Journal of the Royal Statistical Society
SERIES B Statistical Methodology

J. R. Statist. Soc. B (2012)
74, *Part* 4, *pp.* 697–719

Local polynomial regression for symmetric positive definite matrices

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

University of North Carolina at Chapel Hill, USA

[Received August 2010. Final revision October 2011]

Summary. Local polynomial regression has received extensive attention for the non-parametric 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 matrices as responses that lie in a Riemannian manifold with covariate in Euclidean space. The primary motivation and application of the methodology proposed 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 symmetric positive definite 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 fibre tracts in a study of human immunodeficiency virus.

Keywords: Bandwidth; Cross-validation; Intrinsic conditional expectation; Local polynomial regression; Symmetric positive definite matrix

1. Introduction

Symmetric positive definite (SPD) matrix-valued data occur in a wide variety of important applications. For instance, in computational anatomy, an SPD deformation vector $(JJ^{\mathrm{T}})^{1/2}$ 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 three-dimensional pixel) of an imaging space. In functional magnetic resonance imaging, an 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

Address for correspondence: Hongtu Zhu, Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7420, USA.
E-mail: hzhu@bios.unc.edu

 1369–7412/12/74697

(2006) proposed several parametric models for analysing 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 non-parametric estimator for the common density function of a random sample of positive definite matrices. Zhu *et al.* (2009) developed a semiparametric 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 analysing random SPD matrices including the asymptotic properties of the non-parametric 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 an SPD matrix response S, given a covariate vector x from a set of observations $(x_1, S_1), \ldots, (x_n, S_n)$, where the x_i can be either univariate or multivariate. In practice, x can be the arc length of a specific fibre tract (e.g. right internal capsule tract), the co-ordinates in the three-dimensional imaging space or demographic variables such as age. Important applications of ILPR include smoothing diffusion tensors along fibre tracts and smoothing diffusion and deformation tensor fields. Another application is quantifying the change of diffusion and deformation tensors as well as the interregional functional connectivity matrix across groups and over time.

Relative to the existing literature on the analysis of SPD matrices, we make several important contributions in this paper.

(a) 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 covariate. We also derive an approximation of a cross-validation method for bandwidth selection.
(b) Theoretically, we compare the trace metric and the Log-Euclidean metric and establish the asymptotic properties of the ILPR estimators (ILPREs) corresponding to each metric.
(c) 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 non-parametric analysis of random SPD matrix-valued data. In Section 3, we compare the trace metric and the Log-Euclidean metric and derive their ILPREs. 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 ILPREs via simulation studies in Section 5. We analyse 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.

The data that are analysed in the paper and the programs that were used to analyse them can be obtained from

`http://www.blackwellpublishing.com/rss`

2. Intrinsic local polynomial regression for symmetric positive definite matrices

In this section, we develop a general framework for using intrinsic local polynomial regression

in the analysis of SPD matrices and we shall 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 the matrix exponential and logarithm. For any matrix $A \in \mathrm{Sym}(m)$, its matrix exponential is given by $\exp(A) = \Sigma_{k=0}^{\infty} A^k/k! \in \mathrm{Sym}^+(m)$. 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 non-parametric 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.

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

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

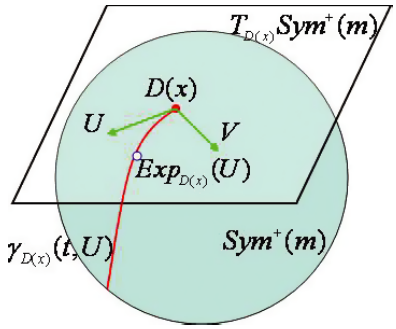

Fig. 1. Graphical illustration of the geometrical structure of $\mathrm{Sym}^+(m)$ near $D(x)$

is called the geodesic distance between $D_1(x)$ and $D_2(x)$, which is denoted $g\{D_1(x), D_2(x)\}$, which satisfies

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

We define $\mathcal{E}_D(X)$ to be $\mathrm{Log}_{D(X)}(S)$ in $T_{D(X)}\,\mathrm{Sym}^+(m)$. Statistically, $\mathcal{E}_D(X)$ can be regarded as the residual of S relative to $D(X)$. Let $\mathrm{vecs}(C) = (c_{11}, c_{21}, c_{22}, \ldots, c_{m1}, \ldots, c_{mm})^{\mathrm{T}}$ 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 $D(x) \in \mathrm{Sym}^+(m)$ such that

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

where O_m is the $m \times m$ matrix with all elements 0 and the expectation is taken componentwise with respect to the multivariate random vector $\mathrm{vecs}\{\mathrm{Log}_{D(x)}(S)\}$. In fact, equation (2) characterizes intrinsic means (Bhattacharya and Patrangenaru, 2005).

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

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

In equation (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.

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

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

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

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

Local polynomial regression has received extensive attention for the non-parametric 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.

How do we define the intrinsic local polynomial regression for estimating $D(X)$ in equation (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 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

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

i.e. we have

$$\begin{aligned} Y(x) = \phi_{D(x_0)}[\mathrm{Log}_{D(x_0)}\{D(x)\}] \in T_{I_m}\,\mathrm{Sym}^+(m), \\ \mathrm{Log}_{D(x_0)}\{D(x)\} = \phi^{-1}_{D(x_0)}\{Y(x)\}, \end{aligned} \tag{5}$$

where $\phi^{-1}_{D(x_0)}(\cdot)$ 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 series expansion as follows:

$$\mathrm{Log}_{D(x_0)}\{D(x)\} = \phi^{-1}_{D(x_0)}\{Y(x)\} \approx \phi^{-1}_{D(x_0)}\left\{ \sum_{k=1}^{k_0} Y^{(k)}(x_0)(x - x_0)^k \right\}, \tag{6}$$

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

$$D(x) \approx \mathrm{Exp}_{D(x_0)}\left[\phi^{-1}_{D(x_0)}\left\{ \sum_{k=1}^{k_0} Y^{(k)}(x_0)(x - x_0)^k \right\} \right] = D\{x, \alpha(x_0), k_0\}, \tag{7}$$

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

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

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

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

$$\hat{\alpha}_{\mathrm{I}}(x_0; h) = \arg\min_{\alpha(x_0)} [G_n\{\alpha(x_0)\}]. \tag{9}$$

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

We propose to use a leave-one-out cross-validation method for bandwidth selection owing to its conceptual simplicity. Let $\hat{D}^{(-i)}_{\mathrm{I}}(x_i; h)$ be the estimate of $D(x_i)$ that is 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

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

The optimal h, which is denoted by $\hat{h}$, can be obtained by minimizing $\mathrm{CV}(h)$. However, since computing $\hat{D}^{(-i)}_{\mathrm{I}}(x_i; h)$ for all i can be computationally prohibitive, we suggest use of the first-order approximation of $\mathrm{CV}(h)$, 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. Intrinsic local polynomial regression under Log-Euclidean metric and trace metric

As discussed in Dryden *et al.* (2009), various metrics can be defined for tangent vectors on $T_{D(x)}\,\mathrm{Sym}^+(m)$. 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 $\exp(\cdot)$ and $\log(\cdot)$ 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 $D(x) \in \mathrm{Sym}^+(m)$ acting on an infinitesimal displacement $U \in T_{D(x)}\,\mathrm{Sym}^+(m)$ (Arsigny *et al.*, 2007). The Log-Euclidean metric on $\mathrm{Sym}^+(m)$ is defined as

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

where U and V are in $T_{D(x)}\,\mathrm{Sym}^+(m)$. 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 $\log\{D(x)\} \in \mathrm{Sym}(m)$ acting on an infinitesimal displacement $A \in T_{\log\{D(x)\}}\,\mathrm{Sym}(m)$ (Arsigny *et al.*, 2007). The Riemannian exponential and logarithm maps are respectively given by

$$\begin{aligned} \mathrm{Exp}_{D(x),\mathrm{L}}(U) &= \exp[\log\{D(x)\} + \partial_{D(x)} \log .(U)], \\ \mathrm{Log}_{D(x),\mathrm{L}}(S) &= \partial_{\log\{D(x)\}} \exp .[\log(S) - \log\{D(x)\}]. \end{aligned} \tag{12}$$

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

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

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

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

Combining equations (12) and (14) yields

$$\begin{aligned} Y(x) = \phi_{D(x_0),\mathrm{L}}[\mathrm{Log}_{D(x_0),\mathrm{L}}\{D(x)\}] &= \log\{D(x)\} - \log\{D(x_0)\}, \\ D(x) &= \exp[\log\{D(x_0)\} + Y(x)]. \end{aligned} \tag{15}$$

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

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

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

To compute the ILPRE, we use Taylor series expansion to expand $\log\{D(x)\}$ at x_0 as follows:

$$\log\{D(x)\} \approx \sum_{k=0}^{k_0} \log\{D(x_0)\}^{(k)}(x - x_0)^k = \log[D_{\mathrm{L}}\{x, \alpha_{\mathrm{L}}(x_0), k_0\}], \tag{17}$$

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

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

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

Let $e_{k_0+1,\,i}$ be the k_0+1 unit vector having 1 in the ith entry and 0 elsewhere. Let

$$e_{k_0+1,\,i}^{\mathrm{T}} \left\{ \sum_{j=1}^{n} K_h(x_j - x) \mathcal{X}_j(x)^{\otimes 2} \right\}^{-1} K_h(x_i - x)\, \mathcal{X}_i(x) = a_i(x).$$

The cross-validation score $\mathrm{CV}(h)$ can be simplified as

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

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

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

For the Log-Euclidean metric, we use $\mathrm{GCV}(h)$ to select the bandwidth throughout this paper.

3.2. Trace metric

We review some basic facts about the theory of the trace metric (Schwartzman, 2006; Lang, 1999; Terras, 1988; Fletcher *et al.*, 2004; Batchelor *et al.*, 2005; Pennec *et al.*, 2006). We add the notation '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

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

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

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

The geodesic distance between $D(x)$ and S, which is denoted by $g_{\mathrm{T}}\{D(x), S\}$, is given by

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

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

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

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

Thus, combining equations (22) and (24) yields

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

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

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

$$D(x) \approx G(x_0) \exp\left\{\sum_{k=1}^{k_0} Y^{(k)}(x_0)(x - x_0)^k\right\} G(x_0)^{\mathrm{T}} = D_{\mathrm{T}}\{x, \alpha_{\mathrm{T}}(x_0), k_0\}, \tag{26}$$

where $\alpha_{\mathrm{T}}(x_0)$ contains all unknown parameters in $G(x_0)$ and $Y^{(k)}(x_0)$ for $k = 1, \ldots, k_0$. Thus, we can compute $\hat{\alpha}_{\mathrm{IT}}(x_0; h)$ by minimizing $G_n\{\alpha_{\mathrm{T}}(x_0)\}$. Under the trace metric, minimizing $G_n\{\alpha_{\mathrm{T}}(x_0)\}$ is computationally challenging when $k_0 > 0$, since $G_n\{\alpha_{\mathrm{T}}(x_0)\}$ is not convex and may have multiple local minimizers. Thus, standard gradient methods, which strongly depend on the starting value of $\alpha_{\mathrm{T}}(x_0)$, do not perform well for optimizing $G_n\{\alpha_{\mathrm{T}}(x_0)\}$ when $k_0 > 0$. Hence, we develop an annealing evolutionary stochastic approximation Monte Carlo algorithm (see Liang (2010) for good discussion) for computing $\hat{\alpha}_{\mathrm{IT}}(x_0; h)$. Details can be found in the on-line supplementary document.

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

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

where $\hat{D}_{\mathrm{IT}}(x; h, k_0) = D_{\mathrm{T}}\{x, \hat{\alpha}_{\mathrm{IT}}(x_0; h), k_0\}$. The $\mathrm{CV}_{\mathrm{T}}(h)$ is close to Akaike's information criterion (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, the ILPRE is almost equivalent to the LPRE for a multivariate response in Euclidean space. Thus, we can generalize the existing theory of the LPRE (Fan and Gijbels, 1996; Wand and Jones, 1995). Moreover, we only present the consistency and asymptotic normality of the ILPRE for interior points, since the asymptotic properties of the 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 $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^{\mathrm{T}}$ for any vector or matrix $\mathbf{a}$ and I_q be an identity matrix of size $q = m(m+1)/2$. Let $\mathcal{H} = \mathrm{diag}(1, h, \ldots, h^{k_0}) \otimes I_q$. Let $\mathbf{u} = (u_1, \ldots, u_{k_0})^{\mathrm{T}}$ and $\mathbf{v} = (v_1, \ldots, v_{k_0})^{\mathrm{T}}$ be $k_0 \times 1$ vectors, where $u_k = \int x^k K(x)\, \mathrm{d}x$ and $v_k = \int x^k K(x)^2\, \mathrm{d}x$ for $k \geqslant 0$. Let $\mathcal{U}_0 = (u_{i+j})$ and $\mathcal{V}_0 = (v_{i+j})$ for $0 \leqslant i, j \leqslant k_0$ be two $(k_0+1) \times (k_0+1)$ matrices for $0 \leqslant i, j \leqslant k_0$. 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 $\mathcal{M}(x_0; h) = (\mathcal{M}_1(x_0; h)^{\mathrm{T}}, \ldots, \mathcal{M}_{k_0+1}(x_0; h)^{\mathrm{T}})^{\mathrm{T}}$, in which we have

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

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 $f_X(\cdot)$. Under the Log-Euclidean metric and conditions 1–4 in Appendix A, we have the following results.

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

(b) For $k_0=0$, under an additional condition 10 in Appendix A and that $f_X^{(1)}(x)$ is continuous in a neighbourhood of x_0, we have

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

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

(c) For $k_0>0$, under the conditions of part (b), we have

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

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

Theorem 1 delineates the asymptotic properties of $\hat{\alpha}_{\mathrm{IL}}(x_0;h)$ for $k_0\geqslant 0$, 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 $\hat{D}_{\mathrm{IL}}(x_0;h,0)$ are closely related to those of the Nadaraya–Watson estimator when both response and covariate are in Euclidean space (Fan, 1992). Since $\mathrm{vecs}[\log\{\hat{D}_{\mathrm{IL}}(x_0;h,k_0)\}]$ is a subvector of $\hat{\alpha}_{\mathrm{IL}}(x_0;h)$, we calculate the asymptotic average mean-squared error AMSE conditional on $\mathbf{x}=\{x_1,\ldots,x_n\}$ as

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

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

$$\mathrm{AMISE}[\log\{\hat{D}_{\mathrm{IL}}(\cdot;h,k_0)\}]=\int\mathrm{AMSE}[\log\{\hat{D}_{\mathrm{IL}}(x;h,k_0)\}]\,w(x)\,\mathrm{d}x.$$

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

By theorem 1, part (c),

$$\mathrm{AMSE}[\log\{\hat{D}_{\mathrm{IL}}(x_0;h,0)\}]=v_0\{nh\,f_X(x_0)\}^{-1}\,\mathrm{tr}\{\Sigma_{\mathcal{E}_D}(x_0)\}+h^4u_2^2\,\mathrm{tr}\{(\mathrm{vecs}[0.5\log\{D(x_0)\}^{(2)}+f_X^{(1)}(x_0)\,f_X(x_0)^{-1}\log\{D(x_0)\}^{(1)}])^{\otimes 2}\}.$$

For the intrinsic local linear estimator, $\mathrm{AMSE}[\log\{\hat{D}_{\mathrm{IL}}(x_0;h,1)\}]$ is given by $0.25h^4u_2^2\times\mathrm{tr}\{(\mathrm{vecs}[\log\{D(x_0)\}^{(2)}])^{\otimes 2}\}+v_0\{nh\,f_X(x_0)\}^{-1}\,\mathrm{tr}\{\Sigma_{\mathcal{E}_D}(x_0)\}$. 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, $h^2 u_2 f_X^{(1)}(x_0) f_X(x_0)^{-1} \times \mathrm{vecs}[\log\{D(x_0)\}^{(1)}]$, which depends on the marginal density $f_X(\cdot)$. Subsequently, we can obtain the optimal bandwidths, whose detailed expression can be found in the supplementary document.

4.2. Trace metric

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

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

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

where G is an $m \times m$ lower triangle matrix, $S \in \mathrm{Sym}^+(m)$ and $Y \in \mathrm{Sym}(m)$. Let $\alpha = (\alpha_G, \alpha_Y)$, in which $\alpha_G = \mathrm{vecs}(G)$ and $\alpha_Y = \mathrm{vecs}(Y)$. 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

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

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

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 1. Let $\mathcal{U}_2 = (u_{i+j})$ and $\mathcal{V}_2 = (v_{i+j})$ for $1 \leqslant i, j \leqslant k_0$ be two $k_0 \times k_0$ matrices. We define

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

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

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

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

Theorem 2. Suppose that x_0 is an interior point of $f_X(\cdot)$. Under the trace metric and conditions 1–8 in Appendix A, we have the following results.

(a) There are solutions $\hat{\alpha}_{\mathrm{IT}}(x_0; h)$ to the equation $\partial G_n\{\alpha_{\mathrm{T}}(x_0)\}/\partial \alpha_{\mathrm{T}}(x_0) = \mathbf{0}$ such that

$$\mathcal{H}\{\hat{\alpha}_{\mathrm{IT}}(x_0; h) - \alpha_{\mathrm{T}}(x_0)\}$$

converges to $\mathbf{0}$ in probability as $n \to \infty$.

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

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

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

(c) For $k_0 > 0$, if condition 9 in Appendix A is also true, we have

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

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

$$\mathcal{N}(x) = \begin{pmatrix} u_0\,\Psi_1(x) & \mathbf{u}\otimes\Psi_2(x) \\ \mathbf{u}^{\mathrm{T}}\otimes\Psi_2(x)^{\mathrm{T}} & \mathcal{U}_2\otimes\Psi_3(x) \end{pmatrix},$$

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

Theorem 2 delineates the asymptotic bias, covariance and asymptotic normality of $\hat{\alpha}_{\mathrm{IT}}(x_0; h)$ for $k_0 \geqslant 0$. On the basis of theorem 2, it is straightforward to derive the asymptotic bias, covariance and asymptotic normality of $\hat{D}_{\mathrm{IT}}(x_0; h, k_0)$ for $k_0 \geqslant 0$. Moreover, to have a direct comparison between the trace and Log-Euclidean metrics, we calculate the asymptotic biases and covariances of $\log\{\hat{D}_{\mathrm{IT}}(x_0; h, k_0)\}$ under these two metrics. Subsequently, we calculate AMSE[$\log\{\hat{D}_{\mathrm{IT}}(x_0; h, k_0)\}$] and AMISE[$\log\{\hat{D}_{\mathrm{IT}}(\cdot; h, k_0)\}$] for a given weight function $w(x)$. Minimizing AMSE[$\log\{\hat{D}_{\mathrm{IT}}(x_0; h, k_0)\}$] and AMISE[$\log\{\hat{D}_{\mathrm{IT}}(x_0; h, k_0)\}$] 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 $\hat{D}_{\mathrm{IT}}(x_0; h, 0)$ and the local linear estimator $\hat{D}_{\mathrm{IT}}(x_0; h, 1)$. It follows from the delta method that AMSE[$\log\{\hat{D}_{\mathrm{IT}}(x_0; h, 0)\}$] can be approximated as

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

where $G_D(x_0) = (\partial\mathrm{vec}[\log\{G(x_0)^{\otimes 2}\}]/\partial\mathrm{vecs}\{G(x_0)\}^{\mathrm{T}}]^{\mathrm{T}}$. The asymptotic bias and variance of $\hat{D}_{\mathrm{IT}}(x_0; h, 0)$ are similar to those of the Nadaraya–Watson estimator when the response is in Euclidean space (Fan, 1992). For the intrinsic local linear estimator,

$$\mathrm{AMSE}[\log\{\hat{D}_{\mathrm{IT}}(x_0; h, 1)\}] = 0.25h^4 u_2^2\,\mathrm{tr}([G_D(x_0)^{\mathrm{T}}\,\Psi_1(x_0)^{-1}\Psi_2^{\mathrm{T}}(x_0)\,\mathrm{vecs}\{Y^{(2)}(x_0)\}]^{\otimes 2}) + (nh)^{-1}\,\mathrm{tr}\{G_D(x_0)^{\otimes 2}\,\Omega_0(x_0)\}.$$

We consider the ILPRE near the edge of the support of $f_X(x)$. Without loss of generality, we assume that the design density $f_X(\cdot)$ has a bounded support [0, 1] and consider the left-hand boundary point $x_0 = dh$ for some positive constant d. The asymptotic consistency and normality of the ILPRE are valid for the boundary points after slight modifications to the definitions of u_k and v_k. Denote $u_{k,d} = \int_{-d}^{\infty} x^k\,K(x)\,\mathrm{d}x$ and $v_{k,d} = \int_{-d}^{\infty} x^k\,K^2(x)\,\mathrm{d}x$. 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 $\mathbf{c}_{k_0+2,d} = (u_{k_0+2,d}, \ldots, u_{2k_0+1,d})^{\mathrm{T}}$ and $\aleph_d(0+) = (u_{k_0+1,d}\,\Psi_2(0+), \mathbf{c}_{k_0+2,d}\otimes\Psi_3(0+))^{\mathrm{T}}\,\mathrm{vecs}\{Y^{(k_0+1)}(0+)\}$. For the boundary points, we have the following asymptotic results under the trace metric.

Theorem 3. Suppose that $x_0 = dh$ is a left-hand boundary point of $f_X(\cdot)$. Under the trace metric and conditions 1–8 in Appendix A, we have the following results.

(a) There are solutions, which are denoted by $\hat{\alpha}_{\mathrm{IT}}(x_0; h)$, to the equation

$$\partial G_n\{\alpha_{\mathrm{T}}(x_0)\}/\partial\alpha_{\mathrm{T}}(x_0) = \mathbf{0}$$

such that $\mathcal{H}\{\hat{\alpha}_{\mathrm{IT}}(x_0, h) - \alpha_{\mathrm{T}}(x_0)\}$ converges to $\mathbf{0}$ in probability as $n \to \infty$.

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

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

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

(c) For $k_0 > 0$, if the condition in Appendix A is also true, conditioning on $\mathbf{x} = \{x_1, \ldots, x_n\}$, we have

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

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

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

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

It follows from theorem 3, parts (b) and (c), that, when x_0 is at the boundary, the asymptotic average mean-squared errors of intrinsic local constant and linear estimators are respectively $\mathrm{AMSE}[\log\{\hat{D}_{\mathrm{IT}}(0+;h,0)\}]=O_p(h^2+n^{-1}h^{-1})$ and $\mathrm{AMSE}[\log\{\hat{D}_{\mathrm{IT}}(0+;h,1)\}]=O_p(h^4+n^{-1}h^{-1})$. 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 various 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 DTIs.

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

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

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

(a) *Riemannian log-normal model*: $S_i=G(x_i)\exp(\varepsilon_i)\,G(x_i)^{\mathrm{T}}$ follows the Riemannian log-normal distribution, where $D(x_i)=G(x_i)^{\otimes 2}$ and $\varepsilon_i\in\mathrm{Sym}(3)$ 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, $k=1,\ldots,31$, with b-factor $b_k=1000$ s mm^{-1} and four baselines with $b_k=0$ s mm^{-1} for $k=32,\ldots,35$. The baseline signal intensity W_0 was set at 1500. For a given diffusion tensor $D(x_i)$, $\varepsilon_{\mathrm{R},k}$ and $\varepsilon_{\mathrm{I},k}$ were independently simulated from a Gaussian random generator with mean 0 and standard deviation 60. The diffusion-

weighted signal was calculated as $W_{i,k} = \sqrt{([W_0 \exp\{-b_k r_k D(x_i) r_k\} + \varepsilon_{\mathrm{R},k}]^2 + \varepsilon_{\mathrm{I},k}^2)}$ for $k = 1, \ldots, 35$. 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 that were 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, whereas 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, $i = 1, \ldots, 50$, independently from an $N(0, 0.25)$ 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

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

Figs 2(a)–2(d) present the true SPD matrix data and a set of simulated data $\{(x_i, S_i) : i = 1, \ldots, 50\}$ under the three noise models. Each SPD matrix S_i at the point x_i is geometrically represented by an ellipsoid. In this representation, the lengths of the semiaxes of the ellipsoid equal the square root of the eigenvalues of an SPD matrix, whereas the eigenvectors define the direction of the three axes. In DTIs, the ellipsoidal representation is used to represent the local Brownian motion of water molecules in the brain. Isotropic diffusion is represented by a sphere, whereas 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 $\mathrm{AGD} = n^{-1}\Sigma_{i=1}^{n} g\{\hat{D}(x_i), D(x_i)\}$, 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 $\mathrm{LAGD}(x_i) = \Sigma_{j=1}^{100} g\{\hat{D}_{(j)}(x_i), D(x_i)\}/100$, where $\hat{D}_{(j)}(x_i)$ is the estimated SPD matrix at x_i based on the jth simulated replication. Although we chose all the three metrics for calculating AGD and LAGD, we present only those based on the Euclidean metric for brevity. The results for 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 various metrics and noise distributions. Figs 2(e)–2(m) display a set of the estimated SPD functions using local linear regression estimators under the three metrics for the three different noise models. Inspecting Figs 2(k)–2(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 that is shown in Fig. 2(d). However, for the other two noise models, our intrinsic local linear regression methods visually outperform the local linear regression based on

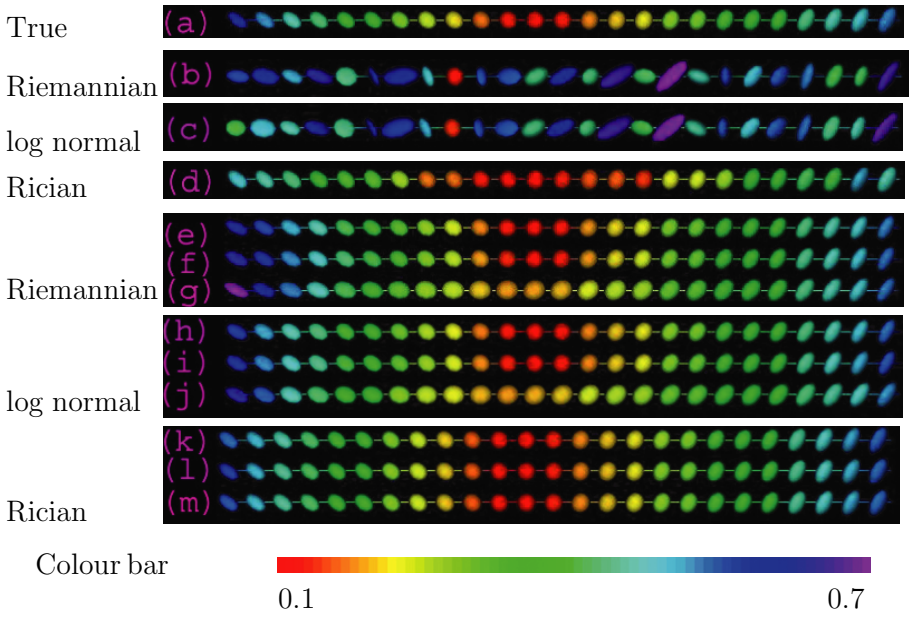

Fig. 2. Ellipsoidal representations of (a) the true SPD matrix data along the design points, the simulated SPD data along the design points under the three noise models (b) Riemannian log-normal, (c) log-normal and (d) Rician, and estimated SPD matrix data along the design points by using the three smoothing methods (e), (h), (k) ILPR under the trace metric, (f), (i), (l) ILPR under the Log-Euclidean metric and (g), (j), (m) LPR under the Euclidean metric, and under the three noise models (e), (f), (g) Riemannian log-normal, (h), (i), (j) log-normal and (k), (l), (m) Rician, coloured with FA-values defined in equation (37) in Section 5.4

the Euclidean metric. In particular, a clear swelling effect is observed for the Euclidean metric (Figs 2(g) and 2(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 that was given in Pasternak *et al.* (2010). However, our findings also suggest that both the trace and the Log-Euclidean metrics are appropriate for the non-parametric 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 non-parametric 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 Fig. 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 corresponding 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

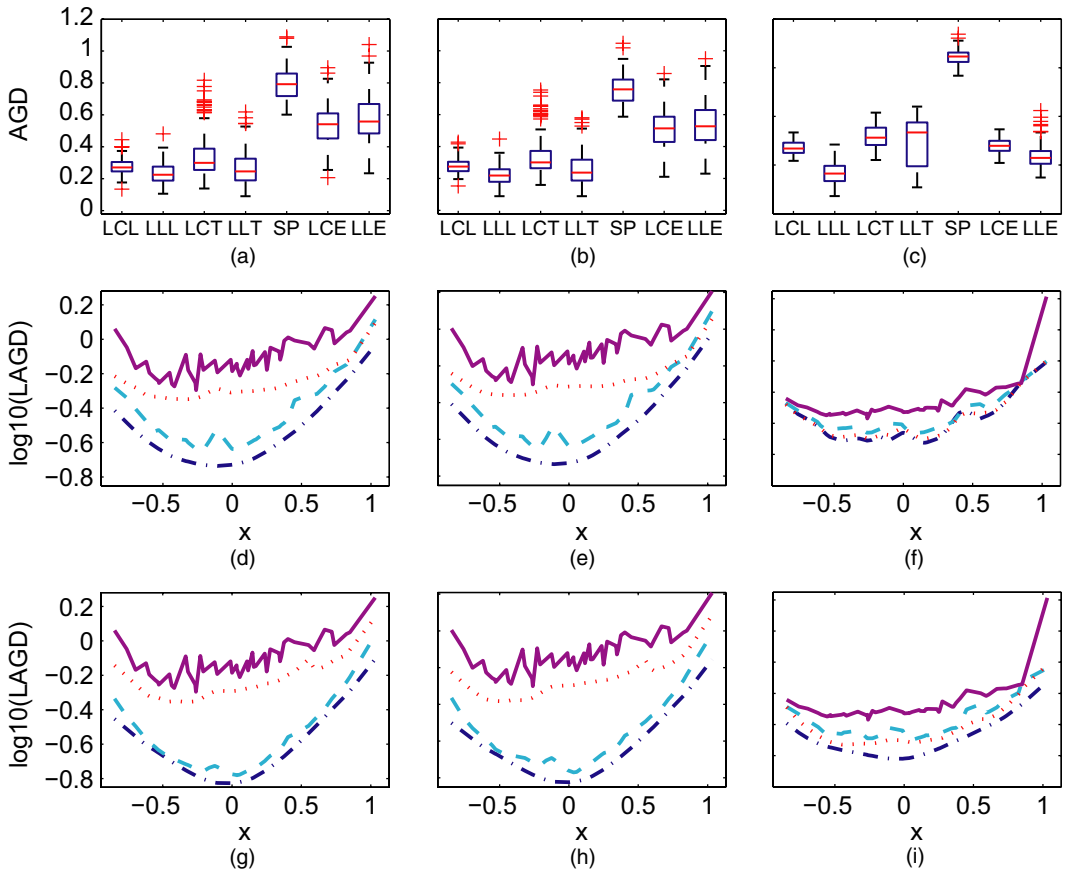

Fig. 3. Comparisons of the local constant and linear estimators under the three metrics and the tensor spline estimators under the three noise models: (a), (b), (c) boxplots of 1000 AGD obtained from seven estimators (LCL, local constant estimator under the Log-Euclidean metric; LCT, local constant estimator under the trace metric; LCE, local constant estimator under the Euclidean metric; LLL, local linear estimator under the Log-Euclidean metric; LLT, local linear estimator under the trace metric; LLE, local linear estimator under the Euclidean metric; SP, tensor spline estimator); (d), (e), (f) log(LAGD) curves based on LCL (– · – ·), LCT (– – –), LCE (·····) and SP (———); (g), (h), (i) log (LAGD) curves based on LLL (– · – ·), LLT (– – –), LLE (·····) and SP (———); (a), (d), (g) Riemannian log-normal noise; (b), (e), (h) log-normal noise; (c), (f), (i) Rician noise

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 level of noise. 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, whereas the tensor spline method does not converge. Thus, Fig. 4 presents only the results under

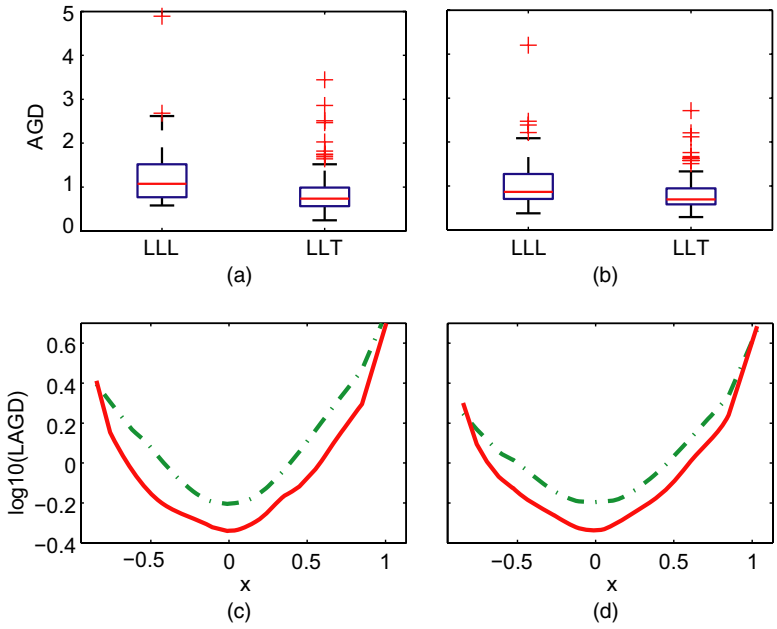

Fig. 4. Comparison of the intrinsic local linear estimators under the Log-Euclidean and trace metrics for the first two noise models at a higher level of noise (it shows that, at a high level of noise, the intrinsic local linear estimators under the trace metric slightly outperform those under the Log-Euclidean metric for the first two noise models): (a) boxplots of AGDs for LLL and LLT, Riemannian log-normal noise; (b) boxplots of AGDs for LLL and LLT, log-normal noise; (c) log(LAGD) curves of LLL (———) and LLT (· – · –), Riemannian log-normal noise; (d) log(LAGD) curves of LLL (———) and LLT (· – · –), log-normal noise

the trace and Log-Euclidean metrics. Inspecting Fig. 4 reveals that, when the level of noise 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 0 (all eigenvalues are the same) and 1 (two eigenvalues equal 0) and is given by

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

with eigenvalues λ_1, λ_2 and λ_3 and their average $\bar{\lambda}$. We compared two methods for smoothing FAs, which here are referred to as method A and method B. Method A first calculates the FAs from all SPD matrices and then uses the classic local linear regression in Euclidean space to smooth the FAs. Method B first applies the intrinsic local linear estimator to smooth SPD matrices and then calculates smoothed FA-curves on the basis of the smoothed SPD matrices. We further divided method B into three methods according to smoothing methods for SPD

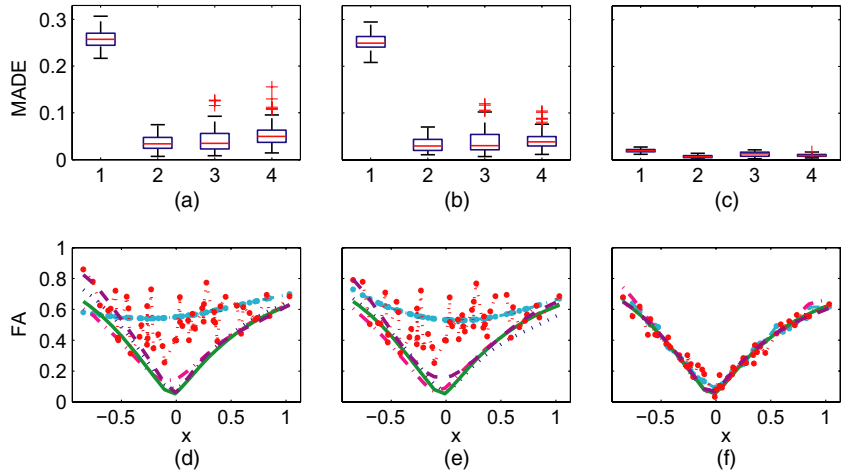

Fig. 5. Boxplots of the MADEs by using the four smoothing methods 1–4 representing the first–fourth methods based on 100 replications for the three noise models (a) Riemannian log-normal, (b) log-normal and (c) Rician, and smoothed FA-curves for the realizations with median MADE for the three noise models (d) Riemannian log-normal, (e) log-normal and (f) Rician: ··•··, raw FA-curve; ———, true FA-curve; –•–, estimated FA-curve for the first method; · · · · · · ·, estimated FA-curve for the second method; · – ·–, FA-curve for the third method; – – –, FA-curve for the fourth method

matrices under three metrics: the 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, which is defined as $\text{MADE} = n^{-1}\Sigma_{i=1}^{n}|\text{FA}(x_i) - \widehat{\text{FA}}(x_i)|$, where $\text{FA}(x_i)$ and $\widehat{\text{FA}}(x_i)$ are respectively the true and estimated FA-values across all design points.

Fig. 5 reveals that method B outperforms method A under 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 that is gained by directly smoothing DTI data over the post-smoothing method A. On the basis of the medians of MADEs (see Figs 5(d)–5(f)), method A cannot faithfully reconstruct the trend of the FA-curve for 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 0.

6. Human immunodeficiency virus imaging data

The aim of this analysis is to assess the integrity of white matter in human immunodeficiency virus (HIV) by using DTIs and our ILPRE. 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 the ILPRE along with its documentation will be accessible from the Web site `http://www.bios.unc.edu/research/bias`. We considered 46 subjects with 28 HIV positive subjects (20 males and eight females whose mean age is 40.0 years with standard deviation 5.6 years) and 18 healthy controls (nine males and nine females whose mean age is 41.2 years with standard deviation 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. 46 contiguous slices with a slice thickness of 2 mm covered a field of view of 256 mm^2 with an

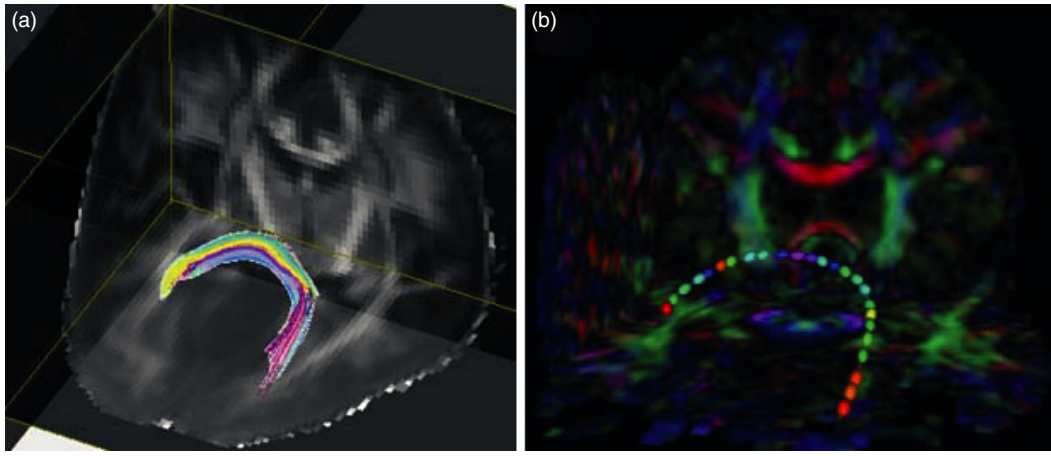

Fig. 6. Splenium of the *corpus callosum* in the analysis of the HIV DTI data and (b) ellipsoidal representation of the full tensors coloured with FA-values on the fibre tract from a selected subject

isotropic voxel size of $2 \times 2 \times 2$ mm^3. High resolution T1-weighted images were acquired by using a three-dimensional magnetization-prepared rapid acquisition with gradient echo 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 to be significantly different between the HIV positive and control groups, we examine the finite sample performance of our method by using this fibre tract. The tensors along the tract were extracted by using methodology described in Zhu *et al.* (2010). Fig. 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:

(a) registration and atlas construction,
(b) fibre tracking on the atlas and
(c) collection of tensor data on the atlas fibre 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. Fig. 7 shows the raw and estimated tensors along the fibre tracts from one subject. It is observed from the ellipsoidal representation of the diffusion tensor data (Fig. 7(a)) that the data are noisy. Figs 7(b)–7(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 fibre 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-hand side of the middle row. The estimated tensors in the middle row are very anisotropic when using the method under the Euclidean metric compared with 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 to apply standard statistical methods to carry out statistical inference on these diffusion measures directly. 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.

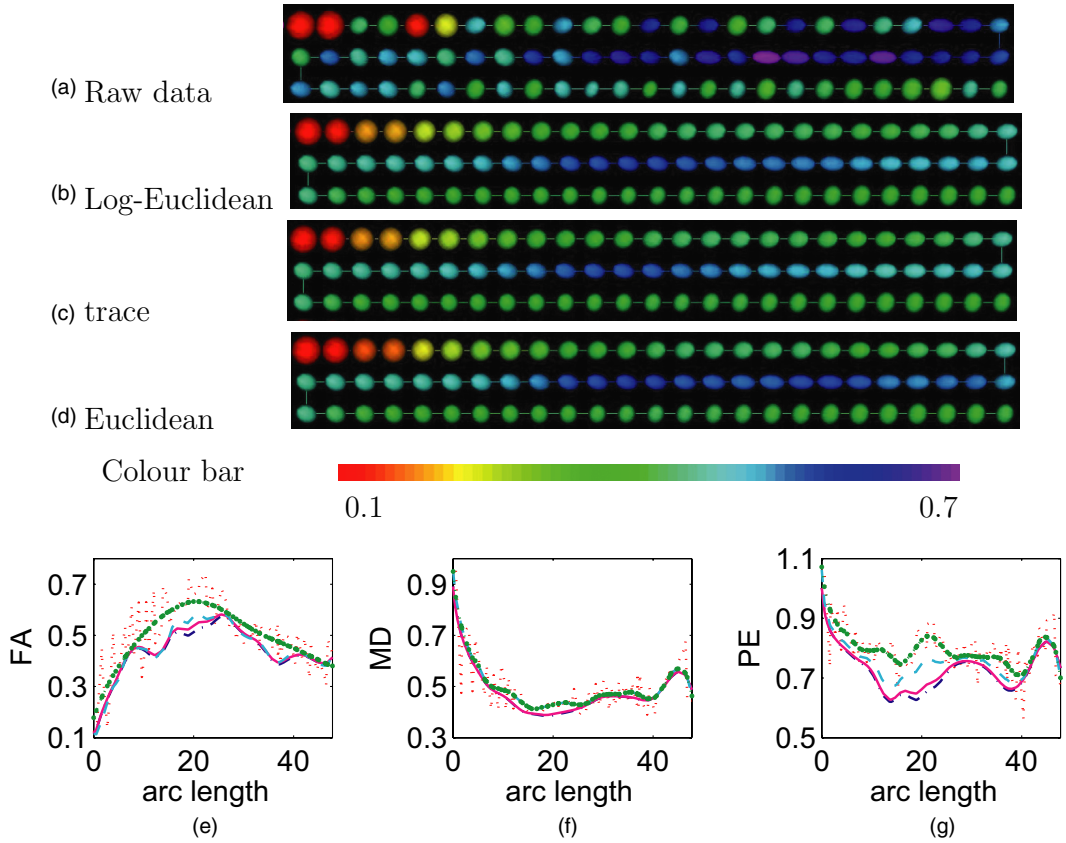

Fig. 7. (a) Ellipsoidal representations of the diffusion tensor data and estimated tensors by using the intrinsic local linear regression under (b) the Log-Euclidean, (c) trace and (d) Euclidean metrics along the splenium of the *corpus callosum*, coloured with FA-values (the estimated tensors in the middle right-hand part are more anisotropic by using the method under the Euclidean metric; each set of three 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) and (e) FAs, (f) MDs and (g) PEs derived from the raw tensor data (·····) and estimated tensors by using the intrinsic local linear regression under the trace (· – · –), Log-Euclidean (———) and Euclidean (– – –) metrics as the function of arc length along the splenium of the *corpus callosum*, with estimated FA-, MD- and PE-functions along the splenium of the *corpus callosum* by using the standard local linear regression for scalars (··•··)

Similarly to the simulation of Section 5.4, we applied method A to smooth FA-, MD- and PE-values along the selected tract directly and then we compared them with method B based on the trace, Log-Euclidean and Euclidean metrics. Figs 7(e)–7(g) show that there is no large difference for methods B under the three metrics. Both method A and method 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 the fact that FA- and PE-values are biased because of the well-known 'sorting' bias in estimating the eigenvalues of diffusion tensors (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. To detect meaningful group differences, registration is crucial. The 46 HIV DTI data that were used in our studies, including the splenium tracts and diffusion tensors on them, were registered in the same atlas space. Figs 8(a)–8(f) display the estimated mean diffusion

tensors along the fibre tract for the two groups by using intrinsic local linear regression for SPD matrices under both the Log-Euclidean and the trace metrics and also using 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 correspond to the differences in colour in Fig. 8(g), and the geodesic distances between estimated mean tensors at each point along the tract in Fig. 8(h). This result agrees with previous DTI findings that the splenium of the *corpus callosum* has been detected as abnormal for the HIV group (Filippi *et al.*, 2001; Chen *et al.*, 2009).

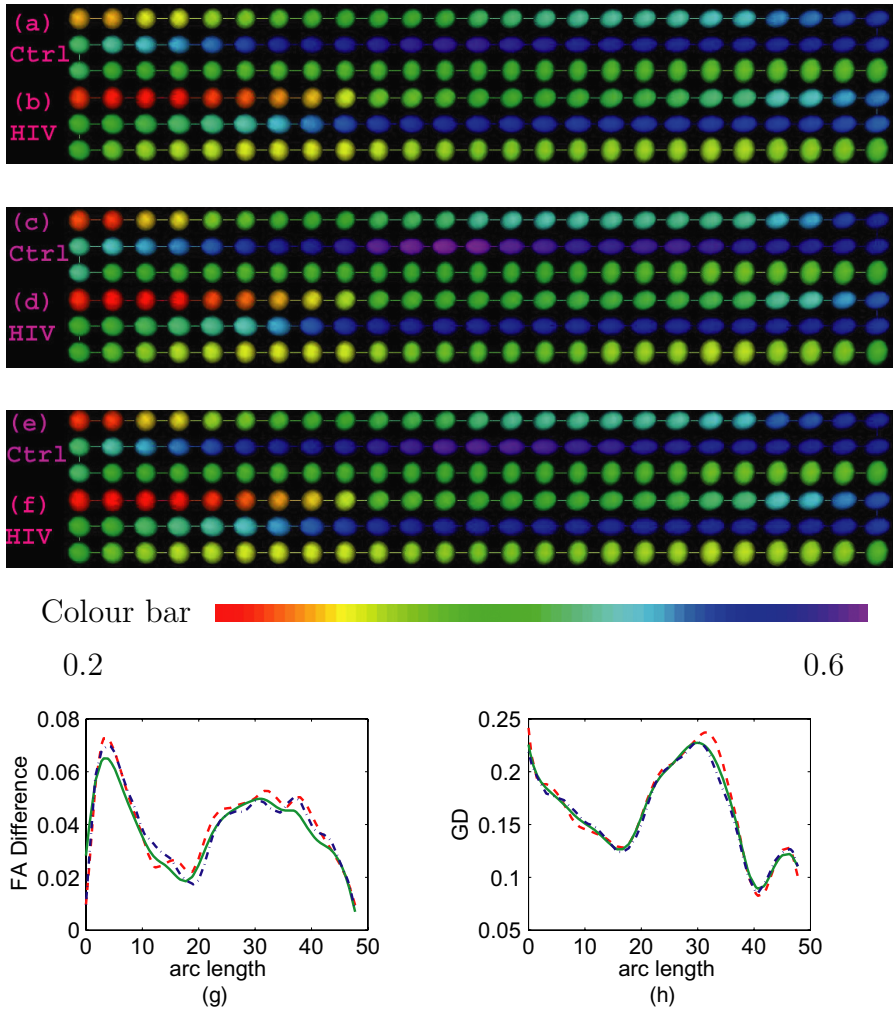

Fig. 8. Ellipsoidal representations of estimated mean tensors along the splenium of the *corpus callosum* for the control and HIV groups by using the intrinsic local linear regression under (a), (b) the Log-Euclidean, (c), (d) trace and (e), (f) Euclidean metrics coloured with FA-values (each set of three rows in (a)–(f) represents one track 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) and (g) FA-differences and (h) geodesic distances GD between the mean diffusion tensors of the HIV and control groups along the splenium of the *corpus callosum* under the Log-Euclidean (———), trace (– – –) and Euclidean (·–·–) metrics

7. Conclusion and discussion

We have systematically investigated the ILPR 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 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 of 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 splenium *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 that is proposed here and theory may be extended to other non-parametric 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 the tensor splines of Barmpoutis *et al.* (2007), it would be interesting to develop free-knot regression splines for SPD matrices and to 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 non-parametric methods, such as the ILPR, for the analysis of manifold-valued data and examination of the asymptotic properties of non-parametric estimates under different metrics should be pursued in future research.

Acknowledgements

We thank the Joint Editor, Associate Editor, two referees and Martha Skup for valuable suggestions, which helped to improve our presentation greatly. We are also grateful to Martin Styner, Pew-Thian Yap and Zhexing Liu for their help with the visualizations. This work was supported in part by a National Science Foundation grant and National Institutes of Health grants.

Appendix A: 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

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

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

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

Condition 2. The regression function $D(x) \in \mathrm{Sym}^+(m)$ has a continuous (k_0+1)th-order derivative in a neighbourhood of x_0.

Condition 3. The bandwidth $h \to 0$ and $nh \to \infty$.

Condition 4. The design density $f_X(\cdot)$ is continuous in a neighbourhood of x_0 and $f_X(x_0) > 0$.

Condition 5. The conditional density $f(S|X=x)$ is continuous in a neighbourhood of x_0.

Condition 6. $E[\partial_\alpha^2 \psi\{S, G, Y(X)\}|X=x]$ and $E([\partial_\alpha \psi\{S, G, Y(X)\}]^{\otimes 2}|X=x)$ are continuous in a neighbourhood of x_0.

Condition 7. The matrix

$$\mathcal{N}(x) = \begin{pmatrix} u_0 \Psi_1(x) & \mathbf{u} \otimes \Psi_2(x) \\ \mathbf{u}^{\mathrm{T}} \otimes \Psi_2(x)^{\mathrm{T}} & \mathcal{U}_2 \otimes \Psi_3(x) \end{pmatrix}$$

is positive definite in a neighbourhood of x_0.

Condition 8. Let $\|\cdot\|$ be the L_2-norm of a matrix, η_0 be a lower triangle matrix, $\eta_1 \in \mathrm{Sym}(m)$ and $U_\delta = \{(\eta_0, \eta_1) : \|\eta_0\|^2 + \|\eta_1\|^2 \leqslant \delta^2\}$. As $\delta \to 0$,

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

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

are uniformly in x in a neighbourhood of x_0.

Condition 9. There is a $b > 0$ such that $E[\|\partial_\alpha \psi\{S, G(x_0), Y(X)\}\|^{b+2}|X=x]$ is bounded in a neighbourhood of x_0.

Condition 10. $\Sigma_{\mathcal{E}_D}(x) = \mathrm{cov}\{\mathcal{E}_D(X)|X=x\}$ is continuous in a neighbourhood of x_0 and there is a $b > 0$ such that $E\{\|\mathcal{E}_D(X)\|^{b+2}|X=x\}$ is bounded in a neighbourhood of x_0.

Assumptions 1–10 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 condition 1 is not essential and can be removed if we put restriction on the tail of $K(\cdot)$. Condition 2 ensures that $Y(x) = \log\{G(x_0)^{-1} D(x) G(x_0)^{-\mathrm{T}}\}$, $G(x)$ and $\log\{D(x)\}$ have a continuous (k_0+1)th-order derivative in a neighbourhood of x_0. Moreover, assume that $f_X(\cdot)$ has a bounded support $[0, 1]$. All assumptions can be easily modified when x_0 is a boundary point, say left-hand boundary point $x_0 = dh$ or right-hand boundary point $x_0 = 1 - dh$ for some $d > 0$. For instance, we require that conditions 2–10 hold in the left-hand neighbourhood of 0 or the right-hand neighbourhood of 1. For condition 2, we also need to introduce $f_X(0+)$ as x_0 is the left-hand boundary point and $f_X(1-)$ as x_0 is the right-hand boundary point. For condition 7, $\mathcal{N}(x)$ is also needed to make some modifications. For simplicity, we omit these details.

References

Anderson, T. W. (2003) *An Introduction to Multivariate Statistical Analysis*, 3rd edn. Hoboken: Wiley.

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

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

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

Batchelor, P., Moakher, M., Atkinson, D., Calamante, F. and Connelly, A. (2005) A rigorous framework for diffusion tensor calculus. *Magn. Resnce Med.*, **53**, 221–225.

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

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

Davis, B. C., Bullitt, E., Fletcher, P. T. and Joshi, S. (2010) Population shape regression from random design data. *Int. J. Comput. Vis.*, **90**, 255–266.
Dryden, I. L., Koloydenko, A. and Zhou, D. (2009) Non-euclidean statistics for covariance matrices with applications to diffusion tensor imaging. *Ann. Appl. Statist.*, **3**, 1102–1123.
Fan, J. (1992) Design-adaptive nonparametric regression. *J. Am. Statist. Ass.*, **87**, 998–1004.
Fan, J. and Gijbels, I. (1996) *Local Polynomial Modelling and Its Applications*. London: Chapman and Hall.
Fan, J., Gijbels, I., Hu, T.-C. and Huang, L.-S. (1996) A study of variable bandwidth selection for local polynomial regression. *Statist. Sin.*, **6**, 113–127.
Filippi, C., Ulug, A., Ryan, E., Ferrando, S. and van Gorp, W. (2001) Diffusion tensor imaging of patients with hiv and normal-appearing white matter on mr images of the brain. *Am. J. Neuroradiol.*, **22**, 277–283.
Fingelkurts, A. A., Fingelkurts, A. A. and Kähkönen, S. (2005) Functional connectivity in the brain—is it an elusive concept? *Neurosci. Biobehav. Rev.*, **28**, 827–836.
Fletcher, P. and Joshi, S. (2007) Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Process.*, **87**, 250–262.
Fletcher, P., Joshi, S., Lu, C. and Pizer, S. (2004) Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Trans. Med. Imgng*, **23**, 995–1005.
Grenander, U. and Miller, M. I. (2007) *Pattern Theory from Representation to Inference*. Oxford: Oxford University Press.
Hall, P., Marron, J. S. and Park, B. U. (1992) Smoothed cross-validation. *Probab. Theor. Reltd Flds*, **92**, 1–20.
Hardle, W., Hall, P. and Marron, J. S. (1992) Regression smoothing parameters that are not far from their optimum. *J. Am. Statist. Ass.*, **87**, 227–233.
Jones, M., Marron, J. S. and Sheather, S. (1996) A brief survey of bandwidth selection for density estimation. *J. Am. Statist. Ass.*, **91**, 401–407.
Jupp, P. E. and Kent, J. T. (1987) Fitting smooth paths to spherical data. *Appl. Statist.*, **36**, 34–46.
Kim, P. T. and Richards, D. S. (2010) *Deconvolution Density Estimation on Spaces of Positive Definite Symmetric Matrices*. Hayward: Institute of Mathematical Statistics.
Kume, A., Dryden, I. L. and Le, H. (2007) Shape-space smoothing splines for planar landmark data. *Biometrika*, **94**, 513–528.
Lang, S. (1999) *Fundamentals of Differential Geometry*. New York: Springer.
Liang, F. (2010) Evolutionary stochastic approximation monte carlo for global optimization. *Statist. Comput.*, **21**, 375–393.
Park, B. U. and Marron, J. S. (1990) Comparision of data-driven bandwidth selectors. *J. Am. Statist. Ass.*, **85**, 66–72.
Pasternak, O., Sochen, N. and Basser, J. P. (2010) The effect of metric selection on the analysis of diffusion tensor mri data. *NeuroImage*, **49**, 2190–2204.
Pennec, X., Fillard, P. and Ayache, N. (2006) A riemannian framework for tensor computing. *Int. J. Comput. Visn*, **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. *Appl. Statist.*, **36**, 325–331.
Rice, J. (1984) Bandwidth choice for nonparametric regression. *Ann. Statist.*, **12**, 1215–1230.
Sakamoto, Y., Ishiguro, M. and Kitagawa, G. (1999) *Akaike Information Criterion Statistics*. New York: Springer.
Sangalli, L. M., Secchi, P., Vantini, S. and Veneziani, A. (2009) Efficient estimation of three-dimensional curves and their derivatives by free-knot regression splines applied to the analysis of inner carotid artery centrelines. *Appl. Statist.*, **58**, 285–306.
Schwartzman, A. (2006) Random ellipsoids and false discovery rates: statistics for diffusion tensor imaging data. *PhD Thesis*. Stanford University, Stanford.
Terras, A. (1988) *Harmonic Analysis on Symmetric Spaces and Applications II*. Berlin: Springer.
Wand, M. P. and Jones, M. C. (1995) *Kernel Smoothing*. London: Chapman and Hall.
Zhu, H., Chen, Y., Ibrahim, J. G., Li, Y and Lin, W. (2009) Intrinsic regression models for positive-definite matrices with applications to diffusion tensor imaging. *J. Am. Statist. Ass.*, **104**, 1203–1212.
Zhu, H., Styner, M., Tang, N., Liu, Z., Lin, W. and Gilmore, J. (2010) Frats: functional regression analysis of dti tract statistics. *IEEE Trans. Med. Imgng*, **29**, 1039–1049.
Zhu, H., Zhang, H., Ibrahim, J. G. and Peterson, B. G. (2007) Statistical analysis of diffusion tensors in diffusion-weighted magnetic resonance image data (with discussion). *J. Am. Statist. Ass.*, **102**, 1085–1102.

Supporting information

Additional 'supporting information' may be found in the on-line version of this article:

'Local Polynomial Regression for Symmetric Positive Definite Matrices (Supplementary Report)'.