Complex Elliptical distributions with application to shape analysis

Athanassios C. Micheas\textsuperscript{1}, Dipak K. Dey\textsuperscript{2} and Kanti V. Mardia\textsuperscript{3}

Abstract: We introduce a general class of complex elliptical distributions on a complex sphere that includes many of the most commonly used distributions, like the complex Watson, Bingham, angular central Gaussian and several others. We study properties of this family of distributions and apply the distribution theory for modeling shapes in two dimensions. We develop maximum likelihood and Bayesian methods of estimation to describe shape and obtain confidence bounds and credible regions for shapes. The methodology is illustrated through an example where estimation of shape of mouse vertebrae is desired.

Key words and phrases: Bayesian Estimation, Complex elliptical family of distributions, Complex Watson shape distribution, Credible set, Modal Shape, Shape distributions.

AMS Classification: 62H99, 62E99

1 Introduction

The study of the shape of an object is essential in many scientific contexts, from biology and medicine, to image analysis, geography, archaeology, genetics and many more. The field of geometrical shape analysis was initially developed from a biological point of view by Thompson (1917), while the mathematical concept of shape was defined by Kendall (1977) as "the geometrical information that remains when location, scale and rotational effects are filtered out from an object". Then two objects will have the same shape if they can be translated, rescaled and rotated to each other so that they match exactly, in the sense that they are similar objects.

Shapes are typically described by locating a finite number of points from the object which are called landmarks. A landmark is a point of correspondence on each object that matches between and within populations.

To formulate these ideas, let $z^o = (z_1, z_2, ..., z_k)^T$ denote the landmarks of an object in the plane, where $z_i$, $i = 1, ..., k$, is complex, with real part the x-coordinate and imaginary part the y-coordinate of the $i^{th}$ landmark. When performing analysis of shape, we take out the effects of translation, dilation and rotation in order to obtain a correct geometrical view of the shape.

In order to remove translation, it is sufficient to take any set of $k$ independent contrasts of the shape $z^o$. Thus, define $z_H = H z^o$, the Helmertized landmarks of $z^o$, where $H$ is the last $(k - 1)$ rows of the Helmert matrix. In order to remove scaling effects, we simply take the standardized version of $z_H$, i.e., we define $z_S = \frac{z_H}{\|z_H\|}$, where $\|z_H\|^2 = z_H^* z_H$, and $z_H^*$ denotes the conjugate transpose of $z_H$. The complex vector $z_S$ is referred to as the pre-shape of the shape $z^o$. Note that $z_S$ lies in the complex sphere, and hence the use of distributions on the complex sphere of $(k - 1)$-dimensions ($CS^{k-1}$), is the natural approach in shape analysis.

\textsuperscript{1}Athanassios C. Micheas is Assistant Professor, Department of Statistics, University of Missouri-Columbia.
\textsuperscript{2}Dipak K. Dey is Professor, Department of Statistics, University of Connecticut.
\textsuperscript{3}Kanti V. Mardia is Professor, Department of Statistics, University of Leeds, Leeds, UK.
Sometimes it is useful to preserve k-dimensions usually for sake of presentation. This can be done by taking $z_c = H^T z_S$.

Finally, in order to remove rotation, we choose only at the final stage, an appropriate angle $\psi$, that satisfies specific needs of the analysis. Since we have removed rotation effects, we need a stochastic mechanism for the pre-shape that eliminates rotational effects. Hence, the distribution of a pre-shape $z_S$ must be invariant under scalar rotation, with $z_S$ and $e^{i\psi} z_S$ having the same distribution for any $\psi$. This is the property of complex symmetry for a complex distribution and provides a mathematically appealing framework as we show below. Such distributions are quite useful in practice as well since they eliminate automatically rotational effects that occur in datasets. However, if the goal of the experiment is to capture rotational effects in the data instead of the assessment of shape, this would not be a desired property.

Hence, starting with an original configuration $z^o$, we can remove effects of location, scale, and rotation by taking

$$z = e^{i\psi_o} \frac{Hz^o}{\|Hz^o\|}$$

for some real $\psi_o$.

In statistical shape analysis we are interested in the development of methods that help obtain a modal (average) shape that summarizes the shape information contained in a population of shapes; compare populations of shapes; employ an appropriate probability distribution that would describe shape stochastically; capture shape variability through Credible/Confidence Sets and finally classify a new shape into one of two or more populations of shapes via discriminant rules. There are several methods that can be employed to answer these questions. Landmark based as well as some non-landmark based methodologies can be found in the books by Bookstein (1991) and Stoyan and Stoyan (1994, part II). A more mathematical treatment can be found in the book by Small (1996). For a review on the methods that have appeared in the literature, we refer to the book by Dryden and Mardia (1998) and the references therein. Some of the contributions in the literature include Kendall (1977, 1989), Bookstein (1996), Kent (1994, 1997), Mardia and Dryden (1999) and Micheas and Dey (2003, 2005). In Micheas and Dey (2005), the authors provide methods for assessing differences in populations of shapes based on modal shapes, while in Micheas and Dey (2003) a treatment of the classification problem is given. Here, we are mainly interested in obtaining an object with shape that will summarize the shapes of the objects in the population as well as capture shape variability.

In section 2 we introduce the elliptical family of distributions on the complex sphere and explore various important cases that are used to model shape. We also investigate properties of this family of shape distributions. Section 3 is concerned with estimation of pre-shapes from a classical as well as a Bayesian perspective. The methods are later applied in section 4, for the mouse vertebrae data. Section 5 contains some concluding remarks.

2 Complex Elliptical family of distributions

The elliptical family of distributions is quite common in modeling multivariate data, since it contains many standard multivariate models. It was defined and explored by Kelker
(1970); for more details on the real elliptical family of distributions and its applications we refer to Fang and Zhang (1990), Gupta and Varga (1993) and Fang and Li (1999).

We will construct the complex elliptical family of distributions via extension of the elliptical family of real distributions, which as we will see, have applications in shape analysis. For an alternative construction of this family we refer the reader to Krishnaiah and Lin(1986).

Assume that a vector $\mathbf{W} = \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}$ has a $2p$-variate real elliptical distribution with mean vector $\mu = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}$ and covariance matrix $\Sigma = \begin{bmatrix} \Gamma & -\Phi \\ \Phi & \Gamma \end{bmatrix}$, where $\Gamma$ is positive definite and $\Phi = -\Phi^T$ (skew symmetric).

The density of the elliptical family of real distributions is defined to be of the form

$$f(\mathbf{w}|\Sigma) = c_p|\Sigma|^{-\frac{1}{2}}g((\mathbf{w} - \mu)^T\Sigma^{-1}(\mathbf{w} - \mu)),$$

where $|\Sigma|$ the determinant of the matrix $\Sigma$ and $c_p$ the normalizing constant that is computed by the usual condition for the generator function $g(.)$,

$$\int_0^\infty c_p\frac{\pi^p}{\Gamma(p)}u^{p-1}g(u)du = 1.$$ 

The usual definition of the distribution that appears in the literature, requires $\Sigma$ to be positive definite and the generator function $g$ to be decreasing. The case of a negative definite $\Sigma$ can be easily obtained by selecting a generator $h(x) = g(-x)$, thus forcing $h$ to be increasing. For what follows we will consider $\Sigma$ to be positive definite, the generator function $g$ increasing or decreasing and we write $\mathbf{W} \sim El_p(\nu, \Sigma; g)$ to denote the $p$-variate real elliptical distribution.

We consider an extension of the real elliptical family of distributions to the complex elliptical family of distributions, similar to the extension of the Normal distribution to the complex Normal distribution, as defined in Anderson(1971). More details on this distribution can be found in Andersen et al.(1995).

Recall that if $\mathbf{W} \sim El_p(\nu, \Sigma; g)$ then $\mathbf{AW} \sim El_{p\text{rank}(A)}(A\nu, APA^T; g)$. Let $A = \begin{bmatrix} \mathbf{I}_p & \mathbf{I}_p \end{bmatrix}$ and define $\mathbf{Z} = \mathbf{AW} = \mathbf{X} + i\mathbf{Y}$, $\mu = A\nu = \mu_x + i\mu_y = E(Z)$ and $\mathbf{Q} = APA^* = 2\Sigma + 2\Phi i$. Then $\mathbf{Z}$ is said to have a complex elliptical distribution with mean $\mu$ and matrix $\mathbf{Q}$, where $\mathbf{Q}$ is Hermitian, where $\mathbf{Q}^* = \mathbf{Q}$, with $\mathbf{Q}^*$ the conjugate transpose of $\mathbf{Q}$. The density will be in this case of the form

$$f(z|\mu, \mathbf{Q}) = c(\lambda)g((z - \mu)^*\mathbf{Q}^{-1}(z - \mu))$$

where $\lambda = (\lambda_1, ..., \lambda_p)$ the vector of eigenvalues of $\mathbf{Q}^{-1}$ and $c(\lambda)$ is the normalizing constant. We write $\mathbf{Z} \sim CEl_p(\mu, \mathbf{Q}; g)$, to denote the complex elliptical distribution.

Most properties of the real elliptical family of distributions, are carried over by this construction, to the complex elliptical family of distributions. We list some of these properties below.

- If $\mathbf{Z} \sim CEl_p(\mu, \Sigma; g)$ then $\mathbf{BZ} \sim CEl_r(B\mu, B\Sigma B^*; g_r)$, where $\mathbf{B}$ is an $r \times p$ matrix of rank $r \leq p$. 

3
• If $Z = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \sim CEl_p\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}\right)$, $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$ with $\Sigma$ Hermitian, $Z_1$ a $k \times 1$ vector and $Z_2$ a $(p-k) \times 1$ vector then $Z_1 \sim CEl_k(\mu_1, \Sigma_{11}; g_k)$ and $Z_2 \sim CEl_{p-k}(\mu_2, \Sigma_{22}; g_{p-k})$.

• The Uniform distribution on $CS^p$, is the unique distribution on $CS^p$ which is invariant under unitary transformations. The proof follows from the corresponding result for the real spherical family of distributions, see Eaton (1977) for invariance and Dempster (1969) for uniqueness.

• If $Z \sim CEl_p(0_p, I_p; \mu; g)$ then $Z_{[Z]} \sim CEl_p(0_p, I_p; \mu; g_1(x) = g(1))$, i.e., $Z_{[Z]}$ is the Uniform distribution on $CS^p$. The proof of this property follows from the latter property above.

The following Theorem, provides a useful result about the complex elliptical family of distributions. Recall that the analog of this in the case of the real elliptical family of distributions is that if $x \sim El_p(\mu, A; g)$, then the distribution of the quadratic $q = x^TA^{-1}x$, is $f(q) = k_p\frac{2^p}{\Gamma(p)}q^{p-1}g(q)$, where $q > 0$, $A$ positive definite and $g$ decreasing.

**Theorem 2.1** Assume that $z \sim CEl_p(0, \Sigma; g)$. The distribution of the quadratic form $q = z^*\Sigma^{-1}z$ is real and given by

$$f(q) = c_p\frac{2^p}{\Gamma(p)}q^{p-1}g(2q),$$

where $q \geq 0$, $\Sigma^{-1}$ is positive definite and $g$ is decreasing. The constant $c_p$ satisfies

$$c_p^{-1} = \int_0^\infty \frac{2^p}{\Gamma(p)}q^{p-1}g(2q)\,dq.$$

**Proof.** The proof is given in the appendix. ■

### 2.1 Complex Elliptical family of shape distributions

Consider $k$-landmarks and assume that we have taken location and scale effects out of the original shapes $z^0$, that is we have the pre-shapes $z_S = \frac{z^0}{\|z^0\|}$ in the $(k-1)$-dimensional complex sphere, with $z_S \in CS^{k-1}$, $z_S^*z_S = 1$. In order to remove rotation effects, we consider any complex distribution $f(\cdot)$ on $CS^{k-1}$, that has the property $f(e^\theta z_S) = f(z_S)$, thus invariant under rotations of the pre-shape $z_S$. Consequently, if an object is rotated, it has the same density and will contribute in an identical way to inference as the same object in the original rotation. This property makes the distribution suitable for shape analysis.

The density of a shape distribution is given by

$$f(z_S | \lambda) = C(\lambda)g(z_S^*\Sigma^{-1}z_S) = C(\lambda)g(\sum_{i=1}^{k-1} \lambda_i z_S^i z_S^i),$$

(2.1)
where \( z_S = (z_{s_1}, \ldots, z_{s_{k-1}})^T \in CS^{k-1} \), the pre-shape, \( \Sigma^{-1} \) is a Hermitian, positive (or negative) definite matrix with \( g \) decreasing (or increasing), \( \lambda = (\lambda_1, \ldots, \lambda_{k-1}) \) the vector of eigenvalues of \( \Sigma^{-1} \) and \( C(\Lambda) \) is the normalizing constant. We will write \( SEl_{k-1}(0, \Sigma; g) \) to denote a shape elliptical distribution on the complex sphere, where \( S \) stands for shape. Notice that \( f(e^{\Phi} z_S) = f(z_S), \forall \psi \in \mathcal{R} \) and thus a distribution of this form is suitable for shape analysis, and the distributions \( SEl_{k-1}(0, \Sigma; g) \) and \( SEl_{k-1}(0, \Lambda; g) \), where \( \Lambda^{-1} = diag(\lambda_1, \ldots, \lambda_k) \), have the same normalizing constant.

Now, in order to model pre-shapes in the \((k - 1)\)-dimensional complex sphere we obtain shape distributions by considering conditioning a complex elliptical distribution to have unit modulus or by defining a complex distribution on the complex sphere directly, assuming an underlying complex distribution. In addition any complex distribution has its corresponding real distribution. Thus shape distributions are constructed here by any of the two following ways:

- Let \( W = \left( \begin{array}{c} X \\ Y \end{array} \right) \sim Ell_{2(k-1)}(0, \begin{bmatrix} \Gamma & -\Phi \\ \Phi & \Gamma \end{bmatrix}; g) \) and define \( U = X + iY \sim C\ell_{k-1}(0, \Sigma = 2\Gamma + 2\Phi; g) \). The shape distribution is obtained in this case by conditioning the complex distribution to have unit modulus, i.e., \( Z_S = U = X + iY \mid ||U|| = 1 \sim SEl_{k-1}(0, \Sigma; g), Z_S \in CS^{k-1} \). Notice that the generator function \( g \) remains the same, whereas the normalizing constant will change.

- We define directly \( Z_S \sim SEl_{k-1}(0, \Sigma; g), Z_S \in CS^{k-1} \). In this case, the underlying real elliptical distribution lies in \( S^{2(k-1)} \), the real \( 2(k - 1) \) dimensional sphere.

The complex elliptical family of distributions provides a stochastic mechanism for modeling shapes that is first: a generalization to the most commonly used shape distributions in the literature thus far, like the offset normal, complex Watson and complex Bingham distributions, and second: it can be used for modeling shapes, where the isotropic covariance structure for the data is implausible (an assumption required for the complex Watson distribution).

Mardia and Dryden(1999) mention the use of a similar class of shape distributions, when an isotropic covariance structure can be assumed for the shape data; they consider densities of the form

\[
c(A) \exp \{ \varphi(z^* A z) \}
\]

where \( A \) is Hermitian and \( \varphi \) a generator function. This class of distributions is a reparameterization of (2.1), and can be treated in a same way as (2.1).

Lemma 2.1, given in the appendix, provides computation of the normalizing constant and the quadratic form \( q = z_S^* \Sigma^{-1} z_S \).

Now, the most likely pre-shape is the one that maximizes \( f(z_S|\lambda) \) subject to \( ||z_S|| = 1 \). It is easily seen, that for \( g \) strictly increasing (decreasing) the modal shape is the (least) dominant eigenvector of \( \Sigma^{-1} \), denoted by \( \gamma_{k-1} (\gamma_1) \). Hence the most likely pre-shape depends on the eigenvectors of \( \Sigma^{-1} \) and thus we concentrate on the estimation of \( \Sigma^{-1} \).

Next we briefly present some of the most important shape distributions which have been effectively used in the literature and are special cases of the elliptical family of shape
distributions. All the constants for the shape distributions below, can be easily obtained using Lemma 2.1.

First consider the complex Bingham distribution on \( CS^{k-1} \), which is denoted by \( CB_{k-1}(\Sigma) \), and has probability density function

\[
f(z_S|\lambda) = \left(2\pi^{-k+1} \sum_{j=1}^{k-1} \frac{\exp(\lambda_j)}{\prod_{i \neq j}(\lambda_i - \lambda_j)} \right)^{-1} \exp(z^*_S \Sigma z_S), \; z_S \in CS^{k-1}
\]

where \( \Sigma \) is Hermitian. The complex Bingham distribution provides a nice framework for the analysis of two dimensional shape data and it was proposed and studied by Kent(1994). Notice that the matrices \( \Sigma \) and \( \Sigma + \alpha I \) define the same complex Bingham distribution, since \( z^*_S z_S = 1 \).

The complex Bingham distribution is a special case of the complex elliptical family of distributions, if we let in (2.1) \( g(x) = e^x \). Dryden and Mardia(1998), provide the maximum likelihood estimators of the eigenvalues of \( \Sigma^{-1} \) based on \( S \), the sample variance-covariance matrix of a sample of \( n \) pre-shapes \( z_1, \ldots, z_n \), where \( S = \sum_{i=1}^n z_i z_i^* \). Finally, the complex Bingham distribution, can be obtained through conditioning a zero mean complex multivariate Normal distribution to have unit modulus. More precisely, if \( z \sim CN_{k-1}(0, A) \) then \( \|z\| = 1 \sim CB_{k-1}(-A^{-1}) \). Since the generator function is increasing, the modal shape in this case will be the dominant eigenvector of \( \Sigma \).

The complex Watson distribution on \( CS^{k-1} \) defined by Mardia and Dryden(1999), is denoted by \( CW_{k-1}(\mu, \xi) \), and has probability density function

\[
f(z_S|\xi, \mu) = \left(2\pi^{-k+1}\xi^{2-k}\left[\exp(\xi) - \sum_{r=0}^{k-3} \frac{\xi^r}{r!}\right] \right)^{-1} \exp(\xi \|z_S^*\|^2), \; z_S \in CS^{k-1}
\]

where \( \xi \in \mathcal{R} \) is the concentration parameter and \( \mu \) the mean pre-shape. If \( \xi > 0 \) the modes are obtained at \( e^{i\theta} \mu, 0 \leq \theta < 2\pi \), which is useful in shape analysis for representing population shape. If \( \xi < 0 \) the distribution has modes at all vectors orthogonal to \( \mu \) and for \( \xi = 0 \) the distribution is reduced to the uniform distribution on \( CS^{k-1} \). The complex Watson distribution is a special case of the complex Bingham distribution if we let \( \Sigma \) in (2.2) have two distinct eigenvalues, that is we let \( \Sigma = -\xi(I_{k-1} - \mu\mu^*) \). All eigenvalues are equal except for the largest one which is different. The derivation of the normalizing constant \( C(\lambda) \), even though the complex Watson distribution is a special case of the complex Bingham distribution, is not identical to the derivation of the constant for the Bingham distribution. For further details we refer the reader to Mardia and Dryden(1999). The authors also provide the maximum likelihood estimators for \( \xi \) and \( \mu \). The generator function is increasing, and hence the modal shape will be the dominant eigenvector of \( \Sigma = -\xi(I_{k-1} - \mu\mu^*) \), which is \( \mu \).

Finally, assume that shape is modeled by a complex angular central Gaussian distribution. Then the density is given by

\[
f(z_S|\Sigma) = \frac{(k - 2)!}{2\pi^{k-1}|\Sigma|^{1-k}} (z^*_S \Sigma^{-1} z_S)^{-(k-1)}
\]

This distribution was studied first in a shape analysis context, by Kent(1994). The matrix \( \Sigma^{-1} \) is positive definite and Hermitian. Notice that the matrices \( \Sigma^{-1} \) and \( c\Sigma^{-1} \), where \( c \) is
some constant, define the same distribution. Kent (1997), showed that this distribution is more resistant to outliers than the complex Bingham and complex Watson distributions and he used EM-algorithm to compute the MLE of $\Sigma^{-1}$. It is a special case of elliptical family of distributions, with generator function

$$g(x) = x^{-(k-1)}, \ k > 1.$$  

Notice that the generator function is decreasing, and hence the modal shape will be the least dominant eigenvector of $\Sigma^{-1}$.

3 Estimation of pre-shape

Consider $n$-shapes, $z_1^n, z_2^n, \ldots, z_n^n$, randomly selected from a true shape and let $z_1, z_2, \ldots, z_n$, denote the $n$ corresponding pre-shapes respectively. We assume that the $n$ pre-shapes are identically distributed with density given in (2.1), and hence $z_i \sim SEL_{k-1}(0, \Sigma; g), i = 1, \ldots, n$.

Assume that the joint distribution of the sample can be written as

$$f(z|\Sigma) = C(\lambda_1, \ldots, \lambda_{k-1}; \lambda_1, \ldots, \lambda_{k-1})g(z^*(I_n \otimes \Sigma^{-1})z) = C(\lambda; \ldots; \lambda)g(\sum_{j=1}^{n} z_j^* \Sigma^{-1} z_j)$$

where $\lambda^T = [\lambda_1, \ldots, \lambda_{k-1}]$ the vector of eigenvalues of $\Sigma^{-1}$, and $z = (z_1, z_2, \ldots, z_n)$. The constant $C(\lambda; \ldots; \lambda) = C(\lambda_1, \ldots, \lambda_{k-1}; \lambda_1, \ldots, \lambda_{k-1})$ will depend on $\lambda$ since the eigenvalues of $I_n \otimes \Sigma^{-1}$ are the same as those of $\Sigma^{-1}$ with their multiplicity increased by $n$. In the case where the generator $g$ is of the form $g(x) = e^x$ or $g(x) = e^{-x}$, (3.1) coincides with the likelihood obtained by $n$ i.i.d. pre-shapes. In this case we get the most commonly used complex Watson, Bingham or offset Normal shape distributions. Thus, (3.1) is a generalization of these important cases.

The assumption of (3.1) does not come without criticism although it has been successfully used in the literature (see for example Sutradhar and Ali, 1989, Micheas and Dey, 2005). Moreover, it is the usual approach employed when working with samples from elliptical distributions, as illustrated in Fang and Zhang (1990). The major advantages of (3.1) are mathematical convenience and a weaker assumption than that of an independent sample, that may be more appropriate in certain scenarios.

3.1 Classical approach to estimation of pre-shape

We prove the following theorem, which provides calculation of the modes of $\lambda$ and their corresponding eigenvectors.

**Theorem 3.1** Assume that the generator function $g$ is strictly monotone. Then the modes of $\lambda_1 < \lambda_2 < \ldots < \lambda_{k-1}$ and $\gamma_1, \gamma_2, \ldots, \gamma_{k-1}$, the eigenvalues and corresponding eigenvectors of $\Sigma^{-1}$, are given by

$$\hat{\gamma}_i = v_i,$$
and \( \hat{\lambda} \) is obtained by finding the solution of the equation

\[
\frac{\partial \log C(\lambda; \ldots; \lambda)}{\partial \lambda_i} \bigg|_{\lambda=\hat{\lambda}} + \frac{\partial \log g(\sum_{j=1}^{k-1} \lambda_j \zeta_j)}{\partial \lambda_i} \bigg|_{\lambda=\hat{\lambda}} = 0
\]

for all \( i = 1, 2, \ldots, k-1 \), respectively, where \( \lambda = (\lambda_1, \lambda_2, \ldots, \lambda_{k-1}) \), \( 0 < \zeta_1 < \zeta_2 < \ldots < \zeta_{k-1} \) and \( \psi_1, \psi_2, \ldots, \psi_{k-1} \), are the eigenvalues and corresponding eigenvectors of \( S = \sum_{j=1}^{n} z_j z_j^* \).

**Proof.** The proof is given in the appendix. \( \blacksquare \)

Notice that in the case where the generator \( g \) is of the form \( g(x) = e^x \) or \( g(x) = e^{-x} \), the estimates above, are the maximum likelihood estimators of \( \lambda_1 < \lambda_2 < \ldots < \lambda_{k-1} \) and \( \gamma_1, \gamma_2, \ldots, \gamma_{k-1} \), the eigenvalues and corresponding eigenvectors of \( \Sigma^{-1} \).

### 3.2 Estimation of pre-shape from a Bayesian viewpoint

The Bayesian approach to estimating \( \Sigma^{-1} \), suggests that we have some prior belief \( \pi(\cdot) \), and we update this information from the data. Then we estimate \( \Sigma^{-1} \) by computing the mode of the posterior distribution, the MAP (Maximum a Posteriori) estimate. Often we consider a squared error loss function for estimation of \( \Sigma^{-1} \) and obtain the posterior mean as the Bayes estimator. Typically, exact calculation of the posterior distribution is not that straightforward. Then in order to obtain the mean of the posterior distribution \( \pi(\Sigma^{-1}|z) \), \( z = (z_1, z_2, \ldots, z_n) \) we use Monte Carlo method. The algorithm is as follows:

**Step 1.** Generate \( \Sigma_{(1)}^{-1}, \Sigma_{(2)}^{-1}, \ldots, \Sigma_{(L)}^{-1} \sim \pi(\cdot) \), where \( L \) is a large integer, where \( \pi(\cdot) \) the prior distribution of \( \Sigma^{-1} \).

**Step 2.** Compute

\[
f(z|\Sigma_{(i)}^{-1}) = \prod_{j=1}^{n} C(\lambda_{(i)}) g(z_j^* \Sigma_{(i)}^{-1} z_j)
\]

with \( \lambda_{(i)} = (\lambda_1, \ldots, \lambda_{k-1}) \) the vector of eigenvalues for \( \Sigma_{(i)}^{-1} \), for all \( i = 1, 2, \ldots, L \).

**Step 3.** The posterior estimate will be given by

\[
\Sigma^{-1} = \hat{E}(\Sigma^{-1}|z) \approx \frac{\sum_{i=1}^{L} f(z|\Sigma_{(i)}^{-1}) \Sigma_{(i)}^{-1}}{\sum_{i=1}^{L} f(z|\Sigma_{(i)}^{-1})}
\]  

(3.3)

Modeling the prior beliefs for \( \Sigma^{-1} \) is somewhat restricted, in the sense that there does not exist a wide range of distributions for random matrices. For a treatment on the subject of modeling priors for random matrices, we refer the reader to Daniels and Kass(1999).

Instead we consider modelling \( \lambda = (\lambda_1, \lambda_2, \ldots, \lambda_{k-1}) \) and \( \gamma_1, \gamma_2, \ldots, \gamma_{k-1} \), the eigenvalues and corresponding eigenvectors of \( \Sigma^{-1} \). This approach provides more flexibility and allows us to create more complex models to better describe our beliefs. Recall that when the generator function \( g \) is strictly increasing, the eigenvector \( \gamma_{k-1} \), that corresponds to the
largest eigenvalue of $\Sigma^{-1}$, is viewed as the average axis of the data. We desire a full posterior analysis for $\gamma_{k-1}$. From (3.1) we can write the joint distribution of a sample of pre-shapes in the form

$$f(z|\lambda, \gamma_1, \gamma_2, \ldots, \gamma_{k-1}) = C(\lambda; \ldots; \lambda)g(\sum_{j=1}^{n} z_j^T \Sigma^{-1} z_j) = C(\lambda; \ldots; \lambda)g(\sum_{j=1}^{k-1} \lambda_j \sum_{i=1}^{n} \gamma_i^* z_i^T \gamma_j),$$

and thus

$$f(z|\lambda, \gamma_1, \gamma_2, \ldots, \gamma_{k-1}) = C(\lambda; \ldots; \lambda)g(\sum_{j=1}^{k-1} \lambda_j \sum_{i=1}^{n} \|z_i^* \gamma_j\|^2) \quad (3.4)$$

or

$$f(z|\lambda, \gamma_1, \gamma_2, \ldots, \gamma_{k-1}) = C(\lambda; \ldots; \lambda)g(\sum_{j=1}^{k-1} \lambda_j \gamma_j^* S \gamma_j) \quad (3.5)$$

where $S = \sum_{i=1}^{n} z_i z_i^*$ and $z = (z_1, \ldots, z_n)$.

Now assume that $\gamma_{k-1} \sim SEL_{k-1}(0, \Sigma_{\text{prior}}; h)$, where the generator function $h$ is such that we obtain the conjugate prior for $\gamma_{k-1}$. For simplicity here we estimate the nuisance parameters $\lambda$ and $\gamma_j$, $j = 1, \ldots, k - 2$, by their corresponding MLE’s, although a marginal approach could be employed, that is, integrate out the nuisance parameters. Thus

$$f(z|\gamma_{k-1}) = C(\hat{\lambda}; \ldots; \hat{\lambda})g(\sum_{j=1}^{k-2} \hat{\lambda}_j \sum_{i=1}^{n} \|z_i^* \gamma_j\|^2 + \hat{\lambda}_{k-1} \sum_{i=1}^{n} \|z_i^* \gamma_{k-1}\|^2)$$

$$= C(\hat{\lambda}; \ldots; \hat{\lambda})g(\sum_{j=1}^{k-2} \hat{\lambda}_j \sum_{i=1}^{n} \|z_i^* \gamma_j\|^2 + \hat{\lambda}_{k-1} \sum_{i=1}^{n} \gamma_i^* z_i^T \gamma_j \gamma_{k-1})$$

$$= C_1(\hat{\lambda})g_1(\sum_{i=1}^{n} \gamma_i^* z_i \gamma_{k-1}) = C_1(\hat{\lambda})g_1(\gamma_{k-1}^* S \gamma_{k-1})$$

where $S = \sum_{i=1}^{n} z_i z_i^*$ and $g_1(.)$ a new generator function. Hence the posterior distribution of $\gamma_{k-1}|z$ must be $SEL_{k-1}(0, \Sigma_{\text{post}}; h)$ with density

$$\pi(\gamma_{k-1}|z) = C_2(\lambda_{\text{post}})h(\gamma_{k-1}^* \Sigma_{\text{post}}^{-1} \gamma_{k-1})$$

where $\lambda_{\text{post}}$ is the vector of eigenvalues of $\Sigma_{\text{post}}^{-1}$, with $\Sigma_{\text{post}}^{-1}$ known. $\Sigma_{\text{post}}^{-1}$ depends on both $\Sigma_{\text{prior}}$ and $S$. The generator function $h$ is such that

$$h(\gamma_{k-1}^* \Sigma_{\text{post}}^{-1} \gamma_{k-1}) = g_1(\gamma_{k-1}^* S \gamma_{k-1})h(\gamma_{k-1}^* \Sigma_{\text{prior}}^{-1} \gamma_{k-1})$$

and

$$\int h(\gamma_{k-1}^* \Sigma_{\text{post}}^{-1} \gamma_{k-1})d\gamma_{k-1} = |C_2(\lambda_{\text{post}})|^{-1}.$$

Thus all the analysis should be based on this posterior distribution. Then the MAP estimate, is the dominant eigenvector $\gamma_{k-1}$ of $\Sigma_{\text{post}}^{-1}$, when $h$ is increasing and the least dominant eigenvector $\gamma_1$ when $h$ decreasing. This provides a point estimate for modal
shape. Moreover, a 100(1 − α)% credible region \( R \) can be obtained for modal shape \( \gamma \), by solving for \( R \) in
\[
\int_R \pi(\gamma|z_1, \ldots, z_n)\,d\gamma = 1 - \alpha.
\]

Conjugacy provides a natural and appealing framework for Bayesian analysis in this context. Unfortunately, when considering more complicated models for the data, conjugate priors are hard to find or they might not even exist.

The following Theorem, allows us to consider any prior distribution, by providing a method for generating random vectors from a complex multivariate posterior distribution when the distribution of a real valued quadratic form is available. Since the examples we consider in section 4, depend on generating from a complex normal distribution, which is much easier to do, we do not use this approach here. We provide this theorem for completeness, as an alternative approach to generating values from more complicated than the complex normal, distributions.

**Theorem 3.2** Assume that \( x \sim SE_{\nu}(0, \Sigma; g) \) and define \( q = x^\ast \Sigma^{-1} x \). Assume that \( q_0 \) is a realization from
\[
f(q|\lambda) \propto g(q),
\]
where \( 0 \leq q \leq \sum_{j=1}^p \lambda_j \) if \( \Sigma^{-1} \) is positive definite and \( \sum_{j=1}^p \lambda_j \leq q \leq 0 \) if \( \Sigma^{-1} \) is negative definite. Then \( x = \sum_{i=1}^p \pm r_i \sqrt{\lambda_i} \lambda_i^{-\frac{1}{2}} v_i \), is a realization from \( x \sim SE_{\nu}(0, \Sigma; g) \), where \( \lambda_1 < \ldots < \lambda_p, v_1, \ldots, v_p \), are the eigenvalues and the corresponding eigenvectors of \( \Sigma^{-1} \) and \( \sum_{i=1}^p r_i = 1, 0 \leq r_i \leq 1, i = 1, \ldots, p \), some constants.

**Proof.** The proof is given in the appendix. ■

We consider the complex Watson shape distribution in order to illustrate the methods under a well known and commonly used model in the literature. Assuming that the pre-shapes of the original data are modeled by a complex Watson distribution, the likelihood of independent pre-shapes is of the form
\[
f(z|\Sigma) = C(\xi) \exp(\xi \sum_{i=1}^n \|z_i^\ast \mu\|^2)
\]
\[
= C(\xi) \exp(\xi^\ast \Sigma \mu), \ z = (z_1, \ldots, z_n), z_i \in CS^{k-1}.
\]

Here \( \Sigma = -\xi(1_{k-1} - \mu^\ast \mu) \), \( \xi > 0 \), with \( \Sigma \) negative definite and the generator function increasing. For what follows, we assume that \( \xi \) is known.

The conjugate prior is the complex Bingham distribution of the form
\[
\pi(\mu) \propto \exp(\mu^\ast \Sigma_{prior} \mu), \ \mu \in CS^{k-1}
\]
for some negative definite and Hermitian \( \Sigma_{prior} \), assumed to be known.
Then the posterior distribution is given by

$$
\pi(\mu | z) \propto \exp(\xi \mu^* S \mu) \exp(\mu^* \Sigma_{prior} \mu) \\
\propto \exp(\mu^* (\xi S + \Sigma_{prior}) \mu) \\
\propto \exp(\mu^* \Sigma_{post} \mu)
$$

(3.6)

which is a complex Bingham distribution with $\Sigma_{post} = \xi S + \Sigma_{prior}$. The MAP estimate in this case is the dominant eigenvector of $\Sigma_{post}$.

4 Application of methods

To illustrate our methods, we will investigate a biology experiment, where in order to assess the effects of body weight on the shape of mouse vertebrae, three groups of mice were collected: small, large and control. For further details we refer the reader to Dryden and Mardia (1998) and the references therein.

We consider only the small group of mouse vertebrae, that corresponds to mice selected for their small body weight. There are $n = 23$ small bones and $k = 6$ landmarks. In Figure 1, all 23 bones are displayed. For each landmark of each shape we create a complex number having as real part the x-coordinate and imaginary part the y-coordinate of the landmark. We also remove location and scale effects as described in previous sections. The figure contains the post-Helmertized versions of the pre-shapes rotated in such a way as to have the two furthest away landmarks of the shapes on the same horizontal line. We notice that we have high concentrations at each landmark for these shapes.

We assume that the pre-shapes are modeled according to a complex Watson distribution. This assumption will be reasonable when an isotropic covariance structure for the data is
plausible, where all except the largest eigenvalue of the covariance matrix are approximately equal. First we notice that a plot of the data in Figure 1, displays generally circular scatters of points for each landmark, which is required for an isotropic model like the complex Watson shape distribution.

Moreover, the eigenvalues of the complex sum of squares and products matrix $S = \sum_{i=1}^n z_i z_i^*$, are given by 22.905766887, 0.071857876, 0.012812261, 0.005388358 and 0.004174619. Notice that all eigenvalues except for the largest one are approximately equal. This provides us with another indication that the complex Watson model is appropriate to model the small vertebrae data. Also notice that the largest eigenvalue 22.905766887 is very close to $n = 23$, which provides another indication of high concentrations.

Recall that the posterior distribution of the MAP pre-shape is a complex Bingham distribution, $\mu z^*_S \sim C B_{k-1}(\Sigma_{\text{post}} - \xi S + \Sigma_{\text{prior}})$. The MAP estimate will be the dominant eigenvector of $\Sigma_{\text{post}}$. We select $\xi$ and $\Sigma_{\text{prior}}$ such that $\Sigma_{\text{post}} = \xi S + \Sigma_{\text{prior}}$ is negative definite and Hermitian.

If we let $\Sigma_{\text{prior}} = -c I_{k-1}$, for some constant $c > 0$, then the prior reduces to the uniform distribution on the $(k - 1)$-dimensional complex sphere. In this case the MAP estimate is equal to the MLE, which is the dominant eigenvector of $S$.

In Figure 2, we have the original data displayed along with the MLE pre-shape, that has value

$$MILE = (-0.0413 + 0.715i, -0.245 + 0.056i, -0.403 - 0.037i, -0.072 - 0.088i, 0.495 + 0.0i).$$

We can obtain 99% confidence ellipsoids for the average axis of the data, assuming a bivariate normal distribution at each landmark. Figure 3 displays the average in solid line and the confidence ellipsoids at each landmark.

Now, we let $\xi = 1$ and $\Sigma_{\text{prior}} = -c R$, where $R = [r_{ij}]$, with $r_{ii} = 4.9$, $r_{ij} = -4i$, for $j > i$, and $r_{ij} = 4i$, for $i > j$, $i, j = 1, 2, \ldots, 5$. We search for $c$ that yields negative...
Figure 3: 99% confidence ellipsoids for each landmark

definite and Hermitian $\Sigma_{post} = S + \Sigma_{prior}$. We can easily check that any $c \geq 4.6287$ will give a valid $\Sigma_{post}$. Increasing the value of this constant will allow us to increase variability of the generated values from the posterior distribution while closer to this bound, we will have highly concentrated generated shapes about the MAP estimate. We select $c_w = 4.629$ and generate 500 pre-shapes from $CB_{k-1}(\Sigma_{post} = S - 4.629R)$.

The MAP estimate of pre-shape is the dominant eigenvector of $\Sigma_{post} = S - 4.629R$, that has value

$$MAP = (-0.020 + 0.719i, -0.229 + 0.065i, -0.390 - 0.011i, -0.059 - 0.063i, 0.514 + 0.0i).$$

In Figure 4, we have the original data displayed along with the MLE and MAP pre-shapes. Using the generated values from the posterior distribution, we can compute a Bayesian credible set by considering an ordering of all generated shapes. The bounds of the $100(1-a)\%$ credible set correspond to the $a^{th}$ and $(1-a)^{th}$ percentiles of the ordered shapes. We order the generated shapes in the following way: we say that pre-shape $z_1$ is smaller than the pre-shape $z_2$ if and only if the post-Helmertized version of $z_1$ is contained within the post-Helmertized version of shape $z_2$. If two shapes intersect then we say they have the same order.

In Figure 5, we display the 90% HPD Credible Set along with the original data. The post-Helmertized lower and upper bounds are

$$lower = (-0.174 - 0.586i, -0.131 + 0.528i, 0.167 + 0.096i, 0.501 + 0.021i, 0.161 - 0.058i, -0.501 + 0.021i)$$
Figure 4: MLE = dotted line, MAP = solid line

and

\[
\text{upper} = (-0.223 - 0.747i, -0.168 + 0.674i, 0.213 + 0.123i, \\
0.639 + 0.027i, 0.205 - 0.074i, -0.640 + 0.027i)
\]

respectively.

To assert sensitivity, under the complex Watson model, to parameters \( \xi \) and \( \Sigma_{\text{prior}} \), satisfying at the same time all necessary conditions for \( \Sigma_{\text{prior}} \) and \( \Sigma_{\text{post}} \), we computed and compared MAP estimates, for a range of values of \( \xi \) and \( \Sigma_{\text{prior}} \). We considered \( \xi = .1, .5 \) and 1 and corresponding \( c \) ranging from .4629 to 15.462, 2.314 to 17.314, and 4.629 to 24.629, respectively. With this structure for \( R \), there is a relationship between \( \xi \) and the smallest constant we use in \( \Sigma_{\text{prior}} = -cR \), in order to obtain a valid \( \Sigma_{\text{post}} \). More precisely, any reduction of \( \xi \) will result in the same reduction of the value of the smallest constant \( c \). We also noticed that increasing the values of \( \xi \) will yield a more resistant MAP estimate to the constant \( c \) we choose. This is not unusual, since \( \xi \) is the concentration parameter of the complex Watson distribution that we used in modelling this data.

5 Conclusion

We introduced and investigated shape distributions on the complex sphere. The natural extension we used, of the real elliptical family of distributions to the complex elliptical family of distributions, provided an appealing framework for shape analysis, and we generalized the methodology typically used in this context. Results from the real elliptical family were used to produce similar results for the complex elliptical family of distributions.

We reported only but a few models for the mouse vertebrae problem because of space constraint. Using our methodology, one can easily apply a large variety of statistical shape distributions to any shape analysis situation.
Appendix

**Proof of Theorem 2.1** Since \( Z = X + iY \sim \mathcal{E}_{l_p}(0, \Sigma = 2\Gamma + 2\Phi i = Q + Ri; g) \), with \( Q = 2\Gamma, R = 2\Phi \), we have

\[
    f(z|\Sigma) = c(\lambda)g(z^*\Sigma^{-1}z),
\]

where \( \lambda \) the vector of eigenvalues of \( \Sigma \). The underlying real elliptical distribution of \( W = \begin{pmatrix} X \\ Y \end{pmatrix} \) will be

\[
    W \sim \mathcal{E}_{l_2}(0, P = \begin{bmatrix} \Gamma & -\Phi \\ \Phi & \Gamma \end{bmatrix}; g).
\]

It is easy to see that

\[
    \Sigma^{-1} = (2\Gamma + 2\Phi i)^{-1} = (Q + Ri)^{-1}
    = (Q + RQ^{-1}R)^{-1} - iQ^{-1}R(Q + RQ^{-1}R)^{-1}
\]

with \( \Sigma^{-1} \) Hermitian. Then \( g = z^*\Sigma^{-1}z = x^T(Q + RQ^{-1}R)^{-1}x + y^T(Q + RQ^{-1}R)^{-1}y \), using the fact that \( \Sigma^{-1} \) is Hermitian. Now it can be shown that

\[
    P^{-1} = \begin{bmatrix}
    \Gamma & -\Phi \\ \Phi & \Gamma
    \end{bmatrix}^{-1}
    = \begin{bmatrix}
    (\Gamma + \Phi\Gamma^{-1}\Phi)^{-1} & \Gamma^{-1}\Phi(\Gamma + \Phi\Gamma^{-1}\Phi)^{-1} \\
    -\Gamma^{-1}\Phi(\Gamma + \Phi\Gamma^{-1}\Phi)^{-1} & (\Gamma + \Phi\Gamma^{-1}\Phi)^{-1}
    \end{bmatrix}
\]
with $P^{-1}$ skew symmetric, and thus
\[
\Gamma^{-1} \Phi (\Gamma + \Phi \Gamma^{-1} \Phi)^{-1} = -((-\Gamma^{-1} \Phi (\Gamma + \Phi \Gamma^{-1} \Phi)^{-1})^T
= (\Gamma + \Phi \Gamma^{-1} \Phi)^{-1} \Phi^T \Gamma^{-1} = -((\Gamma + \Phi \Gamma^{-1} \Phi)^{-1} \Phi \Gamma^{-1}).
\]
Hence
\[
u = W^T P^{-1} W = x^T (\Gamma + \Phi \Gamma^{-1} \Phi)^{-1} x + y^T (\Gamma + \Phi \Gamma^{-1} \Phi)^{-1} y
= 2x^T (Q + R \gamma^{-1} R)^{-1} x + 2y^T (Q + R \gamma^{-1} R)^{-1} y
= 2z^* \Sigma^{-1} z
\]
using the fact that $P^{-1}$ is skew symmetric, and since $W \sim E L_{2p}(0, P = \begin{bmatrix} \Gamma & -\Phi \\ \Phi & \Gamma \end{bmatrix}; g)$ we have
\[
f(w|P) = c_p |P|^{-\frac{p}{2}} g(w^T P^{-1} w),
\]
and hence
\[
f(u) = \frac{\pi^{\frac{p}{2}}}{\Gamma(p)} c_p u^{\frac{p}{2} - 1} g(u).
\]
Thus, the distribution of the quadratic form $q = z^* \Sigma^{-1} z = \frac{u}{\gamma}$ will be
\[
f(q) = \frac{\pi^{p}}{\Gamma(p)} 2^{p-1} c_p q^{p-1} g(2q)2
= \frac{2^p \pi^p}{\Gamma(p)} c_p q^{p-1} g(2q),
\]
where $q \geq 0$ and $g$ decreasing. Finally,
\[
c_p^{-1} = \int_0^\infty \frac{2^p \pi^p}{\Gamma(p)} q^{p-1} g(2q) dq.
\]

**Lemma 2.1** Assume that $Z = (Z_1, ..., Z_{k-1}) \sim E L_{k-1}(0, \Sigma; g)$, $Z \in CS^{k-1}$. Then the constant in (2.1) is given by
\[
C^{-1}(\lambda) = 2\pi^{k-1} \sum_{j=1}^{k-1} \left[ \prod_{i \neq j}^{k-1} (\lambda_j - \lambda_i)^{-1} \right] G^{[k-2]}(\lambda_j)
\]
where $G^{[k-2]}(x)$ is the anti-derivative of $g$ of order $k - 2$, i.e.,
\[
\frac{d^{k-2}}{dx^{k-2}} [G^{[k-2]}(x)] = g(x).
\]
The distribution of the quadratic form $q = z^* \Sigma^{-1} z$ will be
\[
f(q|\lambda) = [G^{[1]}(\lambda_o) - G^{[1]}(0)]^{-1} g(q), \ 0 \leq q \leq \lambda_o,
\]
16
when $\Sigma^{-1}$ is positive definite, and
\[
f(q|\lambda) = [G_q^\lambda(0) - G_q^\lambda(\lambda_o)]^{-1} g(q), \; \lambda_o \leq q \leq 0,
\]
when $\Sigma^{-1}$ is negative definite, where $\lambda_o = \sum_{j=1}^{k-1} \lambda_j$.

**Proof of Lemma 2.1** Assume that $Z = (Z_1, \ldots, Z_{k-1}) \sim SE_{k-1}(0, \Sigma; g)$, and consider Kent’s(1994) polar coordinates. We have
\[
\text{Re}(Z_j) = s_j^\frac{k}{2} \cos(\theta_j),
\]
\[
\text{Im}(Z_j) = s_j^\frac{k}{2} \sin(\theta_j), \; j = 1, \ldots, k-1
\]
where $s = (s_1, \ldots, s_{k-1}) \in S_{k-1} = \{(s_1, \ldots, s_{k-1}) : \sum_{j=1}^{k-1} s_j = 1, 0 \leq s_j \leq 1\}$, the $k-1$ dimensional unit simplex, and $\theta_j \in [0, 2\pi)$, $j = 1, \ldots, k-1$. We notice that \(\sum_{j=1}^{k-1} z_j^s = \sum_{j=1}^{k-1} s_j = 1\), and $s$ is independent of $\theta = (\theta_1, \ldots, \theta_{k-1})$. Furthermore, any $z = (z_1, \ldots, z_{k-1}) \in CS^{k-1}$ can be identified by some $s \in S_{k-1}$ and $\theta \in [0, 2\pi)^{k-1}$, since $CS^{k-1} = S_{k-1} \times [0, 2\pi)^{k-1}$. The Jacobian of the transformation $z \mapsto (s, \theta)$, is $2^{2-k}$, and hence we obtain the joint distribution of $(s, \theta)$ from (2.1) as
\[
df(s, \theta|\lambda) = 2^{2-k} C(\lambda) g\left(\sum_{j=1}^{k-1} \lambda_j s_j\right) ds_1 \ldots ds_{k-2} d\theta_1 \ldots d\theta_{k-1}.
\]
Integrating $\theta$ out, we obtain
\[
f(s|\lambda) = 2\pi^{k-1} C(\lambda) g\left(\sum_{j=1}^{k-1} \lambda_j s_j\right), \; s \in S_{k-1}.
\]
For $s' = (s_1, \ldots, s_{k-2}) \in S'_{k-2} = \{s' = (s_1, \ldots, s_{k-2}) : 0 \leq \sum_{j=1}^{k-2} s_j \leq 1, 0 \leq s_j \leq 1, j = 1, \ldots, k-2\}$, we have
\[
f(s'|\lambda) = 2\pi^{k-1} C(\lambda) g\left(\sum_{j=1}^{k-2} \lambda_j s_j + \lambda_{k-1} s_{k-1}\right)
\]
\[
= 2\pi^{k-1} C(\lambda) g\left(\sum_{j=1}^{k-2} \lambda_j s_j + \lambda_{k-1}(1 - s_1 - \ldots - s_{k-2})\right)
\]
\[
= 2\pi^{k-1} C(\lambda) g\left(\sum_{j=1}^{k-2} (\lambda_j - \lambda_{k-1}) s_j + \lambda_{k-1}\right).
\]
The derivation of the normalizing constant follows after lengthy algebraic manipulations, using induction on \( k - 1 \). We have

\[
C^{-1}(\lambda) = 2\pi^{k-1} \sum_{j=1}^{k-1} \left[ \prod_{\ell=1}^{k-1} (\lambda_j - \lambda_{\ell})^{-1} \right] G^{[k-1]}(\lambda_j)
\]

where \( G^{[k-1]}(x) \) such that \( \frac{d^{k-2}}{dx^{k-2}} \left[ G^{[k-2]}(x) \right] = g(x) \).

Next we consider the distribution of the quadratic form. We split the proof in two parts by considering first \( \Sigma^{-1} \) positive definite and later \( \Sigma^{-1} \) negative definite. Assume first that \( \Sigma^{-1} \) is positive definite, with eigenvalues \( 0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{k-2} \leq \lambda_{k-1} \). We have

\[
q = z^* \Sigma^{-1} z = \sum_{j=1}^{k-1} \lambda_j s_j = \sum_{j=2}^{k-1} (\lambda_j - \lambda_1) s_j + \lambda_1. \quad \text{Let } q_j = (\lambda_j - \lambda_1) s_j, \quad j = 3, \ldots, k - 1.
\]

Then the Jacobian of the transformation \((s_2, \ldots, s_{k-1}) \mapsto (q_3, \ldots, q_{k-1})\) is given by

\[
\frac{\partial s_2 \cdots \partial s_{k-1}}{\partial q_3 \partial s_3 \cdots \partial q_{k-1}} = \prod_{j=2}^{k-1} (\lambda_j - \lambda_1)^{-1}
\]

and hence

\[
f(q, q_3, \ldots, q_{k-1}|\lambda) = 2\pi^{k-1} C(\lambda) g(q) \prod_{j=2}^{k-1} (\lambda_j - \lambda_1)^{-1}
\]

where \((q, q_3, \ldots, q_{k-1}) \in T_{k-1} = \{(q, q_3, \ldots, q_{k-1}) : 0 \leq q_j \leq \lambda_j - \lambda_1, \quad j = 3, \ldots, k - 1, q - \lambda_2 \leq \sum_{j=3}^{k-1} q_j \leq q - \lambda_1\}. \quad \text{Let } Q_{k-1}^d = \{(q_3, \ldots, q_{k-1}) : 0 \leq q_j \leq \lambda_j - \lambda_1, \quad j = 3, \ldots, k - 1, q - \lambda_2 \leq \sum_{j=3}^{k-1} q_j \leq q - \lambda_1\}. \quad \text{Then the distribution of the quadratic form } q = z^* \Sigma^{-1} z, \text{ is given by}

\[
f(q|\lambda) = 2\pi^{k-1} C(\lambda) \left[ \prod_{j=2}^{k-1} (\lambda_j - \lambda_1)^{-1} \right] Vol(Q_{k-1}^d) g(q),
\]

\[
0 \leq q \leq \sum_{j=1}^{k-1} \lambda_j = \lambda_\omega.
\]

Since

\[
\left[ Vol(Q_{k-1}^d) \right]^{-1} = \int_0^{\lambda_\omega} 2\pi^{k-1} C(\lambda) \left[ \prod_{j=2}^{k-1} (\lambda_j - \lambda_1)^{-1} \right] g(q) dq
\]

\[
= 2\pi^{k-1} C(\lambda) \left[ \prod_{j=2}^{k-1} (\lambda_j - \lambda_1)^{-1} \right] \left[ G^{[k]}(\lambda_\omega) - G^{[k]}(0) \right],
\]

we finally have

\[
f(q|\lambda) = \left[ G^{[k]}(\lambda_\omega) - G^{[k]}(0) \right]^{-1} g(q), \quad 0 \leq q \leq \lambda_\omega.
\]
Next, assume that $\Sigma^{-1}$ is negative definite, i.e., $\lambda_1 \leq \lambda_2 \ldots \leq \lambda_{k-2} \leq \lambda_{k-1} \leq 0$. We follow similar steps as in the positive definite case, and consider the transformation

\[ q = z^* \Sigma^{-1} z = \sum_{j=1}^{k-1} \lambda_j s_j = \sum_{j=1}^{k-2} (\lambda_j - \lambda_{k-1}) s_j + \lambda_{k-1}, \quad q_j = (\lambda_j - \lambda_{k-1}) s_j, \quad j = 1, \ldots, k-3. \]

We have

\[ q = \sum_{j=1}^{k-3} q_j + (\lambda_{k-2} - \lambda_{k-1}) s_{k-2} + \lambda_{k-1} \iff s_{k-2} = \frac{q - \sum_{j=1}^{k-3} q_j - \lambda_{k-1}}{\lambda_{k-2} - \lambda_{k-1}}. \]

Then the Jacobian of the transformation $(s_1, \ldots, s_{k-2}) \mapsto (q, q_1, \ldots, q_{k-3})$ is given by

\[
\frac{\partial s_1 \ldots \partial s_{k-2}}{\partial q \partial q_1 \ldots \partial q_{k-3}} = \prod_{j=1}^{k-2} (\lambda_j - \lambda_{k-1})^{-1}
\]

and hence

\[
f(q, q_1, \ldots, q_{k-3} | \lambda) = 2\pi^{k-1} C(\lambda) g(q) \prod_{j=1}^{k-2} (\lambda_j - \lambda_{k-1})^{-1}
\]

where $(q, q_1, \ldots, q_{k-3}) \in T_{k-3} = \{(q, q_1, \ldots, q_{k-3}) : \lambda_j - \lambda_{k-1} \leq q_j \leq 0, \quad j = 1, \ldots, k-3, \quad q - \lambda_{k-1} \leq \sum_{j=1}^{k-3} q_j \leq q - \lambda_{k-2}\}$. Let $Q_{k-1}^n = \{(q_1, \ldots, q_{k-3}) : \lambda_j - \lambda_{k-1} \leq q_j \leq 0, \quad j = 1, \ldots, k-3, \quad q - \lambda_{k-1} \leq \sum_{j=1}^{k-3} q_j \leq q - \lambda_{k-2}\}$. Thus the distribution of the quadratic form $q = z^* \Sigma^{-1} z$, is given by

\[
f(q | \lambda) = 2\pi^{k-1} C(\lambda) \left[ \prod_{j=1}^{k-2} (\lambda_j - \lambda_{k-1})^{-1} \right] \text{Vol}(Q_{k-1}^n) g(q),
\]

\[
\lambda_o = \sum_{j=1}^{k-1} \lambda_j \leq q \leq 0.
\]

Since

\[
[\text{Vol}(Q_{k-1}^n)]^{-1} = \int_{\lambda_o}^{0} 2\pi^{k-1} C(\lambda) \left[ \prod_{j=1}^{k-2} (\lambda_j - \lambda_{k-1})^{-1} \right] g(q) dq \]

\[
= 2\pi^{k-1} C(\lambda) \left[ \prod_{j=1}^{k-2} (\lambda_j - \lambda_{k-1})^{-1} \right] [G^{[1]}(0) - G^{[1]}(\lambda_o)],
\]

we finally have

\[
f(q | \lambda) = [G^{[1]}(0) - G^{[1]}(\lambda_o)]^{-1} g(q), \quad \lambda_o \leq q \leq 0.
\]

19
Proof of Theorem 3.1 It follows from (3.1) that the log-likelihood function is

\[
\log L(\lambda_i, \gamma_i, i = 1, \ldots, k-1 | z) = \log f(z | \Sigma^{-1})
\]

\[
= \log C(\lambda; \ldots; \lambda) + \log [g(\sum_{j=1}^{n} z_j^* \Sigma^{-1} z_j)]
\]

\[
= \log C(\lambda; \ldots; \lambda) + \log [g(tr(\Sigma^{-1} z_j z_j^*))]
\]

\[
= \log C(\lambda; \ldots; \lambda) + \log [g(tr(\Sigma^{-1} \sum_{j=1}^{n} z_j z_j^*))]
\]

\[
= \log C(\lambda; \ldots; \lambda) + \log [g(tr(\Sigma^{-1} S))]
\]

and finally

\[
\log L(\lambda_i, \gamma_i, i = 1, \ldots, k-1 | z) = \log C(\lambda; \ldots; \lambda) + \log [g(\sum_{j=1}^{k-1} \lambda_j \gamma_j^* S \gamma_j)]. 
\] (3.2)

Following Dryden and Mardia(1998), we hold \( \lambda \) constant, and if \( g \) is strictly increasing then the mode of \( \gamma_i \) is

\[
\arg \max f(z | \Sigma) \|_{\gamma_i} = \arg \max \left( \sum_{j=1}^{k-1} \lambda_j \gamma_j^* S \gamma_j \right)_{\|_{\gamma_i}} = \arg \max \sum_{j=1}^{k-1} \lambda_j \gamma_j^* S \gamma_j = v_i.
\]

Similarly, when \( g \) is strictly decreasing then the mode of \( \gamma_i \) is

\[
\arg \max f(z | \Sigma) \|_{\gamma_i} = \arg \max \left( \sum_{j=1}^{k-1} \lambda_j \gamma_j^* S \gamma_j \right)_{\|_{\gamma_i}} = \arg \min \sum_{j=1}^{k-1} \lambda_j \gamma_j^* S \gamma_j = v_i.
\]

A proof of the statements above can be easily obtained by considering standard results on Hermitian matrices, see for example Horn and Johnson(1985), section 4.2.

Now from (3.2) it follows that

\[
\log L(\lambda_i, \tilde{\gamma}_i, i = 1, \ldots, k-1 | z) = \log C(\lambda; \ldots; \lambda) + \log [g(\sum_{j=1}^{k-1} \tilde{\gamma}_j^* S \tilde{\gamma}_j)]
\]

\[
= \log C(\lambda; \ldots; \lambda) + \log [g(\sum_{j=1}^{k-1} \lambda_j \zeta_j)]
\]

since \( \tilde{\gamma}_j^* S \tilde{\gamma}_j = \nu_j^* S \nu_j = \zeta_j \), for all \( j = 1, 2, \ldots, k-1 \).
Hence the mode of $\lambda_i$ is obtained by finding the solution of the equation

$$
\frac{\partial \log C(\lambda; \ldots; \lambda)}{\partial \lambda_i} \bigg|_{\lambda=\hat{\lambda}} + \frac{\partial \log g(\sum_{j=1}^{k-1} \lambda_j \zeta_j)}{\partial \lambda_i} \bigg|_{\lambda=\hat{\lambda}} = 0.
$$

**Proof of Theorem 3.2** From Lemma 2.1 we have that the distribution of the quadratic form is given by

$$
f(q|\lambda) \propto g(q),
$$

where $\lambda_1 < \lambda_2 < \cdots < \lambda_{p-1} < \lambda_p$, are the eigenvalues of $\Sigma^{-1}$. Let $q_0$ be a realization from $f(q|\lambda)$. Since $q_0 = x^*\Sigma^{-1}x$, the points of the ellipsoid $\{\tilde{x} : q_0 = x^*\Sigma^{-1}x\}$, will be realizations from $x \sim SEL_p(0, \Sigma; g)$. From the spectral decomposition theorem, we can deduce that all axes of the ellipsoid are of the form $\pm \sqrt{\nu} \lambda^{-\frac{1}{2}} v$, where $\lambda$ and $v$ are some eigenvalue and corresponding eigenvector of the matrix $\Sigma^{-1}$, and the sign denotes different directions. To obtain any point in the ellipsoid we take a convex combination of these axes. Thus $x$ can be written in the form $x = \sum_{i=1}^{p} r_i \sqrt{\nu} \lambda^{-\frac{1}{2}} v$, where $\sum_{i=1}^{p} r_i = 1$, $0 \leq r_i \leq 1$, $i = 1, \ldots, p$, some constants.

**References**


21


