A multidimensional scaling approach to shape analysis

BY IAN L. DRYDEN
School of Mathematical Sciences, University of Nottingham, University Park, Nottingham, NG7 2RD, U.K.
Ian.Dryden@nottingham.ac.uk

ALFRED KUME
Institute of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, CT2 7NF, U.K.
a.kume@kent.ac.uk

HUILING LE
School of Mathematical Sciences, University of Nottingham, University Park, Nottingham, NG7 2RD, U.K.
Huiling.Le@nottingham.ac.uk

AND ANDREW T.A. WOOD
School of Mathematical Sciences, University of Nottingham, University Park, Nottingham, NG7 2RD, U.K.
Andrew.Wood@nottingham.ac.uk

SUMMARY

We propose an alternative to Kendall's shape space for reflection shapes of configurations in $\mathbb{R}^m$ with k labelled vertices, where reflection shape consists of all the geometric information that is invariant under compositions of similarity and reflection transformations. The proposed approach embeds the space of such shapes into the space $\mathcal{P}(k-1)$ of $(k-1) \times (k-1)$ real symmetric positive semi-definite matrices, which is the closure of an open subset of a Euclidean space, and defines mean shape as the natural projection of Euclidean means in $\mathcal{P}(k-1)$ on to the embedded copy of the shape space. This approach has strong connections with multidimensional scaling, and the mean shape so defined gives good approximations to other commonly used definitions of mean shape. We also use standard perturbation arguments for eigenvalues and eigenvectors to obtain a central limit theorem which then enables the application of standard statistical techniques to shape analysis in $m \geqslant 2$ dimensions.

Some key words: Central limit theorem; Procrustes mean shape; Reflection shape; Tangent space projection.

1. Introduction

In many applications one is interested in the shape of an object, where location, rotation and scale can be ignored. Statistical analysis of shape is commonly based on either Kendall's (1984) or Bookstein's (1986) shape spaces. Both of these two versions of shape spaces are curved rather than flat, and so standard statistical results on Euclidean spaces cannot usually be applied directly to shape analysis. In the past two decades or so, much progress has been made both in theory and applications. For example, the classical method of taking arithmetic averages is inappropriate for the estimation of the mean shape and, among other possibilities, the 'partial Procrustes estimator' and the 'full Procrustes estimator' (Goodall, 1991; Kent, 1992) have been proposed and widely used in practice. Also, if the shapes of the data configurations are highly concentrated, we can project the shapes of the data configurations on to, in effect, the tangent space to the shape space; to be more specific, we could use a Procrustes projection to project the data on to the tangent space at the Procrustes mean shape. Then, since the tangent space at a point of a Riemannian manifold is a Euclidean space, we may apply techniques that are suitable for Euclidean data to the projected data.

Since most shape spaces are unfamiliar spaces, as explained in Kendall et al. (1999), it is not always easy to work with these concepts in practice, especially when $m \geqslant 3$. Except for the full Procrustes estimator for planar shapes, there is no closed form for Procrustes means: their computation is based on computer algorithms which can take a long time to run for large samples of data. The tangent projection technique is also restricted to concentrated data in order to obtain a reasonable conclusion. In this paper, we propose an alternative to the existing approaches to statistical reflection shape analysis, where the reflection shape of an object consists of all the geometric information that is invariant under compositions of similarity and reflection transformations. Our approach is based on an embedding of the reflection shape space in a suitable space of matrices, which has the advantage of making many computations more familiar and easily handled. It also gives an easily computable mean shape and the comparison of mean shapes so defined with other mean shapes shows that, in most cases, the former are good approximations to the latter. The central limit theorem that we shall establish allows us to apply many standard statistical results to statistical analysis of shape. Moreover, the fact that this representation applies only to reflection shapes, rather than shapes, is a relatively mild restriction as it is essentially equivalent to working on 'one half' of shape space and, for the majority of applications, this is automatically the case: there are not many applications where one needs to consider both a shape and its reflection. Nevertheless, reflection information can be recovered in the analysis with the use of a suitable parity function on the reflection shape space, taking values $+1$ and -1, say. Separate analyses would then be performed on the subsample of shapes with parity $+1$ and the complementary subsample with parity -1.

The multidimensional scaling approach described in this paper has a number of connections with other work: Kent (1994, §7) discusses this approach when $m = 2$ and notes that it extends readily to higher dimensions; Chikuse & Jupp (2004), who discuss tests of uniformity in shape spaces, consider essentially the same projection on to the embedded shape space as we do, and also discuss the Bingham distribution in this context; and the Euclidean distance matrix analysis of Lele (1993) and Lele & Richtsmeier (2001) is also closely related, in that both approaches use the relevant parts of the space of symmetric positive semidefinite matrices to represent their objects. However, the work of Lele (1993) and Lele & Richtsmeier (2001) focuses on form, i.e. size-and-shape, whereas we focus exclusively on shape. Our definition of mean shape ensures that it lies in the space of reflection shapes and what it gives is in general not a simple projection of the mean form of Lele (1993) and Lele & Richtsmeier (2001). The central limit theorem that we shall establish is related to our mean shape. It holds for any distribution on the reflection space in common use, and in particular those induced from landmarks.

Other related work includes Bhattacharya & Patrangenaru (2003, 2005), Bandulasiri & Patrangenaru (2005) and an unpublished 2006 Texas Tech University Ph.D. thesis by A. Bandulasiri. This body of work has close connections with the developments here, but there are also important differences. The most substantial of these differences is that, while they work in a general differential-geometric framework, we work in spaces of matrices and exploit useful structure which allows us to represent relevant tangent spaces as linear subspaces of the original matrix spaces.

In summary, the main purpose of this paper is to develop a computationally convenient framework for inference for shapes in $m \geqslant 3$ dimensions. This approach is particularly useful when the sample size, n, or the number of landmarks, k, is large.

2. A representation of reflection shape space

The reflection shape of a configuration in $\mathbb{R}^m$ with k labelled vertices, where without loss of generality we shall assume that $k > m$, is its equivalence class under compositions of translations, scalings, rotations and reflections. Therefore, the space of reflection shapes of configurations in $\mathbb{R}^m$ with k labelled vertices is the quotient space by a reflection of the corresponding shape space.

Following standard practice in shape analysis, we represent each configuration in $\mathbb{R}^m$ with k labelled vertices by an $m \times k$ matrix where its ith column comprises the coordinates of the ith vertex of the configuration. Then, using the standard Helmert submatrix widely used in shape analysis to remove the effect of translation, see e.g. Dryden & Mardia (1998), the class of configurations which differ from each other only by translations and scaling can be represented by an $m \times (k-1)$ real matrix X with $\mathrm{tr}(X^\top X) = 1$. The space $\mathcal{S}_m^k$ of such matrices is called the pre-shape sphere and is identical with the unit sphere in $\mathbb{R}^{m(k-1)}$; see Kendall et al. (1999, p. 3). In particular, the tangent space $\mathcal{T}_X(\mathcal{S}_m^k)$ to $\mathcal{S}_m^k$ at X can be identified with $\{Y \in \mathcal{X}_m^{k-1} \mid \mathrm{tr}(X^\top Y) = 0\}$, where $\mathcal{X}_m^{k-1}$ is the space of $m \times (k-1)$ real matrices.

Let $\mathcal{P}(k)$ denote the space of $k \times k$ positive semidefinite real symmetric matrices and let

$$\mathcal{P}_m(k) = \{P \in \mathcal{P}(k) \,|\, 1 \leqslant \text{rank}(P) \leqslant m,\ \text{tr}(P) = 1\}.$$

In both of these spaces, we define distance in terms of the Euclidean norm $||A|| = \{\text{tr}(A^\top A)\}^{1/2}$ in standard fashion.

Consider the map

$$\pi:\ \mathcal{S}_m^k \longrightarrow \mathcal{P}(k-1);\ \ X \mapsto X^\top X. \tag{1}$$

The image of π is $\mathcal{P}_m(k-1)$ and $\pi(X_1) = \pi(X_2)$ if and only if $X_1 = TX_2$ for some $T \in \mathcal{O}(m)$, where $\mathcal{O}(m)$ is the space of $m \times m$ orthogonal matrices. It then follows from an argument similar to that in Carne (1990) that $\mathcal{P}_m(k-1)$ is homeomorphic to the reflection shape space of configurations in $\mathbb{R}^m$ with k labelled vertices. We shall accordingly use $\mathcal{P}_m(k-1)$ to represent that shape space. This representation is an embedding of the reflection shape space into a Euclidean space. Note that a similar representation for shape space has been used in Kendall (1990) for the investigation of the behaviour of shape diffusion, and in Chikuse & Jupp (2004) for the testing of uniformity of reflection shapes. Another similar representation for shape space has been widely used in statistics for planar shapes; see for example Kent (1992) and Bhattacharya & Patrangenaru (2003, 2005), where the points of $\mathcal{S}_2^k$ are identified with $(k-1)$-dimensional complex unit vectors and the corresponding shape space is represented by the space of $(k-1) \times (k-1)$ complex Hermitian projection matrices of rank 1.

For the purpose of the following statistical analysis, we need to identify the tangent space to $\mathcal{P}_m(k-1)$. To this end, we note that, for any $X \in \mathcal{S}_m^k$ and $Y \in \mathcal{T}_X\,(\mathcal{S}_m^k)$, if $X(t)$ is a curve in $\mathcal{S}_m^k$ with $X(0) = X$ and initial tangent vector Y, that is, $\dot{X}(0) = Y$, then

$$\begin{aligned} \left.\frac{d\pi\{X(t)\}}{dt}\right|_{t=0} &= \left.\frac{d\{X(t)^\top X(t)\}}{dt}\right|_{t=0} \\ &= \dot{X}(0)^\top X(0) + X(0)^\top \dot{X}(0) = Y^\top X + X^\top Y. \end{aligned}$$

Hence, the differential $d\pi(X)$ of π at X is

$$d\pi(X):\ \mathcal{T}_X\,(\mathcal{S}_m^k) \to \mathcal{T}_{X^\top X}\{\mathcal{P}_m(k-1)\};\ \ \ Y \mapsto Y^\top X + X^\top Y. \tag{2}$$

It is useful to decompose $\mathcal{T}_X\,(\mathcal{S}_m^k)$ into the orthogonal sum of the vertical subspace, whose vectors correspond to directions in which shape, as opposed to pre-shape, does not change, and its orthogonal complement, the horizontal subspace; see Kendall et al. (1999, p. 109). Since the vertical subspace, $\mathcal{V}_X$, of $\mathcal{T}_X\,(\mathcal{S}_m^k)$ at X is the kernel of the differential $d\pi(X)$ of π at X, it can be written as

$$\begin{aligned} \mathcal{V}_X &= \{Y \in \mathcal{X}_m^{k-1} \,|\, \text{tr}(X^\top Y) = 0,\ Y^\top X + X^\top Y = 0_{k-1,k-1}\} \\ &= \{Y \in \mathcal{X}_m^{k-1} \,|\, \text{tr}(X^\top Y) = 0,\ \ \text{tr}(Y^\top XS) = 0 \text{ for all symmetric } S\}, \end{aligned}$$

and its orthogonal complement, the horizontal subspace $\mathcal{H}_X$, is given by

$$\mathcal{H}_X = \{XS \,|\, \mathrm{tr}(X^\top XS) = 0,\ S = S^\top\}.$$

The restriction of $d\pi(X)$ to $\mathcal{H}_X$ is a bijection from $\mathcal{H}_X$ to the tangent space $\mathcal{T}_{X^\top X}\{\mathcal{P}_m(k-1)\}$ to $\mathcal{P}_m(k-1)$ at $X^\top X$, and hence (2) shows that $\mathcal{T}_{X^\top X}\{\mathcal{P}_m(k-1)\}$ can be identified as

$$\mathcal{T}_{X^\top X}\{\mathcal{P}_m(k-1)\} = \{X^\top XS + SX^\top X \,|\, \mathrm{tr}(X^\top XS) = 0,\ S = S^\top\}. \tag{3}$$

To simplify the identification of (3), we use the spectral decomposition of the symmetric matrix $X^\top X$. That provides an orthonormal basis $\{u_i \,|\, 1 \leqslant i \leqslant k-1\}$ of $\mathbb{R}^{k-1}$ comprising the eigenvectors of $X^\top X$ and the corresponding nonnegative eigenvalues $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_m \geqslant 0 = \ldots = 0$, with $\sum_{i=1}^m \lambda_i = 1$, such that $X^\top X = \sum_{i=1}^m \lambda_i u_i u_i^\top = U\Lambda U^\top$, where U denotes the orthogonal matrix whose ith column is u_i and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_m, 0, \ldots, 0)$. Since

$$\{u_i u_i^\top \,|\, 1 \leqslant i \leqslant k-1\} \bigcup \{u_i u_j^\top + u_j u_i^\top \,|\, 1 \leqslant i < j \leqslant k-1\}$$

expresses a basis for the space of $(k-1) \times (k-1)$ real symmetric matrices in terms of the orthonormal basis $\{u_i \,|\, 1 \leqslant i \leqslant k-1\}$, it follows from (3) that, if $\mathrm{rank}(X) = m$, then

$$\begin{aligned}&\mathcal{T}_{X^\top X}\{\mathcal{P}_m(k-1)\}\\ &\quad = \left\{ S \,\middle|\, S = U \begin{pmatrix} A_m & B \\ B^\top & 0_{k-1-m,k-1-m} \end{pmatrix} U^\top,\ A_m = A_m^\top,\ \mathrm{tr}(A_m) = 0 \right\}.\end{aligned} \tag{4}$$

Clearly, $\mathcal{T}_{X^\top X}\{\mathcal{P}_m(k-1)\}$ is a Euclidean space of dimension $\frac{1}{2}m(2k-m-1)-1$. Moreover, (4) implies that, for Λ defined as above,

$$\mathcal{T}_\Lambda\,\{\mathcal{P}_m(k-1)\} = \left\{ S \,\middle|\, S = \begin{pmatrix} A_m & B \\ B^\top & 0_{k-1-m,k-1-m} \end{pmatrix},\ A_m = A_m^\top,\ \mathrm{tr}(A_m) = 0 \right\} \tag{5}$$

and that

$$\mathcal{T}_{U\Lambda U^\top}\{\mathcal{P}_m(k-1)\} = U\ \mathcal{T}_\Lambda\,\{\mathcal{P}_m(k-1)\}\, U^\top, \tag{6}$$

with conjugation by U being an isometry, with respect to the induced Euclidean metrics, between the two tangent spaces. When $\mathrm{rank}(X) = r < m$, the above statements still hold provided m is replaced by r.

For statistical analysis, it is sometimes more convenient to express matrices in the tangent space $\mathcal{T}_{X^\top X}\{\mathcal{P}_m(k-1)\}$ as column vectors of dimension $\frac{1}{2}m(2k-m-1)-1$. We do this by first considering $\mathcal{T}_\Lambda\,\{\mathcal{P}_m(k-1)\}$. Using (5), we map $S \in \mathcal{T}_\Lambda\,\{\mathcal{P}_m(k-1)\}$, in a nonstandard manner, to the column vector $\tilde{q}_S$, of dimension $\frac{1}{2}m(2k-m-1)$, given by the elements on and above the diagonal of S, excluding those that are always zero:

$$\tilde{q}_S = (s_{11}, \ldots, s_{mm}, 2^{\frac{1}{2}}s_{12}, \ldots, 2^{\frac{1}{2}}s_{1,k-1}, 2^{\frac{1}{2}}s_{23}, \ldots, 2^{\frac{1}{2}}s_{2,k-1}, \ldots, 2^{\frac{1}{2}}s_{m,m+1}, \ldots, 2^{\frac{1}{2}}s_{m,k-1})^\top.$$

The factor $2^{\frac{1}{2}}$ appears before the components corresponding to the nondiagonal elements of S since these elements contribute twice to the square of the norm $\|S\|^2 = \mathrm{tr}(S^2)$. Then, to account for the constraint $\mathrm{tr}(S) = 0$, we define

$$q_S = \tilde{H}\tilde{q}_S\,, \tag{7}$$

where

$$\tilde{H} = \begin{pmatrix} H_{m-1,m} & 0 \\ 0 & I_{\frac{1}{2}m(2k-m-3)} \end{pmatrix},$$

and $H_{m-1,m}$ is the standard Helmert submatrix defined in Dryden & Mardia (1998, p. 34). It is easy to see that the map $S \mapsto q_S$ is an isometry. Thus, a representation of $\mathcal{T}_\Lambda\,\{\mathcal{P}_m(k-1)\}$ in terms of column vectors of dimension $\frac{1}{2}m(2k-m-1)-1$ is given by

$$\left\{q_S\ :\ S \in \mathcal{T}_\Lambda\,\{\mathcal{P}_m(k-1)\}\right\}. \tag{8}$$

To obtain the corresponding representation for the tangent space $\mathcal{T}_{X^\top X}\{\mathcal{P}_m(k-1)\}$, where $X^\top X = U\Lambda U^\top$, we take the following orthonormal basis of $\mathcal{T}_\Lambda\,\{\mathcal{P}_m(k-1)\}$:

$$W_i = \begin{cases} \{i(i+1)\}^{-\frac{1}{2}}\left(iE_{i+1,i+1} - \sum\limits_{j=1}^{i} E_{jj}\right) & 1 \leqslant i \leqslant m-1 \\[2ex] 2^{-\frac{1}{2}}\left(E_{j,i-\ell_j+j} + E_{i-\ell_j+j,j}\right) & \ell_j < i \leqslant \ell_{j+1},\ 1 \leqslant j \leqslant m, \end{cases} \tag{9}$$

where $1 \leqslant i \leqslant \frac{1}{2}m(2k-m-1)-1$; $\ell_j = m + (j-1)k - j(j+1)/2$; and E_{st} is the $(k-1)\times(k-1)$ matrix whose (s,t)th entry is 1 and whose other entries are all 0. Here, we have taken the standard orthonormal basis of the space of symmetric matrices and adjusted the diagonal members, in the manner of Helmert, to ensure that they have trace zero. Then the ith component of q_S is equal to $\mathrm{tr}(SW_i)$, the component of S with respect to W_i, and so the representation (8) is the column vector of coordinates of the matrix S in $\mathcal{T}_\Lambda\,\{\mathcal{P}_m(k-1)\}$ with respect to the basis (9). On the other hand, if we use the isometry (6) to determine the orthonormal basis

$$\{UW_iU^\top \mid 1 \leqslant i \leqslant \tfrac{1}{2}m(2k-m-1)-1\}$$

of $\mathcal{T}_{X^\top X}\{\mathcal{P}_m(k-1)\}$, then the coordinate vector of $S \in \mathcal{T}_{X^\top X}\{\mathcal{P}_m(k-1)\}$ with respect to this basis is the same as that, $q_{U^\top SU}$, of its image $U^\top SU$ in $\mathcal{T}_\Lambda\,\{\mathcal{P}_m(k-1)\}$ with respect to the basis (9), and this will be our chosen column vector representation for matrices in $\mathcal{T}_{X^\top X}\{\mathcal{P}_m(k-1)\}$.

Finally, we consider, for any $P \in \mathcal{P}_m(k-1)$, the orthogonal projection $\psi_{X^\top X}(P)$ of $P - X^\top X$ on to the tangent space $\mathcal{T}_{X^\top X}\{\mathcal{P}_m(k-1)\}$. Write

$$\tilde{\psi} : P = \begin{pmatrix} A_m & B \\ B^\top & C \end{pmatrix} \mapsto \begin{pmatrix} A_m & B \\ B^\top & 0 \end{pmatrix}.$$

Then, it can be checked that, when $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_m, 0, \ldots, 0)$ as before,

$$\begin{aligned} \psi_\Lambda(P) &= \tilde{\psi}(P) - \Lambda - \frac{1}{m}\mathrm{tr}\{\tilde{\psi}(P) - \Lambda\} \begin{pmatrix} I_m & 0 \\ 0 & 0_{k-1-m,k-1-m} \end{pmatrix} \\ \psi_{X^\top X}(P) &= U\,\psi_\Lambda(U^\top PU)\,U^\top, \end{aligned} \tag{10}$$

with the second formula following from the first on account of the isometry (6). Note that $\psi_{X^\top X}(P)$ so defined is symmetric, has zero trace and lies in (3). Moreover, in the notation of (7),

$$\tilde{q}_{U^\top \psi_{X^\top X}(P) U} = \tilde{q}_{\psi_\Lambda(U^\top PU)}, \qquad q_{U^\top \psi_{X^\top X}(P) U} = q_{\psi_\Lambda(U^\top PU)}. \tag{11}$$

3. Multidimensional scaling mean reflection shape

Using the representation $\mathcal{P}_m(k-1)$ of the reflection shape space, we may define the mean reflection shape of a random configuration as follows.

If $P \in \mathcal{P}(k-1)$ has spectral decomposition $P = V\Omega V^\top$, where $\Omega = \mathrm{diag}(\omega_1, \ldots, \omega_{k-1})$ consists of the ordered eigenvalues $\omega_1 \geqslant \ldots \geqslant \omega_{k-1} \geqslant 0$ of P and $V = (v_1, \ldots, v_{k-1})$ is the matrix whose columns are the corresponding orthonormal eigenvectors, then $P = \sum_{i=1}^{k-1} \omega_i v_i v_i^\top$ and, if we write ϕ for the natural projection of $\mathcal{P}(k-1) \setminus \{0\}$ on to $\mathcal{P}_m(k-1)$, then

$$\phi(P) = \frac{1}{\omega_1 + \ldots + \omega_m} \sum_{i=1}^{m} \omega_i v_i v_i^\top.$$

Note that this projection of P on to $\mathcal{P}_m(k-1)$ is unique if and only if $\omega_m > \omega_{m+1}$.

Let X be a random matrix in $\mathcal{S}_m^k$ and write $\Xi = E(X^\top X)$ for the symmetric nonnegative definite $(k-1) \times (k-1)$ matrix which is the expectation of $X^\top X$.

Definition 1. *If X is a random matrix in $\mathcal{S}_m^k$, then $\phi(\Xi)$ is called the mean ϕ-shape of X.*

Remark 1. A condition for the mean ϕ-shape to be unique is now given. Write $U\Delta U^\top$ for the spectral decomposition of Ξ, where $\Delta = \mathrm{diag}(\delta_1, \ldots, \delta_{k-1})$ and $\delta_1 \geq \ldots \geq \delta_{k-1} \geq 0$. Since

$$\mathrm{tr}\{E(X^\top X)\} = E\{\mathrm{tr}(X^\top X)\} = 1,$$

it follows that $\Xi \in \mathcal{P}_m(k-1)$, and therefore a mean ϕ-shape of X can always be defined, and is given by

$$\phi(\Xi) = \sum_{i=1}^{m} \lambda_i u_i u_i^\top,$$

where $\lambda_i = \delta_i/(\delta_1 + \ldots + \delta_m)$. The mean ϕ-shape is unique if and only if $\delta_m \neq \delta_{m+1}$. When $\delta_m = \delta_{m+1}$, $\phi(\Xi)$ consists of a set of matrices in $\mathcal{P}_m(k-1)$, rather than a single matrix; see Zeizold (1977).

The mean ϕ-shape so defined is the 'extrinsic' mean reflection shape with respect to the embedded copy $\mathcal{P}_m(k-1)\backslash\mathcal{P}_{m-1}(k-1)$ of the nondegenerate component of the reflection shape space in $\mathcal{P}(k-1)$, in the sense of Bhattacharya & Patrangenaru (2003, 2005) and Hendriks & Landsman (1998). If $m = 2$, we can represent the pre-shapes by $(k-1)$-dimensional complex unit column vectors z. Then, if we embed the shape space into the space of $(k-1) \times (k-1)$ Hermitian matrices as the space of $(k-1) \times (k-1)$ Hermitian projection matrices of rank 1 as discussed in the previous section, the widely used full Procrustes mean shapes are just the projections of $E(zz^*)$ on to such a space, where z^* denotes the transpose of the complex conjugate of z. Hence, the definition of mean ϕ-shape has parallels to that of the full Procrustes mean of planar shapes. Also, there is some similarity between the mean ϕ-shape and that of Lele's Euclidean distance matrix analysis mean form; see Lele (1993). The latter involves a correction for bias under Gaussian models, but is appropriate for mean form rather than mean shape.

In addition to the obvious advantage of being easy to compute, the mean ϕ-shape so defined has the following basic properties.

(i) For any fixed $(k-1)\times(k-1)$ matrix A of full rank, the mean ϕ-shape of XA is $\phi(A^\top \Xi A)$. In particular, if $A \in \mathcal{O}(k-1)$, then the mean ϕ-shape of XA is $A^\top\phi(\Xi)A$ and so the mean ϕ-shape of XU is diag$(\lambda_1, \ldots, \lambda_m, 0, \ldots, 0)$.

(ii) If X is a uniform random matrix on $\mathcal{S}_m^{k-1}$ then, by symmetry, Ξ is a diagonal matrix and its diagonal entries are all equal and so $\Xi = (k-1)^{-1}I_{k-1}$. In this case, the mean ϕ-shape is not unique if $k-1 > m$, and the set of mean ϕ-shapes is identical to the Grassmannian of m-planes in $(k-1)$-space. Note that, if the vertices of a configuration are independent and identically distributed with a $N(\mu, \sigma^2 I_m)$ distribution, then the corresponding X is a uniform random matrix.

(iii) Suppose that $X_1, \ldots, X_n$ are random configurations such that $X_1^\top X_1, \ldots, X_n^\top X_n$ are independent and identically distributed with the same distribution as $X^\top X$. Write $\hat{\Xi} = n^{-1}\sum_{i=1}^{n} X_i^\top X_i$, with corresponding spectral decomposition $\hat{\Xi} = \hat{U}\hat{\Delta}\hat{U}^\top$. Assume for simplicity that $\delta_m \neq \delta_{m+1}$, so that the mean ϕ-shape of X is uniquely defined. Then $\hat{\Xi}$ is a strongly consistent estimator of Ξ and so the sample mean ϕ-shape, $\phi(\hat{\Xi})$, is a consistent estimator of the mean ϕ-shape $\phi(\Xi)$, by continuity of the map ϕ on $\mathcal{P}_m(k-1) \setminus \mathcal{P}_{m-1}(k-1)$.

Kent (1994) introduced the sample multidimensional scaling mean for two-dimensional data, and some discussion for higher dimensions was given in Dryden & Mardia (1998, pp. 281-2). These authors described this type of mean for both mean form and mean reflection shape, where the mean form incorporates scale information while the mean reflection shape does not. In the current paper, we focus on mean reflection shape only.

4. A central limit theorem and standard errors

Assume that X is a random matrix in $\mathcal{S}_m^k$, where $k > m$, and that X has rank m with probability one. Let $X_1, \ldots, X_n$ be such that $X_1^\top X_1, \ldots, X_n^\top X_n$ are independent and identically distributed with the same distribution as $X^\top X$. As in the previous section, write $\Xi = E(X^\top X)$ and $\hat{\Xi} = n^{-1} \sum_{i=1}^{n} X_i^\top X_i$, with corresponding spectral decompositions $\Xi = U \Delta U^\top$ and $\hat{\Xi} = \hat{U} \hat{\Delta} \hat{U}^\top$.

By the multivariate central limit theorem, the distribution of $G_n = n^{1/2}(\hat{\Xi} - \Xi)$ converges to a Gaussian distribution with zero mean. Let $G = (g_{ij})$ denote a symmetric random matrix with this limiting Gaussian distribution. The covariance matrix of G can be determined by

$$\operatorname{cov}(g_{ij}, g_{st}) = \operatorname{cov}(x_i^\top x_j, x_s^\top x_t) = E(x_i^\top x_j x_s^\top x_t) - E(x_i^\top x_j)\, E(x_s^\top x_t), \tag{12}$$

where x_i denotes the ith column of X.

Write $G^u = U^\top G U = (g_{ij}^u)$. Then G^u is also a symmetric Gaussian random matrix.

Theorem 1. *If $\delta_m \neq \delta_{m+1}$, then*

$$n^{1/2}\{\phi(\hat{\Xi}) - \phi(\Xi)\} \to Z,$$

in distribution, where Z is a symmetric Gaussian matrix with zero mean given by

$$Z = \frac{1}{\delta_1 + \ldots + \delta_m} \left\{ U \tilde{G}^u U^\top - \left(\sum_{j=1}^{m} g_{jj}^u \right) \phi(\Xi) \right\}, \tag{13}$$

and where the entries of the symmetric matrix $\tilde{G}^u = (\tilde{g}_{ij}^u)$ are determined by

$$\tilde{g}_{ij}^u = \begin{cases} g_{ij}^u & 1 \leqslant i \leqslant j \leqslant m \\ \dfrac{\delta_i}{\delta_i - \delta_j} g_{ij}^u & 1 \leqslant i \leqslant m < j \leqslant k-1 \\ 0 & m < i \leqslant j \leqslant k-1. \end{cases} \tag{14}$$

Proof. For simplicity, we assume that Ξ has distinct eigenvalues. However, the following argument, as well as the result, generalizes to the case with only the stated assumption $\delta_m \neq \delta_{m+1}$.

For $j = 1, \ldots, m$, let $\lambda_i = \delta_i / (\delta_1 + \ldots + \delta_m)$ as before and define $\hat{\lambda}_i$ in terms of the $\hat{\delta}_i$ in parallel fashion. From the identity

$$\hat{\lambda}_i \hat{u}_i \hat{u}_i^\top - \lambda_i u_i u_i^\top = \lambda_i (\hat{u}_i \hat{u}_i^\top - u_i u_i^\top) + (\hat{\lambda}_i - \lambda_i) u_i u_i^\top + (\hat{\lambda}_i - \lambda_i)(\hat{u}_i \hat{u}_i^\top - u_i u_i^\top) \tag{15}$$

it follows that

$$n^{1/2}\{\phi(\hat{\Xi}) - \phi(\Xi)\} = n^{1/2} \sum_{i=1}^{m} (\hat{\lambda}_i - \lambda_i) u_i u_i^\top + n^{1/2} \sum_{i=1}^{m} \lambda_i (\hat{u}_i \hat{u}_i^\top - u_i u_i^\top) + o_p(1).$$

Since

$$n^{1/2}(\hat{\lambda}_i - \lambda_i) = n^{1/2}\frac{\hat{\delta}_i - \delta_i}{\delta_1 + \ldots + \delta_m} - n^{1/2}\frac{\lambda_i}{\delta_1 + \ldots + \delta_m}\sum_{j=1}^{m}(\hat{\delta}_j - \delta_j) + o_p(1)$$

and since, using Watson (1983, Appendix B), in distribution,

$$n^{1/2}\left(\hat{u}_i\hat{u}_i^\top - u_i u_i^\top\right) \to \sum_{j\neq i}\frac{g_{ij}^u(u_j u_i^\top + u_i u_j^\top)}{\delta_i - \delta_j}, \tag{16}$$

$$n^{1/2}\left(\hat{\delta}_i - \delta_i\right) \to g_{ii}^u,$$

we have that

$$\begin{aligned} n^{1/2}\{\phi(\hat{\Xi}) - \phi(\Xi)\} \quad = \quad & \frac{n^{1/2}}{\delta_1 + \ldots + \delta_m}\sum_{i=1}^{m}(\hat{\delta}_i - \delta_i)u_i u_i^\top \\ & - \frac{n^{1/2}}{\delta_1 + \ldots + \delta_m}\sum_{j=1}^{m}(\hat{\delta}_j - \delta_j)\sum_{i=1}^{m}\lambda_i u_i u_i^\top \\ & + n^{1/2}\sum_{i=1}^{m}\lambda_i(\hat{u}_i\hat{u}_i^\top - u_i u_i^\top) + o_p(1) \end{aligned}$$

converges in distribution to

$$\begin{aligned} & \frac{1}{\delta_1 + \ldots + \delta_m}\left\{\sum_{i=1}^{m} g_{ii}^u u_i u_i^\top + \sum_{i=1}^{m}\delta_i \sum_{\substack{1\leqslant j\leqslant k-1 \\ j\neq i}} g_{ij}^u \frac{u_j u_i^\top + u_i u_j^\top}{\delta_i - \delta_j}\right\} \\ & \qquad - \frac{1}{\delta_1 + \ldots + \delta_m}\left(\sum_{j=1}^{m} g_{jj}^u\right)\phi(\Xi) \\ & \quad = \frac{1}{\delta_1 + \ldots + \delta_m}\left\{\sum_{i,j=1}^{m} g_{ij}^u u_i u_j^\top + \sum_{i=1}^{m}\delta_i \sum_{j=m+1}^{k-1} g_{ij}^u \frac{u_j u_i^\top + u_i u_j^\top}{\delta_i - \delta_j}\right\} \\ & \qquad - \frac{1}{\delta_1 + \ldots + \delta_m}\left(\sum_{j=1}^{m} g_{jj}^u\right)\phi(\Xi) \\ & \quad = \frac{1}{\delta_1 + \ldots + \delta_m}\left\{U\tilde{G}^u U^\top - \left(\sum_{j=1}^{m} g_{jj}^u\right)\phi(\Xi)\right\}, \end{aligned}$$

as required. □

The expression (14) shows that $\tilde{G}^u$ is a symmetric matrix with the bottom right-hand $(k-1-m)\times(k-1-m)$ submatrix equal to zero and (13) shows that $\mathrm{tr}(Z) = 0$. It then follows from (4) that the limit Gaussian matrix Z is actually a random matrix on the tangent space $\mathcal{T}_{\phi(\Xi)}\{\mathcal{P}_m(k-1)\}$.

Note also that the convergence stated in the theorem is not the convergence of the tangent space projection of the difference between the sample mean and true mean and so it is not a special case of the central limit theorem presented in Bhattacharya & Patrangenaru (2003). However, for the tangent projection $\psi_{\phi(\Xi)}\{\phi(\hat{\Xi})\}$, defined via (10), of $\phi(\hat{\Xi}) - \phi(\Xi)$ on to the tangent space to $\mathcal{P}_m(k-1)$ at the mean ϕ-shape $\phi(\Xi)$, we have the following result.

COROLLARY 1. *If* $\delta_m \neq \delta_{m+1}$, *then*

$$n^{1/2}\psi_{\phi(\Xi)}\{\phi(\hat{\Xi})\} \to Z,$$

in distribution, where Z *is given as in Theorem* 1.

Proof. It follows from Theorem 1 that

$$n^{1/2}\{\phi(U^\top\hat{\Xi}U) - \Lambda\} \to U^\top ZU,$$

in distribution. On the other hand, $\tilde{\psi}$ is a linear operator leaving any symmetric matrix with zero bottom right-hand $(k-1-m)\times(k-1-m)$ submatrix fixed, from which it follows that $\tilde{\psi}(\Lambda) = \Lambda$ and $\tilde{\psi}(U^\top ZU) = U^\top ZU$. Hence, we have, in distribution, that

$$n^{1/2}[\tilde{\psi}\{\phi(U^\top\hat{\Xi}U)\} - \Lambda] = n^{1/2}\tilde{\psi}\{\phi(U^\top\hat{\Xi}U) - \Lambda\} \to \tilde{\psi}(U^\top ZU) = U^\top ZU,$$

$$n^{1/2}\mathrm{tr}[\tilde{\psi}\{\phi(U^\top\hat{\Xi}U)\} - \Lambda] \to \mathrm{tr}(U^\top ZU) = 0.$$

By the first identity in (10), this shows that

$$n^{1/2}\psi_\Lambda\{\phi(U^\top\hat{\Xi}U)\} \to U^\top ZU,$$

in distribution, and then the required result follows from the second identity in (10). □

To analyze the standard errors, we express the projection $\psi_{\phi(\Xi)}\{\phi(\hat{\Xi})\}$ of $\phi(\hat{\Xi}) - \phi(\Xi)$ on to the tangent space to $\mathcal{P}_m(k-1)$ at the mean shape $\phi(\Xi)$ as a column vector $q_{U^\top\psi_{\phi(\Xi)}\{\phi(\hat{\Xi})\}U}$, of dimension $h = \frac{1}{2}m(2k-m-1)-1$, defined in (7). By (11), we have $\tilde{q}_{U^\top\psi_{\phi(\Xi)}\{\phi(\hat{\Xi})\}U} = \tilde{q}_{\psi_\Lambda\{\phi(U^\top\hat{\Xi}U)\}}$ and then it follows from the corollary that $n^{1/2}\tilde{q}_{\psi_\Lambda\{\phi(U^\top\hat{\Xi}U)\}}$ converges in distribution to $\tilde{q}_{U^\top ZU}$, and that $\tilde{q}_{U^\top ZU}$ is a Gaussian random vector with zero mean. Let Γ be the covariance matrix of $q_{U^\top ZU}$, assumed to be of full rank. Then, if $\tilde{\Gamma}$ is the $(h+1)\times(h+1)$ covariance matrix of $\tilde{q}_{U^\top ZU}$, we have $\Gamma = \tilde{H}\tilde{\Gamma}\tilde{H}^\top$. It now follows from Corollary 1 that

$$nq^\top_{U^\top\psi_{\phi(\Xi)}\{\phi(\hat{\Xi})\}U}\Gamma^{-1}q_{U^\top\psi_{\phi(\Xi)}\{\phi(\hat{\Xi})\}U} = nq^\top_{\psi_\Lambda\{\phi(U^\top\hat{\Xi}U)\}}\Gamma^{-1}q_{\psi_\Lambda\{\phi(U^\top\hat{\Xi}U)\}} \tag{17}$$

has a limiting χ^2_h distribution. An asymptotically equivalent, but more useful, version of (17) is obtained by interchanging the observed and true quantities to give

$$nq^\top_{\hat{U}^\top\psi_{\phi(\hat{\Xi})}\{\phi(\Xi)\}\hat{U}}\hat{\Gamma}^{-1}q_{\hat{U}^\top\psi_{\phi(\hat{\Xi})}\{\phi(\Xi)\}\hat{U}} = nq^\top_{\psi_{\hat{\Lambda}}\{\phi(\hat{U}^\top\Xi\hat{U})\}}\hat{\Gamma}^{-1}q_{\psi_{\hat{\Lambda}}\{\phi(\hat{U}^\top\Xi\hat{U})\}},$$

where $\hat{\Gamma}$ is the natural sample analogue of Γ. This statistic also has a limiting χ^2_h distribution. It can either be used directly for inference or can be used in a bootstrap procedure for one or several samples. Explicit expressions for Γ and $\hat{\Gamma}$ are given in the Appendix.

An alternative possibility, not considered further here, is to represent the population and sample mean ϕ-shapes as unit vectors and then use bootstrap procedures for unit vectors which have been developed for inference in one and several samples; see Fisher et al. (1996), Bhattacharya & Patrangenaru (2003, 2005) and Amaral et al. (2007). To be more specific, we may represent $\phi(\Xi) = \sum_{i=1}^{m} \lambda_i u_i u_i^\top$, where $\lambda_1 \geqslant \ldots \geqslant \lambda_m \geqslant 0$ and $\sum_{i=1}^{m} \lambda_i = 1$, as a unit vector $(\lambda_1^{1/2} u_1^\top, \ldots, \lambda_m^{1/2} u_m^\top)^\top$, with a corresponding unit vector representation $(\hat{\lambda}_1^{1/2} \hat{u}_1^\top, \ldots, \hat{\lambda}_m^{1/2} \hat{u}_m^\top)^\top$ for the sample mean shape.

5. Comparisons

5·1. *Comparing mean shape estimators*

We now discuss the relationship between the sample mean ϕ-shape defined in this paper and various commonly-used mean shape estimators, in particular, the full and partial Procrustes sample mean shapes. Suppose that $X_1, \ldots, X_n$ are given on the pre-shape sphere $\mathcal{S}^k_m$. For a general pre-shape $X \in \mathcal{S}^k_m$, write $[X]$ for the corresponding shape, where the latter is defined as the equivalence class of pre-shapes with the same shape as X; see Dryden & Mardia (1998, p. 56) and Kendall et al. (1999, p. 12). Then we denote by $[X_0]$ the mean shape of $[X_1], \ldots, [X_n]$, with respect to the penalty function g, if $[X_0]$ minimizes

$$\sum_{i=1}^{n} g(\rho_i), \tag{18}$$

where $\rho_i = \rho([X_i], [X])$ is the Riemannian distance between $[X_i]$ and a variable shape $[X]$, and the penalty function g is usually taken to be positive and increasing with $g(0) = 0$; see Kent (1992) and Dryden & Mardia (1998, pp. 87-95), where various candidates for g have been proposed. In particular, the full Procrustes mean shape corresponds to $g(\rho) = \sin^2(\rho)$, the partial Procrustes mean shape corresponds to $g(\rho) = \sin^2(\rho/2)$, and the mean shape with respect to the Riemannian distance corresponds to $g(\rho) = \rho^2$. The mean shape with respect to the penalty function g is identical to the maximum likelihood estimate of shape for the rotationally symmetric shape distribution with density

$$c_g(\kappa)^{-1} \exp\{-\kappa g(\rho)\}, \tag{19}$$

with respect to the uniform measure; see Dryden (1991) and Dryden & Mardia (1998, p. 198).

Note that $[X]$ is a shape, as opposed to a reflection shape, whereas the ϕ-shape, as defined in this paper, corresponds to the identification of a shape and its reflection, $[X]$ and $[X^R]$, say;

given X, we may choose X^R to be a matrix of the form AX, where A is any matrix in $\mathcal{O}(m)$ with determinant -1. Then the reflection shape of X is given by the union of the equivalence classes $[X]$ and $[X^R]$.

Let X_0 be a given point in $\mathcal{S}_m^k$ and write ρ_i for the Riemannian distance between the shapes of X_i and X_0. Then, we may express each X_i in terms of X_0 as

$$R_i X_i = \cos\rho_i X_0 + \sin\rho_i Z_i, \tag{20}$$

for some $R_i \in \mathcal{SO}(m)$, where $\mathcal{SO}(m)$ denotes the space of $m \times m$ rotation matrices, and $Z_i \in \mathcal{S}_m^{k-1}$ is such that $\mathrm{tr}(X_0 Z_i^\top) = 0$ and $X_0 Z_i^\top$ is symmetric; see Kendall et al. (1999, pp. 107-11). Note that Z_i is in fact the normalized Procrustes tangent projection of X_i at X_0; see Kent & Mardia (2001).

We first state a necessary condition, in terms of ρ_i and Z_i, for the shape of X_0 to be the mean shape of $X_1, \ldots, X_n$, with respect to the penalty function g.

LEMMA 1. *Assume that $\rho_i < \pi/2$ for $i = 1, \ldots, n$. If $[X_0]$ is the mean shape of $[X_1], \ldots, [X_n]$ with respect to the continuously differentiable penalty function g, then*

$$\sum_{i=1}^n g'(\rho_i)\, Z_i = 0. \tag{21}$$

In particular, if $[X_0]$ is the full Procrustes mean shape of $[X_1], \ldots, [X_n]$, then

$$\sum_{i=1}^n \sin(2\rho_i)\, Z_i = 0; \tag{22}$$

and if $[X_0]$ is the partial Procrustes mean shape of $[X_1], \ldots, [X_n]$, then

$$\sum_{i=1}^n \sin\rho_i\, Z_i = 0. \tag{23}$$

Proof. If we regard the Riemannian distance ρ_i as a function of the shape $[X_0]$, the corresponding Z_i is just the horizontal lift of the tangent vector $\mathrm{grad}\,\rho_i$ to the tangent space to $\mathcal{S}_m^k$ at X_0. On the other hand, $[X_0]$ is the mean shape of $[X_1], \ldots, [X_n]$ with respect to the penalty function g if and only if the function (18) achieves its global minimum at the shape $[X_0]$. This implies that if $[X_0]$ is the mean shape of $[X_1], \ldots, [X_n]$ with respect to g then, at $[X_0]$, we have

$$\mathrm{grad}\left(\sum_{i=1}^n g(\rho_i)\right) = \sum_{i=1}^n g'(\rho_i)\,\mathrm{grad}\rho_i = 0.$$

Since the horizontal lift is a linear isometry, the above is equivalent to the horizontal lift of the tangent vector $\sum_{i=1}^n g'(\rho_i)\,\mathrm{grad}\,\rho_i$ to the tangent space at X_0 being a zero vector; that is,

$$\sum_{i=1}^n g'(\rho_i)\, Z_i = 0,$$

as required.

Taking $g(\rho) = \sin^2(\rho)$ and $\sin^2(\frac{1}{2}\rho)$, we obtain the special conditions (22) and (23) for the full and partial Procrustes mean shapes respectively. □

One application of the lemma is to compare, in small neighborhoods, the sample mean reflection shape calculated using the penalty functions g, and the sample mean ϕ-shape.

THEOREM 2. *Assume that the penalty function g has the property that, for some constants $\alpha \neq 0$ and β, $g'(\rho) = \alpha\rho + \beta\rho^2 + o(\rho^2)$ as $\rho \to 0$. Let the shape of $X_0 \in \mathcal{S}_m^k$ denote the sample mean shape with respect to g of a sample of pre-shapes $X_1, \ldots, X_n$, and write $\rho_i = \rho([X_0], [X_i])$, $i = 1, \ldots, n$, for the Riemannian distance between $[X_0]$ and $[X_i]$. Fix n and suppose that $\max_{i=1}^n \rho_i \leq \epsilon$ where ϵ is small. Then $\|X_0^\top X_0 - \phi(\hat{\Xi})\| = O(\epsilon^2)$, where $\phi(\hat{\Xi})$ is the sample mean ϕ-shape of $X_1, \ldots, X_n$.*

Proof. On the one hand we have, by (20),

$$\begin{aligned}
\hat{\Xi} &= \frac{1}{n}\sum_{i=1}^n X_i^\top X_i \\
&= \frac{1}{n}\sum_{i=1}^n (\cos\rho_i)^2 X_0^\top X_0 + \frac{1}{n}\sum_{i=1}^n (\sin\rho_i\cos\rho_i)(X_0^\top Z_i + Z_i^\top X_0) \\
&\quad + \frac{1}{n}\sum_{i=1}^n (\sin\rho_i)^2 Z_i^\top Z_i \\
&= X_0^\top X_0 + \frac{1}{n}\sum_{i=1}^n \rho_i(X_0^\top Z_i + Z_i^\top X_0) \\
&\quad - \frac{1}{n}\sum_{i=1}^n \rho_i^2 X_0^\top X_0 + \frac{1}{n}\sum_{i=1}^n \rho_i^2 Z_i^\top Z_i + o(\epsilon^2)
\end{aligned}$$

and, on the other hand, the equality (21) holds by the Lemma 1. However, under the assumptions of Theorem 2, (21) becomes

$$\sum_{i=1}^n \rho_i Z_i = -\frac{\beta}{\alpha}\sum_{i=1}^n \rho_i^2 Z_i + o(\epsilon^2),$$

so that we can simplify (24) to

$$\hat{\Xi} = X_0^\top X_0 - \frac{\beta}{\alpha\, n}\sum_{i=1}^n \rho_i^2(X_0^\top Z_i + Z_i^\top X_0) - \frac{1}{n}\sum_{i=1}^n \rho_i^2 X_0^\top X_0 + \frac{1}{n}\sum_{i=1}^n \rho_i^2 Z_i^\top Z_i + o(\epsilon^2). \quad (24)$$

We now apply the map ϕ to each side of (24). Using standard perturbation analysis for eigen-expansions of matrices, see for example Sibson (1979), an application of ϕ to the right-hand side (24) yields $X_0^\top X_0 + O(\epsilon^2)$. Therefore $\|\phi(\hat{\Xi}) - X_0^\top X_0\| = O(\epsilon^2)$ and the proof is complete. □

The following tables present a numerical comparison of mean ϕ-shapes with the full and partial Procrustes means using simulated samples of the reflection shapes of random configurations having distributions of the type $N(\mu, \sigma^2 I_m \otimes I_k)$, where μ is a configuration of k labelled vertices in $\mathbb{R}^m$. In particular, we look at the cases where the pre-shape of μ is an $m \times (k-1)$ matrix with singular value decomposition $U(\Lambda/\|\Lambda\|, 0_{m,k-m-1})V^\top$, where Λ is of the form

$$\Lambda = \begin{cases} \text{diag}\{1, \ell\} & m = 2 \\ \text{diag}\{1, 1, \ell\} & m = 3, \end{cases}$$

$\|\Lambda\|$ denotes the Euclidean norm of Λ, $U \in \mathcal{O}(m)$ and $V \in \mathcal{SO}(k-1)$. Without loss of generality, we assume that $U = I_m$ and $V = I_{k-1}$. Note that, if $\ell = 1$, the shape of μ is that which is furthest from the collinearity set of the shape space; see Kendall et al. (1999, p. 130) for details. As ℓ increases, the distance between the shape of μ and the collinearity set decreases, and it approaches zero as ℓ tends to infinity. A simulation example is presented in Table 1. For sample sizes of $n = 30$ and $n = 100$, and for various values of σ and of ℓ which determine respectively the variance parameter of the induced shape distribution and the Riemannian distance of the shape of μ from the corresponding collinearity set, we generate 100 random samples of size n. For each such sample, we calculate its mean ϕ-shape as proposed in this paper, denoted by M, its partial Procrustes mean shape with reflection, denoted by P, and its full Procrustes mean shape with reflection, denoted by F. For 100 simulated datasets the corresponding mean and standard deviation of the Riemannian distances $\tilde{\rho}$ on the reflection shape space between the different estimators are given.

INSERT TABLE 1 ABOUT HERE

The results in the table clearly show that these means are similar to each other for $\sigma = 0.01$ and $\sigma = 0.1$, although the mean ϕ-shape is further from the other two Procrustes estimators on average. The mean ϕ-shape is particularly far away when $\ell = 100$, i.e. near the collinearity set. The mean ϕ-shape is not particularly close when $\sigma = 0.5$, in which case the small-distance assumptions do not hold. The findings are similar for both sample sizes, with standard deviations usually smaller for the larger sample size, as expected, and average Riemannian distance a little smaller too. Exceptions can be seen for the near-collinear case $\ell = 100$ for larger variations where the performance is fairly similar for both sample sizes.

The difference in speed of calculation of the estimators can be considerable for large sample sizes. For example, in a simulation study consisting of 100 Monte Carlo runs using R on a $2.8GHz$ Linux PC, with $k = 5, m = 3, n = 10000, \sigma = 0.1$ and $\ell = 1$, the following results were obtained. For the mean ϕ-shape calculation, the mean time was 5·12 seconds with a standard deviation of 0·22, and the corresponding mean time and standard deviation for the full Procrustes mean calculation were 180·94 seconds and 1·177, respectively.

5·2. *Relationship of Procrustes coordinates with ϕ-shape tangent coordinates*

In this section we focus on the relationship between the ϕ-shape tangent coordinates and those obtained using the Procrustes tangent projection. The situation that we are concerned with is addressed in Theorem 2, where the sample points in the pre-shape space are concentrated. In particular, let T denote the coordinates of the Procrustes tangent projection of $[X]$ on to the tangent space at $[M]$, where X and M are two elements in the pre-shape space. We assume that $[X]$ is close to $[M]$, which is formalized as $\|T\| = O(\epsilon)$ for small ϵ. From Kent & Mardia (2001) it follows that

$$T = \Gamma X - \cos\rho\left([X],[M]\right) M, \tag{25}$$

where $\Gamma \in \mathcal{O}\left(m\right)$ is such that $\Gamma X M^T$ is symmetric. Since $\mathrm{tr}(M^\top T) = 0$ and $\mathrm{tr}(M^\top M) = 1$, it follows that $\mathrm{tr}(M^\top \Gamma X) = \cos\rho\left([X],[M]\right)$, so that $\mathrm{tr}(T^\top T) = 1 - \{\cos\rho\left([X],[M]\right)\}^2$. Therefore, since by assumption $\|T\| = O(\epsilon)$, we deduce that $\cos\rho\left([X],[M]\right) = 1 + O(\epsilon^2)$, and so

$$T = \Gamma X - M + O(\epsilon^2).$$

LEMMA 2. *If T is the Procrustes tangent projection of the shape of X at the shape of M then for any $V \in \mathcal{O}(m)$ and $U \in \mathcal{O}(k-1)$ the corresponding projection of $V^\top XU$ at the shape of $V^\top MU$ is $V^\top TU$.*

Proof. This result follows immediately from (25) and the fact that the map in pre-shape space

$$Y \mapsto V^\top Y U \quad V \in \mathcal{O}(m) \quad \text{and} \quad U \in \mathcal{O}(k-1)$$

induces an isometry in the corresponding shape space. □

LEMMA 3. *If T is Procrustes tangent projection of $[X]$ at the tangent space of $[M]$, then if $\|T\|$=$O(\epsilon)$ the corresponding tangent space projection $\psi_{M^\top M}(X^\top X)$ satisfies*

$$\psi_{M^\top M}(X^\top X) = M^\top T + T^\top M + O(\epsilon^2). \tag{26}$$

Proof. Using the singular value decomposition we may write $M = VM_1U^\top$ where $V \in \mathcal{O}(m)$, $U \in \mathcal{O}(k-1)$ and

$$M_1 = \left(\mathrm{diag}(\lambda_1, \ldots, \lambda_m), 0_{m\times(k-m-1)}\right).$$

Therefore, from (10),

$$\psi_{M^\top M}(X^\top X) = U\psi_{M_1^\top M_1}(U^\top X_1^\top X_1 U)U^\top, \tag{27}$$

where $X_1 = V^\top XU$. If T_1 is the Procrustes tangent projection of $[X_1]$ at $[M_1]$, then it follows from Lemma 2 that $T_1 = V^\top TU$, where $\|T_1\| = \|T\| = O(\epsilon)$.

Therefore, from the relationship

$$T_1 = \Gamma_1 X_1 - M_1 + O(\epsilon^2),$$

where $\|T_1\| = O(\epsilon)$ and $\Gamma_1 \in \mathcal{O}(m)$ is chosen so that $\Gamma_1 X_1 M_1^\top$ is symmetric, it follows that

$$X_1 = \Gamma_1^\top (T_1 + M_1) + O(\epsilon^2),$$

and so

$$
\begin{aligned}
X_1^\top X_1 &= (T_1 + M_1)^\top (T_1 + M_1) + O(\epsilon^2) \\
&= T_1^\top T_1 + M_1^\top M_1 + M_1^\top T_1 + T_1^\top M_1 + O(\epsilon^2) \\
&= M_1^\top M_1 + M_1^\top T_1 + T_1^\top M_1 + O(\epsilon^2), \qquad (28)
\end{aligned}
$$

where the last step is a consequence of the fact that $T_1^\top T_1 = O(\epsilon^2)$.

Note that, by (10),

$$
\begin{aligned}
\psi_{M_1^\top M_1}(X_1^\top X_1) &= \tilde{\psi}(X_1^\top X_1) - M_1^\top M_1 \\
&\quad - \frac{1}{m}\mathrm{tr}\{\tilde{\psi}(X_1^\top X_1) - M_1^\top M_1\} \begin{pmatrix} I_m & 0 \\ 0 & 0_{m\times(k-m-1)} \end{pmatrix}. \qquad (29)
\end{aligned}
$$

Since $M_1 = \big(\mathrm{diag}(\lambda_1, \ldots \lambda_m), 0_{m\times(k-m-1)}\big)$, we see that

$$
\tilde{\psi}(M_1^\top M_1) = M_1^\top M_1, \qquad \tilde{\psi}(M_1^\top T_1 + T_1^\top M_1) = M_1^\top T_1 + T_1^\top M_1,
$$

so that

$$
\tilde{\psi}(X_1^\top X_1) = M_1^\top M_1 + M_1^\top T_1 + T_1^\top M_1 + O(\epsilon^2). \qquad (30)
$$

Moreover, as T_1 is in the tangent space of $[M_1]$, $\mathrm{tr}(M_1^\top T_1) = 0$, which implies that

$$
\mathrm{tr}\left(\tilde{\psi}(X_1^\top X_1) - M_1^\top M_1\right) = O(\epsilon^2). \qquad (31)
$$

Therefore, equations (30) and (31) imply that (29) simplifies to

$$
\begin{aligned}
\psi_{M_1^\top M_1}(X_1^\top X_1) &= M_1^\top M_1 + M_1^\top T_1 + T_1^\top M_1 + O(\epsilon^2) - M_1^\top M_1 \\
&= M_1^\top T_1 + T_1^\top M_1 + O(\epsilon^2),
\end{aligned}
$$

where $T_1 = O(\epsilon)$. Finally, since $T = VT_1U^\top$ and $M = VM_1U^\top$, and from (10),

$$
\begin{aligned}
\psi_{M^\top M}(X^\top X) &= U\psi_{M_1^\top M_1}(U^\top X^\top XU)U^\top \\
&= U\psi_{M_1^\top M_1}(X_1^\top X_1)U^\top \\
&= U\{M_1^\top T_1 + T_1^\top M_1 + O(\epsilon^2)\}U^\top \\
&= M^\top T + T^\top M + O(\epsilon^2),
\end{aligned}
$$

which concludes the proof. □

5·3. *Discussion*

Theorem 2 and Lemma 3 together indicate that for concentrated samples the coordinates of the observations for the two types of tangent projection are related by a fixed linear transformation which depends on the projection point. Therefore, any statistical procedure which is invariant with respect to linear transformation of the observations, such as Hotelling T^2 tests, will produce similar outcomes using either set of coordinates when the data are highly concentrated. This finding is particularly useful given that the calculation of the coordinates of the ϕ-shape tangent projection is, in general, much quicker than that of the Procrustes tangent coordinates when $m \geqslant 3$. Principal components analysis based on the ϕ-shape tangent coordinates may be used to study variability in the sample, as an alternative to Procrustes tangent coordinates; a numerical example is given in §7.

It is debatable which tangent coordinate system is to be preferred. However, one point to bear in mind is that isotropy in the Procrustes tangent coordinate systen does not imply, nor is it implied by, isotropy in the ϕ-shape tangent coordinate system. Moreover, one might expect the difference in these isotropy assumptions to have a tendency to be greater when the eigenvalues of $MM^\top$ in Lemma 3 differ appreciably. Arguably, isotropy is accommodated more naturally and transparently in the Procrustes tangent coordinate system than in the ϕ-shape tangent coordinate system, so in some circumstances we may prefer to use Procrustes tangent coordinates. However, even if the Procrustes system is preferred, the mean ϕ-shape provides a computationally convenient way to obtain approximate Procrustes tangent coordinates via Lemma 3, assuming that the data are highly concentrated.

6. Mean ϕ-shape and the Bingham distribution

The Bingham distribution distribution for a random matrix $X \in \mathcal{S}_m^{k-1}$ with unit norm has density

$$f(X|B) = c(B)^{-1} \exp\left\{\operatorname{tr}(XBX^\top)\right\} \tag{32}$$

with respect to the volume measure on $\mathcal{S}_m^{k-1}$, where $c(B)$ is the normalizing constant given by

$$c(B) = \int_{\mathcal{S}_m^k} \exp\left\{\operatorname{tr}(XBX^\top)\right\} dX = {}_1F_1\left(\frac{1}{2}, \frac{m(k-1)}{2}; B \otimes I_m\right),$$

${}_1F_1(1/2, m(k-1)/2; \cdot)$ is a hypergeometric function of matrix argument and $\otimes$ denotes the Kronecker product; see Bingham (1974) and Mardia & Jupp (2000). If B has the spectral decomposition $B = V\Omega V^\top$, where $\Omega = \operatorname{diag}(\omega_1, \ldots, \omega_{k-1})$ and $\omega_1 \geqslant \ldots \geqslant \omega_{k-1}$, then

$$\phi(\Xi) = V\operatorname{diag}(\lambda_1, \ldots, \lambda_m, 0, \ldots, 0)V^\top,$$

where

$$\lambda_i = \frac{\partial c(B)}{\partial \omega_i} \bigg/ \sum_{j=1}^m \frac{\partial c(B)}{\partial \omega_j}.$$

To see this, let $Y = XV$. Then, Y has the Bingham distribution with density

$$f(Y|\Omega) = c(\Omega)^{-1} \exp\left\{\operatorname{tr}(Y\Omega Y^\top)\right\},$$

where $c(\Omega) = c(B)$. It can be checked that $E(Y^\top Y)$ is diagonal and so, by property (i) in §3, the eigenvectors of $E(X^\top X)$ are the same as those of B, that is, $U = V$. On the other hand, $E(Y^\top Y) = \Delta$, where Δ is, as before, the diagonal matrix whose diagonal entries are the eigenvalues of $E(X^\top X)$. Then, since

$$\delta_i = c(\Omega)^{-1} \int_{\mathcal{S}_m^k} \sum_{s=1}^{m} y_{si}^2 \exp\left(\sum_{\ell=1}^{m} \sum_{j=1}^{k-1} y_{\ell j}^2 \omega_j \right) dY,$$

the δ_i are functions of Ω satisfying the condition that

$$\delta_i = c(\Omega)^{-1} \frac{\partial c(\Omega)}{\partial \omega_i}.$$

Note that the density (32) can also be expressed in vector form as

$$f(X|B) = c(B)^{-1} \exp\{\mathrm{vec}(X)^\top (B \otimes I_m) \mathrm{vec}(X)\}.$$

If $X_1, \ldots, X_n$ is a random sample from the Bingham distribution (32), then the ordered eigenvectors of the sample mean $\hat{\Xi}$ are the maximum likelihood estimators of the ordered eigenvectors of B, where the ordering of the eigenvectors is given by the ordering of the corresponding eigenvalues, and $\hat{\Omega}$, the maximum likelihood estimator of Ω, is that Ω which maximizes

$$\ell(\Omega) = -\log c(\Omega) + \sum_{j=1}^{k} \omega_j \hat{\delta}_j.$$

Note that Chikuse & Jupp (2004) discuss this Bingham distribution in relation to tests for uniformity when using a Euclidean embedding similar to (1).

The complex Bingham distribution has been used by Kent (1994) as a model for statistical analysis of planar shapes. Approximations to the normalizing constant $c(B)$ have been studied in Kume & Wood (2005).

7. Application

7·1. *Brain surface dataset*

In Brignell et al. (2007), a dataset of cortical surfaces of brains of schizophrenia patients and controls obtained from magnetic resonance scans is considered. The dataset consists of a very large number $k = 62501$ of pseudo-landmarks located on the cortical surface in $m = 3$ dimensions. In Brignell et al. (2007), the primary interest was in studying the asymmetry of the brain, and the data were regarded as being in fixed registrations. We shall now consider overall shape analysis of the cortical surfaces, where translation, rotation and scale can be ignored.

We have a total of 74 scans consisting of 44 controls and 30 schizophrenia patients. It is of interest to assess whether or not there is a significant difference in mean cortical surface shape between the two groups. We shall use the ϕ-shape defined in this paper, and compare the results with Procrustes analysis. Recall that the eigenvectors and eigenvalues of $X^\top X$ can be computed using the eigenvalues and eigenvectors of $XX^\top$, see for example Brignell et al. (2007), and so the mean ϕ-shape can be calculated quickly here.

First of all the mean ϕ-shape and Procrustes mean shape were calculated for the whole pooled dataset. The Riemannian shape distance between the Procrustes mean and mean ϕ-shape of the pooled data is 0·00023 while the diameter of the pooled sample is 0·08779. Since this difference is small there is no visible difference between the Procrustes mean and mean ϕ-shape. The distances between the Procrustes mean and mean ϕ-shape are 0·00023 for the control group and 0·00024 for the schizophrenia group.

We now consider comparisons between the control and schizophrenia groups. The distances between the means of each group are also quite small. The distance between the mean ϕ-shape of the control versus Schizophrenia groups is 0·00790 and and the corresponding distance between the Procrustes mean shapes is 0·00789.

In Fig. 1 we show the exaggerated differences between the mean ϕ-shapes of each group. Note that the corresponding figure for the Procrustes means is visibly identical.

INSERT FIGURE 1 ABOUT HERE

7·2. *Principal component analysis of the tangent space coordinates*

Even though the distance between the pooled mean estimators is very small, the ϕ-shape tangent projection of the pooled sample at the pooled mean ϕ-shape is noticeably different from the Procrustes tangent space projection at the pooled Procrustes mean, as expected from §5·2. Fig. 2 shows the first two principal components of the data at the tangent spaces projections. In particular, for the Procrustes tangent space projection the first five principal components explain 20·77+14·20+8·24+5·70+4·20=53·11% of the variation, whereas for the mean ϕ-shape they explain 25·60+12·35+ 9·61+6·97+4·26=58·78% of variation. The ϕ-shape tangent space coordinates are most simply calculated using principal coordinate analysis in this application. In particular, given the pre-shapes $Z_1, \dots, Z_n$, all pairs of distances are calculated in the embedded Euclidean space, i.e.

$$d_{ij} = \|Z_i^\top Z_i - Z_j^\top Z_j\|, \quad 1 \leqslant i, j \leqslant n,$$

and principal coordinate analysis is calculated using these pairwise distances (Mardia et al., 1979, p. 405).

7·3. *Two sample hypothesis test*

We performed a hypothesis test to examine if the mean shape is the same in each group. The corresponding values of the F statistic based on the two-sample Goodall test, as described in Dryden & Mardia (1998, p. 162), are 0·9989 for the mean ϕ-shape and 0·9369 for the Procrustes mean, and applying a permutation test using this statistic leads to p-values of 0·54 and 0·53, respectively. Hence there is no evidence for an overall shape difference between the two groups. Note that the assumptions for this F test are based on an assumption of isotropy in the relevant tangent space; for relevant discussion, see §5·3.

INSERT FIGURE 2 ABOUT HERE

Acknowledgement

We are grateful to three reviewers for constructive comments which have led to improvements in the presentation, and to the U.K. Engineeing and Physical Sciences Research Council for financial support.

Appendix

Calculation of Γ *and* $\hat{\Gamma}$

We now derive explicit expressions for $\Gamma = \text{cov}(q_{U^\top ZU})$ and $\hat{\Gamma}$, a natural sample analogue of Γ, where $q_S = \tilde{H}\tilde{q}_S$ for $S \in \mathcal{T}_\Lambda\{\mathcal{P}_m(k-1)\}$; see (7) and Theorem 1 for notation. First, observe that $\tilde{q}_S = C\,\text{vec}(S)$ where the vec operator stacks columns in the usual way, and C is a constant matrix of dimension $\{\frac{1}{2}m(2k-m-1)\} \times (k-1)^2$. In the case $m = 3$, the appropriate specification of $C = (c_{ij})$ is as follows: $c_{11} = c_{2,k+1} = c_{3,2k+1} = 1$; $c_{ij} = 2^{1/2}$ for $i = 4, \ldots, k+1$ and $j = i-2$; $c_{ij} = 2^{1/2}$ for $i = k+2, \ldots, 2k-2$ and $j = i$; $c_{ij} = 2^{1/2}$ for $i = 2k-1, \ldots, 3k-6$ and $j = i+3$; and $c_{ij} = 0$ otherwise.

It follows from (13) that

$$U^\top ZU = \frac{1}{\delta_1 + \ldots + \delta_m}\left\{\tilde{G}^u - \text{tr}(\tilde{G}^u)\phi(\Delta)\right\}, \qquad \text{(A1)}$$

where $U^\top ZU \in \mathcal{T}_{\phi(\Delta)}\{\mathcal{P}_m(k-1)\}$. From the definition of $\tilde{G}^u$ in (14) and of G^u just above the statement of Theorem 1, it is seen that

$$\text{vec}(\tilde{G}^u) = L\,\text{vec}(G^u) = L\,\text{vec}(U^\top GU) = L\,(U^\top \otimes U^\top)\text{vec}(G), \qquad \text{(A2)}$$

where we have made use of some basic properties of the Kronecker product; see for example Mardia et al. (1979, p. 460). The matrix L is of dimension $(k-1)^2 \times (k-1)^2$ and is defined by $L = \text{diag}(\ell_{11}, \ldots, \ell_{k-1,1}, \ell_{2,1}, \ldots, \ell_{k-1,k-1})$, i.e. the ordering of the diagonal elements ℓ_{ij}, $i, j = 1, \ldots, k-1$, of L is the same as the ordering imposed by the vec operator, and

$$\ell_{ij} = \begin{cases} 1 & \text{if } 1 \leqslant i, j \leqslant m \\ \delta_i/(\delta_i - \delta_j) & \text{if } 1 \leqslant i \leqslant m < j \leqslant k-1 \\ \delta_j/(\delta_j - \delta_i) & \text{if } 1 \leqslant j \leqslant m < i \leqslant k-1 \\ 0 & \text{otherwise.} \end{cases}$$

Also,

$$\text{tr}(\tilde{G}^u) = a^\top \text{vec}(U^\top G U) = a^\top (U^\top \otimes U^\top)\text{vec}(G), \tag{A3}$$

where $a = (a_{11}, \ldots, a_{k-1,1}, a_{2,1}, \ldots, a_{k-1,k-1})^\top$ is a column vector of dimension $(k-1)^2$, with elements a_{ij}, $i, j = 1, \ldots, k-1$, arranged in the same order as that imposed by the vec operator, and with $a_{ii} = 1$ for $i = 1, \ldots, m$, and $a_{ij} = 0$ otherwise. Consequently, using (A1) – (A3), we find that

$$\begin{aligned}
\text{vec}(U^\top Z U) &= \frac{1}{\delta_1 + \ldots + \delta_m}\left[\text{vec}(\tilde{G}^u) - \text{tr}(\tilde{G}^u)\text{vec}\{\phi(\Delta)\}\right] \\
&= \frac{1}{\delta_1 + \ldots + \delta_m}\left[L(U^\top \otimes U^\top)\text{vec}(G) - \text{vec}\{\phi(\Delta)\}a^\top(U^\top \otimes U^\top)\text{vec}(G)\right] \\
&= K(U^\top \otimes U^\top)\text{vec}(G),
\end{aligned}$$

where

$$K = \frac{1}{\delta_1 + \ldots + \delta_m}[L - \text{vec}\{\phi(\Delta)\}a^\top]. \tag{A4}$$

Therefore, $\Gamma = \text{cov}(q_{U^\top Z U})$ is given by

$$\Gamma = \tilde{H} C K (U^\top \otimes U^\top)\Sigma(U \otimes U)K^\top C^\top \tilde{H}^\top,$$

where $\Sigma = \text{cov}\{\text{vec}(G)\}$, with a typical element of Σ given by (12), and $\tilde{H}$ and C are as above.

A natural sample analogue of Σ is given by

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} \text{vec}(X_i^\top X_i - \hat{\Xi})\text{vec}(X_i^\top X_i - \hat{\Xi})^\top,$$

where $\hat{\Xi} = n^{-1}\sum_{i=1}^{n} X_i^\top X_i$. Therefore, a natural sample analogue of Γ is given by

$$\hat{\Gamma} = \tilde{H} C \hat{K}(\hat{U}^\top \otimes \hat{U}^\top)\hat{\Sigma}(\hat{U} \otimes \hat{U})\hat{K}^\top C^\top \tilde{H}^\top, \tag{A5}$$

where $\hat{U}$ is obtained from the spectral decomposition $\hat{\Xi} = \hat{U}\hat{\Delta}\hat{U}^\top$, $\hat{K}$ is the same as K in (A4), but with $\hat{\Delta} = \text{diag}(\hat{\delta}_1, \ldots, \hat{\delta}_{k-1})$ replacing Δ, and $\hat{\delta}_i$ replacing δ_i, $i = 1, \ldots, m$. If the population mean ϕ-shape is unique, in the sense explained in Remark 1, then $\hat{\Gamma}$ in (A5) is a consistent estimator of Γ.

References

Amaral, G.J.A., Dryden, I.L. & Wood, A.T.A. (2007). Pivotal bootstrap methods for k-sample problems in directional statistics and shape analysis. *J. Am. Statist. Assoc.* **102**, 695–707.

Bandulasiri, A. & Patrangenaru, V. (2005). Algorithms for nonparametric inference on shape manifolds. *Proceedings of the Joint Statistical Meetings 2005, Minneapolis*, 1617–22.

BHATTACHARYA, R.N. & PATRANGENARU, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds, I. *Ann. Statist.* **31**, 1–29.

BHATTACHARYA, R.N. & PATRANGENARU, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds, II. *Ann. Statist.* **33**, 1225–59.

BINGHAM, C. (1974). An antipodally symmetric distribution on the sphere. *Ann. Statist.* **2**, 1201–5.

BRIGNELL, C.J., DRYDEN, I.L, GATTONE, S.A., PARK, B, LEASK, S., BROWNE, W.J. & FLYNN, S. (2007). Surface shape analysis, with an application to brain cortical surface analysis in schizophrenia. Submitted for publication.

BOOKSTEIN, F. L. (1986). Size and shape spaces for landmark data in two dimensions (with Discussion), *Statist. Sci.* **1**, 181–242.

CARNE, T.K. (1990). The geometry of shape spaces. *Proc. Lond. Math. Soc.* **61**, 407–32.

CHIKUSE, Y. & JUPP, P.E. (2004). A test of uniformity on shape spaces. *J. Mult. Anal.* **88**, 163–76.

DRYDEN, I.L. (1991). Discussion of 'Procrustes methods in the statistical analysis of shape' by C.R. Goodall, *J. R. Statist. Soc.* B **53**, 327-8.

DRYDEN, I.L. & MARDIA, K.V. (1998). *Statistical Shape Analysis*. Chichester: John Wiley.

FISHER, N.I, HALL, P., JING, B-Y. & WOOD, A.T.A. (1996). Improved pivotal methods for constructing confidence regions with directional data. *J. Am. Statist. Assoc.* **91**, 1062–70.

GOODALL, C.R. (1991). Procrustes methods in the statistical analysis of shape (with Discussion). *J. R. Statist. Soc.* B **53**, 285-339.

HENDRIKS, H. & LANDSMAN, Z. (1998). Mean location and sample mean location on manifolds: asymptotics, tests, confidence regions. *J. Mult. Anal.* **67**, 227–43.

KENDALL, D.G. (1984). Shape manifolds, Procrustean metrics and complex projective spaces. *Bull. Lond. Math. Soc.* **16**, 81–121.

KENDALL, D.G., BARDEN, D., CARNE, T.K. & LE, H. (1999). *Shape and Shape Theory*. Chichester: John Wiley.

KENDALL, W.S. (1990). The diffusion of Euclidean shape. In *Disorder in Physical Systems*, Ed. by G.R. Grimmett and D.J.A. Welch, pp. 203-17. Oxford: Oxford University Press.

KENT, J.T. (1992). New directions in shape analysis. In *The Art of Statistical Science*, Ed. K.V. Mardia, pp. 115-27. Chichester: John Wiley.

KENT, J.T. (1994). The complex Bingham distribution and shape analysis. *J. R. Statist. Soc.* B **56**, 285–99.

KENT, J.T. & MARDIA, K.V. (2001). Shape, Procrustes tangent projections and bilateral symmetry. *Biometrika* **88**, 469-85.

KUME, A. & WOOD, A.T.A. (2005). Saddlepoint approximations for the Bingham and Fisher-Bingham normalising constants. *Biometrika* **92**, 465-76.

LELE, S. (1993). Euclidean distance matrix analysis (EDMA): estimation of mean form and mean form difference. *Math. Geol.* **25**, 573–602.

LELE, S.R. & RICHTSMEIER, J.T. (2001). *An Invariant Approach to Statistical Shape Analysis.* London: Chapman and Hall/CRC.

MARDIA, K.V. & JUPP, P.E. (2000). *Directional Statistics*. Chichester: John Wiley.

MARDIA, K.V. & KENT, J.T. & BIBBY, J.M. (1979). *Multivariate Analysis*. London: Academic Press.

SIBSON, R. (1979). Studies in the robustness of multidimensional scaling: perturbational analysis of classical scaling. *J.R. Statist. Soc.* B **41**, 217-29.

WATSON, G.S. (1983). *Statistics on Spheres*. University of Arkansas Lecture Notes in the Mathematical Sciences, Vol. 6. New York: John Wiley.

ZIEZOLD, H. (1977). On expected figures and a strong law of large numbers for random elements in quasi-metric spaces. In *Trans. Seventh Prague Conf. Info. Theory, Statist. Decision Functions, Random Processes*, Vol A, pp. 591–602. Dordrecht: Reidel.

σ	n	ℓ	$\tilde{\rho}$(M,P)	$\tilde{\rho}$(M,F)	$\tilde{\rho}$(P,F)
0·01	30	1	0·00009 (0·00003)	0·00008 (0·00002)	0·00005 (0·00001)
0·01	30	2·5	0·00039 (0·00005)	0·00038 (0·00004)	0·00005 (0·00001)
0·01	30	100	0·01064 (0·00205)	0·01059 (0·00300)	0·00061 (0·00111)
0·1	30	1	0·00923 (0·00274)	0·00792 (0·00240)	0·00436 (0·00146)
0·1	30	2·5	0·03705 (0·00582)	0·03679 (0·00560)	0·00455 (0·00129)
0·1	30	100	0·11054 (0·03524)	0·11046 (0·03546)	0·00550 (0·00416)
0·5	30	1	0·17477 (0·12023)	0·17513 (0·11816)	0·05103 (0·04606)
0·5	30	2·5	0·23811 (0·14863)	0·22598 (0·13284)	0·06044 (0·07165)
0·5	30	100	0·25954 (0·14862)	0·23886 (0·13492)	0·06337 (0·07555)
0·01	100	1	0·00005 (0·00001)	0·00004 (0·00001)	0·00003 (0·00001)
0·01	100	2·5	0·00038 (0·00003)	0·00038 (0·00002)	0·00003 (0·00001)
0·01	100	100	0·00881 (0·00111)	0·00877 (0·00107)	0·00035 (0·00047)
0·1	100	1	0·00511 (0·00137)	0·00453 (0·00129)	0·00246 (0·00070)
0·1	100	2·5	0·03474 (0·00258)	0·03453 (0·00236)	0·00259 (0·00063)
0·1	100	100	0·12185 (0·03790)	0·12107 (0·03740)	0·00509 (0·00696)
0·5	100	1	0·09443 (0·05135)	0·09362 (0·05030)	0·03325 (0·01740)
0·5	100	2·5	0·21089 (0·13551)	0·20869 (0·13141)	0·04515 (0·05817)
0·5	100	100	0·24782 (0·13402)	0·23138 (0·12681)	0·05623 (0·07129)

Table 1: Simulation study for $k = 5$ points in $m = 3$ dimensions. The parameter ℓ indicates the particular mean shape, σ is the standard deviation and n the sample size. For each such sample, we calculate its mean ϕ-shape, denoted by M, its partial Procrustes mean shape with reflection, denoted by P, and its full Procrustes mean shape with reflection, denoted by F. The corresponding Riemannian distances $\tilde{\rho}$ on the reflection shape space between these means are calculated and the mean value from 100 simulations is given, with standard deviation in brackets.

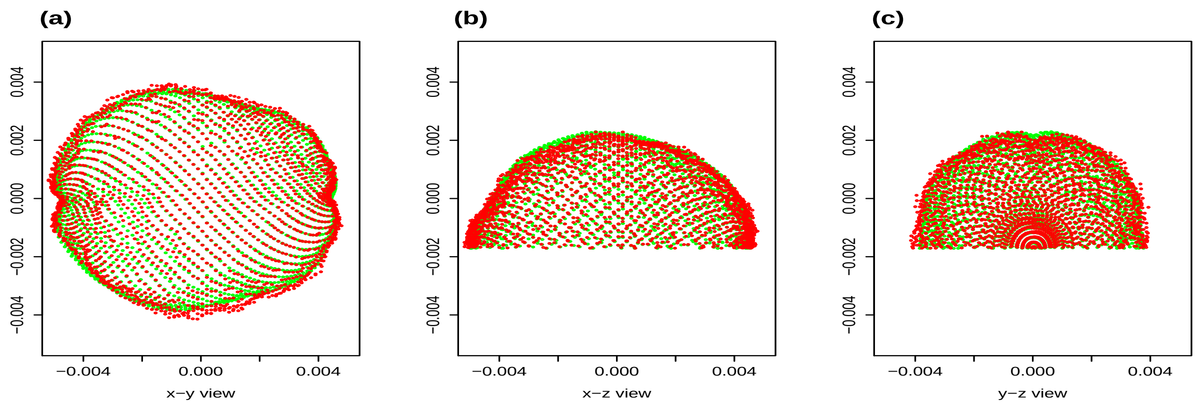

Figure 1: Brain surface dataset. Configurations representing the mean ϕ-shape for the control group (grey) and the schizophrenia group (black), where their differences are magnified 10 times from their common mean configuration. Figures 1(a), 1(b) and 1(c) give the views, respectively, from above, from the side and from behind. The configurations were initially Procrustes matched.

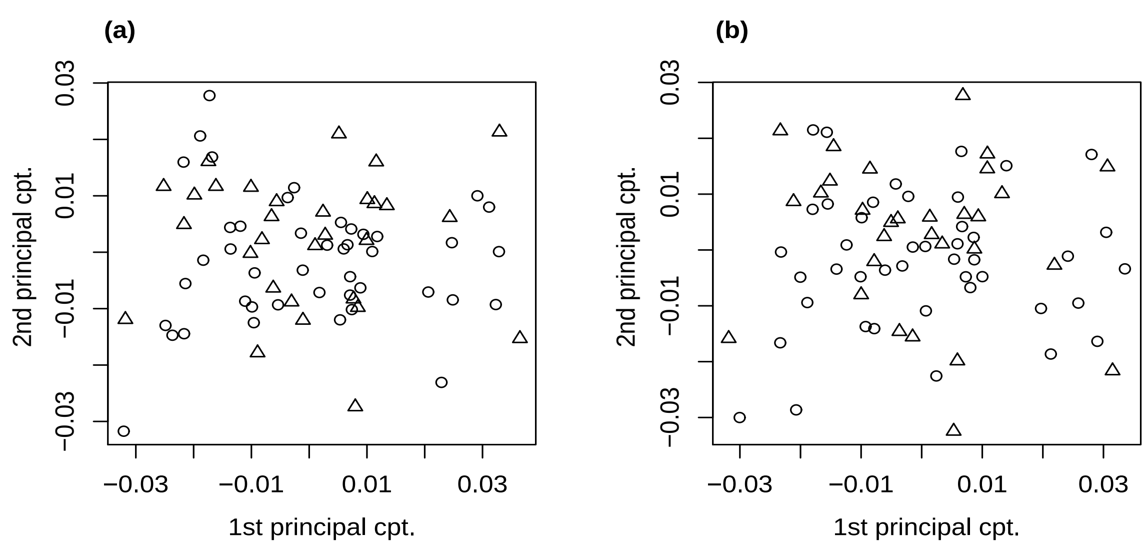

Figure 2: Brain surface dataset. Fig. 2(a) gives a scatterplot of the first two principal component scores in the mean ϕ-shape tangent space, while Fig. 2(b) gives the corresponding scatterplot in the Procrustes tangent space. Controls are given by circles; Schizophrenia patients are given by triangles