](#).

As the reader can check, the shape densities are series of zonal polynomials of the form

$$\sum_{t=0}^{\infty} \frac{f(t, \text{tr} \mathbf{X}, \text{tr} \mathbf{Y})}{t!} \sum_{\kappa} \frac{C_{\kappa}(\mathbf{X}) C_{\kappa}(\mathbf{Y})}{(a)_{\kappa} C_{\kappa}(\mathbf{I})}, \quad (9)$$

Table 1Grades of evidence corresponding to values of the BIC^* difference.

BIC^* difference	Evidence
0–2	Weak
2–6	Positive
6–10	Strong
> 10	Very strong

**Fig. 1.** Mouse vertebra.

which has hypergeometric series

$$\sum_{t=0}^{\infty} \frac{1}{t!} \sum_{\kappa} \frac{C_{\kappa}(\mathbf{X})C_{\kappa}(\mathbf{Y})}{(a)_{\kappa} C_{\kappa}(\mathbf{I})}, \quad (10)$$

as a particular case; the expression (10) was not computable for decades. The work of Koev and Edelman [12] solved the problem and it let the computation of the hypergeometric series by truncation of the series (10) until the coefficient for large degrees are zero under certain tolerance. The cited algorithm gives the coefficients of the series (10), then, we can modified the routine for hypergeometric series in order to compute the shape densities (at the same computational costs) multiplying each coefficient of the series by the required function $f(t, \text{tr}\mathbf{X}, \text{tr}\mathbf{Y})$ in (9).

Based on some statistical criteria, our approach demands the selection of the *best* model for the landmark data, which must be chosen among several available laws. In order to decide which the elliptical model is the best one, different criteria have been employed for the model selection. We shall consider a modification of the BIC^* statistic as discussed in Yang and Yang [18], and which was first achieved by Rissanen [17] in a coding theory framework. The modified BIC^* is given by:

$$BIC^* = -2\mathfrak{L}(\tilde{\boldsymbol{\mu}}, \tilde{\sigma}^2, h) + n_p(\log(n+2) - \log 24),$$

where $\mathfrak{L}(\tilde{\boldsymbol{\mu}}, \tilde{\sigma}^2, h)$ is the maximum of the log-likelihood function, n is the sample size and n_p is the number of parameters to be estimated for each particular shape density.

Now, one of the goal of the shape analysis searches the best elliptical distribution, among a set of proposed models, the modified BIC^* criterion suggests to choose the model for which the modified BIC^* receives its smallest value. In addition, as proposed by Kass and Raftery [11] and Raftery [16], the following selection criteria have been employed in order to compare two contiguous models in terms of its corresponding modified BIC^* (Table 1).

4.2. Example

The mouse vertebra problem is studied in detail by Dryden and Mardia [5] under a number of approaches (see also [14]). The experiment considers the second thoracic vertebra T2 of two groups of mice: large and small. The mice are selected and classified according to large or small body weight, respectively; in this case, the sample consists of 23 large and small bones (the data can be found in [5, p. 313–316]). It is of interest to study shape differences between the two groups. The vertebrae are digitized and summarized in six mathematical landmarks which are placed at points of high curvature, see Fig. 1; they are symmetrically selected by measuring the extreme positive and negative curvature of the bone. See [5] for more details.

Recall that for an elliptical density with generator $h(\cdot)$

$$\mathbf{X} \sim \mathcal{E}_{N \times K}(\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X \otimes \boldsymbol{\Theta}, h),$$

Table 2

The maximum likelihood estimators for the large group under the Gaussian model.

Trunc.	$\tilde{\mu}_{11}$	$\tilde{\mu}_{12}$	$\tilde{\mu}_{21}$	$\tilde{\mu}_{22}$	$\tilde{\mu}_{31}$	$\tilde{\mu}_{32}$	$\tilde{\mu}_{41}$
20	−21.5062	32.5168	25.7488	7.8718	21.8521	−9.0516	23.7604
40	−17.4833	−54.5969	23.0921	−8.2256	45.0737	−10.0328	14.9999
60	−24.7805	−63.4184	35.8392	−9.0398	31.9230	−15.0596	11.4528
80	−21.7392	−78.4991	46.2870	−22.9791	28.7150	−10.7543	12.3936
100	−21.0730	−79.2151	36.3872	−14.6742	57.5817	−8.8868	15.6882
120	−22.3784	−76.3768	39.3598	−20.2723	62.2152	−6.5210	14.3961
140	−22.3784	−76.3768	39.3598	−20.2723	62.2152	−6.5210	14.3961
160	−22.3784	−76.3768	39.3598	−20.2723	62.2152	−6.5210	14.3961
Trunc.	$\tilde{\mu}_{42}$	$\tilde{\mu}_{51}$	$\tilde{\mu}_{52}$	$\tilde{\sigma}^2$	Time	Iter.	BIC*
20	23.3636	6.4774	21.5046	128.0642	14	332	−323.6726
40	19.5390	−29.3441	20.6329	112.9101	24	341	−445.7070
60	14.6941	−44.1216	21.1937	94.0917	41	415	−519.1377
80	10.0026	−58.0929	7.8845	97.7155	56	440	−571.4590
100	14.9611	−43.1221	19.9050	80.1962	58	353	−611.9326
120	17.9921	−32.2970	23.6736	66.5318	72	362	−644.3250
140	17.9921	−32.2970	23.6736	66.5318	82	362	−644.3250
160	17.9921	−32.2970	23.6736	66.5318	92	362	−644.3250

where $\mu = \mathbf{L}\mu_x$, then

$$\mathbf{Y} \sim \mathcal{E}_{N-1 \times K}(\mu \Theta^{-1/2}, \Sigma \otimes \mathbf{I}_K, h),$$

with $\Sigma = \mathbf{L}\Sigma_x\mathbf{L}'$. In the mouse vertebra experiment, the objective is to find the maximum likelihood estimators (MLE) of the disk shape μ_w associated with

$$\mu = \begin{pmatrix} \mu_{11} & \mu_{12} \\ \mu_{21} & \mu_{22} \\ \mu_{31} & \mu_{32} \\ \mu_{41} & \mu_{42} \\ \mu_{51} & \mu_{52} \end{pmatrix},$$

and the scale parameter σ^2 defined in the isotropy assumption $\Theta = \mathbf{I}_K$ and $\Sigma = \sigma^2 \mathbf{I}_{N-1}$.

This estimation is obtained in the two independent populations, the small and large groups; first by assuming a Gaussian model and afterward by considering a Kotz model indexed by $T = 2$ and $R = 1/2$. The general procedure is the following: Let $\mathcal{L}(\tilde{\mu}, \tilde{\sigma}^2, h)$ be the log-likelihood function of a given group model. The maximization of the likelihood function $\mathcal{L}(\tilde{\mu}, \tilde{\sigma}^2, h)$ is obtained in this paper by using the *Nelder–Mead Simplex Method*, which is an unconstrained multivariable function using a derivative-free method; specifically, we apply the routine *fminsearch* implemented by the software MatLab.

The initial value for the algorithm is the sample mean of the elliptical matrix variables $\mathbf{Y} \sim \mathcal{E}_{N-1 \times K}(\mu \Theta^{-1/2}, \Sigma \otimes \mathbf{I}_K, h)$, and the median of the mean of sample variances of the Y 's entries. However, an open problem of Koev and Edelman [12] is the relationship between the convergence and the truncation of the series. Concretely, how many terms we need to consider in the series (9) in order to reach some fixed tolerance for convergence. A numerical solution consists of optimizing the log-likelihood, by increasing the truncation until, the MLE's and the maximum of the function reach an equilibrium, which depends on the standard accuracy and tolerance of the *fminsearch* routine. We try the truncations 20, 40, 60, 80, 100, 120, 140 and 160, and we note that after the truncation 120 the solutions stabilize. The MLE's for the small and large groups under the Kotz ($T = 2$) and Gaussian models are given in Tables 2–5, respectively. Tables also show the modified BIC* value, the number of iterations for obtaining the convergence and the time in seconds for each optimization. It is important to note that the MLE's obtains its stability after the truncation 120.

The computations were performed with a processor Intel(R) Core(TM)2 Duo CPU, E7400@2.80 GHz, and 296 GB of RAM.

Fig. 2 show the behavior of the maximum of the log-likelihood when the number of iterations is increased. In this case we use a truncation of 120, and again, we note that the log-likelihood is bounded for a very small number of iterations in each particular model. Note that for each model, the small and large group has a very low difference in modified BIC* (or in maximum log-likelihood), this can anticipate an equality in disk mean (shape mean). Now, the behavior of the two models under each group can be compared in Fig. 3, which shows simultaneously the results of Tables 2–5.

Remark 2. The researcher may be tempted to determine the optimal value of T for the general subfamily of Kotz, however this parameter cannot be estimated. This is exactly the same situation about the MLE of the degrees of freedom of a multivariate t distribution, see [19,20]. Thus, the parameter T of the Kotz type I distribution must be assigned by the researcher according to the knowledge of the experiment, the landmark perturbation. Once the parameter T is fixed, the MLE of μ and σ^2 can be obtained, see [20, p.402].

Table 3The maximum likelihood estimators for the *small* group under the Gaussian model.

Trunc.	$\tilde{\mu}_{11}$	$\tilde{\mu}_{12}$	$\tilde{\mu}_{21}$	$\tilde{\mu}_{22}$	$\tilde{\mu}_{31}$	$\tilde{\mu}_{32}$	$\tilde{\mu}_{41}$
20	25.1764	−10.5396	3.5501	−0.0007	2.1005	1.1282	15.7191
40	7.8879	3.1710	27.4954	−27.1054	−7.5426	0.5810	19.6497
60	2.3931	−27.7470	34.2670	−16.0397	−9.9515	0.4755	17.2967
80	0.1066	38.6322	39.5829	−13.2103	12.5744	0.4127	24.8427
100	0.5334	45.2792	41.8668	−15.0407	18.8989	0.4111	26.6190
120	−1.0147	50.1763	48.0020	−18.6039	3.6564	0.4352	27.9120
140	−1.0147	50.1763	48.0020	−18.6039	3.6564	0.4352	27.9120
160	−1.0147	50.1763	48.0020	−18.6039	3.6564	0.4352	27.9120
Trunc.	$\tilde{\mu}_{42}$	$\tilde{\mu}_{51}$	$\tilde{\mu}_{52}$	$\tilde{\sigma}^2$	Time	Iter.	BIC*
20	23.1174	−5.9329	16.9229	54.1118	24	634	−324.6589
40	22.0756	16.6381	17.4819	43.9233	33	494	−445.8297
60	24.2432	14.8582	13.5573	33.9149	41	419	−518.3844
80	23.4112	−5.4370	12.5073	33.2953	51	384	−569.8220
100	20.9887	3.6446	10.7249	31.6317	64	376	−609.4062
120	24.5017	22.3645	11.5564	33.4367	77	385	−640.8712
140	24.5017	22.3645	11.5564	33.4367	86	385	−640.8712
160	24.5017	22.3645	11.5564	33.4367	96	385	−640.8712

Table 4The maximum likelihood estimators for the *large* group under the Kotz model with $T = 2$.

Trunc.	$\tilde{\mu}_{11}$	$\tilde{\mu}_{12}$	$\tilde{\mu}_{21}$	$\tilde{\mu}_{22}$	$\tilde{\mu}_{31}$	$\tilde{\mu}_{32}$	$\tilde{\mu}_{41}$
20	−23.2617	36.9700	29.3963	10.8432	20.1714	−10.3610	25.2874
40	−20.5881	−57.7015	25.5610	−8.2083	37.1944	−7.5667	15.5154
60	−19.3968	−69.2566	41.9566	−1.7093	32.1251	−20.2930	13.2026
80	−23.4741	−78.1775	38.0658	−15.7006	29.9428	−15.8666	9.5216
100	−21.6964	−80.6907	32.4012	−11.8184	62.4844	−9.6885	14.4352
120	−4.9957	−101.1453	58.9666	−1.1555	48.8317	−10.5520	26.0149
140	−4.9957	−101.1453	58.9666	−1.1555	48.8317	−10.5520	26.0149
160	−4.9957	−101.1453	58.9666	−1.1555	48.8317	−10.5520	26.0149
Trunc.	$\tilde{\mu}_{42}$	$\tilde{\mu}_{51}$	$\tilde{\mu}_{52}$	$\tilde{\sigma}^2$	Time	Iter.	BIC*
20	22.2545	8.8990	21.0667	124.9701	37	339	−332.9439
40	18.4433	−40.8777	24.0442	108.6229	78	388	−459.9510
60	13.8409	−48.3278	16.6046	94.2873	126	436	−536.9652
80	12.9628	−63.1469	14.0281	86.9775	181	478	−592.1343
100	16.8285	−43.8257	22.9390	75.4232	171	346	−634.9891
120	2.6940	−61.5385	−0.1305	86.3932	273	467	−669.3856
140	2.6940	−61.5385	−0.1305	86.3932	315	467	−669.3856
160	2.6940	−61.5385	−0.1305	86.3932	356	467	−669.3856

Table 5The maximum likelihood estimators for the *small* group under the Kotz model with $T = 2$.

Trunc.	$\tilde{\mu}_{11}$	$\tilde{\mu}_{12}$	$\tilde{\mu}_{21}$	$\tilde{\mu}_{22}$	$\tilde{\mu}_{31}$	$\tilde{\mu}_{32}$	$\tilde{\mu}_{41}$
20	20.4498	−12.6391	6.0558	−15.0367	7.8694	0.7216	23.9342
40	0.8019	−17.7149	30.6213	−17.3600	−5.5309	0.5108	20.2581
60	0.6518	−29.7923	38.3334	−14.8532	−6.5375	0.4000	18.1787
80	−0.4220	41.1346	43.8091	−14.8731	8.2122	0.3782	24.4121
100	−0.6762	47.7239	47.1979	−14.2917	14.5796	0.3822	25.4677
120	−1.3297	54.0273	48.4254	−18.8360	11.3155	0.4174	28.2719
140	−1.3297	54.0273	48.4254	−18.8360	11.3155	0.4174	28.2719
160	−1.3297	54.0273	48.4254	−18.8360	11.3155	0.4174	28.2719
Trunc.	$\tilde{\mu}_{42}$	$\tilde{\mu}_{51}$	$\tilde{\mu}_{52}$	$\tilde{\sigma}^2$	Time	Iter.	BIC*
20	17.6146	−3.8497	20.1945	54.5595	56	555	−333.9200
40	25.3876	3.0270	11.3346	34.5103	82	408	−460.0469
60	25.2224	9.6159	13.8208	32.1565	122	418	−536.1662
80	24.6921	6.3171	9.9801	32.7261	156	395	−590.4314
100	23.9736	8.7093	10.4126	31.6234	191	389	−632.3758
120	23.8349	26.4755	11.5351	33.2120	251	426	−665.8183
140	23.8349	26.4755	11.5351	33.2120	288	426	−665.8183
160	23.8349	26.4755	11.5351	33.2120	326	426	−665.8183

Now, compare the small group under the two models. In the small group we see that the Kotz model with parameter $T = 2$ has the smallest BIC^* (−665.8183), and the corresponding modified BIC^* difference with the Gaussian model

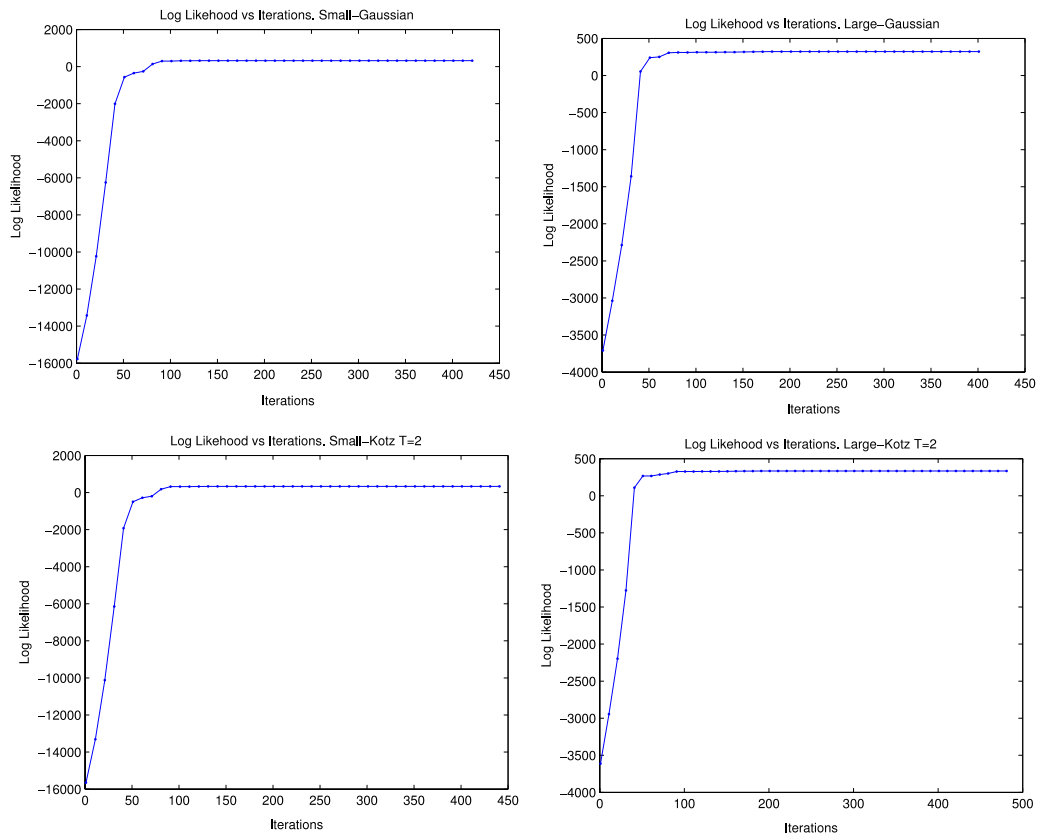


Fig. 2. Behaviors of log-likelihood functions in terms of the iteration number of the fminsearch routine.

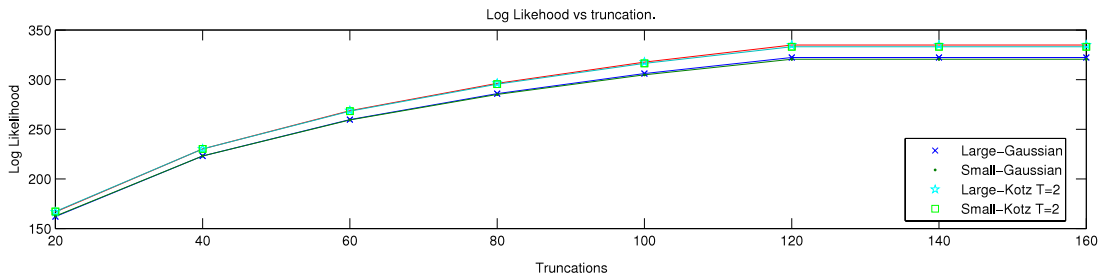


Fig. 3. Behaviors of log-likelihood functions in terms of the truncations.

(−640.8712) is 25 which is rated as a very strong difference, see Tables 3 and 5; in the same way, the large group obtains the smallest BIC^* under the Kotz model (−669.3856) with parameter $T = 2$ meanwhile the Gaussian model has a BIC^* of −644.3250, and again the difference −25 between the two models is classified as very strong, see Tables 2 and 4. In both cases, the true models of the data maybe have tails that are weighted more or less than that of Gaussian model or that the shape distribution present greater or smaller degree of kurtosis than the Gaussian model.

Remark 3. 1. We have used this example of the literature to illustrate the proposed approach; moreover, based on the modified BIC^* , we found that the Kotz distribution (with $T = 2$) is the best model in this experiment. However, suppose that an expert in the area of application knows that the landmarks have a Gaussian distribution, then the classical theory of shape (based on normality) must be applied in such case. Alternatively, if the expert in the area of application suspects that the landmarks do not follow a Gaussian distribution, then our approach can be applied. In this case the expert has the necessary tools to choose an elliptical model (as an alternative to the Gaussian distribution) according to the characteristics of the sample, which reveal and/or support a non-Gaussian distribution, i.e. to select a distribution with more or less heavy tails, with more or less kurtosis than the Gaussian density; among many others possible characteristics.

2. Also, we highlight that the optimizations of this work involves the computations of hundreds of zonal polynomials. This can be carried out by using the efficient algorithm of Koev and Edelman [12] for hypergeometric series, then after some modifications we can evaluate and optimize the required exact log-likelihood. Once the series is truncated at some value t , the algorithm runs over order partitions, of one and two parts, from 0 to t and the so termed zonal polynomials, indexed by these partitions, need to be evaluated, a task which requires a large number of computations. However, the optimization routine requires only a few minutes to obtain the MLE's, as the reader can check in Tables 2–5.

Once the best models are selected for the small and large groups, we can test equality in shape mean between the two independent populations. In this experiment we have: two independent samples of 23 bones; 10 population shape parameters and 1 covariance parameter to estimate for each group. Namely, if $L(\boldsymbol{\mu}_s, \boldsymbol{\mu}_l, \sigma_s^2, \sigma_l^2)$ is the likelihood, where $\boldsymbol{\mu}_s, \boldsymbol{\mu}_l, \sigma_s^2, \sigma_l^2$ represent the mean of shapes and scale parameters of the small and large group, respectively, then we want to test: $H_0 : \boldsymbol{\mu}_s = \boldsymbol{\mu}_l$ vs $H_a : \boldsymbol{\mu}_s \neq \boldsymbol{\mu}_l$. Thus $-2 \log \Lambda = 2 \sup_{H_a} \log L(\boldsymbol{\mu}_s, \boldsymbol{\mu}_l, \sigma_s^2, \sigma_l^2) - 2 \sup_{H_0} \log L(\boldsymbol{\mu}_s, \boldsymbol{\mu}_l, \sigma_s^2, \sigma_l^2)$, and according to Wilk's theorem $-2 \log \Lambda \sim \chi_{10}^2$ under H_0 .

Using fminsearch with a truncation of 160 (convergence at 436 iterations in 18 min) we obtained that:

$$-2 \log \Lambda = 2(668.0510) - 2(659.5534) = 16.9952,$$

this is the same result when the series were truncated at 120 and 140. Since the p-value for the test is

$$P(\chi_{10}^2 \geq 16.9952) = 0.0745,$$

then we have evidence that the small and large mouse vertebrae are different in shape mean. Mardia and Dryden [14] studied this problem with a Gaussian model and Bookstein coordinates (see [5]) and they obtained for the same experiment (under a different transformation) an approximate p-value of zero ($P(\chi_8^2 \geq 127.75)$). Our test does not reject the equality of shape mean with a very small p-value, perhaps, because we are using the SV decomposition and/or a non-Gaussian model; so an important study is suggested in order to clarify which transformation is appropriate for this experiment.

Note that the MLE's given by Tables 2–5 correspond to the matrix $\boldsymbol{\mu}$ in $\mathbf{Y} \sim \mathcal{E}_{N \times K}(\boldsymbol{\mu}, \boldsymbol{\Sigma} \otimes \mathbf{I}_K, h)$, we can use this information and the transformations

$$\mathbf{LX}\boldsymbol{\Theta}^{-1/2} = \mathbf{Y} = \mathbf{V}'\mathbf{D}\mathbf{H} = \mathbf{rV}'\mathbf{W}\mathbf{H} = \mathbf{rV}'\mathbf{W}(\mathbf{u})\mathbf{H},$$

to obtain the disk shape $\boldsymbol{\mu}_w$.

This example deserves a detailed study about some important facts, i.e. the distribution of $-2 \log \Lambda$ for small samples, the truncation of the series, global optimization methods, etc. These problems will be considered in a subsequent work.

It is important to note that the general densities derived here, apply to any elliptical model; some classical elliptical densities such as Kotz, Pearson II and VII, Bessel, Jensen-logistic, can be obtained explicitly. However, they demand the computation of the k th derivative of the generator elliptical function $h(\cdot)$; this is not a trivial fact, but for the above mentioned families, the required formulas are available in [1].

Finally, the isotropic densities are more tractable because they are expanded in terms of zonal polynomials, instead of non-isotropic distributions which require some additional conditions on the number of landmarks in order to obtain a known integral expanded in terms of invariant polynomials. The general densities expanded in terms of invariant polynomials, seem non-computable at this date for large degrees.

Acknowledgments

The authors wish to thank the Associate Editor and the anonymous reviewers for their constructive comments on the preliminary version of this paper. This research work was supported by University of Medellin (Medellin, Colombia) and Universidad Autónoma Agraria Antonio Narro (México), joint grant No. 469, SUMMA group. Also, the first author was partially supported by IDI-Spain, Grant Nos FQM2006-2271 and MTM2008-05785. Finally, F. Caro thanks the project No. 105657 of CONACYT, México.

Appendix. Proofs

We present in this section all the proofs for the lemmas, theorems and corollaries given in Sections 2 and 3.

A.1. Proof of Lemma 2

1. From Lemma 9.5.3 p. 397 in [15] we have

$$\int_{\mathbf{H} \in \mathcal{V}_{n,K}} [\text{tr}(\mathbf{Y} + \mathbf{X}\mathbf{H})]^p (\mathbf{H}\mathbf{d}\mathbf{H}') = \frac{2^n \pi^{Kn/2}}{\Gamma_n[K/2]} \int_{\mathcal{O}(K)} [\text{tr}(\mathbf{Y} + \mathbf{X}\mathbf{H})]^p (\mathbf{d}\mathbf{H}).$$

Furthermore, for $\text{tr}\mathbf{Y} \neq 0$ and $|(\text{tr}\mathbf{Y})^{-1} \text{tr}\mathbf{X}\mathbf{H}| < 1$

$$[\text{tr}(\mathbf{Y} + \mathbf{X}\mathbf{H})]^p = (\text{tr}\mathbf{Y})^p \sum_{f=0}^{\infty} \frac{(p)_f}{f!} (\text{tr}\mathbf{Y})^{-f} (\text{tr}\mathbf{X}\mathbf{H})^f.$$

Now from Eqs. (46) and (22) in [10] it follows that

$$\int_{\mathbf{H} \in \mathcal{V}_{n,K}} [\text{tr}(\mathbf{Y} + \mathbf{X}\mathbf{H})]^p (\mathbf{H}\mathbf{d}\mathbf{H}') = \frac{2^n \pi^{Kn/2}}{\Gamma_n[K/2]} \sum_{f=0}^{\infty} \sum_{\lambda} \frac{(p)_{2f} (\text{tr}\mathbf{Y})^{-2f}}{(2f)!} \frac{\left(\frac{1}{2}\right)_f}{\left(\frac{1}{2}K\right)_{\lambda}} C_{\lambda}(\mathbf{X}\mathbf{X}'),$$

the result follows, noting that $\left(\frac{1}{2}\right)_f / (2f)! = 1/(4^f f!)$ and that $C_{\lambda}(a\mathbf{X}\mathbf{X}') = a^f C_{\lambda}(\mathbf{X}\mathbf{X}')$.

2. This follows by expanding the exponentials in series of powers and by applying (22) and (27) from [10]. $\square$

A.2. Proof of Theorem 1

Let be $\mathbf{\Omega} = \mathbf{\Sigma}^{-1} \mathbf{\mu} \mathbf{\Theta}^{-1} \mathbf{\mu}'$, with $\mathbf{\mu} = \mathbf{L}\mathbf{\mu}_X$, $\mathbf{\Sigma} = \mathbf{L}\mathbf{\Sigma}_X\mathbf{L}'$, then the density of $\mathbf{Y}$ is given by

$$f_{\mathbf{Y}}(\mathbf{Y}) = \frac{1}{|\mathbf{\Sigma}|^{K/2}} h \left[\text{tr}(\mathbf{\Sigma}^{-1} \mathbf{Y} \mathbf{Y}' + \mathbf{\Omega}) - 2 \text{tr} \mathbf{\mu}' \mathbf{\Sigma}^{-1} \mathbf{Y} \right].$$

Now, make the change of variables $\mathbf{Y} = \mathbf{V}'\mathbf{D}\mathbf{H}$, so, by Lemma 1, the joint density function of $\mathbf{V}$, $\mathbf{D}$, $\mathbf{H}$ is

$$dF_{\mathbf{V},\mathbf{D},\mathbf{H}}(\mathbf{V}, \mathbf{D}, \mathbf{H}) = \frac{2^{-n} |\mathbf{D}|^{N-1+K-2n} \prod_{i < j} (D_i^2 - D_j^2)}{|\mathbf{\Sigma}|^{K/2}} (\mathbf{V}d\mathbf{V}') (d\mathbf{D}) h \left[\text{tr}(\mathbf{\Sigma}^{-1} \mathbf{V}' \mathbf{D}^2 \mathbf{V} + \mathbf{\Omega}) - 2 \text{tr} \mathbf{\mu}' \mathbf{\Sigma}^{-1} \mathbf{V}' \mathbf{D} \mathbf{H} \right] (\mathbf{H}d\mathbf{H}').$$

Expanding in power series

$$\begin{aligned} dF_{\mathbf{V},\mathbf{D},\mathbf{H}}(\mathbf{V}, \mathbf{D}, \mathbf{H}) &= \frac{2^{-n} |\mathbf{D}|^{N-1+K-2n} \prod_{i < j} (D_i^2 - D_j^2)}{|\mathbf{\Sigma}|^{K/2}} (\mathbf{V}d\mathbf{V}') (d\mathbf{D}) \\ &\times \sum_{t=0}^{\infty} \frac{1}{t!} h^{(t)} \left[\text{tr}(\mathbf{\Sigma}^{-1} \mathbf{V}' \mathbf{D}^2 \mathbf{V} + \mathbf{\Omega}) \right] \left[\text{tr}(-2 \text{tr} \mathbf{\mu}' \mathbf{\Sigma}^{-1} \mathbf{V}' \mathbf{D} \mathbf{H}) \right]^t (\mathbf{H}d\mathbf{H}'). \end{aligned}$$

From Lemma 2

$$\int_{\mathbf{H} \in \mathcal{V}_{n,K}} \left[\text{tr}(-2 \text{tr} \mathbf{\mu}' \mathbf{\Sigma}^{-1} \mathbf{V}' \mathbf{D} \mathbf{H}) \right]^{2t} (\mathbf{H}d\mathbf{H}') = \frac{2^n \pi^{nK/2}}{\Gamma_n[K/2]} \sum_{\kappa} \frac{\left(\frac{1}{2}\right)_t 4^t}{\left(\frac{1}{2}K\right)_{\kappa}} C_{\kappa}(\mathbf{\Omega} \mathbf{\Sigma}^{-1} \mathbf{V}' \mathbf{D}^2 \mathbf{V}).$$

Observing that $\left(\frac{1}{2}\right)_t 4^t / (2t)! = 1/t!$, the marginal joint density of $\mathbf{V}$, $\mathbf{D}$ is given by

$$dF_{\mathbf{V},\mathbf{D}}(\mathbf{V}, \mathbf{D}) = \frac{\pi^{nK/2} |\mathbf{D}|^{N-1+K-2n} \prod_{i < j} (D_i^2 - D_j^2)}{\Gamma_n\left(\frac{K}{2}\right) |\mathbf{\Sigma}|^{K/2}} \sum_{t=0}^{\infty} \frac{h^{(2t)} \left[\text{tr}(\mathbf{\Sigma}^{-1} \mathbf{V}' \mathbf{D}^2 \mathbf{V} + \mathbf{\Omega}) \right]}{t! \left(\frac{1}{2}K\right)_{\kappa}} C_{\kappa}(\mathbf{\Omega} \mathbf{\Sigma}^{-1} \mathbf{V}' \mathbf{D}^2 \mathbf{V}) (\mathbf{V}d\mathbf{V}') (d\mathbf{D}). \quad \square$$

A.3. Proof of Theorem 2

The joint density of $\mathbf{V}$, $\mathbf{D}$ is

$$dF_{\mathbf{V},\mathbf{D}}(\mathbf{V}, \mathbf{D}) = \frac{\pi^{nK/2} |\mathbf{D}|^{N-1+K-2n} \prod_{i < j} (D_i^2 - D_j^2)}{\Gamma_n[K/2] |\mathbf{\Sigma}|^{K/2}} (d\mathbf{D}) (\mathbf{V}d\mathbf{V}') \sum_{t=0}^{\infty} \sum_{\kappa} \frac{h^{(2t)} \left[\text{tr}(\mathbf{\Sigma}^{-1} \mathbf{V}' \mathbf{D}^2 \mathbf{V} + \mathbf{\Omega}) \right]}{t! \left(\frac{1}{2}K\right)_{\kappa}} C_{\kappa}(\mathbf{\Omega} \mathbf{\Sigma}^{-1} \mathbf{V}' \mathbf{D}^2 \mathbf{V}).$$

Assuming that $h^{(2t)}(\cdot)$ can be expanded in power series,

$$\begin{aligned} h^{(2t)} \left[\text{tr} \mathbf{\Sigma}^{-1} \mathbf{V}' \mathbf{D}^2 \mathbf{V} + \text{tr} \mathbf{\Omega} \right] &= \sum_{l=0}^{\infty} \frac{h^{(2t+l)} [\text{tr} \mathbf{\Omega}]}{l!} \left[\text{tr} \mathbf{\Sigma}^{-1} \mathbf{V}' \mathbf{D}^2 \mathbf{V} \right]^l \\ &= \sum_{l=0}^{\infty} \sum_{\theta} \frac{h^{(2t+l)} [\text{tr} \mathbf{\Omega}]}{l!} C_{\theta}(\mathbf{\Sigma}^{-1} \mathbf{V}' \mathbf{D}^2 \mathbf{V}), \end{aligned}$$

where $C_{\theta}(\mathbf{A})$ is the zonal polynomial corresponding to the partition $\theta = (l_1, \dots, l_{\alpha})$, with $\sum_{i=1}^{\alpha} l_i = l$.

From [2, Eq. (4.13)], the integration of $dF_{\mathbf{V},\mathbf{D}}(\mathbf{V}, \mathbf{D})$ with respect to $\mathbf{V} \in \mathcal{V}_{n,N-1}$ results

$$\int_{\mathbf{V} \in \mathcal{V}_{n,N-1}} C_{\kappa}(\mathbf{\Omega} \mathbf{\Sigma}^{-1} \mathbf{V}' \mathbf{D}^2 \mathbf{V}) C_{\theta}(\mathbf{\Sigma}^{-1} \mathbf{V}' \mathbf{D}^2 \mathbf{V}) (\mathbf{V}d\mathbf{V}') = \frac{2^n \pi^{n(N-1)/2}}{\Gamma_n[(N-1)/2]} \sum_{\phi \in \theta \cdot \kappa} \frac{C_{\phi}^{\theta, \kappa}(\mathbf{\Sigma}^{-1}, \mathbf{\Omega} \mathbf{\Sigma}^{-1}) C_{\phi}^{\theta, \kappa}(\mathbf{D}^2, \mathbf{D}^2)}{C_{\phi}(\mathbf{I}_{N-1})}.$$

And by Davis [2] Eq. (5.1), we have that

$$\begin{aligned} f_{\mathbf{D}}(\mathbf{D}) &= \frac{2^n \pi^{n(N-1+K)/2} |\mathbf{D}|^{N-1+K-2n} \prod_{i < j} (D_i^2 - D_j^2)}{\Gamma_n[K/2] \Gamma_n[(N-1)/2] |\Sigma|^{K/2}} \\ &= \sum_{\theta, \kappa} \sum_{\phi \in \theta \cdot \kappa} \frac{h^{(2t)}(\text{tr} \Omega) \Delta_{\phi}^{\theta, \kappa} C_{\phi}(\mathbf{D}^2) C_{\phi}^{\theta, \kappa}(\Sigma^{-1}, \Omega \Sigma^{-1})}{t!!! \left(\frac{1}{2}K\right)_{\kappa} C_{\phi}(\mathbf{I}_{N-1})}. \quad \square \end{aligned}$$

A.4. Proof of Corollary 1

From Theorem 2

1. $\Omega = \Sigma^{-1} \mu \Theta^{-1} \mu' = \frac{1}{\sigma^2} \mu \Theta^{-1} \mu'$.
2. From [2], Eq. (5.7),

$$\begin{aligned} C_{\phi}^{\theta, \kappa}(\Sigma^{-1}, \Omega \Sigma^{-1}) &= C_{\phi}^{\theta, \kappa}\left(\frac{1}{\sigma^2} \mathbf{I}_{N-1}, \frac{1}{\sigma^2} \Omega\right), \\ &= \left(\frac{1}{\sigma^2}\right)^{2t+l} C_{\phi}^{\theta, \kappa}(\mathbf{I}_{N-1}, \Omega), \\ &= \left(\frac{1}{\sigma^2}\right)^{2t+l} \frac{\Delta_{\phi}^{\theta, \kappa} C_{\phi}(\mathbf{I}_{N-1}) C_{\kappa}(\Omega)}{C_{\kappa}(\mathbf{I}_{N-1})}. \end{aligned}$$

Therefore the second line of (3), denoted by J , it is simplified as follows:

$$\begin{aligned} J &= \sum_{\theta, \kappa} \sum_{\phi \in \theta \cdot \kappa} \frac{h^{(2t+l)}(\text{tr} \Omega) \Delta_{\phi}^{\theta, \kappa} C_{\phi}(\mathbf{D}^2) C_{\phi}^{\theta, \kappa}(\Sigma^{-1}, \Omega \Sigma^{-1})}{t!!! \left(\frac{1}{2}K\right)_{\kappa} C_{\phi}(\mathbf{I}_{N-1})}, \\ &= \sum_{\theta, \kappa} \sum_{\phi \in \theta \cdot \kappa} \frac{h^{(2t+l)}(\text{tr} \Omega) \left(\Delta_{\phi}^{\theta, \kappa}\right)^2 C_{\phi}(\mathbf{D}^2) C_{\kappa}(\Omega)}{t!!! (\sigma^2)^{2t+l} \left(\frac{1}{2}K\right)_{\kappa} C_{\kappa}(\mathbf{I}_{N-1})}. \end{aligned}$$

Note that $\sum_{\phi \in \theta \cdot \kappa} \left(\Delta_{\phi}^{\theta, \kappa}\right)^2 C_{\phi}(\mathbf{D}^2) = C_{\kappa}(\mathbf{D}^2) C_{\theta}(\mathbf{D}^2)$, see [2, Eq. 5.10]. Thus

$$\begin{aligned} J &= \sum_{\theta, \kappa} \frac{h^{(2t+l)}(\text{tr} \Omega) C_{\kappa}(\mathbf{D}^2) C_{\theta}(\mathbf{D}^2) C_{\kappa}(\Omega)}{t!!! (\sigma^2)^{2t+l} \left(\frac{1}{2}K\right)_{\kappa} C_{\kappa}(\mathbf{I}_{N-1})}, \\ &= \sum_{\kappa} \sum_{l=0}^{\infty} \frac{h^{(2t+l)}(\text{tr} \Omega) C_{\kappa}(\mathbf{D}^2) C_{\kappa}(\Omega)}{t!!! (\sigma^2)^{2t+l} \left(\frac{1}{2}K\right)_{\kappa} C_{\kappa}(\mathbf{I}_{N-1})} \sum_{\theta} C_{\theta}(\mathbf{D}^2), \\ &= \sum_{\kappa} \sum_{l=0}^{\infty} \frac{h^{(2t+l)}(\text{tr} \Omega) C_{\kappa}(\mathbf{D}^2) C_{\kappa}(\Omega)}{t!!! (\sigma^2)^{2t+l} \left(\frac{1}{2}K\right)_{\kappa} C_{\kappa}(\mathbf{I}_{N-1})} (\text{tr} \mathbf{D}^2)^l. \end{aligned}$$

Now, observe that $h(v) = \sum_{l=0}^{\infty} \frac{h^{(l)}(a)}{l!} (v-a)^l$, with $a = \text{tr} \Omega$, $v = \text{tr} \Omega + \frac{1}{\sigma^2} \text{tr} \mathbf{D}^2$, and $h(v) = h^{(2t)}(v)$. Thus

$$\begin{aligned} J &= \sum_{\kappa} \frac{h^{(2t)}(\text{tr} \Omega + \frac{1}{\sigma^2} \text{tr} \mathbf{D}^2) C_{\kappa}(\mathbf{D}^2) C_{\kappa}(\Omega)}{t! (\sigma^2)^{2t} \left(\frac{1}{2}K\right)_{\kappa} C_{\kappa}(\mathbf{I}_{N-1})}, \\ &= \sum_{\kappa} \frac{h^{(2t)}(\text{tr} \Omega + \frac{1}{\sigma^2} \text{tr} \mathbf{D}^2) C_{\kappa}(\mathbf{D}^2) C_{\kappa}(\Omega)}{t! \left(\frac{1}{2}K\right)_{\kappa} C_{\kappa}(\mathbf{I}_{N-1})}. \quad \square \end{aligned}$$

A.5. Proof of Theorem 3

From Theorem 1

$$dF_{\mathbf{V}, \mathbf{D}}(\mathbf{V}, \mathbf{D}) = \frac{\pi^{nK/2} |\mathbf{D}|^{N-1+K-2n} \prod_{i < j} (D_i^2 - D_j^2)}{\Gamma_n[K/2] |\Sigma|^{K/2}} (d\mathbf{D})(\mathbf{V} d\mathbf{V}') \sum_{t=0}^{\infty} \sum_{\kappa} \frac{h^{(2t)}[\text{tr}(\Sigma^{-1} \mathbf{V} \mathbf{D}^2 \mathbf{V})] C_{\kappa}(\Omega \Sigma^{-1} \mathbf{V} \mathbf{D}^2 \mathbf{V})}{t! \left(\frac{1}{2}K\right)_{\kappa}}.$$

Let $\mathbf{W} = \text{diag}(l_1^*, \dots, l_n^*)$, $l_i^* = D_i/r$, $r = \|\mathbf{D}\| = \|\mathbf{Y}\|$, then $dF_{\mathbf{V}, \mathbf{W}}(\mathbf{V}, \mathbf{W})$ is given by

$$= \frac{\pi^{nK/2} |r\mathbf{W}|^{N-1+K-2n} r^m J(\mathbf{u}) \prod_{i < j} r^2 (l_i^{*2} - l_j^{*2})}{\Gamma_n \left[\frac{1}{2}K \right] |\Sigma|^{K/2}} (\mathbf{dW})(\mathbf{VdV}') \sum_{t=0}^{\infty} \sum_{\kappa} \frac{h^{(2t)} \left[r^2 \text{tr} (\Sigma^{-1} \mathbf{V}' \mathbf{W}^2 \mathbf{V}) \right] C_{\kappa} (r^2 \Omega \Sigma^{-1} \mathbf{V}' \mathbf{W}^2 \mathbf{V})}{t! \left(\frac{1}{2}K \right)_{\kappa}}.$$

Note that

1. $|r\mathbf{W}|^{N-1+K-2n} = r^{n(N-1+K-2n)}$.
2. $\prod_{i < j} r^2 (l_i^{*2} - l_j^{*2}) = (r^2)^{n(n-1)/2} \prod_{i < j} (l_i^{*2} - l_j^{*2})$.
3. $C_{\kappa} (r^2 \Omega \Sigma^{-1} \mathbf{V}' \mathbf{W}^2 \mathbf{V}) = r^{2t} C_{\kappa} (\Omega \Sigma^{-1} \mathbf{V}' \mathbf{W}^2 \mathbf{V})$.

Collection powers of r and defining

$$r = \frac{s}{(\text{tr} \Sigma^{-1} \mathbf{V}' \mathbf{W}^2 \mathbf{V})^{\frac{1}{2}}}, \quad \text{with } dr = \frac{ds}{(\text{tr} \Sigma^{-1} \mathbf{V}' \mathbf{W}^2 \mathbf{V})^{\frac{1}{2}}}$$

hence

$$\begin{aligned} & \int_0^{\infty} r^{n(N+K-n-1)+2t-1} h^{(2t)} (r^2 \text{tr} \Sigma^{-1} \mathbf{V}' \mathbf{W}^2 \mathbf{V}) dr \\ &= \int_0^{\infty} \left(\frac{s}{(\text{tr} \Sigma^{-1} \mathbf{V}' \mathbf{W}^2 \mathbf{V})^{\frac{1}{2}}} \right)^{n(N+K-n-1)+2t-1} h^{(2t)} (s^2) \frac{ds}{(\text{tr} \Sigma^{-1} \mathbf{V}' \mathbf{W}^2 \mathbf{V})^{\frac{1}{2}}} \\ &= (\text{tr} \Sigma^{-1} \mathbf{V}' \mathbf{W}^2 \mathbf{V})^{-n(N+K-n-1)/2+t} \int_0^{\infty} s^{n(N+K-n-1)+2t-1} h^{(2t)} (s^2) ds. \end{aligned}$$

Thus the marginal density of $dF_{\mathbf{W}}(\mathbf{W})$ is given by

$$\begin{aligned} & \frac{\pi^{nK/2} |\mathbf{W}|^{N-1+K-2n} J(\mathbf{u}) \prod_{i < j} (l_i^{*2} - l_j^{*2})}{\Gamma_n [K/2] |\Sigma|^{K/2}} \sum_{t=0}^{\infty} \sum_{\kappa} \frac{1}{t! \left(\frac{1}{2}K \right)_{\kappa}} \int_{\mathbf{V} \in \mathcal{V}_{n, N-1}} C_{\kappa} (\Omega \Sigma^{-1} \mathbf{V}' \mathbf{W}^2 \mathbf{V}) \\ & \times (\text{tr} \Sigma^{-1} \mathbf{V}' \mathbf{W}^2 \mathbf{V})^{-n(N+K-n-1)/2+t} (\mathbf{VdV}') \int_0^{\infty} s^{n(N+K-n-1)+2t-1} h^{(2t)} (s^2) ds. \end{aligned}$$

Now, let be

$$\begin{aligned} J &= \int_{\mathbf{V} \in \mathcal{V}_{n, N-1}} \frac{C_{\kappa} (\Omega \Sigma^{-1} \mathbf{V}' \mathbf{W}^2 \mathbf{V}) (\mathbf{VdV}')}{(\text{tr} \Sigma^{-1} \mathbf{V}' \mathbf{W}^2 \mathbf{V})^{n(N+K-n-1)/2+t}}, \\ &= \int_{\mathbf{V} \in \mathcal{V}_{n, N-1}} \frac{C_{\kappa} (\Omega \Sigma^{-1} \mathbf{V}' \mathbf{W}^2 \mathbf{V})}{\sum_{\theta} C_{\theta} (\Sigma^{-1} \mathbf{V}' \mathbf{W}^2 \mathbf{V})} (\mathbf{VdV}'), \end{aligned}$$

where the number of landmarks N are selected in such way that $n(N+K-n-1)/2+t$ is a positive integer, then $\theta = (l_1, \dots, l_{\alpha})$ is a partition of $n(N+K-n-1)/2+t$, and $\sum_{i=1}^{\alpha} l_i = n(N+K-n-1)/2+t$. Then we obtain the desired result. $\square$

A.6. Proof of Theorem 4

The result is obtained from (5) taking $\Sigma = \sigma^2 \mathbf{I}$ and observing that $|\Sigma| = |\sigma^2 \mathbf{I}| = (\sigma^2)^{(N-1)}$;

$$\text{tr} \Sigma^{-1} \mathbf{V}' \mathbf{W}^2 \mathbf{V} = \frac{1}{\sigma^2} \text{tr} \mathbf{V}' \mathbf{W}^2 \mathbf{V} = \frac{1}{\sigma^2} \text{tr} \mathbf{W}^2 \mathbf{V} \mathbf{V}' = \frac{1}{\sigma^2} \text{tr} \mathbf{W}^2 = \frac{1}{\sigma^2},$$

recalling that $\|\mathbf{W}\| = 1$. Hence

$$\begin{aligned} J &= \int_{\mathbf{V} \in \mathcal{V}_{n, N-1}} \frac{C_{\kappa} (\Omega \Sigma^{-1} \mathbf{V}' \mathbf{W}^2 \mathbf{V}) (\mathbf{VdV}')}{(\text{tr} \Sigma^{-1} \mathbf{V}' \mathbf{W}^2 \mathbf{V})^{n(N+K-n-1)/2+t}} \\ &= (\sigma^2)^{n(N+K-n-1)+t} \int_{\mathbf{V} \in \mathcal{V}_{n, N-1}} C_{\kappa} \left(\frac{1}{\sigma^2} \Omega \mathbf{V}' \mathbf{W}^2 \mathbf{V} \right) (\mathbf{VdV}') \\ &= \frac{(\sigma^2)^{n(N+K-n-1)+t} 2^n \pi^{n(N-1)/2} C_{\kappa} \left(\frac{1}{\sigma^2} \Omega \right) C_{\kappa} (\mathbf{W}^2)}{\Gamma_n [(N-1)/2] C_{\kappa} (\mathbf{I}_{N-1})}. \end{aligned}$$

Finally, (6) is obtained making the change of variable $s = r/\sigma$ with $ds = dr/\sigma$ in (5) and observing that $C_\kappa \left(\frac{1}{\sigma^2} \Omega \right) C_\kappa (\mathbf{W}^2) = C_\kappa (\Omega) C_\kappa \left(\frac{1}{\sigma^2} \mathbf{W}^2 \right)$.

Alternatively, let $l_i^* = D_i/r$, $\text{tr} \mathbf{D}^2 = \sum_{i=1}^n D_i^2 = r^2$ in (4), therefore

$$f_{\mathbf{W}}(\mathbf{W}) = \frac{2^n \pi^{n(N+K-1)/2} r^m J(\mathbf{u}) \prod_{i=1}^n (r l_i^*)^{N-1+K-2n} \prod_{i < j} r^2 (l_i^{*2} - l_j^{*2})}{\Gamma_n[K/2] \Gamma_n[(N-1)/2] (\sigma^2)^{(N-1)K/2}} \sum_{t=0}^{\infty} \sum_{\kappa} \frac{h^{(2t)}(\text{tr} \Omega + r^2/\sigma^2) C_\kappa \left(\frac{r^2}{\sigma^2} \mathbf{W}^2 \right) C_\kappa (\Omega)}{t! \left(\frac{1}{2} K \right)_\kappa C_\kappa (\mathbf{I}_{N-1})}.$$

Observe that

1. $\prod_{i=1}^n (r l_i^*)^{N-1+K-2n} = r^{n(N-1+K-2n)} \prod_{i=1}^n l_i^{*N-1+K-2n}$.
2. $\prod_{i < j} r^2 (l_i^{*2} - l_j^{*2}) = r^{n(n-1)} \prod_{i < j} (l_i^{*2} - l_j^{*2})$.
3. $C_\kappa \left(\frac{r^2}{\sigma^2} \mathbf{W}^2 \right) = r^{2t} C_\kappa \left(\frac{1}{\sigma^2} \mathbf{W}^2 \right)$.

Collecting powers of r we have

$$\int_0^\infty r^{n(N+K-n-1)+2t-1} h^{(2t)}(\text{tr} \Omega + r^2/\sigma^2) dr. \quad \square$$

A.7. Proof of Corollary 2

Start with

$$dF_{\mathbf{Y}}(\mathbf{Y}) = \frac{1}{|\Sigma|^{K/2}} h[\text{tr} \Sigma^{-1} \mathbf{Y} \mathbf{Y}'] (d\mathbf{Y}).$$

The joint density of $\mathbf{V}$, $\mathbf{D}$, $\mathbf{H}$ is

$$dF_{\mathbf{V}, \mathbf{D}, \mathbf{H}}(\mathbf{V}, \mathbf{D}, \mathbf{H}) = \frac{\prod_{i < j} (D_i^2 - D_j^2) (d\mathbf{D})}{2^n |\mathbf{D}|^{-(N-1+K-2n)} |\Sigma|^{K/2}} h[\text{tr} \Sigma^{-1} \mathbf{V}' \mathbf{D}^2 \mathbf{V}] (\mathbf{H} d\mathbf{H}') (\mathbf{V} d\mathbf{V}').$$

Recalling that

$$\int_{\mathbf{H} \in \mathcal{V}_{n,K}} (\mathbf{H}' d\mathbf{H}) = \frac{2^n \pi^{nK/2}}{\Gamma_n[K/2]}$$

we have that

$$dF_{\mathbf{V}, \mathbf{D}}(\mathbf{V}, \mathbf{D}) = \frac{\pi^{nK/2} |\mathbf{D}|^{N-1+K-2n} \prod_{i < j} (D_i^2 - D_j^2) (d\mathbf{D})}{\Gamma_n[K/2] |\Sigma|^{K/2}} h[\text{tr} \Sigma^{-1} \mathbf{V}' \mathbf{D}^2 \mathbf{V}] (\mathbf{V} d\mathbf{V}').$$

Then the required density is given by

$$f_{\mathbf{D}}(\mathbf{D}) = \frac{|\mathbf{D}|^{N-1+K-2n} \prod_{i < j} (D_i^2 - D_j^2)}{\pi^{-nK/2} \Gamma_n[K/2] |\Sigma|^{K/2}} \int_{\mathbf{V} \in \mathcal{V}_{n,N-1}} h[\text{tr} \Sigma^{-1} \mathbf{V}' \mathbf{D}^2 \mathbf{V}] (\mathbf{V} d\mathbf{V}').$$

Integrating with respect to $\mathbf{V}$

$$\begin{aligned} \int_{\mathcal{V}_{n,N-1}} h[\text{tr} \Sigma^{-1} \mathbf{V}' \mathbf{D}^2 \mathbf{V}] (\mathbf{V} d\mathbf{V}') &= \int_{\mathcal{V}_{n,N-1}} \sum_{l=0}^{\infty} \frac{h^{(l)}(0)}{l!} (\text{tr} \Sigma^{-1} \mathbf{V}' \mathbf{D}^2 \mathbf{V})^l (\mathbf{V} d\mathbf{V}') \\ &= \sum_{l=0}^{\infty} \sum_{\theta} \frac{h^{(l)}(0)}{l!} \int_{\mathcal{V}_{n,N-1}} C_{\theta}(\Sigma^{-1} \mathbf{V}' \mathbf{D}^2 \mathbf{V}) (\mathbf{V} d\mathbf{V}') \\ &= \sum_{l=0}^{\infty} \sum_{\theta} \frac{h^{(l)}(0)}{l!} \frac{2^n \pi^{n(N-1)/2} C_{\theta}(\Sigma^{-1}) C_{\theta}(\mathbf{D}^2)}{\Gamma_n[\frac{1}{2}(N-1)] C_{\theta}(\mathbf{I}_{N-1})} \\ &= \frac{2^n \pi^{n(N-1)/2}}{\Gamma_n[(N-1)/2]} \sum_{l=0}^{\infty} \sum_{\theta} \frac{h^{(l)}(0)}{l!} \frac{C_{\theta}(\Sigma^{-1}) C_{\theta}(\mathbf{D}^2)}{C_{\theta}(\mathbf{I}_{N-1})}. \end{aligned}$$

Then,

$$f_{\mathbf{D}}(\mathbf{D}) = \frac{2^n \pi^{n(N+K-1)/2} |\mathbf{D}|^{N-1+K-2n} \prod_{i < j} (D_i^2 - D_j^2)}{\Gamma_n[K/2] \Gamma_n[(N-1)/2] |\Sigma|^{K/2}} \sum_{l=0}^{\infty} \sum_{\theta} \frac{h^{(l)}(0)}{l!} \frac{C_{\theta}(\Sigma^{-1}) C_{\theta}(\mathbf{D}^2)}{C_{\theta}(\mathbf{I}_{N-1})}.$$

Alternatively, from Theorem 2, if we take

1. $h^{(2t+l)}(\text{tr} \Sigma) = h^{(l)}(0)$,
2. $\Delta_{\phi}^{\theta, \kappa} = \Delta_{\phi}^{\theta, \theta} = \frac{C_{\theta}(\mathbf{I})}{C_{\theta}(\mathbf{I})} = 1$,
3. $C_{\phi}(\mathbf{D}^2) = C_{\theta}(\mathbf{D}^2)$, $C_{\phi}(\mathbf{I}) = C_{\theta}(\mathbf{I})$,
4. $C_{\phi}^{\theta, \kappa}(\Sigma^{-1}, \Sigma \Sigma^{-1}) = C_{\phi}^{\theta, \kappa}(\Sigma^{-1}, 0) = C_{\theta}(\Sigma^{-1})$,

the required result follows. $\square$

References

- [1] F.J. Caro-Lopera, J.A. Díaz-García, G. González-Farías, Noncentral elliptical configuration density, *J. Multivariate Anal.* 101 (2009) 32–43.
- [2] A.W. Davis, Invariant polynomials with two matrix arguments, extending the zonal polynomials, in: P.R. Krishnaiah (Ed.), *Multivariate Analysis V*, North-Holland, 1980, pp. 287–299.
- [3] A.W. Davis, On the construction of a class of invariant polynomials in several matrices, extending the zonal polynomials, *Ann. Inst. Statist. Math.* 33 (1981) 297–313.
- [4] J.A. Díaz-García, R. Gutiérrez-Jáimez, K.V. Mardia, Wishart and Pseudo-Wishart distributions and some applications to shape theory, *J. Multivariate Anal.* 63 (1997) 73–87.
- [5] I.L. Dryden, K.V. Mardia, *Statistical Shape Analysis*, John Wiley and Sons, Chichester, 1989.
- [6] K.T. Fang, Y.T. Zhang, *Generalized Multivariate Analysis*, Science Press, Springer-Verlag, Beijing, 1990.
- [7] C.G. Goodall, Procrustes methods in the statistical analysis of shape with discussion, *J. Roy. Statist. Soc. Series B* 53 (1991) 285–339.
- [8] C.R. Goodall, K.V. Mardia, Multivariate aspects of shape theory, *Ann. Statist.* 21 (1993) 848–866.
- [9] A.K. Gupta, T. Varga, *Elliptically Contoured Models in Statistics*, Kluwer Academic Publishers, Dordrecht, 1993.
- [10] A.T. James, Distributions of matrix variate and latent roots derived from normal samples, *Ann. Math. Statist.* 35 (1964) 475–501.
- [11] R.E. Kass, A.E. Raftery, Bayes factor, *J. Amer. Statist. Assoc.* 90 (1995) 773–795.
- [12] P. Koev, A. Edelman, The efficient evaluation of the hypergeometric function of a matrix argument, *Math. Comp.* 75 (2006) 833–846.
- [13] H.L. Le, D.G. Kendall, The Riemannian structure of Euclidean spaces: a novel environment for statistics, *Ann. Statist.* 21 (1993) 1225–1271.
- [14] K.V. Mardia, I.L. Dryden, The statistical analysis of shape data, *Biometrika* 76 (1989) 271–281.
- [15] R.J. Muirhead, *Aspects of multivariate statistical theory*, in: *Wiley Series in Probability and Mathematical Statistics*, John Wiley & Sons, Inc, 1982.
- [16] A.E. Raftery, Bayesian model selection in social research, *Sociol. Methodol.* 25 (1995) 111–163.
- [17] J. Rissanen, Modelling by shortest data description, *Automatica* 14 (1978) 465–471.
- [18] Ch.Ch. Yang, Ch.Ch. Yang, Separating latent classes by information criteria, *J. Classification* 24 (2007) 183–203.
- [19] A. Zellner, *Theory of Probability*, Clarendon Press, Oxford, 1967.
- [20] A. Zellner, Bayesian and non-Bayesian analysis of the regression model with multivariate student *t* error terms, *JASA* 71 (354) (1976) 400–405.