SPARSE FUNCTIONAL PRINCIPAL COMPONENT ANALYSIS IN HIGH DIMENSIONS

Xiaoyu Hu and Fang Yao

Peking University

Supplementary Material

S1 Proofs of main results

Proof of Proposition 1. We provide the proof of the generalized version of Proposition 1 which allows different basis functions among processes. The univariate orthonormal basis representation for each random process is $X_j = \sum_{l=1}^{\infty} \theta_{jl} b_{jl}$. Recall that $G(s,t) = E\{\boldsymbol{X}(s)\boldsymbol{X}(t)^{\mathrm{T}}\} \in \mathbb{R}^{p\times p}$ and $\int G(s,t)\boldsymbol{\psi}_k(s)ds = \lambda_k \boldsymbol{\psi}_k(t)$. Then we have

$$\begin{aligned}
\left\{\int G(s,t)\boldsymbol{\psi}_k(s)ds\right\}_j &= \sum_{j'=1}^{p}\int \mathrm{cov}\{X_{j'}(s), X_j(t)\}\psi_{kj'}(s)ds \\
&= \sum_{j'=1}^{p}\int \sum_{l'=1}^{\infty}\sum_{m=1}^{\infty} \mathrm{cov}(\theta_{j'l'}, \theta_{jm}) b_{j'l'}(s) b_{jm}(t)\psi_{kj'}(s)ds \\
&= \sum_{j'=1}^{p}\sum_{l'=1}^{\infty}\sum_{m=1}^{\infty} \mathrm{cov}(\theta_{j'l'}, \theta_{jm}) b_{jm}(t)\int b_{j'l'}(s)\psi_{kj'}(s)ds \\
&= \lambda_k \psi_{kj}(t). \qquad \text{(S1.1)}
\end{aligned}$$

Denote $u_{kjl} = \int_{\mathcal{T}} b_{jl}(t)\psi_{kj}(t)dt$. Multiplying both sides by $b_{jl}(t)$ and then integrating both sides over t yields

$$\sum_{j'=1}^{p}\sum_{l'=1}^{\infty}\sum_{m=1}^{\infty}\text{cov}(\theta_{j'l'},\theta_{jm})\int b_{jm}(t)b_{jl}(t)dt\int b_{j'l'}(s)\psi_{kj'}(s)ds = \lambda_k\int b_{jl}(t)\psi_{kj}(t)dt,$$
$$\sum_{j'=1}^{p}\sum_{l'=1}^{\infty}\text{cov}(\theta_{j'l'},\theta_{jl})u_{kj'l'} = \lambda_k u_{kjl}. \qquad \text{(S1.2)}$$

Combining (S1.1) and (S1.2), the eigenfunctions $\boldsymbol{\psi}_k$ are

$$\psi_{kj}(t) = \sum_{l=1}^{\infty} u_{kjl}b_{jl}(t), \quad t\in\mathcal{T}, j=1,\ldots,p, k=1,2,\ldots.$$

Obviously, $\sum_{j=1}^{p}\sum_{l=1}^{\infty}u_{kjl}^2 = 1$ and $\sum_{j=1}^{p}\sum_{l=1}^{\infty}u_{kjl}u_{k'jl} = 0$ for $k\neq k'$. And the scores are

$$\begin{aligned}\eta_k &= \ <\boldsymbol{X}(t),\boldsymbol{\psi}_k(t)> &&= \textstyle\sum_{j=1}^{p}\int X_j(t)\psi_{kj}(t)dt\\ &= \textstyle\sum_{j=1}^{p}\int\sum_{l=1}^{\infty}\theta_{jl}b_{jl}(t)\psi_{kj}(t)dt &&= \textstyle\sum_{j=1}^{p}\sum_{l=1}^{\infty}\theta_{jl}u_{kjl}.\end{aligned}$$

□

For convenience, we suppress the subscript $\mathbb{H}$ in inner product and norm operations when there is no ambiguity.

Proof of Theorem 1. Recall that $\tilde{X}_j = \sum_{l=1}^{N_j}\theta_{jl}b_l$ and $\tilde{G}(s,t) = E\{\tilde{\boldsymbol{X}}(s)\tilde{\boldsymbol{X}}(t)^{\mathrm{T}}\}$, $\tilde{\lambda}_k$ and $\tilde{\boldsymbol{\psi}}_k$ are corresponding eigenvalues and eigenfunctions respectively. First we provide the bound of $|||\tilde{G}-G|||^2$ which is important in the sequel.

$$|||\tilde{G}-G|||^2 \le \int\int\sum_{j=1}^{p}\sum_{j'=1}^{p}\{\tilde{G}_{jj'}(s,t)-G_{jj'}(s,t)\}^2dsdt$$

$$
\begin{aligned}
&= \int\int \sum_{j=1}^{p}\sum_{j'=1}^{p}\{E\tilde{X}_j(s)\tilde{X}_{j'}(t) - EX_j(s)X_{j'}(t)\}^2 dsdt \\
&\lesssim \sum_{j=1}^{p}\sum_{j'=1}^{p}\Big(\int\int [E\tilde{X}_j(s)\{\tilde{X}_{j'}(t) - X_{j'}(t)\}]^2 dsdt \\
&\quad + \int\int [EX_{j'}(t)\{\tilde{X}_j(s) - X_j(s)\}]^2 dsdt\Big) \\
&\leq E\|\tilde{\boldsymbol{X}}\|^2 E\|\tilde{\boldsymbol{X}} - \boldsymbol{X}\|^2 + E\|\boldsymbol{X}\|^2 E\|\tilde{\boldsymbol{X}} - \boldsymbol{X}\|^2. \qquad \text{(S1.3)}
\end{aligned}
$$

Use the notations I_n^- and I_n^+ defined in Lemma S1. Denote the event $\{I_n^- \subset \hat{I} \subset I_n^+\}$ by A_n. By Lemma S1, we have $P(\limsup A_n) = 1$. Under the weak l_q sparsity, $E\|\boldsymbol{X}\|^2 = O(1)$. On the event A_n, we have

$$
\begin{aligned}
E\|\tilde{\boldsymbol{X}} - \boldsymbol{X}\|^2 &\leq \sum_{(j,l)\notin I_n^-} \sigma_{jl}^2 \leq \sum_{j,l}(m^{-1}\sigma^2 a_+ \alpha_n) \wedge \sigma_{(j)(l)}^2 \\
&\asymp \sum_{j=1}^{g_n}\sum_{l=N_j+1}^{\infty} \sigma_{(j)(l)}^2 + \sum_{j=g_n+1}^{p} V_{(j)} = I + II.
\end{aligned}
$$

It is obtained that $II = O(g_n^{1-2/q})$. Based on the weak l_q sparsity and Lemma 1, we have

$$
I \asymp \sum_{j=1}^{g_n} j^{-\frac{2}{q(2\alpha+1)}} \left(\frac{1}{m}\sqrt{\frac{\log p}{n}}\right)^{2\alpha/(2\alpha+1)}.
$$

Next we consider the following cases about the first term I based on relationship between two types of sparsity q and α to obtain the final results.

- If $q(2\alpha+1) > 2$, then $I = O\left[\left\{m^{-1}(\log p/n)^{1/2}\right\}^{2\alpha/(2\alpha+1)} g_n^{1-2/q(2\alpha+1)}\right] = O(g_n^{1-2/q})$. Combining I and II yields $E\|\tilde{\boldsymbol{X}} - \boldsymbol{X}\|^2 = O(g_n^{1-2/q})$.

- If $q(2\alpha+1)=2$, then

$$I=\{m^{-1}(\log p/n)^{1/2}\}^{2\alpha/(2\alpha+1)}\sum_{j=1}^{g_n} j^{-1}=O\left[\{m^{-1}(\log p/n)^{1/2}\}^{2\alpha/(2\alpha+1)}\log(g_n)\right].$$

  Combining I and II yields $E\|\tilde{\boldsymbol{X}}-\boldsymbol{X}\|^2=O\left[\{m^{-1}(\log p/n)^{1/2}\}^{2\alpha/(2\alpha+1)}\log(g_n)\right]$.

- If $q(2\alpha+1)<2$, then we have $I=O\left[\{m^{-1}(\log p/n)^{1/2}\}^{2\alpha/(2\alpha+1)}\right]$.

  Combining I and II yields $E\|\tilde{\boldsymbol{X}}-\boldsymbol{X}\|^2=O\left[\{m^{-1}(\log p/n)^{1/2}\}^{2\alpha/(2\alpha+1)}\right]$.

From the bound on covariance (S1.3), according to the result of Theorem 1 in Hall and Hosseini-Nasab (2006), we arrive at the desired results. □

Proof of Theorem 2. Recall that g_n denotes the number of retained processes. First, we prove that the measurement error is negligible and then it suffices to quantify the error $|||\tilde{G}-\hat{G}|||$ on the event A_n, where $\hat{G}_{jj'}(s,t)=n^{-1}\sum_{i=1}^n \check{x}_{ij}(s)\check{x}_{ij'}(t)$ and $\check{x}_{ij}=\sum_{l=1}^{N_j}\tilde{\theta}_{ijl}b_l$. Observe that

$$\tilde{\epsilon}_{ijl}=\sum_{k=1}^m \epsilon_{ijk}\int_{t_{k-1}}^{t_k} b_l(t)dt.$$

Then we have $\mathrm{var}(\tilde{\epsilon}_{ijl})=\sigma^2\sum_{k=1}^m\left\{\int_{t_{k-1}}^{t_k} b_l(t)dt\right\}^2$.

Denote $\tilde{\Delta}=\mathrm{diag}(\mathrm{var}(\tilde{\epsilon}_{11}),\ldots,\mathrm{var}(\tilde{\epsilon}_{1N_1}),\ldots,\mathrm{var}(\tilde{\epsilon}_{p1}),\ldots,\mathrm{var}(\tilde{\epsilon}_{pN_p}))$ and Δ is a $N\times N$ diagonal matrix whose elements are all $1/m$ where $N=\sum_{j=1}^p N_j$. Note that

$$\left|\sum_{k=1}^m\left\{\int_{t_{k-1}}^{t_k} b_l(t)dt\right\}^2-\frac{1}{m}\right|$$

$$
\begin{aligned}
&= \left|\sum_{k=1}^{m}\int_{t_{k-1}}^{t_k}\int_{t_{k-1}}^{t_k} b_l(t)\{b_l(s)-b_l(t)\}dtds\right| \\
&\le \frac{L}{m^2}\int_0^1 b_l(t)dt,
\end{aligned}
$$

where the last inequality follows from the Condition 8. Note that from Lemma 1, we have $N = o_p(g_n m^2)$ under Condition 5. Then we have $\|\tilde{\Delta} - \Delta\|_F = o_p(g_n^{1/2}m^{-1})$.

On the event A_n,

$$
\begin{aligned}
|||\tilde{G}-\hat{G}|||^2 &= \int\int\sum_{j=1}^{p}\sum_{j'=1}^{p}\{\tilde{G}_{jj'}(s,t)-\hat{G}_{jj'}(s,t)\}^2 dsdt \\
&= \int\int\sum_{j=1}^{p}\sum_{j'=1}^{p}\left\{n^{-1}\sum_{i=1}^{n}\check{x}_{ij}(s)\check{x}_{ij'}(t) - n^{-1}\sum_{i=1}^{n}\tilde{x}_{ij}(s)\tilde{x}_{ij'}(t)\right. \\
&\quad \left. + n^{-1}\sum_{i=1}^{n}\tilde{x}_{ij}(s)\tilde{x}_{ij'}(t) - E\tilde{X}_j(s)\tilde{X}_{j'}(t)\right\}^2 dsdt \\
&\le 4\sum_{j=1}^{g_n}\sum_{j'=1}^{g_n}\int\int\left[n^{-1}\sum_{i=1}^{n}\{\check{x}_{ij}(s)-\tilde{x}_{ij}(s)\}\check{x}_{ij'}(t)\right]^2 dsdt \\
&\quad +4\sum_{j=1}^{g_n}\sum_{j'=1}^{g_n}\int\int\left[n^{-1}\sum_{i=1}^{n}\tilde{x}_{ij}(s)\{\tilde{x}_{ij'}(t)-\check{x}_{ij'}(t)\}\right]^2 dsdt \\
&\quad +2\int\int\sum_{j=1}^{g_n}\sum_{j'=1}^{g_n}\left\{n^{-1}\sum_{i=1}^{n}\tilde{x}_{ij}(s)\tilde{x}_{ij'}(t) - E\tilde{X}_j(s)\tilde{X}_{j'}(t)\right\}^2 dsdt \\
&= I + II + III. \qquad \text{(S1.4)}
\end{aligned}
$$

To bound the term I and II, note that

$$
\int\int\left[n^{-1}\sum_{i=1}^{n}\{\check{x}_{ij}(s)-\tilde{x}_{ij}(s)\}\check{x}_{ij'}(t)\right]^2 dsdt
$$

$$\leq n^{-2}\left(\sum_{i=1}^{n}\left[\int\int\{\check{x}_{ij}(s)-\tilde{x}_{ij}(s)\}^2\check{x}_{ij'}^2(t)dsdt\right]^{1/2}\right)^2$$

$$= n^{-2}\left(\sum_{i=1}^{n}\left[\|\tilde{x}_{ij}-\check{x}_{ij}\|_{L^2}\left\{\int\check{x}_{ij'}(t)^2dt\right\}^{1/2}\right]\right)^2, \qquad \text{(S1.5)}$$

where the first inequality follows from the triangle inequality. Similarly, we have

$$\int\int\left[n^{-1}\sum_{i=1}^{n}\tilde{x}_{ij}(s)\{\tilde{x}_{ij'}(t)-\check{x}_{ij'}(t)\}\right]^2dsdt$$

$$\leq n^{-2}\left(\sum_{i=1}^{n}\left[\|\tilde{x}_{ij'}-\check{x}_{ij'}\|_{L^2}\left\{\int\tilde{x}_{ij}^2(s)ds\right\}^{1/2}\right]\right)^2. \qquad \text{(S1.6)}$$

Using Bessel's inequality and Condition 2, we may prove that

$$\|\tilde{x}_{ij}-\check{x}_{ij}\|_{L^2}\leq\|x_{ij}^*-x_{ij}\|_{L^2}=O_p\left(\frac{1}{m}\right). \qquad \text{(S1.7)}$$

So we have $I=O_p(g_n/m^2)$ and $II=O_p(g_n/m^2)$. To bound the term III,

$$n^{-2}\int\int E\left\{\sum_{i=1}^{n}\{\tilde{x}_{ij}(s)\tilde{x}_{ij'}(t)-E\tilde{X}_j(s)\tilde{X}_{j'}(t)\}\right\}^2dsdt$$

$$\leq n^{-2}\int\int\sum_{i=1}^{n}E\{\tilde{x}_{ij}^2(s)\tilde{x}_{ij'}^2(t)\}dsdt$$

$$= O_p\left(\frac{1}{n}\right),$$

where the last equality follows from Condition 2. Thus, combining together yields that $|||\hat{G}-\tilde{G}|||=O_p(n^{-1/2}+g_n^{1/2}m^{-1})$.

Case 1. If $\gamma>1/(2-q)$, the parametric rate dominates while the discretization error is negligible, $|||\tilde{G}-\hat{G}|||=O_p(n^{-1/2})$. In this case, we adopt techniques in Hall and Horowitz (2007) and Kong et al. (2016) to obtain sharper bounds.

Define $\hat{\Delta} = |||\hat{G} - \tilde{G}|||$. We find that, for $k = 1, \ldots, r_n$,

$$\tilde{\lambda}_k - \tilde{\lambda}_{k+1} \geq |\lambda_k - \lambda_{k+1} - 2\Delta| \geq Ck^{-a-1},$$

which holds according to Condition 9, where $\Delta = |||G - \tilde{G}|||$. Denote

$$\mathcal{J}_n = \{\tilde{\lambda}_k - \tilde{\lambda}_{k+1} > 2/(2 - \sqrt{2})\hat{\Delta}, k = 1, \ldots, r_n\}.$$

The set $\mathcal{J}_n$ means that the distance of adjacent ordered eigenvalues does not fall below $2/(2 - \sqrt{2})\hat{\Delta}$, $P(\mathcal{J}_n) \to 1, n \to \infty$ is implied by Condition 10. For some constant C, define the set

$$\mathcal{F}_n = \{(\hat{\lambda}_{k_1} - \hat{\lambda}_{k_2})^{-2} \leq 2(\tilde{\lambda}_{k_1} - \tilde{\lambda}_{k_2})^{-2} \leq Cr_n^{2(a+1)}, k_1, k_2 = 1, \ldots, r_n, k_1 \neq k_2\}.$$

For $k_1 \neq k_2$, $|\hat{\lambda}_{k_1} - \tilde{\lambda}_{k_1}| \leq \hat{\Delta} < (1 - \sqrt{2}/2) \min_{k_1 \neq k_2} |\tilde{\lambda}_{k_1} - \tilde{\lambda}_{k_2}|$ gives that

$$\begin{aligned} |\hat{\lambda}_{k_1} - \tilde{\lambda}_{k_2}| &= |\hat{\lambda}_{k_1} - \tilde{\lambda}_{k_1} + \tilde{\lambda}_{k_1} - \tilde{\lambda}_{k_2}| \\ &\geq |\tilde{\lambda}_{k_1} - \tilde{\lambda}_{k_2}| - |\hat{\lambda}_{k_1} - \tilde{\lambda}_{k_1}| \\ &\geq |\tilde{\lambda}_{k_1} - \tilde{\lambda}_{k_2}| - \hat{\Delta}. \end{aligned}$$

Then we have $P(\mathcal{F}_n) \to 1$ as $n \to \infty$. By (5.16) in Hall and Horowitz (2007), one has $\|\hat{\boldsymbol{\psi}}_k - \tilde{\boldsymbol{\psi}}_k\|^2 \leq 2\hat{u}_k^2$ where $\hat{u}_k^2 = \sum_{l:l\neq k}(\hat{\lambda}_k - \tilde{\lambda}_l)^{-2}\{\int \hat{\boldsymbol{\psi}}_k^{\mathrm{T}}(\hat{G} - \tilde{G})\tilde{\boldsymbol{\psi}}_l\}^2$.

By Lemma 1 in Kong et al. (2016), we have

$$\hat{u}_k^2 \le 4 \sum_{l:l\neq k} (\tilde{\lambda}_k - \tilde{\lambda}_l)^{-2} \left\{ \int \tilde{\boldsymbol{\psi}}_k^{\mathrm{T}} (\tilde{G} - \hat{G}) \tilde{\boldsymbol{\psi}}_l \right\}^2 + 2C r_n^{2(a+1)} \|\hat{\boldsymbol{\psi}}_k - \tilde{\boldsymbol{\psi}}_k\|^2 \hat{\Delta}^2.$$

Plugging this into $\|\hat{\boldsymbol{\psi}}_k - \tilde{\boldsymbol{\psi}}_k\|^2 \le 2\hat{u}_k^2$, we find that

$$(1 - 4C r_n^{2(a+1)} \hat{\Delta}^2) \|\hat{\boldsymbol{\psi}}_k - \tilde{\boldsymbol{\psi}}_k\|^2 \le 8 \sum_{l:l\neq k} (\tilde{\lambda}_k - \tilde{\lambda}_l)^{-2} \left\{ \int \tilde{\boldsymbol{\psi}}_k^{\mathrm{T}} (\tilde{G} - \hat{G}) \tilde{\boldsymbol{\psi}}_l \right\}^2.$$

As $r_n^{2(a+1)} \hat{\Delta}^2 = o_p(1)$, we have

$$\|\hat{\boldsymbol{\psi}}_k - \tilde{\boldsymbol{\psi}}_k\|^2 \le 8 \sum_{l:l\neq k} (\tilde{\lambda}_k - \tilde{\lambda}_l)^{-2} \left\{ \int \tilde{\boldsymbol{\psi}}_k^{\mathrm{T}} (\tilde{G} - \hat{G}) \tilde{\boldsymbol{\psi}}_l \right\}^2,$$

by analogy to (5.22) in Hall and Horowitz (2007), $E\left[\sum_{l:l\neq k} (\tilde{\lambda}_k - \tilde{\lambda}_l)^{-2} \left\{ \int \tilde{\boldsymbol{\psi}}_k^{\mathrm{T}} (\tilde{G} - \hat{G}) \tilde{\boldsymbol{\psi}}_l \right\}^2\right] = O(k^2 n^{-1})$ holds uniformly in $k = 1, \ldots, r_n$.

Case 2. If $\gamma \le 1/(2-q)$, the discretization error dominates, $|||\hat{G} - \tilde{G}||| = O_p(g_n^{1/2} m^{-1})$. With the result of Theorem 1 in Hall and Hosseini-Nasab (2006), the final results are established. □

S2 Theoretical results on recovery

We can represent the trajectories using estimated eigenfunctions. It is of interest to investigate the theoretical performance of recovered processes. To provide more insights of the sampling frequency of m on the results, we directly characterize the discretization error. For recovered curves, one has the following decomposition:

$$\|\hat{\boldsymbol{x}}_i^{r_n} - \boldsymbol{x}_i\|_{\mathbb{H}} \leq \|\boldsymbol{x}_i^{r_n} - \boldsymbol{x}_i\|_{\mathbb{H}} + \|\tilde{\boldsymbol{x}}_i^{r_n} - \boldsymbol{x}_i^{r_n}\|_{\mathbb{H}} + \|\hat{\boldsymbol{x}}_i^{r_n} - \tilde{\boldsymbol{x}}_i^{r_n}\|_{\mathbb{H}},$$

where $\hat{\boldsymbol{x}}_i^{r_n} = \sum_{k=1}^{r_n} \hat{\eta}_{ik}\hat{\boldsymbol{\psi}}_k$, $\tilde{\boldsymbol{x}}_i^{r_n} = \sum_{k=1}^{r_n} \tilde{\eta}_{ik}\tilde{\boldsymbol{\psi}}_k$ and $\boldsymbol{x}_i^{r_n} = \sum_{k=1}^{r_n} \eta_{ik}\boldsymbol{\psi}_k$. In the righthand, the first and second terms can be both viewed as approximation errors, while the third term is seen as an estimation error. Denote $\tilde{\eta}_{ik} =< \tilde{\boldsymbol{x}}_i, \tilde{\boldsymbol{\psi}}_k >_{\mathbb{H}}$. Under the weak l_q sparsity, we consider the most interesting case where $0 < q < 1$ (Bruckstein et al., 2009). For more detailed interpretation, one can refer to the discussion following Theorems 1 and 2.

Theorem S1 (Approximation Error for recovery under weak l_q)**.** *Under the Conditions in Theorem 1, if* $< \boldsymbol{\psi}_k, \tilde{\boldsymbol{\psi}}_k >_{\mathbb{H}} \geq 0$*, then uniformly for* $k = 1, \ldots, r_n$*,*

Case 1. When $q(\alpha+1) > 2$*,*

$$\begin{aligned} |\tilde{\eta}_{ik} - \eta_{ik}| &= O_p\left(k^{a+1} g_n^{1/2-1/q}\right), \\ \|\tilde{\boldsymbol{x}}_i^{r_n} - \boldsymbol{x}_i^{r_n}\|_{\mathbb{H}} &= O_p\left(r_n^{a+3/2} g_n^{1/2-1/q}\right). \end{aligned}$$

Case 2. When $q(\alpha+1) = 2$*,*

$$\begin{aligned} |\tilde{\eta}_{ik} - \eta_{ik}| &= O_p\left\{k^{a+1}(m^{-1}\sqrt{\log p/n})^{\alpha/(2\alpha+1)}(\log g_n)^{1/2}\right\}, \\ \|\tilde{\boldsymbol{x}}_i^{r_n} - \boldsymbol{x}_i^{r_n}\|_{\mathbb{H}} &= O_p\left\{r_n^{a+3/2}(m^{-1}\sqrt{\log p/n})^{\alpha/(2\alpha+1)}(\log g_n)^{1/2}\right\}. \end{aligned}$$

Case 3. When $q(\alpha+1)<2$,

$$\begin{aligned}|\tilde{\eta}_{ik}-\eta_{ik}| &= O_p\left\{k^{a+1}(m^{-1}\sqrt{\log p/n})^{\alpha/(2\alpha+1)}\right\},\\ \|\tilde{\boldsymbol{x}}_i^{r_n}-\boldsymbol{x}_i^{r_n}\|_{\mathbb{H}} &= O_p\left\{r_n^{a+3/2}(m^{-1}\sqrt{\log p/n})^{\alpha/(2\alpha+1)}\right\},\end{aligned}$$

Moreover,

$$\|\boldsymbol{x}_i-\boldsymbol{x}_i^{r_n}\|_{\mathbb{H}}=O_p\left(r_n^{1-a}\right).$$

Theorem S2 (Estimation Error for recovery under weak l_q)**.** *Under Conditions in Theorem 2, if* $<\hat{\boldsymbol{\psi}}_k,\tilde{\boldsymbol{\psi}}_k>_{\mathbb{H}}\geq 0$, *then uniformly for* $k=1,\ldots,r_n$,

Case 1. When $\gamma>1/(2-q)$,

$$|\tilde{\eta}_{ik}-\hat{\eta}_{ik}|=O_p(kn^{-1/2}+k^{a/2}m^{-1/2}),$$

$$\|\tilde{\boldsymbol{x}}_i^{r_n}-\hat{\boldsymbol{x}}_i^{r_n}\|_{\mathbb{H}}=O_p(r_n^{3/2}n^{-1/2}+r_n^{(a+1)/2}m^{-1/2}),\ i=1,\ldots,n.$$

Case 2. When $(1-\beta)/2<\gamma\leq 1/(2-q)$,

$$|\tilde{\eta}_{ik}-\hat{\eta}_{ik}|=O_p(k^{a+1}g_n^{1/2}m^{-1}+k^{a/2}m^{-1/2}),$$

$$\|\tilde{\boldsymbol{x}}_i^{r_n}-\hat{\boldsymbol{x}}_i^{r_n}\|_{\mathbb{H}}=O_p(r_n^{a+3/2}g_n^{1/2}m^{-1}+r_n^{(a+1)/2}m^{-1/2}),\ i=1,\ldots,n.$$

S3 Proofs of lemmas and auxiliary results

Define two non-random sets $I_n^-=\{(j,l),j=1,\ldots,p;l=1,\ldots,s_n:\sigma_{jl}^2>m^{-1}\sigma^2a_+\alpha_n\}$ and $I_n^+=\{(j,l),j=1,\ldots,p;l=1,\ldots,s_n:\sigma_{jl}^2>$

$m^{-1}\sigma^2 a_- \alpha_n\}$. Recall that $\hat{I} = \{(j,l), j = 1,\ldots,p; l = 1,\ldots,s_n : \hat{\sigma}^2_{jl} \geq m^{-1}\sigma^2(1+\alpha_n)\}$.

Lemma S1. *For sufficiently large n, $I_n^- \subset \hat{I} \subset I_n^+$ almost surely.*

Proof. Recall that $\tilde{\sigma}^2_{jl} = \mathrm{var}(\tilde{\theta}_{jl})$. Observe that

$$\begin{aligned}\hat{\theta}_{ijl} &= \int_{\mathcal{T}} x_{ij}(t) b_l(t) dt + \sum_{k=1}^{m} \int_{t_{k-1}}^{t_k} \{x_{ij}(t_k) - x_{ij}(t)\} b_l(t) dt + \tilde{\epsilon}_{ijl} \\ &= \theta_{ijl} + z_{ijl} + \tilde{\epsilon}_{ijl}.\end{aligned}$$

We have $\tilde{\theta}_{ijl} = \theta_{ijl} + z_{ijl}$ and $\tilde{\sigma}^2_{jl} = \sigma^2_{jl} + \mathrm{var}(z_{ijl}) + \mathrm{cov}(\theta_{ijl}, z_{ijl})$. Under the Lipschitz condition, we have $\mathrm{var}(z_{ijl}) = O(m^{-2})$. So $\tilde{\sigma}^2_{jl} = \sigma^2_{jl} + O(m^{-2}) + O(\sigma_{jl}/m)$. First we state the results from Johnstone (2001) that

$$\begin{aligned}\mathrm{pr}\{\chi^2_n \leq n(1-\epsilon)\} &\leq \exp(-n\epsilon^2/4), \quad 0 \leq \epsilon \leq 1, \\ \mathrm{pr}\{\chi^2_n \geq n(1+\epsilon)\} &\leq \exp(-3n\epsilon^2/16), \quad 0 \leq \epsilon < 1/2.\end{aligned}$$

Denote $\bar{M}_n \overset{d}{=} \chi^2_n/n$ where $x \overset{d}{=} y$ means that x has the same distribution as y. $|I|$ denotes the cardinality of set I. Then,

$$\begin{aligned}P_n^- &= \mathrm{pr}(I_n^- \notin \hat{I}) \\ &= \mathrm{pr}\left[\{(j,l) \in I_n^- : \hat{\sigma}^2_{jl} \leq m^{-1}\sigma^2(1+\alpha_n)\}\right] \\ &\leq \textstyle\sum_{(j,l)\in I_n^-} \mathrm{pr}\{\hat{\sigma}^2_{jl} \leq m^{-1}\sigma^2(1+\alpha_n)\}, \qquad \text{subadditivity} \\ &= \textstyle\sum_{(j,l)\in I_n^-} \mathrm{pr}\{\bar{M}_n \leq (1+\alpha_n)/(1+m\tilde{\sigma}^2_{jl}/\sigma^2)\}, \quad \hat{\sigma}^2_{jl} \sim (m^{-1}\sigma^2 + \tilde{\sigma}^2_{jl})\chi^2_n/n \\ &\leq |I_n^-| \mathrm{pr}\{\bar{M}_n \leq (1+\alpha_n)/(1+(1+o(1))a_+\alpha_n)\} \\ &= |I_n^-| \mathrm{pr}(\bar{M}_n \leq 1-\epsilon_n) \leq |I_n^-| \exp(-n\epsilon_n^2/4),\end{aligned}$$

where $\epsilon_n = \{a_+(1+o(1))-1\}\alpha_n/\{1+(1+o(1))a_+\alpha_n\}$ and the second inequality holds because $\tilde\sigma_{jl}^2/\sigma_{jl}^2 \to 1$ for all $(j,l) \in I_n^-$ under Condition 5. We have $n\epsilon_n^2 \approx \{(a_+-1)^2\alpha_0^2\log(ps_n)\}/(1+a_+\alpha_n)^2 \geq (a_+-1)^2\alpha'\log(ps_n)$ where α' is slightly smaller than α_0^2. Let $\alpha_+'' = (a_+-1)^2\alpha'/4$, then $P_n^- \leq (ps_n)^{1-\alpha_+''}$. If $\alpha_0 \geq \sqrt{12}$, then $\alpha_+'' \geq 3$ for suitable $a_+ > 2$. Similarly, we have

$$
\begin{aligned}
P_n^+ &= \mathrm{pr}(\hat I \notin I_n^+) \\
&\leq \textstyle\sum_{(j,l)\notin I_n^+} \mathrm{pr}\{\hat\sigma_{jl}^2 \geq m^{-1}\sigma^2(1+\alpha_n)\} \\
&\leq \textstyle\sum_{(j,l)\notin I_n^+} \mathrm{pr}\{\bar M_n \geq m^{-1}\sigma^2(1+\alpha_n)/(m^{-1}\sigma^2+\tilde\sigma_{jl}^2)\}, \quad \hat\sigma_{jl}^2 \sim (m^{-1}\sigma^2+\tilde\sigma_{jl}^2)\chi_n^2/n \\
&\leq ps_n\mathrm{pr}\{\bar M_n \geq (1+\alpha_n)/(1+(a_-+o(1))\alpha_n)\} \\
&\leq ps_n\mathrm{pr}(\bar M_n \geq 1+\epsilon_n'), \\
&\leq ps_n\exp(-3n\epsilon_n'^2/16),
\end{aligned}
$$

where $\epsilon_n' = \{1-o(1)-a_-\}\alpha_n/\{1+(o(1)+a_-)\alpha_n\}$ and the third inequality holds because $m(\tilde\sigma_{jl}^2-\sigma_{jl}^2) = o(\alpha_n)$ for all $(j,l) \notin I_n^+$ under Condition 5. We have $n\epsilon_n'^2 \approx \{(1-a_-)^2\alpha_0^2\log(ps_n)\}/(1+a_-\alpha_n)^2 \geq (1-a_-)^2\alpha'\log(ps_n)$ where α' is slightly smaller than α_0^2. Let $\alpha_-'' = 3(1-a_-)^2\alpha'/16$, then $P_n^+ \leq (ps_n)^{1-\alpha_-''}$. If $\alpha_0 \geq \sqrt{12}$, then $\alpha_-'' > 2$ for suitable $0 < a_- < 1-\sqrt{8/9}$. By a Borel-Cantelli argument, the result follows from the bounds on P_n^- and P_n^+. □

Proof of Lemma 1. It is straightforward to obtain the bounds on cardinality of I_n^- and I_n^+ based on sparsity assumptions. Combing Lemma S1 yields the final results. □

Proof of Theorem S1. For the approximated scores $\tilde{\eta}_k$,

$$\begin{aligned} |\tilde{\eta}_{ik} - \eta_{ik}| &= |<\tilde{\boldsymbol{x}}_i, \tilde{\boldsymbol{\psi}}_k>_{\mathbb{H}} - <\boldsymbol{x}_i, \boldsymbol{\psi}_k>_{\mathbb{H}}| \\ &= |<\tilde{\boldsymbol{x}}_i, \tilde{\boldsymbol{\psi}}_k - \boldsymbol{\psi}_k>_{\mathbb{H}} + <\tilde{\boldsymbol{x}}_i - \boldsymbol{x}_i, \boldsymbol{\psi}_k>_{\mathbb{H}}| \\ &\leq \|\tilde{\boldsymbol{x}}_i\|\|\tilde{\boldsymbol{\psi}}_k - \boldsymbol{\psi}_k\| + \|\tilde{\boldsymbol{x}}_i - \boldsymbol{x}_i\|. \end{aligned}$$

For the approximated curves $\tilde{\boldsymbol{x}}_i^{r_n}$,

$$\begin{aligned} \|\tilde{\boldsymbol{x}}_i^{r_n} - \boldsymbol{x}_i^{r_n}\| &= \left\| \sum_{k=1}^{r_n} (\tilde{\eta}_{ik}\tilde{\boldsymbol{\psi}}_k - \eta_{ik}\boldsymbol{\psi}_k) \right\| \\ &\leq \left\| \sum_{k=1}^{r_n} \eta_{ik}(\tilde{\boldsymbol{\psi}}_k - \boldsymbol{\psi}_k) \right\| + \left\| \sum_{k=1}^{r_n} \tilde{\boldsymbol{\psi}}_k(\tilde{\eta}_{ik} - \eta_{ik}) \right\| \\ &\leq \sum_{k=1}^{r_n} |\eta_{ik}|\|\tilde{\boldsymbol{\psi}}_k - \boldsymbol{\psi}_k\| + \left\{ \sum_{k=1}^{r_n} (\tilde{\eta}_{ik} - \eta_{ik})^2 \right\}^{1/2}. \end{aligned}$$

Under the weak l_q sparsity $\|\tilde{\boldsymbol{x}}_i\| = O_p(1)$. According to Theorem 1, we establish the final results. □

Proof of Theorem S2. For the estimated scores, we have

$$|\tilde{\eta}_{ik} - \hat{\eta}_{ik}| \leq \|\tilde{\boldsymbol{x}}_i\|\|\tilde{\boldsymbol{\psi}}_k - \hat{\boldsymbol{\psi}}_k\| + |\hat{\mathbf{u}}_k^{\mathrm{T}}(\boldsymbol{\theta}_{i\hat{I}} - \hat{\boldsymbol{\theta}}_{i\hat{I}})|.$$

where $\mathbf{u}_k$ is the kth eigenvector of $\Sigma = E(\boldsymbol{\theta}_{\hat{I}}\boldsymbol{\theta}_{\hat{I}}^{\mathrm{T}})$ and the inequality follows from Proposition 1. To quantify the second term in the righthand, we have $|\theta_{ijl} - \hat{\theta}_{ijl}| = O_p(m^{-1/2})$ for all j, l by simple calculation.

Under weak l_q sparsity, we assume $\alpha > 1/2$ and we consider the most interesting case where $0 < q < 1$. We have $\|\mathbf{u}_k\|_1 = O(k^{a/2})$, then we have $|\hat{\mathbf{u}}_k^{\mathrm{T}}(\boldsymbol{\theta}_{i\hat{I}} - \hat{\boldsymbol{\theta}}_{i\hat{I}})| = O_p(k^{a/2}m^{-1/2})$.

We have $\|\tilde{\boldsymbol{x}}_i\| = O_p(1)$ under weak l_q sparsity. According to the results in Theorem 2, we establish results in Theorem S2. □

S4 Processes under the l_0 sparsity

We consider the case where only a small fraction of processes contain signals and the rest do not. Here the l_0 sparsity is in the sense of $\|\mathbf{V}\|_0 = g \ll p$. It is assumed w.l.o.g. that the first g processes contain signals with comparable energies and $V_j \equiv 0$, for $j = g+1, \ldots, p$. Moreover, the variances of coefficients for these g processes satisfy (2.2).

S4.1 Regularity conditions

Conditions S1 and S2 concern the approximation error and estimation error, respectively, under the l_0 sparsity. Note that under the l_0 setting, we do not require g to be finite generally. Thus, there exists a little difference about those conditions under these two settings.

Condition S1. $r_n^{a+1} g \left(m^{-1}\sqrt{\log p/n}\right)^{\alpha/(2\alpha+1)} = o(1)$.

Condition S2. $\max\left\{r_n^{a+1} g m^{-1}, r_n^{a+1} g n^{-1/2}\right\} = o(1)$.

To ensure that the g significant processes are consistently estimable, under the l_0 case, the signals should not be too small.

Condition S3. $\min_{j\in\{1,\ldots,g\}} \max_l \sigma_{jl}^2 \gg m^{-1}\sqrt{\log p/n}$.

S4.2 Theoretical results under the l_0 sparsity

In this section, we provide theoretical results for estimating multivariate eigenfunctions under the l_0 sparsity.

Lemma S2. *Under the l_0 sparsity, Conditions 1-2, 4-7 and S3, there exists a constant $C > 0$ such that $N_j \leq C\left(m^{-1}\sqrt{\log p/n}\right)^{-1/(2\alpha+1)}$ almost surely for $j = 1,\ldots,g$ and $N_j \stackrel{a.s.}{\to} 0$ for $j = g+1,\ldots,p$.*

Lemma S2 implies the consistent selection property, that is, all the g processes, and only those, are selected almost surely as $n \to \infty$. Without additional assumptions on the energy, it is clear that $N_j, j = 1,\ldots,g$ share the same order. From the proof of Theorems S3 and S4, we also know that

$$\|\hat{G}(s,t) - G(s,t)\|_{\mathbb{H}} = O_p\left\{g(m^{-1}\sqrt{\log p/n})^{\alpha/(2\alpha+1)} + gn^{-1/2} + gm^{-1}\right\}.$$

These three parts in the rates of convergence correspond to bias caused by thresholding, covariance estimation error and discretization error, respectively. Consequently, the rates of convergence for estimated eigenfunctions are obtained, and presented as approximation and estimation error, respectively.

Theorem S3 (Approximation Error)**.** *Under l_0 sparsity, if Conditions 1–7, S1 and S3 hold and $< \boldsymbol{\psi}_k, \tilde{\boldsymbol{\psi}}_k >_{\mathbb{H}} \geq 0$, then uniformly for $k = 1, \ldots, r_n$,*

$$\|\tilde{\boldsymbol{\psi}}_k - \boldsymbol{\psi}_k\|_{\mathbb{H}} = O\left\{k^{a+1} g \left(m^{-1}\sqrt{\log p/n}\right)^{\alpha/(2\alpha+1)}\right\}, \qquad a.s..$$

The approximation error is caused by excluding the small variances in the subset selection step. Due to the correct selection property, this error is associated with g, the number of retained coefficients N_j and the variance decaying rate α. To be specific, the term $\left(m^{-1}\sqrt{\log p/n}\right)^{\alpha/(2\alpha+1)}$, that is, $N_j^{-\alpha}$, is determined by excluding coordinates with small variances and the additional term k^{a+1} is attributed to the increasing error of approximating higher order eigenelements $\boldsymbol{\psi}_k, k = 1, \ldots, r_n$. Next we quantify the estimation error, where we consider two cases depending on whether the discretization error can be asymptotically negligible. Recall that γ quantifies the sampling rate $m = O(n^\gamma)$, where $\gamma > (1-\beta)/2$ and $p = O\{\exp(n^\beta)\}$ for $0 < \beta < 1$.

Theorem S4 (Estimation Error)**.** *Under l_0 sparsity, if Conditions 1–8, S2–S3 hold and $< \hat{\boldsymbol{\psi}}_k, \tilde{\boldsymbol{\psi}}_k >_{\mathbb{H}} \geq 0$, then uniformly for $k = 1, \ldots, r_n$, we have the following.*

Case 1. When $\gamma > 1/2$,

$$\|\tilde{\boldsymbol{\psi}}_k - \hat{\boldsymbol{\psi}}_k\|_{\mathbb{H}} = O_p\left(kgn^{-1/2}\right).$$

Case 2. When $(1-\beta)/2 < \gamma \leq 1/2$,

$$\|\tilde{\boldsymbol{\psi}}_k - \hat{\boldsymbol{\psi}}_k\|_{\mathbb{H}} = O_p\left(k^{a+1}gm^{-1}\right).$$

The correct selection property implied by Lemma S2 makes it sufficient to consider the estimation error of a small set of retained processes. Note that the estimation error does not involve the term N_j, as we quantify the discretization error of retained coefficients via retained processes using Bessel's inequality. The sampling rate γ plays an important role in the rates of convergence, which exhibits the phase transition phenomenon at $\gamma = 0.5$. For more detailed interpretation, one can refer to the discussion following Theorems 1 and 2.

Theorem S5 (Approximation Error for recovery under l_0)**.** *Under Conditions in Theorem S3, if* $< \boldsymbol{\psi}_k, \tilde{\boldsymbol{\psi}}_k >_{\mathbb{H}} \geq 0$, *then uniformly for* $k = 1, \ldots, r_n$,

$$|\tilde{\eta}_{ik} - \eta_{ik}| \;=\; O_p\left\{k^{a+1}g^{3/2}\left(m^{-1}\sqrt{\log p/n}\right)^{\frac{\alpha}{2\alpha+1}}\right\},$$

Moreover,

$$\|\boldsymbol{x}_i - \boldsymbol{x}_i^{r_n}\|_{\mathbb{H}} = O_p\left(gr_n^{1-a}\right),$$

$$\|\tilde{\boldsymbol{x}}_i^{r_n} - \boldsymbol{x}_i^{r_n}\|_{\mathbb{H}} = O_p\left\{r_n^{a+3/2}g^{3/2}\left(m^{-1}\sqrt{\log p/n}\right)^{\frac{\alpha}{2\alpha+1}}\right\}.$$

Theorem S6 (Estimation error for recovery under l_0)**.** *Under Conditions in Theorem S4, if* $< \hat{\boldsymbol{\psi}}_k, \tilde{\boldsymbol{\psi}}_k >_{\mathbb{H}} \geq 0$*, then uniformly for* $k = 1, \ldots, r_n$*,*

Case 1. When $\gamma > 1/2$*,*

$$|\tilde{\eta}_{ik} - \hat{\eta}_{ik}| = O_p\left(kg^{3/2}n^{-1/2} + k^{a/2}gm^{-1/2}\right),$$

$$\|\tilde{\boldsymbol{x}}_i^{r_n} - \hat{\boldsymbol{x}}_i^{r_n}\|_{\mathbb{H}} = O_p\left(r_n^{3/2}g^{3/2}n^{-1/2} + r_n^{(a+1)/2}gm^{-1/2}\right), \ i = 1, \ldots, n.$$

Case 2. When $(1-\beta)/2 < \gamma \leq 1/2$*,*

$$|\tilde{\eta}_{ik} - \hat{\eta}_{ik}| = O_p\left(k^{a+1}g^{3/2}m^{-1} + k^{a/2}gm^{-1/2}\right),$$

$$\|\tilde{\boldsymbol{x}}_i^{r_n} - \hat{\boldsymbol{x}}_i^{r_n}\|_{\mathbb{H}} = O_p\left(r_n^{a+3/2}g^{3/2}m^{-1} + r_n^{(a+1)/2}gm^{-1/2}\right), \ i = 1, \ldots, n.$$

It is straightforward to quantify the approximation error based on Theorem S3. For estimation error, we need to carefully investigate both the discretization and measurement errors. Basically, the first term in the rates of convergence mainly depend on the estimation of eigenfunctions, and the additional term is attributed to the measurement error. For consistent estimators of scores and recovery, we assume $\alpha > 1/2$.

S4.3 Proofs under the l_0 sparsity

Proof of Lemma S2. It is straightforward to obtain the bounds on cardinality of I_n^- and I_n^+ based on sparsity assumptions. Combing Lemma S1 yields the final results. □

Proof of Theorem S3. On the event A_n, based on the l_0 sparsity, we have

$$E\|\tilde{\boldsymbol{X}} - \boldsymbol{X}\|^2 = \sum_{(j,l)\notin \hat{I}} \sigma_{jl}^2 \leq \sum_{(j,l)\notin I_n^-} \sigma_{jl}^2 = O\left\{ g\left(\frac{1}{m}\sqrt{\frac{\log p}{n}}\right)^{\frac{2\alpha}{2\alpha+1}} \right\}.$$

According to the result of Theorem 1 in Hall and Hosseini-Nasab (2006),

$$\|\tilde{\boldsymbol{\psi}}_k - \boldsymbol{\psi}_k\| \leq 8^{1/2} k^{a+1} \left[\int\int \sum_{j=1}^{p}\sum_{j'=1}^{p} \{\tilde{G}_{jj'}(s,t) - G_{jj'}(s,t)\}^2 dsdt \right]^{1/2}.$$

So, with (S1.3), we have

$$\|\tilde{\boldsymbol{\psi}}_k - \boldsymbol{\psi}_k\| = O\left\{ k^{a+1} g\left(m^{-1}\sqrt{\log p/n}\right)^{\frac{\alpha}{2\alpha+1}} \right\}, \qquad a.s..$$

□

Proof of Theorem S4. Recall that g_n denotes the number of retained processes. Under the weak l_q sparsity, we consider to bound the terms in (S1.4) replacing g by g_n. Combining (S1.5) and (S1.7), using Cauchy-Schwarz inequality and Chebyshev's inequality, we may prove that

$$\begin{aligned} I &= \sum_{j=1}^{g}\sum_{j'=1}^{g} \int\int \left[n^{-1} \sum_{i=1}^{n} \{\check{x}_{ij}(s) - \tilde{x}_{ij}(s)\} \check{x}_{ij'}(t) \right]^2 dsdt \\ &= O_p(g^2 m^{-2}). \end{aligned}$$

With (S1.6) and (S1.7), we have $II = O_p(g^2 m^{-2})$. Using Cauchy-Schwarz inequality, we deduce that $III = O_p(g^2 n^{-1})$. Thus, we obtain $|||\hat{G} - \tilde{G}||| = O_p(gn^{-1/2} + gm^{-1})$.

Case 1. If $\gamma > 1/2$, the parametric rate dominates while discretization error is negligible, $|||\hat{G} - \tilde{G}||| = O_p(gn^{-1/2})$.

With similar arguments as proof of Theorem 2, we have $\tilde\lambda_k - \tilde\lambda_{k+1} \geq Ck^{-a-1}, k = 1, \ldots, r_n$ and $\|\hat{\boldsymbol\psi}_k - \tilde{\boldsymbol\psi}_k\|^2 \leq 8\sum_{l:l\neq k}(\tilde\lambda_k - \tilde\lambda_l)^{-2}\left\{\int \tilde{\boldsymbol\psi}_k^{\mathrm{T}}(\tilde G - \hat G)\tilde{\boldsymbol\psi}_l\right\}^2$, by analogy to (5.22) in Hall and Horowitz (2007), $E\left[\sum_{l:l\neq k}(\tilde\lambda_k - \tilde\lambda_l)^{-2}\left\{\int \tilde{\boldsymbol\psi}_k^{\mathrm{T}}(\tilde G - \hat G)\tilde{\boldsymbol\psi}_l\right\}^2\right] = O(k^2g^2n^{-1})$ holds uniformly in $k = 1, \ldots, r_n$.

Case 2. If $\gamma \leq 1/2$, the discretization error dominates, $|||\hat G - \tilde G||| = O_p(gm^{-1})$. With the result of Theorem 1 in Hall and Hosseini-Nasab (2006), the final results are established. □

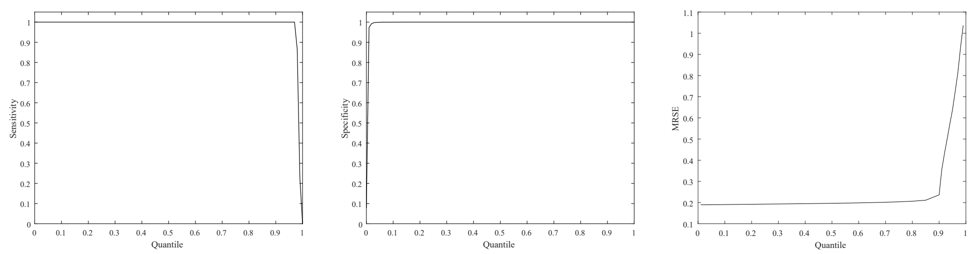
Proof of Theorem S5. Under the l_0 sparsity, $\|\tilde{\boldsymbol x}_i\| = O_p(g^{1/2})$. Based on Theorem S3 and following similar arguments to prove Theorem S1, we obtain final results. □

Proof of Theorem S6. Note that $\tilde\lambda_k u_{kjl}^2 \leq \sigma_{jl}^2$. Under l_0 sparsity, we assume $\alpha > 1/2$, then we have $\|\mathbf u_k\|_1 = O(k^{a/2}g)$. Thus, $|\hat{\mathbf u}_k^{\mathrm T}(\boldsymbol\theta_{i\hat I} - \hat{\boldsymbol\theta}_{i\hat I})| = O_p(k^{a/2}gm^{-1/2})$.

Moreover, we have $\|\tilde{\boldsymbol x}_i\| = O_p(g^{1/2})$ under l_0 sparsity. According to the results in Theorem S4, we establish the final results following similar arguments in proof of Theorem S2. □

S4.4 Simulation under the l_0 sparsity

Let $p = 50$, 100, 200 and the number of processes containing signals $g = 2$, 10, respectively. The underlying true signals $x_{ij}(t_{ijk}) = \sum_{l=1}^{s}\theta_{ijl}\phi_l(t_{ijk})$ for



(a) (b) (c)

Figure 1: The results for the l_0 sparsity setting on sensitivity (a), specificity (b), and MRSE (c), where p=100, g=10.

$j = 1, \ldots, g$, and the rest $x_{ij}(t_{ijk}) = 0$. Denote $\boldsymbol{\theta} = (\theta_{11}, \ldots, \theta_{1s}, \ldots, \theta_{g1}, \ldots, \theta_{gs})^{\mathrm{T}}$. The coefficients $\boldsymbol{\theta}_i$ are generated from $N(\mathbf{0}, C)$, where $C = VDV^{\mathrm{T}}$ with an orthonormal matrix V and a diagonal matrix with diagonal entries $D_{\nu\nu} = 16\nu^{-7/3}, \nu = 1, \ldots, gs$. The dependence between coefficients leads to correlated processes.

To evaluate the correct selection performance under the l_0 sparsity, we use the specificity and sensitivity criteria, defined as Specificity $= TN/(TN+FP)$, Sensitivity $= TP/(TP+FN)$, where TP and TN are abbreviations for true positives and true negatives, respectively, that is, the number of processes containing signals and the rest processes correctly identified by our method, similarly FP and FN stand for false positives and false negatives.

Only results with $p = 100, g = 10$ are reported, while other results revealing similar patterns are not presented for conciseness. We use $s_n = 54$

Table 1: The MSE with standard errors in parentheses for the first four eigenfunctions and the comparison of average computation time for a full sample recovery under the l_0 sparsity with $p = 100$, where the quantile $\rho = 0.5$ in our method.

	$\boldsymbol{\psi}_1$	$\boldsymbol{\psi}_2$	$\boldsymbol{\psi}_3$	$\boldsymbol{\psi}_4$
sFPCA	.057(.005)	.087(.019)	.127(.038)	.239(.134)
MFPCA	.072(.006)	.155(.023)	.286(.043)	.493(.116)
Average computation times for recovery (second)				
s_n	14	24	34	44
sFPCA	1.220	2.018	3.052	4.440
MFPCA	10.55	28.04	70.04	141.1

in the l_0 setting for presented results. In the l_0 sparsity setting, when the underlying complexity is known, the Specificity and Sensitivity analyses in Figures 1(a) and 1(b) clearly support an adequate choice of ρ that covers a broad range to yield correct selection. Moreover, the performance of recovery is quite stable with suitable ρ as shown in Figure 1(c). The above findings suggest that a slightly large ρ is preferred if model parsimony is of main concern. We see from Table 1 that our method with $\rho = 0.5$ clearly outperforms the HG method.

The design for classification follows the previous generation mechanism with $D_{\nu\nu} = 3\nu^{-2}, \nu = 1, \ldots, gs,\ g = 2$, while the mean functions are

Table 2: The averages of misclassification rates on testing samples with standard errors in parentheses across different r_n and the average computation time under the l_0 sparsity. The square brackets show the average model complexity of the proposed method with standard errors in parentheses.

Method	r_n 2	5	8	12	15	Time (second)
sFPCA +LDA	22.80(4.07) [2.00(.00)]	9.95(2.51) [2.01(.10)]	9.84(2.42) [2.00(.00)]	9.97(2.49) [2.00(.00)]	9.94(2.48) [2.00(.00)]	1.92
MFPCA +LDA	27.16(4.50)	18.57(4.14)	18.15(4.13)	17.96(4.18)	17.58(4.02)	51.90
UFPCA +ROAD	29.11(6.02)	11.98(6.34)	11.43(5.56)	11.53(5.48)	11.55(5.46)	43.15

generated in the same way as that in the weak l_q setting.

Bibliography

Bruckstein, A. M., D. L. Donoho, and M. Elad (2009). From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Review 51*(1), 34–81.

Hall, P. and J. L. Horowitz (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics 35*(1), 70–91.

Hall, P. and M. Hosseini-Nasab (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68*(1), 109–126.

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *The Annals of statistics 29*(2), 295–327.

Kong, D., K. Xue, F. Yao, and H. H. Zhang (2016). Partially functional linear regression in high dimensions. *Biometrika 103*(1), 147–159.