Principal Geodesic Analysis for the Study of Nonlinear Statistics of Shape

P. Thomas Fletcher*, *Student Member, IEEE*, Conglin Lu, Stephen M. Pizer, *Senior Member, IEEE*, and Sarang Joshi

***Abstract*—A primary goal of statistical shape analysis is to describe the variability of a population of geometric objects. A standard technique for computing such descriptions is principal component analysis. However, principal component analysis is limited in that it only works for data lying in a Euclidean vector space. While this is certainly sufficient for geometric models that are parameterized by a set of landmarks or a dense collection of boundary points, it does not handle more complex representations of shape. We have been developing representations of geometry based on the medial axis description or m-rep. While the medial representation provides a rich language for variability in terms of bending, twisting, and widening, the medial parameters are not elements of a Euclidean vector space. They are in fact elements of a nonlinear Riemannian symmetric space. In this paper, we develop the method of principal geodesic analysis, a generalization of principal component analysis to the manifold setting. We demonstrate its use in describing the variability of medially-defined anatomical objects. Results of applying this framework on a population of hippocampi in a schizophrenia study are presented.**

***Index Terms*—Deformable models, medial geometry, statistical shape analysis.**

I. Introduction

STATISTICAL shape analysis [1]–[3] is emerging as an important tool for understanding anatomical structures from medical images. Given a set of training images, the goal is to model the geometric variability of the anatomical structures within a class of images. Statistical models give an efficient parameterization of the geometric variability of anatomy. These models can provide shape constraints during image segmentation [4]. Statistical descriptions of shape are also useful in understanding the processes behind growth and disease [5].

Deformable model approaches represent the underlying geometry of the anatomy and then use a statistical analysis to describe the variability of that geometry. Several different geometric representations have been used to model anatomy. Bookstein [6] uses landmarks to capture the important geometric features. The active shape model (ASM) of Cootes and Taylor [4] represents an object's geometry as a dense collection of boundary points. Cootes *et al.* [7] have augmented their models to include the variability of the image information as well as shape. Kelemen *et al.* [8] use a spherical harmonic (SPHARM) decomposition of the object geometry.

In all of these approaches, the underlying geometry is parameterized as a Euclidean vector space. The training data is given as a set of vectors $x_1, \ldots, x_N$ in a vector space V. For ASMs, each vector is constructed by concatenation of the boundary points in an object. For SPHARMs, each vector is constructed as the concatenation of the coefficients of a SPHARM surface representation of the object. An average object vector is computed as the linear average of the training set

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i.$$

Principal component analysis (PCA) [9] is then used to find an efficient parameterization of the model variability. This is accomplished by computing an eigenanalysis of the sample covariance matrix

$$S = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)(x_i - \mu)^T.$$

If $v_k, k = 1, \ldots, d$ are the ordered eigenvectors of the quadratic form S with corresponding eigenvalues λ_k, then a new object within the realm of statistically feasible shapes is parameterized by

$$x = \mu + \sum_{k=1}^{d} \alpha_k v_k$$

where the $\alpha_k \in \mathbb{R}$ weight the modes of variation.

Shape is often defined as the geometry of objects that is invariant under global translation, rotation, and scaling. To ensure that the variability being computed is from shape changes only, an important preprocessing step of any shape analysis technique is to align the training objects to a common position, orientation, and scale. The common alignment technique used is Procrustes alignment [10], which seeks to minimize, with respect to global translation, rotation and scaling, the sum-of-squared distances between corresponding data points.

While most work on the statistical analysis of shape has focused on linear methods, there has been some work on statistical methods for nonlinear geometric data. Pennec [11] defines Gaussian distributions on a manifold as probability densities that minimize information. Related work includes the statistical analysis of directional data [12] and the study of shape spaces as complex projective spaces [13].

Manuscript received January 5, 2004; revised April 28, 2004. This work was supported by the National Institutes of Health (NIH) under Grant P01 CA47982 and Grant EB02779. This is an extended version of research first presented at the 2003 Information Processing in Medical Imaging Conference (IPMI) [18]. The Guest Editors responsible for coordinating the review of this paper and recommending its publication were C. J. Taylor and J. A. Noble. *Asterisk indicates corresponding author*.

*P. T. Fletcher is with the Medical Image Display and Analysis Group, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514 USA (e-mail: fletcher@cs.unc.edu).

C. Lu, S. M. Pizer, and S. Joshi are with the Medical Image Display and Analysis Group, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514 USA.

Digital Object Identifier 10.1109/TMI.2004.831793

Linear shape models treat shape changes as combinations of local translations. Shape changes can also be usefully considered as combinations of local translations, rotations, and magnifications. Following that point of view, in our previous work [14], [15] we have developed methodology based on medial descriptions called m-reps. The medial representation provides a powerful framework for describing shape variability in intuitive terms such as local thickness, bending, and widening. However, the medial parameters are not elements of a Euclidean space. Therefore, the standard linear techniques of shape analysis, namely linear averaging and PCA, do not apply. In this paper, we show that the medial parameters are in fact elements of a certain type of manifold known as a Riemannian symmetric space. We then show how the standard shape analysis techniques can be generalized to handle manifold data. First, we describe how averages can be computed on a manifold. Next, we develop a new method named *principal geodesic analysis* (PGA), a generalization of PCA, for describing the variability of data on a manifold.

In Section II, we review the necessary theory of symmetric spaces and m-reps, showing that m-rep models are parameterized by a symmetric space. Section III describes how mean values are computed on manifolds. PGA is developed in Section IV as a method for describing the variability of data on a manifold. The statistical methods are then applied to the study of medially-defined anatomical shapes in Section V.

II. Background Theory

A. M-Rep Overview

The medial representation used in this paper is based on the medial axis of Blum [16]. In this framework, a geometric object is represented as a set of connected continuous medial manifolds. For three-dimensional (3-D) objects these medial manifolds are formed by the centers of all spheres that are interior to the object and tangent to the object's boundary at two or more points. The medial description is defined by the centers of the inscribed spheres and by the associated vectors, called *spokes*, from the sphere centers to the two respective tangent points on the object boundary. Each continuous segment of the medial manifold represents a medial figure. In this paper, we focus on 3-D objects that can be represented by a single medial figure.

We sample the medial manifold over a spatially regular lattice. The elements of this lattice are called *medial atoms*. A medial atom (Fig. 1) is defined as a 4-tuple $\mathbf{m} = \{\mathbf{x}, r, \mathbf{n}_0, \mathbf{n}_1\}$, consisting of: $\mathbf{x} \in \mathbb{R}^3$, the center of the inscribed sphere; $r \in \mathbb{R}^+$, the local width defined as the common spoke length; $\mathbf{n}_0, \mathbf{n}_1 \in S^2$, the two unit spoke directions (here S^2 is the sphere in $\mathbb{R}^3$ with radius one). The medial atom implies two opposing boundary points, $\mathbf{y}_0, \mathbf{y}_1$, called *implied boundary points*, which are given by

$$\mathbf{y}_0 = \mathbf{x} + r\mathbf{n}_0, \quad \mathbf{y}_1 = \mathbf{x} + r\mathbf{n}_1. \tag{1}$$

The surface normals at the implied boundary points $\mathbf{y}_0, \mathbf{y}_1$ are given by $\mathbf{n}_0, \mathbf{n}_1$, respectively.

We point out that in our previous work [17], [18] we parameterized medial atoms with a position $x \in \mathbb{R}^3$, a radius $r \in \mathbb{R}^+$, an orthonormal frame $(\mathbf{b}, \mathbf{b}^\perp, \mathbf{n}) \in SO(3)$, and an object angle $\theta \in [0, \pi)$ (see Fig. 1). Here the vector $\mathbf{b}$ points in the direction of the spoke bisector, $\mathbf{n}$ is normal to the medial sheet, and $\mathbf{b}^\perp$ is chosen to complete the orthonormal frame. The object angle θ is the half-angle between the two spokes. This representation has the drawback that medial atoms may not have a unique representation. For example, consider a medial atom with object angle $\theta = (\pi/2)$, i.e., the spokes are aligned in opposing directions. In this case, the frame may be rotated arbitrarily about the vector $\mathbf{n}$ without changing the medial atom. The representation presented in this paper, replacing the frame and object angle with two spoke directions, does not suffer from such multiplicities. For generic atoms with $\theta \in (0, (\pi/2))$ there is a well-defined conversion between the two representations.

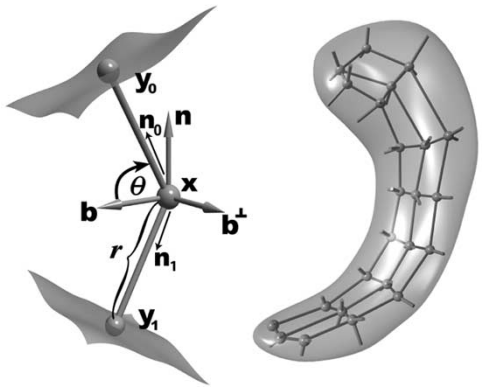

Fig. 1. Medial atom with a cross section of the boundary surface it implies (left). An m-rep model of a hippocampus and its boundary surface (right).

For 3-D figures (Fig. 1), the lattice of medial atoms is a quadrilateral mesh $\mathbf{m}_{ij}, (i, j) \in [1, m] \times [1, n]$. The sampling density of medial atoms in a lattice is inversely proportional to the radius of the medial description. Given an m-rep figure, we fit a smooth boundary surface to the model. We use a subdivision surface method [19] that interpolates the boundary positions and normals implied by each atom.

A medial atom as defined above is a point on the manifold $\mathcal{M}(1) = \mathbb{R}^3 \times \mathbb{R}^+ \times S^2 \times S^2$. Moreover, an m-rep model consisting of n medial atoms may be considered as a point on the manifold $\mathcal{M}(n) = \prod_{i=1}^{n} \mathcal{M}(1)$, i.e., the direct product of n copies of $\mathcal{M}(1)$. The space $\mathcal{M}(n)$ is a particular type of manifold known as a Riemannian symmetric space, which simplifies certain geometric computations, such as computing geodesics and distances. These concepts will be instrumental in our development of PGA, and we review them now.

B. Riemannian Manifolds

A *Riemannian metric* on a manifold M is a smoothly varying inner product $\langle \cdot, \cdot \rangle$ on the tangent space $T_x M$ at each point $x \in M$. The norm of a vector $v \in T_x M$ is given by $\|v\| = \langle v, v \rangle^{(1/2)}$. Given a smooth curve segment in M, its length is computed by integrating the norm of the tangent vectors along the curve. The Riemannian distance between two points $x, y \in M$, denoted $d(x, y)$, is defined as the minimum length over all possible smooth curves between x and y. A *geodesic curve* is a curve that locally minimizes the length between points. An *isometry* of M is a diffeomorphic map $\Phi : M \to M$ that preserves the Riemannian distance, i.e.,

$d(x,y) = d(\Phi(x), \Phi(y))$, for all $x, y \in M$. A manifold is said to be *complete* if all geodesics extend indefinitely. This is an important property because it implies that between any two points there exists a length-minimizing geodesic.

Given a tangent vector $v \in T_xM$, there exists a unique geodesic, $\gamma_v(t)$, with v as its initial velocity. The Riemannian exponential map, denoted Exp_x, maps v to the point at time one along the geodesic γ_v. The geodesic has constant speed equal to $\|d\gamma_v/dt\|(t) = \|v\|$ and, thus, the exponential map preserves distances from the initial point, i.e., $d(x, \mathrm{Exp}_x(v)) = \|v\|$. The exponential map is a diffeomorphism in a neighborhood of zero, and its inverse in this neighborhood is the Riemannian log map, denoted Log_x. Thus, for a point y in the domain of Log_x the geodesic distance between x and y is given by

$$d(x,y) = \|\mathrm{Log}_x(y)\|. \tag{2}$$

C. Lie Groups and Symmetric Spaces

Briefly, a Riemannian symmetric space is a connected manifold M such that at each point the mapping that reverses geodesics through that point is an isometry. For a detailed treatment of symmetric spaces, see the standard texts [20], [21]. Common examples of symmetric spaces are Euclidean spaces, $\mathbb{R}^n$, spheres, S^n, and hyperbolic spaces, H^n. Symmetric spaces, and the methods for computing geodesics and distances on them, arise naturally from Lie group actions on manifolds.

A Lie group G is a differentiable manifold that also forms an algebraic group, where the two group operations

$$\begin{array}{lcll} \mu\colon (x,y) \mapsto xy & : & G \times G \to G & \text{Multiplication} \\ \iota\colon x \mapsto x^{-1} & : & G \to G & \text{Inverse} \end{array}$$

are differentiable mappings (the symbol μ used in this way should not be confused with the mean). Many common geometric transformations of Euclidean space form Lie groups. For example, rotations, translations, magnifications, and affine transformations of $\mathbb{R}^n$ all form Lie groups. More generally, Lie groups can be used to describe transformations of smooth manifolds.

Given a manifold M and a Lie group G, a *smooth group action* of G on M is a smooth mapping $G \times M \to M$, written $(g,x) \mapsto g \cdot x$, such that for all $g, h \in G$, and all $x \in M$ we have $e \cdot x = x$, and $(gh) \cdot x = (g \cdot (h \cdot x))$, where e is the identity element of G. The group action should be thought of as a transformation of the manifold M, just as matrices are transformations of Euclidean space.

The *orbit* of a point $x \in M$ is defined as $G(x) = \{g \cdot x\colon g \in G\}$. In the case that M consists of a single orbit, we call M a *homogeneous space* and say that the group action is *transitive*. The *isotropy subgroup* of x is defined as $G_x = \{g \in G : g \cdot x = x\}$, i.e., G_x is the subgroup of G that leaves the point x fixed.

Let H be a closed Lie subgroup of the Lie group G. Then the *left coset* of an element $g \in G$ is defined as $gH = \{gh : h \in H\}$. The space of all such cosets is denoted G/H and is a smooth manifold. There is a natural bijection $G(x) \cong G/G_x$ given by the mapping $g \cdot x \mapsto gG_x$. Now let M be a symmetric space and choose an arbitrary base point $p \in M$. We can always write M as a homogeneous space $M = G/G_p$, where G is a connected group of isometries of M, and the isotropy subgroup G_p is compact. The fact that G is a group of isometries means that $d(x,y) = d(g \cdot x, g \cdot y)$, for all $x, y \in M, g \in G$.

As an example consider the symmetric space S^2, the sphere in $\mathbb{R}^3$. Rotations of the sphere are a smooth group action by the group Lie $SO(3)$, the 3×3 rotation matrices. We choose the base point to be the north pole, $p = (0,0,1) \in S^2$. It is easy to see that the orbit of p is the entire sphere. Thus, S^2 is a homogeneous space. Also, the isotropy subgroup of p is the group of all rotations about the z-axis, which can be identified with the group of two-dimensional rotations, $SO(2)$. Therefore, S^2 is naturally identified with the quotient space $SO(3)/SO(2)$.

Finally, we turn to the symmetric space of medial atoms, $\mathcal{M}(1) = \mathbb{R}^3 \times \mathbb{R}^+ \times S^2 \times S^2$. The group $G = \mathbb{R}^3 \times \mathbb{R}^+ \times SO(3) \times SO(3)$ acts smoothly on $\mathcal{M}(1)$. Let $g = (\mathbf{v}, s, \mathbf{R}_0, \mathbf{R}_1)$ be an element of G and $\mathbf{m} = (\mathbf{x}, r, \mathbf{n}_0, \mathbf{n}_1)$ be a medial atom. Then the group action is given by

$$g \cdot \mathbf{m} = (\mathbf{x} + \mathbf{v}, s \cdot r, \mathbf{R}_0 \cdot \mathbf{n}_0, \mathbf{R}_1 \cdot \mathbf{n}_1).$$

This action is transitive, and we can choose a base atom p with center $\mathbf{x} = 0$, radius $r = 1$, and both spoke directions, $\mathbf{n}_0, \mathbf{n}_1$, equal to $(0,0,1)$. The isotropy subgroup, G_p, is given by $\{0\} \times \{1\} \times SO(2) \times SO(2)$. Thus, the medial atom space can be thought of as the quotient $\mathcal{M}(1) = \mathbb{R}^3 \times \mathbb{R}^+ \times (SO(3)/SO(2)) \times (SO(3)/SO(2))$.

Other examples of symmetric spaces are the compact Lie groups, such as the rotation groups, $SO(n)$, and the Euclidean groups, $\mathbb{R}^n$. These groups act on themselves transitively by their group multiplication. Thus, the geodesics for such a Lie group at the identity are its one-parameter subgroups.

D. Geodesics

Geodesics on a symmetric space $M = G/G_p$ are computed through the group action. Since G is a group of isometries acting transitively on M, it suffices to consider only geodesics starting at the base point p. For an arbitrary point $x \in M$, geodesics starting at x are of the form $g \cdot \gamma$, where $x = g \cdot p$ and γ is a geodesic with $\gamma(0) = p$. Geodesics are the image of the action of a one-parameter subgroup of G acting on the base point p.

Returning to the sphere, S^2, the geodesics at the base point $p = (0,0,1)$ are the great circles through p, i.e., the meridians. These geodesics are realized by the group action of a one-parameter subgroup of $SO(3)$. Such a subgroup consists of all rotations about a fixed axis in $\mathbb{R}^3$ perpendicular to p. We consider a tangent vector in T_pS^2 as a vector $v = (v_1, v_2, 0)$ in the x-y plane. Then the exponential map is given by

$$\mathrm{Exp}_p(v) = \left(v_1 \cdot \frac{\sin\|v\|}{\|v\|},\ v_2 \cdot \frac{\sin\|v\|}{\|v\|}, \cos\|v\| \right) \tag{3}$$

where $\|v\| = \sqrt{v_1^2 + v_2^2}$. This equation can be derived as a sequence of two rotations that rotate the base point $p = (0,0,1)$ to the point $\mathrm{Exp}_p(v)$. The first is a rotation about the y-axis by an angle of $\phi_y = \|v\|$. The second, aligning the geodesic with the tangent vector v, is a rotation about the z-axis by an angle of ϕ_z, where $\cos(\phi_z) = v_1/\|v\|$ and $\sin(\phi_z) = v_2/\|v\|$.

The corresponding log map for a point $x = (x_1, x_2, x_3) \in S^2$ is given by

$$\mathrm{Log}_p(x) = \left(x_1 \cdot \frac{\theta}{\sin\theta}, x_2 \cdot \frac{\theta}{\sin\theta} \right) \tag{4}$$

where $\theta = \arccos(x_3)$ is the spherical distance from the base point p to the point x. Notice that the antipodal point $-p$ is not in the domain of the log map.

III. Means on Manifolds

The first step in extending statistical methods to manifolds is to define the notion of a mean value. In this section, we formulate two different notions of means on manifolds. We then describe a method for computing the mean of a collection of data on a manifold. Throughout this section we consider only manifolds that are connected and have a complete Riemannian metric.

A. Intrinsic Versus Extrinsic Means

Given a set of points $x_1, \ldots, x_N \in \mathbb{R}^d$, the arithmetic mean $\bar{x} = (1/N)\sum_{i=1}^{N} x_i$ is the point that minimizes the sum-of-squared Euclidean distances to the given points, i.e.,

$$\bar{x} = \arg\min_{x \in \mathbb{R}^d} \sum_{i=1}^{N} \|x - x_i\|^2.$$

Since a general manifold M may not form a vector space, the notion of an additive mean is not necessarily valid. However, like the Euclidean case, the mean of a set of points on M can be formulated as the point which minimizes the sum-of-squared distances to the given points. This formulation depends on the definition of distance. One way to define distance on M is to embed it in a Euclidean space and use the Euclidean distance between points. This notion of distance is extrinsic to M, that is, it depends on the ambient space and the choice of embedding. Given an embedding $\Phi : M \to \mathbb{R}^d$, define the *extrinsic mean* [22] of a collection of points $x_1, \ldots, x_N \in M$ as

$$\mu_\Phi = \arg\min_{x \in M} \sum_{i=1}^{N} \|\Phi(x) - \Phi(x_i)\|^2.$$

Given the above embedding of M, we can also compute the arithmetic (Euclidean) mean of the embedded points and then project this mean onto the manifold M. This projected mean is equivalent to the above definition of the extrinsic mean (see [23]). Define a projection mapping $\pi : \mathbb{R}^d \to G$ as

$$\pi(x) = \arg\min_{y \in M} \|\Phi(y) - x\|^2.$$

Then, the extrinsic mean is given by

$$\mu_\Phi = \pi\left(\frac{1}{N}\sum_{i=1}^{N} \Phi(x_i)\right).$$

A more natural choice of distance is the Riemannian distance on M. This definition of distance depends only on the intrinsic geometry of M. We now define the *intrinsic mean* of a collection of points $x_1, \ldots, x_N \in M$ as the minimizer in M of the sum-of-squared Riemannian distances to each point. Thus, the intrinsic mean is

$$\mu = \arg\min_{x \in M} \sum_{i=1}^{N} d(x, x_i)^2 \tag{5}$$

where $d(\cdot, \cdot)$ denotes Riemannian distance on M. This is the definition of a mean value that we use in this paper.

The idea of an intrinsic mean goes back to Fréchet [24], who defines it for a general metric space. The properties of the intrinsic mean on a Riemannian manifold have been studied by Karcher [25]. Moakher [26] compares the properties of the intrinsic and extrinsic mean for the group of 3-D rotations. Since the intrinsic mean is defined in (5) as a minimization problem, its existence and uniqueness are not ensured. However, Kendall [27] shows that the intrinsic mean exists and is unique if the data is well-localized.

B. Computing the Intrinsic Mean

Computation of the intrinsic mean involves solving the minimization problem in (5). We will assume that our data $x_1, \ldots, x_n \in M$ lies in a sufficiently small neighborhood so that a unique solution is guaranteed. We must minimize the sum-of-squared distance function

$$f(x) = \frac{1}{2N}\sum_{i=1}^{N} d(x, x_i)^2.$$

We now describe a gradient descent algorithm, first proposed by Pennec [11], for minimizing f. Using the assumption that the x_i lie in a strongly convex neighborhood, i.e., a neighborhood U such that any two points in U are connected by a unique geodesic contained completely within U, Karcher [25] shows that the gradient of f is

$$\nabla f(x) = -\frac{1}{N}\sum_{i=1}^{N} \mathrm{Log}_x(x_i).$$

The gradient descent algorithm takes successive steps in the negative gradient direction. Given a current estimate μ_j for the intrinsic mean, the equation for updating the mean by taking a step in the negative gradient direction is

$$\mu_{j+1} = \mathrm{Exp}_{\mu_j}\left(\frac{\tau}{N}\sum_{i=1}^{N} \mathrm{Log}_{\mu_j}(x_i)\right)$$

where τ is the step size.

Because the gradient descent algorithm only converges locally, care must be taken in the choices of the initial estimate of the mean μ_0 and the step size τ. Since the data is assumed to be well-localized, a reasonable choice for the initial estimate μ_0 is one of the data points, say x_1. The choice of τ is somewhat harder and depends on the manifold M. Buss and Fillmore [28] prove for data on spheres, a value of $\tau = 1$ is sufficient. Notice that if M is a vector space, the gradient descent algorithm with $\tau = 1$ is equivalent to linear averaging and, thus, converges in a single step. If $M = \mathbb{R}^+$, the Lie group of positive reals under multiplication, the algorithm with $\tau = 1$ is equivalent to the geometric average and again converges in a single step.

In summary, we have the following algorithm for computing the intrinsic mean of manifold data.

Algorithm 1: Intrinsic Mean
Input: $x_1, \ldots, x_N \in M$
Output: $\mu \in M$, the intrinsic mean
$\mu_0 = x_1$
Do
$\Delta\mu = \frac{\tau}{N}\sum_{i=1}^{N} \mathrm{Log}_{\mu_j} x_i$
$\mu_{j+1} = \mathrm{Exp}_{\mu_j}(\Delta\mu)$
While $\|\Delta\mu\| > \epsilon$.

IV. Principal Geodesic Analysis

Although averaging methods on manifolds have previously been studied, principal component analysis has not been developed for manifolds. We present a new method called *principal geodesic analysis* (PGA), a generalization of principal component analysis to manifolds. We start with a review of PCA in Euclidean space. Consider a set of points $x_1, \ldots, x_N \in \mathbb{R}^d$ with zero mean. Principal component analysis seeks a sequence of linear subspaces that best represent the variability of the data. To be more precise, the intent is to find a orthonormal basis $\{v_1, \ldots, v_d\}$ of $\mathbb{R}^d$, which satisfies the recursive relationship

$$v_1 = \arg\max_{||v||=1} \sum_{i=1}^{N} (v \cdot x_i)^2 \tag{6}$$

$$v_k = \arg\max_{||v||=1} \sum_{i=1}^{N} \sum_{j=1}^{k-1} (v_j \cdot x_i)^2 + (v \cdot x_i)^2. \tag{7}$$

In other words, the subspace $V_k = \text{span}(\{v_1, \ldots, v_k\})$ is the k-dimensional subspace that maximizes the variance of the data projected to that subspace. The basis $\{v_k\}$ is computed as the set of ordered eigenvectors of the sample covariance matrix of the data.

Now turning to manifolds, consider a set of points $x_1, \ldots, x_N$ on a manifold M. Our goal is to describe the variability of the x_i in a way that is analogous to PCA. Thus, we will project the data onto lower dimensional subspaces that best represent the variability of the data. This requires first extending three important concepts of PCA into the manifold setting:

- *Variance*: Following the work of Fréchet, we define the sample variance of the data as the expected value of the squared Riemannian distance from the mean.
- *Geodesic subspaces*: The lower dimensional subspaces in PCA are linear subspaces. For general manifolds we extend the concept of a linear subspace to that of a *geodesic submanifold*.
- *Projection*: In PCA, the data is projected onto linear subspaces. We define a projection operator for geodesic submanifolds, and show how it may be efficiently approximated.

We now develop each of these concepts in detail.

A. Variance

The variance σ^2 of a real-valued random variable x with mean μ is given by the formula

$$\sigma^2 = \mathcal{E}[(x - \mu)^2]$$

where $\mathcal{E}$ denotes expectation. It measures the expected localization of the variable x about the mean. When dealing with a vector-valued random variable $\mathbf{x}$ in $\mathbb{R}^d$ with mean $\boldsymbol{\mu}$, the variance is replaced by a covariance matrix

$$\Sigma = \mathcal{E}[(\mathbf{x} - \mu)(\mathbf{x} - \mu)^T].$$

However, this definition is not valid for general manifolds again since vector space operations do not exist for such spaces.

The definition of variance we use comes from Fréchet [24], who defines the variance of a random variable in a metric space as the expected value of the squared distance from the mean. That is, for a random variable x in a metric space with intrinsic mean μ, the variance is given by

$$\sigma^2 = \mathcal{E}[d(\mu, x)^2].$$

Thus, given data points $x_1, \ldots, x_N$ on a complete, connected manifold M, we define the sample variance of the data as

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} d(\mu, x_i)^2 = \frac{1}{N} \sum_{i=1}^{N} ||\text{Log}_\mu(x_i)||^2 \tag{8}$$

where μ is the intrinsic mean of the x_i.

Notice that if M is a vector space, then the variance definition in (8) is given by the trace of the sample covariance matrix, i.e., the sum of its eigenvalues. It is in this sense that this definition captures the total variation of the data.

B. Geodesic Submanifolds

The next step in generalizing PCA to manifolds is to generalize the notion of a linear subspace. A geodesic is a curve that is locally the shortest path between points. In this way, a geodesic is the generalization of a straight line. Thus, it is natural to use a geodesic curve as the one-dimensional subspace, i.e., the analog of the first principal direction in PCA.

In general, if N is a submanifold of a manifold M, geodesics of N are not necessarily geodesics of M. For instance, the sphere S^2 is a submanifold of $\mathbb{R}^3$, but its geodesics are great circles, while geodesics of $\mathbb{R}^3$ are straight lines. A submanifold H of M is said to be geodesic at $x \in H$ if all geodesics of H passing through x are also geodesics of M. For example, a linear subspace of $\mathbb{R}^d$ is a submanifold geodesic at 0. Submanifolds geodesic at x preserve distances to x. This is an essential property for PGA because variance is defined as the average squared distance to the mean. Thus, submanifolds geodesic at the mean will be the generalization of the linear subspaces of PCA.

C. Projection

The projection of a point $x \in M$ onto a geodesic submanifold H of M is defined as the point on H that is nearest to x in Riemannian distance. Thus, we define the projection operator $\pi_H : M \to H$ as

$$\pi_H(x) = \arg\min_{y \in H} d(x, y)^2.$$

Since projection is defined by a minimization, there is no guarantee that the projection of a point exists or that it is unique. However, by restricting to a small enough neighborhood about the mean, we can be assured that projection is unique for any submanifold geodesic at the mean.

Projection onto a geodesic submanifold can be approximated linearly in the tangent space of M. Let $H \subset M$ be a geodesic submanifold at a point $p \in M$ and $x \in M$ a point to be projected onto H. Then the projection operator is approximated by

$$\begin{aligned} \pi_H(x) &= \arg\min_{y \in H} ||\text{Log}_x(y)||^2 \\ &\approx \arg\min_{y \in H} ||\text{Log}_p(x) - \text{Log}_p(y)||^2. \end{aligned}$$

Notice that $\text{Log}_p(y)$ is simply a vector in T_pH. Thus, we may rewrite the approximation in terms of tangent vectors as

$$\text{Log}_p(\pi_H(x)) \approx \arg\min_{v \in T_pH} ||\text{Log}_p(x) - v||^2.$$

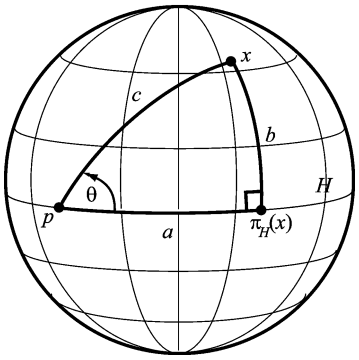

Fig. 2. The spherical triangle used in the calculation of the projection operator for S^2.

But this is simply the minimization formula for linear projection of $\mathrm{Log}_p(x)$ onto the linear subspace T_pH. So, if $v_1, \ldots, v_k$ is an orthonormal basis for T_pH, then the projection operator can be approximated by the formula

$$\mathrm{Log}_p(\pi_H(x)) \approx \sum_{i=1}^{k} \langle v_i, \mathrm{Log}_p(x) \rangle. \tag{9}$$

Analyzing the quality of the approximation in (9) may be difficult for general manifolds. Here we demonstrate the error computations for the special case of the sphere S^2. Let H be a geodesic (i.e., a great circle) through a point $p \in S^2$. Given a point $x \in S^2$, we wish to compute its true projection onto H and compare that with the approximation in the tangent space T_pS^2. Thus, we have the spherical right triangle as shown in Fig. 2. We know the hypotenuse length $c = d(p,x)$ and the angle θ, and we want to derive the true projection, which is given by the side length a. We use the following two relations from the laws of spherical trigonometry:

$$\cos c = (\cos a)(\cos b)$$
$$\frac{\sin b}{\sin \theta} = \sin c.$$

Solving for a in terms of the hypotenuse c and the angle θ, we have

$$a = \arccos\left(\frac{\cos c}{\sqrt{1 - (\sin\theta \sin b)^2}} \right).$$

The tangent-space approximation in (9) is equivalent to solving for the corresponding right triangle in $\mathbb{R}^2$. Using standard Euclidean trigonometry, the tangent-space approximation (9) gives

$$a \approx c \cos\theta.$$

For nearby data, i.e., small values for c, this gives a good approximation. For example, for $c < (\pi/4)$ the maximum absolute error is 0.07 rad. However, the error can be significant for far away points, i.e., as c approaches $(\pi/2)$.

D. Computing Principal Geodesic Analysis

We are now ready to define PGA for data $x_1, \ldots, x_N$ on a connected, complete manifold M. Our goal, analogous to PCA, is to find a sequence of nested geodesic submanifolds that maximize the projected variance of the data. These submanifolds are called the *principal geodesic submanifolds*.

Let $T_\mu M$ denote the tangent space of M at the intrinsic mean μ of the x_i. Let $U \subset T_\mu M$ be a neighborhood of 0 such that projection is well-defined for all geodesic submanifolds of $\mathrm{Exp}_\mu(U)$. We assume that the data is localized enough to lie within such a neighborhood. The principal geodesic submanifolds are defined by first constructing an orthonormal basis of tangent vectors $v_1, \ldots, v_d \in T_\mu M$ that span the tangent space $T_\mu M$. These vectors are then used to form a sequence of nested subspaces $V_k = \mathrm{span}(\{v_1, \ldots, v_k\}) \cap U$. The principal geodesic submanifolds are the images of the V_k under the exponential map: $H_k = \mathrm{Exp}_\mu(V_k)$. The first principal direction is chosen to maximize the projected variance along the corresponding geodesic

$$v_1 = \arg\max_{||v||=1} \sum_{i=1}^{N} ||\mathrm{Log}_\mu(\pi_H(x_i))||^2 \tag{10}$$

where

$$H = \mathrm{Exp}_\mu(\mathrm{span}(\{v\}) \cap U).$$

The remaining principal directions are defined recursively as

$$v_k = \arg\max_{||v||=1} \sum_{i=1}^{N} ||\mathrm{Log}_\mu(\pi_H(x_i))||^2 \tag{11}$$

where

$$H = \mathrm{Exp}_\mu(\mathrm{span}(\{v_1, \ldots, v_{k-1}, v\}) \cap U).$$

If we use (9) to approximate the projection operator π_H in (10) and (11), we get

$$v_1 \approx \arg\max_{||v||=1} \sum_{i=1}^{N} \langle v, \mathrm{Log}_\mu(x_i) \rangle^2$$

$$v_k \approx \arg\max_{||v||=1} \sum_{i=1}^{N} \sum_{j=1}^{k-1} \langle v_j, \mathrm{Log}_\mu(x_i) \rangle^2 + \langle v, \mathrm{Log}_\mu(x_i) \rangle^2.$$

The above minimization problem is simply the standard principal component analysis in $T_\mu M$ of the vectors $\mathrm{Log}_\mu(x_i)$, which can be seen by comparing the approximations above to the PCA equations, (6) and (7). Thus, an algorithm for approximating the PGA of data on a manifold is as follows.

Algorithm 2: Principal Geodesic Analysis
Input: $x_1, \ldots, x_N \in M$
Output: Principal directions, $v_k \in T_\mu M$
Variances, $\lambda_k \in \mathbb{R}$
$\mu =$ intrinsic mean of $\{x_i\}$ (Algorithm 1)
$u_i = \mathrm{Log}_\mu(x_i)$
$\mathbf{S} = (1/N) \sum_{i=1}^{N} u_i u_i^T$
$\{v_k, \lambda_k\} =$ eigenvectors/eigenvalues of $\mathbf{S}$.

V. Application to M-reps

We now apply the statistical framework presented above for general manifolds to the statistical analysis of m-rep models of anatomical objects. That is, we will apply the mean and PGA algorithms to the symmetric space $\mathcal{M}(n)$, representing m-rep models with n atoms. The initial data is a set of m-rep models $\mathbf{M}_1, \ldots, \mathbf{M}_N \in \mathcal{M}(n)$ that have been fit to a particular class of objects in a training set of images. As is the case with other

shape analysis methods, since we are interested in studying the variability of shape alone, we must first align the models to a common position, orientation, and scale. We present an m-rep alignment algorithm that minimizes the sum-of-squared geodesic distances between models, i.e., has the desirable property that it minimizes the same metric as is used in the definition of the mean and principal geodesics, but over the global similarity transformations of alignment. Next the mean and PGA algorithms are adapted to the specific case of m-rep models.

The results of these techniques are demonstrated on a set of 86 m-rep models of hippocampi from a schizophrenia study. A subset of 16 of these models are displayed as surfaces in Fig. 3. The m-rep models were automatically generated by the method described in [29], which chooses the medial topology and sampling that is sufficient to represent the population of objects. The models were fit to expert segmentations of the hippocampi from MRI data. The average distance error from the m-rep boundary to the original segmentation boundary ranged from 0.14 mm and 0.27 mm with a mean error of 0.17 mm. This is well within the original MRI voxel size (0.9375 mm $\times$ 0.9375 mm 1.5 mm). The sampling on each m-rep was 3×8, making each model a point on the symmetric space $\mathcal{M}(24)$. Since the dimensionality of $\mathcal{M}(1)$ is 8, the total number of dimensions required to represent the hippocampus models is 192.

A. The Exponential and Log Maps for M-Reps

Before we can apply the statistical techniques for manifolds developed in the previous sections, we must define the exponential and log maps for the symmetric space $\mathcal{M}(n)$, the space of m-rep models with n atoms. We begin with a discussion of the medial atom space $\mathcal{M}(1) = \mathbb{R}^3 \times \mathbb{R}^+ \times S^2 \times S^2$. Let $p = (0, 1, p_0, p_1) \in \mathcal{M}(1)$ be the base point, where $p_0 = p_1 = (0, 0, 1)$ are the base points for the spherical components. The tangent space for $\mathcal{M}(1)$ at the base point p can be identified with $\mathbb{R}^8$. We write a tangent vector $u \in T_p\mathcal{M}(1)$ as $u = (\mathbf{x}, \rho, v_0, v_1)$, where $\mathbf{x} \in \mathbb{R}^3$ is the positional tangent component, $\rho \in \mathbb{R}$ is the radius tangent component, and $v_0, v_1 \in \mathbb{R}^2$ are the spherical tangent components. The exponential map for $\mathcal{M}(1)$ is now the direct product of the exponential map for each component. The exponential map for $\mathbb{R}^3$ is simply the identity map, for $\mathbb{R}$ it is the standard real exponential function, and for S^2 it is the spherical exponential map given in (3). Thus, for $\mathcal{M}(1)$ we have

$$\mathrm{Exp}_p(u) = \big(\mathbf{x}, e^{\rho}, \mathrm{Exp}_{p_0}(v_0), \mathrm{Exp}_{p_1}(v_1)\big)$$

where the two Exp maps on the right-hand side (RHS) are the spherical exponential maps. Likewise, the log map of a point $\mathbf{m} = (\mathbf{x}, r, \mathbf{n}_0, \mathbf{n}_1)$ is the direct product map

$$\mathrm{Log}_p(\mathbf{m}) = \big(\mathbf{x}, \log r, \mathrm{Log}_{p_0}(\mathbf{n}_0), \mathrm{Log}_{p_1}(\mathbf{n}_1)\big)$$

where the two Log maps on the RHS are the spherical log maps given by (4). Finally, the exponential and log maps for the m-rep model space $\mathcal{M}(n)$ are just the direct products of n copies of the corresponding maps for the medial atom space $\mathcal{M}(1)$.

Notice that the position, radius, and orientations are not in the same units. For the PGA calculations in Section IV, we scale the radius and sphere components in the Riemannian metric to be commensurate with the positional components. The scaling factor for both components is the average radius over all corresponding medial atoms in the population. Thus, the norm of the vector $u = T_p\mathcal{M}(1)$ becomes

$$\|u\| = (\|\mathbf{x}\|^2 + \bar{r}^2(\rho^2 + \|v_1\|^2 + \|v_2\|^2))^{\frac{1}{2}}$$

where $\bar{r}$ is the average radius over all corresponding medial atoms. Using this norm and the formula for Riemannian distance (2), the distance between two atoms $\mathbf{m}_1, \mathbf{m}_2 \in \mathcal{M}(1)$ is given by

$$d(\mathbf{m}_1, \mathbf{m}_2) = \|\mathrm{Log}_{\mathbf{m}_1}(\mathbf{m}_2)\|. \tag{12}$$

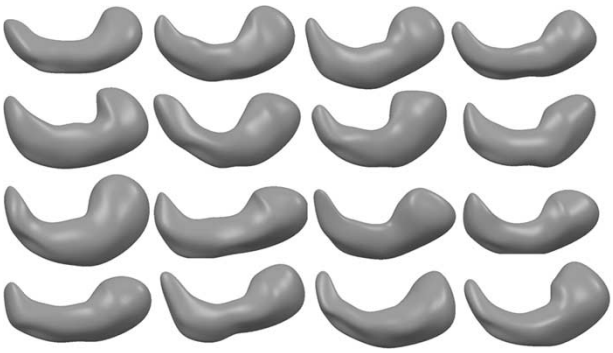

Fig. 3. The surfaces of 16 of the 86 original hippocampus m-rep models.

B. M-rep Alignment

To globally align objects described by boundary points to a common position, orientation, and scale, the standard method is the Procrustes method [10]. Procrustes alignment minimizes the sum-of-squared distances between corresponding boundary points, the same metric used in defining the mean and principal components. We now develop an analogous alignment procedure based on minimizing sum-of-squared geodesic distances on $\mathcal{M}(n)$, the symmetric space of m-rep objects with n atoms.

Let $\mathbf{S} = (s, \mathbf{R}, \mathbf{w})$ denote a similarity transformation in $\mathbb{R}^3$ consisting of a scaling by $s \in \mathbb{R}^+$, a rotation by $\mathbf{R} \in SO(3)$, and a translation by $\mathbf{w} \in \mathbb{R}^3$. We define the action of $\mathbf{S}$ on a medial atom $\mathbf{m} = (\mathbf{x}, r, \mathbf{n}_0, \mathbf{n}_1)$ by

$$\mathbf{S}\cdot\mathbf{m} = \mathbf{S}\cdot(\mathbf{x}, r, \mathbf{n}_0, \mathbf{n}_1) = (s\mathbf{R}\cdot\mathbf{x}+\mathbf{w}, sr, \mathbf{R}\cdot\mathbf{n}_0, \mathbf{R}\cdot\mathbf{n}_1). \tag{13}$$

Now the action of $\mathbf{S}$ on an m-rep object $\mathbf{M} = \{\mathbf{m}_i : i = 1, \ldots, n\}$ is simply the application of $\mathbf{S}$ to each of $\mathbf{M}$'s medial atoms

$$\mathbf{S}\cdot\mathbf{M} = \{\mathbf{S}\cdot\mathbf{m}_i : i = 1, \ldots, n\}. \tag{14}$$

It is easy to check from (1) that this action of $\mathbf{S}$ on $\mathbf{M}$ also transforms the implied boundary points of $\mathbf{M}$ by the similarity transformation $\mathbf{S}$.

Consider a collection $\mathbf{M}_1, \ldots, \mathbf{M}_N \in \mathcal{M}(n)$ of m-rep objects to be aligned, each consisting of n medial atoms. We write $\mathbf{m}_{\alpha i}$ to denote the ith medial atom in the αth m-rep object. Notice that the m-rep parameters, which are positions, orientations, and scalings, are in different units. Before we apply PGA to the

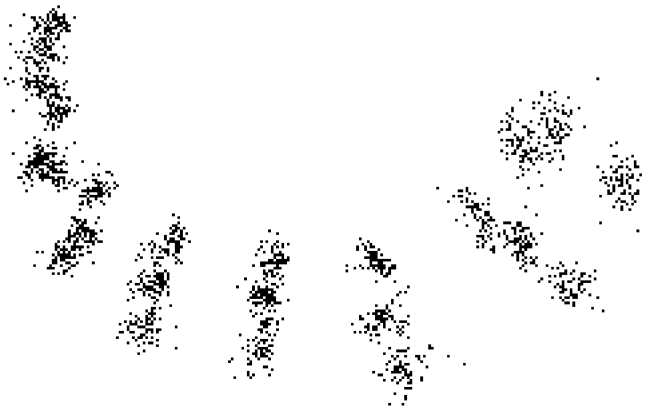

Fig. 4. The 86 aligned hippocampus m-reps, shown as overlayed medial atom centers.

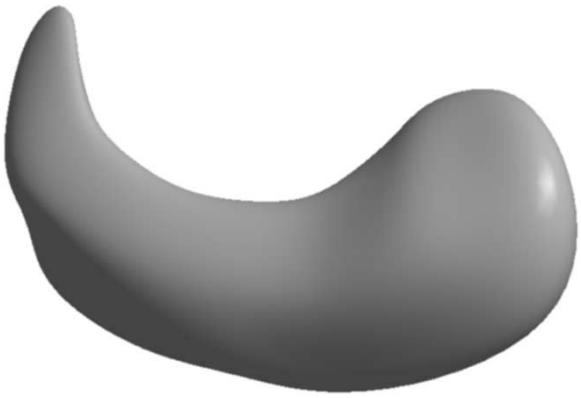

Fig. 5. The surface of the mean hippocampus m-rep.

m-reps, it is necessary to make the various parameters commensurate. This is done by scaling the log rotations and log radii by the average radius value of the corresponding medial atoms. The squared-distance metric between two m-rep models $\mathbf{M}_i$ and $\mathbf{M}_j$ becomes

$$d(\mathbf{M}_i, \mathbf{M}_j)^2 = \sum_{\alpha=1}^{n} d(\mathbf{m}_{\alpha i}, \mathbf{m}_{\alpha j})^2 \tag{15}$$

where the $d(\cdot,\cdot)$ for medial atoms on the RHS is given by (12).

The m-rep alignment algorithm finds the set of similarity transforms $\mathbf{S}_1, \ldots, \mathbf{S}_N$ that minimize the total sum-of-squared distances between the m-rep figures:

$$d(\mathbf{S}_1, \ldots, \mathbf{S}_N; \mathbf{M}_1, \ldots, \mathbf{M}_N) = \sum_{i=1}^{N} \sum_{j=1}^{i} d(\mathbf{S}_i \cdot \mathbf{M}_i, \mathbf{S}_j \cdot \mathbf{M}_j)^2. \tag{16}$$

Following the algorithm for generalized Procrustes analysis for objects in $\mathbb{R}^3$, minimization of (16) proceeds in stages.

Algorithm 3: M-rep Alignment

1. *Translations*. First, the translational part of each $\mathbf{S}_i$ in (16) is minimized once and for all by centering each m-rep model. That is, each model is translated so that the average of it's medial atoms' positions is the origin.

2. *Rotations and Scalings*. The ith model, $\mathbf{M}_i$, is aligned to the mean of the remaining models, denoted μ_i. The alignment is accomplished by a gradient descent algorithm on $SO(3) \times \mathbb{R}^+$ to minimize $d(\mu_i, \mathbf{S}_i \cdot \mathbf{M}_i)^2$. The gradient is approximated numerically by a central differences scheme. This is done for each of the N models.

3. *Iterate*. Step 2 is repeated until the metric (16) cannot be further minimized.

The result of applying the m-rep alignment algorithm to the 86 hippocampus m-rep models is shown in Fig. 4. The resulting aligned figures are displayed as overlayed medial atom centers. Since the rotation and scaling step of the alignment algorithm is a gradient descent algorithm, it is important to find a good starting position. Thus, the alignment was initialized by first aligning the m-rep models with the Procrustes method applied to the implied boundary points of the m-rep models.

C. M-rep Averages

Algorithm 1 can be adapted for computing means of m-rep models by taking the manifold to be the symmetric space $\mathcal{M}(n)$. Since this is a direct product space, the algorithm will converge if each of the components converge. Notice that each of the $\mathbb{R}^3$ and $\mathbb{R}^+$ components in $\mathcal{M}(n)$ converge in a single iteration since they are commutative Lie groups. A step size of $\tau = 1$ is sufficient to ensure that the S^2 components converge as well. Also, care must be taken to ensure that the data is contained in a small enough neighborhood that the minimum in (5) is unique. For the $\mathbb{R}^3$ and $\mathbb{R}^+$ components there is no restriction on the spread of the data. However, for the S^2 components the data must lie within a neighborhood of radius $(\pi/2)$ (see [28]), i.e., within an open hemisphere. This is a reasonable assumption for the aligned m-rep models, whose spoke directions for corresponding atoms are fairly localized, and we have not experienced in practice any models that do not fall within such constraints. We now have the following algorithm for computing the intrinsic mean of a collection of m-rep models.

Algorithm 4: M-rep Mean

Input: $\mathbf{M}_1, \ldots, \mathbf{M}_N \in \mathcal{M}(n)$, m-rep models
Output: $\mu \in \mathcal{M}(n)$, the intrinsic mean

$\mu_0 = \mathbf{M}_1$
Do
$\Delta\mu = (1/N) \sum_{i=1}^{N} \mathrm{Log}_{\mu_j} \mathbf{M}_i$
$\mu_{j+1} = \mathrm{Exp}_{\mu_j}(\Delta\mu)$
While $||\Delta\mu|| > \epsilon$.

Fig. 5 shows the surface of the resulting intrinsic mean of the 86 aligned hippocampus m-rep models computed by Algorithm 4. The maximum difference in the rotation angle from the mean in either of the S^2 components was 0.1276 for the entire data set. Thus, the data falls well within a neighborhood of radius $(\pi/2)$ as required.

One might be tempted to simplify the statistical computations by treating a medial atom as three points in $\mathbb{R}^3$: the center point $\mathbf{x}$, and the two implied boundary points $\mathbf{y}_0, \mathbf{y}_1$. With this linear

representation, the symmetric space mean algorithm involving geodesic computations is replaced by a simpler linear average. However, linear averaging produces invalid medial atoms. To demonstrate this we computed a linear average of the atoms at a corresponding location in the hippocampus mesh across the population. This average was compared to the symmetric space average described in this paper. The resulting two medial atoms are shown in Fig. 6. The symmetric space mean is a valid medial atom, while the linear average is not because the two spoke vectors do not have equal length. The ratio of the two spoke lengths in the linear average is 1.2 to 1.

D. M-rep PGA

The PGA algorithm for m-rep models is a direct adaptation of Algorithm 2. The only concern is to check that the data is localized enough for the projection operator to be unique. That is, we must determine the neighborhood U used in (10) and (11). Again there is no restriction on the $\mathbb{R}^3$ and $\mathbb{R}^+$ components. For S^2 components it is also sufficient to consider a neighborhood with radius $(\pi/2)$. Therefore, there are no further constraints on the data than those discussed for the mean. Also, we can expect the projection operator to be well-approximated in the tangent space, given the discussion of the error in Section IV.C and the fact that the data lie within 0.1276 rad. from the mean. Finally, the computation of the PGA of a collection of m-rep models is as follows.

Algorithm 5: M-rep PGA
Input: M-rep models, $\mathbf{M}_1, \ldots, \mathbf{M}_N \in \mathcal{M}(n)$
Output: Principal directions, $v_k \in T_\mu \mathcal{M}(n)$
Variances, $\lambda_k \in \mathbb{R}$
$\mu =$ intrinsic mean of $\{\mathbf{M}_i\}$ (Algorithm 4)
$\mathbf{u}_i = \mathrm{Log}_\mu(\mathbf{M}_i)$
$\mathbf{S} = (1/N)\sum_{i=1}^{N} \mathbf{u}_i \mathbf{u}_i^T$
$\{v_k, \lambda_k\} =$ eigenvectors/eigenvalues of $\mathbf{S}$.

Analogous to linear PCA models, we may choose a subset of the principal directions v_k that is sufficient to describe the variability of the m-rep shape space. New m-rep models may be generated within this subspace of typical objects. Given a set of real coefficients $\alpha = (\alpha_1, \ldots, \alpha_d)$, we generate a new m-rep model by

$$\mathbf{M}(\alpha) = \mathrm{Exp}_\mu \left(\sum_{k=1}^{d} \alpha_k v_k \right) \tag{17}$$

where α_k is chosen to be within $[-3\sqrt{\lambda_k}, 3\sqrt{\lambda_k}]$.

The m-rep PGA algorithm was applied to the aligned hippocampus data set. Fig. 7 displays the first three modes of variation as the implied boundaries of the m-reps generated from PGA coefficients $\alpha_k = -3\sqrt{\lambda_k}, -1.5\sqrt{\lambda_k}, 0, 1.5\sqrt{\lambda_k}, 3\sqrt{\lambda_k}$. A plot of the eigenvalues and their cumulative sums is given in Fig. 8. The first 30 modes capture 95 percent of the total variability, which is a significant reduction from the original 192 dimensions of the hippocampus m-rep model.

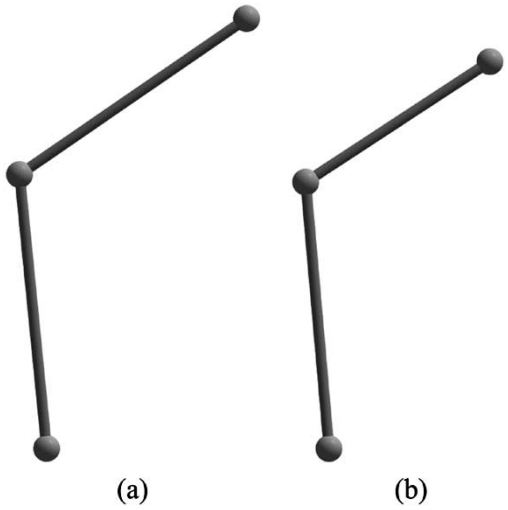

Fig. 6. The resulting average of corresponding medial atoms in the hippocampus models using (a) symmetric space averaging and (b) linear averaging. Notice that the linear average is not a valid medial atom as the two spokes do not have equal length.

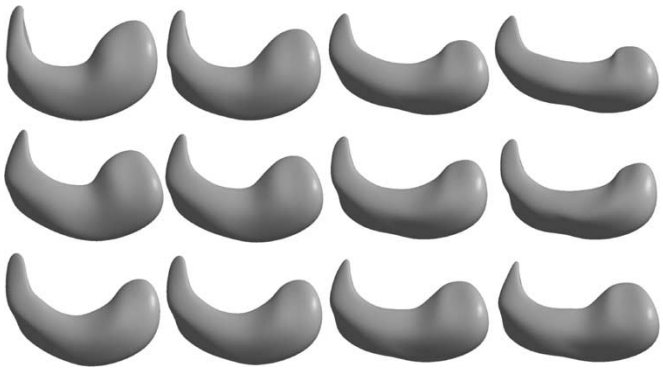

Fig. 7. The first three PGA modes of variation for the hippocampus m-reps. From left to right are the PGA deformations for $-3, -1.5, 1.5$, and $3 \times \sqrt{\lambda_i}$.

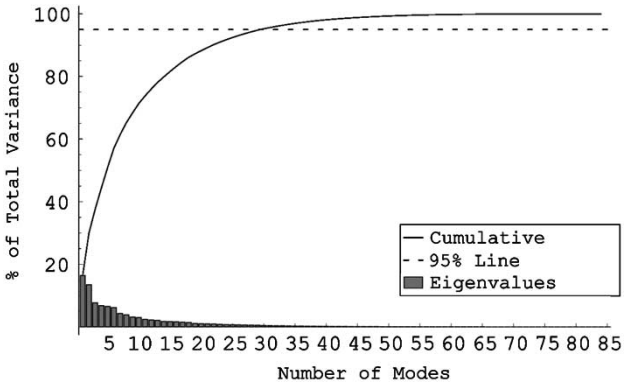

Fig. 8. A plot of the eigenvalues from the modes of variation and their cumulative sums.

VI. Discussion

We presented a new approach to describing shape variability through PGA of medial representations. While m-rep parameters are not linear vector spaces, we showed that they are elements of a Riemannian symmetric space. We developed PGA as a method for efficiently describing the variability of data on

a manifold. The statistical methods for computing averages and principal geodesic analyses were applied to the study of shape from m-rep models.

We believe the methods presented in this paper will have application well beyond m-reps. PGA is a promising technique for describing the variability of data that is inherently nonlinear. As Lie groups such as translations, rotations, and scalings are common entities in image analysis and computer vision, statistical analysis on Lie groups is a promising area for future applications. Also, statistics on linear models may benefit from the addition of nonlinear information. For instance, the point distribution model [4] might be augmented with surface normals, represented as points on a sphere, and handled under the PGA framework.

There is a method called principal curves [30], which has a similar name to PGA. However, the two methods are only loosely related. Principal curves are smooth curves that are fit to data in Euclidean space by minimizing the sum-of-squared Euclidean distances to the data. PGA on the other hand concerns data that lie on a manifold, rather than Euclidean space. Principal geodesic submanifolds are intrinsic to the underlying space, and they minimize sum-of-squared geodesic distances in that space.

A. Application to Image Segmentation

We now briefly describe how PGA can be used to guide a deformable model image segmentation. We are given an image I and we want to fit an m-rep model M to a particular object within the image. The m-rep model M has been trained, i.e., a PGA has been computed, on a training population of m-rep models fit to known objects. PGA will be used to restrict the shape of the model to statistically feasible instances of the object. Following a Bayesian framework, we maximize the log-posterior

$$\log p(\mathbf{M} \mid I) \propto \log p(I \mid \mathbf{M}) + \log p(\mathbf{M}).$$

We do not discuss here the image log-likelihood term, $\log p(I \mid \mathbf{M})$. This term, along with other details of m-rep segmentation, can be found in [15]. The segmentation is initialized by placing the mean model in the image. The model's geometry is deformed within the image by simultaneously optimizing over the parameters α_k in (17) and over global position, orientation, and scale. The geometric log-prior is defined as the squared Mahalanobis distance

$$\log p(\mathbf{M}) \propto \sum_{k=1}^{d} \frac{\alpha_k^2}{\lambda_k}.$$

This segmentation strategy has been implemented, and preliminary tests on CT images of the kidney have been promising. Production of further results and a quantitative analysis of the quality of the segmentations are in progress.

B. Future Work

Another application of PGA is to the statistical analysis of diffusion tensor images. A diffusion tensor is a 3×3 real, symmetric, positive-definite matrix. The space of all such matrices forms a symmetric space $P(3) = GL^+(3)/SO(3)$, where $GL^+(3)$ denotes the Lie group of all 3×3 matrices with positive determinant. We envision using PGA for studying the statistical variability of diffusion tensor images across patients in a statistical atlas framework.

We plan to extend our analysis to more complex m-rep models. This includes objects consisting of several figures, i.e., objects that have a branched medial axis. Also, we intend to handle scenes containing multiple objects. Preliminary work in this area has produced a statistical liver model consisting of several connected figures and a heart model built from multiple objects. In addition to the segmentation application mentioned in the previous section, we foresee an application to shape discrimination, for example, for the separation of the hippocampi into schizophrenics and controls. This requires the development of statistical techniques for discrimination on symmetric spaces.

ACKNOWLEDGMENT

The authors acknowledge J. Lieberman and G. Gerig, UNC Psychiatry, for providing hippocampal data from a clinical Schizophrenia study which was supported by Stanley Foundation and by the UNC-MHNCRC (MH33127). They would like to thank J. Townsend for providing the m-rep segmentations of the data. They would like to thank U. Grenander, A. Srivastava, and A. Budhiraja for useful discussions regarding statistics on Lie groups.

REFERENCES

[1] I. Dryden and K. Mardia, *Statistical Shape Analysis*. New York: Wiley, 1998.
[2] D. G. Kendall, "A survey of the statistical theory of shape," *Statist. Sci.*, vol. 4, no. 2, pp. 87–120, 1989.
[3] C. G. Small, *The Statistical Theory of Shape*. Berlin, Germany: Springer-Verlag, 1996.
[4] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models—Their training and application," *Comput. Vis. Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.
[5] J. Csernansky, S. Joshi, L. Wang, J. Haller, M. Gado, J. Miller, U. Grenander, and M. Miller, "Hippocampal morphometry in schizophrenia via high dimensional brain mapping," in *Proc. Nat. Acad. Sci.*, 1998, pp. 11 406–11 411.
[6] F. L. Bookstein, "Size and shape spaces for landmark data in two dimensions (with discussion)," *Statistical Science*, vol. 1, no. 2, pp. 181–242, 1986.
[7] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," in *Proc. 5th Eur. Conf. Computer Vision*, 1998, pp. 484–498.
[8] A. Kelemen, G. Székely, and G. Gerig, "Elastic model-based segmentation of 3-D neuroradiological data sets," *IEEE Trans. Med. Imag.*, vol. 10, pp. 828–839, Oct. 1999.
[9] I. T. Jolliffe, *Principal Component Analysis*. Berlin, Germany: Springer-Verlag, 1986.
[10] C. Goodall, "Procrustes methods in the statistical analysis of shape," *J. Roy. Statist. Soc.*, vol. 53, no. 2, pp. 285–339, 1991.
[11] X. Pennec, "Probabilities and statistics on Riemannian manifolds: Basic tools for geometric measurements," presented at the *IEEE Workshop on Nonlinear Signal and Image Processing*, Antalaya, Turkey, 1999.
[12] K. V. Mardia, *Directional Statistics*. New York: Wiley, 1999.
[13] D. G. Kendall, "Shape manifolds, Procrustean metrics, and complex projective spaces," *Bull. Lond. Math. Soc.*, vol. 16, pp. 18–121, 1984.
[14] S. Joshi, S. Pizer, P. T. Fletcher, P. Yushkevich, A. Thall, and J. S. Marron, "Multiscale deformable model segmentation and statistical shape analysis using medial descriptions," *IEEE Trans. Med. Imag.*, vol. 21, pp. 538–550, May 2002.
[15] S. M. Pizer, P. T. Fletcher, S. Joshi, A. Thall, J. Z. Chen, Y. Fridman, D. S. Fritsch, A. G. Gash, J. M. Glotzer, M. R. Jiroutek, C. Lu, K. E. Muller, G. Tracton, P. Yushkevich, and E. L. Chaney, "Deformable m-reps for 3D medical image segmentation," *Int. J. Comput. Vis.*, vol. 55, no. 2–3, pp. 85–106, 2003.

[16] H. Blum and R. Nagel, "Shape description using weighted symmetric axis features," *Pattern Recognit.*, vol. 10, no. 3, pp. 167–180, 1978.

[17] P. T. Fletcher, C. Lu, and S. Joshi, "Statistics of shape via principal geodesic analysis on Lie groups," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2003, pp. 95–101.

[18] P. T. Fletcher, S. Joshi, C. Lu, and S. M. Pizer, "Gaussian distributions on Lie groups and their application to statistical shape analysis," in *Lecture Notes in Computer Science*. Berlin, Germany: Springer-Verlag, 2003, vol. 2732, Information Processing in Medical Imaging, pp. 450–462.

[19] A. Thall. (2002) Fast C^2 Interpolating Subdivision Surfaces Using Iterative Inversion of Stationary Subdivision Rules. Univ. North Carolina, Dept. Comput. Sci. Tech. Rep. [Online]. Available: http://midag.cs.unc.edu/pub/papers/Thall_TR02-001.pdf

[20] W. M. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry*, 2nd ed. New York: Academic, 1986.

[21] S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces*. New York: Academic, 1978.

[22] U. Grenander, M. I. Miller, and A. Srivastava, "Hilbert-Schmidt lower bounds for estimators on matrix Lie groups for ATR," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, pp. 790–802, Aug. 1998.

[23] A. Srivastava and E. Klassen, "Monte-Carlo extrinsic estimators of manifold-valued parameters," *IEEE Trans. Signal Processing*, vol. 50, pp. 299–308, Feb. 2002.

[24] M. Fréchet, "Les éléments aléatoires de nature quelconque dans un espace distancié," *Ann. Inst. H. Poincaré*, no. 10, pp. 215–310, 1948.

[25] H. Karcher, "Riemannian center of mass and mollifier smoothing," *Commun. Pure Appl. Math.*, vol. 30, no. 5, pp. 509–541, 1977.

[26] M. Moakher, "Means and averaging in the group of rotations," in *SIAM J. Matrix Anal. Applicat.*, vol. 24, 2002, pp. 1–16.

[27] W. S. Kendall, "Probability, convexity, and harmonic maps with small image—I: Uniqueness and fine existence," *Proc. Lond. Math. Soc.*, vol. 3, no. 61, pp. 371–406, 1990.

[28] S. R. Buss and J. P. Fillmore, "Spherical averages and applications to spherical splines and interpolation," in *ACM Trans. Graphics*, vol. 20, 2001, pp. 95–126.

[29] M. Styner and G. Gerig, "Medial models incorporating object variability for 3D shape analysis," in *Inform. Processing Med. Imag.*, 2001, pp. 502–516.

[30] T. Hastie and W. Stuetzle, "Principal curves," *J. Amer. Statist. Assoc.*, vol. 84, pp. 502–516, 1986.