*Submitted to the Annals of Statistics*
arXiv: arXiv:1705.06226

# PRINCIPAL COMPONENT ANALYSIS FOR FUNCTIONAL DATA ON RIEMANNIAN MANIFOLDS AND SPHERES

By Xiongtao Dai*† and Hans-Georg Müller†

*University of California, Davis*

Functional data analysis on nonlinear manifolds has drawn recent interest. Sphere-valued functional data, which are encountered for example as movement trajectories on the surface of the earth, are an important special case. We consider an intrinsic principal component analysis for smooth Riemannian manifold-valued functional data and study its asymptotic properties. Riemannian functional principal component analysis (RFPCA) is carried out by first mapping the manifold-valued data through Riemannian logarithm maps to tangent spaces around the time-varying Fréchet mean function, and then performing a classical multivariate functional principal component analysis on the linear tangent spaces. Representations of the Riemannian manifold-valued functions and the eigenfunctions on the original manifold are then obtained with exponential maps. The tangent-space approximation through functional principal component analysis is shown to be well-behaved in terms of controlling the residual variation if the Riemannian manifold has nonnegative curvature. Specifically, we derive a central limit theorem for the mean function, as well as root-$n$ uniform convergence rates for other model components, including the covariance function, eigenfunctions, and functional principal component scores. Our applications include a novel framework for the analysis of longitudinal compositional data, achieved by mapping longitudinal compositional data to trajectories on the sphere, illustrated with longitudinal fruit fly behavior patterns. Riemannian functional principal component analysis is shown to be superior in terms of trajectory recovery in comparison to an unrestricted functional principal component analysis in applications and simulations and is also found to produce principal component scores that are better predictors for classification compared to traditional functional functional principal component scores.

**1. Introduction.** Methods for functional data analysis in a linear function space (Wang, Chiou and Müller 2016) or on a nonlinear submanifold

*Corresponding author

†Research supported by NSF grants DMS-1407852 and DMS-1712864. We thank FlightAware for the permission to use the flight data in Section 5.

*MSC 2010 subject classifications:* Primary 62G05; secondary 62G20, 62G99

*Keywords and phrases:* Compositional Data, Dimension Reduction, Functional Data Analysis, Functional Principal Component Analysis, Principal Geodesic Analysis, Riemannian Manifold, Trajectory, Central Limit Theorem, Uniform Convergence

(Lin and Yao 2017) have been much studied in recent years. Growth curve data (Ramsay and Silverman 2005) are examples of functions in a linear space, while densities (Kneip and Utikal 2001) and longitudinal shape profiles (Kent et al. 2001) lie on nonlinear manifolds. Since random functions usually lie in an intrinsically infinite dimensional linear or nonlinear space, dimension reduction techniques, in particular functional principal component analysis, play a central role in representing the random functions (Petersen and Müller 2016a) and in other supervised/unsupervised learning tasks. Methods for analyzing non-functional data on manifolds have also been well developed over the years, such as data on spheres (Fisher, Lewis and Embleton 1987), Kendall's shape spaces (Kendall et al. 2009; Huckemann, Hotz and Munk 2010), and data on other classical Riemannian manifolds (Cornea et al. 2017); for a comprehensive overview of nonparametric methods for data on manifolds see Patrangenaru and Ellingson (2015). Specifically, versions of principal component analysis methods that adapt to the Riemannian or spherical geometry, such as principal geodesic analysis (Fletcher et al. 2004) or nested spheres (Huckemann and Eltzner 2016), have substantially advanced the study of data on manifolds.

However, there is much less known about functional data, i.e., samples of random trajectories, that assume values on manifolds, even though such data are quite common. An example is Telschow, Huckemann and Pierrynowski (2016), who considered the extrinsic mean function and warping for functional data lying on SO(3). Examples of data lying on a Euclidean sphere include geographical data (Zheng 2015) on $S^2$, directional data on $S^1$ (Mardia and Jupp 2009), and square-root compositional data (Huckemann and Eltzner 2016), for which we will study longitudinal/functional versions in Section 4. Sphere-valued functional data naturally arise when data on a sphere have a time component, such as in recordings of airplane flight paths or animal migration trajectories. Our main goal is to extend and study the dimension reduction that is afforded by the popular functional principal component analysis (FPCA) in Euclidean spaces to the case of samples of smooth curves that lie on a smooth Riemannian manifold, taking into account the underlying geometry.

Specifically, Riemannian Functional Principal Component Analysis (RFPCA) is shown to serve as an intrinsic principal component analysis of Riemannian manifold-valued functional data. Our approach provides a theoretical framework and differs from existing methods for functional data analysis that involve manifolds, e.g., a proposed smooth principal component analysis for functions whose domain is on a two-dimensional manifold, motivated by signals on the cerebral cortex (Lila, Aston and Sangalli 2016),

nonlinear manifold representation of $L^2$ random functions themselves lying on a low-dimensional but unknown manifold (Chen and Müller 2012), or functional predictors lying on a smooth low-dimensional manifold (Lin and Yao 2017). While there have been closely related computing and application oriented proposals, including functional principal components on manifolds in discrete time, a systematic approach and theoretical analysis within a statistical modeling framework does not exist yet, to the knowledge of the authors. Specifically, in the engineering literature, dimension reduction for Riemannian manifold-valued motion data has been considered (Rahman et al. 2005; Tournier et al. 2009; Anirudh et al. 2015), where for example in the latter paper the time axis is discretized, followed by multivariate dimension reduction techniques such as principal component analysis on the logarithm mapped data; these works emphasize specific applications and do not provide theoretical justifications. The basic challenge is to adapt inherently linear methods such as functional principal component analysis (FPCA) to curved spaces.

RFPCA is an approach intrinsic to a given smooth Riemannian manifold and proceeds through time-varying geodesic submanifolds on the given manifold by minimizing total residual variation as measured by geodesic distance on the given manifold. Since the mean of manifold-valued functions in the $L^2$ sense is usually extrinsic, i.e., does not lie itself on the manifold in general, for an intrinsic analysis the mean function needs to be carefully defined, for which we adopt the intrinsic Fréchet mean, assuming that it is uniquely determined. RFPCA is implemented by first mapping the manifold valued trajectories that constitute the functional data onto the linear tangent spaces using logarithm maps around the mean curve at a current time $t$ and then carrying out a regular FPCA on the linear tangent space of log-mapped data. Riemannian functional principal component (RFPC) scores, eigenfunctions, and finite-truncated representations of the log-mapped data are defined on the tangent spaces and finite-truncated representations of the data on the original manifold are then obtained by applying exponential maps to the log-mapped finite-truncated data. We develop implementation and theory for RFPCA and provide additional discussion for the important special case where the manifold is the Euclidean sphere, leading to Spherical Principal Component Analysis (SFPCA), in Section 2 below, where also estimation methods are introduced. SFPCA differs from principal component analysis on spheres (e.g., Jung, Dryden and Marron 2012; Huckemann and Eltzner 2016), as these are not targeting functional data that consist of a sample of time-dependent trajectories.

Theoretical properties of the proposed RFPCA are discussed in Sec-

tion 3. Proposition 1 states that the residual variance for a certain finite-dimensional time-varying geodesic manifold representation under the geodesic distance is upper bounded by the $L^2$ residual variance of the log-mapped data. The classical $L^2$ residual variance can be easily calculated and provides a convenient upper bound of the residual variance under the geodesic distance. A uniform central limit theorem for Riemannian manifold-valued functional data is presented in Theorem 1. Corollary 1 and Theorem 2 provide asymptotic supremum convergence rates of the sample-based estimates of the mean function, covariance function, and eigenfunctions to their population targets under proper metrics, and the convergence rate for the sample FPC scores to their population targets is in Theorem 3. We also provide a consistency result for selecting the number of components used according to a criterion that is analogous to the fraction of variance explained (FVE) criterion in Corollary 3. Proofs are in the Appendix and the Supplementary Materials.

An important application for SFPCA is the principal component analysis for longitudinal compositional data, which we will introduce in Section 4, where we show that longitudinal compositional data can be mapped to functional trajectories that lie on a Euclidean sphere. We demonstrate a specific application for longitudinal compositional data in Section 5 for behavioral patterns for fruit flies that are mapped to $S^4$, where we show that the proposed SFPCA outperforms conventional FPCA. A second example concerns a sample of flight trajectories from Hong Kong to London, which are functional data on $S^2$. In this second example SFPCA also outperforms more conventional approaches and illustrates the interpretability of the proposed RFPCA. For the flight trajectory example, we demonstrate that the FPC scores produced by the RFPCA encode more information for classification purposes than those obtained by the classical FPCA in an $L^2$ functional space. These data examples are complemented by simulation studies reported in Section 6.

## 2. Functional principal component analysis for random trajectories on a Riemannian manifold.

2.1. *Preliminaries.* We briefly review the basics of Riemannian geometry essential for the study of Riemannian manifold-valued functions; for further details, see, e.g., Chavel (2006). For a smooth manifold $\mathcal{M}$ with dimension $d$ and tangent spaces $T_p\mathcal{M}$ at $p \in \mathcal{M}$, a *Riemannian metric* on $\mathcal{M}$ is a family of inner products $g_p : T_p\mathcal{M} \times T_p\mathcal{M} \to \mathbb{R}$ that varies smoothly over $p \in \mathcal{M}$. Endowed with this Riemannian metric, $(\mathcal{M}, g)$ is a *Riemannian manifold*. The *geodesic distance* $d_{\mathcal{M}}$ is the metric on $\mathcal{M}$ induced by $g$. A *geodesic* is a

locally length minimizing curve. The *exponential map* at $p \in \mathcal{M}$ is defined as $\exp_p(v) = \gamma_v(1)$ where $v \in T_p\mathcal{M}$ is a tangent vector at $p$, and $\gamma_v$ is a unique geodesic with initial location $\gamma_v(0) = p$ and velocity $\gamma_v'(0) = v$. If $(\mathcal{M}, d_{\mathcal{M}})$ is a complete metric space, then $\exp_p$ is defined on the entire tangent space $T_p\mathcal{M}$. The exponential map $\exp_p$ is a diffeomorphism in a neighborhood of the origin of the tangent space; the *logarithm map* $\log_p$ is the inverse of $\exp_p$. The *radius of injectivity* $\mathrm{inj}_p$ at $p \in \mathcal{M}$ is the radius of the largest ball about the origin of $T_p\mathcal{M}$, on which $\exp_p$ is a diffeomorphism (Figure 1, left panel). If $\mathcal{N}$ is a submanifold of $\mathcal{M}$ with Riemannian metric $h_p : T_p\mathcal{N} \times T_p\mathcal{N} \to \mathbb{R}$, $(u, v) \mapsto g_p(u, v)$ for $u, v \in T_p\mathcal{N}$ induced by $g$, then $(\mathcal{N}, h)$ is a *Riemannian submanifold* of $(\mathcal{M}, g)$.

We consider a $d$-dimensional complete Riemannian submanifold $\mathcal{M}$ of a Euclidean space $\mathbb{R}^{d_0}$ for $d \leq d_0$, with a geodesic distance $d_{\mathcal{M}}$ on $\mathcal{M}$ induced by the Euclidean metric in $\mathbb{R}^{d_0}$, and a probability space $(\Omega, \mathcal{A}, P)$ with sample space $\Omega$, $\sigma$-algebra $\mathcal{A}$, and probability measure $P$. With $\mathcal{X} = \{x : \mathcal{T} \to \mathcal{M} \mid x \in \mathcal{C}(\mathcal{T})\}$ denoting the sample space of all $\mathcal{M}$-valued continuous functions on a compact interval $\mathcal{T} \subset \mathbb{R}$ and $\mathcal{B}(\mathcal{V})$ the Borel $\sigma$-algebra of a space $\mathcal{V}$, the $\mathcal{M}$-valued random functions $X(t, \omega)$ are $X : \mathcal{T} \times \Omega \to \mathcal{M}$, such that $X(\cdot, \omega) \in \mathcal{X}$. Here $\omega \mapsto X(\cdot, \omega)$ and $X(t, \cdot)$ are measurable with respect to $\mathcal{B}(\mathcal{X})$ and $\mathcal{B}(\mathcal{M})$, respectively, with $\mathcal{B}(\mathcal{X})$ generated by the supremum metric $d_{\mathcal{X}} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, $d_{\mathcal{X}}(x, y) = \sup_{t \in \mathcal{T}} d_{\mathcal{M}}(x(t), y(t))$, for investigating the rates of uniform convergence. In the following, all vectors $v$ are column vectors and we write $X(t)$, $t \in \mathcal{T}$, for $\mathcal{M}$-valued random functions, $\|\cdot\|_E$ for the Euclidean norm, and $\mathbb{H} = \{v : \mathcal{T} \to \mathbb{R}^{d_0}, \int_{\mathcal{T}} v(t)^T v(t) dt < \infty\}$ for the ambient $L^2$ Hilbert space of $\mathbb{R}^{d_0}$ valued square integrable functions, equipped with the inner product $\langle v, u \rangle = \int_{\mathcal{T}} v(t)^T u(t) dt$ and norm $\|v\| = \langle v, v \rangle^{1/2}$ for $u, v \in \mathbb{H}$.

2.2. *Riemannian functional principal component analysis.* As intrinsic population mean function for the $\mathcal{M}$-valued random function $X(t)$, we consider the intrinsic Fréchet mean $\mu_{\mathcal{M}}(t)$ at each time point $t \in \mathcal{T}$, where

$$M(p, t) = E[d_{\mathcal{M}}(X(t), p)^2], \quad \mu_{\mathcal{M}}(t) = \operatorname*{arg\,min}_{p \in \mathcal{M}} M(p, t), \tag{1}$$

and we assume the existence and the uniqueness of the Fréchet means $\mu_{\mathcal{M}}(t)$. The mean function $\mu_{\mathcal{M}}$ is continuous due to the continuity of the sample paths of $X$, as per Proposition 2 below. One could consider an alternative definition for the mean function, $\mu_G = \arg\min_\mu F(\mu)$, where $F(\mu) = E[\int_{\mathcal{T}} d_{\mathcal{M}}(X(t), \mu(t))^2 dt]$, which coincides with $\mu_{\mathcal{M}}$ under a continuity assumption; we work with $\mu_{\mathcal{M}}$ in (1), as it matches the approach in functional PCA and allows us to investigate uniform convergence. The goal

of RFPCA is to represent the variation of the infinite dimensional object $X$ around the mean function $\mu_{\mathcal{M}}$ in a lower dimensional submanifold, in terms of a few principal modes of variation, an approach that has been successful to represent random trajectories in the Hilbert space $L^2$ (Castro, Lawton and Sylvestre 1986; Ramsay and Silverman 2005; Wang, Chiou and Müller 2016).

Given an arbitrary system of $K$ orthonormal basis functions, $\Psi_K = \{\psi_k \in \mathbb{H} \mid \psi_k(t) \in T_{\mu_{\mathcal{M}}(t)}, \langle \psi_k, \psi_l \rangle = \delta_{kl},\ k,l = 1,\ldots,K\}$, $\delta_{kl} = 1$ if $k = l$ and 0 otherwise, with values at each time $t \in \mathcal{T}$ restricted to the $d$-dimensional tangent space $T_{\mu_{\mathcal{M}}(t)}$, which we identify with $\mathbb{R}^{d_0}$ for convenience, we define the $K$ dimensional time-varying geodesic submanifold

(2)
$$\mathcal{M}_K(\Psi_K) \coloneqq \{x \in \mathcal{X},\ x(t) = \exp_{\mu_{\mathcal{M}}(t)}(\sum_{k=1}^{K} a_k \psi_k(t)) \text{ for } t \in \mathcal{T} \mid a_k \in \mathbb{R}\}.$$

Here $\mathcal{M}_K(\Psi_K)$ plays an analogous role to the linear span of a set of basis functions in Hilbert space, with expansion coefficients or coordinates $a_k$.

In the following we suppress the dependency of $\mathcal{M}_K$ on the basis functions. With projections $\Pi(x, \mathcal{M}_K)$ of an $\mathcal{M}$-valued function $x \in \mathcal{X}$ onto time-varying geodesic submanifolds $\mathcal{M}_K$,

$$\Pi(x, \mathcal{M}_K) \coloneqq \operatorname*{arg\,min}_{y \in \mathcal{M}_K} \int_{\mathcal{T}} d_{\mathcal{M}}(y(t), x(t))^2 dt,$$

the best $K$-dimensional approximation to $X$ minimizing the geodesic projection distance is the geodesic submanifold that minimizes

$$F_S(\mathcal{M}_K) = E \int_{\mathcal{T}} d_{\mathcal{M}}(X(t), \Pi(X, \mathcal{M}_K)(t))^2 dt \tag{3}$$

over all time-varying geodesic submanifolds generated by $K$ basis functions.

As the minimization of (3) is over a family of submanifolds (or basis functions), this target is difficult to implement in practice, except for simple situations, and therefore it is expedient to target a modified version of (3) by invoking tangent space approximations. This approximation requires that the log-mapped random functions

$$V(t) = \log_{\mu_{\mathcal{M}}(t)}(X(t))$$

are almost surely well-defined for all $t \in \mathcal{T}$, which will be the case if trajectories $X(t)$ are confined to stay within the radius of injectivity at $\mu_{\mathcal{M}}(t)$ for all $t \in \mathcal{T}$. We require this constraint to be satisfied, which will be the case

for many manifold-valued trajectory data, including the data we present in Section 5. Then $V$ is a well-defined random function that assumes its values on the linear tangent space $T_{\mu_{\mathcal{M}}(t)}$ at time $t$. Identifying $T_{\mu_{\mathcal{M}}(t)}$ with $\mathbb{R}^{d_0}$, we may regard $V$ as a random element of $\mathbb{H}$, the $L^2$ Hilbert space of $\mathbb{R}^{d_0}$ valued square integrable functions, and thus our analysis is independent of the choice of the coordinate systems on the tangent spaces. A practically tractable optimality criterion to obtain manifold principal components is then to minimize

$$F_V(\mathcal{V}_K) = E(\|V - \Pi(V, \mathcal{V}_K)\|^2) \tag{4}$$

over all $K$-dimensional linear subspaces $\mathcal{V}_K(\psi_1, \dots, \psi_K) = \{\sum_{k=1}^K a_k \psi_k \mid a_k \in \mathbb{R}\}$ for $\psi_k \in \mathbb{H}$, $\psi_k(t) \in T_{\mu_{\mathcal{M}}(t)}$, and $k = 1, \dots, K$. Minimizing (4) is immediately seen to be equivalent to a multivariate functional principal component analysis (FPCA) in $\mathbb{R}^{d_0}$ (Chiou, Chen and Yang 2014).

Under mild assumptions, the $L^2$ mean function for the log-mapped data $V(t) = \log_{\mu_{\mathcal{M}}(t)}(X(t))$ at the Fréchet means is zero by Theorem 2.1 of Bhattacharya and Patrangenaru (2003). Consider the covariance function $G$ of $V$ in the $L^2$ sense, $G : \mathcal{T} \times \mathcal{T} \to \mathbb{R}^{d_0^2}$, $G(t, s) = \operatorname{cov}(V(t), V(s)) = E(V(t)V(s)^T)$, and its associated spectral decomposition,
$G(t, s) = \sum_{k=1}^\infty \lambda_k \phi_k(t) \phi_k(s)^T$, where the $\phi_k \in \mathbb{H} : \mathcal{T} \to \mathbb{R}^{d_0}$ are the orthonormal vector-valued eigenfunctions and $\lambda_k \geq 0$ the corresponding eigenvalues, for $k = 1, 2, \dots$. One obtains the Karhunen-Loève decomposition (see for example Hsing and Eubank 2015),

$$V(t) = \sum_{k=1}^\infty \xi_k \phi_k(t), \tag{5}$$

where $\xi_k = \int_{\mathcal{T}} V(t)\phi_k(t)dt$ is the $k$th Riemannian functional principal component (RFPC) score, $k = 1, 2, \dots$. A graphical demonstration of $X(t)$, $V(t)$, and $\phi_k(t)$ is in the right panel of Figure 1. In practice, one can use only a finite number of components and target truncated representations of the tangent space process. Employing $K \in \{0, 1, 2, \dots\}$ components, set

$$V_K(t) = \sum_{k=1}^K \xi_k \phi_k(t), \quad X_K(t) = \exp_{\mu_{\mathcal{M}}(t)} \left( \sum_{k=1}^K \xi_k \phi_k(t) \right), \tag{6}$$

where for $K = 0$ the values of the sums are set to 0, so that $V_0(t) = 0$ and $X_0(t) = \mu_{\mathcal{M}}(t)$. By classical FPCA theory, $V_K$ is the best $K$-dimensional approximation to $V$ in the sense of being the minimizing projection $\Pi(V, \mathcal{V}_K)$ for (4). The truncated representation $X_K(t)$, $t \in \mathcal{T}$ of the original $\mathcal{M}$-valued

random function is well-defined for $K = 0, 1, \ldots$ if $\mathcal{M}$ is complete, by the Hopf–Rinow theorem (see, e.g., Chavel 2006). We note that these definitions are independent of the choice of coordinate system on $T_{\mu_{\mathcal{M}}(t)}$.

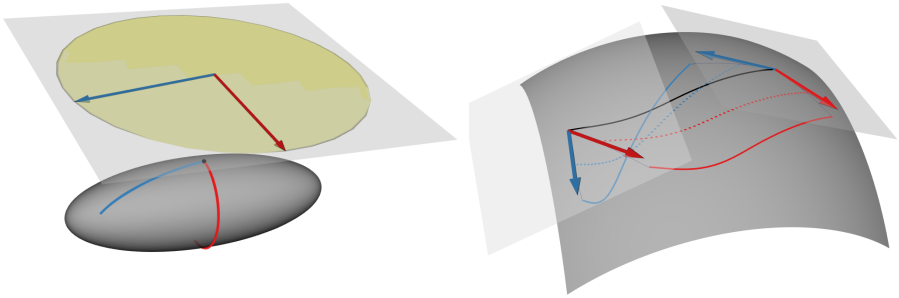


FIG 1. *Left panel: Two tangent vectors $v$ (red and blue arrows) in the tangent ball (yellow) centered at $p$ (black dot) with radius $\mathrm{inj}_p$, and their geodesics $\{\exp_p(tv) \mid t \in [0,1]\}$ (red and blue lines). Right panel: Two trajectories $X(t)$ (red and blue solid curves), corresponding tangent vectors $V(t)$ at $t = 0, 1$ (arrows), and the first two eigenfunctions (red dotted, $\phi_1$, and blue dotted, $\phi_2$) mapped onto $\mathcal{M}$ by the exponential maps. The red trajectory has a large score on $\phi_1$, while the blue one has a large score on $\phi_2$. The mean function is the black curve.*

To quantify how well $X_K$ approximates $X$, in analogy to Petersen and Müller (2016a), we define for $K = 0, 1, \ldots$ the residual variance as

$$U_K = E \int_{\mathcal{T}} d_{\mathcal{M}}(X(t), X_K(t))^2 dt, \tag{7}$$

and the fraction of variance explained (FVE) by the first $K$ components as

$$\mathrm{FVE}_K = \frac{U_0 - U_K}{U_0}. \tag{8}$$

A commonly used criterion for choosing the number of included components $K^*$ is to select the smallest $K$ such that FVE exceeds a specified threshold $0 < \gamma < 1$ of variance explained,

$$K^* = \min \{K : \mathrm{FVE}_K \geq \gamma\}. \tag{9}$$

Common choices for the FVE threshold $\gamma$ are 0.9 or 0.95 in finite sample situations or $\gamma$ increasing with sample size for asymptotic considerations.

2.3. *Spherical functional principal component analysis.* An important special case occurs when random trajectories lie on $\mathcal{M} = S^d$, the Euclidean sphere in $\mathbb{R}^{d_0}$ for $d_0 = d + 1$, with the Riemannian geometry induced by the Euclidean metric of the ambient space. Then the proposed RFPCA specializes to spherical functional principal component analysis (SFPCA). We briefly review the geometry of Euclidean spheres. The geodesic distance $d_{\mathcal{M}}$ on the sphere is the great-circle distance, i.e. for $p_1, p_2 \in \mathcal{M} = S^d$

$$d_{\mathcal{M}}(p_1, p_2) = \cos^{-1}({p_1}^T p_2).$$

A geodesic is a segment of a great circle that connects two points on the sphere. For any point $p \in \mathcal{M}$, the tangent space $T_p\mathcal{M}$ is identified by $\{v \in \mathbb{R}^{d_0} \mid v^T p = 0\} \subset \mathbb{R}^{d_0}$, with the Euclidean inner product. Letting $\|\cdot\|_E$ be the Euclidean norm in the ambient Euclidean space $\mathbb{R}^{d_0}$, then for a tangent vector $v$ on the tangent space $T_p\mathcal{M}$, the exponential map is

$$\exp_p(v) = \cos(\|v\|_E)p + \sin(\|v\|_E)\frac{v}{\|v\|_E}.$$

The logarithm map $\log_p : \mathcal{M}\backslash\{-p\} \to T_p\mathcal{M}$ is the inverse of the exponential map,

$$\log_p(q) = \frac{u}{\|u\|_E} d_{\mathcal{M}}(p, q),$$

where $u = q - (p^T q)\, p$, and $\log_p$ is defined everywhere with the exception of the antipodal point $-p$ of $p$ on $\mathcal{M}$. The radius of injectivity is therefore $\pi$. The sectional curvature of a Euclidean sphere is constant.

2.4. *Estimation.* Consider a Riemannian manifold $\mathcal{M}$ and $n$ independent observations $X_1, \ldots, X_n$, which are $\mathcal{M}$-valued random functions that are distributed as $X$, where we assume that these functions are fully observed for $t \in \mathcal{T}$. Population quantities for RFPCA are estimated by their empirical versions, as follows. Sample Fréchet means $\hat{\mu}_{\mathcal{M}}(t)$ are obtained by minimizing $M_n(\cdot, t)$ at each $t \in \mathcal{T}$,

$$M_n(p, t) = \frac{1}{n}\sum_{i=1}^{n} d_{\mathcal{M}}(X_i(t), p)^2, \quad \hat{\mu}_{\mathcal{M}}(t) = \operatorname*{arg\,min}_{p \in \mathcal{M}} M_n(p, t). \tag{10}$$

We estimate the log-mapped data $V_i$ by $\hat{V}_i(t) = \log_{\hat{\mu}_{\mathcal{M}}(t)}(X_i(t))$, $t \in \mathcal{T}$; the covariance function $G(t, s)$ by the sample covariance function $\hat{G}(t, s) = n^{-1}\sum_{i=1}^{n} \hat{V}_i(t)\hat{V}_i(s)^T$ based on $\hat{V}_i$, for $t, s \in \mathcal{T}$; the $k$th eigenvalue and eigenfunction pair $(\lambda_k, \phi_k)$ of $G$ by the eigenvalue and eigenfunction $(\hat{\lambda}_k, \hat{\phi}_k)$ of $\hat{G}$; and the $k$th RFPC score of the $i$th subject $\xi_{ik} = \int_{\mathcal{T}} V_i(t)\phi_k(t)dt$ by

$\hat{\xi}_{ik} = \int_{\mathcal{T}} \hat{V}_i(t)\hat{\phi}_k(t)dt$. The $K$-truncated processes $V_{iK}$ and $X_{iK}$ for the $i$th subject $X_i$ are estimated by

$$\hat{V}_{iK}(t) = \sum_{k=1}^{K} \hat{\xi}_{ik}\hat{\phi}_k(t), \quad \hat{X}_{iK}(t) = \exp_{\hat{\mu}_{\mathcal{M}}(t)}\left(\sum_{k=1}^{K} \hat{\xi}_{ik}\hat{\phi}_k(t)\right), \tag{11}$$

where again for $K = 0$ we set the sums to 0. The residual variance $U_K$ as in (7), the fraction of variance explained $\mathrm{FVE}_K$ as in (8), and the optimal $K^*$ as in (9) are respectively estimated by

$$\hat{U}_K = \frac{1}{n}\sum_{i=1}^{n}\int_{\mathcal{T}} d_{\mathcal{M}}(X_i(t), \hat{X}_{iK}(t))^2 dt, \tag{12}$$

$$\widehat{\mathrm{FVE}}_K = \frac{\hat{U}_0 - \hat{U}_K}{\hat{U}_0}, \tag{13}$$

$$\hat{K}^* = \min\{K : \widehat{\mathrm{FVE}}_K \geq \gamma\}. \tag{14}$$

Further details about the algorithms for implementing SFPCA can be found in the Supplementary Materials. Sometimes functional data $X(t)$ are observed only at densely spaced time points and observations might be contaminated with measurement errors. In these situations one can presmooth the observations using smoothers that are adapted to a Riemannian manifold (Jupp and Kent 1987; Lin et al. 2016), treating the presmoothed curves as fully observed underlying curves.

**3. Theoretical properties of Riemannian Functional Principal Component Analysis.** We need the following assumptions (A1)–(A2) for the Riemannian manifold $\mathcal{M}$, and (B1)–(B6) for the $\mathcal{M}$-valued process $X(t)$.

(A1) $\mathcal{M}$ is a closed Riemannian submanifold of a Euclidean space $\mathbb{R}^{d_0}$, with geodesic distance $d_{\mathcal{M}}$ induced by the Euclidean metric.

(A2) The sectional curvature of $\mathcal{M}$ is nonnegative.

Assumption (A1) guarantees that the exponential map is defined on the entire tangent plane, and thus that $X_K(t)$ as in (6) is well-defined, while the curvature condition (A2) implies that geodesics starting from the same point tend to converge, so that the geodesic distance between $X_K(t)$ and $X(t)$ is bounded by the Euclidean distance of their tangent vectors. These assumptions are satisfied for example by Euclidean spheres $S^d$. For the following recall $M(p,t)$ and $M_n(p,t)$ are defined as in (1) and (10).

(B1) Trajectories $X(t)$ are continuous for $t \in \mathcal{T}$ almost surely.

(B2) For all $t \in \mathcal{T}$, $\mu_{\mathcal{M}}(t)$ and $\hat{\mu}_{\mathcal{M}}(t)$ exist and are unique, the latter almost surely.

(B3) Almost surely, trajectories $X(t)$ lie in a compact set $S_t \subset B_{\mathcal{M}}(\mu_{\mathcal{M}}(t), r)$ for $t \in \mathcal{T}$, where $B_{\mathcal{M}}(\mu_{\mathcal{M}}(t), r) \subset \mathcal{M}$ is an open ball centered at $\mu_{\mathcal{M}}(t)$ with radius $r < \inf_{t\in\mathcal{T}} \mathrm{inj}_{\mu_{\mathcal{M}}(t)}$.

(B4) For any $\epsilon > 0$,

$$\inf_{t\in\mathcal{T}} \inf_{p:\, d_{\mathcal{M}}(p,\mu_{\mathcal{M}}(t))>\epsilon} M(p,t) - M(\mu_{\mathcal{M}}(t),t) > 0.$$

(B5) For $v \in T_{\mu_{\mathcal{M}}(t)}\mathcal{M}$, define $g_t(v) = M(\exp_{\mu_{\mathcal{M}}(t)}(v), t)$. Then

$$\inf_{t\in\mathcal{T}} \lambda_{\min}\Big(\frac{\partial^2}{\partial v^2} g_t(0)\Big) > 0,$$

where $\lambda_{\min}(A)$ is the smallest eigenvalue of a square matrix $A$.

(B6) Let $L(x)$ be the Lipschitz constant of a function $x$, i.e. $L(x) = \sup_{t\neq s} d_{\mathcal{M}}(x(t), x(s))/|t-s|$. Then $E(L(X)^2) < \infty$ and $L(\mu_{\mathcal{M}}) < \infty$.

Smoothness assumptions (B1) and (B6) for the sample paths of the observations are needed for continuous representations, while existence and uniqueness of Fréchet means (B2) are prerequisites for an intrinsic analysis that are commonly assumed (Bhattacharya and Patrangenaru 2003; Petersen and Müller 2016b) and depend in a complex way on the type of manifold and probability measure considered. Assumptions (B4) and (B5) characterize the local behavior of the criterion function $M$ around the minima and are standard for M-estimators (Bhattacharya and Lin 2017). Condition (B3) ensures that the geodesic between $X(t)$ and $\mu_{\mathcal{M}}(t)$ is unique, ensuring that the tangent vectors do not switch directions under small perturbations of the base point $\mu_{\mathcal{M}}(t)$. It is satisfied for example for the sphere $\mathcal{M} = S^d$, if the values of the random functions are either restricted to the positive quadrant of the sphere, as is the case for longitudinal compositional data as in Section 4, or if the samples are generated by $\exp_{\mu_{\mathcal{M}}(t)}(\sum_{k=1}^{\infty} \xi_k \phi_k(t))$ with bounded eigenfunctions $\phi_k$ and small scores $\xi_k$ such that $\sup_{t\in\mathcal{T}} |\sum_{k=1}^{\infty} \xi_k \phi_k(t)| \leq r$. In real data applications, (B3) is justified when the $\mathcal{M}$-valued samples cluster around the intrinsic mean function, as exemplified by the flight trajectory data that we study in Subsection 5.2.

The following result justifies the tangent space RFPCA approach, as the truncated representation is found to be well-defined, and the residual variance for the optimal geodesic submanifold representation bounded by that for the classical FPCA on the tangent space.

PROPOSITION 1. *Under (A1),* $X_K(t) = \exp_{\mu_{\mathcal{M}}(t)}(V_K(t))$ *is well-defined for* $K = 1, 2, \ldots$ *and* $t \in \mathcal{T}$. *If further (A2) is satisfied, then*

(15)
$$\min_{\mathcal{M}_K} \int_{\mathcal{T}} d_{\mathcal{M}}(X(t), \Pi(X, \mathcal{M}_K)(t))^2 dt \le \int_{\mathcal{T}} d_{\mathcal{M}}(X(t), X_K(t))^2 dt \le \|V - V_K\|^2 .$$

The first statement is a straightforward consequence of the Hopf-Rinow theorem, while the inequalities imply that the residual variance using the best $K$-dimensional time-varying geodesic manifold approximation under geodesic distance (the left hand term) is bounded by that of the geodesic manifold produced by the proposed RFPCA (the middle term), where the latter is again bounded by the residual variance of a linear tangent space FPCA under the familiar Euclidean distance (the right hand term). The r.h.s. inequality in (15) affirms that the tangent space FPCA serves as a gauge to control the preciseness of finite-dimensional approximation to the processes under the geodesic distance. An immediate consequence is that $U_K \to 0$ as $K \to \infty$ for the residual variance $U_K$ in (7), implying that the truncated representation $X_K(t)$ is consistent for $X(t)$ when the sectional curvature of $\mathcal{M}$ is nonnegative. The l.h.s. inequality gets tighter as the samples $X(t)$ lie closer to the intrinsic mean $\mu_{\mathcal{M}}(t)$, where such closeness is not uncommon, as demonstrated in Section 5. The r.h.s. inequality is a consequence of the Alexandrov–Toponogov theorem for comparing geodesic triangles.

Asymptotic properties for the estimated model components for RFPCA are studied below.

PROPOSITION 2. *Under (A1) and (B1)–(B4),* $\mu_{\mathcal{M}}(t)$ *is continuous,* $\hat{\mu}_{\mathcal{M}}(t)$ *is continuous with probability tending to 1 as* $n \to \infty$, *and*

$$\sup_{t \in \mathcal{T}} d_{\mathcal{M}}(\hat{\mu}_{\mathcal{M}}(t), \mu_{\mathcal{M}}(t)) = o_p(1). \tag{16}$$

Under additional assumptions (B5) and (B6), the consistency in (16) of the sample intrinsic mean $\hat{\mu}_{\mathcal{M}}(t)$ as an estimator for the true intrinsic mean $\mu_{\mathcal{M}}(t)$ can be strengthened through a central limit theorem on $\mathcal{C}_d(\mathcal{T})$, where $\mathcal{C}_d(\mathcal{T})$ is the space of $\mathbb{R}^d$-valued continuous functions on $\mathcal{T}$. Let $\tau : U \to \mathbb{R}^d$ be a smooth or infinitely differentiable chart of the form $\tau(q) = \log_{p_0}(q)$, with $U = B_{\mathcal{M}}(p_0, r_0)$, $p_0 \in \mathcal{M}$, and $r_0 < \mathrm{inj}_{p_0}$, identifying tangent vectors in $\mathbb{R}^d$. Define chart distance $d_\tau : \tau(U) \times \tau(U) \to \mathbb{R}$ by $d_\tau(u, v) = d_{\mathcal{M}}(\tau^{-1}(u), \tau^{-1}(v))$, its gradient $T(u, v) = [T_j(u, v)]_{j=1}^d = [\partial d_\tau(u, v)/\partial v^j]_{j=1}^d$, Hessian matrix $H(u, v)$ with $(j, l)$th element $H_{jl}(u, v) = \partial^2 d_\tau^2(u, v)/\partial v_j \partial v_l$, and $\Lambda(t) = E[H(\tau(X(t)), \tau(\mu_{\mathcal{M}}(t)))]$.

THEOREM 1. *Suppose that $\mu_{\mathcal{M}}(t)$ and $X(t)$ are contained in the domain of $\tau$ for $t \in \mathcal{T}$, the latter almost surely, and (A1) and (B1)–(B6) hold. Then*

$$\sqrt{n}[\tau(\hat{\mu}_{\mathcal{M}}) - \tau(\mu_{\mathcal{M}})] \xrightarrow{L} Z, \tag{17}$$

*where $Z$ is a Gaussian process with sample paths in $\mathcal{C}_d(\mathcal{T})$, mean zero, and covariance $G_\mu(t,s) = \Lambda^{-1}(t)G_T(t,s)\Lambda^{-1}(s)$, where $G_T(t,s) = E[T(\tau(X(t)), \tau(\mu_{\mathcal{M}}(t)))T(\tau(X(s)), \tau(\mu_{\mathcal{M}}(s)))^T]$, and all quantities are well-defined.*

**Remark 1.** The first condition in Theorem 1 is not restrictive, since it holds at least piecewise on some finite partition of $\mathcal{T}$. More precisely, due to the compactness guaranteed by (A1), (B3), and Proposition 2, there exists a finite partition $\{\mathcal{T}_j\}_{j=1}^N$ of $\mathcal{T}$ such that $\mu_{\mathcal{M}}(t)$ and $X(t)$ are contained in $B_{\mathcal{M}}(\mu_{\mathcal{M}}(t_j), r_j)$, for $t \in \mathcal{T}_j$, $t_j \in \mathcal{M}$ and $r_j < \operatorname{inj}_{\mu_{\mathcal{M}}(t_j)}$, $j = 1, \ldots, N < \infty$. One can then define $\tau = \tau_j := q \mapsto \log_{\mu_{\mathcal{M}}(t_j)}(q)$ for $t \in \mathcal{T}_j$ and apply Theorem 1 on the $j$th piece, for each $j$.

COROLLARY 1. *Under (A1) and (B1)–(B6),*

$$\sup_{t\in\mathcal{T}} d_{\mathcal{M}}(\hat{\mu}_{\mathcal{M}}(t), \mu_{\mathcal{M}}(t)) = O_p(n^{-1/2}). \tag{18}$$

**Remark 2.** The intrinsic dimension $d$ is only reflected in the rate constant but not the speed of convergence. Our situation is analogous to that of estimating the mean of Euclidean-valued random functions (Bosq 2000), or more generally, Fréchet regression with Euclidean responses (Petersen and Müller 2016b), where the speed of convergence does not depend on the dimension of the Euclidean space, in contrast to common nonparametric regression settings (Lin et al. 2016; Lin and Yao 2017). The root-$n$ rate is not improvable in general since it is the optimal rate for mean estimates in the special Euclidean case.

An immediate consequence of Corollary 1 is the convergence of the log-mapped data.

COROLLARY 2. *Under (A1) and (B1)–(B6), for $i = 1, \ldots, n$,*

$$\sup_{t\in\mathcal{T}} \|\hat{V}_i(t) - V_i(t)\|_E = O_p(n^{-1/2}). \tag{19}$$

In the following, we use the Frobenius norm $\|A\|_F = \operatorname{tr}(A^T A)^{1/2}$ for any real matrices $A$, and assume that the eigenspaces associated with positive

eigenvalues $\lambda_k > 0$ have multiplicity one. We obtain convergence of covariance functions, eigenvalues, and eigenfunctions on the tangent spaces, i.e., the consistency of the spectral decomposition of the sample covariance function, as follows.

THEOREM 2. *Assume (A1) and (B1)–(B6) hold. Then*

$$\sup_{t,s\in\mathcal{T}} \left\| \hat{G}(t,s) - G(t,s) \right\|_F = O_p(n^{-1/2}), \tag{20}$$

$$\sup_{k\in\mathbb{N}} |\hat{\lambda}_k - \lambda_k| = O_p(n^{-1/2}), \tag{21}$$

*and for each* $k = 1, 2, \ldots$ *with* $\lambda_k > 0$,

$$\sup_{t\in\mathcal{T}} \|\hat{\phi}_k(t) - \phi_k(t)\|_E = O_p(n^{-1/2}). \tag{22}$$

Our next result provides the convergence rate of the RFPC scores and is a direct consequence of Corollary 2 and Theorem 2.

THEOREM 3. *Under (A1) and (B1)–(B6), if* $\lambda_K > 0$ *for some* $K \geq 1$, *then for each* $i = 1, \ldots, n$ *and* $k = 1, \ldots, K$,

$$|\hat{\xi}_{ik} - \xi_{ik}| = O_p(n^{-1/2}), \tag{23}$$

$$\sup_{t\in\mathcal{T}} \|\hat{V}_{iK}(t) - V_{iK}(t)\|_E = O_p(n^{-1/2}). \tag{24}$$

To demonstrate asymptotic consistency for the number of components selected according to the FVE criterion, we consider an increasing sequence of FVE thresholds $\gamma = \gamma_n \uparrow 1$ as sample size $n$ increases, which leads to a corresponding increasing sequence of $K^* = K_n^*$, where $K^*$ is the smallest number of eigen-components that explains the fraction of variance $\gamma = \gamma_n$. One may show that the number of components $\hat{K}^*$ selected from the sample is consistent for the true target $K^*$ for a sequence $\gamma_n$. This is formalized in the following Corollary 3, which is similar to Theorem 2 in Petersen and Müller (2016a), where also specific choices of $\gamma_n$ and the corresponding sequences $K^*$ were discussed. The proof is therefore omitted. Quantities $U_0$, $U_K$, $K^*$, $\hat{U}_0$, $\hat{U}_K$, $\hat{K}^*$ that appear below were defined in (7)–(9) and (12)–(14).

COROLLARY 3. *Assume (A1)–(A2) and (B1)–(B6) hold. If the eigenvalues* $\lambda_1 > \lambda_2 > \cdots > 0$ *are all distinct, then there exists a sequence* $0 < \gamma_n \uparrow 1$ *such that*

$$\max_{1\leq K\leq K^*} \left| \frac{\hat{U}_0 - \hat{U}_K}{\hat{U}_0} - \frac{U_0 - U_K}{U_0} \right| = o_p(1), \tag{25}$$

*and therefore*

$$P(\hat{K}^* \neq K^*) = o(1). \tag{26}$$

**4. Longitudinal compositional data analysis.** Compositional data represent proportions and are characterized by a vector $\mathbf{y}$ in the simplex

$$\mathcal{C}^{J-1} = \{\mathbf{y} = [y_1, \ldots, y_J] \in \mathbb{R}^J \mid y_j \geq 0,\, j = 1, \ldots, J;\, \sum_{j=1}^{J} y_j = 1\},$$

requiring that the nonnegative proportions of all $J$ categories sum up to one. Typical examples include the geochemical composition of rocks or other data that consist of recorded percentages. Longitudinal compositional data arise when the compositional data for the same subject are collected repeatedly at different time points. If compositions are monitored continuously, each sample path of longitudinal compositional data is a function $y : \mathcal{T} \to \mathcal{C}^{J-1}$. Analyses of such data, for example from a prospective ophthalmology study (Qiu, Song and Tan 2008) or the surveillance of the composition of antimicrobial use over time (Adriaenssens et al. 2011), have drawn both methodological and practical interest, but as of yet there exists no unifying methodology for longitudinal compositional data, to the knowledge of the authors.

A direct application of standard Euclidean space methods, viewing longitudinal compositional data as unconstrained functional data vectors (Chiou, Chen and Yang 2014), would ignore the non-negativity and unit sum constraints and therefore the resulting multivariate FPCA representation moves outside of the space of compositional data, diminishing the utility of such simplistic approaches. There are various transformation that have been proposed over the years for the analysis of compositional data to enforce the constraints, for example log-ratio transformations such as $\log(y_j/y_J)$ for $j = 1, \ldots, J-1$, after which the data are treated as Euclidean data (Aitchison 1986), which induces the Aitchison geometry on the interior of the simplex $\mathcal{C}^{J-1}$. However, these transformations cannot be defined when some of the elements in the composition are zeros, either due to the discrete and noisy nature of the observations or when the true proportions do contain actual zeros, as is the case in the fruit fly behavior pattern data that we study in Subsection 5.1 below.

We propose to view longitudinal compositional data as a special case of multivariate functional data under constraints, specifically as realizations of

a compositional process over time,

(27)
$$Y(t) \in \{[Y_1(t), \ldots, Y_J(t)] \in \mathbb{R}^J \mid Y_j \in L^2(\mathcal{T}),\, Y_j(t) \geq 0,\, \sum_{j=1}^{J} Y_j(t) = 1\},$$

where the component functions will also be assumed to be continuous on their domain $\mathcal{T}$. To include the entire simplex $\mathcal{C}^{J-1}$ in our longitudinal compositional data analysis, we apply square root transformations to the longitudinal compositional data $Y(t) = [Y_1(t), \ldots, Y_J(t)]$, obtaining

$$X(t) = [X_1(t), \ldots, X_J(t)] = [Y_1(t)^{1/2}, \ldots, Y_J(t)^{1/2}]. \tag{28}$$

A key observation is that the values of $X(t)$ lie on the positive quadrant of a hypersphere $S^{J-1}$ for $t \in \mathcal{T}$, as $X_j(t) \geq 0$ and $\sum_{j=1}^{J} X_j(t)^2 = 1$. There is no problem with zeros as with some other proposed transformations for compositional data. It is then a natural approach to consider a spherical geometry for the transformed data $X(t)$. A square-root transformation and the spherical geometry for non-longitudinal compositional data were previously considered by Huckemann and Eltzner (2016). Now, since $X(t)$ assumes its values on a quadrant of the sphere $S^{J-1}$, processes $X(t)$ fall into the framework of the proposed SFPCA, as described in Subsection 2.3.

Concerning the theoretical properties of SFPCA of longitudinal compositional data, the conditions on the Riemannian manifold $\mathcal{M}$ needed for RFPCA are easily seen to be satisfied, due to the geometry of the Euclidean sphere and the positive quadrant constraint. We conclude

COROLLARY 4. *Under (B1) and (B4)–(B6), Propositions 1 and 2, Theorems 1–3, and Corollaries 1–3 hold for the Spherical Functional Principal Component Analysis (SFPCA) of longitudinal compositional data $X(t)$ in (28).*

## 5. Data applications.

5.1. *Fruit fly behaviors.* To illustrate the proposed SFPCA based longitudinal compositional data analysis, we consider the lifetime behavior pattern data of *D. melanogaster* (common fruit fly, Carey et al. 2006). The behavioral patterns of each fruit fly was observed instantaneously 12 times each day during its entire lifetime, and for each observation one of the five behavioral patterns, feeding, flying, resting, walking, and preening, was recorded. We analyzed the behavioral patterns in the first 30 days since eclosion for $n = 106$ fruit flies with uncensored observations, aiming to characterize and

represent age-specific behavioral patterns of individual fruit flies. For each fruit fly, we observed the behavioral counts $[Z_1(t), \ldots, Z_5(t)]$ for the five behaviors at time $t \in \mathcal{T} = [0, 30]$, where the time unit is day since eclosion, and $\sum_{j=1}^{5} Z_j(t) = 12$ is the constrained total number of counts at each time $t$, with $0 \leq Z_j(t) \leq 12$ for each $j$ and $t$. Since the day-to-day behavioral data are noisy, we presmoothed the counts $Z_j(t)$ of the $j$th behavior pattern over time for $j = 1, \ldots, 5$, using a Nadaraya–Watson kernel smoother (Nadaraya 1964; Watson 1964) with Epanechnikov kernel and a bandwidth of five days. The smoothed data were subsequently divided by the sum of the smoothed component values at each $t$, yielding a functional vector $Y(t) = [Y_1(t), \ldots, Y_5(t)]$, with $Y_j(t) \geq 0$ for all $j$ and $t$ and $\sum_{j=1}^{5} Y_j(t) = 1$ for $t \in \mathcal{T}$, thus corresponding to longitudinal compositional data.

Following the approach described in Section 4, we model the square-root composition proportions $X(t) = [Y_1(t)^{1/2}, \ldots, Y_5(t)^{1/2}]$ with SFPCA. The trajectories $X(t)$ and the fitted trajectories for 12 randomly selected fruit flies by SFPCA with $K = 5$ components are demonstrated in Figure 2, and the mean function and the first five eigenfunctions of the corresponding SFPCA in Figure 3. While resting and walking behaviors were often observed, flying and preening occurred more rarely. SFPCA with $K = 5$ components explains 81.7% of total variation and is seen to provide a reasonable fit to the data. The eigenfunctions obtained from SFPCA have a natural interpretation: The first eigenfunction corresponds to the overall contrast of resting and moving (mainly flying and walking) over all days of observation; the second eigenfunction is a contrast of all behaviors in the early (0–15 days) and the late (16–30 days) periods; and the third eigenfunction mainly reflects the intensity of the feeding behavior in the first 25 days.

The fraction of variance explained by the first $K$ components (FVE) as in (13) for SFPCA and for $L^2$ FPCA is in Table 1, where $L^2$ FPCA is conventional multivariate FPCA (Ramsay and Silverman 2005), which ignores the compositional constraints. The proposed SPFCA has larger FVE given any number of included components $K$. It is seen to be more parsimonious than $L^2$ FPCA and it respects the compositional constraints, in contrast to conventional FPCA. To explain 95% of total variation, 14 components are needed for SFPCA, but 18 for $L^2$ FPCA.

TABLE 1
*FVE (%) by the first K components for the fruit fly data.*

| $K$ | 1 | 2 | 3 | 4 | 5 | 10 | 15 | 20 | 25 |
|---|---|---|---|---|---|---|---|---|---|
| SFPCA | 51.7 | 66.7 | 73.1 | 78.3 | 81.7 | 91.8 | 96.4 | 98.4 | 99.2 |
| $L^2$ FPCA | 48.8 | 62.9 | 68.3 | 71.5 | 77.3 | 87.5 | 92.7 | 96.4 | 98.0 |

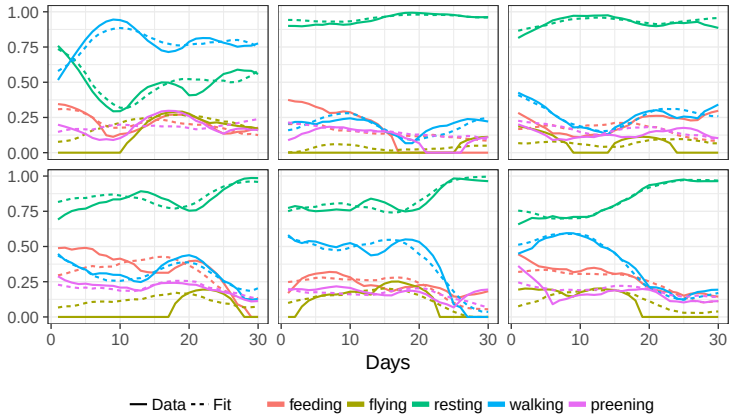



FIG 2. *The original data (solid lines) and SFPCA fitted trajectories (dashed lines) for six randomly selected fruit flies, for $K = 5$ selected components.*

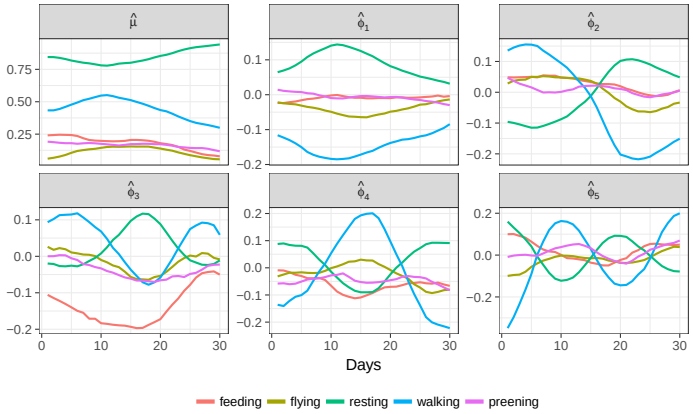



FIG 3. *The estimated mean functions $\hat{\mu}$ and the first five estimated spherical eigenfunctions $\hat{\phi}_1$ to $\hat{\phi}_5$ for the fly data, which together explain* 81.7% *of the total variation. The components explain, respectively, 51.7%, 15.0%, 6.5%, 5.2%, and* 3.4%.

5.2. *Flight trajectories.* A second data example concerns the trajectories of 969 commercial flights from Hong Kong to London from Jun 14, 2016 to Oct 13, 2016, of which 237 were operated by British Airways (BAW), 612 by Cathay Pacific (CPA), and 119 by Virgin Atlantic (VIR). The data were collected from the website of FlightAware (www.flightaware.com) and included longitude, latitude, date, and time, etc. for the whole flight, where the location was densely and accurately tracked by ground based Automatic

Dependent Surveillance–Broadcast (ADS–B) receivers. For each flight we set the takeoff time to be time 0 and the landing time to be time 1, excluding taxi time. To obtain smooth curves from the occasionally noisy data, we pre-smoothed the longitude–latitude data using kernel local linear smoothing with a very small bandwidth and then mapped the longitude–latitude trajectories onto a unit sphere $S^2$. Trajectory data of this kind on geographical spaces corresponding to the surface of the earth that may be approximated by the sphere $S^2$ have drawn extensive interest in computer science and machine learning communities (Zheng 2015; Anirudh et al. 2017). The preprocessed flight trajectories are visualized in Figure 4, indicating that the flight trajectories from the three airlines overlap and are thus not easy to discriminate. We apply RFPCA in the SFPCA version to summarize and represent the flight trajectories, and to predict the operating airline based on the RFPC scores as predictors.

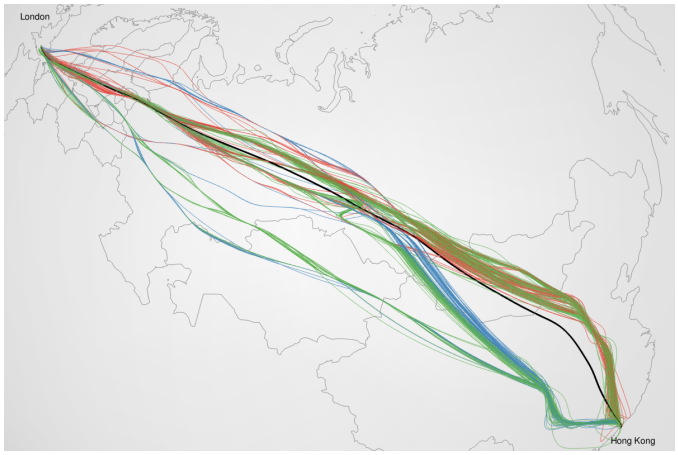



FIG 4. *Flight trajectories from Hong Kong to London, colored by airline (red, British Airways; green, Cathay Pacific; blue, Virgin Atlantic), with the mean trajectory (bold black).*

The estimated mean function and the first three modes of variation obtained by SFPCA are shown in Figure 5, where the $k$th mode of variation is defined as $\exp_{\mu_{\mathcal{M}}(t)}(3\sqrt{\lambda_k}\phi_k(t))$ for $k = 1, 2, 3$. The first mode of variation (red) corresponds to the overall direction of deviation from the mean function (northeast vs southwest), and has roughly constant speed. We connect the second (green) and the third (blue) modes of variation and the mean function using thin gray lines at a regular grid of time in order to display speed information in the corresponding eigenfunctions. Both the second and the third eigenfunctions represent a cross from the northeast to the south-

west at approximately one third of the trip, but they incorporate different speed information. The second eigenfunction encodes an overall fast trip starting to the north, while the third encodes a medium speed start to the south and then a speed up after crossing to the north. The FVE for RFPCA using the first $K = 3$ eigenfunctions is 95%, indicating a reasonably good approximation of the true trajectories.

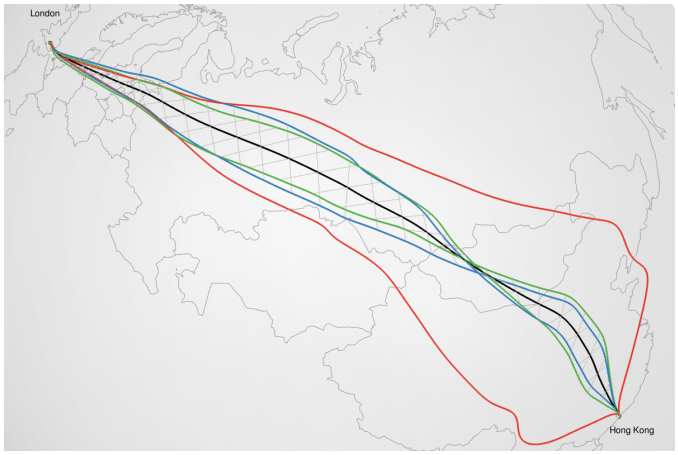



FIG 5. *The mean function (black) and the first three modes of variation defined as* $\exp_{\mu_{\mathcal{M}}(t)}(3\sqrt{\lambda_k}\phi_k(t))$, $k = 1, 2, 3$ *(red, green, and blue, respectively) produced by SFPCA. The second and the third modes of variation were joined to the time-varying mean function at a regular grid of time points to show the "speed" of the eigenfunctions. Both the second and the third eigenfunctions represent a cross from the northeast to the southwest at approximately one third of the trip, but they incorporate different speed information as shown by the thin gray lines. The first three eigenfunctions together explain in total 95% and each explain 72.9%, 13.2%, and 8.9%, respectively, of total variation.*

We next compared the FVE by SFPCA and $L^2$ FPCA for $K = 1, \ldots, 10$ under the geodesic distance $d_{\mathcal{M}}$, as summarized in Table 2. Here SFPCA was applied on the spherical data on $S^2$, while $L^2$ FPCA was based on the latitude–longitude data in $\mathbb{R}^2$. Again SFPCA has higher FVE than the conventional $L^2$ FPCA for all choices of $K$, especially small $K$, where SFPCA shows somewhat better performance in terms of trajectory recovery.

TABLE 2
*The FVE (%) by the first K components for the proposed SFPCA and the* $L^2$ *FPCA for the flight data.*

| $K$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| SFPCA | 72.9 | 86.1 | 95.0 | 96.3 | 97.0 | 97.7 | 98.3 | 98.7 | 99.0 | 99.2 |
| $L^2$ FPCA | 71.2 | 84.9 | 94.6 | 96.1 | 96.8 | 97.4 | 98.1 | 98.4 | 98.8 | 99.1 |

We also aimed to predict the airline (BAW, CPA, and VIR) from an observed flight path by feeding the FPC scores obtained from either the proposed SFPCA or from the traditional $L^2$ FPCA into different multivariate classifiers, including linear discriminant analysis (LDA), logistic regression, and support vector machine (SVM) with radial basis kernel. For each of 200 Monte Carlo runs, we randomly selected 500 flights as training set for training and tuning and used the rest as test set to evaluate classification performance. The number of components $K$ for each classifier was either fixed at 10, 15, 20, 25, 30, or selected by five-fold cross-validation (CV). The results for prediction accuracy are in Table 3. The SFPCA based classifiers performed better or at least equally well as the $L^2$ FPCA based classifiers for nearly all choices of $K$ and classifier, where among the classifiers SVM performed best.

TABLE 3

*A comparison of airline classification accuracy (%) from observed flight trajectories, using the first $K$ components for SFPCA and $L^2$ FPCA (columns), with $K$ either fixed or chosen by CV, for various classifiers (rows). All standard errors for the accuracies are below 0.12%. The numbers in parenthesis are the number of components chosen by CV. S stands for SFPCA and L for $L^2$ FPCA; LDA, linear discriminant analysis; MN, multinomial logistic regression; SVM, support vector machine.*

| | $K=10$ | | $K=15$ | | $K=20$ | | $K=25$ | | $K=30$ | | $K$ chosen by CV | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | S | L | S | L | S | L | S | L | S | L | S | L |
| LDA | 76.9 | 75.8 | 79.6 | 78.4 | 81.9 | 81.5 | 82.7 | 82.5 | 83.5 | 82.3 | 83.2 (28.0) | 82.2 (26.2) |
| MN | 78.5 | 76.0 | 81.8 | 79.4 | 83.8 | 82.7 | 84.6 | 84.0 | 85.2 | 83.6 | 84.8 (27.5) | 83.7 (25.7) |
| SVM | 82.3 | 80.9 | 84.3 | 82.5 | 86.3 | 85.2 | 86.1 | 86.2 | 86.3 | 85.7 | 86.2 (24.6) | 85.8 (25.0) |

**6. Simulations.** To investigate the performance of trajectory recovery for the proposed RFPCA, we considered two scenarios of Riemannian manifolds: The Euclidean sphere $\mathcal{M} = S^2$ in $\mathbb{R}^3$, and the special orthogonal group $\mathcal{M} = \mathrm{SO}(3)$ of $3 \times 3$ rotation matrices, viewed as a Riemannian submanifold of $\mathbb{R}^{3\times 3}$. We compared three approaches: The Direct (D) method, which directly optimizes (3) over all time-varying geodesic submanifolds $\mathcal{M}_K$ and therefore serves as a gold standard, implemented through discretization; the proposed RFPCA method (R) and the classical $L^2$ FPCA method (L), which ignores the Riemannian geometry. In the direct method, the sample curves and time-varying geodesic submanifolds are discretized onto a grid of 20 equally-spaced time points, and a quasi-Newton algorithm is used to maximize the criterion function (3). We used FVE as our evaluation criterion, where models were fitted using $n = 50$ or 100 independent samples.

We briefly review the Riemannian geometry for the special orthogonal group $\mathcal{M} = \mathrm{SO}(N)$. The elements of $\mathcal{M}$ are $N \times N$ orthogonal matrices

with determinant 1, and the tangent space $T_p\mathcal{M}$ is identified with the collection of $N \times N$ skew-symmetric matrices. For $p, q \in \mathcal{M}$ and skew-symmetric matrices $u, v \in T_p\mathcal{M}$, the Riemannian metric is $\langle u, v\rangle = \operatorname{tr}(u^T v)$ where $\operatorname{tr}(\cdot)$ is the matrix trace; the Riemannian exponential map is $\exp_p(v) = \operatorname{Exp}(v)p$ and the logarithm map is $\log_p(q) = \operatorname{Log}(qp^{-1})$, where Exp and Log denote the matrix exponential and logarithm; the geodesic distance is $d_{\mathcal{M}}(p, q) = \left\|\operatorname{Log}(qp^{-1})\right\|_F$. For $N = 3$, the tangent space $T_p\mathcal{M}$ is 3-dimensional and can be identified with $\mathbb{R}^3$ through (Chavel 2006) $\iota : \mathbb{R}^3 \to T_p\mathcal{M}$, $\iota(a, b, c) = [0, -a, -b; a, 0, -c; b, c, 0]$.

The sample curves $X$ were generated as $X : \mathcal{T} = [0, 1] \to \mathcal{M}$, $X(t) = \exp_{\mu_{\mathcal{M}}(t)}(\sum_{k=1}^{20} \xi_k \phi_k(t))$, with mean function $\mu_{\mathcal{M}}(t) = \exp_{[0,0,1]}(2t, 0.3\pi \sin(\pi t), 0)$ for $\mathcal{M} = S^2$, and $\mu_{\mathcal{M}}(t) = \exp(\iota(2t, 0.3\pi \sin(\pi t), 0))$ for $\mathcal{M} = \mathrm{SO}(3)$. For $k = 1, \ldots, 20$, the RFPC scores $\xi_k$ were generated by independent Gaussian distributions with mean zero and variance $0.07^{k/2}$. The eigenfunctions were $\phi_k(t) = 2^{-1/2} R_t[\zeta_k(t/2), \zeta_k((t + 1)/2), 0]^T$ for $\mathcal{M} = S^2$ and $\phi_k(t) = 6^{-1/2}\iota(\zeta_k(t/3), \zeta_k((t+1)/3), \zeta_k((t+2)/3))$ for $\mathcal{M} = \mathrm{SO}(3)$, $t \in [0, 1]$, where $R_t$ is the rotation matrix from $[0, 0, 1]$ to $\mu_{\mathcal{M}}(t)$, and $\{\zeta_k\}_{k=1}^{20}$ is the orthonormal Legendre polynomial basis on $[0, 1]$. A demonstration of ten sample curves, the mean function, and the first three eigenfunctions for $\mathcal{M} = S^2$ is shown in Figure 6.

We report the mean FVE by the first $K = 1, \ldots, 4$ components for the investigated FPCA methods in Table 4, as well as the running time, based on 200 Monte Carlo repeats. The true FVEs for $K = 1, \ldots, 4$ components were 73.5%, 93.0%, 98.1%, and 99.5%, respectively. The proposed RFPCA method had higher FVE and thus outperformed the $L^2$ FPCA in all scenarios and for all $K$, which is expected since RFPCA takes into account the curved geometry. This advantage leads to a more parsimonious representation, e.g., in the $\mathcal{M} = S^2$ and $n = 100$ scenario, the average $K$ required by RFPCA to achieve at least FVE$> 0.95$ is one less than that for $L^2$ FPCA. The performance advantage of RFPCA over $L^2$ FPCA is larger for $\mathcal{M} = S^2$ than for $\mathcal{M} = \mathrm{SO}(3)$, since the former has larger sectional curvature (1 vs 1/8). The Direct method was as expected better than RFPCA (also for SO(3), which is not explicit in the table due to rounding), since the former optimizes the residual variation under the geodesic distance, the true target, while the latter uses the more tractable surrogate residual variation target (4) for $L^2$ distance on the tangent spaces.

Each experiment was run using a single processor (Intel Xeon E5-2670 CPU @ 2.60GHz) to facilitate comparisons. Both RFPCA and $L^2$ FPCA are quite fast in the and take only a few seconds, though RFPCA is 1.5–3 times slower, depending on the Riemannian manifold $\mathcal{M}$. The Direct method, how-

ever, was several magnitudes slower than RFPCA, due to the unstructured optimization problem, while for RFPCA spectral decomposition provides an effective solution. The slim performance gain for the Direct method as compared to RFPCA does not justify the huge computational effort.

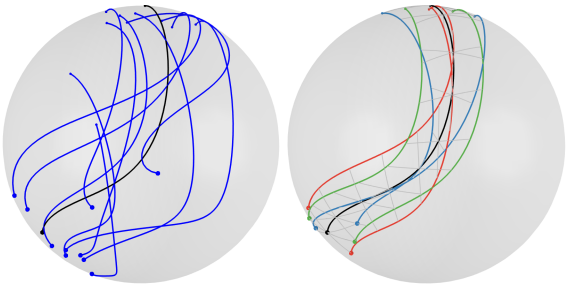


FIG 6. *Left: Ten randomly generated samples (dark blue) for $\mathcal{M} = S^2$. Right: The first three eigenfunctions (red, green, and blue, respectively) multiplied by 0.2 and then exponentially mapped from the mean function (solid black). Light gray lines connect the mean function and the eigenfunctions at 10 equally spaced time points. Small dots denote $t = 0$ and large dots $t = 1$.*

TABLE 4

*A comparison of mean FVE (%) and running time in the simulation study. D, direct optimization of (3) through discretization; R, RFPCA; L, $L^2$ FPCA. The standard errors of the FVEs for all three methods were below 0.32%.*

| | | $K=1$ | | | $K=2$ | | | $K=3$ | | | $K=4$ | | | Time (seconds) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\mathcal{M}$ | $n$ | D | R | L | D | R | L | D | R | L | D | R | L | D | R | L |
| $S^2$ | 50 | 74.3 | 74.1 | 71.4 | 93.0 | 92.9 | 89.6 | 98.1 | 97.9 | 93.8 | 99.5 | 99.2 | 97.5 | 5e3 | 0.72 | 0.24 |
| | 100 | 74.0 | 73.8 | 70.9 | 92.9 | 92.8 | 89.2 | 98.0 | 97.9 | 93.1 | 99.4 | 99.2 | 97.3 | 1e4 | 1.01 | 0.38 |
| SO(3) | 50 | 73.1 | 73.1 | 72.2 | 92.8 | 92.8 | 91.6 | 98.1 | 98.1 | 96.3 | 99.5 | 99.5 | 98.1 | 2e3 | 3.67 | 2.46 |
| | 100 | 72.9 | 72.9 | 71.8 | 92.6 | 92.6 | 91.3 | 98.0 | 98.0 | 96.1 | 99.5 | 99.5 | 97.9 | 4e3 | 6.58 | 4.94 |

## APPENDIX: PROOFS

PROOF OF PROPOSITION 1. Since $\mathcal{M}$ is a closed subset of $\mathbb{R}^{d_0}$ with the induced Riemannian metric by the Euclidean metric, $\mathcal{M}$ is complete. By the Hopf–Rinow theorem (see, e.g., Chavel 2006), $\mathcal{M}$ is geodesically complete, i.e. for all $p \in \mathcal{M}$, the exponential map $\exp_p$ is defined on the entire tangent space $T_p\mathcal{M}$. Therefore $X_K(t) = \exp_{\mu_{\mathcal{M}}(t)}(V_K(t))$ is well-defined.

The first inequality in (15) holds by the definition of projection $\Pi$. The second inequality follows from Alexandrov–Toponogov theorem (e.g., Theorem IX.5.1 in Chavel 2006), which states if two geodesic triangles $T_1$ and $T_2$ on complete Riemannian manifolds $\mathcal{M}_1$ and $\mathcal{M}_2$, where $\mathcal{M}_1$ has uniformly higher sectional curvature than $\mathcal{M}_2$, have in common the length of two sides

and the angle between the two sides, then $T_1$ has a shorter third side than $T_2$. This is applied to triangles $(X(t), \mu_{\mathcal{M}}(t), X_K(t))$ on $\mathcal{M}$ and $(V(t), 0, V_K(t))$ on $T_{\mu_{\mathcal{M}}(t)}$, identified with a Euclidean space. □

For the following proofs we consider the set

$$\mathcal{K} = \overline{\bigcup_{t\in\mathcal{T}} B_{\mathcal{M}}(\mu_{\mathcal{M}}(t), 2r)} \subset \mathcal{M}, \tag{29}$$

where $B_{\mathcal{M}}(p, l)$ is an open $d_{\mathcal{M}}$-geodesic ball of radius $l > 0$ centered at $p \in \mathcal{M}$, and $\overline{A}$ denotes the closure of a set $A$. Under (B1) and (B3), $\mathcal{K}$ is closed and bounded and thus is compact, with diameter $R = \sup_{p,q\in\mathcal{K}} d_{\mathcal{M}}(p, q)$. Then $\mu_{\mathcal{M}}(t), \hat{\mu}_{\mathcal{M}}(t), X(t) \in \mathcal{K}$ for all $t \in \mathcal{T}$. For the asymptotic results we will consider the compact set $\mathcal{K}$.

Proof of Proposition 2. To obtain the uniform consistency results of $\hat{\mu}_{\mathcal{M}}(t)$, we need to show

$$\sup_{t\in\mathcal{T}} \sup_{p\in\mathcal{K}} |M_n(p, t) - M(p, t)| = o_p(1), \tag{30}$$

$$\sup_{t\in\mathcal{T}} |M_n(\hat{\mu}_{\mathcal{M}}(t), t) - M(\mu_{\mathcal{M}}(t), t)| = o_p(1), \tag{31}$$

and for any $\epsilon > 0$, there exist $a = a(\epsilon) > 0$ such that

$$\inf_{t\in\mathcal{T}} \inf_{p:\, d_{\mathcal{M}}(p,\mu_{\mathcal{M}}(t))>\epsilon} [M_n(p, t) - M(\mu_{\mathcal{M}}(t), t)] \geq a - o_p(1). \tag{32}$$

Then by (31) and (32), for any $\delta > 0$, there exists $N \in \mathbb{N}$ such that $n \geq N$ implies the event

$$\begin{aligned} E = &\{\sup_{t\in\mathcal{T}} |M_n(\hat{\mu}_{\mathcal{M}}(t), t) - M(\mu_{\mathcal{M}}(t), t)| \leq a/3\} \cap \\ &\{\inf_{t\in\mathcal{T}} \inf_{p:\, d_{\mathcal{M}}(p,\mu_{\mathcal{M}}(t))>\epsilon} [M_n(p, t) - M(\mu_{\mathcal{M}}(t), t)] \geq 2a/3\} \end{aligned}$$

holds with probability greater than $1 - \delta$. This implies that on $E$, $\sup_{t\in\mathcal{T}} d_{\mathcal{M}}(\hat{\mu}_{\mathcal{M}}(t), \mu_{\mathcal{M}}(t)) \leq \epsilon$, and therefore the consistency of $\hat{\mu}_{\mathcal{M}}$.

Proof of (30): We first obtain the auxiliary result

$$\lim_{\delta\downarrow 0} E\left[\sup_{|t-s|<\delta} d_{\mathcal{M}}(X(t), X(s))\right] = 0 \tag{33}$$

by dominated convergence, (B1), and the boundedness of $\mathcal{K}$ (29). Note that for any $p, q, w \in \mathcal{K}$,

$$|d_{\mathcal{M}}(p, w)^2 - d_{\mathcal{M}}(q, w)^2| = |d_{\mathcal{M}}(p, w) + d_{\mathcal{M}}(q, w)| \cdot |d_{\mathcal{M}}(p, w) - d_{\mathcal{M}}(q, w)| \leq 2R d_{\mathcal{M}}(p, q)$$

by the triangle inequality, where $R$ is the diameter of $\mathcal{K}$. Then

$$\begin{aligned}\sup_{\substack{|t-s|<\delta\\ p,q\in\mathcal{K}\\ d_{\mathcal{M}}(p,q)<\delta}} |M_n(p,t)-M_n(q,s)| &\le \sup_{\substack{|t-s|<\delta\\ p,q\in\mathcal{K}\\ d_{\mathcal{M}}(p,q)<\delta}} |M_n(p,s)-M_n(q,s)| \\ &\quad + \sup_{\substack{|t-s|<\delta\\ p,q\in\mathcal{K}\\ d_{\mathcal{M}}(p,q)<\delta}} |M_n(p,t)-M_n(p,s)| \\ &\le 2R\delta + \frac{2R}{n}\sum_{i=1}^{n} \sup_{|t-s|<\delta} d_{\mathcal{M}}(X_i(t),X_i(s)) \\ &= 2R\delta + 2RE\left[\sup_{|t-s|<\delta} d_{\mathcal{M}}(X(t),X(s))\right] + o_p(1),\end{aligned}$$

where the last equality is due to the weak law of large numbers (WLLN). Due to (33), the quantity in the last display can be made arbitrarily close to zero (in probability) by letting $\delta \downarrow 0$ and $n \to \infty$. Therefore, for any $\epsilon > 0$ and $\eta > 0$, there exist $\delta > 0$ such that

$$\limsup_{n\to\infty} P\Big(\sup_{\substack{|t-s|<\delta\\ p,q\in\mathcal{K}\\ d_{\mathcal{M}}(p,q)<\delta}} |M_n(p,t)-M_n(q,s)| > \epsilon\Big) < \eta,$$

proving the asymptotic equicontinuity of $M_n$ on $\mathcal{K}\times\mathcal{T}$. This and the pointwise convergence of $M_n(p,t)$ to $M(p,t)$ by the WLLN imply (30) by Theorem 1.5.4 and Theorem 1.5.7 of van der Vaart and Wellner (1996).

Proof of (31): Since $\hat{\mu}_{\mathcal{M}}(t)$ and $\mu_{\mathcal{M}}(t)$ are the minimizers of $M_n(\cdot,t)$ and $M(\cdot,t)$, respectively, $|M_n(\hat{\mu}_{\mathcal{M}}(t),t) - M(\mu_{\mathcal{M}}(t),t)| \le \max(M_n(\mu_{\mathcal{M}}(t),t) - M(\mu_{\mathcal{M}}(t),t), M(\hat{\mu}_{\mathcal{M}}(t),t) - M_n(\hat{\mu}_{\mathcal{M}}(t),t)) \le \sup_{p\in\mathcal{K}} |M_n(p,t) - M(p,t)|$. Take suprema over $t\in\mathcal{T}$ and then apply (30) to obtain (31).

Proof of (32): Fix $\epsilon > 0$ and let $a = a(\epsilon) = \inf_{t\in\mathcal{T}} \inf_{p:\, d_{\mathcal{M}}(p,\mu_{\mathcal{M}}(t))>\epsilon} [M(p,t) - M(\mu_{\mathcal{M}}(t),t)] > 0$. For small enough $\epsilon$,

$$\begin{aligned}&\inf_{t\in\mathcal{T}} \inf_{p:\, d_{\mathcal{M}}(p,\mu_{\mathcal{M}}(t))>\epsilon} [M_n(p,t) - M(\mu_{\mathcal{M}}(t),t)] \\ &\quad = \inf_{t\in\mathcal{T}} \inf_{\substack{p\in\mathcal{K},\\ d_{\mathcal{M}}(p,\mu_{\mathcal{M}}(t))>\epsilon}} [M_n(p,t) - M(\mu_{\mathcal{M}}(t),t)] \\ &\quad = \inf_{t\in\mathcal{T}} \inf_{\substack{p\in\mathcal{K},\\ d_{\mathcal{M}}(p,\mu_{\mathcal{M}}(t))>\epsilon}} [M(p,t) - M(\mu_{\mathcal{M}}(t),t) + M_n(p,t) - M(p,t)] \\ &\quad \ge a - \sup_{t\in\mathcal{T}} \sup_{\substack{p\in\mathcal{K}\\ d_{\mathcal{M}}(p,\mu_{\mathcal{M}}(t))>\epsilon}} |M_n(p,t) - M(p,t)| = a - o_p(1),\end{aligned}$$

where the first equality is due to $\hat{\mu}_{\mathcal{M}}(t) \in \mathcal{K}$ and the continuity of $M_n$, the inequality to (B4), and the last equality to (30). For the continuity of $\mu_{\mathcal{M}}$, note for any $t_0, t_1 \in \mathcal{T}$,

$$\begin{aligned}|M(\mu_{\mathcal{M}}(t_1), t_0) - M(\mu_{\mathcal{M}}(t_0), t_0)| &\leq |M(\mu_{\mathcal{M}}(t_1), t_1) - M(\mu_{\mathcal{M}}(t_0), t_0)| \\ &\quad + |M(\mu_{\mathcal{M}}(t_1), t_0) - M(\mu_{\mathcal{M}}(t_1), t_1)| \\ &\leq \sup_{p \in \mathcal{K}} |M(p, t_1) - M(p, t_0)| + 2RE[d_{\mathcal{M}}(X(t_0), X(t_1))] \\ &\leq 4RE[d_{\mathcal{M}}(X(t_0), X(t_1))] \to 0\end{aligned}$$

as $t_1 \to t_0$ by (B1), where the second inequality is due to the fact that $\mu_{\mathcal{M}}(t_l)$ minimizes $M(\cdot, t_l)$ for $l = 0, 1$. Then by (B4), $d_{\mathcal{M}}(\mu_{\mathcal{M}}(t_1), \mu_{\mathcal{M}}(t_0)) \to 0$ as $t_1 \to t_0$, proving the continuity of $\mu_{\mathcal{M}}$. The continuity for $\hat{\mu}_{\mathcal{M}}$ is similarly proven by in probability arguments. □

Proof of Theorem 1. The proof idea is similar to that of Theorem 2.1 in Bhattacharya and Patrangenaru (2005). To lighten notations, let $Y(t) = \tau(X(t))$, $Y_i(t) = \tau(X_i(t))$, $\nu(t) = \tau(\mu_{\mathcal{M}}(t))$, and $\hat{\nu}(t) = \tau(\hat{\mu}_{\mathcal{M}}(t))$. The squared distance $d_{\mathcal{M}}(p, q)^2$ is smooth at $(p, q)$ if $d_{\mathcal{M}}(p, q) < \text{inj}_p$, due to the smoothness of the exponential map (Chavel 2006, Theorem I.3.2). Then $d_\tau(u, v)^2$ is smooth on the compact set $\{(u, v) \in \tau(U) \times \tau(U) \subset \mathbb{R}^d \times \mathbb{R}^d \mid d_{\mathcal{M}}(\tau^{-1}(u), \tau^{-1}(v)) \leq r\}$ and thus $T(Y(t), \nu(t))$ and $H(Y(t), \nu(t))$ are well defined, by (B3) and since the domain $U$ of $\tau$ is bounded. Define

$$h_t(v) = E[d_\tau(Y(t), v)^2], \tag{34}$$

$$h_{nt}(v) = \frac{1}{n} \sum_{i=1}^{n} d_\tau(Y_i(t), v)^2. \tag{35}$$

Since $\nu(t)$ is the minimal point of (34),

$$E[T_j(Y(t), \nu(t))] = E\left[ \frac{\partial}{\partial v_j} d_\tau^2(Y(t), v) \Big|_{v=\nu(t)} \right] = \frac{\partial}{\partial v_j} h_t(\nu(t)) = 0, \tag{36}$$

for $j = 1, \ldots, d$. Similarly, differentiating (35) and applying Taylor's theorem,

$$0 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} T_j(Y_i(t), \hat{\nu}(t))$$

(37)
$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} T_j(Y_i(t), \nu(t)) + \sum_{l=1}^{d} \sqrt{n}[\hat{\nu}_l(t) - \nu_l(t)] \frac{1}{n} \sum_{i=1}^{n} H_{jl}(Y_i(t), \nu(t)) + R_{nj}(t),$$

where $\hat{\nu}_l(t)$ and $\nu_l(t)$ are the $l$th component of $\hat{\nu}(t)$ and $\nu(t)$, and
(38)
$$R_{nj}(t) = \sum_{l=1}^{d} \sqrt{n}[\hat{\nu}_l(t) - \nu_l(t)] \frac{1}{n} \sum_{i=1}^{n} \left[ H_{jl}(Y_i(t), \tilde{\nu}_{jl}(t)) - H_{jl}(Y_i(t), \nu(t)) \right],$$

for some $\tilde{\nu}_{jl}(t)$ lying between $\hat{\nu}_l(t)$ and $\nu_l(t)$.

Due to the smoothness of $d_\tau^2$, (B3), and (B6), for $j, l = 1, \ldots, d$,

$$E \sup_{t \in \mathcal{T}} T_j(Y_i(t), \nu(t))^2 < \infty, \quad E \sup_{t \in \mathcal{T}} H_{jl}(Y_i(t), \nu(t))^2 < \infty, \tag{39}$$

$$\lim_{\epsilon \downarrow 0} E \sup_{t \in \mathcal{T}} \sup_{\|\theta - \nu(t)\| \le \epsilon} |H_{jl}(Y(t), \theta) - H_{jl}(Y(t), \nu(t))| = 0. \tag{40}$$

By (B6), we also have $\lim_{\epsilon \downarrow 0} E \sup_{|t-s|<\epsilon} |H_{jl}(Y(t), \nu(t)) - H_{jl}(Y(s), \nu(s))| \to 0$, which implies the asymptotic equicontinuity of $n^{-1} \sum_{i=1}^{n} H_{jl}(Y_i(t), \nu(t))$ on $t \in \mathcal{T}$, and thus

$$\sup_{t \in \mathcal{T}} \left| \frac{1}{n} \sum_{i=1}^{n} H_{jl}(Y_i(t), \nu(t)) - E[H_{jl}(Y_i(t), \nu(t))] \right| = o_p(1), \tag{41}$$

by Theorem 1.5.4 and Theorem 1.5.7 of van der Vaart and Wellner (1996). In view of (39)–(41) and Proposition 2, we may write (37) into matrix form

$$[\Lambda(t) + E_n(t)] \sqrt{n}[\hat{\nu}(t) - \nu(t)] = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} T(Y_i(t), \nu(t)), \tag{42}$$

where $\Lambda(t) = E[H(Y(t), \nu(t))]$ and $E_n(t)$ is some random matrix with $\sup_{t \in \mathcal{T}} \|E_n(t)\|_F = o_p(1)$. By (B6), $T_j(Y_i(t), \nu(t))$ is Lipschitz in $t$ with a square integrable Lipschitz constant, so one can apply a Banach space central limit theorem (Jain and Marcus 1975)

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} T(Y_i, \nu) \xrightarrow{L} W, \tag{43}$$

where $W$ is a Gaussian process with sample paths in $\mathcal{C}_d(\mathcal{T})$, mean 0, and covariance $G_T(t, s) = E[T(Y(t), \nu(t)) T(Y(s), \nu(s))^T]$.

We conclude the proof by showing

$$\inf_{t\in\mathcal{T}} \lambda_{\min}(\Lambda(t)) > 0. \tag{44}$$

Let $\phi_t(v) = \log_{\mu_{\mathcal{M}}(t)}(v)$, $f_t = \phi_t \circ \tau^{-1}$, and $g_t(v) = E[d_{\mathcal{M}}(X(t), \exp_{\mu_{\mathcal{M}}(t)}(v))^2]$, so $h_t(v) = g_t(f(v))$. Observe

$$\frac{\partial^2}{\partial v_j \partial v_l} h_t(v) = \left(\frac{\partial}{\partial v_j} f_t(v)\right)^T \frac{\partial^2}{\partial v^2} g_t(v) \left(\frac{\partial}{\partial v_l} f_t(v)\right) + \frac{\partial}{\partial v} g_t(v)^T \frac{\partial^2}{\partial v_j \partial v_l} f_t(v). \tag{45}$$

The second term vanishes at $v = \nu(t)$ by (36), so in matrix form

$$\Lambda(t) = \frac{\partial^2}{\partial v^2} h_t(\nu(t)) = \left(\frac{\partial}{\partial v} f_t(\nu(t))\right)^T \frac{\partial^2}{\partial v^2} g_t(0) \left(\frac{\partial}{\partial v} f_t(\nu(t))\right). \tag{46}$$

The gradient of $f_t$ is nonsingular at $\nu(t)$ since it is a local diffeomorphism. Then $\Lambda(t)$ is positive definite for all $t \in \mathcal{T}$ by (B5), and (44) follows by continuity.

□

Proof of Corollary 1. Note $d_{\mathcal{M}}(\hat{\mu}_{\mathcal{M}}(t), \mu_{\mathcal{M}}(t)) = d_\tau(\hat{\nu}(t), \nu(t))$. By Taylor's theorem around $v = \nu(t)$,

$$d_\tau(\nu(t), \hat{\nu}(t))^2 = [\hat{\nu}(t) - \nu(t)]^T \left[\frac{\partial}{\partial v^2} d_\tau^2(\nu(t), v)\Big|_{v=\tilde{\nu}(t)}\right] [\hat{\nu}(t) - \nu(t)],$$

where $\tilde{\nu}(t)$ lies between $\hat{\nu}(t)$ and $\nu(t)$, since $d_\tau^2(u, v)$ and $\partial d_\tau^2(u, v)/\partial v$ both vanish at $u = v$. The result then follows from Theorem 1, Remark 1, and Proposition 2. □

## SUPPLEMENTARY MATERIAL

**Supplement to "Principal Component Analysis for Functional Data on Riemannian Manifolds and Spheres"**
(doi: COMPLETED BY THE TYPESETTER; .pdf). In the Supplementary Materials, we provide proofs of Corollary 2, Theorem 2, and Corollary 4; algorithms for RFPCA of compositional data; and additional simulations.

## REFERENCES

Adriaenssens, N., Coenen, S., Versporten, A., Muller, A., Minalu, G., Faes, C., Vankerckhoven, V., Aerts, M., Hens, N. and Molenberghs, G. (2011). European Surveillance of Antimicrobial Consumption (ESAC): Outpatient antibiotic use in Europe (1997–2009). *Journal of Antimicrobial Chemotherapy* **66** vi3–vi12.

AFSARI, B. (2011). Riemannian $L^p$ center of mass: Existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society* **139** 655–673.

AITCHISON, J. (1986). *The Statistical Analysis of Compositional Data*. Chapman & Hall, London.

ANIRUDH, R., TURAGA, P., SU, J. and SRIVASTAVA, A. (2015). Elastic functional coding of human actions: From vector-fields to latent variables. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 3147–3155.

ANIRUDH, R., TURAGA, P., SU, J. and SRIVASTAVA, A. (2017). Elastic functional coding of Riemannian trajectories. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39** 922–936.

BHATTACHARYA, R. and LIN, L. (2017). Omnibus CLTs for Fréchet means and nonparametric inference on non-Euclidean spaces. *Proceedings of the American Mathematical Society* **145** 413–428.

BHATTACHARYA, R. and PATRANGENARU, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds - I. *Annals of Statistics* **31** 1–29.

BHATTACHARYA, R. and PATRANGENARU, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds -II. *Annals of statistics* **33** 1225–1259.

BOSQ, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*. Springer-Verlag, New York.

CAREY, J. R., PAPADOPOULOS, N. T., KOULOUSSIS, N. A., KATSOYANNOS, B. I., MÜLLER, H.-G., WANG, J.-L. and TSENG, Y.-K. (2006). Age-specific and lifetime behavior patterns in *Drosophila melanogaster* and the Mediterranean fruit fly, *Ceratitis capitata*. *Experimental Gerontology* **41** 93–97.

CASTRO, P. E., LAWTON, W. H. and SYLVESTRE, E. A. (1986). Principal modes of variation for processes with continuous sample curves. *Technometrics* **28** 329-337.

CHAVEL, I. (2006). *Riemannian Geometry: A Modern Introduction*. Cambridge University Press, Cambridge.

CHEN, D. and MÜLLER, H.-G. (2012). Nonlinear manifold representations for functional data. *Annals of Statistics* **40** 1–29.

CHIOU, J.-M., CHEN, Y.-T. and YANG, Y.-F. (2014). Multivariate functional principal component analysis: A normalization approach. *Statistica Sinica* **24** 1571–1596.

CORNEA, E., ZHU, H., KIM, P. and IBRAHIM, J. G. (2017). Regression models on Riemannian symmetric spaces. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79** 463–482.

FISHER, N. I., LEWIS, T. and EMBLETON, B. J. J. (1987). *Statistical Analysis of Spherical Data*. Cambridge University Press, Cambridge.

FLETCHER, P. T., LU, C., PIZER, S. M. and JOSHI, S. (2004). Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on Medical Imaging* **23** 995–1005.

HSING, T. and EUBANK, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley, Hoboken.

HUCKEMANN, S. F. and ELTZNER, B. (2016). Backward nested descriptors asymptotics with inference on stem cell differentiation. *ArXiv e-prints*. arXiv:1609.00814.

HUCKEMANN, S., HOTZ, T. and MUNK, A. (2010). Intrinsic shape analysis: Geodesic PCA for Riemannian manifolds modulo isometric Lie group actions. *Statistica Sinica* **20** 1–58.

JAIN, N. C. and MARCUS, M. B. (1975). Central limit theorems for C(S)-valued random variables. *Journal of Functional Analysis* **19** 216–231.

JUNG, S., DRYDEN, I. L. and MARRON, J. S. (2012). Analysis of principal nested spheres. *Biometrika* **99** 551–568.

Jupp, P. E. and Kent, J. T. (1987). Fitting smooth paths to spherical data. *Journal of the Royal Statistical Society. Series C* **36** 34–46.

Kendall, D. G., Barden, D., Carne, T. K. and Le, H. (2009). *Shape and Shape Theory*. Wiley, Hoboken.

Kent, J. T., Mardia, K. V., Morris, R. J. and Aykroyd, R. G. (2001). Functional models of growth for landmark data. In *Proceedings in Functional and Spatial Data Analysis* 109–115.

Kneip, A. and Utikal, K. J. (2001). Inference for density families using functional principal component analysis. *Journal of the American Statistical Association* **96** 519–542. MR1946423

Lila, E., Aston, J. A. D. and Sangalli, L. M. (2016). Smooth principal component analysis over two-dimensional manifolds with an application to neuroimaging. *The Annals of Applied Statistics* **10** 1854–1879.

Lin, Z. and Yao, F. (2017). Functional regression with unknown manifold structures. *ArXiv e-prints*. arXiv:1704.03005.

Lin, L., Thomas, B. S., Zhu, H. and Dunson, D. B. (2016). Extrinsic local regression on manifold-valued data. *Journal of the American Statistical Association* to appear.

Mardia, K. V. and Jupp, P. E. (2009). *Directional Statistics*. John Wiley & Sons, Hoboken.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and Its Applications* **9** 141–142.

Patrangenaru, V. and Ellingson, L. (2015). *Nonparametric Statistics on Manifolds and Their Applications to Object Data Analysis*. CRC Press, Boca Raton.

Petersen, A. and Müller, H.-G. (2016a). Functional data analysis for density functions by transformation to a Hilbert space. *The Annals of Statistics* **44** 183–218.

Petersen, A. and Müller, H. G. (2016b). Fréchet regression for random objects. *ArXiv e-prints*. arXiv:1608.03012.

Qiu, Z., Song, X. K. and Tan, M. (2008). Simplex mixed-effects models for longitudinal proportional data. *Scandinavian Journal of Statistics* **35** 577–596.

Rahman, I. U., Drori, I., Stodden, V. C., Donoho, D. L. and Schröder, P. (2005). Multiscale representations for manifold-valued data. *Multiscale Modeling & Simulation* **4** 1201–1232.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*, Second ed. Springer, New York. MR2168993

Telschow, F. J. E., Huckemann, S. F. and Pierrynowski, M. R. (2016). Functional inference on rotational curves and identification of human gait at the knee joint. *ArXiv e-prints*. arXiv:1611.03665.

Tournier, M., Wu, X., Courty, N., Arnaud, E. and Reveret, L. (2009). Motion compression using principal geodesics analysis. In *Computer Graphics Forum* **28** 355–364.

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York.

Wang, J.-L., Chiou, J.-M. and Müller, H.-G. (2016). Functional data analysis. *Annual Review of Statistics and its Application* **3** 257–295.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā Series A* **26** 359–372. MR0185765

Zheng, Y. (2015). Trajectory data mining: An overview. *ACM Transactions on Intelligent Systems and Technology* **6** 29:1–29:41.

## SUPPLEMENTARY MATERIALS

### S1. Additional Proofs.

PROOF OF COROLLARY 2.

$$\begin{aligned}\sup_{t\in\mathcal{T}}\|\hat{V}_i(t) - V_i(t)\|_E &= \sup_{t\in\mathcal{T}}\|\log_{\mu_{\mathcal{M}}(t)}(X_i(t)) - \log_{\hat{\mu}_{\mathcal{M}}(t)}(X_i(t))\|_E \\ &\lesssim \sup_{t\in\mathcal{T}}|d_{\mathcal{M}}(\hat{\mu}_{\mathcal{M}}(t), \mu_{\mathcal{M}}(t))|,\end{aligned}$$

where the last inequality is due to (B3) and the fact that $\log_p(q)$ is continuously differentiable in $(p, q)$ (Theorem I.3.2 in Chavel 2006). □

PROOF OF THEOREM 2. Denote $\tilde{G}(t,s) = \frac{1}{n}\sum_{i=1}^n V_i(t)V_i(s)^T$. Then

$$\begin{aligned}\sup_{t,s\in\mathcal{T}}\left\|\hat{G}(t,s) - G(t,s)\right\|_F &\leq \sup_{t,s\in\mathcal{T}}\left\|\hat{G}(t,s) - \tilde{G}(t,s)\right\|_F + \sup_{t,s\in\mathcal{T}}\left\|\tilde{G}(t,s) - G(t,s)\right\|_F \\ &\leq \frac{1}{n}\sum_{i=1}^n \sup_{t,s\in\mathcal{T}}\left\|\hat{V}_i(t)\hat{V}_i(s)^T - V_i(t)V_i(s)^T\right\|_F + \sup_{t,s\in\mathcal{T}}\left\|\frac{1}{n}\sum_{i=1}^n V_i(t)V_i(s)^T - G(t,s)\right\|_F\end{aligned} \tag{47}$$

Since $\sup_{t,s\in\mathcal{T}}\left\|V_i(t)V_i(s)^T\right\|_F < R^2$, viewing $V_i(t)V_i(s)^T$ as random elements in $L_\infty(\mathcal{T}\times\mathcal{T}, \mathbb{R}^{d^2})$ the second term is $O_p(n^{-1/2})$ by Theorem 2.8 in Bosq (2000). For the first term, note

$$\begin{aligned}\left\|\hat{V}_i(t)\hat{V}_i(s)^T - V_i(t)V_i(s)^T\right\|_F &\leq \left\|(\hat{V}_i(t) - V_i(t))\hat{V}_i(s)^T\right\|_F + \left\|V_i(t)(\hat{V}_i(s) - V_i(s))^T\right\|_F \\ &\leq \|\hat{V}_i(s)\|_E\|\hat{V}_i(t) - V_i(t)\|_E + \|V_i(t)\|_E\|\hat{V}_i(s) - V_i(s)\|_E \\ &\lesssim \sup_{t\in\mathcal{T}} d_{\mathcal{M}}(\hat{\mu}_{\mathcal{M}}(t), \mu_{\mathcal{M}}(t)),\end{aligned}$$

where the second inequality is due to the properties of the Frobenius norm, and the last is due to Corollary 2 and (B3). Therefore, by Corollary 1 the first term in (47) is $O_p(n^{-1/2})$ and (20) follows. Result (21) follows from applying Theorem 4.2.8 in Hsing and Eubank (2015) and from the fact that the operator norm is dominated by the Hilbert-Schmidt norm.

To prove (22), Theorem 5.1.8 in Hsing and Eubank (2015) and Bessel's inequality imply

$$\left\|\hat{\phi}_k - \phi_k\right\| = O_p(n^{-1/2}). \tag{48}$$

Then note that for any $t \in \mathcal{T}$,

$$\|\hat{\phi}_k(t) - \phi_k(t)\|_E = \|\int \frac{1}{\hat{\lambda}_k}\hat{G}(t,s)\hat{\phi}_k(s)ds - \int \frac{1}{\lambda_k}G(t,s)\phi_k(s)ds\|_E$$

$$
\begin{aligned}
&= \|(\frac{1}{\hat{\lambda}_k} - \frac{1}{\lambda_k}) \int \hat{G}(t,s)\hat{\phi}_k(s)ds + \frac{1}{\lambda_k}\int \hat{G}(t,s)(\hat{\phi}_k(s) - \phi_k(s)) \\
&+ (\hat{G}(t,s) - G(t,s))\phi_k(s)ds\|_E \\
&= O_p\left(\left|\frac{1}{\hat{\lambda}_k} - \frac{1}{\lambda_k}\right| + \left\|\hat{\phi}_k - \phi_k\right\| + \sup_{t,s\in\mathcal{T}} \left\|\hat{G}(t,s) - G(t,s)\right\|_F\right),
\end{aligned}
$$

which is of order $O_p(n^{-1/2})$ by (21), (48), and (20). Since the r.h.s. does not involve $t$, taking suprema on both sides over $t \in \mathcal{T}$ concludes the proof. □

Proof of Corollary 4. Conditions (A1)–(A2) and (B3) hold for longitudinal compositional data analysis, while (B2) holds by Theorem 2.1 in Afsari (2011). □

**S2. Algorithms for the RFPCA of Compositional Data.** The following Algorithms 1–3 are provided as examples for RFPCA applied to longitudinal compositional data $Z(t)$ or spherical trajectories $X(t)$. For longitudinal compositional data $Z(t)$, we initialize by defining $X(t)$ as the componentwise square root of $Z(t)$, which then lies on a Euclidean sphere $S^d$. We assume the trajectories are observed at $t = t_j = (j-1)/(m-1)$ for $j = 1, \ldots, m$, and all vectors are by default column vectors. Very similar algorithms for SFPCA have also been proposed by Anirudh et al. (2017).

The time complexity for Algorithm 1 is $O(nmf(d) + nm^2d^2 + (md)^3)$, where $f(d)$ is the cost for calculating a Fréchet mean in Line 2, which is typically $O(nd)$ or $O(nd^2)$ for gradient descent or quasi-Newton type optimizers per iteration, respectively. The most demanding computational step for the multivariate FPCA is $O(nm^2d^2)$ for Line 7 and $O((md)^3)$ for the eigendecomposition in Line 8. The computational cost for Algorithm 2 is $O(md)$ and that for Algorithm 3 is $O(nmd)$.

**Algorithm 2:** Truncated $K$-dimensional representations

**Data:** $\hat{\mu}_{\mathcal{M}}(t)$, $\{(\hat{\xi}_{ik}, \hat{\phi}_k(t))\}_{k=1}^K$
**Result:** $\hat{X}_{iK}(t)$, $\hat{V}_{iK}(t)$
1 **for** $j \in \{1, \ldots, m\}$ **do**
2 $\quad \hat{V}_{iK}(t_j) \leftarrow \sum_{k=1}^K \hat{\xi}_{ik}\hat{\phi}_k(t_j)$
3 $\quad \hat{X}_{iK}(t_j) \leftarrow$
$\quad\quad \cos(\|\hat{V}_{iK}(t_j)\|_E)\hat{\mu}_{\mathcal{M}}(t) + \sin(\|\hat{V}_{iK}(t_j)\|_E)\|\hat{V}_{iK}(t_j)\|_E^{-1}\hat{V}_{iK}(t_j)$
4 **end**

**Algorithm 1:** Spherical functional principal component analysis (SFPCA)

**Data:** $S^d$-valued trajectories $X_1(t), \ldots, X_n(t)$
**Result:** $\hat{\mu}_{\mathcal{M}}(t)$, $\hat{V}_i(t)$, $\hat{\xi}_{ik}$, $\hat{\phi}_k(t)$, $\hat{\lambda}_k$, for $i = 1, \ldots, n$ and $k = 1, \ldots, K$
`// Obtain the intrinsic mean function and tangent vectors`
**for** $j \in \{1, \ldots, m\}$ **do**
  $\hat{\mu}_{\mathcal{M}}(t_j) \leftarrow \arg\min_{p \in S^d} n^{-1} \sum_{i=1}^{n} [\cos^{-1}(p^T X_i(t_j))]^2$
  **for** $i \in \{1, \ldots, n\}$ **do**
    $\hat{V}_i(t_j) = \frac{u}{\sqrt{u^T u}} \cos^{-1}(\hat{\mu}_{\mathcal{M}}(t_j)^T X_i(t_j))$, where
    $u = X_i(t_j) - (\hat{\mu}_{\mathcal{M}}(t_j)^T X_i(t_j)) \mu_{\mathcal{M}}(t_j)$
  **end**
**end**
`// A multivariate FPCA. Vec(A) stacks the columns of A.`
$\hat{\mathbf{V}}_i \leftarrow [\hat{V}_i(t_1), \ldots, \hat{V}_i(t_m)]^T$, $\hat{\mathbf{G}} \leftarrow n^{-1} \sum_{i=1}^{n} \mathtt{Vec}(\hat{\mathbf{V}}_i)\mathtt{Vec}(\hat{\mathbf{V}}_i)^T$
$[\omega, \boldsymbol{\Psi}] \leftarrow \mathtt{Eigen}(\hat{\mathbf{V}}_i)$, for eigenvalues $\omega = [\omega_1, \ldots, \omega_m]^T$ and eigenvectors $\boldsymbol{\Psi} = [\psi_1, \ldots, \psi_m]$
**for** $k \in \{1, \ldots, K\}$ **do**
  Write $\hat{\boldsymbol{\Phi}}_k = [\hat{\phi}_k(t_1), \ldots, \hat{\phi}_k(t_m)]^T$, $\mathtt{Vec}(\hat{\boldsymbol{\Phi}}_k) \leftarrow m^{1/2}\psi_k$, $\hat{\lambda}_k \leftarrow m^{-1}\omega_k$, $\hat{\xi}_{ik} \leftarrow m^{-1}\mathtt{Vec}(\hat{\mathbf{V}}_i)^T \mathtt{Vec}(\hat{\boldsymbol{\Phi}}_k)$
**end**

**S3. Additional simulations.** We conducted an additional simulation study to investigate the scalability of the RFPCA algorithms to higher dimensions $d$, on the unit sphere $\mathcal{M} = S^d$ in $\mathbb{R}^{d+1}$ for $d = 5, 10, 15, 20$. Table 5 shows that the RFPCA scales well for larger dimensions in terms of running time, and its relative disadvantage in speed as compared to the $L^2$ FPCA becomes smaller as $d$ and $n$ get larger.

The samples were generated in the same fashion as in the main text, except for the mean function $\mu_{\mathcal{M}}(t) = \exp_{p_0}(2(d-1)^{-1/2}t, \ldots, 2(d-1)^{-1/2}t, 0.3\pi \sin(\pi t), 0)$, and eigenfunctions $\phi_k(t) = d^{-1/2} R_t[\zeta_k(t/d), \ldots, \zeta_k(t/d+$

**Algorithm 3:** Calculate FVE

**Data:** Outputs from Algorithm 1
**Result:** $\widehat{\mathrm{FVE}}_K$
$\hat{U}_0 \leftarrow n^{-1} \sum_{i=1}^{n} \int_{\mathcal{T}} d_{\mathcal{M}}{}^2(X_i(t), \hat{\mu}_{\mathcal{M}}(t))dt$
**for** $i \in \{1, \ldots, n\}$ **do**
  Use Algorithm 2 to obtain $\hat{X}_{iK}(t)$
**end**
$\hat{U}_K \leftarrow n^{-1} \sum_{i=1}^{n} \int_{\mathcal{T}} d_{\mathcal{M}}{}^2(X_i(t), \hat{X}_{iK}(t))dt$
$\widehat{\mathrm{FVE}}_K = (\hat{U}_0 - \hat{U}_K)/\hat{U}_0$

$(d-1)/d), 0]^T$, where $p_0 = [0, \ldots, 0, 1]$ and $R_t$ is the rotation matrix from $p_0$ to $\mu_{\mathcal{M}}(t)$.

TABLE 5
*A comparison of mean running time for $S^d$. The standard errors are below 2% of the means.*

| | $n = 50$ | | | | $n = 100$ | | | | $n = 200$ | | | | $n = 400$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $d$ | 5 | 10 | 15 | 20 | 5 | 10 | 15 | 20 | 5 | 10 | 15 | 20 | 5 | 10 | 15 | 20 |
| RFPCA | 1.3 | 1.7 | 2.1 | 2.7 | 1.9 | 2.6 | 3.3 | 4.3 | 3.1 | 4.6 | 5.7 | 7.4 | 5.9 | 7.8 | 10.5 | 13.2 |
| $L^2$ FPCA | 0.4 | 0.7 | 1.0 | 1.5 | 0.8 | 1.2 | 1.8 | 2.4 | 1.4 | 2.5 | 3.5 | 4.4 | 3.0 | 4.4 | 6.3 | 8.2 |

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA, DAVIS
DAVIS, CA 95616 USA
E-MAIL: dai@ucdavis.edu
E-MAIL: hgmueller@ucdavis.edu