Principal Components Geodesics for Planar Shape Spaces

Stephan Huckemann,[*][†] Thomas Hotz[‡][§]

Abstract

In this paper a numerical method to compute principal component geodesics for Kendall's planar shape spaces - which are essentially complex projective spaces - is presented. Underlying is the notion of principal component analysis based on geodesics for non-Euclidean manifolds as proposed in an earlier paper by Huckemann and Ziezold (2006). Currently, principal component analysis for shape spaces is done on the basis of a Euclidean approximation. In this paper, using well studied datasets and numerical simulations, these approximation errors are discussed. Overall, the error distribution is rather dispersed. The numerical findings back the notion that the Euclidean approximation is good for highly concentrated data. For low concentration, however, the error can be strongly notable. This is in particular the case for a small number of landmarks. For highly concentrated data, stronger anisotropicity and a larger number of landmarks may also increase the error.

Keywords: Shape Analysis, Principal Component Analysis, Riemannian Manifolds, Geodesics, Complex Projective Space, Complex Bingham Distribution, Complex Watson Distribution

AMS 2000 Subject Classification: Primary 60D05
Secondary 62H11, 53C22

1 Introduction

The statistical study of geometrical shapes by extracting finite landmark configurations and mapping to metric quotient-spaces was independently triggered

by Bookstein (1978), Kendall (1977) and Ziezold (1977). Different concepts of "size" lead to different landmark based shape spaces, cf. Small (1996). Taking "size" as the Euclidean norm of a matrix, the resulting quotient spaces are commonly called *Kendall's shape spaces*. They are obtained as follows. Given a sample of original geometrical objects, from each object a landmark configuration matrix is extracted. In a first step, normalizing for location and size, this matrix is mapped to a point on the so called *pre-shape sphere*. In a second step, filtering out rotation, the pre-shape sphere is projected to the *shape space*, containing only the "shape information" of the original object. Shape spaces can be rather complicated. When planar objects are considered, the projection from the pre-shape sphere filtering out rotation is the well known *Hopf fibration* mapping to a complex projective space. This is a non-Euclidean positive curvature manifold.

As this geometry naturally arises through the subsequent normalization and projection steps it is very reasonable to consider statistical features based on that geometry as the "true" shape features. In a non-Euclidean geometry, however, statistical features such as means and principal components are solutions to non-linear problems and hence much harder to compute, than their kin that linearly compute in Euclidean space. Recently, general investigations by Bhattacharya and Patrangenaru (2003), Bhattacharya and Patrangenaru (2005) and methods to compute means on positive curvature manifolds with respect to the intrinsic metric have become available, cf. Le (2001). Most statistical methods, however, currently still rely on embeddings (cf. e.g. Micheas and Dey (2005), Chikuse and Jupp (2004)) or on projections onto suitable Euclidean spaces, such as principal component analysis (PCA), cf. e.g. Dryden and Mardia (1998). The latter approach of projecting the data into the tangent space of an extrinsic mean or mapping it under the inverse Riemannian exponential into the tangent space of an intrinsic mean (cf. e.g. Fletcher et al. (2004)) has been particularly popular as it allows for the use of standard statistical methodology. The prevailing conviction to the moment is that

$$\text{for highly concentrated data the Euclidean approximation} \atop \text{reflects very well intrinsic shape features.} \tag{1}$$

This statement, however, has not been investigated. We note that asymptotic considerations, cf. Hendriks and Landsman (1998), show that the spread of the extrinsic mean is in first order supported by its tangent space projection when embedding both manifold and tangent space in a Euclidean space. In order to investigate the above statement, however, we cannot argue asymptotically since (1) is a statement concerning the spread of the data on the shape manifold and not the spread of the mean.

It is the aim of this paper to provide for numerical methods of PCA for planar shape spaces based on geodesics in the intrinsic metric, i.e. without resorting to embeddings and/or projections to Euclidean spaces, and, on that basis to discuss the above statement. To this end the paper is structured as follows:

Section 2 reviews the concepts of PCA based on geodesics from Huckemann and Ziezold (2006). For the problem at hand, a key feature of non-Euclidean geometry is the fact, that *principal component geodesics* (PCGs) do not pass through *intrinsic means* (IMs). Hence, in order to find a first PCG, we have to minimize over both offset and initial velocity of competing geodesics. At this point we note that Fletcher et al. (2004) consider "medial manifolds" for shape representation. For these manifolds they determine "principal geodesics" which are PCGs that are constrained to pass through the IM.

Section 3 defines Kendall's shape spaces. Also, the current method of PCA by embedding into and projecting onto suitable Euclidean spaces is sketched briefly. In fact, this method computes approximations to PCGs that are constrained to pass through the *extrinsic mean* (EM).

Section 4 is devoted to the *optimal positioning* (cf. Ziezold (1977), also called *alignment* or *registration*) of configurations with respect to shape distance. The Hopf-fibration gives rise to so called *horizontal* and *vertical great circles* on the pre-shape sphere. From the detailed study of these circles, results for optimal positioning of data to a given geodesic are derived and the space of geodesics of planar shape spaces is determined.

Section 5 reformulates the problem of finding PCGs and IMs as an extremal problem under constraints, as has been done in a previous paper by Huckemann and Ziezold (2006) for spheres only. This leads in a natural way to an algorithmic approach. Having provided for a representation of the space of geodesics in the preceeding section, the numerical implementation can be pulled back to the pre-shape sphere.

Section 6 addresses the above conviction (1) by comparing *sums of squared distances* (SSDs) of datasets to *principal components* (PCs) obtained by Euclidean approximation on the one hand and obtained by minimization based on the intrinsic metric on the other hand. Also, mutual distances between *extrinsic mean* (EM), *intrinsic mean* (IM) and *principal component geodesic mean* (PM) are discussed. The discussion is based on existing data and on simulations using the complex Bingham distribution. Foci of the discussion are the correlations between concentration and anisotropy of the distribution as well as the number of landmarks on the one side and the relative improvement of the geodesic fit vs. the Euclidean fit on the other side.

2 PCA Based on Geodesics

For an m-dimensional Riemannian manifold M with induced metric $d(\cdot, \cdot)$ consider both

$$E\big(d(X, p)^2\big) \quad \text{and} \tag{2}$$

3

$$E\big(d(X,\gamma)^2\big) \tag{3}$$

for $p \in M$, a M-valued random variable X and a geodesic $\gamma \in G(M) := \{\gamma : \gamma$ is a geodesic on M, maximal w.r.t. inclusion$\}$.

A point $\overline{p} \in M$ minimizing (2) is called an *intrinsic mean* (IM) to X (also called an *intrinsic Fréchet mean*) and a geodesic $\gamma_1 \in G(M)$ minimizing (3) is called a *first principal component geodesic* (PCG) to X. A geodesic $\gamma_2 \in G(M)$ that minimizes (3) over all geodesics $\gamma \in G(M)$ that have at least one point in common with γ_1 and that are orthogonal to γ_1 at all points in common with γ_1 is called a *second PCG* to X. Every point $\hat{p}$ that minimizes (2) over all common points p of γ_1 and γ_2 is called a *principal component geodesic mean* (PM).

Given a first and a second PCG γ_1 and γ_2 with PM $\hat{p}$, a geodesic γ_3 is a *third PCG* if it minimizes (3) over all geodesics that meet γ_1 and γ_2 orthogonally at $\hat{p}$. Analogously, PCGs of higher order are defined.

In our context, one main feature of non-Euclidean geometry is the fact that in general, due to curvature, the IM will differ from the PM, even more, the IM will not come to lie on any of the PCGs. Also, due to curvature, different non-Euclidean generalizations of variance are possible. Cf. Huckemann and Ziezold (2006) for a detailed discussion.

3 Kendall's Shapes Spaces

Considering spheres in matrix spaces modulo the similarity transformation group acting on columns leads to *Kendall's* shape spaces. Denote by

$.^T$	the transposition of matrices,
$M(m,k)$	all real matrices with m rows and k columns identified with Euclidean $\mathbb{R}^{mk}$, i.e. with the inner product $\langle a,b \rangle := Tr(ab^T), \|a\| = \sqrt{\langle a,a \rangle}$,
$O(m)$	the orthogonal group in $M(m,m)$,
$SO(m)$	the special orthogonal group in $M(m,m)$.

3.1 Definition and Metric

Landmark based shape analysis is based on configurations consisting of k labelled vertices in $\mathbb{R}^m$ called *landmarks* that do not all coincide. Each configuration is a point in $M(m,k)$. Normalizing for location and size, these configurations are mapped by a Helmert matrix to the *pre-shape space* (cf. e.g. Dryden and Mardia (1998))

$$S_m^k := \{s \in M(m, k-1) : \|s\| = 1\}$$

which can be regarded as a hyper-sphere in Euclidean $\mathbb{R}^{m(k-1)}$. Additionally filtering out rotation leads to the definition of *shape space*: Define on S_m^k a smooth left action of $SO(m)$ by

$$gs \quad := \quad (gs_1, \ldots, gs_{k-1}) \in S_m^k \tag{4}$$

for $g \in SO(m)$ and $s = (s_1, \ldots, s_{k-1}) \in S_m^k$. Then the orbit $[s] = \{gs: g \in SO(m)\}$ is the *shape* of $s \in S$ and the topological quotient

$$\Sigma_m^k := S_m^k / SO(m)$$

is called the *shape space*.

Shape spaces of one-dimensional objects are just the corresponding pre-shape spheres as $SO(1) = \{id\}$ is trivial. In case of $m = 2$ shape spaces are complex projective spaces of real dimension $2(k-2)$, as we will see below. In particular Σ_2^3 is isometrical with the two-dimensional sphere $S^2(\frac{1}{2})$ of radius $1/2$. For $m \geq 3$ shape spaces no longer have an overall manifold structure compatible with its quotient topology, yet they still are metric Hausdorff spaces (cf. Kendall et al. (1999) for a detailed discussion).

The pre-shape sphere naturally carries the spherical metric, that is the induced metric of the Riemannian embedding

$$S_m^k \quad \hookrightarrow \quad M(m, k-1).$$

The tangent space at $s \in S_m^k$ is given by

$$T_s S_m^k \quad = \quad \{v \in M(m, k-1) : \langle v, s \rangle = 0\}.$$

The action of $SO(m)$ induces an isometric mapping of tangent spaces $dg : T_s S_m^k \to T_{gs} S_m^k$ given by

$$dg\, v \quad = \quad gv, \quad v \in T_s S_m^k. \tag{5}$$

For any two $a, b \in S_m^k$ the spherical distance is

$$0 \;\leq\; d(a,b) \;=\; 2 \cdot \arcsin\left(\frac{\sqrt{\langle a-b, a-b\rangle}}{2}\right) \;=\; \arccos\langle a, b\rangle \;\leq\; \pi. \tag{6}$$

Geodesics on spheres are precisely the great circles

$$\gamma: t \mapsto a \cos t + b \sin t$$

for any $a, b \in S_m^k$ with $\langle a, b \rangle = 0, t \in \mathbb{R}$. We denote this above great circle γ through a with initial unit velocity b by $\gamma_{a,b}$. The spherical distance of a point $p \in S_m^k$ to the great circle $\gamma_{a,b}$ is (cf. Huckemann and Ziezold (2006), Proposition 3.1)

$$0 \;\leq\; d(p, \gamma_{a,b}) \;=\; \arccos\sqrt{\langle p, a\rangle^2 + \langle p, b\rangle^2} \;\leq\; \frac{\pi}{2}\,, \tag{7}$$

the unique spherical projection of p onto $\gamma_{a,b}$ is given by (cf. Huckemann and Ziezold (2006), Corollary 3.2)

$$\frac{\langle a, p\rangle a + \langle b, p\rangle b}{\sqrt{\langle a, p\rangle^2 + \langle b, p\rangle^2}} \tag{8}$$

whenever $d(p, \gamma_{a,b}) < \frac{\pi}{2}$.

The spherical metric projects naturally to a metric on the shape space

$$d_\Sigma([a], [b]) := \inf_{g \in SO(m)} d(ga, b) \ \text{ for } a, b \in S_2^k.$$

The *shape distance*

$$d_\Sigma(a, b) := d_\Sigma([a], [b])$$

will be used for pre-shapes as well. Two pre-shapes $p, x \in S_m^k$ are in *optimal position* to each other (cf. Ziezold (1977)) if

$$d(p, x) = d_\Sigma(p, x).$$

As $SO(m)$ is compact, any pre-shape p can be rotated into optimal position to a given pre-shape x. We will denote the optimally rotated version of p with respect to x by p^x. p^x is also called the *partial Procrustes fit* of p onto x (cf. Dryden and Mardia (1998)). Optimal positioning is reflexive and symmetric but not transitive, i.e. for $x, p, q \in S_m^k$ in general $q^{p^x} \neq q^x$ (cf. Ziezold (1977), Example 2). Moreover in general, p^x is not uniquely determined (cf. Kendall et al. (1999), p. 121).

Similarly, a pre-shape $p \in S_m^k$ is said to be in *optimal position to a great circle* $\gamma_{x,v}$ with $x, v \in S_m^k$, $\langle x, v\rangle = 0$, if

$$d(p, \gamma_{x,v}) = \inf_{g \in SO(m)} d(gp, \gamma_{x,v}) =: d_\Sigma(p, \gamma_{x,v}).$$

3.2 Euclidean PCA for Shape Spaces

In addition to the spherical metric two extrinsic distances are considered on the pre-shape sphere: For $a, b \in S_m^k$ their distance in the *chordal metric* is given by $\|a - b\| = \sqrt{2(1 - \langle a, b\rangle)}$ wheras their distance in the *residual pseudo-metric* is $\|a - \langle a, b\rangle b\| = \sqrt{1 - \langle a, b\rangle^2}$. The former is obtained from embedding into the ambient Euclidean $M(m, k-1)$, the latter can be viewed as obtained from first orthogonally projecting to the tangent space of the pre-shape sphere and then by embedding this tangent space in the ambient Euclidean space.

Any two pre-shapes $p, x \in S_m^k$ that are in optimal position to each other are also in optimal position to each other with respect to chordal metric and vice versa. With respect to the residual pseudo-metric also $-p$ is in optimal position to x.

6

Given pre-shapes $p_1, \ldots, p_N \in S_m^k$ call a pre-shape $\widetilde{x} \in S_m^k$ a *pre-shape of an extrinsic mean* (EM) *shape*, or equivalently, a *pre-shape of a partial Procrustes mean shape*, if

$$\min_{g_1, \ldots, g_N \in SO(m)} \sum_{i=1}^{N} \|g_i p_i - \widetilde{x}\|^2 \;=\; \min_{x \in S} \left(\min_{g_1, \ldots, g_N \in SO(m)} \sum_{i=1}^{N} \|g_i p_i - x\|^2 \right).$$

Such a mean is computed with standard algorithms, cf. Gower (1975) and Ziezold (1994): In every step the estimate of the mean is updated to the projection onto the pre-shape sphere of the Euclidean mean to that data, that is optimally positioned to the previous estimate.

Euclidean PCA for Kendall's shape spaces is performed as follows, see e.g. Dryden and Mardia (1998): having found a pre-shape $\widetilde{x} \in S_m^k$ of an EM shape, all data points are brought into optimal position to $\widetilde{x}$ and projected onto the tangent space $T_{\widetilde{x}} S$. Then with respect to the residuals

$$r_i \;:=\; p_i^{\widetilde{x}} - \widetilde{x} \, \langle p_i^{\widetilde{x}}, \widetilde{x} \rangle \;\in\; \mathbb{R}^{mk}, \;\; i = 1, \ldots, N,$$

standard PCA is performed.

Alternatively (cf. also Dryden and Mardia (1998)), using instead the residual pseudo-metric, the so-called *full Procrustes mean* will be obtained. For two-dimensional data a pre-shape of the full Procrustes mean can be determined directly. It is any unit-length eigenvector to the largest eigenvalue of a suitable complex data matrix that is introduced by (26) in Section 6.1. The other eigenvectors then qualify as principal components to the residual data.

Instead of a Procrustes mean following Le (2001), an intrinsic mean may be computed (cf. Section 5.4). Then, standard PCA can be performed to the data mapped under the inverse Riemannian exponential to the tangent space at the intrinsic mean (cf. e.g. Fletcher et al. (2004)).

4 Planar Shape Spaces

Suppose that we have k planar landmarks $\begin{pmatrix} x_j \\ y_j \end{pmatrix} \in \mathbb{R}^2$, $j = 1, \ldots, k$. Introducing complex notation define

$$z_j = x_j + i y_j \in \mathbb{C}, \; j = 1, \ldots, k.$$

The landmark wise action (4) of

$$SO(2) \;=\; \left\{ \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} : \phi \in \mathbb{R} \right\}$$

on $\mathbb{R}^{2k}$ is then equivalent to the scalar action of $S^1 = \{e^{i\phi} : \phi \in \mathbb{R}\}$ on $\mathbb{C}^k$. With the Euclidean norm

$$\|z\|^2 := |z_1|^2 + \ldots + |z_k|^2$$

for $z = (z_1, \ldots, z_k) \in \mathbb{C}^k$ we have that the pre-shape sphere of k labelled two-dimensional landmarks is given by

$$S^{2(k-1)-1} := \{z = (z_1, \ldots, z_{k-1}) \in \mathbb{C}^{k-1} : \|z\|^2 = 1\}.$$

The shape space is then the well known complex projective space of real dimension $2(k-2)$ given by the *Hopf-fibration*:

$$\Sigma_2^k \cong S^{2(k-1)-1}/S^1 = P\mathbb{C}^{k-2}.$$

From now on we will use complex notation exclusively and $SO(2)$ will be identified with S^1 in the above obvious way. Then the Euclidean inner product of $z = (z_1, \ldots, z_{k-1})$ with $w = (w_1, \ldots, w_{k-1})$, where $z_j = x_j + iy_j$, $w_j = u_j + iv_j$, rewrites to

$$\langle z, w \rangle = \sum_{j=1}^{n} (x_j u_j + y_j v_j) = \sum_{j=1}^{n} \Re(z_j \overline{w}_j) = \Re(zw^*)$$

where $w^* = (\overline{w}_1, \ldots, \overline{w}_n)^T$, as usual. Note that

$$\langle z, iw \rangle = -\langle iz, w \rangle. \tag{9}$$

4.1 Horizontal and Vertical Great Circles

At every point $z \in S_2^k \subset \mathbb{C}^{k-1}$ of the pre-shape sphere, the tangent space decomposes into one so called *vertical* direction iz - that direction points into the fiber $[z]$ - and into $2(k-2)$ linearly independent so called *horizontal* directions, those are orthogonal to the fiber. In fact there is one vertical geodesic at z, all other geodesics starting off horizontally remain horizontal throughout their course (this latter fact is true for any shape space cf. Kendall et al. (1999), p. 109). More precisely we have as an immediate consequence of (9):

Proposition 4.1. *Let $z \in S_2^k \subset \mathbb{C}^{k-1}$. Then*

(i) $\langle z, iz \rangle = 0$ *and* $[z] = \{z \cos t + iz \sin t : t \in \mathbb{R}\}$,

(ii) *for any $v \in S_2^k$ with $\langle z, v \rangle = 0$ we have for all $t \in \mathbb{R}$ that*

$$\langle \dot{\gamma}_{z,v}(t), \ i\,\gamma_{z,v}(t) \rangle = \langle \dot{\gamma}_{z,v}(0), \ i\,\gamma_{z,v}(0) \rangle = \langle v, iz \rangle.$$

A geodesic $\gamma_{z,v}$ with $z, v \in S_2^k$, $\langle z, v \rangle = 0$ is hence *horizontal* if and only if the *horizontality condition*

$$\langle v, iz \rangle = 0 \tag{10}$$

is satisfied.

Remark 4.2. *Every horizontal geodesic $\gamma_{z,v}$ meets twice the vertical geodesic $\gamma_{z,iz}$, namely at z and $-z$.*

4.2 Optimal Positioning

In order to position optimally we have to minimize the distance along a fiber. The first assertion of the following theorem - a consequence of e.g. (8) and Proposition 4.1 (i) - we take from Ziezold (1977):

Theorem 4.3. *Let $x, p, v \in S$ with $\langle x, v \rangle = 0$. Then*

(i) x and $e^{it}p$ are in optimal position if and only if

$$
e^{it} = \begin{cases} \frac{\langle x,p \rangle + i \langle x, ip \rangle}{\sqrt{\langle x,p \rangle^2 + \langle x, ip \rangle^2}}, & if \quad \langle x,p \rangle^2 + \langle x, ip \rangle^2 \neq 0 \\ arbitrary, & if \quad \langle x,p \rangle^2 + \langle x, ip \rangle^2 = 0. \end{cases}
$$

(ii) $\gamma_{x,v}$ and $e^{it}p$ are in optimal position if and only if

$$
e^{it} = \begin{cases} \pm \frac{D + iC}{\sqrt{D^2 + C^2}}, & if \quad C \neq 0 \\ \pm 1, & if \quad C = 0, A^2 \neq B^2 \\ arbitrary, & if \quad C = 0, A^2 = B^2 \end{cases}
$$

with

$$
A^2 = \langle x,p \rangle^2 + \langle v, p \rangle^2, \qquad B^2 = \langle x, ip \rangle^2 + \langle v, ip \rangle^2,
$$
$$
D = 2\Big(\langle x, p \rangle \langle x, ip \rangle + \langle v, p \rangle \langle v, ip \rangle \Big), \qquad C = B^2 - A^2 + \sqrt{(B^2 - A^2)^2 + D^2}.
$$

Proof. To prove the second assertion note that minimizing (7) is equivalent to maximizing

$$
\cos^2\big(d(e^{it}p, \gamma_{x,v})\big) = c^2 A^2 + s^2 B^2 + csD \tag{11}
$$

with $c = \cos t, s = \sin t$ and A, B, D defined as above. This is a function of t with period π, it suffices thus to consider $-\frac{\pi}{2} < t \leq \frac{\pi}{2}$ only.
In case of $D \neq 0 = A^2 - B^2$ (i.e. $C \neq 0$) we have the solution

$$
t = \text{sign}(D)\,\frac{\pi}{4}.
$$

If $D = 0 = A^2 - B^2$ then (11) is constant along the path $t \to e^{it}p$. In case $D = 0 \neq A^2 - B^2$ we have the solutions

$$
t = \frac{\pi}{2} \quad \text{if} \quad B^2 > A^2, \quad \text{then } C \neq 0\,,
$$
$$
t = 0 \quad \text{if} \quad A^2 > B^2, \quad \text{then } C = 0\,.
$$

Now assume that $D \neq 0 \neq A^2 - B^2$. Computing the first and second derivative of (11) with respect to t observe that we have a maximum if and only if

$$
\left.\begin{array}{rcl} \tan(2t) = \frac{2cs}{c^2 - s^2} & = & \frac{D}{A^2 - B^2}, \\ -cs \frac{(A^2 - B^2)^2}{D} & < & csD. \end{array}\right\} \tag{12}
$$

From the above inequality we infer that $\text{sign}(\sin 2t) = \text{sign}(D)$, thus from the above equation: $\text{sign}(\cos 2t) = \text{sign}(A^2 - B^2)$. Solving the above equation we find that precisely taking the positive square root below meets the two conditions of (12)

$$ t \;=\; \arctan\left(\frac{1}{D}\Big(B^2 - A^2 + \sqrt{(B^2 - A^2)^2 + D^2}\Big)\right). $$

This yields the assertion. $\qquad\square$

4.3   The Space of Geodesics of Planar Shape Space

At this point the space of geodesics $G(\Sigma_m^k)$ for $m = 1, 2$ can be easily determined by an application of the concept of horizontal and vertical geodesics on the pre-shape sphere. Using standard arguments from differential geometry introduce

$O_2\big(2(k-1)\big)$ $:= \{(e_1, e_2) \in \mathbb{C}^{k-1} \times \mathbb{C}^{k-1} : \langle e_l, e_j \rangle = \delta_{ij}, 1 \le l, j \le 2\}$, a
  Stiefel manifold of real dimension $4k - 7$.

and

$\widetilde{O}_2\big(2(k-1)\big)$ $:= \{(e_1, e_2) \in O_2\big(2(k-1)\big) : \langle e_2, ie_1 \rangle = 0\}$ a sub-manifold
  of real dimension $4k - 8$.

Then every pair $(z, v) \in O_2\big(2(k-1)\big)$ defines a geodesic $\gamma_{z,v}$ on S_2^k. In particular, if $(z, v) \in \widetilde{O}_2\big(2(k-1)\big)$ then by (10) the geodesic is horizontal. In order to single out the equivalence class corresponding to the same geodesic consider the free action of $O(2)$ from the right on both $O_2(2k)$ and $\widetilde{O}_2(2k)$ defined by

$$ (e_1, e_2)\begin{pmatrix} a & -b \\ \varepsilon b & \varepsilon a \end{pmatrix} \;=\; (ae_1 + \varepsilon b e_2, -be_1 + \varepsilon a e_2), \qquad (13) $$

where $a^2 + b^2 = 1 = \varepsilon^2$. The quotient $O_2\big(2(k-1)\big)/O(2) =: G_2\big(2(k-1)\big)$ is a *Grassmanian manifold* of real dimension $4k - 8$ which gives

$$ G(S_2^k) \;\cong\; G_2\big(2(k-1)\big). $$

With a similar consideration we obtain the space of geodesics of linear shapes

$$ G(\Sigma_1^k) \;\cong\; G_2(k-1) $$

of dimension $2k - 6$. For planar shape spaces define the sub-manifold

$$ \widetilde{G}_2\big(2(k-1)\big) := \widetilde{O}_2\big(2(k-1)\big)/O(2) \;\subset\; G_2\big(2(k-1)\big), $$

of dimension $4k - 9$. Underlying the above (and also the below) reasoning is the following theorem (cf. Abraham and Marsden (1978), p. 266):

Theorem 4.4. *A smooth and free action of a compact Lie group G on a smooth manifold M induces a natural manifold structure on M/G compatible with its quotient topology.*

We now investigate the corresponding action of $S^1 \cong SO(2)$ on both $O_2\big(2(k-1)\big)$ and $\widetilde{O}_2\big(2(k-1)\big)$. It is, in fact, a free action from the left by components:

$$e^{i\phi}(e_1, e_2) \quad = \quad (e^{i\phi}e_1, e^{i\phi}e_2), \tag{14}$$

that commutes with the action of $O(2)$. Here $e^{i\phi}e_j$ $(j = 1, 2)$ is defined as in (4). Under the Hopf fibration, horizontal geodesics on S_2^k project to geodesics on Σ_2^k and geodesics on Σ_2^k lift to horizontal geodesics on S_2^k. This is a general fact for Riemannian submersions, cf. O'Neill (1983), p. 212. Hence, every geodesic δ on Σ_2^k starting at $[z]$ can be lifted to a unique horizontal great circle $\gamma_{z,v}$ on S_2^k starting at z. Then, for any $e^{i\phi} \in S^1$, we have that the great circle $\gamma_{e^{i\phi}z,w}$ on S_2^k through $e^{i\phi}z$ projects to the same geodesic δ if and only if

$$w \quad = \quad e^{i\phi}v,$$

cf. (5). Thus

$$G(\Sigma_2^k) \quad \cong \quad \widetilde{G}_2\big(2(k-1)\big)/S^1$$

as topological spaces. Unfortunately we cannot straightforward conclude that the right hand side is a manifold since the action of S^1 is no longer free on either $G_2\big(2(k-1)\big)$ or $\widetilde{G}_2\big(2(k-1)\big)$ as Remark 4.2 teaches. However, $S^0 = \{id, -id\} \subset S^1 \cap O(2)$ acts freely on $\widetilde{O}_2\big(2(k-1)\big)$ (even discretely) giving rise to a manifold

$$\widetilde{O}_2^0\big(2(k-1)\big) \quad := \quad \widetilde{O}_2\big(2(k-1)\big)/S^0$$

of the same dimension $4k - 8$. Both S^1 and $O(2)$ act freely on $\widetilde{O}_2^0\big(2(k-1)\big)$ and their actions commute. Moreover, S^1 also acts freely on

$$\widetilde{G}_2^0\big(2(k-1)\big) \quad := \quad \widetilde{O}_2^0\big(2(k-1)\big)/O(2)$$

which is, as a topological quotient the same as $\widetilde{G}_2\big(2(k-1)\big)$. With this we obtain

Theorem 4.5. *The space of all geodesics on planar shape space can be given a manifold structure*

$$G(\Sigma_2^k) \quad \cong \quad \widetilde{G}_2\big(2(k-1)\big)/S^1$$

of dimension $4k - 10$.

5 Algorithms for PCGs and IMs on Planar Shape Spaces

For this section assume that N planar configurations are mapped to pre-shapes $p_1, \ldots, p_N \in S_2^k \subset \mathbb{C}^{k-1}$ and to shapes $[p_1], \ldots, [p_N] \in \Sigma_2^k$. In order to find PCGs $\gamma^* \in G(\Sigma_2^k)$ an objective function F given by (3) will be minimized under suitable vector valued constraining conditions $\Phi = 0$. A standard method to

solve such a minimization problem is to project a gradient descent path onto the constraint surface:

$$x^{(n)} \mapsto x^{(n+1)} := \pi\big(x^{(n)} - \varepsilon^{(n)} dF(x^{(n)})\big).$$

Here, $\varepsilon^{(n)} > 0$ is sufficiently small and π gives the orthogonal projection to the surface S determined by $\Phi = 0$.

We propose another method here. Introducing a vector valued Lagrange multiplier λ, every minimum will then solve

$$dF + \lambda^T d\Phi \quad = \quad 0. \tag{15}$$

This equation will yield in a natural way fixed point equations for the parameters of minimizing geodesics. These in turn will lead to numerical algorithms solving the constrained optimization problem. In practice we have a rather quick convergence to a (possibly local) minimum when applying the below derived algorithms to numerical data in section 6. Analogously, minimizing (2) will lead to a fixed point equation for the IM.

Let us first discuss convergence issues arising from the actions of $O(2)$ and S^1. The ambiguity due to the right action of $O(2)$ on the space of spherical geodesics given by (13) can be overcome by adding appropriate constraining conditions (cf. Huckemann and Ziezold (2006)). The ambiguity stemming from the left action of S^1 on the horizontal geodesics (14) is handled similarly below. We now give a heuristic reasoning, why the algorithms of our approach usually converge faster to a local minimum rather than a method based on projecting a gradient descent path. In view of the subtle argument of Le (2001) (concerning the convergence of an algorithm for the IM which we quote at the end of Section 5.4) a mathematically sound derivation of conditions for convergence is obviously beyond the scope of this paper. For the specific minimization problems below the constraining surfaces determined by $\Phi = 0$ are subsets of a direct product of spheres. If, to simplify the argument, we had a single two-sphere determined by $\Phi(x) = \langle x, x \rangle - 1$, $x \in \mathbb{R}^2$, then the fixed point equation $dF(x) = -2\lambda x$ naturally yields the algorithm $x^{(n+1)} = -dF(x^{(n)})/\|dF(x^{(n)})\|$. Obviously, the less curved the level lines of F are, the faster is the convergence. In case of high curvature, however, convergence might not occur. In Section 6 below, for all data examples and for essentially all of our simulations we observed convergence.

In the following, for values $0 < \zeta_j < 1$, $j = 1, \ldots, N$ the formula

$$\xi_j \quad := \quad -\frac{1}{2\zeta_j} \frac{d}{d\zeta_j} \arccos^2 \zeta_j \quad = \quad \frac{\arccos \zeta_j}{\zeta_j \sqrt{1 - \zeta_j^2}} \tag{16}$$

and $\xi_j := 1$ for $\zeta_j = 1$ will be useful.

5.1 The First PCG

By Theorem 4.5, minimizing (3) over $\gamma \in G(\Sigma_2^k)$ is equivalent to finding a horizontal geodesic $\gamma_{x,v}$ minimizing the sum of squared distances to the fibers of the pre-shape data. Letting $t := (t_1, \ldots, t_N) \in [0, 2\pi)^N$ and taking (7) into account we have the objective function

$$
\begin{aligned}
F(x, v, t) &:= \sum_{j=1}^{N} d\big(\gamma_{x,v}, \ e^{it_j} p_j\big)^2 \\
&= \sum_{j=1}^{N} \arccos^2 \sqrt{\langle x, e^{it_j} p_j \rangle^2 + \langle v, e^{it_j} p_j \rangle^2}
\end{aligned}
$$

and the constraining condition $x, v \in \widetilde{O}_2\big(2(k-1)\big)$, i.e.

$$
\Phi(x, v, t) := \begin{pmatrix} \langle x, x \rangle - 1 \\ \langle v, v \rangle - 1 \\ 2\langle x, v \rangle \\ 2\langle ix, v \rangle \end{pmatrix} \overset{!}{=} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.
$$

Letting $q_j = e^{it_j} p_j$, $x, v \in \mathbb{C}^{k-1}$ and $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ the Lagrange equation (15) rewrites to

$$
\left.\begin{array}{rcl}
\sum_{j=1}^{N} \xi_j \, \langle x, q_j \rangle \, q_j &=& \lambda_1 x + \lambda_3 v - \lambda_4 iv \\
\sum_{j=1}^{N} \xi_j \, \langle v, q_j \rangle \, q_j &=& \lambda_2 v + \lambda_3 x + \lambda_4 ix \\
\frac{\partial}{\partial t_j} \, d(\gamma_{x,v}, e^{it_j} p_j) &=& 0 \quad (j = 1, \ldots N)
\end{array}\right\}
\tag{17}
$$

where $\zeta_j = \zeta_j(x, v) := \sqrt{\langle x, q_j \rangle^2 + \langle v, q_j \rangle^2}$ and $\xi_j = \xi_j(x, v)$ from (16). With Theorem 4.3 we can directly solve the last line by putting each p_j $(j = 1, \ldots, N)$ into optimal position $q_j = q_j(x, v) = e^{it_j(x,v)} p_j$ to a competing $\gamma_{x,v}$. We may assume that the optimal positioned data q_j is sufficiently close to $\gamma_{x,v}$ which means that $\zeta_j \neq 0$, $j = 1, \ldots, N$. The Lagrange multipliers can be obtained from (17):

$$
\left.\begin{array}{rcccl}
\sum_{j=1}^{N} \xi_j \, \langle x, q_j \rangle^2 &=& \lambda_1(x, v) &=& \lambda_1 \\
\sum_{j=1}^{N} \xi_j \, \langle v, q_j \rangle^2 &=& \lambda_2(x, v) &=& \lambda_2 \\
\sum_{j=1}^{N} \xi_j \, \langle x, q_j \rangle \langle v, q_j \rangle &=& \lambda_3(x, v) &=& \lambda_3 \\
\sum_{i=1}^{N} \xi_j \, \langle ix, q_j \rangle \langle v, q_j \rangle &=& \lambda_4(x, v) &=& \lambda_4
\end{array}\right\}.
\tag{18}
$$

Introducing

$$
\begin{aligned}
G_a = G_a(x, v) &:= \sum_{j=1}^{N} \xi_j \, \langle a, q_j \rangle \, q_j, \\
s = s(x, v) &:= \big(\lambda_1 \lambda_2 - \lambda_3^2 - \lambda_4^2\big)^{-1}, \\
\Psi_1(x, v) &:= s \, \big(\lambda_2 G_x - \lambda_3 G_v + \lambda_4 \, iG_v\big), \\
\Psi_2(x, v) &:= s \, \big(\lambda_1 G_v - \lambda_3 G_x - \lambda_4 \, iG_x\big),
\end{aligned}
$$

13

then x^*, v^* solve (17) and (18) if and only if they satisfy the fixed point equations $x^* = \Psi_1(x^*, v^*)$ and $v^* = \Psi_2(x^*, v^*)$. Note that, the linear combinations of the two gradients and the multiplication by s can be viewed as an approximation to the projection to the surface determined by $\Phi = 0$. At (x^*, v^*) this approximation is exact. Hence, we propose the following algorithm to determine (x^*, v^*):

Starting with initial values, e.g.

$$
\begin{aligned}
x^{(0)} \ := \ p_1, \quad v^{(0)} \ := \ & \text{unit horizontal projection of } (p_2 - p_1) \\
= \ & \frac{p_2 - p_1 - \langle p_2 - p_1, p_1\rangle p_1 - \langle p_2 - p_1, ip_1\rangle ip_1}{\|p_2 - p_1 - \langle p_2 - p_1, p_1\rangle p_1 - \langle p_2 - p_1, ip_1\rangle ip_1\|}
\end{aligned}
$$

obtain

$$
x^{(n+1)}, \quad v^{(n+1)} \quad \text{from} \quad x^{(n)}, \quad v^{(n)} \quad \text{for } n \geq 0
$$

by computing $q_j := e^{it_j} p_j \in [p_j], 1 \leq j \leq N$ in optimal position with respect to $\gamma_{x^{(n)}, v^{(n)}}$ according to Theorem 4.3 and by setting

$$
\begin{aligned}
x^{(n+1)} \ = \ & \frac{\Psi_1(x^{(n)}, v^{(n)})}{\|\Psi_1(x^{(n)}, v^{(n)})\|}, \\
v^{(n+1)} \ = \ & \text{unit horizontal projection:} \\
= \ & \frac{\Psi_2(x^{(n)}, v^{(n)}) - \langle \Psi_2(x^{(n)}, v^{(n)}), x^{(n+1)}\rangle \, x^{(n+1)} - \langle \Psi_2(x^{(n)}, v^{(n)}), ix^{(n+1)}\rangle \, ix^{(n+1)}}{\|\Psi_2(x^{(n)}, v^{(n)}) - \langle \Psi_2(x^{(n)}, v^{(n)}), x^{(n+1)}\rangle \, x^{(n+1)} - \langle \Psi_2(x^{(n)}, v^{(n)}), ix^{(n+1)}\rangle \, ix^{(n+1)}\|} \, .
\end{aligned}
$$

Note that, additionally to the ambiguity due to the action of $O(2)$ there is also an ambiguity due to the action of S^1. This can be overcome by bringing $x^{(n+1)}$ into optimal position $e^{it} x^{(n+1)}$ to $x^{(n)}$ in every iteration step. Additionally then, according to (5), the tangent vector $v^{(n+1)}$ at $x^{(n+1)}$ has to be rotated by the same amount to the tangent vector $e^{it} v^{(n+1)}$ at $e^{it} x^{(n+1)}$ so that $(x^{(n+1)}, v^{(n+1)})$ and $(e^{it} x^{(n+1)}, e^{it} v^{(n+1)})$ determine the same geodesic in shape space.

5.2 The Second PCG

Having found a horizontal great circle $\gamma_1 = \gamma_{x,v}$ determined by $x, v \in \widetilde{O}_2\big(2(k-1)\big)$ projecting to a first principal component geodesic on Σ_2^k suppose that $\gamma_2(t) = \gamma_{y,w}(t) = y \cos t + w \sin t$ with $y = y(\tau) = x \cos \tau + v \sin \tau$ for some suitable $\tau \in \mathbb{R}$ projects to a second principal component geodesic. Let us denote the velocity of γ_1 at y by

$$
z \ = \ z(\tau) \ = v \cos \tau - x \sin \tau \, .
$$

The objective function to be minimized is now

$$
\begin{aligned}
F(\tau, w) \ := \ & \sum_{j=1}^{N} d(q_j, \gamma_{y(\tau), w})^2 \\
= \ & \sum_{j=1}^{N} \arccos^2 \sqrt{\langle y, q_j\rangle^2 + \langle w, q_j\rangle^2}
\end{aligned}
$$

14

where, of course, $q_j = q_j(\tau, w) \in [p_j]$ is in optimal position to $\gamma_{y(\tau),w}$. Obviously $\langle y, w \rangle = 0 = \langle z, w \rangle \Leftrightarrow \langle x, w \rangle = 0 = \langle v, w \rangle$, hence, the constraining function can now be taken as

$$\Phi(\tau, w) \;:=\; \begin{pmatrix} 2\langle x, w \rangle \\ 2\langle v, w \rangle \\ \langle w, w \rangle - 1 \\ 2\langle iy, w \rangle \end{pmatrix} \;\overset{!}{=}\; \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Introducing a Lagrange multiplier $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ the Lagrange equation (15) rewrites to

$$\left. \begin{array}{rcl} \sum_{j=1}^{N} \xi_j \, \langle w, q_j \rangle \, q_j & = & \lambda_1 x + \lambda_2 v + \lambda_3 w + \lambda_4 iy \\ \sum_{j=1}^{N} \xi_j \, \langle y, q_j \rangle \, \langle z, q_j \rangle & = & \lambda_4 \langle iz, w \rangle \end{array} \right\} \qquad (19)$$

where now $\zeta_j := \sqrt{\langle y, q_j \rangle^2 + \langle w, q_j \rangle^2}$ and ξ_j from (16). Introducing

$$G_{a,b} \;:=\; \sum_{j=1}^{N} \xi_j \, \langle a, q_j \rangle \, \langle b, q_j \rangle, \qquad c \;:=\; \cos\tau, \qquad s \;:=\; \sin\tau$$

we have for the Lagrange multipliers

$$\left. \begin{array}{rcl} G_{w,x} & = & \lambda_1 \\ G_{w,v} & = & \lambda_2 \\ G_{w,w} - \lambda_4 c\langle w, ix \rangle - \lambda_4 s\langle w, iv \rangle & = & \lambda_3 \\ G_{w,ix} - \lambda_3 \langle w, ix \rangle & = & c\lambda_4 \\ G_{w,iv} - \lambda_3 \langle w, iv \rangle & = & s\lambda_4 \end{array} \right\}. \qquad (20)$$

With this, the last line of (19) rewrites to

$$\begin{array}{rcl} (c^2 - s^2)\, G_{x,v} + cs\, (G_{v,v} - G_{x,x}) & = & c\lambda_4 \langle iv, w \rangle - s\lambda_4 \langle ix, w \rangle \\ & = & G_{w,ix}\langle iv, w \rangle - G_{w,iv}\langle ix, w \rangle. \quad (21) \end{array}$$

From (21) and the first line of (19) two fixed point equations for τ^* and w^* solving (19) and (20) can be derived. It is, however, very convenient to alter the algorithm such that in every step, τ is set to zero, which means that x and v are updated in every step to y and z. Then, in particular $\langle ix, w \rangle = 0$, and, linearizing (21), set

$$\Psi_1(x, v, w) \;:=\; \sum_{i=1}^{N} \xi_i \, \langle w, q_i \rangle \, q_i - G_{w,x}\, x - G_{w,v}\, v - G_{w,ix}\, ix,$$

$$\Psi_2(x, v, w) \;:=\; \frac{G_{w,ix}\langle iv, w \rangle - G_{x,v}}{G_{v,v} - G_{x,x}},$$

to obtain the following algorithm:

Starting with initial values, e.g.

$$x^{(0)} \;:=\; x, \quad v^{(0)} \;:=\; v, \quad w^{(0)} \;:=\; iv,$$

obtain

$$x^{(n+1)}, \quad v^{(n+1)}, w^{(n+1)} \quad \text{from} \quad x^{(n)}, \quad v^{(n)}, \quad w^{(n)} \quad \text{for } n \geq 0$$

by computing $q_j := e^{it_j} p_j \in [p_j], 1 \leq j \leq N$ in optimal position with respect to $\gamma_{x^{(n)}, w^{(n)}}$ according to Theorem 4.3 and by setting

$$
\begin{aligned}
\tau &:= \Psi_2(x^{(n)}, v^{(n)}, w^{(n)}) \\
x^{(n+1)} &:= x^{(n)} \cos \tau + v^{(n)} \sin \tau \\
v^{(n+1)} &:= v^{(n)} \cos \tau - x^{(n)} \sin \tau \\
w^{(n+1)} &:= \frac{\Psi_1(x^{(n+1)}, v^{(n+1)}, w^{(n+1)})}{\|\Psi_1(x^{(n+1)}, v^{(n+1)}, w^{(n+1)})\|}
\end{aligned}
$$

Having thus found x^*, v^*, w^* note that $\hat{x} := x^*$ is a representative of a PM (principal component geodesic mean) on Σ_2^k. With $v_1 := v^*$ and $v_2 := w^*$ we have the two horizontal geodesics

$$\gamma_1 := \gamma_{\hat{x}, v_1}, \quad \gamma_2 := \gamma_{\hat{x}, v_2}$$

mapping to a first and a second PCG on Σ_2^k. For simplicity set

$$x := \hat{x}.$$

5.3 Higher Order PCGs

Suppose that we have found horizontal great circles $\gamma_{x, v_1}, \ldots, \gamma_{x, v_{r-1}}$, $3 \leq r \leq 2k - 4$ mapping to PCGs on Σ_2^k. For any r-th order PCG on the shape space there is a horizontal geodesic through x determined by a single horizontal direction v that is orthogonal to all preceeding directions at x. Introducing $\zeta_j := \sqrt{\langle x, p_j \rangle^2 + \langle v, p_j \rangle^2}$ and ξ_j according to (16) and Lagrange multipliers $\lambda_0, \ldots, \lambda_r, \lambda_{r+1} \in \mathbb{R}$ the corresponding Lagrange equation (15) is given by

$$\sum_{j=1}^{N} \xi_j \langle v, q_j \rangle q_j = \lambda_0 x + \sum_{s=1}^{r-1} \lambda_s v_s + \lambda_r v + \lambda_{r+1} i x. \tag{22}$$

Starting with a suitable $v^{(0)}$ we thus compute $v^{(n+1)}$ from $v^{(n)}$ by the following algorithm which follows in a natural way from (22):

$$
\left.
\begin{aligned}
q_j &\in [p_j] \text{ in optimal position to } \gamma_{x, v^{(n)}} \\
z^{(n+1)} &:= \sum_{j=1}^{N} \xi_j^{(n)} \langle v^{(n)}, q_j \rangle q_j, \\
\lambda_0 &:= \langle z^{(n+1)}, x \rangle, \\
\lambda_s &:= \langle z^{(n+1)}, v_s \rangle, \quad 1 \leq s < r, \\
\lambda_r &:= \sum_{j=1}^{N} \xi_j^{(n)} \langle v^{(n)}, q_j \rangle^2, \\
\lambda_{r+1} &:= \langle z^{(n+1)}, ix \rangle, \\
v^{(n+1)} &:= \operatorname{sign}(\lambda_r) \frac{z^{(n+1)} - \lambda_0 x - \sum_{s=1}^{r-1} \lambda_s v_s - \lambda_{r+1} ix}{\|z^{(n+1)} - \lambda_0 x - \sum_{s=1}^{r-1} \lambda_s v_s - \lambda_{r+1} ix\|}
\end{aligned}
\right\}.
$$

16

5.4 The Intrinsic Mean

Here we determine a pre-shape in the fiber minimizing (2). Hence we minimize

$$F(x) := \min_{t \in [0,\pi)^N} \sum_{j=1}^{N} d(e^{it_j} p_j, x)^2$$

over a single variable $x \in \mathbb{C}^{k-1}$ under the constraining condition $\|x\|^2 = 1$. The shape of a minimizing pre-shape is then a minimizing shape. Arguing as in Section 5.1, rotate p_j by Theorem 4.3 into optimal position $q_j \in [p_j]$ to x, and, set $\zeta_j := \langle x, q_j \rangle$ and $\xi_j = \frac{\arccos \zeta_j}{\sqrt{1-\zeta_j^2}}$ analogously to (16). Then, use (6) to obtain with a single Lagrange multiplier $\lambda \in \mathbb{R}$ the equations

$$\sum_{j=1}^{N} \xi_j q_j = \lambda x, \quad \sum_{j=1}^{N} \xi_j \langle q_j, x \rangle = \lambda.$$

Thus with $\Psi(x) := \text{sign}(\lambda) \sum_{j=1}^{N} \xi_j q_j$ we have the following algorithm for the intrinsic mean

$$\left. \begin{array}{rcl} x^{(n)} & \mapsto & x^{(n+1)} \\ x^{(n+1)} & = & \frac{\Psi(x^{(n)})}{\|\Psi(x^{(n)})\|} \end{array} \right\}. \tag{23}$$

For every iteration all $q_j \in [p_j]$ are rotated into optimal position to $x^{(n)}$.

We note that Le (2001) derived a general algorithm for the computation of intrinsic means based on the Riemannian exponential map for a gradient descent minimization of (2), building on the work of Karcher (1977). Setting $\widetilde{\Psi}(x) := \sum_{j=1}^{N} \xi_j q_j - \lambda x$ her algorithm for the intrinsic mean is the following:

$$y^{(n)} \mapsto y^{(n+1)} = y^{(n)} \cos \|\widetilde{\Psi}(y^{(n)})\| + \frac{\widetilde{\Psi}(y^{(n)})}{\|\widetilde{\Psi}(y^{(n)})\|} \sin \|\widetilde{\Psi}(y^{(n)})\|.$$

Of course as above, for every iteration all $q_j \in [p_j]$ are rotated into optimal position to $y^{(n)}$. In particular, Le (2001, Theorem 4) proved convergence in $d_{\Sigma_2^k}$ to the pre-shape of an intrinsic mean $\overline{p}$ if the data is contained in a geodesic ball on Σ_2^k of radius $\frac{3\pi}{40}$ and center $y^{(0)}$.

6 Numerical Application

In the preceeding sections we have provided for a theoretical concept of PCA based on geodesics and for methods to compute PC geodesics for planar shape spaces. Now, PCA based on Euclidean approximations and PCA based on

geodesics will be compared with one another.

By definition, PCs obtained by Euclidean approximation are straight lines in the tangent space of the pre-shape sphere at a pre-shape of the EM (extrinsic mean). These straight lines project (by the inverse projection to the tangent space) to great circles through the pre-shape of the EM, in fact they will be almost horizontal geodesics. Both methods, PCA based on Euclidean approximation and PCA based on geodesics minimize sums of squared distances. However, the subtle difference lies in the two following facts: First, the former method minimizes only over geodesics passing through an EM. In contrast the latter method does not preassign a mean, rather, when finding the second PC geodesic the PM (principal component geodesic mean) is determined. Recall that, different from both EM and PM is yet the IM (intrinsic mean). Secondly, the former method minimizes over Euclidean projections of intrinsic distances, the latter method minimizes over intrinsic distances itself. It is this intrinsic metric, in which the comparison is performed. Hence, the first PC geodesic will never fit the data worse than the projection of the first PC obtained by Euclidean approximation. In the following the improvement of fit will be quantified.

To this end we determine for a given dataset of pre-shapes $p_1, \ldots, p_N \in S_2^k$ the first PC based on Euclidean approximation and denote the corresponding great circle by γ_e. We also determine a horizontal great circle γ_g that corresponds to the first PC geodesic. The absolute improvement of the fit is given by

$$\sum_{j=1}^N d_\Sigma(p_j, \gamma_e)^2 - \sum_{j=1} d_\Sigma(p_i, \gamma_g)^2 \ \geq \ 0 \,.$$

We normalize this quantity by dividing by the "true" geodesic fit. The improvement of fit will be measured in percent:

$$f \ = \ f(p_1, \ldots, p_N) \quad := \quad \left(\frac{\sum_{j=1}^N d_\Sigma(p_j, \gamma_e)^2}{\sum_{j=1}^N d_\Sigma(p_i, \gamma_g)^2} - 1 \right) 100 \,. \qquad (24)$$

Here the distance

$$d_\Sigma(p_j, \gamma_{x,v}) \ = \ \arccos^2 \sqrt{\langle x, e^{it_j} p_j \rangle^2 + \langle v, e^{it_j} p_j \rangle^2}$$

is the shape distance; the optimal e^{it_j} are given by Theorem 4.3, (ii). Also, for pre-shapes $\widetilde{p}$, $\overline{p}$ and $\hat{p}$ of the EM, IM and PM, resp., the mutual shape distances

$$\left. \begin{array}{rcl} I\_P \ = \ I\_P(p_1, \ldots, p_N) & := & d_\Sigma(\overline{p}, \hat{p}) \,, \\ E\_P \ = \ E\_P(p_1, \ldots, p_N) & := & d_\Sigma(\widetilde{p}, \hat{p}) \,, \\ I\_E \ = \ I\_E(p_1, \ldots, p_N) & := & d_\Sigma(\overline{p}, \widetilde{p}) \end{array} \right\} \qquad (25)$$

are investigated.

18

As stated in the "prevailing conviction" (1), *data concentration* is expected to be correlated with the goodness of fit obtained by Euclidean approximation. We measure data concentration as in Dryden and Mardia (1998): Consider the complex Hermitian $(k-1) \times (k-1)$ sum of squares and products data matrix

$$\sum_{j=1}^{N} p_j p_j^* \tag{26}$$

(here p_j^* denotes the conjugate transpose of the complex column vector $p_j \in S_2^k$, p_j is the pre-shape of the j-th configuration). Since $p_j \in S_2^k$, the sum of all eigenvalues is equal to N. The data is concentrated about the fiber that is determined by the eigenvector to the largest eigenvalue l_1. This fiber is the *full Procrustes mean shape* to the data. Hence,

$$\varepsilon \quad := \quad 1 - \frac{l_1}{N} \tag{27}$$

can be used as a measure for data concentration.

6.1   Euclidean vs. Geodesic PCA for Typical Datasets

In this section the above quantities are computed for three well studied datasets which are taken from Dryden and Mardia (1998); there a detailed discussion of the datasets can be found. The respective values of the above quantities f, I_P, E_P, I_E and ε are recorded in Table 1. These datasets reflect typical features in distribution and landmark assessment:

Rat skulls measures cranium growth of 18 rat specimen measured at eight specific dates in their early life. On each individual rat 8 landmarks are assigned to anatomically specific locations on a planar cranium section. The shapes of each individual rat strongly follow geodesics in shape space (cf. also Le and Kume (2000) and Huckemann and Ziezold (2006)).

Mouse vertebrae comprises of 6 landmarks placed at mathematically relevant locations (maximal extension and extremal curvature of contour) on a planar section of a single thoracic mouse vertebra. This dataset is composed of three sub-groups: 23 small mouse specimen, 23 large mouse specimen and 30 controls. The shapes of specimen in each of the first two groups accumulate significantly around the mean shape of the group.

Digits 3 is a dataset containing 30 handwritten digits "three" used for postcode recognition. 5 Landmarks have been placed at mathematically relevant locations and 8 more ("pseudo"-)landmarks have been evenly distributed within between on the contour. Compared to the previous datasets, these shapes are more dispersed. Both methods of PCA reveal that 5 PCs are required to explain approx. 90% of data variation.

Table 1: *Displaying for various datasets of respective sample size N with k landmarks the percentage of improvement of the fit by the first PC geodesic versus the fit based on the first PC obtained by Euclidean approximation denoted by f, cf. (24). Also, the mutual shape distances between the various means are shown, cf. (25). The last column records the concentration measure from (27) for the various datasets indicating concentration about the full Procrustes mean shape. Small values indicate high concentration.*

Dataset	k	N	f	I_P	E_P	I_E	ε
Rat skulls	8	144	0.000211	$6.66e-05$	$5.2e-05$	$2.19e-05$	0.0052
Mouse vertebrae	6	76	0.000267	$6.72e-05$	$6.08e-05$	$1.57e-05$	0.0051
Digits 3	13	30	0.201	0.0095	0.0081	0.00154	0.075

We see from Table 1 that for the first two datasets, the values obtained by PCA based on Euclidean approximation are very close to the values obtained by PCA based on geodesics. For the third dataset, however, the difference in values is notable. For all three datasets, similar to the observation in Hucke-mann and Ziezold (2006), EM and IM are much closer to each other than to the PM. Obviously, high concentration correlates with a high precision of the Euclidean approximation. More subtly we note, that the dataset "rats skulls" is approximated slightly better than the dataset "mouse vertebrae" even though both sets have about the same concentration about their mean. This could be explained by the difference in distribution type: the first dataset is strongly anisotropically distributed along the first PCG wheras the second dataset is more isotropically distributed. In the following investigation, the anisotropicity of the data distribution will also be included.

6.2 Euclidean vs. Geodesic PCA for Simulated Datasets

In order to broaden the discussion, we now simulate random shapes to analyse more carefully the distributions of the above introduced percentage of improvement f and the various inter-mean distances. A prominent distribution for simulation of planar shapes in the pre-shape sphere is the *complex Bingham* distribution, proposed by Kent (1994). We briefly summarize from the extensive discussion in Dryden and Mardia (1998), pp. 111: A random pre-shape Z with density

$$\pi(z) \quad \propto \quad e^{z^* A z}, \quad z \in S_2^k$$

is complex $(k-1)$-Bingham distributed if A is a complex $(k-1) \times (k-1)$ Hermitian matrix. Obviously, the density is invariant under the action of S^1, hence this distribution maps to a distribution on the shape space Σ_2^k. With $z = (x_1 + iy_1, \ldots, x_{k-1} + iy_{k-1})^T$ and $x = (x_1, y_1, \ldots, x_{k-1}, y_{k-1})^T$ we see that the complex Bingham distribution is a special case of a real $(2k-2)$-Bingham distribution:

$$\pi(x) \quad \propto \quad e^{x^T B x}, \quad x \in \mathbb{R}^{2k-2}, \|x\| = 1$$

with a suitable matrix $B \in M(2k-2, 2k-2)$. As with the complex sum of squares and products data matrix (26) in the preceeding section, the eigenspace to the largest eigenvalue λ of A, if simple, is the modal shape to the distribution. Usually instead of A the matrix $A - \lambda I$ with eigenvalues $\lambda_1 = 0 \geq \lambda_2 \geq \ldots \geq \lambda_{k-1}$ is considered which yields the same distribution. In case of *high concentration*, i.e. $\lambda_2 \ll 0$ there is a MLE (cf. Dryden and Mardia (1998), p. 116):

$$\hat{\lambda}_j \approx -\frac{N}{l_j}, \quad j = 2, \ldots, k-1, \tag{28}$$

with the eigenvalues l_j of the complex Hermitian sum of squares and products data matrix (26) of a sample of size N. Inspired by (28) and

$$l_1 = N(1 - \varepsilon), \quad l_j = N\varepsilon \, a(j) \quad \text{for } j = 2, \ldots, k-1$$

with $\sum_{j=2}^{k-1} a(j) = 1$, we let for the simulation with suitable $\alpha > 0$

$$\lambda_1 = 0, \quad \lambda_j = -\frac{1}{a(j)\alpha} \quad \text{for } j = 2, \ldots, k-1 \tag{29}$$

be the eigenvalues of A. We consider three cases ($2 \leq j \leq k-1$):

(I) $a(j) = \frac{1}{k-2}$ giving an isotropic distribution about the modal shape, the resulting Bingham distribution with only two eigenvalues is called the *Watson* distribution,

(II) $a(j) = \frac{2(k-j)}{(k-1)(k-2)}$ leading to a "mildly" anisotropic distribution with linearly decaying eigenvalues, and

(III) $a(j) = \frac{6(k-j)^2}{(k-1)(k-1)(2k-3)}$, which yields a "more strongly" anisotropic distribution (quadratic decay).

In order to be able to compare over shapes with differing numbers of landmarks we proceed as follows: For every $k \in \{3, 8, 12\}$ and isotropicity model (I) - (III) we choose α such that the empirical $1 - l_1/N \approx \varepsilon \in \{0.8, 0.1, 0.01\}$. For a small landmark number such as $k = 3$ it is not possible with the Bingham distribution to generate arbitrary low concentrations l_1/N. Hence for $k = 3$, samples having $\varepsilon \in \{0.4, 0.1, 0.01\}$ have been generated instead. Of course for $k = 3$, all three models (I) - (III) agree with one another. For every simulation 100 samples, each sample containing $N = 30$ pre-shapes on S_2^k have been created. For A a diagonal matrix with descending eigenvalues $\lambda_1 = 0 \geq \lambda_2 \geq \ldots \geq \lambda_{k-1}$ as in (29) was chosen. Then, the matrix B of the corresponding real Bingham distribution is also diagonal with non-positive, non-increasing double eigenvalues.

As the distribution for the improvement f of the SSD of the data to the first PC is non-symmetric with heavy tail, Table 2 records the median and the distance

Table 2: *Depicting in every row of each box the percentage of improvement f from (24) based on 100 samples each. The first entry gives the median over the sample, the second the distance between median and upper quartile. For every one of the three distribution models, for every concentration value ε and for every landmark number k, separate samples have been generated.*

k	ε	isotropic (I)		anisotropic linear (II)		quadratic (III)	
3	0.4	26.4	+24.2				
	0.1	0.742	+0.768				
	0.01	0.00585	+0.00915				
8	0.8	6.45	+1.71	5.36	+0.741	6.54	+0.76
	0.1	0.0135	+0.00619	0.0124	+0.00943	0.114	+0.0433
	0.01	0.000189	$+8.07e-05$	0.000191	+0.000162	0.000537	+0.000386
12	0.8	4.46	+1.16	5.03	+1.12	5.59	+1.60
	0.1	0.00661	+0.00312	0.016	+0.00615	0.121	+0.054
	0.01	0.000246	+0.000139	0.00127	+0.00179	0.00832	+0.0044

Table 3: *Displaying shape distances between the various means as defined in (25). In every box the first column denotes the median of the shape distances, the second column denotes the distance between upper quartile and median. Underlying for each box is a sample of 100 shapes corresponding to configurations with $k = 8$ landmarks of the respective distribution and concentration value ε.*

| ε		isotropic (I)		anisotropic linear (II)		quadratic (III)	
0.8	I_P	0.592	+0.0958	0.585	+0.117	0.603	+0.102
	E_P	0.576	+0.113	0.55	+0.142	0.567	+0.136
	E_I	0.0369	+0.00997	0.0374	+0.0148	0.0378	+0.0141
0.1	I_P	0.00160	+0.000426	0.00199	+0.000756	0.00283	+0.00171
	E_P	0.00151	+0.000386	0.00188	+0.000671	0.00263	+0.00152
	E_I	0.000432	$+8.07e-05$	0.000447	+0.000139	0.00065	+0.000237
0.01	I_P	$7.54e-05$	$+2.83e-05$	$8.12e-05$	$+1.78e-05$	0.000592	+0.000141
	E_P	$7.37e-05$	$+2.25e-05$	$7.85e-05$	$+1.7e-05$	0.000548	+0.000138
	E_I	$3e-05$	$+6.59e-06$	$1.64e-05$	$+3.77e-06$	0.000114	$+5.77e-05$
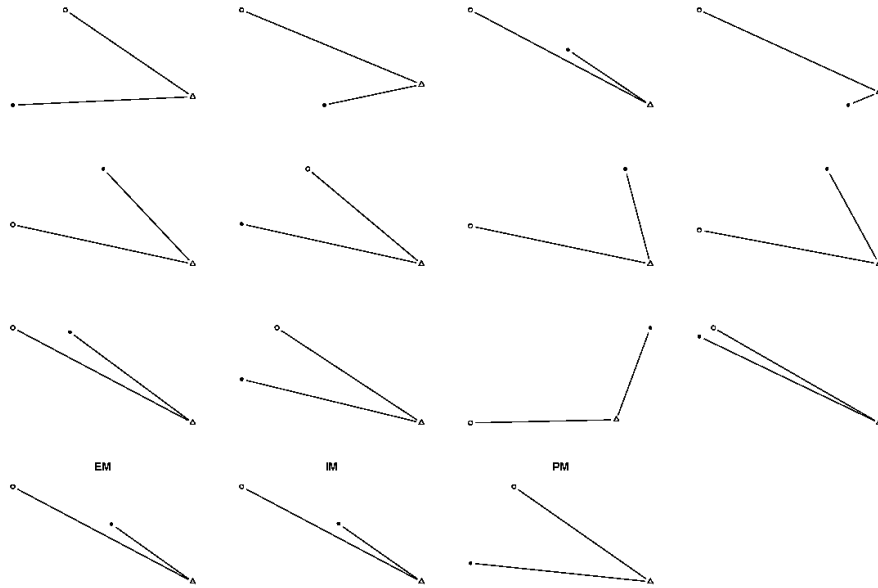
Figure 1: 12 *aligned triangles with concentration $\varepsilon = 0.33$ (top 3 rows). In the bottom row, the EM (extrinsic mean), IM (intrinsic mean) and the PM (principal component mean) are shown.*

of the median to the upper quartile of f for varying concentrations measures, distribution types and landmark numbers. For the same reason Table 3 also records for the mutual distances between the various means, the median and the distance between median and upper quartile. As the same pattern prevails rather independent of the number of landmarks, only results from simulations with with $k = 8$ landmarks are reported.

6.3   Visualizing Triangles with Low Concentration

In a last experiment, 12 triangles with low concentration $\varepsilon = 0.33$ have been generated. The 12 aligned triangles together with their various means are displayed in Figure 1. Notably EM and IM look very similar, however, the PM is different. In Figure 2 the projection of the 12 triangles to their first PC geodesic is shown. In contrast, Figure 3 depicts the projection to the first PC obtained by Euclidean approximation. Again, the difference between PCA based on geodesics and PCA based on Euclidean approximation is very clearly visible.
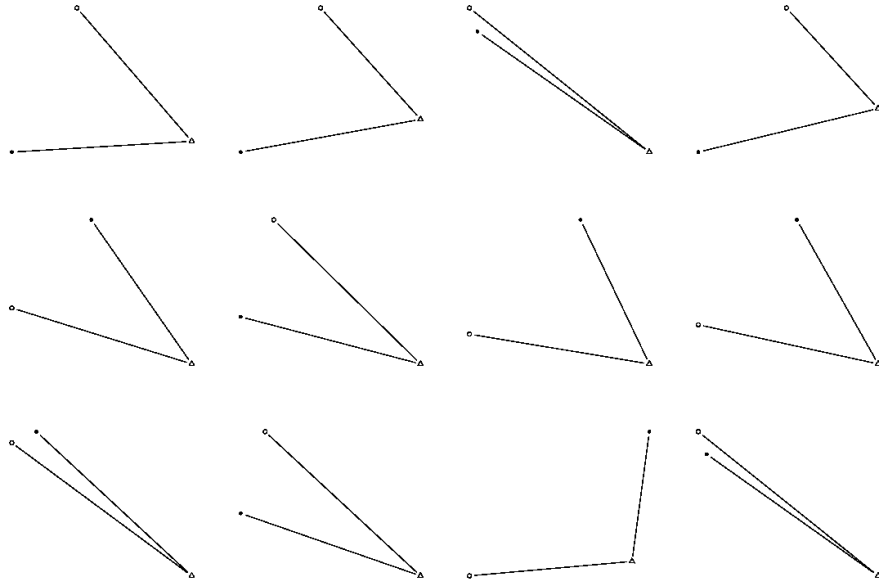
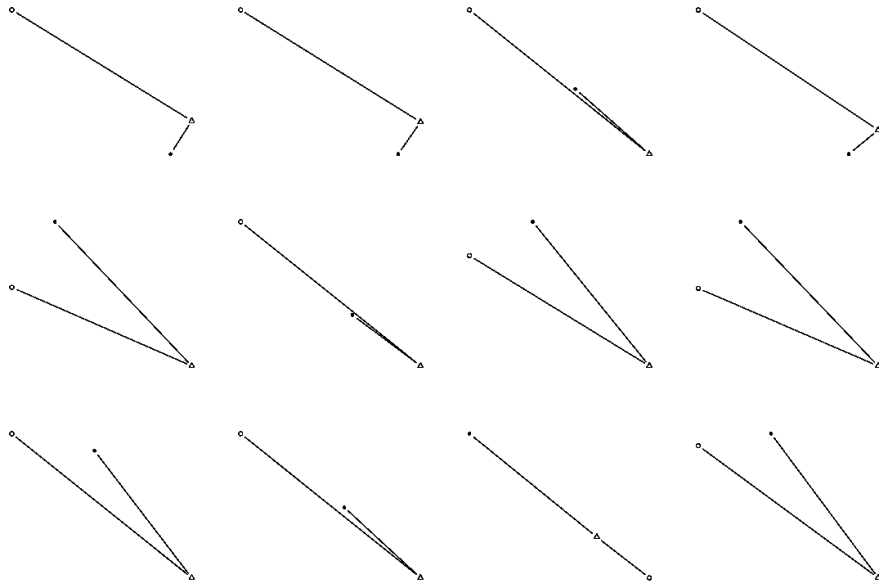Figure 2: *The projection of the above* 12 *aligned triangles to the first PC geodesic.*

Figure 3: *The projection of the above* 12 *aligned triangles to the first PC obtained by Euclidean approximation.*

6.4 Discussion

Based on the preceeding simulations a qualitative answer to the vague proposition (1) can be given. Although in Tables 2 and 3 a high volatility can be noted, several trends are clearly visible.

- With increasing concentration, i.e. decreasing ε, the first PC obtained by Euclidean approximation and also the extrinsic mean are moving closer to the values obtained by PCA based on geodesics.

- For low concentrations (ε large) the error by Euclidean approximation is notable, in particular, for $k = 3$ the error is huge. This last finding may be explained by the fact, that the triangular shape space is in a way the most curved of all planar shape spaces: it has constant Gauss curvature 4 whereas all higher dimensional planar shape spaces have Gauss curvatures ranging from 1 to 4. Also, for low concentrations, the distance between the various means is high (the maximal distance in shape space is $\pi/2$).

- Increasing anisotropicity increases the error by Euclidean approximation. This effect is very well visible for higher concentrations and less visible for low concentrations. Similarly, with increasing anisotropicity the means move apart.

- For low concentrations, increasing the number of landmarks results in a decrease of the error by Euclidian approximation.

- For high concentrations and a higher number of landmarks, the situation is reversed. Increasing the number of landmarks also increases the error by Euclidean approximation. This effect also increases with anisotropicity.

- The distance between the various means is far less sensitive to the number of landmarks.

- In accordance with the observeration in Huckemann and Ziezold (2006) for specials distributions on Σ_2^3 we have a general trend: Intrinsic and extrinsic mean are much closer to each other then to the principal component mean.

Secondly, the features in Figures 1 - 3 are discussed: The first PC geodesic captures notably more features than the first PC based on Euclidean approximation. The original triangles number 5 and number 6 (first and second in second row) are almost mirrored images of each other. This is very well reflected by the projection to the first PC geodesic, the first PC obtained by Euclidean approximation fails to capture this feature. In fact, the first PC geodesic seems to record the width and orientation of the inscribed angle and can thus very well discriminate between triangles, 7 and 12 (the third in the second row and the fourth in the third row), say. In contrast, the projections of these two triangles to the first Euclidean PC are similar. Only triangles 3 and 4 (the penultimate and the ultimate of the first row) are obviously better captured by the Euclidean PC. Slightly better captured are also triangles number 2 and 9 (the second of

the first row and the first of the third row). All other triangles are far better captured by the first PC geodesic.

We should note a difference between the (real) datasets of Section 6.1 and the above simulated datasets. All three real datasets appear as a mixture of the above distribution models. The first few eigenvalues of the sum of squares and products data matrix (26) tend to decay quadratically or stronger. Higher order eigenvalues, however, tend to decay less quickly.

In a final remark we note a limitation to the complex Bingham distribution. With every eigenvalue of the complex Hermitian model matrix comes a pair of eigenvalues for the corresponding real model matrix. Hence, data strongly following a single geodesic such as the dataset "rats skulls" cannot be modeled well using the Bingham distribution.

7 Conclusion

In this paper an algorithmic approach to perform PCA based on intrinsic geodesics for data on Kendall's planar shape spaces has been developed. With this method, PCA based on geodesics and PCA based on Euclidean approximation can now be compared. We have therefore considered typical distributions for two-dimensional shapes. The prevailing conviction (1) was empirically validated for highly concentrated data: the data fit by the first PC obtained by Euclidean approximation serves as a good estimate for the true geodesic fit. For low concentration, as expected, the estimate based on Euclidean approximation can be notably far away from the fit by the first PC geodesic. In particular this is the case for data with few landmarks. For higher concentration, modeling with the Bingham distribution, the error by Euclidean approximation increases notably with the anisotropicity of the data and the number of landmarks.

Acknowledgment

References

Abraham, R., Marsden, J. E., 1978. Foundations of Mechanics, 2nd Edition. Benjamin-Cummings.

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

Bhattacharya, R. N., Patrangenaru, V., 2005. Large sample theory of intrinsic and extrinsic sample means on manifolds II. Ann. Statist. 33 (3), 1225–1259.

Bookstein, F. L., 1978. The Measurement of Biological Shape and Shape Change, 2nd Edition. Vol. 24 of Lecture Notes in Biomathematics. Springer-Verlag, New York.

Chikuse, Y., Jupp, P. E., 2004. A test of uniformity on shape spaces. Journal of Multivariate Analysis 88, 163–176.

Dryden, I. L., Mardia, K. V., 1998. Statistical Shape Analysis. Wiley, Chichester.

Fletcher, P. T., Lu, C., Pizer, S. M., Joshi, S. C., 2004. Principal geodesic analysis for the study of nonlinear statistics of shape. IEEE Transactions on Medical Imaging 23 (8), 995–1005.

Gower, J. C., 1975. Generalized procrustes analysis. Psychometrika 40, 33–51.

Hendriks, H., Landsman, Z., 1998. Mean location and sample mean location on manifolds: asymptotics, tests, confidence regions. Journal of Multivariate Analysis 67, 227–243.

Huckemann, S., Ziezold, H., 2006. Principal component analysis for Riemannian manifolds with an application to triangular shape spaces. Adv. Appl. Prob. (SGSA) 38 (2), 299–319.

Karcher, H., 1977. Riemannian center of mass and mollifier smoothing. Communications on pure and applied mathematics XXX, 509–541.

Kendall, D. G., 1977. The diffusion of shape. Adv. Appl. Prob. 9, 428–430.

Kendall, D. G., Barden, D., Carne, T. K., Le, H., 1999. Shape and Shape Theory. Wiley, Chichester.

Kent, J., 1994. The complex Bingham distribution and shape analysis. Journal of the Royal Statistical Society, Series B 56, 285–299.

Le, H., 2001. Locating Fréchet means with an application to shape spaces. Adv. Appl. Prob. (SGSA) 33, 324–338.

Le, H., Kume, A., 2000. Detection of shape changes in biological features. Journal of Microscopy 200, 140–147.

Micheas, A. C., Dey, D. K., 2005. Modelling shape distributions and inferences for assessing differences in shapes. Journal of Multivariate Analysis 92, 257–280.

O'Neill, B., 1983. Semi-Riemannian Geometry. Academic Press, New York.

Small, C. G., 1996. The Statistical Theory of Shape. Springer-Verlag, New York.

Ziezold, H., 1977. Expected figures and a strong law of large numbers for random elements in quasi-metric spaces. Trans. 7th Prague Conf. Inf. Theory, Stat. Dec. Func., Random Processes A, 591–602.

Ziezold, H., 1994. Mean figures and mean shapes applied to biological figure and shape distributions in the plane. Biom. J. (36), 491–510.