MODELING LONGITUDINAL DATA ON RIEMANNIAN MANIFOLDS

Xiongtao Dai¹, Zhenhua Lin² and Hans-Georg Müller²

Department of Statistics, Iowa State University, Ames, IA 50011 USA
Department of Statistics, University of California, Davis
Davis, CA 95616 USA

October 2018

ABSTRACT

When considering functional principal component analysis for sparsely observed longitudinal data that take values on a nonlinear manifold, a major challenge is how to handle the sparse and irregular observations that are commonly encountered in longitudinal studies. Addressing this challenge, we provide theory and implementations for a manifold version of the principal analysis by conditional expectation (PACE) procedure that produces representations intrinsic to the manifold, extending a well-established version of functional principal component analysis targeting sparsely sampled longitudinal data in linear spaces. Key steps are local linear smoothing methods for the estimation of a Fréchet mean curve, mapping the observed manifold-valued longitudinal data to tangent spaces around the estimated mean curve, and applying smoothing methods to obtain the covariance structure of the mapped data. Dimension reduction is achieved via representations based on the first few leading principal components. A finitely truncated representation of the original manifold-valued data is then obtained by mapping these tangent space representations to the manifold. We show that the proposed estimates of mean curve and covariance structure achieve state-of-the-art convergence rates. For longitudinal emotional well-being data for unemployed workers as an example of time-dynamic compositional data that are located on a sphere, we demonstrate that our methods lead to interpretable eigenfunctions and principal component scores, which are defined on tangent spaces. In a second example, we analyze the body shapes of wallabies by mapping the relative size of their body parts onto a spherical pre-shape space. Compared to standard functional principal component analysis, which is based on Euclidean geometry, the proposed approach leads to improved trajectory recovery for sparsely sampled data on nonlinear manifolds.

KEYWORDS: Longitudinal Compositional Data, Data on Spheres, Dimension Reduction, Functional Data Analysis, Principal Component Analysis, Sparse and Irregular Data.

Research supported by NSF Grant DMS-1712864
1. INTRODUCTION

Functional data are usually considered as elements of a Hilbert space (Horvath & Kokoszka, 2012; Hsing & Eubank, 2015; Wang et al., 2016), a linear space with Euclidean geometry, where typical tools include functional principal component analysis (Kleffe, 1973; Hall & Hosseini-Nasab, 2006; Chen & Lei, 2015) and functional regression (Hall & Horowitz, 2007; Kong et al., 2016; Kneip et al., 2016). Considerably less work has been done on the analysis of nonlinear functional data, which are increasingly encountered in practice, such as $SO(3)$-valued functional data (Telschow et al., 2016), recordings of densely sampled trajectories on the sphere, including flight trajectories (Anirudh et al., 2017; Dai & Müller, 2018), or functions residing on unknown manifolds (Chen & Müller, 2012).

Since functional data are intrinsically infinite-dimensional, dimension reduction is a necessity, and a convenient and popular tool for this is functional principal component analysis, which is geared towards linear functional data and is not suitable for functional data on nonlinear manifolds, for which Dai & Müller (2018) investigated an intrinsic Riemannian Functional Principal Component Analysis (Riemannian FPCA) for functions taking values on a nonlinear Euclidean submanifold, with a Fréchet type mean curve. The concept of Fréchet mean as a minimizer of the Fréchet function extends the classical mean in Euclidean spaces to data on Riemannian manifolds (Patrangenaru et al., 2018). Using Riemannian logarithm maps, data on manifolds can be mapped into tangent spaces identified with hyperplanes in the ambient space of the manifold. Then Riemannian FPCA can be conducted on the mapped data, where the Fréchet mean and Riemannian logarithm maps reflect the curvature of the underlying manifold, yielding representations that are intrinsic to the manifold, which is an advantage over extrinsic approaches.

A challenge is that functional data are often sparsely observed, i.e. each function is only recorded at an irregular grid consisting of a few points. Such sparse recordings are routinely encountered in longitudinal studies (Verbeke et al., 2014). For example, in a longitudinal survey of unemployed workers in New Jersey (Krueger & Mueller, 2011) that we analyze...
in Section 4, the number of longitudinal responses available per subject is less than 4 in more than a half of the subjects. For sparsely observed longitudinal/functional data such as these, observations need to be pooled across subjects in order to obtain sensible estimates of mean and covariance functions, as the data available for individual subjects are so sparse that meaningful smoothing is not possible. This pooling idea is at the core of the principal analysis by conditional expectation (PACE) approach (Yao et al., 2005), whereas for densely sampled functional data one can apply individual curve smoothing or cross-sectional strategies (Zhang & Chen, 2007).

An special case of longitudinal data are longitudinal compositional data, where at each time point one observes fractions or percentages for each of a fixed number of categories, which add up to one. Such data occur in many applications, eg., repeated voting, with counts transformed into percentages of votes for items, consumer preferences in terms of what fraction prefers a certain item, microbiome (Li, 2015), online prediction markets, soil or air composition over time, mood assessment, and shape analysis, where we will study data of the latter two types of longitudinal data in Section 4. While a classical approach for compositional data is to apply the simplex geometry in the form of the Aitchison geometry (Aitchison, 1986) or a variant (Egozcue et al., 2003; Talská et al., 2018), a disadvantage is that a baseline category needs to be identified, which cannot have null outcomes, due to the need to form quotients; this is is especially difficult to satisfy in longitudinal studies, where null outcomes may fluctuate between categories.

Motivated by the need to analyze sparsely sampled longitudinal data as for example found in the emotional well-being data collected in a longitudinal survey for unemployed workers in New Jersey and containing a substantial proportion of null outcomes, we develop a Riemannian principal analysis method geared towards sparsely and irregularly observed Riemannian functional data.

The main contributions of this paper are three-fold:

(1) We develop a principal component analysis for longitudinal compositional data, which we also illustrate with sparsely sampled body shape growth curves of Tammar wallabies,
extending the scope of the approach of Dai & Müller (2018), which only applies to densely observed data. To our knowledge, no methods exist yet for the analysis of longitudinal data on manifolds.

(2) We extend Fréchet regression Petersen & Müller (2018) to functional data, while the approach in Petersen & Müller (2018) was restricted to nonfunctional data as dependence between repeated measurements is not taken into account.

(3) Concerning theoretical analysis, we extend the techniques developed in (Li & Hsing, 2010; Zhang & Wang, 2016) to manifold-valued data and obtain rates of uniform convergence for the mean function. The lack of a vector space structure makes this technically challenging.

To obtain intrinsic representations of the unobserved trajectories on a nonlinear Riemannian manifold from sparsely observed longitudinal data, we first pool data from all subjects to obtain estimates for the mean and covariance function, and then obtain estimates of the individual principal components and trajectories by Best Linear Unbiased Prediction (BLUP). We employ a manifold local linear smoothing approach to estimate the Fréchet mean curve, extending the approach of Petersen & Müller (2018) for sparsely observed Riemannian functional data. Local linear smoothing was originally studied in the context of Euclidean non-functional data (Fan & Gijbels, 1996) and later has been extended to curved non-functional data (Yuan et al., 2012). Observations of each function are then mapped into the tangent spaces around the estimated mean curve via Riemannian logarithm maps. As the log-mapped observations are vectors in the ambient space of the manifold, we proceed by adopting a scatterplot smoothing approach to estimate the covariance structure of the log-mapped data and then obtain a finitely truncated representation, where the principal component scores are estimated by PACE, or sometimes integration, depending on the sparseness of the observations available per function. Finally, a finite-dimensional representation for the original data is obtained by applying Riemannian exponential maps that pull the log-mapped data back to the manifold.
2. METHODOLOGY

2.1. Statistical Model

Let \( \mathcal{M} \) be a \( d \)-dimensional, connected and geodesically complete Riemannian submanifold of \( \mathbb{R}^D \), where \( d \) and \( D \) are positive integers such that \( d \leq D \). The dimension \( d \) is the intrinsic dimension of the manifold \( \mathcal{M} \), while \( D \) is the ambient dimension. The Riemannian metric \( \langle \cdot, \cdot \rangle \) on \( \mathcal{M} \), which defines a scalar product \( \langle \cdot, \cdot \rangle_p \) for the tangent space \( T_p \mathcal{M} \) at each point \( p \in \mathcal{M} \), is induced by the canonical inner product of \( \mathbb{R}^D \), and it also induces a geodesic distance function \( d_\mathcal{M} \) on \( \mathcal{M} \). A brief introduction to Riemannian manifolds can be found in the appendix of Dai & Müller (2018), see also Lang (1995) and Lee (1997).

We define a \( \mathcal{M} \)-valued Riemannian random process, or simply Riemannian random process \( X(t) \), as a \( D \)-dimensional vector-valued random process defined on a compact domain \( \mathcal{T} \subset \mathbb{R} \) such that \( X(t) \in \mathcal{M} \), where we assume that the process \( X \) is of second-order, in the sense that for every \( t \in \mathcal{T} \), there exists \( p \in \mathcal{M} \), potentially depending on \( t \), such that the Fréchet variance \( M(p,t) := \mathbb{E}d_\mathcal{M}^2(p,X(t)) \) is finite. For a fixed \( t \), if \( p \) is a point on \( \mathcal{M} \) satisfying \( M(p,t) = \inf_{q \in \mathcal{M}} M(q,t) \), then \( p \) is a Fréchet mean of \( X \) at \( t \). Under conditions described in Bhattacharya & Patrangenaru (2003), the Fréchet mean of a random variable on a manifold exists and is unique, which we shall assume for \( X(t) \) at all \( t \in \mathcal{T} \).

\((X0) \quad X \) is of second-order, and the Fréchet mean curve \( \mu(t) \) exists and is unique.

Formally, we define the unique Fréchet mean function \( \mu \) by

\[
\mu(t) = \arg \min_{p \in \mathcal{M}} M(p,t), \quad t \in \mathcal{T}.
\]  

(1)

As \( \mathcal{M} \) is geodesically complete, by the Hopf–Rinow theorem, its exponential map \( \text{Exp}_p \) at each \( p \) is defined on the entire \( T_p \mathcal{M} \). To make \( \text{Exp}_p \) injective, define the domain \( \mathcal{D}_p \) to be the interior of the collection of tangent vectors \( v \in T_p \mathcal{M} \) such that if \( \gamma(t) = \text{Exp}_p(tv) \) is a geodesic emanating from \( p \) with the direction \( v \), then \( \gamma([0,1]) \) is a minimizing geodesic.
Then on the domain $\mathcal{D}_p$ the map $\text{Exp}_p$ is injective, and its image is denoted by $\text{Im}(\text{Exp}_p)$. The Riemannian logarithm map at $p$, denoted by $\text{Log}_p$, is the inverse of $\text{Exp}_p$ restricted to $\text{Im}(\text{Exp}_p)$. Specifically, if $q = \text{Exp}_p v$ for some $v \in \mathcal{D}_p$, then $\text{Log}_p q = v$. To study the covariance structure of the random process $X$ on tangent spaces, we will assume

(X1) For some constant $\epsilon_0 > 0$, $\Pr\{X(t) \in \mathcal{M}\backslash(\mathcal{M}\backslash\text{Im}(\text{Exp}_{\mu(t)}))^\epsilon_0 \text{ for all } t \in \mathcal{T}\} = 1$,

where $A^\epsilon_0$ denotes the set $\bigcup_{p \in A} \{q \in \mathcal{M} : d_{\mathcal{M}}(p, q) < \epsilon_0\}$.

This condition requires $X(t)$ to stay away from the cut locus of $\mu(t)$ uniformly for all $t \in \mathcal{T}$, which is necessary for the logarithm map $\text{Log}_{\mu(t)}$ to be well defined, and is not needed if $\text{Exp}_{\mu(t)}$ is injective on $T_{\mu(t)} \mathcal{M}$ for all $t$. In the special case of a $d$-dimensional unit sphere $\mathbb{S}^d$, if $X(t)$ is continuous and the distribution of $X(t)$ vanishes at an open set with positive volume that contains $\mathcal{M}\backslash\text{Im}(\text{Exp}_{\mu(t)})$, (X1) holds. Under (X0) and (X1), $\text{Log}_{\mu(t)} X(t)$ is almost surely defined for all $t \in \mathcal{T}$. We will write $L(t)$ to denote the $\mathbb{R}^D$-valued random process $\text{Log}_{\mu(t)} X(t)$ and refer to $L$ as the log process of $X$.

An important observation (Bhattacharya & Patrangenaru, 2003) is that $E L(\cdot) \equiv 0$. Furthermore, the second-order condition on $X$ passes on to $L$, i.e., $E \|L(t)\|_2^2 = E d_{\mathcal{M}}^2(\mu(t), X(t)) < \infty$ for every $t \in \mathcal{T}$, where $\| \cdot \|_2$ denotes the canonical Euclidean norm in $\mathbb{R}^D$. This enables us to define the covariance function of $L$ by

$$
\Gamma(s, t) = E\{L(s)L(t)^T\}, \quad s, t \in \mathcal{T}. \tag{2}
$$

This covariance function admits the eigendecomposition

$$
\Gamma(s, t) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s) \phi_k^T(t),
$$

where $\phi_k$ are orthonormal, $\lambda_k \geq \lambda_{k+1}$, and $\sum_{k=1}^{\infty} \lambda_k < \infty$. The logarithm process $L$ has the Karhunen-Loève expansion

$$
L(t) = \sum_{k=1}^{\infty} \xi_k \phi_k(t),
$$
where

$$\xi_k = \int_{\mathcal{T}} L(t)^T \phi_k(t) dt$$

are uncorrelated random variables such that \( \mathbb{E}\xi_k = 0 \) and \( \mathbb{E}\xi_k^2 = \lambda_k \).

A finite-truncated representation of \( X \) intrinsic to the manifold is then given by

$$X_K(t) := \text{Exp}_{\mu(t)} L_K(t), \quad L_K(t) = \sum_{k=1}^{K} \xi_k \phi_k(t)$$

for some integer \( K \geq 0 \). It was demonstrated in Dai & Müller (2018) that this representation is superior in terms of trajectory approximation for densely/completely observed manifold valued functional data compared to functional principal component analysis (FPCA), which is not adapted to the intrinsic manifold curvature, and for the same reason the scores \( \xi_k \) are better predictors for classification tasks when compared to traditional FPCs.

Suppose \( X_1, \ldots, X_n \) are i.i.d. realizations of a \( \mathcal{M} \)-valued Riemannian random process \( X \). To reflect the situation in longitudinal studies, we assume that each \( X_i \) is only recorded at \( m_i \) random time points \( T_{i,1}, \ldots, T_{i,m_i} \in \mathcal{T} \), and each observation \( X_i(T_{ij}) \) is furthermore corrupted by some intrinsic random noise. More specifically, we observe \( \mathbb{D}_n = \{(T_{ij}, Y_{ij}) : i = 1, 2, \ldots, n, j = 1, 2, \ldots, m_i \} \) such that \( T_{ij} \overset{i.i.d.}{\sim} f \) for some density \( f \) supported on \( \mathcal{T} \), and the \( T_{ij} \) are independent of the \( X_i \). Furthermore, conditional on \( X_i \) and \( T_{i,1}, \ldots, T_{i,m_i} \), the noisy observations \( Y_{ij} = \text{Exp}_{\mu(T_{ij})} \{L_i(T_{ij}) + \varepsilon_i(T_{ij})\} \) are independent, where \( \varepsilon_i(T_{ij}) \in T_{\mu(T_{ij})} \mathcal{M} \) is independent of \( X_i \), with isotropic variance \( \sigma^2 \) and \( \mathbb{E}\{\varepsilon_i(T_{ij}) \mid T_{ij}\} \equiv 0 \). As \( \mathbb{E}\{L_i(T_{ij}) \mid T_{ij}\} \equiv 0 \), the assumption on \( \varepsilon \) implies that \( \mathbb{E}\{\text{Log}_{\mu(T_{ij})} Y_{ij} \mid T_{ij}\} \equiv 0 \).

2.2. Estimation

For the case of sparse functional or longitudinal data that are the focus of this paper, it is not possible to estimate the mean curve using the cross-sectional approach of Dai & Müller (2018), as repeated observations at the same time \( t \) are not available. Instead we develop a new method, for which we harness Fréchet regression (Petersen & Müller, 2018). Fréchet
regression was developed for independent measurements, and for our purposes we need to study an extension that is valid for the case of repeated measurements.

For any $t \in T$ and $K_{h_{\mu}}$, where the kernel $K(\cdot)$ is a symmetric density function and $h_{\mu} > 0$ is a sequence of bandwidths, with $K_{h_{\mu}}(x) = h_{\mu}^{-1}K(x/h_{\mu})$, we define the local weight function at $t$ by

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

where $\hat{u}_k(t) = \sum_{i=1}^n w_i \sum_{j=1}^{m_i} K_{h_{\mu}}(T_{ij} - t)(T_{ij} - t)^k$ for $k = 0, 1, 2$, and $\hat{\sigma}_0^2(t) = \hat{u}_0(t)\hat{u}_2(t) - \hat{u}_1^2(t)$. Defining the double-weighted Fréchet function

$$
Q_n(y, t) = \sum_{i=1}^n w_i \sum_{j=1}^{m_i} \hat{\omega}(T_{ij}, t, h)d_M^2(Y_{ij}, y),
$$

which includes weights $w_i$ for individual subjects satisfying $\sum_{i=1}^n m_i w_i = 1$, we estimate the mean trajectory $\mu(t)$ by

$$
\hat{\mu}(t) = \arg \min_{y \in M} Q_n(y, t).
$$

Note that for the Euclidean special case, where $M = \mathbb{R}^D$, $Q_n$ coincides with the loss function used in Zhang & Wang (2016) for linear functional data.

For the choice of the weights $w_i$, two options have been studied in the Euclidean special case. One is to assign equal weight to each observation, i.e., $w_i = 1/(n\bar{m})$ with $\bar{m} = n^{-1} \sum_{i=1}^n m_i$, used in Yao et al. (2005). The other is to assign equal weight to each subject, i.e., $w_i = 1/(nm_i)$, as proposed in Li & Hsing (2010). We refer to the former scheme as “OBS” and to the latter as “SUBJ”, following Zhang & Wang (2016), who found that the OBS scheme is generally preferrable for non-dense functional data; the SUBJ scheme performs better for ultra-dense data; and an intermediate weighting scheme that is in between OBS and SUBJ performs at least as well as the OBS and SUBJ schemes in the Euclidean case. The latter corresponds to the choice $w_i = \alpha/(n\bar{m}) + (1 - \alpha)/(nm_i)$ for a constant $\alpha = c_2/(c_1+c_2)$ with $c_1 = 1/(\bar{m}h_{\mu}) + \bar{m}_2/\bar{m}^2$ and $c_2 = 1/(\bar{m}_H h_{\mu}) + 1$, where $\bar{m}_2 = n^{-1} \sum_{i=1}^n m_i^2$ and $\bar{m}_H = n/ \sum_{i=1}^n m_i^{-1}$, and we refer to this choice as INTM.
To estimate the covariance structure, we first map the original data into tangent spaces, setting \( \hat{L}_{ij} = \log_{\tilde{\mu}(T_{ij})} Y_{ij} \) and treating \( \hat{L}_{ij} \) as a column vector in \( \mathbb{R}^D \). To smooth \( D \times D \) matrices \( \Gamma_{ijl} = \hat{L}_{ij} \hat{L}_{il}^T \) for \( j \neq l \), we extend the scatterplot smoother (Yao et al., 2005) to matrix-valued data by finding minimizing \( D \times D \) matrices \( \hat{A}_0, \hat{A}_1 \) and \( \hat{A}_2 \) according to

\[
(\hat{A}_0, \hat{A}_1, \hat{A}_2) = \arg \min_{A_0, A_1, A_2} \sum_{i=1}^{n} v_i \sum_{1 \leq j \neq l \leq m_i} \|\Gamma_{ijl} - A_0 - (T_{ij} - s)A_1 - (T_{il} - t)A_2\|_F^2 K_{h_F}(T_{ij} - s)K_{h_F}(T_{il} - t),
\]

where in the above weighted least squares error minimization step \( \|\cdot\|_F \) is the matrix Frobenius norm, \( h_F > 0 \) is a bandwidth, and \( v_i \) are weights with \( \sum_{i=1}^{n} m_i(m_i - 1)v_i = 1 \).

For the OBS weight scheme, \( v_i = 1/\sum_{i=1}^{n} m_i(m_i - 1) \), for the SUBJ scheme, \( v_i = 1/[nm_i(m_i - 1)] \), while for INTM, \( v_i = \alpha/\sum_{i=1}^{n} m_i(m_i - 1) + (1 - \alpha)/[nm_i(m_i - 1)] \) for a constant \( \alpha = c_2/(c_1 + c_2) \) with \( c_1 = 1/(\bar{m}_2 h_F^2) + \bar{m}_3/(\bar{m}_2^2 h_F) + \bar{m}_4/\bar{m}_2^2 \) and \( c_2 = 1/(\bar{m}_Q h_F^2) + 1/(\bar{m}_H h_F) + 1 \), where \( \bar{m}_k = n^{-1} \sum_{i=1}^{n} m_i^k \) and \( \bar{m}_Q = n/\sum_{i=1}^{n} m_i^{-2} \), in analogy to Zhang & Wang (2016). We then use \( \hat{A}_0 \) as obtained in (5) as an estimate of the population covariance function \( \Gamma(s,t) \). For \( s = t \), the minimization is over symmetric matrices \( A_1, A_2 \) and symmetric semi-positive definite matrices \( A \). Estimates for the eigenfunctions \( \phi_k \) and \( \lambda_k \) of \( \Gamma \) are then obtained by the corresponding eigenfunctions \( \hat{\phi}_k \) and eigenvalues \( \hat{\lambda}_k \) of \( \hat{\Gamma} \).

In applications, one needs to choose appropriate bandwidths \( h_\mu \) and \( h_F \), as well as the number of included components \( K \). To select \( h_\mu \) for smoothing the mean function \( \mu \), we adopt a generalized cross-validation (GCV) criterion

\[
\text{GCV}(h) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m_i} d_M^2(\hat{\mu}(T_{ij}), Y_{ij})}{(1 - K_{h}(0)/N)^2},
\]

where \( N = \sum_{i=1}^{n} m_i \) is the total number of observations, and then choose \( h_\mu \) as the minimizer of \( \text{GCV}(h) \). While a similar GCV strategy can be adopted to select the bandwidth for the covariance function \( \Gamma \), we propose to employ the simpler choice \( h_F = 2h_\mu \), which we found to perform well numerically and which is computationally efficient.
To determine the number of components $K$ included in the finite-truncated representation (4), it is sensible to consider the fraction of variation explained (FVE)

$$FVE(K) = \frac{\sum_{k=1}^{K} \lambda_k}{\sum_{j=1}^{\infty} \lambda_j}, \quad \hat{FVE}(K) = \frac{\sum_{k=1}^{K} \hat{\lambda}_k}{\sum_{j=1}^{\infty} \hat{\lambda}_j},$$

choosing the number of included components as the smallest $K$ such that the FVE exceeds a specified threshold $0 < \gamma < 1$,

$$K^* = \min\{K : FVE(K) \geq \gamma\}, \quad \hat{K}^* = \min\{K : \hat{FVE}(K) \geq \gamma\},$$

where common choices of $\gamma$ are 0.90, 0.95, and 0.99.

### 2.3. Riemannian Functional Principal Component Analysis Through Conditional Expectation

The unobserved scores $\xi_{ik}$ need to be estimated from the discrete samples $\{(T_{ij}, X_{ij})\}_{j=1}^{m_i}$ or log-mapped samples $\{(T_{ij}, L_{ij})\}_{j=1}^{m_i}$. Approximating (3) by numerical integration is not feasible when the number of repeated measurements per curve is small, in analogy to the Euclidean case (Yao et al., 2005; Kraus, 2015). We therefore propose Riemannian Functional Principal Component Analysis Through Conditional Expectation (RPACE), generalizing the PACE procedure of Yao et al. (2005) for tangent-vector valued processes, where we apply best linear unbiased predictors (BLUP) to estimate the $\xi_{ik}$, obtaining the RFPC scores

$$\tilde{\xi}_{ik} = \mathcal{B}[\xi_{ik} | L_i] = \lambda_k \phi_{ik}^T \Sigma_{L_i}^{-1} L_i.$$

Here $\mathcal{B}$ denotes the best linear unbiased predictor. Writing $\text{Vec}(.\cdot)$ for the vectorization operation, $L_i = \text{Vec}([L_{i1}, \ldots, L_{im_i}])$ are the vectorized concatenated log-mapped observations for subject $i$, $\tilde{L}_i = \text{Vec}([L_i(T_{i1}), \ldots, L_i(T_{im_i})])$, $\phi_{ik} = \text{Vec}([\phi_k(T_{i1}), \ldots, \phi_k(T_{im_i})])$, and $\Sigma_{L_i} = \mathbb{E}(L_i L_i^T) = \mathbb{E}(\tilde{L}_i \tilde{L}_i^T) + \sigma^2 I$, where $I$ is the identity matrix. The entry of $\mathbb{E}(\tilde{L}_i \tilde{L}_i^T)$
corresponding to \( E([L_i(T_{ij})]_l)[L_i(T_{il})]_m) \) is \([\Gamma(T_{ij}, T_{ik})]_{lm} \), where \([v]_a \) and \([A]_{ab} \) denote the \( a \)th or \((a, b)\)th entry in a vector \( v \) or matrix \( A \), respectively. Substituting corresponding estimates for the unknown quantities in (8), we obtain plug-in estimates for \( \xi_{ik} \),

\[
\hat{\xi}_{ik} = \hat{\lambda}_k \hat{\phi}_{ik}^T \hat{\Sigma}_{L_i}^{-1} \hat{L}_i, \tag{9}
\]

where \( \hat{\Sigma}_{L_i} = \hat{E}(\hat{L}_i \hat{L}_i^T) + \hat{\sigma}^2 I; \hat{E}(\hat{L}_i \hat{L}_i^T), \hat{\lambda}_k, \) and \( \hat{\phi}_{ik} \) are obtained from \( \hat{\Gamma} \), the minimizer of (5), and we define \( \hat{\sigma}^2 = \sum_{i=1}^n \sum_{j=1}^{m_i} (ndm_i)^{-1} \text{Tr}(L_{ij}L_{ij}^T - \hat{\Gamma}(T_{ij}, T_{ij})) \), where \( \text{Tr}(A) \) denotes the trace of a matrix \( A \). The \( K \)-truncated processes

\[
L_{iK}(t) = \sum_{k=1}^K \xi_{ik} \phi_k(t), \quad X_{iK}(t) = \sum_{k=1}^K \text{Exp}_\mu(t)(L_{iK}(t)) \tag{10}
\]

are estimated by

\[
\hat{L}_{iK}(t) = \sum_{k=1}^K \hat{\xi}_{ik} \hat{\phi}_k(t), \quad \hat{X}_{iK}(t) = \sum_{k=1}^K \text{Exp}_{\hat{\mu}(t)}(\hat{L}_{iK}(t)). \tag{11}
\]

The BLUP estimate \( \tilde{\xi}_{ik} \) coincides with the conditional expectation \( E[\xi_{ik} \mid L_i] \), or the best prediction of \( \xi_{ik} \) given observation \( L_i \), if the joint distribution of \( (\xi_{ik}, L_i) \) is elliptically contoured (Fang et al., 1990, Theorem 2.18), with the Gaussian distribution as the most prominent example.

### 3. ASYMPTOTIC PROPERTIES

To derive the asymptotic properties of the estimates in Section 2, in addition to conditions (X0) and (X1), we require the following assumptions.

(M0) The domain \( \mathcal{T} \) is compact and the manifold \( \mathcal{M} \) is a bounded submanifold of \( \mathbb{R}^D \).

(K0) The kernel function \( K \) is a Lipschitz continuous symmetric probability density function on \([-1, 1]\).
Almost surely, the sample paths $X(\cdot)$ are twice continuously differentiable.

Note that the boundedness assumption on the manifold can be relaxed by imposing additional conditions on the random process $X$, or by requiring a compact support for $X(t)$, $t \in T$. The assumptions on the manifold are satisfied for our data applications in Section 4 where the manifolds under consideration are spheres.

To state the next assumption, we define the following quantities. Let $\omega(s, t, h) = \frac{1}{\sigma_0} K_h(s - t)\{u_2(t) - u_1(t)(s - t)\}$, where $u_k(t) = \mathbb{E}\{K_h(T - t)(T - t)^k\}$, $k = 0, 1, 2$, and $\sigma_0^2(t) = u_0(t)u_2(t) - u_1^2(t) > 0$ for all $t$ by the Cauchy–Schwarz inequality. Note that the finiteness of $u_k$ is implied by the Lipschitz continuity of the kernel function $K$ and the compactness of the domain $T$. Define $\tilde{Q}_h(p, t) = \mathbb{E}\{\omega(T, t, h)d^2_M(Y, p)\}$ and $\tilde{\mu}(t) = \arg\min_{y \in M}\tilde{Q}_h(y, t)$.

(L0) The Fréchet mean functions $\mu$, $\tilde{\mu}$, and $\hat{\mu}$ exist and are unique, the latter almost surely for all $n$.

(L1) The density $f(t)$ of the random times $T$ when measurements are made is positive and twice continuously differentiable for $t \in T$.

Recall that $T_p\mathcal{M}$ denotes the tangent space at $p \in \mathcal{M}$ and $\text{Exp}_p$ is the Riemannian exponential map at $p$, which maps a tangent vector $v \in T_p\mathcal{M}$ onto the manifold $\mathcal{M}$. For $p \in \mathcal{M}$, define a real-valued function $G_p(v, t) = M(\text{Exp}_p v, t)$, $v \in T_p\mathcal{M}$ and $t \in T$, where $M(p, t) = \mathbb{E}d^2_M(p, X(t))$ is the Frechét variance function defined in Section 2.1. We assume

(L2) The Hessian of $G_p(\cdot, t)$ at $v = 0$ is uniformly positive definite along the mean function, i.e.,

$$\inf_{t \in T} \lambda_{\min}\left(\frac{\partial^2}{\partial v^2} G_p(t)(v, t) \mid v = 0\right) > 0.$$ 

Conditions (L0) is necessary to ensure a consistent estimate of the mean curve using $M$-estimation theory, while (L1) is a design density condition; both are standard in the literature (Zhang & Wang, 2016; Petersen & Müller, 2018). On a Riemannian manifold $\mathcal{M}$ with sectional curvature at most $K$, (L0) and (L2) are satisfied if the support of $X(t)$ is
within $B(\mu(t), \pi/(2K))$, where $B(p, r)$ is a geodesic ball with center $p \subset M$ and radius $r$ (Bhattacharya & Bhattacharya, 2012); this specifically holds for longitudinal compositional data mapped to the positive orthant of a unit sphere. The next two conditions impose certain convergence rates for $h_\mu$ and $h_\Gamma$, respectively. For simplicity, we shall assume $m_i \equiv m$, noting that results paralleling those in Zhang & Wang (2016) can be obtained for the general case.

(H1) \quad h_\mu \to 0 \text{ and } (\log n)/(nmh_\mu) \to 0.

(H2) \quad h_\Gamma \to 0, \ (\log n)/(nm^2h_\Gamma^2) \to 0, \text{ and } (\log n)/(nmh_\Gamma) \to 0.

The following result establishes the uniform convergence rate for estimates $\hat{\mu}$.

**Theorem 1.** Assume conditions $(X0)$–$(X2)$, $(M0)$, $(K0)$, $(L0)$–$(L2)$ and $(H1)$ hold. Then

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

(12)

This result shows that the estimate $\hat{\mu}$ enjoys the same rate as the one obtained in Zhang & Wang (2016) for the Euclidean case, even in the presence of curvature. The rate in (12) has three terms that correspond to three regimes that are characterized by the growth rate of $m$ relative to the sample size: (1) When $m \ll (n/\log n)^{1/4}$, the observations per curve are sparse, and the optimal choice $h_\mu \asymp (nm/\log n)^{-1/5}$ yields $\sup_{t \in T} d_M(\hat{\mu}(t), \mu(t)) = O_P \left( (nm/\log n)^{-2/5} \right)$; (2) When $m \asymp (n/\log n)^{1/4}$, corresponding to an intermediate case, the optimal choice $h_\mu \asymp (n/\log n)^{-1/4}$ leads to the uniform rate $O_P \left( \{(\log n)/n\}^{1/2} \right)$ for $\hat{\mu}$; (3) When $m \gg (n/\log n)^{1/4}$, the observations are dense, and any choice $h_\mu = o \left( (n/\log n)^{-1/4} \right)$ gives rise to the uniform rate $O_P \left( \{(\log n)/n\}^{1/2} \right)$. The transition from (1) to (3) is akin to a phase transition, similar to the one observed in Hall et al. (2006).

The next result concerns the uniform rate for the estimator $\hat{\Gamma}$ of $\Gamma$, the covariance function of the log-mapped data, extending a result of Zhang & Wang (2016) for the Euclidean case to curved functional data.
Theorem 2. Assume conditions (X0)–(X2), (M0), (K0), (L0)–(L2), (H1) and (H2) hold. Then

\[
\sup_{s,t \in T} \|\hat{\Gamma}(s,t) - \Gamma(s,t)\|_F^2 = O_P \left( h_\mu^4 + h_\Gamma^4 + \frac{\log n}{nmh_\mu} + \frac{\log n}{nm^2h_\Gamma} + \frac{\log n}{nmh_\Gamma} \right). \tag{13}
\]

Again, the above rate gives rise to three regimes that are determined by the growth rate of \( m \) relative to the sample size: (1) When \( m \ll (n/\log n)^{1/4} \), the observations per curve are sparse, and with the optimal choice \( h_\mu \asymp (nm/\log n)^{-1/5} \) and \( h_\Gamma \asymp (nm^2/\log n)^{-1/6} \), one has \( \sup_{s,t \in T} \|\hat{\Gamma}(s,t) - \Gamma(s,t)\|_F = O_P \left( (nm^2/\log n)^{-1/3} \right) \); (2) When \( m \asymp (n/\log n)^{1/4} \), with the optimal choice \( h_\mu \asymp h_\Gamma \asymp (n/\log n)^{-1/4} \), the uniform rate for \( \hat{\Gamma} \) is \( O_P \left( \{(\log n)/n\}^{1/2} \right) \); (3) When \( m \gg (n/\log n)^{1/4} \), the observations are dense, and any choice \( h_\mu, h_\Gamma = o \left( (n/\log n)^{-1/4} \right) \) yields the uniform rate \( O_P \left( \{(\log n)/n\}^{1/2} \right) \).

Furthermore, according to Lemma 4.2 of Bosq (2000), one has \( \sup_k |\hat{\lambda}_k - \lambda_k| \leq \|\hat{\Gamma} - \Gamma\|_{HS} \). It can also be shown that \( \|\hat{\Gamma} - \Gamma\|_{HS} \leq |T| \sup_{s,t \in T} \|\hat{\Gamma}(s,t) - \Gamma(s,t)\|_F \), where \(|T|\) denotes the Lebesgue measure of \( T \). Therefore, the rate for \( \hat{\Gamma} \) provides a convergence rate for all estimated eigenvalues \( \hat{\lambda}_k \). Furthermore, according to Lemma 4.3 of Bosq (2000), if \( \lambda_{k-1} \neq \lambda_k \) and \( \lambda_k \neq \lambda_{k+1} \), then \( \|\hat{\phi}_k - \phi_k\|^2 \leq c_k \|\hat{\Gamma} - \Gamma\|^2_{HS} \), where \( c_1 = 8(\lambda_1 - \lambda_2)^{-2} \) and \( c_k = 8 \max \{ (\lambda_{k-1} - \lambda_k)^{-2}, (\lambda_k - \lambda_{k+1})^{-2} \} \) for \( k \geq 2 \). Again, by utilizing the fact that \( \|\hat{\Gamma} - \Gamma\|_{HS} \leq |T| \sup_{s,t \in T} \|\hat{\Gamma}(s,t) - \Gamma(s,t)\|_F \), we can derive the convergence rate for \( \hat{\phi}_k \). For example, if we assume polynomial decay of eigenvalue spacing, i.e., \( a_1 k^{-\beta} \leq \lambda_k - \lambda_{k+1} \leq a_2 k^{-\beta} \) for some constants \( a_2 \geq a_1 > 0 \) and \( \beta > 1 \), then \( \|\hat{\phi}_k - \phi_k\|^2 = O_P(k^{2\beta} \gamma_n) \) where \( \gamma_n \) is the rate that appears on the right hand side of (13), and the \( O_P \) term is uniform for all \( k \).

4. DATA APPLICATIONS

4.1. Emotional Well-Being for Unemployed Workers

We demonstrate RPACE for the analysis of longitudinal mood compositional data. These data were collected in the Survey of Unemployed Workers in New Jersey (Krueger & Mueller,
2011), conducted in the fall of 2009 and the beginning of 2010, during which the unemployment rate in the US peaked at 10% after the financial crisis of 2007–2008. A stratified random sample of unemployed workers were surveyed weekly for up to 12 weeks. Questionnaires included an entry survey, which assessed demographics, household characteristics and income, and weekly followups, including job search activities and emotional well-being. In each followup questionnaire, participants were asked to report the percentage of time they spent in each of the four moods: bad, low/irritable, mildly pleasant, and good. The overall weekly response rate was around 40%; see Krueger & Mueller (2011).

We analyzed a sample of \( n = 4771 \) unemployed workers enrolled in the study, who were not offered a job during the survey period. The measurement of interest \( Y(t) = \left[ Y_1(t), \ldots, Y_4(t) \right] \) is the longitudinal mood composition, where \( Y_j(t) \) is the proportion of time a subject spent in the \( j \)th mood in the previous 7 days, \( j = 1, \ldots, 4 \), recorded on day \( t \in [0, 84] \) since the start of the study. The number of responses per subject ranged from 1 to 12, so the data is a mixture of very sparse and mildly sparse longitudinal observations; for 25% of all subjects only one response was recorded. As subjects responded at different days of the week, the observation time points were also irregular. The sparsity and irregularity of the observations poses difficulties for classical analyses and prevents the application of the presmooth-and-then-analyze method (Dai & Müller, 2018), motivating the application of RPACE, which is geared towards such sparse and irregularly sampled manifold-valued functional data.

We applied RPACE for the square-root transformed compositional data

\[
X(t) = \{X_1(t), \ldots, X_4(t)\} = \{\sqrt{Y_1(t)}, \ldots, \sqrt{Y_4(t)}\},
\]

which lie on the sphere \( S^3 \) for \( t \in [0, 84] \), since compositional data are non-negative and sum to 1, using bandwidths \( h_\mu = 18 \) and \( h_\Gamma = 36 \) days, as selected by GCV, and the Epanechnikov kernel \( K(x) = 0.75(1 - x^2) \) on \([-1, 1]\). The mood composition trajectories for four randomly selected subjects are displayed in the left panel of Figure 1. The solid dots
Figure 1: Left: Longitudinal mood compositional data for four randomly selected unemployed workers, with raw observations shown as dots and fitted trajectories by RPACE using 8 eigen-components shown as solid curves. Overlapping dots were slightly jittered vertically. Right: The overall mean function.

denote the reported moods, which are slightly jittered vertically if they overlap, and dashed curves denote the fitted trajectories when selecting $K = 8$ components, selected according to the FVE criterion (7) with threshold $\gamma = 0.99$, which is a reasonable choice in view of the large sample size. A substantial proportion of the mood compositions is zero, which is no problem for the square-root transformation approach in contrast to the alternative log-ratio transformation (Aitchison, 1986), which is undefined when the baseline category is 0.

As the self-reported moods contain substantial aberrations from smooth trajectories that we view as noise, the fitted trajectories do not go through the raw observations, and are drawn towards the observations for subjects with more repeated measurements. The mean trajectory is displayed in the right panel of Figure 1, indicating that the emotional well-being of subjects tends to deteriorate as the period of unemployment lengthens, with an overall increase in the proportion of bad mood and a decrease in the proportion of good mood.

The first four eigenfunctions for mood composition trajectories are shown in Figure 2, where the first eigenfunction corresponds to the overall contrast between neutral-to-positive mood (good and mild) and negative moods (low and bad); the second eigenfunction rep-
Figure 2: The first four eigenfunctions for the mood composition data, with Fraction of Variation Explained (FVE) displayed in the panel subtitles.

represents emotional stability, which is a contrast between more neutral moods and extreme emotions (good and bad); the third eigenfunction corresponds to a shift of mood compositions to more positive moods, namely from bad to low and from mild to good; the fourth eigenfunction encodes an increase of positive feelings and a decrease of negative ones over time. Here it is important to note that the sign of the eigenfunctions is arbitrary and could be reversed. The first four eigenfunctions together explain 95% of the total variation.

As an example to demonstrate that the scores obtained from RFPC are useful for downstream tasks such as regression, we explored the association between the second RFPC score, corresponding to the proportion of extreme moods, and annual household income in 2008, a measure of financial stability. We extracted the RFPC scores for each subject and constructed kernel density estimates for $\xi_2$ within each income category; see Figure 3. Participants with higher household income before losing their job and thus higher financial stability tend to have higher emotion stability, as demonstrated by the right-shifted distributions of $\xi_2$ and larger means (colored dots). The relationship between prior income and emotional stability appears to be nonlinear especially for the lower income groups.
Figure 3: The distributions of the second RFPC score, encoding emotion stability, visualized as densities in dependence on the annual household income in 2008. Colored dots indicate the mean of $\xi_2$ for each income group.

4.2. Wallaby Body Shape Growth

Quantifying the shapes of organisms has been a long-standing statistical and mathematical problem (Thompson, 1942; Kendall et al., 2009). We apply RPACE to analyze the longitudinal development of body shapes of a sample of Tammar wallabies (*Macropus eugenii*), a small macropod native to Australia (data courtesy of Dr Jeff Wood, CSIRO Biometrics Unit INRE, Canberra, and data cleaning and corrections were performed by Professor Heike Hofmann, Department of Statistics, Iowa State University, in 2008). For each of $n = 40$ measured wallabies from two locations, longitudinal measurements of the length (in inches) of six body parts $Y_{ij} = (\text{Head, Ear, Arm, Leg, Foot, Tail})_{ij}$ were available at age $T_{ij}$ in the first 380 days after birth, for $i = 1, \ldots, 40$ and $j = 1, \ldots, n_i$. The measurement time points for the wallabies were highly irregular, and the number of measurements per wallaby varied from 1 to 26, with 14 wallabies having no more than 7 measurements. Typical measurement patterns with mixed sparse and dense observations for each curve are shown in the left panel of Figure 4. This measurement scheme requires methodology that can handle the high degree of irregularity in the measurement times.
Figure 4: Left: Longitudinal shape observations of four randomly selected wallabies, where dots denote raw observations, and solid curves indicates fitted trajectories by 7 components that together explain 90% of total variation. Right: The mean function of all trajectories.

To quantify shapes of wallabies, we normalized the length measurements $Y_{ij}$ by the Euclidean norm, obtaining $X_{ij} = Y_{ij}/\|Y_{ij}\|_2$, thus emphasizing the relative size of each body part expressed as a percentage of total size, leading to longitudinal preshape data (Kendall et al., 2009) that lie on a sphere. The $X_{ij}$ are shape characteristics of wallabies at their respective age. RPACE was then applied to the transformed data $(T_{ij}, X_{ij})$ with bandwidths $h_{\mu} = 18.3$ and $h_{\Gamma} = 36.6$ selected by GCV, using the Epanechnikov kernel. The Fréchet mean trajectory as displayed in the right panel of Figure 4 shows that relative to the body size, the tail becomes larger, while head, arm and ear lengths become relatively smaller throughout the first year of birth; leg and foot lengths increase from birth to roughly 6 months, where relative leg length development peaks before that of foot development.

To decompose the variation of individual shape trajectories, the first three eigenfunctions are displayed in Figure 5. The first eigenfunction corresponds to an overall contrast between tail and other body part development, and the second eigenfunction has a large component in the initial tail growth. The first two eigenfunctions together explain 64% of total variation, showing that tail length is a main driving force for shape differences. Pairwise scatterplots of the first three RFPC scores are shown in Figure 6, where each point stands for a single wallaby, and their patterns indicate two different geographic locations. Shape development
differences between locations were mostly reflected in the second component, corresponding to initial tail growth, while the first and third components were less dissimilar.

5. SIMULATION STUDIES

We demonstrate the performance of the sparse RPACE method for scenarios with varying sample size, sparsity, and manifolds, for which we choose $\mathcal{M} = S^2$ or $\text{SO}(3)$. Here $S^2$ is the 2-sphere and $\text{SO}(3)$ is the manifold consisting of the $3 \times 3$ orthogonal matrices with determinant 1. For each random trajectory $X_i(t)$ on $\mathcal{M}$, $i = 1, \ldots, n$, we sample $m_i$ observations $(T_{ij}, X_{ij})$, $j = 1, \ldots, m_i$, where $T_{ij}$ follows a uniform distribution on $\mathcal{T} = [0, 1]$. The number of observations $m_i$ follows a discrete uniform distribution on $\{1, \ldots, m_{\text{max}}\}$, where $m_{\text{max}}$ is the maximum number of observations per curve that differs between scenarios.

The sparse observations were generated according to $X_{ij} = \text{Exp}_{\mu(T_{ij})}(L_i(T_{ij}) + \epsilon_{ij}), L_i(t) = \sum_{k=1}^{20} \xi_{ik} \phi_k(t)$, with manifold-specific mean function $\mu(t)$ and eigenfunctions $\phi_j(t)$; RFPC scores $\xi_{ik}$ that follow independent Gaussian distributions with mean zero and variance $\lambda_k = 0.05^k/3$, for $k = 1, \ldots, 20$; and independent Gaussian errors $\epsilon_{ij}$ with mean 0 and isotropic variance $\sigma^2 = 0.01$ on the tangent space $T_{\mu(T_{ij})}$. The cumulative FVE for the first $K = 1, \ldots, 6$ components, defined as $\sum_{j=1}^{K} \lambda_j / \sum_{k=1}^{\infty} \lambda_k$, are 63.2%, 86.4%, 95.0%, 98.15%, 99.3%, and 99.8%, respectively. For $\mathcal{M} = S^2$, we set $\mu(t) = \text{Exp}_p(\nu(t))$ where $p = [0, 0, 1]$ and $\nu(t) = [2t/2^{1/2}, 0.3\pi \sin(\pi t), 0]$; eigenfunctions $2^{-1/2}R_t[\zeta_k(t/2), \zeta_k((t + 1)/2), 0]^T$, with $R_t$ being the rotation matrix from $p$ to $\mu(t)$, and $\{\zeta_k\}_{k=1}^{20}$ the orthonormal cosine basis on $[0, 1]$. For $\mathcal{M} = \text{SO}(3)$, $\mu(t) = \text{expm}(\iota(2t, 0.3\pi \sin(\pi t), 0))$ and $\phi_k(t) = 3^{-1/2}l(\zeta_k(t/3), \zeta_k((t + 1)/3), \zeta_k((t + 2)/3))$, where $\text{expm}$ is the matrix exponential and $\iota : \mathbb{R}^3 \to \mathbb{R}^{3 \times 3}$ maps a vector $v$ to a skew-symmetric matrix whose lower diagonal elements (ordered by column) are $v$.

We investigated three settings with varying sparsity and sample size: Scenario 1 (baseline): $n = 100, m_{\text{max}} = 20$; Scenario 2 (sparse): $n = 100, m_{\text{max}} = 5$; Scenario 3 (small $n$): $n = 50, m_{\text{max}} = 20$.

Three different FPCA approaches were evaluated for these scenarios, namely an extrinsic
Figure 5: The first four eigenfunctions for the wallaby shapes, with Fraction of Variation Explained (FVE) displayed in the panel subtitles.

Figure 6: Pairwise scatter plots of the first three RFPC scores for the Wallaby data, where different point patterns represent Wallabies from two distinct geographic locations.
FPCA by Chiou et al. (2014), an extrinsic multivariate FPCA via componentwise FPCA (CFPCA) by Happ & Greven (2018), and the proposed RPACE. The extrinsic FPCA is a multivariate FPCA applied to the sparse manifold-valued data as if they are objects in the ambient Euclidean space. In the CFPCA approach one first fits an FPCA to each of the $D$ components functions, and then applies a second PCA to the pooled component scores to obtain summarized scores and multidimensional eigenfunctions. For sample trajectories with small variation around the mean, the extrinsic FPCA methods (FPCA and CFPCA) can be regarded as linear approximations to RPACE. The Epanechnikov kernel was used for the smoothers, with bandwidth $h_\mu$ selected by GCV and $h_\Gamma = 2h_\mu$.

Using 200 Monte Carlo experiments, we report the average Root Mean Integrated Squared Errors (RMISE) for the fitted trajectories, defined as

$$
\text{RMISE} = \sqrt{\frac{1}{200} \sum_{b=1}^{200} \frac{1}{D} \int_0^1 d_M(\hat{X}_K(t), X(t))^2 dt},
$$

for $K = 1, \ldots, 6$ in Table 1. Since the fitted trajectories using FPCA and CFPCA lie in the ambient space but not on $M$, we projected them back to the manifold by normalizing the norm of $\hat{X}_K(t)$ for $M = S^2$ or the eigenvalues of the matrix representation of $\hat{X}_K(t)$ for $M = \text{SO}(3)$. RPACE was the overall best performer across the various scenarios, especially for the more parsimonious models. Scenario 2 (sparse) is considerably more difficult than Scenario 1, and smaller models with $K \leq 4$ performed better. RPACE also works well for the smaller sample size $n = 50$ in Scenario 3.

**APPENDIX**

**Proofs of Main Results**

**Proof of Theorem 1.** According to Lemma 3, where all auxiliary results are given in the next section,

$$
\sup_{p \in M} \sup_{t \in T} |\hat{Q}_{h_\mu}(p, t) - M(p, t)| = O(h_\mu^2).
$$

(14)
Table 1: RIMSE for fitted trajectories. The standard errors were smaller than $3 \times 10^{-3}$ for all cases: $X_E$, extrinsic Functional Principal Component Analysis (Chiou et al., 2014); $X_C$, componentwise Functional Principal Component Analysis (Happ & Greven, 2018); $X_R$, proposed Riemannian Principal Component Analysis by Conditional Expectation (RPACE).

| Scenario 1 (baseline) | Scenario 2 (sparse) | Scenario 3 (small n) |
|-----------------------|----------------------|-----------------------|
| $\mathcal{M} = S^2$  | $\mathcal{M} = S^2$  | $\mathcal{M} = S^2$  |
| $K$                   | $X_E$                | $X_E$                |
| 1                     | 0.23                 | 0.23                 |
| 2                     | 0.12                 | 0.12                 |
| 3                     | 0.08                 | 0.08                 |
| 4                     | 0.05                 | 0.05                 |
| 5                     | 0.05                 | 0.05                 |
| 6                     | 0.04                 | 0.04                 |

| $X_C$ | $X_C$ | $X_R$ | $X_C$ | $X_R$ | $X_E$ | $X_C$ | $X_R$ | $X_E$ | $X_C$ | $X_R$ | $X_E$ | $X_C$ | $X_R$ |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 0.24  | 0.24  | 0.22  | 0.26  | 0.27  | 0.24  | 0.26  | 0.27  | 0.24  | 0.23  | 0.21  | 0.26  | 0.27  | 0.24  |
| 0.12  | 0.12  | 0.09  | 0.16  | 0.17  | 0.14  | 0.15  | 0.16  | 0.12  | 0.12  | 0.09  | 0.12  | 0.12  | 0.09  |
| 0.08  | 0.08  | 0.04  | 0.13  | 0.15  | 0.11  | 0.11  | 0.13  | 0.08  | 0.08  | 0.05  | 0.08  | 0.08  | 0.04  |
| 0.05  | 0.06  | 0.04  | 0.11  | 0.14  | 0.10  | 0.09  | 0.11  | 0.07  | 0.05  | 0.06  | 0.04  | 0.05  | 0.05  |
| 0.05  | 0.05  | 0.04  | 0.10  | 0.13  | 0.10  | 0.08  | 0.10  | 0.07  | 0.05  | 0.05  | 0.04  | 0.04  | 0.02  |
| 0.04  | 0.05  | 0.04  | 0.10  | 0.13  | 0.10  | 0.08  | 0.10  | 0.07  | 0.04  | 0.05  | 0.04  | 0.03  | 0.04  |

With Lemma 4, by similar arguments as in Theorem 3 in Petersen & Müller (2018),

$$\sup_{t \in T} d^2_M(\bar{\mu}(t), \mu(t)) = O(h^4_{\mu})$$  \hspace{1cm} (15)

as $h_{\mu} \to 0$ for $h_{\mu} = O(n^{-1/2})$.

This result, combined with Lemma 6, yields (12) for $m_i \equiv m$. The proof for the general case follows the same lines. \hfill \Box

**Proof of Theorem 2.** We prove the theorem for $m_i \equiv m$, while the proof for the general case is similar. We will use $h$ to denote $h_{1\Gamma}$ throughout the proof. Observe

$$\hat{\Gamma}(s, t) = \frac{(S_{20}S_{02} - S_{11}^2)R_{00} - (S_{10}S_{02} - S_{01}S_{11})R_{10} + (S_{10}S_{11} - S_{01}S_{20})R_{01}}{(S_{20}S_{02} - S_{11}^2)S_{00} - (S_{10}S_{02} - S_{01}S_{11})S_{10} + (S_{10}S_{11} - S_{01}S_{20})S_{01}},$$

where for $a, b = 0, 1, 2$,

$$S_{ab} = \sum_{i=1}^{n} v_i \sum_{1 \leq j \neq l \neq m_i} K_h(T_{ij} - s)K_h(T_{il} - t) \left( \frac{T_{ij} - s}{h} \right)^a \left( \frac{T_{ij} - t}{h} \right)^b,$$

$$R_{ab} = \sum_{i=1}^{n} v_i \sum_{1 \leq j \neq l \neq m_i} K_h(T_{ij} - s)K_h(T_{il} - t) \left( \frac{T_{ij} - s}{h} \right)^a \left( \frac{T_{ij} - t}{h} \right)^b \Gamma_{ijl}.$$

Let $\delta_{ij} = (\text{Log}_\mu(T_{ij})Y_{ij})(\text{Log}_\mu(T_{il})Y_{il})^T$. Then
\[ R_{00} = R'_{00} + (\log \hat{\mu}(T_{ij}) Y_{ij} - \log \mu(T_{ij}) Y_{ij}) (\log \hat{\mu}(T_{il}) Y_{il})^T + \\
\log \hat{\mu}(T_{ij}) (\log \hat{\mu}(T_{il}) Y_{il} - \log \mu(T_{il}) Y_{il})^T + \\
(\log \hat{\mu}(T_{ij}) Y_{ij} - \log \mu(T_{ij}) Y_{ij}) (\log \hat{\mu}(T_{il}) Y_{il} - \log \mu(T_{il}) Y_{il})^T, \]

where

\[ R'_{ab} = \sum_{i=1}^{n} v_i \sum_{1 \leq j \neq l \neq m_i} K_h(T_{ij} - s)K_h(T_{il} - t) \left( \frac{T_{ij} - s}{h} \right)^a \left( \frac{T_{ij} - t}{h} \right)^b \delta_{ijl}. \]

Given the smoothness of \( \log p q \) with respect to \( p \) and the compactness of \( M \),

\[ \| R_{00} - R'_{00} \|_F^2 \leq c \sup_{t \in T} d_M^2(\hat{\mu}(t), \mu(t)) = O_P(h_{\mu}^4 + \frac{\log n}{n} + \frac{\log n}{nmh_{\mu}}), \]

with similar results for \( R_{ab} \) for \( a, b = 0, 1, 2. \) Setting

\[ \tilde{\Gamma}(s, t) = (S_{20}S_{02} - S_{11}^2) R'_{00} - (S_{10}S_{02} - S_{01}S_{11}) R'_{10} + (S_{10}S_{11} - S_{01}S_{20}) R'_{01} \\
(S_{20}S_{02} - S_{11}^2)S_{00} - (S_{10}S_{02} - S_{01}S_{11})S_{10} + (S_{10}S_{11} - S_{01}S_{20})S_{01}, \]

we have

\[ \| \tilde{\Gamma}(s, t) - \Gamma(s, t) \|_F^2 = O_P(h_{\mu}^4 + (\log n)(n^{-1}m^{-1}h^{-2} + n^{-1}h^{-1} + n^{-1})), \]

whence by the same argument as in Theorem 5.2 in Zhang & Wang (2016), \( \sup_{s, t \in T} \| \tilde{\Gamma}(s, t) - \Gamma(s, t) \|_F^2 = O_P \{ h^4 + (\log n)(n^{-1}m^{-1}h^{-2} + n^{-1}h^{-1} + n^{-1}) \}, \) and the result follows.

\[ \square \]

**Technical Lemmas**

**Lemma 3.** Assume conditions \((X0), (X1), (X2), (M0), (K0), (L0), (L1) \) and \((H1)\) hold. Then for any \( \epsilon > 0, \)

\[ \inf_{t \in T} \inf_{\epsilon < \delta_\cdot} \{ M(y, t) - M(\mu(t), t) \} > 0. \quad (16) \]

\[ \sup_{p \in \mathcal{M}} \sup_{t \in T} | \tilde{Q}_h(p, t) - M(p, t) | = O(h_{\mu}^2) = o(1), \quad (17) \]

\[ \sup_{t \in T} d_\mathcal{M}(\tilde{\mu}(t), \mu(t)) = o(1). \quad (18) \]
Proof. Equation (16) is implied by (L0). For (17), first obtain the auxiliary result

\[ u_k(t) = O(h_{\mu}^k), \]  

(19)

for \( k = 0, 1, 2 \), where the \( O(h_{\mu}^k) \) term is uniform over \( t \in T \) and is bounded away from 0 for \( k = 0, 2 \). This is due to change of variables and a Taylor expansion for \( f \),

\[
\mathbb{E}\{K_{h_{\mu}}(T-t)\left(\frac{T-t}{h_{\mu}}\right)^k\} = \int_{-t/h_{\mu}}^{(1-t)/h_{\mu}} s^k K(s)f(t+h_{\mu}s)ds \\
= (f(t) + O(h))\int_{\max(-1,-t/h_{\mu})}^{\min(1,(1-t)/h_{\mu})} s^k K(s)ds \\
= O(1),
\]

where the \( O(1) \) term is uniform over \( t \) and bounded away from 0 for \( k = 0, 2 \) by (L1). Then

\[
\hat{Q}_{h_{\mu}}(p,t) - M(p,t) = \mathbb{E}\{\mathbb{E}\{d_M^2(Y,p)|T\}\omega(T,t,h_{\mu})\} - M(p,t) \\
= \mathbb{E}\{M(p,T)\omega(T,t,h_{\mu})\} - M(p,t) \\
= \mathbb{E}\{\frac{\partial}{\partial t}M(p,t)(T-t)\omega(T,t,h_{\mu})\} + \mathbb{E}\{\frac{\partial^2}{\partial t^2}M(p,\vartheta)(T-t)^2\omega(T,t,h)\} \\
= \frac{1}{u_2(t)u_0(t) - u_1(t)^2}[u_2(t)\mathbb{E}\left\{\frac{\partial^2}{\partial t^2}M(p,\vartheta)K_{h_{\mu}}(T-t)(T-t)^2\right\} - u_0(t)\mathbb{E}\left\{\frac{\partial^2}{\partial t^2}M(p,\vartheta)K_{h_{\mu}}(T-t)(T-t)^3\right\}] \\
= O(h_{\mu}^2),
\]  

(20)

where \( \vartheta \) is between \( T \) and \( t \), the third equality is due to applying Taylor’s theorem on \( M(p,\cdot) \), the fourth to \( \mathbb{E}[(T-t)\omega(T,t,h_{\mu})] = 0 \), and the last to (19) and the continuity and boundedness in \( \partial^2 M/\partial t^2(p,t) \), as implied by (M0) and (X2). Note that the rate (20) is uniform over \( p \in \mathcal{M} \) and \( t \in T \), so we obtain (17).

By M-estimation theory (e.g. Corollary 3.2.3 in van der Vaart & Wellner, 1996), (17) and (16) imply (18). \( \square \)
Lemma 4. Under conditions \((M0), (K0), (L0)-(L2), (X0)-(X2)\), and \(h_\mu \to 0\), there exist constants \(C > 0\) and \(\eta > 0\) such that for all \(t \in \mathcal{T}\),

\[
M(y, t) - M(\mu(t), t) \geq Cd_M^2(y, \mu(t)),
\]

\[
\liminf_{n} \{\bar{Q}_h(y, t) - \bar{Q}_h(\bar{\mu}(t), t) - Cd_M^2(y, \bar{\mu}(t))\} \geq 0,
\]

if \(d_M(y, \mu(t)) < \eta\) and \(d_M(y, \bar{\mu}(t)) < \eta\), respectively.

Proof. Recall \(G_p(v, t) = M(\text{Exp}_p v, t)\) as defined in \((L2)\), and define \(v = \text{Log}_{\mu(t)}(y)\). For each \(t \in \mathcal{T}\), apply a Taylor expansion to obtain

\[
M(y, t) - M(\mu(t), t) = G_{\mu(t)}(v, t) - G_{\mu(t)}(0, t)
\]

\[
= \left(\frac{\partial^2}{\partial v^2} G_{\mu(t)}(v^*, t)\right) v v^*_{\mu(t)}
\]

\[
\geq \lambda_{\min} \left(\frac{\partial^2}{\partial v^2} G_{\mu(t)}(v^*, t)\right) \langle v, v \rangle_{\mu(t)}
\]

\[
= \lambda_{\min} \left(\frac{\partial^2}{\partial v^2} G_{\mu(t)}(v^*, t)\right) d_M^2(y, \mu(t))
\]

where \(v^*\) is between 0 and \(v\). There exists \(\eta > 0\) such that for \(\langle v, v \rangle^{1/2} < \eta\),

\[
\lambda_{\min} \left(\frac{\partial^2}{\partial v^2} G_{\mu(t)}(v^*, t)\right) \geq C,
\]

by \((L2)\) and the smoothness of \(G_{\mu(t)}\), where \(C = \lambda_{\min} \left(\frac{\partial^2 G_{\mu(t)}}{\partial v^2}(0, t)\right)/2\), and the inequality holds uniformly over \(t\). This then implies \((21)\).

For \((22)\), applying iterated expectations and Taylor’s theorem, we obtain

\[
\frac{\partial^2}{\partial y^2} \bar{Q}_h(y, t) - \frac{\partial^2}{\partial y^2} M(y, t) = O(h_\mu^2),
\]

where the \(O(h_\mu^2)\) term is uniform over \(y \in \mathcal{M}\) and \(t \in \mathcal{T}\), similar to the proof of the uniform consistency of \(\bar{Q}_{h_\mu}\) to \(M(p, t)\) in Lemma 3. Define \(H_p(v, t) = \bar{Q}_h(\text{Exp}_p v, t)\) and \(\bar{v} = \text{Log}_p(y)\).
With (L2) this implies
\[
\inf \liminf_{t \in T} \liminf_{n \to \infty} \lambda_{\min} \left( \frac{\partial^2}{\partial v^2} H_{\hat{\mu}(t)}(v, t) \mid v = 0 \right) = \inf \liminf_{t \in T} \liminf_{n \to \infty} \lambda_{\min} \left( \frac{\partial^2}{\partial y^2} \tilde{Q}_h(y, t) \mid y = \tilde{\mu}(t) \right) > 0, \tag{25}
\]
where the inequality is due to (24). Then
\[
\tilde{Q}_h(y, t) - \tilde{Q}_h(\tilde{\mu}(t), t) = H_{\hat{\mu}(t)}(\hat{v}, t) - H_{\hat{\mu}(t)}(0, t)
\]
\[
= \left( \frac{\partial^2}{\partial v^2} H_{\hat{\mu}(t)}(\hat{v}^*, t) \right) \langle \hat{v}, \tilde{v} \rangle_{\hat{\mu}(t)}
\]
\[
\geq \lambda_{\min} \left( \frac{\partial^2}{\partial v^2} H_{\hat{\mu}(t)}(\hat{v}^*, t) \right) d_{M}^2(y, \tilde{\mu}(t))
\]
\[
= \lambda_{\min} \left( \frac{\partial^2}{\partial v^2} H_{\hat{\mu}(t)}(\hat{v}^*, t) \right) d_{M}^2(y, \tilde{\mu}(t))
\]
By (23) and (24), the last term is not smaller than \(Cd_{M}^2(y, \tilde{\mu}(t))\) for large enough \(n\). Therefore by taking \(\liminf\) and \(\infimum\) over \(t\) we obtain (22).

\[\square\]

**Lemma 5.** Suppose \(B_\delta(p)\) is an open ball centered at \(p \in M\) with radius \(\delta > 0\), and denote the covering number of \(B_\delta(p)\) with \(\epsilon\)-balls by \(N(\epsilon, B_\delta(p), d_M)\). Then condition (M0) implies
\[
\int_0^1 \sup_{t \in T} \sqrt{1 + \log N(\delta, B_\delta(\mu(t)), d_M)} d\epsilon = O(1) \quad \text{as} \quad \delta \to 0.
\]

**Proof.** This is a consequence of Proposition 3 of Petersen & Müller (2018). \[\square\]

**Lemma 6.** Suppose \(m_i = m\) and (M0), (K0), (L0), (L1), (X0), (X1) and (X2) hold. If \(h_\mu \to 0\) and \(nmh_\mu \to \infty\), then
\[
\sup_{t \in T} d_M^2(\hat{\mu}(t), \tilde{\mu}(t)) = O_P \left( \frac{\log n}{n} + \frac{\log n}{nmh_\mu} \right).
\]
Proof. We first establish
\[
\sup_{t \in T} |\hat{u}_k(t) - u_k(t)| = O_P \left( \sqrt{\log n} \sqrt{h^{2k-1}n^{-1}m^{-1} + h^{2k}n^{-1}} \right),
\]
with proof analogous to that of Lemma 5 of Zhang & Wang (2016). Following a similar argument as in Lemma 2 in Petersen & Müller (2018), by (26), Lemma 4 and 5, one obtains
\[
\sup_{t \in T} \mathcal{d}_M^2(\hat{\mu}(t), \hat{\mu}(t)) = o_P(1).
\]

To derive the rate of convergence, we set
\[
\hat{\omega}_{ij}(t) = K_h(T_{ij} - t)\{\hat{u}_0 - \hat{u}_1(T_{ij} - t)\}/\hat{\sigma}_0^2
\]
and \(\tilde{\omega}_{ij}(t) = K_h(T_{ij} - t)\{u_0 - u_1(T_{ij} - t)\}/\sigma_0^2\). With \(S_n(y, t) = \hat{Q}_n(y, t) - \hat{Q}_h(y, t)\) and \(D_{ij}(y, t) = \mathcal{d}_M^2(Y_{ij}, y) - \mathcal{d}_M^2(Y_{ij}, \hat{\mu}(t))\),
\[
|S_n(y, t) - S_n(\hat{\mu}(t), t)| \leq \left| \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \{\hat{\omega}_{ij}(t) - \tilde{\omega}_{ij}(t)\} D_{ij}(y, t) \right|
+ \left| \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m [\hat{\omega}_{ij}(t) D_{ij}(y, t) - E\{\tilde{\omega}_{ij}(t) D_{ij}(y, t)\}] \right|
\equiv A_1(y, t) + A_2(y, t).
\]

For any \(\delta > 0\), using the boundedness of \(d_M\) and (26), one can deduce that
\[
\sup_{t \in T} \sup_{d_M(y, \hat{\mu}(t)) < \delta} A_1(y, t) = O_P(\delta \sqrt{\log n / \sqrt{nmh}}), \text{ with a universal constant for all } \delta > 0.
\]

Thus, for
\[
B_R = \left\{ \sup_{t \in T} \sup_{d_M(y, \hat{\mu}(t)) < \delta} \left| \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \{\hat{\omega}_{ij}(t) - \tilde{\omega}_{ij}(t)\} D_{ij}(y, t) \right| \leq R\delta \sqrt{\log n / \sqrt{nmh}} \right\}
\]
for some \(R > 0\), it holds that \(\Pr(B_R^c) \to 0\). For the second term, we employ a similar argument as in Lemma 5 of Zhang & Wang (2016) to show that \(E \sup_{t \in T} \sup_{d_M(y, \hat{\mu}(t)) < \delta} A_2(y, t) = \)
Thus, 

\[ E \left\{ \mathbb{I}_{B_R} \sup_{t \in T} \sup_{d_M(y, \bar{\mu}(t)) < \delta} |S_n(y, t) - S_n(\bar{\mu}(t), t)| \right\} \leq a\delta \sqrt{\log n \sqrt{1/n + 1/nm}}, \]

where \( a \) is a constant depending on \( R \). To finish, set \( r_n = (\sqrt{\log n \sqrt{1/n + 1/nm}})^{-1} \) and define \( S_{k,n}(t) = \{ y : 2^{k-1} \leq r_n d(y, \bar{\mu}(t)) \leq 2^k \} \). Let \( \eta \) be as in Lemma 4, and \( \tilde{\eta} = \eta/2 \). Then for any positive integer \( W \),

\[
\Pr \left( \sup_{t \in T} r_n d_M(\bar{\mu}(t), \tilde{\mu}(t)) > 2^W \right) \leq \Pr(B_R^c) + \Pr(2 \sup_{t \in T} d_M(\bar{\mu}(t), \tilde{\mu}(t)) > \eta_2) + \sum_{k \geq W} \Pr \left( \left\{ \sup_{t \in T} \sup_{y \in S_{k,n}(t)} |S_n(y, t) - S_n(\bar{\mu}(t), t)| \geq c \frac{2^{(k-1)}}{r_n^2} \right\} \cap B_R \right),
\]

where \( c > 0 \) is some constant, and the second term goes to zero for any \( \eta > 0 \) according to (27). Since \( d_M(y, \bar{\mu}(t)) \leq 2^k/r_n \) on \( S_{k,n}(t) \), this implies that the sum on the right-hand side of the above inequality is bounded by

\[
4ac \sum_{k \geq W} \frac{2^{-k} r_n^{-1} \sqrt{\log n \sqrt{1/n + 1/nm}}}{r_n^2} \leq \sum_{k \geq W} 2^k \to 0
\]

as \( W \to 0 \). Therefore,

\[
\sup_{t \in T} d_M^2(\bar{\mu}(t), \tilde{\mu}(t)) = O_P \left( \frac{\log n}{n} + \frac{\log n}{nmh} \right).
\]

References

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

Anirudh, R., Turaga, P., Su, J. & Srivastava, A. (2017). Elastic functional coding of Riemannian trajectories. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39**, 922–936.
Bhattacharya, A. & Bhattacharya, R. (2012). Nonparametric inference on manifolds: with applications to shape spaces, vol. 2. Cambridge University Press.

Bhattacharya, R. & Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. The Annals of Statistics 31, 1–29.

Bosq, D. (2000). Linear Processes in Function Spaces. Lecture Notes in Statistics. Springer.

Chen, D. & Müller, H. (2012). Nonlinear manifold representations for functional data. The Annals of Statistics 40, 1–29.

Chen, K. & Lei, J. (2015). Localized functional principal component analysis. Journal of the American Statistical Association 110, 1266–1275.

Chiou, J.-M., Chen, Y.-T. & Yang, Y.-F. (2014). Multivariate functional principal component analysis: A normalization approach. Statistica Sinica 24, 1571–1596.

Dai, X. & Müller, H.-G. (2018). Principal component analysis for functional data on Riemannian manifolds and spheres. Annals of Statistics 46, 3334–3361.

Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G. & Barceló-Vidal, C. (2003). Isometric logratio transformations for compositional data analysis. Mathematical Geology 35, 279–300.

Fan, J. & Gijbels, I. (1996). Local Polynomial Modelling and its Applications. London: Chapman & Hall.

Fang, K., Kotz, S. & Ng, K. (1990). Symmetric multivariate and related distributions. Monographs on statistics and applied probability. London: Chapman and Hall.

Hall, P. & Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. The Annals of Statistics 35, 70–91.

Hall, P. & Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68, 109–126.

Hall, P., Müller, H.-G. & Wang, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. The Annals of Statistics 34, 1493–1517.

Happ, C. & Greven, S. (2018). Multivariate functional principal component analysis for data observed on different (dimensional) domains. Journal of the American Statistical Association 113, 649–659.

Horvath, L. & Kokoszka, P. (2012). Inference for Functional Data with Applications. New York: Springer.

Hsing, T. & Eubank, R. (2015). Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators. Wiley.

Kendall, D., Barden, D., Carne, T. & Le, H. (2009). Shape and Shape Theory. Hoboken: Wiley.

Kleffe, J. (1973). Principal components of random variables with values in a separable Hilbert space. Statistics: A Journal of Theoretical and Applied Statistics 4, 391–406.

Kneip, A., Poss, D. & Sarda, P. (2016). Functional linear regression with points of impact. Annals of Statistics 44, 1–30.

Kong, D., Xue, K., Yao, F. & Zhang, H. H. (2016). Partially functional linear regression
in high dimensions. *Biometrika* 103, 147–159.

**Kraus, D.** (2015). Components and completion of partially observed functional data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 77, 777–801.

**Krueger, A. B. & Mueller, A.** (2011). Job search, emotional well-being, and job finding in a period of mass unemployment: Evidence from high-frequency longitudinal data. *Brookings Papers on Economic Activity* 2011, 1–57.

**Lang, S.** (1995). *Differential and Riemannian Manifolds*. New York: Springer.

**Lee, J. M.** (1997). *Riemannian Manifolds: An Introduction to Curvature*. New York: Springer-Verlag.

**Li, H.** (2015). Microbiome, metagenomics, and high-dimensional compositional data analysis. *Annual Review of Statistics and Its Application* 2, 73–94.

**Li, Y. & Hsing, T.** (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics* 38, 3321–3351.

**Patrangenaru, V., Bubenik, P., Paige, R. L. & Osborne, D.** (2018). Topological data analysis for object data. *arXiv preprint arXiv:1804.10255*.

**Petersen, A. & Müller, H.-G.** (2018). Fréchet regression for random objects with Euclidean predictors. *Annals of Statistics, to appear (arXiv preprint arXiv:1608.03012)*.

**Talská, R., Menafoglio, A., Machalová, J., Hron, K. & Fišerová, E.** (2018). Compositional regression with functional response. *Computational Statistics & Data Analysis* 123, 66–85.

**Thompson, D.** (1942). *On Growth and Form*. Dover Publications.

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

**Verbeke, G., Fieuws, S., Molenberghs, G. & Davidian, M.** (2014). The analysis of multivariate longitudinal data: A review. *Statistical Methods in Medical Research* 23, 42–59.

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

**Yao, F., Müller, H.-G. & Wang, J.-L.** (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* 100, 577–590.

**Yuan, Y., Zhu, H., Lin, W. & Marron, J. S.** (2012). Local polynomial regression for symmetric positive definite matrices. *Journal of Royal Statistical Society: Series B (Statistical Methodology)* 74, 697–719.

**Zhang, J.-T. & Chen, J.** (2007). Statistical inferences for functional data. *The Annals of Statistics* 35, 1052–1079.

**Zhang, X. & Wang, J. L.** (2016). From sparse to dense functional data and beyond. *The Annals of Statistics* 44, 2281–2321.