Statistica Sinica: Supplement

Robust Inverse Regression for Multivariate Elliptical Functional Data

Eftychia Solea, Eliana Christou, and Jun Song

Queen Mary University of London,

University of North Carolina at Charlotte, and Korea University

The supplementary consists of four main sections. Section S1 provides the proofs of the main results of the paper, unless it is explicitly stated there that the proof is omitted. The proofs appear in the same order as the one they are presented in the main manuscript. Section S2 provides details on the development of the algorithms outlined in Section 5 of the paper. Finally, Sections S3 and S4 present additional simulation results that are not presented in the paper due to limited space. Note that, all the equations in the supplementary are labeled as (S1), (S2), and so on, while all the lemmas that appear exclusively here are labelled as "S" followed by a number (such as Lemma S1). This is to distinguish them from their counterparts in the paper, such as Lemma 1.

S1 Proofs of Main Results

S1.1 Useful Lemmas

Lemma S1. *Suppose* $A, B \in \mathscr{B}(\oplus_{i=1}^{p}\mathscr{H}_i)$ *are self adjoint and invertible operators. Then*

$$A^{-1} - B^{-1} = (A-B)A^{-2} + B^{-2}(A-B) - B^{-2}A^{2}(A-B)A^{-2} \\ - B^{-2}(A-B)B^{2}A^{-2} - B^{-2}A(A-B)BA^{-2}.$$

If ϵ_n and δ_n are two sequences of positive numbers such that $\epsilon_n/\delta_n \to 0$, then we write $\epsilon_n \prec \delta_n$ or $\delta_n \succ \epsilon_n$. If the sequence ϵ_n/δ_n either goes to zero or is bounded, then we write $\epsilon_n \preceq \delta_n$ or $\delta_n \succeq \epsilon_n$. The next lemma reveals the role played by Tychonoff regularization in the asymptotic order of magnitude and is given in Li and Solea (2018).

Lemma S2. *For any self adjoint operator* A, $\epsilon_n \prec 1$, *and* $a > 0$, $b > 0$, *we have* $\|(A+\epsilon_n I)^{-b}A^{a}\| = O(\epsilon_n^{\min\{0,a-b\}})$. *If* $\hat{A}_n$ *is a sequence of self adjoint random operator with* $\|\hat{A}_n\| = O_P(1)$, *then* $\|(\hat{A}_n+\epsilon_n I)^{-b}\hat{A}_n^{a}\| = O_P(\epsilon_n^{\min\{0,a-b\}})$.

Lemma S3. *If* A *and* B *are self adjoint and invertible linear operators,* *then*

$$A^{-1/2} - B^{-1/2} = A^{-3/2}(B^{3/2} - A^{3/2})B^{-1/2} + A^{-3/2}(A - B)$$
$$= A^{-1/2}(B^{3/2} - A^{3/2})B^{-3/2} + (A - B)B^{-3/2}.$$

S1.2 Proof of Lemma 1

We extend the arguments of Han and Liu (2018) to the functional setting. Since $X, \tilde{X} \sim \mathcal{E}_p(\mu_X, \Sigma, \varphi)$, then there exists a bounded linear operator $A : \oplus_{i=1}^p \mathscr{H}_i \mapsto \mathbb{R}^d$, $d \geq 1$, such that $AX \sim \mathscr{E}_d(A\mu_X, A\Sigma A^*, \varphi)$ and $A\tilde{X} \sim \mathscr{E}_d(A\mu_X, A\Sigma A^*, \varphi)$. By independence of X and $\tilde{X}$, we have

$$\mathbb{E}[\exp\{it^{\intercal} A(X - \tilde{X})\}] = \varphi^2\{t^{\intercal}(A\Sigma A^*)t\}.$$

Therefore, $X - \tilde{X} \sim \mathcal{E}_p(0, \Sigma, \varphi^2)$. Further, by the characterization of elliptical random elements, $X - \tilde{X} \overset{d}{=} S'N$, where S' and N are independent, S' is a nonnegative random variable such that $\mathbb{E}(S'^2) = 1$, and N is a Gaussian random element with zero mean and with the same covariance operator as $X - \tilde{X}$. Then,

$$\mathbb{E}\left\{\frac{(X - \tilde{X}) \otimes (X - \tilde{X})^{\intercal}}{\|X - \tilde{X}\|^2_{\oplus\mathscr{H}}}\right\} = \mathbb{E}\left\{\frac{(S'N) \otimes (S'N)^{\intercal}}{\|S'N\|^2_{\oplus\mathscr{H}}}\right\} = \mathbb{E}\left(\frac{N \otimes N^{\intercal}}{\|N\|^2_{\oplus\mathscr{H}}}\right). \tag{S1.1}$$

Since $X \sim \mathcal{E}_p(\mu_X, \Sigma, \varphi)$, then $X - \mu_X \overset{d}{=} SN$ for some S nonnegative random variable, independent of N, and

$$\mathbb{E}\Big\{\frac{(SN)\otimes(SN)^{\intercal}}{\|SN\|^2_{\oplus\mathscr{H}}}\Big\} = \mathbb{E}\Big\{\frac{(X-\mu_X)\otimes(X-\mu_X)^{\intercal}}{\|X-\mu_X\|^2_{\oplus\mathscr{H}}}\Big\}. \tag{S1.2}$$

Combining (S1.1) and (S1.2) completes the proof. □

S1.3 Proof of Theorem 2

The proof uses similar steps as in Chen et al. (2022) and it suffices to show that

$$\mathrm{span}\{T_{XX}\beta_1, \ldots, T_{XX}\beta_K\} = \mathrm{span}\{\Sigma_{XX}\beta_1, \ldots, \Sigma_{XX}\beta_K\}.$$

Let $h \in \oplus_{i=1}^p \mathscr{H}_i$ and $B = (\beta_1, \ldots, \beta_K)$. Then, by the spectral decomposition $T_{XX} = \sum_{r=1}^{\infty} \delta_r \psi_r \otimes \psi_r$, we have

$$\begin{aligned}
BT_{XX}h &= B\sum_{r=1}^{\infty} \delta_r \langle \psi_r, h\rangle_{\oplus\mathscr{H}}\psi_r = B\sum_{r=1}^{\infty}\mathbb{E}\Big(\frac{\gamma_r Y_r^2}{\sum_{k=1}^{\infty}\gamma_k Y_k^2}\Big)\langle\psi_r, h\rangle_{\oplus\mathscr{H}}\psi_r \\
&= B\sum_{r=1}^{\infty}\mathbb{E}\Big(\frac{Y_r^2}{\sum_{k=1}^{\infty}\gamma_k Y_k^2}\Big)\gamma_r\langle\psi_r, h\rangle_{\oplus\mathscr{H}}\psi_r \\
&= B\sum_{r=1}^{\infty}\mathbb{E}\Big(\frac{Y_1^2}{\sum_{k=1}^{\infty}\gamma_k Y_k^2}\Big)\gamma_r\langle\psi_r, h\rangle_{\oplus\mathscr{H}}\psi_r \\
&= B\sum_{r=1}^{\infty}\theta\gamma_r\langle\psi_r, h\rangle_{\oplus\mathscr{H}}\psi_r = B\sum_{r=1}^{\infty}\gamma_r\langle\psi_r, \theta h\rangle_{\oplus\mathscr{H}}\psi_r = B\Sigma_{XX}(\theta h),
\end{aligned}$$

where $\theta = \mathbb{E}\Big(Y_1^2/\sum_{k=1}^{\infty}\gamma_k Y_k^2\Big)$ and the fourth equality follows by the fact that $Y_r, r = 1, \ldots,$ are independent standard normal random variables. This completes the proof. $\square$

S1.4 Proof of Theorem 3

The Karhunen-Loève decomposition (2.3) of X gives $X - \mu_X = \sum_{r=1}^{\infty} \gamma_r^{1/2} \rho_r \, \psi_r$. Let $\rho_r^* = \gamma_r^{1/2} \rho_r$ and write

$$\mathbb{E}(X|Y) - \mathbb{E}(X) = \mathbb{E}\Big(\sum_{r=1}^{\infty} \rho_r^* \psi_r |Y\Big) = \sum_{r=1}^{\infty} \mathbb{E}(\rho_r^*|Y)\psi_r. \tag{S1.3}$$

Let $\{\psi_1, \ldots, \psi_p\}$ be an orthonormal basis in $\oplus_{i=1}^{p} \mathscr{H}_i$, for some $p < \infty$ and $p \leq p' \leq \infty$ for a large integer p'. By Proposition 2.2 and Theorem 1 of Wang et al. (2022) it is enough to show that the coefficients $\{\mathbb{E}(\rho_r^*|Y)\psi_r, r \geq 1\}$ are coordinate-wise symmetric. In other words, $\mathbb{E}(X|Y)$ is functional coordinate symmetric (FCS) and hence weakly FCS.

Note that, since X is an elliptical element in $\oplus_{i=1}^{p} \mathscr{H}_i$, it is also FCS; see Proposition 2.2 in Wang et al. (2022). Hence,

$$(\rho_1^*, \ldots, \rho_p^*)^\intercal = (\langle X - \mu_X, \psi_1 \rangle_{\oplus \mathscr{H}}, \ldots, \langle X - \mu_X, \psi_p \rangle_{\oplus \mathscr{H}})^\intercal = \Omega_\psi Z_\psi, \tag{S1.4}$$

where Ω_ψ is a $p \times p'$ matrix such that $\Omega_\psi \Omega_\psi^\intercal = I_p$ and Z_ψ is a p-dimensional random vector such that $GZ_\psi = Z_\psi$, where G is a diagonal matrix with diagonal elements $G_{ii} \in \{-1, 1\}$. Then, using equation (S1.3) we obtain

$$(\langle \mathbb{E}(X|Y) - \mathbb{E}(X), \psi_1 \rangle_{\oplus \mathscr{H}}, \ldots, \langle \mathbb{E}(X|Y) - \mathbb{E}(X), \psi_p \rangle_{\oplus \mathscr{H}})^\intercal$$

$$= (\mathbb{E}(\rho_1^*|Y), \ldots, \mathbb{E}(\rho_p^*|Y))^\intercal = \mathbb{E}((\rho_1^*, \ldots, \rho_p^*)^\intercal|Y) = \Omega_\psi \mathbb{E}(Z_\psi|Y),$$

where the last equality follows from (S1.4). Then, $\mathbb{E}(GZ_\psi|Y)$ is coordinate-wise symmetric since $G\mathbb{E}(Z_\psi|Y) = \mathbb{E}(GZ_\psi|Y) = \mathbb{E}(Z_\psi|Y)$ and the proof is complete. □

S1.5 Proof of Proposition 1

First we show that $T_{XX}^{\dagger\frac{1}{2}} T_{XX|Y} T_{XX}^{\dagger\frac{1}{2}}$ is well-defined. For any $u \in \mathcal{R}_{T_{XX}^{\frac{1}{2}}}$, we need to show that $T_{XX|Y} T_{XX}^{\dagger\frac{1}{2}} u \in \mathcal{R}_{T_{XX}^{\frac{1}{2}}}$, i.e $\sum_{i=1}^{\infty} \delta_i^{-1} |\langle T_{XX|Y} T_{XX}^{\dagger\frac{1}{2}} u, \psi_i \rangle|^2 < \infty$. Recall the definitions of $T_{XX}^{\dagger\frac{1}{2}} = \sum_{i=1}^{\infty} \delta^{\ -1/2} \psi_i \otimes \psi_i$ and $T_{XX|Y} = \sum_{=1}^{\infty} \sum_{j=1}^{\infty} R_{ij} \psi_i \otimes \psi_j$, where

$$R_{ij} = \mathbb{E}\left[\frac{\{\mathbb{E}(\rho_i^*|Y) - \mathbb{E}(\rho_i^*|\tilde{Y})\}\{\mathbb{E}(\rho_j^*|Y) - \mathbb{E}(\rho_j^*|\tilde{Y})\}}{\sum_{r=1}^{\infty} \{\mathbb{E}(\rho_r^*|Y) - \mathbb{E}(\rho_r^*|\tilde{Y})\}^2} \right].$$

Let $u \in \mathcal{R}_{T_{XX}^{\frac{1}{2}}}$. By the orthonormality of the $\{\psi_i\}_{i\geq 1}$ and the Cauchy-Schwartz inequality we have

$$\begin{aligned} \textstyle\sum_{i=1}^{\infty} \delta_i^{-1} |\langle T_{XX|Y} T_{XX}^{\dagger\frac{1}{2}} u, \psi_i \rangle|^2 &\leq \textstyle\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \frac{R_{ij}^2}{\delta_i \delta_j} \sum_{\ell=1}^{\infty} |\langle \psi_\ell, u \rangle|^2 \\ &\leq \textstyle\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \frac{R_{ij}^2}{\delta_i \delta_j} \|u\|_{\mathscr{H}}^2 < \infty \end{aligned}$$

due to Assumption 3 and the last inequality is due to Parseval's identity. Thus, the operator $T_{XX}^{\dagger\frac{1}{2}} T_{XX|Y} T_{XX}^{\dagger\frac{1}{2}}$ is a well-defined operator from $\mathcal{R}_{T_{XX}^{1/2}}$ to $\mathcal{R}^{-1}_{T_{XX}^{1/2}}$. Next, we show that the eigenfunctions $\eta_1, \ldots, \eta_K$ of $T_{XX}^{\dagger\frac{1}{2}} T_{XX|Y} T_{XX}^{\dagger\frac{1}{2}}$ are well-defined in $\oplus_{i=1}^{p} \mathscr{H}_i$. Since η_k is an eigenfunction, we need to show that $T_{XX}^{\dagger\frac{1}{2}} T_{XX|Y} T_{XX}^{\dagger\frac{1}{2}} \eta_k$ belongs to $\mathcal{R}_{T_{XX}^{1/2}}$, for any $k = 1, \ldots, K$. By the orthonormality of the $\{\psi_i\}_{i\geq 1}$ and the Cauchy-Schwartz inequality, we have for any k, by

$$\textstyle\sum_{i=1}^{\infty} \delta_i^{-1} |\langle T_{XX}^{\dagger\frac{1}{2}} T_{XX|Y} T_{XX}^{\dagger\frac{1}{2}} \eta_k, \psi_i \rangle|^2 \leq \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \frac{R_{ij}^2}{\delta_i^2 \delta_j} \|\eta_k\|_{\mathscr{H}}^2 = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \frac{R_{ij}^2}{\delta_i^2 \delta_j} < \infty$$

due to the fact that $\|\eta_k\|_{\mathscr{H}}^2 = 1$ and by Assumption 3. Therefore, $\eta_k \in \mathcal{R}_{T_{XX}^{1/2}}$.

□

S1.6 Proof of Theorem 5

We extend the proof of Proposition 2 in Chen et al. (2022) to random operators in $\mathscr{B}(\oplus_{i=1}^{p} \mathscr{H}_i)$. First, we decompose $\hat{T}_{XX|Y} - T_{XX|Y} = A_1 + A_2 + A_3$, where

$$\begin{aligned} A_1 = \frac{2}{n(n-1)} &\textstyle\sum_{1 \leq u < u' \leq n} \frac{\{\mathbf{m}(Y_u) - \mathbf{m}(Y_{u'})\} \otimes \{\mathbf{m}(Y_u) - \mathbf{m}(Y_{u'})\}^{\intercal}}{\|\mathbf{m}(Y_u) - \mathbf{m}(Y_{u'})\|_{\oplus\mathscr{H}}^2} \\ &- \mathbb{E}\Big[\frac{\{\mathbf{m}(Y) - \mathbf{m}(Y')\} \otimes \{\mathbf{m}(Y) - \mathbf{m}(Y')\}^{\intercal}}{\|\mathbf{m}(Y) - \mathbf{m}(Y')\|_{\oplus\mathscr{H}}^2}\Big], \end{aligned}$$

$$A_2 = \frac{2}{H(H-1)} \sum_{1 \le h < h' \le H} \frac{\{\mu_{X|Y}(h) - \mu_{X|Y}(h')\} \otimes \{\mu_{X|Y}(h) - \mu_{X|Y}(h')\}^{\mathsf{T}}}{\|\mu_{X|Y}(h) - \mu_{X|Y}(h')\|^2_{\oplus \mathscr{H}}}$$
$$- \frac{2}{n(n-1)} \sum_{1 \le u < u' \le n} \frac{\{\mathbf{m}(Y_u) - \mathbf{m}(Y_{u'})\} \otimes \{\mathbf{m}(Y_u) - \mathbf{m}(Y_{u'})\}^{\mathsf{T}}}{\|\mathbf{m}(Y_u) - \mathbf{m}(Y_{u'})\|^2_{\oplus \mathscr{H}}},$$
$$A_3 = \frac{2}{H(H-1)} \sum_{1 \le h < h' \le H} \frac{\{\hat{\mu}_{X|Y}(h) - \hat{\mu}_{X|Y}(h')\} \otimes \{\hat{\mu}_{X|Y}(h) - \hat{\mu}_{X|Y}(h')\}^{\mathsf{T}}}{\|\hat{\mu}_{X|Y}(h) - \hat{\mu}_{X|Y}(h')\|^2_{\oplus \mathscr{H}}}$$
$$- \frac{2}{H(H-1)} \sum_{1 \le h < h' \le H} \frac{\{\mu_{X|Y}(h) - \mu_{X|Y}(h')\} \otimes \{\mu_{X|Y}(h) - \mu_{X|Y}(h')\}^{\mathsf{T}}}{\|\mu_{X|Y}(h) - \mu_{X|Y}(h')\|^2_{\oplus \mathscr{H}}}.$$

Then, $\|\hat{T}_{XX|Y} - T_{XX|Y}\|_{\text{op}} \le \|A_1\|_{\text{op}} + \|A_2\|_{\text{op}} + \|A_3\|_{\text{op}}$, and therefore, it is enough to derive the convergence rates of $\|A_1\|_{\text{op}}$, $\|A_2\|_{\text{op}}$, and $\|A_3\|_{\text{op}}$.

First, use Theorem 4 to get

$$\|A_1\|_{\text{op}} = O_p(n^{-1/2}). \tag{S1.5}$$

Next, decompose the second term at the right hand side of A_2 into

$$\frac{2}{n(n-1)} \sum_{1 \le h < h' \le H} \sum_{j=1}^{\ell} \sum_{k=1}^{\ell} \frac{\{\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{h'k})\} \otimes \{\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{h'k})\}^{\mathsf{T}}}{\|\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{h'k})\|^2_{\oplus \mathscr{H}}}$$
$$+ \frac{2}{n(n-1)} \sum_{h=1}^{H} \sum_{1 \le j < k \le \ell} \frac{\{\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{hk})\} \otimes (\mathbf{m}\{Y_{hj}) - \mathbf{m}(Y_{hk})\}^{\mathsf{T}}}{\|\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{hk})\|^2_{\oplus \mathscr{H}}}$$

and write A_2 as $A_{21} + A_{22}$, where

$$A_{21} = \frac{2}{H(H-1)} \sum_{1 \le h < h' \le H} \frac{\{\mu_{X|Y}(h) - \mu_{X|Y}(h')\} \otimes \{\mu_{X|Y}(h) - \mu_{X|Y}(h')\}^{\mathsf{T}}}{\|\mu_{X|Y}(h) - \mu_{X|Y}(h')\|^2_{\oplus \mathscr{H}}}$$
$$- \frac{2}{n(n-1)} \sum_{1 \le h < h' \le H} \sum_{j=1}^{\ell} \sum_{k=1}^{\ell} \frac{\{\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{h'k})\} \otimes \{\mathbf{m}(Y_{hj}) - m(Y_{h'k})\}^{\mathsf{T}}}{\|\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{h'k})\|^2_{\oplus \mathscr{H}}},$$
$$A_{22} = \frac{2}{n(n-1)} \sum_{h=1}^{H} \sum_{1 \le j < k \le \ell} \frac{\{\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{hk})\} \otimes \{\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{hk})\}^{\mathsf{T}}}{\|\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{hk})\|^2_{\oplus \mathscr{H}}}.$$

Then, $\|A_2\|_{\text{op}} \le \|A_{21}\|_{\text{op}} + \|A_{22}\|_{\text{op}}$ and it is enough to derive the convergence rates of $\|A_{21}\|_{\text{op}}$ and $\|A_{22}\|_{\text{op}}$. By noting that $\mu_{X|Y}(h) = \mathbb{E}(X|Y \in J_h) =$

$\mathbb{E}(X|Y \in J_{(hj)}) = \mathbf{m}(Y_{hj})$, the first term of A_{21} can be rewritten as

$$\frac{2}{H(H-1)\ell^2}\sum_{1\le h<h'\le H}\sum_{j=1}^{\ell}\sum_{k=1}^{\ell}\frac{\{\mathbf{m}(Y_{hj})-\mathbf{m}(Y_{h'j})\}\otimes\{\mathbf{m}(Y_{hk})-\mathbf{m}(Y_{h'k})\}^{\intercal}}{\|\mu_{X|Y}(h)-\mu_{X|Y}(h')\|^2_{\oplus\mathscr{H}}}.$$

Moreover, by Jensen's inequality

$$\mathbb{E}(\|\mu_{X|Y}(h)\|^2_{\oplus\mathscr{H}}) \le \mathbb{E}\{\mathbb{E}(\|X\|^2|Y \in J_h)\} = \mathbb{E}(\|X\|^2) < \infty.$$

Thus, $\|\mu_{X|Y}(h) - \mu_{X|Y}(h')\|^2_{\oplus\mathscr{H}} = O_p(1)$.

Now, further decompose A_{21} into $A_{211} + A_{212} + A_{213} + A_{214}$, where

$$A_{211} = \frac{2}{H(H-1)\ell^2}\sum_{1\le h<h'\le H}\sum_{j=1}^{\ell}\sum_{k=1}^{\ell}\{\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{hk})\} \otimes \{\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{hk})\}^{\intercal},$$

$$A_{212} = \frac{2}{H(H-1)\ell^2}\sum_{1\le h<h'\le H}\sum_{j=1}^{\ell}\sum_{k=1}^{\ell}\{\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{hk})\} \otimes \{\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{h'k})\}^{\intercal},$$

$$A_{213} = \frac{2}{H(H-1)\ell^2}\sum_{1\le h<h'\le H}\sum_{j=1}^{\ell}\sum_{k=1}^{\ell}\{\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{h'k})\} \otimes \{\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{h'k})\}^{\intercal},$$

$$A_{214} = \frac{2}{H(H-1)\ell^2}\sum_{1\le h<h'\le H}\sum_{j=1}^{\ell}\sum_{k=1}^{\ell}\{\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{h'k})\} \otimes \{\mathbf{m}(Y_{h'j}) - \mathbf{m}(Y_{h'k})\}^{\intercal}.$$

Since $\|A_{21}\|_{\text{op}} \le \|A_{211}\|_{\text{op}} + \|A_{212}\|_{\text{op}} + \|A_{213}\|_{\text{op}} + \|A_{214}\|_{\text{op}}$ it is enough to derive the convergence rates of $\|A_{21k}\|_{\text{op}}$, $k = 1, 2, 3, 4$.

For A_{211}, note that

$$A_{211} = \frac{2}{H\ell^2}\sum_{h=1}^{H}\sum_{j=1}^{\ell}\sum_{k=1}^{\ell}\{\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{hk})\} \otimes \{\mathbf{m}(Y_{hk}) - \mathbf{m}(Y_{hj})\}^{\intercal},$$

where $A_{211} = 0$ when $j = k$. Therefore, we can rewrite A_{211} as

$$\begin{aligned} A_{2111} + A_{2112} = &\frac{2}{H\ell^2}\sum_{h=1}^{H}\sum_{1\le j<k\le\ell}\{\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{hk})\} \otimes \{\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{hk})\}^{\intercal} \\ &+ \frac{2}{H\ell^2}\sum_{h=1}^{H}\sum_{1\le j<k\le\ell}\{\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{hk})\} \otimes \{\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{hk})\}^{\intercal}. \end{aligned}$$

Hence, $\|A_{211}\|_{\text{op}} \leq \|A_{2111}\|_{\text{op}} + \|A_{2112}\|_{\text{op}}$. Below we derive the convergence rate of A_{2112}. The convergence rate of A_{2111} is derived using similar arguments, and thus we omit the details. For the derivation we use similar arguments as in the proof of Theorem 1 of Zhu and Ng (1995). In particular, arrange the inner double summation and move the summation over h to obtain

$$\begin{aligned}\|A_{2112}\| &\leq \frac{2}{n\ell}\sum_{h=1}^{H}\sum_{m=1}^{\ell-1}\sum_{k=1}^{\ell-m}\|\mathbf{m}(Y_{h(k+m)}) - \mathbf{m}(Y_{hk})\|_{\oplus\mathscr{H}}\|\mathbf{m}(Y_{h(k+m)}) - \mathbf{m}(Y_{hk})\|_{\oplus\mathscr{H}} \\ &= \frac{2}{n\ell}\sum_{m=1}^{\ell-1}\sum_{d=1}^{m} C_{dm},\end{aligned}$$

where

$$C_{dm} = \sum_{h=1}^{H}\sum_{*}\|\mathbf{m}(Y_{h(d+km)}) - \mathbf{m}(Y_{h(d+(k-1)m)})\|^2_{\oplus\mathscr{H}},$$

and the summation $*$ is over d subject to the restriction $d + km \leq \ell - m$. Next, for any δ such that $0 < \delta < 1/2$, we partition the outer sum over h into three intervals $[1, H\delta]$, $[H\delta + 1, H(1-p)]$ and $[H(1-p)+1, H]$. Then, $C_{dm} = C^1_{dm} + C^2_{dm} + C^3_{dm}$. Under the conditions of Theorem 5 and following the same arguments as in the proof of Lemma A.3 in Hsing and Carroll (1992), we can show that for each C^i_{dm}, $i = 1, 2, 3$, the maximum over d and m has order $O_p(n^{1/2})$. Hence, $\|A_{2112}\|_{\text{op}}$ has order $O_p(n^{-1/2})$. Therefore,

$$\|A_{211}\|_{\text{op}} \leq \|A_{2111}\|_{\text{op}} + \|A_{2112}\|_{\text{op}} = O_p(n^{-1/2}).$$

Next, we consider A_{212}. Using the same arguments as with A_{211}, we can decompose A_{212} as

$$A_{2121} + A_{2122} =$$
$$\frac{2}{H(H-1)\ell^2}\sum\nolimits_{1\le h<h'\le H}\sum\nolimits_{1\le k<j\le \ell}\{\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{hk})\} \otimes \{\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{hk})\}^{\intercal}$$
$$+ \frac{2}{H(H-1)\ell^2}\sum\nolimits_{1\le h<h'\le H}\sum\nolimits_{1\le j<k\le \ell}\{\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{hk})\} \otimes \{\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{h'k})\}^{\intercal}.$$

Again, it suffices to consider the term A_{2122} since the terms A_{2121} and A_{2122} are essentially identical. Note that,

$$\|A_{2122}\|_{\mathrm{op}} \le \frac{2}{H(H-1)\ell^2}\sum\nolimits_{1\le h<h'\le H}\sum\nolimits_{1\le k<j\le \ell}\|\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{hk})\|_{\oplus\mathscr{H}}\|\mathbf{m}(Y_{h'j}) - \mathbf{m}(Y_{hk})\|_{\oplus\mathscr{H}}$$

which can be written as

$$\|A_{2122}\|_{\mathrm{op}} = \frac{2}{n(H-1)\ell}\sum\nolimits_{m'=1}^{\ell-1}\sum\nolimits_{b=1}^{m'} C_{bm'am},$$

where

$$C_{bm'am} \quad = \quad \sum\nolimits_{m=1}^{H-1}\sum\nolimits_{a=1}^{m}\sum\nolimits_{*}\sum\nolimits_{**}\Big[\|\mathbf{m}(Y_{a+hm,b+jm'}) - \mathbf{m}(Y_{a+(h-1)m,b+(j-1)m'})\|_{\oplus\mathscr{H}}$$
$$\times\|\mathbf{m}(Y_{a+(h-1)m,b+jm'}) - \mathbf{m}(Y_{a+(h-1)m,b+(j-1)m'})\|_{\oplus\mathscr{H}}\Big],$$

where the summation $*$ is over h subject to the restriction $a+hm \le H-m$ and the summation $**$ is over j subject to the restriction $b+jm' \le \ell - m'$. Parallel to the proof of Lemma A.3 in Hsing and Carroll (1992), if we can choose small δ such that $0 < \delta < 1/2$ and divide the outer summation over

h into three summations over $[1,(H-1)\delta]$, $[(H-1)\delta+1,(H-1)(1-\delta)]$ and $[(H-1)(1-\delta)+1,H-1]$ such that $C^1_{bm'am}+C^2_{bm'am}+C^3_{bm'am}$, then we can show that the maximum over a,b,m and m' is of the order $O_p(n^{1/2})$. Hence, the term A_{2122} is of order $O_p(n^{-1/2})$. Therefore, $\|A_{212}\|_{\mathrm{op}}=O_p(n^{-1/2})$.

Similarly, we can derive $\|A_{214}\|_{\mathrm{op}}=O_p(n^{-1/2})$. For A_{213}, we have

$$\|A_{213}\|_{\mathrm{op}} \le \|A_{2131}\|_{\mathrm{op}} + \|A_{2132}\|_{\mathrm{op}},$$

where

$$A_{2131}=\frac{4}{H(H-1)\ell^2}\sum\nolimits_{1\le h<h'\le H}\sum\nolimits_{1\le j<k\le \ell}\|\mathbf{m}(Y_{hj})-\mathbf{m}(Y_{h'k})\|^2_{\oplus\mathscr{H}},$$
$$A_{2132}=\frac{4}{H(H-1)\ell^2}\sum\nolimits_{1\le h<h'\le H}\sum\nolimits_{j=1}^{\ell}\|\mathbf{m}(Y_{hj})-\mathbf{m}(Y_{h'j})\|^2_{\oplus\mathscr{H}}.$$

It can be shown using the same arguments for the terms A_{211} and A_{212}, that $A_{2131}=O_p(n^{-1/2})$. For A_{2132},

$$\begin{aligned}
\|A_{2132}\| &\le \frac{4}{H(H-1)\ell^2}\sum\nolimits_{m=1}^{H-1}\sum\nolimits_{h=1}^{H-m}\sum\nolimits_{j=1}^{\ell}\|\mathbf{m}(Y_{h+m,j})-\mathbf{m}(Y_{hj})\|^2_{\oplus\mathscr{H}}\\
&\le \frac{4}{H(H-1)\ell^2}\sum\nolimits_{m=1}^{H-1}\sum\nolimits_{n-m}^{h=1}\|\mathbf{m}(Y_{h+m})-\mathbf{m}(Y_{h})\|^2_{\oplus\mathscr{H}}\\
&\le \frac{4}{H(H-1)\ell^2}\sum\nolimits_{m=1}^{H-1}\sum\nolimits_{h_1=1}^{m}\sum\nolimits_{h_2=1}^{n-1}\|\mathbf{m}(Y_{h_2+1})-\mathbf{m}(Y_{h_2})\|^2_{\oplus\mathscr{H}}\\
&\le \frac{4}{H\ell^2}\sum\nolimits_{h_2=1}^{n-1}\|\mathbf{m}(Y_{h_2+1})-\mathbf{m}(Y_{h_2})\|^2_{\oplus\mathscr{H}}\\
&\le \frac{4}{n\ell}\sum\nolimits_{h_2=1}^{n-1}\|\mathbf{m}(Y_{h_2+1})-\mathbf{m}(Y_{h_2})\|^2_{\oplus\mathscr{H}}=O_p(n^{-1/2}),
\end{aligned}$$

by Lemma A.3 in Hsing and Carroll (1992).

Combining the results of $\|A_{21k}\|_{\rm op}$, $k = 1, 2, 3, 4$, yields that the first term at the right hand side of A_{21} is of order $O_p(n^{-1/2})$. Next, for the second term at the right hand side of A_{21}, we have

$$\left\| \frac{\{\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{h'k})\} \otimes \{\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{h'k})\}^{\intercal}}{\|\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{h'k})\|_{\oplus \mathscr{H}}} \right\|_{\rm op} \le 1.$$

Hence,

$$\left\| \frac{2}{n(n-1)} \sum_{1\le h<h'\le H} \sum_{j=1}^{\ell} \sum_{k=1}^{\ell} \frac{\{\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{h'k})\} \otimes \{\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{h'k})\}^{\intercal}}{\|\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{h'k})\|_{\oplus \mathscr{H}}} \right\|_{\rm op} \le \frac{2}{n(n-1)} = O_p(n^{-1}).$$

Therefore, $\|A_{21}\|_{\rm op} = O_p(n^{-1/2} + n^{-1}) = O_p(n^{-1/2})$.

Next, we consider $\|A_{22}\|_{\rm op}$. Since

$$\left\| \frac{\{\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{hk})\} \otimes \{\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{hk})\}^{\intercal}}{\|\mathbf{m}(Y_{hj}) - \mathbf{m}(Y_{hk})\|_{\oplus \mathscr{H}}} \right\|_{\rm op} \le 1,$$

then $\|A_{22}\|_{\rm op}$ is of order $O_p(n^{-1})$. By combining the results for $\|A_{21}\|_{\rm op}$ and $\|A_{22}\|_{\rm op}$ we have

$$\|A_2\|_{\rm op} = O_p(n^{-1/2}). \tag{S1.6}$$

For A_3, first we introduce some notation. Let $M_{hh'} = \mu_{X|Y}(h) - \mu_{X|Y}(h')$ and $\hat{M}_{hh'} = \hat{\mu}_{X|Y}(h) - \hat{\mu}_{X|Y}(h')$. Following the same calculations as in Chen et al. (2022), we can decompose $A_3 = A_{31} + A_{32}$ where

$$A_{31} = \frac{\hat{M}_{hh'} \otimes \hat{M}_{hh'}^{\intercal}}{\|\hat{M}_{hh'}\|^2_{\oplus\mathscr{H}}} - \frac{\hat{M}_{hh'} \otimes \hat{M}_{hh'}^{\intercal}}{\|M_{hh'}\|^2_{\oplus\mathscr{H}}},$$
$$A_{32} = \frac{\hat{M}_{hh'} \otimes \hat{M}_{hh'}^{\intercal}}{\|M_{hh'}\|^2_{\oplus\mathscr{H}}} - \frac{M_{hh'} \otimes M_{hh'}^{\intercal}}{\|M_{hh'}\|^2_{\oplus\mathscr{H}}}.$$

Hence, $\|A_3\| \leq \|A_{31}\|_{\text{op}} + \|A_{32}\|_{\text{op}}$. For A_{31}, we obtain after some calculations,

$$\begin{aligned}
\|A_{31}\|_{\text{op}} &\leq \|\hat{M}_{hh'} \otimes \hat{M}_{hh'}^{\intercal}\|_{\text{op}} \left| \frac{\|M_{hh'}\|^2_{\oplus\mathscr{H}} - \|\hat{M}_{hh'}\|^2_{\oplus\mathscr{H}}}{\|\hat{M}_{hh'}\|^2_{\oplus\mathscr{H}} \|M_{hh'}\|^2_{\oplus\mathscr{H}}} \right| \\
&= \|\hat{M}_{hh'} \otimes \hat{M}_{hh'}^{\intercal}\|_{\text{op}} \left| \frac{(\|M_{hh'}\|_{\oplus\mathscr{H}} + \|\hat{M}_{hh'}\|_{\oplus\mathscr{H}})}{\|\hat{M}_{hh'}\|^2_{\oplus\mathscr{H}} \|M_{hh'}\|^2_{\oplus\mathscr{H}}} \right| \left| \|M_{hh'}\|_{\oplus\mathscr{H}} - \|\hat{M}_{hh'}\|_{\oplus\mathscr{H}} \right| \\
&= \|\hat{M}_{hh'} \otimes \hat{M}_{hh'}^{\intercal}\|_{\text{op}} \left| \frac{(\|M_{hh'}\|_{\oplus\mathscr{H}} + \|\hat{M}_{hh'}\|_{\oplus\mathscr{H}})}{\|\hat{M}_{hh'}\|^2_{\oplus\mathscr{H}} \|M_{hh'}\|^2_{\oplus\mathscr{H}}} \right| \|M_{hh'} - \hat{M}_{hh'}\|_{\oplus\mathscr{H}}.
\end{aligned}$$

Meanwhile, by the Central Limit Theorem for independent and identically distributed Hilbert-valued random elements (Itô and Nisio 1968),

$$\|\hat{\mu}_{X|Y}(h) - \mu_{X|Y}(h)\|_{\oplus\mathscr{H}} = O_p(n^{-1/2}),$$

which implies $\|\hat{\mu}_{X|Y}(h)\|_{\oplus\mathscr{H}} = O_p(1)$. Moreover,

$$\|\hat{M}_{hh'} - M_{hh'}\|_{\oplus\mathscr{H}} = \|\{\hat{\mu}_{X|Y}(h) - \hat{\mu}_{X|Y}(h')\} - \{\mu_{X|Y}(h) - \mu_{X|Y}(h')\}\|_{\oplus\mathscr{H}} = O_p(n^{-1/2}),$$

which implies $\|\hat{M}_{hh'}\|_{\oplus\mathscr{H}} = \|\hat{\mu}_{X|Y}(h) - \hat{\mu}_{X|Y}(h')\|_{\oplus\mathscr{H}} = O_p(1)$. So $\|A_{31}\|_{\text{op}} = O_p(n^{-1/2})$. Finally, decompose $\|A_{32}\|_{\text{op}}$ as

$$\frac{1}{\|M_{hh'}\|^2_{\oplus\mathscr{H}}} \|(\hat{M}_{hh'} - M_{hh'}) \otimes \hat{M}_{hh'}^{\intercal}\|_{\text{op}} + \frac{1}{\|M_{hh'}\|^2_{\oplus\mathscr{H}}} \|M_{hh'} \otimes (\hat{M}_{hh'} - M_{hh'})^{\intercal}\|_{\text{op}}.$$

It is easy to see that both terms are of order $O_p(n^{-1/2})$. Hence, $\|A_{32}\|_{\text{op}} = O_p(n^{-1/2})$. Therefore,

$$\|A_3\|_{\text{op}} = O_p(n^{-1/2}). \tag{S1.7}$$

Combining (S1.5), (S1.6) and (S1.7), we have,

$$\|A_1\|_{\text{op}} + \|A_2\|_{\text{op}} + \|A_3\|_{\text{op}} = O_p(n^{-1/2}).$$

□

S1.7 Proof of Theorem 6

First, we define the following intermediate operator

$$M^{(\epsilon_n)} = (T_{XX} + \epsilon_n I)^{-\frac{1}{2}} T_{XX|Y} (T_{XX} + \epsilon_n I)^{-\frac{1}{2}}.$$

Then, by the triangle inequality, we have the following decomposition

$$\|\hat{M}^{(\epsilon_n)} - M\|_{\text{op}} \leq \|\hat{M}^{(\epsilon_n)} - M^{(\epsilon_n)}\|_{\text{op}} + \|M^{(\epsilon_n)} - M\|_{\text{op}} = I_1 + I_2.$$

For I_1, note that $\hat{M}^{(\epsilon_n)} - M^{(\epsilon_n)}$ is further decomposed as $A_1 + A_2 + A_3$, where

$$A_1 = [(\hat{T}_{XX} + \epsilon_n I)^{-1/2} - (T_{XX} + \epsilon_n I)^{-1/2}]\hat{T}_{XX|Y}(\hat{T}_{XX} + \epsilon_n I)^{-1/2},$$

$$A_2 = (T_{XX} + \epsilon_n I)^{-1/2}(\hat{T}_{XX|Y} - T_{XX|Y})(\hat{T}_{XX} + \epsilon_n I)^{-1/2},$$

$$A_3 = (T_{XX} + \epsilon_n I)^{-1/2} T_{XX|Y}[(\hat{T}_{XX} + \epsilon_n I)^{-1/2} - (T_{XX} + \epsilon_n I)^{-1/2}].$$

Decompose A_1 further as

$$[(\hat{T}_{XX}+\epsilon_n I)^{-1/2}-(T_{XX}+\epsilon_n I)^{-1/2}](\hat{T}_{XX|Y}-T_{XX|Y})(\hat{T}_{XX}+\epsilon_n I)^{-1/2}$$
$$+[(\hat{T}_{XX}+\epsilon_n I)^{-1/2}-(T_{XX}+\epsilon_n I)^{-1/2}]T_{XX|Y}(\hat{T}_{XX}+\epsilon_n I)^{-1/2}=A_{11}+A_{12}.$$

By Lemma S3

$$(\hat{T}_{XX}+\epsilon_n I)^{-1/2}-(T_{XX}+\epsilon_n I)^{-1/2}=$$
$$[(\hat{T}_{XX}+\epsilon_n I)^{-1/2}\{(\hat{T}_{XX}+\epsilon_n I)^{3/2}-(T_{XX}+\epsilon_n I)^{3/2}\}+\hat{T}_{XX}-T_{XX}](T_{XX}+\epsilon_n I)^{-3/2}. \tag{S1.8}$$

Thus,

$$\|(\hat{T}_{XX}+\epsilon_n I)^{-1/2}-(T_{XX}+\epsilon_n I)^{-1/2}\|_{\text{op}}$$
$$\leq [\|(\hat{T}_{XX}+\epsilon_n I)^{-1/2}\|_{\text{op}}\{\|(\hat{T}_{XX}+\epsilon_n I)^{3/2}-(T_{XX}+\epsilon_n I)^{3/2}\|_{\text{op}}+\|\hat{T}_{XX}-T_{XX}\|_{\text{op}}\}]$$
$$\times\|(T_{XX}+\epsilon_n I)^{-3/2}\|_{\text{op}}.$$

Moreover, by Lemma 8 in Fukumizu et al. (2007),

$$\|(\hat{T}_{XX}+\epsilon_n I)^{3/2}-(T_{XX}+\epsilon_n I)^{3/2}\|_{\text{op}}=O_p(n^{-1/2}). \tag{S1.9}$$

Hence, by relation (S1.9), Theorem 5 and the fact that

$$\|(\hat{T}_{XX}+\epsilon_n I)^{-1/2}\|_{\text{op}}=O_p(\epsilon_n^{-1/2}),\ \text{and}\ \|(T_{XX}+\epsilon_n I)^{-1/2}\|_{\text{op}}=O_p(\epsilon_n^{-1/2}), \tag{S1.10}$$

we obtain

$$\|A_{11}\|_{\text{op}}=O_p(n^{-1}\epsilon_n^{-5/2}). \tag{S1.11}$$

For A_{12}, by Assumption 5

$$A_{12} = [(\hat{T}_{XX} + \epsilon_n I)^{-1/2} - (T_{XX} + \epsilon_n I)^{-1/2}] T_{XX}^{1+\beta} D_{XX} T_{XX}^{1+\beta} (\hat{T}_{XX} + \epsilon_n I)^{-1/2}.$$

By (S1.8),

$$A_{12} = A_{121} A_{122},$$

where

$$A_{121} = [(\hat{T}_{XX} + \epsilon_n I)^{-1/2} \{(\hat{T}_{XX} + \epsilon_n I)^{3/2} - (T_{XX} + \epsilon_n I)^{3/2}\} + \hat{T}_{XX} - T_{XX}] \times (T_{XX} + \epsilon_n I)^{-3/2} T_{XX}^{1+\beta},$$

$$A_{122} = D_{XX} T_{XX}^{1+\beta} (\hat{T}_{XX} + \epsilon_n I)^{-1/2}.$$

Now, by Lemma S2,

$$\|(T_{XX} + \epsilon_n I)^{-3/2} T_{XX}^{1+\beta}\|_{\text{op}} = O_p(\epsilon_n^{\min\{0, -1/2+\beta\}}).$$

Hence, $\|A_{121}\|_{\text{op}} = O_p(n^{-1/2} \epsilon_n^{\min\{-1/2, -1+\beta\}})$. For A_{122}, first decompose as

$$\|A_{122}\|_{\text{op}} \leq \|D_{XX}\|_{\text{op}} \|T_{XX}^{1+\beta} \{(\hat{T}_{XX} + \epsilon_n I)^{-1/2} - (T_{XX} + \epsilon_n I)^{-1/2}\}\|_{\text{op}} \times \|D_{XX}\|_{\text{op}} \|T_{XX}^{1+\beta} (T_{XX} + \epsilon_n I)^{-1/2}\|_{\text{op}}.$$

By applying Lemma S2,

$$\|A_{122}\|_{\text{op}} = O_p(n^{-1/2} \epsilon_n^{\min\{-1/2, -1+\beta\}}) + O_p(1) = O_p(1),$$

where we used the fact $\epsilon_n^{\min\{-1/2, -1+\beta\}} \prec \epsilon_n^{-1} \prec n^{2/5}$. It follows

$$\|A_{12}\|_{\text{op}} \leq \|A_{121}\|_{\text{op}} \|A_{122}\|_{\text{op}} = O_p(n^{-1/2} \epsilon_n^{\min\{-1/2, -1+\beta\}}). \tag{S1.12}$$

Combining (S1.11) and (S1.12) we obtain

$$\|A_1\|_{\text{op}} \leq \|A_{11}\|_{\text{op}} + \|A_{12}\|_{\text{op}} \leq O_p(n^{-1}\epsilon_n^{-5/2} + n^{-1/2}\epsilon_n^{\min\{-1/2,-1+\beta\}}). \quad \text{(S1.13)}$$

For A_2, by Theorem 5 and relation (S1.10), it follows immediately that

$$\|A_2\|_{\text{op}} = O_p(n^{-1/2}\epsilon_n^{-1}). \quad \text{(S1.14)}$$

Finally, for the term A_3, by Assumption 5,

$$\begin{aligned} A_3 &= (T_{XX} + \epsilon_n I)^{-1/2} T_{XX}^{1+\beta} D_{XX} T_{XX}^{1+\beta} \{(\hat{T}_{XX} + \epsilon_n I)^{-1/2} - (T_{XX} + \epsilon_n I)^{-1/2}\} \\ &= A_{31} A_{32}, \end{aligned}$$

where

$$\begin{aligned} A_{31} &= (T_{XX} + \epsilon_n I)^{-1/2} T_{XX}^{1+\beta}, \\ A_{32} &= D_{XX} T_{XX}^{1+\beta} \{(\hat{T}_{XX} + \epsilon_n I)^{-1/2} - (T_{XX} + \epsilon_n I)^{-1/2}\}. \end{aligned}$$

By Lemma S2,

$$\|A_{31}\| = O_p(\epsilon_n^{\min\{0,-1+\beta\}}) = O_p(1). \quad \text{(S1.15)}$$

Using the same arguments for deriving $\|A_{121}\|_{\text{op}}$, we have

$$\|A_{32}\| = O_p(n^{-1/2}\epsilon_n^{\min\{-1/2,-1+\beta\}}). \quad \text{(S1.16)}$$

Thus, from (S1.15) and (S1.16),

$$\|A_3\| = O_p(n^{-1/2}\epsilon_n^{\min\{-1/2,-1+\beta\}}). \quad \text{(S1.17)}$$

Combining (S1.13), (S1.14) and (S1.17), and using the fact $\epsilon_n^{\min\{-1/2,-1+\beta\}} \prec \epsilon_n^{-1}$,

$$I_1 = O_p(n^{-1}\epsilon_n^{-5/2} + n^{-1/2}\epsilon_n^{-1}) \tag{S1.18}$$

Next, we consider I_2. First, we decompose $M^{(\epsilon_n)} - M$ into $B_1 + B_2 + B_3$, where

$$B_1 = \{(T_{XX} + \epsilon_n I)^{-1/2} - T_{XX}^{-1/2}\}T_{XX|Y}\{(T_{XX} + \epsilon_n I)^{-1/2} - T_{XX}^{-1/2}\},$$

$$B_2 = \{(T_{XX} + \epsilon_n I)^{-1/2} - T_{XX}^{-1/2}\}T_{XX|Y}T_{XX}^{-1/2},$$

$$B_3 = T_{XX}^{-1/2}T_{XX|Y}\{(T_{XX} + \epsilon_n I)^{-1/2} - T_{XX}^{-1/2}\}.$$

Hence, $I_2 \leq \|B_1\|_{\text{op}} + \|B_2\|_{\text{op}} + \|B_3\|_{\text{op}}$. For B_1, by the smoothness Assumption 5,

$$\|B_1\|_{\text{op}} \leq \|\{(T_{XX} + \epsilon_n I)^{-1/2} - T_{XX}^{-1/2}\}T_{XX}^{1+\beta}\|_{\text{op}}\|T_{XX}^{1+\beta}\{(T_{XX} + \epsilon_n I)^{-1/2} - T_{XX}^{-1/2}\}\|_{\text{op}}.$$

By Lemma S3 and the fact that $T_{XX} + \epsilon_n I$ and T_{XX} commute

$$\begin{aligned}\{(T_{XX} + \epsilon_n I)^{-1/2} - T_{XX}^{-1/2}\}T_{XX}^{1+\beta} = {}& \{T_{XX}^{3/2} - (T_{XX} + \epsilon_n I)^{3/2}\}(T_{XX} + \epsilon_n I)^{-3/2}T_{XX}^{1/2+\beta} \\ & + \epsilon_n(T_{XX} + \epsilon_n I)^{-3/2}T_{XX}^{1+\beta}.\end{aligned} \tag{S1.19}$$

By Lemma 8 in Fukumizu et al. (2007),

$$\|T_{XX}^{3/2} - (T_{XX} + \epsilon_n I)^{3/2}\|_{\text{op}} = O_p(\epsilon_n). \tag{S1.20}$$

Moreover, by Lemma S2,

$$\|(T_{XX}+\epsilon_n I)^{-3/2}T_{XX}^{1/2+\beta}\|_{\mathrm{op}} = O(\epsilon_n^{\min\{0,-1+\beta\}})$$
$$\|(T_{XX}+\epsilon_n I)^{-3/2}T_{XX}^{1+\beta}\|_{\mathrm{op}} = O(\epsilon_n^{\min\{0,-1/2+\beta\}}). \tag{S1.21}$$

Thus, by (S1.19), (S1.20) and (S1.21),

$$\|\{(T_{XX}+\epsilon_n I)^{-1/2} - T_{XX}^{-1/2}\}T_{XX}^{1+\beta}\|_{\mathrm{op}} = O(\epsilon_n^{\min(1,\beta)}).$$

Similarly, $\|T_{XX}^{1+\beta}\{(T_{XX}+\epsilon_n I)^{-1/2} - T_{XX}^{-1/2}\}\|_{\mathrm{op}} = O(\epsilon_n^{\min(1,\beta)})$. Hence,

$$\|B_1\|_{\mathrm{op}} = O(\epsilon_n^{\min(1,\beta)}). \tag{S1.22}$$

By noting that $\|D_{XX}T_{XX}^{1+\beta}\|_{\mathrm{op}} = O(1)$, we can use the same arguments to obtain

$$\|B_2\|_{\mathrm{op}} = O(\epsilon_n^{\min(1,\beta)}), \quad \|B_3\|_{\mathrm{op}} = O(\epsilon_n^{\min(1,\beta)}). \tag{S1.23}$$

By combining (S1.22) and (S1.23), we have

$$I_2 = O(\epsilon_n^{\min(1,\beta)}). \tag{S1.24}$$

Relations (S1.18) and (S1.24) complete the proof. □

S1.8 Proof of Theorem 7

For the proof of Theorem 7, we need the following lemma.

Lemma S4. *Under the assumptions of Theorem 6, $n^{-1/3} \prec \epsilon_n \prec 1$ and $\beta > 1$, we have*

$$\|(\hat{T}_{XX} + \epsilon_n I)^{-1/2}\hat{M}^{(\epsilon_n)} - T_{XX}^{-1/2}M\|_{\mathrm{op}} = O_p(n^{-1/2}\epsilon_n^{-3/2} + \epsilon_n^{\min(1,\beta-1)}).$$

Proof of Lemma S4. First, by Assumption 5, the term $\|(\hat{T}_{XX}+\epsilon_n I)^{-1/2}\hat{M}^{(\epsilon_n)} - T_{XX}^{-1/2}M\|_{\mathrm{op}}$ is bounded from above by $\Theta_1 + \Theta_2 + \Theta_3 + \Theta_4$, where

$$\Theta_1 = \|(\hat{T}_{XX} + \epsilon_n I)^{-1}(\hat{T}_{XX|Y} - T_{XX|Y})(\hat{T}_{XX} + \epsilon_n I)^{-1/2}\|_{\mathrm{op}},$$

$$\Theta_2 = \|[(\hat{T}_{XX} + \epsilon_n I)^{-1}T_{XX}^{1+\beta} - T_{XX}^{1+\beta}]D_{XX}[T_{XX}^{1+\beta}(\hat{T}_{XX} + \epsilon_n I)^{-1/2} - T_{XX}^{1/2+\beta}]\|_{\mathrm{op}},$$

$$\Theta_3 = \|[(\hat{T}_{XX} + \epsilon_n I)^{-1}T_{XX}^{1+\beta} - T_{XX}^{1+\beta}]D_{XX}T_{XX}^{1/2+\beta}\|_{\mathrm{op}},$$

$$\Theta_4 = \|T_{XX}^{\beta}D_{XX}[T_{XX}^{1+\beta}(\hat{T}_{XX} + \epsilon_n I)^{-1/2} - T_{XX}^{1/2+\beta}]\|_{\mathrm{op}}.$$

By Theorem 5, the first term Θ_1 is of order $O_p(n^{-1/2}\epsilon_n^{-3/2})$. By Lemma S7 and Lemma S8 in the Supplementary of Li and Song (2022), we have,

$$\|(\hat{T}_{XX} + \epsilon_n I)^{-1/2}T_{XX}^{1+\beta} - T_{XX}^{1/2+\beta}\|_{\mathrm{op}} = O_p(n^{-1/2}\epsilon_n^{-1/2} + \epsilon_n^{\min(1,\beta)}), \tag{S1.25}$$

$$\|(\hat{T}_{XX} + \epsilon_n I)^{-1}T_{XX}^{1+\beta} - T_{XX}^{\beta}\|_{\mathrm{op}} = O_p(n^{-1/2}\epsilon_n^{-1} + \epsilon_n^{\min(1,\beta-1)}). \tag{S1.26}$$

Hence,

$$\Theta_3 \leq \|[(\hat{T}_{XX} + \epsilon_n I)^{-1}T_{XX}^{1+\beta} - T_{XX}^{1+\beta}]\|_{\mathrm{op}}\|D_{XX}T_{XX}^{1/2+\beta}\|_{\mathrm{op}},$$

where the second norm on the right is of the order $O(1)$, and the first norm is of the order $O_p(n^{-1/2}\epsilon_n^{-1} + \epsilon_n^{\min(1,\beta-1)})$ by (S1.26). Similarly, $\Theta_4 =$

$O_p(n^{-1/2}\epsilon_n^{-1/2} + \epsilon_n^{\min(1,\beta)})$ by (S1.25). Finally, for Θ_2, we have

$$\Theta_2 = O_p(n^{-1/2}\epsilon_n^{-1} + \epsilon_n^{\min(1,\beta-1)})O(1)O_p(n^{-1/2}\epsilon_n^{-1/2} + \epsilon_n^{\min(1,\beta)}).$$

Hence,

$$\begin{aligned}
&\|(\hat{T}_{XX} + \epsilon_n I)^{-1/2}\hat{M}^{(\epsilon_n)} - T_{XX}^{-1/2}M\|_{\text{op}} \\
&= O_p(n^{-1/2}\epsilon_n^{-3/2} + n^{-1/2}\epsilon_n^{-1} + \epsilon_n^{\min(1,\beta-1)} + n^{-1/2}\epsilon_n^{-1/2} + \epsilon_n^{\min(1,\beta)}) \\
&= O_p(n^{-1/2}\epsilon_n^{-3/2} + \epsilon_n^{\min(1,\beta-1)}),
\end{aligned}$$

because $n^{-1/2}\epsilon_n^{-1} \prec n^{-1/2}\epsilon_n^{-3/2}$, $n^{-1/2}\epsilon_n^{-1/2} \prec n^{-1/2}\epsilon_n^{-1}$ and $\epsilon_n \prec \epsilon_n^{\min(1,\beta-1)}$. □

We now move to the proof of Theorem 7. First, by definition

$$\begin{aligned}
\|\hat{\beta}_\ell - \beta_\ell\|_{\oplus\mathscr{H}} &= \|(\hat{T}_{XX} + \epsilon_n I)^{-1/2}\hat{\eta}_\ell - T_{XX}^{-1/2}\eta_\ell\|_{\oplus\mathscr{H}} \\
&= \|(\hat{T}_{XX} + \epsilon_n I)^{-1/2}\hat{M}^{(\epsilon_n)}(\hat{\zeta}_\ell^{-1}\hat{\eta}_\ell) - T_{XX}^{-1/2}M(\zeta_\ell^{-1}\eta_\ell)\|_{\oplus\mathscr{H}}.
\end{aligned}$$

The right-hand side above is bounded from above by $\Lambda_1 + \Lambda_2 + \Lambda_3$, where

$$\begin{aligned}
\Lambda_1 &= \|[(\hat{T}_{XX} + \epsilon_n I)^{-1/2}\hat{M}^{(\epsilon_n)} - T_{XX}^{-1/2}M](\hat{\zeta}_\ell^{-1}\hat{\eta}_\ell - \zeta_\ell^{-1}\eta_\ell)\|_{\oplus\mathscr{H}}, \\
\Lambda_2 &= \|[(\hat{T}_{XX} + \epsilon_n I)^{-1/2}\hat{M}^{(\epsilon_n)} - T_{XX}^{-1/2}M]\zeta_\ell^{-1}\eta_\ell\|_{\oplus\mathscr{H}}, \\
\Lambda_3 &= \|T_{XX}^{-1/2}M(\hat{\zeta}_\ell^{-1}\hat{\eta}_\ell - \zeta_\ell^{-1}\eta_\ell)\|_{\oplus\mathscr{H}}.
\end{aligned}$$

By Lemma S4, Λ_2 is of the order $O_p(n^{-1/2}\epsilon_n^{-3/2} + \epsilon_n^{\min(1,\beta-1)})$. For the term, Λ_3, by Assumption 5, we have

$$\Lambda_3 \le \|T_{XX}^{1+\beta}D_{XX}T_{XX}^{1/2+\beta}\|_{\text{op}}\|(\hat{\zeta}_\ell^{-1}\hat{\eta}_\ell - \zeta_\ell^{-1}\eta_\ell)\|_{\oplus\mathscr{H}} \tag{S1.27}$$

The first norm on the right-hand side of (S1.27) is of order $O(1)$. The second norm on the right-hand side of (S1.27) is upper bounded by

$$|(\hat{\zeta}_\ell^{-1} - \zeta_\ell^{-1})| \|\hat{\eta}_\ell - \eta_\ell\|_{\oplus\mathscr{H}} + \zeta_\ell^{-1} \|\hat{\eta}_\ell - \eta_\ell\|_{\oplus\mathscr{H}} + |\hat{\zeta}_\ell^{-1} - \zeta_\ell^{-1}| \|\eta_\ell\|_{\oplus\mathscr{H}},$$

which is of order $O_p(n^{-1}\epsilon_n^{-5/2} + n^{-1/2}\epsilon_n^{-1} + \epsilon_n)$ by Corollary 1 and $\beta > 1$. Finally, Λ_1 is of the order $O_p(n^{-1/2}\epsilon_n^{-3/2} + \epsilon_n^{\min(1,\beta-1)}) O_p(n^{-1}\epsilon_n^{-5/2} + n^{-1/2}\epsilon_n^{-1} + \epsilon_n)$. Hence,

$$\begin{aligned} \|\hat{\beta}_\ell - \beta_\ell\|_{\oplus\mathscr{H}} &= O_p(n^{-1}\epsilon_n^{-5/2} + n^{-1/2}\epsilon_n^{-1} + \epsilon_n + n^{-1/2}\epsilon_n^{-3/2} + \epsilon_n^{\min(1,\beta-1)}) \\ &= O_p(n^{-1}\epsilon_n^{-5/2} + n^{-1/2}\epsilon_n^{-3/2} + \epsilon_n^{\min(1,\beta-1)}). \end{aligned}$$

Finally, by the Cauchy-Schwarz inequality,

$$\begin{aligned} |\langle \hat{\beta}_k, X\rangle_{\oplus\mathscr{H}} - \langle \beta_k, X\rangle_{\oplus\mathscr{H}}| &\leq \|\hat{\beta}_\ell - \beta_\ell\|_{\oplus\mathscr{H}} \|X\|_{\oplus\mathscr{H}} \\ &\leq O_p(n^{-1}\epsilon_n^{-5/2} + n^{-1/2}\epsilon_n^{-3/2} + \epsilon_n^{\min(1,\beta-1)}). \end{aligned}$$

□

S2 Detailed development of the algorithm

S2.1 Coordinate mapping

In practice, the functions $X_u(t)$, $u = 1, \ldots, n$, cannot be observed for all $t \in T$. Instead, they are observed on a finite set of points, say $t_{u1}, \ldots, t_{uN_u}$, and they need to be estimated using the observed data $\{(t, X_u(t)) : t = t_{u1}, \ldots, t_{uN_u}\}$. Commonly used methods for estimating $X_u(t)$ are smoothing splines or reproducing kernel Hilbert spaces, both of which can be formulated as projections onto a finite-dimensional Hilbert space.

For $i = 1, \ldots, p$, assume that $\mathscr{H}_i$ is spanned by a finite set of functions $\{g_1^i, \cdots, g_{k_n}^i\}$, and each X_u^i, $u = 1, \ldots, n$, can be approximated by a linear combination, say $c_{u1}^i g_1^i + \cdots + c_{uk_n}^i g_{k_n}^i$, where $c_{u1}^i, \ldots, c_{uk_n}^i$ are constants. Strictly speaking, we should use a different notation to denote this approximation, such as $\hat{X}_u^i$. However, we do not make this distinction to keep the notation simple. Because constants are irrelevant to our discussion, we will restrict our attention to $\overline{\text{ran}}(\hat{\Sigma}_{X^i X^i})$, which can be shown to be the $(k_n - 1)$-dimensional subspace of $\mathscr{H}_i$ spanned by $\{g_1^i - \bar{g}^i, \ldots, g_{k_n}^i - \bar{g}^i\} \equiv \{b_1^i, \ldots, b_{k_n}^i\}$, where $\bar{g}^i = k_n^{-1} \sum_{k=1}^{k_n} g_k^i$. We denote this set by $\mathscr{B}_i$ and the space spanned by it by $\mathscr{G}_i$. For an integer m, let I_m be the $m \times m$ identity matrix, 1_m be the m-dimensional vector whose entries are 1, and $Q_m = I_m - 1_m 1_m^{\mathsf{T}}/m$ be the projection onto the orthogonal complement of the subspace spanned by 1_m.

For each $i = 1, \ldots, p$, let $\mathbb{K}_i$ be the Gram matrix of the set $\{g_1^i, \ldots, g_{k_n}^i\}$ for $\mathscr{H}_i$; that is, $\mathbb{K}_i = \{\langle g_k^i, g_\ell^i \rangle_{\mathscr{H}_i}\}_{k,\ell=1}^{k_n}$. Then the Gram matrix of the set $\mathscr{B}_i$ is $Q_{k_n} \mathbb{K}_i Q_{k_n} \equiv G_i$. Each member f of $\mathscr{G}_i$ is a linear combination of $b_1^i, \ldots, b_{k_n}^i$. The vector of linear coefficients is called the coordinate of f with respect to $\mathscr{B}_i$, and is written as $[f]_{\mathscr{B}_i}$. For any operator $A \in \mathscr{B}(\mathscr{G}_i, \mathscr{G}_j)$, the matrix $([Ab_1^i]_{\mathscr{B}_j}, \ldots, [Ab_{kn}^i]_{\mathscr{B}_j})$ is called the coordinate of A, and is written as ${}_{\mathscr{B}_j}[A]_{\mathscr{B}_i}$. The mapping ${}_{\mathscr{B}_j}[\cdot]_{\mathscr{B}_i} : A \mapsto {}_{\mathscr{B}_j}[A]_{\mathscr{B}_i}$ is called the coordinate mapping and has the following properties, which can be derived from the basic facts listed in Horn and Johnson (1985); see also Theorem 5 of Solea and Li (2022).

S2.2 Coordinate representation of $\hat{T}_{XX}$ and $\hat{T}_{XX|Y}$

We first compute the coordinate representations of $\hat{T}_{XX}$ and $\hat{T}_{XX|Y}$ with respect to $\mathscr{B} = \oplus_{i=1}^p \mathscr{B}_i$. By properties 2, 4 and 5 of Lemma S5 in the supplementary material, the coordinate of $\hat{T}_{XX}$ is

$$[\hat{T}_{XX}] = \mathbb{E}_n \left(\frac{[X - \tilde{X}][X - \tilde{X}]^\intercal}{[X - \tilde{X}]^\intercal G [X - \tilde{X}]} \right) G = \Omega G, \tag{S2.1}$$

where $G = \mathrm{diag}(G_i : i = 1, \ldots p)$, $[X] \in \mathbb{R}^{pk_n}$ is the coordinate representation of X whose jth block is the coordinate representation of X^j, and

$$\Omega = \mathbb{E}_n \left(\frac{[X - \tilde{X}][X - \tilde{X}]^\intercal}{[X - \tilde{X}]^\intercal G [X - \tilde{X}]} \right) = \frac{2}{n(n-1)} \sum_{1 \le u < u' \le n} \frac{[X_u - X_{u'}][X_u - X_{u'}]^\intercal}{[X_u - X_{u'}]^\intercal G [X_u - X_{u'}]}. \tag{S2.2}$$

Note that $\Omega \in \mathbb{R}^{pk_n \times pk_n}$ is a symmetric matrix. We now compute the coordinate representation of $\hat{T}_{XX|Y}$ with respect to $\mathscr{B} = \oplus_{i=1}^p \mathscr{B}_i$. For a fixed $h = 1, \ldots, H$, by property 2 of Lemma S5,

$$[\hat{\mu}_{X|Y}(h)] = \frac{[\mathbb{E}_n\{XI(Y \in J_h)\}]}{\mathbb{E}_n\{I(Y \in J_h)\}} = \frac{\mathbb{E}_n\{[X]I(Y \in J_h)\}}{\mathbb{E}_n\{I(Y \in J_h)\}}. \tag{S2.3}$$

Thus, following similar arguments as with the coordinate representation of $\hat{T}_{XX}$, we obtain

$$[\hat{T}_{XX|Y}] = \frac{2}{H(H-1)} \sum_{1\leq h<h'\leq H} \left\{ \frac{[\hat{\mu}_{X|Y}(h)-\hat{\mu}_{X|Y}(h')][\hat{\mu}_{X|Y}(h)-\hat{\mu}_{X|Y}(h')]^{\intercal}}{[\hat{\mu}_{X|Y}(h)-\hat{\mu}_{X|Y}(h')]^{\intercal} G [\hat{\mu}_{X|Y}(h)-\hat{\mu}_{X|Y}(h')]} \right\} G = \Lambda G, \tag{S2.4}$$

where

$$\Lambda = \frac{2}{H(H-1)} \sum_{1\leq h<h'\leq H} \left\{ \frac{[\hat{\mu}_{X|Y}(h)-\hat{\mu}_{X|Y}(h')][\hat{\mu}_{X|Y}(h)-\hat{\mu}_{X|Y}(h')]^{\intercal}}{[\hat{\mu}_{X|Y}(h)-\hat{\mu}_{X|Y}(h')]^{\intercal} G [\hat{\mu}_{X|Y}(h)-\hat{\mu}_{X|Y}(h')]} \right\}. \tag{S2.5}$$

Lemma S5. *Suppose*

a. $\mathscr{G}_1, \ldots, \mathscr{G}_p$ *are finite-dimensional Hilbert spaces with spanning systems* $\mathscr{B}_1, \cdots, \mathscr{B}_p$, G_1 *is the Gram matrix of* $\mathscr{B}_1$, *and* $\mathscr{B} = \oplus_{i=1}^{p} \mathscr{B}_i$;

b. $T, T_1, T_2 \in \mathscr{B}(\mathscr{G}_1, \mathscr{G}_2)$, $U \in \mathscr{B}(\mathscr{G}_2, \mathscr{G}_3)$, $V \in \oplus_{i,j=1}^{p} \mathscr{B}(\mathscr{G}_i, \mathscr{G}_j)$, $W \in \mathscr{B}(\mathscr{G}_1, \mathscr{G}_1)$, $h \in \mathscr{G}_1$, $f \in \mathscr{G}_1$, $g \in \mathscr{G}_2$, $\alpha_1, \alpha_2 \in \mathbb{R}$.

Then,

1. *(evaluation)* $[Th]_{\mathscr{B}_2} = ({}_{\mathscr{B}_2}[T]_{\mathscr{B}_1})\,[h]_{\mathscr{B}_1}$;
2. *(linearity)* ${}_{\mathscr{B}_2}[\alpha_1 T_1 + \alpha_2 T_2]_{\mathscr{B}_1} = \alpha_1({}_{\mathscr{B}_2}[T_1]_{\mathscr{B}_1}) + \alpha_2({}_{\mathscr{B}_2}[T_2]_{\mathscr{B}_1})$;
3. *(composition)* ${}_{\mathscr{B}_3}[UT]_{\mathscr{B}_1} = ({}_{\mathscr{B}_3}[U]_{\mathscr{B}_2})\,({}_{\mathscr{B}_2}[T]_{\mathscr{B}_1})$;
4. *(inner product)* $\langle h, f\rangle_{\mathscr{G}_1} = ([h]_{\mathscr{B}_1})^{\intercal} G_1 ([f]_{\mathscr{B}_1})$;
5. *(tensor product)* ${}_{\mathscr{B}_2}[g \otimes h]_{\mathscr{B}_1} = [g]_{\mathscr{B}_2}\,[h]_{\mathscr{B}_1}^{\intercal} G_1$;
6. *(inverse)* *If* $W \in \mathscr{B}(\mathscr{G}_1, \mathscr{G}_1)$ *is self-adjoint and* G_1 *is the Gram matrix of* $\mathscr{G}_1$, *then for any* $c \in \mathbb{R}$ *for which* $(-1)^c \in \mathbb{R}$, ${}_{\mathscr{B}_1}[W^c]_{\mathscr{B}_1} =$

$G_1^{\dagger 1/2}(G_1^{1/2}{}_{\mathscr{B}_1}[W]_{\mathscr{B}_1}G_1^{\dagger 1/2})^c G_1^{1/2}$;

7. *(operator matrix)* *If* $\mathscr{B}$ *is a spanning system of* $\oplus_{i=1}^p \mathscr{G}_i$, *then* ${}_{\mathscr{B}}[V]_{\mathscr{B}}$ *is a block diagonal matrix with diagonal blocks* ${}_{\mathscr{B}_i}[V_i]_{\mathscr{B}_i}$,*, where* V_i *is the Gram matrix of* $\mathscr{G}_i$, $i = 1, \ldots, p$.

S2.3 Algorithm for R-FSIR

To derive the coordinate representation of the eigenvalue problem (3.7), we use the property 4 of Lemma S5 and write $\langle \eta, T_{XX}^{\dagger\frac{1}{2}} T_{XX|Y} T_{XX}^{\dagger\frac{1}{2}} \eta \rangle_{\oplus\mathscr{H}} = [\eta]^\intercal G[T_{XX}^{\dagger\frac{1}{2}}][T_{XX|Y}][T_{XX}^{\dagger\frac{1}{2}}][\eta]$, where $\langle \eta, \eta \rangle_{\oplus\mathscr{H}} = [\eta]^\intercal G[\eta]$ and $\langle \eta, \eta_\ell \rangle_{\oplus\mathscr{H}} = [\eta]^\intercal G[\eta_\ell]$, $\ell = 1, \ldots, K$. Set $[\eta] = G^{\dagger\frac{1}{2}} v$. Then, (3.7) can be implemented as the following standard eigenvalue problem with respect to v,

$$
\begin{aligned}
&\text{maximize} \quad v^\intercal G^{\dagger\frac{1}{2}} G[T_{XX}^{\dagger\frac{1}{2}}][T_{XX|Y}][T_{XX}^{\dagger\frac{1}{2}}] G^{\dagger\frac{1}{2}} v \\
&\text{subject to} \quad v^\intercal v = 1,\ v^\intercal v_\ell = 0,\ \ell = 1, \ldots, K-1.
\end{aligned}
\tag{S2.6}
$$

Moreover, using property 6 of Lemma S5 and relation (S2.1), we have

$$
[T_{XX}^{\dagger\frac{1}{2}}] = G^{\dagger\frac{1}{2}}(G^{\frac{1}{2}}[T_{XX}]G^{\dagger\frac{1}{2}})^{\dagger\frac{1}{2}} G^{\frac{1}{2}} = G^{\dagger\frac{1}{2}}(G^{\frac{1}{2}}\Omega G^{\frac{1}{2}})^{\dagger\frac{1}{2}} G^{\frac{1}{2}}. \tag{S2.7}
$$

Therefore, using (S2.4) and (S2.7), the matrix $G^{\dagger\frac{1}{2}} G[T_{XX}^{\dagger\frac{1}{2}}][T_{XX|Y}][T_{XX}^{\dagger\frac{1}{2}}] G^{\dagger\frac{1}{2}}$ in (S2.6) can be expressed as

$$
M\Lambda M^\intercal = (G^{\frac{1}{2}}\Omega G^{\frac{1}{2}})^{\dagger\frac{1}{2}} G^{\frac{1}{2}} \Lambda G^{\frac{1}{2}} (G^{\frac{1}{2}}\Omega G^{\frac{1}{2}})^{\dagger\frac{1}{2}}, \tag{S2.8}
$$

where $M = (G^{\frac{1}{2}}\Omega G^{\frac{1}{2}})^{\dagger\frac{1}{2}} G^{\frac{1}{2}}$. Thus, we aim to find the first K eigenvectors, $v_1, \ldots, v_K$, of the matrix $M\Lambda M^\intercal$. We then obtain $[\hat{\beta}_\ell] = G^{\dagger\frac{1}{2}}(G^{\frac{1}{2}}\Omega G^{\frac{1}{2}})^{\dagger\frac{1}{2}} v_\ell$,

$\ell = 1, \ldots, K$, and the sufficient predictors are

$$\hat{\beta}_\ell = v_\ell^\intercal (G^{\frac{1}{2}} \Omega G^{\frac{1}{2}})^{\dagger \frac{1}{2}} G^{\dagger \frac{1}{2}} b, \quad \ell = 1, \ldots, K, \tag{S2.9}$$

where $b = (b^{1\top}, \ldots, b^{p\top})^\top$ and $b^i = (b_1^i, \ldots, b_{k_n}^i)$. We summarize the algorithm below

1. For each $X_u, u = 1, \ldots, n$, calculate the centered version $X_u - \mathbb{E}_n X_u$, compute the coordinates $[X_u]$ relative to the basis $\mathscr{B}$ of $\oplus_{i=1}^p \mathscr{H}_i$, and derive the gram matrix $G = Q_{k_n} \mathbb{K} Q_{k_n}$ of the basis $\mathscr{B}$.
2. Divide the range of Y into H equal slices, $J_1, \ldots, J_H$.
3. For each $h = 1, \ldots, H$, compute $[\hat{\mu}_{X|Y}(h)]$ according to (S2.3).
4. Compute the matrices Ω and Λ, as defined in (S2.2) and (S2.5), respectively.
5. Compute the matrix $M \Lambda M^\intercal$ in (S2.8) and its K eigenvectors, $v_1, \ldots, v_K$.
6. Obtain the sufficient predictors $\hat{\beta}_\ell, \ell = 1, \ldots, K$ according to (S2.9).

S2.4 Order determination

In order to determine the dimension K of the central subspace, we use the CVBIC criterion introduced by Li et al. (2011). Let

$$G_n(k) = \textstyle\sum_{i=1}^k \hat{\nu}_i - a \hat{\nu}_1 n^{-1/4} \log(n) k, \tag{S2.10}$$

where $\hat{\nu}_i$ is the ith largest eigenvalue of the matrix representation of R-FSIR given in (S2.8), and a is a constant determined by the LOOCV criterion as follows. For each fixed a, maximize the criterion (S2.10) over $k = 0, \ldots, k_n p$ to obtain $\hat{K}(a)$. Let $\langle \hat{\beta}_1 X \rangle_{\oplus \mathscr{H}}, \ldots, \langle \hat{\beta}_{\hat{K}(a)}, X \rangle_{\oplus \mathscr{H}}$ be the sufficient predictors. Then, the optimal a is the one that minimizes the LOOCV score defined as $\text{LOOCV}(a) = \sum_{u=1}^{n} \{Y_u - \hat{m}_{(-u)}(Z_u)\}^2$, where $Z_u = (\langle \hat{\beta}_1, X_u \rangle_{\oplus \mathscr{H}}, \ldots, \langle \hat{\beta}_{\hat{K}(a)}, X_u \rangle_{\oplus \mathscr{H}})$ and $\hat{m}_{(-u)}(\cdot)$ is some nonparametric estimate of the conditional expectation $\mathbb{E}(Y|Z)$ based on the sample with the (Y_u, X_u) observation removed. The implementation of the CVBIC criterion relies on a nonparametric estimate of the conditional expectation $m_{(-u)}(Z_u)$. This can lead to the use of mutli-dimensional kernel estimator, which can suffer from the curse of dimensionality. However, we note that, in our simulations, we estimate the conditional expectation using additive models, which do not require the computation of multi-dimensional kernels.

S3 Effect of n, p and H

In this section, we present some additional results to investigate the performance of R-FSIR for a variety of combinations of (n, p, H). We assume Models II and IV, where X is simulated as described in Section 6, ϵ is generated according to a standard normal distribution, and the scores follow a Gaussian and a Cauchy distribution. We consider $(n, p, H) \in \{200, 400, 1000\} \times \{10, 20\} \times \{5, 10, 20\}$. Tables 1-4 report the observed, over the 100 simulation runs, means and standard deviations (in parenthesis) of the multiple correlation when no outliers are present (upper part) and when outliers are added as described in Section 6 (lower part). We observe that the efficiency of both methods increases with the sample size n and decreases with p. Interestingly, R-FSIR seems to be more sensitive to increasing p for the Gaussian case than FSIR. However, for the Cauchy distribution, R-FSIR outperforms FSIR and is more robust to the changes on p. As far as the number of slices, we observe that both methods are not sensitive to the choice of H. Finally, R-FSIR is more robust to outliers than FSIR, especially for the Cauchy case; see Tables 2 and 4.

S4 Estimation of structural dimension

In order to investigate the performance of the CVBIC order-determination procedure (S2.10), we use Model II, where X is simulated as described in Section 6, ϵ is generated according to a standard normal distribution, and the scores follow a Gaussian and a Cauchy distribution. Note that $K = 1$ for Model II. Table 5 reports the number of times, over the 100 simulation results, the CVBIC correctly estimates K.

Table 1: Mean (and standard deviation) of the multiple correlation for Model II, Gaussian distributed scores, with no outliers (upper) and with outliers added (lower) for Study 2

number of outliers	p	H	$n = 200$		$n = 400$		$n = 1000$	
			FSIR	R-FSIR	FSIR	R-FSIR	FSIR	R-FSIR
$m = 0$	10	5	0.91 (0.03)	0.89 (0.1)	0.96 (0.01)	0.96 (0.01)	0.98 (0.004)	0.98 (0.004)
		10	0.90 (0.03)	0.88 (0.06)	0.96 (0.01)	0.96 (0.01)	0.98 (0.004)	0.99 (0.004)
		20	0.87 (0.07)	0.84 (0.1)	0.96 (0.02)	0.96 (0.01)	0.98 (0.004)	0.99 (0.004)
	20	5	0.77 (0.07)	0.41 (0.24)	0.90 (0.03)	0.89 (0.04)	0.96 (0.006)	0.96 (0.006)
		10	0.75 (0.08)	0.42 (0.22)	0.90 (0.02)	0.90 (0.03)	0.97 (0.006)	0.97 (0.005)
		20	0.64 (0.20)	0.44 (0.23)	0.90 (0.03)	0.89 (0.03)	0.97 (0.007)	0.97 (0.007)
$m = 40$	10	5	0.12 (0.09)	0.48 (0.21)	0.10 (0.07)	0.53 (0.27)	0.18 (0.07)	0.56 (0.27)
		10	0.15 (0.11)	0.60 (0.24)	0.10 (0.07)	0.72 (0.21)	0.18 (0.07)	0.56 (0.28)
		20	0.13 (0.10)	0.73 (0.22)	0.10 (0.08)	0.89 (0.13)	0.18 (0.08)	0.70 (0.29)
	20	5	0.20 (0.12)	0.25 (0.19)	0.20 (0.10)	0.65 (0.21)	0.07 (0.04)	0.67 (0.17)
		10	0.18 (0.13)	0.29 (0.19)	0.16 (0.09)	0.75 (0.17)	0.06 (0.05)	0.74 (0.14)
		20	0.15 (0.11)	0.32 (0.21)	0.13 (0.09)	0.87 (0.12)	0.06 (0.04)	0.84 (0.11)

S5 Neuroimaging data application

To illustrate the performance of the methodology we use an fMRI dataset, obtained from the ADHD-200 Consortium (http://fcon_1000.projects.nitrc.org/indi/adhd200/index.html), consisting of resting-state fMRI and anatomical datasets of children with and without ADHD aggregated across 8 independent imaging sites. For our analysis, we consider the resting-state fMRI of the New York University Child Study Center. This dataset includes 222 subjects, of which 99 are the controls and the rest are diagnosed with ADHD. The ADHD group is further divided into the

Table 2: Mean (and standard deviation) of the multiple correlation for Model II, Cauchy distributed scores, with no outliers (upper) and with outliers added (lower) for Study 2

number of outliers	p	H	$n = 200$		$n = 400$		$n = 1000$	
			FSIR	R-FSIR	FSIR	R-FSIR	FSIR	R-FSIR
$m = 0$	10	5	0.14 (0.1)	0.90 (0.19)	0.14 (0.08)	0.89 (0.18)	0.14 (0.09)	0.91 (0.18)
		10	0.12 (0.1)	0.87 (0.19)	0.14 (0.11)	0.83 (0.24)	0.12 (0.10)	0.85 (0.22)
		20	0.13 (0.13)	0.77 (0.27)	0.13 (0.12)	0.81 (0.27)	0.12 (0.12)	0.80 (0.24)
	20	5	0.10 (0.06)	0.85 (0.23)	0.11 (0.07)	0.88 (0.22)	0.09 (0.06)	0.91 (0.16)
		10	0.10 (0.08)	0.87 (0.20)	0.09 (0.07)	0.88 (0.20)	0.10 (0.08)	0.88 (0.20)
		20	0.11 (0.10)	0.81 (0.26)	0.08 (0.08)	0.83 (0.24)	0.09 (0.10)	0.81 (0.27)
$m = 40$	10	5	0.59 (0.30)	0.89 (0.19)	0.67 (0.29)	0.89 (0.19)	0.67 (0.28)	0.89 (0.21)
		10	0.69 (0.33)	0.85 (0.24)	0.64 (0.33)	0.87 (0.22)	0.60 (0.35)	0.83 (0.24)
		20	0.66 (0.36)	0.81 (0.26)	0.70 (0.31)	0.81 (0.25)	0.74 (0.33)	0.85 (0.21)
	20	5	0.59 (0.29)	0.87 (0.23)	0.62 (0.31)	0.87 (0.21)	0.67 (0.30)	0.90 (0.19)
		10	0.61 (0.34)	0.90 (0.19)	0.67 (0.31)	0.86 (0.23)	0.70 (0.30)	0.84 (0.24)
		20	0.66 (0.34)	0.80 (0.27)	0.68 (0.32)	0.83 (0.23)	0.73 (0.33)	0.85 (0.22)

ADHD Combined group (77 subjects), the ADHD Inattentive group (44 subjects) and the ADHD Hyperactive group (2 subjects); we use the 77 subjects in the ADHD Combined group for our analysis. Moreover, 5 subjects were removed from the ADHD Combined group because of significant amount of missing observations, resulting in $n = 72$ subjects. Technical details regarding the sample and the scanning parameters can be found at the ADHD-200 Consortium.

The dataset was preprocessed by the NeuroBureau community using the Athena pipeline. 116 brain regions-of-interest (ROI) were constructed

Table 3: Mean (and standard deviation) of the multiple correlation for Model IV, Gaussian distributed scores, with no outliers (upper) and with outliers added (lower) for Study 2

number of outliers	p	H	$n = 200$		$n = 400$		$n = 1000$	
			FSIR	R-FSIR	FSIR	R-FSIR	FSIR	R-FSIR
$m = 0$	10	5	1.61 (0.1)	1.63 (0.14)	1.78 (0.06)	1.81 (0.06)	1.90 (0.03)	1.91 (0.02)
		10	1.70 (0.07)	1.70 (0.10)	1.84 (0.04)	1.85 (0.04)	1.93 (0.02)	1.93 (0.02)
		20	1.67 (0.12)	1.60 (0.20)	1.84 (0.04)	1.85 (0.04)	1.93 (0.02)	1.94 (0.02)
	20	5	1.41 (0.12)	1.35 (0.18)	1.63 (0.07)	1.64 (0.09)	1.81 (0.04)	1.83 (0.03)
		10	1.50 (0.08)	1.33 (0.18)	1.72 (0.05)	1.73 (0.05)	1.87 (0.03)	1.87 (0.03)
		20	1.41 (0.13)	1.21 (0.19)	1.73 (0.04)	1.71 (0.05)	1.87 (0.02)	1.88 (0.02)
$m = 40$	10	5	1.16 (0.12)	1.42 (0.20)	1.18 (0.08)	1.53 (0.15)	1.24 (0.05)	1.74 (0.09)
		10	1.15 (0.16)	1.24 (0.19)	1.20 (0.09)	1.46 (0.20)	1.27 (0.04)	1.77 (0.06)
		20	1.10 (0.16)	1.18 (0.18)	1.19 (0.11)	1.48 (.19)	1.27 (0.05)	1.79 (0.06)
	20	5	0.91 (0.10)	1.20 (0.17)	0.98 (0.08)	1.25 (0.18)	1.08 (0.05)	1.58 (0.1)
		10	0.91 (0.11)	0.53 (0.39)	1.02 (0.10)	1.24 (0.17)	1.11 (0.04)	1.60 (0.09)
		20	0.88 (0.10)	0.62 (0.40)	1.00 (0.11)	1.25 (0.16)	1.13 (0.05)	1.64 (0.08)

for the preprocessed resting-state fMRI using the anatomical labelling atlas (AAL) developed by Craddock et al. (2012). fMRI time series were extracted for each of the 116 regions by averaging all voxels time series within each region at each time point, resulting in 172 time points for each of the 116 regions for each subject. Hence, for each subject we have 116 different regional fMRI time series, observed at 172 time points. The AAL atlas and the regional fMRI time series are publicly available at NITRC (www.nitrc.org).

Table 4: Mean (and standard deviation) of the multiple correlation for Model IV, Cauchy distributed scores, with no outliers (upper) and with outliers added (lower) for Study 2

number outliers	p	H	$n = 200$		$n = 400$		$n = 1000$	
			FSIR	R-FSIR	FSIR	R-FSIR	FSIR	R-FSIR
$m = 0$	10	5	0.17 (0.07)	1.55 (0.36)	0.17 (0.07)	1.52 (0.37)	0.18 (0.10)	1.61 (0.36)
		10	0.16 (0.09)	1.60 (0.36)	0.15 (0.08)	1.63 (0.33)	0.16 (0.07)	1.62 (0.34)
		20	0.18 (0.12)	1.53 (0.37)	0.18 (0.12)	1.63 (0.33)	0.17 (0.10)	1.70 (0.31)
	20	5	0.09 (0.04)	1.50 (0.39)	0.09 (0.04)	1.60 (0.37)	0.08 (0.04)	1.54 (0.36)
		10	0.10 (0.05)	1.52 (0.36)	0.09 (0.06)	1.55 (0.38)	0.08 (0.05)	1.60 (0.34)
		20	0.11 (0.08)	1.54 (0.37)	0.08 (0.06)	1.62 (0.33)	0.09 (0.06)	1.57 (0.37)
$m = 40$	10	5	0.17 (0.01)	1.59 (0.34)	0.14 (0.05)	1.58 (0.36)	0.16 (0.08)	1.58 (0.35)
		10	0.19 (0.15)	1.59 (0.38)	0.15 (0.07)	1.61 (0.35)	0.17 (0.09)	1.63 (0.32)
		20	0.16 (0.11)	1.60 (0.35)	0.14 (0.1)	1.60 (0.34)	0.14 (0.09)	1.63 (0.35)
	20	5	0.10 (0.08)	1.46 (0.38)	0.08 (0.03)	1.55 (0.36)	0.08 (0.03)	1.50 (0.38)
		10	0.11 (0.08)	1.55 (0.37)	0.09 (0.05)	1.64 (0.35)	0.07 (0.04)	1.65 (0.35)
		20	0.11 (0.1)	1.56 (0.38)	0.08 (0.05)	1.62 (0.36)	0.07 (0.05)	1.60 (0.33)

Table 5: % of correct order determination by FSIR and R-FSIR for Gaussian and Cauchy distributed scores with no outliers (upper) and with outliers added (lower) for Study 3

	Gaussian		Cauchy	
	FSIR	R-FSIR	FSIR	R-FSIR
$m = 0$	82	72	84	85
$m = 40$	74	92	84	79

Figure 1 shows the smoothed spline fMRI curves with outliers in red, as detected by the magnitude-shape plot of Dai and Genton (2018), implemented using the R package `fdaoutlier`. Figure 2 shows the boxplots of the first two principal components for two regions of interest. It is evident

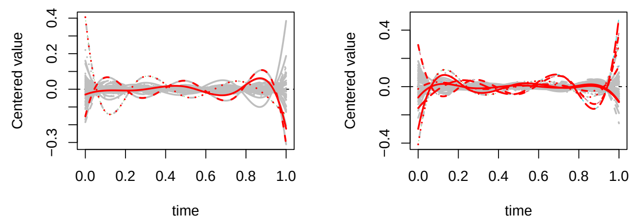

Figure 1: The smoothed fMRI curves, where the red curves represent outliers, for two brain regions of interest (left: Precentral right region, right: Superior frontal gyrus orbital right region).

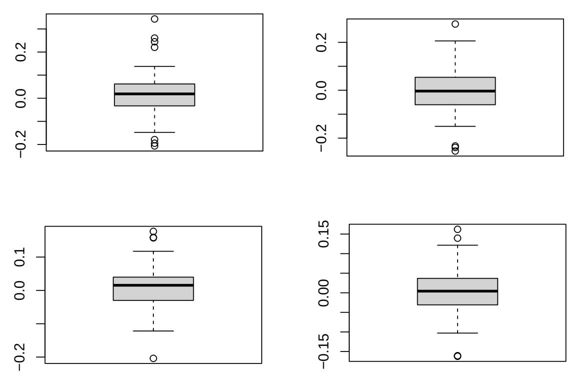

Figure 2: Boxplots of the first two scores for two brain regions of interest (left: Precentral right region, right: Superior frontal gyrus orbital right region).

that the marginal distributions are heavy-tailed.

Bibliography

Chen, X., J. Zhang, and W. Zhou (2022). High-dimensional elliptical sliced inverse regression in non-gaussian distributions. *Journal of Business & Economic Statistics 40*(3), 1204–1215.

Craddock, R. C., G. A. James, P. E. Holtzheimer, X. P. Hu, and H. S. Mayberg (2012). A whole brain fmri atlas generated via spatially constrained spectral clustering. *Human brain mapping 33*(8), 1914–1928.

Dai, W. and M. G. Genton (2018). Multivariate functional data visualization and outlier detection. *Journal of Computational and Graphical Statistics 27*(4), 923–934.

Fukumizu, K., F. R. Bach, and A. Gretton (2007). Statistical consistency of kernel canonical correlation analysis. *Journal of Machine Learning Research 8*(2), 361–383.

Han, F. and H. Liu (2018). Eca: High-dimensional elliptical component analysis in non-gaussian distributions. *Journal of the American Statistical Association 113*(521), 252–268.

Horn, R. A. and C. R. Johnson (1985). *Matrix Analsysis.* Cambridge University Press.

Hsing, T. and R. J. Carroll (1992). An asymptotic theory for sliced inverse regression. *The Annals of Statistics 20*(2), 1040–1061.

Li, B., A. Artemiou, and L. Li (2011). Principal support vector machines for linear and nonlinear sufficient dimension reduction. *The Annals of Statistics 39*(6), 3182–3210.

Li, B. and E. Solea (2018). A nonparametric graphical model for functional data with application to brain networks based on fmri. *Journal of the American Statistical Association 113*(524), 1637–1655.

Li, B. and J. Song (2022). Dimension reduction for functional data based on weak conditional moments. *The Annals of Statistics 50*(1), 107–128.

Solea, E. and B. Li (2022). Copula gaussian graphical models for functional data. *Journal of the American Statistical Association 117*(538), 781–793.

Wang, G., S. Liu, F. Han, and C.-Z. Di (2022). Robust functional principal component analysis via a functional pairwise spatial sign operator. *Biometrics, https://doi.org/10.1111/biom.13695*.

Zhu, L.-X. and K. W. Ng (1995). Asymptotics of sliced inverse regression. *Statistica Sinica 5*(2), 727–736.