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

Online Supplement

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

for the Alzheimer's Disease Neuroimaging Initiative.†

August 17, 2017

*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 Proofs of Propositions

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

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

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

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, $\left\|f^{(j)}\right\|_2^2 \leq B$ for all $j = 1, \ldots, p$. 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 $\boldsymbol{t} \in \mathcal{T}$. Define $K := \max_{i,j=1,\ldots,p} K_{ij}$ with K_{ij} as in (5). Then

$$\begin{aligned}\|g(\boldsymbol{t})\|^2 &\leq \sum\nolimits_{j=1}^{p} \left(\sum\nolimits_{i=1}^{p} \int_{\mathcal{T}_i} \left|C_{ij}(s_i, t_j) f^{(i)}(s_i)\right| \mathrm{d}s_i\right)^2 \\ &\overset{\text{Hölder}}{\leq} \sum\nolimits_{j=1}^{p} \left(\sum\nolimits_{i=1}^{p} \left[\int_{\mathcal{T}_i} C_{ij}(s_i, t_j)^2 \mathrm{d}s_i\right]^{1/2} \left[\int_{\mathcal{T}_i} f^{(i)}(s_i)^2 \mathrm{d}s_i\right]^{1/2}\right)^2 \\ &\leq \sum\nolimits_{j=1}^{p} \left(\sum\nolimits_{i=1}^{p} K^{1/2} B^{1/2}\right)^2 = p^3 K B < \infty.\end{aligned}$$

- $\mathcal{Z}$ 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

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

for all $i, j = 1, \dots, p$. Set $\delta := \min_{i,j=1,\dots,p} \delta_{ij}$ and let $\|\boldsymbol{t} - \boldsymbol{t}^*\|_{\mathcal{T}} < \delta$. Clearly, $\left\|t_j - t_j^*\right\| < \delta$ for all $j = 1, \dots, p$ and for $g \in \mathcal{Z}$ it holds

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

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 $\left\|t_j - t_j^*\right\| < \delta_j$. Then, as $\nu_m > 0$ and $\left\|\psi_m^{(i)}\right\|_2 \leq |||\psi_m||| = 1$,

$$\begin{aligned}
\left|\psi_m^{(j)}(t_j) - \psi_m^{(j)}(t_j^*)\right| &= \frac{1}{\nu_m} \left|\sum\nolimits_{i=1}^{p} \int_{\mathcal{T}_i} \left[C_{ij}(s_i, t_j) - C_{ij}(s_i, t_j^*)\right] \psi_m^{(i)}(s_i) \mathrm{d}s_i\right| \\
&\leq \frac{1}{\nu_m} \tilde{\varepsilon} \sum\nolimits_{i=1}^{p} \int_{\mathcal{T}_i} \left|\psi_m^{(i)}(s_i)\right| \mathrm{d}s_i \overset{\text{Hölder}}{\leq} \frac{\tilde{\varepsilon}}{\nu_m} \sum\nolimits_{i=1}^{p} \left\|\psi_m^{(i)}\right\|_2 \lambda(\mathcal{T}_i)^{1/2} \\
&\leq \frac{\tilde{\varepsilon}}{\nu_m} \sum\nolimits_{i=1}^{p} \lambda(\mathcal{T}_i)^{1/2} = \frac{\varepsilon}{2} < \varepsilon.
\end{aligned}$$

□

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

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

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

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

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

$$\begin{aligned}
0 &\leq \langle\langle \Gamma f, f \rangle\rangle - \langle\langle \Gamma_M f, f \rangle\rangle \\
&= \lambda\left(B_{t^*}\left(\tfrac{1}{n}\right)\right)^{-2} \int_{B_{t^*}(\frac{1}{n})} \int_{B_{t^*}(\frac{1}{n})} C_{jj}(s_j, t_j) - \sum\nolimits_{m=1}^{M} \nu_m \psi_m^{(j)}(s_j) \psi_m^{(j)}(t_j) \mathrm{d}s_j\ \mathrm{d}t_j \\
&\to C_{jj}(t^*, t^*) - \sum\nolimits_{m=1}^{M} \nu_m \psi_m^{(j)}(t^*) \psi_m^{(j)}(t^*) \quad \text{for } n \to \infty
\end{aligned}$$

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

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

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 $g := \left(0, \ldots, 0, g^{(j)}, 0, \ldots, 0\right)$, which is clearly in $\mathcal{H}$. Therefore

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

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,

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

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 $s,t \in \mathcal{T}_j$. □

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 $X(\boldsymbol{t}) = \sum_{m=1}^{\infty} \rho_m \psi_m(\boldsymbol{t}),\ \boldsymbol{t} \in \mathcal{T}$ with random variables $\rho_m = \langle\langle X, \psi_m \rangle\rangle$. Hence for $m,n \in \mathbb{N}$

1. $\mathbb{E}\left(\rho_m\right) = \sum\nolimits_{j=1}^{p} \int_{\mathcal{T}_j} \mathbb{E}\left(X^{(j)}(t_j)\right) \psi_m^{(j)}(t_j) \mathrm{d}t_j = 0$, since $\mathbb{E}\left(X^{(j)}(t_j)\right) = 0$ for all $t_j \in \mathcal{T}_j$, $j = 1, \ldots, p$ by assumption.

2. $$\begin{aligned} \mathrm{Cov}(\rho_m, \rho_n) &= \mathbb{E}\left(\sum\nolimits_{i=1}^{p} \int_{\mathcal{T}_i} X^{(i)}(s_i)\psi_m^{(i)}(s_i)\mathrm{d}s_i \cdot \sum\nolimits_{j=1}^{p} \int_{\mathcal{T}_j} X^{(j)}(t_j)\psi_n^{(j)}(t_j)\mathrm{d}t_j\right) \\ &= \sum\nolimits_{j=1}^{p} \int_{\mathcal{T}_j} \sum\nolimits_{i=1}^{p} \int_{\mathcal{T}_i} C_{ij}(s_i,t_j)\psi_m^{(i)}(s_i)\mathrm{d}s_i\ \psi_n^{(j)}(t_j)\mathrm{d}t_j \\ &= \sum\nolimits_{j=1}^{p} \int_{\mathcal{T}_j} \nu_m \psi_m^{(j)}(t_j)\psi_n^{(j)}(t_j)\mathrm{d}t_j = \nu_m \delta_{mn}. \end{aligned}$$

3. Let $X_{\lceil M \rceil}(\boldsymbol{t}) := \sum\nolimits_{m=1}^{M} \rho_m \psi_m(\boldsymbol{t}) = \sum\nolimits_{m=1}^{M} \left[\sum\nolimits_{i=1}^{p} \int_{\mathcal{T}_i} X^{(i)}(s_i)\psi_m^{(i)}(s_i)\mathrm{d}s_i\right] \psi_m(\boldsymbol{t})$ for

$\boldsymbol{t} \in \mathcal{T}$ be the truncated Karhunen-Loève representation of X. Then

$$\begin{aligned}
\mathbb{E}\left(\left\|X(\boldsymbol{t}) - X_{\lceil M \rceil}(\boldsymbol{t})\right\|^2\right) &= \sum\nolimits_{j=1}^{p}\left[\mathbb{E}\left(X^{(j)}(t_j)^2\right) - 2\mathbb{E}\left(X^{(j)}(t_j)X^{(j)}_{\lceil M \rceil}(t_j)\right) + \mathbb{E}\left(X^{(j)}_{\lceil M \rceil}(t_j)^2\right)\right] \\
&= \sum\nolimits_{j=1}^{p}\left[C_{jj}(t_j,t_j) - 2\sum\nolimits_{m=1}^{M}\sum\nolimits_{i=1}^{p}\int_{\mathcal{T}_i} C_{ij}(s_i,t_j)\psi_m^{(i)}(s_i)\mathrm{d}s_i\ \psi_m^{(j)}(t_j)\right. \\
&\quad \left. + \sum\nolimits_{m=1}^{M}\sum\nolimits_{n=1}^{M}\sum\nolimits_{i=1}^{p}\int_{\mathcal{T}_i}\sum\nolimits_{k=1}^{p}\int_{\mathcal{T}_k} C_{ki}(u_k,s_i)\psi_n^{(k)}(u_k)\mathrm{d}u_k\ \psi_m^{(i)}(s_i)\mathrm{d}s_i\ \psi_m^{(j)}(t_j)\psi_n^{(j)}(t_j)\right] \\
&= \sum\nolimits_{j=1}^{p}\left[C_{jj}(t_j,t_j) - 2\sum\nolimits_{m=1}^{M}\nu_m\psi_m^{(j)}(t_j)\psi_m^{(j)}(t_j)\right. \\
&\quad \left. + \sum\nolimits_{m=1}^{M}\sum\nolimits_{n=1}^{M}\sum\nolimits_{i=1}^{p}\int_{\mathcal{T}_i}\nu_n\psi_n^{(i)}(s_i)\psi_m^{(i)}(s_i)\mathrm{d}s_i\psi_m^{(j)}(t_j)\psi_n^{(j)}(t_j)\right] \\
&= \sum\nolimits_{j=1}^{p}\left[C_{jj}(t_j,t_j) - \sum\nolimits_{m=1}^{M}\nu_m\psi_m^{(j)}(t_j)^2\right] \to 0 \quad \text{for } M \to \infty
\end{aligned}$$

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

$$\begin{aligned}
\left(\Gamma^{(j)}\tilde{\phi}^{(j)}\right)(t) &= \int_{\mathcal{T}_j}\mathrm{Cov}(X^{(j)}(s), X^{(j)}(t))\tilde{\phi}^{(j)}(s)\mathrm{d}s \\
&= \int_{\mathcal{T}_j}\sum\nolimits_{m=1}^{M}\nu_m\psi_m^{(j)}(s)\psi_m^{(j)}(t)\tilde{\phi}^{(j)}(s)\mathrm{d}s \stackrel{!}{=} \lambda^{(j)}\tilde{\phi}^{(j)}(t), \qquad (14)
\end{aligned}$$

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

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

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

$$\boldsymbol{A^{(j)}}\boldsymbol{u^{(j)}_m} = \lambda^{(j)}_m \boldsymbol{u^{(j)}_m}.$$

Let $\lambda^{(j)}_1 \geq \ldots \geq \lambda^{(j)}_{M_j} > 0,\ M_j \leq M$ be the non-zero eigenvalues of $\boldsymbol{A^{(j)}}$. They are at the same time the only non-zero eigenvalues of $\Gamma^{(j)}$. The associated eigenfunctions $\tilde{\phi}^{(j)}_m$ of $\Gamma^{(j)}$ are parametrized by the eigenvectors $\boldsymbol{u^{(j)}_m}$ via (Zemyan, 2012, Chapter 1.3.)

$$\Gamma^{(j)}\tilde{\phi}^{(j)}_m = \sum\nolimits_{n=1}^{M} \nu_n^{1/2}\psi^{(j)}_n[\boldsymbol{u^{(j)}_m}]_n = \lambda^{(j)}_m\tilde{\phi}^{(j)}_m \quad\Leftrightarrow\quad \tilde{\phi}^{(j)}_m = \left(\lambda^{(j)}_m\right)^{-1}\sum\nolimits_{n=1}^{M} \nu_n^{1/2}\psi^{(j)}_n[\boldsymbol{u^{(j)}_m}]_n.$$

Since $\langle\tilde{\phi}^{(j)}_m, \tilde{\phi}^{(j)}_n\rangle_2 = \left(\lambda^{(j)}_m\lambda^{(j)}_n\right)^{-1} \boldsymbol{u^{(j)}_m}^\top \boldsymbol{A^{(j)}}\boldsymbol{u^{(j)}_n} = \left(\lambda^{(j)}_m\right)^{-1}\delta_{mn}$, orthonormal eigenfunctions $\phi^{(j)}_m$ are given by

$$\phi^{(j)}_m = {\lambda^{(j)}_m}^{1/2}\tilde{\phi}^{(j)}_m = \left(\lambda^{(j)}_m\right)^{-1/2}\sum\nolimits_{n=1}^{M} \nu_n^{1/2}\psi^{(j)}_n[\boldsymbol{u^{(j)}_m}]_n.$$

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

$$\xi^{(j)}_m = \langle X^{(j)}, \phi^{(j)}_m\rangle_2 = \left(\lambda^{(j)}_m\right)^{-1/2}\sum\nolimits_{n=1}^{M} \nu_n^{1/2}\left[\boldsymbol{u^{(j)}_m}\right]_n \sum\nolimits_{k=1}^{M}\rho_k\langle\psi^{(j)}_n, \psi^{(j)}_k\rangle_2$$

and $\mathbb{E}(\xi^{(j)}_m) = 0,\ \mathrm{Cov}(\xi^{(j)}_m, \xi^{(j)}_n) = \lambda^{(j)}_m\delta_{mn}$.

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

$$\begin{aligned}(\Gamma\psi)^{(j)}(t_j) &= \sum\nolimits_{k=1}^{p}\int_{\mathcal{T}_k}\mathrm{Cov}\left(X^{(k)}(s_k), X^{(j)}(t_j)\right)\psi^{(k)}(s_k)\mathrm{d}s_k \\ &= \sum\nolimits_{k=1}^{p}\sum\nolimits_{l=1}^{M_j}\sum\nolimits_{n=1}^{M_k}\underbrace{\mathrm{Cov}\left(\xi^{(j)}_l, \xi^{(k)}_n\right)}_{=:Z^{(jk)}_{ln}}\phi^{(j)}_l(t_j)\underbrace{\int_{\mathcal{T}_k}\phi^{(k)}_n(s_k)\psi^{(k)}(s_k)\mathrm{d}s_k}_{=:c^{(k)}_n} \overset{!}{=} \nu\psi^{(j)}(t_j).\end{aligned}$$

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

$$\sum_{k=1}^{p}\sum_{l=1}^{M_j}\sum_{n=1}^{M_k} Z_{ln}^{(jk)}\phi_l^{(j)}(t_j)c_n^{(k)} = \nu\psi^{(j)}(t_j) \tag{15}$$

$$\Rightarrow \int_{\mathcal{T}_j}\phi_m^{(j)}(t_j)\cdot\sum_{k=1}^{p}\sum_{l=1}^{M_j}\sum_{n=1}^{M_k} Z_{ln}^{(jk)}\phi_l^{(j)}(t_j)c_n^{(k)}\mathrm{d}t_j = \int_{\mathcal{T}_j}\phi_m^{(j)}(t_j)\cdot\nu\psi^{(j)}(t_j)\mathrm{d}t_j$$

$$\Leftrightarrow \sum_{k=1}^{p}\sum_{n=1}^{M_k} Z_{mn}^{(jk)}c_n^{(k)} = \nu c_m^{(j)}$$

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

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

with matrices $\boldsymbol{Z^{(jk)}} \in \mathbb{R}^{M_j\times M_k}$ and $\boldsymbol{c^{(j)}} \in \mathbb{R}^{M_j}$. The last equation is again an eigenequation for the symmetric (and since $\nu \geq 0$) positive semidefinite block matrix $\boldsymbol{Z} \in \mathbb{R}^{M_+\times M_+}$. Let $\nu_1 \geq \ldots \geq \nu_M > 0,\ M \leq M_+$ be the non-zero eigenvalues of $\boldsymbol{Z}$. 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 $\boldsymbol{c_1},\dots,\boldsymbol{c_M}$ associated with $\nu_1,\dots,\nu_M$:

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

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

$$\langle\langle\psi_m,\psi_n\rangle\rangle = \sum_{j=1}^{p}\langle\psi_m^{(j)},\psi_n^{(j)}\rangle_2 = \sum_{j=1}^{p}\sum_{l=1}^{M_j}[\boldsymbol{c_m}]_l^{(j)}[\boldsymbol{c_n}]_l^{(j)} = \boldsymbol{c_n}^\top\boldsymbol{c_m} = \delta_{mn}.$$

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

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

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

$\square$

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

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

In the following, use $C_{ij}^{[M]}(s_i, t_j) := \mathrm{Cov}(X^{[M](i)}(s_i), X^{[M](j)}(t_j))$ 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)

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

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

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

converges uniformly to 0 for $t_j \in \mathcal{T}_j$ and $M_j \to \infty$. 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

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

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

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

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

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

Hence it holds that

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

The main proof is now in three steps:

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

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

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

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

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

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

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

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

where $\lambda(\mathcal{T}_i)$ is the Lebesuge 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 $M_1,\ldots,M_p\to\infty$, the limit $M_j\to\infty$ 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 $M_1,\ldots,M_p\to\infty$.

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

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

for $T=\max_{j=1,\ldots,p}\lambda(\mathcal{T}_j)$ and $C=\max_{j=1,\ldots,p}\left\|C_{jj}\right\|_\infty$. 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 non-zero 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

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

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

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

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

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

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

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

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

for $M_1, \ldots, M_p \to \infty$, 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 $\left|\left|\left|X - X^{[M]}\right|\right|\right|$ converges in the second mean to 0, thus $\left|\left|\left|X - X^{[M]}\right|\right|\right| = o_p(1)$. Finally, this leads to

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

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

$\square$

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

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

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

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

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

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

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

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

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

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

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

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 $\boldsymbol{Z}$ and $\hat{\boldsymbol{Z}}$ are both symmetric matrices in $\mathbb{R}^{M_+ \times M_+}$ it holds (cf. Horn and Johnson, 1991, Chapter 3.7)

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

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

$$\begin{aligned}
\left| Z_{ln}^{(jk)} - \hat{Z}_{ln}^{(jk)} \right| &= \left| \text{Cov}\left(\xi_l^{(j)}, \xi_n^{(k)}\right) - \frac{1}{N}\sum_{i=1}^N \hat{\xi}_{i,l}^{(j)}\hat{\xi}_{i,n}^{(k)} - \frac{1}{N(N-1)}\sum_{i=1}^N \hat{\xi}_{i,l}^{(j)}\hat{\xi}_{i,n}^{(k)} \right| \\
&\overset{(A5)}{\leq} \left| \text{Cov}\left(\langle X^{(j)}, \phi_l^{(j)}\rangle_2, \langle X^{(k)}, \phi_n^{(k)}\rangle_2\right) - \frac{1}{N}\sum_{i=1}^N \langle X_i^{(j)}, \hat{\phi}_l^{(j)}\rangle_2 \cdot \langle X_i^{(k)}, \hat{\phi}_n^{(k)}\rangle_2 \right| \\
&\quad + \frac{1}{N(N-1)}\sum_{i=1}^N \left| \langle X_i^{(j)}, \hat{\phi}_l^{(j)}\rangle_2 \right| \left| \langle X_i^{(k)}, \hat{\phi}_n^{(k)}\rangle_2 \right| \\
&\leq \Bigg| \int_{\mathcal{T}_j}\int_{\mathcal{T}_k} \mathbb{E}\left(X^{(j)}(s)X^{(k)}(t)\right)\phi_l^{(j)}(s)\phi_n^{(k)}(t)\mathrm{d}s\ \mathrm{d}t - \\
&\qquad\qquad \int_{\mathcal{T}_j}\int_{\mathcal{T}_k} \frac{1}{N}\sum_{i=1}^N \left(X_i^{(j)}(s)X_i^{(k)}(t)\right)\hat{\phi}_l^{(j)}(s)\hat{\phi}_n^{(k)}(t)\mathrm{d}s\ \mathrm{d}t \Bigg| \\
&\quad + \frac{1}{N(N-1)}\sum_{i=1}^N \left\|X_i^{(j)}\right\|_2 \left\|\hat{\phi}_l^{(j)}\right\|_2 \left\|X_i^{(k)}\right\|_2 \left\|\hat{\phi}_n^{(k)}\right\|_2 \\
&\overset{(21)}{=} \left| \int_{\mathcal{T}_j}\int_{\mathcal{T}_k} \text{Cov}\left(X^{(j)}(s), X^{(k)}(t)\right)\phi_l^{(j)}(s)\phi_n^{(k)}(t) - \widehat{\text{Cov}}\left(X^{(j)}(s), X^{(k)}(t)\right)\hat{\phi}_l^{(j)}(s)\hat{\phi}_n^{(k)}(t)\mathrm{d}s\ \mathrm{d}t \right| \\
&\quad + \frac{1}{N(N-1)}\sum_{i=1}^N O_p(1)\cdot 1 \cdot O_p(1) \cdot 1 \\
&= \left| \int_{\mathcal{T}_j}\int_{\mathcal{T}_k} C_{jk}(s,t)\left[\phi_l^{(j)}(s)\phi_n^{(k)}(t) - \hat{\phi}_l^{(j)}(s)\hat{\phi}_n^{(k)}(t)\right]\mathrm{d}s\ \mathrm{d}t \right| \\
&\quad + \left| \int_{\mathcal{T}_j}\int_{\mathcal{T}_k} \left[C_{jk}(s,t) - \hat{C}_{jk}(s,t)\right]\hat{\phi}_l^{(j)}(s)\hat{\phi}_n^{(k)}(t)\mathrm{d}s\ \mathrm{d}t \right| + O_p(N^{-1})
\end{aligned}$$

$$
\begin{aligned}
&\leq \int_{\mathcal{T}_j}\int_{\mathcal{T}_k} C_{jj}(s,s)^{1/2}C_{kk}(t,t)^{1/2}\left|\phi_l^{(j)}(s)\phi_n^{(k)}(t)-\hat{\phi}_l^{(j)}(s)\hat{\phi}_n^{(k)}(t)\right|\mathrm{d}s\ \mathrm{d}t\\
&\quad+\int_{\mathcal{T}_k}\left|\left((\Gamma^{(jk)}-\hat{\Gamma}^{(jk)})\hat{\phi}_l^{(j)}\right)(t)\hat{\phi}_n^{(k)}(t)\right|\mathrm{d}t+O_p(N^{-1})\\
&\leq \|C_{jj}\|_\infty^{1/2}\|C_{kk}\|_\infty^{1/2}\left(\int_{\mathcal{T}_j}\int_{\mathcal{T}_k}\left|\phi_l^{(j)}(s)\right|\left|\phi_n^{(k)}(t)-\hat{\phi}_n^{(k)}(t)\right|\mathrm{d}s\ \mathrm{d}t+\right.\\
&\qquad\qquad\qquad\qquad\qquad\left.\int_{\mathcal{T}_j}\int_{\mathcal{T}_k}\left|\phi_l^{(j)}(s)-\hat{\phi}_l^{(j)}(s)\right|\left|\hat{\phi}_n^{(k)}(t)\right|\mathrm{d}s\ \mathrm{d}t\right)\\
&\quad+\left\|(\Gamma^{(jk)}-\hat{\Gamma}^{(jk)})\hat{\phi}_l^{(j)}\right\|_2\left\|\hat{\phi}_n^{(k)}\right\|_2+O_p(N^{-1})\\
&\leq \left(\|C_{jj}\|_\infty\|C_{kk}\|_\infty\lambda(\mathcal{T}_j)\lambda(\mathcal{T}_k)\right)^{1/2}\left(\left\|\phi_l^{(j)}\right\|_2\left\|\phi_n^{(k)}-\hat{\phi}_n^{(k)}\right\|_2+\left\|\phi_l^{(j)}-\hat{\phi}_l^{(j)}\right\|_2\left\|\hat{\phi}_n^{(k)}\right\|_2\right)\\
&\quad+\left\|\Gamma^{(jk)}-\hat{\Gamma}^{(jk)}\right\|_{\mathrm{op}}\left\|\hat{\phi}_l^{(j)}\right\|_2\left\|\hat{\phi}_n^{(k)}\right\|_2+O_p(N^{-1})\\
&=\left(\|C_{jj}\|_\infty\|C_{kk}\|_\infty\lambda(\mathcal{T}_j)\lambda(\mathcal{T}_k)\right)^{1/2}\left(\left\|\phi_n^{(k)}-\hat{\phi}_n^{(k)}\right\|_2+\left\|\phi_l^{(j)}-\hat{\phi}_l^{(j)}\right\|_2\right)+\left\|\Gamma^{(jk)}-\hat{\Gamma}^{(jk)}\right\|_{\mathrm{op}}+O_p(N^{-1})\\
&=O_p(\Delta_{M_j}^{(j)}r_N^\Gamma)+O_p(\Delta_{M_k}^{(k)}r_N^\Gamma)+O_p(N^{-1/2})+O_p(N^{-1})=O_p(\max(\Delta_{M_j}^{(j)}r_N^\Gamma,\Delta_{M_k}^{(k)}r_N^\Gamma,N^{-1/2})).
\end{aligned}
$$

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

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

□

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

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

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

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

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

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 $\lambda_{\max}(\boldsymbol{A})$ denotes the maximal eigenvalue of a symmetric matrix $\boldsymbol{A}$. 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:

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

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

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

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

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

$\left|\left|\left|X_i^{[M]}\right|\right|\right|$ is bounded in probability using (16) and analogous arguments as for $|||X|||$

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 $m = 1, \ldots, M_j$, let $\varepsilon > 0$ and $c := (\frac{2\lambda_1^{(j)}}{\varepsilon})^{1/2} < \infty$. Then

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

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

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

In total,

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

□

2 Simulation – Additional Results

2.1 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 $f_1, \ldots, f_M$ on the interval $[0, 2]$. Next, choose split points $0 = T_1 < T_2 < \ldots < T_p < T_{p+1} = 2$ and shift values $\eta_1, \ldots, \eta_p \in \mathbb{R}$ 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 $\sigma_1, \ldots, \sigma_p \in \{-1, 1\}$, the multivariate eigenfunctions are given by their elements

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

The constuction process is illustrated in Fig. 1. Clearly, $\{\psi_m,\ m = 1, \ldots, M\}$ is an orthonormal system in $\mathcal{H} = L^2(\mathcal{T}_1) \times \ldots \times L^2(\mathcal{T}_p)$. 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. 2

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

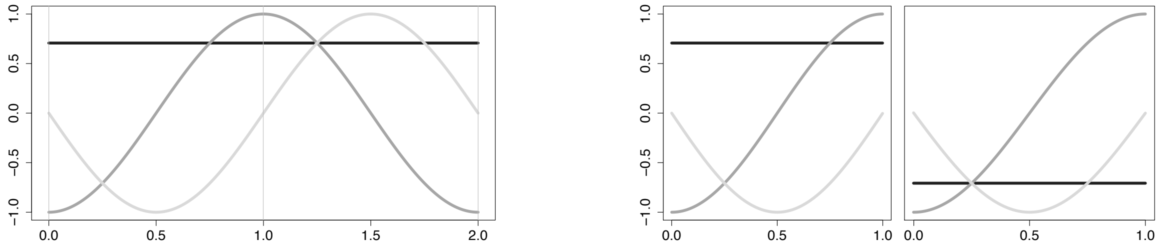

Figure 1: Illustration of the construction of the multivariate eigenfunctions ψ_m for the first setting. Left: The first $M = 3$ functions of the Fourier basis on $[0, 2]$ with one split point. Right: The shifted pieces multiplied with random signs form the first three bivariate eigenfunctions.

$[-1, 1]$. The eigenfunctions are defined via

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

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. 3.

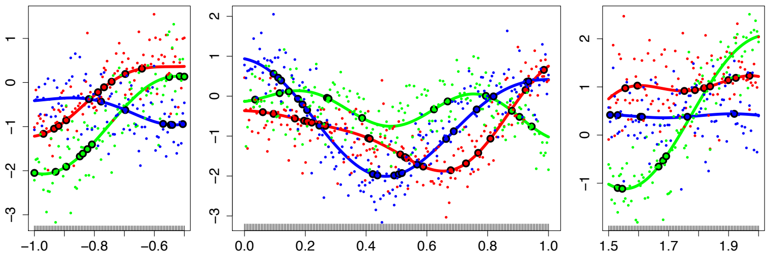

Figure 2: Three examples for simulated data in simulation setting 2 based on the leading $M = 8$ Fourier basis functions and exponential eigenvalue decay. Left: $x_i^{(1)}$, Middle: $x_i^{(2)}$, Right: $x_i^{(3)}$. Solid lines show the realizations x_i, small points are the corresponding data with measurement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

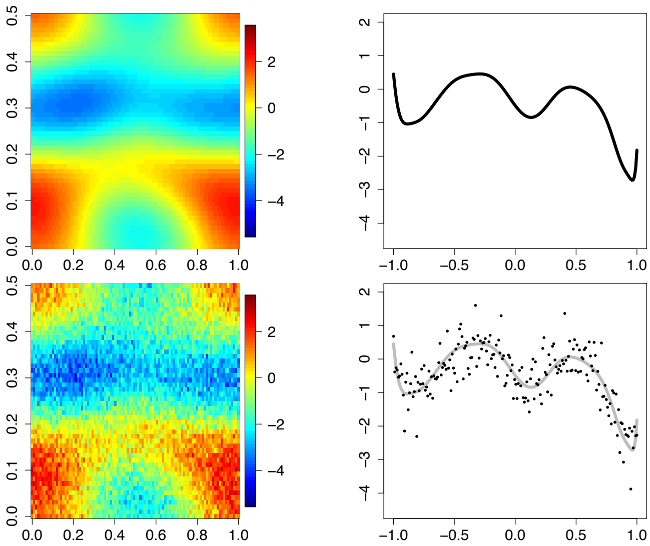

Figure 3: Examples for simulated data in the third simulation setting (cf. Section 4.2) consisting of images ($x_i^{(1)}$, left) and functions ($x_i^{(2)}$, right) without (1st row) and with measurement error (2nd row).

2.2 Example Fits

Table 1: 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$_{\mathrm{RS}}$, $\sigma^2 = 0$; simulation median: $< 10^{-3}$) and 0.710 (MFPCA$_{\mathrm{RS}}$, $\sigma^2 = 0.25$; simulation median: 0.480). The results for the corresponding eigenfunctions are given in Fig. 4.

$m =$	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$_{\mathrm{RS}}$ ($\sigma^2 = 0$)	1.140	0.500	0.315	0.248	0.127	0.090	0.048	0.034
MFPCA$_{\mathrm{RS}}$ ($\sigma^2 = 0.25$)	1.139	0.504	0.317	0.252	0.130	0.091	0.048	0.035

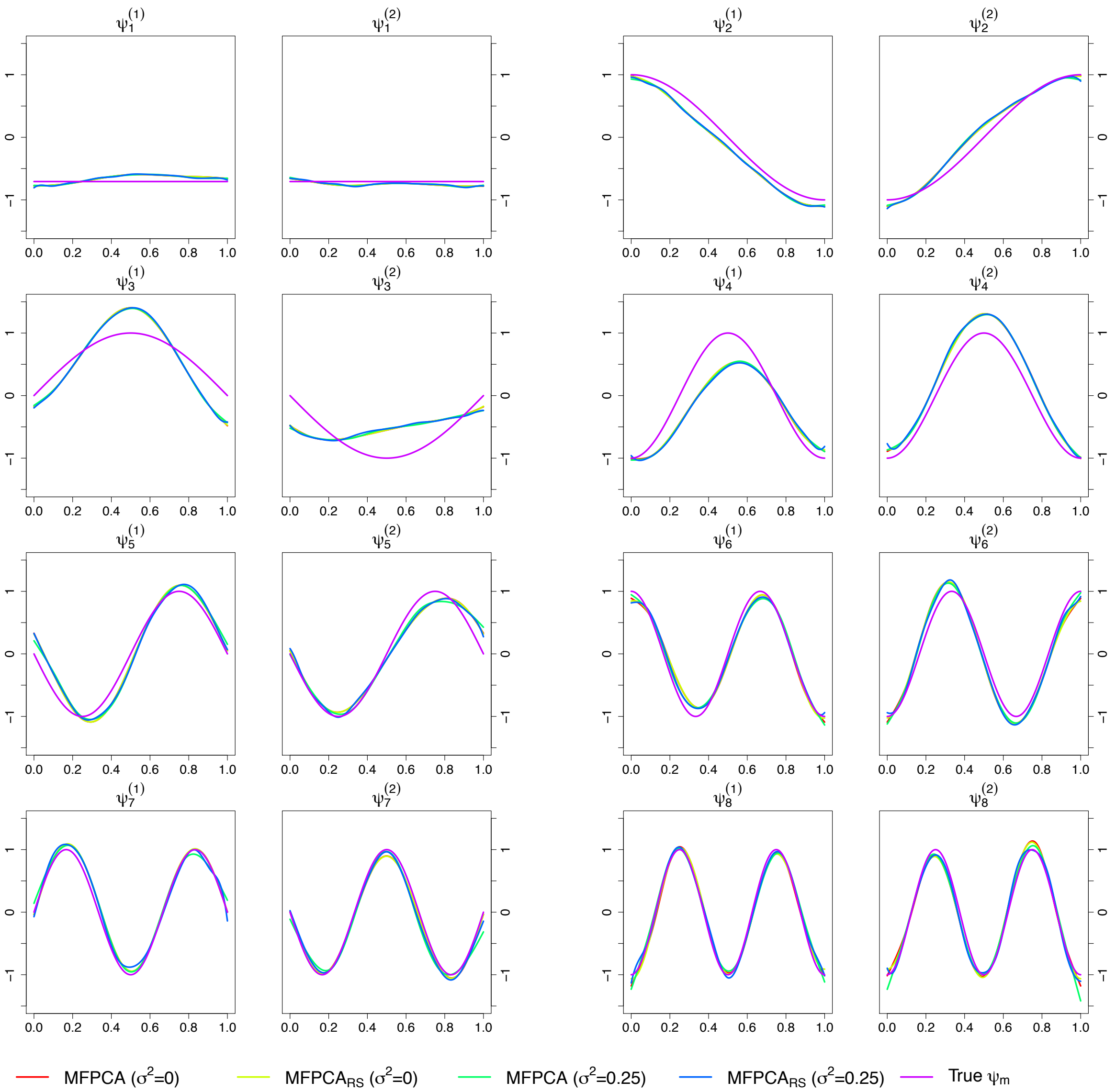

Figure 4: True and estimated eigenfunctions for the first setting based on one example replication with $N = 250$ observations. The results for the corresponding eigenfunctions are given in Table 1.

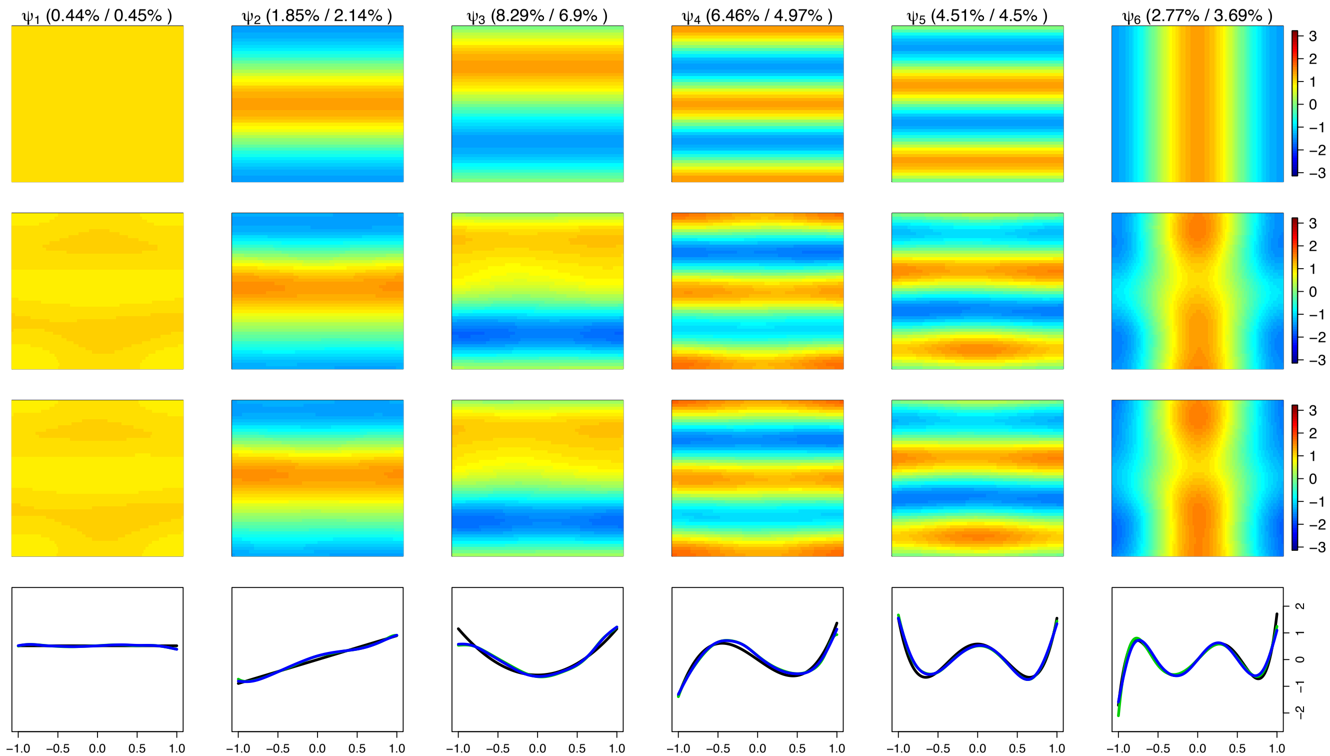

Figure 5: Exemplary result for one replication in simulation setting 3 and MFPCA based on univariate spline expansions (results for the eigenfunctions $\psi_1, \dots, \psi_6$). The first row shows the true first elements $\psi_m^{(1)}$, the second/third row gives the results of $\hat{\psi}_m^{(1)}$ for data without/with measurement error. In the fourth row, the second elements $\psi_m^{(2)}$ of the true eigenfunctions are shown in black and the corresponding estimates for data without/with measurement error are shown in green/blue. Percentages in the titles give the relative errors $\mathrm{Err}(\hat{\psi}_m)$ for the estimates based on data without/with measurement error. The reconstruction error MRSE is 0.40%/2.25% for data without/with measurement error (simulation median: 0.40%/2.05%). Results for the eigenfunctions $\psi_7, \dots, \psi_{12}$ are shown in Fig. 6.

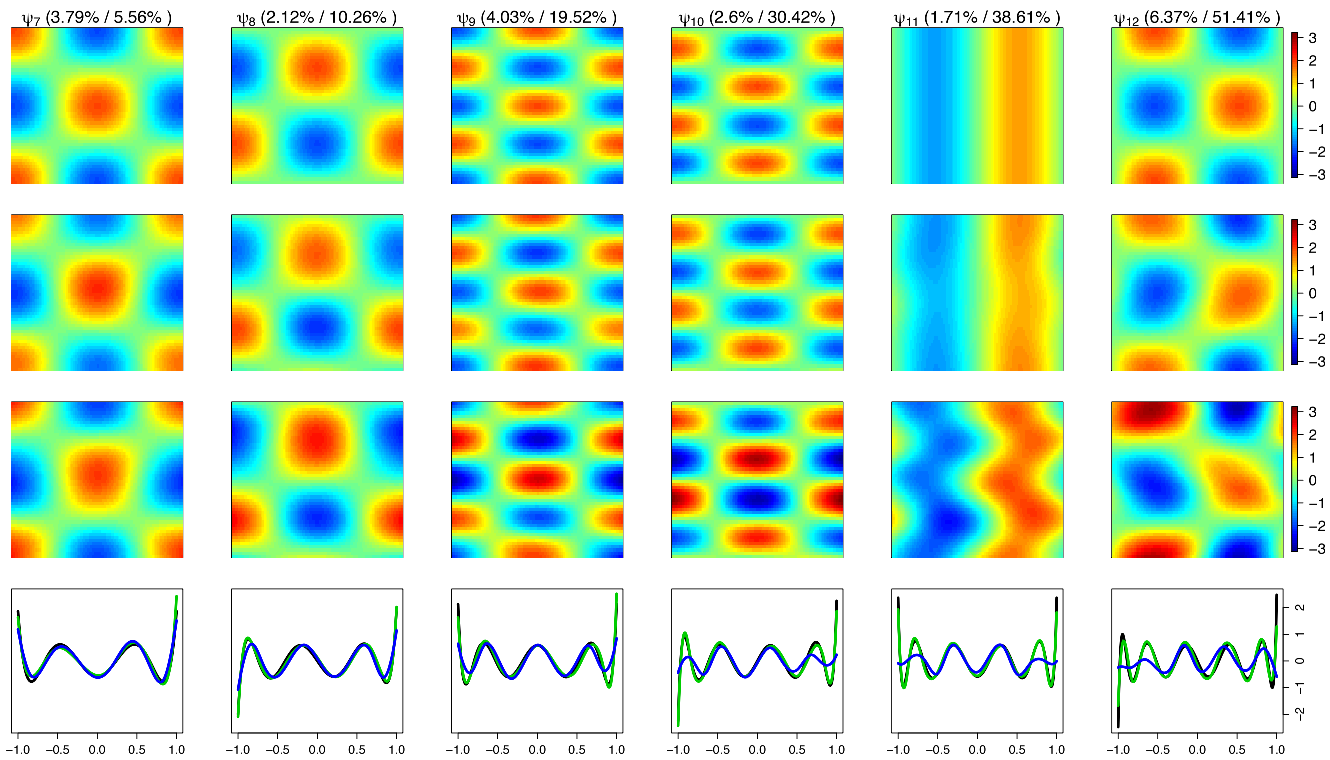

Figure 6: Exemplary result for one replication in simulation setting 3 (results for the eigenfunctions $\psi_7, \ldots, \psi_{12}$). Please refer to Fig. 5 for details.

2.3 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 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. 7 and Table 2. The number of estimated multivariate eigenvalues/eigenfunctions is for all 100 datasets $\hat{M} = 4$ for pve $= 0.75$, $\hat{M} = 6$ for 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. pve $= 0.95$ for $m = 5$, cf. Fig. 7), the additional benefit of using more univariate eigenfunctions (pve > 0.95) becomes negligible. If, in contrast, the univariate

FPCA does not contain enough information (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.

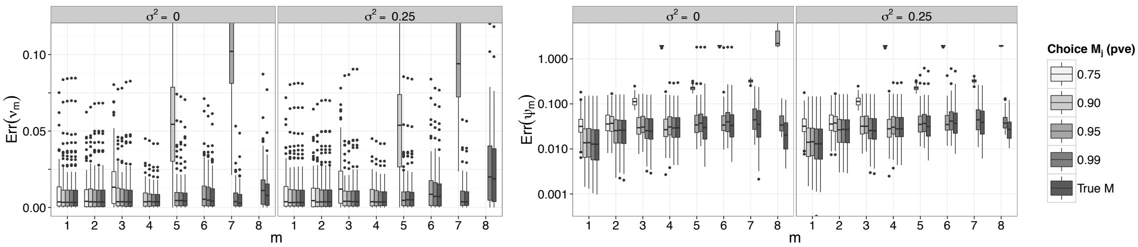

Figure 7: Relative errors for estimated eigenvalues (left) and eigenfunctions (right, log-scale) for the sensitivity analysis. Extreme values cut off for better comparability.

Table 2: 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

2.4 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, 2017). 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. 8 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. 6). 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. 9). 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. 9). 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. 9, 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.

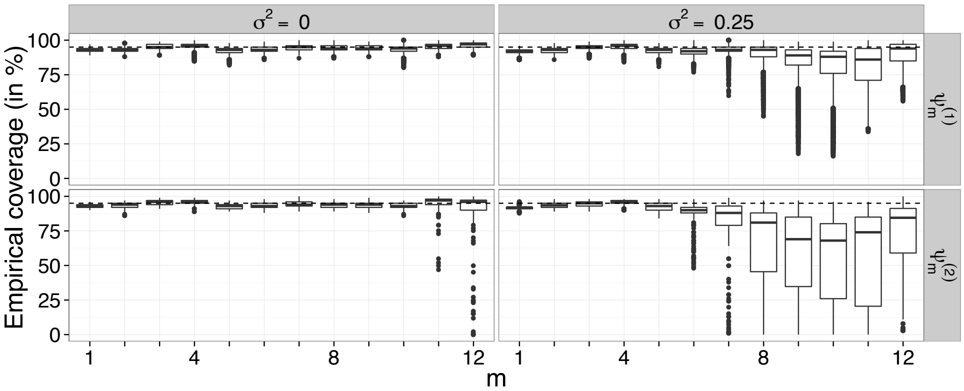

Figure 8: Empirical coverages from the bootstrap simulation study for data without ($\sigma^2 = 0$) and with ($\sigma^2 = 0.25$) measurement error. The boxplots show the pointwise coverage of the bootstrap confidence bands aggregated over the corresponding domains for both elements of the true eigenfunctions $\psi_m,\ m = 1, \ldots, 12$ (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%.

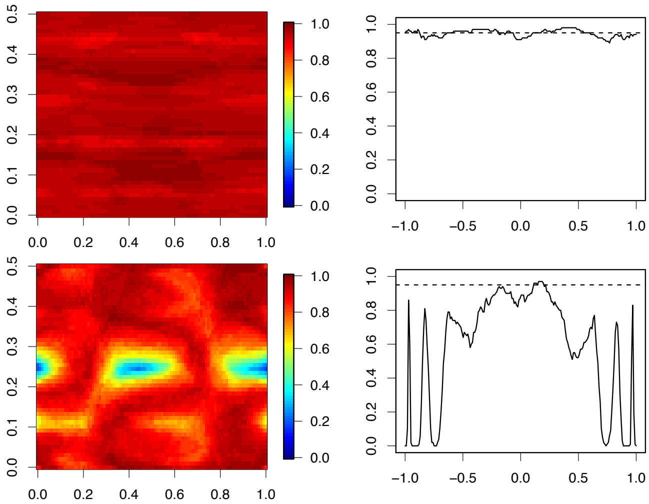

Figure 9: Exemplary results from the bootstrap simulation study for data observed with measurement error. The first row shows the estimated coverages for the third eigenfunction, the second row shows the estimated coverages for the eigenfunction of order 9 (see also Fig. 8). The first column corresponds 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%.

3 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. 10. For the new approach, the multivariate principal components are calculated based on univariate FPCA with $M_1 = M_2 = 5$ principal components. For

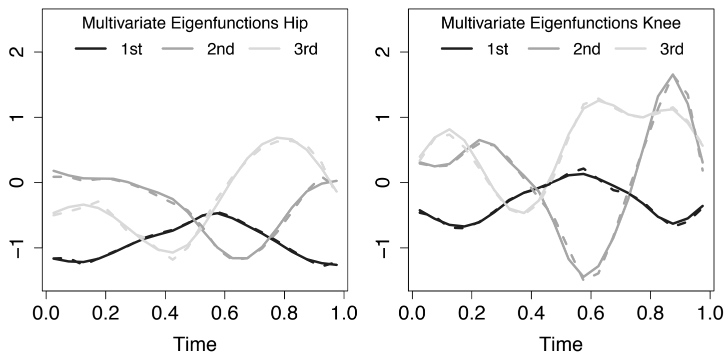

Figure 10: The first three estimated bivariate eigenfunctions for the gait data set. Solid lines show the results of the new MFPCA approach, dashed lines correspond to the approach of Ramsay and Silverman (2005). The functions have been reflected, if necessary, for comparison purposes.

MFPCA_{RS}, the observed functions are presmoothed 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.

References

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

Happ, C. (2017). *MFPCA: Multivariate Functional Principal Component Analysis.* R package version 1.1.

Horn, R. A. and Johnson, C. R. (1991). *Topics in matrix analysis.* Cambridge University Press, Cambridge.

Kato, T. (1976). *Perturbation Theory for Linear Operators.* Springer, Berlin, 2 edition.

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.

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. ed. edition.

Rudin, W. (1987). *Real and complex analysis*. McGraw-Hill, New York, 3 edition.

Werner, D. (2011). *Funktionalanalysis*. Springer, Berlin, 7 edition.

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.