Local Polynomial Regression for Symmetric Positive Definite Matrices

Ying Yuan, Hongtu Zhu †, Weili Lin, J. S. Marron

University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA.

Summary. Local polynomial regression has received extensive attention for the nonparametric estimation of regression functions when both the response and the covariate are in Euclidean space. However, little has been done when the response is in a Riemannian manifold. We develop an intrinsic local polynomial regression estimate for the analysis of symmetric positive definite (SPD) matrices as responses that lie in a Riemannian manifold with covariate in Euclidean space. The primary motivation and application of the proposed methodology is in computer vision and medical imaging. We examine two commonly used metrics, including the trace metric and the Log-Euclidean metric on the space of SPD matrices. For each metric, we develop a cross-validation bandwidth selection method, derive the asymptotic bias, variance, and normality of the intrinsic local constant and local linear estimators, and compare their asymptotic mean square errors. Simulation studies are further used to compare the estimators under the two metrics and to examine their finite sample performance. We use our method to detect diagnostic differences between diffusion tensors along fiber tracts in a study of human immunodeficiency virus.

1. Introduction

Symmetric positive-definite (SPD) matrix-valued data occur in a wide variety of important applications. For instance, in computational anatomy, a SPD deformation vector $(JJ^T)^{1/2}$ is computed to capture the directional information of shape change decoded in the Jacobian matrices J at each location in an image (Grenander and Miller, 2007). In diffusion tensor imaging (Basser et al., 1994), a 3×3 SPD diffusion tensor, which tracks the effective diffusion of water molecules, is estimated at each voxel (a 3 dimensional (3D) pixel) of an imaging space. In functional magnetic resonance imaging, a SPD covariance matrix is calculated to delineate

functional connectivity between different neural assemblies involved in achieving a complex cognitive task or perceptual process (Fingelkurts et al., 2005). In classical multivariate statistics, a common research focus is to model and estimate SPD covariance matrices for multivariate measurements, longitudinal data, and time series data among many others (Pourahmadi, 2000; Anderson, 2003).

Despite the popularity of SPD matrix-valued data, only a handful of methods have been developed for the statistical analysis of SPD matrices as response variables in a Riemannian manifold. In the medical imaging literature (Fletcher and Joshi, 2007; Batchelor et al., 2005; Pennec et al., 2006), various image processing methods have recently been developed to segment, deform, interpolate, extrapolate and regularize diffusion tensor images (DTIs). Schwartzman (2006) proposed several parametric models for analyzing SPD matrices and derived the distributions of several test statistics for comparing differences between the means of the two (or multiple) groups of SPD matrices. Kim and Richards (2010) developed a nonparametric estimator for the common density function of a random sample of positive definite matrices. Zhu et al. (2009) developed a semi-parametric regression model with SPD matrices as responses in a Riemannian manifold and the covariates in a Euclidean space. Barmpoutis et al. (2007) and Davis et al. (2010) proposed tensor splines and local constant regressions for interpolating DTI tensor fields, but they did not address several important issues of analyzing random SPD matrices including the asymptotic properties of the nonparametric estimate proposed. All these methods for SPD matrices discussed above are based on the trace metric (or affine invariant metric) in the SPD space (Lang, 1999; Terras, 1988). Recently, Arsigny et al. (2007) proposed a Log-Euclidean metric and showed its excellent theoretical and computational properties. Dryden et al. (2009) compared various metrics of the space of SPD matrices and their properties.

To the best of our knowledge, this is the first paper to develop an intrinsic local polynomial regression (ILPR) model for estimating an intrinsic *conditional expectation* of a SPD matrix response, S, given a covariate vector x from a set of observations $(x_1, S_1), \cdots, (x_n, S_n)$, where the x_i can be either univariate or multivariate. In practice, x can be the arc-length of a specific fiber tract (e.g., right internal capsule tract), the coordinates in the 3D imaging space, or demographic variables such as age. Important applications of ILPR include smoothing diffusion tensors along fiber tracts and smoothing diffusion and deformation tensor fields. Another application is quantifying the change of diffusion and deformation tensors as well as the inter-regional functional connectivity matrix across groups and over time.

Relative to the existing literature on the analysis of SPD matrices, we make several impor-

tant contributions in this paper.

- To account for the curved nature of the SPD space, we propose the ILPR method for estimating the intrinsic conditional expectation of random SPD responses given the co-variate. We also derive an approximation of a cross-validation method for bandwidth selection.

- Theoretically, we compare the trace metric and the Log-Euclidean metric and establish the asymptotic properties of the ILPR estimators corresponding to each metric.

- Theoretically and numerically, we examine the effect that the use of different metrics has on statistical inference in the SPD space.

The rest of the paper is organized as follows. In Section 2, we develop the ILPR method and a cross-validated bandwidth method for nonparametric analysis of random SPD matrix-valued data. In Section 3, we compare the trace metric and the Log-Euclidean metric and derive their ILPR estimators. We investigate the asymptotic properties of the estimators proposed under the Log-Euclidean metric and the estimators under the trace metric in Sections 4.1 and 4.2, respectively. We examine the finite sample performance of the ILPR estimators via simulation studies in Section 5. We analyze a real data set to illustrate a real-world application of the proposed ILPR method in Section 6 before offering some concluding remarks in Section 7.

2. Intrinsic Local Polynomial Regression for SPD Matrices

In this section, we develop a general framework for using intrinsic local polynomial regression in the analysis of SPD matrices and will examine two examples in Section 3. Let $\mathrm{Sym}^+(m)$ and $\mathrm{Sym}(m)$ be, respectively, the set of $m \times m$ SPD matrices and the set of $m \times m$ symmetric matrices with real entries. The space $\mathrm{Sym}(m)$ is a Euclidean space with the Frobenius metric (or Euclidean inner product) given by $\mathrm{tr}(A_1 A_2)$ for any $A_1, A_2 \in \mathrm{Sym}(m)$, whereas $\mathrm{Sym}^+(m)$ is a Riemannian manifold, which will be detailed below. There is a one-to-one correspondence between $\mathrm{Sym}(m)$ and $\mathrm{Sym}^+(m)$ through matrix exponential and logarithm. For any matrix $A \in \mathrm{Sym}(m)$, its matrix exponential is given by $\exp(A) = \sum_{k=0}^{\infty} A^k/k! \in \mathrm{Sym}^+(m)$. Conversely, for any matrix $S \in \mathrm{Sym}^+(m)$, there is a $\log(S) = A \in \mathrm{Sym}(m)$ such that $\exp(A) = S$.

Standard nonparametric regression models for responses in the Euclidean space estimate $E(S|X = x)$. However, for a random S in a curved space, one cannot directly define the conditional expectation of S given $X = x$ with the usual expectation in Euclidean space. We are interested in answering the following question.

(Q1) How do we define an intrinsic conditional expectation of S at each x, denoted by $D(x)$, in $\text{Sym}^+(m)$?

To appropriately define $D(x)$, we review some basic facts about the geometrical structure of $\text{Sym}^+(m)$ near $D(x)$ (Lang, 1999; Terras, 1988). See Figure 1 for a graphical illustration. We first introduce the tangent vector and tangent space at $D(x)$ in $\text{Sym}^+(m)$. For a small scalar $\delta > 0$, let $C(t)$ be a differentiable map from $(-\delta, \delta)$ to $\text{Sym}^+(m)$ passing through $C(0) = D(x)$. A tangent vector at $D(x)$ is defined as the derivative of the smooth curve $C(t)$ with respect to t evaluated at $t = 0$. The set of all tangent vectors at $D(x)$ forms the tangent space of $\text{Sym}^+(m)$ at $D(x)$, denoted as $T_{D(x)}\text{Sym}^+(m)$, which can be identified with $\text{Sym}(m)$. The $T_{D(x)}\text{Sym}^+(m)$ is equipped with an inner product $\langle \cdot, \cdot \rangle$, called a Riemannian metric, which varies smoothly from point to point. For instance, one may use the Frobenuis metric as a Riemannian metric. Two additional Riemannian metrics for $\text{Sym}^+(m)$ will be given in Section 3. For a given Riemannian metric, we can calculate $\langle U, V \rangle$ for any U and V on $T_{D(x)}\text{Sym}^+(m)$ and then we can calculate the length of a smooth curve $C(t) : [t_0, t_1] \to \text{Sym}^+(m)$, which equals $\int_{t_0}^{t_1} \sqrt{\langle \dot{C}(t), \dot{C}(t) \rangle} dt$, where $\dot{C}(t)$ is the derivative of $C(t)$ with respect to t. A geodesic is a smooth curve on $\text{Sym}^+(m)$ whose tangent vector does not change in length or direction as one moves along the curve. For a $U \in T_{D(x)}\text{Sym}^+(m)$, there is a unique geodesic, denoted by $\gamma_{D(x)}(t; U)$, whose domain contains $[0, 1]$, such that $\gamma_{D(x)}(0; U) = D(x)$ and $\dot{\gamma}_{D(x)}(0; U) = U$. The Riemannian exponential mapping $\text{Exp}_{D(x)} : T_{D(x)}\text{Sym}^+(m) \to \text{Sym}^+(m)$ of the tangent vector U is defined as $\text{Exp}_{D(x)}(U) = \gamma_{D(x)}(1; U)$. The inverse of the Riemannian exponential map $\text{Log}_{D(x)}(\cdot) = \text{Exp}_{D(x)}^{-1}(\cdot)$ is called the Riemannian logarithmic map from $\text{Sym}^+(m)$ to a vector in $T_{D(x)}\text{Sym}^+(m)$. Finally, the shortest distance between two points $D_1(x)$ and $D_2(x)$ in $\text{Sym}^+(m)$ is called the geodesic distance between $D_1(x)$ and $D_2(x)$, denoted as $g(D_1(x), D_2(x))$, which satisfies

$$g(D_1(x), D_2(x))^2 = \langle \text{Log}_{D_1(x)}(D_2(x)), \text{Log}_{D_1(x)}(D_2(x)) \rangle. \tag{1}$$

We define $\mathcal{E}_D(X)$ to be $\text{Log}_{D(X)}(S)$ in $T_{D(X)}\text{Sym}^+(m)$. Statistically, $\mathcal{E}_D(X)$ can be regarded as the residual of S relative to $D(X)$. Let $\text{vecs}(C) = (c_{11}, c_{21}, c_{22}, \cdots, c_{m1}, \cdots, c_{mm})^T$ be an $m(m+1)/2 \times 1$ vector for any $m \times m$ symmetric matrix $C = (c_{ij})$. Thus, the intrinsic conditional expectation of S at $X = x$ is defined as $D(x) \in \text{Sym}^+(m)$ such that

$$E\{\text{Log}_{D(X)}(S)|X = x\} = O_m, \tag{2}$$

where O_m is the $m \times m$ matrix with all elements zeros and the expectation is taken component-wise with respect to the multivariate random vector $\text{vecs}(\text{Log}_{D(x)}(S))$. In fact, (2) characterizes intrinsic means (Bhattacharya and Patrangenaru, 2005).
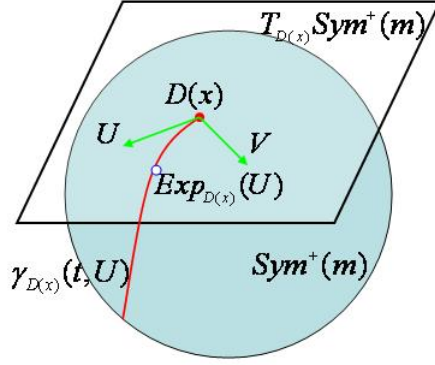
Fig. 1. Graphical illustration of the geometrical structure of $\text{Sym}^+(m)$ near $D(x)$.

Suppose that $(x_i, S_i), i = 1, \cdots, n$ is an independent and identically distributed random sample, where $S_i \in \text{Sym}^+(m)$. For notational simplicity, we focus on a univariate covariate throughout the paper. We are interested in using the observed data $\{(x_i, S_i), i = 1, \cdots, n\}$ to estimate $D(X)$ defined in (2) at each $X = x_0$. By ignoring the Riemannian metric introduced in $T_{D(X)}\text{Sym}^+(m)$, we can directly minimize a weighted least square criterion based on the metric related to the regular Frobenius inner product, which is given by

$$L_n(D(x_0)) = \sum_{i=1}^{n} K_h(x_i - x_0)\text{tr}(\text{Log}_{D(x_0)}(S_i)^2). \tag{3}$$

In (3), $K_h(u) = K(u/h)h^{-1}$, in which h is a positive scalar, and $K(\cdot)$ is a kernel function such as the Epanechnikov kernel (Fan and Gijbels, 1996; Wand and Jones, 1995). However, it is unclear whether the estimate, which minimizes $L_n(D(x_0))$, is truly consistent or not. Therefore, we are interested in solving the second question below.

(Q2) How do we use the observed data to consistently estimate $D(X)$ in (2) at each $X = x_0$?

For a specific Riemannian metric, we consider estimating $D(X)$ at $X = x_0$ by minimizing a weighted intrinsic least square criterion, denoted by $G_n(D(x_0))$ given by

$$\sum_{i=1}^{n} K_h(x_i - x_0)\langle \text{Log}_{D(x_0)}(S_i), \text{Log}_{D(x_0)}(S_i)\rangle = \sum_{i=1}^{n} K_h(x_i - x_0)g(D(x_0), S_i)^2. \tag{4}$$

Directly minimizing $G_n(D(x_0))$ with respect to $D(x_0)$ leads to a weighted intrinsic mean of $S_1, \cdots, S_n \in \text{Sym}^+(m)$ at x_0, denoted by $\hat{D}_I(x_0)$ (Bhattacharya and Patrangenaru, 2005). It will be shown below that $\hat{D}_I(x_0)$ is truly a consistent estimate of $D(x_0)$.

Local polynomial regression has received extensive attention for the nonparametric estimation of regression functions when both response and covariate are in Euclidean space Fan and Gijbels (1996); Wand and Jones (1995). However, little has been done on developing local polynomial regression when the response is in a Riemannian manifold and the covariates are in Euclidean space. Therefore, we are interested in solving a third question below.

(Q3) How do we define the intrinsic local polynomial regression for estimating $D(X)$ in (2) at each $X = x_0$?

We propose the intrinsic local polynomial regression for estimating $D(X)$ at $X = x_0$ as follows. Since $D(x)$ is in the curved space, we cannot directly expand $D(x)$ at x_0 by using a Taylor's series expansion. Instead, we consider the Riemannian logarithmic map of $D(x)$ at $D(x_0)$ in $T_{D(x_0)}\mathrm{Sym}^+(m)$. Let I_m be an $m \times m$ identity matrix. Since $\mathrm{Log}_{D(x_0)}(D(x))$ for different x_0 are in different tangent spaces, we may transport them from $T_{D(x_0)}\mathrm{Sym}^+(m)$ to the same tangent space $T_{I_m}\mathrm{Sym}^+(m)$ through a parallel transport given by

$$\phi_{D(x_0)} : T_{D(x_0)}\mathrm{Sym}^+(m) \to T_{I_m}\mathrm{Sym}^+(m).$$

That is, we have

$$Y(x) = \phi_{D(x_0)}(\mathrm{Log}_{D(x_0)}(D(x))) \in T_{I_m}\mathrm{Sym}^+(m) \text{ and } \mathrm{Log}_{D(x_0)}(D(x)) = \phi_{D(x_0)}^{-1}(Y(x)), \quad (5)$$

where $\phi_{D(x_0)}^{-1}(\cdot)$ is the inverse map of $\phi_{D(x_0)}(\cdot)$. Moreover, since $Y(x_0) = \phi_{D(x_0)}(O_m) = O_m$ and $Y(x)$ are in the same space $T_{I_m}\mathrm{Sym}^+(m)$, we expand $Y(x)$ at x_0 by using the Taylor's series expansion as follows:

$$\mathrm{Log}_{D(x_0)}(D(x)) = \phi_{D(x_0)}^{-1}(Y(x)) \approx \phi_{D(x_0)}^{-1}(\sum_{k=1}^{k_0} Y^{(k)}(x_0)(x - x_0)^k), \quad (6)$$

where k_0 is an integer and $Y^{(k)}(x)$ is the $k-$th derivative of $Y(x)$ with respect to x divided by $k!$. Equivalently, $D(x)$ can be approximated by

$$D(x) \approx \mathrm{Exp}_{D(x_0)}(\phi_{D(x_0)}^{-1}(\sum_{k=1}^{k_0} Y^{(k)}(x_0)(x - x_0)^k)) = D(x, \alpha(x_0), k_0), \quad (7)$$

where $\alpha(x_0)$ contains all unknown parameters in $\{D(x_0), Y^{(1)}(x_0), \cdots, Y^{(k_0)}(x_0)\}$.

To estimate $\alpha(x_0)$, we substitute the approximation of $D(x)$ in (7) into (4) to obtain $G_n(\alpha(x_0))$, which is given by

$$G_n(\alpha(x_0)) = \sum_{i=1}^{n} K_h(x_i - x_0)g(\mathrm{Exp}_{D(x_0)}(\phi_{D(x_0)}^{-1}(\sum_{k=1}^{k_0} Y^{(k)}(x_0)(x - x_0)^k)), S_i)^2. \quad (8)$$

Subsequently, we calculate an intrinsic weighted least square estimator of $\alpha(x_0)$ defined by

$$\hat{\alpha}_I(x_0; h) = \text{argmin}_{\alpha(x_0)} G_n(\alpha(x_0)). \tag{9}$$

Then we can calculate $D(x, \hat{\alpha}_I(x_0; h), k_0)$, denoted by $\hat{D}_I(x, h)$, as an intrinsic local polynomial regression estimator (ILPRE) of $D(x)$. When $k_0 = 0$, $D(x, \hat{\alpha}_I(x_0; h), 0)$ is exactly the intrinsic local constant estimator of $D(x_0)$ considered in Davis et al. (2010).

We propose using a leave-one-out cross validation method for bandwidth selection due to its conceptual simplicity. Let $\hat{D}_I^{(-i)}(x_i; h)$ be the estimate of $D(x_i)$ obtained by minimizing $G_n(\alpha(x_i))$ with (x_i, S_i) deleted for a given bandwidth h and all i. The cross-validation score is defined as follows:

$$\text{CV}(h) = n^{-1} \sum_{i=1}^{n} g(S_i, \hat{D}_I^{(-i)}(x_i; h))^2. \tag{10}$$

The optimal h, denoted by $\hat{h}$, can be obtained by minimizing $\text{CV}(h)$. However, since computing $\hat{D}_I^{(-i)}(x_i; h)$ for all i can be computationally prohibitive, we suggest to use the first-order approximation of $\text{CV}(h)$, whose details will be given below under each specific metric. Although it is possible to develop other bandwidth selection methods, such as plug-in and bootstrap methods (Rice, 1984; Park and Marron, 1990; Hall et al., 1992; Hardle et al., 1992), we must deal with additional computational and theoretical challenges, which will be left for future research.

3.   ILPR under Log-Euclidean Metric and Trace Metric

As discussed in Dryden et al. (2009), various metrics can be defined for tangent vectors on $T_{D(x)}\text{Sym}^+(m)$. To assess the effect of different metrics on ILPREs, we develop ILPR under two commonly used metrics, including the Log-Euclidean metric and the trace metric.

3.1.   Log-Euclidean Metric

In this section, we review some basic facts about the theory of the Log-Euclidean metric, details of which have been given in Arsigny et al. (2007). We introduce the notation 'L' into some necessary quantities under the Log-Euclidean metric. We use $\exp(.)$ and $\log(.)$ to represent the matrix exponential and the matrix logarithm, respectively, whereas we use Exp and Log to represent the Riemannian exponential and logarithm maps, respectively. Let $\partial_{D(x)} \log .(U)$ be the differential of the matrix logarithm at $D(x) \in \text{Sym}^+(m)$ acting on an infinitesimal displacement $U \in T_{D(x)}\text{Sym}^+(m)$ (Arsigny et al., 2007). The Log-Euclidean metric

on $\text{Sym}^+(m)$ is defined as

$$\langle U, V \rangle = \text{tr}(\{\partial_{D(x)} \log .(U)\}\{\partial_{D(x)} \log .(V)\}), \tag{11}$$

where U and V are in $T_{D(x)}\text{Sym}^+(m)$. The geodesic $\gamma_{D(x),L}(t, U)$ is given by $\exp(\log(D(x)) + t\partial_{D(x)} \log .(U))$ for any $t \in R$. Let $\partial_{\log(D(x))} \exp .(A)$ be the differential of the matrix exponential at $\log(D(x)) \in \text{Sym}(m)$ acting on an infinitesimal displacement $A \in T_{\log(D(x))}\text{Sym}(m)$ (Arsigny et al., 2007). The Riemannian exponential and logarithm maps are, respectively, given by

$$\text{Exp}_{D(x),L}(U) = \exp(\log(D(x)) + \partial_{D(x)} \log .(U)), \tag{12}$$
$$\text{Log}_{D(x),L}(S) = \partial_{\log(D(x))} \exp .(\log(S) - \log(D(x))).$$

The geodesic distance between $D(x)$ and S is uniquely given by

$$g_L(D(x), S) = \sqrt{tr[\{\log(D(x)) - \log(S)\}^{\otimes 2}]}. \tag{13}$$

We consider two SPD matrices $D(x)$ and $D(x_0)$. For any $U_{D(x_0)} \in T_{D(x_0)}Sym^+(m)$, the parallel transport $\phi_{D(x_0),L} : T_{D(x_0)}Sym^+(m) \to T_{I_m}\text{Sym}^+(m)$ is defined by

$$\phi_{D(x_0),L}(U_{D(x_0)}) = \partial_{D(x_0)} \log .(U_{D(x_0)}) \in T_{I_m}\text{Sym}^+(m). \tag{14}$$

Combining (12) and (14) yields

$$Y(x) = \phi_{D(x_0),L}(\text{Log}_{D(x_0),L}(D(x))) = \log(D(x)) - \log(D(x_0)),$$
$$D(x) = \exp(\log(D(x_0)) + Y(x)). \tag{15}$$

In this case, $\mathcal{E}_D(X) = \log(S) - \log(D(X))$ and $E\{\log(S)|X = x\} = \log(D(x))$.

Let $\text{vec}(A) = (a_{11}, ..., a_{1m}, a_{21}, ..., a_{2m}, \cdots, a_{m1}, \cdots, a_{mm})^T$ be the vectorization of an $m \times m$ matrix $A = (a_{ij})$. Under the Log-Euclidean metric, $G_n(D(x_0))$ in (4) can be written as

$$G_n(D(x_0)) = \sum_{i=1}^{n} K_h(x_i - x_0)\text{tr}[\{\log(D(x_0)) - \log(S_i)\}^2]. \tag{16}$$

To compute the ILPR estimator, we use the Taylor's series expansion to expand $\log(D(x))$ at x_0 as follows:

$$\log(D(x)) \approx \sum_{k=0}^{k_0} \log(D(x_0))^{(k)}(x - x_0)^k = \log(D_L(x, \alpha_L(x_0), k_0)), \tag{17}$$

where $\alpha_L(x_0)$ contains all unknown parameters in $\log(D(x_0))^{(k)}$ for $k = 0, \cdots, k_0$. We compute $\hat{\alpha}_{IL}(x_0; h)$ by minimizing $G_n(D_L(x, \alpha_L(x_0), k_0))$. It can be shown that $\hat{\alpha}_{IL}(x_0; h)$ has the explicit expression as

$$\hat{\alpha}_{IL}(x_0; h) = \text{vec}(\{\sum_{i=1}^{n} K_h(x_i - x_0)\mathcal{X}_i(x_0)^{\otimes 2}\}^{-1} \sum_{i=1}^{n} K_h(x_i - x_0)\mathcal{X}_i(x_0)\text{vecs}(\log(S_i))^T), \tag{18}$$

where $\mathcal{X}_i(x) = (1, (x_i - x), \cdots, (x_i - x)^{k_0})^T$. By substituting $\hat{\alpha}_{IL}(x_0; h)$ into $D_L(x, \alpha_L(x_0), k_0)$, we have $\hat{D}_{IL}(x; h, k_0) = D_L(x, \hat{\alpha}_{IL}(x_0; h), k_0)$.

Let $e_{k_0+1,i}$ be the $(k_0 + 1)$ unit vector having 1 in the i-th entry and 0 elsewhere. Let $e_{k_0+1,i}^T \{\sum_{j=1}^n K_h(x_j - x)\mathcal{X}_j(x)^{\otimes 2}\}^{-1} K_h(x_i - x)\mathcal{X}_i(x) = a_i(x)$. The cross-validation score CV(h) can be simplified as follows:

$$\mathrm{CV}(h) = n^{-1} \sum_{i=1}^n g_L(S_i, \hat{D}_{IL}(x_i; h))^2 / \{1 - a_i(x_i)\}^2. \tag{19}$$

Replacing $a_i(x_i)$ in (19) by the average of $a_1(x_1), \cdots, a_n(x_n)$, we can get the generalized cross-validation (GCV) score as follows:

$$\mathrm{GCV}(h) = n^{-1} \sum_{i=1}^n g_L(S_i, \hat{D}_{IL}(x_i; h))^2 / \{1 - \sum_{i=1}^n a_i(x_i)/n\}^2. \tag{20}$$

Without special saying, for the Log-Euclidean metric, we use GCV(h) to select the bandwidth throughout this paper.

3.2. Trace Metric

We review some basic facts about the theory of trace metric (Schwartzman, 2006; Lang, 1999; Terras, 1988; Fletcher et al., 2004; Batchelor et al., 2005; Pennec et al., 2006). We add the notation of 'T' into some necessary geometric quantities under the trace metric. Under the trace metric, an inner product of U and V in $T_{D(x)}\mathrm{Sym}^+(m)$ is defined as

$$\langle U, V \rangle = \mathrm{tr}(UD(x)^{-1}VD(x)^{-1}). \tag{21}$$

The geodesic $\gamma_{D(x),T}(t; U)$ is given by $G(x)\exp(tG(x)^{-1}UG(x)^{-T})G(x)^T$ for any t, where $G(x)$ is any square root of $D(x)$ such that $D(x) = G(x)G(x)^T$. The Riemannian exponential and logarithm maps are, respectively, given by

$$\mathrm{Exp}_{D(x),T}(U) = \gamma_{D(x),T}(1; U) = G(x)\exp(G(x)^{-1}UG(x)^{-T})G(x)^T,$$
$$\mathrm{Log}_{D(x),T}(S) = G(x)\log(G(x)^{-1}SG(x)^{-T})G(x)^T. \tag{22}$$

The geodesic distance between $D(x)$ and S, denoted by $g_T(D(x), S)$, is given by

$$\sqrt{\mathrm{tr}\{\log^2(G(x)^{-1}SG(x)^{-T})\}} = \sqrt{\mathrm{tr}\{\log^2(S^{-1/2}D(x)S^{-T/2})\}}, \tag{23}$$

where $S^{1/2}$ is any square root of S.

We consider two SPD matrices $D(x)$ and $D(x_0) = G(x_0)G(x_0)^T$. For any $U_{D(x_0)} \in T_{D(x_0)}\mathrm{Sym}^+(m)$, the parallel transport $\phi_{D(x_0),T}$ is defined by

$$\phi_{D(x_0),T}(U_{D(x_0)}) = G(x_0)^{-1}U_{D(x_0)}G(x_0)^{-T} \in T_{I_m}\mathrm{Sym}^+(m). \tag{24}$$

Thus, combining (22) and (24) yields

$$Y(x) = \phi_{D(x_0),T}(\text{Log}_{D(x_0),T}(D(x))) = \log(G(x_0)^{-1}D(x)G(x_0)^{-T}),$$
$$D(x) = G(x_0)\exp(Y(x))G(x_0)^T. \tag{25}$$

In this case, $\mathcal{E}_D(X) = \log(G(X)^{-1}SG(X)^{-T})$.

To compute the ILPR estimator, we use the Taylor's series expansion to expand $Y(x)$ at x_0 as follows:

$$D(x) \approx G(x_0)\exp(\sum_{k=1}^{k_0} Y^{(k)}(x_0)(x-x_0)^k)G(x_0)^T = D_T(x, \alpha_T(x_0), k_0), \tag{26}$$

where $\alpha_T(x_0)$ contains all unknown parameters in $G(x_0)$ and $Y^{(k)}(x_0)$ for $k = 1, \cdots, k_0$. Thus, we can compute $\hat{\alpha}_{IT}(x_0; h)$ by minimizing $G_n(\alpha_T(x_0))$. Under the trace metric, minimizing $G_n(\alpha_T(x_0))$ is computationally challenging when $k_0 > 0$, since $G_n(\alpha_T(x_0))$ is not convex and may have multiple local minimizers. Thus, standard gradient methods, which strongly depend on the starting value of $\alpha_T(x_0)$, do not perform well for optimizing $G_n(\alpha_T(x_0))$ when $k_0 > 0$. Hence, we develop an annealing evolutionary stochastic approximation Monte Carlo algorithm (see Liang (2011) for good discussion) for computing $\hat{\alpha}_{IT}(x_0; h)$. Details can be found in the supplementary document.

To simplify the computation of $\text{CV}_T(h)$, we suggest the first-order approximation to $\text{CV}_T(h)$ as follows:

$$\text{CV}_T(h) \approx n^{-1}\sum_{i=1}^n g_T(S_i, \hat{D}_{IT}(x_i; h, k_0))^2 + 2p_n(h), \tag{27}$$

where $\hat{D}_{IT}(x; h, k_0) = D_T(x, \hat{\alpha}_{IT}(x_0; h), k_0)$. The $\text{CV}_T(h)$ is close to Akaike's information criterion (AIC) (Sakamoto et al., 1999) and $p_n(h)$ can be regarded as the number of degrees of freedom. The explicit form of $p_n(h)$ is presented in the supplementary document.

4. Asymptotic Properties

We derive the asymptotic properties of ILPREs, such as asymptotic normality, under the Log-Euclidean and trace metrics. Furthermore, we systematically compare the intrinsic local constant and linear estimators under each metric and between the two metrics.

4.1. *Log-Euclidean Metric*

Under the Log-Euclidean metric, ILPRE is almost equivalent to the LPR estimator for multi-variate response in Euclidean space. Thus, we can generalize the existing theory of the local

polynomial regression estimator (Fan and Gijbels, 1996; Wand and Jones, 1995). Moreover, we only present the consistency and asymptotic normality of ILPRE for interior points, since the asymptotic properties of ILPRE for boundary points are similar to those for interior points in Euclidean space (Fan and Gijbels, 1996).

To proceed, we need some additional notation. Let $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$ for any vector or matrix $\mathbf{a}$ and I_q be an identity matrix of size $q = m(m+1)/2$. Let $\mathcal{H} = diag(1, h, \cdots, h^{k_0}) \otimes I_q$. Let $\mathbf{u} = (u_1, \cdots, u_{k_0})^T$ and $\mathbf{v} = (v_1, \cdots, v_{k_0})^T$ be $k_0 \times 1$ vectors, where $u_k = \int x^k K(x) dx$ and $v_k = \int x^k K(x)^2 dx$ for $k \geq 0$. Let $\mathcal{U}_0 = (u_{i+j})$ and $\mathcal{V}_0 = (v_{i+j})$ for $0 \leq i, j \leq k_0$ be two $(k_0 + 1) \times (k_0 + 1)$ matrices for $0 \leq i, j \leq k_0$. Let $f_X(x)$ and $f_X^{(1)}(x)$ be the marginal density function of X and its first-order derivative with respect to x, respectively. We define $\mathcal{M}(x_0; h) = (\mathcal{M}_1(x_0; h)^T, \cdots, \mathcal{M}_{k_0+1}(x_0; h)^T)^T$, in which we have

$$\mathcal{M}_k(x_0; h) = \begin{cases} u_{k_0+k} \text{vecs}(\log\{D(x_0)\}^{(k_0+1)}), \\ \qquad \text{for even} \quad k_0 + k \text{ as } 0 < k \leq k_0 + 1; \\ h u_{k_0+k+1} \text{vecs}(\log\{D(x_0)\}^{(k_0+1)} \log(f_X(x_0))^{(1)} + \log\{D(x_0)\}^{(k_0+2)}(k_0 + 2)^{-1}), \\ \qquad \text{for odd} \quad k_0 + k. \end{cases}$$

We have the following results, whose proof is similar to that of Theorem 2 in the supplementary document.

Theorem 1. Suppose that x_0 is an interior point of $f_X(.)$. Under the Log-Euclidean metric and conditions (C1)-(C4) in the appendix, we have the following results.

(i) $\mathcal{H}\{\hat{\alpha}_{IL}(x_0; h) - \alpha_L(x_0)\}$ converges to $\mathbf{0}$ in probability as $n \to \infty$.

(ii) For $k_0 = 0$, under an additional condition (C10) in the appendix and that $f_X^{(1)}(x)$ is continuous in a neighborhood of x_0, we have

$$\sqrt{nh}[\mathcal{H}\{\hat{\alpha}_{IL}(x_0; h) - \alpha_L(x_0)\} - h^2 u_2 \text{vecs}(0.5 \log\{D(x_0)\}^{(2)} + \frac{f_X^{(1)}(x_0)}{f_X(x_0)} \log\{D(x_0)\}^{(1)})] \\ \to^L N\{\mathbf{0}, \Sigma_0(x_0)\}, \quad (28)$$

where $\Sigma_0(x_0) = f_X^{-1}(x_0) v_0 \Sigma_{\mathcal{E}_D}(x_0)$ with $\Sigma_{\mathcal{E}_D}(x) = Cov(\text{vecs}[\log(S) - \log\{D(x)\}]|X = x)$ and $\to^L$ denotes convergence in distribution.

(iii) For $k_0 > 0$, under the conditions of Theorem 1 (ii), we have

$$\sqrt{nh}[\mathcal{H}\{\hat{\alpha}_{IL}(x_0; h) - \alpha_L(x_0)\} - \frac{h^{k_0+1}}{(k_0 + 1)!}(\mathcal{U}_0^{-1} \otimes I_q)\mathcal{M}(x_0; h)] \to^L N\{\mathbf{0}, \Sigma(x_0)\}, \quad (29)$$

where $\Sigma(x_0) = f_X^{-1}(x_0)(\mathcal{U}_0^{-1}\mathcal{V}_0\mathcal{U}_0^{-1}) \otimes \Sigma_{\mathcal{E}_D}(x_0)$.

Theorem 1 delineates the asymptotic properties of $\hat{\alpha}_{IL}(x_0; h)$ for $k_0 \geq 0$, which covers the asymptotic properties of the intrinsic local constant and linear estimators of $D(x_0)$ as $k_0 = 0, 1$.

In particular, the asymptotic bias and variance of $\hat{D}_{IL}(x_0; h, 0)$ are closely related to those of the Nadaraya-Watson estimator when both response and covariate are in Euclidean space (Fan, 1992). Since $\text{vecs}(\log\{\hat{D}_{IL}(x_0; h, k_0)\})$ is a subvector of $\hat{\alpha}_{IL}(x_0; h)$, we calculate the asymptotic average mean squared error (AMSE) conditional on $\mathbf{x} = \{x_1, \ldots, x_n\}$ as

$$\text{AMSE}(\log\{\hat{D}_{IL}(x_0; h, k_0)\}) = E\{\text{tr}([\log\{\hat{D}_{IL}(x_0; h, k_0)\} - \log\{D(x_0)\}]^2)|\mathbf{x}\}.$$

Furthermore, for a given weight function $w(x)$, we may consider a constant bandwidth that minimizes the asymptotic average mean integrated squared error (AMISE) as

$$\text{AMISE}(\log\{\hat{D}_{IL}(.; h, k_0)\}) = \int \text{AMSE}(\log\{\hat{D}_{IL}(x; h, k_0)\})w(x)dx.$$

Finally, we can calculate the asymptotically optimal local bandwidth, denoted by $h_{opt,L}(x_0; k_0)$, for minimizing $\text{AMSE}(\log\{\hat{D}_{IL}(x_0; h, k_0)\})$ and the optimal bandwidth, denoted by $h_{opt,L}(k_0)$, for minimizing $\text{AMISE}(\log\{\hat{D}_{IL}(.; h, k_0)\})$.

By Theorem 1 (iii), $\text{AMSE}(\log\{\hat{D}_{IL}(x_0; h, 0)\})$ equals $v_0\{nhf_X(x_0)\}^{-1}\text{tr}\{\Sigma_{\mathcal{E}_D}(x_0)\} + h^4 u_2^2\text{tr}\{(\text{vecs}[0.5\log\{D(x_0)\}^{(2)} + f_X^{(1)}(x_0)f_X(x_0)^{-1}\log\{D(x_0)\}^{(1)}])^{\otimes 2}\}$. For the intrinsic local linear estimator, $\text{AMSE}(\log\{\hat{D}_{IL}(x_0; h, 1)\})$ is given by $0.25h^4 u_2^2\text{tr}\{(\text{vecs}[\log\{D(x_0)\}^{(2)}])^{\otimes 2}\} + v_0\{nhf_X(x_0)\}^{-1}\text{tr}\{\Sigma_{\mathcal{E}_D}(x_0)\}$. Intrinsic local constant and linear estimators have the same asymptotic covariance and their differences are concerned only with their biases. The local constant estimator has one more term $h^2 u_2 f_X^{(1)}(x_0)f_X(x_0)^{-1}\text{vecs}[\log\{D(x_0)\}^{(1)}]$, which depends on the marginal density $f_X(.)$. Subsequently, we can get the optimal bandwidths, whose detailed expression can be found in the supplementary document.

4.2. *Trace Metric*

Under the trace metric, since ILPRE is different from the LPR estimator for multivariate response in Euclidean space, we study the consistency and asymptotic normality of ILPRE for both interior and boundary points.

We need to introduce some notation for discussion. Consider a function

$$\psi(S, G, Y) = g_T(S, G\exp(Y)G^T)^2, \tag{30}$$

where G is an $m \times m$ lower triangle matrix, $S \in \text{Sym}^+(m)$, and $Y \in \text{Sym}(m)$. Let $\alpha = (\alpha_G, \alpha_Y)$, in which $\alpha_G = \text{vecs}(G)$ and $\alpha_Y = \text{vecs}(Y)$. Let $\partial_\alpha\psi(S, G, Y)$ and $\partial_\alpha^2\psi(S, G, Y)$ be the first and second order derivatives of $\psi(S, G, Y)$ with respect to α, respectively. By substituting $Y(X)$

into $\partial_\alpha \psi(S, G, Y)$ and $\partial_\alpha^2 \psi(S, G, Y)$ and using the decomposition of $\alpha = (\alpha_G, \alpha_Y)$, we define

$$
\begin{pmatrix} \Psi_1(x) & \Psi_2(x) \\ \Psi_2(x)^T & \Psi_3(x) \end{pmatrix} = E\{\partial_\alpha^2 \psi(S, G, Y(X))|X = x\},
$$

$$
\begin{pmatrix} \Psi_{11}(x) & \Psi_{12}(x) \\ \Psi_{12}(x)^T & \Psi_{22}(x) \end{pmatrix} = E[\{\partial_\alpha \psi(S, G, Y(X))\}^{\otimes 2}|X = x],
$$

where the expectation is taken with respect to S given $X = x$. Let $\mathbf{1}_{k_0}$ be a $k_0 \times 1$ column vector with all elements ones. Let $\mathcal{U}_2 = (u_{i+j})$ and $\mathcal{V}_2 = (v_{i+j})$ for $1 \leq i, j \leq k_0$ be two $k_0 \times k_0$ matrices. We define

$$
\aleph(x_0; h) = (w_1(x_0; h)^T \Psi_2(x_0), \mathbf{w}(x_0; h)\{\mathbf{1}_{k_0} \otimes \Psi_3(x_0)\})^T
$$

and $\mathbf{w}(x_0; h) = (w_2(x_0; h)^T, \cdots, w_{k_0+1}(x_0; h)^T)$, in which

$$
w_k(x_0; h) = \begin{cases} u_{k_0+k}\mathrm{vecs}(Y^{(k_0+1)}(x_0)) & \text{for even } k_0 + k \text{ as } 0 < k \leq k_0 + 1, \\ hu_{k_0+k+1}\mathrm{vecs}(Y^{(k_0+1)}(x_0) \log(f_X(x_0))^{(1)} + Y^{(k_0+2)}(x_0)(k_0 + 2)^{-1}) & \text{for odd } k_0 + k. \end{cases}
$$

Finally, let $\alpha_T(x) = (\mathrm{vecs}\{G(x)\}^T, \mathrm{vecs}\{Y^{(1)}(x)\}^T, \cdots, \mathrm{vecs}\{Y^{(k_0)}(x)\}^T)^T$.

Theorem 2. Suppose that x_0 is an interior point of $f_X(\cdot)$. Under the trace metric and conditions (C1)-(C8) in the appendix, we have the following results.

(i) There exist solutions $\hat{\alpha}_{IT}(x_0; h)$ to equation $\partial G_n(\alpha_T(x_0))/\partial \alpha_T(x_0) = \mathbf{0}$ such that $\mathcal{H}\{\hat{\alpha}_{IT}(x_0; h) - \alpha_T(x_0)\}$ converges to $\mathbf{0}$ in probability as $n \to \infty$.

(ii) For $k_0 = 0$, if $f_X^{(1)}(x)$ is continuous in a neighborhood of x_0, then we have

$$
\sqrt{nh}[\mathcal{H}\{\hat{\alpha}_{IT}(x_0; h) - \alpha_T(x_0)\} - h^2 u_2 \mathrm{vecs}\{G^{(1)}(x_0)\frac{f_X^{(1)}(x_0)}{f_X(x_0)} + 0.5G^{(2)}(x_0)\}] \to^L N\{\mathbf{0}, \Omega_0(x_0)\}, \quad (31)
$$

where $\Omega_0(x_0) = u_0^{-2} f_X^{-1}(x_0) v_0 \Psi_1(x_0)^{-1} \Psi_{11}(x_0) \Psi_1(x_0)^{-1}$.

(iii) For $k_0 > 0$, if condition (C9) in the appendix is also true, we have

$$
\sqrt{nh}[\mathcal{H}\{\hat{\alpha}_{IT}(x_0; h) - \alpha_T(x_0)\} - \frac{h^{k_0+1}}{(k_0 + 1)!}\mathcal{N}(x_0)^{-1}\aleph(x_0; h)] \to^L N\{\mathbf{0}, \Omega(x_0)\}, \quad (32)
$$

where $\Omega(x_0) = f_X^{-1}(x_0)\mathcal{N}(x_0)^{-1}\mathcal{N}^*(x_0)\mathcal{N}(x_0)^{-1}$ and $\mathcal{N}(x)$ and $\mathcal{N}^*(x)$ are, respectively, given by

$$
\mathcal{N}(x) = \begin{pmatrix} u_0\Psi_1(x) & \mathbf{u} \otimes \Psi_2(x) \\ \mathbf{u}^T \otimes \Psi_2(x)^T & \mathcal{U}_2 \otimes \Psi_3(x) \end{pmatrix}, \mathcal{N}^*(x) = \begin{pmatrix} v_0\Psi_{11}(x) & \mathbf{v} \otimes \Psi_{12}(x) \\ \mathbf{v}^T \otimes \Psi_{12}(x)^T & \mathcal{V}_2 \otimes \Psi_{22}(x) \end{pmatrix}.
$$

Theorem 2 delineates the asymptotic bias, covariance, and asymptotic normality of $\hat{\alpha}_{IT}(x_0; h)$ for $k_0 \geq 0$. Based on Theorem 2, it is straightforward to derive the asymptotic bias, covariance, and asymptotic normality of $\hat{D}_{IT}(x_0; h, k_0)$ for $k_0 \geq 0$. Moreover, to have a direct

comparison between the trace and Log-Euclidean metrics, we calculate the asymptotic biases and covariances of $\log\{\hat{D}_{IT}(x_0; h, k_0)\}$ under these two metrics. Subsequently, we calculate AMSE($\log\{\hat{D}_{IT}(x_0; k_0)\}$) and AMISE($\log\{\hat{D}_{IT}(.; h, k_0)\}$) for a given weight function $w(x)$. Minimizing AMSE($\log(\hat{D}_{IT}(x_0; h, k_0))$) and AMISE($\log(\hat{D}_{IT}(x_0; h, k_0))$), respectively, leads to the optimal bandwidths, whose detailed expressions can be found in the supplementary document.

We are interested in comparing the asymptotic properties of the intrinsic local constant $\hat{D}_{IT}(x_0; h, 0)$ and the local linear estimator $\hat{D}_{IT}(x_0; h, 1)$. It follows from the delta method that AMSE($\log\{\hat{D}_{IT}(x_0; h, 0)\}$) can be approximated as

$$h^4 u_2^2 \text{tr}([G_D(x_0)^T \text{vecs}\{G^{(1)}(x_0) f_X^{(1)}(x_0) f_X(x_0)^{-1} + 0.5 G^{(2)}(x_0)\}]^{\otimes 2})$$
$$+ (nh)^{-1} \text{tr}\{G_D(x_0)^{\otimes 2} \Omega_0(x_0)\} + o(h^4 + (nh)^{-1}), \tag{33}$$

where $G_D(x_0) = \{\partial \text{vec}(\log(G(x_0)^{\otimes 2}))/\partial \text{vecs}(G(x_0))^T\}^T$. The asymptotic bias and variance of $\hat{D}_{IT}(x_0; h, 0)$ are similar to those of the Nadaraya-Watson estimator when response is in Euclidean space (Fan, 1992). For the intrinsic local linear estimator, AMSE($\log(\hat{D}_{IT}(x_0; h, 1))$) equals $0.25 h^4 u_2^2 \text{tr}[\{G_D(x_0)^T \Psi_1(x_0)^{-1} \Psi_2^T(x_0) \text{vecs}(Y^{(2)}(x_0))\}^{\otimes 2}] + (nh)^{-1} \text{tr}\{G_D(x_0)^{\otimes 2} \Omega_0(x_0)\}$.

We consider ILPRE near the edge of the support of $f_X(x)$. Without loss of generality, we assume that the design density $f_X(.)$ has a bounded support $[0, 1]$ and consider the left-boundary point $x_0 = dh$ for some positive constant d. The asymptotic consistency and normality of IL-PRE are valid for the boundary points after slight modifications on the definitions of u_k and v_k. Denote $u_{k,d} = \int_{-d}^{\infty} x^k K(x) dx$ and $v_{k,d} = \int_{-d}^{\infty} x^k K^2(x) dx$. Correspondingly, $\mathbf{u}, \mathcal{U}_2, \mathcal{V}_2, \mathcal{U}_0$ and $\mathcal{V}_0$ are replaced by $\mathbf{u}_d, \mathcal{U}_{2,d}, \mathcal{V}_{2,d}, \mathcal{U}_{0,d}$ and $\mathcal{V}_{0,d}$, respectively. Let $\mathbf{c}_{k_0+2,d} = (u_{k_0+2,d}, \cdots, u_{2k_0+1,d})^T$ and $\aleph_d(0+) = (u_{k_0+1,d}\Psi_2(0+), \mathbf{c}_{k_0+2,d} \otimes \Psi_3(0+))^T \text{vecs}(Y^{(k_0+1)}(0+))$. For the boundary points, we have the following asymptotic results under the trace metric.

Theorem 3. Suppose that $x_0 = dh$ is a left boundary point of $f_X(.)$. Under the trace metric and conditions (C1)-(C8) in the appendix, we have the following results.

(i) There exist solutions, denoted by $\hat{\alpha}_{IT}(x_0; h)$, to the equation $\partial G_n(\alpha_T(x_0))/\partial \alpha_T(x_0) = \mathbf{0}$ such that $\mathcal{H}\{\hat{\alpha}_{IT}(x_0, h) - \alpha_T(x_0)\}$ converges to $\mathbf{0}$ in probability as $n \to \infty$.

(ii) For $k_0 = 0$, conditioning on $\mathbf{x} = \{x_1, \cdots, x_n\}$, we have

$$\sqrt{nh}[\mathcal{H}\{\hat{\alpha}_{IT}(0+; h) - \alpha_T(0+)\} - h u_{0,d}^{-1} u_{1,d} G^{(1)}(0+)] \to^L N(\mathbf{0}, \Omega_{0,d}(0+)), \tag{34}$$

where $\Omega_{0,d}(0+) = f_X^{-1}(0+) u_{0,d}^{-2} v_{0,d} \Psi_1(0+)^{-1} \Psi_{11}(0+) \Psi_1(0+)^{-1}$.

(iii) For $k_0 > 0$, if condition (C9) in the appendix is also true, conditioning on $\mathbf{x} = \{x_1, \cdots, x_n\}$,

we have

$$\sqrt{nh}[\mathcal{H}\{\hat{\alpha}_{IT}(0+;h) - \alpha_T(0+)\} - \frac{h^{k_0+1}}{(k_0+1)!}\mathcal{N}_d(0+)^{-1}\aleph_d(0+)] \to^L N(\mathbf{0}, \Omega_d(0+)), \qquad (35)$$

where $\Omega_d(0+) = f_X^{-1}(0+)\mathcal{N}_d(0+)^{-1}\mathcal{N}_d^*(0+)\mathcal{N}_d(0+)^{-1}$ and $\mathcal{N}_d(0+)$ and $\mathcal{N}_d^*(0+)$ are, respectively, given by

$$\mathcal{N}_d(0+) = \begin{pmatrix} u_{0,d}\Psi_1(0+) & \mathbf{u}_d \otimes \Psi_2(0+) \\ \mathbf{u}_d^T \otimes \Psi_2(0+)^T & \mathcal{U}_{2,d} \otimes \Psi_3(0+) \end{pmatrix},$$

$$\mathcal{N}_d^*(0+) = \begin{pmatrix} v_{0,d}\Psi_{11}(0+) & \mathbf{v}_d^T \otimes \Psi_{12}(0+) \\ \mathbf{v}_d \otimes \Psi_{12}^T(0+) & \mathcal{V}_{2,d} \otimes \Psi_{22}(0+) \end{pmatrix}.$$

It follows from Theorem 3 (ii) and (iii) that when x_0 is at the boundary, the asymptotic average mean squared errors of intrinsic local constant and linear estimators are, respectively, $\text{AMSE}(\log\{\hat{D}_{IT}(0+;h,0)\}) = O_p(h^2 + n^{-1}h^{-1})$ and $\text{AMSE}(\log\{\hat{D}_{IT}(0+;h,1)\}) = O_p(h^4 + n^{-1}h^{-1})$. The rate of convergence for the intrinsic local constant estimator at boundary points is slower than that at interior points, and thus the intrinsic local constant estimator suffers from the well-known boundary effects. However, the intrinsic local linear estimator adapts automatically at the boundary points and its rate of convergence is not influenced by the location of points. Thus, the intrinsic local linear (or polynomial) estimators share the same property of automatic adaptation to the boundary points as the local polynomial estimators in Euclidean space (Fan and Gijbels, 1996).

5. Simulation

We conducted four sets of Monte Carlo simulations to examine the finite sample performance of ILPREs for SPD matrices under different metrics and noise distributions. It should be emphasized that these simulation studies are intended to have wide applications of SPDs, and thus they are deliberately not limited to DTI.

We set $m = 3$ and assumed that the true SPD matrix function has the following form:

$$D(x) = \exp(\begin{pmatrix} -0.1(x+0.1) & 0.2(x+0.1) & \sin(0.75x) \\ 0.2(x+0.1) & 0.6(x+0.1) & -0.4(x+0.1) \\ \sin(0.75x) & -0.4(x+0.1) & 0.5(x+0.1) \end{pmatrix}).$$

We considered three noise distributions including a Riemannian log-normal distribution, a log-normal distribution, and the Rician distribution. We used the Rician noise to simulate the ideal noise in diffusion tensor imaging. The three noise models are given as follows:

- (a) *Riemannian log normal model*: $S_i = G(x_i)\exp(\epsilon_i)G(x_i)^T$ follows the Riemannian log normal distribution, where $D(x_i) = G(x_i)^{\otimes 2}$ and $\epsilon_i \in \mathrm{Sym}(3)$ follows a symmetric matrix variate normal distribution $N(0, \Sigma)$, in which Σ is a covariance matrix (Schwartzman, 2006).

- (b) *Log normal model*: $\log(S_i)$ follows a symmetric matrix variate normal distribution $N[\log\{D(x_i)\}, \Sigma]$.

- (c) *Rician noise model*: this noise model is commonly used to simulate ideal noise in diffusion weighted images (Zhu et al., 2007). The diffusion-weighted signal was simulated for 31 gradient directions $r_k, k = 1, \cdots, 31$ with b-factor $b_k = 1000\mathrm{s/mm}$ and four baselines with $b_k = 0\mathrm{s/mm}$ for $k = 32, \cdots, 35$. The baseline signal intensity W_0 was set at 1500. For a given diffusion tensor $D(x_i)$, $\epsilon_{R,k}$ and $\epsilon_{I,k}$ were independently simulated from a Gaussian random generator with mean zero and standard deviation 60. The diffusion-weighted signal was calculated as $W_{i,k} = \sqrt{[W_0 \exp\{-b_k r_k D(x_i) r_k\} + \epsilon_{R,k}]^2 + \epsilon_{I,k}^2}$ for $k = 1, \cdots, 35$. Subsequently, the weighted least squares estimate was used to estimate S_i (Zhu et al., 2007).

For each simulated data set, we considered three metrics including the trace metric, the Log-Euclidean metric, and the Euclidean metric. For the trace and Log-Euclidean metrics, we calculated the intrinsic local constant and linear estimators developed above for each data set. By following the arguments in Pasternak et al. (2010), we employed the Euclidean metric for estimated diffusion tensors. Under the Euclidean metric, we applied the standard local constant and linear regression methods to estimate the SPD matrix function for each simulated data set, while the bandwidth was selected by using its corresponding generalized cross-validation method. For comparison, we also included a tensor spline method (Barmpoutis et al., 2007) based on the trace metric.

We first generated $n = 50$ design points $x_i, i = 1, \cdots, 50$ independently from a $N(0, 0.25)$ distribution. Then we calculated $D(x_i)$ and used it to simulate S_i according to one of the three noise distributions (a)-(c). Unless stated otherwise, the covariance matrix Σ for the noise models (a) and (b) was set as follows:

$$\Sigma_1 = 2 * \begin{pmatrix} 0.3 & 0.049 & 0.052 \\ 0.049 & 0.2 & 0.0424 \\ 0.052 & 0.0424 & 0.1 \end{pmatrix}. \tag{36}$$

Figure 2 (a)-(d) presents the true SPD matrix data and a set of simulated data $\{(x_i, S_i) : i = 1, \cdots, 50\}$ under the three noise models. Each SPD matrix S_i at the point x_i is geometrically
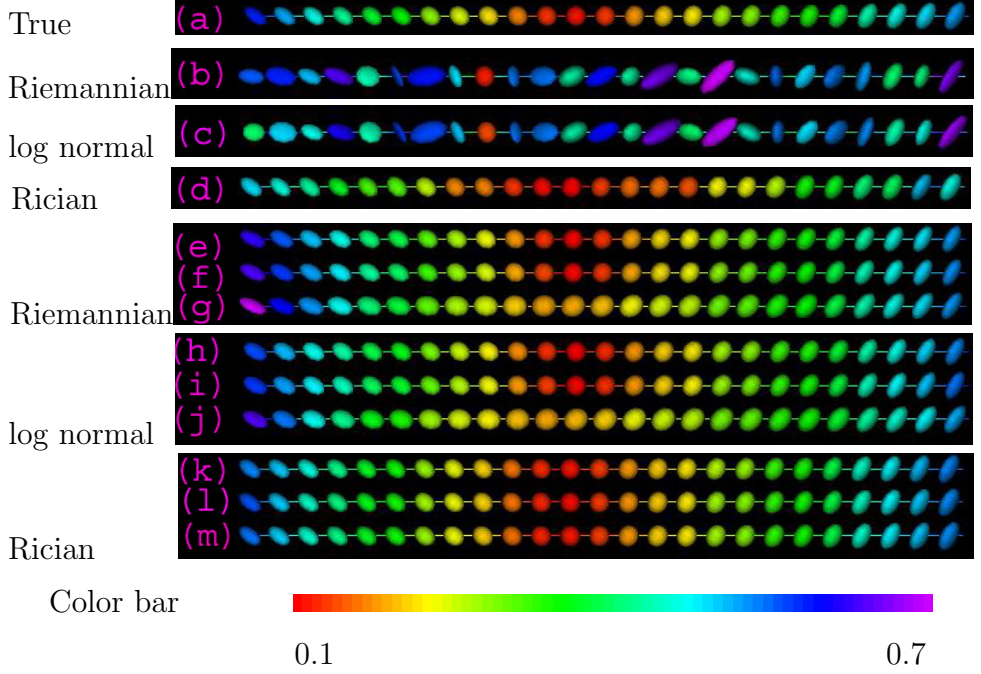
Fig. 2. Ellipsoidal representations of (a) the true SPD matrix data along the design points; simulated SPD matrix data along the design points under the three different noise models: (b) Riemannian log normal, (c) log normal and (d) Rician noise models; and estimated SPD matrix data along the design points using three smoothing methods: (e), (h) and (k): ILPR under the trace metric; (f), (i) and (l): ILPR under the Log-Euclidean metric; and (g), (j), (m): LPR under the Euclidean metric; and under the three different noise models: (e)-(g): Riemannian log normal model; (h)-(j): log normal model; and (k)-(m): Rician noise model, colored with FA values defined in (37).

represented by an ellipsoid. In this representation, the lengths of the semiaxes of the ellipsoid equal the square root of the eigenvalues of a SPD matrix, while the eigenvectors define the direction of the three axes. In DTI, the ellipsoidal representation is used to represent the local brownian motion of water molecules in the brain. Isotropic diffusion is represented by a sphere, while anisotropic diffusion is represented by an anisotropic ellipsoid. We simulated 100 data sets for each scenario. Note that the Rician noise level is visually less variable than the relatively high levels of the other two.

To compare different smoothing methods for SPD matrices under different scenarios, we calculated two summary statistics including an Average Geodesic Distance (AGD) over all design points and a Local Average Geodesic Distance (LAGD) at each design point. Specifically,

AGD is defined as $\text{AGD} = n^{-1} \sum_{i=1}^{n} g(\hat{D}(x_i), D(x_i))$, where $\hat{D}(x_i)$ is an estimated $D(x_i)$ based on a specific smoothing method. At each sample point x_i, LAGD is given by $\text{LAGD}(x_i) = \sum_{j=1}^{100} g(\hat{D}_{(j)}(x_i), D(x_i))/100$, where $\hat{D}_{(j)}(x_i)$ is the estimated SPD matrix at x_i based on the j-th simulated replication. Although we chose all the three metrics for calculating AGD and LAGD, we only present those based on the Euclidean metric for the sake of space. The results of AGD and LAGD for the other two metrics are included in the supplementary document.

5.1. Simulation 1

The first set of simulations compared the finite sample performance of the intrinsic local linear estimators under different metrics and noise distributions. Figure 2(e)-(m) displays a set of the estimated SPD functions using local linear regression estimators under the three metrics for the three different noise models. Inspecting Figure 2 (k)-(m) reveals that under the Rician noise model, all three metrics perform well in recovering the true SPD function. This is not surprising given the relatively low noise level shown in Figure 2(d). However, for the other two noise models, our intrinsic local linear regression methods visually outperform the local linear regression based on the Euclidean metric. In particular, a clear swelling effect is observed for the Euclidean metric (Figure 2 (g) and (j)). This indicates the importance of appropriate metric selection according to the distribution of a specific SPD data set, which is partially in agreement with the suggestion given in Pasternak et al. (2010). However, our findings also suggest that both the trace and Log-Euclidean metrics are appropriate for the nonparametric analysis of SPD matrices for all three noise distributions. This also agrees with the findings in the medical imaging literature (Fletcher and Joshi, 2007; Batchelor et al., 2005; Pennec et al., 2006) on the interpolation and extrapolation of diffusion tensor fields. It should be noted that the simulation studies in Pasternak et al. (2010) solely focus on the effect of metric on the estimated diffusion tensors and their associated scalar measures, such as the apparent diffusion coefficient. Thus, the recommendation in Pasternak et al. (2010) may not apply to the nonparametric analysis of SPD matrices.

5.2. Simulation 2

The second set of simulations compared local constant estimators with local linear estimators under the three metrics and the three noise distributions. In addition, we also compared all local regression methods with the tensor spline estimators in Barmpoutis et al. (2007). Inspecting Figures 3 reveals the following findings. As expected, under all metrics, the local linear estimator is superior to the local constant estimator. Also, our ILPREs outperform the cor-

responding estimators under the Euclidean metric and the tensor spline estimators under the noise models (a) and (b). For the Rician noise model, our ILPREs under the Log-Euclidean metric slightly outperform those under the trace and Euclidean metrics. Moreover, the local constant and linear estimators outperform the tensor spline estimators under all noise distributions. The variations of AGDs for ILPREs under the trace metric are larger than those under the Log-Euclidean metric under all three noise distributions. The U shape of the LAGD curves indicates that interior points have smaller LAGDs than those near the boundaries since there are more design points in the interior than at the boundaries.

5.3. Simulation 3

The third set of simulation studies compared the finite sample performance of the intrinsic local linear estimators under the trace, Log-Euclidean and Euclidean metrics and also tensor spline method in Barmpoutis et al. (2007) at a higher noise level. Specifically, we assumed $\Sigma = 4\Sigma_1$ for the covariance matrix of $N(0, \Sigma)$ in the noise models (a) and (b). At high noise levels, most local linear estimators cannot retain the positive definiteness under the Euclidean metric, while the tensor spline method does not converge. Thus, Figure 4 only presents the results under the trace and Log-Euclidean metrics. Inspecting this figure reveals that when the noise level is high, the intrinsic local linear estimators under the trace metric slightly outperform those under the Log-Euclidean metric under the noise models (a) and (b).

5.4. Simulation 4

The fourth set of simulation studies examined the importance and effect of directly smoothing SPDs on some SPD-derived scalar summary measures under the three noise models. We considered a well-known scalar measure derived from a 3×3 SPD matrix, called fractional anisotropy (FA), which describes the variation of the three eigenvalues of a 3×3 SPD matrix. FA is a scalar value between zero (all eigenvalues are the same) and one (two eigenvlues equal 0) and given by

$$FA = \sqrt{\frac{3\{(\lambda_1 - \bar\lambda)^2 + (\lambda_2 - \bar\lambda)^2 + (\lambda_3 - \bar\lambda)^2\}}{2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}} \tag{37}$$

with eigenvalues $\lambda_1, \lambda_2, \lambda_3$ and their average $\bar\lambda$. We compared two different methods for smoothing FA's, here referred to as method A and method B, respectively. The method A first calculates the FA's from all SPD matrices and then uses the classic local linear regression in Euclidean space to smooth the FA's. The method B first applies the intrinsic local linear estimator to smooth SPD matrices and then calculates smoothed FA curves based on the smoothed SPD
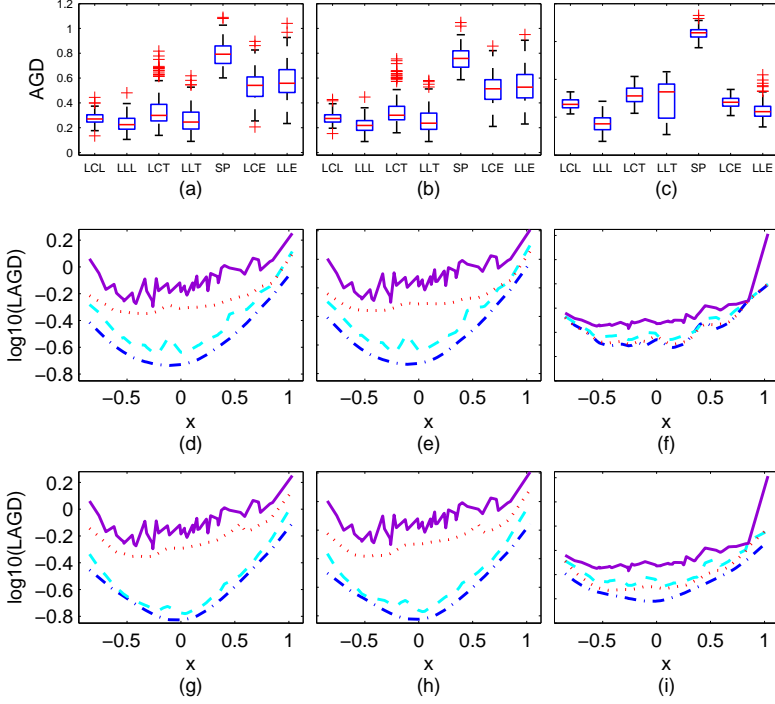
Fig. 3. Comparisons of the local constant and linear estimators under the three metrics and the tensor spline estimators under the three noise models. Panels (a)-(c) the boxplots of $1000\times$AGDs obtained from seven different estimators, where LCL, LCT, and LCE, respectively, represent the local constant estimators under the Log-Euclidean, trace and Euclidean metrics, where LLL, LLT, and LLE , respectively, represent the corresponding local constant and linear estimators under the metrics, and where SP represents the tensor spline estimator. Panels (d)-(f) of the second row show the $\log_{10}$(LAGD) curves based on LCL (dash-dotted line), LCT (dashed line), LCE (dotted line), and SP (solid line). Panels (g)-(i) of the third row show the $\log_{10}$(LAGD) curves based on LLL (dash-dotted line), LLT (dashed line), LLE (dotted line), and SP (solid line). The columns correspond to the three noise models: column 1: Riemannian log normal; column 2: log normal; and column 3: Rician.
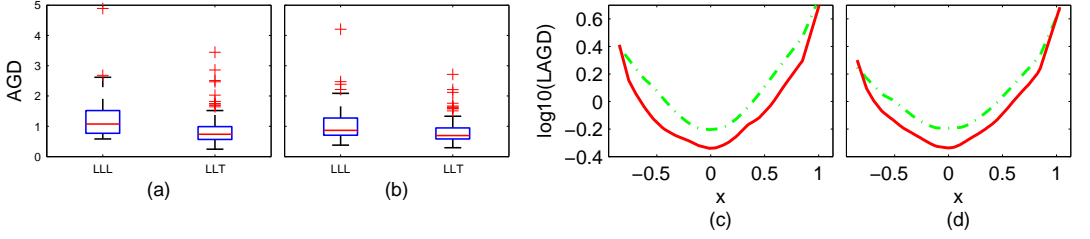
Fig. 4. Comparison of the intrinsic local linear estimators under the Log-Euclidean and trace metrics for the first two noise models at a higher noise level: Panels (a) and (c): Riemannian log normal; Panels (b) and (d): log normal; Panels (a) and (b): the boxplots of AGDs for LLL and LLT; Panels (c) and (d): the $\log_{10}$(LAGD) curves of LLL and LLT. It shows that at a high noise level, the intrinsic local linear estimators under the trace metric slightly outperform those under the Log-Euclidean metric for the first two noise models.

matrices. We further divided method B into three methods according to smoothing methods for SPD matrices under three metrics: trace (method 2), Log-Euclidean (method 3) and Euclidean (method 4) metrics. We assessed each method's performance via the Mean Absolute Deviation Error (MADE) defined by $\text{MADE} = n^{-1} \sum_{i=1}^{n} |FA(x_i) - \widehat{FA}(x_i)|$, where $FA(x_i)$ and $\widehat{FA}(x_i)$ are, respectively, the true and estimated FA values across all design points.

Figure 5 reveals that method B outperforms method A under the noise models (a) and (b). For the Rician noise model, methods A and B are fairly comparable, but method B based on the Log-Euclidean metric is slightly better. This may indicate the potential improvement gained by directly smoothing DT data over the post smoothing method A. Based on the medians of MADEs (see Figure 5(d)-(f)), method A cannot faithfully reconstruct the trend of the FA curve for the noise models (a) and (b), whereas method B can accurately estimate the FA curve and reveal its critical features such as the valley. It should be noted that the true FA value at the valley does not equal zero.

6. HIV Imaging Data

The aim of this analysis is to assess the integrity of white matter in human immunodeficiency virus (HIV) by using DTI and our IPLRE. This clinical study was approved by the Institutional Review Board of the University of North Carolina at Chapel Hill. A sample data set and the code for ILPRE along with its documentation will be accessible from the website http://www.bios.unc.edu/research/bias. We considered 46 subjects with 28 HIV+ subjects (20 males and 8 females whose mean age is 40.0 with SD 5.6 years) and 18 healthy controls (9
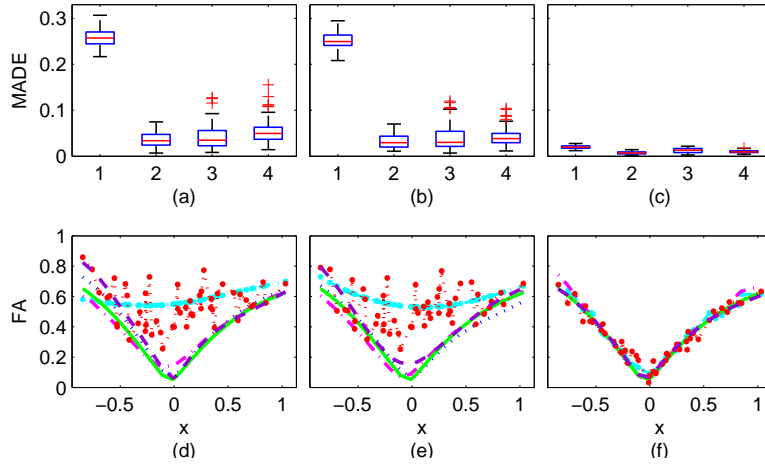
Fig. 5. Boxplot of the MADE's using the four smoothing methods 1-4 representing the first-fourth methods based on 100 replications for three noise models: (a) Riemannian log normal model; (b) log normal model; and (c) Rician model. Smoothed FA curves for the realizations with median MADE for three noise models: (d) Riemannian log normal model; (e) log normal model; and (f) Rician model. In panels (d)-(f), the raw FA curve is the dotted line with circle, true FA curve is the solid line, the estimated FA curve for the first method is the dash-dotted line with circle, the estimated FA curve for the second method is the dotted line, the estimated FA curve for the third method is dash-dotted line and the estimated FA curve for the fourth method is dashed line.

males and 9 females whose mean age is 41.2 with SD 7.4 years). Diffusion-weighted images and T1 weighted images were acquired for each subject. The diffusion tensor acquisition scheme includes 18 repeated measures of six non-collinear directions, (1,0,1), (-1,0,1), (0,1,1), (0,1,-1), (1,1,0), and (-1,1,0) at a b-value of 1000 s/mm^2 and a $b = 0$ reference scan. Forty-six contiguous slices with a slice thickness of 2 mm covered a field of view (FOV) of 256 mm^2 with an isotropic voxel size of $2 \times 2 \times 2$ mm^3. High resolution T1 weighted (T1W) images were acquired using a 3D MP-RAGE sequence. A weighted least square estimation method was used to construct the diffusion tensors (Zhu et al., 2007). Since in the previous DTI findings, the diffusion tensors in the splenium of the corpus callosum were found significantly different between the HIV+ and control groups, we examine the finite sample performance of our method by using this fiber tract. The tensors along the tract were extracted using methodology described in Zhu et al. (2010). Figure 6 displays the splenium of the corpus callosum and the ellipsoidal representation of the full tensors on that tract from one selected subject. This involves three steps: (i) registration and atlas construction, (ii) fiber tracking on the atlas and (iii) collection of tensor data on the atlas fiber tracts.

We calculated the intrinsic local linear estimator of the SPD matrices along this selected tract for each subject under the trace and Log-Euclidean metrics and also calculated the local linear estimator under the Euclidean metric. See Figure 7 for the raw and estimated tensors along the fiber tracts from one subject. It is observed from the ellipsoidal representation of diffusion tensor data (Figure 7 (a)) that the data are noisy. Figure 7 (b)-(e) show that the tensors are more spherical at the beginning with low FA values and more anisotropic in the middle part with high FA values. The methods under the three metrics reveal very similar trend of diffusion tensors changing along the fiber tract, especially in the first row and the last row. This agrees with our simulation results that for diffusion tensor data, all three metrics are comparable. However, some differences appear on the right side of the middle row. The estimated tensors in the middle row are very anisotropic when using the method under the Euclidean metric compared to the other two metrics.

In many applications, it is common to calculate some tensor-derived diffusion measures, including FA, the trace of a diffusion tensor, called MD, and the largest eigenvalue of a diffusion tensor, called PE, based on noisy diffusion tensor data and then apply standard statistical methods to directly carry out statistical inference on these diffusion measures. Since these scalar measures do not capture all information in the full diffusion tensor, they can decrease the sensitivity of detecting subtle changes of the white matter structure. Similar to the simulation of Section 5.4, we applied method A to directly smooth FA, MD and PE values along the
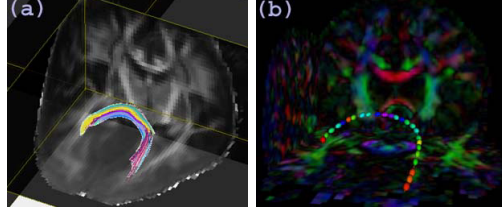
Fig. 6. (a)The splenium of the corpus callosum in the analysis of HIV DTI data. (b)The ellipsoidal representation of full tensors colored with FA values on the fiber tract from a selected subject.

selected tract and then we compared them with method B based on the trace, Log-Euclidean, and Euclidean metrics. Figure 7(e)-(g) shows that there is no large difference for methods B under the three metrics. Both methods A and B perform the same for smoothing MD data, whereas they perform differently for smoothing PE and FA curves, especially in the middle part from 10 to 25. This seems to be caused by that fact that FA and PE values are biased due to the well known 'sorting' bias in estimating the eigenvalues of DT (Zhu et al., 2007), whereas the estimated MD value is unbiased.

Finally, we estimated the mean diffusion tensor curve for each of the two groups: HIV and control groups. In order to detect meaningful group differences, registration is crucial. The 46 HIV DTI data used in our studies, including the splenium tracts and diffusion tensors on them, were registered in the same atlas space. Figure 8(a)-(f) displays the estimated mean diffusion tensors along the fiber tract for the two groups using the intrinsic local linear regression for SPD matrices under both the Log-Euclidean and trace metrics and also using the local linear regression for SPD matrices under the Euclidean metric. We can observe some obvious changes of diffusion tensors of HIV subjects along the splenium corpus callosum compared with those in the control group. We also calculated the differences of FA values derived from the estimated mean diffusion tensors, which corresponds to the color differences in Figure 8 (g), and the geodesic distances between estimated mean tensors at each point along the tract in Figure 8 (h). This result agrees with previous DTI findings that the spleninum of the corpus callosum has been detected as abnormal for the HIV group (Filippi et al. (2001) and Chen et al. (2009)).

7. Conclusion and Discussion

We have systematically investigated the intrinsic local polynomial regression methods under the trace and Log-Euclidean metrics on the space of SPD matrices. Many issues still merit further research. The proposed cross validation bandwidth selector is straightforward and
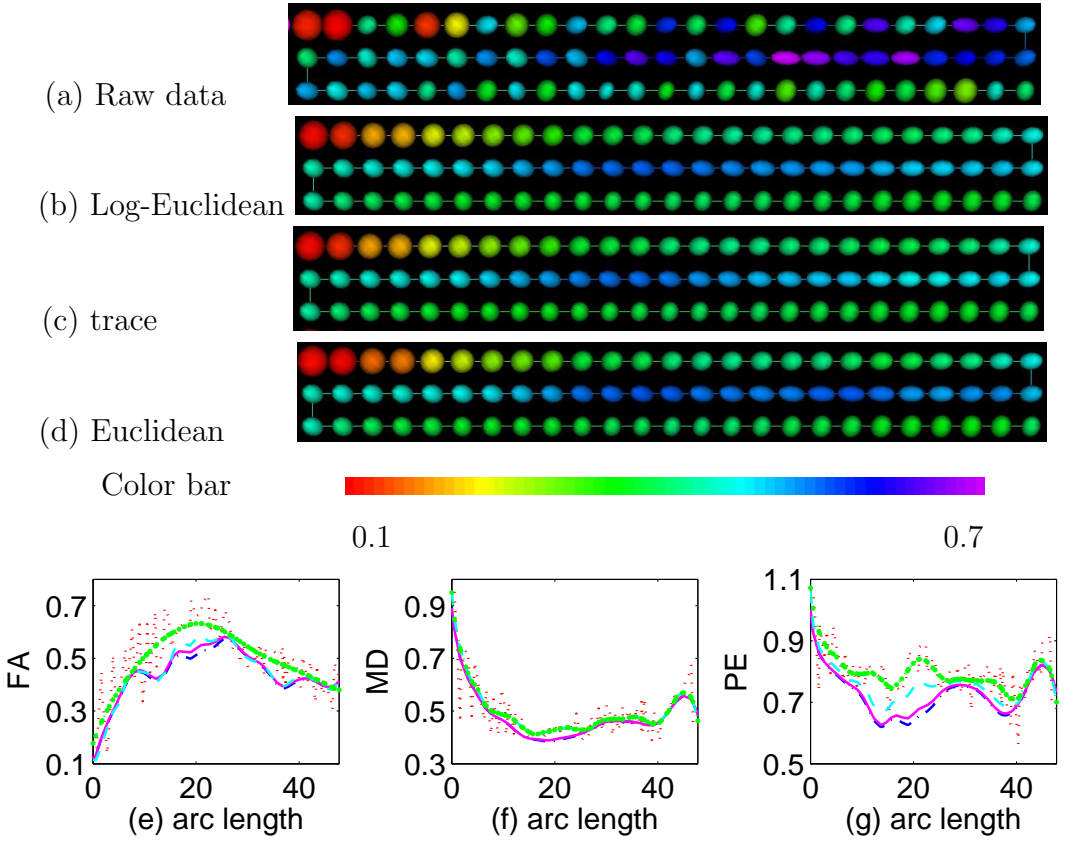
Fig. 7. (a) The ellipsoidal representations of the diffusion tensor data and estimated tensors using the intrinsic local linear regression under the (b) Log-Euclidean, (c) trace and (d) Euclidean metrics along the splenium of the corpus callosum, colored with FA values. The estimated tensors in the middle right part are more anisotropic using the method under the Euclidean metric. Each set of 3 rows in (a)-(d) represents one tract of tensors and the three rows are read from left to right in the top row, right to left in the middle row and then left to right in the bottom row. (e) FA's , (f) MD's and (g) PE's derived from the raw tensor data (dot line) and estimated tensors using the intrinsic local linear regression under the trace (dash-dot line), Log-Euclidean (solid line) and Euclidean (dashed line) metrics as the function of arc-length along the splenium of the corpus callosum. Estimated FA, MD and PE function along the splenium of the corpus callosum by using the standard local linear regression for scalars (dotted line with circles).
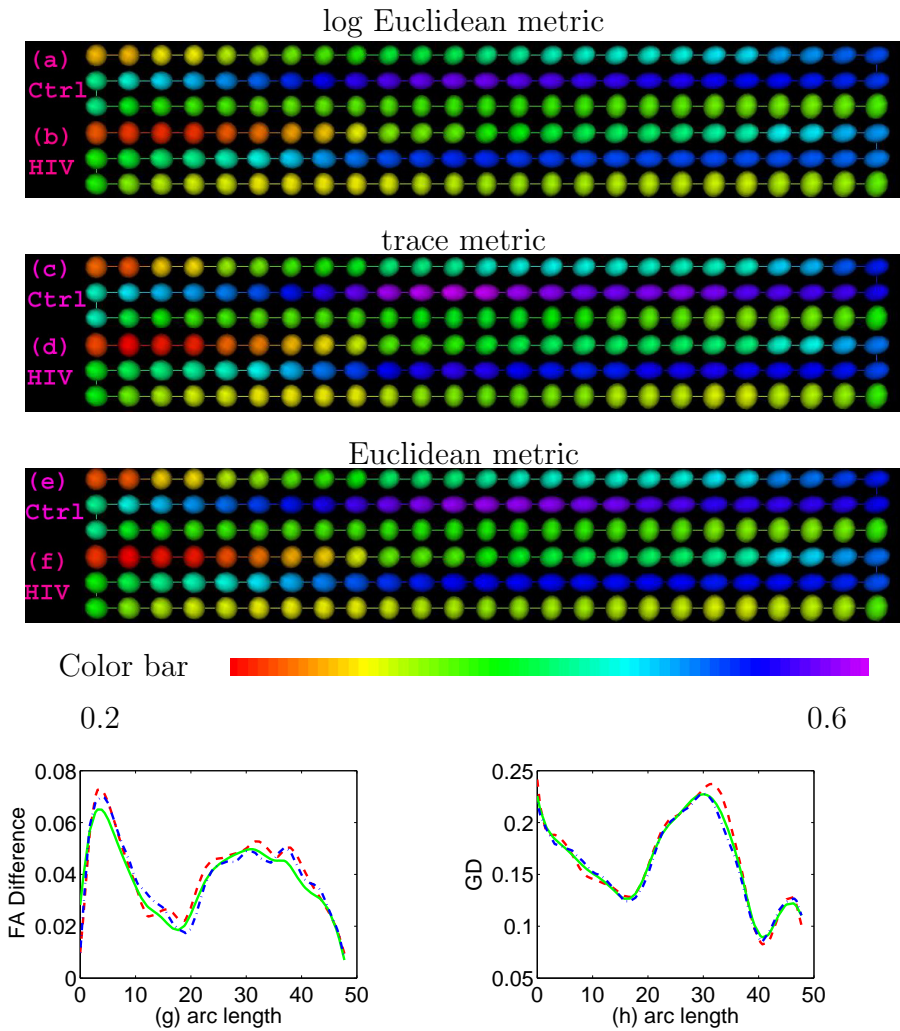
Fig. 8. Ellipsoidal representations of estimated mean tensors along the splenium of the corpus callosum for the control and HIV groups using the intrinsic local linear regression under the Log-Euclidean ((a) and (b)), trace ((c) and (d)) and Euclidean ((e) and (f)) metrics colored with FA values. Each set of 3 rows in (a)-(d) represents one tract of tensors and the three rows are read from left to right in the top row, right to left in the middle row and then left to right in the bottom row. (g) FA differences and (h) geodesic distances (GD) between the mean diffusion tensors of HIV and control groups along the splenium of the corpus callosum under the Log-Euclidean (the solid line), trace (the dashed line) and Euclidean (dash-dot line) metrics.

relatively simple to derive and implement for SPD matrix variate data. However, the relatively high variance of the cross validation bandwidth selector is regarded widely as an impediment to its good performance (Jones et al., 1996; Hardle et al., 1992). It would be great interest to develop variable bandwidth selection methods to capture complicated variations of SPD matrices in the covariate space and better bandwidth selection methods to reduce the variability of cross validation (Fan et al., 1996). From the average diffusion tensor curves for the HIV and control groups in Section 6, we can observe some obvious changes in diffusion tensors of HIV subjects along the spleninium corpus callosum compared with the control group. A topic of future interest should propose tests for comparing the differences across multiple groups of SPD curves by considering varying-coefficient models and additive models among others. The real applicability of the log normal and Riemannian log normal noise models remains unclear. It is of great interest to explore the fit of those noise models to real data in different applications.

Finally, the ILPR method proposed here and theory may be extended to other nonparametric methods (e.g., tensor splines) and manifold-valued data, such as directional data and rotation matrices. For instance, although we have compared our ILPR with Barmpoutis et al. (2007)'s tensor splines, it would be interesting to develop free-knot regression splines for SPD matrices and compare them with our ILPR both theoretically and numerically (Sangalli et al., 2009). Moreover, a smooth spline method based on unrolling and unwrapping procedures in Riemannian manifolds has been developed for fitting smooth curves to spherical data, rotation matrices, and planar landmark data (Jupp and Kent, 1987; Prentice, 1987; Kume et al., 2007). Development of other nonparametric methods, such as ILPR, for the analysis of manifold-valued data and examination of the asymptotic properties of nonparametric estimates under different metrics should be pursued in future research.

Acknowledgement

References

Anderson, T. W. (2003). *An introduction to multivariate statistical analysis (3rd ed.).* Wiley Series in Probability and Statistics.

Arsigny, V., P. Fillard, X. Pennec, and N. Ayache (2007). Geometric means in a novel vector space structure on symmetric positive definite matrices. *SIAM. J. Matrix Anal. Appl. 29*, 328–347.

Barmpoutis, A., B. C. Vemuri, T. M. Shepherd, and J. R. Forder (2007). Tensor splines for interpolation and approximation of dt-mri with applications to segmentation of isolated rat hippocampi. *IEEE Transations on Medical Imaging 26*, 1537–1546.

Basser, P. J., J. Mattiello, and D. LeBihan (1994). Mr diffusion tensor spectroscopy and imaging. *Biophysical Journal 66*, 259–267.

Batchelor, P., M. Moakher, D. Atkinson, F. Calamante, and A. Connelly (2005). A rigorous framework for diffusion tensor calculus. *Magnetic Resonance in Medicine 53*, 221–225.

Bhattacharya, R. and V. Patrangenaru (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds-ii. *Annals of Statistics 33*, 1225–1259.

Chen, Y. S., H. T. An, H. Y.and Zhu, T. Stone, J. K. Smith, C. Hall, E. Bullitt, D. G. Shen, and W. L. Lin (2009). White matter abnormalities revealed by diffusion tensor imaging in non-demented and demented hiv+ patients. *NeuroImage 47*, 1154 –1162.

Davis, B. C., E. Bullitt, P. T. Fletcher, and S. Joshi (2010). Population shape regression from random design data. *International Journal of Computer Vision 90*, 255–266.

Dryden, I. L., A. Koloydenko, and D. Zhou (2009). Non-euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Annals of Applied Statistics 3*, 1102–1123.

Fan, J. and I. Gijbels (1996). *Local polynomical modelling and its applications.* Cahpman and Hall.

Fan, J., I. Gijbels, T.-C. Hu, and L.-S. Huang (1996). A study of variable bandwidth selection for local polynomial regression. *Statist. Sinica 6*(1), 113–127.

Filippi, C., A. Ulug, E. Ryan, S. Ferrando, and W. van Gorp (2001). Diffusion tensor imaging of patients with hiv and normal-appearing white matter on mr images of the brain. *Ajnr: Am. J. Neuroradiol 22*, 277–283.

Fingelkurts, A. A., A. A. Fingelkurts, and S. Kahkonen. (2005). Functional connectivity in the brain-is it an elusive concepts? *Neuroscience and Biobehavioral Reviews 28*, 827–836.

Fletcher, P. and S. Joshi (2007). Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Processing 87*, 250–262.

Fletcher, P., S. Joshi, C. Lu, and S. Pizer (2004). Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on Medical Imaging 23*, 995–1005.

Grenander, U. and M. I. Miller (2007). *Pattern Theory From Representation to Inference.* Oxford University Press.

Hall, P., J. S. Marron, and B. U. Park (1992). Smoothed cross-validation. *Probability Theory and Related Fields 92*, 1–20.

Hardle, W., P. Hall, and J. S. Marron (1992). Regression smoothing parameters that are not far from their optimum. *Journal of American Statistical Association 87*, 227–233.

Jones, M., J. Marron, and S. Sheather (1996). A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association 91*, 401–407.

Jupp, P. E. and J. T. Kent (1987). Fitting smooth paths to spherical data. *Applied Statistics 36*, 34–46.

Kim, P. T. and D. S. Richards (2010). *Deconvolution density estimation on spaces of positive definite symmetric matrices.* IMS Lecture Notes Monograph Series. A Festschrift of Tom Hettmansperger.

Kume, A., I. L. Dryden, and H. Le (2007). Shape-space smoothing splines for planar landmark data. *Biometrika 94*, 513–528.

Lang, S. (1999). *Fundamentals of Differential Geometry*, Volume 191 of *Graduate Texts in Mathematics.* Springer Verlag, New York.

Liang, F. (2011). Annealing evolutionary stochastic approximation monte carlo for global optimization. *Statistics and Computing 21*, 375–393.

Park, B. U. and J. S. Marron (1990). Comparison of data-driven bandwidth selectors. *Journal of American Statistical Association 85*, 66–72.

Pasternak, O., N. Sochen, and J. P. Basser (2010). The effect of metric selection on the analysis of diffusion tensor mri data. *Neuroimage 49*, 2190–2204.

Pennec, X., P. Fillard, , and N. Ayache (2006). A riemannian framework for tensor computing. *International Journal of Computer Vision 66*, 41–66.

Pourahmadi, M. (2000). Maximum likelihood estimation of generalized linear models for multivariate normal covariance matrix. *Biometrika 87*, 425–435.

Prentice, M. J. (1987). Fitting smooth paths to rotation data. *Journal of the Royal Statistical Society. Series C (Applied Statistics) 36*, 325–331.

Rice, J. (1984). Bandwidth choice for nonparametric regression. *Annals of Statistics 12*, 1215–1230.

Sakamoto, Y., M. Ishiguro, and G. Kitagawa (1999). *Akaike Information Criterion Statistics (Mathematics and its Applications)*. Springer.

Sangalli, L., P. Secchi, S. Vantini, and A. Veneziani (2009). Efficient estimation of three-dimensional curves and their derivatives by free-knot regression splines, applied to the analysis of inner carotid artery centrelines. *Journal of the Royal Statistical Society, Ser. C 58*, 285–306.

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

Terras, A. (1988). *Harmonic Analysis on Symmetric Spaces and Applications II*. Springer-Verlag, Berlin, Heidelberg and New York.

Wand, M. P. and M. C. Jones (1995). *Kernel Smoothing*. London: Chapman and Hall.

Zhu, H. T., Y. S. Chen, J. G. Ibrahim, Y. M. Li, and W. L. Lin (2009). Intrinsic regression models for positive-definite matrices with applications to diffusion tensor imaging. *Journal of the American Statistical Association 104*, 1203–1212.

Zhu, H. T., M. Styner, N. S. Tang, Z. X. Liu, W. L. Lin, and J. Gilmore (2010). Frats: Functional regression analysis of dti tract statistics. *IEEE Transactions on Medical Imaging 29*, 1039–1049.

Zhu, H. T., H. P. Zhang, J. G. Ibrahim, and B. G. Peterson (2007). Statistical analysis of diffusion tensors in diffusion-weighted magnetic resonance image data (with discussion). *Journal of the American Statistical Association 102*, 1085–1102.

Appendix. Assumptions

The following assumptions are needed to facilitate development of our methods, although they are not the weakest possible conditions. We need some notation. Recall that $\psi(S, G, Y) = g_T(S, G \exp(Y)G^T)^2$ and $\alpha = (\alpha_G, \alpha_Y)$, where G is an $m \times m$ lower triangle matrix, $S \in \mathrm{Sym}^+(m)$, $Y \in \mathrm{Sym}(m)$, $\alpha_G = \mathrm{vecs}(G)$, and $\alpha_Y = \mathrm{vecs}(Y)$. We define

$$\begin{pmatrix} \partial_{\alpha_G}\psi(S, G, Y) \\ \partial_{\alpha_Y}\psi(S, G, Y) \end{pmatrix} = \begin{pmatrix} \psi_G(S, G, Y) \\ \psi_Y(S, G, Y) \end{pmatrix},$$

$$\begin{pmatrix} \partial^2_{\alpha_G}\psi(S, G, Y) & \partial^2_{\alpha_G\alpha_Y}\psi(S, G, Y) \\ \partial^2_{\alpha_Y\alpha_G}\psi(S, G, Y) & \partial^2_{\alpha_Y}\psi(S, G, Y) \end{pmatrix} = \begin{pmatrix} \psi_{GG}(S, G, Y) & \psi_{GY}(S, G, Y) \\ \psi_{YG}(S, G, Y) & \psi_{YY}(S, G, Y) \end{pmatrix}.$$

(C1) The kernel function $K(.)$ is a continuous symmetric probability density function with bounded support, say $[-1, 1]$.

(C2) The regression function $D(x) \in \mathrm{Sym}^+(m)$ has a continuous $(k_0 + 1)$-th order derivative in a neighborhood of x_0.

(C3) The bandwidth h tends to zero and $nh \to \infty$.

(C4) The design density $f_X(.)$ is continuous in a neighborhood of x_0 and $f_X(x_0) > 0$.

(C5) The conditional density $f(S|X = x)$ is continuous in a neighborhood of x_0.

(C6) $E\{\partial^2_\alpha\psi(S, G, Y(X))|X = x\}$ and $E[\{\partial_\alpha\psi(S, G, Y(X))\}^{\otimes 2}|X = x]$ are continuous in a neighborhood of x_0.

(C7) The matrix $\mathcal{N}(x) = \begin{pmatrix} u_0\Psi_1(x) & \mathbf{u} \otimes \Psi_2(x) \\ \mathbf{u}^T \otimes \Psi_2(x)^T & \mathcal{U}_2 \otimes \Psi_3(x) \end{pmatrix}$ is positive definite in a neighborhood of x_0.

(C8) Let $\|.\|$ be the L_2 norm of a matrix, η_0 be a lower triangle matrix, $\eta_1 \in \mathrm{Sym}(m)$, and $U_\delta = \{(\eta_0, \eta_1) : \|\eta_0\|^2 + \|\eta_1\|^2 \leq \delta^2\}$. As $\delta \to 0$,

$$E\{\sup_{U_\delta}\|\partial^2_\alpha\psi(S, G(x_0) + \eta_0, Y(X) + \eta_1) - \partial^2_\alpha\psi(S, G(x_0), Y(X))\||X = x\} = o(1),$$

$$E\{\sup_{U_\delta}\|\partial_\alpha\psi(S, G(x_0) + \eta_0, Y(X) + \eta_1) - \partial_\alpha\psi(S, G(x_0), Y(X))$$
$$-\partial^2_\alpha\psi(S, G(x_0), Y(X))(\mathrm{vecs}(\eta_0)^T, \mathrm{vecs}(\eta_1)^T)^T\||X = x\} = o(\delta),$$

are uniformly in x in a neighborhood of x_0.

(C9) There exists a $b > 0$ such that $E\{\|\partial_\alpha \psi(S, G(x_0), Y(X))\|^{b+2}|X = x\}$ is bounded in a neighborhood of x_0.

(C10) The $\Sigma_{\mathcal{E}_D}(x) = \text{Cov}\{\mathcal{E}_D(X)|X = x\}$ is continuous in a neighborhood of x_0 and there exists a $b > 0$ such that $E\{\|\mathcal{E}_D(X)\|^{b+2}|X = x\}$ is bounded in a neighborhood of x_0.

Remark: Assumptions (C1)-(C10) are standard conditions for ensuring the asymptotic properties of local polynomial estimators when x_0 is an interior point of $f_X(\cdot)$ (Fan and Gijbels, 1996; Wand and Jones, 1995). Some conditions can be released with additional technicalities of proofs. For instance, the bounded support restriction on $K(\cdot)$ in (C1) is not essential and can be removed if we put restriction on the tail of $K(\cdot)$. Condition (C2) ensures that $Y(x) = \log(G(x_0)^{-1}D(x)G(x_0)^{-T})$, $G(x)$, and $\log(D(x))$ have a continuous $(k_0 + 1)$-th order derivative in a neighborhood of x_0. Moreover, assume that $f_X(.)$ has a bounded support $[0, 1]$. All assumptions can be easily modified when x_0 is a boundary point, say left boundary point $x_0 = dh$ or right boundary point $x_0 = 1 - dh$ for some $d > 0$. For instance, we require that conditions (C2)-(C10) hold in the left neighborhood of 0 or the right neighborhood of 1. For condition (C2), we also need to introduce $f_X(0+)$ as x_0 is the left boundary point and $f_X(1-)$ as x_0 is the right boundary point. For condition (C7), $\mathcal{N}(x)$ is also needed to make some modifications. For simplicity, we omit these details.