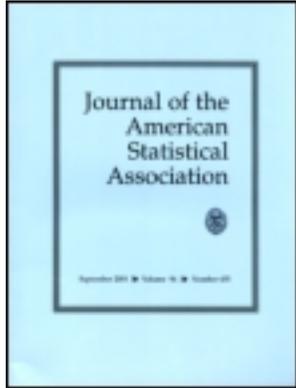


This article was downloaded by: [Ohio State University Libraries]

On: 04 February 2013, At: 11:51

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Journal of the American Statistical Association

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/uasa20>

Statistical Modeling of Curves Using Shapes and Related Features

Sebastian Kurtek^a, Anuj Srivastava^a, Eric Klassen^b & Zhaohua Ding^c

^a Department of Statistics, Florida State University, Tallahassee, FL, 32306

^b Department of Mathematics, Florida State University, Tallahassee, FL, 32306

^c Institute of Imaging Science, Vanderbilt University, Nashville, TN, 37232

Version of record first published: 08 Oct 2012.

To cite this article: Sebastian Kurtek, Anuj Srivastava, Eric Klassen & Zhaohua Ding (2012): Statistical Modeling of Curves Using Shapes and Related Features, Journal of the American Statistical Association, 107:499, 1152-1165

To link to this article: <http://dx.doi.org/10.1080/01621459.2012.699770>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Statistical Modeling of Curves Using Shapes and Related Features

Sebastian KURTEK, Anuj SRIVASTAVA, Eric KLASSEN, and Zhaohua DING

Motivated by the problems of analyzing protein backbones, diffusion tensor magnetic resonance imaging (DT-MRI) fiber tracts in the human brain, and other problems involving curves, in this study we present some statistical models of parameterized curves, in \mathbb{R}^3 , in terms of combinations of features such as shape, location, scale, and orientation. For each combination of interest, we identify a representation manifold, endow it with a Riemannian metric, and outline tools for computing sample statistics on these manifolds. An important characteristic of the chosen representations is that the ensuing comparison and modeling of curves is invariant to how the curves are parameterized. The nuisance variables, including parameterization, are removed by forming quotient spaces under appropriate group actions. In the case of shape analysis, the resulting spaces are quotient spaces of Hilbert spheres, and we derive certain wrapped truncated normal densities for capturing variability in observed curves. We demonstrate these models using both artificial data and real data involving DT-MRI fiber tracts from multiple subjects and protein backbones from the Shape Retrieval Contest of Non-rigid 3D Models (SHREC) 2010 database.

KEY WORDS: DT-MRI fibers; Path-straightening; Protein backbones; Shape analysis; Square-root function; Square-root velocity function.

1. PROBLEM BACKGROUND

Statistical analysis of image data employs the use of shape as an important characteristic for understanding objects in images. The prime examples of such applications come from computer vision and medical imaging, but these problems also appear in many other branches of science, including bioinformatics, geology, anthropology, biometrics, and paleontology. In these applications, one extracts objects from the data using different preprocessing techniques and studies the shapes of these objects using automated tools. While shape itself is important in many applications, there is also a need to study shape in conjunction with other features (Dryden and Mardia 1998). For example, in the studies of biological growth, it seems important to measure the overall size of the growing objects, in addition to their changing shapes. Similarly, in certain anatomical structures, it is important to take their relative locations and orientations into account while deciding on their normality or abnormality. If we consider a continuous curve, then all its physical characteristics can be summarized using *shape*, *scale*, *location*, and *orientation*. While it is convenient to work with parameterized curves, a parameterization is merely for the convenience of analysis and is not an intrinsic property of a curve, like the previous four properties. Our goal in this article is to develop a statistical Riemannian framework for analyzing curves that can incorporate some subset of these properties, depending upon the needs of an application. This allows for generating meaningful comparisons between subjects and populations as well as for performing classification. Furthermore, it provides tools for computation of

statistics such as the mean and covariance. To demonstrate these ideas in a concrete setting, we shall investigate the tasks of comparing, clustering, and classifying protein backbones as well as white-matter fibers in the human brain, obtained from diffusion tensor magnetic resonance imaging (DT-MRI) data.

1.1 Past Approaches

There is a rich literature on statistical analysis of shapes. Kendall (1984) defined shape as a characteristic of an object that is invariant to its rigid motions and global scaling. Since then, there has been a systematic development of tools, using ideas from geometry and statistics, to model and analyze shapes; some treatments of this theory include Dryden and Mardia (1998) and Small (1996). An important property of this theory was that objects under study were represented by *landmarks*, a finite collection of labeled points. These ordered points represent strategic locations along objects, most commonly along their boundaries, that roughly capture their shapes. Using the landmark representation of objects, different combinations of features have been studied. For example, Dryden and Mardia (1992) studied the joint shape and size features of objects. Very often, such as in medical image analysis, the landmarks are detected by human experts and the remaining analysis is performed by automated procedures. However, overabundance of digital data, especially image data, is prompting the need for a different kind of shape analysis. If one considers the boundaries of objects in images (two-dimensional, three-dimensional, or n -dimensional), these boundaries form curves and surfaces and this requires tools for shape analysis of curves and surfaces. Since these objects are infinite-dimensional, a different set of tools is needed.

There has been an increasing interest in shape analysis of continuous objects, especially curves. Consequently, there is now a significant amount of literature on shapes of continuous curves as elements of infinite-dimensional Riemannian manifolds.

Sebastian Kurtek is Ph.D. student, Department of Statistics, Florida State University, Tallahassee, FL 32306 (E-mail: skurtek@stat.fsu.edu). Anuj Srivastava is Professor, Department of Statistics, Florida State University, Tallahassee, FL 32306 (E-mail: anuj@stat.fsu.edu). Eric Klassen is Professor, Department of Mathematics, Florida State University, Tallahassee, FL 32306 (E-mail: klassen@math.fsu.edu). Zhaohua Ding is Professor, Institute of Imaging Science, Vanderbilt University, Nashville, TN 37232 (E-mail: zhaohua.ding@vanderbilt.edu). The authors thank Dr Adam W. Anderson for providing the DT-MRI data used in this work [National Institutes of Health (NIH) grant: R01NS058639]. This research was also supported by the grant ONR N00014-09-1-0664 and the National Science Foundation (NSF) grant DMS-0915003(AS).

Zahn and Roskies (1972) represented arc-length parameterized curves using angle functions and studied the angle functions using Fourier analysis. Younes (1998) defined shape spaces of *planar* curves in a Riemannian framework. Klassen et al. (2004) restricted to arc-length parameterized planar curves and derived numerical algorithms for computing geodesics between *closed* curves, the first ones to directly do so in the space of closed curves. Among other things, they applied this framework to statistical modeling and analysis using large databases of shapes (Srivastava et al. 2005). Michor and Mumford (2006) and Sundaramoorthi et al. (2011) studied several choices of Riemannian metrics in spaces of closed, planar curves for the purpose of comparing their shapes. Mio, Srivastava, and Joshi (2007) presented a family of elastic metrics that quantified the relative amounts of bending and stretching needed to deform shapes into each other. The difficulty with such an elastic shape analysis is that these metrics vary from point to point and only numerical methods are possible. More recently, Joshi et al. (2007) and Srivastava et al. (2011) presented a special representation of curves, called the square-root velocity function (SRVF), under which a specific element of the elastic family of metrics becomes an \mathbb{L}^2 metric and simplifies the shape analysis. This article elaborates on that idea and combines the shape feature with some other features to study different aspects of curves.

The main challenge in performing any analysis of parameterized curves is that the shape analysis should be invariant not only to rigid motions and global scalings, but also to their parameterizations. *A key idea propagated by this body of work is that the feature analysis of parameterized curves is invariant to how they are parameterized.* The choice of a shape representation and of a Riemannian metric are critically important for improved understanding, physical interpretations, and efficient computing. This article uses particularly convenient representations that enable simple physical interpretations of the resulting deformations. These representations are motivated by the well-known Fisher–Rao metric, used previously in information geometry, which imposes a Riemannian structure in the space of probability densities (Rao 1945; Amari 1985). Taking the positive square root of probability densities results in a simple Euclidean structure where geodesics, distances, and statistics are straightforward to compute (Bhattacharya 1943). A similar idea was introduced by Younes (1998) for curves only in \mathbb{R}^2 using complex analysis, and later used in Younes et al. (2008) for studying shapes of *planar closed* curves under an elastic metric.

1.2 Our Approach

The main idea in this approach is to use some specific representations of curves, termed *square-root functions* (SRFs) and *square-root velocity functions* (SRVFs). These representations are important for multiple reasons. First, and most important, a well-known Riemannian metric, called the elastic metric, becomes the standard \mathbb{L}^2 metric under the latter representation, which greatly simplifies the Riemannian analysis of curves. Second, under these representations and the \mathbb{L}^2 metric, any reparameterization of curves does not change the geodesic distance between them. Along with an optimization step on the reparameterization group, this results in the desired invariance to reparameterizations of the curves. Last, these representations are useful because of their versatility. The framework is easily altered to include other features of the curves, such as scale, orientation, and position, in the analysis. Preliminary results from this approach were presented in a conference paper (Mani et al. 2010). As a demonstration of these similarity measures, we will cluster neuronal fiber tracts generated from DT-MRI data based on different features and classify protein backbones based on their shape.

Given the ability to compute geodesics between curves using different features, under the chosen Riemannian metrics, we are able to define the first two moments, the mean and the covariance, in the spaces of such curves. This is quite convenient because given the two central moments, one can define a Gaussian-type distribution in these spaces. The basic idea is to define a truncated Gaussian distribution in a tangent space of the manifold and then map it back to the manifold. We provide a derivation of such distributions under the exponential map and the stereographic projection, and we use these explicit distributions for classification of protein backbones. Comparison and classification of protein structures is an important tool for understanding the evolutionary relationships between proteins and for predicting protein functions (Liu, Srivastava, and Zhang 2011). The main contributions of this article include:

(1) building upon previously presented theory on shape analysis of parameterized curves using the SRVF representation (Joshi et al. 2007; Srivastava et al. 2011) by incorporating other features such as scale and orientation in the analysis;

(2) introduction of the SRF representation for the situation where position is an important feature, and numerical algorithms for computing geodesics between curves under this representation; and

(3) derivation of analytical expressions for truncated Gaussian densities in feature spaces under the exponential map and the stereographic projection.

- (1) building upon previously presented theory on shape analysis of parameterized curves using the SRVF representation (Joshi et al. 2007; Srivastava et al. 2011) by incorporating other features such as scale and orientation in the analysis;
- (2) introduction of the SRF representation for the situation where position is an important feature, and numerical algorithms for computing geodesics between curves under this representation; and
- (3) derivation of analytical expressions for truncated Gaussian densities in feature spaces under the exponential map and the stereographic projection.

The rest of this article is organized as follows. In Section 2, we introduce two mathematical representations of curves: SRF and SRVF. In Section 3, we present results of computing geodesics between curves with different combinations of features. In Section 4, we present clustering results for DT-MRI brain fiber data. In Section 5, we discuss mean computations of curves and empirical analysis of curve data in tangent spaces. In Section 6, we provide explicit density functions in the spaces of curves using the exponential map and the stereographic projection. In that section, we also provide protein backbone classification results obtained using Gaussian-type models. We conclude the article with a short summary in Section 7.

2. REPRESENTATION OF CURVES

Consider the set of all smooth parameterized curves in \mathbb{R}^3 . The set of all rigid rotations of these curves is $\text{SO}(3)$, the special orthogonal group given by $\text{SO}(3) = \{\mathbf{O} \in \mathbb{R}^{3 \times 3} \mid \mathbf{O}^T \mathbf{O} = \mathbf{I}, \det(\mathbf{O}) = +1\}$. The set of all reparameterizations of curves is $\Gamma = \{\gamma : [0, 1] \rightarrow [0, 1] \mid \gamma(0) = 0, \gamma(1) = 1, \gamma \text{ is a diffeomorphism}\}$, while the sets of all possible scales

and translations are \mathbb{R}_+ and \mathbb{R}^3 , respectively. Let $\beta : [0, 1] \rightarrow \mathbb{R}^3$ be a parametrized curve, and let $\|\beta\| = \sqrt{\int_0^1 |\beta(t)|^2 dt}$ be its \mathbb{L}^2 norm. Note that the norm in the integrand is the standard Euclidean norm in \mathbb{R}^3 . The set of such square-integrable curves will be denoted by $\mathbb{L}^2([0, 1], \mathbb{R}^3)$, or simply \mathbb{L}^2 . For any $\gamma \in \Gamma$, the composition $(\beta \circ \gamma)(t) = \beta(\gamma(t))$ is a reparameterization of β . If β_1 and β_2 are two arbitrary elements of \mathbb{L}^2 , then we can compute the \mathbb{L}^2 distance between them using $\|\beta_1 - \beta_2\|$. What makes the analysis of curves using their coordinate representation difficult is the fact that, in general, $\|\beta_1 - \beta_2\| \neq \|\beta_1 \circ \gamma - \beta_2 \circ \gamma\|$. This inequality has deep repercussions, one of them being that the task of making the analysis invariant to reparameterizations of curves becomes very difficult. To address this issue, we propose two representations.

Definition 1.

- (1) Square-root function (SRF): Define the SRF to be a function $h : [0, 1] \rightarrow \mathbb{R}^3$ given by $h(t) = \sqrt{|\dot{\beta}(t)|}\beta(t)$, where β is a smooth parameterized curve. If the curve β is reparameterized by a $\gamma \in \Gamma$, its SRF changes to $h(t) \mapsto (h, \gamma)(t) \equiv h(\gamma(t))\sqrt{\dot{\gamma}(t)}$.
- (2) Square-root velocity function (SRVF) (Srivastava et al. 2011): Define a continuous mapping: $F : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ according to

$$F(v) \equiv \begin{cases} v/\sqrt{|v|} & \text{if } |v| \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

Then, define SRVF to be a function $q : [0, 1] \rightarrow \mathbb{R}^3$, where $q(t) \equiv F(\dot{\beta}(t)) = \dot{\beta}(t)/\sqrt{|\dot{\beta}(t)|}$, where β is a smooth parameterized curve. This representation includes those curves whose parameterization can become singular in the analysis. If the curve β is reparameterized by a $\gamma \in \Gamma$, its SRVF changes to $q(t) \mapsto (q, \gamma)(t) \equiv q(\gamma(t))\sqrt{\dot{\gamma}(t)}$.

It can be shown that if the curves are smooth, then the space of SRFs and SRVFs is a subset of \mathbb{L}^2 . Let the action of $\text{SO}(3)$ on \mathbb{L}^2 be given by $(\mathbf{O}, f)(t) = \mathbf{O}f(t)$ and the action of Γ be given by $(f, \gamma)(t) = f(\gamma(t))\sqrt{\dot{\gamma}(t)}$. Note that this action is different from $\beta(\gamma(t))$ mentioned above. Although we have not yet defined a metric in the space of SRFs and SRVFs, we will use the \mathbb{L}^2 metric. The discussion and motivation behind the choice of this metric are provided in later sections. Then, the following two results hold.

Lemma 1. The actions of Γ and $\text{SO}(3)$ on \mathbb{L}^2 commute.

Lemma 2. The actions of Γ and $\text{SO}(3)$ on \mathbb{L}^2 are by isometries and, therefore, $\|\mathbf{O}(f_1, \gamma) - \mathbf{O}(f_2, \gamma)\| = \|f_1 - f_2\|$ for all $f_1, f_2 \in \mathbb{L}^2, \gamma \in \Gamma$, and $\mathbf{O} \in \text{SO}(3)$.

The proofs of these lemmas can be found in Srivastava et al. (2011). Depending on the choice of features, we will use either SRF or SRVF for analyzing curves. It turns out that, for the purpose of shape analysis of curves, there are several reasons for choosing the SRVF over SRF. First, by the virtue of using $\dot{\beta}$ instead of β , the SRVF q is invariant to translation of β , which is useful in shape analysis. Second, for every $q \in \mathbb{L}^2$, there exists a curve β (unique up to a translation) such that the given q is the SRVF of that β . In fact, this curve can be obtained using

the equation $\beta(t) = \int_0^t q(s)|q(s)|ds$. The recovery of the curve β from its SRF h is not as simple since it requires solving a higher-order ordinary differential equation. Therefore, we use h only in situations where the position of curves is to be included in the analysis. The geodesic computations using SRVFs are straightforward and presented in Srivastava et al. (2011), but not so straightforward when using SRFs. Consequently, we will introduce a numerical technique to compute geodesic paths between curves when using the SRF representation.

2.1 Connections With Fisher–Rao Metric

The choice of SRVF as a representation is motivated in part by the work on the Fisher–Rao metric to form Riemannian manifolds of probability density functions (pdfs). We briefly illustrate these connections. Let \mathcal{P} represent the Banach manifold of all pdfs on $[0, 1]$. This is a manifold with a boundary since any density function whose value is zero for any $t \in [0, 1]$ is a boundary element. For a point p in the interior of \mathcal{P} , the tangent space $T_p(\mathcal{P})$ is given by $\{\delta p : [0, 1] \rightarrow \mathbb{R} \mid \int_0^1 \delta p(t)dt = 0\}$. For any $\delta p_1, \delta p_2 \in T_p(\mathcal{P})$, the Fisher–Rao metric is given by $\langle \delta p_1, \delta p_2 \rangle_p = \int_0^1 \delta p_1(t)\delta p_2(t)(1/p(t))dt$ (Rao 1945; Amari 1985; Vos and Kass 1997). An important property of this metric is that a reparameterization of densities does not change distances between them, that is, the action of reparameterization is by isometries (Čencov 1982). On the other hand, it is relatively difficult to compute geodesic paths under this metric since it differs from point to point. Bhattacharya (1943) showed that if we define $q(t) = +\sqrt{p(t)}$, then two things happen. First, the space of all such representations $\{q : [0, 1] \rightarrow \mathbb{R}_{\geq 0} \mid \int_0^1 |q(t)|^2 dt = 1\}$ is now an orthant in a unit sphere. Second, since $\delta q(t) = (1/(2\sqrt{p(t)}))\delta p(t)$, the Fisher–Rao metric becomes the standard \mathbb{L}^2 metric under this representation. That is, $\langle \delta q_1, \delta q_2 \rangle = \int_0^1 \delta q_1(t)\delta q_2(t)dt$. Under the \mathbb{L}^2 metric on a unit sphere, it is straightforward to compute geodesic paths and geodesic distances since they are simply arcs and arc lengths, respectively.

To see the connection with curves and their SRVF representations, consider smooth curves in $\mathbb{R}^3, \beta : [0, 1] \rightarrow \mathbb{R}^3$. The velocity vector at any point has two components: speed $p(t) = |\dot{\beta}(t)|$ and direction $\theta(t) = \dot{\beta}(t)/|\dot{\beta}(t)|$. Note that $\theta(t) \in \mathbb{S}^2$ for all t . What is the extension of the Fisher–Rao metric to this higher-dimensional case? Mio, Srivastava, and Joshi (2007) provided an answer that they called an *elastic metric*:

$$\begin{aligned} & \langle (\delta p_1, \delta \theta_1), (\delta p_2, \delta \theta_2) \rangle_{(p, \theta)} \\ &= a \int_0^1 \delta p_1(t)\delta p_2(t)(1/p(t))dt \\ &+ b \int_0^1 \langle \delta \theta_1(t), \delta \theta_2(t) \rangle p(t)dt \text{ for } a, b > 0. \end{aligned} \tag{1}$$

Here, the first term measures the variations in the speed function (it is identical to the Fisher–Rao metric), while the second term measures the variations in the direction function. The constants a and b are positive and provide relative weights to the two terms. It was shown by Mio, Srivastava, and Joshi (2007) that this metric is invariant to reparameterizations of curves and is particularly useful in shape analysis. To make a connection with SRVFs, we can rewrite $q(t) = \sqrt{p(t)}\theta(t)$ and its variation $\delta q(t) = (\delta p(t)/(2\sqrt{p(t)}))\theta(t) + \sqrt{p(t)}\delta \theta(t)$. If we compute the

\mathbb{L}^2 metric between such variations, we get

$$\langle \delta q_1, \delta q_2 \rangle_{(p,\theta)} = (1/4) \int_0^1 \delta p_1(t) \delta p_2(t) \langle \theta(t), \theta(t) \rangle (1/p(t)) dt + \int_0^1 \langle \delta \theta_1(t), \delta \theta_2(t) \rangle p(t) dt, \quad (2)$$

which is identical to the elastic metric [Equation (1)] for $a = 1/4$ and $b = 1$. This analysis shows that the SRVF is analogous to the square-root representation used by Bhattacharya, for simplifying the Fisher–Rao metric in the space of pdfs, but extended to handle curves in \mathbb{R}^n , and in particular \mathbb{R}^3 .

2.2 Unit Length Curves and Hilbert Hypersphere

So far, we have identified \mathbb{L}^2 as the representation space for SRFs and SRVFs. If we impose additional constraints on the curves, such as unit length, then the representation space naturally restricts. In the case of SRFs, it is difficult to make such restrictions explicit, so we continue to use the full space \mathbb{L}^2 there. However, in the case of SRVFs, we can gain by using constraints explicitly. If a curve β is of length 1, then $\int_0^1 |\dot{\beta}(t)| dt = \int_0^1 |q(t)|^2 dt = 1$. In other words, the \mathbb{L}^2 norm of q is 1. The set of all such SRVFs is a unit hypersphere in \mathbb{L}^2 : $\mathbb{S}_\infty = \{q : [0, 1] \rightarrow \mathbb{R}^3 \mid \int_0^1 |q(t)|^2 dt = 1\}$; it is also a Hilbert manifold (Lang 1999).

Since \mathbb{L}^2 is a vector space, its tangent space $T_f(\mathbb{L}^2) = \mathbb{L}^2$ for all $f \in \mathbb{L}^2$. The \mathbb{L}^2 Riemannian metric is given by: for any $v_1, v_2 \in \mathbb{L}^2$, we have $\langle v_1, v_2 \rangle = \int_0^1 \langle v_1(t), v_2(t) \rangle dt$, where the inner product in the integrand is the Euclidean inner product in \mathbb{R}^3 . For any $f_1, f_2 \in \mathbb{L}^2$, the geodesic path connecting them is given by a “straight line”:

$$\psi_\tau^l(f_1, f_2) = (1 - \tau)f_1 + \tau f_2, \quad \tau \in \mathbb{R}. \quad (3)$$

A geodesic on \mathbb{L}^2 can also be characterized in terms of a “shooting direction” $v \in \mathbb{L}^2$ as $\psi_\tau(f_1, v) = f_1 + \tau v$. For any $v \in \mathbb{L}^2$, the exponential map at $f_1 \in \mathbb{L}^2$, $\exp : \mathbb{L}^2 \mapsto \mathbb{L}^2$ is given by $\exp_{f_1}(v) = f_1 + v$. Also, for any $f_2 \in \mathbb{L}^2$, the inverse of the exponential map, denoted by $\exp_{f_1}^{-1} : \mathbb{L}^2 \mapsto \mathbb{L}^2$, is computed as $\exp_{f_1}^{-1}(f_2) = f_2 - f_1$. The geodesic distance between any two points is given by the \mathbb{L}^2 norm of their difference: $d(f_1, f_2) = \|f_1 - f_2\|$.

A Riemannian structure can be inherited on \mathbb{S}_∞ in a straightforward fashion. For a $f \in \mathbb{S}_\infty$, the tangent space at f is given by $T_f(\mathbb{S}_\infty) = \{v \in \mathbb{L}^2 \mid \langle v, f \rangle = 0\}$, that is, the tangents to a sphere at a point are orthogonal to the direction representing that point. If we impose the \mathbb{L}^2 metric in the tangent spaces of \mathbb{S}_∞ , we get a Riemannian structure. The realization that \mathbb{S}_∞ is a sphere is important because the differential geometry of a sphere is well known and the geodesics can be written analytically. For any $f_1, f_2 \in \mathbb{S}_\infty$ (such that f_1 and f_2 are not antipodal), the geodesic path between them in \mathbb{S}_∞ , for $\tau \in [0, 1]$, is given by

$$\psi_\tau^s(f_1, f_2) = (1/\sin(\theta))[\sin(\theta - \tau\theta)f_1 + \sin(\tau\theta)f_2], \quad \theta = \cos^{-1}(\langle f_1, f_2 \rangle). \quad (4)$$

The value θ is actually the geodesic distance between them under the \mathbb{L}^2 Riemannian metric. A geodesic on \mathbb{S}_∞ can also be characterized in terms of a tangent direction $v \in T_{f_1}(\mathbb{S}_\infty)$ as $\psi_\tau(f_1, v) = \cos(\tau\|v\|)f_1 + \sin(\tau\|v\|)(v/\|v\|)$. For any $v \in T_{f_1}(\mathbb{S}_\infty)$, the exponential map at $f_1 \in \mathbb{S}_\infty$, $\exp : T_{f_1}(\mathbb{S}_\infty) \mapsto \mathbb{S}_\infty$, is given by $\exp_{f_1}(v) = \cos(\|v\|)f_1 + \sin(\|v\|)(v/\|v\|)$. Also, the inverse of the exponential map, denoted by $\exp_{f_1}^{-1} : \mathbb{S}_\infty \mapsto T_{f_1}(\mathbb{S}_\infty)$, is given by $\exp_{f_1}^{-1}(f_2) = (\theta/\sin(\theta))(f_2 - \cos(\theta)f_1)$, where $\theta = \cos^{-1}(\langle f_1, f_2 \rangle)$.

3. GEODESICS ON FEATURE SPACES

Under this setup, we are going to discuss a number of scenarios, each corresponding to a certain combination of physical properties of the curves. We remind the reader that we are interested in four physical properties of curves: shape, size, location, and orientation, and all of them are independent of parameterization. By using SRVFs, we are already invariant to locations and, thus, are focusing on the remaining three properties.

Using the SRVF representation, we can study four feature spaces: \mathcal{S}_1 (shape + orientation + scale), \mathcal{S}_2 (shape + scale), \mathcal{S}_3 (shape + orientation), and \mathcal{S}_4 (shape). The corresponding spaces and metrics for each of these cases are summarized in Table 1. For the sake of brevity, we will describe the case of \mathcal{S}_3 only. The application of this theory to the other cases is similar.

In case we are interested in analyzing the shapes and orientations of curves, the curves are rescaled to have unit length and thus lie on \mathbb{S}_∞ . To remove the additional variability of

Table 1. Description of different feature spaces for the SRVF representation

Features	Shape + orientation + scale	Shape + scale
Prespace	\mathbb{L}^2	\mathbb{L}^2
Equivalence classes	$[q] = \text{closure}\{(q, \gamma) \mid \gamma \in \Gamma\}$	$[q] = \text{closure}\{\mathbf{O}(q, \gamma) \mid \mathbf{O} \in \text{SO}(3), \gamma \in \Gamma\}$
Feature space	$\mathcal{S}_1 = \{[q] \mid q \in \mathbb{L}^2\}$	$\mathcal{S}_2 = \{[q] \mid q \in \mathbb{L}^2\}$
Distances	$d_1([q_1], [q_2]) = \min_{\gamma \in \Gamma} \ q_1 - (q_2, \gamma)\ $	$d_2([q_1], [q_2]) = \min_{\mathbf{O} \in \text{SO}(3), \gamma \in \Gamma} \ q_1 - \mathbf{O}(q_2, \gamma)\ $
Geodesics	Equation (3) between q_1 and q_2^*	Equation (3) between q_1 and q_2^*
Features	Shape + orientation	Shape
Prespace	\mathbb{S}_∞	\mathbb{S}_∞
Equivalence classes	$[q] = \text{closure}\{(q, \gamma) \mid \gamma \in \Gamma\}$	$[q] = \text{closure}\{\mathbf{O}(q, \gamma) \mid \mathbf{O} \in \text{SO}(3), \gamma \in \Gamma\}$
Feature space	$\mathcal{S}_3 = \{[q] \mid q \in \mathbb{S}_\infty\}$	$\mathcal{S}_4 = \{[q] \mid q \in \mathbb{S}_\infty\}$
Distances	$d_3([q_1], [q_2]) = \min_{\gamma \in \Gamma} \cos^{-1}(\langle q_1, (q_2, \gamma) \rangle)$	$d_4([q_1], [q_2]) = \min_{\mathbf{O} \in \text{SO}(3), \gamma \in \Gamma} \cos^{-1}(\langle q_1, \mathbf{O}(q_2, \gamma) \rangle)$
Geodesics	Equation (4) between q_1 and q_2^*	Equation (4) between q_1 and q_2^*

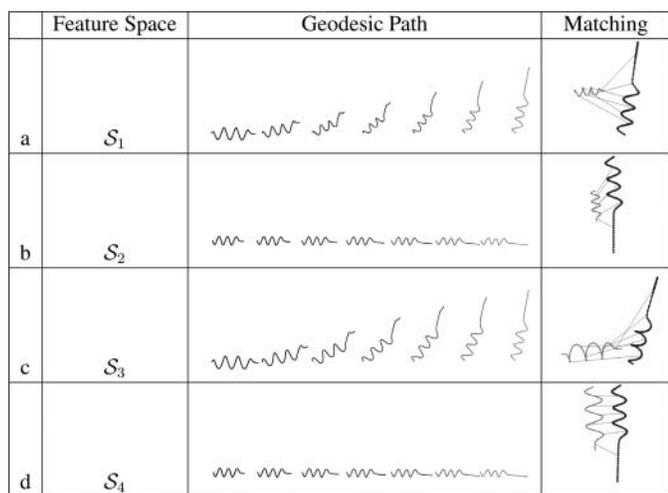


Figure 1. An example of geodesic paths in different feature spaces.

reparameterization, we use the notion of equivalence classes. We would like to form the quotient space as the set of all equivalence classes in the prespace under the action of Γ . Recall that this action is given by $(q, \gamma)(t) \equiv q(\gamma(t))\sqrt{\dot{\gamma}(t)}$. The trouble is that although this group action is by isometries, the orbits are not closed. The reason for this is that the space of diffeomorphisms is not closed with respect to the \mathbb{L}^2 metric, since a sequence of diffeomorphisms might approach a map that is not a diffeomorphism under this metric. To resolve this theoretical difficulty, we propose that instead of using the orbits, we use their closures. This issue is described in more detail in Srivastava et al. (2011). The shape + orientation space is defined as the set of all closed orbits, and we can compute the distance in that space using a minimization problem over Γ , which is solved with the dynamic programming algorithm (the Appendix). We refer to the optimal reparameterization as γ^* and to (q_2, γ^*) as q_2^* . Figure 1 presents some examples of geodesics between artificial helical curves using the SRVF representation and different feature spaces. The artificial curves, which are used for demonstration throughout the article, were formed in the following manner $\beta = (R \cos(2\pi t \eta\rho), R \sin(2\pi t \eta\rho), 2\pi t)^T$, $t \in [0, 1]$, where R , η , and ρ control the radius, number, and length of the spirals, respectively.

When we need to include the position feature in the analysis, we will use the SRF representation. While this representation is quite convenient for including position in the feature set, the main drawback is that it is not straightforward to reconstruct a

curve from its SRF. Thus, the computation of geodesic paths is not simple as in the SRVF cases. For this purpose, we use a numerical method called path-straightening, which was previously introduced to compute geodesic paths in the space of closed curves (Klassen and Srivastava 2006; Srivastava et al. 2011). In that case, the path-straightening algorithm was derived for computing geodesics in a complicated space under a standard metric. In this article, we present this algorithm for computing geodesics in a vector space under a complicated metric. The general idea is to endow the space of parameterized curves with a Riemannian metric, which is the pullback of the \mathbb{L}^2 metric from the space of SRFs. Then, it is possible to write down an energy under this metric, which is used to find geodesic paths in the space of curves. We present some more details on the derivation in the Appendix. The space and metric for the SRF representation are summarized in Table 2.

Next, we show some geodesics between curves under the SRF representation. First, we show a few examples that verify that the path-straightening algorithm works correctly. In each of these examples, we initialize the path, with the endpoints being the same curve ($F(0) = F(1) = \beta$), and interpolate the interior points. Thus, the geodesic path in each of these cases should be a constant path ($F(t) = \beta$ for all t) and the energy of this path should be zero. We note that the curves here are allowed to have different shapes, scales, orientations, and positions. The results are displayed in Figure 2. In both examples, the resulting geodesic paths are visibly constant.

The next set of examples, displayed in Figure 3, show some geodesics computed in S_5 (shape + orientation + scale + position). We note that the features are nicely preserved along these geodesics. This suggests that finding optimal reparameterizations (i.e., removing parameterization variability) plays an important role in comparisons of curves, and improved matching between curves yields a more accurate measure of their differences.

4. CLUSTERING OF DT-MRI BRAIN FIBERS

An important problem in the analysis of curves is their clustering by taking different features under consideration. Since we have defined geodesic distances between curves under different feature sets, we can use these distances for clustering. In this section, we present some clustering results using DT-MRI brain fiber data.

A DT-MRI scan of a brain generates a field of 3×3 tensor matrices that describe the constraints on local Brownian

Table 2. Description of the shape + orientation + scale + position feature space for the SRF representation

Prespace	\mathcal{B}
Metric	$\langle\langle p, m \rangle\rangle_\beta = \langle dh_\beta(p), dh_\beta(m) \rangle_{\mathbb{L}^2}$
Equivalence classes	$[\beta] = \text{closure}\{\beta \circ \gamma \mid \gamma \in \Gamma\}$
Feature space	$S_5 = \{[\beta] \mid \beta \in \mathcal{B}\}$
Distances	$d_s([\beta_1], [\beta_2]) = \min_{\gamma \in \Gamma} \left(\min_{F: [0,1] \rightarrow \mathcal{B}} \left(\int_0^1 \langle\langle F_t, F_t \rangle\rangle_F^{(1/2)} dt \right) \right)$ $F(0) = \beta_1$ $F(1) = (\beta_2 \circ \gamma)$
Geodesics	Path-straightening

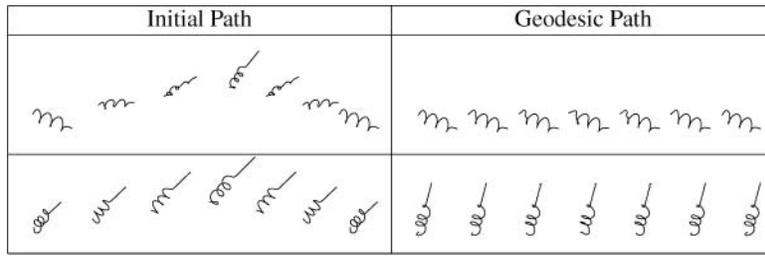


Figure 2. Geodesic paths using path-straightening.

motion of water molecules. Tractography is the primary technique for finding fiber tracts; it uses the principle that the principal diffusion directions coincide with the local tangent directions of fibrous tissues. Integration of the principal diffusion directions permits entire fiber tracts to be delineated. The brain fibers used in this article are a part of the human language circuit that connects Broca’s and Wernicke’s areas in the left hemisphere. Previous studies have shown that there are two major connection routes, ventral and dorsal, respectively, between these two regions (Morgan et al. 2009). Although these two routes are widely recognized, connecting patterns between the two language areas are more complicated. Some studies have demonstrated that there can be a third connection route, while others have suggested that these routes can be further partitioned into finer clusters that connect different subregions of the language areas. It should be pointed out that this clustering is solely data-driven—whether or not it represents the true physiological structure remain unknown. One way to validate this is to correlate the clustering results with functional magnetic resonance imaging (fMRI) data, which can be used to study functional connections between the language areas. While validations with fMRI appear very appealing due to the unanimously accepted notion of structure–function coupling, it is certainly beyond the scope of the current article and thus is not included.

The datasets considered here consist of 388 fibers from four subjects: subject 1 has 176 total fibers, subject 2 has 68 total fibers, subject 3 has 48 total fibers, and subject 4 has 88 total fibers. Since we consider the locations of fibers as an important feature, we use SRFs (and distance d_5) to compute pairwise distances between them. We use a hierarchical clustering method with a complete linkage function (the distance between two clusters corresponds to the maximum distance between cluster members). As described in the previous paragraph, there are two or three connection routes between the Broca’s and Wernicke’s regions. Thus, we began by separating each subject’s set of fibers into three clusters. The results are shown in Figure 4. We see a clear separation of clusters based on distances, which take into account shape, translation, scale, and orientation. To enhance the display of clusters, we have used the multidimensional scaling

(MDS) method to obtain two-dimensional coordinates of each fiber and display them as a scatterplot. The distance matrices and the MDS plots for all of the subjects indicate that there are three main fiber clusters present. Visual inspection indicates that this is in fact the case, and we see that fiber position plays a very important role in these examples. Looking at the plotted fibers and distance matrix for subject 1, the separation of two of the clusters is not very clear. It may be possible to combine the blue and red clusters in this case. All of the clusters for subject 2 contain curves with very different shapes. In the case of subject 3, the blue and red clusters differ most in shape, while the green and the red/blue clusters differ a lot in their position. It is reasonable that the red and blue clusters are very closely related. In the case of subject 4, the curves in the red and blue clusters have similar shapes but very different lengths and positions.

5. MEAN COMPUTATION AND DATA ANALYSIS

In this section, we develop tools for statistical analysis of curve data. In particular, we define and compute means of curves under different features and study the sample sets on tangent bundles using principal component analysis (PCA). More ambitiously, we are interested in capturing variability associated with curves within object classes using probability models. This means that we want to learn the model from a sample or, in the case of parametric models, to estimate the parameters of the model from a sample. Because the computation of geodesics (required for computation of statistics) between SRFs is not a straightforward task, this section presents derivations and results for SRVFs only.

5.1 Mean Computation and Examples

We start by defining and estimating the first two central moments: the Karcher mean \bar{q} and the covariance in the tangent space at the Karcher mean in the feature space of interest. Let $\{\beta_1, \beta_2, \dots, \beta_n\}$ be a given collection of curves with SRFVs $\{q_1, q_2, \dots, q_n\}$. Considering them as elements of the space \mathbb{L}^2 , we can compute their extrinsic average as $\bar{q}_{\text{ext}} = (1/n) \sum_{i=1}^n q_i$ and project them into the appropriate preimage, if necessary.

The sample Karcher mean is given by: $\bar{q} = \operatorname{argmin}_{[q] \in \mathcal{S}} \sum_{i=1}^n d([q], [q_i])^2$. Here, d denotes the geodesic distance in the appropriate quotient space \mathcal{S} under the elastic metric. A gradient-based approach for finding the Karcher mean is given in several places (Klassen et al. 2004; Le 2001), and is repeated here for convenience. This iteration is based on the fact that the gradient of the geodesic distance squared

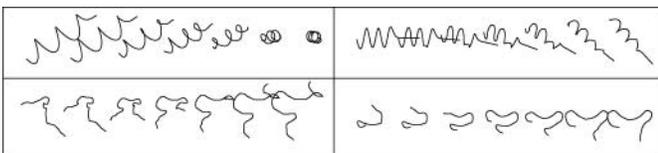


Figure 3. Computation of geodesic paths in the feature space \mathcal{S}_5 .

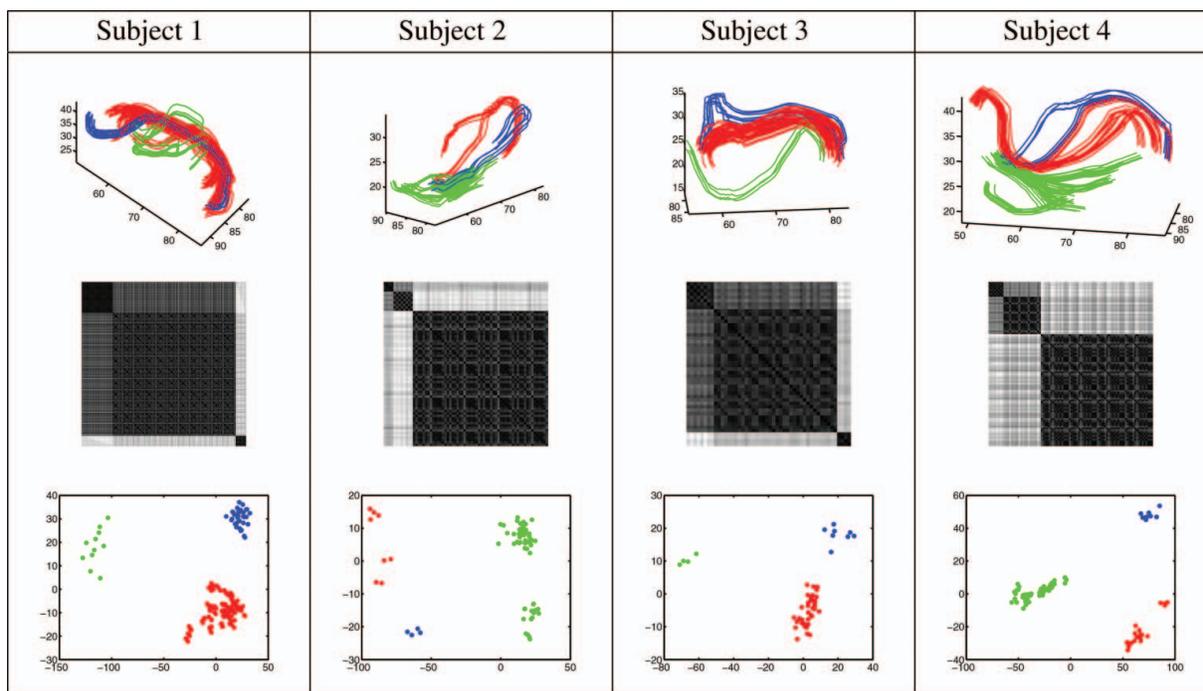


Figure 4. Top: cluster of fibers. Middle: d_5 distance matrix between fibers. Bottom: two-dimensional MDS plot (different colors denote different fiber clusters). The online version of this figure is in color.

between any two points, with respect to one of the points, is the inverse exponential map of the other at the first point. The exponential and inverse exponential maps for different spaces are given in Section 2.2.

Algorithm 1 Karcher mean on M .

- Let \bar{q}_0 be an initial estimate of the Karcher mean. Set $j = 0$.
1. For each $i = 1, \dots, n$, compute $v_i = \exp_{\bar{q}_j}^{-1}(q_i)$.
 2. Compute the average direction $\bar{v} = (1/n) \sum_{i=1}^n v_i$.
 3. If $\|\bar{v}\|$ is small, then stop. Else, update using $\bar{q}_{j+1} = \exp_{\bar{q}_j}(\epsilon \bar{v})$, $\epsilon = 0.3$.
 4. Set $j = j + 1$ and return to Step 1

Figure 5 displays some experimental results (helical curves) from mean computation for artificial data. The intrinsic means (sample means obtained using the intrinsic method) appear to be good representatives of the data, since they all have three loops.

In addition, for this artificial example, it is useful to compare our method with the standard landmark method. Figure 5 also shows extrinsic means for this set of curves. We note that although the shape and shape + scale extrinsic means are roughly representative of the data given, they are not as accurate in describing the actual variability in the data as the means generated using our methodology. On the other hand, the shape + orientation and shape + orientation + scale means are not even representative of the data. This shows a clear advantage in using our elastic intrinsic method in statistical modeling of curves.

In Figure 6, we display Karcher mean computation results for brain fiber data. These data consist of a subset of 46 DT-MRI fiber curves from the corpus callosum, a region of the human brain. In clinical applications, it may be of interest to build mean templates based on fibers in different populations. Thus, here, we consider computing averages of brain fibers based on different features. We observe similar results as when we used artificial data. The shape only (\mathcal{S}_4) and shape + scale (\mathcal{S}_2) means have a very similar shape but very different lengths.

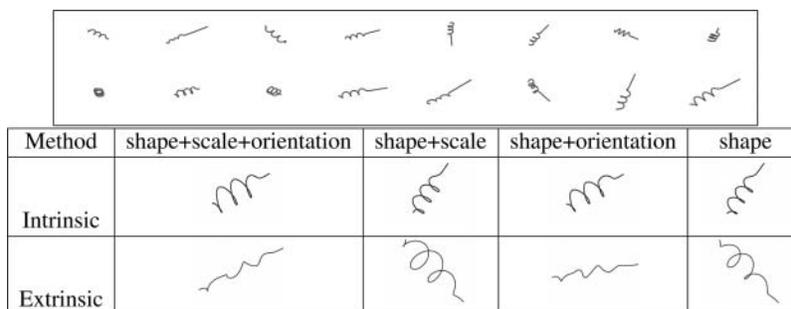


Figure 5. Top: artificial helical curves. Bottom: sample means obtained using the intrinsic method and the extrinsic method via landmark representation. (Note: scales altered for improved displays.)

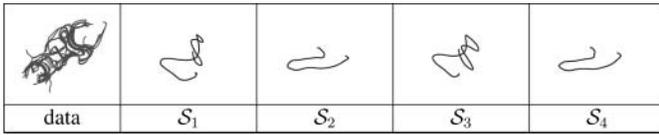


Figure 6. Mean curves for a cluster of fiber tracts. (Note: scales altered for improved displays.)

Both of these averages have a horseshoe shape, which is a clear representative of the fibers in the given data. We caution the reader that they have been scaled in the figure for an improved display of their shapes. Also, the shape + orientation (S_3) and shape + orientation + scale (S_1) mean curves are very similar in shape and orientation and only vary in scale.

5.2 Computation of Karcher Covariance

Once the sample Karcher mean has been computed, the evaluation of the Karcher covariance is as follows. Let $v_i = \exp_{\bar{q}}^{-1}(q_i)$, $i = 1, 2, \dots, n$, $v_i \in T_{\bar{q}}(\mathcal{S})$. Then, the covariance kernel can be defined as a function $K_q : [0, 1] \times [0, 1] \rightarrow \mathbb{R}$, given by $K_q(\omega, \tau) = (1/(n - 1)) \sum_{i=1}^n \langle v_i(\omega), v_i(\tau) \rangle$. Note that the trace of this covariance function is proportional to the Karcher variance, that is, $\int_0^1 K_q(\tau, \tau) d\tau \propto \sum_{i=1}^n d(\bar{q}, q_i)^2$. In practice, since the curves have to be sampled with a finite number of points, say m , the resulting covariance matrices are finite-dimensional. Often, the observation size n is much less than m and, consequently, n controls the degree of variability in the stochastic model. In the case of learning shape models from the observations, one can reach an efficient basis for $T_{\bar{q}}(\mathcal{S})$ using the traditional PCA as follows. Let $V \in \mathbb{R}^{3m \times n}$ be the observed tangent data matrix with n observations and m sample points in \mathbb{R}^3 on each tangent. Let $K \in \mathbb{R}^{3m \times 3m}$ be the Karcher covariance matrix, and let $K = U \Sigma U^T$ be its singular value decomposition (SVD). The submatrix formed by the first r columns of U , call it U_1 , spans the principal subspace of the observed data and provides the observations of the principal coefficients as $C = U_1^T V \in \mathbb{R}^{r \times 3m}$. (The choices of r and m depend on the application of interest.)

We can visualize the variations in curves along dominant directions in $T_{\bar{q}}(\mathcal{S}_3)$ or $T_{\bar{q}}(\mathcal{S}_4)$ by computing straight lines along these directions and projecting them on \mathcal{S}_3 or \mathcal{S}_4 using the exponential map. We do this for the data displayed in Figures 5

Table 3. Percentage of variation explained by first 10 principal components for (a) artificial data and (b) brain fiber data

	Shape	Shape + orientation	Shape + scale	Shape + orientation + scale
(a)	83.2	86.2	94.9	91.4
(b)	74.8	79.3	78.5	82.1

and 6. Figure 7 shows the top three directions (displayed along the rows) by rendering curves along the path $\exp_{\bar{q}}(\tau v)$ for τ from -1 to $+1$, where $v \doteq \sqrt{\Sigma_{ii}} U_i$. The middle curve, with $\tau = 0$, is obviously the mean shape \bar{q} . In Table 3, we report the percentage of variation that was explained by the first 10 principal components in each case.

6. EXPLICIT MODELS ON SHAPE SPACES

Our next goal is to present some explicit probability models in spaces of curves for use in future statistical analyses. There are several types of models that can be developed. One distinction is between parametric and nonparametric models, an example of the latter being the kernel density estimators on manifolds (Pelletier 2005). From a computational point of view, at least for the purposes of sampling from the distributions and using Monte Carlo estimators, it seems more efficient to have parametric models. A second distinction is between intrinsic and extrinsic models. The former restricts the analysis completely to the underlying manifold, while the latter embeds the manifold in a larger vector space and imposes models in the larger space. We have not used the extrinsic models in our problem since it is difficult to find appropriate embeddings for the quotient spaces of the kind we are dealing with. Therefore, we are interested in parametric, intrinsic probability models for curves that result in analytical expressions for the pdfs. Some examples of parametric probability models for spherical manifolds were presented by Mardia and Jupp (2000). Our approach will be to define densities in principal subspaces of a tangent space and then map them back to the shape space. In the case of a finite-dimensional sphere, there are at least two interesting mappings from a tangent plane to the sphere: the exponential map and the stereographic projection, as shown in Figure 8. In this

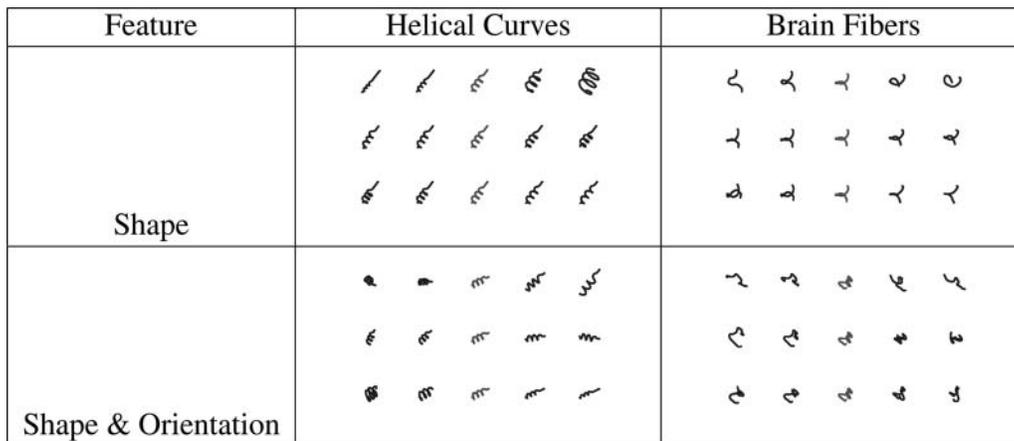


Figure 7. Principal modes of variation with the Karcher mean in the center.

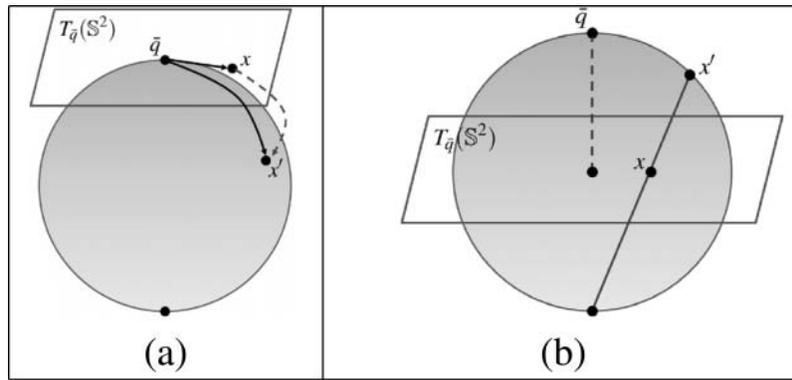


Figure 8. Map from $T_{\bar{q}}(\mathbb{S}^2)$ to \mathbb{S}^2 for wrapped densities: exponential (a) and stereographic (b).

article, we explore the use of both of these mappings in inducing densities on a sphere.

6.1 Wrapped Gaussian Distribution

In case our goal is only to sample from a probability model and we do not need an explicit density function, we can do so using the Gaussian density in the tangent space and the exponential map. Let $\mathbf{K} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^T$ be the SVD of the sample covariance matrix, as earlier, where $\mathbf{\Sigma}$ is a diagonal matrix of singular values and \mathbf{U} is an orthogonal matrix, with each column representing an eigenvector of \mathbf{K} . Then, a multivariate Gaussian model for the tangent vector rearranged as a long vector $\mathbf{v} \in \mathbb{R}^{3m}$ is given by $\mathbf{v} = \sum_{i=1}^n z_i \sqrt{\Sigma_{ii}} \mathbf{U}_i$, where $z_i \sim N(0, 1)$ iid. One can restructure the elements of \mathbf{v} to form a matrix of size $\mathbb{R}^{3 \times m}$ and approximate an element of $T_{\bar{q}}(\mathcal{S})$; call it v . This random v can then be projected in the shape space using the exponential map $v \mapsto \exp_{\bar{q}}(v)$ to obtain a random curve. This provides a technique for sampling from the wrapped Gaussian models on \mathcal{S}_3 and \mathcal{S}_4 . Figure 9 displays sampling results from the wrapped Gaussian distribution on the shape and shape + orientation spaces. We note that all of the samples seem valid. That is, the computed models correctly summarize the variability in the given data.

In case we are interested in reaching analytical expressions for densities in quotient spaces, we need to introduce some changes in this model. Although our spaces are infinite-dimensional, we will start with the finite-dimensional sphere, \mathbb{S}^k .

6.2 Truncated Gaussian Under the Exponential Map

We start with the exponential map for projecting points in a tangent space $T_{\bar{q}}(\mathbb{S}^k)$ to \mathbb{S}^k as follows: $\exp_{\bar{q}} : T_{\bar{q}}(\mathbb{S}^k) \rightarrow \mathbb{S}^k$ is given by: for any $v \in T_{\bar{q}}(\mathbb{S}^k)$, $\exp_{\bar{q}}(v) = \cos(\|v\|)\bar{q} +$

$(\sin(\|v\|)/\|v\|)v$. To transfer the density from $T_{\bar{q}}(\mathbb{S}^k)$ to \mathbb{S}^k , we need an invertible map. To make the exponential map invertible, we need to restrict the density to a subset of $T_{\bar{q}}(\mathbb{S}^k)$ using truncation. While choosing a sphere of radius π around the origin will be sufficient for this purpose, this will lead to a singularity in the resulting density at the point antipodal to the mean. To avoid this, we will truncate the density at a smaller distance, say $\pi/2$, from the origin in the tangent space. In the case of \mathcal{S}_4 (shape), it can be shown that the upper bound on all distances is $\pi/2$, which implies that this is a natural restriction of the pdf for this feature space. Thus, $\exp_{\bar{q}}$ becomes invertible in this domain and its inverse is given by: for $p \in \mathbb{S}^k$, $\exp_{\bar{q}}^{-1}(p) = (\theta/\sin(\theta))(p - \bar{q}\cos(\theta))$, where $\theta = \cos^{-1}(\langle p, \bar{q} \rangle)$ and where p and \bar{q} are viewed as vectors in \mathbb{R}^{k+1} . To induce the truncated Gaussian density on a sphere, using the exponential map, we need to compute the determinant of the Jacobian of this map. As a first step, we will derive the induced density for $k = 2$ and then extend it to a general k .

Let $\bar{q} \in \mathbb{S}^2$ be the mean of the induced density, and let w_1, w_2 form an orthonormal basis of $T_{\bar{q}}(\mathbb{S}^2)$. The set $\{\bar{q}, w_1, w_2\}$ forms an orthogonal basis for \mathbb{R}^3 . Now, we can identify any element v of $T_{\bar{q}}(\mathbb{S}^2)$ with its coordinates $x = (x_1, x_2) \in \mathbb{R}^2$ such that $v = x_1 w_1 + x_2 w_2$. Define a truncated bivariate normal density on $T_{\bar{q}}(\mathbb{S}^2)$ using its identification with \mathbb{R}^2 : $f(x) = (1/Z_2)e^{-(1/2)x^T K^{-1}x} \mathbf{1}_{\|x\| \leq \pi/2}(x)$, where Z_2 is simply the normalizing constant. Next, we map this density onto \mathbb{S}^2 using the exponential map so that the origin of the tangent space coincides with the mean \bar{q} . For a $\theta \in \mathbb{R}$, and the point θw_1 in $T_{\bar{q}}(\mathbb{S}^2)$, the exponential map equation becomes $\theta w_1 \mapsto \exp_{\bar{q}}(\theta w_1) \equiv \cos(\theta)\bar{q} + \sin(\theta)w_1$. Let this point on \mathbb{S}^2 be called p . We need to establish an orthogonal basis for the tangent space $T_p(\mathbb{S}^2)$, and

Shape			Shape+Orientation		
Helical Curves	Brain Fibers	Helical Curves	Brain Fibers		

Figure 9. Samples from the wrapped Gaussian distribution.

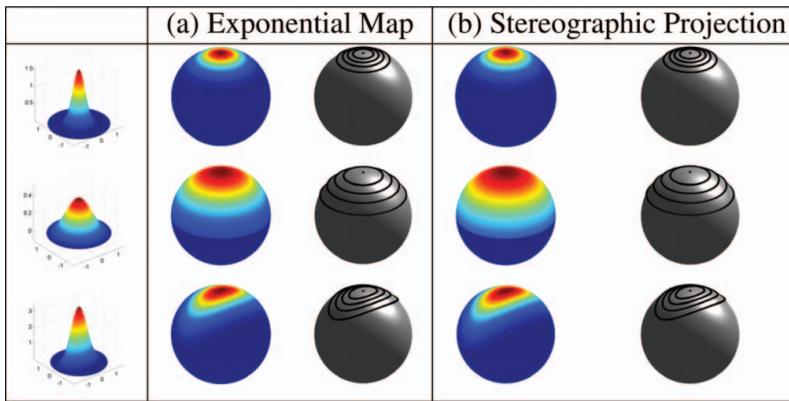


Figure 10. The first two examples (rows) show isotropic densities for small and large values of σ . The third example (row) shows a nonisotropic density. The online version of this figure is in color.

the vectors $\mathbf{b}_1 = (-\sin(\theta)\bar{\mathbf{q}} + \cos(\theta)\mathbf{w}_1)$ and $\mathbf{b}_2 = \mathbf{w}_2$ provide a convenient orthonormal basis. To derive the differential of the map $\exp_{\bar{\mathbf{q}}}$, denoted $\exp_{\bar{\mathbf{q}}^*}$, we take each of the basis elements of $T_{\bar{\mathbf{q}}}(\mathbb{S}^2)$ and map them to $T_p(\mathbb{S}^2)$ under $\exp_{\bar{\mathbf{q}}^*}$. The resulting determinant of the Jacobian matrix turns out to be $(\sin(\theta)/\theta)$.

Now, we can write the expression for the induced density. For a point $\mathbf{r} \in \mathbb{S}^2$, the local coordinates of the inverse map in $T_{\bar{\mathbf{q}}}(\mathbb{S}^2)$ are: for $i = 1, 2$, $x_i(\mathbf{r}) = \langle \mathbf{w}_i, \exp_{\bar{\mathbf{q}}}^{-1}(\mathbf{r}) \rangle = (\theta/\sin(\theta))(\langle \mathbf{w}_i, \mathbf{r} \rangle - \langle \mathbf{w}_i, \cos(\theta)\bar{\mathbf{q}} \rangle) = (\theta/\sin(\theta))\langle \mathbf{w}_i, \mathbf{r} \rangle = (\theta/\sin(\theta))\mathbf{W}^T \mathbf{r}$, where $\theta = \cos^{-1}(\langle \mathbf{r}, \bar{\mathbf{q}} \rangle)$ and $\mathbf{W} = [\mathbf{w}_1 \ \mathbf{w}_2] \in \mathbb{R}^{3 \times 2}$. Then, the induced truncated normal density on \mathbb{S}^2 is given by

$$f(\mathbf{r}; \bar{\mathbf{q}}, \mathbf{K}) = (1/Z_2)(\theta/\sin(\theta)) \exp\left\{-\frac{1}{2}(\theta/\sin(\theta))^2 \times (\mathbf{r}^T \mathbf{K}_w^{-1} \mathbf{r})\right\} \mathbf{1}_{\theta \leq \pi/2}, \tag{5}$$

where $\mathbf{K}_w^{-1} = \mathbf{W}\mathbf{K}^{-1}\mathbf{W}^T$. Three examples of a general f are shown in Figure 10(a). In each case, we show the original density in \mathbb{R}^2 (left column), the graph of f on \mathbb{S}^2 , and the level curves of f on \mathbb{S}^2 .

This framework can be extended to \mathbb{S}^k in a straightforward manner. Once again, let the mean $\bar{\mathbf{q}}$ be identified with the north pole of \mathbb{S}^k , and let $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_k \in \mathbb{R}^{k+1}$ form an orthonormal basis of $T_{\bar{\mathbf{q}}}(\mathbb{S}^k)$. Then, we can associate with any $\mathbf{v} \in T_{\bar{\mathbf{q}}}(\mathbb{S}^k)$ its coordinates $\mathbf{x} = (x_1, x_2, \dots, x_k)$, where $x_i = \langle \mathbf{v}, \mathbf{w}_i \rangle$. A truncated multivariate normal density in $T_{\bar{\mathbf{q}}}(\mathbb{S}^k)$, in terms of the local coordinates, is given by $f(\mathbf{x}) = (1/Z_k)e^{-(1/2)\mathbf{x}^T \mathbf{K}^{-1} \mathbf{x}} \mathbf{1}_{\|\mathbf{x}\| \leq \pi/2}(\mathbf{x})$, where \mathbf{K} is a $k \times k$ covariance matrix and Z_k is the normalizing constant. Without loss of generality, we can write an arbitrary point $\mathbf{v} \in T_{\bar{\mathbf{q}}}(\mathbb{S}^k)$ as $\theta \mathbf{w}_1$, with $\theta \in \mathbb{R}$ (since we can always rotate \mathbf{w}_i s so that \mathbf{w}_1 aligns with \mathbf{v}). The corresponding point on \mathbb{S}^k under the exponential map is given by $\mathbf{p} \equiv \cos(\theta)\bar{\mathbf{q}} + \sin(\theta)\mathbf{w}_1$. A convenient orthogonal basis for the tangent space $T_p(\mathbb{S}^k)$ is given by $\mathbf{b}_1 = (-\sin(\theta)\bar{\mathbf{q}} + \cos(\theta)\mathbf{w}_1)$, $\mathbf{b}_2 = \mathbf{w}_2, \dots, \mathbf{b}_k = \mathbf{w}_k$. To derive the differential of the map $\exp_{\bar{\mathbf{q}}}$, denoted $\exp_{\bar{\mathbf{q}}^*}$, we take each of the basis elements of $T_{\bar{\mathbf{q}}}(\mathbb{S}^k)$ and map them to $T_p(\mathbb{S}^k)$ under $\exp_{\bar{\mathbf{q}}^*}$. The resulting determinant of the Jacobian matrix is $(\sin(\theta)/\theta)^{k-1}$. For the derivation of $\exp_{\bar{\mathbf{q}}^*}$, refer to the Appendix.

Therefore, we can write the induced density on \mathbb{S}^k for a point \mathbf{r} , in terms of its local coordinates, as

$$f(\mathbf{r}; \bar{\mathbf{q}}, \mathbf{K}) = (1/Z_k)(\theta/\sin(\theta))^{(k-1)} \exp\left\{-\frac{1}{2} \times (\theta/\sin(\theta))^2 \mathbf{r}^T \mathbf{K}_w^{-1} \mathbf{r}\right\} \mathbf{1}_{\theta \leq \pi/2}, \tag{6}$$

where $\theta = \cos^{-1}(\langle \mathbf{r}, \bar{\mathbf{q}} \rangle)$, $\mathbf{K}_w^{-1} = \mathbf{W}\mathbf{K}^{-1}\mathbf{W}^T$, and $\mathbf{W} = [\mathbf{w}_1 \ \mathbf{w}_2 \ \dots \ \mathbf{w}_k] \in \mathbb{R}^{(k+1) \times k}$.

6.2.1 Shape Space \mathcal{S}_4 . Our goal is to develop explicit probability models in shape spaces of curves. We will take \mathcal{S}_4 as an example; it is a quotient space of a Hilbert sphere. To make the connection with \mathbb{S}^k , we need to find a relevant submanifold of \mathcal{S}_4 that can be identified with \mathbb{S}^k , for some k , and apply the previous steps. We accomplish this as follows. Take a $\bar{\mathbf{q}} \in \mathbb{S}_\infty$ and restrict to the subspace M , which is the orthogonal complement of $T_{\bar{\mathbf{q}}}([\bar{\mathbf{q}}])$ in $T_{\bar{\mathbf{q}}}(\mathbb{S}_\infty)$. Then, find a k -dimensional subspace M_k of M , using PCA or some other dimension reduction technique. Identify $\bar{\mathbf{q}}$ with the point $\mathbf{n} = (1, 0, \dots, 0)$ (the origin of this subspace). Then, the exponential map can be used to project elements of M_k into the desired sphere \mathbb{S}^k . Let $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_k$ denote an orthonormal basis of a subspace M_k (of $T_{\bar{\mathbf{q}}}(\mathbb{S}_\infty)$) of interest. We can represent any point $\mathbf{v} \in M_k$ as an element of \mathbb{R}^k using $\mathbf{v} = \sum_{i=1}^k x_i \mathbf{w}_i$. It can be shown that $u \doteq \exp_{\mathbf{n}}(\mathbf{v}) = c(\mathbf{x})\mathbf{n} + s(\mathbf{x}) \sum_{i=1}^k x_i \mathbf{w}_i$, where $c(\mathbf{x}) = \cos(\|\mathbf{x}\|)$ and $s(\mathbf{x}) = \sin(\|\mathbf{x}\|)/\|\mathbf{x}\|$. Such a point $u \in \mathbb{S}_\infty$ can alternatively be expressed using the vector $\mathbf{r} = (c(\mathbf{x}), s(\mathbf{x})x_1, \dots, s(\mathbf{x})x_k) \in \mathbb{S}^k$, which provides a way to identify a subset of \mathbb{S}_∞ with \mathbb{S}^k . Consequently, we can write a probability density on $\mathbb{S}^k \subset \mathbb{S}_\infty$ for any $u \in \mathbb{S}_\infty$ as

$$f(u) = (1/Z_k)(\theta/\sin(\theta))^{(k-1)} e^{-(1/2)\mathbf{x}(u)^T \mathbf{K}^{-1} \mathbf{x}(u)}, \tag{7}$$

where $\theta = \cos^{-1}(\langle \mathbf{r}, \mathbf{n} \rangle) = \|\mathbf{x}\|$ and $x_i(u) = (\theta/\sin(\theta)) \langle \mathbf{w}_i, \mathbf{r} \rangle$.

Figure 11 (top row) displays random samples from the truncated Gaussian distribution under the exponential map on the shape and shape + orientation spaces. Also note that these samples reflect the properties of interest. When only shape is considered, with invariance to position, orientation, scale, and parameterization, the samples vary in shape only. The rest of the properties are exactly the same for each of the curves generated by the model. When shape and orientation are of interest,

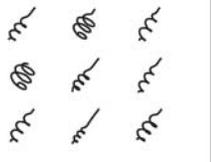
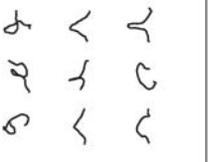
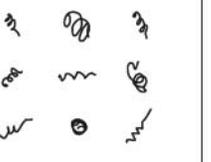
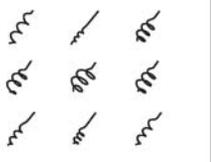
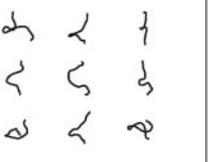
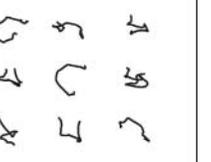
Shape		Shape+Orientation	
			
			
Helical Curves	Brain Fibers	Helical Curves	Brain Fibers

Figure 11. Samples from truncated Gaussian distribution under the exponential map (top row) and the stereographic projection (bottom row).

the generated curves vary in those properties only—that is, the scales and positions of the curves are fixed. It is important to note that in the original artificial data, all of the curves have three helices. This is evident in all of the generated curves as well. The samples based on the brain fiber data also provide valid observations under our model.

6.3 Truncated Gaussian Under the Stereographic Projection

Another possibility to map a pdf from \mathbb{R}^k to \mathbb{S}^k is by using the stereographic projection. Once a point $\bar{q} \in \mathbb{S}^k$ is identified with the origin in \mathbb{R}^k , a stereographic projection is specified as follows. Imagine a plane \mathbb{R}^k intersecting the sphere across the equator and the point s being the diametrically opposite point of \bar{q} on \mathbb{S}^k . Then, for any point v in \mathbb{R}^k , draw a straight line from s to v ; the intersection of this line with \mathbb{S}^k is called the stereographic projection of v . This is depicted pictorially in Figure 8(b). For the sake of brevity, we will only provide the final expression for the truncated Gaussian distribution on \mathbb{S}^k :

$$f(\mathbf{x}) = (1/(Z_k(1 + x_1)^k)) \exp\{-(1/2)\mathbf{x}^T \mathbf{K}^{-1} \mathbf{x}\} \mathbf{1}_{\|\mathbf{x}\| \leq \pi/2}(\mathbf{x}). \tag{8}$$

This has been accomplished earlier by Dortet-Bernadet (2008).

Figure 10(b) displays some illustrations of truncated Gaussian distributions under the stereographic projection for \mathbb{S}^2 . We

see that, in general, the truncated Gaussian distribution maps nicely. Nonetheless, we note that if the greatest eigenvalue of the covariance matrix becomes large, a unimodal density in the tangent space may map to a bimodal density on the sphere due to the nature of the Jacobian term.

Figure 11 (bottom row) displays random samples from the truncated Gaussian distribution under the stereographic projection on the shape and shape + orientation spaces. Once again, we see valid samples based on the given data.

6.4 Classification Using Gaussian Models on \mathcal{S}_4

To demonstrate the use of Gaussian-type distributions on \mathcal{S}_4 (shape only), we present classification results for a subset of the SHREC (Shape Retrieval Contest of Non-rigid 3D Models) 2010 protein dataset (Mavridis et al. 2010). Given a sample of l shapes in the training data for n classes, to perform classification, our approach is to first compute the sample Karcher mean of each class, $\bar{q}_i, i = 1, \dots, n$, in a leave-one-out manner and estimate the corresponding shooting vectors for each shape in the given data $v_i^j, i = 1, \dots, n, j = 1, \dots, l$. To classify each shape, we first remove its shooting vector from the set and train the covariance, $\mathbf{K}_i, i = 1, \dots, n$, for each of the classes. As discussed earlier, we perform PCA using the SVD of $\mathbf{K}_i = \mathbf{U}_i \mathbf{\Sigma}_i \mathbf{U}_i^T$. We restrict to a small number of principal components ($k = 8$) to keep the modeling efficient, and project each

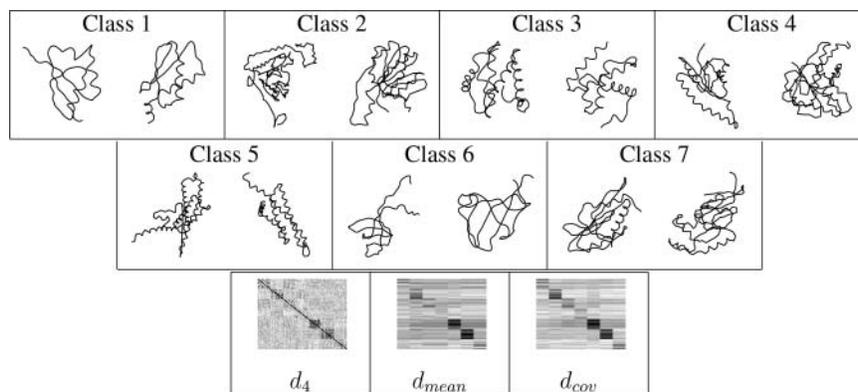


Figure 12. Top: sample proteins. Bottom: distance matrices for protein data.

Table 4. Leave-one-out nearest-neighbor classification results for protein data

Distance	d_4	d_{mean}	d_{cov}	Procrustes
Class. performance	87.1%	68.6%	88.6%	77.5%

shooting vector onto the orthonormal basis generated by them using $\mathbf{v}_{i,\text{proj}}^j = \sum_{m=1}^k c_i^{j,m} \mathbf{U}_i^m$, where $c_i^{j,m} = \langle \mathbf{v}_i^j, \mathbf{U}_i^m \rangle$. Then, to classify the removed shape, we compute a log-likelihood of each class as $D_i^j = (1/2) \mathbf{c}_i^j \boldsymbol{\Sigma}_i^{-1} \mathbf{c}_i^{jT} + (\|\mathbf{v}_i^j - \mathbf{v}_{i,\text{proj}}^j\|/\epsilon) + (1/2) \log(|\boldsymbol{\Sigma}_i|)$, where $\epsilon > 0$ is small (although this can be treated as a parameter in the analysis, we simply choose a value of 0.1). From now on, we will refer to this quantity as d_{cov} . We compare the classification results obtained using d_{cov} with those obtained using all pairwise shape distances (d_4 introduced in Section 3), a simple distance to the mean ($d_{\text{mean}} = d_4$ between each curve and the class means computed in a leave-one-out framework), and all pairwise Procrustes distances.

We consider the problem of classifying shapes of protein backbones selected from the SHREC 2010 database. The data consist of the first seven training classes, with 10 proteins in each. Two examples from each class are displayed in the top part of Figure 12. It is important to note that the protein shape classification problem is not a simple one, and that it is hard to separate the different classes in these data even through visual inspection. In the bottom part of Figure 12, we show the three distance matrices obtained using the different distance measures. The distance matrix obtained using d_{mean} and all pairwise distances (d_4) show little structure. The distance matrix generated using the log-likelihood method d_{cov} (computed in a leave-one-out manner) shows better separation of the seven classes. We compare all of the above performance measures with a leave-one-out nearest-neighbor classification rate based on a distance computed using rigid alignment only (Procrustes analysis, function `procrustes.m` in MATLAB). The leave-one-out nearest-neighbor classification performance results are summarized in Table 4. The log-likelihood method provides the highest classification rate for this protein dataset (in bold in Table 4). Furthermore, elastic shape analysis performs better than standard Procrustes analysis, which is an indication of improved feature (α helices, β sheets, etc.) matching through optimization over the reparameterization group.

7. CONCLUSION

We have proposed a methodology for the analysis of three-dimensional curves that allows for inclusion of any combination of shape, scale, orientation, and position in the analysis. The SRVF and SRF representations allow for invariance under the action of the reparameterization group, thus making deformations of curves elastic. This is an important feature because it allows us to generate more natural and smooth geodesic paths and more representative statistical models. In addition, we have presented a derivation of Gaussian models in shape spaces of curves under the exponential map and the stereographic projection. There are many applications of such a framework and we present ones from medical imaging in clustering of DT-

MRI brain fibers and bioinformatics for classification of protein backbones based on Gaussian log-likelihoods.

APPENDIX

A.1 Path-Straightening for SRF Representation

For a given parameterized curve β , its corresponding SRF is defined as $h_\beta(s) = \sqrt{|\dot{\beta}(s)|} \beta(s)$. As stated in Section 2, the space of all SRFs is a subset of \mathbb{L}^2 , and the actions of the reparameterization and rotation groups in this space under the \mathbb{L}^2 metric are by isometries. Thus, we will use the pullback of this metric from the space of SRFs to define geodesics in the space of parameterized curves (we will refer to this space as \mathcal{B}). It is important to note that Γ acts on \mathcal{B} on the right by composition, $\beta \circ \gamma$. The differential of the mapping h at β , denoted by dh_β , given a vector field $p \in T_\beta(\mathcal{B})$ and a scalar $r \in \mathbb{R}$, is given by $dh_\beta(p) = d/dr|_{r=0} h(\beta + rp) = (1/2\sqrt{|\dot{\beta}|})(\dot{\beta} \cdot \dot{p})\beta + \sqrt{|\dot{\beta}|}p$. We use this differential function to define a Riemannian metric on $T_\beta(\mathcal{B})$ as follows.

Definition A.1. For a $\beta \in \mathcal{B}$ and $p, m \in T_\beta(\mathcal{B})$, define the inner product:

$$\langle p, m \rangle_\beta \equiv \langle dh_\beta(p), dh_\beta(m) \rangle_{\mathbb{L}^2}, \quad (\text{A.1})$$

where the inner product on the right side is the standard inner product in \mathbb{L}^2 , as indicated.

With this metric, \mathcal{B} becomes a Riemannian manifold and we want to compute a geodesic path between two points, say β_1 and β_2 , in \mathcal{B} . Substituting the expression for dh_β in Equation (A.1), we obtain:

$$\begin{aligned} \langle p, m \rangle_\beta &= \langle dh_\beta(p), dh_\beta(m) \rangle_{\mathbb{L}^2} \\ &= \langle (1/2|\dot{\beta}|^{3/2})(\dot{\beta} \cdot \dot{p})\beta \\ &\quad + \sqrt{|\dot{\beta}|}p, (1/2|\dot{\beta}|^{3/2})(\dot{\beta} \cdot \dot{m})\beta + \sqrt{|\dot{\beta}|}m \rangle \\ &= \langle (1/4|\dot{\beta}|^3)(\dot{\beta} \cdot \dot{p})(\dot{\beta} \cdot \dot{m})\beta, \beta \rangle + \langle (1/2|\dot{\beta}|)(\dot{\beta} \cdot \dot{m})p \\ &\quad + (\dot{\beta} \cdot \dot{p})m, \beta \rangle + \langle |\dot{\beta}|p, m \rangle. \end{aligned}$$

This is the detailed form of the Riemannian metric imposed on \mathcal{B} for forming geodesic paths.

Let $F : [0, 1] \rightarrow \mathcal{B}$ denote a path in \mathcal{B} indexed by t such that $F(0) = \beta_1$ and $F(1) = \beta_2$. For example, we can initialize F using a straight line between β_1 and β_2 in \mathcal{B} under the \mathbb{L}^2 metric. The energy of this path is

$$\begin{aligned} E[F] &= \int_0^1 \langle F_t, F_t \rangle_F dt = \int_0^1 \langle dh_F(F_t), dh_F(F_t) \rangle_{\mathbb{L}^2} dt \\ &= \int_0^1 \langle ((1/4|\dot{F}|^3)(\dot{F} \cdot \dot{F})^2 F, F)_{\mathbb{L}^2} + \langle (1/|\dot{F}|)(\dot{F} \cdot \dot{F})F_t, F \rangle_{\mathbb{L}^2} \\ &\quad + \langle |\dot{F}|F_t, F_t \rangle_{\mathbb{L}^2} dt \\ &= \int_0^1 \int_0^1 \langle ((1/4|\dot{F}|^3)(\dot{F} \cdot \dot{F})^2 (F \cdot F) + (1/|\dot{F}|)(\dot{F} \cdot \dot{F})(F_t \cdot F) \\ &\quad + |\dot{F}|(F_t \cdot F_t)) ds dt, \end{aligned}$$

where we have used Equation (A.1) to reach the second equality. In this expression, we have suppressed the argument t for all quantities. Also, \dot{F} denotes the partial derivative with respect to s (curve parameterization), and F_t denotes the partial derivative with respect to t (path parameterization). To “straighten” F , we are going to use the gradient of E . It is well known that a critical point of E is a geodesic path. Let $G \in \mathcal{G}$ be a direction of perturbation of the path F . Here, \mathcal{G} is the space of all such perturbations. Then, the directional derivative of E in the direction G is given by, for $\epsilon \in \mathbb{R}$, $\nabla E_F(G) = (d/d\epsilon)E(F + \epsilon G)|_{\epsilon=0}$. The energy of the perturbed path is $E[F + \epsilon G] = \int_0^1 \langle dh_{F+\epsilon G}(F_t + \epsilon G_t), dh_{F+\epsilon G}(F_t + \epsilon G_t) \rangle_{\mathbb{L}^2} dt$. For

the sake of brevity, we only provide the final expression of the directional derivative of the energy, rather than its full derivation:

$$\begin{aligned} \nabla E_F(G) = & \int_0^1 \int_0^1 (((-3(\dot{F} \cdot \dot{G}))/4|\dot{F}|^5))(\dot{F} \cdot \dot{F}_i)^2(F \cdot F) \\ & + (1/(2|\dot{F}|^3))(\dot{F} \cdot \dot{F}_i)(\dot{F} \cdot \dot{G}_i + \dot{F}_i \cdot \dot{G})(F \cdot F) \\ & + (1/(2|\dot{F}|^3))(\dot{F} \cdot \dot{F}_i)^2(F \cdot G) \\ & + (-\dot{F} \cdot \dot{G})/|\dot{F}|^3(\dot{F} \cdot \dot{F}_i)(F_i \cdot F) + (1/|\dot{F}|)(\dot{F} \cdot \dot{G}_i \\ & + \dot{F}_i \cdot \dot{G})(F_i \cdot F) + (1/|\dot{F}|)(\dot{F} \cdot \dot{F}_i)(F_i \cdot G \\ & + F \cdot G_i) + ((\dot{F} \cdot \dot{G})/|\dot{F}|)(F_i \cdot F_i) + 2|\dot{F}|(F_i \cdot G_i)dsdt. \end{aligned}$$

To approximate the gradient of the path-straightening energy, we will use an orthonormal basis of \mathcal{G} , $p = \{p_i | i = 1, 2, \dots\}$ and set $\nabla E_F = \sum_{i=1}^{\infty} (\nabla E_F(p_i))p_i$. The basis p is formed using products of a modified Fourier basis, where the vector fields at $t = 0$ and $t = 1$ are zero. This modification is necessary because we do not want to change the starting and end points of the paths.

A.2 Optimization Over Nuisance Groups

In several of the representations mentioned in the article, one needs an optimization over certain shape-preserving transformation groups. The scale and translation can be easily removed in the representation process, by using rescaled SRVFs, but the process is different for removing the rotation and reparameterization groups. These are removed by solving optimization problems on these respective groups. In this section, we describe these two optimizations.

If we fix the parameterizations of curves, then their rotation alignment is straightforward, using Procrustes analysis. The solution to the problem $\mathbf{O}^* = \operatorname{argmin}_{\mathbf{O} \in \text{SO}(3)} \|\mathbf{q}_1 - \mathbf{O}\mathbf{q}_2\|$ is given by $\mathbf{O}^* = \mathbf{U}\mathbf{V}^T$, where $\mathbf{U}\mathbf{S}\mathbf{V}^T$ is the SVD of the 3×3 matrix $\mathbf{A} = \int_0^1 \mathbf{q}_1(t)\mathbf{q}_2(t)^T dt$. If the determinant of \mathbf{A} is negative, one needs to modify \mathbf{V} by changing the sign of its last column before multiplying by \mathbf{U} to obtain the optimal rotation \mathbf{O}^* .

Here, given a fixed rotation, we need to solve the optimization problem $\gamma^* = \operatorname{argmin}_{\gamma \in \Gamma} \|\mathbf{q}_1 - \sqrt{\gamma}(q_2 \circ \gamma)\|$. Since there is no analytical solution to this problem in general, we discuss a computational solution, called the dynamic programming algorithm. In this approach, the parameter domain $[0, 1]$ is discretized using a finite partition, for example, using k equally spaced points. This leads to a $k \times k$ grid on the square $[0, 1]^2$. Γ is actually a set of paths that start at the bottom-left corner of $[0, 1]^2$ and reach the top-right corner in such a way that the instantaneous directions are always strictly between 0 and $\pi/2$. In the approximate problem, one restricts to only piecewise linear paths that pass through the nodes of the $k \times k$ grid and the slopes are always strictly between 0 and ∞ . Since the cost function is defined by the \mathbb{L}^2 distance and, thus, is additive over the path $(t, \gamma(t))$, the dynamic programming algorithm applies here. The interested readers can refer to a textbook, for example, Bertsekas (1995).

A.3 Differential of the Exponential Map

Case of S^2 : The mapping of $\mathbf{w}_1, \exp_{n^*}(\mathbf{w}_1)$, is given by

$$\begin{aligned} \partial/\partial t|_{t=0} \exp_n(\theta \mathbf{w}_1 + t \mathbf{w}_1) &= \partial/\partial t|_{t=0} (\cos(\theta + t)\mathbf{n} + \sin(\theta + t)\mathbf{w}_1) \\ &= -\sin(\theta)\mathbf{n} + \cos(\theta)\mathbf{w}_1 = \mathbf{1}\mathbf{b}_1 + \mathbf{0}\mathbf{b}_2. \end{aligned}$$

The mapping of $\mathbf{w}_2, \exp_{n^*}(\mathbf{w}_2)$, is given by

$$\begin{aligned} \partial/\partial t|_{t=0} \exp_n(\theta \mathbf{w}_1 + t \mathbf{w}_2) &= \partial/\partial t|_{t=0} (\cos(\sqrt{\theta^2 + t^2})\mathbf{n} + \sin(\sqrt{\theta^2 + t^2}) \\ &\times ((\theta \mathbf{w}_1 + t \mathbf{w}_2)/\sqrt{\theta^2 + t^2})) = (-\sin(\sqrt{\theta^2 + t^2})(t/\sqrt{\theta^2 + t^2})\mathbf{n} \\ &+ \cos(\sqrt{\theta^2 + t^2})(t/\sqrt{\theta^2 + t^2})((\theta \mathbf{w}_1 + t \mathbf{w}_2)/(\sqrt{\theta^2 + t^2})) \\ &+ \sin(\sqrt{\theta^2 + t^2})(\mathbf{w}_2/\sqrt{\theta^2 + t^2}) - \sin(\sqrt{\theta^2 + t^2}) \end{aligned}$$

$$\begin{aligned} \times (\theta \mathbf{w}_1 + t \mathbf{w}_2)(t/(\theta^2 + t^2)^{3/2})|_{t=0} &= (\sin(\theta)/\theta)\mathbf{w}_2 \\ &= \mathbf{0}\mathbf{b}_1 + (\sin(\theta)/\theta)\mathbf{b}_2. \end{aligned}$$

Case of S^k : The mapping of $\mathbf{w}_1, \exp_{n^*}(\mathbf{w}_1)$, is given by

$$\begin{aligned} \frac{\partial}{\partial t} \Big|_{t=0} \exp_n(\theta \mathbf{w}_1 + t \mathbf{w}_1) &= \frac{\partial}{\partial t} \Big|_{t=0} (\cos(\theta + t)\mathbf{n} + \sin(\theta + t)\mathbf{w}_1) \\ &= -\sin(\theta)\mathbf{n} + \cos(\theta)\mathbf{w}_1 = \mathbf{1}\mathbf{b}_1. \end{aligned}$$

The mapping of $\mathbf{w}_i, i > 1, \exp_{n^*}(\mathbf{w}_i)$, is given by

$$\begin{aligned} \partial/\partial t|_{t=0} \exp_n(\theta \mathbf{w}_1 + t \mathbf{w}_i) &= \partial/\partial t|_{t=0} (\cos(\sqrt{\theta^2 + t^2})\mathbf{n} + \sin(\sqrt{\theta^2 + t^2}) \\ &\times ((\theta \mathbf{w}_1 + t \mathbf{w}_i)/\sqrt{\theta^2 + t^2})) = \frac{\sin(\theta)}{\theta} \mathbf{w}_i = \frac{\sin(\theta)}{\theta} \mathbf{b}_i. \end{aligned}$$

[Received February 2011. Revised April 2012.]

REFERENCES

Amari, S. (1985), *Differential Geometric Methods in Statistics* (Lecture Notes in Statistics, Vol. 28), Berlin: Springer. [1153,1154]

Bertsekas, D. P. (1995), *Dynamic Programming and Optimal Control*, Belmont, MA: Athena Scientific. [1164]

Bhattacharya, A. (1943), "On a Measure of Divergence Between Two Statistical Populations Defined by Their Probability Distributions," *Bulletin of Calcutta Mathematical Society*, 35, 99–109. [1153,1154]

Čencov, N. N. (1982), *Statistical Decision Rules and Optimal Inferences* (Translations of Mathematical Monographs, Vol. 53), Providence, RI: American Mathematical Society. [1154]

Dortet-Bernadet, J. L. (2008), "Model-Based Clustering on the Unit Sphere With an Illustration Using Gene Expression Profiles," *Biostatistics*, 9(1), 66–80. [1162]

Dryden, I. L., and Mardia, K. V. (1992), "Size and Shape Analysis of Landmark Data," *Biometrika*, 79(1), 57–68. [1152]

— (1998), *Statistical Shape Analysis*, New York: Wiley. [1152,1157]

Joshi, S. H., Klassen, E., Srivastava, A., and Jermyn, I. H. (2007), "A Novel Representation for Riemannian Analysis of Elastic Curves in \mathbb{R}^n ," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–7. [1153]

Kendall, D. G. (1984), "Shape Manifolds, Procrustean Metrics and Complex Projective Spaces," *Bulletin of London Mathematical Society*, 16, 81–121. [1152]

Klassen, E., and Srivastava, A. (2006), "Geodesics Between 3D Closed Curves Using Path-Straightening," in *European Conference on Computer Vision*, Heidelberg: Springer, pp. 95–106. [1156]

Klassen, E., Srivastava, A., Mio, W., and Joshi, S. H. (2004), "Analysis of Planar Shapes Using Geodesic Paths on Shape Spaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(3), 372–383. [1153]

Lang, S. (1999), *Fundamentals of Differential Geometry*, New York: Springer. [1155]

Le, H. (2001), "Locating Frechet Means With Application to Shape Spaces," *Advances in Applied Probability*, 33(2), 324–338. [1157]

Liu, W., Srivastava, A., and Zhang, J. (2011), "A Mathematical Framework for Protein Structure Comparison," *PLoS Computational Biology*, 7(2), e1001075. [1153]

Mani, M., Kurtsek, S., Barillot, C., and Srivastava, A. (2010), "A Comprehensive Riemannian Framework for the Analysis of White Matter Fiber Tracts," in *IEEE International Symposium on Biomedical Imaging*, pp. 1101–1104. [1153]

Mardia, K. V., and Jupp, P. E. (2000), *Directional Statistics*, Chichester: Wiley. [1159]

Mavridis, L., Venkatraman, V., Ritchie, D., Morikawa, H., Andonov, R., Cornu, A., Malod-Dognin, N., Nicolas, J., Temerinac-Ott, M., Reisert, M., Burkhardt, H., Axenopoulos, A., and Daras, P. (2010), "SHREC'10 Track: Protein Models," in *Eurographics Workshop on 3D Object Retrieval - 3DOR 2010*. [1162]

Michor, P. W., and Mumford, D. (2006), "Riemannian Geometries on Spaces of Plane Curves," *Journal of European Mathematical Society*, 8, 1–48. [1153]

Mio, W., Srivastava, A., and Joshi, S. H. (2007), "On Shape of Plane Elastic Curves," *International Journal of Computer Vision*, 73(3), 307–324. [1153,1154]

- Morgan, V. L., Mishra, A., Newton, A. T., Gore, J. C., and Ding, Z. (2009), "Integrating Functional and Diffusion Magnetic Resonance Imaging for Analysis of Structure-Function Relationship in the Human Language Network," *PLoS ONE*, 4(8), e6660. [1157]
- Pelletier, B. (2005), "Kernel Density Estimation on Riemannian Manifolds," *Statistics & Probability Letters*, 73, 297–304. [1159]
- Rao, C. R. (1945), "Information and Accuracy Attainable in the Estimation of Statistical Parameters," *Bulletin of the Calcutta Mathematical Society*, 37, 81–91. [1153,1154]
- Small, C. G. (1996), *The Statistical Theory of Shape*, Berlin: Springer. [1152]
- Srivastava, A., Joshi, S. H., Mio, W., and Liu, X. (2005), "Statistical Shape Analysis: Clustering, Learning and Testing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4), 590–602. [1153]
- Srivastava, A., Klassen, E., Joshi, S. H., and Jermyn, I. H. (2011), "Shape Analysis of Elastic Curves in Euclidean Spaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33, 1415–1428. [1153,1154,1156]
- Sundaramoorthi, G., Mennucci, A., Soatto, S., and Yezzi, A. (2011), "A New Geometric Metric in the Space of Curves, and Applications to Tracking Deforming Objects by Prediction and Filtering," *SIAM Journal of Imaging Science*, 4(1), 109–145. [1153]
- Vos, P. W., and Kass, R. E. (1997), *Geometrical Foundations of Asymptotic Inference*, New York: Wiley-Interscience. [1154]
- Younes, L. (1998), "Computable Elastic Distance Between Shapes," *SIAM Journal of Applied Mathematics*, 58(2), 565–586. [1153]
- Younes, L., Michor, P. W., Shah, J., Mumford, D., and Lincei, R. (2008), "A Metric on Shape Space With Explicit Geodesics," *Matematica E Applicazioni*, 19(1), 25–57. [1153]
- Zahn, C. T., and Roskies, R. Z. (1972), "Fourier Descriptors for Plane Closed Curves," *IEEE Transactions on Computers*, 21(3), 269–281. [1153]