2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) | 978-1-6654-4509-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/CVPR46437.2021.01133

Geo-FARM: Geodesic Factor Regression Model for Misaligned Pre-shape Responses in Statistical Shape Analysis

Chao Huang, Anuj Srivastava, Rongjie Liu
Department of Statistics, Florida State University
Tallahassee, FL 32306, USA
chaohuang@stat.fsu.edu,anuj@stat.fsu.edu,rliu3@fsu.edu

Abstract

The problem of using covariates to predict shapes of objects in a regression setting is important in many fields. A formal statistical approach, termed Geodesic regression model, is commonly used for modeling and analyzing relationships between Euclidean predictors and shape responses. Despite its popularity, this model faces several key challenges, including (i) misalignment of shapes due to pre-processing steps, (ii) difficulties in shape alignment due to imaging heterogeneity, and (iii) lack of spatial correlation in shape structures. This paper proposes a comprehensive geodesic factor regression model that addresses all these challenges. Instead of using shapes as extracted from pre-registered data, it takes a more fundamental approach, incorporating alignment step within the proposed regression model and learns them using both pre-shape and covariate data. Additionally, it specifies spatial correlation structures using low-dimensional representations, including latent factors on the tangent space and isotropic error terms. The proposed framework results in substantial improvements in regression performance, as demonstrated through simulation studies and a real data analysis on Corpus Callosum contour data obtained from the ADNI study.

1. Introduction

The field of statistical analysis and modeling of shapes has seen tremendous research and progress. This research is driven by strong applications in computer vision, bioinformatics, computational anatomy, forensics, computer graphics, and so on. Numerous important scientific endeavors have sought to analyze the shapes of objects and investigate their correlations with objects' functionality in large-scale datasets [10, 36, 40, 29]. Shape is broadly defined to be a characteristic that is left after certain *nuisance or shape-preserving* transformations, such as rotations, translations and scale, have been removed [6, 34, 21], with the result that shape representation spaces are nonlinear, high-dimensional, and have quotient space geometry. The last property stems from the need to be invariant to certain shape-preserving transformations as rotations, translations and scale. Consequently, shapes are represented by *orbits* under transformation groups, rather than as points in a pre-shape space. Together, these properties make shape spaces as non-traditional domains for statistical formulations, including definitions of shape statistics (*e.g.* mean and variability) [15], clustering analysis [35], classification [8], testing differences in populations [2] and some others.

In recent years, *shape regression analysis* – the use of shape variables in statistical regression models – has attracted considerable attention. Consequently several approaches have been developed to model the relationships between shape responses and some Euclidean covariates of interest [25, 24, 32, 14, 17, 22, 7, 42, 33, 43]. The past approaches can be classified in two broad categories: *extrinsic regression* and *intrinsic regression*. In the extrinsic regression framework, the shape responses are usually embedded onto a higher dimensional Euclidean space, where classical regression models in that space are applied, and then the estimated models and predictions are projected back onto the original shape space [25, 24]. However, these approaches face some drawbacks including (i) lack in preservation the local shape geometry and (ii) non-guaranteed existence of an inverse and continuous embedding map to the shape space [37].

In contrast, the intrinsic approaches are natural generalizations of regression models from Euclidean spaces to non-Euclidean shape geometries, typically using exponential maps and tangent space representations [32, 14, 17, 22, 7, 42, 33, 43]. To understand this approach better, let $\{f_i^y, x_i\}_{i=1}^n$ be the observed data, where f_i^y is an element of Kendall's shape space $\mathcal{S}$, and $x_i \in \mathbb{R}^p$ is a Euclidean variable. Ideally f_i^y, representing a shape, should be an orbit $[f_i^y]$ of a pre-shape space under the rotation group. However, the use of quotient space geometry in specifiy-

978-1-6654-4509-2/21/$31.00 ©2021 IEEE
DOI 10.1109/CVPR46437.2021.01133

ing regression models is difficult and has not been pursued. Instead, the common approach is to take a representative element of the pre-shape space, aligned or rotated appropriately through some pre-processing steps. The presumption is that through this pre-processing the nuisance transformations have been filtered out. Then, one can apply a commonly used *geodesic regression model*:

$$f_i^y = \text{Exp}(\kappa(x_i), \epsilon_i), \quad \kappa(x_i) \in \mathcal{S}, \epsilon_i \in T_{\kappa(x_i)}\mathcal{S}, \qquad (1)$$

where $\text{Exp}(\kappa(x_i), \cdot) : T_{\kappa(x_i)}\mathcal{S} \to \mathcal{S}$ is the exponential map at $\kappa(x_i)$, and $T_{\kappa(x_i)}\mathcal{S}$ is the corresponding tangent space (some useful concepts from differential geometry can be found in Appendix A). Model (1) involves two key terms: the conditional mean shape $\kappa(x_i)$ and the error $\epsilon_i \in T_{\kappa(x_i)}\mathcal{S}$. The conditional mean shape $\kappa(x_i)$ can be treated as a link function including the typical parametric setting, i.e., $\text{Exp}(\mu, \mathbf{B}x_i)$, [14, 22, 43] and some other nonparametric settings [32, 7]. Similarly, the error term ϵ_i can be specified using parametric [14], semi-parametric [7], or completely nonparametric models [22].

The main issue here lies in using (pre-aligned) elements of pre-shape space, rather than actual orbits as shape representations. It is a well-known mathematical fact that optimal rotations alignments of objects can not be achieved via pre-processing. For instance, the optimal alignments of objects A and B to object C, respectively, do not result in optimal alignments between objects A and B themselves! Rotations have to be solved for during pairwise shape comparisons. That is why shape spaces are typically quotient spaces, and not subsets, of pre-shape spaces. This fundamental issue leads to three key limitations of geodesic regression models: **(i) Misalignment issue in pre-aligned responses.** In practice one observes raw images rather than getting the shape data directly. The shapes are extracted from the images using a pre-processing step – these steps are increasingly being performed using deep learning networks. Even when the image data are pre-registered and assumed to be well aligned, the shapes extracted from this image data exhibit mis-alignment and can even be noisy [1, 42, 33], which negatively affects regression performance. To illustrate this issue, we fitted the popular geodesic regression model to some simulated data later in the paper (detailed simulation settings can be found in Section 3.2), and reached an estimated "baseline shape" presented in Figure 1, where the estimate is found to be biased when the shape data contains misalignment variability. **(ii) Non-optimal alignment due to imaging heterogeneity.** Since most pre-alignment approaches are implemented on imaging data, the presence of imaging heterogeneity [19] causes the nuisance transformations to be correlated with some covariates of interest, *e.g.*, gender and age, which makes the pre-alignment non-optimal and adversely affects the regression performance. **(iii) Lack of spatial correlation structure in modeling.** Most existing methods assume that stochastic terms have isotropic variability [14, 22]. Although the spatial correlation is considered in [7] via introduction of a random weighted matrix on the tangent space, its implementation suffered due to heavy computational burden incurred in choosing the optimal weighted matrix.

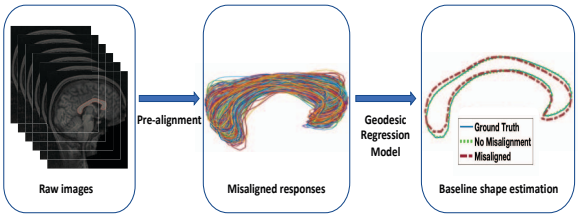

Figure 1. Example of misalignment issue in pre-aligned responses.

This aim of this paper is to propose a **Geo**desic **FA**ctor **R**egression **M**odel (**Geo-FARM**) that addresses all three challenges. Specifically, the main contributions of this paper are: **(i).** Instead of treating objects extracted from pre-aligned functional data as shapes, we treat them as pre-shapes, i.e., coordinate data after filtering out only the location and scale effects. We incorporate full shapes inside regression models as proper orbits. In practical terms, the rotational alignments are applied on pre-shapes and learned inside the regression model itself. **(ii).** The spatial correlation structure in our Geo-FARM is established as a low-dimensional representation, including latent factors through a factor analysis framework on the tangent space and error term modeled using the isotropic Riemannian Normal (RN) distribution [28, 14]. **(iii).** A Monte Carlo Expectation-Maximization (MCEM) algorithm is used to develop the estimation procedure for both parameters and nuisance transformations. In addition, hypothesis testing problems are discussed to investigate the significance of some covariates of interest on the shape responses. **(iv).** The efficacy of our Geo-FARM is assessed using Monte Carlo simulations and a real data example on corpus callosum contour data obtained from the ADNI study. A MATLAB-based companion software will be released to the public through **GitHub**.

2. Method

2.1. Pre-shape space for planar curves

Let $\mathbf{L} \in \mathcal{L}_{2,k}$ be a $2 \times k$ matrix whose k columns denote k landmarks from a 2-dimensional object. After removing the translation and scaling of elements in $\mathcal{L}_{2,k}$, one reaches the pre-shape space defined as $\mathcal{S}_2^k = \{\mathbf{L} \in \mathcal{L}_{2,k} : \sum_{j=1}^k L_{i,j} = 0,\ i = 1, 2,\ \|\mathbf{L}\|_F = 1\}$, where $\|\cdot\|_F$ is the Frobenius norm. $\mathcal{S}_2^k$ is not the shape space since the pre-shape is not invariant to the action of the rotation group, $SO(2)$. Noting that there exists an one-to-one map $f(\cdot)$ such that $\mathcal{S}_2^k$ is equivalent to the unit hypersphere $\mathbb{S}^{m-1}$ with $m = 2k-2$, one can utilize

all the geometric properties on hyperspheres to analyze pre-shapes [36]. Specifically, Let $y \in \mathcal{L}_{2,k}$ and $f(y)$ be a point on a $(m\text{-}1)$-dimensional sphere $\mathbb{S}^{m-1}$, and v be a tangent vector at $f(y)$. The exponential map is given by

$$\mathrm{Exp}(f(y), v) = \cos(\|v\|) f(y) + \frac{\sin(\|v\|)}{\|v\|} v. \tag{2}$$

For another point $f(y') \in \mathbb{S}^{m-1}$, the inverse exponential map, or log map, between $f(y)$ and $f(y')$ is given by

$$\mathrm{Log}(f(y), f(y')) = \frac{\mathbb{P}(y, y')}{\|\mathbb{P}(y, y')\|} \arccos\langle f(y), f(y')\rangle, \tag{3}$$

where $\mathbb{P}(y, y') = f(y') - f(y)\langle f(y), f(y')\rangle$. In addition, the parallel transport of the tangent vector v from $T_{f(y)}\mathbb{S}^{m-1}$ to $T_{f(y')}\mathbb{S}^{m-1}$ can be derived as

$$\Gamma_{f(y)}^{f(y')}[v] = v - \frac{2v^T f(y')}{\|f(y) + f(y')\|^2}(f(y) + f(y')). \tag{4}$$

2.2. Geo-FARM

Assume that $\{y_i\}_{i=1}^n$ are observed from the pre-shape space $\mathcal{S}_2^k$ and $\{x_i\}_{i=1}^n$ are from a Euclidean space $\mathbb{R}^p$. In order to simultaneously handle the nuisance transformations and establish the relationship between pre-shape responses and Euclidean covariates, a novel geodesic regression model is proposed as:

$$f(y_i * g_i) = \mathrm{Exp}\left(\mathrm{Exp}(f(\mu), \mathbf{B}x_i), \epsilon_i\right), \tag{5}$$

where $g_i \in SO(2)$ denotes the rotation group action that forms the individual nuisance transformation. $\mu \in \mathcal{S}_2^k$ and $f(\mu) \in \mathbb{S}^{m-1}$ is the base point. All the columns in $\mathbf{B}$, i.e., $\{\beta_j\}_{j=1}^p$, are tangent vectors at $T_{f(\mu)}\mathbb{S}^{m-1}$, representing the effects from the predictors $\{x_i\}_{i=1}^n$.

In order to establish the spatial correlation structure, our Geo-FARM integrates model (5) with a factor analysis framework generalized from Euclidean space to the pre-shape space:

$$f(y_i * g_i)|x_i, z_i \sim \mathrm{RN}\left(\kappa(x_i, z_i), \sigma\right), \; z_i \sim \mathrm{N}(0, \mathrm{I}_q), \tag{6}$$

where $\kappa(x_i, z_i) = \mathrm{Exp}(f(\mu), \mathbf{B}x_i + \mathbf{\Lambda}z_i)$ and RN is the isotropic RN distribution [14]. The proposed factor analysis framework builds the correlation structure with a low rank representation ($q \ll m$) including (i) a low number of latent factors represented by the columns of $\mathbf{\Lambda}$, i.e., $\{\alpha_j \in T_{f(\mu)}\mathbb{S}^{m-1}\}_{j=1}^q$, and (ii) stochastic error term ϵ_i that follows the isotropic RN distribution. Then, the joint probability density function for $(f(y_i * g_i), z_i)$ is given by

$$h(f(y_i * g_i), z_i) = \phi(z_i)\mathcal{C}(\sigma) \exp\{-\frac{1}{2\sigma} \times \\ \|\mathrm{Log}\left(\mathrm{Exp}(f(\mu), \mathbf{B}x_i + \mathbf{\Lambda}z_i), f(y_i * g_i)\right)\|^2\}, \tag{7}$$

where $\phi(\cdot)$ is the q-dimensional standard normal distribution density function. In addition, the model identifiability of our Geo-FARM is guaranteed under certain conditions.

Proposition 2.1 *Consider the probability density function of GEO-FARM, i.e., $\varrho(f(y * g), \Theta) \doteq \int h(f(y * g), z)dz$. Given the nuisance transformation g and the number of latent factors q, if the design matrix $\mathbf{X} = (x_1, \ldots, x_n)^T$ is full row rank, the density function $\varrho(f(y * g), \Theta)$ is generically identifiable in the parameter space.*

A graphical illustration of our Geo-FARM is presented in Figure 2. In summary, there are several advantages of our

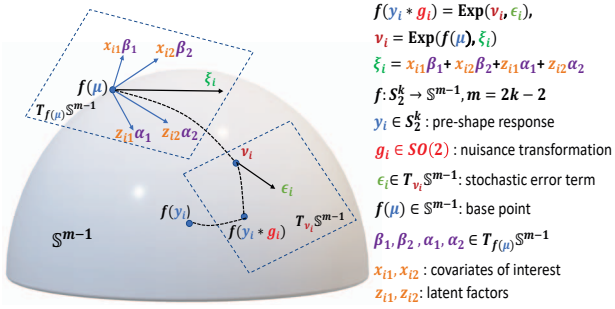

Figure 2. Graphical illustration of Geo-FARM.

Geo-FARM: **(i)** Compared to the existing geodesic regression models for well aligned responses [22, 7], our Geo-FARM can successfully deal with the misalignment issue in pre-aligned responses via introducing the individual nuisance transformation g_i. In addition, compared to the traditional preprocessing approaches, the alignment of each pre-shape in our Geo-FARM can be refined since the nuisance transformation g_i is learned based on all the available information including not only the response y_i but also the other covariates x_i. **(ii)** Through the factor analysis frame in our Geo-FARM, the variability among pre-shapes in model (5) can be expressed by two parts: latent variables from a low dimensional space in the tangent space $T_{f(\mu)}\mathbb{S}^{m-1}$; and stochastic error terms with isotropic variance structure in the tangent space $T_{\mathrm{Exp}(f(\mu), Bx_i + \Lambda z_i)}\mathbb{S}^{m-1}$. Compared to the general RN distribution [28], the number of parameters that specify the correlation structure has been reduced a lot from $m(m+1)/2$ to $mq + 1$, where $q \ll m$. **(iii)** Instead of considering the nuisance transformation $g_i \in SO(m)$ for responses $f(y_i) \in \mathbb{S}^{m-1}$, our Geo-FARM treats g_i as a rotation group action on the pre-shape $y_i \in \mathcal{S}_2^k$, which avoids the computational burden caused by the high dimensional structure of nuisance transformations in [42], Specifically, g_i can be represented by the 2-dimensional orthogonal matrix: $O_i(\psi) = \begin{bmatrix} \cos(\psi) & -\sin(\psi) \\ \sin(\psi) & \cos(\psi) \end{bmatrix}$, where $\psi \in [-\pi, \pi]$. Therefore, the dimension of the nuisance transformation group is reduced from $m(m-1)/2$ to 1.

2.3. Estimation procedure

The maximum likelihood estimate (MLE) of Θ including the parameters of interest, $\Upsilon \doteq \{\mu, \mathbf{B}, \mathbf{\Lambda}\}$, the nuisance parameter, σ, and the nuisance transformations, $G \doteq \{g_i\}_{i=1}^n$, can be derived through the following complete log-likelihood function based optimization problem:

$$\hat{\Theta} = \text{argmax}_{\Upsilon,\sigma}\, n \log \mathcal{C}(\sigma) - \frac{1}{2\sigma} \times \sum_{i=1}^n \inf_{g_i} \|\text{Log}\left(\text{Exp}(f(\mu), \mathbf{B}x_i + \mathbf{\Lambda}z_i), f(y_i * g_i)\right)\|^2. \quad (8)$$

Noting that the presence of optimization over g_i inside the summation ensures that the nuisance transformation is removed using all available information (e.g., y_i and x_i) instead of a pre-processing over y_i. Since the latent variables $\{z_i\}_{i=1}^n$ are unobserved, the objective function in (8) is often intractable in practice. Here the EM algorithm [9] is considered by iteratively applying two steps: E-step and M-step. Specifically, at the t-th iteration, instead of the complete log-likelihood function, the optimization problem in M-step is established based on the Q-function $Q(\Theta|\Theta^{(t)})$ calculated in E-step,

$$\sum_{i=1}^n \mathbb{E}_{z_i}\left[\log h(f(y_i * g_i), z_i)|y_i, x_i, \Theta^{(t)}\right], \quad (9)$$

which is the expectation of complete log-likelihood function with respect to the latent variables given the observed data and the current estimate $\Theta^{(t)}$. However, the conditional expectation in (9) does not yield a closed-form solution, which brings about difficulties in M-step [41]. To address this issue, we consider the MCEM algorithm [23], where the Q-function is approximated via Monte Carlo techniques.

Monte Carlo E-step: at the t-th iteration, given the current estimate $\Theta^{(t)}$, we consider the approximated Q-function:

$$\tilde{Q}(\Theta|\Theta^{(t)}) \propto \frac{1}{n_z}\sum_{j=1}^{n_z}\sum_{i=1}^{n} \log h(f(y_i * g_i), z_i^j), \quad (10)$$

where $\{z_i^j\}_{j=1}^{n_z}$ are generated from the conditional distribution $p(z_i|y_i, x_i, \Theta^{(t)})$ via the Hamiltonian Monte Carlo (HMC) sampling method [26]. According to HMC method, we set up the Hamiltonian dynamic system first. The Hamiltonian function can be written as

$$H(z_i, r) = U(z_i) + \tfrac{1}{2} r_i^T r_i,$$
$$U(z_i) = -\log p(z_i|y_i, x_i, \Theta^{(t)}), \quad (11)$$

where $U(z_i)$ is called the potential energy function. The other item $\frac{1}{2} r_i^T r_i$ is called the kinetic energy, where r_i are auxiliary momentum variables drawn independently from $\text{N}(0, \text{I}_q)$. Because of the introduction of r_k, the Hamiltonian dynamics can be established as

$$\frac{dz_i}{dt} = r_i, \quad \frac{dr_i}{dt} = -\nabla_{z_i} U(z_i). \quad (12)$$

Then the approximation solution to (12) can be obtained via the Leap Frog numerical integration method [26] if the item $\nabla_{z_i} U(z_i)$ is calculated. In fact, the gradient term $\nabla_{z_i} U(z_i)$ can be derived as below

$$z_i - \sigma^{(t)^{-1}} \mathbf{\Lambda}^{(t)^T} d_v \text{Exp}(u, v)^\dagger \text{Log}\left(\text{Exp}(u, v), f_i^{(t)}\right), \quad (13)$$

where $u = f(\mu^{(t)})$, $v = \mathbf{B}^{(t)} x_i + \mathbf{\Lambda}^{(t)} z_i$, $f_i^{(t)} = f(y_i * g_i^{(t)})$, and $d_v \text{Exp}(u, v)$ is the derivative of $\text{Exp}(u, v)$ with respect to v, and $\dagger$ represents the adjoint of a linear operator. For spheres, the adjoint derivative has an analytical expression, i.e., $d_v \text{Exp}(u, v)\dagger w = \sin(\|v\|)\|v\|^{-1} w^\perp + w^\top$ where $w^\perp$ and $w^\top$ denote the components of w that are orthogonal and tangent to v, respectively.

The performance of standard HMC method is highly sensitive to two user-specified parameters: a step size ε and a desired number of steps l. In particular, if l is too small then the algorithm exhibits undesirable random walk behavior, while if l is too large the algorithm wastes computation. Compared with the standard HMC, the No-U-Turn Sampler (NUTS) [18], an extension to standard HMC, can avoid setting the tuning parameter l. Specifically, NUTS uses a recursive algorithm to build a set of likely candidate points that spans a wide swath of the target distribution, stopping automatically when it starts to double back and retrace its steps. Because of this, NUTS is adopted in this paper. The details of NUTS algorithm is omitted here, and readers can refer to Section 3 and Algorithm 3 in [18].

M-step: the updated estimate $\Theta^{(t+1)}$ can be obtained via maximizing the approximated Q-function $\tilde{Q}(\Theta|\Theta^{(t)})$. In order to solve the optimization problem above, here an iterative approach is adopted, where one updates the estimates of Υ, σ, or G while keeping the other fixed. Thus, we first focus on techniques for estimating these quantities separately.

Updating $\Upsilon^{(t+1)}$ while keeping $G^{(t)}$ and $\sigma^{(t)}$ fixed. Given $G^{(t)}$ and $\sigma^{(t)}$, $\Theta^{(t+1)}$ is updated via the proximal alternating linearized minimization (PALM) algorithm [4, 42], where Υ is iteratively updated through the gradient functions of $\tilde{Q}(\Theta|\Theta^{(t)})$. Specifically, let c be a positive constant while K_μ, K_{β_j}, and K_{α_j} be the Lipschitz constants of $\nabla_{f(\mu)}\tilde{Q}$, $\nabla_{\beta_j}\tilde{Q}$, and $\nabla_{\alpha_j}\tilde{Q}$, respectively. Given the update $\Theta^{(t,k)} = \{\Upsilon^{(t,k)}, G^{(t)}, \sigma^{(t)}\}$ at the k-th iteration, the iterations of PALM algorithm is provided in Algorithm 1. Finally, $\Theta^{(t)}$ in E-step is updated by $\Theta^{(t+1)} = \Theta^{(t,k+1)}$ when certain iteration stopping criterion is satisfied, e.g., $\|\Upsilon^{(t,k+1)} - \Upsilon^{(t,k)}\| < 10^{-4}$.

Updating $G^{(t+1)}$ while keeping $\Upsilon^{(t+1)}$ and $\sigma^{(t)}$ fixed. Since each $g_i \in SO(2)$ can be written as the orthogonal matrix $O(\psi_i)$ related to one parameter $\psi_i \in [-\pi, \pi]$, $g_i^{(t+1)}$ is updated via minimizing the following objective function:

$$E(\psi_i|\Upsilon^{(t+1)}) = \sum_{l=1}^{n_z} \left\|\text{Log}\left(f_{i,l}^{(t+1)}, f(O_i(\psi)y_i)\right)\right\|^2, \quad (14)$$

Algorithm 1: PALM algorithm in M-step

while stopping criterion not satisfied **do**
- Update $f(\mu^{(t,k+1)})$ by $\text{Exp}\left(f(\mu^{(t,k)}), -(cK_\mu)^{-1}\nabla_{f(\mu)}\tilde{Q}\right)$;
- Update $\beta_j^{(t,k+1)}, j = 1, \ldots, p$, by $\Gamma_{f(\mu^{(t,k)})}^{f(\mu^{(t,k+1)})}\left[\beta_j^{(t,k)} - (cK_{\beta_j})^{-1}\nabla_{\beta_j}\tilde{Q}\right]$;
- Update $\alpha_j^{(t,k+1)}, j = 1, \ldots, q,$, by $\Gamma_{f(\mu^{(t,k)})}^{f(\mu^{(t,k+1)})}\left[\alpha_j^{(t,k)} - (cK_{\alpha_j})^{-1}\nabla_{\alpha_j}\tilde{Q}\right]$;
- Set $k = k + 1$.

where $f_{i,l}^{(t+1)} = \text{Exp}(f(\mu^{(t+1)}), \mathbf{B}^{(t+1)}x_i + \mathbf{\Lambda}^{(t+1)}z_i^l)$, and $\psi_i \in [-\pi, \pi]$. This univariate minimization problem can be solved based on the *Golden Section Search* algorithm [30] and implemented by the MATLAB function *fminbnd*.

Updating $\sigma^{(t+1)}$ while keeping $\Upsilon^{(t+1)}$ and $G^{(t+1)}$ fixed. We first define that $\eta = -1/2\sigma$ and $\varphi(\eta) = -\log \mathcal{C}(\sigma)$. Then given $\Upsilon^{(t+1)}$ and $G^{(t+1)}$, $\sigma^{(t+1)}$ is updated via solving the following problem related to η:

$$\min_{\eta} \varphi(\eta) - \frac{\eta}{n_z}\sum_{i=1}^{n} E(\psi_i^{(t+1)}|\Upsilon^{(t+1)}). \tag{15}$$

where the normalization term $\varphi(\eta)$ defined for $\mathbb{S}^{m-1}$ can be derived with a closed form [41]. Noting that $\varphi(\eta)$ is a strictly convex function with respect to η, therefore the solution to problem (15) exists and is unique [31], which can be solved based on the *Newton-Raphson* algorithm [5] and implemented by the MATLAB function *fminunc*.

Now we summarize the overall estimation procedure for our Geo-FARM in the following Algorithm 2.

Algorithm 2: MCEM algorithm for Geo-FARM

Data: pre-shapes $\{y_i\}_{i=1}^n$ and covariates $\{x_i\}_{i=1}^n$
Result: estimation of Θ
Initialization: $\Theta^{(0)}$ and $t = 0$
while stopping criterion not satisfied **do**
- **Monte Carlo E-step**
- Sample $\{z_i^j\}_{i,j}$ via HMC method;
- Calculate the approximated Q-function in (10);
- **M-step**
- Update $\Upsilon^{(t+1)}$ in Algorithm 1;
- Update $G^{(t+1)}$ by minimizing (14);
- Update $\sigma^{(t+1)}$ by minimizing (15);
- Set $t = t + 1$;

Output: $\hat{\Theta} = \Theta^{(t)}$.

Here we end the estimation procedure with discussions on some other issues in MCEM algorithm including the initialization in MCEM algorithm and criterion for choosing the number of latent factors.

Initialization in MCEM algorithm. Since the MCEM algorithm is an iterative procedure, its performance strongly depends on starting points. For our Geo-FARM, good initialization is crucial for finding the estimates due to the presence of multiple local maxima of the likelihood function. Here we consider using the $random$ MCEM algorithm, where multiple starting points are chosen and the point with the highest log-likelihood function is chosen as the starting point. In simulation studies and real data analysis, the estimation procedure in [22] is considered for initializing base point parameter $f(\mu)$ and coefficients $\mathbf{B}$, while the factor analysis method is performed in the tangent space $T_{f(\mu)}\mathbb{S}^{m-1}$ to initialize the factor loading matrices $\mathbf{\Lambda}$ and the nuisance parameter σ.

Determining the number of latent factors. Since the number of latent factors, q, is unknown, the 2-fold cross predictive log-likelihood method is considered through an exhaustive search as our model selection criterion [20]. However, according to the simulation results in Section 3.2, our Geo-FARM is not sensitive to the choice of number of latent variable. In particular, even when the model is misspecified (*e.g.* the values of q is larger than 0 but incorrectly selected), our Geo-FARM still performs well in terms of its estimation accuracy. Therefore, in our simulation studies and real data analysis, we prefixed the number of latent factors as 1, which will release the potential computational burden due to the large value assigned for q.

2.4. Hypothesis testing

In medical image data analysis, people are interested in investigating the relationship between the pre-shape responses and some covariates of interest. This type of scientific questions are formulated into the following hypothesis testing problem on each $\beta_j, j = 1, \ldots, p$:

$$\mathbf{H_0} : \beta_j = 0 \text{ vs. } \mathbf{H_1} : \beta_j \neq 0. \tag{16}$$

For this testing problem, we consider using the Wald test statistic [7]. Specifically, for testing problem (16), the test statistic is constructed as

$$T_j = \boldsymbol{e}_j^T\hat{\mathbf{B}}^T[\boldsymbol{e}_j \otimes \mathrm{I}_m]\hat{\Omega}_B[\boldsymbol{e}_j \otimes \mathrm{I}_m]^T\hat{\mathbf{B}}\boldsymbol{e}_j, , \tag{17}$$

where $\boldsymbol{e}_j$ is a $p \times 1$ vector with the j-th element 1 and rest 0. $\hat{\Omega}_B$ is the corresponding partition of estimated inverse covariance matrix of all the estimated parameters in Geo-FARM. It can be shown that T_j is asymptotically χ^2 distributed with $m - 1$ degree of free under $\mathbf{H_0}$. In practice, when the sample size is not on a large scale, the parametric bootstrap procedure can be used to derive the empirical distribution of T_j and the p-value [11].

On the other side, since we claim that our Geo-FARM can help refine the alignment as the nuisance transformations are learned based on both pre-shapes and covariates, we would like to check if the mean shape extracted from our Geo-FARM is significantly different from that extracted from the pre-aligned data. In order to achieve this goal, The Hoteling T^2 test statistic is adopted and implemented through a bootstrap hypothesis testing approach [2] and its R package *shapes* (http://cran.r-project.org/web/packages/shapes/index.html).

3. Numerical experiments

The corpus callosum (CC) contains homotopic and heterotopic interhemispheric connections and is essential for communication between the two cerebral hemispheres [27, 20]. Connection with most of the cortex has made CC a target of investigation of brain integrity in Alzheimer's disease [3, 12, 7]. In this section, both simulation studies and real data analysis are conducted to assess the performance of Geo-GARM using the CC contour data of ADNI-1 study. In simulation studies, the synthetic data is generated from the real dataset.

3.1. Data preprocessing

We use *FreeSurfer* [13] to process each T1-weighted MRI, including motion correction, non-parametric non-uniform intensity normalization, affine transform to the MNI305 atlas, intensity normalization, skullstripping, and automatic subcortical segmentation. Some quality control procedures are done on each output image data. Then, through the package *CCSeg* [38], each T1-weighted MRI image and tissue segmentation results are used to extract the planar CC contour data on the midsagittal slice , which contains 50 landmarks (See Figure 3).

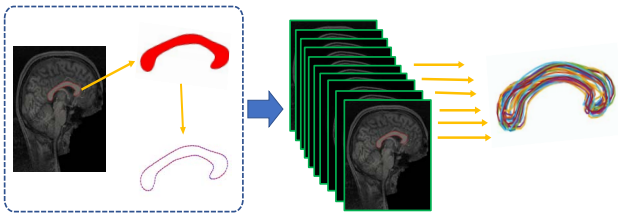

Figure 3. CC contour data preprocessing.

Table 1. Demographic information of preprocessed ADNI-1 CC contour data, including gender and age (in years).

	NC (223)	AD (186)	All (409)
Gender (F/M)	107/116	88/98	195/214
Avg. Age (Std.)	76.2 (4.9)	75.4 (7.4)	75.9 (6.2)

After the quality control, we obtain CC contour data from 409 subjects including 223 normal control (NC) and 186 AD. The demographic information of the preprocessed CC contour data set is presented in Table 1.

3.2. Simulation studies

We generate the data $\{x_i, y_i, g_i\}_{i=1}^n$ from the model (6) as follows. Without special saying, we set $n = 100$. In order to mimic the ADNI-1 CC contour data, the true values of parameters in (6) are learned from the real data itself. Specifically, we first fit Geo-FARM to the CC contour data of all the normal controls in ADNI-1, where two predictors, i.e., gender (x_{i1}) and normalized age (x_{i2}), are included. Then, we use the obtained parameter estimators of $\mu, \mathbf{B}, \mathbf{\Lambda}$, and σ as their true values in our simulation setting. Meanwhile, the number of latent factors is also learned from the model fitting, i.e., $q = 2$. Next, the covariates x_{i1} and x_{i2} are generated according to their data types, i.e., x_{i1} is generated from Bernoulli distribution with parameter $p = 0.5$, while x_{i2} is generated from uniform distribution $U(0, 1)$. In order to generate the sphere data $f(y_i * g_i)$ from the RN distribution, the sampling algorithm proposed for spherical normal distribution [16] is considered. After that, we generate the nuisance transformations g_i via sampling the rotation angle ψ in orthogonal matrix $O_i(\psi)$. Specifically, given an upper bound $\bar{\psi}$, the rotation angle ψ for each data point is uniformly generated from $[-\bar{\psi}, \bar{\psi}]$. In order to investigate the effect of nuisance transformations on the estimation performance, multiple scenarios are considered here via setting different values of $\bar{\psi}$, i.e., $\bar{\psi} \in \{0, 5, 10, 15, 20, 25, 30\}$. Finally, we generate 50 datasets for each simulation scenario.

Here two other approaches are considered here for comparison: (i) multivariate general linear models (MGLM) on Riemannian manifolds [22] and (ii) multivariate regression with gross errors (MRGE) on manifold-valued data [42]. Specifically, MGLM can be treated as an generalization of multiple linear regression models from Euclidean space to Riemannian space, while MRGE aims to improve MGLM by considering gross errors on manifold responses. For comparison, we introduce several loss functions here: (i) the sum of squared geodesic errors (SSGE), i.e., $\sum_{i=1}^n \|\mathrm{Log}(\mathrm{Exp}(f(\mu), \mathbf{B}x_i), \mathrm{Exp}(f(\hat{\mu}), \hat{\mathbf{B}}x_i))\|^2$, to assess the prediction accuracy of the pre-shapes; (ii) the norm $\|\mathbf{B} - \Gamma_{f(\hat{\mu})}^{f(\mu)}\hat{\mathbf{B}}\|_F$ to assess the estimation accuracy of $\hat{\mathbf{B}}$; (iii) geodesic distance between $f(\mu)$ and $f(\hat{\mu})$, i.e., $\|\mathrm{Log}(f(\mu), f(\hat{\mu}))\|$ to assess the estimation accuracy of $\hat{\mu}$; and (iv) the median absolute deviation (MAD) of $\triangle_\psi \doteq \{|\psi_i - \hat{\psi}_i|\}_{i=1}^n$, i.e., $\mathrm{median}(\triangle_\psi)$, to assess the detection accuracy of nuisance rotations. The simulation results for all different scenarios are presented in Figure 4. It can be found that (i) when the pre-shapes are slightly misaligned, all the three methods perform well; (ii) when the misalignment is getting severe (i.e., $\bar{\psi}$ increases), our Geo-FARM

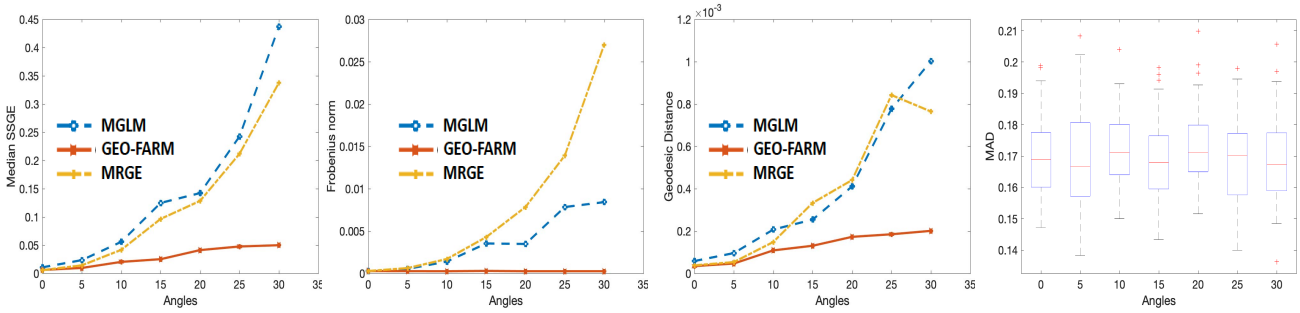

Figure 4. Comparison of three methods (MGLM, Geo-FARM, and MRGE) for different settings of $\bar{\psi}$ ($\bar{\psi} \in \{0, 5, 10, 15, 20, 25, 30\}$): (a) Median SSGE; (b) Frobenius norm $\|B - \Gamma_{f(\hat{\mu})}^{f(\mu)} \hat{B}\|_F$; (c) geodesic distance between $f(\mu)$ and $f(\hat{\mu})$; and (d) MAD of $\triangle_\psi$.

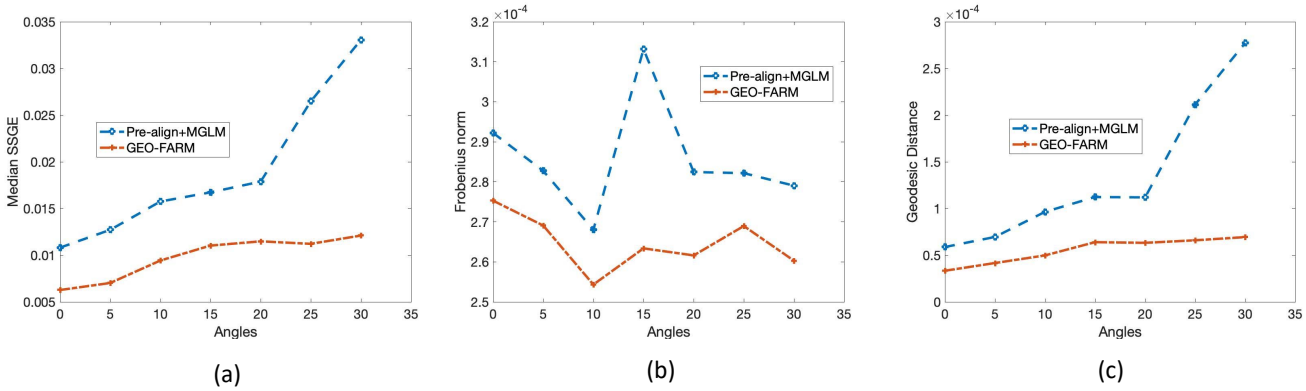

Figure 5. Comparison the performance of Geo-FARM with MGLM on pre-aligned pre-shapes with $\psi_i = \tilde{\psi}_i x_{i1}$ and $\tilde{\psi}_i \sim U(-\bar{\psi}, \bar{\psi})$: (a) Median SSGE; (b) Frobenius norm $\|B - \Gamma_{f(\hat{\mu})}^{f(\mu)} \hat{B}\|_F$; and (c) geodesic distance between $f(\mu)$ and $f(\hat{\mu})$.

is more robust compared to other two approaches in terms of both prediction accuracy and parameter estimation accuracy; (iii) although MRGE considers detecting and correcting the gross errors on pre-shapes through penalization approach, its performance is almost as worse as MGLM because it is very sensitive to the data structure and choice of tuning parameter; (iv) Geo-FARM shows great performance in detecting the nuisance rotations for all different scenarios. Therefore, our Geo-FARM shows its great power in both misalignment detection and parameter estimation.

Next, we compare our Geo-FARM with the typical procedure that conduct the geodesic regression analysis on pre-aligned manifold valued data [39]. Specifically, for comparison with our Geo-FARM, the simulated pre-shapes are first aligned to their Karcher mean, then the MGLM is adopted to conduct the regression analysis on the pre-aligned data. Here the individual rotation angle ψ_i is generated through $\psi_i = \tilde{\psi}_i * x_{i1}$, where $\tilde{\psi}_i$ is uniformly generated from $[-\bar{\psi}, \bar{\psi}]$. This simulation mechanism indicates that the nuisance transformations are correlated to the covariate information. The simulation results for all different scenarios are presented in Figure 5. It can be found that, our Geo-FARM outperforms the pre-alignment based approach in terms of the estimation accuracy, which means that our Geo-FARM benefits from the fact that the estimate of the nuisance transformation can be refined based on all the available information including pre-shapes and covariates.

Finally, we investigate the robustness of our Geo-FARM in choosing the number of latent factors. Three different scenarios are considered here: $\bar{\psi} = 0$, $\bar{\psi} = 10$, and $\bar{\psi} = 20$. We manually specified four different values for q, i.e., 0, 1, 2, and 3, where 2 is the ground truth in our simulation settings. The four loss functions defined above are considered here as well. The simulation results for different choices of q are presented in Figure 6. Couple of findings are listed here: (i) when no latent factors are specified in our Geo-FARM (an isotropic covariance structure used instead), the estimation performance gets worse and worse when the misalignment is getting severe; (ii) when the latent factor structure is considered but the number of factors is misspecified, the estimation performance of our Geo-FARM almost keeps the same as that when q is set to the true value. Therefore, the latent factor structure in our Geo-FARM is of great importance while the choice of the number of factors is not that critical with respect to the estimation performance.

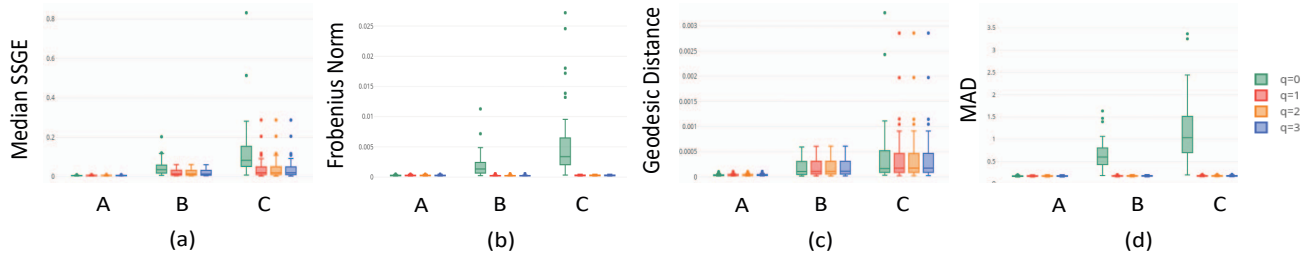

Figure 6. Comparison the performance of Geo-FARM for different choices of q (i.e., 0, 1, 2, and 3, where 2 is the ground truth in our simulation settings) under three scenarios (Scenario A: $\bar{\psi} = 0$, Scenario B: $\bar{\psi} = 10$, and Scenario C: $\bar{\psi} = 20$): (a) Median SSGE; (b) Frobenius norm $\|B - \Gamma_{f(\hat{\mu})}^{f(\mu)} \hat{B}\|_F$; (c) geodesic distance between $f(\mu)$ and $f(\hat{\mu})$; and (d) MAD of $\triangle_\psi$.

3.3. Real data analysis

Here we apply Geo-FARM on ADNI CC CC contour data, where the pre-shape data were extracted by removing the translation and scaling information from the original landmarks. In addition, three covariates of interest are included in the regression model, i.e., gender, age (standardized), diagnostic status (AD), and an interaction term age×AD. Then a sequence of estimated shapes from age 50 to age 95 is presented for each of the four subgroups (1. Male & NC; 2. Female & NC; 3. Male & AD; and 4. Female & AD) in Figure 7.

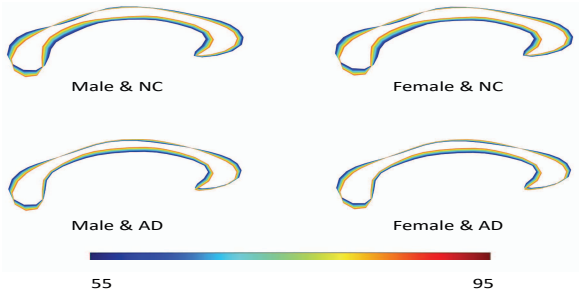

Figure 7. Estimated shapes from age 55 (blue) to age 95 (red) is presented for each of the four subgroups (1. Male & NC; 2. Female & NC; 3. Male & AD; and 4. Female & AD).

Comparing the sequences of estimated shapes from different sub-groups, some differences can be found between the normal control groups and AD groups. In order to investigate the relationship between the pre-shape responses and all the covariates of interest. The hypothesis testing problem is (16) considered and the test statistics along with the p-values are reported in Table 2. It can be found that, there are strong AD effect and age-dependent AD effect on the shape responses while the gender effect is not significant.

Finally, we would like to compare the pre-aligned pre-shapes and the post-aligned pre-shapes derived from Geo-FARM in terms of the mean shape. Here we adopt the bootstrap hypothesis testing approach through the R package *shapes*, where the number of bootstrap is set to 500. Then the test statistic is 0.0209 and the related p-value is 0.0412,

Table 2. Hypothesis testing results.

	gender	age	AD	age×AD
test stat.	101.21	118.37	131.61	128.75
p-value	0.365	0.0693	0.0111	0.0172

which indicates that there is significant difference between the pre-aligned pre-shapes and the post-aligned pre-shapes in terms of the mean shape. In other word, our Geo-FARM does borrow covariate information in learning the nuisance transformations.

4. Conclusion

This paper proposes a geodesic factor regression model for misaligned pre-shape responses, where the additional nuisance rotational effects are built within the proposed model and learned based on both pre-shapes and covariates of interest. In addition, the spatial correlation structure is specified through a low dimensional representation including latent variables on the tangent space and isotropic error terms. Both Monte Carlo simulation studies and real data analysis on ADNI CC contour data show that the proposed model outperforms most existing approaches.

A. Preliminaries from differential geometry

Let $\mathcal{M}$ be a $d_{\mathcal{M}}$-dimensional complete Riemannian manifold with distance function $\text{dist}_{\mathcal{M}}$. We denote the tangent space at $y \in \mathcal{M}$ by $T_y\mathcal{M}$ and the inner product of $u, v \in T_y\mathcal{M}$ by $\langle u, v \rangle$. For any $v \in T_y\mathcal{M}$, there is a unique geodesic curve $\gamma : [0, 1] \to \mathbb{R}$, with initial conditions $\gamma(0) = y$ and $\gamma'(0) = v$. The geodesic is only guaranteed to exist in a neighborhood of y, where the largest neighborhood is denoted by $\mathcal{N}_y \in \mathcal{M}$. The exponential map at y, $\text{Exp}(y, \cdot) : T_y\mathcal{M} \to \mathcal{N}_y$, is locally diffeomorphic and defined as $\text{Exp}(y, v) = \gamma(1)$. The log map $\text{Log}(y, \cdot) : \mathcal{N}_y \to T_y\mathcal{M}$ is defined as the inverse of exponential map. For any $y' \in \mathcal{N}_y$, the Riemannian distance $\text{dist}_{\mathcal{M}}(y, y') = \|\text{Log}(y, y')\|$.

References

[1] K. Ahn, J. Derek Tucker, W. Wu, and A. Srivastava. Elastic handling of predictor phase in functional regression models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 324–331, 2018.

[2] GJ A Amaral, IL Dryden, and Andrew T A Wood. Pivotal bootstrap methods for k-sample problems in directional statistics and shape analysis. *Journal of the American Statistical Association*, 102(478):695–707, 2007.

[3] Alvin H Bachman, Sang Han Lee, John J Sidtis, and Babak A Ardekani. Corpus callosum shape and size changes in early alzheimer's disease: a longitudinal mri study using the oasis brain database. *Journal of Alzheimer's Disease*, 39(1):71–78, 2014.

[4] Jérôme Bolte, Shoham Sabach, and Marc Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming*, 146(1-2):459–494, 2014.

[5] Joseph-Frédéric Bonnans, Jean Charles Gilbert, Claude Lemaréchal, and Claudia A Sagastizábal. *Numerical optimization: theoretical and practical aspects*. Springer Science & Business Media, 2006.

[6] F. L. Bookstein. *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge University Press, 1991.

[7] Emil Cornea, Hongtu Zhu, Peter Kim, Joseph G Ibrahim, and Alzheimer's Disease Neuroimaging Initiative. Regression models on riemannian symmetric spaces. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(2):463–482, 2017.

[8] Luciano da Fona Costa and Roberto Marcond Cesar Jr. *Shape classification and analysis: theory and practice*. CRC Press, 2018.

[9] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977.

[10] Ian L Dryden and Kanti V Mardia. *Statistical shape analysis: with applications in R*, volume 995. John Wiley & Sons, 2016.

[11] Bradley Efron and Robert J Tibshirani. *An introduction to the bootstrap*. CRC press, 1994.

[12] Sahar Elahi, Alvin H Bachman, Sang Han Lee, John J Sidtis, Babak A Ardekani, and Alzheimer's Disease Neuroimaging Initiative. Corpus callosum atrophy rate in mild cognitive impairment and prodromal alzheimer's disease. *Journal of Alzheimer's Disease*, 45(3):921–931, 2015.

[13] Bruce Fischl. Freesurfer. *Neuroimage*, 62(2):774–781, 2012.

[14] P. T. Fletcher. Geodesic regression and the theory of least squares on riemannian manifolds. *International journal of computer vision*, 105(2):171–185, 2013.

[15] P Thomas Fletcher, Conglin Lu, Stephen M Pizer, and Sarang Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE transactions on medical imaging*, 23(8):995–1005, 2004.

[16] Søren Hauberg. Directional statistics with the spherical normal distribution. In *2018 21st International Conference on Information Fusion (FUSION)*, pages 704–711. IEEE, 2018.

[17] Jacob Hinkle, Prasanna Muralidharan, P Thomas Fletcher, and Sarang Joshi. Polynomial regression on riemannian manifolds. In *European Conference on Computer Vision*, pages 1–14. Springer, 2012.

[18] Matthew D Hoffman and Andrew Gelman. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.*, 15(1):1593–1623, 2014.

[19] Chao Huang. *Advanced Statistical Learning Methods for Heterogeneous Medical Imaging Data*. PhD thesis, The University of North Carolina at Chapel Hill, 2019.

[20] Chao Huang, Martin Styner, and Hongtu Zhu. Clustering high-dimensional landmark-based two-dimensional shape data. *Journal of the American Statistical Association*, 110(511):946–961, 2015.

[21] D. G. Kendall, D. Barden, T. K. Carne, and H. Le. *Shape and shape theory*. Wiley, 1999.

[22] H. J. Kim, N. Adluru, M. D. Collins, M. K. Chung, B. B. Bendlin, S. C. Johnson, R. J. Davidson, and V. Singh. Multivariate general linear models (mglm) on riemannian manifolds with applications to statistical analysis of diffusion weighted images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2705–2712, 2014.

[23] Richard A Levine and George Casella. Implementations of the monte carlo em algorithm. *Journal of Computational and Graphical Statistics*, 10(3):422–439, 2001.

[24] Lizhen Lin, Niu Mu, Pokman Cheung, David Dunson, et al. Extrinsic gaussian processes for regression and classification on manifolds. *Bayesian Analysis*, 14(3):887–906, 2019.

[25] Lizhen Lin, Brian St. Thomas, Hongtu Zhu, and David B Dunson. Extrinsic local regression on manifold-valued data. *Journal of the American Statistical Association*, 112(519):1261–1273, 2017.

[26] Radford M. Neal. MCMC using Hamiltonian dynamics. In Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng, editors, *Handbook of Markov Chain Monte Carlo*, chapter 5, pages 113–162. Chapman and Hall/CRC, 2011.

[27] Lynn K Paul, Warren S Brown, Ralph Adolphs, J Michael Tyszka, Linda J Richards, Pratik Mukherjee, and Elliott H Sherr. Agenesis of the corpus callosum: genetic, developmental and functional aspects of connectivity. *Nature Reviews Neuroscience*, 8(4):287–299, 2007.

[28] Xavier Pennec. Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127, 2006.

[29] Xavier Pennec, Stefan Sommer, and Tom Fletcher. *Riemannian Geometric Statistics in Medical Image Analysis*. Academic Press, 2019.

[30] William H Press, Saul A Teukolsky, William T Vetterling, and Brian P Flannery. *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge university press, 2007.

[31] Salem Said, Hatem Hajri, Lionel Bombrun, and Baba C Vemuri. Gaussian distributions on riemannian symmetric

spaces: statistical learning with structured covariance matrices. *IEEE Transactions on Information Theory*, 64(2):752–772, 2017.

[32] Xiaoyan Shi, Martin Styner, Jeffrey Lieberman, Joseph G Ibrahim, Weili Lin, and Hongtu Zhu. Intrinsic regression models for manifold-valued data. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 192–199. Springer, 2009.

[33] Ha-Young Shin and Hee-Seok Oh. Robust geodesic regression. *arXiv preprint arXiv:2007.04518*, 2020.

[34] C. G. Small. *The Statistical Theory of Shape*. Springer, 1996.

[35] Anuj Srivastava, Shantanu H Joshi, Washington Mio, and Xiuwen Liu. Statistical shape analysis: Clustering, learning, and testing. *IEEE Transactions on pattern analysis and machine intelligence*, 27(4):590–602, 2005.

[36] A. Srivastava and E. Klassen. *Functional and Shape Data Analysis*. Springer Series in Statistics, 2016.

[37] Dimosthenis Tsagkrasoulis and Giovanni Montana. Random forest regression for manifold-valued responses. *Pattern Recognition Letters*, 101:6–13, 2018.

[38] Clement Vachet, Benjamin Yvernault, Kshamta Bhatt, Rachel G Smith, Guido Gerig, Heather Cody Hazlett, and Martin Styner. Automatic corpus callosum segmentation using a deformable active fourier contour model. In *SPIE Medical Imaging*, volume 8317, pages 831707–831707–7. International Society for Optics and Photonics, 2012.

[39] J. L. Wang, J. M. Chiou, and H. G. Müller. Functional data analysis. *Annual Review of Statistics and Its Application*, 3:257–295, 2016.

[40] Laurent Younes. *Shapes and diffeomorphisms*, volume 171. Springer, 2010.

[41] Miaomiao Zhang and Tom Fletcher. Probabilistic principal geodesic analysis. In *Advances in Neural Information Processing Systems*, pages 1178–1186, 2013.

[42] X. Zhang, X. Shi, Y. Sun, and L. Cheng. Multivariate regression with gross errors on manifold-valued data. *IEEE transactions on pattern analysis and machine intelligence*, 41(2):444–458, 2018.

[43] Youshan Zhang. Bayesian geodesic regression onriemannian manifolds. *arXiv preprint arXiv:2009.05108*, 2020.