The Annals of Statistics
2019, V l. 47, N . 6, 3533 3577
h ://d i. g/10.1214/18-AOS1787

INTRINSIC RIEMANNIAN FUNCTIONAL DATA ANALYSIS[1]

BY ZHENHUA LIN* AND FANG YAO

National University of Singapore and Peking University*

In thi k e de el a n el and f ndati nal f am k f ana- l ing gene al Riemannian f ncti nal data, in a tic la a ne de el ment f ten Hilbe t ace al ng c e n a manif ld. S ch ace enable t de i e Ka h nen L e e an i n f Riemannian and m ce e . Thi f am k al feat e an a ach t c m a e bject f m diffe ent ten Hilbe t ace , hich a e the a f a m t tic anal i in Riemannian f ncti nal data anal i . B il n int in ic ge met ic c nce t ch a ec- t eld, Le i-Ci ita c nnecti n and a allel t an t n Riemannian man- if ld , the de el ed f am k a lie t n t nl E clidean bmanif ld b t al manif ld ith t a nat al ambient ace. A a licati n f thi f am k, e de el int in ic Riemannian f ncti nal inci al c m nent anal i (iRFPCA) and int in ic Riemannian f ncti nal linea eg e i n (iR- FLR) that a e di tinct f m thei t aditi nal and ambient c nte a t . We al ide e timati n ced e f iRFPCA and iRFLR, and in e tigate thei a m t tic e tie ithin the int in ic ge met . N me ical e f mance i ill t ated b im lated and eal e am le .

1. Introduction. F ncti nal data anal i (FDA) ad ance b tantiall in the a t decade , a the a id de el ment f m de n techn l g enable c llecting m e and m e data c ntin l e time. The e i ich lite at e anning m e than e ent ea n thi t ic, incl ding de el ment n f nc- ti nal inci al c m nent anal i ch a Da i , P e and R main (1982), Hall and H eini-Na ab (2006), Kleffe (1973), Ra (1958), Sil e man (1996), Ya , M lle and Wang (2005a), Zhang and Wang (2016), and ad ance n f nc- ti nal linea eg e i n ch a Hall and H it (2007), K ng et al. (2016), Ya , M lle and Wang (2005b), Y an and Cai (2010), am ng man the . F a th gh e ie f the t ic, e efe eade t the e ie a ticle Wang, Chi

Recei ed Oct be 2017; e i ed Oct be 2018.

[1]Fang Ya ' e ea ch i a tiall ted b Nati nal Nat al Science F ndati n f China G ant 11871080, a Di ci line C n t cti n F nd at Peking Uni e it and Ke Lab at f Mathematical Ec n mic and Q antitati e Finance (Peking Uni e it), Mini t f Ed cati n. Data e e ided b the H man C nnect me P ject, WU-Minn C n ti m (P inci al In e tigat : Da id Van E en and Kamil Ug bil; 1U54MH091657) f nded b the 16 NIH In tit te and Cente that t the NIH Bl e int f Ne cience Re ea ch, and b the McD nnell Cente f S tem Ne cience at Wa hingt n Uni e it .

MSC2010 subject classifications. 62G05, 62J05.

Key words and phrases. F ncti nal inci al c m nent, f ncti nal linea eg e i n, int in ic Riemannian Ka h nen L e e an i n, a allel t an t, ten Hilbe t ace.

and Müller (2016) and monographs Ferraty and Vieu (2006), Hsing and Eubank (2015), Kokoszka and Reimherr (2017), Ramsay and Silverman (2005) for comprehensive treatments on classic functional data analysis. Although traditionally functional data take values in a vector space, more data of nonlinear structure arise and should be properly handled in a nonlinear space. For instance, trajectories of bird migration are naturally regarded as curves on a sphere which is a nonlinear Riemannian manifold, rather than the three-dimensional vector space $\mathbb{R}^3$. Another example is the dynamics of brain functional connectivity. The functional connectivity at a time point is represented by a symmetric positive-definite matrix (SPD). Then the dynamics shall be modeled as a curve in the space of SPDs that is endowed with either the affine-invariant metric (Moakher (2005)) or the Log-Euclidean metric (Arsigny et al. (2006/07)) to avoid the swelling effect (Arsigny et al. (2006/07)). Both metrics turn SPD into a nonlinear Riemannian manifold. In this paper, we refer this type of functional data as *Riemannian functional data*, which are functions taking values on a Riemannian manifold and modeled by *Riemannian random processes*, that is, we treat Riemannian trajectories as realizations of a Riemannian random process.

Analysis of Riemannian functional data is not only challenged by the infinite dimensionality and compactness of covariance operator from functional data, but also obstructed by the *nonlinearity* of the range of functions, since manifolds are generally not vector spaces and render many techniques relying on linear structure ineffective or inapplicable. For instance, if the sample mean curve is computed for bird migration trajectories as if they were sampled from the ambient space $\mathbb{R}^3$, this naive sample mean in general does not fall on the sphere of earth. For manifolds of tree-structured data studied in Wang and Marron (2007), as they are naturally not Euclidean submanifolds which refer to Riemannian submanifolds of a Euclidean space in this paper, the naive sample mean can not even be defined from ambient spaces, and thus a proper treatment of manifold structure is necessary. While the literature for Euclidean functional data is abundant, works involving nonlinear manifold structure are scarce. Chen and Müller (2012) and Lin and Yao (2019) respectively investigate representation and regression for functional data living in a low-dimensional nonlinear manifold that is embedded in an infinite-dimensional space, while Lila, Aston and Sangalli (2016) focuses principal component analysis on functional data whose domain is a two-dimensional manifold. None of these deal with functional data that take values on a nonlinear manifold, while Dai and Müller (2018) is the only endeavor in this direction for Euclidean submanifolds.

As functional principal component analysis (FPCA) is an essential tool for FDA, it is of importance and interest to develop this notion for Riemannian functional data. Since manifolds are in general not vector spaces, classic covariance functions/operators do not exist naturally for a Riemannian random process. A strategy that is often adopted, for example, Shi et al. (2009) and Cornea et al. (2017), to overcome the lack of vectorial structure is to map data on the manifold into tangent spaces via Riemannian logarithm map defined in Section 2.2. As tangent spaces at different points are different vector spaces, in order to handle observations from different tangent spaces, some existing works assume a Euclidean ambient space for the manifold and identify tangent vectors as Euclidean vectors. This strategy is adopted by Dai and Müller (2018) on Riemannian functional data such as compositional data modeled on the unit sphere for the first time. Specifically, they assume that functional data are sampled from a time-varying geodesic submanifold, where at a given time point, the functions take values on a geodesic submanifold of a common manifold. Such a common manifold is further assumed to be a Euclidean submanifold that allows to identify all tangent spaces as hyperplanes in a common Euclidean space (endowed with the usual Euclidean inner product). Then, with the aid of Riemannian logarithm map, Dai and Müller (2018) are able to transform Riemannian functional data into Euclidean ones while accounting for the intrinsic curvature of the underlying manifold.

To avoid confusion, we distinguish two different perspectives to deal with Riemannian manifolds. One is to work with the manifold under consideration without assuming an ambient space surrounding it or an isometric embedding into a Euclidean space. This perspective is regarded as *completely intrinsic*, or simply *intrinsic*. Although generally difficult to work with, it can fully respect all geometric structure of the manifold. The other one, referred to as *ambient* here, assumes that the manifold under consideration is isometrically embedded in a Euclidean ambient space, so that geometric objects such as tangent vectors can be processed within the ambient space. For example, from this point of view, the local polynomial regression for SPD proposed by Yuan et al. (2012) is intrinsic, while the aforementioned work by Dai and Müller (2018) takes the ambient perspective.

Although it is possible to account for some of geometric structure in the ambient perspective, for example, the curved nature of manifold via Riemannian logarithm map, several issues arise due to manipulation of geometric objects such as tangent vectors in the ambient space. First, the essential dependence on an ambient space restricts potential applications. It is not immediately applicable to manifolds that are not a Euclidean submanifold or do not have a natural isometric embedding into a Euclidean space, for example, the Riemannian manifold of $p \times p$ ($p \geq 2$) SPD matrices endowed with the affine-invariant metric (Moakher (2005)) which is not compatible with the $p(p+1)/2$-dimensional Euclidean metric. Second, although an ambient space provides a common stage for tangent vectors at different points, operation on tangent vectors from this ambient perspective can potentially violate the intrinsic geometry of the manifold. To illustrate this, consider comparison of two tangent vectors at different points (this comparison is needed in the asymptotic analysis of Section 3.2; see also Section 2.4). From the ambient perspective, taking the difference of tangent vectors requires moving a tangent vector parallelly *within the ambient space* to the base point of the other tangent vector. However, the resultant tangent vector after movement in the ambient space is generally not a tangent vector for the base point of the other tangent vector; see the left panel of Figure 1 for a geometric illustration. In another word, the ambient difference

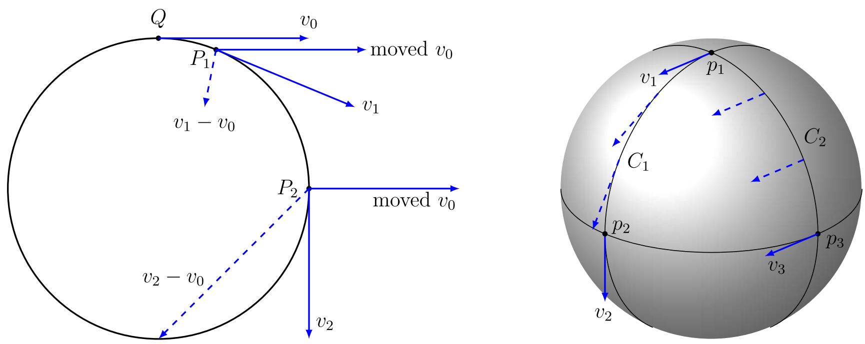


FIG. 1. *Left panel: illustration of ambient movement of tangent vectors. The tangent vector v_0 at the point Q of a unit circle embedded in a Euclidean plane is moved to the point P_1 and P_2 within the ambient space. v_1 (resp. v_2) is a tangent vector at P_1 (resp. P_2). The differences $v_1 - v_0$ and $v_2 - v_0$ are not tangent to the circle at P_1 and P_2, respectively. If v_0, v_1 and v_2 have the same length, then the intrinsic parallel transport of v_0 to P_k shall coincide with v_k, and $\mathcal{P}v_0 - v_k = 0$, where $k = 1, 2$ and $\mathcal{P}$ represents the parallel transport on the unit circle with the canonical metric tensor. Thus, $\|\mathcal{P}v_0 - v_k\|_{\mathbb{R}^2} = 0$. However, $\|v_0 - v_k\|_{\mathbb{R}^2} > 0$, and this nonzero value completely results from the departure of the Euclidean geometry from the unit circle geometry. The ambient discrepancy $\|v_0 - v_1\|_{\mathbb{R}^2}$ is small as P_1 is close to P, while $\|v_0 - v_2\|_{\mathbb{R}^2}$ is large since P_2 is far away from Q. Right panel: illustration of parallel transport. A tangent vector v_1 at the point p_1 on the unit sphere is parallelly transported to the point p_2 and p_3 along curves C_1 and C_2, respectively. During parallel transportation, the transported tangent vector always stays within the tangent spaces along the curve.*

f tangent ect at different int i n t an intr in ic ge metric bject n the manif ld, and the de art re fr m intr in ic ge metr can tentiall affect the tati tical ef cac and/ r ef cienc . La tl , ince manif ld might be embedded int m re than ne ambient ace, the inter retati n f tati tical re lt e eciall de end n the ambient ace and c ld be mi leading if ne d e n t ch e the ambient ace a r riatel .

In the a er, e de el a c m letel intr in ic frame rk that r ide a f n-dati nal the r f r general Riemannian f ncti nal data that a e the a f r the de el ment f intr in ic Riemannian f ncti nal rinci al c m nent anal i and intr in ic Riemannian f ncti nal linear regre i n, am ng ther tential a lica-ti n . The ke b ilding bl ck i a ne c ncet f *tensor Hilbert space* al ng a c r e n the manif ld, hich i de cribed in Secti n 2. On ne hand, r a - r ach e erience dramaticall ele ated technical challenge relati e t the am-bient c nter art . F r e am le, ith t an ambient ace, it i n ntri ial t er-cei e and handle tangent ect r . On the ther hand, the ad antage f the intrin-ic er ecti e are at lea t three f ld, in c ntra t t ambient a r ache . Fir t, r re lt immediatel a l t man im rtant Riemannian manif ld that are n t n at rall a E clidean bmanif ld b t c mm nl een in tati tical anal i and machine learning, ch a the af rementi ned SPD manif ld and Gra mannian manif ld . Sec nd, r frame rk feat re a n el intr in ic r al f r c herent comparison of objects from different tensor Hilbert spaces on the manifold, and hence make the asymptotic analysis sensible. Third, results produced by our approach are invariant to embeddings and ambient spaces, and can be interpreted independently, which avoid potential misleading interpretation in practice.

As important applications of the proposed framework, we develop intrinsic Riemannian functional principal component analysis (iRFPCA) and intrinsic Riemannian functional linear regression (iRFLR). Specifically, estimation procedures of intrinsic eigenstructure are provided and their asymptotic properties are investigated within the intrinsic geometry. For iRFLR, we study a Riemannian functional linear regression model, where a scalar response intrinsically and linearly depends on a Riemannian functional predictor through a *Riemannian slope function*, a concept that is formulated in Section 4, along with the concept of linearity in the context of Riemannian functional data. We present an FPCA-based estimator and a Tikhonov estimator for the Riemannian slope function and explore their asymptotic properties, where the proposed framework of tensor Hilbert space again plays an essential role.

The rest of the paper is structured as follows. The foundational framework for intrinsic Riemannian functional data analysis is laid in Section 2. Intrinsic Riemannian functional principal component analysis is presented in Section 3, while intrinsic Riemannian functional regression is studied in Section 4. In Section 5, numerical performance is illustrated through simulations, and an application to Human Connectome Project analyzing functional connectivity and behavioral data is provided.

2. Tensor Hilbert space and Riemannian random process. In this section, we first define the concept of tensor Hilbert space and discuss its properties, including a mechanism to deal with two different tensor Hilbert spaces at the same time. Then, random elements on tensor Hilbert spaces are investigated, with the proposed intrinsic Karhunen–Loève expansion for the random elements. Finally, practical computation with respect to an orthonormal frame is given. Throughout this section, we assume a d-dimensional, connected and geodesically complete Riemannian manifold $\mathcal{M}$ equipped with a Riemannian metric $\langle\cdot,\cdot\rangle$, which defines a scalar product $\langle\cdot,\cdot\rangle_p$ for the tangent space $T_p\mathcal{M}$ at each point $p \in \mathcal{M}$. This metric also induces a distance function $d_{\mathcal{M}}$ on $\mathcal{M}$. A preliminary for Riemannian manifolds can be found in the Appendix. For a comprehensive treatment on Riemannian manifolds, we recommend the introductory text by Lee (1997) and also Lang (1995).

2.1. *Tensor Hilbert spaces along curves.* Let μ be a measurable curve on a manifold $\mathcal{M}$ and parameterized by a compact domain $\mathcal{T} \subset \mathbb{R}$ equipped with a finite measure υ. A vector field V along μ is a map from $\mathcal{T}$ to the tangent bundle $T\mathcal{M}$ such that $V(t) \in T_{\mu(t)}\mathcal{M}$ for all $t \in \mathcal{T}$. It is seen that the collection of vector fields V along μ is a vector space, where the vector addition between two vector field V_1 and V_2 is a vector field U such that $U(t) = V_1(t) + V_2(t)$ for all $t \in \mathcal{T}$, and the scalar multiplication between a real number a and a vector field V is a vector field U such that $U(t) = aV(t)$ for all $t \in \mathcal{T}$. Let $\mathscr{T}(\mu)$ be the collection of (equivalence classes of) measurable vector fields V along μ such that $\|V\|_{\mu} := \{\int_{\mathcal{T}} \langle V(t), V(t)\rangle_{\mu(t)} \, d\upsilon(t)\}^{1/2} < \infty$ with identification between V and U in $\mathscr{T}(\mu)$ (or equivalently, V and U are in the same equivalence class) when $\upsilon(\{t \in \mathcal{T} : V(t) \neq U(t)\}) = 0$. Then $\mathscr{T}(\mu)$ is turned into an inner product space by the inner product $\langle\!\langle V, U \rangle\!\rangle_{\mu} := \int_{\mathcal{T}} \langle V(t), U(t)\rangle_{\mu(t)} \, d\upsilon(t)$, with the induced norm given by $\|\cdot\|_{\mu}$. Moreover, we have that:

THEOREM 1. *For a measurable curve μ on $\mathcal{M}$, $\mathscr{T}(\mu)$ is a separable Hilbert space.*

We call the space $\mathscr{T}(\mu)$ the *tensor Hilbert space* along μ, as tangent vectors are a special type of tensor and the above Hilbertian structure can be defined for tensor fields along μ. The above theorem is of paramount importance, in the sense that it suggests $\mathscr{T}(\mu)$ to serve as a cornerstone for Riemannian functional data analysis for two reasons. First, as shown in Section 2.2, via Riemannian logarithm maps, a Riemannian random process may be transformed into a tangent-vector-valued random process (called log-process in Section 2.2) that can be regarded as a random element in a tensor Hilbert space. Second, the rigorous theory of functional data analysis formulated in Hsing and Eubank (2015) by random elements in separable Hilbert spaces fully applies to the log-process.

One distinct feature of the tensor Hilbert space is that, different curves that are even parameterized by the same domain give rise to different tensor Hilbert spaces. In practice, one often needs to deal with two different tensor Hilbert spaces at the same time. For example, in the next subsection we will see that under some conditions, a Riemannian random process X can be conceived as a random element on the tensor Hilbert space $\mathscr{T}(\mu)$ along the intrinsic mean curve μ. However, the mean curve is often unknown and estimated from a random sample of X. Since the sample mean curve $\hat{\mu}$ generally does not agree with the population one, quantities of interest such as covariance operator and their sample versions are defined on two different tensor Hilbert spaces $\mathscr{T}(\mu)$ and $\mathscr{T}(\hat{\mu})$, respectively. For statistical analysis, one needs to compare the sample quantities with their population counterparts and hence involve objects such as covariance operators from two different tensor Hilbert spaces.

In order to intrinsically quantify the discrepancy between objects of the same kind from different tensor Hilbert spaces, we utilize the Levi-Civita connection (Lee (1997), page 18) associated with the Riemannian manifold $\mathcal{M}$. The Levi-Civita connection is uniquely determined by the Riemannian structure. It is the only torsion-free connection compatible with the Riemannian metric. Associated with this connection is a unique parallel transport operator $\mathcal{P}_{p,q}$ that smoothly transport tangent vectors at p along a curve to q and preserve the inner product. We shall emphasize that the parallel transportation is performed intrinsically. For instance, tangent vectors being transported always stay tangent to the manifold during transportation, which is illustrated by the right panel of Figure 1. Although transport operator $\mathcal{P}_{p,q}$ depends on the curve connecting p and q, there exists a canonical choice of the curve connecting two points, which is the minimizing geodesic between p and q (under some conditions, almost surely the minimizing geodesic is unique between two points randomly sampled from the manifold). The smoothness of parallel transport also implies that if p and q are not far apart, then the initial tangent vector and the transported one stay close (in the space of tangent bundle endowed with the Sasaki metric (Sasaki (1958))). This feature is desirable for our purpose, as when sample mean $\hat{\mu}(t)$ approaches to $\mu(t)$, one expects a tangent vector at $\hat{\mu}(t)$ converges to its transported version at $\mu(t)$. Owing to these nice properties of parallel transport, it becomes an ideal tool to construct a mechanism of comparing objects from different tensor Hilbert spaces as follows.

Suppose f and h are two measurable curves on $\mathcal{M}$ defined on $\mathcal{T}$. Let $\gamma_t(\cdot) := \gamma(t, \cdot)$ be a family of smooth curves that is parameterized by the interval $[0, 1]$ (the way of parameterization here does not matter) and connects $f(t)$ to $h(t)$, that is, $\gamma_t(0) = f(t)$ and $\gamma_t(1) = h(t)$, such that $\gamma(\cdot, s)$ is measurable for all $s \in [0, 1]$. Suppose $v \in T_{f(t)}\mathcal{M}$ and let V be a smooth vector field along γ_t such that $\nabla_{\dot{\gamma}} V = 0$ and $V(0) = v$, where ∇ denotes the Levi-Civita connection of the manifold $\mathcal{M}$. The theory of Riemannian manifolds shows that such a vector field V uniquely exists. This gives rise to the parallel transporter $\mathcal{P}_{f(t),h(t)} : T_{f(t)}\mathcal{M} \to T_{h(t)}\mathcal{M}$ along γ_t, defined by $\mathcal{P}_{f(t),h(t)}(v) = V(1)$. In other words, $\mathcal{P}_{f(t),h(t)}$ parallelly transports v to $V(1) \in T_{h(t)}\mathcal{M}$ along the curve γ_t. As the parallel transporter determined by the Levi-Civita connection, $\mathcal{P}$ preserves the inner product of tangent vectors along transportation, that is, $\langle u, v\rangle_{f(t)} = \langle \mathcal{P}_{f(t),h(t)}u, \mathcal{P}_{f(t),h(t)}v\rangle_{h(t)}$ for $u, v \in T_{f(t)}\mathcal{M}$. Then we can define the parallel transport of vector fields from $\mathscr{T}(f)$ to $\mathscr{T}(h)$, denoted by $\Gamma_{f,h}$, $(\Gamma_{f,h}U)(t) = \mathcal{P}_{f(t),h(t)}(U(t))$ for all $U \in \mathscr{T}(f)$ and $t \in \mathcal{T}$. One immediately sees that $\Gamma_{f,h}$ is a linear operator on tensor Hilbert spaces. Its adjoint, denoted by $\Gamma^*_{f,h}$, is a map from $\mathscr{T}(h)$ to $\mathscr{T}(f)$ and is given by $\langle\langle U, \Gamma^*_{f,h}V\rangle\rangle_f = \langle\langle \Gamma_{f,h}U, V\rangle\rangle_h$ for $U \in \mathscr{T}(f)$ and $V \in \mathscr{T}(h)$. Let $\mathscr{C}(f)$ denote the set of all Hilbert–Schmidt operators on $\mathscr{T}(f)$, which is a Hilbert space with the Hilbert–Schmidt norm $|||\cdot|||$

smooth and $\gamma(,s)$ *is measurable. Then the following statements regarding* $\Gamma_{f,h}$ *and* $\Phi_{f,h}$ *hold.*

1. *The operator* $\Gamma_{f,h}$ *is a unitary transformation from* $\mathscr{T}(f)$ *to* $\mathscr{T}(h)$.
2. $\Gamma_{f,h}^{*} = \Gamma_{h,f}$. *Also*, $\|\Gamma_{f,h}U - V\|_h = \|U - \Gamma_{h,f}V\|_f$.
3. $\Gamma_{f,h}(\mathcal{A}U) = (\Phi_{f,h}\mathcal{A})(\Gamma_{f,h}U)$.
4. *If* $\mathcal{A}$ *is invertible, then* $\Phi_{f,h}\mathcal{A}^{-1} = (\Phi_{f,h}\mathcal{A})^{-1}$.
5. $\Phi_{f,h}\sum_k c_k\varphi_k \otimes \varphi_k = \sum_k c_k(\Gamma_{f,h}\varphi_k) \otimes (\Gamma_{f,h}\varphi_k)$, *where* c_k *are scalar constants, and* $\varphi_k \in \mathscr{T}(f)$.
6. $|||\Phi_{f,h}\mathcal{A} - \mathcal{B}|||_h = |||\mathcal{A} - \Phi_{h,f}\mathcal{B}|||_f$.

We define $U \ominus_\Gamma V := \Gamma_{f,h}U - V$ for $U \in \mathscr{T}(f)$ and $V \in \mathscr{T}(h)$, and $\mathcal{A} \ominus_\Phi \mathcal{B} := \Phi_{f,h}\mathcal{A} - \mathcal{B}$ for operators $\mathcal{A}$ and $\mathcal{B}$. To quantify the discrepancy between an element U in $\mathscr{T}(f)$ and another one V in $\mathscr{T}(h)$, we can use the quantity $\|U \ominus_\Gamma V\|_h$. Similarly, we adopt $|||\mathcal{A} \ominus_\Phi \mathcal{B}|||_h$ as a discrepancy measure for two covariance operators $\mathcal{A}$ and $\mathcal{B}$. These quantities are intrinsic as they are built on intrinsic geometric concepts. In light of Proposition 2, they are symmetric under the parallel transport, that is, transporting $\mathcal{A}$ to $\mathcal{B}$ yields the same discrepancy measure as transporting $\mathcal{B}$ to $\mathcal{A}$. We also note that, when $\mathcal{M} = \mathbb{R}^d$, the difference operators $\ominus_\Gamma$ and $\ominus_\Phi$ reduce to their regular vector and operator difference, that is, $U \ominus_\Gamma V$ becomes $U - V$, while $\mathcal{A} \ominus_\Phi \mathcal{B}$ becomes $\mathcal{A} - \mathcal{B}$. Therefore, $\ominus_\Gamma$ and $\ominus_\Phi$ can be viewed as generalizations of their regular vector and operator difference to a Riemannian setting. One shall note that Γ and Φ depend on the choice of the family of curves γ, a canonical choice of which is discussed in Section 3.2.

2.2. *Random elements on tensor Hilbert spaces*. Let X be a Riemannian random process. In order to introduce the concept of intrinsic mean function for X, we define a family of functions indexed by t:

$$F(p,t) = \mathbb{E}d_{\mathcal{M}}^2(X(t),p), \quad p \in \mathcal{M}, t \in \mathcal{T}. \tag{1}$$

For a fixed t, if there exists a unique $q \in \mathcal{M}$ that minimizes $F(p,t)$ over all $p \in \mathcal{M}$, then q is called the intrinsic mean (also called Fréchet mean) at t, denoted by $\mu(t)$, that is,

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

As required for intrinsic analysis, we assume the following condition.

A.1 The intrinsic mean function μ exists.

We refer readers to Bhattacharya and Patrangenaru (2003) and Afsari (2011) for conditions under which the intrinsic mean of a random variable on a general manifold exists and is unique. For example, according to Cartan–Hadamard theorem, if the manifold is simply connected and complete with nonpositive sectional curvature, then intrinsic mean function always exists and is unique as long as for all $t \in \mathcal{T}$, $F(p,t) < \infty$ for some $p \in \mathcal{M}$.

Since $\mathcal{M}$ is geodesically complete, by Hopf–Rinow theorem (Lee (1997), page 108), its exponential map Exp_p at each p is defined on the entire $T_p\mathcal{M}$. As Exp_p might not be injective, in order to define its inverse, we restrict Exp_p to a subset of the tangent space $T_p\mathcal{M}$. Let $\mathrm{Cut}(p)$ denote the set of all tangent vectors $v \in T_p\mathcal{M}$ such that the geodesic $\gamma(t) = \mathrm{Exp}_p(tv)$ fails to be minimizing for $t \in [0, 1+\epsilon)$ for each $\epsilon > 0$. Now, we define Exp_p only on $\mathscr{D}_p := T_p\mathcal{M}\backslash\mathrm{Cut}(p)$. The image of Exp_p, denoted by $\mathrm{Im}(\mathrm{Exp}_p)$, consists of points q in $\mathcal{M}$, such that $q = \mathrm{Exp}_p v$ for some $v \in \mathscr{D}_p$. In this case, the inverse of Exp_p exists and is called Riemannian logarithm map, which is denoted by Log_p and maps q to v. We shall make the following assumption:

A.2 $\Pr\{\forall t \in \mathcal{T} : X(t) \in \mathrm{Im}(\mathrm{Exp}_{\mu(t)})\} = 1$.

Then, $\mathrm{Log}_{\mu(t)} X(t)$ is almost surely defined for all $t \in \mathcal{T}$. The condition is superfluous if $\mathrm{Exp}_{\mu(t)}$ is injective for all t, like the manifold of $m \times m$ SPD endowed with the affine-invariant metric.

In the sequel we shall assume X satisfies conditions A.1 and A.2. An important observation is that the log-process $\{\mathrm{Log}_{\mu(t)} X(t)\}_{t \in \mathcal{T}}$ (denoted by $\mathrm{Log}_\mu X$ for short) is a random vector field along μ. If we assume continuity for the sample paths of X, then the process $\mathrm{Log}_\mu X$ is measurable with respect to the product σ-algebra $\mathscr{B}(\mathcal{T}) \times \mathscr{E}$ and the Borel algebra $\mathscr{B}(T\mathcal{M})$, where $\mathscr{E}$ is the σ-algebra of the probability space. Furthermore, if $\mathbb{E}\|\mathrm{Log}_\mu X\|_\mu^2 < \infty$, then according to Theorem 7.4.2 of Hsing and Eubank (2015), $\mathrm{Log}_\mu X$ can be viewed as a tensor Hilbert space $\mathscr{T}(\mu)$ valued random element. Observing that $\mathbb{E}\mathrm{Log}_\mu X = 0$ according to Theorem 2.1 of Bhattacharya and Patrangenaru (2003), the intrinsic covariance operator for $\mathrm{Log}_\mu X$ is given by $\mathcal{C} = \mathbb{E}(\mathrm{Log}_\mu X \otimes \mathrm{Log}_\mu X)$. This operator is nuclear and self-adjoint. It then admits the following eigendecomposition (Hsing and Eubank (2015), Theorem 7.2.6):

$$\mathcal{C} = \sum_{k=1}^{\infty} \lambda_k \phi_k \otimes \phi_k \tag{2}$$

with eigenvalues $\lambda_1 \geq \lambda_2 \cdots \geq 0$ and orthonormal eigenelements ϕ_k that form a complete orthonormal system for $\mathscr{T}(\mu)$. Also, with probability one, the log-process of X has the following Karhunen–Loève expansion:

$$\mathrm{Log}_\mu X = \sum_{k=1}^{\infty} \xi_k \phi_k \tag{3}$$

with $\xi_k := \langle\!\langle X, \phi_k \rangle\!\rangle_\mu$ being uncorrelated and centered random variables. Therefore, we obtain the intrinsic Riemannian Karhunen–Loève (iRKL) expansion for

X given by

$$X(t) = \operatorname{Exp}_{\mu(t)} \sum_{k=1}^{\infty} \xi_k \phi_k(t). \tag{4}$$

The elements ϕ_k are called intrinsic Riemannian functional principal components (iRFPC), while the variables ξ_k are called intrinsic iRFPC scores. This result is summarized in the following theorem whose proof is already contained in the above derivation and hence omitted. We shall note that the continuity assumption on sample paths can be weakened to piecewise continuity.

THEOREM 3 (Intrinsic Karhunen–Loève representation). *Assume that X satisfies assumptions A.1 and A.2. If sample paths of X are continuous and $\mathbb{E}\|\operatorname{Log}_\mu X\|_\mu^2 < \infty$, then the intrinsic covariance operator $\mathcal{C} = \mathbb{E}(\operatorname{Log}_\mu X \otimes \operatorname{Log}_\mu X)$ of $\operatorname{Log}_\mu X$ admits the decomposition (2), and the random process X admits the representation (4).*

In practice, the series at (4) is truncated at some positive integer K, resulting in a truncated intrinsic Riemannian Karhunen–Loève expansion of X, given by $X_K = \operatorname{Exp}_\mu W_K$ with $W_K = \sum_{k=1}^{K} \xi_k \phi_k$. The quality of the approximation of X_K for X is quantified by $\int_{\mathcal{T}} d_{\mathcal{M}}^2(X(t), X_K(t))\, d\upsilon(t)$, and can be shown by a method similar to Dai and Müller (2018) that if the manifold has nonnegative sectional curvature everywhere, then $\int_{\mathcal{T}} d_{\mathcal{M}}^2(X(t), X_K(t))\, d\upsilon(t) \le \|\operatorname{Log}_\mu X - W_K\|_\mu^2$. For manifolds with negative sectional curvatures, such inequality in general does not hold. However, for Riemannian random process X that almost surely lies in a compact subset of $\mathcal{M}$, the residual $\int_{\mathcal{T}} d_{\mathcal{M}}^2(X(t), X_K(t))\, d\upsilon(t)$ can be still bounded by $\|\operatorname{Log}_\mu X - W_K\|_\mu^2$ up to a scaling constant.

PROPOSITION 4. *Assume that conditions A.1 and A.2 hold, and the sectional curvature of $\mathcal{M}$ is bounded from below by $\kappa \in \mathbb{R}$. Let $\mathcal{K}$ be a subset of $\mathcal{M}$. If $\kappa \ge 0$, we let $\mathcal{K} = \mathcal{M}$, and if $\kappa < 0$, we assume that $\mathcal{K}$ is compact. Then, for some constant $C > 0$, $d_{\mathcal{M}}(P, Q) \le \sqrt{C}|\operatorname{Log}_O P - \operatorname{Log}_O Q|$ for all $O, P, Q \in \mathcal{K}$. Consequently, if $X \in \mathcal{K}$ almost surely, then $\int_{\mathcal{T}} d_{\mathcal{M}}^2(X(t), X_K(t))\, d\upsilon(t) \le C\|\operatorname{Log}_\mu X - W_K\|_\mu^2$.*

2.3. *Computation in orthonormal frames.* In practical computation, one might want to work with specific orthonormal bases for tangent spaces. A choice of orthonormal basis for each tangent space constitutes an orthonormal frame on the manifold. In this section, we study the representation of the intrinsic Riemannian Karhunen–Loève expansion under an orthonormal frame and formulas for change of orthonormal frames.

Let $\mathbf{E} = (E_1, \ldots, E_d)$ be a continuous orthonormal frame, that is, each E_j is a vector field of $\mathcal{M}$ such that $\langle E_j(p), E_j(p)\rangle_p = 1$ and $\langle E_j(p), E_k(p)\rangle_p = 0$ for $j \ne k$ and all $p \in \mathcal{M}$. At each point p, $\{E_1(p), \ldots, E_d(p)\}$ form an orthonormal basis for $T_p\mathcal{M}$. The coordinate of $\mathrm{Log}_{\mu(t)} X(t)$ with respect to $\{E_1(\mu(t)),\dots,E_d(\mu(t))\}$ is denoted by $Z_{\mathbf{E}}(t)$, with the subscript $\mathbf{E}$ indicating its dependence on the frame. The resulting process $Z_{\mathbf{E}}$ is called the $\mathbf{E}$-*coordinate process* of X. Note that $Z_{\mathbf{E}}$ is a regular $\mathbb{R}^d$ valued random process defined on $\mathcal{T}$, and classic theory in Hsing and Eubank (2015) applies to $Z_{\mathbf{E}}$. For example, its $\mathcal{L}^2$-norm is defined by $\|Z_{\mathbf{E}}\|_{\mathcal{L}^2} = \{\mathbb{E}\int_{\mathcal{T}} |Z_{\mathbf{E}}(t)|^2 \,dt\}^{1/2}$, where $|\cdot|$ denotes the canonical norm on $\mathbb{R}^d$. One can show that $\|Z_{\mathbf{E}}\|^2_{\mathcal{L}^2} = \mathbb{E}\|\mathrm{Log}_\mu X\|^2_\mu$. Therefore, if $\mathbb{E}\|\mathrm{Log}_\mu X\|^2_\mu < \infty$, then the covariance function exists and is $d \times d$ matrix-valued, quantified by $\mathcal{C}_{\mathbf{E}}(s,t) = \mathbb{E}\{Z_{\mathbf{E}}(s) Z_{\mathbf{E}}(t)^T\}$ (Balakrishnan (1960), Kelly and Root (1960)), noting that $\mathbb{E} Z_{\mathbf{E}}(t) = 0$ as $\mathbb{E}\,\mathrm{Log}_{\mu(t)} X(t) = 0$ for all $t \in \mathcal{T}$. Also, the vector-valued Mercer's theorem implies the eigendecomposition

$$\mathcal{C}_{\mathbf{E}}(s,t) = \sum_{k=1}^{\infty} \lambda_k \phi_{\mathbf{E},k}(s) \phi^T_{\mathbf{E},k}(t), \tag{5}$$

with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots$ and corresponding eigenfunctions $\phi_{\mathbf{E},k}$. Here, the subscript $\mathbf{E}$ in $\phi_{\mathbf{E},k}$ is to emphasize the dependence on the chosen frame. One can see that $\phi_{\mathbf{E},k}$ is a coordinate representation of ϕ_k, that is, $\phi_k = \phi^T_{\mathbf{E},k}\mathbf{E}$.

The coordinate process $Z_{\mathbf{E}}$ admits the vector-valued Karhunen–Loève expansion

$$Z_{\mathbf{E}}(t) = \sum_{k=1}^{\infty} \xi_k \phi_{\mathbf{E},k}(t) \tag{6}$$

under the assumption of mean square continuity of $Z_{\mathbf{E}}$, according to Theorem 7.3.5 of Hsing and Eubank (2015), where $\xi_k = \int_{\mathcal{T}} Z^T_{\mathbf{E}}(t)\phi_{\mathbf{E},k}(t)\,d\upsilon(t)$. While the covariance function and eigenfunctions of $Z_{\mathbf{E}}$ depend on frames, λ_k and ξ_k in (4) and (6) are not, which justifies the absence of $\mathbf{E}$ in their subscripts and the use of the same notation for eigenvalues and iRFPC scores in (2), (4), (5) and (6). This follows from the formula for change of frames that we shall develop below.

Suppose $\mathbf{A} = (A_1,\dots,A_d)$ is another orthonormal frame. Change from $\mathbf{E}(p) = \{E_1(p),\dots,E_d(p)\}$ to $\mathbf{A}(p) = \{A_1(p),\dots,A_d(p)\}$ can be characterized by a unitary matrix $\mathbf{O}_p$. For example, $\mathbf{A}(t) = \mathbf{O}^T_{\mu(t)}\mathbf{E}(t)$ and hence $Z_{\mathbf{A}}(t) = \mathbf{O}_{\mu(t)} Z_{\mathbf{E}}(t)$ for all t. Then the covariance function of $Z_{\mathbf{A}}$ is given by

$$\begin{aligned} \mathcal{C}_{\mathbf{A}}(s,t) &= \mathbb{E}\{Z_{\mathbf{A}}(s) Z^T_{\mathbf{A}}(t)\} \\ &= \mathbb{E}\{\mathbf{O}_{\mu(s)} Z_{\mathbf{E}}(s) Z^T_{\mathbf{E}}(t) \mathbf{O}^T_{\mu(t)}\} \\ &= \mathbf{O}_{\mu(s)} \mathcal{C}_{\mathbf{E}}(s,t) \mathbf{O}^T_{\mu(t)}, \end{aligned} \tag{7}$$

and consequently,

$$\mathcal{C}_{\mathbf{A}}(s,t) = \sum_{k=1}^{\infty} \lambda_k \{\mathbf{O}_{\mu(s)}\phi_{\mathbf{E},k}(s)\}\{\mathbf{O}_{\mu(t)}\phi_{\mathbf{E},k}(t)\}^T.$$

From the above calculation, we immediately see that λ_k are also eigenvalues of $\mathcal{C}_{\mathbf{A}}$. Moreover, the eigenfunction associated with λ_k for $\mathcal{C}_{\mathbf{A}}$ is given by

(8) $$\phi_{\mathbf{A},k}(t) = \mathbf{O}_{\mu(t)}\phi_{\mathbf{E},k}(t).$$

Also, the variable ξ_k in (4) and (6) is the functional principal component score for $Z_{\mathbf{A}}$ associated with $\phi_{\mathbf{A},k}$, as seen by $\int_{\mathcal{T}} Z_{\mathbf{A}}^T(t)\phi_{\mathbf{A},k}(t)\,\mathrm{d}\upsilon(t) = \int_{\mathcal{T}} Z_{\mathbf{E}}^T(t)\mathbf{O}_{\mu(t)}^T\mathbf{O}_{\mu(t)}\phi_{\mathbf{E},k}(t)\,\mathrm{d}\upsilon(t) = \int_{\mathcal{T}} Z_{\mathbf{E}}^T(t)\phi_{\mathbf{E},k}(t)\,\mathrm{d}\upsilon(t)$. The following proposition summarizes the above results.

PROPOSITION 5 (Invariance principle). *Let X be a $\mathcal{M}$-valued random process satisfying conditions A.1 and A.2. Suppose $\mathbf{E}$ and $\mathbf{A}$ are measurable orthonormal frames that are continuous on a neighborhood of the image of μ, and $Z_{\mathbf{E}}$ denotes the $\mathbf{E}$-coordinate log-process of X. Assume $\mathbf{O}_p$ is the unitary matrix continuously varying with p such that $\mathbf{A}(p) = \mathbf{O}_p^T\mathbf{E}(p)$ for $p \in \mathcal{M}$.*

1. *The $\mathcal{L}^r$-norm of $Z_{\mathbf{E}}$ for $r > 0$, defined by $\|Z_{\mathbf{E}}\|_{\mathcal{L}^r} = \{\mathbb{E} \times \int_{\mathcal{T}} |Z_{\mathbf{E}}(t)|^r\,\mathrm{d}\upsilon(t)\}^{1/r}$, is independent of the choice of frames. In particular, $\|Z_{\mathbf{E}}\|_{\mathcal{L}^2}^2 = \mathbb{E}\|\mathrm{Log}_\mu X\|_\mu^2$ for all orthonormal frames $\mathbf{E}$.*
2. *If $\mathbb{E}\|\mathrm{Log}_\mu X\|_\mu^2 < \infty$, then the covariance function of $Z_{\mathbf{E}}$ exists for all $\mathbf{E}$ and admits decomposition of (5). Also, (2) and (5) are related by $\phi_k(t) = \phi_{\mathbf{E},k}^T(t)\mathbf{E}(\mu(t))$ for all t, and the eigenvalues λ_k coincide. Furthermore, the eigenvalues of $\mathcal{C}_{\mathbf{E}}$ and the principal component scores of Karhunen–Loève expansion of $Z_{\mathbf{E}}$ do not depend on $\mathbf{E}$.*
3. *The covariance functions $\mathcal{C}_{\mathbf{A}}$ and $\mathcal{C}_{\mathbf{E}}$ of respectively $Z_{\mathbf{A}}$ and $Z_{\mathbf{E}}$, if they exist, are related by (7). Furthermore, their eigendecompositions are related by (8) and $Z_{\mathbf{A}}(t) = \mathbf{O}_{\mu(t)}Z_{\mathbf{E}}(t)$ for all $t \in \mathcal{T}$.*
4. *If $\mathbb{E}\|\mathrm{Log}_\mu X\|_\mu^2 < \infty$ and sample paths of X are continuous, then the scores ξ_k (6) coincide with the iRFPC scores in (4).*

We conclude this subsection by emphasizing that the concept of covariance function of the log-process depends on the frame $\mathbf{E}$, while the covariance operator, eigenvalues, eigenelements and iRFPC scores do not. In particular, the scores ξ_k, which are often the input for further statistical analysis such as regression and classification, are invariant to the choice of coordinate frames. An important consequence of the invariance principle is that, the scores can be safely computed in any convenient coordinate frame without altering the subsequent analysis.

2.4. *Connection to the special case of Euclidean submanifolds.* Our framework applies to general manifolds that include Euclidean submanifolds as special examples to which the methodology of Dai and Müller (2018) also applies. When the underlying manifold is a d-dimensional submanifold of the Euclidean space $\mathbb{R}^{d_0}$ with $d < d_0$, we recall that the tangent space at each point is identified as a d-dimensional linear subspace of $\mathbb{R}^{d_0}$. For such Euclidean manifolds, Dai and Müller

(2018) treat the log-process of X as a $\mathbb{R}^{d_0}$-valued random process, and derive the representation for the log-process (equation (5) in their paper) within the ambient Euclidean space. This is distinctly different from our intrinsic representation (3) based on the theory of tensor Hilbert spaces, despite their similar appearance. For instance, equation (5) in their work can only be defined for Euclidean submanifolds, while ours is applicable to general Riemannian manifolds. Similarly, the covariance function defined in Dai and Müller (2018), denoted by $C^{\mathrm{DM}}(s,t)$, is associated with the ambient log-process $V(t) \in \mathbb{R}^{d_0}$, that is, $C^{\mathrm{DM}}(s,t) = \mathbb{E}V(s)^T V(t)$. Such an ambient covariance function can only be defined for Euclidean submanifolds but not to general manifolds.

Nevertheless, there are connections between the ambient method of Dai and Müller (2018) and our framework when $\mathcal{M}$ is a Euclidean submanifold. For instance, the mean curves are intrinsically defined in the same way in both works. For the covariance structure, although our covariance function $\mathcal{C}_{\mathbf{E}}$ is a $d \times d$ matrix-valued function while $C^{\mathrm{DM}}(s,t)$ is a $d_0 \times d_0$ matrix-valued function, they both represent the intrinsic covariance operator when $\mathcal{M}$ is a Euclidean submanifold. To see this, first, we observe that the ambient log-process $V(t)$ as defined in Dai and Müller (2018) at the time t, although is ambiently d_0-dimensional, lives in a d-dimensional linear subspace of $\mathbb{R}^{d_0}$. Second, the orthonormal basis $\mathbf{E}(t)$ for the tangent space at $\mu(t)$ can be realized by a $d_0 \times d$ full-rank matrix $\mathbf{G}_t$ by concatenating vectors $E_1(\mu(t)), \ldots, E_d(\mu(t))$. Then $U(t) = \mathbf{G}_t^T V(t)$ is the $\mathbf{E}$-coordinate process of X. This implies that $\mathcal{C}_{\mathbf{E}}(s,t) = \mathbf{G}_s^T C^{\mathrm{DM}}(s,t)\mathbf{G}_t$. On the other hand, since $V(t) = \mathbf{G}_t U(t)$, one has $C^{\mathrm{DM}}(s,t) = \mathbf{G}_t \mathcal{C}_{\mathbf{E}}(s,t)\mathbf{G}_t^T$. Thus, $\mathcal{C}_{\mathbf{E}}$ and C^{DM} determine each other and represent the same object. In light of this observation and the invariance principle stated in Proposition 5, when $\mathcal{M}$ is a Euclidean submanifold, C^{DM} can be viewed as the ambient representation of the intrinsic covariance operator $\mathcal{C}$, while $\mathcal{C}_{\mathbf{E}}$ is the coordinate representation of $\mathcal{C}$ with respect to the frame $\mathbf{E}$. Similarly, the eigenfunctions ϕ_k^{DM} of C^{DM} are the ambient representation of the eigenelements ϕ_k of $\mathcal{C}$. The above reasoning also applies to sample mean function and sample covariance structure. Specifically, when $\mathcal{M}$ is a Euclidean submanifold, our estimator for the mean function discussed in Section 3 is identical to the one in Dai and Müller (2018), while the estimators for the covariance function and eigenfunctions proposed in Dai and Müller (2018) are the ambient representations of our estimators stated in Section 3.

However, when quantifying the discrepancy between the population covariance structure and its estimator, Dai and Müller (2018) adopt the Euclidean difference as a measure. For instance, they use $\hat{\phi}_k^{\mathrm{DM}} - \phi_k^{\mathrm{DM}}$ to represent the discrepancy between the sample eigenfunctions and the population eigenfunctions, where $\hat{\phi}_k^{\mathrm{DM}}$ is the sample version of ϕ_k^{DM}. When $\hat{\mu}(t)$, the sample version of $\mu(t)$, is not equal to $\mu(t)$, $\hat{\phi}_k^{\mathrm{DM}}(t)$ and $\phi_k^{\mathrm{DM}}(t)$ belong to different tangent spaces. In such case, the Euclidean difference $\hat{\phi}_k^{\mathrm{DM}} - \phi_k^{\mathrm{DM}}$ is a Euclidean vector that does not belong to the tangent space at either $\hat{\mu}(t)$ or $\mu(t)$, as illustrated in the left panel of Figure 1. In other words, the Euclidean difference of ambient eigenfunctions does not to be the geometry of the manifold, hence might not serve well to measure the intrinsic discrepancy. In particular, the measure $\|\hat{\phi}_k^{\mathrm{DM}}-\phi_k^{\mathrm{DM}}\|_{\mathbb{R}^{d_0}}$ might completely result from the departure of the ambient Euclidean geometry from the manifold, rather than the intrinsic discrepancy between the sample and population eigenfunctions, as demonstrated in the left panel of Figure 1. Similar reasoning applies to $\hat{C}^{\mathrm{DM}}-C^{\mathrm{DM}}$. In contrast, we base on Proposition 2 to propose an intrinsic measure to characterize the intrinsic discrepancy between a population quantity and its estimator in Section 3.2.

3. Intrinsic Riemannian functional principal component analysis.

3.1. *Model and estimation.* Suppose X admits the intrinsic Riemannian Karhunen–Loève expansion (4), and $X_1,\dots,X_n$ are a random sample of X. In the sequel, we assume that trajectories X_i are fully observed. In the case that data are densely observed, each trajectory can be individually interpolated by using regression techniques for manifold valued data, such as Steinke, Hein and Schölkopf (2010), Cornea et al. (2017) and Petersen and Müller (2019). This way the densely observed data could be represented by their interpolated surrogates, and thus treated as if they were fully observed curves. When data are sparse, delicate information pooling of observations across different subjects is required. The development of such methods is substantial and beyond the scope of this paper.

In order to estimate the mean function μ, we define the finite-sample version of F in (1) by

$$F_n(p,t)=\frac{1}{n}\sum_{i=1}^{n}d_{\mathcal{M}}^2(X_i(t),p).$$

Then, an estimator for μ is given by

$$\hat{\mu}(t)=\arg\min_{p\in\mathcal{M}}F_n(p,t).$$

The computation of $\hat{\mu}$ depends on the Riemannian structure of the manifold. Readers are referred to Cheng et al. (2016) and Salehian et al. (2015) for practical algorithms. For a subset A of $\mathcal{M}$, A^{ϵ} denotes the set $\bigcup_{p\in A}B(p;\epsilon)$, where $B(p;\epsilon)$ is the ball with center p and radius ϵ in $\mathcal{M}$. We use $\mathrm{Im}^{-\epsilon}(\mathrm{Exp}_{\mu(t)})$ to denote the set $\mathcal{M}\backslash\{\mathcal{M}\backslash\mathrm{Im}(\mathrm{Exp}_{\mu(t)})\}^{\epsilon}$. In order to define $\mathrm{Log}_{\hat{\mu}}X_i$, at least with a dominant probability for a large sample, we shall assume a slightly stronger condition than A.2:

A.2′ There is some constant $\epsilon_0>0$ such that $\Pr\{\forall t\in\mathcal{T}: X(t)\in\mathrm{Im}^{-\epsilon_0}(\mathrm{Exp}_{\mu(t)})\}=1$.

Then, combining the fact $\sup_t|\hat{\mu}(t)-\mu(t)|=o_{a.s.}(1)$ that we will show later, we conclude that for a large sample, almost surely, $\mathrm{Im}^{-\epsilon}(\mathrm{Exp}_{\mu(t)})\subset\mathrm{Im}(\mathrm{Exp}_{\hat{\mu}(t)})$ for all $t\in\mathcal{T}$. Therefore, under this condition, $\mathrm{Log}_{\hat{\mu}(t)}X_i(t)$ is well-defined almost surely for a large sample.

The intrinsic Riemannian covariance operator is estimated by its finite-sample version

$$\hat{\mathcal{C}} = \frac{1}{n}\sum_{i=1}^{n}(\mathrm{Log}_{\hat{\mu}} X_i) \otimes \mathrm{Log}_{\hat{\mu}} X_i).$$

This sample intrinsic Riemannian covariance operator also admits an intrinsic eigendecomposition $\hat{\mathcal{C}} = \sum_{k=1}^{\infty} \hat{\lambda}_k \hat{\phi}_k \otimes \hat{\phi}_k$ for $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq 0$. Therefore, the estimates for the eigenvalues λ_k are given by $\hat{\lambda}_k$, while the estimates for ϕ_k are given by $\hat{\phi}_k$. These estimates can also be conveniently obtained under a frame, due to the invariance principle stated in Proposition 5. Let $\mathbf{E}$ be a chosen orthonormal frame and $\hat{\mathcal{C}}_{\mathbf{E}}$ be the sample covariance function based on $\hat{Z}_{\mathbf{E},1}, \ldots, \hat{Z}_{\mathbf{E},n}$, where $\hat{Z}_{\mathbf{E},i}$ is the coordinate process of $\mathrm{Log}_{\hat{\mu}(t)} X_i(t)$ under the frame $\mathbf{E}$ with respect to $\hat{\mu}$. We can then obtain the eigendecomposition $\hat{\mathcal{C}}_{\mathbf{E}}(s,t) = \sum_{k=1}^{\infty} \hat{\lambda}_k \hat{\phi}_{\mathbf{E},k}(s) \hat{\phi}_{\mathbf{E},k}(t)^T$, which yields $\hat{\phi}_k(t) = \hat{\phi}_{\mathbf{E},k}^T(t)\mathbf{E}(t)$ for $t \in \mathcal{T}$. Finally, the truncated process for X_i is estimated by

$$\hat{X}_i^{(K)} = \mathrm{Exp}_{\hat{\mu}} \sum_{k=1}^{K} \hat{\xi}_{ik} \hat{\phi}_k, \tag{9}$$

where $\hat{\xi}_{ik} = \langle\!\langle \mathrm{Log}_{\hat{\mu}} X_i, \hat{\phi}_k \rangle\!\rangle_{\hat{\mu}}$ are estimated iRFPC scores. The above truncated iRKL expansion can be regarded as generalization of the representation (10) in Dai and Müller (2018) from Euclidean submanifolds to general Riemannian manifolds.

3.2. *Asymptotic properties*. To quantify the difference between $\hat{\mu}$ and μ, it is natural to consider the square geodesic distance $d_{\mathcal{M}}(\hat{\mu}(t), \mu(t))$ as a measure of discrepancy. For the asymptotic properties of $\hat{\mu}$, we need the following regularity conditions.

B.1 The manifold $\mathcal{M}$ is connected and complete. In addition, the exponential map $\mathrm{Exp}_p : T_p\mathcal{M} \to \mathcal{M}$ is surjective at every point $p \in \mathcal{M}$.

B.2 The sample paths of X are continuous.

B.3 F is finite. Also, for all compact subsets $\mathcal{K} \subset \mathcal{M}$, $\sup_{t\in\mathcal{T}} \sup_{p\in\mathcal{K}} \mathbb{E} d_{\mathcal{M}}^2(p, X(t)) < \infty$.

B.4 The image $\mathcal{U}$ of the mean function μ is bounded, that is, the diameter is finite, $\mathrm{diam}(\mathcal{U}) < \infty$.

B.5 For all $\epsilon > 0$, $\inf_{t\in\mathcal{T}} \inf_{p: d_{\mathcal{M}}(p,\mu(t)) \geq \epsilon} F(p,t) - F(\mu(t),t) > 0$.

To state the next condition, let $V_t(p) = \mathrm{Log}_p X(t)$. The calculus of manifolds suggests that $V_t(p) = -d_{\mathcal{M}}(p, X(t))\, \mathrm{grad}_p\, d_{\mathcal{M}}(p, X(t)) = \mathrm{grad}_p(-d_{\mathcal{M}}^2(p, X(t))/2)$, where grad_p denotes the gradient operator at p. For each $t \in \mathcal{T}$, let H_t denote the Hessian of the real function $d_{\mathcal{M}}^2(\cdot, X(t))/2$, that is, for vector fields U and W on $\mathcal{M}$,

$$\langle H_t U, W\rangle(p) = \langle -\nabla_U V_t, W\rangle(p) = \mathrm{Hess}_p\left(\frac{1}{2} d_{\mathcal{M}}^2(p, X(t))\right)(U, W).$$

B.6 $\inf_{t\in\mathcal{T}}\{\lambda_{\min}(\mathbb{E}H_t)\} > 0$, where $\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue of an operator or matrix.

B.7 $\mathbb{E}L(X)^2 < \infty$ and $L(\mu) < \infty$, where $L(f) := \sup_{s\neq t} d_{\mathcal{M}}(f(s), f(t))/|s-t|$ for a real function f on $\mathcal{M}$.

The assumption B.1 regarding the property of manifolds is met in general, for example, the d-dimensional unit sphere $\mathbb{S}^d$, SPD manifolds, etc. By the Hopf–Rinow theorem, the condition also implies that $\mathcal{M}$ is geodesically complete. Conditions similar to B.2, B.5, B.6 and B.7 are made in Dai and Müller (2018). The condition B.4 is a weaker requirement for the mean function and is automatically satisfied if the manifold is compact, while B.3 is analogous to standard moment conditions in *the* $(of that)$ ΓΔΦΔΔ$T f\phi$ ΓΠΓΥTΠΓΥ $B\Phi$

Also, the continuity of $\mu(t)$ and $\hat{\mu}(t)$ implies the continuity of $\gamma(\cdot,\cdot)$ and hence the measurability of $\gamma(\cdot,s)$ for each $s \in [0,1]$. By Proposition 2, one sees that $\Phi\hat{\mathcal{C}} = n^{-1}\sum_{i=1}^{n}(\Gamma\hat{V}_i \otimes \Gamma\hat{V}_i)$, recalling that $\hat{V}_i = \mathrm{Log}_{\hat{\mu}} X_i$ is a vector field along $\hat{\mu}$. It can also be seen that $(\hat{\lambda}_k, \Gamma\hat{\phi}_k)$ are eigenpairs of $\Phi\hat{\mathcal{C}}$. These identities match our intuition that the transported sample covariance operator ought to be an operator derived from transported sample vector fields, and that the eigenfunctions of the transported operator are identical to the transported eigenfunctions.

To state the asymptotic properties for the eigenstructure, we define

$$\eta_k = \min_{1\le j\le k}(\lambda_j - \lambda_{j+1}), \qquad J = \inf\{j \ge 1 : \lambda_j - \lambda_{j+1} \le 2|||\hat{\mathcal{C}} \ominus_\Phi \mathcal{C}|||_\mu\},$$

$$\hat{\eta}_j = \min_{1\le j\le k}(\hat{\lambda}_j - \hat{\lambda}_{j+1}), \qquad \hat{J} = \inf\{j \ge 1 : \hat{\lambda}_j - \hat{\lambda}_{j+1} \le 2|||\hat{\mathcal{C}} \ominus_\Phi \mathcal{C}|||_\mu\}.$$

THEOREM 7. *Assume that every eigenvalue λ_k has multiplicity one, and conditions* A.1, A.2′ *and* B.1–B.7 *hold. Suppose tangent vectors are parallel transported along minimizing geodesics for defining the parallel transporters* Γ *and* Φ. *If* $\mathbb{E}\|\mathrm{Log}_\mu X\|_\mu^4 < \infty$, *then* $|||\hat{\mathcal{C}} \ominus_\Phi \mathcal{C}|||_\mu^2 = O_P(n^{-1})$. *Furthermore,* $\sup_{k\ge1}|\hat{\lambda}_k - \lambda_k| \le |||\hat{\mathcal{C}} \ominus_\Phi \mathcal{C}|||_\mu$ *and for all* $1 \le k \le J-1$,

$$\|\hat{\phi}_k \ominus_\Gamma \phi_k\|_\mu^2 \le 8|||\hat{\mathcal{C}} \ominus_\Phi \mathcal{C}|||_\mu^2/\eta_k^2. \tag{10}$$

If (J, η_j) *is replaced by* $(\hat{J}, \hat{\eta}_j)$, *then* (10) *holds with probability* 1.

In this theorem, (10) generalizes Lemma 4.3 of Bosq (2000) to the Riemannian setting. Note that the intrinsic rate for $\hat{\mathcal{C}}$ is optimal. Also, from (10) one can deduce the optimal rate $\|\hat{\phi}_k \ominus_\Gamma \phi_k\|_\mu^2 = O_P(n^{-1})$ for a fixed k. We stress that these results apply not only to Euclidean submanifolds, but also to *general* Riemannian manifolds.

4. Intrinsic Riemannian functional linear regression.

4.1. *Regression model and estimation.* Classical functional linear regression for Euclidean functional data is well studied in the literature, that is, the model relating a scalar response Y and a functional predictor X by $Y = \alpha + \int_{\mathcal{T}} X(t)\beta(t)\,d\upsilon(t) + \varepsilon$, where α is the intercept, β is the slope function and ε represents measurement errors, for example, Cardot, Ferraty and Sarda (2003), Cardot, Mas and Sarda (2007), Hall and Horowitz (2007) and Yuan and Cai (2010), among others. However, for Riemannian functional data, both $X(t)$ and $\beta(t)$ take values in a manifold and hence the product $X(t)\beta(t)$ is not well defined. Rewriting the model as $Y = \alpha + \langle\langle X, \beta\rangle\rangle_{\mathcal{L}^2} + \varepsilon$, where $\langle\langle\cdot,\cdot\rangle\rangle_{\mathcal{L}^2}$ is the canonical inner product of the $\mathcal{L}^2$ square integrable functions, we propose to replace $\langle\langle\cdot,\cdot\rangle\rangle_{\mathcal{L}^2}$ by the inner product on the tensor Hilbert space $\mathscr{T}(\mu)$, and define the following Riemannian functional linear regression model:

$$Y = \alpha + \langle\langle \mathrm{Log}_\mu X, \mathrm{Log}_\mu \beta\rangle\rangle_\mu + \varepsilon, \tag{11}$$

where we require conditions A.1 and A.2. Note that β is a manifold valued function defined on $\mathcal{T}$, namely the *Riemannian slope function* of the model (11), and this model is linear in terms of $\mathrm{Log}_{\mu(t)}\,\beta(t)$. We stress that the model (11) is intrinsic to the Riemannian structure of the manifold.

According to Theorem 2.1 of Bhattacharya and Patrangenaru (2003), the process $\mathrm{Log}_{\mu(t)}\,X(t)$ is centered at its mean function, that is, $\mathbb{E}\,\mathrm{Log}_{\mu(t)}$

submanifold, an argument similar to that in Section 2.4 can show that, if one treats X as an ambient random process and adopts the FPCA and Tikhonov regularization approaches (Hall and Horowitz (2007)) to estimate the slope function β in the ambient space, then the estimates are the ambient representation of our estimates $\mathrm{Log}_{\hat{\mu}}\hat{\beta}$ and $\mathrm{Log}_{\hat{\mu}}\tilde{\beta}$ in (12) and (13), respectively.

4.2. *Asymptotic properties.* In order to derive convergence of the iRFPCA estimator and the Tikhonov estimator, we shall assume the sectional curvature of the manifold is bounded from below by κ to exclude pathological cases. The compact support condition on X in the case $\kappa < 0$ might be relaxed to weaker assumptions on the tail decay of the distribution of $\mathrm{Log}_{\mu} X$. Such weaker condition does not provide more insight for our derivation, but complicate the proofs significantly, which is not pursued further.

C.2 If $\kappa < 0$, X is assumed to lie in a compact subset $\mathcal{K}$ almost surely. Moreover, errors ε_i are identically distributed with zero mean and variance not exceeding a constant $C > 0$.

The following conditions concern the spacing and the decay rate of eigenvalues λ_k of the covariance operator, as well as the strength of the signal b_k. They are standard in the literature of functional linear regression, for example, Hall and Horowitz (2007).

C.3 For $k \ge 1$, $\lambda_k - \lambda_{k+1} \ge Ck^{-\alpha-1}$.
C.4 $|b_k| \le Ck^{-\varrho}$, $\alpha > 1$ and $(\alpha+1)/2 < \varrho$.

Let $\mathcal{F}(C, \alpha, \varrho)$ be the collection of distributions f of (X, Y) satisfying conditions C.2–C.4. The following theorem establishes the convergence rate of the iRFPCA estimator $\hat{\beta}$ for the class of models in $\mathcal{F}(C, \alpha, \varrho)$.

THEOREM 8. *Assume that conditions* A.1, A.2′, B.1–B.7 *and* C.1–C.4 *hold. If* $K \asymp n^{1/(4\alpha+2\varrho+2)}$, *then*

$$\lim_{c\to\infty} \lim_{n\to\infty} \sup_{f\in\mathcal{F}} P_f \left\{ \int_{\mathcal{T}} d^2_{\mathcal{M}}\big(\hat{\beta}(t), \beta(t)\big)\, \mathrm{d}\upsilon(t) > cn^{-\frac{2\varrho-1}{4\alpha+2\varrho+2}} \right\} = 0.$$

For the Tikhonov estimator $\tilde{\beta}$, we have a similar result. Instead of conditions C.3–C.4, we make the following assumptions, which again are standard in the functional data literature.

C.5 $k^{-\alpha} \le C\lambda_k$.

THEOREM 9. *Assume that conditions A.1, A.2′, B.1–B.7, C.1–C.2 and C.5–C.6 hold. If* $\rho \asymp n^{-\alpha/(\alpha+2\varrho)}$, *then*

$$\lim_{c\to\infty}\lim_{n\to\infty}\sup_{f\in\mathcal{G}} \mathrm{P}_f\left\{\int_{\mathcal{T}} d_{\mathcal{M}}^2\big(\tilde{\beta}(t),\beta(t)\big)\,\mathrm{d}\upsilon(t) > cn^{-\frac{2\varrho-\alpha}{2\varrho+\alpha}}\right\} = 0.$$

It is important to point out that the theory in Hall and Horowitz (2007) is formulated for Euclidean functional data and hence does not apply to Riemannian functional data. In particular, their proof machinery depends on the linear structure of the sample mean function $n^{-1}\sum_{i=1}^n X_i$ for Euclidean functional data. However, the intrinsic empirical mean generally does not admit an analytic expression, which hinges derivation of the optimal convergence rate. We leave the refinement on minimax rates of iRFPCA and Tikhonov estimators to future research. Note that model (11) can be extended to include a finite and fixed number of scalar predictors with slight modification, and the asymptotic properties of $\hat{\beta}$ and $\tilde{\beta}$ remain unchanged.

5. Numerical examples.

5.1. *Simulation studies.* We consider two manifolds that are frequently encountered in practice[2]. The first one is the unit sphere $\mathbb{S}^d$ which is a compact nonlinear Riemannian submanifold of $\mathbb{R}^{d+1}$ for a positive integer d. The spheres can be used to model compositional data, as exhibited in Dai and Müller (2018) which also provide details of the geometry of $\mathbb{S}^d$. Here we consider the case of $d=2$. The sphere $\mathbb{S}^2$ consists of points $(x,y,z)\in\mathbb{R}^3$ satisfying $x^2+y^2+z^2=1$. Since the intrinsic Riemannian geometry of $\mathbb{S}^2$ is the same as the one inherited from its ambient space (referred to as ambient geometry hereafter), according to the discussion in Section 2.4, the ambient approach to FPCA and functional linear regression yields the same results as our intrinsic approach.

The other manifold considered is the space of $m\times m$ symmetric positive definite matrices, denoted by $\mathrm{Sym}_\star^+(m)$. The space $\mathrm{Sym}_\star^+(m)$ includes nonsingular covariance matrices which naturally arise from the study of DTI data (Dryden, the affine-invariant metric has a negative sectional curvature, and thus the Fréchet mean is unique if it exists. In our simulation, we consider $m = 3$. We emphasize that the affine-invariant geometry of $\mathrm{Sym}_\star^+(m)$ is different from the geometry inherited from the linear space $\mathrm{Sym}(m)$. Thus, the ambient RFPCA of Dai and Müller (2018) might yield inferior performance on this manifold.

We simulate data as follows. First, the time domain is set to $\mathcal{T} = [0, 1]$. The mean curves for $\mathbb{S}^2$ and $\mathrm{Sym}_\star^+(m)$ are, respectively, $\mu(t) = (\sin\varphi(t)\cos\theta(t), \sin\varphi(t)\sin\theta(t), \cos\varphi(t))$ with $\theta(t) = 2t^2 + 4t + 1/2$ and $\varphi(t) = (t^3 + 3t^2 + t + 1)/2$, and $\mu(t) = (t^{0.4}, 0.5t, 0.1t^{1.5}; 0.5t, t^{0.5}, 0.5t; 0.1t^{1.5}, 0.5t, t^{0.6})$ that is a 3×3 matrix. The Riemannian random processes are produced in accordance to $X = \mathrm{Exp}(\sum_{k=1}^{20} \sqrt{\lambda_k}\xi_k\phi_k)$, where $\xi_k \overset{\text{i.i.d.}}{\sim} \mathrm{Uniform}(-\pi/4, \pi/4)$ for $\mathbb{S}^2$ and $\xi_k \overset{\text{i.i.d.}}{\sim} N(0, 1)$ for $\mathrm{Sym}_\star^+(m)$. We set iRFPC $\phi_k(t) = (A\psi_k(t))^T\mathbf{E}(t)$, where $\mathbf{E}(t) = (E_1(\mu(t)), \dots, E_d(\mu(t)))$ is an orthonormal frame over the path μ, $\psi_k(t) = (\psi_{k,1}(t), \dots, \psi_{k,d}(t))^T$ with $\psi_{k,j}$ being orthonormal Fourier basis functions on $\mathcal{T}$, and A is an orthonormal matrix that is randomly generated but fixed throughout all simulation replicates. We take $\lambda_k = 2k^{-1.2}$ for all manifolds. Each curve $X(t)$ is observed at $M = 101$ regular design points $t = 0, 0.01, \dots, 1$. The slope function is $\beta = \sum_{k=1}^{K} c_k\phi_k$ with $c_k = 3k^{-2}/2$. Two different types of distribution for ε in (11) are considered, namely, normal and Student's t distribution with degree of freedom $\mathrm{df} = 2.1$. Note that the latter is a heavy-tailed distribution, with a smaller df suggesting a heavier tail and $\mathrm{df} > 2$ ensuring the existence of variance. In addition, the noise ε is scaled to make the signal-to-noise ratio equal to 2. Three different training sample sizes are considered, namely, 50, 150 and 500, while the sample size for test data is 5000. Each simulation setup is repeated independently 100 times.

First, we illustrate the difference between the intrinsic measure and the ambient counterpart for the discrepancy of two random objects residing on different tangent spaces, through the example of sphere manifold $\mathbb{S}^2$ and the route to iRFPCs. Recall that the metric of $\mathbb{S}^2$ agrees with its ambient Euclidean geometry, so that both iRFPCA and RFPCA essentially yield the same estimates for iRFPCs. We propose to use the intrinsic root mean integrated squared error (iRMISE) $\{\mathbb{E}\|\hat{\phi}_k \ominus_\Gamma \phi_k\|_\mu^2\}^{1/2}$ to characterize the difference between ϕ_k and its estimator $\hat{\phi}_k$, while Dai and Müller (2018) adopt the ambient RMISE (aRMISE) $\{\mathbb{E}\|\hat{\phi}_k - \phi_k\|_{\mathbb{R}^{d_0}}^2\}^{1/2}$, as discussed in Section 2.4. The numerical results of iRMISE and aRMISE for $\hat{\phi}_1$ and $\hat{\phi}_2$, as well as the RMISE for $\hat{\mu}$, are shown in Table 1. We see that, when n is small and hence $\hat{\mu}$ is not sufficiently close to μ, the difference between iRMISE and aRMISE is visible, while such difference decreases as the sample size grows and $\hat{\mu}$ converges to μ. In particular, aRMISE is always larger than iRMISE since aRMISE contains an additional ambient component that is not intrinsic to the manifold, as illustrated on the left panel of Figure 1.

We now use iRMISE to assess the performance of iRFPCA by comparing to the ambient counterpart RFPCA proposed by Dai and Müller (2018). Table 2 presents

TABLE 1

The root mean integrated squared error (RMISE) of the estimation of the mean function, and the intrinsic RMISE (iRMISE) and the ambient RMISE (aRMISE) of the estimation for the first two eigenfunctions in the case of $\mathbb{S}^2$ manifold. The Monte Carlo standard error based on 100 simulation runs is given in parentheses

	$n = 50$		$n = 150$		$n = 500$	
μ	0.244 (0.056)		0.135 (0.029)		0.085 (0.019)	
	iRMISE	aRMISE	iRMISE	aRMISE	iRMISE	aRMISE
ϕ_1	0.279 (0.073)	0.331 (0.078)	0.147 (0.037)	0.180 (0.042)	0.086 (0.022)	0.106 (0.027)
ϕ_2	0.478 (0.133)	0.514 (0128)	0.264 (0.064)	0.287 (0.061)	0.147 (0.044)	0.167 (0.042)

the results for the top 5 eigenelements. The first observation is that iRFPCA and RFPCA yield the same results on the manifold $\mathbb{S}^2$, which numerically verifies our discussion in Section 2.4. We notice that in Dai and Müller (2018) the quality of estimation of principal components is not evaluated, likely due to the lack of a proper tool to do so. In contrast, our framework of tensor Hilbert space provides an intrinsic gauge (e.g., iRMISE) to naturally compare vector fields along different curves. For the case of $\mathrm{Sym}_\star^+(m)$ which is not a Euclidean submanifold, the iRFPCA produces more accurate estimation than RFPCA. In particular, as sample size grows, the estimation error for iRFPCA decreases quickly, while the error of RFPCA persists. This coincides with our intuition that when the geometry induced from the ambient space is not the same as the intrinsic geometry, the ambient RFPCA incurs loss of statistical efficiency, or even worse, inconsistent estimation. In summary, the results on $\mathrm{Sym}_\star^+(m)$ numerically demonstrate that the RFPCA proposed by Dai and Müller (2018) does not apply to manifolds that do not have an ambient space or whose intrinsic geometry differs from its ambient geometry, while our iRFPCA are applicable to such Riemannian manifolds.

For functional linear regression, we adopt iRMISE to quantify the quality of the estimator $\hat{\beta}$ for slope function β, and assess the prediction performance by prediction RMSE on an independent test dataset. For comparison, we also fit the functional linear model using the principal components produced by RFPCA (Dai and Müller (2018)), and hence we refer to this competing method as RFLR. For both methods, the tuning parameter which is the number of principal components included for $\hat{\beta}$, is selected by using an independent validation data of the same size of the training data to ensure fair comparison between two methods. The simulation results are presented in Table 3. As expected, we observe that on $\mathbb{S}^2$ both methods produce the same results. For the SPD manifold, in terms of estimation, we see that iRFLR yields far better estimators than RFLR does. Particularly, we again observe that, the quality of RFLR estimators does not improve significantly when sample

TABLE 2

Intrinsic root integrated mean squared error (iRMISE) of estimation for eigenelements. The first column denotes the manifolds, where $\mathbb{S}^2$ is the unit sphere and $\mathrm{Sym}^{+}_{\star}(m)$ is the space of $m \times m$ symmetric positive-definite matrices endowed with the affine-invariant metric. In the second column, $\phi_1, \dots, \phi_5$ are the top five intrinsic Riemannian functional principal components. Columns 3–5 are (iRMISE) of the iRFPCA estimators for $\phi_1, \dots, \phi_5$ with different sample sizes, while columns 5–8 are iRMISE for the RFPCA estimators. The Monte Carlo standard error based on 100 simulation runs is given in parentheses

		iRFPCA			RFPCA		
Manifold	FPC	$n = 50$	$n = 150$	$n = 500$	$n = 50$	$n = 150$	$n = 500$
$\mathbb{S}^2$	ϕ_1	0.279 (0.073)	0.147 (0.037)	0.086 (0.022)	0.279 (0.073)	0.147 (0.037)	0.086 (0.022)
	ϕ_2	0.475 (0.133)	0.264 (0.064)	0.147 (0.044)	0.475 (0.133)	0.264 (0.064)	0.147 (0.044)
	ϕ_3	0.647 (0.153)	0.389 (0.120)	0.206 (0.054)	0.647 (0.153)	0.389 (0.120)	0.206 (0.054)
	ϕ_4	0.818 (0.232)	0.502 (0.167)	0.261 (0.065)	0.818 (0.232)	0.502 (0.167)	0.261 (0.065)
	ϕ_5	0.981 (0.223)	0.586 (0.192)	0.329 (0.083)	0.981 (0.223)	0.586 (0.192)	0.329 (0.083)
$\mathrm{Sym}^{+}_{\star}(m)$	ϕ_1	0.291 (0.105)	0.155 (0.046)	0.085 (0.025)	0.707 (0.031)	0.692 (0.021)	0.690 (0.014)
	ϕ_2	0.523 (0.203)	0.283 (0.087)	0.143 (0.040)	0.700 (0.095)	0.838 (0.113)	0.684 (0.055)
	ϕ_3	0.734 (0.255)	0.418 (0.163)	0.206 (0.067)	0.908 (0.116)	0.904 (0.106)	0.981 (0.039)
	ϕ_4	0.869 (0.251)	0.566 (0.243)	0.288 (0.086)	0.919 (0.115)	1.015 (0.113)	0.800 (0.185)
	ϕ_5	1.007 (0.231)	0.699 (0.281)	0.378 (0.156)	0.977 (0.100)	1.041 (0.140)	1.029 (0.058)

TABLE 3

Estimation quality of slope function β and prediction of y on test datasets. The second column indicates the distribution of noise, while the third column indicates the manifolds, where $\mathbb{S}^2$ is the unit sphere and $\mathrm{Sym}^+_\star(m)$ is the space of $m \times m$ symmetric positive-definite matrices endowed with the affine-invariant metric. Columns 4–6 are performance of the iRFLR on estimating the slope curve β and predicting the response on new instances of predictors, while columns 7–9 are performance of the RFLR method. Estimation quality of the slope curve is quantified by intrinsic root mean integrated squared errors (iRMISE), while the performance of prediction on independent test data is measured by root mean squared errors (RMSE). The Monte Carlo standard error based on 100 simulation runs is given in parentheses

			iRFLR			RFLR		
			$n = 50$	$n = 150$	$n = 500$	$n = 50$	$n = 150$	$n = 500$
Estimation	normal	$\mathbb{S}^2$	0.507 (0.684)	0.164 (0.262)	0.052 (0.045)	0.507 (0.684)	0.164 (0.262)	0.052 (0.045)
		SPD	1.116 (2.725)	0.311 (0.362)	0.100 (0.138)	2.091 (0.402)	1.992 (0.218)	1.889 (0.126)
	$t(2.1)$	$\mathbb{S}^2$	0.575 (0.768)	0.183 (0.274)	0.053 (0.050)	0.575 (0.768)	0.183 (0.274)	0.053 (0.050)
		SPD	1.189 (2.657)	0.348 (0.349)	0.108 (0.141)	2.181 (0.439)	1.942 (0.209)	1.909 (0.163)
Prediction	normal	$\mathbb{S}^2$	0.221 (0.070)	0.135 (0.046)	0.083 (0.019)	0.221 (0.070)	0.135 (0.046)	0.083 (0.019)
		SPD	0.496 (0.184)	0.284 (0.092)	0.165 (0.062)	0.515 (0.167)	0.328 (0.083)	0.248 (0.047)
	$t(2.1)$	$\mathbb{S}^2$	0.251 (0.069)	0.142 (0.042)	0.088 (0.020)	0.251 (0.069)	0.142 (0.042)	0.088 (0.020)
		SPD	0.532 (0.189)	0.298 (0.097)	0.172 (0.066)	0.589 (0.185)	0.360 (0.105)	0.268 (0.051)

size increases, in contrast to estimators based on the proposed iRFLR. For prediction, iRFLR outperforms RFLR by a significant margin. Interestingly, comparing to estimation of slope function where the RFLR estimator is much inferior to the iRFLR one, the prediction performance by RFLR is relatively closer to that by iRFLR. We attribute this to the smoothness effect brought by the integration in model (11). Nevertheless, although the integration cancels out certain discrepancy between the intrinsic and the ambient geometry, the loss of efficiency is inevitable for the RFLR method that is bound to the ambient space. In addition, we observe that the performance of both methods for Gaussian noise is slightly better than that in the case of heavy-tailed noise.

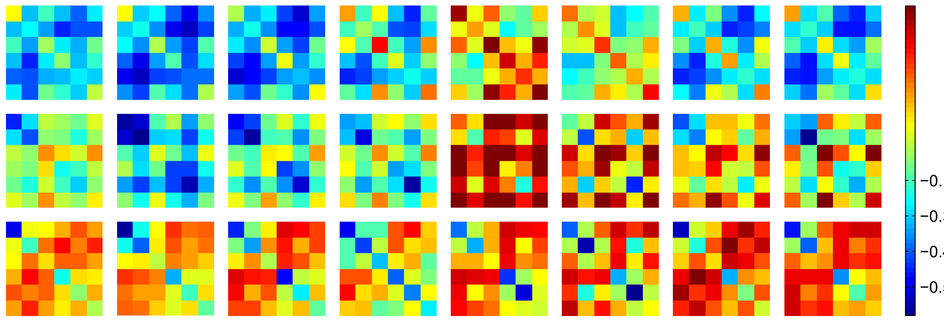
5.2. *Data application.* We apply the proposed iRFPCA and iRFLR to analyze the relationship between functional connectivity and behavioral data from the HCP 900 subjects release (Essen et al. (2013)). Although neural effects on language (Binder et al. (1997)), emotion (Phan et al. (2002)) and fine motor skills (Dayan and Cohen (2011)) have been extensively studied in the literature, scarce is the exploration on human behaviors that do not seem related to neural activities, such as endurance. Nevertheless, a recent research by Raichlen et al. (2016) suggests that endurance can be related to functional connectivity. Our goal is to study the endurance performance of subjects based on their functional connectivity.

The data consist of $n = 330$ subjects who are healthy young adults, in which each subject is asked to walk for two minutes and the distance in feet is recorded. Also, each subject participates in a motor task, where participants are asked to act according to presented visual cues, such as tapping their fingers, squeezing their toes or moving their tongue. During the task, the brain of each subject is scanned and the neural activities are recorded at 284 equispaced time points. After preprocessing, the average BOLD (blood-oxygen-level dependent) signals at 68 different brain regions are obtained. The details of the experiment and data acquisition can be found in the reference manual of WU-Minn HCP 900 Subjects Data Release that is available on the website of human connectome project.

Our study focuses on $m = 6$ regions that are related to the primary motor cortex, including precentral gyrus, Broca's area, etc. At each design time point t, the functional connectivity of the ith subject is represented by the covariance matrix $S_i(t)$ of BOLD signals from regions of interest (ROI). To practically compute $S_i(t)$, let V_{it} be an m-dimensional column vector that represents the BOLD signals at time t from the m ROIs of the ith subject. We then adopt a local sliding window approach (Park et al. (2017)) to compute $S_i(t)$ by

$$S_i(t) = \frac{1}{2h+1} \sum_{j=t-h}^{t+h} (V_{ij} - \bar{V}_{it})(V_{ij} - \bar{V}_{it})^T \quad \text{with } \bar{V}_{it} = \frac{1}{2h+1} \sum_{j=t-h}^{t+h} V_{ij},$$

where h is a positive integer that represents the length of the sliding window to compute $S_i(t)$ for $t = h+1, h, \ldots, 284-h$. Without loss of generality, we reparameterize each $S_i(\cdot)$ from the domain $[h+1, 284-h]$ to $[1, 284-2h]$. In practice,


−0.2
−0.3
−0.4
−0.5

speed and strength are included as baseline covariates, selected by the forward-stepwise selection method (Hastie, Tibshirani and Friedman (2009), Section 3.3). Among these covariates, gender and age are in accordance with the common sense about endurance, while gait speed and muscle strength could be influential since endurance is measured by the distance walked in 6 minutes. Our primary interest is to assess the significance of the functional predictor when effects of the baseline covariates is controlled.

To fit the intrinsic functional linear model, we adopt the cross-validation procedure to select the number of components to be included in representing the Riemannian functional predictor and the Riemannian slope function β. For assessment, we conduct 100 runs of 10-fold cross-validation, where in each run we permute the data independently. In each run, the model is fitted on 90% data and the MSE for predicting the walking distance is computed on the other 10% data for both iRFLR and RFLR methods. The fitted intrinsic Riemannian slope function $\mathrm{Log}_{\hat{\mu}}\hat{\beta}$ displayed in the bottom panel of Figure 2 shows the pattern of weight changes. The MSE for iRFLR is reduced by around 9.7%, compared to that for RFLR. Moreover, the R^2 for iRFLR is 0.338, with a p-value 0.012 based on a permutation test of 1000 permutations, which is significant at level 5%. In contrast, the R^2 for RFLR drops to 0.296 and the p-value is 0.317 that does not show significance at all.

APPENDIX A: BACKGROUND ON RIEMANNIAN MANIFOLD

We introduce geometric concepts related to Riemannian manifolds from an intrinsic perspective without referring to any ambient space.

A smooth manifold is a differentiable manifold with all transition maps being C^∞ differentiable. For each point p on the manifold $\mathcal{M}$, there is a linear space $T_p\mathcal{M}$ of tangent vectors which are derivations. A derivation is a linear map that sends a differentiable function on $\mathcal{M}$ into $\mathbb{R}$ and satisfies the Leibniz property. For example, if D_v is the derivation associated with the tangent vector v at p, then $D_v(fg) = (D_v f) \cdot g(p) + f(p) \cdot D_v(g)$ for any $f, g \in A(\mathcal{M})$, where $A(\mathcal{M})$ is a collection of real-valued differentiable functions on $\mathcal{M}$. For submanifolds of a Euclidean space $\mathbb{R}^{d_0}$ for some $d_0 > 0$, tangent vectors are often perceived as vectors in $\mathbb{R}^{d_0}$ that are tangent to the submanifold surface. If one interprets a Euclidean tangent vector as a directional derivative along the vector direction, then Euclidean tangent vectors coincide with our definition of tangent vectors on a general manifold. The linear space $T_p\mathcal{M}$ is called the tangent space at p. The disjoint union of tangent spaces at each point constitutes the tangent bundle, which is also equipped with a smooth manifold structure induced by $\mathcal{M}$. The tangent bundle of $\mathcal{M}$ is conventionally denoted by $T\mathcal{M}$. A (smooth) vector field V is a map from $\mathcal{M}$ to $T\mathcal{M}$ such that $V(p) \in T_p\mathcal{M}$ for each $p \in \mathcal{M}$. It is also called a smooth section of $T\mathcal{M}$. Noting that a tangent vector is a tensor of type $(0, 1)$, a vector field can be viewed as a kind of tensor field, which assigns a tensor to each point on $\mathcal{M}$. A vector field along a curve $\gamma : I \to \mathcal{M}$ on $\mathcal{M}$ is a map V from an interval $I \subset \mathbb{R}$ to $T\mathcal{M}$ such that $V(t) \in T_{\gamma(t)}\mathcal{M}$. For a smooth function from a manifold $\mathcal{M}$ and to another manifold $\mathcal{N}$, the differential $\mathrm{d}\varphi_p$ of f at $p \in \mathcal{M}$ is a linear map from $T_p\mathcal{M}$ to $T_{\varphi(p)}\mathcal{N}$, such that $\mathrm{d}\varphi_p(v)(f) = D_v(f \circ \varphi)$ for all $f \in A(\mathcal{M})$ and $v \in T_p\mathcal{M}$.

An affine connection ∇ on $\mathcal{M}$ is a bilinear mapping that sends a pair of

$L(\gamma)$ over all continuously differentiable curves joining p and q. For a connected and complete Riemannian, given two points on the manifold, there is a minimizing geodesic connecting these two points.

APPENDIX B: IMPLEMENTATION FOR $\mathrm{Sym}_\star^+(m)$

Given $\mathrm{Sym}_\star^+(m)$-valued functional data $X_1,\dots,X_n$, below we briefly outline the numerical steps to perform iRFPCA. The computation details for $\mathbb{S}^d$ can be found in Dai and Müller (2018).

Step 1. Compute the sample Fréchet mean $\hat\mu$. As there is no analytic solution, the recursive algorithm developed by Cheng et al. (2016) can be used.

Step 2. Select an orthonormal frame $\mathbf{E}=(E_1,\dots,E_d)$ along $\hat\mu$. For $\mathrm{Sym}_\star^+(m)$, at each $S\in\mathrm{Sym}_\star^+(m)$, the tangent space $T_S\mathrm{Sym}_\star^+(m)$ is isomorphic to $\mathrm{Sym}(m)$. This space has a canonical linearly independent basis $e_1,\dots,e_d$ with $d=m(m+1)/2$, defined in the following way. For an integer $k\in[1,d]$, let N_1 be the largest integer such that $N_1(N_1+1)/2\le k$. Let $N_2=k-N_1(N_1-1)/2$. Then e_k is defined as the $m\times m$ matrix that has 1 at (N_1,N_2), 1 at (N_2,N_1) and 0 elsewhere. Because the inner product on the space $T_{\hat\mu(t)}\mathrm{Sym}_\star^+(m)$ is given by

$$\mathrm{tr}\left(\hat\mu(t)^{-1/2}U\hat\mu(t)V\hat\mu(t)^{-1/2}\right)$$

for $U,V\in T_S\mathrm{Sym}_\star^+(m)$, in general this basis is not orthonormal in $T_{\hat\mu(t)}\mathrm{Sym}_\star^+(m)$. To obtain an orthonormal basis of $T_{\hat\mu(t)}\mathrm{Sym}_\star^+(m)$ for any given t, we can apply the Gram–Schmidt procedure on the basis $e_1,\dots,e_d$. The orthonormal bases obtained in this way smoothly vary with t and hence form an orthonormal frame of $\mathrm{Sym}_\star^+(m)$ along $\hat\mu$.

Step 3. Compute the $\mathbf{E}$-coordinate representation $\hat Z_{\mathbf{E},i}$ of each $\mathrm{Log}_{\hat\mu}X_i$. For $\mathrm{Sym}_\star^+(m)$, the logarithm map at a generic $S\in\mathrm{Sym}_\star^+(m)$ is given by $\mathrm{Log}_S(Q)=S^{1/2}\log(S^{-1/2}QS^{-1/2})S^{1/2}$ for $Q\in\mathrm{Sym}_\star^+(m)$, where log denotes the matrix logarithm function. Therefore,

$$\mathrm{Log}_{\hat\mu(t)}X_i(t)=\hat\mu(t)^{1/2}\log\big(\hat\mu(t)^{-1/2}X_i(t)\hat\mu(t)^{-1/2}\big)\hat\mu(t)^{1/2}.$$

Using the orthonormal basis $E_1(t),\dots,E_d(t)$ obtained in the previous step, one can compute the coefficient $\hat Z_{\mathbf{E},i}(t)$ representation of $\mathrm{Log}_{\hat\mu(t)}X_i(t)$ for any given t.

Step 4. Compute the first K eigenvalues $\hat\lambda_1,\dots,\hat\lambda_K$ and eigenfunctions $\hat\phi_{\mathbf{E},1},\dots,\hat\phi_{\mathbf{E},K}$ of the empirical covariance function $\hat{\mathcal{C}}_{\mathbf{E}}(s,t)=n^{-1}\sum_{i=1}^n\hat Z_{\mathbf{E},i}(s)\times\hat Z^T_{\mathbf{E},i}(t)$. This step is generic and does not involve the manifold structure. For $d=1$, the classic univariate FPCA method such as Hsing and Eubank (2015) can be employed to derive the eigenvalues and eigenfunctions of $\hat{\mathcal{C}}_{\mathbf{E}}$. When $d>1$, each observed coefficient function $\hat Z_{\mathbf{E},i}(t)$ is vector-valued. FPCA for vector-valued functional data can be performed by the method developed in Happ and Greven (2018) or Wang (2008).

Step 5. Compute the scores $\hat{\xi}_{ik} = \int \hat{Z}^T_{\mathbf{E},i}(t)\hat{\phi}_{\mathbf{E},k}(t)\,\mathrm{d}t$. Finally, compute the approximation of X_i by the first K principal components using

$$\hat{X}_i^K(t) = \mathrm{Exp}_{\hat{\mu}(t)} \sum_{k=1}^{K} \hat{\xi}_{ik}\hat{\phi}^T_{\mathbf{E},k}(t)\mathbf{E}(t),$$

where for $\mathrm{Sym}^+_\star(m)$, the exponential map at a generic S is given by

$$\mathrm{Exp}_S(U) = S^{1/2}\exp\big(S^{-1/2}US^{-1/2}\big)S^{1/2}$$

for $U \in T_S\mathrm{Sym}^+_\star(m)$, where exp denotes the matrix exponential function.

APPENDIX C: PROOFS OF MAIN THEOREMS

PROOF OF THEOREM 1. We first show that $\mathscr{T}(\mu)$ is a Hilbert space. It is sufficient to prove that the inner product space $\mathscr{T}(\mu)$ is complete. Suppose $\{V_n\}$ is a Cauchy sequence in $\mathscr{T}(\mu)$. We will later show that there exists a subsequence $\{V_{n_k}\}$ such that

(14) $$\sum_{k=1}^{\infty}\big|V_{n_{k+1}}(t) - V_{n_k}(t)\big| < \infty, \quad \upsilon\text{-a.s.}$$

Since $T_{\mu(t)}\mathcal{M}$ is complete, the limit $V(t) = \lim_{k\to\infty} V_{n_k}(t)$ is υ-a.s. well defined and in $T_{\mu(t)}\mathcal{M}$. Fix an $\epsilon > 0$ and choose N such that $n, m \geq M$ imply $\|V_n - V_m\|_\mu \leq \epsilon$. Fatou's lemma applying to the function $|V(t) - V_m(t)|$ implies that if $m \geq N$, then $\|V - V_m\|^2_\mu \leq \liminf_{k\to\infty}\|V_{n_k} - V_m\|^2_\mu \leq \epsilon^2$. This shows that $V - V_m \in \mathscr{T}(\mu)$. Since $V = (V - V_m) + V_m$, we see that $V \in \mathscr{T}(\mu)$. The arbitrariness of ϵ implies that $\lim_{m\to\infty}\|V - V_m\|_\mu = 0$. Because $\|V - V_n\|_\mu \leq \|V - V_m\|_\mu + \|V_m - V_n\|_\mu \leq 2\epsilon$, we conclude that V_n converges to V in $\mathscr{T}(\mu)$.

It remains to show (14). To do so, we choose $\{n_k\}$ so that $\|V_{n_k} - V_{n_{k+1}}\|_\mu \leq 2^{-k}$. This is possible since V_n is a Cauchy sequence. Let $U \in \mathscr{T}(\mu)$. By Cauchy–Schwarz inequality, $\int_{\mathcal{T}} |U(t)| \cdot |V_{n_k}(t) - V_{n_{k+1}}(t)|\,\mathrm{d}\upsilon(t) \leq \|U\|_\mu\|V_{n_k} - V_{n_{k+1}}\|_\mu \leq 2^{-k}\|U\|_\mu$. Thus, $\sum_k \int_{\mathcal{T}} |U(t)| \cdot |V_{n_k}(t) - V_{n_{k+1}}(t)|\,\mathrm{d}\upsilon(t) \leq \|U\|_\mu < \infty$. Then (14) follows, because otherwise, if the series diverges on a set A with $\upsilon(A) > 0$, then a choice of U such that $|U(t)| > 0$ for $t \in A$ contradicts the above inequality.

Now let $\mathbf{E}$ be a measurable orthonormal frame. For every element $U \in \mathscr{T}(\mu)$, the coordinate representation of U with respect to $\mathbf{E}$ is denoted by $U_{\mathbf{E}}$. One can see that $U_{\mathbf{E}}$ is an element in the Hilbert space $\mathcal{L}^2(\mathcal{T}, \mathbb{R}^d)$ of square integrable $\mathbb{R}^d$-valued measurable functions with norm $\|f\|_{\mathcal{L}^2} = \{\int_{\mathcal{T}} |f(t)|^2\,\mathrm{d}\upsilon(t)\}^{1/2}$ for $f \in \mathcal{L}^2(\mathcal{T}, \mathbb{R}^d)$. If we define the map $\Upsilon : \mathscr{T}(\mu) \to \mathcal{L}^2(\mathcal{T}, \mathbb{R}^d)$ by $\Upsilon(U) = U_{\mathbf{E}}$, we can immediately see that Υ is a linear map. It is also surjective, because for any $f \in \mathcal{L}^2(\mathcal{T}, \mathbb{R}^d)$, the vector field U along μ given by $U_f(t) = f(t)\mathbf{E}(\mu(t))$ for

$t \in \mathcal{T}$ is an element in $\mathscr{T}(\mu)$, since $\|U_f\|_\mu = \|f\|_{\mathcal{L}^2}$. It can be also verified that Υ preserves the inner product. Therefore, it is a Hilbertian isomorphism. Since $\mathcal{L}^2(\mathcal{T}, \mathbb{R}^d)$ is separable, the isomorphism between $\mathcal{L}^2(\mathcal{T}, \mathbb{R}^d)$ and $\mathscr{T}(\mu)$ implies that $\mathscr{T}(\mu)$ is also separable. □

PROOF OF PROPOSITION 2. The regularity conditions on f, h and γ ensure that Γ and Φ are measurable. Parts 1, 2 and 6 can be deduced from the fact that $\mathcal{P}_{f,h}$ is a unitary operator between two finite-dimensional real Hilbert spaces and its inverse is $\mathcal{P}_{h,f}$. To reduce notational burden, we shall suppress the subscripts f, h from $\Gamma_{f,h}$ and $\Phi_{f,h}$ below. For Part 3,

$$(\Phi\mathcal{A})(\Gamma U) = \Gamma(\mathcal{A}\Gamma^*\Gamma U)) = \Gamma(\mathcal{A}U).$$

To prove Part 4, assume $V \in \mathscr{T}(g)$. Then, noting that $\Gamma(\Gamma^* V) = V$ and $\Gamma^*(\Gamma U) = U$, we have

$$\begin{aligned}(\Phi\mathcal{A})((\Phi\mathcal{A}^{-1})V) &= (\Phi\mathcal{A})(\Gamma(\mathcal{A}^{-1}\Gamma^* V)) \\ &= \Gamma(\mathcal{A}\Gamma^*(\Gamma(\mathcal{A}^{-1}\Gamma^* V))) \\ &= \Gamma(\mathcal{A}\mathcal{A}^{-1}\Gamma^* V) = \Gamma(\Gamma^* V) = V\end{aligned}$$

and

$$\begin{aligned}(\Phi\mathcal{A}^{-1})(\Phi\mathcal{A}V) &= (\Phi\mathcal{A}^{-1})(\Gamma(\mathcal{A}\Gamma^* V)) \\ &= \Gamma(\mathcal{A}^{-1}\Gamma^*(\Gamma(\mathcal{A}\Gamma^* V))) \\ &= \Gamma(\mathcal{A}^{-1}\mathcal{A}\Gamma^* V) = \Gamma(\Gamma^* V) = V.\end{aligned}$$

Part 5 is seen by the following calculation: for $V \in \mathscr{T}(g)$,

$$\begin{aligned}\Big(\Phi_{f,g}\sum c_k\varphi_k \otimes \varphi_k\Big)V &= \Gamma\Big(\sum c_k \langle\!\langle \varphi_k, \Gamma^* V\rangle\!\rangle_f \varphi_k\Big) \\ &= \sum c_k \langle\!\langle \phi_k, \Gamma^* V\rangle\!\rangle_f \Gamma\varphi_k \\ &= \sum c_k \langle\!\langle \Gamma\varphi_k, V\rangle\!\rangle_g \Gamma\varphi_k \\ &= \Big(\sum c_k \Gamma\varphi_k \otimes \Gamma\varphi_k\Big)V.\end{aligned}$$ □

PROOF OF PROPOSITION 4. The case $\kappa \geq 0$ is already given by Dai and Müller (2018) with $C = 1$. Suppose $\kappa < 0$. The second statement follows from the first one if we let $O = \mu(t)$, $P = X(t)$ and $Q = X_K(t)$ for any fixed t and note that C is independent of t.

For the first statement, the inequality is clearly true if $P = O$, $Q = O$ or $P = Q$. Now suppose O, P and Q are distinct points on $\mathcal{M}$. The minimizing geodesic curves between these points form a geodesic triangle on $\mathcal{M}$. By Toponogov's theorem (the hinge version), $d_{\mathcal{M}}(P, Q) \leq d_{\mathbb{M}_\kappa}(P', Q')$, where $\mathbb{M}_\kappa$ is the model space with constant sectional curvature κ. For $\kappa < 0$, it is taken as the hyperboloid with curvature κ. Let $a = d_{\mathcal{M}}(O, P)$, $b = d_{\mathcal{M}}(O, Q)$ and $c = d_{\mathcal{M}}(P, Q)$. The interior angle of geodesics connecting O to P and O to Q is denoted by γ. Denote $\delta = \sqrt{-\kappa}$, the law of cosines on $\mathbb{M}_\kappa$ gives

$$\cosh(\delta c) = \{\cosh(\delta a)\cosh(\delta b) - \sinh(\delta a)\sinh(\delta b)\} + \{\sinh(\delta a)\sinh(\delta b)(1 - \cos\gamma)$$

where $C = \{(2BD + \sqrt{2B})/\delta\}^2$, or in other words, $d_{\mathcal{M}}(P, Q) \le \sqrt{C}|\mathrm{Log}_O P - \mathrm{Log}_O Q|$. $\square$

PROOF OF PROPOSITION 5. Part 1 follows from a simple calculation. To lighten notation, let $\mathbf{f}^T\mathbf{E}$ denote $\mathbf{f}^T(\cdot)\mathbf{E}(\mu(\cdot))$ for a $\mathbb{R}^d$ valued function defined on $\mathcal{T}$. Suppose $\phi_{\mathbf{E},k}$ is the coordinate of ϕ_k under $\mathbf{E}$. Because

$$
\begin{aligned}
(\mathcal{C}_{\mathbf{E}}\phi_{\mathbf{E},k})^T\mathbf{E} &= \mathbb{E}\langle Z_{\mathbf{E}}, \phi_{\mathbf{E},k}\rangle Z_{\mathbf{E}}\mathbf{E} \\
&= \mathbb{E}\langle\!\langle \mathrm{Log}_\mu X, \phi_k\rangle\!\rangle_\mu \mathrm{Log}_\mu X \\
&= \lambda_k\phi_k = \lambda_k\phi_{\mathbf{E},k}^T\mathbf{E},
\end{aligned}
$$

one concludes that $\mathcal{C}_{\mathbf{E}}\phi_{\mathbf{E},k} = \lambda_k\phi_{\mathbf{E},k}$ and hence $\phi_{\mathbf{E},k}$ is an eigenfunction of $\mathcal{C}_{\mathbf{E}}$ corresponding to the eigenvalue λ_k. Other results in Part 2 and 3 have been derived in Section 3. The continuity of X and $\mathbf{E}$, in conjunction with $\mathbb{E}\|\mathrm{Log}_\mu X\|_\mu^2 < \infty$, implies that $Z_{\mathbf{E}}$ is a mean square continuous random process and the joint measurability of X passes to $Z_{\mathbf{E}}$. Then $Z_{\mathbf{E}}$ can be regarded as random element of the Hilbert space $\mathcal{L}^2(\mathcal{T}, \mathscr{B}(\mathcal{T}), \upsilon)$ that is isomorphic to $\mathscr{T}(\mu)$. Also, the isomorphism maps $Z_{\mathbf{E}}$ to X for each ω in the sample space. Then, Part 4 follows from Theorem 7.4.3 of Hsing and Eubank (2015). $\square$

PROOF OF THEOREM 6. The strong consistency stated in Part 2 is an immediate consequence of Lemma 12. For Part 1, to prove the continuity of μ, $t \in \mathcal{T}$. Let $\mathcal{K} \supset \mathcal{U}$ be compact. By B.3, $c := \sup_{p\in\mathcal{K}} \sup_{s\in\mathcal{T}} \mathbb{E}d_{\mathcal{M}}^2(p, X(s)) < \infty$. Thus,

$$
\begin{aligned}
&\bigl|F(\mu(t), s) - F(\mu(s), s)\bigr| \\
&\quad \le \bigl|F(\mu(t), t) - F(\mu(s), s)\bigr| + \bigl|F(\mu(t), s) - F(\mu(t), t)\bigr| \\
&\quad \le \sup_{p\in\mathcal{K}} \bigl|F(p, t) - F(p, s)\bigr| + 2c\mathbb{E}d_{\mathcal{M}}\bigl(X(s), X(t)\bigr) \\
&\quad \le 4c\mathbb{E}d_{\mathcal{M}}\bigl(X(s), X(t)\bigr).
\end{aligned}
$$

The continuity assumption of sample paths implies $\mathbb{E}d_{\mathcal{M}}(X(s), X(t)) \to 0$ as $s \to t$. Then by condition B.5, $d_{\mathcal{M}}(\mu(t), \mu(s)) \to 0$ as $s \to t$, and thus the continuity of μ follows. The uniform continuity follows from the compactness of $\mathcal{T}$. Given Lemma 11 and 12, the a.s. continuity of $\hat{\mu}$ can be derived in a similar way. The first statement of Part 4 is a corollary of Part 3, while the second statement follows from the first one and the compactness of $\mathcal{T}$. It remains to show Part 3 in order to conclude the proof, as follows.

Let $V_{t,i}(p) = \operatorname{Log}_p X_i(t)$ and $\gamma_{t,p}$ be the minimizing geodesic from $\mu(t)$ to $p \in \mathcal{M}$ at unit time. The first-order Taylor series expansion at $\mu(t)$ yields

$$
\begin{aligned}
&\mathcal{P}_{\hat{\mu}(t),\mu(t)} \sum_{i=1}^{n} V_{t,i}(\hat{\mu}(t)) \\
(16) \qquad &= \sum_{i=1}^{n} V_{t,i}(\mu(t)) + \sum_{i=1}^{n} \nabla_{\gamma'_{t,\hat{\mu}(t)}(0)} V_{t,i}(\mu(t)) + \Delta_t(\hat{\mu}(t))\gamma'_{t,\hat{\mu}(t)}(0) \\
&= \sum_{i=1}^{n} V_{t,i}(\mu(t)) - \sum_{i=1}^{n} H_t(\mu(t))\gamma'_{t,\hat{\mu}(t)}(0) + \Delta_t(\hat{\mu}(t))\gamma'_{t,\hat{\mu}(t)}(0),
\end{aligned}
$$

where an expression for Δ_t is provided in the proof of Lemma 10.

Since $\sum_{i=1}^{n} V_{t,i}(\hat{\mu}(t)) = \sum_{i=1}^{n} \operatorname{Log}_{\hat{\mu}(t)} X_i(t) = 0$, we deduce from (16) that

$$
(17) \qquad \frac{1}{n}\sum_{i=1}^{n} \operatorname{Log}_{\mu(t)} X_i(t) - \left(\frac{1}{n}\sum_{i=1}^{n} H_{t,i}(\mu(t)) - \frac{1}{n}\Delta_t(\hat{\mu}(t))\right) \operatorname{Log}_{\mu(t)} \hat{\mu}(t) = 0.
$$

By LLN, $\frac{1}{n}\sum_{i=1}^{n} H_{t,i}(\mu(t)) \to \mathbb{E}H_t(\mu(t))$ in probability, while $\mathbb{E}H_t(\mu(t))$ is invertible for all t by condition B.6. In light of Lemma 10, this result suggests that with probability tending to one, for all $t \in \mathcal{T}$, $\frac{1}{n}\sum_{i=1}^{n} H_{t,i}(\mu(t)) - \frac{1}{n}\Delta_t(\hat{\mu}(t))$ is invertible, and also

$$
\left(\frac{1}{n}\sum_{i=1}^{n} H_{t,i}(\mu(t)) - \frac{1}{n}\Delta_t(\hat{\mu}(t))\right)^{-1} = \{\mathbb{E}H_t(\mu(t))\}^{-1} + o_P(1),
$$

and according to (17),

$$
\operatorname{Log}_{\mu(t)} \hat{\mu}(t) = \{\mathbb{E}H_t(\mu(t))\}^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} \operatorname{Log}_{\mu(t)} X_i(t)\right) + o_P(1),
$$

where the $o_P(1)$ term does not depend on t. Given this, we can now conclude the proof of Part 3 by applying a central limit theorem in Hilbert space (Aldous (1976)) to establish that the process $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\mathbb{E}H_t(\mu(t))\}^{-1} \operatorname{Log}_{\mu(t)} X_i(t)$ converges to a Gaussian measure on tensor Hilbert space $\mathscr{T}(\mu)$ with covariance operator $\mathcal{C}(\cdot) = \mathbb{E}(\langle\!\langle V, \cdot \rangle\!\rangle_\mu V)$ for a random element $V(t) = \{\mathbb{E}H_t(\mu(t))\}^{-1} \operatorname{Log}_{\mu(t)} X(t)$ in the tensor Hilbert space $\mathscr{T}(\mu)$. $\square$

PROOF OF THEOREM 7. Note that

$$
\begin{aligned}
\Phi\hat{\mathcal{C}} - \mathcal{C} &= n^{-1}\sum(\Gamma \operatorname{Log}_{\hat{\mu}} X_i) \otimes (\Gamma \operatorname{Log}_{\hat{\mu}} X_i) - \mathcal{C} \\
&= n^{-1}\sum(\operatorname{Log}_{\mu} X_i) \otimes (\operatorname{Log}_{\mu} X_i) - \mathcal{C} \\
&\quad + n^{-1}\sum(\Gamma \operatorname{Log}_{\hat{\mu}} X_i - \operatorname{Log}_{\mu} X_i) \otimes (\operatorname{Log}_{\mu} X_i)
\end{aligned}
$$

$$
\begin{aligned}
&+ n^{-1}\sum(\mathrm{Log}_{\mu} X_i)\otimes(\Gamma\,\mathrm{Log}_{\hat\mu} X_i - \mathrm{Log}_{\mu} X_i)\\
&+ n^{-1}\sum(\Gamma\,\mathrm{Log}_{\hat\mu} X_i - \mathrm{Log}_{\mu} X_i)\otimes(\Gamma\,\mathrm{Log}_{\hat\mu} X_i - \mathrm{Log}_{\mu} X_i)\\
\equiv{}& A_1 + A_2 + A_3 + A_4.
\end{aligned}
$$

For A_2, it is seen that

$$
\begin{aligned}
|||A_2|||_{\mathrm{HS}}^2 \le \text{const.}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}&\big(\|\mathrm{Log}_{\mu} X_i\|_{\mu}^2 + \|\mathrm{Log}_{\mu} X_j\|_{\mu}^2\big)\\
&\times\big(\|\Gamma\mathrm{Log}_{\hat\mu} X_i - \mathrm{Log}_{\mu} X_i\|_{\mu}^2 + \|\Gamma\mathrm{Log}_{\hat\mu} X_j - \mathrm{Log}_{\mu} X_j\|_{\mu}^2\big).
\end{aligned}
$$

With smoothness of $d_{\mathcal{M}}^2$, continuity of μ and compactness of $\mathcal{T}$, one can show that $\sup_{t\in\mathcal{T}}\|H_t(\mu(t))\| < \infty$. By the uniform consistency of $\hat\mu$, with the same Taylor series expansion in (16) and the technique in the proof of Lemma 10, it can be established that $n^{-1}\sum_{i=1}^{n}\|\Gamma\mathrm{Log}_{\hat\mu} X_i - \mathrm{Log}_{\mu} X_i\|_{\mu}^2\|\mathrm{Log}_{\mu} X_i\|_{\mu}^2 \le \text{const.}(1 + o_P(1))\sup_{t\in\mathcal{T}} d_{\mathcal{M}}^2(\hat\mu(t),\mu(t))$. Also note that by LLN, $n^{-1}\sum_{j=1}^{n}\|\mathrm{Log}_{\mu} X_j\|_{\mu}^2 = O_P(1)$. Then, with Part 4 of Theorem 6,

$$
|||A_2|||_{\mathrm{HS}}^2 \le \text{const.}\{4 + o_P(1) + O_P(1)\}\sup_{t\in\mathcal{T}} d_{\mathcal{M}}^2\big(\hat\mu(t),\mu(t)\big) = O_P(1/n).
$$

Similar calculation shows that $|||A_3|||_{\mathrm{HS}}^2 = O_P(1/n)$ and $|||A_4|||_{\mathrm{HS}}^2 = O_P(1/n^2)$. Now, by Dauxois, Pousse and Romain (1982), $\|n^{-1}\sum(\mathrm{Log}_{\mu} X_i)\otimes(\mathrm{Log}_{\mu} X_i) - \mathcal{C}\|_{\mathrm{HS}}^2 = O_P(1/n)$. Thus, $\|\Phi\hat{\mathcal{C}} - \mathcal{C}\|_{\mathrm{HS}}^2 = O_P(1/n)$. According to Parts 1 & 5 of Proposition 2, $\hat\lambda_k$ are also eigenvalues of $\Phi\hat{\mathcal{C}}$. The results for $\hat\lambda_k$ and (J,δ_j) follow from Bosq (2000). Those for $(\hat J,\hat\delta_j)$ are due to $\sup_{k\ge1}|\hat\lambda_k - \lambda_k| \le |||\hat{\mathcal{C}}\ominus_{\Phi}\mathcal{C}|||_{\mathrm{HS}}$. $\square$

PROOF OF THEOREM 8. In this proof, both $o_P(\cdot)$ and $O_P(\cdot)$ are understood to be uniform for the class $\mathcal{F}$. Let $\check\beta = \mathrm{Exp}_{\mu}\sum_{k=1}^{K}\hat b_k\Gamma\hat\phi_k$. Then

$$
d_{\mathcal{M}}^2(\hat\beta,\beta) \le 2d_{\mathcal{M}}^2(\hat\beta,\check\beta) + 2d_{\mathcal{M}}^2(\check\beta,\beta).
$$

The first term is of order $O_p(1/n)$ uniform for the class $\mathcal{F}$, according to a technique similar to the one in the proof of Lemma 10, as well as Theorem 6 (note that the results in Theorem 6 are uniform for the class $\mathcal{F}$). Then the convergence rate is established if one can show that

$$
d_{\mathcal{M}}^2(\check\beta,\beta) = O_P\big(n^{-\frac{2\varrho-1}{4\alpha+2\varrho+2}}\big),
$$

which follows from

$$
\Big\|\sum_{k=1}^{K}\hat b_k\Gamma\hat\phi_k - \sum_{k=1}^{\infty} b_k\phi_k\Big\|_{\mu}^2 = O_P\big(n^{-\frac{2\varrho-1}{4\alpha+2\varrho+2}}\big) \tag{18}
$$

and Proposition 4. It remains to show (18).

We first observe that because $b_k \le Ck^{-\varrho}$,

(19)
$$\left\| \sum_{k=1}^{K} \hat{b}_k \Gamma \hat{\phi}_k - \sum_{k=1}^{\infty} b_k \phi_k \right\|_{\mu}^{2} \le 2 \left\| \sum_{k=1}^{K} \hat{b}_k \Gamma \hat{\phi}_k - \sum_{k=1}^{K} b_k \phi_k \right\|_{\mu}^{2} + O\big(K^{-2\varrho+1}\big).$$

Define

$$A_1 = \sum_{k=1}^{K} (\hat{b}_k - b_k)\phi_k, \qquad A_2 = \sum_{k=1}^{K} b_k (\Gamma \hat{\phi}_k - \phi_k),$$
$$A_3 = \sum_{k=1}^{K} (\hat{b}_k - b_k)(\Gamma \hat{\phi}_k - \phi_k).$$

Then

$$\left\| \sum_{k=1}^{K} \hat{b}_k \Gamma \hat{\phi}_k - \sum_{k=1}^{K} b_k \phi_k \right\|_{\mu}^{2} \le 2\|A_1\|_{\mu}^2 + 2\|A_2\|_{\mu}^2 + 2\|A_3\|_{\mu}^2.$$

It is clear that the term A_3 is asymptotically dominated by A_1 and A_2. Note that the compactness of X in condition C.2 implies $\mathbb{E}\|\mathrm{Log}_{\mu} X\|_{\mu}^4 < \infty$. Then, by Theorem 7, for A_2, we have the bound

$$\|A_2\|_{\mu}^2 \le 2 \sum_{k=1}^{K} b_k^2 \|\Gamma \hat{\phi}_k - \phi_k\|_{\mu}^2 = \left\{ \right.$$

$K \asymp n^{1/(4\alpha+2\varrho+2)}$, one can conclude that

$$\|A_1\|_\mu^2 = \sum_{k=1}^{K} \mu\Delta \Bigl(\frac{\lambda_k - \hat{\lambda}_k}{\hat{\lambda}_k} b_k$$

$$
\begin{aligned}
&= -(I_\mu + \mathcal{C}_\rho^+ \Delta)^{-1} \mathcal{C}_\rho^+ \Delta \mathrm{Log}_\mu \beta - (I_\mu + \mathcal{C}_n^+ \Delta)^{-1} \mathcal{C}_\rho^+ \Delta (\mathcal{C}_\rho^+ \chi_n - \mathrm{Log}_\mu \beta) \\
&\equiv A_{n211} + A_{n212}.
\end{aligned}
$$

By Theorem 7, $|||\Delta|||_\mu = O_P(1/n)$. Also, one can see that $|||(I_\mu + \mathcal{C}_\rho^+ \Delta)^{-1}|||_\mu = O_P(1)$, with the assumption that $\rho^{-1}/n = o(1)$. Also, $|||(I_\mu + \mathcal{C}_\rho^+ \Delta)^{-1} \mathcal{C}_\rho^+ \Delta|||_{op} = O_P(\rho^{-2}/n)$. Using the similar technique in Hall and Horowitz (2005), we can show that $\|\mathcal{C}_\rho^+ \chi_n - \mathrm{Log}_\mu \beta\|_\mu^2 = O_P(n^{-(2\varrho-1)/(2\varrho+\alpha)})$, and hence conclude that $\|A_{n212}\|_\mu^2 = O_P(n^{-(2\varrho-1)/(2\varrho+\alpha)})$. For A_{n211},

$$
\begin{aligned}
\|A_{n211}\|_\mu^2 &= \|(I_\mu + \mathcal{C}_n^+ \Delta)^{-1} \mathcal{C}_n^+ \Delta \mathrm{Log}_\mu \beta\|_\mu^2 \\
&\le |||(I_\mu + \mathcal{C}_n^+ \Delta)^{-1}|||_{op}^2 |||\mathcal{C}_n^+ \Delta|||_{op}^2 \|\mathrm{Log}_\mu \beta\|_\mu^2 \\
&= O_P(n^{-(2\varrho-\alpha)/(2\varrho+\alpha)}).
\end{aligned}
$$

Combining all results above, we deduce that $\|\Gamma(\hat{\mathcal{C}}^+ \hat{\chi}) - \mathrm{Log}_\mu \beta\|_\mu^2 = O_P(n^{-(2\varrho-\alpha)/(2\varrho+\alpha)})$ and thus

$$
d_{\mathcal{M}}^2(\mathrm{Exp}_\mu \Gamma(\hat{\mathcal{C}}^+ \hat{\chi}), \beta) = O_P(n^{-(2\varrho-\alpha)/(2\varrho+\alpha)}),
$$

according to condition C.2 and Proposition 4. □

APPENDIX D: ANCILLARY LEMMAS

LEMMA 10. $\sup_{t \in \mathcal{T}} n^{-1} \|\Delta_t(\hat{\mu}(t))\| = o_P(1)$, *where* Δ_t *is as in* (16).

PROOF. With the continuity of μ and compactness of $\mathcal{T}$, the existence of local smooth orthonormal frames (e.g., Proposition 11.17 of Lee (2013)) suggests that we can find a finite sequence $\mathcal{T}_1, \dots, \mathcal{T}_m$ for $\mathcal{T}$ such that there exists a smooth orthonormal frame $b_{j,1}, \dots, b_{j,d}$ for the jth piece $\{\mu(t) : t \in \mathrm{cl}(\mathcal{T}_j)\}$ of μ, where $\mathrm{cl}(A)$ denotes the topological closure of a set A. For fixed $t \in \mathcal{T}_j$, by mean value theorem, it can be shown that

$$
\Delta_t(\hat{\mu}(t))U = \sum_{r=1}^{d} \sum_{i=1}^{n} (\mathcal{P}_{\gamma_{t,\hat{\mu}(t)}(\theta_t^{r,j}),\mu(t)} \nabla_U W_{t,i}^{r,j}(\gamma_{t,\hat{\mu}(t)}(\theta_t^{r,j})) - \nabla_U W_{t,i}^{r,j}(\mu(t))) \tag{20}
$$

for $\theta_t^{r,j} \in [0,1]$ and $W_{t,i}^{r,j} = \langle V_{t,i}, e_t^{r,j} \rangle e_t^{r,j}$, where $e_t^{1,j}, \dots, e_t^{d,j}$ is the orthonormal frame extended by parallel transport of $b_{j,1}(\mu(t)), \dots, b_{j,d}(\mu(t))$ along minimizing geodesic.

Take $\epsilon = \epsilon_n \downarrow 0$ as $n \to \infty$. For each j, by the same argument of Lemma 3 of Kendall and Le (2011), together with continuity of μ and the continuity f the frame $b_{j,1}, \dots, b_{j,d}$, e can ch e a c ntin iti e ρ_t^j ch that, $\hat{\mu}(t) \in B(\mu(t), \rho_t^j)$ and f $p \in B(\mu(t), \rho_t^j)$ here $B(q, \rho)$ den te the ball n $\mathcal{M}$ centered at q ith radi ρ,

$$
\begin{aligned}
&\|\mathcal{P}_{p,\mu(t)} \nabla W_{t,i}^{r,j}(p) - \nabla W_{t,i}^{r,j}(\mu(t))\| \\
&\quad \le (1 + 2\epsilon\rho_t^j) \qquad \|\mathcal{P}_{q,\mu(t)} \nabla V_{t,i}(q) - \nabla V_{t,i}(\mu(t))\| \\
&\qquad\qquad {}_{q \in B(\mu(t), \rho_t^j)} \\
&\qquad + 2\epsilon(\|V_{t,i}(\mu(t))\| + \rho_t^j \|\nabla V_{t,i}(\mu(t))\|).
\end{aligned}
$$

In the ab e, p la a le f $\gamma_{t,\hat{\mu}(t)}(\theta_t^{r,j})$ in (20). Let $\rho^j = \mathrm{ma}\ \{\rho_t : t \in \mathrm{cl}(\mathcal{T}_j)\}$ and $\rho_{\mathrm{ma}} = \mathrm{ma}_{\ j}\ \rho^j$. We then ha e

$$
\begin{aligned}
&{}_{t \in \mathcal{T}} \|\Delta_t(\hat{\mu}(t))\| \\
&\quad \le \mathrm{ma}_{j} \ {}_{t \in \mathcal{T}_j} \|\Delta_t(\hat{\mu}(t))\| \\
&\quad = \sum_{r=1}^{d} \sum_{i=1}^{n} \mathrm{ma}_{j} \ {}_{t \in \mathcal{T}_j} \|\mathcal{P}_{\gamma_{t,\hat{\mu}(t)}(\theta_t^{r,j}),\mu(t)} \nabla_U W_{r,i}^{t,j}(\gamma_{t,\hat{\mu}(t)}(\theta_t^{r,j})) - \nabla_U W_{r,i}^{t,j}(\mu(t))\| \\
&\quad \le d(1 + 2\epsilon\rho_{\mathrm{ma}})m \sum_{i=1}^{n} {}_{t \in \mathcal{T}} {}_{q \in B(\mu(t), \rho_{\mathrm{ma}})} \|\mathcal{P}_{q,\mu(t)} \nabla V_{t,i}(q) - \nabla V_{t,i}(\mu(t))\| \qquad (21) \\
&\qquad + 2d\epsilon \ {}^{n}
\end{aligned}
$$

so

$$\frac{1}{n}\sum_{i=1}^{n}\sup_{t\in\mathcal{T}}\|V_{t,i}(\mu(t))\| = O_P(1). \tag{24}$$

For the second term in (22), the compactness of $\mathcal{T}$, the Lipschitz condition of B.7 and smoothness of $d_{\mathcal{M}}$ also imply that $\mathbb{E}\sup_{t\in\mathcal{T}}\|\nabla V_{t,i}(\mu(t))\| = \mathbb{E}\sup_{t\in\mathcal{T}}\|H_t(\mu(t))\| < \infty$. Consequently, by LLN,

$$\frac{1}{n}\sum_{i=1}^{n}\sup_{t\in\mathcal{T}}\|\nabla V_{t,i}(\mu(t))\| = O_P(1). \tag{25}$$

Combining (23), (24) and (25), with $\epsilon = \epsilon_n \downarrow 0$, one concludes that $\sup_{t\in\mathcal{T}} n^{-1}\|\Delta_t(p)\| = o_P(1)$. □

LEMMA 11. *Suppose conditions A.1 and B.1–B.3 hold. For any compact subset $\mathcal{K}\subset\mathcal{M}$, one has*

$$\sup_{p\in\mathcal{K}}\sup_{t\in\mathcal{T}}|F_n(p,t) - F(p,t)| = o_{\text{a.s.}}(1).$$

PROOF. By applying the uniform SLLN to $n^{-1}\sum_{i=1}^{n} d_{\mathcal{M}}(X_i(t), p_0)$, for a given $p_0\in\mathcal{K}$,

$$\begin{aligned}\sup_{p\in\mathcal{K}}\sup_{t\in\mathcal{T}}\frac{1}{n}\sum_{i=1}^{n} d_{\mathcal{M}}(X_i(t), p) &\le \sup_{t\in\mathcal{T}}\frac{1}{n}\sum_{i=1}^{n} d_{\mathcal{M}}(X_i(t), p_0) + \sup_{p\in\mathcal{K}} d_{\mathcal{M}}(p_0, p)\\ &\le \sup_{t\in\mathcal{T}}\mathbb{E}d_{\mathcal{M}}(X(t), p_0) + \operatorname{diam}(\mathcal{K}) + o_{\text{a.s.}}(1).\end{aligned}$$

Therefore, there exists a set $\Omega_1\subset\Omega$ such that $\Pr(\Omega_1) = 1$, $N_1(\omega) < \infty$ and for all $n \ge N_1(\omega)$,

$$\sup_{p\in\mathcal{K}}\sup_{t\in\mathcal{T}}\frac{1}{n}\sum_{i=1}^{n} d_{\mathcal{M}}(X_i(t), p) \le \sup_{t\in\mathcal{T}}\mathbb{E}d_{\mathcal{M}}(X(t), p_0) + \operatorname{diam}(\mathcal{K}) + 1 := c_1 < \infty,$$

since $\sup_{t\in\mathcal{T}}\mathbb{E}d_{\mathcal{M}}(X(t), p_0) < \infty$ by condition B.3. Fix $\epsilon > 0$. By the inequality $|d^2_{\mathcal{M}}(x,p) - d^2_{\mathcal{M}}(x,q)| \le \{d_{\mathcal{M}}(x,p) + d_{\mathcal{M}}(x,q)\}d_{\mathcal{M}}(p,q)$, for all $n \ge N_1(\omega)$ and $\omega\in\Omega_1$,

$$\sup_{p,q\in\mathcal{K}:d_{\mathcal{M}}(p,q)<\delta_1}\sup_{t\in\mathcal{T}}|F_{n,\omega}(p,t) - F_{n,\omega}(q,t)| \le 2c_1\delta_1 = \epsilon/3$$

with $\delta_1 := \epsilon/(6c_1)$. Now, let $\delta_2 > 0$ be chosen such that $\sup_{t\in\mathcal{T}}|F(p,t) - F(q,t)| < \epsilon/3$ if $p, q\in\mathcal{K}$ and $d_{\mathcal{M}}(p,q) < \delta_2$. Suppose $\{p_1,\ldots,p_r\}\subset\mathcal{K}$ is a δ-net in $\mathcal{K}$ with $\delta := \min\{\delta_1,\delta_2\}$. Applying uniform SLLN again, there exists a set Ω_2 such that $\Pr(\Omega_2) = 1$, $N_2(\omega) < \infty$ for all $\omega\in\Omega_2$, and

$$\max_{j=1,\ldots,r}\sup_{t\in\mathcal{T}}|F_{n,\omega}(p_j,t) - F(p_j,t)| < \epsilon/3$$

for all $n \geq N_2(\omega)$ with $\omega \in \Omega_2$. Then, for all $\omega \in \Omega_1 \cap \Omega_2$, for all $n \geq \max\{N_1(\omega), N_2(\omega)\}$, we have

$$\begin{aligned}
&\sup_{p\in\mathcal{K}}\sup_{t\in\mathcal{T}}\left|F_{n,\omega}(p,t) - F(p,t)\right| \\
&\quad \leq \sup_{p\in\mathcal{K}}\sup_{t\in\mathcal{T}}\left|F_{n,\omega}(p) - F_{n,\omega}(u_p)\right| + \sup_{p\in\mathcal{K}}\sup_{t\in\mathcal{T}}\left|F_{n,\omega}(u_p,t) - F(u_p,t)\right| \\
&\qquad + \sup_{p\in\mathcal{K}}\sup_{t\in\mathcal{T}}\left|F(u_p,t) - F(p,t)\right| \\
&\quad < \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon,
\end{aligned}$$

and this concludes the proof. $\square$

LEMMA 12. *Assume conditions A.1 and B.1–B.5 hold. Given any $\epsilon > 0$, there exists $\Omega' \subset \Omega$ such that $\Pr(\Omega') = 1$ and for all $\omega \in \Omega'$, $N(\omega) < \infty$ and for all $n \geq N(\omega)$, $\sup_{t\in\mathcal{T}} d_{\mathcal{M}}(\hat{\mu}_\omega(t), \mu(t)) < \epsilon$.*

PROOF. Let $c(t) = F(\mu(t), t) = \min\{F(p,t) : p \in \mathcal{M}\}$ and $\mathcal{N}(t) := \{p : d_{\mathcal{M}}(p, \mu(t)) \geq \epsilon\}$. It is sufficient to show that there exist $\delta > 0$ and $N(\omega) < \infty$ for all $\omega \in \Omega'$, such that for all $n \geq N(\omega)$,

$$\sup_{t\in\mathcal{T}}\left\{F_{n,\omega}\big(\mu(t),t\big) - c(t)\right\} \leq \delta/2 \quad \text{and} \quad \inf_{t\in\mathcal{T}}\left\{\inf_{p\in\mathcal{N}(t)} F_{n,\omega}(p,t) - c(t)\right\} \geq \delta.$$

This is because the above inequalities suggest that for all $t \in \mathcal{T}$, $\inf\{F_{n,\omega}(p,t) : p \in \mathcal{M}\}$ is not attained at p with $d_{\mathcal{M}}(p,\mu(t)) \geq \epsilon$, and hence $\sup_{t\in\mathcal{T}} d_{\mathcal{M}}(\hat{\mu}_\omega(t), \mu(t)) < \epsilon$.

Let $\mathcal{U} = \{\mu(t) : t \in \mathcal{T}\}$. We first show that there exist a compact set $\mathcal{A} \supset \mathcal{U}$ and $N_1(\omega) < \infty$ for some $\Omega_1 \subset \Omega$ such that $\Pr(\Omega_1) = 1$, and both $F(p,t)$ and $F_{n,\omega}(p,t)$ are greater than $c(t) + 1$ for all $p \in \mathcal{M}\backslash\mathcal{A}$, $t \in \mathcal{T}$ and $n \geq N_1(\omega)$. This is trivially true when $\mathcal{M}$ is compact, by taking $\mathcal{A} = \mathcal{M}$. Now assume $\mathcal{M}$ is noncompact. By the inequality $d_{\mathcal{M}}(x,q) \geq |d_{\mathcal{M}}(q,y) - d_{\mathcal{M}}(y,x)|$, one has

$$\begin{aligned}
&\mathbb{E}d^2_{\mathcal{M}}\big(X(t),q\big) \\
&\quad \geq \mathbb{E}\{d^2_{\mathcal{M}}\big(q,\mu(t)\big) + d^2_{\mathcal{M}}\big(X(t),\mu(t)\big) - 2d_{\mathcal{M}}\big(q,\mu(t)\big)d_{\mathcal{M}}\big(X(t),\mu(t)\big)\},
\end{aligned}$$

and by Cauchy–Schwarz inequality,

$$F(q,t) \geq d^2_{\mathcal{M}}\big(q,\mu(t)\big) + F\big(\mu(t),t\big) - 2d_{\mathcal{M}}\big(q,\mu(t)\big)\{F\big(\mu(t),t\big)\}^{1/2}.$$

Similarly,

$$F_{n,\omega}(q,t) \geq d^2_{\mathcal{M}}\big(q,\mu(t)\big) + F_{n,\omega}\big(\mu(t),t\big) - 2d_{\mathcal{M}}\big(q,\mu(t)\big)\{F_{n,\omega}\big(\mu(t),t\big)\}^{1/2}.$$

Now, we take q at a sufficiently large distance Δ from $\mathcal{U}$ such that $F(q,t) > c(t) + 1$ on $\mathcal{M}\backslash\mathcal{A}$ for all t, where $\mathcal{A} := \overline{\{q : d_{\mathcal{M}}(q,\mathcal{U}) \leq \Delta\}}$ (Heine–Borel property yield c m act ne f $\mathcal{A}$, ince it i b nded and cl ed). Since $F_{n,\omega}(\mu(t),t)$ c n er ge t $F(\mu(t),t)$ nif rml n $\mathcal{T}$ a. . b Lemma 11, e can nd a et $\Omega_1 \subset \Omega$ ch that P $(\Omega_1) = 1$ and $N_1(\omega) < \infty$ f r $\omega \in \Omega_1$, and $F_{n,\omega}(q,t) > c(t) + 1$ n $\mathcal{M}\backslash\mathcal{A}$ f r all t and $n \geq N_1(\omega)$.

Finall , let $\mathcal{A}_\epsilon(t) := \{p \in \mathcal{A} : d_{\mathcal{M}}(p,\mu(t)) \geq \epsilon\}$ and $c_\epsilon(t) := \min\{F(p,t) : p \in \mathcal{A}_\epsilon\}$. Then $\mathcal{A}_\epsilon(t)$ i c m act and b c nditi n B.5, $\inf_t\{c_\epsilon(t) - c(t)\} > 2\delta > 0$ f r me c n tant δ. B Lemma 11, ne can nd a et $\Omega_2 \subset \Omega$ ith P $(\Omega_2) = 1$ and $N_2(\omega) < \infty$ f r $\omega \in \Omega_2$, ch that f r all $n \geq N_2(\omega)$, (i) $_t\{F_{n,\omega}(\mu(t),t) - c(t)\} \leq \delta/2$ and (ii) $\inf_t \inf_{p\in\mathcal{A}_\epsilon(t)}\{F_{n,\omega}(p,t) - c(t)\} > \delta$. Since $_t\{F_{n,\omega}(p,t) - c(t)\} > 1$ n $\mathcal{M}\backslash\mathcal{A}$ f r all $n \geq N_1(\omega)$ ith $\omega \in \Omega_1$, e c ncl de that $\inf_t\{F_{n,\omega}(p,t) - c(t)\} > \min\{\delta, 1\}$ f r all $p \in \mathcal{A}_\epsilon \cup (\mathcal{M}\backslash\mathcal{A})$ if $n \geq$ ma $\{N_1(\omega), N_2(\omega)\}$ f r $\omega \in \Omega_1 \cap \Omega_2$. The r f i c m leted b n ting that $\Omega_1 \cap \Omega_2$ can er e the Ω' e are l king f r. □

REFERENCES

AFSARI, B. (2011). Riemannian L^p center f ma : E i tence, ni ene , and c n e it . *Proc. Amer. Math. Soc.* **139** 655 673. MR2736346

ALDOUS, D. J. (1976). A characteri ati n f Hilbert ace ing the central limit the rem. *J. Lond. Math. Soc.* (2) **14** 376 380. MR0443017

ARSIGNY, V., FILLARD, P., PENNEC, X. and AYACHE, N. (2006/07). Ge metric mean in a n el ect r ace tr ct re n mmetric iti e-de nite matrice . *SIAM J. Matrix Anal. Appl.* **29** 328 347. MR2288028

BALAKRISHNAN, A. V. (1960). E timati n and detecti n the r f r m lti le t cha tic r ce e . *J. Math. Anal. Appl.* **1** 386 410. MR0131323

BHATTACHARYA, R. and PATRANGENARU, V. (2003). Large am le the r f intrin ic and e trin ic am le mean n manif ld . I. *Ann. Statist.* **31** 1 29. MR1962498

BINDER, J. R., FROST, J. A., HAMMEKE, T. A., COX, R. W., RAO, S. M. and PRIETO, T. (1997). H man brain lang age area identi ed b f ncti nal magnetic re nance imaging. *J. Neurosci.* **17** 353 362.

BOSQ, D. (2000). *Linear Processes in Function Spaces*. *Lecture Notes in Statistics* **149**. S ringer, Ne Y rk. MR1783138

CARDOT, H., FERRATY, F. and SARDA, P. (2003). S line e timat r f r the f ncti nal linear m del. *Statist. Sinica* **13** 571 591. MR1997162

CARDOT, H., MAS, A. and SARDA, P. (2007). CLT in f ncti nal linear regre i n m del . *Probab. Theory Related Fields* **138** 325 361. MR2299711

CHEN, D. and M LLER, H.-G. (2012). N nlinear manif ld re re entati n f r f ncti nal data. *Ann. Statist.* **40** 1 29. MR3013177

CHENG, G., HO, J., SALEHIAN, H. and VEMURI, B. C. (2016). Rec r i e c m tati n f the Frechet mean n n n- iti el c r ed Riemannian manif ld ith a licati n . In *Riemannian Computing in Computer Vision* 21 43. S ringer, Cham. MR3444345

CORNEA, E., ZHU, H., KIM, P. and IBRAHIM, J. G. (2017). Regre i n m del n Riemannian mmetric ace . *J. R. Stat. Soc. Ser*

DAYAN, E. and COHEN, L. G. (2011). Neuroplasticity subserving motor skill learning.

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

PARK, J. E., JUNG, S. C., RYU, K. H., OH, J. Y., KIM, H. S., CHOI, C. G., KIM, S. J. and SHIM, W. H. (2017). Differences in dynamic and static functional connectivity between young and elderly healthy adults. *Neuroradiol.* **59** 781–789.

PETERSEN, A. and MÜLLER, H.-G. (2019). Fréchet regression for random objects with Euclidean predictors. *Ann. Statist.* **47** 691–719. MR3909947

PHANA, K. L., WAGER, T., T

DEPARTMENT OF STATISTICS
AND APPLIED PROBABILITY
NATIONAL UNIVERSITY OF SINGAPORE
SINGAPORE 117546
SINGAPORE
E-MAIL: tal @n .ed . g

DEPARTMENT OF PROBABILITY AND STATISTICS
SCHOOL OF MATHEMATICAL SCIENCES
CENTER FOR STATISTICAL SCIENCE
PEKING UNIVERSITY
BEIJING
CHINA
E-MAIL: f a @math. k .ed .cn