Intrinsic Regression Models for Manifold-Valued Data*

Xiaoyan Shi, **Martin Styner**, **Jeffrey Lieberman**, **Joseph G. Ibrahim**, **Weili Lin**, and **Hongtu Zhu**

Department of Biostatistics, Radiology, Psychiatry and Computer Science, and Biomedical Research Imaging Center, University of North Carolina at Chapel Hill Department of Psychiatry, Columbia University

Abstract

In medical imaging analysis and computer vision, there is a growing interest in analyzing various manifold-valued data including 3D rotations, planar shapes, oriented or directed directions, the Grassmann manifold, deformation field, symmetric positive definite (SPD) matrices and medial shape representations (m-rep) of subcortical structures. Particularly, the scientific interests of most population studies focus on establishing the associations between a set of covariates (e.g., diagnostic status, age, and gender) and manifold-valued data for characterizing brain structure and shape differences, thus requiring a regression modeling framework for manifold-valued data. The aim of this paper is to develop an intrinsic regression model for the analysis of manifold-valued data as responses in a Riemannian manifold and their association with a set of covariates, such as age and gender, in Euclidean space. Because manifold-valued data do not form a vector space, directly applying classical multivariate regression may be inadequate in establishing the relationship between manifold-valued data and covariates of interest, such as age and gender, in real applications. Our intrinsic regression model, which is a semiparametric model, uses a link function to map from the Euclidean space of covariates to the Riemannian manifold of manifold data. We develop an estimation procedure to calculate an intrinsic least square estimator and establish its limiting distribution. We develop score statistics to test linear hypotheses on unknown parameters. We apply our methods to the detection of the difference in the morphological changes of the left and right hippocampi between schizophrenia patients and healthy controls using medial shape description.

1 Introduction

Statistical analysis of manifold-valued data has gained a great deal of attention in neuroimaging applications [1], [2], [3], [4], [5], [6], [7], [8], [9]. Examples of manifold-valued data that we encounter in medical imaging analysis include the Grassmann manifold, planar shapes, deformation field, symmetric positive definite (SPD) matrices and the medial shape representations (m-rep) of subcortical structures (Fig. 1). Some review papers on the applications of manifold-valued data in medical imaging can be found in a recent special issue of **NeuroImage** [10]. However, the existing statistical methods for manifold-valued

data are primarily developed to estimate intrinsic and extrinsic means, to estimate the structure of population variability, to carry out principal geodesic analysis, and to compare intrinsic (or extrinsic) means across two or more groups [11], [9], [12], [13], [14], [15], [16], [17].

Little literature exists for regression analyses of manifold-valued data. The existing parametric and nonparametric regression models for manifold-valued data were primarily developed for 2 (or 3) dimensional directional data [6], [18]. In parametric regression of directional data, parametric distributions, such as the Von Mises distribution, are commonly assumed for directional data, whereas it can be very challenging to assume useful parametric distributions for other manifold-valued data, such as SPD matrices and the m-rep, which can characterize the feature (e.g., shape) of real imaging data [15]. In the nonparametric analysis of manifold-valued data, although smoothing splines have been developed for directional data and planar landmark data, it is computationally difficult to generalize such smoothing splines to other manifold-valued data [6]. Recently, local constant regressions have been developed for manifold-valued data, but these regression models are defined with respect to either the Frechet mean or the geometric median [2], [4].

According to the best of our knowledge, this is the first paper that develops a semiparametric regression model with manifold-valued data as responses on a Riemannian manifold and a set of covariates, such as time, gender, and diagnostic status, in Euclidean space. Our regression model are solely based on the first-order moment, thus avoiding specifying any parametric distributions. We propose an inference procedure to estimate the regression coefficients in this semi-parametric model. We establish asymptotic properties, including consistency and asymptotic normality, of the estimates of the regression coefficients. We develop score statistics to test linear hypotheses on unknown parameters. Finally, we illustrate the application of our statistical methods to the detection of the difference in morphological changes of the hippocampi between schizophrenia patients and healthy controls in a neuroimaging study of schizophrenia.

2 Method

2.1 Review of Regression Models

We consider a dataset that is composed of a response y_i and a $p \times 1$ covariate vector $\mathbf{x}_i$ for $i = 1, \cdots, n$. Responses may be continuous observations in classical linear models, such as age, weight, income, and they may be discrete or ordinal observations, such as differing severity of diseases and disease status (patients v.s. healthy subjects). Covariates may be quantitative, such as age, or qualitative, such as handiness, gender, and presence of risk factors (yes/no).

Regression models often include two key elements: a **link function** $\mu_i(\boldsymbol{\beta}) = g(\mathbf{x}_i, \boldsymbol{\beta})$ and a **residual** $\varepsilon_i = y_i - \mu_i(\boldsymbol{\beta})$, where $\boldsymbol{\beta}$ is a $q \times 1$ vector of regression coefficients and $g(\cdot, \cdot)$ is an known mapping from $R^p \times R^q$ to R. Regression models assume that

$$E[\varepsilon_i | \mathbf{x}_i] = 0 \text{ for all } i = 1, \cdots, n, \quad (1)$$

where the expectation denotes the conditional expectation of ε given $\mathbf{x}$. Moreover, nonparametric regressions include a **link function** $\mu_i = g(\mathbf{x}_i)$, in which $g(\cdot)$ is an unknown function, and a **residual** $\varepsilon_i = y_i - g(\mathbf{x}_i)$, for which equation (1) holds.

To carry out statistical inference on β (or $g(\cdot)$), we need at least three statistical methods. The first one is an estimation method for calculating parameter estimate of β, denoted by $\hat{\beta}$. Various estimation methods include maximum likelihood estimation, robust estimation, estimating equations, among many others. The second is to prove that $\hat{\beta}$ is a consistent estimator of β and has certain asymptotic distribution (e.g., normal). The third is to develop test statistics for testing the hypotheses:

$$H_0 : \mathbf{h}_0(\beta) = \mathbf{b}_0 \, \text{vs.} \, H_1 : \mathbf{h}_0(\beta) \neq \mathbf{b}_0, \quad (2)$$

where $\mathbf{h}_0(\cdot)$ is an $r \times 1$ vector function and $\mathbf{b}_0$ is an $r \times 1$ specified vector. In most applications, we are interested in testing $\mathbf{h}_0(\beta) = \mathbf{H}\beta = \mathbf{b}_0$ for a given $r \times q$ matrix $\mathbf{H}$ [18], [7], [8].

2.2 Intrinsic Regression for Manifold-Valued Data

We formally develop an intrinsic regression model for manifold-valued responses and covariates of interest from n observations. Suppose we observe a dataset $\{(S_i, \mathbf{x}_i) : i = 1, \cdots, n\}$, where S_i are points on a Riemannian manifold $\mathcal{S}$ and $\mathbf{x}_i$ are covariates of interest in Euclidean space.

The intrinsic regression model first involves modeling a 'conditional mean' of an manifold-valued response S_i given $\mathbf{x}_i$, denoted by $\mu_i(\beta) = g(\mathbf{x}_i, \beta)$, where $g(\cdot, \cdot)$, called **link function**, is a map from $R^p \times R^q$ to the manifold $\mathcal{S}$. Note that we just borrow the term 'conditional mean' from Euclidean space. Given two points $\mathbf{S}_i$ and $\mu_i(\beta)$ on the manifold $\mathcal{M}$, the intrinsic regression model also define the **residual** or 'difference' between $\mathbf{S}_i$ and $\mu_i(\beta)$ to ensure that $\mu_i(\beta)$ is the proper 'conditional mean' of $\mathbf{S}_i$ given $\mathbf{x}_i$. At $\mu_i(\beta)$, we have a tangent space of the manifold $\mathcal{S}$, denoted by $T_{\mu_i(\beta)}\mathcal{S}$, which is a Euclidean space representing a first order approximation of the manifold $\mathcal{S}$ near $\mu_i(\beta)$. Then, we calculate the projection of $\mathbf{S}_i$ onto $T_{\mu_i(\beta)}\mathcal{S}$, denoted by $\mathrm{Log}_{\mu_i(\beta)}(\mathbf{S}_i)$, which can be regarded as the difference between $\mathbf{S}_i$ and $\mu_i(\beta)$ for $i = 1, \cdots, n$. If $\mathcal{S}$ is a Euclidean space, then $\mathrm{Log}_{\mu_i(\beta)}(\mathbf{S}_i) = \mathbf{S}_i - \mu_i(\beta)$.

The intrinsic regression model for manifold-valued data is then defined by

$$E[\mathrm{Log}_{\mu_i(\beta)}(\mathbf{S}_i) | \mathbf{x}_i] = 0, \quad (3)$$

for $i = 1, \cdots, n$, where the expectation is taken with respect to the conditional distribution of $\mathbf{S}_i$ given $\mathbf{x}_i$. Model (3) does not assume any parametric distribution for $\mathbf{S}_i$ given $\mathbf{x}_i$, and thus it allows for a large class of distributions [15]. In addition, our model (3) does not assume homogeneous variance across all i. This is also desirable for the analysis of imaging measures, such as diffusion tensors, because between-subject and between-voxel variability in the imaging measures can be substantial.

2.3 Estimation

We calculate an intrinsic least squares estimator (ILSE) of the parameter vector $\boldsymbol{\beta}$, denoted by $\hat{\boldsymbol{\beta}}$, by minimizing the total residual sum of squares given by

$$G_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} d(\mathbf{S}_i, \mu_i(\boldsymbol{\beta}))^2 = \sum_{i=1}^{n} <<\mathrm{Log}_{\mu_i(\boldsymbol{\beta})}(\mathbf{S}_i), \mathrm{Log}_{\mu_i(\boldsymbol{\beta})}(\mathbf{S}_i)>>, \quad (4)$$

where $<< \cdot, \cdot >>$ is an inner product of two tangent vectors in $T_{\mu_i}(\boldsymbol{\beta})^{\mathcal{S}}$ and $d(\cdot, \cdot)$ is the Riemannian distance function on $\mathcal{S}$. Thus, let $G_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} d(\mathbf{S}_i, \mu_i(\boldsymbol{\beta}))^2$, $\hat{\beta}$, $\hat{\boldsymbol{\beta}}$ solves the estimating equations given by

$$\partial_{\beta} G_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} \partial_{\beta} d(\mathbf{S}_i, \mu_i(\boldsymbol{\beta}))^2 = \mathbf{0}, \quad (5)$$

where   denotes partial differentiation with respect to a parameter vector, such as $\boldsymbol{\beta}$. The ILSE is closely related to the intrinsic mean $\hat{\mu}_{IM}$ of $\mathbf{S}_1, \cdots, \mathbf{S}_n \in \mathcal{S}$, which is defined as

$$\hat{\mu}_{IM} = \mathrm{argmin}_{\mu} \sum_{i=1}^{n} d(\mu, \mathbf{S}_i)^2. \quad (6)$$

In this case, μ_i is independent of i and covariates of interest. Moreover, under some conditions, we can establish consistency and asymptotically normality of $\hat{\boldsymbol{\beta}}$. A Newton-Raphson algorithm is developed to obtain $\hat{\boldsymbol{\beta}}$. Let $\partial_{\beta} G_n(\boldsymbol{\beta})$ and $\partial_{\beta}^2 G_n(\beta)$, respectively, be the first- and second-order partial derivatives of $G_n(\boldsymbol{\beta})$. We iterates $\beta^{(t+1)} = \beta^{(t)} + \rho \{ -\partial_{\beta}^2 G_n(\beta^{(t)}) \}^{-1} \partial_{\beta} G_n(\beta^{(t)})$, where $0 < \rho = 1/2^{k_0}$ 1 for some k_0 0 is chosen such that $G_n(\boldsymbol{\beta}^{(t+1)})$ $G_n(\boldsymbol{\beta}^{(t)})$. We stop the Newton-Raphson algorithm when the absolute difference between consecutive $\boldsymbol{\beta}^{(t)}$'s is smaller than a predefined small number, say 10^{-4}. Finally, we set $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}^{(t)}$. In addition, because $-\partial_{\beta}^2 G_n(\beta^{(t)})$ may not be positive definite, we use $E[-\partial_{\beta}^2 G_n(\beta^{(t)})]$ instead of $-\partial_{\beta}^2 G_n(\beta^{(t)})$ in order to stabilize the Newton-Raphson algorithm.

2.4 Hypotheses and Test Statistics

In medical analysis, most scientific questions of interest involve a comparison of manifold-valued data across diagnostic groups or detecting change in manifold-valued data across time [8], [19]. These scientific questions usually can be formulated as follows:

$$H_0 : \mathbf{H}\boldsymbol{\beta} = \boldsymbol{b}_0 \, \mathrm{vs.} \, H_1 : \mathbf{H}\boldsymbol{\beta} \neq \boldsymbol{b}_0. \quad (7)$$

We test the null hypothesis $H_0 : \mathbf{H}\boldsymbol{\beta} = \mathbf{b}_0$ using a score test statistic W_n defined by

$$W_n = \mathbf{L}_n{}^T \hat{\mathbf{I}}^{-1} \mathbf{L}_n, \quad (8)$$

where $\mathbf{L}_n = n^{-1/2} \sum_{i=1}^n \hat{\mathbf{U}}_i(\tilde{\beta})$ and $\hat{\boldsymbol{I}} = n^{-1} \sum_{i=1}^n \hat{\mathbf{U}}_i(\tilde{\beta}) \hat{\mathbf{U}}_i(\tilde{\beta})^T$, in which $\tilde{\boldsymbol{\beta}}$ denotes the estimate of $\boldsymbol{\beta}$ under H_0 and $\hat{\mathbf{U}}_i(\tilde{\boldsymbol{\beta}})$ is associated with $_{\boldsymbol{\beta}} G_n(\boldsymbol{\beta})$. It can be shown that W_n is asymptotically χ^2 distributed.

2.5 Positive Definitive Matrices

We develop an intrinsic regression for SPDs. We introduce the tangent vector and tangent space at any $\mu \in \text{Sym}^+(m)$, the space of SPDs [8]. The tangent space of $\text{Sym}^+(m)$ at μ, denoted by $T_\mu\text{Sym}^+(m)$, is identified with a copy of $\text{Sym}(m)$, the space of symmetric matrices. Then we consider the scaled Frobenius inner product of any two tangent vectors Y_μ and Z_μ in $T_\mu\text{Sym}^+(m)$, which is defined by $<< Y_\mu, Z_\mu >> = \text{tr}(Y_\mu\mu^{-1}Z_\mu\mu^{-1})$. Given the inner product, we can formally construct the Riemannian geometry of $\text{Sym}^+(m)$ [8].

We consider the **link function** $\mu(\mathbf{x}, \boldsymbol{\beta})$ using the Cholesky decomposition of $\mu(\mathbf{x}, \boldsymbol{\beta})$. For the i–th observation, through a lower triangular matrix $\mathbf{C}_i(\boldsymbol{\beta}) = \mathbf{C}(\mathbf{x}_i, \boldsymbol{\beta}) = (C_{jk}(\mathbf{x}_i, \boldsymbol{\beta}))$, the Cholesky decomposition of $\mu(\mathbf{x}_i, \boldsymbol{\beta})$ equals $\mu(\mathbf{x}_i, \boldsymbol{\beta}) = \mu_i(\boldsymbol{\beta}) = \mathbf{C}_i(\boldsymbol{\beta})\mathbf{C}_i(\boldsymbol{\beta})^T$. We must specify the explicit forms of $C_{jk}(\mathbf{x}_i, \boldsymbol{\beta})$ for all $j \quad k$ in order to determine all entries in $\mu_i(\boldsymbol{\beta})$. As an illustration, for $m = 2$, we may choose the 2×2 matrix $\mathbf{C}_i(\boldsymbol{\beta})$ with $C_{11}(\mathbf{x}_i, \beta) = \exp(\mathbf{z}_i^T\beta_{(1)})$, $C_{12}(\mathbf{x}_i, \boldsymbol{\beta}) = 0$, $C_{21}(\mathbf{x}_i, \beta) = \mathbf{z}_i^T\beta_{(2)}$, and $C_{22}(\mathbf{x}_i, \beta) = \exp(\mathbf{z}_i^T\beta_{(3)})$, where $\mathbf{z}_i = (1, \mathbf{x}_i^T)^T$ and $\boldsymbol{\beta}_{(k)}$ for $k = 1, 2, 3$ are subvectors of $\boldsymbol{\beta}$. We introduce a definition of '**residuals**' to ensure that $\mu_i(\boldsymbol{\beta})$ is the proper 'conditionalmean' of S_i given $\mathbf{x}_i$. Then, we calculate the **residual** $\text{Log}_{\mu_i(\boldsymbol{\beta})}(S_i)$ given by $\mathbf{C}_i(\boldsymbol{\beta}) \log(\mathbf{C}_i(\boldsymbol{\beta})^{-1}S_i\mathbf{C}_i(\boldsymbol{\beta})^{-T})\mathbf{C}_i(\boldsymbol{\beta})^T$. The intrinsic regression is defined in (3).

The first- and second-order derivatives of $G_n(\boldsymbol{\beta})$ are given as follows. The a-th element of $_{\boldsymbol{\beta}} G_n(\boldsymbol{\beta})$ is given by $-2\sum_{i=1}^n \text{tr}\{\hat{\text{a}}\check{\text{D}}\check{\text{G}}_i(\beta)\mathbf{C}_i(\beta)^{-1}\partial_{\beta_a}\mu_i(\beta)\mathbf{C}_i(\beta)^{-T}\}$, where $\mathcal{E}_i(\boldsymbol{\beta}) = \log(\mathbf{C}_i(\boldsymbol{\beta})^{-1}S_i\mathbf{C}_i(\boldsymbol{\beta})^{-T})$ and $_{\beta_a} = / \beta_a$. The (a, b)-th element of $\partial_\beta^2 G_n(\beta)$ is given by

$$-2\sum_{i=1}^n \text{tr}\{\partial_{\beta_b}\hat{\text{a}}\check{\text{D}}\check{\text{G}}_i(\beta)[\mathbf{C}_i(\beta)^{-1}\partial_{\beta_a}\mathbf{C}_i(\beta) + \partial_{\beta_a}\mathbf{C}_i(\beta)^T\mathbf{C}_i(\beta)^{-T}]\} - 2\sum_{i=1}^n \text{tr}\{\hat{\text{a}}\check{\text{D}}\check{\text{G}}_i(\beta)\partial_{\beta_b}[\mathbf{C}_i(\beta)^{-1}\partial_{\beta_a}\mathbf{C}_i(\beta)$$
$$+\partial_{\beta_a}\mathbf{C}_i(\beta)^T\mathbf{C}_i(\beta)^{-T}]\}$$

where $\partial_{\beta_a\beta_b}^2 = \partial^2/\partial\beta_a\partial\beta_b$ and $\partial_{\beta_b}\hat{\text{a}}\check{\text{D}}\check{\text{G}}_i(\beta) = \int_0^1 \mathbf{h}(s, \beta)ds$, in which

$$\mathbf{h}(s, \beta) = \{[\tilde{\mathbf{S}}_i(\beta) - \mathbf{I}_3]s + \mathbf{I}_3\}^{-1}\partial_{\beta_b}\tilde{\mathbf{S}}_i(\beta)\{[\tilde{\mathbf{S}}_i(\beta) - \boldsymbol{I}_3]s + \boldsymbol{I}_3\}^{-1} \quad (9)$$

and $\tilde{\mathbf{S}}_i(\boldsymbol{\beta}) = \mathbf{C}_i(\boldsymbol{\beta})^{-1}S_i\mathbf{C}_i(\boldsymbol{\beta})^{-T}$.

2.6 Median Representation

We develop an intrinsic regression for m-reps. An m-rep model consisting of k medial atoms can be considered as the direct product of k copies of $M(1) = R^3 \times R^+ \times S(2) \times S(2)$, that is $M(k) = \prod_{i=1}^k M(1)$, where $S(2)$ is the sphere in R^3 with radius one [5]. We introduce a tangent space $T_PM(1)$ at the point $P = (O, r, \mathbf{n}_0, \mathbf{n}_1)$, where $O \in R^3$, $r \in R^+$, and $\mathbf{n}_0$ and $\mathbf{n}_1 \in S(2)$. The tangent vector $U \in T_PM(1)$ takes the form $U = (U_0, U_r, U_{\mathbf{n}_0}, U_{\mathbf{n}_1})$, where $U_0 \in R^3$, $U_r \in R$, $U_{\mathbf{n}_i} \in R^3$ and $U_{\mathbf{n}_i}^T\mathbf{n}_i = 0$ for $i = 0, 1$. The inner product of any two tangent vectors

$U^{(0)}$ and $U^{(1)}$ in $T_P M(1)$ is defined by $<< U^{(0)}, U^{(1)} >> = U^{(0)T} U^{(1)}$. The geodesic distance between P and $P_1 = (O_1, r_1, \mathbf{n}_{0,1}, \mathbf{n}_{1,1})$ in $M(1)$ is uniquely given by

$$\sqrt{(O - O_1)^T (O - O_1) + (\log(r) - \log(r_1))^2 + [\arccos(\mathbf{n}_0^T \mathbf{n}_{0,1})]^2 + [\arccos(\mathbf{n}_1^T \mathbf{n}_{1,1})]^2}.$$

To introduce an intrinsic regression for m-rep, we need to define a **link function** $\mu(\mathbf{x}, \boldsymbol{\beta}) = (\mu_O(\mathbf{x}, \boldsymbol{\beta}), \mu_r(\mathbf{x}, \boldsymbol{\beta}), \mu_0(\mathbf{x}, \boldsymbol{\beta}), \mu_1(\mathbf{x}, \boldsymbol{\beta}))^T \in M(1)$, which is a 10×1 vector. For instance, we may set $\mu_O(\mathbf{x}, \boldsymbol{\beta}) = (\mathbf{x}^T \boldsymbol{\beta}_1, \mathbf{x}^T \boldsymbol{\beta}_2, \mathbf{x}^T \boldsymbol{\beta}_3)^T$ and $\mu_r(\mathbf{x}, \boldsymbol{\beta}) = \exp(\mathbf{x}^T \boldsymbol{\beta}_4)$. A link function for $\mu_k(x, \boldsymbol{\beta}) = (\mu_{0k}^x(\boldsymbol{\beta}), \mu_{0k}^y(\boldsymbol{\beta}), \mu_{0k}^z(\boldsymbol{\beta}))(k = 0, 1)$ is based on the stereographic projection given by

$$\frac{\mu_{0k}^x}{1 - \mu_{0k}^z} = g_5(\mathbf{x}^T \boldsymbol{\beta}_{5,k}) \text{ and } \frac{\mu_{0k}^y}{1 - \mu_{0k}^z} = g_6(\mathbf{x}^T \boldsymbol{\beta}_{6,k}), \quad (10)$$

where $g_5(\cdot)$ and $g_6(\cdot)$ are known link functions and $\boldsymbol{\beta}_{5,k}$ and $\boldsymbol{\beta}_{6,k}$ are subvectors of $\boldsymbol{\beta}$. The **residual** $\mathrm{Log}_{\mu(\mathbf{x}, \boldsymbol{\beta})}(P)$ is given by

$$(O - \mu_O(\mathbf{x}, \boldsymbol{\beta}), \log(r/\mu_r(\mathbf{x}, \boldsymbol{\beta})), \mathrm{Log}_{\mu_0(\mathbf{x}, \beta)}(U_{\mathbf{n}_0}), \mathrm{Log}_{\mu_1(\mathbf{x}, \beta)}(U_{n_1})),$$

where $\mathrm{Log}_{\mu_0(\mathbf{x}, \boldsymbol{\beta})}(U_{\mathbf{n}_0}) = \arccos(\mu_0(\mathbf{x}, \boldsymbol{\beta})^T U_{\mathbf{n}_0}) \mathbf{v}/\|\mathbf{v}\|_2$, in which $\mathbf{v} = U_{\mathbf{n}_0} - (\mu_0(\mathbf{x}, \boldsymbol{\beta})^T U_{\mathbf{n}_0}) \mu_0(\mathbf{x}, \boldsymbol{\beta})$.

3 Results

To demonstrate our regression method, we applied our methods to the m-rep shape of the hippocampus structure in the left and right brain hemisphere in schizophrenia patients and healthy controls, collected at 14 academic medical centers in North America and western Europe [19]. There were 56 healthy controls and 238 schizophrenia patients who met the following criteria: age 16 to 40 years; onset of psychiatric symptoms before age 35; diagnosis of schizophrenia, schizophreniform, or schizoaffective disorder according to DSM-IV criteria; and various treatment and substance dependence conditions.

We investigated the difference of m-rep shape between schizophrenia patients and healthy controls while controlling for other factors, such as gender and age. The hippocampi m-rep shape at the 24 medial atoms of the left and right brain were used as the response in our intrinsic regression model. Covariates of interest include Whole Brain Volume (WBV), race (Caucasian, African American and others), age (in years), gender, and diagnostic status (patient or control).

We tested the diagnostic status effect on the whole m-rep structure. We presented the color-coded p-values of the diagnostic status effects across the atoms of both the left and right reference hippocampi in Fig 2 (a) and (b) and the corresponding adjusted p-values using false discovery rate were shown in Fig 2 (c) and (d). We observed large significance area in the left hippocampus, and some in the right hippocampus even after correcting for multiple comparisons.

4 Discussion

We have developed an intrinsic regression model for the analysis of manifold-valued data as responses in a Riemannian manifold and their association with a set of covariates. We have developed an estimation procedure to calculate the intrinsic least square estimates. We have developed score statistics for testing linear hypotheses on unknown parameters. We plan to apply our method to other manifold-valued data including the Grassmann manifold, planar shapes, and deformation field.

References

1. Pennec X. Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements. Journal of Mathematical Imaging and Vision. 2006; 25:127–154.

2. Davis, BC.; Bullitt, E.; Fletcher, PT.; Joshi, S. Population shape regression from random design data; IEEE 11th International Conference on Computer Vision; 2007.

3. Pennec X, Fillard P, Ayache N. A riemannian framework for tensor computing. International Journal of Computer Vision. 2006; 66:41–66.

4. Fletcher PT, Venkatasubramanian S, Joshi S. The geometric median on riemannian manifolds with application to robust atlas estimation. NeuroImage. 2009; 45:S143–S152. [PubMed: 19056498]

5. Yushkevich P, Fletcher P, Joshi S, Thall A, Pizer SM. Continuous medial representations for geometric object modeling in 2d and 3d. Image and Vision Computing. 2003; 21:17–28.

6. Jupp PE, Kent JT. Fitting smooth paths to spherical data. Applied Statistics. 1987; 36:34–46.

7. Schwartzman A, Robert FD, Taylor JE. Cross-subject comparison of principal diffusion direction maps. Magnetic Resonance in Medicine. 2005; 53:1423–1431. [PubMed: 15906307]

8. Schwartzman, A. Ph. D. thesis. Stanford University; 2006. Random ellipsoids and false discovery rates: Statistics for diffusion tensor imaging data.

9. Huckemann S, Hotz T, Munk A. Intrinsic shape analysis: Geodesic pca for riemannian manifolds modulo isometric lie group actions (with discussions). Statistica Sinica. 2009

10. Miller MI, Qiu A. The emerging discipline of computational functional anatomy. NeuroImage. 2009; 45:S16–S39. [PubMed: 19103297]

11. Fletcher P, Joshi S, Lu C, Pizer S. Principal geodesic analysis for the study of nonlinear statistics of shape. IEEE Transactions on Medical Imaging. 2004; 23:995–1005. [PubMed: 15338733]

12. Dryden, I.; Mardia, K. Statistical shape analysis. New York: John Wiley and Sons;

13. Dryden I, Mardia K. Multivariate shape analysis. Sankhya. 1993; 55:460–480.

14. Kendall DG. Shape manifolds, procrustean metrics and complex projective spaces. Bull. Lond. Math. Soc. 1984; 16:81–121.

15. Kent J. The complex bingham distribution and shape analysis. Journal of the Royal Statistical Society B. 1994; 56:285–299.

16. Kim PT, Koo JY. Statistical inverse problems on manifolds. Journal of Fourier Analysis and Applications. 2005; 11:639–653.

17. Fréchet M. Les éléments aléatoires de nature quelconque dans un espace distancé. Ann. Inst. H. Poincaré. 1948; 10:215–230.

18. Mardia, K.; Jupp, PE. Directional statistics. John Wiley: Academic Press; 1983.

19. Styner M, Lieberman JA, Pantazis D, Gerig G. Boundary and medial shape analysis of the hippocampus in schizophrenia. Medical Image Analysis. 2004; 4:197–203. [PubMed: 15450215]

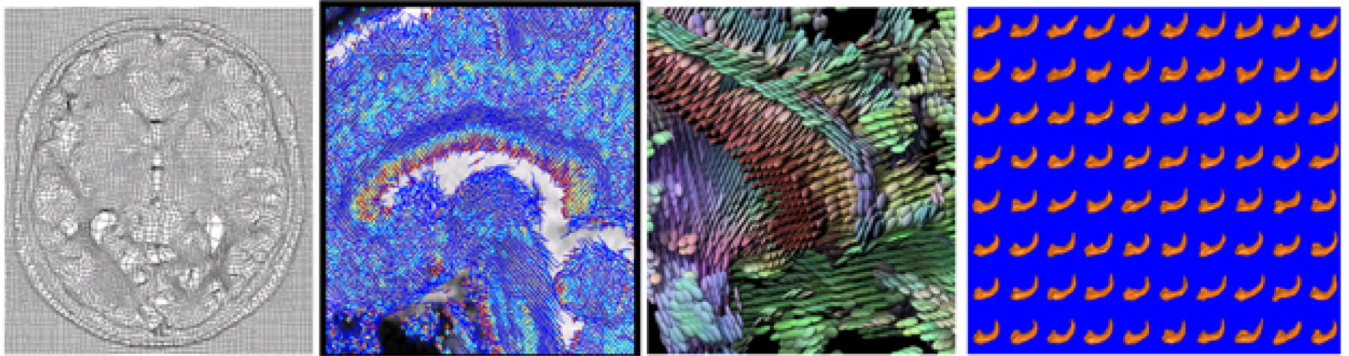**Fig. 1.**
Four different manifold-valued data (from the left to the right): deformation field reflecting brain deformation obtained from the registration of either diffusion tensor images (DTIs) or T1 magnetic resonance images (T1 MRIs); principal direction (PD) field reflecting fiber orientations obtained from DTIs; diffusion tensor field reflecting water diffusion along fiber tracts from DTIs; medial shape representations of hippocampi from multiple subjects obtained from the segmented T1 MRIs
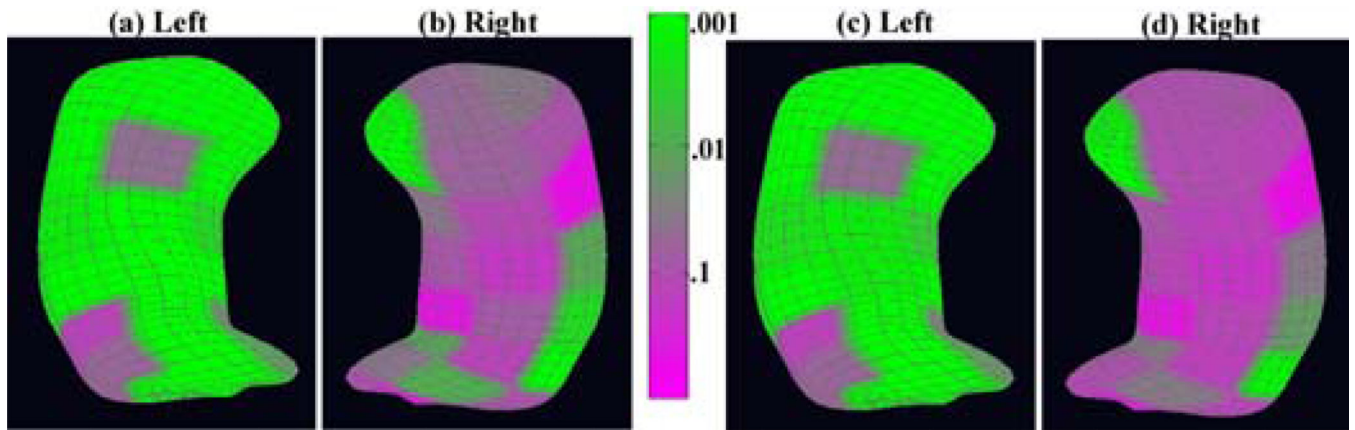
Fig. 2.
Results for the m-rep shape analysis result mapped to the surface of the hippocampal schizophrenia study: the color-coded uncorrected *p*–value maps of the diagnostic status effects for (a) the left hippocampus and (b) the right hippocampus; the corrected *p*–value maps for (c) the left hippocampus and (d) the right hippocampus after correcting for multiple comparisons