Rejoinder to the Discussion of "Intrinsic Shape Analysis: Geodesic Principal Component Analysis for Riemannian Manifolds under Isometric Lie Group Actions"

Stephan Huckemann*[†], Thomas Hotz[‡] and Axel Munk

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

*Corresponding Author, Institute for Mathematical Stochastics, Georgia Augusta Universiät Göttingen, Maschmühlenweg 8-10, D-37073 Göttingen, phone: +49-551-39 13 514, fax: +49-551-39 13 505, e-mail: huckeman@uni-math.gwdg.de

[†]Supported by Deutsche Forschungsgemeinschaft Grant MU 1230/10-1.

[‡]Supported by Deutsche Forschungsgemeinschaft, Graduiertenkolleg 1023 "Identifikation in mathematischen Modellen"

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 et al. (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 et al. (2003), Fuchs and Scherzer (2008) as well as Sen et al. (2008)), but can also be used to model prealigned landmark-based shape data (cf. Dryden (2005) as well as Hotz et al. (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 et al. (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.

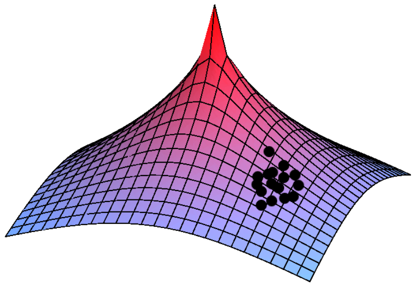
In his discussion, J.T. Kent re-analyzes our crown data in a fascinating and simple manner. While the impact of unbounded curvature for data near singular shapes is quite dramatic when the data involve reflections, i.e. symmetries w.r.t. a singular shape, J.T. Kent correctly points out that if this symmetry 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 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.

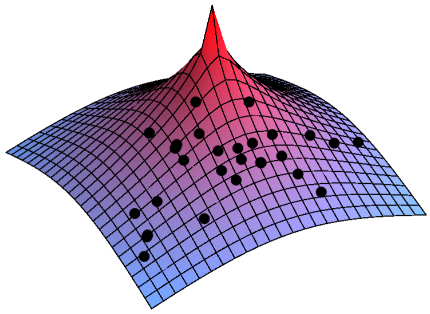
Flat data are concentrated enough such that one can treat them as if they were observed on a flat (i.e. Euclidean) space; clearly, the higher the curvature in this region, the more concentrated the data set needs to be. For this kind of data, all means (Procrustes mean, IM and PM) are close to one another 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.

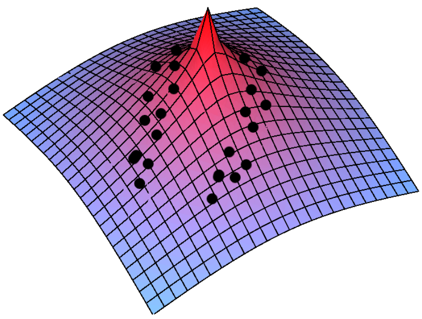
Looped data are more severely affected by the space's curvature than curved data. While curvature is more of a local feature in curved data, for looped data global effects of curvature play a dominating role. In consequence, such data may not feature parsimonious data descriptors in a meaningful way 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.



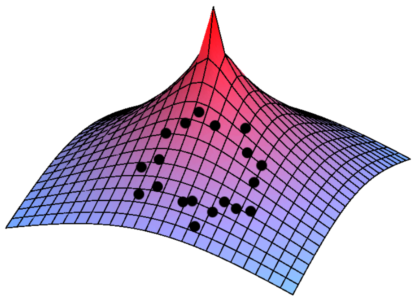
(a) *Flat data.*



(b) *Curved data.*



(c) *Looped data.*



(d) *Incongruous data.*

Figure 1: *Visualizing typical data types on a cone-like surface with unbounded curvature.*

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 et al. (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

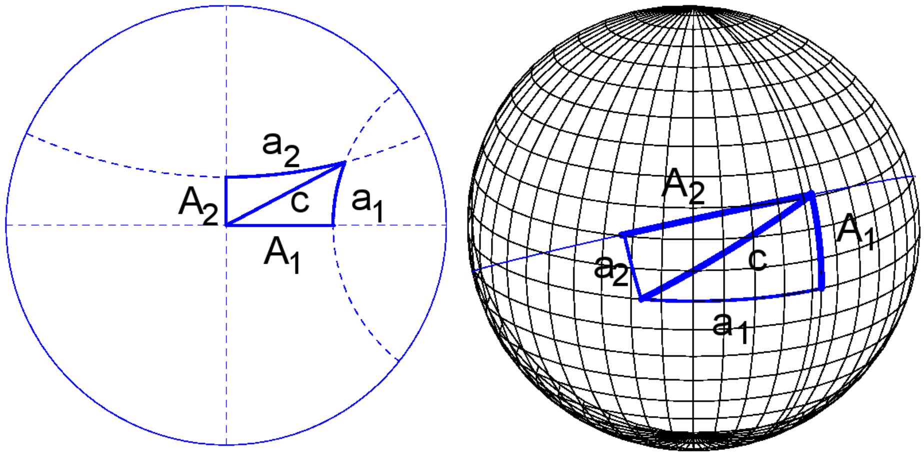
In this section we remark on the very profound comments in the contributions of R.N. Bhattacharya and V. Patrangenaru concerning the role of means and geodesic variance. As is well known (cf. Karcher (1977)), uniqueness of intrinsic means can only be assured under restrictive conditions involving bounds 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 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.



(a) *Hyperbolic disk.* (b) *Two-sphere.*

Figure 2: *Pythagoras Theorem on constant sectional curvature manifolds.*

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 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

$$\begin{aligned} y \sin x &\leq \sin(yx), \quad 0 \leq x \leq \pi\,, & (1)\\ y \tan x &\geq \tan(yx), \quad 0 \leq x \leq \frac{\pi}{2}\,, & (2)\\ y \tan x &\leq \tan(yx), \quad \frac{\pi}{2} \leq x \leq \pi\,, & (3) \end{aligned}$$

$$
\begin{aligned}
y \sinh z &\geq \sinh(yz), \quad z \geq 0\,, &(4)\\
y \tanh z &\leq \tanh(yz), \quad z \geq 0\,. &(5)
\end{aligned}
$$

for all $0 \leq y \leq 1$. The inequalities are strict unless $y = 0, 1$ or $x = 0$ (for (1) and (2)) or $z = 0$ (for (4) and (5)). This can be seen by verifying equality at $y = 0, 1$ and by verifying that the r.h.s. of (1) and (5) are strictly concave in y, while the r.h.s. of (2) and (4) are strictly convex. For (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 (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)

$$
\begin{aligned}
a_1^2 + a_2^2 &= \arcsin^2(\sin\alpha_1 \sin c) + \arcsin^2(\sin\alpha_2 \sin c)\\
&\leq c^2 \sin^2\alpha_1 + c^2 \cos^2\alpha_1 = c^2\,,
\end{aligned}
$$

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

$$
\begin{aligned}
A_1^2 + A_2^2 &= \arccos^2\left(\frac{\cos c}{\sqrt{1 - \sin^2 c\, \cos^2\alpha_1}}\right) + \arccos^2\left(\frac{\cos c}{\sqrt{1 - \sin^2 c\, \cos^2\alpha_2}}\right)\\
&\geq c^2 \sin^2\alpha_1 + c^2 \sin^2\alpha_2 = c^2\,.
\end{aligned}
$$

Here, we used

$$
\begin{aligned}
\frac{\cos^2 c}{1 - \sin^2 c\, \cos^2\alpha_j} &= \frac{\cos^2 c}{\cos^2 c + \sin^2 c\, \sin^2\alpha_j}\\
&= \frac{1}{1 + \sin^2\alpha_j\, \tan^2 c}\\
&\left.\begin{matrix}\leq\\ \geq\end{matrix}\right\} \cos^2(c \sin\alpha_j) \;\text{ for }\; \left\{\begin{matrix} 0 \leq c \leq \frac{\pi}{2}\\ \frac{\pi}{2} \leq c \leq \pi \end{matrix}\right. ,
\end{aligned}
$$

which is a consequence of (2) and (3). Equality holds again if and only if p lies on one of the two geodesics.

Nows suppose that M is a hyperbolic space. We have the hyperbolic laws of sine and cosine:

$$\frac{\sinh a_j}{\sin\alpha_j} = \sinh c, \quad \cosh A_j = \frac{\cosh c}{\cos a_{j'}}, \;\; \{j, j'\} = \{1, 2\}\,.$$

Then with the argument above, accordingly modified using (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

$$\begin{aligned}
\cosh a_{j'} &= \sqrt{1+\sinh^2 a_{j'}} &=& \sqrt{1+\sinh^2 c\, \cos^2\alpha_j}\ ,\\
\frac{\cosh^2 c}{1+\sinh^2 c\, \cos^2\alpha_j} &&=& \frac{\cosh^2 c}{\cosh^2 c - \sinh^2 c\, \sin^2\alpha_j}\\
&&=& \frac{1}{1-\sin^2\alpha_j\ \tanh^2 c}\\
&&\leq& \cosh^2(c\, \sin\alpha_j)\,,
\end{aligned}$$

which is a consequence of (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

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		
$2.7e-02$	$1.2e-02$	$7.2e-03$	$4.7e-03$	$3.3e-03$	...	$-1.3e-05$

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.*

challenging goal when confronting the non-linear structure of shape spaces. In a similar vein, additive models cannot directly be generalized to shape spaces, because in general these spaces lack a (natural) commutative operation. In Huckemann et al. (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 $\{(x_1, \dots, x_{m+1}) \in \mathbb{R}^{m+1} : \sum_{j=1}^{m+1} x_j^2 = 1, -\pi/4 \leq x_1 \leq \pi/4\}$ 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 would determine 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 et al. (2006) or Wardetzky et al. (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 et al. (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 et al. (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** (1), 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. In press.

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

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

Bookstein, F. L. (1986). Size and shape spaces for landmark data in two dimensions (with discussion). Statistical Science **1** (2), 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** (4), 1643–1665.

Dryden, I. L., and Mardia, K. V. (1993). Multivariate shape analysis. Sankhyā Series 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** (2), 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. International Journal of Computer Vision **79** (2), 119–135.

Grenander, U., and Miller, M. (1994). Representation of knowledge in complex systems. Journal of the Royal Statistical Society, Series B **56** (4), 549–603.

Hastie, T., and Stuetzle, W. (1989). Principal curves. Journal of the American Statistical Association **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. Journal of Multivariate Analysis **67**, 227–243.

Hotz, T., Huckemann, S., Gaffrey, D., Munk, A., and Sloboda, B. (2009). Shape spaces for pre-alingend star-shaped objects in studying the growth of plants. Journal of the Royal Statistical Society, Series C, to appear.

Huckemann, S. (2009). Parallel shape deformation. Preprint.

Huckemann, S., and Hotz, T. (2009). Principal components geodesics for planar shape spaces. Journal of Multivariate Analysis **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 Transactions on Pattern Analysis and Machine Intelligence. To appear.

Karcher, H. (1977). Riemannian center of mass and mollifier smoothing. Communications on Pure and Applied Mathematics **XXX**, 509–541.

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

Kent, J. T., Mardia, K. V., and McDonnell, P. (2006). The complex Bingham quartic distribution and shape analysis. Journal Of The Royal Statistical Society Series B **68** (5), 747–765.

Kim, P. T., Koo, J.-Y., and Luo, Z.-M. (2009). Weyl eigenvalue asymptotics and sharp adaptation on vector bundles. Journal of Multivariate Analysis. To appear.

Kume, A., Dryden, I., and Le, H. (2007). Shape space smoothing splines for planar landmark data. Biometrika **94** (3), 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. Journal of the London Mathematical Society **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. Int. J. Comput. Vision **55** (2-3), 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** (4), 342–366.

Ruymgaart, F. H., and Yang, S. (1997). Some applications of Watson's perturbation approach to random matrices. Journal of Multivariate Analysis **60** (1), 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. pp. 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.