Generalized PCA via the backward stepwise approach in image analysis

Sungkyu Jung, Xiaoxiao Liu, J. S. Marron and Stephen M. Pizer

Abstract Principal component analysis (PCA) for various types of image data is analyzed in terms of the forward and backward stepwise viewpoints. In the traditional forward view, PCA and approximating subspaces are constructed from lower dimension to higher dimension. The backward approach builds PCA in the reverse order from higher dimension to lower dimension. We see that for manifold data the backward view gives much more natural and accessible generalizations of PCA. As a backward stepwise approach, composite Principal Nested Spheres, which generalizes PCA, is proposed. In an example describing the motion of the lung based on CT images, we show that composite Principal Nested Spheres captures landmark data more succinctly than forward PCA methods.

1 Introduction

Principal component analysis (PCA) is a widely used data exploration method in a variety of fields, for many purposes including dimension reduction and visualization of important data structures. In image analysis, the dimensionality of objects under investigation is usually very high, so dimension reduction through PCA is essential in some analysis; see for example, [18].

Sungkyu Jung
University of North Carolina at Chapel Hill, e-mail: sungkyu@email.unc.edu

Xiaoxiao Liu
University of North Carolina at Chapel Hill, e-mail: sharonxx@cs.unc.edu

J. S. Marron
University of North Carolina at Chapel Hill, e-mail: marron@email.unc.edu

Stephen M. Pizer
University of North Carolina at Chapel Hill, e-mail: pizer@cs.unc.edu

The classical PCA is based on the Euclidean properties of vector space, especially inner products and orthogonality. PCA is easily applicable for the many data types with these properties, an example of which is Functional PCA ([19, 20]), where the data set consists of smooth curves and the goal is to understand the variation in a set of curves. By a basis expansion of curves, the Euclidean properties are still well-defined, and the Functional PCA is a complete analog of the classical PCA.

Two useful viewpoints on PCA are the forward and backward stepwise approaches. In the traditional forward view, PCA is constructed from lower dimension to higher dimension. In the backward point of view, PCA is constructed in reverse order from higher to lower dimensions. These two approaches are equivalent in Euclidean space but lead to different methodologies in non-Euclidean data discussed next.

A growing number of data types are non-Euclidean, so the classical PCA idea does not apply. This paper focuses on the *mildly non-Euclidean* data, which are also referred to as manifold data, as in that context, the data objects are on the surface of a curved manifold forming a feature space. Data on curved manifolds have long been investigated. Among those the following are best studied:

Directional data Angles or directions lie on the unit circle or the unit sphere (or a hemisphere), which include wind or ocean current directions, orientation of cracks on rocks, and directions from the earth to celestial objects. A substantial amount of literature can be found in the area of circular, angular or directional statistics, see [3], [14].

Statistical shape space Landmark-based shape analysis analyzes data lying on special manifolds. A shape is defined as an equivalence class under translation and rotation, scaling in many cases and sometimes reflection. Thus, shape spaces are constructed by removing the translation, scale, and rotation from the set of landmarks, as proposed and investigated by both Kendall [12] and Bookstein [1] and described well in [2].

Due to advances in technology, a demand to analyze different types of manifold data is growing. These modern data are mostly from medical imaging and include

Medial shape representations Shapes of 2-*d* or 3-*d* objects are represented in a parametric model, called *m-reps* in short, including directions and angles as parameters. The data space here is a manifold that is a direct product of Euclidean space and unit spheres. See [21].

Diffusion Tensor Imaging DTI [17] is a recently developed and widely studied MRI technique that measures the diffusion of water molecules in a biological object. Random motion of water molecules in each voxel of an image is represented by a 3-*d* tensor, i.e. a non-negative definite 3×3 matrix. Each tensor lies in a lower dimensional sub-manifold of $\mathbf{R}^9$ since it has to be non-negative definite. DTI data, consisting of multiple tensors, thus naturally lie in a manifold.

Diffeomorphisms A common methodology for comparing shapes in image analysis is to use diffeomorphisms ([9], [8]), i.e. smooth space warping functions. This method delivers a new approach to shape analysis. A shape is considered

as a distortion (i.e. diffeomorphism) of some template. Thus a set of shapes is represented as a set of diffeomorphisms and the variation in the population of diffeomorphisms can be studied to understand variation in shapes. The set of diffeomorphisms forms a very high dimensional manifold.

Conventional statistical analysis, including PCA, is not directly applicable to these manifold data. On the other hand, there is a growing need of PCA-like methods, because the dimensionality of the data space is often very high. Previous approaches for generalized PCA to manifold data are listed and discussed in Section 2. Many commonly used methods can be viewed as the forward approach. However, [15] also points out that the backward viewpoint is seen to provide much more natural and accessible analogues of PCA than the standard view. This is discussed in Section 2.2.

Section 3 is devoted to proposing a methodology of generalized PCA to the surface point distribution model (PDM). The method, *composite PNS*, can be viewed as an extension of Principal Nested Spheres, proposed by [10] and discussed in Section 3.1, which also can be viewed as a backward generalization of PCA to manifold data. The procedure of the proposed method is illustrated in Section 3.2.

Advantages of the proposed method are presented by some experimental results in Section 3.3. We use this approach to describe the motion of the lung using landmark data extracted from CT images. We show that composite Principal Nested Spheres captures more variation of this landmark data in fewer dimensions than the standard PCA.

2 Forward and backward stepwise view of PCA

The forward and backward stepwise views of PCA, either in Euclidean space or for manifolds, are discussed in this section.

2.1 Mathematical development for Euclidean PCA

Let $X_1,\ldots,X_n$ be d-dimensional column vectors that are inputs for Euclidean PCA. The data matrix is formed by aggregating the data vectors: $\mathbf{X} = [X_1,\ldots,X_n]$. A *forward stepwise* view to Euclidean PCA is understood by increasing the dimension of the best approximating (affine) subspace, as described in the following steps:

1. Find a center point that best represents the data, by minimizing the sum of squared distances: the empirical mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.
2. Find a line, an affine 1-d subspace, that best approximates the data, again by minimizing the sum of squared distances from the data to the line. Pythagorean theorem shows that this line must pass through the sample mean $\bar{X}$. This affine one dimensional subspace can be written with a direction vector u_1 so that

$$AS_1^1 = \{\bar{X} + cu_1 : c \in \mathbf{R}\}.$$

The direction u_1 is the first principal component (PC) direction. The orthogonal projections of the data X_i onto AS_1^1 are then in the form $\bar{X} + c_i u_1$, which are the best rank 1 approximation of the data. The amount of deviation c_i from the center is called PC scores.

3. Next find a line in the affine subspace orthogonal to u_1, that best represents the data. The line is denoted with the second PC direction vector u_2 by $AS_2^1 = \{\bar{X} + cu_2 : c \in \mathbf{R}\}$. Since u_1 and u_2 are orthogonal, the best two dimensional approximation of the data is contained in the affine 2-d subspace

$$AS^2 = AS_1^1 \otimes AS_2^1 = \{\bar{X} + c_1 u_1 + c_2 u_2 : c_1, c_2 \in \mathbf{R}\},$$

where $\otimes$ represents the direct product. PC scores for the second PC are found again through the projections of the data onto AS_2^1 (or AS^2), similar to the 1-d case.

4. Higher order components can be found iteratively for $k = 3, 4, \ldots, d$, that results in k-dimensional affine subspaces

$$AS^k = \otimes_{j=1}^k AS_j^1 = \{\bar{X} + \textstyle\sum_{j=1}^k c_j u_j : c_1, \ldots, c_k \in \mathbf{R}\}.$$

In this forward formulation of PCA the best approximating affine subspaces are constructed from the lowest dimension to higher dimension, i.e.

$$\{\bar{X}\} \subset AS_1^1 \subset AS^2 \subset \cdots \subset AS^d.$$

This formulation is most useful for heuristic understanding of the method. A practical formulation uses the fact that the PC direction vectors u_j are eigenvectors of the sample covariance matrix $S = \frac{1}{n-1}(\mathbf{X} - \bar{X})(\mathbf{X} - \bar{X})^T$ or the left singular vectors of the centered data matrix $(\mathbf{X} - \bar{X})$.

The viewpoint that seems most useful for generalization of PCA to manifold data is the *backward stepwise* view. In backward PCA, principal components are found in reverse order, i.e. AS_ks are fitted from the largest dimension, which leads to

$$\mathbf{R}^d = AS^d \supset AS^{d-1} \supset \cdots \supset AS_1^1 \supset \{\bar{X}\}.$$

In particular, AS^{d-1} is found from AS^d by removing the PC direction $u_d \in AS^d$, which is orthogonal to AS^{d-1}. Deriving AS^{d-1} from AS^d is equivalent to the finding of the $(d-1)$-dimensional linear subspace of AS^d that minimizes the sum of squared distances. The projections X_i^P of X_i to AS^{d-1} are then the best $(d-1)$-dimensional approximation of the data, and the signed length of projections (from X_is) become the PC scores. An application of the Pythagorean theorem yields that AS^{d-2} can be found in the same manner from the projections X_i^P.

In Euclidean space, the forward and backward approaches are equivalent. However, in non-Euclidean spaces, this is usually not the case, and the choice of viewpoint affects the generalizations of PCA.

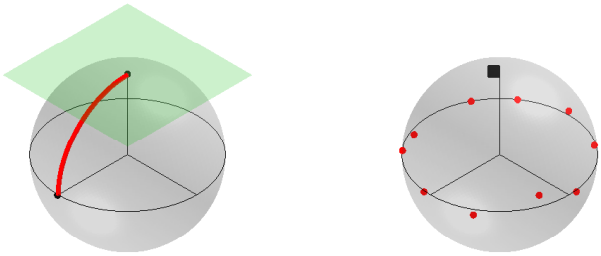

Fig. 1 (Left) The usual unit sphere S^2 with a geodesic segment (great circle segment) joining the north pole and a point in the equator. The tangent plane at the north pole is also depicted. (Right) Plot of 10 points along the equator with random perturbation and the geodesic mean (black square) near the north pole illustrates the case where the geodesic means on S^2 does not represent the data well.

2.2 PCA approaches for manifold data

A widely used approach to manifold PCA, called Principal Geodesic Analysis (PGA, [5]), generalizes PCA in a forward stepwise manner. The first step in PGA is to find a center point for the manifold data. While the sample mean (i.e. the average) is not defined, a useful notion for generalization of mean is the Fréchet mean, defined as a minimizer of $\min_{x\in\mathscr{M}} \sum_{i=1}^{n} \rho^2(x,x_i)$, where $\mathscr{M}$ is the data space and ρ is a metric defined on $\mathscr{M}$. The Fréchet mean is widely applicable, since it only requires a metric on the manifold. In Euclidean space, the sample mean is the Fréchet mean with the usual metric $\rho(x,y) = \|x-y\|$. On curved manifolds, distances are commonly measured along geodesics. A geodesic is an analog of straight lines in Euclidean space; it is roughly defined as the shortest path between two points (see Fig. 1). The geodesic distance function measures the shortest arc length between two points. With the geodesic distance as its metric, the Fréchet mean is often called geodesic mean.

Having the geodesic mean as the center point in PGA, the second step is to find a geodesic (instead of a line) that best represents the data, among all geodesics that pass through the geodesic mean. The higher order components are again geodesics that are orthogonal (in a sense) to the lower order geodesics. In practice, these geodesic components are computed through the projection of the data onto the tangent space at the geodesic mean. The PGA and similarly defined forward approaches are developed for various types of data; see e.g. [5] for m-reps data, [4] for DTI data, and [2] for shape data.

However, there has been a concern that the geodesic mean and tangent space approximation can be very poor. As a simple example, consider the usual unit sphere S^2 in $\mathbf{R}^3$ and the data distributed uniformly along the equator of the sphere as illustrated in Fig. 1. In this case, the equator itself is the geodesic that best represents the data. However, the geodesic mean is located at near the north or the south pole, far

from any data. PGA finds principal geodesics through this geodesic mean, which fail to effectively describe the variation in the data.

This observation motivated [7] to propose Geodesic PCA (GPCA). In GPCA, the geodesic mean or any pre-determined mean is no longer used; instead it finds the best approximating geodesic among all possible candidates. A center point of the data is then found in the first geodesic component. In the equator example above, GPCA finds the equator as the first component. GPCA can be viewed as a backward approach, particularly when applied to S^2, since the center point is found last. In higher dimensional manifolds, for example in hyperspheres $S^p, p > 2$ and Kendall's shape spaces (see [6]), GPCA does not appear to be fully backward, since the method is built by considering lower dimensional components first, only with an exception for center point. Nevertheless, the advantage of the method indeed comes from the backward viewpoint, i.e. from reversing the order of the first two steps.

Another method that can be viewed as the backward stepwise approach is Principal Arc Analysis (PAA), proposed by [11], which is a non-geodesic generalization of PCA. PAA is motivated by data distributed along a *small circle* on S^2. Since the major variation is no longer along a geodesic, no geodesic based methods including PGA and GPCA capture the variation effectively. PAA begins with the full sphere S^2 and finds the small circle as the best fitting 1-d approximation of the data, followed by a center point contained in the small circle. PAA was shown to provide just this type of effective data approximation in S^2 and also in m-reps data in [11].

In generalizations of PCA for higher dimensional manifolds, including hyperspheres S^p and Kendall's shape spaces, the backward stepwise principle led to a new fully backward generalization of PCA: Principal Nested Spheres (PNS), proposed by [10]. In taking the backward approach, it inherits the advantages of GPCA. In using non-geodesic approximation, it inherits advantages of PAA. A detailed description of the method can be found in Section 3.1. PNS has been shown to provide more representative description of the data (compared to other forward stepwise approaches) in a number of standard examples in [10]. A discussion of application of PNS to Euclidean data, in Section 3, shows how beneficial a backward generalization of PCA could be even for Euclidean data.

3 Method

In this section, a method for Euclidean data that possesses the advantage of backward generalization of PCA is discussed. In particular, when the dataset is a set of the surface point distribution models (PDM) representing the shape of an object, the backward generalization of PCA to shape space, Principal Nested Spheres (PNS), fits well. We summarize PNS in more detail and discuss the composite PNS, followed by experimental results which shows that the composite PNS gives more effective description of the PDMs in lower dimension than Euclidean PCA.

3.1 Principal Nested Spheres

The analysis of PNS is summarized in this section. PNS is essentially a decomposition method for hyperspheres and Kendall's shape space, which generalizes PCA in a non-geodesic way. Detailed geometric properties and statistical discussions of PNS can be found at [10]. As mentioned in Section 2.2, the first step in PNS is to reduce the dimension d of S^d to $d-1$. Specifically, we wish to find the best approximating sub-manifold of dimension $d-1$. PNS solves this problem with a flexible class of sub-manifolds in the form of nested spheres.

A k-dimensional nested sphere A_k of S^d is nested within (i.e. sub-manifold of) higher dimensional nested spheres; and A_k itself can be thought of as a k-dimensional sphere. As an example, A_{d-1} of S^d is defined with an axis $v_1 \in S^d$ and distance $r_1 \in (0, \pi/2]$ as follows,

$$A_{d-1}(v_1, r_1) = \{x \in S^d : \rho_d(v_1, x) = r_1\},$$

where ρ_d is the geodesic distance function defined on S^d. The parameter v_1 drives the 'direction' that is not contained in A_{d-1}. In relation to the backward view of Euclidean PCA in Section 2.1, the direction coincides to u_d, which is orthogonal to AS^{d-1}. The distance from v_1 to any point in A_{d-1} is r_1, which is responsible for the curvature of A_{d-1}. This flexibility of curvature in A_{d-1} allows PNS to capture a certain form of non-geodesic variation.

Lower dimensional nested spheres are defined similarly. Since A_{d-1} is essentially a sphere, A_{d-2} can be defined again with a pair (v_2, r_2) and in a way that A_{d-2} is nested within A_{d-1}. Iteratively, one can continue to build a sequence of nested spheres $S^d \supset A_{d-1} \supset \cdots \supset A_1$.

In PNS with a data set $X_1, \ldots, X_n \in S^d$, the pair (v, r) of nested spheres is fitted to the data iteratively so that the fitted nested spheres represent the data. [10] proposed minimizing the sum of squared distances to the data, i.e. the $d-1$ dimensional PNS is

$$\hat{A}_{d-1} = \operatorname{argmin} \sum_{i=1}^{n} \rho_d(A_{d-1}, X_i)^2, \tag{1}$$

where $\rho_d(A_{d-1}, X_i)$ is defined as follows. Each X_i can be projected on A_{d-1} along the minimal geodesic that joins X_i to A_{d-1}. Denote X_i^P for the projection. The length of the minimal geodesic is the distance, that is $\rho_d(A_{d-1}, X_i) = \rho_d(X_i^P, X_i)$. Note that each observation gets a signed residual $z_{d,i}$.

The second (or the $d-2$ dimensional) PNS is found with the projections X_i^P. Since X_i^P's are on $\hat{A}_{d-1}$, one can use the method (1) by treating $\hat{A}_{d-1}$ and $\{X_i^P\}$ as S^{d-1} and $\{X_i\}$, respectively. Simply put, $\hat{A}_{d-2}$ is fitted to X_i^P's by minimizing the sum of squared distances. In general, we recursively find the sequence of PNS from the (iteratively) projected data.

The lowest level principal nested sphere $\hat{A}_1$ is then a small circle, with intrinsic dimension 1. The Fréchet mean of $X_1^P, \ldots, X_n^P \in \hat{A}_1$ is used as the best 0-dimensional

representation of the data in the framework of PNS. Denote the Fréchet mean as $\hat{A}_0$, and keep the signed deviations $z_{1,i}$ of X_i^P for later use.

As a result, PNS constructs the sequence of the best approximating sub-manifolds

$$S^d \supset \hat{A}_{d-1} \supset \cdots \supset \hat{A}_1 \supset \{\hat{A}_0\},$$

for every dimension. The backward principle is essential to PNS, since the forward stepwise generalizations of PCA are not be equivalent to PNS (see Section 2.2) and are even not clearly defined for non-geodesic variation.

Furthermore, we wish to represent the data in an Euclidean space for further analysis (e.g. the method of composite PNS, discussed later in Section 3.2). Recall that in the procedure above, we have collected the signed residuals $z_{k,i}$. The *Euclidean-type representation* of the data by PNS is obtained by combining those residuals into a $p \times n$ data matrix

$$\mathscr{Z} = \begin{pmatrix} z_{1,1} & \cdots & z_{1,n} \\ \vdots & \ddots & \vdots \\ z_{d,1} & \cdots & z_{d,n} \end{pmatrix},$$

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

The procedure is computationally fast in a moderate size of dimension and samples, when using the computational algorithm proposed in [10] for the optimization task (1). However in the high dimension low sample size situation where for example $p > 1000$ and $n < 100$, strict application of the iterative procedure results in a very slow computation. [10] have shown that the intrinsic dimensionality of the data can be reduced to $n-1$ without losing any information and that the first $d-n$ PNS can be found trivially by an application of singular value decomposition. This fact is used when it applies, including the experiments in Section 3.3.

3.2 Application of PNS to scaled point distribution models

The surface point distribution model (PDM) representing the shape of a human organ (or other solid object) is denoted by $X_i = [x_1(i), \ldots, x_p(i)]^T$ where $x_j(i) = (x_{ij}, y_{ij}, z_{ij})$ is the jth point on the surface and p is the number of points on the surface. The subscript i denotes the ith sample or time point, and let n be the total number of time points. A scaled PDM (SPDM) is derived from a PDM by moving each point towards some designated center point by some fixed factor such that the sum of squared distances from the center point is unity. Thus an SPDM is a PDM that lies on a unit hypersphere, which reflects the shape. The PCA-like analysis of such data should reflect not only variability on the hypersphere but also the correlation between the scale factor, which reflects the size, and the shape.

Wanting to apply a backward generalization of PCA, we might think to use PNS, but it applies only to the variability on the hypersphere. In the composite PNS we

propose, the variables are separated into the size and the shape variables. The dimension of the shape space is reduced by PNS. Then the size variable is post-combined with the result of PNS, to incorporate the correlation between size and shape.

A procedure for the composite PNS is as follows:

1. (Centering) Let $\widetilde{X}_i = X_i - \frac{1}{np}\sum_{ij} x_j(i)$ be the ith uniformly translated PDM, so that $\frac{1}{np}\sum_{ij}\widetilde{x}_j(i) = (0,0,0)$.
2. (Scaling) Let $S_i = (\sum_{j=1}^{p} \left\|\widetilde{x}_j(i)\right\|^2)^{\frac{1}{2}}$ be the size of the ith PDM, measured by the sum of squared distances to the center. The scaled PDM is $\widetilde{X}_i^* = \widetilde{X}_i/S_i$, so that the size of $\widetilde{X}_i^*$ is 1 for all i. Then the pair $(\widetilde{X}_i^*, S_i)$ represents the shape and size of X_i, respectively.
3. (Shape analysis by PNS) Find principal nested spheres and PNSmean, as described in the previous subsection with inputs $\{\widetilde{X}_i^*\}$, and denote the resulting Euclidean-type representation as an $m \times n$ matrix $\mathscr{Z} = (z_{ki})$, where z_{ki} is the ith sample's deviation from the PNSmean along the kth principal arc, and $m \leq n-1$ is the number of components, which may be chosen by practitioners.
4. (Size analysis in log scale) Since the size S_i is strictly positive, it makes most sense to compare variability in a log scale. Let $\bar{S}_n = (\prod_{i=1}^{n} S_i)^{\frac{1}{n}}$ be the geometric mean of the size, which is the exponential of the arithmetic mean of $\log(S_i)$. Define $S_i^* = \log(S_i/\bar{S}_n)$.
5. (Composite space for shape-and-size of PDM) In order to incorporate the correlation between the size variables S_i^* and the shape variables $\mathscr{Z}$, define a composite data matrix
$$\mathscr{Z}_s = \begin{pmatrix} \mathscr{S} \\ \mathscr{Z} \end{pmatrix},$$
where $\mathscr{S} = (S_1^*, \ldots, S_n^*)$ and each column contains the size (S_i) and shape $(z_{1i}, \ldots, z_{mi})$ information of each sample.
6. (Principal arcs and scores) Let the spectral decomposition of the $(m+1)$-dimensional square matrix $\frac{1}{n-1}\mathscr{Z}_s\mathscr{Z}_s^T$ be $U\Lambda U^T$, where $U = [u_1, \ldots, u_{m+1}]$ is the orthogonal matrix of eigenvectors and Λ is the diagonal matrix of eigenvalues $\lambda_1, \ldots, \lambda_{m+1}$. Similar to the conventional PCA, the eigenvectors u_k represent the direction of important variation in the space of $\mathscr{Z}_s$ which leads to the *principal arc* when converted back to the original space of PDMs. Likewise, the eigenvalues λ_k represent the variation contained in each component. *Principal Arc scores* for each component are computed by $u_k^T\mathscr{Z}_s$, which is the vector of the kth scores of all n samples.

The analysis of composite PNS can be used in a same fashion as Euclidean PCA is used. Both provides a nested sequence of subspaces (or sub-manifolds) for dimension reduction, and PC scores (or PA scores) that are important for visualization of important data structure, and for further analysis such as PC regression.

The advantage of composite PNS comes from the flexible class of sub-manifolds instead of subspaces. As shown in Section 3.3, the proposed method gives more effective decomposition of the space compared to Euclidean PCA and PGA.

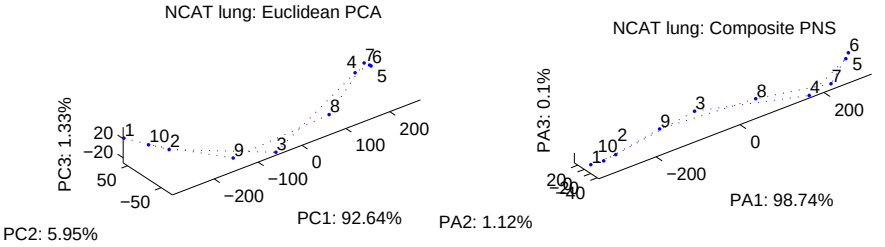

Fig. 2 (Left) Scatterplot of NCAT lung data by PC scores in the first three components of Euclidean PCA. Time points are labeled as 0-9 in the scatterplot and the proportion of variance contained in each component appears in the labels of axes. Major variation in the data is non-linear. (Right) Scatterplot of the NCAT lung data by PA scores of composite PNS. The non-linear variation is captured in the first principal arc, and thus the variation appears linear. The first component in composite PNS contains more variation (98.74% of the total variation) than 92.64% of PCA.

3.3 Experimental results

Respiratory motion analysis in the lung is important for understanding the motilities of tumors and various organs in the lung of an individual patient for radiation therapy applications. The PDM of the lung boundary is used as the surrogate signal for characterizing the respiratory motion [13]. The usual PCA has been used for extracting shape deformation statistics of a patient's lungs from a sequence of 3-d CT images collected at ten time points within one breathing cycle. In preparation for PCA (also for composite PNS) on this data set, the geometric correspondence over the the training samples is optimized via an entropy-based correspondence algorithm [16].

We consider two examples. The first data set is from 4D Nurbs-based Cardiac-Torso (NCAT) phantom thorax CTs, which were produced at ten phases sampled in one breathing cycle. The second data set is from Respiration-correlated CT of a real patient. The CT data sets are provided by a 4-slice scanner (lightSpeed GX/i, GE Medical System), acquiring repeat CT images for a complete respiratory cycle at each couch position while recording patient respiration (Real-time Position Management System, Varian Medical Systems). The CT images are retrospectively sorted (GE Advantage 4D) to produce a series of 3D images at different respiratory time points.

The difficulty of the problem is two-fold; the dimension is very high ($d = 10550$, which could be much higher depending on the number of points on the surface) while the sample size is small ($n = 10$) and the major variation is non-linear, as shown in Fig. 2 for the NCAT data sets.

Fig. 2 shows scatter plots of NCAT lung data by the usual PCA (in the left panel) and by composite PNS (in the right panel). The dimension of the data space is reduced to 3 to visualize the structure of major variation. The non-linear variation apparent in the PCA subspace is represented as a linear motion in the sub-manifold of composite PNS. In particular, the quadratic motion in the PC 1–2 plane is efficiently

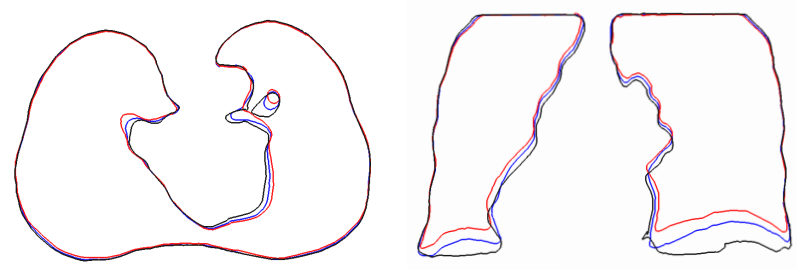

Fig. 3 Axial view (left) and coronal view (right) of boundaries of lungs. Illustrated is the variation of shapes captured in the first principal arc.

captured by the 1-dimensional principal arc. Observe that the sum of variances contained PC 1–2 is roughly the amount of variation in the first principal arc.

The data set from the real patient gives a similar result, where the cumulative proportions of variances in the first three sub-manifolds (96.38%, 97.79%, and 98.63%, respectively) are higher than those of PCA (93.52%, 96.25% and 97.74%).

The major lung motions contained in the first principal arc is illustrated in Fig. 3. We show the coronal and axial slices of lungs corresponding to the PNSmean and ± 1.5 standard deviations.

We also measure the discrepancy between the PDM at each time point and its 1-*d* approximation by PCA or composite PNS. The discrepancy here is computed by the square root of sum of squared distances between corresponding points. In the patient lung data, the discrepancy of 1-*d* approximations by composite PNS is uniformly smaller than that by PCA, as summarized in Table 1.

Table 1 Discrepancy of 1-*d* approximations at each time point of the real patient lung motion.

time point	1	2	3	4	5	6	7	8	9	10
PCA	65.2	69.9	88.7	77.7	38.9	74.4	44.1	69.8	74.6	57.6
composite PNS	38.2	66.9	66.1	55.6	37.8	36.7	30.4	63.0	60.2	44.6

4 Conclusion

The backward PCA approaches have proven useful for dimension reduction of nonlinear manifolds. In particular, PNS enjoys the advantages of the fully backward approach that enable it to yield more succinct description of the data, as shown in the example of size and SPDM shape changes with application to the lung motion. Image analysis benefits from taking attention to analysis of shapes, and thus statis-

tical analysis in that domain might be beneficial. Particularly, the idea of PNS can be generalized to a variety of applications over both computer vision and medical imaging.

References

1. Bookstein, F.L.: Morphometric tools for landmark data. Cambridge University Press, Cambridge (1991). Geometry and biology, Reprint of the 1991 original
2. Dryden, I.L., Mardia, K.V.: Statistical shape analysis. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester (1998)
3. Fisher, N.I.: Statistical analysis of circular data. Cambridge University Press, Cambridge, UK (1993)
4. Fletcher, P.T., Joshi, S.: Riemannian geometry for the statistical analysis of diffusion tensor data. Signal Processing **87**(2), 250–262 (2007)
5. Fletcher, P.T., Lu, C., Pizer, S.M., Joshi, S.: Principal geodesic analysis for the study of nonlinear statistics of shape. IEEE Trans. Medical Imaging **23**, 995–1005 (2004)
6. Huckemann, S., Hotz, T., Munk, A.: Intrinsic shape analysis: Geodesic PCA for Riemannian manifolds modulo isometric lie group actions. Statistica Sinica **20**(1), 1–58 (2010)
7. Huckemann, S., Ziezold, H.: Principal component analysis for Riemannian manifolds, with an application to triangular shape spaces. Adv. in Appl. Probab. **38**(2), 299–319 (2006)
8. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. Neuroimage **23**, S151–160 (2004)
9. Joshi, S.C., Miller, M.I.: Landmark matching via large deformation diffeomorphisms. IEEE Trans. Image Process. **9**(8), 1357–1370 (2000)
10. Jung, S., Dryden, I.L., Marron, J.S.: Analysis of Principal Nested Spheres. Submitted in Biometrika (2010)
11. Jung, S., Foskey, M., Marron, J.S.: Principal arc analysis on direct productmanifolds,. to appear in Annals of Applied Statistics (2010)
12. Kendall, D.G., Barden, D., Carne, T.K., Le, H.: Shape and shape theory. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester (1999)
13. Liu, X., Oguz, I., Pizer, S.M., Mageras, G.S.: Shape-correlated deformation statistics for respiratory motion prediction in 4D lung. SPIE Medical Imaging **7625** (2010)
14. Mardia, K.V., Jupp, P.E.: Directional statistics. John Wiley & Sons Ltd. (2000)
15. Marron, J.S., Jung, S., Dryden, I.L.: Speculation on the generality of the backward stepwise view of pca. In: Proceedings of MIR 2010: 11th ACM SIGMM International Conference on Multimedia Information Retrieval, Association for Computing Machinery, Inc., Danvers, MA, 227-230. (2010)
16. Oguz, I., Cates, J., Fletcher, T., Whitaker, R., Cool, D., Aylward, S., Styner, M.: Cortical correspondence using entropy-based particle systems and local features. In: Biomedical Imaging, ISBI 2008. 5th IEEE International Symposium, 1637-1640 (2008)
17. Pennec, X., Fillard, P., Ayache, N.: A riemannian framework for tensor computing. International Journal of Computer Vision **66**(1), 41–66 (2006)
18. Rajamani, K.T., Styner, M.A., Talib, H., Zheng, G., Nolte, L.P., Ballester, M.A.G.: Statistical deformable bone models for robust 3d surface extrapolation from sparse data. Medical Image Analysis **11**, 99–109 (2007)
19. Ramsay, J.O., Silverman, B.W.: Applied Functional Data Analysis: Methods and Case Studies. Springer, New York (2002)
20. Ramsay, J.O., Silverman, B.W.: Functional Data Analysis, 2nd ed. edn. Springer, New York (2005)
21. Siddiqi, K., Pizer, S.M.: Medial Representations: Mathematics, Algorithms and Applications. Springer (2008)