Supplementary material for the manuscript entitled "Interpretable Functional Principal Component Analysis"

Zhenhua Lin, Liangliang Wang and Jiguo Cao

S.1 Addtional Simulation Study and Data Application

S.1.1 Simulation 2

In this simulation study, 500 curves are simulated with the same settings as in the previous simulation study except that the three true FPCs, $\xi_k(t)$, $k = 1, 2, 3$, are simulated as $\xi_1(t) = \sin 2\pi t$, $\xi_2(t) = \sin 4\pi t$, and $\xi_3(t) = \sin 6\pi t$, where $t \in [0, 1]$. All three of these FPCs are designed such that they are nonzero in almost the entire domain. We estimate the three FPCs using our iFPCA method and the classic regularized FPCA method. We use the same cubic B-spline basis functions and the same method to choose tuning parameters as in Simulation 1. The simulation is implemented with 100 simulation replicates. In this scenario, the iFPCA method is expected to not perform as well as the classic regularized FPCA method. But we will see that the iFPCA method is still quite competitive with the classic regularized FPCA method.

S.1.2 Data Application: Canadian Weather Data

The Canadian weather data contains daily temperature recorded at 35 Canadian cities in one year (Ramsay and Silverman, 2005). A plot of this data is given in Figure S.1. The functional principal component analysis is used to detect the time intervals in which the daily temperature curves have major variations.

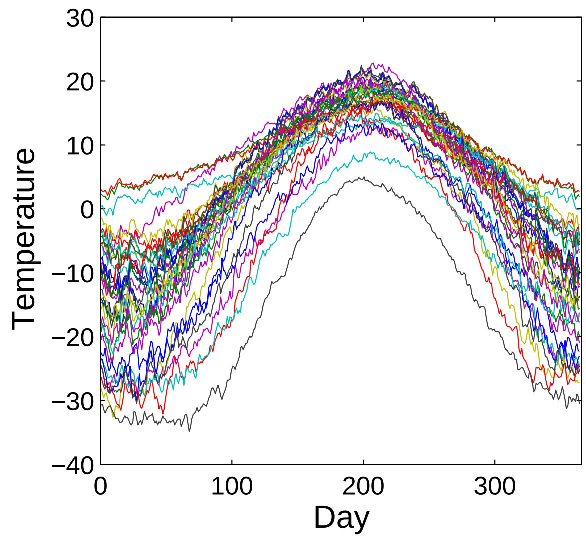

Figure S.1: Daily temperature recorded at 35 Canadian cities in one year. This figure appears in color in the electronic version of this article.

Figure S.2 displays the first three estimated FPCs for the Canadian weather data. The classic regularized FPCA method are again nonzero in the entire time interval, while the estimated three iFPCs are only nonzero in a short time interval. Hence, the estimated three iFPCs can help us identify the precise time intervals in which the 35 curves have major variations.

The estimated first iFPC is only nonzero in $[1, 150] \cup [247, 365]$, which indicates that the first major variation of the 35 temperature curves happens in the spring, fall and winter. Note that the first iFPC explains 82.7% variance of the 35 temperature curves. It can be interpreted as about 82.7% fluctuation of the temperature locates at these three seasons. The remaining variance that is not captured by the first iFPC is mostly picked up by the estimated second iFPC. This iFPC is only nonzero in [132,257], which indicates that the rest of temperature fluctuation mostly occurs in the summer. The percentages of explained variances of iFPCs are comparable to those of the classic regularized FPCs.

Figure S.3 displays the estimated FPCs using our iFPCA method and the classic regularized FPCA method in one random simulation replicate. It shows that the estimated FPCs using the iFPCA method are very close to the true FPCs. The bottom three panels in Figure S.3 also show the pointwise mean squared errors (MSE) of the estimated FPCs using the two methods. The

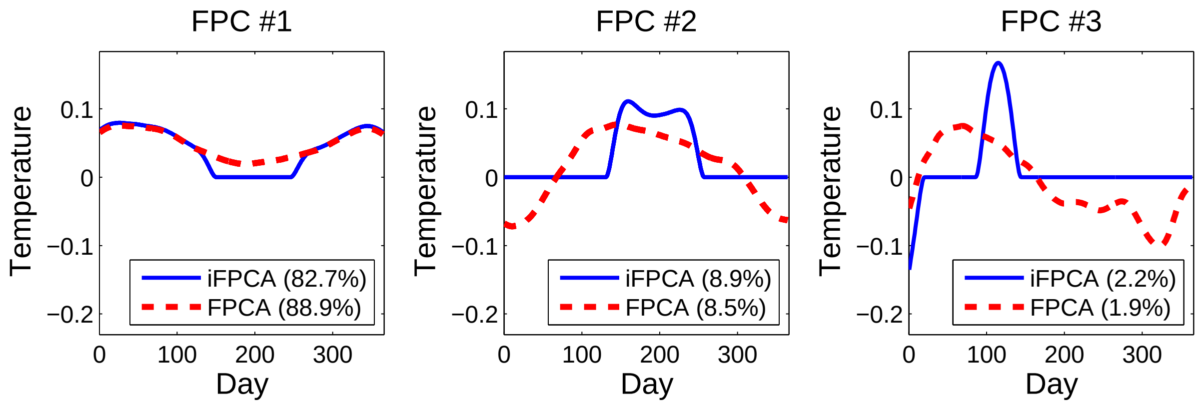

Figure S.2: The estimated three FPCs using our iFPCA method (solid lines) and the classic regularized FPCA method (dashed lines) for the Canadian weather data. Numbers in the parentheses are percentages of variance explained by the FPCs. This figure appears in color in the electronic version of this article.

pointwise MSE of the estimated FPCs using the iFPCA method is only slightly larger than those using the classic regularized FPCA method.

Figure S.4 shows that the average cumulative percentages of variance explained by the estimated FPCs using our iFPCA method are almost the same as those estimated by the classic regularized FPCA method. The differences of the cumulative percentages of explained variance using the two methods in 100 simulation replicates are no larger than 6%.

S.2 Some Details of Computation

S.2.1 Compute The Leading Eigenvalue

When p is relatively small, the Cholesky decomposition can be used to compute all eigenvalues and eigenvectors of the GEP $\mathbf{Q}_\alpha \mathbf{a}_\alpha = \lambda \mathbf{G}_\alpha \mathbf{a}_\alpha$. When p is large, the Lanczos algorithm is preferred since it can take only $O(qp^2 \log p)$ operations (Kuczyński and Woźniakowski, 1992; Leyk and Woźniakowski, 1996) to accurately estimate the leading eigenvalue and its eigenvector of the GEP $\mathbf{Q}_\alpha \mathbf{a}_\alpha = \lambda \mathbf{G}_\alpha \mathbf{a}_\alpha$, where q is a constant factor depending on the tolerance level. When p is large, q is typically much smaller than p (i.e. typically, $q \ll p$). Therefore, in total it takes $O(p^2 \log p)$ operations for the AMVL approach to approximately solve the problem (7iFPCA Algorithmequation.3.7).

The greedy backward elimination algorithm based on AVML criterion to compute the first iFPC is outlined in Algorithm 1. Note that in Algorithm 1, the update to $\mathbf{G}_\alpha^{-1}$ and $\mathbf{G}_\alpha^{-1}\mathbf{Q}_\alpha$ in each iteration can be done in $O(p^2)$ operations (see Section S2 in the supplementary file). Therefore, the computational complexity of the greedy backward elimination algorithm is $O(p^3 \log p)$ operations by noting that $\tau \leq p$. The iFPCA algorithm is outlined in Algorithm 2, which uses the projection deflation described in Section 3 in the supplementary file to iteratively compute $\widehat{\mathbf{a}}_k$ by applying the greedy backward elimination algorithm (Algorithm 1) on matrices $\mathbf{Q}_k$ and $\mathbf{G}$ until the first m iFPCs are obtained. In total, the iFPCA algorithm takes $O(nmp^2 + m^2p^2 + mp^3 \log p)$ operations to complete. When $n = O(p)$ and m is fixed, the computational complexity of the iFPCA algorithm grows in the order of $O(p^3 \log p)$. This complexity is almost the same as the regularized FPCA described in Ramsay and Silverman (2005), which requires at least $O(p^3)$ operations.

S.2.2 Efficiently Update $\mathbf{G}_\alpha^{-1}$ and $\mathbf{G}_\alpha^{-1}\mathbf{Q}_\alpha$

Suppose

$$\mathbf{Q}_{\alpha\cup\{i\}} = \begin{pmatrix} \mathbf{Q}_\alpha & \mathbf{q}_i \\ \mathbf{q}_i^T & w \end{pmatrix}, \qquad \mathbf{G}_{\alpha\cup\{i\}}^{-1} = \begin{pmatrix} \mathbf{F} & \mathbf{f}_i \\ \mathbf{f}_i^T & y \end{pmatrix}, \qquad \mathbf{G}_{\alpha\cup\{i\}}^{-1}\mathbf{Q}_{\alpha\cup\{i\}} = \begin{pmatrix} \mathbf{P}_1 & \mathbf{p}_2 \\ \mathbf{p}_3^T & c \end{pmatrix}.$$

Then it is easy to check that

$$\mathbf{G}_\alpha^{-1} = \mathbf{F} - \frac{1}{y}\mathbf{f}_i\mathbf{f}_i^T.$$

Therefore, if $\mathbf{G}_{\alpha\cup\{i\}}^{-1}$ is already computed, then only $O(p^2)$ additional operations are required to compute $\mathbf{G}_\alpha^{-1}$. Also,

$$\mathbf{G}_\alpha^{-1}\mathbf{Q}_\alpha = \mathbf{F}\mathbf{Q}_\alpha - \frac{1}{y}\mathbf{f}_i\mathbf{f}_i^T\mathbf{Q}_\alpha = (\mathbf{F}\mathbf{Q}_\alpha + \mathbf{f}_i\mathbf{q}_i^T) - y^{-1}\mathbf{f}_i(\mathbf{f}_i^T\mathbf{Q}_\alpha + y\mathbf{q}_i^T) = \mathbf{P_1} - y^{-1}\mathbf{f}_i\mathbf{p}_3^T,$$

since

$$\mathbf{G}_{\alpha\cup\{i\}}^{-1}\mathbf{Q}_{\alpha\cup\{i\}} = \begin{pmatrix} \mathbf{F} & \mathbf{f}_i \\ \mathbf{f}_i^T & y \end{pmatrix}\begin{pmatrix} \mathbf{Q}_\alpha & \mathbf{q}_i \\ \mathbf{q}_i^T & w \end{pmatrix} = \begin{pmatrix} \mathbf{F}\mathbf{Q}_\alpha + \mathbf{f}_i\mathbf{q}_i^T & \mathbf{F}\mathbf{q}_i + w\mathbf{f}_i \\ \mathbf{f}_i^T\mathbf{Q}_\alpha + y\mathbf{q}_i^T & \mathbf{f}_i^T\mathbf{q}_i + wy \end{pmatrix} = \begin{pmatrix} \mathbf{P}_1 & \mathbf{p}_2 \\ \mathbf{p}_3^T & c \end{pmatrix}.$$

Therefore, $\mathbf{G}_\alpha^{-1}\mathbf{Q}_\alpha$ can also be computed in $O(p^2)$ operations provided that $\mathbf{G}_{\alpha\cup\{i\}}^{-1}\mathbf{Q}_{\alpha\cup\{i\}}$ is computed already.

S.2.3 Projection Deflation

Recall that $\hat{r}_i^k$ denotes the residual of $\hat{X}_i$ perpendicular to the space spanned by the first $k-1$ iFPCs $\hat{\xi}_1, \hat{\xi}_2, \ldots, \hat{\xi}_{k-1}$, i.e., $\hat{r}_i^k = \hat{X}_i - \sum_{j=1}^{k-1} b_{ij}\hat{\xi}_j$. However, as the iFPCs $\hat{\xi}_1, \hat{\xi}_2, \ldots, \hat{\xi}_{k-1}$ may not be orthogonal in the inner product $\langle\cdot,\cdot\rangle$, the coefficients b_{ij}'s are dependent on the index k. In other words, we have to recompute b_{ij}'s for each $\hat{r}_i^k$, which is computationally inefficient. To address this issue, we use the Gram-Schmidt process to orthogonalize $\hat{\xi}_1, \hat{\xi}_2, \ldots, \hat{\xi}_{k-1}$ as we progressively compute iFPCs.

Suppose $\eta_1, \eta_2, \ldots, \eta_{k-1}$ are orthogonal functions that are produced by Gram–Schmidt process on $\widehat{\xi}_1, \widehat{\xi}_2, \ldots, \widehat{\xi}_{k-1}$. Instead of representing the residual $\widehat{r}_i^k$ by $\widehat{r}_i^k = \widehat{X}_i - \sum_{j=1}^{k-1} b_{ij}\widehat{\xi}_j$, we write $\widehat{r}_i^k = \widehat{X}_i - \sum_{j=1}^{k-1} b_{ij}\eta_j$ and each $\eta_\nu = \sum_{j=1}^{p} h_{\nu j}\phi_j$ for $\nu = 1, 2, \ldots$. Recall that the element $s_{ij}^{(k)} = \mathbf{S}_k(i,j)$ of the matrix $\mathbf{S}_k$ is the coefficient of $\widehat{r}_i^k$ with respect to the jth basis function ϕ_j. Suppose we already have $\mathbf{S}_k, \eta_1, \eta_2, \ldots, \eta_{k-1}, \widehat{\xi}_k$ and we want to compute $\mathbf{S}_{k+1}$ (note that $\mathbf{S}_1 = \mathbf{S}$). We first continue the Gram-Schmidt process on $\eta_1, \eta_2, \ldots, \eta_{k-1}$ and $\widehat{\xi}_k = \widehat{\mathbf{a}}_k^T\boldsymbol{\phi}$ to obtain η_k. It is equivalent to computing the residual ϑ of $\widehat{\xi}_k$ with respect to the space spanned by $\eta_1, \eta_2, \ldots, \eta_{k-1}$ and then normalize ϑ. Let $c_\nu\eta_\nu$ be the projection of $\widehat{\xi}_k$ on $\eta_\nu = \mathbf{h}_\nu^T\boldsymbol{\phi}$. Then $\langle\widehat{\xi}_k - c_\nu\eta_\nu, \eta_\nu\rangle = 0$, which implies $\widehat{\mathbf{a}}_k\mathbf{W}\mathbf{h}_\nu = c_\nu\mathbf{h}_\nu^T\mathbf{W}\mathbf{h}_\nu$ and hence

$$c_\nu = \frac{\widehat{\mathbf{a}}_k\mathbf{W}\mathbf{h}_\nu}{\mathbf{h}_\nu^T\mathbf{W}\mathbf{h}_\nu}.$$

Let $\mathbf{c}$ denote the vector $(c_1, c_2, \ldots, c_k)^T$. Then $\vartheta = \widehat{\xi}_k - \sum_{\nu=1}^{k} c_\nu\eta_\nu = (\widehat{\mathbf{a}}_k^T - \mathbf{c}^T\mathbf{H})\boldsymbol{\phi}$ where $\mathbf{H}(\nu, j) = h_{\nu j}$. By normalizing ϑ to have unit length, we obtain η_k. This process takes $O(kp^2)$ operations to complete.

To compute the residual of $\widehat{X}_i$ perpendicular to the subspace spanned by the first k iFPCs, we only need to compute the residual of $\widehat{r}_i^k$ with respect to η_k since by assumption $\widehat{r}_i^k \perp \eta_j$ and $\eta_k \perp \eta_j$ for all $j < k$. Thus, the residual $\widehat{r}_i^{k+1}$ is found to make $\widehat{r}_i^k = b_{ik}\eta_k + \widehat{r}_i^{k+1}$ hold for some scalar b_{ik} and $\langle\widehat{r}_i^{k+1}, \eta_k\rangle = 0$. In matrix form, it becomes $\mathbf{S}^k\boldsymbol{\phi} = \mathbf{b}\mathbf{h}_k^T\boldsymbol{\phi} + \mathbf{S}^{k+1}\boldsymbol{\phi}$ and $\mathbf{S}^{k+1}\mathbf{W}\mathbf{h}_k = \mathbf{0}$, where

$\mathbf{h}_k = (h_{k1}, h_{k2}, \dots, h_{kp})^T$. They together give

$$\mathbf{b} = \frac{\mathbf{S}_k \mathbf{W} \mathbf{h}_k}{\mathbf{h}_k^T \mathbf{W} \mathbf{h}_k}, \qquad \mathbf{S}_{k+1} = \mathbf{S}_k - \frac{\mathbf{S}_k \mathbf{W} \mathbf{h}_k \mathbf{h}_k^T}{\mathbf{h}_k^T \mathbf{W} \mathbf{h}_k}.$$

This process takes $O(np^2)$ operations to complete. Thus, in total it takes $O(np^2 + kp^2)$ operations to compute $\mathbf{S}_{k+1}$.

S.2.4 Cross-Validation for Selecting Tuning Parameters τ_k's

Suppose the first $(k-1)$ iFPCs, $\widehat{\xi}_1(t), \dots, \widehat{\xi}_{k-1}(t)$, have been obtained, and we are now estimating the k-th iFPC $\widehat{\xi}_k(t)$. To determine τ_k, for each positive integer $\tau_k = 1, 2, \dots, p$ (Recall that p is the number of basis functions, and is the maximum possible value of τ_k.), we compute the corresponding CV score. For example, if a K-fold CV procedure is employed, then the CV score for τ_k is

$$CV(\tau_k) = \sum_{j=1}^{K} \sum_{i \in \mathcal{P}_j} \|r_i^{(-j)}(\tau_k)\|_2^2,$$

where $\mathcal{P}_j$ is the j-th partition used in the K-fold CV procedure, and $r_i^{(-j)}(\tau_k)$ is the residual of $X_i(t)$ orthogonal to the span of the iFPCs $\widehat{\xi}_1(t), \dots, \widehat{\xi}_{k-1}(t)$, and $\widehat{\xi}_k^{(-j)}(t)$. Here $\widehat{\xi}_k^{(-j)}(t)$ is the k-th iFPC estimated from all sample curves excluding those in the j-th partition.

S.3 Proof of Proposition 1

Proof. First of all, note that when $\mathbf{B}$ is invertible, the generalized eigenproblem $\mathbf{A}\mathbf{z} = \lambda \mathbf{B}\mathbf{z}$ is equivalent to the ordinary eigenproblem $\mathbf{B}^{-1}\mathbf{A}\mathbf{z} = \lambda \mathbf{z}$. Write the matrices $\mathbf{A}$,$\mathbf{B}$, $\mathbf{B}^{-1}\mathbf{A}$ and the vector $\mathbf{v}$ in block forms so that

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_\alpha & \mathbf{a}_i \\ \mathbf{a}_i^T & w \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} \mathbf{B}_\alpha & \mathbf{b}_i \\ \mathbf{b}_i^T & z \end{pmatrix}, \qquad \mathbf{B}^{-1} = \begin{pmatrix} \mathbf{F} & \mathbf{f}_i \\ \mathbf{f}_i^T & y \end{pmatrix}, \qquad \mathbf{v} = \begin{pmatrix} \mathbf{v}_\alpha \\ v_i \end{pmatrix}.$$

Nothing that $\mathbf{B}_\alpha^{-1} = \mathbf{F} - \frac{1}{y}\mathbf{f}_i\mathbf{f}_i^T$, we have

$$\mathbf{v}_\alpha^T \mathbf{B}_\alpha^{-1}\mathbf{A}_\alpha\mathbf{v}_\alpha = \mathbf{v}_\alpha^T(\mathbf{F}\mathbf{A}_\alpha - y^{-1}\mathbf{f}_i\mathbf{f}_i^T\mathbf{A}_\alpha)\mathbf{v}_\alpha \leq \lambda(\mathbf{A}_\alpha, \mathbf{B}_\alpha)\|\mathbf{v}_\alpha\|_2^2 = \lambda(\mathbf{A}_\alpha, \mathbf{B}_\alpha)(1 - v_i^2),$$

or equivalently,

$$\mathbf{v}_\alpha^T(\mathbf{F}\mathbf{A}_\alpha + \mathbf{f}_i\mathbf{a}_i^T)\mathbf{v}_\alpha - y^{-1}\mathbf{v}_\alpha^T\mathbf{f}_i(\mathbf{f}_i^T\mathbf{A}_\alpha + y\mathbf{a}_i^T)\mathbf{v}_\alpha \leq \lambda(\mathbf{A}_\alpha, \mathbf{B}_\alpha)(1 - v_i^2). \tag{S.1}$$

Now write $\mathbf{B}^{-1}\mathbf{A}$ in a block form so that

$$\mathbf{B}^{-1}\mathbf{A} = \begin{pmatrix} \mathbf{P}_1 & \mathbf{p}_2 \\ \mathbf{p}_3^T & c \end{pmatrix}.$$

Then $\mathbf{F}\mathbf{A}_\alpha + \mathbf{f}_i\mathbf{a}_i^T = \mathbf{P}_1$ and $\mathbf{f}_i^T\mathbf{A}_\alpha + y\mathbf{a}_i^T = \mathbf{p}_3^T$ since

$$\mathbf{B}^{-1}\mathbf{A} = \begin{pmatrix} \mathbf{F} & \mathbf{f}_i \\ \mathbf{f}_i^T & y \end{pmatrix}\begin{pmatrix} \mathbf{A}_\alpha & \mathbf{a}_i \\ \mathbf{a}_i^T & w \end{pmatrix} = \begin{pmatrix} \mathbf{F}\mathbf{A}_\alpha + \mathbf{f}_i\mathbf{a}_i^T & \mathbf{F}\mathbf{a}_i + w\mathbf{f}_i \\ \mathbf{f}_i^T\mathbf{A}_\alpha + y\mathbf{a}_i^T & \mathbf{f}_i^T\mathbf{a}_i + wy \end{pmatrix} = \begin{pmatrix} \mathbf{P}_1 & \mathbf{p}_2 \\ \mathbf{p}_3^T & c \end{pmatrix}. \tag{S.2}$$

From (S.2) and $\mathbf{v}^T\mathbf{B}^{-1}\mathbf{A}\mathbf{v} = \lambda(\mathbf{A}, \mathbf{B})$, we have

$$\mathbf{v}_\alpha^T\mathbf{P}_1\mathbf{v}_\alpha + v_i\mathbf{p}_3^T\mathbf{v}_\alpha + v_i\mathbf{v}_\alpha^T\mathbf{p}_2 + v_i^2 c = \lambda(\mathbf{A}, \mathbf{B}).$$

Thus,

$$\mathbf{v}_\alpha^T\mathbf{P}_1\mathbf{v}_\alpha = \lambda(\mathbf{A}, \mathbf{B}) - \left(v_i\mathbf{v}_\alpha^T(\mathbf{p}_2 + \mathbf{p}_3) + v_i^2 c\right).$$

With (S.1), this gives

$$\lambda(\mathbf{A}_\alpha, \mathbf{B}_\alpha) \geq \frac{\lambda(\mathbf{A}, \mathbf{B}) - v_i\mathbf{v}_\alpha^T(\mathbf{p}_2 + \mathbf{p}_3) - v_i^2 c - y^{-1}\mathbf{v}_\alpha^T\mathbf{f}_i\mathbf{p}_3^T\mathbf{v}_\alpha}{1 - h^2},$$

or

$$\lambda(\mathbf{A}, \mathbf{B}) - \lambda(\mathbf{A}_\alpha, \mathbf{B}_\alpha) \leq \frac{y^{-1}\mathbf{v}_\alpha^T\mathbf{f}_i\mathbf{p}_3^T\mathbf{v}_\alpha + v_i\mathbf{v}_\alpha^T(\mathbf{p}_2 + \mathbf{p}_3) + v_i^2 c - v_i^2\lambda(\mathbf{A}, \mathbf{B})}{1 - v_i^2}.$$

Noting that $\mathbf{v}_\alpha^T \mathbf{p}_3 + v_i c = v_i \lambda(\mathbf{A}, \mathbf{B})$, we have further

$$\lambda(\mathbf{A}, \mathbf{B}) - \lambda(\mathbf{A}_\alpha, \mathbf{B}_\alpha) \leq \frac{v_i y^{-1} \mathbf{f}_i^T \mathbf{v}_\alpha [\lambda(\mathbf{A}, \mathbf{B}) - c] + v_i \mathbf{v}_\alpha^T \mathbf{p}_2}{1 - v_i^2}. \tag{S.3}$$

This proves Theorem 1. □

S.4 Proofs of Theorem 2, 3, 4

The proofs for Theorem 2, 3, 4 are organized as follows. We first outline our proof strategy for Theorem 2. Then we briefly introduce some results important to our proofs from Qi and Zhao (2011). After that, we start our proofs by first proving Lemma 1. This lemma not only serves as a technical tool in the course of proving Lemma 3, but also is of independent interest. Roughly speaking, it asserts two finite dimensional Hilbert spaces are close to each other if their bases are close to each other. Lemma 3 is our primary lemma. Some technical details of the proof of this lemma are moved to Lemma 4 and Lemma 5 so that the mainstream of the proof is manifest. Once Lemma 3 is established, we are ready and proceed to prove Theorem 2 and Theorem 3. In order to prove Theorem 4, we first establish three auxiliary lemmas, namely, Lemma 6, Lemma 7 and Lemma 8. At last, we prove the Theorem 4.

In Qi and Zhao (2011), $\sqrt{n}(\widetilde{\xi}_k - \xi_k)$ has been shown to converge to a Gaussian random element. As $\sqrt{n}(\widehat{\xi}_k - \xi_k) = \sqrt{n}(\widehat{\xi}_k - \widetilde{\xi}_k) + \sqrt{n}(\widetilde{\xi}_k - \xi_k)$, if we can show

$$\|\widehat{\xi}_k - \widetilde{\xi}_k\| = o_p(n^{-1/2}), \tag{S.4}$$

then $\sqrt{n}(\widehat{\xi}_k - \widetilde{\xi}_k)$ converges to 0 in probability and hence by Slutsky's theorem, $\sqrt{n}(\widehat{\xi}_k - \xi_k)$ and $\sqrt{n}(\widetilde{\xi}_k - \xi_k)$ converge to the same random element in distribution. Instead of directly proving (S.4), we consider a more general problem. Take a scalar sequence $\{\iota_n\}$ convergent to zero and we choose $\rho_k = o(\iota_n)$ for all $k = 1, 2, \ldots, m$. When $\iota_n = n^{-1/2}$ we recover the assumption on $\rho_k' s$ in (*i*) of Theorem 2, and when $\iota_n = n^{-1}$, we recover the assumption on ρ_k's in (*ii*) of Theorem 2. In other words, we will show that $|\widehat{\lambda}_k - \lambda_k|$ and $\|\widehat{\xi}_k - \widetilde{\xi}_k\|^2$ have a convergence rate at least as fast as ρ_k. That is, $|\widehat{\lambda}_k - \lambda_k| = O_p(\rho_k)$ and $\|\widehat{\xi}_k - \widetilde{\xi}_k\|^2 = O_p(\rho_k)$.

In our proof, we need the "half-smoothing" operator mentioned in Silverman (1996) and studied more thoroughly in Qi and Zhao (2011). Here we cite some results from Qi and Zhao (2011). Recall that $\mathcal{D}^2$ is the second-derivative operator on W_2^2. Let $\mathcal{D}_*^2$ be the Hilbert adjoint of $\mathcal{D}^2$. Define $\mathcal{S}^2 = (\mathcal{I} + \gamma \mathcal{D}_*^2 \mathcal{D}^2)^{-1}$. The inverse $\mathcal{S}^{-1}$ of $\mathcal{S}$ exists and is a self-adjoint operator. By Lemma 6 of Qi and Zhao (2011), $\{(\widetilde{\lambda}_k, \mathcal{S}^{-1}\widetilde{\xi}_k) : k \in \mathbb{N}\}$ are eigenvalues and eigenfunctions of the compact positive operator $\mathcal{S}\widehat{\mathcal{C}}\mathcal{S}$ and moreover, there are no other eigenvalues for $\mathcal{S}\widehat{\mathcal{C}}\mathcal{S}$. Furthermore, under the assumption that the eigenvalues $\lambda_1 > \lambda_2 > \cdots > \lambda_m$ of the operator $\mathcal{C}$ have multiplicity one, $\widetilde{\lambda}_1, \widetilde{\lambda}_2, \ldots, \widetilde{\lambda}_m$ have multiplicity one almost surely. Therefore, we can assume $\widetilde{\lambda}_1 > \widetilde{\lambda}_2 > \cdots > \widetilde{\lambda}_m$ without affecting the asymptotic properties. As $\mathcal{S}\widehat{\mathcal{C}}\mathcal{S}$ is a compact self-adjoint positive operator, its eigenfunctions $\{\mathcal{S}^{-1}\widetilde{\xi}_k : k \in \mathbb{N}\}$ form an orthogonal basis of $L^2(\mathscr{I})$. Based on this fact, we can conclude that $\{\widetilde{\xi}_1, \widetilde{\xi}_2, \ldots\}$ is an orthonormal basis of W_2^2 equiped with the inner product $\langle \cdot, \cdot \rangle_\gamma$, as the operator $\mathcal{S} : L^2(\mathscr{I}) \to W_2^2$ is one-to-one. Also, for each $\eta \in W_2^2$, we have $\langle \mathcal{S}^{-1}\eta, \mathcal{S}^{-1}\eta \rangle = \langle \eta, \eta \rangle_\gamma$. Therefore, we can derive the following useful equation

$$\langle \widetilde{\xi}_i, \widehat{\mathcal{C}}\widetilde{\xi}_j \rangle = \langle \mathcal{S}^{-1}\widetilde{\xi}_i, (\mathcal{S}\widehat{\mathcal{C}}\mathcal{S})\mathcal{S}^{-1}\widetilde{\xi}_j \rangle = \widetilde{\lambda}_j \langle \mathcal{S}^{-1}\widetilde{\xi}_i, \mathcal{S}^{-1}\widetilde{\xi}_j \rangle = \begin{cases} \widetilde{\lambda}_j & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases} \tag{S.5}$$

Now we start our proofs. Let $\widetilde{\mathcal{P}}^\perp$ be the projection onto $\widetilde{\xi}_1, \widetilde{\xi}_2, \ldots, \widetilde{\xi}_{k-1}$, with respect to the inner product $\langle \cdot, \cdot \rangle_\gamma$. Let $\widehat{\mathcal{P}}^\perp$ be the projection onto $\widehat{\xi}_1, \widehat{\xi}_2, \ldots, \widehat{\xi}_{k-1}$ with respect to the same inner product structure. Note that we are dealing with two inner products in W_2^2: the one inheritent from $L^2(\mathscr{I})$ and the modified one $\langle \cdot, \cdot \rangle_\gamma$. These two projections defined in terms of the inner product $\langle \cdot, \cdot \rangle_\gamma$. Intuitively, when each $\widehat{\xi}_j$ is close to $\widetilde{\xi}_j$, the projections $\widehat{\mathcal{P}}$ and $\widetilde{\mathcal{P}}$ are expected to be close to each other in the operator norm $\| \cdot \|_\gamma$ defined as

$$\|\mathcal{U}\|_\gamma = \sup_{\eta \in W_2^2, \|\eta\|_\gamma = 1} \|\mathcal{U}\eta\|_\gamma.$$

The following lemma gives a more precise statement of this intuition. In the sequel the notation $o(1)$ is understood as some quantity converge to zero almost surely (with probability one).

Lemma 1. *Fix any* $\gamma \geq 0$.

1. *If* $\langle \widehat{\xi}_j, \widetilde{\xi}_j \rangle_\gamma^2 = 1 + o_p(\iota_n)$ *for all* $j < k$, *then* $\|\widehat{\mathcal{P}}^\perp - \widetilde{\mathcal{P}}^\perp\|_\gamma = o_p(\sqrt{\iota_n})$ *and* $\|\widehat{\mathcal{P}} - \widetilde{\mathcal{P}}\|_\gamma = o_p(\sqrt{\iota_n})$.
2. *If* $\langle \widehat{\xi}_j, \widetilde{\xi}_j \rangle_\gamma^2 = 1 + o(1)$ *for all* $j < k$, *then* $\|\widehat{\mathcal{P}}^\perp - \widetilde{\mathcal{P}}^\perp\|_\gamma = o(1)$ *and* $\|\widehat{\mathcal{P}} - \widetilde{\mathcal{P}}\|_\gamma = o(1)$.

These conclusions hold even when $\widehat{\xi}_1, \widehat{\xi}_2, \ldots, \widehat{\xi}_{k-1}$ are not orthogonal with respect to the inner product $\langle \cdot, \cdot \rangle_\gamma$.

Before the proof, we shall point out that in the above lemma claim 1 does not imply claim 2, as the convergence in 1 is "in probability" while the one in 2 is "almost surely".

Proof. We first prove 1. For each $j < k$, expanding $\widehat{\xi}_j$ by the basis $\widetilde{\xi}_1, \widetilde{\xi}_2, \ldots$ we assume $\widehat{\xi}_j = \sum_{i=1}^{\infty} b_{ji} \widetilde{\xi}_i$ where $b_{ji} = \langle \widehat{\xi}_j, \widetilde{\xi}_i \rangle_\gamma$. First note that

$$\sum_{i \neq j} b_{ji}^2 = \sum_{i=1}^{\infty} b_{ji}^2 - b_{jj}^2 = \|\widehat{\xi}_j\|_\gamma^2 - b_{jj}^2 = 1 - b_{jj}^2 = 1 - \langle \widehat{\xi}_j, \widetilde{\xi}_j \rangle_\gamma^2. \tag{S.6}$$

This implies

$$\sum_{i \neq j} b_{ji}^2 = o_p(\iota_n) \tag{S.7}$$

as by assumption $\langle \widehat{\xi}_j, \widetilde{\xi}_j \rangle_\gamma^2 = 1 + o_p(\iota_n)$. Hence,

$$\|\widetilde{\mathcal{P}}^\perp \widehat{\xi}_j - \widehat{\xi}_j\|_\gamma = \sqrt{\sum_{i \geq k} b_{ji}^2} = o_p(\sqrt{\iota_n}). \tag{S.8}$$

for each $j < k$. Let ζ be any element in W_2^2 such that $\|\zeta\|_\gamma = 1$. The projection of ζ into the subspace spanned by $\widehat{\xi}_1, \widehat{\xi}_2, \ldots, \widehat{\xi}_{k-1}$ can be written as $\widehat{\mathcal{P}}^\perp \zeta = \sum_{j=1}^{k-1} c_j \widehat{\xi}_j$ (if $\widehat{\xi}_i = g \widehat{\xi}_j$ for some scalar g and $i < j$, then we set $c_j = 0$). Let $\eta = \zeta - \widehat{\mathcal{P}}^\perp \zeta$. Then for each $j < k$,

$$0 = \langle \eta, \widehat{\xi}_j \rangle_\gamma = \sum_{i=1}^{\infty} b_{ji} \langle \eta, \widetilde{\xi}_i \rangle_\gamma = b_{jj} \langle \eta, \widetilde{\xi}_j \rangle_\gamma + \sum_{i \neq j} b_{ji} \langle \eta, \widetilde{\xi}_i \rangle_\gamma.$$

Thus,

$$|b_{jj} \langle \eta, \widetilde{\xi}_j \rangle_\gamma| = \left| \sum_{i \neq j} b_{ji} \langle \eta, \widetilde{\xi}_i \rangle_\gamma \right|. \tag{S.9}$$

On ther other hand, by Cauchy-Schwarz inequality

$$\left|\sum_{i\neq j} b_{ji}\langle \eta, \widetilde{\xi}_i\rangle_\gamma\right|^2 \leq \left(\sum_{i\neq j} b_{ji}^2\right)\left(\sum_{i\neq j}\langle \eta, \widetilde{\xi}_i\rangle_\gamma^2\right) \leq \left(\sum_{i\neq j} b_{ji}^2\right)\|\eta\|_\gamma^2 \leq \left(\sum_{i\neq j} b_{ji}^2\right). \tag{S.10}$$

Combining (S.9) and (S.10) we have

$$0 \leq |b_{jj}\langle \eta, \widetilde{\xi}_j\rangle|_\gamma^2 \leq \left(\sum_{i\neq j} b_{ji}^2\right).$$

In light of (S.7), this implies

$$|\langle \eta, \widetilde{\xi}_j\rangle_\gamma|^2 \leq |b_{jj}|^{-2}\left(\sum_{i\neq j} b_{ji}^2\right) = o_p(\iota_n)$$

because of $b_{jj}^2 = 1 + o(\iota_n)$. Thus,

$$\|\widetilde{\mathcal{P}}^\perp \eta\|_\gamma = \sqrt{\sum_{j<k}\langle \eta, \widetilde{\xi}_j\rangle_\gamma^2} = o_p(\sqrt{\iota_n}).$$

This result, combined with (S.8), gives

$$\begin{aligned}
\|\widehat{\mathcal{P}}^\perp \zeta - \widetilde{\mathcal{P}}^\perp \zeta\|_\gamma &= \left\|\sum_{j<k} c_j \widehat{\xi}_j - \sum_{j<k} c_j \widetilde{\mathcal{P}}^\perp \widehat{\xi}_j - \widetilde{\mathcal{P}}^\perp \eta\right\|_\gamma \\
&\leq \sum_{j=1}^{k-1} |c_j| \|\widetilde{\mathcal{P}}^\perp \widehat{\xi}_j - \widehat{\xi}_j\|_\gamma + \|\widetilde{\mathcal{P}}^\perp \eta\|_\gamma \\
&= \sum_{j=1}^{k-1} |c_j| o_p(\sqrt{\iota_n}) + o_p(\sqrt{\iota_n}).
\end{aligned} \tag{S.11}$$

When $\widehat{\xi}_1, \widehat{\xi}_2, \ldots, \widehat{\xi}_{k-1}$ are orthogonal, $c_j = \langle \zeta, \widehat{\xi}_j \rangle_\gamma$ and hence $|c_j| \le 1$ for each $j < k$. In this case,

$$\sum_{j=1}^{k-1} |c_j| o_p(\sqrt{\iota_n}) + o_p(\sqrt{\iota_n}) = o_p(\sqrt{\iota_n}). \tag{S.12}$$

When $\widehat{\xi}_1, \widehat{\xi}_2, \ldots, \widehat{\xi}_{k-1}$ are not orthogonal, using ϑ to denote $\widehat{\mathcal{P}}^\perp \zeta$, we have

$$\sum_{j<k} c_j \sum_{i \ge 1} b_{ji} \widetilde{\xi}_i = \sum_{i \ge 1} \langle \widehat{\mathcal{P}}^\perp \zeta, \widetilde{\xi}_i \rangle_\gamma \widetilde{\xi}_i,$$

so that $\sum_{j<k} c_j b_{ji} = \langle \vartheta, \widetilde{\xi}_i \rangle_\gamma$ for all $i \ge 1$. Let $\mathbf{c} = (c_1, c_2, \ldots, c_{k-1})^T$, $\mathbf{B}$ be a $(k-1) \times (k-1)$ matrix such that $\mathbf{B}(i,j) = b_{ji}$ for $1 \le i, j \le k-1$, and $\mathbf{h} = (\langle \vartheta, \widetilde{\xi}_1 \rangle_\gamma, \langle \vartheta, \widetilde{\xi}_2 \rangle_\gamma, \ldots, \langle \vartheta, \widetilde{\xi}_{k-1} \rangle_\gamma)^T$. Then by a matrix form we have $\mathbf{Bc} = \mathbf{h}$. Because of $\sum_{i \ne j} b_{ji}^2 = o_p(\iota_n)$ and $b_{jj}^2 = 1 + o_p(\iota_n)$, $\mathbf{B}(j,i) = b_{ji} = o_p(1)$ for $i \ne j$ and and $\mathbf{B}(j,j) = 1 + o_p(1)$. By Lemma 2 (see below), $\mathbf{c} = \mathbf{h} + o_p(1)\mathbf{1}$. Therefore, we still have

$$\sum_{j=1}^{k-1} |c_j| o_p(\sqrt{\iota_n}) + o_p(\sqrt{\iota_n}) = o_p(\sqrt{\iota_n}). \tag{S.13}$$

Combining (S.11), (S.12) and (S.13), we have

$$\|\widehat{\mathcal{P}}^\perp \zeta - \widetilde{\mathcal{P}}^\perp \zeta\|_\gamma = o_p(\sqrt{\iota_n}).$$

Since this holds for every $\zeta \in W_2^2$ such that $\|\zeta\|_\gamma = 1$, we have $\|\widetilde{\mathcal{P}}^\perp - \widehat{\mathcal{P}}^\perp\|_\gamma = o_p(\sqrt{\iota}_n)$. Since $\mathcal{I} = \widetilde{\mathcal{P}} + \widetilde{\mathcal{P}}^\perp = \widehat{\mathcal{P}} + \widehat{\mathcal{P}}^\perp$, we also have $\|\widetilde{\mathcal{P}} - \widehat{\mathcal{P}}\|_\gamma = o_p(\sqrt{\iota_n})$.

The above proof is still valid if we replace $o_p(\cdot)$ by $o(1)$. Thus, 2 is also true. □

Lemma 2. *Suppose $\mathbf{B}_n$ is a $q \times q$ matrix such that $\mathbf{B}_n^{ij} = o_p(1)$ for $i \ne j$ and $\mathbf{B}_n^{ii} = 1 + o_p(1)$. If $\mathbf{c}_n$ is a sequence of vectors in $\mathbb{R}^q$ and $\mathbf{h}$ is a vector such that $\mathbf{B}_n \mathbf{c}_n = \mathbf{h}$, where $\mathbf{h}$ is fixed. Then $\mathbf{c}_n = \mathbf{h} + o_p(1)\mathbf{1}$.*

Proof. By assumption, for any fixed $\epsilon > 0$, for sufficiently large n, we have $\mathbf{Pr}\{|\mathbf{B}_n^{ij}| > \epsilon\} < \epsilon$ for all $i \ne j$ and $\mathbf{Pr}\{|\mathbf{B}_n^{ii}| > 1 + \epsilon\} < \epsilon$ for all $i = 1, 2, \ldots, q$. Suppose Ω is the sample space. Then this means we can find a subset Ω_1 of Ω such that $\mathbf{Pr}(\Omega_1) \ge 1 - q^2\epsilon$ and for $\omega \in \Omega_1$ and sufficiently large n, $|\mathbf{B}_n^{ij}(\omega)| \leq \epsilon$ for all $i \neq j$ and $1 - \epsilon \leq |\mathbf{B}_n^{ii}(\omega)| \leq 1 + \epsilon$ for all $i = 1, 2, \ldots, q$. When ϵ is taken to be sufficiently small, $\mathbf{B}_n$ is invertible for sufficiently large n, since the determinant of $\mathbf{B}_n$ is bounded away from 0. In this case, since $\mathbf{B}_n(\omega) \to \mathbf{I}$, we have $\mathbf{B}_n^{-1}(\omega) \to \mathbf{I}$, since the set of invertible $q \times q$ matrices is a Banach algebra, where the inverse operator $\mathbf{B} \to \mathbf{B}^{-1}$ is contintuous. Thus, $\mathbf{c}_n(\omega) = \mathbf{B}_n^{-1}\mathbf{h} \to \mathbf{h}$ for all $\omega \in \Omega_1$, or in other words, $\mathbf{c}_n = \mathbf{h} + o_p(1)\mathbf{1}$. □

Lemma 3. *Suppose $\rho_k = o(\iota_n)$ where $\iota_n \to 0$. For each $k = 1, 2, \ldots m$, we have*

$$\widehat{\lambda}_k \overset{a.s.}{\to} \widetilde{\lambda}_k, \quad \widehat{\lambda}_k - \widetilde{\lambda}_k = o_p(\iota_n); \tag{S.14}$$

and

$$\langle \widehat{\xi}_k, \widetilde{\xi}_k \rangle_\gamma^2 \overset{a.s.}{\to} 1, \quad \langle \widehat{\xi}_k, \widetilde{\xi}_k \rangle_\gamma^2 = 1 + o_p(\iota_n). \tag{S.15}$$

Proof. We adopt the same strategy of proof by Silverman (1996) to use mathematical induction on the statement $\mathscr{H}_k$ of two convergences (S.14) and (S.15). In what follows we will prove $\mathscr{H}_k$ holds by assuming $\mathscr{H}_j$ holds for each $j < k$.

On the one hand,

$$\begin{aligned}
\widetilde{\lambda}_k &= \langle \widetilde{\mathcal{P}}\widetilde{\xi}_k, \widehat{\mathcal{C}}\widetilde{\mathcal{P}}\widetilde{\xi}_k \rangle \geq \langle \widetilde{\mathcal{P}}\widehat{\xi}_k, \widehat{\mathcal{C}}\widetilde{\mathcal{P}}\widehat{\xi}_k \rangle &\text{(S.16)}\\
&= \langle \widehat{\mathcal{P}}\widehat{\xi}_k, \widehat{\mathcal{C}}\widehat{\mathcal{P}}\widehat{\xi}_k \rangle + o_p(\iota_n) &\text{(S.17)}\\
&= \widehat{\lambda}_k + o_p(\iota_n), &\text{(S.18)}
\end{aligned}$$

where the (S.17) is due to Lemma 4 (see below). On the other hand,

$$\begin{aligned}
\widehat{\lambda}_k &= \langle \widehat{\mathcal{P}}\widehat{\xi}_k, \widehat{\mathcal{C}}\widehat{\mathcal{P}}\widehat{\xi}_k \rangle - \rho_k S(\widehat{\xi}_k) + \rho_k S(\widehat{\xi}_k) \\
&\geq \langle \widehat{\mathcal{P}}\widetilde{\xi}_k, \widehat{\mathcal{C}}\widehat{\mathcal{P}}\widetilde{\xi}_k \rangle - \rho_k S(\widetilde{\xi}_k) + \rho_k S(\widehat{\xi}_k) &\text{(S.19)}\\
&= \langle \widetilde{\mathcal{P}}\widetilde{\xi}_k, \widehat{\mathcal{C}}\widetilde{\mathcal{P}}\widetilde{\xi}_k \rangle + o_p(\iota_n) &\text{(S.20)}\\
&= \widetilde{\lambda}_k + o_p(\iota_n), &\text{(S.21)}
\end{aligned}$$

where the inequality (S.19) comes from the fact that $\widehat{\xi}_k$ maximizes $\langle \widehat{\mathcal{P}}\xi, \widehat{\mathcal{C}}\widehat{\mathcal{P}}\xi \rangle - \rho_k S(\xi)$ among all $\xi \in W_2^2$ such that $\|\xi\|_\gamma = 1$ and the equality (S.20) is due to Lemma 5 (see below) and $\rho_k = o(\iota_n)$.

From (S.18) and (S.21) we conclude that $|\widehat{\lambda}_k - \widetilde{\lambda}_k| = o_p(\iota_n)$. If in the above we replace all $o_p(\iota_n)$ with $o(1)$, the above proof still holds. Thus, $\widehat{\lambda}_k \overset{a.s.}{\to} \widetilde{\lambda}_k$.

We now proceed to prove (S.15). First note that

$$\begin{aligned}
\langle \widetilde{\mathcal{P}}\widehat{\xi}_k, \widehat{\mathcal{C}}\widetilde{\mathcal{P}}\widehat{\xi}_k \rangle &= \langle \sum_{j\geq k} \langle \widehat{\xi}_k, \widetilde{\xi}_j \rangle_\gamma \widetilde{\xi}_j, \widehat{\mathcal{C}} \sum_{j\geq k} \langle \widehat{\xi}_k, \widetilde{\xi}_j \rangle_\gamma \widetilde{\xi}_j \rangle \\
&= \sum_{i,j\geq k} \langle \widehat{\xi}_k, \widetilde{\xi}_i \rangle_\gamma \langle \widehat{\xi}_k, \widetilde{\xi}_j \rangle_\gamma \langle \widetilde{\xi}_i, \widehat{\mathcal{C}}\widetilde{\xi}_j \rangle \\
&= \sum_{j\geq k} \widetilde{\lambda}_j \langle \widehat{\xi}_k, \widetilde{\xi}_j \rangle_\gamma^2,
\end{aligned}$$

where the last equality is again due to (S.5). Also, since $\widetilde{\mathcal{P}}$ is a projection, $\|\widetilde{\mathcal{P}}\widehat{\xi}_k\|_\gamma^2 \leq \|\widehat{\xi}_k\|_\gamma = 1$. Then

$$\widetilde{\lambda}_k - \langle \widetilde{\mathcal{P}}\widehat{\xi}_k, \widehat{\mathcal{C}}\widetilde{\mathcal{P}}\widehat{\xi}_k \rangle \geq \widetilde{\lambda}_k \|\widetilde{\mathcal{P}}\widehat{\xi}_k\|_\gamma^2 - \sum_{j\geq k} \widetilde{\lambda}_j \langle \widehat{\xi}_k, \widetilde{\xi}_j \rangle_\gamma^2 = \sum_{j>k} (\widetilde{\lambda}_k - \widetilde{\lambda}_j) \langle \widehat{\xi}_k, \widetilde{\xi}_j \rangle_\gamma^2 \geq 0 \tag{S.22}$$

as $\widetilde{\lambda}_k > \widetilde{\lambda}_j$ for all $j > k$. On the other hand, $\widehat{\lambda}_k = \widetilde{\lambda}_k + o(\iota_n)$ and by Lemma 4,

$$\langle \widetilde{\mathcal{P}}\widehat{\xi}_k, \widehat{\mathcal{C}}\widetilde{\mathcal{P}}\widehat{\xi}_k \rangle = \langle \widehat{\mathcal{P}}\widehat{\xi}_k, \widehat{\mathcal{C}}\widehat{\mathcal{P}}\widehat{\xi}_k \rangle + o_p(\iota_n) = \widehat{\lambda}_k + o_p(\iota_n) = \widetilde{\lambda}_k + o_p(\iota_n). \tag{S.23}$$

With (S.22), this implies

$$\sum_{j>k} (\widetilde{\lambda}_k - \widetilde{\lambda}_j) \langle \widehat{\xi}_k, \widetilde{\xi}_j \rangle_\gamma^2 = o_p(\iota_n).$$

Thus,

$$\begin{aligned}
\sum_{j>k} \langle \widehat{\xi}_k, \widetilde{\xi}_j \rangle_\gamma^2 &= \frac{1}{\widetilde{\lambda}_k - \widetilde{\lambda}_{k+1}} \sum_{j>k} (\widetilde{\lambda}_k - \widetilde{\lambda}_{k+1}) \langle \widehat{\xi}_k, \widetilde{\xi}_j \rangle_\gamma^2 \\
&\leq \frac{1}{\widetilde{\lambda}_k - \widetilde{\lambda}_{k+1}} \sum_{j>k} (\widetilde{\lambda}_k - \widetilde{\lambda}_j) \langle \widehat{\xi}_k, \widetilde{\xi}_j \rangle_\gamma^2 \\
&= o_p(\iota_n) \qquad \text{(S.24)}
\end{aligned}$$

since $\widetilde{\lambda}_k - \widetilde{\lambda}_j > 0$ almost surely for all $j > k$. Also, we have

$$\begin{aligned}
\sqrt{\sum_{j<k} \langle \widehat{\xi}_k, \widetilde{\xi}_j \rangle_\gamma^2} &= \sqrt{\sum_{j<k} \langle \widetilde{\mathcal{P}}^\perp \widehat{\xi}_k, \widetilde{\mathcal{P}}^\perp \widetilde{\xi}_j \rangle_\gamma^2} \\
&\leq \sum_{j<k} |\langle \widetilde{\mathcal{P}}^\perp \widehat{\xi}_k, \widetilde{\mathcal{P}}^\perp \widetilde{\xi}_j \rangle_\gamma| \\
&= \sum_{j<k} \Big(|\langle \widehat{\mathcal{P}}^\perp \widehat{\xi}_k, \widetilde{\mathcal{P}}^\perp \widetilde{\xi}_j \rangle_\gamma| + o_p(\sqrt{\iota_n}) \Big) && (\widehat{\mathcal{P}}^\perp \widehat{\xi}_k = 0) \\
&= o_p(\sqrt{\iota_n}).
\end{aligned}$$

Thus,

$$\sum_{j<k} \langle \widehat{\xi}_k, \widetilde{\xi}_j \rangle_\gamma^2 \leq o_p(\iota_n).$$

This result, with (S.24) together, gives the desired conclusion (S.15):

$$\langle \widehat{\xi}_k, \widetilde{\xi}_k \rangle_\gamma^2 = 1 - \sum_{j<k} \langle \widehat{\xi}_k, \widetilde{\xi}_j \rangle_\gamma^2 - \sum_{j>k} \langle \widehat{\xi}_k, \widetilde{\xi}_j \rangle_\gamma^2 = 1 + o_p(\iota_n).$$

The above proof is still valid if we replace $o_p(\cdot)$ by $o(1)$. Thus, we complete the proof. □

Lemma 4. *Suppose $\mathscr{H}_j$ holds for each $j < k$. Then $\langle \widetilde{\mathcal{P}} \widehat{\xi}_k, \widehat{\mathcal{C}} \widetilde{\mathcal{P}} \widehat{\xi}_k \rangle = \langle \widehat{\mathcal{P}} \widehat{\xi}, \widehat{\mathcal{C}} \widehat{\mathcal{P}} \widehat{\xi}_k \rangle + o_p(\iota_n)$ and $\langle \widetilde{\mathcal{P}} \widehat{\xi}_k, \mathcal{C} \widetilde{\mathcal{P}} \widehat{\xi}_k \rangle = \langle \widehat{\mathcal{P}} \widehat{\xi}, \mathcal{C} \widehat{\mathcal{P}} \widehat{\xi}_k \rangle + o(1)$.*

Proof. Clearly, $\widetilde{\mathcal{P}} \widehat{\xi}_k = \widehat{\xi}_k - \widetilde{\mathcal{P}}^\perp \widehat{\xi}_k$. Thus

$$\langle \widetilde{\mathcal{P}} \widehat{\xi}_k, \widehat{\mathcal{C}} \widetilde{\mathcal{P}} \widehat{\xi}_k \rangle = \langle \widehat{\xi}_k, \widehat{\mathcal{C}} \widehat{\xi}_k \rangle - \langle \widehat{\xi}_k, \widehat{\mathcal{C}} \widetilde{\mathcal{P}}^\perp \widehat{\xi}_k \rangle - \langle \widetilde{\mathcal{P}}^\perp \widehat{\xi}_k, \widehat{\mathcal{C}} \widehat{\xi}_k \rangle + \langle \widetilde{\mathcal{P}}^\perp \widehat{\xi}_k, \widehat{\mathcal{C}} \widetilde{\mathcal{P}}^\perp \widehat{\xi}_k \rangle. \tag{S.25}$$

First, note that

$$\langle \widehat{\xi}_k, \widehat{\mathcal{C}} \widehat{\xi}_k \rangle = \langle \widehat{\mathcal{P}} \widehat{\xi}_k, \widehat{\mathcal{C}} \widehat{\mathcal{P}} \widehat{\xi}_k \rangle \tag{S.26}$$

as $\widehat{\xi}_k = \widehat{\mathcal{P}} \widehat{\xi}_k$. Next, expanding $\widehat{\xi}_k$ by the basis $\{\widetilde{\xi}_j : j \in \mathbb{N}\}$ we get $\widehat{\xi}_k = \sum_{j \geq 1} \langle \widehat{\xi}_k, \widetilde{\xi}_j \rangle_\gamma \widetilde{\xi}_j$. Thus, $\widetilde{\mathcal{P}}^\perp \widehat{\xi}_k = \sum_{j<k} \langle \widehat{\xi}_k, \widetilde{\xi}_j \rangle_\gamma \widetilde{\xi}_j$, and

$$\langle\widehat{\xi}_k, \widehat{\mathcal{C}}\widetilde{\mathcal{P}}^{\perp}\widehat{\xi}_k\rangle = \langle\sum_{j\geq 1}\langle\widehat{\xi}_k, \widetilde{\xi}_j\rangle_\gamma\widetilde{\xi}_j, \widehat{\mathcal{C}}\sum_{j<k}\langle\widehat{\xi}_k, \widetilde{\xi}_j\rangle_\gamma\widetilde{\xi}_j\rangle \tag{S.27}$$

$$= \langle\sum_{j\geq 1}\langle\widehat{\xi}_k, \widetilde{\xi}_j\rangle_\gamma\widetilde{\xi}_j, \mathcal{S}^{-1}\mathcal{S}\widehat{\mathcal{C}}\mathcal{S}\mathcal{S}^{-1}\sum_{j<k}\langle\widehat{\xi}_k, \widetilde{\xi}_j\rangle_\gamma\widetilde{\xi}_j\rangle \tag{S.28}$$

$$= \langle\sum_{j\geq 1}\langle\widehat{\xi}_k, \widetilde{\xi}_j\rangle_\gamma\mathcal{S}^{-1}\widetilde{\xi}_j, (\mathcal{S}\widehat{\mathcal{C}}\mathcal{S})\sum_{j<k}\langle\widehat{\xi}_k, \widetilde{\xi}_j\rangle_\gamma\mathcal{S}^{-1}\widetilde{\xi}_j\rangle \tag{S.29}$$

$$= \langle\sum_{j\geq 1}\langle\widehat{\xi}_k, \widetilde{\xi}_j\rangle_\gamma\mathcal{S}^{-1}\widetilde{\xi}_j, \sum_{j<k}\langle\widehat{\xi}_k, \widetilde{\xi}_j\rangle_\gamma\widetilde{\lambda}_j\mathcal{S}^{-1}\widetilde{\xi}_j\rangle \tag{S.30}$$

$$= \sum_{j<k}\widetilde{\lambda}_j\langle\widehat{\xi}_k, \widetilde{\xi}_j\rangle_\gamma^2\langle\mathcal{S}^{-1}\widetilde{\xi}_j, \mathcal{S}^{-1}\widetilde{\xi}_j\rangle \tag{S.31}$$

$$= \sum_{j<k}\widetilde{\lambda}_j\langle\widehat{\xi}_k, \widetilde{\xi}_j\rangle_\gamma^2, \tag{S.32}$$

where from (S.28) to (S.29) we rely one the self-adjointness of $\mathcal{S}^{-1}$, from (S.29) to (S.30) we use the fact that $\mathcal{S}^{-1}\widetilde{\xi}_j$ is an eigenfunction of $\mathcal{S}\widehat{\mathcal{C}}\mathcal{S}$, from (S.30) to (S.31) we rely on the orthogonality of eigenfunctions $\mathcal{S}^{-1}\widetilde{\xi}_1, \mathcal{S}^{-1}\widetilde{\xi}_2, \ldots$, and from (S.31) to (S.32) we use the identity $\langle\mathcal{S}^{-1}\widetilde{\xi}_j, \mathcal{S}^{-1}\widetilde{\xi}_j\rangle = \|\widetilde{\xi}_j\|_\gamma = 1$. By Lemma 1,

$$\langle\widehat{\xi}_k, \widetilde{\xi}_j\rangle_\gamma = \langle\widehat{\mathcal{P}}\widehat{\xi}_k, \widetilde{\mathcal{P}}^{\perp}\widetilde{\xi}_j\rangle_\gamma = \langle\widetilde{\mathcal{P}}\widehat{\xi}_k, \widetilde{\mathcal{P}}^{\perp}\widetilde{\xi}_j\rangle_\gamma + o_p(\sqrt{\iota_n}) = o_p(\sqrt{\iota_n}).$$

Thus,

$$\sum_{j<k}\widetilde{\lambda}_j\langle\widehat{\xi}_k, \widetilde{\xi}_j\rangle_\gamma^2 = \sum_{j<k}\widetilde{\lambda}_j o_p(\iota_n) = o_p(\iota_n). \tag{S.33}$$

Combining (S.32) and (S.33), we establish

$$\langle\widehat{\xi}_k, \widehat{\mathcal{C}}\widetilde{\mathcal{P}}^{\perp}\widehat{\xi}_k\rangle = o_p(\iota). \tag{S.34}$$

Similarly,

$$\langle\widetilde{\mathcal{P}}^{\perp}\widehat{\xi}_k, \widehat{\mathcal{C}}\widehat{\xi}_k\rangle = o_p(\iota_n), \tag{S.35}$$

and

$$\langle\widetilde{\mathcal{P}}^{\perp}\widehat{\xi}_k, \widehat{\mathcal{C}}\widetilde{\mathcal{P}}^{\perp}\widehat{\xi}_k\rangle = o_p(\iota_n). \tag{S.36}$$

Combining (S.25), (S.26), (S.34), (S.35) and (S.36) together, we prove that $\langle \widetilde{\mathcal{P}}\widehat{\xi}_k, \widehat{\mathcal{C}}\widetilde{\mathcal{P}}\widehat{\xi}_k\rangle = \langle \widehat{\mathcal{P}}\widehat{\xi}, \widehat{\mathcal{C}}\widehat{\mathcal{P}}\widehat{\xi}_k\rangle + o_p(\iota_n)$. Replacing $o_p(\cdot)$ by $o(1)$ gives the proof of $\langle \widetilde{\mathcal{P}}\widehat{\xi}_k, \widehat{\mathcal{C}}\widetilde{\mathcal{P}}\widehat{\xi}_k\rangle = \langle \widehat{\mathcal{P}}\widehat{\xi}, \widehat{\mathcal{C}}\widehat{\mathcal{P}}\widehat{\xi}_k\rangle + o(1)$. □

Lemma 5. *Suppose $\mathscr{H}_j$ holds for all $j < k$. Then $\langle \widehat{\mathcal{P}}\widetilde{\xi}_k, \widehat{\mathcal{C}}\widehat{\mathcal{P}}\widetilde{\xi}_k\rangle = \langle \widetilde{\xi}_k, \widehat{\mathcal{C}}\widetilde{\xi}_k\rangle + o_p(\iota_n)$ and $\langle \widehat{\mathcal{P}}\widetilde{\xi}_k, \widehat{\mathcal{C}}\widehat{\mathcal{P}}\widetilde{\xi}_k\rangle = \langle \widetilde{\xi}_k, \widehat{\mathcal{C}}\widetilde{\xi}_k\rangle + o(1)$.*

Proof. Clearly $\widehat{\mathcal{P}}\widetilde{\xi}_k = \widetilde{\xi}_k - \widehat{\mathcal{P}}^{\perp}\widetilde{\xi}_k = \widetilde{\xi}_k - \sum_{j<k}\langle \widehat{\xi}_j, \widetilde{\xi}_k\rangle_\gamma \widehat{\xi}_j$. Thus,

$$\langle \widehat{\mathcal{P}}\widetilde{\xi}_k, \widehat{\mathcal{C}}\widehat{\mathcal{P}}\widetilde{\xi}_k\rangle = \langle \widetilde{\xi}_k, \widehat{\mathcal{C}}\widetilde{\xi}_k\rangle - \langle \widetilde{\xi}_k, \widehat{\mathcal{C}}\widehat{\mathcal{P}}^{\perp}\widetilde{\xi}_k\rangle - \langle \widehat{\mathcal{P}}^{\perp}\widetilde{\xi}_k, \widehat{\mathcal{C}}\widetilde{\xi}_k\rangle + \langle \widehat{\mathcal{P}}^{\perp}\widetilde{\xi}_k, \widehat{\mathcal{C}}\widehat{\mathcal{P}}^{\perp}\widetilde{\xi}_k\rangle, \tag{S.37}$$

based on which we again use the divide-and-conquer strategy. First of all,

$$\langle \widetilde{\xi}_k, \widehat{\mathcal{C}}\widehat{\mathcal{P}}^{\perp}\widetilde{\xi}_k\rangle = \langle \widetilde{\xi}_k, \widehat{\mathcal{C}}\sum_{j<k}\langle \widehat{\xi}_j, \widetilde{\xi}_k\rangle_\gamma \widehat{\xi}_j\rangle = \sum_{j<k}\langle \widehat{\xi}_j, \widetilde{\xi}_k\rangle_\gamma \langle \widetilde{\xi}_k, \widehat{\mathcal{C}}\widehat{\xi}_j\rangle.$$

For each $j < k$, we have

$$\begin{aligned}
\langle \widetilde{\xi}_k, \widehat{\mathcal{C}}\widehat{\xi}_j\rangle &= \langle \widetilde{\xi}_k, \widehat{\mathcal{C}}\sum_{i\geq 1}\langle \widehat{\xi}_j, \widetilde{\xi}_i\rangle_\gamma \widetilde{\xi}_i\rangle \\
&= \sum_{i\geq 1}\langle \widehat{\xi}_j, \widetilde{\xi}_i\rangle_\gamma \langle \widetilde{\xi}_k, \widehat{\mathcal{C}}\widetilde{\xi}_i\rangle \\
&= \sum_{i\geq 1}\langle \widehat{\xi}_j, \widetilde{\xi}_i\rangle_\gamma \langle \widetilde{\xi}_k, \widetilde{\mathcal{C}}\widetilde{\xi}_i\rangle \\
&= \widetilde{\lambda}_k \langle \widehat{\xi}_j, \widetilde{\xi}_k\rangle_\gamma,
\end{aligned}$$

where the last equality is due to (S.5). Thus,

$$\langle \widetilde{\xi}_k, \widehat{\mathcal{C}}\widehat{\mathcal{P}}^{\perp}\widetilde{\xi}_k\rangle = \widetilde{\lambda}_k \sum_{j<k}\langle \widehat{\xi}_j, \widetilde{\xi}_k\rangle_\gamma^2.$$

By Lemma 1, we also have

$$\langle \widehat{\xi}_j, \widetilde{\xi}_k\rangle_\gamma = \langle \widehat{\mathcal{P}}^{\perp}\widehat{\xi}_j, \widetilde{\mathcal{P}}\widetilde{\xi}_k\rangle_\gamma = \langle \widetilde{\mathcal{P}}^{\perp}\widehat{\xi}_j, \widetilde{\mathcal{P}}\widetilde{\xi}_k\rangle_\gamma + o_p(\sqrt{\iota_n}) = o(\sqrt{\iota_n}). \tag{S.38}$$

Therefore,

$$\langle\widetilde{\xi}_k, \widehat{\mathcal{C}}\widehat{\mathcal{P}}^{\perp}\widetilde{\xi}_k\rangle = \widetilde{\lambda}_k \sum_{j<k} \langle\widehat{\xi}_j, \widetilde{\xi}_k\rangle_\gamma^2 = o(\iota_n). \tag{S.39}$$

Similarly,

$$\langle\widehat{\mathcal{P}}^{\perp}\widetilde{\xi}_k, \widehat{\mathcal{C}}\widetilde{\xi}_k\rangle = o_p(\iota_n). \tag{S.40}$$

Also, we have

$$\begin{aligned}
|\langle\widehat{\mathcal{P}}^{\perp}\widetilde{\xi}_k, \widehat{\mathcal{C}}\widehat{\mathcal{P}}^{\perp}\widetilde{\xi}_k\rangle| &= \sum_{i,j<k} \langle\widehat{\xi}_i, \widetilde{\xi}_k\rangle_\gamma \langle\widehat{\xi}_j, \widetilde{\xi}_k\rangle_\gamma \langle\widehat{\xi}_i, \widehat{\mathcal{C}}\widehat{\xi}_j\rangle \\
&\le \|\widehat{\mathcal{C}}\| \sum_{i,j<k} \langle\widehat{\xi}_i, \widetilde{\xi}_k\rangle_\gamma \langle\widehat{\xi}_j, \widetilde{\xi}_k\rangle_\gamma \\
&= o_p(\iota_n).
\end{aligned} \tag{S.41}$$

by (S.38) and $\|\widehat{\mathcal{C}} - \mathcal{C}\| < \infty$ almost surely. Combining (S.37), (S.39), (S.40) and (S.41), we prove the $\langle\widehat{\mathcal{P}}\widetilde{\xi}_k, \widehat{\mathcal{C}}\widehat{\mathcal{P}}\widetilde{\xi}_k\rangle = \langle\widetilde{\xi}_k, \widehat{\mathcal{C}}\widetilde{\xi}_k\rangle + o_p(\iota_n)$. Replacing $o_p(\cdot)$ by $o(1)$ we then prove $\langle\widehat{\mathcal{P}}\widetilde{\xi}_k, \widehat{\mathcal{C}}\widehat{\mathcal{P}}\widetilde{\xi}_k\rangle = \langle\widetilde{\xi}_k, \widehat{\mathcal{C}}\widetilde{\xi}_k\rangle + o(1)$. □

We are now ready to prove Theorem 2 and Theorem 3.

Proof of Theorem 2. As $\sqrt{n}(\widehat{\xi}_k - \xi_k) = \sqrt{n}(\widehat{\xi}_k - \widetilde{\xi}_k) + \sqrt{n}(\widetilde{\xi}_k - \xi_k)$, by Lemma 3, when $\rho_k = o(n^{-1})$ for all $k = 1, 2, \ldots, m$, we have

$$\begin{aligned}
\|\widehat{\xi}_k - \widetilde{\xi}_k\|^2 \le \|\widehat{\xi}_k - \widetilde{\xi}_k\|_\gamma^2 &= \|\widehat{\xi}_k\|_\gamma^2 + \|\widetilde{\xi}_k\|_\gamma^2 - 2\langle\widehat{\xi}_k, \widetilde{\xi}_k\rangle_\gamma = 2 - 2\langle\widehat{\xi}_k, \widetilde{\xi}_k\rangle_\gamma \\
&= \Big(\langle\widehat{\xi}_k, \widetilde{\xi}_k\rangle_\gamma - 1\Big)^2 + 1 - \langle\widehat{\xi}_k, \widetilde{\xi}_k\rangle_\gamma^2 = o_p(n^{-1}) + o_p(n^{-1}) \\
&= o_p(n^{-1}),
\end{aligned}$$

by noting that $\langle\widehat{\xi}_k, \widetilde{\xi}_k\rangle_\gamma = 1 + o_p(n^{-1/2})$. In other words, $\|\widehat{\xi}_k - \widetilde{\xi}_k\| \le o_p(n^{-1/2})$. Thus $\sqrt{n}(\widehat{\xi}_k - \widetilde{\xi}_k)$ converges to 0 in probability for each $k = 1, 2, \ldots, m$. So does the joint vector $\{\sqrt{n}(\widehat{\xi}_k - \widetilde{\xi}_k) : k = 1, 2, \ldots, m\}$. By Slutsky's theorem, $\{\sqrt{n}(\widehat{\xi}_k - \xi_k) : k = 1, 2, \ldots, m\}$ and $\{\sqrt{n}(\widetilde{\xi}_k - \xi_k) : k = 1, 2, \ldots, m\}$ converge to the same random element in distribution. As Qi and Zhao (2011) have shown that the latter converges to a Gaussian random element when $\gamma = o_p(n^{-1})$, $\{\sqrt{n}(\widehat{\xi}_k - \xi_k) : k = 1, 2, \ldots, m\}$ converges to the same Gaussian random element as well.

By a similar argument, we can show that when $\gamma = o_p(n^{-1/2})$ and $\rho_k = o_p(n^{-1/2})$ for $k = 1, 2, \ldots, m$, $\{\sqrt{n}(\widehat{\lambda}_k - \lambda_k) : k = 1, 2, \ldots, m\}$ converges to a Gaussian distribution with mean 0. □

Proof of Theorem 3. When $\rho_k = o(1)$, then by Lemma 3, $\|\widehat{\xi}_k - \widetilde{\xi}_k\| \leq \|\widehat{\xi}_k - \widetilde{\xi}_k\|_\gamma \stackrel{a.s.}{\to} 0$. In Silverman (1996), it has been shown that $\|\widetilde{\xi}_k - \xi_k\| \stackrel{a.s.}{\to} 0$ under the condition that $\gamma \to 0$ as $n \to \infty$. Thus,

$$\|\widehat{\xi}_k - \xi_k\| \leq \|\widehat{\xi}_k - \widetilde{\xi}_k\| + \|\widetilde{\xi}_k - \xi_k\| \leq \|\widehat{\xi}_k - \widetilde{\xi}_k\|_\gamma + \|\widetilde{\xi}_k - \xi_k\| \stackrel{a.s.}{\to} 0.$$

By a similar argument, we can show that $\widehat{\lambda}_k \stackrel{a.s.}{\to} \lambda_k$. □

Below we proceed to prove three auxiliary results in preparing proof of Theorem 4. When using projection deflation, the estimated $\widehat{\xi}_k$'s are not orthogonal in $\langle \cdot, \cdot \rangle_\gamma$. Also, note that the residuals of curves X_i's are defined in terms of inner product $\langle \cdot, \cdot \rangle$ (we can not project X_i's in terms of the modified inner product since X_i might not in the domain of $\mathcal{D}^2$). To reduce the notational burden, we reuse $\widehat{\mathcal{P}}$ to denote the projection perpendicular to $\widehat{\xi}_1, \widehat{\xi}_2, \ldots, \widehat{\xi}_{k-1}$ in the space $L^2(\mathscr{I})$. In this case, $\widehat{\mathcal{P}}$ is a self-adjoint operator on $L^2(\mathscr{I})$. Recall that $\xi_1, \xi_2, \ldots$ denote the eigenfunctions of $\mathcal{C}$. Let $\mathcal{P}$ denote the projection perpendicular to $\xi_1, \xi_2, \ldots, \xi_{k-1}$.

Lemma 6. *If $\langle \widehat{\xi}_j, \xi_j \rangle^2 = o(1)$ and $\|\widehat{\xi}_j\| = 1 + o(1)$ for all $j < k$, then*

$$\|\widehat{\mathcal{P}}^\perp - \mathcal{P}^\perp\| = o(1)$$

and

$$\|\widehat{\mathcal{P}} - \mathcal{P}\| = o(1).$$

Proof. The proof is almost along the lines of proof of Lemma 1, with $\widetilde{\xi}_j$ replaced by ξ_j, $\langle \cdot, \cdot \rangle_\gamma$ replaced by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|_\gamma$ replaced by $\| \cdot \|$. □

Recall that we use H_{k-1} to denote the subspace spanned by the first $k - 1$ eigenfunctions $\xi_1, \xi_2, \ldots, \xi_{k-1}$ of the covariance operator $\mathcal{C}$. It is easy to see that the operator $\mathcal{C}_k = \mathcal{P}\mathcal{C}\mathcal{P}$ has eigenfunctions $\xi_k, \xi_{k+1}, \ldots$ and eigenvalues $\lambda_k > \lambda_{k+1} > \cdots \geq 0$ in the subspace $H_{k-1}^{\perp}$. Its null space is exactly H_{k-1}. The following lemma shows that if an operator $\mathcal{U}$ is sufficiently close to $\mathcal{C}_k$ and $\langle \xi, \mathcal{U}\xi \rangle$ is close to λ_k enough, then projection of the function ξ of unit length on the null space of $\mathcal{C}_k$ must be close to zero.

Lemma 7. *Let $\{\mathcal{U}_n\}$ be a sequence of operators on $L^2(\mathscr{I})$ such that $\sup_{\|\xi\|_\gamma=1} |\langle \xi, \mathcal{U}_n \xi \rangle - \langle \xi, \mathcal{C}_k \xi \rangle| = o(1)$. If ξ_n^* is a function such that $\langle \xi_n^*, \mathcal{U}_n \xi_n^* \rangle = \lambda_k + o(1)$ and $\|\xi^*\|_\gamma = \|\xi^*\| + o(1)$, then $\|\mathcal{P}^{\perp} \xi_n^*\| = o(1)$.*

Proof. First of all, we have $\xi_n^* = \mathcal{P}^{\perp}\xi_n^* + \mathcal{P}\xi_n^*$. Then by Cauchy-Schwarz inequality,

$$\langle \mathcal{P}^{\perp}\xi_n^*, \mathcal{U}_n \mathcal{P}^{\perp}\xi_n^* \rangle = \langle \mathcal{P}^{\perp}\xi_n^*, \mathcal{U}_n \mathcal{P}^{\perp}\xi_n^* \rangle - \langle \mathcal{P}^{\perp}\xi_n^*, \mathcal{C}_k \mathcal{P}^{\perp}\xi_n^* \rangle = \langle \mathcal{P}^{\perp}\xi_n^*, (\mathcal{U}_n - \mathcal{C}_k) \mathcal{P}^{\perp}\xi_n^* \rangle = o(1)$$

as $\sup_{\|\xi\|_\gamma=1} |\langle \xi, \mathcal{U}_n \xi \rangle - \langle \xi, \mathcal{C}_k \xi \rangle| = o(1)$. Similarly, $\langle \mathcal{P}^{\perp}\xi_n^*, \mathcal{U}_n \mathcal{P}\xi_n^* \rangle = o(1)$, $\langle \mathcal{P}\xi_n^*, \mathcal{U}_n \mathcal{P}^{\perp}\xi_n^* \rangle = o(1)$, and $\langle \mathcal{P}\xi_n^*, \mathcal{U}_n \mathcal{P}\xi_n^* \rangle = \langle \mathcal{P}\xi_n^*, \mathcal{C}_k \mathcal{P}\xi_n^* \rangle + o(1)$. Thus,

$$\begin{aligned}\langle \xi_n^*, \mathcal{U}_n \xi_n^* \rangle &= \langle \mathcal{P}\xi_n^*, \mathcal{U}_n \mathcal{P}\xi_n^* \rangle + \langle \mathcal{P}\xi_n^*, \mathcal{U}_n \mathcal{P}^{\perp}\xi_n^* \rangle + \langle \mathcal{P}^{\perp}\xi_n^*, \mathcal{U}_n \mathcal{P}\xi_n^* \rangle + \langle \mathcal{P}^{\perp}\xi_n^*, \mathcal{U}_n \mathcal{P}^{\perp}\xi_n^* \rangle \\ &= \langle \mathcal{P}\xi_n^*, \mathcal{C}_k \mathcal{P}\xi_n^* \rangle + o(1). \end{aligned} \tag{S.42}$$

On the other hand, by the assumption $\langle \xi_n^*, \mathcal{U}_n \xi_n^* \rangle = \lambda_k + o(1)$. Combining this with (S.42) gives $\langle \mathcal{P}\xi_n^*, \mathcal{C}_k \mathcal{P}\xi_n^* \rangle = \lambda_k + o(1)$. Since

$$\langle \frac{\mathcal{P}\xi_n^*}{\|\mathcal{P}\xi_n^*\|}, \frac{\mathcal{C}_k \mathcal{P}\xi_n^*}{\|\mathcal{P}\xi_n^*\|} \rangle \leq \lambda_k,$$

we have $\lambda_k + o(1) = \langle \mathcal{P}\xi_n^*, \mathcal{C}_k \mathcal{P}\xi_n^* \rangle \leq \lambda_k \|\mathcal{P}\xi_n^*\|^2$. Since $\|\mathcal{P}\xi_n^*\|^2 \leq \|\xi_n^*\|^2 \leq 1$, we must have $\|\mathcal{P}\xi_n^*\| = 1 + o(1)$, which implies that $\|\mathcal{P}^{\perp}\xi_n^*\| = o(1)$ since $\|\xi_n^*\| = 1 + o(1)$. □

Lemma 8. *If $\mathrm{E}(\|X\|^2) < \infty$ and $\langle \widehat{\xi}_j, \xi_j \rangle^2 = o(1)$ for all $j < k$, then*

$$\sup_{\|\xi\|_\gamma=1} |\langle \xi, \widehat{\mathcal{C}}_k \xi \rangle - \langle \xi, \mathcal{C}_k \xi \rangle| = o(1).$$

Proof. For each $\xi \in W_2^2$,

$$\begin{aligned}
\langle \xi, \widehat{\mathcal{C}}_k \xi \rangle &= \int_{\mathscr{I}} \xi(s) \left(\int_{\mathscr{I}} \frac{1}{n} \sum_{i=1}^n \widehat{X}_i^k(s) \widehat{X}_i^k(t) \xi(t) \, \mathrm{d}t \right) \mathrm{d}s \\
&= \frac{1}{n} \sum_{i=1}^n \left(\int_{\mathscr{I}} \xi(s) \widehat{\mathcal{P}} X_i(s) \, \mathrm{d}s \right) \left(\int_{\mathscr{I}} \xi(s) \widehat{\mathcal{P}} X_i(s) \, \mathrm{d}t \right) \\
&= \frac{1}{n} \sum_{i=1}^n \langle \xi, \widehat{\mathcal{P}} X_i \rangle^2 \\
&= \frac{1}{n} \sum_{i=1}^n \langle \widehat{\mathcal{P}} \xi, X_i \rangle^2 .
\end{aligned}$$

On ther other hand,

$$\begin{aligned}
\langle \mathcal{P}\xi, \widehat{\mathcal{C}} \mathcal{P} \xi \rangle &= \int_{\mathscr{I}} \mathcal{P}\xi(s) \left(\int_{\mathscr{I}} \frac{1}{n} \sum_{i=1}^n X_i(s) X_i(t) \mathcal{P}\xi(t) \, \mathrm{d}t \right) \mathrm{d}s \\
&= \frac{1}{n} \sum_{i=1}^n \langle \mathcal{P}\xi, X_i \rangle^2 .
\end{aligned}$$

By Lemma 6,

$$\sup_{\|\xi\|_\gamma = 1} |\langle \widehat{\mathcal{P}}\xi, X_i \rangle - \langle \mathcal{P}\xi, X_i \rangle| \le \sup_{\|\xi\|_\gamma = 1} \|X_i\| \|(\mathcal{P} - \widehat{\mathcal{P}})\xi\| = o(1) \|X_i\|.$$

Thus,

$$\sup_{\|\xi\|_\gamma = 1} |\langle \widehat{\mathcal{P}}\xi, X_i \rangle^2 - \langle \mathcal{P}\xi, X_i \rangle^2| = o(1) \|X_i\|^2 .$$

Therefore,

$$\langle \xi, \widehat{\mathcal{C}}_k \xi \rangle - \langle \mathcal{P}\xi, \widehat{\mathcal{C}} \mathcal{P}\xi \rangle = o(1) \frac{1}{n} \sum_{i=1}^n \|X_i\|^2 = o(1),$$

since by the strong law of large number, we have $\frac{1}{n} \sum \|X_i\|^2 \to \mathrm{E}\|X\|^2 < \infty$. Also, since $\|\mathcal{C} - \widehat{\mathcal{C}}\| = o(1)$, we have

$$\langle \xi, \widehat{\mathcal{C}}_k \xi \rangle - \langle \mathcal{P}\xi, \mathcal{C} \mathcal{P}\xi \rangle = \langle \xi, \widehat{\mathcal{C}}_k \xi \rangle - \langle \xi, \mathcal{C}_k \xi \rangle = o(1)$$

for all $\xi \in W_2^2$ such that $\|\xi\|_\gamma = 1$. □

of Theorem 4. The paradigm of our proof is almost identical to the one in Silverman (1996). Again, we use mathematical induction on the statement $\mathscr{H}_k'$ of statements:

$$\widehat{\lambda}_k \overset{a.s.}{\to} \lambda_k, \tag{S.43}$$

$$\|\widehat{\xi}_k\| \to 1; \tag{S.44}$$

and

$$\langle \widehat{\xi}_k, \xi_k \rangle^2 \overset{a.s.}{\to} 1. \tag{S.45}$$

Assume $\mathscr{H}_j'$ holds for $j = 1, 2, \ldots, k-1$ and we are going to prove $\mathscr{H}_k'$.

We first establish (S.43). On the one hand,

$$\lambda_k = \langle \xi_k, \mathcal{C}_k \xi_k \rangle \geq \langle \xi_k^*, \mathcal{C}_k \xi_k^* \rangle \geq \langle \widehat{\xi}_k, \mathcal{C}_k \widehat{\xi}_k \rangle \tag{S.46}$$

$$= \langle \widehat{\xi}_k, \widehat{\mathcal{C}}_k \widehat{\xi}_k \rangle + o(1) = \widehat{\lambda}_k + o(1) \tag{S.47}$$

because of Lemma and 8. On the other hand,

$$\widehat{\lambda}_k = \langle \widehat{\xi}_k, \widehat{\mathcal{C}}_k \widehat{\xi}_k \rangle - \rho_k S(\widehat{\xi}_k) + \rho_k S(\widehat{\xi}_k)$$

$$\geq \langle \frac{\xi_k}{\|\xi_k\|_\gamma}, \frac{\widehat{\mathcal{C}}_k \xi_k}{\|\xi_k\|_\gamma} \rangle - \rho_k S(\xi_k) + \rho_k S(\widehat{\xi}_k) \tag{S.48}$$

$$= \langle \frac{\xi_k}{\|\xi_k\|_\gamma}, \frac{\mathcal{C}_k \xi_k}{\|\xi_k\|_\gamma} \rangle + o(1) - \rho_k S(\xi_k) + \rho_k S(\widehat{\xi}_k) \tag{S.49}$$

$$= \frac{\lambda_k}{\|\xi_k\|_\gamma^2} + o(1) = \lambda_k + o(1), \tag{S.50}$$

where the inequality (S.48) comes from the fact that $\widehat{\xi}_k$ maximizes $\langle \xi, \widehat{\mathcal{C}}_k \xi \rangle - \rho_k S(\xi)$ among all $\xi \in W_2^2$ such that $\|\xi\|_\gamma = 1$ and the equality (S.49) is due to Lemma 8. From (S.47) and (S.50) we conclude that $\widehat{\lambda}_k - \lambda_k = o(1)$.

Second, we prove (S.44). The proof is actually identical to the one by Silverman (1996). We repeat it here for the sake of completeness. As all inequalities in (S.46) tend to equality, it follows

$$\|\widehat{\xi}_k\|^2 = \frac{\langle \widehat{\xi}_k, \mathcal{C}_k \widehat{\xi}_k \rangle}{\langle \xi_k^*, \mathcal{C}_k \xi_k^* \rangle} \overset{a.s.}{\to} 1,$$

so that $\|\widehat{\xi}_k\| \overset{a.s.}{\to} 1$.

Finally, we proceed to prove (S.45). First note that

$$\langle \xi_k^*, \mathcal{C}_k \xi_k^* \rangle = \sum_{j \geq k} \lambda_j \langle \xi_k^*, \xi_j \rangle^2.$$

Also, note that $\|\mathcal{P}\xi_k^*\| \leq \|\xi_k^*\| \leq 1$. Then

$$\lambda_k - \langle \xi_k^*, \mathcal{C}_k \xi_k^* \rangle \geq \lambda_k \|\mathcal{P}\xi_k^*\|^2 - \sum_{j \geq k} \lambda_j \langle \xi_k^*, \xi_j \rangle^2 = \sum_{j > k} (\lambda_k - \lambda_j) \langle \xi_k^*, \xi_j \rangle^2 \geq 0 \tag{S.51}$$

as $\lambda_k > \lambda_j$ for all $j > k$. On the other hand, by (S.46), (S.47) and (S.43),

$$\lambda_k - \langle \xi_k^*, \mathcal{C}_k \xi_k^* \rangle = o(1). \tag{S.52}$$

Thus,

$$\sum_{j > k} \langle \xi_k^*, \xi_j \rangle^2 = o(1) \tag{S.53}$$

since $\lambda_k - \lambda_j > 0$ for all $j > k$. By Lemma 7 we obtain $\|\mathcal{P}^\perp \xi_k^*\| \overset{a.s.}{\to} 0$. This result, with (S.53) together, gives the desired conclusion (S.45):

$$\langle \xi_k^*, \xi_k \rangle^2 = 1 - \|\mathcal{P}^\perp \xi_k^*\|^2 - \sum_{j > k} \langle \xi_k^*, \xi_j \rangle^2 \overset{a.s.}{\to} 1.$$

□

References

Kuczyński, J. and Woźniakowski, H. (1992). Estimating the largest eigenvalue by the power and lanczos algorithms with a random start. *SIAM Journal on Matrix Analysis and Applications* **13,** 1094–1122.

Leyk, Z. and Woźniakowski, H. (1996). Estimating a largest eigenvector by polynomial algorithms with a random start. Technical Report CUCS-023-96, Columbia University.

Qi, X. and Zhao, H. (2011). Some theoretical properties of Silverman's method for smoothed functional principal component analysis. *Journal of Multivariate Analysis* **102,** 741–767.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer Series in Statistics. Springer, 2nd edition.

Silverman, B. W. (1996). Smoothed functional principal components analysis by choice of norm. *The Annals of Statistics* **24,** 1–24.

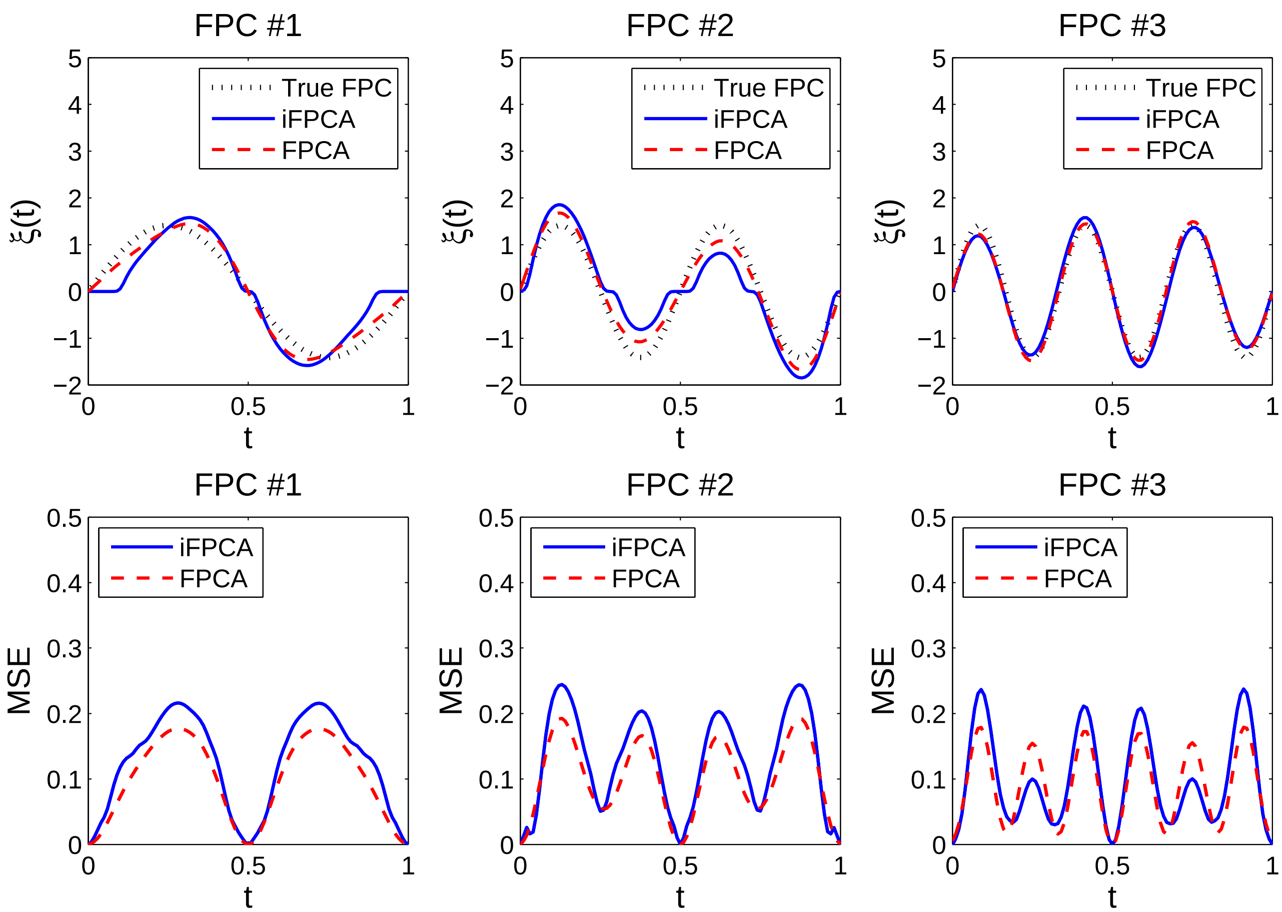

Figure S.3: The top three panels display the estimated FPCs using our iFPCA method and the classic regularized FPCA method in one random simulation replicate in Simulation 2. The bottom three panels show the pointwise mean squared errors (MSE) of the estimated FPCs using the two methods. This figure appears in color in the electronic version of this article.

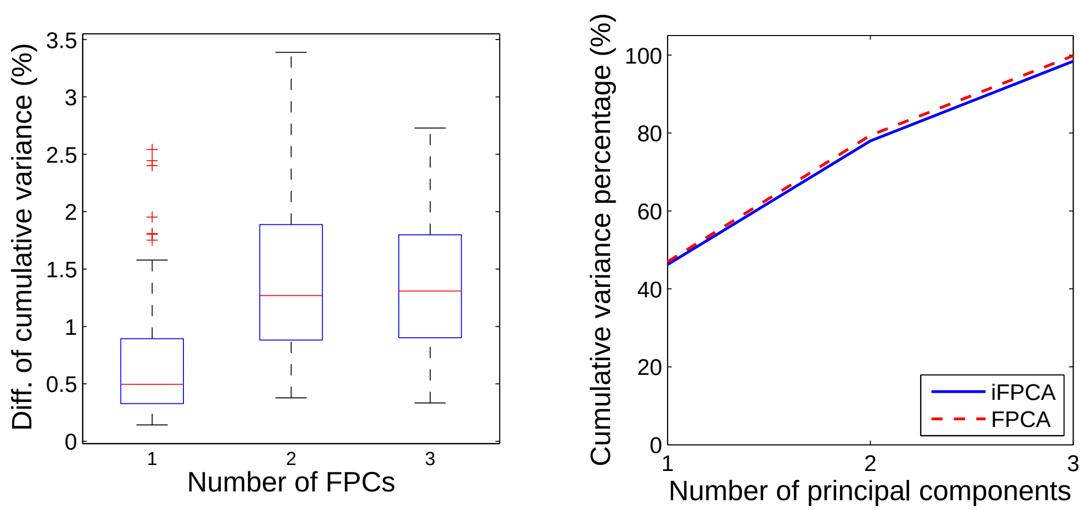

Figure S.4: The left panel shows the average cumulative percentages of variance explained by the estimated FPCs using our iFPCA method in comparison with the classic regularized FPCA method in 100 simulation replicates in Simulation 2. The right panel displays the boxplot of the loss of cumulative percentages of variance explained by the estimated FPCs using our iFPCA method compared with the classic regularized FPCA method in 100 simulation replicates. This figure appears in color in the electronic version of this article.