Object Statistics on Curved Manifolds

Stephen M. Pizer* and J.S. Marron[†]*
University of North Carolina at Chapel Hill, Departments of Computer Science* and Statistics and Operations Research[†]

To appear as Chapter 6 in
G. Zheng, S. Li, G. Székley. *Statistical Shape and Deformation Analysis*. Elsevier, 2017.

Abstract

It is valuable to understand an object in 2-space or 3-space as abstractly being a point in a feature space that is a high-dimensional curved manifold. Thus, statistical operations on objects would be expected to be more effective if they recognize that the manifold is curved. As a result the common statistical methods applied to objects, which all assume Euclidean metrics, would not be fully effective when directly applied. These methods for flat (Euclidean) space include those for Gaussian probability density estimation (e.g., mean estimation and principal component analysis), for classification (e.g., support vector machine and distance-weighted discrimination), and for hypothesis testing (e.g., t-test and permutation test).

We present a number of forms of geometric object models: point distribution models, spherical harmonic models, normal vectors models, deformation models, and skeletal models. For each we show how an object thereby modeled can be understood to live on a curved manifold.

Then after summarizing the aforementioned statistical methods in Euclidean spaces, we present a variety of methods that use the fact that the manifolds are curved. These include methods that work directly on the curved manifold and methods of Euclideanizing the features before applying the Euclidean methods, in particular by Composite Principal Nested Spheres and Polysphere Principal Component Analysis. We describe how the Euclideanization approach allows statistics on change of shape, as well as statistics on shape. We also describe methods for improving correspondence of the primitives in the geometric models for the objects in the training or target cases, and as well a method that is independent of reparametrization of the boundary.

After describing methods for comparing object model / statistical method combinations, we give the results of a number of evaluations. These indicate that for object data those statistical methods that recognize that the manifold is curved provide superior results than those that assume a Euclidean space and that geometric models that include geometry of higher order than positions alone, such as boundary normal directions, provide superior results to those that are only based on positions.

1. Objectives of Object Statistics

We understand an object to be a region of space, typically with a smooth, or at least piecewise smooth, boundary (Fig. 1). Examples are an anatomic entity such as the kidney or the brain structure called the hippocampus or an everyday object such as a table. While objects can have holes, we will concentrate on objects with no holes, i.e., with spherical topology in 3D or circular topology in 2D.

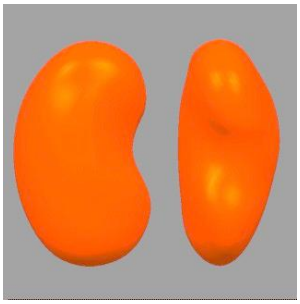


Figure 1. An example of a smooth object, seen from two points of view (by permission of T. Fletcher)

The statistical objectives on training samples from populations of such objects include probability distribution estimation, classification, and hypothesis testing. For classification we will restrict ourselves to discrimination between two classes. Likewise, in hypothesis testing we will restrict ourselves to finding object features that have statistically significant differences between two classes. In this chapter we will be concerned with not just whether there *are* differences but with where on the object they are and what geometric type they have, e.g., bulging, bending, or twisting.

The standard statistical methods for reaching these objectives assume that the feature tuples describing these entities lie in a Euclidean space, i.e., a "flat manifold": the Pythagorean Theorem applies. This is the case for the main method of Gaussian probability distribution estimation, Principal Component Analysis (PCA); the main methods for classification, including Support Vector Machines (SVM) and Fisher linear discriminants; and the main methods for hypothesis testing, including the t-test.

2. Objects live on curved manifolds

But as explained below, geometric object properties (GOPs) do not lie on an abstract flat manifold. Rather, that manifold is best understood as curved, so the standard statistical methods do not strictly apply. The following discusses some of the most common GOPs and explains why they live in curved spaces.

a. Point distribution models

The simplest GOPs are tuples of locations, most commonly spatially sampling the boundary of the object (Fig. 2). These are called "point distribution models (PDMs)" [Cootes and Taylor 1993; Cooper, Cootes, Taylor, Graham 1995]. We will call PDMs 0th order GOPs. Let $\underline{\mathbf{z}} = (\underline{x}_1, \underline{x}_2, \ldots \underline{x}_m)$ be such a tuple; it has dimension $\mathcal{D}$m for $\mathcal{D}$-dimensional objects. They initially appear to have Euclidean properties. However, they typically need centering: let us center the tuple by subtracting its center of mass (average of the $\underline{x}_i$) from each location. Then let us compute γ = the square root of the sum of squares of the distances of the centered locations to the origin, taking γ as a GOP. Finally, let us normalize each centered location by dividing it by γ to produce a GOP formed by the tuple of normalized, centered locations:
$((x_{11}'', x_{12}'', \ldots, x_{1\mathcal{D}}''), (x_{21}'', x_{22}'', \ldots, x_{2\mathcal{D}}''), \ldots, (x_{m1}'', x_{m2}'', \ldots, x_{m\mathcal{D}}''))$, where the first subscript indexes the points and in the second subscript "1" refers to the x dimension, "2" refers to the y dimension, …, and "$\mathcal{D}$" refers to the $\mathcal{D}^{\text{th}}$ spatial dimension. This GOP satisfies $\sum_{i=1}^{m}\sum_{j=1}^{\mathrm{D}}\left(x_{ij}\right)^2 = 1$, which is the equation for a point on a unit sphere of dimension $\mathcal{D}$m-1 (a circle is a sphere of dimension 1). Strictly, this sphere is of dimension $\mathcal{D}$m-$\mathcal{D}$-1: $S^{\mathcal{D}\text{m-}\mathcal{D}\text{-}1}$, since $\mathcal{D}$ dimensions have been taken up by the centering. Thus a population of that GOP consists of points on a unit sphere, a curved manifold.

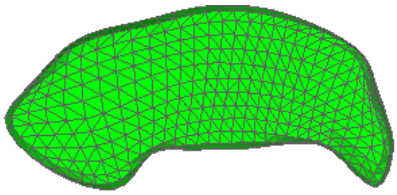


Figure 2. An object boundary in 3D represented by the PDM formed by the vertices of the triangular tiles. The inter-vertex connections are strictly not part of the PDM.

Moreover, the scale parameter γ also is not Euclidean because it cannot take on negative values. Different scales are related to each other by a multiplicative relation, not an additive one. We will produce a "Euclideanized" version of γ by taking its logarithm, and then we will statistically center this value log γ, by subtracting its mean over the training cases. This yields a Euclideanized GOP $\bar{\gamma}\ log(\gamma/\bar{\gamma}\ )$, where $\bar{\gamma}$ is the geometric mean of the training cases' γs.

We conclude that even PDMs, which initially appear to live on a flat space, can be better understood to live on the Cartesian product of a sphere and a 1D flat space, $\mathcal{R}^1$, once the scale parameter is Euclideanized using the logarithm.

b. Spherical harmonic models

A PDM on a normalized, centered object with spherical topology can be alternatively represented by orthonormal basis functions mapping the unit sphere to the boundary of the object (Fig. 3). In 2D these functions are the Fourier (sinusoidal) basis functions, and in 3D they are the spherical harmonics [Kazhdan, Funkhouser & Rusinkiewicz 2003]. Then a particular object's boundary is described by a weighted sum of these basis

functions; that is, it is represented by the ($\mathcal{D}$m-$\mathcal{D}$-1)-dimensional tuple of the weights: the coefficients of the basis functions. One should realize that these mappings vary according to the parameterization of the object surface.

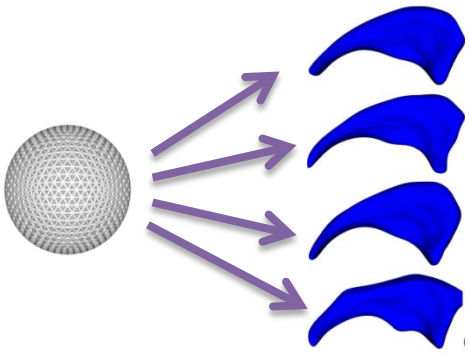


Figure 3. Mapping a sphere to objects via spherical harmonics (by permission of M. Styner)

While these coefficient tuples are normally taken to lie in a Euclidean space, Parseval's theorem guarantees that when they are derived from normalized, centered objects, their sum of squares is the same as the sum of squares of the PDM-points' distances to the origin, i.e., has the value 1. Thus these coefficient tuples also live on a sphere of dimension $\mathcal{D}$m-$\mathcal{D}$-1: $\boldsymbol{S}^{\mathcal{D}\text{m-}\mathcal{D}\text{-}1}$.

c. Normals distribution models

Derivatives of 0^{th} order boundary properties yield a normal vector **u** at each boundary point: 1^{st} order properties. Intuition suggests that the tuple of normal directions $\underline{\mathbf{z}} = (\mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_m)$ better captures the shape (by encoding surface information from a small neighborhood) than the positions themselves.

But each normal is a direction vector; it lives on a unit sphere of dimension $\mathcal{D}$-1: $S^{\mathcal{D}-1}$. Thus a tuple of normals lives on a Cartesian product of unit $\mathcal{D}$-1–dimensional spheres, a curved space.

A view on objects given by Srivastava et al. [Srivastava, Klassen, Joshi, Jermyn 2011] also yields a representation by boundary normals. Their objective was to consider every object's boundary as an equivalence class over the spherical parameterizations described in section 2d. They wished to build a distance function between pairs of such equivalence classes as a distance between the corresponding objects. Section 3c discusses the use of this non-Euclidean distance function for object statistics. The use of this non-Euclidean metric is equivalent to regarding a model as living on a curved manifold.

Srivastava et al. first studied such distances modulo parameterization on objects in 2D [Klassen, Srivastava, Mio & Joshi 2004], and then they extended the method to objects in

3D [Kurtek, Srivastava, Klassen & Laga 2013]. In both cases they showed that a natural distance measure with these properties could be built upon Euclidean comparisons across the boundary if it was represented by a dense sampling of the boundary normal, as follows:

$$\|[q_1],[[q_2]]\| = min_{O\in SO^n,\gamma\in\Gamma}\ d_c\left(q_1, O(q_2\circ\gamma)\right)/\sqrt{\dot{\gamma}}\,,$$

where [q] refers to the equivalence class of object representations that are reparametrizations of q, γ is a reparametrization of the object boundary, Γ is the set of all reparametrizations, $\mathcal{SO}^n$ is the set of rotations, and d_c is the L^2 norm on the boundary normals distribution.

Models of higher spatial derivative order than 1, e.g., those based on curvatures, are also possible, but they are not common and will not be discussed here.

d. Deformation models

Another representation of an object is the deformation of Euclidean space from an atlas for the object, i.e., a reference object formed by a collection of voxels (pixels in 2D) assigned to values 1 if they are interior to the reference object and 0 if they are exterior to that object. This deformation is commonly understood as a displacement vector **d** at each voxel, so the object representation consists of an M-tuple of **d** values, where M is the number of voxels. That is, the representation is of dimension ∂M. Alternatively, the deformation of each pixel is represented by a curved path formed by series of short "velocity" vectors over pseudo-time divided into *m* intervals, whereby the object representation is of dimension ∂mM.

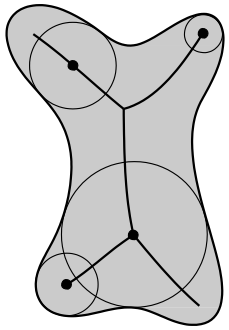
Usually these deformation models are understood to live in Euclidean space. Theoretically, this is justified by arguing that deformations are geodesics on a curved space and geodesics from the reference model can be understood as an initializing direction on that curved space, which can be understood as a tangent space, i.e., a flat space [Beg 2005]. However, if we look at the component displacement or velocity vectors, it is more geometrically intuitive to see each of them as a direction in ∂ dimensions and a length. The direction lives on a ∂-1-dimensional unit sphere: $\mathcal{S}^{\partial\text{-}1}$. Each length *l*, like the aforementioned scale factors γ, requires the application of the logarithm to produce $\bar{l}\,log\left(l/\bar{l}\right)$ before it lives in a flat space ($\bar{l}$ is the geometric mean of the lengths at that voxel). This formulation understands an object as living in a space $(\mathcal{S}^{\partial\text{-}1})^M \times \mathcal{R}^M$ (after Euclideanization) or of $(\mathcal{S}^{\partial\text{-}1})^{mM} \times \mathcal{R}^{mM}$, depending on whether the deformation is represented by displacements or velocities.

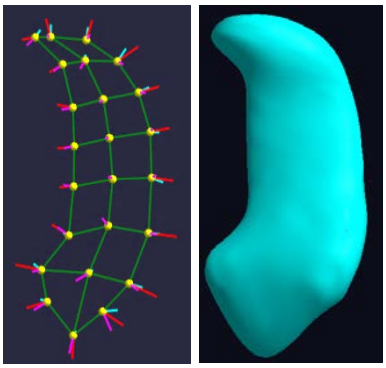
e. Skeletal models

An object has an interior, not just a boundary, and the connections from one boundary position to the other "opposite" it across the object are intuitively important. The length of these connections captures the notion of local object width. Skeletal models (Fig. 4) [Siddiqi and Pizer 2008] capture these relations by having the following two components:

1) A skeleton: a curve in 2D and a surface in 3D ideally halfway between the connected boundary points. This skeleton is a sort of collapsed version of the boundary and thus has the same topology as the boundary; the fold curve (points in 3D) of that surface divides one side of the boundary from the other.
2) Non-crossing vectors that we call "spokes" extending from each skeletal point to the boundary. The lengths of these spokes capture the object half-widths. The mathematics of such models, algorithms for computing them from the object boundary, and applications of these are covered at length in [Siddiqi and Pizer 2008].

a) b)





c)

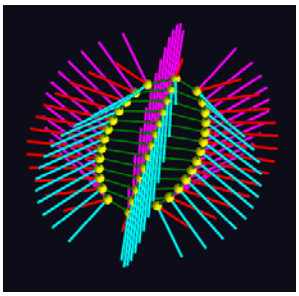


Fig. 4. a) An object boundary in 2D and its skeletal (here, medial) model; the medial model is built from the bitangent circles; b) left: a skeletal surface for a hippocampus with selected spokes -- the curve along the exterior of the mesh forms the curve at which the skeletal surface is folded; right: the boundary implied by it. c) An s-rep for an ellipsoid in 3D (by permission of J. Vicory).

i. Medial models

The earliest forms of such skeletal models were called "medial models". The most common form of medial model was due to Blum [1973]. The Blum medial axis consists of a skeleton which is a doubling (two-sided, folded version) of the locus of spheres bitangent to the object boundary and interior to the object, together with spokes formed by the radii connecting the sphere center to the points of boundary bitangency. These spokes are orthogonal to the boundary. The spokes proceeding from the fold curve touch the boundary at crest points (a type of curvature extremum [Koenderink 1990]).

The Blum medial model for a typical object is highly branching, with different branching patterns for different objects in a population. This makes statistical analysis of these

models problematic. Pruning the branches to form a representation with a fixed branching pattern over a population usually makes the boundary implied by the locus of spoke vector ends too far from the true object boundary to be useful.

Two forms of medial model computer representations have been developed: one due to Pizer et al. produced by sampling the medial locus and associated spokes, and another due to Yushkevich based on parameterizing splines. Both are described in [Pizer ch 8 of Siddiqi and Pizer 2008]. While some statistical analysis has been done using each, the s-reps below have been shown to be more successful for statistical analysis.

ii. S-reps

S-reps are skeletal models that are designed for statistical analysis by fixing the object branching pattern but still having the skeletally implied boundary close to the actual object boundary. They do this by still having the structure of a folded surface but relaxing the conditions of having the skeletal surface be precisely centered from the boundary and of having the spokes precisely orthogonal to the object boundary; instead they fit a reference model [Pizer, Jung, Goswami, Vicory, Zhao, Chaudhuri, Damon, Huckemann, Marron 2013, Tu, Yang, Vicory, Zhang, Pizer, Styner 2015 according to an objective function that rewards fit to the object boundary and penalizes deviations from mediality and spoke-to-boundary orthogonality. A computer representation is formed by sampling the skeleton and associated spokes (Fig. 2b). There are spokes on the skeletal manifold as well as on the skeletal fold curve; the latter proceed to crest points on the implied boundary. Thus this representation consists of a tuple of *n* spokes, each consisting of a skeletal point **p**, a spoke direction (unit vector) ***U***, and a spoke length *r*. In implementation the skeletal locus is sampled, assuring that spokes on the fold are included. Such representations live on the manifold formed by the Cartesian product of the PDM manifold of the *n* skeletal points: $\mathcal{S}^{3n-4} \times \mathcal{R}^1$, the *n* 2-spheres $\mathcal{S}^2$ on which the spoke directions live, and $\mathcal{R}^n$, containing the Euclideanized spoke length values $\bar{r}\, log(r/\bar{r})$, where $\bar{r}$ is the geometric mean of the lengths of *that* spoke over the population.

3. Statistical analysis background

This section reviews the basic statistical methods used in the shape analysis contexts of this paper. Section 3.a considers the case of conventional Euclidean data. Section 3.b describes fully intrinsic approaches. Tangent plane methods are described in Section 3.c.

a. Statistical analysis in Euclidean space

Several statistical tasks are routinely addressed in shape analysis, i.e. summarizing and modeling populations of shapes, usually based on a sample from that population. The first task is understanding *centerpoint*, which is usually calculated as the sample mean. After the center is understood, the next task is to consider the variation about that center. Because shape statistics are typically High Dimension, Low Sample Size problems (using terminology from [Hall, Marron & Neemon 2005]) in nature, standard quantifications of variability such as the *covariance matrix* are hard to estimate. Thus it is natural to analyze variability using *Principal Component Analysis* (PCA, Fig. 5)). See [Jolliffe 2002] for a good introduction and discussion of many important aspects of PCA.

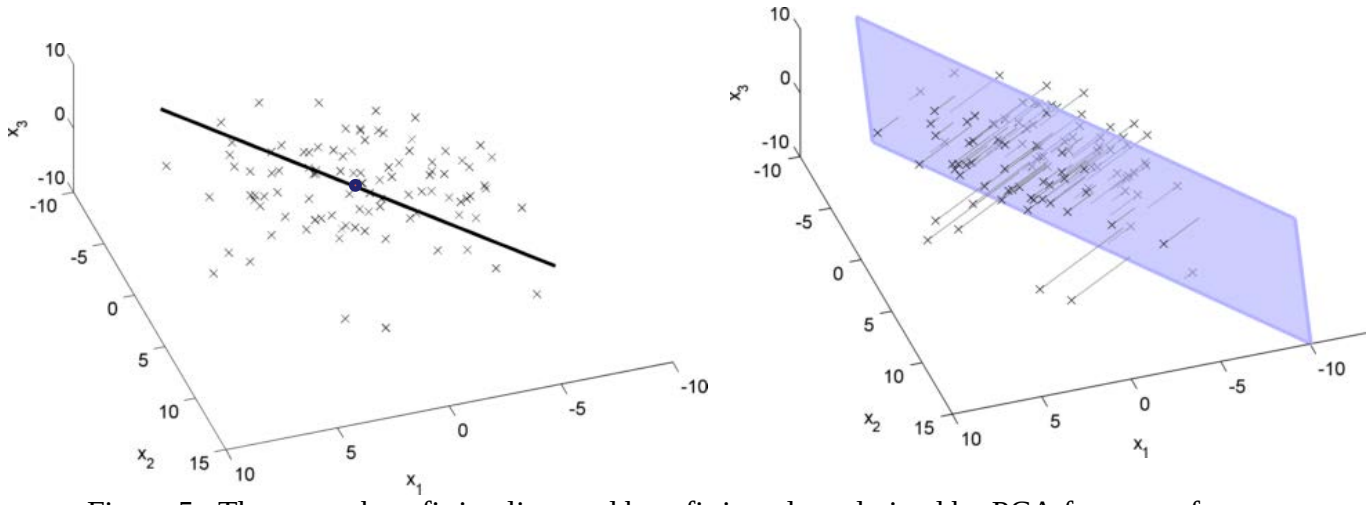


Figure 5. The mean, best fitting line, and best fitting plane derived by PCA for a set of observations in $\mathcal{R}^3$.

The third task, useful for many purposes, for example incorporating external information such as anatomical structure into an analysis through Bayes-like methods, is *probability distribution modeling*. Gaussian distributions are far and away the most commonly used, mostly because many natural distributions are approximately Gaussian due to the Central Limit Theorem. A second reason that Gaussian distributions commonly appear in shape analysis is their Bayes conjugacy properties; that is, Gaussian priors and likelihoods lead to Gaussian posteriors. This is important as it allows closed form calculations instead of the complicated simulation-based approaches used in most modern Bayes analyses.

A fourth task is statistical *classification* (also called *discrimination*). For this task training data, with known class labels, is given and is used to develop a classification rule for assigning new data to one of the classes. For Euclidean data objects, there are many methods available; see [Duda, Hart & Stork 2001] for a good overview.

The most common methods for classification are based on Euclidean spaces. Particularly widely used these days is the method called support vector machines (SVM); see, for example, [Schölkopf & Smola, 2002] for detailed discussion. SVM is based on optimizing the gap in feature space between the training cases in the two classes. A more statistically efficient method called Distance-Weighted Discrimination (DWD), still in a Euclidean feature space was developed by [Marron, Todd & Ahn 2007]; its efficiency derives from it using all of the training data, not just those near the gap. Both SVM and DWD yield a direction in feature space that optimally separates the classes (Fig. 6). Classification then involves projecting the feature tuple onto the separation direction and deriving the class or the class probability from the resulting scalar value.

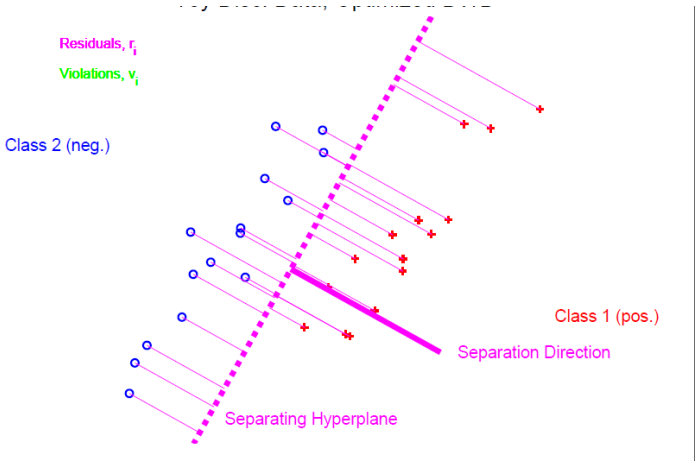


Figure 6. The idea of the separation direction determined from SVM or DWD (each method will determine a somewhat different separation direction)

The above tools are useful in situations where the full set of GOPs is available, and they are preferable in that case. But in some situations only pairwise distances between the data objects, i.e., the *distance matrix*, are available. Some analysis can still be done in that case. For example, a notion of centerpoint that can still be computed is the *Fréchet mean*, proposed by [Fréchet 1948]; which has a number of synonyms including *barycenter*. The Fréchet mean is actually defined, for an arbitrary metric space, as the point which minimizes the sum of squared distances to the data objects (or points, since the arg minimum may not be unique). There also is an analog of PCA, called *multi-dimensional scaling*, see [Torgerson 1952] and [Gower 1966]. The main idea here is to solve an optimization problem whose solution is a set of points in Euclidean space whose distance matrix best approximates the given one. When the input distance matrix is Euclidean, multi-dimensional scaling results in scores that are the same as PC scores. Multi-dimensional scaling scores can also provide a surrogate population for further analysis, but an important issue that will be central to the following discussion is that for non-Euclidean (e.g., manifold-based shapes) data, this represents an approximation. In some situations, the approximation is adequate. Situations where that approximation is inadequate and alternate approaches give better results are the focus of Section 4.

b. Fully extrinsic analysis on manifolds

The challenge of statistical analysis of populations lying on manifolds has been addressed in a number of different ways. The diversity of these can be seen already in the simple but nontrivial case where the data objects lie on the unit circle. That topic is often called *directional data* and naturally arises in contexts where the data objects are angles such as wind or magnetic field directions. As seen in the monographs of [Mardia 2015] and [Fisher 1995], there is a large literature on the statistical analysis of directional data. In that context, even for the simplest statistical summary of mean or center point of a data set, there are divergent reasonable choices, which have been characterized as extrinsic and intrinsic (see [Bhattacharya & Patrangenaru 2003, 2005] and [Patrangenaru &

Ellingson 2015] for an overview of these issues). In general, extrinsic summaries are first computed in the ambient space (the Euclidean space in which the manifold is embedded) and then projected back to the nearest point on the manifold. Thus the extrinsic mean for directional data is the vector average in $\mathcal{R}^2$ of the data vectors, which is projected to the circle by using the angular part of its polar coordinate representation. Intrinsic summaries strive to work more within the manifold.

Extrinsic versions of PCA are not widely used in shape analysis, perhaps because there is potential for very strong distortions while projecting the resulting PC scores back to the manifold. Hotz [2013] did a detailed comparison of the intrinsic and extrinsic mean for data that consists of points in the unit circle. A number of criteria were considered, and each form of mean had its relative strengths and weaknesses.

c. Distance-based statistical analysis methods

For data lying on a manifold, a natural distance is the geodesic distance (the length of the shortest path along the manifold). Computing the matrix of pairwise distances between the collection of data points leads to approximate analyses as described in Section 3.a. This includes the Fréchet mean, which in this case is also called the geodesic mean.

An important example of the non-uniqueness of the Fréchet mean is a data set that is distributed widely around the equator of S^2, the 2D sphere, where both the north and south poles are minimizers of the Fréchet sum of squares. Neither of these means is close to the data in that example.

Another distance-based analysis approach, which is a variation of the Euclideanization idea discussed above, is to represent the data using the multi-dimensional scaling scores of each object to each of the objects and then to proceed with standard Euclidean analysis methods.

d. Tangent plane statistical analysis methods

The first generation of intrinsic analogs of PCA for the analysis of manifold data are based on the definition of a manifold, which is a surface in the ambient space which is smooth in the sense of having an approximating hyperplane (in the sense of shrinking neighborhoods) at every point. In this spirit, [Fletcher, Lu, Pizer, & Joshi 2004] proposed *Principal Geodesic Analysis* (PGA) (Fig. 7). This is based on the plane that is tangent at the Fréchet mean. The data on the surface of the manifold are represented as points in the tangent plane using the *Log map*. PCA is then performed there, and the resulting eigenvectors and summarized data are mapped back into the manifold using the *Exponential map*. The corresponding scores give another type of Euclideanization.

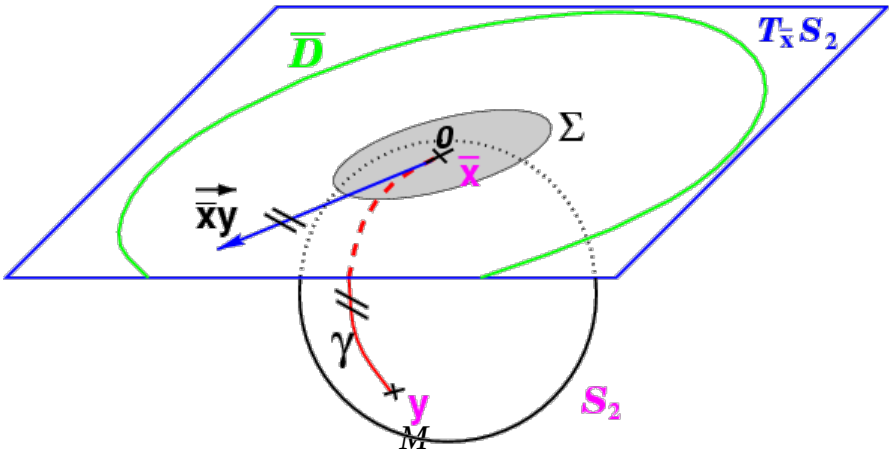


Figure 7. Tangent plane statistical analysis (by permission from X. Pennec)

4. Advanced Statistical Methods for Manifold Data

While much useful shape analysis has been done using the statistical methods described in Section 3, for many data sets large gains in statistical efficiency have been realized through the development of more sophisticated methods. The $\mathcal{S}^2$ example given in Section 3.c above, where the data lies very near the equator of a sphere, illustrates an important aspect of this challenge. While the data are essentially *one dimensional* in nature (since they just follow a single geodesic), the PGA requires *two* components (where the projections follow a circle) to appropriately describe the variation in the data. This type of consideration has motivated a search for more statistically efficient approaches to statistical analysis of data lying on a manifold. The first of these is *Principal Nested Spheres* (PNS), motivated and described in Section 4.a. Extension of this to more complicated manifolds, such as the polyspheres central to s-rep shape representations is given in Section 4.b. Section 4.c discusses a yet more efficient approach to polysphere analysis involving a high-dimensional spherical approximation followed by a PNS analysis.

a. Principal Nested Spheres

In the case of data lying in a high dimensional sphere $\mathcal{S}^k$ embedded in $\mathcal{R}^{k+1}$, a useful intrinsic version of PCA is *Principal Nested Spheres* (PNS), proposed by [Jung et al. 2012]. The central idea is to iteratively find a nested (through dimension) series of subspheres, each of which provides an optimal fit to the data (Fig. 8). In particular, at each step the dimension of the approximation is reduced by 1, finding the subsphere which best fits the data in the sense of minimum sum of squared residuals, measured using arc length along the surface of the sphere. The signed residuals are also saved as PNS scores for that component. The concatenation of these scores, over the dimensions, becomes the PNS Euclideanization of each data point. The advantages of this approach

are tractability, since each lower dimensional manifold is determined by the imposition of a single (usually easy to find) constraint, and statistical efficiency.

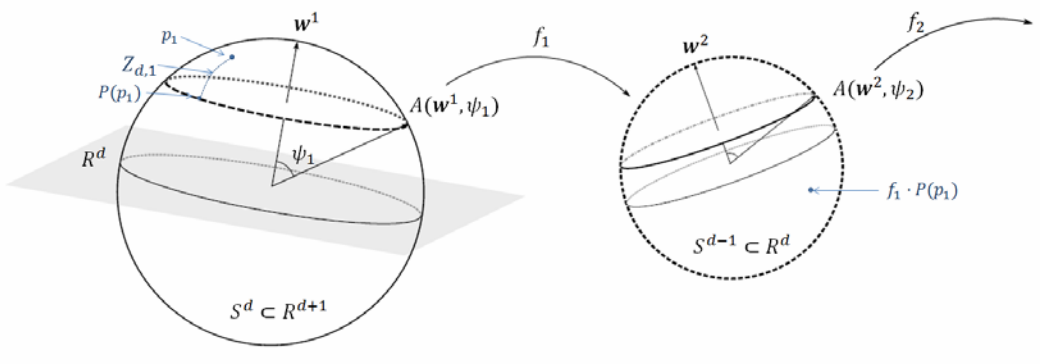


Figure 8. An optimal small subsphere, geodesic distances ($Z_{d,1)}$ forming scores, and projections onto the subsphere (by permission from S. Jung)

One reason that PNS was an important statistical landmark is that it motivated the more general idea of *Backwards PCA* as a general paradigm for finding principal components in non-Euclidean data contexts. The full generality of this idea can be found in [Damon & Marron 2014]. A key concept is that the general utility of backwards PCA follows from thinking of PCA in terms of a nested series of constraints. Backwards tends to be easier to work with because from that viewpoint, the constraints can be found sequentially, instead of needing to know the full set and then sequentially relaxing them. This idea is seen to generate (or have the potential to generate) useful analogs of PCA in a variety of other non-Euclidean settings such as Nonnegative Matrix Factorization, Manifold Learning, and on other manifolds.

The S^2 example given above, with data widely distributed around the equator, also illustrates a sense in which the Fréchet mean can be a poor choice that is not at all representative of the data. The *backwards mean* is an intrinsic mean that is much more representative of the data in that example than the Fréchet mean. Generally this is computed by taking one more step in PNS. In particular, the backward mean is the Fréchet mean of the rank 1 circular representation of the data, which can then be viewed as the best backwards rank 0 approximation.

b. Composite Principal Nested Spheres

[Pizer, Jung, Goswami, Vicory, Zhao, Chaudhuri, Damon Huckemann & Marron 2013] proposed extending PNS to manifolds that involve products of spheres, such as those for various shape representations discussed in Section 2 above, using the idea of Composite Principal Nested Spheres (CPNS). The idea here is to first develop the PNS representation for each spherical component, and then concatenate these, together with Euclidean components, into a large Euclidean representation, followed by PCA on the result.

Of importance is the commensuration between the components before the PCA is done. In preparation for the comparisons between using original features of PDMs of hippocampi and the alternative Euclideanized features in [Hong, Vicory, Schulz, Styner, Marron & Pizer 2016], an experimental paradigm was designed to determine the most reasonable commensuration on the Euclideanized features, more precisely, on the scale factor and the sphere-resident features derived from the centered, normalized PDMs. First, the features were transformed to be in the same units: in our case the log-transformed version of the scale factor γ (i.e., $\bar{\gamma}\ log(\gamma/\bar{\gamma})$) had units of millimeters, and the Euclideanized shape features derived by PNS from the high-dimensional unit sphere on which the scaled PDM tuple for each case live had units of radians (unitless values θ_i for each of the dimensions of the unit sphere). Thus we multiplied each PNS-derived feature by $\bar{\gamma}$ to put them into units of distance. The problem then is to determine the scalar factor to commensurate the feature capturing scale with the PNS-derived features. We determined that scalar factor by creating a new population that would have a non-varying shape consistent with those in the original population and a scale variation that was the same as those in the original population. To do this, we formed the new population by applying the measured log-transformed γ values for each case to the hippocampus of median scale scaled to have γ=1. By comparing the total variances of the original and created populations, respectively, one could determine the correct commensuration factor between $\bar{\gamma}\ log(\gamma/\bar{\gamma})$) and the Euclideanized features from PNS, namely, $\bar{\gamma}\theta_i$. The experiment concluded that the correct commensuration factor was 1.0, up to sample variation. This idea of separately treating scale can be used for problems of commensuration of other types of variation.

c. Polysphere PCA

Data spaces that are products of spheres, such as the skeletal model spaces of Section 2.e, have been called polyspheres by [Eltzner et al. 2015]. That paper goes on to propose a new method, *Polyshpere PCA*. This is a potentially large improvement over CPNS which allows a more flexible modelling of the dependence between features than is done by the Gaussian PCA on the Euclideanized data in CPNS. This is achieved by approximating the polysphere manifold with a higher dimensional tangent ellipsoid, projecting the data onto that, then squashing the ellipsoid into a sphere, and using PNS on that. On skeletal shape data, this approach has been shown to give a lower dimensional probability distribution representation than is available from CPNS.

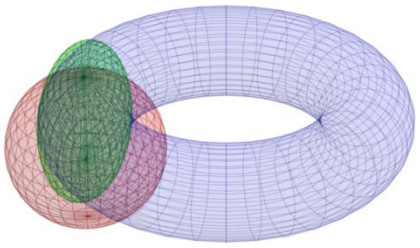


Figure 9. The ellipsoid tangent to a torus, which is a polysphere $S^1 \times S^1$, together with the sphere into which the ellipsoid is squashed (by permission of B. Eltzner)

d. Barycentric Subspace Analysis

An important new general view of PCA analogs for manifold data has been provided by [Pennec 2016]. A fundamental observation motivating that work is that both the backwards methods (e.g., PNS) that were explicitly described in [Marron & Damon 2014], and forward methods such as PGA and Geodesic PCA [Huckemann, Hotz & Munk 2010] rely upon greedy sequential searches (in opposite directions in some sense). This new work proposes the appealing idea of a simultaneous search. The goal of such a search is an appropriate analog of a properly nested sequence of subspaces that is found by a complete (in the sense of simultaneously over subspaces of all ranks) PCA in Euclidean space. This analog is called a *flag*, which is a sequence of *nested* (where those of lower dimension are contained in the higher dimension members, as in Euclidean PCA) sub-manifolds. Appropriate components for the flag subspaces in manifolds are constructed as *barycentric subspaces* which are affine combinations (a generalization of Fréchet, Karcher or exponential weighted averages where negative weights allow appropriate extrapolation) of a set of *reference points*. Nesting of the manifolds is elegantly achieved through an appropriate nesting of the control points. In this mechanism geodesics and thus inter-object distances are nicely handled by restriction to regions bounded by *cut loci* associated with each of the reference points and determined by the original object's topology. The computational implications of this property are still a matter of research.

e. Bayes methods for manifold data

Motivated by the desire to develop Bayes statistical methods on manifolds, Fletcher [2013] has noted that the statistical approaches based on Gaussians in Euclidean spaces follow three criteria:

1) that the probability is *shift-equivariant* (this is the mathematical term; engineers call this "shift-invariant"). That is, for all vectors **v**, $p(\underline{x} \mid \underline{\mu}, \sigma) = p(\underline{x}+\mathbf{v} \mid \underline{\mu}+\mathbf{v}, \sigma)$.
2) that the probability distribution is *scale-equivariant* (for engineers: "scale-invariant"). That is, for all scale factors s, $p(\underline{x} \mid \underline{\mu}, \sigma) = (1/s^d)\, p(s\underline{x} \mid s\underline{\mu}, s\sigma)$, where *d* is the dimension of the space.
3) that the Fréchet mean is the mode of the probability distribution. That is,

   $\underset{\underline{\mu} \in \Re^d}{\arg min} \sum_{i=1}^{N} \left|\underline{x}_i - \underline{\mu}\right|^2 = \underset{\underline{\mu} \in \Re^d}{\arg max} \prod_{i=1}^{N} p\left(\underline{x}_i \middle| \underline{\mu}, \Sigma\right)$, where *d* is the dimension of the variables $\underline{x}$ and N is the number of points.

He gives requirements for the same properties to be followed for probability distributions on a curved manifold. When these requirements are satisfied, operations such as probability distribution estimation, regression, classification, and hypothesis testing can be accomplished by means very similar to those used in Euclidean spaces.

f. Manifold techniques based on Brownian motion

[Sommer 2015] understands Gaussian distributions as formed by Brownian motion, with a covariance that is stationary in space. He derives a curved-manifold counterpart to

Brownian motion in Euclidean space by augmenting the positions on the surface with a fitted frame (coordinate system) that is transported from the origin of the motion along a curve on the surface by parallel transport. This allows the definition of stationary covariance that is dependent on the curve, and this in turn allows the definition of a Brownian-produced probability that is dependent on the curve. Based on this careful mathematics, an implementation calculating the most probable curve from the origin to any other point on the manifold, given the covariance matrix, has been completed (Fig. 10), and an implementation calculating the regression curve to a set of data points, given the covariance matrix, is in progress.

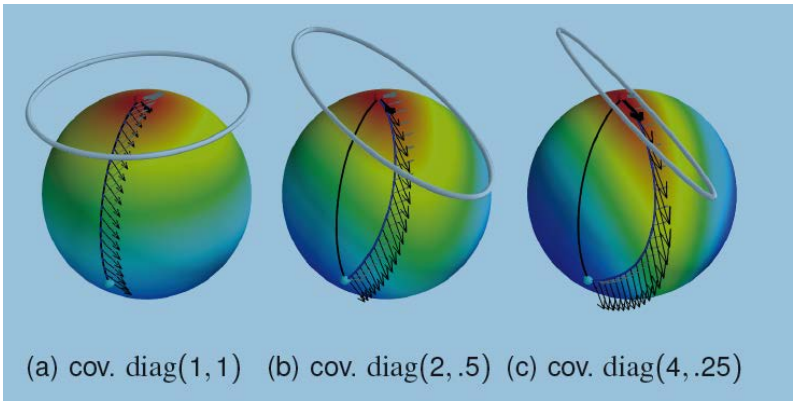


Figure 10. Most probable paths on a sphere with different Brownian covariances. In each subfigure the curve without arrows is the geodesic path between the points. (by permission of S. Sommer)

e. Manifold Classification

Use of DWD with object data is often done by Euclideanizing the object features and then applying DWD to the result. As shown in [Hong, Vicory, Schulz, Styner, Marron, Pizer 2016], this can improve classification accuracy over using the object features directly in DWD.

A large open research problem is the development of intrinsic classification methods on manifolds. A first attempt, using SVM ideas, can be found in [Sen, Foskey, Marron & Styner 2008] (Fig. 11). Given the major improvements in PCA from using intrinsic ideas discussed in Sections 4a-c, we believe that large gains in classification error rates can be made by creative work in this direction.

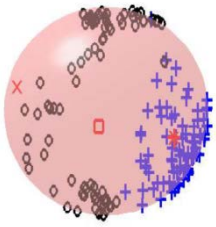


Figure 11. Classification on the manifold by optimizing two control points, shown as "×", such that the separation direction is the geodesic between the control points [Sen, Foskey, Marron & Styner 2008]

5. Correspondence

Statistics on objects depend on a good correspondence between the primitives on each object in the training or target set with the others. For example, it would not do to have the position or normal direction at a fingertip on one hand object in the set be considered the corresponding GOP to that at the knuckle in another hand object in the set. Putting things in correspondence can be understood as a reparametrization of the object representation. But how can this correspondence be determined? Roughly, it can be done in a geometric way or a statistical way.

The geometric way of producing correspondence works by finding the reparametrization on objects that brings them closest to each other in ambient space. This requires a metric between a pair of objects. For example, this could be the L^2 norm of the GOP differences between corresponding positions, with the integration across the object (e.g., its boundary or its skeletal surface). This metric is best if it recognizes that the objects lie on a curved manifold. With such a metric, one can find the correspondence between two objects by reparametrizing the second to minimize the metric between the two. This approach can be applied between all pairs of objects in the training set, or it can be applied between a reference object and each of the objects in the training set. Such a method is discussed in section a below.

The statistical means of producing correspondence is based on properties of the whole set of training objects, in particular on the probability distribution estimated from those objects. The concept is that miscorrespondence widens that distribution, so correspondence optimization should involve reparameterizing each object in the set so as to optimally narrow the distribution, while each object's GOP is still a good descriptor of the whole object (Fig. 12). Such methods are discussed in section b below.

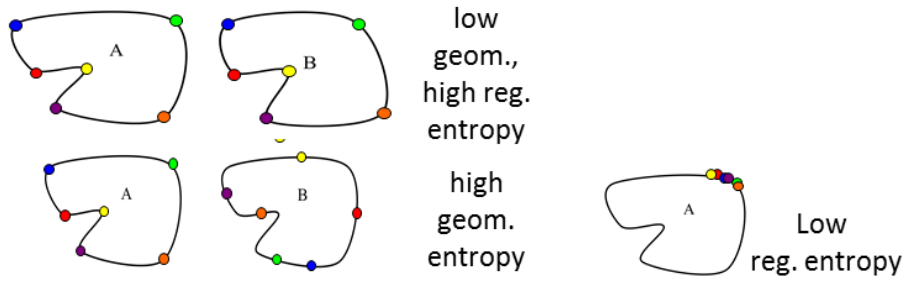


Figure 12. Correspondence: good (low geometric entropy) vs. bad (high geometric entropy); irregular sampling (low regularity entropy) vs. regular sampling (high regularity entropy)

a. Correspondence via reparametrization-insensitive metrics

As discussed earlier in section 1.c, [Klassen, Srivastava, Mio & Joshi 2004] produced a method for objects in 2D that allowed a metric between equivalence classes of objects over reparametrizations. The mathematics required that the comparison be over derivatives of the object boundary. It follows that Kurtek's [Kurtek, Srivastava, Klassen

& Laga 2013] generalization to 3D had to be over normal directions on the boundary. These metrics are seen to be invariant with respect to reparametrization, which allows them to be easily adapted to the quotient space under reparametrizations, That gives a notion of shape as orbits, i.e, equivalence classes with respect to reparametrization.

The method for computing the metric of the difference between two objects described in section 1.c involves picking a particular parametrization of one of the objects and finding the parametrization of the other yielding the closest field of surface normals (together with a global rotation). It thereby yields a correspondence between the locations on the surfaces of the object pair.

Finding the Fréchet mean equivalence class, and a central representer of the class gives a template mean representative. Their method allows a distance to be calculated between a reference object, e.g., the template mean, and each object in the training set. As just described, it allows a distance to be calculated between objects, so the multidimensional scaling approach can be used.

b. Correspondence via entropy minimization

The tightness of a probability distribution can be measured in many ways, but the ones that have turned out to be the most effective are based on information theory. Taylor and his team [Davies 2003] pioneered a form based on minimum description length (MDL), and an almost equivalent form was based on entropy, as developed in Whitaker's laboratory [Cates 2007]. The basic idea is to fit the population of GOPs by a Gaussian and minimize its entropy over all object reparameterizations.

The entropy of an n-dimensional Gaussian is given by
n [½ ln(2π+1) + the mean of the principal variances]. The difficulty is that when the idea is applied to a probability distribution estimated from data by a PCA or PCA-like method, the successive sorted principal variances get successively smaller and eventually are dominated by noise in the data, and each negative logarithm of these small principal variances is very large, so these noise-induced principal variances dominate the population principal variances in the entropy calculation. Cates et al. [2007] dealt with this problem by adding a small constant to each principal variance. [Tu, Styner, Vicory, Paniagua, Prieto, Yang, Pizer 2015] dealt with it by cutting off the series when the ratio of the principal variance to the total variance fell below a certain threshold.

Cates et al. [2007] showed that entropy minimization for GOPs that were a boundary PDM could yield improvements in hypothesis testing significances. In their work boundary points were slid along the boundary to minimize entropy. [Tu, Styner, Vicory, Paniagua, Prieto, Yang, Pizer 2015] developed the same idea for s-reps, whereby the skeletal spokes were slid along the skeletal surface to minimize entropy. They showed (Fig. 13) the surprising result that, according to the measures described in section 6c on a training set of hippocampi, when the s-rep-based correspondence method was applied, the boundary points at the ends of the spokes had superior statistical properties (specificity and generalization, see section 6.c) than those produced by the Cates boundary point shifting method.

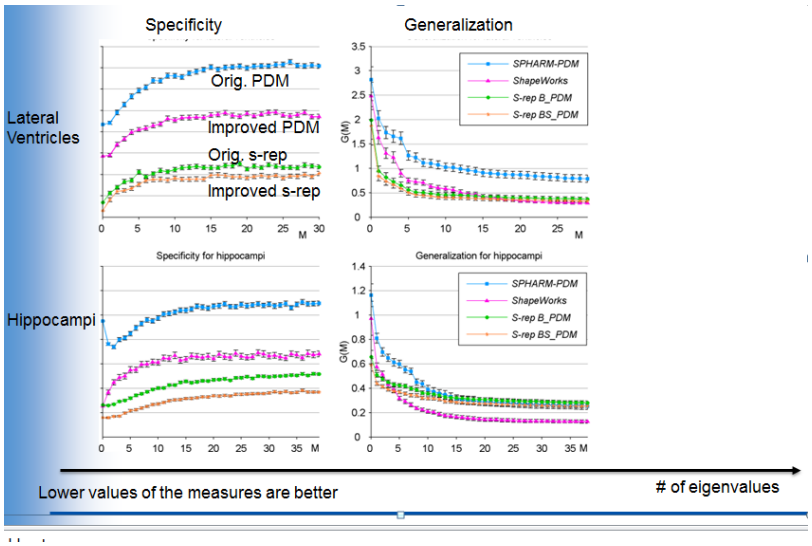


Figure 13. Generalization and specificity improvements due to s-rep-based and PDM-based correspondence optimization

6. How to compare representations and statistical methods

Given all of the different shape representations discussed in Section 2, and the analytic approaches discussed in Sections 3 and 4, an important issue is comparison of them. One basis for this is knowledge of how they work, which can lead to sensible choices in the wide variety of shape problems that one may encounter. But it is also interesting to consider quantitative bases for comparison, which is done in this section.

a. Classification accuracy

When a data set is naturally grouped into 2 (or perhaps more) subsets, e.g., pathology vs. normal controls, various methods can be compared on the basis of *classification accuracy*. The classification problem starts with a group of labeled data called the *training set*, and the goal is to develop a rule for classifying new observations. Classification accuracy is simply the rate of correct classifications, either for an independent test set, or using some variation of the cross-validation idea. See [Duda, Hart & Stork 2012] for access to the large literature on classification.

b. Hypothesis testing power

A related approach to comparing shape methodologies, again based on two well labeled subgroups in the data, is to construct a formal hypothesis test for the difference between the groups. Quantification of the difference then follows from the level of statistical significance, allowing a different type of comparison of the relative merits of various approaches to the analysis. This was done to good effect in [Schulz, Pizer, Marron & Godtliebsen 2016].

c. Specificity, generalization, compactness

In object statistics work in medicine two common measures of a probability distribution derived from data via its eigenmodes are specificity and generalization [Davies 2003]. Specificity is a measure of how well the estimated probability distribution represents only

valid instances of the object. It is computed as the average distance between random samples in the computed shape space with their nearest members of the data. Generalization is a measure of how close new instances of the object are to the probability distribution estimated from the training cases. It is calculated by computing a shape space, spanned by the eigenmodes, on all but one of the training cases and computing the distance between the last shape and its projection onto this shape space. Thus both specificity and generalization are measured in units of GOP differences, e.g., positional differences when the GOP is a position tuple or normal direction differences when the GOP is a normal direction.

Compactness is a measure of the tightness of a probability distribution. The entropy of the distribution or the determinant of its covariance matrix (total variance) are often used.

d. Compression into few modes of variation

The Euclidean PCA decomposition of data into modes of variation is also usefully understood from a signal processing viewpoint. In the case of a low dimensional signal, in the presence of noise (which is high dimensional by its definition of spreading energy across the spectrum). PCA provides a data driven basis (in the sense of linear algebra) that puts as much of the low dimensional signal as possible into a few basis elements with largest variance. Many of the gains in statistical efficiency, such as those discussed in Section 4, can be understood in terms providing better signal compression in these terms. However, this analogy fails in the presence of even moderate noise in Euclidean data, and the problem appears even more strongly in non-Euclidean contexts such as shape analysis. In particular, the standard Euclidean assumption of all of the signal being present on in the first few eigenvalues is usually misleading because both nonlinear shape signals and noise typically spread some signal power among many eigenvalues. The result is that very noisy data can be measured to require fewer eigenmodes to achieve a given fraction of total variance than less noisy data. Hence standard *dimension reduction* approaches, based on "total signal power" or "percent of variation explained", are usually inappropriate.

Yet there is still a natural desire to think in terms of *effective dimensionality*, i.e., the concept that the true underlying signal has much less variation than is present in noisy data. Ideas based in *random matrix theory* are promising avenues for research in this direction. See [Yao, Bai & Zheng 2015] for good discussion of using this powerful theory in the context of Euclidean PCA. A first important part of this theory is the asymptotic probability distribution of the full collection of eigenvalues under a pure noise model, called the Marčenko-Partur distribution [Marčenko & Pastur 1967]. Second is the corresponding limiting distribution of the largest eigenvalue, called the Tracey Widom distribution [Tracy & Widom 1994].

e. Quality in application, esp. segmentation

One more approach to comparing shape analysis methods is to study their impact when used for various applications. An important application has been to the *segmentation* problem in medical image analysis, where the goal is to find the region of an image occupied by an object such as a particular organ. A series of successive improvements in

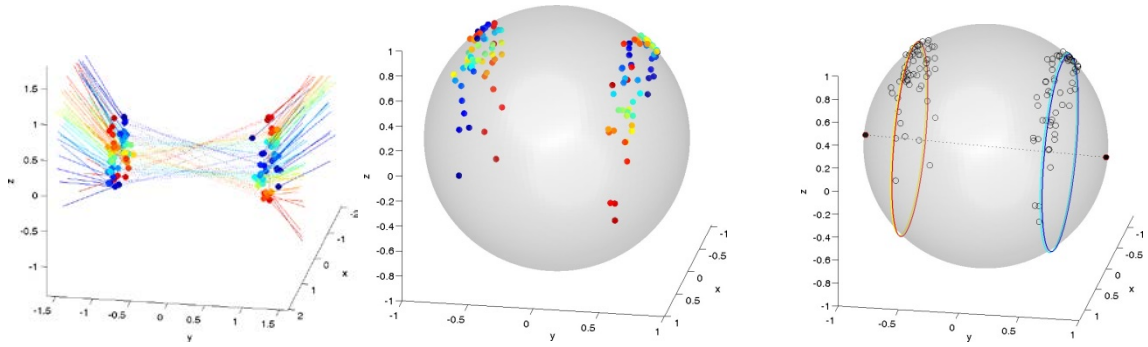
shape analysis resulting in improved image segmentation can be found in Pizer et al. [1999, 2001], Fletcher et al. [2004], Rao et al. [2005], Gorczowski et al. [2007], Pizer et al. [2013] and Vicory et al. [2014, 2016].

7. Results of classification, hypothesis testing, and probability distribution estimation

This section reviews some recent work, with a number of specific applications of the above ideas.

a. Generalized rotations

[Schulz, Jung, Huckemann, Pierrynowski, Marron & Pizer 2015] studied statistics on generalized rotations as characterized by boundary normal directions at a small number of boundary locations. The generalized rotations they considered were global rotation, bending, and twisting. They showed that given a collection of cases with one of these rotations but with statistically varying angles of rotation about a fixed axis, they could derive the axis as well as variance of the rotation angle. The axis derivation was based on the realization that the data fell on coaxial small circles on the sphere of directions (Fig. 14). Next, they studied compositions of two of these types of generalized rotations, each about its own axis. Finally, they developed hypothesis tests for establishing statistical significance of variation in these directions and used it for new scientific insights into human knee movement.



(a) Random twists (b) Random directions (c) Estimated axis

Figure 14. Random twists applied to two normals, those normal on the 2-sphere, and the twist axis estimated from the co-axial circles.

b. Classification of schizophrenia via hippocampus s-reps

[Hong, Vicory, Schulz, Styner, Marron & Pizer 2016], compared PDMs vs. s-reps and Euclideanization vs. direct Euclidean analysis of the ambient space coordinate values in classifying a hippocampus as to whether it was from a typical individual or from a first-episode schizophrenic. They showed that, according to areas under the Receiver Operating Characteristic curve (ROC) [Hanley & McNeil 1982], Euclideanizing boundary PDMs produced better classification than without Euclideanization and that Euclideanized s-reps produced better classification than either of the boundary-based analyses (Fig 15). They also showed the usefulness of displaying the variation of the

object along the vector in the Euclideanized feature space passing through the pooled mean and in the separation direction of the classes (Fig. 16).

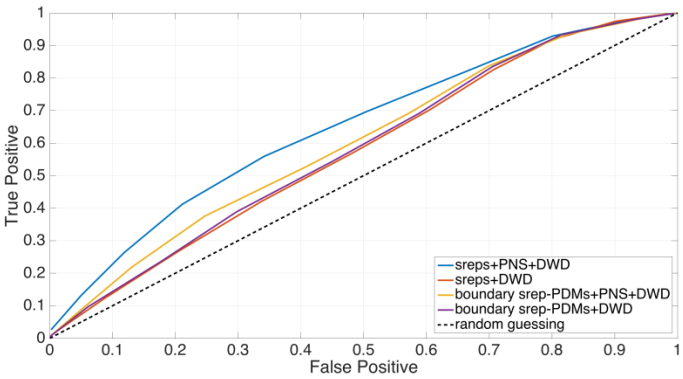


Figure15. ROCs for classifying hippocampi as to schizophrenia vs. normal using s-reps and PDMs, each with original features and their Euclideanized counterparts.

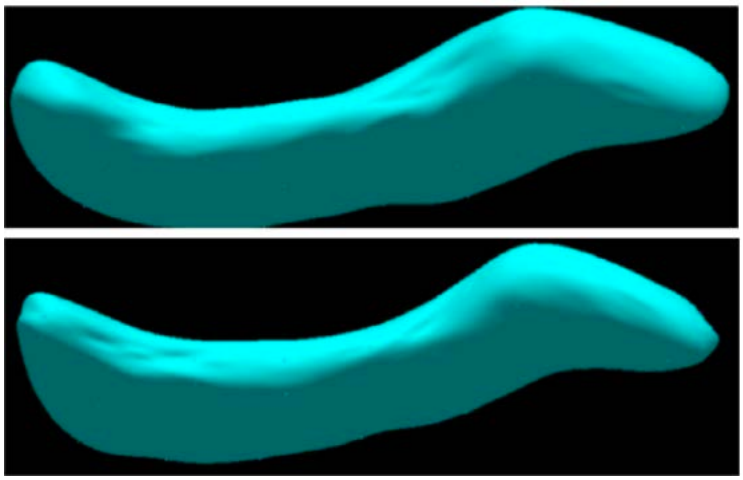


Fig. 16. Boundaries implied by s-reps of hippocampi at (top) 2 standard deviations from the pooled mean in one direction and (bottom) at 2 standard deviations from the pooled mean in the other direction, for classification between first episode schizophrenics and controls.

c. Hypothesis testing via s-reps

[Schulz, Pizer, Marron & Godtliebsen 2016] demonstrated the ability to test hypotheses on shape variation between two objects using s-rep features. They showed how to do not only global hypothesis tests but also GOP-by-GOP and location-by-location tests. The method involved permutation tests that recognize that means benefit from being backward means and that GOP differences need to use a metric appropriate for curved manifolds. This work reported a new method for compensating for the correlations of these GOPs. With this approach they were able to analyze which locations on hippocampi of first-episodes schizophrenics had statistically significant GOP differences from controls, and which GOPs had those statistically significant differences. Thus they found important shape differences between the classes.

d. Shape change statistics

[Vicory 2016] studied statistics on the change of object shape between two stages. Realizing that statistics on change requires transporting each object pair such that the starting object was at a fixed place on the manifold of objects, he proposed solving this problem by pooling all 2n objects in the n shape pairs in the training set, producing a polar system by PNS on the pool of GOP tuples, Euclideanizing each object according to that polar system, and then producing Euclidean (ordinary) differences between the Euclideanized features of each pair.

Vicory studied this method in two applications. The first was on ellipsoids that were randomly bent and/or twisted. The GOPs in this study were the boundary point coordinates provided by spherical harmonic analysis. He found that when the simulated data was produced from random samples from a single transformation type, analysis either with or without Euclideanization yielded a single correct eigenmode, but the analysis with Euclideanization allowed more accurate estimation of an object's bending or twisting angle and, when the mode of variation was visualized, the object moved more naturally than using PCA alone. He also found that when the simulated data came from random bending cascaded with twisting, both types of analysis yielded two modes of variation but Euclideanization allowed a more correct estimate of the relative variances and mean deformations closer to the expected deformations than their PCA-alone counterparts.

Vicory's second test was on prostates segmented from MRI and the same prostate within 3D transrectal ultrasound; the latter was deformed by the ultrasound transducer. The shape change eigenmodes were used as a shape space in which to segment the prostate from the ultrasound image, given the prostate shape of the patient derived from his MRI. The GOPs he used were the skeletal points, spoke lengths, and spoke directions from fitted s-reps. He found that using the shape change space resulting from Euclideanization followed by subtraction yielded more accurate segmentations than when a shape space was formed by s-rep feature differences applied to a mean of the prostates in the training MRIss and then CPNS was applied to the resulting objects. For a target segmentation using that second method, the feature difference in prostate from the patient's MRI from the MRI-based training mean was applied to the ultrasound mean, which became the starting point for the segmentation within the scale space.

Recently Hong is applying this shape change Euclideanization approach to two classes of pairs of shapes, at two different ages. Pooling all 4 of his hippocampus objects to yield a Euclideanization is yielding informative shape differences between children at high risk for autism who do not develop autistic symptoms and those at high risk who do develop such symptoms.

e. Tu correspondence evaluation

[Tu, Styner, Vicory, Paniagua, Prieto, Yang, Pizer 2015] compared analysis approaches through focusing simultaneously on specificity, generalization and compactness. Measuring according to specificity and generalization, they compared probability

distributions derived from boundary PDMs on hippocampi in multiple ways. Briefly, they showed notable improvements in specificity and generalization (Fig. 13) when improving PDMs' correspondence derived via spherical harmonics by Cates' entropy minimization method and, as compared to the results of correspondence improvements via PDMs, notable improvements in specificity with little change in generalization when improving s-reps fitted from a common reference s-rep.

Conclusions

We have shown that object shapes can be considered to reside on curved manifolds and have presented a number of methods for representing the objects and calculating statistics that recognize that curvature. We have presented methods for comparing these representations and statistical methods and have given evidence that doing the statistics recognizing the manifold curvature frequently provides more effective analysis. We have also showed improvements when the object representation captures aspects of boundary normal directions and object widths. The study of how to do statistics on curved manifolds is an active area of research, and we anticipate numerous advances in the next few years.

Acknowledgments

We are grateful for contributions to this paper by Hyo-Young Choi, Benjamin Eltzner, Thomas Fletcher, Junpyo Hong, Thomas Hotz, Sungkyu Jung, Xavier Pennec, Jörn Schulz, Stefan Sommer, Anuj Srivastava, Liyun Tu, Stephan Huckemann, and Jared Vicory.

References:

Beg, M. F., Miller, M. I., Trouvé, A., & Younes, L. (2005). Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International journal of computer vision*, *61*(2), 139-157.

Bhattacharya, R., & Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *Annals of statistics*, 1-29.

Bhattacharya, R., & Patrangenaru, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds: II. *Annals of statistics*, 1225-1259.

Blum, H. (1973). Biological shape and visual science (Part I). *Journal of theoretical Biology*, *38*(2), 205-287.

Cates, J., P. T. Fletcher, M. Styner, M. Shenton, and R. Whitaker, "Shape modeling and analysis with entropy-based particle systems," in *Proc. Inf. Proc. Med. Imag.*, 2007, vol. 20, pp. 333-345.

Cooper, D.H. T.F. Cootes, C.J. Taylor and J. Graham (1995), "Active shape models—their training and application", Computer Vision and Image Understanding (61): 38–59

T.F.Cootes, C.J.Taylor (1993) Active Shape Model Search using Local Grey-Level Models: A Quantitative Evaluation, in Proc. British Machine Vision Conference, (Ed. J.Illingworth), BMVA Press, 639-648.

Damon, J., & Marron, J. S. (2014). Backwards principal component analysis and principal nested relations. *Journal of Mathematical Imaging and Vision*, *50*(1-2), 107-114.

Davies, RH, C.J.Twining, P.D.Allen, T.F.Cootes and C.J.Taylor (2003). Building optimal 2D Statistical Shape Models", *Image and Vision Computing*, Vol.21,.117-82

Duda, R. O., Hart, P. E., & Stork, D. G. (2012). *Pattern classification*. John Wiley & Sons.

Eltzner, B., Jung, S., & Huckemann, S. (2015, October). Dimension Reduction on Polyspheres with Application to Skeletal Representations. In *International Conference on Networked Geometric Science of Information* (pp. 22-29). Springer International Publishing.

Fisher, N. I. (1995). *Statistical analysis of circular data*. Cambridge University Press.

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

Fletcher, P.T. (2013) Geodesic regression and the theory of least squares on Riemannian manifolds, IJCV, 105 (2), pp. 171-185.

Fréchet, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. In *Annales de l'institut Henri Poincaré* (Vol. 10, No. 4, pp. 215-310).

Gorczowski, K., Styner, M., Jeong, J. Y., Marron, J. S., Piven, J., Hazlett, H. C., ... & Gerig, G. (2007, June). Statistical shape analysis of multi-object complexes. In *2007 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1-8). IEEE.

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, *53*(3-4), 325-338.

Hall, P., Marron, J. S., & Neeman, A. (2005). Geometric representation of high dimension, low sample size data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(3), 427-444.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*(1), 29-36.

Hong, J., Vicory, J., Schulz, J., Styner, M., Marron, J., & Pizer, S. M. (2016). Classification of medically imaged objects via s-rep statistics. *Med. Image Anal.*, 31, 37–45.

Hotz, T. (2013), Extrinsic vs Intrinsic Means on the Circle, *Geometric Science of Information*, vol. 8085, Nielsen, F. and Barbaresco, F. eds. Lecture Notes in Computer Science, Springer-Verlag,433-440.

Huckemann, S., Hotz, T., Munk, A. (2010), Intrinsic shape analysis: Geodesic principal component analysis for Riemannian manifolds modulo Lie group actions. Discussion paper with rejoinder. *Statistica Sinica*, **20**, 1-100.

Jolliffe, I. (2002). *Principal component analysis*. John Wiley & Sons, Ltd.

Jung, S., Dryden, I. L., & Marron, J. S. (2012). Analysis of principal nested spheres. *Biometrika*, *99*(3), 551-568.

Kazhdan, M., Funkhouser, T., & Rusinkiewicz, S. (2003, June). Rotation invariant spherical harmonic representation of 3 d shape descriptors. In *Symposium on geometry processing* (Vol. 6, pp. 156-164).

Klassen, E., Srivastava, A., Mio, M., & Joshi, S. H. (2004). Analysis of planar shapes using geodesic paths on shape spaces. *IEEE transactions on pattern analysis and machine intelligence, 26*(3), 372-383.

Koenderink, J. J. (1990). *Solid Shape*, MIT Press.

Kurtek, S., Srivastava, A., Klassen, E., & Laga, H. (2013, May). Landmark-Guided Elastic Shape Analysis of Spherically-Parameterized Surfaces. In *Computer graphics forum* (Vol. 32, No. 2pt4, pp. 429-438). Blackwell Publishing Ltd.

Marčenko, V. A., & Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, *1*(4), 457.

Marron, J. S., Todd, M. J., & Ahn, J. (2007). Distance-weighted discrimination. *Journal of the American Statistical Association*, *102*(480), 1267-1271.

Mardia, K. V. (2014). *Statistics of directional data*. Academic Press.

Patrangenaru, V., & Ellingson, L. (2015). Nonparametric Statistics on Manifolds and Their Applications to Object Data Analysis.

Pennec, Xavier. "Barycentric Subspace Analysis on Manifolds." *arXiv preprint arXiv:1607.02833* (2016).

Pizer, S. M., Fritsch, D. S., Yushkevich, P. A., Johnson, V. E., & Chaney, E. L. (1999). Segmentation, registration, and measurement of shape variation via image object shape. *IEEE transactions on medical imaging*, *18*(10), 851-865.

Pizer, S. M., Joshi, S., Fletcher, P. T., Styner, M., Tracton, G., & Chen, J. Z. (2001, October). Segmentation of single-figure objects by deformable M-reps. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 862-871). Springer Berlin Heidelberg.

Pizer, S.M., Jung, S., Goswami, D., Vicory, J., Zhao, X., Chaudhuri, R., Damon, J.N., Huckemann, S. and Marron, J.S. (2013). Nested sphere statistics of skeletal models. In *Innovations for Shape Analysis* (pp. 93-115). Springer Berlin Heidelberg.

Rao, M., Stough, J., Chi, Y. Y., Muller, K., Tracton, G., Pizer, S. M., & Chaney, E. L. (2005). Comparison of human and automatic segmentations of kidneys from CT images. *International Journal of Radiation Oncology* Biology* Physics*, *61*(3), 954-960.

Schölkopf, B., & Smola, A. J. (2002). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press.

Schulz, J., Jung, S., Huckemann, S., Pierrynowski, M., Marron, J. S., & Pizer, S. M. (2015). Analysis of rotational deformations from directional data. *Journal of Computational and Graphical Statistics*, *24*(2), 539-560.

Schulz, J., Pizer, S. M., Marron, J. S., & Godtliebsen, F. (2016). Non-linear Hypothesis Testing of Geometric Object Properties of Shapes Applied to Hippocampi. *Journal of Mathematical Imaging and Vision*, *54*(1), 15-34.

Sen, S. K., Foskey, M., Marron, J. S., & Styner, M. A. (2008, May). Support vector machine for data on manifolds: An application to image analysis. In *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on* (pp. 1195-1198). IEEE.

Siddiqi, K., & Pizer, S. (Eds.). (2008). *Medial representations: mathematics, algorithms and applications* (Vol. 37). Springer Science & Business Media.

Sommer, S. (2015). Anisotropic Distributions on Manifolds: Template Estimation and Most Probable Paths, Lecture Notes in Computer Science, vol. 9123, S. Ourselin et al., eds., *Information Processing in Medical Imaging*, Springer, 193-204.

Srivastava, A., Klassen, E., Joshi, S. H., & Jermyn, I. H. (2011). Shape analysis of elastic curves in euclidean spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(7), 1415-1428.

Torgerson, W. S. (1952). Multidimensional scaling: I. Theory and method. *Psychometrika*, *17*(4), 401-419.

Tracy, C. A., & Widom, H. (1994). Level-spacing distributions and the Airy kernel. *Communications in Mathematical Physics*, *159*(1), 151-174.

Tu, L, D,. Yang, J. Vicory, X. Zhang, S.M. Pizer, M. Styner (2015). Fitting skeletal object models using spherical harmonics based template warping. *Signal Processing Letters*, vo. 22, no. 12, IEEE, 2269-2273.

Tu, L., Styner, M., Vicory, J., Paniagua, B., Prieto, J. C., Yang, D., & Pizer, S. M. (2015, March). Skeletal shape correspondence via entropy minimization. In *SPIE Medical Imaging* (pp. 94130U-94130U). International Society for Optics and Photonics.

Vicory, J., Foskey, M., Fenster, A., Ward, A., & Pizer, S. M. (2014, June). Prostate Segmentation From 3DUS Using Regional Texture Classification and Shape Differences. In *Symposium on Statistical Shape Models & Applications* (p. 24).

Vicory, J. (2016). *Shape Deformation Statistics and Regional Texture-Based Appearance Models for Segmentation.* PhD dissertation, University of North Carolina at Chapel Hill (on the internet at http://midag.cs.unc.edu: PhD dissertations button).

Yao, J., Bai, Z., & Zheng, S. (2015). *Large Sample Covariance Matrices and High-Dimensional Data Analysis* (No. 39). Cambridge University Press.