ADVANCED BIOSTATISTICAL METHODS FOR CURVED AND CENSORED BIOMEDICAL DATA

Emil A. Cornea

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics in the Gillings School of Global Public Health.

Chapel Hill
2014

Approved by:

Joseph G. Ibrahim

Hongtu T. Zhu

Donglin Zeng

Amy Herring

Martin Styner

ABSTRACT

EMIL A. CORNEA: Advanced Biostatistical Methods for Curved and Censored Biomedical Data
(Under the direction of Joseph G. Ibrahim and Hongtu T. Zhu)

This research was dedicated to analyze two types of biomedical data: curved data lying on a manifold and censored survival data from clinical trials.

The main part of the research aims at developing a general regression framework for the analysis of a manifold-valued response in a Riemannian symmetric space (RSS) and its association with Euclidean covariates of interest, such as age. Such data arises frequently in medical imaging, computational biology, and computer vision, among many others. We developed an intrinsic regression model solely based on an intrinsic conditional moment assumption, avoiding specifying any parametric distribution on RSS. We proposed various link functions from the Euclidean space of covariates to the RSS of responses. We constructed parameter estimates and test statistics, and determined their asymptotic distributions and geometric invariant properties. Simulation studies were used to evaluate the finite sample properties of our method. We applied our model to investigate the association between covariates, including gender, age, and diagnosis, and the shape of the Corpus Callosum contours from the Alzheimer's Disease Neuroimaging Initiative dataset, in both cross-sectional and longitudinal cases.

In oncology clinical trials, progression-free survival (PFS) has been a key endpoint to support licensing approval, and it is recommended to have the investigator's tumor assessments verified by an independent review committee blinded to study treatments, especially in open-label studies. Agreement between these evaluations may vary for

subjects with short or long PFS, while there exist no such statistical quantities that can completely account for this temporal pattern of agreements. We proposed a new method to assess temporal agreement between two time-to-event endpoints, assuming they have a positive probability of being identical. Overall scores of agreement over a period of time are also proposed. We used maximum likelihood estimation to infer the proposed agreement measures using empirical data, accounting for different censoring mechanisms including reader's censoring (event from one reader dependently censored by event from the other reader). The proposed method is demonstrated to perform well in small-sample via extensive simulation studies and is illustrated through a head and neck cancer trial.

To My Daughters, Christiana and Andreea,

and

In Loving Memory of My Brother, Florin.

ACKNOWLEDGMENTS

I would like to take this opportunity to express my sincere appreciation to all the people who encouraged and supported me throughout the years.

In particular, I would like to extend my deepest gratitude to Dr. Joseph Ibrahim and Dr. Hongtu Zhu, my advisors, for their inspired guidance, encouragement throughout the ups and downs of this work, and generous help far beyond conventional limits, both academic and non-academic, that made the completion of this dissertation possible. The experience I have gained working with them is invaluable.

I would also like to acknowledge the deepest influence on my mathematical background and the great appreciation of my former advisors, the late Professors Björn Jawerth and Martin Jurchescu.

I thank my committee members Dr. Amy Herring and Dr. Donglin Zeng for numerous, very enriched and helpful discussions throughout the past years. Thanks are also extended to Dr. Martin Styner, who served as a member of my committee.

Finally, I thank my parents for their continuous encouragement, support, trust, and unconditional love even from a great distance. And of course great thanks go out to my family, especially my little daughter Andreea, for being my greatest supporter.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: LITERATURE REVIEW

We first review the literature on curved data in non-Euclidean spaces, such as differential manifolds. We then focus on the existing literature on agreement measures in bivariate time-to-event censored data.

1.1 Manifold-Valued Data

Statistical inference for distributions on manifolds is a broad discipline with wide ranging applications. Its study has gained momentum, due to its applications in biosciences and medicine, geosciences, astronomy, computer vision and image analysis, electrical engineering, and other fields.

Statisticians are working more and more with nonlinear object data, regarding the observations as points on manifolds. The idea of a data analysis on abstract metric space goes back to a visionary paper by Fréchet (1948), where he suggested to analyze object data on separable metric spaces rather than just on linear spaces in an effort to accommodate a large variety of *elements* (objects) that need to be analyzed statistically.

Fréchet's ideas got exploited much later, especially with the increase of computational capabilities, so that his approach could be followed numerically. The first examples of data analysis on manifolds are due to Watson (1983), for directional data analysis (on spheres and real projective spaces), to Kendall (1984a), for similarity shape data analysis (on complex projective spaces), and to Chang (1988), for tectonic plates data analysis (on groups of rotations).

During late 1990s and early 2000s, the focus of modern statistical methodology was on data analysis on more complicated sample spaces, including some Lie groups - Kim

(2000), Stiefel manifolds - Hendriks and Landsman (1998), projective shape manifolds (product of real projective spaces) - Patrangenaru (2001), or on affine shape manifolds (Grassmannians) - Patrangenaru and Mardia (2003).

The common features of all these sample spaces are reflected in the fact that they are all *homogeneous spaces*. Given data in a homogeneous space to be analyzed, it is statistician's choice of selecting an appropriate Riemannian structure on the sample space, that would best address the data analysis.

In the areas of directional data analysis or shape data analysis that dominated object data analysis in its initial phase, the homogeneous spaces considered as sample spaces were compact spaces. In recent years, the attention has turned to brain imaging data and size-and-shape data analysis in structural genomics for which the sample spaces considered are noncompact (see Bandulasiri et al. (2009a) and Bandulasiri et al. (2009b)). One object of our research is to develop a unified framework for data analysis on compact and noncompact Riemanian homogeneous spaces.

Statistical analysis on general homogeneous spaces was first considered in the context of density estimation with the pioneering paper by Beran (1968). This line of research was also pursued for function estimation on Lie groups via Fourier analysis by Kim (1998; 2000), Lesosky et al. (2008), and Koo and Kim (2008a). Such methodologies found applications in medical imaging, robotics, and polymer science (Yarman and Yazici (2003; 2005), Koo and Kim (2008b)).

Data analysis of data taking values in an homogeneous spaces that admit a noncompact Riemannian structure appeared first in *Diffusion Tensor Imaging* (DTI) and in *Cosmic Microwave Background* (CMB) radiation (Schwartzman et al. 2008b). Both areas lead to analysis of random objects on the set of symmetric positive definite matrices $Sym^+(m)$. The main statistical techniques used for DTI data were parametric in nature (Schwartzman 2006, Schwartzman et al. 2008a;b). Most recently, Haff et al.

(2011), Osborne and Patrangenaru (2011), Osborne (2012), Zhu et al. (2009a), Yuan et al. (2012a), Yuan et al. (2012b), and Shi et al. (2012) used the geometric approach of Lie group action, which is related to the approach in our research.

Data belonging to some m-dimensional compact submanifold M of Euclidean space $\mathbf{R}^N$ appear in many areas of natural sciences. Directional statistics, image analysis, vector cardiography in medicine, orientational statistics, plate tectonics, astronomy, and shape analysis comprise a (by no means exhaustive) list of examples. Research in the statistical analysis of such data is well documented in the pioneering book by Mardia (1972) and more recently in the book by Mardia and Jupp (2000). In these books, as well as in many research papers , the primary emphasis is placed on the analysis of data on a circle or a sphere S^d. These are the simplest examples of compact manifolds and do not manifest the generic features of statistical inference intrinsic to compact submanifolds of Euclidean spaces.

There is an immense literature that has been devoted to shape representation, shape descriptors, or shape signatures in computer vision. An overview of several shape space models with Riemannian manifold structure, including some shape representation methods, shape spaces structures (from metric spaces to manifolds), and applications that result from such models, is given in Younes (2010). There are many ways to define the notion of shape. To simplify, a shape can be interpreted as the boundary of a two- or three-dimensional object. A shape representation is a function that assigns to a given shape a well-defined mathematical feature that will simplify further algorithms and analysis. For example, the shape of an m-dimensional object is represented by $k > m$ points in $\mathbf{R}^m$ called landmarks, which represent k locations on an object. The configuration of k landmarks is called a k-ad. The choice of landmarks is generally made with expert help in the particular field of application. Depending on the way the data are collected or recorded, the appropriate shape of a k-ad is its orbit under

a group of transformations. For example, one may look at k-ads modulo size and Euclidean rigid body motions of translation and rotation. The analysis of shapes under this invariance was pioneered by Kendall (1977; 1984b) and Bookstein (1978). Kendall identified a shape with the orbit under m-dimensional rotations of a k-ad centered at the origin and scaled to have unit size. The resulting shape spaces are called similarity shape spaces, or Kendall's shape space, and denoted by Σ_m^k. A fairly comprehensive account of parametric inference on these spaces, with many references to the literature, may be found in Dryden and Mardia (1998). When the orbits are considered under all orthogonal transformations and scaling, the resulting shape spaces are called reflection shape spaces and denoted by $R\Sigma_m^k$. A computation of the extrinsic mean reflection shape, which has remained unresolved in earlier works, was given by Bhattacharya (2008) in arbitrary dimensions, enabling one to extend nonparametric inference on Kendall type shape manifolds from 2D to higher dimensions.

Data belonging to noncompact submanifold M appears in many applications including Diffusion Tensor Imaging (DTI), Cosmic Microwave Background radiation (Schwartzman et al. 2008b), medical imaging, brain imaging (Zhu et al. 2009a, Osborne and Patrangenaru 2011, Ellingson et al. 2012), structural genomics (Bandulasiri et al. 2009a;b), computational anatomy, and statistics. A specific such a manifold is the space $\mathrm{Sym}^+(m)$ of $m \times m$ symmetric positive-definite (SPD) matrices defined as follows.

$$Sym^+(m) = \{S \in Sym(m) \,:\, \mathbf{x}^\top S\mathbf{x} > 0, \text{ for all nonzero } \mathbf{x} \in \mathbf{R}^m\},$$

where $Sym(m) \subset M(m, \mathbf{R})$ denote the linear space of all $m \times m$ symmetric matrices, and $M(m, \mathbf{R})$ is the space of all $m \times m$ matrices with real entries.

The aim of our research is to develop a general regression framework for the analysis

of manifold-valued response in a Riemannian symmetric space (RSS) and its relationship with covariates of interest, such as age, in Euclidean space. Such manifold-valued data, such as directional data and symmetric positive-definite matrices, arises frequently in medical imaging, computational biology, molecular imaging, surface modeling, and computer vision, among many others. However, little has been done when the response is in a general RSS. We develop an intrinsic regression model solely based on an intrinsic conditional moment assumption, avoiding specifying any parametric distribution in RS space. We propose various link functions to map from the Euclidean space of covariates to the the RS space 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 on unknown parameters. Simulations studies are used to evaluate the finite sample property of our methods and a real data set from Alzheimer's Disease Neuroimaging Initiative (ADNI) database is analyzed to illustrate the use of our test statistics.

1.2 Statistical Analysis of Manifold-Valued Data in the Literature

1.2.1 Fréchet Mean Set

The earliest literature on statistical analysis on spaces other than Euclidean ones goes back to mid 20-th century, when Fréchet (1948) introduced the notion of the mean for a distribution on a metric space (M, d). Let Q be a probability distribution on (M, d), i.e. on the Borel σ-algebra of$M$.

Definition 1.2.1. *The Fréchet mean set of Q, denoted C_Q, is the set of all minimizers of the Fréchet function F on M defined by*

$$F(p) = \int_M d^2(y, p)Q(dy),$$

assuming that $F(p) < \infty$ for some $p \in M$. If there is a unique minimizer, say μ_F, then $C_Q = \{\mu_F\}$ and μ_F is called the Fréchet mean of Q.

The Fréchet mean set C_Q is a natural index of location for the probability distribution Q.

Since their original definition as metrical means by Fréchet in 1948 such means have found much interest. Independently, on Riemannian manifolds with respect to the Riemannian metric, Kobayashi and Nomizu (1996) defined the corresponding metrical means in the original space as *centers of gravity*. With application in landmark based shape analysis in mind, Ziezold (1977) extended the concept to quasi-metrical means (in the sense that the distance function $d : M \times M \to [0, \infty)$ is not required to be symmetric). Further generalizations can be found in Huckemann (2011).

Applications of this concept for data analysis did not get too much traction until it could have been followed numerically. Due to the high computational complexity on the manifolds, it was not before mid-late 1980s when data analyses on manifolds were performed.

The concept of variation was introduced in Bhattacharya and Patrangenaru (2002), where it has been referred to as the total variance, as the minimum value of F on M. It is called the *Fréchet variation* of Q and denoted by V.

Definition 1.2.2. *If $Y_1, Y_2, ..., Y_n$ are independent and identically distributed (iid) M-valued random variables defined on some probability space $(\Omega, \mathcal{F}, \mathcal{P})$ with common distribution Q, and $\hat{Q}_n := 1/n \sum_{i=1}^{n} \delta_{Y_i}$ is the corresponding empirical distribution, then the Fréchet mean set of $\hat{Q}_n$ is called the sample Fréchet mean set, denoted by $C_{\hat{Q}_n}$. The Fréchet variation of $\hat{Q}_n$ is called the sample Fréchet variation and denoted by $\hat{V}_n$.*

Existence of the Fréchet mean set. It was shows that under mild assumptions, the minimum value of F on M is attained, thereby proving that the Fréchet mean set

is nonempty, as proved by Bhattacharya and Patrangenaru (2003) and Bhattacharya (2008).

Proposition 1.2.1. *Suppose (M, d) is a metric space such that every closed and bounded subset of M is compact (the Heine-Borel property). If the Fréchet function F of Q is finite for some $p \in M$, then C_Q is nonempty and compact.*

Consistency of the Fréchet mean set. An important question on estimation of location is that of consistency. Two different Strong-Consistency results for Fréchet sample means have been obtained by Ziezold (1977), Theorem I, for quasi-metrical means on separable spaces (i.e. containing a dense countable subset) and by Bhattacharya and Patrangenaru (2003), Theorem 2.3, for metrical means on spaces with Heine-Borel property (i.e. that every closed bounded set is compact).

Theorem 1.2.1. *(Strong Consistency - Ziezold (1977)) Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and (M, d) a separable quasi-metric space. Let $Y_1, Y_2, \ldots$ be independent, identically distributed random M-valued variables, $Y_i : \Omega \to M$, $i = 1, 2, \ldots$, with common distribution Q, such that $F(\mathrm{p}) < \infty$ for some $\mathrm{p} \in M$. Then for almost all $\omega \in \Omega$*

$$\bigcap_{n=1}^{\infty} \overline{\bigcup_{k=n}^{\infty} C_{Q_k(\omega)}} \subseteq C_Q,$$

where $\overline{A}$ denotes the closure of the set A.

Theorem 1.2.2. *(Uniform Strong Consistency - Bhattacharya and Patrangenaru (2003)) Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and (M, d) a metric space such that every closed and bounded subset of M is compact (the Heine-Borel property). Let $Y_1, Y_2, \ldots$ be independent, identically distributed random M-valued variables, $Y_i : \Omega \to M$, $i = 1, 2, \ldots$, with common distribution Q, such that $F(\mathrm{p}) < \infty$ for some $\mathrm{p} \in M$. Then given any*

$\varepsilon > 0$ and almost all $\omega \in \Omega$, there exist a number $N = N(\varepsilon, \omega)$ such that

$$\bigcup_{k=N}^{\infty} C_{\hat{Q}_k(\omega)} \subseteq \{\mathrm{p} \in M : d(C_Q, \mathrm{p}) \leq \varepsilon\}.$$

In particular, if $C_Q = \{\mu_F\}$, then the sample Fréchet mean μ_{F_n} (any measurable selection from $C_{\hat{Q}_n}$) is a strongly consistent estimator of μ_F.

Bhattacharya and Patrangenaru, 2003,, in Remark 2.5, note that Ziezold's strong-consistency implies theirs for compact metric spaces M, but not for noncompact M.

The following strong consistency of $\hat{V}_n$ as an estimator of V is due to Bhattacharya and Bhattacharya (2008).

Theorem 1.2.3. *Suppose (M, d) is a metric space with Heine-Borel property and F is finite on M. Then $\hat{V}_n$ is a strongly consistent estimator of V .*

Asymptotic theories for the Fréchet means on manifolds was established in Bhattacharya and Patrangenaru (2003; 2005) and Huckemann (2011).

Central Limit Theorems. The "δ-method" allows to formulate Central-Limit Theorems (CLTs) for differentiable images of random variables. The following definition of CTL for random manifold-valued variables was given in Huckemann (2011).

Definition 1.2.3. *Let M be a smooth m-dimensional manifold. We say that a M-valued estimator $\mu_n(\omega)$ of $\mu \in M$ satisfies a Central-Limit-Theorem (CLT), if in any local chart (U, ϕ) near μ there is a suitable $m \times m$ matrix A_ϕ and a $m \times m$ positive semi-definite symmetric matrix Σ_ϕ such that*

$$\sqrt{n}\, A_\phi \left(\phi(\mu_n) - \phi(\mu)\right) \xrightarrow{d} N_m(\mathbf{0}, \Sigma_\phi)$$

in distribution as $n \to \infty$.

In most applications A_ϕ is non-singular, then as a consequence of the "δ-method", for any other chart (U', ϕ') near μ, we have

$$A_{\phi'}^{-1}\Sigma_{\phi'}\left(A_{\phi'}^{-1}\right)^\top = J(\phi' \circ \phi^{-1})_{\phi(\mu)} A_\phi^{-1}\Sigma_\phi \left(A_\phi^{-1}\right)^\top J(\phi' \circ \phi^{-1})_{\phi(\mu)}^\top$$

where $J(\cdot)_a$ denotes the Jacobi-matrix of the first order derivatives at a.

Suppose now that d is a metric on a differentiable manifold M, and $\mathrm{p} \to d(\mathrm{q}, \mathrm{p})$ is at least twice continuously differentiable on M, for any $q \in M$. We can only expect a CLT for the Fréchet mean set to hold under additional regularity conditions concerning the expectation of derivatives of d. To this end, in a local chart (U, ϕ) on M, denote by $\mathrm{grad}_2 d(\mathrm{q}, \mathrm{p})^2$ the gradient of $u \to d^2(\mathrm{q}, \phi^{-1}(u))$ at $u = \phi^{-1}(p)$ for $\mathrm{p} \in U$, and $\mathrm{Hess}_2 d(\mathrm{q}, \mathrm{p})^2$ the corresponding Hessian matrix of the second order derivatives. We mention here two CLT results.

First, Bhattacharya and Patrangenaru (2005) establish a CLT when there is a unique Fréchet mean μ_F of Q and there is a local (U, ϕ) that almost covers M, i.e. $Q(U) = 1$.

Theorem 1.2.4. *(CLT for Fréchet means - Bhattacharya and Patrangenaru (2005)) Let Q be a probability measure on a differentiable manifold M endowed with a metric $d(\cdot, \cdot)$ such that (M, d) has the Heine-Borel property. Let Y_i, $i = 1, 2, \ldots$, be i.i.d. random variables on M with common distribution Q, and $\mu_{n,F}$ be a measurable selection from the Fréchet mean set (w.r.t. d) of the empirical $\hat{Q}_n = \frac{1}{n}\sum_{i=1}^n \delta_{Y_i}$. Assume*

(i) the Fréchet mean μ_F exists,

(ii) there exists a local chart (U, ϕ) such that $Q(U) = 1$,

(iii) the map $u \to d^2(\mathrm{q}, \phi^{-1}(u))$ is twice differentiable on $\phi(U)$,

(iv) $E_Q(\|\mathrm{grad}_2 d(Y_1, \mu_F)^2\|^2) < \infty$, $E_Q\left(\left|(\mathrm{Hess}_2 d(Y_1, \mu_F)^2)_{k,l}\right|^2\right) < \infty$, and

(v) $E_Q\left(\sup_{\mathrm{p}\,:\,d(\mathrm{p},\mu_F)<\varepsilon}\left|(\mathrm{Hess}_2 d(Y_1,\mathrm{p})^2)_{k,l}-(\mathrm{Hess}_2 d(Y_1,\mu_F)^2)_{k,l}\right|\right)\to 0$ *as* $\varepsilon\to 0$, $k,l=1,\ldots,m$.

Then,

(a) $\mu_{n,F}$ *is a consistent estimator of* μ_F, *and*

(b) $\sqrt{n}A_\phi(\phi(\mu_{n,F})-\phi(\mu_F))\xrightarrow{d} N_m(\mathbf{0},\Sigma_\phi)$,

with

$$A_\phi = E_Q(\mathrm{Hess}_2 d(Y_1,\mathrm{p})^2) \;\; \textit{and} \;\; \Sigma_\phi = \mathrm{Cov}(\mathrm{grad}_2 d(Y_1,\mu_F)^2). \tag{1.1}$$

The assumption above on the existence of a local chart (U,ϕ) such that $Q(U)=1$ is less restrictive than it may seem. If $\mathbf{m}$ is a Riemannian structure on M and Q is absolutely continuous with respect to the volume measure, then, for any given $\mathrm{p}\in M$, the complement U of the cut locus $C(\mathrm{p})$, is the domain of definition of such a local chart, $\phi=\mathrm{Log}_\mathrm{p}$, the Riemannian logarithmic map of M at p. For example, when $M=S^m$, the unit sphere in $\mathbf{R}^{m+1}$ with the canonical Riemanian structure, for a given $\mathrm{p}\in S^m$, the maximal normal chart centered at p is given by $U=S^m\setminus\{-\mathrm{p}\}$ and $\phi(y)=\mathrm{Log}_\mathrm{p}(y)=(y-(\mathrm{p}^\top y)\mathrm{p})\arccos(\mathrm{p}^\top y)/\sqrt{1-(\mathrm{p}^\top y)^2}$. Any probability measure Q on S^m that has no mass concentrated at $-\mathrm{p}$ satisfies $Q(U)=1$.

Second, the most general CLT result to date is due to Huckemann (2011). The so-called Fréchet ρ-mean set of a probability distribution Q on M is defined in a more general setting, where the distance d is replaced by a continuous function $\rho: M\times P\to[0,\infty)$, where M is a topological space and (P,d_P) is a *space with distance* (in the sense that P is a topological space and $d_P: P\times P\to[0,\infty)$ is a continuous map that vanishes on the diagonal $\{(p,p): p\in P\}$). Fréchet ρ-mean set of Q is a subset of P. In this setting, Huckemann assumes the following two properties for ρ: (c1) *continuity* in the

second argument *uniform* over the first argument, and (c2) a version of *coercivity* in the second argument: *there are* $p_0 \in P$ *and* $C > 0$ *such that* $\Pr(\rho(Y, p_0) < C) > 0$, *for* $Y \sim Q$, *and that such that for every sequence* $p_n \in P$ *with* $d_P(p_0, p_n) \to \infty$ *there is a sequence* $K_n \to \infty$ *with* $\rho(y, p_n) > K_n$ *for all* $y \in M$ *with* $\rho(y, p_0) < C$; *moreover, if* $p_n \in P$ *with* $d_P(p^*, p_n) \to \infty$ *for some* $p^* \in P$, *then* $d_P(p_0, p_n) \to \infty$. Both properties are valid if $M = P$ and ρ is a quasi-metric. Huckemann showed that property (c1) implies the strong consistency, assuming that the Fréchet function is finite at least at one point in P (Theorem 3.4 in Huckemann (2011)). He also showed that, when P enjoys the Heine-Borel property and the Fréchet ρ-mean set is nonempty, the conditions (c1) and (c2) together imply uniform strong consistency (Theorem 3.5 in Huckemann (2011)). It may be pointed out that it is the assumption of some symmetries of P that often causes the Fréchet ρ-mean set to contain more than one element (see, e.g., Proposition 2.2 in Bhattacharya and Patrangenaru (2005)). This situation is taken care of too in Huckemann's CLT result as follows.

Theorem 1.2.5. *(CLT - Theorem 3.8 in Huckemann (2011)) Assume* $P = R$, *where* R *is a smooth manifold, and there is a self-understood discrete group* H *acting smoothly on* R *such that* $\{p' \in R : d_R(p, p') = 0\} = \{hp : h \in H\}$, *for any* $p \in R$. *Suppose that* μ *is a point in the Fréchet* ρ*-mean set unique up to the action of* H *on the manifold* R *(i.e. Fréchet* ρ*-mean set equals* $H\mu$*) with respect to a continuous* $\rho : M \times R \to [0, \infty)$, ρ^2 *smooth in the second argument, satisfying uniform consistency, where* M *is a topological space. Let* Y_i, $i = 1, 2, \ldots$, *be i.i.d. random variables on* M *with common distribution* Q. *If*

(i) probability measure Q *on* M *has compact support or*

(ii) in a suitable local chart on R *near* μ, $E_Q\left(\mathrm{grad}_2\rho(Y_1, \mu)^2\right)$ *exists,* $Cov_Q\left(\mathrm{grad}_2\rho(Y_1, \mu)^2\right)$ *exists, and* $E_Q\left(\mathrm{Hess}_2\rho(Y_1, \nu)^2\right)$ *exists for* ν *near* μ *and is continuous at* $\nu = \mu$,

then for any measurable choice μ_n^o in the sample Fréchet ρ-mean set there is a sequence $h_n \in H$ such that $\mu_n = h_n \mu_n^o$ satisfies a CLT. In a suitable local chart (U, ϕ) on R near μ the corresponding matrices in CLT are given by $A_\phi = E_Q\left(\text{Hess}_2\rho(Y_1, \mu)^2\right)$ and $\Sigma_\phi = Cov_Q\left(\text{grad}_2\rho(Y_1, \mu)^2\right)$.

1.2.2 Intrinsic and Extrinsic Mean Sets

If M is a Riemannian manifold, the Fréchet mean (set) with respect to the geodesic distance $d = d_{\mathbf{m}}$ is defined to be the *intrinsic mean (set)* of Q and denoted $\mu_I(Q)$. The corresponding Fréchet variation is called the *intrinsic variation* of Q and denoted $V_I(Q)$.

In the case of M is connected C^∞ Riemannian manifold with a metric tensor $\mathbf{m}$ and geodesic distance $d_{\mathbf{m}}$, with $(M, d_{\mathbf{m}})$ a complete metric space, Theorem 2.1 in Bhattacharya and Patrangenaru (2003) shows that (i) the intrinsic Fréchet mean set is compact, (ii) for each point μ in the intrinsic mean set, the Euclidean mean of the of the distribution on the tangent space at μ of the Riemanian logarithmic map is zero, and (iii) in the case of simply connected M of nonpositive curvature, the intrinsic mean exists if $F(\cdot)$ is finite. A particular case of this result, when M is a Bookstein's shape space of labeled triangles, with Riemannian metric of constant negative curvature is due to Le and Kume (2000). From a result of Karcher (1977) it follows that if the distribution is sufficiently concentrated then the intrinsic mean exists. For complete Riemannian manifolds, it seems that the sharpest uniqueness result to date for the intrinsic mean is due to Afsari (2011): *if Q is supported in a ball of radius less than the geodesic convexity radius of $(M, \mathbf{m})$, then the minimizer of F_Q (for $d = d_{\mathbf{m}}$) is unique.* For planar shape space $\mathbf{CP}^{k-2}$, a useful necessary and sufficient condition for the existence of an intrinsic mean is proved by Le (1998) for distributions Q which are absolutely continuous (with respect to the volume measure) with a density that is a

function only of the distance from a given point.

Much of the literature in the field deals with special cases of what it is called *extrinsic mean*, perhaps because of the difficulties involved in proving the existence of an intrinsic mean and in computing the intrinsic sample mean, even when it exists. It is simpler both mathematically and computationally to carry out an *extrinsic* analysis on M, by embedding it into some Euclidean space $\mathbf{R}^N$ via some map $J : M \to \mathbf{R}^N$ such that both J and its derivative are injective, and for which $J(M)$ has the induced topology from $\mathbf{R}^N$. Then J induces the metric $d_J(x, y) = \|J(x)-J(y)\|_{\mathbf{R}^N}$ on M, where $\|\cdot\|_{\mathbf{R}^N}$ denotes Euclidean norm ($\|u\|^2_{\mathbf{R}^N} = \sum_{i=1}^N u_i^2$, for any $u = (u_1, u_2, ..., u_N)^\top$). This is called the *extrinsic distance* on M. Among the possible embeddings, one seeks out *equivariant embeddings* which preserve many of the geometric features of M. For a Lie group H acting on a manifold M, an embedding $J : M \to \mathbf{R}^N$ is *H-equivariant*, if there exists a group homomorphism $\varphi : H \to GL(N, \mathbf{R})$ such that $J(a \cdot p) = \varphi(a)J(p)$ for all $p \in M$ and all $a \in H$. Here, $GL(N, \mathbf{R})$ is the general linear group of all $N \times N$ non-singular matrices. The extrinsic mean and variations of a probability distribution Q on M are defined (see below) with respect to the embedding J. The notion of extrinsic mean on a manifold was introduced independently by Hendriks and Landsman (1998) and Patrangenaru (1998), and later considered in detail in Bhattacharya and Patrangenaru (2003; 2005).

The Fréchet mean set of Q with respect to the distance d_J is called the *extrinsic mean set* of Q and the Fréchet variation of Q is called the *extrinsic variation* of Q. If Y_i, $i = 1, \ldots, n$, are iid observations from Q, then the Fréchet mean set of $\hat{Q}_n$ is called the *sample extrinsic mean set* and the Fréchet variation of $\hat{Q}_n$ is called the *sample extrinsic variation*.

In case $J(M) = \tilde{M}$ is a closed subset of $\mathbf{R}^N$, for every $u \in \mathbf{R}^N$, there exists a compact set of points in $\tilde{M}$ whose distance from u is the smallest among all points in

$\tilde{M}$. This set is the set of projections of u on $\tilde{M}$ and denote it by $P_{\tilde{M}}(u) = \{x \in \tilde{M} : \|x - u\|_{\mathbf{R}^N} \leq \|y - u\|_{\mathbf{R}^N}$ for any $y \in \tilde{M}\}$. If this set is a singleton, u is said to be a *nonfocal* point of $\mathbf{R}^N$ (with respect to $\tilde{M}$), otherwise it is said to be a *focal* point of $\mathbf{R}^N$.

For a given embedding J, Bhattacharya and Patrangenaru (2003) established the relationships between the extrinsic mean and variation of Q on M and the mean and variation of the push forward probability distribution $\tilde{Q} = Q \circ J^{-1}$ of Q onto $\mathbf{R}^N$. They showed that the extrinsic mean set of Q is given by $J^{-1}(P_{\tilde{M}}(\tilde{\mu}))$, where $\tilde{\mu}$ is the mean of $\tilde{Q}$, and the extrinsic variation of Q is given by

$$V = \int_{\mathbf{R}^N} \|x - \tilde{\mu}\|^2 \tilde{Q}(dx) + \|\tilde{\mu} - \mu\|^2,$$

where $\mu \in P_{\tilde{M}}(\tilde{\mu})$.

An asymptotic theory for the intrinsic and extrinsic means on manifolds was established in Bhattacharya and Patrangenaru (2003; 2005). Asymptotic distributions of extrinsic sample means were derived. Explicit computations of these means of $\tilde{Q}_n$ and their asymptotic dispersions were carried out for distributions on the sphere S^d (directional spaces), real projective space $\mathbf{RP}^{N-1}$ (axial spaces), and $\mathbf{CP}^{k-2}$ (planar shape spaces). Nonparametric inference procedures for estimation and testing problems for sample Fréchet means on manifolds were also derived and bootstrap methods for these problems were presented, with applications to distributions on S^d, $\mathbf{RP}^{N-1}$, $\mathbf{CP}^{k-2}$ with respect to Veronese-Whitney embeddings, and a 3-dimensional shape space Σ_3^4. A detailed theory of shape-spaces Σ_m^k can be found in the pioneering paper by Kendall (1984b), and a brief description of these spaces is presented below.

Central Limit Theorems for Intrinsic Means. In the case Q is a probability distribution on M whose support is compact and is contained in a local chart, an immediate CLT result for intrinsic sample means follows from Theorem 1.2.4 above.

Corollary 1.2.1. *Let $(M, \mathbf{m})$ be a m-dimensional Riemannian manifold and $d = d_{\mathbf{m}}$ be the geodesic distance. Let Q be a probability distribution on M whose support is compact and is contained in a local chart (U, ϕ). Assume that (i) the intrinsic mean μ_I exists, (ii) the map $u \to d^2(\mathrm{q}, \phi^{-1}(u))$ is twice continuously differentiable on $\phi(U)$ for each $\mathrm{q} \in U$ and A_ϕ and Σ_ϕ defined as in (1.1). Then the conclusion of Theorem 1.2.4 holds for the intrinsic sample mean $\mu_{n,I}$.*

Two main CLT results for intrinsic means are established by Bhattacharya and Patrangenaru (2005). One is relative to normal charts on M when the support of Q is in a geodesically convex ball whose radius depends on the sectional curvature of M, and the second when the uniqueness of the intrinsic mean is assumed. The other allows for multiple intrinsic means due to the invariance of Q under a group of symmetries.

Theorem 1.2.6. *Let $(M, \mathbf{m})$ be a Riemannian manifold and let $d = d_{\mathbf{m}}$ be the geodesic distance. Let Q be a probability measure on M whose support is contained in a closed geodesic ball $\overline{B_r} = \overline{B_r}(\mathrm{p}_0)$ with center p_0 and radius r which is disjoint from the cut locus $C(\mathrm{p}_0)$. Assume $r < \frac{\pi}{4K}$, where K^2 is the supremum of the sectional curvatures in $\overline{B_r}$ if this supremum is positive, or zero if this supremum is nonpositive. Then*

(a) the intrinsic mean μ_I (of Q) exists, and

(b) then the CLT from the conclusion of Theorem 1.2.4 holds for any measurable intrinsic sample mean $\mu_{n,I}$ of $\hat{Q}_n = \frac{1}{n}\sum_{i=1}^n \delta_{Y_i}$, under the normal chart $\phi = \mathrm{Log}_{\mathrm{p}_0}$.

We note here, that if the supremum of the sectional curvatures (of a complete manifold M) is nonpositive, and support of Q is contained in $\overline{B_r}$, for some $r > 0$, then the hypotheses of Theorem 1.2.6 are satisfied, and the conclusions (a) and (b) hold. One may apply this theorem even with $r = \infty$.

The assumptions in Theorem 1.2.6 on the support of Q for the existence of μ_I are too restrictive for general applications. But without additional structures they cannot

be entirely dispensed with, as can be shown by letting Q be the uniform distribution on the equator of S^2. For example, as it was mentioned above, Le (1998) gives necessary and sufficient conditions for the existence of the intrinsic mean μ_I of an absolutely continuous (w.r.t to the volume measure) Q on the projective space $\mathbf{CP}^{k-2}$, $k \geq 3$, with radially symmetric density. The standard Riemannian structure on $\mathbf{CP}^{k-2}$ is induced by the circular arc metric, i.e. the geodesic distance d is given by $\cos(d([\boldsymbol{\zeta}], [\boldsymbol{\xi}])) = |\boldsymbol{\zeta}^\top \boldsymbol{\xi}|$, where $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_{k-1})^\top, \boldsymbol{\xi} = (\xi_1, \ldots, \xi_{k-1})^\top \in \mathbf{C}^{k-1}$, with $\sum_i |\zeta_i|^2 = \sum_i |\xi_i|^2 = 1$. Let $f([\zeta])$ be the density function of Q with respect to the volume measure $d\omega$ on $\mathbf{C}^{k-2}$. If $f(\cdot)$ can be expressed as a non-increasing function of the distance d of $[\boldsymbol{\zeta}]$ from a fixed point $[\boldsymbol{\mu}]$ and is strictly decreasing on a set of positive measure, then $[\boldsymbol{\mu}]$ is the unique intrinsic Fréchet mean of Q.

Thus, assuming the that the intrinsic mean is unique, the following result may be more generally applicable than Theorem 1.2.6.

Theorem 1.2.7. *(CLT - Theorem 2.3 in Bhattacharya and Patrangenaru (2005)) Let Q be absolutely continuous with respect to the volume measure on a Riemannian manifold $(M, \mathbf{m})$. Assume that μ_I is exists, the integrability conditions (iv) in Theorem 1.2.4 hold, the Hessian matrix A_ϕ of the Fréchet function F with respect to a local chart ϕ near μ_I is is nonsingular, and the covariance matrix Σ_ϕ of the* $\mathrm{grad}_2 d(Y_1, \mu_I)^2$ *is nonsingular. Then $\sqrt{n}\,(\phi(\mu_n) - \phi(\mu_I)) \xrightarrow{d} N(\mathbf{0}, A_\phi^{-1} \Sigma_\phi A_\phi^{-\top})$.*

1.3 Longitudinal Data on Manifolds

Many scientific questions can be expressed in terms of changes or alterations of a dynamical process. In camera surveillance, one aims at distinguishing normal from abnormal behaviors behind the large variety of the shapes and the motions of the silhouettes in video sequences. In neuroscience, one studies the neurodevelopment or the neurodegeneration of the brain and its related structures. Every brain has

a different shape, whereas its maturation may follow some common patterns that one would like precisely to describe and quantify. The computational anatomy, an emerging discipline at the interface of geometry, statistics and image analysis, aims at modeling and analyzing the biological shape of tissues and organs. The goal is to estimate representative organ anatomies across diseases, populations, species or ages, to model the organ development across time (growth or aging), to establish their variability, and to correlate this variability information with other functional, genetic or structural information. In clinical studies, one wants to characterize anatomical or functional changes due to disease progression, clinical intervention or therapy. In cardiac imaging, one looks for abnormal patterns in the heart motion. What make these questions so challenging is the nature of the object of interest and that it changes in appearance in different situations. The observed object's features, such as the shape, are is inherently nonlinear and high-dimensional. Because of this, manifold representations of the data have proven to be effective. These problems can be addressed by statistical analysis of longitudinal data that takes values in a Riemannian manifold. The Riemannian structure provides useful tools to carry out such analyzes.

Longitudinal analysis differs from the usual cross-sectional variability analysis in that it takes into account the inherent correlation of repeated measurements of the same individuals. It must also provide a model of how an individual subject's trajectory changes relative to another subject. At the population level, we typically analyze how the subjects are distributed within a group by estimating a mean configuration and its variance. For longitudinal data, the mean configuration may be a "mean growth scenario", which averages the growth patterns in the population. The analysis of its variance explains how each subject's trajectory differs from the mean growth scenario. Such a statistical approach based on mean and variance is well-known for scalar measurements and for analysis of cross-sectional shape data, for which the mean is usually

called “template” or “atlas”. The extension of these concepts for longitudinal shape data is challenging, as there is no consensus about how to combine shape changes over time and shape changes across subjects.

There has been extensive research for the analysis of Euclidean longitudinal data in the last few decades. Diggle et al. (2002) and Fitzmaurice et al. (2008) provided a comprehensive overview of various models and methods for the analysis of longitudinal data, among others. In the analysis of longitudinal data, three types of models are commonly used: mixed effects models, GEE models, and transitional models.

Recent work suggests that attempts to describe anatomical shapes using flat Euclidean spaces undermines our ability to represent natural biological variability (Fletcher et al. 2004a, Grenander and Miller 1998).

A number of longitudinal growth models have been developed to provide this type of analysis to time-series imagery of a single subject (e.g., Beg (2004); Clatz et al. (2005), Miller (2004); Thompson et al. (2000)). While these methods provide important results, their use is limited by their reliance on longitudinal data, which can be impractical to obtain for many medical studies. Also, while these methods allow for the study of an individualâĂŹs anatomy over time, they do not apply when the average growth for a population is of interest.

Related work in longitudinal analysis includes several approaches in the setting of diffeomorphic transformations, which form an infinite-dimensional manifold, applied to image sequences. Durrleman et al. (2009) construct spatiotemporal image atlases from longitudinal data. Qiu et al. (2009) use parallel translation to bring individual trajectories to a common point for comparison. Lorenzi et al. (2011) use a hierarchical model on stationary velocity fields, in a framework that does not include a Riemannian metric on the manifold of diffeomorphisms. An important shortcoming of these approaches is that they do not model distances between trajectories. This makes it

difficult to compare the differences in trends of two groups, or even to rigorously define the concept of the variance of a population of trends.

The aim of this research is to extended our framework of intrinsic regression models for cross-sectional data to fixed and random effect models for the analysis of manifold-valued measures from longitudinal studies.

1.4 Agreement Assessment in Bivariate Time-to-Event Times: Motivation

Assessing agreement is often of interest in clinical studies and biomedical sciences to evaluate the similarity of measurements produced by different methods on the same subjects. For examples, when a new assay or instrument is developed, it is important to assess whether the new assay or instrument can reproduce the results of a traditional method. Additionally, the strength of agreement can help researchers decide whether a simple measurement is an acceptable replacement for a more expensive gold standard method and whether measurements obtained by different methods or instruments are comparable or not. Given the importance of agreement studies in biomedical sciences, there has been extensive literature on measuring agreement for categorical and continuous outcomes, e.g. Cohen (1960; 1968), Kraemer (1980), Kraemer et al. (2002), Lin (1989; 2000) among others.

The time to an event is the main outcome variable of interest in many medical and public health studies, including the number of years from an exposure to onset of a disease, age of onset of an illness, and the number of weeks from entry into a clinical trial to remission of a disease. A potential problem in all such researches is the fact that raters may disagree about the occurrence of events, or raters who agree on the occurrence of an event may disagree on the time to the event. Such inconsistencies also can occur in studies involving age-of-onset of certain illnesses or disorders. However, very limited work has been done to assess agreement between time-to-event data with

censored observations.

Censoring is an issue complicating the assessment of reliability of time to event data. The presence of censoring raises problems in the application of commonly used measures of reliability or reproducibility. The nature of censoring depends on the study design. Right censored data can be produced in follow up studies if no event is reported by the time the study ends or if a subject drops out. Left-censored data may occur, for example, in a study of age-of-onset when subjects of varying ages are diagnosed as having the disease at baseline but no information is available on age of onset. In such studies, age-of-onset is censored by baseline age.

There are several agreement measures for assessing dependence or agreement for censored bivariate time-to-event data in the literature. The rank-based Kendall's coefficient of concordance τ (Section 4.2, Hougaard (2000)) is a measure of overall dependence. It is simple and can be estimated non-parametrically for censored data, but does not measure the degree of agreement at a single time point. In the aforementioned extreme case, the data achieves the maximum value of Kendall's τ, even that there is no agreement of PD time at all. Liu et al. (2005) provided an estimation method of the concordance correlation coefficient for time-to-event data. It is a correlation type of measure and has the same issue as Kendall's τ. Guo and Manatunga (2009) proposed a modified weighted kappa coefficient to measure agreement between discrete bivariate survival times, but it requires discretization of continuous outcomes. Amit et al. (2011) also proposed two discrepancy rates defined as the simple proportion of subjects whose PD time is strictly greater by a reader than the other. These rates are naïve approaches by definition and do not fully utilize the temporal nature of time-to-event data. Guo et al. (2013) proposed a new agreement measure, which is formulated as the chance-corrected concordance between survival processes on the absolute distance scale.

Our research in this area is motivated by a small phase 2 head and neck cancer trial. Progression-free survival (PFS), the time from ramdomization until disease progression (PD) or death, is a key endpoint to support licensing approval. The PD is determined by the investigator (local assessment), and also assessed by an independent review committee (IRC) blinded to the study (central assessment). Among a random subset of 92 subjects followed-up in the trial, the local assessment yields 82 local PFS events while the central assessment gives 72 events and the number of agreed events is 35. We propose a new method to assess temporal agreement between two time-to-event endpoints, where the two event times are assumed to have a positive probability of being identical. This method measures agreement in terms of the two event times being identical at a given time or both being greater than a given time. Overall scores of agreement over a period of time are also proposed.

1.5 Agreement Assessment in Bivariate Time-to-Event Times: Background

1.5.1 Bivariate Dependence Measures

Correlation coefficient

A traditional way of evaluating dependence in a bivariate distribution is by means of the correlation coefficient (Pearson correlation), defined as

$$\rho(T_1, T_2) = \frac{\text{Cov}(T_1, T_2)}{[\text{Var}(T_1)\text{Var}(T_2)]^{1/2}}.$$

The correlation is undefined when the variances in the denominator are 0 or infinite. An alternative expression based on the bivariate survivor function, $S(t_1, t_2) = Pr(T_1 > t_1, T_2 > t_2)$ is $\text{Cov}(T_1, T_2) = \int_0^\infty \int_0^\infty [S(t_1, t_2) - S_1(t_1)S_2(t_2)]\, dt_1 dt_2$, where $S_1(t_1) = S(t_1, 0)$ and $S_2(t_2) = S(0, t_2)$ are the marginal survival functions. Similarly, in the denominator the variance can be expressed as $\text{Var}(T_1) = 2\int_0^\infty \int_0^\infty tS_1(t)\, dt -$

$\left[\int_0^\infty S_1(t)\,dt\right]^2$. The properties of the measure $\rho(T_1, T_2)$ are that the range is $[-1, 1]$, with the values ± 1 if and only if T_1 and T_2 depend linearly each other, i.e. $T_2 = a + bT_1$, in which case $\rho = 1$ for $b > 0$ and -1 for $b < 0$. When T_1 and T_2 are independent, then $\rho = 0$. In general, the reverse statement is not true. However, a value of 0 for ρ implies that there is no linear correlation between the variables. More generally, note that $(T_{i1} - \bar{T}_1)(T_{i2} - \bar{T}_2)$ is positive if and only if T_{i1} and T_{i2} lie on the same side of their respective means. Thus the correlation coefficient is positive if T_{i1} and T_{i2} tend to be simultaneously greater than, or simultaneously less than, their respective means. The correlation coefficient is negative if T_{i1} and T_{i2} tend to lie on opposite sides of their respective means. The measure ρ is symmetric and invariant under linear transformations, that is, $\rho(T_1, T_2) = \rho(T_2, T_1)$ and $\rho(T_1, a + bT_2) = \rho(T_1, T_2)$ for any a, b, with $b > 0$. In addition, it is L_2-continuous, in the sense that $\rho(T_{n,1}, T_{n,2}) \to \rho(T_2, T_1)$, if $T_{n,1} \xrightarrow{L_2} T_1$ and $T_{n,2} \xrightarrow{L_2} T_2$, as $n \to \infty$. For normally distributed random variable, the correlation coefficient is intimately related to the conditional distributions. The conditional mean and variance are

$$\begin{aligned} E[T_2|T_1] &= E[T_2] - \rho\{\mathrm{Var}(T_2)/\mathrm{Var}(T_1)\}^{1/2}\{T - 1 - E[T_1]\}, \\ \mathrm{Var}(T_2|T_1) &= 1 - \rho^2\mathrm{Var}(T_2). \end{aligned}$$

The correlation is very well suited for measuring the linear dependence in the bivariate normal distribution.

The estimation of ρ with complete data is straightforward, just substitute the empirical mean and variances into the defining equation, i.e.

$$\hat{\rho} = \frac{\sum_{i=1}^n (T_{i1} - \bar{T}_1)(T_{i2} - \bar{T}_2)}{\{[\sum_{i=1}^n (T_{i1} - \bar{T}_1)^2][\sum_{i=1}^n (T_{i2} - \bar{T}_2)^2]\}^{1/2}}$$

The properties of this estimate are well known, when the distribution is bivariate normal, but it is not known too much for general distributions.

Kendall's Coefficient of Concordance

Another popular dependence measure is Kendall's coefficient concordance which is defined as a rank-based correlation type measure. It was suggested first by Fechner in 1897 and later rediscovered by Kendall in 1938, who examined it more completely. It is a simple measure of concordance, which does not require the assumption of normality. For a set of n independent observed values (T_{i1}, T_{i2}), $i = 1, \ldots, n$, of a bivariate variable (T_1, T_2), τ is defined as

$$\tau = E\{\text{sign}[(T_{11} - T_{21})(T_{12} - T_{22})]\},$$

where sign(x) is the sign of x, -1 for $x < 0$, 0 for $x = 0$, and 1 for $x > 0$. A more transparent formulation for continuous failure times is $\tau = 2p - 1$, where p is the probability that in two pairs the order of the first coordinates is the same as the order of the second coordinates, i.e. $p = Pr[(T_{11} - T_{21})(T_{12} - T_{22}) > 0]$. Kendall's τ has the same nice properties as the correlation coefficient and in addition τ is unchanged by both linear and nonlinear monotonic transformations. If the agreement between the two rankings is perfect (i.e. the two rankings are the same), then the coefficient τ has value 1. If the disagreement between the two rankings is perfect (i.e. one ranking is the reverse of the other), then the coefficient τ has value -1. If T_1 and T_2 are independent, then $\tau = 0$.

Alternatively, τ can be evaluated by integration of the bivariate survivor function,

$$\tau = 4 \iint f(t_1, t_2) S(t_1, t_2)\, dt_1 dt_2 - 1$$

For bivariate normal distribution, $\tau = 2\sin^{-1}(\rho)/\pi$, where ρ is the correlation coefficient.

Kendall's concordance coefficient can be estimated non-parametrically for complete data, by considering each combination of two pairs, scoring each concordant pair as 1, each discordant pair as -1, and each tie as 0. This is then normalized in a similar way as a standard deviation. In the case of complete data, for each pair i and i', let set $a_{ii'} = sign(T_{i1} - T_{i'1})$ and $b_{ii'} = sign(T_{i2} - T_{i'2})$, $i, i' = 1, \ldots, n$. the sores for the first coordinate and second coordinate, respectively. Then, the estimation formula is

$$\hat{\tau} = \frac{1}{n(n-1)} \sum_{i,i'=1;i\neq i'}^{n} a_{ii'} b_{ii'},$$

for data without ties, and

$$\hat{\tau} = \frac{\sum_{i,i'=1;i\neq i'}^{n} a_{ii'} b_{ii'}}{[(\sum_{i,i'=1;i\neq i'}^{n} a_{ii'}^2)(\sum_{i,i'=1;i\neq i'}^{n} b_{ii'}^2)]^{1/2}}.$$

for data with ties.

In the case of ties or censored data, the latter formula can be used with the scores $a_{ii'}$ and $b_{ii'}$ modified to account for censoring. A first suggestion is to use a score (a, b) of 0 in all cases, when the failure is not certain, due to censoring. This gives so-called simple estimate. This will typically underestimate the dependence, and it does not account for the fact that as censoring happens at different times, there is a higher probability of early death for the subject censored first. In 1974, Brown et al. (1974), proposed the use of a score based on the marginal Kaplan-Meier estimate. This is called adjusted

estimate. The values for the a scores are

$$\begin{aligned} a_{ij} &= \delta_{i1}\delta_{i'1}\text{sign}(T_{i1}-T_{i'1}) + (1-\delta_{i1})\delta_{i'1}\{2[\hat{S}_1(T_{i'1})/\hat{S}_1(T_{i1})]^{I(T_{i1}<T_{i'1})}-1\} \\ &\quad +\delta_{i1}(1-\delta_{i'1})\{1-2[\hat{S}_1(T_{i1})/\hat{S}_1(T_{i'1})]^{I(T_{i1}>T_{i'1})}\} \\ &\quad +(1-\delta_{i1})(1-\delta_{i'1})\{[\hat{S}_1(T_{i'1})/\hat{S}_1(T_{i1})]^{I(T_{i1}<T_{i'1})}-[\hat{S}_1(T_{i1})/\hat{S}_1(T_{i'1})]^{I(T_{i1}>T_{i'1})}\} \end{aligned}$$

The values for the b score are similar, just using the second coordinate for T and δ.

Spearman's Correlation Coefficient

An alternative measure is Spearman's correlation coefficient, denoted ρ_S, which was suggested by Spearman in 1904. It is a nonparametric measure, which is independent of marginal transformations and it is more like an ordinary correlation, in the sense that it accounts for the values, and not just the order of the observations. It is defined for arbitrary continuous marginal distributions by the formula

$$\rho_S = 12\int_0^1\int_0^1 S(S_1^{-1}(u), S_2^{-1}(v))\,dudv - 3. \tag{1.2}$$

When the marginals are uniform on $[0,1]$ the definition becomes

$$\rho_S = 12\int_0^1\int_0^1 S(S_1(u), S_2(v))\,dudv - 3 = 12\int_0^1\int_0^1 uvf(u,v)\,dudv - 3,$$

which corresponds to the correlation coefficient. These expression are not simple to integrate in most of the models, but they are simple to handle by numerical integration. Equation (1.2) does not work for marginal distributions that are not continuous as the inverse marginal survival function may be not well-defined. This is a particular issue in survival data, where it is not certain that the event will happen, because there is a mass point at infinity, and ρ_S cannot be evaluated.

ρ_S assesses how well the relationship between two variables can be described using a monotonic function. If there are no repeated data values, a perfect Spearman correlation of $+1$ or âĹŠ1 occurs when each of the variables is a perfect monotone function of the other. It can be be estimated for complete data, by considering the marginal ranks, (R_{i1}, R_{i2}), as follows.

$$\hat{\rho}_S = \frac{\sum_{i=1}^n [R_{i1} - (n+1)/2][R_{i2} - (n+1)/2]}{n(n^2-1)/12} = 1 - \frac{6\sum_{i=1}^n d_{ij}^2}{n(n^2-1)},$$

where $d_{ij} = R_{i1} - R_{i2}$, is the difference between ranks. Here, the two coordinates T_1 and T_2 are ordered separately, that is, R_{i1} is the rank of T_{i1} among $T_{11}, \ldots, T_{n1}$, and similarly for R_{i2}. In fact, this empirical formula was the original suggestion of Spearman, in 1904. For the bivariate normal distribution, it can be calculated from the correlation coefficient by $\rho_S = 6\sin^{-1}(\rho/2)/\pi$. Spearman's ρ_S can be approximated in terms of the Kendall's τ via $\rho_S = 3\tau/2$, which is valid for most distributions.

Cross Ratio

In familial examples, researches tend to believe that genetic influences may exist only in early ages. The global measures, such as Kendall's tau, is not ideal for addressing the concepts of early/late dependence. To address the question of local dependence, we need measures which evaluate dependence at a single time point, such as the cross ratio. For continuous (T_1, T_2), define the bivariate hazard function $\lambda(t_1, t_2) = f(t_1, t_2)/S(t_1, t_2)$. The cross ratio at (t_1, t_2) is defined as

$$\begin{aligned} \theta(t_1, t_2) &= \frac{\lambda(T_1 = t_1 | T_2 = t_2)}{\lambda(T_1 = t_1 | T_2 > t_2)} = \frac{f(t_1, t_2)}{\partial_{s_2} S(t_1, s_2)|_{s_2 = t_2}} \div \frac{\partial_{s_1} S(s_1, t_2)|_{s_1 = t_1}}{S(t_1, t_2)} \\ &= \frac{S(t_1, t_2) f(t_1, t_2)}{\partial_{s_1} S(s_1, t_2)|_{s_1 = t_1} \partial_{s_2} S(t_1, s_2)|_{s_2 = t_2}} \end{aligned}$$

The cross ratio $\theta(t_1, t_2)$ is interpreted as the ratio of one's failure risk at time t_1 if his/her partner is known to have failed versus survived at time t_2. The cross ratio measures the degree of dependence between T_1 and T_2, where independence is implied by $\theta(t_1, t_2) = 1$. When two failure times are exchangeable, such as the failure times from (identical) twins, the cross ratio is symmetric with respect to the two components; that is, the cross ratio for (T_1, T_2) is the same as the cross ratio for (T_2, T_1)

1.5.2 Copula Model for Bivariate Failure Times

One of the earliest family of distributions for correlated bivariate measurements is the Copula family, in which the marginal distributions are uniform on the unit interval. The Copula family includes many popular bivariate failure time models and has gained considerable attention in statistical literature because of its flexibility in building stochastic models. A copula is used as a general way of formulating a multivariate distribution in such a way that various general types of dependence can be represented. The approach to formulating a multivariate distribution using a copula is based on the idea that a simple transformation can be made of each marginal variable in such a way that each transformed marginal variable has a uniform distribution. Once this is done, the dependence structure can be expressed as a multivariate distribution on the obtained uniforms, and a copula is precisely a multivariate distribution on marginally uniform random variables.

Suppose that $C(u_1, u_2)$ is a joint cumulative distribution function with density $c(u_1, u_2)$ on $[0, 1] \times [0, 1]$, that is, $C : [0, 1][0, 1][0, 1]$, and $C(0, u_2) = C(u_1, 0) = 0$, $C(u, 1) = C(1, u) = u$. Let (T_1, T_2) denote the paired failure times, (S_1, S_2) and (f_1, f_2) denote the corresponding marginal survival and density functions. Then the joint survival function of (T_1, T_2) in the Copula family is given by $S(t_1, t_2) = C(S_1(t_1), S_2(t_2))$ and its density by $f(t_1, t_2) = c(S_1(t_1), S_2(t_2))f_1(t_1)f_2(t_2)$. The copula C contains all

information on the dependence structure between T_1 and T_2, whereas the marginal survival functions $S_i(t_i)$ contain all information on marginal distributions.

Archimedian Copula model.

The survival function in this subclass has the following form

$$S(t_1, t_2) = \phi[\phi^{-1}(S_1(t_1)) + \phi^1(S_2(t_2))],$$

where $0 \leq \phi \leq 1$, $\phi(0) = 1$, $\phi' < 0$, $phi > 0$ (a convex decreasing function). If ϕ is a Laplace transform of some distribution (of W), $\phi(t) = E(e^{tW})$, the Archimedian copula model reduces to the proportional frailty model.

Gaussian Copula Model.

For a given correlation matrix $\Sigma \in R^{2\times 2}$, the Gaussian copula with parameter matrix Σ can be written as

$$C(u_1, u_2) = \Phi_\Sigma^{-1}(\Phi^{-1}(u_1), \Phi^{-1}(u_2)),$$

where Φ^{-1} is the inverse cumulative distribution function of the univariate standard normal distribution and Φ_Σ is the joint cumulative function of bivariate normal distribution with mean zero and covariance matrix equal to the correlation matrix Σ.

The Copula models can be formulated by the marginal distributions and Copula. This two-step approach of modeling is convenient because many tractable models are readily available for the marginal distributions. Also, the Copula models make sense for illustrating dependence. Other Copula models include Clayton's Family, Frank's Family, Positive stable copula, etc. (Clayton, 1978; Hougaard, 1986; Frank,1979).

1.5.3 Frailty Models for Multivariate Failure Times

A commonly used approach to model multivariate failure times, the frailty model, is to specify independence among multivariate failure times conditional on an unobserved positive-valued variable, W, called *frailty*. Assume that the hazard function of T_{ij} given $W_i = w$ (frailty) is $\lambda_j(t_j|W_i = w) = w\lambda_{0j}(t_j)$, which is a proportional frailty model with the baseline hazard function $\lambda_{0j}()$. Let $B_j(\cdot)$ be the corresponding survival function for $\lambda_{0j}(\cdot)$.

Univariate Inference

The survival function of T_{ij} given by $W_i = w$ is $S(t_j|W_i = w) = B_j(t_j)^w$ and the multivariate survival function of $(T_{i1}, \ldots, T_{im})$ given $W_i = w$ by $S(t_1, \ldots, t_m|W_i = w) = \prod_{j=1}^{m} B_j(t_j)^w$. Thus, the unconditional survival function of T_{ij} is $S_j(t_j) = \phi(\log B_j(t_j))$, where $\phi(\cdot)$ is the Laplace transform of the random variable W_i, i.e., $\phi(t) = E(e^{tW_i})$.

By extending the proportional hazards model, a more general setting of the proportional frailty model can be expressed as $\lambda_j(t_j; x_{ij}, w_i) = w_i\lambda_{0j}(t_j)\exp(\beta x_{ij})$, for $j = 1, \ldots, m$.

Bivariate Inference

The bivariate survival function satisfies $S(t_1, t_2) = \int [B_1(t_1)B_2(t_2)]^w dF_W(w)$ where F_W denotes the the frailty cumulative distribution function of W . It follows that $S(t_1, t_2) = \phi(\log B_1(t_1) \log B_2(t_2))$, where $\phi(\cdot)$ is the Laplace transform of the random variable W.

Bivariate distributions generated by frailty models are seen to be a subclass of the Archimedean distributions (Genest and MacKey, 1989, American Statistician). With $B_j(t_j) = \exp[\phi^1(S_j(t_j)]$ and $\phi(\cdot)$ a Laplace transform, the bivariate distribution can be

written as

$$S(t_1, t_2) = \int \prod_{i=1}^{2} \exp[w\phi^1(S_j(t_j))] dF_W(w) = \phi[\phi^{-1}(S_1(t_1) + \phi^{-1}(S_2(t_2))].$$

Gamma frailty models (Clayton model). Assume that the frailty W follows a Gamma distribution with mean 1 and variance $\alpha > 0$. The corresponding Laplace transform is $\phi(u) = (1 + u)^1$. The failure times (T_1, T_2) are positively correlated when $\alpha > 0$ and independent when $\alpha = 0$. The joint survival function can be written as $S(t_1, t_2) = [S_1(t_1)^\alpha + S_2(t_2)^\alpha 1]^{1/\alpha}$.

Stable frailty models. Hougaard (1986) proposed a class of multivariate model, where the frailty W follows the positive stable distribution with parameter α so that the Laplace transform is $\phi(u) = \exp(ua)$, $0 < a < 1$. The corresponding joint survival function is $S(t_1, t_2) = \exp([(\log S_1(t_1))^{1/a} + (\log S_2(t_2))^{1/a}]^a)$. A notable property of the stable frailty model is that if the conditional hazards are proportional, then the hazard in the marginal distributions are also proportional, but with different baseline hazards and regression coefficients.

1.5.4 Agreement Measures for Time-to-Event Times

For time-to-event data, rater concordance has been calculated with two separate analyses. Cohen's kappa (Cohen 1960) was first calculated for ratings of event occurrence. Then, among subjects where raters agreed that the event occurred, the intraclass correlation coefficient (ICC) (Lord and Novick 1968) was estimated for the observed times to an event. This method does not properly evaluate concordance when data are censored because the ICC estimate is biased upward when only using the time data for consistently reported events. In addition, because it is less likely for a rater to observe event occurrence in a short follow-up than in a longer follow-up, between-subject variation in length of follow up may impact on the kappa estimate for agreement in reported

event occurrence. Finally, when reported rates are very large or small, kappa could be low due to the well-known impact of extreme marginal probabilities on agreement measures, see Feinstein and Cicchetti (1990).

Cohen (1960) introduced the kappa coefficient of agreement as simply the proportion of chance-expected disagreements which do not occur, or alternatively, it is the proportion agreement after chance agreement is removed from consideration:

$$\kappa = \frac{p_o - p_c}{1 - p_c},$$

where p_o is the probability of observed agreement among raters and p_c is the hypothetical probability of chance agreement. Or, equivalently, in terms of frequencies, $\kappa = \frac{f_o - f_c}{1 - f_c}$. If the raters are in complete agreement then $\kappa = 1$. if there is no agreement among the rates other than what would be expected by chance, then $\kappa = 0$. Later, in (Cohen 1968), he further developed κ to a *weighted* kappa. This was motivated by studies in which it is the sense of the investigator that some disagreements in assignments by two raters, are of greater gravity than others. Cohen's kappa and weighted kappa are the most commonly used measures of the concordance of qualitative and ordinal ratings between two rates, adjusting for chance of agreements. For extensions and generalization of kappa we refer to Kraemer et al. (2002) and J. M. Williamson et al. (2000).

In 1989, Lin first proposed the concordance correlation coefficient (CCC) to evaluate reliability of quantitative ratings between two raters (see Lin (1989)), and then in 2000 he generalized it to measure overall agreement among multiple raters (see Lin (2000)). Later, Barnhart et al. (2002) defined the concordance correlation coefficient as follows. Let random variable T_j be the rating from the j-th rater with mean μ_j and variance σ_j^2, $j = 1, \ldots, m$, and $\sigma_{jk} = \rho_{jk}\sigma_j\sigma_k$ be the covariance of the ratings T_j and T_k for $j \neq k$.

Then CCC is defined as

$$CCC(m) = \frac{\sum_{j=1}^{m}\sum_{k=1,k\neq j}^{m}\sigma_{ij}}{(m-1)\sum_{j=1}^{m}\sigma_j^2 + m\sum_{j=1}^{m}(\mu_j-\bar{\mu})^2}$$

equivalent to Lin's generalized CCC. Its possible values are ranging between $-1/(m-1)$ and 1. For the bivariate case, $m = 2$, $CCC(2) = \frac{2\sigma_{12}}{\sigma_1^2+\sigma_2^2+(\mu_1-\mu_2)^2}$ and the values are ranging between -1 and 1. When dealing with with data subject to censoring, the likelihood-based approach to estimate CCC is attractive for the advantages that censoring can be easily accommodated and the estimates have good properties. As a function of the first two moments of rating measures, the CCC can be estimated for censored data, using likelihood-based estimation method under the assumptions of random censoring and parametric distribution models for the ratings of time to event, see Liu et al. (2005).

Guo et al. (2013) proposes a framework for assessing agreement based on survival processes that can be viewed as a natural representation of time-to-event outcomes. Their agreement measure is formulated as the chance-corrected concordance between survival processes. The key idea is to represent survival outcomes through survival processes as, $U_j(t) = I(T_j \geq t)$, $j = 1, 2$, where T_1 and T_2 are the survival times of the same subject based on different methods and I is the indicator function. The agreement measure is based on the concordance between the survival processes over a finite range of $[0, a]$, and it is defined by

$$\rho_{sp}(a) = 1 - \frac{E\left\{\int_0^a [U_1(t)-U_2(t)]^2\,dt\right\}}{E\left\{\int_0^a [U_1(t)-U_2(t)]^2\,dt \,|\, U_1, U_2 \text{ independent}\right\}}.$$

The measure $\rho_{sp}(a)$ is well defined with nonzero denominator as long as $S_j(a) < 1$ for $j = 1$ or 2, where $S_j(\cdot)$ is the marginal survival function of T_j. It provides a novel perspective for studying the relationship between correlated survival outcomes

and offers an appealing interpretation as the agreement between survival times on the absolute distance scale. The upper bound of $\rho_{sp}(a)$ is 1 which is achieved when there is perfect agreement between the two survival processes within the specified range, i.e. $\Pr(U_1(t) = U_2(t)) = 1$ for all $t \in [0, a]$. When the two survival processes are statistically independent, $\rho_{sp}(a)$ is 0 representing no agreement beyond that expected by chance. When the discrepancy between the two survival processes is even larger than what is expected by chance, $\rho_{sp}(a)$ is negative. The lower bound of $\rho_{sp}(a)$ is achieved under the following conditions: (i) $S(t,t)F(t,t) = 0$ for all $t \in [0, a]$; and (ii) $S_1(t) = S_2(t)$ or all $t \in [0, a]$. Under these conditions, $\rho_{sp}(a)$ reaches its lower bound which is a function of the common marginal survival functions and the specified range $[0, a]$. Nonparametric estimates of $\rho_{sp}(a)$ are derived, and it is shown that the estimators are strongly consistent and asymptotically normal.

CHAPTER 2: REGRESSION MODELS ON RIEMANNIAN SYMMETRIC SPACES

2.1 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 Figure 2.1 for four different examples of manifold-valued data. Regression analysis is a fundamental statistical tool for relating a response variable to a set of covariates, such as age. In particular, when both the response and the covariate are in Euclidean space, the classical linear regression model and its variants have been widely used in various fields (McCullagh and A.Nelder 1989, Fahrmeir and Tutz 2001). However, when the response is in RSS and the covariate is in Euclidean space, developing regression models for this type of data raises both computational and theoretical challenges. The aim of our research is to develop a general regression framework to address these challenges.

Certain types of manifolds are encountered more frequently in practice. As an illustration, we discuss three of them and their applications as follows.

1. **Unit Sphere and Quotient Spaces of Spheres**: Directional data on the unit sphere in R^k, denoted by $S^{k-1} = \{x \in R^k : ||x|| = 1\}$, are routinely encountered in a wide variety of disciplines, where $||\cdot||$ denotes the L_2 norm. For example, 3-dimensional directions arise in analyzing the directions of moving objects (Mardia and Jupp 2000, Healy and Kim 1996, Kim 1998). In the landmark-based shape analysis of objects, 2-dimensional (2D) objects are represented by configurations

of salient points and after removing the translation and scale, the space of all such configurations with k landmarks is S^{2k-3} (Dryden and Mardia 1998). In some problems, the spaces of interest are quotient spaces of spheres rather than the spheres themselves. For example, in the shape analysis of k landmarks, after removing the rotation variability, the space of all 2D-configurations becomes the complex projective space $CP^{k-2} = S^{2k-3}/S^1$ (Kendall 1984a, Kendall et al. 1999, Dryden and Mardia 1998, Huckemann et al. 2010).

2. **Matrix Lie Groups**: The transformation group, denoted by $\mathrm{GL}(k)$, is the set of $k \times k$ invertible matrices, and its subgroups are important in many situations involving the actions of these groups on objects in Euclidean space. For example, a transformation of the type $x \to Ax + \mathbf{b}$ with $\mathbf{b} \in R^k$ (i) is called an *affine* transformation when $A \in \mathrm{GL}(k)$, (ii) is called a *volume-preserving* transformation when $A \in \mathrm{SL}(k) = \{A : \det(A) = 1\}$, and (iii) is called a *similarity* transformation when $A \in \mathrm{SO}(k) = \{A : A^T A = AA^T = I_k\}$, in which I_k is a $k \times k$ identity matrix. In the problem of tracking and recognizing objects in video data, their poses to the camera are important. The pose of a rigid object is conveniently represented as an element of SO(2) (or SO(3)) for planar objects (3-dimensional (3D) objects) (Grenander et al. 1998, Moakher 2002).

3. **Quotient Spaces of Matrix Lie Groups (RSS)**: Any RSS can be regarded as a quotient space of Lie groups. The most prominent examples in this category are the Grassmann and Stiefel manifolds that are used in orthogonal transformations. The S^k can be viewed as a quotient space of $\mathrm{SO}(k+1)$. In many applications, the statistical analysis of symmetric positive-definite matrices is essential. The manifold of $k \times k$ symmetric positive definite matrices $\mathrm{Sym}^+(k)$ can be viewed as a quotient space of $\mathrm{GL}(k)$, namely it can be identified with $\mathrm{GL}(k)/O(k)$, where $O(k)$ is the matrix group of orthogonal transformations of R^k. The $\mathrm{Sym}^+(k)$

manifold, which can also be regarded as a Lie group, occurs in a wide variety of important applications including diffusion tensor imaging, functional and structural connectivity, and computational anatomy, among others (Fletcher and Joshi 2007, Zhu et al. 2009a, Grenander and Miller 1998, Dryden et al. 2009).

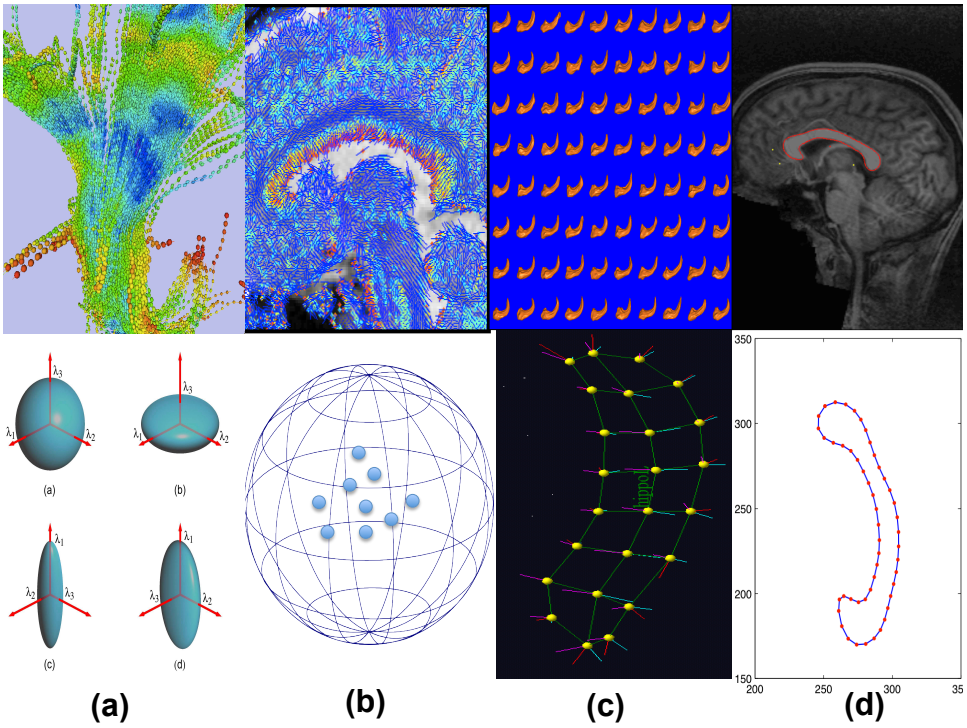

Figure 2.1: Examples of manifold-valued data: (a) diffusion tensors along white matter fiber bundles and their ellipsoid representations; (b) principal direction map of a selected slice and their directional representations on $S^2(1)$; (c) median representations and median atoms; and (d) automatic corpus callosum segmentation and its contour and landmarks of a selected subject.

Little has been done on the regression analyses of manifold-valued 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. 2004b, Dryden and Mardia 1998, Huckemann et al. 2010, Younes 2010, Osborne et al. 2013). In contrast, even for the 'simplest'

directional data, there is a sparse literature on regression modeling of a single directional response and a set of covariates (Mardia and Jupp 2000, Jupp and Mardia 1989). In addition, these regression models of directional data are primarily based on a specific parametric distribution, such as the von Mises-Fisher distribution (Mardia 1975, Mardia and Jupp 2000, Kent 1982, Presnell et al. 1998). 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. Recently, Zhu and his coauthors have proposed semiparametric intrinsic regression models for manifold-valued response data lying in $\mathrm{Sym}^+(k)$ and S^2 (Shi et al. 2012; 2009, Zhu et al. 2009a).

In case the covariate is scalar and continuous (e.g. time), the problem is similar to the that of fitting smooth trajectories to time-indexed point on a manifold. There is a quite large amount of work in applied mathematics community on fitting splines to observed points on a Riemannian manifold (Samir et al. 2012, Su et al. 2012, Muralidharan and Fletcher 2012, Machado and Leite 2006, Machado et al. 2010). Recently, Samir et. al. developed an optimization framework for searching for smoothing splines on manifolds based on a metric-based steepest-descend gradient algorithm. Su and his coauthors apply Samir's methods to some general manifolds relevant to problems in computer vision, like $SO(3)$, 3×3 SPDs space, and Kendall's landmark-based shape space (Su et al. 2012). In contrast, there is little work done in the case of multidimensional covariates.

While aside of the situation considered here, but still related, we note there is some literature on regression where both the responses and covariates lie on a manifold, mainly for for the case of $3D$-spherical data (Chang 1986; 1989, Downs 2003, Rosenthal et al. 2014).

An intriguing question is whether there is a general regression framework for manifold-valued response in a RSS and covariates in a multidimensional Euclidean space. The aim of this research work 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.

- We propose an intrinsic regression model solely based on an intrinsic conditional moment for the response in a RSS, thus avoiding specifying any parametric distributions in a general RSS. The model handles more than one Euclidean covariate.
- We develop a procedure for estimating the regression coefficients in this intrinsic model.
- We develop several test statistics for testing linear hypotheses of the regression coefficients.
- We develop a general asymptotic framework for the estimates of the regression coefficients and test statistics.
- We systematically investigate the geometrical properties (e.g., chart invariance) of these parameter estimates and test statistics.

The presentation of this work is organized as follows. In Section 2.2, we review the basic notion and concepts of Riemannian geometry. In Section 2.3, we propose the intrinsic regression models and develop estimation and inference procedures. In Section 2.4, we examine several specific RSS's and propose various link functions to map from the Euclidean space of covariates to the RSS of responses. In Section 2.5, we examine a set of simulation studies with the known ground truth to examine the finite sample performance of the test statistics. In Section 2.6, we introduced the ADNI dataset and apply the estimation and inference procedures to investigate the association between

the shape of Corpus Callosum (CC) contours and covariates including gender, age, and diagnosis, in a Alzheimer's disease study from the ADNI data. a attention deficit hyperactivity disorder study. Finally, we conclude with a discussion in Section 2.7. Technical conditions and proofs are deferred to the Appendix.

2.2 Differential Geometry - An Introduction

We briefly review some basic facts about the theory of Riemannian geometry and present more technical details in the Appendix. The reader can refer to (Spivak 1979, Lang 1999, Boothby 1986, do Carmo 1992, Pennec 2006) for more details.

A Riemannian manifold $(\mathcal{M}, \mathrm{m})$ is a smooth manifold $\mathcal{M} \subset R^{d_{\mathcal{M}}}$ together with an inner product m, where $d_{\mathcal{M}}$ is the dimension of $\mathcal{M}$. We first introduce the tangent vector and tangent space at $\mathrm{p} \in \mathcal{M}$. For a small scalar $\delta > 0$, let $\gamma(t)$ be a differentiable map from $(-\delta, \delta)$ to $\mathcal{M}$ passing through $\gamma(0) = \mathrm{p}$. A tangent vector at p is defined as the derivative of the smooth curve $\gamma(t)$ with respect to t evaluated at $t = 0$. The set of all tangent vectors at p forms the tangent space of $\mathcal{M}$ at p, denoted as $T_{\mathrm{p}}\mathcal{M}$. The $T_{\mathrm{p}}\mathcal{M}$ is equipped with an inner product $\mathrm{m_p}$, called a Riemannian metric, which varies smoothly from point to point. If $\mathcal{M}$ is complete, the *exponential map* at p is defined on the tangent space $T_{\mathrm{p}}\mathcal{M}$ by $\mathrm{Exp}_{\mathrm{p}}^{\mathcal{M}}(V) = \gamma(1; \mathrm{p}, V)$, where $\gamma(1; \mathrm{p}, V)$ is the geodesic with $\gamma(0; \mathrm{p}, V) = \mathrm{p}$ and $\gamma'(0; \mathrm{p}, V) = V$. An open subset $\mathcal{U}$ of $\mathcal{M}$ containing p is a normal chart near p if $\mathrm{Exp}_{\mathrm{p}}^{\mathcal{M}}$ is a diffeomorphism on an open neighborhood $\mathcal{V}$ of the origin in $T_{\mathrm{p}}\mathcal{M}$ onto $\mathcal{U}$ with $\mathcal{V}$ such that $tV \in \mathcal{V}$ for $0 \leq t \leq 1$ and $V \in \mathcal{V}$. The inverse map is the *logarithmic map* at p, denoted by $\mathrm{Log}_{\mathrm{p}}^{\mathcal{M}}$. Then, for $\mathrm{q} \in \mathcal{U}$, $\mathrm{dist}_{\mathcal{M}}(\mathrm{p}, \mathrm{q}) = \|\mathrm{Log}_{\mathrm{p}}^{\mathcal{M}}(\mathrm{q})\|_{\mathrm{p}}$. The *radius of injectivity* of $\mathcal{M}$ at p, denoted by $\rho^*(\mathcal{M}, \mathrm{p})$, is the largest $r > 0$ such that $\mathrm{Exp}_{\mathrm{p}}^{\mathcal{M}}$ is a diffeomorphism on the open ball $B(\mathbf{0}, r) \subset T_{\mathrm{p}}\mathcal{M}$ onto an open set in $\mathcal{M}$ near p. Any basis in the tangent space $T_{\mathrm{p}}\mathcal{M}$ induces an isomorphism from $T_{\mathrm{p}}\mathcal{M}$ to $R^{d_{\mathcal{M}}}$,

and then the logarithmic map $\mathrm{Log}_{\mathrm{p}}$ provides a local chart near p. If $T_{\mathrm{p}}\mathcal{M}$ is endowed with an orthonormal basis, such a chart is called a *normal chart* and the coordinates are called *normal coordinates*.

A Lie group G is a group together with a smooth manifold structure such that the operations of multiplication $(a, b) \mapsto ab$ and inversion $a \mapsto a^{-1}$ are smooth maps. Throughout this work, we only consider finite dimensional Lie groups. The *exponential map* of G at its identity element, denoted by e, is defined as $\mathrm{Exp}^G(\mathbf{v}) = \gamma^G(1; \mathbf{v})$ for any $\mathbf{v} \in T_eG$, where $\gamma^G(\cdot; \mathbf{v}) : R \to G$ is the unique one-parameter subgroup of G whose tangent vector at e is equal to $\mathbf{v}$. For $a \in G$, the *exponential map* of G at a is defined by $\mathrm{Exp}^G_a \circ L_{a*} = L_a \circ \mathrm{Exp}^G_e$, where L_a is the left multiplication by a and L_{a*} is its tangent map. Many common geometric transformations of Euclidean spaces including rotations, translations, dilations, and affine transformations on R^d form Lie groups. In general, Lie groups can be used to describe transformations of smooth manifolds.

A 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 RSS's 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. Given a smooth manifold $\mathcal{M}$ and a Lie group G, a *smooth group action of G on $\mathcal{M}$* is a smooth mapping $G \times \mathcal{M} \to \mathcal{M}$, $(a, \mathrm{p}) \mapsto a \cdot \mathrm{p}$ such that $e \cdot \mathrm{p} = \mathrm{p}$ and $(aa') \cdot \mathrm{p} = a \cdot (a' \cdot \mathrm{p})$ for all $a, a' \in G$ and all $\mathrm{p} \in \mathcal{M}$. The group action should be interpreted as a group of transformations of the manifold $\mathcal{M}$, namely, $\{L_a\}_{a \in G}$ such that $L_a : \mathcal{M} \to \mathcal{M}$, $L_a(\mathrm{p}) = a \cdot \mathrm{p}$ for $\mathrm{p} \in \mathcal{M}$ and $a \in G$. The L_a is a smooth transformation on $\mathcal{M}$ and its inverse is denoted by $L_{a^{-1}}$. Given $\mathrm{p} \in \mathcal{M}$, let ι_{p} denote the action of G on the point $\mathrm{p} \in \mathcal{M}$ such that $\iota_{\mathrm{p}} : G \to \mathcal{M}$, $\iota_{\mathrm{p}}(a) = a \cdot \mathrm{p} = L_a(\mathrm{p})$ for all $a \in G$. Thus, ι_{p} is a smooth map from G into $\mathcal{M}$. The *orbit* of a point $\mathrm{p} \in \mathcal{M}$ is defined as $G(\mathrm{p}) = \{a \cdot \mathrm{p} \,|\, a \in G\}$. The orbits form a partition

of $\mathcal{M}$. Specifically, two points $\mathrm{p}, \mathrm{p}' \in \mathcal{M}$ are equivalent if there exists $a \in G$ such that $a \cdot \mathrm{p} = \mathrm{p}'$. If $\mathcal{M}$ consists of a single orbit, the group action is *transitive* or G acts *transitively* on $\mathcal{M}$, and we call $\mathcal{M}$ as a *homogeneous* space. The *isotropy* subgroup of a point $\mathrm{p} \in \mathcal{M}$ is defined as $G_{\mathrm{p}} = \{a \in G \,|\, a \cdot \mathrm{p} = \mathrm{p}\}$.

When a Lie group G acts smoothly on a smooth manifold $\mathcal{M}$, for any $\mathrm{p} \in \mathcal{M}$, there is a natural bijection from the orbit $G(\mathrm{p})$ onto the quotient manifold given by the smooth mapping $a \cdot \mathrm{p} \mapsto aG_{\mathrm{p}}$ such that $G(\mathrm{p}) \cong G/G_{\mathrm{p}}$, where $\cong$ denotes the bijection. Let G be a connected group of isometries of the RSS $\mathcal{M}$ such that $\mathrm{dist}_{\mathcal{M}}(\mathrm{p}, \mathrm{p}') = \mathrm{dist}_{\mathcal{M}}(a \cdot \mathrm{p}, a \cdot \mathrm{p}')$, for all $\mathrm{p}, \mathrm{p}' \in \mathcal{M}$ and all $a \in G$. For any $\mathrm{p} \in \mathcal{M}$, the RSS $\mathcal{M}$ can always be viewed as a homogeneous space, $\mathcal{M} \cong G/G_{\mathrm{p}}$, and the isotropy subgroup G_{p} is compact.

From now on, we will assume that the manifold $\mathcal{M}$ is a RSS and $\mathcal{M} = G/G_{\mathrm{p}}$ 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 $\mathcal{M}$. Due 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. For any point $y \in \mathcal{M}$, geodesics starting from y are of the form $a \cdot \gamma(\cdot)$, where $\gamma(\cdot)$ is a geodesic starting from p, $\gamma(0) = \mathrm{p}$ and $y = a \cdot \mathrm{p}$ for some $a \in G$. Due to the local uniqueness of geodesics, if $y = a' \cdot \mathrm{p}$ for some other $a' \in G$, then $a \cdot \gamma(\cdot) = a' \cdot \gamma(\cdot)$. Geodesics on $\mathcal{M}$ starting from p are the images of the action of a 1-parameter subgroup of G acting on the base point p. In other words, for any geodesic γ on $\mathcal{M}$, $\gamma(\cdot) : R \to \mathcal{M}$, starting from p, there exists a 1-parameter subgroup $c(\cdot) : R \to G$ such that $\gamma(t) = c(t) \cdot \mathrm{p}$ for all $t \in R$.

2.3 Intrinsic Regression Model

Let $(\mathcal{M}, \mathrm{m})$ 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. Let $\mathrm{p} \in \mathcal{M}$ be a base point of $\mathcal{M}$ and $\rho = \rho^*_{\mathcal{M}}$ be the

radius of injectivity of $\mathcal{M}$.

2.3.1 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 $\mathbf{x}_i$ is a $d_{\mathbf{x}} \times 1$ vector of covariates. Our objective is to introduce an intrinsic regression model for RSS responses and covariates of interest from n subjects.

The specification of the intrinsic regression model involves three key steps including (i) a link function mapping from the space of covariates to $\mathcal{M}$, (ii) the definition of a residual, and (iii) the action of transporting all residuals to a common space. First, we explicitly formalize the link function. From now on, all covariates have been centered to have mean zero. We consider a single-center link function given by

$$\boldsymbol{\mu}(\mathbf{x}, \mathrm{q}, \boldsymbol{\beta}) : R^{d_{\mathbf{x}}} \times \mathcal{M} \times R^{d_{\boldsymbol{\beta}}} \to M, \tag{2.1}$$

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

$$\boldsymbol{\mu}(\mathbf{0}, \mathrm{q}, \boldsymbol{\beta}) = \boldsymbol{\mu}(\mathbf{x}, \mathrm{q}, \mathbf{0}) = \mathrm{q}. \tag{2.2}$$

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 center (or 'mean') $\mathrm{q} \in \mathcal{M}$. When all the covariates are equal to zero, the link function is independent of the regression coefficients and reduces to the center $\mathrm{q} \in \mathcal{M}$.

More generally, we may consider a multicenter link function to account for the presence of discrete covariates, such as gender and diagnostic group. Let $\mathbf{x}_i = (\mathbf{x}_{i,C}, \mathbf{x}_{i,D})$, where $\mathbf{x}_{i,D}$ and $\mathbf{x}_{i,C}$ are, respectively, a $d_{\mathbf{x},D} \times 1$ vector of all the discrete covariates and

a $d_{\mathbf{x},C} \times 1$ vector of all the continuous covariates and their potential interactions with $\mathbf{x}_{i,D}$. We may introduce a center for each covariate class based on $\mathbf{x}_{i,D}$ (McCullagh and A.Nelder 1989). In this case, we may define the multicenter link function as follows:

$$\boldsymbol{\mu}(\mathbf{x}, \mathrm{q}(\mathbf{x}_D), \boldsymbol{\beta}) : R^{d_{\mathbf{x}}} \times \mathcal{M}^{d_D} \times R^{d_{\boldsymbol{\beta}}} \to M, \tag{2.3}$$

where d_D is an integer associated with the number of covariate classes and $\boldsymbol{\beta}$ is primarily associated with continuous covariates. Moreover, it is assumed that $\boldsymbol{\mu}(\mathbf{x}, \mathrm{q}, \boldsymbol{\beta})$ satisfies a multicenter property as follows:

$$\boldsymbol{\mu}((\mathbf{0}, \mathbf{x}_D), \mathrm{q}(\mathbf{x}_D), \boldsymbol{\beta}) = \boldsymbol{\mu}(\mathbf{x}, \mathrm{q}(\mathbf{x}_D), \mathbf{0}) = \mathrm{q}(\mathbf{x}_D). \tag{2.4}$$

When the regression coefficients vector $\boldsymbol{\beta}$ equals $\mathbf{0}$, the link function is independent of continuous covariates and reduces to $\mathrm{q}(\mathbf{x}_D)$ in $\mathcal{M}$. When all continuous covariates are equal to zero, the link function is independent of the regression coefficients and reduces to the center $\mathrm{q}(\mathbf{x}_D)$ in $\mathcal{M}$. For notational simplicity, we focus on (2.1) from now on and as the extension to (2.3) is trivial.

Secondly, we introduce a definition of "residual" to ensure that $\boldsymbol{\mu}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta})$ is the proper "conditional mean" of y_i given $\mathbf{x}_i$, which is the key concept of many regression models (McCullagh and A.Nelder 1989, Fahrmeir and Tutz 2001). 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 zero. Given the points y_i and $\boldsymbol{\mu}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta})$ on the RSS $\mathcal{M}$, we need to define the residual as "a difference" between y_i and $\boldsymbol{\mu}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta})$. Assume that y_i and $\boldsymbol{\mu}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta})$ are "close enough" to each other in the sense that there is an open ball $B(0, \rho) \subset T_{\boldsymbol{\mu}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta})}\mathcal{M}$

such that for all $i = 1, \ldots, n$,

$$y_i \in \mathrm{Exp}_{\boldsymbol{\mu}(\mathbf{x}_i,\mathrm{q},\boldsymbol{\beta})}\left(B(0,\rho)\right) \text{ or } \mathrm{Log}_{\boldsymbol{\mu}(\mathbf{x}_i,\mathrm{q},\boldsymbol{\beta})}(y_i) \subset B(0,\rho).$$

Thus, $\mathrm{Log}_{\boldsymbol{\mu}(\mathbf{x}_i,\mathrm{q},\boldsymbol{\beta})}(y_i)$ may make it a good candidate to play the role of a 'residual'. These residuals, however, lie on different tangent spaces to $\mathcal{M}$, so it is difficult to carry out a multivariate analysis of these residuals.

Thirdly, since $\mathcal{M}$ is a RSS, this enables us to "transport" all the residuals, separately, to a common space, say $T_{\mathrm{p}}\mathcal{M}$, by exploiting 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 $\boldsymbol{\mu}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta})$ can be joined in $\mathcal{M}$ by a geodesic, denoted by $\gamma_i(t; \mathrm{q}, \boldsymbol{\beta}) = \gamma(t; \mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta})$, satisfying $\gamma_i(0; \mathrm{q}, \boldsymbol{\beta}) = \mathrm{p}$ and $\gamma_i(1; \mathrm{q}, \boldsymbol{\beta}) = \boldsymbol{\mu}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta})$. Moreover, γ can be seen as the action of a one-parameter subgroup of G and $\gamma_i(t; \mathrm{q}, \boldsymbol{\beta}) = c_i(t; \mathrm{q}, \boldsymbol{\beta}) \cdot \mathrm{p}$ for $t \in R$, where $c_i(t; \mathrm{q}, \boldsymbol{\beta}) = c(t; \mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta}) : R \to G$. Thus, since $L_{a*}^{-1} = L_{a^{-1}*}$ on $T\mathcal{M}$ for $a \in G$, we have

$$\begin{aligned} &L^{-1}_{c(1;\mathbf{x}_i,\mathrm{q},\boldsymbol{\beta})*}(\mathrm{Log}_{\boldsymbol{\mu}(\mathbf{x}_i,\mathrm{q},\boldsymbol{\beta})}(y_i)) = L_{c(1;\mathbf{x}_i,\mathrm{q},\boldsymbol{\beta})^{-1}*}(\mathrm{Log}_{\boldsymbol{\mu}(\mathbf{x}_i,\mathrm{q},\boldsymbol{\beta})}(y_i)) \\ &= \mathrm{Log}_{\mathrm{p}}\left(L_{c(1;\mathbf{x}_i,\mathrm{q},\boldsymbol{\beta})^{-1}}(y_i)\right) = \mathrm{Log}_{\mathrm{p}}\left(c(1;\mathbf{x}_i,\mathrm{q},\boldsymbol{\beta})^{-1} \cdot y_i\right) \in T_{\mathrm{p}}M. \end{aligned}$$

We define the *rotated residual* $\mathcal{E}(y_i, \mathbf{x}_i; \mathrm{q}, \boldsymbol{\beta})$ of $y_i \in \mathcal{M}$ with respect to $\boldsymbol{\mu}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta})$ as the parallel transport of the actual residual, $\mathrm{Log}_{\boldsymbol{\mu}(\mathbf{x}_i,\mathrm{q},\boldsymbol{\beta})}(y_i)$, along the geodesic from the conditional mean, $\boldsymbol{\mu}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta})$, to the base point p. That is,

$$\mathcal{E}(y_i, \mathbf{x}_i; \mathrm{q}, \boldsymbol{\beta}) = \mathcal{E}_i(\mathrm{q}, \boldsymbol{\beta}) := \mathrm{Log}_{\mathrm{p}}\left(c(1; \mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta})^{-1} \cdot y_i\right) \in T_{\mathrm{p}}M \tag{2.5}$$

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

$\mathcal{M}$ is defined by

$$E[\mathcal{E}(y_i, \mathbf{x}_i; \mathrm{q}_*, \boldsymbol{\beta}_*) | \, \mathbf{x}_i] = 0, \tag{2.6}$$

where $(\mathrm{q}_*, \boldsymbol{\beta}_*)$ denotes the true value of $(\mathrm{q}, \boldsymbol{\beta})$ and the expectation is taken with respect to the conditional distribution of y_i given $\mathbf{x}_i$. Model (2.6) is equivalent to $E[\mathrm{Log}_{\boldsymbol{\mu}(\mathbf{x}_i, \mathrm{q}_*, \boldsymbol{\beta}_*)}(y_i) | \, \mathbf{x}_i] = 0$ for $i = 1, \ldots, n$, since $L_{c(1;\mathbf{x}_i, \mathrm{q}_*, \boldsymbol{\beta}_*)^{-1}*}$ is an isomorphism of linear spaces (invariant under the metric m) between the fibers of $T\mathcal{M}$. 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 semi-parametric, since we do not restrict the joint distribution of $(y, \mathbf{x})$ except by the conditional moment restriction in (2.6).

2.3.2 Generalized Method of Moment Estimators

We consider the generalized method of moment estimator (GMM estimator) to estimate the unknown parameters in model (2.6) (Hansen 1982, Newey 1993, Korsholm 1999). By identifying $T_{\mathrm{p}}\mathcal{M}$ with $R^{d_\mathcal{M}}$, we may assume that $\mathcal{E}$ is a function with values in $R^{d_\mathcal{M}}$. Let $h(\mathbf{x}; \mathrm{q}, \boldsymbol{\beta})$ be a $s \times d_\mathcal{M}$ matrix of functions of $(\mathbf{x}, \mathrm{q}, \boldsymbol{\beta})$ with $s \geq d_\mathcal{M} + d_{\boldsymbol{\beta}}$ and W_n be a random sequence of positive definite $s \times s$ weight matrices. It follows from (2.6) that

$$E\{h(\mathbf{x}_i; \mathrm{q}_*, \boldsymbol{\beta}_*) E[\mathcal{E}(y_i, \mathbf{x}_i; \mathrm{q}_*, \boldsymbol{\beta}_*) | \mathbf{x}_i]\} = 0. \tag{2.7}$$

We define $\mathcal{Q}_n(\mathrm{q}, \boldsymbol{\beta})$ to be

$$\left[\mathbb{P}_n(h(\mathbf{x}; \mathrm{q}, \boldsymbol{\beta}) \mathcal{E}(y, \mathbf{x}; \mathrm{q}, \boldsymbol{\beta}))\right]^\top W_n \left[\mathbb{P}_n(h(\mathbf{x}; \mathrm{q}, \boldsymbol{\beta}) \mathcal{E}(y, \mathbf{x}; \mathrm{q}, \boldsymbol{\beta}))\right],$$

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 $(\hat{\mathrm{q}}_{\mathrm{GMM}}, \hat{\boldsymbol{\beta}}_{\mathrm{GMM}})$, or simply $(\hat{\mathrm{q}}, \hat{\boldsymbol{\beta}})$, of $(\mathrm{q}, \boldsymbol{\beta})$ associated with $(h(\cdot,\cdot,\cdot), W_n)$ is defined as

$$(\hat{\mathrm{q}}_{\mathrm{GMM}}, \hat{\boldsymbol{\beta}}_{\mathrm{GMM}}) = \underset{(\mathrm{q},\boldsymbol{\beta}) \in M \times R^{d_{\boldsymbol{\beta}}}}{\operatorname{argmin}} \mathcal{Q}_n(\mathrm{q}, \boldsymbol{\beta}). \tag{2.8}$$

Under some conditions detailed below, we can show the first order asymptotic properties of $(\hat{\mathrm{q}}_{\mathrm{GMM}}, \hat{\boldsymbol{\beta}}_{\mathrm{GMM}})$ 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 $l = 1, \ldots;$ $\mathbf{v}^{\otimes 2} = \mathbf{v}\mathbf{v}^T$ for any vector $\mathbf{v}$; $V = \mathrm{Var}[h(\mathbf{x}; \mathrm{q}_*, \boldsymbol{\beta}_*)\mathcal{E}(y, \mathbf{x}; \mathrm{q}_*, \boldsymbol{\beta}_*)]$; I_d is a $d \times d$ identity matrix; and $\xrightarrow{d}$ and $\xrightarrow{p}$, respectively, denote convergence in distribution and in probability. We obtain the following theorems, whose detailed proofs can be found in the supplementary document.

Theorem 2.3.1. *Assume that $(y_i, \mathbf{x}_i)$, $i = 1, \ldots, n$, are iid random variables in $\mathcal{M} \times R^{d_{\mathbf{x}}}$. Let $(\mathrm{q}_*, \boldsymbol{\beta}_*)$ be the exact value of the parameters satisfying (2.6). Let $\{W_n\}_n$ be a random sequence of $s \times s$ symmetric positive semi-definite matrices, with $s \geq d_{\mathcal{M}} + d_{\boldsymbol{\beta}}$.*

(a) Assume

- *(C1) $W_n \xrightarrow{p} W$, as $n \to \infty$, where W is a positive semi-definite matrix;*
- *(C2) $W\, E[h(\mathbf{x}; \mathrm{q}, \boldsymbol{\beta})\mathcal{E}(y, \mathbf{x}; \mathrm{q}, \boldsymbol{\beta})] = 0$ only for $(\mathrm{q}, \boldsymbol{\beta}) = (\mathrm{q}_*, \boldsymbol{\beta}_*)$;*
- *(C3) $(\mathrm{q}_*, \boldsymbol{\beta}_*) \in \Theta^o$, where $\Theta = K \times B$ is compact in $\mathcal{M} \times R^{d_{\boldsymbol{\beta}}}$;*
- *(C4) the family of measurable functions $\{(y, \mathbf{x}) \mapsto h(\mathbf{x}; \mathrm{q}, \boldsymbol{\beta})\mathcal{E}(y, \mathbf{x}; \mathrm{q}, \boldsymbol{\beta}) : (\mathrm{q}, \boldsymbol{\beta}) \in \Theta\}$ forms a Glivenko-Cantelli class;*
- *(C5) $\inf_{(\mathrm{q},\boldsymbol{\beta}) \in \Theta : dist_{\mathcal{M}}(\mathrm{q},\mathrm{q}_*) + \|\boldsymbol{\beta} - \boldsymbol{\beta}_*\| \geq \varepsilon} \|W\, E[h(\mathbf{x}; \mathrm{q}, \boldsymbol{\beta})\mathcal{E}(y, \mathbf{x}, \mathrm{q}; \boldsymbol{\beta})]\| > 0$, for every $\varepsilon > 0$.*

Then, $(\hat{\mathrm{q}}, \hat{\boldsymbol{\beta}})$ *in (2.8) is consistent in probability, as* $n \to \infty$.

(b) In addition to the assumptions (C1)-(C4), assume that in some neighborhood of $(\mathrm{q}_*, \boldsymbol{\beta}_*)$*, denoted by* $N_*(\delta)$*, we have*

(C6) $h(\mathbf{x}; \mathrm{q}, \boldsymbol{\beta})$ *and* $\mathcal{E}(y, \mathbf{x}; \mathrm{q}, \boldsymbol{\beta})$ *are continuously twice differentiable with probability one;*

(C7) $0 < E[\|h(\mathbf{x}; \mathrm{q}, \boldsymbol{\beta})\mathcal{E}(y, \mathbf{x}; \mathrm{q}, \boldsymbol{\beta})\|^2] < \infty$ *and* W *is positive definite;*

(C8) $E[\sup_{(\mathrm{q},\boldsymbol{\beta})\in N_*(\delta)} \|\nabla_{(\mathrm{q},\boldsymbol{\beta})}[h_j(\mathbf{x}; \mathrm{q}, \boldsymbol{\beta})\mathcal{E}(y, \mathbf{x}; \mathrm{q}, \boldsymbol{\beta})]\|] < \infty$*, where* $h_j(\mathbf{x}; \mathrm{q}, \boldsymbol{\beta})$ *is the* j*-th row of matrix* $h(\mathbf{x}; \mathrm{q}, \boldsymbol{\beta})$*,* $j = 1, \ldots, s$*;*

(C9) for a chart (U, ϕ) *near* q_* *(and, thus for any chart near* q_**),*

$$G_\phi = G_{\phi,h} = E\left[h(\mathbf{x}; \mathrm{q}_*, \boldsymbol{\beta}_*)\partial_{(\mathbf{t},\boldsymbol{\beta})}\mathcal{E}(y, \mathbf{x}; \phi^{-1}(\mathbf{t}), \boldsymbol{\beta}_*)\big|_{t=\phi(\mathrm{q}_*)}\right]$$

is a $s \times (d_{\mathcal{M}} + d_{\boldsymbol{\beta}})$ *matrix with full column rank;*

(C10) $E\left[\left\|\partial^2_{(\mathbf{t},\boldsymbol{\beta})}\left[h_{j\ell}(\mathbf{x}; \phi^{-1}(\mathbf{t}), \boldsymbol{\beta}_*)\mathcal{E}_\ell(y, \mathbf{x}; \phi^{-1}(\mathbf{t}), \boldsymbol{\beta}_*)\right]\big|_{t=\phi(\mathrm{q}_*)}\right\|\right] < \infty$ *for* $j = 1, \ldots, s$ *and* $\ell = 1, \ldots, d_{\mathcal{M}}$.

Then, for any local chart (U, ϕ) *on* $\mathcal{M}$ *near* q_**, as* $n \to \infty$*, we have*

$$\sqrt{n}\left[(\phi(\hat{\mathrm{q}})^\top, \hat{\boldsymbol{\beta}}^\top)^\top - (\phi(\mathrm{q}_*)^\top, \boldsymbol{\beta}_*^\top)^\top\right] \xrightarrow{d} N_{d_{\mathcal{M}}+d_{\boldsymbol{\beta}}}(\mathbf{0}, \Sigma_\phi), \tag{2.9}$$

where $\Sigma_\phi = (G_\phi^\top W G_\phi)^{-1} G_\phi^\top W V W G_\phi (G_\phi^\top W G_\phi)^{-1}$*. Moreover, for any other chart* (U, ϕ') *near* q_**, we have*

$$\Sigma_{\phi'} = diag(J(\phi' \circ \phi^{-1})_{\phi(\mathrm{q}_*)}, I_{d_{\boldsymbol{\beta}}})\Sigma_\phi diag(J(\phi' \circ \phi^{-1})_{\phi(\mathrm{q}_*)}, I_{d_{\boldsymbol{\beta}}})^\top, \tag{2.10}$$

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

Remark 1: Assumptions (C1)-(C10) are generalizations of standard conditions for ensuring first order asymptotic properties of M-estimators (van der Vaart 1998). Some conditions can be weakened with some intricate arguments. For instance, the uniqueness condition (C2) can be relaxed, since the solution set of the equation

$$W\, E[h(\mathbf{x}; \mathrm{q}, \boldsymbol{\beta})\mathcal{E}(y, \mathbf{x}, \mathrm{q}; \boldsymbol{\beta})] = 0$$

may have multiple elements. It can be shown that the distance between $(\hat{\mathrm{q}}, \hat{\boldsymbol{\beta}})$ and the solution set converges to zero in probability. However, if different true values $(\mathrm{q}_*, \boldsymbol{\beta}_*)$ are isolated points, we can establish the same asymptotic properties of $(\hat{\mathrm{q}}, \hat{\boldsymbol{\beta}})$ as Theorem 2.3.1 for $(\mathrm{q}_*, \boldsymbol{\beta}_*)$, which are close to $(\hat{\mathrm{q}}, \hat{\boldsymbol{\beta}})$.

Remark 2: When Θ is compact and the functions $h(\mathbf{x}; \mathrm{q}, \boldsymbol{\beta})\mathcal{E}(y, \mathbf{x}; \mathrm{q}_*, \boldsymbol{\beta}_*)$ are continuous for every $(y, \mathbf{x})$, the following condition

$$E[\sup_{(\mathrm{q}, \boldsymbol{\beta})} \|h(\mathbf{x}; \mathrm{q}, \boldsymbol{\beta})\mathcal{E}(y, \mathbf{x}; \mathrm{q}, \boldsymbol{\beta})\|] < \infty$$

together with (C2) implies both (C4) and (C5). Also note that if the conditional mean link function $f(\mathbf{x}, \mathrm{q}, \boldsymbol{\beta})$ is continuous in $(\mathrm{q}, \boldsymbol{\beta}) \in M \times R^{d_{\boldsymbol{\beta}}}$ uniformly with respect to $\mathbf{x}$, then $c(1, \mathbf{x}, \mathrm{q}, \boldsymbol{\beta}) \in G$ is continuous in $(\mathrm{q}, \boldsymbol{\beta})$ uniformly with respect to $\mathbf{x}$, which yield that the functions $(\mathrm{q}, \boldsymbol{\beta}) \to \mathcal{E}(y, \mathbf{x}; \mathrm{q}, \boldsymbol{\beta})$ are uniformly continuous with respect to $\mathbf{x}$. If, in addition, $\boldsymbol{\mu}(\mathbf{x}, \mathrm{q}, \boldsymbol{\beta})$ and $h(\mathbf{x}; \mathrm{q}, \boldsymbol{\beta})$ are twice continuously differentiable in $(\mathrm{q}, \boldsymbol{\beta})$ in some neighborhood of $(\mathrm{q}_*, \boldsymbol{\beta}_*)$ with probability one, then (C6) holds.

Theorem 2.3.1 establishes the first-order asymptotic properties of $(\hat{\mathrm{q}}_{\mathrm{GMM}}, \hat{\boldsymbol{\beta}}_{\mathrm{GMM}})$ for the intrinsic regression model (2.6). Theorem 2.3.1 (a) establishes the consistency of $(\hat{\mathrm{q}}, \hat{\boldsymbol{\beta}})$. The consistency result does not depend on the local chart. Theorem 2.3.1 (b) establishes the asymptotic normality of $(\phi(\hat{\mathrm{q}}), \hat{\boldsymbol{\beta}})$ 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 $d_{\boldsymbol{\beta}} \times d_{\boldsymbol{\beta}}$ submatrix of $\Sigma_{\phi'}$ that the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}$ does not depend on the chart. However, the asymptotic normality of $\hat{\mathrm{q}}$ does depend on a specific chart. A consistent estimator of the asymptotic covariance matrix Σ_ϕ is given by

$$(\widehat{G}_\phi^\top W_n \widehat{G}_\phi)^{-1} \widehat{G}_\phi^\top W_n \widehat{V} W_n \widehat{G}_\phi (\widehat{G}_\phi^\top W_n \widehat{G}_\phi)^{-1},$$

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

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

where $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$ for any matrix or vector $\mathbf{a}$. This estimator is also compatible with the manifold structure of $\mathcal{M}$.

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

$$\begin{aligned} (\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I) &= \underset{(\mathrm{q},\boldsymbol{\beta}) \in M \times R^{d_{\boldsymbol{\beta}}}}{\operatorname{argmin}} \mathcal{G}_{I,n}(\mathrm{q}, \boldsymbol{\beta}) \\ &= \underset{(\mathrm{q},\boldsymbol{\beta}) \in M \times R^{d_{\boldsymbol{\beta}}}}{\operatorname{argmin}} \sum_{i=1}^n \mathrm{dist}_{\mathcal{M}}(y_i, \boldsymbol{\mu}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta}))^2. \end{aligned} \tag{2.11}$$

According to (2.2), the ILSE is closely related to the intrinsic mean $\hat{\mathrm{q}}_{IM}$ of $y_1, \cdots, y_n \in \mathcal{M}$, which is defined as

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

Recall that $\boldsymbol{\mu}(\mathbf{0}, \mathrm{q}, \boldsymbol{\beta})$ is independent of both $\boldsymbol{\beta}$ and the covariates of interest.

The $(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)$ can be regarded as a special case of the GMM estimator when we set $W_n = I_{d_{\mathcal{M}}+d_\beta}$ and $h(\mathbf{x}, \mathrm{q}, \boldsymbol{\beta}) = \begin{pmatrix} (L_{c(1;\mathbf{x}_i,\mathrm{q},\boldsymbol{\beta})^{-1}.*}(\partial_{t_1}\boldsymbol{\mu}(\mathbf{x}, \phi^{-1}(\mathbf{t}), \boldsymbol{\beta})_{|_{\mathbf{t}=\phi(\mathrm{q})}}))^\top \\ \vdots \\ (L_{c(1;\mathbf{x}_i,\mathrm{q},\boldsymbol{\beta})^{-1}.*}(\partial_{t_{d_{\mathcal{M}}}}\boldsymbol{\mu}(\mathbf{x}, \phi^{-1}(\mathbf{t}), \boldsymbol{\beta})_{|_{\mathbf{t}=\phi(\mathrm{q})}}))^\top \\ (L_{c(1;\mathbf{x}_i,\mathrm{q},\boldsymbol{\beta})^{-1}.*}(\partial_{\beta_1}\boldsymbol{\mu}(\mathbf{x}, \mathrm{q}, \boldsymbol{\beta})))^\top \\ \vdots \\ (L_{c(1;\mathbf{x}_i,\mathrm{q},\boldsymbol{\beta})^{-1}.*}(\partial_{\beta_{d_{\boldsymbol{\beta}}}}\boldsymbol{\mu}(\mathbf{x}, \mathrm{q}, \boldsymbol{\beta})))^\top \end{pmatrix} \in R^{(d_{\mathcal{M}}+d_{\boldsymbol{\beta}})\times d_{\mathcal{M}}}$, where each row is in $R^{1\times d_{\mathcal{M}}}$ via the identification $T_{\mathrm{p}}\mathcal{M} \cong R^{d_{\mathcal{M}}}$, and (U, ϕ) is a chart on $\mathcal{M}$

Specifically, $(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)$ solves a set of estimating equations given by

$$
\begin{aligned}
\mathbf{0} &= \partial_{t_\ell}\mathcal{G}_{I,n}(\phi^{-1}(\mathbf{t}), \boldsymbol{\beta})_{|_{\mathbf{t}=\phi(\mathrm{q})}} \\
&= -2\mathbb{P}_n[< \partial_{t_\ell}\boldsymbol{\mu}(\cdot, \phi^{-1}(\mathbf{t}), \boldsymbol{\beta})_{|_{\mathbf{t}=\phi(\mathrm{q})}}, \mathrm{Log}_{\boldsymbol{\mu}(\cdot,\mathrm{q},\boldsymbol{\beta})}(\cdot) >_{T_{\boldsymbol{\mu}(\cdot,\mathrm{q},\boldsymbol{\beta})}\mathcal{M}}] \\
&= -2\mathbb{P}_n[< L_{c(1;\mathbf{x}_i,\mathrm{q},\boldsymbol{\beta})^{-1}.*}(\partial_{t_\ell}\boldsymbol{\mu}(\cdot, \phi^{-1}(\mathbf{t}), \boldsymbol{\beta})_{|_{\mathbf{t}=\phi(\mathrm{q})}}), \mathcal{E}(\cdot, \cdot; \mathrm{q}, \boldsymbol{\beta}) >_{T_{\mathrm{p}}\mathcal{M}}], \\
\mathbf{0} &= \partial_{\beta_k}\mathcal{G}_{I,n}(\mathrm{q}, \boldsymbol{\beta}) \\
&= -2\mathbb{P}_n[< \partial_{\beta_\ell}\boldsymbol{\mu}(\cdot, \mathrm{q}, \boldsymbol{\beta}), \mathrm{Log}_{\boldsymbol{\mu}(\cdot,\mathrm{q},\boldsymbol{\beta})}(\cdot) >_{T_{\boldsymbol{\mu}(\cdot,\mathrm{q},\boldsymbol{\beta})}\mathcal{M}}] \\
&= -2\mathbb{P}_n[< L_{c(1;\mathbf{x}_i,\mathrm{q},\boldsymbol{\beta})^{-1}.*}(\partial_{\beta_k}\boldsymbol{\mu}(\cdot, \mathrm{q}, \boldsymbol{\beta})), \mathcal{E}(\cdot, \cdot; \mathrm{q}, \boldsymbol{\beta}) >_{T_{\mathrm{p}}\mathcal{M}}],
\end{aligned}
$$

for $\ell = 1, \ldots, d_{\mathcal{M}}$ and $k = 1, \ldots, d_{\boldsymbol{\beta}}$. Theoretically, it follows from Theorem 2.3.1 that under model (2.6), $(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)$ enjoys the first-order asymptotic properties as well.

2.3.3 Efficient GMM 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 $W^{opt} = V^{-1}$, and the use of $W_n = W^{opt}$ leads to the most efficient estimator in the class of all GMM estimators obtained 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 $h^{opt}(\cdot)$ is.

To address this question, we introduce some notation. For a chart (U, ϕ) on $\mathcal{M}$ near q_*, let

$$\begin{aligned} D_\phi(\mathbf{x}) &= E[\partial_{(\mathbf{t},\boldsymbol{\beta})}\mathcal{E}(y, \mathbf{x}; \phi^{-1}(\mathbf{t}), \boldsymbol{\beta}_*)\big|_{\mathbf{t}=\phi(\mathrm{q}_*)}|\mathbf{x}]^\top,\ h^*_\phi(\mathbf{x}) = D_\phi(\mathbf{x})\Omega(\mathbf{x})^{-1}, \\ W^*_\phi &= E[D_\phi(\mathbf{x})\Omega(\mathbf{x})^{-1}D_\phi(\mathbf{x})^\top]^{-1},\ \Omega(\mathbf{x}) = \mathrm{Var}(\mathcal{E}(y, \mathbf{x}; \mathrm{q}_*, \boldsymbol{\beta}_*)|\mathbf{x}). \end{aligned}$$

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

Theorem 2.3.2. *Suppose that (C2)-(C8) and (C10) hold for $h^*_\phi(\mathbf{x})$ and W^*_ϕ. Assume that*

(C11) $\Omega(\mathbf{x})$ is a $d_\mathcal{M} \times d_\mathcal{M}$ positive-definite matrix with probability 1;

(C12) $E[D_\phi(\mathbf{x})\Omega(\mathbf{x})^{-1}D_\phi(\mathbf{x})^\top]$ is a positive-definite matrix.

We have the following results:

(i) $(\hat{\mathrm{q}}^*, \hat{\boldsymbol{\beta}}^*)$ is asymptotically normally distributed with zero mean and covariance W^*_ϕ;

(ii) $(\hat{\mathrm{q}}^*, \hat{\boldsymbol{\beta}}^*)$ is optimal among all GMM estimators for model (2.6);

(iii) $(\hat{\mathrm{q}}^*, \hat{\boldsymbol{\beta}}^*)$ is independent of the chart.

Theorem 2.3.2 characterizes the optimality of $h^*_\phi(\mathbf{x})$ and W^*_ϕ among regular GMM estimators for model (2.6). Geometrically, $(\hat{\mathrm{q}}^*, \hat{\boldsymbol{\beta}}^*)$ is independent of the chart. Specifically, for any other chart (U, ϕ') near q_*, we have

$$\begin{aligned} D_{\phi'}(\mathbf{x}) &= \mathrm{diag}\left([J(\phi' \circ \phi^{-1})_{\phi(\mathrm{q}_*)}]^{-1}, I_{d_{\boldsymbol{\beta}}}\right)^\top D_\phi(\mathbf{x}), \\ h^*_{\phi'}(\mathbf{x}) &= \mathrm{diag}\left([J(\phi' \circ \phi^{-1})_{\phi(\mathrm{q}_*)}]^{-1}, I_{d_{\boldsymbol{\beta}}}\right)^\top h^*_\phi(\mathbf{x}), \\ W^*_{\phi'} &= \mathrm{diag}(J(\phi' \circ \phi^{-1})_{\phi(\mathrm{q}_*)}, I_{d_{\boldsymbol{\beta}}})W^*_\phi \mathrm{diag}(J(\phi' \circ \phi^{-1})_{\phi(\mathrm{q}_*)}, I_{d_{\boldsymbol{\beta}}})^\top. \end{aligned}$$

Thus, the quadratic form in (2.8) 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 $(\hat{\mathrm{q}}^*, \hat{\boldsymbol{\beta}}^*)_\phi$ 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 of $(\mathrm{q}, \boldsymbol{\beta})$, such as $(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)$. The second step is to plug $(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)$ into the functions $\mathcal{E}_i(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)$ and $\partial_{(\mathbf{t},\boldsymbol{\beta})}\mathcal{E}(y_i, \mathbf{x}_i; \phi^{-1}(\mathbf{t}), \hat{\boldsymbol{\beta}}_I)|_{\mathbf{t}=\phi(\hat{\mathrm{q}}_I)}$, for all i and then use them to construct the nonparametric 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_0-th order satisfying $\int K(\mathbf{u})d\mathbf{u} = 1$, $\int u_s^l K(\mathbf{u})d\mathbf{u} = 0$, for any $s = 1, \ldots, d_{\mathbf{x}}$ and $1 \leq l < l_0$, and $\int u_s^{l_0} K(\mathbf{u})d\mathbf{u} \neq 0$, where $\mathbf{u} = (u_1, \ldots, u_{d_{\mathbf{x}}})^T$. Let $K_\tau(u) = \tau^{-1}K(u/\tau)$, where τ is a bandwidth. Then, a nonparametric estimator of $D_\phi(\mathbf{x})$ can be written by

$$\widehat{D}_\phi(\mathbf{x})^\top = \sum_{i=1}^{n} \omega_i(\mathbf{x}; \tau)\partial_{(\mathbf{t},\boldsymbol{\beta})}\mathcal{E}(y_i, \mathbf{x}_i; \phi^{-1}(\mathbf{t}), \hat{\boldsymbol{\beta}}_I)|_{\mathbf{t}=\phi(\hat{\mathrm{q}}_I)}, \tag{2.12}$$

where $\omega_i(\mathbf{x}; \tau) = K_\tau(\mathbf{x} - \mathbf{x}_i)/\{\sum_{k=1}^n K_\tau(\mathbf{x} - \mathbf{x}_k)\}$. Although we may construct a nonparametric estimator of $\Omega(\mathbf{x})$ similar to (2.12), we have found that even for moderate $d_{\mathbf{x}}$, such an estimator is numerically unstable. Instead, we approximate $\Omega(\mathbf{x}_i) = \mathrm{Var}(\mathcal{E}(y, \mathbf{x}; \mathrm{q}_*, \boldsymbol{\beta}_*)|\mathbf{x} = \mathbf{x}_i)$ by its mean $V_{E*} = \mathrm{Var}(\mathcal{E}(y, \mathbf{x}; \mathrm{q}_*, \boldsymbol{\beta}_*))$. In this case, $h^*_\phi(\mathbf{x})$ and W^*_ϕ, respectively, reduce to

$$h^*_{E,\phi}(\mathbf{x}) = D_\phi(\mathbf{x})V_{E*}^{-1},\ W^*_{E,\phi} = \{E[D_\phi(\mathbf{x})V_{E*}^{-1}\Omega(\mathbf{x})V_{E*}^{-1}D_\phi(\mathbf{x})^T]\}^{-1}. \tag{2.13}$$

For any local chart (U, ϕ) with $\hat{\mathrm{q}}_I \in U$, we construct the estimators of $h^*_{E,\phi}$ and

$W^*_{E,\phi}$ as follows. Let $\hat{V}(\mathrm{q},\boldsymbol{\beta}) = \mathbb{P}_n\mathcal{E}(y,\mathbf{x};\mathrm{q},\boldsymbol{\beta})^{\otimes 2}$, we have

$$\begin{aligned}\hat{h}_{E,\phi}(\mathbf{x}_i) &= \hat{h}_{E,\phi,i} = \widehat{D}_\phi(\mathbf{x}_i)\hat{V}(\hat{\mathrm{q}}_I,\hat{\boldsymbol{\beta}}_I)^{-1},\\ \widehat{W}_{E,\phi} &= \left\{\mathbb{P}_n[\hat{h}_{E,\phi}(\mathbf{x})\mathcal{E}(y,\mathbf{x};\hat{\mathrm{q}}_I,\hat{\boldsymbol{\beta}}_I)]^{\otimes 2}\right\}^{-1}. \end{aligned} \tag{2.14}$$

Then, we substitute $\hat{h}_{E,\phi}$ and $\widehat{W}_{E,\phi}$ into (2.8) and then calculate the GMM estimator of $(\mathrm{q},\boldsymbol{\beta})$, denoted by $(\hat{\mathrm{q}}_E,\hat{\boldsymbol{\beta}}_E)$. Similar to $(\hat{\mathrm{q}}^*,\hat{\boldsymbol{\beta}}^*)$, it can be shown that $(\hat{\mathrm{q}}_E,\hat{\boldsymbol{\beta}}_E)$ is independent of the chart (U,ϕ) on $\mathcal{M}$ near q_* with $\hat{\mathrm{q}}_I \in U$. For sufficiently large n, $\mathrm{dist}_{\mathcal{M}}(\hat{\mathrm{q}}_I,\mathrm{q}_*) < \rho^*(\mathcal{M},\mathrm{p}_*)$ and any maximal normal chart on $\mathcal{M}$ centered at $\hat{\mathrm{q}}_I$ contains the true value q_* with probability approaching one. Computationally, an annealing evolutionary stochastic approximation Monte Carlo algorithm is developed to compute $(\hat{\mathrm{q}}_I,\hat{\boldsymbol{\beta}}_I)$ and $(\hat{\mathrm{q}}_E,\hat{\boldsymbol{\beta}}_E)$. See the supplementary document for details.

We calculate a one-step linearized estimator of $(\mathrm{q},\boldsymbol{\beta})$, denoted by $(\tilde{\mathrm{q}}_E,\tilde{\boldsymbol{\beta}}_E)$, to approximate $(\hat{\mathrm{q}}_E,\hat{\boldsymbol{\beta}}_E)$. Computationally, the linearized estimator does not require iteration, whereas, theoretically, it shares the first asymptotic properties with $(\hat{\mathrm{q}}_E,\hat{\boldsymbol{\beta}}_E)$ as shown below. Specifically, in the chart (U,ϕ) near $\hat{\mathrm{q}}_I$, we have

$$\begin{aligned}&(\tilde{\mathbf{t}}^\top_{E,\phi},\tilde{\boldsymbol{\beta}}^\top_{E,\phi})^\top - (\phi(\hat{\mathrm{q}}_I)^\top,\hat{\boldsymbol{\beta}}_I^\top)^\top \quad = \\ &\left\{-\mathbb{P}_n[\hat{h}_{E,\phi}(\mathbf{x})\partial_{(\mathbf{t},\boldsymbol{\beta})}\mathcal{E}(y,\mathbf{x};\phi^{-1}(t),\hat{\boldsymbol{\beta}}_I)|_{\mathbf{t}=\phi(\hat{\mathrm{q}}_I)}]\right\}^{-1}\mathbb{P}_n\left[\hat{h}_{E,\phi}(\mathbf{x})\mathcal{E}_i(y,\mathbf{x};\hat{\mathrm{q}}_I,\hat{\boldsymbol{\beta}}_I)\right].\end{aligned} \tag{2.15}$$

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

$$\begin{aligned}&(\tilde{\mathbf{t}}^\top_{E,\phi'},\tilde{\boldsymbol{\beta}}^\top_{E,\phi'})^\top - (\phi'(\hat{\mathrm{q}}_I)^\top,\hat{\boldsymbol{\beta}}_I^\top)^\top\\ &\quad= \begin{pmatrix} J_{\phi(\hat{\mathrm{q}}_I)}(\phi'\circ\phi^{-1}) & 0\\ 0 & I_{d_{\boldsymbol{\beta}}}\end{pmatrix}[(\tilde{\mathbf{t}}^\top_{E,\phi},\tilde{\boldsymbol{\beta}}^\top_{E,\phi})^\top - (\phi(\hat{\mathrm{q}}_I)^\top,\hat{\boldsymbol{\beta}}_I^\top)^\top].\end{aligned}$$

Thus, $\tilde{\boldsymbol{\beta}}_{E,\phi}$ is independent of the chart ϕ and $\{\tilde{\mathbf{t}}_{E,\phi} - \phi(\hat{\mathrm{q}}_I)\,|\,\phi \text{ is a chart on } M\}$ defines

a tangent vector to $\mathcal{M}$ at $\hat{\mathrm{q}}_I$. Moreover, if ϕ and ϕ' are maximal normal charts centered at $\hat{\mathrm{q}}_I$, 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 $\hat{\mathrm{q}}_I$ 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 centered at $\hat{\mathrm{q}}_I$. Finally, we can establish the first order asymptotic properties of $(\tilde{\mathrm{q}}_E, \tilde{\boldsymbol{\beta}}_E)$ as follows.

Theorem 2.3.3. *Assume that (C2)-(C11) are valid and*

(C13) $(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)$ *is a* $\sqrt{n}$*-consistent estimator of* $(\mathrm{q}_*, \boldsymbol{\beta}_*)$*;*

(C14) (U, ϕ) *is a maximal normal chart on* $\mathcal{M}$ *centered at* $\hat{\mathrm{q}}_I$*, with* $\mathrm{q}_* \in U$*;*

(C15) V_{E*}*,* $G_{\phi,h^*_{E,\phi}} = E[D_\phi(\mathbf{x})V_{E*}^{-1}D_\phi(\mathbf{x})^\top]$ *and* $W^*_{E,\phi}$ *are positive-definite;*

(C16) $\sup_{(\mathrm{q},\boldsymbol{\beta})\in\Theta} \|\mathcal{E}(y,\mathbf{x};\mathrm{q},\boldsymbol{\beta})\| + \sum_{l=1}^2 \sup_{(\mathrm{q},\boldsymbol{\beta})\in B((\mathrm{q}_*,\boldsymbol{\beta}_*),\delta)} \|\partial^l_{(\mathrm{q},\boldsymbol{\beta})}\mathcal{E}(y,\mathbf{x};\mathrm{q},\boldsymbol{\beta})\| \le f_0(y,\mathbf{x})$ *with* $E[f_0(y,\mathbf{x})^4] < \infty$ *for a small* $\delta > 0$*;*

(C17) $\sup_{(\mathrm{q},\boldsymbol{\beta})\in B((\mathrm{q}_*,\boldsymbol{\beta}_*),\delta)} \|E[\partial^l_{(\mathrm{q},\boldsymbol{\beta})}\mathcal{E}(y,\mathbf{x};\mathrm{q},\boldsymbol{\beta})|\mathbf{x}]\| \le f_1(\mathbf{x})$ *for* $l = 1$ *and* 2 *and* $\delta > 0$ *with* $E[f_0(y,\mathbf{x})^4 f_1(\mathbf{x})^2] < \infty$

(C18) $\mathbb{P}_n\|\widehat{D}_\phi(\mathbf{x}) - D_\phi(\mathbf{x})\|^2 = o_p(1)$*, and* $\sqrt{n}\mathbb{P}_n\{[\widehat{D}_\phi(\mathbf{x}) - D_\phi(\mathbf{x})] \otimes \mathcal{E}(y,\mathbf{x};\mathrm{q}_*,\boldsymbol{\beta}_*)\} = o_p(1)$.

As $n \to \infty$*, we have the following results:*

$$\sqrt{n}\,[(\phi(\tilde{\mathrm{q}}_E)^\top, \tilde{\boldsymbol{\beta}}_E^\top)^\top - (\phi(\mathrm{q}_*)^\top, \boldsymbol{\beta}_*^\top)^\top] \xrightarrow{d} N_{d_\mathcal{M}+d_{\boldsymbol{\beta}}}(\mathbf{0}, \Sigma_{E,\phi}), \tag{2.16}$$

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*

$$\begin{aligned}\widehat{\Sigma}_{E,\phi} \;=\; & n^{-1}\left\{\mathbb{P}_n[\widehat{D}_\phi(\mathrm{x})\widehat{V}(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)^{-1}\widehat{D}_\phi(\mathrm{x})^\top]\right\}^{-1} \\ & \times\left\{\mathbb{P}_n[\widehat{D}_\phi(\mathrm{x})\widehat{V}(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)^{-1}\mathcal{E}(y, \mathbf{x}, \tilde{\mathrm{q}}_I, \tilde{\boldsymbol{\beta}}_I)^{\otimes 2}\widehat{V}(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)^{-1}\widehat{D}_\phi(\mathrm{x})^\top]\right\} \\ & \times\left\{\mathbb{P}_n[\widehat{D}_\phi(\mathrm{x})\widehat{V}(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)^{-1}\widehat{D}_\phi(\mathrm{x})^\top]\right\}^{-1},\end{aligned} \tag{2.17}$$

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

Theorem 2.3.3 establishes the first-order asymptotic properties of $(\tilde{\mathrm{q}}_E, \tilde{\boldsymbol{\beta}}_E)$. If $\Omega(\mathbf{x}) = \Omega$ for a constant matrix Ω, then it follows from Theorems 2.3.2 and 2.3.3 that $(\tilde{\mathrm{q}}_E, \tilde{\boldsymbol{\beta}}_E)$ is optimal. If $\Omega(\mathbf{x})$ does not vary dramatically as a function of $\mathbf{x}$, then $(\tilde{\mathrm{q}}_E, \tilde{\boldsymbol{\beta}}_E)$ is nearly optimal. If $\Omega(\mathbf{x})$ varies dramatically as a function of $\mathbf{x}$, one can replace $\widehat{V}(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)$ in (2.14) by $\widehat{\Omega}(\mathbf{x}_i)$ to obtain $\hat{h}_{E,\phi}(\mathbf{x}) = \widehat{D}_\phi(\mathbf{x})\widehat{\Omega}(\mathbf{x})^{-1}$ and $\widehat{W}_{E,\phi} = \{\mathbb{P}_n[\hat{h}_{E,\phi}(\mathbf{x})\mathcal{E}(y, \mathbf{x}; \hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)]^{\otimes 2}\}^{-1}$, where $\widehat{\Omega}(\mathbf{x}_i)$ is a consistent estimator of $\Omega(\mathbf{x}_i)$ for all i, then the optimality of $(\tilde{\mathrm{q}}_E, \tilde{\boldsymbol{\beta}}_E)$ still holds. We have the following theorem.

Theorem 2.3.4. *Assume that (C2)-(C17) are valid and*

(C19) $\sqrt{n}\mathbb{P}_n\{[\widehat{D}_\phi(\mathbf{x})\widehat{\Omega}(\mathbf{x})^{-1} - D_\phi(\mathbf{x})\Omega(\mathbf{x})^{-1}]\mathcal{E}(y, \mathbf{x}; \mathrm{q}_*, \boldsymbol{\beta}_*)\} = o_p(1)$,
$\mathbb{P}_n\|\widehat{D}_\phi(\mathbf{x}) - D_\phi(\mathbf{x})\|^4 = o_p(1)$, $\mathbb{P}_n\|\widehat{\Omega}(\mathbf{x})^{-1} - \Omega(\mathbf{x})^{-1}\|^4 = o_p(1)$.

Then, as $n \to \infty$*, we have*

$$\sqrt{n}\,[(\phi(\tilde{\mathrm{q}}_E)^\top, \tilde{\boldsymbol{\beta}}_E^\top)^\top - (\phi(\mathrm{q}_*)^\top, \boldsymbol{\beta}_*^\top)^\top] \xrightarrow{d} N_{d_{\mathcal{M}}+d_{\boldsymbol{\beta}}}(\mathbf{0}, \Sigma_\phi^*), \tag{2.18}$$

in which Σ_ϕ^* *is given in Theorem 2.3.2. If we set*

$$\widehat{\Sigma}_{E,\phi} = n^{-1}\left\{\mathbb{P}_n[\widehat{D}_\phi(\mathbf{x})\widehat{\Omega}(\mathbf{x})^{-1}\widehat{D}_\phi(\mathbf{x})^\top]\right\}^{-1}, \tag{2.19}$$

then $n\widehat{\Sigma}_{E,\phi}$ *is a consistent estimator of* Σ^*_ϕ.

We note that our two-stage method estimates jointly q and $\boldsymbol{\beta}$. This is different than the naïve idea of estimating first q by the sample Fréchet mean $\hat{\mathrm{q}}_F$ of y_i's, and then estimate $\boldsymbol{\beta}$ using the residuals at $\hat{\mathrm{q}}_F$; the asymptotic efficiency is not guaranteed.

2.3.4 Hypotheses Testing

Many scientific questions involve in 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:

$$H_0^{(1)} : \mathrm{C}_0\boldsymbol{\beta} = \mathbf{b}_0 \quad \text{vs.} \quad H_1^{(1)} : \mathrm{C}_0\boldsymbol{\beta} \neq \mathbf{b}_0, \tag{2.20}$$

$$H_0^{(2)} : \mathrm{q} = \mathrm{q}_0 \quad \text{vs.} \quad H_1^{(2)} : \mathrm{q} \neq \mathrm{q}_0, \tag{2.21}$$

where C_0 is a $r \times d_{\boldsymbol{\beta}}$ 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 multicenter link function, we may be interested in testing whether all $\mathrm{q}(\mathbf{x}_D)$ are independent of $\mathbf{x}_D$.

We develop several test statistics for testing the hypotheses given in (2.20) and (2.21). Firstly, we consider the Wald test statistic for testing $H_0^{(1)}$ against $H_1^{(1)}$ in (2.20), which is given by

$$\begin{aligned} W^{(1)}_{n,\phi} &= (\mathrm{C}_0\tilde{\boldsymbol{\beta}}_E - \mathbf{b}_0)^\top \left[(\mathbf{0}\ \mathrm{C}_0)\hat{\Sigma}_{E,\phi}(\mathbf{0}\ \mathrm{C}_0)^\top\right]^{-1} (\mathrm{C}_0\tilde{\boldsymbol{\beta}}_E - \mathbf{b}_0) \\ &= (\mathrm{C}_0\tilde{\boldsymbol{\beta}}_E - \mathbf{b}_0)^\top \left[\mathrm{C}_0\hat{\Sigma}_{E,\phi;22}\mathrm{C}_0^\top\right]^{-1} (\mathrm{C}_0\tilde{\boldsymbol{\beta}}_E - \mathbf{b}_0), \end{aligned}$$

where $\hat{\Sigma}_{E,\phi}$ is given in Theorem 2.3.3 or Theorem 2.3.4 , 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.

Secondly, we consider the Wald test statistic for testing the hypotheses given in (2.21) when there is a local chart (U, ϕ) on $\mathcal{M}$ containing both $\hat{\mathrm{q}}_E$ and q_0. Specifically, the Wald test statistic for testing (2.21) is defined by

$$W_{n,\phi}^{(2)} = (\phi(\tilde{\mathrm{q}}_E) - \phi(\mathrm{q}_0))^\top \left[(I_{d_\mathcal{M}} \ \mathbf{0}) \hat{\Sigma}_{E,\phi} (I_{d_\mathcal{M}} \ \mathbf{0})^\top \right]^{-1} (\phi(\tilde{\mathrm{q}}_E) - \phi(\mathrm{q}_0)).$$

Thirdly, we develop an intrinsic Wald test statistic, that is independent of the chart, for testing the hypotheses given in (2.21). We consider the asymptotic covariance estimator $\widehat{\Sigma}_{E,\phi}$ based on $\tilde{\mathrm{q}}_E$ and its upper-left $d_\mathcal{M} \times d_\mathcal{M}$ submatrix $\widehat{\Sigma}_{E,\phi;11}$. Since both are compatible with the manifold structure of $\mathcal{M}$, $\widehat{\Sigma}_{E,\phi;11}$ defines a unique non-degenerate linear map $\widehat{\Sigma}_{E;11}(\cdot)$ from the tangent space $T_{\tilde{\mathrm{q}}_E}\mathcal{M}$ of $\mathcal{M}$ at $\tilde{\mathrm{q}}_E$ onto itself, which is independent of the chart (U, ϕ). In a maximal normal chart centered at $\tilde{\mathrm{q}}_E$, then in any such normal chart, the Wald test statistic for testing (2.21) is given by

$$W_{\mathcal{M},n}^{(2)} = \mathrm{m}_{\tilde{\mathrm{q}}_E}((\widehat{\Sigma}_{E;11})^{-1}(\mathrm{Log}_{\tilde{\mathrm{q}}_E}\mathrm{q}_0), \mathrm{Log}_{\tilde{\mathrm{q}}_E}\mathrm{q}_0).$$

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

Theorem 2.3.5. *Let (U, ϕ) be a local chart on $\mathcal{M}$ so that $\tilde{\mathrm{q}}_E, \mathrm{q}_* \in U$. Assume that all conditions in Theorem 2.3.3 hold. Under the corresponding null hypothesis, we have the following results:*

(i) $W_{n,\phi}^{(1)}$ and $W_{n,\phi}^{(2)}$ are asymptotically distributed as χ_r^2 and $\chi_{d_\mathcal{M}}^2$, respectively;

(ii) $W_{n,\phi}^{(1)}$ is independent of the chart (U, ϕ);

(iii) for any other local chart (U, ϕ') with $\tilde{\mathrm{q}}_E$ and q_0 in U,

$$W_{n,\phi'}^{(2)} = W_{n,\phi}^{(2)} + o_p(1).$$

(iv) For any normal chart (U, ϕ) centered at $\tilde{\mathrm{q}}_E$, $W^{(2)}_{n,\phi} = W^{(2)}_{\mathcal{M},n}$.

Theorem 2.3.5 has several important implications. Theorem 2.3.5 (i) characterizes the asymptotic null distributions of $W^{(1)}_{n,\phi}$ and $W^{(2)}_{n,\phi}$. Theorem 2.3.5 (ii) shows that $W^{(1)}_{n,\phi}$ does not depend the choice of the chart (U, ϕ) on $\mathcal{M}$. Theorem 2.3.5 (iii) shows that $W^{(2)}_{n,\phi'}$ and $W^{(2)}_{n,\phi}$ are asymptotically equivalent for any two local charts. Theorem 2.3.5 (iv) shows that $W^{(2)}_{n,\phi'}$ can be used to construct an intrinsic test statistic.

We consider a local alternative framework for (2.20) and (2.21) as follows:

$$H_0^{(1)} : \mathrm{C}_0\boldsymbol{\beta} = \mathbf{b}_0 \quad \text{vs.} \quad H_{1,n}^{(1)} : \mathrm{C}_0\boldsymbol{\beta} = \mathbf{b}_0 + \delta/\sqrt{n} + o(1/\sqrt{n}), \tag{2.22}$$

$$H_0^{(2)} : \mathrm{q} = \mathrm{q}_0 \quad \text{vs.} \quad H_{1,n}^{(2)} : \mathrm{q} = \mathrm{Exp}_{\mathrm{q}_0}(\mathbf{v}/\sqrt{n} + o(1/\sqrt{n})), \tag{2.23}$$

where δ and $\mathbf{v}$ are specified (and fixed) in R^r and $T_{\mathrm{q}_0}\mathcal{M}$, 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 2.3.6. *Let (U, ϕ) be a local chart on $\mathcal{M}$ so that $\tilde{\mathrm{q}}_E, \mathrm{q}_* \in U$. Assume that all conditions in Theorem 2.3.3 hold. Under the local alternatives (2.22) and (2.23), we have the following results:*

(i) Under $H^{(1)}_{1,n}$, $W^{(1)}_{n,\phi}$ is asymptotically distributed as noncentral χ^2_r with noncentrality parameter $\delta^\top \left[\mathrm{C}_0\hat{\Sigma}_{E,\phi;22}\mathrm{C}_0^\top\right]^{-1}\delta$.

(ii) Under $H^{(2)}_{1,n}$, $W^{(2)}_{n,\phi}$ is asymptotically distributed as noncentral $\chi^2_{d_\mathcal{M}}$, with noncentrality parameter

$$J(\phi \circ \mathrm{Exp}_{\mathrm{q}_0})_0(\mathbf{v})^\top \left[\hat{\Sigma}_{E,\phi;11}\right]^{-1} J(\phi \circ \mathrm{Exp}_{\mathrm{q}_0})_0(\mathbf{v}).$$

The noncentrality parameter does not depend on the choice of the coordinate system at q_0. Here, $J(f)_a$ denotes the Jacobian matrix of map f at a.

(iii) Under $H_{1,n}^{(2)}$, $W_{\mathcal{M},n}^{(2)}$ is asymptotically distributed as noncentral $\chi^2_{d_{\mathcal{M}}}$, with noncentrality parameter

$$\mathrm{m}_{\tilde{\mathrm{q}}_E}((\widehat{\Sigma}_{E;11})^{-1}(J(\mathrm{Log}_{\tilde{\mathrm{q}}_E})_{\mathrm{q}_0}(\mathbf{v})), (J(\mathrm{Log}_{\tilde{\mathrm{q}}_E})_{\mathrm{q}_0}(\mathbf{v}))).$$

The noncentrality parameter does not depend on the choice of the coordinate systems at $\tilde{\mathrm{q}}_E$ and q_0, respectively.

We consider another scenario that there are no local charts on $\mathcal{M}$ containing both $\tilde{\mathrm{q}}_E$ and q_0. In this case, we restate the hypotheses $H_0^{(2)}$ and $H_1^{(2)}$ as follows:

$$H_0^{(2)} : \mathrm{dist}_{\mathcal{M}}(\mathrm{q}, \mathrm{q}_0) = 0 \quad \text{vs.} \quad H_1^{(2)} : \mathrm{dist}_{\mathcal{M}}(\mathrm{q}, \mathrm{q}_0) \neq 0. \tag{2.24}$$

We propose a geodesic test statistic given by

$$W_{dist} = \mathrm{dist}_{\mathcal{M}}(\tilde{\mathrm{q}}_E, \mathrm{q}_0)^2, \tag{2.25}$$

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

Theorem 2.3.7. *Assume that all conditions in Theorem 2.3.5 hold.*

(a) Under $H_0^{(2)}$, nW_{dist} is asymptotically weighted chi-square $\chi^2(\lambda_1, \ldots, \lambda_{d_{\mathcal{M}}})$ distributed, where the weights $\lambda_1, \ldots, \lambda_{d_{\mathcal{M}}}$ are the eigenvalues of the matrix $\Sigma_{E,\mathrm{Log}_{\mathrm{q}_0},11}$, which is the upper-left $d_{\mathcal{M}} \times d_{\mathcal{M}}$ submatrix of the asymptotic covariance matrix $\Sigma_{E,\mathrm{Log}_{\mathrm{q}_0}}$ of $\tilde{\mathrm{q}}_E$ in a normal chart centered at q_0. Moreover, the weights are independent, up to a permutation, of the choice of the normal chart centered at q_*.

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

$$\sqrt{n}(W_{dist} - dist_{\mathcal{M}}(\mathrm{q}_*, \mathrm{q}_0)^2) \xrightarrow{d} N_{d_{\mathcal{M}}}(\mathbf{0}, D_{dist}^{\top} \Sigma_{E,\mathrm{Log}_{\mathrm{q}_*},11} D_{dist}),$$

where D_{dist} *is the column vector representation of* $\mathrm{grad}_{\mathrm{q}_*}(dist(\cdot, \mathrm{q}_0)^2)$ *with respect to the orthonormal basis of* $T_{\mathrm{q}_*}\mathcal{M}$ *associated with the normal chart used to represent the asymptotic covariance of* $\tilde{\mathrm{q}}_E$ *as the matrix* $\Sigma_{E,\mathrm{Log}_{\mathrm{q}_*}}$. *In particular, when* q_0 *is close to* q_*, *then*

$$\sqrt{n}(W_{dist} - dist_{\mathcal{M}}(\mathrm{q}_*, \mathrm{q}_0)^2) \xrightarrow{d} N_{d_{\mathcal{M}}}(\mathbf{0}, 4[\mathrm{Log}_{\mathrm{q}_*}\mathrm{q}_0]^\top \Sigma_{E,\mathrm{Log}_{\mathrm{q}_*},11}[\mathrm{Log}_{\mathrm{q}_*}\mathrm{q}_0]).$$

Theorem 2.3.7 establishes the asymptotic distribution of W_{dist} when $\tilde{\mathrm{q}}_E$ and q_0 do not belong to the same chart of $\mathcal{M}$. In practice, the covariance matrix $\Sigma_{E,\mathrm{Log}_{\mathrm{q}_*},11}$ is not available, since $\Sigma_{E,\mathrm{Log}_{\mathrm{q}_*}}$ is not known; it also depends on the unknown true value $\boldsymbol{\beta}_*$, so we may use the estimate $\widehat{\Sigma}_{E,\mathrm{Log}_{\mathrm{q}_*}}$ as defined in Theorems 2.3.3 and 2.3.4 . Therefore, under the null hypothesis, the asymptotic distribution of W_{dist} can be approximated by the weighted chi-square distribution $\chi^2(\hat{\lambda}_1, \ldots, \hat{\lambda}_{d_{\mathcal{M}}})$, in which the weights $\hat{\lambda}_1, \ldots, \hat{\lambda}_{d_{\mathcal{M}}}$ are the eigenvalues of the covariance matrix $(\widehat{\Sigma}_{E,\mathrm{Log}_{\mathrm{q}_0}})_{11}/n$.

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 only consider the ILSE estimator of $(\mathrm{q}, \boldsymbol{\beta})$, denoted by $(\mathrm{q}_0, \tilde{\boldsymbol{\beta}}_I)$, under the *null* hypothesis $H_0^{(2)}$. For any chart (U, ϕ) on $\mathcal{M}$ with $\mathrm{q}_0 \in U$, we define

$$\begin{aligned} F_{\phi i} &= (F_{\phi i,1}^\top, F_{\phi i,2}^\top)^\top = \partial_{(\mathbf{t},\boldsymbol{\beta})}\mathrm{dist}_{\mathcal{M}}(f(\mathbf{x}_i, \phi^{-1}(\mathbf{t}), \boldsymbol{\beta}), y_i)^2\big|_{\mathbf{t}=\phi(\mathrm{q}_0),\tilde{\boldsymbol{\beta}}_I}, \\ \mathrm{U}_\phi &= \begin{pmatrix} \mathrm{U}_{\mathbf{tt}} & \mathrm{U}_{\mathbf{t}\boldsymbol{\beta}} \\ \mathrm{U}_{\boldsymbol{\beta}\mathbf{t}} & \mathrm{U}_{\boldsymbol{\beta}\beta} \end{pmatrix} = \sum_{i=1}^n \partial^2_{(\mathbf{t},\boldsymbol{\beta})}\mathrm{dist}_{\mathcal{M}}(f(\mathbf{x}_i, \phi^{-1}(\mathbf{t}), \boldsymbol{\beta}), y_i)^2\big|_{\mathbf{t}=\phi(\mathrm{q}_0),\tilde{\boldsymbol{\beta}}_I}, \end{aligned}$$

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 $W_{SC,\phi}$ reduces to

$$W_{SC,\phi} = (\sum_{i=1}^{n} F_{\phi i,1})^{\top} \tilde{\Sigma}_{\phi,\mathrm{q}}^{-1} (\sum_{i=1}^{n} F_{\phi i,1}), \tag{2.26}$$

where $\tilde{\Sigma}_{\phi,\mathrm{q}} = (I_{d_{\mathcal{M}}}, -\mathrm{U}_{\mathbf{t}\boldsymbol{\beta}}\mathrm{U}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1})[\sum_{i=1}^{n}(F_{\phi i} - \overline{F_{\phi}})^{\otimes 2}](I_{d_{\mathcal{M}}}, -\mathrm{U}_{\mathbf{t}\boldsymbol{\beta}}\mathrm{U}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1})^{\top}$, in which $\overline{F_{\phi}} = n^{-1}\sum_{i=1}^{n} F_{\phi i}$. Theoretically, we can establish the asymptotic distribution of $W_{SC,\phi}$ under the null hypothesis.

Theorem 2.3.8. *Assume that all conditions in Theorem 2.3.5 hold. We have the following results:*

(i) For any suitable local chart (U, ϕ), the score test statistic $W_{SC,\phi}$ is asymptotically distributed as $\chi^2_{d_{\mathcal{M}}}$ under the null hypothesis $H_0^{(2)}$.

(ii) Under $H_0^{(2)}$, for any other local chart (U, ϕ') with $\mathrm{q}_0 \in U$, we have

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

Theorem 2.3.8 establishes the asymptotic distribution of $W_{SC,\phi}$ for $\mathcal{M}-$valued data.

2.4 Examples

We investigate the intrinsic regression model for four specific RSS's and include several other examples in the supplementary document.

2.4.1 Symmetric Positive-definite Matrices

We review some basic facts about the geometric structure of $\mathrm{Sym}^{+}(k)$ (Schwartzman 2006, Lang 1999, Terras 1988, Fletcher et al. 2004b, Batchelor et al. 2005, Zhu et al. 2009a, Yuan et al. 2012, Osborne et al. 2013). Let $\mathrm{Sym}(k)$ be the set of $k \times k$ symmetric matrices with real entries, which is a topological linear space of dimension $k(k+1)/2$.

The $\mathrm{Sym}^+(k)$ is an open subset of $\mathrm{Sym}(k)$ and $T_{\mathrm{q}}\mathrm{Sym}^+(k)$ is a copy of $\mathrm{Sym}(k)$ for $\mathrm{q} \in \mathrm{Sym}^+(k)$. Let $\mathrm{q} = C_{\mathrm{q}} C_{\mathrm{q}}^\top$ be the Cholesky decomposition of q, where C_q is a lower triangular matrix with strictly positive diagonal entries. Then, for $\mathrm{q}, \mathrm{q}' \in \mathrm{Sym}^+(k)$, the map $(\mathrm{q}, \mathrm{q}') \to \mathrm{q} \circ \mathrm{q}' := C_{\mathrm{q}} \mathrm{q}' C_{\mathrm{q}}^\top$ induces a (non-commutative) Lie group structure on $\mathrm{Sym}^+(k)$, denoted by G. The unit element of G is the identity matrix I_k and the inverse of a matrix $\mathrm{q} \in G$ with respect to the operation on G is $\mathrm{q}^{\sim 1} = C_{\mathrm{q}}^{-1} C_{\mathrm{q}}^{-\top}$. The Lie group G can be entirely covered with a single chart. We also have $L_{\mathrm{q}}(\mathrm{q}') = C_{\mathrm{q}} \mathrm{q}' C_{\mathrm{q}}^\top$ and $L_{\mathrm{q}*}(A) = C_{\mathrm{q}} A C_{\mathrm{q}}^\top$ for $\mathrm{q}, \mathrm{q}' \in \mathrm{Sym}^+(k)$ and $A \in \mathrm{Sym}(k)$. The associated Lie algebra is $\mathfrak{sym}(k) = \mathrm{Sym}(k)$ with the bracket map being $[A_1, A_2] = A_1 A_2 - A_2 A_1$ for $A_1, A_2 \in \mathrm{Sym}(k)$. Let $\exp(\cdot)$ and $\log(\cdot)$ be, respectively, the matrix exponential and logarithm. The manifold exponential at I_k, Exp_{I_k}, is the matrix exponential $\exp(\cdot)$ and its inverse map is $\mathrm{Log}_{I_k} = \mathrm{Exp}_{I_k}^{-1} = \log(\cdot)$. For $A \in \mathrm{Sym}(k)$ and $\mathrm{q}' \in \mathrm{Sym}^+(k)$, we have

$$\begin{aligned} \mathrm{Exp}_{\mathrm{q}}(A) &= (L_{\mathrm{q}} \circ \mathrm{Exp}_{I_k} \circ L_{\mathrm{q}^{\sim 1}*})(A) = C_{\mathrm{q}} \exp(C_{\mathrm{q}}^{-1} A C_{\mathrm{q}}^{-\top}) C_{\mathrm{q}}^\top, \\ \mathrm{Log}_{\mathrm{q}}(\mathrm{q}') &= \mathrm{Exp}_{\mathrm{q}}^{-1}(\mathrm{q}') = C_{\mathrm{q}} \log(C_{\mathrm{q}}^{-1} \mathrm{q}' C_{\mathrm{q}}^{-\top}) C_{\mathrm{q}}^\top. \end{aligned}$$

We consider the trace norm $\|A\| = \sqrt{\mathrm{tr}(A^2)}$ on $Sym(k)$, identified as $T_{I_k}\mathrm{Sym}^+(k)$. This norm is actually the 2-norm of the in R^{k^2} of the vectorized form of the matrix. This allows to introduce the following metric on $\mathrm{Sym}^+(k)$

$$< A_1, A_2 >_{\mathrm{q}} := < L_{\mathrm{q}^{\sim 1}} * (A_1), L_{\mathrm{q}^{\sim 1}} * (A_2) >_{I_k} = \mathrm{tr}(A_1 \mathrm{q}^{-1} A_2 \mathrm{q}^{-1}),$$

for $A_1, A_2 \in T_{\mathrm{q}}\mathrm{Sym}^+(k)$ and $\mathrm{q} \in \mathrm{Sym}^+(k)$. This metric induces a Riemannian structure on the group $\mathrm{Sym}^+(k)$, and the above $\mathrm{Exp}_{\mathrm{q}}$ and $\mathrm{Log}_{\mathrm{q}}$ are the Riemannian exponential and logarithmic maps, respectively. The curve $t \to \gamma(t; \mathrm{q}, A) := \mathrm{Exp}_{\mathrm{q}}(tA)$ is the geodesic curve starting from q with initial tangent vector $A \in T_{\mathrm{q}}\mathrm{Sym}^+(k)$. The radius of injectivity is $\rho^*(\mathrm{Sym}(k)) = \rho^*(\mathrm{Sym}(k), I_k) = \infty$.

We introduce the intrinsic regression model for $\mathrm{Sym}^+(k)-$valued responses. Suppose that we observe $\{(y_i, \mathbf{x}_i) : i = 1, \ldots, n\}$, where $y_i \in \mathrm{Sym}^+(k)$ for all i. We define a function $f(\mathbf{x}, \boldsymbol{\beta})$ given by

$$f(\cdot, \cdot) : R^{d_{\mathbf{x}}} \times R^{d_\beta} \to R^{k(k+1)/2} \text{ with } f(\mathbf{0}, \cdot) = f(\cdot, \mathbf{0}) = \mathbf{0}.$$

An example of $f(\cdot, \cdot)$ is $f(\mathbf{x}_i, \boldsymbol{\beta}) = B\mathbf{x}_i$ (Zhu et al. 2009a), where B is an $k(k+1)/2 \times d_{\mathbf{x}}$ matrix of regression coefficients and $\boldsymbol{\beta}$ includes all components of B. Let $\{E_{j\ell} : 1 \leq \ell \leq j \leq k\}$ be the canonical basis of $\mathrm{Sym}(k)$, where $E_{j\ell}$ is the $m \times m$ matrix with the (j, ℓ) and (ℓ, j) entries being 1 and 0 otherwise; let $f(\mathbf{x}_i, \boldsymbol{\beta})_{j(j-1)/2+\ell}$ be the $j(j-1)/2 + \ell$-th component of $f(\mathbf{x}_i, \boldsymbol{\beta})$. We consider a single-center link function given by

$$\boldsymbol{\mu}(\mathbf{x}, \mathrm{q}, \boldsymbol{\beta}) = \mathrm{Exp}_{\mathrm{q}}(\mathbf{u}(\mathbf{x}_i, \boldsymbol{\beta})) = C_{\mathrm{q}} \exp(C_{\mathrm{q}}^{-1} \mathbf{u}(\mathbf{x}_i, \boldsymbol{\beta}) C_{\mathrm{q}}^{-\top}) C_{\mathrm{q}}^{\top},$$

where $\mathbf{u}(\mathbf{x}_i, \boldsymbol{\beta}) = \sum_{j=1}^{k} \sum_{\ell=1}^{j} f(\mathbf{x}_i, \boldsymbol{\beta})_{j(j-1)/2+\ell} E_{j\ell}$ and $\mathrm{q} = C_{\mathrm{q}} C_{\mathrm{q}}^{\top} \in \mathrm{Sym}^+(k)$ is the 'center'. The *rotated residual* is given by

$$\mathcal{E}(y_i, \mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta}) = \log(C_i(\mathrm{q}, \boldsymbol{\beta})^{-1} y_i C_i(\mathrm{q}, \boldsymbol{\beta})^{-\top}),$$

where $C_i(\mathrm{q}, \boldsymbol{\beta}) C_i(\mathrm{q}, \boldsymbol{\beta})^{\top}$ is the Cholesky decomposition of $f(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta})$.

2.4.2 Special Orthogonal Group $\mathrm{SO}(k)$

We review some basic facts about the geometric structure of $\mathrm{SO}(k)$ (Grenander et al. 1998, Moakher 2002, Gallier and Xu 2002). This is a compact (C^∞) submanifold of $R^{k \times k}$ of dimension $k(k-1)/2$ as well as a Lie group with respect to matrix multiplication. The unit element of $\mathrm{SO}(k)$ is the identity matrix I_k and its associated Lie algebra $\mathfrak{so}(k) = T_{I_k}\mathrm{SO}(k)$ is the linear space of all $k \times k$ skew-symmetric matrices q, i.e.

$\mathrm{q}^\top = -\mathrm{q}$, denoted by SkewSym($k$). For $\mathrm{q} \in \mathrm{SO}(k)$, $T_\mathrm{q}\mathrm{SO}(k)$ is given by

$$T_\mathrm{q}\mathrm{SO}(k) = \{A \in R^{k\times k} : A^\top = -\mathrm{q}^\top A \mathrm{q}^\top\} = \mathrm{q}\ \mathrm{SkewSym}(k).$$

We consider the trace metric on $T_\mathrm{q}\mathrm{SO}(k)$. The trace metric is also a left-invariant Riemannian metric on SO(k). Specifically, since $\mathrm{q}\,\mathrm{q}^\top = I_k$, for $A_1, A_2 \in T_{I_m}\mathrm{SO}(k)$, we have

$$< \mathrm{q}A_1, \mathrm{q}A_2 >_\mathrm{q} = \mathrm{tr}[(\mathrm{q}A_1)^\top(\mathrm{q}A_2)] = \mathrm{tr}(A_1^\top A_2) =< A_1, A_2 >_{I_k} .$$

The Lie exponential map at I_k is given by the usual matrix exponentiation. Although the Lie logarithmic map at I_k has a closed form, the formula for a general k is quite complicated. We present the Lie logarithmic map for $k = 2$ and 3 in the supplementary document. Generally, the Lie exponential map of $A \in T_\mathrm{q}\mathrm{SO}(k)$ at $\mathrm{q} \in \mathrm{SO}(k)$ and its corresponding Lie logarithmic map are, respectively, given by

$$\mathrm{Exp}_\mathrm{q}(A) = \mathrm{q}\,\mathrm{Exp}_{I_k}(\mathrm{q}^\top A) = \mathrm{q}\exp(\mathrm{q}^\top A) \text{ and } \mathrm{Log}_\mathrm{q}(\mathrm{q}') = \mathrm{q}\,\mathrm{Log}_{I_k}(\mathrm{q}^\top \mathrm{q}').$$

We introduce the intrinsic regression model for SO(k)−valued responses. Suppose that we observe $\{(y_i, \mathbf{x}_i) : i = 1, \ldots, n\}$, where $y_i \in \mathrm{SO}(k)$ for all i. We define a function $f(\mathbf{x}, \boldsymbol{\beta})$ given by

$$f(\cdot,\cdot) : R^{d_\mathbf{x}} \times R^{d_\beta} \to R^{k(k-1)/2} \text{ with } f(\mathbf{0},\cdot) = f(\cdot,\mathbf{0}) = \mathbf{0}.$$

An example of $f(\cdot,\cdot)$ is $f(\mathbf{x}_i, \boldsymbol{\beta}) = B_1\mathbf{x}_i$, where B_1 is a $k(k-1)/2 \times d_\mathbf{x}$ matrix of regression coefficients and $\boldsymbol{\beta}$ includes all components of B_1. Let $\{\tilde{E}_{j\ell} : 1 \le \ell \le j \le k\}$ be the basis of SkewSym(k), where $\tilde{E}_{j\ell}$ is a $k \times k$ matrix with the (j,ℓ) and (ℓ, j) entries being $(-1)^{j+\ell-1}$ and $(-1)^{j+\ell}$, respectively, and 0 otherwise. Let $\mathrm{q} \in \mathrm{SO}(k)$ be

the 'center', and we consider a single-center link function given by

$$\boldsymbol{\mu}(\mathbf{x}, \mathrm{q}, \boldsymbol{\beta}) = \mathrm{Exp}_{\mathrm{q}}(\mathbf{u}(\mathbf{x}_i, \boldsymbol{\beta})) = \mathrm{q}\exp(\mathbf{u}(\mathbf{x}_i, \boldsymbol{\beta})),$$

where $\mathbf{u}(\mathbf{x}_i, \boldsymbol{\beta}) = \sum_{j=2}^{k}\sum_{\ell=1}^{j-1} f(\mathbf{x}_i, \boldsymbol{\beta})_{(j-2)(j-1)/2+\ell}\tilde{E}_{j\ell} \in \mathrm{SkewSym}(k)$. The *rotated residual* is given by

$$\mathcal{E}(y_i, \mathbf{x}_i; \mathrm{q}, \boldsymbol{\beta}) = \mathrm{Log}_{I_k}(\exp(-\mathbf{u}(\mathbf{x}_i, \boldsymbol{\beta}))\mathrm{q}^\top y_i).$$

The explicit form of $\mathcal{E}(y_i, \mathbf{x}_i; \mathrm{q}, \boldsymbol{\beta})$ for $k = 2, 3$ can be found in the supplementary document.

2.4.3 The Unit Sphere S^k

We review some basic facts about the geometric structure of S^k in R^{k+1} (Shi et al. 2012, Mardia and Jupp 2000, Healy and Kim 1996, Huckemann et al. 2010). For $\mathrm{q} \in S^k$, $T_{\mathrm{q}}S^k$ is given by $T_{\mathrm{q}}S^k = \{\mathbf{v} \in R^{k+1} : \mathbf{v}^\top \mathrm{q} = 0\}$. For two points q and q′ in S^k with $\mathrm{q} \neq -\mathrm{q}'$, let $R_{\mathrm{q}',\mathrm{q}}(t) \in SO(k+1)$ be the rotation matrix in the subspace generated by q′ and q that rotates q′ towards q by the angle $t \in R$, if $\mathrm{q}' \neq \mathrm{q}$, and $R_{\mathrm{q}',\mathrm{q}}(t) = I$, $t \in R$, if $\mathrm{q}' = \mathrm{q}$. Here we use the usual convention that $R_{\mathrm{q}',\mathrm{q}}(t)$ rotates q′ towards q, if $t > 0$, and $R_{\mathrm{q}',\mathrm{q}}(t)$ rotates q′ away from q, if $t < 0$. The rotation $R_{\mathrm{q}',\mathrm{q}}(\cdot)$ takes the form

$$\begin{aligned} R_{\mathrm{q}',\mathrm{q}}(t)\mathbf{v} &= \mathbf{v} - (\mathbf{v}^\top \mathrm{q}')\mathrm{q}' - (\mathbf{v}^\top \tilde{\mathrm{q}})\tilde{\mathrm{q}} \\ &+ [(\mathbf{v}^\top \mathrm{q}')\cos t - (\mathbf{v}^\top \tilde{\mathrm{q}})\sin t]\mathrm{q}' + [(\mathbf{v}^\top \mathrm{q}')\sin t + (\mathbf{v}^\top \tilde{\mathrm{q}})\cos t]\tilde{\mathrm{q}} \end{aligned}$$

for $\mathbf{v} \in R^{k+1}$, where $\tilde{\mathrm{q}} = \frac{\mathrm{q} - (\mathrm{q}^\top \mathrm{q}')\mathrm{q}'}{\sqrt{1-(\mathrm{q}^\top \mathrm{q}')^2}}$. Thus, $(-\pi, \pi) \ni t \mapsto R_{\mathrm{q}',\mathrm{q}}(t) \cdot \mathrm{q}'$ is the unique geodesic curve in S^k joining q′ with q. More precisely, $R_{\mathrm{q}',\mathrm{q}}(0) \cdot \mathrm{q}' = \mathrm{q}'$ and $R_{\mathrm{q}',\mathrm{q}}(\arccos(\mathrm{q}^\top \mathrm{q}')) \cdot \mathrm{q}' = \mathrm{q}$. We consider standard multiplication as the inner

product. For $\mathbf{v} \in T_{\mathrm{q}}S^k$, the Riemannian Exponential map is given by $\mathrm{Exp}_{\mathrm{q}}(\mathbf{v}) = \cos(\|\mathbf{v}\|)\mathrm{q} + \sin(\|\mathbf{v}\|)\mathbf{v}/\|\mathbf{v}\|$. If q and q′ are not antipodal, the Riemannian Logarithmic map is given by $\mathrm{Log}_{\mathrm{q}}(\mathrm{q}') = \arccos(\mathrm{q}^T\mathrm{q}')\mathbf{v}/\|\mathbf{v}\|$, where $\mathbf{v} = \mathrm{q}' - (\mathrm{q}^T\mathrm{q}')\mathrm{q} \neq 0$.

We introduce the intrinsic regression model for S^k-valued responses. Suppose that we observe $\{(y_i, \mathbf{x}_i) : i = 1, \ldots, n\}$, where $y_i \in S^k$ for all i. We define a function $f(\mathbf{x}, \boldsymbol{\beta})$ given by

$$f(\cdot,\cdot) : R^{d_{\mathbf{x}}} \times R^{d_\beta} \to R^k \text{ with } f(\mathbf{0},\cdot) = f(\cdot,\mathbf{0}) = \mathbf{0}.$$

An 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. Without loss of generality, we fix the "north pole" $\mathrm{p} = (0, \ldots, 0, 1)^\top \in R^{k+1}$ as a base point. The $(k+1) \times 1$ vectors $\mathbf{e}_j$, with a 1 at the j-th component and a 0 otherwise, $j = 1, \ldots, k$, form an orthonormal basis in $T_{\mathrm{p}}S^k$. Let $\mathrm{q} \in S^k$ be the 'center', and we consider the following two examples of single-center link functions

$$\boldsymbol{\mu}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta}) = \mathrm{Exp}_{\mathrm{q}}\left(\sum_{j=1}^{k} f(\mathbf{x}_i, \boldsymbol{\beta})_j c_{\mathrm{p},\mathrm{q}}(\arccos(\mathrm{p}^\top\mathrm{q}))\mathbf{e}_j\right) \tag{2.27}$$

and

$$\boldsymbol{\mu}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta}) = c_{-\mathrm{p},\mathrm{q}}(\arccos(-\mathrm{p}^\top\mathrm{q}))(T^{-1}_{st,-\mathrm{p}}((f(\mathbf{x}_i, \boldsymbol{\beta})^\top, -1)^\top)), \tag{2.28}$$

where $T_{st,-\mathrm{p}}$ is the stereographic projection mapping from $S^k \backslash \{\mathrm{p}\}$ onto the d-dimensional hyperplane $R^k \times \{-1\}$. The *rotated residual* is given by

$$\begin{aligned}\mathcal{E}(y_i, \mathbf{x}_i; \mathrm{q}, \boldsymbol{\beta}) &= R_{\mathrm{p}_0, \boldsymbol{\mu}(\mathbf{x}_i,\mathrm{q},\boldsymbol{\beta})}(\arccos(\mathrm{p}_0^\top \boldsymbol{\mu}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta})))^{-1}(\mathrm{Log}_{\boldsymbol{\mu}(\mathbf{x}_i,\mathrm{q},\boldsymbol{\beta})} y_i) \\ &= \mathrm{Log}_{\mathrm{p}_0}(R_{\mathrm{p}_0, \boldsymbol{\mu}(\mathbf{x}_i,\mathrm{q},\boldsymbol{\beta})}(-\arccos(\mathrm{p}_0^\top \boldsymbol{\mu}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta}))) \cdot y_i). \end{aligned} \tag{2.29}$$

2.4.4 Kendall's Planar Shape Space Σ_2^k

We review the definition and some basic facts about the geometric structure of the shape space Σ_2^k formed by k landmarks in R^2, $k > 2$ (Kendall 1984a, Kendall et al. 1999, Dryden and Mardia 1998, Huckemann et al. 2010, Su et al. 2012). Geometrical planar objects are studied by placing $k > 2$ landmarks at specific locations of each object, usually on the boundary of the object. Then each object is described by a $k \times 2$ matrix $x \in R^{k\times 2}$, each row x^i denoting the coordinates of a point in R^2, $i = 1, \ldots, k$. It is often convenient to identify points in R^2 with complex numbers, i.e. $x^i \equiv z^i = x^{i,1} + jx^{i,2} \in C$, where $j = \sqrt{-1}$. In this representation, a configuration x of k landmarks is an element $z \in C^k$. We remove the translations by restricting to those elements of C^k whose average is zero, $\sum_{i=1}^k z^i = 0$, and the scale variability by rescaling the matrix to have norm one, $\|z\|^2 = \bar{z}^\top z = \sum_{i=1}^k z^i \bar{z^i} = 1$, where the "overline" denotes complex conjugation. Thus, we obtain a set $\mathcal{D}^k = \{z = (z^1, \ldots, z^k)^\top \in C^k \,|\, 1/n \sum_{i=1}^k z^i = 0, \|z\| = 1\}$ called the pre-shape space. Here, $\mathcal{D}^k$ is a unit sphere and we can utilize the geometry of a sphere to analyze points on it. Thus, $\mathcal{D}^k$ has the canonical structure of a real Riemannian manifold of real dimension $(2k-3)$, with the metric induced by the standard inner product on $R^{k\times 2}$ which is the real part of the complex inner product on C^k. The tangent space of $\mathcal{D}^k$ at a point z is $T_z\mathcal{D}^k = \{\mathbf{v} = (v^1, \ldots, v^k)^\top \in C^k \,|\, Re(\bar{z}^\top v) = 0,\ 1/k \sum_{i=1}^k v^i = 0\}$ and the geodesic distance on $\mathcal{D}^k$ is the spherical distance $d_{\mathcal{D}^k}(z, z') = \arccos(Re(\bar{z'}^\top z))$. The special unitary group $G = \mathrm{SU}(\mathcal{V}) \cong \mathrm{SU}(k-1) \subset \mathrm{SU}(k)$ acts transitively on $\mathcal{D}^k$, where $\mathcal{V}$ is the complex-orthogonal complement of $span_{C^k}\{(1, \ldots, 1)^\top\}$ in C^k. $\mathcal{V}$ has complex dimension $k-1$. $\mathrm{SU}(k-1)$ is a real Lie group of dimension $(k-1)^2 - 1$. The isotropy subgroup of z is $G_z \cong \mathrm{SU}(k-2)$. Thus, $\mathcal{D}^k$ is a Riemannian symmetric space.

To get the shape space we remove the planar rotations of pre-shapes. For $z \in \mathcal{D}^k$, let $[z]$ be the set of all planar rotations of a configuration z according to $[z] = \{z' = e^{j\theta} z | \theta \in S^1\}$. One defines an equivalence relation on $\mathcal{D}^k$ by setting all elements of the

set $[z]$ as equivalent, i.e. $z \sim z'$ if there is an angle θ such that $z' = e^{j\theta}z$. The set of all such equivalence classes is the quotient space $\mathcal{D}^k/S^1$. This space is called *Kendall's planar shape space* and is denoted by Σ_2^k. Since S^1 acts freely on $\mathcal{D}^k$, i.e. the only element of S^1 whose action has fixed points is the unit element of S^1, then the quotient space Σ_2^k is a $(2k-4)$-dimensional real Riemannian manifold. In fact, this space can be identified with a complex projective space CP^{k-2}. Since, $z \sim z'$ implies $Uz \sim Uz'$, for any $z, z' \in \mathcal{D}^k$ and any $U \in \mathrm{SU}(\mathcal{V})$, the group G acts transitively on Σ_e^k as well, and the isotropy subgroup is $G_{[z]} \cong \mathrm{SU}(k-2) \times S^1$. The natural Riemannian structure on Σ_2^k (as CP^{k-2}) is given by the Fubini-Study metric, which is defined as follows.

The tangent space of Σ_2^k at a point $\mathrm{q} = [z_\mathrm{q}]$, with $z_\mathrm{q} \in \mathcal{D}^k$, is

$$\begin{aligned} T_\mathrm{q}\Sigma_2^k &= \{\mathbf{v} = (v^1, \ldots, v^k)^\top \in C^k \,|\, Re\left(\overline{(e^{j\theta}z_\mathrm{q})}^\top \mathbf{v}\right) = 0,\, \theta \in S^1,\, 1/k \sum_{i=1}^k v^i = 0\} \\ &= \{\mathbf{v} = (v^1, \ldots, v^k)^\top \in C^k \,|\, \overline{z_\mathrm{q}}^\top \mathbf{v} = 0,\, 1/k \sum_{i=1}^k v^i = 0\}, \end{aligned}$$

and it is equipped with the complex inner product induced from C^k, that is, $< \mathbf{v}, \mathbf{w} >_\mathrm{q} := \overline{\mathbf{w}}^\top \mathbf{v}$, for $\mathbf{v}, \mathbf{w} \in T_\mathrm{q}\Sigma_2^k$, which is well-defined.

A geodesic between two elements $\mathrm{q}_1, \mathrm{q}_2 \in \Sigma_2^k$, with $\mathrm{q}_i = [z_{\mathrm{q}_i}]$, $i = 1, 2$, is given by a spherical geodesic on $\mathcal{D}^k$ between z_{q_1} and $z^*_{\mathrm{q}_2}$, where $z^*_{\mathrm{q}_2} = e^{j\theta^*} z_{\mathrm{q}_2}$ and θ^* is the optimal rotational alignment of z_{q_2} to z_{q_1} given by $\bar{z}_{\mathrm{q}_2}^\top z_{\mathrm{q}_1} = e^{j\theta^*} |\bar{z}_{\mathrm{q}_2}^\top z_{\mathrm{q}_1}|$. The geodesic distance on Σ_2^k between $\mathrm{q}_1, \mathrm{q}_2$, $d_{\Sigma_2^k}(\mathrm{q}_1, \mathrm{q}_2)$, is the spherical distance $d_{\mathcal{D}_m^k}(z_{\mathrm{q}_1}, z^*_{\mathrm{q}_2}) = \arccos(\overline{z^*_{\mathrm{q}_2}}^\top z_{\mathrm{q}_1}) = \arccos(|\overline{z_{\mathrm{q}_2}}^\top z_{\mathrm{q}_1}|)$. The definitions of both the geodesics and geodesic distance are independent of the choice of representatives for the equivalence classes q_1 and q_2. For $\mathbf{v} \in T_\mathrm{q}\Sigma_2^k$, the Riemannian Exponential map is given by $\mathrm{Exp}_\mathrm{q}(\mathbf{v}) = \cos(\|\mathbf{v}\|) z_q + \sin(\|\mathbf{v}\|) \frac{\mathbf{v}}{\|\mathbf{v}\|}$. The exponential map is well-defined and it is a bijection on the set of $[(z_\mathrm{q}, \mathbf{v})]$ so that $\|\mathbf{v}\| \in [0, \frac{\pi}{2})$. The Riemannian Logarithmic map is given by $\mathrm{Log}_\mathrm{q}(\mathrm{q}') = \arccos(|\overline{z_{\mathrm{q}'}}^\top z_\mathrm{q}|)\mathbf{v}/\|\mathbf{v}\| = \frac{r}{\sin(r)}\mathbf{v}$, where $\mathbf{v} = z^*_{\mathrm{q}'} - |\overline{z_{\mathrm{q}'}}^\top z_\mathrm{q}| z_\mathrm{q}$, $r = d_{\Sigma_2^k}(\mathrm{q}, \mathrm{q}')$,

and $z^*_{\mathrm{q}'}$ is the optimal alignment of $z_{\mathrm{q}'}$ to z_{q}. It is easy to check that all the definitions above are independent of the choice of representatives for the corresponding equivalence classes.

Note that with respect to a chosen complex orthonormal basis $\{Z_1, \ldots, Z_{k-2}\}$ for $T_{\mathrm{p}}\Sigma_2^k$, the normal chart ϕ centered at p has the expression

$$\begin{aligned} \phi(\mathrm{q}) &= \boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_{k-2})^\top \in C^{k-2}, \\ &\cong \mathbf{t} = (t_1, \ldots, t_{2k-4})^\top \in R^{2k-4} \end{aligned} \tag{2.30}$$

where $\zeta_\ell = t_{2\ell-1} + jt_{2\ell}$ and

$$\zeta_\ell = \frac{r}{\sin(r)} e^{j\theta}\, \overline{Z_\ell}^\top z_{\mathrm{q}},\ r = d_{\Sigma_2^k}(\mathrm{q}, \mathrm{p}) = \arccos(|\overline{z_{\mathrm{q}}}^\top z_{\mathrm{p}}|),\ e^{j\theta} = \frac{\overline{z_{\mathrm{q}}}^\top z_{\mathrm{p}}}{|\overline{z_{\mathrm{q}}}^\top z_{\mathrm{p}}|}, \tag{2.31}$$

$\ell = 1, \ldots, k-2$, (see Bhattacharya and Bhattacharya (2008)).

We introduce the intrinsic regression model for Σ_2^k-valued responses. Suppose that we observe $\{(y_i, \mathbf{x}_i) : i = 1, \ldots, n\}$, where $y_i \in \Sigma_2^k$ and $\mathbf{x}_i \in R^{d_{\mathbf{x}}}$, for all i. We define a function $f(\mathbf{x}, \boldsymbol{\beta})$ given by

$$f(\cdot, \cdot) : R^{d_{\mathbf{x}}} \times R^{d_\beta} \to R^{2k-4} \text{ with } f(\mathbf{0}, \cdot) = f(\cdot, \mathbf{0}) = \mathbf{0}. \tag{2.32}$$

An example of $f(\cdot, \cdot)$ is $f(\mathbf{x}_i, \boldsymbol{\beta}) = B\mathbf{x}_i$, where B is a $(2k-4) \times d_{\mathbf{x}}$ matrix of regression coefficients and $\boldsymbol{\beta}$ includes all components of B. We fix a point $\mathrm{p}_0 \in \Sigma_2^k$, as the base point, and $\{Z_1, \ldots, Z_{k-2}\}$ an orthonormal basis for $T_{\mathrm{p}_0}\Sigma_2^k$. For example, set $\mathrm{p}_0 = [z_0]$, with $z_0 = (1/\sqrt{2}, -1/\sqrt{2}, 0, \ldots, 0)^\top \in R^n \subset C^n$, and $Z_\ell \in R^n$ be the vector with $\frac{1}{\sqrt{(\ell+1)(\ell+2)}}$ as the first $(\ell+1)$ components, with $\frac{\ell+1}{\sqrt{(\ell+1)(\ell+2)}}$ as the $(\ell+2)$-th component, and 0 otherwise, for $\ell = 1, \ldots, k-2$. Thus, $\{Z_1, \ldots, Z_{k-2}\}$ forms a complex orthonormal basis of $T_{\mathrm{p}_0}\Sigma_2^k$ when viewed as a complex linear space, and, equivalently,

$\{Z_1, jZ_1, \ldots, Z_{k-2}, jZ_{k-2}\}$ forms a real orthonormal basis of $T_{\mathrm{p}_0}\Sigma_2^k$ when viewed as a real linear space. Letting $\mathrm{q} \in \Sigma_2^k$ be the 'center', we consider a single-center link function given by

$$\boldsymbol{\mu}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta}) = \mathrm{Exp}_{\mathrm{q}}(\mathbf{u}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta})) \in \Sigma_2^k, \tag{2.33}$$

where $\mathbf{u}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta}) \in T_{\mathrm{q}}\Sigma_2^k$. An example of $\mathbf{u}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta})$ is given by

$$\mathbf{u}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta}) = \left[\sum_{\ell=1}^{k-2}(f(\mathbf{x}_i, \boldsymbol{\beta})_{2\ell-1} + jf(\mathbf{x}_i, \boldsymbol{\beta})_{2\ell})U_{z_{\mathrm{p}_0}, z_{\mathrm{q}}^*}Z_\ell\right] \in T_{\mathrm{q}}\Sigma_2^k, \tag{2.34}$$

where $\mathrm{p}_0 = [z_{\mathrm{p}_0}]$, $\mathrm{q} = [z_{\mathrm{q}}]$, with $z_{\mathrm{p}_0}, z_{\mathrm{q}} \in \mathcal{D}^k$. Here, for $z_1, z_2 \in \mathcal{D}^k$, $U_{z_1,z_2} \in \mathrm{SU}(\mathcal{V}) \subset SU(k)$ denotes the unique special unitary map in the subspace generated by z_1 and z_2 that maps z_1 onto z_2. The map U_{z_1,z_2} takes the form

$$\begin{aligned} U_{z_1,z_2}\mathbf{v} &= \mathbf{v} - (\overline{z_1}^\top\mathbf{v})z_1 - (\overline{\widetilde{z}_2}^\top\mathbf{v})\widetilde{z}_2 \\ &\quad + \left((\overline{z_1}^\top z_2)(\overline{z_1}^\top\mathbf{v}) - \sqrt{1-|\overline{z_1}^\top z_2|^2}\,(\overline{\widetilde{z}_2}^\top\mathbf{v})\right)z_1 \\ &\quad + \left(\sqrt{1-|\overline{z_1}^\top z_2|^2}\,(\overline{z_1}^\top\mathbf{v}) + \overline{(\overline{z_1}^\top z_2)}(\overline{\widetilde{z}_2}^\top\mathbf{v})\right)\widetilde{z}_2, \end{aligned} \tag{2.35}$$

for $\mathbf{v} \in C^k$, where $\widetilde{z}_2 = \frac{z_2 - (\overline{z_1}^\top z_2)z_1}{\sqrt{1-|\overline{z_1}^\top z_2|^2}}$. Thus, $U_{\mathrm{q}_1,\mathrm{q}_2}\mathbf{v} := U_{z_{\mathrm{q}_1}, z_{\mathrm{q}_2}^*}\mathbf{v} \in \mathcal{V}$, $\mathbf{v} \in \mathcal{V}$, and $U_{\mathrm{q}_1,\mathrm{q}_2}\mathrm{q} := [U_{z_{\mathrm{q}_1}, z_{\mathrm{q}_2}^*}z_{\mathrm{q}}] \in \Sigma_2^k$, $\mathrm{q} \in \Sigma_2^k$, are well defined, independently of the choice of representatives $z_{\mathrm{q}_1}, z_{\mathrm{q}_2}$, and z_{q} for $\mathrm{q}_1, \mathrm{q}_2$, and q, respectively. The *rotated residual* is given by

$$\mathcal{E}(y_i, \mathbf{x}_i; \mathrm{q}, \boldsymbol{\beta}) = U^{-1}_{\mathrm{p}_0, \boldsymbol{\mu}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta})}(\mathrm{Log}_{\boldsymbol{\mu}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta})}y_i) = \mathrm{Log}_{\mathrm{p}_0}(\overline{U_{\mathrm{p}_0, \boldsymbol{\mu}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta})}}^\top y_i). \tag{2.36}$$

We consider the *intrinsic* model

$$E(\mathcal{E}(y_i, \mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta}) \,|\, \mathbf{x}_i) = 0, \quad i = 1, \ldots, n. \tag{2.37}$$

2.5 Simulation Studies

We generated $(y_i, \mathbf{x}_i) \in S^2 \times R$ as follows. We generate directional data in S^2 according to (2.29) and (2.28) given at the end of subsection 2.4.3. Moreover, we set $f(\mathbf{x}_i, \boldsymbol{\beta}) = B\mathbf{x}_i$, in which $\mathbf{x}_i \in R$ and $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top \in R^2$, and $\mathrm{q} = (u, v, w) \in S^2 \setminus \{\mathrm{p}_0\}$. We independently simulated the x_i's from the standard normal distribution $N(0, 1)$ and fixed the true values of $(\mathrm{q}, \boldsymbol{\beta})$ to be $(\mathrm{q}_*, \boldsymbol{\beta}_*) = ((2/3, 2/3, 1/3)^\top, (1, 1)^\top)$. The parameterization is given by $t_1 = u/(1-w)$ and $t_2 = v/(1-w)$, and thus q_* corresponds to $\mathbf{t}_* = (t_{1*}, t_{2*}) = \mathbf{1}_2 = (1, 1)^\top$. We generated residuals $\mathcal{E}_i(y_i, \mathbf{x}_i; \mathrm{q}, \boldsymbol{\beta})$ from a $N_2(\mathbf{0}, 0.5\{\rho_1 \mathbf{1}_2 \mathbf{1}_2^\top + (1-\rho_1) I_2\})$ distribution on the tangent space, $T_{\mathrm{p}_0} S^2$, at the north pole $\mathrm{p}_0 = (0, 0, 1)^\top$, rotated the residuals onto $T_{\boldsymbol{\mu}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta})} S^2$, and then used the exponential map $\mathrm{Exp}_{\boldsymbol{\mu}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta})}(\cdot)$ to calculate the response y_i. We set $n = 40$, 80, and 120 in order to examine the finite sample performance of the parameter estimates and their covariance estimates.

We first compared the biases and the root-mean-square errors of the two estimates: $(\hat{\mathbf{t}}_I, \hat{\boldsymbol{\beta}}_I)$ and $(\hat{\mathbf{t}}_E, \hat{\boldsymbol{\beta}}_E)$. As seen in Table 2.1, $\hat{\mathbf{t}}_E$ and $\hat{\boldsymbol{\beta}}_E$ have smaller root-mean-square errors than $\hat{\mathbf{t}}_I$ and $\hat{\boldsymbol{\beta}}_I$ for every component of $\mathbf{t}$ and $\boldsymbol{\beta}$, respectively, confirming that $(\hat{\mathbf{t}}_E, \hat{\boldsymbol{\beta}}_E)$ is more efficient. We calculated the biases, root-mean-square error, and means of the estimated standard error estimates of $\hat{\beta}_E$ (Table 2.2). All relative efficiencies (the ratio of root-mean-square error to standard deviation) are close to 1.0, indicating that our covariance estimates are pretty accurate. As expected, the root-mean-square error and relative efficiency improve as the sample size increases. We also note that for β_1 and β_2, $\hat{\boldsymbol{\beta}}_E$ is more biased than $\hat{\boldsymbol{\beta}}_I$.

Table 2.1: Bias ($\times 10^{-3}$) and MS($\times 10^{-2}$) of $(\hat{\mathbf{t}}_I, \hat{\boldsymbol{\beta}}_I)$ and $(\hat{\mathbf{t}}_E, \hat{\boldsymbol{\beta}}_E)$. Bias denotes the bias of the mean of the estimates; MS denotes the root-mean-square error.

		$n = 40$		$n = 80$		$n = 120$	
		Bias	MS	Bias	MS	Bias	MS
I	t_1	5.29	13.91	5.86	9.18	6.24	7.83
E		4.67	12.97	5.10	9.09	6.22	7.20
I	β_1	5.86	18.39	5.98	12.45	4.16	10.05
E		7.40	15.39	6.16	11.46	5.42	9.05
I	t_2	4.35	13.73	6.48	8.97	3.64	8.62
E		4.21	12.84	4.36	8.56	3.05	8.04
I	β_2	6.94	18.64	4.49	12.09	5.06	10.04
E		7.35	14.52	4.74	10.85	5.64	9.98

Table 2.2: Bias ($\times 10^{-3}$), MS($\times 10^{-2}$), SD($\times 10^{-2}$), and RE of $(\hat{\mathbf{t}}_E, \hat{\boldsymbol{\beta}}_E)$. Bias denotes the bias of the mean of the estimates; MS denotes the root-mean-square error; SD denotes the mean of the standard deviation estimates; RE denotes the relative efficiency, which is the ratio of MS over SD.

	$n = 40$				$n = 80$				$n = 120$			
	Bias	MS	SD	RE	Bias	MS	SD	RE	Bias	MS	SD	RE
t_1	4.67	12.97	14.36	0.90	5.10	9.09	9.59	0.95	6.22	7.20	7.55	0.95
β_1	7.40	15.39	17.69	0.87	6.16	11.46	12.58	0.91	5.42	9.05	9.33	0.97
t_2	4.21	12.84	14.44	0.89	4.36	8.56	9.32	0.92	3.05	8.04	7.94	1.01
β_2	7.35	14.52	17.59	0.83	4.74	10.85	12.54	0.87	5.64	9.98	10.32	0.97

Table 2.3: Comparisons of the rejection rates for Wald test statistics. Three different sample sizes $n \in \{40, 80, 120\}$ and 2000 simulated datasets were used for each case and two significance levels, 5% and 1%, were considered.

	$n = 40$		$n = 80$		$n = 120$	
	5%	1%	5%	1%	5%	1%
1	0.062	0.019	0.056	0.015	0.054	0.014
1.2	0.251	0.101	0.430	0.218	0.599	0.364
1.4	0.644	0.413	0.898	0.716	0.977	0.927
1.6	0.842	0.680	0.994	0.952	0.998	0.994
1.8	0.965	0.899	0.998	0.994	1.000	1.000

To examine the finite sample performance of the Wald statistic $W^{(1)}_{n,\phi}$, we used the same setup except that we varied the value of β_2. To assess the Type I and II error rates for W_n, we tested the following hypotheses

$$H_0 : \beta_2 = 1 \text{ and } H_1 : \beta_2 \neq 1.$$

We set β_2 at 1.0, 1.2, 1.4, 1.6, and 1.8, respectively. Then, we set $n = 40$, 80, and 120 and simulated 2000 datasets for each case. The Wald statistic $W^{(1)}_{n,\phi}$ performs reasonably well for relatively small sample sizes (Table 2.3). The Type I error rates are not too excessive even for both the 5% and 1% significance levels at $n = 40$. Increasing the sample size can increase the statistical power in rejecting the null hypothesis.

Note: In order to demonstrate the performance of our method, we compared it with more naïve approaches that ignore nonlinearity or deal with it in a simple minded fashion. One example of such an approach is to compute a Fréchet mean of the observed response variable, map the responses to the tangent space at this mean, using the the logarithmic map, then apply a standard (Euclidean) regression framework in the tangent space, and then map the results back on the manifold via the exponential map. In Appendix, we provided an example of simulated data which illustrate that our method outperforms the naïve method in terms of both prediction accuracy and

estimation efficiency, when the distribution of the responses strongly depends on the covariates and the covariates have substantial variation.

2.6 Real Data Example - ADNI Corpus Callosum Shape Data

2.6.1 The ADNI Data

Alzheimer's disease (AD) is a disorder of cognitive and behavioral 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. In most people with Alzheimer's, symptoms first appear after age 60. AD affects almost 50% of those over the age of 85 and is the sixth leading cause of death in the US.

The corpus callosum (CC), 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 of high interest in many neuro-imaging studies.of neuro-devlopmental pathology. Individual differences in CC, and their possible implications regarding interhemispheric connectivity, have been investigated in last several decades (Witelson 1985, Paul et al. 2007). There is a substantial work suggesting that CC plays an important role in neurological normal individuals and its integrity in clinical populations diagnosed with AD (Preti et al. 2012, Di Paola et al. 2010, Lebel et al. 2010).

We consider the CC contour data of ADNI dataset[1] "ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute

[1] *http://adni.loni.usc.edu.* Data used in preparation of this thesis were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the work presented here. 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.*

of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profitt organizations, as a $60 million, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California, San Francisco. ADNI is the result of efforts of many coinvestigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in the protocols for ADNI-1, ADNI-2 and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see www.adni-info.org."

The CC shape data were processed for each subject in ADNI Dataset as follows. We used *FreeSurfer* package[2] (Dale et al. 1999) to process each subject to segment the T1-weighted MRI. Then the intracranial volume (ICV) information was calculated from the output of *FreeSurfer* package, while the midsagittal CC area was calculated in the *CCseg* package, which is measured by using subdivisions in Witelson (1989)

[2]http://surfer.nmr.mgh.harvard.edu/

Table 2.4: Demographic information about processed ADNI CC shape dataset, including disease status, age, and gender.

Disease status	Num.	Range of age in years(mean)	Gender (female/male)
Normal Healthy Control	223	62-90 (76.25)	107/116
AD	186	55-92 (75.42)	88/98

motivated by neuro-histological studies. Finally, each T1-weighted MRI image and tissue segmentation were used as the input files of *CCSeg* package to extract the planar CC shape data, which contains 50 landmarks.

Our problem of interest is to investigate the association between the shape of the CC contours and Euclidean covariates, including gender, age, and AD diagnosis.

2.6.2 Intrinsic Regression Model

We processed the CC shape data for each of the 409 subject in the ADNI dataset as follows. Each planar CC contour data contain 50 landmarks. The demographic information about the processed CC shape data set is presented in Table 2.4.

We are interested in whether the CC shape information is a promising bio-marker for the diagnosis of AD and may provide a clue to the topological spread of disease. We investigate that by using regression models for the Kendall's planar shape space, described in Section 2.4.4. The dataset we work with contains one observation for each of the $n = 409$ subjects. Each observation consists of a CC planar contour Y_i with 50 landmarks, as the response variable, and a vector $(x_{i,1}, x_{i,2}, x_{i,3})^\top$ of three covariates: *gender* ($x_{i,1} = 0$ -female, 1 - male), *age* ($x_{i,2}$), and *diagnosis* ($x_{i,3} = 0$ -normal, 1 - AD), $i = 1, \ldots, 409$. For computational reason, we pre-processed the responses to reduce the number of landmarks, so the dimensionality of the Kendall's planar shape is reduced too. We subsampled the CC contours by keeping every other landmark and those with the top 20 variance and/or top 25 curvature, which resulted in k=32 landmarks. Each CC contour Y_i is specified as a $k \times 2$ real matrix, each row representing the planar

coordinates of a landmark in R^2 on the CC contour for the i-th subject. Each such Y_i can be regarded as a complex vector Z_i in C^{32}, via the standard identification $R^2 \cong C$, $Z_i = Y_{i,1} + jY_{i,2}$. After removing the translations and normalizing to the unit 2-norm, each CC contour Y_i can be view as an element $z_i \in \mathcal{D}^{32}$, and after removing the $2D$-rotations, as an element $y_i \in \Sigma_2^{32} \cong CP^{30}$, $y_i = [z_i]$ (see Section 2.4.4). Note that here $k = 32$. Thus, $dim_R \Sigma_2^{32} = 60$ and $dim_C \Sigma_2^{32} = 30$.

We consider a model like (2.37) with *gender, age, diagnosis,* and the *age*diagnosis* interaction as covariates. First, we standardized the covariates x_1, x_2, and x_3, and set $x_{i,4} = x_{i,2} * x_{i,3}$, and $\mathbf{x}_i = (x_{i,1}, x_{i,2}, x_{i,3}, x_{i,4})^\top$, $i = 1, \ldots, n$. Here $d_{\mathbf{x}} = 4$. Initially, we started with the base point $\mathrm{p}_0 = [z_0]$ and the orthonormal basis $\{Z_1, \ldots, Z_{30}\}$ for $T_{\mathrm{p}_0}\Sigma_2^{32}$ as defined in Section 2.4.4, below Eq. (2.32). As the parameter estimates depends on the choice of the base point, later, we redefined it "closer" to the data. First, we projected the responses y_i's onto the tangent space $T_{\mathrm{p}_0}\Sigma_2^{32}$ via the logarithmic map $\mathrm{Log}_{\mathrm{p}_0}$. Then, in the tangent space, we computed the mean of those projections, and we projected it back onto the manifold Σ_2^{32} via $\mathrm{Exp}_{\mathrm{p}_0}$. This new point is used to redefine the base point p_0. As for the orthonormal basis for the tangent space at the new base point, we just rotate the initial basis to the new location via the tangent map of the $2D$-unitary map $U_{\mathrm{p}_{0,initial},\mathrm{p}_{0,new}}$ (see Eq. (2.35)), which is in fact the parallel transport of the basis at the initial base point along the geodesic path from the initial base point to the the new one. The model parameters are $(\mathrm{q}, \boldsymbol{\beta}) \in \Sigma_2^{32} \times R^{240}$ (here, $d_{\boldsymbol{\beta}} = 240$). The intercept q is specified by $\mathrm{q} = \phi_{\mathrm{p}_0}^{-1}(\mathbf{t}) := \mathrm{Exp}_{\mathrm{p}_0}\left(\sum_{\ell=1}^{30}(t_{2\ell-1} + jt_{2\ell})Z_\ell\right)$, $\mathbf{t} = (t_1, \ldots, t_{60})^\top \in R^{60}$.This ϕ_{p_0} is the normal chart centered at p_0 defined by the (complex) orthonormal basis $\{Z_1, \ldots, Z_{30}\}$ for $T_{\mathrm{p}_0}\Sigma_2^{32}$, as defined in (2.30) and (2.31). The regression coefficients $\boldsymbol{\beta} = (\boldsymbol{\beta}^{(g)\top}, \boldsymbol{\beta}^{(a)\top}, \boldsymbol{\beta}^{(d)\top}, \boldsymbol{\beta}^{(ad)\top})^\top$, with $\boldsymbol{\beta}^{(g)} = (\beta_1, \ldots, \beta_{60})^\top$ corresponding to the gender covariate, $\boldsymbol{\beta}^{(a)} = (\beta_{61}, \ldots, \beta_{120})^\top$ corresponding to the age, $\boldsymbol{\beta}^{(d)} = (\beta_{121}, \ldots, \beta_{180})^\top$ corresponding to the diagnosis, and $\boldsymbol{\beta}^{(ad)} = (\beta_{181}, \ldots, \beta_{240})^\top$

corresponding to the interaction age*diagnosis. Actually, the model has 300 parameters to be estimated, namely, $(\mathbf{t}^\top, \boldsymbol{\beta}^\top)^\top \in R^{300}$. We denoted

$$B = \begin{pmatrix} \beta_1 + j\beta_2 & \beta_{61} + j\beta_{62} & \beta_{121} + j\beta_{122} & \beta_{181} + j\beta_{182} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{59} + j\beta_{60} & \beta_{119} + j\beta_{120} & \beta_{179} + j\beta_{180} & \beta_{239} + j\beta_{240} \end{pmatrix} \in C^{30\times 4}.$$

Each column of B corresponds to a covariate. The function in (2.32) becomes $f(\mathbf{x}_i, \boldsymbol{\beta}) = B\mathbf{x}_i$, and the single-center link function (2.33) is now

$$\boldsymbol{\mu}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta}) = \mathrm{Exp}_{\mathrm{q}}\left([U_{\mathrm{p_0,q}} Z_1, \ldots, U_{\mathrm{p_0,q}} Z_{30}]\, B\mathbf{x}_i\right) \in \Sigma_2^{32}.$$

The intrinsic regression model (2.37) is

$$E\left[\mathrm{Log}_{\mathrm{p_0}}\left(\overline{U_{\mathrm{p_0},\boldsymbol{\mu}(\mathbf{x}_i,\mathrm{q},\boldsymbol{\beta})}}^{\top} y_i\right) \,\middle|\, \mathbf{x}_i\right] = \mathbf{0}, \quad i = 1, \ldots, 409. \tag{2.38}$$

2.6.3 Results

Note that Σ_2^{32} is a compact manifold of diameter π. The response data, Y_i, $i = 1, \ldots, 409$, lie in an open set of Σ_2^{32} of diameter less than 0.25. The two farthest away responses are the CC contours of subjects 130 and 382, a 60-year old AD male and a 77-year old normal healthy man, respectively. In Stage 1, we calculated the ILSE, $(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I) = (\phi_{\mathrm{p_0}}^{-1}(\hat{\mathbf{t}}_I), \hat{\boldsymbol{\beta}}_I)$, with respect to the normal chart $\phi_{\mathrm{p_0}}$, as defined in Eq. (2.11). Then, we used it in Stage2, to find the efficient estimator $(\tilde{\mathrm{q}}_E, \tilde{\boldsymbol{\beta}}_E) = (\phi_{\mathrm{p_0}}^{-1}(\tilde{\mathbf{t}}_E), \tilde{\boldsymbol{\beta}}_E)$ defined in Section 2.3.3, Eq. (2.15).

The intercept estimates $\hat{\mathrm{q}}_I$ and $\tilde{\mathrm{q}}_E$ are estimates of the sample Fréchet mean of Y_i's. The two estimates are very close to each other, with $dist(\hat{\mathrm{q}}_I, \tilde{\mathrm{q}}_E) < 0.0005$. The responses are within a distance less than 0.145 from the sample Fréchet mean.

The estimates $\hat{\boldsymbol{\beta}}_I$ and $\tilde{\boldsymbol{\beta}}_E$ of regression coefficients and their standard deviations are displayed in Figure 2.2. The efficiency gain in Stage II is measured by the relative reduction in the variances-covariances of $\boldsymbol{\beta}_E$ relative to the to the variances of $\boldsymbol{\beta}_I$, which is shown in Figure 2.3. There is an overall average variance relative reduction of about 16.77%. For the $age * diagnosis$ interaction coefficients, $\boldsymbol{\beta}^{(ad)}$, the average variance relative reduction of about 12.25%, while for the gender coefficients $\boldsymbol{\beta}^{(g)}$ about 19.98% variance relative reduction.

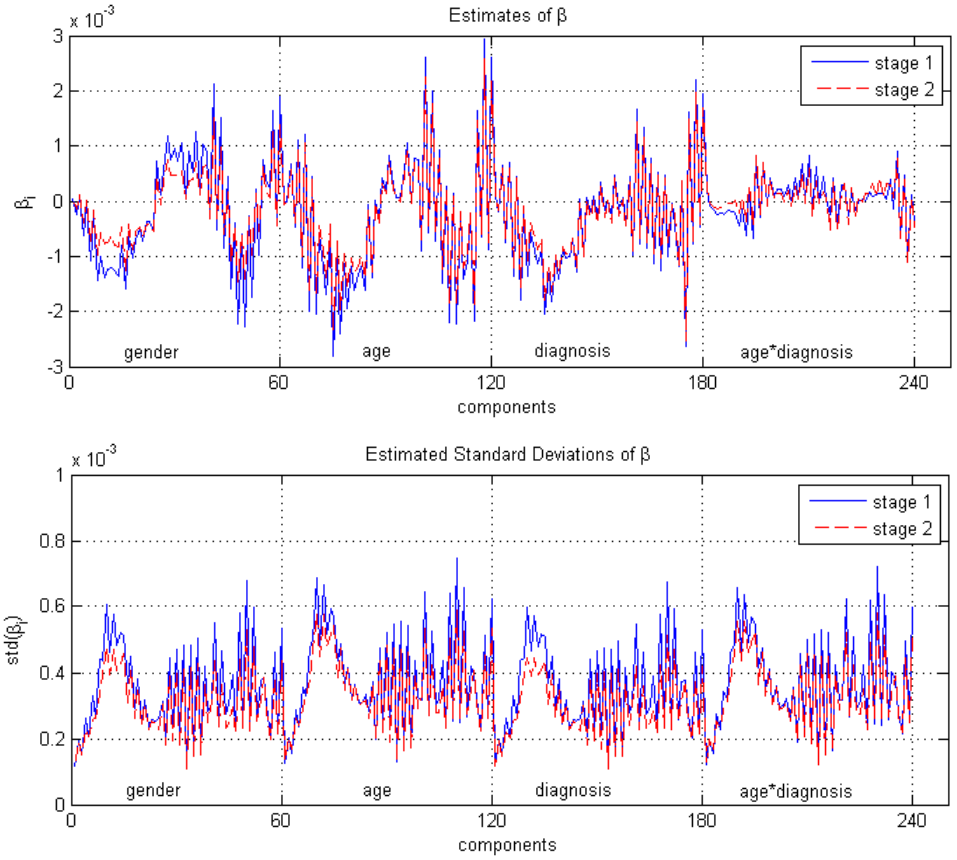

Figure 2.2: Regression coefficient estimates and their standard deviations from stage I and stage II.

We have visually checked the estimated rotated residuals in order to ensure the goodness of fit of our model to the real dataset. The rotated residuals are shown in Figures 2.4 and 2.5. Visually, we cannot find any outliers in the dataset. However,

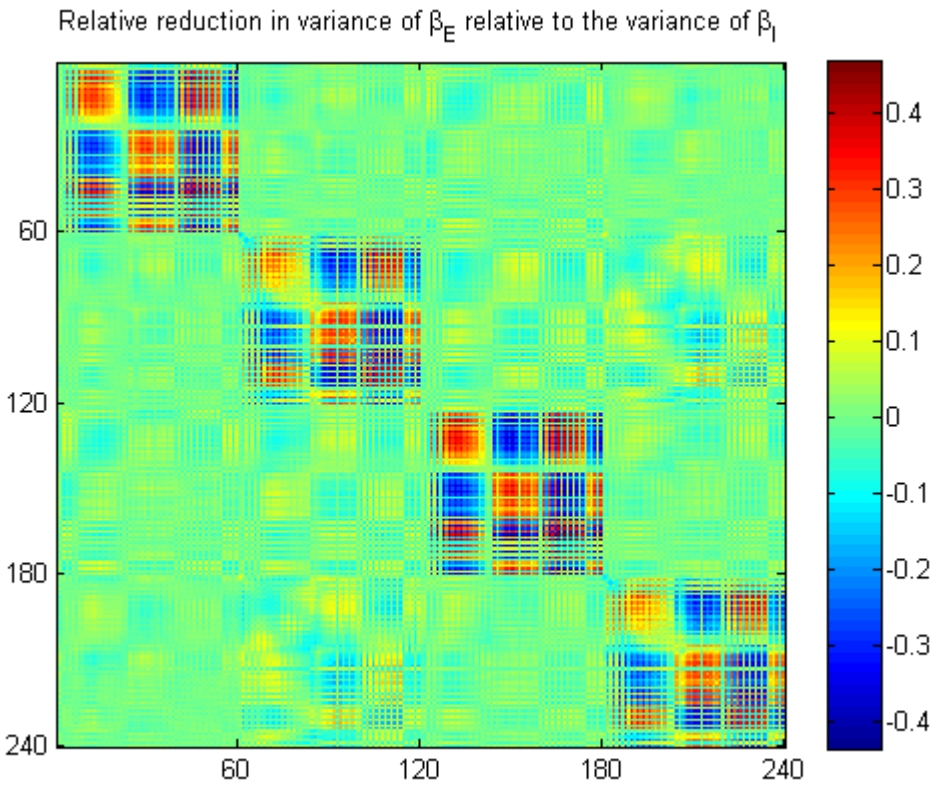

Figure 2.3: The relative reduction in the variances-covariances of $\boldsymbol{\beta}_E$ relative to the to the variances of $\boldsymbol{\beta}_I$. In average, there is about a 16.77% relative decrease in variances in average; 12.25% for $\boldsymbol{\beta}^{(ad)}$ and 19.98% for $\boldsymbol{\beta}^{(g)}$.

it is very interesting to develop some goodness-of-fit statistics. Technically, diagnostic measures, such as residuals, to identify potential outliers and high leverage points, and some residual processes to formally test whether the conditional mean assumption (2.6) is valid can be developed (Zhu et al. 2009b; 2008). A such work will be part of a separate paper.

To assess whether there is or not an age dependent diagnosis effect and/or gender effect on the shape of the CC contour, we performed two sets of hypothesis testings. First, we tested the *null* hypothesis $H_0 : \boldsymbol{\beta}^{(ad)} = \mathbf{0}_{60}$ versus the *alternative* $H_1 : \boldsymbol{\beta}^{(ad)} \neq \mathbf{0}_{60}$ in both stages I and II. The Wald test statistics are $W_I^{(ad)} = 130.00$ and

$W_E^{(ad)} = 98.20$, in stage I and II, respectively, and they are χ^2_{60} distributed. The p-values are < 0.0001 and 0.001, in stage I and II, respectively. Thus, the test statistics are significant at both 0.05 and 0.01 levels of significance, in both stages I and II. The data contains enough evidence to reject the *null*, that is, to reject that there is no age dependent diagnosis effect on the shape of the CC contours. The mean age-trajectories for normal and AD groups within each gender group, based on the stage II estimates, are shown in Figure. 2.6. A similar graph yields using stage I estimates (not displayed here). It can be observed that there is a difference in shape along the inner side of the posterior splenium and isthmus subregions, for subjects age 50-70 in both male and female groups. The splenium seems to be less rounded and the isthmus thinner in subjects with AD than in normal healthy subjects. For subjects 75-95, it seems to be no difference in the shape of CC between AD and normal healthy subjects. The display is consistent with the result of the above hypothesis testing that there is a significant age dependent difference in the CC shape between AD and normal subjects.

Second, we tested the *null* hypothesis $H_0 : \boldsymbol{\beta}^{(g)} = \mathbf{0}_{60}$ versus the *alternative* $H_1 : \boldsymbol{\beta}^{(g)} \neq \mathbf{0}_{60}$ in both stages I and II. The Wald test statistics are $W_I^{(g)} = 63.78$ and $W_E^{(g)} = 73.34$, in stage I and II, respectively, and they are χ^2_{60} distributed. The p-values are 0.345 and 0.116, in stage I and II, respectively. Thus, the test statistics are not significant at the 0.05 level of significance, in both stages I and II. The data does not contain enough evidence to reject the *null*, that is, it fails to reject that there is no gender effect on the shape of the CC contours. The mean age-trajectories for female and male groups within each diagnosis group, based on the stage II estimates, are shown in Figure 2.7. A similar graph yields using stage I estimates (not displayed here). The display shows that at any age (50-95) the CC contours have very similar shapes for males and females, which is consistent with the results of the hypothesis testing, that there is no significant gender effect on the CC shape.

2.7 Discussion

We have developed a general statistical framework for intrinsic regression models of responses valued in a Riemannian symmetric space in general, and Lie groups in particular, and their association with a set of covariates in a Euclidean space. The intrinsic regression models are based on the generalized method of moment 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, including both one-center and multicenter link functions as special cases. Essentially, the covariates are first mapped to the tangent bundle to the Riemmanian 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 ILSE, $(\widehat{\mathrm{q}}_I, \widehat{\boldsymbol{\beta}}_I)$, of $(\mathrm{q}, \boldsymbol{\beta})$, in the Stage I of the estimation process, and a one-step procedure to search for the efficient estimator $(\tilde{\mathrm{q}}_E, \tilde{\boldsymbol{\beta}}_E)$ in Stage II. Our simulation study for three-dimensional directional data with covariates demonstrates that the relative efficiency of the Stage II estimator improves as the sample size increases. We have provided a simulated dataset example which illustrates that our method outperforms the naïve method in terms of both prediction accuracy and estimation efficiency, when the distribution of the responses strongly depends on the covariates and the covariates have substantial variation.

Our regression framework covers various link functions including both one-center and multicenter link functions as special cases. We believe that our method should outperform the naïve method in terms of both prediction accuracy and estimation efficiency, when the distribution of the responses strongly depends on the covariates and the covariates have substantial variation.

There are still many outstanding issues for further research. One major issue is to construct goodness-of-fit statistics for testing for possible model misspecifications in

(2.6). Another important issue is to develop diagnostic measures for assessing the influence of individual observations in the semiparametric regression for manifold-valued data. We intend to dedicate future work to study these issues. The current work has also motivated our interest for developing nonparametric Bayesian regression models for manifold-valued data.

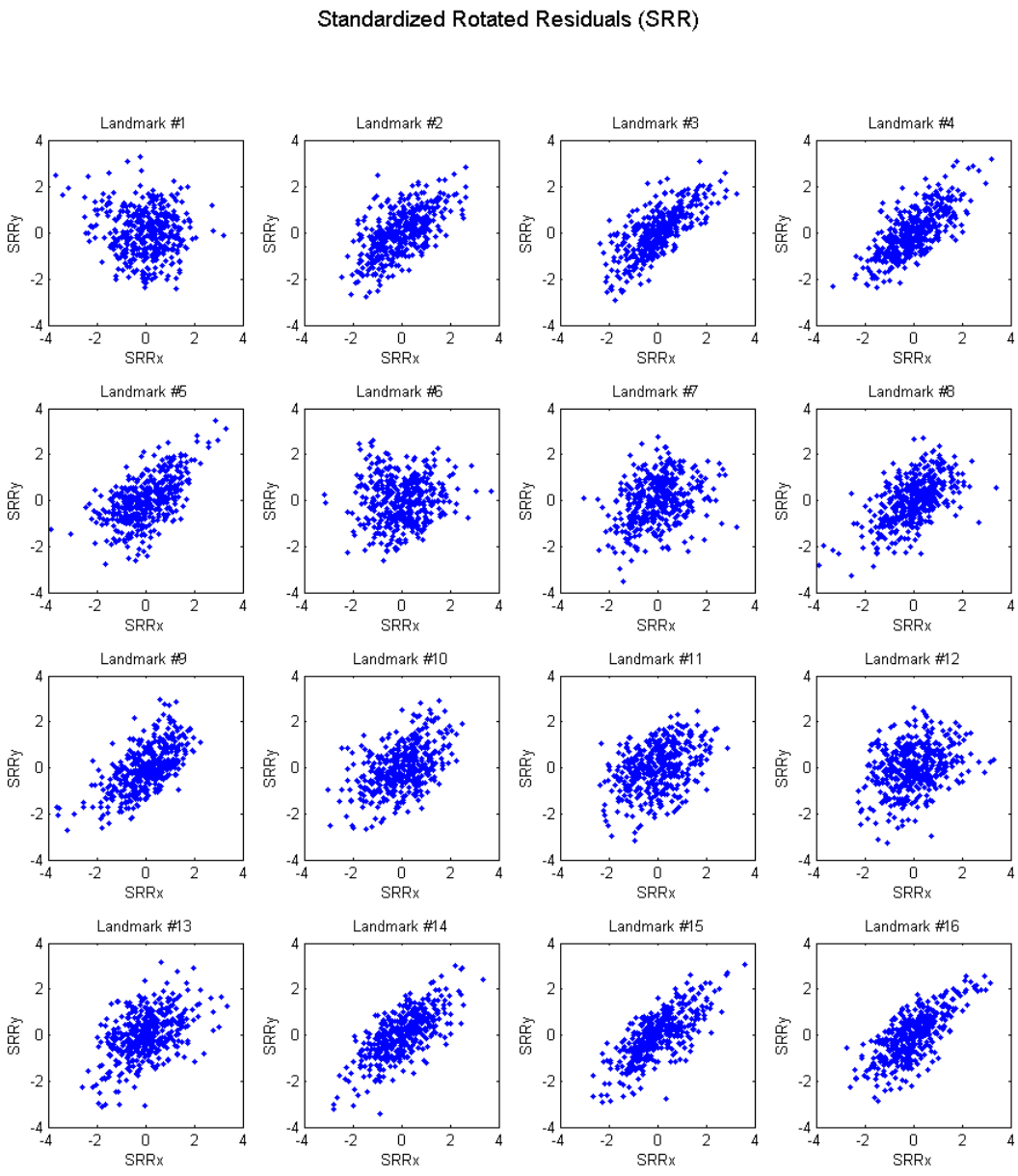

Figure 2.4: The plots of the rotated residuals at the first 16 landmarks

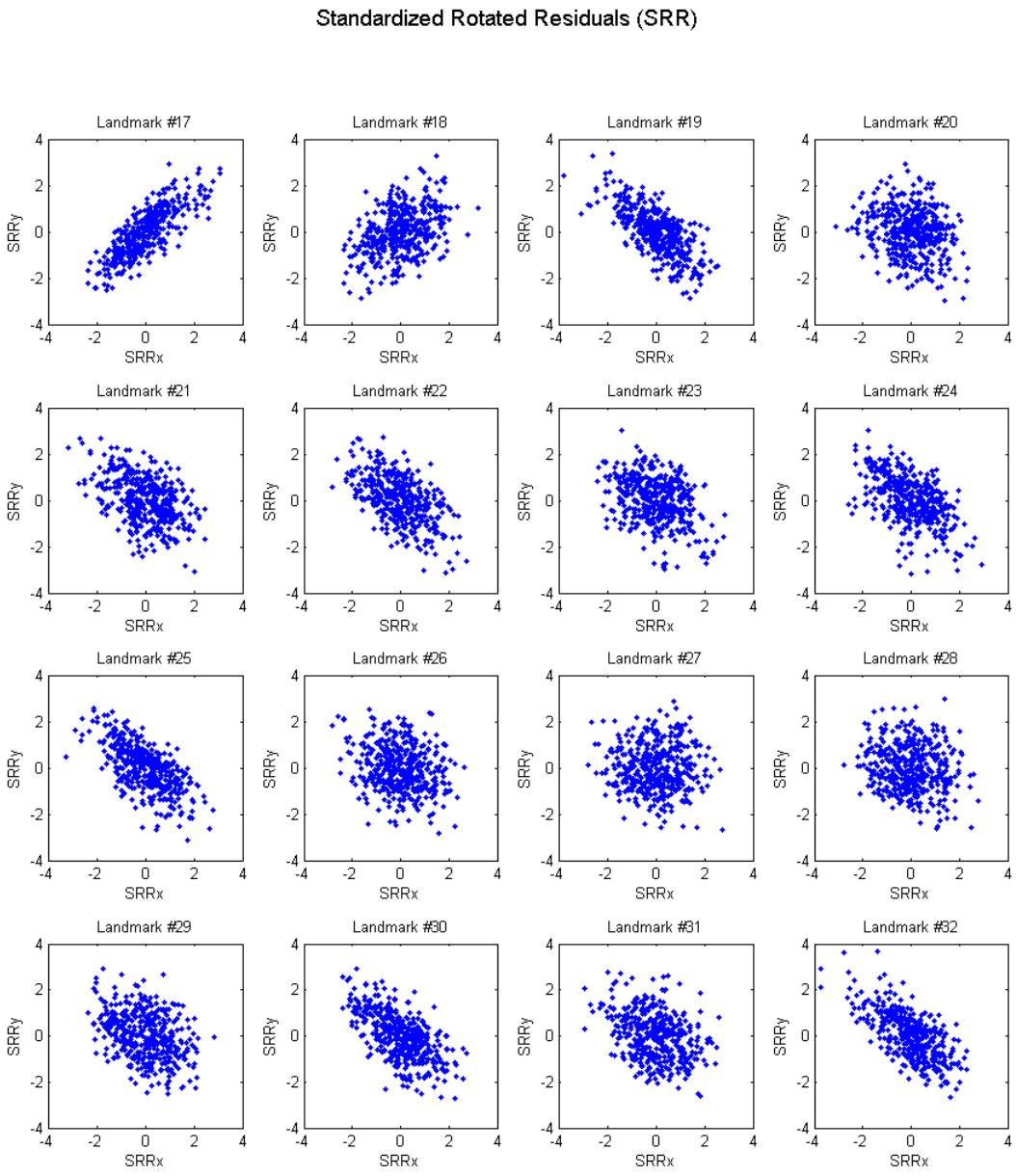

Figure 2.5: The plots of the rotated residuals at the last 16 landmarks

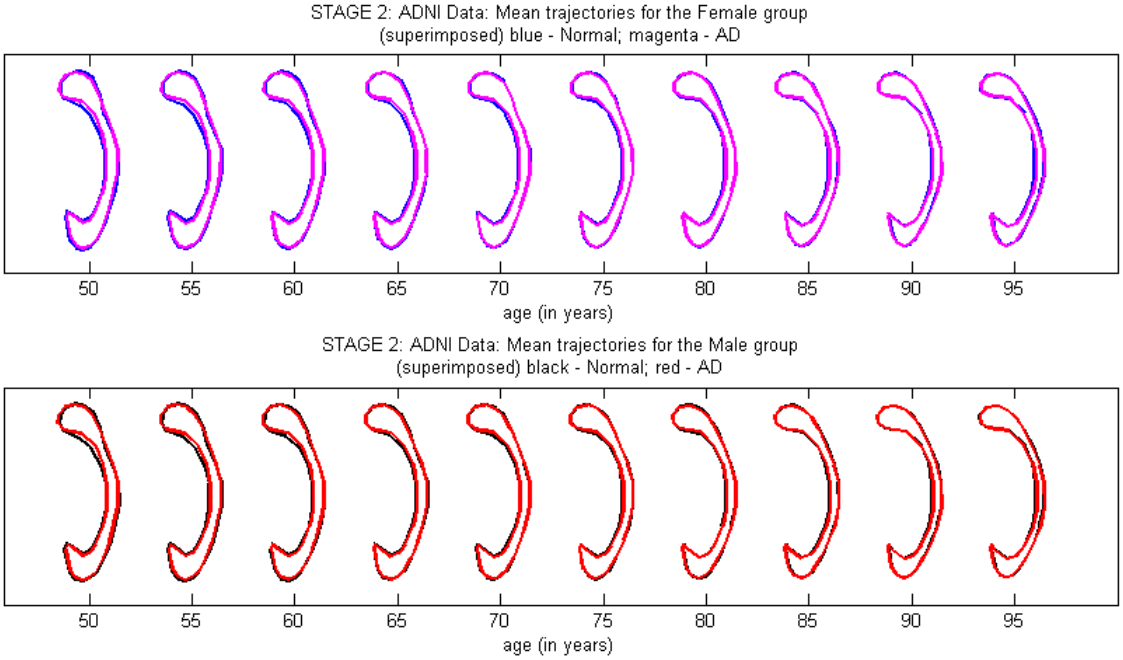

Figure 2.6: Age-trajectories of the intrinsic mean shapes by diagnosis within each gender group, based on the stage II parameter estimates.

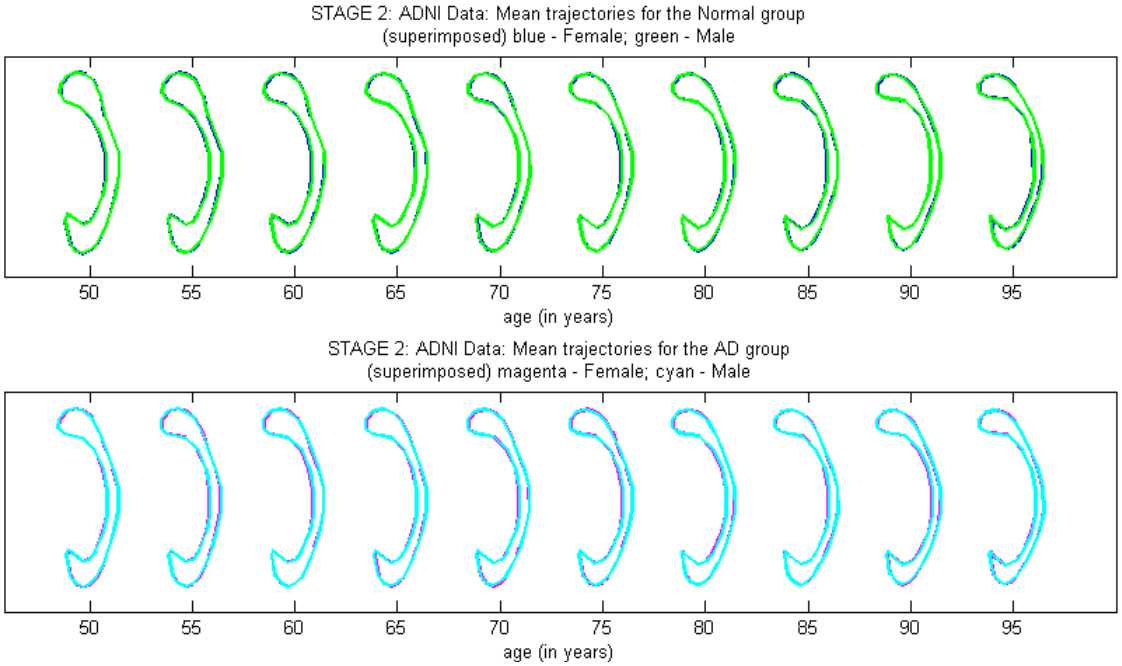

Figure 2.7: Age-trajectories of the intrinsic mean shapes by gender within each diagnosis group, based on the stage II parameter estimates.

CHAPTER 3: LONGITUDINAL DATA ANALYSIS ON RIEMANNIAN MANIFOLDS

3.1 Introduction

Longitudinal data arises in many applications in which the goal is to understand changes in individual entities over time. Repeated measurements of anatomical shape are examples of longitudinal data that takes values in a Riemannian manifold. A driving application of such data is to characterize anatomical shape changes and to distinguish between trends in anatomy that are healthy versus those that are due to disease, adjusted for a set of covariates. The individual shape trajectory is considered as a perturbation of the mean trajectory for the population. For example, the computational anatomy, an emerging discipline at the interface of geometry, statistics and image analysis, aims at modeling and analyzing the biological shape of tissues and organs. The goal is to estimate representative organ anatomies across diseases, populations, species or ages, to model the organ development across time (growth or aging), to establish their variability, and to correlate this variability information with other functional, genetic or structural information

A longitudinal study tracks changes in individuals by repeatedly collecting measurements over time. Longitudinal studies are popular in medicine, where the goal is to understand change processes, such as healthy development, aging, or disease progression. Often, shape is the quantity of interest being tracked. For example, understanding changes in neuroanatomy is a critical goal in the study of degenerative diseases such as Alzheimer's and in developmental disorders such as autism. Longitudinal shape data also arises in various branches of biology, such as as evolutionary biology, where the

evolution of the shapes of bones in the fossil record is of interest. The main challenge for these studies is that shape, i.e., the geometry of an object that is invariant to rotation, scaling, and translation, is inherently nonlinear and high-dimensional. Because of this, manifold representations of shape have proven to be effective. Therefore, analysis of shape changes necessitates the development of models for dealing with manifold-valued longitudinal data. Such models would also benefit other applications that involve serial collection of manifold data, such as directional data, transformation groups, and tensors.

Related to the longitudinal data analysis problem is the regression problem. However, regression does not model individual changes and is not appropriate for analyzing longitudinal data. Instead, regression models are used for describing cross-sectional data, where only one data point per individual is available. Several authors have proposed methods for regression on manifolds. Jupp and Kent (1987) proposed an unrolling method on shape spaces. Regression analysis on the group of diffeomorphisms has been proposed as growth models by Miller (2004), nonparametric regression by Davis et al. (2010), and second-order splines by Trouvé and Vialard (2010). Recently, parametric models of regression, where the regression function is a geodesic curve, have been introduced independently by Fletcher (2013) and Niethammer et al. (2011). Shi et al. (2012) proposed a semiparametric regression model with multiple covariates for median representation of subcortical structures. Cornea et al. proposed a general framework for intrinsic regression models model with multiple Euclidean covariates for Riemannian symmetric space response data.

Related work in longitudinal analysis includes several approaches in the setting of diffeomorphic transformations, which form an infinite-dimensional manifold, applied to image sequences. Durrleman et al. (2009) constructed spatiotemporal image atlases from longitudinal data. Qiu et al. (2009) used parallel translation to bring individual

trajectories to a common point for comparison. Lorenzi et al. (2011) used a hierarchical model on stationary velocity fields, in a framework that does not include a Riemannian metric on the manifold of diffeomorphisms. An important shortcoming of these approaches is that they do not model distances between trajectories. This makes it difficult to compare the differences in trends of two groups, or even to rigorously define the concept of the variance of a population of trends.

The aim of this work is to extend the framework of intrinsic regression models, proposed by Cornea et al., to fixed and random effect models for the analysis of manifold-valued measures from longitudinal studies.

3.2 Intrinsic Regression Model

Let $(\mathcal{M}, \mathrm{m})$ be an (C^∞) RSS of dimension $d_\mathcal{M}$ and geodesically complete with an inner product $\mathrm{m_p}$ and G be a Lie group of isometries acting smoothly and transitively on $\mathcal{M}$ with the identity element e. Let $\mathrm{p} \in \mathcal{M}$ be a base point of $\mathcal{M}$ and $\rho = \rho^*_\mathcal{M}$ be the radius of injectivity of $\mathcal{M}$.

3.2.1 Formulation of Intrinsic Fixed Effect Regression Model

Consider a longitudinal response variable, y_i, for the i-th subject, $i = 1, \ldots, n$, taking values in $\mathcal{M}$. The observed values of the response is denoted y_{ij}, corresponding to the j-th observation of the i-th subject, $j = 1, \ldots, r_i$. Let t_i be the denote the independent variable, typically time, with the value $t_{ij} \in \mathbf{R}$ corresponding to the y_{ij}. Let $\mathbf{x}_{ij} = (x_{ij,1}, \ldots, x_{ij,d_\mathbf{x}})^\top$ be an Euclidean $d_\mathbf{x}$-dimensional vector of covariates corresponding to the i-th subject at the j-th observation. Our objective is to introduce intrinsic regression models for longitudinal data $(y_{ij}, t_{ij}, \mathbf{x}_{ij})$, $i = 1, \ldots, n$, $j = 1, \ldots, r_i$. When $j = 1$, $(y_{i1}, t_{i1}, \mathbf{x}_{i1})$ is the data at the baseline for the i-th subject.

Intrinsic Fixed Effect Model (IFEM). The specification of IFEM involves three key steps: (i) a link function mapping from the space of covariates to the manifold $\mathcal{M}$, (ii) the definition of residual, and (ii) the action of transporting all residuals to a common space.

First, we explicitly formalize the link function. From now on, all covariates in $\mathbf{x}_{ij}$ have been standardized and the times t_{ij} have been normalized by the standard deviation and shifted so that $t_{i1} = 0$ represents the time at the baseline. We consider a single-center link function given by

$$\boldsymbol{\mu}(t, \mathbf{x}, \mathrm{q}, \boldsymbol{\beta}) : R \times R^{d_x} \times \mathcal{M} \times R^{d_\beta} \to \mathcal{M}, \tag{3.1}$$

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

$$\boldsymbol{\mu}(0, \mathbf{0}, \mathrm{q}, \boldsymbol{\beta}) = \boldsymbol{\mu}(t, \mathbf{x}, \mathrm{q}, \mathbf{0}) = \mathrm{q}. \tag{3.2}$$

When the regression coefficient vector $\boldsymbol{\beta}$ equals $\mathbf{0}$, the link function is independent of time and covariates and thus, it reduces to a single center (or "'mean"') q.

Secondly, we introduce the concept of "residual" to ensure that $\boldsymbol{\mu}(t_{ij}, \mathbf{x}_{ij}, \mathrm{q}, \boldsymbol{\beta})$ is the proper "conditional mean" of y_{ij} given t_{ij} and $\mathbf{x}_{ij}$, which is the key concept of regression models. Given the points y_{ij} and $\boldsymbol{\mu}(t_{ij}, \mathbf{x}_{ij}, \mathrm{q}, \boldsymbol{\beta})$ in $\mathcal{M}$, we define the residual as " a difference" between y_{ij} and $\boldsymbol{\mu}(t_{ij}, \mathbf{x}_{ij}, \mathrm{q}, \boldsymbol{\beta})$. Assumming that y_{ij} and $\boldsymbol{\mu}(t_{ij}, \mathbf{x}_{ij}, \mathrm{q}, \boldsymbol{\beta})$ are close enough to each other, namely,

$$d_{\mathcal{M}}(y_{ij}, \boldsymbol{\mu}(t_{ij}, \mathbf{x}_{ij}, \mathrm{q}, \boldsymbol{\beta})) < \rho.$$

Thus, as in the cross-sectional case, $\mathrm{Log}_{\boldsymbol{\mu}(t_{ij},\mathbf{x}_{ij},\mathrm{q},\boldsymbol{\beta})} y_{ij}$ may make it a good candidate to play the role of "residual." As these residuals lie on different tangent spaces to $\mathcal{M}$, it is difficult to carry out a multivariate analysis of these residuals.

Thirdly, since $\mathcal{M}$ is a Riemannian manifold, this enable us to "transport" all residuals, separately, to a common tangent space, say $T_{\mathrm{p}}\mathcal{M}$ at a base point p, by using the parallel transport induced by the metric $\mathbf{m}$. Indeed, the base point p and the point $\boldsymbol{\mu}_{ij}(\mathrm{q},\boldsymbol{\beta}) := \boldsymbol{\mu}(t_{ij},\mathbf{x}_{ij},\mathrm{q},\boldsymbol{\beta})$ can be joined in $\mathcal{M}$ by a unique geodesic curve $\gamma_{ij}(\,\cdot\,;\mathrm{q},\boldsymbol{\beta})$, with unit velocity and satisfying $\gamma_{ij}(0;\mathrm{q},\boldsymbol{\beta}) = \mathrm{p}$ and $\gamma_{ij}(d_{ij};\mathrm{q},\boldsymbol{\beta}) = \boldsymbol{\mu}_{ij}(\mathrm{q},\boldsymbol{\beta})$, where d_{ij} is the distance in $\mathcal{M}$ from p to $\boldsymbol{\mu}_{ij}(\mathrm{q},\boldsymbol{\beta})$. Let $\Gamma_{\mu_{ij}(\mathrm{q},\boldsymbol{\beta})\to\mathrm{p}}$ denote the parallel transport along $\gamma_{ij}(\,\cdot\,;\mathrm{q},\boldsymbol{\beta})$ from $T_{\boldsymbol{\mu}_{ij}(\mathrm{q},\boldsymbol{\beta})}\mathcal{M}$ to $T_{\mathrm{p}}\mathcal{M}$.

We define the *transported (repositioned) residual* of $y_{ij} \in \mathcal{M}$ with respect to $\boldsymbol{\mu}_{ij}(\mathrm{q},\boldsymbol{\beta})$ as

$$\begin{aligned}
\mathcal{E}(y_{ij},t_{ij},\mathbf{x}_{ij};\mathrm{q},\boldsymbol{\beta}) &= \mathcal{E}_{ij}(\mathrm{q},\boldsymbol{\beta}) = \Gamma_{\mu_{ij}(\mathrm{q},\boldsymbol{\beta})\to\mathrm{p}}(\mathrm{Log}_{\mu_{ij}(\mathrm{q},\boldsymbol{\beta})} y_{ij}) \\
&= L_{c_{ij}(1;\mathrm{q},\boldsymbol{\beta})^{-1}*}(\mathrm{Log}_{\mu_{ij}(\mathrm{q},\boldsymbol{\beta})} y_{ij}) \\
&= \mathrm{Log}_{\mathrm{p}}(c_{ij}(1;\mathrm{q},\boldsymbol{\beta})^{-1}\cdot y_{ij}) \in T_{\mathrm{p}}\mathcal{M}
\end{aligned} \tag{3.3}$$

$i = 1,\ldots,n$ and $j = 1,\ldots,r_i$, where $c_{ij}(\,\cdot\,;\mathrm{q},\boldsymbol{\beta})$ is the one-parameter subgroup of G such that $\boldsymbol{\mu}_{ij}(\mathrm{q},\boldsymbol{\beta}) = c_{ij}(1;\mathrm{q},\boldsymbol{\beta})\cdot\mathrm{p}$, and $T_{\mathrm{p}}\mathcal{M} \equiv R^{d_{\mathcal{M}}}$ via an orthonormal basis. The *intrinsic fixed effects* on $\mathcal{M}$ is defined by

$$E[\mathcal{E}(y_{ij},t_{ij},\mathbf{x}_{ij};\mathrm{q}_*,\boldsymbol{\beta}_*)\,|\,\mathbf{x}_{ij},t_{ij}] = \mathbf{0}, \tag{3.4}$$

$i = 1,\ldots,n$ and $j = 1,\ldots,n_i$, where $(\mathrm{q}_*,\boldsymbol{\beta}_*)$ denote the true value of $(\mathrm{q},\boldsymbol{\beta})$ and the distribution is taken with respect to the conditional distribution of y_{ij} given t_{ij} and $\mathbf{x}_{ij}$. The model is essentially semi-parametric, since we do not restrict the joint distribution of $(y,t,\mathbf{x})$ except by the conditional moment in (3.4). This IFEM is actually the intrinsic

model from Section in which the time is incorporated as an additional covariate. Thus, the same estimation procedure may apply here as well.

Let $(\hat{\mathrm{q}}, \hat{\boldsymbol{\beta}})$ be an estimator of $(\mathrm{q}_*, \boldsymbol{\beta}_*)$. Then, $t \to \boldsymbol{\mu}(t, \mathbf{x}(t), \hat{\mathrm{q}}, \hat{\boldsymbol{\beta}})$ is the estimated mean trajectory at the population level of an individual whose covariates at time t have the values $\mathbf{x}(t)$, while $t \to \boldsymbol{\mu}(t, \mathbf{x}(t), \mathrm{q}_*, \boldsymbol{\beta}_*)$ is the true trajectory at the population level.

We note that the ILSE $(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)$ coincides with the MLE under the assumption that $\mathcal{E}_{ij}(\mathrm{q}, \boldsymbol{\beta})|\mathrm{q}, \boldsymbol{\beta} \overset{iid}{\sim} N_{d_{\mathcal{M}}}(\mathbf{0}, \sigma^2 I)$ in $T_{\mathrm{p}}\mathcal{M} \cong R^{d_{\mathcal{M}}}$.

3.3 Longitudinal ADNI Corpus Callosum Shape Data Example

We use again the AD study and the CC shape data from the ADNI database, which were described in Section 2.6.1 above.

3.3.1 Intrinsic Fixed Effects Model

We are interested in characterizing the change of the CC contour shape as a function of three covariates including gender, age, and AD diagnosis at the baseline and the time from baseline. We focused on $n = 400$ subjects with 214 healthy controls (HCs) and 186 AD patients at baseline of the ADNI1 database. We observed a CC planar contour y_{ij} with 32 landmarks, three time independent clinical variables including gender ($x_{i,1} = 0$ for female and 1 for male), age ($x_{i,2}$), and diagnosis ($x_{i,3} = 0$ for normal control and 1 for AD) assessed at the baseline, and the times t_{ij} measured from the baseline, for $i = 1, \ldots, 409$, and $j = 1, \ldots, n_i$. Here, n_i is the number of measurements for the i-th subjects, which for these data is between 1 and 6. Each subject has at least one measurement - at the baseline $t_{i1} = 0$ - and at most 6 repeated measurements over a period no longer than 52 months. The demographic information and number of measurements of these 400 subjects are presented in Table 3.1 and Figure 3.1.

Throughout the rest of the section we use the notation and definitions for Sections

Table 3.1: Demographic information about processed longitudinal ADNI CC shape dataset, including disease status, age, gender, and number of measurements.

Disease status	Num.	Range of age in years(mean)	Gender (female/male)	Range of # of measurements (mean)
Healthy Control (HC)	214	62-90 (76.09)	107/107	1-6 (3.79)
AD	186	55-92 (75.42)	88/98	1-4 (2.90)

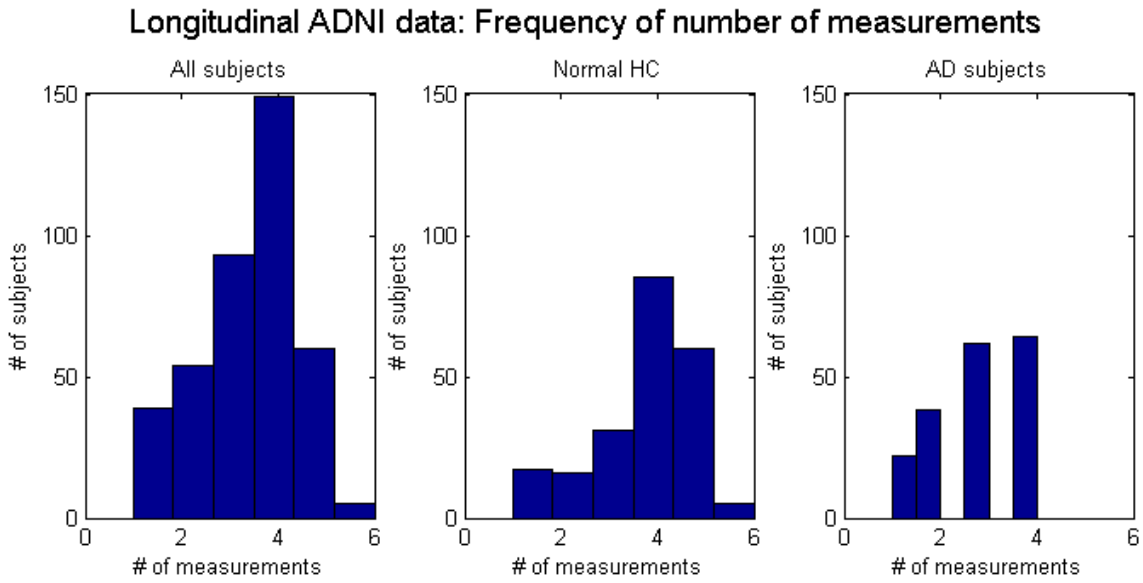

Figure 3.1: Longitudinal ADNI Date: Histogram of the number of measurements.

2.4.4 and 2.6.2.

We consider an intrinsic fixed effects regression model with $y_{ij} \in \Sigma_2^{32}$ as a response vector shape and a vector of four covariates including gender, age, diagnosis, age*diagnosis, time, time*age, and time*diagnosis, that is,
$x_{ij} = (x_{i,1}; x_{i,2}; x_{i,3}; x_{i,4}, x_{ij,5}, x_{ij,6}, x_{ij,7})^T$ with $x_{i,4} = x_{i,2}x_{i,3}$, $x_{ij,5} = t_{ij}$, $x_{ij,6} = t_{ij}x_{i,2}$, and $x_{ij,7} = t_{ij}x_{i,3}$. We used a single-center link function with model parameters $(q, \boldsymbol{\beta}) \in \Sigma_2^{32} \times R^{420}$ as follow (here, $d_{\boldsymbol{\beta}} = 420$). The intercept q is specified by $\mathrm{q} = \phi_{\mathrm{p_0}}^{-1}(\boldsymbol{\tau}) := \mathrm{Exp}_{\mathrm{p_0}}\left(\sum_{\ell=1}^{30}(\tau_{2\ell-1} + j\tau_{2\ell})Z_\ell\right)$, where $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_{60})^\top \in R^{60}$. The regression coefficients $\boldsymbol{\beta}$ includes seven 60×1 subvectors $\boldsymbol{\beta}^{(g)}$, $\boldsymbol{\beta}^{(a)}$, $\boldsymbol{\beta}^{(d)}$, $\boldsymbol{\beta}^{(ad)}$,$\boldsymbol{\beta}^{(t)}$, $\boldsymbol{\beta}^{(ta)}$, and $\boldsymbol{\beta}^{(td)}$, which correspond to $x_{ij,1}, \ldots, x_{ij,6}$, and $x_{ij,7}$, respectively. Therefore,

there are 480 unknown parameters in $(\boldsymbol{\tau}^\top, \boldsymbol{\beta}^\top)^\top \in R^{480}$. We define a 30×7 complex matrix as

$$B = \begin{pmatrix} \beta_1 + j\beta_2 & \beta_{61} + j\beta_{62} & \dots & \beta_{361} + j\beta_{362} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{59} + j\beta_{60} & \beta_{119} + j\beta_{120} & \dots & \beta_{419} + j\beta_{420} \end{pmatrix} \in C^{30\times 7}.$$

The single-center link function is given by

$$\boldsymbol{\mu}(t_{ij}, \mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta}) = \mathrm{Exp}_{\mathrm{q}}\left([U_{\mathrm{p,q}}Z_1, \dots, U_{\mathrm{p,q}}Z_{30}]\, B\mathbf{x}_{ij}\right) \in \Sigma_2^{32}.$$

Finally, our intrinsic fixed effect regression model is defined by

$$E\left[\mathrm{Log}_{\mathrm{p}}\left(\overline{U}^T_{\mathrm{p},\boldsymbol{\mu}(t_{ij},\mathbf{x}_i,\mathrm{q},\boldsymbol{\beta})}\, y_i\right) \,\middle|\, \mathbf{x}_i, t_{ij}\right] = \mathbf{0} \text{ for } i = 1, \dots, 400,\ j = 1, \dots, n_i. \tag{3.5}$$

3.3.2 Results

Recall that Σ_2^{32} is compact and its diameter is π. The 1352 response data, y_{ij}, $i = 1, \dots, 400$, $j = 1, \dots, n_i$, lie in an open subset of Σ_2^{32} of diameter less than 0.26. The within-subject response data sets have diameter ranging between 0 and 0.1155, with mean and median equal to 0.0382 and 0.0380, respectively. The maximum diameter occurs for the 285-th subject, who is a 71-year old normal HC female with 4 measurements, at the baseline and 14, 25, and 37 months later, respectively. Among the AD subjects, the largest diameter of within-subject response data is 0.1041, and it occurs for the 185-th subject, who is a 87-year old AD male with 3 measurements, at the baseline, and 7 and 13 months. The average follow-up period of time is about 22.11 months, with the maximum of 52 months.

Initially, we fitted the expanded model of (3.5) which includes the three-way interaction time*age*diagnosis. We found out that the effect of that term is non-significant

and we remove it from the model. We ended up fitting the model (3.5). We calculated the efficient estimator $(\tilde{\mathrm{q}}, \tilde{\boldsymbol{\beta}}) = (\phi_{p_0}^{-1}(\tilde{\boldsymbol{\tau}}_E, \tilde{\boldsymbol{\beta}}_E)$ defined in Section 2.3.3, Eq. (2.15). The intercept estimate $\tilde{\mathrm{q}}$ is an estimate of the Fréchet mean of y_{ij}'s, and distance between them is less than 0.0004. The responses are within a distance less than 0.15 from the sample Frëchet mean, and in an average distance of 0.057. The estimates $\tilde{\boldsymbol{\beta}}$ of the regression coefficients and their standard deviations are shown in Figure 3.2. The estimated covariance matrix of $\tilde{\boldsymbol{\beta}}_E$ is shown in Figure 3.3.

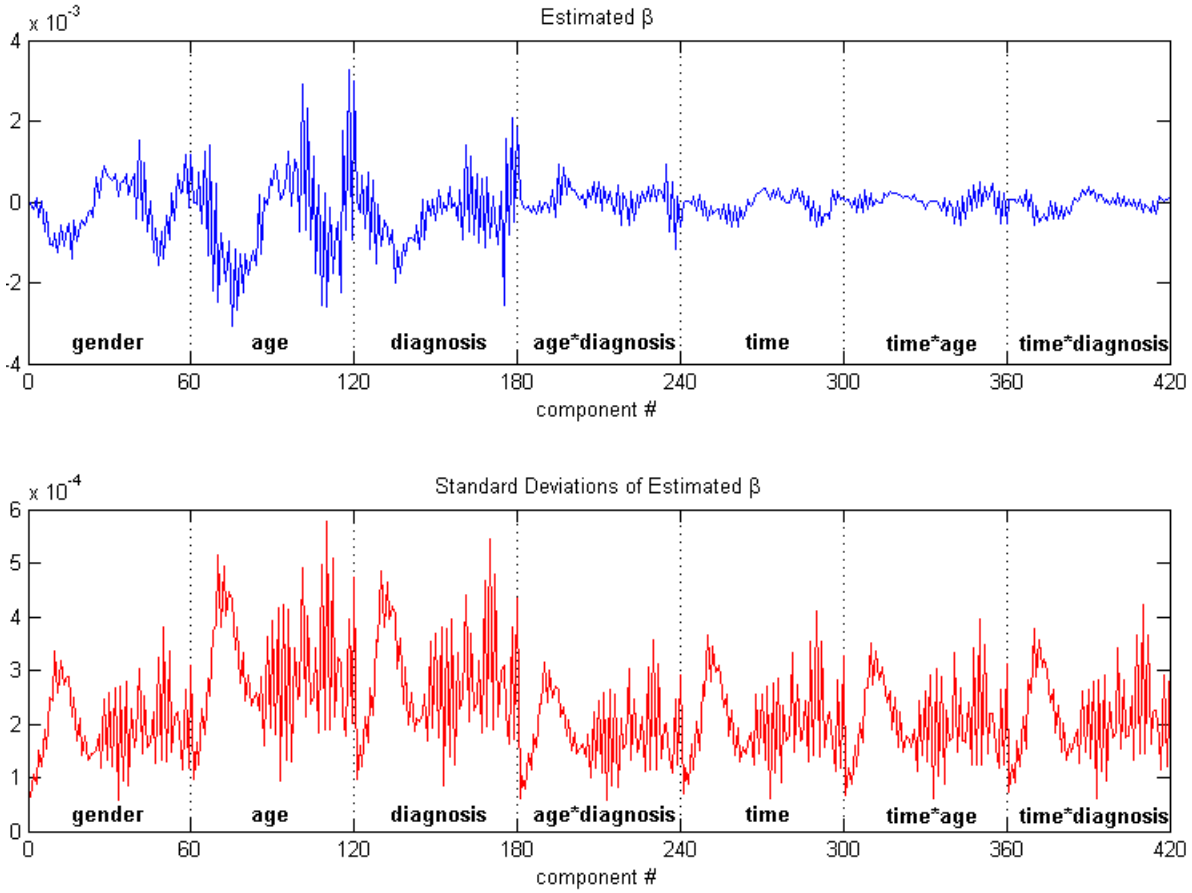

Figure 3.2: Longitudinal ADNI Data: Estimated regression coefficients and their standard deviations.

We have visually inspected the standardized estimated rotated residuals in order to

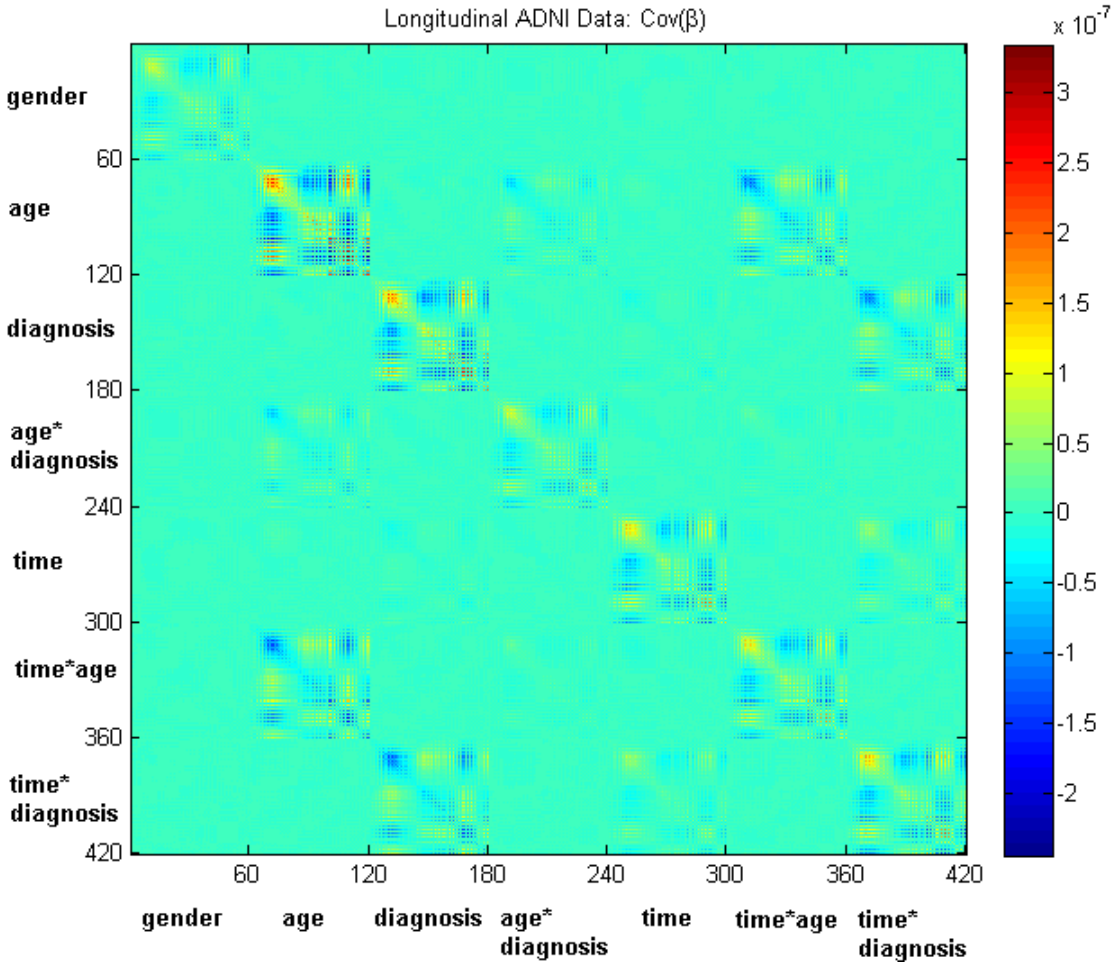

Figure 3.3: Longitudinal ADNI Data: Estimated regression coefficients and their standard deviations.

ensure the goodness-of-fit of our model to the real data. The standardized estimated rotated residuals are shown in Figures 3.4 and 3.5. Visually, we cannot find ant outlier in the dataset.

To assess whether there are or not age dependent time and/or diagnosis dependent effects and/or age dependent diagnosis effect on the shape of the CC contour, we performed several sets of hypothesis testings. First, we tested the *null* hypothesis $H_0 : \boldsymbol{\beta}^{(ta)} = \mathbf{0}_{60}$ versus the *alternative* $H_1 : \boldsymbol{\beta}^{(ta)} \neq \mathbf{0}_{60}$. The Wald test statistic is $W^{(ta)} = 73.77$ and it is χ^2_{60} distributed. The p-values are 0.109. Thus, the test statistic

is not significant at both 0.05 and 0.01 levels of significance. There is not enough evidence in the data to reject the *null*, that is, that these data do not provide evidence that there is an age dependent time effect on the shape of the CC contours. One reason for that it might be the relative short follow-up period and few subject-measurements.

Second, we tested the *null* hypothesis $H_0 : \boldsymbol{\beta}^{(td)} = \mathbf{0}_{60}$ versus the *alternative* $H_1 : \boldsymbol{\beta}^{(td)} \neq \mathbf{0}_{60}$. The Wald test statistic is $W^{(td)} = 83.2$ and it is χ^2_{60} distributed. The p-values are 0.024. Thus, the test statistic is significant at 0.05, but not at 0.01, level of significance. The data contains enough evidence to reject the *null*, that is, to reject that there is no diagnosis dependent time effect on the shape of the CC contours.

Third, we tested the *null* hypothesis $H_0 : \boldsymbol{\beta}^{(ad)} = \mathbf{0}_{60}$ versus the *alternative* $H_1 : \boldsymbol{\beta}^{(ad)} \neq \mathbf{0}_{60}$. The Wald test statistic is $W^{(td)} = 80.12$ and it is χ^2_{60} distributed. The p-values are 0.042. Thus, the test statistic is significant at 0.05, but not at 0.01, level of significance. The data contains enough evidence to reject the *null*, that is, to reject that there is no age dependent diagnosis effect on the shape of the CC contours.

Based on these tests, the data show enough evidence to conclude that the AD subject trajectories at the population level have a CC shape that depends of the age at the baseline and the CC shape changes over time.

To illustrate the subject-trajectories of conditional means at the population level, we selected three subjects, the one with the most spread responses overall (who is a normal HC female), the one with the most spread responses within the AD group (who is a male), and one of those that have the largest number, 6, of measurements (who is normal HC male). At the top of Figures 3.6, 3.7, and 3.8, there are shown the trajectories of the subject and his/her would-be opposite diagnosis counterpart subject, i.e. an hypothetical subject who has the same covariate values, but for diagnosis which is the opposite one. We see that there is a diagnosis effect on the evolution of the conditional mean shapes over time, adjusted for all the other baseline covariate at the same

values. At the bottom of Figures 3.6, 3.7, and 3.8 there are displayed the conditional mean shapes at the time of measurements together with the observed response shapes. Visually, we observed a large variability between the within-subject response and the corresponding conditional mean at the population level. These suggest the need to account for subject specific random-effects. In this ongoing research, the next aim is to define the framework of *intrinsic random-effects models (IREM)* for manifold data.

3.4 Simulation Study

We generate 3-D directional longitudinal data with random effects, $(y_{ij}, x_{ij}, t_{ij}) \in S^2 \times R \times [0, \infty)$, $i = 1, \ldots, n$, $j = 1, \ldots, n_i$, as follows. Let $p = (0, \ldots, 0, -1) \in S^2$ be the "South" pole. We set $f(x_{ij}, t_{ij}, \boldsymbol{\beta}) = B(x_{ij}, t_{ij})^\top$, in which $x_{ij} \in R$, $t_{ij} \in [0, \infty)$, and $B = \left[\begin{smallmatrix} \beta_1 & \beta_3 \\ \beta_2 & \beta_4 \end{smallmatrix}\right]$, and $\mathrm{q} = (u, v, w) \in S^2 \setminus \{-p\}$. The subject conditional mean at the population level is defined by $\boldsymbol{\mu}(t_{ij}, x_{ij}, \mathrm{q}, \boldsymbol{\beta}) = T^{-1}_{st,q}\left(\mathrm{R}_{p,q}[(f(x_{ij}, t_{ij}, \boldsymbol{\beta})^\top, -1)^\top]\right)$. We set m the maximum number of observations per subject. We independently simulate x_i, $i = 1, \ldots, n$, from the standard normal distribution $N(0, 1)$, and n_i's as numbers between 1 and m. Each subject has at least one observation (at the baseline) and no more than m observations. For each subject i, we set $t_{i1} = 0$, the time at the baseline, and we randomly sample without replacement $n_i - 1$ numbers from $\{s/(m-1) : s = 1, \ldots, m-1\}$, sort them in increasing order, and label them $t_{i2}, \ldots, t_{i,n_i}$, if $n_i > 1$. For each $j = 1, \ldots, n_i$, we set $x_{ij} = x_i$, i.e. x_{ij} is a time-independent covariate; its value is set at the baseline. We fix the true values of $(\mathrm{q}, \boldsymbol{\beta})$ to be $(\mathrm{q}_*, \boldsymbol{\beta}_*) = ((2/3, 2/3, 1/3)^\top, (1, 1, 1, 1)^\top)$. The chart ϕ, given by the stereographic projection from the "North" pole, $-p$, yields the parametrization $\tau_1 = u/(1-w)$ and $\tau_2 = v/(1-w)$, and thus q_* corresponds to $\boldsymbol{\tau}_* = (1, 1)^\top$. We independently generated $\mathcal{V}_i$'s in R^6 from the multivariate normal distribution $N_6\left(\mathbf{0}, 0.1\left(I_3 \otimes \left[\begin{smallmatrix} 1 & 0.2 \\ 0.2 & 1 \end{smallmatrix}\right]\right)\right)$ and set the random effects $\mathcal{U}_i = (\mathcal{V}_{i,1} + \mathcal{V}_{i,3} x_i + \mathcal{V}_{i,5} t_{ij}, \mathcal{V}_{i,2} + \mathcal{V}_{i,4} x_i + \mathcal{V}_{i,6} t_{ij}, -1)^\top \in T_{\mathrm{p}} S^2$, rotate

Table 3.2: Bias ($\times 10^{-2}$) and MS($\times 10^{-2}$) of $(\hat{\mathbf{t}}_I, \hat{\boldsymbol{\beta}}_I)$ and $(\hat{\mathbf{t}}_E, \hat{\boldsymbol{\beta}}_E)$. Bias denotes the bias of the mean of the estimates; MS denotes the root-mean-square error.

		$n = 40$		$n = 80$	
		Bias	MS	Bias	MS
I	τ_1	2.94	6.13	0.55	4.04
E		2.69	5.15	0.32	3.45
I	τ_2	1.69	6.34	-0.28	5.05
E		1.56	5.71	-0.06	4.91
I	β_1	3.05	8.12	1.15	7.25
E		3.29	8.02	0.90	6.11
I	β_2	-0.39	9.68	-1.00	7.18
E		0.10	10.11	-0.47	6.44
I	β_3	-0.22	7.55	0.08	4.97
E		0.36	7.09	-0.05	4.57
I	β_4	0.73	6.80	-1.31	5.89
E		1.68	6.81	-0.99	5.10

them onto $T_{\boldsymbol{\mu}(t_{ij},x_{ij},\mathrm{q}_*,\boldsymbol{\beta}_*)}S^2$, and then map them on the S^2 via the inverse stereographic projection from $-\boldsymbol{\mu}(t_{ij}, x_{ij}, \mathrm{q}, \boldsymbol{\beta})$ to calculate the subject-specific conditional mean $\boldsymbol{\eta}(t_{ij}, x_{ij}, \mathrm{q}, \boldsymbol{\beta}, \mathcal{U}_i)$. Finally, we generate the error measurements residuals $\mathcal{E}_{ij}^o$ from the distribution $N_2\left(\mathbf{0}, 0.2 \left[\begin{smallmatrix} 1 & 0.2 \\ 0.2 & 1 \end{smallmatrix}\right]\right)$ on $T_{\mathrm{p}}S^2$, rotate them onto $T_{\boldsymbol{\eta}(t_{ij},x_{ij},\mathrm{q}_*,\boldsymbol{\beta}_*,\mathcal{U}_i)}S^2$, and then map them on the S^2 via the inverse stereographic projection from $-\boldsymbol{\eta}(t_{ij}, x_{ij}, \mathrm{q}_*, \boldsymbol{\beta}_*, \mathcal{U}_i)$ to calculate the responses y_{ij}. We set n=40,80 and m=8 in order to examine the finite sample performance of the parameter estimates.

We compared the biases and the root-mean-square errors of the two estimates $(\hat{\boldsymbol{\tau}}_I, \hat{\boldsymbol{\beta}}_I)$ and $(\tilde{\boldsymbol{\tau}}_E, \tilde{\boldsymbol{\beta}}_E)$. As seen in Table 3.2, $\tilde{\boldsymbol{\tau}}_E$ and $\tilde{\boldsymbol{\beta}}_E$ have smaller root-mean-square errors than $\hat{\boldsymbol{\tau}}_I$ and $\hat{\boldsymbol{\beta}}_I$ for every component of $\boldsymbol{\tau}$ and $\boldsymbol{\beta}$, respectively, confirming that $(\tilde{\boldsymbol{\tau}}_E, \tilde{\boldsymbol{\beta}}_E)$ is more efficient.

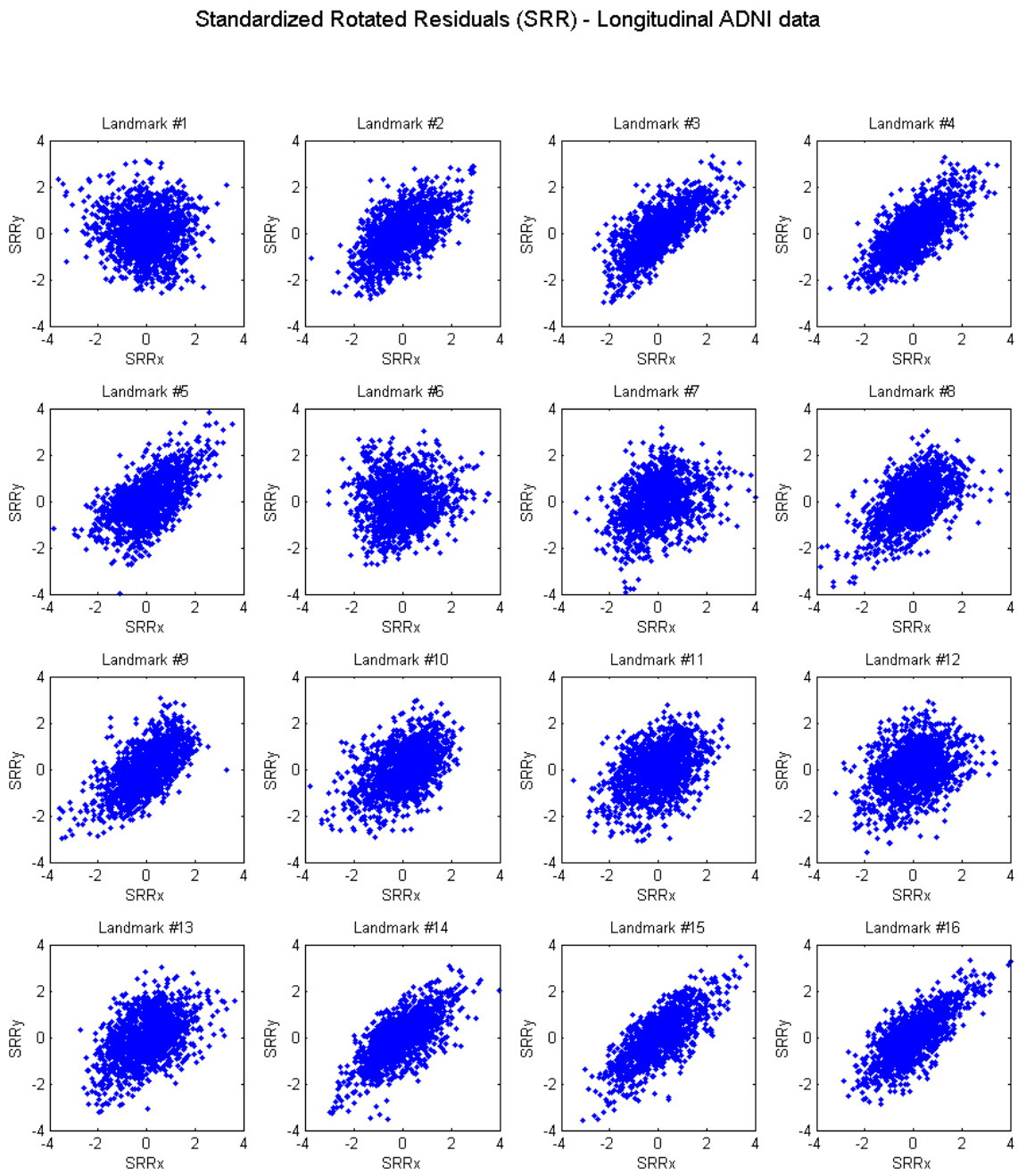

Figure 3.4: Longitudinal ADNI Data: The plots of the rotated residuals at the first 16 landmarks

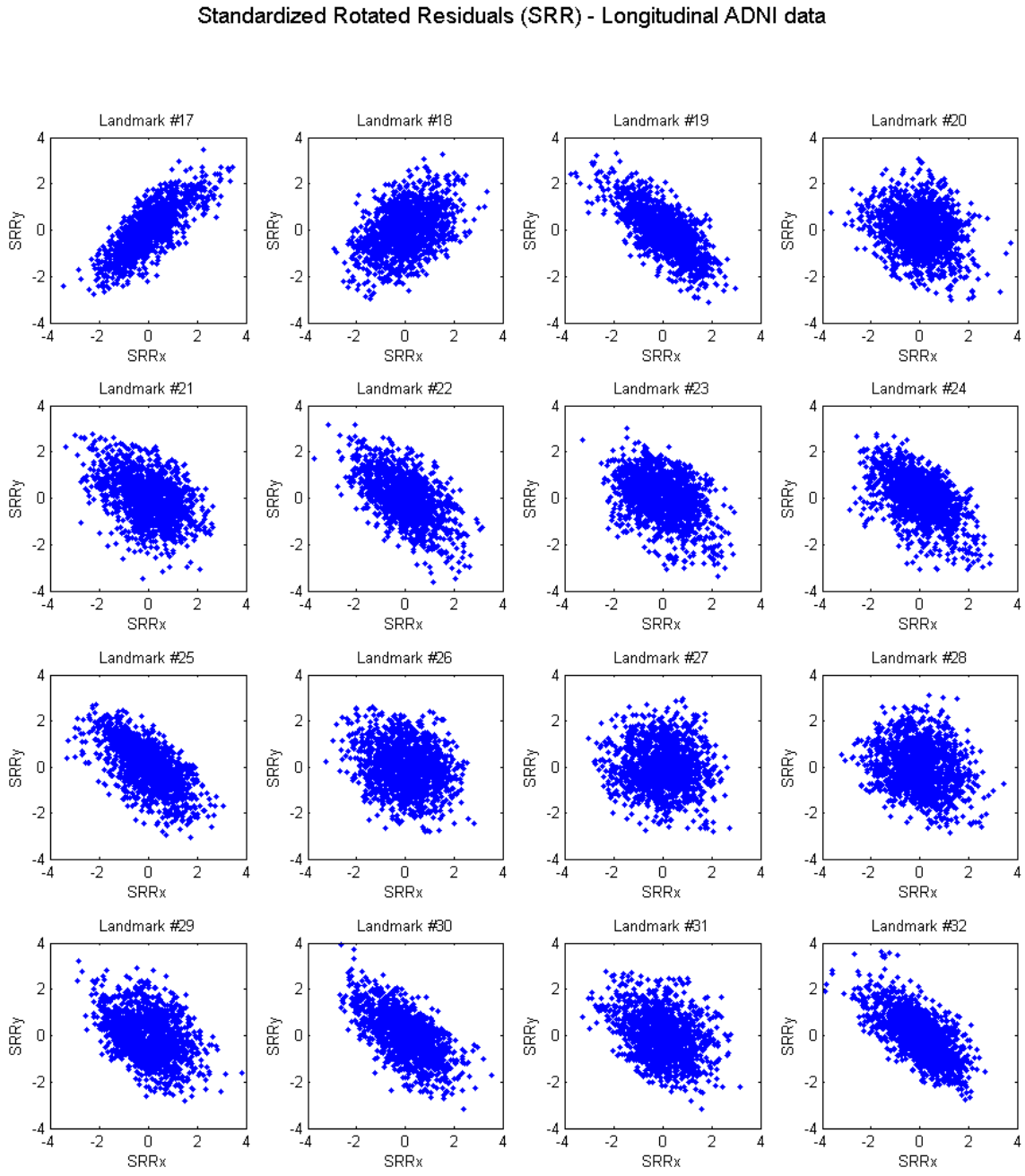

Figure 3.5: Longitudinal ADNI Data: The plots of the rotated residuals at the last 16 landmarks

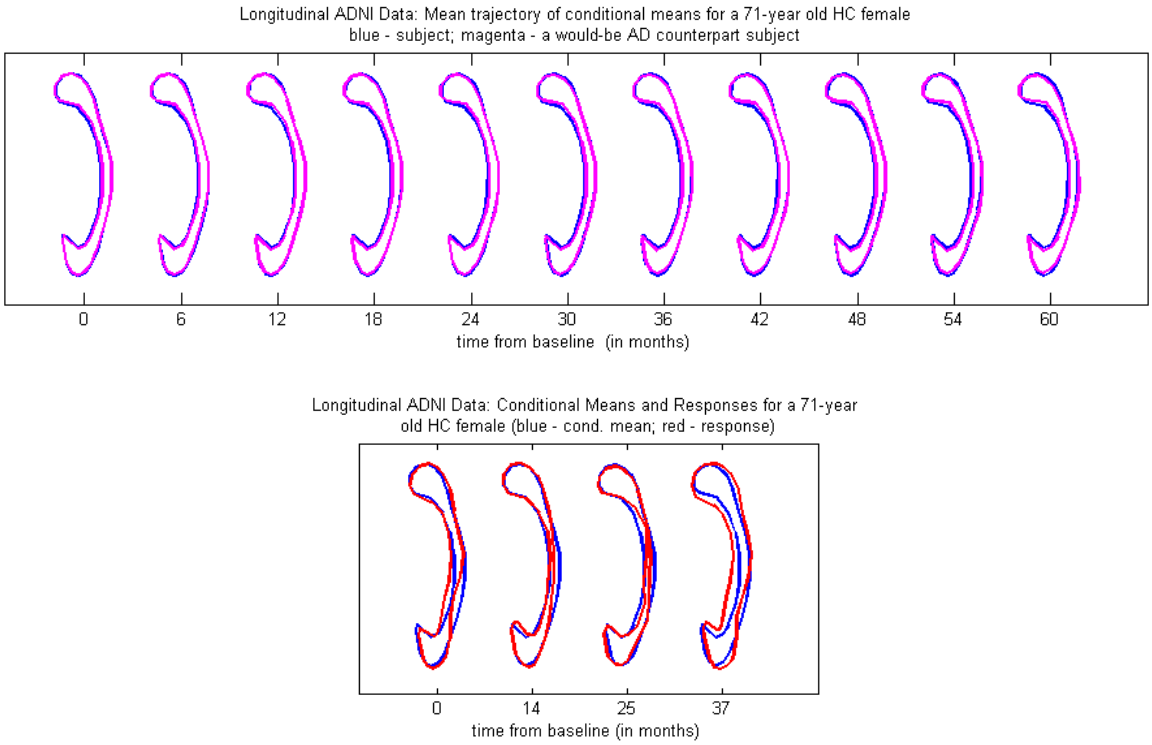

Figure 3.6: *Top:* Conditional mean trajectories of a 71-year old normal HC female and her would-be AD counterpart subject. *Bottom:* The conditional mean and observed shapes at the measurement times.

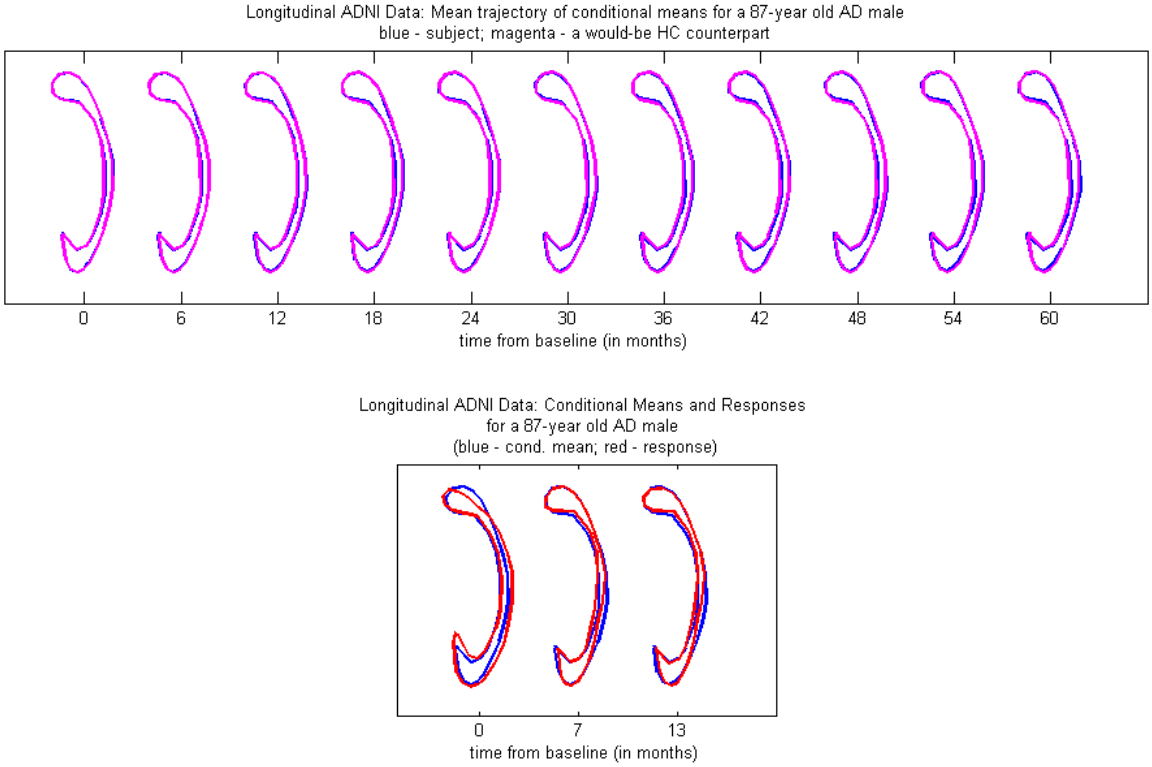

Figure 3.7: *Top:* Conditional mean trajectories of a 87-year old AD male and his would-be HC counterpart subject. *Bottom:* The conditional mean and observed shapes at the measurement times.

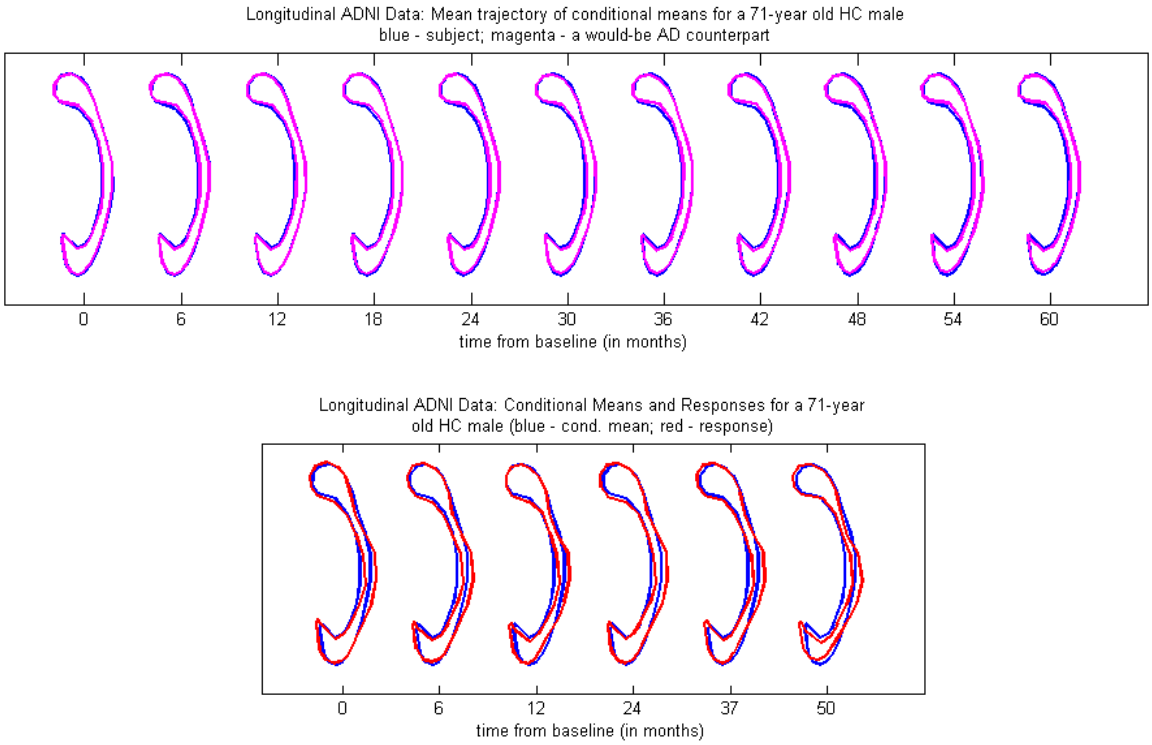

Figure 3.8: *Top:* Conditional mean trajectories of a 71-year old normal HC male and his would-be AD counterpart subject. *Bottom:* The conditional mean and observed shapes at the measurement times.

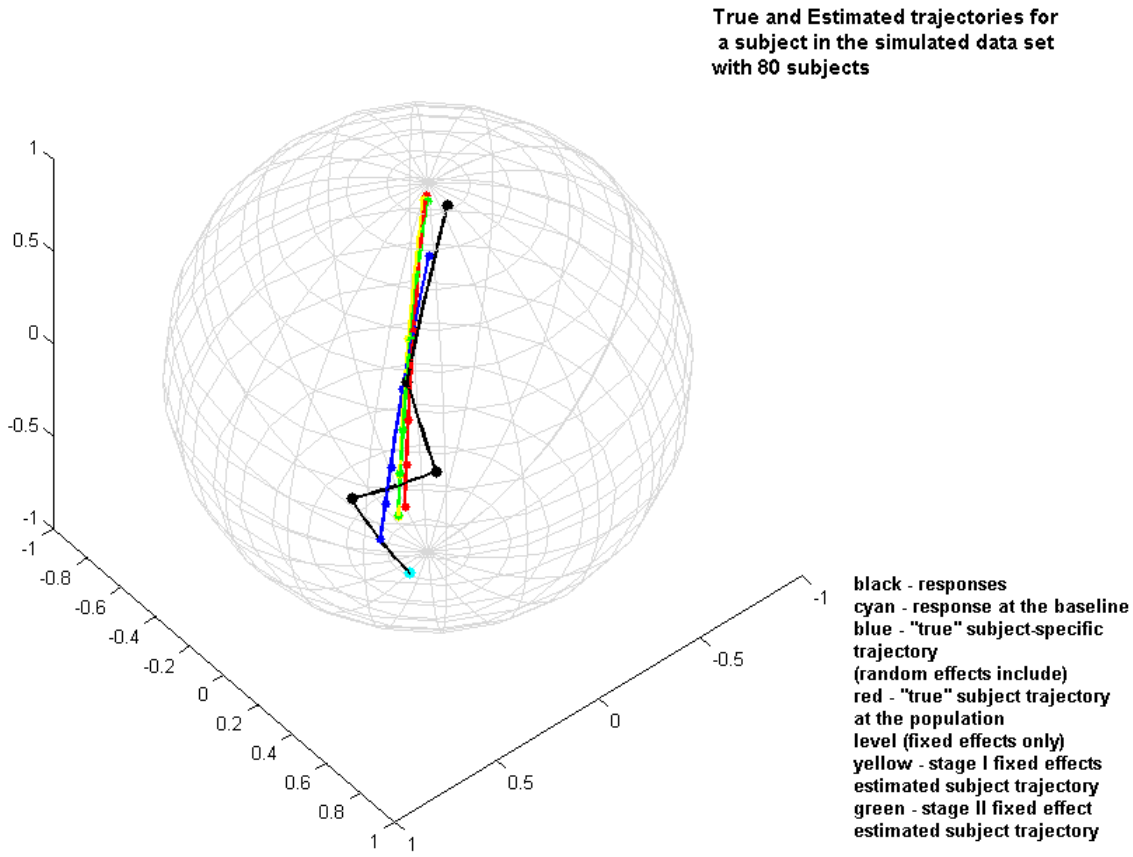

Figure 3.9: Longitudinal Simulation Study: Plots of the "true" and estimated fixed effect trajectories for a subject, based on the longitudinal simulated data on S^2.

CHAPTER 4: ASSESSING TEMPORAL AGREEMENT BETWEEN CENTRAL AND LOCAL PROGRESSION-FREE SURVIVAL TIMES

4.1 Introduction

Progression-free survival (PFS) has been a key endpoint to support licensing approval in some cancer settings (e.g. FDA (2007)). PFS is generally defined as the time from randomization until disease progression (PD) or death. In oncology clinical trials, radiologic tumor assessments are usually performed at pre-defined intervals until PD as determined by an investigator. When PFS is the primary or co-primary endpoint, it is recommended to have tumor assessments verified by an independent review committee (IRC) blinded to study treatments, especially in open-label studies (see FDA (2007)). The primary analysis of PFS for these studies may be based on the IRC's evaluation and a sensitivity analysis may be performed using the investigators' evaluation, or vice versa. It is generally considered reassuring about the lack of reader-evaluation bias if treatment effect estimates from the investigators' and IRC's evaluations agree.

Historically, less attention has been paid to the agreement (or disagreement) between the tumor assessments by the investigators and the IRC itself, compared to its impact on the treatment effect estimate. The agreement is of interest for several reasons. First, an agreement in treatment effect estimates does not necessarily mean the individual time-to-event times are the same or close. As pointed out in Amit et al. (2011), hazard ratio estimates based on the investigators' and IRC's evaluations were highly concordant in their meta-analysis, but there existed a marked degree of individual patient discordance in the timing or occurrence of PD. In an extreme case, when the PD time by the IRC's evaluation and the investigators' evaluation differ by a constant for all subjects, the

treatment effects measured by the hazard ratio are identical from both data sources, even though the evaluations of PD times never agree. Second, the median PFS time by treatment arm is often of great interest to clinicians. In the aforementioned extreme case, the median PFS times differ by the same constant between two data sources, in either treatment arm. Last, it is widely recognized that the investigators' evaluation of PD is potentially biased when the patients' treatment assignments are known to the investigators. Although the IRC is blinded to study treatments, and hence expected to reduce observation bias in the investigators' evaluation, results of the IRC's review may be subject to a new bias introduced by a special type of informative censoring: in many clinical trials, tumor assessments are discontinued when PD is determined by investigators, which precludes an IRC evaluation of PD in subsequent assessments. The IRC's evaluation of PD can be dependently censored by the investigator's evaluation of PD, which results in the IRC's evaluation of PFS be dependently censored if death is not further observed for the subject. Hence, informative censoring occurs for the IRC's evaluation of PFS endpoint. In this paper, we refer to this type of informative censoring as Reader Censoring (RC). Therefore, it is desirable to have a measure that can directly quantify the agreement between the IRC's and the investigators' evaluations, and also develop a robust estimator of an agreement measure in the presence of RC.

There are several agreement measures for assessing dependence or agreement for censored bivariate time-to-event data in the literature. The rank-based Kendall's coefficient of concordance τ (Section 4.2, Hougaard (2000)) is a measure of overall dependence. It is simple and can be estimated non-parametrically for censored data, but does not measure the degree of agreement at a single time point. In the aforementioned extreme case, the data achieves the maximum value of Kendall's τ, even though there is no agreement of PD time at all. Liu et al. (2005) provided an estimation method of the concordance correlation coefficient for time-to-event data. It is a correlation type of

measure and has the same issue as Kendall's τ. Guo and Manatunga (2009) proposed a modified weighted kappa coefficient to measure agreement between discrete bivariate survival times, but it requires discretization of continuous outcomes. Amit et al. (2011) also proposed two discrepancy rates defined as the simple proportion of subjects whose PD time is strictly greater by a reader than the other. These rates are naïve approaches by definition and do not fully utilize the temporal nature of time-to-event data. Guo et al. (2013) proposed a new agreement measure, which is formulated as the chance-corrected concordance between survival processes on the absolute distance scale.

In this paper, we propose a new method to assess temporal agreement between two time-to-event endpoints, where the two event times are assumed to have a positive probability of being identical. This method measures agreement in terms of the two event times being identical at a given time or both being greater than a given time. Overall scores of agreement over a period of time are also proposed. While the agreement measures are defined based on the underlying distributions of the two time-to-event endpoints, the proposed estimation method provides unbiased estimates of the proposed agreement measures (at a given time and overall) in the presence of RC in the observed data, for example, the IRC's evaluation of PFS. Although the focus of this paper is not to assess the impact of this agreement to the estimated treatment effect, the proposed measures can be applied to combined treatment arms as well as within each treatment arm to help assess the potential bias in estimating the treatment effect.

The development of our method is motivated by a small phase 2 head and neck cancer trial. A random subset of all randomized subjects is used here for demonstration purpose. Among 92 subjects followed-up in the trial, the local assessment yields 82 local PFS events while the central assessment gives 72 events and the number of agreed events is 35. In this paper, we will apply our development method to assess temporal

agreement of the PFS between the local and central assessments in this trial dataset.

The work presentation in this chapter is organized as follows. In Section 4.2, we present the methodology and notation for the underlying two time-to-event distributions. Furthermore, we propose time-varying agreement measures to assess the temporal agreement between two time-to-event endpoints. In Section 4.3, we propose a mixture bivariate survival model and, in Section 4.4, we describe an EM algorithm to estimate the agreement measures induced by this model. Section 4.5 studies the small-sample performance of the proposed methods through extensive simulation studies. A real data example from the aforementioned head and neck cancer study is analyzed in Section 4.6. Finally, we conclude with a discussion in Section 4.7.

4.2 Time-Varying Agreement Measures

Let T_c and T_l denote the two time-to-event endpoints under consideration (e.g. PFS per central and local assessment). The idea of defining an agreement measure is such that the analysis results based on comparing the hazard rates of the two event times should indicate similarity when the agreement between the two event times is high. In censored data analysis, the estimates of the hazard ratios, for example, fitting the Cox proportional hazards model or performing log-rank tests, are fully dependent on the behavior of the subjects at risk at each time t. In other words, the analysis relies on sufficient information regarding the number of events and the number of subjects at risk at any time t. Therefore, if T_l and T_c agree perfectly, we expect (pretending T_l and T_c to be discrete)

$$P(T_l = t) = P(T_c = t), \quad P(T_l \geq t) = P(T_c \geq t),$$

or equivalently,

$$P(T_c = t | T_l = t) = 1, \quad P(T_c > t | T_l > t) = 1.$$

This motivates us to define the following two time-varying agreement measures:

$$pA(t) = P(T_c = t|T_l = t), \quad nA(t) = P(T_c > t|T_l > t).$$

The above two quantities can be interpreted in terms of sensitivity and specificity as well. Given that the local event has not occurred prior to time t, at time t, T_l either occurs ("disease") or not ("non-disease"). Treating T_c as a diagnostic measure, we then observe that pA is the sensitivity measure while nA is the specificity measure in the usual medical diagnostic context.

We vary time t from 0 to the maximum follow-up time T_E to obtain the curves $pA(t)$ and $nA(t)$. For practical purpose, it will be convenient to use a summary quantity of the curves to assess an overall agreement between the two types of events. To this end, we define the weighted area under these curves, denoted $wAUC(pA)$ and $wAUC(nA)$, respectively, by

$$\begin{aligned} wAUC(pA) &= \int_0^{T_E} pA(t) f_{T_l}(t)dt \Big/ \int_0^{T_E} f_{T_l}(t)dt \,, \\ wAUC(nA) &= \int_0^{T_E} nA(t) f_{T_l}(t)dt \Big/ \int_0^{T_E} f_{T_l}(t)dt \,, \end{aligned}$$

where $f_{T_l}(t)$ is the probability density function of T_l. The areas $wAUC(pA)$ and $wAUC(nA)$ can be used to measure the overall agreement. The measure $wAUC(pA)$ over the entire follow-up period of time is actually the agreement probability $P(T_c = T_l)$.

4.3 Mixture Models for Estimating Time-Varying Agreements

To estimate the time-varying agreement measures $pA(t)$ and $nA(t)$ using censored data, we introduce a mixture model for the bivariate distribution of (T_c, T_l). Specifically, we assume that there is some positive chance for T_c and T_l to match. We define the

agreement probability as $p = \Pr(T_c = T_l)$ and assume $p > 0$. Thus, we model the joint distribution by considering two different situations: given $T_c = T_l$, $\Pr(T_c > t | T_c = T_l) = S(t)$; while given $T_c \neq T_l$, we assume (T_c, T_l) follows a continuous bivariate distribution with cumulative distribution function (CDF) denoted by $Q(t, s)$. Therefore, for any t and s with $t \leq s$, we can write down the joint distribution for (T_c, T_l) as

$$\begin{aligned}
\Pr(T_c \leq t, T_l \leq s) &= \Pr(T_c \leq t, T_l \leq t, T_c = T_l) + \Pr(T_c \leq t, T_l \leq s, T_c \neq T_l) \\
&= p\,[1 - S(t)] + (1-p)\Pr(T_c \leq t, T_l \leq s | T_c \neq T_l) \\
&= p\,[1 - S(t)] + (1-p)Q(t, s).
\end{aligned}$$

By symmetry, for any t and s,

$$\Pr(T_c \leq t, T_l \leq s) = p\,[1 - S(\min(t, s))] + (1-p)Q(t, s).$$

Another way of understanding this distribution is to introduce a latent Bernoulli variable $B = I(T_c = T_l)$. Then, given $B = 1$, $T_c = T_l$ follows the distribution $1 - S(t)$ while given $B = 0$, (T_c, T_l) follows the distribution $Q(t, s)$. This latent variable will be useful for developing the EM algorithm for empirical estimation later.

With practical sample sizes, estimating $S(t)$ non-parametrically may not be numerically stable. On the other hand, estimating $Q(t, s)$ nonparametrically is not feasible if there exists RC in the observed data. Therefore, in the following development, we will adopt a general class of parametric distributions for the estimation. Specifically, we will use the Weibull distribution to model $S(t)$ while assume $Q(t, s)$ to be from a copula distribution (e.g. see Joe (1997), Durante and Sempi (2010)). We take $S(t) = \exp\left[-(\lambda_0 t)^{\alpha_0}\right]$, which is the survival function of the Weibull distribution $Weibull(\alpha_0, \lambda_0)$, with shape parameter α_0 and inverse scale parameter λ_0. We assume $Q(t, s)$ is from the following copula distribution: the CDFs of the marginal distributions

of T_l and T_c given $T_l \neq T_c$ are $1 - \exp\left[-(\lambda_l t)^{\alpha_l}\right]$ and $1 - \exp\left[-(\lambda_c t)^{\alpha_c}\right]$, respectively; the joint distribution of T_l and T_c is given as

$$\left\{\Phi^{-1}\left(1 - e^{-(\lambda_c T_c)^{\alpha_c}}\right), \Phi^{-1}\left(1 - e^{-(\lambda_l T_l)^{\alpha_l}}\right)\right\} \sim N\left(0, \left[\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right]\right),$$

where $\Phi(\cdot)$ is the CDF of the univariate standard normal distribution and ρ describes their correlation. For this setting, we obtain

$$Q(t,s) = P\left(Z_1 \leq \Phi^{-1}\left(1 - e^{-(\lambda_c t)^{\alpha_c}}\right), Z_2 \leq \Phi^{-1}\left(1 - e^{-(\lambda_l s)^{\alpha_l}}\right)\right),$$

where (Z_1, Z_2) follows the above bivariate normal distribution.

With the above distributions, we can easily calculate the proposed time-varying agreement measures as

$$\begin{aligned} pA(t) &= \frac{p f_{\alpha_0,\lambda_0}(t)}{p f_{\alpha_0,\lambda_0}(t) + (1-p) f_{\alpha_l,\lambda_l}(t)} \\ &= \frac{p\alpha_0\lambda_0(\lambda_0 t)^{\alpha_0 - 1} e^{-(\lambda_0 t)^{\alpha_0}}}{p\alpha_0\lambda_0(\lambda_0 t)^{\alpha_0 - 1} e^{-(\lambda_0 t)^{\alpha_0}} + (1-p)\alpha_l\lambda_l(\lambda_l t)^{\alpha_l - 1} e^{-(\lambda_l t)^{\alpha_l}}}, \end{aligned}$$

and

$$\begin{aligned} nA(t) &= \frac{p S_{\alpha_0,\lambda_0}(t) + (1-p)\left[S_{\alpha_l,\lambda_l}(t) + S_{\alpha_c,\lambda_c}(t) - 1 + Q(t,t)\right]}{p S_{\alpha_0,\lambda_0}(t) + (1-p) S_{\alpha_l,\lambda_l}(t)} \\ &= \frac{p e^{-(\lambda_0 t)^{\alpha_0}} + (1-p)\left[e^{-(\lambda_l t)^{\alpha_l}} + e^{-(\lambda_c t)^{\alpha_c}} - 1 + Q(t,t)\right]}{p e^{-(\lambda_0 t)^{\alpha_0}} + (1-p) e^{-(\lambda_l t)^{\alpha_l}}}, \end{aligned}$$

where $f_{\alpha,\lambda}(t)$ and $S_{\alpha,\lambda}(t)$ are the density and, respectively, the survival functions of the Weibull distribution $Weibull(\alpha, \lambda)$.

As an illustrative example, we use some distribution settings from the simulation section and plot the curves of these agreement measures in Figure 4.1, where we fix $\rho = 0.5$ while varying $p = 0$, 0.2, 0.4, 0.6, 0.8, 1, or we fix $p = 0.5$ while varying ρ,

$\rho = 0$, 0.2, 0.4, 0.6, 0.8, while all the other parameters are set to the values in Tables 4.1 & 4.2. Values for pA and nA at selected times are also given in Tables 4.1 & 4.2. Furthermore, we plot the weighted area under the curves from 0 to T_E, where T_E varies from 0.5 to 3.0, $p = 0.2$, 0.4, 0.6, 0.8, and $\rho = 0.5$, in Figure 4.2. For comparison, we also plot the Kendall's τ values over $[0, T_E]$. Values for $wAUC(pA)$ and $wAUC(nA)$ for selected T_E's are also given in Table 4.3. Tables 4.1 and 4.2 and Figure 4.1 show that for given T_E, $pA(t)$ and $nA(t)$ vary over time. Table 4.3 and Figure 4.2 show that by varying T_E, $wAUC(pA)$ and $wAUC(nA)$ vary as well and become flat after most events have occurred.

4.4 Observed Likelihood and Inference

In general, the observed data from n i.i.d subjects are

$$(Y_{ci} = \min(T_{ci}, C_{ci}), Y_{li} = \min(T_{li}, C_{li}), \Delta_{ci}, \Delta_{li}), \quad i = 1, ..., n, \tag{4.1}$$

where C_{ci} and C_{li} are the respective censoring times for T_{ci} and T_{li}, $\Delta_{li} = I(T_{li} \leq C_{li})$, and $\Delta_{ci} = I(T_{ci} \leq C_{ci})$. We can allow the situation when the RC is present, in which case the observed data are

$$(Y_{ci} = \min\left(T_{ci}, \min(C_{ci}, \{(1 - \Delta_{li})\infty + \Delta_{li}T_{li}\})\right), Y_{li} = \min(T_{li}, C_{li}), \Delta_{ci}, \Delta_{li}), \tag{4.2}$$

$i = 1, \ldots, n$, where C_{ci} and C_{li} are the respective censoring times for T_{ci} and T_{li}, $\Delta_{li} = I(T_{li} \leq C_{li})$, and $\Delta_{ci} = I(T_{ci} \leq \min(C_{ci}, \{(1 - \Delta_{li})\infty + \Delta_{li}T_{li}\})$. Here, we use the convention $0 \cdot \infty = 0$.

Assuming that (C_{ci}, C_{li}) is noninformative for the joint distribution of (T_{ci}, T_{li}), we can write down the observed likelihood function as follows. In the case of data without RC, for a given observation $(y_{ci}, y_{li}, \delta_{ci}, \delta_{li})$, the likelihood contributions from

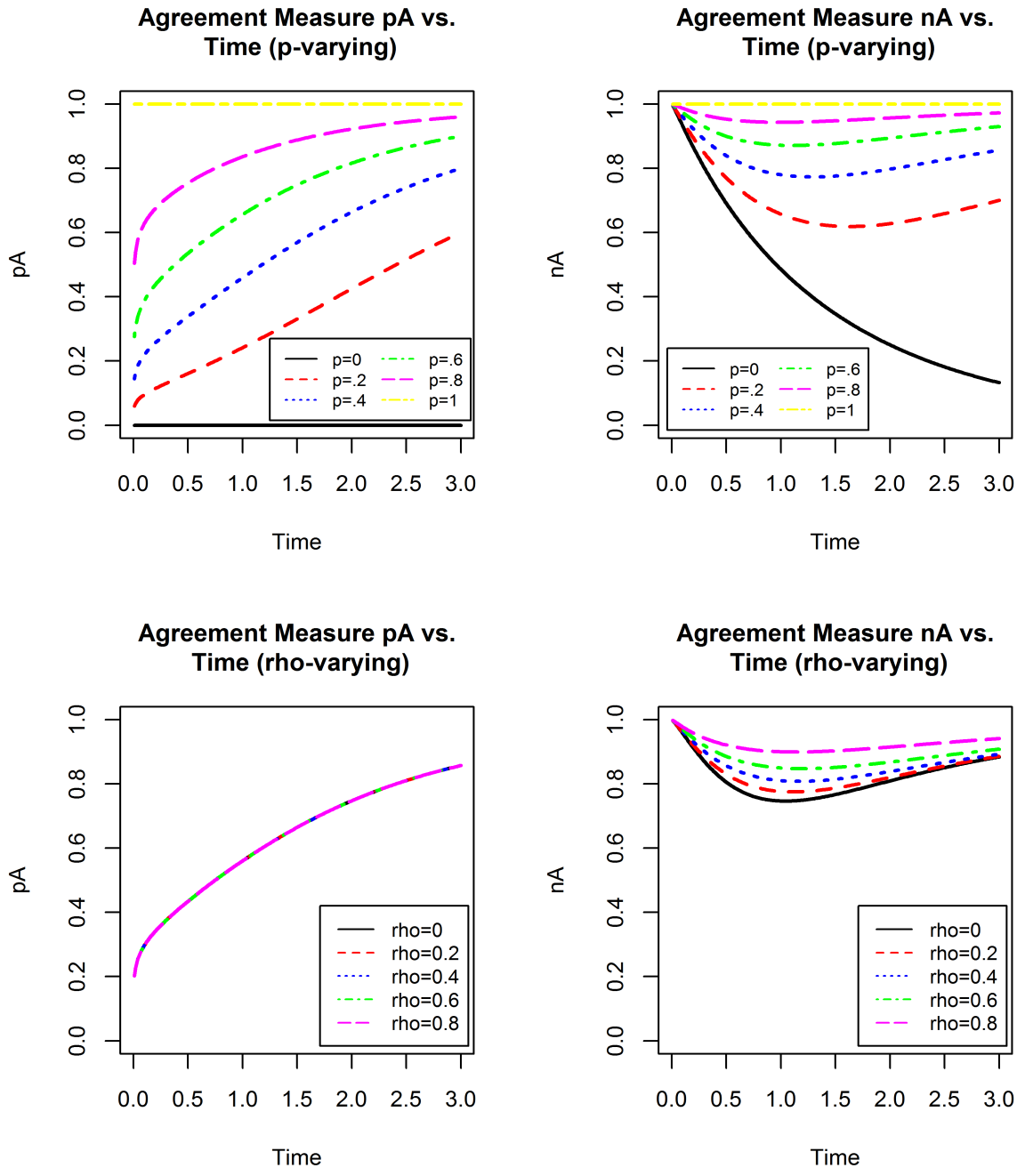

Figure 4.1: Plots of the curves of the agreement measures pA and nA. *Top:* $\rho = 0.5$ fixed and p varying, $p = 0$, 0.2, 0.4, 0.6, 0.8, 1; *bottom:* $p = 0.5$ fixed and ρ varying, $\rho = 0$, 0.2, 0.4, 0.6, 0.8; all other parameters being set to the values as in Tables 4.1 & 4.2, and $T_E = 3.0$.

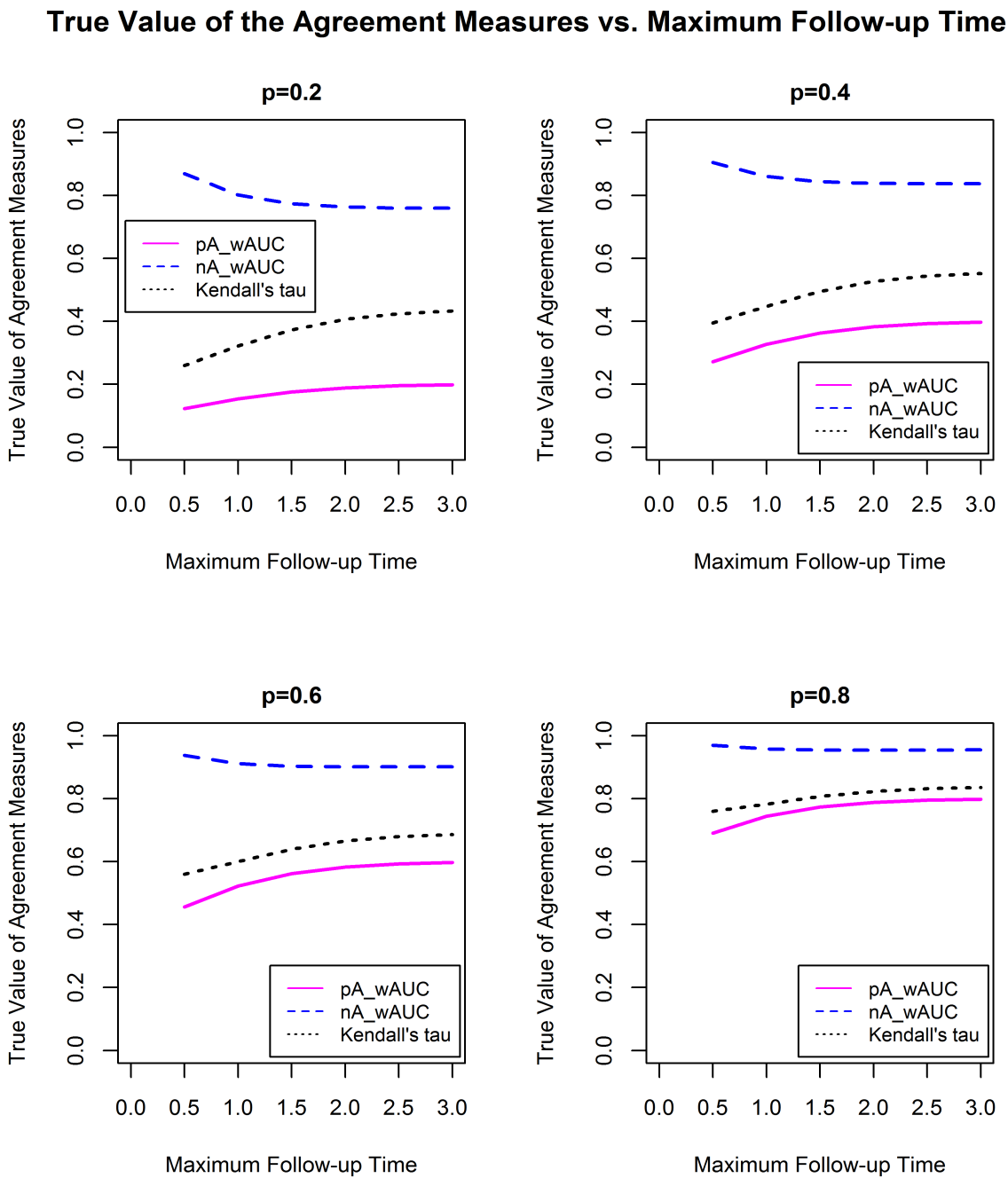

Figure 4.2: Plots of the areas, $wAUC(pA)$ and $wAUC(nA)$, under the agreement measure curves pA and nA, respectively, and the Kendall's coefficient of concordance τ.

each subject are from one of four groups:

$$\begin{aligned}
(\delta_{ci}=1,\delta_{li}=1) &: \Pr(T_c=y_{ci},T_l=y_{li})\Pr(C_c\geq y_{ci},C_l\geq y_{li}),\\
(\delta_{ci}=1,\delta_{li}=0) &: \Pr(T_c=y_{ci},T_l>y_{li})\Pr(C_c\geq y_{ci},C_l=y_{li}),\\
(\delta_{ci}=0,\delta_{li}=1) &: \Pr(T_c>y_{ci},T_l=y_{li})\Pr(C_c=y_{ci},C_l\geq y_{li}),\\
(\delta_{ci}=0,\delta_{1i}=0) &: \Pr(T_c>y_{ci},T_l>y_{li})\Pr(C_c=y_{ci},C_l=y_{li}).
\end{aligned} \tag{4.3}$$

In the case of data with RC, it comes from one of the following four groups:

$$\begin{aligned}
(\delta_{ci}=1,\delta_{li}=1) &: \Pr(T_c=y_{ci},T_l=y_{li})\Pr(\min(C_c,y_{li})\geq y_{ci},C_l\geq y_{li}),\\
(\delta_{ci}=1,\delta_{li}=0) &: \Pr(T_c=y_{ci},T_l>y_{li})\Pr(C_c\geq y_{ci},C_l=y_{li}),\\
(\delta_{ci}=0,\delta_{li}=1) &: \Pr(T_c>y_{ci},T_l=y_{li})\Pr(\min(C_c,y_{li})=y_{ci},C_l\geq y_{li}),\\
(\delta_{ci}=0,\delta_{1i}=0) &: \Pr(T_c>y_{ci},T_l>y_{li})\Pr(C_c=y_{ci},C_l=y_{li}).
\end{aligned} \tag{4.4}$$

It is clear from the above likelihood formulation that no matter if the censoring is RC or even partially RC (i.e., only some central events may be censored by local events), we only need the likelihood from the observed local/central events to make inference about the agreement measures.

Specifically, we maximize the observed likelihood function where the part concerning the joint distribution of (C_c, C_l) can be dropped. The maximization can be carried out via the EM algorithm by treating the latent status, $B_i = I(T_{ci} = T_{li})$ as missing data.

Then the complete-data likelihood function for the above expression corresponds to:

$$\begin{aligned}
(\delta_{ci}=1,\delta_{li}=1) &: \{p\, f(y_{li})\}^{B_i}\{(1-p)q(y_{ci},y_{li})\}^{1-B_i},\\
(\delta_{ci}=1,\delta_{li}=0) &: \{p\, f(y_{ci})\}^{B_i}\left\{(1-p)\int_{v>y_{li}} q(y_{ci},v)dv\right\}^{1-B_i},\\
(\delta_{ci}=0,\delta_{li}=1) &: \{p\, f(y_{li})\}^{B_i}\left\{(1-p)\int_{u>y_{ci}} q(u,y_{li})du\right\}^{1-B_i},\\
(\delta_{ci}=0,\delta_{1i}=0) &: \{p\, S(\max(y_{ci},y_{li}))\}^{B_i}\{(1-p)Q^c(y_{ci},y_{li})\}^{1-B_i},
\end{aligned} \tag{4.5}$$

where $f(t)$ is the derivative of $-S(t)$, $q(u,v)$ is $\partial^2 Q(u,v)/\partial u\partial v$, and $Q^c(t,s)$ is the probability $P(T_c>t, T_l>s)$, which is equal to $1-Q(t,\infty)-Q(\infty,s)+Q(t,s)$. The complete-data log-likelihood function becomes

$$\begin{aligned}
&\textstyle\sum_{i=1}^n \delta_{ci}\delta_{li}\{B_i\log f(y_{li})+(1-B_i)\log q(y_{ci},y_{li})\}\\
&+\textstyle\sum_{i=1}^n \delta_{ci}\delta_{li}\{B_i\log p+(1-B_i)\log(1-p)\}\\
&+\textstyle\sum_{i=1}^n \delta_{ci}(1-\delta_{li})\left\{B_i\log f(y_{ci})+(1-B_i)\log\int_{\nu>y_{li}} q(y_{ci},\nu)d\nu\right\}\\
&+\textstyle\sum_{i=1}^n \delta_{ci}(1-\delta_{li})\{B_i\log p+(1-B_i)\log(1-p)\}\\
&+\textstyle\sum_{i=1}^n (1-\delta_{ci})\delta_{li}\left\{B_i\log f(y_{li})+(1-B_i)\log\int_{\mu>y_{ci}} q(\mu,y_{li})d\mu\right\}\\
&+\textstyle\sum_{i=1}^n (1-\delta_{ci})\delta_{li}\{B_i\log p+(1-B_i)\log(1-p)\}\\
&+\textstyle\sum_{i=1}^n (1-\delta_{ci})(1-\delta_{li})\{-B_i\log S(\max(y_{ci},y_{li}))+(1-B_i)\log Q^c(y_{ci},y_{li})\}\\
&+\textstyle\sum_{i=1}^n (1-\delta_{ci})(1-\delta_{li})\{B_i\log p+(1-B_i)\log(1-p)\}.
\end{aligned} \tag{4.6}$$

For the E-step, we calculate the posterior probability of $B_i=1$, denoted by p_i,

given the observed data as follows:

$$
\begin{aligned}
\delta_{ci}=1,\delta_{li}=1 \quad &: \quad p_i = I(y_{ci}=y_{li});\\
\delta_{ci}=1,\delta_{li}=0 \quad &: \quad p_i = \frac{I(y_{li}<y_{ci})f(y_{ci})p}{I(y_{li}<y_{ci})f(y_{ci})p+\int_{\nu>y_{li}} q(y_{ci},\nu)d\nu(1-p)};\\
\delta_{ci}=0,\delta_{li}=1 \quad &: \quad p_i = \frac{I(y_{ci}<y_{li})f(y_{li})p}{I(y_{ci}<y_{li})f(y_{li})p+\int_{\mu>y_{ci}} q(\mu,y_{li})d\mu(1-p)};\\
\delta_{ci}=0,\delta_{li}=0 \quad &: \quad p_i = \frac{S(\max(y_{ci},y_{li}))p}{S(\max(y_{ci},y_{li}))p+Q^c(y_{ci},y_{li})(1-p)}.
\end{aligned}
\tag{4.7}
$$

For the M-step, we estimate the parameters in $S(t)$ by maximizing

$$
\begin{aligned}
\textstyle\sum_{i=1}^n p_i \{&\delta_{ci}\delta_{li}\log f(y_{li})+\delta_{ci}(1-\delta_{li})\log f(y_{ci})+(1-\delta_{ci})\delta_{li}\log f(y_{li})\\
&+(1-\delta_{ci})(1-\delta_{li})\left(-\log S(\max(y_{ci},y_{li}))\right)\},
\end{aligned}
\tag{4.8}
$$

and the parameters in $Q(u,v)$ by maximizing

$$
\begin{aligned}
\sum_{i=1}^n (1-p_i)\Bigg\{&\delta_{ci}\delta_{li}\log q(y_{ci},y_{li})+\delta_{ci}(1-\delta_{li})\int_{\nu>y_{li}} q(y_{ci},\nu)d\nu\\
&+\ (1-\delta_{ci})\delta_{li}\int_{\mu>y_{ci}} q(\mu,y_{li})d\mu+(1-\delta_{ci})(1-\delta_{li})\log Q^c(y_{ci},y_{li})\Bigg\}.
\end{aligned}
\tag{4.9}
$$

We estimate p by $\sum_{i=1}^n p_i/n$. We iterate between the E- and M-steps until convergence. For inference, the information is calculated based on the Louis (1982) formula and its inverse is used as the sample estimate of the asymptotic covariance of the maximum likelihood estimators.

4.5 Simulation Study

We conducted two sets of simulations, one for data with RC and the other for data without RC, to evaluate the accuracy of the parameter estimates, their associated

variance estimates, and the estimates of the agreement measures $pA(t)$ and $nA(t)$ for our proposed method. We also evaluated the performance of the true pA, nA, $wAUC(pA)$, and $wAUC(nA)$. In addition, we also evaluated the performance of Kendall's τ and its estimate. All computations were done in R.

For these sets of simulations, we assume that along the diagonal (i.e., $T_c = T_l$) (T_c, T_l) is parametrically modeled by the Weibull distribution with shape parameter α_0 and inverse scale parameter λ_0, while outside of the diagonal (i.e., $T_c \neq T_l$), T_c and T_l marginally follow Weibull distributions with parameters α_c, λ_c and, respectively, α_l, λ_l, and their joint distribution is modeled by the copula distribution based on the bivariate normal distribution with zero mean, unit variances, and correlation parameter ρ, as in the example of Section 4.3. Recall that we denoted the agreement probability, $P(T_c = T_l)$, by p. For a sample size n, we generated the simulated data as follows. We generated $B_i \in \{0, 1\}$, $i = 1, \ldots, n$, independently from $Bernoulli(p)$. The subjects with $B_i = 1$ have the event times along the diagonal, i.e. $T_{ci} = T_{li}$, while those with $B_i = 0$ have the event times outside the diagonal, i.e. $T_{ci} \neq T_{li}$. Given $B_i = 1$, the common event time $T_{ci} = T_{li}$ is generated as a random sample from a $Weibull(\alpha_0, \lambda_0)$ distribution, while given $B_i = 0$, (T_{ci}, T_{li}) is generated so that $\left(\Phi^{-1}(1 - e^{-(\lambda_c T_{ci})^{\alpha_c}}), \Phi^{-1}(1 - e^{-(\lambda_l T_{li})^{\alpha_l}})\right)$ is a random sample from the bivariate normal distribution $N_2(\mathbf{0}, \Sigma)$, with $\Sigma = \left[\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right]$. For the time-to-event without RC scenario, the simulated data $(Y_{ci}, Y_{li}, \Delta_{ci}, \Delta_{li})$ are obtained by censoring T_{ci} and T_{li} with a single censoring time C_i, i.e. $C_{ci} = C_{li}$, and an end time T_E (maximum follow-up time). That is, $Y_{ci} = \min(T_{ci}, C_i, T_E)$, $Y_{li} = \min(T_{li}, C_i, T_E)$, $\Delta_{ci} = I(T_{ci} \leq \min(C_i, T_E))$, and $\Delta_{li} = I(T_{li} \leq \min(C_i, T_E))$, $i = 1, \ldots, n$. The censoring times (C_i's) are independent of the T's and independently distributed from $Weibull(\alpha_{cens}, \lambda_{cens})$, $i = 1, \ldots, n$. For the time-to-event data with RC scenario, in addition to the time-to-event data without RC case, the central time T_{ci} is further censored by the local time T_{li}, $i = 1, \ldots, n$, as presented in Section 4.4.

For each simulation setting, we considered sample sizes of $n = 100$ and 200 and varied the agreement proportion p from 0.2 to 0.8. Furthermore, we let $\lambda_0 = 0.90$, $\alpha_0 = 1.50$, $\lambda_c = 1.35$, $\alpha_c = 1.20$, $\lambda_l = 1.40$, $\alpha_l = 1.30$, and $\rho = 0.50$. For censoring, we set $\alpha_{cens} = 1.10$, $\lambda_{cens} = 0.75$, and $T_E = 3.0$ resulting in 36.4% to 43.8% censoring rates. The parametric estimation is based on the Weibull distribution. In each case, we generated 500 datasets. For each simulated dataset, we applied the EM algorithm to obtain the maximum likelihood estimators for the model parameters and their standard errors using the Louis formula. Additionally, we calculated the agreement measures $pA(t)$ and $nA(t)$, at time points $t = 0.5$, 1.5, and 2.5. The standard errors for the latter quantities were obtained using the delta method.

Table 4.1 gives the results from the first simulation setting with RC, and Table 4.2 gives the results from the second simulation setting without RC, where column "SD" denotes the empirical standard deviations of the obtained estimates, column "SE" is the estimated standard errors by the Louis formula, and column "CP" is the 95% coverage probability. Both tables indicate that the proposed estimation and inference work well in small samples: the biases are small, the estimated standard errors agree well with the empirical standard deviations, and the coverage probabilities are close to the nominal level. The coverage probabilities of the estimated $pA(t)$ or $nA(t)$ tend to be larger for the time point at the tail but are improved when increasing the sample size. Additionally, for each agreement probability, $p = 0.2$, 0.6, 0.8, we calculated the estimates and biases of the weighted areas under the curves, $wAUC(pA)(T_E)$ and $wAUC(nA)(T_E)$, of pA and nA respectively, and Kendall's coefficient of concordance $\tau(T_E)$ as functions of the maximum follow-up time T_E, in the time-to-event data with or without RC. The results for maximum follow-up times $T_E = 0.5$, 1.5, 2.5, and 3.0. are presented in Table 4.3. These results show that the estimates of these values are quite accurate and that $wAUC(pA)$ approximates the true p for large enough T_E when

most events have occurred.

Finally, we conducted some simulations to examine the robustness of the estimated pA or nA to the model misspecification. In the simulations, instead of using the copula model for Q, we used the following distribution. Assume that T_c is marginally $Weibull(\alpha_c, \lambda_c)$ distributed and $Pr(T_l = T_c | T_c = t) = p(t)$, where $p(t)$, $0 \leq p(t) \leq 1$, for $t \geq 0$, is a given probability function. Given $T_l \neq T_c$, we assume T_l given $T_c = t$ satisfies $T_l = \exp(X)$ where X is normally distributed with mean $\log(t)$ and variance σ^2. We set $\lambda_C = 1.35$, $\alpha_C = 1.2$, $p(t) = ae^{-rt}$, with $a = 0.8$ and $r = 0.4$, and $\sigma^2 = 0.5$. For this setting, the true value of agreement probability is $p = 0.620$. For this simulation study, we considered the sample size $n = 200$, and we generated 500 datasets. The results are shown in Table 4.4. The biases of $wAUC(pA)$ and $wAUC(nA)$ are still relatively small, even though the actual joint distribution of (T_c, T_l) given $T_c \neq T_l$ is not of copula type. The agreement measure nA seems to be more sensitive to this model assumption than pA when $t \geq 1$.

4.6 Head and Neck Cancer Trial

We consider data from a relatively small phase 2 head and neck cancer clinical trial with $n = 92$ subjects (a random subset of all randomized subjects for demonstration purpose). The PFS times (in days) were recorded from two different sources, an IRC's evaluation (central) and the investigators' evaluation (local). The range of the central PFS is from 0 to 861 days and for the local PFS times is from 0 to 832 days. There are 72 central and 82 local events, respectively. For the purpose of analyzin the data, we exclude the subjects with zero central time or zero local time or zero both times from the analysis. There are 89 subjects with nonzero observed times. For the purpose of this analysis, we rescaled the times-to-event to years, i.e., we divided the event times (in

Table 4.1: Summary of simulation results from data with RC

			$n = 100$				$n = 200$			
p	Par	True	Mean	SD	SE	CP	Mean	SD	SE	CP
0.2	p	0.20	0.201	0.053	0.055	0.96	0.203	0.039	0.039	0.94
	λ_0	0.90	0.978	0.264	0.307	0.91	0.921	0.165	0.187	0.96
	α_0	1.50	1.741	0.520	0.526	0.92	1.596	0.298	0.299	0.94
	λ_C	1.35	1.348	0.220	0.225	0.95	1.359	0.156	0.159	0.95
	α_C	1.20	1.219	0.167	0.164	0.94	1.213	0.112	0.115	0.95
	λ_L	1.40	1.400	0.160	0.166	0.95	1.410	0.124	0.117	0.93
	α_L	1.30	1.332	0.148	0.146	0.94	1.316	0.101	0.101	0.95
	ρ	0.50	0.507	0.130	0.138	0.99	0.499	0.097	0.098	0.95
	$pA(0.5)$	0.161	0.168	0.054	0.058	0.96	0.162	0.038	0.039	0.95
	$pA(1.5)$	0.331	0.312	0.151	0.149	0.99	0.335	0.103	0.101	0.97
	$pA(2.5)$	0.517	0.418	0.314	0.308	0.998	0.479	0.251	0.257	0.99
	$nA(0.5)$	0.771	0.775	0.046	0.044	0.95	0.773	0.032	0.032	0.95
	$nA(1.5)$	0.620	0.611	0.153	0.158	0.96	0.622	0.113	0.114	0.96
	$nA(2.5)$	0.659	0.588	0.284	0.296	0.91	0.619	0.242	0.254	0.89
0.6	p	0.60	0.594	0.063	0.062	0.96	0.597	0.042	0.044	0.97
	λ_0	0.90	0.925	0.114	0.124	0.95	0.911	0.079	0.082	0.95
	α_0	1.50	1.544	0.226	0.216	0.94	1.532	0.151	0.148	0.93
	λ_C	1.35	1.385	0.344	0.340	0.94	1.355	0.228	0.234	0.95
	α_C	1.20	1.277	0.283	0.256	0.92	1.229	0.171	0.169	0.95
	λ_L	1.40	1.438	0.269	0.251	0.92	1.408	0.164	0.175	0.96
	α_L	1.30	1.365	0.222	0.216	0.92	1.332	0.145	0.147	0.95
	ρ	0.50	0.494	0.167	0.211	0.996	0.496	0.131	0.143	0.99
	$pA(0.5)$	0.535	0.529	0.075	0.076	0.96	0.532	0.052	0.054	0.97
	$pA(1.5)$	0.748	0.753	0.118	0.106	0.97	0.749	0.073	0.075	0.96
	$pA(2.5)$	0.865	0.802	0.221	0.182	0.98	0.832	0.148	0.133	0.97
	$nA(0.5)$	0.900	0.900	0.030	0.030	0.96	0.900	0.022	0.022	0.94
	$nA(1.5)$	0.877	0.870	0.094	0.085	0.96	0.874	0.061	0.061	0.96
	$nA(2.5)$	0.913	0.848	0.194	0.164	0.94	0.875	0.138	0.119	0.95
0.8	p	0.80	0.794	0.050	0.051	0.96	0.800	0.037	0.037	0.94
	λ_0	0.90	0.914	0.095	0.098	0.94	0.904	0.071	0.070	0.95
	α_0	1.50	1.547	0.189	0.184	0.94	1.516	0.122	0.126	0.94
	λ_C	1.35	1.436	0.478	0.537	0.95	1.377	0.312	0.365	0.96
	α_C	1.20	1.308	0.373	0.421	0.96	1.251	0.242	0.256	0.95
	λ_L	1.40	1.450	0.363	0.386	0.93	1.423	0.245	0.273	0.95
	α_L	1.30	1.424	0.338	0.340	0.93	1.365	0.215	0.222	0.95
	ρ	0.50	0.512	0.179	0.373	1.00	0.493	0.161	0.219	1.00
	$pA(0.50)$	0.754	0.745	0.067	0.069	0.95	0.751	0.051	0.049	0.95
	$pA(1.5)$	0.888	0.891	0.071	0.071	0.98	0.892	0.049	0.050	0.98
	$pA(2.5)$	0.945	0.907	0.151	0.113	0.98	0.926	0.095	0.078	0.97
	$nA(0.5)$	0.953	0.951	0.021	0.022	0.97	0.953	0.014	0.015	0.97
	$nA(1.5)$	0.948	0.941	0.059	0.054	0.97	0.947	0.039	0.037	0.96
	$nA(2.5)$	0.965	0.922	0.136	0.114	0.96	0.941	0.090	0.072	0.96

Table 4.2: Summary of simulation results from data without RC

			$n = 100$				$n = 200$			
p	Par	True	Mean	SD	SE	CP	Mean	SD	SE	CP
0.2	p	0.20	0.202	0.053	0.055	0.95	0.202	0.039	0.038	0.94
	λ_0	0.90	0.966	0.300	0.308	0.91	0.920	0.165	0.175	0.95
	α_0	1.50	1.715	0.489	0.517	0.91	1.596	0.303	0.304	0.93
	λ_C	1.35	1.354	0.169	0.174	0.95	1.362	0.125	0.122	0.93
	α_C	1.20	1.215	0.133	0.133	0.94	1.212	0.096	0.093	0.95
	λ_L	1.40	1.405	0.154	0.165	0.96	1.409	0.125	0.116	0.93
	α_L	1.30	1.329	0.142	0.145	0.95	1.315	0.102	0.101	0.95
	ρ	0.50	0.505	0.103	0.103	0.93	0.502	0.071	0.073	0.95
	$pA(0.5)$	0.161	0.166	0.053	0.056	0.97	0.162	0.038	0.038	0.94
	$pA(1.5)$	0.331	0.320	0.147	0.143	0.99	0.334	0.104	0.099	0.97
	$pA(2.5)$	0.517	0.436	0.308	0.298	0.99	0.479	0.252	0.242	0.98
	$nA(0.5)$	0.771	0.775	0.045	0.043	0.94	0.773	0.031	0.031	0.95
	$nA(1.5)$	0.620	0.616	0.148	0.150	0.94	0.621	0.112	0.109	0.94
	$nA(2.5)$	0.659	0.594	0.281	0.287	0.87	0.616	0.242	0.243	0.87
0.6	p	0.60	0.594	0.061	0.062	0.96	0.598	0.042	0.045	0.97
	λ_0	0.90	0.924	0.113	0.121	0.96	0.910	0.077	0.086	0.95
	α_0	1.50	1.539	0.221	0.216	0.95	1.533	0.151	0.155	0.93
	λ_C	1.35	1.366	0.254	0.260	0.94	1.363	0.175	0.187	0.96
	α_C	1.20	1.262	0.211	0.200	0.93	1.230	0.141	0.140	0.94
	λ_L	1.40	1.436	0.253	0.251	0.93	1.410	0.160	0.181	0.97
	α_L	1.30	1.365	0.221	0.216	0.93	1.334	0.144	0.148	0.95
	ρ	0.50	0.490	0.145	0.152	0.99	0.500	0.102	0.108	0.96
	$pA(0.5)$	0.535	0.528	0.074	0.076	0.97	0.532	0.052	0.054	0.97
	$pA(1.5)$	0.748	0.752	0.114	0.105	0.97	0.751	0.072	0.074	0.97
	$pA(2.5)$	0.865	0.807	0.207	0.180	0.98	0.835	0.140	0.133	0.97
	$nA(0.5)$	0.900	0.900	0.029	0.030	0.96	0.900	0.022	0.021	0.95
	$nA(1.5)$	0.877	0.869	0.092	0.085	0.96	0.874	0.061	0.061	0.96
	$nA(2.5)$	0.913	0.846	0.194	0.165	0.94	0.875	0.137	0.122	0.95
0.8	p	0.80	0.791	0.046	0.051	0.96	0.801	0.036	0.035	0.95
	λ_0	0.90	0.915	0.090	0.098	0.95	0.903	0.070	0.067	0.95
	α_0	1.50	1.549	0.194	0.185	0.94	1.516	0.124	0.125	0.94
	λ_C	1.35	1.387	0.387	0.404	0.95	1.372	0.262	0.273	0.95
	α_C	1.20	1.290	0.325	0.303	0.95	1.254	0.210	0.202	0.94
	λ_L	1.40	1.468	0.367	0.395	0.94	1.422	0.242	0.258	0.94
	α_L	1.30	1.398	0.334	0.323	0.93	1.361	0.215	0.218	0.95
	ρ	0.50	0.506	0.163	0.219	1.00	0.493	0.139	0.157	0.99
	$pA(0.5)$	0.754	0.744	0.067	0.069	0.95	0.752	0.051	0.048	0.95
	$pA(1.5)$	0.888	0.891	0.065	0.070	0.98	0.893	0.048	0.049	0.98
	$pA(2.5)$	0.945	0.904	0.151	0.113	0.97	0.927	0.091	0.074	0.96
	$nA(0.5)$	0.953	0.954	0.020	0.020	0.97	0.953	0.014	0.014	0.96
	$nA(1.5)$	0.948	0.941	0.060	0.053	0.97	0.947	0.038	0.036	0.97
	$nA(2.5)$	0.965	0.914	0.151	0.113	0.95	0.941	0.091	0.071	0.96

Table 4.3: True values, estimates, and biases of the weighted area under the curve of the agreement measures pA and nA, and of Kendall's coefficient of concordance τ, in both data with RC and without RC cases, for maximum follow-up times $T_E = 0.5$, 1.5, 2.5, 3.0, agreement probabilities $p = 0.2$, 0.6, 0.8, sample size $n = 200$, and all the other parameters as in Tables 4.1 and 4.2.

p	Agreement Measure	Maximum Follow-up Time	True Value	Data with RC Estimate	Data with RC Bias	Data without RC Estimate	Data without RC Bias
$p = 0.2$	$wAUC(pA)$	0.5	0.1224	0.1202	-0.0023	0.1197	-0.0027
		1.5	0.1759	0.1814	0.0055	0.1809	0.0050
		2.5	0.1955	0.1997	0.0042	0.1992	0.0038
		3.0	0.1984	0.2017	0.0034	0.2013	0.0029
	$wAUC(nA)$	0.5	0.8696	0.8695	-0.0001	0.8695	-0.0001
		1.5	0.7734	0.7721	-0.0013	0.7720	-0.0013
		2.5	0.7602	0.7580	-0.0022	0.7579	-0.0023
		3.0	0.7595	0.7570	-0.0025	0.7569	-0.0026
	τ	0.5	0.2602	0.3108	0.0506	0.4237	0.1634
		1.5	0.3727	0.3457	-0.0270	0.4530	0.0803
		2.5	0.4244	0.3476	-0.0769	0.4536	0.0291
		3.0	0.4328	0.3477	-0.0851	0.4536	0.0208
$p = 0.6$	$wAUC(pA)$	0.5	0.4557	0.4520	-0.0037	0.4520	-0.0037
		1.5	0.5615	0.5599	-0.0016	0.5602	-0.0012
		2.5	0.5931	0.5911	-0.0020	0.5917	-0.0014
		3.0	0.5975	0.5952	-0.0024	0.5958	-0.0018
	$wAUC(nA)$	0.5	0.9379	0.9376	-0.0003	0.9375	-0.0004
		1.5	0.9027	0.9018	-0.0009	0.9021	-0.0006
		2.5	0.9009	0.9005	-0.0004	0.9010	9.5e-05
		3.0	0.9012	0.9009	-0.0004	0.9014	0.0002
	τ	0.5	0.5597	0.5876	0.0280	0.6532	0.0936
		1.5	0.6392	0.6370	-0.0022	0.6925	0.0532
		2.5	0.6797	0.6390	-0.0407	0.6934	0.0138
		3.0	0.6863	0.6391	-0.0472	0.6934	0.0072
$p = 0.8$	$wAUC(pA)$	0.5	0.6906	0.6896	-0.0010	0.6900	-0.0006
		1.5	0.7735	0.7719	-0.0015	0.7726	-0.0009
		2.5	0.7954	0.7954	-0.0000	0.7959	0.0006
		3.0	0.7983	0.7984	9.4e-05	0.7990	0.0007
	$wAUC(nA)$	0.5	0.9695	0.9692	-0.0002	0.9696	9.9e-05
		1.5	0.9543	0.9551	0.0008	0.9554	0.0011
		2.5	0.9543	0.9564	0.0021	0.9566	0.0023
		3.0	0.9546	0.9568	0.0023	0.9570	0.0024
	τ	0.5	0.7596	0.7727	0.0131	0.8038	0.0443
		1.5	0.8068	0.8103	0.0035	0.8359	0.0291
		2.5	0.8315	0.8117	-0.0199	0.8369	0.0054
		3.0	0.8356	0.8118	-0.0238	0.8369	0.0014

Table 4.4: Summary of simulation results for data from the distribution described at the end of Section 4.5, for which the model is misspecified. Sample size is $n = 200$.

Maximum Follow-up Time	Agreement Measure: $wAUC(pA)$ True Value	Estimate	Bias	Rel. Bias (%)	$wAUC(nA)$ True Value	Estimate	Bias	Rel. Bias (%)
0.50	0.6963	0.7176	0.0213	3.06	0.9763	0.9709	-0.0055	-0.56
1.00	0.6824	0.6727	-0.0098	-1.43	0.9596	0.9338	-0.0258	-2.69
1.50	0.6677	0.6511	-0.0167	-2.50	0.9524	0.9063	-0.0461	-4.84
2.00	0.6550	0.6416	-0.0134	-2.04	0.9499	0.8910	-0.0589	-6.20
2.50	0.6452	0.6379	-0.0073	-1.13	0.9492	0.8842	-0.0650	-6.85
3.00	0.6383	0.6367	-0.0016	-0.25	0.9491	0.8817	-0.0674	-7.10

days) by 365.25. The event times are now in the range [0,3], similar to the time range in our simulation studies. The maximum follow-up time is $T_E = 2.36$. A descriptive data summary is shown in Table 4.5. The probability of exact agreement of T_c and T_l is at least 30%.

We fit the model described in Section 4.3. The estimates, standard deviations, and 95% confidence intervals for the parameters $p, \lambda_0, \alpha_0, \lambda_c, \alpha_c, \lambda_l, \alpha_l, \rho$ are shown in Table 4.6. The estimate of the agreement probability p is $\hat{p} = 0.413$. The estimates of the agreement measures $pA(t)$ and $nA(t)$, together with their 95% pointwise confidence bands, are shown in Table 4.7 and displayed in Figure 4.3. The weighted area under the estimated curves $\widehat{pA}(t)$ and $\widehat{nA}(t)$ are 0.412 and 0.843. The $wAUC$s are calculated over the study duration, that is, over the time period from 0 to the last event/censoring time, $T_E = 2.36$. Table 4.8 shows the estimated values of $wAUC(pA)$, $wAUC(nA)$, and Kendall's coefficient of concordance τ. The statistic τ only reflect the rank correlation between the two outcomes, T_c and T_l, at any time, while our measures $pA(t)$ and $nA(t)$ give the exact agreement probabilities at any time t. Our measures are more informative than the Kendall's coefficient τ.

The results indicate that both the local and central times agreed well at the beginning and end of this study. The relatively higher agreement at the end of the study

Table 4.5: Head and neck cancer trial data information at times $t = 0.5, 1.5, 2.5$. ($n = 89$)

Time		Average percentage of subjects (%)		
		$Y_C = Y_L$	$Y_C < Y_L$	$Y_C > Y_L$
0.5	Both Censored	28.09	0.00	1.12
	Central Event Only	1.12	11.24	1.12
	Local Event Only	6.74	0.00	6.74
	Both Events	31.46	12.36	0.00
1.5	Both Censored	4.49	0.00	1.12
	Central Event Only	1.12	3.37	1.12
	Local Event Only	11.24	0.00	7.87
	Both Events	37.08	24.72	7.86
2.5	Both Censored	2.25	0.00	1.12
	Central Event Only	1.12	2.25	1.12
	Local Event Only	11.24	0.00	4.49
	Both Events	39.33	25.84	11.24

is mostly due to death for subjects without either local or central PD. The local and central diagnosis for the non-event cases were similar. The plots in Figure 4.3 indicate a large variability of the agreements over time; thus, it is necessary to account for this temporally variable agreement in any further analysis.

Table 4.6: Parameter estimates, standard deviations, and 95% confidence intervals of the parameters for the head and neck cancer trial data. (sample size $n = 89$).

Parameter	PFS			
	Estimated Value	Standard Deviation (Louis)	95% Confidence Interval lower bound	upper bound
p	0.413	0.053	0.314	0.520
λ_0	2.209	0.419	1.522	3.205
α_0	0.998	0.129	0.774	1.286
λ_C	1.284	0.194	0.955	1.727
α_C	1.088	0.140	0.845	1.401
λ_L	1.463	0.139	1.215	1.762
α_L	1.664	0.192	1.327	2.087
ρ	0.612	0.118	0.330	0.794

Table 4.7: Estimated values, standard deviations, and 95% confidence intervals of the agreement measures pA and nA for the head and neck cancer trial data..

Agreement Measure	Time	PFS Estimated Value	PFS Standard Deviation (Louis)	PFS 95% Confidence Interval lower bound	PFS 95% Confidence Interval upper bound
pA	0.50	0.320	0.059	0.217	0.445
	1.00	0.264	0.072	0.147	0.426
	1.50	0.359	0.193	0.097	0.744
	2.00	0.599	0.391	0.058	0.973
	2.36	0.800	0.377	0.038	0.998
nA	0.50	0.808	0.044	0.708	0.880
	1.00	0.770	0.078	0.585	0.888
	1.50	0.826	0.124	0.465	0.963
	2.00	0.924	0.128	0.254	0.998
	2.36	0.972	0.074	0.136	0.999

Table 4.8: Estimated values of the weighted area under the agreement measures pA and nA, and of the Kendall's coefficient of concordance τ for the head and neck cancer trial data.

Maximum Follow-up Time	PFS Estimated Values $wAUC(pA)$	PFS Estimated Values $wAUC(nA)$	PFS Estimated Values τ
0.50	0.512	0.890	0.617
1.00	0.425	0.849	0.587
1.50	0.410	0.842	0.600
2.00	0.411	0.843	0.609
2.36	0.412	0.843	0.613

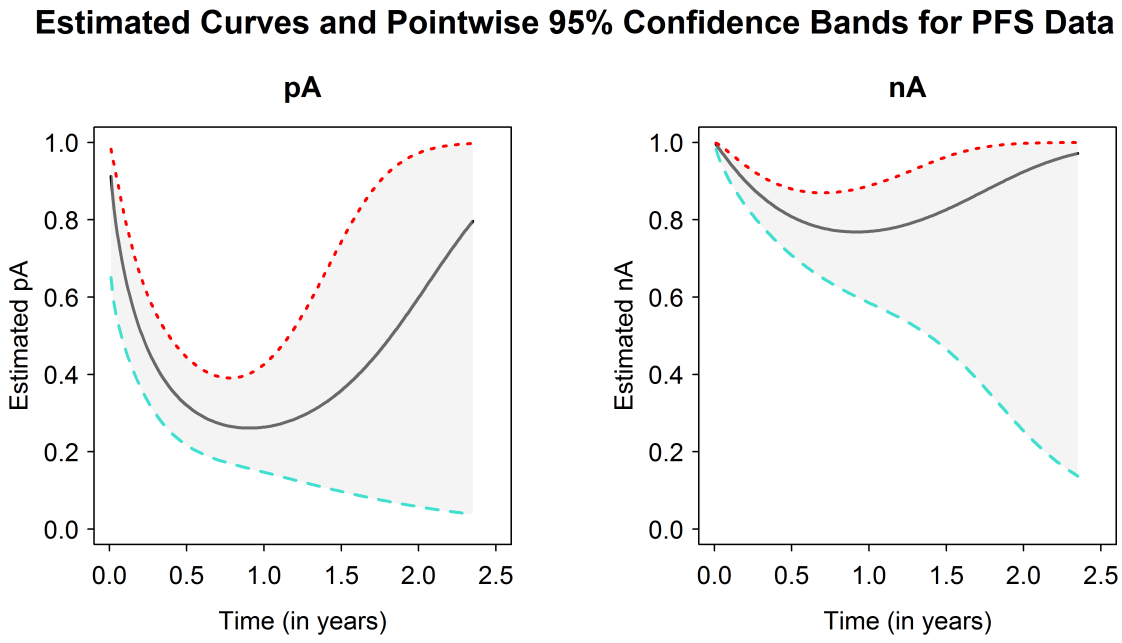

Figure 4.3: Plots of the estimated curves of the agreement measures pA and nA, together with the pointwise 95% confidence bands, over the study duration for PFS data. The weighted AUCs are $wAUC(pA) = 0.412$ and $wAUC(nA) = 0.843$.

4.7 Discussion

We have proposed a new method to assess agreement between two time-to-event endpoints, where the two event times are assumed to have a positive probability of being identical. Such an assumption is motivated by type of data we targeted. For example, the head and neck cancer trial data show that the probability of exact agreement of the two event times is at least 30%. We can view our model as a two-pattern mixture model, so $T_c = T_l$ is one pattern. Although, it is not uncommon that the two evaluations are based on measurements which are usually performed at predefined time intervals, there might be a large amount of variation in the two observed event times, with $T_c \neq T_l$, with T_c and T_l continuous random variables, is another pattern (see head and neck cancer trial data). Other approaches may be envisioned for modeling same type of

data. For example, one can set up the two event times in a discrete failure time setting, by considering a model arising from a single copula model, which is then discretized according to the observation windows and estimate accordingly. However, such an approach is not the scope of this paper. Nonetheless, it is an interesting alternative approach and can be addressed in a separate research.

Our method measures agreement in terms of two event times of being identical at a given time point or both being greater than a given time. The proposed agreement measures are based on underlying joint distribution of the time-to-event endpoints. We estimate the proposed agreement measures using a mixture parametric distribution based on the Weibull distributions, which are general enough to cover a wide range of applications. However, it is of interest to further relax this parametric assumption while consider even more flexible or nonparametric distributions for both $S(t)$ and $Q(t, s)$ in future work. Sieve estimation based on splines may be useful for carrying out inference, but will increase the computation complexity. Although our agreement measures are defined as the conditional probability of the central event given the local event, this definition can be similarly defined as the reverse conditional probability or the summation of both conditional probabilities, especially when there is no preference to either event.

Our approach is fully likelihood-based and the informative censoring RC is naturally accounted for in the likelihood formulation. So, RC has no impact on the validity of our method. As shown in the simulation study, the relative biases of the proposed agreement measures for data without RC and data with RC are very similar. Our method yields robust estimators of the agreement measures in the presence of RC. However, if the censoring is truly informative (depends on potential outcome value), modeling mechanism is necessary to obtain a valid inference.

The proposed method can be directly applied to other commonly used endpoints

for oncology clinical trials, for example, time-to-progression and disease free survival (FDA (2007)), to assess their agreement between IRC's and investigators' evaluation. In addition, by jointly modeling the IRC's and investigators' assessments, the proposed estimation method can provide less-biased estimates of the parameters in the assumed underlying distribution than those based on IRC's assessment alone. One application is to apply the method within each treatment arm and then make inference about treatment effect with the estimated underlying distribution. High values (e.g., $> 80\%$) of the measure $wAUC(pA)$ in both treatment and control arms provide a high assurance of assessing treatment effects using the investigator's evaluations only, which implies investigator's reading data are robust and reliable to estimate treatment effects; on the contrast, lower values require the need for the IRC's readings in clinical trials. Another application is for event projection. The primary analysis in most oncology trials are event driven (i.e., based on planned number of events). When the primary analysis is based on IRC's evaluation, event modeling based on the estimated underlying distribution of the IRC's data from joint modeling of IRC's and investigator's data provide more accurate projection of future events than that from modeling of IRC's data alone.

The proposed temporal agreement measurements are potentially useful for future trials when central assessment is needed for auditing purposes, since time-to-event outcomes such as PFS, are usually subject to measurement error. For example, it was reported in Dodd et al. (2011) that a discrepancy in PFS times between local and central assessment could range from 23% to 36%. Recently, there has been some development to address this discrepancy, see Dodd et al. (2011; 2013). With the development of our temporal agreement measurements, we will be able to not only capture the overall agreement, but also assess the temporal discrepancy between these two assessments; therefore, the differences in the analyses based on local and central assessments will

be quantified more precisely. Furthermore, the proposed temporal agreement measurements will also be useful to help design more effective and efficient audit strategies. For instance, if the disagreement occurs most often for long survivals, an audit should then be applied to those subjects. In future work, we will examine how to incorporate this temporal agreement measure into designing an audit strategy to correct bias due to imprecise local assessments.

CHAPTER 5: CONCLUSIONS

We have developed a general statistical framework for intrinsic regression models of responses valued in a Riemannian symmetric space in general, and Lie groups in particular, and their association with a set of covariates in a Euclidean space. The intrinsic regression models are based on the generalized method of moment 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 Riemmanian 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 ILSE, $(\widehat{\mathrm{q}}_I, \widehat{\boldsymbol{\beta}}_I)$, of $(\mathrm{q}, \boldsymbol{\beta})$, in the Stage I of the estimation process, and a one-step procedure to search for the efficient estimator $(\tilde{\mathrm{q}}_E, \tilde{\boldsymbol{\beta}}_E)$ in Stage II. Our simulation study and real data analysis demonstrate that the relative efficiency of the Stage II estimator improves as the sample size increases.

There are still many outstanding issues for further research. One major issue is to construct goodness-of-fit statistics for testing for possible model misspecifications in (2.6). Another important issue is to develop diagnostic measures for assessing the influence of individual observations in the semiparametric regression for manifold-valued data. Third, there is some interest in developing regression models where both responses and covariates lie on the same manifold (or different manifolds) (Chang 1986; 1989, Downs 2003, Rosenthal et al. 2014). Mathematically, most developments above

are still valid with some modifications. Fourth, there is a great interest in developing nonparametric Bayesian regression models for manifold-valued response data and multiple covariates (Bhattacharya and Dunson 2010; 2012). Fifth, smoothing spline methods and local polynomial regression have been developed for the non-parametric estimation of regression functions for manifold-valued response data given a continuous covariate in Euclidean space (Samir et al. 2012, Su et al. 2012, Muralidharan and Fletcher 2012, Machado and Leite 2006, Machado et al. 2010, Yuan et al. 2012). We will develop an intrinsic local polynomial regression (or smoothing spline) estimate for manifold-valued responses with multiple covariates and examine its asymptotic properties. These extensions are of great interest for our future research.

We extended the framework of intrinsic regression models to fixed effect models for the analysis of manifold-valued data from longitudinal data. We performed a detailed longitudinal data analysis of a corpus callosum shape data from the Alzheimer's Disease Neuroimaging Initiative database. Our investigation reveals that there is an association between the the CC shape and the Alzheimer's disease, which depends on the age, while gender seems to have little or no relevance. The splenium seems to be less rounded and the isthmus thinner in patients with Alzheimer's disease than in normal healthy patients. We also found that the evolution of the CC shape over time is different in patients with AD than normal healthy patients. A large variability between the within-subject response and the corresponding conditional mean at the population level was detected as well. Therefore, developing intrinsic random effects models for manifold-valued is of a great interest and need. Our ongoing research efforts aim to develop such models.

We have proposed a new method to assess agreement between two time-to-event endpoints, where the two event times are assumed to have a positive probability being identical. Our method measures agreement in terms of two event times of being identical

at a given time point or both being greater than a given time. The proposed agreement measures are based on underlying joint distribution of the time-to-event endpoints. We estimate the proposed agreement measures using a mixture parametric distribution based on the Weibull distributions, which are general enough to cover a wide range of applications. However, it is of interest to further relax this parametric assumption while consider even more flexible or nonparametric distributions for both $S(t)$ and $Q(t, s)$ in future work. Sieve estimation based on splines may be useful for carrying out inference, but will increase the computation complexity. Although our agreement measures are defined as the conditional probability of the central event given the local event, this definition can be similarly defined as the reverse conditional probability or the summation of both conditional probabilities, especially when there is no preference to either event.

The proposed method can be directly applied to other commonly used endpoints for oncology clinical trials, for example, time-to-progression and disease free survival to assess their agreement between IRC's and investigators' evaluation. In addition, by jointly modeling the IRC's and investigators' assessments, the proposed estimation method can provide less-biased estimates of the parameters in the assumed underlying distribution than those based on IRC's assessment alone. One application is to apply the method within each treatment arm and then make inference about treatment effect with the estimated underlying distribution. One other application is for event projection. The primary analysis in most oncology trials are event driven (i.e., based on planned number of events). When the primary analysis is based on IRC's evaluation, event modeling based on the estimated underlying distribution of the IRC's data from joint modeling of IRC's and investigator's data provide more accurate projection of future events than that from modeling of IRC's data alone.

The proposed temporal agreement measurements are potentially useful for future

trials when central assessment is needed for auditing purposes, since time-to-event outcomes such as PFS, are usually subject to measurement error. With the development of our temporal agreement measurements, we will be able to not only capture the overall agreement, but also assess the temporal discrepancy between these two assessments; therefore, the differences in the analysis based on local and central assessments will be quantified more precisely. Furthermore, the proposed temporal agreement measurements will also be useful to help design more effective and efficient audit strategies.

APPENDIX : TECHNICAL DETAILS FOR CHAPTER 2

A.1 Differential Geometry

A.1.1 Technical Details

Riemannian Metric, Distance, and Geodesics

A Riemannian manifold $(\mathcal{M}, \mathrm{m})$ is a smooth manifold $\mathcal{M}$ together with a metric m. The $\mathrm{m} = (\mathrm{m_p})_{\mathrm{p}\in\mathcal{M}}$ is a family of inner products m_p on the tangent space $T_\mathrm{p}\mathcal{M}$ of $\mathcal{M}$ at $\mathrm{p} \in \mathcal{M}$, and for any smooth vector fields X and Y on an open set $U \subset \mathcal{M}$, the real valued map $\mathrm{p} \mapsto m_\mathrm{p}(X_\mathrm{p}, Y_\mathrm{p})$ is smooth on U. Let $d_\mathcal{M}$ be the dimension of $\mathcal{M}$. The tangent space $T_\mathrm{p}\mathcal{M}$ is isomorphic to $R^{d_\mathcal{M}}$. For a local chart (U, ϕ), U is an open subset of $\mathcal{M}$ and there is a homeomorphism $\phi : U \to \phi(U) \subset R^{d_\mathcal{M}}$, where $\phi(U)$ is an open set containing $\phi(\mathrm{p}) = \mathbf{t} = (t^1, \ldots, t^{d_\mathcal{M}})$. Let ∂_j denote the tangent vector with respect to the coordinate curves $\partial/\partial t^j$ for $j = 1, \ldots, d_\mathcal{M}$. The vector fields $\frac{\partial}{\partial \mathbf{t}} = (\partial_1, \ldots, \partial_{d_\mathcal{M}})$ induce a basis at each of the tangent spaces $T_{\phi^{-1}(\mathbf{t})}\mathcal{M}$ for $\mathbf{t} \in \phi(U)$. In this basis, the metric can be expressed by a symmetric positive definite matrix $\mathrm{M}_\phi(\mathbf{t}) = [\mathrm{m}_{jk}(\mathbf{t})]$, where $\mathrm{m}_{j,k}(\mathbf{t}) = m_{\phi^{-1}(\mathbf{t})}(\partial_j, \partial_k)$. The matrix $\mathrm{M}_\phi(\mathbf{t})$ is called the local representation of the Riemannian metric in the chart (U, ϕ), and for any $\mathrm{p} \in U$, the inner product of $\mathbf{v}$ and $\mathbf{w} \in T_\mathrm{p}\mathcal{M}$ is given by $m_p(\mathbf{v}, \mathbf{w}) = \tilde{\mathbf{v}}^\top \mathrm{M}_\phi(\phi(\mathrm{p}))\tilde{\mathbf{w}}$, where $\tilde{\mathbf{v}} = (v^1, \ldots, v^{d_\mathcal{M}})^\top$ and $\tilde{\mathbf{w}} = (w^1, \ldots, w^{d_\mathcal{M}})^\top$ are the representations of $\mathbf{v}$ and $\mathbf{w}$, respectively, in the chart (U, ϕ), i.e., $\mathbf{v} = \sum_{j=1}^{d_\mathcal{M}} v^j \partial_j$.

The length $\ell(\gamma)$ of a C^1-curve $\gamma : [t_0, t_1] \to \mathcal{M}$ on a Riemannian manifold $\mathcal{M}$ is defined by $\ell(\gamma) = \int_{t_0}^{t_1} \sqrt{\mathrm{m}_{\gamma(t)}(\gamma'(t), \gamma'(t))}\, dt$. The length of a continuous, piecewise smooth curve on $\mathcal{M}$ is defined as the sum of the lengths of its smooth pieces. The *geodesic distance* $\mathrm{dist}_\mathcal{M}(\mathrm{p}, \mathrm{q})$ between p and $\mathrm{q} \in \mathcal{M}$ is defined as the infimum of $L(\gamma)$ taken over all continuous, piecewise smooth curves $\gamma : [a, b] \to \mathcal{M}$ with $\gamma(a) = \mathrm{p}$ and $\gamma(b) = \mathrm{q}$. The Riemannian manifold $(\mathcal{M}, \mathrm{dist}_\mathcal{M})$ is a metric space and geodesics are

then, by definition, the locally distance-minimizing paths. The geodesics are the curves satisfying the second order differential system in the chart (U, ϕ) given by

$$\ddot{\gamma}^j + \sum_{l',l} \Gamma^j_{l'l} \dot{\gamma}^{l'} \dot{\gamma}^l = 0,$$

where $\Gamma^j_{l'l} = 0.5 \sum_{j'} \mathrm{m}^{j'j} (\partial_{l'} \mathrm{m}_{li} + \partial_{j'} \mathrm{m}_{ll'} - \partial_l \mathrm{m}_{j'l'})$ are the Christoffel symbols of the first kind.

For $\mathrm{p} \in \mathcal{M}$ and $\mathbf{v}$ in $T_\mathrm{p}\mathcal{M}$, there exists a unique geodesic $\gamma = \gamma(\cdot; \mathrm{p}, \mathbf{v}) : I \to \mathcal{M}$ satisfying $\gamma(0) = \mathrm{p}$ and $\gamma'(0) = V$, where I is a maximal open interval in R containing 0. Moreover, γ depends smoothly on both p and $\mathbf{v}$. In general, I may not be all of R. The manifold is said to be *geodesically complete* if the maximal interval I is the entire real line R for all geodesics. For example, the Euclidean space R^n and the unit sphere S^n are geodesically complete manifolds, while $R \setminus \{0\}$ is not. The Hopf-Rinow-De Rham theorem states that a geodesically complete Riemannian manifold is complete as a metric space with the distance induced by the Riemannian metric, and that there always exists at least one distance minimizing geodesic between any two points of the manifold.

Exponential and Logarithmic Maps

For a general Riemannian manifold, given a vector $\mathbf{v}$ in $T_\mathrm{p}\mathcal{M}$ and a real number $\tau \in R$, we have that $\gamma(t; \mathrm{p}, \tau\mathbf{v}) = \gamma(t\tau; \mathrm{p}, \mathbf{v})$, for all $t \in R$ with $t\tau$ in the definition domain of $\gamma(\cdot; \mathrm{p}, \mathbf{v})$. Therefore, for a tangent vector $\mathbf{v} \in T_\mathrm{p}\mathcal{M}$ with $\|\mathbf{v}\|_\mathrm{p} := (m_\mathrm{p}(\mathbf{v}, \mathbf{v}))^{1/2} < r(\mathrm{p})$ for some small $r(\mathrm{p}) > 0$, the geodesic starting from p and with initial velocity $\mathbf{v}$ is defined on an interval containing $[0, 1]$. The manifold *exponential map* at a point $\mathrm{p} \in \mathcal{M}$, $\mathrm{Exp}^\mathcal{M}_\mathrm{p} : B_\mathrm{p}(\mathbf{0}, r(\mathrm{p})) \to \mathcal{M}$ is defined by $\mathrm{Exp}^\mathcal{M}_\mathrm{p}(V) = \gamma(1; \mathrm{p}, V)$ for $V \in B_\mathrm{p}(0, r(\mathrm{p}))$, where, $B_\mathrm{p}(\mathbf{0}, r(\mathrm{p}))$ denotes the ball of radius $r(\mathrm{p})$ centered at the origin in $T_\mathrm{p}\mathcal{M}$. The

exponential map $\mathrm{Exp}_{\mathrm{p}}^{\mathcal{M}}$ is a locally smooth diffeomorphism around $\mathbf{0} \in T_{\mathrm{p}}\mathcal{M}$, i.e. there is a $r_*(\mathrm{p}) \in (0, r(\mathrm{p}))$ such that $\mathrm{Exp}_{\mathrm{p}}^{\mathcal{M}}$ is a diffeomorphism from $B_{\mathrm{p}}(\mathbf{0}, r_*(\mathrm{p}))$ into $\mathcal{M}$. The inverse map is denoted by $\mathrm{Log}_{\mathrm{p}}^{\mathcal{M}}$ and it provides normal coordinates on $\mathcal{M}$ around p.

For $\mathrm{q} \in \mathrm{Exp}_{\mathrm{p}}^{\mathcal{M}}(B_{\mathrm{p}}(\mathbf{0}, r_*(\mathrm{p})))$, the geodesic distance from p to q can be expressed as $\mathrm{dist}_{\mathcal{M}}(\mathrm{p}, \mathrm{q}) = \|\mathrm{Log}_{\mathrm{p}}\mathrm{q}\|_{\mathrm{p}}$, and thus $\mathrm{Exp}_{\mathrm{p}}^{\mathcal{M}}(B_{\mathrm{p}}(\mathbf{0}, r_*(\mathrm{p})))$ is the ball $B^{\mathcal{M}}(\mathrm{p}, r^*(\mathrm{p}))$ in $\mathcal{M}$, with the induced distance, centered at p of radius $r^*(\mathrm{p})$. As the tangent space $T_{\mathrm{p}}\mathcal{M}$ is isomorphic to $R^{d_{\mathcal{M}}}$, the logarithmic map $\mathrm{Log}_{\mathrm{p}}$ provides a local chart near p. If the tangent space of $\mathcal{M}$ at p is endowed with an orthonormal basis, then such a chart is called a *normal chart* and the coordinates are called *normal coordinates*.

Cut Locus and Radius of Injectivity

From now on, we will assume that the manifold $\mathcal{M}$ is geodesically complete, and thus the exponential map Exp_p is defined on the entire tangent space $T_{\mathrm{p}}\mathcal{M}$. A geodesic $\gamma(t; \mathrm{p}, \mathbf{v})$ is either always minimizing the distance from p to $\gamma(t; \mathrm{p}, \mathbf{v})$ from $t = 0$ to ∞, or it is minimizing up to a finite point t_0 and no more thereafter. In the latter case, the point $\gamma(t_0; \mathrm{p}, \mathbf{v})$ is called a *cut point* for the geodesic $\gamma(\cdot; \mathrm{p}, \mathbf{v})$ and the tangent vector $t_0\mathbf{v}$ is called a *tangential cut point*. The set of cut points of all geodesics starting from p is called the *cut locus of* p and denoted by $C(\mathrm{p}) \subset \mathcal{M}$. The set of corresponding tangent vectors is called the *tangential cut locus of* p and denoted by $\mathcal{C}(\mathrm{p}) \subset T_{\mathrm{p}}\mathcal{M}$. We have $C(\mathrm{p}) = \mathrm{Exp}_{\mathrm{p}}(\mathcal{C}(\mathrm{p}))$ and thus, the maximal definition domain of the normal chart centered at p is the domain $\mathcal{D}(\mathrm{p}) \subset T_{\mathrm{p}}\mathcal{M}$ containing $\mathbf{0}$ and bounded by $\mathcal{C}(\mathrm{p})$. The domain $\mathcal{D}(\mathrm{p})$ is connected and star-shaped with respect to the origin and its image via $\mathrm{Exp}_{\mathrm{p}}$ is the entire manifold except the cut locus of p (Pennec 2006). Hence the normal

chart centered at p is given by

$$\mathrm{Log}_{\mathrm{p}} : D(\mathrm{p}) = M \setminus C(\mathrm{p}) \to \mathcal{D}(\mathrm{p}) \subset R^{d_{\mathcal{M}}}.$$

Here $T_{\mathrm{p}}\mathcal{M}$ is endowed with an orthonormal basis and identified with $R^{d_{\mathcal{M}}}$. The size of this chart is quantified by the *radius of injectivity of* $\mathcal{M}$ *at* p, $\rho^*(\mathcal{M}, \mathrm{p}) = \mathrm{dist}_{T_{\mathrm{p}}}(\mathbf{0}, \mathcal{C}(\mathrm{p}))$, which is the maximal radius of origin centered balls in $T_{\mathrm{p}}\mathcal{M}$ on which the exponential map is one-to-one. The radius of injectivity $\rho^*_{\mathcal{M}}$ of the manifold $\mathcal{M}$ is the infimum of the radii of injectivity at all points over the manifold. For example, in the case of Euclidean space R^d, the maximal definition domain of the normal chart is $D(\mathbf{t}) = R^d$, for all $\mathbf{t} \in R^d$, and therefore the radius of injectivity is $\rho^*_{R^d} = \infty$. In the case of the unit sphere S^k, the Riemannian metric induced by the canonical inner product on R^{d+1}, the cut locus of a point $\mathrm{p} \in S^k$ is $C(\mathrm{p}) = \{-\mathrm{p}\}$, and the tangential cut locus is $\mathcal{C}(\mathrm{p}) = S^{d-1}(\pi) \subset T_{\mathrm{p}}S^k$. Therefore, we have $\mathcal{D}(\mathrm{p}) = B(\mathbf{0}, \pi) \subset T_{\mathrm{p}}S^k$, $D(\mathrm{p}) = S^k \setminus \{-\mathrm{p}\}$, and $\rho^*(S^k, \mathrm{p}) = \pi$ for all points p on S^k. Thus, the radius of injectivity of S^k is $\rho^*_{S^k} = \pi$.

Taylor's Series Expansion of Real Functions on Riemannian Manifolds

Let $f : M \to R$ be a smooth real-valued function. The gradient $\mathrm{grad}_{\mathrm{p}} f$ of f at point p is the linear form on $T_{\mathrm{p}}\mathcal{M}$. Thus, it can be uniquely identified with a vector in $T_{\mathrm{p}}\mathcal{M}$ via the inner product $\mathrm{m}_{\mathrm{p}}(\cdot, \cdot)$ such that $\mathrm{grad}_{\mathrm{p}} f(\mathbf{v})$ corresponds to the directional derivative $\partial_{\mathbf{v}} f$. In a local chart (U, ϕ) near p with $\phi(\mathrm{p}) = \mathbf{0}$, the expression of the gradient is

$$\mathrm{grad}_{\phi^{-1}(\mathbf{t})} f = \mathrm{M}_{\phi}^{-1}(\mathbf{t}) \frac{\partial (f \circ \phi^{-1})^{\top}}{\partial \mathbf{t}} = \sum_{l=1}^{d_{\mathcal{M}}} \mathrm{m}^{jl}(\mathbf{t}) \partial_l (f \circ \phi^{-1}).$$

The *Hessian* of f at p in a local chart (U, ϕ) near p is given by

$$\mathrm{Hess}_{\phi^{-1}(\mathbf{t})} f = \sum_{j,j'=1}^{d_{\mathcal{M}}} \{\partial_{jj'}(f \circ \phi^{-1}) - \sum_{l=1}^{d_{\mathcal{M}}} \Gamma^{l}_{jj'} \partial_l (f \circ \phi^{-1})\} dt^j dt^{j'}.$$

Let ϕ_{p} be a normal chart at p, i.e. $\phi_{\mathrm{p}}(\mathrm{q}) = \mathrm{Log}_{\mathrm{p}}(\mathrm{q})$, and $f_{\mathrm{p}} = f \circ \mathrm{Exp}_{\mathrm{p}}$. Thus, $f_{\mathrm{p}}(0) = f(\mathrm{p})$. The Taylor's series expansion of $f_{\mathrm{p}}(\mathbf{v})$ around $\mathbf{0}$ is given by

$$f_{\mathrm{p}}(\mathbf{v}) = f_{\mathrm{p}}(0) + J_{f_{\mathrm{p}},0}\mathbf{v} + \frac{1}{2}\mathbf{v}^{\top} H_{f_{\mathrm{p}},0}\mathbf{v} + O(\|\mathbf{v}\|^3),$$

where $J_{f_{\mathrm{p}},0} = [\partial_j f_{\mathrm{p}}(0)]$ and $H_{f_{\mathrm{p}},0} = [\partial_{jj'} f_{\mathrm{p}}(0)]$. In a normal chart, $J_{f_{\mathrm{p}},0}$ reduces to $\mathrm{grad}_{\mathrm{p}} f^{\top}$, and the Christoffel symbols vanishes at the origin such that $H_{f_{\mathrm{p}},0}$ corresponds to the Hessian $\mathrm{Hess}_{\mathrm{p}} f$ of f at p. Thus, for all $\mathbf{v} \in \mathcal{D}(\mathrm{p})$, we have

$$f(\mathrm{Exp}_{\mathrm{p}}(\mathbf{v})) = f(\mathrm{p}) + \mathrm{grad}_{\mathrm{p}} f(\mathbf{v}) + \frac{1}{2}\mathrm{Hess}_{\mathrm{p}} f(\mathbf{v}, \mathbf{v}) + O(\|\mathbf{v}\|^3). \tag{A.1}$$

Lie Groups

A Lie group G is a group together with a smooth manifold structure such that the group operations are compatible with the smooth structure, that is, the operations of multiplication $(a, b) \mapsto ab$ and inversion $a \mapsto a^{-1}$ are smooth maps. Let G be a C^{∞} Lie group of dimension d_G and with the identity element e. Let $T_a G$ be the tangent space of G at $a \in G$, which is a d_G dimensional linear space, and let TG be the tangent bundle on G, which itself is a $2d_G$ dimensional manifold. For $a \in G$, let L_a and R_a be, respectively, the *left* and *right multiplications* by a, which are defined by

$$L_a : G \to G, \quad L_a(b) = ab, \quad b \in G,$$
$$R_a : G \to G, \quad R_a(b) = ba, \quad b \in G.$$

These maps are C^∞-diffeomorphisms and the inverses are $L_a^{-1} = L_{a^{-1}}$ and $R_a^{-1} = R_{a^{-1}}$, respectively. They include maps of the tangent bundle to itself given by $L_{a*} : T_hG \to T_{ah}G$ and $R_{a*} : T_hG \to T_{ha}G$ for $a, h \in G$. They are C^∞-diffeomorphisms and their inverses are, respectively, $L_{a*}^{-1} = L_{a^{-1}*}$ and $R_{a*}^{-1} = R_{a^{-1}*}$. Moreover, for any $b \in G$, we have $T_{ab}G = L_{a*}(T_bG)$ and $T_{ba}G = R_{a*}(T_bG)$. The fiber map $L_{a*,b}$ (or $R_{a*,b}$) is the restriction and corestriction of L_{a*} (or R_{a*}) to T_bG and is a linear isomorphism from T_bG onto $T_{ab}G$ (or $T_{ba}G$) with their inverse $L_{a*,b}^{-1} = L_{a^{-1}*,ab}$ (or $R_{a*,b}^{-1} = R_{a^{-1}*,ba}$).

A Lie group is equipped with a canonical vector-valued one form, the so called *Maurer-Cartan form* $\omega(X_a) = L_{a^{-1}*}(X_a)$ for $X_a \in T_aG$. Thus, the tangent bundle to G is trivial $TG \cong G \times T_eG$. A *left-invariant* vector is completely defined by its value at the group unity e. In particular, there is an isomorphism between the tangent space at the origin and left-invariant vector fields. Since the Lie bracket of such fields is again a left-invariant vector field, the Lie algebra structure on vector fields is inherited by the tangent space at the origin T_eG. This algebra is called the Lie algebra of the group G and it is denoted by $\mathfrak{g}$. We also have $T_aG = L_{a*}(\mathfrak{g})$, for any $a \in G$.

The *exponential map of G at the unity e* is the map $\mathrm{Exp}_e^G : \mathfrak{g} \to G$ defined as follows. For $\mathbf{v} \in \mathfrak{g}$, the exponential of $\mathbf{v}$ is defined by $\mathrm{Exp}_e^G(\mathbf{v}) = \gamma^G(1; \mathbf{v})$, where $\gamma^G(\cdot; \mathbf{v}) : R \to G$ is the unique one-parameter subgroup of G with $\gamma^G(0; \mathbf{v}) = e$ and $\frac{d}{dt}\gamma^G(0; \mathbf{v}) = \mathbf{v}$. It follows easily from the chain rule that $\mathrm{Exp}_e^G(t\mathbf{v}) = \gamma^G(t; \mathbf{v})$. The map $\gamma^G(\cdot; \mathbf{v})$ may be constructed as the integral curve of either the left- or right-invariant vector field associated with $\mathbf{v}$. The integral curve exists for all real parameters followed by left- or right-translation of the solution near zero. Therefore, Exp_e^G is globally defined on $\mathfrak{g}$ with $\mathrm{Exp}_e^G(0) = e$, and $\mathrm{Exp}_e^G(-\mathbf{v}) = (\mathrm{Exp}_e^G(\mathbf{v}))^{-1}$ for $\mathbf{v} \in \mathfrak{g}$. Moreover, the exponential map Exp_e^G is a local C^∞-diffeomorphism around $\mathbf{0} \in \mathfrak{g} = T_eG$.

For $a \in G$, the *exponential map of G at a*, Exp_a^G, is the unique map from T_aG into

G that satisfies the following condition

$$\mathrm{Exp}_a^G \circ L_{a*} = L_a \circ \mathrm{Exp}_e^G \tag{A.2}$$

on $\mathfrak{g}$. Therefore, Exp_a^G is globally defined on T_aG, $\mathrm{Exp}_a^G(0) = a$, and Exp_a^G is a local C^∞-diffeomorphism around $0 \in T_aG$. Assume that $X_1, \ldots, X_{d_G}$ is a given basis for $\mathfrak{g} = T_eG$. Any $\mathbf{v} \in T_eG$ can be uniquely written as

$$\mathbf{v} = \sum_{\ell=1}^{d_G} v^\ell X_\ell.$$

Let $X_{1,a}, \ldots, X_{d_G,a}$ be the (unique) left invariant tangent vector fields with values $X_1, \ldots, X_{d_G}$ at e, i.e $X_{\ell,a} = L_{a*}(X_\ell)$ for $\ell = 1, \ldots, d_G$, and $a \in G$. Then, $X_{1,a}, \ldots, X_{d_G,a}$ at a form a basis for T_aG, for all $a \in G$, so they define a *trivialization of the tangent bundle* of G as follows:

$$f : TG \to G \times R^{d_G}, \qquad f(\sum_{\ell=1}^{d_G} c^\ell \mathbf{X}_{\ell,a}) = (a, (c^1, \ldots c^{d_G})). \tag{A.3}$$

Let $< \cdot, \cdot >_e$ be an inner product on T_eG and $< \cdot, \cdot >$ be the Riemannian metric defined as in (A.4). A *Riemannian* or *pseudo-Riemannian metric* on a Lie group G is *left invariant* if it is preserved under every left multiplication L_a, that is,

$$< \mathbf{v}, \mathbf{w} >_b = < L_{a,*}(\mathbf{v}), L_{a,*}(\mathbf{w}) >_{ab}, \quad \text{for } \mathbf{v}, \mathbf{w} \in T_bG, \text{ and } b, a \in G.$$

A left-invariant metric is uniquely defined by its restriction to the tangent space to the group at unity, hence by an inner product on $\mathfrak{g}$. Therefore, any inner product $< \cdot, \cdot >_e$ on T_eG can be extended to a (unique) left invariant Riemannian metric

$< \cdot, \cdot >= \{< \cdot, \cdot >_a\}_{a \in G}$ on G, namely

$$< X, Y >_a := < L_{a^{-1}*}(X), L_{a^{-1}*}(Y) >_e, \quad X, Y \in T_a G,\ a \in G. \tag{A.4}$$

The associated norm is denoted by $\| \cdot \|_a$ and $\|X\|_a = < X, X >_a^{1/2} = \|L_{a^{-1}*}(X)\|_e$.

It is easy to see that the exponential maps Exp_e^G and Exp_a^G defined as above using the algebraic structure of G coincide with the manifold exponential maps defined when G is viewed as a Riemannian manifold (with a left-invariant metric). Moreover, the maximal domain on which the exponential map Exp_a^G is one-to-one is $\mathcal{D}(a) = L_{a,*}(\mathcal{D}(e))$ and so $\rho^*(G, e) = \rho^*(G, a)$, for all $a \in G$. Therefore, the radius of injectivity of G is $\rho_G^* = \rho^*(G, e)$. Let Log_e^G and Log_a^G be the inverse maps of Exp_e^G and Exp_a^G. We have that $\mathrm{Log}_e^G(b^{-1}) = -\mathrm{Log}_e^G(b)$ provided $b \in D(e)$. For $b \in D(a) = \mathrm{Exp}_a^G(\mathcal{D}(a))$, the geodesic distance from b to a can be expressed as

$$\mathrm{dist}_G(b, a) = \|\mathrm{Log}_a^G(b)\|_a = \|\mathrm{Log}_e^G(a^{-1}b)\|_e. \tag{A.5}$$

Riemannian Symmetric Spaces

A map $f : M \to \mathcal{M}$ defined on a neighborhood of p $\in \mathcal{M}$ is said to be a *geodesic symmetry* if it fixes the point p and reverses geodesics through that point, i.e. if $\gamma(\cdot)$ is a geodesic with $\gamma(0) =$ p and $f(\gamma(t)) = \gamma(-t)$, for any t. A Riemannian symmetric space RSS is a connected Riemannian manifold $\mathcal{M}$ with the property that at each point, geodesic symmetries are isometric or distance preserving (Boothby 1986, Helgason 1978). They arise in a wide variety of situations in both mathematics and physics. Basic examples of RSS's are Euclidean spaces, R^d, spheres, S^k, projective spaces, PR^d, and hyperbolic spaces, H^d, each with their standard Riemannian metric. Symmetric spaces arise naturally from Lie group actions on manifolds. Many common geometric transformations of Euclidean spaces - rotations, translations, dilations, and

affine transformations on R^d - form Lie groups. In general, Lie groups can be used to describe transformations of smooth manifolds.

Given a smooth manifold $\mathcal{M}$ and a Lie group G, a *smooth group action of G on $\mathcal{M}$* is a smooth mapping $G \times \mathcal{M} \to \mathcal{M}$, $(a, y) \mapsto a \cdot y$, such that $e \cdot y = y$ and $(ab) \cdot y = a \cdot (b \cdot y)$ for all $a, b \in G$ and all $y \in \mathcal{M}$, where e is the identity element of G. The group action should be interpreted as a group of transformations of the manifold $\mathcal{M}$, namely, $\{L_a\}_{a \in G}$, where L_a is the action of the group element a on $\mathcal{M}$, $L_a : \mathcal{M} \to \mathcal{M}$, $L_a(y) = a \cdot y$ for $y \in \mathcal{M}$ and $a \in G$. L_a is a smooth diffeomorphism on $\mathcal{M}$ and its inverse is $L_{a^{-1}}$. Given $y \in \mathcal{M}$ a point on $\mathcal{M}$, let ι_y denote the action of G on the point y, i.e. $\iota_y : G \to \mathcal{M}$, $\iota_y(a) = a \cdot y = L_a(y)$ for all $a \in G$. The ι_y is a smooth map from G into $\mathcal{M}$. For example, for any Lie group G, the group multiplication defines a group action of G on itself, and the action of an element a on the group itself is exactly the left-multiplication by a. Another example is $\mathrm{SO}(d)$, which is a Lie group and it acts on R^d as rotations, i. e. $R \cdot \mathbf{y} = R\mathbf{y}$ for all $R \in \mathrm{SO}(d)$ and $\mathbf{y} \in R^d$.

We now introduce some common concepts related to group actions. The *orbit* of a point $y \in \mathcal{M}$ is defined as $G(y) = \{a \cdot y \,|\, a \in G\}$. The orbits form a partition of $\mathcal{M}$, and we say that two points $y, y' \in \mathcal{M}$ are equivalent if they belong to the same orbit. In the case that $\mathcal{M}$ consists of a single orbit, we say that 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 y of $\mathcal{M}$ is defined as $G_y = \{a \in G \,|\, a \cdot y = y\}$. For example, for the action of the group, $\mathrm{SO}(2)$, the isotropy subgroup of the zero vector is $G_{\mathbf{0}} = \mathrm{SO}(2)$ and for any non-zero vector $\mathbf{y} \in R^2$, the isotropy group $G_{\mathbf{y}}$ reduces to the trivial subgroup $\{I_2\}$.

Let H be a closed Lie subgroup of the Lie group G. Then the *left coset* of an element $a \in G$ is defined by $aH = \{ah \,|\, h \in H\}$. The space of all such cosets is called a quotient space of the group G with respect to the subgroup H, denoted by G/H, and it is a smooth manifold with the quotient topology. When a Lie group G acts smoothly on a

smooth manifold $\mathcal{M}$ for any $y \in \mathcal{M}$, there is a natural bijection from the orbit $G(y)$ onto the quotient manifold given by the mapping $a \cdot y \mapsto aG_y$, which is well-defined and smooth, so $G(y) \cong G/G_y$.

Now let $\mathcal{M}$ be a symmetric space and choose an arbitrary base point $\mathrm{p} \in \mathcal{M}$. We can always view $\mathcal{M}$ as a homogeneous space $\mathcal{M} \cong G/G_{\mathrm{p}}$, where G is a connected group of isometries of $\mathcal{M}$ and the isotropy subgroup G_{p} is compact. We call G a *group of isometries* of $\mathcal{M}$ if for all $a \in G$, $\mathrm{dist}_{\mathcal{M}}(y, z) = \mathrm{dist}_{\mathcal{M}}(a \cdot y, a \cdot z)$ for all $y, z \in \mathcal{M}$. Any Lie group G can be viewed as a symmetric space with a Riemannian structure induced by an inner product on T_eG, and G acting on itself by left multiplication. Obviously, this action is transitive and the isotropy subgroups are trivial, i.e. $G_a = \{e\}$, for all $a \in G$.

A very common example of a symmetric space is S^2, which is a 2-dimensional compact Riemannian manifold. The Lie group, SO(3), of all rotations in R^3 acts smoothly and transitively on S^2. For example, let us choose the north pole $\mathrm{p} = (0, 0, 1) \in S^2$ as the base point. It is easy to see that the orbit of p is the entire sphere and thus S^2 is a homogeneous space. The isotropy subgroup of p is the group of all rotations about the z-axis in R^3, which can be identified with the group of 2D rotations, SO(2). Hence, S^2 can be naturally identified with the quotient space SO(3)/SO(2). Similarly, the k-dimensional unit sphere, S^k, can be identified as the quotient space $\mathrm{SO}(k+1)/\mathrm{SO}(k)$. The sphere S^k is a compact Riemannian manifold.

Other examples of symmetric spaces can be obtained by taking Cartesian products of symmetric spaces. Consider two manifolds $\mathcal{M}_1$ and $\mathcal{M}_2$ and two Lie groups G_1 and G_2 so that G_j acts transitively on $\mathcal{M}_j$ for $j = 1, 2$. Thus, the group $G = G_1 \times G_2$ is a Lie group and acts transitively on the manifold $\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2$. Given a base point $\mathrm{p} = (\mathrm{p}_1, \mathrm{p}_2)$ in $\mathcal{M}$, the isotropy subgroup of p in G is $G_{\mathrm{p}} = G_{1,\mathrm{p}_1} \times G_{2,\mathrm{p}_2}$. Thus, we can write $\mathcal{M}_1 \times \mathcal{M}_2$ as a homogeneous space $G/G_{\mathrm{p}} = (G_1/G_{1,\mathrm{p}_1}) \times (G_2/G_{2,\mathrm{p}_2})$.

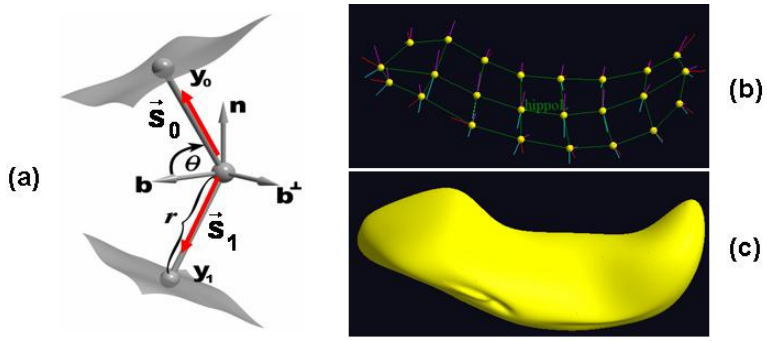

Figure A.1: (a) A medial representation model $\mathbf{m} = (\mathbf{O}, r, \mathbf{s}_0, \mathbf{s}_1)$ at an atom, where $\mathbf{O}$ is the center of the inscribed sphere, r is the common spoke length, abd $\{\mathbf{s}_0, \mathbf{s}_1\}$ are the two unit spoke directions; (b) a skeleton of a hippocampus with 24 medial atoms,; (c) the smoothed surface of the hippocampus.

An example of a symmetric space used in the study of 3D geometric objects is the space of medial atoms, $\mathcal{M} = R^3 \times R^+ \times S^2 \times S^2$ (Shi et al. 2012). See Figure A.1 for an illustration. The group $G = R^3 \times R^+ \times \mathrm{SO}(3) \times \mathrm{SO}(3)$ acts smoothly on $\mathcal{M}$. For an element $a = (O', r', R_0, R_1) \in G$ and an medial atom $\mathrm{q} = (O, r, \mathbf{s}_0, \mathbf{s}_1) \in \mathcal{M}$, the group action is defined by

$$a \cdot \mathrm{q} = (O + O', rr', R_0\mathbf{s}_0, R_1\mathbf{s}_1),$$

which is a transitive action. Consider the atom p located at $O = (0, 0, 0)$ with radius $r = 1$ and spokes $\mathbf{s}_0 = \mathbf{s}_1 = (0, 0, 1)$. Then, the isotropy subgroup of p is $G_\mathrm{p} = \{0\} \times \{1\} \times \mathrm{SO}(2) \times \mathrm{SO}(2)$, and we can write the medial atom space as the quotient space $\mathcal{M} = R^3 \times R^+ \times (\mathrm{SO}(3)/\mathrm{SO}(2)) \times (\mathrm{SO}(3)/\mathrm{SO}(2))$.

From now on, it is assumed that the manifold $\mathcal{M}$ is a symmetric space, $\mathcal{M} = G/G_\mathrm{p}$ 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 $\mathcal{M}$. Due 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. For an arbitrary point $y \in \mathcal{M}$, geodesics starting from y are of the form $a \cdot \gamma(\cdot)$,

where $\gamma(\cdot)$ is a geodesic starting from p with $\gamma(0) = \mathrm{p}$ and $y = a \cdot \mathrm{p}$ for some $a \in G$. Due to the local uniqueness of geodesics, if $y = a' \cdot \mathrm{p}$ for some other $a' \in G$, then $a \cdot \gamma(\cdot) = a' \cdot \gamma(\cdot)$.

Geodesics on $\mathcal{M}$ starting from p are the images of the action of a 1-parameter subgroup of G acting on the base point p. In other words, for any geodesic γ on $\mathcal{M}$, $\gamma(\cdot) : I \to \mathcal{M}$, starting from p, there exists a 1-parameter subgroup $c(\cdot) : R \to G$ such that $\gamma(t) = c(t) \cdot \mathrm{p}$ for all $t \in I$. The manifold exponential map $\mathrm{Exp}_{\mathrm{p}}^{\mathcal{M}}$ at the base point p is defined by

$$\mathrm{Exp}_{\mathrm{p}}^{\mathcal{M}}(t\mathbf{v}) = \gamma(t; \mathrm{p}, \mathbf{v}) = c(t; e, \mathbf{u}) \cdot \mathrm{p},$$

where $\gamma(0; \mathrm{p}, \mathbf{v}) = \mathrm{p}$, $\frac{d}{dt}\gamma(0; \mathrm{p}, \mathbf{v}) = \mathbf{v} \in T_{\mathrm{p}}\mathcal{M}$, $c(0; e, \mathbf{u}) = e$, and $\frac{d}{dt}c(0; \mathrm{p}, \mathbf{u}) = \mathbf{u} \in T_e G$ with $\mathbf{u}$ so that $\iota_{\mathrm{p}*,e}(\mathbf{u}) = \mathbf{v}$ for small $t \in R$. That is,

$$\mathrm{Exp}_{\mathrm{p}}^{\mathcal{M}}(t\, \iota_{\mathrm{p}*,e}(\mathbf{u})) = \mathrm{Exp}_{\mathrm{p}}^{G}(t\mathbf{u}) \cdot \mathrm{p},$$

for all $\mathbf{u} \in T_e G$ and $t \in R$ with small $\|t\mathbf{u}\|$.

Moreover, the manifold exponential map $\mathrm{Exp}_{\mathrm{q}}^{\mathcal{M}}$ of $\mathcal{M}$ at a point q is defined by

$$\mathrm{Exp}_{\mathrm{q}}^{\mathcal{M}}(L_{a*,\mathrm{p}}\mathbf{v}) = a \cdot \mathrm{Exp}_{\mathrm{p}}^{\mathcal{M}}(\mathbf{v})$$

for any $a \in G$ with $\mathrm{q} = a \cdot \mathrm{p}$ and any small $\mathbf{v} \in T_{\mathrm{p}}\mathcal{M}$, where L_a is the action of the element a on the points of $\mathcal{M}$. Due to the uniqueness of geodesics, if $\mathrm{q} = a_1 \cdot \mathrm{p} = a_2 \cdot \mathrm{p}$ with $a_1, a_2 \in G$ and $\mathbf{w} = L_{a_1*,\mathrm{p}}(\mathbf{v}_1) = L_{a_2*,\mathrm{p}}(\mathbf{v}_2)$ with $\mathbf{v}_1, \mathbf{v}_2 \in T_{\mathrm{p}}\mathcal{M}$, then $a_1 \cdot \mathrm{Exp}_{\mathrm{p}}^{\mathcal{M}}(\mathbf{v}_1) = a_2 \cdot \mathrm{Exp}_{\mathrm{p}}^{\mathcal{M}}(\mathbf{v}_2)$. Since G is a group of isometries on $\mathcal{M}$, the radius of injectivity $\mathrm{Exp}_{\mathrm{q}}^{\mathcal{M}}$ of $\mathcal{M}$ at q is independent of the point q, so $\rho_{\mathcal{M}}^* = \rho^*(\mathcal{M}, \mathrm{p})$.

The unit sphere S^k is a compact Riemannian manifold of dimension d and injectivity

radius $\rho = \pi$. The tangent space at $\mathrm{q} \in S^k$ is

$$T_{\mathrm{q}} S^k = \{\mathbf{v} \in R^{k+1} : \mathbf{v}^\top \mathrm{q} = 0\}.$$

The tangent space is endowed with the metric tensor from R^{k+1}, $m_{\mathrm{q}}(\mathbf{v}_1, \mathbf{v}_2) = \mathbf{v}_1^\top \mathbf{v}_2$ for all $\mathbf{v}_1, \mathbf{v}_2 \in T_{\mathrm{q}} S^k$. The geodesic distance between two points $\mathrm{q}_1, \mathrm{q}_2 \in S^k$ is given by $\mathrm{dist}_{\mathcal{M}}(\mathrm{q}_1, \mathrm{q}_2) = \arccos(\mathrm{q}_1^\top \mathrm{q}_2)$, which lies between 0 and π. The exponential map takes the form

$$\mathrm{Exp}_{\mathrm{q}} : T_{\mathrm{q}} S^k \to S^k, \ \mathrm{Exp}_{\mathrm{q}}(\mathbf{v}) = \cos(\|\mathbf{v}\|)\mathrm{q} + \frac{\sin(\|\mathbf{v}\|)}{\|\mathbf{v}\|}\mathbf{v}.$$

It is a diffeomorphism from $B(0, \pi) \subset T_{\mathrm{q}} S^k$ onto $S^k \setminus \{-\mathrm{q}\}$, and the logarithmic map is given by

$$\mathrm{Log}_{\mathrm{q}} : S^k \setminus \{-\mathrm{q}\} \to B(0, \pi), \ \mathrm{Exp}_{\mathrm{q}}(\mathrm{q}_1) = \frac{\arccos(\mathrm{q}_1^\top \mathrm{q})}{\sqrt{1 - (\mathrm{q}_1^\top \mathrm{q})^2}} \left(\mathrm{q}_1 - (\mathrm{q}_1^\top \mathrm{q})\mathrm{q}\right),$$

for all $\mathrm{q}_1 \in S^k$ with $\mathrm{q}_1 \neq -\mathrm{q}$.

A.1.2 Unit circle S^1 in the complex plane

Let $S^1 = \{z = \cos(\phi) + j \sin(\phi) : \phi \in R\}$ be the unit circle in the complex plane $\mathbf{C}$, where $j = \sqrt{-1}$. The S^1 with the usual multiplication of complex numbers forms a compact 1-dimensional C^∞ Lie group with 1 as the unity. The tangent space of S^1 at $a = \cos(\theta_0) + j \sin(\theta_0) \ \in S^1$ is given by $T_a(S^1) = \{t\,(-\sin(\theta_0) + j \cos(\theta_0)) : t \in R\}$, which is a 1-dimensional real linear subspace of $\mathbf{C}$ formed by all $z = z_x + j z_y$'s that are orthogonal to a as vectors in R^2. The Lie algebra of S^1 is $T_1 S^1 = \{jt : t \in R\}$ and the exponential map at unity is given by $\mathrm{Exp}_1(jt) = e^{jt} = \cos t + j \sin t$. Thus, we have

$$\mathrm{Exp}_a\left(t\,(-\sin(\theta_0) + j \cos(\theta_0))\right) = \cos(t + \theta_0) + j \sin(t + \theta_0).$$

Geometrically, Exp_a "wraps" the tangent line at a around the circle, and thus the injectivity radius ρ equals π.

Suppose that we observe $\{(y_i, \mathbf{x}_i) : i = 1, \ldots, n\}$, where $y_i = \cos(\phi_i) + j\sin(\phi_i) \in S^1$ for all i. We define

$$I(\cdot, \cdot) : R^{d_{\mathbf{x}}} \times R^{d_\beta} \to R \text{ with } I(\mathbf{0}, \cdot) = I(\cdot, \mathbf{0}) = 0.$$

For an $a \in S^1$, we consider a single-center link function and its corresponding *rotated residual*, which are, respectively, given by

$$\begin{aligned} f(\mathbf{x}_i, a, \boldsymbol{\beta}) &= a\, e^{jI(\mathbf{x}_i, \boldsymbol{\beta})} = e^{j(\theta_0 + I(\mathbf{x}_i, \boldsymbol{\beta}))}, \\ \mathcal{E}_i(a, \boldsymbol{\beta}) &= j(\phi_i - \theta_0 - I(\mathbf{x}_i, \boldsymbol{\beta}))_{\text{mod } 2\pi}, \end{aligned}$$

where $t_{\text{mod } 2\pi}$ is the unique number in $(-\pi, \pi]$ so that $t - t_{\text{mod } 2\pi} \in 2\pi z$. Thus, the intrinsic regression model is written as

$$E[\mathcal{E}_i(a, \boldsymbol{\beta}) |\, \mathbf{x}_i] = 0, \qquad i = 1, \ldots, n. \tag{A.6}$$

A.1.3 Lie Logarithmic Maps of SO(2) and SO(3)

When $k = 2$, $SO(2)$ is the set of all 2×2 matrices of the form $\begin{pmatrix} x & -y \\ y & x \end{pmatrix}$ with $x^2 + y^2 = 1$ for $x, y \in R$. The group $SO(2)$ of rotations in R^2 is isomorphic with S^1. The canonical isomorphism is $\begin{pmatrix} x & -y \\ y & x \end{pmatrix} \to z = x + jy$. A 2×2 skew-symmetric matrix B can be written as $B = \lambda J$, where $\lambda \in R$ and

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

There is a canonical isomorphism from the linear space $T_{I_2}SO(2)$ into the space of pure imaginary numbers, jR, namely, $\lambda J \mapsto j\lambda$ for $\lambda \in R$. It can be shown that

$$e^B = e^{\lambda J} = \cos(\lambda) I_2 + \sin(\lambda) J.$$

and, since $J^2 = -I_2$, it follows

$$\cos(\lambda) = \frac{1}{2}\mathrm{tr}(e^B) \qquad \sin(\lambda) = -\frac{1}{2}\mathrm{tr}(e^B J).$$

Thus e^B determines λ uniquely up to an additive multiple of 2π.

Given a rotation matrix $O \in SO(2)$, the Lie logarithmic map at I_2 of O is given by

$$\mathrm{Log}_{I_2}(O) = \lambda J, \tag{A.7}$$

where $\cos(\lambda) = 0.5\mathrm{tr}(R)$ and $\sin(\lambda) = -0.5\mathrm{tr}(RJ)$ for $\lambda \in (-\pi, \pi]$. Thus, when $SO(2)$ is endowed with the trace metric, it follows immediately that the radius of injectivity of $SO(2)$ is $\rho^*_{SO(2)} = \sqrt{2}\,\pi$.

When $k = 3$, a 3×3 skew-symmetric matrix B is of the form

$$B = \begin{pmatrix} 0 & -c_1 & c_2 \\ c_1 & 0 & -c_3 \\ -c_2 & c_3 & 0 \end{pmatrix},$$

and letting $\lambda = \sqrt{c_1^2 + c_2^2 + c_3^2}$, we have the well-known Rodrigues formula

$$e^B = I_3 + \frac{\sin(\lambda)}{\lambda} B + \frac{[1 - \cos(\lambda)]}{\lambda^2} B^2.$$

It may be more convenient to normalize B such that one can write $B = \lambda B_1$ (or,

equivalently, $B_1 = B/\lambda$, assuming $\lambda \neq 0$). In this case, e^B can be written as

$$e^B = e^{\lambda B_1} = I_3 + \sin\lambda\, B_1 + (1 - \cos\lambda) B_1^2.$$

Observing that $\mathrm{tr}(e^B) = 1 + 2\cos(\lambda)$ and $0.5[e^B - (e^B)^\top] = \sin(\lambda)\, B_1$, the logarithmic map at I_3 of a rotation $O \in SO(3)$ is given by

$$\mathrm{Log}_{I_3}(O) = \lambda B_1, \tag{A.8}$$

where $\lambda = \arccos((\mathrm{tr}(O) - 1)/2)$ and $B_1 = (O - O^\top)/(2\sin\lambda)$. When $\lambda = 0$ or $\lambda = \pi$, the above formulae cannot be used. When $\lambda = 0$, we have $O = I_3$ and $B_1 = 0$, so $\mathrm{Log}_I(O) = \mathrm{Log}_I(I_3) = O_3$, and Eq. (A.8) still holds. When $\lambda = \pi$, we need to find B_1 such that $B_1^2 = \frac{1}{2}(O - I_3)$. As B_1 is a skew-symmetric matrix, this amounts to solving a simple system of equations with three unknowns. When $SO(3)$ is endowed with the trace metric, elementary calculations on the Rodrigues formula yield that Exp_I is one-to-one on the ball $B(\mathbf{0}, \sqrt{2}\,\pi)$ in $T_{I_2}SO(3)$, but not on any ball $B(\mathbf{0}, \rho)$ with $\rho > \sqrt{2}\,\pi$. Therefore, the radius of injectivity of $SO(3)$ is $\rho^*_{SO(3)} = \sqrt{2}\,\pi$.

Suppose we observe an element $\mathrm{q}_i \in SO(k)$ and a $d_x \times 1$ covariate vector $\mathbf{x}_i$ for $i = 1, \ldots, n$. We consider an intercept rotation matrix O. Then, for a given map $f(\mathbf{x}_i, \boldsymbol{\beta})$ with $f(\cdot,\cdot) : R^{d_x} \times R^{d_\beta} \to R^{k(k-1)/2}$ with $f(\mathbf{0}, \cdot) = \mathbf{0}$, let $\Lambda(\mathbf{x}_i, \boldsymbol{\beta}) = \sum_{k=2}^{m} \sum_{\ell=1}^{k-1} f(\mathbf{x}_i, \boldsymbol{\beta})_{(k-1)(k-2)/2+\ell} X_{k\ell} \in \mathrm{SkewSym}(k)$, and consider the "directional" matrix $\mathbf{u}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta}) = \mathrm{q}\Lambda(\mathbf{x}_i, \boldsymbol{\beta})$ as a tangent vector to $SO(k)$ at $\mathrm{q} \in SO(k)$. By considering the "conditional mean"

$$\mu(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta}) = \mathrm{Exp}_{\mathrm{q}}(\mathbf{u}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta})) = \mathrm{q}\exp(\Lambda(\mathbf{x}_i, \boldsymbol{\beta})),$$

the intrinsic residual is given by

$$\mathcal{E}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta}) = \mathcal{E}_i(\mathrm{q}, \boldsymbol{\beta}) = \mathrm{Log}(e^{-\Lambda(\mathbf{x}_i, \boldsymbol{\beta})} \mathrm{q}^\top \mathrm{q}_i).$$

When $k = 2$, both the manifold $SO(2)$ and the linear space SkewSym(2) have dimension 1, so $f(\mathbf{x}_i, \boldsymbol{\beta})$ is a scalar map. We have

$$\begin{aligned} \Lambda(\mathbf{x}_i, \boldsymbol{\beta}) &= f(\mathbf{x}_i, \boldsymbol{\beta}) J = \begin{pmatrix} 0 & -f(\mathbf{x}_i, \boldsymbol{\beta}) \\ f(\mathbf{x}_i, \boldsymbol{\beta}) & 0 \end{pmatrix} \\ \mathrm{q} &= \begin{pmatrix} \cos(\theta_0) & -\sin(\theta_0) \\ \sin(\theta_0) & \cos(\theta_0) \end{pmatrix}, \ \mathrm{q}_i = \begin{pmatrix} \cos(\theta_i) & -\sin(\theta_i) \\ \sin(\theta_i) & \cos(\theta_i) \end{pmatrix} \end{aligned}$$

where θ_i are in $(-\pi, \pi]$ for $i = 0, \ldots, n$. The "conditional mean" becomes

$$\mu(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta}) = \begin{pmatrix} \cos(\theta_0 + f(\mathbf{x}_i, \boldsymbol{\beta})) & -\sin(\theta_0 + f(\mathbf{x}_i, \boldsymbol{\beta})) \\ \sin(\theta_0 + f(\mathbf{x}_i, \boldsymbol{\beta})) & \cos(\theta_0 + f(\mathbf{x}_i, \boldsymbol{\beta})) \end{pmatrix}$$

and the "intrinsic residual" is

$$\mathcal{E}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta}) = (\phi_i - \theta_0 - \mathrm{I}(\mathbf{x}_i, \boldsymbol{\beta}))_{\mathrm{mod}\ 2\pi} J.$$

We observe that we recapture, via the canonical isomorphism between $SO(2)$ and S^1, the intrinsic model presented in Example 2. When $SO(2)$ is endowed with the trace metric, the Riemannian distance on $SO(2)$ between two rotations is a constant multiple of the Riemannian distance on S^1 between their counterparts in S^1; the multiplicative factor is $\sqrt{2}$.

When $k = 3$, the manifold $SO(3)$ and the linear space SkewSym(3) have dimension

3, and the above "conditional mean" becomes

$$\mu(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta}) = \mathrm{q}\left(I_3 + \frac{\sin\lambda(\mathbf{x}_i,\boldsymbol{\beta})}{\lambda(\mathbf{x}_i,\boldsymbol{\beta})}\Lambda(\mathbf{x}_i,\boldsymbol{\beta}) + \frac{(1-\cos\lambda(\mathbf{x}_i,\boldsymbol{\beta}))}{\lambda(\mathbf{x}_i,\boldsymbol{\beta})^2}\Lambda(\mathbf{x}_i,\boldsymbol{\beta})^2\right),$$

where $\lambda(\mathbf{x}_i,\boldsymbol{\beta}) = \|f(\mathbf{x}_i,\boldsymbol{\beta})\|$, and the "intrinsic residual" is

$$\begin{aligned}\mathcal{E}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta}) \;=\; & \frac{\theta_i(\mathrm{q},\boldsymbol{\beta})}{2\sin(\theta_i(\mathrm{q},\boldsymbol{\beta}))} \times \\ & \big[(\mathrm{q}^\top \mathrm{q}_i - \mathrm{q}_i^\top \mathrm{q}) \\ & - \frac{\sin(\lambda(\mathbf{x}_i,\boldsymbol{\beta}))}{\lambda(\mathbf{x}_i,\boldsymbol{\beta})}\left(\Lambda(\mathbf{x}_i,\boldsymbol{\beta})\mathrm{q}^\top\mathrm{q}_i + \mathrm{q}_i^\top\mathrm{q}\Lambda(\mathbf{x}_i,\boldsymbol{\beta})\right) \\ & + \frac{[1-\cos(\lambda(\mathbf{x}_i,\boldsymbol{\beta}))]}{\lambda(\mathbf{x}_i,\boldsymbol{\beta})^2}\left(\Lambda(\mathbf{x}_i,\boldsymbol{\beta})^2\mathrm{q}^\top\mathrm{q}_i - \mathrm{q}_i^\top\mathrm{q}\Lambda(\mathbf{x}_i,\boldsymbol{\beta})^2\right)\Big],\end{aligned}$$

where $\theta_i(\mathrm{q},\boldsymbol{\beta})$ is given by

$$\begin{aligned}&\arccos\left\{\frac{1}{2}\left[\mathrm{tr}\left(\mathrm{q}^\top\mathrm{q}_i\right) - \frac{\sin(\lambda(\mathbf{x}_i,\boldsymbol{\beta}))}{\lambda(\mathbf{x}_i,\boldsymbol{\beta})}\mathrm{tr}\left(\Lambda(\mathbf{x}_i,\boldsymbol{\beta})\mathrm{q}^\top\mathrm{q}_i\right)\right.\right. \\ &\left.\left. + \frac{(1-\cos(\lambda(\mathbf{x}_i,\boldsymbol{\beta})))}{\lambda(\mathbf{x}_i,\boldsymbol{\beta})^2}\mathrm{tr}\left(\Lambda(\mathbf{x}_i,\boldsymbol{\beta})^2\mathrm{q}^\top\mathrm{q}_i\right)\right] - 1\right\}.\end{aligned}$$

A.2 Proofs for Chapter 2

Proof of Theorem 2.3.1. (a) Let

$$\mathcal{Q}(\mathrm{q},\boldsymbol{\beta}) := (E[h(\mathbf{x},\mathrm{q},\boldsymbol{\beta})\mathcal{E}(y,\mathbf{x},\mathrm{q},\boldsymbol{\beta})])^\top W(E[h(\mathbf{x},\mathrm{q},\boldsymbol{\beta})\mathcal{E}(y,\mathbf{x},\mathrm{q},\boldsymbol{\beta})]) \tag{A.9}$$

and $\mathcal{Q}_n(\mathrm{q},\boldsymbol{\beta})$ as in (2.8). Conditions $(C1)$ and $(C4)$ imply that $\sup_{(\mathrm{q},\boldsymbol{\beta})}|\mathcal{Q}_n(\mathrm{q},\boldsymbol{\beta}) - \mathcal{Q}(\mathrm{q},\boldsymbol{\beta})| \xrightarrow{p} 0$, while conditions $(C2)$, $(C3)$, and $(C5)$ yield

$$\mathcal{Q}(\mathrm{q}_*,\boldsymbol{\beta}_*) = 0 < \inf_{(\mathrm{q},\boldsymbol{\beta}):dist_M(\mathrm{q},\mathrm{q}_*)+\|\boldsymbol{\beta}-\boldsymbol{\beta}_*\|\geq\epsilon}\mathcal{Q}(\mathrm{q},\boldsymbol{\beta}),$$

for any $\epsilon > 0$. Thus, the consistency of $(\hat{\mathrm{q}}, \hat{\boldsymbol{\beta}})$ in (2.8) follows from Theorem 5.7 in van der Vaart (1998) (van der Vaart 1998) applied to $-\mathcal{Q}_n$.

(b) In the case M is an open subset of the Euclidean space $\mathbf{R}^d$ and $\phi = \mathrm{id}$, the CLT in (2.9) is a classical result, e.g. see Newey (1993) (Newey 1993). Under the assumption $(C6)$, the GMM estimator $(\hat{\mathrm{q}}, \hat{\boldsymbol{\beta}})$ is a zero of the real vector-valued function $\mathcal{H}_n(\mathrm{q}, \boldsymbol{\beta}) := \frac{1}{2}\frac{\partial}{\partial(\mathrm{q},\boldsymbol{\beta})}\mathcal{Q}_n(\mathrm{q}, \boldsymbol{\beta})$, i.e.

$$\begin{aligned}\mathcal{H}_n(\mathrm{q}, \boldsymbol{\beta}) &= \left[\frac{\partial}{\partial(\mathrm{q}, \boldsymbol{\beta})}\left(\mathbb{P}_n(h(\cdot, \mathrm{q}, \boldsymbol{\beta})\mathcal{E}(\cdot, \cdot, \mathrm{q}, \boldsymbol{\beta}))\right)\right]^\top W_n \mathbb{P}_n(h(\cdot, \mathrm{q}, \boldsymbol{\beta})\mathcal{E}(\cdot, \cdot, \mathrm{q}, \boldsymbol{\beta})) \\ &= \left[\mathbb{P}_n\left(\frac{\partial}{\partial(\mathrm{q}, \boldsymbol{\beta})}\left[h(\cdot, \mathrm{q}, \boldsymbol{\beta})\mathcal{E}(\cdot, \cdot, \mathrm{q}, \boldsymbol{\beta})\right]\right)\right]^\top W_n \mathbb{P}_n(h(\cdot, \mathrm{q}, \boldsymbol{\beta})\mathcal{E}(\cdot, \cdot, \mathrm{q}, \boldsymbol{\beta})).\end{aligned} \tag{A.10}$$

The CLT (2.9) follows from the first order Taylor expasion of $\mathcal{H}_n(\mathrm{q}, \boldsymbol{\beta})$ around the true value $(\mathrm{q}_*, \boldsymbol{\beta}_*)$ evaluated at the estimator $(\hat{\mathrm{q}}, \hat{\boldsymbol{\beta}})$. Indeed, under $(C6)$ the function $\mathcal{H}_n(\cdot, \cdot)$ is continously differentiable, and its first order Taylor expansion gives

$$\begin{aligned}0 = \mathcal{H}_n(\hat{\mathrm{q}}, \hat{\boldsymbol{\beta}}) &= \mathcal{H}_n(\mathrm{q}_*, \boldsymbol{\beta}_*) + \frac{\partial}{\partial(\mathrm{q}, \boldsymbol{\beta})}\mathcal{H}_n(\mathrm{q}_*, \boldsymbol{\beta}_*)[(\hat{\mathrm{q}}^\top, \hat{\boldsymbol{\beta}}^\top)^\top - (\mathrm{q}_*^\top, \boldsymbol{\beta}_*^\top)^\top] \\ &+ O_p(\|(\hat{\mathrm{q}}^\top, \hat{\boldsymbol{\beta}}^\top)^\top - (\mathrm{q}_*^\top, \boldsymbol{\beta}_*^\top)^\top\|^2) \\ &= \mathcal{H}_n(\mathrm{q}_*, \boldsymbol{\beta}_*) + \left[\frac{\partial}{\partial(\mathrm{q}, \boldsymbol{\beta})}\mathcal{H}_n(\mathrm{q}_*, \boldsymbol{\beta}_*) + o_p(1)\right][(\hat{\mathrm{q}}^\top, \hat{\boldsymbol{\beta}}^\top)^\top - (\mathrm{q}_*^\top, \boldsymbol{\beta}_*^\top)^\top]\end{aligned}$$

Solve and we get

$$\sqrt{n}\,((\hat{\mathrm{q}}^\top, \hat{\boldsymbol{\beta}}^\top)^\top - (\mathrm{q}_*^\top, \boldsymbol{\beta}_*^\top)^\top) = -\left[\left(\frac{\partial}{\partial(\mathrm{q}, \boldsymbol{\beta})}\mathcal{H}_n(\mathrm{q}_*, \boldsymbol{\beta}_*)\right)^{-1} + o_p(1)\right]\left(\sqrt{n}\,\mathcal{H}_n(\mathrm{q}_*, \boldsymbol{\beta}_*)\right). \tag{A.11}$$

Condition $(C7)$ implies that the variance matrix V is well-defined and has finite entries. Thus, the standard multivariate CLT for the random variables $h(\mathrm{x}_i, \mathrm{q}, \boldsymbol{\beta})\mathcal{E}(y_i, \mathrm{x}_i, \mathrm{q}_*, \boldsymbol{\beta}_*)$,

$i = 1, \ldots, n$, yields

$$\sqrt{n}\, \mathbb{P}_n(h(\cdot, \mathrm{q}_*, \boldsymbol{\beta}_*)\mathcal{E}(\cdot, \cdot, \mathrm{q}_*, \boldsymbol{\beta}_*)) \xrightarrow{d} N_s(\mathbf{0}, V). \tag{A.12}$$

Under the model (2.7), we have

$$E\left[\frac{\partial}{\partial(\mathbf{t}, \boldsymbol{\beta})}\left[h(\mathbf{x}; \phi^{-1}(\mathbf{t}), \boldsymbol{\beta}_*)\mathcal{E}(y, \mathbf{x}; \phi^{-1}(\mathbf{t}), \boldsymbol{\beta}_*)\right]\Big|_{t=\phi(\mathrm{q}_*)}\right] = G_\phi.$$

Thus, the condition $(C8)$ and the Law of Large Numbers imply

$$\mathbb{P}_n(\frac{\partial}{\partial(\mathrm{q}, \boldsymbol{\beta})}\left[h(\cdot, \mathrm{q}, \boldsymbol{\beta})\mathcal{E}(\cdot, \cdot; \mathrm{q}_*, \boldsymbol{\beta}_*)\right]) \xrightarrow{p} G_\phi. \tag{A.13}$$

Using this together with the condition $(C1)$ and (A.12) in (A.10), by the Sluzky's theorem, we get

$$\sqrt{n}\mathcal{H}_n(\mathrm{q}_*, \boldsymbol{\beta}_*) \xrightarrow{d} N_{d_M + d_{\boldsymbol{\beta}}}(\mathbf{0}, G_\phi^\top W V W G_\phi). \tag{A.14}$$

On the other hand, denoting $G_n(\mathrm{q}, \boldsymbol{\beta}) = \mathbb{P}_n\left(\frac{\partial}{\partial(\mathrm{q}, \boldsymbol{\beta})}\left[h(\cdot; \mathrm{q}, \boldsymbol{\beta})\mathcal{E}(\cdot, \cdot; \mathrm{q}, \boldsymbol{\beta})\right]\right)$, we have

$$\begin{aligned}\frac{\partial}{\partial(\mathrm{q}, \boldsymbol{\beta})}\mathcal{H}_n(\mathrm{q}_*, \boldsymbol{\beta}_*) &= G_n(\mathrm{q}_*, \boldsymbol{\beta}_*)^\top W_n G_n(\mathrm{q}_*, \boldsymbol{\beta}_*) \\ &+ \left(I_{d_M + d_{\boldsymbol{\beta}}} \otimes \left[W_n \mathbb{P}_n(h(\cdot; \mathrm{q}_*, \boldsymbol{\beta}_*)\mathcal{E}(\cdot, \cdot; \mathrm{q}_*, \boldsymbol{\beta}_*))\right]\right)^\top \frac{\partial \mathrm{Vec}\,(G_n)}{\partial(\mathrm{q}, \boldsymbol{\beta})}(\mathrm{q}_*, \boldsymbol{\beta}_*),\end{aligned}$$

where $\otimes$ is the matrix Kronecker product. By the Law of Large Numbers, $G_n(\mathrm{q}_*, \boldsymbol{\beta}_*) \xrightarrow{p} G_\phi$ and $\mathbb{P}_n(h(\cdot; \mathrm{q}_*, \boldsymbol{\beta}_*)\mathcal{E}(\cdot, \cdot; \mathrm{q}_*, \boldsymbol{\beta}_*)) \xrightarrow{p} 0$, and by the Slutky's theorem and the assumption $(C10)$, the second term on the right-hand side of the above equality is $o_p(1)$ and. By the continuous mapping theorem, it follows that

$$\left(\frac{\partial}{\partial(\mathrm{q}, \boldsymbol{\beta})}\mathcal{H}_n(\mathrm{q}_*, \boldsymbol{\beta}_*)\right)^{-1} \xrightarrow{p} (G_\phi^\top W G_\phi)^{-1}. \tag{A.15}$$

Conditions $(C7)$ and $(C9)$ assure that $G_\phi^\top W G_\phi$ is a $(d_M + d_{\boldsymbol{\beta}}) \times (d_M + d_{\boldsymbol{\beta}})$ invertible matrix. Thus, the CLT (2.9) follows by using (A.14) and (A.15) in (A.11).

Let's now consider the case when M is a RS space. For a chart (U, ϕ) on M near q_*, let $\mathbf{t}_* = \phi(\mathrm{q}_*)$. As $(\hat{\mathrm{q}}, \hat{\boldsymbol{\beta}})$ is a consistent estimator for $(\mathrm{q}_*, \boldsymbol{\beta}_*)$, thus $\hat{\mathrm{q}} \in U$ with probability approaching one, as $n \to \infty$. Let denote $\hat{\mathbf{t}} = \phi(\hat{\mathrm{q}})$, when $\hat{\mathrm{q}} \in U$. By the continuous mapping theorem, $(\phi(\hat{\mathrm{q}})^\top, \hat{\boldsymbol{\beta}}^\top)^\top$ is a consistent estimator for $(\phi(\mathrm{q}_*)^\top, \boldsymbol{\beta}_*^\top)^\top$ in $R^{(d_M + d_{\boldsymbol{\beta}})}$. The conditions $(C6) - (C10)$ hold for $(\phi(\mathrm{q}_*)^\top, \boldsymbol{\beta}_*^\top)^\top$ and functions $(\mathbf{t}^\top, \boldsymbol{\beta}^\top)^\top \mapsto h(\mathbf{x}, \phi^{-1}(\mathbf{t}), \boldsymbol{\beta}) \mathcal{E}(y, \mathbf{x}; \phi^{-1}(\mathbf{t}), \boldsymbol{\beta})$. The Euclidean case applied to $(\mathbf{t}^\top, \boldsymbol{\beta}^\top)^\top \mapsto \mathcal{H}_n(\phi^{-1}(\mathbf{t}), \boldsymbol{\beta})$ yields the CLT (2.9). Note that here, in contrast with the Euclidean case, $\mathcal{H}_n(\phi^{-1}(\mathbf{t}), \boldsymbol{\beta})$ is a function of random variables $(y_i, \mathbf{x}_i)$ with y_i being manifold-valued random variables. However, for all $(\mathbf{t}, \boldsymbol{\beta})$, the random variables $\mathcal{E}(y_i, \mathbf{x}_i, \phi^{-1}(\mathbf{t}), \boldsymbol{\beta})$ and $\frac{\partial}{\partial(\mathbf{t}, \boldsymbol{\beta})} \mathcal{E}(y_i, \mathbf{x}_i, \phi^{-1}(\mathbf{t}), \boldsymbol{\beta})$, $i = 1, \ldots, n$, are real vector-valued variables, so all arguments still hold.

The relationship between $\Sigma_{\phi'}$ and Σ_ϕ, i.e. the compatibility of the covariance matrix with the manifold structure of M, follows from the chain rule applied to $\phi' = (\phi' \circ \phi^{-1}) \circ \phi$ near q_*. □

We note that when Θ is compact and the functions

$$(\mathrm{q}, \boldsymbol{\beta}) \to h(\mathbf{x}; \mathrm{q}, \boldsymbol{\beta}) \mathcal{E}(y, \mathbf{x}; \mathrm{q}_*, \boldsymbol{\beta}_*)$$

are continuous for every (y, x), then the following condition

$$(C4') \qquad E[\sup_{(\mathrm{q}, \boldsymbol{\beta})} \|h(\mathbf{x}; \mathrm{q}, \boldsymbol{\beta}) \mathcal{E}(y, \mathbf{x}; \mathrm{q}, \boldsymbol{\beta})\|_{T_{\mathrm{p}} M}] < \infty,$$

together with $(C2)$, implies both $(C4)$ and $(C5)$. Also note that if the conditional mean link function $F(\mathbf{x}, \mathrm{q}, \boldsymbol{\beta})$ is continuous in $(\mathrm{q}, \boldsymbol{\beta}) \in M \times \mathbf{R}^{d_{\boldsymbol{\beta}}}$ uniformly with respect to $\mathbf{x}$, then the one-parameter subgroups $c(\cdot, \mathbf{x}, \mathrm{q}, \boldsymbol{\beta})$ of G, the definition of the intrinsic

residuals $\mathcal{E}$, have the property that $c(1, \mathbf{x}, \mathrm{q}, \boldsymbol{\beta}) \in G$ are continuous in $(\mathrm{q}, \boldsymbol{\beta})$ uniformly with respect to $\mathbf{x}$, which yield that the functions $(\mathrm{q}, \boldsymbol{\beta}) \to \mathcal{E}(y, \mathbf{x}; \mathrm{q}, \boldsymbol{\beta})$ are uniformly continuous with respect to $\mathbf{x}$. If, in addition, $F(\mathbf{x}, \mathrm{q}, \boldsymbol{\beta})$ and $h(\mathbf{x}; \mathrm{q}, \boldsymbol{\beta})$ is twice continuously differentiable in $\mathrm{q}, \boldsymbol{\beta})$ in some neighborhood of $(\mathrm{q}_*, \boldsymbol{\beta}_*)$ with probability one, then $(C6)$ holds.

Optimal GMM estimator given an instrumental function h. Given h such that $V = \mathrm{Var}[h(\mathbf{x}; \mathrm{q}, \boldsymbol{\beta})\mathcal{E}(y, \mathbf{x}; \mathrm{q}_*, \boldsymbol{\beta}_*)]$ is a $s \times s$ positive-definite matrix and taking $W_n = W_h^{opt} = W^{opt} = V^{-1}$ will result in the most efficient estimator in the class of all GMM estimators that use h and its asymptotic variance is $\Sigma_{\phi,h}^{opt} = \Sigma_\phi^{opt} = (G_\phi^\top W^{opt} G_\phi)^{-1}$. Efficiency means that an estimator will have the smallest possible variance (for two symmetric positive semi-definite matrices S and T, we say that matrix S is greater than matrix T and write $S \geq T$, if the matrix $S - T$ is positive semi-definite), i.e for any given weight matrix W the variance Σ_ϕ is greater than Σ_ϕ^{opt}. We have that the variance difference

$$\begin{aligned}
\Sigma_\phi - \Sigma_\phi^{opt} &= (G_\phi^\top W G_\phi)^{-1} G_\phi^\top W V W G_\phi (G_\phi^\top W G_\phi)^{-1} - (G_\phi^\top V^{-1} G_\phi)^{-1} \\
&= (G_\phi^\top W G_\phi)^{-1} G_\phi^\top W V^{1/2} \\
&\quad \times \left(I - V^{-1/2} G_\phi (G_\phi^\top V^{-1} G_\phi)^{-1} G_\phi^\top V^{-1/2}\right) \\
&\quad \times V^{1/2} W G_\phi (G_\phi^\top W G_\phi)^{-1} \\
&= A(I - B)A',
\end{aligned}$$

where A, B are $(d_M + d_{\boldsymbol{\beta}}) \times (d_M + d_{\boldsymbol{\beta}})$ matrices, B symmetric, and $B^2 = B$. Thus, the variance difference is of the form $A(I-B)(I-B)'A'$, which is positive semi-definite, so $\Sigma_\phi \geq \Sigma_\phi^{opt}$. Therefore, given h, the use of $W_n = W^{opt} = V^{-1}$ yields the GMM estimator with the smallest asymptotic covariance matrix among those using the same h.

Proof of Theorem 2.3.2. The asymptotic normality of $(\hat{\mathrm{q}}^*, \hat{\boldsymbol{\beta}}^*)$ follows from Theorem

2.3.1 (here, $s = d_M + d_{\boldsymbol{\beta}}$) with covariance matrix $\Sigma_\phi^* = (G_\phi^{*\top} W_\phi^* G_\phi^*)^{-1}$, where

$$\begin{aligned} G_\phi^* &= E\left[h_\phi^*(\mathbf{x}) \frac{\partial}{\partial(\mathbf{t}, \boldsymbol{\beta})} \mathcal{E}(y, \mathbf{x}; \phi^{-1}(\mathbf{t}), \boldsymbol{\beta}_*)\big|_{\mathbf{t}=\phi(\mathrm{q}_*)} | \mathbf{x}\right] \\ &= E[D_\phi(\mathbf{x}) \Omega(\mathbf{x})^{-1} D_\phi(\mathbf{x})^\top]. \end{aligned}$$

and

$$\begin{aligned} W_\phi^* &= \left(\mathrm{Var}[h_\phi^*(\mathbf{x}) \mathcal{E}(y, \mathbf{x}; \mathrm{q}_*, \boldsymbol{\beta}_*)]\right)^{-1} \\ &= \left(E\left[\mathrm{Var}\left(h_\phi^*(\mathbf{x}) \mathcal{E}(y, \mathbf{x}; \mathrm{q}_*, \boldsymbol{\beta}_*) \,|\, \mathbf{x}\right)\right]\right)^{-1} \\ &= \left(E[h_\phi^*(\mathbf{x}) \mathrm{Var}\left(\mathcal{E}(y, \mathbf{x}; \mathrm{q}_*, \boldsymbol{\beta}_*) \,|\, \mathbf{x}\right) h_\phi^{*\top}(\mathbf{x})]\right)^{-1} \\ &= \left(E[D_\phi(\mathbf{x}) \Omega(\mathbf{x})^{-1} \Omega(\mathbf{x}) \Omega(\mathbf{x})^{-1} D_\phi(\mathbf{x})^\top]\right)^{-1} \\ &= \left(E\left[D_\phi(\mathbf{x}) \Omega(\mathbf{x})^{-1} D_\phi(\mathbf{x})^\top\right]\right)^{-1} = G_\phi^{*-1}. \end{aligned}$$

We also have $G_\phi^{*\top} = G_\phi^*$, since $\Omega(\mathbf{x})$ is a symmetric positive definite matrix. Thus,

$$\Sigma_\phi^* = (G_\phi^* G_\phi^{*-1} G_\phi^*)^{-1} = G_\phi^{*-1} = \left(E[D_\phi(\mathbf{x}) \Omega(\mathbf{x})^{-1} D_\phi(\mathbf{x})^\top]^\top\right)^{-1}.$$

Finally, to prove the optimality of the GMM estimator $(\hat{\mathrm{q}}^*, \hat{\boldsymbol{\beta}}^*)$ among all estimators minimizing the quadratic form in the definition (2.8), it is enough to show that the variance matrix $\Sigma_{\phi,h}^{opt}$ is greater than Σ_ϕ^* or, equivalently, $\Sigma_\phi^{*\,-1} - \Sigma_{\phi,h}^{opt\,-1} \geq 0$, in the sense of positive semi-definite matrices, for any $s \times d_M$ matrix-valued function $h(\mathbf{x}; \mathrm{q}, \boldsymbol{\beta})$. With the notations introduced in this theorem, we have $G_{\phi,h} = E[h D_\phi^\top]$, $W_h^{opt} = E[h \Omega h^\top]^{-1}$,

and $\Sigma_{\phi,h}^{opt} = (E[hD_\phi^\top]^\top E[h\Omega h^\top]^{-1} E[hD_\phi^\top])^{-1}$. Thus, a simple calculation shows

$$\begin{aligned}
&\Sigma_\phi^{*-1} - \Sigma_{\phi,h}^{opt\,-1} = \\
&\quad = E[D_\phi \Omega^{-1} D_\phi^\top] - E[hD_\phi^\top]^\top E[h\Omega h^\top]^{-1} E[hD_\phi^\top] \\
&\quad = E\left[\left(D_\phi - \Omega h^\top E[h\Omega h^\top]^{-1} E[hD_\phi^\top]\right)^\top \Omega^{-1} \left(D_\phi - \Omega h^\top E[h\Omega h^\top]^{-1} E[hD_\phi^\top]\right)\right] \\
&\quad \geq 0,
\end{aligned}$$

since Ω^{-1} is a symmetric positive definite matrix. Therefore, the instrumental function h_ϕ^*, together with its associated optimal weight matrix W_ϕ^*, yields the optimal GMM estimator among all estimators minimizing the quadratic form in the definition (2.8).

□

Proof of Theorem 2.3.3. We use the following short notations, $\hat{D}_i := \hat{D}(\mathbf{x}_i)$, $D_i := D(\mathbf{x}_i)$, $\hat{h}_{E,\phi,i} := \hat{h}_{E,\phi}(\mathbf{x}_i)$, $h_{E,\phi,i}^* := h_{E,\phi}^*(\mathbf{x}_i)$, $\mathcal{E}_i(\mathrm{q}, \boldsymbol{\beta}) := \mathcal{E}(y_i, \mathbf{x}_i; \mathrm{q}, \boldsymbol{\beta})$, and $\Delta_i(\mathrm{q}, \boldsymbol{\beta}) = \partial \mathcal{E}_i(\mathrm{q}, \boldsymbol{\beta}) / \partial(\mathrm{q}, \boldsymbol{\beta})$, for $i = 1, \ldots, n$. For a matrix A let $\|A\|$ be the Frobenius norm, i.e. $\|A\|^2 = \mathrm{tr}(A^\top A)$.

Consider any $(\bar{\mathrm{q}}, \bar{\boldsymbol{\beta}}) \in \Theta$ such that $dist_M(\bar{\mathrm{q}}, \mathrm{q}_*) + \|\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}_*\| = o_p(1)$. With probability approaching one $(\bar{\mathrm{q}}, \bar{\boldsymbol{\beta}}) \in N_*(\delta)$,

$$\begin{aligned}
&\|\mathbb{P}_n(\hat{h}_{E,\phi}(\cdot) \frac{\partial}{\partial(\mathbf{t}, \boldsymbol{\beta})} \mathcal{E}(\cdot, \cdot; \phi^{-1}(\mathbf{t}), \bar{\boldsymbol{\beta}})\big|_{\phi(\bar{\mathrm{q}})}) - \mathbb{P}_n(h_{E,\phi}^*(\cdot) \frac{\partial}{\partial(\mathbf{t}, \boldsymbol{\beta})} \mathcal{E}(\cdot, \cdot; \phi^{-1}(\mathbf{t}), \bar{\boldsymbol{\beta}})\big|_{\phi(\bar{\mathrm{q}})})\| \\
&= \|\frac{1}{n} \sum_{i=1}^n (\hat{h}_{E,\phi,i} - h_{E,\phi,i}^*) \Delta_i(\bar{\mathrm{q}}, \bar{\boldsymbol{\beta}})\| \leq \frac{1}{n} \sum_{i=1}^n \|\hat{h}_{E,\phi,i} - h_{E,\phi,i}^*\| \|\Delta_i(\bar{\mathrm{q}}, \bar{\boldsymbol{\beta}})\| \\
&\leq \frac{1}{n} \sum_{i=1}^n \|\hat{h}_{E,\phi,i} - h_{E,\phi,i}^*\| M_1(y_i, \mathbf{x}_i) \\
&\leq \left(\frac{1}{n} \sum_{i=1}^n \|\hat{h}_{E,\phi,i} - h_{E,\phi,i}^*\|^2\right)^{1/2} \left(\frac{1}{n} \sum_{i=1}^n M_1(y_i, \mathbf{x}_i)^2\right)^{1/2} \\
&= \left(\frac{1}{n} \sum_{i=1}^n \|\hat{h}_{E,\phi,i} - h_{E,\phi,i}^*\|^2\right)^{1/2} O_p(1),
\end{aligned} \tag{A.16}$$

by assumption $E[M_1(y,\mathbf{x})^2] < \infty$. Also $E[\|D(\mathbf{x})\|^2] < \infty$. As $(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)$ is consistent, $dist_{M\times\mathbf{R}^{d_{\boldsymbol{\beta}}}}((\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I),(\mathrm{q}_*, \boldsymbol{\beta}_*)) = o_p(1)$, by the uniform LLN and $V(\mathrm{q}_*, \boldsymbol{\beta}_*)$ nonsingular, then with probability approaching one, $\widehat{V}(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)$ is nonsingular, $\widehat{V}(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)^{-1} = V(\mathrm{q}_*, \boldsymbol{\beta}_*)^{-1} + o_p(1)$. Then,

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} \|\hat{h}_{E,\phi,i} - h^*_{E,\phi,i}\|^4 = \frac{1}{n}\sum_{i=1}^{n} \|\hat{D}_{\phi,i}\hat{V}(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)^{-1} - D_{\phi,i}V(\mathrm{q}_*, \boldsymbol{\beta}_*)^{-1}\|^4 \\
&\quad \le \frac{1}{n}\sum_{i=1}^{n} C\left(\|(\hat{D}_{\phi,i} - D_{\phi,i})\hat{V}(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)^{-1}\|^4 + \|D_{\phi,i}(\hat{V}(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)^{-1} - V(\mathrm{q}_*, \boldsymbol{\beta}_*)^{-1})\|^4\right) \\
&\quad \le C\|\hat{V}(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)^{-1}\|^4 \left(\frac{1}{n}\sum_{i=1}^{n} \|\hat{D}_{\phi,i} - D_{\phi,i}\|^4\right) \\
&\qquad\qquad + C\|\hat{V}(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)^{-1} - V(\mathrm{q}_*, \boldsymbol{\beta}_*)^{-1})\|^4 \left(\frac{1}{n}\sum_{i=1}^{n} \|D_{\phi,i}\|^4\right) \\
&\quad = O_p(1)o_p(1) + o_p(1)O_p(1) = o_p(1),
\end{aligned} \tag{A.17}
$$

where C is an absolute positive constant.

Note that $\sup_{(\mathrm{q},\boldsymbol{\beta})\in N(\delta)} \|h^*_{E,\phi,i}\Delta_i(\mathrm{q},\boldsymbol{\beta})\| \le \|D_{\phi,i}\|\|V(\mathrm{q}_*, \boldsymbol{\beta}_*)^{-1}\|M_1(y_i, \mathbf{x}_i)$, by the uniform LLN $\frac{1}{n}\sum_{i=1}^{n} h^*_{E,\phi,i}\Delta_i(\bar{\mathrm{q}}, \bar{\boldsymbol{\beta}}) = E[D_\phi(x)V(\mathrm{q}_*, \boldsymbol{\beta}_*)^{-1}D_\phi(x)^\top] + o_p(1) = G_{\phi,h^*_{E,\phi}} + o_p(1)$. Then by (A.16) and (A.17), it follows

$$
\frac{1}{n}\sum_{i=1}^{n} \hat{h}_{E,\phi,i}\Delta_i(\bar{\mathrm{q}}, \bar{\boldsymbol{\beta}}) = G_{\phi,h^*_{E,\phi}} + o_p(1) \tag{A.18}
$$

In particular, for $(\bar{\mathrm{q}}, \bar{\boldsymbol{\beta}}) = (\mathrm{q}_*, \boldsymbol{\beta}_*)$, we have $\frac{1}{n}\sum_{i=1}^{n} \hat{D}_{\phi,i}\widehat{V}(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)^{-1}D_{\phi,i}^\top = G_{\phi,h^*_{E,\phi}} + o_p(1)$. Also,

$$
\begin{aligned}
&\|\frac{1}{n}\sum_{i=1}^{n}(\hat{D}_{\phi,i}\widehat{V}(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)\hat{D}_{\phi,i}^\top - \hat{D}_{\phi,i}\widehat{V}(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)^{-1}D_{\phi,i}^\top)\| \\
&\le (n^{-1}\sum_{i=1}^{n} \|\hat{h}_{E,\phi,i}\|^2)^{1/2}(n^{-1}\sum_{i=1}^{n} \|\hat{D}_{\phi,i} - D_{\phi,i}\|^2)^{1/2} \\
&= O_p(1)o_p(1) = o_p(1).
\end{aligned}
$$

It follows that

$$\frac{1}{n}\sum_{i=1}^{n}\hat{D}_{\phi,i}\widehat{V}(\hat{\mathrm{q}}_I,\hat{\boldsymbol{\beta}}_I)^{-1}\hat{D}_{\phi,i}^{\top}=G_{\phi,h^*_{E,\phi}}+o_p(1). \tag{A.19}$$

Next, note that for matrices A_1 and A_2 , $\|A_1A_1^{\top}-A_2A_2^{\top}\|=\|(A_1-A_2)(A_1-A_2)^{\top}+A_2(A_1-A_2)^{\top}+(A_1-A_2)A_2^{\top}\|\leq\|A_1-A_2\|^2+2\|A_1-A_2\|\|A_2\|$.

Let $\hat{B}_i=\hat{h}_{E,\phi,i}\boldsymbol{\mathcal{E}}_i(\hat{\mathrm{q}}_I,\hat{\boldsymbol{\beta}}_I)$ and $\tilde{B}_i=h^*_{E,\phi,i}\boldsymbol{\mathcal{E}}_i(\hat{\mathrm{q}}_I,\hat{\boldsymbol{\beta}}_I)$. Then,

$$\begin{aligned}
&\|\frac{1}{n}\sum_{i=1}^{n}\left(\hat{h}_{E,\phi,i}\boldsymbol{\mathcal{E}}_i(\hat{\mathrm{q}}_I,\hat{\boldsymbol{\beta}}_I)^{\otimes 2}\hat{h}_{E,\phi,i}^{\top}-h^*_{E,\phi,i}\boldsymbol{\mathcal{E}}_i(\hat{\mathrm{q}}_I,\hat{\boldsymbol{\beta}}_I)^{\otimes 2}h^{*\top}_{E,\phi,i}\right)\| \\
&\quad\leq\frac{1}{n}\sum_{i=1}^{n}\|\hat{B}_i\hat{B}_i^{\top}-\tilde{B}_i\tilde{B}_i^{\top}\|\leq\frac{1}{n}\sum_{i=1}^{n}\|\hat{B}_i-\tilde{B}_i\|^2+\frac{2}{n}\sum_{i=1}^{n}\|\hat{B}_i-\tilde{B}_i\|\|\tilde{B}_i\| \\
&\quad\leq\frac{1}{n}\sum_{i=1}^{n}\|\hat{h}_{E,\phi,i}-h^*_{E,\phi,i}\|^2M_0(y_i,\mathbf{x}_i)^2 \\
&\qquad+2\left(\frac{1}{n}\sum_{i=1}^{n}\|\hat{h}_{E,\phi,i}-h^*_{E,\phi,i}\|^2M_0(y_i,\mathbf{x}_i)^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\|h^*_{E,\phi,i}\|^2M_0(y_i,\mathbf{x}_i)^2\right)^{1/2} \\
&\quad\leq\left(\frac{1}{n}\sum_{i=1}^{n}\|\hat{h}_{E,\phi,i}-h^*_{E,\phi,i}\|^4\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}M_0(y_i,\mathbf{x}_i)^4\right)^{1/2} \\
&\qquad+2\left(\frac{1}{n}\sum_{i=1}^{n}\|\hat{h}_{E,\phi,i}-h^*_{E,\phi,i}\|^4\right)^{1/4}\left(\frac{1}{n}\sum_{i=1}^{n}\|h^*_{E,\phi,i}\|^4\right)^{1/4}\left(\frac{1}{n}\sum_{i=1}^{n}M_0(y_i,\mathbf{x}_i)^4\right)^{1/2} \\
&\quad=(o_p(1)O_p(1))^{1/2}+(o_p(1)O_p(1))^{1/4}O_p(1)^{1/2}=o_p(1).
\end{aligned} \tag{A.20}$$

Note that $\sup_{(\mathrm{q},\boldsymbol{\beta})\in N(\delta)}\|h^*_{E,\phi,i}\boldsymbol{\mathcal{E}}_i(\mathrm{q},\boldsymbol{\beta})^{\otimes 2}h^{*\top}_{E,\phi,i}\|\leq\|D_{\phi,i}\|^2\|V(\mathrm{q}_*,\boldsymbol{\beta}_*)^{-1}\|^2M_0(y_i,\mathbf{x}_i)^2$, by the uniform LLN

$$\frac{1}{n}\sum_{i=1}^{n}h^*_{E,\phi,i}\boldsymbol{\mathcal{E}}_i(\hat{\mathrm{q}}_I,\hat{\boldsymbol{\beta}}_I)^{\otimes 2}h^{*\top}_{E,\phi,i}=E[D_{\phi}(\mathbf{x})V(\mathrm{q}_*,\boldsymbol{\beta}_*)^{-1}\Omega(\mathbf{x})V(\mathrm{q}_*,\boldsymbol{\beta}_*)^{-1}D(_{\phi}\mathbf{x})^{\top}]+o_p(1).$$

Then, from (A.20), we have

$$\frac{1}{n}\sum_{i=1}^{n}\hat{h}_{E,\phi,i}\mathcal{E}_i(\hat{\mathrm{q}}_I,\hat{\boldsymbol{\beta}}_I)^{\otimes 2}\hat{h}_{E,\phi,i}^{\top}=E[D_\phi(\mathbf{x})V(\mathrm{q}_*,\boldsymbol{\beta}_*)^{-1}\Omega(\mathbf{x})V(\mathrm{q}_*,\boldsymbol{\beta}_*)^{-1}D_\phi(\mathbf{x})^{\top}]+o_p(1) \tag{A.21}$$

The conclusion (2.17) that $\widehat{\Sigma}_{E,\phi}$ is a consistent estimator for $\Sigma_{E,\phi}$ follows from (A.19), (A.21), the continuous mapping theorem and the Slutzky's theorem.

Next, with A_j denoting the j-th row of a matrix A, we have

$$\begin{aligned}
&\big|n^{-1/2}\sum_{i=1}^{n}(\hat{D}_{\phi,i;j}-D_{\phi,i;j})\widehat{V}(\hat{\mathrm{q}}_I,\hat{\boldsymbol{\beta}}_I)^{-1}\mathcal{E}_i(\mathrm{q}_*,\boldsymbol{\beta}_*)\big| \\
&\quad=\big|\mathrm{tr}[\widehat{V}(\hat{\mathrm{q}}_I,\hat{\boldsymbol{\beta}}_I)^{-1}\ n^{-1/2}\sum_{i=1}^{n}\mathcal{E}_i(\mathrm{q}_*,\boldsymbol{\beta}_*)(\hat{D}_{\phi,i;j}-D_{\phi,i;j})]\big| \\
&\quad\le C\ \|\widehat{V}(\hat{\mathrm{q}}_I,\hat{\boldsymbol{\beta}}_I)^{-1}\|\ \|n^{-1/2}\sum_{i=1}^{n}\mathcal{E}_i(\mathrm{q}_*,\boldsymbol{\beta}_*)(\hat{D}_{\phi,i;j}-D_{\phi,i;j})\| \\
&\quad\le O_p(1)\|n^{-1/2}\sum_{i=1}^{n}(\hat{D}_{\phi,i}-D_{\phi,i})\otimes\mathcal{E}_i(\mathrm{q}_*,\boldsymbol{\beta}_*)\| \\
&\quad=O_p(1)o_p(1)=o_p(1).
\end{aligned} \tag{A.22}$$

Also, we have $E[D_\phi(\mathbf{x})\otimes\mathcal{E}(y,\mathbf{x};\mathrm{q}_*,\boldsymbol{\beta}_*)]=0$ and
$E[\|n^{-1/2}\sum_{i=1}^{n}D_{\phi,i}\otimes\mathcal{E}_i(\mathrm{q}_*,\boldsymbol{\beta}_*)\|^2]=E[\|D_\phi(\mathbf{x})\|^2\|\mathcal{E}(y,\mathbf{x};\mathrm{q}_*,\boldsymbol{\beta}_*)\|^2]<\infty$. It follows that

$$\begin{aligned}
&\big|n^{-1/2}\sum_{i=1}^{n}D_{\phi,i,j}(\widehat{V}(\hat{\mathrm{q}}_I,\hat{\boldsymbol{\beta}}_I)^{-1}-V(\mathrm{q}_*,\boldsymbol{\beta}_*)^{-1})\mathcal{E}_i(\mathrm{q}_*,\boldsymbol{\beta}_*)\big| \\
&\quad=\left|\mathrm{tr}\left[(\widehat{V}(\hat{\mathrm{q}}_I,\hat{\boldsymbol{\beta}}_I)^{-1}-V(\mathrm{q}_*,\boldsymbol{\beta}_*)^{-1})n^{-1/2}\sum_{i=1}^{n}\mathcal{E}_i(\mathrm{q}_*,\boldsymbol{\beta}_*)D_{\phi,i,j}\right]\right| \\
&\quad\le C\ \|\widehat{V}(\hat{\mathrm{q}}_I,\hat{\boldsymbol{\beta}}_I)^{-1}-V(\mathrm{q}_*,\boldsymbol{\beta}_*)^{-1}\|\ \|n^{-1/2}\sum_{i=1}^{n}\mathcal{E}_i(\mathrm{q}_*,\boldsymbol{\beta}_*)D_{\phi,i,j}\| \\
&\quad\le o_p(1)\|n^{-1/2}\sum_{i=1}^{n}D_{\phi,i}\otimes\mathcal{E}_i(\mathrm{q}_*,\boldsymbol{\beta}_*)\|=o_p(1).
\end{aligned} \tag{A.23}$$

Since this is true for each j, $j = 1, \ldots, d_M + d_{\boldsymbol{\beta}}$, it follows that

$$\|n^{-1/2} \sum_{i=1}^{n} (\hat{h}_{E,\phi,i} - h^*_{E,\phi,i}) \mathcal{E}_i(\mathrm{q}_*, \boldsymbol{\beta}_*)\| = o_p(1). \tag{A.24}$$

From (2.15), using the Taylor expansion of $\sum_{i=1}^{n} h^*_{E,\phi,i} \mathcal{E}_i(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)$ at $(\mathrm{q}_*, \boldsymbol{\beta}_*)$, twice (A.18) with $(\bar{\mathrm{q}}, \bar{\boldsymbol{\beta}}) = (\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)$ and $(\bar{\mathrm{q}}, \bar{\boldsymbol{\beta}}) = (\mathrm{q}_*, \boldsymbol{\beta}_*)$, respectively, (A.24), and the Slutzky's theorem, it follows that

$$\begin{aligned}
&\sqrt{n}[(\phi(\tilde{\mathrm{q}}_E)^\top, \tilde{\boldsymbol{\beta}}_E^\top)^\top - (\phi(\mathrm{q}_*)^\top, \boldsymbol{\beta}_*^\top)^\top] \\
&\quad = \left[\frac{1}{n} \sum_{i=1}^{n} \hat{h}_{E,\phi,i} \frac{\partial \mathcal{E}_i}{\partial(\mathbf{t}, \boldsymbol{\beta})} (\phi^{-1}(t), \hat{\boldsymbol{\beta}}_I) \Big|_{\mathbf{t} = \phi(\hat{\mathrm{q}}_I)} \right]^{-1} \left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{h}_{E,\phi,i} \mathcal{E}_i(\mathrm{q}_*, \boldsymbol{\beta}_*) \right] + o_p(1).
\end{aligned}$$

The first conclusion, (2.16), of the theorem follows now from the above equation by standard central limit arguments, namely, $n^{-1/2} \sum_{i=1}^{n} h^*_{E,\phi,i} \mathcal{E}_i(\mathrm{q}_*, \boldsymbol{\beta}_*) \xrightarrow{d} N(0, W^*_{E,\phi})$, with $W^*_{E,\phi}$ as defined in (2.13), and then using (A.18) with $(\bar{\mathrm{q}}, \bar{\boldsymbol{\beta}}) = (\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)$ and the Slutzky's theorem. This completes the proof. □

Proof of Theorem 2.3.4. The proof follows similar steps as in Theorem 4.3 with straightforward modifications, which for brevity are omitted. □

Proof of Theorem 2.3.5. Let (U, ϕ) be a chart on $\mathcal{M}$ near q_*.

(i) We only prove the result for $W^{(2)}_{n,\phi}$ as follows. Under $H^{(2)}_0$, the true value q_* equals q_0 and (U, ϕ) is a chart near q_0. As $\tilde{\mathrm{q}}_E$ is a consistent estimator for q, it follows that $\tilde{\mathrm{q}}_E \in U$, for n large enough, with probability approaching 1. From the CLT for $\tilde{\mathrm{q}}_E$, we have that, under $H^{(2)}_0$,

$$\sqrt{n}\,(\phi(\tilde{\mathrm{q}}_E) - \phi(\mathrm{q}_0)) \xrightarrow{d} N_{d_{\mathcal{M}}}(\mathbf{0}, (I_{d_{\mathcal{M}}}\ \mathbf{0}) \Sigma_{E,\phi} (I_{d_{\mathcal{M}}}\ \mathbf{0})^T)$$

As $n\widehat{\Sigma}_{E,\phi} \xrightarrow{p} \Sigma_{E,\phi}$, by the continuous mapping theorem, we get

$$\frac{1}{\sqrt{n}}\left[(I_{d_{\mathcal{M}}}\ \mathbf{0})\widehat{\Sigma}_{E,\phi}(I_{d_{\mathcal{M}}}\ \mathbf{0})^T\right]^{-1/2} \xrightarrow{p} \left[(I_{d_{\mathcal{M}}}\ \mathbf{0})\Sigma_{E,\phi}(I_{d_{\mathcal{M}}}\ \mathbf{0})^T\right]^{-1/2}.$$

Then, using Slutzky's theorem, we have

$$\left[(I_{d_{\mathcal{M}}}\ \mathbf{0})\widehat{\Sigma}_{E,\phi}(I_{d_{\mathcal{M}}}\ \mathbf{0})^T\right]^{-1/2}(\phi(\tilde{\mathrm{q}}_E) - \phi(\mathrm{q}_0)) \xrightarrow{d} N_{d_{\mathcal{M}}}(\mathbf{0}, I_{d_{\mathcal{M}}})$$

which implies $W^{(2)}_{n,\phi} \xrightarrow{d} \chi^2_{d_{\mathcal{M}}}$.

(ii) Since $\tilde{\boldsymbol{\beta}}_E$ and the lower-right $d_{\boldsymbol{\beta}} \times d_{\boldsymbol{\beta}}$ submatrix of $\widehat{\Sigma}_{E,\phi}$ are independent of the chart (U, ϕ), so is $W^{(1)}_{n,\phi}$.

(iii) Let (U', ϕ') be another chart near q_0 with $\hat{\mathrm{q}}_E \in U'$. A Taylor's series expansion of the transition function $\phi' \circ \phi$ about $\phi(\mathrm{q}_0)$ shows that $\phi'(\mathrm{q}_E) - \phi'(\mathrm{q}_0) = (J(\phi' \circ \phi)_{\phi(\mathrm{q}_0)} + o_p(1))(\phi(\mathrm{q}_E) - \phi(\mathrm{q}_0))$. Let $\hat{\mathrm{q}}$ be the consistent estimator of q that the asymptotic covariance estimator $\widehat{\Sigma}_{E,\phi}$ is based on. As $\widehat{\Sigma}_{E,\phi}$ is compatible with the manifold structure of $\mathcal{M}$ and $J(\phi' \circ \phi)_{\phi(\hat{\mathrm{q}})} = J(\phi' \circ \phi)_{\phi(\mathrm{q}_0)} + o_p(1)$, we have

$$\begin{aligned}
W^{(2)}_{n,\phi'} &= [\phi(\tilde{\mathrm{q}}_E) - \phi(\mathrm{q}_0)]^T [J(\phi' \circ \phi)_{\phi(\mathrm{q}_0)} + o_p(1)]^T \\
&\quad \times \left[\left(J(\phi' \circ \phi)_{\phi(\hat{\mathrm{q}})}\ \mathbf{0}\right) \widehat{\Sigma}_{E,\phi} \left(J(\phi' \circ \phi)_{\phi(\hat{\mathrm{q}})}\ \mathbf{0}\right)^T\right]^{-1} \\
&\quad \times [J(\phi' \circ \phi)_{\phi(\mathrm{q}_0)} + o_p(1)][\phi(\tilde{\mathrm{q}}_E) - \phi(\mathrm{q}_0)] \\
&= [\phi(\tilde{\mathrm{q}}_E) - \phi(\mathrm{q}_0)]^T [J(\phi' \circ \phi)^{-1}_{\phi(\mathrm{q}_0)} + o_p(1)]^{-\top} \\
&\quad \times \left[\left(J(\phi' \circ \phi)_{\phi(\mathrm{q}_0)} + o_p(1)\ \mathbf{0}\right) \widehat{\Sigma}_{E,\phi} \left(J(\phi' \circ \phi)_{\phi(\mathrm{q}_0)} + o_p(1)\ \mathbf{0}\right)^T\right]^{-1} \\
&\quad \times [J(\phi' \circ \phi)^{-1}_{\phi(\mathrm{q}_0)} + o_p(1)]^{-1} [\phi(\tilde{\mathrm{q}}_E) - \phi(\mathrm{q}_0)] \\
&= [\phi(\tilde{\mathrm{q}}_E) - \phi(\mathrm{q}_0)]^T \left[(I_{d_{\mathcal{M}}}\ \mathbf{0})\, \widehat{\Sigma}_{E,\phi}\, (I_{d_{\mathcal{M}}}\ \mathbf{0})^T + o_p(1)\right]^{-1} [\phi(\tilde{\mathrm{q}}_E) - \phi(\mathrm{q}_0)] \\
&= W^{(2)}_{n,\phi} + o_p(1).
\end{aligned}$$

Thus, $W^{(2)}_{n,\phi'}$ and $W^{(2)}_{n,\phi}$ are asymptotically equivalent.

(iv) Let ϕ and ϕ' be two normal charts on $\mathcal{M}$ centered at $\tilde{\mathrm{q}}_E$. Thus, $\phi(\cdot) = A \circ \mathrm{Log}_{\tilde{\mathrm{q}}_E}(\cdot)$ and $\phi'(\cdot) = A' \circ \mathrm{Log}_{\tilde{\mathrm{q}}_E}(\cdot)$, where $A, A' : T_{\tilde{\mathrm{q}}_E}\mathcal{M} \to R^{d_{\mathcal{M}}}$ are two isomorphisms of linear spaces induced by the coordinates with respect to two orthonormal bases of $T_{\tilde{\mathrm{q}}_E}\mathcal{M}$. Therefore, $\phi'(\cdot) = O\phi(\cdot)$, where $O = A'A^{-1}$ corresponds to an orthonormal matrix, and $\widehat{\Sigma}_{E,\phi';11} = O\widehat{\Sigma}_{E,\phi;11}O^T$. Thus, $\widehat{\Sigma}_{E;11} := A^{-1} \circ \widehat{\Sigma}_{E,\phi;11} \circ A$ is independent of the chart ϕ and is a 1-1 linear map from $T_{\tilde{\mathrm{q}}_E}\mathcal{M}$ onto itself. Since A preserves the inner product, we have

$$\begin{aligned} W^{(2)}_{n,\phi} &= \mathrm{tr}\{[\widehat{\Sigma}^{-1}_{E,\phi;11}A(\mathrm{Log}_{\tilde{\mathrm{q}}_E}(\mathrm{q}_0))]^\top A(\mathrm{Log}_{\tilde{\mathrm{q}}_E}(\mathrm{q}_0))\} \\ &= \mathrm{tr}\{[A((\widehat{\Sigma}_{E;11})^{-1}\mathrm{Log}_{\tilde{\mathrm{q}}_E}(\mathrm{q}_0))]^\top A(\mathrm{Log}_{\tilde{\mathrm{q}}_E}(\mathrm{q}_0))\} \\ &= \mathrm{m}_{\tilde{\mathrm{q}}_E}(\widehat{\Sigma}_{E;11})^{-1}\mathrm{Log}_{\tilde{\mathrm{q}}_E}(\mathrm{q}_0), \mathrm{Log}_{\tilde{\mathrm{q}}_E}(\mathrm{q}_0)) = W^{(2)}_{M,n}. \end{aligned}$$

□

Proof of Theorem 2.3.6. The proof follows from a straightforward application of a Taylor's series expansion and Slutzky's theorem. We only prove (ii). We have that, under $H^{(2)}_{1,n}$,

$$\sqrt{n}\,(\phi(\tilde{\mathrm{q}}_E) - \phi(\mathrm{q}_n)) \xrightarrow{d} N_{d_{\mathcal{M}}}(\mathbf{0}, (I_{d_{\mathcal{M}}}\ \mathbf{0})\Sigma_{E,\phi}(I_{d_{\mathcal{M}}}\ \mathbf{0})^T),$$

where $\mathrm{q}_n = \mathrm{Exp}_{\mathrm{q}_0}(\mathbf{v}/\sqrt{n} + o(1/\sqrt{n}))$. As $n\widehat{\Sigma}_{E,\phi} \xrightarrow{p} \Sigma_{E,\phi}$, by the continuous mapping theorem, we get $n\widehat{\Sigma}_{E,\phi;11} \xrightarrow{p} \Sigma_{E,\phi;11}$ Then, using Slutzky's theorem, we have

$$\left[\widehat{\Sigma}_{E,\phi;11}\right]^{-1/2}(\phi(\tilde{\mathrm{q}}_E) - \phi(\mathrm{q}_n)) \xrightarrow{d} N_{d_{\mathcal{M}}}(\mathbf{0}, I_{d_{\mathcal{M}}})$$

From Taylor's series expansion of the map $\phi \circ \mathrm{Exp}_{\mathrm{q}_0}$ at $\mathbf{0}$, we have

$$\sqrt{n}(\phi(\mathrm{q}_n) - \phi(\mathrm{q}_0)) = J(\phi \circ \mathrm{Exp}_{\mathrm{q}_0})_{\mathbf{0}}(\mathbf{v}) + o(1).$$

Thus, again using Slutzky's theorem, we obtain that

$$\left[\widehat{\Sigma}_{E,\phi;11}\right]^{-1/2} (\phi(\tilde{\mathrm{q}}_E) - \phi(\mathrm{q}_0)) \xrightarrow{d} N_{d_{\mathcal{M}}}([\Sigma_{E,\phi;11}]^{-1/2} J(\phi \circ \mathrm{Exp}_{\mathrm{q}_0})_{\mathbf{0}}(\mathbf{v}), I_{d_{\mathcal{M}}}),$$

which implies that, under $H_{1,n}^{(2)}$, $W_{n,\phi}^{(2)}$ converges in distribution to a noncentral $\chi^2_{d_{\mathcal{M}}}$ with noncentrality parameter

$$J(\phi \circ \mathrm{Exp}_{\mathrm{q}_0})_0(\mathbf{v})^T \left[\hat{\Sigma}_{E,\phi;11}\right]^{-1} J(\phi \circ \mathrm{Exp}_{\mathrm{q}_0})_0(\mathbf{v}).$$

(iii) It follows from (ii) applied to a normal chart $\phi = \mathrm{Log}_{\tilde{\mathrm{q}}_E}$ near $\tilde{\mathrm{q}}_E$.

□

Proof of Theorem 2.3.7. Consider a Taylor's series expansion of the real-valued function $\mathrm{dist}_{\mathcal{M}}(\mathrm{q}, \mathrm{q}_0)^2$ around the point q_*

$$\begin{aligned} \mathrm{dist}_{\mathcal{M}}(\mathrm{q}, \mathrm{q}_0)^2 = \mathrm{dist}_{\mathcal{M}}(\mathrm{q}_*, \mathrm{q}_0)^2 &+ \mathrm{grad}_{\mathrm{q}_*}(\mathrm{dist}_{\mathcal{M}}(\cdot, \mathrm{q}_0)^2)(\mathrm{Log}_{\mathrm{q}_*}(\mathrm{q})) \\ &+ \frac{1}{2}\mathrm{Hess}_{\mathrm{q}_*}(\mathrm{dist}_{\mathcal{M}}(\cdot, \mathrm{q}_0)^2)(\mathrm{Log}_{\mathrm{q}_*}(\mathrm{q}), \mathrm{Log}_{\mathrm{q}_*}(\mathrm{q})) \\ &+ O(\|\mathrm{Log}_{\mathrm{q}_*}(\mathrm{q})\|^3), \end{aligned}$$

for any q in a normal chart centered at q_* with $\mathrm{dist}_{\mathcal{M}}(\mathrm{q}, \mathrm{q}_*) < \rho^*_{\mathcal{M}}$. The result depending on which method is used, implies

$$\sqrt{n}\,\mathrm{Log}_{\mathrm{q}_*}(\tilde{\mathrm{q}}_E) \xrightarrow{d} N_{d_{\mathcal{M}}}(\mathbf{0}, \Sigma_{E,\mathrm{Log}_{\mathrm{q}_*};11}), \tag{A.25}$$

where $\Sigma_{E,\mathrm{Log}_{\mathrm{q}_*}}$ is the matrix representation of the asymptotic covariance matrix of $\tilde{\mathrm{q}}_E$ with respect to the orthonormal basis in $T_{\mathrm{q}_*}\mathcal{M}$ associated with the normal chart under consideration. The squared distance function becomes $\mathrm{dist}_{\mathcal{M}}(\mathrm{Exp}_{\mathrm{q}_*}(\mathbf{v}), \mathrm{q}_*)^2 = \|\mathbf{v}\|^2_{T_{\mathrm{q}_*}M}$, for any $\mathbf{v} \in T_{\mathrm{q}_*}\mathcal{M}$ with $\|\mathbf{v}\|_{T_{\mathrm{q}_*}M} < \rho^*_{\mathcal{M}}$, and the matrix representation of its Hessian at the chart center q_* is the identity matrix $I_{d_{\mathcal{M}}}$, with respect to any orthonormal basis of $T_{\mathrm{q}_*}\mathcal{M}$.

(a) Under the null hypothesis $H_0^{(2)}$, $\tilde{\mathrm{q}}_E$ belongs to a normal chart centered at q_0 with probability approaching one, and

$$W_{dist} = \mathrm{m}_{\mathrm{q}_0}(\mathrm{Log}_{\mathrm{q}_0}(\hat{\mathrm{q}}_E), \mathrm{Log}_{\mathrm{q}_0}(\hat{\mathrm{q}}_E)) = (\mathrm{Log}_{\mathrm{q}_0}(\hat{\mathrm{q}}_E))^T(\mathrm{Log}_{\mathrm{q}_0}(\hat{\mathrm{q}}_E)),$$

when $\mathrm{Log}_{\mathrm{q}_0}(\hat{\mathrm{q}}_E)$ is expressed in the orthonormal basis of $T_{\mathrm{q}_0}\mathcal{M}$ associated with the normal chart centered at $\mathrm{q}_* = \mathrm{q}_0$. From this and (A.25), it follows that $nW_{dist} \xrightarrow{d} \chi^2(\lambda_1, \ldots, \lambda_{d_{\mathcal{M}}})$, where $\lambda_1, \ldots, \lambda_{d_{\mathcal{M}}}$ are the eigenvalues of the matrix $\Sigma_{E,\mathrm{Log}_{\mathrm{q}_0},11}$. Let $\Sigma_{E,\mathrm{Log}_{\mathrm{q}_0}}$ and $\Sigma'_{E,\mathrm{Log}_{\mathrm{q}_0}}$ be the matrix representations of the asymptotic covariance matrix of $\tilde{\mathrm{q}}_E$ in two normal charts centered at q_0. Then $\Sigma'_{E,\mathrm{Log}_{\mathrm{q}_0},11} = O\Sigma_{E,\mathrm{Log}_{\mathrm{q}_0},11}O^T$, for some $d_{\mathcal{M}} \times d_{\mathcal{M}}$ orthogonal matrix O, so the eigenvalues $\lambda_1, \ldots, \lambda_{d_{\mathcal{M}}}$ are independent of the normal chart.

(b) Under the alternative hypothesis $H_1^{(2)}$, from the Taylor's series expansion above, we have

$$\begin{aligned} W_{dist} - \mathrm{dist}_{\mathcal{M}}(\mathrm{q}_*, \mathrm{q}_0)^2 &= \mathrm{grad}_{\mathrm{q}_*}(\mathrm{dist}_{\mathcal{M}}(\cdot, \mathrm{q}_0)^2)[\mathrm{Log}_{\mathrm{q}_*}(\tilde{\mathrm{q}}_E) + O_p(\|\mathrm{Log}_{\mathrm{q}_*}\tilde{\mathrm{q}}_E\|^2)] \\ &= [D^T_{dist} + o_p(1)]\mathrm{Log}_{\mathrm{q}_*}(\tilde{\mathrm{q}}_E). \end{aligned}$$

Using Slutzky's theorem, we get

$$\sqrt{n}(W_{dist} - \mathrm{dist}_{\mathcal{M}}(\mathrm{q}_*, \mathrm{q}_0)^2) \xrightarrow{d} N_{d_{\mathcal{M}}}(\mathbf{0}, D^T_{dist}\Sigma_{E,\mathrm{Log}_{\mathrm{q}_*},11}\, D_{dist}).$$

In the case when q_0 is close to q_* so that q_0 is in a normal chart centered at q_*, then $D_{dist} = \mathrm{grad}_{\mathrm{q}_*}(\mathrm{dist}_{\mathcal{M}}(\cdot, \mathrm{q}_0)^2) = -2\mathrm{Log}_{\mathrm{q}_*}\mathrm{q}_0$ and we have

$$\sqrt{n}(W_{dist} - \mathrm{dist}_{\mathcal{M}}(\mathrm{q}_*, \mathrm{q}_0)^2) \xrightarrow{d} N_{d_{\mathcal{M}}}(\mathbf{0}, 4[\mathrm{Log}_{\mathrm{q}_*}\mathrm{q}_0]^T \Sigma_{E,\mathrm{Log}_{\mathrm{q}_*},11}\,[\mathrm{Log}_{\mathrm{q}_*}\mathrm{q}_0],$$

which completes the proof. □

Proof of Theorem 2.3.8. We introduce some notation. For any chart (U, ϕ) on $\mathcal{M}$ with $\mathrm{q}_0 \in U$, we define $F^*_{\phi i}$ and U^* in a similar way as $F_{\phi i}$ and U^*, respectively, by replacing $(\mathrm{q}_0, \tilde{\boldsymbol{\beta}}_I)$ with $(\mathrm{q}_*, \boldsymbol{\beta}_*)$. That is,

$$F^*_{\phi i} = (F^{*\top}_{\phi i,1}, F^{*\top}_{\phi i,2})^\top = \partial_{(\mathbf{t},\boldsymbol{\beta})}\mathrm{dist}_{\mathcal{M}}(f(\mathbf{x}_i, \phi^{-1}(\mathbf{t}), \boldsymbol{\beta}), y_i)^2\big|_{\mathbf{t}=\phi(\mathrm{q}_*),\boldsymbol{\beta}_*},$$
$$\mathrm{U}^* = \begin{pmatrix} \mathrm{U}^*_{\mathbf{tt}} & \mathrm{U}^*_{\mathbf{t}\boldsymbol{\beta}} \\ \mathrm{U}^*_{\boldsymbol{\beta}\mathbf{t}} & \mathrm{U}^*_{\boldsymbol{\beta}\boldsymbol{\beta}} \end{pmatrix} = \sum_{i=1}^n \partial^2_{(\mathbf{t},\boldsymbol{\beta})}\mathrm{dist}_{\mathcal{M}}(f(\mathbf{x}_i, \phi^{-1}(\mathbf{t}), \boldsymbol{\beta}), y_i)^2\big|_{\mathbf{t}=\phi(\mathrm{q}_*),\boldsymbol{\beta}_*},$$

where the subcomponents $F^{*\top}_{\phi i,1}$ and $F^{*\top}_{\phi i,2}$ correspond to $\mathbf{t}$ and $\boldsymbol{\beta}$, respectively.

(i) The key idea in deriving the asymptotic distribution of $W_{SC,\phi}$ consists of two steps. In Step 1, using a Taylor's series expansion of $\sum_{i=1}^n \tilde{\mathrm{U}}_{i,2}$ at $(\phi(\mathrm{q}_*), \boldsymbol{\beta}_*)$, we can show that, under the *null* hypothesis $H_0^{(2)}$,

$$\tilde{\boldsymbol{\beta}}_I - \boldsymbol{\beta}_* = (-\mathrm{U}^*_{\boldsymbol{\beta}\boldsymbol{\beta}})^{-1} \sum_{i=1}^n F^*_{\phi i,2} + O_p(n^{-1}).$$

In Step 2, under $H_0^{(2)}$, we expand $\sum_{i=1}^n F_{\phi i,1}$ at $(\phi(\mathrm{q}_*), \boldsymbol{\beta}_*)$ to get

$$\begin{aligned}
\sum_{i=1}^n F_{\phi i,1} &= \sum_{i=1}^n F^*_{\phi i,1} + \mathrm{U}^*_{\mathbf{t}\boldsymbol{\beta}}(\tilde{\boldsymbol{\beta}}_I - \boldsymbol{\beta}_*)[1+o_p(1)] \\
&= \sum_{i=1}^n F^*_{\phi i,1} - \mathrm{U}^*_{\mathbf{t}\boldsymbol{\beta}}\mathrm{U}^{*-1}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\sum_{i=1}^n F^*_{\phi i,2})[1+o_p(1)] \\
&= (I_{d_{\mathcal{M}}}, -\mathrm{U}^*_{\mathbf{t}\boldsymbol{\beta}}(\mathrm{U}^*_{\boldsymbol{\beta}\boldsymbol{\beta}})^{-1}) \left(\sum_{i=1}^n F^*_{\phi i}\right)[1+o_p(1)].
\end{aligned}$$

Thus, by using Slutzky's theorem, we have

$$(I_{d_{\mathcal{M}}}, -\mathrm{U}^*_{\mathbf{t}\boldsymbol{\beta}}(\mathrm{U}^*_{\boldsymbol{\beta}\boldsymbol{\beta}})^{-1})\frac{1}{\sqrt{n}}\left(\sum_{i=1}^n F^*_{\phi i}\right) \xrightarrow{d} N_{d_{\mathcal{M}}}(\mathbf{0}, \Sigma_{\phi,\mathrm{q}_*}),$$

where $\Sigma_{\phi,\mathrm{q}_*}$ is given by

$$E\Big\{[(I_{d_{\mathcal{M}}}, -\mathrm{U}^*_{\mathbf{t}\boldsymbol{\beta}}(\mathrm{U}^*_{\boldsymbol{\beta}\boldsymbol{\beta}})^{-1})\partial_{\mathbf{t},\boldsymbol{\beta}}\mathrm{dist}_{\mathcal{M}}(f(\mathbf{x}, \phi^{-1}(\mathbf{t}), \boldsymbol{\beta}_*), y)^2\big|_{\mathbf{t}=\phi(\mathrm{q}_*)}]^{\otimes 2}\Big\}.$$

Since $\tilde{\Sigma}_{\phi,\mathrm{q}} = [n^{-1}\sum_{i=1}^n [(I_{d_{\mathcal{M}}}, -\mathrm{U}_{\mathbf{t}\boldsymbol{\beta}}\mathrm{U}^{-1}_{\boldsymbol{\beta}\boldsymbol{\beta}})(F_{\phi i} - \overline{F}_\phi)]^{\otimes 2}] \xrightarrow{p} \Sigma_\phi$, it follows from the continuous mapping theorem and Slutzky's theorem that under $H_0^{(2)}$, the score test statistic $W_{SC,\phi} = (\sum_{i=1}^n F_{\phi i,1})^\top \tilde{\Sigma}^{-1}_{\phi,\mathrm{q}}(\sum_{i=1}^n F_{\phi i,1}) \xrightarrow{d} \chi^2_{d_{\mathcal{M}}}$.

(ii) Let (U', ϕ') be another chart on $\mathcal{M}$ with $\mathrm{q}_0 \in U'$. Under $H_0^{(2)}$, by the chain rule, we have $F_{\phi',i} = \mathrm{diag}(J(\phi\circ\phi'^{-1})_{\phi'(\mathrm{q}_0)}, I_{d_{\boldsymbol{\beta}}})^T F_{\phi,i}$ and $\mathrm{U}_{\phi',\mathbf{t}'\boldsymbol{\beta}} = J(\phi\circ\phi'^{-1})^T_{\phi'(\mathrm{q}_0)}\mathrm{U}_{\phi,\mathbf{t}\boldsymbol{\beta}}$. It immediately follows that

$$\begin{aligned}
F_{\phi' i,1} &= J(\phi\circ\phi'^{-1})^T_{\phi'(\mathrm{q}_0)} F_{\phi i,1}, \\
\tilde{\Sigma}_{\phi',\mathrm{q}} &= J(\phi\circ\phi'^{-1})^T_{\phi'(\mathrm{q}_0)} \tilde{\Sigma}_{\phi,\mathrm{q}} J(\phi\circ\phi'^{-1})_{\phi'(\mathrm{q}_0)},
\end{aligned}$$

which implies $W_{SC,\phi'} = W_{SC,\phi}$. Thus, the score test statistic $W_{SC,\phi}$ is independent of the chart (U, ϕ) near q_0. □

A.3 RSS method versus more naïve approaches

More naïve approaches that ignore nonlinearity or deal with it in a simple minded fashion may be considered. One example of such an approach is to compute a Fréchet mean of the observed response variable, map the responses to the tangent space at this mean, using the the logarithmic map, then apply a standard (Euclidean) regression framework in the tangent space, and then map the results back on the manifold via the exponential map. We simulated circular data, i.e. $\mathcal{M} = S^1$, to illustrate that our method outperforms the naïve method in terms of both prediction accuracy and estimation efficiency, when the distribution of the responses strongly depends on the covariates and the covariates have substantial variation.

Let $\mathcal{M} = S^1$ be the unit circle in R^2 and the âĂIJtrueâĂİ data $y_1^*, \ldots, y_{16}^* \in S^1$ be the vertices of a polygon with all the sides of equal length, except for the side y_8^*, y_9^* which has a longer length, and it is symmetric about the diameter through the East point $E(1, 0)$. For simplicity, we distorted the regular polygon to ensure that there is no rotational invariance of data and the FrÃľchet mean of y_i^*âĂŹs is unique, in this case the point E. Let x_i be the covariate associated with the response y_i^*, x_i be the polar angle of y_i^* from the East point. The covariates are assumed fixed and known. We set $x_i = 0.9\frac{\pi}{8}(2i - 1)$ for $i = 1, \ldots, 8$, and $x_i = -0.9\frac{\pi}{8}[2(16 - i) + 1]$ for $i = 9, \ldots, 16$. We generated noisy data y_i, $i = 1, \ldots, 16$, from the âĂIJtrueâĂİ data as follows. We introduced a random $N(0, \sigma^2)$ noise around 0 in the tangent space to S^1 at each y_i^* and then project it onto S^1 via the exponential map at y_i^* to obtain the observed data y_i. We consider different levels of noise by varying σ and the link

Table A.1: Comparison of the performance of RSS model and naïve method for the simulated circular data in Section A.3.

	True Values	RSS model estimates	Naïve method estimates
q	$(1, 0)$	$(0.999, -0.001)$	$(0.001, 0.999)$
β	1	0.869	-0.070
SSR	0	17.840	38.759

function: $\boldsymbol{\mu}(x, \mathrm{q}, \beta) : (-pi, \pi) \times S^1 \times R \to S^1$ given by

$$\begin{aligned}
\boldsymbol{\mu}(x, \mathrm{q}, \beta) &= T^{-1}_{st;-\mathrm{q}}(\beta T_{st;-\mathrm{q}} R_{E,q}(\cos x, \sin x)) \\
&= R_{E,q}(\cos(2\tan^{-1}(\beta \tan(x/2))), \sin(2\tan^{-1}(\beta \tan(x/2)))) \\
&= R_{E,q}\left(\frac{1-\beta^2\tan^2(x/2)}{1+\beta^2\tan^2(x/2)}, \frac{2\beta\tan(x/2)}{1+\beta^2\tan^2(x/2)}\right),
\end{aligned}$$

where $T_{st;-\mathrm{q}}$ is the stereographic projection of S^1 from the point $-\mathrm{q}$ onto the tangent space of S^1 at q and $R_{\mathrm{p,q}}$ is the $2D$-rotation that maps p over q. Note that $T_{st;-E}(\cos x, \sin x) = (1, 2\tan(x/2))$, $\boldsymbol{\mu}(0, \mathrm{q}, \beta) = \boldsymbol{\mu}(x, \mathrm{q}, 0) = \mathrm{q}$, and $\boldsymbol{\mu}(x, \mathrm{q}, 1) = R_{E,q}((\cos x, \sin x)) = (\cos(x + x_{\mathrm{q}}), \sin(x + x_{\mathrm{q}}))$, where x_{q} is the polar angle of q from the East point. For the "true" data, we have $y_i^* = \boldsymbol{\mu}(x_i, E, 1)$, for $i = 1, \ldots, 16$. The true values of q and β are $\mathrm{q}_* = E = (1, 0)$ and $\beta_* = 1$, respectively.

To compare the performance of our model and the naïve method, we consider the circular data presented above with random standard normal distributed noise ($\sigma = 1$). Under our model, we estimate q and σ simultaneously, while under the naïve method, q is estimated first, as the FrÃľchet mean of the observed data, and thenβ. Table A.1 and Figure A.2 below display the results from both methods, indicating that our method outperforms the naïve method.

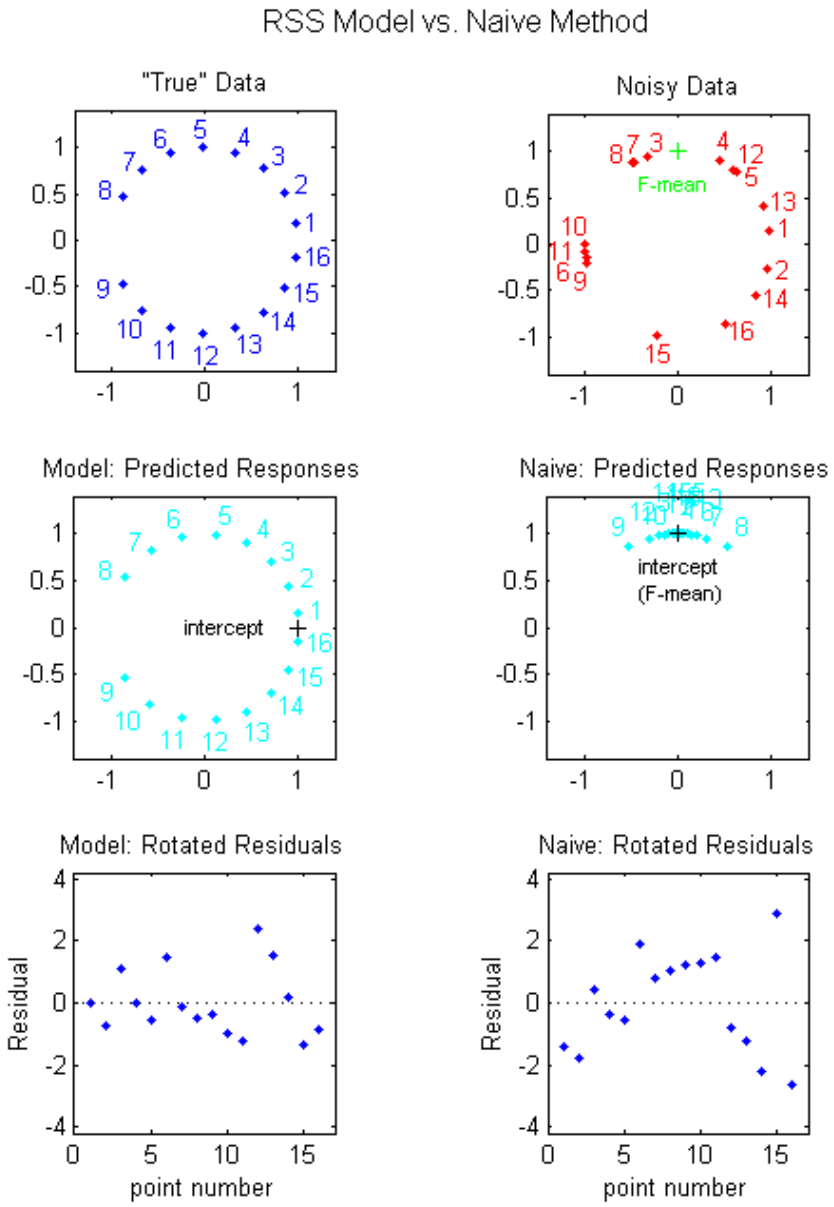

Figure A.2: RSS model vs. naïve method: predicted responses an rotated residuals, for the simulated circular data in Section A.3.

A.4 Annealing evolutionary stochastic approximation Monte Carlo

We now develop an annealing evolutionary stochastic approximation Monte Carlo algorithm for computing $\hat{\boldsymbol{\theta}}_I = (\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)$ and $\hat{\boldsymbol{\theta}}_E = (\hat{\mathrm{q}}_E, \hat{\boldsymbol{\beta}}_E)$. Quite recently, the stochastic approximation Monte Carlo algorithm (Liang et al. 2010) has been proposed in the

literature as a general simulation technique, which possesses a nice feature in that the moves are self-adjustable and thus not likely to get trapped by local energy minima. The annealing evolutionary SAMC algorithm (Liang et al. 2010) represents a further improvement of stochastic approximation Monte Carlo for optimization problems by incorporating some features of simulated annealing and the genetic algorithm into its search process.

Like the genetic algorithm, annealing evolutionary stochastic approximation Monte Carlo works on a population of samples. Let $\boldsymbol{\theta}^l = (\boldsymbol{\theta}_{(1)}, \ldots, \boldsymbol{\theta}_{(l)})$ denote the population, where l is the population size, and $\boldsymbol{\theta}_{(k)} = (\theta_{k1}, \ldots, \theta_{kp_\theta})$ is a p_θ-dimensional vector called an individual or chromosome in terms of genetic algorithms. Thus, the minimum of the objective function $\mathcal{Q}_n(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$, can be obtained by minimizing the function $U(\boldsymbol{\theta}^l) = \sum_{k=1}^{l} \mathcal{Q}_n(\boldsymbol{\theta}_{(k)})$. An unnormalized Boltzmann density can be defined for the population as follows,

$$\psi(\boldsymbol{\theta}^l) = \exp\left\{-U(\boldsymbol{\theta}^l)/\tau\right\}, \quad \boldsymbol{\theta}^l \in \Theta^l, \tag{A.26}$$

where $\tau = 1$ is called the temperature, and $\Theta^l = \Theta \times \cdots \times \Theta$ is a product sample space. The sample space can be partitioned according to the function $U(\boldsymbol{\theta}^l)$ into b subregions: $\mathbb{E}_1 = \{\boldsymbol{\theta}^l : U(\boldsymbol{\theta}^l) \leq \delta_1\}$, $\mathbb{E}_2 = \{\boldsymbol{\theta}^l : \delta_1 < U(\boldsymbol{\theta}^l) \leq \delta_2\}$, $\cdots$, $\mathbb{E}_{b-1} = \{\boldsymbol{\theta}^l : \delta_{b-2} < U(\boldsymbol{\theta}^l) \leq \delta_{b-1}\}$, and $\mathbb{E}_b = \{\boldsymbol{\theta}^l : U(\boldsymbol{\theta}^l) > \delta_{b-1}\}$, where $\delta_1 < \delta_2 < \ldots < \delta_{b-1}$ are $b-1$ known real numbers. We note that here the sample space is not necessarily partitioned according to the function $U(\boldsymbol{\theta}^l)$, for example, the function $\lambda(\boldsymbol{\theta}^l) = \min\{\mathcal{Q}_n(\boldsymbol{\theta}_{(1)}), \ldots, \mathcal{Q}_n(\boldsymbol{\theta}_{(l)})\}$ also works.

Let $\varpi(\delta)$ denote the index of the subregion that a sample with energy $U(\boldsymbol{\theta}^l)$ belongs to. For example, if $\boldsymbol{\theta}^l \in E_j$, then $\varpi(U(\boldsymbol{\theta}^l)) = j$. Let $\mathcal{B}^{(t)}$ denote the sample space at iteration t. The algorithm initiates its search in the entire sample space $\mathcal{B}_0 = \bigcup_{j=1}^{b} E_j$,

and then iteratively searches in the set

$$\mathcal{B}_t = \bigcup_{j=1}^{\varpi(U_{\min}^{(t)}+\aleph)} \mathbb{E}_j, \quad t = 1, 2, \ldots, \tag{A.27}$$

where $U_{\min}^{(t)}$ is the best function value obtained until iteration t, and $\aleph > 0$ is a user specified parameter which determines the broadness of the sample space at each iteration. Note that in this method, the sample space shrinks iteration by iteration. To ensure the convergence of the algorithm to the set of global minima, the moves at each itertaion are required to admit the following distribution as the invariant distribution,

$$f_{w^{(t)}}(\boldsymbol{\theta}^l) \propto \sum_{j=1}^{\varpi(U_{\min}^{(t)}+\aleph)} \frac{\psi(\boldsymbol{\theta}^l)}{e^{w_j^{(t)}}} I(\boldsymbol{\theta}^l \in \mathbb{E}_j), \tag{A.28}$$

where $w_j^{(t)}$ are the working parameters which will be updated from itertaion to iteration as described in the algorithm below.

The annealing evolutionary stochastic approximation Monte Carlo includes five types of moves, the MH-Gibbs mutation, K-point mutation, K-point crossover, snooker crossover, and linear crossover operators. See Liang et al. (2010) for the details of the moves. Let $\rho_1, \ldots, \rho_5$, $0 < \rho_k < 1$ and $\sum_{k=1}^5 \rho_k = 1$, denote the respective working probabilities of the five types of moves. The algorithm can be summarized as follows.

The algorithm:

(a) (Initialization) Partition the sample space $\mathcal{B}^l$ into b disjoint subregions $\mathbb{E}_1, \ldots, \mathbb{E}_b$; choose the threshold value $\aleph$ and the working probabilities $\rho_1, \ldots, \rho_5$; initialize a population $\boldsymbol{\theta}^{l(0)}$ at random; and set $w^{(0)} = (w_1^{(0)}, \ldots, w_b^{(0)}) = (0, 0, \ldots, 0)$, $\mathcal{B}_0^l = \bigcup_{j=1}^b \mathbb{E}_j$, $U_{\min}^{(0)} = U(\boldsymbol{\theta}^{l(0)})$ and $t = 0$. Let $\mathcal{W}$ be a compact set in R^b.

(b) (Sampling) Update the current population $\boldsymbol{\theta}^{l(t)}$ using the MH-Gibbs mutation,

K-point mutation, K-point crossover, snooker crossover, and linear crossover operators according to the respective working probabilities.

(c) (Working weight updating) Update the working weight $w^{(t)}$ by setting

$$w_j^* = w_j^{(t)} + \gamma_{t+1} H_j(w^{(t)}, \boldsymbol{\theta}^{l(t+1)}), \quad j = 1, \ldots, \varpi(U_{\min}^{(t)} + \aleph),$$

where $H_j(w^{(t)}, \boldsymbol{\theta}^{l(t+1)}) = I(\boldsymbol{\theta}^{l(t+1)} \in \mathbb{E}_j)$ for the crossover operators, $H_j(w^{(t)}, \boldsymbol{\theta}^{l(t+1)}) = \sum_{k=1}^{l} I(\boldsymbol{\theta}^{l(t+1,k)} \in \mathbb{E}_j)/l$ for the mutation operators, and γ_{t+1} is called the gain factor. If $w^* \in \mathcal{W}$, set $w^{(t+1)} = w^*$; otherwise, set $w^{(t+1)} = w^* + c^*$, where $c^* = (c^*, \ldots, c^*)$ and c^* is chosen such that $w^* + c^* \in \mathcal{W}$.

(d) (Termination Checking) Check the termination condition, e.g., whether a fixed number of iterations has been reached. Otherwise, set $t \to t + 1$ and go to step (b).

In this article, we follow Liang et al. (2010) to set $\rho_1 = \rho_2 = 0.05$, $\rho_3 = \rho_4 = \rho_5 = 0.3$, and the gain factor sequence

$$\gamma_t = \frac{t_0}{\max(t_0, t)}, \quad t = 0, 1, 2, \ldots, \tag{A.29}$$

with $t_0 = 5000$. In general, a large value of t_0 will allow the sampler to reach all the subregions very quickly even for a large system. As shown in Liang et al. (2010), it can converge weakly toward a neighboring set of global minima of $U(\boldsymbol{\theta}^l)$ in the space of energy. More precisely, the sample $\boldsymbol{\theta}^{l(t)}$ converges in distribution to a random population with the density function

$$f_w(\boldsymbol{\theta}^l) \propto \sum_{j=1}^{\varpi(U_{\min}+\aleph)} \frac{\psi(\boldsymbol{\theta}^l)}{\int_{\mathbb{E}_j} \psi(\boldsymbol{\theta}^l) d\boldsymbol{\theta}^l} I(x \in \mathbb{E}_j), \tag{A.30}$$

where $U_{\min}$ is the global minimum value of $U(\boldsymbol{\theta})$,

Regarding the setting of other parameters, we have the following suggestions. In the algorithm, the moves are reduced to the Metropolis-Hastings moves Metropolis et al. (1953), Hastings (1970) within the same subregions. Hence, the sample space should be partitioned such that the MH moves within the same subregion have a reasonable acceptance rate. In this article, we set $\delta_{j+1} - \delta_j \equiv 0.2$ for $j = 1, \ldots, b-1$.

The crossover operator has been modified to serve as a proposal for the moves, and it is no longer as critical as to the genetic algorithm. Hence, the population size l is usually set to a moderate number, ranging from 10 to 100. Since $\aleph$ determines the size of the neighboring set toward which the method converges, $\aleph$ should be chosen carefully for efficiency of the algorithm. If $\aleph$ is too small, it may take a long time for the algorithm to locate the global minima. In this case, the sample space may contain a lot of separated regions, and most of the proposed transitions will be rejected if the proposal distribution is not spread out enough. If $\aleph$ is too large, it may also take a long time for the algorithm to locate the global energy minimum due to the broadness of the sample space. In practice, the values of l and $\aleph$ can be determined through a trial and error process based on the diagnosis for the convergence of the algorithm. If it fails to converge, the parameters should be tuned to larger values. The convergence of the method can be diagnosed by examining the difference of the patterns of the working weights obtained in multiple runs. In this article, we set $l = 50$ and $\aleph = 50$.

REFERENCES

Afsari, B. (2011), "Riemannian L^p center of mass: existence, uniqueness, and convexity," *Proc. Amer. Math. Soc.*, 139, 655–673.

Amit, O., Mannino, F., Stone, A., Bushnell, W., Denne, J., Helterbrand, J., and Burger, H. (2011), "Blinded independent central review of progression in cancer clinical trials: Results from a meta-analysis," *European Journal of Cancer*, 47, 1772 – 1778.

Bandulasiri, A., Bhattacharya, R. N., and Patrangenaru, V. (2009a), "Nonparametric inference for extrinsic means on size-and-(reflection)-shape manifolds with applications in medical imaging," *J. Multivar. Anal.*, 100, 1867–1882.

Bandulasiri, A., Patrangenaru, V., Su, J., and Zhang, J. (2009b), "Applications of Nonparametric Statistics on Reflection Shape Manifolds and Reflection Size-and-Shape Manifolds," in *Proceedings of the Joint Statistical Meetings 2008, Denver, CO.*, pp. 2769 – 2776.

Barnhart, H. X., Haber, M., and Song, J. (2002), "Overall Concordance Correlation Coefficient for Evaluating Agreement Among Multiple Observers," *Biometrics*, 58, 1020–1027.

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

Beg, M. F. (2004), "Variational and computational methods for flows of diffeomorphisms in image matching and growth in computational anatomy," Ph.D. thesis, Johns Hopkins University.

Beran, R. J. (1968), "Testing for Uniformity on a Compact Homogeneous Space," *Journal of Applied Probability*, 5, pp. 177–195.

Bhattacharya, A. (2008), "Statistical analysis on manifolds: a nonparametric approach for inference on shape spaces," *Sankhyā*, 70, 223–266.

Bhattacharya, A. and Bhattacharya, R. (2008), "Statistics on Riemannian Manifolds: Asymptotic Distribution and Curvature," *Proceedings of the American Mathematical Society*, 136, pp. 2959–2967.

Bhattacharya, A. and Dunson, D. (2012), "Nonparametric Bayes classification and hypothesis testing on manifolds," *Journal of Multivariate Analysis*, 111, 1 – 19.

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

Bhattacharya, R. and Patrangenaru, V. (2002), "Nonparametic estimation of location and dispersion on Riemannian manifolds," *Journal of Statistical Planning and Inference*, 108, 23 – 35, c.R. Rao 80th Birthday Felicitation Volume, Part II.

— (2003), "Large sample theory of intrinsic and extrinsic sample means on manifolds. I," *Ann. Statist.*, 31, 1–29.

— (2005), "Large sample theory of intrinsic and extrinsic sample means on manifolds. II," *Ann. Statist.*, 33, 1225–1259.

Bookstein, F. (1978), *The Measurement of Biological Shape and Shape Change.*, Lecture Notes in Biomathematics (24), Springer, Berlin.

Boothby, W. M. (1986), *An introduction to differentiable manifolds and Riemannian geometry*, vol. 120 of *Pure and Applied Mathematics*, Orlando, FL: Academic Press Inc., 2nd ed.

Brown, B. W. J., Hollander, M., and Korwar, R. M. (1974), "Nonparametric tests of independence of censored data, with applications to heart transplant studies," *Reliability and Biometry, Statistical Analysis of Lifelength, F. Proschan and RJ Serfling*, 327–354.

Chang, T. (1986), "Spherical Regression," *The Annals of Statistics*, 14, 907–924.

— (1988), "Estimating the Relative Rotation of Two Tectonic Plates from Boundary Crossings," *Journal of the American Statistical Association*, 83, pp. 1178–1183.

— (1989), "Spherical regression with errors in variables," *Ann. Statist.*, 17, 293–306.

Clatz, O., Sermesant, M., Bondiau, P.-Y., Delingette, H., Warfield, S. K., Malandain, G., and Ayache, N. (2005), "Realistic simulation of the 3d growth of brain tumors in mr images coupling diffusion with biomechanical deformation," *IEEE TRANSACTIONS ON MEDICAL IMAGING*, 1334–1346.

Cohen, J. (1960), "A coefficient of agreement for nominal scales," *Educational Psychological Measurement*, 20, 37–46.

— (1968), "Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit," *Psychological Bulletin*, 70, 213–220.

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

Davis, B., Fletcher, P., Bullitt, E., and Joshi, S. (2010), "Population Shape Regression from Random Design Data," *International Journal of Computer Vision*, 90, 255–266.

Di Paola, M., Di Iulio, F., Cherubini, A., Blundo, C., Casini, A. R., Sancesario, G., Passafiume, D., Caltagirone, C., and Spalletta, G. (2010), "When, where and how corpus callosum changes in MCI and AD: a multimodal MRI study," *Neurology*, 74, 1136–1142.

Diggle, P. J., Heagerty, P., Liang, K. Y., and Zeger, S. L. (2002), *Analysis of Longitudinal Data*, Oxford University Press.

do Carmo, M. P. (1992), *Riemannian geometry*, Mathematics: Theory & Applications, Boston, MA: Birkhäuser Boston Inc., translated from the second Portuguese edition by Francis Flaherty.

Dodd, L., Korn, E., Freidlin, B., Gu, W., Abrams, J., Bushnell, W., Canetta, R., Doroshow, J., Gray, R., and Sridhara, R. (2013), "An addit strategy for time-to-event outcomes measured with error: Application to five randomized controlled trials in oncology," *Clinical Trials*, 10, 754–760.

Dodd, L., Korn, E., Freidlin, B., Robert, G., and Bhattacharya, S. (2011), "An addit strategy for progression-free survival," *Biometrics*, 67, 1092–1099.

Downs, T. D. (2003), "Spherical regression," *Biometrika*, 90, 655–668.

Dryden, I. L., Koloydenko, A., and Zhou, D. (2009), "Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging," *Ann. Appl. Stat.*, 3, 1102–1123.

Dryden, I. L. and Mardia, K. V. (1998), *Statistical shape analysis*, Chichester: John Wiley & Sons Ltd.

Durante, F. and Sempi, C. (2010), "Copula Theory: An Introduction," in *Copula Theory and Its Applications*, eds. Jaworski, P., Durante, F., HÃd'rdle, W. K., and Rychlik, T., Springer Berlin Heidelberg, Lecture Notes in Statistics, pp. 3–31.

Durrleman, S., Pennec, X., Trouvé, A., Gerig, G., and Ayache, N. (2009), "Spatiotemporal Atlas Estimation for Developmental Delay Detection in Longitudinal Datasets," in *Proceedings of the 12th International Conference on Medical Image Computing and Computer-Assisted Intervention: Part I*, Berlin, Heidelberg: Springer-Verlag, MICCAI '09, pp. 297–304.

Ellingson, L., Groisser, D., Osborne, D., Patrangenaru, V., and Schwartzman, A. (2012), "Nonparametric Bootstrap of Sample Means of Positive-Definite Matrices with an Application to Diffusion-Tensor-Imagining Data Analysis," Tech. rep., Florida State University, Tallahassee, Florida.

Fahrmeir, L. and Tutz, G. (2001), *Multivariate Statistical Modelling Based on Generalized Linear Models*, Springer Series in Statistics, Springer-Verlag Inc.

FDA (2007), "Guidance for Industry: Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics," .

Feinstein, A. R. and Cicchetti, D. V. (1990), "High agreement but low Kappa: I. the problems of two paradoxes," *Journal of Clinical Epidemiology*, 43, 543 – 549.

Fitzmaurice, G., Davidian, M., Molenberghs, G., and Verbeke, G. (2008), *Longitudinal Data Analysis*, Boca Raton, Florida: Chapman & Hall/CR.

Fletcher, P., Joshi, S., Lu, C., and Pizer, S. (2004a), "Principal Geodesic Analysis for the Study of Nonlinear Statistics of Shape," *IEEE Transactions on Medical Imaging*, 23, 995–1005.

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

Fletcher, P. T. (2013), "Geodesic Regression and the Theory of Least Squares on Riemannian Manifolds," *International Journal of Computer Vision*, 105, 171–185.

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

Fréchet, M. (1948), "Les éléments aléatoires de nature quelconque dans un espace distancié," *Ann. Inst. H.Poincaré, Sect. B, Prob. et Stat.*, 10, 235–310.

Gallier, J. and Xu, D. (2002), "Computing Exponentials of Skew-Symmetric Matrices and Logarithms of Orthogonal Matrices," *Int. J. Robot. Autom.*, 17, 1–11.

Grenander, U., Miller, M., and Srivastava, A. (1998), "Hilbert-Schmidt lower bounds for estimators on matrix lie groups for ATR," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20, 790–802.

Grenander, U. and Miller, M. I. (1998), "Computational anatomy: an emerging discipline," *Quart. Appl. Math.*, 56, 617–694, current and future challenges in the applications of mathematics (Providence, RI, 1997).

Guo, Y., Li, R., Peng, L., and Manatunga, A. K. (2013), "New agreement measures based on survival processes," *Biometrics*, 1–9.

Guo, Y. and Manatunga, A. K. (2009), "Measuring Agreement of Multivariate Discrete Survival Times Using a Modified Weighted Kappa Coefficient," *Biometrics*, 65, 125–134.

Haff, L. R., Kim, P. T., Koo, J. Y., and Richards, D. S. P. (2011), "Minimax Estimation for Mixtures of Wishart Distributions," *Ann. Statist.*, 39, 3417–3440.

Hansen, L. P. (1982), "Large sample properties of generalized method of moments estimators," *Econometrica*, 50, 1029–1054.

Hastings, W. K. (1970), "Monte Carlo sampling methods using Markov chains and their application," *Biometrika*, 57, 97–100.

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*, vol. 80 of *Pure and Applied Mathematics*, New York: Academic Press Inc. [Harcourt Brace Jovanovich Publishers].

Hendriks, H. and Landsman, Z. (1998), "Mean Location and Sample Mean Location on Manifolds: Asymptotics, Tests, Confidence Regions," *Journal of Multivariate Analysis*, 67, 227–243.

Hougaard, P. (2000), *Analysis of multivariate survival data*, Statistics for Biology and Health, New York: Springer-Verlag.

Huckemann, S. (2011), "Inference on 3D Procrustes means: tree bole growth, rank deficient diffusion tensors and perturbation models," *Scand. J. Stat.*, 38, 424–446.

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. Patt. Anal. Mach. Intell.*, 32, 593–603.

J. M. Williamson, J. M., Manatunga, A. K., and Lipsitz, S. R. (2000), "Modeling kappa for measuring dependent categorical agreement data," *Biostatistics*, 1, 191–202.

Joe, H. (1997), *Multivariate models and dependence concepts*, vol. 73 of *Monographs on Statistics and Applied Probability*, London: Chapman & Hall.

Jupp, P. E. and Kent, J. T. (1987), "Fitting Smooth Paths to Spherical Data," *Applied Statistics*, 36, 34–46.

Jupp, P. E. and Mardia, K. V. (1989), "A Unified View of the Theory of Directional Statistics, 1975-1988," *International Statistical Review / Revue Internationale de Statistique*, 57, 261–294.

Karcher, H. (1977), "Riemannian center of mass and mollifier smoothing," *Communications on Pure and Applied Mathematics*, 30, 509–541.

Kendall, D. G. (1977), "The Diffusion of Shape," *Advances in Applied Probability*, 9, pp. 428–430.

— (1984a), "Shape manifolds, Procrustean metrics, and complex projective spaces," *Bull. London Math. Soc.*, 16, 81–121.

— (1984b), "Shape Manifolds, Procrustean Metrics and Complex Projective Spaces," *Bull. Lond. Math. Soc.*, 16, 81–121.

Kendall, D. G., Barden, D., Carne, T. K., and Le, H. (1999), *Shape and Shape Theory*, Chichester: John Wiley & Sons Ltd.

Kent, J. T. (1982), "The Fisher-Bingham distribution on the sphere," *J. Roy. Statist. Soc. Ser. B*, 44, 71–80.

Kim, P. T. (1998), "Deconvolution density estimation on SO(N)," *Ann. Statist.*, 26, 1083–1102.

— (2000), "On the Characteristic Function of the Matrix von MisesâĂŤFisher Distribution with Application to SO(N)âĂŤDeconvolution," in *High Dimensional Probability II*, eds. GinÃĺ, E., Mason, D., and Wellner, J., BirkhÃďuser Boston, vol. 47 of *Progress in Probability*, pp. 477–492.

Kobayashi, S. and Nomizu, K. (1996), *Foundations of differential geometry. Vol. II*, Wiley Classics Library, New York: John Wiley & Sons Inc., reprint of the 1969 original, A Wiley-Interscience Publication.

Koo, J.-Y. and Kim, P. (2008a), "Asymptotic Minimax Bounds for Stochastic Deconvolution Over Groups," *Information Theory, IEEE Transactions on*, 54, 289 –298.

Koo, J.-Y. and Kim, P. T. (2008b), "Sharp adaptation for spherical inverse problems with applications to medical imaging," *J. Multivar. Anal.*, 99, 165–190.

Korsholm, L. (1999), *The GMM estimator versus the semiparametric efficient score estimator under conditional moment restrictions*, Working paper series, University of Aarhus, Department of Economics, Building 350.

Kraemer, H. C. (1980), "Extension of the Kappa Coefficient," *Biometrics*, 36, pp. 207–216.

Kraemer, H. C., V. S. Periyakoil, V. S., and Noda, A. (2002), "Kappa coefficients in medical research," *Statistics in Medicine*, 21, 2109–2129.

Lang, S. (1999), *Fundamentals of differential geometry*, vol. 191 of *Graduate Texts in Mathematics*, New York: Springer-Verlag.

Le, H. (1998), "On the consistency of procrustean mean shapes," *Adv. in Appl. Probab.*, 30, 53–63.

Le, H. and Kume, A. (2000), "The Fréchet mean shape and the shape of the means," *Adv. in Appl. Probab.*, 32, 101–113.

Lebel, C., Caverhill-Godkewitsch, S., and Beaulieu, C. (2010), "Age-related regional variations of the corpus callosum identified by diffusion tensor tractography," *NeuroImage*, 52, 20 – 31.

Lesosky, M., Kim, P. T., and Kribs, D. W. (2008), "Regularized deconvolution on the 2D-Euclidean motion group," *Inverse Problems*, 24, 055017, 15.

Liang, F., Liu, C., and Carroll, R. (2010), *Advanced Markov chain Monte Carlo: Learning from Past Samples.*, Wiley.

Lin, L.-K. (1989), "A concordance correlation coefficient to evaluate reproducibility," *Biometrics*, 45, 255âĂŞ268.

— (2000), "A note on the concordance correlation coefficient," *Biometrics*, 56, 324 âĂŞ325.

Liu, X., Du, Y., Teresi, J., and Hasin, D. S. (2005), "Concordance correlation in the measurements of time to event," *Statistics in Medicine*, 24, 1409–1420.

Lord, F. and Novick, M. (1968), *Statistical theories of mental test scores*, Addison-Wesley series in behavioral science, Addison-Wesley Pub. Co.

Lorenzi, M., Ayache, N., Frisoni, G. B., and Pennec, X. (2011), "Mapping the Effects of Ab142 Levels on the Longitudinal Changes in Healthy Aging: Hierarchical Modeling Based on Stationary Velocity Fields," in *MICCAI 2011*, eds. Peters, T., Fichtinger, G., and Martel, A., Springer, Heidelberg, LNCS.

Louis, T. A. (1982), "Finding the Observed Information Matrix when Using the EM Algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, 44, pp. 226–233.

Machado, L. and Leite, F. S. (2006), "Fitting smooth paths on Riemannian manifolds," *Int. J. Appl. Math. Stat.*, 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. (1972), *Statistics of directional data*, London: Academic Press, probability and Mathematical Statistics, No. 13.

— (1975), "Statistics of directional data," *J. Roy. Statist. Soc. Ser. B*, 37, 349–393.

Mardia, K. V. and Jupp, P. E. (2000), *Directional statistics*, Chichester: John Wiley & Sons Ltd.

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

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953), "Equations of state calculations by fast computing machine," *Journal of Chemical Physics*, 21, 1087–1091.

Miller, M. I. (2004), "Computational anatomy: shape, growth, and atrophy comparison via diffeomorphisms," *NeuroImage*, 23, Supplement 1, S19 – S33, mathematics in Brain Imaging.

Moakher, M. (2002), "Means and averaging in the group of rotations," *SIAM J. Matrix Anal. Appl.*, 24, 1–16 (electronic).

Muralidharan, P. and Fletcher, P. (2012), "Sasaki metrics for analysis of longitudinal data on manifolds," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 1027–1034.

Newey, W. K. (1993), "Efficient estimation of models with conditional moment restrictions," in *Econometrics*, Amsterdam: North-Holland, vol. 11 of *Handbook of Statist.*, pp. 419–454.

Niethammer, M., Huang, Y., and Vialard, F.-X. (2011), "Geodesic Regression for Image Time-Series," in *Medical Image Computing and Computer-Assisted Intervention âĂŞ MICCAI 2011*, eds. Fichtinger, G., Martel, A., and Peters, T., Springer Berlin Heidelberg, vol. 6892 of *Lecture Notes in Computer Science*, pp. 655–662.

Osborne, D., Patrangenaru, V., Ellingson, L., Groisser, D., and Schwartzman, A. (2013), "Nonparametric two-sample tests on homogeneous Riemannian manifolds, Cholesky decompositions and Diffusion Tensor Image analysis," *Journal of Multivariate Analysis*, 119, 163 – 175.

Osborne, D. E. (2012), "Nonparametric data analysis on manifolds with applications in medical imaging," Thesis, Florida State University, Tallahassee, Florida.

Osborne, D. E. and Patrangenaru, V. (2011), "Nonparametric two-sample tests on homogeneous Riemannian manifolds, Cholesky decompositions and dyslexia detection from diffusion tensor imaging outputs," Tech. rep., Florida State University, Tallahassee, Florida.

Patrangenaru, V. (1998), "Asymptotic Statistics on Manifolds," Dissertation, Indiana University.

— (2001), "New large sample and bootstrap methods on shape spaces in high level analysis of natural images," *Comm. Statist. Theory Methods*, 30, 1675–1693, statistics: reflections on the past and visions for the future (San Antonio, TX, 2000).

Patrangenaru, V. and Mardia, K. V. (2003), "Affine Shape Analysis and Image Analysis," in *Proceedings of the Leeds Annual Statistics Research Workshop 2003.*, pp. 57 – 62.

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," *Nature Reviews Neuroscience*, 8, 287–299.

Pennec, X. (2006), "Intrinsic statistics on Riemannian manifolds: basic tools for geometric measurements," *J. Math. Imaging Vision*, 25, 127–154.

Presnell, B., Morrison, S. P., and Littell, R. C. (1998), "Projected multivariate linear models for directional data," *J. Amer. Statist. Assoc.*, 93, 1068–1077.

Preti, M. G., Baglio, F., Laganà, M. M., Griffanti, L., Nemni, R., Clerici, M., Bozzali, M., and Baselli, G. (2012), "Assessing Corpus Callosum Changes in Alzheimer's Disease: Comparison between Tract-Based Spatial Statistics and Atlas-Based Tractography," *PLoS ONE*, 7.

Qiu, A., Albert, M., Younes, L., and Miller, M. I. (2009), "Time sequence diffeomorphic metric mapping and parallel transport track time-dependent shape changes," *NeuroImage*, 45, S51 – S60, mathematics in Brain Imaging.

Rosenthal, M., Wu, W., Klassen, E., and Srivastava, A. (2014), "Spherical Regression Models Using Projective Linear Transformations," *JASA*, 17.

Samir, C., Absil, P.-A., Srivastava, A., and Klassen, E. (2012), "A Gradient-Descent Method for Curve Fitting on Riemannian Manifolds," *Foundations of Computational Mathematics*, 12, 49–73.

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

Schwartzman, A., Dougherty, R. F., and Taylor, J. E. (2008a), "False discovery rate analysis of brain diffusion direction maps," *Ann. Appl. Stat.*, 2, 153–175.

Schwartzman, A., Mascarenhas, W. F., and Taylor, J. E. (2008b), "Inference for eigenvalues and eigenvectors of Gaussian symmetric matrices," *Ann. Statist.*, 36, 2886–2919.

Shi, X., Styner, M., J., L., Ibrahim, J. G., Lin, W., and Zhu, H. (2009), "Intrinsic regression models for manifold-value data," *International Conference on Medical Imaging Computing and Computer Assisted Intervention (MICCAI)*, 5762, 192–199.

Shi, X., Zhu, H., Ibrahim, J. G., Liang, F., Liberman, J., and Styner, M. (2012), "Intrinsic regression models for median representation of subcortical structures," *Journal of American Statistical Association*, 107, 12–23.

Spivak, M. (1979), *A Comprehensive Introduction to Differential Geometry. Vol. I*, Wilmington, Del.: Publish or Perish Inc., 2nd ed.

Su, J., Dryden, I., Klassen, E., Le, H., and Srivastava, A. (2012), "Fitting smoothing splines to time-indexed, noisy points on nonlinear manifolds," *Image and Vision Computing*, 30, 428 – 442.

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

Thompson, P. M., Giedd, J. N., Woods, R. P., MacDonald, D., Evans, A. C., and Toga, A. W. (2000), "Growth patterns in the developing brain detected by using continuum mechanical tensor maps," *Nature*, 404, 190–193.

Trouvé, A. and Vialard, F.-X. (2010), "A second-order model for time-dependent data interpolation: Splines on shape spaces," *MICCAI SYIA Workshop*.

van der Vaart, A. W. (1998), *Asymptotic Statistics*, vol. 3, Cambridge: Cambridge University Press.

Watson, G. S. (1983), *Statistics on spheres*, University of Arkansas Lecture Notes in the Mathematical Sciences, 6, New York: John Wiley & Sons Inc., a Wiley-Interscience Publication.

Witelson, S. F. (1985), "The Brain connection: the Corpus Callosum is Larger in Left-Handers," *Science*, 229, 665–668.

— (1989), "Hand and sex differences in isthmus and genu of the human corpus callosum: a postmortem morphological study," *Brain*, 112, 799–835.

Yarman, C. and Yazici, B. (2003), "Radon transform inversion via Wiener filtering over the Euclidean motion group," in *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, vol. 2, pp. II – 811–14 vol.3.

— (2005), "Radon Transform Inversion Based on Harmonic Analysis of the Euclidean Motion Group," in *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, vol. 2, pp. 481 – 484.

Younes, L. (2010), *Shapes and Diffeomorphisms*, Springer.

Yuan, Y., Zhu, H., Lin, W. L., and Marron, J. S. (2012), "Local polynomial regression for symmetric positive definitive matrices." *JRSS, Series B.*, 74, 697–719.

— (2012a), "Local polynomial regression for symmetric positive definite matrices." *JRSS, Series B.*, 74, 697–719.

Yuan, Y., Zhu, H., Styner, M., Gilmore, J., and Marron, J. (2012b), "Varying coefficient model for modeling diffusion tensors along white matter tracts." *Annals of Applied Statistics*, in press.

Zhu, H., Chen, Y., Ibrahim, J. G., Li, Y., Hall, C., and Lin, W. (2009a), "Intrinsic regression models for positive-definite matrices with applications to diffusion tensor imaging," *J. Amer. Statist. Assoc.*, 104, 1203–1212.

Zhu, H., Ibrahim, J. G., and Shi, X. (2009b), "Diagnostic measures for generalized linear models with missing covariates," *Scand. J. Stat.*, 36, 686–712.

Zhu, H., Ibrahim, J. G., Tang, N., and Zhang, H. (2008), "Diagnostic measures for empirical likelihood of general estimating equations," *Biometrika*, 95, 489–507.

Ziezold, H. (1977), "On expected figures and a strong law of large numbers for random elements in quasi-metric spaces," in *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the Eighth European Meeting of Statisticians (Tech. Univ. Prague, Prague, 1974), Vol. A*, Dordrecht: Reidel, pp. 591–602.