INTRINSIC SHAPE ANALYSIS:
GEODESIC PCA FOR RIEMANNIAN MANIFOLDS
MODULO ISOMETRIC LIE GROUP ACTIONS

Stephan Huckemann, Thomas Hotz and Axel Munk

Georgia Augusta Universiät Göttingen

Abstract: A general framework is laid out for principal component analysis (PCA) on quotient spaces that result from an isometric Lie group action on a complete Riemannian manifold. If the quotient is a manifold, geodesics on the quotient can be lifted to horizontal geodesics on the original manifold. Thus, PCA on a manifold quotient can be pulled back to the original manifold. In general, however, the quotient space may no longer carry a manifold structure. Still, horizontal geodesics can be well-defined in the general case. This allows for the concept of generalized geodesics and orthogonal projection on the quotient space as the key ingredients for PCA. Generalizing a result of Bhattacharya and Patrangenaru (2003), geodesic scores can be defined outside a null set. Building on that, an algorithmic method to perform PCA on quotient spaces based on generalized geodesics is developed. As a typical example where non-manifold quotients appear, this framework is applied to Kendall's shape spaces. In fact, this work has been motivated by an application occurring in forest biometry where the current method of Euclidean linear approximation is unsuitable for performing PCA. This is illustrated by a data example of individual tree stems whose Kendall shapes fall into regions of high curvature of shape space: PCs obtained by Euclidean approximation fail to reflect between-data distances and thus cannot correctly explain data variation. Similarly, for a classical archeological data set with a large spread in shape space, geodesic PCA allows new insights that have not been available under PCA by Euclidean approximation. We conclude by reporting challenges, outlooks, and possible perspectives of intrinsic shape analysis.

Key words and phrases: Extrinsic mean, forest biometry, geodesics, intrinsic mean, Lie group actions, non-linear multivariate statistics, orbifolds, orbit spaces, principal component analysis, Riemannian manifolds, shape analysis,

1. Introduction

In this paper, we illustrate a new approach for applying classical statistical methods to multivariate non-linear data. In two examples occurring in the statistical study of shape of three dimensional geometrical objects, we illustrate that the current methods of PCA by linear Euclidean approximation are unsuitable if such data in non-linear spaces fall into regions of high curvature, or if they

have a large spread. In the following we give an overview of the background of relevant previous work, and an introduction to the building blocks of our work.

Euclidean and Non-Euclidean Data. Over the last century, multivariate statistics for *Euclidean* data structures has been the target of intensive research, leading mainly to *linear* statistical methods of analysis. More recently, a growing demand can be observed for methods treating multivariate data on spaces with a natural *non-Euclidean* structure. We mention statistical estimation problems of and on manifolds as they arise in various applications, e.g., Kim and Koo (2005), estimation of manifolds, e.g., de Silva and Carlsson (2004) and Bubenik and Kim (2007), or statistical inference in shape analysis, e.g., Munk, Paige, Pang, Patrangenaru and Ruymgaart (2007), which often require a generalization of the underlying space to a quotient of a manifold, e.g., Kendall, Barden, Carne, and Le (1999), or even more general structures such as semimetrical spaces, e.g., Schmidt, Töppe, Cremers and Boykov (2007). Mostly, such data have been dealt with by linear approximations and quite some advances have been achieved. The aim of this paper is to explore the limitations of such linearizations, and to provide a methodology that may be applied when linear approximations fail to capture the non-Euclidean nature of the data. We emphasize that we do not claim to solve these issues in full generality, rather that we would like to direct the interest of the readers to this ambitious research program while we restrict our presentation to quotients of manifolds under a Lie group action.

Extrinsic, Euclidean and Intrinsic Methods. A very powerful tool of traditional Euclidian multivariate statistical analysis is *principal component analysis* (PCA). It aims to reduce the dimensionality of the data, and yields a hierarchy of major directions explaining the main sources of data variation. This raises the question of designing a similar tool for data on non-Euclidean spaces. In Table 1 we give an overview of various methods developed in the past and proposed in this paper to tackle that question. Following the idea of linearization this can be done by performing PCA in a Euclidean tangent space (whenever it exists) of an underlying space. Usually, the tangent space at an *extrinsic mean* (EM) is chosen, the latter in the manifold case being an orthogonal projection of the Euclidean mean onto the manifold in an ambient space, e.g., Hendriks, Landsman and Ruymgaart (1996), as well as Hendriks and Landsman (1998) or, in a more general case, being a *Procrustes mean*, cf., Gower (1975). Often it seems more natural to define an *intrinsic mean* (IM), i.e., a minimizer of the squared intrinsic distance to the data (Kobayashi and Nomizu (1969, p.109), and Karcher (1977)), where the intrinsic distance is usually determined by the Riemannian structure induced either by the subject matter or by the specific construction as it is the case for shape spaces (e.g., Le (2001), Bhattacharya and Patrangenaru (2003, 2005), as well as Klassen, Srivastava, Mio, and Joshi (2004)). The relationship

Table 1. Terminology and description of various approaches of PCA for non-Euclidean data. For the Euclidean methods on quotients such as Procrustes analysis, usually the tangent space of the original space with data optimally positioned w.r.t. the mean is used. For the intrinsic methods on quotients, we use generalized geodesics and submanifolds.

Term	Description
extrinsic mean (EM)	Procrustes mean or,
	projection of the Euclidean mean for manifolds
intrinsic mean (IM)	minimizer of expected squared intrinsic distance
Euclidean PCA (EPCA)	based on the empirical covariance matrix
	in a tangent space
general Procrustes analysis (GPA)	EPCA of data projected to the
	tangent space at an EM
principal geodesic analysis (PGA)	EPCA of data mapped under the inverse
	Riemann exponential to the tangent space
	at the IM
geodesic PCA (GPCA)	PCA based on minimization of intrinsic
	residual distances to geodesics
geodesic principal components (GPCs)	minimizing geodesics of GPCA
principal component mean (PM)	intersection point of first and second GPC
restricted GPCA	GPCA while requiring that all GPCs pass
	through the IM
manifold PCA (MPCA)	PCA based on non-nested submanifolds
	totally geodesic (at a point) of increasing
	dimension, determined by minimizing
	intrinsic residual distances

between EM and IM is not obvious, and not well understood. We mention that in our applications the EM is a fairly good approximation to the IM. Currently, PCA in the tangent space of either mean is performed in a *Euclidean* manner, either by some projection of the data to the tangent space at the EM, or by mapping the data under the inverse Riemann exponential map to the tangent space at the IM. The mapped data serve as the basis for computing the empirical covariance matrix, and hence the PCA. The well known *general Procrustes analysis* (GPA) for quotients, such as Kendall's shape spaces, is based on this procedure by orthogonally projecting the data to the tangent space at an EM, see Gower (1975), Goodall (1991), Cootes, Taylor, H. and Graham (1992) and Kent (1994). Alternatively in *principal geodesic analysis* (PGA), the data is mapped under the inverse Riemann exponential at the IM, see Fletcher, Lu, Pizer and Joshi (2004). In fact, intrinsic distances between data and mean are equal (under the inverse exponential) or approximately equal (in case of orthogonal projection) to the respective distances in the tangent space. When curvature is present this is not the case for between-data distances, which carry the additional information

extracted by PCA. Therefore extrinsic and Euclidean methods, as developed so far, are well suited for statistical analysis focussing on the mean, but may fail to capture the additional information required for PCA. A meaningful PCA has to take into account potential high curvature; obviously, any method relying on a Euclidean linearization of tangent space will not perform well in such cases. In this paper we develop the notion of *geodesic principal components* (GPCs) based on the intrinsic distance of data to geodesics that reflect the manifold curvature. For quotient spaces this requires the notion of generalized geodesics.

Shape Spaces. In many statistical applications, data on a sub-manifold of Euclidean or Hilbert space are considered up to an isometric smooth Lie group action. Very prominently, this is the situation in the field of statistical shape analysis: *similarity shapes* are defined modulo the group of similarity transformations, e.g., Bookstein (1978) and Kendall (1984), *affine shapes* modulo the affine group, e.g., Ambartzumian (1990, Chap. 4), and *projective shapes* modulo the general projective group, e.g., Goodall and Mardia (1999), as well as Mardia and Patrangenaru (2001). More precisely, in order to study the *shape* of a geometrical object, either a finite number of landmarks at specific locations or a bounding contour or surface is extracted and mapped to a point in a suitable Euclidean or Hilbert space. When considering similarity shapes, usually size and location information is removed by mapping onto a unit-sphere called *pre-shape-space.* Then, rotation information is removed by further mapping to elements of the quotient of the pre-shape space modulo an orthogonal group action, cf., Section 5.1. An overview of many newly developed shape space models can be found in Krim and Yezzi (2006). Earlier, finite dimensional, landmark-based shape spaces have been extensively studied; we mention the monographs of Small (1996), Dryden and Mardia (1998), as well as Kendall et al. (1999).

GPCA in Shape Analysis. One main field of application of statistical shape analysis is the study of shapes of biological entities. Methods for Kendall's landmark-based similarity shape spaces have led Le and Kume (2000) to the belief that

<center>"biological shapes evolve mainly along geodesics".</center>

In joint research with the Institute for Forest Biometry and Informatics at the University of Göttingen studying the growth of individual tree stems, the biological geodesic-hypothesis is of high interest. As we will see in Section 6.1, Euclidean PCA is not applicable since the shapes in question come to lie in a region of Kendall's shape space with unbounded curvature. This is due to the fact that shapes of tree stems are roughly degenerate long straight line segments that are invariant under rotations orthogonal to the stem, and thus correspond to

singularities of the space. Clearly, many more objects in biological research such as protein structures and cell filaments are nearly one-dimensional, whereas their shape change extends into all three spatial directions. All such shapes fall into high curvature regions of shape space rendering the current methods of Euclidean PCA unsuitable.

In addition to high curvature effects, oscillation of geodesics may cause Euclidean approximations to deviate considerably from the respective intrinsic quantities. This effect is illustrated by an example in Section 6.2 of less concentrated data in regions of lesser curvature.

It is the objective of this work to propose *geodesic principal component analysis* (GPCA) that is dependent on the intrinsic structure only, and independent of a specific linearization due to an embedding into or projection onto Euclidean space. This can then be used in general on quotient spaces

1. to carry out PCA in high curvature regions and near locations where the quotient space ceases to carry a natural manifold structure,

2. to include the effects of oscillation of geodesics for less concentrated data in PCA, and

3. as a tool for detection of curvature within a data sample.

This task faces several challenges from differential geometry, statistical theory, and numerical optimization. In this work, we introduce some key concepts and major results. Many issues are beyond the scope of this paper and leave room for further research and for discussion in Section 7. This extends in particular to numerical performance and convergence issues of the algorithms employed. In our implementation, we have used standard numerical methods to locate minima. Further research is certainly necessary to derive specific fast-converging algorithms.

Throughout this paper we assume that a random variable is given on a quotient space $Q = M/G$ that arises from the isometric action of a Lie group G on a complete Riemannian manifold M. Since we know not of any application involving non-Hausdorff quotients due to non-proper actions of groups, we assume that G is compact (which is in fact somewhat more restrictive than a proper action, cf., Section 2.2). Then the quotient carries a natural metric structure, it is even locally a manifold away from *singular* locations.

The outline of this paper is as follows. In the next section we provide some background from differential geometry: at every $p \in M$ the tangent space decomposes into a horizontal and a vertical subspace. In fact, for geodesics on M, horizontality at one point is equivalent to horizontality at all points. Calling projections of horizontal geodesics on M *generalized geodesics* on Q (as in

Kendall et al. (1999, pp.109-113) for Kendall's shape spaces) we obtain a family of curves qualifying for principal components. Orthogonal projection - the corner stone of GPCA - will be defined by lifting to the manifold. In Appendix A we show that focal points, these are points with multivalent projection, and foci form a null set on the quotient. Hastie and Stuetzle (1989, p.515) had proved this fact for one-dimensional, and Bhattacharya and Patrangenaru (2003, p.12) for arbitrary submanifolds of Euclidean space. Thus, geodesic projections on generalized geodesics are uniquely determined up to a set of measure zero. We note that *medial axes*, introduced early to shape analysis by Blum and Nagel (1978), and currently of high interest in computer vision and in shape representation, e.g., Pizer, Siddiqi, Székely, Damon, and Zucker (2003), as well as Fuchs and Scherzer (2007), are taken from foci and focal points. This section is concluded by pointing to the possible oscillating and not-everywhere minimizing nature of geodesics.

In Section 3 we elaborate on basic statistical quantities on quotients as above. Unlike for Euclidean geometry in which means, variance and principal components have several equivalent characterizations that allow for an explicit computation, in general each characterization leads to an essentially different generalization on the quotient, which in turn leads to an optimization problem that can only be solved numerically. We motivate our definition based on the minimization of residual distances. Close inspection shows that the first geodesic principal component (GPC), defined as the geodesic approximating the data best, may no longer pass through the IM, cf., Huckemann and Ziezold (2006). This fact leads to a third generalization of a mean which we call a *principal component mean* (PM); it will play a crucial role in the sections to come.

In Section 4 we lay out how to obtain sample GPCs for general quotients by pulling the numerical computation back to the manifold M. The algorithmic ansatz based on Lagrange-minimization is twofold: first computing the quotient-space distance to horizontal geodesics, thus determining *optimally positioned* data points, and second, finding all *horizontal* directions at a given data point and choosing a suitable iterate.

In Section 5 an implementable algorithm for Kendall's shape spaces is provided. Along the way we give a new and constructive proof for the fact that every singular shape can be approached by a geodesic along which some sectional curvatures are unbounded. Also, we further discuss oscillating and not-everywhere minimizing geodesics.

In the penultimate section we illustrate the effects of unbounded curvature and oscillating, not-everywhere minimizing geodesics with some exemplary 3D datasets. High curvature is encountered in the previously introduced dataset of tree stems. We find near singular shapes where

1. approximating the IM by the EM is fairly accurate; however,

2. Euclidean PCA fails to catch essential features of the shape distribution that appear under GPCA.

A classical dataset of iron age fibulae from Hodson, Sneath and Doran (1966) serves as an example for oscillating and not-everywhere minimizing geodesics in lower curvature regions. As above, for this dataset the EM and the IM are rather close to one another. Due to oscillation, however, Euclidean PCA again fails to recognize essential features only found by geodesic PCA. This gives new results characterizing the temporal evolution of shape of these iron age brooches.

Only when both ambient curvature is low and data concentration is high, as is demonstrated by a third data set of macaque skulls, is the Euclidean approximation valid.

We note that Fletcher et al. (2004) have also proposed principal component analysis for manifolds based on geodesics. However, they require GPCs to pass through the IM and compute them by Euclidean approximation only. With our method of GPCA, the restricted GPCs through the IM can be computed as well. For applications in high curvature regions this constraint makes the restricted method as unsuitable as the Euclidean approximation. It is the additional effort to determine the location of the PM that is considerably far from the EM and IM that is crucial to the success of our method of GPCA in such cases.

2. Lie Group Action, Horizontal Geodesics and Optimal Positioning

In this section we collect well-known facts from Riemannian geometry, (e.g., Abraham and Marsden (1978), Bredon (1972), Helgason (1962), and Lang (1999)) and simple consequences thereof that are necessary to introduce notation, formulate and build up our method of GPCA. We give a comprehensive introduction not found elsewhere, as these results are not easily accessable for statisticians.

Throughout this paper we consider a connected Riemannian manifold M and a Lie group G, with Lie algebra $\mathfrak{g}$ and unit element e, acting smoothly on M. The action will be denoted by $p \overset{g}{\mapsto} gp$ for $p \in M, g \in G$. We also assume for the entire paper that the action is *effective*, i.e., that for every $g \neq e$ there is a $p = p_g \in M$ with $gp \neq p$. As usual $d_M(\cdot, \cdot)$ denotes the distance on M induced by the Riemannian metric.

We remark in advance that in many recent shape space models (e.g., Krim and Yezzi (2006)) infinite-dimensional Hilbert manifolds are considered. These are limits of finite-dimensional manifolds on which numerical computations are carried out. Even though many of the following results are also true in the general case of an infinite-dimensional Banach Lie group acting on an infinite-dimensional Riemannian Hilbert manifold, we note that a cornerstone of our efforts, the

existence of geodesics of minimal length, Section 2,1 below, is false in general; a counter-example can be found in Lang (1999, pp.226-227). In the following we mention explictly if a result holds only for finite-dimensional manifolds.

2.1. Riemannian metric and projection to the quotient

Denote by $\Gamma(M)$ the space of all maximal (w.r.t. inclusion) geodesics on M. The *Hopf-Rinow Theorem* asserts that on a complete Riemannian manifold geodesics $t \to \gamma(t)$ are defined for all $t \in \mathbb{R}$. Also, if M is finite-dimensional, any two points p_1, p_2 can be joined by a geodesic of length $d(p_1, p_2)$.

The Riemannian metric is denoted as usual by $p \mapsto \langle Z_p, W_p \rangle$, $p \in M$ for $Z, W \in T(M)$. Here $T(M)$ is the module of smooth vector fields on M, and $Z_p \in T_pM$ is the value of Z in the tangent space T_pM of M at $p \in M$. $\mathrm{d}g : T_pM \to T_{gp}M$, which denotes the differential induced by the action, is an isomorphism. The action of G is called *isometric* if

$$\langle Z_p, W_p \rangle = \langle (\mathrm{d}gZ)_{gp}, (\mathrm{d}gW)_{gp} \rangle \quad \forall p \in M, \ g \in G, \ Z, W \in T(M).$$

Then,

$$\gamma \text{ geodesic} \Leftrightarrow g\gamma \text{ geodesic} \ \forall g \in G.$$

For $p \in M$ let $[p] = \{gp : g \in G\}$ be the *fiber* (or *orbit*) of p, and let $I_p = \{g \in G : gp = p\}$ the *isotropy* group at p. Then $[p]$ is a sub-manifold of M (locally an embedding, but in general not globally) that is diffeomorphic to G/I_p. G is said to be acting *freely* on M if all isotropy groups consist of the unit element only, i.e., $[p] \cong G \ \forall p \in M$.

The tangent space T_pM of M at p decomposes into a *vertical subspace* $T_p[p]$, that is the tangent space of the fiber, and an orthogonal *horizontal subspace* H_pM,

$$T_pM = T_p[p] \oplus H_pM.$$

A curve $t \mapsto \gamma(t)$ on M is called *horizontal* (*vertical*) at t_0 if its derivative there is horizontal (vertical), i.e., $\dot{\gamma}(t_0) \in H_{\gamma(t_0)}M$ ($\dot{\gamma}(t_0) \in T_{\gamma(t_0)}[\gamma(t_0)]$). Denote by $\Gamma^H(M)$ the space of all geodesics that are horizontal everywhere.

The Riemann exponential $\exp_p$ maps a sufficiently small tangent vector $v \in T_pM$ to the point $\gamma_{p,v}(1) \in M$ when $\gamma_{p,v}$ is the geodesic through $p = \gamma_{p,v}(0)$ with initial velocity $v = \dot{\gamma}_{p,v}(0)$, i.e.,

$$\exp_p(tv) := \gamma_{p,v}(t).$$

Every point p_0 on a Riemannian manifold has a *normal neighborhood* U, i.e., for all $p \in U \ \exists r_p > 0$ such that the inverse exponential $\log_p := (\exp_p)^{-1}$ is

well defined on the *geodesic ball* $B_{r_p}(p) := \exp_p(\{v \in T_pM : \|v\| < r_p\})$ and $U \subset B_{r_p}(x)$. The *Gauss Lemma* asserts that $\exp_p$-images of spheres in T_pM are orthogonal to geodesics through p.

Let

$$\pi \colon M \to M/G := \{[p] : p \in M\} \tag{2.1}$$

be the *canonical projection* to the quotient space equipped with the quotient topology. Note that π is both open and continuous. Then

$$d_{M/G}([p_1], [p_2]) := \inf_{g,h \in G} d_M(gp_1, hp_2) \quad \forall [p_1], [p_2] \in M/G$$

is a quasi-metric on M/G. In case of an isometric action we have that any geodesic segment γ joining p_1 and p_2 has the same length as the geodesic segment $g\gamma$ joining gp_1 and gp_2. Hence in case of an isometric action, $d_M(gp_1, gp_2) = d_M(p_1, p_2) \; \forall \; p_1, p_2 \in M, \; g \in G$, and thus

$$\inf_{g \in G} d_M(gp_1, p_2) = d_{M/G}([p_1], [p_2]) \quad \forall p_1, p_2 \in M.$$

2.2. The slice theorem and killing vector fields

In all applications we know of, M/G is a Hausdorff space which means that all fibers $[p]$ are closed in M. This is the case if G acts *properly* on M, i.e., if for all $p_n, p, p' \in M, \; g_n \in G, \; n \in \mathbb{N}$ with $g_n p_n \to p', \; p_n \to p$:

$$g_n \text{ has a point of accumulation } g \in G \text{ with } gp = p'.$$

A sufficient condition for a proper action is that G is compact. Even if M/G is Hausdorff the dimensions of the fibers may vary along M. Then, M/G will fail to have a natural manifold structure. This is the case for Kendall's shape spaces of three and higher-dimensional configurations. In case of a Lie group G acting isometrically and properly on a finite-dimensional manifold M, Mostov's *Slice Theorem*, cf., Palais (1960, p.108) and Palais (1961), asserts that for an open disk D about the origin in H_p, the twisted product $G \times_{I_p} D$ is diffeomorphic to a tubular neighborhood of $[p]$ in M. As a consequence,

$$I_{p'} \text{ is a subgroup of } I_p \quad \text{for } p' \in \exp_p(D). \tag{2.2}$$

Hence in case of a free action, H_pM is locally diffeomorphic to M/G at $[p]$. Then, M/G has a unique manifold structure compatible with its quotient topology (Abraham and Marsden (1978, p.266)), making the projection (2.1) a *Riemannian submersion*. Moreover then, any vector field $Z \in T(M/G)$ has a unique *horizontal lift* $\widetilde{Z} \in T(M)$, i.e., $\widetilde{Z}_p \in H_pM \; \forall \; p \in M$. Also, every smooth curve

$t \to \gamma(t)$ on M/G through $\gamma(t_0) = [p]$ has a unique horizontal lift $t \to \widetilde{\gamma}(t)$ through p.

If G is compact, then any inner product on $\mathfrak{g}$ can be extended to a bi-invariant Riemannian metric on G making all the curves $t \mapsto \mathrm{Exp}(tv)$ geodesics on G that are defined for all $t \in \mathbb{R}$. Here $v \in \mathfrak{g}$ is arbitrary and $\mathrm{Exp} : \mathfrak{g} \to G$ denotes the *Lie-exponential*. In most applications G is a transformation group and $\mathfrak{g}$ is equipped with the standard Euclidean inner product. The Lie-exponential is then the exponential function for matrices. Note that $t \to \mathrm{Exp}(tv)p$ is usually not geodesic on M.

The action of G on M gives rise to a natural mapping $\alpha : \mathfrak{g} \to T(M)$ defined by the homomorphism:

$$\alpha_p : \mathfrak{g} \to T_p[p]$$
$$v \mapsto \frac{d}{dt}\Big|_{t=0}\Big(\mathrm{Exp}(tv)p\Big)^\cdot$$

In case of an isometric action, every $\alpha(v)$ is a *Killing vector field* on M. Since the local flow of a Killing vector field X is isometric, one can show

$$\frac{d}{dt}\Big\langle X_{\gamma(t)}, \dot{\gamma}(t)\Big\rangle = 0 \tag{2.3}$$

for all geodesics $t \mapsto \gamma(t)$.

2.3. Generalized geodesics and optimal positioning

As an immediate consequence of (2.3) we have

Theorem 2.1. *Let M be a Riemannian manifold and G be a Lie group acting isometrically on M. Then a geodesic on M that is horizontal at one point is horizontal at all points.*

Due to the fact that Killing vector fields are in general not of constant modulus, a similar statement for vertical geodesics is not true. In fact, for Kendall's shape spaces of configurations of dimension $m \geq 3$, there are geodesics that are vertical at isolated points only (cf., Example 5.1).

As done in Kendall et al. (1999, pp.109-113) for Kendall's shape spaces, the concept of geodesics can thus be pushed forward also to non-manifold quotients:

Definition 2.2. Given a quotient $\pi : M \to M/G =: Q$, where M is a Riemannian manifold and G a Lie group acting isometrically on M, call a curve δ on Q a *generalized geodesic* on Q if it is the projection of a horizontal geodesic on M.

$$\Gamma(Q) := \{\delta = \pi \circ \gamma : \gamma \in \Gamma^H(M)\}$$

is the space of generalized geodesics on Q. For $\pi \circ \gamma$ we also write $[\gamma]$.

Generalized geodesics can be lifted to horizontal geodesics just as in the submersion case: for $\delta \in \Gamma(Q)$ with $\delta(0) = q$ there is, given $p \in q$, a unique lift $\gamma \in \Gamma^H(M)$ such that $\pi \circ \gamma = \delta$ and $\gamma(0) = p$. If $\eta \in \Gamma^H(M)$ is any other lift of δ, then $\exists g \in G$ such that $\eta(\cdot) = g\,\gamma(\cdot)$. Two generalized geodesics through a common point are *orthogonal there* if their lifts through one (and thus any) common point are orthogonal there.

Definition 2.3(Ziezold (1977)). Given a manifold M and a Lie group G acting isometrically on M, points $p_1, p_\in M$, and $g \in G$, call the point gp_1 in *optimal* position to p_2 if

$$d_M(gp_1, p_2) = d_{M/G}([p_1], [p_2]).$$

Also, gp is said to be in *optimal position* to a curve γ on M if

$$d_M(gp, \gamma) = d_{M/G}([p], \pi \circ \gamma).$$

If G is compact then any point can be brought into optimal position to a given point and a curve, respectively. Moreover, if gp_1 is in optimal position to p_2 then $g^{-1}p_2$ is in optimal position to p_1, and gp_1 and p_2 are called *registered*. Note that in general, optimally positioned points will not be uniquely determined. Moreover, the relation *being in optimal position* may not be transitive, see Ziezold (1977, p.602).

Theorem 2.4. *Let M be a Riemannian manifold and G a Lie group acting isometrically on M. Then any geodesic joining two points in optimal position is horizontal.*

Proof. Suppose that there is a geodesic $t \mapsto \gamma(t)$ joining $p = \gamma(0)$ and $p' = \gamma(1)$ in optimal position to each other. Then, $\log_p$ can be defined in a neighborhood U containing $\{\gamma(t) : 0 \leq t \leq 1\}$. Moreover, let $s \to \delta(s)$ be any smooth curve in $[p']$ through $p' = \delta(0)$. Then, the image of δ under $\log_p$ is a curve outside $\log_p(U) \cap \{v \in T_pM : \|v\| < d(p, p')\}$ touching at $\log_p(p')$. By the Gauss Lemma, cf., Section 2.1, the image curve of δ is hence orthogonal to the straight line through 0 and $\log_p(p')$ which is the image of γ. As δ was arbitrary, γ is thus horizontal at p', and by Theorem 2.1 it is a horizontal geodesic.

The converse, that any two points on a horizontal geodesic segment are in optimal position, is not even true in general for arbitrarily close points, cf., Theorem 5.4 (b).

As a consequence of the Hopf-Rinow Theorem (Section 2.1) and Theorem 2.4 we have the following.

Corollary 2.5. *Let M be a finite dimensional complete Riemannian manifold, and G a compact Lie group acting isometrically on M and $Q = M/G$. Then any two $q_1, q_2 \in Q$ are joined by a generalized geodesic of length $d_Q(q_1, q_2)$.*

2.4. Orthogonal projection and principal orbit theorem

On the quotient $Q = M/G$ of a complete Riemannian manifold we can thus define *orthogonal projection*: an *orthogonal projection* q_δ of $q \in Q$ onto $\delta \in \Gamma(Q)$ is the fiber $[p'_\gamma]$ of an orthogonal projection p'_γ of p' onto γ. Here $\gamma \in \Gamma^H(M)$ is an arbitrary lift of δ, $p \in q$, and $p' = gp$ is p put into optimal position with respect to γ. The orthogonal projection may be multivalued at some points (e.g., on a sphere, when projecting a pole to the equator); these form a set of measure zero as we shall see. Recall that a subset $A \subset M$ of a finite dimensional manifold has *zero measure* in M if for every local chart (u, U) of M the set $u(U \cap A)$ has Lebesgue measure zero. A set $B \subset Q$ has measure zero in Q if its lift $\pi^{-1}(B) \subset M$ has measure zero in M.

The following theorem is a consequence of Lemma A.2 and Theorem A.5 which is stated and proven in the Appendix.

Theorem 2.6. *Let G be a compact Lie group acting isometrically on a finite-dimensional Riemannian manifold M. Given a generalized geodesic δ on Q then the orthogonal projection q_δ is unique for all $q \in Q$ up to a set of measure zero.*

Call $M^* := \{p \in M : I_p = \{id\}\}$ the *regular space* (w.r.t. the quotient $M/G = Q$) and $M^o := \{p \in M : I_p \neq \{id\}\}$ the *singular space*. Since we assume an effective action, $M^* \neq \emptyset$. Below, we see that M^* is a manifold, hence by Section 2.2 the projection to the *regular quotient*

$$\pi|_{M^*} : M^* \to Q^* := M^*/G$$

is a Riemannian submersion. Some sectional curvatures of Q^* may tend to infinity when approaching a singular point, cf., Theorem 5.2. The assertion (a) of the following theorem is part of the *Principal Orbit Theorem*, cf., Bredon (1972, p.179); the assertion (b) follows from Lemma A.2 in the Appendix.

Theorem 2.7. *Let G be a compact Lie group acting isometrically and effectively on a finite-dimensional Riemannian manifold M. Then*

(a) *M^* and Q^* are open and dense in M, Q, respectively, and*

(b) *any geodesic on M that meets M^* has at most isolated points in M^o.*

2.5. Not-everywhere minimizing and oscillating geodesics

We call a generalized geodesic $\delta \in \Gamma(Q)$

(a) *everywhere-minimizing* if for all two points q_1, q_2 on δ the generalized geodesic segment of minimal length between q_1 and q_2 exists and is contained in δ,

(b) *oscillating* if there is a point $q \in Q$ such that $t \mapsto d_Q(q, \delta(t))$ has more than one strict local minimum,

(c) *recurrent* if there is a period $\tau > 0$ such that $\delta(t + \tau) = \delta(t)$ for all $t \in \mathbb{R}$,

(d) *asymptotic* if there is another generalized geodesic in $\Gamma(Q)$ which is approached asymptotically by δ.

Examples of Riemannian manifolds embedded in Euclidean space.

1. On a sphere, all geodesics are recurrent, non-oscillating and everywhere-minimizing.

2. On a proper ellipsoid, every meridional geodesic (from pole to pole) is not everywhere-minimizing, as two equatorial points on it are joined by the shorter equatorial geodesic.

3. In general, geodesics on a torus are oscillating and not-everywhere minimizing. Infinitely many oscillating geodesics are recurrent and infinitely many are dense.

4. On more complicated manifolds, say surfaces of revolution generated by a function with zero curvature at a critical point, there are non-recurrent asymptotic geodesics approaching equatorial geodesics, cf., e.g., Borzellino, Jordan-Squire, Petrics and Sullivan (2007).

If geodesics on M are not too ill-behaved then so are generalized geodesics on the quotient.

Remark 2.8. If all geodesics on M are recurrent, then all generalized geodesics on $Q = M/G$ are recurrent.

Projections of everywhere-minimizing geodesics, however, may lose this property near singularities, cf., Theorem 5.4.

3. PCA Based on Generalized Geodesics for Quotients Arising from Isometric Lie Group Actions

We first ponder different approaches to principal components on a quotient space. Then, having motivated our selection, detailed specific definitions follow.

3.1. Generalizations of PCs to non-Euclidean spaces

In a Euclidean space, principal components can be equivalently defined by minimizing the variance of the residuals or by maximizing the variance of the projections. Also, PCs are *nested* in the following sense: given a distribution in $\mathbb{R}^m$, the s-dimensional linear subspace approximating the distribution best (by minimizing sum of squared distances) is the linear space spanned by the first s

principal components. We note that the mean, which is the zero-dimensional subspace approximating best, can only be found by minimizing residuals.

In a non-Euclidean space, parametric submanifolds qualify naturally as candidates for principal components. For one-dimensional components, geodesics come into mind. Higher dimensional components would then be sub-manifolds spanned by geodesics, totally geodesic at a point, as proposed by Fletcher and Joshi (2007) and computed in approximation: eigenspaces of the respective covariance matrix in the tangent space at the IM are mapped to geodesics and submanifolds totally geodesic at the IM under the Riemann exponential.

Let us now inspect which building blocks of Euclidean PCA generalize to non-Euclidean spaces in a numerically feasible manner.

Since the straight line minimizing residual variance (the sum of squared distances) is uniquely determined in Euclidean geometry, save for special cases, we expect in a non-Euclidean geometry, also "some" uniqueness of (generalized) geodesics defined by minimizing residual variance. Nestedness of PCs based on residuals, however, cannot be expected, as the IM will in general no longer come to lie on a first (generalized) principal component geodesic, cf., Huckemann and Ziezold (2006).

Alternatively, let us consider straight lines and geodesics, respectively, maximizing the sum of squared distances of projections to a variable offset. In Euclidean geometry, save for special cases, such PCs and their offsets (the projection of the mean) are uniquely determined only up to a common translation orthogonal to the respective PC. In a non-Euclidean geometry, save for special cases due to curvature, we again expect uniqueness. In case of small curvature, however, we expect poor convergence properties for numerical algorithms. Indeed, experiments with data on unit-spheres, using algorithms based on maximizing projected variance derived along the lines of the algorithms developed below, often feature slow or no convergence at all. Even worse, on manifolds or quotients with recurrent geodesics (e.g., a great circle on a sphere or generalized geodesics on Kendall's shape spaces), the desired maximum is usually local and not global. Algorithms, moreover, may converge to the global maximum attained at an offset near the antipode of the mean. To overcome this difficulty, nestedness can be required again, cf., Fletcher and Joshi (2007). To our knowledge, it is unknown whether geodesics obtained by maximizing projection variance are nested in general. Numerical experiments on spheres hint to the contrary, that the intrinsic mean does not lie on the geodesic maximizing projected variance. It would be interesting to search for an explicit example asserting this phenomenon analytically as well.

For these reasons, we consider the *minimization of the residual variance* to be the natural approach for a non-Euclidean concept of PCA. We note that in

the manifold case, our approach locally gives manifold PCs totally geodesic at a point, even though we minimize residual variance w.r.t. each single spanning GPC individually and not to the whole manifold PC. The higher-order (≥ 2) manifold PCs are then nested again.

Since the concept of generalized geodesics for quotients extends naturally to generalized sub-manifolds (one possible definition is in Appendix A), it would be an interesting task to develop methods for non-nested residual higher-order manifold PCA as well.

3.2. Geodesic PCA based on residuals

Throughout this section, let $\pi : M \to M/G =: Q$ be the canonical projection of a complete Riemannian manifold M on which a Lie group G acts isometrically. With the induced quasi-metric $d_Q(\cdot, \cdot) = d(\cdot, \cdot)$. on Q consider, if finite,

$$E\left(d(X, q)^2\right) \quad \text{and} \tag{3.1}$$

$$E\left(d(X, \delta)^2\right) \tag{3.2}$$

for $q \in Q$, a generalized geodesic $\delta \in \Gamma(Q)$, and a Q-valued random variable X. For Kendall's shape spaces, cf., Section 5, these quantities are finite; we assume this in the following.

In applications, it is often desirable to assume that X is continuously distributed on Q with respect to the projection of Riemannian volume. If M is of finite dimension m then from any non-vanishing m-form, a Riemannian volume can be constructed. By definition, such a non-vanishing m-form exists if and only if M is orientable. If M is non-orientable, a Riemannian volume can be defined locally only. If possible then, in order to have continuity, one would assume that the support of X is contained in the projection of a subset of M which supports a non-vanishing m-form.

A point $\mu_I \in Q$ minimizing (3.1) is called an *intrinsic mean* (IM) of X with *total intrinsic variance*

$$V_{int}X := E\left(d(X, \mu_I)^2\right).$$

Due to positive sectional curvatures, the IM may not be uniquely determined. E.g., this is the case for a uniform distribution on a sphere. For this reason, Kendall shapes of two-dimensional triangles with i.i. standard multi-normally distributed landmarks have no mean, cf., Dryden and Mardia (1998, p.126). In general on manifolds, non-positive sectional curvature or sufficient concentration ensure the uniqueness of the IM. In particular, on a positive sectional curvature manifold M, if the support of a distribution of a random variable Y is contained

in a geodesic ball $B_r(p)$ for some $p \in M$, and if $B_{4r}(p)$ is contained in a normal neighborhood U on which positive sectional curvatures are bounded by $\kappa > 0$, then the condition $r < (\pi/(2\kappa))$ ensures that Y has a unique IM, see Karcher (1977) and Le (2001).

Now again, let us consider a random variable X on Q. In view of Theorem 2.7, one might be tempted to neglect the part of the distribution of X near the singularity set $Q \setminus Q^*$, in applications. However, since sectional curvatures may be unbounded when approaching the singularity set (Section 5.2), uniqueness of intrinsic means on Q cannot be expected in general, not even for concentrated distributions.

Definition 3.1. A generalized geodesic $\delta_1 \in \Gamma(Q)$ minimizing (3.2) is called a *first generalized geodesic principal component (GPC)* of X. A generalized geodesic $\delta_2 \in \Gamma(Q)$ that minimizes (3.2) over all generalized geodesics $\delta \in \Gamma(Q)$ that have at least one point in common with δ_1 and that are orthogonal to δ_1 at all points in common with δ_1 is called a *second GPC* of X.

Every point μ_P that minimizes (3.1) over all common points q of δ_1 and δ_2 is called a *principal component mean (PM)*. Given a first and a second GPC δ_1 and δ_2 with PM μ_P, a generalized geodesic δ_3 is a *third GPC* if it minimizes (3.2) over all generalized geodesics that meet δ_1 and δ_2 orthogonally at μ_P. Analogously, *GPCs of higher order* are minimizing generalized geodesics through the PM, passing orthogonally to all lower order GPCs.

One main feature of non-Euclidean geometry is the fact that in general, due to curvature, the IM will differ from the PM, cf., Huckemann and Ziezold (2006) for a detailed discussion.

Given a generalized geodesic δ of X, denote by $X^{(\delta)}$ the orthogonal projection of X onto δ. We call it the *marginal* or the *geodesic score* of X on δ. By virtue of Theorem 2.6, geodesic scores are uniquely defined up to a null set on Q. A minimizer $\mu_I^{(\delta)}$ on δ of the function $q \mapsto E(d(X^{(\delta)}, q)^2)$ on the GPC δ will be called an *intrinsic mean of X on the generalized geodesic δ*.

In order to define variance, recall that variance in Euclidean space can be obtained equivalently by considering projections or by considering residuals each of which, in non-Euclidean geometry, yield different results, however. Suppose we are given GPCs $\delta_1, \delta_2, \ldots$ with PM μ_P. Let $m \in \mathbb{N} \cup \{\infty\}$ be the dimension of M. Then, define the *geodesic variance* explained by the s-th GPC, $1 \le s \le m$, $s < \infty$, as *obtained by projection*

$$V_{proj}^{(s)} X := E\Big(d(X^{(\delta_s)}, \mu_P)^2\Big), \tag{3.3}$$

with total variance

$$V_{proj}X := \sum_{s=1}^{m} V_{proj}^{(s)}X,$$

if finite. In the finite-dimensional case $m < \infty$, we also have the *geodesic variance* explained by the s-th GPC as *obtained by residuals*

$$V_{res}^{(s)}X := E\left(\frac{1}{m-1}\sum_{j=1}^{m} d(X,\delta_j)^2 - d(X,\delta_s)^2\right),$$

with the respective total variance

$$V_{res}X := \sum_{s=1}^{m} V_{res}^{(s)}X.$$

Mixing both approaches yields (again for any $m \in \mathbb{N} \cup \{\infty\}$) the definition of *mixed geodesic variance*

$$V_{mix}X := E\left(d(X^{(\delta_1)},\mu_p^{(\delta_1)})^2\right) + E\left(d(X,\delta_1)^2\right)$$
$$= E\left(d(X^{(\delta_1)},\mu_p^{(\delta_1)})^2\right) + E\left(d(X,X^{(\delta_1)})^2\right).$$

Finally, for $m < \infty$,

$$CX := \frac{V_{proj}X - V_{res}X}{V_{int}} \tag{3.4}$$

can be taken for a measure of *curvature present in X*. In a Euclidean space we have $CX = 0$. On a positive sectional curvature manifold (which tends to pull geodesics together) we expect $CX \geq 0$, whereas on a negative sectional curvature manifold (which tends to push geodesics apart) we would have $CX \leq 0$. For a distribution that mainly follows a generalized geodesic we expect CX to be small even with high absolute sectional curvatures of the surrounding space.

4. Finding Sample GPCs: Computational Issues

In this section an algorithmic method to compute sample GPCs on a quotient $Q = M/G$ is proposed. We assume that the manifold M of finite or infinite dimension is implicitly defined by

$$M = \{x \in \mathbb{H} : \phi(x) = 0\},$$
$$T_xM = \{v \in \mathbb{H} : \mathrm{d}\phi(x)v = 0\}, \quad x \in M,$$

for a suitable smooth function $\phi : \mathbb{H} \to \mathbb{R}^n$ with $\mathrm{d}\phi(x) : \mathbb{H} \to \mathbb{R}^n$ having full rank for all $x \in M$. Here $\mathbb{H}$ denotes a suitable Euclidean or Hilbert space of dimension $> n$.

In landmark-based shape analysis, for example, usually $\mathbb{H}$ is a finite-dimensional matrix space, $n = 1$, and ϕ defines a unit-hypersphere, cf., Section 5.1. For the shape space of closed planar curves of Klassen et al. (2004) using *direction functions* θ, $\mathbb{H}$ is the space of Fourier series, $n = 3$, and

$$\phi(\theta) = \left(\int_0^{2\pi} \cos\theta(s)ds, \int_0^{2\pi} \sin\theta(s)ds, \int_0^{2\pi} \theta(s)ds \right)$$

defines a subspace of codimension 3. Here, M itself contains only shape information, the action of $G = SO(2)$ on M allows for different choices of initial points. In view of landmark-based shape analysis, this corresponds to additionally filter out cyclic relabelling of landmarks on each curve separately. For numerical feasibility, only finitely many Fourier coefficients are used.

In general, we assume that only a finite-dimensional subspace $\mathbb{H} = \mathbb{R}^d$ is considered and that $\phi : \mathbb{R}^d \to \mathbb{R}^{d-m}$, $d > m$, yields an m-dimensional manifold M.

Further, let G be a Lie group of finite dimension l acting isometrically on M. We assume that a similar representation is possible as well:

$$G = \{ g \in \mathbb{R}^f : \chi(g) = 0 \},$$

for a suitable smooth function $\chi : \mathbb{R}^f \to \mathbb{R}^{f-l}$, $f - l > 0$, and $d\chi$ of full rank on G. In all applications, G will be a compact transformation group.

In shape analysis, $\mathbb{H}$ is usually the *configuration space* (e.g., centered configurations) and M the *pre-shape space*. The function ϕ is responsible for removing size information. Sometimes, as noted above in Klassen et al. (2004), the configuration space and more aspects of shape invariance are also defined implicitly.

In our setup M is closed and thus complete. Therefore, cf., Section 2.1, maximal geodesics $t \mapsto \gamma(t)$ are defined for all $t \in \mathbb{R}$. Denote by $\langle \cdot, \cdot \rangle$ the Riemannian metric on $T_x M$ which is, in the case of an isometric embedding, the standard Euclidean inner product. By $\gamma_{x,v}$ denote the unique geodesic on M determined by $\gamma_{x,v}(0) = x, \dot{\gamma}_{x,v}(0) = v$. Here (x, v) is an element of the *tangent bundle* $TM := \cup_{x \in M} \{x\} \times T_x M$.

Furthermore, suppose that N data points $x_1, \ldots, x_N \in M$ are given that project to $[x_1], \ldots, [x_N] \in Q = M/G$. With the knowledge of the preceeding sections, finding generalized geodesics on M/G that minimize the squared distances to $[x_1], \ldots, [x_N]$ as in (3.2) is equivalent to finding horizontal geodesics on M that minimize squared distances to optimally positioned data points.

Thus, we have to develop two separate sets of algorithms: the first puts points into optimal position to points and horizontal geodesics; the second computes minimizing horizontal geodesics. Therefore, our goal is the minimization of an objective function under certain constraints. In this section we derive the

corresponding Lagrange equations from which fixed point equations can be obtained. Algorithms for concrete situations can be naturally derived from the latter, as illustrated in Section 5.

4.1. Optimally positioning

Optimally Positioning with Respect to a Point. In order to bring $x \in M$ into optimal position $g^* x$ to a given data point $y \in M$, we have to find

$$g^* = \operatorname{argmin}_{g \in G} d_M(gx, y)^2.$$

Letting $H : G \to [0, \infty) : g \to d(gx, y)^2$, we have hence to

$$\text{find} \quad g^* \in \mathbb{R}^f \quad \text{such that}$$
$$H(g^*) = \inf\{H(g) \colon g \in \mathbb{R}^f \quad \text{with} \quad \chi(g) = 0\}. \tag{4.1}$$

A standard method to solve this non-linear extremal problem under constraints consists in employing Lagrange multipliers. Every solution $g \in \mathbb{R}^f$ of (4.1) also solves

$$\mathrm{d}H + \lambda^T \mathrm{d}\chi = 0$$

for suitable $\lambda \in \mathbb{R}^{f-l}$. In some cases, such as for Kendall's similarity shape spaces, (4.1) can be solved explicitly, cf., Section 5.3.

Optimally Positioning with Respect to a Geodesic. Here we are given a data point $x \in M$ and a geodesic γ on M. In order to find $g^* \in G$ placing $g^* x$ into optimal position to γ, minimize the objective function

$$H_1(g) := d_M(gx, \gamma)^2$$

for $g \in \mathbb{R}^f$ under the constraint $\chi(g) = 0$. This will again be achieved using the method of Lagrange multipliers by solving

$$dH_1 + \lambda^T d\chi = 0$$

for $g \in \mathbb{R}^f$ and $\lambda \in \mathbb{R}^{f-l}$. Alternatively, by solving (4.1), a two-stage minimization is possible: for every t find

$$g(t) := \operatorname{argmin}_{g \in G} d_M(gx, \gamma(t))^2;$$

minimize

$$H_2(t) := d_M\Big(g(t)x, \gamma(t)\Big)$$

over t in a suitable interval such that the geodesic $t \to \gamma(t)$ is traversed once.

4.2. The vertical space at a given offset

In order to determine all horizontal geodesics it is necessary to find all horizontal directions at a given offset, i.e., all directions that are orthogonal to the vertical subspace there. To this end, we explicitly determine an orthogonal base for the vertical space $T_x[x]$ at a given point $x \in M$. Recall the homomorphism $\alpha_x : \mathfrak{g} \to T_x[x]$ from Section 2.2, suppose that $l_x = \dim(G/I_x)$, and that $e_1, \ldots, e_l$ is an arbitrary but fixed base for $\mathfrak{g}$. In general, even in case of a free action, the image of an orthogonal base in $\mathfrak{g}$ will no longer be orthogonal for $T_x[x]$. Hence determine an independent system $w_1 := \alpha_x^{-1}(v_1), \ldots, w_{l_x} := \alpha_x^{-1}(v_{l_x})$ in $\mathfrak{g}$ where $v_1, \ldots, v_{l_x}$ are obtained from $\alpha_x(e_1), \ldots, \alpha_x(e_{l_x})$ by a Gram-Schmidt ortho-normalization:

$$v_1 := \frac{\alpha_x(e_{k_1})}{\|\alpha_x(e_{k_1})\|}, \qquad \text{where } k_1 \text{ is the smallest index s.t. } \alpha_x(e_{k_1}) \neq 0$$

$$\vdots$$

$$v_{l_x} := \frac{\alpha_x(e_{k_{l_x}}) - \sum_{j=1}^{l_x-1} \langle \alpha_x(e_{k_{l_x}}), v_j \rangle v_j}{\|\alpha_x(e_{k_{l_x}}) - \sum_{j=1}^{l_x-1} \langle \alpha_x(e_{k_{l_x}}), v_j \rangle v_j\|} \quad \text{where } k_{l_x} \text{ is the smallest index s.t. } \alpha_x(e_{k_{l_x}}) \notin$$
$$\text{span}\{v_1, \ldots, v_{l_x-1}\}$$

The result will be denoted by the homomorphism $\beta_x : \mathfrak{g} \to T_x[x]$ defined by

$$\beta_x(e_{k_j}) = v_j = \alpha_x(w_j) \qquad j = 1, \ldots, l_x. \tag{4.2}$$

Furthermore, define the mapping $\psi : TM = \cup_{x \in M}\{x\} \times T_x M \to \mathbb{R}^{l_x}$ by

$$\psi(x, v) := \begin{pmatrix} \langle \beta_x(e_{k_1}), v \rangle \\ \vdots \\ \langle \beta_x(e_{k_{l_x}}), v \rangle \end{pmatrix}. \tag{4.3}$$

Then we have that a geodesic $\gamma_{x,v}$ is horizontal if and only if $\psi(x, v) = 0$.

4.3. Minimizing horizontal geodesics

We derive three different types of Lagrange equations, one for the first, one for the second, and one for all subsequent geodesics. In passing, we also give an equation for the intrinsic mean not involving the Riemann exponential function, as opposed to the algorithm of Le (2001). In order to compute the variance by projection (3.3) we also compute the intrinsic mean on a geodesic.

First Sample GPC. Define the objective function by parameterizing (3.2) with

$$F(x, v) := \sum_{i=1}^{N} d_{M/G}(\pi \circ \gamma_{x,v}, [p_i])^2 = \sum_{i=1}^{N} d_M(\gamma_{x,v}, g_i^* p_i)$$

for suitable $g_i^* \in G$, $i = 1, \ldots, N$, placing p_i into optimal position w. r. t. $\gamma_{x,v}$.

Every unit speed horizontal geodesic $\gamma_{x,v}$ on M is uniquely determined by an offset $x \in M$, an initial direction $v \in T_x M$ of unit length, i.e., $\langle v, v \rangle = 1$, and the horizontality condition $\psi(x, v) = 0$. Hence, define the constraining function

$$\Phi_1 \; : \; \mathbb{R}^d \times \mathbb{R}^d \; \rightarrow \; \mathbb{R}^{2d-2m+l_x+1}$$

$$(x, v) \mapsto \begin{pmatrix} \phi(x) \\ \mathrm{d}\phi(x)v \\ \langle v, v \rangle - 1 \\ \psi(x, v) \end{pmatrix}. \tag{4.4}$$

Finding a first GPC on M/G is thus equivalent to solving the extremal problem

$$\text{find} \;\; (x^*, v^*) \in \mathbb{R}^d \times \mathbb{R}^d \;\; \text{such that}$$
$$F(x^*, v^*) = \inf\{F(x, v) \colon x, \; v \in \mathbb{R}^d \;\; \text{with} \;\; \Phi_1(x, v) = 0\}.$$

Again, employ a Lagrange multiplier $\lambda \in \mathbb{R}^{2d-2m+l_x+1}$ and obtain from

$$\mathrm{d}F + \lambda^T \mathrm{d}\Phi_1 = 0 \tag{4.5}$$

two fixed point equations which naturally yield an algorithm to determine the solution (x^*, v^*), cf., Section 5.4.

Second Sample GPC and Sample PM. Given a horizontal lift $t \mapsto \gamma_{x,v}(t)$ of a first GPC, a suitable horizontal lift of a second GPC must pass through a point $y = \gamma_{x,v}(\tau)$ with an initial direction $w \in H_y M$ orthogonal to $\dot\gamma_{x,v}(\tau)$. Hence, with

$$\Phi_2 \; : \; \mathbb{R} \times \mathbb{R}^d \; \rightarrow \; \mathbb{R}^{d-m+l_x+2}$$

$$(\tau, w) \mapsto \begin{pmatrix} \mathrm{d}\phi\big(\gamma_{x,v}(\tau)\big)w \\ \langle \dot\gamma_{x,v}(\tau), w \rangle \\ \langle w, w \rangle - 1 \\ \psi(\gamma_{x,v}(\tau), w) \end{pmatrix}$$

and $F_2(\tau, w) := F(\gamma_{x,v}(\tau), w)$, finding a second GPC is equivalent to solving the extremal problem

$$\text{find} \;\; (\hat\tau, \hat w) \in \mathbb{R} \times \mathbb{R}^d \;\; \text{such that}$$
$$F_2(\hat\tau, \hat w) = \inf\{F_2(\tau, w) \colon \tau \in \mathbb{R}, \; w \in \mathbb{R}^d \;\; \text{with} \;\; \Phi_2(\tau, w) = 0\}.$$

This will again be achieved by the method of Lagrange multipliers by solving

$$\mathrm{d}F_2 + \lambda^T \mathrm{d}\Phi_2 = 0 \tag{4.6}$$

for $\tau \in \mathbb{R}$, $w \in \mathbb{R}^d$, and $\lambda \in \mathbb{R}^{d-m+l_x+2}$. For convenience, having found $\hat{\tau}$ and $\hat{w}$, let $v_2 := \hat{w}$ and rewrite $\gamma_{x,v}$ as $\gamma_{\hat{x},v_1}$ where $\hat{x} := \gamma_{x,v}(\hat{\tau})$ and $v_1 := \dot{\gamma}_{x,v}(\hat{\tau})$. Note that $[\hat{x}]$ is a sample PM on Q.

Higher Order Sample GPCs. All GPCs on Q of order r, $3 \leq r \leq m$, pass through the sample PM $[\hat{x}] \in Q$, i.e., each is determined only by a horizontal initial direction $v_r \in \mathbb{R}^d$ at offset $\hat{x}$. In particular, v_r is perpendicular to the horizontal lifts of all lower order GPCs at $\hat{x}$.

Suppose now that we have already found suitable $r - 1 \geq 2$ horizontal lifts $\gamma_{\hat{x},v_1}, \ldots, \gamma_{\hat{x},v_{r-1}}$ through $\hat{x}$ of GPCs on Q. Then, defining

$$\Phi_r \; : \; \mathbb{R}^d \; \rightarrow \; \mathbb{R}^{d-m+l_x+r}$$

$$v \; \mapsto \; \begin{pmatrix} \mathrm{d}\phi(\hat{x})v \\ \langle v, v_1 \rangle \\ \vdots \\ \langle v, v_{r-1} \rangle \\ \langle v, v \rangle - 1 \\ \psi(\hat{x}, v) \end{pmatrix}$$

and $F_3(v) := F(\hat{x}, v)$, finding a suitable horizontal lift of a j-th GPC is equivalent to solving the extremal problem

$$\text{find} \quad v_r \in \mathbb{R}^d \quad \text{such that}$$
$$F_3(v_r) = \inf\{F_3(v) : v \in \mathbb{R}^d \quad \text{with} \quad \Phi_r(v) = 0\}.$$

As before, this leads to the task of solving the equation

$$\mathrm{d}F_3 + \lambda^T \mathrm{d}\Phi_r = 0 \tag{4.7}$$

for $v \in \mathbb{R}^d$ and $\lambda \in \mathbb{R}^{d-m+l_x+r}$.

Sample IM. In a similar fashion, a representative $\overline{x}$ of an IM can be found. For this purpose consider

$$T(x) := \sum_{i=1}^{N} d_Q([x], [p_i])^2 = \sum_{i=1}^{N} d_M(x, h_i^* p_i)^2$$

with suitable $h_i^* \in G$ for $i = 1, \ldots, n$, putting $h_i^* p_i$ into optimal position to x. Then, finding a representative of an IM is equivalent to solving the extremal problem

$$\text{find} \quad \overline{x} \in \mathbb{R}^d \quad \text{such that}$$
$$T(\overline{x}) = \inf\{T(x) : x \in \mathbb{R}^d \quad \text{with} \quad \phi(x) = 0\}.$$

The method of Lagrange multipliers yields

$$\mathrm{d}T + \lambda^T \mathrm{d}\phi = 0$$

for $x \in \mathbb{R}^d$ and $\lambda \in \mathbb{R}^{d-m}$.

Sample IM on a GPC. Here, given a horizontal lift $t \mapsto \gamma(t) := \gamma_{x,v}(t)$ of a GPC $\pi \circ \gamma$ on Q, we want to find a representative $\overline{x}^\gamma = \gamma(\overline{t})$ of a point $[\overline{x}^\gamma]$ on $\pi \circ \gamma$ best approximating the orthogonal projections $[q_i]$ of the data points $[p_i]$ onto $\pi \circ \gamma$ $(i = 1, \ldots, N)$. Bringing p_i into optimal position $g_i^* p_i$ with respect to $\gamma_{x,v}$, observe that the orthogonal projections q_i^* of $g_i^* p_i$ onto $\gamma_{x,v}$ are representatives of $[q_i]$, $i = 1, \ldots, N$. Their intrinsic mean $\overline{x}^\gamma$ on $\gamma_{x,v}$ is obviously a representative for the intrinsic mean of $[q_i]$ $(i = 1, \ldots, N)$ on $[\gamma]$. This leads to an unconstrained extremal problem for

$$T_1(t) := \sum_{i=1}^{N} d_M(\gamma_{x,v}(t), q_i^*)^2$$

in one variable $t \in \mathbb{R}$.

5. Application: Kendall's Shape Spaces

We will now illustrate the generic method developed by explicitly determining functions H, H_1, etc., and suitable χ, ϕ and ψ for Kendall's shape spaces; those are defined in the beginning. Then we apply the method of PCA based on generalized geodesics and give explicit algorithms.

5.1. Kendall's shape space

Kendall's shape spaces are spheres in a matrix space modulo similarity transformations; in some ways they are generalizations of complex projective spaces. Denote by

$M(m, k)$ all real matrices having m rows and k columns with the Euclidean structure of $\mathbb{R}^{mk}$, i.e., the inner product $\langle a, b \rangle := \mathrm{tr}(ab^T)$, $\|a\| := \sqrt{\langle a, a \rangle}$,

$\mathfrak{gl}(m) := M(m, m)$, the Lie algebra of the general linear group $GL(m)$; the Lie exponential is then simply the matrix exponential $\mathrm{Exp}(A) = e^A$ for $A \in \mathfrak{gl}(m)$,

$O(m)$ the orthogonal group in $GL(m)$,

$\mathfrak{o}(m)$ the Lie algebra of $O(m)$, i.e., the skew symmetric matrices in $\mathfrak{gl}(m)$,

$SM(m)$ the orthogonal complement of $\mathfrak{o}(m)$, i.e., the symmetric matrices in $\mathfrak{gl}(m)$,

$SO(m) := \mathrm{Exp}(\mathfrak{o}(m))$, the special orthogonal group in $GL(m)$ of dimension $(m(m-1))/2$,

$i_m := \mathrm{diag}(1, \ldots, 1) \in SO(m)$, the identity matrix (the unit element).

Labelled landmark-based shape analysis is based on *configurations* consisting of $k \geq m + 1$ labelled vertices in $\mathbb{R}^m$ called *landmarks* that do not all coincide. A configuration is thus a matrix with k columns, each an m-dimensional landmark vector. Disregarding center and size, these configurations are mapped to the *pre-shape space* sphere

$$S_m^k := \left\{ x \in M(m, k-1) \colon \|x\| = 1 \right\}.$$

This can be done by, say, multiplying by a sub-Helmert matrix, cf., Dryden and Mardia (1998) for a detailed discussion of this and other normalization methods. The pre-shape sphere will be equipped with the natural spherical Riemannian metric, i.e., $T_x S_m^k$ is identified with the Euclidean space $\{v \in M(m, k-1) : \langle x, v \rangle = 0\}$.

In order to filter out rotation information, define on S_m^k a smooth action of $SO(m)$ by the usual matrix multiplication $S_m^k \xrightarrow{g} S_m^k : x \mapsto gx$ for $g \in SO(m)$. The orbit $[x] = \{gx \colon g \in SO(m)\}$ is the *shape* of $x \in S_m^k$. The quotient

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

is called Kendall's similarity *shape space*. Since $SO(m)$ is compact, this is a Hausdorff space, cf., Section 2.2. Horizontal and vertical subspace can be explicitly determined: $v \in H_x S_m^k$ if and only if $\mathrm{tr}(vx^T h^T) = 0 \ \forall \ h \in \mathfrak{o}(m)$, i.e.,

$$v \in H_x \Leftrightarrow vx^T \in SM(m),$$

cf., Kendall et al. (1999, p.109). The complete situation is given by

$$
\begin{array}{ccccc}
\mathfrak{gl}(m) & = & \mathfrak{o}(m) & \oplus & SM(m) \\
 & & \downarrow \cdot x & & \uparrow \cdot x^T \\
T_x S_m^k \oplus N_x S_m^k & = & T_x[x] & \oplus & \overbrace{H_x S_m^k \oplus N_x S_m^k}
\end{array}
\tag{5.1}
$$

Here, $N_x S_m^k := \{\lambda x : \lambda \in \mathbb{R}\}$ is the *normal space* of the pre-shape sphere. The first map is surjective, and for $\mathrm{rank}(x) \geq m - 1$, i.e., $I_x = \{i_m\}$, the second map is also surjective.

Note that the differential mapping of tangent spaces $\mathrm{d}g : T_s S_m^k \to T_{gs} S_m^k$ is given by

$$\mathrm{d}gv = gv, \quad v \in T_x M, \ g \in SO(m). \tag{5.2}$$

Unit speed geodesics on the pre-shape sphere are precisely the great circles

$$\gamma_{x,v}(t) := x \cos t + v \sin t \tag{5.3}$$

through an offset $x = \gamma_{x,v}(0) \in S_m^k$ with initial velocity $v = \dot{\gamma}_{x,v}(0) \in S_m^k$, $\langle x, v \rangle = 0$. For any $p, q \in S_m^k$, the spherical distance is given by

$$0 \leq d_{S_k^m}(p, q) = 2 \cdot \arcsin\left(\frac{\sqrt{\langle p - q, p - q \rangle}}{2}\right)$$

$$= \arccos\langle p, q \rangle = \arccos\left(\operatorname{tr}(pq^T)\right) \leq \pi.$$

The distance of a point $p \in S_m^k$ to the great circle $\gamma_{x,v}$ is given by

$$0 \leq d(p, \gamma_{x,v}) = \arccos\sqrt{\langle p, x \rangle^2 + \langle p, v \rangle^2} \leq \frac{\pi}{2}, \tag{5.4}$$

and the orthogonal projection of p onto $\gamma_{x,v}$ by

$$\frac{\langle x, p \rangle x + \langle v, p \rangle v}{\sqrt{\langle x, p \rangle^2 + \langle v, p \rangle^2}}.$$

Let us now return to the action of $SO(m)$ on S_k^m, cf., Kendall et al. (1999) for a detailed discussion.

In case of $m = 1$ the action is trivial, i.e., $\Sigma_1^k \cong S_1^k$.

In case of $m = 2$ the action of $SO(m)$ on S_2^k is just the scalar action of $SO(2) \cong S^1 \subset \mathbb{C}$ on the $(2k-3)$ - dimensional pre-shape sphere naturally embedded in complex vector space

$$S_2^k \cong S^{2k-3} \subset \mathbb{C}^{k-1}.$$

The quotient map is then the well known *Hopf fibration*, leading to complex projective space of dimension $k - 2$:

$$\Sigma_2^k \cong S^{2k-3}/S^1 = P\mathbb{C}^{k-2}.$$

In case of $m \geq 3$ a pre-shape $s \in S_m^k$ with $0 < \operatorname{rank}(s) = r < m - 1$ will be invariant under some rotation group, a non-trivial isotropy group of dimension $r - m - 1$. For this reason, cf., Section 2.2, the shape spaces Σ_m^k ($m \geq 3$) have no natural manifold structure.

Rotating a pre-shape $p' \in S_m^k$ into optimal position to a given pre-shape $p \in S_m^k$ can be accomplished via *pseudo singular value decomposition*

$$p'p^T = u\mu v^T$$

where $u, v \in SO(m)$ and $\mu = \operatorname{diag}(\mu_1, \ldots, \mu_n)$ with $\mu_1 \geq \ldots \mu_{m-1} \geq |\mu_m| \geq 0$. Then

$$g^* := vu^T$$

puts p' into optimal position g^*p' to p (e.g., Kendall et al. (1999, p.114)). We note that the rotation g^*, and thus g^*p', is uniquely determined up to a set of

measure zero. More precisely: only in case of p', p regular and $\mu_{n+1} + \mu_n > 0$ is the rotation g^* uniquely determined (cf., Kendall et al. (1999, p.121)).

5.2. The generalized geodesics of Kendall's shape spaces

Recall Theorem 2.7 to note that $SO(m)$ acts freely on the open *regular pre-shape sphere*

$$S^{*k}_m := \left\{ x \in S^k_m : \operatorname{rank}(x) \geq m - 1 \right\},$$

which is open and dense in S^k_m, making the projection to *regular shape space*

$$S^{*k}_m \to \Sigma^{*k}_m := S^{*k}_m / SO(m) \subset \Sigma^k_m$$

a Riemannian submersion. Generalized geodesics in shape space restricted to regular shape space are geodesics in the usual sense. In shape space, a generalized geodesic through a regular shape $[x] \in \Sigma^{*k}_m$ is either a single geodesic in Σ^{*k}_m, or the union of geodesics in Σ^{*k}_m and isolated singular shapes in $\Sigma^k_m \setminus \Sigma^{*k}_m$ (cf., Lemma A.2. in the Appendix). For planar shape spaces Σ^k_2, the fibers of S^k_2 are spanned by single vertical geodesics. In general this is not the case.

Example 5.1. For $3 \leq m < k$, a geodesic may be vertical only at an isolated point. Consider

$$x = \frac{1}{\sqrt{m}}(i_m | 0), \quad v = \frac{1}{\sqrt{2}}(w | 0) \in S^k_m \quad \text{with} \quad w = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ \hline 0 & 0 \end{pmatrix} \in \mathfrak{o}(m),$$

and the geodesic $t \mapsto \gamma_{x,v}(t) = x \cos t + v \sin t$. By (5.1), this is vertical if and only if $\exists w_t \in \mathfrak{o}(m)$ with $v \cos t - x \sin t = w_t(x \cos t + v \sin t)$. $w_0 = w$ yields verticality at $t = 0$. For $0 < t < \pi$ however, verticality would imply $-\sqrt{2} i_m = \sqrt{m} w_t w$, and this is impossible.

As $[x] \in \Sigma^{*k}_m$ tends to $[y] \in \Sigma^k_m \setminus \Sigma^{*k}_m$, some sectional curvatures at $[x]$ tend to infinity, see Kendall et al. (1999, pp.149-156). We give a new, short, and constructive proof for this fact.

Theorem 5.2. *In shape space Σ^k_m, $k > m > 2$, every singular shape can be approached from regular shape space Σ^{*k}_m by a generalized geodesic along which some sectional curvatures are unbounded.*

Proof. In the situation of a Riemannian submersion $S^{*k}_m \to \Sigma^{*k}_m$ we have O'Neill's formula (Lang (1999, p.393)) for the respective curvatures and, in particular, in case of sectional curvatures at $[x] \in \Sigma^{*k}_m$ of any two orthonormal vector

fields $X, Y \in T(\Sigma^{*k}_{m})$,

$$\text{curv}_{\Sigma^*}(X,Y)_{[x]} = \text{curv}_{S^*}(\widetilde{X},\widetilde{Y})_x + \frac{3}{4}\sum_{1\leq r<l\leq m}\left\langle V_{rl},[\widetilde{X},\widetilde{Y}]\right\rangle^2_x$$

$$= 1 + \frac{3}{4}\sum_{1\leq r<l\leq m}\left\langle V_{rl},[\widetilde{X},\widetilde{Y}]\right\rangle^2_x. \tag{5.5}$$

Here $\widetilde{X}$ and $\widetilde{Y}$ denote the horizontal lifts as in Section 2.2, $[\cdot,\cdot]$ denotes the *Lie bracket* and the V_{rl} $(1 \leq r < l \leq m)$ constitute a base system for $T([x])$ orthonormal at x. For any vector fields V, G, H we have the well-known (Lang (1999, pp. 126-127))

$$\langle V,[G,H]\rangle = \omega[G,H]$$
$$= G\langle V,H\rangle - H\langle V,G\rangle - 2d\omega(G,H),$$

where ω is the one-form dual (w.r.t. the Riemannian structure) to V, and $d\omega$ denotes its exterior derivative. The above reduces to

$$\langle V,[G,H]\rangle = -2d\omega(G,H)$$

if V is vertical and G, H are horizontal. In view of (5.5), in order to prove the theorem it suffices thus to provide on S^{*m}_{k} for

(a) a horizontal geodesic $x(t) = p\cos t + v\sin t$, $x(t) \in S^{*m}_{k}$ for $0 < t < \pi$, $p \in S^{m}_{k} \setminus S^{*m}_{k}$, and

(b) unit length vector fields V, G, H such that V is vertical and G, H are horizontal along $x(t)$ such that

(c) $\lim_{t\to 0} d\omega(G,H) \to \infty$ for the dual ω of V.

In fact it suffices to give an example for Σ^{*4}_{3}, as this can be embedded isometrically in all higher dimensional shape spaces, cf., Kendall et al. (1999, p.29). For any singular shape $[p] \in \Sigma^{4}_{3} \setminus \Sigma^{*4}_{3}$ all landmarks are on a single line segment, hence we pick w.l.o.g. a pre-shape representative of the form

$$p = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \alpha & \beta & \gamma \end{pmatrix},$$

with $0 \neq \gamma$, $\alpha^2 + \beta^2 + \gamma^2 = 1$. In the following, in order to verify the respective properties "horizontal" and "vertical" we make repeated use of the decomposition

(5.1). First note that

$$v = \frac{1}{\sqrt{\alpha^2 + \gamma^2}} \begin{pmatrix} \gamma & 0 & -\alpha \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

is a unit length horizontal vector at p, hence $x(t) = p \cos t + v \sin t$ satisfies the requirement (a). Requirement (b) is met by the unit length vertical field

$$V = \frac{1}{\sqrt{\sum_{j=1}^{k-1}(x_{1j}^2 + x_{2j}^2)}} \sum_{r=1}^{k-1}(x_{1r}\partial_{2r} - x_{2r}\partial_{1r})$$

and the two constant vector fields

$$G = a\partial_{11} + \partial_{12} + b\partial_{13}, \quad H = a\partial_{21} + \partial_{22} + b\partial_{23}, \quad a = -\frac{\beta\alpha}{\alpha^2 + \gamma^2}, \quad b = -\frac{\beta\gamma}{\alpha^2 + \gamma^2},$$

horizontal along $x(t)$. In order to verify (c), consider the exterior derivative of the dual to V:

$$d\omega = -\frac{\sum_{j=1}^{k-1}(x_{1j}d^{1j} + x_{2j}d^{2j})}{\sqrt{\sum_{j=1}^{k-1}(x_{1j}^2 + x_{2j}^2)}^3} \wedge \sum_{r=1}^{k-1}(x_{1r}d^{2r} - x_{2r}d^{1r}) + \frac{2\sum_{r=1}^{k-1}(d^{1r} \wedge d^{2r})}{\sqrt{\sum_{j=1}^{k-1}(x_{1j}^2 + x_{2j}^2)}}.$$

Along $x(t)$ we have

$$d\omega = \frac{1}{(\alpha^2 + \gamma^2)^{3/2} \sin t} \Big((2\alpha^2 + \gamma^2)d^{11} \wedge d^{21} + \alpha\gamma(d^{11} \wedge d^{23} + d^{13} \wedge d^{21})$$
$$+ 2(\alpha^2 + \gamma^2)d^{12} \wedge d^{22} + (2\gamma^2 + \alpha^2)d^{13} \wedge d^{23} \Big),$$

yielding, as required,

$$d\omega(G, H) = \frac{1}{(\alpha^2 + \gamma^2)^{3/2} \sin t} \to \infty.$$

We note that V as introduced above is, in case of dimension $m = 2$, the only (up to the sign) vertical unit length vector field. The exterior derivative of its dual is then simply

$$d\omega = 2\sum_{j=1}^{k-1} d^{1j} \wedge d^{2j},$$

and hence $0 \leq |d\omega(G, H)| \leq 1$ for any unit length horizontal fields G, H. Then (5.5) yields the sectional curvature of the complex projective spaces: Σ_2^k has for

$k = 3$ constant sectional curvature 4, whereas for $k \geq 4$, the sectional curvatures assume all values between 1 and 4. Observe that only the *complex* curvature of Σ_2^k is constantly 4 also for $k \geq 4$.

In order to determine the space of (generalized) geodesics $\Gamma(\Sigma_m^k)$, introduce for any $m, k \in \mathbb{N}$, $k > m$,

$$O_2(m, k) := \quad \{(e_1, e_2) \in M(m, k) \times M(m, k) \colon \langle e_i, e_j \rangle = \delta_{ij}, \, 1 \leq i, \, j \leq 2\}, \text{ an}$$
orthonormal Stiefel manifold of dimension $2mk - 3$,

$$O_2^H(m, k) := \quad \{(e_1, e_2) \in O_2(m, k) \colon e_2 e_1^T \in SM(m)\} \text{ a sub-manifold of dimension } [2mk - 3 - (m(m-1))/2].$$

We thus have surjective mappings $O_2(m, k - 1) \rightarrow \Gamma(S_m^k)$, $O_2^H(m, k - 1) \rightarrow \Gamma^H(S_m^k) : (x, v) \mapsto \gamma_{x,v}$. Under the action of $O(2)$ from the right, given by

$$(e_1, e_2) \begin{pmatrix} a & -b \\ \varepsilon b & \varepsilon a \end{pmatrix} = (ae_1 + \varepsilon be_2, -be_1 + \varepsilon ae_2), \quad a^2 + b^2 = 1 = \varepsilon^2,$$

pairs defining the same great circle are mapped onto each other. This action is free on both $O_2(m, k - 1)$ and $O_2^H(m, k - 1)$, hence we can identify $\Gamma(S_m^k)$ with the Grassmannian $G_2(m, k - 1) := O_2(m, k - 1)/O(2)$, and $\Gamma^H(S_m^k)$ with the sub-manifold

$$G^H(m, k - 1) := O_2^H(m, k - 1)/O(2)$$

of dimension $2m(k - 1) - 4 - m(m - 1)/2$. On $O_2^H(m, k - 1)$ there is also a free action of $SO(m)$ from the left, defined component-wise,

$$g(e_1, e_2) = (ge_1, ge_2),$$

that commutes with the right action of $O(2)$. (5.2) and (5.3) imply that if $(x, v) \in O_2^H(m, k - 1)$ determines a horizontal geodesic $\gamma_{x,v}$ on S_m^k projecting to a generalized geodesic δ on Σ_m^k, then the horizontal geodesic determined by $(gx, gv) \in O_2^H(m, k - 1)$ for given $g \in SO(m)$ projects to the same δ. For even dimensions m, $\{id, -id\}$ is contained in every isotropy group on $G^H(m, k - 1)$. Choosing suitably a regular pre-shape e_1 and a singular pre-shape e_2, we see that on $G^H(m, k - 1)$ we have an effective action, of $SO(m)/\{id, -id\}$ for m even, and of $SO(m)$ for m odd, respectively. Furthermore, the action is free for $m = 2$. Hence by Sections 2.2 and 2.4, we have the following.

Theorem 5.3. *The space of all (generalized) geodesics on Kendall's shape space* Σ_m^k *can be given the structure of the canonical quotient*

$$\Gamma(\Sigma_m^k) \cong G_2^H(m, k - 1)/SO(m),$$

with manifold part of dimension $2m(k - 1) - 4 - m(m - 1)$. *For* $m = 2$, $\Gamma(\Sigma_2^k)$ *is a manifold.*

We now turn to the properties of (generalized) geodesics described in Section 2.5. Not all the pathological cases may occur on Kendall's shape spaces.

Theorem 5.4. *The following hold:*

(a) *for $m < k$, all (generalized) geodesics on Kendall's shape spaces Σ_m^k are recurrent;*

(b) *any generalized geodesic $t \to \delta(t)$ on Σ_m^k, $2 < m < k$, with horizontal lift of the form $\gamma : t \to x \cos t + v \sin t$ with $x \in S_m^k \setminus S_m^{*k}$, $v \in S_m^{*k}$, and $\delta(t) = [\gamma(t)] \neq [\gamma(-t)] = \delta(-t)$, is not everywhere-minimizing in any neighborhood of $[x]$;*

(c) *on Kendall's planar shape spaces Σ_2^k ($k \geq 3$), all geodesics are everywhere-minimizing as well as non-oscillating.*

Proof. The first assertion (a) is a consequence of the fact that (generalized) geodesics on Σ_m^k are projections of horizontal great circles which are recurrent on the pre-shape sphere, cf., Remark 2.8.

For the second assertion (b), consider $y = x \cos t + v \sin t$ and $z = x \cos t - v \sin t$, $[y] \neq [z]$, for $t \neq 0$ arbitrary small. Then it suffices to show that z is not in optimal position w.r.t. y, since then $d_{\Sigma_m^k}([y], [z]) < 2t$ and δ is not everywhere-minimizing. W.l.o.g. we may assume that

$$
x = \begin{pmatrix} \alpha_1^T \\ \vdots \\ \alpha_m^T \end{pmatrix}, \quad v = \begin{pmatrix} \beta_1^T \\ \vdots \\ \beta_m^T \end{pmatrix},
$$

with $\alpha_i \neq 0$ and $\alpha_i^T \beta_j = \alpha_j^T \beta_i$ for $1 \leq i, j \leq r \leq m-2$, $\alpha_i = 0$ for $i = r+1, \ldots, m$, and with $\beta_1, \ldots, \beta_{m-1}$ non-vanishing. Then $y z^T$ is symmetric with diagonal vector

$$
\begin{pmatrix} \|\alpha_1\|^2 \cos^2 t - \|\beta_1\|^2 \sin^2 t \\ \vdots \\ \|\alpha_r\|^2 \cos^2 t - \|\beta_r\|^2 \sin^2 t \\ -\|\beta_{r+1}\|^2 \sin^2 t \\ \vdots \\ -\|\beta_m\|^2 \sin^2 t \end{pmatrix}.
$$

By hypothesis, for all $0 < |t| < \pi$, at least one of the entries is negative, hence y and z cannot be in optimal position w.r.t. one another.

To prove assertion (c), let $t \mapsto \gamma(t) = x \cos t + v \sin t$ be a horizontal great circle on the pre-shape sphere $S_2^k \ni x, v$. In order to see that $[\gamma]$ is everywhere-minimizing, it suffices to show that $d_{\Sigma_2^k}([x], [p]) = t$ for $p = xc + vs$, $c = \cos t$, $s = \sin t$, and $t \in (0, \pi/2)$. With the notation of Huckemann and Hotz (2009), the fiber $[x]$ of x is given by the points $\alpha x + \beta i x \in S_2^k$, $\alpha^2 + \beta^2 = 1$ (ix denotes a distinct unit vector spanning the vertical space at x). Then,

$$d_{\Sigma_2^k}([x], [p]) = \min_{\alpha^2+\beta^2=1} d_{S_2^k}(\alpha x + \beta i x, cx + sv) = \min_{\alpha^2+\beta^2=1} \arccos(\alpha c) = t.$$

From this, and the fact that $[\gamma_{x,v}(t)] = [-\gamma_{x,v}(t)] = [\gamma_{x,v}(t+\pi)]$, we infer at once that $[\gamma]$ is of length π.

To see that the projection of γ to Σ_2^k is non-oscillating consider, for arbitrary $p \in S_2^k$,

$$\left(\cos \left(d_{\Sigma_2^k}([p], [\gamma(t)]) \right) \right)^2$$

$$= \max_{\alpha^2+\beta^2=1} \left(\cos \left(d_{S_2^k}(\alpha p + \beta i p, x \cos t + v \sin t) \right) \right)^2$$

$$= \max_{\alpha^2+\beta^2=1} \left(\alpha \Big(\langle x, p \rangle \cos t + \langle v, p \rangle \sin t \Big) + \beta \Big(\langle x, ip \rangle \cos t + \langle v, ip \rangle \sin t \Big) \right)^2$$

$$= \Big(\langle x, p \rangle \cos t + \langle v, p \rangle \sin t \Big)^2 + \Big(\langle x, ip \rangle \cos t + \langle v, ip \rangle \sin t \Big)^2,$$

which is either constantly zero or a π-periodic function.

Remark 5.5. Numerical examples (cf., also Section 6.2) for $3 \leq m < k$ show that generalized geodesics on Σ_m^k are usually oscillating and not-everywhere minimizing.

As a consequence of Theorem 5.4, intrinsic means feature some similarities to intrinsic means on a cone.

Corollary 5.6. *The intrinsic sample mean $\hat{\mu}_I$ of two non-degenerate shapes $[y]$ and $[z]$, with sufficiently close mirrored locations to a degenerate shape $[x]$, is closer to $[x]$ but unequal to $[x]$:*

$$0 < d_{\Sigma_m^k}(\hat{\mu}_I, [x]) < d_{\Sigma_m^k}([z], [x]) = d_{\Sigma_m^k}([y], [x]).$$

Proof. W.l.o.g we use the notation of y and z as in the proof of (b) in Theorem

5.4. Then for $0 < t = d_{\Sigma_m^k}([z],[x]) = d_{\Sigma_m^k}([y],[x])$ sufficiently small,

$$
gz = \begin{pmatrix}
\alpha_1^T \cos t - \beta_1^T \sin t \\
\vdots \\
\gamma_r^T \cos t - \beta_r \sin t \\
\beta_{r+1}^T \sin t \\
\vdots \\
\beta_{m-1}^T \sin t \\
\epsilon \beta_m^T \sin t
\end{pmatrix}
$$

with $\epsilon = (-1)^{m-r-1}$, is z brought into optimal position w.r.t. y. Since, for two configurations only, the extrinsic sample mean coincides with the intrinsic sample mean, a pre-shape $w \in \hat{\mu}_I$ is given by

$$
\frac{y + gz}{\|y + gz\|} = \frac{1}{\sqrt{\cos^2 t + \beta^2 \sin^2 t}}
\begin{pmatrix}
\alpha_1^T \cos t \\
\vdots \\
\gamma_r^T \cos t \\
\beta_{r+1}^T \sin t \\
\vdots \\
\beta_{m-1}^T \sin t \\
\frac{1+\epsilon}{2} \beta_m^T \sin t
\end{pmatrix}
$$

with indeed

$$
0 < d_{\Sigma_m^k}(\hat{\mu}_I, [x]) = \arccos \frac{\cos t}{\sqrt{\cos^2 t + \beta^2 \sin^2 t}} < t,
$$

where, by hypothesis,

$$
0 < \beta^2 := \|\beta_{r+1}\|^2 + \cdots + \|\beta_{m-1}\|^2 + \frac{1+\epsilon}{2}\|\beta_m\|^2 < 1.
$$

5.3. Optimally positioning w.r.t. horizontal great circles

The maximal shape distance of any two shapes $[p], [p'] \in \Sigma_m^k$ is given by $\pi/2$. Incidentally this is also the maximal possible distance of a shape to a generalized geodesic as the example

$$
p = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad
x = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad
v = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}
$$

teaches: $d_{\Sigma_2^4}([p], [\gamma_{x,v}(t)]) = \pi/2$ is constant.

Now, given a pre-shape $p \in S_m^k$ and a unit speed horizontal geodesic $\gamma_{x,v}$ on S_m^k, we adapt the methods of Section 4.1 to obtain $g^* \in SO(m)$ putting g^*p into optimal position to $\gamma_{x,v}$. According to (5.4), the objective function to be minimized is given by

$$H_1(g) = \arccos \sqrt{\left(\operatorname{tr}(gpx^T) \right)^2 + \left(\operatorname{tr}(gpv^T) \right)^2}.$$

Equivalently, the simpler

$$\widetilde{H}_1(g) := \left(\operatorname{tr}(gpx^T) \right)^2 + \left(\operatorname{tr}(gpv^T) \right)^2$$

will be maximized. For odd dimensions m we may equivalently maximize over $O(m)$, since $\widetilde{H}_1(g) = \widetilde{H}_1(-g)$ and for $g \in O(m)$ either g or $-g \in SO(m)$. Alternatively, with $g(t)$ in optimal position to $\gamma_{x,v}(t)$,

$$\widetilde{H}_2(t) := \operatorname{tr}\left(g(t)px^T \right) \cos(t) + \operatorname{tr}\left(g(t)pv^T \right) \sin(t)$$

can be maximized over $[0, 2\pi)$ for odd m, and over $[0, \pi)$ for even m. Since

$$\widetilde{H}_2(t) \leq \left(\cos d_S(g(t)p, \gamma_{x,v}(t)) \right)^2 = \widetilde{H}_1\left(g(t) \right),$$

we obtain at once that if g^*p is in optimal position to $\gamma_{x,v}$ then it is also in optimal position to its orthogonal projection onto the geodesic.

Theorem 5.7. *Given $p \in S_m^k$ and a horizontal geodesic $\gamma_{x,v}$ on S_m^k, let $g^* = g(t^*) \in SO(m)$ put p in optimal position g^*p to $\gamma_{x,v}$, as well as in optimal position to $\gamma_{x,v}(t^*)$ for some $t^* \in [0, 2\pi)$. Then*

$$\tan(t^*) = \frac{\operatorname{tr}(g^*pv^T)}{\operatorname{tr}(g^*px^T)}.$$

The corresponding algorithm alternates between orthogonal projection and pairwise optimally positioning.

Basic algorithm to put p into optimal position to a horizontal great circle $\gamma_{x,v}$:

Starting with a suitable $t^{(0)}$ compute, for $n \geq 0$,

$$g^{(n+1)} := uv^T,$$

where u and v are from a pseudo singular value decomposition of $p\gamma_{x,v}(t^{(n)})^T = u\mu v^T$, and update

$$t^{(n+1)} := \arctan \frac{\operatorname{tr}(g^{(n+1)}pv^T)}{\operatorname{tr}(g^{(n+1)}px^T)}.$$

In general $\widetilde{H}_2$ has several local maxima that can be accessed by choosing different suitable starting values $t^{(0)}$. As correct optimal positioning is crucial to the validity of the algorithms for computing GPCs, diligent care has to be taken to obtain a global maximum.

In order to jump out of local maxima we propose to add another optimization step to the algorithm.

Diagonal-optimization algorithm to put p into optimal position to a horizontal great circle $\gamma_{x,v}$:

Starting with a suitable $t^{(0)}$ compute, for $n \geq 0$,

$$g^{(n+1)} := u\epsilon^* v^T$$

where, as before, u and v are from a pseudo singular value decomposition of $p\gamma_{x,v}(t^{(n)}) = u\mu v^T$ and

$$\epsilon^* \in E_m = \Big\{\epsilon = \mathrm{diag}(\epsilon_1, \ldots, \epsilon_m) : \det(\epsilon) = 1, \epsilon_j \in \{-1, 1\}$$
$$(j = 1, \ldots, m)\Big\}$$

is chosen to minimize $d_S(u\epsilon v^T p, \gamma_{x,v})$. Then set

$$t^{(n+1)} := \arctan \frac{\mathrm{tr}(g^{(n+1)} pv^T)}{\mathrm{tr}(g^{(n+1)} px^T)}.$$

Note that there are four elements in E_3 and, in general, $\sum_{0 \leq j < m/2} \binom{m}{2j}$ elements in E_m.

Numerical experiments show that the diagonal-optimization algorithm converges to the global maximum in the majority of cases. Further research is necessary to develop a faster and more reliable method.

5.4. Algorithms for GPCA and means for Kendall's shape spaces

In this section we compute *sample* GPCs and *sample* means algorithmically from a data sample. For brevity we omit the prefix "sample" in the following.

Throughout this section suppose we have N m-dimensional configurations, each with k landmarks. We map these to the pre-shapes $p_1, \ldots, p_N \in S_m^k$. For most experimental situations we may assume with probability 1 that the isotropy groups in question are trivial, i.e., $I_x = \{i_m\}$ at pre-shapes x where iterations are performed, cf., Theorem 2.7. Hence, the dimension of the fiber is maximal: $l_x = l$. In the general case of varying l_x, the dimension of the vertical space will have to be computed anew for every iteration step.

Picking an arbitrary but fixed base $e_1, \ldots, e_l$, $l = (m(m-1))/2$, for $\mathfrak{o}(m)$ note that $\alpha_x(v) = vx$ for the mapping defined in Section 2.2 Defining β_x as in (4.2), obtain a base $w_1 := w_1(x), \ldots, w_l := w_l(x)$ for $\mathfrak{o}(m)$ mapping to an orthogonal base $w_1 x, \ldots, w_l x$ for the vertical space $T_x[x]$. With these, define the function ψ as in (4.3). For every offset $x = x_0$ the base $w_1, \ldots, w_l$ will be computed anew. When taking derivatives with respect to x at an offset x_0 we assume $w_1, \ldots, w_l$ to be constant (still being a base in an open neighborhood of x_0) and mapping to an orthogonal base at offset x_0.

The First GPC. Recall (5.4) to observe that the objective function from (4.5) is given by

$$F(x,v) := \sum_{j=1}^{N} d_{S_m^k}(q_j, \gamma_{(x,v)}) = \sum_{j=1}^{N} \arccos^2 \sqrt{\langle x, q_j \rangle^2 + \langle v, q_j \rangle^2}$$

for $x, v \in M(m, k-1)$, where $q_j \in [p_j]$ are in optimal position to $\gamma_{x,v}$. The constraining function (4.4) can be taken as

$$\Phi_1(x,v) = \begin{pmatrix} \langle x, x \rangle - 1 \\ 2\langle x, v \rangle \\ \langle v, v \rangle - 1 \\ 2\langle w_1 x, v \rangle \\ \vdots \\ 2\langle w_l x, v \rangle \end{pmatrix}.$$

Abbreviate $\zeta_j := \sqrt{\langle x, q_j \rangle^2 + \langle v, q_j \rangle^2}$, $\xi_j := (\arccos \zeta_j)/(\zeta_j \sqrt{1 - \zeta_j^2})$, and let $\xi :=$ 1 for $\zeta = 1$. We assume that the optimally positioned data q_j do not have the maximal distance $\pi/2$ to $\gamma_{x,v}$, which means that $\zeta_i \neq 0$. Computing (4.5) we have, with a Lagrange multiplier $\lambda = (\lambda_1, \ldots, \lambda_{l+3})$, that

$$\sum_{i=1}^{N} \xi_i \langle x, q_i \rangle q_i = \lambda_1 x + \lambda_2 v + \sum_{j=1}^{l} \lambda_{j+3} w_j^T v, \quad \text{and}$$

$$\sum_{i=1}^{N} \xi_i \langle v, q_i \rangle q_i = \lambda_2 x + \lambda_3 v + \sum_{j=1}^{l} \lambda_{j+3} w_j x,$$

where

$$\sum_{i=1}^{N} \xi_i \langle x, q_i \rangle^2 = \lambda_1,$$

$$\sum_{i=1}^{N} \xi_i \langle x, q_i \rangle \langle v, q_i \rangle = \lambda_2,$$

$$\sum_{i=1}^{N} \xi_i \langle v, q_i \rangle^2 = \lambda_3, \quad \text{and}$$

$$\sum_{i=1}^{N} \xi_i \langle v, q_i \rangle \langle w_j x, q_i \rangle = \lambda_{j+3} \quad \text{for} \quad 1 \leq j \leq l.$$

Letting

$$\Psi_1(x, v) := \sum_{i=1}^{N} \xi_i \langle x, q_i \rangle q_i - \lambda_2 v - \sum_{j=1}^{l} \lambda_{j+3} w_j^T v$$

$$\Psi_2(x, v) := \sum_{i=1}^{N} \xi_i \langle v, q_i \rangle q_i - \lambda_2 x - \sum_{j=1}^{l} \lambda_{j+3} w_j x$$

we obtain the following.

Algorithm for (x^*, v^*) determining a first GPC:

Starting with initial values, e.g.,

$x^{(0)} := p_1, \quad v^{(0)} := \quad$ unit horizontal projection of $p_2 - p_1$ at $x^{(0)},$

obtain

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

by computing $q_j \in [p_j]$, $1 \leq j \leq N$, in optimal position with respect to $\gamma_{x^{(n)}, v^{(n)}}$ according to Section 5.3, and by setting

$$x^{(n+1)} := \frac{\Psi_1(x^{(n)}, v^{(n)})}{\|\Psi_1(x^{(n)}, v^{(n)})\|},$$

$v^{(n+1)} := \quad$ unit horizontal projection of $\Psi_2(x^{(n)}, v^{(n)})$ at $x^{(n+1)}.$

The unit horizontal projection of v at $x \in S_m^k$ is, of course, given by

$$\frac{z}{\|z\|} \quad \text{where} \quad z := v - \langle x, v \rangle x - \sum_{j=1}^{l} \Big\langle w_j(x)x, v \Big\rangle w_j(x)x.$$

The Second GPC and PM. Given a horizontal great circle $\gamma_1 = \gamma_{x,v}$ mapping to a first GPC determined by $x, v \in S_m^k$, $\Phi_1(x, v) = 0$, suppose that $\gamma_2(t) = \gamma_{y,w}(t) = y \cos t + w \sin t$, with

$$y = y(\tau) = x \cos \tau + v \sin \tau = \gamma_1(\tau)$$

for some suitable $\tau \in \mathbb{R}$, is a horizontal great circle projecting to a second GPC. According to Section 4.3, the second GPC is then obtained by minimizing the objective function

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

over $(\tau, w) \in \mathbb{R} \times M(m, k-1)$, where $q_j \in [p_j]$, $i = 1, \ldots, N$, are in optimal position to $\gamma_{y,w}$. Defining

$$z = z(\tau) = v \cos \tau - x \sin \tau = \dot{\gamma}_1(\tau)$$

and inspecting the first two rows of the constraining condition $\Phi_2(\tau, w) = 0$, observe that we may alternatively employ

$$\Phi_2(\tau, w) = \begin{pmatrix} 2\langle w, x \rangle \\ 2\langle w, v \rangle \\ \langle w, w \rangle - 1 \\ 2\langle w_1 y, w \rangle \\ \vdots \\ 2\langle w_l y, w \rangle \end{pmatrix}.$$

As before, abbreviate $\zeta_j := \sqrt{\langle y, q_j \rangle^2 + \langle w, q_j \rangle^2}$, $\xi_j := (\arccos \zeta_j)/(\zeta_j \sqrt{1 - \zeta_j^2})$, and compute (4.6) with a Lagrange multiplier $\lambda = (\lambda_1, \ldots, \lambda_{l+3})$:

$$\left. \begin{array}{rl} \sum_{i=1}^{N} \xi_i \langle w, q_i \rangle q_i & = \lambda_1 x + \lambda_2 v + \lambda_3 w + \sum_{j=1}^{l} \lambda_{j+3} w_j y, \\ \sum_{i=1}^{N} \xi_i \langle y, q_i \rangle \langle z, q_i \rangle & = \sum_{j=1}^{l} \lambda_{j+3} \langle w_j z, w \rangle \end{array} \right\} . \tag{5.6}$$

Also abbreviating

$$G(a, b) := \sum_{i=1}^{N} \xi_i \langle a, q_i \rangle \langle b, q_i \rangle, \quad \text{and} \quad A(a, b) := \sum_{j=1}^{l} \lambda_{j+3} \langle w_j a, b \rangle,$$

we have, for the Lagrange multipliers,

$$G(w, x) - \sum_{j=1}^{l} \lambda_{j+3} \langle w_j y, x \rangle = \lambda_1,$$

$$G(w, v) - \sum_{j=1}^{l} \lambda_{j+3} \langle w_j y, v \rangle = \lambda_2,$$

$$G(w, w) - \sum_{j=1}^{l} \lambda_{j+3} \langle w_j y, w \rangle = \lambda_3, \quad \text{and}$$

$$G(w, w_j y) - \lambda_1 \langle x, w_j y \rangle - \lambda_2 \langle v, w_j y \rangle - \lambda_3 \langle w, w_j y \rangle = \lambda_{j+3}, \quad 1 \le j \le l.$$

It seems convenient to alter the algorithm such that for every step, τ is set to zero, i.e., x and v are updated for every step to y and z. Then $\lambda_{j+3} = G(w, w_j y)$ and we consider, corresponding to (5.6),

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

$$- \sum_{j=1}^{d} G(w_j x, w) w_j x,$$

$$\Psi_2(x, v, w) := \quad \text{the } \tau \text{ of smallest absolute value satisfying}$$

$$G(x, v) \cos 2\tau + \frac{\sin 2\tau}{2} \Big(G(v, v) - G(x, x) \Big)$$

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

By definition, $\Psi_2(x, v, w)$ is orthogonal to x and v and it is horizontal. This leads to the following.

Algorithm for determining a second GPC:

Starting with some initial values, e.g.,

$$x^{(0)} := x, \quad v^{(0)} := v, \quad w^{(0)} := \frac{z}{\|z\|} \quad \text{where} \quad w := (p_2 - p_1) \quad \text{and}$$

$$z := w - \langle x^{(0)}, w \rangle x^{(0)} - \langle v^{(0)}, w \rangle v^{(0)} - \sum_{j=1}^{l} \langle w_j x^{(0)}, w \rangle w_j x^{(0)},$$

obtain

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

by computing $q_j \in [p_j]$, $1 \leq j \leq N$, in optimal position with respect to $\gamma_{x^{(n)}, w^{(n)}}$ according to Section 5.3, and by setting

$$\tau := \Psi_2(x^{(n)}, v^{(n)}, w^{(n)}),$$
$$x^{(n+1)} := x^{(n)} \cos \tau + v^{(n)} \sin \tau,$$
$$v^{(n+1)} := v^{(n)} \cos \tau - x^{(n)} \sin \tau, \quad \text{and}$$
$$w^{(n+1)} := \frac{\Psi_1(x^{(n)}, v^{(n)}, w^{(n)})}{\|\Psi_1(x^{(n)}, v^{(n)}, w^{(n)})\|}.$$

Having thus found x^*, v^* and w^*, recall from Section 4.3 that $\hat{x} := x^*$ is a representative of a PM on Σ_k^m. With $v_1 := v^*$ and $v_2 := w^*$, we have the two horizontal geodesics

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

projecting to a first and a second GPC on Σ_k^m. For simplicity set $x := \hat{x}$.

Higher Order GPCs. Suppose that we have found horizontal great circles $\gamma_{x,v_1}, \ldots, \gamma_{x,v_{r-1}}$, $3 \leq r \leq m$, mapping to GPCs on Σ_m^k. Then the Lagrange equation (4.7), for a horizontal great circle $\gamma_{x,v}$ projecting to a r-th order GPC on Σ_m^k, is given by

$$\sum_{i=1}^{N} \xi_i \langle v, q_i \rangle q_i = \lambda_0 x + \sum_{s=1}^{r-1} \lambda_s v_s + \lambda_r v + \sum_{j=1}^{l} \lambda_{r+3} w_j x$$

with $\zeta_i := \sqrt{\langle x, q_i \rangle^2 + \langle v, q_i \rangle^2}$, q_i in optimal position to $\gamma_{x,v}$, $\xi_i := (\arccos \zeta_i)/(\zeta_i \sqrt{1 - \zeta_i^2})$, and suitable Lagrange multipliers $\lambda_0, \ldots, \lambda_{r+l+1} \in \mathbb{R}$ that are computed as before. We then have the following.

Algorithm for determining an r-th order GPC, $r \geq 3$:

Starting with an initial value, e.g.,

$$v^{(0)} := \frac{z}{\|z\|} \quad \text{where} \quad w := (p_2 - p_1) \quad \text{and}$$

$$z := w - \langle x, w \rangle x - \sum_{s=1}^{r-1} \langle v_s, w \rangle v_s - \sum_{j=1}^{l} \langle w_j x, w \rangle w_j x,$$

obtain

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

by computing $q_j \in [p_j]$, $1 \leq j \leq N$, in optimal position with respect to $\gamma_{x,v^{(n)}}$ according to Section 5.3, and by setting

$$z^{(n+1)} := \sum_{i=1}^{N} \xi_i^{(n)} \langle v^{(n)}, q_i \rangle q_i,$$

$$\lambda_0 := \langle z^{(n+1)}, x \rangle,$$

$$\lambda_s := \langle z^{(n+1)}, v_s \rangle, \quad 1 \leq s < r,$$

$$\lambda_r := \sum_{i=1}^{N} \xi_i^{(n)} \langle v^{(n)}, q_i \rangle^2,$$

$$\lambda_{j+r+1} := \langle z^{(n+1)}, w_j x \rangle, \quad 1 \leq j \leq l, \quad \text{and}$$

$$v^{(n+1)} := \text{sign}(\lambda_r) \frac{z^{(n+1)} - \lambda_0 x - \sum_{s=1}^{j-1} \lambda_s v_s - \sum_{j=1}^{l} \lambda_{j+r+1} w_j x}{\|z^{(n+1)} - \lambda_0 x - \sum_{s=1}^{j-1} \lambda_s v_s - \sum_{j=1}^{l} \lambda_{j+r+1} w_j x\|}.$$

The IM and the IM on a GPC. The computation of an IM according to Section 4.3 can be carried out analogously. With $\zeta_j := \langle x, q_j \rangle$, $\xi_j = (\arccos \zeta_j)/(\sqrt{1 - \zeta_j^2})$, $\sum_{j=1}^{N} \xi_j \langle q_j, x \rangle = \lambda$ and $\Psi(x) := \text{sign}(\lambda) \sum_{j=1}^{N} \xi_j q_j$, we have the following algorithm:

$$\left. \begin{array}{rcl} x^{(n)} & \mapsto & x^{(n+1)} \\[2mm] x^{(n+1)} & = & \frac{\Psi(x^{(n)})}{\|\Psi(x^{(n)})\|} \end{array} \right\}.$$

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

Computing an IM on a GPC is equivalent to computing an IM on a horizontal great circle on a sphere with respect to projections of optimally positioned points according to Section 4.3. The corresponding algorithm can be found in Huckemann and Ziezold (2006).

6. Data Examples

In conclusion, we apply our methods to three typical data sets. The first example, from forest biometry, features concentrated and nearly degenerate shapes. In the second and third examples we consider classical data sets: regular, less concentrated shapes from an archaeological site, and regular concentrated shapes of macaque skulls. For all data sets we computed

GPCA as laid out in this paper, and compared it with

restricted GPCA by restricting the GPCs to pass through the sample IM with our algorithms accordingly simplified,

Euclidean PCA at the IM (PGA) by computing the covariance matrix of the data mapped under the inverse Riemann exponential to the tangent space at the sample IM, cf., Fletcher et al. (2004),

Euclidean PCA at the EM (GPA) by computing the covariance matrix of the data orthogonally projected to the tangent space of the Procrustes sample mean, cf., Dryden and Mardia (1998, Chap. 5).

Again for brevity we omit the prefix "sample" in the following. All GPCs, means and variances found below are in fact *sample* GPCs, *sample* means and *sample* variances.

6.1. Nearly degenerate shapes

In collaboration with the Institute for Forest Biometry and Informatics at the University of Göttingen, the influence on the shape of tree stems of certain external and internal factors is studied. Of particular interest is the influence by competition with nearby trees that commences when the crowns meet. We consider here a dataset of tree stems of five Douglas fir trees collected at an experimental site in the Netherlands, cf., Gaffrey and Sloboda (2000) and Table 5 in Appendix B. Since crown competition will be less reflected by lower tree rings, tetrahedral shapes have been extracted by placing landmarks at the top of the tree, as well as on the center of the stem-disk at middle height, and at the maximal and minimal radius of this disk, cf., Figure 1. Since the heights in questions are about 10 meters and the radii around 10 centimeters, all shapes are close to one-dimensional line segments. Shape change, however, occurs in all three spatial directions. As is visible in Table 2, relative data variation along the respective PCs, and distances of the data to the first PC differ considerably between GPCA, restricted GPCA and Euclidean PCA.

Thus, with this data-set we are very close to the situation described in Theorem 5.2: the shape space is locally spanned by vector fields (among others) whose sectional curvatures tend to infinity when approaching the singularity. Moreover, note that the difference between "variance explained by projection" and distances to generalized geodesics (corresponding to "variance explained by residuals") can be taken as a measure for curvature present in the data. As the data is concentrated on a geodesic (93.58% of the variation is along the first GPC), the empirical value of CX (2.466) must be generated by a high curvature of the surrounding space, cf., (3.4). The overall variance rmiv (the root of the total intrinsic variance divided by sample size) of the data-set is small. Still, for the same reason, the PM $\overline{x}_P$ is considerably far away from the EM $\overline{x}_E$ and the IM $\overline{x}_I$, the latter three being rather close to each other, in fact the distance
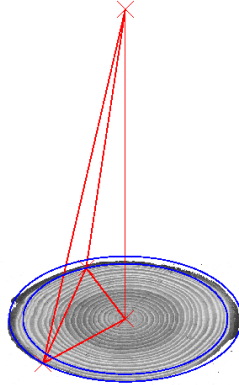
Figure 1. Extracting a tetrahedral shape from a Douglas fir tree stem. Distorted view.

between PM and IM is about of the size of rmiv. This means that the PM is well within the data spread:

$$\text{rmiv} = 0.00342, \quad d_{\Sigma_3^4}(\overline{x}_P, \overline{x}_I) = 0.002126, \quad d_{\Sigma_3^4}(\overline{x}_E, \overline{x}_I) = 2.107e - 08.$$

Therefore the classical methods of GPA, Euclidean PCA at the EM, and Euclidean PCA at the IM are practically identical. Also, when considering residuals and projections to generalized geodesics, Euclidean PCA and GPCA restricted to the IM are almost equivalent. On the other hand, due to the flexibility of GPCA in choosing the PM, the first GPC approximates the data far better than the first PCs of the other methods; in fact the approximation of the data by the first PC of restricted GPCA is worse by a factor larger than 3. Most notably, however, Euclidean PCA and the almost equivalent GPCA restricted through the IM fail to recognize the non-trivial fact that shape change for this data set is apparently one-dimensional. This finding suggests that shape variation of nearby trees of the same kind in interaction, is essentially one-dimensional, cf., Hotz, Huckemann, Gaffrey, Munk, and Sloboda (2009).

6.2. Non-concentrated regular shapes

In a second illustration, consider a data-set of 28 brooches (fibulae) from an Iron Age grave site in Münsingen, Switzerland, closely studied by Hodson et al. (1966) among others. As the cemetery grew over a large period of time, it is reasonable to believe that brooches at different locations originated in different time epochs. Thus, five individual temporal groups have been proposed. In order to study shape change, at each of the three-dimensional brooches, four landmarks have been assigned to specific "anatomical" locations. Small (1996, Section 3.5)

Table 2. Displaying for five shapes of Douglas fir trees the percentages of
variance explained by PCA based on generalized geodesics (first box), PCA
based on generalized geodesics while requiring that all GPCs pass through
the intrinsic mean (second box), PCA by computing the covariance matrix in
the tangent space of the IM under the inverse Riemann exponential (third
box), and PCA by computing the covariance matrix of the data orthogo-
nally projected to the tangent space of the EM (bottom box). In the first
two boxes, the line labelled "by projection" gives the percentages of vari-
ance obtained by projection, cf., (3.3); the line labelled "rmssd" reports the
square-root of the mean of the squared shape-distances of the data to the
respective generalized geodesic. In the lower two boxes the values obtained
by "Euclidean projection" are the precentages of eigenvalues of the respec-
tive covariance matrices under the inverse Riemann exponential at the IM
(Euclidean PCA at IM), and under orthogonal projection to the tangent
space at the EM (Euclidean PCA at EM). Under "geodesic projection" the
variances obtained by projection (as above) to the respective generalized
geodesics corresponding to the eigenvectors of the respective covariance ma-
trices are reported. No values are reported for the ultimate eigenvectors as
they point no longer into horizontal space. Similarly "rmssd" gives again
the square-root of the mean of the squared shape-distances of the data to
the respective generalized geodesics.

GPCA	GPC1	GPC2	GPC3	GPC4	GPC5
by projection	93.58 %	6.38 %	0.04193 %	5.309e-06 %	2.505e-06 %
rmssd	0.0002818	0.00183	0.002405	0.002420	0.002420
GPCA through the IM	GPC1	GPC2	GPC3	GPC4	GPC5
by projection	50.23 %	49.50 %	0.2744 %	2.197e-05 %	2.390e-07 %
rmssd	0.0009775	0.001077	0.003392	0.0034	0.0034
Euclidean PCA at IM	PC1	PC2	PC3	PC4	PC5
Euclidean projection	67.33 %	32.43 %	0.24 %	5.366e-06 %	4.295e-15 %
geodesic projection	50.98 %	48.62 %	0.4076 %	6.041e-05 %	
rmssd	0.001046	0.001240	0.003388	0.0034	
Euclidean PCA at EM	PC1	PC2	PC3	PC4	PC5
Euclidean projection	67.33 %	32.43 %	0.2401 %	5.597e-06 %	1.494e-31 %
geodesic projection	50.98 %	48.62 %	0.4073 %	6.157e-05 %	
rmssd	0.001046	0.001240	0.003388	0.0034	

has applied *principal coordinate analysis* to *Procrustes innerpoint distances* of
a planar lateral view of the landmarks. Table 3 displays the relative variances
explained by the respective five PCs in Kendall's shape space Σ_3^4.

As visible, the results of the two geodesic methods, which are this time very
similar, differ from the results of the two Euclidean approaches, which are again
almost identical. Euclidean PCA leads to the belief that the variation of brooch
shape is essentially explained by two PCs.

Two dominating components (minimizing residual variance) can also be

Table 3. Displaying for the Münsingen brooch data-set the results of the various methods of PCA with the notation of Table 2.

GPCA	GPC1	GPC2	GPC3	GPC4	GPC5
by projection	29.16 %	13.48 %	48.69 %	8.152 %	0.5145 %
rmssd	0.1178	0.1270	0.1586	0.1681	0.1720
GPCA through the IM	GPC1	GPC2	GPC3	GPC4	GPC5
by projection	28.12 %	12.72 %	51.59 %	7.085 %	0.4866 %
rmssd	0.1193	0.1266	0.1554	0.1685	0.1709
Euclidean PCA at IM	PC1	PC2	PC3	PC4	PC5
Euclidean projection	48.95 %	37.96 %	10.11 %	2.236 %	0.7397 %
geodesic projection	8.932 %	51.29 %	34.14 %	0.4577 %	5.173 %
rmssd	0.1228	0.1241	0.1567	0.1703	0.1711
Euclidean PCA at EM	PC1	PC2	PC3	PC4	PC5
Euclidean projection	49.53 %	37.11 %	10.32 %	2.287 %	0.7598 %
geodesic projection	8.95 %	50.99 %	34.27 %	0.4671 %	5.326 %
rmssd	0.1228	0.1355	0.1631	0.1704	0.1717
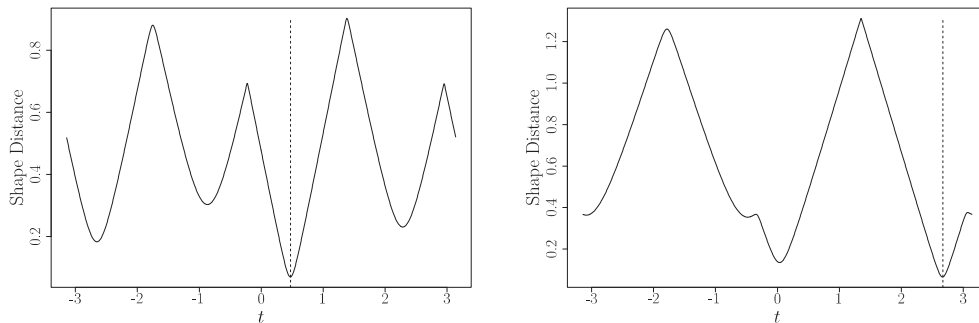
Figure 2. Shape distance (vertical axis) to shapes of the second (left) and the twenty-second (right) fibula (the sample has 28 fibulae), to shapes along the generalized geodesic $\delta(t), -\pi \leq t \leq \pi$, (horizontal axis) determined by the second Euclidean PC at the EM. The dashed line marks the global minimum, its horizontal location gives the score.

found by the two geodesic methods; however, the corresponding GPCs point in directions different from the directions of the Euclidean PCs in the tangent space of the EM (or IM, resp.) and approximate the data w.r.t. the intrinsic metric better. This fact can be explained by curvature, and even more by oscillation and the not-everywhere minimizing property of the respective generalized geodesics corresponding to the Euclidean PC directions, cf., Figure 2. There, for two fibulae, oscillation of the generalized geodesic corresponding to the second Euclidean PC is visible. The left image shows a score close to zero, the right image depicts a score close to π though there is a local (slightly higher)
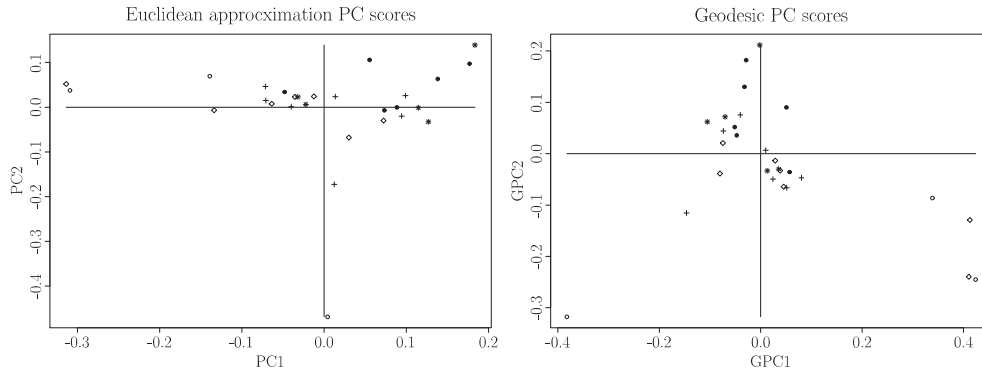
Figure 3. Scores of the five groups of the Münsingen brooch data on the first two PCs. Top: Euclidean PCA at the EM, bottom: geodesic method. The oldest group is depicted by filled circles, the second oldest by stars, the middle group by crosses, the second youngest by diamonds and the youngest by circles. The scaling on the coordinate axes is depicted in Euclidean projection (top image) and arclength on shape space (bottom image, the largest distance in shape space is $\pi/2$). Note the different scaling on the axes.

minimum close to zero. For this reason, variance explained by projection may increase with higher order PCs for all methods, even though higher order PCs have higher residual variance. This hints again at a smaller explanatory significance of variance obtained by projection, in particular for the Euclidean approximations. Analysis of residuals will then add to the significance of the findings.

In the Euclidean approximations a trend of temporal evolution can be identified as the strongest component, see Figure 3: shapes move in time from right to left along the first PC, cf., also Small (1996, p.94). Using the geodesic method, this trend is also visible in the second component: shapes move in time from top to bottom. As the stronger first principal component, however, a temporal increase in shape diversification can be identified. This latter observation is validated by a plot of residual distances in Figure 4.

In conclusion of this example, we note that the curvature estimate (3.4) for this data set is $CX = 0.1217$, much smaller than in the previous example. For this reason, as visible in Table 3, GPCA restricted through the IM approximates unrestricted GPCA rather well. Equivalently, in relation to the data spread (rmiv) and the again comparatively small distance between EM and IM, the PM is relatively closer to both IM and EM than in the previous example:

$$\text{rmiv} = 0.1734, \quad d_{\Sigma_3^4}(\overline{x}_P, \overline{x}_I) = 0.01936, \quad d_{\Sigma_3^4}(\overline{x}_E, \overline{x}_I) = 0.002291.$$
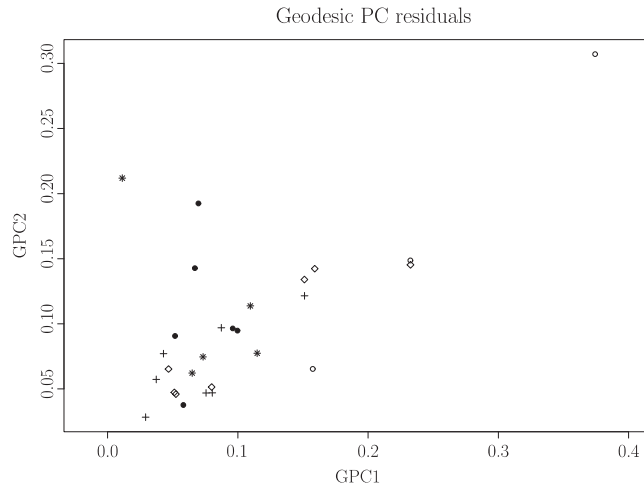
Figure 4. Residual distances to the first two GPCs for the five groups from Figure 3 of the Münsingen brooch data-set.

6.3. Concentrated regular shapes

In a final example we consider a data set of seven anatomical landmarks chosen on the skulls of nine male and nine female macaque specimen, cf., Dryden and Mardia (1998). Male specimen show a greater variance, the difference between mean shapes (EM) of both sexes is not significant, however, cf., Dryden and Mardia (1998, p.159). As in the previous examples, Table 4 reports the results of the various methods of PCA. In this example we observe no difference between the three approaches, rendering the Euclidean approximation valid. In fact, the curvature estimate (3.4) takes the very small value $CX = 0.0001924$, and the root of the total intrinsic variance divided by sample size is also of small size, rmiv $= 0.05954$. Due to these facts, all means are nearly coincident and the various methods of PCA give (almost) identical results, cf., Table 4.

7. Discussion

With this paper, we aimed at providing a method that allows one to perform non-linear multivariate statistics in cases where the data live on spaces with a non-Euclidean intrinsic structure. Furthermore, we wanted this method to rely on the given intrinsic structure alone, thereby avoiding any linear approximation. We were particularly interested in a rather general quotient space occurring from an isometric action of a Lie group on a Riemannian manifold. For such spaces, we proposed a method as desired, namely a PCA based on the spaces' intrinsic structure. A typical application lies in the statistical analysis of shapes which has until now been performed almost exclusively by linear approximations. An

Table 4. Displaying for a data set of macaque skulls the results of the various methods of PCA with the notation of Table 2.

GPCA	GPC1	GPC2	GPC3	GPC4	GPC5
by projection	31.15 %	20.10 %	14.67 %	10.57 %	6.223 %
rmssd	0.06158	0.06636	0.06858	0.07021	0.0719
GPCA through the IM	GPC1	GPC2	GPC3	GPC4	GPC5
by projection	31.15 %	20.10 %	14.67 %	10.57 %	6.223 %
rmssd	0.06158	0.06636	0.06858	0.07021	0.0719
Euclidean PCA at IM	PC1	PC2	PC3	PC4	PC5
Euclidean projection	31.18 %	20.10 %	14.67 %	10.58 %	6.216 %
geodesic projection	31.15 %	20.10 %	14.67 %	10.57 %	6.223 %
rmssd	0.06158	0.06636	0.06858	0.07021	0.0719
Euclidean PCA at EM	PC1	PC2	PC3	PC4	PC5
Euclidean projection	31.13 %	20.11 %	14.68 %	10.59 %	6.223 %
geodesic projection	31.15 %	20.10 %	14.67 %	10.57 %	6.223 %
rmssd	0.06158	0.06636	0.06858	0.07021	0.0719

approach respecting the non-linearity of the intrinsic metric leads to non-linear optimization problems, which pose specific numerical challenges. The methods derived in this work are applicable to a wide variety of non-manifold shape spaces which usually occur when studying three- and higher-dimensional shapes. Taking Kendall's shape spaces as an example we have illustrated how this methodology provides for explicit algorithms.

We have presented PCA based solely on the intrinsic structure in contrast to the current methods of PCA by embedding into and projecting to Euclidean space; our motivation was threefold. First, the intrinsic approach may be the only method available nearsingularities of the quotient. On Kendall's shape spaces this is the case when mean shapes are nearly singular. Recall from the remarks following Theorem 5.2 that this cannot happen with planar shapes, i.e., for Σ_2^k. It is, however, the case when samples of three-dimensional objects fall into high curvature regions of the regular shape space, as happens in biological applications. This phenomenon which we encountered in the study of shapes of individual tree stems (cf., Section 6.1) was the initial motivation for this research. Second, our approach may serve as a basis for a larger study to compare the validity of the much simpler linear approximation methods. Such a study was carried out by Huckemann and Hotz (2009) for Kendall's planar shape spaces. Third, we consider our work a preliminary basis for further development of non-linear statistical analysis working exclusively intrinsically.

At this point let us sum up our results on quotients with singularities such as Kendall's shape spaces concerning GPCA, restricted GPCA through the IM, Euclidean PCA at the IM (PGA), and Euclidean PCA at the EM (general Procrustes analysis), cf., Section 6. We found the following.

1. Due to the vicinity of IM and EM and the proximity of Euclidean and intrinsic distances, Euclidean PCA at the IM and Euclidean PCA at the EM are practically equivalent.

2. Due to unbounded curvature and oscillations of generalized geodesics, Euclidean PCA may fail to recognize data features that occur under GPCA. This is the case for

   (a) large data-spread, even with little curvature present: then GPCA may well be approximated by restricted GPCA; and

   (b) data within high curvature regions: then restricted GPCA may also fail to recognize data features that occur under GPCA due to the fact that IM and PM may differ considerably when high curvature is present.

3. Curvature within the data can be estimated by CX from (3.4).

We view our methods and results as an early contribution to the ambitious task of carrying over statistical methods from linear Euclidean spaces to manifolds and more general spaces. We would like to conclude by pointing out open problems and research directions we consider to be of high importance.

Exploring Effects of Non-Euclidean Geometry. As we have seen, the geometric structure of generalized geodesics for Kendall's shape spaces for three-dimensional configurations is far more complicated than it is for two-dimensional configurations. In particular, the location and the descriptive nature of the PM deserves to be studied more closely. Also, studying oscillation effects certainly requires more research. In case of Kendall's shape spaces, an upper bound on the number of local minima of the distance of a given shape to shapes along a generalized geodesic depends linearly on $\dim(SO(m)/I_{\gamma(t)})$ with the isotropy group $I_{\gamma(t)}$ at $\gamma(t)$. Note that by Lemma A.2 this dimension is constant a.e. on $t \in [0, 2\pi)$. In general, conditions on geodesics and the data support D (the geodesic convex hull of the data, say) can be found that ensure that geodesics and geodesic segments in D are everywhere minimizing.

Computing Curvature Present in the Data. Even with good numerical algorithms, GPCA will be computationally more costly than Euclidean PCA. It would be helpful to develop diagnostic tools for the evaluation of the benefit of GPCA over Euclidean PCA. CX as introduced in (3.4) is such an indicator; however, CX is available only after computing the GPCA. An alternative could be a numerical method to compute intrinsic sectional curvature, say at a mean of the data sample, using suitable horizontal and vertical vector fields.

Degenerate Means. Sample means of non-degenerate shapes grouping around a degenerate shape may move closer to the degenerate shape, cf., Corollary 5.6. It would be interesting to determine under what circumstances (e.g., symmetry conditions on the distribution) the intrinsic population mean of concentrated non-degenerate shapes may be unique and/or degenerate.

Numerics. Crucial to the success of GPCA is a reliable method for optimal positioning w.r.t. a horizontal geodesic. The results of this paper have been validated by highly time consuming brute force methods to ensure that local minima found are in fact global. In order to undertake larger studies in the future, considerable improvement of the numerical methods is essential. The numerical problem can be attacked either by travelling along a goedesic i.e., essentially along S^1, or by optimizing over $SO(m)$; both are problems of non-convex optimization, cf., Section 5.3. The objective function corresponding to the former can be non-differentiable, thus posssibly leading to very narrow minima; the objective function corresponding to the latter can be expressed as a multivariate polynomial. To the knowledge of the authors these problems encountered have found little attention in numerical analysis in the past.

Extensions to Other Shape Spaces. Obviously, the methods developed here extend to many other shape spaces as well. If horizontal geodesics and distances to them are given analytically, then our method can be applied directly. Our approach may be applicable as well if geodesics and distances to geodesics can only be computed numerically. Such numerical methods have been proposed, e.g., by Miller, Trouvé, and Younes (2006) or Schmidt, Clausen and Cremers (2006), for general models, such as e.g., Michor and Mumford (2006) or Klassen et al. (2004), for closed 2D curves, or closed 2D and 3D contours based on medial axes, e.g., Pizer et al. (2003), Fletcher et al. (2004), as well as Fuchs and Scherzer (2007).

Manifold PCA. On the conceptual side, as pointed out in Section 3.1, non-nested intrinsic PCA based on (generalized, cf., Appendix A) manifolds appears as a natural extension of our methods. One might consider manifolds totally geodesic at a point as well as manifolds totally geodesic at all points. In particular for Kendall's shape spaces of three- and higher-dimensional objects, the former may be numerically hard to compute, and the latter may be available only for some dimensions. We also note that parametrized spaces obtained thereby may be manifolds only locally.

Inference. In this paper we did not address the issue of inference. Classical and functional PCA allows for manifold inferential tools, such as tests and confidence bands on the components (see e.g., Kneip and Utikal (2001) or Munk et al. (2007)). These approaches are all asymptotic in nature and a linearization of the underlying estimator typically leads to satisfactory results, often for quite small samples. This does not seem to be the case anymore in the present context, rather high curvature effects suggest that (first order) asymptotic considerations will fail. Model selection, as it is used for the automatic selection of the number of required components (see e.g., Hsieh (2007) in the context of nonlinear PCA), is another related issue. It would be of great importance to transfer these ideas to GPCA. Finally, we believe that there is much room for sharp risk bounds for the geodesic principal components as it has been investigated e.g., by Zwald, Bousquet and Blanchard (2004) in the context of kernel PCA.

Acknowledgement

Appendix A: Foci and Focal Points

In this section we use the notation and hypotheses of Section 2: a compact Lie group G acts isometrically on a complete finite-dimensional and connected Riemannian manifold M giving rise to the canonical quotient $\pi : M \to M/G =: Q$. In order to define *generalized submanifolds* on Q as well as *foci and focal points*, we introduce some additional notation and results, cf., Bredon (1972, p.182).

A point $p \in M$ is of *orbit type* (G/H) if $I_{gp} = H$ for some $g \in G$. This is equivalent to saying that all isotropy groups in the fiber $[p]$ are conjugates of H. The union

$$M^{(H)} := \{p \in M : I_{gp} = H \ \text{ for some } \ g \in G\}$$

of all points of equal orbit type (G/H) - if not void - is a submanifold of M, and $\overline{M^{(H)}} \setminus M^{(H)}$ consists of points of *smaller orbit type*, i.e., $H \subset I_{gp'}$, $H \neq I_{gp'}$

for every $p' \in \overline{M^{(H)}} \setminus M^{(H)}$ with suitable $g = g_{p'} \in G$. In particular, $Q^{(H)} := M^{(H)}/G$ is a manifold. Recall that $Q^{(\{id\})} = Q^*$ is the manifold part of Q.

It is well-known that every point has a neighborhood in which only finitely many orbit types occur. This can easily be seen by an inductive argument relying on the Slice Theorem (cf., Section 2.2) and Lemma A.2 below. Since manifolds are separable (i.e., contain a countable dense set) there are only countably many orbit types on M.

Definition A.1. We call $K \subset Q$ a *generalized submanifold* of Q if $K \cap Q^{(I_p)}$ is a submanifold of $Q^{(I_p)}$ for every $[p] \in K$.

Since $\pi|_{M^{(I_p)}} : M^{(I_p)} \to Q^{(I_p)}$ is a Riemannian submersion, $L^{(I_p)} := \pi^{-1}(K \cap Q^{(I_p)}) \subset M^{(I_p)}$ is a submanifold of M for all $[p] \in K$, for a generalized submanifold K. In particular, every submanifold K of Q^* is a generalized manifold. The notion of generalized submanifolds includes generalized geodesics.

Lemma A.2. *The point set L of a horizontal geodesic contains a point p of maximal orbit type. Then $L^{I_p} := L \cap M^{(I_p)} = \{p' \in L : I_{p'} = I_p\}$ is a submanifold, $L \setminus L^{I_p}$ is contained in $\overline{M^{(I_p)}} \setminus M^{(I_p)}$, and consists only of isolated points.*

Proof. Assume that $\gamma(t) = \exp_{p_0}(tv)$ with $p_0 = \gamma(0)$ and suitable $v \in H_{p_0}M$ is a parametrization of the point set L of a horizontal geodesic. Moreover let $H_t := I_{\gamma(t)}$. From (2.2) we obtain $H_t \subset H_0$ for t sufficiently small. Since

$$g\Big(\exp_{p_0}(tv)\Big) = \exp_{gp_0}(t\mathrm{d}gv),$$

we have at once $g \in H_t$ for $t \neq 0$ sufficiently small if and only if $\mathrm{d}gv = v$. Hence $H_t = H_{t'}$ for all sufficiently small non-zero t, t'. This yields that the orbit type on L near p_0 is constant and maximal, except for at most the isolated point p_0. W.l.o.g assume that $p_0 = p$ is of locally maximal orbit type. Since $\gamma(t) = \exp_p(tv)$ is a global parametrization of L, we have that $H_0 \subset H_t$ for all $t \in \mathbb{R}$. This argument can be applied to any point of locally maximal orbit type, hence $H_0 = H_t$ for all $t \in \mathbb{R}$ except for at most isolated points. $\square$

Now, consider an arbitrary submanifold L of M and an arbitrary generalized submanifold K of Q. Recall the normal bundle $NL = \cup_{p \in L}\{p\} \times N_pL$ of L in M. With the Riemann exponential of M we have the well defined *endpoint map*

$$\phi : NL \to M$$
$$(p, v) \mapsto \phi(p, v) = \exp_p(v).$$

Also, we have the set of *orthogonal projections*

$$K_{[p]} := \Big\{[p^*] \in K : d_Q([p], [p^*]) = \inf_{[p'] \in K} d_Q([p], [p'])\Big\}$$

of $[p] \in Q \setminus K$ onto K.

Definition A.3. Call

$$
\begin{array}{lll}
p \in M \setminus L \ \ \text{a } \textit{focus} \text{ of } \ L & \text{if} & \exists (p', v') \in NL \ \ \text{such that} \ \ \exp_{p'}(v') = p \\
& & \text{and } (\mathrm{d}\phi)_{(p', v')} \text{ is singular,} \\
[p] \in Q \setminus K \ \ \text{a } \textit{focal point} \text{ of } \ K & \text{if} & |K_{[p]}| > 1.
\end{array}
$$

Note that foci are points with locally stationary distance to a subset of L.

If $G = \{id\}$ and $M = \mathbb{R}^m$, then these definitions agree with Bhattacharya and Patrangenaru (2003, p.2 and p.12). Also for $M = \mathbb{R}^m$, foci have been introduced as "focal points" by Milnor (1969, p.32) and Hastie and Stuetzle (1989, p.514) call the focal points of a one-dimensional submanifold of Euclidean space the *ambiguity set*. We first give some illustrations in the Euclidean plane.

1. The center of a circle is both a focus and a focal point to that circle; the "foci" of a non-circular ellipse are its foci, the open line segment between them constitute its focal points, none of which is a focus.

2. Consider a suitably smoothed version of the union of a circular segment and a line segment

$$
\left\{ -\frac{1}{\sqrt{2}} \right\} \times \left[-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right] \cup \left\{ (x, y) \in \mathbb{R}^2 : x \geq -\frac{1}{\sqrt{2}}, x^2 + y^2 = 1 \right\}.
$$

Then the origin is its focus, the minimal distance to the origin, however, is attained at $(-1/\sqrt{2}, 0)$, from which the origin is reached by the endpoint map with non-singular derivative.

3. In order to see that the set of foci of a closed manifold is not necessarily closed, consider a suitably smoothed version M of the union $\cup_{n=2}^{\infty}(K_n \cup L_n)$ of circular segments K_n and line segments L_n defined by

$$
K_n := \left\{ (x, y) \in \mathbb{R} : \left(x - \frac{1}{n} \right)^2 + y^2 = R_n^2, \ x > 0, \ |y| \leq 2 \right\}, \quad R_n := n - \frac{1}{n}
$$

$$
L_n := \begin{cases} \left\{ (x, -2) : \dfrac{1}{n} + \sqrt{R_n^2 - 4} \leq x \leq \dfrac{1}{n+1} + \sqrt{R_{n+1}^2 - 4} \right\} & n \ \text{odd} \\[3mm] \left\{ (x, 2) : \dfrac{1}{n} + \sqrt{R_n^2 - 4} \leq x \leq \dfrac{1}{n+1} + \sqrt{R_{n+1}^2 - 4} \right\} & n \ \text{even.} \end{cases}
$$

Then, all $p_n = (1/n, 0)$ with $n \geq 2$ are foci of M; the origin, their limit point, however is not a focus.

We are concerned with *minimizing foci* which form a closed subset for closed manifolds as we shall see.

Definition A.4. A focus p of L will be called *minimizing* if there is $(p', v) \in NL$ such that $\exp_{p'}(v) = p$, $(d\phi)_{(p',v)}$ is singular and $d_Q(p, L) = \|v\|$.

The following is a generalization of Theorem 3.2 of Bhattacharya and Patrangenaru (2003, p.12), which is a generalization of Proposition 6 of Hastie and Stuetzle (1989, p.515).

Theorem A.5. *Let G be a compact Lie group acting isometrically on a complete finite-dimensional Riemannian manifold M, and let L be an arbitrary submanifold of M, K an arbitrary generalized submanifold of $Q = M/G$. Then*

(a) *the set of foci and focal points of L has measure zero in M,*

(b) *if L is closed then the set of minimizing foci and focal points of L is closed in M,*

(c) *the set of focal points of K has measure zero in Q.*

Proof. Throughout the proof we use the notation introduced above. Moreover let

$$
\begin{array}{ll}
\mathcal{L}^0 & \text{be the set of foci of } \mathrm{L}, \\
\mathcal{L}^{00} & \text{be the set of minimizing foci of } \mathrm{L}, \\
\mathcal{L} & \text{be set the of focal points of } \mathrm{L}, \text{ and} \\
\mathcal{K} & \text{be set the of focal points of } \mathrm{K}.
\end{array}
$$

From the following Claims I and II we obtain the assertion (a). Claims III and IV yield assertion (b), and Claim V assertion (c).

Claim I: $\mathcal{L}^0$ is of measure zero. By definition, $\mathcal{L}^0$ is the set of critical points of the endpoint map $\phi : NL \to M$. As in Milnor (1969, p.33), Sard's theorem ensures that the critical points have measure zero in M.

Claim II: $\mathcal{L}^1 := \mathcal{L} \setminus \mathcal{L}^0$ is of measure zero. From every focus $p \in \mathcal{L}$ we have the set of orthogonal projections L_p. Conversely, for every point $p' \in L_p \subset M$, there is a unique normal vector $v = v(p, p')$ such that $p = \exp_{p'}(v)$. If $p' \in L_{p_1} \cap L_{p_2}$ for focal points $p_1 \neq p_2$, then obviously $v(p_1, p')$ and $v(p_2, p')$ cannot be collinear. Hence, there is a subset $A \subset CL$ of the cylindrical manifold

$$CL := \{(q, v) \in NL : \|v\| = 1\}$$

around L in NL and a mapping $\psi : A \to NL$, $(q, v) \to (q, t(q, v)v)$ such that $\chi := \phi \circ \psi : A \to \mathcal{L}^1$ is surjective. We show that χ is locally homeomorphic. Then, $\mathcal{L}^1 \subset M$ is the locally homeomorphic image of a subset A of a set of measure zero CL in NL with $\dim(NL) = \dim(M)$; hence $\mathcal{L}^1$ is a set of measure zero in M.

Continuity of χ follows from continuity of t: let $A \ni (q_n, v_n) \to (q, v) \in A$ and $t_n = t_n(q_n, v_n) \to t_0$ with focal points $\mathcal{L}^1 \ni p_n := \exp_{q_n}(t_n v_n) \to \exp_q(t_0 v) = p$. By continuity of distance, $d_M(p, q) = d_M(p, L)$. By hypothesis $\widetilde{p} = \exp_q(t(q, v)v)$ is focal with $d_M(\widetilde{p}, L) = d(\widetilde{p}, q)$. This yields $t_0 = t(q, v)$.

Moreover, $\chi^{-1} = \rho \circ \phi^{-1}|_{\mathcal{L}^1}$ is locally well defined and continuous as it is the composition of the continuous projection

$$\rho : \left\{ (q, v) \in NL : \|v\| \neq 0 \right\} \to CL, \quad (q, v) \mapsto \left(q, \frac{v}{\|v\|} \right),$$

and a locally diffeomorphic inverse branch ϕ^{-1} around $\mathcal{L}^1$; by hypothesis we stay away from the singularity set $\mathcal{L}^0$ of ϕ.

Claim III: $\mathcal{L}^{00}$ is closed. Consider sequences $p_n \in \mathcal{L}^{00}$, $p'_n \in L$, and $v_n \in T_{p'_n} L$ such that $p_n = \exp_{p'_n}(v_n)$ and $\mathrm{d}\phi_{(p'_n, v_n)}$ is singular. If $p_n \to p \in M$ then, as we are considering minimizing foci only, $\|v_n\|$ is bounded. As a consequence, p'_n has a point of accumulation $p' \in M$ which is in L if the latter is closed. By continuity there is a $v \in N_{p'} L$ with $\exp_{p'}(v) = p$ and $\|v\| = d_M(p, L)$. Again by continuity, $\mathrm{d}\phi_{q,v}$ is singular, proving that $p \in \mathcal{L}^{00}$.

Claim IV: $\mathcal{L}^{00} \cup \mathcal{L}$ is closed. In order to complete the argument that the set of minimizing foci and focal points is closed, we consider a non-focal limit point $p \in M$ of a sequence of focal points $p_n \in M$ and show that it is a minimizing focus. Indeed if $q_n, q'_n \in L_{p_n}$, $q_n \neq q'_n$, we may assume that $q_n, q'_n \in L$ have a common point of accumulation $q \in L$ (since p is non-focal) and that

$$\exp_{q_n}(t_n v_n) = p_n = \exp_{q'_n}(t_n v'_n) \tag{A.1}$$

for suitable unit length $v_n \in N_{q_n} L$, $v'_n \in N_{q'_n} L$, and $t_n = d_M(q_n, L) > 0$. Again, t_n has a point of accumulation t and v_n, v'_n have a point of accumulation v such that

$$\exp_q(tv) = p, \quad t = d_M(p, L).$$

Hence, from (A.1) we infer at once that $\mathrm{d}\phi_{(p, tv)}$ is singular and thus p is a minimizing focus.

Claim V: $\mathcal{K}$ is of measure zero in Q. Since there are only countably many orbit types in M, by hypothesis there is an index set $I \subset \mathbb{N}$ such that $K = \cup_{i \in I} K_i$, with submanifolds $K_i := K \cap Q^{(H_i)}$ of $Q^{(H_i)}$ and closed subgroups H_i of G for $i \in I$. As noted in the beginning of this section, the inverse projection $L_i := \pi^{-1}(K_i)$ of each of these is a submanifold of M. Moreover, every point $p \in [p]$, where $[p]$

is focal for K, is focal for $L_i \cup L_j$ with suitable indices $i, j \in I$. As the argument of claim II is local in nature, it can be applied to every L_i with

$$\widetilde{\mathcal{L}}_i^1 := \left\{ p \in \mathcal{L}^1 \setminus \mathcal{L}_i^0 : L_p \cap L_i \neq \emptyset \right\}$$

instead of $\mathcal{L}_i^1 \setminus \mathcal{L}_i^0$; here $L = \cup_{i \in I} L^i$ and $\mathcal{L}_i^0, \mathcal{L}_i^1$ denote the foci and focal points, resp. of L_i. This yields that the set of focal points for $L = \cup_{i \in I} L_i$ is of measure zero in M. As this is the inverse projection of $\mathcal{K}$, the latter is of measure zero in Q.

Appendix B: Tree-Stem Data Set

Table 5. Tetrahedral configurations from the stems of 5 Douglas fir trees. Units are given in meters.

Landmarks	1	2	3	4
Tree 1	27	14	14	14
	0	0	0.05662873	-0.09531028
	0	0	-0.0674875	0.03469011
Tree 2	29.7	17.6	17.6	17.6
	0	0	-0.07479217	0.0587687
	0	0	-0.04318128	0.1017904
Tree 3	32.5	18.2	18.2	18.2
	0	0	0.1427532	-0.1581646
	0	0	-3.496329e-17	0.09131639
Tree 4	28	18	18	18
	0	0	0.05264516	0.01673678
	0	0	-0.06274006	0.094919
Tree 5	24.3	16.1	16.1	16.1
	0	0	-0.0591466	0.04722901
	0	0	-0.03414831	0.08180305

References

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

Ambartzumian, R. V. (1990). *Factorization, Calculus and Geometric Probability*. Cambridge University Press.

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

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

Blum, H. and Nagel, R. N. (1978). Shape description using weighted symmetric axis features. *Pattern Recognition* **10**, 167-180.

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

Borzellino, J., Jordan-Squire, C., Petrics, G. and Sullivan, D. (2007). On the existence of infinitely many closed geodesics on orbifolds of revolution. Submitted.

Bredon, G. E. (1972). *Introduction to Compact Transformation Groups*. Vol. 46 of Pure and Applied Mathematics. Academic Press.

Bubenik, P. and Kim, P. T. (2007). A statistical approach to persistent homology. *Homology Homotopy Appl.* **9**, 337-362.

Cootes, T. F., Taylor, C. J., H., C. D. and Graham, J. (1992). Training models of shape from sets of examples. In *Britisch Machine Vision Conference*, 9-18, Springer, Berlin.

de Silva, V. and Carlsson, G. (2004). Topological estimation using witness complexes. In *SPBG04 Symposium on Point-Based Graphics*, 157-166.

Dryden, I. L. and Mardia, K. V. (1998). *Statistical Shape Analysis*. Wiley, Chichester.

Fletcher, P. T. and Joshi, S. C. (2007). Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Processing* **87**, 250-262.

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

Fuchs, M. and Scherzer, O. (2007). Regularized reconstruction of shapes with statistical a priori knowledge. Technical Report 51, FSP 092.

Gaffrey, D. and Sloboda, B. (2000). Dynamik der Stammorphologie: Ansätze zur quantifizierung und modellierung der assimilatakollation. In *Proceedings Deutscher Verband Forstl*. Forschungsanstalten, Sektion Forstl. Biometrie und Informatik. 12. Tagung, 7-26, Göttingen.

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

Goodall, C. R. and Mardia, K. V. (1999). Projective shape analysis. *J. Graphical and Computational Statist.* **8**, 143-168.

Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika* **40**, 33-51.

Hastie, T. and Stuetzle, W. (1989). Principal curves. *J. Amer. Statist. Assoc.* **84**, 502-516.

Helgason, S. (1962). *Differential Geometry and Symmetric Spaces*. Academic Press, New York.

Hendriks, H. and Landsman, Z. (1998). Mean location and sample mean location on manifolds: asymptotics, tests, confidence regions. *J. Multivariate Anal.* **67**, 227-243.

Hendriks, H., Landsman, Z. and Ruymgaart, F. (1996). Asymptotic behaviour of sample mean direction for spheres. *J. Multivariate Anal.* **59**, 141-152.

Hodson, F. R., Sneath, P. H. and Doran, J. E. (1966). Some experiments in the numerical analysis of archeological data. *Biometrika* **53**, 411-324.

Hotz, T., Huckemann, S., Gaffrey, D., Munk, A. and Sloboda, B. (2009). Shape spaces for prealingend star-shaped objects in studying the growth of plants. *J. Roy. Statist. Soc. Ser. C*, to appear.

Hsieh, W. W. (2007). Nonlinear principal component analysis of noisy data. *Neural Netw.* **20**, 434-443.

Huckemann, S. and Hotz, T. (2009). Principal component geodesics for planar shape spaces. *J. Multivariate Anal.* **100**, 699-714.

Huckemann, S. and Ziezold, H. (2006). Principal component analysis for Riemannian manifolds with an application to triangular shape spaces. *Adv. Appl. Probab.* **38**, 299-319.

Karcher, H. (1977). Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math.* **XXX**, 509-541.

Kendall, D. G. (1984). Shape manifolds, procrustean metrics and complex projective spaces. *Bull. Lond. Math. Soc.* **16**, 81-121.

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

Kent, J. (1994). The complex Bingham distribution and shape analysis. *J. Roy. Statist. Soc. Ser. B* **56**, 285-299.

Kim, P. T. and Koo, J.-Y. (2005). Statistical inverse problems on manifolds. *J. Fourier Anal. Appl.* **11**, 639-653.

Klassen, E., Srivastava, A., Mio, W. and Joshi, S. (2004). Analysis on planar shapes using geodesic paths on shape spaces. *IEEE Trans. Pattern Anal. Mach. Intell.* **26**, 372-383.

Kneip, A. and Utikal, K. J. (2001). Inference for density families using functional principal component analysis (with discussion). *J. Amer. Statist. Assoc.* **96**, 519-542.

Kobayashi, S. and Nomizu, K. (1969). *Foundations of Differential Geometry*, Vol. II. Wiley, Chichester.

Krim, H. and Yezzi, A. J. J. E. (2006). Statistics and analysis of shapes. *Modeling and Simulation in Science, Engineering and Technology.* Birkhäuser, Boston.

Lang, S. (1999). *Fundamentals of Differential Geometry.* Springer.

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

Le, H. and Kume, A. (2000). Detection of shape changes in biological features. *J. Microsc.* **200**, 140-147.

Mardia, K. and Patrangenaru, V. (2001). On affine and projective shape data analysis. In *Functional and Spatial Data Analysis*, Proceedings of the 20th LASR Workshop (Edited by K. V. Mardia and R. G. Aykroyd), 39-45.

Michor, P. W. and Mumford, D. (2006). Riemannian geometries on spaces of plane curves. *J. European Math. Soc.* **8**, 1-48.

Miller, M. I., Trouvé, A. and Younes, L. (2006). Geodesic shooting for computational anatomy. *J. Math. Imaging Vis.* **24**, 209-228.

Milnor, J. W. (1969). *Morse Theory.* Princeton University Press, Princton, 3rd printing with corrections.

Munk, A., Paige, R., Pang, J., Patrangenaru, V. and Ruymgaart, F. (2007). The one- and multi-sample problem for functional data with application to projective shape analysis. *J. Multivariate Anal.* 815-833.

Palais, R. S. (1960). Slices and equivariant embeddings. In *Seminar on Transformation Groups*, (Edited by A. Borel), **46** in Annals of Mathematics Studies, 101-115. Princeton Univ. Press, Princton NJ.

Palais, R. S. (1961). On the existence of slices for actions of non-compact Lie groups. *Ann. Math.* 2*nd Ser.* **73**, 295-323.

Pizer, S. M., Siddiqi, K., Székely, G., Damon, J. N. and Zucker, S. W. (2003). Multiscale medial loci and their properties. *Int. J. Comput. Vision* **55**, 155-179.

Schmidt, F. R., Clausen, M. and Cremers, D. (2006). Shape matching by variational computation of geodesics on a manifold. In: Pattern Recognition (Proc. DAGM), Vol. 4174 of LNCS, 142-151, Springer, Berlin, Germany.

Schmidt, F. R., Töppe, E., Cremers, D. and Boykov, Y. (2007). Intrinsic mean for semi-metrical shape retrieval via graph cuts. In *DAGM-Symposium*, (Edited by F. A. Hamprecht, C. Schnörr and B. Jähne), Vol. 4713 of Lecture Notes in Computer Science, pp.446-455, Springer.

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

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

Zwald, L., Bousquet, O., and Blanchard, G. (2004). Learning theory. In *COLT. Vol* 3120 *of Lecture Notes in Computer Science*, (Edited by J. Shawe-Taylor and Y. Singer), 594-608, Springer.

Institute for Mathematical Stochastics, Georgia Augusta Universiät Göttingen, Maschmühlenweg 8-10, D-37073 Göttingen, Germany.

E-mail: huckeman@math.uni-goettingen.de

Institute for Mathematical Stochastics, Georgia Augusta Universiät Göttingen, Maschmühlenweg 8-10, D-37073 Göttingen, Germany.

E-mail: hotz@math.uni-goettingen.de

Institute for Mathematical Stochastics, Georgia Augusta Universiät Göttingen, Maschmühlenweg 8-10, D-37073 Göttingen, Germany.

E-mail: munk@math.uni-goettingen.de

COMMENT

Rabi N. Bhattacharya

University of Arizona

The authors Huckemann, Hotz, and Munk should be commended for their mathematically elegant and substantial theory of *geodesic principal components analysis* (GPCA) based on the novel idea of *generalized geodesics.* A particularly important aspect of this theory is its ability to overcome some of the vexing technical problems associated with Kendall's shape space Σ_m^k for $m > 2$. We expect this theory to be further developed into an effective statistical tool for feature selection and classification.

We here discuss mainly certain aspects of the paper as they relate to statistical analysis on shape manifolds with landmarks-based data-a field pioneered by Kendall (1977, 1984), and Bookstein (1978). The problem of identifying a shape distribution, or discriminating between two or more shape distributions, arises in many fields including morphometrics (Bookstein (1978)), archaeology (Kendall

(1984)), medical imaging and diagnostics (Bandulasiri, Bhattacharya and Patrangenaru (2009), Hufnagel, Pennec, Ehrhardt, Ayache and Handels (2008)), and machine vision or robotics (Munk, Paige, Pang, Patrangenaru and Ruymgaart (2008), and Patrangenaru, Liu and Sugathadasa (2010)). A fairly comprehensive account of the parametric statistical theory is provided in Dryden and Mardia (1998), where important contributions due to these authors, as well as Kent (1992), Goodall (1991), and Prentice (1984), and many others, may be found. Our focus here is on nonparametric methods based on the so-called Fréchet means of the distributions. Recall that a *Fréchet mean* of a distribution on a metric space is the minimizer, if unique, of the expected squared distance from a point. On a manifold, two appropriate distances are (1) the extrinsic distance, the Euclidean distance induced from a proper equivariant embedding in a vector space, and (2) the intrinsic distance, the geodesic distance on a manifold endowed with a natural Riemannian metric. The respective Fréchet minimizers are called the *extrinsic mean* and the *intrinsic mean.* Since (a) the question of uniqueness of the Fréchet minimizer is more easily resolved for the extrinsic distance than the intrinsic distance, and (b) computations for the intrinsic analysis are harder than those for the extrinsic analysis, one may sometimes prefer statistical tests based on the sample extrinsic mean. Our main references for this theory are Bhattacharya and Patrangenaru (2003, 2005), Hendriks and Landsman (1998, 2007), Mardia and Patrangenaru (2005), Bhattacharya and Bhattacharya (2008) Bhattacharya (2009), and Bandulasiri et al. (2009).

The main objective of the article under discussion by Huckemann, Hotz and Munk is to develop an intrinsic analog of the Euclidean principal components analysis (PCA) on the quotient of a Riemannian manifold M by the isometric action of a Lie group G. Among important examples are Kendall's shape spaces Σ_m^k of k-ads in $\mathbb{R}^m$. Recall that a k-*ad* in $\mathbb{R}^m$ is a set of k points, or landmarks, in $\mathbb{R}^m$, excluding those where all k points are the same. One may identify the *pre-shape of a k-ad* as a centered and scaled (to unity) k-ad, i.e., a point in S^d, with $d = m(k-1) - 1$. The shape of the k-ad is then the orbit of the preshape under rotation in $\mathbb{R}^m$. Thus in this case $M = S^d$ and $G = SO(m)$, the special orthogonal group of rotations in $\mathbb{R}^m$. The most important cases are the planar shape space of k-ads ($m = 2$), and the 3D shape space of k-ads ($m = 3$). The planar shape space Σ_2^k, with the plane identified as the complex plane, is then the *complex projective space* $\mathbb{C}P^{k-2}$ - a familiar and important example of a compact manifold in differential geometry. Nonparametric statistical analysis may be carried out for distributions on this space following Bhattacharya and Patrangenaru (2003, 2005), and Bhattacharya and Bhattacharya (2008). For the case $m = 3$, however, the action of the group $SO(3)$ is *not free* (Small (1996), and Kendall, Barden, Carne and Le (1999)). Indeed, if a k-ad consists of collinear

points in $\mathbb{R}^3$, then its preshape lies on a line, and all rotations around this line keep the k-ad fixed. But no nonlinear k-ad is left fixed by a rotation other than the identity. Thus in this case M is not a differential manifold; the orbits of preshapes have different dimensions at different points, so that the horizontal tangent spaces on M do not have the same dimension everywhere. This problem gets worse with higher dimensions m. This difficulty has impeded statistical analysis of $3D$ shapes in the past, although a proper embedding for extrinsic analysis has been found recently by Bandulasiri and Patrangenaru (2005) and Bandulasiri et al. (2009), switching to the full orthogonal group $O(m)$ in place of $SO(m)$. The embedding was also obtained independently in Dryden, Kume, Le and Wood (2008). The extrinsic mean for this so-called *Schoenberg embedding* was obtained by Bhattacharya (2009), who carried out nonparametric analysis for the two-sample problem on this space referred to as *reflection shape space.* For intrinsic analysis the problem is not resolved by taking out collinear points from the space. For, whether one considers $SO(m)$ or $O(m)$ as the group G, the quotient space is then incomplete with respect to the geodesic distance, and the curvature of the manifold tends to infinity near the holes created by taking out exceptional points. Thus standard results on the existence and uniqueness due to Karcher (1977) do not hold (also see Kendall et al. (1999)).

Although a general theory of intrinsic means is still not available for such manifolds, the article by Huckemann et al. shows that one can still consider the original compact space Σ_m^k, or more generally the quotient of a complete Riemannian manifold with respect the action of an isometric Lie group G on it, construct generalized geodesics on it as images under the projection map of horizontal geodesics on the ambient manifold. It would be interesting to explore extensions on such quotient spaces for (i) a Karcher type theory for the existence of a (unique) Fréchet mean, and (ii) intrinsic analysis of one-sample and two-sample problems developed in Bhattacharya and Patrangenaru (2005) and Bhattacharya and Bhattacharya (2008). In any case, however that endeavor turns out, Huckemann, Hotz and Munk have provided a generalization of the theory of principal components (PC) on these quotient spaces in general. The detailed application of this theory to Kendall's shape spaces Σ_m^k is a particularly important contribution of these authors. The merits and some limitations of this theory developed so far are discussed below.

For an absolutely continuous distribution Q on a complete Riemannian manifold, let $\tilde{Q}$ denote the image of Q under the exponential map at the intrinsic mean IM (assumed unique) of Q. Fletcher (2004) considered an analog of PCs given by images under the exponential map of the principal components of the covariance matrix of $\tilde{Q}$. Huckemann et al. have pointed out that on manifolds (and more generally, on quotients of manifolds with respect to isometric Lie

group actions), a more natural procedure might be to define the *first generalized principal component* (GPC) as the generalized geodesic whose expected squared (pseudo-) distance from the random X (with distribution Q) is the smallest. As a justification, it has been pointed out that the intrinsic mean may not lie near this first GPC, as some examples show. Hence from the point of view of either a generalization of Euclidean PCA, or feature selection on manifolds, it may be more natural to define the first GPC as the one closest to the mass distribution Q. The *second* GPC is then defined to be the generalized geodesic that minimizes the (pseudo-) squared distance among all those intersecting the first one orthogonally. A principal *component mean* (PM) is then defined to be a point which, among all points of intersection of the first two GPCs, minimizes the expected squared (pseudo-) distance from X. The subsequent GPCs minimize expected squared (pseudo-) distance from X among all GPCs orthogonal to the preceding ones, and passing through the PM.

Huckemann et al. have shown that, outside of null sets, the GPCs are uniquely defined. In contrast, the intrinsic mean is not known to be uniquely defined except under restrictive support conditions on Q. However, the definition of the PM depends only on the first two GPCs and, therefore, seems a bit arbitrary in the general case. Of course, all Euclidean PCs go through the (intrinsic) mean; but that need not be the general model when considering manifolds. From this point of view the approach of Fletcher (2004) and Fletcher and Joshi (2007) seems more natural. Partly because of this, it is not clear to us if the so-called "total variance" and its decomposition is quite natural. By the way, is it obvious that the "geodesic variances" $V_{res}X$ "explained by the s-th GPC" are all nonnegative?

Huckeman el. al. develop innovative algorithms, to compute the GPC's by assuming that the orbifold M/G is defined in terms of a submanifold M of a Hilbert space $\mathbb{H}$ as the set of zeroes of a differentiable function $\phi : \mathbb{H} \to \mathbb{R}^k$, with a differential of maximal rank on M. They derive their computations using a two-stage minimization for the *optimal positioning* of a point $x \in M$ with respect to a geodesic on M, based on previous work by Huckemann and Ziezold (2006). Their computations are carried out in detail in the case of Kendall's shape spaces, in the line of theory developed in Small (1996). Huckeman et. al. compare their PC methods for two types of data: the "brooch data" (Small (1996)), located around a regular point on an shape orbifold, and the "fir tree" data, located a region closer to a singular point on a shape orbifold. We will call these "regular data" and "singular data". We do not agree with the authors' view that in the "brooch" example their approach is better in analyzing variability of the data. Perhaps from the point of feature selection, their approach fares better here (as it does in the "fir tree" example), but not from the point of view of dimension reduction for purposes of prediction.

A more serious issue is the complete lack of consideration of problems of statistical inference based on samples, namely, an asymptotic distribution theory for sample GPCs, and the corresponding geodesic variances.

Acknowledgements

References

Bandulasiri, A. and Patrangenaru, V. (2005). Algorithms for nonparametric inference on shape manifolds, *Proceedings of the Joint Statistical Meetings* 2005, *Minneapolis, MN*, 1617-1622.

Bandulasiri, A., Bhattacharya, R. N. and Patrangenaru, V. (2009). Nonparametric inference for extrinsic means on size-and-shape manifolds with applications in medical imaging. *J. Multivariate Anal.* **100**, 1867-1882.

Bhattacharya, A. (2009). Statistical analysis on manifolds: A nonparametric approach for inference on shape spaces. *Sankhyā, Ser. A.* in press.

Bhattacharya, R. N. and Bhattacharya, A. (2008). Statistics on Riemannian manifolds: Asymptotic distribution and curvature. *Proc. Amer. Math. Soc.* **136**, 2957-2967.

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

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

Bookstein, F. L. (1978). *The Measurement of Biological Shape and Shape Change.* Lecture Notes in Biomathematics. **24**. Springer, Berlin.

Dryden, I. L., Kume, A., Le, H. and Wood, A. T. A. (2008). A multidimensional scaling approach to shape analysis. *Biometrika* **95**, 779-798.

Dryden, I. L. and Mardia, K. V. (1998). *Statistical Shape Analysis.* Wiley, Chichester.

Fletcher, P. T. (2004). Statistical variability in nonlinear spaces: Application to shape analysis and DT-MRI. Ph.D Dissertation, Computer Science, University of North Carolina, Chapel Hill.

Fletcher, P. T. and Joshi, S. C. (2007). Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Processing* **87**, 250-262.

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

Hendriks, H. and Landsman, Z. (1998). Mean location and sample mean location on manifolds: asymptotics, tests, confidence regions. *J. Multivariate Anal.* **67** , 227-243.

Hendriks, H. and Landsman, Z. (2007). Asymptotic data analysis on manifolds. *Ann. Statist.* **35**, 109-131.

Huckemann, S. and Ziezold, H. (2006). Principal component analysis for Riemannian manifolds with an application to triangular shape spaces.*Ann. Appl. Probab.* **38**, 299-319.

Hufnagel, H., Pennec, X., Ehrhardt, J., Ayache, N. and Handels, H. (2008). Generation of a statistical shape model with probabilistic point correspondences and EM-ICP. *International Journal for Computer Assisted Radiology and Surgery* **2** 5 , 265-273.

Karcher, H. (1977). Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math.* XXX, 509-541.

Kendall, D. G. (1977). The diffusion of shape. *Ann. Appl. Probab.* **2**, 428-430.

Kendall, D. G. (1984). Shape manifolds, Procrustean metrics, and complex projective spaces. *Bull. London Math. Soc.* **16**, 81-121.

Kendall, D. G., Barden, D., Carne, T. K. and Le, H. (1999). *Shape and Shape Theory.* Wiley, New York.

Kent, J. T. (1992). New directions in shape analysis. *The Art of Statistical Science*, 115-127. Wiley.

Mardia, V. and Patrangenaru, V. (2005). Directions and projective shapes. *Ann. Statist.* **33**, 1666-1699.

Munk, A., Paige, R., Pang, J., Patrangenaru, V. and Ruymgaart, F. H. (2008). The one and multisample problem for functional data with applications to projective shape analysis. *J. Multivariate Anal.* **99**, 815-833.

Patrangenaru, V., Liu, X. and Sugathadasa, S. (2010). Nonparametric 3D projective shape estimation from Pairs of 2D Images - I, in memory of W.P. Dayawansa. *J. Multivariate Anal.* **101**, 11-31

Prentice, M. J. (1984). A distribution-free method of interval estimation for unsigned directional data. *Biometrika* **71**, 147-154.

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

University of Arizona, Department of Mathematics, Tucson, AZ 85721, U.S.A.

E-mail: rabi@math.arizona.edu

COMMENT

Sungkyu Jung, Mark Foskey and J. S. Marron

The University of North Carolina at Chapel Hill

This paper proposes a novel PCA scheme for variables on a manifold. The authors tackled an interesting problem. In particular when the mass of a random quantity or a set of data is mostly aligned along a sufficiently long arc of a geodesic, the intrinsic and extrinsic mean no longer provide a useful notion of the center of mass due to the curvature of the manifold. The authors defined a new notion of mean, the Principal Component mean (PM), as the point where the first and second geodesic PCs meet. Regardless of the curvature of the manifold, this Geodesic PCA (GPCA) works well when most of the mass is along a geodesic.
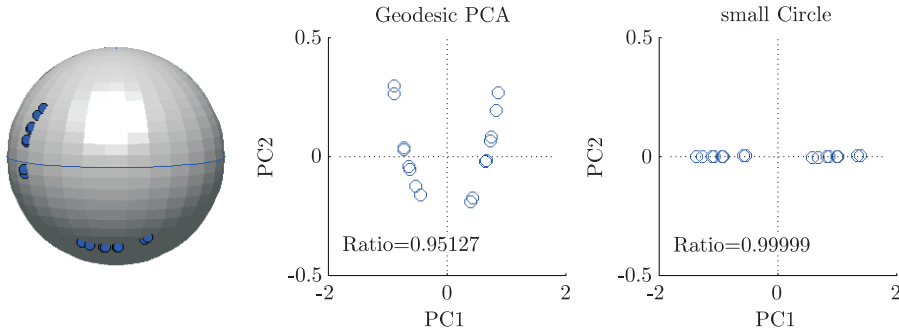
Figure 1. Sampling variation of the S^2-valued variable in m-reps and projections on various PCs. (i) The data points are mostly along a small circle, which is not a geodesic. (ii) The method of GPCA: A set of data in (i) is projected to the tangent space at PM together with the two Geodesic PCs. (iii) Alternative method: The data set is projected on circles δ_1 and δ_2. The proportion of variance explained by the first PC is shown as *ratios*.

The manifold data we encounter are typically S^2 (and $(S^2)^d$). In the data that we work with, we have noticed a tendency towards variation along smaller circles than geodesics, so we point out variations of the PCA method of this paper that better capture variation along a non-geodesic curve on manifolds.

The medial representations of shapes (m-reps, see e.g., Siddiqi and Pizer (2008), Fletcher, Lu, Pizer and Joshi (2004)) lie in direct products of S^2, S^1, R, and R^+. In particular, in an example of randomly generated m-reps, that resemble human organs, from a simulator discussed in Jeong, Stough, Marron and Pizer (2008), we have observed S^2-valued variables distributed along a small circle, which is not a geodesic (left panel of Figure 1.) The projection of the S^2-valued data on geodesic PCs (center panel of Figure 1) results in a curvy form, where the first PC captures only approximately 95% of the variation. The method of PCA at IM (Fletcher et al. (2004)) also shows similar performance for this data set. We suggest an improvement, due to Gray, Geiser and Geiser (1980), that captures 100% of the variation of the first PC. This allows a non-geodesic curve to be a principal curve.

To illustrate this idea, let the simple manifold S^2 be our space. Let $x_1, \ldots, x_n \in S^2$ be the data points. We wish to consider the principal modes of variation that are not necessarily along a geodesic, where the constraint of δ_1 being geodesics is relaxed to include all circles. Then a small circle δ_1 minimizing residual variance $\sum_i d(x_i, \delta_1)^2$, where d is the metric on S^2, is defined to be the first mode of variation, i.e., the first PC. Furthermore, in the spirit of Definition 3.1, sequentially find δ_2 to minimize $\sum_i d(x_i, \delta_2)^2$ subject to orthogonally crossing δ_1. This entails

that δ_2 is a geodesic. Then we can define a new center of mass as the intersection of δ_1 and δ_2, say p, which minimizes $\sum_i d(x_i, p)^2$. The last plot in Figure 1 shows the projection of the data onto δ_1 and δ_2, where the projection on δ_2 is not an orthogonal projection, but a projection along the direction of δ_1.

These ideas, including extension to important manifolds such as $(S^2)^d$, will be further developed in Jung, Foskey and Marron (2009).

References

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

Jeong, J.-Y., Stough, J. V., Marron, J. S. and Pizer, S. M. (2008). Conditional-mean initialization using neighboring objects in deformable model segmentation., *SPIE Medical Imaging*, February.

Jung, S., Foskey, M. and Marron, J. S. (2009). Principal arc analysis on direct product manifolds, manuscript in progress.

Gray, N. H., Geiser, P. A. and Geiser, J. R. (1980). On the least-squares fit of small and great circles to spherically projected orientation data. *Mathematical Geology* **12**, 173-184.

Siddiqi, K. and Pizer, S. (2008). *Medial Representations: Mathematics, Algorithms and Applications*. Springer.

Department of Statistics and Operations Research, The University of North Carolina at Chapel Hill, Chapel Hill, NC27514, U.S.A.

E-mail: sungkyu@email.unc.edu

Department of Radiation Oncology, The University of North Carolina at Chapel Hill, Chapel Hill, NC27514, U.S.A.

E-mail: mark_foskey@unc.edu

Department of Statistics and Operations Research, The University of North Carolina at Chapel Hill, Chapel Hill, NC27514, U.S.A.

E-mail: marron@email.unc.edu

COMMENT

John T. Kent

University of Leeds

1. Introduction

Modern statistical shape analysis started with the seminal papers by Kendall (1984) and Bookstein (1986). The subject has greatly developed over the past

20+ years and in many ways is now a mature area of Statistics, especially landmark-based shape analysis. It has applications spanning biology (morphometrics), computer vision, robotics, archaeology, and medical imaging. However, as demonstrated in the paper by Huckemann, Hotz and Munk (2009) (referred to below as HHM), there are still substantial issues to be tackled in shape analysis.

In this contribution, we look at some of the suggestions in HHM for principal component analysis (PCA) in more detail. First, these ideas are made more concrete through some simple illustrations, and then their implications for the tree crown data are explored through a re-analysis of that data. Finally some general points are made about the statistical modelling of shape data.

One of the main points made by HHM is that on manifolds and related spaces, dimension reduction is more subtle than it is in Euclidean space. In Euclidean space there is just one way to do PCA: find the mean vector and the covariance matrix of the data and represent the variability about the mean in one (or a small number of) dimensions.

On manifolds there are two distinct idealized types of PCA, which can be illustrated most easily on the usual unit sphere in $\mathbb{R}^3$.

- Type I. This type of PCA mimics Euclidean PCA. The data are concentrated about a mean direction on the sphere with a roughly elliptical pattern of variability. The more eccentric the ellipse, the more successfully can the data be reduced to a one-dimensional summary.
- Type II. This type of PCA has no equivalent in Euclidean PCA. The data are concentrated on or near a great circle such as the equator and are uniformly distributed around this circle. In this case there is no mean direction; but, treating a great circle as a one-dimensional object, there is again a reduction to a one-dimensional summary of the data.

Figure 1 illustrates the two types of behavior. HHM point out that small circles about a singularity point in shape space Σ_m^k, $m \geq 3$, will form geodesics. Thus they effectively argue that if data are clustered about a singularity point, it may be the case that Type II PCA will be more appropriate than a Type I PCA. See the next section for an example.

2. Application to Tree Crowns

For the "tree crown" data, each configuration can be represented as a 3×4 matrix of 3 coordinates for 4 landmarks. The first row represents the vertical component z of each landmark, and the second and third rows represent the east-west and north-south directions, x and y, say. The four columns represent the four landmarks. The first landmark is the crown itelf, assumed to lie on the vertical axis $x = y = 0$, say, of the tree, and the remaining 3 landmarks lie in a

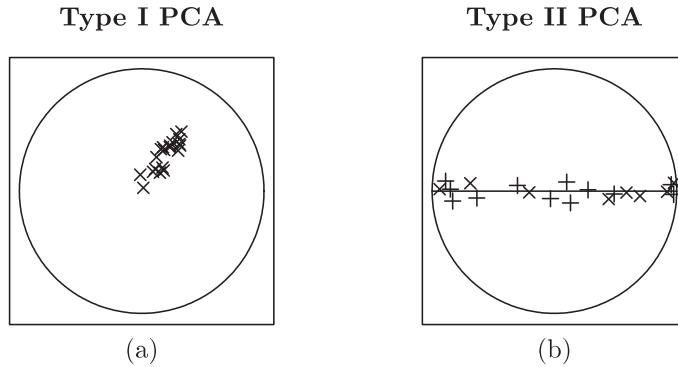**Type I PCA**　　　　　　　**Type II PCA**

(a)　　　　　　　　　　　(b)

Figure 1. Two examples of data on the sphere: (a) data concentrated with an elliptical pattern of variation; (b) data concentrated near the equator with a uniform distribution around the equator, where "×" denotes data on the front of the sphere, and "+" data on the back.

horizontal planar cross-section of the trunk $z = 0$, say, given by the center of the tree rings and points on the circumference of the trunk minimally and maximally distant from the center, respectively.

This description is already standardized in terms of location and partially standardized in terms of rotation. To finish the standardization of rotation, rotate the horizontal plane $z = 0$ about the origin so that the landmark representing the minimal radius of the trunk lies on the positive x axis (with $y = 0$). Finally scale the configuration so that the crown has vertical value $z = 1$. This standardization is not quite a Procrustes registration: in particular, the centroid is not at the origin and the centroid size is not quite equal to 1. However, it lets us understand the structure of the data more intuitively.

Thus the shape of each configuration can be represented as a matrix

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & \epsilon & \delta_1 \\ 0 & 0 & 0 & \delta_2 \end{bmatrix},$$

where $\epsilon > 0$, and where ϵ, δ_1, δ_2 are small and vary from one tree to another.

There are $n = 5$ configurations in the dataset. Plots of landmarks 2, 3, 4 in the plane $z = 0$ are shown in Figure 2. Note that the angle at the origin between landmarks 3 and 4 is clockwise for configurations 1, 2, 5 and counterclockwise for configurations 3, 4.

Figure 3(a) shows the first two principal component scores for a PCA in Procrustes tangent coordinates. As noted by HHM, the proportion of variance
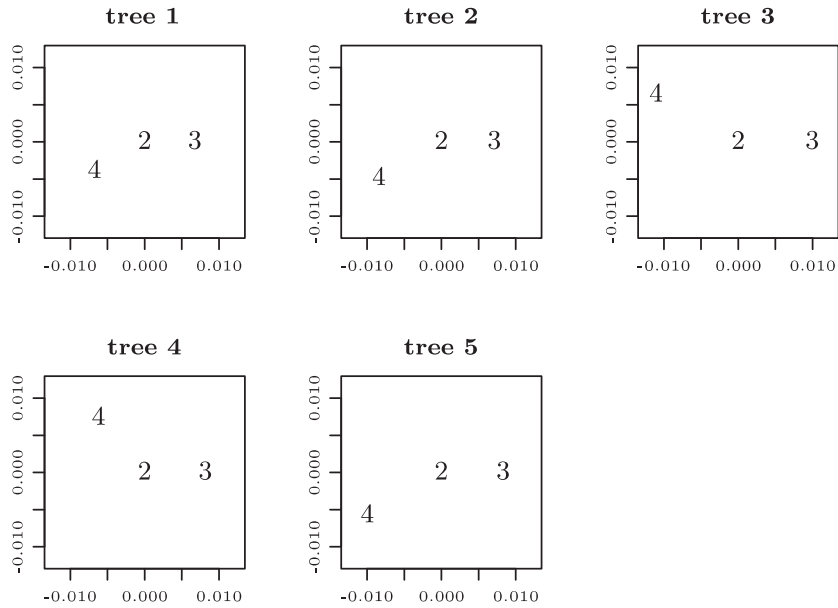
Figure 2. Cross-section of the trunk for each of the 5 trees. Landmark 2 lies at the center of the tree; landmark 3 indicates the shortest radius to the circumference of the tree, rotated to lie on the positive horizontal axis; landmark 4 lies on the longest radius.
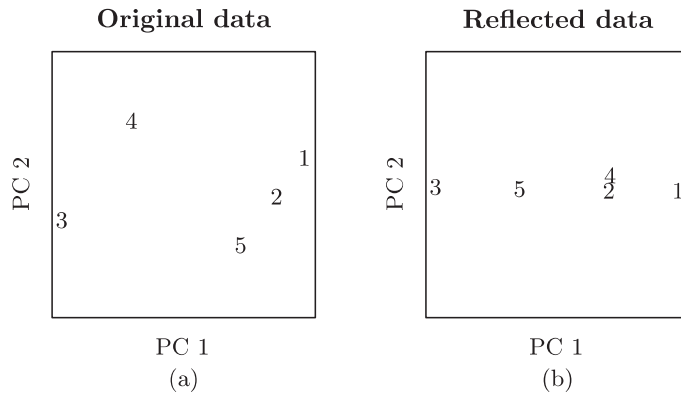

Figure 3. Principal component scores for the tree crown data based Procrustes tangent coordinates: (a) original data; (b) data modified by reflecting configurations 3 and 4.

explained by the first two components is $67.3 + 32.4 = 99.8\%$. Note that configurations 1,2,5 approximately lie on one line, and configurations 3 and 4 on a

parallel line. With a bit more imagination we can regard all the configurations as lying approximately uniformly dispersed around the circumference of a circle centered at the origin. Thus, we can see why a Type I PCA needs two components to explain the data, whereas a Type II PCA needs only 1 principal component. These observations help us understand why HHM have recommended a Type II PCA for this dataset.

3. Reservations about Shape Analysis on Tree Crown Data

The tree crown example provides a nice illustration of what can happen near a singular point in shape space. However, there are several limitations in using this example as a serious example of statistical shape analysis.

(a) *Data are not configurations.* Shape analysis is not really appropriate for this sort of data. In particular, the "crown" is not a true 3-dimensional landmark, but only a one-dimensional height.

(b) *Form vs. shape.* Further, removing size removes the relationship between crown height and trunk measurements. A study of form (= size + shape) is likely to be of more scientific interest than a study of shape.

(c) *Reflection shape.* Two of the configurations (3 and 4) are roughly reflections of the other three (1, 2, and 5). In particular, the angles between the maximal and minimal radii in degrees are $-150°$, $-150°$, $150°$, $130°$, $-150°$, respectively. Aligning configurations with respect to reflection as well as similarity transformations removes much of the variability seen in the data and removes the need for a Type II PCA. A Type I PCA with one retained component now suffices; see Figure 3(b). Further the PC score is highly correlated with relative trunk size ($r = -0.99$), where relative trunk size is defined by $\sqrt{\epsilon^2 + \delta_1^2 + \delta_2^2}$.

4. Data Analysis vs. Statistical Models

The paper focuses on *data analysis* — dimension reduction in the context of a particular set of data. Alternatively, we can look at *statistical models* which accommodate some degree of concentration in lower dimensions. Note that the sphere is an important special case of shape space since, up to a factor of 2, it can be identified with the shape space Σ_2^3 of triangles in the plane.

- Sphere, Type I PCA: FB5 distribution (analogous to bivariate normal).
- Sphere, Type II PCA: Bingham distribution (girdle case).
- Σ_m^k, $m = 2$, Type I PCA: CBQ distribution (analogous to multivariate normal).

- $m > 2$ (Types I and II models) or $m \geq 2$ (Type II models): more work to be done.

Details about standard distributions for directional data, such as the 5-parameter Fisher-Bingham (FB5) and the Bingham distributions, can be found in Fisher, Lewis and Embleton (1987) and Mardia and Jupp (1999). See Kent, Mardia and McDonnell (2006) for the complex Bingham-quartic (CBQ) distribution.

5. Longitudinal (Growth) Data

HHM focus on samples of independent, identically distributed data, where only the scatter of the data is of interest. However, for longitudinal data, the path through shape space is also important. HMM suggest such data will often lie on a geodesic, but such models are too simplistic in many practical situations.

For example, consider a set of rat growth data in the plane $\mathbb{R}^2$, described and analyzed in Bookstein (1991) and Kent, Mardia, Morris and Aykroyd (2001). The data are obtained from a two-dimensional midsagittal section of the calvarium (the skull without the lower jaw). There is complete information on $N = 18$ rats at $H = 8$ times (or ages) on $K = 8$ landmarks. To facilitate the model fitting, we replace the actual age by a "pseudo-age" given by the average centroid size at each age and ignore any differences between the individual rats.

One possible growth model in Procrustes tangent space takes the tensor product form

$$v_{jt} = \sum_{\alpha=1}^{p} \sum_{\beta=1}^{q} a_{\alpha\beta} f_{(\alpha)}(\mu_j) g_{(\beta)}(t).$$

Here $v_{jt} \in \mathbb{R}^2$ denotes the Procrustes tangent coordinates for landmark j at time t about a mean configuration μ with landmarks μ_j. The functions $f_{(\alpha)}$ and $g_{(\beta)}$ represent modes of change in space and time. In this example we used principal splines; these functions are analogous to polynomials of increasing degree. In particular, the first two principal splines in space correspond to a linear transformation of the plane. The first principal spline in time corresponds to linear growth, and the second principal spline in time is approximately quadratic.

Figure 4 illustrates an analysis of the rat data, with the growth in parts (a), (b), and (c) exaggerated by a factor of 5 for visibility. Note the raw data in (a) show curvature in their growth paths at each landmark. A geodesic in shape space (part (b)) implies a constant speed straight line growth path at each landmark which does not capture this curvature. The best-fitting model, labelled a "special model", is illustrated in part (c) and can be described as

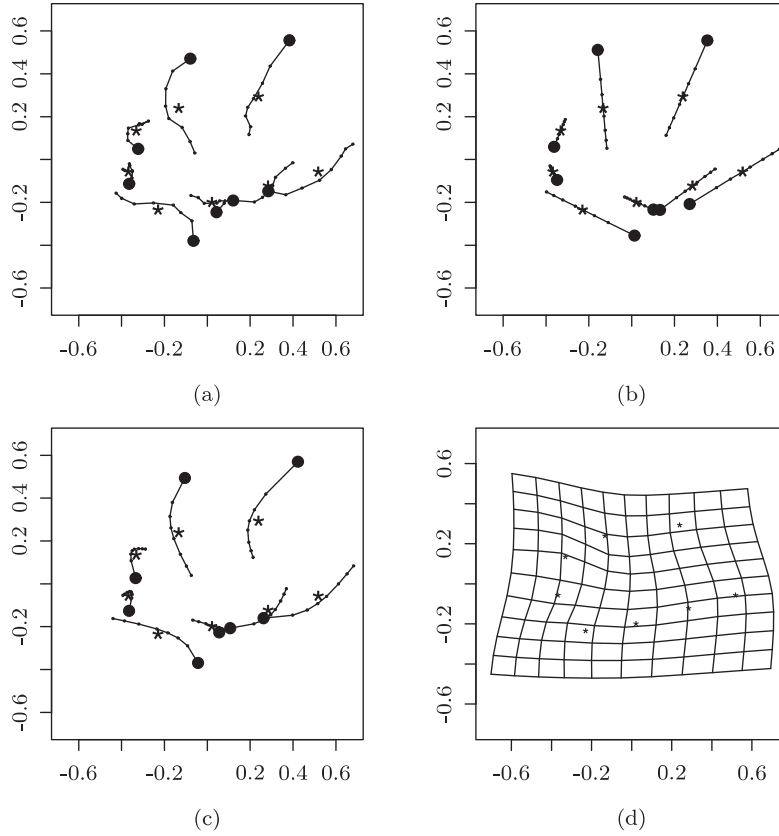(general space $\times$ linear time) + (linear space) $\times$ (quadratic time).

Figure 4. Analysis of the rat data: (a) raw data, (b) geodesic growth model, (c) special growth model. Each "*" represents a landmark of the the Procrustes mean shape μ, and each closed circle represents the position of a landmark at the initial time. Part (d) shows the grid deformation, without an expansion factor, between the initial and final times for the special model. For parts (a), (b), (c), the growth patterns have been expanded by a factor of 5 for clarity.

6. Conclusions

This has been a fascinating paper. New types of data on manifolds and quotient spaces are of growing interest. Dimension reduction, especially near singularities in shape space, provides interesting challenges. In this setting it is important to emphasize the distinction between Type I and Type II PCA.

References

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

Bookstein, F. L. (1991). *Morphometric Tools for Landmark Data.* Cambridge University Press.

Fisher, N. I., Lewis, T. and Embleton, B. J. J. (1987). *Statistical Analysis of Spherical Data.* Cambridge University Press.

Huckemann, S., Hotz, T. and Munk, A. (2009). Intrinsic shape analysis: geodesic PCA for Riemannian manifolds modulo isometric Lie group actions. *Statist. Sinica* **20**, 1-100.

Kendall, D. G. (1984). Shape manifolds, procrustean metrics and complex projective spaces. *Bull. London Math. Soc.* **16**, 81-121.

Kent, J. T., Mardia, K. V., Morris, R. J. and Aykroyd, R. G. (2001). Functional models of growth for landmark data. In *Proceedings in Functional and Spatial Data Analysis*, (Edited by K. V. Mardia and R. G. Aykroyd), 109-115, Leeds University Press.

Kent, J. T., Mardia, K. V. and McDonnell, P. (2006). The complex Bingham quartic distribution and shape analysis. *J. Roy. Statist. Soc. Ser. B* **68**, 747-765.

Mardia, K. V. and Jupp, P. E. (1999). *Directional Statistics.* Wiley, Chichester.

Department of Statistics, University of Leeds, Leeds LS2 9JT, UK.

E-mail: j.t.kent@leeds.ac.uk

COMMENT

Peter T. Kim and Ja-Yong Koo

University of Guelph and Korea University

Modern statistics is undergoing an evolution as data are being realized in ways that have not been realized before. The current paper under discussion is an example of such an evolution and, therefore, as a profession we have to face such challenges. What we would like to bring forth in this discussion is whether statistics can adapt to this evolution by keeping some elements of how the subject has functioned in the past while, at the same time, being amenable to modern data that are increasingly complicated. In particular, we would like to note that the normal distribution is central to statistics and permeates the teaching of statistics through the fundamental frequentist notion, i.e., the Central Limit Theorem, and as a way of performing statistical estimation through maximum likelihood estimation. This is certainly well-defined in the case of Euclidean space, but once we leave such a structure, the situation quickly becomes unclear.

This is apparent in directional statistics where the sample space is the unit sphere. Statisticians have looked for reasonable ways of expressing the "normal"

distribution for estimation and inference on the sphere. Indeed, early attempts by Fisher (1953) led to the von Mises–Fisher distribution; this resulted in many papers exploring the idea of a mean direction and concentration. Other directional distributions have been introduced to model various peculiarities associated with directional data and a very compelling formalism was put forward by Beran (1979) in which the author demonstrates that all these distributions are special cases of exponentiating spherical harmonic expansions on the sphere. Nevertheless, the idea of a normal distribution was not satisfactorily resolved. A formal treatment was explored in Hartman and Watson (1974) where this very question is detailed in the title of their paper. What we learn from there is that if we define a normal, or Gaussian, distribution on the sphere to mean a diffusion on the sphere, then the von-Mises–Fisher distribution is not it. In fact, Kim and Koo (2000) demonstrates in what way they are different.

We ask what one could mean by the normal distribution so that it fits the classical definition and can adapt itself to a more modern environment. We begin with the Euclidean case and use the notation $G_t(x)$ to mean the Gaussian density with mean 0, and variance t. The Gaussian density satisfies the so called heat equation

$$\frac{\partial}{\partial t} = \frac{1}{2}\frac{\partial^2}{\partial x^2}, \qquad (1.1)$$

where $t > 0$ and $x \in \mathbb{R}$. A solution to (1.1) is a function $u : (0, \infty) \times \mathbb{R} \to \mathbb{R}$ and $u(x, 0) = f(x)$ is the initial condition. It is then straightforward to show that the solution has the form $u(x, t) = G_t * f(x)$ where, for two functions f and h, the convolution is $f * h(x) = \int f(x - y)h(y)dy$. If we then specify $f(x) = \delta_\mu(x)$, where δ_μ is the Dirac delta function centered at μ, then the symbolic way of writing a normal distribution with mean $\mu \in \mathbb{R}$ and $\sigma > 0$ as $N(\mu, \sigma)$, is $G_{\sigma^2} * \delta_\mu(x)$.

The above is the argument used in Hartman and Watson (1974) for the case of the sphere and it can be extended to more general structures. Consider the same question on a Riemannian manifold $\mathbb{M}$. By following their implicit argument, we are led to the Laplacian

$$\Delta = \frac{1}{\sqrt{|g|}} \sum_{j,k} \partial_j (g^{jk} \sqrt{|g|} \, \partial_k), \qquad (1.2)$$

where ∂_j is in reference to the partial derivative with respect to the $j-$th coordinate, $g = (g_{jk})$ are the metric tensors, whose inverse is written (g^{jk}), and $|g|$ is its determinant for a fixed $x \in \mathbb{M}$. The corresponding heat equation on $\mathbb{M}$ is

$$\frac{\partial}{\partial t} = \frac{1}{2}\Delta, \qquad (1.3)$$

where again we have as its fundamental solution the Gaussian density $G_t(x)$, for $x \in \mathbb{M}$. Now if the manifold is equipped with a group structure so that it

is a Lie group, as in the paper under discussion, then one can further mimic the Euclidean version. In particular consider a solution $u : (0, \infty) \times \mathbb{M} \to \mathbb{M}$ with initial condition $u(x, 0) = f(x)$. The solution can be written as $G_t * f(x)$, where here for f and h, the convolution is $f * h(x) = \int f(xy^{-1})h(y)dy$ to reflect the possible non-commutative group structure. Therefore on a Lie group we can define the normal distribution with unknown mean $\mu \in \mathbb{M}$ and $\sigma > 0$ by $G_{\sigma^2} * \delta_\mu(x)$, where now $\delta_\mu(x)$ is the Dirac delta function on $\mathbb{M}$ centered at $\mu \in \mathbb{M}$.

Can we live with such a definition? First the fact that we need a group structure may seem restrictive, a second point is that we do not know whether there are central limit theorems available, and third, there is the question of how to do estimation, in particular, maximum likelihood estimation.

As a short answer to the first, we note that most manifolds we encounter are homogeneous spaces of Lie groups. In fact if one has a transitive group of isometries on a manifold, the manifold is a homogeneous space of the Lie group. In fact, the sphere is a homogeneous space of the Lie group of rotation matrices. Consequently, one can work over the covering space and project down to the homogeneous space where things should be manageable.

As to the second issue there have been past attempts to generalize central limit theorems in ways so that the limiting distribution is the Gaussian density as described in the more general setting. This has mainly centered on the space of positive definite symmetric matrices that can be realized as the homogeneous space of $\mathbb{GL}(m)$, the Lie group of $m \times m$ non-singular matrices, modulo $\mathbb{O}(m)$, the $m \times m$ group of orthogonal matrices. By taking matrix products of iid random matrices and spectrally rescaling by $\sqrt{n}$, one can show convergence in distribution to the Gaussian distribution on $\mathbb{GL}(m)/\mathbb{O}(m)$. Consequently, non-Euclidean central limit theorems abound. See Terras (1988), Richards (1988) and Graczyk (1992).

Perhaps the most difficult question is how does one do maximum likelihood estimation. What is crucial to realize in this question is the rather delicate point that, although $G_{\sigma^2} * \delta_\mu$ can be written down in closed form for $\mathbb{R}$, such is not the case for a general $\mathbb{M}$. The only characterization of the Gaussian density that permeates across different structures is the notion that if $\{(\varphi_\lambda, \lambda), \lambda \in \Lambda\}$ is the spectral resolution of the Laplacian Δ, i.e., $\Delta \varphi_\lambda = \lambda \varphi_\lambda$, then the Fourier transform of the Gaussian density has a specific formulation,

$$\widehat{G}_t(\lambda) = \int_{\mathbb{M}} G_t(x)\overline{\varphi_\lambda(x)}dx = e^{(\lambda/2)t}, \tag{1.4}$$

where $\lambda \in \Lambda$, and overline means complex conjugation. Furthermore, under certain conditions for two functions $f, h : \mathbb{M} \to \mathbb{R}$,

$$\widehat{f * h}(\lambda) = \widehat{f}(\lambda)\,\widehat{h}(\lambda). \tag{1.5}$$

Thus in the case of what we are calling the Gaussian density

$$\widehat{G_{\sigma^2} * \delta_\mu}(\lambda) = \widehat{G_{\sigma^2}}(\lambda)\,\widehat{\delta_\mu}(\lambda) = e^{(\lambda/2)\sigma^2}\overline{\varphi_\lambda(\mu)}. \tag{1.6}$$

Let us see if this makes sense in the more familar Euclidean setting of $\mathbb{R}$. In particular $\Delta = \partial^2/\partial x^2$ and $\varphi_\lambda(x) = e^{-ikx}$ for $k \in \mathbb{R}$. Now $\Delta e^{-ikx} = -k^2 e^{-ikx}$, therefore $\lambda = -k^2$, for $k \in \mathbb{R}$. Applying this to (1.6), it turns out that

$$\widehat{G_{\sigma^2} * \delta_\mu}(\lambda) = e^{ik\mu - (k^2/2)\sigma^2}, \tag{1.7}$$

for $k \in \mathbb{R}$, which is precisely the characteristic function of $G_t * \delta_\mu(x)$ or a $N(\mu, \sigma)$ probability density.

Now based on the fact that if we have a random sample $X_1, \ldots, X_n$ from the Gaussian density $G_{\sigma^2} * \delta_\mu$, one can form the empirical characteristic function

$$\widehat{G}_{\sigma^2}^n(\lambda) = \frac{1}{n}\sum_{j=1}^n \overline{\varphi_\lambda(X_j)} \tag{1.8}$$

for $\lambda \in \Lambda$ along the lines of what is done for the Euclidean case, see Bhattacharya and Patrangenaru (2003). This then allows one to construct maximum likelihood estimation procedures based on the empirical characteristic function, and this is laid out in Beran and Fisher (1998); Bhattacharya (2008), Bandulasiri, Bhattacharya and Patrangenaru (2009) and more generally in Anotoniadis, Feuerverger and Goncalves (2006). Therefore it is evident that, even though we cannot write the normal distribution down in closed form on $\mathbb{M}$, it is still possible to have that concept with unknown mean and variance, and to perform maximum likelihood estimation through the empirical characteristic function. One can then consider issues related to such things as efficiency.

In summarizing this discussion, it is encouraging to see that statistical techniques beyond the case of Euclidean space are gaining momentum. As in the paper being discussed, they are contributing to the evolution of statistics. Our discussion is an attempt to bridge the evolving nature of statistics with some "bread and butter" tools of statistics in the hope that statistics can adapt to the changing environment.

References

Anotoniadis, A., Feuerverger, A. and Goncalves, P. (2006). Wavelet based estimation for univariate stable laws. *Ann. Inst. Statist. Math.* **58**, 779-807.

Beran, R. (1979). Exponential models for directional data. *Ann. Statist.* **6**, 1162-1178.

Feuerverger, A. (1990). An efficiency result for the empirical characteristic function in stationary time series models. *Canad. J. Statist.* **18**, 155-161.

Feuerverger, A. and McDunnough, P. (1981a). On some Fourier methods of inference. *J. Amer. Statist. Assoc.* **76**, 379-387.

Feuerverger, A. and McDunnough, P. (1981b). On the efficiency of empirical characteristic function procedures. *J. Roy. Statist. Soc. Ser. B* **43**, 20-27.

Feuerverger, A. and Mureika, R. A. (1977). The empirical characteristic function and its applications. *Ann. Statist.* **5**, 88-97.

Fisher, R. A. (1953). Dispersion on a sphere. *Proc. Roy. Soc. Ser. A* **217**, 295-300.

Graczyk, P. (1992). A central limit theorem on the space of positive definite symmetric matrices. *Ann. Inst. Fourier* (*Grenoble*) **42**, 857-874.

Hartman, P. and Watson, G. (1974). "Normal" distribution functions on spheres and the modified Bessel functions. *Ann. Probab.* **2**, 593-607.

Kim, P. T., Koo, J. Y. (2000). Directional mixture models and optimal estimation of the mixing density. *Canad. J. Statist.* **28**, 383-398.

Richards, D. St. P. (1988). The central limit theorem on spaces of positive definite matrices. *J. Multivariate Anal.* **29**, 326-332.

Terras, A. (1988). *Harmonic Analysis of Symmmetric Spaces and Applications II.* Springer-Verlag, New York.

Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario N1G 2W1 Canada.

E-mail: pkim@uoguelph.ca

Department of Statistics, Korea University, Anam-Dong Sungbuk-Ku, Seoul 136-701, Korea.

E-mail: jykoo@korea.ac.kr

COMMENT

Manik C Mukherjee and Atanu Biswas

Netajinagar Vidyamandir and Indian Statistical Institute

In the present discussion we discuss the transition from principal component analysis (PCA) to geodesic principal component analysis (GPCA) – the path of the journey from geometry to Lie group topology.

A primary goal of statistical shape analysis is to describe the variability of a population of geometric objects. A standard technique for computing such descriptions is PCA. However, as PCA is limited in that it only works for data lying in a Euclidean vector space, this is certainly sufficient for geometric models that

are parameterized by a set of landmarks or a dense collection of boundary points, but it does not handle more complex representations of shape. PCA was extended to nonlinear PCA (NLPCA) in different directions (see, e.g., Kramer (1991), and Lawrence (2005)), but still that is under the usual Euclidean geometry set-up. Principal geodesic analysis is a generalization of PCA to a non-Euclidean, non-linear setting of manifolds suitable for use with shape descriptors such as medical representations. Huckemann, Hotz and Munk provided an extremely informative discussion on various aspects of GPCA including the procedure to find different ordered GPCs. They illustrated that their approach outperfoms the usual PCA. A notable ommission in their study is any resampling technique. Dam, Fletcher, Pizer, Tracton and Rosenman (2004) carried out prostate shape modeling based on principal geodesic analysis bootstrapping. Surely, resampling is a practical possibility, and it would be quite useful to find some comparison with that, along with some theoretical study.

Fletcher, Lu, Pizer and Joshi (2004) laid an excellent platform to discuss the GPCA. The underlying philosophy, set-up, analysis, and the required mathematics are very different!! Huckemann, Hotz and Munk explained basic mathematical preliminaries in great details but, as a reader, one would like to have not only the changed technology but an account of the flow of that change that might help a statistician appreciate the use of topology in GPCA. It is not only about Lie groups as a tool, the philosophy should be blended with the philosophy of GPCA. Next we focus on going from geometry to topology in order to appreciate the journey from PCA to GPCA.

Mathematics is the language of 'Nature'. Biostatistics mainly deals with the application of different statistical methods to data collected from a biological frame of reference in a systematic way. We always try to handle the collected data through a rigorous analytical system. 'Mathematical Analysis' (real, complex, functional, numerical) wants a minute observation and scrutiny on the desired results. For this purpose, knowledge of Calculus, algebra (linear and abstract), and geometry are essentially needed. In order to analyze the characteristics of a sample in sample space we first have to locate its position, and hence geometrical sense is required. The basic concepts of Euclidean space helps. As research is a continuous process, our views are extended from one stage to another and this is done by different kinds of operators (raising/maximizing or lowering/minimizing) transformations (viz. translation, shifting, rotation or helocoidal (mixed) form). This comes from the notion of orphism (homomorphism, isomorphism, auto-morphism, heomorphism, diffeomorphism, etc.). From finer to finest form, we are moving from Euclidean space to non-Euclidean space, Riemannian space to non-Riemannian space.

In order to broaden our outlook we give more emphasis on 'structure'. So the place of geometry is occupied by topology. To measure the distance between two

objects 'metric' has appeared. As a consequence, we observe that by means of the usual metric a topology on $\mathcal{R}^n$ (n-dimensional real Euclidean space) is defined that allows us to regard it as a Hausdorff space (where distinct points will get distinct images). We consider a topological space and then we take a non-empty open subspace of $\mathcal{R}^n$. This homomorphism is called a 'chart' in the topological subspace. A topological space is called locally Euclidean at a point if there exists a chart on a neighborhood of that point. So a Hausdorff space which is locally Euclidean at each point is called a 'manifold'. Evidently, homeomorphism gives us the guarantee of continuity. The notion of analytic function (smooth) comes which is infinitely differentiable. In this way in a topological space the concept of continuity, differentiability, connectedness, etc., are developed. On a curved surface the smallest distance between two points is calculated by 'geodesic'.

For analyzing GPCA, the idea of manifold is not sufficient. We extend it into differentiable manifold just to make our journey complete through continuity. Lie groups appear in differential manifolds with algebric form. In our literature, the tree of differential geometry which takes birth from the basic ideas of two pioneers, Gauss and Riemann, has blossomed with many beautiful flowers on different branches of tensor analysis (Riemannian, Ricci, Einstein tensor, etc.).

The concepts of tangent bundle, fiber (path) bundle are developed in the tangent space of $\mathcal{R}^n$. The tangent bundle of a manifold, embedded manifold, and c^∞ manifold are explained by means of projection maps (orthogonal) by many eminent mathematicians. In the modern phase, ideas of Euclidean classical concept are developed by new techniques and notations. These tools (both theoretical and practical) are frequently used in stochastic processes also. In Abraham and Marsden (1978), applications of differential geometry are available to explain the uniqueness of the intrinsic and extrinsic mean as measures of location of probability measures on Riemannian manifolds. Asymptotic dispersions are obtained for distribution on the direction spaces, real projective spaces, and planar shape spaces accepting the notion of PCA on a linear system of Euclidean data. But in Ambartzumian (1990) mainly, the idea was extended by GPCA on nonlinear systems with high curvature.

We finally observe that from Darwin to Kendall shape analysis, from Dalton to Dirac space, everywhere accepts the new thoughts and ideas which are cast on the screen of development of our society. Theoretical and classical methods are well-served by the new practical approach. Today we are echoing the Aristotelian postulate on Euclidean space: what applies to the whole applies also to the parts.

References

Abraham, R. and Marsden, J. E. (1978). *Foundations of Mechanics*, 2nd edition. Benjamin-Cummings.

Ambartzumian, R. V. (1990). *Factorization, Calculus and Geometric Probability*. Cambridge University Press.

Dam, E., Fletcher, P. T., Pizer, S. M., Tracton, G. and Rosenman, J. (2004). Prostate shape modeling based on principal geodesic analysis bootstrapping. In *Medical Image Computing and Computer-Assisted Intervention MICCAI* 2004, 1008-1016. Springer, Berlin/Heidelberg.

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

Kramer, M. A. (1991). Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal* **37**, 233-243.

Lawrence, N. (2005). Probabilistic non-linear principal component analysis with Gaussian process latent variable models. *J. Mach. Learn. Res.* **6**, 1783-1816.

Netajinagar Vidyamandir, Netajinagar, Kolkata – 700 092, India.

E-mail: maniknnv@gmail.com

Applied Statistics Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata – 700 108, India.

E-mail: atanu@isical.ac.in

COMMENT

Vic Patrangenaru

Florida State University

1. Introduction

In their paper "Intrinsic Shape Analysis: Geodesic PCA For Riemannian Manifolds Modulo Isometric Lie Group Actions", Huckemann, Hotz, and Munk study a class of *principal curves* of random objects on *orbifolds* that are quotients of Riemannian manifolds by actions of group of isometries. These principal curves are least expected square distance generalized geodesics that explain the overall location of a data set on such an orbifold. Algorithms for computing generalized principal components are designed and carried out on Kendall's shape spaces Σ_m^k for $m > 2$ (Kendall (1984)).

2. Discussion

We discuss mainly aspects of the paper from the perspective of statistics on manifolds. The article requires a reasonable knowledge of Riemannian geometry

including geodesic and curvature computations at the level of Kobayashi and Nomizu (1969), and basic knowledge of PCA and statistical shape analysis.

Recall that an orbifold is a space of orbits of a group action on a manifold. In the paper the action is that of a group of isometries of a Riemannian manifold, and the resulting quotient orbifold can be thought of as a manifold with some well-behaved singularities. One of the most important contributions in the paper is on the behavior of generalized principal components for distributions around singular points.

In the following, all manifolds are assumed to be connected. Let ρ be a distance on a complete manifold $\mathcal{M}$ and $Q = P_X$ be a probability measure associated with a random object on $\mathcal{M}$. Ziezold (1994) defined the *Fréchet mean set* of X as the set of minimizers of the Fréchet function $F_\rho(x) = E(\rho^2(X, x)) = \int_{\mathcal{M}} \rho^2(y, x) Q(dy)$, assuming the integral is finite. Note that a point $x \in \mathcal{M}$ can be regarded as a 0-dimensional submanifold of $\mathcal{M}$, leading to the following.

Definition 2.1. Let $\mathcal{S}(\mathcal{M}, p)$ be the set of all closed $p-$dimensional connected submanifolds of $\mathcal{M}$, and the p-dimensional Fréchet function, $F_{Q,p} : \mathcal{S}(\mathcal{M}, p) \to [0, \infty]$, be

$$F_{Q,p}(N) = E(\rho^2(X, N)) = \int_{\mathcal{M}} \rho^2(x, N) Q(dx). \tag{2.1}$$

A submanifold $N_0 \in S(\mathcal{M}, p)$ is said to be a *principal submanifold* if N_0 is a minimizer of $F_{Q,p}$. If $X_1, \ldots, X_n$ is a random sample from Q, a sample principal submanifold is a submanifold $\hat{N}_0 \in \mathcal{S}(\mathcal{M}, p)$, minimizer of $F_{\hat{Q}_n,p}$, where $\hat{Q}_n$ is the empirical distribution of $X_1, \ldots, X_n$.

Note that when $p = 1$, principal submanifolds correspond to principal curves in the sense of Hastie and Stuetzle (1989). The notion of principal submanifold is nevertheless too general; indeed, given a sample $x_1, \ldots, x_n$, of observations from Q, for any $p \geq 1$ there are infinitely many sample principal submanifolds $\hat{N}$, since there are many submanifolds $\hat{N}$ containing the set $\{x_1, \ldots, x_n\}$, and for any such submanifold $F_{\hat{Q}_n,p}(\hat{N}) = 0$. It follows that this definition does not insure consistency. It is then preferable to restrict $F_{Q,p}$ to convenient types of submanifolds, satisfying certain second order (partial) differential equations, so that such families of manifolds depend continuously on a finite number of parameters, and consistency follows. For example in the Riemannian case, for $p = 1$, one may consider principal geodesics on $\mathcal{M}$, and for $p = 2$, one may consider principal minimal surfaces or principal totally geodesic surfaces if $\mathcal{M}$ is a symmetric space, etc.

Remark 2.1. Note that if $(\mathcal{M}, \rho)$ is the Euclidean space $(\mathbb{R}^p, \rho_0)$, geodesics are straight lines, totally geodesic surfaces are planes, etc. From the Pythagorean theorem, it follows that for any random vector X on $\mathbb{R}^p$, principal lines, principal

planes, etc. are spanned by principal components and they all go through the mean vector μ_X (for a proof, see for example Hastie and Stuetzle (1989)).

Nevertheless, the common Euclidean thinking that PC's go through the mean vector of a distribution (see Remark 2.1), led Fletcher, Lu, Pizer and Joshi (2004) to assume, by definition, that principal geodesics lines go through the intrinsic mean. On an arbitrary Riemannian manifold however, unlike in Euclidean geometry, the Pythagorean theorem fails, therefore there is no reason to believe that a principal geodesic contains the intrinsic mean. This was the starting point of *principal geodesic analysis* on Riemannian manifolds in Huckemann and Ziezold (2006), where a first example of a principal geodesic that does not go through the intrinsic mean was given. Huckemann et al. now build on the idea of principal geodesic on a manifold in Huckemann and Ziezold (2006), expand it, and give additional data-driven examples of geodesic PCA on Kendall's shape spaces (in 3D). Note that in shape analysis, the Kendall shape space Σ_3^k, the set of direct similarity shapes of nontrivial k-ads (configurations of k labeled nonidentical points) in $\mathbb{R}^3$, is often described as the quotient of the *preshape sphere* $\mathbb{S}^{3k-4} \subset \mathbb{R}^{3k-1}$ by the diagonal isometric action of the orthogonal group $SO(3)$. Since the isotropy groups of shapes of configurations of collinear k-ads are notrivial, such shapes are singular points on the shape orbifold Σ_3^k. Bearing in mind that this type of orbifold arises in statistical shape analysis, the authors consider in general the geometry of a quotient M/G of the canonical projection

$$\pi : M \to M/G, \tag{2.2}$$

of a complete Riemannian manifold M by a Lie group G that acts by isometries. Note that if d is the Riemannian distance on M, the distance between two orbits (fibers of π), $d_{M/G}(Gx, Gy) =: inf_{\{z \in Gy\}} d_M(x, z)$ is well defined, and *horizontal geodesics* on M (orthogonal to the fibers) project onto a so-called *generalized geodesic* on M/G. The authors define a *first generalized principal component* of a probability measure Q on the orbifold M/G in (2.2) as minimizer of a Fréchet-like function $F_{Q,1}$ defined on the set $\mathcal{G}(M/G, 1)$ of generalized geodesics, given by

$$F_{Q,1}(N) = E(d_{M/G}^2(X, N)) = \int_{\mathcal{M}} d_{M/G}^2(x, N) Q(dx). \tag{2.3}$$

A *first generalized principal component* (*GPC*) is a minimizing generalized geodesic of $F_{Q,1}$ in (2.3). Its sample counterpart is a generalized geodesic that yields the least sum of square residuals to the sample points. The *second GPC* is a generalized geodesic minimizing $F_{Q,1}$ in (2.3) over all generalized geodesics meeting orthogonally the first GPC. A *principal component mean* (*PM*) μ_P is a minimizer of the expected square distance to the points of intersection of the

first and second GPC. An example of a distribution on an orbifold M/G for which the first two GPC's meet at more than one point would been helpful here. The $m - th$ GPC, for $m > 2$, is defined inductively as a minimizing generalized geodesic of $F_{Q,1}$ in (2.3) over all generalized geodesics going through μ_P that are orthogonal on the first $m - 1$ GPC's.

The authors show that the PM and GPC's are well defined outside null sets. In the Euclidean case the intrinsic mean and the PM coincide, and, given that Huckemann et al.'s definition, although technical, falls under the general characterization of principal submanifolds, we find it more natural than that of Fletcher et al. (2004).

An interesting concept introduced by Huckeman et al. is that of *curvature CX present* in a random object X, given in their equation (7). Note that planar Kendall shape spaces have positive constant holomorphic curvature. It would be useful if the authors would prove their conjecture claiming that on a positive sectional curvature manifold $CX \geq 0$, while on a manifold of negative sectional curvature $CX \leq 0$, at least in the case $M/G = \Sigma_2^k = \mathbb{C}P^{k-2}$.

Much of the paper is dedicated to computations of generalized geodesics, and in particular of sample GPC's in Kendall's shape spaces. Huckemann et al. assume that M is a submanifold of a Hilbert space $\mathbb{H}$, which is globally defined as set of zeroes of a *submersion* $\phi : \mathbb{H} \to \mathbb{R}^k$. Generalized geodesics on M/G are represented by *horizontal geodesics* on M that are orthogonal to the orbits of K. Their computations are based on a two-stage minimization for the *optimal positioning* of a point $x \in M$ with respect to a geodesic on M, following ideas in Small (1996) or in Kendall, Barden, Carne and Le (1999). In particular a result in Kendall et al. (1999) on the unboundedness of the curvature along a geodesic approaching a singular point on a shape space in 3D, is given an elegant proof here.

Huckemann et al. apply their GPC methodology to data analysis on 3D shape orbifolds for both data on the regular part of the shape space, as well as for data that is closer to the singular part of this orbifold. Their methodology fares better than competing PCA techniques on manifolds, especially for data close to singular points on shape spaces, shapes of almost collinear landmarks in $\mathbb{R}^3$.

Historically, statistical shape analysis, including inference, was based on an equivariant embedding of the shape manifold in an Euclidian space of matrices. Unlike for intrinsic means, there are necessary and sufficient conditions for existence of the Fréchet means with respect to such extrinsic distances (see Goodall (1991), Kent (1992), Dryden and Mardia (1998), Bhattacharya and Patrangenaru (2003, 2005), Bhattacharya and Bhattacharya (2008), Bhattacharya (2008) and Bandulasiri, Bhattacharya and Patrangenaru (2009)). This remark is also

valid for directional and axial data analysis (Watson (1983), Prentice (1984), Fisher, Hall, Jing and Wood (1996), Beran and Fisher (1998)). In view of Definition 2.1, which can be formulated for any distance, the authors could have had pursued an extrinsic GPCA on manifolds. Such an approach would not be restricted to concentrated data on manifolds. An extrinsic approach might be also instrumental in developing in general an asymptotic theory for GPC's on manifolds.

Acknowledgements

References

Bhattacharya, A. and Bhattacharya, R. (2008). Nonparametric statistics on manifolds with applications to shape. Inst. Math. Stat. Collect., **3**, Pushing the limits of contemporary statistics: contributions in honor of Jayanta K. Ghosh, 282-301.

Bandulasiri, A., Bhattacharya, R. N. and Patrangenaru, V. (2009). Nonparametric inference for extrinsic means on size-and-shape manifolds with applications in medical imaging. *J. Multivariate Anal.* **100**, 1867-1882.

Beran, R. and Fisher, N. I. (1998). Nonparametric comparison of mean axes. *Ann. Statist.* **26**, 472-493.

Bhattacharya, A. (2008). Statistical analysis on manifolds: A nonparametric approach for inference on shape spaces. *Sankhyā, Ser. A.* **70**, 0-43.

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

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

Dryden, I. L. and Mardia, K. V. (1998). *Statistical Shape Analysis.* Wiley, Chichester.

Fisher, N. I., Hall, P., Jing, B. Y. and Wood, A. T. A. (1996). Improved pivotal methods for constructing confidence regions with directional data. *J. Amer. Statist. Assoc.* **91**, 1062-1070.

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

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

Hastie, T. and Stuetzle, W. (1989). Principal curves. *J. Amer. Statist. Assoc.* **84**, 502- 516.

Huckemann, S. and Ziezold, H. (2006). Principal component analysis for Riemannian manifolds with an application to triangular shape spaces. *Adv. Appl. Probab.* **38**, 299-319.

Kendall, D. G. (1984). Shape manifolds, Procrustean metrics, and complex projective spaces. *Bull. London Math. Soc.* **16**, 81-121.

Kendall, D. G., Barden, D., Carne, T. K. and Le, H. (1999). *Shape and Shape Theory.* Wiley, New York.

Kent, J. T. (1992). New directions in shape analysis. *The Art of Statistical Science*, 115-127. Wiley, Chichester.

Kobayashi, S. and Nomizu, K. (1969). *Foundations of Differential Geometry.* Vol. II.

Prentice, M. J. (1984). A distribution-free method of interval estimation for unsigned directional data. *Biometrika* **71**, 147-154.

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

Watson, G. S. (1983). *Statistics on Spheres.* John Wiley., New York.

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

Department of Statistics, Florida State University, Tallahassee, FL, 32308, U.S.A.

E-mail: vic@stat.fsu.edu

REJOINDER

Stephan Huckemann, Thomas Hotz and Axel Munk

Georgia Augusta Universiät Göttingen

1. Introduction

The authors wish to thank the discussants for their very interesting and stimulating contributions indicating various directions for future research and clarifying issues raised in our contribution. It seems that the following three major topics

1. "Simple and Parsimonious Descriptors for Shape Data",

2. "Shape Space Geometry", and

3. "Statistical Inference for Shape Spaces"

emerge from the ample comments provided by the discussants. These comments have been given from the individual perspectives of expertise in quite different fields which interestingly allow to connect originally disjoint strains of thoughts. For this reason we organize our rejoinder by following these specific issues and perspectives, rather than by addressing each contribution separately and thus loosing valuable aspects of this stimulating discussion.

1. Simple and Parsimonious Data-Descriptors

One goal of classical Euclidean statistics is to effectively describe data using low dimensional descriptors, not least to make them more interpretable. To such ends principal components analysis (PCA) is often employed, as its variance decomposition yields zero-dimensional (means), one-dimensional (first PC), and higher dimensional data descriptors. We emphasize that *simplicity* of data-descriptors is of value in itself, e.g., linear models may not model real life situations satisfactorily but their use for understanding and handling by a practitioner are beyond doubt.

As J.T. Kent elaborated upon in his comment, for some shape data, variance decomposition, dimension reduction, and arbitrary dimensional data-descriptors may be inappropriate concepts, and tools likewise. Thus for most data applications on a torus, almost every (w.r.t. the induced canonical Riemannian measure on the space of geodesics) "one-dimensional" geodesic data-descriptor is dense, i.e. is two-dimensional in effect, so hardly giving a "parsimonious" description of the data. Thus, for data sufficiently spread out on a torus, meaningful one-dimensional data descriptors may prove difficult to define. Hence, as a general phenomenon on arbitrary shape spaces, there may not be meaningful data-descriptors of any desired dimensionality. Moreover as demonstrated by our contribution, for a given data-set, data-descriptors of varying dimensionality may have little in common. In the second subsection, we follow and extend the classification of data proposed by J.T. Kent in his contribution.

Beforehand, however, we elaborate on the first issue, namely, that a reasonable objective of (intrinsic) data analysis consists in finding *parsimonious* data-descriptors that allow for the essential tasks of statistics to be performed; in particular, R.N. Bhattacharya in his contribution asked for feature selection, classification, and prediction – dimension reduction might also be added. Clearly, whether a descriptor is indeed parsimonious depends on one's aims. This well-known fact has been recently discussed in mathematical rigour by Yang (2005) among many others, viewing model selection as the search for a parsimonious model. This dependence will resurface when we discuss inference. An alternative intrinsic approach, based on directly adapting the geometry itself to suit the data, is proposed in Section 2.5.

Our contribution may be seen as proposing that, when working in a specific non-Euclidean geometry, parsimony is most naturally achieved by using data-descriptors based on the space's intrinsic geometry, most prominently based on geodesics. Choosing such intrinsic descriptors can be justified when the data at hand are in some way *congruent* to the underlying geometry. Since obviously some data – e.g., as presented by S. Jung, M. Foskey, and J. S. Marron

– are *incongruous* to the underlying geometry, a thorough investigation of this assumption is necessary.

1.1. Data-descriptors

In our contribution we used generalized geodesics to obtain parsimonious data descriptors. These generalized geodesics were taken from an underlying canonical geometric structure. Often, there is a unique canonical structure stemming from the subsequent immersions and submersions defining the shape space, e.g. for Kendall's shape spaces; this structure is given by immersing a hypersphere in a Euclidean space and subsequently submersing it w.r.t. the special orthogonal group action. Sometimes, however, there is more than one canonical structure, e.g., for the spaces of geodesics on Kendall's shape spaces (cf., Theorem 5.3) at least two different canonical structures come to mind. For the first structure, the Grassmannian involved is viewed as a quotient of a Stiefel manifold; for the other more simple structure, only quotients w.r.t. orthogonal groups are considered (cf., Edelman, Arias, and Smith (1998)). As remarked on by many discussants, one can generalize to other geometries as well.

A very interesting approach by S. Jung, M. Foskey and J.S. Marron is to retain the original spherical geometry but include arbitrary circles for principal components. While computationally not much harder to obtain than great circles, arbitrary circles allow for more flexibility in adapting to data on spheres and direct products thereof, which are not only the common ingredients of medial axes based shape manifolds (e.g Pizer, Siddiqi, Székely, Damon, and Zucker (2003), Fuchs and Scherzer (2008), as well as Sen, Foskey, Marron, and Styner (2008)), but can also be used to model prealigned landmark-based shape data (cf., Dryden (2005) as well as Hotz, Huckemann, Gaffrey, Munk, and Sloboda (2009)). We note that circles on spheres are curves of constant curvature thereby generalizing great circles which are curves with constant curvature zero. Extending this approach is a challenging project; one may as well investigate curves of constant non-zero curvature on general shape spaces, and build a principal component analysis on them.

J.T. Kent also proposed allowing more freedom for choosing one-dimensional descriptors. In particular, he demonstrated how time series of shape data showing the growth of rats can be successfully modeled by employing tensor products of e.g., principal splines. Kume, Dryden, and Le (2007) also model higher-order curves on manifolds to describe shape data by splines. We remark, however, that these models serve a different primary purpose, namely to describe longitudinal data as opposed to i.i.d. observations for which PCA is usually employed.

In the Appendix we touched on a third approach to parsimony based on simple descriptors, namely considering totally geodesic submanifolds or submanifolds totally geodesic at least at a point. We emphasize that the former may not exist for arbitrary dimension, the latter may only locally be manifolds. Noting this idea of employing higher-dimensional submanifolds, V. Patrangenaru suggested considering principal submanifolds, as introduced by Hastie and Stuetzle (1989), e.g., from the class of minimal surfaces and their higher dimensional analogs. This challenging topic certainly deserves further research.

1.2. Limitations to dimension reduction

Depending on their distribution, not all data on manifolds may warrant descriptions of any desired dimension. In order to facilitate discussions about which descriptors are appropriate for what data sets, J.T. Kent distinguished between *Type I* and *Type II* data. Data of Type I can effectively be analyzed by mimicking Euclidean PCA; in particular, the first principal component describing the main direction of data variation passes through the mean. In contrast on compact spaces, data may be spread along recurrent geodesics; for such Type II data, the concept of a data mean is meaningless and the first principal component is the most parsimonious data descriptor. A typical example of Type II data is given by a sample of a girdle distribution around the equator of a sphere. Obviously, the means, i.e. the north- and south-pole, constitute no reasonable (zero-dimensional) summaries of the data.

In his discussion, J.T. Kent re-analyzes our crown data in a fascinating and simple manner. While the impact of unbounded curvature for data near singular shapes is quite dramatic when the data involve reflections, i.e. symmetries w.r.t. a singular shape, J.T. Kent correctly points out that if this symmetry is removed, the impact of curvature may be considerably reduced; for the "tree-crown" data at hand, *Type II* data thus are transformed into *Type I* data. Nonetheless, whether the symmetry may be removed is a question of the research's aims; often, identifying symmetric objects is not permissible.

We believe that introducing the distinction between *Type I* and *Type II* is a very enlightening clarification of the fundamentals of our endeavor. While being inspired by J.T. Kent's classification, we would like to distinguish more subtly between *flat data* (his Type I), *curved data*, and *looped data*, as we feel that for data on general manifolds there are more than two types of situations to be treated distinctly. We note that J.T. Kent's definition of data of Type II is a special case of our definition of *looped data*. Also, data – be they Euclidean or on a manifold – may be *incongruous* with the geometric structure of the space. Typical data sets of the four types are depicted in Figure 1.
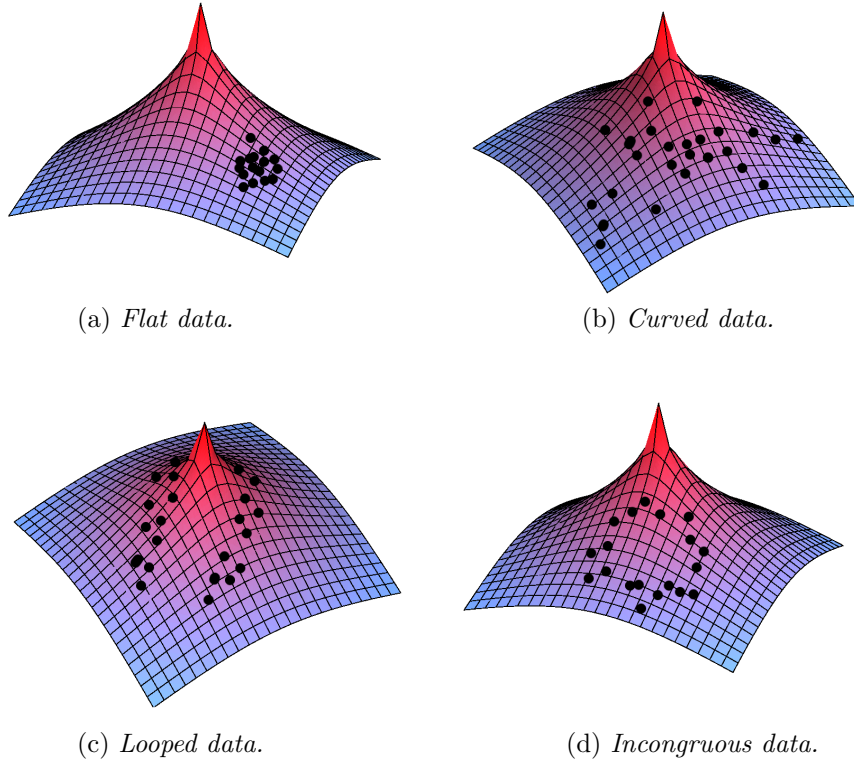
(a) *Flat data.*   (b) *Curved data.*

(c) *Looped data.*   (d) *Incongruous data.*

Figure 1. Visualizing typical data types on a cone-like surface with un-bounded curvature.

Flat data are concentrated enough such that one can treat them as if they were observed on a flat (i.e. Euclidean) space; clearly, the higher the curvature in this region, the more concentrated the data set needs to be. For this kind of data, all means (Procrustes mean, IM and PM) are close to one another and the first GPC passes nearly through the IM. In consequence within some approximation, the total variance decomposes, mimicking Euclidean PCA. This type of data has been characterized as Type I by J.T. Kent. In this situation, general Procrustes analysis (GPA), principal geodesic analysis (PGA), and geodesic principal component analysis (GPCA) yield similar results.

Curved data spread further than flat data, so the space's curvature needs to be taken into account. Such data have their Procrustes mean and IM much closer to one another than to the PM, and are considerably distant from the first GPC. Still, the means represent meaningful zero-dimensional data

descriptors. For this type of data, however, a decomposition of *total variance* into variances explained by geodesics is inappropriate.

Looped data are more severely affected by the space's curvature than curved data. While curvature is more of a local feature in curved data, for looped data global effects of curvature play a dominant role. In consequence, such data may not feature parsimonious data descriptors in a meaningful way for any given dimensionality, e.g., a meaningful mean may not exist because the data are spread around an equator of a sphere or because data "loop" around a singularity of the space. Often, the IM and the Procrustes mean can still be computed and they are far from the PM. Then GPA and PGA can be performed as well, usually yielding similar results yet very different from the results obtained by GPCA.

Incongruous data possess features that are not easily modelled using descriptors derived from the space's (intrinsic) geometry. One might as well say that geometry-based models do not fit the data. Similar to looped data, parsimonious data descriptors may not exist in a meaningful way for any given dimensionality. In contrast to looped data however, this is not a consequence of the geometry of the underlying space, rather the data do not conform to the given geometry. A typical example of incongruous data is given by clusters, e.g., isotropically arranged w.r.t. their common center such that only the zero- and the full-dimensional data descriptor are meaningful, or by data along a circle in a two-dimensional Euclidean space (only the zero- and two-dimensional data descriptor are meaningful).

In a way, for flat data one may work and think Euclidean, for curved data one must abandon the Euclidean concept of nested variance decomposition, while for looped data, one has to additionally give up the quest for reduction to arbitrarily low dimension. As clearly illustrated by S. Jung, M. Foskey, and J.S. Marron, incongruous data cannot be tackled well with models based on the intrinsic geometry alone. Note that curved and looped data can only occur on non-flat spaces.

Examples for flat, looped and incongruous data. The classical data set of "macaque skulls" (Dryden and Mardia, 1993) is a typical example for flat data, intrinsic variance (data dispersion) and *data curvature* (CX) both are low, as illustrated in Section 6.3 of our contribution. *Flat* Kendall shapes of two-dimensional objects can be modelled well by complex Bingham distributions (cf., Kent (1994)) if the dispersion is small (cf., Huckemann and Hotz (2009)). More realistic models are achieved by using quartic complex Bingham distributions (cf.,

Kent, Mardia, and McDonnell (2006)) of low dispersion. Low dispersion and low curvature distinguish flat from looped data, as the latter feature a notably high data curvature.

While for looped data there is no need for low intrinsic variance – e.g. for a girdle distribution around a sphere as pointed out by J.T. Kent – small intrinsic variance with high data curvature indicates proximity to a singularity as in our "tree crown" example. Recall that shape spaces may feature unbounded curvature in regions of bounded diameter. In Figure 1(c) we illustrate the latter situation of low intrinsic variance for looped data in a simplified two-dimensional geometry (Σ_3^4 is five-dimensional). Note the subtle difference from *incongruous data* on the same surface, depicted in Figure 1(d). For the former, the singularity is surrounded by the data which lie "intrinsically along a straight line" (i.e. along a geodesic); it is the space that generates the loop. The latter data surround a regular region of the surface where the looped structure is not caused by the space's geometry. Typical examples of incongruous data modelled with non-geodesic descriptors are illustrated in the contribution of S. Jung, M. Foskey, and J.S. Marron, as well as by Hastie and Stuetzle (1989).

The "brooch" data are curved. Let us now reconsider the brooch data from Section 6.2. Again, the IM or the Procrustes mean serve well as one-dimensional data-descriptors. Recall that the first GPC captures the dominant mode of data-variation, namely diversification that is found neither by GPA nor PGA. The differences between these methods' results show the data not to be flat; this is also visible from the considerable data curvature (CX).

In some way, the second GPC seems "parallel" to the generalized geodesic determined by the covariance matrix obtained from GPA or PGA, as it catches precisely that mode of data-variation. Since every single brooch shape is closer to any other brooch shape than to its reflected shape, this data-set contains "no reflections" as opposed to the crown data which are indeed looped. We conclude that this data set is curved.

2. Shape Space Geometry

In this section we remark on the very profound comments in the contributions of R.N. Bhattacharya and V. Patrangenaru concerning the role of means and geodesic variance. As is well known (cf., Karcher (1977)), uniqueness of intrinsic means can only be assured under restrictive conditions involving bounds on curvature. In effect, for data in high curvature regions as in our "tree-crown" example, a theoretical argument giving the uniqueness of intrinsic means that we naively computed presents quite a challenge. It seems even more difficult,

yet similarly important for a thorough foundation of analysis of nearly degenerate data to derive general conditions on the uniqueness of geodesic principal components and their intersection point, the principal component mean.

Inspired by V. Patrangenaru, we first establish that the *data curvature CX* has the same sign as the sectional curvature on constant curvature manifolds. As a consequence, MANOVA cannot directly be applied to data on manifolds, rather we propose a combination of local variance decomposition coupled with parallel transport in the second subsection.

The third subsection addresses the concern of R.N. Bhattacharya that geodesic variances explained by the s-th GPC may be negative.

In the fourth subsection we take up the non-trivial issue of variance obtained by projection on Kendall's shape spaces. At this point we would like to clarify that *principal geodesic analysis* (PGA), as introduced by Fletcher and Joshi (2004), which has been cited by R.N. Bhattacharya as well as by M.C. Mukherjee and A. Biswas, is almost equivalent to *general Procrustes analysis* (PGA). In both approaches the eigenvectors of the covariance matrix computed from the data mapped to the tangent space at some mean determine the principal components; the difference is that PGA employs the IM whereas GPA uses the Procrustes mean. The similarity of the approaches follows from the fact that the means are usually very close to one another. However, Fletcher and Joshi (2004) also suggested that one could define GPCs by maximizing the projected variance. The fourth section is also intended to clarify why we consider this problematic; we rather agree with V. Patrangenaru that minimizing the residual variance under no constraining condition appears far more natural.

In the concluding two subsections we suggest altering the Riemannian structure, based on the comments of P. T. Kim and J.-Y. Koo; finally, leaving Riemannian geometry altogether, we propose a version of *extrinsic PCA*, as triggered by V. Patrangenaru's comments.

2.1. *Data curvature* estimated by CX

Here we consider data on a manifold M of constant positive or constant negative sectional curvature, i.e., on a sphere or on a hyperbolic space, respectively. Recall that spherical shape spaces have been studied by Dryden (2005) as well as Hotz et al. (2009); for hyperbolic shape spaces we refer to studies of Bookstein (1991), Le and Small (1999), Le and Barden (2001), as well as Kume and Le (2002). The situation underlying the following lemma is depicted in Figure 2.

Lemma 2.1. (Spherical and Hyperbolic Theorem of Pythagoras). *Suppose that two geodesics γ_1 and γ_2, on a constant sectional curvature manifold M with*

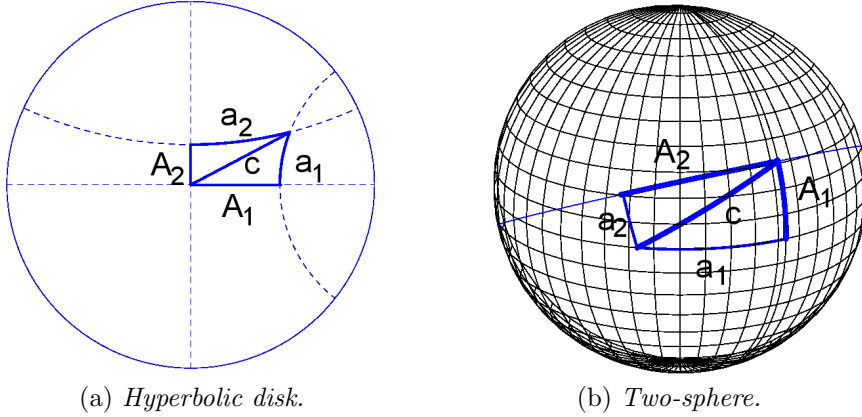(a) *Hyperbolic disk.*   (b) *Two-sphere.*

Figure 2. Pythagoras Theorem on constant sectional curvature manifolds.

intrinsic metric d, meet orthogonally at $\mu \in M$, and that $p \in M$ is contained in the surface spanned by γ_1 and γ_2. For $j = 1, 2$, let $a_j = d(p^{(\gamma_j)}, p)$, $A_j = d(p^{(\gamma_j)}, \mu)$, and $c = d(\mu, p)$ where the orthogonal projection of p to γ_j is $p^{(\gamma_j)}$ assumed to be well defined. Then

$$a_1^2 + a_2^2 \leq c^2 \leq A_1^2 + A_2^2$$

on spheres, whereas

$$a_1^2 + a_2^2 \geq c^2 \geq A_1^2 + A_2^2$$

on hyperbolic spaces. The inequalities are strict unless p lies on γ_1 or γ_2.

Proof. First note that

$$y \sin x \leq \sin(yx), \quad 0 \leq x \leq \pi, \tag{2.1}$$

$$y \tan x \geq \tan(yx), \quad 0 \leq x \leq \frac{\pi}{2}, \tag{2.2}$$

$$y \tan x \leq \tan(yx), \quad \frac{\pi}{2} \leq x \leq \pi, \tag{2.3}$$

$$y \sinh z \geq \sinh(yz), \quad z \geq 0, \tag{2.4}$$

$$y \tanh z \leq \tanh(yz), \quad z \geq 0, \tag{2.5}$$

for all $0 \leq y \leq 1$. The inequalities are strict unless $y = 0, 1$ or $x = 0$ (for (2.1) and (2.2)) or $z = 0$ (for (2.4) and (2.5)). This can be seen by verifying equality at $y = 0, 1$ and by verifying that the r.h.s. of (2.1) and (2.5) are strictly concave in y, while the r.h.s. of (2.2) and (2.4) are strictly convex. For (2.3) note that the r.h.s. is strictly concave for $\pi/2 \leq yx \leq \pi$, and $y \tan x < 0 < \tan(yx)$ otherwise. Now for $j = 1, 2$, denote by $\alpha_j \in [0, \pi/2]$ the angle between the geodesic γ_j and the geodesic from μ to p. Note that $\sin \alpha_1 = \cos \alpha_2$.

Assume that M is a sphere. From the spherical law of the sine we have

$$\frac{\sin a_j}{\sin \alpha_j} = \sin c,$$

giving with (2.1) (e.g. for the first term set $x = c$ and $y = \sin \alpha_1$, and for $c > \pi/2$ use the monotonicity of arcsin at $(\pi - x)y$ instead)

$$a_1^2 + a_2^2 = \arcsin^2(\sin \alpha_1 \sin c) + \arcsin^2(\sin \alpha_2 \sin c)$$
$$\leq c^2 \sin^2\alpha_1 + c^2 \cos^2\alpha_1 = c^2,$$

as desired. Note that equality holds if and only if $c = 0$ or $\sin \alpha_1 = 0, 1$, i.e. iff p lies on one of the two geodesics. From the spherical law of the cosine we have

$$\cos A_1 = \frac{\cos c}{\cos a_2} = \frac{\cos c}{\sqrt{1 - \sin^2 c \, \cos^2\alpha_1}}, \quad \cos A_2 = \frac{\cos c}{\sqrt{1 - \sin^2 c \, \cos^2\alpha_2}}$$

giving

$$A_1^2 + A_2^2 = \arccos^2\left(\frac{\cos c}{\sqrt{1 - \sin^2 c \, \cos^2\alpha_1}}\right) + \arccos^2\left(\frac{\cos c}{\sqrt{1 - \sin^2 c \, \cos^2\alpha_2}}\right)$$
$$\geq c^2 \sin^2\alpha_1 + c^2 \sin^2\alpha_2 = c^2.$$

Here, we used

$$\frac{\cos^2 c}{1 - \sin^2 c \, \cos^2\alpha_j} = \frac{\cos^2 c}{\cos^2 c + \sin^2 c \, \sin^2\alpha_j}$$
$$= \frac{1}{1 + \sin^2\alpha_j \, \tan^2 c}$$
$$\left.\begin{array}{c}\leq\\\geq\end{array}\right\} \cos^2(c \sin \alpha_j) \quad \text{for} \quad \left\{\begin{array}{c}0 \leq c \leq \frac{\pi}{2}\\\frac{\pi}{2} \leq c \leq \pi\end{array}\right.,$$

which is a consequence of (2.2) and (2.3). Equality holds again if and only if p lies on one of the two geodesics.

Now suppose that M is a hyperbolic space. We have the hyperbolic laws of sine and cosine:

$$\frac{\sinh a_j}{\sin \alpha_j} = \sinh c, \quad \cosh A_j = \frac{\cosh c}{\cos a_{j'}}, \quad \{j, j'\} = \{1, 2\}.$$

Then with the argument above, accordingly modified using (2.4), we have at once $a_1^2 + a_2^2 \geq c^2$. The other inequality, $A_1^2 + A_2^2 \leq c^2$, follows from an analogous argument using

$$\cosh a_{j'} = \sqrt{1 + \sinh^2 a_{j'}} = \sqrt{1 + \sinh^2 c \, \cos^2\alpha_j},$$

$$\frac{\cosh^2 c}{1 + \sinh^2 c \, \cos^2 \alpha_j} = \frac{\cosh^2 c}{\cosh^2 c - \sinh^2 c \, \sin^2 \alpha_j}$$
$$= \frac{1}{1 - \sin^2 \alpha_j \, \tanh^2 c}$$
$$\leq \cosh^2(c \, \sin \alpha_j) \,,$$

which is a consequence of (2.5). Equality holds again if and only if p lies on one of the two geodesics.

We say that a random variable X on a quotient space $Q = M/G$ *admits a unique GPCA* if all population GPCs and the population PM exist and are uniquely determined, and if the orthogonal projections $X^{(\delta)}$ to all GPCs δ are a.s. well defined.

As a consequence of Theorem 2.6 of our contribution, every random variable absolutely continuous w.r.t. the measure induced by the Riemannian measure on M features a.s. well defined orthogonal projections to a given generalized geodesic.

Recall that every submanifold of a constant curvature manifold spanned by geodesics through a common point is totally geodesic. Hence, an inductive argument relying on Lemma 2.1 gives at once the following.

Theorem 2.2. *Suppose that a random variable X on a constant curvature manifold M admits a unique GPCA. Then $CX = 0$ for zero sectional curvature, $CX \geq 0$ for positive sectional curvature, and $CX \leq 0$ for negative sectional curvature. The inequalites are strict if and only if X does not exclusively assume values on its GPCs a.s.*

This settles the issue raised by V. Patrangenaru in the special case of constant curvature manifolds.

2.2. Variance decomposition and multiple effects models

Variance decomposition, and hence dimension reduction, in Euclidean space is based on the Pythagoras Theorem that has $CX = 0$. For random variables spread out on spaces involving curvature this decomposition poses difficulties, as will be further elaborated on below. The approach of classical MANOVA and *multiple effects models* can be thought of as a combination of variance decomposition locally and comparison via the connection of tangent spaces, i.e. affine parallel transport. Obviously on compact spaces, *parallelism* can only be a local concept. Translating an intuitive notion of similar shape variation into, say, *parallel* data variation (as begun in Huckemann (2009)), seems like another challenging goal when confronting the non-linear structure of shape spaces. In a

Table 1. Variance explained by residuals. Top row: five-dimensional shapes of tree crowns; middle row: five-dimensional data of iron age brooches; and bottom row: nine-dimensional data of macaque skulls. For the latter data only the ultimate variance is negative.

GPC1	GPC2	GPC3	GPC4	GPC5	...	GPC9
$2.6e-05$	$9.4e-06$	$-2.8e-06$	$-3.1e-06$	$-3.1e-06$		
0.40226	0.33941	0.08660	-0.00035	-0.03718		
$2.7e-02$	$1.2e-02$	$7.2e-03$	$4.7e-03$	$3.3e-03$	...	$-1.3e-05$

similar vein, additive models cannot directly be generalized to shape spaces, because in general these spaces lack a (natural) commutative operation. In Huckemann, Hotz, and Munk (2009), we discuss generalizations of classical fixed effects models toward *intrinsic MANOVA*.

2.3. Variance explained by residuals

In Euclidean geometry due to the Pythagoras Theorem, $V_{res}^{(s)} \geq 0$ for all $1 \leq s \leq m$. For higher dimensions m, the variance $V_{res}^{(s)}$ *explained by the s-th GPC obtained by residuals* can be viewed as the difference between the mean squared distance to all GPCs and the squared distanced to the s-th GPC. In view of the Pythagoras Theorem for constant curvature spaces, cf., Lemma 2.1 for this reason, higher order variances may be negative. This effect increases with dimension, dispersion, and anisotropy. Numerical experiments for data on m-spheres give negative variance $V_{res}^{(m)}$ "explained" by the ultimate GPC in more than 80 % of the simulations for

(a) data uniformly distributed on a quarter sphere $\{(x_1,\ldots,x_{m+1}) \in \mathbb{R}^{m+1} : \sum_{j=1}^{m+1} x_j^2 = 1, -\pi/4 \leq x_1 \leq \pi/4\}$ for $m \geq 7$; and

(b) data highly anisotropically distributed following a spherical Bingham distribution with eigenvalues $0, 0, 0, -10^4$, i.e. $m = 3$ (see e.g., Mardia and Jupp (2000, Section 9.4.3)).

Obviously, for $m = 2$ and any data admitting a unique GPCA, both $V_{res}^{(1)}$ and $V_{res}^{(2)}$ are non-negative. For the shape data considered in our contribution the individual variances explained by residuals are depicted in Table 1.

Summarizing, we can say that the tendency of higher order residual variances to be negative increases with dimension, dispersion, anisotropy, and data curvature (CX).

2.4. Geodesic scores

Suppose that $p^{(\delta)} = x \sin \alpha + v \cos \alpha$ is a pre-shape of the orthogonal projection of a shape $[p] \in \Sigma_m^k$ to a generalized geodesic δ through a principal

component mean $[x] \in \Sigma_m^k$ with initial velocity $v \in H_x S_m^k$ and geodesic score $t = \arctan(\alpha)$. As we have seen, even for concentrated data, due to oscillation $|t|$ can be large. If one determined generalized geodesics by maximizing sums of squared geodesic scores as proposed by Fletcher and Joshi (2004), this effect would be enlarged giving non-interpretable geodesic scores. For an example one may think of data on a torus where there will be geodesics that allow infinite scores while staying arbitrarily close to the data.

2.5. Data driven Riemannian metrics

As pointed out by most of the discussants (cf., Section 1.1), many data-sets are approximated much better by non-geodesic curves than by geodesics. In view of parsimony and the interpretation of the geometry of the shape space as reflecting an "elastic shape energy" (cf., Bookstein (1986) as well as Grenander and Miller (1994)), one might boldly want to alter the canonical geometry of the shape space according to the data to be modelled. In their very interesting contribution P.T. Kim and J.-Y. Koo pointed to the fact that the geometric structure is equivalently described by the Laplace operator, which in turn is characterized by its eigenfunctions and eigenvalues. Recent applications to image understanding and shape analysis have successfully exploited this fact, e.g., Reuter, Wolter, and Peinecke (2006) or Wardetzky, Mathur, Kälberer, and Grinspun (2007). Under a statistical paradigm, these relations may be used to obtain a *data-driven adaption of the metric*; a challenging endeavor that may provide further insight, e.g., into biological growth, by finding the suitable geometry for a "geodesic hypothesis" to hold. Indeed it is well known that for some applications (e.g., Kume et al. (2007)), certain classes of curves non-geodesic w.r.t. the canonical metric fit biological growth data much better than geodesic curves. Possibly, a framework can be utilized which has been laid out in Kim, Koo, and Luo (2009) for a different statistical estimation problem, though in a very similar context.

2.6. Extrinsic PCA

Finally, we comment on V. Patrangenaru's plea for extrinsic analysis. As illustrated in Bandulasiri, Bhattacharya and Patrangenaru (2009) for Kendall's three-dimensional reflection shape space, the *Schönberg embedding* allows for extrinsic methods for the manifold part of the quotient. The intriguing fact about extrinsic methods – if available – is that means and principal components can be directly computed in Euclidean space and are mapped orthogonally back to the manifold and the tangent space at the former, respectively. For Kendall's three-dimensional shapes non-invariant under reflections, a suitable embedding seems not at hand. Moreover, in general, a canonical approach to extrinsic PCA seems not obvious; e.g., one could define extrinsic PCs by projecting straight lines of

the ambient space. Building such an *extrinsic PCA* at least on spaces with a "benign" embedding seems like an interesting and challenging goal to pursue.

3. Statistical Inference

Several discussants bemoaned "the complete lack of consideration of problems of statistical inference" (R. N. Bhattacharya). Indeed, GPCA so far only gives a parsimonious description of the data, but it does highlight the difficulties already associated with descriptive statistics of shape data that need to be understood before attempting to do inference. Nonetheless, we summarize some of the discussants' suggestions for moving forward and mention some recent developments in this direction.

As J.T. Kent points out, there is a need for "more work to be done" developing *distributions* on shape spaces, especially for higher-dimensional shapes. Such distributions are necessary to perform what is commonly known as *parametric* statistics where one starts by specifying a probabilistic model for one's data in order to infer about the model's parameters after observing the data. One promising approach to obtaining a generalization of a Gaussian distribution on manifolds was mentioned by P.T. Kim and J.-Y. Koo, viewing this "Gaussian" distribution as the solution of a diffusion equation with an adequately defined Laplacian. They then propose to use likelihood methods for statistical inference by means of the corresponding empirical characteristic function.

If one wants to avoid distributional assumptions about the data, *nonparametric* methods need to be employed. M. C. Mukherjee and A. Biswas suggest the use of resampling techniques to this end. A common technique for proving the validity of, say, the bootstrap for inference requires a central limit theorem (CLT) for the statistic in question. For a mean on a manifold, this is indeed available, see e.g., Hendriks and Landsman (1996, 1998), as well as Bhattacharya and Patrangenaru (2003, 2005) for the extrinsic and intrinsic mean. Such results are relatively easy to obtain since they make use of the fact that the mean's distribution gets more and more concentrated asymptotically, hence allowing for a Euclidean approximation. For PCA, matters are more difficult since PCs by definition extend into the manifold – possibly even worse, onto the non-manifold part of the quotient – and hence do not allow for a Euclidean approximation, even asymptotically; for the Euclidean case see e.g., Anderson (1963) or Ruymgaart and Yang (1997). More involved resampling techniques will be necessary here, and the asymptotic distribution of the GPCs and its bootstrap analog appears to us a very interesting challenge for the future.

Although statistical inference on manifolds raises difficulties, successful attempts have been made for specific statistical models, especially for *flat data* (cf., above). The latter e.g., allow for one-way analysis of variance, testing the

hypothesis of no difference between the groups, see e.g., Dryden and Mardia (1998) where the analysis is performed in the tangent space; R.N. Bhattacharya discussed intrinsic treatments of one- and two-sample problems in his contribution. Recently, Huckemann et al. (2009) have developed an intrinsic two-way MANOVA for groups of flat data for which Euclidean approximation in a single tangent space is not necessarily appropriate, i.e., where the entire data set is not necessarily flat.

While in the past most efforts have focused on flat data, we currently witness an increased interest in developing methodology for curved data. Due to the aforementioned difficulties, many of the existing tools are only descriptive but there are first results allowing to do inference for such data, e.g., based on CLTs of intrinsic means. Curved data will certainly remain an issue of intense research in the near future, calling for the careful generalization of existing techniques to spaces where curvature has to be taken into account. This is especially important for Kendall's shape spaces that feature non-constant curvature, or even unbounded curvature for three- and higher-dimensional shapes. For looped data, however, many concepts and views that have been developed for flat data will no longer be applicable, so new ideas are needed to address the challenges such data sets pose. This certainly requires fresh ways of thinking, opening the field of shape analysis toward hitherto uncharted territory.

References

Anderson, T. (1963). Asymptotic theory for principal component analysis. *Ann. Math. Statist.* **34**, 122-148.

Bandulasiri, A., Bhattacharya, R. N. and Patrangenaru, V. (2009). Nonparametric inference for extrinsic means on size-and-(reflection)-shape manifolds with applications in medical imaging. *J. Multivariate Anal.* **100**, 1867-1882.

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

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

Bookstein, F. L. (1986). Size and shape spaces for landmark data in two dimensions (with discussion). *Statist. Sci.* **1**, 181-222.

Bookstein, F. L. (1991). *Morphometric Tools for Landmark Data: Geometry and Biology.* Cambridge University Press, Cambridge.

Dryden, I. L. (2005). Statistical analysis on high-dimensional spheres and shape spaces. *Ann. Statist.* **33**, 1643-1665.

Dryden, I. L. and Mardia, K. V. (1993). Multivariate shape analysis. *Sankhyā Ser. A* **55**, 460-480.

Dryden, I. L. and Mardia, K. V. (1998). *Statistical Shape Analysis.* Wiley, Chichester.

Edelman, A., Arias, T. A. and Smith, S. T. (1998). The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.* **20**, 303-353.

Fletcher, P. T. and Joshi, S. C. (2004). Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors. *ECCV Workshops CVAMIA and MMBIA*, 87-98.

Fuchs, M. and Scherzer, O. (2008). Regularized reconstruction of shapes with statistical a priori knowledge. *Internat. J. Comput. Vision* **79**, 119-135.

Grenander, U. and Miller, M. (1994). Representation of knowledge in complex systems. *J. Roy. Statist. Soc. Ser. B* **56**, 549-603.

Hastie, T. and Stuetzle, W. (1989). Principal curves. *J. Amer. Statist. Assoc.* **84** (406), 502-516.

Hendriks, H. and Landsman, Z. (1996). Asymptotic behaviour of sample mean location for manifolds. *Statistics & Probability Letters* **26**, 169-178.

Hendriks, H. and Landsman, Z. (1998). Mean location and sample mean location on manifolds: asymptotics, tests, confidence regions. *J. Multivariate Anal.* **67**, 227-243.

Hotz, T., Huckemann, S., Gaffrey, D., Munk, A. and Sloboda, B. (2009). Shape spaces for prealingend star-shaped objects in studying the growth of plants. *J. Roy. Statist. Soc. Ser. C*, to appear.

Huckemann, S. (2009). Parallel shape deformation. Preprint.

Huckemann, S. and Hotz, T. (2009). Principal components geodesics for planar shape spaces. *J. Multivariate Anal.* **100**, 699-714.

Huckemann, S., Hotz, T. and Munk, A. (2009). Intrinsic MANOVA for Riemannian manifolds with an application to Kendall's space of planar shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* To appear.

Karcher, H. (1977). Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math.* **XXX**, 509-541.

Kent, J. (1994). The complex Bingham distribution and shape analysis. *J. Roy. Statist. Soc. Ser. B* **56**, 285-299.

Kent, J. T., Mardia, K. V. and McDonnell, P. (2006). The complex Bingham quartic distribution and shape analysis. *J. Roy. Statist. Soc. Ser. B* **68**, 747-765.

Kim, P. T., Koo, J.-Y. and Luo, Z.-M. (2009). Weyl eigenvalue asymptotics and sharp adaptation on vector bundles. *J. Multivariate Anal.* To appear.

Kume, A., Dryden, I. and Le, H. (2007). Shape space smoothing splines for planar landmark data. *Biometrika* **94**, 513-528.

Kume, A. and Le, H. (2002). On Fréchet means in simplex shape space. *Adv. Appl. Prob. (SGSA)* **35**, 885-897.

Le, H. and Barden, D. (2001). On simplex shape spaces. *J. London Math. Soc.* **64**, 301-512.

Le, H. and Small, C. (1999). Multidimensional scaling of simplex shapes. *Pattern Recognition* **32**, 1601-1613.

Mardia, K. V. and Jupp, P. E. (2000). *Directional Statistics*. Wiley, New York.

Pizer, S. M., Siddiqi, K., Székely, G., Damon, J. N. and Zucker, S. W. (2003). Multiscale medial loci and their properties. *Internat. J. Comput. Vision* **55**, 155-179.

Reuter, M., Wolter, F.-E. and Peinecke, N. (2006). Laplace-Beltrami spectra as "shape-DNA" of surfaces and solids. *Computer-Aided Design* **38**, 342-366.

Ruymgaart, F. H. and Yang, S. (1997). Some applications of Watson's perturbation approach to random matrices. *J. Multivariate Anal.* **60**, 48-60.

Sen, S. K., Foskey, M., Marron, J. S. and Styner, M. A. (2008). Support vector machine for data on manifolds: An application to image analysis. In *Proceedings of the* 2008 *IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Paris, France,* May 14-17, 2008, 1195-1198.

Wardetzky, M., Mathur, S., Kälberer, F. and Grinspun, E. (2007). Discrete Laplace operators:
    No free lunch. Symposium on Geometry Processing, 33-37.

Yang, Y. (2005). Can the strengths of AIC and BIC be shared? A conflict between model
    identification and regression estimation. *Biometrika* **92**, 937-950.

Institute for Mathematical Stochastics, Georgia Augusta Universiät Göttingen, Maschmühlenweg
8-10, D-37073 Göttingen, Germany.

E-mail: huckeman@math.uni-goettingen.de

Institute for Mathematical Stochastics, Georgia Augusta Universiät Göttingen, Maschmühlenweg
8-10, D-37073 Göttingen, Germany.

E-mail: hotz@math.uni-goettingen.de

Institute for Mathematical Stochastics, Georgia Augusta Universiät Göttingen, Maschmühlenweg
8-10, D-37073 Göttingen, Germany.

E-mail: munk@math.uni-goettingen.de