Int J Comput Vis (2013) 105:171–185
DOI 10.1007/s11263-012-0591-y

Geodesic Regression and the Theory of Least Squares on Riemannian Manifolds

P. Thomas Fletcher

Received: 14 December 2011 / Accepted: 17 October 2012 / Published online: 16 November 2012

Abstract This paper develops the theory of geodesic regression and least-squares estimation on Riemannian manifolds. Geodesic regression is a method for finding the relationship between a real-valued independent variable and a manifold-valued dependent random variable, where this relationship is modeled as a geodesic curve on the manifold. Least-squares estimation is formulated intrinsically as a minimization of the sum-of-squared geodesic distances of the data to the estimated model. Geodesic regression is a direct generalization of linear regression to the manifold setting, and it provides a simple parameterization of the estimated relationship as an initial point and velocity, analogous to the intercept and slope. A nonparametric permutation test for determining the significance of the trend is also given. For the case of symmetric spaces, two main theoretical results are established. First, conditions for existence and uniqueness of the least-squares problem are provided. Second, a maximum likelihood criteria is developed for a suitable definition of Gaussian errors on the manifold. While the method can be generally applied to data on any manifold, specific examples are given for a set of synthetically generated rotation data and an application to analyzing shape changes in the corpus callosum due to age.

Keywords Geodesic regression · Manifold statistics · Shape analysis

1 Introduction

Regression analysis is a fundamental statistical tool for determining how a measured variable is related to one or more potential explanatory variables. The most commonly used regression model is linear regression, due to its simplicity, ease of interpretation, and ability to model many phenomena. However, if the response variable takes values on a nonlinear manifold, a linear model is not applicable. Such manifold-valued measurements arise in many applications, including those involving directional data, transformations, tensors, and shape. For example, in biology and medicine it is often critical to understand processes that change the shape of anatomy. The difficulty is that shape variability is inherently high-dimensional and nonlinear. An effective approach to capturing this variability has been to parameterize shape as a manifold, or shape space.

Statistical analysis of manifold data has been developed by several authors. The seminal work of Fréchet (1948) generalized the concept of expectation from linear spaces to general metric spaces. This opened up the possibility of computing a sample mean statistic from a set of data on a manifold using the geodesic distance as metric. The Fréchet mean of a set of points, $y_1, \ldots, y_N$, in a Riemannian manifold M is given by

$$\mu = \arg\min_{y \in M} \sum_{i=1}^{N} d(y, y_i)^2,$$

where d is the geodesic distance between points on M. This equation generalizes the principle of least squares to the metric space setting. Karcher (1977) provided conditions guaranteeing the existence and uniqueness of the Fréchet mean, which were later improved by Kendall (1990).

Second-order statistics on manifolds, such as principal geodesic analysis (Fletcher et al. 2003), a generalization of principal components analysis to manifolds, and Gaussian covariance estimation (Pennec 2006), have also been developed and applied in the domain of image analysis. The

P. Thomas Fletcher (✉)
University Of Utah, Salt Lake City, UT 84112, USA
e-mail: fletcher@sci.utah.edu

work of Sommer et al. (2010) developed an exact solution to principal geodesic analysis, which utilized much of the same Jacobi field machinery that is needed here for geodesic regression. Also, the geodesic principal components analysis of Huckemann et al. (2010) finds best fitting geodesics to data living in certain quotient spaces, such as Kendall's shape space. A related method is the geodesic curve modeling of Kenobi et al. (2010), where shape changes over time are modeled as a geodesic curve. In each of these works, geodesics model variation in data that is not labeled with an independent scalar parameter. This is a related, but distinctly different problem from geodesic regression, which seeks to find the relationship between an independent scalar parameter (such as time) and a dependent, manifold-valued variable. Related work includes statistical analysis of directional data (e.g., spheres) (Mardia and Jupp 2000) and analysis on shape manifolds (Dryden and Mardia 1998), where statistics are derived from probability distributions on specific manifolds (for example, the Fisher-von Mises distribution on spheres).

Several works have studied the regression problem on manifolds. Jupp and Kent (1987) propose an unrolling method on the sphere, which was later extended to shape spaces by Kume et al. (2007). Regression analysis on the group of diffeomorphisms has been proposed as growth models by Miller (2004), nonparametric regression by Davis et al. (2007), and second order splines by Trouvé and Vialard (2010). Durrleman et al. (2009) construct spatiotemporal image atlases from longitudinal data. Finally, Shi et al. (2009) proposed a semiparametric model with multiple covariates for manifold response data. None of these methods provide a direct generalization of linear regression to manifolds.

The purpose of this paper is to develop such a generalization, called geodesic regression, which models the relationship between an independent scalar variable and a dependent manifold-valued random variable as a geodesic curve. Like linear regression, the advantages of this model are its simplicity and ease of interpretation. As will be shown, the geodesic regression model also leads to a straightforward generalization of the R^2 statistic and a hypothesis test for significance of the estimated geodesic trend. This paper is an expanded exposition of the geodesic regression method first introduced in Fletcher (2011). Niethammer et al. (2011) independently proposed geodesic regression for the case of diffeomorphic transformations of image time series. In the current paper, we advance the theory of geodesic regression and least-squares estimation by proving two main results. First, we show that under certain conditions, the least-squares estimation problem for geodesic regression has a unique solution. Second, we formulate the least-squares problem as a maximum likelihood estimation.

2 Multiple Linear Regression

Before formulating geodesic regression on general manifolds, we begin by reviewing a particular case of multiple linear regression in $\mathbb{R}^n$. Here we are interested in the relationship between a non-random *independent* variable $X \in \mathbb{R}$ and a random *dependent* variable Y taking values in $\mathbb{R}^n$. A multiple linear model of this relationship is given by

$$Y = \alpha + X\beta + \epsilon, \tag{1}$$

where $\alpha \in \mathbb{R}^n$ is an unobservable *intercept* parameter, $\beta \in \mathbb{R}^n$ is an unobservable *slope* parameter, and ϵ is an $\mathbb{R}^n$-valued, unobservable random variable representing the error. Geometrically, this is the equation of a one-dimensional line through $\mathbb{R}^n$ (plus noise), parameterized by the scalar variable X. For the purposes of generalizing to the manifold case, it is useful to think of α as the starting point of the line and β as a velocity vector.

Given realizations of the above model, i.e., data $(x_i, y_i) \in \mathbb{R} \times \mathbb{R}^n$, for $i = 1, \ldots, N$, the least squares estimates, $\hat{\alpha}$, $\hat{\beta}$, for the intercept and slope are computed by solving the minimization problem

$$(\hat{\alpha}, \hat{\beta}) = \arg\min_{(\alpha,\beta)} \sum_{i=1}^{N} \|y_i - \alpha - x_i\beta\|^2 . \tag{2}$$

This equation can be solved analytically, yielding

$$\hat{\beta} = \frac{\frac{1}{N}\sum x_i\, y_i - \bar{x}\,\bar{y}}{\sum x_i^2 - \bar{x}^2},$$
$$\hat{\alpha} = \bar{y} - \bar{x}\,\hat{\beta},$$

where $\bar{x}$ and $\bar{y}$ are the sample means of the x_i and y_i, respectively. If the errors in the model are drawn from distributions with zero mean and finite variance, then these estimators are unbiased and consistent.

3 Geodesic Regression

Let $y_1, \ldots, y_N$ be points on a smooth Riemannian manifold M, with associated scalar values $x_1, \ldots, x_N \in \mathbb{R}$. The goal of geodesic regression is to find a geodesic curve γ on M that best models the relationship between the x_i and the y_i. Just as in linear regression, the speed of the geodesic will be proportional to the independent parameter corresponding to the x_i. Estimation will be set up as a least-squares problem, where we want to minimize the sum-of-squared Riemannian distances between the model and the data. A schematic of the geodesic regression model is shown in Fig. 1.

Before formulating the model, we review a few basic concepts of Riemannian geometry. We will write an element of the tangent bundle as the pair $(p, v) \in TM$, where p is a

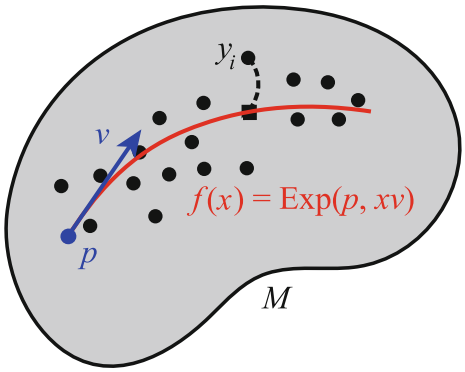

Fig. 1 Schematic of the geodesic regression model

point in M and $v \in T_pM$ is a tangent vector at p. Recall that for any $(p, v) \in TM$ there is a unique geodesic curve γ, with initial conditions $\gamma(0) = p$ and $\gamma'(0) = v$. This geodesic is only guaranteed to exist locally. When γ is defined over the interval $[0, 1]$, the exponential map at p is defined as $\mathrm{Exp}_p(v) = \gamma(1)$. In other words, the exponential map takes a position and velocity as input and returns the point at time one along the geodesic with these initial conditions. The exponential map is locally diffeomorphic onto a neighborhood of p. Let $V(p)$ be the largest such neighborhood. Then within $V(p)$ the exponential map has an inverse, the Riemannian log map, $\mathrm{Log}_p : V(p) \to T_pM$. For any point $q \in V(p)$ the Riemannian distance function is given by $d(p, q) = \|\mathrm{Log}_p(q)\|$. It will be convenient to include the point p as a parameter in the exponential and log maps, i.e., define $\mathrm{Exp}(p, v) = \mathrm{Exp}_p(v)$ and $\mathrm{Log}(p, q) = \mathrm{Log}_p(q)$.

Notice that the tangent bundle TM serves as a convenient parameterization of the set of possible geodesics on M. An element $(p, v) \in TM$ provides an intercept p and a slope v, analogous to the α and β parameters in the multiple linear regression model (1). In fact, β is a vector in the tangent space $T_\alpha\mathbb{R}^n \cong \mathbb{R}^n$, and thus (α, β) is an element of the tangent bundle $T\mathbb{R}^n$. Now consider an M-valued random variable Y and a non-random variable $X \in \mathbb{R}$. The generalization of the multiple linear model to the manifold setting is the *geodesic model*,

$$Y = \mathrm{Exp}(\mathrm{Exp}(p, Xv), \epsilon), \tag{3}$$

where ϵ is a random variable taking values in the tangent space at $\mathrm{Exp}(p, Xv)$. Notice that for Euclidean space, the exponential map is simply addition, that is, $\mathrm{Exp}(p, v) = p + v$. Thus, the geodesic model coincides with (1) when $M = \mathbb{R}^n$.

3.1 Least Squares Estimation

Consider a realization of the model (3): $(x_i, y_i) \in \mathbb{R} \times M$, for $i = 1, \ldots, N$. Given this data, we wish to find estimates of the parameters $(p, v) \in TM$. First, define the sum-of-squared error of the data from the geodesic given by (p, v) as

$$E(p, v) = \frac{1}{2} \sum_{i=1}^{N} d(\mathrm{Exp}(p, x_i v), y_i)^2. \tag{4}$$

Following the ordinary least squares minimization problem given by (2), we formulate a least squares estimator of the geodesic model as a minimizer of the above sum-of-squares energy, i.e.,

$$(\hat{p}, \hat{v}) = \arg\min_{(p,v)} E(p, v). \tag{5}$$

Again, notice that this problem coincides with the ordinary least squares problem when $M = \mathbb{R}^n$.

Unlike the linear setting, the least squares problem in (5) for a general manifold M will typically not yield an analytic solution. Instead we derive a gradient descent algorithm. Computation of the gradient of (4) will require two parts: the derivative of the Riemannian distance function and the derivative of the exponential map. Fixing a point $p \in M$, the gradient of the squared distance function is $\nabla_x d(p, x)^2 = -2\mathrm{Log}_x(p)$ for $x \in V(p)$.

The derivative of the exponential map $\mathrm{Exp}(p, v)$ can be separated into a derivative with respect to the initial point p and a derivative with respect to the initial velocity v. To do this, first consider a variation of geodesics given by $c_1(s, t) = \mathrm{Exp}(\mathrm{Exp}(p, su_1), tv(s))$, where $u_1 \in T_pM$ defines a variation of the initial point along the geodesic $\eta(s) = \mathrm{Exp}(p, su_1)$. Here we have also extended $v \in T_pM$ to a vector field $v(s)$ along η via parallel translation. This variation is illustrated on the left side of Fig. 2. Next consider a variation of geodesics $c_2(s, t) = \mathrm{Exp}(p, su_2 + tv)$, where $u_2 \in T_pM$. (Technically, u_2 is a tangent to the tangent space, i.e., an element of $T_v(T_pM)$, but there is a natural isomorphism $T_v(T_pM) \cong T_pM$.) The variation c_2 produces a "fan" of geodesics as seen on the right side of Fig. 2. These variations lead to *Jacobi fields*, which are solutions to the second order equation

$$\frac{D^2}{dt^2} J(t) + R(J(t), \quad \gamma'(t))\ \gamma'(t) = 0, \tag{6}$$

where R is the Riemannian curvature tensor. For more details on the derivation of the Jacobi field equation and the curvature tensor, see for instance (do Carmo 1992).

Now the derivatives of $\mathrm{Exp}(p, v)$ with respect to p and v are given by

$$d_p\mathrm{Exp}(p, v) \cdot u_1 = \frac{d}{ds} c_1(s, t)\Big|_{s=0} = J_1(1)$$
$$d_v\mathrm{Exp}(p, v) \cdot u_2 = \frac{d}{ds} c_2(s, t)\Big|_{s=0} = J_2(1),$$

where $J_i(t)$ are Jacobi fields along the geodesic $\gamma(t) = \mathrm{Exp}(p, tv)$. The initial conditions for the two Jacobi fields

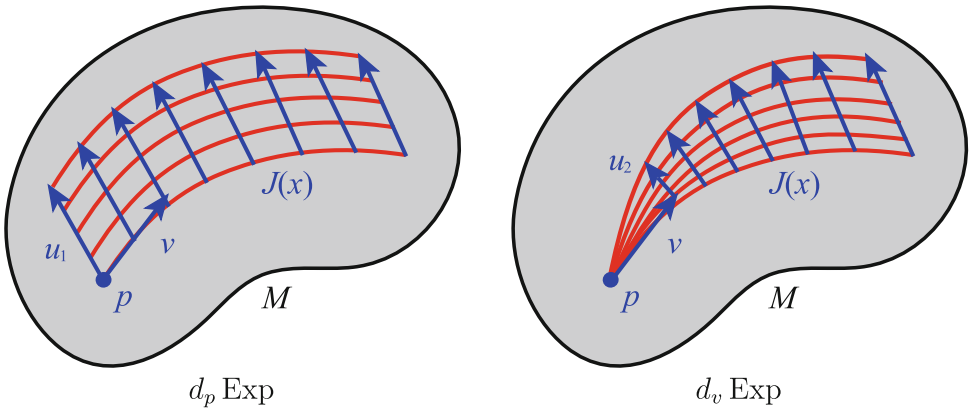

Fig. 2 Jacobi fields as derivatives of the exponential map

above are $J_1(0) = u_1$, $J_1'(0) = 0$ and $J_2(0) = 0$, $J_2'(0) = u_2$, respectively. If we decompose the Jacobi field into a component tangential to γ and a component orthogonal, i.e., $J = J^\top + J^\perp$, the tangential component is linear: $J^\top(t) = u_1^\top + t u_2^\top$. Therefore, the only challenge is to solve for the orthogonal component. Finally, the gradient of the sum-of-squares energy in (4) is givenby

$$\nabla_p E = -\sum_{i=1}^{N} d_p \mathrm{Exp}(p, x_i v)^\dagger \epsilon_i, \tag{7}$$

$$\nabla_v E = -\sum_{i=1}^{N} x_i \, d_v \mathrm{Exp}(p, x_i v)^\dagger \epsilon_i, \tag{8}$$

where we have defined

$$\epsilon_i = \mathrm{Log}(\mathrm{Exp}(p, x_i v), y_i),$$

and we have taken the adjoint of the exponential map derivative, e.g., defined by

$$\langle d_p \mathrm{Exp}(p, v) u, w \rangle = \langle u, d_p \mathrm{Exp}(p, v)^\dagger w \rangle.$$

As we will see in the next section, formulas for Jacobi fields and their respective adjoint operators can often be derived analytically for many useful manifolds.

3.2 R^2 Statistics and Hypothesis Testing

In regression analysis the most basic question one would like to answer is whether the relationship between the independent and dependent variables is significant. A common way to test this is to see if the amount of variance explained by the model is high. For geodesic regression we will measure the amount of explained variance using a generalization of the R^2 statistic, or coefficient of determination, to the manifold setting. To do this, we first define predicted values of y_i and the errors ϵ_i as

$$\hat{y}_i = \mathrm{Exp}(\hat{p}, x_i \hat{v}),$$
$$\hat{\epsilon}_i = \mathrm{Log}(\hat{y}_i, y_i),$$

where $(\hat{p}, \hat{v})$ are the least squares estimates of the geodesic parameters defined above. Note that the $\hat{y}_i$ are points along the estimated geodesic that are the best predictions of the y_i given only the x_i. The $\hat{\epsilon}_i$ are the residuals from the model predictions to the true data.

Now to define the total variance of data, $y_1, \ldots, y_N \in M$, we use the Fréchet variance, intrinsically defined by

$$\mathrm{var}(y_i) = \min_{y \in M} \frac{1}{N} \sum_{i=1}^{N} d(y, y_i)^2.$$

The unexplained variance is the variance of the residuals, $\mathrm{var}(\hat{\epsilon}_i) = \frac{1}{N} \sum \|\hat{\epsilon}_i\|^2$. From the definition of the residuals, it can be seen that the unexplained variance is the mean squared distance of the data to the model, i.e., $\mathrm{var}(\hat{\epsilon}_i) = \frac{1}{N} \sum d(\hat{y}_i, y_i)^2$. Using these two variance definitions, the generalization of the R^2 statistic is then given by

$$R^2 = 1 - \frac{\text{unexplained variance}}{\text{total variance}} = 1 - \frac{\mathrm{var}(\hat{\epsilon}_i)}{\mathrm{var}(y_i)}. \tag{9}$$

Fréchet variance coincides with the standard definition of variance when $M = \mathbb{R}^n$. Therefore, it follows that the definition of R^2 in (9) coincides with the R^2 for linear regression when $M = \mathbb{R}^n$. Also, because Fréchet variance is always nonnegative, we see that $R^2 \leq 1$, and that $R^2 = 1$ if and only if the residuals to the model are exactly zero, i.e., the model perfectly fits the data. Finally, it is clear that the residual variance is always smaller than the total variance, i.e., $\mathrm{var}(\hat{\epsilon}_i) \leq \mathrm{var}(y_i)$. This is because we could always choose $\hat{p}$ to be the Fréchet mean and $v = 0$ to achieve $\mathrm{var}(\hat{\epsilon}_i) = \mathrm{var}(y_i)$. Therefore, $R^2 \geq 0$, and it must lie in the interval [0, 1], as is the case for linear models.

We now describe a permutation test for testing the significance of the estimated slope term, $\hat{v}$. Notice that if we constrain v to be zero in (5), then the resulting least squares estimate of the intercept, $\hat{p}$, will be the Fréchet mean of the y_i. The desired hypothesis test is whether the fraction of unexplained variance is significantly decreased by also estimating

v. The null hypothesis is $H_0 : v = 0$, which is the case if the unexplained variance in the geodesic model is equal to the total variance. Under the null hypothesis, there is no relationship between the X variable and the Y variable. Therefore, the x_i are exchangeable under the null hypothesis, and a permutation test may randomly reorder the x_i data, keeping the y_i fixed. Estimating the geodesic regression parameters for each random permutation of the x_i, we can calculate a sequence of R^2 values, $R_1^2, \ldots, R_m^2$, which approximate the sampling distribution of the R^2 statistic under the null hypothesis. Computing the fraction of the R_k^2 that are greater than the R^2 estimated from the unpermuted data gives us a p-value.

4 Geodesic Regression on Symmetric Spaces

In this section we derive two results on the theory of geodesic regression and least squares estimation for the case when the Riemannian manifold under consideration is a symmetric space. The motivation is to extend desirable properties of least-squares estimates of linear regression models to the analogous least-squares estimates of geodesic regression on symmetric spaces. The first result provides conditions for existence and uniqueness of the geodesic least-squares estimator. This is analogous to the existence and uniqueness result for the sample Fréchet mean given by Karcher (1977) and extended by Kendall (1990). The second result is a characterization of the least-squares estimation problem as a maximum likelihood estimator under a suitable probability distribution for the errors. In the process, we give an explicit formula for the second derivative of the exponential map on a symmetric space.

4.1 Symmetric Spaces

Symmetric spaces make up an important class of manifolds with "nice" geometric properties. Many useful manifolds are symmetric spaces, including constant curvature spaces (Euclidean spaces, hyperbolic spaces, and spheres), compact Lie groups (rotation groups, unitary groups, etc.), the spaces of linear subspaces in $\mathbb{R}^n$ (Grassmannians) and orthonormal frames in $\mathbb{R}^n$ (Stiefel manifolds), complex projective space (Kendall's 2D shape space), and the spaces of positive-definite, symmetric matrices. This section provides an overview of the basic properties of symmetric spaces. More details can be found in standard references (Boothby 1986), including a complete classification of symmetric spaces (Helgason 1978).

Recall that an *isometry* of a Riemannian manifold is a diffeomorphism $\sigma : M \to M$ that preserves the Riemannian metric, or equivalently, such that $d(x, y) = d(\sigma(x), \sigma(y))$ for all $x, y \in M$. Also, the map σ is called *involutive* if it is its own inverse, i.e., $\sigma \circ \sigma = \mathrm{id}$. We are now ready to give the definition of a symmetric space.

Definition 1 A symmetric space is a Riemannian manifold M such that for each point $x \in M$, there exists an involutive isometry σ_x that fixes x and reverses geodesics passing through x.

Symmetric spaces have several geometric properties that will be important for proving the theoretical properties of least-squares estimation in this section. The first property is that a symmetric space M is also a *homogeneous space*, which means that the group of isometries of M act transitively. In other words, given two points $x, y \in M$, there exists an isometry ϕ such that $\phi(x) = y$. The second property is that symmetric spaces are always complete, which according to the Hopf-Rinow theorem is equivalent to saying that geodesics continue for all time or that between any two points there exists at least one minimizing geodesic.

For the purposes of geodesic regression, the most important property of symmetric spaces is the fact that the Jacobi field equations can be given in closed form. Not only does this result in efficient algorithms for computing least-squares estimates, but it will allow us to compute second derivatives of the exponential map in the next section. This stems from the fact that the covariant derivative of the Riemannian curvature tensor is zero, i.e., $\nabla R = 0$.

To write formulas for the Jacobi fields on a symmetric space, consider a variation of geodesics, $\alpha(s, t) = \mathrm{Exp}(p(t), sV(t))$, where $p : (-\epsilon, \epsilon) \to M$ is a geodesic, and V is a smooth vector field along p that satisfies $D^2V/dt^2 = 0$. Define the coordinate vector fields

$$V(s,t) = \frac{d}{ds}\alpha(s,t),$$
$$J(s,t) = \frac{d}{dt}\alpha(s,t).$$

For a fixed $t = t_0$, we have that $\gamma_{t_0}(s) = \alpha(s, t_0)$ is a geodesic with velocity $V(s, t_0)$, and $J(s, t_0)$ is a Jacobi field along γ_{t_0}.

Let $L(t) = \|V(s,t)\|$, which is constant in s because $DV/ds = 0$. Let $T(s,t) = V(s,t)/\|V(s,t)\|$ be the unit tangent vector field. Then the subspace normal to T can be decomposed as an orthonormal frame $E_i(s,t)$, $i = 2, \ldots, n$, with $\langle T, E_i \rangle = 0$ and $\|E_i\| = 1$, such that $DE_i/ds = 0$ and the sectional curvature $K(E_i, J) = \kappa_i$ is constant. Such a basis vector field exists due to the fact that $\nabla R = 0$ on a symmetric space.

Now the Jacobi field J can be written as

$$J(s,t) = \sum_{i=2}^{n} \left(f_i(s)X_i(t) + g_i(s)Y_i(t)\right) E_i(t) + \left(Z(t) + sW(t)\right) T(s,t), \tag{10}$$

where X_i, Y_i are functions representing the orthogonal components of J, and Z, W are functions representing the

tangential components of J. Defining $k_i = \sqrt{|\kappa_i|}L$, the functions f_i, g_i are given by

$$f_i(s) = \begin{cases} \cos(sk_i), & \text{if } \kappa_i > 0 \\ \cosh(sk_i), & \text{if } \kappa_i < 0 \\ 1, & \text{if } \kappa_i = 0 \end{cases}$$

$$g_i(s) = \begin{cases} \frac{1}{k_i}\sin(sk_i), & \text{if } \kappa_i > 0 \\ \frac{1}{k_i}\sinh(sk_i), & \text{if } \kappa_i < 0 \\ s, & \text{if } \kappa_i = 0 \end{cases}$$

When the context is clear, we will suppress the s, t parameters in the notation for simplicity.

4.2 Second Derivative of the Exponential Map

As we have seen, the gradients of the geodesic sum-of-squares energy in (4) are given by the first derivative of the exponential map. To prove the convexity of the least-squares problem, we will need to compute the second derivative of the sum-of-squares energy, which involves the second derivative of the exponential map. Theorem 1 below provides explicit equations for the second derivative of the exponential map on a general symmetric space. To the best of our knowledge, these formulas have not appeared anywhere previously in the literature. We will use the notation established above for the Jacobi field formula (10). We know that the first derivative of the exponential map is given by the endpoint of the Jacobi field,

$$\frac{d}{dt}\text{Exp}(p(t),\ V(t))\Big|_{t=0} = J(1,0).$$

The goal now is to compute the second derivative

$$\frac{D}{dt}\frac{d}{dt}\text{Exp}(p(t),\ V(t))\Big|_{t=0} = \frac{DJ}{dt}(1,0).$$

Theorem 1 *Let M be a symmetric space, and consider a geodesic $p : (-\epsilon, \epsilon) \to M$ and a smooth vector field V along p, satisfying $D^2V/dt^2 = 0$. Let E_i be a parallel, orthonormal basis as above with corresponding sectional curvatures κ_i. Then the second derivative of the exponential map on M is given by*

$$\frac{D}{dt}\frac{d}{dt}Exp(p(t),\ V(t))\Big|_{t=0} = \sum_{i=2}^{n} a_i T + \sum_{i=2}^{n} b_i E_i,$$

where for $\kappa_i > 0$,

$$a_i = \frac{k_i}{2L}\sin(2k_i)\left(X_i^2 - \frac{Y_i^2}{k_i^2}\right) - \frac{\cos(2k_i)}{L}X_iY_i + \frac{Y_i^2}{L} + \frac{X_iY_i}{L},$$

$$b_i = -\frac{2k_i}{L}\sin(k_i)X_iW - \frac{k_i}{L}\sin(k_i)X_iZ + \frac{2}{L}\cos(k_i)Y_iW - \frac{2}{k_iL}\sin(k_i)Y_iW,$$

and for $\kappa_i < 0$,

$$a_i = \frac{k_i}{2L}\sinh(2k_i)\left(X_i^2 - \frac{Y_i^2}{k_i^2}\right) + \frac{\cosh(2k_i)}{L}X_iY_i - \frac{Y_i^2}{L} - \frac{X_iY_i}{L},$$

$$b_i = \frac{2k_i}{L}\sinh(k_i)X_iW + \frac{k_i}{L}\sinh(k_i)X_iZ + \frac{2}{L}\cosh(k_i)Y_iW - \frac{2}{k_iL}\sinh(k_i)Y_iW,$$

and for $\kappa_i = 0$,

$$a_i = b_i = 0.$$

The proof of Theorem 1 is rather lengthy and can be found in Appendix A.

4.3 Existence and Uniqueness

We now prove a local existence and uniqueness result for the least-squares estimates, $\hat{p}, \hat{v}$, of geodesic regression. To do this, we show that the least-squares criterion is convex within a local neighborhood of the data, that is, the second derivative of the energy functional (4) is positive. Given a variation of the initial conditions, a geodesic $p : (-\epsilon, \epsilon) \to M$ and a vector field V along p, with $D^2V/Dt^2 = 0$, we want

$$\frac{d^2}{dt^2}E(p(t),\ V(t)) > 0.$$

We will see that this is true for geodesics that are sufficiently close to the data points. To this end, define the *diameter* of a set $A \subset M$ as $\text{diam}(A) = \max_{x,y\in A} d(x, y)$. Also, we say that the set $A \subset$ *of* M is *geodesically convex* if between any two points $x, y \in A$, there exists a unique, minimal geodesic between x and y. This condition will be needed to ensure that the Riemannian log map and Jacobi fields are always well-defined.

For local existence and uniqueness, we will consider the class of geodesics that are sufficiently close to the data points. Given a real number $r > 0$, we say that a geodesic $\gamma \in M$ is *r-close* to the data (x_i, y_i) if $d(\gamma(x_i), y_i) \leq r$ and γ does not pass through the cut locus of M. Throughout we will also assume without loss of generality that one of the $x_i = 0$, which we can achieve by a constant shift in the parameterization. This just ensures that the estimate for the intercept point p at $x = 0$ will remain close to the data. The final assumption is that we avoid the degenerate case where all x_i are equal (which would also be a problem for linear regression).

Now let's consider the general formula for the second derivative of the regression energy. First, define

$$\eta_i(t) = \text{Exp}(p(t), x_iV(t)),$$
$$\epsilon_i(t) = \text{Log}\,(\eta_i(t), y_i)\,.$$

Using this notation, we can write the geodesic regression energy from (4) as

$$E(p(t),\ V(t)) = \frac{1}{2}\sum_{i=1}^{N} \|\epsilon_i(t)\|^2.$$

Taking derivatives in t, we have

$$\frac{dE}{dt} = -\sum_{i=1}^{N}\left\langle \epsilon_i, \frac{d\eta_i}{dt}\right\rangle,$$

$$\frac{d^2E}{dt^2} = \sum_{i=1}^{N} -\left\langle \frac{D\epsilon_i}{dt}, \frac{d\eta_i}{dt}\right\rangle - \left\langle \epsilon_i, \frac{D}{dt}\frac{d\eta_i}{dt}\right\rangle. \tag{11}$$

To prove this second derivative is positive at $t = 0$, we will show that each term inside the sum is positive. As before, we consider the geodesic variation given by

$$\alpha(s,t) = \mathrm{Exp}(p(t),\ sx_iV(t)).$$

Define the coordinate vector field $J_i = d\alpha/dt$. Notice that the derivatives of η_i are

$$\frac{d\eta_i}{dt}(0) = J_i(1,0), \qquad \frac{D}{dt}\frac{d\eta_i}{dt}(0) = \frac{DJ_i}{dt}(1,0),$$

and this second derivative can be computed using the result from Theorem 1 (taking care to account for the factor of x_i).

The derivative of ϵ_i has the following form

$$\frac{D\epsilon_i}{dt}(0) = d_1\mathrm{Log}(\eta_i(0), y_i)\frac{d\eta_i}{dt}(0) = \nabla_{\epsilon_i} Q(0)$$

where $d_1\mathrm{Log}$ denotes the differential of Log with respect to the first parameter. The derivative of Log above can be computed as the derivative of a Jacobi field Q along the geodesic from $\eta_i(t)$ to y_i, with boundary conditions $Q(0) = (d\eta_i/dt)(0)$ and $Q(1) = 0$. (It helps to think of this derivative as the "inverse" of the process we used to compute the derivative of Exp with respect to its first argument, see Fig. 2.)

With these pieces in place, we are ready to prove convexity of the geodesic regression least-squares problem. We begin with the positive curvature case.

Theorem 2 *Let M be a symmetric space with positive sectional curvature bounded above by $K_M > 0$. Define $r = \pi/(4\sqrt{K_M})$. Let $x_1, \ldots, x_N \in \mathbb{R}$, not all equal, and $y_1, \ldots, y_N \in M$, where the y_i are contained in a geodesically convex set $A \subset M$ with $\mathrm{diam}(A) < r$. Then the least-squares estimates, $\hat{p}, \hat{v}$, given by (5), exist and are unique within the class of r-close geodesics.*

Theorem 3 *Let M be a symmetric space with negative sectional curvature bounded below by $K_M < 0$. Define $r = 1/\sqrt{|K_M|}$. Let $x_1, \ldots, x_N \in \mathbb{R}$, not all equal, and $y_1, \ldots, y_N \in M$, where the y_i are contained in a geodesically convex set $A \subset M$ with $\mathrm{diam}(A) < r$. Then the least-squares estimates, $\hat{p}, \hat{v}$, given by (5), exist and are unique within the class of r-close geodesics.*

Proof (Theorem 2 and 3) We first note that the set of $(p, v) \in TM$ that generate the r-close geodesics to the data forms a compact subset of TM. This set is also nonempty because the point $(\mu, 0) \in TM$, where μ is the Fréchet mean of the y_i, generates an r-close geodesic (of length zero). The Fréchet mean of the y_i exists thanks to the result by Karcher (1977), using the bounds on the diameter of the data in the positive curvature case.

To prove that (11) is greater than zero, it suffices to show that each term inside the sum is positive. Thus, we consider only one data point at a time, and drop the subscript i for convenience. Also, let T, E_j be an orthonormal frame as in the setup for (10). It is clear that we only need to consider one E_j at a time, and we thus drop the j subscript as well, and work in the dimensions spanned by the geodesic tangent, T, and a single normal vector, E. Thus, we let $J(s, 0)$ denotes the Jacobi field corresponding to the geodesic variation α, projected onto the plane spanned by E and T.

Following the same Jacobi field decomposition as (10), we write Q as

$$Q(s) = \sum_{i=2}^{n}\{f_i(s)A_i(s) + g_i(s)B_i(s)\} + C(s) + sD(s),$$

for appropriate parallel vector fields A_i, B_i, C, D, where f_i, g_i are defined as above for the corresponding sectional curvatures, κ_i. Note that the basis vectors for Q and J are not the same because they live along different geodesics. Given the boundary conditions $Q(0) = (d\eta/dt)(0) = J(1, 0)$ and $Q(1) = 0$, we have the following relationship for the normal and tangential components of $(D\epsilon/dt)(0) = \nabla_\epsilon Q(0) = \sum_i B_i(0) + D(0)$,

$$B_i(0) = -\frac{f_i(1)}{g_i(1)}A_i(0), \qquad D(0) = -C(0).$$

Now, the first term in (11) (evaluated at $t = 0$) is

$$-\left\langle \frac{D\epsilon}{dt}, \frac{d\eta}{dt}\right\rangle\Big|_{t=0} = \sum_{i=2}^{n}\frac{f_i(1)}{g_i(1)}\|A_i(0)\|^2 + \|C(0)\|^2.$$

Focusing on the positive curvature case (Theorem 2), we derive a lower bound for the first term in (11). Define the function $h(x) = x\cos(x)/\sin(x)$. It is straightforward to verify that in the range $x \in (0, \pi/4)$, we have $h(x) > \pi/4$. Combining this with the constraint on the diameter of the data, which tells us that $\|\epsilon\| < \pi/(4\sqrt{\kappa_i})$, we arrive at the bound,

$$\frac{f_i(1)}{g_i(1)} = \frac{\sqrt{\kappa_i}\|\epsilon\|\cos(\sqrt{\kappa_i}\|\epsilon\|)}{\sin(\sqrt{\kappa_i}\|\epsilon\|)} > \frac{\pi}{4}, \quad \text{when } K_M > 0.$$

Combining this with $Q(0) = \sum_i A_i(0) + C(0)$, we have the following bound on the first term in (11),

$$-\left\langle \frac{D\epsilon}{dt}, \frac{d\eta}{dt} \right\rangle \Big|_{t=0} > \frac{\pi}{4} \|Q(0)\|^2, \quad \text{when } K_M > 0. \tag{12}$$

Similarly, in the case of negative curvature (Theorem 3), we use the fact that $x \cosh(x)/\sinh(x) > 1$. This can be verified from the Taylor series of sinh and cosh . This gives us the bound,

$$-\left\langle \frac{D\epsilon}{dt}, \frac{d\eta}{dt} \right\rangle \Big|_{t=0} > \|Q(0)\|^2, \quad \text{when } K_M < 0. \tag{13}$$

The magnitude of the second term in (11) is bounded by

$$\left| \left\langle \epsilon, \frac{D}{dt}\frac{d\eta}{dt} \right\rangle \right|_{t=0} < |a||\langle \epsilon, T\rangle| + |b||\langle \epsilon, E\rangle|, \tag{14}$$

where a, b are the factors in the second derivative of the exponential map given in Theorem 1. The proof proceeds by comparing common factors in the bounds on the two pieces of (11), i.e., showing the right-hand side of (14) is less than the left-hand side of (12) or (13), depending on the curvature. For example, collecting the terms that contain the factor X^2 in the positive curvature case, we get the following:

$$\frac{\pi}{4} \cos(\sqrt{\kappa}L)^2 X^2 > \sqrt{\kappa}\|\epsilon\| \sin(\sqrt{\kappa}L)\cos(\sqrt{\kappa}L) X^2.$$

This inequality can be seen by using $\cos(x) > \sin(x)$ for $x \in [0, \pi/4)$ and $\sqrt{\kappa}\|\epsilon\| < \pi/4$. The other terms that appear in the derivation proceed in a similar fashion. □

4.4 Maximum Likelihood

In Euclidean spaces, least-squares estimation is equivalent to maximum likelihood of the model parameters under an assumption of Gaussian errors. In this section, we extend this connection to symmetric spaces by formulating the geodesic regression least-squares problem as a maximum likelihood estimation of the model parameters. We begin by defining a suitable generalization of a "Gaussian" distribution on a Riemannian manifold. Several previous works have defined generalizations of Gaussian distributions on manifolds. Grenander (1963) defines a generalization of the Gaussian distribution to Lie groups and homogeneous spaces as a solution to the heat equation. Pennec (2006) defines a generalization of the Gaussian distribution in the tangent space to the mean (up to the cut locus of the mean).

We will use the following pdf on a connected Riemmanian manifold M.

$$f(y; \mu, \sigma) = \frac{1}{C(\mu, \sigma)} \exp\left(-\frac{d(y, \mu)^2}{2\sigma^2}\right), \tag{15}$$

where

$$C(\mu, \sigma) = \int_M \exp\left(-\frac{d(y, \mu)^2}{2\sigma^2}\right) dy.$$

This formulation is related to the one given by Pennec (2006). A key difference is that the above pdf is not restricted to be truncated at the cut locus. This is possible because the Riemannian distance function is defined past the cut locus, even when the log map is not unique or not defined. Also, our pdf does not include a covariance matrix, which would require again mapping to the tangent space at the mean. In some sense, this makes the definition above more intrinsic to the geometry defined by the Riemannian metric, at the cost of some flexibility.

We will say that the geodesic regression model has *Gaussian errors* if the random variable Y is conditionally distributed according to (15), that is, if

$$p(y|X = x) = f(y; \text{Exp}(p, xv), \sigma). \tag{16}$$

The log-likelihood under this distribution is given by

$$l(p, v, \sigma; \{(x_i, y_i)\}) = N \log C(\text{Exp}(p, x_i v), \sigma) - \sum_{i=1}^{N} \frac{1}{2\sigma^2} d(y_i, \text{Exp}(p, x_i v))^2.$$

Theorem 4 *Let M be a symmetric space, $x_1, \ldots, x_N \in \mathbb{R}$, and $y_1, \ldots, y_N \in M$, such that the criteria in Theorems 2 and 3 for existence and uniqueness of the least-squares estimates hold. The solution to the least-squares problem defined by (5) is equivalent to the maximum likelihood estimator of the geodesic regression model with Gaussian errors, i.e., if Y is assumed to be conditionally distributed by (16).*

Proof First, notice that the normalizing factor $C(\mu, \sigma)$ in (15) is constant in μ. This is because the integral defining it is invariant to isometries (being defined purely in terms of the distance), and because M is a symmetric space, we can move the distribution from μ to any other point by an isometry.

Now, the gradients of the log-likelihood function with respect to the parameters p and v are

$$\nabla_p l = \frac{1}{\sigma^2} \sum_{i=1}^{N} d_p \text{Exp}(p, x_i v)^\dagger \text{Log}(\text{Exp}(p, x_i v) y_i)$$

$$\nabla_v l = \frac{1}{\sigma^2} \sum_{i=1}^{N} x_i\, d_v \text{Exp}(p, x_i v)^\dagger \text{Log}(\text{Exp}(p, x_i v), y_i).$$

But these are equivalent to the negatives of the gradients (7), (8) of the least-squares problem (with a constant scaling by σ^{-2}). Therefore, both problems attain extrema at the same values. □

It's important to note that for general Riemannian manifolds, this result will not hold. The proof above uses the homogeneity property of a symmetric space to show that the normalizing factor $C(\mu, \sigma)$ is constant in μ. This will not happen for Riemannian manifolds that are not homogeneous, causing the gradient of the log-likelihood to possibly be non-zero at the least-squares estimate.

5 Experimental Results

5.1 Regression of 3D Rotations

5.1.1 Overview of Unit Quaternions

We represent 3D rotations as the unit quaternions, $\mathbb{Q}_1$. A quaternion is denoted as $q = (a, v)$, where a is the "real" component and $v = bi + cj + dk$. Geodesics in the rotation group are given simply by constant speed rotations about a fixed axis. Let $e = (1, 0)$ be the identity quaternion. The tangent space $T_e\mathbb{Q}_1$ is the vector space of quaternions of the form $(0, v)$. The tangent space at an arbitrary point $q \in \mathbb{Q}_1$ is given by right multiplication of $T_e\mathbb{Q}_1$ by q. The Riemannian exponential map is $\mathrm{Exp}_q((0, v) \cdot q) = (\cos(\theta/2),\ 2v \cdot \sin(\theta/2)/\theta) \cdot q$, where $\theta = 2\|v\|$. The log map is given by $\mathrm{Log}_q((a, v) \cdot q) = (0, \theta v/\|v\|) \cdot q$, where $\theta = \arccos(a)$.

Being a unit sphere, $\mathbb{Q}_1$ has constant sectional curvature $K = 1$. In this case the orthogonal component of the Jacobi field equation (6) along a geodesic $\gamma(t)$ has the analytic solution

$$J(t)^{\perp} = u_1(t)\cos(Lt) + u_2(t)\frac{\sin(Lt)}{L},$$

where u_1, u_2 are parallel vector fields along γ, with initial conditions $u_1(0) = J(0)^{\perp}$ and $u_2(0) = J'(0)^{\perp}$, and $L = \|\gamma'\|$. While the Jacobi field equation gives us the differential of the exponential map, we really need the adjoint of this operator for geodesic regression. However, from the above equation it is clear that $d_p\mathrm{Exp}$ and $d_v\mathrm{Exp}$ are both self-adjoint operators. That is, the above Jacobi field equation provides us both the differential and its adjoint.

5.1.2 Geodesic Regression of Simulated Rotation Data

To test the geodesic regression least squares estimation on $\mathbb{Q}_1$, synthetic rotation data was simulated according to the geodesic model (3). The intercept was the identity rotation: $p = (1, 0, 0, 0)$, and the slope was a rotation about the z-axis: $v = (0, 0, 0, \pi/4)$. The x_i data were drawn from a uniform distribution on $[0, 1]$. The errors in the model were generated from an isotropic Gaussian distribution in the tangent space, with $\sigma = \pi/8$. The resulting data (x_i, y_i) were used to compute estimates of the parameters $(\hat{p}, \hat{v})$. This experiment was repeated 1,000 times each for sample sizes $N = 2^k, k = 1, \ldots, 8$. We would expect that as the sample size increases, the mean squared error (MSE) in the estimates $(\hat{p}, \hat{v})$, relative to the true parameters, would approach zero. The MSE is defined as

$$\mathrm{MSE}(\hat{p}) = \frac{1}{M}\sum_{i=1}^{M} d(\hat{p}_i, p)^2,$$

$$\mathrm{MSE}(\hat{v}) = \frac{1}{M}\sum_{i=1}^{M} \|\hat{v}_i \cdot (\hat{p}_i^{-1} p) - v\|^2,$$

where $M = 1{,}000$ is the number of repeated trials, and $(\hat{p}_i, \hat{v}_i)$ is the estimate from the ith trial. Notice the multiplication by $(\hat{p}_i^{-1} p)$ in the second equation is a right-translation of $\hat{v}_i$ to the tangent space of p. Figure 3 shows plots of the resulting MSE for the slope and intercept estimates. As expected, the MSE approaches zero as sample size increases, indicating at least empirically that the least squares estimates are consistent.

5.2 Regression in Shape Spaces

One area of medical image analysis and computer vision that finds the most widespread use of Riemannian geometry is the analysis of shape. Dating back to the groundbreaking work of Kendall (1984) and Bookstein (1986), modern shape analysis is concerned with the geometry of objects that is invariant to rotation, translation, and scale. This typically results in representing an object's shape as a point in a nonlinear Riemannian manifold, or *shape space*. Recently, there has been a great amount of interest in Riemannian shape analysis, and several shape spaces for 2D and 3D objects have been proposed (Fletcher et al. 2003; Klassen et al. 2004; Michor and Mumford 2006; Younes 1998).

We choose here to use Kendall's shape space, but geodesic regression is applicable to other shape spaces as well. It could also be applied to spaces of diffeomorphisms, using the Jacobi field calculations given by Younes (2006). We note that Kendall's shape space is in some sense the most "agnostic" shape space available. By this we mean that there are no assumptions put on the geometry of the objects, other than that we want to remove the effects of translation, rotation, and scale. This simplicity is reflected in the fact that Kendall's shape space is a Riemannian symmetric space (and thus all the theory of Sect. 3 applies). More complicated assumptions on shapes, e.g., that they be smooth curves or surfaces, or that they be diffeomorphic to a fixed template, lead to more difficult shape spaces and computations. Of course, there are many cases where this added complexity is important for shape modeling. However, given the relative ease of computations in Kendall's shape space, and its well-developed statistical theory, it serves as a very useful shape space for many applications in statistical shape analysis.

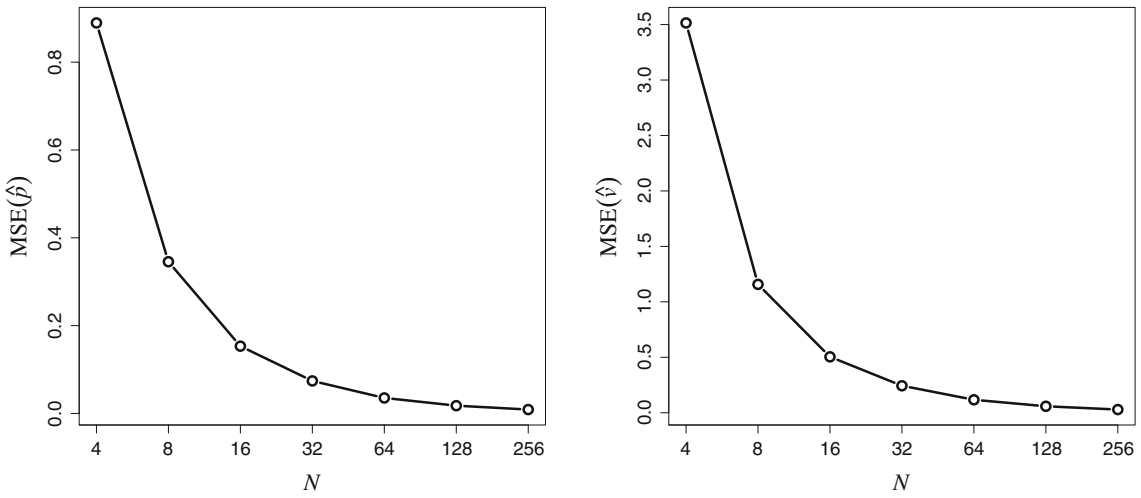

Fig. 3 Results for simulated rotation data: MSE of the geodesic regression estimates for the intercept (*left*) and slope (*right*) as a function of sample size

5.2.1 Overview of Kendall's Shape Space

We begin with derivations of the necessary computations for geodesic regression on Kendall's shape space. A configuration of k points in the 2D plane is considered as a complex k-vector, $z \in \mathbb{C}^k$. Removing translation, by requiring the centroid to be zero, projects this point to the linear complex subspace $V = \{z \in \mathbb{C}^k : \sum z_i = 0\}$, which is equivalent to the space $\mathbb{C}^{k-1}$. Next, points in this subspace are deemed equivalent if they are a rotation and scaling of each other, which can be represented as multiplication by a complex number, $\rho e^{i\theta}$, where ρ is the scaling factor and θ is the rotation angle. The set of such equivalence classes forms the complex projective space, $\mathbb{C}P^{k-2}$. As Kendall points out, there is no unique way to identify a shape with a specific point in complex projective space. However, if we consider that the geodesic regression problem only requires computation of exponential/log maps and Jacobi fields, we can formulate these computations without making an explicit identification of shapes with points in $\mathbb{C}P^{k-2}$.

Thus, we think of a centered shape $x \in V$ as representing the complex line $L_x = \{z \cdot x : z \in \mathbb{C} \backslash \{0\}\}$, i.e., L_x consists of all point configurations with the same shape as x. A tangent vector at $L_x \in V$ is a complex vector, $v \in V$, such that $\langle x, v \rangle = 0$. The exponential map is given by rotating (within V) the complex line L_x by the initial velocity v, that is,

$$\mathrm{Exp}_x(v) = \cos\theta \cdot x + \frac{\|x\| \sin\theta}{\theta} \cdot v, \quad \theta = \|v\|. \tag{17}$$

Likewise, the log map between two shapes $x, y \in V$ is given by finding the initial velocity of the rotation between the two complex lines L_x and L_y. Let $\pi_x(y) = x \cdot \langle x, y \rangle / \|x\|^2$ denote the projection of the vector y onto x. Then the log map is given by

$$\mathrm{Log}_x(y) = \frac{\theta \cdot (y - \pi_x(y))}{\|y - \pi_x(y)\|}, \quad \theta = \arccos \frac{|\langle x, y \rangle|}{\|x\| \|y\|}. \tag{18}$$

Notice that we never explicitly project a shape onto $\mathbb{C}P^{k-2}$. This has the effect that shapes computed via the exponential map (17) will have the same orientation and scale as the base point x. Also, tangent vectors computed via the log map (18) are valid only at the particular representation x (and not at a rotated or scaled version of x). This works nicely for our purposes and implies that shapes along the estimated geodesic will have the same orientation and scale as the intercept shape, $\hat{p}$.

The sectional curvature of $\mathbb{C}P^{k-2}$ can be computed as follows. Let u, w be orthonormal vectors at a point $p \in \mathbb{C}P^{k-2}$. These vectors may be thought of as vectors in $\mathbb{C}^{k-1} \cong \mathbb{R}^{2k-2}$. Writing the vector w as $w = (w_1, \ldots, w_{2k-2})$, define the operator

$$j(w) = (-w_k, \ldots, -w_{2k-2}, w_1, \ldots, w_{k-1}).$$

(This is just multiplication by $i = \sqrt{-1}$ if we take w as a complex vector with the $k-1$ real coordinates listed first.) Using this operator, the sectional curvature is given by

$$K(u, w) = 1 + 3\langle u, j(w) \rangle^2.$$

When $k = 3$, $\mathbb{C}P^1$ is the space of triangle shapes and is isomorphic to the sphere, S^2, with radius $1/2$, and thus has constant sectional curvature, $K = 4$. For $k > 3$, $\mathbb{C}P^{k-2}$ has sectional curvature in the interval $K \in [1, 4]$. Furthermore, let $u \in T_p\mathbb{C}P^{k-2}$ be any unit length vector. If we decompose the tangent space into an orthonormal basis $e_1, \ldots, e_{2k-2}$, such that $e_1 = j(u)$, then we have $K(u, e_1) = 4$ and $K(u, e_i) = 1$ for $i > 1$. This leads to the following procedure for computing the Jacobi field equation on $\mathbb{C}P^{k-2}$

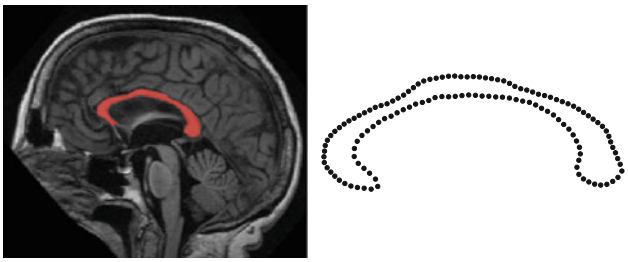

Fig. 4 Corpus callosum segmentation and boundary point model for one subject

along a geodesic γ. Given initial conditions for $J(0)^{\perp}$ and $J'(0)^{\perp}$, decompose $J(0)^{\perp} = u_1 + w_1$, so that u_1 is orthogonal to $j(\gamma')$ and w_1 is tangential to $j(\gamma')$. Do the same for $J'(0)^{\perp} = u_2 + w_2$. As before, extend these vectors to parallel fields, $u_i(t)$, $w_i(t)$, along γ. Then the orthogonal component of the Jacobi field along γ is given by

$$J(t)^{\perp} = u_1(t)\cos(Lt) + u_2(t)\frac{\sin(Lt)}{L} + w_1(t)\cos(2Lt) + w_2(t)\frac{\sin(2Lt)}{2L}.$$

As was the case for rotations, both $d_p\text{Exp}$ and $d_v\text{Exp}$ are self-adjoint operators.

5.2.2 Application to Corpus Callosum Aging

The corpus callosum is the major white matter bundle connecting the two hemispheres of the brain. A midsagittal slice from a magnetic resonance image (MRI) with segmented corpus callosum is shown in Fig. 4. Several studies have shown that the volume of the corpus callosum decreases with normal aging (Driesen and Raz 1995). However, less is known about how the *shape* of the corpus callosum changes with age. Understanding shape changes may provide a deeper understanding of the anatomical and biological processes underlying aging. For example, does the corpus callosum shrink uniformly in size, or do certain regions deteriorate faster than others? This type of question can be answered by geodesic regression in shape spaces.

To understand age-related changes in the shape of the corpus callosum, geodesic regression was applied to corpus callosum shape data derived from the OASIS database (www.oasis-brains.org). The data consisted of MRI from 32 subjects with ages ranging from 19 to 90 years old. The corpus callosum was segmented in a midsagittal slice using the ITK SNAP program (www.itksnap.org). The boundaries of these segmentations were sampled with 128 points using ShapeWorks (www.sci.utah.edu/software.html). This algorithm generates a sampling of a set of shape boundaries while enforcing correspondences between different point models within the population. An example of a segmented corpus callosum and the resulting boundary point model is shown in Fig. 4. The entire collection of input shapes and their ages is

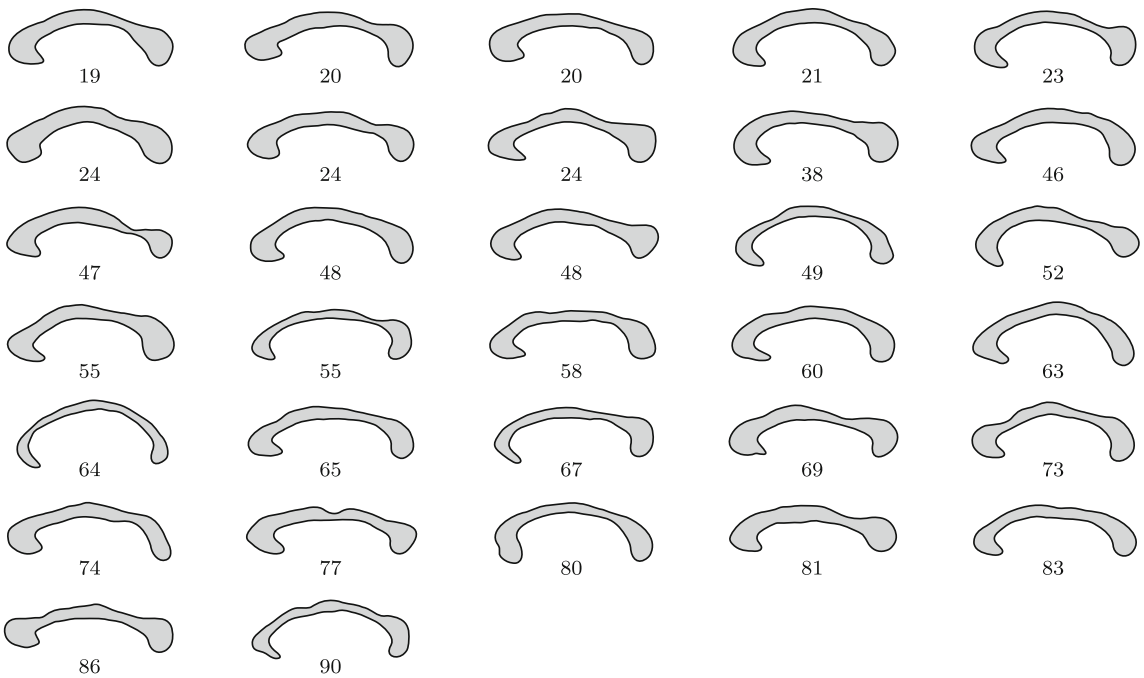

Fig. 5 The input corpus callosum shape data and corresponding subject ages in years

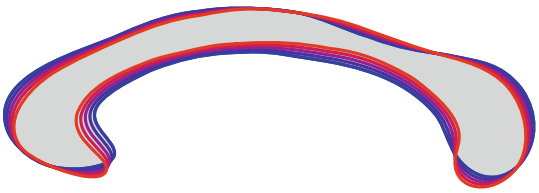

Fig. 6 Geodesic regression of the corpus callosum. The estimated geodesic is shown as a sequence of shapes from age 19 (*blue*) to age 90 (*red*)

shown in Fig. 5 (boundary points have been connected into a boundary curve for visualization purposes). Each of these preprocessing steps were done without consideration of the subject age, to avoid any bias in the data generation.

Geodesic regression was applied to the data (x_i, y_i), where x_i was the ith subject's age, and y_i was the ith subject's corpus callosum, generated as above and represented as a point in Kendall's shape space. First, the average age of the group, $\bar{x}$, was subtracted from each x_i, which was done to make the intercept term correspond to the shape at the mean age, rather than the shape at age zero, which would be far outside the data range. Least squares estimates $(\hat{p}, \hat{v})$ were generated according to (5), and using the above calculations for $\mathbb{C}P^{k-2}$. The resulting estimated geodesic is shown in Fig. 6 as a sequence of shapes: $\hat{\gamma}(t_k) = \mathrm{Exp}(\hat{p}, (t_k - \bar{x})\hat{v})$, for $t_k = 19, 36, 54, 72, 90$. The shape trend shows a very clear thinning of the corpus callosum, with the largest effects in the posterior part of the body and in the genu (anterior end).

Finally, the statistical significance of the estimated trend was tested using the permutation test described in Sect. 3.2, using 10,000 permutations. The p-value for the significance of the slope estimate, $\hat{v}$, was $p = 0.009$. The coefficient of determination (for the unpermuted data) was $R^2 = 0.12$. The low R^2 value must be interpreted carefully. It says that age only describes a small fraction of the shape variability in the corpus callosum. This is not surprising: we would expect the intersubject variability in corpus callosum shape to be difficult to fully describe with a single variable (age). However, this does not mean that the age effects are not important. In fact, the low p-value says that the estimated age changes are highly unlikely to have been found by random chance.

6 Discussion

This paper presented a method for geodesic regression of data on a Riemannian manifold. A least-squares criteria for geodesic regression was formulated using using intrinsic Riemannian distances, and a gradient decent algorithm for its computation involving Jacobi fields was developed. While these methods are applicable for data on general Riemannian manifolds, several theoretical properties were derived for the important class of symmetric spaces. The two main contributions to the theory of geodesic regression on symmetric spaces are (1) a guarantee of local existence and uniqueness to the least-squares geodesic regression problem and (2) a formulation of the least-squares problem as a maximum-likelihood estimation under an appropriate definition of Gaussian error distributions. Another significant contribution was the derivation of a formula for the second derivative of the exponential map on symmetric spaces (Theorem 1).

There are several potential avenues for future work. First, the theoretical contributions are limited to symmetric spaces, and it would be of great interest to see if these results could be extended to more general classes of Riemannian manifolds. This seems to be, however, a difficult proposition. A main step in the proof of the local convexity of the least-squares problem involves the computation of the second derivative of the exponential map, which turned out to be computed in closed-form for symmetric spaces. In order to extend this to more general manifolds, one would need to prove a bound on the second derivative of the exponential map.

Even in the case of symmetric spaces, there is likely room for improvement of the convexity results. For example, are the curvature-based bounds on the diameter of the data tight, or would it be possible to guarantee convexity under a wider spread of the data? It is instructive to compare these bounds for geodesic regression to those for the Fréchet mean. In the positive curvature case, the diameter is constrained to a factor of $\pi/(4\sqrt{K})$, which is equivalent to the bound for the mean developed by Karcher (1977). However, this bound for the mean was improved to $\pi/(2\sqrt{K})$ by Kendall (1990), and it would be interesting to see if this same improvement could be achieved for geodesic regression. The bound for the negative curvature case is somewhat more surprising because there is no bound for the mean in this case, i.e., the least-squares problem for the Fréchet mean is globally convex in the negative curvature case (assuming the manifold is geodesically convex). However, the geodesic regression problem can certainly fail to be convex with large enough distances between the geodesic and the data point. Another possible area for improvement is in the condition that the solution be in the class of r-close geodesics. The drawback of this constraint is that it is feasible that a *better* solution may exist outside the class of r-close geodesics.

There is also much to be done to develop the practical application of geodesic regression. For example, one can imagine diagnostics for regression fits, such as testing for outliers or influential points, non-geodesicity, or non-Gaussianity. This would get to the question of whether a geodesic curve is an appropriate model for a particular data set. Another practical extension would be the development of hypothesis tests for comparing the geodesic trends of two populations. Also, two very recent extensions of this

work are longitudinal models based on geodesic trajectories (Muralidharan and Fletcher 2012) and higher-order polynomial regression (Hinkle et al. 2012), of which, geodesic regression is a special case. Another possible extension to a more flexible class of regression curves would be a geodesic model including a nonlinear reparameterization of time, such as a growth curve. For instance, this would allow shape changes that asymptotically slow down, rather than proceed at a constant pace.

Acknowledgments This work was supported by NSF CAREER Grant 1054057.

Appendix A: Proof of Theorem 1

When $\kappa = 0$, it is clear that the vector field J changes linearly in t, giving the desired result $DJ/dt = 0$. Therefore, it suffices to consider only the cases where $\kappa \neq 0$. Let's first consider the case where $\kappa > 0$. Then we can write

$$\begin{aligned} J(s,t) &= \cos\left(s\sqrt{\kappa}L(t)\right) X(t)E(s,t) \\ &\quad + \frac{\sin\left(s\sqrt{\kappa}L(t)\right)}{\sqrt{\kappa}} Y(t)E(s,t) \\ &\quad + Z(t)T(s,t) + sW(t)T(s,t). \end{aligned}$$

Our goal is to compute the normal and tangential components of DJ/dt. We will use the identities

$$\begin{aligned} \left\langle \frac{DJ}{dt}, T \right\rangle &= \frac{1}{L}\left\langle \frac{DJ}{dt}, V \right\rangle \\ &= \frac{1}{L}\left(\frac{d}{dt}\langle J, V\rangle - \left\langle J, \frac{DV}{dt}\right\rangle \right). \end{aligned} \tag{19}$$

$$\left\langle \frac{DJ}{dt}, E \right\rangle = \frac{d}{dt}\langle J, E\rangle - \left\langle J, \frac{DE}{dt}\right\rangle, \tag{20}$$

Tangential Component: We start by noting that

$$\frac{DV}{dt} = \frac{D}{dt}\frac{d\alpha}{ds} = \frac{D}{ds}\frac{d\alpha}{dt} = \frac{DJ}{ds},$$

and we can compute, using $k = \sqrt{\kappa}L$,

$$\frac{DV}{dt} = \frac{DJ}{ds} = -k\sin(sk)XE + \cos(sk)YE + WT. \tag{21}$$

This gives the second term in (19),

$$\begin{aligned} \left\langle J, \frac{DV}{dt}\right\rangle &= \frac{1}{2}\sin(2sk)\left(\frac{1}{k}Y^2 - kX^2\right) \\ &\quad + \cos(2sk)XY + WZ + sW^2. \end{aligned}$$

The first term in (19) is given by

$$\begin{aligned} \frac{d}{dt}\langle J, V\rangle &= \frac{d}{dt}(LZ + sLW) \\ &= \frac{dL}{dt}(Z + sW) + L\frac{dZ}{dt} + sL\frac{dW}{dt}. \end{aligned}$$

We now compute each of the derivatives in this equation. Starting with dL/dt, and using the previous result in (21), we get

$$\frac{dL}{dt} = \frac{d}{dt}\|V\| = \frac{1}{L}\left\langle \frac{DV}{dt}, V\right\rangle = W.$$

Using the fact that T is a unit vector field, we get

$$\begin{aligned} \frac{DT}{dt} &= \left\langle \frac{DT}{dt}, E\right\rangle E = -\frac{1}{L}\left\langle \frac{DV}{dt}, E\right\rangle E \\ &= \left(\sin\left(s\sqrt{\kappa}L\right)X - \frac{\cos\left(s\sqrt{\kappa}L\right)}{L}Y\right)E. \end{aligned}$$

Evaluating this at $s = 0$, gives

$$\frac{DT}{dt}(0,t) = -\frac{Y(t)}{L(t)}E(0,t)$$

Denoting τ_t as parallel translation along $p(t)$, we can write

$$\begin{aligned} Z(t) &= \langle Z(0)\,\tau_t T(0,0) + X(0)\,\tau_t E(0,0),\ T(0,t)\rangle, \\ W(t) &= \langle W(0)\,\tau_t T(0,0) + Y(0)\,\tau_t E(0,0),\ T(0,t)\rangle. \end{aligned}$$

Using this, and the fact that Z, W are constant in s, allows us to compute

$$\begin{aligned} \frac{dZ}{dt} &= -\frac{XY}{L}, \\ \frac{dW}{dt} &= -\frac{Y^2}{L}. \end{aligned}$$

We put these together to get

$$\frac{d}{dt}\langle J, V\rangle = WZ - XY + sW^2 - sY^2.$$

Finally, the tangential component of DJ/dt is given by

$$\begin{aligned} \left\langle \frac{DJ}{dt}, T\right\rangle &= \frac{\sqrt{\kappa}}{2}\sin(2sL)\left(X^2 - \frac{Y^2}{\kappa L^2}\right) \\ &\quad - \frac{\cos\left(2s\sqrt{\kappa}L\right)}{L}XY \\ &\quad + \frac{sY^2}{L} + \frac{XY}{L}. \end{aligned} \tag{22}$$

Normal Component: Similar to the computation above for Z, W, but now for the normal components X, Y, we get

$$\begin{aligned} X(t) &= \langle X(0)\,\tau_t E(0,0) + Z(0)\,\tau_t T(0,0),\ E(0,t)\rangle, \\ Y(t) &= \langle Y(0)\,\tau_t E(0,0) + W(0)\tau_t T(0,0),\ E(0,t)\rangle. \end{aligned}$$

Using the fact that E is a unit vector field, we get

$$\frac{DE}{dt} = \left\langle \frac{DE}{dt}, T \right\rangle T = -\frac{1}{L}\left\langle \frac{DV}{dt}, E \right\rangle T$$
$$= \left(\sin(sk)X - \frac{\cos(sk)}{L}Y\right)T.$$

Evaluating this at $s = 0$, gives

$$\frac{DE}{dt}(0,t) = -\frac{Y(t)}{L(t)}T(0,t)$$

Again, using the fact that X, Y are constant in s, this gives us

$$\frac{dX}{dt} = -\frac{YZ}{L},$$
$$\frac{dY}{dt} = -\frac{YW}{L}.$$

The first term in (20) is calculated as

$$\frac{d}{dt}\langle J, E\rangle = \frac{d}{dt}\left(\cos\left(s\sqrt{\kappa}L\right)X + \frac{\sin\left(s\sqrt{\kappa}L\right)}{\sqrt{\kappa}L}Y\right)$$
$$= -s\sqrt{\kappa}\sin\left(s\sqrt{\kappa}L\right)XW - \frac{\cos\left(s\sqrt{\kappa}L\right)}{L}YZ$$
$$+ \frac{s\sqrt{\kappa}L\cos\left(s\sqrt{\kappa}L\right) - 2\sin\left(s\sqrt{\kappa}L\right)}{\sqrt{\kappa}L^2}YW.$$

Again, using the fact that E is a unit vector field, we have

$$\frac{DE}{dt} = \frac{1}{L}\left\langle \frac{DE}{dt}, V \right\rangle T = -\frac{1}{L}\left\langle E, \frac{DV}{dt} \right\rangle T$$
$$= \left(\sqrt{\kappa}\sin\left(s\sqrt{\kappa}L\right)X - \frac{\cos\left(s\sqrt{\kappa}L\right)}{L}Y\right)T$$

The second term in (20) is now given by

$$\left\langle J, \frac{DE}{dt} \right\rangle = \left(\sqrt{\kappa}\sin\left(s\sqrt{\kappa}L\right)X - \frac{\cos\left(s\sqrt{\kappa}L\right)}{L}Y\right)$$
$$\times (Z + sW).$$

Putting this together, we get the normal component of DJ/dt to be

$$\left\langle \frac{DJ}{dt}, E \right\rangle = -2s\sqrt{\kappa}\sin\left(s\sqrt{\kappa}L\right)XW$$
$$-\sqrt{\kappa}\sin\left(s\sqrt{\kappa}L\right)XZ$$
$$+\frac{2s}{L}\cos\left(s\sqrt{\kappa}L\right)YW$$
$$-\frac{2}{\sqrt{\kappa}L^2}\sin\left(s\sqrt{\kappa}L\right)YW. \tag{23}$$

Negative Sectional Curvature: Now consider the case when the sectional curvature is negative, i.e., $\kappa < 0$. The Jacobi field is given by

$$J(s,t) = \cosh\left(s\sqrt{-\kappa}L\right)X(t)E(s,t)$$
$$+\frac{\sinh\left(s\sqrt{-\kappa}L\right)}{\sqrt{-\kappa}L}Y(t)E(s,t)$$
$$+Z(t)T(s,t) + sW(t)T(s,t).$$

The derivation of DJ/dt in this case proceeds almost identically to the positive curvature case, taking care to handle the sign difference when differentiating cosh . The result is

$$\left\langle \frac{DJ}{dt}, T \right\rangle = \frac{\sqrt{-\kappa}}{2}\sinh\left(2s\sqrt{-\kappa}L\right)\left(X^2 - \frac{Y^2}{\kappa L^2}\right)$$
$$+\frac{\cosh\left(2s\sqrt{-\kappa}L\right)}{L}XY$$
$$-s\frac{Y^2}{L} - \frac{XY}{L} \tag{24}$$

$$\left\langle \frac{DJ}{dt}, E \right\rangle = 2s\sqrt{-\kappa}\sinh\left(s\sqrt{-\kappa}L\right)XW$$
$$+\sqrt{-\kappa}\sinh\left(s\sqrt{-\kappa}L\right)XZ$$
$$+\frac{2s}{L}\cosh\left(s\sqrt{-\kappa}L\right)YW$$
$$-\frac{2}{\sqrt{\kappa}L^2}\sin\left(s\sqrt{-\kappa}L\right)YW. \tag{25}$$

The final formulas for the second derivative of the exponential map are given by evaluation at $(s,t) = (1,0)$ in Eqs. (22)–(25). □

References

Bookstein, F. L. (1986). Size and shape spaces for landmark data in two dimensions (with discussion). *Statistical Science*, *1*(2), 181–242.

Boothby, W. M. (1986). *An introduction to differentiable manifolds and Riemannian geometry* (2nd ed.). London: Academic Press.

do Carmo, M. (1992). *Riemannian geometry*. Boston: Birkhäuser.

Davis, B., Fletcher, P. T., Bullitt, E., & Joshi, S. (2007). *Population shape regression from random design data*. In: Proceedings of IEEE international conference on computer vision. Rio de Janeiro: IEEE.

Driesen, N., & Raz, N. (1995). The influence of sex, age, and handedness on corpus callosum morphology: A meta-analysis. *Psychobiology*, *23*(3), 240–247.

Dryden, I., & Mardia, K. (1998). *Statistical shape analysis*. New York: John Wiley and Sons.

Durrleman, S., Pennec, X., Trouvé, A., Gerig, G., & Ayache, N. (2009). *Spatiotemporal atlas estimation for developmental delay detection in longitudinal datasets* (pp. 297–304). In: Medical image computing and computer-assisted Intervention intervention. Heidelberg: Springer.

Fletcher, P. T. (2011). *Geodesic regression on Riemannian manifolds* (pp. 75–86). In: MICCAI workshop on mathematical foundations of computational anatomy. Toronto: MFCA.

Fletcher, P. T., Lu, C., & Joshi, S. (2003). *Statistics of shape via principal geodesic analysis on lie groups* (pp. 95–101). In: IEEE CVPR. Madison: IEEE.

Fréchet, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'Institut Henri Poincaré*, *10*(3), 215–310.

Grenander, U. (1963). *Probabilities on algebraic structures*. New York: John Wiley and Sons.

Helgason, S. (1978). *Differential geometry, lie groups, and symmetric spaces*. New York: Academic Press.

Hinkle, J., Fletcher, P. T., Muralidharan, P., Joshi, S., & (2012). *Polynomial regression on Riemannian manifolds*. In: European conference on computer vision. Florence: ECCV.

Huckemann, S., Hotz, T., & Munk, A. (2010). Intrinsic shape analysis: Geodesic principal component analysis for Riemannian manifolds modulo lie group actions. *Statistica Sinica*, *20*, 1–100.

Jupp, P. E., & Kent, J. T. (1987). Fitting smooth paths to spherical data. *Applied Statistics*, *36*(1), 34–46.

Karcher, H. (1977). Riemannian center of mass and mollifier smoothing. *Communication on Pure and Applied Mathematics*, *30*, 509–541.

Kendall, D. G. (1984). Shape manifolds, procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society*, *16*, 18–121.

Kendall, W. S. (1990). Probability, convexity, and harmonic maps with small image I: Uniqueness and fine existence. *Proceedings of the London Mathematical Society*, *3*(61), 371–406.

Kenobi, K., Dryden, I. L., & Le, H. (2010). Shape curves and geodesic modelling. *Biometrika*, *97*(3), 567–584.

Klassen, E., Srivastava, A., Mio, W., & Joshi, S. (2004). Analysis of planar shapes using geodesic paths on shape spaces. *IEEE PAMI*, *26*(3), 372–383.

Kume, A., Dryden, I. L., & Le, H. (2007). Shape-space smoothing splines for planar landmark data. *Biometrika*, *94*(3), 513–528.

Mardia, K. V. (2000). *Directional statistics*. New York: John Wiley and Sons.

Michor, P. W., & Mumford, D. (2006). Riemannian geometries on spaces of plane curves. *Journal of European Mathematical Society*, *8*, 1–48.

Miller, M. (2004). Computational anatomy: Shape, growth, and atrophy comparison via diffeomorphisms. *NeuroImage*, *23*, S19–S33.

Muralidharan, P., & Fletcher, P. T. (2012). *Sasaki metrics for analysis of longitudinal data on manifolds*. In: IEEE conference on computer vision and pattern recognition. Providence: CVPR.

Niethammer, M., Huang, Y., & Viallard, F. X. (2011). *Geodesic regression for image time-series*. In: Proceedings of medical image computing and computer assisted intervention. Toronto: MICCAI.

Pennec, X. (2006). Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, *25*(1), 1–12.

Shi, X., Styner, M., Lieberman, J., Ibrahim, J., Lin, W., & Zhu, H. (2009). Intrinsic regression models for manifold-valued data. *Journal of American Statistical Association*, *5762*, 192–199.

Sommer, S., Lauze, F., Nielsen, M. (2010). The differential of the exponential map, Jacobi fields and exact principal geodesic, analysis. Retrieved October 7, 2010 from arXiv:10081902v3.

Trouvé, A., & Vialard, F. X. (2010). *A second-order model for time-dependent data interpolation: Splines on shape spaces*. In: MICCAI STIA workshop. Beijing: MICCAI.

Younes, L. (1998). Computable elastic distances between shapes. *SIAM Journal of Applied Mathematics*, *58*, 565–586.

Younes, L. (2006). Jacobi fields in groups of diffeomorphisms and applications. *Quarterly of Applied Mathematics*. *65*, 113–134.