

Analysis of Principal Nested Spheres

BY SUNGKYU JUNG

Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.

sungkyu@email.unc.edu

IAN L. DRYDEN

*School of Mathematical Sciences, University of Nottingham, Nottingham, NG7 2RD, UK and
Department of Statistics, University of South Carolina, Columbia, SC 29208, USA.*

Ian.Dryden@nottingham.ac.uk

AND J. S. MARRON

Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.

marron@unc.edu

SUMMARY

A general framework for a novel non-geodesic decomposition of high dimensional spheres or high dimensional shape spaces for planar landmarks is discussed. The decomposition, *Principal Nested Spheres*, finds a sequence of submanifolds with decreasing intrinsic dimensions, which can be interpreted as an analogue of Principal Component Analysis (PCA). In a number of real datasets, an apparent one dimensional mode of variation curving through more than one geodesic component is captured in the lowest dimensional Principal Nest Sphere (PNS). While analysis of PNS provides intuitive and flexible decomposition of the high dimensional sphere, an interesting special case of PNS results in finding principal geodesics, similar to those from previous approaches to manifold PCA. An adaptation of PNS to Kendall's shape space is discussed, and a computational algorithm for fitting PNS is proposed. The result of PNS provides a coordinate system to visualize the data structure, and an intuitive summary of principal modes of variation, as exemplified by several interesting spherical and shape data sets.

Some key words: dimension reduction; Kendall's shape space; manifold; principal arc; principal component analysis; spherical data.

1. INTRODUCTION

This paper proposes a general framework for a novel decomposition of a high dimensional sphere, which is the sample space of directions (Fisher (1993), Fisher et al. (1993), Mardia & Jupp (2000)) and pre-shapes in Kendall's statistical theory of landmark shapes (Kendall (1984), Dryden & Mardia (1998)). The proposed decomposition method, Principal Nested Spheres (PNS), is a flexible extension of Principal Component Analysis (PCA) for curved manifolds. PCA provides an effective means of analyzing the main modes of variation of the dataset and also gives a basis for dimension reduction. There have been a number of extensions of PCA to

manifold-valued data, most of which find principal geodesics (Fletcher et al. (2004), Huckemann & Ziezold (2006), Huckemann et al. (2010), Kenobi et al. (2010)).

A geodesic on a manifold is a shortest path between two points and can be understood as analogues of straight lines in Euclidean space. In particular, a geodesic on a sphere is a great circle path. The precise definition of geodesics on a sphere can be found in Eq. (A1), Appendix 8.

There has been a concern that when non-geodesic variation is major and apparent, the geodesic based PCA does not give a fully effective decomposition of the space. As an example, a dataset of shapes representing human movements, discussed later in Section 6.3 and introduced in Kume et al. (2007), is plotted in Fig. 1, using the first two principal component directions. In this dataset and many other interesting real data sets, the major one dimensional variation of the data curves through at least two components, and thus at least two dimensions are needed to explain the major variation. PNS decomposes the data space in a way that the major one dimensional variation is linearly represented, as shown in the bottom of Fig. 1.

For a unit d -sphere S^d , which is the set of unit vectors in \mathbb{R}^{d+1} , PNS gives a decomposition of S^d that captures the non-geodesic variation in a lower dimensional sub-manifold. The decomposition sequentially provides the best k -dimensional approximation \mathfrak{A}_k of the data for each $k = 0, 1, \dots, d - 1$. \mathfrak{A}_k is called the k -dimensional PNS, since it is essentially a sphere and is *nested* within (i.e. a sub-manifold of) the higher dimensional PNS. The sequence of PNS is then

$$\mathfrak{A}_0 \subset \mathfrak{A}_1 \subset \dots \subset \mathfrak{A}_{d-1} \subset S^d.$$

Since the preshape space of two dimensional landmark based shapes is also a hypersphere, the method can be readily applied to shape data, with some modifications (see Section 5). The analysis of PNS provides intuitive approximations of the directional or shape data for every dimension, captures the non-geodesic variation, and provides intuitive visualization of the major variability in terms of shape changes.

The procedure of fitting PNS involves iterative reduction of the dimensionality of the data. We first fit a $d - 1$ dimensional *subsphere* \mathfrak{A}_{d-1} of S^d that best approximates the data. This subsphere is not necessarily a great sphere (i.e. a sphere with radius 1, analogous to the great circle for S^2), which makes the resulting decomposition non-geodesic. Nevertheless, \mathfrak{A}_{d-1} can be treated as if it was the unit $(d - 1)$ -sphere by some geometric facts discussed in Section 2.1 and in Appendix 8 in greater detail. Each data point has an associated residual, which is the geodesic distance to its projection on \mathfrak{A}_{d-1} . Then for the data projected onto the subsphere, we continue to search for the best fitting $d - 2$ dimensional subsphere. These steps are iterated to find lower dimensional PNS. A detailed discussion of the procedure is in Section 2. For visualization and further analysis, we obtain an Euclidean-type representation of the data, essentially consisting of the residuals of each level. The first two coordinates of this representation, related to the one and two dimensional PNS, applied to the human movement data are plotted in Fig. 1.

In Fig. 1, PNS (bottom panel) has less curving variation. The proportion of variance in the 1- d PNS is almost the proportion of the sum of the first two geodesic component variances. That is, the variation explained by two geodesic components is attained in only one component of the PNS. Moreover, the graph in the top panel is indeed obtained by a special case of PNS analysis, which is similar to the geodesic-based PCA, as discussed in Section 2.6.

The procedure of PNS tends to find smaller spheres than the great sphere. Since this may cause an overfitting of the data, we developed a test procedure that can be applied to each layer to prevent the overfitting (see Section 3). A computational scheme for fitting PNS is proposed in Section 4. Necessary considerations and modifications for planar shape data are discussed in Section 5. In Section 6, we describe applications of the method to several interesting real datasets.

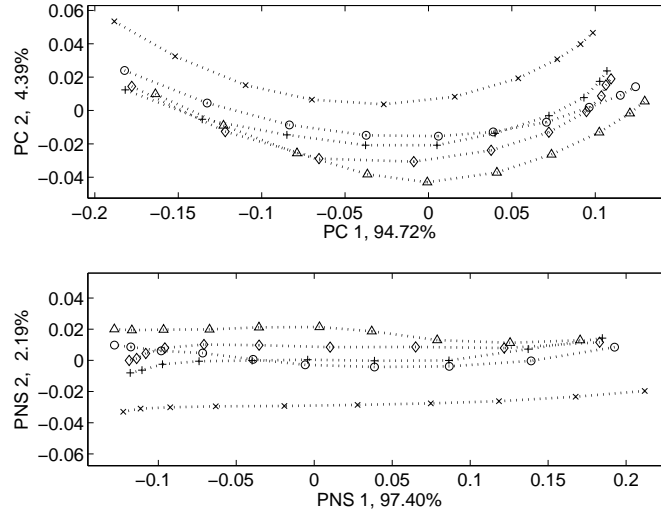


Fig. 1. Human movement data: (top) Scatter plot of two major geodesic components, where the different symbols represent different tasks, and samples for each task are interpolated. (bottom) Scatter plot of the first two PNS. The number% is the percent variance explained. The first PNS captures more of the interesting variation.

We conclude with discussions. Appendix (Sections 8 & 9) contains some geometric background to help define PNS, and proofs of the theorems.

2. PRINCIPAL NESTED SPHERES

2.1. Geometry of Nested Spheres

We begin with describing essential geometric background for nested spheres. More detailed discussion of the arguments in this section can be found in Appendix 8.

For a unit sphere S^d , a geodesic joining any two points is a great circle joining the two points. A natural distance function on S^d is the Riemannian (geodesic) distance function $\rho_d(\cdot, \cdot)$ defined as the length of the shortest great circle segment joining $\mathbf{x}, \mathbf{y} \in S^d$, $\rho_d(\mathbf{x}, \mathbf{y}) = \cos^{-1}(\mathbf{x}^T \mathbf{y})$.

A sequence of nested spheres of S^d is roughly a set of lower dimensional submanifolds that are essentially spheres. In the following, the precise form of nested spheres is introduced. We first define a subsphere of S^d , which induces the nested spheres.

DEFINITION 1. A subsphere A_{d-1} of S^d is defined by an orthogonal axis $\mathbf{v} \in S^d$ and a distance $r \in (0, \pi/2]$, as follows:

$$A_{d-1}(\mathbf{v}, r) = \{\mathbf{x} \in S^d : \rho_d(\mathbf{v}, \mathbf{x}) = r\},$$

where $\rho_d(\cdot, \cdot)$ is the geodesic distance function on S^d , $d \geq 2$.

The subsphere A_{d-1} can be viewed as an intersection of $S^d \subset \mathbb{R}^{d+1}$ and an affine d dimensional hyperplane, $\{\mathbf{x} \in \mathbb{R}^{d+1} : \mathbf{v}^T \mathbf{x} - \cos(r) = 0\}$. In other words, A_{d-1} is identified with a “slicing” of S^d with the affine hyperplane, an example of which is illustrated as a shaded plane in Fig 2. A subsphere A_{d-1} is indeed a $d - 1$ dimensional nested sphere \mathfrak{A}_{d-1} of S^d .

The subsphere A_{d-1} is isomorphic to S^{d-1} , as shown in Proposition 1, so we can treat the subsphere as a unit sphere S^{d-1} . This is done by an isomorphism $f_1 : A_{d-1} \rightarrow S^{d-1}$ and its

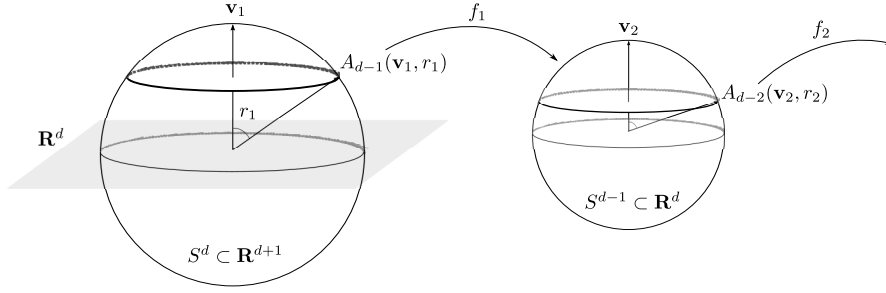


Fig. 2. The subsphere $A_{d-1}(\mathbf{v}_1, r_1)$ in S^d and its relation to S^{d-1} , through the isomorphism f_1 . Recursively, $A_{d-2}(\mathbf{v}_2, r_2)$ is found in S^{d-1} , and is isomorphic to S^{d-2} with the isomorphism f_2 .

inverse f_1^{-1} , defined in Eq. 1 below and also depicted in Fig. 2. Now a subsphere A_{d-2} of S^{d-1} can be obtained by applying Def. 1 with dimension d reduced by 1. For a general subsphere A_{d-k} of S^{d-k+1} , where $k = 1, \dots, d-1$, we also use the isomorphic transformation $f_k : A_{d-k} \rightarrow S^{d-k}$ and its inverse f_k^{-1} . Let $m = d - k + 1$, so that the subsphere $A_{d-k} \in S^m \subset \mathbb{R}^{m+1}$. The transformations are defined by $\mathbf{v}_k \in S^m$ and $r_k \in (0, \pi/2]$ as

$$\begin{aligned} f_k(\mathbf{x}) &= \frac{1}{\sin(r_k)} \mathbf{R}^-(\mathbf{v}_k) \mathbf{x}, \quad \mathbf{x} \in A_{d-k}, \\ f_k^{-1}(\mathbf{x}^\dagger) &= \mathbf{R}^T(\mathbf{v}_k) \begin{bmatrix} \sin(r_k) \cdot \mathbf{x}^\dagger \\ \cos(r_k) \end{bmatrix}, \quad \mathbf{x}^\dagger \in S^{d-k}, \end{aligned} \quad (1)$$

where $\mathbf{R}(\mathbf{v}_k)$ is the $(m+1) \times (m+1)$ rotation matrix that moves \mathbf{v}_k to the north pole (see Appendix 8), and $\mathbf{R}^-(\mathbf{v}_k)$ is the $m \times (m+1)$ matrix consisting of the first m rows of $\mathbf{R}(\mathbf{v}_k)$.

The subspheres A_{d-k} are defined in different spaces (in S^{d-k+1} for each k). A nested sphere is defined by the subsphere located in the original space S^d .

DEFINITION 2. A $d - k$ dimensional nested sphere \mathfrak{A}_{d-k} of S^d is defined as

$$\mathfrak{A}_{d-k} = \begin{cases} f_1^{-1} \circ \dots \circ f_{k-1}^{-1}(A_{d-k}) & \text{if } k = 2, \dots, d-1 \\ A_{d-1} & \text{if } k = 1 \end{cases}$$

A $d - k$ dimensional nested sphere \mathfrak{A}_{d-k} is indeed identified with a slicing of S^d by a $d - k + 1$ dimensional affine hyperplane. Note that, however, we work with each S^{d-k} , as it is logically simple in terms of dimensionality reduction as described in Section 2.3.

2.2. The Best Fitting Subsphere

Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be samples in S^d , $d \geq 2$. We first define the residual ξ of \mathbf{x} from a subsphere $A_{d-1}(\mathbf{v}_1, r_1)$ of S^d as the signed length of the minimal geodesic that joins \mathbf{x} to A_{d-1} . Then $\xi = \rho_d(\mathbf{x}, \mathbf{v}_1) - r_1$. The sign of ξ is negative if \mathbf{x} is in the interior of the geodesic ball corresponding to A_{d-1} , and is positive if \mathbf{x} is in the exterior.

The *best fitting subsphere* $\hat{A}_{d-1} \equiv A_{d-1}(\hat{\mathbf{v}}_1, \hat{r}_1)$ is found by minimizing the sum of squares of residuals of the data points to \hat{A}_{d-1} . In other words, $\hat{\mathbf{v}}_1$ and \hat{r}_1 minimize

$$\sum_{i=1}^n \xi_i(\mathbf{v}_1, r_1)^2 = \sum_{i=1}^n \{\rho_d(\mathbf{x}_i, \mathbf{v}_1) - r_1\}^2, \quad (2)$$

193 among all $\mathbf{v}_1 \in S^d$, $r_1 \in (0, \pi/2]$.

194 Note that the method can be extended using other objective functions, e.g. sum of absolute
195 deviations for more robust fitting.

196 Each \mathbf{x}_i can be projected on \widehat{A}_{d-1} along the minimal geodesic that joins \mathbf{x} to \widehat{A}_{d-1} . The
197 projection of \mathbf{x} onto A_{d-1} is defined as

$$198 \quad P(\mathbf{x}; A_{d-1}(\mathbf{v}, r)) = \frac{\sin(r)\mathbf{x} + \sin(\rho_d(\mathbf{x}, \mathbf{v}) - r)\mathbf{v}}{\sin\{\rho_d(\mathbf{x}, \mathbf{v})\}}. \quad (3)$$

199 Denote $\mathbf{x}^P = P(\mathbf{x}; \widehat{A}_{d-1}) \in \widehat{A}_{d-1}$ for the projected \mathbf{x} . We use the isomorphism $\hat{f}_1 \equiv$
200 $f(\hat{\mathbf{v}}_1, \hat{r}_1)$, as defined in Eq. (1), to transform \widehat{A}_{d-1} to S^{d-1} so that $\hat{f}_1(\mathbf{x}_i^P) \in S^{d-1}$.

201 2.3. The sequence of Principal Nested Spheres

202 The sequence of PNS are fully meaningful when they are in the same space. On the other hand,
203 utilizing the isomorphic spaces of the nested spheres, i.e. the unit spheres, makes the process
204 simpler. Therefore the procedure to find the sample PNS consists of iteratively finding the best
205 fitting subsphere and mapping to the original space.

206 The $d-1$ dimensional sample PNS $\widehat{\mathfrak{A}}_{d-1}$ is the same as the best fitting subsphere
207 $A_{d-1}(\hat{\mathbf{v}}_1, \hat{r}_1)$ because both are in the original space S^d . The second layer, the $d-2$ dimen-
208 sional sample PNS, is obtained from the subsphere that best fits $\hat{f}_1(\mathbf{x}_i^P) \in S^{d-1}$. The best fitting
209 subsphere $A_{d-2}(\hat{\mathbf{v}}_2, \hat{r}_2)$ is then mapped to S^d by the relevant isomorphism f_1^{-1} and becomes
210 $\widehat{\mathfrak{A}}_{d-2}$.

211 In general, we recursively find the sequence of best fitting subspheres from the projected and
212 transformed samples, i.e. $\mathbf{x} \mapsto \hat{f}_k(P(\mathbf{x}; \widehat{A}_{d-k}))$. In the ‘ k th level’, where we fit a subsphere
213 from S^{d-k+1} , we denote the best fitting subsphere as $\widehat{A}_{d-k} \equiv A_{d-k}(\hat{\mathbf{v}}_k, \hat{r}_k)$ and keep residuals
214 $\xi_i \equiv \xi_{i,d-k}$, $i = 1, \dots, n$, for later use as analogs of principal component scores.

215 The lowest level best fitting subsphere \widehat{A}_1 is then a small circle isomorphic to S^1 . No further
216 sphere or circle can be used to reduce the dimensionality. Instead, we find the Fréchet mean
217 (Fréchet (1944, 1948) and Karcher (1977)) \widehat{A}_0 of $\mathbf{x}_1^\dagger, \dots, \mathbf{x}_n^\dagger$ (the projected and transformed
218 samples in S^1) which can be thought of as a best 0-dimensional representation of the data in the
219 framework of PNS. The Fréchet mean \widehat{A}_0 is defined as the minimizer of the squared distances to
220 the \mathbf{x}_i^\dagger s, i.e.

$$221 \quad \widehat{A}_0 = \operatorname{argmin}_{\mathbf{x} \in S^1} \sum_{i=1}^n \rho_1(\mathbf{x}, \mathbf{x}_i^\dagger)^2.$$

222 The Fréchet mean is unique when the support of \mathbf{x}_i^\dagger is a proper subset of a half circle in S^1 , which
223 is often satisfied in practice. If there are multiple Fréchet means, then careful inspection of the
224 data must be followed. A typical case for having multiple means is that the data are uniformly
225 distributed on the circle. If this is the case, then $\widehat{\mathfrak{A}}_0$ can be chosen to be any solution of the above
226 criterion, and since it does not summarize the data well we may not lay much emphasis on \widehat{A}_0 .

227 The sequence of best fitting subspheres including the \widehat{A}_0 can be located in the original space
228 S^d , as follows.

229 DEFINITION 3. *The sequence of sample Principal Nested Spheres in S^d is then*
230 $\{\widehat{\mathfrak{A}}_0, \widehat{\mathfrak{A}}_1, \dots, \widehat{\mathfrak{A}}_{d-1}\}$, where

$$231 \quad \widehat{\mathfrak{A}}_{d-k} = \begin{cases} \hat{f}_1^{-1} \circ \dots \circ \hat{f}_{k-1}^{-1}(\widehat{A}_{d-k}) & \text{if } k = 2, \dots, d, \\ \widehat{A}_{d-1} & \text{if } k = 1. \end{cases}$$

232

We call $\widehat{\mathfrak{A}}_0$ the PNSmean.

2.4. Euclidean-type representation

We wish to represent the data in an Euclidean space for visualization and further analysis.

Recall that in the k th level of the procedure, we have collected the signed residuals which we denote by $\xi_{i,d-k}$, $i = 1, \dots, n$. These were measured by the metric ρ_{d-k} in a space different from S^d . Therefore we scale these residuals by multiplying $\prod_{i=1}^{k-1} \sin(\hat{r}_i)$ which makes the magnitude of residuals commensurate (see Proposition 2). We put the scaled residuals in a row vector

$$\Xi(d-k)_{1 \times n} \doteq \prod_{i=1}^{k-1} \sin(\hat{r}_i) [\xi_{1,d-k}, \dots, \xi_{n,d-k}].$$

We further define $\xi_{i,0}$ as the i th sample's signed deviation from \widehat{A}_0 measured by ρ_1 . Similar to before, rescale the deviations and let

$$\Xi(0)_{1 \times n} \doteq \prod_{i=1}^{d-1} \sin(\hat{r}_i) [\xi_{1,0}, \dots, \xi_{n,0}].$$

These commensurate residuals are combined into a $d \times n$ data matrix

$$\widehat{X}_{PNS} = \begin{bmatrix} \Xi(0) \\ \Xi(1) \\ \vdots \\ \Xi(d-1) \end{bmatrix},$$

where each column is the corresponding sample's coordinates in terms of the sample PNS. Each entry in row k works like the k th principal component score.

The data matrix \widehat{X}_{PNS} can be used to visualize the structure of the data. For example, the graph in Fig. 1 is a scatterplot of $\Xi(0)$ and $\Xi(1)$. The variance of each component is defined by the variance of the corresponding residuals. Moreover, conventional multivariate statistics based on Euclidean space can be applied to \widehat{X}_{PNS} for further analysis (e.g. PCA and classification methods).

2.5. Principal Arcs

In analogy to the principal component directions in Euclidean space, or the manifold extension principal geodesics, the *principal arcs* that represent the *direction* of major variations are defined by PNS. These arcs are space curves lying in the manifold S^d , which frequently are not equivalent to any geodesic.

Given a sequence of PNS $\{\widehat{\mathfrak{A}}_0, \widehat{\mathfrak{A}}_1, \dots, \widehat{\mathfrak{A}}_{d-1}\}$, the first principal arc coincides with the 1- d PNS $\widehat{\mathfrak{A}}_1$. This arc may be parameterized by the signed distance from the PNSmean $\widehat{\mathfrak{A}}_0$. In the space of the Euclidean-type representation \widehat{X}_{PNS} , the first principal arc coincides with the direction $\mathbf{e}_1 = (1, 0, \dots, 0)^T$.

The second principal arc lies in $\widehat{\mathfrak{A}}_2$ and is orthogonal to the first principal arc at all points in common. The first and second arcs cross at $\widehat{\mathfrak{A}}_0$ and also at the farthest point from $\widehat{\mathfrak{A}}_0$ on $\widehat{\mathfrak{A}}_1$. The second arc is in general a small circle in S^d but is identified with a great circle in S^2 , the isomorphic space of $\widehat{\mathfrak{A}}_2$. The second principal arc in S^2 must pass through the axis \mathbf{v}_{d-1} in order to be orthogonal to the first. This arc may be parameterized by the signed distance from $\widehat{\mathfrak{A}}_0$, and coincides with the direction \mathbf{e}_2 in the space of \widehat{X}_{PNS} .

289 The higher order principal arcs are defined in the same manner. The k th principal arc can be
 290 defined and identified with the direction \mathbf{e}_k in the space of \widehat{X}_{PNS} . The k th arc is then orthogonal
 291 to the principal arcs of order $1, \dots, k-1$, and passes through the PNSmean.

292 In addition, a space curve in S^d may be parameterized by the coordinates of the Euclidean-type
 293 representation. Fitting a space curve that further smooths the data is a separate issue.

294 2.6. Principal Nested Spheres restricted to Great Spheres

295 An important special case of the PNS is obtained by setting $r = \pi/2$ for each subsphere fitting.
 296 This restriction leads to the nested spheres being great spheres, and the principal arcs become
 297 geodesics. In all data sets we tested, the resulting principal geodesics are similar to the previous
 298 geodesic-based PCA methods. In the following, we indicate this special case as *Principal Nested*
 299 *Great spheres* (PNG).

300 We conjecture that the principal geodesics, found by PNG, are more similar to the Geodesic
 301 Principal Component of Huckemann et al. (2010) than the usual tangent space projection meth-
 302 ods. This is mainly because any pre-determined mean (either geodesic mean or Procrustes mean)
 303 is not used in PNG nor Huckemann's. The PNSmean in this special case is similar to the notion
 304 of mean of Huckemann, and is identical when the sphere has dimension $d = 2$. Although we
 305 have not yet found a significant difference of PNG than previous methods, we point out that the
 306 PCA extension approach of PNG (and PNS) is different from those methods. See Section 7 for
 307 discussion on this.
 308

310 3. PREVENTION OF OVERFITTING BY SEQUENTIAL TESTS

311 3.1. Significance of small sphere fitting

312 In this section, the significance of small spheres against the great sphere is discussed. We
 313 propose a test procedure consisting of two different tests for each level of subsphere fitting.
 314 Similar to the backward regression procedure, sequentially testing small spheres at each layer
 315 may prevent overfitting.
 316

317 There are two cases where a great sphere provides more appropriate fit to the data, yet the
 318 sum of squared residuals is minimized by a small sphere. The first case is where a true major
 319 variation is along a great sphere, an example of which on S^2 is illustrated in Fig 3a. The second
 320 case is when the underlying distribution is isotropic with a single mode, so that there is no major
 321 variation along any direction. An example of such a distribution is $N(\mathbf{0}, \mathbf{I}_k)$ (in linear space), or
 322 the von Mises–Fisher distribution on S^d (Fisher (1953), Mardia & Jupp (2000)), as illustrated in
 323 Fig 3b. In this situation, small spheres centered at the point of isotropy are frequently obtained,
 324 which do not give a useful decomposition.

325 We have developed two different tests to handle these cases. The first is a likelihood ratio
 326 test (LRT) for the detection of the first case above (Fig 3a), which tests the significance of the
 327 reduction of residual variances. The second is a parametric bootstrap test aimed at the second case
 328 above (Fig 3b), which tests the isotropy of the underlying distribution. A detailed description of
 329 the tests is given in the following subsections. A procedure to apply these tests to PNS fitting is
 330 then discussed in Section 3.4.

331 3.2. Likelihood ratio test

332 We define a likelihood ratio statistic for each level to sequentially test the significance of small
 333 sphere fitting against the great sphere.

334 For the k th level of the procedure, where A_{d-k} is fitted to $\mathbf{x}_1, \dots, \mathbf{x}_n \in S^{d-k+1}$, we assume
 335 that the deviations of the samples \mathbf{x}_i from the subsphere $A_{d-k}(\mathbf{v}, r)$ are independent $N(0, \sigma^2)$.
 336

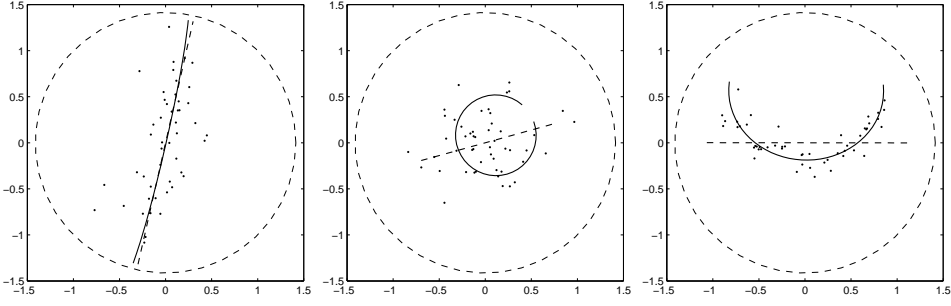


Fig. 3. Simulated data examples on S^2 projected by an equal area projection, and the fitted best small (solid) and great (dotted) spheres, which are arcs in this two dimensional case. (a) The LRT gives the p-value 0.565, while p-value of the bootstrap test is 0. The LRT detects the overfitting. (b) The LRT leads to p-value ≈ 0 , and the bootstrap p-value is 0.82. The bootstrap test detects the overfitting. (c) When the fitted small sphere (circle) is not overfitted, both tests give very small p-values (≈ 0). This assures that the small sphere is not overly fitted.

It makes more sense when a truncated Normal distribution on a range $[-\pi/2, \pi/2]$ is assumed. However unless the data spread too widely (e.g. Uniform on the sphere), the distribution will be approximately Normal. Thus we use the approximate likelihood function of $(\mathbf{v}, r, \sigma^2)$, given by

$$L(\mathbf{v}, r, \sigma^2 | \mathbf{x}_1^n) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (\rho(\mathbf{x}_i, \mathbf{v}) - r)^2\right),$$

where ρ is the geodesic distance function on S^{d-k+1} . The approximate maximum likelihood estimator (m.l.e.) of (\mathbf{v}, r) coincides with $(\hat{\mathbf{v}}, \hat{r})$, the solution of Eq. (2), and the approximate m.l.e. of σ^2 is given by $\hat{\sigma}^2 = n^{-1} \sum_{i=1}^n \xi_i(\hat{\mathbf{v}}, \hat{r})^2$, which is obtained by differentiating the log-likelihood function and setting the derivative equal to zero.

We can test $H_{0a} : r = \pi/2$ (i.e. the great sphere), versus $H_{1a} : r < \pi/2$ (i.e. some small sphere), using a likelihood ratio test. The m.l.e. of $(\mathbf{v}, r, \sigma^2)$ under H_{0a} is given by $(\hat{\mathbf{v}}^0, \pi/2, \hat{\sigma}_0^2)$, where $\hat{\mathbf{v}}^0$ minimizes the sum of squared residuals of Eq. (2) with $r = \pi/2$, and $\hat{\sigma}_0^2 = n^{-1} \sum_{i=1}^n \xi_i(\hat{\mathbf{v}}^0, \pi/2)^2$. The log-likelihood ratio is $(\hat{\sigma}^2/\hat{\sigma}_0^2)^{-n/2}$. Then using Wilks' theorem, for large samples $n \log(\hat{\sigma}^2/\hat{\sigma}_0^2) \approx \chi_1^2$ under H_{0a} , and the test rejects H_{0a} in favor of H_{1a} for large values of $n \log(\hat{\sigma}^2/\hat{\sigma}_0^2)$.

3.3. Parametric bootstrap test

For each level of PNS fitting, suppose $\mathbf{X} \in S^m$ has a distribution function $F_{\mathbf{X}}$. We wish to test for the underlying distribution $F_{\mathbf{X}}$, $H_{0b} : F_{\mathbf{X}}$ is an isotropic distribution with a single mode, versus $H_{1b} : \text{not } H_{0b}$ (i.e. anisotropic). We develop a parametric bootstrap test with an assumption of the von Mises-Fisher distribution. The von Mises-Fisher distribution is an analogue of Normal distribution on the unit sphere with concentration parameter κ and directional parameter μ , denoted as $\text{vMF}(\mu, \kappa)$.

We build a test statistic that is large when $F_{\mathbf{X}}$ is neither isotropic nor having a single mode. For this purpose, we derive the following test statistic. Given $\mathbf{x}_1, \dots, \mathbf{x}_n \in S^m$, estimate the best fitting subsphere $A(\hat{\mathbf{v}}, \hat{r})$ as done in Eq. (2). Let $\zeta_i = \rho_d(\mathbf{x}_i, \hat{\mathbf{v}}) = \cos^{-1}(\mathbf{x}_i^T \hat{\mathbf{v}})$ be the radial distances from the axis of the subsphere. Then the test statistic to use is the coefficient of variation

of ζ ,

$$Z = Z(\mathbf{x}_1, \dots, \mathbf{x}_n) = \frac{\bar{\zeta}}{\text{std}(\zeta)} = \frac{\frac{1}{n} \sum_{i=1}^n \zeta_i}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (\zeta_i - \bar{\zeta})^2}}.$$

The next step is to estimate a null distribution of this statistic. We have assumed that under H_{0b} , $F_{\mathbf{X}}$ is $\text{vMF}(\mu, \kappa)$. The unknown parameters are estimated from the data. μ is estimated by a standard likelihood approach, see Mardia & Jupp (2000). For an estimate of κ , Banerjee et al. (2005) empirically derived an approximation of the m.l.e. of κ . The estimates are

$$\hat{\mu}^{MLE} = \frac{r}{\|r\|} = \frac{\sum_{i=1}^n x_i}{\|\sum_{i=1}^n x_i\|}, \quad \hat{\kappa}^{MLE} \approx \frac{\bar{r}(d+1) - \bar{r}^3}{1 - \bar{r}^2},$$

where $\bar{r} = \frac{\|r\|}{n}$. Then we generate $B \geq 100$ random samples of size n from $\text{vMF}(\hat{\mu}^{MLE}, \hat{\kappa}^{MLE})$ and calculate Z_1, \dots, Z_B . The test rejects H_{0b} with a significance level α if

$$\frac{1}{B} \sum_{i=1}^B 1_{\{Z_i > Z\}} < \alpha.$$

3.4. Application procedure

As discussed in section 2.3, a sequence of sample PNS is obtained by iterative fitting of subspheres. In each layer of subsphere fitting, both of the tests developed in this section will be used, due to the observation in Fig. 3. We first illustrate how these tests are applied to the examples in Fig. 3, then propose a procedure to apply the tests to the PNS fitting procedure.

Some typical data examples on the 2-sphere and the results of the two tests are illustrated in Fig. 3. When the true major variation is along a great circle, as in Fig. 3a, the LRT works well and accepts H_{0a} (great sphere) but the bootstrap test rejects H_{0b} . On the other hand, when the underlying distribution is von Mises–Fisher, the LRT rejects H_{0a} in favor of H_{1a} : small sphere. However, the best fitting small sphere is frequently inappropriate, as shown in Fig. 3b. The bootstrap test accepts H_{0b} and thus can be used to detect such a case. Therefore, in order to prevent an overfitting, we proposed to sequentially apply both tests in each level of subsphere fitting. In a case where a true variation is along a small sphere, both tests reject the null hypotheses, and we assure that the small subsphere is not overfitting.

In each level of subsphere fitting, we use the following testing procedure to test the significance of “small” subsphere fitting.

1. Test H_{0a} versus H_{1a} by the likelihood ratio test. If H_{0a} is accepted, then fit a great sphere with $r = \pi/2$ and proceed to the next layer.
2. If H_{0a} is rejected, then test the isotropy of the distribution by the parametric bootstrap test. If H_{0b} is accepted, then use great spheres for ‘all’ further subsphere fittings.
3. If both tests do not reject the null hypotheses, then use the fitted small sphere for decomposition.

Note that in step 2, when H_{0b} is accepted, we use great sphere fitting not only for the level, but also for all further levels with smaller dimensions. This is because once H_{0b} is accepted, the underlying distribution at the level is assumed to be a von Mises-Fisher. An analogy in Euclidean space is $N(0, I_k)$ where a non-linear mode of variation is meaningless. Therefore, great spheres are used for all further nested spheres, without further application of tests.

Note that for S^d , we test at most $2(d-1)$ hypotheses. This brings us a multiple testing problem, i.e. using significance level $\alpha = 0.05$ for every test may result in a larger overall type I

error. This phenomenon can be treated by, for example, using Bonferroni's correction. Deeper discussion of more advanced treatments, such as False Discovery Rate, is beyond the scope of this paper.

4. COMPUTATIONAL ALGORITHM

The computation of sample PNS involves iterative applications of minimization, projection and transformation. We have given explicit formulas for the projection (Eq. 3) and the transformation (Eq. 1). The least squares problem (Eq. 2) is a constrained non-linear minimization problem. It can be solved by the doubly iterative algorithm described in Jung et al. (2010) with some modifications. The algorithm is best understood in two iterative steps: The outer loop finds the point of tangency to approximate S^d by a tangent space; the inner loop solves an optimization problem in the linear space.

We make use of the exponential map and its inverse for mappings between the manifold and tangent spaces (see Helgason (2001) and Buss & Fillmore (2001)). A tangent space at $\mathbf{p} \in S^m$, $T_{\mathbf{p}}S^m$, is an affine m -dimensional vector space and can be identified by \mathbb{R}^m . Without loss of generality set the point of tangency $\mathbf{p} = \mathbf{e}_{m+1} = (0, \dots, 0, 1)$, because one can use the rotation operator $\mathbf{R}(\mathbf{p})$ to transform \mathbf{p} to \mathbf{e}_{m+1} while preserving all data structure. The exponential map $\text{Exp}_{\mathbf{p}} : T_{\mathbf{p}}S^m \rightarrow S^m$ is defined for $\mathbf{z} \in \mathbb{R}^m \cong T_{\mathbf{p}}S^m$,

$$\text{Exp}_{\mathbf{p}}(\mathbf{z}) = \left(\frac{\sin(\|\mathbf{z}\|)}{\|\mathbf{z}\|} \mathbf{z}^T, \cos(\|\mathbf{z}\|) \right)^T \in S^m.$$

The inverse exponential map (log map) $\text{Log}_{\mathbf{p}} : S^m \rightarrow T_{\mathbf{p}}S^m$ is defined for $\mathbf{x} = (x_1, \dots, x_{m+1})^T \in S^m$,

$$\text{Log}_{\mathbf{p}}(\mathbf{x}) = \frac{\theta}{\sin(\theta)} (x_1, \dots, x_m)^T \in \mathbb{R}^m,$$

where $\cos(\theta) = x_{m+1}$. These mappings preserve the distances to the point of tangency. By using the exponential mapping and its inverse, a hypersphere with radius r in the tangent space corresponds to a subsphere in S^m with distance r . In particular, $A_{m-1}(\mathbf{v}, r)$ is equivalent to the image of $\{\mathbf{x} \in \mathbb{R}^m : \|\mathbf{x}\| = r\}$ by $\text{Exp}_{\mathbf{v}}$.

The algorithm finds a suitable point of tangency \mathbf{v} , which is also the center of the fitted subsphere. Given a candidate \mathbf{v}_0 , the data are mapped to the tangent space $T_{\mathbf{v}_0}S^m$ by the log map. Write $\mathbf{x}_i^\dagger = \text{Log}_{\mathbf{v}_0}(\mathbf{x}_i)$, then the inner loop finds the minimizer of

$$\min_{\mathbf{v}^\dagger, r} \sum_{i=1}^n (\|\mathbf{x}_i^\dagger - \mathbf{v}^\dagger\| - r)^2,$$

which is a non-linear least-squares problem and can be solved numerically by e.g. the Levenberg-Marquardt algorithm (see e.g. Ch.4 of Scales (1985)). The solution \mathbf{v}^\dagger is then mapped to S^m by the exponential map, and becomes the \mathbf{v}_1 . This procedure is repeated until \mathbf{v} converges.

A main advantage of this approach is the reduced difficulty of the optimization task. The inner loop solves an unconstrained problem in a vector space, which is much simpler than the original constrained problem on manifolds. Experience has shown that with a carefully chosen initial value, the algorithm has worked well in a wide range of simulated and real applications.

It becomes increasingly common in modern applied problems that the sample size is less than the dimension of the manifold, i.e. $\mathbf{x}_1, \dots, \mathbf{x}_n \in S^d$ with $n \leq d$, which is frequently referred

481 to as the high dimension, low sample size situation (Hall et al. (2005), Dryden (2005)). In Eu-
 482 clidean space, the dimensionality of the data can be reduced to n without losing any information.
 483 Likewise, the intrinsic dimensionality of the data on the hypersphere can be reduced to $n - 1$,
 484 where additional reduction of 1 occurs because there is no ‘origin’ in S^d . For the simplest yet
 485 intuitive example, let $n = 2$. Then there is a geodesic joining the two points, which is the sub-
 486 manifold containing all information. A generalization of this fact can be made for any $n > 2$, by
 487 the following theorem.

488 **THEOREM 1.** *There exists an $n - 1$ dimensional nested sphere \mathfrak{A}_{n-1} of S^d satisfying $\mathbf{x}_i \in$
 489 \mathfrak{A}_{n-1} for all $i = 1, \dots, n$. Moreover, there exist $\mathfrak{A}_{d-1} \supset \dots \supset \mathfrak{A}_{n-1}$, all of which are great
 490 spheres (i.e. with radius 1).*
 491

492 As can be seen in the proof of the theorem in the Appendix, the singular value decomposition
 493 of the data matrix $[\mathbf{x}_1 \dots \mathbf{x}_n]$ gives the appropriate \mathfrak{A}_{n-1} . Let \mathcal{L}_1^n be the vector space of dimension
 494 n that all data points span. Then, the intersection of \mathcal{L}_1^n and S^d is the $n - 1$ dimensional manifold
 495 \mathfrak{A}_{n-1} .

496 For a faster computation (when $n < d$), we reduce the dimensionality to \mathfrak{A}_{n-1} by the singular
 497 value decomposition, and use the proposed algorithm to fit \mathfrak{A}_{n-2} , and so on.
 498

5. APPLICATION TO PLANAR SHAPE SPACE

5.1. Planar shape space

499 The shape of an object is what is left after removing location, scale, and rotation. The classical
 500 approach in shape analysis (see e.g. Dryden & Mardia (1998)) is to work with (biological) land-
 501 marks of the objects. Each shape determined by a set of landmarks can be represented by a point
 502 in Kendall’s (1984) shape space. A useful approach to understanding the non-Euclidean shape
 503 space is through preshape space, which is a high dimensional sphere. We begin by summarizing
 504 the Kendall’s framework for shape data, followed by a discussion of necessary considerations to
 505 apply PNS to shape space through the preshape space.
 506

507 Consider a set of $k > 2$ landmarks in \mathbb{R}^2 and the corresponding configuration matrix \mathbf{X} , which
 508 is a $k \times 2$ matrix of Cartesian coordinates of landmarks. The preshape of the configuration \mathbf{X}
 509 is invariant under translation and scale, which is given by $\mathbf{Z} = \mathbf{HX} / \|\mathbf{HX}\|$, where \mathbf{H} is the
 510 $(k - 1) \times k$ Helmert sub-matrix (Dryden & Mardia, 1998, p. 34). Provided that $\|\mathbf{HX}\| > 0$,
 511 $\mathbf{Z} \in S^{2(k-1)-1}$. The unit sphere $S^{2(k-1)-1}$ in $\mathbb{R}^{2(k-1)}$ is the space of all possible preshapes, and
 512 is called the preshape space.
 513

514 The shape of a configuration matrix \mathbf{X} can be represented by the equivalence set under rota-
 515 tion, $[\mathbf{Z}] = \{\mathbf{Z}\Gamma : \Gamma \in SO(2)\}$, where $SO(2)$ is the set of all 2×2 rotation matrices. The space
 516 of all possible shapes is then a non-Euclidean space called the shape space and is denoted by Σ_2^k .
 517

518 We also write the preshape \mathbf{Z} as a vectorized version $\mathbf{z} = \text{vec}(\mathbf{Z}^T)$, where $\text{vec}(\mathbf{A})$ is obtain
 519 by stacking the columns of the matrix \mathbf{A} on top of one another. Then the following facts are
 520 well-known (see e.g. Kume et al. (2007) and Dryden & Mardia (1998)).
 521

522 Suppose that $\mathbf{v}, \mathbf{w} \in S^{2(k-1)-1}$ are preshapes satisfying $\mathbf{v}^T \mathbf{w} \geq 0$ and $\mathbf{v}^T \mathbf{M} \mathbf{w} = 0$, where
 523 \mathbf{M} is the $2(k - 1) \times 2(k - 1)$ skew-symmetric matrix consisting of $k - 1$ diagonal blocks

$$524 \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

525 Then the geodesic that joins \mathbf{v} to \mathbf{w} , $\mathbf{Q}(\mathbf{v} \rightarrow \mathbf{w}, \theta)\mathbf{v}$, $\theta \in [0, \pi/2]$ is said to be a horizontal
 526 geodesic, and the Riemannian distance between \mathbf{v}, \mathbf{w} is the same as the Riemannian distance
 527 between the corresponding shapes $[\mathbf{v}]$ and $[\mathbf{w}]$ in Σ_2^k .
 528

For the preshapes $\mathbf{w}, \mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n$, as long as the shapes of those are of interest, we assume without loss of generality that $\mathbf{w}^T \mathbf{v}_i \geq 0$ and $\mathbf{w}^T \mathbf{M} \mathbf{v}_i = 0$.

5.2. Principal Nested Spheres for planar shapes

The intrinsic dimension of the shape space is $2k - 4$, since the degrees of freedom are reduced from $2k$ (of the set of landmarks) by 2 for translation, 1 for scale, and 1 for rotation. This is less than the dimension of the preshape space $d = 2k - 3$. It is thus desired that the $d - 1$ dimensional PNS of S^d leaves no residuals. This is achieved by the theory and practical modifications in this section. In short, Procrustes fit of configurations or preshapes to a common alignment base (e.g. the Procrustes mean) results in the desired decomposition of the shape space.

THEOREM 2. *Suppose the preshapes $\mathbf{w}, \mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n \in S^d$ satisfy $\mathbf{w}^T \mathbf{z}_i \geq 0$ and $\mathbf{w}^T \mathbf{M} \mathbf{z}_i = 0$ for all $i = 1, \dots, n$. Let $\mathbf{w}^* = \mathbf{M} \mathbf{w}$ for M defined above. Then $\mathbf{w}, \mathbf{z}_i \in A_{d-1}(\mathbf{w}^*, \pi/2)$. Moreover, define $hA_{d-1} = \{\mathbf{z} \in A_{d-1} : \mathbf{z}^T \mathbf{w} \geq 0\}$ as a hyper-hemisphere. Then $\mathbf{w}, \mathbf{z}_i \in hA_{d-1}(\mathbf{w}^*, \pi/2)$.*

We have the following comments:

- The dimension of preshape space can be reduced by 1 without loss of any shape information.
- For the nested hemisphere hA_{d-1} ,
 1. the intrinsic distance $\rho_{d-1}^*(\mathbf{w}, \mathbf{z})$ defined on hA_{d-1} (see Proposition 1(b) in Appendix 8) is the same as the Riemannian distance $\rho([\mathbf{w}], [\mathbf{z}])$ in Σ_2^k for any $\mathbf{z} \in hA_{d-1}$.
 2. the tangent space of hA_{d-1} at \mathbf{w} is in fact identical to the horizontal subspace of the tangent space of S^d at \mathbf{w} .
- The hA_{d-1} is closely related to Σ_2^k , but is not identical.

When $k = 3$, the preshape space has dimension $d = 2(k - 1) - 1 = 3$. The corresponding shape space of planar triangles Σ_2^3 is $S^2(\frac{1}{2})$. hA_2 obtained from some $\mathbf{w} \in S^3$ is isometric to a unit hemisphere in \mathbb{R}^3 . A geodesic in Σ_2^3 may or not be identified with a geodesic in hA_2 . A geodesic in Σ_2^3 through $[\mathbf{w}]$ is identified with a geodesic in hA_2 through \mathbf{w} . On the other hand, a set of points in distance $\pi/4$ from $[\mathbf{w}]$ in Σ_2^3 is a geodesic, but is identified with a small circle with center \mathbf{w} and radius $\pi/4$ in hA_2 .

The choice of the alignment base \mathbf{w} is an important issue because the Riemannian distance in hA_{d-1} is the same as the Riemannian distance in the shape space Σ_2^k when compared to \mathbf{w} , i.e.

$$\rho_d(\mathbf{w}, \mathbf{z}) = \rho_\Sigma([\mathbf{w}], [\mathbf{z}]), \text{ for } \mathbf{z} \in hA_{d-1}.$$

Moreover, $\rho_d(\mathbf{z}_1, \mathbf{z}_2)$ for $\mathbf{z}_1, \mathbf{z}_2 \in hA_{d-1}$ is closer to $\rho_\Sigma([\mathbf{z}_1], [\mathbf{z}_2])$ when $\mathbf{z}_1, \mathbf{z}_2$ are close to \mathbf{w} .

In general, we wish to set the alignment base \mathbf{w} as a center of the data. Among many reasonable options of \mathbf{w} , we recommend to use the preshape of the Procrustes mean of the data. Other reasonable candidates for \mathbf{w} are the geodesic mean and the PNSmean. We have tested these options to a number of real and simulation datasets. Setting \mathbf{w} as the PNSmean or the geodesic mean usually takes longer computation time than using the full Procrustes mean, and the resulting decompositions are virtually the same in most cases.

In the following, we describe all candidates of \mathbf{w} in more detail, giving the advantages and disadvantages of each option.

We have first considered use of the PNSmean \mathfrak{A}_0 as the alignment base. \mathfrak{A}_0 is identified with the origin of the coordinate system for the Euclidean representation of data \widehat{X}_{PNS} . Since the PNSmean is estimated from the data, we begin with the preshape of the full Procrustes mean as

an initial guess for \mathfrak{A}_0 and recursively update $\widehat{\mathfrak{A}}_0$ on which preshapes are aligned. The algorithm consists of the following steps.

1. Initialize \mathbf{w} as the preshape of the Procrustes mean of \mathbf{z}_i .
 2. Align $\mathbf{z}_1, \dots, \mathbf{z}_n$ to \mathbf{w} and compute the sample PNSmean $\widehat{\mathfrak{A}}_0$ of aligned \mathbf{z}_i .
 3. If $\rho_d(\mathbf{w}, \widehat{\mathfrak{A}}_0) < \epsilon$, then set $\mathbf{w} = \widehat{\mathfrak{A}}_0$ and stop. Otherwise update $\mathbf{w} = \widehat{\mathfrak{A}}_0$ and go to Step 2.
- Note that in practice, there is no guarantee that this algorithm should converge.

Other candidates of \mathbf{w} are the full Procrustes mean preshape and the geodesic mean of the preshapes. These are relevant to the Fréchet mean, where the geodesic mean is the Fréchet mean with the intrinsic (Riemannian) distance and the Procrustes mean is using the full Procrustes distance which is extrinsic to S^d . Recently, it has been observed that the curvature of the manifold S^d sometimes makes the Fréchet mean inadequate, see e.g. Huckemann et al. (2010). When the Fréchet mean is indeed a useful representation of the data, the PNSmean is usually found at a point close to the Fréchet mean. Note that even if the Fréchet mean is far from the data, the PNSmean is nevertheless located at the appropriate center of the data.

Finally, the tests for overfitting discussed in Section 3 can be applied for the planar shapes case too, as the residuals are all obtained after optimal procrustes rotation.

6. REAL DATA ANALYSIS

6.1. Migration path of an elephant seal

As a simplest example, consider a dataset on the usual sphere S^2 . The dataset consists of $n = 73$ daily location measurements of a migrating female elephant seal, presented in Brillinger & Stewart (1998) and also discussed in Rivest (1999). The seal migrates from the southern California coast to the eastern mid-north Pacific Ocean. Of interest is to investigate whether the seal migrates along a great circle path, i.e. the shortest distance path. Note that Brillinger & Stewart (1998) and Rivest (1999) have analyzed this dataset in greater detail. We briefly re-analyze this data set with our hypothesis test.

Fig. 4 shows the path of the migration, including both forward journey and return trip. Since the dataset in the latitude-longitude coordinates can be converted to points on the unit sphere, it is viewed as a set of points on S^2 and we fitted PNS, with only one nested sphere (circle in this case). We fit the best fitting great circle and small circle with fitted distance $\hat{r} = 75.45^\circ$. The likelihood ratio test developed in Section 3.2 results in p-value 0.0851 (with $H_{0a} : r = 90^\circ$, the great circle). Therefore, the migration is not significantly different at the level $\alpha = 0.05$ from a great circle path, which is consistent with the results from Brillinger & Stewart (1998) and Rivest (1999).

6.2. River and sea sand grains

We consider sand grain outlines that can be parameterized as a set of points in a hypersphere. The dataset was originally analyzed in Kent et al. (2000), and consists of outlines of sand grains in two dimensional view. There are $n_1 = 25$ river and $n_2 = 24$ sea sand grains. We illustrate an application of PNS, and use of the Euclidean-type representation to test for group mean difference.

The outline of each sand grain is represented in polar coordinates (r_1, \dots, r_k) at each equally spaced angle $(\theta_1, \dots, \theta_k)$, with $k = 20$. The scale is removed so that $\sum_{i=1}^k r_i^2 = 1$. The origin for each sand grain is its center of gravity, and we keep the grains fixed in the orientation that they were recorded, rather than removing rotation as in the shape analysis of Kent et al. (2000). With θ_i fixed throughout the samples (as $\theta_i = (i - 1)2\pi/k$), $\mathbf{r} = (r_1, \dots, r_k)$ on the unit $(k-1)$ -sphere

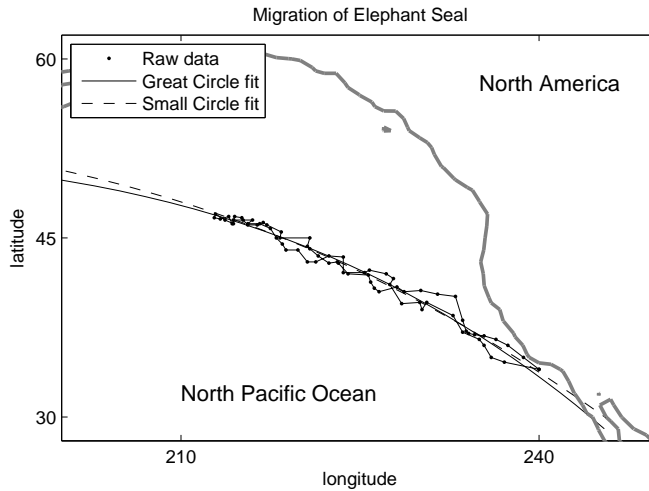


Fig. 4. Daily observations of migration path of an elephant seal, in the latitude-longitude coordinates, and the great circle and small circle fit of the data.

represents the scale invariant profile of a registered sand grain. Note that the size of river sand grains are typically larger than that of sea sand (see Kent et al. (2000)), but this analysis focuses on the variability in the scale invariant profiles of sand grains.

To the 49 ($= n_1 + n_2$) data points on the 19-sphere, we have applied the procedure of sample PNS, with significance level $\alpha = 0.05$ for every test applied. The small sphere is significant for only three layers of the procedure, when fitting A_{18} , A_{17} and A_{11} , with both p-values less than 0.05. The smallest dimension nested sphere \mathfrak{A}_1 has radius 0.8738, suggesting that the captured principal variation is not so much curved than geodesics. The PNS leads to the Euclidean-type representation X_{PNS} of the dataset, in a way that the curved principal arcs are flattened. The first three coordinates in X_{PNS} are used for visualization of major variation as in Fig. 5.

To test the group mean difference between river and sea sand grains, we can use any Euclidean space based test procedure applied to X_{PNS} . Since we do not have any prior information on the underlying distribution, it makes sense to use a nonparametric permutation test. In particular, we use the *DiProPerm* test (*Direction-Projection-Permutation*), described in Wichers et al. (2007). The test finds a direction vector pointing from one group to the other, and computes a t-statistic of the projected values onto the direction. The null distribution of the t-statistic is found by permutation of group labels. We have used the *DiProPerm* test with the *Distance Weighted Discrimination* (DWD) direction (Marron et al. (2007)). The DWD is a classification tool that separates two groups with more generalizability than e.g. the popular SVM (Vapnik (1995)). The subspace found by the DWD direction and first three coordinates of X_{PNS} is illustrated as a scatterplot matrix in Fig. 5. Although the first three coordinates of X_{PNS} do not give a visual separation between the groups, X_{PNS} turns out to be a useful Euclidean space for linear classification methods such as DWD.

DiProPerm tests the null hypothesis of equal group means. In our analysis, the test with 1000 permutations rejects the null hypothesis with p-value 0.0292. The difference of shapes in the overlay of the outlines of sand grains (Fig. 6) is statistically significant.

673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720

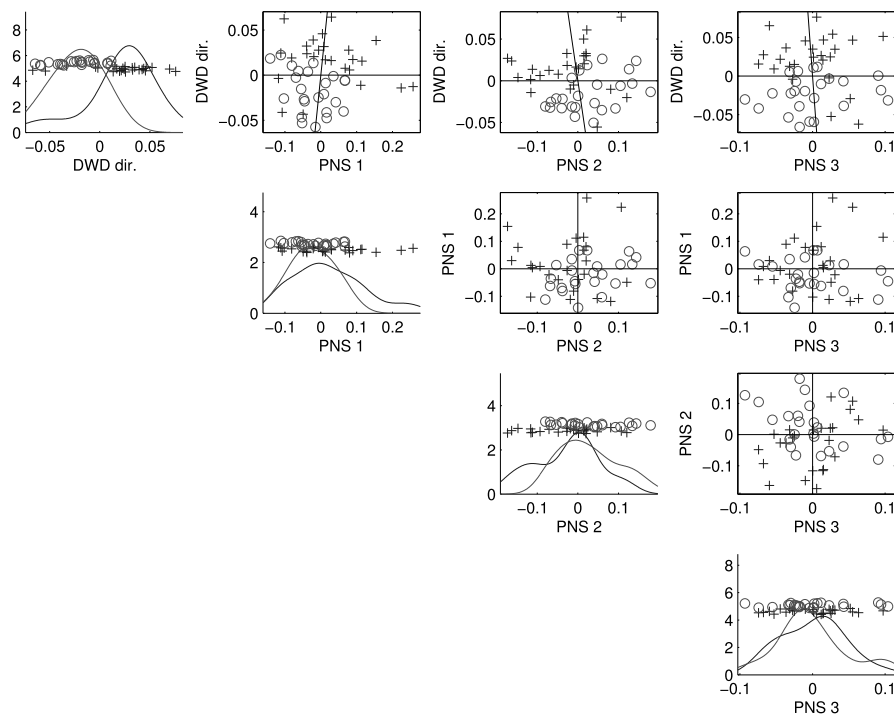


Fig. 5. Scatterplot matrix of sand grain data set, by the DWD direction and the first three coordinates of X_{PNS} . (+: river sand grains, o: sea sand grains) Diagonal entries are jitter plots of one dimensional projections with kernel density estimates for each group. The DWD direction separating the two groups is found in the Euclidean space, X_{PNS} .

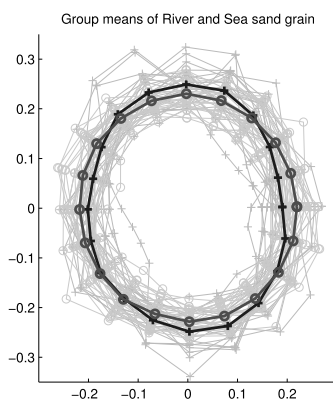


Fig. 6. Overlaid outlines of 25 river sand grains (+) and 24 sea sand grains (o) with the group means (thick outlines) identified with the geodesic mean of each group. The DiProPerm test rejects a null hypothesis of equal group means with p-value 0.0292

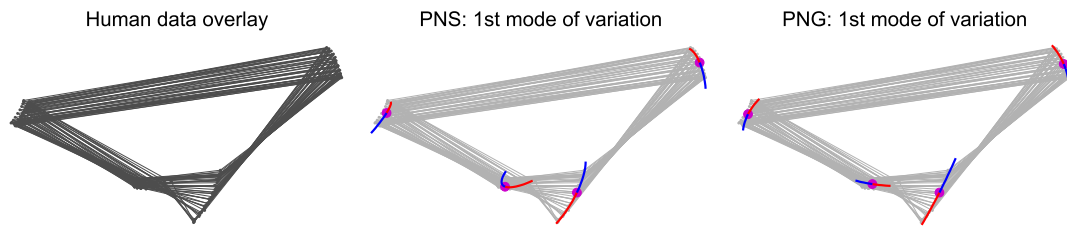


Fig. 7. (left) Procrustes fitted human data, (center) The first principal mode of variation by PNS, showing ± 2 standard deviation from PNSmean. (right) The first principal mode of variation by Principal Nested Great sphere (PNG). PNS allows to capture curving variation. In particular, the variation in the bottom left landmark is more precisely captured in PNS than PNG.

6.3. Human movement

A human movement dataset, introduced in Kume et al. (2007), contains 50 samples of $k = 4$ landmarks (lower back, shoulder, wrist, and index finger). The dataset consists of shape configurations in the plane of a table, of one person performing five different tasks, each observed at ten different time points. The raw data are plotted in Fig. 7.

In the left panel of Fig. 7, overlaid are 50 quadrilaterals, each of which is a shape configuration. Vertices of the quadrilateral are the locations of the landmarks. These 50 samples are Procrustes fitted to each other, i.e. translated, scaled, and rotated to each other so that they are as close as possible.

We have applied PNS and PNG. The fitted nested spheres of PNS have radii 1, 0.7019, 0.3967, and 0.2473 (from the 4-sphere to the 1-sphere, respectively). Note that since the dimension of the corresponding shape space is 4, the 4- d PNS is a great sphere and leaves no residuals, as expected. P-values of the sequential LRT are at most 0.0013, supporting the significance of the fitted PNS. The quadratic form of variation in the PNG coordinates is captured by the 1- d PNS, as illustrated in Fig. 1. The principal mode of variation found by PNS is plotted in Fig. 7, where the four bold dots together represent the shape of the PNSmean, and the curves through the PNSmean illustrate the shape change captured in the first PNS. The curvy form of variation apparent in the raw data are well captured.

Each task can be modeled as a 1- d arc, by applying PNS to the samples corresponding to each task. The results are plotted in Fig. 8. Each task is curving through at least three geodesic components, and is well approximated by the separately fitted PNS.

6.4. Rat skull growth

The shape and size changes of rat skulls are described in Bookstein (1991) and studied by several other authors including Kenobi et al. (2010). The data are eight landmark locations for skulls of 21 laboratory rats observed at eight ages (days 7, 14, 21, 30, 40, 60, 90, and 150). We discard 4 missing samples, and analyze the remaining 164 samples.

A non-geodesic variation curving through three geodesic components, in Fig. 9(a-b), is captured in the 2- d PNS (Fig. 9(c)). The first two principal arcs are plotted in PNG coordinates, showing the non-geodesic variation captured by PNS. The PNS coordinates capture more interesting variability in fewer components and give concise and useful representation of the data. In particular, the shape change due to the growth of the rat is well captured by the first PNS, which

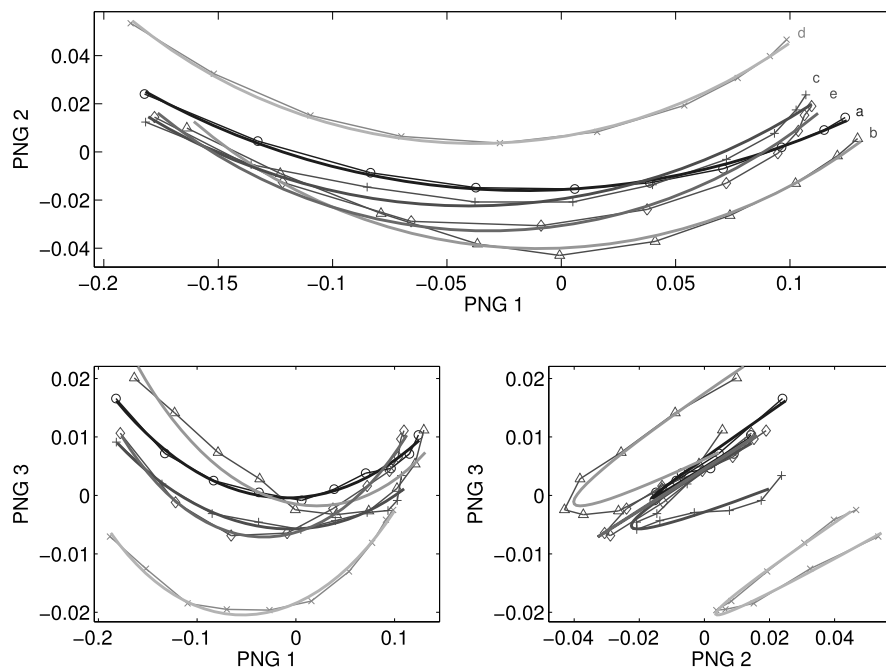


Fig. 8. Human movement data—fitted curves (solid curves) for different tasks labeled a, b, c, d, and e, plotted in PNG coordinates.

can be checked by inspecting the relation to the size of the rat skulls (Fig. 9(d)). The skull size is naturally larger for older rats. The first PNS coordinates are strongly associated with the size with the sample correlation coefficient 0.9705.

The shape change due to the first PNS is illustrated in Fig. 10. The two extreme shape configurations are also overlaid, which shows a typical effect of age.

7. DISCUSSION

A classical approach to PCA for manifold data is a tangent plane approach (see e.g. Section 5.5 of Dryden & Mardia (1998) and Fletcher et al. (2004)). The resulting principal geodesics are then passing through the point of tangency, i.e. the Procrustes mean or the geodesic mean. In their inspiring paper, Huckemann et al. (2010) pointed out that the pre-determined mean may not be a good representation of the data due to the curvature of the manifold. They find the best fitting geodesic that are not necessarily passing through any pre-determined mean. Jung et al. (2010) proposed to use small circles to fit the non-geodesic modes of variation when the sample space is essentially S^2 . Our method builds upon these earlier works, where PNS can be viewed as a high dimensional extension of Jung et al. (2010), and PNG are close to Huckemann et al. (2010).

Our method can be viewed as a backward generalization of PCA. A conventional PCA approach is to begin with lower dimension, e.g. finding the mean first, then a least squares line through the mean, then the second line which determines the best fitting plane and so on. We, on the other hand, sequentially reduce the dimensionality from the full data. These two approaches are equivalent in Euclidean space, but different on manifolds. The backward approach serves as a natural viewpoint for extension of PCA to manifolds. In fact, Huckemann et al. (2010) partly

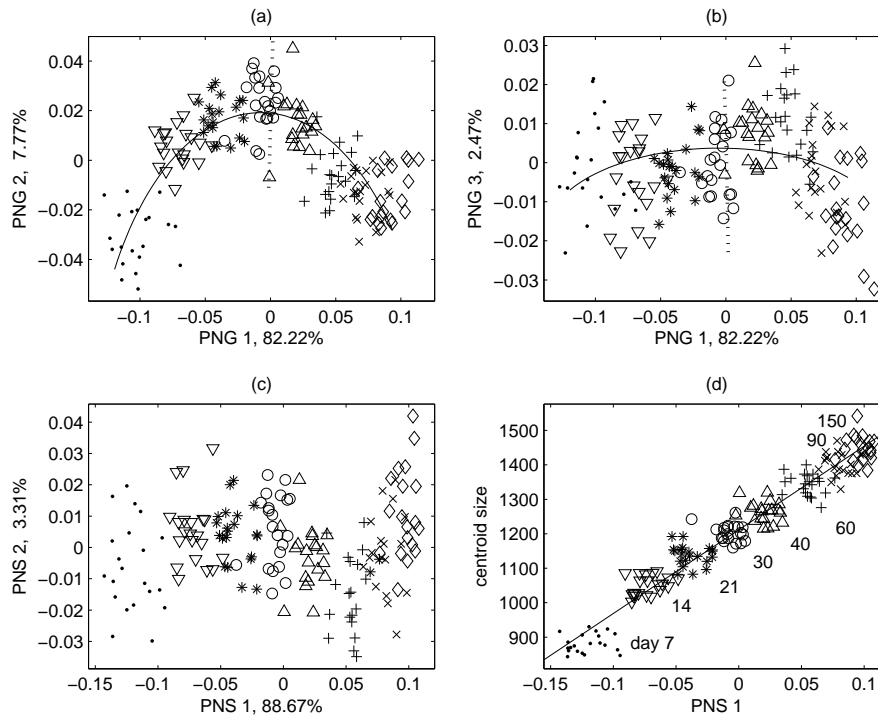


Fig. 9. Rat skull growth: (a-b) data plotted by PNG coordinates. (—) represents the first principal arc, and (· · ·) represents the second principal arc. (c) data plotted by PNS

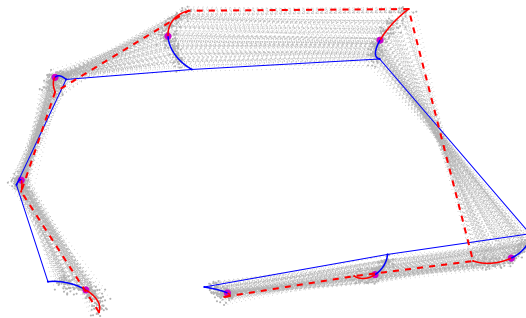


Fig. 10. Rat skull growth: the first principal mode of variation by PNS, (· · ·) represents the shape of a typical young skull (at -2 s.d.), and (—) represents the shape of a typical old rat skull (at $+2$ s.d.).

used the backward approach, since their notion of mean is post-defined by two geodesic components.

Another related area is manifold learning, see e.g. Roweis & Saul (2000) and Tenenbaum et al. (2000), or Principal curves and Principal surfaces (Hastie & Stuetzle (1989) and LeBlanc & Tibshirani (1994)). These methods find a low dimensional representation of the data in a non-parametric way, but there is no notion of *nestedness*. PNS also finds a non-linear low dimensional representation, but by a parametric form, and the findings give non-linear approximations of all dimensions.

865 Versions of PNS can be developed further. In particular, the least-squares fitting in Eq. 2 can
 866 be modified to reducing the sum of squared deviations for more robust fitting. Otherwise, instead
 867 of applying the post-hoc test procedure in Section 3, one can think of adding a penalty term to
 868 Eq. 2 and regularize the overfitting. Moreover, a modification to three or more dimensional shape
 869 space will be very useful.

870 The general idea of PNS can be used to analyze other types of data. For example, analysis
 871 of diffusion tensor images involves a sample space of the symmetric positive definite matrices
 872 (Basser et al. (1994), Pennec et al. (2006), Dryden et al. (2009)). The space of these positive def-
 873 inite matrices is a non-Euclidean, cone-shaped manifold. A tangent space approach in Fletcher
 874 & Joshi (2007) is used as an extension of PCA to this type of manifolds. A backward general-
 875 ization of PCA for diffusion tensors might give a non-geodesic way of dimension reduction. A
 876 backward approach can enhance reliability of dimension reduction when the local approximation
 877 by tangent space fails.

878 8. GEOMETRY OF NESTED SPHERES

881 Geometric properties of nested spheres are discussed in this section. Initially it will be necessary to
 882 introduce a particular type of transformation on the sphere in order to help define PNS. Specifically, we
 883 describe a rotation matrix for moving a dataset on a sphere along a particular minimal geodesic which
 884 retains the interpoint geodesic distances after the transformation. We then describe the subsphere and the
 885 sequence of nested spheres as defined in section 2.1, and discuss the geometric properties of these.

886 8.1. Preliminary Transformations: Rotation matrices

887 Suppose that \mathbf{a} and \mathbf{b} are unit vectors in \mathbb{R}^m and we wish to “move \mathbf{b} to \mathbf{a} along the geodesic path
 888 on the unit sphere in \mathbb{R}^m which connects \mathbf{b} to \mathbf{a} .” Amaral et al. (2007) showed that a rotation matrix is
 889 determined in a natural way.

890 Define $\mathbf{c} = \{\mathbf{b} - \mathbf{a}(\mathbf{a}^T \mathbf{b})\} / \|\mathbf{b} - \mathbf{a}(\mathbf{a}^T \mathbf{b})\|$, where $\|\cdot\|$ denotes the Euclidean norm on \mathbb{R}^m . Provided
 891 that $|\mathbf{a}^T \mathbf{b}| < 1$, \mathbf{c} is well defined. Let $\mathbf{A} = \mathbf{a}\mathbf{c}^T - \mathbf{c}\mathbf{a}^T$. The following lemma is proved in Amaral et al.
 892 (2007).

893 LEMMA 1. Assume that $\mathbf{a}, \mathbf{b} \in \mathbb{R}^m$ are unit vectors such that $|\mathbf{a}^T \mathbf{b}| < 1$, and let \mathbf{A} and \mathbf{c} be defined
 894 as earlier. Then for $\theta \in (0, \pi]$, the matrix

$$895 \mathbf{Q}(\theta) = \exp(\theta \mathbf{A}) = \mathbf{I}_d + \sum_{j=1}^{\infty} \frac{\theta^j}{j!} \mathbf{A}^j$$

896 has the following properties:

- 897 (a) $\mathbf{Q}(\theta)$ is an $m \times m$ rotation matrix,
 898 (b) $\mathbf{Q}(\theta)$ can be written as

$$899 \mathbf{Q}(\theta) = \mathbf{I}_d + \sin(\theta)\mathbf{A} + (\cos(\theta) - 1)(\mathbf{a}\mathbf{a}^T + \mathbf{c}\mathbf{c}^T),$$

- 900 (c) $\mathbf{Q}(\alpha)\mathbf{b} = \mathbf{a}$ for $\alpha = \cos^{-1}(\mathbf{a}^T \mathbf{b})$ and
 901 (d) for any $\mathbf{z} \in \mathbb{R}^m$ such that $\mathbf{a}^T \mathbf{z} = 0$ and $\mathbf{b}^T \mathbf{z} = 0$, we have $\mathbf{Q}\mathbf{z} = \mathbf{z}$.

902 The path of minimum length on the surface of the unit sphere in \mathbb{R}^m connecting \mathbf{b} to \mathbf{a} is given by
 903 $\{\mathbf{x}(\theta) = \mathbf{Q}(\theta)\mathbf{b} : \theta \in [0, \cos^{-1}(\mathbf{a}^T \mathbf{b})]\}$. We write this $\mathbf{Q}(\theta)$ as $\mathbf{Q}(\mathbf{b} \rightarrow \mathbf{a}, \theta)$ and denote $\mathbf{Q}(\mathbf{b} \rightarrow \mathbf{a}) \doteq$
 904 $\mathbf{Q}(\mathbf{b} \rightarrow \mathbf{a}, \cos^{-1}(\mathbf{a}^T \mathbf{b}))$ for the rotation matrix that moves \mathbf{b} to \mathbf{a} . The path defined here is indeed a
 905 minimal geodesic on the sphere. If \mathbf{b} and \mathbf{a} were orthogonal, then

$$906 \mathbf{Q}(\theta)\mathbf{b} = \cos(\theta)\mathbf{b} + \sin(\theta)\mathbf{a}, \quad -\pi/2 < \theta \leq \pi/2, \quad (\text{A1})$$

907 which corresponds to a definition of the unit speed geodesic (Kendall et al. (1999)).

We also define $\mathbf{R}(\mathbf{v})$, for $\mathbf{v} \in \mathbb{R}^m$, as a rotation matrix that rotates \mathbf{v} to the north pole $\mathbf{e}_m = (0, \dots, 0, 1)^T$, i.e.

$$\mathbf{R}(\mathbf{v}) = \mathbf{Q}(\mathbf{v} \rightarrow \mathbf{e}_m).$$

Note that the last row of $\mathbf{R}(\mathbf{v})$ is \mathbf{v}^T . If $\mathbf{v} = \mathbf{e}_m$, then $\mathbf{R}(\mathbf{v}) = \mathbf{I}_m$.

LEMMA 2. Assume that $\mathbf{a}, \mathbf{b} \in \mathbb{R}^m$ are unit vectors such that $|\mathbf{a}^T \mathbf{b}| < 1$, and let $\theta \in (0, 2\pi]$.

(a) Let \mathbf{R} be an $m \times m$ rotation matrix. Then, $\mathbf{Q}(\mathbf{R}\mathbf{b} \rightarrow \mathbf{R}\mathbf{a}, \theta) = \mathbf{R}\mathbf{Q}(\mathbf{b} \rightarrow \mathbf{a}, \theta)\mathbf{R}^T$. Equivalently, $\mathbf{Q}(\mathbf{R}^T \mathbf{b} \rightarrow \mathbf{R}^T \mathbf{a}, \theta) = \mathbf{R}^T \mathbf{Q}(\mathbf{b} \rightarrow \mathbf{a}, \theta)\mathbf{R}$.

(b) Let $\mathbf{a}' = (\mathbf{a}^T, 0)^T$, $\mathbf{b}' = (\mathbf{b}^T, 0)^T$. Then the $(m+1) \times (m+1)$ rotation matrix that moves \mathbf{b}' to \mathbf{a}' is given by

$$\mathbf{Q}(\mathbf{b}' \rightarrow \mathbf{a}', \theta) = \begin{bmatrix} \mathbf{Q}(\mathbf{b} \rightarrow \mathbf{a}, \theta) & \mathbf{0}_{m \times 1} \\ \mathbf{0}_{1 \times m} & 1 \end{bmatrix},$$

where $\mathbf{0}_{m \times n}$ is the $m \times n$ matrix of zeros.

8.2. Geometry of Subsphere

The nested spheres of S^d are lower dimensional submanifolds of S^d , each of which is isomorphic to the unit spheres in different dimensions. The subsphere A_{m-1} of S^m , $m \geq 2$, (Definition 1) induces the nested spheres. The $A_{m-1}(\mathbf{v}, r)$ is the boundary of the geodesic ball in S^m with center \mathbf{v} and radius r . The \mathbf{v} is said to be orthogonal to A_{m-1} in a sense of the following lemma.

LEMMA 3. (a) For any $\mathbf{x}, \mathbf{y} \in A_{m-1}$, $(\mathbf{x} - \mathbf{y})^T \mathbf{v} = 0$.

(b) $\mathbf{x} \in A_{m-1}$ if and only if $\mathbf{v}^T (\mathbf{x} - \cos(r)\mathbf{v}) = 0$ and $\|\mathbf{x}\| = 1$.

A subsphere A_{m-1} is essentially an $(m-1)$ dimensional sphere. The following properties of subspheres give the mathematical background to treat A_{m-1} as S^{m-1} .

PROPOSITION 1. Let $A_{m-1}(\mathbf{v}, r)$ be a subsphere in S^m . Then

(a) (A_{m-1}, ρ_m) is isomorphic to (S^{m-1}, ρ_{m-1}) with an isomorphism $f : A_{m-1} \rightarrow S^{m-1}$ defined by

$$f(\mathbf{x}) = \frac{1}{\sin(r)} \mathbf{R}^-(\mathbf{v})\mathbf{x}, \quad \mathbf{x} \in A_{m-1}$$

with inverse

$$f^{-1}(\mathbf{x}^\dagger) = \mathbf{R}^T(\mathbf{v}) \begin{bmatrix} \sin(r) \cdot \mathbf{x}^\dagger \\ \cos(r) \end{bmatrix}, \quad \mathbf{x}^\dagger \in S^{m-1},$$

where $\mathbf{R}(\mathbf{v})$ is the $(m+1) \times (m+1)$ rotation matrix that moves \mathbf{v} to the north pole, $\mathbf{R}^-(\mathbf{v})$ is the $m \times (m+1)$ matrix consisting of the first m rows of $\mathbf{R}(\mathbf{v})$.

(b) Let $\rho_{m-1}^*(\mathbf{x}, \mathbf{y}) = \sin(r)\rho_{m-1}(f(\mathbf{x}), f(\mathbf{y}))$. Then ρ_{m-1}^* is a metric on A_{m-1} .

(c) (A_{m-1}, ρ_{m-1}^*) is isometric to $(S^{m-1}, \sin(r)\rho_{m-1})$.

(d) The two metrics ρ_m and ρ_{m-1}^* are equivalent, in a sense that the following inequalities

$$\rho_m(\mathbf{x}, \mathbf{y}) \leq \rho_{m-1}^*(\mathbf{x}, \mathbf{y}) \leq \frac{\pi \sin(r)}{2r} \rho_m(\mathbf{x}, \mathbf{y})$$

hold for all $\mathbf{x}, \mathbf{y} \in A_{m-1}$ and both equalities hold if and only if $r = \pi/2$ or $\mathbf{x} = \mathbf{y}$.

(e) $\rho_{m-1}^*(\mathbf{x}, \mathbf{y}) - \rho_m(\mathbf{x}, \mathbf{y}) \leq \pi \sin(r) - 2r$ for all $\mathbf{x}, \mathbf{y} \in A_{m-1}$.

The $\rho_{m-1}^*(\mathbf{x}, \mathbf{y})$ can be interpreted as the length of a minimal arc in A_{m-1} that joins \mathbf{x}, \mathbf{y} . Precisely, the minimal arc is the image by f^{-1} of the minimal geodesic segment joining $f(\mathbf{x})$ and $f(\mathbf{y})$. Let $\mathbf{x}^\dagger = f(\mathbf{x})$, $\mathbf{y}^\dagger = f(\mathbf{y})$. Then the geodesic segment is given by

$$\Gamma = \{\gamma(\theta) = \mathbf{Q}(\mathbf{x}^\dagger \rightarrow \mathbf{y}^\dagger, \theta)\mathbf{x}^\dagger : \theta \in [0, \cos^{-1}(\mathbf{x}^{\dagger T} \mathbf{y}^\dagger)]\}.$$

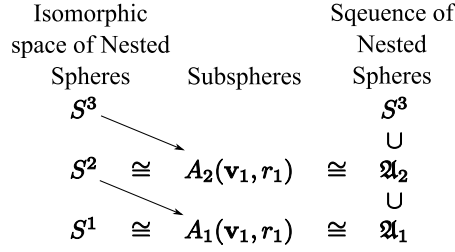


Fig. 11. Hierarchical structure of the sequence of nested spheres of the 3-sphere.

By Lemma 2, we have for any $\theta \in [0, \cos^{-1}(\mathbf{x}^\dagger T \mathbf{y}^\dagger)]$,

$$\begin{aligned}
 f^{-1}(\gamma(\theta)) &= \mathbf{R}(\mathbf{v})^T \begin{bmatrix} \sin(r) \mathbf{Q}(\mathbf{x}^\dagger \rightarrow \mathbf{y}^\dagger, \theta) \mathbf{x}^\dagger \\ \cos(r) \end{bmatrix} \\
 &= \mathbf{R}(\mathbf{v})^T \begin{bmatrix} \mathbf{Q}(\mathbf{x}^\dagger \rightarrow \mathbf{y}^\dagger, \theta) \mathbf{0}_{1 \times m} \\ \mathbf{0}_{m \times 1} & 1 \end{bmatrix} \mathbf{R}(\mathbf{v}) \mathbf{R}(\mathbf{v})^T \begin{bmatrix} \sin(r) \mathbf{x}^\dagger \\ \cos(r) \end{bmatrix} \\
 &= \mathbf{Q}(\mathbf{x}_p \rightarrow \mathbf{y}_p, \theta) \mathbf{x},
 \end{aligned}$$

where

$$\mathbf{x}_p = \mathbf{R}(\mathbf{v})^T \begin{bmatrix} \mathbf{x}^\dagger \\ 0 \end{bmatrix} = \frac{\mathbf{x} - \cos(r) \mathbf{v}}{\sin(r)},$$

and \mathbf{y}_p is defined similarly. One can check that $\rho_m(\mathbf{x}_p, \mathbf{y}_p) = \rho_{m-1}(\mathbf{x}^\dagger, \mathbf{y}^\dagger)$ and $\mathbf{Q}(\mathbf{x}_p \rightarrow \mathbf{y}_p) \mathbf{x} = \mathbf{y}$. Thus the arc $\{\mathbf{Q}(\mathbf{x}_p \rightarrow \mathbf{y}_p, \theta) \mathbf{x} : \theta \in [0, \cos^{-1}(\mathbf{x}_p^T \mathbf{y}_p)]\}$ joins \mathbf{x} to \mathbf{y} and is minimal in A_{m-1} because it is isomorphic to the minimal geodesic Γ .

The difference between ρ_m and ρ_{m-1}^* is due to the fact that the minimal arc for ρ_{m-1}^* is not a geodesic in S^m . If $r < \pi/2$, then the geodesic segment joining \mathbf{x}, \mathbf{y} is always shorter than the minimal arc in A_{m-1} . Since the difference is relatively small for close points (by Proposition 1(d-e)), this difference does not obscure much the underlying structure of the points in S^m .

8.3. Geometry of Nested Spheres

We now define a sequence of *nested spheres* $\{\mathfrak{A}_{d-1}, \mathfrak{A}_{d-2}, \dots, \mathfrak{A}_1\}$ of S^d , $d \geq 2$, with decreasing intrinsic dimensions. We first introduce a sequence of subspheres $A_{d-1}, A_{d-2}, \dots, A_1$ of S^d , which are in different spaces. The $d-1$ dimensional subsphere A_{d-1} of S^d , defined in Definition 1, is in $S^d \in \mathbb{R}^{d+1}$. The second subsphere A_{d-2} is defined from the isomorphic space S^{d-1} of A_{d-1} . Similarly, the lower dimensional subspheres are defined recursively.

DEFINITION 4. A sequence $\{A_{d-1}, A_{d-2}, \dots, A_1\}$ of subspheres is defined recursively as follows:

- (i) A_{d-1} is defined as the subsphere with $\mathbf{v}_1 \in \mathbb{R}^{d+1}$, $r \in (0, \pi/2]$ by Definition 1.
- (ii) For each $k = 2, \dots, d-1$, A_{d-k} is the subsphere defined with $\mathbf{v}_k \in \mathbb{R}^{d-k+2}$, $r_k \in (0, \pi/2]$ from S^{d-k+1} , which is isomorphic to A_{d-k+1} .

The nested spheres are defined in Definition 2, with transformations f_k (1). The geometric interpretation and hierarchical structure of the nested spheres are illustrated in Fig. 2 and 11. The nested sphere \mathfrak{A}_{d-k} can be understood as a shifted $(d-k)$ -sphere, which is orthogonal to k orthogonal directions in the sense of Lemma 4. The following properties summarize some geometric facts of the nested spheres. $\mathbf{x}_{p,k}$ in the lemma can be understood as the projection of \mathbf{x} onto the subspace that is orthogonal to $\mathbf{v}_1^*, \dots, \mathbf{v}_{d-k}^*$.

LEMMA 4. Let $\mathfrak{A}_{d-1}, \dots, \mathfrak{A}_1$ be nested spheres of S^d from a sequence of subspheres $A_{d-k}(\mathbf{v}_k, r_k)$. Then, there exists an orthogonal basis $\mathbf{v}_1^*, \dots, \mathbf{v}_{d-1}^* \in \mathbb{R}^{d+1}$ such that for each $k = 1, \dots, d-1$,

1009 (a) $(\mathbf{x} - \mathbf{y})^T \mathbf{v}_i^* = 0$ for all $i = 1, \dots, k$, $\mathbf{x}, \mathbf{y} \in \mathfrak{A}_{d-k}$,

1010 (b) $\mathbf{x} \in \mathfrak{A}_{d-k}$ if and only if $\mathbf{x}_{p,k}^T \mathbf{v}_j^* = 0$, for all $j = 1, \dots, k$, and $\|\mathbf{x}_{p,k}\| = \prod_{i=1}^k \sin(r_i)$ where

$$1011 \mathbf{x}_{p,k} = \mathbf{x} - \cos(r_1) \mathbf{v}_1^* - \sin(r_1) \cos(r_2) \mathbf{v}_2^* - \dots - \prod_{i=1}^{k-1} \sin(r_i) \cos(r_k) \mathbf{v}_k^*.$$

1014 Moreover, an explicit expression for \mathbf{v}_j^* can be obtained from $\mathbf{v}_1, \dots, \mathbf{v}_j$ as

$$1015 \mathbf{v}_j^\dagger = f_1^{-1} \circ \dots \circ f_{j-1}^{-1}(\mathbf{v}_j) \in \mathfrak{A}_{d-j+1}, \quad (\text{A2})$$

$$1016 \mathbf{v}_j^* = \prod_{i=1}^{j-1} \sin^{-1}(r_i) \{ \mathbf{v}_j^\dagger - \cos(r_1) \mathbf{v}_1^* - \sin(r_1) \cos(r_2) \mathbf{v}_2^* - \dots - \prod_{i=1}^{j-2} \sin(r_i) \cos(r_{j-1}) \mathbf{v}_{j-1}^* \} \quad (\text{A3})$$

1020 A direct consequence of this lemma is that for any nested sphere \mathfrak{A}_{d-k} of S^d can be understood as the
 1021 intersection of a hyperplane \mathcal{H}_k and S^d . The hyperplane \mathcal{H}_k is a $d - k$ dimensional affine subspace that
 1022 is orthogonal to $\mathbf{v}_1^*, \dots, \mathbf{v}_{d-k}^*$.

1024 PROPOSITION 2. Let $\mathfrak{A}_{d-1}, \dots, \mathfrak{A}_1$ be nested spheres of S^d from subspheres $A_{d-k}(\mathbf{v}_k, r_k)$. Then,

1025 (a) $\mathfrak{A}_1 \subsetneq \mathfrak{A}_2 \subsetneq \dots \subsetneq \mathfrak{A}_{d-1} \subsetneq S^d$, where $A \subsetneq B$ means that A is a proper subset of B ,

1026 (b) Let $\rho_{d-k}^*(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^k \sin(r_i) \rho_{d-k}(\mathbf{x}', \mathbf{y}')$, where $\mathbf{x}' = f_k \circ \dots \circ f_1(\mathbf{x})$. Then ρ_{d-k}^* is a metric on
 1027 \mathfrak{A}_{d-k} .

1028 (c) $(\mathfrak{A}_{d-k}, \rho_{d-k}^*)$ is isometric to $(S^{d-k}, \prod_{i=1}^k \sin(r_i) \rho_{d-k})$.

1029 The ρ_d and ρ_{d-k}^* are indeed equivalent metrics. Moreover, one can show that $\rho_{d-k}^*(\mathbf{x}, \mathbf{y})$ is the length of
 1030 a minimal arc in \mathfrak{A}_{d-k} that joins \mathbf{x} and \mathbf{y} .

1034 9. PROOFS AND ADDITIONAL LEMMAS

1035 Proofs for Appendix 8 will be given first. We then return to give proofs for the main article.

1036 *Proof of Lemma 2.* (a) Let $\mathbf{a}_0 = \mathbf{R}\mathbf{a}$, $\mathbf{b}_0 = \mathbf{R}\mathbf{b}$ and $\mathbf{c}_0 = \{\mathbf{b}_0 - \mathbf{a}_0(\mathbf{a}_0^T \mathbf{b}_0)\} / \|\mathbf{b}_0 - \mathbf{a}_0(\mathbf{a}_0^T \mathbf{b}_0)\|$.
 1037 Then $\mathbf{c}_0 = \mathbf{R}\mathbf{c}$, where $\mathbf{c} = \{\mathbf{b} - \mathbf{a}(\mathbf{a}^T \mathbf{b})\} / \|\mathbf{b} - \mathbf{a}(\mathbf{a}^T \mathbf{b})\|$, since $\mathbf{R}\mathbf{R}^T = \mathbf{R}^T \mathbf{R} = \mathbf{I}_m$. Then,

$$1038 \mathbf{Q}(\mathbf{R}\mathbf{b} \rightarrow \mathbf{R}\mathbf{a}, \theta) = \mathbf{R}\{\mathbf{I}_d + \sin(\theta)(\mathbf{a}\mathbf{c}^T - \mathbf{c}\mathbf{a}^T) + (\cos(\theta) - 1)(\mathbf{a}\mathbf{a}^T + \mathbf{c}\mathbf{c}^T)\}\mathbf{R}^T = \mathbf{R}\mathbf{Q}(\mathbf{b} \rightarrow \mathbf{a}, \theta)\mathbf{R}^T.$$

1039 (b) Let \mathbf{c}' be defined similarly for \mathbf{a}' , \mathbf{b}' . We have

$$1040 \mathbf{a}'\mathbf{a}'^T = \begin{bmatrix} \mathbf{a} \\ 0 \end{bmatrix} \begin{bmatrix} \mathbf{a}^T & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{a}\mathbf{a}^T & \mathbf{0}_{m \times 1} \\ \mathbf{0}_{1 \times m} & 0 \end{bmatrix},$$

1041 and $\mathbf{c}'\mathbf{c}'^T$, $\mathbf{a}'\mathbf{c}'^T$ and $\mathbf{c}'\mathbf{a}'^T$ can be expressed in a similar fashion. Then the expression of \mathbf{Q} in Lemma 1(b)
 1042 gives the desired result. \square

1043 *Proof of Lemma 3.* For $\mathbf{x} \in \mathbb{R}^m$ such that $\|\mathbf{x}\| = 1$, $\mathbf{x} \in A_{m-1}$ if and only if $\rho_m(\mathbf{v}, \mathbf{x}) =$
 1044 $\cos^{-1}(\mathbf{v}^T \mathbf{x}) = r$. This is equivalent to $\mathbf{v}^T \mathbf{x} - \cos(r) \mathbf{v}^T \mathbf{v} = 0$ since $\mathbf{v}^T \mathbf{v} = 1$. This proves (b). Write
 1045 $\mathbf{x} - \mathbf{y} = (\mathbf{x} - \cos(r)\mathbf{v}) - (\mathbf{y} - \cos(r)\mathbf{v})$, then the result (a) follows from (b). \square

1046 *Proof of Proposition 1.* We first show that f is a well-defined bijective function. Proofs for (b-e) will
 1047 follow. (a) is then given by (c) and (d).

1048 First note that since $\sin(r) > 0$, f is well defined. For any $\mathbf{x} \in A_{m-1}$, let $\mathbf{x}^\dagger = f(\mathbf{x})$. Then $\mathbf{x}^\dagger \in \mathbb{R}^m$,
 1049 and since $\mathbf{R}^-(\mathbf{v})^T \mathbf{R}^-(\mathbf{v}) = \mathbf{I}_{m+1} - \mathbf{v}\mathbf{v}^T$, we get

$$1050 \|\mathbf{x}^\dagger\|^2 = \frac{1}{\sin^2(r)} \|\mathbf{R}^-(\mathbf{v})\mathbf{x}\|^2 = \frac{1}{\sin^2(r)} \{\mathbf{x}^T \mathbf{x} - (\mathbf{x}^T \mathbf{v})^2\} = \frac{1}{\sin^2(r)} \{1 - \cos^2(r)\} = 1$$

1051
1052
1053
1054
1055
1056

Thus, $\mathbf{x}^\dagger \in S^{m-1}$. Conversely, for any $\mathbf{x}^\dagger \in S^{m-1}$, let $\mathbf{x} = f^{-1}(\mathbf{x}^\dagger)$. Then $\|\mathbf{x}\| = 1$ and

$$\mathbf{v}^T \mathbf{x} = (\mathbf{R}(\mathbf{v})\mathbf{v})^T \begin{bmatrix} \sin(r)\mathbf{x}^\dagger \\ \cos(r) \end{bmatrix} = \cos(r).$$

By Lemma 3(b), $\mathbf{x} \in A_{m-1}$. One can easily show that $f \circ f^{-1}(\mathbf{x}^\dagger) = \mathbf{x}^\dagger$, $f^{-1} \circ f(\mathbf{x}) = \mathbf{x}$. Therefore, f is a well defined bijective function.

Since ρ_{m-1} is a metric and $\sin(r) > 0$, the metric ρ_{m-1}^* is nonnegative and symmetric, and the triangle inequality holds. In addition, since f is bijective, we have $\rho_{m-1}^*(\mathbf{x}, \mathbf{y}) = 0$ if and only if $\mathbf{x} = \mathbf{y}$. This proves (b). With the metric ρ_{m-1}^* , f is an isometry and (c) follows.

To prove (d) and (e), the difference between two metrics for a fixed $r \in (0, \pi/2]$ is given by

$$\rho_{m-1}^*(\mathbf{x}, \mathbf{y}) - \rho_m(\mathbf{x}, \mathbf{y}) = \sin(r) \cos^{-1} \left(\frac{\cos(\rho_m(\mathbf{x}, \mathbf{y})) - \cos^2(r)}{\sin^2(r)} \right) - \rho_m(\mathbf{x}, \mathbf{y}) := h_r\{\rho_m(\mathbf{x}, \mathbf{y})\},$$

for any $\mathbf{x}, \mathbf{y} \in A_{m-1}$. Note that $\max_{\mathbf{x}, \mathbf{y}} \rho_m(\mathbf{x}, \mathbf{y}) = 2r$. Then h_r is a strictly increasing function on $[0, 2r]$ with minimum $h_r(0) = 0$ and the maximum $h_r(2r) = \pi \sin(r) - 2r$. This proves (e) and leads to the first inequality of (d). The second inequality is obtained from observing that

$$\frac{\pi \sin(r)}{2r} \rho_m(\mathbf{x}, \mathbf{y}) - \rho_{m-1}^*(\mathbf{x}, \mathbf{y})$$

is nonnegative and is zero if and only if $\rho_m(\mathbf{x}, \mathbf{y}) = 0$ or $2r$. \square

The following lemmas are useful to prove Lemma 4 and also could be of independent interest.

LEMMA 5. Let \mathbf{v}_j^\dagger and \mathbf{v}_j^* be as defined in (A2-A3). For any $\mathbf{x} \in S^d$ and $k = 1, \dots, d-1$, the following are equivalent:

- (i) $\mathbf{x} \in \mathfrak{A}_{d-k}$.
- (ii) $\mathbf{v}_k^T [f_k \circ \dots \circ f_1(\mathbf{x})] = \cos(r_k)$.
- (iii) For all $j = 1, \dots, k$,

$$\mathbf{x}^T \mathbf{v}_j^\dagger = \prod_{i=1}^{j-1} \sin^2(r_i) \cos(r_j) + \prod_{i=1}^{j-2} \sin^2(r_i) \cos^2(r_{j-1}) + \dots + \cos^2(r_1).$$

- (iv) For all $j = 1, \dots, k$,

$$\mathbf{x}^T \mathbf{v}_j^* = \prod_{i=1}^{j-1} \sin(r_i) \cos(r_j).$$

Proof of Lemma 5. [(i) \Leftrightarrow (ii)] By Definition 2 and since each f_i is bijective, $\mathbf{x} \in \mathfrak{A}_{d-k}$ is equivalent to $f_k \circ \dots \circ f_1(\mathbf{x}) \in A_{d-k}$. By Lemma 3(b), this is also equivalent to (ii).

[(i) \Leftrightarrow (iii)] First note that for any $k = 1, \dots, d-1$, for $\mathbf{y} \in S^{d-k}$,

$$\begin{aligned} f_1^{-1} \circ \dots \circ f_k^{-1}(\mathbf{y}) &= \mathbf{R}^T(\mathbf{v}_1) \begin{bmatrix} \sin(r_1)\{f_2^{-1} \circ \dots \circ f_k^{-1}(\mathbf{y})\} \\ \cos(r_1) \end{bmatrix} \\ &= [\mathbf{R}(\mathbf{v}_1, \dots, \mathbf{v}_k)]^T \begin{bmatrix} \prod_{i=1}^k \sin(r_i) \mathbf{y} \\ \prod_{i=1}^{k-1} \sin(r_i) \cos(r_{k-1}) \\ \vdots \\ \cos(r_1) \end{bmatrix}, \end{aligned}$$

where $\mathbf{R}(\mathbf{v}_1, \dots, \mathbf{v}_k)$ is a rotation matrix defined as

$$\mathbf{R}(\mathbf{v}_1, \dots, \mathbf{v}_k)^T = \mathbf{R}^T(\mathbf{v}_1) \begin{bmatrix} \mathbf{R}^T(\mathbf{v}_2) & \mathbf{0}_{d \times 1} \\ \mathbf{0}_{1 \times d} & 1 \end{bmatrix} \dots \begin{bmatrix} \mathbf{R}^T(\mathbf{v}_k) & \mathbf{0}_{(d+2-k) \times 1} \\ \mathbf{0}_{1 \times (d+2-k)} & 1 \end{bmatrix}.$$

Then

$$\begin{aligned} \mathbf{x}^T \mathbf{v}_j^\dagger &= (f_1^{-1} \circ \cdots \circ f_{j-1}^{-1} \{f_{j-1} \circ \cdots \circ f_1(\mathbf{x})\})^T f_1^{-1} \circ \cdots \circ f_{j-1}^{-1}(\mathbf{v}_j) \\ &= \prod_{i=1}^{j-1} \sin^2(r_i) \{f_{j-1} \circ \cdots \circ f_1(\mathbf{x})\}^T \mathbf{v}_j + \prod_{i=1}^{j-2} \sin^2(r_i) \cos^2(r_{j-1}) + \cdots + \cos^2(r_1) \end{aligned}$$

and the result follows from (ii).

[(i) \Rightarrow (iv)] Since $\mathbf{x} \in \mathfrak{A}_{d-k}$, we have $\mathbf{x}^T \mathbf{v}_1^* = \cos(r_1)$ by definition. Suppose $\mathbf{x}^T \mathbf{v}_j^* = \prod_{i=1}^{j-1} \sin(r_i) \cos(r_j)$ for all $j = 1, \dots, j-1$, then (iii) and canceling terms give

$$\begin{aligned} \mathbf{x}^T \mathbf{v}_j^* &= \mathbf{x}^T \left(\mathbf{v}_j^\dagger - \cos(r_1) \mathbf{v}_1^* - \cdots - \prod_{i=1}^{j-2} \sin(r_i) \cos(r_{j-1}) \mathbf{v}_{j-1}^* \right) \prod_{i=1}^{j-1} \sin^{-1}(r_i) \\ &= \prod_{i=1}^{j-1} \sin(r_i) \cos(r_j). \end{aligned}$$

Thus by induction, (iv) holds.

[(iv) \Rightarrow (iii)] Suppose (iv) holds, then for $j = 1, \dots, k$,

$$\begin{aligned} \mathbf{x}^T \mathbf{v}_j^* - \prod_{i=1}^{j-1} \sin(r_i) \cos(r_j) \\ = \prod_{i=1}^{j-1} \sin^{-1}(r_i) \left(\mathbf{x}^T \mathbf{v}_j^\dagger - \cos^2(r_1) - \cdots - \prod_{i=1}^{j-2} \sin^2(r_i) \cos^2(r_{j-1}) - \prod_{i=1}^{j-1} \sin^2(r_i) \cos(r_j) \right), \end{aligned}$$

which equals to zero if and only if (iii) holds. \square

Proof of Lemma 4. We first show that $\{\mathbf{v}_i^*; i = 1, \dots, d-1\}$ is an orthonormal basis. Note that $\mathbf{v}_1^* = \mathbf{v}_1$, and $\mathbf{v}_2^* = \sin^{-1}(r_1) \{\mathbf{v}_2^\dagger - \cos(r_1) \mathbf{v}_1^*\}$. Since $\mathbf{v}_2^\dagger \in \mathfrak{A}_{d-1}$, by Lemma 5, we have

$$\mathbf{v}_2^{*T} \mathbf{v}_1 = \sin^{-1}(r_1) \{\mathbf{v}_2^{\dagger T} \mathbf{v}_1 - \cos(r_1)\} = 0,$$

and

$$\mathbf{v}_2^{*T} \mathbf{v}_2^* = \sin^{-1}(r_1) \mathbf{v}_2^{*T} \mathbf{v}_2^\dagger = \sin^{-1}(r_1) \{\mathbf{v}_2^{\dagger T} \mathbf{v}_2^\dagger - \cos(r_1) \mathbf{v}_2^\dagger \mathbf{v}_1^*\} = 1.$$

Suppose $\mathbf{v}_i^{*T} \mathbf{v}_j^* = 0$ and $\|\mathbf{v}_i^*\| = \|\mathbf{v}_i^*\| = 1$ for $1 \leq i < j \leq k-1$. Since $\mathbf{v}_k^\dagger \in \mathfrak{A}_{d-k+1}$, by Lemma 5, we have

$$\mathbf{v}_j^{*T} \mathbf{v}_k^* = \prod_{i=1}^{k-1} \sin^{-1}(r_i) \mathbf{v}_j^{*T} \{\mathbf{v}_k^\dagger - \prod_{i=1}^{j-1} \sin(r_i) \cos(r_j) \mathbf{v}_j^*\} = 0,$$

and

$$\begin{aligned} \|\mathbf{v}_k^*\| &= \prod_{i=1}^{k-1} \sin^{-1}(r_i) \mathbf{v}_k^{*T} \mathbf{v}_k^\dagger \\ &= \prod_{i=1}^{k-1} \sin^{-2}(r_i) \mathbf{v}_k^{\dagger T} \{\mathbf{v}_k^\dagger - \cos(r_1) \mathbf{v}_1^* - \cdots - \prod_{i=1}^{k-2} \sin(r_i) \cos(r_{k-1}) \mathbf{v}_{k-1}^*\} = 1. \end{aligned}$$

Thus, by induction, $\mathbf{v}_i^*, i = 1, \dots, d-1$, are orthonormal.

Now for (b), suppose first that $\mathbf{x} \in \mathfrak{A}_{d-k}$. Then by Lemma 5, we get for all $j = 1, \dots, k$

$$\mathbf{x}_{p,k}^T \mathbf{v}_j^* = \mathbf{x}^T \mathbf{v}_j^\dagger - \prod_{i=1}^{j-1} \sin(r_i) \cos(r_j) = 0$$

and

$$\begin{aligned} \|\mathbf{x}_{p,k}\|^2 &= \mathbf{x}^T \mathbf{x}_{p,k} = \mathbf{x}^T \left\{ \mathbf{x} - \cos(r_1) \mathbf{v}_1^* - \cdots - \prod_{i=1}^{k-1} \sin(r_i) \cos(r_k) \mathbf{v}_k^* \right\} \\ &= 1 - \cos^2(r_1) - \cdots - \prod_{i=1}^{k-1} \sin^2(r_i) \cos^2(r_k). \end{aligned}$$

Thus by rearranging terms, $\|\mathbf{x}_{p,k}\| = \prod_{i=1}^k \sin(r_i)$.

Conversely, suppose that $\mathbf{x}_{p,k}^T \mathbf{v}_j^* = 0$ for all $j = 1, \dots, k$ and $\|\mathbf{x}_{p,k}\| = \prod_{i=1}^k \sin(r_i)$. Then since $\mathbf{x}_{p,k}, \mathbf{v}_1^*, \dots, \mathbf{v}_k^*$ are orthogonal to each other,

$$\begin{aligned} \|\mathbf{x}\|^2 &= \left\| \mathbf{x}_{p,k} + \cos(r_1) \mathbf{v}_1^* + \cdots + \prod_{i=1}^{k-1} \sin(r_i) \cos(r_k) \mathbf{v}_k^* \right\|^2 \\ &= \mathbf{x}_{p,k}^T \mathbf{x}_{p,k} + \cos^2(r_1) + \cdots + \prod_{i=1}^{k-1} \sin^2(r_i) \cos^2(r_k) = 1. \end{aligned}$$

One can check that for all $j = 1, \dots, k$

$$\mathbf{x}^T \mathbf{v}_j^* = \left\{ \mathbf{x}_{p,k} + \prod_{i=1}^{j-1} \sin(r_i) \cos(r_j) \mathbf{v}_j^* \right\}^T \mathbf{v}_j^* = \prod_{i=1}^{j-1} \sin(r_i) \cos(r_j),$$

and again by Lemma 5, the result follows. (a) is directly obtained from (b). \square

Proof of Proposition 2. (a) is readily derived by either Lemma 4 or the fact that $A_{m-1} \subsetneq S^m$ for all $m = 2, \dots, d$.

For (b) and (c), it can be easily checked that $f_k \circ \cdots \circ f_1 : \mathfrak{A}_{d-k} \rightarrow S^{d-k}$ is a well defined bijective function. Since ρ_{d-k} is a metric and $\sin(r_i) > 0$, the metric ρ_{d-k} is nonnegative and symmetric, and the triangle inequality holds. In addition, since f is a bijection, $\rho_{d-k}^*(x, y) = 0$ if and only if $x = y$. This proves (b). Then by the definition of ρ_{d-k}^* , $f_k \circ \cdots \circ f_1$ is an isometry and (c) follows. \square

Proof of Theorem 1. Let the singular value decomposition of the $(d+1) \times n$ data matrix $\mathbf{X} = [\mathbf{x}_1 \cdots \mathbf{x}_n]$ be

$$\mathbf{X} = \sum_{i=1}^n \lambda_i \mathbf{u}_i \mathbf{v}_i^T,$$

where λ_i s are the singular values, $\mathbf{V} = [\mathbf{v}_1 \cdots \mathbf{v}_n]$ is such that $\mathbf{V}^T \mathbf{V} = \mathbf{V} \mathbf{V}^T = \mathbf{I}_n$, and $\mathbf{U} = [\mathbf{u}_1 \cdots \mathbf{u}_n \mathbf{u}_{n+1} \cdots \mathbf{u}_{d+1}]$ is such that $\mathbf{U}^T \mathbf{U} = \mathbf{U} \mathbf{U}^T = \mathbf{I}_{d+1}$. Then $\mathcal{U} = \{\mathbf{u}_{n+1}, \dots, \mathbf{u}_{d+1}\}$ is an orthogonal basis set that complements $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$. For any $\mathbf{u} \in \mathcal{U}$, $\mathbf{u}^T \mathbf{X} = [\mathbf{u}^T \mathbf{x}_1 \cdots \mathbf{u}^T \mathbf{x}_n] = \mathbf{0}$. Therefore, $\rho(u, x_i) = \cos^{-1}(u_i^x) = \pi/2$ for all $i = 1, \dots, n$. Write the orthogonal basis \mathcal{U} as $\{v_1^*, \dots, v_{d-n+1}^*\}$ and let $r_1 = \cdots = r_{d-n+1} = \pi/2$. Then by Lemma 4(b), there exist $\mathfrak{A}_{d-1} \supset \cdots \supset \mathfrak{A}_{n-1}$ such that $\mathbf{x}_i \in \mathfrak{A}_{n-1}$. \square

Proof of Theorem 2. Note that $\mathbf{w}^T \mathbf{w}^* = \mathbf{w}^T \mathbf{M} \mathbf{w} = 0$, and for all $\mathbf{z} \in S^d$ such that $\mathbf{w}^T \mathbf{z} \geq 0$ and $\mathbf{w}^T \mathbf{M} \mathbf{z} = 0$, $\mathbf{z}^T \mathbf{w}^* = \mathbf{z}^T \mathbf{M} \mathbf{w} = \mathbf{w}^T \mathbf{M} \mathbf{z} = 0$. Thus $\rho_d(\mathbf{w}, \mathbf{w}^*) = \cos^{-1}(\mathbf{w}^T \mathbf{w}^*) = \pi/2$ and $\rho_d(\mathbf{z}, \mathbf{w}^*) = \pi/2$. Moreover, since $\mathbf{w}^T \mathbf{z} \geq 0$, we have $\mathbf{w}, \mathbf{z} \in hA_{d-1}$. \square

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. Amer. Statist. Assoc.* **102**, 695–707.
- BANERJEE, A., DHILLON, I. S., GHOSH, J. & SRA, S. (2005). Clustering on the unit hypersphere using von Mises-Fisher distributions. *Journal of Machine Learning Research* **6**, 1345–1382.

- 1201 BASSER, P. J., MATTIELLO, J. & LEBIHAN, D. (1994). MR Diffusion Tensor spectroscopy and imaging. *Biophys-*
 1202 *ical Journal* **66**, 259–267.
- 1203 BOOKSTEIN, F. L. (1991). *Morphometric tools for landmark data*. Cambridge: Cambridge University Press. Geom-
 1204 etry and biology, Reprint of the 1991 original.
- 1205 BRILLINGER, D. R. & STEWART, B. S. (1998). Elephant-seal movements: Modelling migration. *Canad. J. Statist.*
 1206 **26**, 431–443.
- 1207 BUSS, S. R. & FILLMORE, J. P. (2001). Spherical Averages and Applications to Spherical Splines and Interpolation.
 1208 *ACM Transactions on Graphics* **20**, 95–126.
- 1209 DRYDEN, I. L. (2005). Statistical analysis on high-dimensional spheres and shape spaces. *Ann. Statist.* **33**, 1643–
 1210 1665.
- 1211 DRYDEN, I. L., KOLOYDENKO, A. & ZHOU, D. (2009). Non-Euclidean statistics for covariance matrices, with
 1212 applications to diffusion tensor imaging. *Annals of Applied Statistics* **3**, 1102–1123.
- 1213 DRYDEN, I. L. & MARDIA, K. V. (1998). *Statistical shape analysis*. Wiley Series in Probability and Statistics.
 1214 Chichester: John Wiley & Sons Ltd.
- 1215 FISHER, N. I. (1993). *Statistical analysis of circular data*. Cambridge: Cambridge University Press.
- 1216 FISHER, N. I., LEWIS, T. & EMBLETON, B. J. J. (1993). *Statistical analysis of spherical data*. Cambridge: Cam-
 1217 bridge University Press. Revised reprint of the 1987 original.
- 1218 FISHER, R. (1953). Dispersion on a sphere. *Proc. Roy. Soc. London. Ser. A.* **217**, 295–305.
- 1219 FLETCHER, P. T. & JOSHI, S. (2007). Riemannian geometry for the statistical analysis of diffusion tensor data.
 1220 *Signal Processing* **87**, 250–262.
- 1221 FLETCHER, P. T., LU, C., PIZER, S. M. & JOSHI, S. (2004). Principal geodesic analysis for the study of nonlinear
 1222 statistics of shape. *IEEE Trans. on Medical Imaging* **23**, 995–1005.
- 1223 FRÉCHET, M. (1944). L'intégrale abstraite d'une fonction abstraite d'une variable abstraite et son application à la
 1224 moyenne d'un élément aléatoire de nature quelconque. *Revue Scientifique*, 483512.
- 1225 FRÉCHET, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. H. Poincaré*
 1226 **10**, 215310.
- 1227 HALL, P., MARRON, J. S. & NEEMAN, A. (2005). Geometric representation of high dimension, low sample size
 1228 data. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67**, 427–444.
- 1229 HASTIE, T. & STUETZLE, W. (1989). Principal curves. *J. Amer. Statist. Assoc.* **84**, 502–516.
- 1230 HELGASON, S. (2001). *Differential geometry, Lie groups, and symmetric spaces*, vol. 34 of *Graduate Studies in*
 1231 *Mathematics*. American Mathematical Society.
- 1232 HUCKEMANN, S., HOTZ, T. & MUNK, A. (2010). Intrinsic shape analysis: Geodesic PCA for Riemannian manifolds
 1233 modulo isometric lie group actions. *to appear in Statistica Sinica* **20**.
- 1234 HUCKEMANN, S. & ZIEZOLD, H. (2006). Principal component analysis for Riemannian manifolds, with an appli-
 1235 cation to triangular shape spaces. *Adv. in Appl. Probab.* **38**, 299–319.
- 1236 JUNG, S., FOSKEY, M. & MARRON, J. S. (2010). Principal arc analysis on direct product manifolds,. *To appear in*
 1237 *Ann. App. Statist.* .
- 1238 KARCHER, H. (1977). Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math.* **30**, 509–541.
- 1239 KENDALL, D. G. (1984). Shape manifolds, Procrustean metrics, and complex projective spaces. *Bull. London Math.*
 1240 *Soc.* **16**, 81–121.
- 1241 KENDALL, D. G., BARDEN, D., CARNE, T. K. & LE, H. (1999). *Shape and shape theory*. Wiley Series in Probabi-
 1242 lity and Statistics. Chichester: John Wiley & Sons Ltd.
- 1243 KENOBI, K., DRYDEN, I. L. & LE, H. (2010). Shape curves and geodesic modelling. *To appear in Biometrika* .
- 1244 KENT, J. T., DRYDEN, I. L. & ANDERSON, C. R. (2000). Using circulant symmetry to model featureless objects.
 1245 *Biometrika* **87**, 527–544.
- 1246 KUME, A., DRYDEN, I. L. & LE, H. (2007). Shape-space smoothing splines for planar landmark data. *Biometrika*
 1247 **94**, 513–528.
- 1248 LEBLANC, M. & TIBSHIRANI, R. (1994). Adaptive principal surfaces. *Journal of the American Statistical Associa-*
 1249 *tion* **89**, 53–64.
- 1250 MARDIA, K. V. & JUPP, P. E. (2000). *Directional statistics*. Wiley Series in Probability and Statistics. John Wiley
 1251 & Sons Ltd.
- 1252 MARRON, J. S., TODD, M. J. & AHN, J. (2007). Distance-weighted discrimination. *J. Amer. Statist. Assoc.* **102**,
 1253 1267–1271.
- 1254 PENNEC, X., FILLARD, P. & AYACHE, N. (2006). A riemannian framework for tensor computing. *International*
 1255 *Journal of Computer Vision* **66**, 41–66.
- 1256 RIVEST, L.-P. (1999). Some linear model techniques for analyzing small-circle spherical data. *Canad. J. Statist.* **27**,
 1257 623–638.
- 1258 ROWEIS, S. & SAUL, L. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science* **290**,
 1259 2323–2326.
- 1260 SCALES, L. E. (1985). *Introduction to nonlinear optimization*. New York: Springer-Verlag.
- 1261 TENENBAUM, J. B., DE SILVA, V. & LANGFORD, J. (2000). A global geometric framework for nonlinear dimen-
 1262 sionality reduction. *Science* **290**, 2319–2322.

1249	VAPNIK, V. N. (1995). <i>The Nature of Statistical Learning Theory</i> . New York: Springer.
1250	WICHERS, L., LEE, C., COSTA, D., WATKINSON, P. & MARRON, J. S. (2007). A functional data analysis approach
1251	for evaluating temporal physiologic responses to particulate matter. Tech. Rep. 5, University of North Carolina at
1252	Chapel Hill, Department of Statistics and Operations Research.
1253	
1254	
1255	
1256	
1257	
1258	
1259	
1260	
1261	
1262	
1263	
1264	
1265	
1266	
1267	
1268	
1269	
1270	
1271	
1272	
1273	
1274	
1275	
1276	
1277	
1278	
1279	
1280	
1281	
1282	
1283	
1284	
1285	
1286	
1287	
1288	
1289	
1290	
1291	
1292	
1293	
1294	
1295	
1296	