

Multivariate Functional Principal Component Analysis for Data Observed on Different (Dimensional) Domains

Clara Happ and Sonja Greven*
Department of Statistics, LMU Munich

for the Alzheimer's Disease Neuroimaging Initiative.†

December 6, 2016

Abstract

Existing approaches for multivariate functional principal component analysis are restricted to data on the same one-dimensional interval. The presented approach focuses on multivariate functional data on different domains that may differ in dimension, e.g. functions and images. The theoretical basis for multivariate functional principal component analysis is given in terms of a Karhunen-Loève Theorem. For the practically relevant case of a finite Karhunen-Loève representation, a relationship between univariate and multivariate functional principal component analysis is established. This offers an estimation strategy to calculate multivariate functional principal components and scores based on their univariate counterparts. For the resulting estimators, asymptotic results are derived. The approach can be extended to finite univariate expansions in general, not necessarily orthonormal bases. It is also applicable for sparse functional data or data with measurement error. A flexible R implementation is available on CRAN. The new method is shown to be competitive to existing approaches for data observed on a common one-dimensional domain. The motivating application is a neuroimaging study, where the goal is to explore how longitudinal trajectories of a neuropsychological test score covary with FDG-PET brain scans at baseline. Supplementary material, including detailed proofs, additional simulation results and software is available online.

Keywords:

Functional data analysis, Multivariate functional data, Dimension reduction, Image analysis

*The authors acknowledge support from the German Research Foundation through Emmy Noether grant GR 3793/1-1.

†Data used in preparation of this article were obtained from the Alzheimers Disease Neuroimaging Initiative (ADNI) database (<http://adni.loni.usc.edu>). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this paper. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf. Data collection and sharing for the neuroimaging data in Section 5 was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI, National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). A detailed list of ADNI funding is available at <http://adni.loni.usc.edu/about/funding/>. The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

1 Introduction

Statistical methods for functional data have become increasingly important in recent years. Functional principal component analysis (FPCA) is one of the key techniques in functional data analysis, as it provides an easily interpretable exploratory analysis of the data. Further, it is an important building block for many statistical models (see e.g. Ramsay and Silverman, 2005). The technical progress in many fields of application allows the collection of more and more data with functional features, often several kinds per observation unit. This encourages the study of multivariate functional data and new methods are required to reveal e.g. joint variation in the different elements.

As a simple motivating example, consider the gait cycle data (Ramsay and Silverman, 2005) shown in Fig. 1. It contains 39 observations of hip and knee angle during a gait cycle on a standardized time interval. Both elements of this bivariate data can be described separately by their first three univariate eigenfunctions that explain 94.4% (hip) and 87.5% (knee) of the total variability in the data. The associated functional principal component scores, however, reveal that there is a non negligible correlation between almost all score pairs of the two elements. The separate FPCA thus captures joint variation between hip and knee angles only indirectly, which makes the interpretation of the FPCA results difficult. Correlated scores can also lead to multicollinearity issues in a subsequent regression analysis (functional principal component regression, e.g. Müller and Stadtmüller, 2005). Multivariate FPCA, by contrast, directly addresses potential covariation between the hip and knee elements. The first three bivariate principal components shown in Fig. 1, which explain 85.3% of the variability in the data, give insight into the main modes of joint variation in the overall gait movement. The corresponding scores do not only allow a more parsimonious representation of the data (one score value per bivariate principal component and per observation), but they are also uncorrelated by construction. Finally, the multivariate functional principal components are more natural to represent multivariate functional data in the sense that they have the same structure as each observation. The extension of FPCA to multivariate functional data is hence of high practical relevance.

Existing approaches for multivariate functional principal component analysis (MFPCA) are restricted to functions observed on the same finite, one-dimensional interval (Ramsay and Silverman, 2005; Jacques and Preda, 2014; Chiou et al., 2014; Berrendero et al., 2011). Except for Berrendero et al. (2011), they all aim at a multivariate functional Karhunen-Loève representation of the data. For data measured e.g. in different units, Jacques and Preda (2014) and Chiou et al. (2014) also discuss normalized versions of MFPCA based on a normalized covariance operator.

The key motivation for this paper is that in practical applications, multivariate functional data are neither restricted to lie on the same interval nor to have one-dimensional domains, e.g. data that consists of functions and images, as in our neuroimaging application. We start by extending the notion of multivariate functional data to the case of different (dimensional) domains for the different elements. Next, the theoretical foundations of MFPCA are provided in terms of a Karhunen-Loève Theorem. For the practically relevant case of a finite or truncated Karhunen-Loève representation, we establish a direct theoretical relationship between univariate and multivariate FPCA. This suggests a simple estimation strategy for multivariate functional principal components and scores based on their univariate counterparts. For data on higher dimensional domains (tensor data, e.g. images), principal component methods have originally been developed in the context of psychometrics (e.g. Tucker, 1966; Carroll and Chang, 1970) and have become particularly important in the machine

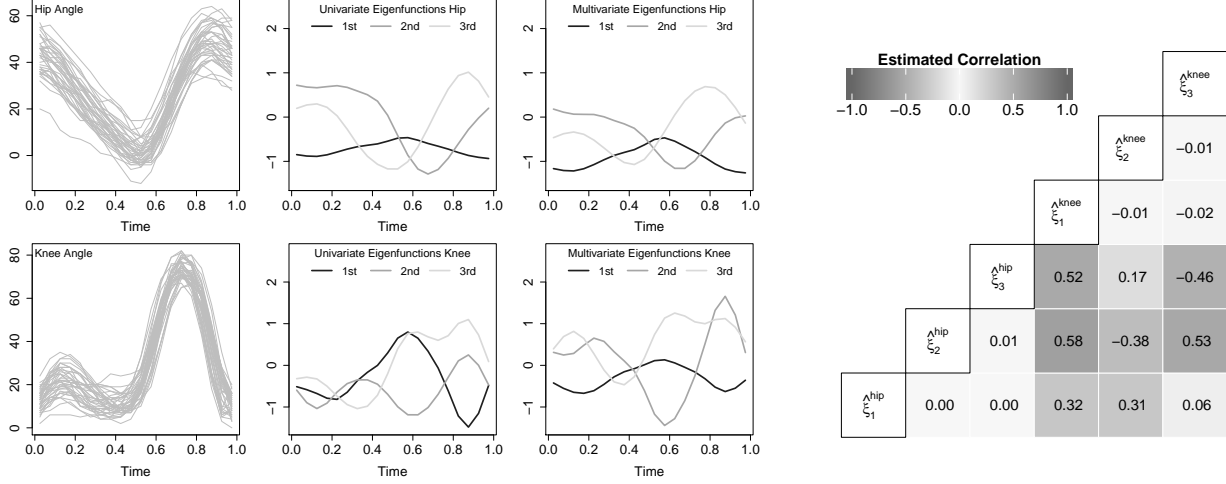


Figure 1: Univariate and multivariate FPCA for the gait cycle data. 1st column: Original data. 2nd column: Results for univariate FPCA, calculated separately. The functions have been reflected, if necessary, and rescaled to have the same norm as the multivariate eigenfunctions for comparison purposes. 3rd column: Results for multivariate FPCA, calculated with the new approach. 4th column: Empirical correlation of the univariate FPCA scores for hip and knee.

learning literature (Coppi and Bolasco, 1989; Lu et al., 2013). Recent approaches for functional or smooth principal component analysis for tensor data have been proposed e.g. in Allen (2013). All these methods can be used as univariate building blocks for MFPCA. The resulting estimators for MFPCA are shown to be consistent under a given set of assumptions. In contrast to most of the existing methods for MFPCA, our new approach can be applied to sparse functional data and data with measurement error. It can be generalized to data available in arbitrary basis expansions and hence includes the MFPCA procedure proposed by Jacques and Preda (2014) as a special case. The new method further allows to incorporate weights for the elements, if they differ in domain, range or variation.

The paper is organized as follows. Section 2 introduces multivariate functional data and gives the theoretical basis for MFPCA. In Section 3 we derive the estimation algorithm for MFPCA based on univariate basis expansions and investigate asymptotic properties of the resulting estimators. The performance of the new method is evaluated in Section 4 in a simulation with different levels of complexity. Section 5 contains the analysis of the motivating neuroimaging dataset. The paper concludes with a discussion and an outlook in Section 6. Supplementary material, containing detailed proofs of all propositions, more simulation results and R code is available online.

2 Theoretical Foundations of Multivariate Functional Data

2.1 Data Structure and Notation

This paper is concerned with *multivariate functional data*, i.e. each observation consists of $p \geq 2$ functions $X^{(1)}, \dots, X^{(p)}$. They may be defined on different domains $\mathcal{T}_1, \dots, \mathcal{T}_p$ with possibly different dimensions. Technically, \mathcal{T}_j must be compact sets in \mathbb{R}^{d_j} , $d_j \in \mathbb{N}$ with finite (Lebesgue-) measure and each element $X^{(j)}: \mathcal{T}_j \rightarrow \mathbb{R}$ is assumed to be in $L^2(\mathcal{T}_j)$.

In analogy to other approaches for multivariate functional data, the different functions are combined in a vector X with

$$X(\mathbf{t}) = \left(X^{(1)}(t_1), \dots, X^{(p)}(t_p) \right) \in \mathbb{R}^p.$$

Note that $\mathbf{t} := (t_1, \dots, t_p) \in \mathcal{T} := \mathcal{T}_1 \times \dots \times \mathcal{T}_p$ is a p -tuple of d_1, \dots, d_p -dimensional vectors and not a scalar. This is a main difference to earlier approaches, as it allows each element $X^{(j)}$ to have a different argument t_j , even in the case of a common one-dimensional domain. In the following, it will be further assumed that

$$\mu(\mathbf{t}) := \mathbb{E}(X(\mathbf{t})) = \left(\mathbb{E}\left(X^{(1)}(t_1)\right), \dots, \mathbb{E}\left(X^{(p)}(t_p)\right) \right) = \mathbf{0} \quad \forall \mathbf{t} \in \mathcal{T}.$$

For $\mathbf{s}, \mathbf{t} \in \mathcal{T}$, define the matrix of covariances $C(\mathbf{s}, \mathbf{t}) := \mathbb{E}(X(\mathbf{s}) \otimes X(\mathbf{t}))$ with elements

$$C_{ij}(s_i, t_j) := \mathbb{E}\left(X^{(i)}(s_i)X^{(j)}(t_j)\right) = \text{Cov}(X^{(i)}(s_i), X^{(j)}(t_j)), \quad s_i \in \mathcal{T}_i, t_j \in \mathcal{T}_j. \quad (1)$$

As noted in Ramsay and Silverman (2005, Chapter 8.5.), a suitable inner product is the basis of all approaches for principal component analysis. For functions $f = (f^{(1)}, \dots, f^{(p)})$ with elements $f^{(j)} \in L^2(\mathcal{T}_j)$ define the space $\mathcal{H} := L^2(\mathcal{T}_1) \times \dots \times L^2(\mathcal{T}_p)$ and

$$\langle\langle f, g \rangle\rangle := \sum_{j=1}^p \langle f^{(j)}, g^{(j)} \rangle_2 = \sum_{j=1}^p \int_{\mathcal{T}_j} f^{(j)}(t_j) g^{(j)}(t_j) dt_j, \quad f, g \in \mathcal{H}. \quad (2)$$

Proposition 1. \mathcal{H} is a Hilbert space with respect to the scalar product $\langle\langle \cdot, \cdot \rangle\rangle$.

Proofs for all propositions are given in the online appendix. The norm induced by $\langle\langle \cdot, \cdot \rangle\rangle$ is denoted by $\|\cdot\|$ ¹. Next, define the covariance operator $\Gamma: \mathcal{H} \rightarrow \mathcal{H}$ with the j -th element of Γf , $f \in \mathcal{H}$ given by

$$(\Gamma f)^{(j)}(t_j) := \sum_{i=1}^p \int_{\mathcal{T}_i} C_{ij}(s_i, t_j) f^{(i)}(s_i) ds_i = \langle\langle C_{\cdot,j}(\cdot, t_j), f \rangle\rangle, \quad t_j \in \mathcal{T}_j. \quad (3)$$

The setting can be generalized to a weighted scalar product on \mathcal{H} , i.e.

$$\langle\langle f, g \rangle\rangle_w := \sum_{j=1}^p w_j \langle f^{(j)}, g^{(j)} \rangle_2, \quad f, g \in \mathcal{H} \quad (4)$$

for some positive weights w_1, \dots, w_p , cf. Ramsay and Silverman (2005, Chapter 10.3. in the context of hybrid data) or Chiou et al. (2014). The associated weighted covariance operator Γ_w is given by its elements $(\Gamma_w f)^{(j)}$ with $f \in \mathcal{H}$ and

$$(\Gamma_w f)^{(j)}(t_j) = \langle\langle C_{\cdot,j}(\cdot, t_j), f \rangle\rangle_w, \quad t_j \in \mathcal{T}_j.$$

¹The L^2 -norm induced by $\langle \cdot, \cdot \rangle_2$ on each $L^2(\mathcal{T}_j)$ is denoted by $\|\cdot\|_2$. Further, $\|\cdot\|$ is the Euclidean norm for vectors and $\|\cdot\|_{\mathcal{T}}$ denotes a norm on \mathcal{T} with $\|\mathbf{t}\|_{\mathcal{T}}^2 = \sum_{j=1}^p \|t_j\|_2^2$ for $t_j \in \mathcal{T}_j \subset \mathbb{R}^{d_j}$, $j = 1, \dots, p$.

The use of weights may be necessary if the elements have quite different domains or ranges or if they exhibit different amounts of variation, in order to obtain multivariate functional principal components that have a meaningful interpretation (Chiou et al., 2014). A weighted scalar product corresponds to a (global) rescaling of the elements by $w_j^{1/2}$. An alternative approach would be pointwise rescaling, e.g. by the inverse of the square root of the pointwise variance $C_{jj}(t_j, t_j)$. This can be seen as normalizing the covariance operator (Chiou et al., 2014; Jacques and Preda, 2014). However, this second approach does not consider the size of the different domains \mathcal{T}_j and would give equal variation per observation point t_j rather than per element j . Moreover, rescaling with the pointwise variance would downweight areas in \mathcal{T}_j with stronger variation, hence areas that might contribute relevant information to the functional principal components. Therefore, only global rescaling by means of a weighted scalar product is considered in the following. The weights have to be chosen prior to the analysis. They can be specified based on expert knowledge or estimated from the data, e.g. based on the variation in each element (see references in Chiou et al., 2014). A sensible choice will always depend on the specific application and the question of interest. One possible solution that is analogous to standardization in multivariate PCA is proposed in the application in Section 5. For the sake of better readability, all following theoretical results are derived for $w_1 = \dots = w_p = 1$, but remain valid in the more general case of different weights. For the estimation algorithm discussed in Section 3.2, MFPCA based on the weighted scalar product is addressed again.

2.2 A Karhunen-Loève Theorem for Multivariate Functional Data

In the following it is shown that under mild conditions, Γ has the same properties as the covariance operator in the univariate case and therefore a Karhunen-Loève representation for multivariate functional data exists. The main difference to existing approaches for data with elements observed on the same (one-dimensional) domain is that in this special case, Γ is an integral operator with positive definite kernel $C(s, t)$. This directly gives all of the desired properties (Saporta, 1981). In the more general case of elements observed on different domains, this is not obviously the case and the properties are shown explicitly.

Proposition 2. *The covariance operator Γ defined in (3) is a linear, self-adjoint and positive operator. If further for all $i, j = 1, \dots, p$ there exist $K_{ij} < \infty$ with*

$$\|C_{ij}(\cdot, t_j)\|_2^2 = \int_{\mathcal{T}_i} C_{ij}(s_i, t_j)^2 ds_i < K_{ij} \quad \forall t_j \in \mathcal{T}_j, \quad (5)$$

and C_{ij} is uniformly continuous in the sense that

$$\forall \varepsilon > 0 \exists \delta_{ij} > 0 : \quad \|t_j - t_j^*\| < \delta_{ij} \quad \Rightarrow \quad |C_{ij}(s_i, t_j) - C_{ij}(s_i, t_j^*)| < \varepsilon \quad \forall s_i \in \mathcal{T}_i,$$

then Γ is a compact operator.

In the remainder of this paper, it is assumed that the C_{ij} satisfy all conditions of Prop. 2 and hence Γ can always be assumed to be a compact positive operator on \mathcal{H} . By the Hilbert-Schmidt Theorem (e.g. Reed and Simon, 1980, Thm. VI.16) it follows that there exists a complete orthonormal basis of eigenfunctions $\psi_m \in \mathcal{H}$, $m \in \mathbb{N}$ of Γ such that

$$\Gamma \psi_m = \nu_m \psi_m \quad \text{and} \quad \nu_m \rightarrow 0 \quad \text{for } m \rightarrow \infty.$$

In particular, since Γ is a positive operator, it may be assumed w.l.o.g. that $\nu_1 \geq \nu_2 \geq \dots \geq 0$. Since ψ_m , $m \in \mathbb{N}$ is an orthonormal basis of \mathcal{H} and Γ is self-adjoint, by the Spectral Theorem (e.g. Werner, 2011, Thm. VI.3.2.) it holds that

$$\Gamma f = \sum_{m=1}^{\infty} \nu_m \langle f, \psi_m \rangle \psi_m \quad \forall f \in \mathcal{H}.$$

The following proposition is a multivariate version of Mercer's Theorem (Mercer, 1909). It plays a key role in the proof of the Karhunen-Loève Theorem (Prop. 4).

Proposition 3 (Mercer's Theorem). *For $j = 1, \dots, p$ and $s_j, t_j \in \mathcal{T}_j$ it holds that*

$$\text{Cov} \left(X^{(j)}(s_j), X^{(j)}(t_j) \right) = C_{jj}(s_j, t_j) = \sum_{m=1}^{\infty} \nu_m \psi_m^{(j)}(s_j) \psi_m^{(j)}(t_j),$$

where the convergence is absolute and uniform.

Proposition 4 (Multivariate Karhunen-Loève Theorem). *Under the assumptions of Prop. 2,*

$$X(\mathbf{t}) = \sum_{m=1}^{\infty} \rho_m \psi_m(\mathbf{t}), \quad \mathbf{t} \in \mathcal{T}, \quad (6)$$

with zero mean random variables $\rho_m = \langle X, \psi_m \rangle$ and $\text{Cov}(\rho_m, \rho_n) = \nu_m \delta_{mn}$. Moreover

$$\mathbb{E} \left(\left\| X(\mathbf{t}) - \sum_{m=1}^M \rho_m \psi_m(\mathbf{t}) \right\|^2 \right) \rightarrow 0 \quad \text{for } M \rightarrow \infty$$

uniformly for $\mathbf{t} \in \mathcal{T}$.

The multivariate Karhunen-Loève representation has an analogous interpretation as in the univariate case (Ramsay and Silverman, 2005, Chapter 8.2.). The eigenvalues ν_m represent the amount of variability in X explained by the single *multivariate functional principal components* ψ_m , while the *multivariate functional principal component scores* ρ_m serve as weights of ψ_m in the Karhunen-Loève representation of X . As the eigenvalues ν_m decrease towards 0, leading eigenfunctions reflect the most important features of X . Truncated Karhunen-Loève expansions, optimal M -dimensional approximations to X ,

$$X_{[M]}(\mathbf{t}) := \sum_{m=1}^M \rho_m \psi_m(\mathbf{t}), \quad \mathbf{t} \in \mathcal{T}, \quad (7)$$

are often used in practice. Single observations x_i of X can then be characterized by their score vectors $(\rho_{i,1}, \dots, \rho_{i,M})$ with $\rho_{i,m} = \langle x_i, \psi_m \rangle$ for further analysis, e.g. for regression (Müller and Stadtmüller, 2005) or clustering (Jacques and Preda, 2014).

3 Multivariate FPCA

3.1 Relationship Between Univariate and Multivariate FPCA for Finite Karhunen-Loève Decompositions

Given the Karhunen-Loève representation of multivariate functional data X as in (6), a natural question is how this representation relates to the univariate Karhunen-Loève representations of the single elements $X^{(j)}$. The following proposition establishes a direct relationship between these two representations if they are both finite, based on the theory of integral equations (Zemlyan, 2012).

Proposition 5. *The multivariate functional vector $X = (X^{(1)}, \dots, X^{(p)})$ in (6) has a finite Karhunen-Loève representation if and only if all univariate elements $X^{(1)}, \dots, X^{(p)}$, have a finite Karhunen-Loève representation. In this case, it holds:*

1. *Given the multivariate Karhunen-Loève representation (6), the positive eigenvalues $\lambda_1^{(j)} \geq \dots \geq \lambda_{M_j}^{(j)} > 0$, $M_j \leq M$ of the univariate covariance operator $\Gamma^{(j)}$ associated with $X^{(j)}$ correspond to the positive eigenvalues of the matrix $\mathbf{A}^{(j)} \in \mathbb{R}^{M \times M}$ with entries*

$$A_{mn}^{(j)} = (\nu_m \nu_n)^{1/2} \langle \psi_m^{(j)}, \psi_n^{(j)} \rangle_2, \quad m, n = 1, \dots, M.$$

The eigenfunctions of $\Gamma^{(j)}$ are given by

$$\phi_m^{(j)}(t_j) = \left(\lambda_m^{(j)} \right)^{-1/2} \sum_{n=1}^M \nu_n^{1/2} [\mathbf{u}_m^{(j)}]_n \psi_n^{(j)}(t_j), \quad t_j \in \mathcal{T}_j, \quad m = 1, \dots, M_j,$$

where $\mathbf{u}_m^{(j)}$ denotes an (orthonormal) eigenvector of $\mathbf{A}^{(j)}$ associated with eigenvalue $\lambda_m^{(j)}$ and $[\mathbf{u}_m^{(j)}]_n$ denotes the n -th entry of this vector. For the univariate scores

$$\xi_m^{(j)} = \langle X^{(j)}, \phi_m^{(j)} \rangle_2 = \left(\lambda_m^{(j)} \right)^{-1/2} \sum_{n=1}^M \nu_n^{1/2} [\mathbf{u}_m^{(j)}]_n \sum_{k=1}^M \rho_k \langle \psi_n^{(j)}, \psi_k^{(j)} \rangle_2.$$

2. *Assuming the univariate Karhunen-Loève representation $X^{(j)} = \sum_{m=1}^{M_j} \xi_m^{(j)} \phi_m^{(j)}$ with $\Gamma^{(j)} \phi_m^{(j)} = \lambda_m^{(j)} \phi_m^{(j)}$ for each element $X^{(j)}$ of X , the positive eigenvalues $\nu_1 \geq \dots \geq \nu_M > 0$ of Γ with $M \leq \sum_{j=1}^p M_j =: M_+$ correspond to the positive eigenvalues of the matrix $\mathbf{Z} \in \mathbb{R}^{M_+ \times M_+}$ consisting of blocks $\mathbf{Z}^{(jk)} \in \mathbb{R}^{M_j \times M_k}$ with entries*

$$Z_{mn}^{(jk)} = \text{Cov} \left(\xi_m^{(j)}, \xi_n^{(k)} \right), \quad m = 1, \dots, M_j, \quad n = 1, \dots, M_k, \quad j, k = 1, \dots, p.$$

The eigenfunctions of Γ are given by their elements

$$\psi_m^{(j)}(t_j) = \sum_{n=1}^{M_j} [\mathbf{c}_m]_n^{(j)} \phi_n^{(j)}(t_j), \quad t_j \in \mathcal{T}_j, \quad m = 1, \dots, M,$$

where $[\mathbf{c}_m]^{(j)} \in \mathbb{R}^{M_j}$ denotes the j -th block of an (orthonormal) eigenvector \mathbf{c}_m of \mathbf{Z} associated with eigenvalue ν_m . The scores are given by

$$\rho_m = \sum_{j=1}^p \sum_{n=1}^{M_j} [\mathbf{c}_m]_n^{(j)} \xi_n^{(j)}.$$

Extensions: The second part of Prop. 5 can be extended in a natural way if univariate elements are expanded in finitely many, not necessarily orthonormal basis functions $b_m^{(j)}$ with coefficients $\theta_m^{(j)}$, i.e.

$$X^{(j)}(t_j) = \sum_{m=1}^{K_j} \theta_m^{(j)} b_m^{(j)}(t_j), \quad t_j \in \mathcal{T}_j. \quad (8)$$

This is a very likely situation in practice, e.g. due to pre-smoothing of noisy observations. Following analogous steps as in the proof of Prop. 5 results in an eigenanalysis problem $\mathbf{B}\mathbf{Q}\mathbf{c} = \nu\mathbf{c}$ as starting point for the MFPCA. Here $\mathbf{B} \in \mathbb{R}^{K_+ \times K_+}$ with $K_+ = \sum_{j=1}^p K_j$ is a block diagonal matrix of scalar products $\langle b_m^{(j)}, b_n^{(j)} \rangle_2$ of univariate basis functions associated with each element $X^{(j)}$. In the special case that all univariate bases are orthonormal (e.g. when using the univariate principal component bases as in Prop. 5), \mathbf{B} equals the identity matrix. The symmetric block matrix \mathbf{Q} with entries $Q_{mn}^{(jk)} = \text{Cov}(\theta_m^{(j)}, \theta_n^{(k)})$ corresponds to \mathbf{Z} in Prop. 5. Although $\mathbf{B}\mathbf{Q}$ is in general not symmetric, its eigenvectors \mathbf{c}_m and eigenvalues ν_m , which are at the same time the eigenvalues of Γ , are real. This can be easily shown using the Cholesky decomposition of the symmetric matrix $\mathbf{B} = \mathbf{R}\mathbf{R}^\top$ and solving $\mathbf{R}^\top \mathbf{Q} \mathbf{R} \tilde{\mathbf{c}} = \nu \tilde{\mathbf{c}}$ with $\tilde{\mathbf{c}} = \mathbf{R}^{-1} \mathbf{c}$. The estimation algorithm for principal components ψ_m and associated scores ρ_m based on this general basis expansion is presented in the next section combined with the case of a weighted scalar product.

3.2 Estimation of Multivariate FPCA

Estimation based on univariate FPCA: The second part of Prop. 5. suggests a simple and natural approach for estimating the MFPCA. After calculation of univariate FPCAs for each element, the estimates can be plugged into the formulae given in Prop. 5. Given de-meaned samples x_1, \dots, x_N of X , the proposed estimation procedure for MFPCA consists of four steps:

1. For each element $X^{(j)}$ estimate a univariate FPCA based on the observations $x_1^{(j)}, \dots, x_N^{(j)}$. This results in estimated eigenfunctions $\hat{\phi}_m^{(j)}$ and scores $\hat{\xi}_{i,m}^{(j)}$, $i = 1, \dots, N$, $m = 1, \dots, M_j$ for suitably chosen truncation lags M_j . As there exist numerous estimation procedures, e.g. for irregularly sampled and sparse data with measurement error (Yao et al., 2005), the multivariate method is also applicable to this kind of data.
2. Define the matrix $\Xi \in \mathbb{R}^{N \times M_+}$, where each row $(\hat{\xi}_{i,1}^{(1)}, \dots, \hat{\xi}_{i,M_1}^{(1)}, \dots, \hat{\xi}_{i,1}^{(p)}, \dots, \hat{\xi}_{i,M_p}^{(p)})$ contains all estimated scores for a single observation. An estimate $\hat{\mathbf{Z}} \in \mathbb{R}^{M_+ \times M_+}$ of the block matrix \mathbf{Z} in Prop. 5 is given by $\hat{\mathbf{Z}} = (N-1)^{-1} \Xi^\top \Xi$.
3. Perform a matrix eigenanalysis for $\hat{\mathbf{Z}}$ resulting in eigenvalues $\hat{\nu}_m$ and orthonormal eigenvectors $\hat{\mathbf{c}}_m$.
4. Estimates for the multivariate eigenfunctions are given by their elements

$$\hat{\psi}_m^{(j)}(t_j) = \sum_{n=1}^{M_j} [\hat{\mathbf{c}}_m]_n^{(j)} \hat{\phi}_n^{(j)}(t_j), \quad t_j \in \mathcal{T}_j, \quad m = 1, \dots, M^+ \quad (9)$$

and multivariate scores can be calculated via

$$\hat{\rho}_{i,m} = \sum_{j=1}^p \sum_{n=1}^{M_j} [\hat{\mathbf{c}}_m]_n^{(j)} \hat{\xi}_{i,n}^{(j)} = \Xi_{i,\cdot} \hat{\mathbf{c}}_m. \quad (10)$$

Finding an appropriate truncation lag M_j in step 1 is a well-known issue in functional data analysis. Common approaches are based on the decrease of the estimated eigenvalues $\hat{\lambda}_m^{(j)}$ (scree-plot, Cattell, 1966) or the percentage of variance explained (e.g. Ramsay and Silverman, 2005, Chapter 8.2.). An optimal number $M \leq M_+$ of multivariate functional principal components can basically be chosen with the same techniques, while the importance of a “correct” choice depends on the specific application: For simply exploratory aims it is less crucial than for subsequent analyses that ignore the information of the eigenvalues (and hence, the proportion of variance explained by the single components) and are based solely on multivariate eigenfunctions or scores, as e.g. clustering or functional principal component regression. For the latter, relevant eigenfunctions can also be selected using model-based approaches such as AIC or cross-validation. The goodness of the resulting MFPCA estimates of course depends on an appropriate choice of M_j , which can also be used as a sensitivity check: If the first M_j eigenfunctions capture all the relevant information in $X^{(j)}$, increasing M_j will add only little information and hence should have only little impact on the results. This relationship is analyzed in a simulation in the online appendix.

Extensions: The estimation algorithm can easily be extended to elements $X^{(j)}$ available in general basis expansions as in (8) and to MFPCA based on a weighted scalar product as in (4). Given weights $w_1, \dots, w_p > 0$ and demeaned observations x_1, \dots, x_N of X with estimated basis function coefficients $\hat{\theta}_{i,m}^{(j)}$ for each element, the eigenanalysis problem to solve is

$$(N-1)^{-1} \mathbf{B} \mathbf{D} \mathbf{\Theta}^\top \mathbf{\Theta} \mathbf{D} \mathbf{c} = \nu \mathbf{c}. \quad (11)$$

The matrix \mathbf{B} is the block diagonal matrix of basis scalar products as in Section 3.1 and $\mathbf{D} = \text{diag}(\mathbf{w}_1^{1/2}, \dots, \mathbf{w}_p^{1/2}) \in \mathbb{R}^{K_+ \times K_+}$ accounts for the weights, where each $w_j^{1/2}$ is repeated K_j times to give $\mathbf{w}_j^{1/2}$. $\mathbf{\Theta} \in \mathbb{R}^{N \times K_+}$ with rows $(\hat{\theta}_{i,1}^{(1)}, \dots, \hat{\theta}_{i,K_1}^{(1)}, \dots, \hat{\theta}_{i,1}^{(p)}, \dots, \hat{\theta}_{i,K_p}^{(p)})$ corresponds to the matrix $\mathbf{\Xi}$ defined in step 2 of the original algorithm and $(N-1)^{-1} \mathbf{\Theta}^\top \mathbf{\Theta}$ is an estimate for \mathbf{Q} introduced in Section 3.1. Given eigenvectors $\hat{\mathbf{c}}_m$ and eigenvalues $\hat{\nu}_m$ for (11), estimated orthonormal eigenfunctions $\hat{\psi}_m$ of Γ_w and associated scores $\hat{\rho}_{i,m}$ can be calculated in analogy to (9) and (10) with $\hat{\mathbf{Q}}_w = (N-1)^{-1} \mathbf{D} \mathbf{\Theta}^\top \mathbf{\Theta} \mathbf{D}$:

$$\begin{aligned} \hat{\psi}_m^{(j)}(t_j) &= \left(w_j \cdot \hat{\nu}_m \hat{\mathbf{c}}_m^\top \hat{\mathbf{Q}}_w \hat{\mathbf{c}}_m \right)^{-1/2} \sum_{k=1}^p \sum_{l=1}^{K_j} \sum_{n=1}^{K_k} [\hat{\mathbf{Q}}_w]_{ln}^{(jk)} [\hat{\mathbf{c}}_m]_n^{(k)} b_l^{(j)}(t_j), \\ \hat{\rho}_{i,m} &= (\hat{\nu}_m)^{1/2} \left(\hat{\mathbf{c}}_m^\top \hat{\mathbf{Q}}_w \hat{\mathbf{c}}_m \right)^{-1/2} \mathbf{\Theta}_{i,\cdot} \mathbf{D} \hat{\mathbf{c}}_m. \end{aligned}$$

Clearly, the original algorithm is obtained as a special case with $\mathbf{\Theta} = \mathbf{\Xi}$, $\mathbf{B} = \mathbf{I}$ (univariate FPCA for each element) and $\mathbf{D} = \mathbf{I}$ (all weights equal to 1). Moreover, the extended algorithm allows to flexibly combine univariate FPCA and general basis expansions for different elements of the multivariate functional data.

If all elements $X^{(j)}$ are defined on the same (one-dimensional) interval and $\mathbf{D} = \mathbf{I}$, expanding each element in a general basis is equivalent to the method of Jacques and Preda (2014). The approach proposed in this paper, however, is more general, as it allows for different intervals as well as for higher-dimensional \mathcal{T}_j and thus basis functions $b_m^{(j)}$.

Implementation: All presented variations of the MFPCA estimation algorithm are implemented in an R package MFPCA (Happ, 2016b). Univariate basis expansions include univariate FPCA (1D), smooth tensor PCA (2D), spline bases (1D/2D) and cosine bases (2D/3D). New bases can be

added easily and in a modular way. The **MFPCA** package is based on the package **funData** (Happ, 2016a) for representing (multivariate) functional data on potentially different dimensional domains.

3.3 Asymptotic Properties

The results of Prop. 5 and the estimators proposed in the previous section have been derived under the assumption of a finite sample size N and a finite Karhunen-Loève representation for each element $X^{(j)}$. This case is relevant in practice, since data is observable only in finite form (finitely many observations, finite resolution) and hence contains only finite information. In this case, the maximal number of principal components which can be estimated is limited to the number of observations N . For a growing number of observations, the truncation limits M_j and thus M_+ may increase with N . All asymptotic examinations hence have to consider the approximation error caused by truncating the univariate Karhunen-Loève representations to finite sums as well as the estimation error. For the eigenfunctions (analogously for the eigenvalues and scores) one hence has the following decomposition:

$$\left\| \psi_m - \hat{\psi}_m \right\| \leq \left\| \psi_m - \psi_m^{[M]} \right\| + \left\| \psi_m^{[M]} - \hat{\psi}_m \right\|.$$

Here ψ_m is the true m -th eigenfunction of the covariance operator Γ and $\hat{\psi}_m$ is the estimator based on the assumption of a finite Karhunen-Loève representation in each element. This assumption is reflected in $\psi_m^{[M]}$, which denotes the m -th eigenfunction of the covariance operator $\Gamma^{[M]}$ associated with $X^{[M]}$ with elements equal to the truncated $X^{(j)}$. These are really the eigenfunctions targeted with the estimation algorithm presented in Section 3.2. The first term on the right hand side of the inequality can be seen as a bias term caused by truncation. It depends on N only implicitly via M_1, \dots, M_p . The second term accounts for the estimation error, thus can be interpreted as a variance term.

Proposition 6 (Approximation Error). *Let $\nu_m^{[M]}$, $m \in \mathbb{N}$ be the eigenvalues of the covariance operator $\Gamma^{[M]}$ associated with $X^{[M]}$ having truncated univariate elements $X^{[M](j)} = \sum_{m=1}^{M_j} \xi_m^{(j)} \phi_m^{(j)}$. Then the approximation error $\left\| X^{[M]} - X \right\|$ converges to 0 in probability for $M_1, \dots, M_p \rightarrow \infty$. For each $m \in \mathbb{N}$, $\nu_m^{[M]}$ converges to ν_m including multiplicity and the total projection $P_m^{[M]}$ of \mathcal{H} onto the eigenspace of $\Gamma^{[M]}$ associated with $\nu_m^{[M]}$ converges in norm to the total projection P_m of \mathcal{H} onto the eigenspace of Γ associated with ν_m .*

In particular, if ν_m and $\nu_m^{[M]}$ both have multiplicity 1 with associated eigenfunctions ψ_m and $\psi_m^{[M]}$, such that $\langle \psi_m, \psi_m^{[M]} \rangle \geq 0$, then

$$\left\| \psi_m^{[M]} - \psi_m \right\| \rightarrow 0 \quad \text{for } M_1, \dots, M_p \rightarrow \infty.$$

The scores $\rho_m^{[M]} := \langle X^{[M]}, \psi_m^{[M]} \rangle$ converge in probability to ρ_m for all $m \in \mathbb{N}$.

In the remainder of this section, all nonzero eigenvalues ν_m are assumed to have multiplicity 1, as then the eigenfunctions $\psi_m^{[M]}$ converge to ψ_m , if their orientation is chosen such that $\langle \psi_m, \psi_m^{[M]} \rangle \geq 0$.

For the estimation error, consider the univariate elements $X^{(j)}$ of X with covariance operator $\Gamma^{(j)}$ and associated eigenvalues $\lambda_m^{(j)}$ and eigenfunctions $\phi_m^{(j)}$, $m = 1, \dots, M_j$. In the following, let

X_1, \dots, X_N be independent copies of X and assume for all $j = 1, \dots, p$

$$\Delta_{M_j}^{(j)} := \sup_{m=1, \dots, M_j} (\lambda_m^{(j)} - \lambda_{m+1}^{(j)})^{-1} < \infty \text{ for every finite } M_j \quad (\text{A1})$$

$$\int_{\mathcal{T}_j} \int_{\mathcal{T}_k} \mathbb{E} \left(X^{(j)}(t_j)^2 X^{(k)}(s_k)^2 \right) ds_k dt_j < \infty \quad \forall k = 1, \dots, p \quad (\text{A2})$$

$$\left\| \Gamma^{(j)} - \hat{\Gamma}^{(j)} \right\|_{\text{op}} = O_p(r_N^\Gamma) \quad (\text{A3})$$

$$\langle \phi_m^{(j)}, \hat{\phi}_m^{(j)} \rangle_2 \geq 0 \text{ for all } m = 1, \dots, M_j \quad (\text{A4})$$

$$\hat{\xi}_{i,m}^{(j)} = \langle X_i^{(j)}, \hat{\phi}_m^{(j)} \rangle_2 \text{ for all } m = 1, \dots, M_j, i = 1, \dots, N \quad (\text{A5})$$

(A1) – (A2) concern theoretical properties of $X^{(j)}$ and $\Gamma^{(j)}$, while (A3) – (A5) depend on the univariate decompositions used. (A1) is a standard assumption in univariate FPCA (Bosq, 2000; Hall and Hosseini-Nasab, 2006). It guarantees that the first M_j univariate eigenvalues of each element all have multiplicity 1. With (A2), the integral operator with kernel $\hat{C}_{jk}(s, t) := N^{-1} \sum_{i=1}^N X_i^{(j)}(s) X_i^{(k)}(t)$ converges to the one with kernel $C_{jk}(s, t)$ with rate $N^{-1/2}$. (A2) is used in combination with (A5) to obtain a convergence rate for the maximal eigenvalue of $\mathbf{Z} - \hat{\mathbf{Z}}$, which, in turn, affects the convergence of the eigenvectors $\hat{\mathbf{c}}_m$ to \mathbf{c}_m (Yu et al., 2015). (A3) ensures that the operator $\hat{\Gamma}^{(j)}$, which is the basis of the univariate FPCA, converges to $\Gamma^{(j)}$ in the operator norm $\|\cdot\|_{\text{op}}$ induced by $\|\cdot\|_2$ with a given rate r_N^Γ . For fully observed data, Hall and Horowitz (2007) show $r_N^\Gamma = N^{-1/2}$, while the approach of Yao et al. (2005) yields $r_N^\Gamma = N^{-1/2}h^{-2}$ in the case of measurement error or irregularly sampled data for a certain bandwidth h . Together with (A1), r_N^Γ gives a convergence rate for the univariate eigenfunctions $\hat{\phi}_m^{(j)}$ (Bosq, 2000, Lemma 4.3). (A4) guarantees that $\hat{\phi}_m^{(j)}$ is an estimator for $\phi_m^{(j)}$ rather than for $-\phi_m^{(j)}$, as eigenfunctions are defined only up to a sign change (Bosq, 2000; Hall and Hosseini-Nasab, 2006). Finally, (A5) is used to formulate the convergence of the estimated scores in terms of convergence rates for the estimated eigenfunctions. If this assumption does not hold (e.g. in Yao et al., 2005), convergence results can still be obtained e.g. by assuming a convergence rate for $\hat{\xi}_{i,m}^{(j)}$ and replacing (A2) by an assumption on the rate of convergence for the maximal eigenvalue of $\mathbf{Z} - \hat{\mathbf{Z}}$.

Proposition 7 (Estimation Error). *Assume (A1) – (A5) hold. Then for $M_{\max} = \max_{j=1, \dots, p} M_j$ and $\Delta_M := \max_{j=1, \dots, p} \Delta_{M_j}^{(j)}$, the maximal eigenvalue of $\mathbf{Z} - \hat{\mathbf{Z}}$ can be characterized by*

$$\lambda_{\max}(\mathbf{Z} - \hat{\mathbf{Z}}) = O_p(M_{\max} \max(N^{-1/2}, \Delta_M r_N^\Gamma)).$$

Using the same notation as in Prop. 6, it holds for all $m = 1, \dots, M_+$ that

$$\begin{aligned} \left| \nu_m^{[M]} - \hat{\nu}_m \right| &= O_p(M_{\max} \max(N^{-1/2}, \Delta_M r_N^\Gamma)), \\ \left\| \psi_m^{[M]} - \hat{\psi}_m \right\| &= O_p(M_{\max}^{3/2} \max(N^{-1/2}, \Delta_M r_N^\Gamma)), \\ \left| \rho_{i,m}^{[M]} - \hat{\rho}_{i,m} \right| &= O_p(M_{\max}^{3/2} \max(N^{-1/2}, \Delta_M r_N^\Gamma)), \\ \left\| X_i^{[M]} - \hat{X}_i^{[M]} \right\| &= O_p(M_{\max} \Delta_M r_N^\Gamma) \end{aligned}$$

with $\hat{X}_i^{[M](j)} = \sum_{m=1}^{M_j} \hat{\xi}_{i,m}^{(j)} \hat{\phi}_m^{(j)}$.

When combining the results of Prop. 6 and Prop. 7, the analogy to bias and variance again becomes apparent: For fixed N , higher values of M_1, \dots, M_p will reduce the approximation error, but simultaneously increase the estimation error, as both M_{\max} and Δ_M increase with M_j . If one assumes for example $M_1 = \dots = M_p = M_{\max} = O(N^\beta)$, $r_N^\Gamma = N^{-1/2}$, and that the eigengaps fulfill $\lambda_m^{(j)} - \lambda_{m+1}^{(j)} \geq C^{-1}m^{-\alpha-1}$ with $\alpha > 1$, $C > 0$ (cf. Hall and Horowitz, 2007), the MFPCA estimators given in Section 3.2 are consistent for $0 < \beta < (2\alpha + 5)^{-1}$.

4 Simulation

We illustrate the performance of our new MFPCA estimation procedure in three settings with increasing complexity:

1. Densely observed bivariate functional data on the same one-dimensional interval.
2. Trivariate functional data on different one-dimensional intervals with different levels of sparsity.
3. Bivariate functional data on different dimensional domains (images and functions).

The first two settings deal with multivariate functional data on one-dimensional domains and are presented together in Section 4.1. Setting 3 is discussed separately in Section 4.2. Examples for simulated data and estimation results for all three settings are given in the online appendix, which also includes two additional simulations (cf. Sections 3.2 and 5). Unless specified otherwise, the MFPCA package (Happ, 2016b) is used for all calculations.

Each setting is based on 100 datasets with $N = 250$ observations of the form

$$x_i(\mathbf{t}) = \sum_{m=1}^M \rho_{i,m} \psi_m(\mathbf{t}) + \varepsilon_i(\mathbf{t}), \quad \varepsilon_i(\mathbf{t}) \stackrel{\text{iid}}{\sim} N_p(0, \sigma^2 \mathbf{I}), \quad \mathbf{t} \in \mathcal{T}, \quad i = 1, \dots, N.$$

In each case, we consider data without ($\sigma^2 = 0$) and with ($\sigma^2 = 0.25$) measurement error. The scores $\rho_{i,m}$ are independent samples from $N(0, \nu_m)$ for eigenvalues with exponential ($\nu_m^{\text{exp}} = \exp(-(m+1)/2)$) or linear ($\nu_m^{\text{lin}} = (M+1-m)/M$) decrease, while the choice of \mathcal{T} , M and ψ_m varies between settings (see Sections 4.1 and 4.2). In all cases, we use unit weights ($w_j = 1$). The accuracy of the resulting estimates $\hat{\nu}_m$ and $\hat{\psi}_m$ is measured by the relative errors $\text{Err}(\hat{\nu}_m) = (\nu_m - \hat{\nu}_m)^2 / \nu_m^2$ and $\text{Err}(\hat{\psi}_m) = \left\| \psi_m - \hat{\psi}_m \right\|^2$. As functional principal components are defined only up to a sign change, the estimate $\hat{\psi}_m$ is reflected, i.e. multiplied by -1 , if $\langle \psi_m, \hat{\psi}_m \rangle < 0$. The goodness of the reconstructed observations $\hat{x}_i = \sum_{m=1}^M \hat{\rho}_{i,m} \hat{\psi}_m^{(j)}$ is evaluated by the mean relative squared error $\text{MRSE} = N^{-1} \sum_{i=1}^N \left(\left\| x_i - \hat{x}_i \right\|^2 / \left\| x_i \right\|^2 \right)$.

4.1 Multivariate Functional Data on One-Dimensional Domains

Setting 1: For the first setting, the first $M = 8$ Fourier basis functions on $[0, 2]$ are split into $p = 2$ parts. The pieces are shifted and multiplied by a random sign to form the elements $\psi_m^{(1)}$ and $\psi_m^{(2)}$ on $\mathcal{T}_1 = \mathcal{T}_2 = [0, 1]$ (for technical details, see online appendix). The observations x_i are sampled on an equispaced grid of $S_1 = S_2 = 100$ sampling points. The MFPCA is based on $M_1 = M_2 = 8$ univariate functional principal components that are calculated by the PACE algorithm (Yao et al., 2005)

with penalized splines to smooth the covariance function, as implemented in the R package **refund** (Crainiceanu et al., 2014). In this simple setting of a common, one-dimensional domain, the new approach can be compared to the method of Ramsay and Silverman (2005), which is implemented in the R package **fda** (Ramsay et al., 2014) and in the following denoted by MFPCA_{RS}. This method involves pre-smoothing of the elements with $K = 15$ cubic spline basis function. MFPCA_{RS} computes score values $\hat{\rho}_{i,m}^{(j)} = \langle x_i^{(j)}, \hat{\psi}_m^{(j)} \rangle_2$ for each observation i and each element j . Since they do not have the same interpretation as the scores in the multivariate Karhunen-Loève representation (Prop. 4), $\sum_{j=1}^p \hat{\rho}_{i,m}^{(j)} = \hat{\rho}_{i,m}$ is used for comparison purposes.

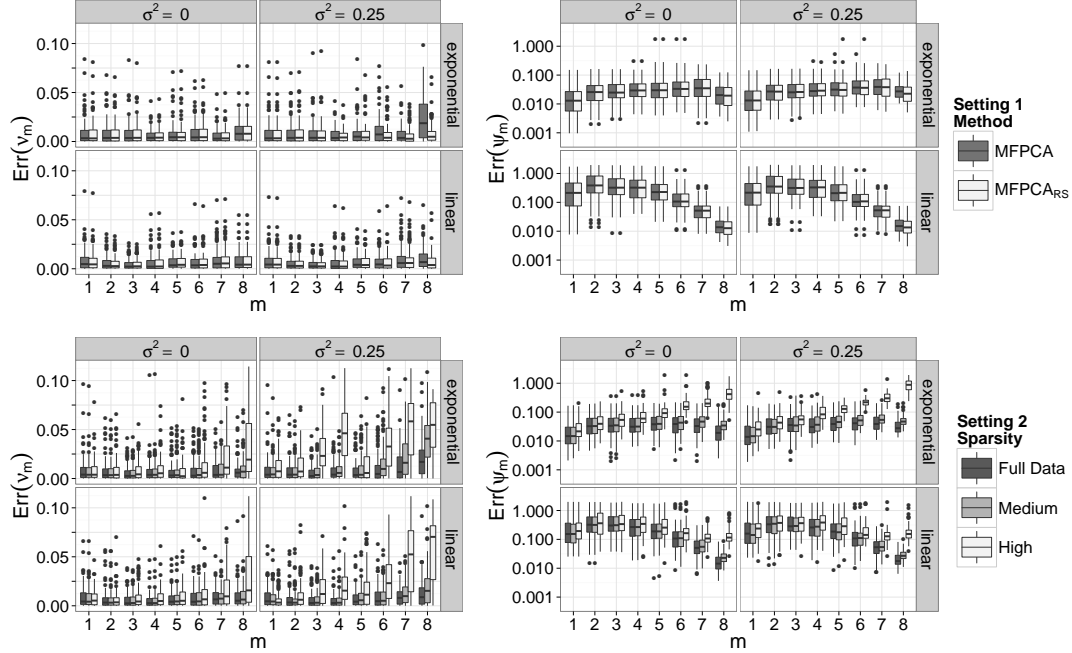


Figure 2: Relative errors for estimated eigenvalues (left) and eigenfunctions (right, log-scale) for simulation settings 1 and 2, depending on eigenvalue decrease, measurement error and estimation method (setting 1) or sparsity (setting 2).

The results for the first setting are shown in Fig. 2 and Table 1. In total, the new approach can compete very well with the existing method of Ramsay and Silverman and gives nearly identical results for synthetic and real data (see online appendix for the gait cycle example). Both techniques mostly have higher errors in ψ_m for linearly decreasing eigenvalues, as in these cases, the eigenfunctions are more often confused, i.e. $\hat{\psi}_m$ is an estimate for e.g. ψ_{m-1} or ψ_{m+1} rather than for ψ_m . In the ideal case of no measurement error, MFPCA_{RS} yields lower MRSE values than the new approach, which might be an effect of MFPCA_{RS} expecting smooth or presmoothed data. For the practically relevant case of data with measurement error, both methods give almost the same prediction errors (cf. Table 1). Simulations based on Legendre polynomials gave very similar results (not shown here).

Setting 2: Here we consider trivariate functional data on $\mathcal{T}_1 = [-1, 0.5]$, $\mathcal{T}_2 = [0, 1]$, $\mathcal{T}_3 = [1.5, 2]$. The eigenfunctions are constructed according to the same scheme as in setting 1 by splitting the

Table 1: Average MRSE (in %) for simulation setting 1 and 2, depending on eigenvalue decrease and measurement error.

Setting		$\sigma^2 = 0$		$\sigma^2 = 0.25$	
		ν_m^{exp}	ν_m^{lin}	ν_m^{exp}	ν_m^{lin}
1	MFPCA	0.006	0.009	0.740	0.355
1	MFPCA _{RS}	$< 10^{-3}$	$< 10^{-3}$	0.720	0.338
2	Full Data	0.004	0.007	0.778	0.367
2	Medium Sparsity	0.164	0.146	2.070	1.102
2	High Sparsity	5.755	4.568	15.365	10.824

first $M = 8$ Fourier basis functions on $[0, 2]$ into $p = 3$ parts, followed by a shift and multiplication with a random sign. The observations are sampled on equidistant grids with $S_1 = S_3 = 50$ and $S_2 = 100$ sampling points. We consider the dense observations as well as sparse variants with medium (50 – 70%) and high (90 – 95% missings) sparsity. The sparsification mechanism is analogous to Yao et al. (2005) and applied to each observation and each element separately. The MFPCA is calculated in the same way as in setting 1, using the PACE approach to estimate $M_1 = M_2 = M_3 = 8$ functional principal components for each element. For data with high sparsity, we set $M_1 = M_3 = 3$ and $M_2 = 5$ to make computation of the univariate FPCA feasible.

The results are given in Fig. 2 and Table 1. Here there is no available competitor. The performance of our MFPCA for full data is very similar to the simpler case of setting 1. Even for a moderate level of sparsity, the new method yields excellent results for most eigenvalues and eigenfunctions at the expense of somewhat higher reconstruction errors. For very sparse data, the leading eigenvalues and eigenfunctions are still estimated well, but the reconstruction error is considerably higher than for the full data. However, this is still acceptable (average MRSE is lower than 16% for all levels of sparsity), bearing in mind that data with high sparsity contains at most 10% of the original information. Again, simulations based on Legendre polynomials gave very similar results (not shown here).

4.2 Multivariate Functional Data Consisting of Functions and Images

Setting 3: Observations are generated based on $M = 25$ principal components, where the image elements $\psi_m^{(1)}$ are formed by tensor products of Fourier basis functions on $\mathcal{T}_1 = [0, 1] \times [0, 0.5]$ and $\psi_m^{(2)}$ are given by Legendre polynomials on $\mathcal{T}_2 = [-1, 1]$. The elements are weighted by random factors $\alpha^{1/2}$ and $(1 - \alpha)^{1/2}$, respectively, with $\alpha \in (0.2, 0.8)$ to ensure orthonormality. For the scores, only exponentially decreasing eigenvalues are used. The observations are discretized using a grid of $S_1 = 100 \times 50$ equidistant points for the image element and $S_2 = 200$ equidistant points for the functions.

We consider the new MFPCA approach based on univariate FPCA as well as non-orthogonal basis functions. In the first case, the eigendecomposition for the image data is calculated with the FCP-TPA algorithm for regularized tensor decomposition (Allen, 2013). The smoothing parameters for penalizing second differences in both image directions are chosen via generalized cross-validation in $[10^{-5}, 10^5]$ (Allen, 2013; Huang et al., 2009). Multivariate FPCA is calculated based on $M_1 = 20$ eigenimages and $M_2 = 15$ univariate eigenfunctions. In the case of general basis functions, image elements are expanded in tensor products of $K_1 = 10 \times 12$ B-splines and the one-dimensional element

is represented in terms of $K_2 = 15$ B-spline basis functions. In the presence of measurement error the univariate expansions are fit with appropriate smoothness penalties (Eilers and Marx, 1996).

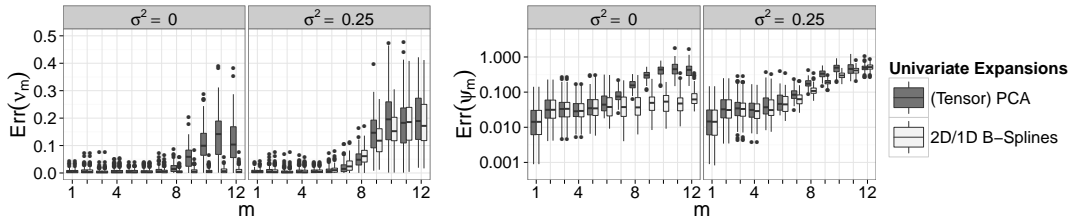


Figure 3: Relative errors for estimated eigenvalues (left) and eigenfunctions (right, log-scale) for simulation setting 3, depending on measurement error and univariate expansions.

The overall results for the first $M = 12$ eigenvalue/eigenvector pairs are given in Fig. 3. Compared to the settings with one-dimensional domain, the errors are slightly higher, in particular for higher order eigenvalues and eigenfunctions. Exemplary results however, show that even in this case, the new approach is still able to capture the important features of the true eigenfunctions well (see online appendix). The results further show that the general approach with spline basis functions performs mostly better than the pure MFPCA approach. Moreover, the truncated Karhunen-Loève representation with $M = 12$ (true $M = 25$) estimated eigenfunctions and scores gives an excellent reconstruction of the original data. The average MRSE is 1.382%/0.398% (PCA/splines) for data without measurement error and 2.233%/2.048% (PCA/splines) for data with measurement error.

5 Application – ADNI Study

In this section, the new method is applied to data from the Alzheimer’s Disease Neuroimaging Initiative study (ADNI), which aims at identifying biomarkers for accurate diagnosis of Alzheimer’s disease (AD) in an early stage (Mueller et al., 2005). We use MFPCA to explore how longitudinal trajectories of a neuropsychological score (ADAS-Cog, a current standard for monitoring AD progression) covary with FDG-PET scans at baseline. The latter are used to assess the glucose metabolism in the brain, which is tightly coupled with neuronal function. As the brain images might be predictive of subsequent cognitive decline, common patterns between these two sources of information would be highly relevant.

Dataset: The dataset considered for MFPCA contains data from all $N = 483$ participants enrolled in ADNI1, having an FDG-PET scan at baseline and at least three ADAS-Cog measurements during follow-up. At baseline, 84 subjects were diagnosed with AD, 302 were suffering from mild cognitive impairment (MCI, in many cases an early stage of AD) and 97 were cognitively healthy elderly controls. The ADAS-Cog trajectories constitute the first element $X^{(1)}$, where high values indicate a high level of cognitive impairment. The measurements contain missings, mostly in the second half of the study period and thus are sparse. The second element $X^{(2)}$ is an axial slice of 93×117 pixels (139.5×175.5 mm²) of FDG-PET scans, containing the Precuneus and temporo-parietal regions. Both are believed to show a strong relation between hypometabolism (reduced brain function) and AD (Blennow et al., 2006). Exemplary data is shown in Fig. 4.

Weighted scalar product: As the ADAS-Cog trajectories and FDG-PET scans differ consid-

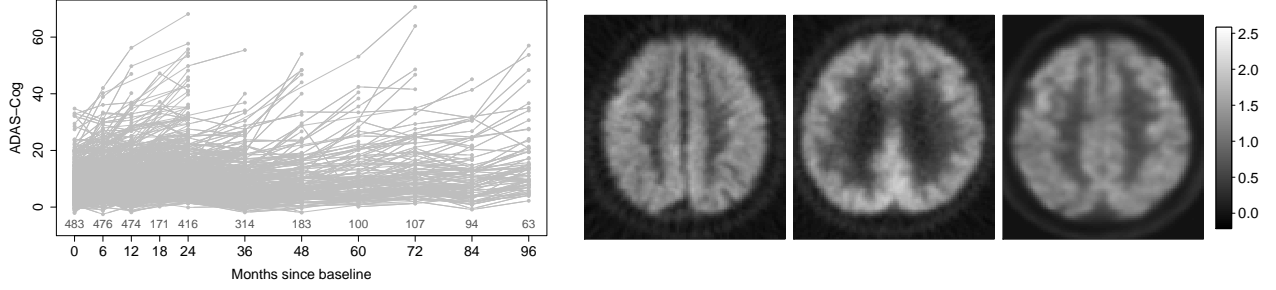


Figure 4: Left: ADAS-Cog trajectories for all $N = 483$ subjects. Numbers above the x-axis give the total number of measurements for each visit. Right: FDG-PET scans for three randomly chosen male subjects (left to right: AD, MCI, normal; diagnosis at baseline).

erably in domain, range and variation (cf. Fig. 4), we use a weighted MFPCA with

$$w_j = \left(\int_{\mathcal{T}_j} \hat{C}_{jj}(t_j, t_j) dt_j \right)^{-1} = \left(\int_{\mathcal{T}_j} \widehat{\text{Var}} \left(X^{(j)}(t_j) \right) dt_j \right)^{-1}, \quad j = 1, 2,$$

where \hat{C}_{jj} is estimated from the data. Using these weights, the integrated variance equals 1 for the rescaled elements $\tilde{X}^{(j)} = w_j^{1/2} X^{(j)}$. All elements thus contribute equal amounts of variation to the analysis, similarly to multivariate PCA, where the data is usually standardized before the analysis. We believe that this a sensible choice for many applications, but there may of course be situations, in which other weighting schemes may be preferable. For example, one could think of data that has two image elements, representing brain regions of different size for the same imaging modality. Here variability is naturally on the same scale and it might be better to keep the information of the size of the individual domains by setting both weights to one. On the other hand, if the images stem from different imaging modalities on the same domain, it might be necessary to correct solely for differences in variation. As a general rule, the weights should be chosen in close coordination with practitioners, considering the objective of the analysis and the data at hand.

Results: The results for the first two multivariate functional principal components, that account for 80.7% of the total weighted variance, are shown in Fig. 5. For the univariate expansions, we use FPCA for $X^{(1)}$ with $M_1 = 3$ principal components (explaining 99.2% of the univariate variance) and 20×15 tensor product B-splines for the images $X^{(2)}$. Fig. 5 further includes pointwise bootstrap confidence bands for the principal components based on 100 nonparametric bootstrap iterations on the level of subjects. The coverage of such confidence bands for data consisting of functions and images has been analyzed in a simulation study, which gave good results, even in the presence of measurement error (see online appendix). The entire analysis for the ADNI data took around 15 minutes on a standard laptop (2.7 GHz, 16 GB RAM) including the calculation of the bootstrap confidence bands and without parallelization.

Almost half of the variability in the data (46.7% of the weighted variance) is explained by the first functional principal component. The ADAS-Cog element – and hence the degree of cognitive impairment – is elevated relative to the mean and increases during follow-up. The FDG-PET element exhibits hypometabolism in the Precuneus and the temporo-parietal regions, i.e. this component reflects reduced brain activity in these regions already at baseline. In total, the first eigenfunction seems to be interpretable as an AD related effect, as the pattern for positive scores perfectly agrees

with medical knowledge about AD progression. This interpretation is supported by the estimated scores, which are mainly positive for people diagnosed with AD by their last visit, while scores of subjects who remained cognitively normal during follow-up are nearly all negative. Persons with MCI have intermediate score values, which is in line with the hypothesis that this diagnosis can constitute a transitional phase between normal ageing and AD.

For the second functional principal component (explains 33.9% of weighted variance), the ADAS-Cog element is nearly constant and has wide bootstrap confidence bands that include zero during the whole follow-up. In contrast, the FDG-PET element differs significantly from zero in almost all voxels (cf. Fig. 5). Hence, this principal component reflects variation in the FDG-PET scans at baseline. Plotting the overall mean plus or minus this component suggests that it can be interpreted as an effect of imperfect registration that manifests in different brain sizes, which are known to correlate with gender (Ruigrok et al., 2014). This hypothesis is supported by the boxplots of the estimated scores in Fig. 5, while scores do not differ notably by diagnosis (not shown here).

Discussion: The results show that MFPCA is able to capture important sources of variation in the data that have a meaningful interpretation from a medical and neuroimaging point of view. An important issue not addressed here is that for ADAS-Cog, there may well be an informative dropout of patients with high score values (cf. Fig. 4). While addressing informative missingness goes beyond the scope of this paper, interpretation of results should take this possibility into account. For instance, it is easily conceivable that $\hat{\psi}_1^{(1)}$ may be underestimating $\psi_1^{(1)}$ towards the end of the study period.

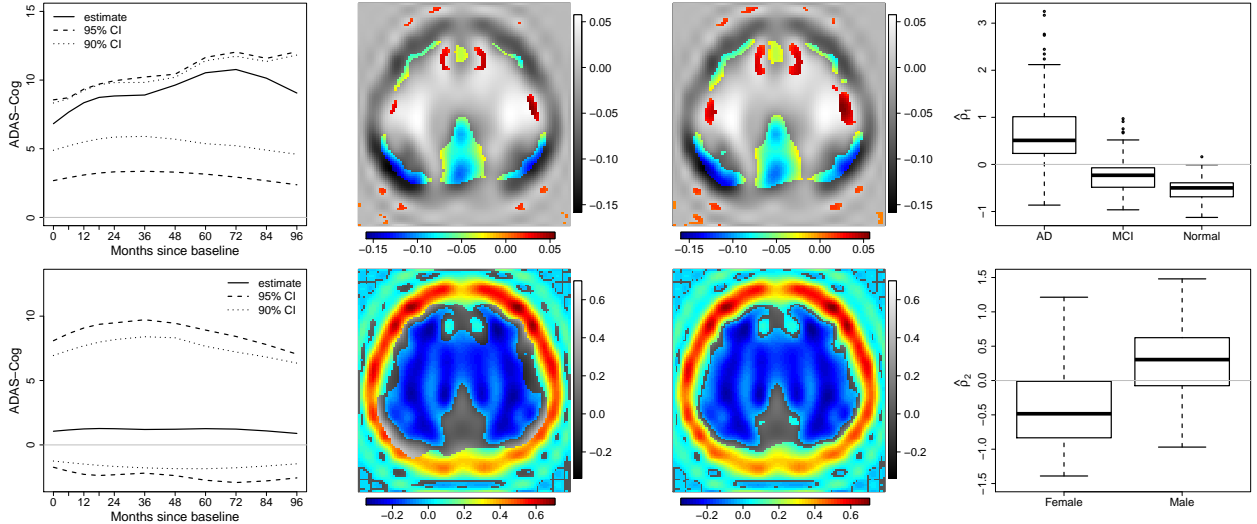


Figure 5: The first two estimated multivariate functional principal components for the ADNI data (1st row: $\hat{\psi}_1$, 2nd row: $\hat{\psi}_2$). Estimates are given with pointwise 95% and 90% bootstrap confidence bands based on 100 nonparametric bootstrap iterations (ADAS-Cog, 1st column: Dashed lines; FDG-PET, 2nd and 3rd column: Pixels with pointwise 95% (left) and 90% (right) confidence bands not including zero in color). Boxplots of the scores (4th column) support the interpretation.

6 Discussion and Outlook

This paper introduces methodology and a practical estimation algorithm for multivariate functional principal component analysis. While other methods for MFPCA are restricted to observations on a common, one-dimensional interval, the new approach is suitable for data on different domains, which may also differ in dimension, such as functions and images. The key results are 1. a Karhunen-Loève Theorem, that establishes the theoretical basis for MFPCA (Prop. 4), 2. an explicit relation between multivariate and univariate FPCA, which serves as a starting point for the estimation (Prop. 5) and 3. asymptotic results for the estimators (Prop. 6 and 7). The estimation algorithm can be extended to expansions of the univariate elements in not necessarily orthonormal bases. This allows to flexibly choose an appropriate basis for each element depending on the data structure, in particular also mixtures of univariate FPCA and general bases. The algorithm is applicable to sparse data or data with measurement error, as well as to images. Notably, the proposed method can be used to calculate smooth univariate functional principal components for data on higher dimensional domains and is hence an alternative to existing methods for tensor PCA (Allen, 2013). The results of MFPCA give insights into simultaneous variation within the data and provide a natural tool for dimension reduction. Moreover, they can be used as a building block for further statistical analyses such as functional clustering methods or functional principal component regression with multiple covariates (cf. Müller and Stadtmüller, 2005, for the univariate case). If the elements differ in domain, range or variation, the new method can incorporate weights, which should be chosen with respect to the question of interest and the data at hand.

Possible extensions of the approach include normalization methods as an alternative to the weighted scalar product, following the ideas in Jacques and Preda (2014) or Chiou et al. (2014) for functions observed on a common interval. However, one should take into account that the domains may have different dimensions and sizes. The concept of MFPCA could further be extended to *hybrid data*, i.e. data consisting of a functional and a vector part (Ramsay and Silverman, 2005, Chapter 10.3.). A natural starting point would be to extend the scalar product suggested by Ramsay and Silverman (2005) in this context to multivariate functional data as proposed in Prop. 1. However, transferring the results for MFPCA shown in this paper requires a careful revision of the concept of the covariance operator and related proofs. Finally, one could think of estimating the multivariate covariance operator directly without computing a univariate decomposition for each element. This operator is typically high-dimensional, making smoothing as well as an eigendecomposition hardly feasible, which is avoided in our two-step approach.

Supplementary Material

The online appendix contains detailed proofs for all propositions, some additional simulation results and R code for reproducing the analysis for the ADNI and gait cycle data based on the R packages `fundata` and `MFPCA` (Happ, 2016a,b).

References

- Allen, G. (2013). Multi-way functional principal components analysis. In *IEEE 5th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 220–223.
- Berrendero, J., Justel, A., and Svarc, M. (2011). Principal components for multivariate functional data. *Computational Statistics & Data Analysis*, 55(9):2619–2634.
- Blennow, K., Leon, M., and Zetterberg, H. (2006). Alzheimers disease. *Lancet*, 368:387–403.
- Bosq, D. (2000). *Linear Processes in Function Spaces*. Springer, New York.
- Carroll, J. D. and Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 35(3):283–319.
- Cattell, R. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1(2):245–276.
- Chiou, J.-M., Yang, Y.-F., and Chen, Y.-T. (2014). Multivariate functional principal component analysis: A normalization approach. *Statistica Sinica*, 24:1571–1596.
- Coppi, R. and Bolasco, S., editors (1989). *Multiway data analysis*. North-Holland, Amsterdam.
- Crainiceanu, C., Reiss, P., Goldsmith, J., Huang, L., Huo, L., and Scheipl, F. (2014). *refund: Regression with Functional Data*. R package version 0.1-14.
- Eilers, P. and Marx, B. (1996). Flexible Smoothing with B-splines and Penalties. *Statistical Science*, 11(2):89–121.
- Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *Annals of Statistics*, 35(1):70–91.
- Hall, P. and Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):109–126.
- Happ, C. (2016a). *funData: An S4 class for functional data in R*. R package version 1.0.
- Happ, C. (2016b). *MFPCA: Multivariate Functional Principal Component Analysis*. R package version 1.0-1.
- Horn, R. A. and Johnson, C. R. (1991). *Topics in matrix analysis*. Cambridge Univ. Press, Cambridge.
- Huang, J. Z., Shen, H., and Buja, A. (2009). The Analysis of Two-Way Functional Data Using Two-Way Regularized Singular Value Decompositions. *Journal of the American Statistical Association*, 104(488):1609–1620.
- Jacques, J. and Preda, C. (2014). Model-based clustering for multivariate functional data. *Computational Statistics & Data Analysis*, 71:92–106.

- Kato, T. (1976). *Perturbation Theory for Linear Operators*. Springer, Berlin, 2. edition.
- Lu, H., Plataniotis, K. N., and Venetsanopoulos, A. (2013). *Multilinear Subspace Learning: Dimensionality Reduction of Multidimensional Data*. CRC press, Boca Raton.
- Mercer, J. (1909). Functions of Positive and Negative Type, and their Connection with the Theory of Integral Equations. *Philosophical Transactions A*, 209(441-458):415–446.
- Mueller, S., Weiner, M., Thal, L., Petersen, R., Jack, C., Jagust, W., Trojanowski, J., Toga, A., and Beckett, L. (2005). Ways toward an early diagnosis in Alzheimer’s disease: The Alzheimer’s Disease Neuroimaging Initiative (ADNI). *Alzheimer’s & Dementia*, 1(1):55–66.
- Müller, H. G. and Stadtmüller, U. (2005). Generalized functional linear models. *Annals of Statistics*, 33(2):774–805.
- Munkres, J. (2000). *Topology*. Prentice-Hall, Upper Saddle River, NJ, 2. edition.
- Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis*. Springer, New York, 2. edition.
- Ramsay, J., Wickham, H., Graves, S., and Hooker, G. (2014). *fda: Functional Data Analysis*. R package version 2.4.4.
- Reed, M. and Simon, B. (1980). *Methods of Modern Mathematical Physics. I: Functional Analysis*. Academic Press, San Diego, rev. and enl. edition.
- Rudin, W. (1987). *Real and complex analysis*. McGraw-Hill, New York, 3. edition.
- Ruigrok, A., Salimi-Khorshidi, G., Lai, M.-C., Baron-Cohen, S., Lombardo, M., Tait, R., and Suckling, J. (2014). A meta-analysis of sex differences in human brain structure. *Neuroscience and biobehavioral reviews*, 39:34–50.
- Saporta, G. (1981). *Méthodes exploratoires d’analyse de données temporelles*. PhD thesis, Université Pierre et Marie Curie, Paris.
- Tucker, L. R. (1966). Some Mathematical Notes on Three-Mode Factor Analysis. *Psychometrika*, 31(3):279–311.
- Werner, D. (2011). *Funktionalanalysis*. Springer, Berlin, 7. edition.
- Yao, F., Müller, H.-G., and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590.
- Yu, Y., Wang, T., and Samworth, R. J. (2015). A useful variant of the Davis-Kahan theorem for statisticians. *Biometrika*, 102(2):315–323.
- Zemyan, S. (2012). *The Classical Theory of Integral Equations*. Birkhäuser, Basel.

Proofs of Propositions

Proof of Prop. 1. \mathcal{H} is a direct sum of the Hilbert spaces $L^2(\mathcal{T}_j)$, $j = 1, \dots, p$, with natural scalar product $\langle\langle \cdot, \cdot \rangle\rangle$ (cf. Reed and Simon, 1980, Chapter II.1.) \square

Proof of Prop. 2.

1. Γ is linear: Follows from the linearity of the scalar product in (3).
2. Γ is self-adjoint: Follows from the symmetry $C_{ij}(s_i, t_j) = C_{ji}(t_j, s_i)$.
3. Γ is positive: Let $f \in \mathcal{H}$. Then

$$\begin{aligned} \langle\langle f, \Gamma f \rangle\rangle &= \sum_{j=1}^p \int_{\mathcal{T}_j} f^{(j)}(t_j) \sum_{i=1}^p \int_{\mathcal{T}_i} \mathbb{E} \left(X^{(i)}(s_i) X^{(j)}(t_j) \right) f^{(i)}(s_i) ds_i dt_j \\ &= \mathbb{E} \left(\sum_{j=1}^p \int_{\mathcal{T}_j} f^{(j)}(t_j) X^{(j)}(t_j) dt_j \right)^2 \geq 0. \end{aligned}$$

4. Γ is compact: Let $\mathcal{B} := \{f \in \mathcal{H} : \|f\|^2 \leq B\}$ be a bounded family in \mathcal{H} for some constant $0 < B < \infty$. Clearly, $\|f^{(j)}\|_2^2 \leq B$ for all $j = 1, \dots, p$. Define the image of \mathcal{B} under Γ by $\mathcal{Z} = \Gamma\mathcal{B} = \{g \in \mathcal{H} : \exists f \in \mathcal{B} \text{ such that } g = \Gamma f\}$, which has the following properties:

- \mathcal{Z} is uniformly bounded: Let $g \in \mathcal{Z}$ and $\mathbf{t} \in \mathcal{T}$. Define $K := \max_{i,j=1,\dots,p} K_{ij}$ with K_{ij} as in (5). Then

$$\begin{aligned} \|g(\mathbf{t})\|^2 &\leq \sum_{j=1}^p \left(\sum_{i=1}^p \int_{\mathcal{T}_i} |C_{ij}(s_i, t_j) f^{(i)}(s_i)| ds_i \right)^2 \\ &\stackrel{\text{H\"older}}{\leq} \sum_{j=1}^p \left(\sum_{i=1}^p \left[\int_{\mathcal{T}_i} C_{ij}(s_i, t_j)^2 ds_i \right]^{1/2} \left[\int_{\mathcal{T}_i} f^{(i)}(s_i)^2 ds_i \right]^{1/2} \right)^2 \\ &\leq \sum_{j=1}^p \left(\sum_{i=1}^p K^{1/2} B^{1/2} \right)^2 = p^3 K B < \infty. \end{aligned}$$

- \mathcal{Z} is equicontinuous: Denote by $\lambda(\mathcal{T}_j)$ the Lebesgue measure of \mathcal{T}_j and let $T = \max_{j=1,\dots,p} \lambda(\mathcal{T}_j)$. For $\varepsilon > 0$, define $\tilde{\varepsilon} := \frac{\varepsilon}{p(pTB)^{1/2}}$. By the continuity assumption for $C_{ij}(s_i, \cdot)$, there exist $\delta_{ij} > 0$ such that

$$\|t_j - t_j^*\| < \delta_{ij} \quad \Rightarrow \quad |C_{ij}(s_i, t_j) - C_{ij}(s_i, t_j^*)| < \tilde{\varepsilon}, \quad \forall s_i \in \mathcal{T}_i. \quad (12)$$

for all $i, j = 1, \dots, p$. Set $\delta := \min_{i,j=1,\dots,p} \delta_{ij}$ and let $\|\mathbf{t} - \mathbf{t}^*\|_{\mathcal{T}} < \delta$. Clearly, $\|t_j - t_j^*\| < \delta$

for all $j = 1, \dots, p$ and for $g \in \mathcal{Z}$ it holds

$$\begin{aligned}
\|g(\mathbf{t}) - g(\mathbf{t}^*)\|^2 &\leq \sum_{j=1}^p \left(\sum_{i=1}^p \left| \int_{\mathcal{T}_i} (C_{ij}(s_i, t_j) - C_{ij}(s_i, t_j^*)) f^{(i)}(s_i) ds_i \right| \right)^2 \\
&\stackrel{\text{H\"older}}{\leq} \sum_{j=1}^p \left(\sum_{i=1}^p \left[\int_{\mathcal{T}_i} (C_{ij}(s_i, t_j) - C_{ij}(s_i, t_j^*))^2 ds_i \right]^{1/2} \left[\int_{\mathcal{T}_i} f^{(i)}(s_i)^2 ds_i \right]^{1/2} \right)^2 \\
&\stackrel{(12)}{<} \sum_{j=1}^p \left(\sum_{i=1}^p \left(\int_{\mathcal{T}_i} \tilde{\varepsilon}^2 ds_i \right)^{1/2} \|f^{(i)}\|_2 \right)^2 \leq p^3 T B \tilde{\varepsilon}^2 = \varepsilon^2.
\end{aligned}$$

By the Theorem of Arzelà-Ascoli (Reed and Simon, 1980, Thm. I.28. and related notes for Chapter I), for each sequence $\{f_n\}_{n \in \mathbb{N}}$ in \mathcal{B} there exists a convergent subsequence $\{g_{n(i)} = \Gamma f_{n(i)}\}_{i \in \mathbb{N}}$ of the corresponding sequence $\{g_n\}_{n \in \mathbb{N}}$ in \mathcal{Z} , which implies that Γ is a compact operator (cf. Reed and Simon, 1980, Chapter VI.5.).

□

Lemma 1. For fixed $m \in \mathbb{N}$ and $j \in \{1, \dots, p\}$, the j -th element $\psi_m^{(j)}$ of the eigenfunction ψ_m is continuous, if $\nu_m > 0$ and C_{ij} is uniformly continuous as in Prop. 2.

Proof. Let $\varepsilon > 0$ and $\tilde{\varepsilon} := \varepsilon \nu_m \left(2 \sum_{j=1}^p \lambda(\mathcal{T}_j)^{1/2} \right)^{-1}$. By the uniform continuity assumption for C_{ij} , there exist $\delta_{ij} > 0$ such that for all $i = 1, \dots, p$ (12) holds. Let $\delta_j = \min_{i=1, \dots, p} \delta_{ij}$ and $\|t_j - t_j^*\| < \delta_j$. Then, as $\nu_m > 0$ and $\|\psi_m^{(i)}\|_2 \leq \|\psi_m\| = 1$,

$$\begin{aligned}
\left| \psi_m^{(j)}(t_j) - \psi_m^{(j)}(t_j^*) \right| &= \frac{1}{\nu_m} \left| \sum_{i=1}^p \int_{\mathcal{T}_i} [C_{ij}(s_i, t_j) - C_{ij}(s_i, t_j^*)] \psi_m^{(i)}(s_i) ds_i \right| \\
&\leq \frac{1}{\nu_m} \tilde{\varepsilon} \sum_{i=1}^p \int_{\mathcal{T}_i} |\psi_m^{(i)}(s_i)| ds_i \stackrel{\text{H\"older}}{\leq} \frac{\tilde{\varepsilon}}{\nu_m} \sum_{i=1}^p \|\psi_m^{(i)}\|_2 \lambda(\mathcal{T}_i)^{1/2} \\
&\leq \frac{\tilde{\varepsilon}}{\nu_m} \sum_{i=1}^p \lambda(\mathcal{T}_i)^{1/2} = \frac{\varepsilon}{2} < \varepsilon.
\end{aligned}$$

□

Proof of Prop. 3. The proof follows the idea in Werner (2011, Chapter VI.4.) for the proof of Mercer's Theorem in the univariate case. From the Spectral Theorem for compact self-adjoint operators (Werner, 2011, Thm. VI.3.2.), it is known that

$$\Gamma f = \sum_{m=1}^{\infty} \nu_m \langle f, \psi_m \rangle \psi_m \quad \forall f \in \mathcal{H}.$$

For $M \in \mathbb{N}$ define

$$\Gamma_M f := \sum_{m=1}^M \nu_m \langle f, \psi_m \rangle \psi_m \quad \forall f \in \mathcal{H}.$$

Then for all $f \in \mathcal{H}$: $\langle \Gamma f, f \rangle - \langle \Gamma_M f, f \rangle = \sum_{m=M+1}^{\infty} \nu_m \langle f, \psi_m \rangle^2 \geq 0$. Let $j \in \{1, \dots, p\}$ and $t^* \in \mathcal{T}_j$. Define $f = (0, \dots, 0, f^{(j)}, 0, \dots, 0)$ with $f^{(j)} = \lambda \left(B_{t^*} \left(\frac{1}{n} \right) \right)^{-1} \mathbf{1}_{B_{t^*} \left(\frac{1}{n} \right)}$ for some $n \in \mathbb{N}$, where

$B_{t^*}(\frac{1}{n})$ is a closed ball in \mathcal{T}_j with center t^* and radius $\frac{1}{n}$, $\lambda(\cdot)$ denotes the Lebesgue measure and $\mathbf{1}$ is the indicator function. Clearly, $f^{(j)} \in L^2(\mathcal{T}_j)$ and $f \in \mathcal{H}$. Therefore

$$\begin{aligned} 0 &\leq \langle \Gamma f, f \rangle - \langle \Gamma_M f, f \rangle \\ &= \lambda(B_{t^*}(\frac{1}{n}))^{-2} \int_{B_{t^*}(\frac{1}{n})} \int_{B_{t^*}(\frac{1}{n})} C_{jj}(s_j, t_j) - \sum_{m=1}^M \nu_m \psi_m^{(j)}(s_j) \psi_m^{(j)}(t_j) ds_j dt_j \\ &\rightarrow C_{jj}(t^*, t^*) - \sum_{m=1}^M \nu_m \psi_m^{(j)}(t^*) \psi_m^{(j)}(t^*) \quad \text{for } n \rightarrow \infty \end{aligned}$$

by the Lebesgue Differentiation Theorem (Rudin, 1987, Thm. 7.10.). As t^* was arbitrary in \mathcal{T}_j , this implies that for all $M \in \mathbb{N}$

$$\sum_{m=1}^M \nu_m \psi_m^{(j)}(t)^2 \leq C_{jj}(t, t) \leq \|C_{jj}\|_\infty < \infty \quad \forall t \in \mathcal{T}_j,$$

since C_{jj} is continuous and \mathcal{T}_j is compact, implying that $\|C_{jj}\|_\infty := \sup_{t \in \mathcal{T}_j} |C_{jj}(t, t)|$ is finite. Using Hölder's inequality

$$\sum_{m=1}^\infty \left| \nu_m \psi_m^{(j)}(s) \psi_m^{(j)}(t) \right| \leq C_{jj}(s, s)^{1/2} C_{jj}(t, t)^{1/2} < \infty,$$

i.e. the series $\tilde{C}_j(s, t) := \sum_{m=1}^\infty \nu_m \psi_m^{(j)}(s) \psi_m^{(j)}(t)$ is absolutely convergent for all $s, t \in \mathcal{T}_j$. In the following, assume $t \in \mathcal{T}_j$ to be fixed. For $\varepsilon > 0$ choose $M \in \mathbb{N}$ such that $\sum_{m=M+1}^\infty \nu_m \psi_m^{(j)}(t)^2 < \varepsilon^2$. Then, again by Hölder's inequality

$$\sum_{m=M+1}^\infty \left| \nu_m \psi_m^{(j)}(s) \psi_m^{(j)}(t) \right| \leq C_{jj}(s, s)^{1/2} \cdot \varepsilon \leq \|C_{jj}\|_\infty^{1/2} \cdot \varepsilon. \quad (13)$$

The upper bound in (13) does not depend on s , hence $\tilde{C}_j(s, t)$ converges uniformly for fixed t . As the eigenfunctions $\psi_m^{(j)}(s)$ are continuous in s for all $m \in \mathbb{N}$ (Lemma 1), $\tilde{C}_j(s, t)$ is also continuous in s (Uniform Limit Theorem, Munkres, 2000, Thm. 21.6.). Define

$$h_j(s) := C_{jj}(s, t) - \tilde{C}_j(s, t), \quad s \in \mathcal{T}_j.$$

Let now $g^{(j)} \in L^2(\mathcal{T}_j)$ and define $g := (0, \dots, 0, g^{(j)}, 0, \dots, 0)$, which is clearly in \mathcal{H} . Therefore

$$\begin{aligned} \int_{\mathcal{T}_j} h_j(s) g^{(j)}(s) ds &= \sum_{i=1}^p \int_{\mathcal{T}_i} C_{ij}(s_i, t) g^{(i)}(s_i) ds_i - \sum_{m=1}^\infty \nu_m \sum_{i=1}^p \int_{\mathcal{T}_i} \psi_m^{(i)}(s_i) g^{(i)}(s_i) ds_i \psi_m^{(j)}(t) \\ &= (\Gamma g)^{(j)}(t) - \sum_{m=1}^\infty \nu_m \langle \psi_m, g \rangle \psi_m^{(j)}(t) \\ &= \sum_{m=1}^\infty \nu_m \langle g, \psi_m \rangle \psi_m^{(j)}(t) - \sum_{m=1}^\infty \nu_m \langle \psi_m, g \rangle \psi_m^{(j)}(t) = 0 \end{aligned}$$

according to the Spectral Theorem. Choosing $g^{(j)} = h_j$ implies $h_j(s) = 0$ for all $s \in \mathcal{T}_j$, as h_j is continuous in s . Therefore,

$$\tilde{C}_j(s, t) = \sum_{m=1}^\infty \nu_m \psi_m^{(j)}(s) \psi_m^{(j)}(t) = C_{jj}(s, t) \quad \forall s \in \mathcal{T}_j.$$

By Dini's Theorem (Werner, 2011, Thm. VI.4.6.) the series $C_{jj}(t, t) = \sum_{m=1}^\infty \nu_m \psi_m^{(j)}(t)^2$ converges uniformly. Hence, M can be chosen independent of t in (13). This implies that $\tilde{C}_j(s, t)$ converges absolutely and uniformly to $C_{jj}(s, t)$ for all $s, t \in \mathcal{T}_j$. \square

Proof of Prop. 4. By the Hilbert-Schmidt Theorem (Reed and Simon, 1980, Thm. VI.16.), the (deterministic) eigenfunctions of Γ form an orthonormal basis of \mathcal{H} , i.e. X can be written in the form $X(\mathbf{t}) = \sum_{m=1}^{\infty} \rho_m \psi_m(\mathbf{t})$, $\mathbf{t} \in \mathcal{T}$ with random variables $\rho_m = \langle\langle X, \psi_m \rangle\rangle$. Hence for $m, n \in \mathbb{N}$

$$1. \mathbb{E}(\rho_m) = \sum_{j=1}^p \int_{\mathcal{T}_j} \mathbb{E}(X^{(j)}(t_j)) \psi_m^{(j)}(t_j) dt_j = 0, \text{ since } \mathbb{E}(X^{(j)}(t_j)) = 0 \text{ for all } t_j \in \mathcal{T}_j, \\ j = 1, \dots, p \text{ by assumption.}$$

$$2. \text{Cov}(\rho_m, \rho_n) = \mathbb{E} \left(\sum_{i=1}^p \int_{\mathcal{T}_i} X^{(i)}(s_i) \psi_m^{(i)}(s_i) ds_i \cdot \sum_{j=1}^p \int_{\mathcal{T}_j} X^{(j)}(t_j) \psi_n^{(j)}(t_j) dt_j \right) \\ = \sum_{j=1}^p \int_{\mathcal{T}_j} \sum_{i=1}^p \int_{\mathcal{T}_i} C_{ij}(s_i, t_j) \psi_m^{(i)}(s_i) ds_i \psi_n^{(j)}(t_j) dt_j \\ = \sum_{j=1}^p \int_{\mathcal{T}_j} \nu_m \psi_m^{(j)}(t_j) \psi_n^{(j)}(t_j) dt_j = \nu_m \delta_{mn}.$$

3. Let $X_{[M]}(\mathbf{t}) := \sum_{m=1}^M \rho_m \psi_m(\mathbf{t}) = \sum_{m=1}^M \left[\sum_{i=1}^p \int_{\mathcal{T}_i} X^{(i)}(s_i) \psi_m^{(i)}(s_i) ds_i \right] \psi_m(\mathbf{t})$ for $\mathbf{t} \in \mathcal{T}$ be the truncated Karhunen-Loève representation of X . Then

$$\mathbb{E}(\|X(\mathbf{t}) - X_{[M]}(\mathbf{t})\|^2) = \sum_{j=1}^p \left[\mathbb{E}(X^{(j)}(t_j)^2) - 2\mathbb{E}(X^{(j)}(t_j) S_M^{(j)}(t_j)) + \mathbb{E}(S_M^{(j)}(t_j)^2) \right] \\ = \sum_{j=1}^p \left[C_{jj}(t_j, t_j) - 2 \sum_{m=1}^M \sum_{i=1}^p \int_{\mathcal{T}_i} C_{ij}(s_i, t_j) \psi_m^{(i)}(s_i) ds_i \psi_m^{(j)}(t_j) \right. \\ \left. + \sum_{m=1}^M \sum_{n=1}^M \sum_{i=1}^p \int_{\mathcal{T}_i} \sum_{k=1}^p \int_{\mathcal{T}_k} C_{ki}(u_k, s_i) \psi_n^{(k)}(u_k) du_k \psi_m^{(i)}(s_i) ds_i \psi_m^{(j)}(t_j) \psi_n^{(j)}(t_j) \right] \\ = \sum_{j=1}^p \left[C_{jj}(t_j, t_j) - 2 \sum_{m=1}^M \nu_m \psi_m^{(j)}(t_j) \psi_m^{(j)}(t_j) \right. \\ \left. + \sum_{m=1}^M \sum_{n=1}^M \sum_{i=1}^p \int_{\mathcal{T}_i} \nu_n \psi_n^{(i)}(s_i) \psi_m^{(i)}(s_i) ds_i \psi_m^{(j)}(t_j) \psi_n^{(j)}(t_j) \right] \\ = \sum_{j=1}^p \left[C_{jj}(t_j, t_j) - \sum_{m=1}^M \nu_m \psi_m^{(j)}(t_j)^2 \right] \rightarrow 0 \quad \text{for } M \rightarrow \infty$$

uniformly for $\mathbf{t} \in \mathcal{T}$ by Prop. 3.

□

Proof of Prop. 5.

1. Let X have a finite Karhunen-Loève representation (7). Then, each element is given by $X^{(j)} = \sum_{m=1}^M \rho_m \psi_m^{(j)}$. For $t \in \mathcal{T}_j$

$$\left(\Gamma^{(j)} \tilde{\phi}^{(j)} \right) (t) = \int_{\mathcal{T}_j} \text{Cov}(X^{(j)}(s), X^{(j)}(t)) \tilde{\phi}^{(j)}(s) ds \\ = \int_{\mathcal{T}_j} \sum_{m=1}^M \nu_m \psi_m^{(j)}(s) \psi_m^{(j)}(t) \tilde{\phi}^{(j)}(s) ds \stackrel{!}{=} \lambda^{(j)} \tilde{\phi}^{(j)}(t), \quad (14)$$

which is a homogenous Fredholm integral equation of the second kind with separable kernel function $K(s, t) = \sum_{m=1}^M \nu_m \psi_m^{(j)}(s) \psi_m^{(j)}(t) = \sum_{m=1}^M a_m(s) b_m(t)$ with continuous functions $a_m(s) = \nu_m^{1/2} \psi_m^{(j)}(s)$, $b_m(t) = \nu_m^{1/2} \psi_m^{(j)}(t)$ (cf. Lemma 1). Following the argumentation in Zemyan (2012, Chapter 1.3.), (14) can be transformed into the matrix eigenequation

$$\mathbf{A}^{(j)} \mathbf{u} = \lambda^{(j)} \mathbf{u}$$

with a symmetric matrix $\mathbf{A}^{(j)} \in \mathbb{R}^{M \times M}$ given by $A_{mn}^{(j)} = \langle a_m, b_n \rangle_2$. Positivity of $\Gamma^{(j)}$ implies $\lambda^{(j)} \geq 0$ and therefore $\mathbf{A}^{(j)}$ is positive semidefinite. Hence it can have at most M strictly positive eigenvalues $\lambda_m^{(j)}$ associated with eigenvectors $\mathbf{u}_m^{(j)} \in \mathbb{R}^M$

$$\mathbf{A}^{(j)} \mathbf{u}_m^{(j)} = \lambda_m^{(j)} \mathbf{u}_m^{(j)}.$$

Let $\lambda_1^{(j)} \geq \dots \geq \lambda_{M_j}^{(j)} > 0$, $M_j \leq M$ be the non-zero eigenvalues of $\mathbf{A}^{(j)}$. They are at the same time the only non-zero eigenvalues of $\Gamma^{(j)}$. The associated eigenfunctions $\tilde{\phi}_m^{(j)}$ of $\Gamma^{(j)}$ are parametrized by the eigenvectors $\mathbf{u}_m^{(j)}$ via (Zemyan, 2012, Chapter 1.3.)

$$\Gamma^{(j)} \tilde{\phi}_m^{(j)} = \sum_{n=1}^M \nu_n^{1/2} \psi_n^{(j)} [\mathbf{u}_m^{(j)}]_n = \lambda_m^{(j)} \tilde{\phi}_m^{(j)} \Leftrightarrow \tilde{\phi}_m^{(j)} = \left(\lambda_m^{(j)} \right)^{-1} \sum_{n=1}^M \nu_n^{1/2} \psi_n^{(j)} [\mathbf{u}_m^{(j)}]_n.$$

Since

$$\langle \tilde{\phi}_m^{(j)}, \tilde{\phi}_n^{(j)} \rangle_2 = \left(\lambda_m^{(j)} \lambda_n^{(j)} \right)^{-1} \mathbf{u}_m^{(j) \top} \mathbf{A}^{(j)} \mathbf{u}_n^{(j)} = \left(\lambda_m^{(j)} \right)^{-1} \delta_{mn},$$

orthonormal eigenfunctions $\phi_m^{(j)}$ are given by

$$\phi_m^{(j)} = \lambda_m^{(j)1/2} \tilde{\phi}_m^{(j)} = \left(\lambda_m^{(j)} \right)^{-1/2} \sum_{n=1}^M \nu_n^{1/2} \psi_n^{(j)} [\mathbf{u}_m^{(j)}]_n.$$

Therefore, $X^{(j)}$ has a finite Karhunen-Loève representation $X^{(j)} = \sum_{m=1}^{M_j} \xi_m^{(j)} \phi_m^{(j)}$ with scores

$$\xi_m^{(j)} = \langle X^{(j)}, \phi_m^{(j)} \rangle_2 = \left(\lambda_m^{(j)} \right)^{-1/2} \sum_{n=1}^M \nu_n^{1/2} [\mathbf{u}_m^{(j)}]_n \sum_{k=1}^M \rho_k \langle \psi_n^{(j)}, \psi_k^{(j)} \rangle_2$$

and $\mathbb{E}(\xi_m^{(j)}) = 0$, $\text{Cov}(\xi_m^{(j)}, \xi_n^{(j)}) = \lambda_m^{(j)} \delta_{mn}$.

2. Assume the functional covariates $X^{(1)}, \dots, X^{(p)}$ do each have a finite Karhunen-Loève representation, i.e. for each $j = 1, \dots, p$: $X^{(j)} = \sum_{m=1}^{M_j} \xi_m^{(j)} \phi_m^{(j)}$. Let $\Gamma\psi = \nu\psi$. Then for all $j = 1, \dots, p$ and $t_j \in \mathcal{T}_j$

$$\begin{aligned} (\Gamma\psi)^{(j)}(t_j) &= \sum_{k=1}^p \int_{\mathcal{T}_k} \text{Cov} \left(X^{(k)}(s_k), X^{(j)}(t_j) \right) \psi^{(k)}(s_k) ds_k \\ &= \sum_{k=1}^p \sum_{l=1}^{M_j} \sum_{n=1}^{M_k} \underbrace{\text{Cov} \left(\xi_l^{(j)}, \xi_n^{(k)} \right)}_{=: Z_{ln}^{(jk)}} \phi_l^{(j)}(t_j) \underbrace{\int_{\mathcal{T}_k} \phi_n^{(k)}(s_k) \psi^{(k)}(s_k) ds_k}_{=: c_n^{(k)}} \stackrel{!}{=} \nu \psi^{(j)}(t_j). \end{aligned}$$

With a similar argumentation as in Zemyan (2012, Chapter 1.3.) it holds

$$\begin{aligned}
& \sum_{k=1}^p \sum_{l=1}^{M_j} \sum_{n=1}^{M_k} Z_{ln}^{(jk)} \phi_l^{(j)}(t_j) c_n^{(k)} = \nu \psi^{(j)}(t_j) \\
& \Rightarrow \int_{\mathcal{T}_j} \phi_m^{(j)}(t_j) \cdot \sum_{k=1}^p \sum_{l=1}^{M_j} \sum_{n=1}^{M_k} Z_{ln}^{(jk)} \phi_l^{(j)}(t_j) c_n^{(k)} dt_j = \int_{\mathcal{T}_j} \phi_m^{(j)}(t_j) \cdot \nu \psi^{(j)}(t_j) dt_j \\
& \Leftrightarrow \sum_{k=1}^p \sum_{n=1}^{M_k} Z_{mn}^{(jk)} c_n^{(k)} = \nu c_m^{(j)}
\end{aligned} \tag{15}$$

for $m = 1, \dots, M_j$ due to orthonormality of $\phi_m^{(j)}$. Since m and j were arbitrarily chosen, this is equivalent to

$$\underbrace{\begin{pmatrix} \mathbf{Z}^{(11)} & \dots & \mathbf{Z}^{(1p)} \\ \vdots & \ddots & \vdots \\ \mathbf{Z}^{(p1)} & \dots & \mathbf{Z}^{(pp)} \end{pmatrix}}_{=: \mathbf{Z}} \underbrace{\begin{pmatrix} \mathbf{c}^{(1)} \\ \vdots \\ \mathbf{c}^{(p)} \end{pmatrix}}_{=: \mathbf{c}} = \nu \begin{pmatrix} \mathbf{c}^{(1)} \\ \vdots \\ \mathbf{c}^{(p)} \end{pmatrix}$$

with matrices $\mathbf{Z}^{(jk)} \in \mathbb{R}^{M_j \times M_k}$ and $\mathbf{c}^{(j)} \in \mathbb{R}^{M_j}$. The last equation is again an eigenequation for the symmetric (and since $\nu \geq 0$) positive semidefinite block matrix $\mathbf{Z} \in \mathbb{R}^{M_+ \times M_+}$. Let $\nu_1 \geq \dots \geq \nu_M > 0$, $M \leq M_+$ be the non-zero eigenvalues of \mathbf{Z} . These are also the only non-zero eigenvalues of Γ and the elements $\psi_m^{(j)}$ of the associated eigenfunctions ψ_m are parametrized by the (orthonormal) eigenvectors $\mathbf{c}_1, \dots, \mathbf{c}_M$ associated with ν_1, \dots, ν_M :

$$\psi_m^{(j)}(t_j) \stackrel{(15)}{=} \frac{1}{\nu_m} \sum_{k=1}^p \sum_{l=1}^{M_j} \sum_{n=1}^{M_k} Z_{ln}^{(jk)} [\mathbf{c}_m]_n^{(k)} \phi_l^{(j)}(t_j) = \sum_{l=1}^{M_j} [\mathbf{c}_m]_l^{(j)} \phi_l^{(j)}(t_j), \quad t_j \in \mathcal{T}_j$$

for $m = 1, \dots, M$, $j = 1, \dots, p$. The eigenfunctions form an orthonormal system with respect to $\langle\langle \cdot, \cdot \rangle\rangle$:

$$\langle\langle \psi_m, \psi_n \rangle\rangle = \sum_{j=1}^p \langle \psi_m^{(j)}, \psi_n^{(j)} \rangle_2 = \sum_{j=1}^p \sum_{l=1}^{M_j} [\mathbf{c}_m]_l^{(j)} [\mathbf{c}_n]_l^{(j)} = \mathbf{c}_n^\top \mathbf{c}_m = \delta_{mn}.$$

The Karhunen-Loève decomposition of X is therefore given by $X = \sum_{m=1}^M \rho_m \psi_m$ with scores

$$\rho_m = \langle\langle X, \psi_m \rangle\rangle = \sum_{j=1}^p \sum_{n=1}^{M_j} [\mathbf{c}_m]_n^{(j)} \xi_n^{(j)}$$

and $\mathbb{E}(\rho_m) = 0$, $\text{Cov}(\rho_m, \rho_n) = \nu_m \delta_{mn}$, $m = 1, \dots, M \leq M_+$. □

Proof of Prop. 6. For $f \in \mathcal{H}$, $\mathbf{t} \in \mathcal{T}$ and $j = 1, \dots, p$, the covariance operator $\Gamma^{[M]}$ associated with $X^{[M]}$ is given by

$$(\Gamma^{[M]} f)^{(j)}(t_j) = \sum_{i=1}^p \int_{\mathcal{T}_i} \text{Cov}(X^{[M](i)}(s_i), X^{[M](j)}(t_j)) f^{(i)}(s_i) ds_i.$$

In the following, use $C_{ij}^{[M]}(s_i, t_j) := \text{Cov}(X^{[M](i)}(s_i), X^{[M](j)}(t_j))$ as short notation for the covariance functions (cf. the definition of C_{ij} in (1) in the paper). Next, recall some well-known results for

univariate functional data: By Mercer's Theorem (Mercer, 1909)

$$C_{jj}^{[M]}(t_j, t_j) = \sum_{m=1}^{M_j} \lambda_m^{(j)} \phi_m^{(j)}(t_j)^2 \nearrow \sum_{m=1}^{\infty} \lambda_m^{(j)} \phi_m^{(j)}(t_j)^2 = C_{jj}(t_j, t_j) \text{ for } M_j \rightarrow \infty, t_j \in \mathcal{T}_j. \quad (16)$$

The univariate Karhunen-Loève Theorem (e.g. Bosq, 2000, Thm 1.5.) states that

$$\mathbb{E} \left[\left| X^{(j)}(t_j) - \sum_{m=1}^{M_j} \xi_m^{(j)} \phi_m^{(j)}(t_j) \right|^2 \right]$$

converges uniformly to 0 for $t_j \in \mathcal{T}_j$ and $M_j \rightarrow \infty$. As both $X^{(j)}$ and $X^{[M](j)}$ have zero mean ($X^{(j)}$ by assumption and $X^{[M](j)}$ since the scores $\xi_m^{(j)}$ have zero mean), this implies

$$\text{Var} \left(X^{(j)}(t_j) - X^{[M](j)}(t_j) \right) \rightarrow 0 \text{ for } M_j \rightarrow \infty. \quad (17)$$

With the assumptions of Prop. 2 it further holds (cf. proof of Prop. 3) that

$$\text{Var}(X^{(j)}(t_j)) = C_{jj}(t_j, t_j) \leq \|C_{jj}\|_{\infty} < \infty. \quad (18)$$

For fixed $s_i \in \mathcal{T}_i$, $t_j \in \mathcal{T}_j$ with $i, j = 1, \dots, p$, these three properties give

$$\begin{aligned} & \left| C_{ij}(s_i, t_j) - C_{ij}^{[M]}(s_i, t_j) \right| \\ & \leq \left| \text{Cov}(X^{(i)}(s_i) - X^{[M](i)}(s_i), X^{(j)}(t_j)) \right| + \left| \text{Cov}(X^{[M](i)}(s_i), X^{(j)}(t_j) - X^{[M](j)}(t_j)) \right| \\ & \stackrel{(16)}{\leq} \underbrace{\text{Var}(X^{(i)}(s_i) - X^{[M](i)}(s_i))^{1/2}}_{\rightarrow 0 \text{ for } M_i \rightarrow \infty (17)} \underbrace{C_{jj}(t_j, t_j)^{1/2}}_{< \infty (18)} + \underbrace{C_{ii}(s_i, s_i)^{1/2}}_{< \infty (18)} \underbrace{\text{Var}(X^{(j)}(t_j) - X^{[M](j)}(t_j))^{1/2}}_{\rightarrow 0 \text{ for } M_j \rightarrow \infty (17)} \end{aligned}$$

Hence it holds that

$$C_{ij}^{[M]}(s_i, t_j) \rightarrow C_{ij}(s_i, t_j) \text{ for } M_i, M_j \rightarrow \infty. \quad (19)$$

The main proof is now in three steps:

1. $\Gamma^{[M]}$ converges in norm to Γ for $M_1, \dots, M_p \rightarrow \infty$: Let $\|\cdot\|_{\text{op}}$ be the operator norm induced by $\|\cdot\|$. Then

$$\begin{aligned} \|\Gamma - \Gamma^{[M]}\|_{\text{op}}^2 &= \sup_{\|f\|=1} \sum_{j=1}^p \int_{\mathcal{T}_j} \left[\sum_{i=1}^p \int_{\mathcal{T}_i} \left(C_{ij}(s_i, t_j) - C_{ij}^{[M]}(s_i, t_j) \right) f^{(i)}(s_i) ds_i \right]^2 dt_j \\ &\leq \sup_{\|f\|=1} \sum_{j=1}^p \int_{\mathcal{T}_j} \left[\sum_{i=1}^p \int_{\mathcal{T}_i} \left| (C_{ij}(s_i, t_j) - C_{ij}^{[M]}(s_i, t_j)) f^{(i)}(s_i) \right| ds_i \right]^2 dt_j \\ &\stackrel{\text{Hölder}}{\leq} \sup_{\|f\|=1} \sum_{j=1}^p \int_{\mathcal{T}_j} \left[\sum_{i=1}^p \|C_{ij}(\cdot, t_j) - C_{ij}^{[M]}(\cdot, t_j)\|_2 \|f^{(i)}\|_2 \right]^2 dt_j \\ &\leq \sum_{j=1}^p \int_{\mathcal{T}_j} \left[\sum_{i=1}^p \|C_{ij}(\cdot, t_j) - C_{ij}^{[M]}(\cdot, t_j)\|_2 \right]^2 dt_j, \end{aligned} \quad (20)$$

where the last equality holds since $\|f^{(i)}\|_2 \leq 1$ for all $f \in \mathcal{H}$ with $\|f\| = 1$. The final bound for $\|\Gamma - \Gamma^{[M]}\|_{\text{op}}^2$ converges to zero for $M_1, \dots, M_p \rightarrow \infty$ by (19), applying the dominated convergence theorem twice: For the norm term in (20) consider

$$\left| C_{ij}(s_i, t_j) - C_{ij}^{[M]}(s_i, t_j) \right| \leq |C_{ij}(s_i, t_j)| + \left| C_{ij}^{[M]}(s_i, t_j) \right| \stackrel{(16)}{\leq} 2C_{ii}(s_i, s_i)^{1/2}C_{jj}(t_j, t_j)^{1/2},$$

thus $\left| C_{ij}(s_i, t_j) - C_{ij}^{[M]}(s_i, t_j) \right|^2 \stackrel{(18)}{\leq} 4\|C_{ii}\|_{\infty}\|C_{jj}\|_{\infty} < \infty$. This upper bound is constant and therefore integrable over \mathcal{T}_i , which implies

$$\lim_{M_i \rightarrow \infty} \left\| C_{ij}(\cdot, t_j) - C_{ij}^{[M]}(\cdot, t_j) \right\|_2 = \left\| C_{ij}(\cdot, t_j) - \lim_{M_i \rightarrow \infty} C_{ij}^{[M]}(\cdot, t_j) \right\|_2.$$

For the outer integral in (20) the results of the norm term give

$$\left(\sum_{i=1}^p \left\| C_{ij}(\cdot, t_j) - C_{ij}^{[M]}(\cdot, t_j) \right\|_2 \right)^2 \leq 4\|C_{jj}\|_{\infty} \left(\sum_{i=1}^p (\|C_{ii}\|_{\infty} \lambda(\mathcal{T}_i))^{1/2} \right)^2,$$

where $\lambda(\mathcal{T}_i)$ is the Lebesgue measure of \mathcal{T}_i as in the Proof of Prop. 2. The term on the right hand side is constant and hence integrable over \mathcal{T}_j , which gives that for $M_1, \dots, M_p \rightarrow \infty$, the limit $M_j \rightarrow \infty$ and the integral over \mathcal{T}_j in (20) can be interchanged. In summary, these results give that $\Gamma^{[M]}$ converges to Γ in norm for $M_1, \dots, M_p \rightarrow \infty$.

2. $\Gamma^{[M]}$ is bounded: Let $f \in \mathcal{H}$. Clearly, $\|f^{(i)}\|_2 \leq \|f\|$ for all $i = 1, \dots, p$ and therefore

$$\begin{aligned} \left\| \Gamma^{[M]} f \right\|^2 &= \sum_{j=1}^p \int_{\mathcal{T}_j} \left(\sum_{i=1}^p \int_{\mathcal{T}_i} C_{ij}^{[M]}(s_i, t_j) f^{(i)}(s_i) ds_i \right)^2 dt_j \\ &\stackrel{\text{H\"older}}{\leq} \sum_{j=1}^p \int_{\mathcal{T}_j} \left(\sum_{i=1}^p \left(\int_{\mathcal{T}_i} |C_{ij}^{[M]}(s_i, t_j)|^2 ds_i \right)^{1/2} \left(\int_{\mathcal{T}_i} |f^{(i)}(s_i)|^2 ds_i \right)^{1/2} \right)^2 dt_j \\ &\stackrel{(16)(18)}{\leq} \sum_{j=1}^p \int_{\mathcal{T}_j} \left(\sum_{i=1}^p \|C_{ii}\|_{\infty}^{1/2} \|C_{jj}\|_{\infty}^{1/2} \lambda(\mathcal{T}_i)^{1/2} \|f^{(i)}\|_2 \right)^2 dt_j \\ &\leq \sum_{j=1}^p \|C_{jj}\|_{\infty} \lambda(\mathcal{T}_j) \left(\sum_{i=1}^p \|C_{ii}\|_{\infty}^{1/2} \lambda(\mathcal{T}_i)^{1/2} \right)^2 \|f\|^2 \leq p^3 C^2 T^2 \|f\|^2, \end{aligned}$$

for $T = \max_{j=1, \dots, p} \lambda(\mathcal{T}_j)$ and $C = \max_{j=1, \dots, p} \|C_{jj}\|_{\infty}$. The value $p^{3/2}CT$ is constant and finite, hence $\Gamma^{[M]}$ is bounded.

3. Convergence results for $\nu_m^{[M]}$, $\psi_m^{[M]}$ and $\rho_m^{[M]}$: In Prop. 2, it was shown that Γ is compact, which implies that this operator is also bounded (Reed and Simon, 1980, Chapter VI.5.). As Γ and $\Gamma^{[M]}$ are both bounded, norm convergence is equivalent to convergence in the generalized sense (Kato, 1976, Chapter IV, §2.6., Thm. 2.23). This implies that the eigenvalues $\nu_m^{[M]}$ of $\Gamma^{[M]}$

converge to the eigenvalues ν_m of Γ including multiplicity (if the multiplicity is finite, which holds for all nonzero eigenvalues, as Γ is compact (cf. Reed and Simon, 1980, Thm. VI.15.)) and the associated total projections converge in norm (Kato, 1976, Chapter IV, §3.5.). If the m -th eigenvalue has multiplicity 1, then the projections on the eigenspaces spanned by ψ_m and $\psi_m^{[M]}$, respectively, are given by

$$P_m f = \langle \psi_m, f \rangle \psi_m, \quad P_m^{[M]} f = \langle \psi_m^{[M]}, f \rangle \psi_m^{[M]}, \quad f \in \mathcal{H}.$$

Without loss of generality one may choose the orientation of ψ_m and $\psi_m^{[M]}$ such that $\langle \psi_m, \psi_m^{[M]} \rangle \geq 0$. In this case, as $\psi_m, \psi_m^{[M]}$ both have norm 1,

$$\begin{aligned} \left\| P_m - P_m^{[M]} \right\|_{\text{op}}^2 &\geq \left\| \langle \psi_m, \psi_m \rangle \psi_m - \langle \psi_m^{[M]}, \psi_m \rangle \psi_m^{[M]} \right\|^2 = \left\| \psi_m - \langle \psi_m^{[M]}, \psi_m \rangle \psi_m^{[M]} \right\|^2 \\ &= \left\| \psi_m \right\|^2 - 2 \langle \psi_m, \langle \psi_m^{[M]}, \psi_m \rangle \psi_m^{[M]} \rangle + \langle \psi_m^{[M]}, \psi_m \rangle^2 \left\| \psi_m^{[M]} \right\|^2 \\ &= (1 - \langle \psi_m^{[M]}, \psi_m \rangle)(1 + \langle \psi_m^{[M]}, \psi_m \rangle) \geq (1 - \langle \psi_m^{[M]}, \psi_m \rangle) \\ &= \frac{1}{2} \cdot \left\| \psi_m - \psi_m^{[M]} \right\|^2. \end{aligned}$$

Norm convergence of the total projections hence implies $\left\| \psi_m - \psi_m^{[M]} \right\| \rightarrow 0$ for $M_1, \dots, M_p \rightarrow \infty$.

To derive convergence of the scores $\rho_m^{[M]}$, note that for $\varepsilon > 0$ and $c := \left(\frac{2}{\varepsilon} \sum_{j=1}^p \|C_{jj}\|_{\infty} \lambda(\mathcal{T}_j) \right)^{1/2}$

$$\begin{aligned} P(\|X\| > c) &\stackrel{\text{Markov}}{\leq} \frac{1}{c^2} \mathbb{E}[\|X\|^2] \stackrel{\text{Fubini}}{=} \frac{1}{c^2} \sum_{j=1}^p \int_{\mathcal{T}_j} \mathbb{E}[X^{(j)}(t_j)^2] dt_j \\ &= \frac{1}{c^2} \sum_{j=1}^p \int_{\mathcal{T}_j} C_{jj}(t_j, t_j) dt_j \leq \frac{1}{c^2} \sum_{j=1}^p \|C_{jj}\|_{\infty} \lambda(\mathcal{T}_j) = \frac{\varepsilon}{2} < \varepsilon, \end{aligned} \quad (21)$$

i.e. the norm of X is bounded in probability. Moreover,

$$\mathbb{E} \left[\left\| X - X^{[M]} \right\|^2 \right] \stackrel{\text{Fubini}}{=} \sum_{j=1}^p \int_{\mathcal{T}_j} \mathbb{E} \left[\left| X^{(j)}(t_j) - X^{[M](j)}(t_j) \right|^2 \right] dt_j \rightarrow 0$$

for $M_1, \dots, M_p \rightarrow \infty$, as the expectation in the integral converges uniformly to 0 and is thus bounded (by univariate Karhunen-Loève). As \mathcal{T}_j has finite measure, the overall integral converges to 0. Hence $\|X - X^{[M]}\|$ converges in the second mean to 0, thus $\|X - X^{[M]}\| = o_p(1)$. Finally, this leads to

$$\begin{aligned} \left| \rho_m - \rho_m^{[M]} \right| &= \left| \langle X, \psi_m \rangle - \langle X^{[M]}, \psi_m^{[M]} \rangle \right| \leq \left| \langle X, \psi_m - \psi_m^{[M]} \rangle \right| + \left| \langle X - X^{[M]}, \psi_m^{[M]} \rangle \right| \\ &\leq \|X\| \left\| \psi_m - \psi_m^{[M]} \right\| + \left\| X - X^{[M]} \right\| \left\| \psi_m^{[M]} \right\| = O_p(1) o(1) + o_p(1) = o_p(1). \end{aligned}$$

i.e. $\rho_m^{[M]}$ converges in probability to ρ_m .

□

Lemma 2. *Under the assumptions of Prop. 7 it holds that*

$$\lambda_{\max}(\mathbf{Z} - \hat{\mathbf{Z}}) \leq O_p(M_{\max} \max(N^{-1/2}, \Delta_M r_N^\Gamma)).$$

with \mathbf{Z} as defined in Prop. 5, $\hat{\mathbf{Z}} = (N-1)^{-1} \mathbf{\Xi}^\top \mathbf{\Xi}$ as in Section 3.2, $M_{\max} = \max_{j=1, \dots, p} M_j$ and $\Delta_M := \max_{j=1, \dots, p} \Delta_{M_j}^{(j)}$.

Proof of Lemma 2. For $j, k = 1, \dots, p$ and $f \in L^2(\mathcal{T}_j)$ define the bounded operator $\Gamma^{(jk)}: L^2(\mathcal{T}_j) \rightarrow L^2(\mathcal{T}_k)$ via

$$\left(\Gamma^{(jk)} f \right) (t) := \int_{\mathcal{T}_j} \text{Cov} \left(X^{(j)}(s), X^{(k)}(t) \right) f(s) ds = \int_{\mathcal{T}_j} C_{jk}(s, t) f(s) ds$$

Analogously, define $\hat{\Gamma}^{(jk)}: L^2(\mathcal{T}_j) \rightarrow L^2(\mathcal{T}_k)$ by

$$\left(\hat{\Gamma}^{(jk)} f \right) (t) := \int_{\mathcal{T}_j} \widehat{\text{Cov}} \left(X^{(j)}(s), X^{(k)}(t) \right) f(s) ds = \int_{\mathcal{T}_j} \hat{C}_{jk}(s, t) f(s) ds$$

with $\hat{C}_{jk}(s, t) := \widehat{\text{Cov}} \left(X^{(j)}(s), X^{(k)}(t) \right) = \frac{1}{N} \sum_{i=1}^N X_i^{(j)}(s) X_i^{(k)}(t)$. If the X_i are independent copies of the process X , it holds

$$\begin{aligned} & \mathbb{E} \left[\int_{\mathcal{T}_j} \int_{\mathcal{T}_k} \left(C_{jk}(s, t) - \hat{C}_{jk}(s, t) \right)^2 ds dt \right] \\ &= \mathbb{E} \left[\frac{1}{N^2} \sum_{i=1}^N \sum_{l=1}^N \int_{\mathcal{T}_j} \int_{\mathcal{T}_k} \left(C_{jk}(s, t) - X_i^{(j)}(s) X_l^{(k)}(t) \right) \left(C_{jk}(s, t) - X_l^{(j)}(s) X_i^{(k)}(t) \right) ds dt \right] \\ &= \frac{1}{N^2} \sum_{i=1}^N \int_{\mathcal{T}_j} \int_{\mathcal{T}_k} \mathbb{E} \left[\left(C_{jk}(s, t) - X_i^{(j)}(s) X_i^{(k)}(t) \right)^2 \right] ds dt \\ &= \frac{1}{N} \int_{\mathcal{T}_j} \int_{\mathcal{T}_k} \mathbb{E} \left[X^{(j)}(s)^2 X^{(k)}(t)^2 \right] - C_{jk}(s, t)^2 ds dt = O(N^{-1}). \end{aligned}$$

The last step follows from the fact that the integral term does not depend on N and is finite by assumption (A2) and the conditions in Prop. 2 for C_{jk} . This implies

$$\left\| \Gamma^{(jk)} - \hat{\Gamma}^{(jk)} \right\|_{\text{op}} \leq \left(\int_{\mathcal{T}_j} \int_{\mathcal{T}_k} \left(C_{jk}(s, t) - \hat{C}_{jk}(s, t) \right)^2 ds dt \right)^{1/2} \stackrel{\text{Markov}}{=} O_p(N^{-1/2}).$$

Recall $Z_{ln}^{(jk)} = \text{Cov}(\xi_l^{(j)}, \xi_n^{(k)})$ and $\hat{Z}_{ln}^{(jk)} = \frac{1}{N-1} \sum_{i=1}^N \hat{\xi}_{i,l}^{(j)} \hat{\xi}_{i,n}^{(k)}$ for $j, k = 1, \dots, p$, $l = 1, \dots, M_j$, $n = 1, \dots, M_k$. As \mathbf{Z} and $\hat{\mathbf{Z}}$ are both symmetric matrices in $\mathbb{R}^{M_+ \times M_+}$ it holds (cf. Horn and Johnson, 1991, Chapter 3.7)

$$\lambda_{\max}(\mathbf{Z} - \hat{\mathbf{Z}}) \leq \max_{j=1, \dots, p} \max_{l=1, \dots, M_j} \sum_{k=1}^p \sum_{n=1}^{M_k} \left| Z_{ln}^{(jk)} - \hat{Z}_{ln}^{(jk)} \right|. \quad (22)$$

Let now $j, k = 1, \dots, p$, $l = 1, \dots, M_j$, $n = 1, \dots, M_k$ be fixed. Assumption (A5) gives

$$\begin{aligned}
\left| Z_{ln}^{(jk)} - \hat{Z}_{ln}^{(jk)} \right| &= \left| \text{Cov} \left(\xi_l^{(j)}, \xi_n^{(k)} \right) - \frac{1}{N} \sum_{i=1}^N \hat{\xi}_{i,l}^{(j)} \hat{\xi}_{i,n}^{(k)} - \frac{1}{N(N-1)} \sum_{i=1}^N \hat{\xi}_{i,l}^{(j)} \hat{\xi}_{i,n}^{(k)} \right| \\
&\stackrel{(A5)}{\leq} \left| \text{Cov} \left(\langle X^{(j)}, \phi_l^{(j)} \rangle_2, \langle X^{(k)}, \phi_n^{(k)} \rangle_2 \right) - \frac{1}{N} \sum_{i=1}^N \langle X_i^{(j)}, \hat{\phi}_l^{(j)} \rangle_2 \cdot \langle X_i^{(k)}, \hat{\phi}_n^{(k)} \rangle_2 \right| \\
&\quad + \frac{1}{N(N-1)} \sum_{i=1}^N \left| \langle X_i^{(j)}, \hat{\phi}_l^{(j)} \rangle_2 \right| \left| \langle X_i^{(k)}, \hat{\phi}_n^{(k)} \rangle_2 \right| \\
&\leq \left| \int_{\mathcal{T}_j} \int_{\mathcal{T}_k} \mathbb{E} \left(X^{(j)}(s) X^{(k)}(t) \right) \phi_l^{(j)}(s) \phi_n^{(k)}(t) ds dt - \int_{\mathcal{T}_j} \int_{\mathcal{T}_k} \frac{1}{N} \sum_{i=1}^N \left(X_i^{(j)}(s) X_i^{(k)}(t) \right) \hat{\phi}_l^{(j)}(s) \hat{\phi}_n^{(k)}(t) ds dt \right| \\
&\quad + \frac{1}{N(N-1)} \sum_{i=1}^N \left\| X_i^{(j)} \right\|_2 \left\| \hat{\phi}_l^{(j)} \right\|_2 \left\| X_i^{(k)} \right\|_2 \left\| \hat{\phi}_n^{(k)} \right\|_2 \\
&\stackrel{(21)}{=} \left| \int_{\mathcal{T}_j} \int_{\mathcal{T}_k} \text{Cov} \left(X^{(j)}(s), X^{(k)}(t) \right) \phi_l^{(j)}(s) \phi_n^{(k)}(t) - \widehat{\text{Cov}} \left(X^{(j)}(s), X^{(k)}(t) \right) \hat{\phi}_l^{(j)}(s) \hat{\phi}_n^{(k)}(t) ds dt \right| \\
&\quad + \frac{1}{N(N-1)} \sum_{i=1}^N O_p(1) \cdot 1 \cdot O_p(1) \cdot 1 \\
&= \left| \int_{\mathcal{T}_j} \int_{\mathcal{T}_k} C_{jk}(s, t) \left[\phi_l^{(j)}(s) \phi_n^{(k)}(t) - \hat{\phi}_l^{(j)}(s) \hat{\phi}_n^{(k)}(t) \right] ds dt \right| \\
&\quad + \left| \int_{\mathcal{T}_j} \int_{\mathcal{T}_k} \left[C_{jk}(s, t) - \hat{C}_{jk}(s, t) \right] \hat{\phi}_l^{(j)}(s) \hat{\phi}_n^{(k)}(t) ds dt \right| + O_p(N^{-1}) \\
&\leq \int_{\mathcal{T}_j} \int_{\mathcal{T}_k} C_{jj}(s, s)^{1/2} C_{kk}(t, t)^{1/2} \left| \phi_l^{(j)}(s) \phi_n^{(k)}(t) - \hat{\phi}_l^{(j)}(s) \hat{\phi}_n^{(k)}(t) \right| ds dt \\
&\quad + \int_{\mathcal{T}_k} \left| \left((\Gamma^{(jk)} - \hat{\Gamma}^{(jk)}) \hat{\phi}_l^{(j)} \right) (t) \hat{\phi}_n^{(k)}(t) \right| dt + O_p(N^{-1}) \\
&\leq \|C_{jj}\|_\infty^{1/2} \|C_{kk}\|_\infty^{1/2} \left(\int_{\mathcal{T}_j} \int_{\mathcal{T}_k} \left| \phi_l^{(j)}(s) \right| \left| \phi_n^{(k)}(t) - \hat{\phi}_n^{(k)}(t) \right| ds dt + \int_{\mathcal{T}_j} \int_{\mathcal{T}_k} \left| \phi_l^{(j)}(s) - \hat{\phi}_l^{(j)}(s) \right| \left| \hat{\phi}_n^{(k)}(t) \right| ds dt \right) \\
&\quad + \left\| (\Gamma^{(jk)} - \hat{\Gamma}^{(jk)}) \hat{\phi}_l^{(j)} \right\|_2 \left\| \hat{\phi}_n^{(k)} \right\|_2 + O_p(N^{-1}) \\
&\leq (\|C_{jj}\|_\infty \|C_{kk}\|_\infty \lambda(\mathcal{T}_j) \lambda(\mathcal{T}_k))^{1/2} \left(\left\| \phi_l^{(j)} \right\|_2 \left\| \phi_n^{(k)} - \hat{\phi}_n^{(k)} \right\|_2 + \left\| \phi_l^{(j)} - \hat{\phi}_l^{(j)} \right\|_2 \left\| \hat{\phi}_n^{(k)} \right\|_2 \right) \\
&\quad + \left\| \Gamma^{(jk)} - \hat{\Gamma}^{(jk)} \right\|_{\text{op}} \left\| \hat{\phi}_l^{(j)} \right\|_2 \left\| \hat{\phi}_n^{(k)} \right\|_2 + O_p(N^{-1}) \\
&= (\|C_{jj}\|_\infty \|C_{kk}\|_\infty \lambda(\mathcal{T}_j) \lambda(\mathcal{T}_k))^{1/2} \left(\left\| \phi_n^{(k)} - \hat{\phi}_n^{(k)} \right\|_2 + \left\| \phi_l^{(j)} - \hat{\phi}_l^{(j)} \right\|_2 \right) + \left\| \Gamma^{(jk)} - \hat{\Gamma}^{(jk)} \right\|_{\text{op}} + O_p(N^{-1}) \\
&= O_p(\Delta_{M_j}^{(j)} r_N^\Gamma) + O_p(\Delta_{M_k}^{(k)} r_N^\Gamma) + O_p(N^{-1/2}) + O_p(N^{-1}) \\
&= O_p(\max(\Delta_{M_j}^{(j)} r_N^\Gamma, \Delta_{M_k}^{(k)} r_N^\Gamma, N^{-1/2})).
\end{aligned}$$

The rate for $\phi_n^{(k)}$ and $\phi_l^{(j)}$ in the last steps is shown at the beginning of the proof of Prop. 7. In total, equation (22), $M_{\max} = \max_{j=1,\dots,p} M_j$ and $\Delta_M := \max_{j=1,\dots,p} \Delta_{M_j}^{(j)}$ give

$$\lambda_{\max}(\mathbf{Z} - \hat{\mathbf{Z}}) \leq O_p(M_{\max} \max(N^{-1/2}, \Delta_M r_N^\Gamma)).$$

□

Proof of Prop. 7. Under assumption (A4) and using the convention $\lambda_0^{(j)} := \infty$, Lemma 4.3. in Bosq (2000) gives for $m = 1, \dots, M_j$

$$\begin{aligned} \left\| \phi_m^{(j)} - \hat{\phi}_m^{(j)} \right\|_2 &\leq 8^{1/2} \left[\min \left(\lambda_m^{(j)} - \lambda_{m+1}^{(j)}, \lambda_{m-1}^{(j)} - \lambda_m^{(j)} \right) \right]^{-1} \left\| \Gamma^{(j)} - \hat{\Gamma}^{(j)} \right\|_{\text{op}} \\ &\leq 8^{1/2} \Delta_{M_j}^{(j)} \left\| \Gamma^{(j)} - \hat{\Gamma}^{(j)} \right\|_{\text{op}} = O_p(\Delta_{M_j}^{(j)} r_N^\Gamma). \end{aligned}$$

Based on this result, Lemma 2 states that $\lambda_{\max}(\mathbf{Z} - \hat{\mathbf{Z}}) \leq O_p(M_{\max} \max(N^{-1/2}, \Delta_M r_N^\Gamma))$ with $\Delta_M := \max_{j=1,\dots,p} \Delta_{M_j}^{(j)}$ and $M_{\max} = \max_{j=1,\dots,p} M_j$.

1. Eigenvalues: Let $\boldsymbol{\xi} \in \mathbb{R}^{M_+}$ with entries $\xi_m^{(j)} = \langle X^{(j)}, \phi_m^{(j)} \rangle_2 = \langle X^{[M](j)}, \phi_m^{(j)} \rangle_2$, $m = 1, \dots, M_j$, $j = 1, \dots, p$. For fixed $m = 1, \dots, M_+$ it holds that

$$\begin{aligned} \left| \nu_m^{[M]} - \hat{\nu}_m \right| &= \left| \text{Var}(\langle X_i^{[M]}, \psi_m^{[M]} \rangle) - \hat{\mathbf{c}}_m^\top \hat{\mathbf{Z}} \hat{\mathbf{c}}_m \right| = \left| \mathbf{c}_m^\top \text{Var}(\boldsymbol{\xi}) \mathbf{c}_m - \hat{\mathbf{c}}_m^\top \hat{\mathbf{Z}} \hat{\mathbf{c}}_m \right| \\ &= \left| (\mathbf{c}_m - \hat{\mathbf{c}}_m)^\top \mathbf{Z} \mathbf{c}_m + \hat{\mathbf{c}}_m^\top \mathbf{Z} (\mathbf{c}_m - \hat{\mathbf{c}}_m) + \hat{\mathbf{c}}_m^\top (\mathbf{Z} - \hat{\mathbf{Z}}) \hat{\mathbf{c}}_m \right| \\ &\leq \|\mathbf{c}_m - \hat{\mathbf{c}}_m\| \cdot \nu_m^{[M]} \|\mathbf{c}_m\| + \lambda_{\max}(\mathbf{Z}) \|\mathbf{c}_m - \hat{\mathbf{c}}_m\| + \lambda_{\max}(\mathbf{Z} - \hat{\mathbf{Z}}) \|\hat{\mathbf{c}}_m\| \\ &= \|\mathbf{c}_m - \hat{\mathbf{c}}_m\| (\nu_m^{[M]} + \nu_1^{[M]}) + \lambda_{\max}(\mathbf{Z} - \hat{\mathbf{Z}}) \\ &\leq \frac{8^{1/2} \lambda_{\max}(\mathbf{Z} - \hat{\mathbf{Z}})}{\min(\nu_{m-1}^{[M]} - \nu_m^{[M]}, \nu_m^{[M]} - \nu_{m+1}^{[M]})} 2\nu_1^{[M]} + \lambda_{\max}(\mathbf{Z} - \hat{\mathbf{Z}}) \\ &= \left[\frac{2^{5/2} \nu_1^{[M]}}{\min(\nu_{m-1}^{[M]} - \nu_m^{[M]}, \nu_m^{[M]} - \nu_{m+1}^{[M]})} + 1 \right] \lambda_{\max}(\mathbf{Z} - \hat{\mathbf{Z}}) \\ &= O_p(M_{\max} \max(N^{-1/2}, \Delta_M r_N^\Gamma)), \end{aligned}$$

as the expression in square brackets converges to a constant $C < \infty$ (cf. Prop. 6 and the fact that ν_m is assumed to have multiplicity 1, see p. 18 in the main paper). Here $\lambda_{\max}(\mathbf{A})$ denotes the maximal eigenvalue of a symmetric matrix \mathbf{A} . The second inequality follows from Corollary 1 in Yu et al. (2015) and the fact that $\nu_m^{[M]} \leq \nu_1^{[M]}$.

2. Eigenfunctions: Consider the j -th element of the m -th eigenfunctions:

$$\begin{aligned}
\left\| \psi_m^{[M](j)} - \hat{\psi}_m^{(j)} \right\|_2 &\leq \sum_{n=1}^{M_j} \left| [\mathbf{c}_m]_n^{(j)} - [\hat{\mathbf{c}}_m]_n^{(j)} \right| \left\| \phi_n^{(j)} \right\|_2 + \left| [\hat{\mathbf{c}}_m]_n^{(j)} \right| \left\| \phi_n^{(j)} - \hat{\phi}_n^{(j)} \right\|_2 \\
&\leq M_j^{1/2} \|\mathbf{c}_m - \hat{\mathbf{c}}_m\| + M_j^{1/2} \|\hat{\mathbf{c}}_m\| O_p(\Delta_{M_j}^{(j)} r_N^\Gamma) \\
&\leq M_j^{1/2} \left(\frac{8^{1/2} \lambda_{\max}(\mathbf{Z} - \hat{\mathbf{Z}})}{\min(\nu_{m-1}^{[M]} - \nu_m^{[M]}, \nu_m^{[M]} - \nu_{m+1}^{[M]})} + O_p(\Delta_{M_j}^{(j)} r_N^\Gamma) \right) \\
&= M_j^{1/2} O_p \left(\max(M_{\max} N^{-1/2}, M_{\max} \Delta_M r_N^\Gamma, \Delta_{M_j}^{(j)} r_N^\Gamma) \right),
\end{aligned}$$

where the last inequality uses again Corollary 1 in Yu et al. (2015). By definition of the norm, the result for the single elements implies

$$\left\| \psi_m^{[M]} - \hat{\psi}_m \right\| = O_p \left(M_{\max}^{3/2} \max(N^{-1/2}, \Delta_M r_N^\Gamma) \right).$$

3. Scores and reconstructed \hat{X} : For $\hat{\rho}_{i,m} = \Xi_{i,\cdot} \hat{\mathbf{c}}_m = \langle \hat{X}_i^{[M]}, \hat{\psi}_m \rangle$ as in Section 3.2 with $\hat{X}_i^{[M](j)} = \sum_{m=1}^{M_j} \hat{\xi}_{i,m}^{(j)} \hat{\phi}_m^{(j)}$,

$$\begin{aligned}
\left| \rho_{i,m}^{[M]} - \hat{\rho}_{i,m} \right| &= \left| \langle X_i^{[M]}, \psi_m^{[M]} - \hat{\psi}_m \rangle + \langle X_i^{[M]} - \hat{X}_i^{[M]}, \hat{\psi}_m \rangle \right| \\
&\leq \left\| X_i^{[M]} \right\| \cdot \left\| \psi_m^{[M]} - \hat{\psi}_m \right\| + \left\| X_i^{[M]} - \hat{X}_i^{[M]} \right\| \cdot \left\| \hat{\psi}_m \right\|.
\end{aligned}$$

$\left\| X_i^{[M]} \right\|$ is bounded in probability using equation (16) and analogous arguments as for $\left\| X \right\|$ in the proof of Prop. 6 (convergence of $\rho_m^{[M]}$). For the second term, note that

$$\begin{aligned}
\left\| X_i^{[M](j)} - \hat{X}_i^{[M](j)} \right\|_2 &= \left\| \sum_{m=1}^{M_j} \xi_{i,m}^{(j)} \phi_m^{(j)} - \hat{\xi}_{i,m}^{(j)} \hat{\phi}_m^{(j)} \right\|_2 \leq \sum_{m=1}^{M_j} \left| \xi_{i,m}^{(j)} \right| \left\| \phi_m^{(j)} - \hat{\phi}_m^{(j)} \right\|_2 + \left| \xi_{i,m}^{(j)} - \hat{\xi}_{i,m}^{(j)} \right| \left\| \hat{\phi}_m^{(j)} \right\|_2 \\
&\leq \sum_{m=1}^{M_j} \left| \xi_{i,m}^{(j)} \right| \left\| \phi_m^{(j)} - \hat{\phi}_m^{(j)} \right\|_2 + \left| \langle X_i^{(j)}, \phi_m^{(j)} \rangle_2 - \langle X_i^{(j)}, \hat{\phi}_m^{(j)} \rangle_2 \right| \\
&\leq \sum_{m=1}^{M_j} \left(\left| \xi_{i,m}^{(j)} \right| + \left\| X_i^{(j)} \right\|_2 \right) \left\| \phi_m^{(j)} - \hat{\phi}_m^{(j)} \right\|_2.
\end{aligned}$$

The univariate scores are uniformly bounded in probability: For $m = 1, \dots, M_j$, let $\varepsilon > 0$ and $c := (\frac{2\lambda_1^{(j)}}{\varepsilon})^{1/2} < \infty$. Then

$$P(|\xi_{i,m}^{(j)}| > c) \stackrel{\text{Markov}}{\leq} \frac{1}{c^2} \mathbb{E}[|\xi_{i,m}^{(j)}|^2] = \frac{1}{c^2} \text{Var}(\xi_{i,m}^{(j)}) = \frac{\lambda_m^{(j)}}{c^2} < \varepsilon.$$

Hence $\left\| X_i^{[M](j)} - \hat{X}_i^{[M](j)} \right\|_2 = M_j O_p(1) O_p(\Delta_{M_j}^{(j)} r_N^\Gamma)$ and

$$\left\| X_i^{[M]} - \hat{X}_i^{[M]} \right\| = O_p(M_{\max} \Delta_M r_N^\Gamma).$$

In total,

$$\begin{aligned}
\left| \rho_{i,m}^{[M]} - \hat{\rho}_{i,m} \right| &\leq \left\| \left\| X_i^{[M]} \right\| \right\| \left\| \psi_m^{[M]} - \hat{\psi}_m \right\| + \left\| \left\| X_i^{[M]} - \hat{X}_i^{[M]} \right\| \right\| \\
&= O_p(1) O_p(M_{\max}^{3/2} \max(N^{-1/2}, \Delta_M r_N^\Gamma)) + O_p(M_{\max} \Delta_M r_N^\Gamma) \\
&= O_p(M_{\max}^{3/2} \max(N^{-1/2}, \Delta_M r_N^\Gamma)).
\end{aligned}$$

□

Simulation – Additional Results

Construction of Eigenfunctions (Technical Details)

Setting 1 and 2: The first two settings of the simulation study consider multivariate functional data where each element has a one-dimensional domain (cf. Section 4.1). As a starting point for the construction of the multivariate eigenfunctions ψ_m with p elements, we use Fourier basis functions f_1, \dots, f_M on the interval $[0, 2]$. Next, choose split points $0 = T_1 < T_2 < \dots < T_p < T_{p+1} = 2$ and shift values $\eta_1, \dots, \eta_p \in \mathbb{R}$ such that $\mathcal{T}_j = [T_j + \eta_j, T_{j+1} + \eta_j]$. In the first setting with $p = 2$, one has $T_1 = 0, T_2 = 1, T_3 = 2$ and $\eta_1 = 0, \eta_2 = 1$, i.e. the functions are cut at $T_2 = 1$, and the second part is shifted to the left by 1 such that $\mathcal{T}_1 = \mathcal{T}_2 = [0, 1]$. Given random signs $\sigma_1, \dots, \sigma_p \in \{-1, 1\}$, the multivariate eigenfunctions are given by their elements

$$\psi_m^{(j)}(t_j) = \sigma_j \cdot f_m|_{[T_j, T_{j+1}]}(t_j - \eta_j), \quad m = 1, \dots, M.$$

The construction process is illustrated in Fig. 6. Clearly, $\{\psi_m, m = 1, \dots, M\}$ is an orthonormal system in $\mathcal{H} = L^2(\mathcal{T}_1) \times \dots \times L^2(\mathcal{T}_p)$. The observations x_i for the simulation are constructed as a truncated Karhunen-Loève expansion, cf. the introduction of Section 4. Exemplary data for the second simulation setting including sparse data and data with measurement error is given in Fig. 7

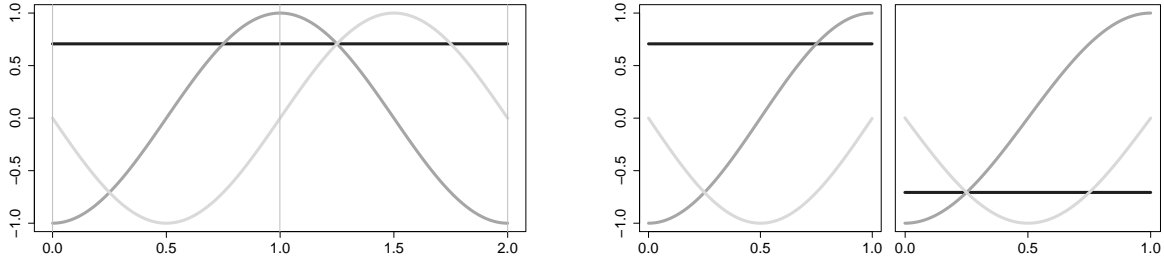


Figure 6: Illustration of the construction of the multivariate eigenfunctions ψ_m for the first setting. Left: The first $M = 3$ functions of the Fourier basis on $[0, 2]$ with one split point. Right: The shifted pieces multiplied with random signs form the first three bivariate eigenfunctions.

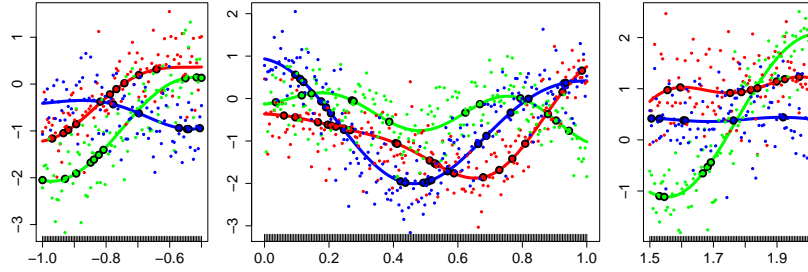


Figure 7: Three examples for simulated data in simulation setting 2 based on the leading $M = 8$ Fourier basis functions and exponential eigenvalue decay. Left: $x_i^{(1)}$, Middle: $x_i^{(2)}$, Right: $x_i^{(3)}$. Solid lines show the realizations x_i , small points are the corresponding data with measurement error, big points mark measurements of the artificially sparsified data (high sparsity level).

Setting 3: The data in the third setting consists of images and functions, hence multivariate functional data with elements having different dimensional domains (cf. Section 4.2). The basic idea here is to find orthonormal bases for each of the domains and to construct the eigenfunctions as weighted combinations of those bases. Specifically, we use five Fourier basis functions $f_{m_1}^{(1,1)}$, $f_{m_2}^{(1,2)}$ on $[0, 1]$ or $[0, 0.5]$, respectively, to form $M = 25$ tensor product functions $f_m^{(1)}$ on $[0, 1] \times [0, 0.5]$ and $M = 25$ Legendre Polynomials $f_m^{(2)}$ on $[-1, 1]$. The eigenfunctions are defined via

$$\begin{aligned}\psi_m^{(1)}(s, t) &= \sqrt{\alpha} f_{m_1}^{(1,1)}(s) \cdot f_{m_2}^{(1,2)}(t), \quad (s, t) \in \mathcal{T}_1 := [0, 1] \times [0, 0.5], \\ \psi_m^{(2)}(t) &= \sqrt{1 - \alpha} f_m^{(2)}(t), \quad t \in \mathcal{T}_2 := [-1, 1]\end{aligned}$$

with a random weight $\alpha \in (0, 1)$. This choice implies that ψ_m forms an orthonormal system in $\mathcal{H} = L^2(\mathcal{T}_1) \times L^2(\mathcal{T}_2)$. In order to avoid extreme weights, α is set to $u_1/(u_1 + u_2)$ with $u_1, u_2 \sim U(0.2, 0.8)$. This construction restricts $\alpha \in (0.2, 0.8)$ and can easily be generalized to the simulation of multivariate functional data with p elements. Example data based on this type of eigenfunctions is shown in Fig. 8.

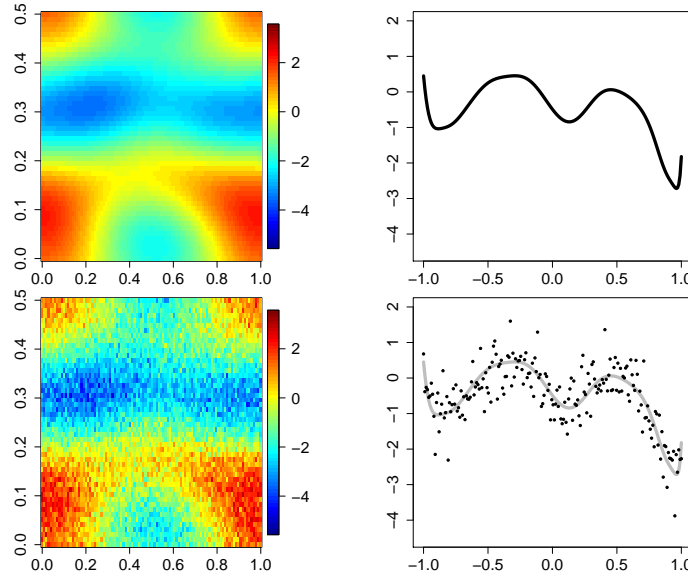


Figure 8: Examples for simulated data in the third simulation setting (cf. Section 4.2) consisting of images ($x_i^{(1)}$, left) and functions ($x_i^{(2)}$, right) without (1st row) and with measurement error (2nd row).

Example Fits

Table 2: True and estimated eigenvalues for the first simulation setting (exponential eigenvalue decay, eigenfunctions based on the first $M = 8$ Fourier basis functions) for one replication with $N = 250$ observations. The reconstruction errors are (in %) 0.007 (MFPCA, $\sigma^2 = 0$; simulation median: 0.008), 0.734 (MFPCA, $\sigma^2 = 0.25$; simulation median: 0.497), $< 10^{-3}$ (MFPCA_{RS}, $\sigma^2 = 0$; simulation median: $< 10^{-3}$) and 0.710 (MFPCA_{RS}, $\sigma^2 = 0.25$; simulation median: 0.480). The results for the corresponding eigenfunctions are given in Fig. 9.

$m =$	1	2	3	4	5	6	7	8
True Eigenvalues	1.000	0.607	0.368	0.223	0.135	0.082	0.050	0.030
MFPCA ($\sigma^2 = 0$)	1.144	0.502	0.316	0.249	0.128	0.090	0.048	0.034
MFPCA ($\sigma^2 = 0.25$)	1.140	0.501	0.316	0.249	0.128	0.087	0.046	0.031
MFPCA _{RS} ($\sigma^2 = 0$)	1.140	0.500	0.315	0.248	0.127	0.090	0.048	0.034
MFPCA _{RS} ($\sigma^2 = 0.25$)	1.139	0.504	0.317	0.252	0.130	0.091	0.048	0.035

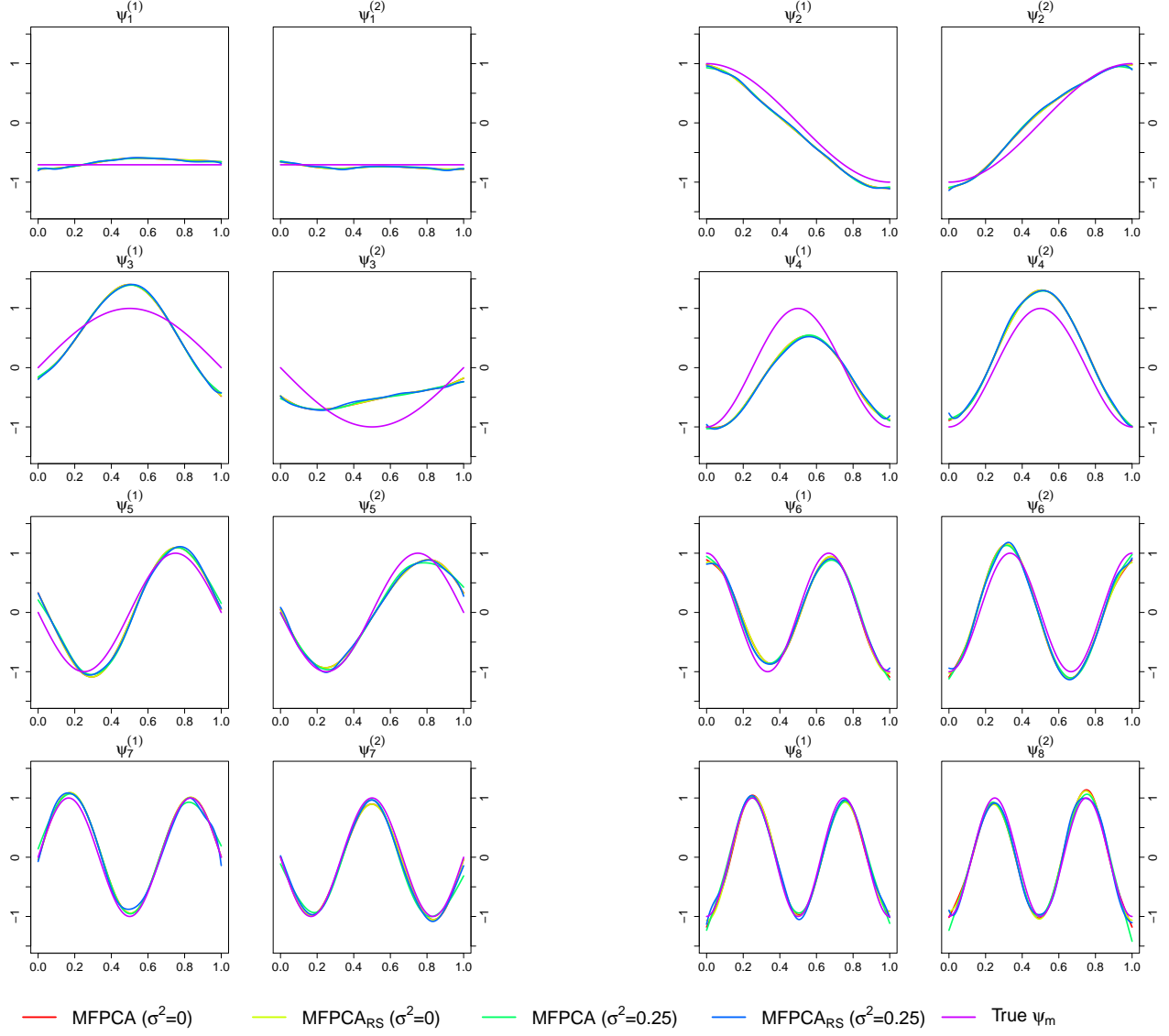
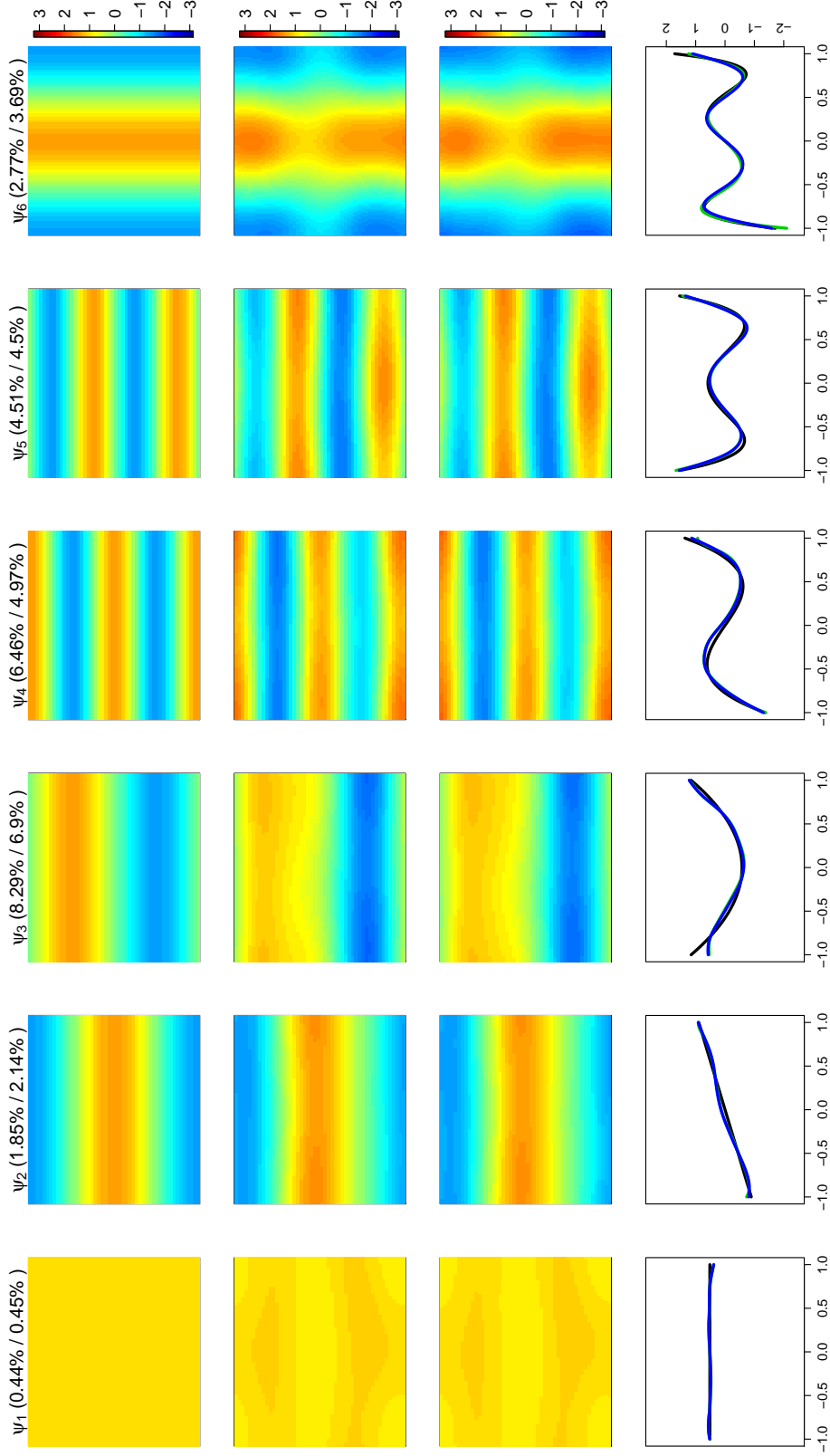


Figure 9: True and estimated eigenfunctions for the first setting based on one example replication with $N = 250$ observations. The results for the corresponding eigenfunctions are given in Table 2.



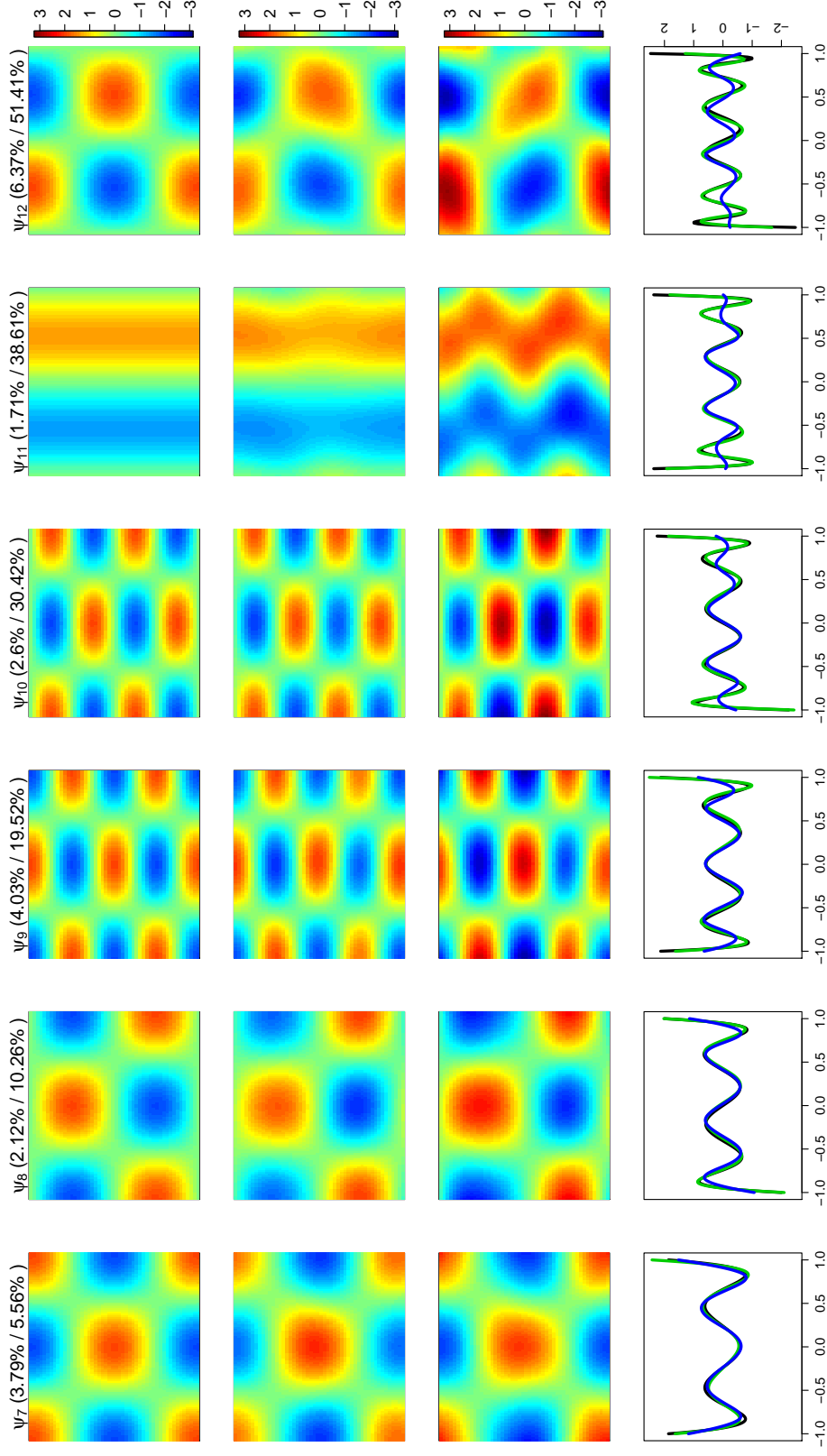


Figure 11: Exemplary result for one replication in simulation setting 3 (results for the eigenfunctions ψ_7, \dots, ψ_{12}). Please refer to Fig. 10 for details.

Sensitivity Analysis

As discussed in Section 3.2, the number M_j of univariate eigenfunctions used for MFPCA clearly has an impact on the results, as they control how much of the information in the univariate elements is used for calculating the multivariate FPCA. A standard approach in functional data analysis for quantifying the amount of information contributed by single eigenfunctions $\phi_m^{(j)}$ is the percentage of variance explained (pve), which is the ratio of the associated eigenvalue $\lambda_m^{(j)}$ and the sum of all eigenvalues. The following simulation systematically examines the sensitivity of the MFPCA result based on the pve of the univariate eigenfunctions.

Simulation Setup: The simulation is based on 100 replications with $N = 250$ observations of bivariate data on the unit interval (cf. setting 1 in Section 4.1), with $M = 8$ Fourier basis functions and exponentially decreasing eigenvalues for simulating the data. The number of univariate eigenfunctions M_1, M_2 for MFPCA is chosen based on $\text{pve} \in \{0.75, 0.90, 0.95, 0.99\}$ for both elements and $M_1 = M_2 = M = 8$ for comparison. The number of multivariate principal component functions is then set to $\min\{M_1 + M_2, M\}$.

Results: The results of the sensitivity analysis are shown in Fig. 12 and Table 3. The number of estimated multivariate eigenvalues/eigenfunctions is for all 100 datasets $\hat{M} = 4$ for $\text{pve} = 0.75$, $\hat{M} = 6$ for $\text{pve} = 0.90$ and $\hat{M} = 8$ in all other cases. The results are as expected: Increasing the pve, and hence the information in the univariate FPCA, improves the estimation accuracy for both, multivariate eigenvalues and eigenfunctions. As a consequence, the reconstruction error reduces with increasing pve. Moreover, for a fixed m , the results show that there is a critical amount of information in univariate FPCA that is needed to describe the multivariate eigenvalues and eigenfunctions well. If this is reached (e.g. $\text{pve} = 0.95$ for $m = 5$, cf. Fig. 12), the additional benefit of using more univariate eigenfunctions ($\text{pve} > 0.95$) becomes negligible. If, in contrast, the univariate FPCA does not contain enough information ($\text{pve} < 0.95$), the error rates for the MFPCA estimates are considerably increased. For fixed pve, the error rates rise abruptly for the last pair of eigenfunctions ($m \in \{\hat{M}_+ - 1, \hat{M}_+\}$). This is due to the fact that in this simulation, the multivariate functional principal components are derived from a Fourier basis. The last two eigenfunctions are hence sine and cosine functions with highest frequency and cannot be represented well by the univariate functions used, as they contain only functions with lower frequency, in other words, they do not contain enough information.

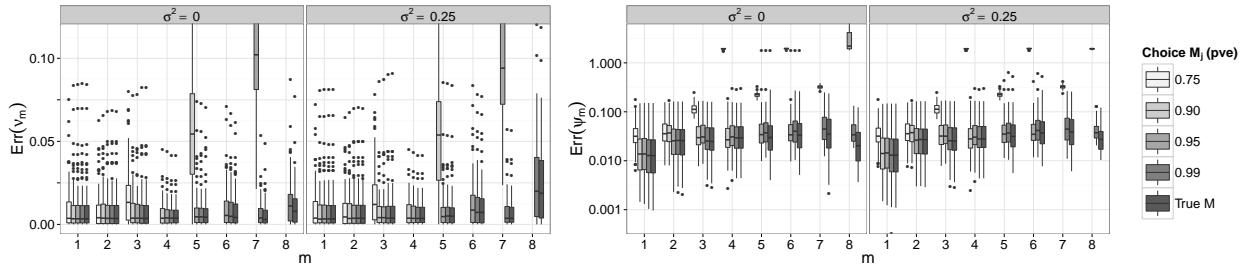


Figure 12: Relative errors for estimated eigenvalues (left) and eigenfunctions (right, log-scale) for the sensitivity analysis. Extreme values cut off for better comparability.

Table 3: Average MRSE (in %) in the sensitivity analysis.

	Choice of M_j (pve)				True M
	0.75	0.90	0.95	0.99	
$\sigma^2 = 0$	23.756	9.075	2.924	0.165	0.006
$\sigma^2 = 0.25$	24.099	9.583	3.593	0.842	0.740

Coverage Analysis of Pointwise Bootstrap Confidence Bands

In Section 5, pointwise bootstrap confidence bands were calculated for the multivariate functional principal components estimated from the ADNI data to quantify the variability in the estimates. The following simulation study examines the coverage properties of such confidence bands.

Simulation Setup: The data generating process is the same as in the simulation in Section 4.2, mimicking the ADNI data that consists of functions on a one-dimensional domain and images. In total, the simulation is based on 100 datasets, all having $N = 250$ observations. Each dataset is considered with and without measurement error. Both elements are represented in terms of B-spline basis functions with appropriate smoothness penalties in the presence of measurement error (cf. Section 4.2). For each dataset and each estimated eigenfunction, a pointwise 95% bootstrap confidence band is calculated based on 100 bootstrap samples on the level of subjects (cf. Section 5). The coefficients of the spline basis decompositions can efficiently be reused when bootstrapping, as the basis is fixed and does not depend on the bootstrap sample. In contrast, the univariate functional principal components for the ADAS-Cog trajectories in the ADNI application have to be re-estimated for each bootstrap sample. This computational aspect is taken into account in the bootstrap implementation in the MFPCA package (Happ, 2016b). Finally, the confidence bands are calculated separately for each element as pointwise percentile bootstrap confidence intervals.

For each eigenfunction and each observation point, the estimated coverage at one point $t_j \in \mathcal{T}_j$ is the percentage of datasets for which the true eigenfunction $\psi_m^{(j)}$ evaluated at t_j is enclosed in the bootstrap confidence band (up to a sign change of the whole function). Fig. 13 shows the estimated coverages of the elements of the eigenfunctions for data with and without measurement error aggregated over the observation points.

Results: If the data is observed without measurement error, the pointwise confidence bands enclose the true functions fairly precisely in 95% of all cases with very little variation between the observation points. For the leading eigenfunctions, the same holds true if the data is observed with measurement error. For higher order eigenfunctions, that explain hardly any variation in the data, the estimated coverage decreases, especially for the second element ($\psi_m^{(2)}$, one-dimensional domain) and shows a much higher variation between the observation points. On the one hand this may be caused by the fact that the true eigenfunctions $\psi_m^{(2)}$ have a stronger curvature for growing m (cf. Fig. 11). Severe undercoverage for higher-order eigenfunctions occurs mainly in regions of high curvature and slope of the eigenfunctions, where the low signal-to-noise level leads to oversmoothing (cf. Fig. 14). On the other hand, the results of Section 4.2 show that the estimates for higher order eigenfunction elements become more inaccurate due to interchanging of eigenfunctions, hence the bootstrap confidence bands can be centered incorrectly. For the image elements $\psi_m^{(1)}$, the bootstrapped confidence bands give much better results, except for some outliers that form spatially smooth outlying regions (see e.g. Fig. 14). This reflects that the pointwise coverages are

not independent, as the true eigenfunctions as well as the confidence bands are smooth: If the function $\psi_m^{(j)}$ lies within the bootstrap confidence band at a point t_j , it is very likely that it will also be inside the confidence band at the neighbouring observation points (analogously for points outside the CI). This relation is highlighted in Fig. 14, which illustrates the coverage rates for ψ_3 (having a good coverage) and ψ_9 (having a rather poor coverage) in the case of measurement error. In summary, the results of the simulation show that the bootstrapped confidence bands give reliable results, in particular for the leading eigenfunctions that explain most of the variation in the data. Moreover, smooth eigenfunctions will have a stabilizing effect for the coverage. However, when interpreting such pointwise confidence bands, one should keep in mind the dependence across neighbouring observation points due to the smoothness of the eigenfunctions.

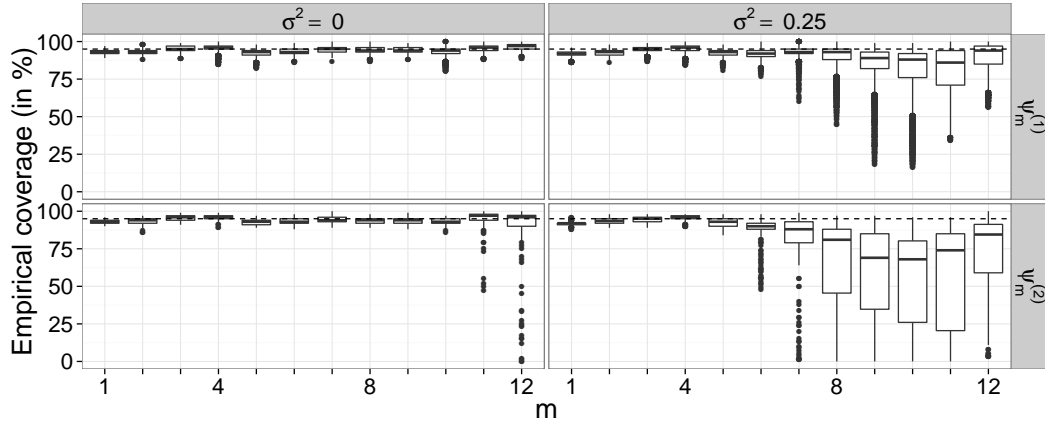


Figure 13: Empirical coverages from the bootstrap simulation study for data without ($\sigma^2 = 0$) and with ($\sigma^2 = 0.25$) measurement error. The boxplots show the pointwise coverage of the bootstrap confidence bands aggregated over the corresponding domains for both elements of the true eigenfunctions ψ_m , $m = 1, \dots, 12$ (1st row: Image element $\psi_m^{(1)}$, 2nd row: Element $\psi_m^{(2)}$ with one-dimensional domain). The dashed line marks a coverage of 95%.

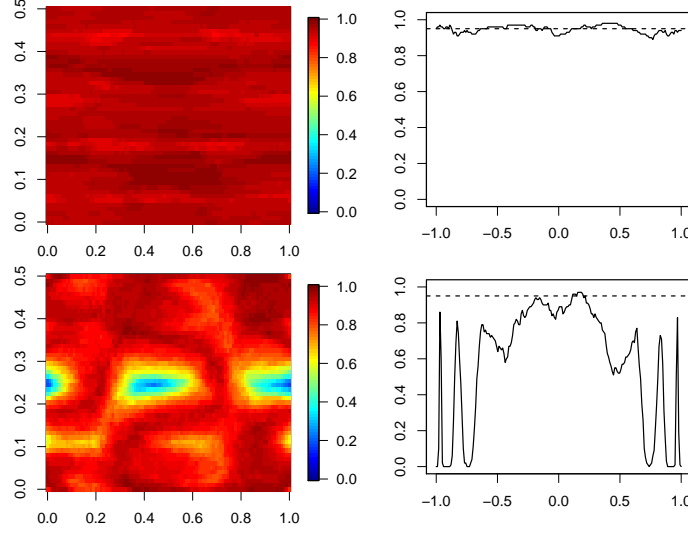


Figure 14: Exemplary results from the bootstrap simulation study for data observed with measurement error. The first row shows the estimated coverages for the third eigenfunction, the second row shows the estimated coverages for the eigenfunction of order 9 (see also Fig. 13). The first column corresponds to the estimated elements $\hat{\psi}_m^{(1)}$ and the second column corresponds to the estimated elements $\hat{\psi}_m^{(2)}$. For the latter, the dashed lines correspond to the nominal level of 95%.

Applications – Gait Cycle Data

For comparison to an existing method in the special case of densely sampled bivariate data on the same one-dimensional interval, the new MFPCA approach is applied to the gait cycle data (cf. Fig. 1 in the main document) and compared to the method of Ramsay and Silverman (2005) as implemented in the R-package `fda` (Ramsay et al., 2014). The results are shown in Fig. 15. For the new approach, the multivariate principal components are calculated based on univariate FPCA with $M_1 = M_2 = 5$ principal components. For MFPCA_{RS}, the observed functions are pre-smoothed using $K = 15$ cubic spline basis functions as in the simulation study (cf. Section 4.1). As for synthetic data, the two methods give nearly identical results.

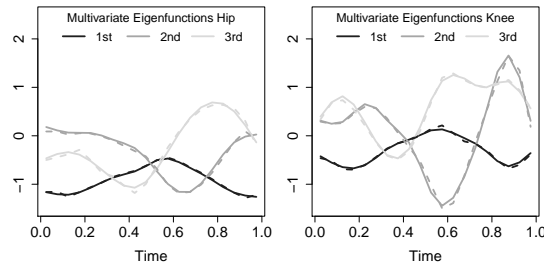


Figure 15: The first three estimated bivariate eigenfunctions for the gait data set. Solid lines show the results of the new MFPCA approach, dashed lines correspond to the approach of Ramsay and Silverman (2005). The functions have been reflected, if necessary, for comparison purposes.