Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: http://amstat.tandfonline.com/loi/uasa20

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

Clara Happ & Sonja Greven

To cite this article: Clara Happ & Sonja Greven (2018) Multivariate Functional Principal Component Analysis for Data Observed on Different (Dimensional) Domains, Journal of the American Statistical Association, 113:522, 649-659, DOI: 10.1080/01621459.2016.1273115

To link to this article: https://doi.org/10.1080/01621459.2016.1273115

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2018, VOL. 113, NO. 522, 649–659, Theory and Methods
https://doi.org/10.1080/01621459.2016.1273115

Taylor & Francis
Taylor & Francis Group

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

Clara Happ and Sonja Greven for the Alzheimer's Disease Neuroimaging Initiative

Department of Statistics, LMU Munich, München, Germany

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, such as 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.

ARTICLE HISTORY
Received September 2015
Revised December 2016

KEYWORDS
Dimension reduction; Functional data analysis; Image analysis; Multivariate functional data

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, for example, joint variation in the different elements.

As a simple motivating example, consider the gait cycle data (Ramsay and Silverman 2005) shown in Figure 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 nonnegligible 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, for example, 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 Figure 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; Berrendero, Justel, and Svarc 2011; Jacques and Preda 2014; Chiou, Yang, and Chen 2014). Except for Berrendero, Justel, and Svarc (2011), they all aim at a multivariate functional Karhunen–Loève representation of the data. For data measured, for example, in different units, Jacques and Preda (2014) and Chiou, Yang, and Chen (2014) also discussed normalized versions of MFPCA based on a normalized covariance operator.

The key motivation for this article is that in practical applications, multivariate functional data are neither restricted to lie on the same interval nor to have one-dimensional domains, for

CONTACT Clara Happ clara.happ@stat.uni-muenchen.de Department of Statistics, LMU Munich, 80539 München, Germany.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

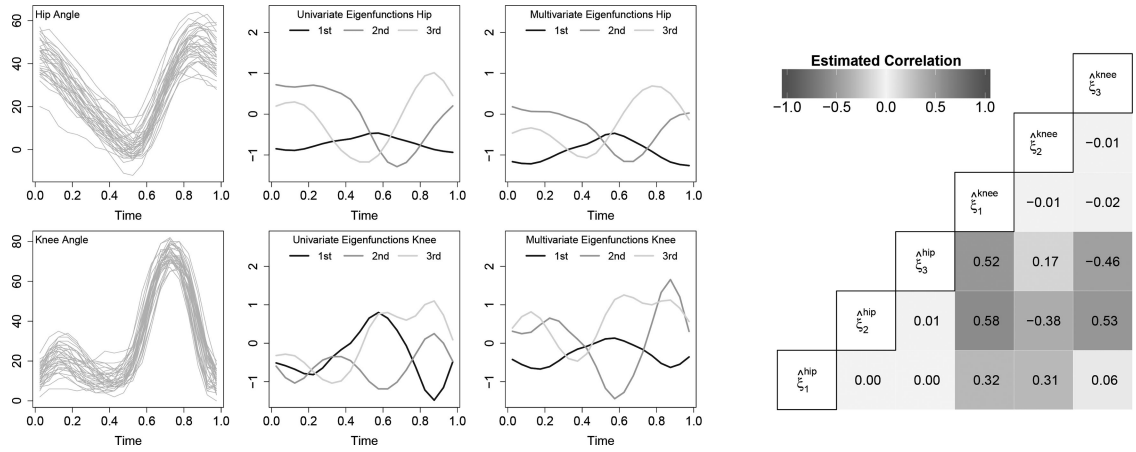

Figure 1. Univariate and multivariate FPCA for the gait cycle data. 1st column: Original data. 2nd column: Results for univariate FPCA, calculated separately. The functions have been reflected, if necessary, and rescaled to have the same norm as the multivariate eigenfunctions for comparison purposes.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

example, data that consist 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 learning literature (Coppi and Bolasco 1989; Lu, Plataniotis, and Venetsanopoulos 2013). Recent approaches for functional or smooth principal component analysis for tensor data have been proposed, for example, 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 article 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 article 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 article is concerned with *multivariate functional data*, that is, each observation consists of $p \geq 2$ functions $X^{(1)}, \ldots, X^{(p)}$. They may be defined on different domains $\mathcal{T}_1, \ldots, \mathcal{T}_p$ 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 $X^{(j)} : \mathcal{T}_j \to \mathbb{R}$ 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

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

Note that $\boldsymbol{t} := (t_1, \ldots, t_p) \in \mathcal{T} := \mathcal{T}_1 \times \cdots \times \mathcal{T}_p$ is a p-tuple of $d_1, \ldots, d_p$-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

$$\mu(\boldsymbol{t}) := \mathbb{E}\left(X(\boldsymbol{t})\right) = \left(\mathbb{E}\left(X^{(1)}(t_1)\right), \ldots, \mathbb{E}\left(X^{(p)}(t_p)\right)\right) = \boldsymbol{0}$$

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

$$C_{ij}(s_i, t_j) := \mathbb{E}\left(X^{(i)}(s_i)X^{(j)}(t_j)\right) = \operatorname{cov}(X^{(i)}(s_i), X^{(j)}(t_j)) \quad (1)$$

with $s_i \in \mathcal{T}_i$, $t_j \in \mathcal{T}_j$.

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

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

for $f, g \in \mathcal{H}$.

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 $|||\cdot|||$. The 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 $\|\boldsymbol{t}\|_{\mathcal{T}}^2 = \sum_{j=1}^p \|t_j\|^2$ for $t_j \in \mathcal{T}_j \subset \mathbb{R}^{d_j}$, $j = 1,\ldots,p$. Next, define the covariance operator $\Gamma : \mathcal{H} \to \mathcal{H}$ with the jth element of Γf, $f \in \mathcal{H}$ given by

$$\begin{aligned}(\Gamma f)^{(j)}(t_j) &:= \sum_{i=1}^{p} \int_{\mathcal{T}_i} C_{ij}(s_i, t_j) f^{(i)}(s_i) ds_i \\ &= \langle\langle C_{\cdot j}(\cdot, t_j), f\rangle\rangle \end{aligned} \tag{3}$$

for $t_j \in \mathcal{T}_j$.

The setting can be generalized to a weighted scalar product on $\mathcal{H}$, that is,

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

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

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

The use of weights may be necessary if the elements have quite different domains or ranges or if they exhibit different amounts of variation, to obtain multivariate functional principal components that have a meaningful interpretation (Chiou, Yang, and Chen 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, for example 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, Yang, and Chen 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, for example based on the variation in each element (see references in Chiou, Yang, and Chen 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 $w_1 = \cdots = w_p = 1$, 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 $i, j = 1,\ldots,p$, there exist $K_{ij} < \infty$ with

$$\left\|C_{ij}(\cdot, t_j)\right\|_2^2 = \int_{\mathcal{T}_i} C_{ij}(s_i, t_j)^2 ds_i < K_{ij} \tag{5}$$

for all, $t_j \in \mathcal{T}_j$, and C_{ij} is uniformly continuous in the sense that

$$\begin{aligned}\forall\, \varepsilon > 0\, \exists\, \delta_{ij} > 0 : \|t_j - t_j^*\| < \delta_{ij} \\ \Rightarrow |C_{ij}(s_i, t_j) - C_{ij}(s_i, t_j^*)| < \varepsilon\end{aligned}$$

for all, $s_i \in \mathcal{T}_i$, then Γ is a compact operator.

In the remainder of this article, it is assumed that the C_{ij} satisfy all conditions of Proposition 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

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

In particular, since Γ is a positive operator, it may be assumed w.l.o.g. that $\nu_1 \geq \nu_2 \geq \cdots \geq 0$. 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

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

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 (Proposition 4).

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

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

where the convergence is absolute and uniform.

Proposition 4 (Multivariate Karhunen–Loève Theorem). Under the assumptions of Proposition 2,

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

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

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

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

The multivariate Karhunen–Loève representation has an analogous interpretation as in the univariate case (Ramsay and Silverman 2005, chap. 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 toward 0, leading eigenfunctions reflect the most important features of X. Truncated Karhunen–Loève expansions, optimal M-dimensional approximations to X,

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

are often used in practice. Single observations x_i of X can then be characterized by their score vectors $(\rho_{i,1}, \ldots, \rho_{i,M})$ with $\rho_{i,m} = \langle\langle x_i, \psi_m \rangle\rangle$ for further analysis, for example, 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 (Zemyan 2012).

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

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

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

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

$$\phi_m^{(j)}(t_j) = \left(\lambda_m^{(j)}\right)^{-1/2} \sum_{n=1}^{M} \nu_n^{1/2} \left[\boldsymbol{u_m^{(j)}}\right]_n \psi_n^{(j)}(t_j)$$

for $t_j \in \mathcal{T}_j$ and $m = 1, \ldots, M_j$,
where $\boldsymbol{u_m^{(j)}}$ denotes an (orthonormal) eigenvector of $\boldsymbol{A^{(j)}}$ associated with eigenvalue $\lambda_m^{(j)}$ and $[\boldsymbol{u_m^{(j)}}]_n$ denotes the nth entry of this vector. For the univariate scores,

$$\begin{aligned}\xi_m^{(j)} &= \left\langle X^{(j)}, \phi_m^{(j)} \right\rangle_2 \\ &= \left(\lambda_m^{(j)}\right)^{-1/2} \sum_{n=1}^{M} \nu_n^{1/2} \left[\boldsymbol{u_m^{(j)}}\right]_n \sum_{k=1}^{M} \rho_k \left\langle \psi_n^{(j)}, \psi_k^{(j)} \right\rangle_2.\end{aligned}$$

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 $\nu_1 \geq \cdots \geq \nu_M > 0$ of Γ with $M \leq \sum_{j=1}^{p} M_j =: M_+$ correspond to the positive eigenvalues of the matrix $\boldsymbol{Z} \in \mathbb{R}^{M_+ \times M_+}$ consisting of blocks $\boldsymbol{Z^{(jk)}} \in \mathbb{R}^{M_j \times M_k}$ with entries

$$Z_{mn}^{(jk)} = \text{cov}\left(\xi_m^{(j)}, \xi_n^{(k)}\right)$$

for $m = 1, \ldots, M_j$, $n = 1, \ldots, M_k$ and $j, k = 1, \ldots, p$. The eigenfunctions of Γ are given by their elements

$$\psi_m^{(j)}(t_j) = \sum_{n=1}^{M_j} [\boldsymbol{c_m}]_n^{(j)} \phi_n^{(j)}(t_j), \quad t_j \in \mathcal{T}_j, \; m = 1, \ldots, M,$$

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

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

Extensions: The second part of Proposition 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 $\theta_m^{(j)}$, that is,

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

This is a very likely situation in practice, for example, due to presmoothing of noisy observations. Following analogous steps as in the proof of Proposition 5 results in an eigenanalysis problem $\boldsymbol{BQc} = \nu \boldsymbol{c}$ as starting point for the MFPCA. Here, $\boldsymbol{B} \in \mathbb{R}^{K_+ \times K_+}$ 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 Proposition 5), $\boldsymbol{B}$ equals the identity matrix. The symmetric block matrix $\boldsymbol{Q}$ with entries $Q_{mn}^{(jk)} = \text{cov}(\theta_m^{(j)}, \theta_n^{(k)})$ corresponds to $\boldsymbol{Z}$ in Proposition 5. Although $\boldsymbol{BQ}$ is in general not symmetric, its eigenvectors $\boldsymbol{c}_m$ and eigenvalues ν_m, which are at the same time the nonnegative eigenvalues of Γ, are real. This can be easily shown using the Cholesky decomposition of the symmetric matrix $\boldsymbol{B} = \boldsymbol{R}\boldsymbol{R}^\top$ and solving $\boldsymbol{R}^\top \boldsymbol{Q} \boldsymbol{R} \tilde{\boldsymbol{c}} = \nu \tilde{\boldsymbol{c}}$ with $\tilde{\boldsymbol{c}} = \boldsymbol{R}^{-1}\boldsymbol{c}$. 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 Proposition 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 formulas given in Proposition 5. Given de-meaned samples $x_1, \ldots, x_N$ 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 $x_1^{(j)}, \ldots, x_N^{(j)}$. This results in estimated eigenfunctions $\hat{\phi}_m^{(j)}$ and scores $\hat{\xi}_{i,m}^{(j)}$, $i = 1, \ldots, N$, $m = 1, \ldots, M_j$ for suitably chosen truncation lags M_j. As there exist numerous estimation procedures, such as for irregularly sampled and sparse data with measurement error (Yao, Müller, and Wang 2005), the multivariate method is also applicable to this kind of data.
2. Define the matrix $\boldsymbol{\Xi} \in \mathbb{R}^{N \times M_+}$, where each row $(\hat{\xi}_{i,1}^{(1)}, \ldots, \hat{\xi}_{i,M_1}^{(1)}, \ldots, \hat{\xi}_{i,1}^{(p)}, \ldots, \hat{\xi}_{i,M_p}^{(p)})$ contains all estimated scores for a single observation. An estimate $\hat{\boldsymbol{Z}} \in \mathbb{R}^{M_+ \times M_+}$ of the block matrix $\boldsymbol{Z}$ in Proposition 5 is given by $\hat{\boldsymbol{Z}} = (N-1)^{-1} \boldsymbol{\Xi}^\top \boldsymbol{\Xi}$.
3. Perform a matrix eigenanalysis for $\hat{\boldsymbol{Z}}$ resulting in eigenvalues $\hat{\nu}_m$ and orthonormal eigenvectors $\hat{\boldsymbol{c}}_{\boldsymbol{m}}$.
4. Estimates for the multivariate eigenfunctions are given by their elements

$$\hat{\psi}_m^{(j)}(t_j) = \sum_{n=1}^{M_j} [\hat{\boldsymbol{c}}_{\boldsymbol{m}}]_n^{(j)} \hat{\phi}_n^{(j)}(t_j) \tag{9}$$

for $t_j \in \mathcal{T}_j$, and $m = 1, \ldots, M^+$ and multivariate scores can be calculated via

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

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 $\hat{\lambda}_m^{(j)}$ (screeplot, Cattell 1966) or the percentage of variance explained (e.g., Ramsay and Silverman 2005, chap. 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, such as clustering or functional principal component regression. For the latter, relevant eigenfunctions can also be selected using model-based approaches such as Akaike information criterion (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 $w_1, \ldots, w_p > 0$ and demeaned observations $x_1, \ldots, x_N$ of X with estimated basis function coefficients $\hat{\theta}_{i,m}^{(j)}$ for each element, the eigenanalysis problem to solve is

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

The matrix $\boldsymbol{B}$ is the block diagonal matrix of basis scalar products as in Section 3.1 and $\boldsymbol{D} = \text{diag}(\boldsymbol{w_1}^{1/2}, \ldots, \boldsymbol{w_p}^{1/2}) \in \mathbb{R}^{K_+ \times K_+}$ accounts for the weights, where each $w_j^{1/2}$ is repeated K_j times to give $\boldsymbol{w_j}^{1/2}$. $\boldsymbol{\Theta} \in \mathbb{R}^{N \times K_+}$ with rows $(\hat{\theta}_{i,1}^{(1)}, \ldots, \hat{\theta}_{i,K_1}^{(1)}, \ldots, \hat{\theta}_{i,1}^{(p)}, \ldots, \hat{\theta}_{i,K_p}^{(p)})$ corresponds to the matrix $\boldsymbol{\Xi}$ defined in step 2 of the original algorithm and $(N-1)^{-1} \boldsymbol{\Theta}^\top \boldsymbol{\Theta}$ is an estimate for $\boldsymbol{Q}$ introduced in Section 3.1. Given eigenvectors $\hat{\boldsymbol{c}}_{\boldsymbol{m}}$ 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 $\hat{\boldsymbol{Q}}_{\boldsymbol{w}} = (N-1)^{-1} \boldsymbol{D} \boldsymbol{\Theta}^\top \boldsymbol{\Theta} \boldsymbol{D}$:

$$\hat{\psi}_m^{(j)}(t_j) = (w_j \cdot \hat{\nu}_m \hat{\boldsymbol{c}}_{\boldsymbol{m}}^\top \hat{\boldsymbol{Q}}_{\boldsymbol{w}} \hat{\boldsymbol{c}}_{\boldsymbol{m}})^{-1/2} \sum_{k=1}^{p} \sum_{l=1}^{K_j} \sum_{n=1}^{K_k} [\hat{\boldsymbol{Q}}_{\boldsymbol{w}}]_{ln}^{(jk)} [\hat{\boldsymbol{c}}_{\boldsymbol{m}}]_n^{(k)} b_l^{(j)}(t_j),$$

$$\hat{\rho}_{i,m} = (\hat{\nu}_m)^{1/2} (\hat{\boldsymbol{c}}_{\boldsymbol{m}}^\top \hat{\boldsymbol{Q}}_{\boldsymbol{w}} \hat{\boldsymbol{c}}_{\boldsymbol{m}})^{-1/2} \boldsymbol{\Theta}_{\boldsymbol{i},\cdot} \boldsymbol{D} \hat{\boldsymbol{c}}_{\boldsymbol{m}}.$$

Clearly, the original algorithm is obtained as a special case with $\boldsymbol{\Theta} = \boldsymbol{\Xi}$, $\boldsymbol{B} = \boldsymbol{I}$ (univariate FPCA for each element) and $\boldsymbol{D} = \boldsymbol{I}$ (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 $\boldsymbol{D} = \boldsymbol{I}$, expanding each element in a general basis is equivalent to the method of Jacques and Preda (2014). The approach proposed in this article, 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 2017b). 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 2017a) for representing (multivariate) functional data on potentially different dimensional domains. An introduction to the software including a case study with real data can be found in Happ (2017c).

3.3. Asymptotic Properties

The results of Proposition 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 are observable only in finite form (finitely many observations, finite resolution) and hence contain only finite information. In

this case, the maximal number of principal components that 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:

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

Here, ψ_m is the true mth 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 mth 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 $M_1, \ldots, M_p$. 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 $|||X^{[M]} - X|||$ converges to 0 in probability for $M_1, \ldots, M_p \to \infty$. 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 $\langle\langle \psi_m, \psi_m^{[M]} \rangle\rangle \geq 0$, then

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

The scores $\rho_m^{[M]} := \langle\langle X^{[M]}, \psi_m^{[M]} \rangle\rangle$ converge to ρ_m in probability 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 $\langle\langle \psi_m, \psi_m^{[M]} \rangle\rangle \geq 0$.

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 $\phi_m^{(j)}$, $m = 1, \ldots, M_j$. In the following, let $X_1, \ldots, X_N$ be independent copies of X and assume for all $j = 1, \ldots, p$

$$\Delta_{M_j}^{(j)} := \sup_{m=1,\ldots,M_j} \left(\lambda_m^{(j)} - \lambda_{m+1}^{(j)}\right)^{-1} < \infty \quad \forall M_j < \infty \tag{12}$$

$$\int_{\mathcal{T}_j} \int_{\mathcal{T}_k} \mathbb{E}\left(X^{(j)}(t_j)^2 X^{(k)}(s_k)^2\right) ds_k \, dt_j < \infty \quad \forall\, k = 1, \ldots, p \tag{13}$$

$$\left\| \Gamma^{(j)} - \hat{\Gamma}^{(j)} \right\|_{\text{op}} = O_p(r_N^\Gamma) \tag{14}$$

$$\langle \phi_m^{(j)}, \hat{\phi}_m^{(j)} \rangle_2 \geq 0 \quad \text{for all } m = 1, \ldots, M_j \tag{15}$$

$$\hat{\xi}_{i,m}^{(j)} = \langle X_i^{(j)}, \hat{\phi}_m^{(j)} \rangle_2 \quad \text{for all } m = 1, \ldots, M_j,\ i = 1, \ldots, N \tag{16}$$

(12)–(13) concern theoretical properties of $X^{(j)}$ and $\Gamma^{(j)}$, while (14)–(16) depend on the univariate decompositions used. (12) 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 (13), 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 $N^{-1/2}$. (13) is used in combination with (16) to obtain a convergence rate for the maximal eigenvalue of $\boldsymbol{Z} - \hat{\boldsymbol{Z}}$, which, in turn, affects the convergence of the eigenvectors $\hat{\boldsymbol{c}}_{\boldsymbol{m}}$ to $\boldsymbol{c}_{\boldsymbol{m}}$ (Yu, Wang, and Samworth 2015). (14) 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) showed $r_N^\Gamma = N^{-1/2}$, while the approach of Yao, Müller, and Wang (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 (12), r_N^Γ gives a convergence rate for the univariate eigenfunctions $\hat{\phi}_m^{(j)}$ (Bosq 2000, Lemma 4.3). (15) 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, (16) 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., Yao, Müller, and Wang 2005), convergence results can still be obtained, for example by assuming a convergence rate for $\hat{\xi}_{i,m}^{(j)}$ and replacing (13) by an assumption on the rate of convergence for the maximal eigenvalue of $\boldsymbol{Z} - \hat{\boldsymbol{Z}}$.

Proposition 7 (Estimation Error). Assume (12)–(16) hold. Then for $M_{\max} = \max_{j=1,\ldots,p} M_j$ and $\Delta_M := \max_{j=1,\ldots,p} \Delta_{M_j}^{(j)}$, the maximal eigenvalue of $\boldsymbol{Z} - \hat{\boldsymbol{Z}}$ can be characterized by

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

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

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

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 Propositions 6 and 7, the analogy to bias and variance again becomes apparent: For fixed N, higher values of $M_1, \ldots, M_p$ 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, $M_1 = \cdots = M_p = M_{\max} = O(N^\beta)$, $r_N^\Gamma = N^{-1/2}$, and that the eigengaps fulfill $\lambda_m^{(j)} - \lambda_{m+1}^{(j)} \geq C^{-1} m^{-\alpha-1}$ with $\alpha > 1$, $C > 0$

(see 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 (see Sections 3.2 and 5). Unless specified otherwise, the `MFPCA` package (Happ 2017b) was used for all calculations.

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

$$x_i(\boldsymbol{t}) = \sum_{m=1}^{M} \rho_{i,m}\psi_m(\boldsymbol{t}) + \boldsymbol{\varepsilon}_i(\boldsymbol{t}), \quad \boldsymbol{\varepsilon}_i(\boldsymbol{t}) \overset{\text{iid}}{\sim} \mathrm{N}_p(0, \sigma^2\boldsymbol{I}),$$

for $\boldsymbol{t} \in \mathcal{T}$ and $i = 1, \ldots, N$.

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 $\mathrm{N}(0, \nu_m)$ for eigenvalues with exponential ($\nu_m^{\exp} = \exp(-(m+1)/2)$) 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 $\mathrm{Err}(\hat{\nu}_m) = (\nu_m - \hat{\nu}_m)^2/\nu_m^2$ and $\mathrm{Err}(\hat{\psi}_m) = \|\|\psi_m - \hat{\psi}_m\|\|^2$. As functional principal components are defined only up to a sign change, the estimate $\hat{\psi}_m$ is reflected, that is, multiplied by -1, if $\langle\langle \psi_m, \hat{\psi}_m \rangle\rangle < 0$. 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 $\mathrm{MRSE} = N^{-1}\sum_{i=1}^{N}(\|\|x_i - \hat{x}_i\|\|^2/\|\|x_i\|\|^2)$.

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, Müller, and Wang 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 by Ramsay and Silverman (2005), which is implemented in the R package `fda` (Ramsay et al. 2014) and in the following denoted by $\mathrm{MFPCA}_{\mathrm{RS}}$. This method involves presmoothing of the elements with $K = 15$ cubic spline basis functions. $\mathrm{MFPCA}_{\mathrm{RS}}$ 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 (Proposition 4), $\sum_{j=1}^{p} \hat{\rho}_{i,m}^{(j)} = \hat{\rho}_{i,m}$ is used for comparison purposes.

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

Setting		$\sigma^2 = 0$		$\sigma^2 = 0.25$	
		$\nu_m^{\exp}$	ν_m^{lin}	$\nu_m^{\exp}$	ν_m^{lin}
1	MFPCA	0.006	0.009	0.740	0.355
1	$\mathrm{MFPCA}_{\mathrm{RS}}$	$< 10^{-3}$	$< 10^{-3}$	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

The results for the first setting are shown in Figure 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 as well as 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, that is, $\hat{\psi}_m$ is an estimate for ψ_{m-1} or ψ_{m+1}, say, rather than for ψ_m. In the ideal case of no measurement error, $\mathrm{MFPCA}_{\mathrm{RS}}$ yields lower MRSE values than the new approach, which might be an effect of $\mathrm{MFPCA}_{\mathrm{RS}}$ expecting smooth or presmoothed data. For the practically relevant case of data with measurement error, both methods give almost the same prediction errors (see 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 first $M = 8$ Fourier basis functions on $[0, 2]$ into missing $p = 3$ 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, Müller, and Wang (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 Figure 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,

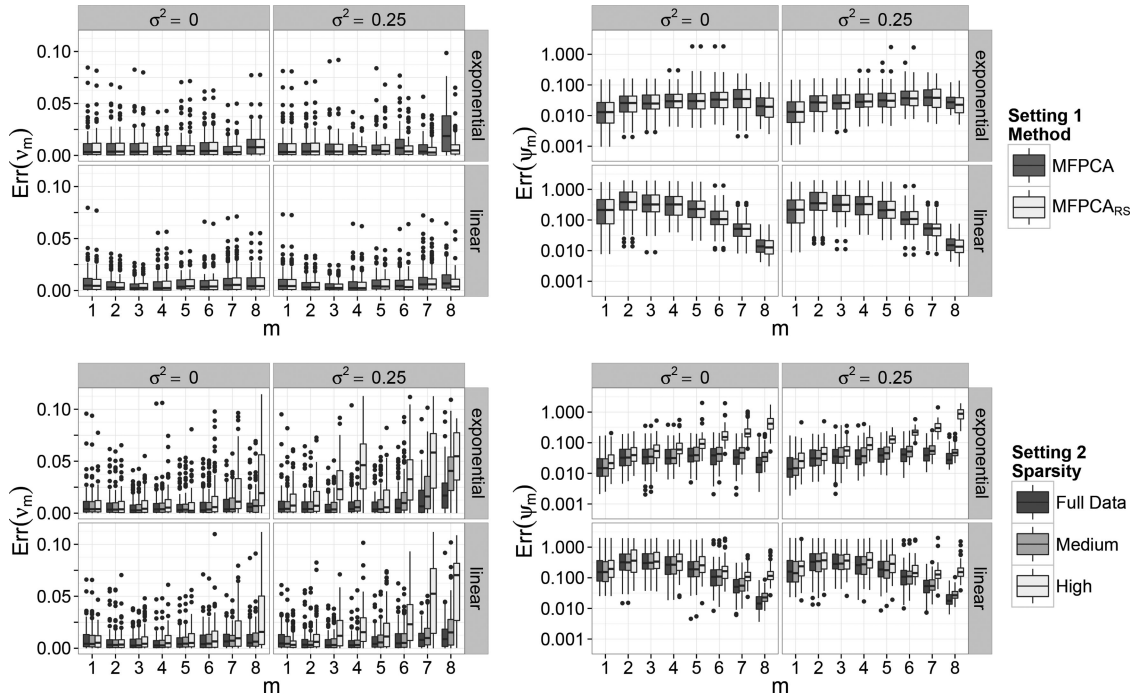

Figure 2. Relative errors for estimated eigenvalues (left) and eigenfunctions (right, log-scale) for simulation settings 1 and 2, depending on eigenvalue decrease, measurement error, and estimation method (setting 1) or sparsity (setting 2).

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 contain 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 nonorthogonal 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 $[10^{-5}, 10^5]$ (Huang, Shen, and Buja 2009; Allen 2013). 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).

The overall results for the first $M = 12$ eigenvalue/eigenvector pairs are given in Figure 3. Compared to the settings with one-dimensional domains, 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 FPCA based 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

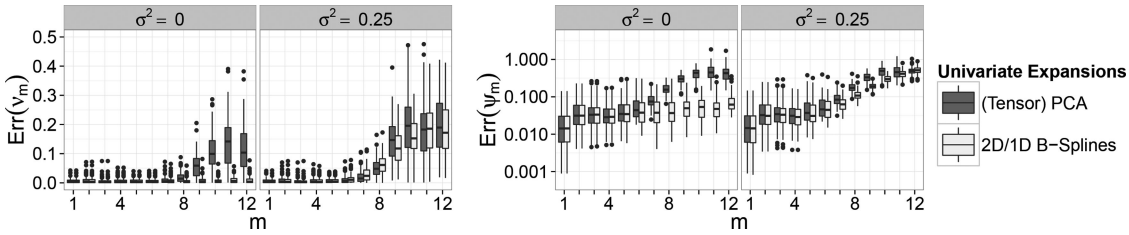

Figure 3. Relative errors for estimated eigenvalues (left) and eigenfunctions (right, log-scale) for simulation setting 3, depending on measurement error and univariate expansions.

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 mm^2) 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, Leon, and Zetterberg 2006). Exemplary data are shown in Figure 4.

Weighted scalar product: As the ADAS-Cog trajectories and FDG-PET scans differ considerably in domain, range, and variation (see Figure 4), we use a weighted MFPCA with

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

for $j = 1, 2$, 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 are 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 site 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 Figure 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 $X^{(2)}$. Figure 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 min 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, that is, 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

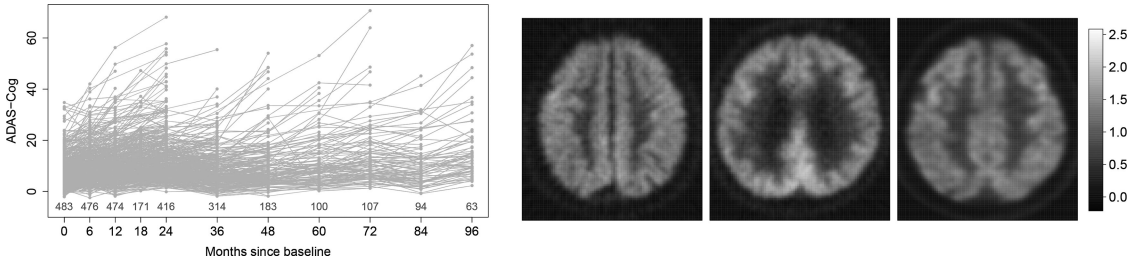

Figure 4. Left: ADAS-Cog trajectories for all $N = 483$ subjects. Numbers above the x-axis give the total number of measurements for each visit. Right: FDG-PET scans for three randomly chosen male subjects (left to right: AD, MCI, normal; diagnosis at baseline).

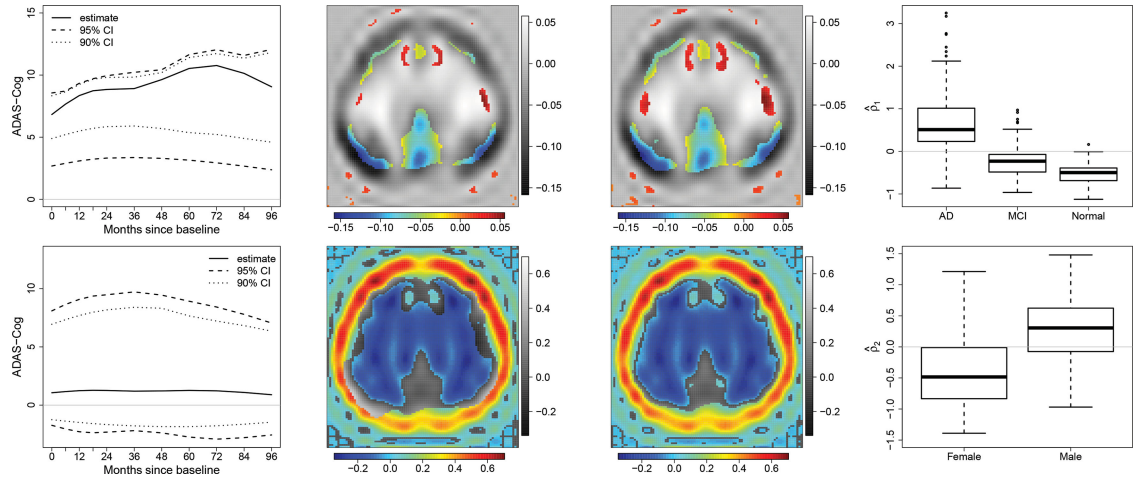

Figure 5. The first two estimated multivariate functional principal components for the ADNI data (1st row: $\hat{\psi}_1$, 2nd row: $\hat{\psi}_2$). Estimates are given with pointwise 95% and 90% bootstrap confidence bands based on 100 nonparametric bootstrap iterations (ADAS-Cog, 1st column: Dashed 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.

zero during the whole follow-up. In contrast, the FDG-PET element differs significantly from zero in almost all voxels (see Figure 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 Figure 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 (see Figure 4). While addressing informative missingness goes beyond the scope of this article, interpretation of results should take this possibility into account. For instance, it is easily conceivable that $\hat{\psi}_1^{(1)}$ may be underestimating $\psi_1^{(1)}$ toward the end of the study period.

6. Discussion and Outlook

This article 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 (Proposition 4), (2) an explicit relation between multivariate and univariate FPCA, which serves as a starting point for the estimation (Proposition 5), and (3) asymptotic results for the estimators (Propositions 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 (see 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 by Jacques and Preda (2014) or Chiou, Yang, and Chen (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*, that is, data consisting of a functional and a vector part (Ramsay and Silverman 2005, chap. 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 Proposition 1. However, transferring the results for MFPCA shown in this article 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 Materials

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 2017a, 2017b).

Acknowledgments

Data used in preparation of this article were obtained from the Alzheimer's 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 article. 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*. A detailed list of ADNI funding is available at *http://adni.loni.usc.edu/about/funding/*.

Funding

The authors acknowledge support from the German Research Foundation through Emmy Noether grant GR 3793/1-1 and would like to thank the LMUMentoring program for financial support to cover the printing costs. 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). 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.

References

Allen, G. (2013), Multi-way Functional Principal Components Analysis, in *IEEE 5th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pp. 220–223. [650,656,658]

Berrendero, J., Justel, A., and Svarc, M. (2011), "Principal Components for Multivariate Functional Data," *Computational Statistics & Data Analysis*, 55, 2619–2634. [649]

Blennow, K., Leon, M., and Zetterberg, H. (2006), "Alzheimer's Disease," *Lancet*, 368, 387–403. [657]

Bosq, D. (2000), *Linear Processes in Function Spaces*, New York: Springer. [654]

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, 283–319. [650]

Cattell, R. (1966), "The Scree Test for the Number of Factors," *Multivariate Behavioral Research*, 1, 245–276. [653]

Chiou, J.-M., Yang, Y.-F., and Chen, Y.-T. (2014), Multivariate Functional Principal Component Analysis: A Normalization Approach," *Statistica Sinica*, 24, 1571–1596. [649,651,658]

Coppi, R., and Bolasco, S. (eds.) (1989), *Multiway Data Analysis*, Amsterdam: North-Holland. [650]

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*. [655]

Eilers, P., and Marx, B. (1996), "Flexible Smoothing with B-splines and Penalties," *Statistical Science*, 11, 89–121. [656]

Hall, P., and Horowitz, J. L. (2007), "Methodology and Convergence Rates for Functional Linear Regression," *Annals of Statistics*, 35, 70–91. [654]

Hall, P., and Hosseini-Nasab, M. (2006), "On Properties of Functional Principal Components Analysis," *Journal of the Royal Statistical Society*, Series B, 68, 109–126. [654]

Happ, C. (2017a), *funData: An S4 Class for Functional Data in R, R Package Version 1.0*. [653,659]

——— (2017b), *MFPCA: Multivariate Functional Principal Component Analysis, R Package Version 1.0-1*. [653,655,659]

——— (2017c), "Object-Oriented Software for Functional Data," *arXiv 1707.02129*. [653]

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, 1609–1620. [656]

Jacques, J., and Preda, C. (2014), "Model-Based Clustering for Multivariate Functional Data," *Computational Statistics & Data Analysis*, 71, 92–106. [649,651,652,653,658]

Lu, H., Plataniotis, K. N., and Venetsanopoulos, A. (2013), *Multilinear Subspace Learning: Dimensionality Reduction of Multidimensional Data*, Boca Raton, FL: CRC Press. [650]

Mercer, J. (1909), "Functions of Positive and Negative Type, and their Connection with the Theory of Integral Equations," *Philosophical Transactions A*, 209, 415–446. [651]

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, 55–66. [656]

Müller, H. G., and Stadtmüller, U. (2005), "Generalized Functional Linear Models," *Annals of Statistics*, 33, 774–805. [649,652,658]

Ramsay, J., and Silverman, B. (2005), *Functional Data Analysis* (2nd ed.), New York: Springer. [649,650,651,652,653,655,658]

Ramsay, J., Wickham, H., Graves, S., and Hooker, G. (2014), *fda: Functional Data Analysis, R Package Version 2.4.4*. [655]

Reed, M., and Simon, B. (1980), *Methods of Modern Mathematical Physics. I: Functional Analysis* (rev. and enl. ed.), San Diego, CA: Academic Press. [651]

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. [658]

Saporta, G. (1981), "Méthodes Exploratoires d'analyse de Données Temporelles," Ph.D. dissertation, Université Pierre et Marie Curie, Paris. [651]

Tucker, L. R. (1966), "Some Mathematical Notes on Three-Mode Factor Analysis," *Psychometrika*, 31, 279–311. [650]

Werner, D. (2011), *Funktionalanalysis* (7th ed.), Berlin: Springer. [651]

Yao, F., Müller, H.-G., and Wang, J.-L. (2005), "Functional Data Analysis for Sparse Longitudinal Data," *Journal of the American Statistical Association*, 100, 577–590. [654,655]

Yu, Y., Wang, T., and Samworth, R. J. (2015), "A Useful Variant of the Davis-Kahan Theorem for Statisticians," *Biometrika*, 102, 315–323. [654]

Zemyan, S. (2012), *The Classical Theory of Integral Equations*, Basel: Birkhäuser. [652]