A nonlinear regression technique for manifold valued data with applications to Medical Image Analysis [*]

Monami Banerjee[1], Rudrasis Chakraborty[1], Edward Ofori[2], Michael S. Okun[3], David E. Vaillancourt[2]
and Baba C. Vemuri[1]

[1]Department of CISE, University of Florida, FL 32611, USA
[2]Department of Applied Physiology and Kinesiology, University of Florida, FL 32611, USA
[3]Department of Neurology, University of Florida, FL 32611, USA
[1]{monami, rudrasis, vemuri}@cise.ufl.edu   [2]{eofori, vcourt}@ufl.edu   [3]okun@neurology.ufl.edu

Abstract

Regression is an essential tool in Statistical analysis of data with many applications in Computer Vision, Machine Learning, Medical Imaging and various disciplines of Science and Engineering. Linear and nonlinear regression in a vector space setting has been well studied in literature. However, generalizations to manifold-valued data are only recently gaining popularity. With the exception of a few, most existing methods of regression for manifold valued data are limited to geodesic regression which is a generalization of the linear regression in vector-spaces. In this paper, we present a novel nonlinear kernel-based regression method that is applicable to manifold valued data. Our method is applicable to cases when the independent and dependent variables in the regression model are both manifold-valued or one is manifold-valued and the other is vector or scalar valued. Further, unlike most methods, our method does not require any imposed ordering on the manifold-valued data. The performance of our model is tested on a large number of real data sets acquired from Alzhiemers and movement disorder (Parkinsons and Essential Tremor) patients. We present an extensive set of results along with statistical validation and comparisons.

1. Introduction

An essential task in any regression-based analysis involves finding the relation between two sets of variables, i.e., the independent and the dependent variables. So, given a training data set $\{x_i, y_i\}$ of independent and dependent variables our goal is to regress between these two sets of variables, i.e., find a function $f : x \to y$ s.t. $y_i = f(x_i)$. If both of these variables are vector valued and if there is a linear relationship between them, i.e., $y_i = ax_i + b$ for some a, b, any linear least square estimator can be used to find this relationship. However, in most real applications, this relationship is rather nonlinear and thus, one resorts to the use of either nonlinear least-squares or tools such as support vector regression [1].

Often, one or the other or both the variables lie on a Riemannian manifold which lacks global vector space structure. This lack of vector space structure means that any linear combination of points on the manifold does not lie on that manifold. For example, in general, a linear combination of points on a hypersphere do not lie on that hypersphere. Moreover, in a vector space the linear relation between $\{x_i, y_i\}$ can be expressed as a straight line, $y_i = ax_i + b$, but on a general Riemannian manifold with non-zero sectional curvature, straight lines correspond to geodesic curves. Hence, even for the linear relation, one can not use the linear least squares method on a general Riemannian manifold. Instead, a sophisticated technique is needed even to find a linear relationship for manifold valued data. This poses a restriction on the direct use of well known vector-space based linear/non-linear least-squares type techniques on the manifold. Hence, finding a relationship between manifold valued variables poses a formidable challenge. One may be tempted to use an *embedding* of the manifold valued variables in the Euclidean space (using the Whitney Embedding [2]) and apply the linear/non-linear regression scheme in the Euclidean settings. But, note that often embedding results in a poor estimation of the underlying relationship. Moreover, the data dimension after the embedding becomes larger. For example, using the *strong Whitney*

embedding, one can embed any n-dimensional manifold in a $2n$-dimensional Euclidean space, i.e., $\mathbf{R}^{2n}$. These problems motivated the research community to seek a regression technique applicable to manifold valued data sets. Now, we will briefly present some earlier work in this context.

The most commonly used regression on Riemannian manifolds is the geodesic regression, where, some notion of ordering is imposed on the manifold-valued data [3, 4, 5, 6, 7]. In [4], Fletcher et al. proposed a geodesic regression technique from real-valued to manifold-valued data. Hong et al. [8] proposed a shooting spline based regression technique on the Grassmannian. In [9], Hinkle et al. proposed a polynomial regression method on Riemannian manifolds in a variational framework. The minimization in their problem requires the solution to a system of covariant differential equations. In [10], authors estimate the correlation between shape and age using manifold regression. Recently, a variational spline regression for the manifold of diffeomorphisms was presented in a large deformation diffeomorphic mapping (LDDMM) setting [11]. In [12], authors formulate the manifold regression problem in a regularized risk minimization setting. In [13], Skwerer proposed a regression scheme in phylogenetic tree spaces (negatively curved spaces). Most of these methods however are applicable to cases where the independent variable in the regression problem is scalar valued. Recently, a multivariate general linear model (MGLM) to regress from real-valued vectors to manifold-valued data was proposed in [7]. Further, in [14], Kim et al. generalized the well known Canonical Correlation Analysis (CCA) to Riemannian manifolds. Both these methods use geodesic regression on the manifold, which is the equivalent of the linear regression in vector-spaces. To the best of our knowledge, there is no nonlinear regression method between manifold valued independent and dependent variables or even from manifold to vector valued data.

Recently in [15], authors proposed a non-linear regression technique where, either of the independent or the dependent variables are manifold-valued. Unlike existing methods, this method *does not* require an ordering of the manifold-valued data. Motivated by their work, in this paper, we propose a novel regression scheme in a more complex setting where both of the variables (independent and dependent) are manifold-valued (lie on the same or different Riemannian manifold(s)). Such problems are commonly encountered in Medical Image Analysis. For e.g., given diffusion tensor images (DTIs) derived from diffusion magnetic resonance data sets, to assess changes caused by pathologies, one needs to warp the given DTI to an atlas DTI. The information in the warp can be captured using the Cauchy deformation tensors that are symmetric and positive definite. This gives another tensor field (one deformation tensor at each voxel). Finding the relation between the lo-

cal water molecule diffusion in the tissue and the changes with respect to a reference template (atlas), provides a way to characterize the population. Several other applications include finding relationship between diffusion and conductance tensor fields in Cardiac imaging etc.

The rest of the paper is organized as follows. In section 2, we briefly present some relevant concepts of Riemannian geometry and notations for subsequent use. In section 3, we propose our manifold to manifold regression technique. We report experimental results of our regression method in section 4. In section 5, we extend our regression method to a new robust regression technique on manifold valued data and present experimental results on synthetic data. Finally, in section 6, we draw the conclusion.

2. Mathematical Preliminaries

In this section, we present some basic definitions of terms from Riemannian Geometry that will be used throughout the paper. For further details on these concepts, we refer the readers to [16]. Let $\mathcal{M}$ be a topological space. A *chart of dimension* $n \geq 0$ on $\mathcal{M}$ is a pair $(U, \phi), U \subset \mathcal{M}$ open, and $\phi : U \to \mathbf{R}^n$ is a diffeomorphism onto an open subset of $\mathbf{R}^n$. An *atlas on* $\mathcal{M}$ is a collection of charts, $\mathcal{U} = \{U_\alpha, \phi_\alpha\}_{\alpha \in A}$ such that each pair of charts is C^∞ compatible and $\{U_\alpha\}$ is an open cover of $\mathcal{M}$. A (smooth) manifold is a pair $(\mathcal{M}, \mathfrak{M})$, where $\mathcal{M}$ is a topological manifold and $\mathfrak{M}$ is an atlas on $\mathcal{M}$. For simplicity, we will use $\mathcal{M}$ to denote a smooth manifold.

Let $p \in \mathcal{M}$. Then, define $S_p = \{\gamma : I \to \mathcal{M}, I \text{ open}, 0 \in I, \gamma(0) = p\}$. Define an equivalence relation $\sim_p$ on S_p as $\gamma_1 \sim_p \gamma_2$ iff $\gamma_1'(0) = \gamma_2'(0)$. Then, *the tangent space of* $\mathcal{M}$ *at* p, denoted by $T_p\mathcal{M}$ is the set $S_p / \sim$, $T_p\mathcal{M} \simeq \mathbf{R}^n$, where $n = dim(\mathcal{M})$. The *tangent bundle* of $\mathcal{M}$ is defined as a set by $T\mathcal{M} = \cup_{p \in \mathcal{M}} T_p\mathcal{M}$.

The Riemannian metric on $\mathcal{M}$ is a field g of smoothly varying inner products on the tangent spaces. A Riemannian manifold, $(\mathcal{M}, g)$ is a manifold equipped with a Riemannian metric g. A *connection* on $T\mathcal{M}$ is a map $\nabla : \Gamma(T\mathcal{M}) \times \Gamma(T\mathcal{M}) \to \Gamma(T\mathcal{M})$ which is $\mathfrak{F}-$linear in the first argument and Leibnitzian in the second argument. There $\exists!$ a torsion-free connection on $T\mathcal{M}$ which respects g. This connection is called the *Levi-Civita connection*. In the rest of the paper, we use $(\mathcal{M}, g)$ to denote a Riemannian manifold equipped with a Levi-Civita connection ∇.

A *vector field along a smooth map* γ is a map $Y : I \to T\mathcal{M}$ such that $Y(t) \in T_{\gamma(t)}\mathcal{M}$. We say that Y is *parallel along* γ if $\nabla_{\gamma'(t)} Y \equiv 0$. Let $Y_0 \in T_{\gamma(t_0)}\mathcal{M}$, where $t_0 \in I$. Then, $\exists!$ parallel vector field Y along γ, such that $Y(t_0) = Y_0$. $Y(t)$ is called the *parallel transport* of $Y(t_0)$ along γ. A *geodesic* is a curve $\gamma : I \to \mathcal{M}$ such that $\nabla_{\gamma'} \gamma' = 0$. For $\mathbf{v} \in T\mathcal{M}$, $\exists!$ geodesic $\gamma_{\mathbf{v}} : I \to \mathcal{M}$ with $\gamma_{\mathbf{v}}' = \mathbf{v}$. If $1 \in I$, i.e., $\gamma_{\mathbf{v}}(1)$ is defined, then we can define $\mathrm{Exp}(\mathbf{v}) = \gamma_{\mathbf{v}}(1)$. So, we can define $\mathrm{Exp} :$

$\{\mathbf{v} \in T\mathcal{M} | \gamma_{\mathbf{v}}(1) \text{ is defined}\} \to \mathcal{M}$. The *geodesic distance function* on a connected manifold $\mathcal{M}$ determined by g is $\rho = \rho_g : \mathcal{M} \times \mathcal{M} \to \mathbf{R}$ by $\rho(p, q) = \inf\{L(\gamma) | \gamma \in \Omega_{p,q}\}$, where $\Omega_{p,q} = \{\gamma : [0, 1] \to \mathcal{M} \text{ piecewise smooth}, \gamma(0) = p, \gamma(1) = q. \ L(\gamma) = \int_0^1 \|\gamma'(t)\| dt$. It's easy to show that ρ_g is a metric and $\mathcal{M}$ is a metric space. It can be shown that, $\forall p \in \mathcal{M}$, $\exists$ open neighborhood W of p, $\epsilon > 0$, such that $\forall q \in W$, $\mathrm{Exp}_q \restriction_{B_\epsilon(0) \subset T_q \mathcal{M}}$ is a diffeomorphism onto an open neighborhood U_q with $q \in U_q$ and $W \subset U_q$. Exp_q is called Riemannian Exponential map. On U_q, Exp_q^{-1} is defined and is called Riemannian inverse Exponential map (denoted by Log). A Riemannian manifold is called *geodesically complete* (or complete) if domain$(\mathrm{Exp}) = T\mathcal{M}$. For a complete manifold, there exists a minimal geodesic between any two points on the manifold.

Without going into the definition and details of *sectional curvature*, for the rest of this paper, we will assume that the data points lie inside a geodesic ball of convexity radius $< \rho$ where $\rho = \frac{1}{2} \min(conv(\mathcal{M}), \frac{\pi}{\sqrt{\Delta}})$ [17] where $conv(\mathcal{M})$ is the convexity radius of $\mathcal{M}$ and Δ is the upper bound on the sectional curvature. This assumption is needed to ensure that the Riemannian ℓ_p center of mass exists and is unique.

3. Problem Formulation and Algorithm

Let $(\mathcal{M}, g^{\mathcal{M}})$ and $(\mathcal{N}, g^{\mathcal{N}})$ be complete Riemannian manifolds [16] with $g^{\mathcal{M}}$ and $g^{\mathcal{N}}$ being the Riemannian metrics on the Riemannian manifolds $\mathcal{M}$ and $\mathcal{N}$, respectively. Given $\{x_i, y_i\}_{i=1}^N \subset \mathcal{M} \times \mathcal{N}$, our goal is to find a function $f : \mathcal{M} \to \mathcal{N}$ s.t., $y_i = f(x_i), \forall i$. Let, $d_{\mathcal{M}} : \mathcal{M} \times \mathcal{M} \to \mathbf{R}$ be the distance function on $\mathcal{M}$, i.e., $d_{\mathcal{M}}(x_i, x_j) = g_{x_i}^{\mathcal{M}}(\mathrm{Log}_{x_i} x_j, \mathrm{Log}_{x_i} x_j)$, where Log is the Riemannian inverse Exponential map (Note that the completeness assumption of manifold ensures that Log map is defined on the entire manifold, but for a manifold which is not geodesically complete, within a geodesic ball of appropriate radius (mentioned in Section 2), the Log map is well defined). Let $d_{\mathcal{N}}$ be the distance function on $\mathcal{N}$ induced by the Riemannian metric $g^{\mathcal{N}}$. Now, we can estimate f by minimizing the following objective function, $E = \frac{1}{N} \sum_{i=1}^N d_{\mathcal{N}}(y_i, \hat{y}_i)^2$, where $\hat{y}_i$ is the predicted y_i defined by,

$$\hat{y}_i = \hat{f}(x_i) = \arg\min_{\mu \in \mathcal{N}} \sum_{j=1}^k \frac{\mathcal{K}(x_i, t_j)}{\sum_{l=1}^k \mathcal{K}(x_i, t_l)} d_{\mathcal{N}}(c_j, \mu)^2 \tag{1}$$

where, $\{c_j\}_{j=1}^k \subset \mathcal{N}$ and $\{t_j\}_{j=1}^k \subset \mathcal{M}$ are the *representatives* on $\mathcal{N}$ and $\mathcal{M}$ respectively. $\mathcal{K} : \mathcal{M} \times \mathcal{M} \to \mathbf{R}$ is the *kernel function* (not necessarily positive-definite). Thus, $\hat{y}$ is approximated as a weighted Fréchet mean (FM) [18] of $\{c_j\}_{j=1}^k$. We use $\{t_j\}_{j=1}^k$ as the cluster representatives (of the given manifold-valued data) and estimate $\{c_j\}_{j=1}^k$ using

the steepest descent on the objective function. The direction of the gradient of E with respect to c_j is given by,

$$D_{c_j} E = -\frac{2}{N} \sum_{i=1}^N \mathrm{Log}_{\hat{y}_i} y_i \ D_{c_j} \hat{y}_i. \tag{2}$$

Where, c_j and $\hat{y}_i$ both are on $\mathcal{N}$, hence, we will use chart maps [16] to compute $D_{c_j} \hat{y}_i$. Let $\{U_\alpha, \Phi_\alpha\}_{\alpha \in \mathcal{I}}$ be the chart map of $\mathcal{N}$. Without loss of generality, assume $c_j \in U_{\alpha_1}$ and $\hat{y}_i \in U_{\alpha_2}$ where, $(U_{\alpha_1}, \Phi_{\alpha_1})$ and $(U_{\alpha_2}, \Phi_{\alpha_2})$ are the corresponding charts. Note that, by definition, chart maps are diffeomorphisms. Given a fixed x_i, from Eq. 1, let us define a function $\mathcal{F} : \mathcal{N}^k \to \mathcal{N}$ by $\hat{y}_i = \mathcal{F}\left(\{c_j\}_{j=1}^k\right)$. Now, we can define $D_{c_j} \hat{y}_i$ as $D_{c_j} \hat{y}_i := D_{\tilde{c}_j} \widetilde{\mathcal{F}}$, where, $\tilde{c}_j = \Phi_{\alpha_1}(c_j)$ and $\widetilde{\mathcal{F}} = \Phi_{\alpha_2} \circ \mathcal{F} \circ \Phi_{\alpha_1}^{-1} : \mathbf{R}^n \to \mathbf{R}^n$, with $n = dim(\mathcal{N})$. Hence, $D_{\tilde{c}_j} \widetilde{\mathcal{F}}$ is the Jacobian of $\widetilde{\mathcal{F}}$.

Note that, $D_{c_j} E \in T_{c_j} \mathcal{N}$, thus, in order to make the RHS of Eq. 2 to be in $T_{c_j} \mathcal{N}$, we use parallel transport of $\mathrm{Log}_{\hat{y}_i} y_i$ from $\hat{y}_i$ to c_j. For a general Riemannian manifold $\mathcal{N}$, as parallel transport is not easy to compute, we can approximate this parallel transport, $\Lambda_{c_j} \mathrm{Log}_{\hat{y}_i} y_i$ by, $\Lambda_{c_j} \mathrm{Log}_{\hat{y}_i} y_i \approx \mathrm{Log}_{c_j} y_i - \mathrm{Log}_{c_j} \hat{y}_i$. Moreover, for a general Riemannian manifold, as the weighted FM is not in closed form for more than two samples, $\widetilde{\mathcal{F}}$ is not in closed form, hence, computation of the Jacobian is not feasible. But, recently, an efficient recursive FM estimator was proposed for several Riemannian manifolds including the manifold of symmetric positive definite matrices, $SPD(m)$ [19, 20], hypersphere, $\mathbf{S}^m$ [21] and the Grassmannian, $Gr(p, m)$ [22]. We use this recursive FM estimator to compute $\widetilde{\mathcal{F}}$ (and the Jabcobian) in closed form for these aforementioned manifolds. But, for other Riemannian manifolds, we approximate Eq. 1 (in the spirit of [7, 23]) by $\hat{y}_i \approx \mathrm{Exp}_p\left(\sum_{j=1}^k \mathcal{K}(t_j, x_i) \mathrm{Log}_p c_j\right)$, where, $p \in \mathcal{N}$ and Exp is the Riemannian Exponential map. For the sake of completeness, we will briefly present the recursive FM estimator formulation here (as given in [19, 21, 22]). Let $\mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_k$ be independent samples drawn from a probability distribution $P(\mathcal{X})$ on a complete manifold $\mathcal{N}$ with a set of associated weights $\{w_j\}_{j=1}^k$ such that $\sum_j w_j = 1$. Then, we define the weighted Fréchet mean estimator M_k by the following recursion:

$$M_1 = \mathcal{X}_1 \tag{3}$$

$$M_{l+1} = \Gamma_{M_l}^{\mathcal{X}_{l+1}}\left(w_{l+1}/(\sum_{i=1}^{l+1} w_i)\right) \tag{4}$$

where, $\Gamma_{M_l}^{\mathcal{X}_{l+1}} : [0, 1] \to \mathcal{N}$ is the shortest geodesic between M_l and $\mathcal{X}_{l+1}$, i.e., $\Gamma_{M_l}^{\mathcal{X}_{l+1}}(0) = M_l$, $\Gamma_{M_l}^{\mathcal{X}_{l+1}}(1) = \mathcal{X}_{l+1}$. Note that, this formulation can be easily extended to any complete Riemannian manifold (or within a geodesic ball

of specific radius of any Riemannian manifold) where a closed form expression of the Riemannian Exponential and Inverse Exponential map exist. But, in order to be a valid FM estimator, one needs to show its consistency which is proved in [19, 21, 22] for $SPD(m)$, $\mathbf{S}^m$ and $Gr(p,m)$ respectively. Hence, for Riemannian manifolds other than the aforementioned three, we will use the approximation discussed above. Now, we sketch our manifold regression algorithm for a general Riemannian manifold below.

Algorithm 1: Nonlinear Regression of Manifold Valued Data, Training Stage.

Input: $\{x_i\}_{i=1}^N \subset \mathcal{M}$, $\{y_i\}_{i=1}^N \subset \mathcal{N}$, $k, b, \eta, \epsilon > 0$
Output: $\{t_j\}_{j=1}^k \subset \mathcal{M}$, $\{c_j\}_{j=1}^k \subset \mathcal{N}$

1 Compute $\{t_j\}_{j=1}^k$ as k cluster centers of $\{x_i\}_{i=1}^N$;
2 Compute $K_{N \times k}$ matrix, where $K_{i,j} = \mathcal{K}(x_i, t_j)$;
3 Initialize $\{c_j\}_{j=1}^k$ as,
  $c_j = \arg\min_{\mu \in \mathcal{N}} \sum_{\{l|x_l \in k_j\}} w_{l,j}\, d_{\mathcal{N}}(y_l, \mu)$, where k_j is the jth cluster, and $w_{l,j} = K_{l,j}/\sum_{\{l|x_l \in k_j\}} K_{l,j}$;
4 Compute $W_{N \times k}$ matrix, where $W_{i,j} = K_{i,j}/\sum_{l=1}^k K_{i,l}$;
5 Compute the objective function $E = \frac{1}{N}\sum_i d_{\mathcal{N}}(y_i, \hat{y}_i)^2$, using the predictions $\hat{y}_i$'s in Eq. (1), and c_j values in line (3) above;
6 $E^{\text{old}} \leftarrow E$, and $flag \leftarrow 1$;
7 **while** $flag = 1$ **do**
8 　Compute $D_{c_j}E, \forall j = 1, \cdots, k$, using Eq. (2);
9 　$c_j^{\text{new}} \leftarrow \text{Exp}_{c_j}(-\eta\, D_{c_j}E)$;
10 　Recompute $\{\hat{y}_i\}_{i=1}^N$ and E, using $\{c_j^{\text{new}}\}_{j=1}^k$;
11 　**if** $\|D_{c_j}E\| < \epsilon, \forall j = 1, \cdots, k$ **then**
12 　　$flag \leftarrow 0$;
13 　**end**
14 　**if** $E < E^{old}$ **then**
15 　　$c_j \leftarrow c_j^{\text{new}}$, and $E^{\text{old}} \leftarrow E$;
16 　**else**
17 　　$c_j^{\text{new}} \leftarrow c_j$, $E \leftarrow E^{\text{old}}$, and $\eta \leftarrow 0.9\,\eta$;
18 　**end**
19 **end**

Algorithm 2: Nonlinear Regression of Manifold Valued Data, Testing Stage.

Input: $x \in \mathcal{M}$, $\{t_j\}_{j=1}^k \subset \mathcal{M}$, $\{c_j\}_{j=1}^k \subset \mathcal{N}$
Output: $\hat{y} \in \mathcal{N}$
1 Compute $\{w_j\}_{j=1}^k$ as $w_j = \mathcal{K}(x, t_j)/\sum_{l=1}^k \mathcal{K}(x, t_l)$;
2 $\hat{y} = \arg\min_{\mu \in \mathcal{N}} \sum_{j=1}^k w_j\, d_{\mathcal{N}}(c_j, \mu)$;

Special Case: $SPD(m) \to \mathbf{S}^n$ **regression**

Here, we present the derivation of our regression formulation where, $\mathcal{M} = SPD(m)$ and $\mathcal{N} = \mathbf{S}^n$. We will use the GL-invariant Riemannian metric on $SPD(m)$, the induced distance is given by $d_{\mathcal{M}}(x,t) = \text{Tr}((\text{Log}(x^{-1}t))^2)$, where Log and Tr are the matrix logarithm and matrix trace operators respectively, and $x, t \in SPD(m)$. On $\mathbf{S}^n$, we chose the arc length metric, and the distance $d_{\mathcal{N}}(y, c) = \arccos(y^t c)$, where $y, c \in \mathbf{S}^n$. Now, using Eq. 1, $\hat{y}_i$ is a weighted FM

of $\{c_j\}$, where the jth weight, $w_j = \mathcal{K}(x_i, t_j)$. In this work, we chose, $\mathcal{K}(x_i, t_j) = \exp(-b\, d_{\mathcal{M}}(x_i, t_j)^2/2\sigma^2)$, b and σ being the kernel parameters. Using the inductive FM estimator presented in [21], $D_{c_j}\hat{y}_i$ is computed as follows. $\hat{y}_i = M_k$, i.e., the kth intrinsic mean estimator, hence $D_{c_j}\hat{y}_i = D_{M_{k-1}}M_k\, D_{M_{k-2}}M_{k-1} \cdots D_{M_j}M_{j+1}\, D_{c_j}M_j$, where $D_{M_{l-1}}M_l = (1 - s_l)\sin((1 - s_l)\,\theta_l)/\sin(\theta_l)\, I_{n+1}$, and $D_{c_l}M_l = s_l\sin(s_l\,\theta_l)/\sin(\theta_l)\, I_{n+1}$. Here, $s_l = w_l/(\sum_{p=1}^l w_p)$, $\theta_l = \arccos(c_l^t M_{l-1})$.

4. Experimental Results

In this section, we evaluate the performance of our proposed nonlinear regression technique via several experiments on real datasets. In all of these experiments, we have used the kernel parameter σ^2 as the data variance. In order to measure the accuracy of the regression, we use the R^2 statistical measure on Riemannian manifolds, described in [4]. The R^2 statistic on a Riemannian manifold captures the fraction of the data variance that can not be explained by the regression model. Let the *unexplained variance* be defined as $\sum_{i=1}^N d_{\mathcal{N}}(y_i, \hat{y}_i)^2$. Then, the R^2 statistic is given by $R^2 = 1 - \frac{unexplained\ variance}{data variance} \in [0, 1]$. The value of the R^2 statistic lies in the interval $[0, 1]$, and a large value indicates a better regression performance. We vary the b value from 1 to 100 in increments of 1 and select the b value using a cross-validation scheme. In order to measure the statistical significance of our results, we perform the following statistical analysis. We executed 100 independent runs with varying b-value and fit a normal distribution to the 100 R^2 statistic values. Let this random variable be denoted by $\mathcal{X}$. Now, we fit a normal distribution to the R^2 statistic from the 100 independent runs with varying b value and a random permutation of the independent variables. Let this random variable be denoted by $\mathcal{Y}$. The null hypothesis is set to, H_0: *mean of $\mathcal{X}$ = mean of $\mathcal{Y}$*. We reject the null hypothesis with a significance level of 0.01 (*this p-value is denoted as p_1 in the rest of this section*).

In the second analysis setting, we consider the b value (denoted by b_M) which yields the largest R^2 statistic (denoted by r_M^2). Now, for $b = b_M$, we execute 100 independent runs by randomly permuting the independent variables and fit a student's t-distribution to these R^2 values. Let this random variable be denoted by $\mathcal{Z}$. The null hypothesis is then set to, $H_0 : r_M^2$ *comes from the distribution of $\mathcal{Z}$*. As before, we reject the null hypothesis with a 0.01 significance level (this p-value is denoted as p_2 in the rest of this section). Moreover as in [4], we performed the following test. For b_M, we execute 100 independent runs by randomly permuting the independent variables and compute the R^2 statistics. Let this 100 R^2 statistics be denoted by Z. Then, we report what fraction of Z is larger than r_M^2 (denoted by the *f-value*). Note that, the *f-value* is in the range

[0, 1], with smaller values being preferred. The f-value is a measure used to see if there is a relationship between the independent and the dependent sets of variables. We now provide the detailed experimental results on two real datasets.

OASIS data [24]: We used the OASIS data [24] to perform several regression tasks. This data consists of 36 T1 magnetic resonance (MR) brain scans of subjects with varying ages in the range of 18 to 96, including early stage AD patients.

We construct two different data representations as follows. (i) First, we segmented the Corpus Callosum (CC) from these MR brain scans. Then, we take points on the 3D boundary of the CC and map it to $\mathbf{S}^{24575}$ using the Schrödinger distance transform (SDT) [25]. (ii) We used a set of landmark points on the boundary of CC and map each of these point sets into the Kendall's shape space ($\mathbf{C}P^n$) [26], which is a *complex projective space*. In this experiment, $n = 249$. First, we seek to model the relationship between $\mathbf{C}P^n$ and the SDT representation of CC shapes. The regression results are given in Table 1 and Table 2. Since this relationship is fairly complex, we can say that the R^2 statistic values are very good even for a small number of control points (k). We vary k over $\lceil N/4 \rceil$, $\lceil N/3 \rceil$, and $\lceil N/2 \rceil$, i.e., we took k to be $9, 12$ and 18 respectively. Note that, if $k = N$, then the model will memorize the data rather than learn from the data. Hence, a very large k is not a good choice. Also note that, the R^2 statistic values for $\mathbf{C}P^n$ to SDT regression are better than the vice-versa case. This is justified because, SDT does not contain enough information to recover the shape while one can recover the shape (up to an rigid transformation) from $\mathbf{C}P^n$ representation. Thus, given a point on $\mathbf{S}^n$, representing an SDT, finding it's corresponding point in the $\mathbf{C}P^n$ representation correctly, is a much harder problem. This justifies the comparatively smaller R^2 statistics in Table 2. Moreover, as evident, since the shape of the CC varies a lot with age, it would be useful to find a regression model from these two representations of CC shapes to age. The regression results from $\mathbf{C}P^n$ and SDT to age are given in Tables 3 4 respectively. From these two tables, one can see that SDT is a better representative of CC shapes than $\mathbf{C}P^n$ for the task of finding a relationship between shape and age. The usefulness of SDT features in finding the relationship between age and CC shapes were also noted in [15], where the authors regressed CC shapes from age information. Moreover, Kendall's shape space representation is a canonical representation of shapes. Hence, the usefulness of regression between SDT and $\mathbf{C}P^n$ is evident. As mentioned before, to the best of our knowledge there does not exist any manifold to manifold (or even manifold to real) regression technique in the literature, so we could not compare the performance of our method with any previous work in this context.

We have given the performance of our regressor for CC

k	b_M	r_M^2	p_1	p_2	f-val
9	63	0.479	< 0.01	< 0.01	0
12	59	0.752	< 0.01	< 0.01	0
18	98	0.754	< 0.01	< 0.01	0

Table 1: *OASIS*: $\mathbf{C}P^n$ to SDT regression

k	b_M	r_M^2	p_1	p_2	f-val
9	77	0.348	< 0.01	< 0.01	0
12	62	0.626	< 0.01	< 0.01	0
18	109	0.631	< 0.01	< 0.01	0

Table 2: *OASIS*: SDT to $\mathbf{C}P^n$ regression

k	b_M	r_M^2	p_1	p_2	f-val
9	100	0.888	< 0.01	< 0.01	0
12	93	0.988	< 0.01	< 0.01	0
18	62	0.987	< 0.01	< 0.01	0

Table 3: *OASIS*: SDT to age ($\mathbf{R}_+$) regression

k	b_M	r_M^2	p_1	p_2	f-val
9	80	0.474	< 0.01	< 0.01	0
12	90	0.737	< 0.01	< 0.01	0
18	56	0.719	< 0.01	< 0.01	0

Table 4: *OASIS*: $\mathbf{C}P^n$ to age ($\mathbf{R}_+$) regression

shapes in Figure 1. The top row contains six sample shapes and the bottom two rows have the corresponding regressed shapes with varying number of control points. We also present the number of control points and R^2 statistics for these reconstructions. The value of b used here is 109. These results provide an evidence of the good performance of our proposed regression method.

Movement Disorder data: This dataset consists of High Angular Resolution Diffusion Image (HARDI) scans from, (i) healthy controls, (ii) patients with essential tremor (ET) and (iii) Parkinson's disease (PD) patients. This data pool contains HARDI scans from 25 controls, 15 ET and 26 PD patients. These HARDI data were acquired using a single-shot spin echo EPI sequence, with repetition time = 7748 ms, echo time = 86 ms, flip angle = 90°, field of view = 224 × 224 mm, voxel size = 2 mm isotropic with no gap between slices ($n = 60$), number of diffusion gradient (monopolar) directions = 64, diffusion gradient timing DELTA/ delta = 42.4/10 ms, b-values: 0, 1000 s/ mm2, fat suppression was performed using SPIR, in-plane, SENSE factor = 2. The dimension of each image is 112 × 112 × 60.
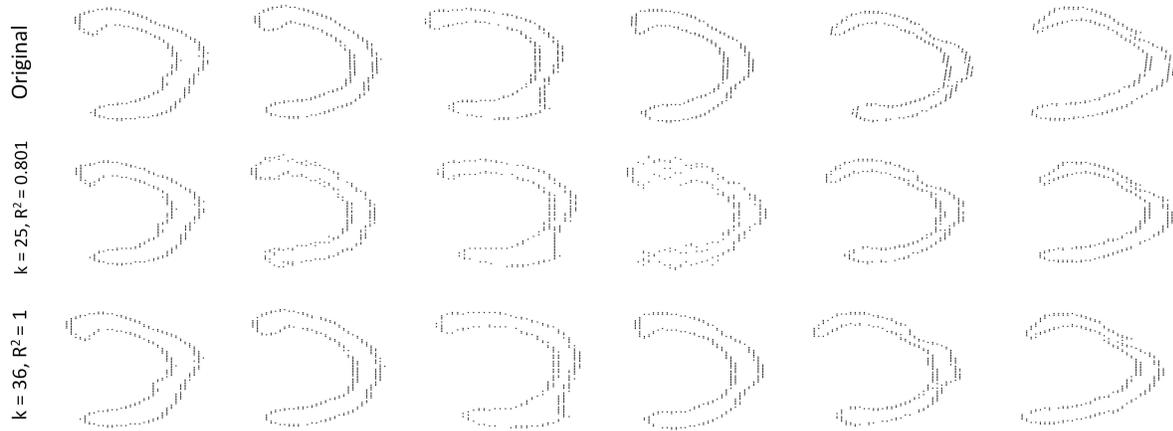
Figure 1: Reconstruction of Corpus Callosum shapes. The top row depicts the original shapes and the bottom two rows depict the regressed shapes.

From each of these images, we identify the region of interest (ROI) (40 voxels in size) containing the Substantia Nigra, a neuroanatomical structure known to be affected most by PD and ET. Then, from the HARDI data within the ROI we compute the ensemble average propagator (EAP), a local probability density function that fully captures the local diffusional characteristics of the tissue. The EAP density function is extracted using methods described in [27] and represented by a Gaussian mixture model, with fixed eigen values for their covariance matrices. This is done to facilitate the representation of multiple neuronal fiber bundles in a voxel and since the fibers are tubular in structure, the eigen values are assumed to satisfy $\lambda_1 > \lambda_2 = \lambda_3$. Thus the degrees of freedom are controlled by the eigen vector orientations. We choose 321 directions for tessellation of the sphere (of directions), using an icosadodecahedron as in [27]. Thus, the EAP field now has a discrete representation of 40 voxels, with each voxel containing a probability vector of size 321.

In morphometric analysis, it is common to use the Cauchy deformation tensor (CDT) field to capture changes in a patient scan with respect to a reference template/ atlas. Thus, in order to capture changes in a patient HARDI scan with respect to the control atlas, we first nonrigidly register each EAP-field estimated from the HARDI data to the EAP atlas and obtain the CDT at each voxel in the ROI, given by $\sqrt{JJ^T}$, where, J is the Jacobian of the non-rigid transformation [28]. The CDT is a symmetric positive definite matrix (SPD) of dimension 3×3 in this case. Hence, for each patient we extract a CDT field of dimension $3 \times 3 \times 40$. In this experiment, we seek to find the relationship between structural information in the form of CDT and clinical measures such as the MDS-UPDRS (Movement Disorder Society's Unified Parkinson's Disease Rating Scale) [29]. The

MDS-UPDRS score is widely used to follow the longitudinal course of PD. These scores are obtained via interviews and clinical observations by an expert. In this experiment, available to us are the MDS-UPDRS scores of 58 subjects in the population under consideration, including 21 controls, 13 ET, and 24 PD patients. This score is a nonnegative integer, with smaller values indicating normality of the patient.

For these 58 patients, we first find a relationship between the EAP field and the MDS-UPDRS score by regressing the EAP against the MDS-UPDRS. We did the similar experiment as before. We vary k to be $\lceil N/4 \rceil$, $\lceil N/3 \rceil$, and $\lceil N/2 \rceil$. The results are given in Table 5. From the results, it is evident that even for small number of control points, i.e., $k = 15$, the R^2 statistic is quite high (very high) and the statistical significance of our result can be ensured by examining the two p-values. Like before the f-value denotes the number of times the regression result on a random permutation of independent variables is better than the reported R^2 statistic. And the small f-values, i.e., 0 signify that there is indeed a relationship between the two sets of variables. So, by examining the combination of these statistical measures reported, one can see that the regression result is good not because there is over fitting but because, there exists a relationship (inferred from f-value) which is captured well (inferred from R^2 statistics and p-values) by our method.

Finally, we present an experiment to find a relation between the two distinct structural representations derived from the HARDI data, namely, the EAP field and Cauchy deformation tensor field. The former captures the local diffusional characteristics of the tissue being imaged and the latter captures the changes between the imaged sample and the control atlas. We first regress the Cauchy deformation tensor field to EAP field and the regression result is given

k	b_M	r_M^2	p_1	p_2	f-val
15	98	0.815	< 0.01	< 0.01	0
20	74	0.868	< 0.01	< 0.01	0
29	25	0.925	< 0.01	< 0.01	0

Table 5: *Movement Disorder*: EAP to MDS-UPDRS ($\mathbf{N} \cup \{0\}$) regression

in Table 6. Note that, here we have taken all the subjects, i.e., sample size is 67. Analogous to the previous experiments we vary k and b. Though the results reported are statistically significant, the R^2 statistics are not very high. The possible reasons behind this comparatively small R^2 statistics are two-fold. (i) The dimensionality of either of the independent or dependent variable is very large. Moreover they lie on different Riemannian manifolds, which are far more "complex" compared to a vector space. (ii) The relationship between Cauchy deformation tensor and EAP is highly nonlinear and complex, and the number of sample points, i.e., 67 is far less compared to the dimension of either of the independent or dependent variable. Hence, in this example with a complex relationship, the relatively small R^2 statistic values are justified.

k	b_M	r_M^2	p_1	p_2	f-val
17	25	0.339	< 0.01	< 0.01	0
23	25	0.426	< 0.01	< 0.01	0
34	20	0.599	< 0.01	< 0.01	0

Table 6: *Movement Disorder*: Cauchy tensor field to EAP field regression

k	b_M	r_M^2	p_1	p_2	f-val
17	15	0.331	< 0.01	< 0.01	0
23	10	0.407	< 0.01	< 0.01	0
34	10	0.573	< 0.01	< 0.01	0

Table 7: *Movement Disorder*: EAP field to Cauchy tensor field regression

Now, we present the regression from EAP field to Cauchy deformation tensor field. The results for this regression are reported in Table 7. From a statistical accuracy view point, these results are analogous to the earlier case of mapping between the Cauchy deformation tensor field to the EAP field regression. This is justified given that both Cauchy and EAP tensor fields are equally good representatives of HARDI data.

Data	Our method			CCA
	k	b_M	r_M^2	CC
$SPD(3) \to SPD(3)$	20	4	0.895	0.771
$\mathbf{S}^9 \to \mathbf{S}^9$	100	5	0.998	0.996

Table 8: Comparative results with CCA

Comparison with CCA [14]: We now compare our manifold regression technique with Cross Correlation Analysis for manifold-valued data by Kim et al. [14]. CCA can be applied to find the relation between two manifold valued data, X and Y, on the same Riemannian manifold. Since for none of our data sets, independent and dependent variables are on the same manifold, we present a comparison for two synthetic data sets. The first data is on $SPD(3)$, here we first draw 1000 samples in $\mathbf{R}^6$ from Normal distribution with 0 mean and small variance. Then we create a 3×3 symmetric matrix from each of these vectors. Each of these symmetric matrices lie on the tangent space anchored at I_3 (the 3×3 identity matrix). As $SPD(n)$ is geodesically complete, we use Exp map to map each tangent vector to a point on $SPD(3)$. Now, for each $A \in SPD(3)$, obtain $B \in SPD(3)$ using RAR^T, where R is a randomly generated matrix in $SO(3)$. Note that, all the eigenvalues of A and B are same. Now, we do regression and CCA between $\{A\}$ and $\{B\}$. For the second data, we randomly generate 1000 points on $\mathbf{S}^9$ and scale and translate these points and project them back on $\mathbf{S}^9$. Then, we applied CCA and regression to these two sets of variables. But, as cross-correlation (CC) and R^2 statistic are different metrics, the rationale between comparing the two is as follows. Note that, R^2 statistic can be thought as a square of the CC between $\{Y\}$ and $\{\hat{Y}\}$. Moreover, for only one independent variable regression, it is also the square of the CC between $\{X\}$ and $\{Y\}$. Hence, there is a relation (though not explicit for the manifold valued cases) between these two measures. Moreover, a high absolute CC value and an R^2 statistic value close to 1 imply there is a strong relation between these two sets of variables. Hence, though they are different metrics, it is still meaningful to compare and contrast them. The comparison results are presented in Table 8. From this table, it is evident that for both the synthetic data sets, our regression method yields a comparatively better performance than CCA.

5. Robust formulation for manifold value regression

In this section, we show how to extend the nonlinear regression formulation presented earlier to cope with outliers in the data. This leads to a new and robust formulation for the regression of manifold valued data. Let $(\mathcal{M}, g^{\mathcal{M}})$ and $(\mathcal{N}, g^{\mathcal{N}})$ be complete Riemannian manifolds. Given

$\{x_i, y_i\}_{i=1}^N \subset \mathcal{M} \times \mathcal{N}$, our goal is to find a function $f : \mathcal{M} \to \mathcal{N}$ s.t., $y_i = f(x_i), \forall i$. But, in this section, we assume that $\{y_i\}$ are corrupted with outliers. It is well known, that an ℓ_1 norm based formulation is much more robust to outliers than the one based on the ℓ_2 norm. Hence, instead of using the weighted Fréchet mean (FM) of $\{c_j\}_{j=1}^k$ as in Eq. 1, we use the *weighted Fréchet median (FMd)*. For $\mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_k \in \mathcal{N}$, with associated weights $\{w_j\}$ with $\sum w_j = 1$, the *weighted FMd*, $\bar{M}$ is defined as follows:

$$\bar{M} = \arg \min_{\mu \in \mathcal{N}} \sum_{j=1}^k w_j \, d_{\mathcal{N}}(\mu, \mathcal{X}_j) \quad (5)$$

Using this formulation, we reformulate Eq. 1 as follows:

$$\hat{y}_i = \hat{f}(x_i) = \arg \min_{\mu \in \mathcal{N}} \sum_{j=1}^k \frac{\mathcal{K}(x_i, t_j)}{\sum_{l=1}^k \mathcal{K}(x_i, t_l)} \, d_{\mathcal{N}}(c_j, \mu) \quad (6)$$

As in the case of the FM, the FMd also has no closed form expression. Hence, we use the incremental FMd computation formulation in spirit of the work in [30]. Note that in [30], the formulation involves a stochastic gradient descent for finding the ℓ_p FM, while our formulation is deterministic, i.e., the next data point is fixed and not selected at random within the neighborhood. The incremental FMd formulation is given below. Assuming the above hypothesis holds, we define the incremental FMd estimator, $\bar{M}_k$ as follows:

$$\bar{M}_1 = \mathcal{X}_1 \quad (7)$$

$$\bar{M}_{l+1} = \operatorname{Exp}_{\bar{M}_l} \left(\frac{w_{l+1}}{\left(\sum_{i=1}^{l+1} w_i \right)} v_l \right) \quad (8)$$

where $v_l = \operatorname{Log}_{\bar{M}_l}(\mathcal{X}_{l+1}) / d_{\mathcal{N}}(\bar{M}_l, \mathcal{X}_{l+1})$. Though this method is deterministic in contrast to [30], one might prove the consistency of $\bar{M}_k$ to the true FMd in a similar way as in [30]. Notice that, the formulation in Eq. 7 requires only the Riemannian Exponential and the inverse Exponential maps, respectively. For example, on $\mathbf{S}^n$, $\bar{M}_{l+1}$ is:

$$\bar{M}_{l+1} = \cos(s_{l+1}) \bar{M}_l + \sin(s_{l+1}) \frac{\mathcal{X}_{l+1} - \bar{M}_l \cos(\theta_l)}{\sin(\theta_l)}, \quad (9)$$

where, $\theta_l = \arccos(\bar{M}_l^t \mathcal{X}_{l+1})$. In the next section, we discuss some experimental results on a synthetic data with outliers.

5.1. Experimental Results of Robust Regression

In this section, we give some preliminary regression results of our Fréchet median (FMd) based regressor to show its effectiveness on synthetic data corrupted with outliers. We compare the performance with our earlier formulation, i.e., using the incremental FM estimator. In our future work,

Outliers (%)	FMd		FM	
	b_M	r_M^2	b_M	r_M^2
0	5	0.280	8	0.349
10	5	0.265	5	0.222
15	5	0.251	5	0.092
20	5	0.262	3	−0.043
25	4	0.350	2	−0.170
30	5	0.377	1	−0.527
50	6	0.354	0	−7.151

Table 9: $SPD(3)$ to $\mathbf{S}^2$ regression with outliers

we will apply this method to real datasets. First, we generated 1000 i.i.d. samples on $SPD(3)$ from a Log-Normal distribution [31] with a variance of 0.25 and an expectation of I_3 (the identity matrix). Then, for each of these matrices, we compute the principal eigen vector which lies on $\mathbf{S}^2$. Then, we added random noise only to a fraction of the dependent variables in order to create outliers, and perform a regression from $SPD(3)$ to $\mathbf{S}^2$. The values of this fraction are varied during the experiments and given in the Table 9. The R^2 statistic results using both the FM (Eq. 6) and the FMd (Eq. 1) formulations are reported in Table 9. We use $N - 1$ cluster representatives, where N is the number of sample points, and perform leave-one-out analysis. We varied b from 1 to 20 in unit increments.

From the results, when there are no outliers, our FM formulation gives a better R^2 statistic, whereas with increasing outliers, the intrinsic FMd yields a far better performance. Moreover, for more than 15% outliers, the R^2 statistic for regression based on the FM formulation is negative. The negative value denotes that the regressor performed worse than the most trivial choice, which is FM of $\{y_i\}$ for any given x value. This result depicts the robustness of our regressor with the ℓ_1 norm and justifies its choice over the FM in this situation.

6. Conclusion

In this paper, we presented a novel nonlinear kernel-based regression technique for manifold-valued data sets. Our method is applicable to a variety of situations inclusive of regression between the manifold valued independent and dependent variables. We presented an extensive set of experiments on MR and diffusion MR scans from Alzheimers and movement disorder patients respectively. Further, we validated our results using the R^2-statistic and permutation tests. As an extension to our regression model, we presented a way to make the regression robust to outliers and showed its performance on synthetic data. Our future efforts will be focused on further developing and experimenting with the robust model.

References

[1] Harris Drucker, Chris JC Burges, et al., "Support vector regression machines," in *NIPS*, 1997, pp. 155–161. 1

[2] Masahisa Adachi and Kiki Hudson, *Embeddings and immersions*, American Mathematical Soc., 2012. 1

[3] Chafik Samir, P-A Absil, et al., "A gradient-descent method for curve fitting on Riemannian manifolds," *Foundations of Comput. Mathematics*, pp. 49–73, 2012. 2

[4] P Thomas Fletcher, "Geodesic regression and the theory of least squares on Riemannian manifolds," *International journal of computer vision*, pp. 171–185, 2013. 2, 4

[5] Yi Hong, Nikhil Singh, et al., "Time-warped geodesic regression," in *MICCAI*, pp. 105–112. 2014. 2

[6] Jia Du, Alvina Goh, et al., "Geodesic regression on orientation distribution functions with its application to an aging study," *Neuroimage*, pp. 416–426, 2014. 2

[7] Hyunwoo J Kim, Barbara B Bendlin, et al., "MGLM on Riemannian manifolds with applications to statistical analysis of diffusion weighted images," in *IEEE CVPR*, 2014, pp. 2705–2712. 2, 3

[8] Yi Hong, Roland Kwitt, et al., "Geodesic regression on the Grassmannian," in *ECCV*, pp. 632–646. 2014. 2

[9] Jacob Hinkle, Prasanna Muralidharan, et al., "Polynomial regression on Riemannian manifolds," in *ECCV*, pp. 1–14. 2012. 2

[10] Bradley C Davis, P Thomas Fletcher, et al., "Population shape regression from random design data," in *IEEE ICCV*, 2007, pp. 1–7. 2

[11] Nikhil Singh and Marc Niethammer, "Splines for diffeomorphic image regression," in *MICCAI*, pp. 121–129. 2014. 2

[12] Florian Steinke and Matthias Hein, "Non-parametric regression between manifolds," in *Advances in Neural Information Processing Systems*, 2009, pp. 1561–1568. 2

[13] S. Skwerer, *Tree Oriented Data Analysis*, Ph.D. thesis, http://arxiv.org/abs/1409.5501, 2014. 2

[14] Hyunwoo J Kim, Nagesh Adluru, et al., "Canonical Correlation analysis on Riemannian Manifolds and its Applications," in *ECCV*, pp. 251–267. 2014. 2, 7

[15] Monami Banerjee, Rudrasis Chakraborty, Edward Ofori, David Vaillancourt, and Baba C Vemuri, "Nonlinear regression on riemannian manifolds and its applications to neuro-image analysis," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015*, pp. 719–727. Springer International Publishing, 2015. 2, 5

[16] Manfredo Perdigao do Carmo Valero, *Riemannian geometry*, 1992. 2, 3

[17] Bijan Afsari, "Riemannian L^p center of mass: Existence, uniqueness, and convexity," *Proceedings of the American Mathematical Society*, vol. 139, no. 2, pp. 655–673, 2011. 3

[18] Maurice Fréchet, "Les éléments aléatoires de nature quelconque dans un espace distancié," *Annales de l'institut Henri Poincaré*, pp. 215–310, 1948. 3

[19] Jeffrey Ho, Guang Cheng, et al., "Recursive Karcher expectation estimators and geometric law of large numbers," in *AISTATS*, 2013, pp. 325–332. 3, 4

[20] K. T. Sturm, "Probability measures on metric spaces of nonpositive curvature," in *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces*, 2003. 3

[21] Hesamoddin Salehian, Rudrasis Chakraborty, Edward Ofori, David Vaillancourt, and Baba C Vemuri, "An efficient recursive estimator of the fréchet mean on a hypersphere with applications to medical image analysis," *MICCAI sponsored workshop MFCA*, 2015. 3, 4

[22] Rudrasis Chakraborty and Baba C Vemuri, "Recursive fréchet mean computation on Grassmannian and its applications to computer vision," *ICCV (accepted)*, 2015. 3, 4

[23] Jeffrey Ho, Yuchen Xie, et al., "On a nonlinear generalization of sparse coding and dictionary learning," in *ICML*, 2013, pp. 1480–1488. 3

[24] Daniel S Marcus, Tracy H Wang, et al., "OASIS: crosssectional mri data in young, middle aged, nondemented, and demented older adults," *Journal of cognitive neuroscience*, pp. 1498–1507, 2007. 5

[25] Yan Deng, Anand Rangarajan, et al., "A Riemannian framework for matching point clouds represented by the Schrodinger distance transform," in *IEEE CVPR*, 2014, pp. 3756–3761. 5

[26] David Kendall, "A survey of the statistical theory of shape," *Stat. Science*, pp. 87–99, 1989. 5

[27] Bing Jian and Baba C Vemuri, "Multi-fiber reconstruction from diffusion mri using mixture of wisharts and sparse deconvolution," in *Information Processing in Medical Imaging*. Springer, 2007, pp. 384–395. 6

[28] Guang Cheng, Baba C Vemuri, Paul R Carney, and Thomas H Mareci, "Non-rigid registration of high angular resolution diffusion images represented by gaussian mixture fields," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009*, pp. 190–197. Springer, 2009. 6

[29] Claudia Ramaker, Johan Marinus, Anne Margarethe Stiggelbout, and Bob Johannes van Hilten, "Systematic evaluation of rating scales for impairment and disability in parkinson's disease," *Movement Disorders*, vol. 17, no. 5, pp. 867–876, 2002. 6

[30] Silvere Bonnabel, "Stochastic gradient descent on riemannian manifolds," *Automatic Control, IEEE Transactions on*, vol. 58, no. 9, pp. 2217–2229, 2013. 8

[31] Armin Schwartzman, *Random ellipsoids and false discovery rates: Statistics for diffusion tensor imaging data*, Ph.D. thesis, Stanford University, 2006. 8