Intrinsic Correlation Analysis for Wasserstein Functional Data

Hang Zhou[1], Zhenhua Lin[2] and Fang Yao[3]

[1]*Department of Statistics, University of California, Davis, U.S.A.*

[2]*Department of Statistics and Applied Probability*

National University of Singapore, Singapore

[3]*School of Mathematical Sciences, Center for Statistical Science,*

Peking University, Beijing, China

Supplementary Material

S1 Proofs of main results

In this section, we provide the proofs of the main results, which contain the main arguments for this type of problems. The proofs of ancillary lemmas and propositions are deferred to Section S2.

Proof of Theorem 1. By the definition (3) and the equation (4),

$$\mathrm{Log}_{\mu(t)}\hat{\mu}(t) = F^{-1}_{\hat{\mu}(t)} \circ F_{\mu(t)} - \mathbf{id} = \frac{1}{n}\sum_{i=1}^{n} F^{-1}_{X_i(t)} \circ F_{\mu(t)} - \mathbf{id} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Log}_{\mu(t)}X_i(t).$$

By Proposition 3.2.14 in Panaretos and Zemel (2020), we have $\mathbf{E}\mathrm{Log}_{\mu(t)}X_i(t) = 0$. Given this, the part (a) is verified by applying the central limit theorem in Hilbert space (Aldous, 1976) that asserts convergence of the process $(\sqrt{n})^{-1}\sum_{i=1}^{\infty}\mathrm{Log}_{\mu(t)}X_i(t)$ to a Gaussian measure on tensor Hilbert space $\mathscr{T}(\mu)$ with covariance operator $\mathbf{C}'(\cdot) = \mathbf{E}(\langle\!\langle U, \cdot\rangle\!\rangle_\mu U)$ defined via the random element $U = \mathrm{Log}_{\mu(t)}X_i(t)$ in $\mathscr{T}(\mu)$. The first statement of the part (b) is a corollary of (a), while the second statement follows from the first one and the compactness of $\mathcal{T}$.

For assertion (c), note that $\mu(t)$ is atomless for each $t \in \mathcal{T}$, which implies that $\hat{\mu}(t)$ is also atomless (Chen et al., 2023). Then we have

$$
\begin{aligned}
\mathcal{P}_{\mathfrak{B}(\hat{\mu},\hat{\mu})}^{\mathfrak{B}(\mu,\mu)}\hat{\mathbf{C}} - \mathbf{C} =& \frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{Log}_\mu X_i\right)\otimes\left(\mathrm{Log}_\mu X_i\right) - \mathbf{C} + \frac{1}{n}\sum_{i=1}^{n}\left(\mathcal{P}_{\hat{\mu}}^{\mu}\mathrm{Log}_{\hat{\mu}}X_i - \mathrm{Log}_\mu X_i\right)\otimes\left(\mathrm{Log}_\mu X_i\right) \\
&+ \frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{Log}_\mu X_i\right)\otimes\left(\mathcal{P}_{\hat{\mu}}^{\mu}\mathrm{Log}_{\hat{\mu}}X_i - \mathrm{Log}_\mu X_i\right) \\
&+ \frac{1}{n}\sum_{i=1}^{n}\left(\mathcal{P}_{\hat{\mu}}^{\mu}\mathrm{Log}_{\hat{\mu}}X_i - \mathrm{Log}_\mu X_i\right)\otimes\left(\mathcal{P}_{\hat{\mu}}^{\mu}\mathrm{Log}_{\hat{\mu}}X_i - \mathrm{Log}_\mu X_i\right) \\
:=& A_1 + A_2 + A_3 + A_4.
\end{aligned}
$$

For A_2, we further have

$$\begin{aligned}|||A_2|||^2_{\mathfrak{B}(\mu,\mu)} \lesssim & \frac{1}{n^2}\sum_{i_1=1}^n\sum_{i_2=1}^n(\|\mathrm{Log}_\mu X_{i_1}\|^2_\mu + \|\mathrm{Log}_\mu X_{i_2}\|^2_\mu) \\ & \times(\|\mathcal{P}^\mu_{\hat\mu}\mathrm{Log}_{\hat\mu}X_{i_1} - \mathrm{Log}_\mu X_{i_1}\|^2_\mu + \|\mathcal{P}^\mu_{\hat\mu}\mathrm{Log}_{\hat\mu}X_{i_2} - \mathrm{Log}_\mu X_{i_2}\|^2_\mu) \\ = & \frac{2}{n}\sum_{i=1}^n \|\mathrm{Log}_\mu X_i\|^2_\mu \|\mathcal{P}^\mu_{\hat\mu}\mathrm{Log}_{\hat\mu}X_i - \mathrm{Log}_\mu X_i\|^2_\mu \\ & + \frac{2}{n^2}\sum_{i_1=1}^n\sum_{i_2=1}^n \|\mathrm{Log}_\mu X_{i_1}\|^2_\mu \|\mathcal{P}^\mu_{\hat\mu}\mathrm{Log}_{\hat\mu}X_{i_2} - \mathrm{Log}_\mu X_{i_2}\|^2_\mu. \end{aligned} \tag{S1.1}$$

For the first term on the right hand side of equation (S1.1),

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^n \|\mathrm{Log}_\mu X_i\|^2_\mu \|\mathcal{P}^\mu_{\hat\mu}\mathrm{Log}_{\hat\mu}X_i - \mathrm{Log}_\mu X_i\|^2_\mu \\ = & \frac{1}{n}\sum_{i=1}^n \|\mathrm{Log}_\mu X_i\|^2_\mu \\ & \times \int \left\langle \mathcal{P}^{\mu(t)}_{\hat\mu(t)}\mathrm{Log}_{\hat\mu(t)}X_i(t) - \mathrm{Log}_{\mu(t)}X_i(t), \mathcal{P}^{\mu(t)}_{\hat\mu(t)}\mathrm{Log}_{\hat\mu(t)}X_i(t) - \mathrm{Log}_{\mu(t)}X_i(t)\right\rangle_{\mu(t)} \mathrm{d}t \\ = & \iint \left(F^{-1}_{\hat\mu(t)} \circ F_{\mu(t)}(u) - u\right)^2 \mathrm{d}F_{\mu(t)}(u)\mathrm{d}t \times \frac{1}{n}\sum_{i=1}^n \|\mathrm{Log}_\mu X_i\|^2_\mu \\ = & \int_{\mathcal{T}} d^2(\hat\mu(t),\mu(t))\mathrm{d}t \times \frac{1}{n}\sum_{i=1}^n \|\mathrm{Log}_\mu X_i\|^2_\mu = O_p(n^{-1}), \end{aligned}$$

where the second equality follows from $\mathcal{P}^{\mu(t)}_{\hat\mu(t)}\mathrm{Log}_{\hat\mu(t)}X_i(t) - \mathrm{Log}_{\mu(t)}X_i(t) = F^{-1}_{\hat\mu(t)} \circ F_{\mu(t)} - \mathbf{id}$, and the last equality is based on the part (b) and $n^{-1}\sum_{i=1}^n \|\mathrm{Log}_\mu X_i\|^2_\mu = O_p(1)$ by the law of large numbers. A similar argument shows that the second term in (S1.1) is of order $O_p(n^{-2})$. Thus $|||A_2|||^2_{\mathfrak{B}(\mu,\mu)} = O_p(n^{-1})$. Analogous calculation shows that $|||A_3|||^2_{\mathfrak{B}(\mu,\mu)} = O_p(n^{-1})$ and $|||A_4|||^2_{\mathfrak{B}(\mu,\mu)} = O_p(n^{-2})$. According to Dauxois et al. (1982), $|||n^{-1}\sum_{i=1}^n \left(\mathrm{Log}_\mu X_i\right) \otimes \left(\mathrm{Log}_\mu X_i\right) - \mathbf{C}|||^2_{\mathfrak{B}(\mu,\mu)} = O_p(n^{-1})$, and consequently,

$\left|\left|\left|\mathcal{P}_{\mathfrak{B}(\hat\mu,\hat\mu)}^{\mathfrak{B}(\mu,\mu)}\hat{\mathbf{C}}-\mathbf{C}\right|\right|\right|_{\mathfrak{B}(\mu,\mu)}^{2}=O_p(n^{-1})$. The result for $\hat\lambda_k$ follows from the perturbation argument in Bosq (2000).

For the part (e), we first note that

$$\begin{aligned}\mathcal{P}_{\mathfrak{B}(\hat\mu,\hat\mu)}^{\mathfrak{B}(\mu,\mu)}\hat{\mathbf{C}}=&\frac{1}{n}\sum_{i=1}^{n}\left(\mathcal{P}_{\hat\mu}^{\mu}\mathrm{Log}_{\hat\mu}X_i\right)\otimes\left(\mathcal{P}_{\hat\mu}^{\mu}\mathrm{Log}_{\hat\mu}X_i\right)\\ =&\frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{Log}_{\mu}X_i-\overline{\mathrm{Log}_{\mu}X}\right)\otimes\left(\mathrm{Log}_{\mu}X_i-\overline{\mathrm{Log}_{\mu}X}\right),\end{aligned}$$

where $\overline{\mathrm{Log}_{\mu}X}=n^{-1}\sum_{i=1}^{n}\mathrm{Log}_{\mu}X_i$.

By the proof of Theorem 5.1.18 in Hsing and Eubank (2015), for all $\{j:|||\Delta|||_{\mathfrak{B}(\mu,\mu)}\leqslant\eta_j/2\}$, we have the following expansion,

$$\begin{aligned}\mathcal{P}_{\hat\mu}^{\mu}\hat\Phi_j-\Phi_j=&\sum_{k\neq j}\frac{\langle\!\langle\Delta\Phi_j,\Phi_k\rangle\!\rangle_{\mu}}{(\lambda_j-\lambda_k)}\Phi_k+\sum_{k\neq j}\frac{\langle\!\langle\Delta(\mathcal{P}_{\hat\mu}^{\mu}\hat\Phi_j-\Phi_j),\Phi_k\rangle\!\rangle_{\mu}}{(\lambda_j-\lambda_k)}\Phi_k\\ &+\sum_{k\neq j}\sum_{s=1}^{\infty}\frac{(\lambda_j-\hat\lambda_j)^s}{(\lambda_j-\lambda_k)^{s+1}}\langle\!\langle\Delta\mathcal{P}_{\hat\mu}^{\mu}\hat\Phi_j,\Phi_k\rangle\!\rangle_{\mu}\Phi_k+\langle\!\langle(\mathcal{P}_{\hat\mu}^{\mu}\hat\Phi_j-\Phi_j),\Phi_j\rangle\!\rangle_{\mu}\Phi_j,\end{aligned}$$

where $\Delta=\mathcal{P}_{\mathfrak{B}(\hat\mu,\hat\mu)}^{\mathfrak{B}(\mu,\mu)}\hat{\mathbf{C}}-\mathbf{C}$. For all $k\neq j$, we have

$$\begin{aligned}\mathbf{E}\langle\!\langle\Delta\Phi_j,\Phi_k\rangle\!\rangle_{\mu}^2=&\mathbf{E}\left\langle\!\!\left\langle\mathcal{P}_{\mathfrak{B}(\hat\mu,\hat\mu)}^{\mathfrak{B}(\mu,\mu)}\hat{\mathbf{C}}\Phi_j,\Phi_k\right\rangle\!\!\right\rangle_{\mu}^2\\ =&\mathbf{E}\left\{\frac{1}{n}\sum_{i=1}^{n}\langle\!\langle\mathrm{Log}_{\mu}X_i-\overline{\mathrm{Log}_{\mu}X_i},\Phi_j\rangle\!\rangle_{\mu}\langle\!\langle\mathrm{Log}_{\mu}X_i-\overline{\mathrm{Log}_{\mu}X_i},\Phi_k\rangle\!\rangle_{\mu}\right\}^2\\ =&\mathbf{E}\left\{\frac{1}{n}\sum_{i=1}^{n}(\xi_{i,k}-\overline{\xi}_k)(\xi_{i,j}-\overline{\xi}_j)\right\}^2\\ \leqslant&C\frac{\mathbb{E}\xi_{ik}^2\xi_{ij}^2}{n}\leqslant C\frac{\sqrt{\mathbb{E}\xi_{ik}^4\mathbb{E}\xi_{ij}^4}}{n}=C\frac{\lambda_k\lambda_j}{n},\end{aligned}\tag{S1.2}$$

where $\overline{\xi}_k=n^{-1}\sum_{i=1}^{n}\xi_{i,k}$ and $\xi_{i,k}$ is the k-th component score of $\mathrm{Log}_{\mu}X_i$. By equation

(S1.2) and Lemma 7 in Dou et al. (2012),

$$\mathbf{E}\left|\left|\left|\sum_{k\neq j}\frac{\langle\!\langle \Delta\Phi_j,\Phi_k\rangle\!\rangle_\mu}{(\lambda_j-\lambda_k)}\Phi_k\right|\right|\right|_\mu^2=\sum_{k\neq j}\frac{\mathbf{E}\langle\!\langle \Delta\Phi_j,\Phi_k\rangle\!\rangle_\mu^2}{(\lambda_j-\lambda_k)^2}\leqslant C\frac{1}{n}\sum_{k\neq j}\frac{\lambda_j\lambda_k}{(\lambda_j-\lambda_k)^2}\leqslant C\frac{j^2}{n},\quad \text{(S1.3)}$$

where C is a constant does not depend on j. From Bessel's inequality and given that $|||\Delta|||_{\mathfrak{B}(\mu,\mu)}<\eta_j/2$,

$$\mathbf{E}\left|\left|\left|\sum_{k\neq j}\frac{\langle\!\langle \Delta(\mathcal{P}_{\hat\mu}^{\mu}\hat\Phi_j-\Phi_j),\Phi_k\rangle\!\rangle_\mu}{(\lambda_j-\lambda_k)}\Phi_k\right|\right|\right|_\mu^2\leqslant\mathbf{E}\frac{|||\Delta|||_{\mathfrak{B}(\mu,\mu)}^2\left|\left|\left|\mathcal{P}_{\hat\mu}^{\mu}\hat\Phi_j-\Phi_j\right|\right|\right|_\mu^2}{(2\eta_j)^2}<\frac{1}{16}\mathbf{E}\|\mathcal{P}_{\hat\mu}^{\mu}\hat\Phi_j-\Phi_j\|_\mu^2. \tag{S1.4}$$

Similarly,

$$\begin{aligned}
&\mathbb{E}\left|\left|\left|\sum_{k\neq j}\sum_{s=1}^{\infty}\frac{(\lambda_j-\hat\lambda_j)^s}{(\lambda_j-\lambda_k)^{s+1}}\langle\!\langle \Delta\mathcal{P}_{\hat\mu}^{\mu}\hat\Phi_j,\Phi_k\rangle\!\rangle_\mu\Phi_k\right|\right|\right|_\mu^2=\mathbb{E}\sum_{k\neq j}\left\{\sum_{s=1}^{\infty}\frac{(\lambda_j-\hat\lambda_j)^s}{(\lambda_j-\lambda_k)^{s+1}}\right\}^2\langle\!\langle \Delta\mathcal{P}_{\hat\mu}^{\mu}\hat\Phi_j,\Phi_k\rangle\!\rangle_\mu^2\\
\leqslant&\mathbb{E}\frac{|||\Delta|||_{\mathfrak{B}(\mu,\mu)}^2}{(2\eta_j-|||\Delta|||_{\mathfrak{B}(\mu,\mu)})^2}\left\{2\sum_{k\neq j}\frac{\langle\!\langle \Delta\mathcal{P}_{\hat\mu}^{\mu}\Phi_j,\Phi_k\rangle\!\rangle_\mu^2}{(\lambda_j-\lambda_k)^2}+2\sum_{k\neq j}\frac{\langle\!\langle \Delta\mathcal{P}_{\hat\mu}^{\mu}(\hat\Phi_j-\Phi_j),\Phi_k\rangle\!\rangle_\mu^2}{(\lambda_j-\lambda_k)^2}\right\}\\
\leqslant&\frac{8}{9}\mathbb{E}\left[\frac{|||\Delta|||_{\mathfrak{B}(\mu,\mu)}^2}{\eta_j^2}\sum_{k\neq j}\frac{\langle\!\langle \Delta\mathcal{P}_{\hat\mu}^{\mu}\Phi_j,\Phi_k\rangle\!\rangle_\mu^2}{(\lambda_j-\lambda_k)^2}+\frac{|||\Delta|||_{\mathfrak{B}(\mu,\mu)}^4}{\eta_j^4}\|\mathcal{P}_{\hat\mu}^{\mu}\hat\Phi_j-\Phi_j\|_\mu^2\right]\\
\leqslant&\frac{2}{9}\mathbb{E}\sum_{k\neq j}\frac{\langle\!\langle \Delta\mathcal{P}_{\hat\mu}^{\mu}\Phi_j,\Phi_k\rangle\!\rangle_\mu^2}{(\lambda_j-\lambda_k)^2}+\frac{1}{18}\mathbb{E}\left|\left|\left|\mathcal{P}_{\hat\mu}^{\mu}\hat\Phi_j-\Phi_j\right|\right|\right|_\mu^2.
\end{aligned} \tag{S1.5}$$

Combing equation (S1.2) to (S1.5), the proof is completed by the fact $\left|\left|\left|\langle\!\langle(\mathcal{P}_{\hat\mu}^{\mu}\hat\Phi_j-\Phi_j),\Phi_j\rangle\!\rangle_\mu\Phi_j\right|\right|\right|_\mu = 1/2\left|\left|\left|\mathcal{P}_{\hat\mu}^{\mu}\hat\Phi_j-\Phi_j\right|\right|\right|_\mu^2$ and Cauchy-Schwarz inequality. □

Proof of Lemma 1. By Taylor expansion, for any real numbers x and x_0,

$$x^{-1}-x_0^{-1}=-x_0^{-2}\left(x-x_0\right)+\hat x^{-3}\left(x-x_0\right)^2,$$

where $\hat{x}$ is some value between x and x_0. Using this fact, we obtain

$$\begin{aligned}\left|\hat{\lambda}_{X,j}^{-1} - \lambda_{X,j}^{-1}\right| =& \lambda_{X,j}^{-2} O_p\left(\left|\hat{\lambda}_{X,j} - \lambda_{X,j}\right|\right) + \left(\lambda_{X,j} - O_p\left(|||\Delta_X|||\right)\right)^{-3} O_p\left(\left|\hat{\lambda}_{X,j} - \lambda_{X,j}\right|^2\right) \\ =& O_p(j^{2a_X}/\sqrt{n}) + O_p(j^{3a_X}/n),\end{aligned}$$

where $\Delta_X = \mathcal{P}_{\mathfrak{B}(\hat{\mu}_X,\hat{\mu}_X)}^{\mathfrak{B}(\mu_X,\mu_X)} \hat{\mathbf{C}}_X - \mathbf{C}_X$ and the last equality follows from the part (d) of Theorem 1. Under the condition Assumption (A.2), we have $j^{3a_X}/n = o(j^{2a_X}/\sqrt{n})$ and the first assertion follows. The second assertion is proved by a similar argument.

□

Proof of Theorem 2. We first introduce two lemmas and theirs proofs can be found in the supplementary material.

Lemma 1. *Under the assumption A.2, we have*

$$\sup_{j \leqslant k_X} j^{-2a_X} \left|\hat{\lambda}_{X,j}^{-1} - \lambda_{X,j}^{-1}\right| = O_p(1/\sqrt{n}) \text{ and } \sup_{j \leqslant k_Y} j^{-2a_Y} \left|\hat{\lambda}_{Y,j}^{-1} - \lambda_{Y,j}^{-1}\right| = O_p(1/\sqrt{n}).$$

Lemma 2. *For* $\eta_{X,k} = \inf_{j \neq k} |\lambda_{X,k} - \lambda_{X,j}|$ *and* $\eta_{Y,k} = \inf_{j \neq k} |\lambda_{Y,k} - \lambda_{Y,j}|$*, under the assumption B.1, one has*

$$\mathbf{P}\left(\frac{1}{2}\eta_{X,k_X} > \left|\left|\left|\mathcal{P}\hat{\mathbf{C}}_X - \mathbf{C}_X\right|\right|\right|\right) \to 1 \qquad \text{and} \qquad \mathbf{P}\left(\frac{1}{2}\eta_{Y,k_Y} > \left|\left|\left|\mathcal{P}\hat{\mathbf{C}}_Y - \mathbf{C}_Y\right|\right|\right|\right) \to 1.$$

To reduce notational burden, we shall suppress the subscripts and superscripts from $\mathcal{P}_{\hat{\mu}_X}^{\mu_X}$ in the sequel. We first show that $\left|\left|\left|(\mathcal{P}\hat{\mathbf{C}}_{Y,k_Y}^{-1})(\mathcal{P}\hat{\mathbf{C}}_{YX}) - \mathbf{C}_Y^{-1}\mathbf{C}_{YX}\right|\right|\right|_{\mathfrak{B}(\mu_X,\mu_Y)}^2 = O_p(n^{(1-2b_Y)/(a_Y+2b_Y)})$, where $\mathbf{C}_{Y,k_Y}^{-1} = \sum_{j=1}^{k_Y} \lambda_{Y,j}^{-1}\Phi_{Y,j}$. Given $h = \sum_{j=1}^{\infty} h_j \Phi_{X,j} \in$

$\mathscr{T}(\mu_X)$ with $\|h\|_{\mu_X} = 1$, we have

$$
\begin{aligned}
(\mathcal{P}\hat{\mathbf{C}}_{Y,k_Y}^{-1})(\mathcal{P}\hat{\mathbf{C}}_{YX})h - \mathbf{C}_Y^{-1}\mathbf{C}_{YX}h =& (\mathcal{P}\hat{\mathbf{C}}_{Y,k_Y}^{-1})(\mathcal{P}\hat{\mathbf{C}}_{YX})h - \mathbf{C}_{Y,k_Y}^{-1}(\mathcal{P}\hat{\mathbf{C}}_{YX})h \\
&+ \mathbf{C}_{Y,k_Y}^{-1}(\mathcal{P}\hat{\mathbf{C}}_{YX})h - \mathbf{C}_{Y,k_Y}^{-1}\mathbf{C}_{YX}h + \mathbf{C}_{Y,k_Y}^{-1}\mathbf{C}_{YX}h - \mathbf{C}_Y^{-1}\mathbf{C}_{YX}h := J_1 + J_2 + J_3.
\end{aligned}
\tag{S1.6}
$$

For the first term in (S1.6), we further have

$$
\begin{aligned}
J_1 =& \sum_{j_2=1}^{k_Y} \lambda_{Y,j_2}^{-1} \left[\sum_{j_1=1}^{\infty} h_{j_1} \left\{ \gamma_{j_1 j_2} - \frac{1}{n}\sum_{i=1}^{n} (\xi_{ij_1} - \overline{\xi}_{j_1})(\eta_{ij_2} - \overline{\eta}_{j_2}) \right\} \right] \Phi_{Y,j_2} \\
&+ \sum_{j_2=1}^{k_Y} \hat{\lambda}_{Y,j_2}^{-1} \left(\sum_{j_1=1}^{\infty} \gamma_{j_1 j_2} h_{j_1} \right) \mathcal{P}\hat{\Phi}_{Y,j_2} - \sum_{j_2=1}^{k_Y} \lambda_{Y,j_2}^{-1} \left(\sum_{j_1=1}^{\infty} \gamma_{j_1 j_2} h_{j_1} \right) \Phi_{Y,j_2} \\
&+ \sum_{j_2=1}^{k_Y} \hat{\lambda}_{Y,j_2}^{-1} \left[\sum_{j_1=1}^{\infty} h_{j_1} \left\{ \frac{1}{n}\sum_{i=1}^{n} (\xi_{ij_1} - \overline{\xi}_{j_1})(\eta_{ij_2} - \overline{\eta}_{j_2}) - \gamma_{j_1 j_2} \right\} \right] \mathcal{P}\hat{\Phi}_{Y,j_2} \\
:=& J_{11} + J_{12} + J_{13},
\end{aligned}
\tag{S1.7}
$$

where $\overline{\xi}_{j_1} = n^{-1}\sum_{i=1}^{n} \xi_{ij_1}$ and $\overline{\eta}_{j_2} = n^{-1}\sum_{i=1}^{n} \eta_{ij_2}$. For J_{11}, it is of order

$$
\begin{aligned}
\mathbf{E}\|J_{11}\|_{\mu_Y}^2 =& \sum_{j_2=1}^{k_Y} \lambda_{Y,j_2}^{-2} \mathbf{E} \left[\sum_{j_1=1}^{\infty} h_{j_1} \left\{ \gamma_{j_1 j_2} - \frac{1}{n}\sum_{i=1}^{n} (\xi_{ij_1} - \overline{\xi}_{j_1})(\eta_{ij_2} - \overline{\eta}_{j_2}) \right\} \right]^2 \\
\leqslant& \sum_{j_2=1}^{k_Y} \lambda_{Y,j_2}^{-2} \sum_{j_1=1}^{\infty} \mathbf{E} \left\{ \gamma_{j_1 j_2} - \frac{1}{n}\sum_{i=1}^{n} (\xi_{ij_1} - \overline{\xi}_{j_1})(\eta_{ij_2} - \overline{\eta}_{j_2}) \right\}^2 \\
\leqslant& \sum_{j_2=1}^{k_Y} \lambda_{Y,j_2}^{-2} \sum_{j_1=1}^{\infty} \frac{C\lambda_{X,j_1}\lambda_{Y,j_2}}{n} \\
=& \frac{C}{n} \sum_{j_2=1}^{k_Y} \lambda_{Y,j_2}^{-1} \sum_{j_1=1}^{\infty} \lambda_{X,j_1} = O_p\left(\frac{k_Y^{a_Y+1}}{n} \right).
\end{aligned}
\tag{S1.8}
$$

By condition A.2 and Cauchy–Schwarz inequality together, the second inequality in the above follows from

$$\mathbf{E}\left\{\frac{1}{n}\sum_{i=1}^{n}(\xi_{ij_1}-\overline{\xi}_{j_1})(\eta_{ij_2}-\overline{\eta}_{j_2})-\gamma_{j_1j_2}\right\}^2 \leqslant C\frac{\mathbf{E}(\xi_{ij_1}\eta_{ij_2})^2}{n} \leqslant \frac{C}{n}\sqrt{\mathbb{E}\xi_{ij_1}^4\eta_{ij_2}^4} \leqslant \frac{C}{n}\lambda_{X,j_1}\lambda_{Y,j_2}.$$

For J_{13}, we have

$$\begin{aligned}
\mathbf{E}\|J_{13}\|_{\mu_Y}^2 =& \mathbf{E}\sum_{j_2=1}^{k_Y}\hat{\lambda}_{Y,j_2}^{-2}\sum_{j_1=1}^{\infty}\left\{\frac{1}{n}\sum_{i=1}^{n}(\xi_{ij_1}-\overline{\xi}_{j_1})(\eta_{ij_2}-\overline{\eta}_{j_2})-\gamma_{j_1j_2}\right\}^2 \\
\leqslant& 4\sum_{j_2=1}^{k_Y}\lambda_{Y,j_2}^{-2}\sum_{j_1=1}^{\infty}\mathbf{E}\left\{\frac{1}{n}\sum_{i=1}^{n}(\xi_{ij_1}-\overline{\xi}_{j_1})(\eta_{ij_2}-\overline{\eta}_{j_2})-\gamma_{j_1j_2}\right\}^2 \\
\leqslant& 4C\sum_{j_2=1}^{k_Y}\lambda_{Y,j_2}^{-2}\sum_{j_1=1}^{\infty}\frac{\lambda_{X,j_1}\lambda_{Y,j_2}}{n} \\
=& \frac{4C}{n}\sum_{j_2=1}^{k_Y}\lambda_{Y,j_2}^{-1}\sum_{j_1=1}^{\infty}\lambda_{X,j_1} = O_p\left(\frac{k_Y^{a_Y+1}}{n}\right),
\end{aligned} \tag{S1.9}$$

where the first inequality comes from Lemma 1. For J_{12}, we divide it into two components by

$$\begin{aligned}
J_{12} =& \sum_{j_2=1}^{k_Y}\hat{\lambda}_{Y,j_2}^{-1}\left(\sum_{j_1=1}^{\infty}\gamma_{j_1j_2}h_{j_1}\right)\mathcal{P}\hat{\Phi}_{Y,j_2} - \sum_{j_2=1}^{k_Y}\lambda_{Y,j_2}^{-1}\left(\sum_{j_1=1}^{\infty}\gamma_{j_1j_2}h_{j_1}\right)\Phi_{Y,j_2} \\
=& \sum_{j_2=1}^{k_Y}\hat{\lambda}_{Y,j_2}^{-1}\left(\sum_{j_1=1}^{\infty}\gamma_{j_1j_2}h_{j_1}\right)\mathcal{P}\hat{\Phi}_{Y,j_2} - \sum_{j_2=1}^{k_Y}\lambda_{Y,j_2}^{-1}\left(\sum_{j_1=1}^{\infty}\gamma_{j_1j_2}h_{j_1}\right)\mathcal{P}\hat{\Phi}_{Y,j_2} \\
&+ \sum_{j_2=1}^{k_Y}\lambda_{Y,j_2}^{-1}\left(\sum_{j_1=1}^{\infty}\gamma_{j_1j_2}h_{j_1}\right)\left(\mathcal{P}\hat{\Phi}_{Y,j_2}-\Phi_{Y,j_2}\right) \\
:=& J_{121}+J_{122}.
\end{aligned}$$

For the first component,

$$\|J_{121}\|_{\mu_Y}^2 = \sum_{j_2=1}^{k_Y} \left(\hat{\lambda}_{Y,j_2}^{-1} - \lambda_{Y,j_2}^{-1}\right)^2 \left(\sum_{j_1=1}^{\infty} h_{j_1}\gamma_{j_1j_2}\right)^2 \leqslant \sum_{j_2=1}^{k_Y} \left(\hat{\lambda}_{Y,j_2}^{-1} - \lambda_{Y,j_2}^{-1}\right)^2 \sum_{j_1=1}^{\infty} \gamma_{j_1j_2}^2$$
$$\leqslant \sup_{j_2 \leqslant k_Y} \left(\hat{\lambda}_{Y,j_2}^{-1} - \lambda_{Y,j_2}^{-1}\right)^2 \sum_{j_2=1}^{k_Y} j^{-2a_Y-2b_Y} = O_p\left(\frac{k_Y^{4a_Y}}{n}\right) O(k_Y^{1-2a_Y-2b_Y}) = o_p\left(\frac{k^{a_Y+1}}{n}\right),$$

where the last inequality is due to the second assertion in Lemma 1 and the assumption B.1. For the second component, we have

$$\begin{aligned}\mathbf{E}\|J_{122}\|_{\mu_Y}^2 =& \mathbf{E}\left\|\sum_{j_2=1}^{k_Y} \lambda_{Y,j_2}^{-1}\left(\sum_{j_1=1}^{\infty}\gamma_{j_1j_2}h_{j_1}\right)\left(\mathcal{P}\hat{\Phi}_{Y,j_2} - \Phi_{Y,j_2}\right)\right\|_{\mu_Y}^2 \\ \leqslant& k_Y \sum_{j_2=1}^{k_Y} \lambda_{Y,j_2}^{-2}\left(\sum_{j_1=1}^{\infty}\gamma_{j_1j_2}h_{j_1}\right)^2 \mathbf{E}\left\|\mathcal{P}\hat{\Phi}_{Y,j_2} - \Phi_{Y,j_2}\right\|_{\mu_Y}^2 \\ \leqslant& C\frac{k_Y}{n}\sum_{j_2=1}^{k_Y}\lambda_{Y,j_2}^{-2} j_2^2 \sum_{j_1=1}^{\infty}\gamma_{j_1j_2}^2 = O\left(\frac{k_Y^{4-2b_Y}}{n}\right) = o\left(\frac{k_Y^{a_Y+1}}{n}\right),\end{aligned}$$

where the last inequality follows from the part (d) of Theorem 1, Lemma 2 and $b_Y > a_Y/2+1$. Together with the first component, this shows that $\|J_{12}\|_{\mu_Y}^2 = o_p(k^{a_Y+1}/n)$. Combing this with (S1.7) to (S1.9), we deduce that $\|J_1\|_{\mu_Y}^2 = O_p\left(k^{(a_Y+1)}/n\right) = O_p\left(n^{-(2b_Y-1)/(a_Y+2b_Y)}\right)$.

Note that $\|J_2\|_{\mu_Y}^2 = \|J_{11}\|_{\mu_Y}^2 = O_p\left(n^{-(2b_Y-1)/(a_Y+2b_Y)}\right)$ and for J_3,

$$\begin{aligned}\|J_3\|_{\mu_Y}^2 =& \sum_{j_2=k_Y+1}^{\infty} \lambda_{Y,j_2}^{-2}\left(\sum_{j_1=1}^{\infty} h_{j_1}\gamma_{j_1j_2}\right)^2 \leqslant \sum_{j_2=k_Y+1}^{\infty}\lambda_{Y,j_2}^{-2}\sum_{j_1=1}^{\infty}\gamma_{j_1j_2}^2 \\ \leqslant& \sum_{j_2=k_Y+1}^{\infty} j_2^{2a_Y} j_2^{-2a_Y-2b_Y} = O\left(k_Y^{1-2b_Y}\right) = O\left(n^{-\frac{2b_Y-1}{a_Y+2b_Y}}\right).\end{aligned} \tag{S1.10}$$

Consequently, $\left|\left|\left|(\mathcal{P}\hat{\mathbf{C}}_{Y,k_Y}^{-1})(\mathcal{P}\hat{\mathbf{C}}_{YX}) - \mathbf{C}_Y^{-1}\mathbf{C}_{YX}\right|\right|\right|_{\mathfrak{B}(\mu_X,\mu_Y)}^2 = O_p(n^{-(2b_Y-1)/(a_Y+2b_Y)})$ by (S1.6). Similarly, we can show $\left|\left|\left|(\mathcal{P}\hat{\mathbf{C}}_{X,k_X}^{-1})(\mathcal{P}\hat{\mathbf{C}}_{XY}) - \mathbf{C}_X^{-1}\mathbf{C}_{XY}\right|\right|\right|_{\mathfrak{B}(\mu_Y,\mu_X)}^2 = O_p(n^{-(2b_X-1)/(a_X+2b_X)})$.

By the part (f) of Proposition 2 and the fact that $\mathcal{P}\hat{\mathbf{C}}_{X,k_X}^{-1}\mathcal{P}\hat{\mathbf{C}}_{XY} = O_p(1)$, we further deduce that

$$\begin{aligned}
&\left|\left|\left|\mathbf{C}_X^{-1}\mathbf{C}_{XY}\mathbf{C}_Y^{-1}\mathbf{C}_{YX} - \mathcal{P}\hat{\mathbf{C}}_{X,k_X}^{-1}\mathcal{P}\hat{\mathbf{C}}_{XY}\mathcal{P}\hat{\mathbf{C}}_{Y,k_Y}^{-1}\mathcal{P}\hat{\mathbf{C}}_{YX}\right|\right|\right|_{\mathfrak{B}(\mu_X,\mu_X)}^2 \\
\leq & 2\left|\left|\left|\mathbf{C}_X^{-1}\mathbf{C}_{XY} - \mathcal{P}\hat{\mathbf{C}}_{X,k_X}^{-1}\mathcal{P}\hat{\mathbf{C}}_{XY}\right|\right|\right|_{\mathfrak{B}(\mu_Y,\mu_X)}^2 \left|\left|\left|\mathbf{C}_Y^{-1}\mathbf{C}_{YX}\right|\right|\right|_{\mathfrak{B}(\mu_X,\mu_Y)}^2 \\
&+ 2\left|\left|\left|\mathcal{P}\hat{\mathbf{C}}_{X,k_X}^{-1}\mathcal{P}\hat{\mathbf{C}}_{XY}\right|\right|\right|_{\mathfrak{B}(\mu_Y,\mu_X)}^2 \left|\left|\left|\mathbf{C}_Y^{-1}\mathbf{C}_{YX} - \mathcal{P}\hat{\mathbf{C}}_{Y,k_Y}^{-1}\mathcal{P}\hat{\mathbf{C}}_{YX}\right|\right|\right|_{\mathfrak{B}(\mu_X,\mu_Y)}^2 \\
= & O_p\left(\max\left\{n^{-\frac{2b_X-1}{a_X+2b_X}}, n^{-\frac{2b_Y-1}{a_Y+2b_Y}}\right\}\right).
\end{aligned}$$

Now we adopt the perturbation argument in Bosq (2000) to establish

$$\|\mathcal{P}\hat{U} - U\|_{\mu_X}^2 = O_p\left(\max\left\{n^{-(2b_X-1)/(a_X+2b_X)}, n^{-(2b_Y-1)/(a_Y+2b_Y)}\right\}\right),$$

and complete the proof by

$$\begin{aligned}
&\|\mathcal{P}\hat{\mathbf{C}}_{Y,k_Y}^{-1}\hat{\mathbf{C}}_{YX}\mathcal{P}\hat{U} - \mathbf{C}_Y^{-1}\mathbf{C}_{YX}U\|_{\mu_Y}^2 \\
\leqslant & 2\left|\left|\left|(\mathcal{P}\hat{\mathbf{C}}_{Y,k_Y}^{-1})(\mathcal{P}\hat{\mathbf{C}}_{YX}) - \mathbf{C}_Y^{-1}\mathbf{C}_{YX}\right|\right|\right|_{\mathfrak{B}(\mu_X,\mu_Y)}^2 \|\mathcal{P}\hat{U}\|_{\mu_X}^2 + 2\left|\left|\left|\mathbf{C}_Y^{-1}\mathbf{C}_{YX}\right|\right|\right|_{\mathfrak{B}(\mu_X,\mu_Y)}^2 \|\mathcal{P}\hat{U} - U\|_{\mu_X}^2 \\
= & O_p\left(\max\left\{n^{-(2b_X-1)/(a_X+2b_X)}, n^{-(2b_Y-1)/(a_Y+2b_Y)}\right\}\right).
\end{aligned}$$

□

Proof of Theorem 3. By the fact $\mathcal{P}\hat{\mathbf{id}}_Y = \mathbf{id}_Y$, it is sufficient to show

$$\sup_{\substack{h\in\mathscr{T}(\mu_X)\\ \|h\|_{\mu_X}=1}} \left\|(\mathcal{P}\hat{\mathbf{C}}_Y + \epsilon_Y\mathbf{id}_Y)^{-1}(\mathcal{P}\hat{\mathbf{C}}_{YX})h - \mathbf{C}_Y^{-1}\mathbf{C}_{YX}h\right\|_{\mu_Y}^2 = O_p(n^{-(2b_Y-1)/(a_Y+2b_Y)}), \tag{S1.11}$$

as then the proof can be completed in analogy to the proof of Theorem 2. We start with the following decomposition

$$
\begin{aligned}
&(\mathcal{P}\hat{\mathbf{C}}_Y + \epsilon_Y \mathbf{id}_Y)^{-1}(\mathcal{P}\hat{\mathbf{C}}_{YX})h - \mathbf{C}_Y^{-1}\mathbf{C}_{YX}h \\
=&(\mathcal{P}\hat{\mathbf{C}}_Y + \epsilon_Y \mathbf{id}_Y)^{-1}(\mathcal{P}\hat{\mathbf{C}}_{YX})h - (\mathbf{C}_Y + \epsilon_Y \mathbf{id}_Y)^{-1}(\mathcal{P}\hat{\mathbf{C}}_{YX})h \\
&+ (\mathbf{C}_Y + \epsilon_Y \mathbf{id}_Y)^{-1}(\mathcal{P}\hat{\mathbf{C}}_{YX})h - (\mathbf{C}_Y + \epsilon_Y \mathbf{id}_Y)^{-1}(\mathbf{C}_{YX})h \\
&+ (\mathbf{C}_Y + \epsilon_Y \mathbf{id}_Y)^{-1}(\mathbf{C}_{YX})h - \mathbf{C}_Y^{-1}\mathbf{C}_{YX}h \\
:=&K_1 + K_2 + K_3.
\end{aligned}
\tag{S1.12}
$$

Since

$$
\begin{aligned}
&\mathbf{E}\left\|(\mathbf{C}_Y + \epsilon_Y \mathbf{id}_Y)^{-1}(\mathbf{C}_Y - \mathcal{P}\hat{\mathbf{C}}_Y)g\right\|_{\mu_Y}^2 \\
=&\mathbf{E}\left\|\sum_{j=1}^{\infty}\frac{1}{\lambda_{Y,j}+\epsilon_Y}g_j\lambda_j\Phi_{Y,j} - \sum_{j_1=1}^{\infty}\frac{1}{\lambda_{Y,j_1}+\epsilon_Y}\sum_{j_2=1}^{\infty}\frac{1}{n}\sum_{i=1}^{n}\eta_{ij_1}\eta_{ij_2}g_{j_2}\Phi_{Y,j_1}\right\|_{\mu_Y}^2 \\
=&\sum_{j_1=1}^{\infty}\left(\frac{1}{\lambda_{Y,j_1}+\epsilon_Y}\right)^2\mathbf{E}\left\{\left(\sum_{j_2=1}^{\infty}\frac{1}{n}\sum_{i=1}^{n}\eta_{ij_1}\eta_{ij_2}g_{j_2}\right) - g_{j_1}\lambda_{Y,j_1}\right\}^2 \\
=&\sum_{j_1=1}^{\infty}\left(\frac{1}{\lambda_{Y,j_1}+\epsilon_Y}\right)^2\left(\frac{2\lambda_{Y,j_1}^2 g_{j_1}^2}{n} + \sum_{j_2=1}^{\infty}\frac{\lambda_{Y,j_1}\lambda_{Y,j_2}}{n}g_{j_2}^2\right) \\
\leqslant&C\frac{1}{n}\sum_{j=1}^{\infty}\frac{\lambda_{Y,j}}{(\lambda_{Y,j}+\epsilon_Y)^2} = O_p\left(\epsilon_Y^{-(1+1/a_Y)}/n\right),
\end{aligned}
$$

for $g = \sum_{j=1}^{\infty} g_j \Phi_{Y,j}$ with $\sum_{j=1}^{\infty} g_j^2 = 1$ and due to the fact that $(\mathcal{P}\hat{\mathbf{C}}_Y + \epsilon_Y \mathbf{id}_Y)^{-1}\hat{\mathbf{C}}_{YX}$ is bounded in probability, we have $\|K_1\|_{\mu_Y}^2 = O_p\left(\epsilon_Y^{-(1+1/a_Y)}/n\right)$.

For the second term,

$$\begin{aligned}\mathbf{E}\|K_2\|_{\mu_Y}^2 =&\mathbf{E}\left\|\sum_{j_2=1}^{\infty}\frac{1}{\lambda_{Y,j_2}+\epsilon_Y}\sum_{j_1=1}^{\infty}\left\{\frac{1}{n}\sum_{i=1}^{n}(\xi_{ij_1}-\overline{\xi}_{j_1})(\eta_{ij_2}-\overline{\eta}_{j_2})-\gamma_{j_1j_2}\right\}h_{j_1}\Phi_{Y,j_2}\right\|_{\mu_Y}^2\\ \leqslant&\sum_{j_2=1}^{\infty}\left(\frac{1}{\lambda_{Y,j_2}+\epsilon_Y}\right)^2\mathbf{E}\sum_{j_1=1}^{\infty}\left\{\frac{1}{n}\sum_{i=1}^{n}(\xi_{ij_1}-\overline{\xi}_{j_1})(\eta_{ij_2}-\overline{\eta}_{j_2})-\gamma_{j_1j_2}\right\}^2\sum_{j_1=1}^{\infty}h_{j_1}^2\\ \leqslant&C\sum_{j_2=1}^{\infty}\left(\frac{1}{\lambda_{Y,j_2}+\epsilon_Y}\right)^2\sum_{j_1=1}^{\infty}\frac{\lambda_{X,j_1}\lambda_{Y,j_2}}{n}\\ \leqslant&\frac{C}{n}\sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty}\frac{\lambda_{X,j_1}\lambda_{Y,j_2}}{(\lambda_{Y,j_2}+\epsilon_Y)^2}=O_p\left(\epsilon_Y^{-(1+1/a_Y)}/n\right).\end{aligned} \tag{S1.13}$$

Finally,

$$\begin{aligned}\|K_3\|_{\mu_Y}^2 =&\|(\mathbf{C}_Y+\epsilon_Y\mathbf{id}_Y)^{-1}(\mathbf{C}_{YX})h-\mathbf{C}_Y^{-1}\mathbf{C}_{YX}h\|_{\mu_Y}^2\\ =&\left\|\sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty}\left(\frac{1}{\lambda_{Y,j_2}+\epsilon_Y}-\frac{1}{\lambda_{Y,j_2}}\right)\gamma_{j_1j_2}h_{j_1}\Phi_{Y,j_2}\right\|_{\mu_Y}^2\\ =&\sum_{j_2=1}^{\infty}\left(\frac{1}{\lambda_{Y,j_2}+\epsilon_Y}-\frac{1}{\lambda_{Y,j_2}}\right)^2\left(\sum_{j_1=1}^{\infty}\gamma_{j_1j_2}h_{j_1}\right)^2\leqslant\sum_{j_2=1}^{\infty}\frac{\epsilon_Y^2}{(\lambda_{Y,j_2}+\epsilon_Y)^2\lambda_{Y,j_2}^2}\sum_{j_1=1}^{\infty}\gamma_{j_1j_2}^2\\ \leqslant&C\epsilon_Y^2\sum_{j_2=1}^{\infty}\frac{j_2^{-2b_Y}}{(\lambda_{Y,j_2}+\epsilon_Y)^2}=O\left(\epsilon_Y^{(2b_Y-1)/a_Y}\right).\end{aligned}$$

When $\epsilon_Y \asymp n^{-a_Y/(a_Y+2b_Y)}$, we have $\epsilon_Y^{-(1+1/a_Y)}/n = \epsilon_Y^{(2b_Y-1)/a_Y}$, and then K_1, K_2 and K_3 are of the same order $n^{-(2b_Y-1)/(a_Y+2b_Y)}$. This establishes (S1.11) and further

$$\left|\left|\left|(\mathcal{P}\hat{\mathbf{C}}_Y+\epsilon_Y\mathbf{id}_Y)^{-1}(\mathcal{P}\hat{\mathbf{C}}_{YX})-\mathbf{C}_Y^{-1}\mathbf{C}_{YX}\right|\right|\right|_{\mathfrak{B}(\mu_X,\mu_Y)}^2=O_p(n^{-(2b_Y-1)/(a_Y+2b_Y)}).$$

□

Remark 1. In the equation (7) in Lian (2014) and the proof of Theorem 2 in Zhou and Chen (2020), both on canonical correlation analysis for Euclidean functional data, the following inequality

$$\mathbf{E}\left\|(\mathbf{C}_Y+\epsilon_Y\mathbf{id}_Y)^{-1}(\hat{\mathbf{C}}_{YX})h-(\mathbf{C}_Y+\epsilon_Y\mathbf{id}_Y)^{-1}(\mathbf{C}_{YX})h\right\|^2\leqslant\frac{1}{n}\mathbf{E}\left\|(\mathbf{C}_Y+\epsilon_Y\mathbf{id}_Y)^{-1}(\mathbf{C}_{YX})h\right\|^2,$$

translated from Lian (2014) and Zhou and Chen (2020) into our notation, taking into account that parallel transport is not needed for the Euclidean case, is used to derive

$$\mathbf{E}\|K_2\|^2\leqslant\frac{1}{n}\sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty}\frac{\gamma_{j_1j_2}^2}{(\lambda_{Y,j_2}+\epsilon_Y)^2}.$$

However, this seems incorrect. For a counterexample, take (ξ_{j_1},η_{j_2}) to follow a joint Gaussian distribution with zero mean and covariance matrix

$$\begin{pmatrix}\lambda_{X,j_1},\gamma_{j_1j_2}\\ \gamma_{j_1j_2},\lambda_{Y,j_2}\end{pmatrix}.$$

Let $h=\Phi_{X,1}$. Then $\mathbf{E}\|K_2\|^2=\sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty}(\lambda_{X,j_1}\lambda_{Y,j_2}+2\gamma_{j_1j_2}^2)/(\lambda_{Y,j_2}+\epsilon_Y)^2$, and under the condition B.2*, the term $n^{-1}\sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty}\lambda_{X,j_1}\lambda_{Y,j_2}/(\lambda_{Y,j_2}+\epsilon_Y)^2$ is asymptotically strictly larger than $n^{-1}\sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty}\gamma_{j_1j_2}^2/(\lambda_{Y,j_2}+\epsilon_Y)^2$. Consequently, there is a gap in the proof of Theorem 1 of Lian (2014), which however may be filled by using our arguments in the above. In contrast, the proof of Theorem 2 in both Lian (2014) and Zhou and Chen (2020) about the convergence rate with respect to an RHKS (reproducing kernel Hilbert space) norm, may not be fixed in the same way. To see this, let $\mathbb{G}_X$ be the RKHS generated by the kernel $\mathbf{C}_X$. Then a bound on

$\mathbf{E}\|K_2\|^2_{\mathbb{G}_X}$ is given by

$$\begin{aligned}\mathbf{E}\|K_2\|^2_{\mathbb{G}_X} \leqslant & \sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty}\frac{1}{\lambda_{X,j_1}}\left(\frac{1}{\lambda_{Y,j_2}+\epsilon_Y}\right)^2 \mathbf{E}\left(\frac{1}{n}\sum_{i=1}^{n}(\xi_{ij_1}-\overline{\xi}_{j_1})(\eta_{ij_2}-\overline{\eta}_{j_2})-\gamma_{j_1j_2}\right)^2\\ \leqslant & \frac{1}{n}\sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty}\frac{\lambda_{X,j_1}\lambda_{Y,j_2}}{\lambda_{X,j_1}}\left(\frac{1}{\lambda_{Y,j_2}+\epsilon_Y}\right)^2,\end{aligned}$$

which diverges with respect to j_1, contrasting the finite bound made in those works.

S2 Proofs of the ancillary results

Proof of Lemma 1. By Taylor expansion, for any real numbers x and x_0,

$$x^{-1}-x_0^{-1} = -x_0^{-2}\left(x-x_0\right)+\hat{x}^{-3}\left(x-x_0\right)^2,$$

where $\hat{x}$ is some value between x and x_0. Using this fact, we obtain

$$\begin{aligned}\left|\hat{\lambda}_{X,j}^{-1}-\lambda_{X,j}^{-1}\right| = & \lambda_{X,j}^{-2}O_p\left(\left|\hat{\lambda}_{X,j}-\lambda_{X,j}\right|\right)+\left(\lambda_{X,j}-O_p\left(|||\Delta_X|||\right)\right)^{-3}O_p\left(\left|\hat{\lambda}_{X,j}-\lambda_{X,j}\right|^2\right)\\ = & O_p(j^{2a_X}/\sqrt{n})+O_p(j^{3a_X}/n),\end{aligned}$$

where $\Delta_X = \mathcal{P}^{\mathfrak{B}(\mu_X,\mu_X)}_{\mathfrak{B}(\hat{\mu}_X,\hat{\mu}_X)}\hat{\mathbf{C}}_X - \mathbf{C}_X$ and the last equality follows from the part (d) of Theorem 1. Under the condition Assumption (A.2), we have $j^{3a_X}/n = o(j^{2a_X}/\sqrt{n})$ and the first assertion follows. The second assertion is proved by a similar argument.

□

Proof of Lemma 2. According to the part (d) of Theorem 1, we have $\left|\left|\left|\mathcal{P}\hat{\mathbf{C}}_X - \mathbf{C}_X\right|\right|\right| = O_p(n^{-1/2})$. Under the assumption Assumption (B.1), $\eta_{X,k_X} \leqslant k_X^{-a_X-1}$ and $n^{1/2}k_X^{-a_X-1} \to$

∞, which implies that

$$\mathbf{P}\left(\frac{1}{2}\eta_{X,k_X} > \left|\left|\left|\mathcal{P}\hat{\mathbf{C}}_X - \mathbf{C}_X\right|\right|\right|\right) \to 1.$$

The second statement follows analogously.

□

Proposition 1 extends the framework in Lin and Yao (2019) developed for (finite-dimensional) Riemannian manifolds to Wasserstein spaces.

Proof of Proposition 1. To show that $\mathscr{T}(\mu)$ is a Hilbert space, it is sufficient to prove that $\mathscr{T}(\mu)$ is complete. Suppose that V_n is a Cauchy sequence in $\mathscr{T}(\mu)$, i.e., $\lim_{N\to\infty}\sup_{m,n\geq N}|||V_n - V_m|||_\mu = 0$, where $|||\cdot|||_\mu$ denotes the norm induced by the inner product in $\mathscr{T}(\mu)$. From this sequence we choose a subsequence n_k such that $\left|\left|\left|V_{n_k} - V_{n_{k+1}}\right|\right|\right|_\mu \leqslant 2^{-k}$. For $U \in \mathscr{T}(\mu)$, by Cauchy–Schwarz inequality,

$$\int_{\mathcal{T}} \|U(t)\|_{\mu(t)} \cdot \|V_{n_k}(t) - V_{n_{k+1}}(t)\|_{\mu(t)}\mathrm{d}t \leqslant |||U|||_\mu \left|\left|\left|V_{n_k} - V_{n_{k+1}}\right|\right|\right|_\mu \leqslant 2^{-k}|||U|||_\mu,$$

where $\|\cdot\|_{\mu(t)}$ denotes the norm induced by the inner product in $\mathrm{Tan}_{\mu(t)}$. Thus

$$\sum_{k=1}^{\infty}\int_{\mathcal{T}} \|U(t)\|_{\mu(t)} \cdot \|V_{n_k}(t) - V_{n_{k+1}}(t)\|_{\mu(t)}\mathrm{d}t \leqslant |||U|||_\mu < \infty,$$

which implies

$$\sum_{k=1}^{\infty}\left\|V_{n_{k+1}}(t) - V_{n_k}(t)\right\|_{\mu(t)} < \infty, \quad \text{a.e..} \tag{S2.14}$$

Since $\mathrm{Tan}_{\mu(t)}$ is complete, for those $t \in \mathcal{T}$ such that (S2.14) holds, the limit $V(t) = \lim_{k\to\infty} V_{n_k}(t)$ is defined and falls into $\mathrm{Tan}_{\mu(t)}$. For any $\epsilon > 0$, choose N_ϵ such that

$n, m \geqslant N_\epsilon$ and $|||V_n - V_m|||_\mu \leqslant \epsilon$. Fatou's lemma applied to the function $\|V_{n_k}(t) - V_{n_{k+1}}(t)\|_{\mu(t)}$ implies that if $m \geqslant N$, then $|||V - V_m|||_\mu^2 \leq \liminf_{k\to\infty} |||V_{n_k} - V_m|||_\mu^2 \leq \epsilon^2$. This shows that $V - V_m \in \mathscr{T}(\mu)$, thus $V = (V - V_m) + V_m \in \mathscr{T}(\mu)$. The arbitrariness of ϵ implies that $\lim_{m\to\infty} |||V - V_m|||_\mu = 0$. From the triangle inequality $|||V - V_n|||_\mu \leqslant |||V - V_m|||_\mu + |||V_m - V_n|||_\mu \leqslant 2\epsilon$, we conclude that V_n converges to V in $\mathscr{T}(\mu)$, and further that $\mathscr{T}(\mu)$ is complete.

To see $\mathscr{T}(\mu)$ is separable, we notice that for each $t \in \mathcal{T}$, $\mathrm{Tan}_{\mu(t)}$ is a separable Hilbert space thus it has an orthonormal basis $\{\Phi_k(t,s)\}_{k=1}^\infty$. Define $\mathbf{O} = \{\{\Phi_k(t,s)\}_{k=1}^\infty | \forall t \in \mathcal{T}\}$, which can be regarded as an orthonormal frame along the curve $\mu(t)$. For every element $U \in \mathscr{T}(\mu)$, define $U_{\mathbf{O}}$ be the coordinate representation of U with respect to $\mathbf{O}$. One can see that $U_{\mathbf{O}}$ is an element in the Hilbert space $\mathcal{L}^2(\mathcal{T}, l^2)$ of square integrable l^2-valued measurable functions with norm $\|f\|_{\mathcal{L}^2} = \{\int |f(t)|_{l^2}^2 dt\}^{1/2}$ for $f \in \mathcal{L}^2(\mathcal{T}, l^2)$, where l^2 is the space of square-summable sequences endowed with the inner product $\langle \mathbf{a}, \mathbf{b}\rangle_{l^2} = \sum_{j=1}^\infty a_j b_j$ and the induced norm $|\cdot|_{l^2}$, for each $\mathbf{a} = (a_1, a_2, \cdots)$ and $\mathbf{b} = (b_1, b_2, \cdots)$. If we define the map $\Upsilon : \mathscr{T}(\mu) \longmapsto \mathcal{L}^2(\mathcal{T}, l^2)$ by $\Upsilon(U) = U_{\mathbf{O}(t)}$, we can immediately see that Υ is a linear map. It is also surjective, because for any $f \in \mathcal{L}^2(\mathcal{T}, l^2)$, the vector field U along μ given by $U(t) = \sum_{k=1}^\infty f_k(t)\Phi_k(t,s)$ for $t \in \mathcal{T}$ is an element in $\mathscr{T}(\mu)$, where $f_k(t)$ denotes the kth component of $f(t)$. It can be verified that Υ preserves the inner product. Therefore, it is a Hilbertian isomorphism. Since $\mathcal{L}^2(\mathcal{T}, l^2)$ is separable, the isomorphism between $\mathcal{L}^2(\mathcal{T}, l^2)$ and $\mathscr{T}(\mu)$ implies separability of $\mathscr{T}(\mu)$. □

The first two statements of Proposition 2 are direct results of Proposition 1 in Chen et al. (2023) and the remaining assertions can be checked by similar arguments in Lin and Yao (2019). Proposition 3 and 4 can be checked by arguments similar to that of He et al. (2003) and Lian (2014). Thus we omit theirs proofs.

S3 Additional simulation results

Figure 2 and 3 present the histograms for selected tuning parameters and behaviors of our proposed estimators with noise level $\sigma = 0.1$ and $\sigma = 0.2$, leading to the same conclusion as in those observations in Section 4. Table 1 presents the IMSE of V.

Following are the setting details of the heavy-tail distribution and Figure 1 presents the convergence rates of weight functions via different tuning parameters.

- Case 3 (Log-normal distribution): We sample $\xi_{ik} \sim 0.1 v_k (V_k M)^{-1} a_i \theta_{ik}$ with $\theta_{ik} \sim \mathrm{LN}(0,1)$ independently for $i = 1, 2, \cdots, n$ and $k = 1, 2, \cdots, K$, and $\eta_{ik} \sim 0.1 v_k (V_k M)^{-1} b_i \vartheta_{ik}$ with $\vartheta_{ik} \sim \mathrm{LN}(0,1)$, except that $\eta_{i2} = 0.5(\xi_{i1} + \xi_{i2}) + 0.1 * \sigma \mathbf{E}(\xi_{i1}^2 + \xi_{i2}^2)\vartheta_{i2}$, where LN(0, 1) denotes the log-normal distribution with mean 0 and standard deviation 1 in the log scale, a_i are the Bernoulli random variables with success probability 1/2, and σ is a constant representing the noise level.

Table 1: IMSE of $\hat{V}$ ($\tilde{V}$, respectively) for different noise levels with tuning parameters chosen by 5-fold CV.

		FPCA				Tikhonov			
	σ	n=50	n=100	n=200	n=500	n=50	n=100	n=200	n=500
Case 1	0.05	.1095	.0774	.0621	.0405	.1943	.1982	.2431	.1849
	0.1	.1554	.1177	.0989	.0657	.2003	.1907	.2361	.2316
	0.2	.2111	.1754	.1661	.1048	.2255	.2268	.2359	.2020
	0.3	.2640	.2340	.2061	.1360	.2549	.2622	.2459	.1997
	0.5	.3869	.3268	.2753	.2187	.2860	.3360	.2939	.2608
Case 2	0.05	.1133	.0769	.0645	.0433	.2040	.1955	.2387	.1925
	0.1	.1568	.1114	.1074	.0663	.2082	.1857	.2414	.2395
	0.2	.2150	.1674	.1632	.1070	.2271	.2241	.2410	.2041
	0.3	.2677	.2274	.2062	.1422	.2578	.2563	.2504	.2029
	0.5	.3847	.3287	.2829	.2085	.2922	.3287	.2917	.2571

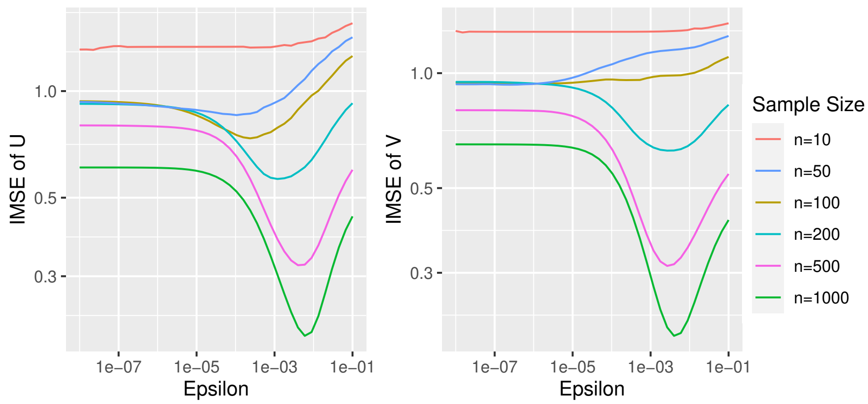

Figure 1: The IMSE for $\tilde{U}$ (left panel) and $\tilde{V}$ (right panel), respectively on different tuning parameters for the and Tikhonov method by the average of 200 Monte Carlo replicates with noise level $\sigma = 0.1$ in Case 3.

References

Aldous, D. J. (1976). A characterisation of Hilbert space using the central limit theorem. Journal of the London Mathematical Society 2(2), 376–380.

Bosq, D. (2000). Linear Processes in Function Spaces: Theory and Applications, Volume 149. Springer Science & Business Media.

Chen, Y., Z. Lin, and H.-G. Müller (2023). Wasserstein regression. J. Amer. Statist. Assoc. 118(542), 869–882.

Dauxois, J., A. Pousse, and Y. Romain (1982). Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. J. Multivar. Anal. 12(1), 136–154.

Dou, W. W., D. Pollard, H. H. Zhou, et al. (2012). Estimation in functional regression for general exponential

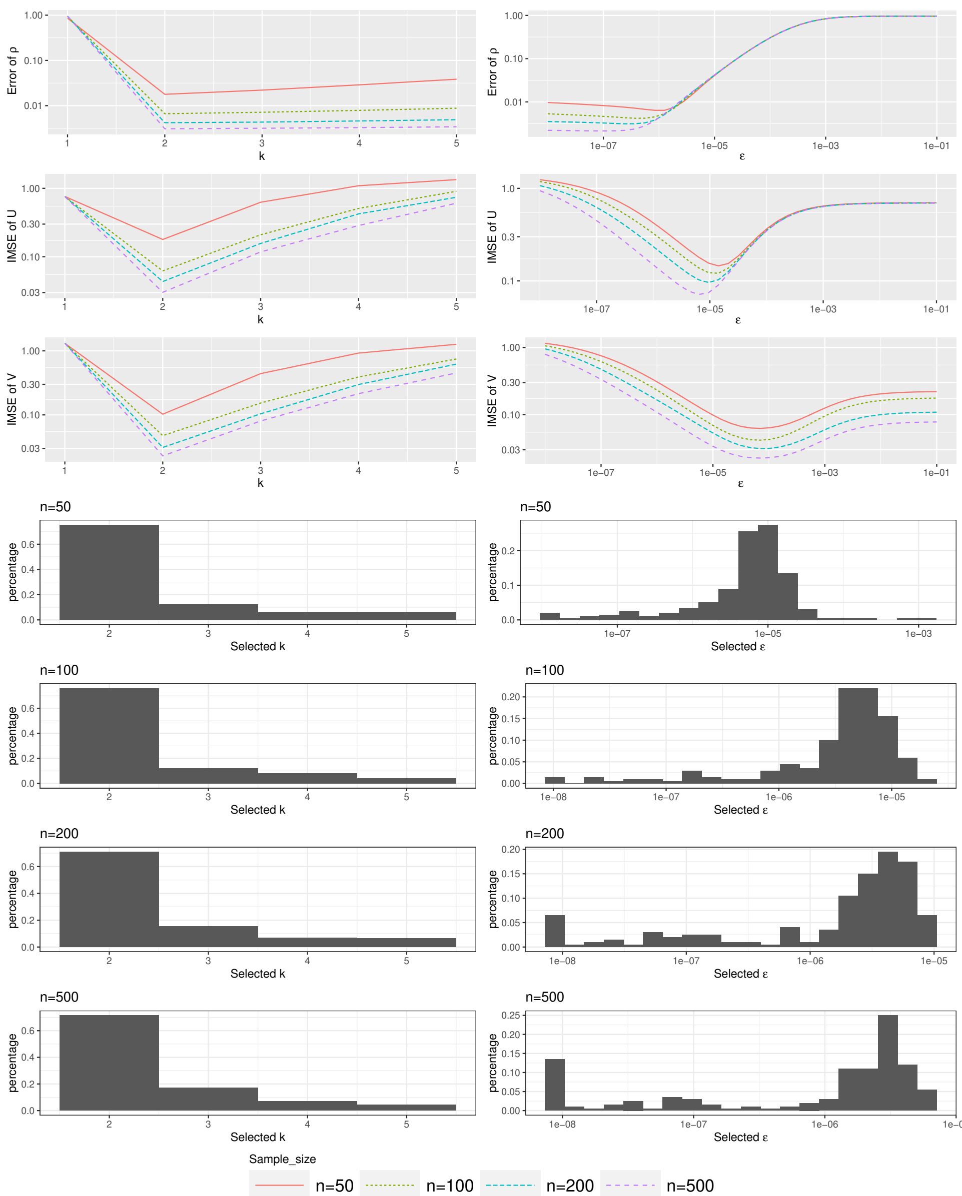

Figure 2: The first three rows show the absolute error for $\hat{\rho} - \rho$ ($\tilde{\rho} - \rho$, respectively) and IMSE for $\hat{U}, \hat{V}$ ($\tilde{U}, \tilde{V}$, respectively) on different tuning parameters and the last four rows present the Histograms of tuning parameters selected by 5-fold CV for the FPCA (left column) and Tikhonov (right column) methods by the average of 200 Monte Carlo replicates with noise level $\sigma = 0.1$ in Case 1.

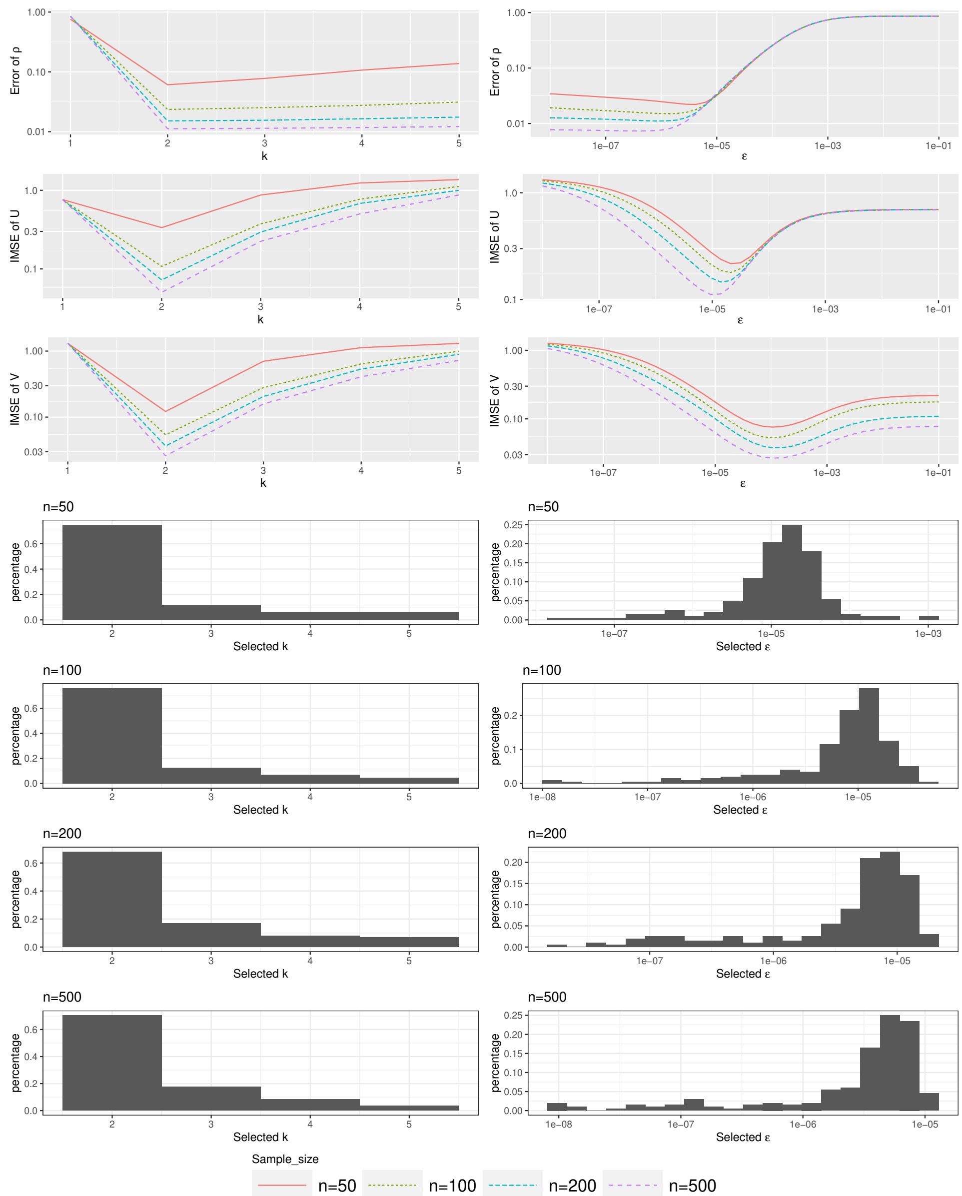

Figure 3: The first three rows show the absolute error for $\hat{\rho} - \rho$ ($\tilde{\rho} - \rho$, respectively) and IMSE for $\hat{U}, \hat{V}$ ($\tilde{U}, \tilde{V}$, respectively) on different tuning parameters and the last four rows present the Histograms of tuning parameters selected by 5-fold CV for the FPCA (left column) and Tikhonov (right column) methods by the average of 200 Monte Carlo replicates with noise level $\sigma = 0.2$ in Case 1.

families. Ann. Stat. 40(5), 2421–2451.

He, G., H.-G. Müller, and J.-L. Wang (2003). Functional canonical analysis for square integrable stochastic processes. J. Multivar. Anal. 85(1), 54–77.

Hsing, T. and R. Eubank (2015). Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators, Volume 997. John Wiley & Sons.

Lian, H. (2014). Some asymptotic properties for functional canonical correlation analysis. J. Stat. Plan. Inference 153, 1–10.

Lin, Z. and F. Yao (2019). Intrinsic Riemannian functional data analysis. Ann. Stat. 47(6), 3533 – 3577.

Panaretos, V. M. and Y. Zemel (2020). An Invitation to Statistics in Wasserstein Space. Springer Nature.

Zhou, Y. and D.-R. Chen (2020). The optimal rate of canonical correlation analysis for stochastic processes. J. Stat. Plan. Inference 207, 276–287.