

INTRINSIC RIEMANNIAN FUNCTIONAL DATA ANALYSIS FOR SPARSE LONGITUDINAL OBSERVATIONS

BY LINGXUAN SHAO^{1,a}, ZHENHUA LIN^{2,b} AND FANG YAO^{3,c}

For the Alzheimer's Disease Neuroimaging Initiative

¹*School of Mathematical Sciences, Center for Statistical Science, Peking University, shao-14@pku.edu.cn*

²*Department of Statistics and Data Science, National University of Singapore, linz@nus.edu.sg*

³*School of Mathematical Sciences, Center for Statistical Science, Peking University, fyao@math.pku.edu.cn*

A new framework is developed to intrinsically analyze sparsely observed Riemannian functional data. It features four innovative components: a frame-independent covariance function, a smooth vector bundle termed *covariance vector bundle*, a parallel transport and a smooth bundle metric on the covariance vector bundle. The introduced intrinsic covariance function links estimation of covariance structure to smoothing problems that involve raw covariance observations derived from sparsely observed Riemannian functional data, while the covariance vector bundle provides a rigorous mathematical foundation for formulating such smoothing problems. The parallel transport and the bundle metric together make it possible to measure fidelity of fit to the covariance function. They also play a critical role in quantifying the quality of estimators for the covariance function. As an illustration, based on the proposed framework, we develop a local linear smoothing estimator for the covariance function, analyze its theoretical properties and provide numerical demonstration via simulated and real data sets. The intrinsic feature of the framework makes it applicable to not only Euclidean submanifolds but also manifolds without a canonical ambient space.

1. Introduction. Functional data are nowadays commonly encountered in practice and have been extensively studied in the literature; for instance, see the monographs [Ferraty and Vieu \(2006\)](#), [Hsing and Eubank \(2015\)](#), [Kokoszka and Reimherr \(2017\)](#), [Ramsay and Silverman \(2005\)](#), as well as the survey papers [Wang, Chiou and Müller \(2016\)](#) and [Aneiros et al. \(2019\)](#), for a comprehensive treatment on functional data analysis. These classic endeavors study functional data in which functions are real- or vector-valued, and thus are challenged by data of functions that do not take values in a vector space. Such data emerge increasingly often, partially due to the rapid development of modern technologies. For example, in the longitudinal study of diffusion tensors, as the tensor measured at a time point is represented by a 3×3 symmetric positive-definite matrix (SPD), the study results in a collection of SPD-valued functions. The space of SPD matrices is not a vector space, and in particular, the usual Euclidean distance on it suffers from the “swelling effect,” which introduces artificial and undesirable inflation of variability in data analysis ([Arsigny et al. \(2006/07\)](#)). Specialized distance functions ([Dryden, Koloydenko and Zhou \(2009\)](#), [Pennec, Fillard and Ayache \(2006\)](#)) or metrics ([Arsigny et al. \(2006/07\)](#), [Lin \(2019\)](#), [Moakher \(2005\)](#)) are required to alleviate or completely eliminate the swelling effect. These metrics turn the space of SPD matrices of a fixed dimension into a nonlinear Riemannian manifold. Data in the form of Riemannian manifold valued functions are termed Riemannian functional data and modeled

Received April 2021; revised January 2022.

MSC2020 subject classifications. Primary 62R10; secondary 62R30.

Key words and phrases. Diffusion tensor, Fréchet mean, intrinsic covariance function, parallel transport, smoothing, vector bundle.

by Riemannian random processes, which are random processes taking values in Riemannian manifolds (Lin and Yao (2019)).

Since the mean and covariance functions are two of the most fundamental concepts in functional data analysis, as many downstream analyses depend on them, it is of particular importance to generalize them to Riemannian functional data. For the mean function, the generalized counterpart is the well-established Fréchet mean function that is adopted in Dai, Lin and Müller (2021), Dai and Müller (2018), Lin and Yao (2019) and is an extension of Fréchet mean. The concept of Fréchet mean in turn generalizes the usual mean of random vectors to manifold-valued random elements, and has been studied in depth by Afsari (2011), Bhattacharya and Patrangenaru (2003), Bhattacharya and Patrangenaru (2005), Pennec (2019), Schötz (2019). Related to estimation of Fréchet mean function is regression on manifold-valued nonfunctional data that was investigated by Cornea et al. (2017), Fletcher (2013), Hinkle, Fletcher and Joshi (2014), Pelletier (2006), Shi et al. (2009), Steinke, Hein and Schölkopf (2010), and more broadly, on metric-space valued data by Faraway (2014), Hein (2009), Lin and Müller (2021), Petersen and Müller (2019), among others.

The genuine challenge comes from modeling and estimating the covariance structure. To tackle nonlinearity of the Riemannian manifold, a strategy commonly employed in the literature is to transform data from the manifold into tangent spaces via Riemannian logarithmic maps, and then to model the covariance via the transformed data. Specifically, at each time point, the associated observations are transformed into the tangent space at the Fréchet mean at that time point. Although tangent spaces of a manifold are linear spaces, and thus provide the desired vector structure, there is one issue to resolve: *Different tangent spaces are distinct vector spaces, and thus their tangent vectors are incomparable, but the covariance involves random tangent vectors from different tangent spaces.* More specifically, the value of the covariance function at a time pair (s, t) involves observations at both s and t , and in the manifold setting, the observations at these time points are often transformed into tangent vectors of distinct tangent spaces.

The above issue is especially pronounced for sparsely observed Riemannian functional data. A common strategy well established in the Euclidean setting for sparse functional data is to smooth the discrete and noisy raw covariance function (Cai and Yuan (2010), Li and Hsing (2010), Yao, Müller and Wang (2005), Zhang and Wang (2016)). However, there are fundamental difficulties in extending this seemingly simple strategy to the manifold setting. First, as previously mentioned, the covariance function involves tangent vectors from different tangent spaces, so that an appropriate definition of covariance between two incomparable random tangent vectors is in order. Second, for the smoothing strategy to work, the underlying covariance function shall possess certain regularity of smoothness, such as continuity or differentiability. However, it is challenging to define and quantify such regularity for covariance of Riemannian functional data. This problem is unique to sparsely observed data; when data are fully observed or sufficiently dense so that each trajectory can be individually recovered, the sample covariance operator serves as an estimate for the covariance structure (Lin and Yao (2019)), which does not require smoothing.

To overcome the above difficulties, in this paper we develop a novel framework to model and estimate the covariance when Riemannian functional data are sparsely and noisily recorded. The proposed framework features four innovative components:

- First, an intrinsic covariance function is developed to characterize covariance between random tangent vectors from distinct tangent spaces. Such covariance function is invariant to manifold parameterization, frame selection and embedding, and is made possible by considering the covariance of two random tangent vectors in different tangent spaces as a linear operator that maps one tangent space into the other. This covariance function does not require reference to a frame, and thus is fundamentally different from the covariance function of coefficients with respect to a frame in Lin and Yao (2019).

- Second, we construct a novel smooth vector bundle from the manifold, termed *covariance vector bundle*, to provide an appropriate mathematical foundation for intrinsic quantification of the regularity of the proposed covariance function, such as continuity, differentiability and smoothness. For example, it makes statements like “find a *smooth* covariance function that minimizes the mean squared error” sensible. In addition, covariance function estimation amounts to smoothing data located in a smooth vector bundle. Although there is a rich literature on smoothing Riemannian manifold-valued data, the study on smoothing data in a vector bundle is still in its infancy.
- Third, a parallel transport on the covariance vector bundle is developed from the intrinsic geometry of the manifold, which also induces a covariant derivative on the bundle. The covariant derivative allows intrinsic definition of derivatives of a function taking values in the covariance vector bundle. Such derivatives are often needed when one analyzes theoretical properties of a smoothing procedure. The parallel transport also enables one to move the raw observations into a common vector space in which classic smoothing methods may apply.
- Fourth, a smooth bundle metric is constructed and plays an essential role in measuring the fidelity of fit to data during estimation and quantifying the quality of an estimator. It is derived from the intrinsic geometry of the underlying Riemannian manifold and utilizes the Hilbert–Schmidt inner product of linear operators between two potentially different Hilbert spaces. Such inner product, mathematically well established (e.g., Definition 2.3.3 and Proposition B.0.7 by Prévôt and Röckner (2007)), is less seen in statistics; the common one is usually for operators that map a Hilbert space into itself.

The intrinsic covariance function and the covariance vector bundle together pave the way for intrinsically smoothing the observed raw covariance function, while the parallel transport and the bundle metric are critical for developing an estimation procedure for sparsely observed Riemannian functional data. As an illustration, we propose an estimator for the covariance function based on local linear smoothing and establish the pointwise and uniform convergence rates of the estimator under various designs, while emphasize that other smoothing techniques such as spline smoothing are also applicable. Other contributions include extending the invariance principle of Lin and Yao (2019) to the sparse design and connecting holonomy theory to statistics via Lemma 4.1 that might be of independent interest.

Our work is clearly set apart from existing endeavors in the scarce literature that emerge only in recent years. Su et al. (2014) first represented each trajectory by its normalized velocity curve and then transported the velocity vectors into a common tangent space. Zhang, Klassen and Srivastava (2018) specifically considered spherical trajectories and developed a data transformation geared to the spherical geometry, while Dai and Müller (2018), Lin and Yao (2019) and Dubey and Müller (2020) studied trajectories on a more general manifold or metric space. All of these works assume fully observed functions, and thus require no smoothing. Dai, Lin and Müller (2021) proposed to smooth the sparsely observed raw covariance by embedding the manifold into a Euclidean space. This approach of using an embedding, although making adaption of classic smoothing techniques to the manifold setting straightforward, does not readily apply to manifolds without a canonical embedding. Moreover, the results and their interpretations may be tied to the chosen embedding; see Section 6 of the Supplementary Material (Shao, Lin and Yao (2022)). In contrast, our framework does not require an embedding, and thus circumvents these drawbacks, though it needs to overcome drastically elevated technical challenges.

We shall emphasize that, the proposed framework is not to replace explicit parameterization in practice, but to make the statistical outcomes invariant to the parameterization and/or frame adopted in computation. For instance, in Section 5 we demonstrate that the proposed

method produces identical results under different parameterizations. In particular, the intrinsicity featured by our method refers to requiring no embedding, rather than no parameterization. This makes our framework immediately applicable to manifolds without a canonical embedding. We demonstrate this feature via the manifold of SPD matrices endowed with the affine-invariant metric in our simulation studies; see Section 5.

The rest of the paper is organized as follows. In Section 2, we construct the covariance vector bundle, a parallel transport and a smooth metric on the bundle. In addition, we formulate the intrinsic concept of covariance function for Riemannian functional data. An estimator for the covariance function from sparsely observed Riemannian functional data is described in Section 3, and its theoretical properties are given in Section 4. Simulation studies are placed in Section 5, followed by an application to longitudinal diffusion tensors in Section 6. All of the proofs are deferred to an online Supplementary Material (Shao, Lin and Yao (2022)) for space economy.

2. Covariance vector bundle.

2.1. Preliminaries. We briefly review concepts from Riemannian manifolds that are essential for our development at a high level, while relegate all formal definitions to the Section 1 of the Supplementary Material (Shao, Lin and Yao (2022)) and refer readers to the introductory text by Lee (1997) for further exposition.

Let \mathcal{M} be a d -dimensional smooth manifold, roughly speaking, a space that locally resembles \mathbb{R}^d and is endowed with a smooth structure. A smooth structure is formally described by a (maximal) smooth atlas on \mathcal{M} , specifically, a collection of pairs (U_α, ϕ_α) that are indexed by an index set J and satisfy the following conditions:

- Each U_α is an open subset of \mathcal{M} and $\bigcup_{\alpha \in J} U_\alpha = \mathcal{M}$;
- Each ϕ_α is a bijective continuous map between U_α and an open set of \mathbb{R}^d ;
- If $U_\alpha \cap U_\beta \neq \emptyset$, then the transition map $\phi_\alpha \circ \phi_\beta^{-1} : \phi_\beta(U_\alpha \cap U_\beta) \rightarrow \phi_\alpha(U_\alpha \cap U_\beta)$ is smooth, that is, infinitely differentiable; we say ϕ_α and ϕ_β are compatible.

The pair (U_α, ϕ_α) or sometimes ϕ_α itself is called a chart (or coordinate map). Intuitively, ϕ_α assigns a local coordinate to each point in U_α . Two atlases are compatible if their union is again an atlas (satisfying the above conditions). An atlas is maximal if it contains any other atlas compatible with it.

Every point in a d -dimensional manifold is associated with a distinct d -dimensional vector space, called the tangent space at the point. In addition, any chart (U_α, ϕ_α) gives rise to a basis for the tangent space at each point in U_α , and the basis smoothly varies with the point with U_α . More generally, one can assign to each tangent space a basis. Such an assignment is called a frame. Tangent spaces at different points of a manifold are conceptually distinct spaces, so that their elements, called tangent vectors, are incomparable; only tangent vectors from the same tangent space are comparable.

A Riemannian manifold is a smooth manifold equipped with a Riemannian metric $\langle \cdot, \cdot \rangle$, which defines an inner product $\langle \cdot, \cdot \rangle_p$ on the tangent space $T_p\mathcal{M}$ at each point $p \in \mathcal{M}$, with the associated norm denoted by $\|v\|_p = \sqrt{\langle v, v \rangle_p}$ for $v \in T_p\mathcal{M}$. The metric, which smoothly varies with p , induces a distance function $d_{\mathcal{M}}$ on \mathcal{M} and turns the manifold into a metric space. A geodesic in a Riemannian manifold is a constant-speed curve of which every sufficiently small segment is the shortest path connecting the endpoints of the segment. At each point $p \in \mathcal{M}$, there is an exponential map Exp_p that maps tangent vectors at p onto the manifold \mathcal{M} . In particular, for each $v \in T_p\mathcal{M}$, $\gamma_v(t) := \text{Exp}_p(tv)$ defines a geodesic. The inverse of Exp_p , when it exists, is called the Riemannian logarithmic map at p and denoted by Log_p .

In statistical analysis, it is desirable to compare the tangent vectors from different tangent spaces. To this end, one may transport the tangent vectors into a common tangent space in

which tangent vectors can be directly compared by vector subtraction. For a Riemannian manifold, there is a unique (parallel) transport associated with the Riemannian metric and realized by Levi–Civita connection. In this paper, unless otherwise stated, parallel transport is performed along shortest geodesics between two points y and z , denoted by \mathcal{P}_y^z , which moves tangent vectors from the tangent space $T_y\mathcal{M}$ to $T_z\mathcal{M}$ in a smooth way and meanwhile preserves the inner product.

A smooth vector bundle, denoted by $\pi : \mathcal{E} \rightarrow \mathcal{M}$ or simply \mathcal{E} , consists of a base smooth manifold \mathcal{M} , a smooth manifold \mathcal{E} called total space and a smooth bundle projection π , such that for every $p \in \mathcal{M}$, the fiber $\pi^{-1}(p)$ is a k -dimensional real vector space, and there is an open neighborhood $U \subset \mathcal{M}$ of p and a diffeomorphism $\Phi : \pi^{-1}(U) \rightarrow U \times \mathbb{R}^k$ satisfying the property that for all $z \in U$, $(\pi \circ \Phi^{-1})(z, v) = z$ for all $v \in \mathbb{R}^k$ and the map $v \mapsto \Phi^{-1}(z, v)$ is a linear isomorphism between \mathbb{R}^k and $\pi^{-1}(z)$. The map Φ is called a local trivialization. A prominent example of vector bundle is the space composed by the union of all tangent spaces of a manifold, which is called the tangent bundle of the manifold, where the tangent space at each point is a fiber. To identify different fibers, one can introduce a parallel transport \mathcal{P} on a vector bundle along a curve γ on the base manifold. Such parallel transport must satisfy the following axioms: (1) \mathcal{P}_p^p is the identity map on $\pi^{-1}(p)$ for all $p \in \mathcal{M}$, (2) $\mathcal{P}_{\gamma(u)}^{\gamma(t)} \circ \mathcal{P}_{\gamma(s)}^{\gamma(u)} = \mathcal{P}_{\gamma(s)}^{\gamma(t)}$ and (3) the dependence of \mathcal{P} on γ , s and t are smooth. The parallel transport \mathcal{P} introduced previously for a Riemannian manifold is indeed a parallel transport on the tangent bundle. In Section 2.4, we shall construct a new type of vector bundle and a parallel transport on it. If for each fiber in a smooth vector bundle there is an inner product and the inner product smoothly varies from fiber to fiber, then the inner products are collectively referred to as a smooth bundle metric. The aforementioned Riemannian metric is indeed a smooth bundle metric on the tangent bundle.

2.2. Riemannian functional data. Functional data in which each function takes values in a Riemannian manifold are termed Riemannian functional data and modeled by the Riemannian random process (Lin and Yao (2019)). Specifically, let \mathcal{M} be a d -dimensional Riemannian manifold and X a \mathcal{M} -valued random process indexed by a compact domain $\mathcal{T} \subset \mathbb{R}$, that is, $X : \mathcal{T} \times \Omega \rightarrow \mathcal{M}$, where Ω is the sample space of the underlying probability space. In reality, measurements of X are often corrupted by noise. To accommodate this common practice, we assume that the actual observable process is Y , which is indexed by the same domain \mathcal{T} .

The process X is said to be of second order, if for each $t \in \mathcal{T}$, $F(p, t) = \mathbf{E}d_{\mathcal{M}}^2(X(t), p) < \infty$ for some $p \in \mathcal{M}$, and hence for all $p \in \mathcal{M}$ due to the triangle inequality. The minimizer of $F(p, t)$, if it exists, is called the Fréchet mean of $X(t)$ and denoted by $\mu(t)$, that is,

$$(1) \quad \mu(t) := \arg \min_{p \in \mathcal{M}} F(p, t).$$

The concept of the Fréchet mean generalizes the mean from the Euclidean space to the Riemannian manifold and plays an important role in analysis of data residing in a Riemannian manifold. Under fairly general conditions, the Fréchet mean exists and is unique (Afsari (2011), Bhattacharya and Patrangenaru (2003), Sturm (2003)), for instance, when the manifold is of nonpositive sectional curvature (page 146, Lee (1997)) or data are located in a small subspace of the manifold. Formally, we make the following assumption.

ASSUMPTION 2.1. The Fréchet mean functions of X and Y exist and are unique.

As the manifold \mathcal{M} is not a vector space, it is challenging to directly study the processes X and Y . A common strategy is to transform them into tangent spaces, in which the vector

structure can facilitate the analysis, via Riemannian logarithmic maps. This requires an additional assumption to ensure the well-posedness of the Riemannian logarithmic maps. For simplicity, we assume the following sufficient condition, which can be relaxed by a delicate formulation via cut locus.¹

ASSUMPTION 2.2. There exists a geodesically convex² subset $\mathcal{Q} \subset \mathcal{M}$ such that $X(t), Y(t) \in \mathcal{Q}$ for all $t \in \mathcal{T}$.

If the manifold is of nonpositive sectional curvature, \mathcal{Q} can be taken to be \mathcal{M} , and thus the above assumption becomes superfluous. Examples of manifolds of this kind include hyperbolic manifolds, tori and the space of symmetric positive-definite matrices endowed with the affine-invariant metric (Moakher (2005)), Log-Euclidean metric (Arsigny et al. (2006/07)) or Log-Cholesky metric (Lin (2019)). An example \mathcal{Q} for Riemannian manifolds of positive sectional curvature is the hypersphere $\mathbb{S}^k = \{(x_0, \dots, x_k) \in \mathbb{R}^{k+1} : x_0^2 + \dots + x_k^2 = 1\}$ or the positive orthant $\mathcal{Q} = \{(x_0, \dots, x_k) \in \mathbb{S}^k : x_j \geq 0 \text{ for all } j = 0, \dots, k\}$, which has applications in compositional data analysis (Dai and Müller (2018)), where k is a positive integer.

Under Assumptions 2.1 and 2.2, the Riemannian logarithmic maps $\text{Log}_{\mu(t)}\{X(t)\}$ and $\text{Log}_{\mu(t)}\{Y(t)\}$ are well defined. In addition, we can further model the observed process by

$$Y(t) = \text{Exp}_{\mu(t)}(\text{Log}_{\mu(t)}\{X(t)\} + \varepsilon(t)),$$

where $\varepsilon(t) \in T_{\mu(t)}\mathcal{M}$ represents the random noise in the tangent space, is independent of X and satisfies $\mathbf{E}\varepsilon(t) = 0$ and $\text{Exp}_{\mu(t)}\varepsilon(t) \in \mathcal{Q}$. With this setup, the mean functions of X and Y are the same, in analogy to the Euclidean case; see Lemma 2.1 below.

LEMMA 2.1. *If Assumptions 2.1 and 2.2 hold, and \mathcal{M} is complete and simply connected, then $\mathbf{E}\{\text{Log}_{\mu(\cdot)} X(\cdot)\} = 0$. In addition, if $Y(t) = \text{Exp}_{\mu(t)}(\text{Log}_{\mu(t)} X(t) + \varepsilon(t))$, where $\varepsilon(t) \in T_{\mu(t)}\mathcal{M}$ is independent of X and satisfies $\mathbf{E}\varepsilon(\cdot) = 0$, then μ is also the Fréchet mean function of Y .*

Now we are ready to model sparsely observed Riemannian functional data. First, the sample functions X_1, \dots, X_n are considered as i.i.d. realizations of X however accessible are their noisy copies Y_1, \dots, Y_n , rather than X_1, \dots, X_n . To further accommodate the practice that functions are recorded at discrete points, we assume each Y_i is only observed at m_i time points $T_{i,1}, \dots, T_{i,m_i} \in \mathcal{T}$. Specifically, the observed data are $\{(T_{ij}, Y_{ij}) \in \mathcal{T} \times \mathcal{M} : 1 \leq i \leq n, 1 \leq j \leq m_i\}$ with $Y_{ij} = \text{Exp}_{\mu(T_{ij})}(\text{Log}_{\mu(T_{ij})}\{X_i(T_{ij})\} + \varepsilon_{ij})$, where the centered random elements $\varepsilon_{ij} \in T_{\mu(T_{ij})}\mathcal{M}$ are independent of each other and also independent of $\{X_i : 1 \leq i \leq n\}$.

2.3. Covariance function of Riemannian functional data. In addition to the Fréchet mean function, the covariance structure of Riemannian functional data is essential for downstream analysis, for instance, functional principal component analysis. In Lin and Yao (2019), the covariance structure is modeled by the covariance operator of $\text{Log}_{\mu(\cdot)} X(\cdot)$ from the random element perspective (Chapter 7, Hsing and Eubank (2015)) and also by the covariance function of $\text{Log}_{\mu(\cdot)} X(\cdot)$ with respect to a frame.³ The covariance operator is not computationally friendly to sparse data, while the frame-dependent covariance function is not compatible with

¹See Section 1 of the Supplementary Material (Shao, Lin and Yao (2022)) for a precise definition.

²A subset in a Riemannian manifold is geodesically convex if for any two points in the subset there is a unique shortest geodesic that is contained in the subset and connects the points.

³See Section 1 of the Supplementary Material (Shao, Lin and Yao (2022)) for a precise definition.

most smoothing methods; see Section 6 of the Supplementary Material (Shao, Lin and Yao (2022)) for more details.

To develop a frame-independent intrinsic concept of the covariance function from the perspective of stochastic processes, we first revisit the covariance between two centered random vectors U and V . When they are in a common Euclidean space, it is classically defined as the matrix $\mathbf{E}(UV^\top)$. When U and V are in different general inner product spaces \mathbb{U} and \mathbb{V} , a matrix representation of the covariance is definable if one picks an orthonormal basis for each of \mathbb{U} and \mathbb{V} . To eliminate the dependence on the orthonormal bases, we take an operator perspective to treat the covariance C of U and V as a linear operator between \mathbb{U} and \mathbb{V} characterized by

$$\langle Cu, v \rangle_{\mathbb{V}} := \mathbf{E}(\langle U, u \rangle_{\mathbb{U}} \langle V, v \rangle_{\mathbb{V}}), \quad \forall u \in \mathbb{U}, v \in \mathbb{V},$$

where $\langle \cdot, \cdot \rangle_{\mathbb{U}}$ and $\langle \cdot, \cdot \rangle_{\mathbb{V}}$ denote the inner products of \mathbb{U} and \mathbb{V} , respectively. To simplify the notation, we write $C = \mathbf{E}(U \otimes V)$.

Observe that $\text{Log}_{\mu(\cdot)} X(\cdot)$ ($\text{Log}_{\mu} X$ for short) is a random vector field along the curve μ with $\mathbf{E}(\text{Log}_{\mu} X) = 0$ according to Lemma 2.1 (also Theorem 2.1 of Bhattacharya and Patrangenaru (2003)). Given that $\text{Log}_{\mu(s)} X(s) \in T_{\mu(s)}\mathcal{M}$ and $\text{Log}_{\mu(t)} X(t) \in T_{\mu(t)}\mathcal{M}$, and both $T_{\mu(s)}\mathcal{M}$ and $T_{\mu(t)}\mathcal{M}$ are Hilbert spaces, we define the covariance function for X by

$$(2) \quad \mathcal{C}(s, t) := \mathbf{E}\{\text{Log}_{\mu(s)} X(s) \otimes \text{Log}_{\mu(t)} X(t)\}, \quad \text{for } (s, t) \in \mathcal{T}^2.$$

This covariance function is clearly *independent* of any frame or coordinate system. This feature fundamentally and distinctly separates (2) from the frame-dependent covariance function (5) defined in Lin and Yao (2019) for the coordinate of $\text{Log}_{\mu} X$ with respect to a frame along the mean function. Moreover, (2) can be viewed as the intrinsic covariance function of the covariance operator \mathbf{C} proposed in Lin and Yao (2019). Specifically, under some measurability or continuity assumption on X and the condition that $\mathbf{E} \int_{\mathcal{T}} \|\text{Log}_{\mu(t)} X(t)\|_{\mu(t)}^2 dt < \infty$, the process $\text{Log}_{\mu} X$ can be regarded as a random element in the Hilbert space

$$\mathcal{F}(\mu) := \left\{ Z : Z(\cdot) \in T_{\mu(\cdot)}\mathcal{M}, \int_{\mathcal{T}} \langle Z(t), Z(t) \rangle_{\mu(t)}^2 dt < \infty \right\}$$

endowed with the inner product $\langle\langle Z_1, Z_2 \rangle\rangle_{\mu} := \int_{\mathcal{T}} \langle Z_1(t), Z_2(t) \rangle_{\mu(t)} dt$ for $Z_1, Z_2 \in \mathcal{F}(\mu)$. The covariance operator $\mathbf{C} : \mathcal{F}(\mu) \rightarrow \mathcal{F}(\mu)$ for X can be defined by

$$(3) \quad \langle\langle Cu, v \rangle\rangle_{\mu} := \mathbf{E}(\langle\langle \text{Log}_{\mu} X, u \rangle\rangle_{\mu} \langle\langle \text{Log}_{\mu} X, v \rangle\rangle_{\mu}) \quad \text{for } u, v \in \mathcal{F}(\mu).$$

The following theorem, which generalizes Theorem 7.4.3 of Hsing and Eubank (2015) to Riemannian random processes, shows that the proposed covariance function induces the covariance operator \mathbf{C} .

THEOREM 2.1. *Let $\mathcal{C}(\cdot, \cdot)$ and \mathbf{C} be defined in (2) and (3), respectively. Suppose that X is mean-square continuous, that is, $\lim_{k \rightarrow \infty} \mathbf{E}d^2(X(t_k), X(t)) = 0$ for any $t \in \mathcal{T}$ and any sequence $\{t_k\}$ in \mathcal{T} converging to t . Also assume that X is jointly measurable, that is, $X : \mathcal{T} \times \Omega \rightarrow \mathcal{M}$ is measurable with respect to the product σ -field on $\mathcal{T} \times \Omega$, where Ω is the sample space of the underlying probability space. Then under Assumptions 2.1 and 2.2, for all $t \in \mathcal{T}$ and $u \in \mathcal{F}(\mu)$, we have*

$$(\mathbf{C}u)(t) = \int_{\mathcal{T}} \mathcal{C}(s, t)u(s) ds.$$

In light of this result, in the sequel we often use the same notation \mathcal{C} to denote both the covariance operator and the covariance function in (2). The proposed covariance function enables estimating the covariance operator \mathbf{C} through estimating $\mathcal{C}(s, t)$ for each $(s, t) \in \mathcal{T} \times \mathcal{T}$ in a frame-independent fashion. The frame-independent feature is of particular importance to deriving a frame-invariant estimate in the more practical scenario that only discrete and noisy observations are available so that smoothing is desirable; see Section 3 for more detail.

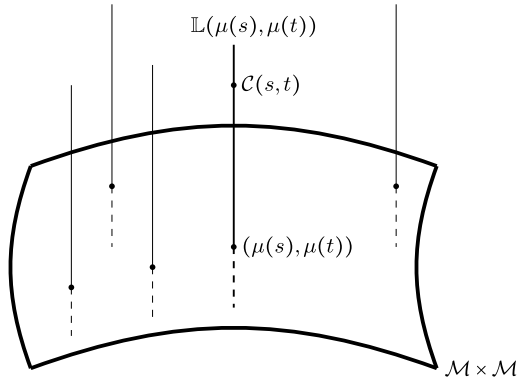


FIG. 1. Illustration of the vector bundle \mathbb{L} . The thick bending parallelogram presents the product manifold $\mathcal{M} \times \mathcal{M}$ and the vertical lines represent fibers. The value of $\mathcal{C}(s, t)$ is located within the fiber $\mathbb{L}(\mu(s), \mu(s))$ at the point $(\mu(s), \mu(t)) \in \mathcal{M} \times \mathcal{M}$.

2.4. *The vector bundle of covariance and parallel transport.* To estimate the covariance function in (2), it seems rather intuitive to perform smoothing over the raw covariance

$$(4) \quad \hat{\mathcal{C}}_{i,jk} := \text{Log}_{\hat{\mu}(T_{ij})} Y_{ij} \otimes \text{Log}_{\hat{\mu}(T_{ik})} Y_{ik} \in \mathbb{L}(\hat{\mu}(T_{ij}), \hat{\mu}(T_{ik})),$$

where $\hat{\mu}$ is an estimate of μ to be detailed in Section 3. The first challenge encountered is that these raw observations $\hat{\mathcal{C}}_{i,jk}$ do not reside in a common vector space. This also gives rise to the second challenge in defining the key concept of *smoothness* of the function \mathcal{C} and its estimate. To circumvent these difficulties, we consider the spaces $\mathbb{L}(p, q)$ consisting of all linear maps from $T_p\mathcal{M}$ to $T_q\mathcal{M}$, and their disjoint union $\mathbb{L} = \bigcup_{(p,q) \in \mathcal{M}^2} \mathbb{L}(p, q)$. Then $\hat{\mathcal{C}}_{i,jk}$ are encompassed by the space \mathbb{L} , and in addition, the covariance function \mathcal{C} is now viewed as an \mathbb{L} -valued function. Although the space \mathbb{L} is not a vector space so that the smoothness is not definable in the classic sense, we observe that \mathbb{L} comes with a canonical smooth structure induced by the manifold \mathcal{M} , and continuity, differentiability and smoothness relevant to statistics can be defined with reference to this smooth structure as follows.

We first observe that \mathbb{L} is a vector bundle on $\mathcal{M} \times \mathcal{M}$, with $\pi : \mathbb{L} \rightarrow \mathcal{M} \times \mathcal{M}$ defined by $\pi(\mathbb{L}(p, q)) = (p, q)$ being the bundle projection and $\mathbb{L}(p, q)$ being the fiber attached to the point $(p, q) \in \mathcal{M} \times \mathcal{M}$; see Figure 1 for a graphical illustration. To define the smoothness structure on \mathbb{L} induced by the manifold \mathcal{M} , let $\{(U_\alpha, \phi_\alpha) : \alpha \in J\}$ for an index set J be an atlas of \mathcal{M} . Recall that each chart (U_α, ϕ_α) gives rise to a smoothly varying basis of $T_p\mathcal{M}$ for each $p \in U_\alpha$. Such basis is denoted by $B_{\alpha,1}(p), \dots, B_{\alpha,d}(p)$. For $(p, q) \in U_\alpha \times U_\beta$, the tensor products $B_{\alpha,j}(p) \otimes B_{\beta,k}(q)$, $j, k = 1, \dots, d$, form a basis for the space $\mathbb{L}(p, q)$. Each element $v \in \mathbb{L}(p, q)$ is then identified with its coefficients v_{jk} with respect to this basis, that is, $v = \sum_{j,k=1}^d v_{jk} B_{\alpha,j}(p) \otimes B_{\beta,k}(q)$. For each $U_\alpha \times U_\beta$, we define the map $\varphi_{\alpha,\beta}(p, q, \sum_{j,k=1}^d v_{jk} B_{\alpha,j}(p) \otimes B_{\beta,k}(q)) = (\phi_\alpha(p), \phi_\beta(q), v_{11}, v_{12}, \dots, v_{dd}) \in \mathbb{R}^{2d+d^2}$, for $(p, q) \in U_\alpha \times U_\beta$. The collection $\{(\pi^{-1}(U_\alpha \times U_\beta), \varphi_{\alpha,\beta}) : (\alpha, \beta) \in J^2\}$ indeed is a smooth atlas that turns \mathbb{L} into a smooth manifold. Moreover, \mathbb{L} is a *smooth* vector bundle with the projection map π and the local trivializations $\Phi_{\alpha,\beta} : \pi^{-1}(U_\alpha \times U_\beta) \rightarrow U_\alpha \times U_\beta \times \mathbb{R}^{d^2}$ defined as $\Phi_{\alpha,\beta}(p, q, \sum_{j,k=1}^d v_{jk} B_{\alpha,j}(p) \otimes B_{\beta,k}(q)) = (p, q, v_{11}, v_{12}, \dots, v_{dd})$.

THEOREM 2.2. *The collection $\{(\pi^{-1}(U_\alpha \times U_\beta), \varphi_{\alpha,\beta}) : (\alpha, \beta) \in J^2\}$ is a smooth atlas on \mathbb{L} . With this atlas, \mathbb{L} is a smooth vector bundle with the smooth projection map π and smooth local trivializations $\Phi_{\alpha,\beta}$. In addition, any compatible atlas of the manifold \mathcal{M} gives rise to the same smooth vector bundle \mathbb{L} .*

With the above smooth structure, the covariance function \mathcal{C} in (2), viewed as an \mathbb{L} -valued function, is said to be κ -times continuously differentiable in (s, t) , if $(\mu(s), \mu(t)) \in U_\alpha \times U_\beta$ implies that $\varphi_{\alpha,\beta}(\mu(s), \mu(t), \mathcal{C}(s, t))$ is κ -times continuously differentiable in (s, t) , where we recall that $\{(\pi^{-1}(U_\alpha \times U_\beta), \varphi_{\alpha,\beta}) : (\alpha, \beta) \in J^2\}$ is a smooth atlas on \mathbb{L} . From this perspective, the constructed vector bundle \mathbb{L} provides a framework to rigorously define the regularity of \mathcal{C} . In this framework, estimating the covariance function \mathcal{C} amounts to smoothing the discrete raw observations $\hat{\mathcal{C}}_{i,jk}$ in the vector bundle \mathbb{L} .

Although the vector bundle \mathbb{L} provides a qualitative framework for defining differentiability or other smoothness regularity, it does not provide a quantitative characterization. Roughly speaking, the smooth vector bundle \mathbb{L} allows one to check whether \mathcal{C} is differentiable or smooth, but not to measure how rapidly \mathcal{C} changes relative to (s, t) . In other words, derivatives that quantify the rate of change of the function \mathcal{C} at a given pair (s, t) and that are consistent across all compatible atlases for \mathbb{L} require an additional structure as follows. We first introduce the parallel transport on the covariance vector bundle \mathbb{L} to identify different fibers and to compare the elements from the fibers. Suppose that $(p_1, q_1), (p_2, q_2) \in \mathcal{M} \times \mathcal{M}$ and $\gamma(t) = (\gamma_p(t), \gamma_q(t))$ is the shortest geodesic connecting (p_1, q_1) to (p_2, q_2) . The parallel transport $\mathcal{P}_{(p_1,q_1)}^{(p_2,q_2)}$ from a fiber $\mathbb{L}(p_1, q_1)$ to another fiber $\mathbb{L}(p_2, q_2)$ is naturally constructed from the parallel transport operators $\mathcal{P}_{p_2}^{p_1}$ and $\mathcal{P}_{q_1}^{q_2}$ on \mathcal{M} by

$$(5) \quad (\mathcal{P}_{(p_1,q_1)}^{(p_2,q_2)} \mathcal{C})(u) := \mathcal{P}_{q_1}^{q_2}(\mathcal{C}(\mathcal{P}_{p_2}^{p_1} u)),$$

where $C \in \mathbb{L}(p_1, q_1)$ and $u \in T_{p_2}\mathcal{M}$. To distinguish between the parallel transport on the manifold and the one on the vector bundle \mathbb{L} , notationally we use the calligraphic symbol \mathcal{P} for the manifold while the script symbol \mathcal{P} for the bundle. The parallel transport \mathcal{P} further determines a covariant derivative⁴ on the bundle.

THEOREM 2.3. *For a tangent vector V of $\mathcal{M} \times \mathcal{M}$ at (p, q) , the map ∇_V defined by*

$$(6) \quad \nabla_V W := \lim_{h \rightarrow 0} \frac{\mathcal{P}_{\gamma(h)}^{\gamma(0)} W(\gamma(h)) - W(\gamma(0))}{h} := \left. \frac{d}{dt} \mathcal{P}_{\gamma(t)}^{\gamma(0)} W(\gamma(t)) \right|_{t=0}$$

for all differentiable section W is a covariant derivative in the direction of V , where γ is a smooth curve⁵ in $\mathcal{M} \times \mathcal{M}$ with initial point $\gamma(0) = (p, q)$ and initial velocity $\gamma'(0) = V$, and a section is any function $W : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{L}$ satisfying $W(p, q) \in \mathbb{L}(p, q)$ for all $(p, q) \in \mathcal{M} \times \mathcal{M}$.

The covariant derivative of a section W can be viewed as the first derivative of the section. It quantifies the rate and direction of change of W at each point in $\mathcal{M} \times \mathcal{M}$. This applies to the covariance function \mathcal{C} since it can be viewed as a section along the surface $\mu \times \mu : \mathcal{T} \times \mathcal{T} \rightarrow \mathcal{M} \times \mathcal{M}$. In addition, the ‘‘partial derivative’’ $\frac{\partial}{\partial s} \mathcal{C}(s, t)|_{s=s_0}$ of $\mathcal{C}(s, t)$ with respect to s at s_0 can be understood as the limit

$$\lim_{h \rightarrow 0} \frac{\mathcal{P}_{\gamma(s_0+h)}^{\gamma(s_0)} \mathcal{C}(s_0 + h, t) - \mathcal{C}(s_0, t)}{h} \in T_{(\mu(s_0), \mu(t))} \mathcal{M}^2$$

with $\gamma(s) = (\mu(s), \mu(t))$. Furthermore, since the derivative $\frac{\partial}{\partial s} \mathcal{C}(s, t)$ is again a section of the vector bundle, one can define the partial derivatives of $\frac{\partial}{\partial s} \mathcal{C}(s, t)$, which can be regarded as the second derivatives of \mathcal{C} . Higher-order derivatives can be defined in a recursive way.

⁴For a definition of the covariant derivative, see Chapter 4 (specifically, page 50) of Lee (1997) or Section 1 of the Supplementary Material (Shao, Lin and Yao (2022)).

⁵It can be shown that the value $\nabla_V W$ depends on V , but not on γ .

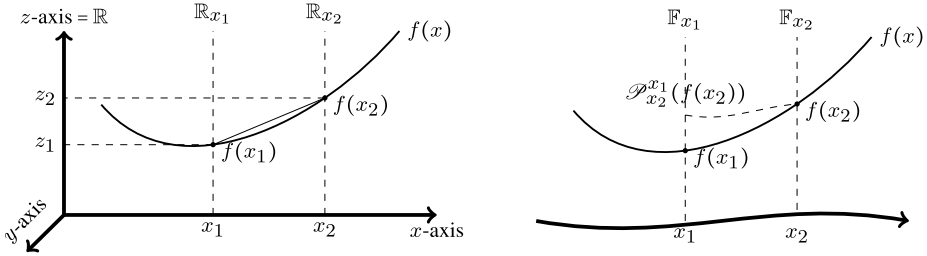


FIG. 2. Illustration of classic differentiation (left) and general covariant derivative (right).

To further illustrate the parallel transport and the induced covariant derivative on the vector bundle $\pi : \mathbb{L} \rightarrow \mathcal{M} \times \mathcal{M}$, consider a simple example in which $\mathcal{M} = \mathbb{R}$ and the bundle \mathbb{L} is then parameterized by $(x, y, z) \in \mathbb{R}^2 \times \mathbb{R}$. Let $g : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{L}$ be a smooth section. For visualization, we fix $y = 0$ and write $f(x) = g(x, 0)$. For the smooth function $f(x)$ shown in Figure 2, the classic definition of the derivative of $f(x)$ at x_1 is

$$\frac{\partial}{\partial x} f(x) \Big|_{x=x_1} := \lim_{x_2 \rightarrow x_1} \frac{f(x_2) - f(x_1)}{x_2 - x_1}.$$

From the perspective of the vector bundle, each point x in the x -axis is attached with a fiber \mathbb{R}_x , which is simply a copy of the z -axis $= \mathbb{R}$. Since $f(x_1) \in \mathbb{R}_{x_1}$ while $f(x_2) \in \mathbb{R}_{x_2}$, the operation $f(x_2) - f(x_1)$ would not be well defined if we did not identify \mathbb{R}_{x_1} with \mathbb{R}_{x_2} . The identification between \mathbb{R}_{x_1} and \mathbb{R}_{x_2} is canonical, and nothing else but parallelly transporting \mathbb{R}_{x_2} to \mathbb{R}_{x_1} . This inspiring observation applies to general manifolds and covariant derivatives. Specifically, the covariant derivative is defined by parallel transporting $f(x_2)$ from the fiber \mathbb{F}_{x_2} into the fiber \mathbb{F}_{x_1} and then performing differentiation therein, that is,

$$\frac{\partial}{\partial x} f(x) \Big|_{x=x_1} := \lim_{x_2 \rightarrow x_1} \frac{\mathcal{P}_{x_2}^{x_1}(f(x_2)) - f(x_1)}{x_2 - x_1} \in \mathbb{F}_{x_1}.$$

Finally, when smoothing the raw covariance function $\hat{C}_{i,jk}$, one needs to quantify the discrepancy between the data and the fit. Such discrepancy is often measured by a distance function or inner product on the data space. Fortunately, the vector bundle \mathbb{L} comes with a natural bundle metric. Specifically, for any $(p, q) \in \mathcal{M} \times \mathcal{M}$, the metric $G_{(p,q)} : \mathbb{L}(p, q) \times \mathbb{L}(p, q) \rightarrow \mathbb{R}$ is defined as the Hilbert–Schmidt inner product, that is,

$$(7) \quad G_{(p,q)}(L_1, L_2) = \sum_{k=1}^d \langle L_1 e_k, L_2 e_k \rangle_q \quad \text{for } L_1, L_2 \in \mathbb{L}(p, q),$$

where e_1, \dots, e_d denotes an orthonormal basis of $T_p \mathcal{M}$. One can show that the definition (7) does not depend on the choice of the orthonormal basis. In fact, G is a smooth bundle metric and the parallel transport (5) defines an isometry between any two fibers, as asserted by the following result.

THEOREM 2.4. *The metric defined in (7) is a vector bundle metric that smoothly varies with $(p, q) \in \mathcal{M} \times \mathcal{M}$ and is preserved by the parallel transport in (5).*

The inner product (7) is defined for linear operators that map a Hilbert space, such as $T_p \mathcal{M}$, to another potentially different Hilbert space, such as $T_q \mathcal{M}$. The inner product of this type, although mathematically well established (e.g., Definition 2.3.3 and Proposition B.0.7 by Prévôt and Röckner (2007)), is less seen in statistics; the commonly used one is usually linear operators that map a Hilbert space into the same Hilbert space. The metric G also

induces a norm, denoted by $\|\cdot\|_{G(p,q)}$ or simply $\|\cdot\|_G$, on each fiber $\mathbb{L}(p, q)$. This norm in turn defines a distance on each fiber $\mathbb{L}(p, q)$ by $\|A - B\|_{G(p,q)}$ for $A, B \in \mathbb{L}(p, q)$, which is an integrated part of the loss function in (8) for estimating the covariance function.

The smooth vector bundle \mathbb{L} together with the covariant derivative (6) and the bundle metric (7), termed *covariance vector bundle* in this paper, paves the way for estimation of the covariance function (2) from sparsely observed Riemannian functional data. The smooth structure and the covariant derivative together provide an intrinsic mechanism to quantify the regularity of \mathcal{C} . For example, it makes meaningful the statement that the second derivatives of $\mathcal{C}(s, t)$ are continuous. In the Euclidean case, statements of this kind are often adopted as assumptions that are fundamental to theoretical analysis of estimators derived from a smoothing method. The developed vector bundle and covariant derivative now enable us to extend such assumptions to the manifold setting, as demonstrated in Section 4.2 where we analyze the theoretical properties of the proposed estimator in Section 3 for the covariance function \mathcal{C} . The parallel transport (5) and the bundle metric (7) allow an intrinsic measure of the discrepancy of objects in the covariance vector bundle. Such measure is critical for finding an estimator for \mathcal{C} and quantifying estimation quality, as illustrated in the following sections.

3. Estimation. The first step is to estimate the mean function, for which we adopt the local linear regression method proposed by Petersen and Müller (2019) and also employed by Dai, Lin and Müller (2021). Define the local weight function

$$\hat{w}(T_{ij}, t, h_\mu) = \frac{1}{\hat{\sigma}_0^2(t)} K_{h_\mu}(T_{ij} - t) \{ \hat{u}_2(t) - \hat{u}_1(t)(T_{ij} - t) \},$$

where $\hat{u}_k(t) = \sum_i \lambda_i \sum_j K_{h_\mu}(T_{ij} - t)(T_{ij} - t)^k$, $\hat{\sigma}_0^2(t) = \hat{u}_0(t)\hat{u}_2(t) - \hat{u}_1^2(t)$ and $K_{h_\mu}(\cdot) = K(\cdot/h_\mu)/h_\mu$ for a kernel function K with bandwidth $h_\mu > 0$. The estimate $\hat{\mu}$ is defined as the minimizer of the weighted function

$$\hat{Q}_n(y, t) = \sum_{1 \leq i \leq n} \lambda_i \sum_{1 \leq j \leq m_i} \hat{w}(T_{ij}, t, h) d_{\mathcal{M}}^2(Y_{ij}, y),$$

that is,

$$\hat{\mu}(t) = \arg \min_{y \in \mathcal{M}} \hat{Q}_n(y, t),$$

where the weights $\{\lambda_i\}_{1 \leq i \leq n}$ are subject-specific and satisfy $\sum_{i=1}^n \lambda_i m_i = 1$. For the Euclidean case $\mathcal{M} = \mathbb{R}$, the objective function $\hat{Q}_n(y, t)$ coincides with the sum of squared error loss used in Zhang and Wang (2016). Two popular choices for λ_i are $\lambda_i = (\sum_{i=1}^n m_i)^{-1}$ (Yao, Müller and Wang (2005)) that assigns equal weight to each observation, and $\lambda_i = (nm_i)^{-1}$ (Li and Hsing (2010)) that assigns equal weight to each subject. Other choices are discussed in Zhang and Wang (2018).

Given the parallel transport introduced in Section 2.4, we are allowed to move the raw covariance $\hat{C}_{i,jk}$ defined in (4) from different fibers into the same fiber and employ the classic local linear smoothing on the transported observations. For $\hat{C}_{i,jk}$ to be well defined, similar to Assumption 2.1, we assume the existence and uniqueness of the empirical mean function $\hat{\mu}$ in the following Assumption 3.1. Such assumption always holds for manifolds of non-positive sectional curvature, or may be replaced by a convexity condition on the distance function when Assumption 2.2 holds. In particular, Assumption 3.1 is satisfied by the manifolds Sym_{AF}^+ and Sym_{LC}^+ of SPD matrices adopted in the simulation studies in Section 5 and data application in Section 6 without additional conditions, as these manifolds have non-positive sectional curvature. The assumption also holds for the sphere \mathbb{S}^2 used in the simulation studies when Assumption 2.2 is fulfilled.

ASSUMPTION 3.1. The estimated mean function $\hat{\mu}(t)$ exists and is unique for each $t \in \mathcal{T}$.

To estimate $\mathcal{C}(s, t)$, the nearby raw observations $\hat{C}_{i,jk}$ are parallelly transported into the fiber $\mathbb{L}(\hat{\mu}(s), \hat{\mu}(t))$, and the estimate $\hat{\mathcal{C}}(s, t)$ is set by $\hat{\mathcal{C}}(s, t) = \hat{\beta}_0$ with

$$(8) \quad (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) = \arg \min_{\beta_0, \beta_1, \beta_2 \in \mathbb{L}(\hat{\mu}(s), \hat{\mu}(t))} \left\{ \sum_i v_i \sum_{j \neq k} \left\| \mathcal{P}_{(\hat{\mu}(T_{ij}), \hat{\mu}(T_{ik}))}^{(\hat{\mu}(s), \hat{\mu}(t))} \hat{C}_{i,jk} - \beta_0 - \beta_1(T_{ij} - s) - \beta_2(T_{ik} - t) \right\|_{G(\hat{\mu}(s), \hat{\mu}(t))}^2 K_{h_C}(s - T_{ij}) K_{h_C}(t - T_{ik}) \right\},$$

where $h_C > 0$ is a bandwidth, $\mathcal{P}_{(\hat{\mu}(T_{ij}), \hat{\mu}(T_{ik}))}^{(\hat{\mu}(s), \hat{\mu}(t))}$ is the parallel transport along minimizing geodesics defined in (5), and the weights $\{v_i\}_{1 \leq i \leq n}$ are subject-specific and satisfy $\sum_{i=1}^n v_i m_i (m_i - 1) = 1$. Similar to the estimation of the mean function, two popular choices for the weights are $v_i = (\sum_{i=1}^n m_i (m_i - 1))^{-1}$ (Yao, Müller and Wang (2005)) that assign equal weight to each observation, and $v_i = (nm_i (m_i - 1))^{-1}$ (Li and Hsing (2010)) that assign equal weight to each subject, while more options are studied in Zhang and Wang (2018).

The objective function in (8) involves only intrinsic concepts and thus is fundamentally different from the objective function in (5) of Dai, Lin and Müller (2021) in which the raw observations $\hat{C}_{i,jk}$ are computed in an ambient space. In addition, the quantities $\hat{C}_{i,jk}$ in (8) are frame-independent, and thus the resulting estimator is invariant to the frame.⁶ This frame-independent feature makes our estimator distinct from the noninvariant estimators discussed Section 6 of the Supplement (Shao, Lin and Yao (2022)).

REMARK 3.1. One might attempt to endow \mathbb{L} with a distance ρ so that the estimation is turned into a regression problem with a metric-space-valued response and the local linear method of Petersen and Müller (2019) can be adopted. Such distance is expected to have the following properties:

- The distance ρ on \mathbb{L} coincides with the fiber metric G for any two points on the same fiber. Specifically, for $L_1, L_2 \in \mathbb{L}(p, q)$, $\rho^2(L_1, L_2) = G_{(p,q)}(L_1 - L_2, L_1 - L_2)$.
- The distance ρ on the zero section $W_0(p, q) = 0 \in \mathbb{L}(p, q)$ coincides with the geodesic distance on $\mathcal{M} \times \mathcal{M}$. Specifically, for $(p_1, q_1), (p_2, q_2) \in \mathcal{M} \times \mathcal{M}$, $\rho(W_0(p_1, q_1), W_0(p_2, q_2)) = d_{\mathcal{M}^2}((p_1, q_1), (p_2, q_2))$.
- When \mathcal{M} is a Euclidean space, especially when $\mathcal{M} = \mathbb{R}$, the estimate derived from Petersen and Müller (2019) under the distance ρ coincides with the classic estimate, that is, the estimate derived from the same method but applied to the observations $\hat{C}_{i,jk} \in \mathbb{R}$ that are treated as real-valued responses.

However, such distance ρ does not exist. On one hand, the positive-definiteness of the distance suggests that $\rho(\hat{C}_{i_1, j_1 k_1}, \hat{C}_{i_2, j_2 k_2}) \neq 0$ as long as $\hat{C}_{i_1, j_1 k_1}, \hat{C}_{i_2, j_2 k_2} \in \mathbb{L}$ reside in different fibers, that is, when $\hat{\mu}(T_{i_1 j_1}) \neq \hat{\mu}(T_{i_2 j_2})$ or $\hat{\mu}(T_{i_1 k_1}) \neq \hat{\mu}(T_{i_2 k_2})$. On the other hand, when $\mathcal{M} = \mathbb{R}$, the quantities $\hat{C}_{i_1, j_1 k_1}$ and $\hat{C}_{i_2, j_2 k_2}$ are treated as real numbers and thus their distance could be zero even when $\hat{\mu}(T_{i_1 j_1}) \neq \hat{\mu}(T_{i_2 j_2})$ or $\hat{\mu}(T_{i_1 k_1}) \neq \hat{\mu}(T_{i_2 k_2})$.

Once an estimate $\hat{\mathcal{C}}$ of the covariance function \mathcal{C} is obtained, according to Theorem 2.1, the intrinsic Riemannian functional principal component proposed in Lin and Yao (2019) can

⁶For the computational purpose, a frame might be adopted, but the resulting estimator is independent of the choice of the frame, since the objective function in (8) does not depend on any frame.

be adopted. Specifically, the eigenvalues $\hat{\lambda}_k$ and eigenfunctions $\hat{\psi}_k$ of \hat{C} can be obtained by eigendecomposition of \hat{C} , for example, via the method described in Section 2.3 of Lin and Yao (2019). For estimation of the scores $\xi_{ik} = \langle \text{Log}_\mu X_i, \psi_k \rangle$ in the intrinsic Karhunen–Loève expansion $\text{Log}_\mu X_i = \sum_{k=1}^\infty \xi_{ik} \psi_k$ proposed in Lin and Yao (2019), numerical approximation to the integral $\text{Log}_\mu X_i, \psi_j \rangle$ is infeasible when the data are sparse. In the Euclidean setting, this issue is addressed by the technique of principal analysis through conditional expectation (PACE, Yao, Müller and Wang (2005)). The technique was also adopted by Dai, Lin and Müller (2021) for their ambient approach to Riemannian functional data analysis on sparsely observed data. To adapt this technique in our intrinsic framework, for each $T_{\hat{\mu}(T_{ij})} \mathcal{M}$, we fix an orthonormal basis $B_{ij,1}, \dots, B_{ij,d}$; in Proposition 3.1, we will show that the computed scores do not depend on the choice of the basis. Then the observations $\text{Log}_{\hat{\mu}(T_{ij})} Y_{ij}$ and the estimated eigenfunctions $\hat{\psi}_k(T_{ij})$ can be represented by their respective coordinate vectors z_{ij} and $g_{k,ij}$ with respect to the basis. Similarly, the estimated covariance function $\hat{C}(T_{ij}, T_{il})$ at (T_{ij}, T_{il}) can be represented by a matrix $C_{i,jl}$ of coefficients. By treating the vectors z_{ij} as \mathbb{R}^d -valued observations, the best linear unbiased predictor (BLUP) of ξ_{ik} is given by

$$(9) \quad \hat{\xi}_{ik} = \hat{\lambda}_k g_{k,i}^\top \Sigma_i^{-1} z_i,$$

where $g_{k,i} = (g_{k,i1}^\top, \dots, g_{k,im_i}^\top)^\top$, $z_i = (z_{i1}^\top, \dots, z_{im_i}^\top)^\top$ and

$$\Sigma_i = \hat{\sigma}^2 \mathbf{I} + \begin{pmatrix} C_{i,11} & C_{i,12} & \cdots & C_{i,1m_i} \\ C_{i,21} & C_{i,22} & \cdots & C_{i,2m_i} \\ \vdots & \vdots & \ddots & \vdots \\ C_{i,m_i1} & C_{i,m_i2} & \cdots & C_{i,m_im_i} \end{pmatrix}$$

with $\hat{\sigma}^2 = \sum_{i=1}^n \sum_{j=1}^{m_i} (ndm_i)^{-1} \text{tr}\{z_{ij} z_{ij}^\top - \hat{C}(T_{ij}, T_{ij})\}$. The following invariance principle shows that the scores $\hat{\xi}_{ik}$ in (9) are invariant to the choice of bases $B_{ij,1}, \dots, B_{ij,d}$. This extends the invariance principle of Lin and Yao (2019) from the fully observed and/or dense design to the sparse case.

PROPOSITION 3.1. *The principal component scores $\hat{\xi}_{ik}$ in (9) do not depend on the choice of the orthonormal bases $\{(B_{ij,1}, \dots, B_{ij,d}) : i = 1, \dots, n, j = 1, \dots, m_i\}$.*

It remains to choose the bandwidths h_μ and h_C . Although the theoretical analysis in the next section sheds light on how to choose them when the sample size is large, to determine appropriate values for them when the sample is limited, we propose the following k -fold cross-validation procedure. For an integer $k \geq 2$, divide the subjects into k partitions, denoted by $P_1, \dots, P_k \subset \{1, \dots, n\}$, of roughly even size. Let $P_{-l} = \{1, \dots, n\} \setminus P_l$ for $l = 1, \dots, k$. For a candidate value h of h_μ , its cross-validation error is computed by

$$\text{CV}(h) = \sum_{l=1}^k \sum_{i \in P_j} \sum_{j=1}^{m_i} d_{\mathcal{M}}^2(\hat{\mu}_{-l}^h(T_{ij}), Y_{ij}),$$

where $\hat{\mu}_{-l}^h$ is the estimated mean function by using the bandwidth h and the data P_{-l} . Among a set of candidate values of h_μ , the one with the minimal cross-validation error is selected. A value for h_C can be selected by the similar procedure. As demonstrated in Section 5, this k -fold cross-validation procedure is numerically effective.

4. Asymptotic properties. In the sequel, we assume $m_1 = \dots = m_n = m$ for a clear exposition; extension to more general cases is technically straightforward (Zhang and Wang (2016)). There are two popular types of designs, namely, the random design in which the design points T_{ij} are i.i.d. sampled from a distribution, and the deterministic design in which T_{ij} are predetermined, and thus nonrandom. For the random design, the following assumption is commonly adopted (Li and Hsing (2010), Yao, Müller and Wang (2005), Zhang and Wang (2016)).

ASSUMPTION 4.1 (Random design). The design points T_{ij} , independent of other random quantities, are i.i.d. sampled from a distribution on \mathcal{T} with a probability density that is bounded away from zero and infinity.

In contrast, the deterministic design is less studied, especially the irregular deterministic design; for instance, Cai and Yuan (2011) considers only a regular deterministic design. In this paper, we consider a deterministic design with the following condition that basically states that the design points are sufficiently irregular. To focus on longitudinal observations, the regular design of a common grid case is studied Section 5 of the Supplement (Shao, Lin and Yao (2022)), and is not included in the following condition.

ASSUMPTION 4.2 (Deterministic design). The design points T_{ij} are nonrandom, and there exist constants $c_2 \geq c_1 > 0$, such that for any interval $A, B \subset \mathcal{T}$ and all $n \geq 1$:

- (a) $\sup_{1 \leq i \leq n} \sum_{j=1}^m 1_{T_{ij} \in A} \leq \max\{c_2 m |A|, 1\}$,
- (b) $c_1 n m |A| - 1 \leq \sum_{i,j} 1_{T_{ij} \in A} \leq \max\{c_2 n m |A|, 1\}$, and
- (c) $c_1 n m^2 |A| |B| - 1 \leq \sum_{i,j,k} 1_{T_{ij} \in A} 1_{T_{ik} \in B} \leq \max\{c_2 n m^2 |A| |B|, 1\}$,

where $|A|$ denotes the length of A .

In many applications, the design points are neither completely random nor completely predetermined. For example, in longitudinal studies the visit of a patient may take place at a time that randomly deviates from the scheduled time. Such design includes both a deterministic part and a random component, which is termed *hybrid design* in this paper. Specifically, suppose all measurements are scheduled to take place in some of the L predetermined points of \mathcal{T} . Without loss of generality, we assume these predetermined points are $\mathcal{A}_L := \{s_k : 1 \leq k \leq L\}$ with $s_k = k/(L+1)$ and the set of all m -element subsets of \mathcal{A}_L is $\mathcal{S}_m = \{\{t_1, \dots, t_m\} : t_1, \dots, t_m \in \mathcal{A}_L \text{ are distinct}\}$. There are $m \leq L$ measurements scheduled at distinct time points $\mathbb{S}_i := \{S_{i1}, \dots, S_{im}\} \in \mathcal{S}_m$ for each curve i . Instead of S_{ij} , the actual measurement takes place at $T_{ij} = S_{ij} + \zeta_{ij}$ for some random variable $\zeta_{ij} \in (-1/(2L+2), 1/(2L+2))$. We postulate the following condition in which we emphasize that S_{i1}, \dots, S_{im} are *not* independent, and thus neither are T_{i1}, \dots, T_{im} . In addition, note that the condition also includes the special case that S_{i1}, \dots, S_{im} are deterministic when $m = L$.

ASSUMPTION 4.3 (Hybrid design).

(1) For each $i = 1, \dots, n$, S_{i1}, \dots, S_{im} are m distinct elements randomly sampled (without replacement) from \mathcal{A}_L . In addition, $\mathbb{S}_1, \dots, \mathbb{S}_n$ are i.i.d. random subsets of \mathcal{A}_L and there are positive constants c_1, c_2 such that $c_1 |\mathcal{S}_m|^{-1} \leq \Pr(\mathbb{S}_1 \in \mathbf{s}) \leq c_2 |\mathcal{S}_m|^{-1}$ for all $\mathbf{s} \in \mathcal{S}_m$.

(2) ζ_{ij} are i.i.d. centered random variables taking values in $(-1/(2L+2), 1/(2L+2))$ and are independent of other random quantities. In addition, there exist universal positive constants c_3 and c_4 such that the probability density f_ζ of ζ_{11} satisfies $c_3 L \leq \inf_s f_\zeta(s) \leq \sup_s f_\zeta(s) \leq c_4 L$.

Although these designs differ in nature, in the next two sections, we show that the estimator with either of these designs, respectively, for the mean function and for covariance function, achieves the same convergence rate under suitable regularity conditions.

4.1. *Mean function.* The pointwise convergence rate of the estimate $\hat{\mu}(t)$ is established in Petersen and Müller (2019), while the uniform convergence rate is derived by Dai, Lin and Müller (2021). For completeness, we include them here, and establish a new local uniform result that is needed in the theoretical analysis of the covariance estimator. First, we require the following assumptions, where the condition 2 may be replaced with tail and moment conditions on the distributions of Y and X at the cost of heavier technicalities. In addition, by modifying our proofs, the compactness in the condition 4 can be replaced with a condition on the decay rate of the kernel function when it moves away from zero, so that noncompact kernels such as Gaussian kernel can be accommodated.

ASSUMPTION 4.4.

- (1) The Riemannian manifold \mathcal{M} is complete and simply connected.⁷
- (2) There exists a compact subset of $\mathcal{K} \subset \mathcal{M}$ such that $\Pr\{X(t), Y(t) \in \mathcal{K} \text{ for all } t \in \mathcal{T}\} = 1$.
- (3) The domain \mathcal{T} is a compact interval.
- (4) The kernel function K is Lipschitz continuous, symmetric, positive on $(-1, 1)$, compactly supported on $[-1, 1]$ and monotonically decreasing on $[0, 1]$.

The following regularity on the mean function or related quantities is adapted from Petersen and Müller (2019) and is specialized to the Riemannian manifold. Part 2 states that the Fréchet mean is well separated from the other points in terms of the Fréchet function $F^*(y, t) := \mathbf{E}d_{\mathcal{M}}^2(X(t), y)$, while part 3 basically amounts to convexity of $F^*(\cdot, t)$ around $\mu(t)$; they hold, for example, when the manifold has nonpositive curvature or the data sufficiently concentrate on a geodesically convex region.

ASSUMPTION 4.5.

- (1) The second partial derivative $\partial_t^2 F^*(y, t)$ is bounded on $\mathcal{K} \times \mathcal{T}$.
- (2) For any $\delta > 0$,

$$\inf_{\substack{d_{\mathcal{M}}(y, \mu(t)) > \delta \\ t \in \mathcal{T}}} \{F^*(y, t) - F^*(\mu(t), t)\} > 0,$$

- (3) There exist $\eta_1 > 0$ and $C_1 > 0$ such that for all $t \in \mathcal{T}$ and all y with $d_{\mathcal{M}}(y, \mu(t)) < \eta_1$,

$$F^*(y, t) - F^*(\mu(t), t) - C_1 d_{\mathcal{M}}(y, \mu(t))^2 \geq 0.$$

The following proposition, whose proof, as well as proofs for other results in this section, is deferred to Sections 2, 3 and 4 of Supplement (Shao, Lin and Yao (2022)), states the pointwise and uniform convergence rates of the estimated mean function, where the pointwise rate is an immediate consequence of the local uniform rate stated in Proposition 4.2. The condition $nh_{\mu} \gtrsim 1$ in the following is only needed for the deterministic design.

⁷See Section 1 of the Supplement (Shao, Lin and Yao (2022)) for a precise definition of simple connectedness.

PROPOSITION 4.1. *Suppose that Assumptions 2.1, 2.2, 3.1, 4.4 and 4.5. Under either of Assumptions 4.1, 4.2 and 4.3, if $h_\mu \rightarrow 0$ and $nmh_\mu \rightarrow \infty$, then for any fixed $t \in \mathcal{T}$,*

$$d_{\mathcal{M}}^2(\mu(t), \hat{\mu}(t)) = O_p\left(h_\mu^4 + \frac{1}{n} + \frac{1}{nmh_\mu}\right),$$

and if $h_\mu \rightarrow 0$, $nh_\mu \gtrsim 1$ and $nmh_\mu/\log n \rightarrow \infty$, then

$$\sup_{t \in \mathcal{T}} d_{\mathcal{M}}^2(\mu(t), \hat{\mu}(t)) = O_p\left(h_\mu^4 + \frac{\log n}{n} + \frac{\log n}{nmh_\mu}\right).$$

To derive the pointwise convergence rate of the estimator $\hat{\mathcal{C}}(s, t)$ in the next subsection, we require a local convergence property of the estimator $\hat{\mu}$. The following Proposition 4.2, which is new in the literature, shows that the local uniform convergence rate is the same as the pointwise rate in Proposition 4.1, and differs from the global uniform convergence rate that has an additional $\log n$ factor. The reason for this phenomenon is that $\mathbf{E}\{K_{h_\mu}(T - t)\} = 1$ at a fixed-point t but $\mathbf{E}\{\sup_{t \in \mathcal{T}} K_{h_\mu}(T - t)\} = 1/h_\mu \rightarrow \infty$. Therefore, the additional $\log n$ factor is needed to offset this explosion in the case of global uniform convergence. In the local case, if $h = O(h_\mu)$, and thus $\mathbf{E}\{\sup_{\tau: |\tau - t| \leq h} K_{h_\mu}(T - \tau)\} = O(h/h_\mu) = O(1)$, then no offset is required. The proposition also directly implies the pointwise rate in Proposition 4.1.

PROPOSITION 4.2. *Suppose that Assumptions 2.1, 2.2, 3.1, 4.4 and 4.5 hold. Under either of Assumptions 4.1, 4.2 and 4.3, if $h_\mu \rightarrow 0$ and $nmh_\mu \rightarrow \infty$, then for any fixed t and $h = O(h_\mu)$,*

$$\sup_{\tau: |\tau - t| \leq h} d_{\mathcal{M}}^2(\mu(\tau), \hat{\mu}(\tau)) = O_p\left(h_\mu^4 + \frac{1}{n} + \frac{1}{nmh_\mu}\right).$$

4.2. *Covariance function.* We start with the following assumption on the regularity of the covariance function \mathcal{C} . As discussed in Section 2.4, such regularity condition in the manifold setting is made precise and meaningful by the constructed covariance vector bundle \mathbb{L} and the covariant derivative ∇ in (6).

ASSUMPTION 4.6. The covariance function \mathcal{C} is twice differentiable and its second derivatives are continuous.

To study the asymptotic properties of the estimator $\hat{\mathcal{C}}$, one of the major challenges that are not encountered in the Euclidean setting of Zhang and Wang (2016) or the ambient case of Dai, Lin and Müller (2021) is to deal with the parallel transport in (8). It turns out that we need to quantify the discrepancy between a tangent vector and the parallelly transported one along a geodesic quadrilateral. We address this issue by the following lemma which may be of independent interest. In particular, the proof of the lemma given in Section 3 of the Supplement (Shao, Lin and Yao (2022)) utilizes holonomy theory that appears new in statistical literature.

LEMMA 4.1. *For a compact subset $\mathcal{G} \subset \mathcal{M}$, there exists a constant $c > 0$ depending only on \mathcal{G} , such that for all $p_1, p_2, q_1, q_2, y \in \mathcal{G}$,*

$$\|\mathcal{P}_{q_1}^{p_1} \mathcal{P}_{q_2}^{q_1} \text{Log}_{q_2} y - \mathcal{P}_{p_2}^{p_1} \text{Log}_{p_2} y\|_{p_1} \leq c(d_{\mathcal{M}}(p_1, q_1) + d_{\mathcal{M}}(p_2, q_2)).$$

With the above regularity condition and lemma, the following theorem establishes the pointwise convergence rate of $\hat{\mathcal{C}}$.

THEOREM 4.1. *Suppose that Assumptions 2.1, 2.2, 3.1, 4.4, 4.5 and 4.6 hold. Under either of Assumptions 4.1, 4.2 and 4.3, if $h_\mu \rightarrow 0$, $h_C = O(h_\mu)$, and $\min\{nmh_\mu, nm^2h_C^2\} \rightarrow \infty$, then for any fixed $s, t \in \mathcal{T}$,*

$$(10) \quad \left\| \mathcal{P}_{(\hat{\mu}(s), \hat{\mu}(t))}^{\mu(s), \mu(t)} \hat{\mathcal{C}}(s, t) - \mathcal{C}(s, t) \right\|_{G(\mu(s), \mu(t))}^2 = O_p \left(h_\mu^4 + h_C^4 + \frac{1}{n} + \frac{1}{nmh_\mu} + \frac{1}{nm^2h_C^2} \right).$$

The rate in the above theorem matches the pointwise rate in the Euclidean setting of Zhang and Wang (2016) in the case of $m_i = m$. Unlike Zhang and Wang (2016), which assumes that the mean function is known in their analysis, we do not need such assumption thanks to the local uniform rate of the mean function stated in Proposition 4.2. In our analysis, the local uniform rate cannot be replaced with the global uniform rate in Proposition 4.1 without introducing an additional $\log n$ factor. Although the condition $h_C = O(h_\mu)$ is required in order to utilize Proposition 4.2, it does not limit the convergence rate, as a proper choice of h_μ and h_C leads to the following rates that still match the rates of Zhang and Wang (2016).

COROLLARY 4.1. *Assume the conditions of Theorem 4.1.*

(a) *When $m \asymp n^{1/4}$ or $m \gg n^{1/4}$, with $h_\mu \asymp h_C \asymp n^{-1/4}$, one has*

$$\left\| \mathcal{P}_{(\hat{\mu}(s), \hat{\mu}(t))}^{\mu(s), \mu(t)} \hat{\mathcal{C}}(s, t) - \mathcal{C}(s, t) \right\|_{G(\mu(s), \mu(t))}^2 = O_p \left(\frac{1}{n} \right).$$

(b) *When $m \ll n^{1/4}$, with $h_\mu \asymp h_C \asymp n^{-1/6}m^{-1/3}$, one has*

$$\left\| \mathcal{P}_{(\hat{\mu}(s), \hat{\mu}(t))}^{\mu(s), \mu(t)} \hat{\mathcal{C}}(s, t) - \mathcal{C}(s, t) \right\|_{G(\mu(s), \mu(t))}^2 = O_p \left(\frac{1}{n^{2/3}m^{4/3}} \right).$$

Like the Euclidean case, a phase transition is observed at $m \asymp n^{1/4}$. With a proper choice of h_μ and h_C , if m grows at least as fast as $n^{1/4}$, it does not impact the convergence rate that is at a parametric order of magnitude, that is, $n^{-1/2}$. Otherwise, the sampling rate m becomes an integrated part of the convergence rate of $\hat{\mathcal{C}}$. In particular, when $m \ll n^{1/4}$, the choice $h_\mu \asymp n^{-1/6}m^{-1/3}$ is required to respect the condition $h_C = O(h_\mu)$. This choice is strictly larger than the optimal choice $h_\mu \asymp (nm)^{-1/5}$ that is implied by Proposition 4.1 in the case of $m \ll n^{1/4}$. This suggests that oversmoothing in the mean function estimation may be needed in order to reach the optimal pointwise rate of the covariance estimator when $m \ll n^{1/4}$. It is interesting to note that, in the literature if a lower-dimensional estimate depends on a higher-dimensional one, undersmoothing the latter often helps the former to attain a better rate. For instance, undersmoothing the two-dimensional covariance surface estimate leads to a better rate of the one-dimensional eigenfunction estimates (Hall, Müller and Wang (2006)). In contrast, the phenomenon in our case is reversed: the two-dimensional covariance function estimate depends on the one-dimensional mean estimate, thus requires oversmoothing the latter instead.

The following results establish the uniform convergence rate of the estimator $\hat{\mathcal{C}}$, where the condition $\min\{nh_\mu, nh_C^2\} \gtrsim 1$ is only needed for the deterministic design.

THEOREM 4.2. *Suppose that Assumptions 2.1, 2.2, 3.1, 4.4, 4.5 and 4.6 hold. Under either of Assumptions 4.1, 4.2 and 4.3, if $\max\{h_\mu, h_C\} \rightarrow 0$, $\min\{nh_\mu, nh_C^2\} \gtrsim 1$ and $\min\{nmh_\mu, nm^2h_C^2\}/\log n \rightarrow \infty$, then*

$$(11) \quad \begin{aligned} & \sup_{(s,t) \in \mathcal{T}^2} \left\| \mathcal{P}_{(\hat{\mu}(s), \hat{\mu}(t))}^{\mu(s), \mu(t)} \hat{\mathcal{C}}(s, t) - \mathcal{C}(s, t) \right\|_{G(\mu(s), \mu(t))}^2 \\ & = O_p \left(h_\mu^4 + h_C^4 + \frac{\log n}{n} + \frac{\log n}{nmh_\mu} + \frac{\log n}{nm^2h_C^2} \right). \end{aligned}$$

COROLLARY 4.2. Assume the conditions of Theorem 4.2.

(a) When $m \asymp n^{1/4}$ or $m \gg n^{1/4}$, with $h_\mu \asymp h_C \asymp n^{-1/4}$, one has

$$\sup_{(s,t) \in \mathcal{T}^2} \left\| \mathcal{P}_{(\hat{\mu}(s), \hat{\mu}(t))}^{(\mu(s), \mu(t))} \hat{C}(s, t) - C(s, t) \right\|_{G_{(\mu(s), \mu(t))}}^2 = O_p \left(\frac{\log n}{n} \right).$$

(b) When $m \ll n^{1/4}$, with $h_\mu \asymp n^{-1/5} m^{-1/5} (\log n)^{1/5}$ and $h_C \asymp n^{-1/6} m^{-1/3} (\log n)^{1/6}$, one has

$$\sup_{(s,t) \in \mathcal{T}^2} \left\| \mathcal{P}_{(\hat{\mu}(s), \hat{\mu}(t))}^{(\mu(s), \mu(t))} \hat{C}(s, t) - C(s, t) \right\|_{G_{(\mu(s), \mu(t))}}^2 = O_p \left(\frac{(\log n)^{2/3}}{n^{2/3} m^{4/3}} \right).$$

These rates again match the uniform rates in Zhang and Wang (2016). They also coincide with the rates⁸ in Dai and Müller (2018). It is interesting to see that, when $m \ll n^{1/4}$, the choice of h_μ in the corollary is the same as the optimal choice implied by Proposition 4.1, which suggests that no oversmoothing is needed in order to reach the optimal uniform rate for the covariance estimator \hat{C} . This is because, the local uniform result of Proposition 4.2 and thus the condition $h_C = O(h_\mu)$ are not required, as the role of Proposition 4.2 in the analysis is now played by Proposition 4.1.

5. Simulation studies. We consider three different manifolds for illustrating the numerical properties of the proposed covariance estimator (8) in Section 3; the numerical performance of the mean estimator can be found in Dai, Lin and Müller (2021). Namely, they are the two-dimensional unit sphere \mathbb{S}^2 , the manifold Sym_{LC}^+ of symmetric positive-definite 2×2 matrices with the Log-Cholesky metric (Lin (2019)) and the manifold Sym_{AF}^+ of symmetric positive-definite 2×2 matrices with the affine-invariant metric (Moakher (2005)), representing manifolds of positive, zero and negative sectional curvature, respectively. Note that although Sym_{LC}^+ and Sym_{AF}^+ share the same collection of matrices, they are endowed with different Riemannian metric tensors, and thus have fundamentally different Riemannian geometry. We set $\mathcal{T} = [0, 1]$. The sampling rate m_i is randomly sampled from $\text{Poisson}(m) + 2$, where $\text{Poisson}(m)$ is a Poisson distribution with parameter m . Conditional on m_i , the time points T_{i1}, \dots, T_{im_i} are i.i.d. sampled from the uniform distribution $\text{Uniform}(0, 1)$. The random process X and its mean and covariance functions are described below.

Sphere \mathbb{S}^2 . We parameterize $\mathbb{S}^2 = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}$ by the polar coordinate system

$$(12) \quad x(u, v) = \cos(u) \sin(v), \quad y(u, v) = \cos(u) \cos(v), \quad z(u, v) = \sin(u)$$

for the latitude $u \in (-\pi/2, \pi/2)$ and longitude $v \in [0, 2\pi)$. This coordinate system also gives rise to a local chart $\phi : U \rightarrow (-\pi/2, \pi/2) \times [0, 2\pi)$ on $V = \mathbb{S}^2 \setminus \{(0, 0, -1), (0, 0, 1)\}$. Let $B_1(t) = \frac{\partial \phi}{\partial u}$ and $B_2(t) = \frac{\partial \phi}{\partial v}$. The random process X is then given by

$$X(t) = \text{Exp}_{\mu(t)}(tZ_1 B_1(t) + tZ_2 B_2(t))$$

with $Z_1, Z_2 \stackrel{i.i.d.}{\sim} \text{Uniform}(-0.1, 0.1)$. The mean curve μ of X is $\mu(t) = \phi(0, \pi t/2) = (\sin(\pi t/2), \cos(\pi t/2), 0)$, which lies on the equator. The covariance function is $C(s, t) = \frac{st}{300} \mathbf{I}_2$ under the frame (B_1, B_2) , where \mathbf{I}_2 denotes the 2×2 identity matrix. The contaminated observations are

$$Y_{ij} = \text{Exp}_{\mu(T_{ij})} \left\{ (T_{ij} Z_{1i} + v_{ij1}) B_1(T_{ij}) + (T_{ij} Z_{2i} + v_{ij2}) B_2(T_{ij}) \right\},$$

⁸Note that the extra term $1/(nmh_C)$ in Dai and Müller (2018) is dominated by $1/n + 1/(nm^2 h_C^2)$ due to the inequality of arithmetic and geometric means, that is, $\sqrt{ab} \leq (a + b)/2$.

where $Z_{1i}, Z_{2i} \stackrel{i.i.d.}{\sim} \text{Uniform}(-0.1, 0.1)$ and $v_{ij1}, v_{ij2} \stackrel{i.i.d.}{\sim} \text{Uniform}(-a, a)$ with $a > 0$ chosen to make $\text{SNR} = 5$ defined by

$$(13) \quad \text{SNR} := \frac{\mathbf{E} \int_{\mathcal{T}} \|\text{Log}_{\mu(t)} X(t)\|_{\mu(t)}^2 dt}{\mathbf{E} \int_{\mathcal{T}} \|\varepsilon(t)\|_{\mu(t)}^2 dt}.$$

Manifold Sym_{LC}^+ . We parameterize Sym_{LC}^+ by the chart

$$\phi : (u, v, w) \rightarrow \begin{pmatrix} e^{2u} & we^u \\ we^u & w^2 + e^{2v} \end{pmatrix},$$

which induces the orthogonal frame formed by $B_1(t) = \frac{\partial \phi}{\partial u}$, $B_2(t) = \frac{\partial \phi}{\partial v}$ and $B_3(t) = \frac{\partial \phi}{\partial w}$. The random process X is set to be

$$\begin{aligned} X(t) &= \text{Exp}_{\mu(t)}(tZ_1B_1(t) + tZ_2B_2(t) + tZ_3B_3(t)) \\ &= \begin{pmatrix} e^{t+tZ_1} & 0 \\ t+tZ_3 & e^{t+tZ_2} \end{pmatrix} \begin{pmatrix} e^{t+tZ_1} & t+tZ_3 \\ 0 & e^{t+tZ_2} \end{pmatrix} \end{aligned}$$

with $Z_1, Z_2, Z_3 \stackrel{i.i.d.}{\sim} \text{Uniform}(-0.1, 0.1)$. The mean curve μ is a geodesic with

$$\mu(t) = \phi(t, t, t) = \begin{pmatrix} e^{2t} & te^t \\ te^t & t^2 + e^{2t} \end{pmatrix}$$

and the covariance function is $\mathcal{C}(s, t) = \frac{st}{300} \mathbf{I}_3$ under the frame (B_1, B_2, B_3) . With $\varepsilon(T_{ij}) = v_{ij1}B_1(T_{ij}) + v_{ij2}B_2(T_{ij}) + v_{ij3}B_3(T_{ij}) \in T_{\mu(T_{ij})}\mathcal{M}$, the contaminated observations are

$$Y_{ij} = \begin{pmatrix} e^{T_{ij}+T_{ij}Z_{1i}+v_{ij1}} & 0 \\ T_{ij} + T_{ij}Z_{3i} + v_{ij3} & e^{T_{ij}+T_{ij}Z_{2i}+v_{ij2}} \end{pmatrix} \begin{pmatrix} e^{T_{ij}+T_{ij}Z_{1i}+v_{ij1}} & T_{ij} + T_{ij}Z_{3i} + v_{ij3} \\ 0 & e^{T_{ij}+T_{ij}Z_{2i}+v_{ij2}} \end{pmatrix},$$

where $Z_{1i}, Z_{2i}, Z_{3i} \stackrel{i.i.d.}{\sim} \text{Uniform}(-0.1, 0.1)$ and $v_{ij1}, v_{ij2}, v_{ij3} \stackrel{i.i.d.}{\sim} \text{Uniform}(-a, a)$ with $a > 0$ set to satisfy $\text{SNR} = 5$ defined in (13).

Manifold Sym_{AF}^+ . We parameterize Sym_{AF}^+ by the chart

$$\phi : (u, v, w) \rightarrow \begin{pmatrix} e^u & w \\ w & e^v \end{pmatrix}$$

which gives rise to the frame formed by $B_1(t) = \frac{\partial \phi}{\partial u}$, $B_2(t) = \frac{\partial \phi}{\partial v}$ and $B_3(t) = \frac{\partial \phi}{\partial w}$. The random process $X(t)$ is set to

$$X(t) = \begin{pmatrix} \frac{1}{4}e^{t+tZ_1} + \frac{3}{4}e^{t+tZ_2} & \frac{\sqrt{3}}{4}e^{t+tZ_1} - \frac{\sqrt{3}}{4}e^{t+tZ_2} \\ \frac{\sqrt{3}}{4}e^{t+tZ_1} - \frac{\sqrt{3}}{4}e^{t+tZ_2} & \frac{3}{4}e^{t+tZ_1} + \frac{1}{4}e^{t+tZ_2} \end{pmatrix},$$

for $Z_1, Z_2 \stackrel{i.i.d.}{\sim} \text{Uniform}(-0.1, 0.1)$. The mean function is $\mu(t) = e^t \mathbf{I}_2$ while the covariance function is $\mathcal{C}(s, t) = \text{diag}\{st/300, st/300, 0\}$ under the frame (B_1, B_2, B_3) , providing an illustration on covariance structure of nonfull rank. With $\varepsilon(T_{ij}) = v_{ij1}B_1(T_{ij}) + v_{ij2}B_2(T_{ij}) \in T_{\mu(T_{ij})}\mathcal{M}$, the contaminated observations are

$$Y_{ij} = \begin{pmatrix} \frac{1}{4}e^{T_{ij}+T_{ij}Z_{1i}+v_{ij1}} + \frac{3}{4}e^{T_{ij}+T_{ij}Z_{2i}+v_{ij2}} & \frac{\sqrt{3}}{4}e^{T_{ij}+T_{ij}Z_{1i}+v_{ij1}} - \frac{\sqrt{3}}{4}e^{T_{ij}+T_{ij}Z_{2i}+v_{ij2}} \\ \frac{\sqrt{3}}{4}e^{T_{ij}+T_{ij}Z_{1i}+v_{ij1}} - \frac{\sqrt{3}}{4}e^{T_{ij}+T_{ij}Z_{2i}+v_{ij2}} & \frac{3}{4}e^{T_{ij}+T_{ij}Z_{1i}+v_{ij1}} + \frac{1}{4}e^{T_{ij}+T_{ij}Z_{2i}+v_{ij2}} \end{pmatrix},$$

TABLE 1
*r*MUIE and its Monte Carlo standard errors under different settings in percentage (%)

Manifold	n	<i>r</i> MUIE			
		$m = 5$	$m = 10$	$m = 20$	$m = 30$
\mathbb{S}^2	100	35.40 (17.50)	27.60 (12.43)	18.97 (7.53)	17.69 (10.91)
	200	26.36 (12.26)	20.72 (13.79)	14.94 (6.07)	13.85 (4.75)
	400	18.04 (8.78)	12.47 (4.41)	10.48 (2.52)	8.30 (4.19)
Sym_{LC}^+	100	41.58 (13.24)	36.70 (37.70)	25.44 (8.42)	22.10 (5.56)
	200	30.36 (10.41)	22.05 (6.51)	20.89 (7.14)	15.51 (3.52)
	400	24.15 (12.30)	14.55 (5.13)	12.47 (4.85)	12.09 (2.46)
Sym_{AF}^+	100	35.40 (17.50)	27.60 (12.43)	18.97 (7.53)	18.77 (7.24)
	200	26.35 (12.26)	20.72 (13.79)	14.94 (6.05)	13.85 (4.75)
	400	18.04 (8.78)	12.49 (4.40)	10.48 (2.52)	8.30 (4.19)

where $Z_{1i}, Z_{2i}, Z_{3i} \stackrel{i.i.d.}{\sim} \text{Uniform}(-0.1, 0.1)$ and $v_{ij1}, v_{ij2} \stackrel{i.i.d.}{\sim} \text{Uniform}(-a, a)$ with $a > 0$ set to satisfy $\text{SNR} = 5$ defined in (13).

We consider different sample sizes and sampling rates, namely, $n = 100, 200, 400$ and $m = 5, 10, 20, 30$. Each simulation is repeated independently 100 times. The kernel adopted is the tricube kernel defined by $K(u) = 70(1 - |u|^3)^3/81$, and the bandwidths h_μ and h_C are selected by the two-fold cross-validation procedure described in Section 3. Estimation quality is measured by relative mean uniform integrated error (*r*MUIE) and relative root mean integrated squared error (*r*RMISE), defined by

$$(14) \quad \begin{aligned} \text{rMUIE} &:= \frac{\mathbf{E} \sup_{s,t \in \mathcal{T}} \|\mathcal{P}_{(\hat{\mu}(s), \hat{\mu}(t))}^{(\mu(s), \mu(t))} \hat{\mathcal{C}}(s, t) - \mathcal{C}(s, t)\|_G}{\sup_{s,t \in \mathcal{T}} \|\mathcal{C}(s, t)\|_G}, \\ \text{rRMISE} &:= \frac{\{\mathbf{E} \int_{\mathcal{T}^2} \|\mathcal{P}_{(\hat{\mu}(s), \hat{\mu}(t))}^{(\mu(s), \mu(t))} \hat{\mathcal{C}}(s, t) - \mathcal{C}(s, t)\|_G^2 ds dt\}^{1/2}}{\{\int_{\mathcal{T}^2} \|\mathcal{C}(s, t)\|_G^2 ds dt\}^{1/2}}. \end{aligned}$$

The results, summarized in Tables 1 and 2, show that the estimation errors in terms of both *r*MUIE and *r*RMISE in percentage decrease as n or m increases, and thus demonstrate the effectiveness of the proposed estimation method. A phase transition phenomenon is also

TABLE 2
*r*RMISE and its Monte Carlo standard errors under different settings in percentage (%)

Manifold	n	<i>r</i> RMISE			
		$m = 5$	$m = 10$	$m = 20$	$m = 30$
\mathbb{S}^2	100	24.22 (8.70)	20.63 (7.97)	16.10 (6.11)	15.68 (6.73)
	200	17.16 (5.73)	14.00 (6.14)	12.10 (4.48)	11.89 (4.55)
	400	11.99 (4.45)	9.29 (3.03)	8.81 (1.81)	6.90 (3.49)
Sym_{LC}^+	100	29.52 (7.20)	25.98 (12.15)	21.66 (6.63)	19.01 (2.98)
	200	21.13 (5.19)	16.27 (3.81)	18.99 (6.05)	13.95 (2.94)
	400	16.29 (4.33)	11.04 (2.54)	10.99 (4.33)	10.08 (2.37)
Sym_{AF}^+	100	24.22 (8.70)	20.63 (7.97)	16.10 (6.11)	15.37 (5.75)
	200	17.16 (5.73)	14.00 (6.14)	13.21 (4.96)	11.89 (4.55)
	400	11.99 (4.45)	10.55 (3.55)	9.81 (1.81)	6.90 (3.49)

observed: When m is increased from 5 to 10 or 20, the errors in terms of both rMUIE and rRMISE decrease substantially, while when m is further increased to 30, the decrease in errors is marginal. This phenomenon, hinted by our theoretical analysis in Section 4, suggests that for a fixed sample size, when $m = 5$ or $m = 10$ the errors are primarily due to the low sampling rate m , while when $m = 30$ or higher the errors are mainly contributed by the sample size.

To numerically verify that the proposed framework is invariant to parameterization, we also computed the estimates with a different parameterization of the manifolds in the above. Specifically, we considered the following additional parameterization called stereographic projection:

$$(15) \quad \varphi : (u, v) \in \mathbb{R}^2 \rightarrow \left(\frac{2u}{u^2 + v^2 + 1}, \frac{2v}{u^2 + v^2 + 1}, \frac{u^2 + v^2 - 1}{u^2 + v^2 + 1} \right) \in \mathbb{S}^2$$

for the sphere \mathbb{S}^2 , parameterizing the matrices generated in the setting of Sym_{LC}^+ by their lower triangular parts instead of their Cholesky factors, and parameterizing the matrices in the setting of Sym_{AF}^+ by their Cholesky factors instead of their lower triangular parts. In addition, to verify that the results are invariant to frames, for each setting, we consider two sets of randomly selected frames for computation. We then found that identical results were obtained under different choices of parameterization and/or frames. This numerically demonstrates that the proposed framework and method are invariant to parameterization and the choice of frames. In addition, the manifold Sym_{AF}^+ does not have a canonical embedding. As a matter of fact, we did not employ an embedding for any of the above manifolds in our studies, demonstrating the intrinsicity of the proposed framework.

6. Application to longitudinal diffusion tensors. We apply the proposed framework to analyze longitudinal diffusion tensors from Alzheimer's Disease Neuroimaging Initiative (ADNI) database. The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. 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). For up-to-date information, see www.adni-info.org.

Diffusion tensor imaging (DTI), a special kind of diffusion-weighted magnetic resonance imaging, has been extensively adopted in brain science to investigate white matter tractography. In a DTI image, each brain voxel is associated with a 3×3 symmetric positive-definite matrix, called diffusion tensor, that characterizes diffusion of water molecules in the voxel. As diffusion of water molecules carries rich information about axons, diffusion tensor imaging has important applications in both clinical diagnostics and scientific research related to brain diseases. From a statistical perspective, diffusion tensors are modeled as random elements in $\text{Sym}_\star^+(3)$, and have been studied extensively, such as Arsigny et al. (2006), Dryden, Koloydenko and Zhou (2009), Fillard et al. (2005), Fletcher and Joshi (2007), Lenglet et al. (2006), Pennec (2020), Pennec, Fillard and Ayache (2006), Zhu et al. (2009), among many others. In these works $\text{Sym}_\star^+(3)$ is endowed with a Riemannian metric or a non-Euclidean distance that aims to alleviate or completely eliminate swelling effect (Arsigny et al. (2006/07)). However, none of them consider the longitudinal aspect of diffusion tensors.

We focus on the hippocampus, a brain region that plays an important role in memory and is central to Alzheimer's disease (Lindberg et al. (2012)), and include in the study subjects with at least four properly recorded DTI images. This results in a sample of $n = 177$ subjects with age ranging from 55.2 to 93.5. Among them, 42 subjects are cognitively normal (CN), while the others (AD) developed one of early mild cognitive impairment, mild cognitive

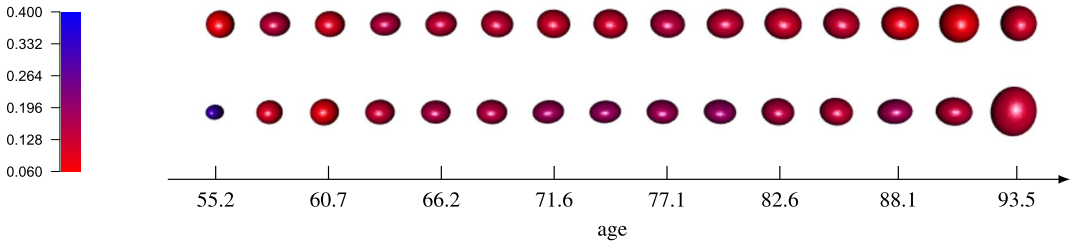


FIG. 3. Mean functions. Top: AD group; bottom: CN group. The color encodes fractional anisotropy.

impairment, late mild cognitive impairment and Alzheimer’s disease. On average, there are $m = 5.5$ DTI scans for each subject, which shows that the data are rather sparsely recorded. A standard procedure that includes denoising, eddy current and motion correction, skull stripping, bias correction and normalization is adopted to preprocess the raw images. Based on the preprocessed DTI images, diffusion tensors are derived. We endow $\text{Sym}_\star^+(3)$ with the Log-Cholesky metric (Lin (2019)) and turn it into a Riemannian manifold of nonpositive sectional curvature. Under the Log-Cholesky framework that avoids swelling effect and meanwhile enjoys computational efficiency, the Fréchet mean of the tensors inside hippocampus is calculated for each DTI scan, which represents a coarse-grain summary of hippocampal diffusion tensors. As we shall see below, this averaged mean tensor is already capable of illuminating some differences of the diffusion dynamics between the AD and CN groups.

The estimated Fréchet mean trajectories are depicted in Figure 3 with the bandwidth 4.2 for the AD group and 5.7 for the CN group, where each tensor is visualized as an ellipsoid whose volume corresponds to the determinant of the tensor. They suggest that, overall the averaged hippocampal diffusion tensor remains rather stable for the CN group; the tensors at age 55.2 and 93.5 that markedly depart from the others could be due to boundary effect, that is, there are relatively less data around the two boundary time points. In contrast, for the AD group, the dynamic tensor varies more substantially, and the diffusion (measured by the determinant of tensors and indicated by volume of ellipsoids) seems larger. Also, the mean trajectory of the AD group exhibits slightly lower fractional anisotropy at each time point. Fractional anisotropy, defined for each 3×3 symmetric positive-definite matrix A by

$$\text{FA} = \sqrt{\frac{3(\rho_1 - \bar{\rho})^2 + (\rho_2 - \bar{\rho})^2 + (\rho_3 - \bar{\rho})^2}{2(\rho_1^2 + \rho_2^2 + \rho_3^2)}},$$

where ρ_1, ρ_2, ρ_3 are eigenvalues of A and $\bar{\rho} = (\rho_1 + \rho_2 + \rho_3)/3$, describes the degree of anisotropy of diffusion of water molecules. It is close to zero unless movement of the water molecules is constrained by structures such as white matter fibers. The below-normal fractional anisotropy might suggest some damage on the hippocampal structure for the AD group.

For the covariance function, Figure 4 shows the first three intrinsic Riemannian functional principal components that are mapped on $\text{Sym}_\star^+(3)$ via the Riemannian exponential maps $\text{Exp}_{\hat{\mu}(t)}$, where the bandwidth is 3.5 for the AD group and 4.5 for the CN group. They respectively account for 40.2%, 22.2% and 7.0% of variance for the AD group, and 40.7%, 19.4% and 8.0% of variance for the CN group. These components, compared side by side in Figure 4, exhibit different patterns between the two cohorts. For instance, the Riemannian functional principal components of the AD group show relatively larger diffusion and more dynamics over time. In addition, they exhibit relatively lower fractional anisotropy, which suggests that individual diffusion tensor trajectories in the AD group tend to deviate from their mean trajectory along the direction with below-normal fractional anisotropy.

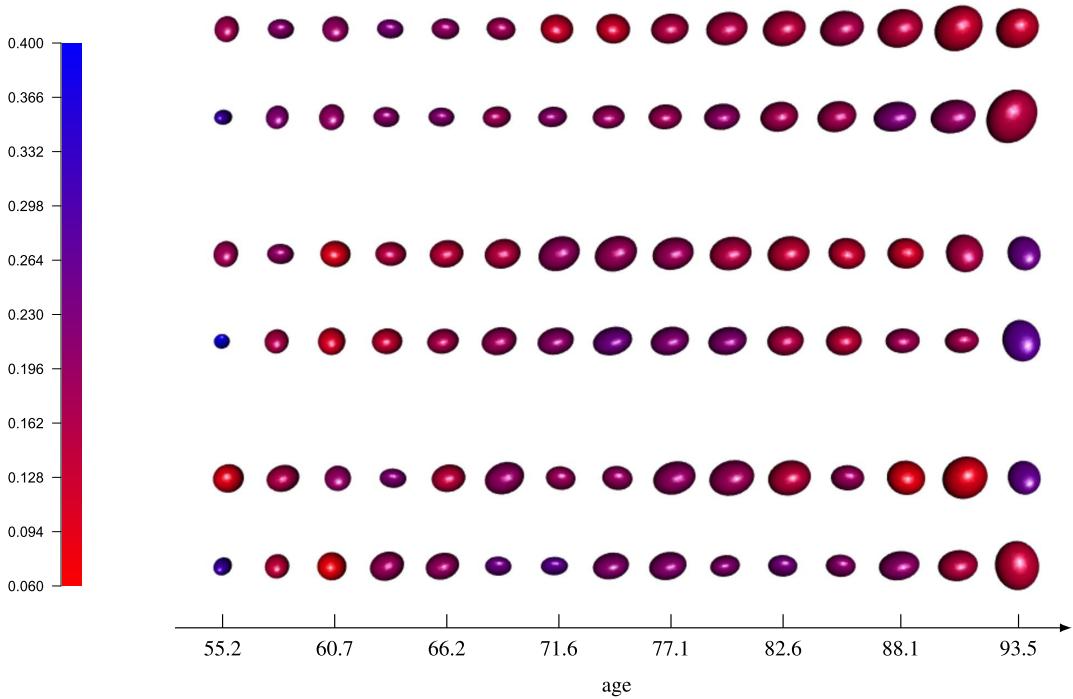


FIG. 4. *The first principal component of AD group (row 1) and CN group (row 2), the second principal component of AD group (row 3) and CN group (row 4), and the third principal component of AD group (row 5) and CN group (row 6). The color encodes fractional anisotropy.*

We conclude this section by the following remarks. Note that, in the above analysis, the averaged hippocampal diffusion tensors do not capture the rich spatial information of all tensors within the hippocampus. To account for such information, all hippocampal diffusion tensors shall be taken into consideration by being modeled as an $\text{Sym}_\star^+(3)$ -valued function defined on the hippocampal region, which is a three-dimensional domain of \mathbb{R}^3 . Along with the temporal dynamics, for each subject there are spatiotemporal Riemannian manifold-valued data, with the sparseness along the temporal direction. Our framework can be extended to analyze such data, but the extension requires substantial development and is left for future study.

In addition, each of the sparse trajectories is only observed in an individual-specific period shorter than the span ($93.6 - 55.2 = 38.4$ years) of the entire study. Functional data of this feature, called functional fragments (Delaigle et al. (2021), Descary and Panaretos (2019)) or functional snippets (Lin, Wang and Zhong (2021)), require special treatment on estimating the covariance structure. Particularly, local smoothing techniques can only estimate the diagonal region of the covariance function for such data, and thus require the additional assumption that the covariance function is supported in the diagonal region, as we have done implicitly in the above analysis. Extension of the estimation method proposed in this paper to functional fragments/snippets is nontrivial, and thus also left for future study.

Acknowledgments. Lingxuan Shao was a visiting student of Zhenhua Lin in National University of Singapore at the time of developing the paper. Lingxuan Shao and Zhenhua Lin are the joint first authors, and Fang Yao is the corresponding author.

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI inves-

tigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

Funding. Fang Yao’s research is supported by National Natural Science Foundation of China Grants 11931001 and 11871080, the National Key R&D Program of China Grant 2020YFE0204200, the LMAM and the Key Laboratory of Mathematical Economics and Quantitative Finance (Peking University), Ministry of Education.

Zhenhua Lin’s research is partially supported by NUS startup Grant R-155-000-217-133.

Data collection and sharing for this project was funded by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012).

SUPPLEMENTARY MATERIAL

Supplement to “Intrinsic Riemannian functional data analysis for sparse longitudinal observations” (DOI: [10.1214/22-AOS2172SUPP](https://doi.org/10.1214/22-AOS2172SUPP); .pdf). The supplementary material (Shao, Lin and Yao (2022)) contains some preliminaries for Riemannian geometry, the asymptotic distribution of the proposed covariance estimator, proofs, theoretical results for the regular design and further illustrations of the invariance property. The code and data are hosted at <https://github.com/linulysses/iRFDA-sparse>.

REFERENCES

- AFSARI, B. (2011). Riemannian L^p center of mass: Existence, uniqueness, and convexity. *Proc. Amer. Math. Soc.* **139** 655–673. [MR2736346 https://doi.org/10.1090/S0002-9939-2010-10541-5](https://doi.org/10.1090/S0002-9939-2010-10541-5)
- ANEIROS, G., CAO, R., FRAIMAN, R., GENEST, C. and VIEU, P. (2019). Recent advances in functional data analysis and high-dimensional statistics. *J. Multivariate Anal.* **170** 3–9. [MR3913024 https://doi.org/10.1016/j.jmva.2018.11.007](https://doi.org/10.1016/j.jmva.2018.11.007)
- ARSIGNY, V., FILLARD, P., PENNEC, X. and AYACHE, N. (2006). Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Magn. Reson. Med.* **56** 411–421.
- ARSIGNY, V., FILLARD, P., PENNEC, X. and AYACHE, N. (2006/07). Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.* **29** 328–347. [MR2288028 https://doi.org/10.1137/050637996](https://doi.org/10.1137/050637996)
- BHATTACHARYA, R. and PATRANGENARU, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *Ann. Statist.* **31** 1–29. [MR1962498 https://doi.org/10.1214/aos/1046294456](https://doi.org/10.1214/aos/1046294456)
- BHATTACHARYA, R. and PATRANGENARU, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds. II. *Ann. Statist.* **33** 1225–1259. [MR2195634 https://doi.org/10.1214/009053605000000093](https://doi.org/10.1214/009053605000000093)
- CAI, T. and YUAN, M. (2010). Nonparametric covariance function estimation for functional and longitudinal data Technical Report Univ. Pennsylvania.
- CAI, T. T. and YUAN, M. (2011). Optimal estimation of the mean function based on discretely sampled functional data: Phase transition. *Ann. Statist.* **39** 2330–2355. [MR2906870 https://doi.org/10.1214/11-AOS898](https://doi.org/10.1214/11-AOS898)
- CORNEA, E., ZHU, H., KIM, P. and IBRAHIM, J. G. (2017). Regression models on Riemannian symmetric spaces. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 463–482. [MR3611755 https://doi.org/10.1111/rssb.12169](https://doi.org/10.1111/rssb.12169)
- DAI, X., LIN, Z. and MÜLLER, H.-G. (2021). Modeling sparse longitudinal data on Riemannian manifolds. *Biometrics* **77** 1328–1341. [MR4357841 https://doi.org/10.1111/biom.13385](https://doi.org/10.1111/biom.13385)
- DAI, X. and MÜLLER, H.-G. (2018). Principal component analysis for functional data on Riemannian manifolds and spheres. *Ann. Statist.* **46** 3334–3361. [MR3852654 https://doi.org/10.1214/17-AOS1660](https://doi.org/10.1214/17-AOS1660)
- DELAIGLE, A., HALL, P., HUANG, W. and KNEIP, A. (2021). Estimating the covariance of fragmented and other related types of functional data. *J. Amer. Statist. Assoc.* **116** 1383–1401. [MR4309280 https://doi.org/10.1080/01621459.2020.1723597](https://doi.org/10.1080/01621459.2020.1723597)
- DESCARY, M.-H. and PANARETOS, V. M. (2019). Recovering covariance from functional fragments. *Biometrika* **106** 145–160. [MR3912388 https://doi.org/10.1093/biomet/asy055](https://doi.org/10.1093/biomet/asy055)
- 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. [MR2750388 https://doi.org/10.1214/09-AOAS249](https://doi.org/10.1214/09-AOAS249)

- DUBEY, P. and MÜLLER, H.-G. (2020). Functional models for time-varying random objects. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 275–327. [MR4084166](#)
- FARAWAY, J. J. (2014). Regression for non-Euclidean data using distance matrices. *J. Appl. Stat.* **41** 2342–2357. [MR3256391](#) <https://doi.org/10.1080/02664763.2014.909794>
- FERRATY, F. and VIEU, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer Series in Statistics. Springer, New York. [MR2229687](#)
- FILLARD, P., ARSIGNY, V., AYACHE, N. and PENNEC, X. (2005). A Riemannian framework for the processing of tensor-valued images. In *Deep Structure, Singularities, and Computer Vision* (O. Fogh Olsen, L. Florack and A. Kuijper, eds.) 112–123. Springer, Heidelberg.
- FLETCHER, P. T. (2013). Geodesic regression and the theory of least squares on Riemannian manifolds. *Int. J. Comput. Vis.* **105** 171–185. [MR3104017](#) <https://doi.org/10.1007/s11263-012-0591-y>
- FLETCHER, T. and JOSHI, S. (2007). Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Process.* **87** 250–262.
- HALL, P., MÜLLER, H.-G. and WANG, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.* **34** 1493–1517. [MR2278365](#) <https://doi.org/10.1214/009053606000000272>
- HEIN, M. (2009). Robust nonparametric regression with metric-space valued output. In *Advances in Neural Information Processing Systems* 718–726.
- HINKLE, J., FLETCHER, P. T. and JOSHI, S. (2014). Intrinsic polynomials for regression on Riemannian manifolds. *J. Math. Imaging Vision* **50** 32–52. [MR3233133](#) <https://doi.org/10.1007/s10851-013-0489-5>
- HSING, T. and EUBANK, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley Series in Probability and Statistics. Wiley, Chichester. [MR3379106](#) <https://doi.org/10.1002/9781118762547>
- KOKOSZKA, P. and REIMHERR, M. (2017). *Introduction to Functional Data Analysis*. Texts in Statistical Science Series. CRC Press, Boca Raton, FL. [MR3793167](#)
- LEE, J. M. (1997). *Riemannian Manifolds: An Introduction to Curvature*. Graduate Texts in Mathematics **176**. Springer, New York. [MR1468735](#) <https://doi.org/10.1007/b98852>
- LENGLET, C., ROUSSON, M., DERICHE, R. and FAUGERAS, O. (2006). Statistics on the manifold of multivariate normal distributions: Theory and application to diffusion tensor MRI processing. *J. Math. Imaging Vision* **25** 423–444. [MR2283616](#) <https://doi.org/10.1007/s10851-006-6897-z>
- LI, Y. and HSING, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Ann. Statist.* **38** 3321–3351. [MR2766854](#) <https://doi.org/10.1214/10-AOS813>
- LIN, Z. (2019). Riemannian geometry of symmetric positive definite matrices via Cholesky decomposition. *SIAM J. Matrix Anal. Appl.* **40** 1353–1370. [MR4032859](#) <https://doi.org/10.1137/18M1221084>
- LIN, Z. and MÜLLER, H.-G. (2021). Total variation regularized Fréchet regression for metric-space valued data. *Ann. Statist.* **49** 3510–3533. [MR4352539](#) <https://doi.org/10.1214/21-aos2095>
- LIN, Z., WANG, J.-L. and ZHONG, Q. (2021). Basis expansions for functional snippets. *Biometrika* **108** 709–726. [MR4298773](#) <https://doi.org/10.1093/biomet/asaa088>
- LIN, Z. and YAO, F. (2019). Intrinsic Riemannian functional data analysis. *Ann. Statist.* **47** 3533–3577. [MR4025751](#) <https://doi.org/10.1214/18-AOS1787>
- LINDBERG, O., WALTERFANG, M., LOOI, J. C. L., MALYKHIN, N., ÖSTBERG, P., ZANDBELT, B., STYNER, M., VELAKOULIS, D., ÖRNDAHL, E. et al. (2012). Shape analysis of the hippocampus in Alzheimer’s disease and subtypes of frontotemporal lobar degeneration. *J. Alzheimer’s Dis.* **30** 355–365.
- MOAKHER, M. (2005). A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.* **26** 735–747. [MR2137480](#) <https://doi.org/10.1137/S0895479803436937>
- PELLETIER, B. (2006). Non-parametric regression estimation on closed Riemannian manifolds. *J. Nonparametr. Stat.* **18** 57–67. [MR2214065](#) <https://doi.org/10.1080/10485250500504828>
- PENNEC, X. (2019). Curvature effects on the empirical mean in Riemannian and affine manifolds: A non-asymptotic high concentration expansion in the small-sample regime. Available at [arXiv:arxiv.org/1906.07418](https://arxiv.org/1906.07418).
- PENNEC, X. (2020). Manifold-valued image processing with SPD matrices. In *Riemannian Geometric Statistics in Medical Image Analysis* 75–134. Elsevier, London. [MR4254214](#) <https://doi.org/10.1016/B978-0-12-814725-2.00010-8>
- PENNEC, X., FILLARD, P. and AYACHE, N. (2006). A Riemannian framework for tensor computing. *Int. J. Comput. Vis.* **66** 41–66. <https://doi.org/10.1007/s11263-005-3222-z>
- PETERSEN, A. and MÜLLER, H.-G. (2019). Fréchet regression for random objects with Euclidean predictors. *Ann. Statist.* **47** 691–719. [MR3909947](#) <https://doi.org/10.1214/17-AOS1624>
- PRÉVÔT, C. and RÖCKNER, M. (2007). *A Concise Course on Stochastic Partial Differential Equations*. Lecture Notes in Math. **1905**. Springer, Berlin. [MR2329435](#)

- RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2168993
- SCHÖTZ, C. (2019). Convergence rates for the generalized Fréchet mean via the quadruple inequality. *Electron. J. Stat.* **13** 4280–4345. MR4023955 <https://doi.org/10.1214/19-EJS1618>
- SHAO, L., LIN, Z. and YAO, F. (2022). Supplement to “Intrinsic Riemannian functional data analysis for sparse longitudinal observations.” <https://doi.org/10.1214/22-AOS2172SUPP>
- SHI, X., STYNER, M., LIEBERMAN, J., IBRAHIM, J. G., LIN, W. and ZHU, H. (2009). Intrinsic regression models for manifold-valued data. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 12* 192–199.
- STEINKE, F., HEIN, M. and SCHÖLKOPF, B. (2010). Nonparametric regression between general Riemannian manifolds. *SIAM J. Imaging Sci.* **3** 527–563. MR2736019 <https://doi.org/10.1137/080744189>
- STURM, K.-T. (2003). Probability measures on metric spaces of nonpositive curvature. In *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces (Paris, 2002)*. *Contemp. Math.* **338** 357–390. Amer. Math. Soc., Providence, RI. MR2039961 <https://doi.org/10.1090/conm/338/06080>
- SU, J., KURTEK, S., KLASSEN, E. and SRIVASTAVA, A. (2014). Statistical analysis of trajectories on Riemannian manifolds: Bird migration, hurricane tracking and video surveillance. *Ann. Appl. Stat.* **8** 530–552. MR3192001 <https://doi.org/10.1214/13-AOAS701>
- WANG, J.-L., CHIOU, J.-M. and MÜLLER, H.-G. (2016). Review of functional data analysis. *Annu. Rev. Stat. Appl.* **3** 257–295.
- YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100** 577–590. MR2160561 <https://doi.org/10.1198/016214504000001745>
- ZHANG, Z., KLASSEN, E. and SRIVASTAVA, A. (2018). Phase-amplitude separation and modeling of spherical trajectories. *J. Comput. Graph. Statist.* **27** 85–97. MR3788303 <https://doi.org/10.1080/10618600.2017.1340892>
- ZHANG, X. and WANG, J.-L. (2016). From sparse to dense functional data and beyond. *Ann. Statist.* **44** 2281–2321. MR3546451 <https://doi.org/10.1214/16-AOS1446>
- ZHANG, X. and WANG, J.-L. (2018). Optimal weighting schemes for longitudinal and functional data. *Statist. Probab. Lett.* **138** 165–170. MR3788733 <https://doi.org/10.1016/j.spl.2018.03.007>
- ZHU, H., CHEN, Y., IBRAHIM, J. G., LI, Y., HALL, C. and LIN, W. (2009). Intrinsic regression models for positive-definite matrices with applications to diffusion tensor imaging. *J. Amer. Statist. Assoc.* **104** 1203–1212. MR2750245 <https://doi.org/10.1198/jasa.2009.tm08096>