Intrinsic Regression Models for Positive-Definite Matrices With Applications to Diffusion Tensor Imaging

Hongtu Zhu, Yasheng Chen, Joseph G. Ibrahim, Yimei Li, Colin Hall, and Weili Lin

The aim of this paper is to develop an intrinsic regression model for the analysis of positive-definite matrices as responses in a Riemannian manifold and their association with a set of covariates, such as age and gender, in a Euclidean space. The primary motivation and application of the proposed methodology is in medical imaging. Because the set of positive-definite matrices do not form a vector space, directly applying classical multivariate regression may be inadequate in establishing the relationship between positive-definite matrices 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 positive-definite matrices. We develop an estimation procedure to calculate parameter estimates and establish their limiting distributions. We develop score statistics to test linear hypotheses on unknown parameters and develop a test procedure based on a resampling method to simultaneously assess the statistical significance of linear hypotheses across a large region of interest. Simulation studies are used to demonstrate the methodology and examine the finite sample performance of the test procedure for controlling the family-wise error rate. We apply our methods to the detection of statistical significance of diagnostic effects on the integrity of white matter in a diffusion tensor study of human immunodeficiency virus. Supplemental materials for this article are available online.

KEY WORDS: Diffusion tensor; Intrinsic regression; Positive-definite matrix; Riemannian manifold; Score statistic.

1. INTRODUCTION

Symmetric positive-definite matrices (SPD) arise in many applications including diffusion tensor imaging (DTI), computational anatomy, and statistics. For instance, DTI, which tracks the effective diffusion of water molecules in the human brain in vivo, contains a 3×3 positive-definite matrix $\mathbf{D}$, called a diffusion tensor, at each voxel (three-dimensional pixel) of an imaging space. Because water molecules tend to diffuse along the pathways of white matter fibers, tracking their diffusion with DTI allows investigators to map the microstructure and organization of those pathways. Specifically, the three eigenvalue-eigenvector pairs $\{(\lambda_i, \mathbf{v}_i) : i = 1, 2, 3\}$ of $\mathbf{D}$ $(\lambda_1 \geq \lambda_2 \geq \lambda_3)$ can quantify the degree and direction, respectively, of diffusivity and the principal directions $(\mathbf{v}_1)$ of the diffusion tensors (DTs) have been widely used to reconstruct the white matter fiber tracts (Basser, Mattiello, and LeBihan 1994a, 1994b; Zhu et al. 2007). Statistical analysis of DTI measures (e.g., eigenvalues and eigenvectors), diffusion tensors, and fiber tracts can reveal disruptions in structural organization and the neuronanatomical connectivity of the brains of persons with neurological and neuropsychiatric illnesses, and substance use disorders (Lim and Helpern 2002; Focke et al. 2008; Vernooij et al. 2008).

H. Zhu is Associate Professor of Biostatistics (E-mail: hzhu@bios.unc.edu), J. G. Ibrahim is Alumni Distinguished Professor of Biostatistics (E-mail: ibrahim@bios.unc.edu), and Y. Li is Ph.D. Student (E-mail: liyimei@email.unc.edu), Department of Biostatistics and Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7420. Y. Chen is Research Fellow (E-mail: yasheng.chen@med.unc.edu) and W. Lin is Professor of Radiology (E-mail: weili.lin@med.unc.edu), Department of Radiology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599. C. Hall is Professor of Neurology, Department of Neurology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 (E-mail: hallc@neurology.unc.edu). This work was supported in part by NSF grants SES-06-43663 and BCS-08-26844 and NIH grants UL1-RR025747-01, R01MH086633, and R21 AG033387 to Dr. Zhu, NIH grants GM 70335 and CA 74015 to Dr. Ibrahim and NIH grants R01NS055754 to Dr. Lin. We thank the Editor, an associate editor, and two referees for helpful suggestions, which have improved substantially the present form of this article.

In computational anatomy, a symmetric positive-definite deformation tensor $\mathbf{S} = (\mathbf{J}\mathbf{J}^T)^{1/2}$ at each location of the image can capture the directional information of shape change decoded in the Jacobian matrices $\mathbf{J}$ (Grenander and Miller 1998; Lepore et al. 2008). The directional shape change information decoded in deformation tensors has been demonstrated to be more discriminative than the determinant of $\mathbf{J}$ in several structural MRI based studies of HIV/AIDS (Lepore et al. 2008). More generally, deformation tensors can be used to detect subtle changes of brain structures across groups and along time. An appropriate statistical analysis of deformation tensors is important for understanding normal brain development, the neural bases of neuropsychiatric disorders, and the joint effects of environmental and genetic factors on brain structure and function. A fundamental issue in many applications is modeling covariance matrices for multivariate data, longitudinal data, spatial data, and time series data (Chiu, Leonard, and Tsui 1996; Pourahmadi 1999, 2000; Anderson 2003).

Despite the extensive use of SPD matrices, a formal statistical framework for using a set of covariates in a Euclidean space to predict SPD matrices as responses has not yet been developed. One may ignore the fact that SPD matrices are in a nonlinear space and then directly apply classical multivariate regression to studying the association between SPD matrices and covariates. However, this method is not adequate in many applications. As an illustration, we consider the "mean" of SPD matrices. In Euclidean space, the mean of SPD matrices is just the empirical average of SPD matrices, which can lead to a swelling effect in imaging processing (Chefdhotel et al. 2004). In contrast, if one regards SPD matrices as points in Riemannian manifold and calculates their Frechét mean (Frechét 1948; Karcher 1977), then the swelling effect completely disappears (Pennec, Fillard, and Ayache 2006). Moreover, even when SPD matrices are regarded as random elements in Riemannian

manifold, the existing statistical methods can only compare differences between the means of the two (or multiple) groups of SPD matrices, which is equivalent to modeling SPD matrices with a single categorical variable (Bhattacharya and Patrangenaru 2003, 2005; Schwartzman 2006; Whitcher et al. 2007). For instance, Schwartzman (2006) has developed several parametric models for SPD matrices and derived distributions of several test statistics for testing differences in two groups of SPD matrices. However, the use of parametric models to characterize SPD matrices in real applications requires further research. Recently, Kim and Richards (2008) have studied the problem of deconvolution density estimation on the space of SPD matrices. Although much effort has been devoted to appropriately modeling covariance matrices in longitudinal data and time series settings (Pourahmadi 1999, 2000), these covariance matrices are not response variables as is the case in medical imaging.

To the best of our knowledge, this is the very first paper that develops a semiparametric regression model with SPD matrices as responses in a Remannian manifold and covariates in Euclidean space. Compared with the existing literature, we extend the two-group model studied by Schwartzman (2006) to allow for general covariates and generalize the normal distributional modeling in Schwartzman (2006) by assuming only that certain appropriately defined residuals have mean zero. Our regression model is semiparametric, and thus it avoids specifying parametric distributions for random SPD matrices. We propose an inference procedure to estimate the regression coefficients in this semiparametric 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 develop a test procedure based on a resampling method to simultaneously assess the statistical significance of linear hypotheses across a large region of interest.

The rest of this paper is organized as follows. We will first review some basic results about the Riemannian geometrical structure for the space of SPD matrices. We will next present the semiparametric regression model and estimation methods for estimating the regression coefficients. Then we will develop score statistics to carry out hypothesis testing and develop a test procedure for correcting for multiple comparisons. Simulation studies will assess the empirical performance of the estimation algorithms and test statistics. Finally, we will analyze a neuroimaging dataset to illustrate an application of these methods, before offering some concluding remarks.

2. THEORY

In this section, we formally introduce a semiparametric regression model for a SPD matrix response, develop an estimation algorithm for estimating regression coefficients, and develop a score statistic for testing linear hypotheses of the regression coefficients. We then present a test procedure based on a resampling method for correcting multiple statistical comparisons.

2.1 Model Formulation

We develop a formal regression model for SPD matrices. Suppose we observe a $m \times m$ SPD matrix $\mathbf{S}_i$ and a $k \times 1$ covariate vector $\mathbf{x}_i$, such as diagnostic status and age, for $i = 1, \ldots, n$.

Let $\mathrm{Sym}^+(m)$ be the set of $m \times m$ SPD matrices, $\boldsymbol{\beta}$ be a $p \times 1$ vector of regression coefficients in R^p and $\Sigma(\cdot, \cdot)$ be a map from $R^k \times R^p$ to $\mathrm{Sym}^+(m)$. The regression model is obtained by modeling the "conditional mean" of $\mathbf{S}_i$ given $\mathbf{x}_i$, denoted by $\Sigma_i(\boldsymbol{\beta}) = \Sigma(\mathbf{x}_i, \boldsymbol{\beta}) \in \mathrm{Sym}^+(m)$. Note that we just borrow the term "conditional mean" from Euclidean space, but we will formalize the notion of a "conditional mean" explicitly below.

Following Pourahmadi (2000), we consider a general specification of $\Sigma(\mathbf{x}, \boldsymbol{\beta})$ using the Cholesky decomposition of $\Sigma(\mathbf{x}, \boldsymbol{\beta})$. For the ith observation, through a lower triangular matrix $\mathbf{C}_i(\boldsymbol{\beta}) = \mathbf{C}(\mathbf{x}_i, \boldsymbol{\beta}) = (c_{jk}(\mathbf{x}_i, \boldsymbol{\beta}))$ (Pourahmadi 2000), the Cholesky decomposition of $\Sigma(\mathbf{x}_i, \boldsymbol{\beta})$ is given by

$$\Sigma(\mathbf{x}_i, \boldsymbol{\beta}) = \Sigma_i(\boldsymbol{\beta}) = \mathbf{C}_i(\boldsymbol{\beta})\mathbf{C}_i(\boldsymbol{\beta})^T. \tag{1}$$

The specification in (1) has been widely used due to computational simplicity (Pourahmadi 1999, 2000). Now we need to specify the explicit forms of $c_{jk}(\mathbf{x}_i, \boldsymbol{\beta})$ for all $j \geq k$ in order to determine all entries in $\Sigma_i(\boldsymbol{\beta})$. These functions $c_{jk}(\mathbf{x}_i, \boldsymbol{\beta})$ in $\mathbf{C}_i(\boldsymbol{\beta})$ can be regarded as link functions in generalized linear models (McCullagh and Nelder 1989). For instance, for $m = 2$, we may choose

$$\mathbf{C}_i(\boldsymbol{\beta}) = \begin{pmatrix} \mathbf{z}_i^T \boldsymbol{\beta}_{(1)} & 0 \\ \mathbf{z}_i^T \boldsymbol{\beta}_{(2)} & \mathbf{z}_i^T \boldsymbol{\beta}_{(3)} \end{pmatrix}, \tag{2}$$

where $\mathbf{z}_i = (1, \mathbf{x}_i^T)^T$ and $\boldsymbol{\beta}_{(j)}$ for $j = 1, 2, 3$ are subvectors of $\boldsymbol{\beta}$. To ensure the uniqueness of the Cholesky decomposition, we may impose additional constraints that $c_{jj}(\mathbf{x}_i, \boldsymbol{\beta}) > 0$ for all $j = 1, \ldots, m$. For instance, we may assume that $\mathbf{C}_i(\boldsymbol{\beta})$ takes the form

$$\begin{pmatrix} (\mathbf{z}_i^T \boldsymbol{\beta}_{(1)})^2 & 0 \\ \mathbf{z}_i^T \boldsymbol{\beta}_{(2)} & (\mathbf{z}_i^T \boldsymbol{\beta}_{(3)})^2 \end{pmatrix} \quad \text{or} \\ \begin{pmatrix} \exp(\mathbf{z}_i^T \boldsymbol{\beta}_{(1)}) & 0 \\ \mathbf{z}_i^T \boldsymbol{\beta}_{(2)} & \exp(\mathbf{z}_i^T \boldsymbol{\beta}_{(3)}) \end{pmatrix}. \tag{3}$$

Compared with (3), the link function in (2) leads to a simpler computation. However, there is an ambiguous sign problem associated with (2), since $\Sigma_i(\boldsymbol{\beta}) = \mathbf{C}_i(\boldsymbol{\beta})\mathbf{P}\mathbf{P}^T\mathbf{C}_i(\boldsymbol{\beta})^T$, where $\mathbf{P} = \mathrm{diag}(\pm 1, \ldots, \pm 1)$. In practice, we can standardize all covariates and impose additional sign constraints on the intercept terms of $\boldsymbol{\beta}_{(1)}$ and $\boldsymbol{\beta}_{(3)}$ to remove this ambiguity. The link functions in (2) and (3) can be easily generalized to $m \geq 3$. When $\mathbf{x}_i = \mathbf{0}$, $\Sigma(\mathbf{0}, \boldsymbol{\beta}) = \mathbf{C}(\mathbf{0}, \boldsymbol{\beta})\mathbf{C}(\mathbf{0}, \boldsymbol{\beta})^T$ represents an intercept matrix, which is similar to the intercept in a linear model.

We may consider two direct reparametrizations of $\Sigma_i(\boldsymbol{\beta})$. The first is a matrix-logarithmic model. If we consider the spectral decomposition of $\Sigma \in \mathrm{Sym}^+(m)$ given by $\mathbf{O}\mathbf{V}\mathbf{O}^T$, where $\mathbf{O}$ is a $m \times m$ orthonormal matrix and $\mathbf{V}$ is a $m \times m$ diagonal matrix with positive entries, then the *matrix logarithm* of Σ is $\log(\Sigma) = \log(\mathbf{O}\mathbf{V}\mathbf{O}^T) = \mathbf{O}\log(\mathbf{V})\mathbf{O}^T$, where $\log(\mathbf{V}) \in \mathrm{diag}(m)$ with diagonal elements equal the logarithm of the diagonal entries of $\mathbf{V}$. Let $\mathrm{vecs}(\mathbf{U})$ represent $(U_{11}, \ldots, U_{1m}, U_{22}, \ldots, U_{2m}, \ldots, U_{mm})^T$ for any $m \times m$ symmetric matrix $\mathbf{U} = (U_{ij})$ and let $\mathbf{l}(\mathbf{x}, \boldsymbol{\beta})$ be a map from $R^k \times R^p$ to $R^{m(m+1)/2}$. The matrix-logarithmic model assumes that the matrix logarithm of $\Sigma_i(\boldsymbol{\beta})$ can be written as

$$\mathrm{vecs}\big[\log(\Sigma_i(\boldsymbol{\beta}))\big] = \mathbf{l}(\mathbf{x}_i, \boldsymbol{\beta}). \tag{4}$$

A graphical illustration is given in Figure 1. For instance, let $\mathbf{B}$ be an $m(m + 1)/2 \times (k + 1)$ matrix. If we set $\mathbf{l}(\mathbf{x}_i, \boldsymbol{\beta}) = \mathbf{B}\mathbf{z}_i$
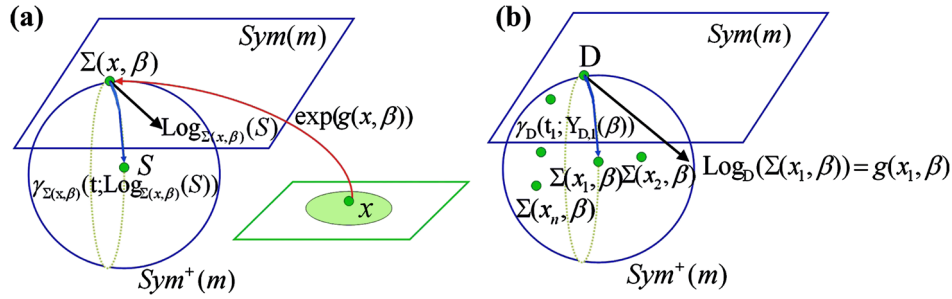
Figure 1. Graphical illustration of two link functions in the (a) matrix-logarithmic and (b) geodesic models. In (a), the link function for the matrix-logarithmic model equals $\Sigma(\mathbf{x}, \boldsymbol{\beta}) = \exp(\mathbf{g}(\mathbf{x}, \boldsymbol{\beta}))$ or $\log(\Sigma(\mathbf{x}, \boldsymbol{\beta})) = \mathbf{g}(\mathbf{x}, \boldsymbol{\beta})$. In (b), the link function for the geodesic model equals $\mathrm{Log}_{\mathbf{D}}(\Sigma(\mathbf{x}, \boldsymbol{\beta})) = \mathbf{g}(\mathbf{x}, \boldsymbol{\beta})$, where $\mathbf{D}$ is the intercept matrix.

and $\boldsymbol{\beta}$ contains all unknown parameters in $\mathbf{B}$, then we obtain the matrix-logarithmic model for the covariance matrix in Chiu, Leonard, and Tsui (1996). For $m = 1$, we have $\mathbf{l}(\mathbf{x}_i, \boldsymbol{\beta}) = (\mathbf{z}_i^T \boldsymbol{\beta}_{(1)})$ and $\Sigma_i(\boldsymbol{\beta}) = \exp(\mathbf{z}_i^T \boldsymbol{\beta}_{(1)}) = \exp(\beta_{1,1} + \mathbf{x}_i^T \boldsymbol{\beta}_{1,(2)})$, where $\boldsymbol{\beta}_{(1)} = (\beta_{1,1}, \boldsymbol{\beta}_{1,(2)}^T)^T$. Thus, $\exp(\beta_{1,1})$ can be represented as an intercept when $\mathbf{x}_i = \mathbf{0}$, whereas $\boldsymbol{\beta}_{1,(2)}$ characterizes the "slope" vector.

The second is a "geodesic" model for $\Sigma(\mathbf{x}_i, \boldsymbol{\beta})$. Intuitively, a geodesic on the manifold $\mathrm{Sym}^+(m)$ is a "straight line" on $\mathrm{Sym}^+(m)$. Let $\mathrm{GL}(m)$ be the set of $m \times m$ matrices of nonzero determinant. For any $\mathbf{C} \in \mathrm{GL}(m)$, we define the *matrix exponential transformation* of $\mathbf{C}$, denoted by $\exp(\mathbf{C})$, as $\exp(\mathbf{C}) = \sum_{k=0}^{\infty} \mathbf{C}^k / k$, where $\mathbf{C}^0 = \mathbf{I}_m$ is the $m \times m$ identity matrix and $\mathbf{C}^k$ is the ordinary matrix multiplication of $\mathbf{C}$ k times. We consider an intercept matrix $\mathbf{D} \in \mathrm{Sym}^+(m)$ and any square root of $\mathbf{D}$, denoted by $\mathbf{B}$, such that $\mathbf{B} \in \mathrm{GL}(m)$ and $\Sigma(\mathbf{0}, \boldsymbol{\beta}) = \mathbf{D} = \mathbf{B}\mathbf{B}^T$, when $\mathbf{x}_i = \mathbf{0}$. Then, for a given $\mathbf{l}(\mathbf{x}_i, \boldsymbol{\beta})$, we consider a "directional" matrix $\mathbf{Y}_D(\mathbf{x}_i, \boldsymbol{\beta}) = \mathbf{Y}_{D,i}(\boldsymbol{\beta})$ in the space of $m \times m$ symmetric matrices, denoted by $\mathrm{Sym}(m)$, such that $\mathrm{vecs}(\mathbf{Y}_{D,i}(\boldsymbol{\beta})) = \mathbf{l}(\mathbf{x}_i, \boldsymbol{\beta})$ and $\mathbf{l}(\mathbf{0}, \boldsymbol{\beta}) = \mathbf{0}$. The geodesic model assumes that

$$\Sigma(\mathbf{x}_i, \boldsymbol{\beta}) = \mathbf{C}_i(\boldsymbol{\beta})\mathbf{C}_i(\boldsymbol{\beta})^T = \mathbf{B}\exp(\mathbf{B}^{-1}\mathbf{Y}_{D,i}(\boldsymbol{\beta})\mathbf{B}^{-T})\mathbf{B}^T \quad (5)$$

for $i = 1, \ldots, n$. As shown in proposition 2.2.7 of Schwartzman (2006), $\Sigma(\mathbf{x}_i, \boldsymbol{\beta})$ is closely related to the geodesic passing through $\mathbf{D} = \mathbf{B}\mathbf{B}^T$ in the direction of $\mathbf{Y}_{D,i}(\boldsymbol{\beta})$ and the specification of $\Sigma(\mathbf{x}_i, \boldsymbol{\beta})$ in (5) is unique. We will formalize the notions of "geodesic" and "direction" in Section 2.2. A graphical illustration is given in Figure 1(b). Because the specifications (4) and (5) involve the matrix exponential transformation, they can be much more computationally difficult than specifications (2) and (3).

We introduce a definition of "residual" to ensure that $\Sigma_i(\boldsymbol{\beta})$ is the proper "conditional mean" of $\mathbf{S}_i$ given $\mathbf{x}_i$. For instance, in the classical linear model, the response is the sum of a regression function and a residual term. Then, the regression function is the conditional mean of the response only when the conditional mean of the residual equals zero. Given two points $\mathbf{S}_i$ and $\Sigma_i(\boldsymbol{\beta})$ on the manifold $\mathrm{Sym}^+(m)$, we need to define the residual or "difference" between $\mathbf{S}_i$ and $\Sigma_i(\boldsymbol{\beta})$. At $\Sigma_i(\boldsymbol{\beta})$, we have a tangent space of the manifold $\mathrm{Sym}^+(m)$, denoted by $T_{\Sigma_i(\boldsymbol{\beta})} \mathrm{Sym}^+(m)$, which is a Euclidean space representing a first-order approximation of the manifold $\mathrm{Sym}^+(m)$

near $\Sigma_i(\boldsymbol{\beta})$. Then, we calculate the projection of $\mathbf{S}_i$ onto $T_{\Sigma_i(\boldsymbol{\beta})} \mathrm{Sym}^+(m)$, denoted by $\mathrm{Log}_{\Sigma_i(\boldsymbol{\beta})}(\mathbf{S}_i)$, which is given by

$$\mathrm{Log}_{\Sigma_i(\boldsymbol{\beta})}(\mathbf{S}_i) = \mathbf{C}_i(\boldsymbol{\beta}) \log(\mathbf{C}_i(\boldsymbol{\beta})^{-1}\mathbf{S}_i\mathbf{C}_i(\boldsymbol{\beta})^{-T})\mathbf{C}_i(\boldsymbol{\beta})^T. \quad (6)$$

Thus, $\mathrm{Log}_{\Sigma_i(\boldsymbol{\beta})}(\mathbf{S}_i)$ on $T_{\Sigma_i(\boldsymbol{\beta})} \mathrm{Sym}^+(m)$ can be regarded as the difference between $\mathbf{S}_i$ and $\Sigma_i(\boldsymbol{\beta})$ for $i = 1, \ldots, n$. Since $\mathrm{Log}_{\Sigma_i(\boldsymbol{\beta})}(\mathbf{S}_i)$ for different i are in different tangent spaces (or Euclidean spaces), we must translate them back to the same tangent space, such as the tangent space at the identity matrix $\mathbf{I}_m$, denoted by $T_{\mathbf{I}_m} \mathrm{Sym}^+(m)$ (Schwartzman 2006). This translation can be done using the group action of $\mathrm{GL}(m)$ introduced in Section 2.2. Specifically, to translate $\mathrm{Log}_{\Sigma_i(\boldsymbol{\beta})}(\mathbf{S}_i) \in T_{\Sigma_i(\boldsymbol{\beta})} \mathrm{Sym}^+(m)$ into $T_{\mathbf{I}_m} \mathrm{Sym}^+(m)$, we follow Schwartzman (2006) to define the *residual* of $\mathbf{S}_i$ with respect to $\Sigma_i(\boldsymbol{\beta})$ as

$$\mathcal{E}_i(\boldsymbol{\beta}) = \mathcal{E}_{\Sigma_i(\boldsymbol{\beta})}(\mathbf{S}_i) = \mathbf{C}_i(\boldsymbol{\beta})^{-1} \mathrm{Log}_{\Sigma_i(\boldsymbol{\beta})}(\mathbf{S}_i)\mathbf{C}_i(\boldsymbol{\beta})^{-T}$$

$$= \log(\mathbf{C}_i(\boldsymbol{\beta})^{-1}\mathbf{S}_i\mathbf{C}_i(\boldsymbol{\beta})^{-T}), \quad (7)$$

for $i = 1, \ldots, n$. The $\mathcal{E}_i(\boldsymbol{\beta})$ are uniquely defined, because the Cholesky decomposition is unique and $\mathrm{Log}_{\Sigma_i(\boldsymbol{\beta})}(\mathbf{S}_i)$ is independent of the choice of the square root of $\Sigma_i(\boldsymbol{\beta})$. Moreover, because all $\mathcal{E}_i(\boldsymbol{\beta})$ are in the same tangent space $T_{\mathbf{I}_m} \mathrm{Sym}^+(m)$ and $T_{\mathbf{I}_m} \mathrm{Sym}^+(m)$ is a Euclidean space, we can apply classical multivariate analysis techniques in Euclidean space to the analysis of $\mathcal{E}_{\Sigma_i(\boldsymbol{\beta})}(\mathbf{S}_i)$ (Anderson 2003).

The intrinsic regression model for SPD matrices is then defined by

$$E[\mathcal{E}_i(\boldsymbol{\beta})|\mathbf{x}_i] = \mathbf{0} \quad \text{or} \quad E[\mathrm{Log}_{\Sigma_i(\boldsymbol{\beta})}(\mathbf{S}_i)|\mathbf{x}_i] = \mathbf{0}, \quad (8)$$

for $i = 1, \ldots, n$, where the expectation is taken with respect to the conditional distribution of $\mathbf{S}_i$ given $\mathbf{x}_i$. Model (8) does not assume any parametric distribution for $\mathbf{S}_i$ given $\mathbf{x}_i$, and thus it allows for a large class of distributions. In contrast, let Ω be an $m \times m$ matrix in $\mathrm{Sym}^+(m)$, Schwartzman (2006, p. 42) considered a symmetric matrix variate normal distribution of $\mathcal{E}_i$ (Gupta and Nagar 2000), whose density is given by

$$p(\mathcal{E}_i; \mathbf{0}, \Omega) = \frac{1}{(2\pi)^m |\Omega|^2} \exp\left(-\frac{1}{2} \mathrm{tr}(\mathcal{E}_i\Omega^{-1})^2\right). \quad (9)$$

Then, using $\mathbf{S}_i = \mathbf{C}_i(\boldsymbol{\beta}) \exp(\mathcal{E}_i)\mathbf{C}_i(\boldsymbol{\beta})^T$, the density of $\mathbf{S}_i$, denoted by $p(\mathbf{S}_i; \boldsymbol{\beta})$, is

$$\frac{J(\mathbf{C}_i(\boldsymbol{\beta})^{-1}\mathbf{S}_i\mathbf{C}_i(\boldsymbol{\beta})^{-T})|\mathbf{C}_i(\boldsymbol{\beta})\mathbf{C}_i(\boldsymbol{\beta})^T|^2}{(2\pi)^{2m}|\Omega|^2}$$

$$\times \exp\left(-\frac{1}{2} \mathrm{tr}(\log(\mathbf{C}_i(\boldsymbol{\beta})^{-1}\mathbf{S}_i\mathbf{C}_i(\boldsymbol{\beta})^{-T})\Omega^{-1})^2\right), \quad (10)$$

where $J(\mathbf{C}_i(\boldsymbol{\beta})^{-1}\mathbf{S}_i\mathbf{C}_i(\boldsymbol{\beta})^{-T})$ is the Jacobian of the transformation $\mathcal{E}_i = \mathbf{C}_i(\boldsymbol{\beta})^{-1}\mathbf{S}_i\mathbf{C}_i(\boldsymbol{\beta})^{-T}$. Because the true distribution of $\mathcal{E}_i$ may deviate from the Gaussian distribution in (9), it can be very restrictive to assume a parametric distribution, such as (10), for $\mathbf{S}_i$. In addition, our model (8) does not assume homogeneous variance across all i. This is also desirable for applications, such as the analysis of imaging data, including diffusion tensor data, because between-subject and between-voxel variability in the imaging measures (e.g., DT) can be substantial.

2.2 Geometrical Structure of $\mathrm{Sym}^+(m)$

We summarize some basic results from Schwartzman (2006) about the geometrical structure of $\mathrm{Sym}^+(m)$ as a Riemannian manifold (Do Carmo 1992; Lang 1999). The space $\mathrm{Sym}^+(m)$ is a submanifold of the Euclidean space $\mathrm{Sym}(m)$. Geometrically, the spaces $\mathrm{Sym}^+(m)$ and $\mathrm{Sym}(m)$ are differentiable manifolds of $m(m+1)/2$ dimensions and they are homeomorphically related by the matrix exponential transformation and logarithm.

We first introduce the tangent vector and tangent space at any $\mathbf{D} \in \mathrm{Sym}^+(m)$. For a small scalar $\delta > 0$, let $\mathbf{C}(t)$ be a differentiable map from $(-\delta, \delta)$ to $\mathrm{Sym}^+(m)$ such that it passes through $\mathbf{C}(0) = \mathbf{D}$. The tangent vector at $\mathbf{D}$ is defined as the derivative of the smooth curve $\mathbf{C}(t)$ with respect to t at $t = 0$. The set of all tangent vectors at $\mathbf{D}$ forms the tangent space of $\mathrm{Sym}^+(m)$ at $\mathbf{D}$, denoted by $T_{\mathbf{D}}\,\mathrm{Sym}^+(m)$. As shown in proposition 2.2.3 of Schwartzman (2006), $T_{\mathbf{D}}\,\mathrm{Sym}^+(m)$ is identified with a copy of $\mathrm{Sym}(m)$. For instance, when $\mathbf{D} = \mathbf{I}_m$, we consider the curve $\mathbf{C}(t) = \exp(t\mathbf{Y}) \in \mathrm{Sym}^+(m)$ satisfying $\mathbf{C}(0) = \mathbf{I}_m$ for any $\mathbf{Y} \in \mathrm{Sym}(m)$ and $t \in (-\delta, \delta)$. Then, the derivative of $\mathbf{C}(t)$ at $t = 0$ is just $\mathbf{Y}$ and thus $T_{\mathbf{I}_m}\,\mathrm{Sym}^+(m) = \mathrm{Sym}(m)$. Second, we introduce an inner product of any two tangent vectors in the same tangent space, which varies smoothly along the manifold. We consider the *scaled Frobenius inner product* of any two tangent vectors $\mathbf{Y}_{\mathbf{D}}$ and $\mathbf{Z}_{\mathbf{D}}$ in $T_{\mathbf{D}}\,\mathrm{Sym}^+(m)$, which is defined by

$$\langle\!\langle \mathbf{Y}_{\mathbf{D}}, \mathbf{Z}_{\mathbf{D}} \rangle\!\rangle = \mathrm{tr}(\mathbf{Y}_{\mathbf{D}}\mathbf{D}^{-1}\mathbf{Z}_{\mathbf{D}}\mathbf{D}^{-1}). \tag{11}$$

Let $\mathbf{B} \in \mathrm{GL}(m)$ be any square root of $\mathbf{D} \in \mathrm{Sym}^+(m)$ such that $\mathbf{D} = \mathbf{B}\mathbf{B}^T$ and let $\gamma_{\mathbf{D}}(t; \mathbf{Y}_{\mathbf{D}})$ be the geodesic on $\mathrm{Sym}^+(m)$ passing through $\mathbf{D}$ in the direction of the tangent vector $\mathbf{Y}_{\mathbf{D}} \in T_{\mathbf{D}}\,\mathrm{Sym}^+(m)$. As shown in proposition 2.2.7 of Schwartzman (2006), $\gamma_{\mathbf{D}}(t; \mathbf{Y}_{\mathbf{D}})$ is uniquely given by

$$\gamma_{\mathbf{D}}(t; \mathbf{Y}_{\mathbf{D}}) = \mathbf{B}\exp(t\mathbf{B}^{-1}\mathbf{Y}_{\mathbf{D}}\mathbf{B}^{-T})\mathbf{B}^T \quad \text{for all } t \in (-\infty, \infty). \tag{12}$$

Given the geodesic $\gamma_{\mathbf{D}}(t; \mathbf{Y}_{\mathbf{D}})$, the Riemannian exponential map $\mathrm{Exp}_D(\mathbf{Y}_{\mathbf{D}})$, which maps the tangent vector $\mathbf{Y}_{\mathbf{D}} \in T_{\mathbf{D}}\,\mathrm{Sym}^+(m)$ to a point $\mathbf{X} \in \mathrm{Sym}^+(m)$, is uniquely defined as

$$\mathbf{X} = \mathrm{Exp}_D(\mathbf{Y}_{\mathbf{D}}) = \gamma_{\mathbf{D}}(1; \mathbf{Y}_{\mathbf{D}}) = \mathbf{B}\exp(\mathbf{B}^{-1}\mathbf{Y}_{\mathbf{D}}\mathbf{B}^{-T})\mathbf{B}^T. \tag{13}$$

When $\mathbf{D} = \mathbf{I}_m$, Exp_D reduces to the matrix exponential $\exp(\cdot)$. Recall that in the geodesic model, we assume that $\Sigma(\mathbf{0}, \boldsymbol{\beta}) = \mathbf{D} = \mathbf{B}\mathbf{B}^T$ and $\Sigma(\mathbf{x}_i, \boldsymbol{\beta}) = \gamma_{\mathbf{D}}(1; \mathbf{Y}_{\mathbf{D},i}(\boldsymbol{\beta}))$, in which $\mathrm{vecs}(\mathbf{Y}_{\mathbf{D},i}(\boldsymbol{\beta})) = \mathbf{l}(\mathbf{x}_i, \boldsymbol{\beta})$ and $\mathbf{l}(\mathbf{0}, \boldsymbol{\beta}) = \mathbf{0}$.

The Riemannian logarithmic map at $\mathbf{D}$, denoted by $\mathrm{Log}_{\mathbf{D}}(\cdot)$, maps $\mathbf{X} \in \mathrm{Sym}^+(m)$ onto the tangent vector $\mathbf{Y}_{\mathbf{D}}$ in $T_{\mathbf{D}}\,\mathrm{Sym}^+(m)$. Specifically, $\mathrm{Log}_{\mathbf{D}}: \mathrm{Sym}^+(m) \to T_{\mathbf{D}}\,\mathrm{Sym}^+(m)$ at $\mathbf{D} = \mathbf{B}\mathbf{B}^T \in \mathrm{Sym}^+(m)$ is uniquely defined as

$$\mathbf{Y}_{\mathbf{D}} = \mathrm{Log}_{\mathbf{D}}(\mathbf{X}) = \mathbf{B}\log(\mathbf{B}^{-1}\mathbf{X}\mathbf{B}^{-T})\mathbf{B}^T. \tag{14}$$

The Riemannian exponential and logarithmic maps satisfy

$$\mathbf{Y}_{\mathbf{D}} = \mathrm{Log}_{\mathbf{D}}(\mathrm{Exp}_{\mathbf{D}}(\mathbf{Y}_{\mathbf{D}})) \quad \text{and} \quad \mathbf{X} = \mathrm{Exp}_{\mathbf{D}}(\mathrm{Log}_{\mathbf{D}}(\mathbf{X})). \tag{15}$$

Recall that in the definition of "conditional residual," we have two points $\mathbf{D} = \Sigma(\mathbf{x}_i, \boldsymbol{\beta})$ and $\mathbf{X} = \mathbf{S}_i$ on $\mathrm{Sym}^+(m)$ and then we define $\mathrm{Log}_{\Sigma(\mathbf{x}_i, \boldsymbol{\beta})}(\mathbf{S}_i)$ on $T_{\Sigma(\mathbf{x}_i, \boldsymbol{\beta})}\,\mathrm{Sym}^+(m)$. Because $\mathrm{Log}_{\Sigma(\mathbf{x}_i, \boldsymbol{\beta})}(\mathbf{S}_i)$ is independent of the square root of $\Sigma(\mathbf{x}_i, \boldsymbol{\beta})$, we can choose the Cholesky decomposition of $\Sigma(\mathbf{x}_i, \boldsymbol{\beta}) = \mathbf{C}_i(\boldsymbol{\beta})\mathbf{C}_i(\boldsymbol{\beta})^T$ in the definition of $\mathrm{Log}_{\Sigma(\mathbf{x}_i, \boldsymbol{\beta})}(\mathbf{S}_i)$.

A nice property of $\mathrm{Sym}^+(m)$ is that a group action of $\mathrm{GL}(m)$ on $\mathrm{Sym}^+(m)$ can relate any two $(\mathbf{X}, \mathbf{D}) \in \mathrm{Sym}^+(m)$ and the tangent spaces at $\mathbf{X}$ and $\mathbf{D}$, respectively. Specifically, the group action of $\mathrm{GL}(m)$ on $\mathrm{Sym}^+(m)$ consists of all transformations $\phi_{\mathbf{B}}(\mathbf{X}) = \mathbf{B}\mathbf{X}\mathbf{B}^T$ for any $\mathbf{B} \in \mathrm{GL}(m)$. For any $(\mathbf{X}, \mathbf{D}) \in \mathrm{Sym}^+(m)$, there exists $\mathbf{B} \in \mathrm{GL}(m)$ such that $\mathbf{B}\mathbf{X}\mathbf{B}^T = \mathbf{D}$. The group action of $\mathrm{GL}(m)$ on $\mathrm{Sym}^+(m)$ induces a group action between $T_{\mathbf{X}}\,\mathrm{Sym}^+(m)$ and $T_{\phi_{\mathbf{B}}(\mathbf{X})}\,\mathrm{Sym}^+(m)$. Explicitly, if $\mathbf{Y} \in T_{\mathbf{X}}\,\mathrm{Sym}^+(m)$, then $\phi_{\mathbf{B}}(\mathbf{Y}) = \mathbf{B}\mathbf{Y}\mathbf{B}^T \in T_{\phi_{\mathbf{B}}(\mathbf{X})}\,\mathrm{Sym}^+(m)$. Particularly, we have applied the group action to translate $\mathrm{Log}_{\Sigma_i(\boldsymbol{\beta})}(\mathbf{S}_i) \in T_{\Sigma_i(\boldsymbol{\beta})}\,\mathrm{Sym}^+(m)$ to $\mathbf{C}_i(\boldsymbol{\beta})^{-1}\mathrm{Log}_{\Sigma_i(\boldsymbol{\beta})}(\mathbf{S}_i) \times \mathbf{C}_i(\boldsymbol{\beta})^{-T} \in T_{\mathbf{C}_i(\boldsymbol{\beta})^{-1}\Sigma_i(\boldsymbol{\beta})\mathbf{C}_i(\boldsymbol{\beta})^{-T}}\,\mathrm{Sym}^+(m) = T_{\mathbf{I}_m}\,\mathrm{Sym}^+(m)$ for $i = 1, \dots, n$.

We consider the geodesic distance between any two points on $\mathrm{Sym}^+(m)$. Let $\mathbf{B} \in \mathrm{GL}(m)$ be any square root of $\mathbf{D}$, the geodesic distance between $\mathbf{D}$ and $\mathbf{X}$ in $\mathrm{Sym}^+(m)$ is uniquely given by

$$d(\mathbf{D}, \mathbf{X}) = \sqrt{\langle\!\langle \mathrm{Log}_{\mathbf{D}}(\mathbf{X}), \mathrm{Log}_{\mathbf{D}}(\mathbf{X}) \rangle\!\rangle}$$
$$= \sqrt{\mathrm{tr}(\log^2(\mathbf{B}^{-1}\mathbf{X}\mathbf{B}^{-T}))}. \tag{16}$$

The geodesic distance has many nice properties. For instance, the geodesic distance is a proper metric satisfying positive definiteness, symmetry, and the triangle inequality. Specifically, $d(\mathbf{D}, \mathbf{X}) = d(\mathbf{X}, \mathbf{D})$. The geodesic distance is also invariant under group actions, that is, $d(\mathbf{D}, \mathbf{X}) = d(\mathbf{B}\mathbf{D}\mathbf{B}^T, \mathbf{B}\mathbf{X}\mathbf{B}^T)$ for any $\mathbf{B} \in \mathrm{GL}(m)$.

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:

$$\hat{\boldsymbol{\beta}} = \operatorname*{argmin}_{\boldsymbol{\beta}} G_n(\boldsymbol{\beta}) = \operatorname*{argmin}_{\boldsymbol{\beta}} \sum_{i=1}^{n} \mathrm{tr}(\mathcal{E}_i(\boldsymbol{\beta})^2). \tag{17}$$

Thus, $\hat{\boldsymbol{\beta}}$ solves the estimating equations given by

$$\partial_{\boldsymbol{\beta}} G_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} \partial_{\boldsymbol{\beta}}\,\mathrm{tr}(\mathcal{E}_i(\boldsymbol{\beta})^2)$$
$$= \sum_{i=1}^{n} \mathrm{tr}\big(\partial_{\boldsymbol{\beta}}\big[\log(\mathbf{C}_i(\boldsymbol{\beta})^{-1}\mathbf{S}_i\mathbf{C}_i(\boldsymbol{\beta})^{-T})\big]^2\big) = \mathbf{0}, \tag{18}$$

where ∂ denotes partial differentiation with respect to a parameter vector, such as $\boldsymbol{\beta}$. The ILSE is closely related to the intrinsic mean Σ_{IM} of $\mathbf{S}_1, \dots, \mathbf{S}_n \in \mathrm{Sym}^+(m)$, which is defined as

$$\hat{\Sigma}_{IM} = \operatorname*{argmin}_{\Sigma} \sum_{i=1}^{n} d(\Sigma, \mathbf{S}_i)^2, \tag{19}$$

where $d(\cdot, \cdot)$ has been defined in equation (16). In this case, both Σ_i and $\mathbf{C}_i(\boldsymbol{\beta})$ are independent of i and the covariates. Let $\boldsymbol{\beta} = (c_{11}, c_{21}, c_{22}, \ldots, c_{m1}, \ldots, c_{mm})$ be the entries of the lower triangular portion of $\mathbf{C}(\boldsymbol{\beta}) = (c_{jk})$ such that $\Sigma = \mathbf{C}(\boldsymbol{\beta})\mathbf{C}(\boldsymbol{\beta})^T$. It follows from (16) that $\sum_{i=1}^{n} \text{tr}(\mathcal{E}_i(\boldsymbol{\beta})^2) = \sum_{i=1}^{n} d(\Sigma, \mathbf{S}_i)^2$, which leads to $\hat{\Sigma}_{IM} = \mathbf{C}(\hat{\boldsymbol{\beta}})\mathbf{C}(\hat{\boldsymbol{\beta}})^T$.

We now develop an annealing optimization algorithm for obtaining $\hat{\boldsymbol{\beta}}$. Our annealing algorithm consists of five steps. In Step 1, we first calculate the Cholesky decompositions of all $\mathbf{S}_i$, denoted by $\mathbf{S}_i = \tilde{\mathbf{C}}_i \tilde{\mathbf{C}}_i^T$, and then we fit a multivariate linear regression model with the lower part of $\tilde{\mathbf{C}}_i$ as the response and the corresponding part of $\mathbf{C}_i(\boldsymbol{\beta})$ as the mean function, which yield the initial value $\boldsymbol{\beta}^{(0)}$. Then, starting from $\boldsymbol{\beta}^{(0)}$, we use the downhill simplex method to search for another initial value $\boldsymbol{\beta}^{(1)}$ (Nelder and Mead 1965). In Step 2, starting from $\boldsymbol{\beta}^{(k)}$, we use the Gibbs sampler coupled with the Metropolis–Hasting algorithm to iteratively draw N_0 (say, 200) dependent observations $\{\boldsymbol{\beta}^{(k,l)} : l = 1, \ldots, N_0\}$ from $\exp(-G_n(\boldsymbol{\beta})/\tau_k)$. Among the $\{\boldsymbol{\beta}^{(k,l)} : l = 1, \ldots, N_0\}$, we find the optimal $\tilde{\boldsymbol{\beta}}^{(k)} = \text{argmin}_{\boldsymbol{\beta}^{(k,l)}}\{G_n(\boldsymbol{\beta})\}$. In Step 3, we use the downhill simplex method with $\tilde{\boldsymbol{\beta}}^{(k)}$ as the starting point to search for $\boldsymbol{\beta}^{(k+1)}$. In Step 4, we iterate between Steps 2 and 3 for $k = 1, \ldots, M_0$ (say, $M_0 = 4$) and increase τ_k at each iteration. We choose τ_1 to achieve an acceptance rate of the Metropolis–Hasting algorithm within the Gibbs sampler around 0.25 and then we slightly reduce the acceptance rate to around 0.15 as k increases. We save the optimal $\boldsymbol{\beta}^B$, which minimizes $G_n(\boldsymbol{\beta})$, among all $\boldsymbol{\beta}^{(k,l)}$ and $\boldsymbol{\beta}^{(k)}$ during the iterations. In Step 5, starting from $\boldsymbol{\beta}^B$, we use the Newton–Raphson algorithm to calculate $\hat{\boldsymbol{\beta}}$ as described below. Compared with other optimization methods, such as the simulated annealing algorithm, we can efficiently find $\hat{\boldsymbol{\beta}}$ due to several nice features. According to our experience, $\boldsymbol{\beta}^{(1)}$ is usually close to $\hat{\boldsymbol{\beta}}$, and we always use the downhill simplex method to search for a new optimal estimator after drawing a small number of Monte Carlo samples. Thus, we avoid using large values of N_0 and M_0.

Let $\partial_{\boldsymbol{\beta}} G_n(\boldsymbol{\beta}^{(t)})$ and $\partial_{\boldsymbol{\beta}}^2 G_n(\boldsymbol{\beta}^{(t)})$, respectively, denote the first and second-order partial derivatives of $G_n(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ evaluated at $\boldsymbol{\beta}^{(t)}$. The Newton–Raphson algorithm yields

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} + \rho\{-\partial_{\boldsymbol{\beta}}^2 G_n(\boldsymbol{\beta}^{(t)})\}^{-1} \partial_{\boldsymbol{\beta}} G_n(\boldsymbol{\beta}^{(t)}). \quad (20)$$

In addition, we choose $0 < \rho = 1/2^{k_0} \leq 1$ for some $k_0 \geq 0$ such that $G_n(\boldsymbol{\beta}^{(t+1)}) \leq G_n(\boldsymbol{\beta}^{(t)})$. The Newton–Raphson algorithm stops when the absolute difference between consecutive $\boldsymbol{\beta}^{(t)}$'s is smaller than a predefined small number, say 10^{-4}. At the final iteration, we set $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}^{(t)}$. In addition, because $-\partial_{\boldsymbol{\beta}}^2 G_n(\boldsymbol{\beta}^{(t)})$ may not be positive definite, we approximate $-\partial_{\boldsymbol{\beta}}^2 G_n(\boldsymbol{\beta}^{(t)})$ in order to stabilize the Newton–Raphson algorithm. Details regarding $\partial_{\boldsymbol{\beta}} G_n(\boldsymbol{\beta}^{(t)})$, and $\partial_{\boldsymbol{\beta}}^2 G_n(\boldsymbol{\beta}^{(t)})$ and its approximation are given in lemma 1 of the supplemental document.

2.4 Asymptotic Properties

We first introduce some notation to describe the limiting behavior of the ILSE for SPDs. Let $\boldsymbol{\beta}_*$ be the true value of $\boldsymbol{\beta}$ such that (8) holds at $\boldsymbol{\beta}_*$. Let $\mathbf{C}_{i*}$ denote $\mathbf{C}_i(\boldsymbol{\beta}_*)$ and $\Sigma_{i*} = \Sigma_i(\boldsymbol{\beta}_*) = \mathbf{C}_{i*}\mathbf{C}_{i*}^T$; $\mathcal{B}$ denotes the parameter space for $\boldsymbol{\beta}$; $\|\cdot\|$ denotes the

Euclidean norm of a vector or a matrix; $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$ for any vector $\mathbf{a}$; and $\rightarrow^L$ denotes convergence in distribution.

We establish consistency and asymptotically normality of $\hat{\boldsymbol{\beta}}$. We obtain the following theorems, whose detailed assumptions and proofs can be found in the supplemental document.

Theorem 1. (a) For model (8), if assumptions (C1), (C2), and (C3) in the supplemental document are true, then $\hat{\boldsymbol{\beta}}$ converges to $\boldsymbol{\beta}_*$ in probability.

(b) For model (8), under assumptions (C1)–(C4), we have

$$\left\{\sum_{i=1}^{n} [\partial_{\boldsymbol{\beta}} \text{tr}(\mathcal{E}_i(\hat{\boldsymbol{\beta}})^2)]^{\otimes 2}\right\}^{-1/2} [-\partial_{\boldsymbol{\beta}}^2 G_n(\hat{\boldsymbol{\beta}})](\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_*)$$

$$\rightarrow^L N(0, \mathbf{I}_p) \quad (21)$$

as $n \rightarrow \infty$.

Theorem 1 has several important applications. Theorem 1(a) establishes weak convergence of $\hat{\boldsymbol{\beta}}$. According to Theorem 1(b), the covariance matrix of $\hat{\boldsymbol{\beta}}$ under model (8) can be consistently estimated by

$$[-\partial_{\boldsymbol{\beta}}^2 G_n(\hat{\boldsymbol{\beta}})]^{-1} \left\{\sum_{i=1}^{n} [\partial_{\boldsymbol{\beta}} \text{tr}(\mathcal{E}_i(\hat{\boldsymbol{\beta}})^2)]^{\otimes 2}\right\} [-\partial_{\boldsymbol{\beta}}^2 G_n(\hat{\boldsymbol{\beta}})]^{-1}. \quad (22)$$

Moreover, we can use Theorem 1(b) to construct confidence cones of $\boldsymbol{\beta}$ and its functions. Since Theorem 1 only establishes the asymptotic properties of $\hat{\boldsymbol{\beta}}$ when the sample size is large, these properties may be inadequate to characterize the finite sample behavior of $\hat{\boldsymbol{\beta}}$ for relatively small sample sizes. In the case of small and moderate sample sizes, we may have to resort to higher-order approximations, such as saddlepoint approximations and bootstrap methods (Davison and Hinkley 1997; Butler 2007).

2.5 Testing Linear Hypotheses

Our choice of hypotheses to test is motivated by scientific questions, which involve a comparison of SPDs across diagnostic groups or detecting change in SPDs across time (Schwartzman 2006; Whitcher et al. 2007; Lepore et al. 2008). These questions usually can be formulated as testing linear hypotheses of $\boldsymbol{\beta}$ as follows:

$$H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{b}_0 \qquad \text{vs.} \qquad H_1 : \mathbf{R}\boldsymbol{\beta} \neq \mathbf{b}_0, \quad (23)$$

where $\mathbf{R}$ is an $r \times p$ matrix of full row rank and $\mathbf{b}_0$ is an $r \times 1$ specified vector. We test the null hypothesis $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{b}_0$ using a score test statistic W_n defined by

$$W_n = \mathbf{L}_{n,\boldsymbol{\mu}}^T \hat{\mathbf{I}}_{\boldsymbol{\mu}\boldsymbol{\mu}}^{-1} \mathbf{L}_{n,\boldsymbol{\mu}}, \quad (24)$$

where $\boldsymbol{\mu} = \mathbf{R}\boldsymbol{\beta}$, $\mathbf{L}_{n,\boldsymbol{\mu}} = n^{-1/2} \sum_{i=1}^{n} \hat{\mathbf{U}}_{i,\boldsymbol{\mu}}(\tilde{\boldsymbol{\beta}})$, and $\hat{\mathbf{I}}_{\boldsymbol{\mu}\boldsymbol{\mu}} = n^{-1} \sum_{i=1}^{n} \hat{\mathbf{U}}_{i,\boldsymbol{\mu}}(\tilde{\boldsymbol{\beta}}) \hat{\mathbf{U}}_{i,\boldsymbol{\mu}}(\tilde{\boldsymbol{\beta}})^T$, in which $\tilde{\boldsymbol{\beta}}$ denotes the estimate of $\boldsymbol{\beta}$ under H_0 and the explicit expressions of $\hat{\mathbf{U}}_{i,\boldsymbol{\mu}}(\tilde{\boldsymbol{\beta}})$ and $\mathbf{L}_{n,\boldsymbol{\mu}}$ are given in the supplemental document (Rotnitzky and Jewell 1990).

Theorem 2. For model (8), if assumptions (C1)–(C4) in the supplemental document are true, then the statistic W_n is asymptotically distributed as $\chi^2(r)$, a chi-square distribution with r degrees of freedom, under the null hypothesis H_0.

An asymptotically valid test can be obtained by comparing sample values of the test statistic with the critical value of the right-hand tail of the $\chi^2(r)$ distribution at a prespecified significance level α. That is, we reject H_0 if $W_n \geq \chi_\alpha^2(r)$, and do not reject H_0 otherwise, where $\chi_\alpha^2(r)$ is the upper α-percentile of the $\chi^2(r)$ distribution.

2.6 Multiple Hypotheses

In imaging applications, we need to test the hypotheses H_0 against H_1 across multiple brain regions or across the many voxels of the imaging volume. Thus, the next step entails using statistical methods (e.g., random field theory, false discovery rates, permutation methods) to adjust p-values for these multiple statistical tests. To test H_0 in all voxels of the region under study, we consider the maximum of the score test statistics given by

$$W_{n,\mathcal{D}} = \max_{d \in \mathcal{D}} W_n(d), \qquad (25)$$

where $W_n(d)$ denotes the score test statistic at voxel d and $\mathcal{D}$ denotes the brain region. The maximum statistic $W_{n,\mathcal{D}}$ plays a crucial role in controlling the family-wise error rate.

We propose to use a test procedure that is based on a resampling method to approximate the null distribution of $W_{n,\mathcal{D}}$ (Kosorok 2003; Lin 2005). The test procedure is implemented as follows:

Step 1. We calculate the score test statistic $W_n(d)$ at each voxel d, and then we compute $W_{n,\mathcal{D}} = \max_{d \in \mathcal{D}} W_n(d)$.

Step 2. Generate n realizations of ε denoted by $\{\varepsilon_i^{(k)} : i = 1, \ldots, n\}$, where ε equals ± 1 with equal probability.

Step 3. Calculate $W_n(d)^{(k)} = \mathbf{L}_{n,\mu}(d)^{(k)T}[\hat{\mathbf{I}}_{\mu\mu}(d)]^{-1} \times \mathbf{L}_{n,\mu}(d)^{(k)}$, and then compute $W_{n,\mathcal{D}}^{(k)} = \max_{d \in \mathcal{D}} W_n(d)^{(k)}$, where $\mathbf{L}_{n,\mu}(d)^{(k)} = n^{-1/2} \sum_{i=1}^n \hat{\mathbf{U}}_{i,\mu}(\tilde{\boldsymbol{\beta}}, d) \varepsilon_i^{(k)}$.

Step 4. Repeat Steps 2–3 K_0 times and calculate $\{W_{n,\mathcal{D}}^{(k)} : k = 1, \ldots, K_0\}$. Finally, the p-value of $W_{n,\mathcal{D}}$ is approximated by $p_{\mu,\mathcal{D}} = K_0^{-1} \sum_{k=1}^{K_0} I(W_{n,\mathcal{D}}^{(k)} \geq W_{n,\mathcal{D}})$. We reject the null hypothesis $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{b}_0$ across all voxels of the region when $p_{\mu,\mathcal{D}}$ is smaller than a prespecified value α, say 0.05.

Step 5. Calculate the adjusted p-value of $W_n(d)$ at each voxel d of the region according to $p_D(d) \approx K_0^{-1} \times \sum_{k=1}^{K_0} I(W_{n,\mathcal{D}}^{(k)} \geq W_n(d))$.

There are several advantages of using the resampling method in the above test procedure. The above procedure needs to compute $\hat{\mathbf{U}}_{i,\mu}(\tilde{\boldsymbol{\beta}}, d)$ and $\hat{\mathbf{I}}_{\mu\mu}$ only once. It also avoids repeated analyses of permuted datasets in the permutation methods, because fitting the regression models for SPD matrices across all voxels of an imaging volume can take up to more than 20 hours for each permuted dataset. In contrast, the proposed resampling method takes less than 5 minutes for $K_0 = 1,000$.

3. SIMULATION STUDIES

We conducted three sets of Monte Carlo simulations. All computations for these simulation studies were done in C++ on an IBM ThinkCentre M50 workstation. The first set of simulations was to evaluate the accuracy of the parameter estimates

and their associated variance estimates for the proposed intrinsic regression model. We set $m = 3$ and generated the simulated data as follows. We considered the Cholesky decomposition (1) of $\Sigma(\mathbf{x}_i, \boldsymbol{\beta})$ with $\text{vecs}(\mathbf{C}_i(\boldsymbol{\beta})^T) = (\mathbf{z}_i^T\boldsymbol{\beta}_1, \mathbf{z}_i^T\boldsymbol{\beta}_2, \mathbf{z}_i^T\boldsymbol{\beta}_3, \mathbf{z}_i^T\boldsymbol{\beta}_4, \mathbf{z}_i^T\boldsymbol{\beta}_5, \mathbf{z}_i^T\boldsymbol{\beta}_6)^T$, in which $\mathbf{z}_i = (1, x_{i1})^T$ was a 2×1 vector of covariates of interest, for $i = 1, \ldots, n$. We generated x_{i1} independently from a Gaussian generator with zero mean and unit variance. Thus, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_6^T)^T$ is a 12×1 vector. Then, we simulated $\mathcal{E}_i$ from a $N(0, \Omega)$ distribution and calculated $\mathbf{C}_i(\boldsymbol{\beta}_*) \exp(\mathcal{E}_i) \mathbf{C}_i(\boldsymbol{\beta}_*)^T$, in which we set $\boldsymbol{\beta}_{k*} = [1 + 0.1 \times (k - 1)]\mathbf{1}_2^T$ for $k = 1, \ldots, 6$. We chose two different Ω's as follows: $\Omega_1 = 0.6\mathbf{I}_3$ and $\Omega_2 = 0.3\mathbf{I}_3 + 0.3\mathbf{1}_3\mathbf{1}_3^T$, where $\mathbf{1}_3 = (1, 1, 1)^T$. For each Ω, we set $n = 20$ and 80 and then simulated 500 data sets for each case. For each simulated dataset, we applied the annealing optimization algorithm with $N_0 = 200$ and $M_0 = 4$, which took an average CPU time of about 30 seconds to obtain $\hat{\boldsymbol{\beta}}$ and $\text{Cov}(\hat{\boldsymbol{\beta}})$ in (22).

Based on 500 parameter estimates, we calculated the bias, the mean of the estimated standard error estimates (SE), the standard deviation of the estimated standard error estimates (SD–SE) and the root-mean-square error (RMS) (Table 1). All relative efficiencies (the ratio of the mean of the standard deviation estimates to the root mean-square error) are close to 1.0, indicating that matrix (22) is an accurate estimate of $\text{diag}(\text{Cov}(\hat{\boldsymbol{\beta}}))$. As expected, the root mean-square error decreases as the sample size increases. Moreover, comparing the results from Ω_1 and Ω_2, we note that increasing the correlation in Ω only slightly decreases the bias and SD of $\hat{\boldsymbol{\beta}}$.

The second set of simulations examined the finite sample performance of the score statistic W_n. We used the same setup as the first set of simulations except that we set $\boldsymbol{\beta} = \mathbf{0}$, and then varied the second component of $\boldsymbol{\beta}_j = (\beta_{j,1}, \beta_{j,2})'$ for $j = 1, \ldots, 6$. We were interested in testing whether the effect of the covariate x_{i1} is significant. Letting $\boldsymbol{\beta}_{\cdot,2} = (\beta_{1,2}, \ldots, \beta_{6,2})^T$, we tested the following hypotheses:

$$H_0 : \boldsymbol{\beta}_{\cdot,2} = \mathbf{0} \qquad \text{and} \qquad H_1 : \boldsymbol{\beta}_{\cdot,2} \neq \mathbf{0}.$$

To assess the type I and II error rates for W_n, we chose four different values for $\boldsymbol{\beta}_{\cdot,2}$: $0 \times \mathbf{1}_6$, $0.2 \times \mathbf{1}_6$, $0.4 \times \mathbf{1}_6$, and $0.6 \times \mathbf{1}_6$, and chose Ω_1 and Ω_2. For each Ω, we set $n = 20, 40$, and 80 and then simulated 1,000 datasets for each case.

The score statistic W_n performs reasonably well for relatively small sample sizes (Table 2). The type I error rates were not excessive even for both the 5% and 1% significance levels at $n = 40$. Increasing the sample size can increase the power of rejecting the null hypothesis. Comparing the results from Ω_1 and Ω_2, we note that increasing the correlation in Ω only slightly influences the finite sample performance of W_n.

In the third set of simulations, we examined the finite sample performance of $W_{n,\mathcal{D}}$. We generated 3×3 SPDs at all 2,500 pixels of a 50 by 50 image slice for each of all n subjects. We used the same setup as the first set of simulations with an exception. We only consider Ω_1, the variance of $\mathcal{E}_i(d)$, at each pixel and for any two pixels d and d', the correlation of $\mathcal{E}_{i,k}(d)$ and $\mathcal{E}_{i,k}(d')$ equals $\rho^{\|d-d'\|}$, where $\rho \in [0, 1]$. We tested the null hypothesis $H_0 : \boldsymbol{\beta}_{\cdot,2}(d) = \mathbf{0}$ at all pixels on the slice. We set $n = 40$ and 80. We first assumed $\boldsymbol{\beta}_{\cdot,2}(d) = \mathbf{0}$ at all pixels on the slice to assess the family-wise error rate. To assess the power, we se-

Table 1. Bias ($\times 10^{-2}$), RMS ($\times 10^{-2}$), SE ($\times 10^{-2}$), SD–SE ($\times 10^{-2}$), and RS of all 12 parameters under Ω_1 and Ω_2. BIAS denotes the bias of the mean of the ILSE estimates; RMS denotes the root-mean-square error; SE denotes the mean of the standard deviation estimates; SD–SE denotes the standard deviation of the standard deviation estimates; RS denotes the ratio of RMS over SD. Two different sample sizes {20, 80} and 500 simulated datasets were used for each case

	$n = 20$					$n = 80$				
	BIAS	RMS	SE	SD–SE	RS	BIAS	RMS	SE	SD–SE	RS
					Ω_1					
β_1	2.60	6.10	6.73	2.20	1.10	0.58	3.59	3.37	0.68	0.94
β_2	1.78	6.10	6.61	1.72	1.08	0.06	3.90	3.61	0.55	0.92
β_3	1.88	7.06	6.96	1.32	0.98	0.69	3.91	3.56	0.35	0.91
β_4	1.15	6.86	6.89	1.30	1.01	0.35	3.83	3.51	0.33	0.92
β_5	3.83	15.34	17.24	4.22	1.12	1.08	8.35	8.58	1.20	1.02
β_6	2.83	15.07	16.97	3.76	1.12	0.54	8.45	8.46	0.95	1.01
β_7	1.43	8.75	8.07	1.42	0.92	−0.37	4.19	4.10	0.39	0.98
β_8	0.48	8.32	7.98	1.40	0.96	−0.44	4.12	4.06	0.38	0.98
β_9	5.14	29.38	32.06	7.57	1.09	1.6	14.84	16.07	2.23	1.08
β_{10}	3.97	28.88	31.59	7.14	1.09	1.05	14.84	15.85	1.77	1.07
β_{11}	3.62	20.32	19.91	3.87	0.98	1.00	10.63	10.15	0.85	0.96
β_{12}	2.69	20.11	19.68	3.83	0.98	0.53	10.48	10.03	0.84	0.96
					Ω_2					
β_1	2.76	6.87	6.73	2.21	0.97	0.59	4.00	3.7	0.57	0.93
β_2	1.81	6.72	6.61	1.63	0.98	0.46	3.97	3.7	0.53	0.93
β_3	1.96	7.74	7.23	1.27	0.93	0.23	3.72	3.54	0.36	0.95
β_4	1.24	7.43	7.15	1.26	0.96	0.03	3.67	3.5	0.36	0.95
β_5	3.08	11.63	12.61	2.85	1.08	1.02	6.87	6.33	0.73	0.92
β_6	1.87	11.78	12.41	2.31	1.05	0.86	6.9	6.3	0.7	0.91
β_7	1.56	8.44	8.26	1.46	0.98	0.11	4.5	4.2	0.34	0.93
β_8	0.62	8.09	8.17	1.44	1.01	0.04	4.5	4.1	0.34	0.91
β_9	2.75	18.90	19.75	4.14	1.04	1.06	10.74	9.93	1.09	0.93
β_{10}	1.33	18.84	19.46	3.68	1.03	0.86	10.79	9.82	0.93	0.91
β_{11}	3.87	16.68	15.49	2.54	0.93	−0.58	7.45	7.44	0.66	0.99
β_{12}	2.88	16.46	15.31	2.51	0.94	−0.91	7.22	7.37	0.65	1.02

lected a region-of-interest (ROI) with 40 pixels on the reference slice and set $\boldsymbol{\beta}_{\cdot,2}(d)$ at $0.3 \times \mathbf{1}_6$, $0.6 \times \mathbf{1}_6$, and $0.9 \times \mathbf{1}_6$ for any point d in the ROI. For each simulated dataset, we applied the annealing optimization algorithm with $N_0 = 200$ and $M_0 = 5$

to the simulated data at each pixel and the test procedure to all 2,500 pixels, which took an average CPU time of about 8 hours.

We used the family-wise error rate as the type I error rate and estimated it based on 100 replications with significance level

Table 2. Comparisons of the rejection rates for score test statistics under Ω_1 and Ω_2. Three differing sample sizes {20, 40, 80} and 1,000 simulated datasets were used for each case and two significance levels, 5% and 1%, were considered. The null and alternative hypotheses were, respectively, given by $H_0 : \boldsymbol{\beta}_{\cdot,2} = \mathbf{0}$ and $H_1 : \boldsymbol{\beta}_{\cdot,2} \neq \mathbf{0}$. Two methods including the resampling method (RE) and χ^2 distribution [$\chi^2(6)$] were used to calculate the rejection rates

	$n = 20$				$n = 40$				$n = 80$			
	5%		1%		5%		1%		5%		1%	
$\boldsymbol{\beta}_{\cdot,2}$	RE	$\chi^2(6)$	RE	$\chi^2(6)$	RE	$\chi^2(6)$	RE	$\chi^2(6)$	RE	$\chi^2(6)$	RE	$\chi^2(6)$
						Ω_1						
$0 \times \mathbf{1}_6$	0.143	0.031	0.037	0	0.067	0.043	0.026	0.007	0.067	0.037	0.017	0.003
$0.2 \times \mathbf{1}_6$	0.513	0.177	0.253	0.011	0.957	0.883	0.796	0.461	1	1	0.991	0.971
$0.4 \times \mathbf{1}_6$	0.597	0.213	0.293	0.022	0.993	0.951	0.832	0.481	1	1	1	1
$0.6 \times \mathbf{1}_6$	0.773	0.442	0.520	0.042	1	1	1	0.983	1	1	1	1
						Ω_2						
$0 \times \mathbf{1}_6$	0.126	0.023	0.037	0	0.063	0.037	0.017	0.003	0.061	0.033	0.013	0.003
$0.2 \times \mathbf{1}_6$	0.581	0.221	0.302	0.010	0.977	0.953	0.851	0.491	1	1	0.991	0.991
$0.4 \times \mathbf{1}_6$	0.602	0.227	0.321	0.032	0.991	0.981	0.871	0.562	1	1	1	1
$0.6 \times \mathbf{1}_6$	0.903	0.51	0.611	0.051	1	1	1	0.981	1	1	1	1

Table 3. Comparisons of the family-wise error rates and average powers for the test procedure under two different correlations $\rho = 0.0$ and 0.5. We considered two different sample sizes $\{40, 80\}$ and 100 simulated datasets for each case at the 5% significance level. In all voxels, the null and alternative hypotheses were, respectively, given by $H_0 : \boldsymbol{\beta}_{\cdot,2}(d) = \mathbf{0}_6$ and $H_1 : \boldsymbol{\beta}_{\cdot,2}(d) \neq \mathbf{0}_6$. We considered four different $\boldsymbol{\beta}_{\cdot,2}(d)$ $\{0.0 \times \mathbf{1}_6, 0.3 \times \mathbf{1}_6, 0.6 \times \mathbf{1}_6, 0.9 \times \mathbf{1}_6\}$ for all voxels within the region of interest, whereas we set $\boldsymbol{\beta}_{\cdot,2}(d) = \mathbf{0}_6$ for all voxels outside the region of interest. FWR denotes the family wise error rate and Apower denotes the average rejection rate for voxels inside the region of interest

| | $n = 40$ | | | | $n = 80$ | | | |
| | $\rho = 0.0$ | | $\rho = 0.5$ | | $\rho = 0.0$ | | $\rho = 0.5$ | |
$\boldsymbol{\beta}_{\cdot,2}(d)$	FWR	Apower	FWR	Apower	FWR	Apower	FWR	Apower
$0.0 \times \mathbf{1}_6$	0.12	0.00	0.06	0.00	0.08	0.00	0.07	0.00
$0.3 \times \mathbf{1}_6$	0.18	0.10	0.12	0.10	0.06	0.56	0.06	0.57
$0.6 \times \mathbf{1}_6$	0.14	0.67	0.06	0.68	0.02	1.00	0.03	1.00
$0.9 \times \mathbf{1}_6$	0.12	0.83	0.10	0.85	0.08	1.00	0.06	1.00

$\alpha = 5\%$. We also calculated the average of the probabilities of rejecting each of the 40 pixels in the ROI as an estimate of the average power using 100 replications and a significance level of $\alpha = 5\%$.

For the test statistic $W_{n,\mathcal{D}}$, our test procedure worked reasonably well for relatively small sample sizes (Table 3). The family-wise error rates for our test procedure were not particularly accurate for $n = 40$ when the correlation is zero; in contrast, they improved at $n = 40$ and $\rho = 0.5$. Thus, the sample size could somewhat influence the finite sample performance of our test procedure, particularly when the sample sizes are small.

4. HIV IMAGING DATA

We assess the integrity of white matter in human immunodeficiency virus (HIV). White matter is one of the three main solid components of the central nervous system (CNS) and is composed of bundles of myelinated nerve cell processes (or axons), which connect various gray matter areas (the locations of nerve cell bodies) of the brain to each other, and carry nerve impulses between neurons. The white matter is important for passing messages between different areas of gray matter within the CNS. After initial HIV infection, the virus is detectable in the CNS before antibodies are detectable in the blood and HIV could be cultured from brain tissue as early as 15 days (Davis et al. 1992; Rausch and Davis 2001). Since DTI can detect the subtle disruption of white matter structural integrity by assessing the degree to which fiber tracts within the white matter have lost their directional organization (Basser, Mattiello, and LeBihan 1994a, 1994b; Lim and Helpern 2002; Focke et al. 2008; Vernooij et al. 2008), DTI may be an important tool for detecting the early CNS HIV involvement.

We considered 47 subjects, in which 29 were HIV+ subjects (21 males and 8 females) and 18 were healthy (9 males and 9 females) controls from a cross-sectional study. The ages of the HIV+ subjects ranged from 30 to 52 years (mean: 40.0, SD: 5.6 years) and those of healthy controls ranged from 27 to 54 years (mean: 41.2, SD: 7.4 years). For each subject, both diffusion-weighted images and T1 weighted images were acquired. Diffusion gradients with a b-value of $1,000$ s/mm^2 were applied in six noncollinear directions, $(1, 0, 1)$, $(-1, 0, 1)$, $(0, 1, 1)$, $(0, 1, -1)$, $(1, 1, 0)$, and $(-1, 1, 0)$. A $b = 0$ reference scan was also obtained for diffusion tensor matrix calculations. Forty-six contiguous slices with a slice thickness of 2 mm covered a field

of view (FOV) of 256×256 mm^2 with an isotropic voxel size of $2 \times 2 \times 2$ mm^3. Eighteen acquisitions were used to improve the signal-to-noise ratio (SNR) in the images. High resolution T1 weighted (T1W) images were acquired using a 3D MP-RAGE sequence. Then, a weighted least squares estimation method was used to construct the diffusion tensors (Basser, Mattiello, and LeBihan 1994b; Zhu et al. 2007).

All images were visually inspected before analysis to ensure no bulk motion. A two-step image registration approach was utilized to spatially normalize the DTI parameters. The first step used a B-spline model based on a bidirectional elastic registration method to align the T1 weighted images of all subjects including both normal controls and HIV+ subjects to the T1 weighted images (template) of an arbitrarily chosen 41-year-old female healthy subject. In order to minimize biases induced by the choice of the template, the symmetry between the image pair to be coregistered was ensured through enforcing the consistency between the forward (e.g., from subject X to the template) and backward (from the template to subject X) transformations. The second step was to align each subject's DTI images to his/her own anatomical T1 weighted images with a 12-parameter affine registration tool in FSL 3.2 (Analysis Group, FMRIB, Oxford, U.K.), so that DTI results can be spatially normalized through the spatial transformation obtained from high resolution T1 elastic registration. The tensor reorientation matrix at each voxel was derived as a rotation matrix approximating the transformation matrix (from T1 registration) through a singular value decomposition (SVD) (Alexander et al. 2001; Xu et al. 2003).

To control for the effects of covariates (diagnosis, age, and gender), we considered model (2) for $\mathbf{C}_i(\boldsymbol{\beta})$ for diffusion tensors at each voxel. The $\mathbf{z}_i = (1, x_{1i}, x_{2i}, x_{3i})^T$ is a 4×1 vector, in which x_{1i} is Age/10, x_{2i} is gender, and x_{3i} denotes the diagnosis (1 HIV+ and 0 Healthy control). Moreover, we limited the statistical analysis within the major white matter regions which contains 17,444 voxels with mean fractional anisotropy (FA) value in normal volunteers greater than 0.4. We applied the annealing optimization algorithm with $N_0 = 200$ and $M_0 = 4$, which took an average CPU time of approximately 80 hours to carry out the statistical analysis.

We detected the statistical significance of diagnosis (or age) on the integrity of white matter at all voxels with FA value greater than 0.4. Here, $\mathbf{R}$ is a 6×24 matrix and $\mathbf{b}_0 = (0, 0, 0, 0, 0, 0)^T$ for the hypotheses on either diagnosis or age.
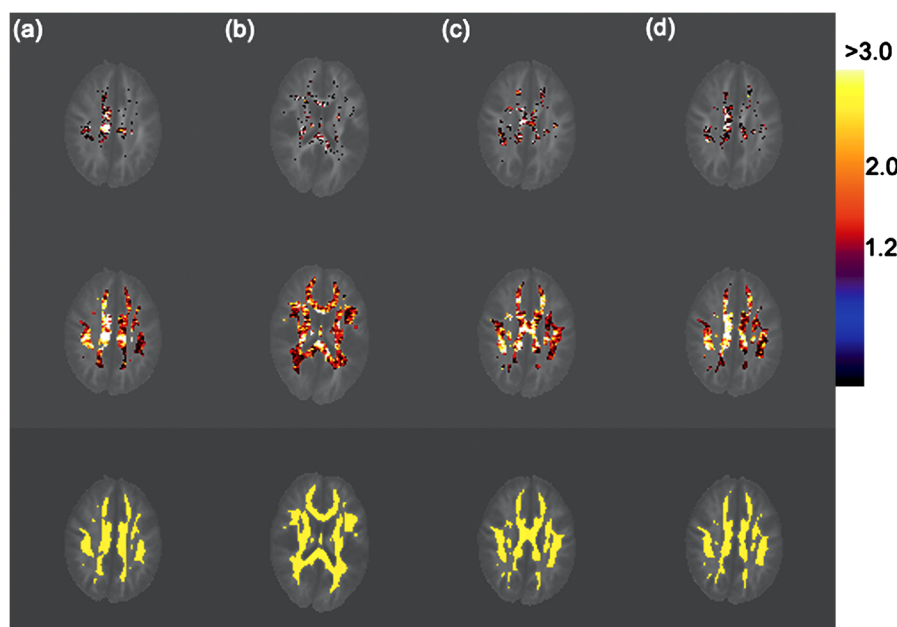
Figure 2. Significance testing of diagnosis effect: color-coded maps of raw and adjusted p-values in four selected ROIs of the reference brain. The color scale reflects the magnitude of the values of $-\log_{10}(P)$, with black to blue representing smaller values (0–1) and red to white representing larger values (1.88–3). Row 1: adjusted $-\log_{10}(P)$ values of the score statistics based on our test procedure for the correction of multiple comparisons. Row 2: raw $-\log_{10}(P)$ values of the score statistics based on a χ^2 distribution. Row 3: selected ROIs with FA values greater than 0.4. After correcting for multiple comparisons, statistically significant diagnosis effects remain in the superior internal capsule area.

The uncorrected p-values based on the score statistics were color coded at each voxel in the selected regions of the reference brain (Figure 2). To correct for multiple comparisons, we applied our test procedure to calculate the adjusted p-value $p_D(d)$ at each voxel in the selected regions of the reference brain (Figure 2). Color-coded maps of p-values using either the uncorrected $p(d)$ alone or the corrected $p_D(d)$ indicated several large-scale diagnosis and age effects for white matter integrity. The test procedure for correcting for multiple hypotheses, however, captured far fewer points of differences in the selected regions of the reference brain (Figure 2). As previously reported in studies of HIV (Kure et al. 1990; Pomara et al. 2001), the patients with HIV+ exhibited the subtle disruption of white matter structural integrity in the internal capsule and inferior longitudinal fasciculus. Our studies based on diffusion tensors thus seem

to confirm these earlier results obtained with FA values (Pomara et al. 2001). We also picked two voxels and presented their diffusion tensors using an ellipsoid representation (Figure 3). We observed significant differences between HIV+ subjects and healthy controls (Figure 3a) and observed different age trends between HIV+ subjects and healthy controls (Figure 3b).

5. DISCUSSION

We have developed a general statistical framework for an intrinsic regression model of positive-definite matrices as responses in a Riemannian manifold and their association with a set of covariates, such as age and gender, in Euclidean space. The intrinsic regression model is based on the first moment of imaging measures and therefore it avoids any parametric assumptions regarding SPD matrices. We have proposed several
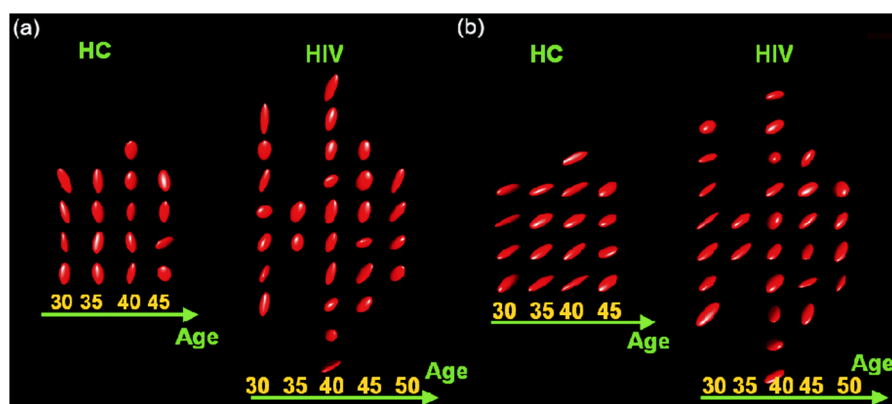

Figure 3. Ellipsoid representation of diffusion tensors from two voxels to illustrate significant effects of diagnosis and age. Panel (a): testing diagnosis effect. Panel (b): testing age effect.

link functions to map covariates in Euclidean space to positive-definite matrices in the Riemannian manifold $\text{Sym}^+(m)$. We have developed an annealing optimization algorithm to search for the ILSE of $\boldsymbol{\beta}$. The test procedure based on the resampling method not only accounts for multiple comparisons across the entire region of interest under investigation, but it also asymptotically preserves the dependence structure among the test statistics. Our simulation studies have demonstrated that the methodology developed here provides relatively accurate control of the family-wise error rate.

We also note several limitations of our procedures. There is a computational issue with the estimation procedure for computing $\hat{\boldsymbol{\beta}}$. Since $G_n(\boldsymbol{\beta})$ is generally not concave and can have many local minima, the annealing optimization algorithm proposed here cannot guarantee to locating the global minimum $\hat{\boldsymbol{\beta}}$ of $G_n(\boldsymbol{\beta})$ for arbitrary link function. More advanced optimization algorithms are needed for more complex link functions. Our test procedure based on $W_{n,\mathcal{D}}$ and the resampling method performs reasonably well for relatively larger sample sizes, say 100. However, further improvement on controlling the family-wise error rates is needed. We may either use other computationally extensive methods, such as permutation method, in order to achieve better control of the family-wise error rates, or use other multiple comparison procedures, such as the false discovery rate.

Many issues still merit further research. One major issue is to develop diagnostic measures for assessing the influence of individual observations in semiparametric regression for SPD matrices. Another major issue is to construct goodness-of-fit statistics for testing possible misspecification in (7). Moreover, it is of interest to develop nonparametric regression methods for SPD matrices (Fan and Gijbels 1996). We will study these issues in our future work.

SUPPLEMENTAL MATERIALS

Supplemental materials include assumptions, proofs of Theorems 1 and 2, and a figure for significance testing of age effects.

[Received February 2008. Revised February 2009.]

REFERENCES

Alexander, D. C., Pierpaoli, C., Basser, P. J., and Gee, J. C. (2001), "Spatial Transformations of Diffusion Tensor Magnetic Resonance Images," *IEEE Transactions on Medical Imaging*, 20, 1131–1139.

Anderson, T. W. (2003), *An Introduction to Multivariate Statistical Analysis* (3rd ed.), *Wiley Series in Probability and Statistics*, Hoboken, NJ: Wiley.

Basser, P. J., Mattiello, J., and LeBihan, D. (1994a), "MR Diffusion Tensor Spectroscopy and Imaging," *Biophysical Journal*, 66, 259–267.

——— (1994b), "Estimation of the Effective Self-Diffusion Tensor From the NMR Spin Echo," *Journal of Magnetic Resonance, Ser. B*, 103, 247–254.

Bhattacharya, R., and Patrangenaru, V. (2003), "Large Sample Theory of Intrinsic and Extrinsic Sample Means on Manifolds-I," *The Annals of Statistics*, 31, 1–29.

——— (2005), "Large Sample Theory of Intrinsic and Extrinsic Sample Means on Manifolds-II," *The Annals of Statistics*, 33, 1225–1259.

Butler, R. W. (2007), *Saddlepoint Approximations With Applications*, New York: Cambridge University Press.

Chefd'hotel, C., Tschumperlé, D., Deriche, R., and Faugeras, O. (2004), "Regularizing Flows for Constrained Matrix-Valued Images," *Journal of Mathematical Imaging and Vision*, 20, 147–162.

Chiu, T. Y. M., Leonard, T., and Tsui, K. W. (1996), "The Matrix-Logarithmic Covariance Model," *Journal of the American Statistical Association*, 91, 198–210.

Davis, L. E., Hjelle, B. L., Miller, V. E., Palmer, D. L., Llewellyn, A. L., Merlin, T. L., Young, S. A., Mills, R. G., Wachsman, W., and Wiley, C. A. (1992), "Early Viral Brain Invasion in Iatrogenic Human Immunodeficiency Virus Infection," *Neurology*, 42, 1736–1739.

Davison, A. C., and Hinkley, D. V. (1997), *Bootstrap Methods and Their Application*, Cambridge: Cambridge University Press.

Do Carmo, M. P. (1992), *Riemannian Geometry*, Boston: Birkhäuser, Basel.

Fan, J., and Gijbels, I. (1996), *Local Polynomial Modeling and Its Applications*, London: Chapman & Hall.

Focke, N. K., Yogarajah, M., Bonelli, S. B., Bartlett, P. A., Symms, M. R., and Duncan, J. S. (2008), "Voxel-Based Diffusion Tensor Imaging in Patients With Mesial Temporal Lobe Epilepsy and Hippocampal Sclerosis," *NeuroImage*, 40, 728–737.

Fréchet, M. (1948), "Les éléments Aléatoires De Nature Quelconque Dans Un Espace Distancé," *Annales de l'Institut Henri Poincaré*, 10, 215–230.

Grenander, U., and Miller, M. I. (1998), "Computational Anatomy: An Emerging Discipline," *Quarterly of Applied Mathematics*, 56, 617–694.

Gupta, A. K., and Nagar, D. K. (2000), *Matrix Variate Distributions*, Boca Raton, FL: Chapman & Hall/CRC.

Karcher, H. (1977), "Riemannian Center of Mass and Mollifier Smoothing," *Communications on Pure and Applied Mathematics*, 30, 509–541.

Kim, P. T., and Richards, D. S. (2008), "Deconvolution Density Estimation on Spaces of Positive Definite Symmetric Matrices," technical report, The Pennsylvania State University, Dept. of Statistics.

Kosorok, M. R. (2003), "Bootstraps of Sums of Independent but Not Identically Distributed Stochastic Processes," *Journal of Multivariate Analysis*, 84, 299–318.

Kure, K., Lyman, W. D., Weidenheim, K. M., and Dickson, D. W. (1990), "Cellular Localization of an HIV-1 Antigen in Subacute AIDS Encephalitis Using an Improved Double-Labeling Immunohistochemical Method," *American Journal of Pathology*, 136, 1085–1092.

Lang, S. (1999), *Fundamentals of Differential Geometry*, New York: Springer-Verlag.

Lepore, N., Brun, C. A., Chou, Y., Chiang, M., Dutton, R. A., Hayashi, K. M., Luders, E., Lopez, O. L., Aizenstein, H. J., Toga, A. W., Becker, J. T., and Thompson, P. M. (2008), "Generalized Tensor-Based Morphometry of HIV/AIDS Using Multivariate Statistics on Deformation Tensors," *IEEE Transactions in Medical Imaging*, 27, 129–141.

Lim, K. O., and Helpern, J. A. (2002), "Neuropsychiatric Applications of DTI—a Review," *NMR in Biomedicine*, 15, 587–593.

Lin, D. Y. (2005), "An Efficient Monte Carlo Approach to Assessing Statistical Significance in Genomic Studies," *Bioinformatics*, 21, 781–787.

McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models* (2nd ed.), London: Chapman & Hall.

Nelder, J. A., and Mead, R. (1965), "A Simplex Method for Function Minimization," *Computer Journal*, 7, 308–313.

Pennec, X., Fillard, P., and Ayache, N. (2006), "A Riemannian Framework for Tensor Computing," *International Journal of Computer Vision*, 66, 41–66.

Pomara, N., Crandalla, D. T., Choia, S. J., Johnson, G., and Lim, K. O. (2001), "White Matter Abnormalities in HIV-1 Infection: A Diffusion Tensor Imaging Study," *Psychiatry Research: Neuroimaging*, 106, 15–24.

Pourahmadi, M. (1999), "Joint Mean–Covariance Models With Applications to Longitudinal Data, I: Unconstrained Parametrization," *Biometrika*, 86, 677–690.

——— (2000), "Maximum Likelihood Estimation of Generalized Linear Models for Multivariate Normal Covariance Matrix," *Biometrika*, 87, 425–435.

Rausch, D. M., and Davis, M. R. (2001), "HIV in the CNS: Pathogenic Relationships to Systemic HIV Disease and Other CNS Diseases," *Journal of Neurovirology*, 7, 85–96.

Rotnitzky, A., and Jewell, N. P. (1990), "Hypothesis Testing of Regression Parameters in Semiparametric Generalized Linear Models for Cluster Data," *Biometrika*, 77, 485–489.

Schwartzman, A. (2006), "Random Ellipsoids and False Discovery Rates: Statistics for Diffusion Tensor Imaging Data," Ph.D. thesis, Stanford University.

Vernooij, M. W., de Groot, M., van der Lugt, A., Ikram, M. A., Krestin, G. P., Hofman, A., Niessen, W. J., and Breteler, M. M. B. (2008), "White Matter Atrophy and Lesion Formation Explain the Loss of Structural Integrity of White Matter in Aging," *NeuroImage*, 43, 470–477.

Whitcher, B., Wisco, J. J., Hadjikhani, N., and Tuch, D. S. (2007), "Statistical Group Comparison of Diffusion Tensors via Multivariate Hypothesis Testing," *Magnetic Resonance in Medicine*, 57, 1065–1074.

Xu, D., Mori, S., Shen, D., van Zijl, P., and Davatzikos, C. (2003), "Spatial Normalization of Diffusion Tensor Fields," *Magnetic Resonance in Medicine*, 50, 175–182.

Zhu, H. T., Zhang, H. P., Ibrahim, J. G., and Peterson, B. G. (2007), "Statistical Analysis of Diffusion Tensors in Diffusion-Weighted Magnetic Resonance Image Data" (with discussion), *Journal of the American Statistical Association*, 102, 1085–1102.