

Chapter 2

Geodesic Regression and Its Application to Shape Analysis

P. Thomas Fletcher

Abstract In this chapter, I present a regression method for modeling the relationship between a manifold-valued random variable and a real-valued independent parameter. The principle is to fit a geodesic curve, parameterized by the independent parameter, that best fits the data. Error in the model is evaluated as the sum-of-squared geodesic distances from the model to the data, and this provides an intrinsic least squares criterion. Geodesic regression is, in some sense, the simplest parametric model that one could choose, and it provides a direct generalization of linear regression to the manifold setting. A generalization of the coefficient of determination and a resulting hypothesis test for determining the significance of the estimated trend is developed. Also, a diagnostic test for the quality of the fit of the estimated geodesic is demonstrated. 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.

2.1 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 widely 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

P.T. Fletcher (✉)
University of Utah, Salt Lake City, 84112 UT, USA
e-mail: fletcher@cs.utah.edu

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 [10] 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, \dots, 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 [12] provided conditions guaranteeing the existence and uniqueness of the Fréchet mean, which were later improved by Kendall [14]. Second-order statistics such as generalizations of principal components analysis [8] and Gaussian covariances [21] have also been developed and applied in the domain of image analysis. Related work includes statistical analysis of directional data (e.g., spheres) [16] and analysis on shape manifolds [5], 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 [11] propose an unrolling method on shape spaces. Regression analysis on the group of diffeomorphisms has been proposed as growth models by Miller [18], nonparametric regression by Davis et al. [2], and second-order splines by Trouvé and Vialard [23]. Durrleman et al. [6] construct spatiotemporal image atlases from longitudinal data. Finally, Shi et al. [22] 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 work is to develop such a generalization, called geodesic regression, which models the relationship between an independent scalar variable with 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 chapter is an expanded exposition of the geodesic regression method first introduced in [7]. Niethammer et al. [20] independently proposed geodesic regression for the case of diffeomorphic transformations of image time series.

2.2 Multiple Linear Regression

Before formulating geodesic regression on general manifolds, we begin by reviewing 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

$$Y = \alpha + X\beta + \epsilon, \quad (2.1)$$

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 $i = 1, \dots, N$, the least squares estimates, $\hat{\alpha}, \hat{\beta}$, for the intercept and slope are computed by solving the minimization problem

$$(\hat{\alpha}, \hat{\beta}) = \arg \min_{(\alpha, \beta)} \sum_{i=1}^N \|y_i - \alpha - x_i \beta\|^2. \quad (2.2)$$

This equation can be solved analytically, yielding

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

$$\hat{\alpha} = \bar{y} - \bar{x} \hat{\beta},$$

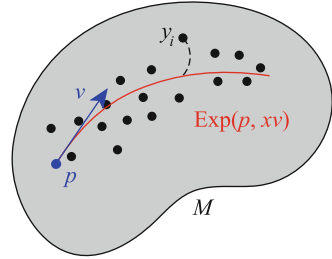
where \bar{x} and \bar{y} are the sample means of the x_i and y_i , respectively. If the errors in the model are drawn from distributions with zero mean and finite variance, then these estimators are unbiased and consistent. Furthermore, if the errors are homoscedastic (equal variance) and uncorrelated, then the Gauss-Markov theorem states that they will have minimal mean-squared error amongst all unbiased linear estimators.

2.3 Geodesic Regression

Let y_1, \dots, y_N be points on a smooth Riemannian manifold M , with associated scalar values $x_1, \dots, x_N \in \mathbb{R}$. 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. 2.1.

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 $(p, v) \in TM$,

Fig. 2.1 Schematic of the geodesic regression model



where p is a point in M and $v \in T_p M$ 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 $\text{Exp}_p(v) = \gamma(1)$. In other words, the exponential map takes a position and velocity as input and returns the point at time 1 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, $\text{Log}_p : V(p) \rightarrow T_p M$. For any point $q \in V(p)$ the Riemannian distance function is given by $d(p, q) = \|\text{Log}_p(q)\|$. It will be convenient to include the point p as a parameter in the exponential and log maps, i.e., define $\text{Exp}(p, v) = \text{Exp}_p(v)$ and $\text{Log}(p, q) = \text{Log}_p(q)$.

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 (2.1). 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*,

$$Y = \text{Exp}(\text{Exp}(p, Xv), \epsilon), \quad (2.3)$$

where ϵ is a random variable taking values in the tangent space at $\text{Exp}(p, Xv)$. Notice that for Euclidean space, the exponential map is simply addition, i.e., $\text{Exp}(p, v) = p + v$. Thus, the geodesic model coincides with (2.1) when $M = \mathbb{R}^n$.

2.3.1 Least Squares Estimation

Consider a realization of the model (2.3): $(x_i, y_i) \in \mathbb{R} \times M$, for $i = 1, \dots, N$. 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

$$E(p, v) = \frac{1}{2} \sum_{i=1}^N d(\text{Exp}(p, x_i v), y_i)^2. \quad (2.4)$$

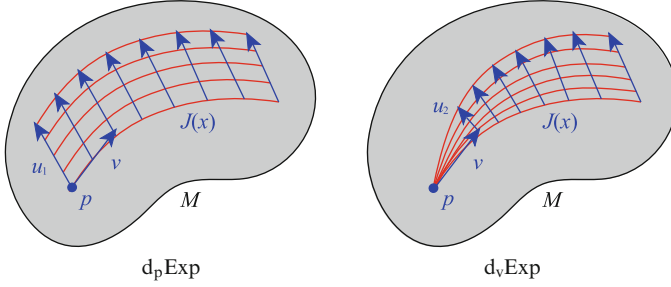


Fig. 2.2 Jacobi fields as derivatives of the exponential map

Following the ordinary least squares minimization problem given by (2.2), we formulate a least squares estimator of the geodesic model as a minimizer of the above sum-of-squares energy, i.e.,

$$(\hat{p}, \hat{v}) = \arg \min_{(p,v)} E(p, v). \tag{2.5}$$

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 (2.5) for a general manifold M will typically not yield an analytic solution. Instead we derive a gradient descent algorithm. Computation of the gradient of (2.4) 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 $\nabla_x d(p, x)^2 = -2\text{Log}_x(p)$ for $x \in V(p)$.

The derivative of the exponential map $\text{Exp}(p, v)$ 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 $c_1(s, t) = \text{Exp}(\text{Exp}(p, su_1), tv(s))$, where $u_1 \in T_p M$ defines a variation of the initial point along the geodesic $\eta(s) = \text{Exp}(p, su_1)$. Here we have also extended $v \in T_p M$ to a vector field $v(s)$ along η via parallel translation. This variation is illustrated on the left side of Fig. 2.2. Next consider a variation of geodesics $c_2(s, t) = \text{Exp}(p, su_2 + tv)$, where $u_2 \in T_p M$. (Technically, u_2 is a tangent to the tangent space, i.e., an element of $T_v(T_p M)$, but there is a natural isomorphism $T_v(T_p M) \cong T_p M$.) The variation c_2 produces a “fan” of geodesics as seen on the right side of Fig. 2.2.

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

$$d_p \text{Exp}(p, v) \cdot u_1 = \left. \frac{d}{ds} c_1(s, t) \right|_{s=0} = J_1(1)$$

$$d_v \text{Exp}(p, v) \cdot u_2 = \left. \frac{d}{ds} c_2(s, t) \right|_{s=0} = J_2(1),$$

where $J_i(t)$ are *Jacobi fields* along the geodesic $\gamma(t) = \text{Exp}(p, tv)$. Jacobi fields are solutions to the second order equation

$$\frac{D^2}{dt^2} J(t) + R(J(t), \gamma'(t)) \gamma'(t) = 0, \quad (2.6)$$

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 [3]. The initial conditions for the two Jacobi fields 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: $J^\top(t) = u_1^\top + tu_2^\top$. Therefore, the only challenge is to solve for the orthogonal component.

Finally, the gradient of the sum-of-squares energy in (2.4) is given by

$$\begin{aligned} \nabla_p E(p, v) &= - \sum_{i=1}^N d_p \text{Exp}(p, x_i v)^\dagger \text{Log}(\text{Exp}(p, x_i v), y_i), \\ \nabla_v E(p, v) &= - \sum_{i=1}^N x_i d_v \text{Exp}(p, x_i v)^\dagger \text{Log}(\text{Exp}(p, x_i v), y_i), \end{aligned}$$

where we have taken the adjoint of the exponential map derivative, e.g., defined by $\langle d_p \text{Exp}(p, v)u, w \rangle = \langle u, d_p \text{Exp}(p, v)^\dagger w \rangle$. 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.

2.3.2 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

$$\begin{aligned} \hat{y}_i &= \text{Exp}(\hat{p}, x_i \hat{v}), \\ \hat{\epsilon}_i &= \text{Log}(\hat{y}_i, y_i), \end{aligned}$$

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, $y_1, \dots, y_N \in M$, we use the Fréchet variance, intrinsically defined by

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

The unexplained variance is the variance of the residuals, $\text{var}(\hat{\epsilon}_i) = \frac{1}{N} \sum \|\hat{\epsilon}_i\|^2$. 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., $\text{var}(\hat{\epsilon}_i) = \frac{1}{N} \sum d(\hat{y}_i, y_i)^2$. Using these two variance definitions, the generalization of the R^2 statistic is then given by

$$R^2 = 1 - \frac{\text{unexplained variance}}{\text{total variance}} = 1 - \frac{\text{var}(\hat{\epsilon}_i)}{\text{var}(y_i)}. \quad (2.7)$$

Fréchet variance coincides with the standard definition of variance when $M = \mathbb{R}^n$. Therefore, it follows that the definition of R^2 in (2.7) 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., $\text{var}(\hat{\epsilon}_i) \leq \text{var}(y_i)$. This is because we could always choose \hat{p} to be the Fréchet mean and $v = 0$ to achieve $\text{var}(\hat{\epsilon}_i) = \text{var}(y_i)$. 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 (2.5), 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 $H_0 : R^2 = 0$, 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, \dots, 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.

2.4 Testing the Geodesic Fit

In any type of regression analysis, a choice is made as to the type of model that is fit to the data, whether it be linear, polynomial, or perhaps nonparametric. An important step in the analysis is to verify that the selected model is in fact

appropriate. In linear regression, for example, one would want to test several assumptions: (1) that the trend in the data is truly linear, (2) that the error is homoscedastic, (3) that the model fit is not led astray by outliers, (4) that the errors are Gaussian distributed, etc. Several graphical and quantitative heuristic tests have been developed to test these assumptions. For a detailed treatment of these methods, see [9].

In this section we develop a diagnostic test of the model assumptions for geodesic regression. We focus on the following question: is a geodesic curve an appropriate model for the relationship between the independent and dependent variables? A geodesic curve is, in some sense, the “straightest” path one can take on a manifold. This begs the question if a more flexible model would do a better job at fitting the data. This is analogous to the model selection problem for real-valued data when one is making the choice between a linear model and something more flexible, such as a higher-order polynomial model. Of course, if a model is made more flexible there is a danger that the data will be overfit. One way to test if a model is “flexible enough” is to plot the residuals of the model versus the independent variable. If the model has captured the relationship between the independent and dependent variables, then the residuals should show no obvious trend. If they do show a trend, then a more flexible model is needed to capture the relationship between the data. However, for regression on manifolds, this is a difficult test to apply because the residuals are high-dimensional tangent vectors and are thus difficult to plot versus the independent variable. One solution might be to plot the magnitude of the residuals instead, but this loses most of the information contained in the residual vectors.

Instead, we will use nonparametric regression as a comparison model to test if a geodesic is sufficient to capture the relationships in the data. Nonparametric regression models, such as the kernel regression method described below, are highly flexible. Because there is no parametric model assumed for the functional relationship between the independent and dependent variables, these models can adapt to highly complex functions given enough data. Given a method to visualize the results of a manifold-valued regression, the diagnostic test is as follows. First, compute both a geodesic regression and a nonparametric regression of the data. Second, visualize the results of both regression methods. If the nonparametric regression trend is similar to the estimated geodesic, then this provides strong evidence that the geodesic model is sufficient. If the nonparametric trend deviates significantly from the estimated geodesic, then this indicates that the geodesic model is too inflexible to capture the relationship between the two variables.

An example of this procedure is given for synthesized univariate data in Fig. 2.3. The left figure shows data generated from a noisy linear trend. In this case the linear model and the nonparametric model give similar answers. The right figure shows data generated from a noisy nonlinear (sinusoidal) trend. Here the nonparametric regression adapts to the nonlinearities in the data, and the inadequacy of the linear trend can be seen as a difference between the two regression models. Of course, in the univariate case we can easily see that a linear trend is inadequate just by plotting the data even without comparing it to a nonparametric regression.

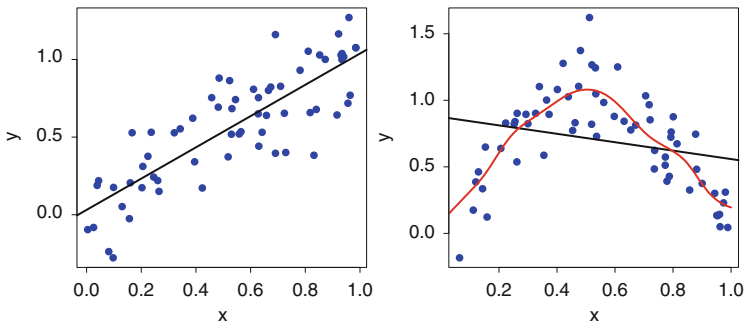


Fig. 2.3 Comparison of linear (*black*) and nonparametric (*red*) regressions as a test of fit. When the data is generated from a linear model (*left*), the two regression methods produce similar results. When the data is generated from a nonlinear model (*right*), the difference in the two models helps detect that a linear model is insufficient

However, for high-dimensional manifolds this type of plot is not available. This is where a comparison to a nonparametric trend is highly useful. In the results below (Sect. 2.6.2) we give an example of how this comparison to nonparametric regression can be used as a diagnostic of model fit in shape analysis applications. The nonparametric regression method that we use for comparison is the one given by Davis et al. [2], which we review now.

2.4.1 Review of Univariate Kernel Regression

Before reviewing the manifold version, we give a quick overview of univariate kernel regression as developed by Nadaraya [19] and Watson [25]. As in the linear regression setting, we are interested in finding a relationship between data $x_1, \dots, x_N \in \mathbb{R}$, coming from an independent variable X , and data $y_1, \dots, y_N \in \mathbb{R}$, representing a dependent variable Y . The model of their relationship is given by

$$Y = f(X) + \epsilon,$$

where f is an arbitrary function, and ϵ is a random variable representing the error. Contrary to linear regression, the function f is not assumed to have any particular parametric form.

Instead, the function f is estimated from the data by local weighted averaging.

$$\hat{f}_h(x) = \frac{\sum_{i=1}^N K_h(x - x_i) y_i}{\sum_{i=1}^N K_h(x - x_i)}.$$

In this equation, K is a function that satisfies $\int K(t) dt = 1$ and $K_h(t) = \frac{1}{h} K(\frac{t}{h})$, with bandwidth parameter $h > 0$. This is the estimation procedure shown in Fig. 2.3 (red curves).

2.4.2 Nonparametric Kernel Regression on Manifolds

The regression method of Davis et al. [2] generalizes the Nadaraya-Watson kernel regression method to the case where the dependent variable lives on a Riemannian manifold, i.e., $y_i \in M$. Here the model is given by

$$Y = \text{Exp}(f(X), \epsilon),$$

where $f : \mathbb{R} \rightarrow M$ defines a curve on M , and $\epsilon \in T_{f(X)}M$ is an error term. As in the univariate case, there are no assumptions on the parametric form of the curve f .

Motivated by the definition of the Nadaraya-Watson estimator as a weighted averaging, the *manifold kernel regression estimator* is defined using a weighted Fréchet sample mean as

$$\hat{f}_h(x) = \arg \min_{y \in M} \frac{\sum_{i=1}^N K_h(x - x_i) d(y, y_i)^2}{\sum_{i=1}^N K_h(x - x_i)}.$$

Notice that when the manifold under study is a Euclidean vector space, equipped with the standard Euclidean norm, the above minimization results in the Nadaraya-Watson estimator.

2.4.3 Bandwidth Selection

It is well known within the kernel regression literature that kernel width plays a crucial role in determining regression results [24]. In particular, it is important to select a bandwidth that captures relevant population-wide changes without either oversmoothing and missing relevant changes or undersmoothing and biasing the results based on individual noisy data points. The ‘Goldie Locks’ method of tuning the bandwidth until the results are most pleasing is a common subjective method for bandwidth selection. However, non-subjective methods may be required, for example, when kernel regression is part of a larger statistical study. A number of automatic kernel bandwidth selection techniques have been proposed for this purpose [24].

One classic method for automatic bandwidth selection is based on least squares cross-validation. This method is easily extended to the manifold regression setting in the following way. The least squares cross-validation estimate for the optimal bandwidth h is defined as

$$\hat{h}_{\text{LSCV}} = \arg \min_{h \in \mathbb{R}^+} \frac{1}{N} \sum_{i=1}^N d \left(\hat{f}_h^{(i)}(x_i), y_i \right)^2,$$

where

$$\hat{f}_h^{(i)}(t) = \arg \min_{y \in M} \left(\frac{\sum_{j=1, j \neq i}^N K_h(x - x_j) d(y, y_j)^2}{\sum_{j=1, j \neq i}^N K_h(x - x_j)} \right)$$

is the manifold kernel regression estimator with the i -th observation left out. This cross-validation method was used to select the bandwidth for the kernel regression example in Fig. 2.3.

2.5 Results: Regression of 3D Rotations

2.5.1 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 $\text{Exp}_q((0, v) \cdot q) = (\cos(\theta/2), 2v \cdot \sin(\theta/2)/\theta) \cdot q$, where $\theta = 2\|v\|$. The log map is given by $\text{Log}_q((a, v) \cdot q) = (0, \theta v/\|v\|) \cdot q$, 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 (2.6) 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 $d_p \text{Exp}$ and $d_p^* \text{Exp}$ are both self-adjoint operators. That is, the above Jacobi field equation provides us both the differential and its adjoint.

2.5.2 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 (2.3). 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 $[0, 1]$.

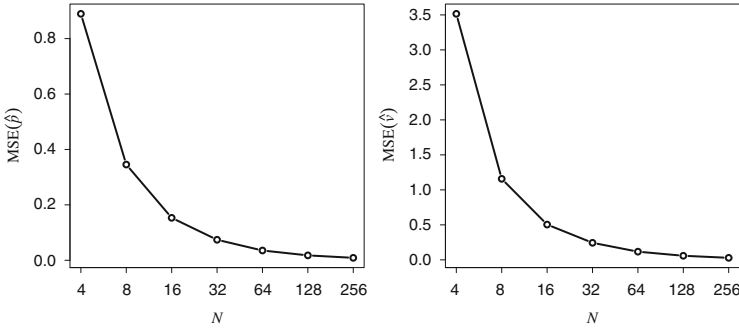


Fig. 2.4 Results for simulated rotation data: MSE of the geodesic regression estimates for the intercept (*left*) and slope (*right*) as a function of sample size

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 $N = 2^k$, $k = 2, 3, \dots, 8$. 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

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

where $M = 1,000$ is the number of repeated trials, and (\hat{p}_i, \hat{v}_i) is the estimate from the i th 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 2.4 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.

2.6 Results: 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 [13] and Bookstein [1], 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 [8, 15, 17, 26]. 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 [27]. In fact, Niethammer et al. [20] recently independently developed geodesic regression for diffeomorphic transformations of image time series. They solve the gradient descent problem with an elegant control theory approach, constraining the regression curve to be a geodesic using Lagrange multipliers. The resulting update to the geodesic's initial conditions involves a numerical integration of an adjoint equation backwards along the geodesic with jump conditions at the data points.

2.6.1 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 $L_x = \{z \cdot x : z \in \mathbb{C} \setminus \{0\}\}$, 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,

$$\text{Exp}_x(v) = \cos \theta \cdot x + \frac{\|x\| \sin \theta}{\theta} \cdot v, \quad \theta = \|v\|. \quad (2.8)$$

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

$$\text{Log}_x(y) = \frac{\theta \cdot (y - \pi_x(y))}{\|y - \pi_x(y)\|}, \quad \theta = \arccos \frac{|\langle x, y \rangle|}{\|x\| \|y\|}. \quad (2.9)$$

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 (2.8) will have the same orientation and scale as the base point x . Also, tangent vectors computed via the log map (2.9) 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 $w = (w_1, \dots, w_{2k-2})$, define the operator

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

(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 , and thus has constant sectional curvature, $K = 1$. 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, \dots, 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 $\mathbb{C}P^{k-2}$ 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.

2.6.2 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. 2.5. Several studies have shown that the volume of the corpus callosum decreases with normal aging [4]. 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 brain database (www.oasis-brains.org). The data consisted of MRI from 32 subjects

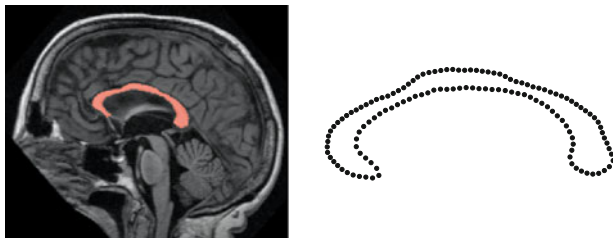


Fig. 2.5 Corpus callosum segmentation and boundary point model for one subject

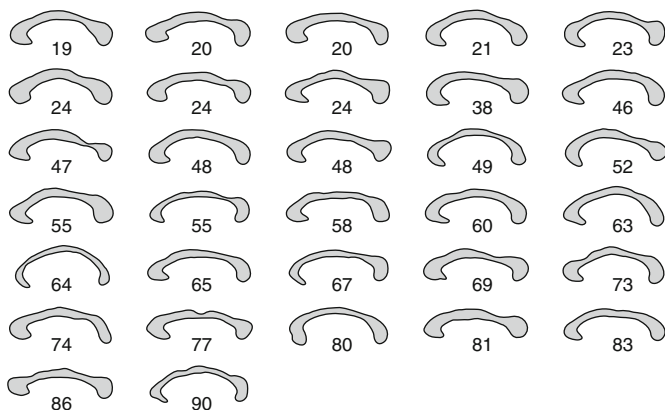


Fig. 2.6 The input corpus callosum shape data and corresponding subject ages in years

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). These 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. 2.5. The entire collection of input shapes and their ages is shown in Fig. 2.6 (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 i th subject's age, and y_i was the i th 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 0, which would be far outside the data range. Least squares estimates (\hat{p}, \hat{v}) were generated according to (2.5), and using the above calculations for $\mathbb{C}P^{k-2}$. The resulting estimated geodesic is shown in Fig. 2.7 as a sequence of shapes: $\hat{y}(t_k) = \text{Exp}(\hat{p}, (t_k - \bar{x})\hat{v})$, for

Fig. 2.7 Geodesic regression of the corpus callosum. The estimated geodesic is shown as a sequence of shapes from age 19 (blue) to age 90 (red)

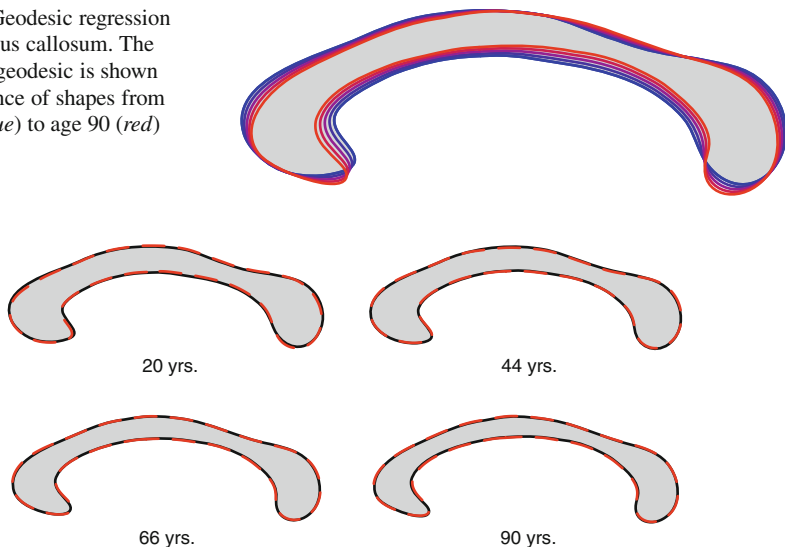


Fig. 2.8 Comparison of geodesic regression (solid black) and nonparametric kernel regression (dashed red) of the corpus callosum shape versus age

$t_k = 19, 36, 54, 72, 90$. 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).

The statistical significance of the estimated trend was tested using the permutation test described in Sect. 2.3.2, 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.

Finally, the appropriateness of the resulting geodesic model fit was tested using a comparison to nonparametric regression, as outlined in Sect. 2.4. First, a nonparametric kernel regression of the corpus callosum data versus age was computed using the method developed by Davis et al. [2] and reviewed in Sect. 2.4.2. The kernel regression was performed on the same Kendall shape space manifold and the bandwidth was chosen automatically using the cross-validation procedure described in Sect. 2.4.3. Next, the resulting corpus callosum shape trend generated by the kernel regression method was compared to the result of the geodesic regression. This was done by again generating shapes from the geodesic model $\hat{\gamma}(t_k)$ at a sequence of ages, t_k , and overlaying the corresponding generated shapes from the kernel regression model at the same ages. The results are plotted for ages

$t_k = 20, 44, 66,$ and 90 (Fig. 2.8). Both regression methods give strikingly similar results. The two regression models at other values of ages, not shown, are also close to identical. This indicates that a geodesic curve does capture the relationship between age and corpus callosum shape, and that the additional flexibility offered by the nonparametric regression does not change the estimated trend. However, even though both methods provide a similar estimate of the trend, the geodesic regression has the advantage that it is simpler to compute and easier to interpret, from the standpoint of the R^2 statistic and hypothesis test demonstrated above.

2.7 Conclusion

We introduced a geodesic regression analysis method for Riemannian manifolds. The geodesic regression model is the natural generalization of linear regression and is parameterized by an intercept and slope term. We also developed a generalization of the R^2 statistic and a permutation test for the significance of the estimated geodesic trend. There are several avenues for future work. First, the hypothesis test presented here could be extended to test for group differences, for example, to test if age-related anatomical changes are different in a disease population compared to controls. Second, theoretical properties of geodesic regression, such as unbiasedness and consistency, would be of interest. Finally, regression diagnostics and model selection procedures need to be developed to assess the appropriateness of a geodesic model for a particular data set.

References

1. Bookstein, F.L.: Size and shape spaces for landmark data in two dimensions (with discussion). *Stat. Sci.* **1**(2), 181–242 (1986)
2. Davis, B., Fletcher, P.T., Bullitt, E., Joshi, S.: Population shape regression from random design data. In: *Proceedings of IEEE International Conference on Computer Vision* (2007)
3. do Carmo, M.: *Riemannian Geometry*. Birkhäuser, Boston (1992)
4. Driesen, N., Raz, N.: The influence of sex, age, and handedness on corpus callosum morphology: a meta-analysis. *Psychobiology* **23**(3), 240–247 (1995)
5. Dryden, I., Mardia, K.: *Statistical Shape Analysis*. Wiley, Chichester (1998)
6. Durrleman, S., Pennec, X., Trounevé, A., Gerig, G., Ayache, N.: Spatiotemporal atlas estimation for developmental delay detection in longitudinal datasets. In: *Medical Image Computing and Computer-Assisted Intervention*, pp. 297–304 (2009)
7. Fletcher, P.T.: Geodesic regression on Riemannian manifolds. In: *MICCAI Workshop on Mathematical Foundations of Computational Anatomy*, pp. 75–86 (2011)
8. Fletcher, P.T., Lu, C., Joshi, S.: Statistics of shape via principal geodesic analysis on Lie groups. In: *IEEE CVPR*, pp. 95–101 (2003)
9. Fox, J.: *Applied Regression Analysis, Linear Models, and Related Methods*. Sage, Thousand Oaks (1997)
10. Fréchet, M.: Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. H. Poincaré* **10**(3), 215–310 (1948)

11. Jupp, P.E., Kent, J.T.: Fitting smooth paths to spherical data. *Appl. Stat.* **36**(1), 34–46 (1987)
12. Karcher, H.: Riemannian center of mass and mollifier smoothing. *Commun. Pure Appl. Math.* **30**, 509–541 (1977)
13. Kendall, D.G.: Shape manifolds, Procrustean metrics, and complex projective spaces. *Bull. Lond. Math. Soc.* **16**, 18–121 (1984)
14. Kendall, W.S.: Probability, convexity, and harmonic maps with small image I: uniqueness and fine existence. *Proc. Lond. Math. Soc.* **3**(61), 371–406 (1990)
15. Klassen, E., Srivastava, A., Mio, W., Joshi, S.: Analysis of planar shapes using geodesic paths on shape spaces. *IEEE PAMI* **26**(3), 372–383 (2004)
16. Mardia, K.V.: *Directional Statistics*. Wiley, Chichester (1999)
17. Michor, P.W., Mumford, D.: Riemannian geometries on spaces of plane curves. *J. Eur. Math. Soc.* **8**, 1–48 (2006)
18. Miller, M.: Computational anatomy: shape, growth, and atrophy comparison via diffeomorphisms. *NeuroImage* **23**, S19–S33 (2004)
19. Nadaraya, E.A.: On estimating regression. *Theory Probab. Appl.* **10**, 186–190 (1964)
20. Niethammer, M., Huang, Y., Vialard, F.-X.: Geodesic regression for image time-series. In: *Proceedings of Medical Image Computing and Computer Assisted Intervention* (2011)
21. Pennec, X.: Intrinsic statistics on Riemannian manifolds: basic tools for geometric measurements. *J. Math. Imaging Vis.* **25**(1), 127–154 (2006)
22. Shi, X., Styner, M., Lieberman, J., Ibrahim, J., Lin, W., Zhu, H.: Intrinsic regression models for manifold-valued data. *J. Am. Stat. Assoc.* **5762**, 192–199 (2009)
23. Trounev, A., Vialard, F.-X.: A second-order model for time-dependent data interpolation: splines on shape spaces. In: *MICCAI STIA Workshop* (2010)
24. Wand, M.P., Jones, M.C.: *Kernel Smoothing*. Number 60 in *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, London/New York (1995)
25. Watson, G.S.: Smooth regression analysis. *Sankhya* **26**, 101–116 (1964)
26. Younes, L.: Computable elastic distances between shapes. *SIAM J. Appl. Math.* **58**, 565–586 (1998)
27. Younes, L.: Jacobi fields in groups of diffeomorphisms and applications. *Q. Appl. Math.* **65**, 113–113 (2006)