Supplementary Material to
Fast Multilevel Functional Principal Component Analysis

Erjia Cui*[1]
Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, Baltimore, MD 21205
ecui1@jhmi.edu

Ruonan Li*
Department of Statistics, North Carolina State University, 2311 Stinson Dr, Raleigh, NC 27607

Ciprian M. Crainiceanu
Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, Baltimore, MD 21205

Luo Xiao
Department of Statistics, North Carolina State University, 2311 Stinson Dr, Raleigh, NC 27607

[1]The * indicates co-first authors

S.1 Details in Smoothing Splines

The penalty matrix is symmetric with $\mathbf{P} = \mathbf{D}_q^T\mathbf{D}_q$, where $\mathbf{D}_q$ is a differencing matrix of difference orders q. Let $\boldsymbol{\theta} = \text{vec}(\boldsymbol{\Theta})$, where $\boldsymbol{\Theta} = (\mathbf{B}^T\mathbf{B}/L + \lambda\mathbf{P})^{-1}(\mathbf{B}^T\widehat{\mathbf{K}}\mathbf{B}/L^2)(\mathbf{B}^T\mathbf{B}/L + \lambda\mathbf{P})^{-1}$. Then, an estimate of $\boldsymbol{\theta}$ is given by minimizing

$$\|\widehat{\mathbf{K}} - \mathbf{B}\boldsymbol{\Theta}\mathbf{B}^T\|_F^2 + \boldsymbol{\theta}^T\mathbf{P}_2\boldsymbol{\theta},$$

where $\|\cdot\|_F$ is the Frobenius norm and $\mathbf{P}_2 = (2\mathbf{B}^T\mathbf{B} + \lambda\mathbf{D}_q^T\mathbf{D}_q) \otimes (\lambda\mathbf{D}_q^T\mathbf{D}_q)$. The above optimization gives the smooth estimate $\widetilde{\mathbf{K}} = \mathbf{B}\boldsymbol{\Theta}\mathbf{B}^T$. FACE selects the smoothing parameter by minimizing the pooled generalized cross validation (PGCV), which has a form of

$$\sum_{i=1}^{I} \|\mathbf{Y}_i - \mathbf{S}\mathbf{Y}_i\|_2^2/\{1 - \text{tr}(\mathbf{S})/J\}^2,$$

where $\mathbf{Y}_i$ is the observed data from subject i, $\|\cdot\|_2$ is the Euclidean norm and $\text{tr}(\mathbf{S})$ denotes the trace of $\mathbf{S}$. FACE provides an easy way to calculate PGCV, which only requires $O(IL + Ic)$ flops. Thus, the FACE algorithm is computationally fast when L is large.

S.2 Eigenfunctions by Multilevel FACE

Define $\widetilde{\mathbf{Y}}_{ij} = [\widetilde{Y}_{ij}(s_1), \ldots, \widetilde{Y}_{ij}(s_L)]^T$ and $\widetilde{\mathbf{Y}} = [\widetilde{\mathbf{Y}}_{11}, \ldots, \widetilde{\mathbf{Y}}_{1J_1}, \ldots, \widetilde{\mathbf{Y}}_{I1}, \ldots, \widetilde{\mathbf{Y}}_{IJ_I}] \in \mathbb{R}^{L\times n}$. The matrix format of $\widehat{K}_T(s,t)$ is then $\widehat{\mathbf{K}}_T = n^{-1}\widetilde{\mathbf{Y}}\mathbf{E}\mathbf{E}^T\widetilde{\mathbf{Y}}^T$, where $\mathbf{E}$ is a block diagonal matrix with the ith block being $\sqrt{nw_i}\mathbf{I}_{J_i}$. Similarly, the matrix format of $\widehat{K}_W(s,t)$ is $\widehat{\mathbf{K}}_W = n^{-1}\widetilde{\mathbf{Y}}\mathbf{R}\mathbf{R}^T\widetilde{\mathbf{Y}}^T$, where $\mathbf{R}$ is an $n \times n$ block diagonal matrix with the ith block being $\sqrt{nv_iJ_i}(\mathbf{I}_{J_i} - \mathbf{1}_{J_i}\mathbf{1}_{J_i}^T/J_i)$. See Section 3.2 in the paper for details of w_i and v_i with $\sum_i J_iw_i = 1$ and $\sum_{i=1}^I J_i(J_i - 1)v_i = 1$.

Before constructing the transformed functional data, we first compute $\mathbf{G} = \int \mathbf{B}(s)\mathbf{B}^T(s)ds$ and perform the decomposition $(\mathbf{B}^T\mathbf{B})^{-1/2}\mathbf{P}(\mathbf{B}^T\mathbf{B})^{-1/2} = \mathbf{O}\text{diag}(\mathbf{s})\mathbf{O}^T$. Let $\mathbf{A}_0 = (\mathbf{B}^T\mathbf{B})^{-1/2}\mathbf{O}$ and $\mathbf{A} = \mathbf{G}^{1/2}\mathbf{A}_0$. We then obtain within-subject eigenfunctions and between-subject eigenfunctions as follows.

1. Apply FACE for the total covariance and obtain $\widetilde{\mathbf{Y}}_{\mathbf{E}} = \mathbf{A}_0^T\mathbf{B}^T(\widetilde{\mathbf{Y}}\mathbf{E})$ and $\boldsymbol{\Sigma}_{\mathbf{S}}^1 = \{\mathbf{I}_c + \lambda_1\text{diag}(\mathbf{s})\}^{-1}$, where λ_1 is the smoothing parameter for total covariance. Compute the decomposition $n^{-1}\mathbf{A}\{\boldsymbol{\Sigma}_{\mathbf{S}}^1\widetilde{\mathbf{Y}}_{\mathbf{E}}\widetilde{\mathbf{Y}}_{\mathbf{E}}^T\boldsymbol{\Sigma}_{\mathbf{S}}^1\}\mathbf{A}^T = \mathbf{U}_{\mathbf{E}}\boldsymbol{\Lambda}_{\mathbf{E}}\mathbf{U}_{\mathbf{E}}^T$.

2. Apply FACE for the within covariance, we have $\widetilde{\mathbf{Y}}_{\mathbf{R}} = \mathbf{A}_0^T\mathbf{B}^T(\widetilde{\mathbf{Y}}\mathbf{R})$ and $\mathbf{\Sigma}_{\mathbf{S}}^2 = \{\mathbf{I}_c + \lambda_2\mathrm{diag}(\mathbf{s})\}^{-1}$, where λ_2 is the smoothing parameter for within covariance. Compute the decomposition $n^{-1}\mathbf{A}\{\mathbf{\Sigma}_{\mathbf{S}}^2\widetilde{\mathbf{Y}}_{\mathbf{R}}\widetilde{\mathbf{Y}}_{\mathbf{R}}^T\mathbf{\Sigma}_{\mathbf{S}}^2\}\mathbf{A}^T = \mathbf{U}_{\mathbf{R}}\mathbf{\Lambda}_{\mathbf{R}}\mathbf{U}_{\mathbf{R}}^T$ to obtain the within-subject eigenfunctions as $\mathbf{\Psi} = \mathbf{B}\mathbf{G}^{-1/2}\mathbf{U}_{\mathbf{R}}$.

3. Construct the decomposition $n^{-1}\mathbf{A}\{\mathbf{\Sigma}_{\mathbf{S}}^1\widetilde{\mathbf{Y}}_{\mathbf{E}}\widetilde{\mathbf{Y}}_{\mathbf{E}}^T\mathbf{\Sigma}_{\mathbf{S}}^1 - \mathbf{\Sigma}_{\mathbf{S}}^2\widetilde{\mathbf{Y}}_{\mathbf{R}}\widetilde{\mathbf{Y}}_{\mathbf{R}}^T\mathbf{\Sigma}_{\mathbf{S}}^2\}\mathbf{A}^T = \mathbf{U}_{\mathbf{F}}\mathbf{\Lambda}_{\mathbf{F}}\mathbf{U}_{\mathbf{F}}^T$ to obtain the between-subject eigenfunctions $\mathbf{\Phi} = \mathbf{B}\mathbf{G}^{-1/2}\mathbf{U}_{\mathbf{F}}$.

S.3 Details in Asymptotic Theory

S.3.1 Notation

We use T to denote the transpose of a matrix. For a matrix $\mathbf{A} = (a_{ij})$, $\|\mathbf{A}\|_{\max} = \max_{i,j}|a_{ij}|$, $\|\mathbf{A}\|_\infty = \max_i \sum_j |a_{ij}|$, and $(\mathbf{A})_+ = (|a_{ij}|)$. We also use $\mathrm{tr}(\mathbf{A})$ to denote the trace of $\mathbf{A}$, i.e., the sum of diagonal elements of $\mathbf{A}$. Let $\mathrm{vec}(\cdot)$ be the matrix operation that stacks the columns of the matrix into a single column vector. The little o and big O notation are with respect to the number of subjects I and we allow the number of observations per curve L to increase with I.

For two scalars, let $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$. The notation $a \preceq b$ denotes that there exists an absolute constant $\delta > 0$ such that $a \leq \delta b$ for sufficiently large I. And $a \simeq b$ means that $a \preceq b$ and $b \preceq a$, i.e., a and b are rate wise equivalent. Also $a \ll b$ means $a = o(b)$ and $a \gg b$ means $b = o(a)$. For two square matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{A} \leq \mathbf{B}$ denotes that $\mathbf{B} - \mathbf{A}$ is positive semidefinite, $\mathbf{A} \preceq \mathbf{B}$ means that there exists an absolute constant $\delta > 0$ such that $\mathbf{A} \leq \delta\mathbf{B}$ for sufficiently large I, and $\mathbf{A} \simeq \mathbf{B}$ means that $\mathbf{A} \preceq \mathbf{B}$ and also $\mathbf{B} \preceq \mathbf{A}$.

S.3.2 A General Result on Asymptotic Theory

We first establish a general result on the convergence rate of the sandwich smoother (Xiao et al., 2013) for covariance function estimation.

Let $\widehat{\mathbf{K}} \in \mathbb{R}^{L\times L}$ be an empirical estimate of $\mathbf{K} = \{K(s_{\ell_1}, s_{\ell_2})\}_{1\leq\ell_1,\ell_2\leq L}$ for some smooth covariance function $K(s,t)$ over $\mathcal{S}^2$. To obtain an estimate of K, we smooth $\widehat{\mathbf{K}}$, which may be noisy, via the sandwich smoother. Let $\mathbf{B}(s) = \{B_1(s), \ldots, B_c(s)\}^T$ be the vector of mth order B-spline bases constructed from $c - m$ equally-spaced interior knots in a bounded interval $\mathcal{S}$. Assume that the fixed design points $\{s_1, \ldots, s_L\}$ are equally-spaced in $\mathcal{T}$. Let $\mathbf{B} = [\mathbf{B}(s_1), \ldots, \mathbf{B}(s_L)]^T \in \mathbb{R}^{L\times c}$ be the design matrix, and $\mathbf{P} \in \mathbb{R}^{c\times c}$ be the qth order penalty matrix in P-splines (Eilers and Marx, 1996). Then the estimate from the sandwich smoother is $\widetilde{K}(s,t) = \mathbf{B}^T(s)\boldsymbol{\Theta}\mathbf{B}(t)$, where $\boldsymbol{\Theta} = (\mathbf{B}^T\mathbf{B}/L + \lambda\mathbf{P})^{-1}(\mathbf{B}^T\widehat{\mathbf{K}}\mathbf{B}/L^2)(\mathbf{B}^T\mathbf{B}/L + \lambda\mathbf{P})^{-1}$. Let $\mathbf{G}_L = \mathbf{B}^T\mathbf{B}/L$ and $\mathbf{V} = \mathbf{G}_L + \lambda\mathbf{P}$. Then $\boldsymbol{\Theta} = \mathbf{V}^{-1}(\mathbf{B}^T\widehat{\mathbf{K}}\mathbf{B}/L^2)\mathbf{V}^{-1}$. Notice that the above estimate is the same as that from the FACE method (Xiao et al., 2016) for smoothing single-level functional data. Assumption 3 ensures the following results.

Lemma S.1 *Suppose that Assumption 3 holds. Then, (a). $L^{-1}h_e^{-1} = O(1)$; (b). $\lambda h_e^{-q} = o(1)$.*

Theorem S.1 *Suppose that Assumption 3 holds. Also assume that $\mathbb{E}(\widehat{\mathbf{K}}) = \mathbf{K} + \delta\mathbf{I}_L$ for some constant $\delta \geq 0$ and $\left\|\mathrm{cov}\left\{\mathrm{vec}(\widehat{\mathbf{K}})\right\}\right\|_{\max} = O(I^{-1})$. If $K \in \mathcal{C}^p(\mathcal{S}^2)$ with $q \leq p \wedge m$, then,*

$$\mathbb{E}\left(\|\widetilde{K} - K\|_{L_2}^2\right) = \delta^2 O(L^{-2}h_e^{-1}) + O(h^{2m}) + o(h^{2p}) + O(\lambda^2 h_e^{-2q}) + O(I^{-1}).$$

In the derived rates in Theorem S.1, the first term $\delta^2 O(L^{-2}h_e^{-1})$ corresponds to the bias (when $\delta > 0$) in the empirical estimate $\widehat{K}$, the terms $O(h^{2m}) + o(h^{2p})$ is the approximation bias of spline functions, the term $O(\lambda^2 h_e^{-2q})$ is the shrinkage bias due to the smoothness penalty, and the last term $O(I^{-1})$ is the variability of the estimate. Except for the first term, the derived rates in Theorem S.1 are the same as those in Theorem 4.1 in Xiao (2020), which considers covariance function estimation using penalized splines for functional data under a fixed common design. For standard functional data without a multilevel structure, it can be easily shown that $\left\|\mathrm{cov}\left\{\mathrm{vec}(\widehat{\mathbf{K}})\right\}\right\|_{\max} = O(I^{-1})$ holds under proper conditions. Thus, Theorem S.1 establishes the convergence rates of the FACE method (Xiao et al., 2016).

S.3.3 Notations for Multilevel Functional Data and Technical Lemmas

Recall that for subject $i = 1, \dots, I$ at visit $j = 1, \dots, J_i$, the model for multilevel functional data is

$$Y_{ij}(s) = X_{ij}(s) + \epsilon_{ij}(s) = \mu(s) + \eta_j(s) + Z_i(s) + W_{ij}(s) + \epsilon_{ij}(s).$$

The data are observed at fixed and equally-spaced grid points $\{s_1, \dots, s_L\}$ so that the actual observations are $\{Y_{ij\ell} = Y_{ij}(s_\ell), 1 \le \ell \le L, 1 \le j \le J_i, 1 \le i \le I\}$.

To simplify theoretical analysis, we shall assume that $\mu(s)$ and $\eta_j(s)$ are known and let $\widetilde{Y}_{ij}(s) = Y_{ij}(s) - \mu(s) - \eta_j(s)$ as well as $\widetilde{Y}_{ij\ell} = \widetilde{Y}_{ij}(s_\ell)$. Moreover, we shall assume that the subjects have the same number of visits so that $J_i = J$ and J is a finite number; then it is sensible to to use the same weights when aggregating the curves, i.e., $w_i = (nJ)^{-1}$ and $v_i = \{nJ(J-1)\}^{-1}$. Let $\widetilde{\mathbf{Y}}_{ij} = [\widetilde{Y}_{ij}(s_1), \dots, \widetilde{Y}_{ij}(s_L)]^T$. Then, the MoM estimators are

$$\begin{aligned}
\widehat{\mathbf{K}}_T &= \frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} \widetilde{\mathbf{Y}}_{ij} \widetilde{\mathbf{Y}}_{ij}^T, \\
\widehat{\mathbf{K}}_W &= \frac{1}{I(J-1)} \sum_{i=1}^{I} \sum_{j=1}^{J} (\widetilde{\mathbf{Y}}_{ij} - \bar{\mathbf{Y}}_{i\cdot})(\widetilde{\mathbf{Y}}_{ij} - \bar{\mathbf{Y}}_{i\cdot})^T,
\end{aligned}$$

where $\widetilde{\mathbf{Y}}_{i\cdot} = J^{-1} \sum_{j=1}^{J} \widetilde{\mathbf{Y}}_{ij}$. Let $\mathbf{\Theta}_T = \mathbf{V}^{-1}(\mathbf{B}^T \widehat{\mathbf{K}}_T \mathbf{B}/L^2)\mathbf{V}^{-1}$, and $\mathbf{\Theta}_W = \mathbf{V}^{-1}(\mathbf{B}^T \widehat{\mathbf{K}}_W \mathbf{B}/L^2)\mathbf{V}^{-1}$. It follows that the estimates from our multilevel FACE method are

$$\begin{aligned}
\widetilde{K}_T(s,t) &= \mathbf{B}^T(s)\mathbf{\Theta}_T\mathbf{B}(t), \\
\widetilde{K}_W(s,t) &= \mathbf{B}^T(s)\mathbf{\Theta}_W\mathbf{B}(t), \\
\widetilde{K}_B(s,t) &= \widetilde{K}_T(s,t) - \widetilde{K}_W(s,t).
\end{aligned}$$

Note that here for simplicity, we assume the same smoothing parameter λ is used for both $\widetilde{K}_T$ and $\widetilde{K}_W$.

Recall that $\mathbf{G}_L = \mathbf{B}^T\mathbf{B}/L$, $\mathbf{G} = \int \mathbf{B}(t)\mathbf{B}^T(t)dt$, $\mathbf{V} = \mathbf{G}_L + \lambda\mathbf{P}$. Also we use $h = c^{-1}$ and $h_e = h \vee \lambda^{1/(2q)}$. We next introduce some lemmas used in our proof.

Lemma S.2 *Suppose that Assumption 3 holds. Then, (a).* $\mathbf{G}_L \simeq h\mathbf{I}_c$*; (b).* $\mathbf{G} \simeq h\mathbf{I}_c$*; (c).* $\|\mathbf{V}^{-1}\|_{\max} = O(h_e^{-1})$*; (d).* $\|\mathbf{V}^{-1}\|_\infty = O(h^{-1})$*; (e).* $\|\mathbf{V}^{-1}\mathbf{P}\|_\infty = O(h_e^{-q})$*; (f).* $\mathrm{tr}(\mathbf{V}^{-r}) = O(h^{-r}h_e^{-1})$ *for a positive integer* $r \ge 2$.

Remark S.1 *Lemma S.2 (a) and (b) hold by Lemma A.1 and Remark A.1 in Xiao (2019). Parts (c), (d), and (e) hold by Proposition 3 and Proposition 4 in Xiao and Nan (2020). Part (f) can be easily established by a proof similar to that Lemma 5.2 in Xiao (2019).*

We also use frequently the following results, which are easy to prove and hence the details are omitted.

Lemma S.3 *Let* $\mathbf{A}_1 \in \mathbb{R}^{c\times c}$, $\mathbf{A}_2 \in \mathbb{R}^{c\times c}$ *and both are symmetric matrices. Then, (a).* $\|\mathbf{A}_1\|_F \leq c\|\mathbf{A}_1\|_{\max}$*; (b).* $\|\mathbf{A}_1\mathbf{A}_2\|_{\max} \leq \|\mathbf{A}_1\|_\infty \|\mathbf{A}_2\|_{\max}$*; (c). If both* $\mathbf{A}_1$ *and* $\mathbf{A}_2$ *are positive semi-definite, and* $\mathbf{A}_2 \simeq h\mathbf{I}_c$*, then (i)* $\|\mathbf{A}_1\mathbf{A}_2\|_F \simeq h\|\mathbf{A}_1\|_F$*, and (ii)* $\mathrm{tr}(\mathbf{A}_1\mathbf{A}_2\mathbf{A}_1) \simeq h\mathrm{tr}(\mathbf{A}_1^2)$.

Lemma S.4 *Let* $g(s,t) = \mathbf{B}^T(s)\mathbf{A}\mathbf{B}(t)$ *for some matrix* $\mathbf{A} \in \mathbb{R}^{c\times c}$*. Then* $\|g\|_{L_2}^2 = \|\mathbf{A}\mathbf{G}\|_F^2$.

Proof S.1 (Proof of Lemma S.4) *We derive that*

$$\begin{aligned} g^2(s,t) &= \left\{\mathbf{B}^T(s)\mathbf{A}\mathbf{B}(t)\right\}^2 \\ &= \mathbf{B}^T(s)\mathbf{A}\mathbf{B}(t)\mathbf{B}^T(t)\mathbf{A}^T\mathbf{B}(s) \\ &= \mathrm{tr}\left\{\mathbf{A}\mathbf{B}(t)\mathbf{B}^T(t)\mathbf{A}^T\mathbf{B}(s)\mathbf{B}^T(s)\right\}. \end{aligned}$$

Thus,

$$\begin{aligned} \|g\|_{L_2}^2 &= \iint g^2(s,t)dsdt \\ &= \iint \mathrm{tr}\left\{\mathbf{A}\mathbf{B}(t)\mathbf{B}^T(t)\mathbf{A}^T\mathbf{B}(s)\mathbf{B}^T(s)\right\} dsdt \\ &= \mathrm{tr}\left(\mathbf{A}\mathbf{G}\mathbf{A}^T\mathbf{G}\right) \\ &= \|\mathbf{A}\mathbf{G}\|_F^2. \end{aligned}$$

The proof is now complete.

Lemma S.5 *Let* $\mathbf{\Gamma} = \mathbf{G}_L^{-1}\mathbf{B}^T\mathbf{K}\mathbf{B}\mathbf{G}_L^{-1}/L^2$ *as in* (S3)*. Suppose that Assumption 3 holds and* $K(s,t)$ *is continuous in* $\mathcal{T}^2$*. Then,*

$$\|\mathbf{\Gamma}\|_{\max} = O(1).$$

Proof S.2 (Proof of Lemma S.5) *Let* $\mathbf{A} = (a_{j_1 j_2}) = \mathbf{B}^T\mathbf{K}\mathbf{B}/L^2$*. Then it is easy to derive that* $\|\mathbf{A}\|_{\max} = O(h^2)$*. Thus,*

$$
\begin{aligned}
\|\boldsymbol{\Gamma}\|_{\max} =& \|\mathbf{G}_L^{-1}\mathbf{A}\mathbf{G}_L^{-1}\|_{\max} \\
\le& \|\mathbf{G}_L^{-1}\|_{\infty}^2 \|\mathbf{A}\|_{\max} \\
\preceq& h^{-2}h^2 \\
\preceq& 1,
\end{aligned}
$$

which concludes the proof.

S.3.4 Proofs of Asymptotic Theory

Proof S.3 (Proof of Theorem S.1) *First note that we shall make frequent uses of results in Lemma S.2 and Lemma S.3. By the bias and variance decomposition, it is easy to derive that*

$$
\mathbb{E}\left(\|\widetilde{K} - K\|_{L_2}^2\right) = \iint \left[\mathbb{E}\left\{\widetilde{K}(s,t)\right\} - K(s,t)\right]^2 dsdt + \iint \operatorname{var}\left\{\widetilde{K}(s,t)\right\} dsdt. \tag{S1}
$$

We first consider the bias and let $b(s,t) = \mathbb{E}\{\widetilde{K}(s,t)\} - K(s,t)$*. Then,*

$$
b(s,t) = \mathbf{B}^T(s)\mathbf{V}^{-1}(\mathbf{B}^T\mathbf{K}\mathbf{B}/L^2)\mathbf{V}^{-1}\mathbf{B}(t) + \delta L^{-1}\mathbf{B}^T(s)\mathbf{V}^{-1}\mathbf{G}_L\mathbf{V}^{-1}\mathbf{B}(t) - K(s,t).
$$

Let

$$
b_0(s,t) = \delta L^{-1}\mathbf{B}^T(s)\mathbf{V}^{-1}\mathbf{G}_L\mathbf{V}^{-1}\mathbf{B}(t). \tag{S2}
$$

To simplify notation, define

$$
\boldsymbol{\Gamma} = \mathbf{G}_L^{-1}\mathbf{B}^T\mathbf{K}\mathbf{B}\mathbf{G}_L^{-1}/L^2. \tag{S3}
$$

Notice that $\mathbf{V}^{-1}\mathbf{G}_L = \mathbf{I} - \mathbf{V}^{-1}(\lambda\mathbf{P})$*. Then,*

$$
\begin{aligned}
b(s,t) &= \mathbf{B}^T(s)\mathbf{V}^{-1}\mathbf{G}_L\boldsymbol{\Gamma}\mathbf{G}_L\mathbf{V}^{-1}\mathbf{B}(t) - K(s,t) + b_0(s,t) \\
&= \mathbf{B}^T(s)\left\{\mathbf{I} - \mathbf{V}^{-1}(\lambda\mathbf{P})\right\}\boldsymbol{\Gamma}\left\{\mathbf{I} - (\lambda\mathbf{P})\mathbf{V}^{-1}\right\}\mathbf{B}(t) - K(s,t) + b_0(s,t) \\
&= \mathbf{B}^T(s)\left\{\boldsymbol{\Gamma} - \mathbf{V}^{-1}(\lambda\mathbf{P})\boldsymbol{\Gamma} - \boldsymbol{\Gamma}(\lambda\mathbf{P})\mathbf{V}^{-1} + \mathbf{V}^{-1}(\lambda\mathbf{P})\boldsymbol{\Gamma}(\lambda\mathbf{P})\mathbf{V}^{-1}\right\}\mathbf{B}(t) - K(s,t) + b_0(s,t).
\end{aligned}
$$

Let

$$\begin{aligned} b_1(s,t) &= \mathbf{B}^T(s)\mathbf{\Gamma}\mathbf{B}(t) - K(s,t), &\text{(S4)}\\ b_2(s,t) &= \mathbf{B}^T(s)\mathbf{V}^{-1}(\lambda\mathbf{P})\mathbf{\Gamma}\mathbf{B}(t), &\text{(S5)}\\ b_3(s,t) &= \mathbf{B}^T(s)\mathbf{V}^{-1}(\lambda\mathbf{P})\mathbf{\Gamma}(\lambda\mathbf{P})\mathbf{V}^{-1}\mathbf{B}(t). &\text{(S6)}\end{aligned}$$

Then,

$$b(s,t) = b_0(s,t) + b_1(s,t) - b_2(s,t) - b_2(t,s) + b_3(s,t). \tag{S7}$$

It follows that

$$\|b\|_{L_2}^2 \le 5\left(\|b_0\|_{L_2}^2 + \|b_1\|_{L_2}^2 + 2\|b_2\|_{L_2}^2 + \|b_3\|_{L_2}^2\right). \tag{S8}$$

We first consider b_0 in (S2) *and derive that*

$$\begin{aligned}\|b_0\|_{L_2}^2 =&\delta^2 L^{-2}\mathrm{tr}\left(\mathbf{V}^{-1}\mathbf{G}_L\mathbf{V}^{-1}\mathbf{G}\mathbf{V}^{-1}\mathbf{G}_L\mathbf{V}^{-1}\mathbf{G}\right)\\ \preceq&\delta^2 L^{-2}h^4\mathrm{tr}(\mathbf{V}^{-4})\\ \preceq&\delta^2 L^{-2}h^4h^{-4}h_e^{-1}\\ \preceq&\delta^2 L^{-2}h_e^{-1}.\end{aligned}$$

So we have

$$\|b_0\|_{L_2}^2 = \delta^2 O(L^{-2}h_e^{-1}). \tag{S9}$$

We now consider b_1 in (S4). *Let $K_\Delta(s,t) = \mathbf{B}^T(s)\mathbf{\Delta}\mathbf{B}(t)$ be the spline function as in Lemma S.3.1 in the supplement of Xiao (2020) and $\mathbf{K}_\Delta = \{K_\Delta(t_{\ell_1}, t_{\ell_2})\}_{1\le\ell_1,\ell_2\le L}$. By equation* (S4) *and the definition of $\mathbf{\Gamma}$ in* (S3), *we derive that*

$$\begin{aligned} b_1(s,t) =&\mathbf{B}^T(s)\left(\mathbf{G}_L^{-1}\mathbf{B}^T\mathbf{K}\mathbf{B}\mathbf{G}_L^{-1}/L^2\right)\mathbf{B}(t) - K(s,t)\\ =&\mathbf{B}^T(s)\left(\mathbf{G}_L^{-1}\mathbf{B}^T\mathbf{K}_\Delta\mathbf{B}\mathbf{G}_L^{-1}/L^2\right)\mathbf{B}(t) + \mathbf{B}^T(s)\left\{\mathbf{G}_L^{-1}\mathbf{B}^T(\mathbf{K}-\mathbf{K}_\Delta)\mathbf{B}\mathbf{G}_L^{-1}/L^2\right\}\mathbf{B}(t) - K(s,t)\\ =&K_\Delta(s,t) - K(s,t) + \mathbf{B}^T(s)\left\{\mathbf{G}_L^{-1}\mathbf{B}^T(\mathbf{K}-\mathbf{K}_\Delta)\mathbf{B}\mathbf{G}_L^{-1}/L^2\right\}\mathbf{B}(t).\end{aligned}$$

In the above derivations, we used the facts that $\mathbf{K}_\Delta = \mathbf{B}\mathbf{\Delta}\mathbf{B}^T$ and $\mathbf{G}_L = \mathbf{B}^T\mathbf{B}/L$. Define

$$\alpha(s,t) = \mathbf{B}^T(s)\left\{\mathbf{G}_L^{-1}\mathbf{B}^T(\mathbf{K}-\mathbf{K}_\Delta)\mathbf{B}\mathbf{G}_L^{-1}/L^2\right\}\mathbf{B}(t).$$

Then,

$$b_1(s,t) = (K_\Delta - K)(s,t) + \alpha(s,t),$$

and it follows that

$$\|b_1\|_{L_2}^2 \le 2\left(\|K_\Delta - K\|_{L_2}^2 + \|\alpha\|_{L_2}^2\right). \tag{S10}$$

By Lemma S.3.1 in the supplement of Xiao (2020),

$$\|K_\Delta - K\|_{L_2}^2 = O(h^{2m}) + o(h^{2p}). \tag{S11}$$

Applying Lemma S.4 to α, we have

$$\|\alpha\|_{L_2}^2 = \|\mathbf{G}_L^{-1}\mathbf{B}^T(\mathbf{K}-\mathbf{K}_\Delta)\mathbf{B}\mathbf{G}_L^{-1}\mathbf{G}/L^2\|_F^2.$$

Let $\mathbf{D} = (d_{j_1j_2}) = \mathbf{B}^T(\mathbf{K}-\mathbf{K}_\Delta)\mathbf{B}/L^2$ so that

$$\|\alpha\|_{L_2}^2 = \|\mathbf{G}_L^{-1}\mathbf{D}\mathbf{G}_L^{-1}\mathbf{G}\|_F^2. \tag{S12}$$

Notice that

$$\begin{aligned} d_{j_1j_2} &= \frac{1}{L^2}\sum_{1\le \ell_1,\ell_2\le L} B_{j_1}(s_{\ell_1})B_{j_2}(s_{\ell_2})\left\{K(s_{\ell_1},s_{\ell_2}) - K_\Delta(s_{\ell_1},s_{\ell_2})\right\} \\ &= \iint B_{j_1}(s)B_{j_2}(t)(K-K_\Delta)(s,t)d_sd_t\left\{F(s)F(t)\right\}, \end{aligned}$$

where $F(t) = L^{-1}\sum_{1\le\ell\le L} 1_{\{s_\ell \le t\}}$. By Lemma S.3.2 in the supplement of Xiao (2020),

$$\|\mathbf{D}\|_{\max} = o(h^{p+2}).$$

Thus, by equation (S12) *and the facts that $\mathbf{G} \simeq h\mathbf{I}$ and $\|\mathbf{G}_L^{-1}\|_\infty = O(h^{-1})$, we derive that*

$$\begin{aligned} \|\alpha\|_{L_2}^2 &\preceq h^2\|\mathbf{G}_L^{-1}\mathbf{D}\mathbf{G}_L^{-1}\|_F^2 \\ &\preceq \|\mathbf{G}_L^{-1}\mathbf{D}\mathbf{G}_L^{-1}\|_{\max}^2 \\ &\preceq \|\mathbf{G}_L^{-1}\|_\infty^4\|\mathbf{D}\|_{\max}^2 \\ &\ll h^{-4}h^{2(p+2)} \\ &\ll h^{2p}. \end{aligned}$$

So $\|\alpha\|_{L_2}^2 = o(h^{2p})$ and together with equations (S10) *and* (S11)*, we have*

$$\|b_1\|_{L_2}^2 = O(h^{2m}) + o(h^{2p}). \tag{S13}$$

For the bias, next we consider b_2 in (S5)*. Applying Lemma S.4 to b_2, we have*

$$\|b_2\|_{L_2}^2 = \|\mathbf{V}^{-1}(\lambda\mathbf{P})\mathbf{\Gamma}\mathbf{G}\|_F^2,$$

By Lemma S.2, $\|\mathbf{V}^{-1}(\lambda\mathbf{P})\|_\infty = O(\lambda h_e^{-q})$. We shall also use the result $\|\mathbf{\Gamma}\|_{\max} = O(1)$ by Lemma S.5. Thus, we derive that

$$\begin{aligned}\|\mathbf{V}^{-1}(\lambda\mathbf{P})\mathbf{\Gamma}\mathbf{G}\|_F^2 \preceq & h^2\|\mathbf{V}^{-1}(\lambda\mathbf{P})\mathbf{\Gamma}\|_F^2 \\ \preceq & \|\mathbf{V}^{-1}(\lambda\mathbf{P})\mathbf{\Gamma}\|_{\max}^2 \\ \preceq & \|\mathbf{V}^{-1}(\lambda\mathbf{P})\|_\infty^2\|\mathbf{\Gamma}\|_{\max}^2 \\ \preceq & \lambda^2 h_e^{-2q}.\end{aligned}$$

So we have derived that

$$\|b_2\|_{L_2}^2 = O(\lambda^2 h_e^{-2q}). \tag{S14}$$

For the bias, finally we consider b_3 in (S6)*. Applying Lemma S.4 to b_3, we have*

$$\|b_3\|_{L_2}^2 = \|\mathbf{V}^{-1}(\lambda\mathbf{P})\mathbf{\Gamma}(\lambda\mathbf{P})\mathbf{V}^{-1}\mathbf{G}\|_F^2.$$

Then,

$$\begin{aligned}\|b_3\|_{L_2}^2 \preceq & \|\mathbf{V}^{-1}(\lambda\mathbf{P})\mathbf{\Gamma}(\lambda\mathbf{P})\mathbf{V}^{-1}\|_{\max}^2 \\ \preceq & \|\mathbf{V}^{-1}(\lambda\mathbf{P})\|_\infty^4\|\mathbf{\Gamma}\|_{\max}^2 \\ \preceq & \lambda^4 h_e^{-4q}.\end{aligned}$$

Note that $\lambda h_e^{-q} = o(1)$ by Lemma S.1. So we have derived that

$$\|b_3\|_{L_2}^2 = o(\lambda^2 h_e^{-2q}). \tag{S15}$$

Now combining equations (S8), (S9), (S13), (S14) *and* (S15)*, we have*

$$\|b\|_{L_2}^2 = \delta^2 O(L^{-2}h_e^{-1}) + O(h^{2m}) + o(h^{2p}) + O(\lambda^2 h_e^{-2q}). \tag{S16}$$

Now we have derived the rate of the bias, we consider the variance of the estimate. We have $\widetilde{K}(s,t) = \mathbf{B}^T(s)\mathbf{\Theta}\mathbf{B}(t) = \{\mathbf{B}(t)\otimes\mathbf{B}(s)\}^T\mathrm{vec}(\mathbf{\Theta})$, where $\mathrm{vec}(\cdot)$ is the matrix *operation that stacks the columns of the matrix into a single column vector. Recall that* $\boldsymbol{\Theta} = \mathbf{V}^{-1}(\mathbf{B}^T\widehat{\mathbf{K}}\mathbf{B}/L^2)\mathbf{V}^{-1}$. *Thus, we derive that*

$$\mathrm{vec}(\boldsymbol{\Theta}) = (\mathbf{B}\mathbf{V}^{-1}/L \otimes \mathbf{B}\mathbf{V}^{-1}/L)^T \mathrm{vec}(\widehat{\mathbf{K}})$$

and also

$$\widetilde{K}(s,t) = (\mathbf{B}\mathbf{V}^{-1}\mathbf{B}(t)/L \otimes \mathbf{B}\mathbf{V}^{-1}\mathbf{B}(s)/L)^T \mathrm{vec}(\widehat{\mathbf{K}})$$

Let $\mathbf{a}(t) = \{a_1(t), \ldots, a_L(t)\}^T = \mathbf{B}\mathbf{V}^{-1}\mathbf{B}(t)/L$ *so that* $a_\ell(t) = \mathbf{B}^T(s_\ell)\mathbf{V}^{-1}\mathbf{B}(t)/L$. *Then,*

$$\widetilde{K}(s,t) = \{\mathbf{a}(t) \otimes \mathbf{a}(s)\}^T \mathrm{vec}(\widehat{\mathbf{K}}).$$

For a matrix $\mathbf{A} = (a_{ij})$, *let* $(\mathbf{A})_+ = (|a_{ij}|)$. *Notice that* $\|(\mathbf{A})_+\|_\infty = \|\mathbf{A}\|_\infty$. *Next let* $\tilde{a}_\ell(t) = \mathbf{B}^T(s_\ell)(\mathbf{V}^{-1})_+\mathbf{B}(t)/L$, *then* $|a_\ell(t)| \le \tilde{a}_\ell(t)$. *Therefore,*

$$\begin{aligned}
\mathrm{var}\left\{\widetilde{K}(s,t)\right\} =& \{\mathbf{a}(t) \otimes \mathbf{a}(s)\}^T \mathrm{cov}\left\{\mathrm{vec}(\widehat{\mathbf{K}})\right\} \{\mathbf{a}(t) \otimes \mathbf{a}(s)\} \\
\le& \left\|\mathrm{cov}\left\{\mathrm{vec}(\widehat{\mathbf{K}})\right\}\right\|_{\max} \left\{\sum_\ell \tilde{a}_\ell(t)\right\}^2 \left\{\sum_\ell \tilde{a}_\ell(s)\right\}^2.
\end{aligned}$$

It follows that

$$\iint \mathrm{var}\left\{\widetilde{K}(s,t)\right\} dsdt \le \left\|\mathrm{cov}\left\{\mathrm{vec}(\widehat{\mathbf{K}})\right\}\right\|_{\max} \left[\int \left\{\sum_\ell \tilde{a}_\ell(t)\right\}^2 dt\right]^2.$$

We derive that

$$\begin{aligned}
\int \left\{\sum_\ell \tilde{a}_\ell(t)\right\}^2 dt =& \int \left\{\mathbf{1}_L^T\mathbf{B}(\mathbf{V}^{-1})_+\mathbf{B}(t)/L\right\}^2 dt \\
=& L^{-2}\mathbf{1}_L^T\mathbf{B}(\mathbf{V}^{-1})_+\mathbf{G}(\mathbf{V}^{-1})_+\mathbf{B}^T\mathbf{1}_L \\
\preceq& L^{-2}h\mathbf{1}_L^T\mathbf{B}(\mathbf{V}^{-1})_+(\mathbf{V}^{-1})_+\mathbf{B}^T\mathbf{1}_L \\
\preceq& L^{-2}\|(\mathbf{V}^{-1})_+\|_\infty^2\|\mathbf{B}^T\mathbf{1}_L\|_{\max}^2 \\
\preceq& L^{-2}h^{-2}L^2h^2 \\
\preceq& 1.
\end{aligned}$$

Therefore,

$$\iint \mathrm{var}\left\{\widetilde{K}(s,t)\right\} dsdt = O\left[\left\|\mathrm{cov}\left\{\mathrm{vec}(\widehat{\mathbf{K}})\right\}\right\|_{\max}\right]. \tag{S17}$$

Combining (S1), (S16) *and* (S17), *the proof is complete.*

Proof S.4 (Proof of Theorem 1) *It suffices to verify that*

$$\left\|\mathrm{cov}\left\{\mathrm{vec}(\widehat{\mathbf{K}}_T)\right\}\right\|_{\max} = O(I^{-1})$$

and

$$\left\|\mathrm{cov}\left\{\mathrm{vec}(\widehat{\mathbf{K}}_W)\right\}\right\|_{\max} = O(I^{-1}).$$

We use $\mathrm{cov}\left\{\mathrm{vec}(\widehat{\mathbf{K}}_T)\right\}$ *as an example, since* $\mathrm{cov}\left\{\mathrm{vec}(\widehat{\mathbf{K}}_W)\right\}$ *can be shown similarly. Notice that the* $(\ell_1 + (\ell_2 - 1) \times L)$*th entry of* $\mathrm{vec}(\widehat{\mathbf{K}}_T)$ *is* $(IJ)^{-1}\sum_{i=1}^{I}\sum_{j=1}^{J}\widetilde{Y}_{ij}(s_{\ell_1})\widetilde{Y}_{ij}(s_{\ell_2})$. *Therefore, the* $(\ell_1 + (\ell_2 - 1) \times L, \ell_3 + (\ell_4 - 1) \times L)$*th entry of* $\mathrm{cov}\left\{\mathrm{vec}(\widehat{\mathbf{K}}_T)\right\}$ *is*

$$E\{\frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\widetilde{Y}_{ij}(s_{\ell_1})\widetilde{Y}_{ij}(s_{\ell_2}) \times \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\widetilde{Y}_{ij}(s_{\ell_3})\widetilde{Y}_{ij}(s_{\ell_4})\} -$$
$$E\{\frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\widetilde{Y}_{ij}(s_{\ell_1})\widetilde{Y}_{ij}(s_{\ell_2})\} \times E\{\frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\widetilde{Y}_{ij}(s_{\ell_3})\widetilde{Y}_{ij}(s_{\ell_4})\}.$$

By Assumption 1, we have

$$\begin{aligned}
&E\{\frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\widetilde{Y}_{ij}(s_{\ell_1})\widetilde{Y}_{ij}(s_{\ell_2}) \times \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\widetilde{Y}_{ij}(s_{\ell_3})\widetilde{Y}_{ij}(s_{\ell_4})\} \\
=&\frac{1}{I^2J^2}E\{\sum_{i=1}^{I}(\sum_{j_1=1}^{J}\sum_{j_2=1}^{J}\widetilde{Y}_{ij_1}(s_{\ell_1})\widetilde{Y}_{ij_1}(s_{\ell_2})\widetilde{Y}_{ij_2}(s_{\ell_3})\widetilde{Y}_{ij_2}(s_{\ell_4}))\} \\
=&\frac{1}{I^2J^2}\sum_{i=1}^{I}(\sum_{j_1=1}^{J}\sum_{j_2=1}^{J}E\{\widetilde{Y}_{ij_1}(s_{\ell_1})\widetilde{Y}_{ij_1}(s_{\ell_2})\widetilde{Y}_{ij_2}(s_{\ell_3})\widetilde{Y}_{ij_2}(s_{\ell_4})\}).
\end{aligned}$$

It is easy to show that

$$E\{\widetilde{Y}_{ij_1}(s_{\ell_1})\widetilde{Y}_{ij_1}(s_{\ell_2})\widetilde{Y}_{ij_2}(s_{\ell_3})\widetilde{Y}_{ij_2}(s_{\ell_4}))\} \le 3^3(\sup_{s\in\mathcal{S}}\mathbb{E}[Z_i^4(s)] + \sup_{j\in\{j_1,j_2\}}(\sup_{s\in\mathcal{S}}\mathbb{E}[W_{ij}^4(s)] + \mathbb{E}[\epsilon_{ij\ell}^4])).$$

using Cauchy-Schwarz inquality. By Assumption 2, we have

$$E\{\frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\widetilde{Y}_{ij}(s_{\ell_1})\widetilde{Y}_{ij}(s_{\ell_2}) \times \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\widetilde{Y}_{ij}(s_{\ell_3})\widetilde{Y}_{ij}(s_{\ell_4})\} = O(I^{-1}),$$

for any $\ell_1, \ell_2, \ell_3, \ell_4$. *Thus* $\left\|\mathrm{cov}\left\{\mathrm{vec}(\widehat{\mathbf{K}}_T)\right\}\right\|_{\max} = O(I^{-1})$.

S.4 Additional Simulation Results

S.4.1 Higher Rank/frequency of Eigenfunctions

We conduct additional simulations to study the effect of having higher ranks on both levels with higher frequency eigenfunctions. We assume the functions are observed on an equally-spaced grid $\{s_1, \ldots, s_L\}$ of $\mathcal{S} = [0, 1]$ such that $s_l = l/L$ for $l = 1, \ldots, L$. Following the simulation setting described in the paper, we assume $J_i = J = 2$ (balanced), $L = 100$, and $I \in \{100, 200, 1000, 5000\}$. In the paper, we assume four eigenfunctions of each level. Here, we increase the ranks of each level and propose the following additional simulation settings:

Higher ranks of level-1 $N_1 = 8$, $N_2 = 4$, where $\lambda^{(1)}_{k_1} = 0.5^{k_1-1}, k_1 = 1, \ldots, 8$ and $\lambda^{(2)}_{k_2} = 0.5^{k_2-1}, k_2 = 1, \ldots, 4$. The true eigenfunctions are

Level 1: $\phi_{k_1}(s) = \{\sqrt{2}\sin(2\pi s), \sqrt{2}\cos(2\pi s), \sqrt{2}\sin(4\pi s), \sqrt{2}\cos(4\pi s), \sqrt{2}\sin(6\pi s), \sqrt{2}\cos(6\pi s), \sqrt{2}\sin(8\pi s), \sqrt{2}\cos(8\pi s)\}$.

Level 2: $\psi_{k_2}(s) = \{1, \sqrt{3}(2s - 1), \sqrt{5}(6s^2 - 6s + 1), \sqrt{7}(20s^3 - 30s^2 + 12s - 1)\}$.

Higher ranks of level-2 $N_1 = 4$, $N_2 = 8$, where $\lambda^{(1)}_{k_1} = 0.5^{k_1-1}, k_1 = 1, \ldots, 4$ and $\lambda^{(2)}_{k_2} = 0.5^{k_2-1}, k_2 = 1, \ldots, 8$. The true eigenfunctions are

Level 1: $\phi_{k_1}(s) = \{\sqrt{2}\sin(2\pi s), \sqrt{2}\cos(2\pi s), \sqrt{2}\sin(4\pi s), \sqrt{2}\cos(4\pi s)\}$.

Level 2: $\psi_{k_2}(s) = \{1, \sqrt{3}(2s-1), \sqrt{5}(6s^2-6s+1), \sqrt{7}(20s^3-30s^2+12s-1), 3(70s^4-140s^3+90s^2-20s+1), \sqrt{11}(252s^5-630s^4+560s^3-210s^2+30s-1), \sqrt{13}(924s^6-2772s^5+3150s^4-1680s^3+420s^2-42s+1), \sqrt{15}(3432s^7-12012s^6+16632s^5-11550s^4+4200s^3-756s^2+56s-1)\}$.

Given that higher ranks of eigenfunctions in each level leads to very small eigenvalues, we evaluate the covariance estimation accuracy in each setting. Specifically, we calculate $\text{RISE}(\mathbf{K}_B) = \int_s\int_t(\widetilde{K}_B(s,t) - K_B(s,t))^2 dtds / \int_s\int_t(K_B(s,t))^2 dtds$ for between-subject covariance, and $\text{RISE}(\mathbf{K}_W) = \int_s\int_t(\widetilde{K}_W(s,t) - K_W(s,t))^2 dtds / \int_s\int_t(K_W(s,t))^2 dtds$ for within-subject covariance, where $\widetilde{K}_B(s,t)$ and $\widetilde{K}_W(s,t)$ are obtained from fast MFPCA or MFPCA method.

The simulation results are shown Table S1. The first block shows the results of baseline setting ($N_1 = 4, N_2 = 4$). The second block shows the results when we double the ranks of level-1 eigenfunctions ($N_1 = 8, N_2 = 4$), and the third block shows the results when we double the ranks of level-2 eigenfunctions ($N_1 = 4, N_2 = 8$). In all simulations settings, we observe similar between-subject covariance estimation accuracy between fast MFPCA and MFPCA, and slightly better within-subject covariance estimation accuracy using proposed fast MFPCA method especially when the number of subjects is small. In addition, the computation of fast MFPCA is much faster than MFPCA, ranging from about 10 times when $I = 100$ to over 100 times when $I = 5000$.

In summary, for higher ranks on both levels with higher frequency eigenfunctions, fast MFPCA exhibit similar or slightly higher estimation accuracy than traditional MFPCA approach, while the computation is orders of magnitude faster. Therefore, fast MFPCA can be applied to very large datasets.

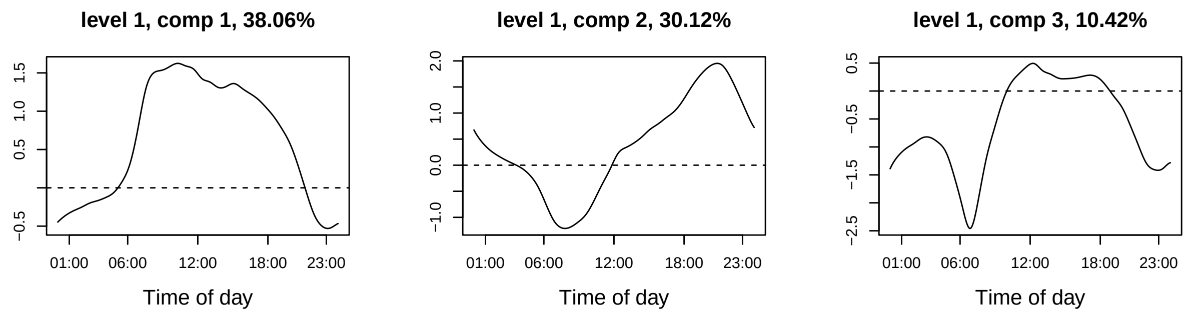
S.4.2 Visit- and Subject-specific Weights

Table S2-S4 show the simulation results using different methods, including 1) fast MFPCA weighted by visit; 2) fast MFPCA weighted by subject; 3) traditional MFPCA. For unbalanced data, fast MFPCA weighted by subject exhibits a better estimation performance of level-1 eigenfunctions than fast MFPCA weighted by visit especially when the number of subjects is small. The performance of both methods is comparable as number of visits per subject or the dimension of the functional domain increase.

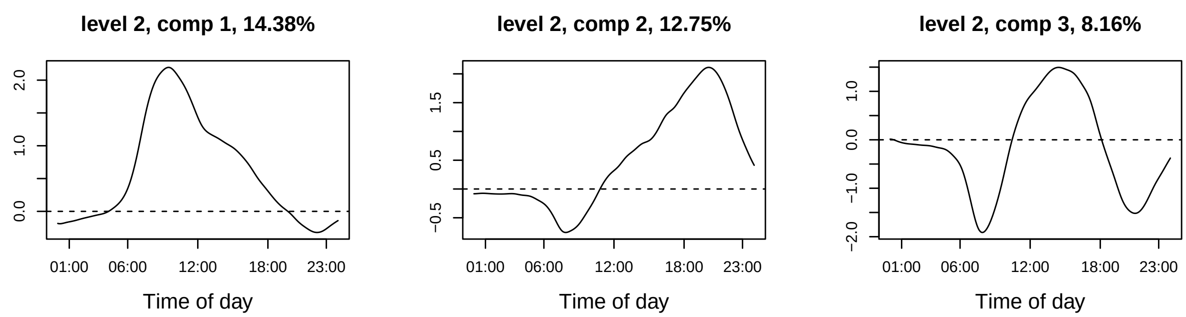
S.5 Additional Application Results

In the paper, we assume that the physical activity profile has a day-of-week-specific shift within each study participant in the NHANES application. Here we perform additional analysis by treating weekday and weekend average as the only two visit-specific shift functions.

Figure S1 shows the first three level-1 and level-2 eigenfunctions under this application setting. The shape of each eigenfunction at each level is almost identical to the shape in our original analysis, and the difference in the proportion of variability explained in each principal component is less than 0.2% compared to our original analysis. Given that the computation time is almost identical between original and this simplified analysis, we conclude that there is no substantial advantage in considering a model with only weekday and weekend average in this study.



(a) The first three estimated level-1 eigenfunctions.



(b) The first three estimated level-2 eigenfunctions.

Figure S1: The top three estimated level-1 (first row) and level-2 (second row) eigenfunctions from the NHANES dataset using fast MFPCA when treating weekday and weekend average as two visit-specific shift functions. The proportion of variability explained in each principal component within each level is shown on the title of each panel.

Table S1: Simulation results of complete balanced data under different ranks of eigenfunctions in each level and different I when $J = 2$ and $L = 100$. The first block shows the results of baseline setting ($N_1 = 4, N_2 = 4$). The second block shows the results when we double the ranks of level-1 eigenfunctions ($N_1 = 8, N_2 = 4$). The third block shows the results when we double the ranks of level-2 eigenfunctions ($N_1 = 4, N_2 = 8$). The computation time ("Time(s)"), MISE of $\mathbf{Y}$ ("MISE($\mathbf{Y}$)"), and RISE of between-subject ("RISE($\mathbf{K}_B$)") and within-subject covariance ("RISE($\mathbf{K}_W$)") reported in the table are median values across 100 replications.

I	Method	$N_1 = 4, N_2 = 4$			
		Time(s)	MISE($\mathbf{Y}$)	RISE($\mathbf{K}_B$)	RISE($\mathbf{K}_W$)
100	fast MFPCA	**0.93**	0.949	0.074	0.029
	MFPCA	**8.71**	0.948	0.074	0.032
200	fast MFPCA	**1.04**	0.951	0.034	0.015
	MFPCA	**17.33**	0.949	0.034	0.016
1000	fast MFPCA	**1.71**	0.953	0.007	0.004
	MFPCA	**122.33**	0.950	0.007	0.003
5000	fast MFPCA	**2.48**	0.953	0.002	0.001
	MFPCA	**320.81**	0.950	0.002	0.001
I	Method	$N_1 = 8, N_2 = 4$			
		Time(s)	MISE($\mathbf{Y}$)	RISE($\mathbf{K}_B$)	RISE($\mathbf{K}_W$)
100	fast MFPCA	**0.84**	0.937	0.065	0.032
	MFPCA	**8.50**	0.939	0.067	0.033
200	fast MFPCA	**0.79**	0.941	0.035	0.015
	MFPCA	**13.17**	0.942	0.036	0.015
1000	fast MFPCA	**0.96**	0.946	0.009	0.003
	MFPCA	**62.79**	0.944	0.009	0.003
5000	fast MFPCA	**2.66**	0.945	0.002	0.001
	MFPCA	**290.42**	0.942	0.002	0.001
I	Method	$N_1 = 4, N_2 = 8$			
		Time(s)	MISE($\mathbf{Y}$)	RISE($\mathbf{K}_B$)	RISE($\mathbf{K}_W$)
100	fast MFPCA	**0.92**	0.927	0.077	0.034
	MFPCA	**9.52**	0.937	0.077	0.034
200	fast MFPCA	**1.12**	0.929	0.036	0.018
	MFPCA	**20.01**	0.937	0.037	0.018
1000	fast MFPCA	**0.89**	0.933	0.007	0.004
	MFPCA	**57.47**	0.938	0.007	0.004
5000	fast MFPCA	**2.77**	0.933	0.002	0.001
	MFPCA	**325.17**	0.938	0.002	0.001

Table S2: Simulation results under different I when $J = 2$ and $L = 100$. The computation time ("Time(s)"), MISE for $\mathbf{Y}$ ("MISE($\mathbf{Y}$)") and eigenfunctions ("MISE($\boldsymbol{\phi}$)", "MISE($\boldsymbol{\psi}$)") reported in the table are median values across 100 simulations.

Design	I	Model	Balanced				Unbalanced			
			Time(s)	MISE($\mathbf{Y}$)	MISE($\boldsymbol{\phi}$)	MISE($\boldsymbol{\psi}$)	Time(s)	MISE($\mathbf{Y}$)	MISE($\boldsymbol{\phi}$)	MISE($\boldsymbol{\psi}$)
Complete	100	fast MFPCA (visit)	**1.81**	0.9450	0.0781	0.0319	**1.74**	0.9450	0.1203	0.0416
		fast MFPCA (subject)	**1.16**	0.9450	0.0781	0.0319	**1.26**	0.9446	0.1175	0.0487
		MFPCA	**14.82**	0.9451	0.0797	0.0315	**21.51**	0.9571	0.1930	0.1286
	200	fast MFPCA (visit)	**1.70**	0.9469	0.0413	0.0182	**1.88**	0.9465	0.0469	0.0229
		fast MFPCA (subject)	**1.48**	0.9469	0.0413	0.0182	**1.52**	0.9464	0.0429	0.0230
		MFPCA	**25.42**	0.9474	0.0377	0.0171	**48.90**	0.9530	0.0780	0.0698
	1000	fast MFPCA (visit)	**2.20**	0.9505	0.0093	0.0075	**2.07**	0.9524	0.0120	0.0063
		fast MFPCA (subject)	**1.42**	0.9505	0.0093	0.0075	**1.50**	0.9496	0.0139	0.0067
		MFPCA	**109.15**	0.9501	0.0073	0.0042	**262.76**	0.9525	0.0146	0.0176
	5000	fast MFPCA (visit)	**4.60**	0.9506	0.0034	0.0043	**5.31**	0.9516	0.0037	0.0046
		fast MFPCA (subject)	**4.84**	0.9506	0.0034	0.0043	**4.63**	0.9489	0.0064	0.0048
		MFPCA	**540.84**	0.9503	0.0019	0.0008	**943.56**	0.9514	0.0036	0.0033
Incomplete	100	fast MFPCA (visit)	**2.16**	0.9003	0.0942	0.0348	**2.03**	0.9010	0.1570	0.0461
		fast MFPCA (subject)	**1.46**	0.8952	0.0891	0.0348	**1.51**	0.8909	0.1094	0.0513
		MFPCA	**3.58**	0.8921	0.0999	0.0614	**5.00**	0.9089	0.2496	0.1644
	200	fast MFPCA (visit)	**2.34**	0.9028	0.0554	0.0198	**2.46**	0.9056	0.0671	0.0278
		fast MFPCA (subject)	**1.50**	0.8987	0.0465	0.0198	**1.62**	0.8974	0.0455	0.0252
		MFPCA	**4.80**	0.8950	0.0489	0.0303	**8.34**	0.9045	0.0917	0.0857
	1000	fast MFPCA (visit)	**5.48**	0.9087	0.0230	0.0081	**5.83**	0.9074	0.0246	0.0074
		fast MFPCA (subject)	**3.88**	0.9042	0.0131	0.0081	**4.07**	0.8998	0.0130	0.0075
		MFPCA	**13.19**	0.9037	0.0102	0.0064	**29.46**	0.9055	0.0179	0.0194
	5000	fast MFPCA (visit)	**19.99**	0.9077	0.0147	0.0048	**22.21**	0.9086	0.0150	0.0051
		fast MFPCA (subject)	**20.93**	0.9037	0.0064	0.0048	**17.25**	0.9042	0.0061	0.0050
		MFPCA	**58.19**	0.9041	0.0022	0.0012	**134.49**	0.9066	0.0043	0.0037

Table S3: Simulation results under different J when $I = 100$ and $L = 100$. The computation time ("Time(s)"), MISE for $\mathbf{Y}$ ("MISE($\mathbf{Y}$)") and eigenfunctions ("MISE($\boldsymbol{\phi}$)", "MISE($\boldsymbol{\psi}$)") reported in the table are median values across 100 simulations. Computation time more than 24 hours is denoted as ∞.

Design	J	Model	Balanced				Unbalanced			
			Time(s)	MISE($\mathbf{Y}$)	MISE($\boldsymbol{\phi}$)	MISE($\boldsymbol{\psi}$)	Time(s)	MISE($\mathbf{Y}$)	MISE($\boldsymbol{\phi}$)	MISE($\boldsymbol{\psi}$)
Complete	2	fast MFPCA (visit)	**1.81**	0.9450	0.0781	0.0319	**1.74**	0.9450	0.1203	0.0416
		fast MFPCA (subject)	**1.16**	0.9450	0.0781	0.0319	**1.26**	0.9446	0.1174	0.0486
		MFPCA	**14.82**	0.9451	0.0797	0.0315	**21.51**	0.9571	0.1930	0.1286
	4	fast MFPCA (visit)	**1.98**	0.9532	0.0547	0.0126	**1.80**	0.9528	0.0634	0.0171
		fast MFPCA (subject)	**1.35**	0.9532	0.0547	0.0126	**1.19**	0.9542	0.0678	0.0189
		MFPCA	**46.93**	0.9537	0.0483	0.0099	**75.87**	0.9597	0.0800	0.0463
	20	fast MFPCA (visit)	**2.33**	0.9614	0.0364	0.0056	**2.66**	0.9618	0.0314	0.0054
		fast MFPCA (subject)	**2.14**	0.9614	0.0364	0.0056	**2.36**	0.9622	0.0301	0.0056
		MFPCA	**10146.82**	0.9617	0.0317	0.0019	**10334.10**	0.9626	0.0346	0.0088
	100	fast MFPCA (visit)	**6.02**	0.9626	0.0335	0.0042	**7.15**	0.9632	0.0332	0.0043
		fast MFPCA (subject)	**4.79**	0.9626	0.0335	0.0042	**5.17**	0.9632	0.0330	0.0043
		MFPCA	∞	-	-	-	∞	-	-	-
Incomplete	2	fast MFPCA (visit)	**2.16**	0.9003	0.0942	0.0348	**2.03**	0.9010	0.1570	0.0461
		fast MFPCA (subject)	**1.46**	0.8952	0.0891	0.0348	**1.51**	0.8909	0.1094	0.0513
		MFPCA	**3.58**	0.8921	0.0999	0.0614	**5.00**	0.9089	0.2496	0.1644
	4	fast MFPCA (visit)	**2.48**	0.9094	0.0616	0.0142	**2.48**	0.9100	0.0784	0.0200
		fast MFPCA (subject)	**1.94**	0.9059	0.0593	0.0141	**1.92**	0.9055	0.0780	0.0218
		MFPCA	**11.07**	0.9147	0.0615	0.0161	**17.14**	0.9173	0.0842	0.0544
	20	fast MFPCA (visit)	**5.88**	0.9158	0.0432	0.0060	**6.08**	0.9162	0.0393	0.0067
		fast MFPCA (subject)	**5.56**	0.9132	0.0380	0.0060	**5.92**	0.9136	0.0301	0.0064
		MFPCA	**733.01**	0.9254	0.0325	0.0037	**903.30**	0.9262	0.0399	0.0099
	100	fast MFPCA (visit)	**21.44**	0.9159	0.0384	0.0044	**23.72**	0.9160	0.0375	0.0046
		fast MFPCA (subject)	**17.40**	0.9137	0.0330	0.0044	**20.81**	0.9136	0.0343	0.0045
		MFPCA	∞	-	-	-	∞	-	-	-

Table S4: Simulation results under different L when $I = 100$ and $J = 2$. The computation time ("Time(s)"), MISE for $\mathbf{Y}$ ("MISE($\mathbf{Y}$)") and eigenfunctions ("MISE($\boldsymbol{\phi}$)", "MISE($\boldsymbol{\psi}$)") reported in the table are median values across 100 simulations. Computation time more than 24 hours is denoted as ∞.

Design	L	Model	Balanced				Unbalanced			
			Time(s)	MISE($\mathbf{Y}$)	MISE($\boldsymbol{\phi}$)	MISE($\boldsymbol{\psi}$)	Time(s)	MISE($\mathbf{Y}$)	MISE($\boldsymbol{\phi}$)	MISE($\boldsymbol{\psi}$)
Complete	100	fast MFPCA (visit)	**1.81**	0.9450	0.0781	0.0319	**1.74**	0.9450	0.1203	0.0416
		fast MFPCA (subject)	**1.16**	0.9450	0.0781	0.0319	**1.26**	0.9446	0.1174	0.0486
		MFPCA	**14.82**	0.9451	0.0797	0.0315	**21.51**	0.9571	0.1930	0.1286
	200	fast MFPCA (visit)	**1.67**	0.9722	0.0804	0.0277	**1.78**	0.9746	0.1193	0.0395
		fast MFPCA (subject)	**1.38**	0.9722	0.0804	0.0277	**1.31**	0.9752	0.1187	0.0446
		MFPCA	**119.92**	0.9733	0.0785	0.0272	**334.66**	0.9830	0.1933	0.1398
	1000	fast MFPCA (visit)	**2.51**	0.9953	0.0758	0.0244	**3.06**	0.9976	0.1145	0.0368
		fast MFPCA (subject)	**1.45**	0.9953	0.0758	0.0244	**1.77**	0.9984	0.1094	0.0416
		MFPCA	**16883.52**	0.9963	0.0784	0.0237	**24395.62**	1.0055	0.1794	0.1058
	5000	fast MFPCA (visit)	**7.79**	0.9990	0.0756	0.0246	**12.40**	1.0023	0.1103	0.0369
		fast MFPCA (subject)	**5.18**	0.9990	0.0756	0.0246	**9.68**	1.0025	0.1086	0.0406
		MFPCA	∞	-	-	-	∞	-	-	-
Incomplete	100	fast MFPCA (visit)	**2.16**	0.9003	0.0942	0.0348	**2.03**	0.9010	0.1570	0.0461
		fast MFPCA (subject)	**1.46**	0.8952	0.0891	0.0348	**1.51**	0.8909	0.1094	0.0513
		MFPCA	**3.58**	0.8921	0.0999	0.0614	**5.00**	0.9089	0.2496	0.1644
	200	fast MFPCA (visit)	**2.52**	0.9515	0.0828	0.0291	**2.41**	0.9498	0.1308	0.0408
		fast MFPCA (subject)	**1.53**	0.9488	0.0822	0.0291	**1.59**	0.9480	0.1161	0.0466
		MFPCA	**17.01**	0.9484	0.0914	0.0368	**26.18**	0.9579	0.2138	0.1426
	1000	fast MFPCA (visit)	**3.35**	0.9923	0.0811	0.0263	**3.86**	0.9949	0.1160	0.0363
		fast MFPCA (subject)	**2.47**	0.9919	0.0796	0.0263	**2.31**	0.9946	0.1119	0.0407
		MFPCA	**1883.75**	0.9903	0.0788	0.0274	**2697.17**	1.0023	0.1889	0.1425
	5000	fast MFPCA (visit)	**10.09**	0.9985	0.0773	0.0247	**14.56**	1.0022	0.1147	0.0374
		fast MFPCA (subject)	**9.56**	0.9985	0.0760	0.0247	**10.61**	1.0025	0.1081	0.0404
		MFPCA	∞	-	-	-	∞	-	-	-

References

Eilers, P. and Marx, B. (1996). Flexible smoothing with b-splines and penalties (with Discussion). *Statistical Science*, 11:89–121.

Xiao, L. (2019). Asymptotic theory of penalized splines. *Electronic Journal of Statistics*, 13(1):747–794.

Xiao, L. (2020). Asymptotic properties of penalized splines for functional data. *Bernoulli*, 26(4):2847–2875.

Xiao, L., Li, Y., and Ruppert, D. (2013). Fast bivariate p-splines: the sandwich smoother. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(3):577–599.

Xiao, L. and Nan, Z. (2020). Uniform convergence of penalized splines. *Stat*, 9(1):e297.

Xiao, L., Zipunnikov, V., Ruppert, D., and Crainiceanu, C. (2016). Fast covariance estimation for high-dimensional functional data. *Statistics and Computing*, 26(1):409–421.