ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistical Methodology
Series B

J. R. Statist. Soc. B (2017)
79, *Part* 2, *pp.* 463–482

Regression models on Riemannian symmetric spaces

Emil Cornea and Hongtu Zhu,

University of North Carolina at Chapel Hill, USA

Peter Kim

University of Guelph, Canada

and Joseph G. Ibrahim

University of North Carolina at Chapel Hill, USA

for the Alzheimer's Disease Neuroimaging Initiative

[Received February 2013. Final revision January 2016]

Summary. The paper develops a general regression framework for the analysis of manifold-valued response in a Riemannian symmetric space (RSS) and its association with multiple covariates of interest, such as age or gender, in Euclidean space. Such RSS-valued data arise frequently in medical imaging, surface modelling and computer vision, among many other fields. We develop an intrinsic regression model solely based on an intrinsic conditional moment assumption, avoiding specifying any parametric distribution in RSS. We propose various link functions to map from the Euclidean space of multiple covariates to the RSS of responses. We develop a two-stage procedure to calculate the parameter estimates and determine their asymptotic distributions. We construct the Wald and geodesic test statistics to test hypotheses of unknown parameters. We systematically investigate the geometric invariant property of these estimates and test statistics. Simulation studies and a real data analysis are used to evaluate the finite sample properties of our methods.

Keywords: Generalized method of moment; Geodesic; Group action; Lie group; Link function; Regression; Riemannian symmetric space

1. Introduction

Manifold-valued responses in curved spaces frequently arise in many disciplines including medical imaging, computational biology and computer vision, among many others. For instance, in medical and molecular imaging, it is interesting to delineate the changes in the shape and anatomy of a molecule. See Fig. 1 for four examples of manifold-valued data. Regression analysis is a fundamental statistical tool for relating a response variable to a covariate, such as age and gender. In particular, when both the response and the covariate(s) are in Euclidean space, the classical linear regression model and its variants have been widely used in various fields (McCullagh and Nelder, 1989). However, when the response is in a Riemannian symmetric space (RSS) and the covariates are in Euclidean space, developing regression models for this type of

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

 1369–7412/17/79463

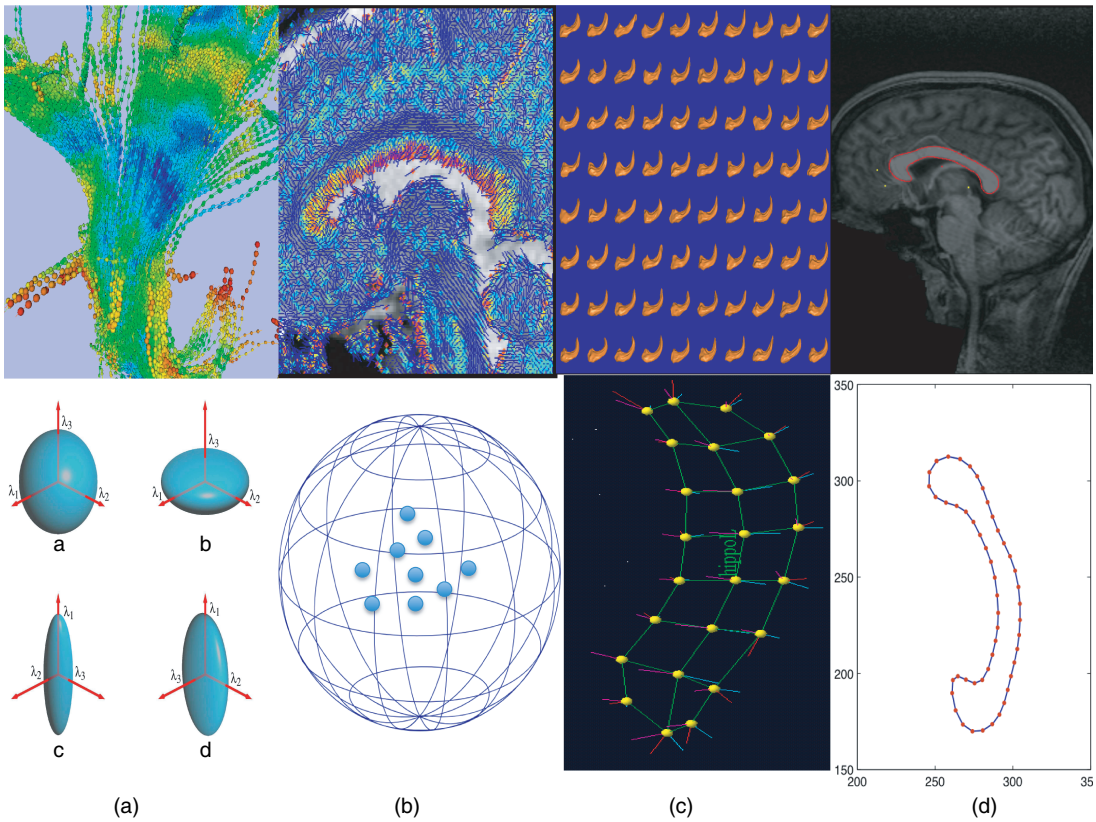


Fig. 1. Examples of manifold-valued data: (a) diffusion tensors along white matter fibre bundles and their ellipsoidal representations; (b) principal direction map of a selected slice for a randomly selected subject and the directional representation of some randomly selected principal directions on S^2; (c) median representations and median atoms of a hippocampus from a randomly selected subject; (d) an extracted contour and landmarks along the contour of the mid-saggital section of the *corpus callosum* from a randomly selected subject

data raises both computational and theoretical challenges. The aim of this paper is to develop a general regression framework to address these challenges.

Little has been done on the regression analyses of manifold-valued response data. The existing statistical methods for general manifold-valued data are primarily developed to characterize the population 'mean' and 'variation' across groups (Bhattacharya and Patrangenaru, 2003, 2005; Fletcher *et al.*, 2004; Dryden and Mardia, 1998; Huckemann *et al.*, 2010). In contrast, even for the 'simplest' directional data, there is a sparse literature on regression modelling of a single directional response and multiple covariates (Mardia and Jupp, 2000). In addition, these regression models of directional data are primarily based on a specific parametric distribution, such as the von Mises–Fisher distribution (Mardia and Jupp, 2000; Kent, 1982). However, it can be very challenging to assume useful parametric distributions for general manifold-valued data, and thus it is difficult to generalize these regression models of directional data to general manifold-valued data except for some specific manifolds (Shi *et al.*, 2009, 2012; Fletcher, 2013; Kim *et al.*, 2014; Zhu *et al.*, 2009). There is also a great interest in developing non-parametric regression models for manifold-valued response data and multiple covariates (Bhattacharya

and Dunson, 2010, 2012; Samir *et al.*, 2012; Su *et al.*, 2012; Muralidharan and Fletcher, 2012; Machado and Leite, 2006; Machado *et al.*, 2010; Yuan *et al.*, 2012).

An intriguing question is whether there is a general regression framework for manifold-valued response in an RSS and covariates in a multi-dimensional Euclidean space. The aim of this paper is to give an affirmative answer to such a question. The theoretical development is challenging but of great interest for carrying out statistical inferences on regression coefficients. We make five major contributions in this paper as follows.

(a) We propose an intrinsic regression model solely based on an intrinsic conditional moment for the response in an RSS, thus avoiding specifying any parametric distributions in a general RSS—the model can handle multiple covariates in Euclidean space.
(b) We develop several 'efficient' estimation methods for estimating the regression coefficients in this intrinsic model.
(c) We develop several test statistics for testing linear hypotheses of the regression coefficients.
(d) We develop a general asymptotic framework for the estimates of the regression coefficients and test statistics.
(e) We systematically investigate the geometrical properties (e.g. chart invariance) of these parameter estimates and test statistics.

The paper is organized as follows. In Section 2, we review the basic notion and concepts of Riemannian geometry. In Section 3, we propose the intrinsic regression models and propose various link functions for several specific RSSs. In Section 4, we develop estimation and test procedures for the intrinsic regression models. In Section 5, we carry out a detailed data analysis on the shape of *corpus callosum* (CC) contours obtained from the 'Alzheimer's disease neuroimaging initiative' (ADNI) study. Finally, we conclude with some discussions in Section 6. Technical conditions, simulation studies, theoretical examples and proofs are deferred to the on-line supplementary document. Our code and data are available from `http://www.bios.unc.edu/research/bias/software.html`.

2. Differential geometry preliminaries

We briefly review some basic facts about the theory of Riemannian geometry and present more technical details in the on-line supplementary document. The reader can refer to Helgason (1978), Spivak (1979) and Lang (1999) for more details.

Let $\mathcal{M}$ be a smooth manifold and $d_{\mathcal{M}}$ be its dimension. A tangent vector of $\mathcal{M}$ at $p \in \mathcal{M}$ is defined as the derivative of a smooth curve $\gamma(t)$ with respect to t evaluated at $t = 0$, denoted as $\dot{\gamma}(0)$, where $\gamma(0) = p$. The tangent space of $\mathcal{M}$ at p is denoted as $T_p\mathcal{M}$ and is the set of all tangent vectors at p.

A Riemannian manifold $(\mathcal{M}, m)$ is a smooth manifold together with a family of inner products, $m = \{m_p\}$, on the tangent spaces $T_p\mathcal{M}$s that vary smoothly with $p \in \mathcal{M}$, and m is called a Riemannian metric. This metric induces a so-called *geodesic distance* $\mathrm{dist}_{\mathcal{M}}$ on $\mathcal{M}$. The geodesics are, by definition, the locally distance minimizing paths. If the metric space $(\mathcal{M}, \mathrm{dist}_{\mathcal{M}})$ is complete, the *exponential map* at p is defined on the tangent space $T_p\mathcal{M}$ by $\mathrm{Exp}_p^{\mathcal{M}}(V) = \gamma(1; p, V)$, where $t \to \gamma(t; p, V)$ is the geodesic with $\gamma(0; p, V) = p$ and $\dot{\gamma}(0; p, V) = V$. $\mathrm{Exp}_p^{\mathcal{M}}$ is well defined near $\mathbf{0}$ and is a diffeomorphism on an open neighbourhood $\mathcal{V}$ of the origin in $T_p\mathcal{M}$ onto $\mathcal{U}$ with $\mathcal{V}$ such that $tV \in \mathcal{V}$ for $0 \leqslant t \leqslant 1$ and $V \in \mathcal{V}$. The inverse map is the *logarithmic map* at p, which is denoted by $\mathrm{Log}_p^{\mathcal{M}}$. Then, for $q \in \mathcal{U}$, $\mathrm{dist}_{\mathcal{M}}(p, q) = \|\mathrm{Log}_p^{\mathcal{M}}(q)\|_{m_p}$. The *radius of injectivity* of $\mathcal{M}$ at p, which is denoted by $\rho^*(\mathcal{M}, p)$, is the largest $r > 0$ such that $\mathrm{Exp}_p^{\mathcal{M}}$ is a diffeomorphism on the open ball $B_{m_p}(\mathbf{0}, r) \subset T_p\mathcal{M}$ onto an open set in $\mathcal{M}$ near p. Any basis in the tangent space

$T_p\mathcal{M}$ induces an isomorphism from $T_p\mathcal{M}$ to $R^{d_\mathcal{M}}$, and then the logarithmic map Log_p provides a local chart near p. If $T_p\mathcal{M}$ is endowed with an orthonormal basis, such a chart is called a *normal chart* and the co-ordinates are called *normal co-ordinates*.

A Lie group G is a group together with a smooth manifold structure such that the operations of multiplication and inversion are smooth maps. Many common geometric transformations of Euclidean spaces that form Lie groups include rotations, translations, dilations and affine transformations on R^d. In general, Lie groups can be used to describe transformations of smooth manifolds.

An RSS is a connected Riemannian manifold $\mathcal{M}$ with the property that, at each point, the mapping that reverses geodesics through that point is an isometry. Examples of RSSs include Euclidean spaces R^k, spheres S^k, projective spaces PR^k and hyperbolic spaces H^k, each with their standard Riemannian metrics. Symmetric spaces arise naturally from Lie group actions on manifolds; see Helgason (1978). Given a smooth manifold $\mathcal{M}$ and a Lie group G, a *smooth group action of* G *on* $\mathcal{M}$ is a smooth mapping $G \times \mathcal{M} \to \mathcal{M}$, $(a, p) \mapsto a \cdot p$ such that $e \cdot p = p$ and $(aa') \cdot p = a \cdot (a' \cdot p)$ for all $a, a' \in G$ and all $p \in \mathcal{M}$, where e is the unit element of G. The group action should be interpreted as a group of transformations of the manifold $\mathcal{M}$, namely $\{L_a\}_{a \in G}$, $L_a(p) = a \cdot p$ for $p \in \mathcal{M}$. L_a is a smooth transformation on $\mathcal{M}$ and its inverse is $L_{a^{-1}}$. The *orbit* of a point $p \in \mathcal{M}$ is defined as $G(p) = \{a \cdot p | a \in G\}$. The orbits form a partition of $\mathcal{M}$. If $\mathcal{M}$ consists of a single orbit, the group action is *transitive* or G acts *transitively* on $\mathcal{M}$, and we call $\mathcal{M}$ a *homogeneous* space. The *isotropy* subgroup of a point $p \in \mathcal{M}$ is defined as $G_p = \{a \in G | a \cdot p = p\}$. When G is a connected group of isometries of the RSS $\mathcal{M}$, $\mathcal{M}$ can always be viewed as a homogeneous space, $\mathcal{M} \cong G/G_p$, and the isotropy subgroup G_p is compact.

From now on, we shall assume that the manifold $\mathcal{M}$ is an RSS and $\mathcal{M} = G/G_p$ with G being a Lie group of isometries acting transitively on $\mathcal{M}$. Geodesics on $\mathcal{M}$ are computed through the action of G on $\mathcal{M}$. Owing to the transitive action of the group G of isometries on $\mathcal{M}$, it suffices to consider only the geodesic starting at the base point p. Geodesics on $\mathcal{M}$ starting from p are the images of the action of a one-parameter subgroup of G acting on the base point p, i.e. for any geodesic γ on $\mathcal{M}$, $\gamma(\cdot) : R \to \mathcal{M}$, starting from p, there is a one-parameter subgroup $c(\cdot) : R \to G$ such that $\gamma(t) = c(t) \cdot p$ for all $t \in R$.

3. Intrinsic regression model

Let $(\mathcal{M}, m)$ be a (C^∞) RSS of dimension $d_\mathcal{M}$ and geodesically complete with an inner product m_p and let G be a Lie group of isometries acting smoothly and transitively on $\mathcal{M}$ with the identity element e.

3.1. Formulation

Consider n independent observations $(y_1, \mathbf{x}_1), \ldots, (y_n, \mathbf{x}_n)$, where y_i is the $\mathcal{M}$-valued response variable and $\mathbf{x}_i = (x_{i1}, \ldots, x_{id_\mathbf{x}})^\mathrm{T}$ is a $d_\mathbf{x} \times 1$ vector of multiple covariates. Our objective is to introduce an intrinsic regression model for RSS responses and multiple covariates of interest from n subjects.

The specification of the intrinsic regression model involves three key steps including

(a) a link function mapping from the space of covariates to $\mathcal{M}$,
(b) the definition of a residual and
(c) the action of transporting all residuals to a common space.

First, we explicitly formalize the link function. From now on, all covariates have been centred to have mean 0. We consider a single-centre link function given by

$$\boldsymbol{\mu}(\mathbf{x}, q, \boldsymbol{\beta}) : R^{d_{\mathbf{x}}} \times \mathcal{M} \times R^{d_{\boldsymbol{\beta}}} \to \mathcal{M}, \tag{1}$$

where $\boldsymbol{\mu}(\mathbf{x}, q, \boldsymbol{\beta})$ is a known link function, $q \in \mathcal{M}$ can be regarded as the intercept or centre and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{d_{\boldsymbol{\beta}}})^{\mathrm{T}}$ is a $d_{\boldsymbol{\beta}} \times 1$ vector of regression coefficients. Moreover, it is assumed that $\boldsymbol{\mu}(\mathbf{x}, q, \boldsymbol{\beta})$ satisfies a single-centre property as follows:

$$\boldsymbol{\mu}(\mathbf{0}, q, \boldsymbol{\beta}) = \boldsymbol{\mu}(\mathbf{x}, q, \mathbf{0}) = q. \tag{2}$$

When the regression coefficient vector $\boldsymbol{\beta}$ equals $\mathbf{0}$, the link function is independent of the covariates and, thus, it reduces to the single centre (or 'mean') $q \in \mathcal{M}$. When all the covariates are equal to 0, the link function is independent of the regression coefficients and reduces to the centre $q \in \mathcal{M}$. An example of the single-centre link function is the geodesic link function in Kim *et al.* (2014) and Fletcher (2013), which is given by

$$\boldsymbol{\mu}(\mathbf{x}, q, \boldsymbol{\beta}) = \mathrm{Exp}_q^{\mathcal{M}}\left(\sum_{k=1}^{d_{\mathbf{x}}} x_{ik} V_k \right), \tag{3}$$

where V_ks are tangent vectors in $T_q\mathcal{M}$ and $\boldsymbol{\beta}$ includes all unknown parameters associated with the tangent vectors.

More generally, we shall consider a multicentre link function to account for the presence of multiple discrete covariates and even a general link function defined as $\boldsymbol{\mu}(\mathbf{x}, \boldsymbol{\theta}) : R^{d_{\mathbf{x}}} \times \Theta \to M$, where $\boldsymbol{\theta}$ is a vector of unknown parameters in a parameter space Θ. For the multicentre link function, $\boldsymbol{\theta}$ contains all unknown intercepts, denoted as $q(\mathbf{x}_D)$, corresponding to each discrete covariate class and all regression parameters $\boldsymbol{\beta}$ corresponding to continuous covariates and their potential interactions with the discrete variables. However, the details on these link functions are presented in the on-line supplementary document, and here, for notational simplicity, we focus on function (1) from now on.

Second, we introduce a definition of 'residual' to ensure that $\boldsymbol{\mu}(\mathbf{x}_i, q, \boldsymbol{\beta})$ is the proper 'conditional mean' of y_i given $\mathbf{x}_i$, which is the key concept of many regression models (McCullagh and Nelder, 1989). For instance, in the classical linear regression model, the response can be written as the sum of the regression function and a residual term and the regression function is the conditional mean of the response only when the conditional mean of the residual is equal to 0. Given the points y_i and $\boldsymbol{\mu}(\mathbf{x}_i, q, \boldsymbol{\beta})$ on an RSS $\mathcal{M}$, we need to define the residual as 'a difference' between y_i and $\boldsymbol{\mu}(\mathbf{x}_i, q, \boldsymbol{\beta})$. Assume that y_i and $\boldsymbol{\mu}(\mathbf{x}_i, q, \boldsymbol{\beta})$ are 'sufficiently close' to each other in the sense that there is an open ball $B(0, \rho) \subset T_{\boldsymbol{\mu}(\mathbf{x}_i, q, \boldsymbol{\beta})}\mathcal{M}$ such that, for all $i = 1, \ldots, n$,

$$y_i \in \mathrm{Exp}_{\boldsymbol{\mu}(\mathbf{x}_i, q, \boldsymbol{\beta})} \{B(0, \rho)\} \tag{4}$$

or

$$\mathrm{Log}_{\boldsymbol{\mu}(\mathbf{x}_i, q, \boldsymbol{\beta})}(y_i) \subset B(0, \rho).$$

However, according to a result in Le and Barden (2014), $\mathrm{Log}_{\boldsymbol{\mu}(\mathbf{x}_i, q, \boldsymbol{\beta})}(y_i)$ is well defined under some very mild conditions, which require that $\int_{\mathcal{M}} \mathrm{dist}_{\mathcal{M}}(p, y_i)^2 \, \mathrm{d}\nu(y_i)$ be finite and achieve a local minimum at $\boldsymbol{\mu}(\mathbf{x}_i, q, \boldsymbol{\beta})$, where $\nu(y_i)$ is any finite measure of y_i on $\mathcal{M}$. Thus, $\mathrm{Log}_{\boldsymbol{\mu}(\mathbf{x}_i, q, \boldsymbol{\beta})}(y_i)$ makes it a good candidate to play the role of a 'residual'. These residuals, however, lie on different tangent spaces from $\mathcal{M}$, so it is difficult to carry out a multivariate analysis of these residuals.

Third, since $\mathcal{M}$ is an RSS, this enables us to 'transport' all the residuals, separately, to a common space, say $T_p\mathcal{M}$, by exploiting the fact that the parallel transport along the geodesics can be expressed in terms of the action of G on $\mathcal{M}$. Indeed, since $\mathcal{M}$ is a symmetric space, the base point p and the point $\boldsymbol{\mu}(\mathbf{x}_i, q, \boldsymbol{\beta})$ can be joined in $\mathcal{M}$ by a geodesic, which can be seen as the action of a one-parameter subgroup $c(t; \mathbf{x}_i, q, \boldsymbol{\beta})$ of G such that $c(1; \mathbf{x}_i, q, \boldsymbol{\beta}) \cdot p = \boldsymbol{\mu}(\mathbf{x}_i, q, \boldsymbol{\beta})$.

We define the *rotated residual* $\mathcal{E}(y_i, \mathbf{x}_i; q, \boldsymbol{\beta})$ of $y_i \in \mathcal{M}$ with respect to $\boldsymbol{\mu}(\mathbf{x}_i, q, \boldsymbol{\beta})$ as the parallel

transport of the actual residual, $\mathrm{Log}_{\boldsymbol{\mu}(\mathbf{x}_i,q,\beta)}(y_i)$, along the geodesic from the conditional mean $\boldsymbol{\mu}(\mathbf{x}_i, q, \boldsymbol{\beta})$, to the base point p, i.e.

$$\mathcal{E}(y_i, \mathbf{x}_i; q, \boldsymbol{\beta}) = \mathcal{E}_i(q, \boldsymbol{\beta}) := \mathrm{Log}_p\{c(1; \mathbf{x}_i, q, \boldsymbol{\beta})^{-1} y_i\} \in T_p M \tag{5}$$

for $i = 1, \ldots, n$, where $T_p\mathcal{M}$ is identified with $R^{d_{\mathcal{M}}}$. The *intrinsic regression model* on $\mathcal{M}$ is defined by

$$E[\mathcal{E}(y_i, \mathbf{x}_i; q_*, \boldsymbol{\beta}_*)|\mathbf{x}_i] = 0, \tag{6}$$

where $(q_*, \boldsymbol{\beta}_*)$ denotes the true value of $(q, \boldsymbol{\beta})$ and the expectation is taken with respect to the conditional distribution of y_i given $\mathbf{x}_i$. Model (6) is equivalent to $E[\mathrm{Log}_{\boldsymbol{\mu}(\mathbf{x}_i,q_*,\boldsymbol{\beta}_*)}(y_i)|\mathbf{x}_i] = 0$ for $i = 1, \ldots, n$, since the tangent map of the action of $c(1; \mathbf{x}_i, q_*, \boldsymbol{\beta}_*)^{-1}$ on $\mathcal{M}$ is an isomorphism of linear spaces (invariant under the metric m) between the fibres of the tangent bundle $T\mathcal{M}$. This model does not assume any parametric distribution for y_i given $\mathbf{x}_i$, and thus it allows for a large class of distributions. The model is essentially semiparametric, since the joint distribution of $(y, \mathbf{x})$ is not restricted except by the zero conditional moment requirement in equation (6).

3.2. A theoretical example: the unit sphere S^k

We investigate the intrinsic regression model for S^k-valued responses and include several other examples in the on-line supplementary document. We review some basic facts about the geometric structure of $\mathcal{M} = S^k = \{x \in R^{k+1} : \|x\|_2 = 1\}$ (Shi *et al.*, 2012; Mardia and Jupp, 2000; Healy and Kim, 1996; Huckemann *et al.*, 2010). For $q \in S^k$, $T_q S^k$ is given by $T_q S^k = \{\mathbf{v} \in R^{k+1} : \mathbf{v}^{\mathrm{T}} q = 0\}$. The canonical Riemannian metric on S^k is that induced by the canonical inner product on R^{k+1}. Under this metric, the geodesic distance between any two points q and q' is equal to $\psi_{q,q'} = \cos^{-1}(q^{\mathrm{T}} q')$. If the points are not antipodal (i.e. $q' \neq -q$), then there is a unique geodesic path that joins them. Therefore, the radius of injectivity is $\rho^*(S^k) = \pi$. For $\mathbf{v} \in T_q S^k$, the Riemannian exponential map is given by $\mathrm{Exp}_q(\mathbf{v}) = \cos(\|\mathbf{v}\|)q + \sin(\|\mathbf{v}\|)\mathbf{v}/\|\mathbf{v}\|$. (Here, $\sin(0)/0 = 1$.) If q and q' are not antipodal, the Riemannian logarithmic map is given by $\mathrm{Log}_q(q') = \cos^{-1}(q^{\mathrm{T}} q')\mathbf{v}/\|\mathbf{v}\|$, where $\mathbf{v} = q' - (q^{\mathrm{T}} q')q \neq 0$.

The special orthogonal group $G = \mathrm{SO}(k+1)$ is a group of isometries on S^k and acts transitively on S^k via the left matrix multiplication. Specifically, the two-dimensional rotation matrix $R_{q,q'} \in \mathrm{SO}(k+1)$ which rotates q to $q' \in S^k$, along the great circle passing through q and q', is given by

$$R_{q,q'} = \mathbf{I}_{k+1} + \sin(\psi_{q,q'})(q'\tilde{q}^{\mathrm{T}} - \tilde{q}q'^{\mathrm{T}}) + \{\cos(\psi_{q,q'}) - 1\}(q'q'^{\mathrm{T}} + \tilde{q}\tilde{q}^{\mathrm{T}}),$$

where $\tilde{q} = \{q - (q^{\mathrm{T}} q')q'\}/\sqrt{\{1 - (q^{\mathrm{T}} q')^2\}}$. Thus, $q' = R_{q,q'} q$ and $R_{q,q'}\mathbf{v} \in T_{q'} S^k$, for any $\mathbf{v} \in T_q S^k$. Moreover, $(-\pi, \pi) \ni t \mapsto c_{q',q}(t) \cdot q'$ is the unique geodesic curve in S^k joining q' with q, where $c_{q',q}(t)$ takes the form $c_{q',q}(t) = \mathbf{I}_{k+1} + \sin(t)(\tilde{q}q'^{\mathrm{T}} - q'\tilde{q}^{\mathrm{T}}) + \{\cos(t) - 1\}(q'q'^{\mathrm{T}} + \tilde{q}\tilde{q}^{\mathrm{T}})$.

Suppose that we observe $\{(y_i, \mathbf{x}_i) : i = 1, \ldots, n\}$, where $y_i \in S^k$ for all i. We introduce the three key components of our intrinsic regression model for S^k-valued responses. First, we consider several examples of the general link function $\boldsymbol{\mu}(\mathbf{x}_i, \boldsymbol{\theta})$. Specifically, without loss of generality, we fix the 'north pole' $p = (0, \ldots, 0, 1)^{\mathrm{T}} \in R^{k+1}$ as a base point. Let $\mathbf{e}_j$ be the $(k+1) \times 1$ vector with a 1 at the jth component and a 0 otherwise for $j = 1, \ldots, k$. Let $q \in S^k$ be the 'centre', and we consider two link functions as follows:

$$\begin{aligned} \boldsymbol{\mu}(\mathbf{x}_i, q, \boldsymbol{\beta}) &= \mathrm{Exp}_q\left[\sum_{j=1}^{k} f(\mathbf{x}_i, \boldsymbol{\beta})_j c_{p,q}\{\cos^{-1}(p^{\mathrm{T}} q)\}\mathbf{e}_j\right], \\ \boldsymbol{\mu}(\mathbf{x}_i, q, \boldsymbol{\beta}) &= c_{-p,q}\{\cos^{-1}(-p^{\mathrm{T}} q)\} T_{st,-p}^{-1}\{(f(\mathbf{x}_i, \boldsymbol{\beta})^{\mathrm{T}}, -1)^{\mathrm{T}}\}, \end{aligned} \tag{7}$$

where $T_{st,-p}$ is the stereographic projection mapping from $S^k \setminus \{p\}$ onto the d-dimensional hyperplane $R^k \times \{-1\}$ and $f(\mathbf{x}, \boldsymbol{\beta})$ is a function mapping from $R^{d_\mathbf{x}} \times R^{d_\beta}$ to R^k with $f(\mathbf{0}, \cdot) = f(\cdot, \mathbf{0}) = \mathbf{0}$. A simple example of $f(\cdot, \cdot)$ is $f(\mathbf{x}_i, \boldsymbol{\beta}) = B\mathbf{x}_i$, where B is a $k \times d_\mathbf{x}$ matrix of regression coefficients and $\boldsymbol{\beta}$ includes all components of B.

Second, we define residuals for our intrinsic regression model. The residual in model (4) requires that y_i is not antipodal to $\boldsymbol{\mu}(\mathbf{x}_i, \boldsymbol{\theta})$. In this case, the residual for the ith subject is given by $\mathrm{Log}_{\boldsymbol{\mu}(\mathbf{x}_i,\boldsymbol{\theta})}(y_i) = \cos^{-1}\{\boldsymbol{\mu}(\mathbf{x}_i, \boldsymbol{\theta})^{\mathrm{T}} y_i\}\mathbf{v}_i / \|\mathbf{v}_i\|$, where $\mathbf{v}_i = y_i - \{\boldsymbol{\mu}(\mathbf{x}_i, \boldsymbol{\theta})^{\mathrm{T}} y_i\}\boldsymbol{\mu}(\mathbf{x}_i, \boldsymbol{\theta})$. However, when $y_i = -\boldsymbol{\mu}(\mathbf{x}_i, \boldsymbol{\theta})$ holds, there is an infinite number of geodesics connecting $\boldsymbol{\mu}(\mathbf{x}_i, \boldsymbol{\theta})$ and y_i. In this case, $\mathrm{Log}_{\boldsymbol{\mu}(\mathbf{x}_i,\boldsymbol{\theta})}(y_i)$ is not uniquely defined, whereas their geodesic distance is unique. However, $\mathrm{Log}_{\boldsymbol{\mu}(\mathbf{x}_i,q,\boldsymbol{\beta})}(y_i)$ is well defined almost surely for any finite measure with no point masses.

Third, we transport all the residuals, separately, to a common space, say $T_p S^k$. The *rotated residual* is given by

$$\mathcal{E}(y_i, \mathbf{x}_i; \boldsymbol{\theta}) = \mathrm{Log}_p(c_{p,\boldsymbol{\mu}(\mathbf{x}_i,\boldsymbol{\theta})}[-\cos^{-1}\{p^{\mathrm{T}}\boldsymbol{\mu}(\mathbf{x}_i, \boldsymbol{\theta})\}]y_i). \tag{8}$$

Our *intrinsic regression model* is defined by the zero conditional mean assumption in model (6) on the above rotated residuals (8). A graphic illustration of the stereographic link functions in expression (7), rotated residual and parallel transport is given in Fig. 2.

Alternatively, we may consider some parametric spherical regression models for spherical responses. As an illustration, we consider the von Mises–Fisher regression model. Specifically, it is assumed that $y_i|\mathbf{x}_i \sim \mathrm{vMF}\{\boldsymbol{\mu}(\mathbf{x}_i, \boldsymbol{\theta}), \kappa\}$ or, equivalently, $R_{\boldsymbol{\mu}(\mathbf{x}_i,\boldsymbol{\theta}),p} y_i|\mathbf{x}_i \sim \mathrm{vMF}(p, \kappa)$, for $i = 1, \ldots, n$, where κ is a positive concentration parameter and is assumed to be known for simplicity. Calculating the maximum likelihood estimate of $\boldsymbol{\theta}$ is equivalent to solving a score equation given by $\Sigma_{i=1}^n y_i^{\mathrm{T}} \partial_{\boldsymbol{\theta}} \boldsymbol{\mu}(\mathbf{x}_i, \boldsymbol{\theta}) = \mathbf{0}$, where $\partial_{\boldsymbol{\theta}} = \partial/\partial\boldsymbol{\theta}$. Since $\|\boldsymbol{\mu}(\mathbf{x}_i, \boldsymbol{\theta})\| = 1$ and $(\partial_{\boldsymbol{\theta}}\boldsymbol{\mu}(\mathbf{x}_i, \boldsymbol{\theta}))^{\mathrm{T}}\boldsymbol{\mu}(\mathbf{x}_i, \boldsymbol{\theta}) = 0$ for all subcomponents of $\boldsymbol{\theta}$, we have

$$y_i^{\mathrm{T}} \partial_{\boldsymbol{\theta}} \boldsymbol{\mu}(\mathbf{x}_i, \boldsymbol{\theta}) = \sqrt{2}\frac{\sqrt{[1 - \{\boldsymbol{\mu}(\mathbf{x}_i, \boldsymbol{\theta})^{\mathrm{T}} y_i\}^2]}}{\cos^{-1}\{\boldsymbol{\mu}(\mathbf{x}_i, \boldsymbol{\theta})^{\mathrm{T}} y_i\}}(R_{\boldsymbol{\mu}(\mathbf{x}_i,\boldsymbol{\theta}),p}\partial_{\boldsymbol{\theta}}\boldsymbol{\mu}(\mathbf{x}_i, \boldsymbol{\theta}))^{\mathrm{T}}\mathcal{E}(y_i, \mathbf{x}_i; \boldsymbol{\theta}), \tag{9}$$

which is a linear combination of the rotated residual.

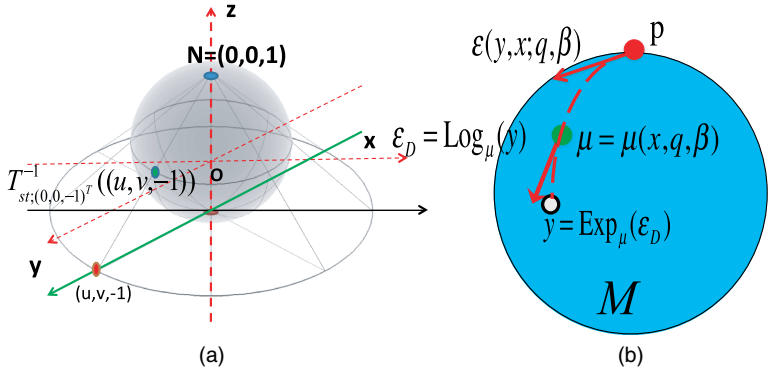


Fig. 2. Graphical illustration of (a) stereographic projection and (b) rotated residual and parallel transport: in (a), N and O denote the north pole (0, 0, 1) and the origin (0, 0, 0) respectively; in (b), y, $\mu = \mu(\mathbf{x}, q, \beta)$, $\epsilon_D = \mathrm{Log}_\mu(y)$, $\mathcal{E}(y, \mathbf{x}; q, \beta)$ and p respectively represent an observation, the conditional mean, the residual, the rotated residual and the base point

4. Estimation and test procedures

4.1. Generalized method-of-moment estimators

We consider the generalized method-of-moment (GMM) estimator to estimate the unknown parameters in model (6) (Hansen, 1982; Newey, 1993; Korsholm, 1999). We may view the $T_p\mathcal{M}$-valued function $\mathcal{E}$ as a function with values in $R^{d_\mathcal{M}}$. Let $h(\mathbf{x}; q, \boldsymbol{\beta})$ be an $s \times d_\mathcal{M}$ matrix of functions of $(\mathbf{x}, q, \boldsymbol{\beta})$ with $s \geqslant d_\mathcal{M} + d_\beta$ and W_n be a random sequence of positive definite $s \times s$ weight matrices. It follows from equation (6) that

$$E[h(\mathbf{x}_i; q_*, \boldsymbol{\beta}_*) E[\mathcal{E}(y_i, \mathbf{x}_i; q_*, \boldsymbol{\beta}_*) | \mathbf{x}_i]] = 0. \tag{10}$$

We define $\mathcal{Q}_n(q, \boldsymbol{\beta}) = (\mathbb{P}_n\{h(\mathbf{x}; q, \boldsymbol{\beta})\,\mathcal{E}(y, \mathbf{x}; q, \boldsymbol{\beta})\})^{\mathrm{T}} W_n (\mathbb{P}_n\{h(\mathbf{x}; q, \boldsymbol{\beta})\,\mathcal{E}(y, \mathbf{x}; q, \boldsymbol{\beta})\})$, where $\mathbb{P}_n\{f(y, \mathbf{x})\} = n^{-1}\sum_{i=1}^n f(y_i, \mathbf{x}_i)$ for a real-vector-valued function $f(y, \mathbf{x})$. The GMM estimator $(\hat{q}_\mathrm{G}, \hat{\boldsymbol{\beta}}_\mathrm{G})$, or simply $(\hat{q}, \hat{\boldsymbol{\beta}})$, of $(q, \boldsymbol{\beta})$ associated with $(h(\cdot, \cdot, \cdot), W_n)$ is defined as

$$(\hat{q}_\mathrm{G}, \hat{\boldsymbol{\beta}}_\mathrm{G}) = \underset{(q, \boldsymbol{\beta}) \in M \times R^{d_\beta}}{\arg\min}\ \mathcal{Q}_n(q, \boldsymbol{\beta}). \tag{11}$$

Under some conditions detailed below, we can show the first-order asymptotic properties of $(\hat{q}_\mathrm{G}, \hat{\boldsymbol{\beta}}_\mathrm{G})$ including consistency and asymptotic normality of GMM estimators. We introduce some notation. Let $\|\cdot\|$ denote the Euclidean norm of a vector or a matrix; $\partial^l f(\mathbf{t}, \boldsymbol{\beta})/\partial(\mathbf{t}, \boldsymbol{\beta})^l = \partial^l_{(\mathbf{t}, \boldsymbol{\beta})} f(\mathbf{t}, \boldsymbol{\beta})$ for $l = 1, \ldots$; $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^{\mathrm{T}}$ for any matrix or vector $\mathbf{a}$;

$$V = \mathrm{var}\{h(\mathbf{x}; q_*, \boldsymbol{\beta}_*)\mathcal{E}(y, \mathbf{x}; q_*, \boldsymbol{\beta}_*)\};$$

I_d is the identity matrix; $\to^{\mathrm{d}}$ and $\to^{\mathrm{p}}$ respectively denote convergence in distribution and in probability. We obtain the following results, whose detailed proofs can be found in the on-line supplementary document.

Theorem 1. Assume that $(y_i, \mathbf{x}_i)$, $i = 1, \ldots, n$, are independently and identically distributed random variables in $\mathcal{M} \times R^{d_\mathbf{x}}$. Let $(q_*, \boldsymbol{\beta}_*)$ be the exact value of the parameters satisfying model (6). Let $\{W_n\}_n$ be a random sequence of $s \times s$ symmetric positive semidefinite matrices with $s \geqslant d_\mathcal{M} + d_\beta$.

(a) Under assumptions (C1)–(C5) in the on-line supplementary document, $(\hat{q}_\mathrm{G}, \hat{\boldsymbol{\beta}}_\mathrm{G})$ in equation (11) is consistent in probability as $n \to \infty$.

(b) Under assumptions (C1)–(C4) and (C6)–(C10) in the supplementary document, for any local chart (U, ϕ) on $\mathcal{M}$ near q_* as $n \to \infty$, we have

$$n^{1/2}[(\phi(\hat{q}_\mathrm{G})^{\mathrm{T}}, \hat{\boldsymbol{\beta}}_\mathrm{G}^{\mathrm{T}})^{\mathrm{T}} - (\phi(q_*)^{\mathrm{T}}, \boldsymbol{\beta}_*^{\mathrm{T}})^{\mathrm{T}}] \xrightarrow{\mathrm{d}} N_{d_\mathcal{M} + d_\beta}(\mathbf{0}, \Sigma_\phi), \tag{12}$$

where $\Sigma_\phi = (G_\phi^{\mathrm{T}} W G_\phi)^{-1} G_\phi^{\mathrm{T}} W V W G_\phi (G_\phi^{\mathrm{T}} W G_\phi)^{-1}$, in which G_ϕ is defined in assumption (C9). Moreover, for any other chart (U, ϕ') near q_*, we have

$$\Sigma_{\phi'} = \mathrm{diag}(J(\phi' \circ \phi^{-1})_{\phi(q_*)}, I_{d_\beta}) \Sigma_\phi \mathrm{diag}(J(\phi' \circ \phi^{-1})_{\phi(q_*)}, I_{d_\beta})^{\mathrm{T}}, \tag{13}$$

where $J(\cdot)_\mathbf{t}$ denotes the Jacobian matrix evaluated at $\mathbf{t}$.

Theorem 1 establishes the first-order asymptotic properties of $(\hat{q}_\mathrm{G}, \hat{\boldsymbol{\beta}}_\mathrm{G})$ for the intrinsic regression model (6). Theorem 1, part (a), establishes the consistency of $(\hat{q}_\mathrm{G}, \hat{\boldsymbol{\beta}}_\mathrm{G})$. The consistency result does not depend on the local chart. Theorem 1, part (b), establishes the asymptotic normality of $(\phi(\hat{q}_\mathrm{G}), \hat{\boldsymbol{\beta}}_\mathrm{G})$ for a specific chart (U, ϕ) and the relationship between the asymptotic covariances $\Sigma_{\phi'}$ and Σ_ϕ for two different charts. It follows from the lower right $d_\beta \times d_\beta$ submatrix of $\Sigma_{\phi'}$ that the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}$ does not depend on the chart. However,

the asymptotic normality of $\hat{q}_{\mathrm{G}}$ does depend on a specific chart. A consistent estimator of the asymptotic covariance matrix Σ_ϕ is given by

$$(\hat{G}_\phi^{\mathrm{T}} W_n \hat{G}_\phi)^{-1} \hat{G}_\phi^{\mathrm{T}} W_n \hat{V} W_n \hat{G}_\phi (\hat{G}_\phi^{\mathrm{T}} W_n \hat{G}_\phi)^{-1}$$

with

$$\hat{G}_\phi = n^{-1} \sum_{i=1}^{n} \left[h(\mathbf{x}_i; \hat{q}, \hat{\boldsymbol{\beta}}) \frac{\partial}{\partial(\mathbf{t}, \boldsymbol{\beta})} \mathcal{E}\{y_i, \mathbf{x}_i; \phi^{-1}(\mathbf{t}), \hat{\boldsymbol{\beta}}\} \Big|_{\mathbf{t}=\phi(\hat{q})} \right]$$

and

$$\hat{V} = n^{-1} \sum_{i=1}^{n} \{h(\mathbf{x}_i; \hat{q}, \hat{\boldsymbol{\beta}})\, \mathcal{E}(y_i, \mathbf{x}_i; \hat{q}, \hat{\boldsymbol{\beta}})\}^{\otimes 2}.$$

This estimator is also compatible with the manifold structure of $\mathcal{M}$.

We consider the relationship between the GMM estimator and the *intrinsic least squares estimator* of $(q, \boldsymbol{\beta})$, which is denoted by $(\hat{q}_{\mathrm{I}}, \hat{\boldsymbol{\beta}}_{\mathrm{I}})$. The $(\hat{q}_{\mathrm{I}}, \hat{\boldsymbol{\beta}}_{\mathrm{I}})$ minimizes the total residual sum of squares $\mathcal{G}_{\mathrm{I},n}(q, \boldsymbol{\beta})$ as follows:

$$(\hat{q}_{\mathrm{I}}, \hat{\boldsymbol{\beta}}_{\mathrm{I}}) = \underset{(q,\boldsymbol{\beta}) \in M \times R^{d_\beta}}{\arg\min}\ \mathcal{G}_{\mathrm{I},n}(q, \boldsymbol{\beta}) = \underset{(q,\boldsymbol{\beta}) \in M \times R^{d_\beta}}{\arg\min} \sum_{i=1}^{n} \mathrm{dist}_{\mathcal{M}}\{y_i, \boldsymbol{\mu}(\mathbf{x}_i, q, \boldsymbol{\beta})\}^2. \tag{14}$$

According to equation (2), the intrinsic least squares estimator is closely related to the intrinsic mean $\hat{q}_{\mathrm{IM}}$ of $y_1, \ldots, y_n \in \mathcal{M}$, which is defined as

$$\hat{q}_{\mathrm{IM}} = \underset{q \in M}{\arg\min} \sum_{i=1}^{n} \mathrm{dist}_{\mathcal{M}}(y_i, q)^2 = \underset{q \in M}{\arg\min} \sum_{i=1}^{n} \mathrm{dist}_{\mathcal{M}}\{y_i, \boldsymbol{\mu}(\mathbf{0}, q, \boldsymbol{\beta})\}^2.$$

Recall that $\boldsymbol{\mu}(\mathbf{0}, q, \boldsymbol{\beta})$ is independent of $\boldsymbol{\beta}$.

The $(\hat{q}_{\mathrm{I}}, \hat{\boldsymbol{\beta}}_{\mathrm{I}})$ can be regarded as a special case of the GMM estimator when we set $W_n = I_{d_{\mathcal{M}}+d_\beta}$ and h's rows $h_j(\mathbf{x}, q, \boldsymbol{\beta}) = (L_{c(1;\mathbf{x},q,\boldsymbol{\beta})^{-1},*}(\partial_{t_j} \boldsymbol{\mu}\{\mathbf{x}, \phi^{-1}(\mathbf{t}), \boldsymbol{\beta}\}|_{\mathbf{t}=\phi(q)}))^{\mathrm{T}}$ for $j = 1, \ldots, d_{\mathcal{M}}$, and $h_{d_{\mathcal{M}}+j}(\mathbf{x}, q, \boldsymbol{\beta}) = (L_{c(1;\mathbf{x},q,\boldsymbol{\beta})^{-1},*}(\partial_{\beta_j} \boldsymbol{\mu}(\mathbf{x}, q, \boldsymbol{\beta})))^{\mathrm{T}}$ for $j = 1, \ldots, d_\beta$, where (U, ϕ) is a chart on $\mathcal{M}$ and each row of $h(\mathbf{x}, q, \boldsymbol{\beta})$ is in $R^{1 \times d_{\mathcal{M}}}$ via the identification $T_p\mathcal{M} \cong R^{d_{\mathcal{M}}}$ corresponding to (U, ϕ). It follows from theorem 1 that, under model (6), $(\hat{q}_{\mathrm{I}}, \hat{\boldsymbol{\beta}}_{\mathrm{I}})$ enjoys the first-order asymptotic properties as well.

4.2. Efficient generalized method-of-moment estimator

We investigate the most efficient estimator in the class of GMM estimators. For a fixed $h(\cdot;\cdot,\cdot)$, the optimal choice of W is $W^{\mathrm{opt}} = V^{-1}$, and the use of $W_n = W^{\mathrm{opt}}$ leads to the most efficient estimator in the class of all GMM estimators obtained by using the same $h(\cdot)$ function (Hansen, 1982). Its asymptotic covariance is given by $(G_\phi V^{-1} G_\phi)^{-1}$. An interesting question is what the optimal choice of $h^{\mathrm{opt}}(\cdot)$ is.

We first introduce some notation. For a chart (U, ϕ) on $\mathcal{M}$ near q_*, let

$$D_\phi(\mathbf{x}) = E[\partial_{(\mathbf{t},\boldsymbol{\beta})} \mathcal{E}\{y, \mathbf{x}; \phi^{-1}(\mathbf{t}), \boldsymbol{\beta}_*\}|_{\mathbf{t}=\phi(q_*)} | \mathbf{x}]^{\mathrm{T}}, \qquad h_\phi^*(\mathbf{x}) = D_\phi(\mathbf{x}) \Omega(\mathbf{x})^{-1},$$
$$W_\phi^* = E[D_\phi(\mathbf{x})\, \Omega(\mathbf{x})^{-1} D_\phi(\mathbf{x})^{\mathrm{T}}]^{-1}, \qquad \Omega(\mathbf{x}) = \mathrm{var}\{\mathcal{E}(y, \mathbf{x}; q_*, \boldsymbol{\beta}_*) | \mathbf{x}\}.$$

Let $(\hat{q}^*, \hat{\boldsymbol{\beta}}^*)$ be the GMM estimator of $(q, \boldsymbol{\beta})$ based on $h_\phi^*(\mathbf{x})$ and W_ϕ^*. Generally, we obtain an optimal result of $h^{\mathrm{opt}}(\cdot)$, which generalizes an existing result for Euclidean-valued responses and covariates (Newey, 1993), as follows.

Theorem 2. Suppose that assumptions (C2)–(C8) and (C10)–(C12) in the on-line supplementary document hold for $h_\phi^*(\mathbf{x})$ and W_ϕ^*. We have the following results.

(a) $(\hat{q}^*, \hat{\boldsymbol{\beta}}^*)$ is asymptotically normally distributed with mean $\mathbf{0}$ and covariance W_ϕ^*;
(b) $(\hat{q}^*, \hat{\boldsymbol{\beta}}^*)$ is optimal among all GMM estimators for model (6);
(c) $(\hat{q}^*, \hat{\boldsymbol{\beta}}^*)$ is independent of the chart.

Theorem 2 characterizes the optimality of $h_\phi^*(\mathbf{x})$ and W_ϕ^* among regular GMM estimators for model (6). Geometrically, $(\hat{q}^*, \hat{\boldsymbol{\beta}}^*)$ is independent of the chart. Specifically, for any other chart (U, ϕ') near q_*, we have

$$\begin{aligned} D_{\phi'}(\mathbf{x}) &= \operatorname{diag}(J(\phi' \circ \phi^{-1})^{-1}_{\phi(q_*)}, I_{d_\beta})^{\mathrm{T}} D_\phi(\mathbf{x}), \\ h^*_{\phi'}(\mathbf{x}) &= \operatorname{diag}(J(\phi' \circ \phi^{-1})^{-1}_{\phi(q_*)}, I_{d_\beta})^{\mathrm{T}} h^*_\phi(\mathbf{x}), \\ W^*_{\phi'} &= \operatorname{diag}(J(\phi' \circ \phi^{-1})_{\phi(q_*)}, I_{d_\beta}) W^*_\phi \operatorname{diag}(J(\phi' \circ \phi^{-1})_{\phi(q_*)}, I_{d_\beta})^{\mathrm{T}}. \end{aligned}$$

Thus, the quadratic form in equation (11) associated with $h^*_{\phi'}(\mathbf{x})$ and $W^*_{\phi'}$ is the same as that which is associated with $h^*_\phi(\mathbf{x})$ and W^*_ϕ. It indicates that the GMM estimator $(\hat{q}^*, \hat{\boldsymbol{\beta}}^*)_\phi$ based on $h^*_\phi(\mathbf{x})$ and W^*_ϕ is independent of the chart (U, ϕ).

The next challenging issue is the estimation of $D_\phi(\mathbf{x})$ and $\Omega(\mathbf{x})$. We may proceed in two steps. The first step is to calculate a $\sqrt{n}$-consistent estimator $(\hat{q}_{\mathrm{I}}, \hat{\boldsymbol{\beta}}_{\mathrm{I}})$ of $(q, \boldsymbol{\beta})$, such as the intrinsic least squares estimator. The second step is to plug $(\hat{q}_{\mathrm{I}}, \hat{\boldsymbol{\beta}}_{\mathrm{I}})$ into the functions $\mathcal{E}_i(\hat{q}_{\mathrm{I}}, \hat{\boldsymbol{\beta}}_{\mathrm{I}})$ and $\partial_{(\mathbf{t},\boldsymbol{\beta})}\{\mathcal{E}(y_i, \mathbf{x}_i; \phi^{-1}(\mathbf{t}), \hat{\boldsymbol{\beta}}_{\mathrm{I}})\}|_{\mathbf{t}=\phi(\hat{q}_I)}$ for all i and then to use them to construct the non-parametric estimates of $D_\phi(\mathbf{x})$ and $\Omega(\mathbf{x})$ (Newey, 1993). Specifically, let $K(\cdot)$ be a $d_{\mathbf{x}}$-dimensional kernel function of the l_0th order satisfying $\int K(u_1, \ldots, u_{d_{\mathbf{x}}}) \mathrm{d}u_1 \ldots \mathrm{d}u_{d_{\mathbf{x}}} = 1$, $\int u_s^l K(u_1, \ldots, u_{d_{\mathbf{x}}}) \mathrm{d}u_1 \ldots \mathrm{d}u_{d_{\mathbf{x}}} = 0$ for any $s = 1, \ldots, d_{\mathbf{x}}$ and $1 \leqslant l < l_0$, and $\int u_s^{l_0} K(u_1, \ldots, u_{d_{\mathbf{x}}}) \mathrm{d}u_1 \ldots \mathrm{d}u_{d_{\mathbf{x}}} \neq 0$. Let $K_\tau(u) = \tau^{-1} \times K(u/\tau)$, where $\tau > 0$ is a bandwidth. Then, a non-parametric estimator of $D_\phi(\mathbf{x})$ can be written as

$$\hat{D}_\phi(\mathbf{x})^{\mathrm{T}} = \sum_{i=1}^n \omega_i(\mathbf{x}; \tau) \partial_{(\mathbf{t},\boldsymbol{\beta})} \mathcal{E}\{y_i, \mathbf{x}_i; \phi^{-1}(\mathbf{t}), \hat{\boldsymbol{\beta}}_{\mathrm{I}}\}|_{\mathbf{t}=\phi(\hat{q}_{\mathrm{I}})}, \tag{15}$$

where $\omega_i(\mathbf{x}; \tau) = K_\tau(\mathbf{x} - \mathbf{x}_i) / \Sigma_{k=1}^n K_\tau(\mathbf{x} - \mathbf{x}_k)$. Although we may construct a non-parametric estimator of $\Omega(\mathbf{x})$ that is similar to equation (15), we have found that, even for moderate $d_{\mathbf{x}}$, such an estimator is numerically unstable. Instead, we approximate

$$\Omega(\mathbf{x}_i) = \operatorname{var}\{\mathcal{E}(y, \mathbf{x}; q_*, \boldsymbol{\beta}_*) | \mathbf{x} = \mathbf{x}_i\}$$

by its mean $V_{E*} = \operatorname{var}\{\mathcal{E}(y, \mathbf{x}; q_*, \boldsymbol{\beta}_*)\}$. In this case, $h^*_\phi(\mathbf{x})$ and W^*_ϕ respectively reduce to

$$\begin{aligned} h^*_{E,\phi}(\mathbf{x}) &= D_\phi(\mathbf{x}) V_{E*}^{-1}, \\ W^*_{E,\phi} &= (E[D_\phi(\mathbf{x}) V_{E*}^{-1} \Omega(\mathbf{x}) V_{E*}^{-1} D_\phi(\mathbf{x})^{\mathrm{T}}])^{-1}. \end{aligned} \tag{16}$$

For any local chart (U, ϕ) with $\hat{q}_{\mathrm{I}} \in U$, we construct the estimators of $h^*_{E,\phi}$ and $W^*_{E,\phi}$ as follows. Let $\hat{V}(q, \boldsymbol{\beta}) = \mathbb{P}_n\{\mathcal{E}(y, \mathbf{x}; q, \boldsymbol{\beta})\}^{\otimes 2}$; we have

$$\begin{aligned} \hat{h}_{E,\phi}(\mathbf{x}_i) &= \hat{D}_\phi(\mathbf{x}_i) \hat{V}(\hat{q}_{\mathrm{I}}, \hat{\boldsymbol{\beta}}_{\mathrm{I}})^{-1}, \\ \hat{W}_{E,\phi} &= [\mathbb{P}_n\{\hat{h}_{E,\phi}(\mathbf{x}) \mathcal{E}(y, \mathbf{x}; \hat{q}_{\mathrm{I}}, \hat{\boldsymbol{\beta}}_{\mathrm{I}})\}^{\otimes 2}]^{-1}. \end{aligned} \tag{17}$$

Then, we substitute $\hat{h}_{E,\phi}$ and $\hat{W}_{E,\phi}$ into equation (11) and then calculate the GMM estimator of $(q, \boldsymbol{\beta})$, which is denoted by $(\hat{q}_E, \hat{\boldsymbol{\beta}}_E)$. Similarly to $(\hat{q}^*, \hat{\boldsymbol{\beta}}^*)$, it can be shown that $(\hat{q}_E, \hat{\boldsymbol{\beta}}_E)$ is independent of the chart (U, ϕ) on $\mathcal{M}$ near q_* with $\hat{q}_{\mathrm{I}} \in U$. For sufficiently large n, $\operatorname{dist}_{\mathcal{M}}(\hat{q}_{\mathrm{I}}, q_*)$ can be made sufficiently small and any maximal normal chart on $\mathcal{M}$ centred at $\hat{q}_{\mathrm{I}}$ contains the true value q_* with probability approaching 1.

We calculate a one-step linearized estimator of $(q, \boldsymbol{\beta})$, which is denoted by $(\tilde{q}_E, \tilde{\boldsymbol{\beta}}_E)$, to approx-

imate $(\hat{q}_E, \hat{\boldsymbol{\beta}}_E)$. Computationally, the linearized estimator does not require iteration, whereas, theoretically, it shares the first-order asymptotic properties with $(\hat{q}_E, \hat{\boldsymbol{\beta}}_E)$ as shown below. Specifically, in the chart (U, ϕ) near $\hat{q}_{\mathrm{I}}$, we have

$$
\begin{aligned}
&(\tilde{\mathbf{t}}_{E,\phi}^{\mathrm{T}}, \tilde{\boldsymbol{\beta}}_{E,\phi}^{\mathrm{T}})^{\mathrm{T}} - (\phi(\hat{q}_{\mathrm{I}})^{\mathrm{T}}, \hat{\boldsymbol{\beta}}_{\mathrm{I}}^{\mathrm{T}})^{\mathrm{T}} \\
&\quad = (-\mathbb{P}_n[\hat{h}_{E,\phi}(\mathbf{x}) \partial_{(\mathbf{t},\boldsymbol{\beta})} \mathcal{E}\{y, \mathbf{x}; \phi^{-1}(t), \hat{\boldsymbol{\beta}}_{\mathrm{I}}\}|_{\mathbf{t}=\phi(\hat{q}_{\mathrm{I}})}])^{-1} \mathbb{P}_n\{\hat{h}_{E,\phi}(\mathbf{x}) \mathcal{E}_i(y, \mathbf{x}; \hat{q}_{\mathrm{I}}, \hat{\boldsymbol{\beta}}_{\mathrm{I}})\}. \qquad (18)
\end{aligned}
$$

Furthermore, if (U', ϕ') is another chart on $\mathcal{M}$ near $\hat{q}_{\mathrm{I}}$, then we have

$$
\begin{aligned}
&(\tilde{\mathbf{t}}_{E,\phi'}^{\mathrm{T}}, \tilde{\boldsymbol{\beta}}_{E,\phi'}^{\mathrm{T}})^{\mathrm{T}} - (\phi'(\hat{q}_{\mathrm{I}})^{\mathrm{T}}, \hat{\boldsymbol{\beta}}_{\mathrm{I}}^{\mathrm{T}})^{\mathrm{T}} \\
&\quad = \begin{pmatrix} J(\phi' \circ \phi^{-1})_{\phi(\hat{q}_{\mathrm{I}})} & 0 \\ 0 & I_{d_\beta} \end{pmatrix} \{(\tilde{\mathbf{t}}_{E,\phi}^{\mathrm{T}}, \tilde{\boldsymbol{\beta}}_{E,\phi}^{\mathrm{T}})^{\mathrm{T}} - (\phi(\hat{q}_{\mathrm{I}})^{\mathrm{T}}, \hat{\boldsymbol{\beta}}_{\mathrm{I}}^{\mathrm{T}})^{\mathrm{T}}\}.
\end{aligned}
$$

Thus, $\tilde{\boldsymbol{\beta}}_{E,\phi}$ is independent of the chart ϕ and $\{\tilde{\mathbf{t}}_{E,\phi} - \phi(\hat{q}_{\mathrm{I}}) | \phi \text{ is a chart on } M\}$ defines a unique tangent vector to $\mathcal{M}$ at $\hat{q}_I$. Moreover, if ϕ and ϕ' are maximal normal charts centred at $\hat{q}_{\mathrm{I}}$, then $\gamma_\phi(\tau) = \phi^{-1}(\tau \tilde{\mathbf{t}}_{E,\phi})$ and $\gamma_{\phi'}(\tau) = \phi'^{-1}(\tau \tilde{\mathbf{t}}_{E,\phi'})$ are two geodesic curves on $\mathcal{M}$ starting from the same point $\hat{q}_{\mathrm{I}}$ with the same initial velocity vector, and thus these two geodesics coincide. Therefore, $\phi^{-1}(\tilde{\mathbf{t}}_{E,\phi})$ is independent of the normal chart ϕ centred at $\hat{q}_{\mathrm{I}}$. Finally, we can establish the first-order asymptotic properties of $(\tilde{q}_E, \tilde{\boldsymbol{\beta}}_E)$ as follows.

Theorem 3. Assume that assumptions (C2)–(C11) and (C13)–(C18) in the on-line supplementary document are valid. As $n \to \infty$, we have the following results:

$$
n^{1/2}\{(\phi(\tilde{q}_E)^{\mathrm{T}}, \tilde{\boldsymbol{\beta}}_E^{\mathrm{T}})^{\mathrm{T}} - (\phi(q_*)^{\mathrm{T}}, \boldsymbol{\beta}_*^{\mathrm{T}})^{\mathrm{T}}\} \xrightarrow{\mathrm{d}} N_{d_{\mathcal{M}}+d_\beta}(\mathbf{0}, \Sigma_{E,\phi}), \qquad (19)
$$

where $\Sigma_{E,\phi} = (G_{\phi,h_{E,\phi}^*} W_{E,\phi}^* G_{\phi,h_{E,\phi}^*})^{-1}$. In addition, $\Sigma_{E,\phi}$ is invariant under the change of coordinates in $\mathcal{M}$ and the asymptotic distribution of $\tilde{\boldsymbol{\beta}}_E$ does not depend on the chart (U, ϕ). Also, if we set

$$
\begin{aligned}
\hat{\Sigma}_{E,\phi} = n^{-1} &[\mathbb{P}_n\{\hat{D}_\phi(\mathbf{x}) \hat{V}(\hat{q}_{\mathrm{I}}, \hat{\boldsymbol{\beta}}_{\mathrm{I}})^{-1} \hat{D}_\phi(\mathbf{x})^{\mathrm{T}}\}]^{-1} \\
&\times [\mathbb{P}_n\{\hat{D}_\phi(\mathbf{x}) \hat{V}(\hat{q}_{\mathrm{I}}, \hat{\boldsymbol{\beta}}_{\mathrm{I}})^{-1} \mathcal{E}(y, \mathbf{x}, \tilde{q}_{\mathrm{I}}, \tilde{\boldsymbol{\beta}}_{\mathrm{I}})^{\otimes 2} \hat{V}(\hat{q}_{\mathrm{I}}, \hat{\boldsymbol{\beta}}_{\mathrm{I}})^{-1} \hat{D}_\phi(\mathbf{x})^{\mathrm{T}}\}] \\
&\times [\mathbb{P}_n\{\hat{D}_\phi(\mathbf{x}) \hat{V}(\hat{q}_{\mathrm{I}}, \hat{\boldsymbol{\beta}}_{\mathrm{I}})^{-1} \hat{D}_\phi(\mathbf{x})^{\mathrm{T}}\}]^{-1}, \qquad (20)
\end{aligned}
$$

then $n\hat{\Sigma}_{E,\phi}$ is a consistent estimator of $\Sigma_{E,\phi}$, i.e. $n\hat{\Sigma}_{E,\phi} \to p\Sigma_{E,\phi}$. This estimator is also compatible with the manifold structure of $\mathcal{M}$.

Theorem 3 establishes the first-order asymptotic properties of $(\tilde{q}_E, \tilde{\boldsymbol{\beta}}_E)$. If $\Omega(\mathbf{x}) = \Omega$ for a constant matrix Ω, then it follows from theorems 2 and 3 that $(\tilde{q}_E, \tilde{\boldsymbol{\beta}}_E)$ is optimal. If $\Omega(\mathbf{x})$ does not vary dramatically as a function of $\mathbf{x}$, then $(\tilde{q}_E, \tilde{\boldsymbol{\beta}}_E)$ is nearly optimal. If $\Omega(\mathbf{x})$ varies dramatically as a function of $\mathbf{x}$, we can replace $\hat{V}(\hat{q}_{\mathrm{I}}, \hat{\boldsymbol{\beta}}_{\mathrm{I}})$ in expression (17) by $\hat{\Omega}(\mathbf{x}_i)$ to obtain $\hat{h}_{E,\phi}(\mathbf{x}) = \hat{D}_\phi(\mathbf{x}) \hat{\Omega}(\mathbf{x})^{-1}$ and $\hat{W}_{E,\phi} = [\mathbb{P}_n\{\hat{h}_{E,\phi}(\mathbf{x}) \mathcal{E}(y, \mathbf{x}; \hat{q}_{\mathrm{I}}, \hat{\boldsymbol{\beta}}_{\mathrm{I}})\}^{\otimes 2}]^{-1}$, where $\hat{\Omega}(\mathbf{x}_i)$ is a consistent estimator of $\Omega(\mathbf{x}_i)$ for all i; then the optimality of $(\tilde{q}_E, \tilde{\boldsymbol{\beta}}_E)$ still holds. We have the following theorem.

Theorem 4. Assume that assumptions (C2)–(C17) and (C19) are valid. Then, as $n \to \infty$, we have

$$
n^{1/2}\{(\phi(\tilde{q}_E)^{\mathrm{T}}, \tilde{\boldsymbol{\beta}}_E^{\mathrm{T}})^{\mathrm{T}} - (\phi(q_*)^{\mathrm{T}}, \boldsymbol{\beta}_*^{\mathrm{T}})^{\mathrm{T}}\} \xrightarrow{\mathrm{d}} N_{d_{\mathcal{M}}+d_\beta}(\mathbf{0}, \Sigma_\phi^*), \qquad (21)
$$

in which Σ_ϕ^* is given in theorem 2. If we set $\hat{\Sigma}_{E,\phi} = n^{-1}[\mathbb{P}_n\{\hat{D}_\phi(\mathbf{x}) \hat{\Omega}(\mathbf{x})^{-1} \hat{D}_\phi(\mathbf{x})^{\mathrm{T}}\}]^{-1}$, then $n\hat{\Sigma}_{E,\phi}$ is a consistent estimator of Σ_ϕ^*.

4.3. Computational algorithm

Computationally, an annealing evolutionary stochastic approximation Monte Carlo algorithm (Liang *et al.*, 2010) is developed to compute $(\hat{q}_{\mathrm{I}}, \hat{\boldsymbol{\beta}}_{\mathrm{I}})$ and $(\hat{q}_E, \hat{\boldsymbol{\beta}}_E)$. See the on-line supplementary document for details. Although some gradient-based optimization methods, such as the quasi-Newton method, have been used to optimize $\mathcal{Q}_n(q, \boldsymbol{\beta})$ (Kim *et al.*, 2014; Fletcher, 2013), we have found that these methods strongly depend on the starting value of $(q, \boldsymbol{\beta})$. Specifically, when $\mathcal{E}(y, \mathbf{x}; q, \boldsymbol{\beta})$ takes a relatively complicated form, $\mathcal{Q}_n(q, \boldsymbol{\beta})$ is generally not convex and can easily converge to local minima. Moreover, we have found that it can be statistically misleading to carry out statistical inference, such as the estimated standard errors of $(\hat{q}_E, \hat{\boldsymbol{\beta}}_E)$, at those local minima. The annealing evolutionary stochastic approximation Monte Carlo algorithm converges fast and is distinguished from many gradient-based algorithms, since it has a nice feature in that the moves are self-adjustable and thus not likely to become trapped by local energy minima. The annealing evolutionary stochastic approximation Monte Carlo algorithm (Liang *et al.*, 2010) represents a further improvement of stochastic approximation Monte Carlo methods for optimization problems by incorporating some features of simulated annealing and the genetic algorithm in its search process.

4.4. Hypotheses testing

Many scientific questions involve the comparison of the $\mathcal{M}$-valued data across groups and subjects and the detection of the change in the $\mathcal{M}$-valued data over time. Such questions usually can be formulated as testing the hypotheses of q and $\boldsymbol{\beta}$. We consider two types of hypotheses as follows:

$$H_0^{(1)} : C_0\boldsymbol{\beta} = \mathbf{b}_0 \quad \textit{versus} \quad H_1^{(1)} : C_0\boldsymbol{\beta} \neq \mathbf{b}_0, \tag{22}$$

$$H_0^{(2)} : q = q_0 \quad \textit{versus} \quad H_1^{(2)} : q \neq q_0, \tag{23}$$

where C_0 is an $r \times d_{\boldsymbol{\beta}}$ matrix of full row rank and q_0 and $\mathbf{b}_0$ are specified in $\mathcal{M}$ and R^r respectively. Further extensions of these hypotheses are definitely interesting and possible. For instance, for the multicentre link function, we may be interested in testing whether all intercepts are independent of the discrete covariate class.

We develop several test statistics for testing the hypotheses given in expression (22) and (23). First, we consider the Wald test statistic for testing $H_0^{(1)}$ against $H_1^{(1)}$ in expression (22), which is given by

$$W_{n,\phi}^{(1)} = (C_0\tilde{\boldsymbol{\beta}}_E - \mathbf{b}_0)^{\mathrm{T}}(C_0\hat{\Sigma}_{E,\phi;22}C_0^{\mathrm{T}})^{-1}(C_0\tilde{\boldsymbol{\beta}}_E - \mathbf{b}_0),$$

where $\hat{\Sigma}_{E,\phi}$ is given in theorem 3 or theorem 4, and $\hat{\Sigma}_{E,\phi;22}$ is its lower right $d_{\boldsymbol{\beta}} \times d_{\boldsymbol{\beta}}$ submatrix. Since $\tilde{\boldsymbol{\beta}}_E$ and its asymptotic covariance matrix are independent of the chart on $\mathcal{M}$, the test statistic $W_{n,\phi}^{(1)}$ is independent of the chart.

Second, we consider the Wald test statistic for testing the hypotheses that are given in expression (23) when there is a local chart (U, ϕ) on $\mathcal{M}$ containing both $\hat{q}_E$ and q_0. Specifically, the Wald test statistic for testing hypothesis (23) is defined by

$$W_{n,\phi}^{(2)} = (\phi(\tilde{q}_E) - \phi(q_0))^{\mathrm{T}}\{(I_{d_{\mathcal{M}}}\ \mathbf{0})\hat{\Sigma}_{E,\phi}(I_{d_{\mathcal{M}}}\ \mathbf{0})^{\mathrm{T}}\}^{-1}(\phi(\tilde{q}_E) - \phi(q_0)).$$

Third, we develop an intrinsic Wald test statistic, that is independent of the chart, for testing the hypotheses that are given in expression (23). We consider the asymptotic covariance estimator $\hat{\Sigma}_{E,\phi}$ based on $\tilde{q}_E$ and its upper left $d_{\mathcal{M}} \times d_{\mathcal{M}}$ submatrix $\hat{\Sigma}_{E,\phi;11}$. Since both are compatible with the manifold structure of $\mathcal{M}$, $\hat{\Sigma}_{E,\phi;11}$ defines a unique non-degenerate linear map $\hat{\Sigma}_{E;11}(\cdot)$ from the tangent space $T_{\tilde{q}_E}\mathcal{M}$ of $\mathcal{M}$ at $\tilde{q}_E$ onto itself, which is independent of the chart (U, ϕ).

In a maximal normal chart centred at $\tilde{q}_E$, then, in any such normal chart, the Wald test statistic for testing hypotheses (23) is given by

$$W^{(2)}_{\mathcal{M},n} = m_{\tilde{q}_E}\{(\hat{\Sigma}_{E;11})^{-1}\mathrm{Log}_{\tilde{q}_E}(q_0), \mathrm{Log}_{\tilde{q}_E}(q_0)\}.$$

We obtain the asymptotic null distributions of $W^{(1)}_{n,\phi}$, $W^{(2)}_{n,\phi}$ and $W^{(2)}_{\mathcal{M},n}$ as follows.

Theorem 5. Let (U, ϕ) be a local chart on $\mathcal{M}$ so that $\tilde{q}_E, q_* \in U$. Assume that all the conditions in theorem 3 hold. Under the corresponding null hypothesis, we have the following results:

(a) $W^{(1)}_{n,\phi}$ and $W^{(2)}_{n,\phi}$ are asymptotically distributed as χ^2_r and $\chi^2_{d_{\mathcal{M}}}$ respectively;
(b) $W^{(1)}_{n,\phi}$ is independent of the chart (U, ϕ);
(c) $W^{(2)}_{n,\phi'} = W^{(2)}_{n,\phi} + o_p(1)$, for any other local chart (U, ϕ') with $\tilde{q}_E$ and q_0 in U;
(d) $W^{(2)}_{n,\phi} = W^{(2)}_{\mathcal{M},n}$, for any normal chart (U, ϕ) centred at $\tilde{q}_E$.

Theorem 5 has several important implications. Theorem 5, part (a), characterizes the asymptotic null distributions of $W^{(1)}_{n,\phi}$ and $W^{(2)}_{n,\phi}$. Theorem 5, part (b), shows that $W^{(1)}_{n,\phi}$ does not depend on the choice of the chart (U, ϕ) on $\mathcal{M}$. Theorem 5, part (c), shows that $W^{(2)}_{n,\phi'}$ and $W^{(2)}_{n,\phi}$ are asymptotically equivalent for any two local charts. Theorem 5, part (d), shows that $W^{(2)}_{n,\phi'}$ can be used to construct an intrinsic test statistic.

We consider a local alternative framework for expressions (22) and (23) as follows:

$$H^{(1)}_0 : C_0\boldsymbol{\beta} = \mathbf{b}_0 \quad \textit{versus} \quad H^{(1)}_{1,n} : C_0\boldsymbol{\beta} = \mathbf{b}_0 + \delta/\sqrt{n} + o(1/\sqrt{n}), \tag{24}$$

$$H^{(2)}_0 : q = q_0 \quad \textit{versus} \quad H^{(2)}_{1,n} : q = \mathrm{Exp}_{q_0}\{\mathbf{v}/\sqrt{n} + o(1/\sqrt{n})\}, \tag{25}$$

where δ and $\mathbf{v}$ are specified (and fixed) in R^r and $T_{q_0}\mathcal{M}$ respectively, and we establish the asymptotic distributions of $W^{(1)}_{n,\phi}$, $W^{(2)}_{n,\phi}$ and $W^{(2)}_{\mathcal{M},n}$ under these local alternatives.

Theorem 6. Let (U, ϕ) be a local chart on $\mathcal{M}$ so that $\tilde{q}_E$, $q_* \in U$. Assume that all conditions in theorem 3 hold. Under the local alternatives (24) and (25), we have the following results.

(a) Under $H^{(1)}_{1,n}$, $W^{(1)}_{n,\phi}$ is asymptotically distributed as non-central χ^2_r with non-centrality parameter $\delta^T(C_0\hat{\Sigma}_{E,\phi;22}C_0^{\mathrm{T}})^{-1}\delta$.
(b) Under $H^{(2)}_{1,n}$, $W^{(2)}_{n,\phi}$ is asymptotically distributed as non-central $\chi^2_{d_{\mathcal{M}}}$, with non-centrality parameter $J(\phi \circ \mathrm{Exp}_{q_0})_0(\mathbf{v})^{\mathrm{T}}(\hat{\Sigma}_{E,\phi;11})^{-1}J(\phi \circ \mathrm{Exp}_{q_0})_0(\mathbf{v})$. The non-centrality parameter does not depend on the choice of the co-ordinate system at q_0. Here, $J(f)_a$ denotes the Jacobian matrix of map f at a.
(c) Under $H^{(2)}_{1,n}$, $W^{(2)}_{\mathcal{M},n}$ is asymptotically distributed as non-central $\chi^2_{d_{\mathcal{M}}}$, with non-centrality parameter $m_{\tilde{q}_E}\{(\hat{\Sigma}_{E;11})^{-1}J(\mathrm{Log}_{\tilde{q}_E})_{q_0}(\mathbf{v}), J(\mathrm{Log}_{\tilde{q}_E})_{q_0}(\mathbf{v})\}$. The non-centrality parameter does not depend on the choice of the co-ordinate systems at $\tilde{q}_E$ and q_0 respectively.

We consider another scenario that there are no local charts on $\mathcal{M}$ containing both $\tilde{q}_E$ and q_0. In this case, we restate hypotheses $H^{(2)}_0$ and $H^{(2)}_1$ as follows:

$$H^{(2)}_0 : \mathrm{dist}_{\mathcal{M}}(q, q_0) = 0 \quad \textit{versus} \quad H^{(2)}_1 : \mathrm{dist}_{\mathcal{M}}(q, q_0) \neq 0. \tag{26}$$

We propose a geodesic test statistic given by

$$W_{\mathrm{dist}} = \mathrm{dist}_{\mathcal{M}}(\tilde{q}_E, q_0)^2, \tag{27}$$

which is independent of the chart (U, ϕ). Theoretically, we can establish the asymptotic distribution of W_{dist} under both the null and the alternative hypotheses as follows.

Theorem 7. Assume that all conditions in theorem 5 hold.

(a) Under $H_0^{(2)}$, nW_{dist} is asymptotically weighted $\chi^2(\lambda_1, \ldots, \lambda_{d_{\mathcal{M}}})$ distributed, where the weights $\lambda_1, \ldots, \lambda_{d_{\mathcal{M}}}$ are the eigenvalues of the matrix $\Sigma_{E,\mathrm{Log}_{q_0},11}$, which is the upper left $d_{\mathcal{M}} \times d_{\mathcal{M}}$ submatrix of the asymptotic covariance matrix $\Sigma_{E,\mathrm{Log}_{q_0}}$ of $\tilde{q}_E$ in a normal chart centred at q_0. Moreover, the weights are independent, up to a permutation, of the choice of the normal chart centred at q_*.

(b) Under the alternative hypothesis, W_{dist} is asymptotically normally distributed and we have

$$n^{1/2}\{W_{\mathrm{dist}} - \mathrm{dist}_{\mathcal{M}}(q_*, q_0)^2\} \xrightarrow{\mathrm{d}} N_{d_{\mathcal{M}}}(\mathbf{0}, D_{\mathrm{dist}}^{\mathrm{T}} \Sigma_{E,\mathrm{Log}_{q_*},11} D_{\mathrm{dist}}),$$

where D_{dist} is the column vector representation of $\mathrm{grad}_{q_*}\{\mathrm{dist}(\cdot, q_0)^2\}$ with respect to the orthonormal basis of $T_{q_*}\mathcal{M}$ associated with the normal chart used to represent the asymptotic covariance of $\tilde{q}_E$ as the matrix $\Sigma_{E,\mathrm{Log}_{q_*}}$. In particular, when q_0 is sufficiently close to q_*, then

$$n^{1/2}\{W_{\mathrm{dist}} - \mathrm{dist}_{\mathcal{M}}(q_*, q_0)^2\} \xrightarrow{\mathrm{d}} N_{d_{\mathcal{M}}}\{\mathbf{0}, 4\mathrm{Log}_{q_*}(q_0)^{\mathrm{T}} \Sigma_{E,\mathrm{Log}_{q_*},11} \mathrm{Log}_{q*}(q_0)\}.$$

Theorem 7 establishes the asymptotic distribution of W_{dist} when $\tilde{q}_{\mathrm{E}}$ and q_0 do not belong to the same chart of $\mathcal{M}$. In practice, the covariance matrix $\Sigma_{E,\mathrm{Log}_{q_*},11}$ is not available, since $\Sigma_{E,\mathrm{Log}_{q_*}}$ is not known; it also depends on the unknown true value $\boldsymbol{\beta}_*$, so we may use the estimate $\hat{\Sigma}_{E,\mathrm{Log}_{q_*}}$ as defined in theorems 3 and 4. Therefore, under the null hypothesis, the asymptotic distribution of W_{dist} can be approximated by the weighted χ^2 distribution $\chi^2(\hat{\lambda}_1, \ldots, \hat{\lambda}_{\mathrm{d}_{\mathcal{M}}})$, in which the weights $\hat{\lambda}_1, \ldots, \hat{\lambda}_{d_{\mathcal{M}}}$ are the eigenvalues of the covariance matrix $(\hat{\Sigma}_{E,\mathrm{Log}_{q_0}})_{11}/n$.

Finally, we develop a score test statistic for testing $H_0^{(2)}$ against $H_1^{(2)}$. An advantage of using the score test statistic is that it avoids the calculation of an estimator under the alternative hypothesis $H_1^{(2)}$. For notational simplicity, we consider only the intrinsic least squares estimator of $(q, \boldsymbol{\beta})$, which is denoted by $(q_0, \tilde{\boldsymbol{\beta}}_{\mathrm{I}})$, under the *null* hypothesis $H_0^{(2)}$. For any chart (U, ϕ) on $\mathcal{M}$ with $q_0 \in U$, we define

$$F_{\phi i} = (F_{\phi i,1}^{\mathrm{T}}, F_{\phi i,2}^{\mathrm{T}})^{\mathrm{T}} = \partial_{(\mathbf{t},\boldsymbol{\beta})} \mathrm{dist}_{\mathcal{M}}[f\{\mathbf{x}_i, \phi^{-1}(\mathbf{t}), \boldsymbol{\beta}\}, y_i]^2|_{\mathbf{t}=\phi(q_0),\tilde{\boldsymbol{\beta}}_{\mathrm{I}}},$$

$$U_\phi = \begin{pmatrix} U_{\mathbf{tt}} & U_{\mathbf{t}\boldsymbol{\beta}} \\ U_{\boldsymbol{\beta}\mathbf{t}} & U_{\boldsymbol{\beta}\boldsymbol{\beta}} \end{pmatrix} = \sum_{i=1}^{n} \partial^2_{(\mathbf{t},\boldsymbol{\beta})} \mathrm{dist}_{\mathcal{M}}[f\{\mathbf{x}_i, \phi^{-1}(\mathbf{t}), \boldsymbol{\beta}\}, y_i]^2|_{\mathbf{t}=\phi(q_0),\tilde{\boldsymbol{\beta}}_{\mathrm{I}}},$$

where the subcomponents $F_{\phi i,1}$ and $F_{\phi i,2}$ correspond to $\mathbf{t}$ and $\boldsymbol{\beta}$ respectively. It can be shown that the score test $W_{\mathrm{SC},\phi}$ reduces to

$$W_{\mathrm{SC},\phi} = \left(\sum_{i=1}^{n} F_{\phi i,1}\right)^{\mathrm{T}} \tilde{\Sigma}_{\phi,q}^{-1} \left(\sum_{i=1}^{n} F_{\phi i,1}\right), \tag{28}$$

where

$$\tilde{\Sigma}_{\phi,q} = (I_{d_{\mathcal{M}}}, -U_{\mathbf{t}\boldsymbol{\beta}} U_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1}) \left\{\sum_{i=1}^{n} (F_{\phi i} - \overline{F_\phi})^{\otimes 2}\right\} (I_{d_{\mathcal{M}}}, -U_{\mathbf{t}\boldsymbol{\beta}} U_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1})^{\mathrm{T}},$$

in which $\overline{F_\phi} = n^{-1}\Sigma_{i=1}^{n} F_{\phi i}$. Theoretically, we can establish the asymptotic distribution of $W_{\mathrm{SC},\phi}$ under the null hypothesis.

Theorem 8. Assume that all conditions in theorem 5 hold. We have the following results.

(a) For any suitable local chart (U, ϕ), the score test statistic $W_{\mathrm{SC},\phi}$ is asymptotically distributed as $\chi^2_{d_{\mathcal{M}}}$ under the null hypothesis $H_0^{(2)}$.
(b) Under $H_0^{(2)}$, for any other local chart (U, ϕ') with $q_0 \in U$, we have

$$W_{\mathrm{SC},\phi'} = W_{\mathrm{SC},\phi}.$$

5. Real data example

5.1. Alzheimer's disease neuroimaging initiative corpus callosum shape data

Alzheimer's disease (AD) is a disorder of cognitive and behavioural impairment that markedly interferes with social and occupational functioning. It is an irreversible, progressive brain disease that slowly destroys memory and thinking skills, and eventually even the ability to carry out the simplest tasks. AD affects almost 50% of those over the age of 85 years and is the sixth leading cause of death in the USA.

The CC, as the largest white matter structure in the brain, connects the left and right cerebral hemispheres and facilitates homotopic and heterotopic interhemispheric communication. It has been a structure of high interest in many neuroimaging studies of neurodevelopmental pathology. Individual differences in the CC and their possible implications regarding interhemispheric connectivity have been investigated over the last several decades (Paul *et al.*, 2007).

We consider the CC contour data obtained from the ADNI study. For each subject in the ADNI data set, the segmentation of the T1-weighted magnetic resonance images and the calculation of the intracranial volume were done in the *FreeSurfer* package (`http://surfer.nmr.mgh.harvard.edu/`) (Dale *et al.*, 1999), whereas the midsagittal CC area was calculated in the *CCseg* package, which was measured by using subdivisions in Witelson (1989) motivated by neurohistological studies. Finally, each T1-weighted magnetic resonance image and tissue segmentation were used as the input files of the CCSeg package to extract the planar CC shape data.

5.2. Intrinsic regression models

We are interested in characterizing the change of the CC contour shape as a function of three covariates including gender, age and AD diagnosis. We focused on $n = 409$ subjects with 223 healthy controls and 186 AD patients at baseline of the ADNI ADNI1 database. We observed a CC planar contour Y_i with 32 landmarks and three clinical variables including gender $x_{i,1}$ (0, female; 1, male), age $x_{i,2}$ and diagnosis $x_{i,3}$ (0, control; 1, AD) for $i = 1, \ldots, 409$. The demographic information is presented in Table 1.

We treat the CC planar contour Y_i as an RSS-valued response in the Kendall planar shape space Σ_2^{32}. The geometric structure of Σ_2^k for $k > 2$ is included in the on-line supplementary document. Each Y_i is specified as a 32×2 real matrix, whose rows represent the planar coordinates of 32 landmarks on Y_i. Moreover, $Y_i = (Y_{i,1} Y_{i,2})$ can be represented as a complex vector $z_i = Y_{i,1} + \mathrm{j} Y_{i,2}$ in C^{32}, where $\mathrm{j} = \sqrt{(-1)}$ and C is the standard complex space. After

Table 1. Demographic information for the processed ADNI CC shape data set including disease status, age and gender

Disease status	*Number*	*Range of age (years) (mean)*	*Females/males*
Healthy control	223	62–90 (76.25)	107/116
AD	186	55–92 (75.42)	88/98

removing the translations and normalizing to the unit 2-norm, each contour Y_i can be viewed as an element $z_i \in \mathcal{D}^{32} = \{z = (z^1, \ldots, z^{32})^{\mathrm{T}} \in C^{32} | \Sigma_{m=1}^{32} z^m = 0 \text{ and } \|z\|_2 = 1\}$. Then, after removing the two-dimensional rotations, we obtain an element $y_i = [z_i]$ in Kendall's planar shape space, $\Sigma_2^{32} = \mathcal{D}^{32}/S^1$, which has dimension 30 and is identified with the complex projective space CP^{30}.

To use our intrinsic regression model, we determined the base point p and an orthonormal basis $\{Z_1, \ldots, Z_{30}\}$ for $T_p\Sigma_2^{32}$ as follows. We initially set $p_0 = [z_0]$ with $z_0 = (1, -1, 0, \cdots, 0)^{\mathrm{T}}/\sqrt{2}$ and an orthonormal basis $\{\tilde{Z}_1, \ldots, \tilde{Z}_{30}\}$ in $T_{p_0}\Sigma_2^{32}$, where

$$\tilde{Z}_l = (1, \ldots, 1, -(l+1), 0, \ldots, 0)^{\mathrm{T}}/\sqrt{\{(l+1)(l+2)\}}.$$

Then, we projected all y_is onto $T_{p_0}\Sigma_2^{32}$ and calculated $\mathrm{Log}_{p_0}(y_i)$ for all i. Finally, we set the base point p as $\mathrm{Exp}_{p_0}\{n^{-1}\Sigma_{i=1}^{n}\mathrm{Log}_{p_0}(y_i)\}$ and then used the parallel transport to rotate the initial basis $\{\tilde{Z}_1, \ldots, \tilde{Z}_{30}\}$ to obtain a new orthonormal basis $\{Z_1, \ldots, Z_{30}\}$ at p.

We consider an intrinsic regression model with $y_i \in \Sigma_2^{32}$ as a response vector and a vector of four covariates including *gender*, *age*, *diagnosis* and the interaction *age*diagnosis*, i.e. $\mathbf{x}_i = (x_{i,1}, x_{i,2}, x_{i,3}, x_{i,4})^{\mathrm{T}}$ with $x_{i,4} = x_{i,2}x_{i,3}$. We used a single-centre link function with model parameters $(q, \boldsymbol{\beta}) \in \Sigma_2^{32} \times R^{240}$ as follows. The intercept q is specified by $q = \phi_p^{-1}(\mathbf{t}) = \mathrm{Exp}_p\{\Sigma_{l=1}^{30}(t_{2l-1} + \mathrm{j}t_{2l})Z_l\}$, where $\mathbf{t} = (t_1, \ldots, t_{60})^{\mathrm{T}} \in R^{60}$. The regression coefficient vector $\boldsymbol{\beta}$ includes four 60×1 subvectors including $\boldsymbol{\beta}^{(\mathrm{g})}$, $\boldsymbol{\beta}^{(\mathrm{a})}$, $\boldsymbol{\beta}^{(\mathrm{d})}$ and $\boldsymbol{\beta}^{(\mathrm{ad})}$, which correspond to gender x_{i1}, age x_{i2}, diagnosis x_{i3} and interaction age*diagnosis, x_{i4}, respectively. Therefore, there are 300 unknown parameters in $(\mathbf{t}^{\mathrm{T}}, \boldsymbol{\beta}^{\mathrm{T}})^{\mathrm{T}}$. We define a 30×4 complex matrix as $B = B_{\mathrm{o}} + \mathrm{j}B_{\mathrm{e}}$, with $B_{\mathrm{o}} = (\boldsymbol{\beta}_{\mathrm{o}}^{(\mathrm{g})}\ \boldsymbol{\beta}_{\mathrm{o}}^{(\mathrm{a})}\ \boldsymbol{\beta}_{\mathrm{o}}^{(\mathrm{d})}\ \boldsymbol{\beta}_{\mathrm{o}}^{(\mathrm{ad})})$, $B_{\mathrm{e}} = (\boldsymbol{\beta}_{\mathrm{e}}^{(\mathrm{g})}\ \boldsymbol{\beta}_{\mathrm{e}}^{(\mathrm{a})}\ \boldsymbol{\beta}_{\mathrm{e}}^{(\mathrm{d})}\ \boldsymbol{\beta}_{\mathrm{e}}^{(\mathrm{ad})}) \in R^{30\times 4}$, where $\boldsymbol{\beta}_{\mathrm{o}}^{(\cdot)}$ and $\boldsymbol{\beta}_{\mathrm{e}}^{(\cdot)}$ are the subvectors of $\boldsymbol{\beta}^{(\cdot)}$ formed by the odd-indexed and even-indexed components respectively, and a link function by $\boldsymbol{\mu}(\mathbf{x}_i, q, \boldsymbol{\beta}) = \mathrm{Exp}_q\{(U_{p,q}Z_1, \ldots, U_{p,q}Z_{30})B\mathbf{x}_i\} \in \Sigma_2^{32}$, $U_{q_1,q_2}\mathbf{v} = U_{z_{q_1}, z_{q_2}^*}\mathbf{v}$, with $q_1 = [z_{q_1}]$, $q_2 = [z_{q_2}]$, $z_{q_2}^* = \exp(\mathrm{j}\theta^*)z_{q_2}$ the optimal rotational alignment of z_{q_2} to z_{q_1}, given by $\bar{z}_{q_2}^{\mathrm{T}}z_{q_1} = \exp(\mathrm{j}\theta^*)|\bar{z}_{q_2}^{\mathrm{T}}z_{q_1}|$, and $U_{z_1,z_2} \in \mathrm{SU}(k)$, the unique special unitary map in the subspace generated by z_1 and z_2 that maps z_1 onto z_2, given by

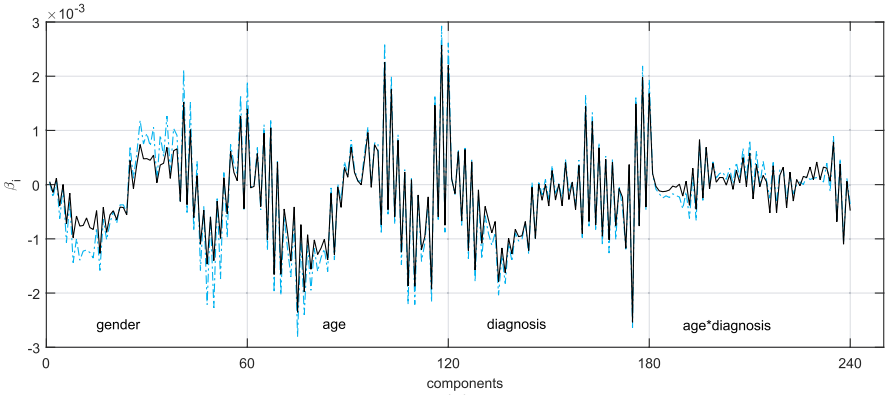
$$\begin{aligned} U_{z_1,z_2}\mathbf{v} = \mathbf{v} &- (\bar{z}_1^{\mathrm{T}}\mathbf{v})z_1 - (\bar{\tilde{z}}_2^{\mathrm{T}}\mathbf{v})\tilde{z}_2 + \{(\bar{z}_1^{\mathrm{T}}z_2)(\bar{z}_1^{\mathrm{T}}\mathbf{v}) - \sqrt{(1 - |\bar{z}_1^{\mathrm{T}}z_2|^2)}(\bar{\tilde{z}}_2^{\mathrm{T}}\mathbf{v})\}z_1 \\ &+ \{\sqrt{(1 - |\bar{z}_1^{\mathrm{T}}z_2|^2)}(\bar{z}_1^{\mathrm{T}}\mathbf{v}) + \overline{(\bar{z}_1^{\mathrm{T}}z_2)}(\bar{\tilde{z}}_2^{\mathrm{T}}\mathbf{v})\}\tilde{z}_2, \end{aligned}$$

$\mathbf{v} \in c_1^k$, in which $\tilde{z}_2 = \{z_2 - (\bar{z}_1^{\mathrm{T}}z_2)z_1\}/\sqrt{(1 - |\bar{z}_1^{\mathrm{T}}z_2|^2)}$, for $z_1, z_2 \in \mathcal{D}^{32}$. Finally, our intrinsic model is defined by $E[\mathrm{Log}_p(\bar{U}_{p,\boldsymbol{\mu}(\mathbf{x}_i,q,\boldsymbol{\beta})}^{\mathrm{T}}y_i)|\mathbf{x}_i] = \mathbf{0}$, for $i = 1, \ldots, 409$. Here, $\bar{U}$ is the matrix of complex conjugate entries of U.

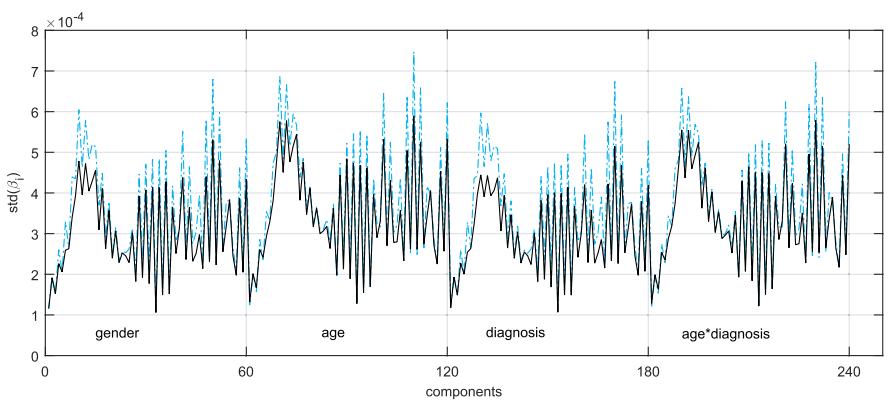
5.3. Results

We first calculated $(\hat{q}_{\mathrm{I}}, \hat{\boldsymbol{\beta}}_{\mathrm{I}}) = (\phi_p^{-1}(\hat{\mathbf{t}}_{\mathrm{I}}), \hat{\boldsymbol{\beta}}_{\mathrm{I}})$ in equation (14) and $(\tilde{q}_E, \tilde{\boldsymbol{\beta}}_E) = (\phi_p^{-1}(\tilde{\mathbf{t}}_{\mathrm{E}}), \tilde{\boldsymbol{\beta}}_E)$ in equation (18). The intercept estimates $\hat{q}_{\mathrm{I}}$ and $\hat{q}_E$ are very close to each other with $\mathrm{dist}_{\Sigma_2^{32}}(\hat{q}_{\mathrm{I}}, \hat{q}_E) < 0.0005$. Second, we compared the efficiency gain in the estimates of $\boldsymbol{\beta}$. The estimates $\hat{\boldsymbol{\beta}}_{\mathrm{I}}$ and $\hat{\boldsymbol{\beta}}_E$ of regression coefficients and their standard deviations are displayed in Figs 3(a) and 3(b). The efficiency gain in stage II is measured by the relative reduction in the variances of $\hat{\boldsymbol{\beta}}_E$ relative to those of $\hat{\boldsymbol{\beta}}_{\mathrm{I}}$, which is shown in Fig. 3(c). There is an average variance relative reduction of about 16.77% across all parameters in $\boldsymbol{\beta}$. There is an average variance relative reduction of about 12.25% for parameters in $\boldsymbol{\beta}^{(\mathrm{ad})}$, whereas there is an average relative reduction of 19.98% for parameters in $\boldsymbol{\beta}^{(\mathrm{g})}$.

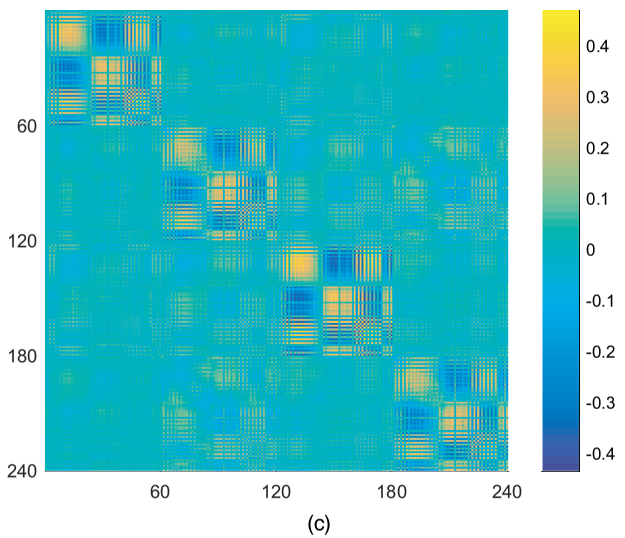
Third, we assessed whether there is an age×diagnosis interaction effect on the shape of the CC contour or not. We tested $H_0 : \boldsymbol{\beta}^{(\mathrm{ad})} = \mathbf{0}_{60}$ *versus* $H_1 : \boldsymbol{\beta}^{(\mathrm{ad})} \neq \mathbf{0}_{60}$. The Wald test statistic



(a)



(b)



(c)

Fig. 3. (a) Regression coefficient estimates (– – –, stage I; ———, stage II) and (b) their standard deviations from stages I (– – –) and II (———), and (c) the relative reduction in the variances of $\hat{\beta}_E$ relative to those of $\hat{\beta}_I$: there is an average of 16.77% relative decrease in variances of all parameters in β, corresponding to 12.25% for $\beta^{(ad)}$ and 19.98% for $\beta^{(g)}$

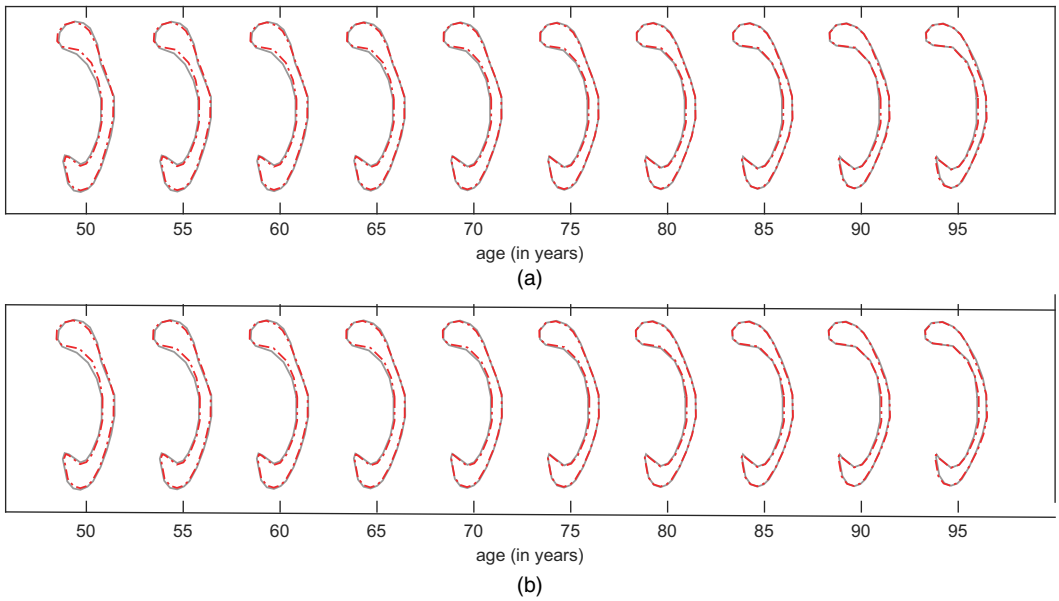


Fig. 4. Age trajectories of the intrinsic mean shapes by diagnosis within each gender group (stage II, ADNI data): (a) female group (———, normal; — · — ·, AD); (b) male group (———, normal; — · — ·, AD)

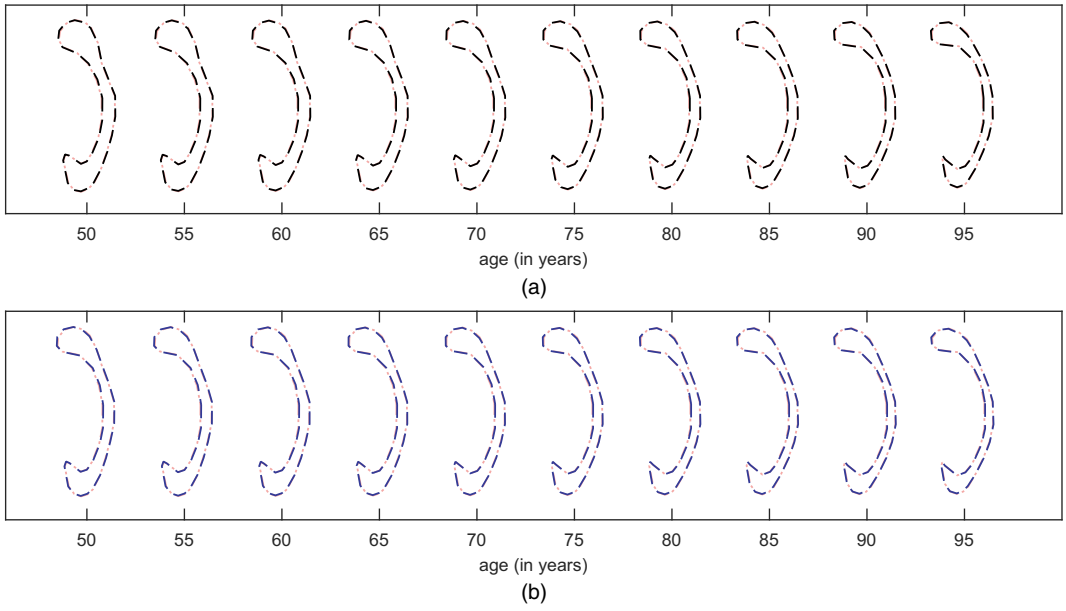


Fig. 5. Age trajectories of the intrinsic mean shapes by gender within each diagnosis group (stage II, ADNI data): (a) normal group (— · — ·, female; – – –, male); (b) AD group (— · — ·, female; — — —, male)

equals $W_{n,\phi}^{(1)} = 98.20$ with its p-value around 0.001. Thus, the data contain enough evidence to reject H_0, indicating that there is a strong age-dependent diagnosis effect on the shape of the CC contours. The mean age-dependent CC trajectories for healthy controls and AD sufferers within each gender group are shown in Fig. 4. It can be observed that there is a difference in

shape along the inner side of the posterior splenium and isthmus subregions in both male and female groups. The splenium seems to be less rounded and the isthmus is thinner in subjects with AD than in healthy controls.

Fourth, we assessed whether there is a gender effect on the shape of the CC contour or not. We tested $H_0 : \boldsymbol{\beta}^{(g)} = \mathbf{0}_{60}$ *versus* $H_1 : \boldsymbol{\beta}^{(g)} \neq \mathbf{0}_{60}$. The Wald test statistic is $W^{(1)}_{n,\phi} = 73.34$ with its p-value 0.116. Thus, it is not significant at the 0.05 level of significance. It may indicate that there is no gender effect on the shape of the CC contours. The mean age-dependent CC trajectories for the female and male groups within each diagnosis group are shown in Fig. 5. We observed similar shapes of CC contours in males and females.

6. Discussion

We have developed a general statistical framework for intrinsic regression models of responses valued in an RSS in general, and Lie groups in particular, and their association with a set of covariates in Euclidean space. The intrinsic regression models are based on the GMM estimator and therefore the models avoid any parametric assumptions regarding the distribution of the manifold-valued responses. We also proposed a large class of link functions to map Euclidean covariates to the manifold of responses. Essentially, the covariates are first mapped to the tangent bundle to the Riemannian manifold, and from there further mapped, via the manifold exponential map, to the manifold itself. We have adapted an annealing evolutionary stochastic algorithm to search for the intrinsic least squares estimator $(\hat{q}_{\mathrm{I}}, \hat{\boldsymbol{\beta}}_{\mathrm{I}})$, of $(q, \boldsymbol{\beta})$, in stage I of the estimation process, and a one-step procedure to search for the efficient estimator $(\tilde{q}_E, \tilde{\boldsymbol{\beta}}_E)$ in stage II. Our simulation study (included in the on-line supplementary document) and real data analysis demonstrate that the relative efficiency of the stage II estimator improves as the sample size increases.

Acknowledgements

We thank the Joint Editor, an Associate Editor, two referees and Professor Huiling Le for valuable suggestions, which helped to improve our presentation greatly. We also thank Dr Chao Huang and Mr Yuai Hua for processing the ADNI CC shape data set. This work was supported in part by National Science Foundation grants and National Institutes of Health grants.

Data used in preparation of this paper were obtained from the ADNI database. As such, the investigators within the ADNI contributed to the design and implementation of the ADNI and/or provided data but did not participate in the analysis or writing of this paper. A complete listing of ADNI investigators can be found at `http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf`.

References

Bhattacharya, A. and Dunson, D. B. (2010) Nonparametric Bayesian density estimation on manifolds with applications to planar shapes. *Biometrika*, **97**, 851–865.

Bhattacharya, A. and Dunson, D. (2012) Nonparametric Bayes classification and hypothesis testing on manifolds. *J. Multiv. Anal.*, **111**, 1–19.

Bhattacharya, R. and Patrangenaru, V. (2003) Large sample theory of intrinsic and extrinsic sample means on manifolds: I. *Ann. Statist.*, **31**, 1–29.

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

Dale, A. M., Fischl, B. and Sereno, M. I. (1999) Cortical surface-based analysis: I, segmentation and surface reconstruction. *Neuroimage*, **9**, 179–194.

Dryden, I. L. and Mardia, K. V. (1998) *Statistical Shape Analysis*. Chichester: Wiley.
Fletcher, P. T. (2013) Geodesic regression and the theory of least squares on Riemannian manifolds. *Int. J. Comput. Visn*, **105**, 171–185.
Fletcher, P. T., Lu, C., Pizer, S. and Joshi, S. (2004) Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Trans. Med. Imgng*, **23**, 995–1005.
Hansen, L. P. (1982) Large sample properties of generalized method of moments estimators. *Econometrica*, **50**, 1029–1054.
Healy, D. M. J. and Kim, P. T. (1996) An empirical Bayes approach to directional data and efficient computation on the sphere. *Ann. Statist.*, **24**, 232–254.
Helgason, S. (1978) *Differential Geometry, Lie Groups, and Symmetric Spaces*. New York: Academic Press.
Huckemann, S., Hotz, T. and Munk, A. (2010) Intrinsic manova for Riemannian manifolds with an application to Kendall's space of planar shapes. *IEEE Trans. Pattn Anal. Mach. Intell.*, **32**, 593–603.
Kent, J. T. (1982) The Fisher–Bingham distribution on the sphere. *J. R. Statist. Soc.* B, **44**, 71–80.
Kim, H. J., Adluru, N., Collions, M. D., Chung, M. K., Bendlin, B. B., Johnson, S. C., Davidson, R. J. and Singh, V. (2014) Multivariate general linear models (mglm) on Riemannian manifolds with applications to statistical analysis of diffusion weighted images. In *Proc. Institute of Electrical and Electronics Engineers Conf. Computer Vision and Pattern Recognition, Columbus*, pp. 2705–2712. New York: Institute of Electrical and Electronics Engineers.
Korsholm, L. (1999) The GMM estimator versus the semiparametric efficient score estimator under conditional moment restrictions. *Working Paper*. Department of Economics, University of Aarhus, Aarhus.
Lang, S. (1999) *Fundamentals of Differential Geometry*. New York: Springer.
Le, H. and Barden, D. (2014) On the measure of the cut locus of a Fréchet mean. *Bull. Lond. Math. Soc.*, **46**, 698–708.
Liang, F., Liu, C. and Carroll, R. J. (2010) *Advanced Markov Chain Monte Carlo: Learning from Past Samples*. New York: Wiley.
Machado, L. and Silva Leite, F. (2006) Fitting smooth paths on Riemannian manifolds. *Int. J. Appl. Math. Statist.*, **4**, 25–53.
Machado, L., Silva Leite, F. and Krakowski, K. (2010) Higher-order smoothing splines versus least squares problems on Riemannian manifolds. *J. Dyn. Control Syst.*, **16**, 121–148.
Mardia, K. V. and Jupp, P. E. (2000) *Directional Statistics*. Chichester: Wiley.
McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*, 2nd edn. London: Chapman and Hall.
Muralidharan, P. and Fletcher, P. (2012) Sasaki metrics for analysis of longitudinal data on manifolds. In *Proc. Institute of Electrical and Electronics Engineers Conf. Computer Vision and Pattern Recognition, Providence*, pp. 1027–1034. New York: Institute of Electrical and Electronics Engineers.
Newey, W. K. (1993) Efficient estimation of models with conditional moment restrictions. In *Handbook of Statistics*, vol. 2, *Econometrics* (eds G. S. Maddala, C. R. Rao and H. D. Vinod), pp. 419–454. Amsterdam: North-Holland.
Paul, L. K., Brown, W. S., Adolphs, R., Tyszka, J. M., Richards, L. J., Mukherjee, P. and Sherr, E. H. (2007) Agenesis of the corpus callosum: genetic, developmental and functional aspects of connectivity. *Nat. Rev. Neursci.*, **8**, 287–299.
Samir, C., Absil, P.-A., Srivastava, A. and Klassen, E. (2012) A gradient-descent method for curve fitting on Riemannian manifolds. *Found. Computnl Math.*, **12**, 49–73.
Shi, X., Styner, M., Lieberman, J., Ibrahim, J. G., Lin, W. and Zhu, H. (2009) Intrinsic regression models for manifold-valued data. In *Proc Int. Conf. Medical Image Computing and Computer-assisted Intervention, London, Sept. 20th–24th* (eds G.-Z. Yang, D. J. Hawkes, D. Rueckert, A. Noble and C. Taylor), pp. 192–199. Berlin: Springer.
Shi, X., Zhu, H., Ibrahim, J. G., Liang, F., Liberman, J. and Styner, M. (2012) Intrinsic regression models for median representation of subcortical structures. *J. Am. Statist. Ass.*, **107**, 12–23.
Spivak, M. (1979) *A Comprehensive Introduction to Differential Geometry*, vol. I, 2nd edn. Wilmington: Publish or Perish.
Su, J., Dryden, I., Klassen, E., Le, H. and Srivastava, A. (2012) Fitting smoothing splines to time-indexed, noisy points on nonlinear manifolds. *Im. Visn Comput.*, **30**, 428–442.
Witelson, S. F. (1989) Hand and sex differences in isthmus and genu of the human corpus callosum: a postmortem morphological study. *Brain*, **112**, 799–835.
Yuan, Y., Zhu, H., Lin, W. and Marron, J. S. (2012) Local polynomial regression for symmetric positive definite matrices. *J. R. Statist. Soc.* B, **74**, 697–719.
Zhu, H., Chen, Y., Ibrahim, J. G., Li, Y., Hall, C. and Lin, W. (2009) Intrinsic regression models for positive-definite matrices with applications to diffusion tensor imaging. *J. Am. Statist. Ass.*, **104**, 1203–1212.

Supporting information

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

'Regression models on Riemannian Symmetric spaces (supplementary report)'.