Supplement to ‘On the Optimality of Functional Sliced Inverse Regression’

Rui Chen[a], Songtao Tian[b,*], Dongming Huang[c], Qian Lin[d], Jun S. Liu[e,**]

[a]*Department of Statistics and Data Science, Tsinghua University*
[b]*Department of Mathematical Sciences, Tsinghua University*
[c]*Department of Statistics and Data Science, National University of Singapore*
[d]*Department of Statistics and Data Science, Tsinghua University; Beijing Academy of Artificial Intelligence, Beijing, 100084, China*
[e]*Department of Statistics, Harvard University, 02138*

Section A - G contain the proofs of lemmas, propositions and theorems. Section H contains additional simulation results of Section 4.

A. Proof of Lemma 1

The proof follows the same argument as in the proof the (Huang et al., 2023, Theorem 1) and we only need to generalize the multivariate result therein to a functional version. We omit the proof here for simplicity.

B. Proof of Proposition 1

Suppose H is any integer greater than the constant K in Assumption 2.

*Co-first author.
**Corresponding author.

For any $\boldsymbol{u} \in \mathbb{S}_{\mathcal{H}}$, it holds that

$$\begin{aligned}
\langle \Gamma_e(\boldsymbol{u}), \boldsymbol{u} \rangle &= \int (\mathbb{E}[\langle \boldsymbol{X}, \boldsymbol{u} \rangle \mid Y = y])^2 dP_Y(y) = \sum_{h:\mathcal{S}_h \in \mathfrak{S}_H(n)} \int_{\mathcal{S}_h} (\mathbb{E}[\langle \boldsymbol{X}, \boldsymbol{u} \rangle \mid Y = y])^2 dP_Y(y) \\
&= \sum_{h:\mathcal{S}_h \in \mathfrak{S}_H(n)} \mathbb{P}(Y \in \mathcal{S}_h) \mathbb{E}\left((\mathbb{E}[\langle \boldsymbol{X}, \boldsymbol{u} \rangle \mid Y = y])^2 \mid Y \in \mathcal{S}_h \right) \\
&= \sum_{h:\mathcal{S}_h \in \mathfrak{S}_H(n)} \delta_h \mathbb{E}[\boldsymbol{m}^2(\boldsymbol{u}) \mid Y \in \mathcal{S}_h]
\end{aligned} \tag{B.1}$$

where $\delta_h := \mathbb{P}(Y \in \mathcal{S}_h)$ and $\mathbf{m}(\boldsymbol{u}) := \langle \mathbf{m}(Y), \boldsymbol{u} \rangle$. Furthermore, it holds that

$$\langle (\overline{\mathbf{m}}_h \otimes \overline{\mathbf{m}}_h(\boldsymbol{u}), \boldsymbol{u} \rangle = \mathbb{E}^2[\mathbf{m}(\boldsymbol{u}) \mid Y \in \mathcal{S}_h] = \mathbb{E}[\mathbf{m}^2(\boldsymbol{u}) \mid Y \in \mathcal{S}_h] - \mathrm{var}(\mathbf{m}(\boldsymbol{u}) \mid Y \in \mathcal{S}_h). \tag{B.2}$$

For any γ-partition $\mathfrak{S}_H(n) := \{\mathcal{S}_h, h = 1, .., H\}$, it holds that

$$\begin{aligned}
&\left| \left\langle \left(\widetilde{\Gamma}_e - \Gamma_e \right)(\boldsymbol{u}), \boldsymbol{u} \right\rangle \right| = \left| \frac{1}{H} \sum_{h:\mathcal{S}_h \in \mathfrak{S}_H(n)} \langle (\overline{\mathbf{m}}_h \otimes \overline{\mathbf{m}}_h(\boldsymbol{u}), \boldsymbol{u} \rangle - \sum_{h:\mathcal{S}_h \in \mathfrak{S}_H(n)} \delta_h \mathbb{E}[\boldsymbol{m}^2(\boldsymbol{u}) \mid Y \in \mathcal{S}_h] \right| \\
\leqslant& \left| \sum_{h:\mathcal{S}_h \in \mathfrak{S}_H(n)} (\frac{1}{H} - \delta_h) \mathbb{E}(\mathbf{m}^2(\boldsymbol{u}) \mid Y \in \mathcal{S}_h) \right| + \left| \frac{1}{H} \sum_{h:\mathcal{S}_h \in \mathfrak{S}_H(n)} \mathrm{var}(\mathbf{m}(\boldsymbol{u}) \mid Y \in \mathcal{S}_h) \right| \\
\leqslant& \frac{1}{\tau - 1} \left| \sum_{h:\mathcal{S}_h \in \mathfrak{S}_H(n)} \delta_h \mathbb{E}[\mathbf{m}^2(\boldsymbol{u}) \mid Y \in \mathcal{S}_h] \right| + \left| \frac{1}{H} \sum_{h:\mathcal{S}_h \in \mathfrak{S}_H(n)} \mathrm{var}(\mathbf{m}(\boldsymbol{u}) \mid Y \in \mathcal{S}_h) \right| \\
\leqslant& \frac{2}{\tau} \langle \Gamma_e(\boldsymbol{u}), \boldsymbol{u} \rangle + \left| \frac{1}{H} \sum_{h:\mathcal{S}_h \in \mathfrak{S}_H(n)} \mathrm{var}(\mathbf{m}(\boldsymbol{u}) \mid Y \in \mathcal{S}_h) \right| \\
\leqslant& \frac{3}{\tau} \langle \Gamma_e(\boldsymbol{u}), \boldsymbol{u} \rangle,
\end{aligned}$$

where the first inequality is due to (B.2), the second inequality is because $1 - \gamma \leqslant H\delta_h$ since $\mathfrak{S}_H(n)$ is a γ-partition, the third is because $\gamma \leqslant \frac{1}{\tau}$, and the last is due to the Assumption 2.

Recall Lemma 18, which states that there is some H' and C such that for any $H > H'$

and $n > \frac{4H}{\gamma} + 1$, the sample sliced partition is a γ-partition with probability at least

$$1 - CH^2\sqrt{n+1}\exp\left(-\frac{\gamma^2(n+1)}{32H^2}\right)$$

for some absolute constant C. Then the proof is completed by choosing $H_0 := \max\{H', K\}$.

□

C. Proof of Lemma 2

By Proposition 1, the first statement in Lemma 2 is a direct corollary of the second one. Now we begin to prove the second statement.

Define an event $\mathtt{E}$ as follows:

$$\mathtt{E} := \{\mathfrak{S}_H(n) \text{ is a } \gamma\text{-partition}\}. \tag{C.1}$$

From the proof of Proposition 1, we know that

$$\mathbb{P}(\mathtt{E}) \geqslant 1 - CH^2\sqrt{n+1}\exp\left(-\frac{\gamma^2(n+1)}{32H^2}\right).$$

and we have the following lemma.

Lemma 3. *Under WSSC, on the event* $\mathtt{E}$*, the following holds:*

$$\left|\left\langle\left(\widetilde{\Gamma}_e - \Gamma_e\right)(\boldsymbol{u}), \boldsymbol{u}\right\rangle\right| \leqslant \frac{3}{\tau}\left\langle\Gamma_e(\boldsymbol{u}), \boldsymbol{u}\right\rangle, \forall \boldsymbol{u} \in \mathbb{S}_{\mathcal{H}}.$$

The rest of our effort will be devoted to establishing the following lemma, which directly implies the second statement in Lemma 2.

Lemma 4. *Suppose that Assumptions 2 and 3 hold. For any fixed integer $H > H_0$ (H_0 is defined in Proposition 1) and any sufficiently large $n > 1 + 4H/\gamma$, we have*

$$\mathbb{E}\left[1_{\mathbb{E}}\left\|\widehat{\Gamma}_e - \widetilde{\Gamma}_e\right\|^2\right] \lesssim \frac{H^2}{n}.$$

Proof of Lemma 4. For any $\boldsymbol{u} \in \mathbb{S}_{\mathcal{H}}$, let $\boldsymbol{u} := \sum_{j\geqslant 1} b_j\phi_j$ with $\sum_{j\geqslant 1} b_j^2 = 1$ where $\{\phi_j\}_{j=1}^{\infty}$ is the eigenfunctions of Γ. Then we have

$$\begin{aligned}
\langle(\widehat{\Gamma}_e - \widetilde{\Gamma}_e)\boldsymbol{u}, \boldsymbol{u}\rangle &= \sum_{i,j\geqslant 1} b_i b_j \langle(\widehat{\Gamma}_e - \widetilde{\Gamma}_e)\phi_i, \phi_j\rangle \\
&\leqslant \left(\sum_{i,j\geqslant 1} b_i^2 b_j^2\right)^{\frac{1}{2}} \left(\sum_{i,j\geqslant 1} \langle(\widehat{\Gamma}_e - \widetilde{\Gamma}_e)\phi_i, \phi_j\rangle^2\right)^{\frac{1}{2}} \\
&= \left(\sum_{i,j\geqslant 1} \langle(\widehat{\Gamma}_e - \widetilde{\Gamma}_e)\phi_i, \phi_j\rangle^2\right)^{\frac{1}{2}}.
\end{aligned}$$

It shows that $\left\|\widehat{\Gamma}_e - \widetilde{\Gamma}_e\right\| \leqslant \left(\sum_{i,j\geqslant 1} \langle(\widehat{\Gamma}_e - \widetilde{\Gamma}_e)\phi_i, \phi_j\rangle^2\right)^{\frac{1}{2}}$.

Let $\xi_{hj} := \langle \overline{\boldsymbol{m}}_h, \phi_j\rangle$ and $\widehat{\xi}_{hj} := \langle \overline{\boldsymbol{X}}_{h,\cdot}, \phi_j\rangle$. The operators $\widehat{\Gamma}_e$ and $\widetilde{\Gamma}_e$ can be written as

$$\begin{aligned}
\widehat{\Gamma}_e =& \frac{1}{H} \sum_{h:\mathcal{S}_h \in \mathfrak{S}_H(n)} \left(\sum_{j\geqslant 1} \widehat{\xi}_{hj}\phi_j\right) \otimes \left(\sum_{j\geqslant 1} \widehat{\xi}_{hj}\phi_j\right), \\
\widetilde{\Gamma}_e =& \frac{1}{H} \sum_{h:\mathcal{S}_h \in \mathfrak{S}_H(n)} \left(\sum_{j\geqslant 1} \xi_{hj}\phi_j\right) \otimes \left(\sum_{j\geqslant 1} \xi_{hj}\phi_j\right),
\end{aligned}$$

respectively. Now we obtain

$$\sum_{i,j\geqslant 1}\langle(\widehat{\Gamma}_e-\widetilde{\Gamma}_e)\phi_i,\phi_j\rangle^2\leqslant\frac{1}{H}\sum_{i,j\geqslant 1}\sum_{h:\mathcal{S}_h\in\mathfrak{S}_H(n)}\left(\widehat{\xi}_{hi}\widehat{\xi}_{hj}-\xi_{hi}\xi_{hj}\right)^2$$
$$=\frac{1}{H}\sum_{i,j\geqslant 1}\sum_{h:\mathcal{S}_h\in\mathfrak{S}_H(n)}\left((\widehat{\xi}_{hi}-\xi_{hi})(\widehat{\xi}_{hj}-\xi_{hj})+\xi_{hj}(\widehat{\xi}_{hi}-\xi_{hi})+\xi_{hi}(\widehat{\xi}_{hj}-\xi_{hj})\right)^2.$$

The lemma is proved if we can show that for any h,

$$A:=\mathbb{E}\left[1_{\mathtt{E}}\sum_{i,j\geqslant 1}\left((\widehat{\xi}_{hi}-\xi_{hi})(\widehat{\xi}_{hj}-\xi_{hj})+\xi_{hj}(\widehat{\xi}_{hi}-\xi_{hi})+\xi_{hi}(\widehat{\xi}_{hj}-\xi_{hj})\right)^2\right]\leqslant\frac{C'H^2}{n}$$

for some positive constant C'.

Note that

$$\begin{aligned}&\mathbb{E}\left[1_{\mathtt{E}}\sum_{i,j\geqslant 1}\left((\widehat{\xi}_{hi}-\xi_{hi})(\widehat{\xi}_{hj}-\xi_{hj})+\xi_{hj}(\widehat{\xi}_{hi}-\xi_{hi})+\xi_{hi}(\widehat{\xi}_{hj}-\xi_{hj})\right)^2\right]\\
\leqslant&3\mathbb{E}\left[1_{\mathtt{E}}\sum_{i,j\geqslant 1}(\widehat{\xi}_{hi}-\xi_{hi})^2(\widehat{\xi}_{hj}-\xi_{hj})^2+\xi_{hj}^2(\widehat{\xi}_{hi}-\xi_{hi})^2+\xi_{hi}^2(\widehat{\xi}_{hj}-\xi_{hj})^2\right]\\
=&3\mathbb{E}\left[1_{\mathtt{E}}\sum_{i,j\geqslant 1}(\widehat{\xi}_{hi}-\xi_{hi})^2(\widehat{\xi}_{hj}-\xi_{hj})^2\right]+6\mathbb{E}\left[1_{\mathtt{E}}\sum_{i,j\geqslant 1}\xi_{hj}^2(\widehat{\xi}_{hi}-\xi_{hi})^2\right]\\
:=&3I+6II.\end{aligned}$$

We first bound the term II. Recall that $\langle\boldsymbol{X},\phi_j\rangle=\xi_j,\forall j$. Then there exists a constant $C_1>0$, such that for all $j\geqslant 1$, we have

$$1_{\mathtt{E}}\xi_{hj}^2=1_{\mathtt{E}}\mathbb{E}^2[\xi_j|Y\in\mathcal{S}_h]\overset{(a)}{\leqslant}1_{\mathtt{E}}\mathbb{E}[\xi_j^2|Y\in\mathcal{S}_h]\overset{(b)}{\leqslant}1_{\mathtt{E}}\mathbb{E}^{1/2}[\xi_j^4|Y\in\mathcal{S}_h]\overset{(c)}{\leqslant}C_1\sqrt{H}\mathbb{E}^{1/2}[\xi_j^4]\overset{(d)}{\leqslant}C\sqrt{H}\lambda_j \tag{C.2}$$

where we have used Jensen inequality for conditional expectation in (a) and (b), Lemma 19 in (c), and Assumption 3 in (d).

Assume $\boldsymbol{X}_{h,j} = \sum_{i=1}^{\infty} \xi_{h,j,i}\phi_i$, then one has

$$
\begin{aligned}
\mathbb{E}[1_{\mathtt{E}}(\widehat{\xi}_{hi} - \xi_{hi})^2] =& \mathbb{E}[1_{\mathtt{E}}\langle \frac{1}{c}\sum_{j=1}^{c} \boldsymbol{X}_{h,j} - \mathbb{E}[\boldsymbol{X}|Y \in \mathcal{S}_h], \phi_i\rangle^2] \\
=& \mathbb{E}[1_{\mathtt{E}}(\frac{1}{c}\sum_{j=1}^{c} \xi_{h,j,i} - \mathbb{E}[\xi_i|Y \in \mathcal{S}_h])^2] \\
=& \mathbb{E}[1_{\mathtt{E}}(\frac{1}{c-1}\sum_{j=1}^{c-1} \xi_{h,j,i} - \mathbb{E}[1_{\mathtt{E}}\xi_i|Y \in \mathcal{S}_h] + \frac{\xi_{h,c,i}}{c} - \frac{1}{c(c-1)}\sum_{j=1}^{c-1} \xi_{h,j,i})^2] \\
\lesssim& \mathbb{E}[1_{\mathtt{E}}(\frac{1}{c-1}\sum_{j=1}^{c-1} \xi_{h,j,i} - \mathbb{E}[1_{\mathtt{E}}\xi_i|Y \in \mathcal{S}_h])^2] + \frac{1}{c^2}\mathbb{E}[1_{\mathtt{E}}\xi_{h,c,i}^2] + \frac{1}{c^2(c-1)}\mathbb{E}[1_{\mathtt{E}}\sum_{j=1}^{c-1} \xi_{h,j,i}^2] \\
\overset{(a)}{\leqslant}& \frac{1}{c-1}\mathbb{E}[1_{\mathtt{E}}\xi_{h,1,i}^2] + \frac{1}{c^2}\mathbb{E}[1_{\mathtt{E}}\xi_{h,1,i}^2] + \frac{1}{c^2}\mathbb{E}[1_{\mathtt{E}}\xi_{h,1,i}^2] \\
\lesssim& \frac{1}{c-1}\mathbb{E}[1_{\mathtt{E}}\xi_{h,1,i}^2] \overset{(b)}{\leqslant} \frac{1}{c-1}\mathbb{E}^{1/2}[1_{\mathtt{E}}\xi_{h,1,i}^4] \\
\overset{(c)}{\lesssim}& \frac{1}{c-1}\sqrt{H}\mathbb{E}^{1/2}[\xi_i^4] \overset{(d)}{\leqslant} \frac{1}{c-1}\sqrt{H}\lambda_i \leqslant C_1' H^{3/2}\lambda_i/n,
\end{aligned}
$$

where we have used Jensen inequality for conditional expectation in (b), Lemma 19 in (c), and Assumption 3 in (d). Furthermore, (a) is based on Lemma 17 and the following derivation:

$$
\begin{aligned}
&\mathbb{E}[(\frac{1}{c-1}\sum_{j=1}^{c-1}\xi_{h,j,i}-\mathbb{E}[\xi_i|Y\in\mathcal{S}_h])^2] \\
&=\mathbb{E}\left[\mathbb{E}[(\frac{1}{c-1}\sum_{j=1}^{c-1}\xi_{h,j,i}-\mathbb{E}[\xi_i|Y\in\mathcal{S}_h])^2 \mid \{\mathcal{S}_{h'}\}_{h'=1}^{H}]\right] \\
&=\mathbb{E}\left[\frac{1}{c-1}\mathrm{var}[\xi_{h,1,i}\mid\{\mathcal{S}_{h'}\}_{h'=1}^{H}]\mid\{\mathcal{S}_{h'}\}_{h'=1}^{H}\right] \\
&\leqslant \frac{1}{c-1}\mathbb{E}[\xi_{h,1,i}^2].
\end{aligned}
$$

Thus

$$
II \leqslant \sum_{i,j\geqslant 1}\frac{C_1H^2\lambda_i\lambda_j}{n} \tag{C.3}
$$

for some sufficiently large C_1.

Next we handle the term I. From (C.2) and Cauchy-Schwarz inequality, one has: $1_{\mathtt{E}}\left|\mathbb{E}[\xi_i^3|Y\in\mathcal{S}_h]\right|\leqslant CH^{3/4}\lambda_i^{3/2}$. By direct calculation of fourth moment (see, e.g., (Angelova, 2012, Theorem 1)), one has

$$
\mathbb{E}[1_{\mathtt{E}}(\widehat{\xi}_{hi}-\xi_{hi})^4]\leqslant C\mathbb{E}[1_{\mathtt{E}}(\frac{1}{c-1}\sum_{j=1}^{c-1}\xi_{h,j,i}-\mathbb{E}[\xi_i|Y\in\mathcal{S}_h])^4]\leqslant C\frac{H^3\lambda_i^2}{n^2}\leqslant C\frac{H^2\lambda_i^2}{n}.
$$

We get

$$
\begin{aligned}
I =&\mathbb{E}\left[1_{\mathtt{E}}\sum_{i,j\geqslant 1}(\widehat{\xi}_{hi}-\xi_{hi})^2(\widehat{\xi}_{hj}-\xi_{hj})^2\right] \\
&\leqslant \sum_{i,j\geqslant 1}\mathbb{E}^{\frac{1}{2}}\left[1_{\mathtt{E}}(\widehat{\xi}_{hi}-\xi_{hi})^4\right]\mathbb{E}^{\frac{1}{2}}\left[1_{\mathtt{E}}(\widehat{\xi}_{hj}-\xi_{hj})^4\right] \\
&\leqslant \sum_{i,j\geqslant 1}\frac{C_2H^2\lambda_i\lambda_j}{n}.
\end{aligned} \tag{C.4}
$$

Since Γ has a finite trace, we can take $C'=6\sum_{i,j\geqslant 1}(C_1+C_2)\lambda_i\lambda_j<\infty$. Then by Equations

(C.3) and (C.4), we have

$$A \leqslant \frac{C'H^2}{n}$$

as required. □

D. Proof of Theorem 1

We first provide the following lemma.

Lemma 5. *Suppose Assumption 2 holds with* $\tau > \frac{6\|\Gamma_e\|}{\lambda_{\min}^+(\Gamma_e)}$*. On the event* $\mathtt{E}$*, we have* $\mathrm{Im}(\Gamma_e) = \mathrm{Im}(\widetilde{\Gamma}_e)$ *and* $\lambda_{\min}^+(\widetilde{\Gamma}_e) = \lambda_d(\widetilde{\Gamma}_e) \geqslant \frac{\lambda_d(\Gamma_e)}{2}$*.*

Proof. Recall that $\widetilde{\Gamma}_e := \frac{1}{H}\sum_{h:\mathcal{S}_h \in \mathfrak{S}_H(n)} \overline{\boldsymbol{m}}_h \otimes \overline{\boldsymbol{m}}_h$ and $\overline{\boldsymbol{m}}_h := \mathbb{E}[\boldsymbol{m}(Y) \mid Y \in \mathcal{S}_h] = \mathbb{E}[\boldsymbol{X}|Y \in \mathcal{S}_h]$. Note that $\overline{\boldsymbol{m}}_h \in \mathcal{S}_e = \mathrm{Im}(\Gamma_e)$ for all $h = 1, \dots, H$. Thus, $\mathrm{Im}(\widetilde{\Gamma}_e) \subseteq \mathrm{Im}(\Gamma_e)$.

Next we prove by contradiction that $\mathrm{Im}(\Gamma_e) \subseteq \mathrm{Im}(\widetilde{\Gamma}_e)$. Assume that there exists a vector $\boldsymbol{u} \in \mathcal{S}_{\mathcal{H}}$ such that $\Gamma_e \boldsymbol{u} \neq 0$ and $\widetilde{\Gamma}_e \boldsymbol{u} = 0$. Since $\tau > \frac{6\|\Gamma_e\|}{\lambda_{\min}^+(\Gamma_e)}$ and the event $\mathtt{E}$ happens, we have

$$\|\Gamma_e(\boldsymbol{u})\| = \|(\Gamma_e - \widetilde{\Gamma}_e)(\boldsymbol{u})\| \leqslant \frac{3}{\tau}\|\Gamma_e\| \leqslant \frac{\lambda_{\min}^+(\Gamma_e)}{2},$$

where the first inequality comes from Lemma 3. This is a contradiction because $0 < \lambda_{\min}^+(\Gamma_e) \leqslant \|\Gamma_e(\boldsymbol{u})\|$. Thus $\mathrm{Im}(\Gamma_e) = \mathrm{Im}(\widetilde{\Gamma}_e)$ and $\mathrm{rank}(\widetilde{\Gamma}_e) = \mathrm{rank}(\Gamma_e) = d$. By Lemma 21, we have

$$|\lambda_d(\widetilde{\Gamma}_e) - \lambda_d(\Gamma_e)| \leqslant \|\widetilde{\Gamma}_e - \Gamma\| \leqslant \frac{3}{\tau}\|\Gamma_e\| \leqslant \frac{\lambda_d(\Gamma_e)}{2}$$

where the second inequality comes from Proposition 1. $\square$

Under the condition of Theorem 1 with $\tau > \frac{6\|\Gamma_e\|}{\lambda_{\min}^+(\Gamma_e)}$ and the event $\mathtt{E}$, then $\mathrm{Im}(\widetilde{\Gamma}_e) = \mathrm{Im}(\Gamma_e)$ and thus $\mathrm{rank}(\widetilde{\Gamma}_e) = \mathrm{rank}(\Gamma_e) = d$. Let $\{\widehat{\mu}_i\}_{i=1}^d$ be the d largest eigenvalues of $\widehat{\Gamma}_e$ with associated eigenfunctions $\{\widehat{v}_i\}_{i=1}^d$. Define

$$\widehat{\Gamma}_e^d := \sum_{i=1}^d \widehat{\mu}_i \widehat{v}_i \otimes \widehat{v}_i. \tag{D.1}$$

We first prove that

$$\mathbb{E}\left[\left\|\widehat{\Gamma}_e^d - \widetilde{\Gamma}_e\right\|^2 1_{\mathtt{E}}\right] \lesssim \frac{H^2}{n}.$$

Using Lemma 21, one can get $\lambda_i^2\left(\widehat{\Gamma}_e\right) \leqslant 2\left(\left\|\widehat{\Gamma}_e - \widetilde{\Gamma}_e\right\|^2 + \lambda_i^2\left(\widetilde{\Gamma}_e\right)\right)$. Since $\mathrm{rank}(\widetilde{\Gamma}_e) = d$, one can get $\lambda_i(\widetilde{\Gamma}_e) = 0,\ i \geqslant d+1$. Thus by Lemma 4, one has $\mathbb{E}\left[1_{\mathtt{E}}\left\|\widehat{\Gamma}_e - \widetilde{\Gamma}_e\right\|^2\right] \lesssim \frac{H^2}{n}$, which implies

$$\mathbb{E}\left[\left|\lambda_i\left(\widehat{\Gamma}_e\right)\right|^2 1_{\mathtt{E}}\right] \lesssim \frac{H^2}{n} \quad (i \geqslant d+1).$$

Furthermore,

$$\begin{aligned}
&\mathbb{E}\left[\left\|\widehat{\Gamma}_e^d - \widetilde{\Gamma}_e\right\|^2 1_{\mathtt{E}}\right] \leqslant 2\left(\mathbb{E}\left[\left\|\widetilde{\Gamma}_e - \widehat{\Gamma}_e\right\|^2 1_{\mathtt{E}} + \left\|\widehat{\Gamma}_e - \widehat{\Gamma}_e^d\right\|^2 1_{\mathtt{E}}\right]\right)\\
=&2\mathbb{E}\left[\left\|\widetilde{\Gamma}_e - \widehat{\Gamma}_e\right\|^2 1_{\mathtt{E}} + \lambda_{d+1}^2(\widehat{\Gamma}_e) 1_{\mathtt{E}}\right] \lesssim \frac{H^2}{n}.
\end{aligned}$$

By Lemma 5, we have $\lambda_{\min}^+(\widetilde{\Gamma}_e) \geqslant \lambda_d(\Gamma_e)/2$.

By Markov inequality, we have

$$
\begin{aligned}
&\mathbb{P}(|\lambda_d(\widehat{\Gamma}_e^d) - \lambda_d(\widetilde{\Gamma}_e)|1_{\mathtt{E}} \leqslant \frac{\lambda_d(\widetilde{\Gamma}_e)}{2}) \geqslant \mathbb{P}(|\lambda_d(\widehat{\Gamma}_e^d) - \lambda_d(\widetilde{\Gamma}_e)|1_{\mathtt{E}} \leqslant \frac{\lambda_d(\Gamma_e)}{4}) \\
\geqslant & 1 - \frac{\mathbb{E}[|\lambda_d(\widehat{\Gamma}_e^d) - \lambda_d(\widetilde{\Gamma}_e)|^2 1_{\mathtt{E}}]}{\lambda_d^2(\Gamma_e)/16} \\
\geqslant & 1 - \frac{\mathbb{E}[\|\widehat{\Gamma}_e^d - \widetilde{\Gamma}_e\|^2 1_{\mathtt{E}}]}{\lambda_d^2(\Gamma_e)/16} \geqslant 1 - C'\frac{H^2}{n}
\end{aligned}
$$

for some constant $C' > 0$. Define $\widetilde{\mathtt{F}} := \{|\lambda_d(\widehat{\Gamma}_e^d) - \lambda_d(\widetilde{\Gamma}_e)|1_{\mathtt{E}} \leqslant \frac{\lambda_d(\widetilde{\Gamma}_e)}{2}\}$, we know under the event $\mathtt{E}\cap\widetilde{\mathtt{F}}$, it hold that $\lambda_d(\widehat{\Gamma}_e^d) \geqslant \lambda_d(\widetilde{\Gamma}_e)/2 \geqslant \lambda_d(\Gamma_e)/4$. Thus we have $\min\left\{\lambda_{\min}^+(\widehat{\Gamma}_e^d), \lambda_{\min}^+(\widetilde{\Gamma}_e)\right\} 1_{\mathtt{E}\cap\widetilde{\mathtt{F}}} \geqslant \lambda_d(\Gamma_e)/4$.

Then Applying $\sin\Theta$ theorem (Lemma 20), we have

$$
\mathbb{E}\left[\left\|P_{\widehat{\mathcal{S}}_e} - P_{\mathcal{S}_e}\right\|^2 1_{\mathtt{E}\cap\widetilde{\mathtt{F}}}\right] \lesssim \frac{\mathbb{E}\left[\left\|\widehat{\Gamma}_e^d - \widetilde{\Gamma}_e\right\|^2 1_{\mathtt{E}\cap\widetilde{\mathtt{F}}}\right]}{\min\left\{\lambda_{\min}^+(\widehat{\Gamma}_e^d), \lambda_{\min}^+(\widetilde{\Gamma}_e)\right\}^2 1_{\mathtt{E}\cap\widetilde{\mathtt{F}}}} \lesssim \frac{H^2}{n}.
$$

Thus

$$
\begin{aligned}
\mathbb{E}\left[\left\|P_{\widehat{\mathcal{S}}_e} - P_{\mathcal{S}_e}\right\|^2\right] =& \mathbb{E}\left[\left\|P_{\widehat{\mathcal{S}}_e} - P_{\mathcal{S}_e}\right\|^2 1_{\mathtt{E}\cap\widetilde{\mathtt{F}}}\right] + \mathbb{E}\left[\left\|P_{\widehat{\mathcal{S}}_e} - P_{\mathcal{S}_e}\right\|^2 1_{(\mathtt{E}\cap\widetilde{\mathtt{F}})^c}\right] \\
\lesssim & \frac{H^2}{n} + dCH^2\sqrt{n+1}\exp\left(-\frac{\gamma^2(n+1)}{32H^2}\right) \\
\lesssim & \frac{H^2}{n}.
\end{aligned}
$$

□

E. Proof of Theorem 2

By Markov inequality and $\mathbb{E}[|\widehat{\lambda}_j - \lambda_j|^2] \lesssim \frac{1}{n}$(see e.g., Equation (5.26) in Hall and Horowitz (2007)), we have

$$\mathbb{P}(|\widehat{\lambda}_j - \lambda_j| \leqslant \frac{\lambda_j}{2}) \geqslant 1 - \frac{\mathbb{E}[|\widehat{\lambda}_j - \lambda_j|^2]}{\lambda_j^2/4} \geqslant 1 - \frac{Cj^{2\alpha}}{n}.$$

Define

$$\mathtt{F} := \left\{\frac{\lambda_i}{2} \leqslant \widehat{\lambda}_i \leqslant \frac{3\lambda_i}{2}, \forall i \in [m]\right\} = \left\{|\widehat{\lambda}_i - \lambda_i| \leqslant \frac{\lambda_i}{2}, \forall i \in [m]\right\}. \tag{E.1}$$

Then we have $\mathbb{P}(\mathtt{F}) \geqslant 1 - \frac{m^{2\alpha+1}}{n}$.

We first need a preparatory theorem.

E.1 A preparatory theorem

Theorem 4. *With the same conditions as in Theorem 2, we have*

$$\mathbb{E}\left[\left\|\widehat{\Gamma}_m^{\dagger}\widehat{\Gamma}_e - \Gamma_m^{\dagger}\widetilde{\Gamma}_e\right\| 1_{\mathtt{E}\cap\mathtt{F}}\right] \lesssim Hn^{\frac{-(2\beta-1)}{2(\alpha+2\beta)}}.$$

Proof. We can decompose $\widehat{\Gamma}_m^{\dagger}\widehat{\Gamma}_e - \Gamma_m^{\dagger}\widetilde{\Gamma}_e$ as follows:

$$\begin{aligned}
&\widehat{\Gamma}_m^{\dagger}\widehat{\Gamma}_e - \Gamma_m^{\dagger}\widetilde{\Gamma}_e \\
=&\Gamma_m^{\dagger}\left(\widehat{\Gamma}_e - \widetilde{\Gamma}_e\right) + (\widehat{\Gamma}_m^{\dagger} - \Gamma_m^{\dagger})\widetilde{\Gamma}_e + \left(\widehat{\Gamma}_m^{\dagger} - \Gamma_m^{\dagger}\right)\left(\widehat{\Gamma}_e - \widetilde{\Gamma}_e\right) \\
:=&B_1 + B_2 + B_3.
\end{aligned} \tag{E.2}$$

Then our theorem is derived directly by the next proposition. $\square$

Proposition 2. *With the same conditions as in Theorem 2, we have*

(1). $\mathbb{E}[\|B_1\|^2 1_{\mathtt{E}}] = \mathbb{E}\left[\left\|\Gamma_m^{\dagger}\left(\widehat{\Gamma}_e - \widetilde{\Gamma}_e\right)\right\|^2 1_{\mathtt{E}}\right] \lesssim H^2 \frac{m^{\alpha+1}}{n} \lesssim H^2 n^{\frac{-(2\beta-1)}{(\alpha+2\beta)}}$;

(2). $\mathbb{E}[\|B_2\|^2 1_{\mathtt{E}}] = \mathbb{E}\left[\left\|\left(\widehat{\Gamma}_m^{\dagger} - \Gamma_m^{\dagger}\right)\widetilde{\Gamma}_e\right\|^2 1_{\mathtt{E}}\right] \lesssim H^2 \frac{m^{\alpha+1}}{n} \lesssim H^2 n^{\frac{-(2\beta-1)}{(\alpha+2\beta)}}$;

(3). $\mathbb{E}[\|B_3\| 1_{\mathtt{E}\cap\mathtt{F}}] = \mathbb{E}\left[\left\|\left(\widehat{\Gamma}_m^{\dagger} - \Gamma_m^{\dagger}\right)\left(\widehat{\Gamma}_e - \widetilde{\Gamma}_e\right)\right\| 1_{\mathtt{E}\cap\mathtt{F}}\right] \lesssim H \frac{m^{(\alpha+1)/2}}{\sqrt{n}} \lesssim H n^{\frac{-(2\beta-1)}{2(\alpha+2\beta)}}$.

Proof of Proposition 2-(1). For any $\boldsymbol{u} = \sum_{i\geqslant 1} b_i \phi_i \in \mathbb{S}_{\mathcal{H}}$ with $\sum_{i\geqslant 1} b_i^2 = 1$, we have

$$\begin{aligned}
\left\|\Gamma_m^{\dagger}\left(\widehat{\Gamma}_e - \widetilde{\Gamma}_e\right)\boldsymbol{u}\right\|^2 &= \sum_{j=1}^{\infty}\left\langle \Gamma_m^{\dagger}\left(\widehat{\Gamma}_e - \widetilde{\Gamma}_e\right)\boldsymbol{u}, \phi_j\right\rangle^2 \\
&= \sum_{j=1}^{\infty}\left(\sum_{i=1}^{\infty} b_i \left\langle \left(\widehat{\Gamma}_e - \widetilde{\Gamma}_e\right)(\phi_i), \Gamma_m^{\dagger}\phi_j\right\rangle\right)^2 \\
&\leqslant \sum_{j=1}^{\infty}\left(\sum_{i=1}^{\infty} b_i^2 \sum_{i=1}^{\infty}\left\langle \left(\widehat{\Gamma}_e - \widetilde{\Gamma}_e\right)(\phi_i), \Gamma_m^{\dagger}\phi_j\right\rangle^2\right) \\
&= \sum_{i,j\geqslant 1}\left\langle \left(\widehat{\Gamma}_e - \widetilde{\Gamma}_e\right)\phi_i, \Gamma_m^{\dagger}\phi_j\right\rangle^2 \\
&= \sum_{j=1}^{m}\frac{1}{\lambda_j^2}\sum_{i\geqslant 1}\left\langle \left(\widehat{\Gamma}_e - \widetilde{\Gamma}_e\right)\phi_i, \phi_j\right\rangle^2.
\end{aligned}$$

Then we obtain

$$\left\|\Gamma_m^{\dagger}\left(\widehat{\Gamma}_e - \widetilde{\Gamma}_e\right)\right\|^2 \leqslant \sum_{j=1}^{m}\frac{1}{\lambda_j^2}\sum_{i\geqslant 1}\left\langle \left(\widehat{\Gamma}_e - \widetilde{\Gamma}_e\right)\phi_i, \phi_j\right\rangle^2.$$

By the proof of Lemma 4, we see that there exists some constant $C > 0$, such that

$$\mathbb{E}\left[1_{\mathtt{E}}\sum_{j=1}^{m}\frac{1}{\lambda_j^2}\sum_{i\geqslant 1}\left\langle\left(\widehat{\Gamma}_e-\widetilde{\Gamma}_e\right)\phi_i,\phi_j\right\rangle^2\right]\leqslant\sum_{j=1}^{m}\frac{1}{\lambda_j^2}\sum_{i\geqslant 1}\frac{CH^2\lambda_i\lambda_j}{n}$$
$$=CH^2\sum_{j=1}^{m}\frac{1}{\lambda_j n}\sum_{i\geqslant 1}\lambda_i$$
$$\leqslant C'\frac{CH^2m^{\alpha+1}}{n},$$

where the last inequality comes from $\lambda_j \geqslant Cj^{-\alpha}$ by Assumption 4. It implies that $\mathbb{E}[1_{\mathtt{E}}\left\|B_1\right\|^2] \lesssim \frac{H^2m^{\alpha+1}}{n} \lesssim H^2 n^{\frac{-(2\beta-1)}{\alpha+2\beta}}$.

Proof of Proposition 2-(2). We can reformulate B_2 to

$$B_2=\left(\widehat{\Gamma}_m^{\dagger}-\Gamma_m^{\dagger}\right)\widetilde{\Gamma}_e=\left(\widehat{\Gamma}_m^{\dagger}-\Gamma_m^{\dagger}\right)\Gamma\Gamma^{-1}\widetilde{\Gamma}_e.$$

When $\mathtt{E}$ happens, it holds that $\mathrm{Im}(\Gamma_e)=\mathrm{Im}(\widetilde{\Gamma}_e)$. For any $\boldsymbol{u}\in\mathcal{H}$ with $\|\boldsymbol{u}\|=1$, we can write

$$\Gamma^{-1}\widetilde{\Gamma}_e(\boldsymbol{u})=\sum_{k=1}^{d}c_k\eta_k$$

with $|c_k| \leqslant \left\|\Gamma^{-1}\widetilde{\Gamma}_e\right\|$ for all $k=1,\ldots,d$ and $\{\eta_k\}_{k=1}^d$ are the generalized eigenfunctions of Γ_e associated with eigenvalues $\{\mu_k\}_{k=1}^d$ (i.e., $\Gamma_e\eta_k=\mu_k\Gamma\eta_k$).

We conclude that

$$\left\|B_2\right\|^2 1_{\mathtt{E}}\leqslant d\sum_{k=1}^{d}\left\|\Gamma^{-1}\widetilde{\Gamma}_e\right\|^2 1_{\mathtt{E}}\left\|(\widehat{\Gamma}_m^{\dagger}-\Gamma_m^{\dagger})\Gamma\eta_k\right\|^2 1_{\mathtt{E}}.$$

Now we need two lemmas:

Lemma 6.

$$\mathbb{E}\left[\left\|(\widehat{\Gamma}_m^\dagger - \Gamma_m^\dagger)\Gamma\eta_k\right\|^2 1_{\mathtt{E}}\right] \lesssim \frac{H^2 m^{\alpha+1}}{n}, \qquad k=1,\dots,d. \tag{E.3}$$

Proof. We first decompose $\left(\widehat{\Gamma}_m^\dagger - \Gamma_m^\dagger\right)\Gamma\eta_k$ as follows:

$$\begin{aligned}\left(\widehat{\Gamma}_m^\dagger - \Gamma_m^\dagger\right)\Gamma\eta_k &= \widehat{\Gamma}_m^\dagger(\widehat{\Gamma} + (\Gamma - \widehat{\Gamma}))(\widehat{\eta}_k + (\eta_k - \widehat{\eta}_k)) - \Gamma_m^\dagger\Gamma\eta_k \\ &= \widehat{\Gamma}_m^\dagger\widehat{\Gamma}\widehat{\eta}_k - \Gamma_m^\dagger\Gamma\eta_k + \widehat{\Gamma}_m^\dagger\widehat{\Gamma}(\eta_k - \widehat{\eta}_k) + \widehat{\Gamma}_m^\dagger(\Gamma - \widehat{\Gamma})\eta_k.\end{aligned} \tag{E.4}$$

Suppose $\eta_k = \sum_{j\geqslant 1} b_{kj}\phi_j$ and $\widehat{\eta}_k = \sum_{j\geqslant 1} b_{kj}\widehat{\phi}_j$. Let $\eta_k^{(m)} = \sum_{j=1}^m b_{kj}\phi_j$ and $\widehat{\eta}_k^{(m)} = \sum_{j=1}^m b_{kj}\widehat{\phi}_j$. Recall that we have introduced the notation $\Pi_m := \sum_{i=1}^m \phi_i \otimes \phi_i$ and $\widehat{\Pi}_m := \sum_{i=1}^m \widehat{\phi}_i \otimes \widehat{\phi}_i$, then it holds that $\widehat{\Gamma}_m^\dagger\widehat{\Gamma}\widehat{\eta}_k = \widehat{\Pi}_m\widehat{\eta}_k = \widehat{\eta}_k^{(m)}$ and similarly, $\Gamma_m^\dagger\Gamma\eta_k = \Pi_m\eta_k = \eta_k^{(m)}$. In addition, $\widehat{\Gamma}_m^\dagger\widehat{\Gamma}\widehat{\eta}_k = \widehat{\Gamma}_m^\dagger\widehat{\Gamma}\widehat{\eta}_k^{(m)}$. Thus,

$$\widehat{\Gamma}_m^\dagger\widehat{\Gamma}\widehat{\eta}_k - \Gamma_m^\dagger\Gamma\eta_k + \widehat{\Gamma}_m^\dagger\widehat{\Gamma}(\eta_k - \widehat{\eta}_k) = \widehat{\eta}_k^{(m)} - \eta_k^{(m)} + \widehat{\Gamma}_m^\dagger\widehat{\Gamma}(\eta_k - \widehat{\eta}_k^{(m)}).$$

Insert this equality into (E.4), we have

$$\begin{aligned}\left(\widehat{\Gamma}_m^\dagger - \Gamma_m^\dagger\right)\Gamma\eta_k &= \widehat{\eta}_k^{(m)} - \eta_k^{(m)} - \widehat{\Gamma}_m^\dagger\widehat{\Gamma}(\widehat{\eta}_k^{(m)} - \eta_k^{(m)} + \eta_k^{(m)} - \eta_k) + \widehat{\Gamma}_m^\dagger(\Gamma - \widehat{\Gamma})\eta_k \\ &= (\boldsymbol{I} - \Pi_m)(\widehat{\eta}_k^{(m)} - \eta_k^{(m)}) - \widehat{\Pi}_m(\eta_k^{(m)} - \eta_k) + \widehat{\Gamma}_m^\dagger(\Gamma - \widehat{\Gamma})\eta_k\end{aligned} \tag{E.5}$$

where $\boldsymbol{I}$ is the identity operator.

We first find a bound for

$$\|(\boldsymbol{I} - \Pi_m)(\widehat{\eta}_k^{(m)} - \eta_k^{(m)})\| \leqslant \|\widehat{\eta}_k^{(m)} - \eta_k^{(m)}\|.$$

Note that

$$\widehat{\eta}_k^{(m)} - \eta_k^{(m)} = \sum_{j=1}^{m} b_{kj}(\widehat{\phi}_j - \phi_j). \tag{E.6}$$

It reduces to analyzing $\left\|\widehat{\phi}_j - \phi_j\right\|$. Note that our predictor $\boldsymbol{X}$ satisfies the assumptions in (Hall and Horowitz, 2007). By Equation (5.22) of (Hall and Horowitz, 2007), we have

$$\mathbb{E}\left[\left\|\widehat{\phi}_j - \phi_j\right\|^2\right] \lesssim j^2/n \tag{E.7}$$

uniformly in $1 \leqslant j \leqslant m$. Substituting it into Equation (E.6) and using Cauchy–Schwarz inequality, we obtain

$$\begin{aligned}\mathbb{E}\left[\left\|\widehat{\eta}_k^{(m)} - \eta_k^{(m)}\right\|^2\right] &= \mathbb{E}\left[\left\|\sum_{j=1}^{m} b_{kj}(\widehat{\phi}_j - \phi_j)\right\|^2\right] \\ &\lesssim \frac{1}{n}\sum_{j=1}^{m} m j^{-2\beta} j^2 \leqslant \frac{m}{n}\sum_{j=1}^{\infty} j^{2-2\beta} \lesssim \frac{m}{n} \lesssim \frac{m^{\alpha+1}}{n}\end{aligned}$$

by Assumption 4. Then we have

$$\mathbb{E}\left[\|(\boldsymbol{I} - \Pi_m)(\widehat{\eta}_k^{(m)} - \eta_k^{(m)})\|^2\right] \lesssim \frac{m^{\alpha+1}}{n}.$$

Next, we bound $\|\widehat{\Pi}_m(\eta_k^{(m)} - \eta_k)\|$ as follows:

$$\begin{aligned}\left\|\widehat{\Pi}_m(\eta_k^{(m)} - \eta_k)\right\|^2 &= \left\|\widehat{\Pi}_m \sum_{j=m+1}^{\infty} b_{kj}\phi_j\right\|^2 \leqslant \left\|\sum_{j=m+1}^{\infty} b_{kj}\phi_j\right\|^2 \\ &= \sum_{j=m+1}^{\infty} b_{kj}^2 \leqslant C \sum_{j=m+1}^{\infty} j^{-2\beta} \asymp m^{1-2\beta} \asymp \frac{m^{\alpha+1}}{n}.\end{aligned}$$

by Assumption 4 and the choice of m in Theorem 2.

Finally, using a similar argument in the proof of Proposition 2 (1), we can easily derive the following inequality:

$$\mathbb{E}\left[\left\|\widehat{\Gamma}_m^{\dagger}\left(\widehat{\Gamma}-\Gamma\right)\right\|^2 1_{\mathtt{E}}\right] \lesssim \frac{H^2 m^{\alpha+1}}{n}$$

by noting that $\mathbb{E}\left[\left\|\widehat{\Gamma}-\Gamma\right\|^2\right] = O(n^{-1})$ (see e.g., Equation (5.9) in Hall and Horowitz (2007)). This proves the desired Equation (E.3) □

Lemma 7. *Suppose Assumption 1 holds. There exists a constant C that depends only on Γ and Γ_e, such that if $\tau > \frac{6\|\Gamma_e\|}{\lambda^+_{\min}(\Gamma_e)}$ and the event $\mathtt{E}$ holds, then the norm of the operator $\Gamma^{-1}\widetilde{\Gamma}_e$ is no greater than C.*

Proof. By Lemma 5, if $\tau > \frac{6\|\Gamma_e\|}{\lambda^+_{\min}(\Gamma_e)}$, then we have $\mathrm{Im}(\Gamma_e) = \mathrm{Im}(\widetilde{\Gamma}_e)$ on the event $\mathtt{E}$. Furthermore,

$$\begin{aligned}\|\Gamma^{-1}\widetilde{\Gamma}_e\|1_{\mathtt{E}} \leqslant& \|\Gamma^{-1}\mid_{\mathcal{S}_e}\|\|\widetilde{\Gamma}_e\|1_{\mathtt{E}} \leqslant \|\Gamma^{-1}\mid_{\mathcal{S}_e}\|\left(\|\widetilde{\Gamma}_e-\Gamma_e\|+\|\Gamma_e\|\right)1_{\mathtt{E}}\\ \leqslant& \|\Gamma^{-1}\mid_{\mathcal{S}_e}\|(\frac{3}{\tau}+1)\|\Gamma_e\| \leqslant \|\Gamma^{-1}\mid_{\mathcal{S}_e}\|(\frac{\lambda^+_{\min}(\Gamma_e)}{2\|\Gamma_e\|}+1)\|\Gamma_e\|.\end{aligned}$$

The last expression is bounded since Γ_e is of rank d and under Assumption 1, $\|\Gamma^{-1}\mid_{\mathcal{S}_e}\|$ is upper bounded since $\Gamma^{-1}\mathcal{S}_e = \mathcal{S}_{Y|\boldsymbol{X}}$.

□

Thanks to these two lemmas, we obtain $\mathbb{E}\left[\|B_2\|^2\right] \lesssim \frac{H^2 m^{\alpha+1}}{n}$, and finish the proof of Part-(2).

Proof of Proposition 2-(3). For the term B_3, we have

$$\widehat{\Gamma}_m^{\dagger} - \Gamma_m^{\dagger} = \sum_{j=1}^{m} (\widehat{\lambda}_j^{-1} - \lambda_j^{-1}) \widehat{\phi}_j \otimes \widehat{\phi}_j + \sum_{j=1}^{m} \lambda_j^{-1} \left(\widehat{\phi}_j \otimes \widehat{\phi}_j - \phi_j \otimes \phi_j \right) =: A_{11} + A_{12}.$$

By Cauchy–Schwarz inequality, we have

$$\begin{aligned} \mathbb{E}^2 \left[\left\| A_{12} \left(\widehat{\Gamma}_e - \widetilde{\Gamma}_e \right) \right\| 1_{\mathbb{E}} \right] \leqslant & \mathbb{E} \left[\| A_{12} \|^2 \right] \mathbb{E} \left[\left\| \left(\widehat{\Gamma}_e - \widetilde{\Gamma}_e \right) \right\|^2 1_{\mathbb{E}} \right] \lesssim \frac{1}{n} \mathbb{E} \left[\| A_{12} \|^2 \right] \\ \lesssim & \frac{1}{n} m \sum_{j=1}^{m} \lambda_j^{-2} \mathbb{E} [\| \widehat{\phi}_j - \phi_j \|^2]. \end{aligned}$$

By Equation (5.22) in Hall and Horowitz (2007), we have $\mathbb{E}\left[\left\|\widehat{\phi}_j - \phi_j\right\|^2\right] \lesssim \frac{j^2}{n}$ holds uniformly for all $j \leqslant m$. Hence, by Assumption 4, we have

$$\frac{1}{n} m \sum_{j=1}^{m} \lambda_j^{-2} \mathbb{E} [\| \widehat{\phi}_j - \phi_j \|^2] \lesssim \frac{m}{n^2} \sum_{j=1}^{m} j^{2+2\alpha} \asymp \frac{m^{4+2\alpha}}{n^2} \lesssim \frac{m^{\alpha+1}}{n}.$$

Also, we have

$$A_{11} = \sum_{j=1}^{m} \frac{\widehat{\lambda}_j - \lambda_j}{\widehat{\lambda}_j \lambda_j} \left(\widehat{\phi}_j \otimes \widehat{\phi}_j - \phi_j \otimes \phi_j \right) + \sum_{j=1}^{m} \frac{\widehat{\lambda}_j - \lambda_j}{\widehat{\lambda}_j \lambda_j} \phi_j \otimes \phi_j =: A_{111} +_{A112} .$$

By direct calculation, we find that

$$\begin{aligned} \left\| A_{111} \left(\widehat{\Gamma}_e - \widetilde{\Gamma}_e \right) \right\| = & \left\| \sum_{j=1}^{m} \frac{\widehat{\lambda}_j - \lambda_j}{\widehat{\lambda}_j \lambda_j} \left(\widehat{\phi}_j \otimes \widehat{\phi}_j - \phi_j \otimes \phi_j \right) \left(\widehat{\Gamma}_e - \widetilde{\Gamma}_e \right) \right\| \\ \leqslant & C \sum_{j=1}^{m} \left| \frac{\widehat{\lambda}_j - \lambda_j}{\widehat{\lambda}_j \lambda_j} \right| \| \phi_j - \hat{\phi}_j \| \| \widehat{\Gamma}_e - \widetilde{\Gamma}_e \|. \end{aligned}$$

Recall the definition of $\mathtt{F}$ in (E.1), we have

$$
\begin{aligned}
&\mathbb{E}^2\left[\left\|A_{111}\left(\widehat{\Gamma}_e-\widetilde{\Gamma}_e\right)\right\|1_{\mathtt{E}}1_{\mathtt{F}}\right] \leqslant C^2 m\sum_{j=1}^{m}\lambda_j^{-2}\mathbb{E}^2[\|\phi_j-\hat{\phi}_j\|\|\widehat{\Gamma}_e-\widetilde{\Gamma}_e\|] \\
\leqslant & C^2 m\sum_{j=1}^{m}\lambda_j^{-2}\mathbb{E}[\|\phi_j-\hat{\phi}_j\|^2]\mathbb{E}[\|\widehat{\Gamma}_e-\widetilde{\Gamma}_e\|^2 1_{\mathtt{E}}] \\
\leqslant & \frac{C^2 m}{n}\sum_{j=1}^{m}\lambda_j^{-2}\mathbb{E}\left[\|\phi_j-\hat{\phi}_j\|^2\right] \lesssim \frac{m}{n^2}\sum_{j=1}^{m} j^{2+2\alpha} \lesssim \frac{m^{\alpha+1}}{n}.
\end{aligned}
$$

For the term A_{112}, we have

$$
\begin{aligned}
\left\|A_{112}\left(\widehat{\Gamma}_e-\widetilde{\Gamma}_e\right)\right\| &= \sup_{\boldsymbol{u}:\|\boldsymbol{u}\|=1}\left\|\sum_{j=1}^{m}\frac{\widehat{\lambda}_j-\lambda_j}{\widehat{\lambda}_j\lambda_j}\phi_j\otimes\phi_j\left(\widehat{\Gamma}_e-\widetilde{\Gamma}_e\right)\boldsymbol{u}\right\| \\
&= \sup_{\boldsymbol{u}:\|\boldsymbol{u}\|=1}\left\|\sum_{j=1}^{m}\frac{\widehat{\lambda}_j-\lambda_j}{\widehat{\lambda}_j\lambda_j}\left\langle\left(\widehat{\Gamma}_e-\widetilde{\Gamma}_e\right)\boldsymbol{u},\phi_j\right\rangle\phi_j\right\| \\
&= \sup_{\boldsymbol{u}:\|\boldsymbol{u}\|=1}\left\|\sum_{j=1}^{m}\frac{\widehat{\lambda}_j-\lambda_j}{\widehat{\lambda}_j\lambda_j}\sum_{i\geqslant 1}u_i\left\langle\left(\widehat{\Gamma}_e-\widetilde{\Gamma}_e\right)\phi_i,\phi_j\right\rangle\phi_j\right\| \quad \left(\boldsymbol{u}=\sum_{i\geqslant 1}u_i\phi_i\right) \\
&\leqslant \left(\sum_{j=1}^{m}\left(\frac{\widehat{\lambda}_j-\lambda_j}{\widehat{\lambda}_j\lambda_j}\right)^2\sum_{i\geqslant 1}\left\langle\left(\widehat{\Gamma}_e-\widetilde{\Gamma}_e\right)\phi_i,\phi_j\right\rangle^2\right)^{1/2}
\end{aligned}
$$

where in the fourth line, we use the Cauchy-Schwarz inequality and the relation that $\sum_{i\geqslant 1}u_i^2 =$

1. Then we have

$$
\begin{aligned}
&\mathbb{E}\left[\left\|A_{112}\left(\widehat{\Gamma}_e-\widetilde{\Gamma}_e\right)\right\|^2 1_{\mathtt{E}\cap\mathtt{F}}\right]\\
\leqslant&\mathbb{E}\left[\left(\sum_{j=1}^{m}\left(\frac{\widehat{\lambda}_j-\lambda_j}{\widehat{\lambda}_j\lambda_j}\right)^2\sum_{i\geqslant 1}\left\langle\left(\widehat{\Gamma}_e-\widetilde{\Gamma}_e\right)\phi_i,\phi_j\right\rangle^2\right)1_{\mathtt{E}\cap\mathtt{F}}\right]\\
\leqslant&\sum_{j=1}^{m}\lambda_j^{-2}\sum_{i\geqslant 1}\mathbb{E}\left[1_{\mathtt{E}}\left\langle\left(\widehat{\Gamma}_e-\widetilde{\Gamma}_e\right)\phi_i,\phi_j\right\rangle^2\right]\\
\leqslant&\sum_{j=1}^{m}\lambda_j^{-2}\sum_{i\geqslant 1}\frac{CH^2\lambda_i\lambda_j}{n}\leqslant CH^2\sum_{j=1}^{m}\frac{1}{\lambda_j n}\sum_{i\geqslant 1}\lambda_i\lesssim\frac{H^2m^{\alpha+1}}{n}.
\end{aligned}
$$

where the third inequality follows the same proof of Lemma 2. Thus, we complete the proof of Proposition 2-(3).

□

E.2 The proof of Theorem 2

Let $T:=\Gamma^{-1}\widetilde{\Gamma}_e\left(\Gamma^{-1}\widetilde{\Gamma}_e\right)^*$ and $\widehat{T}_m=\widehat{\Gamma}_m^{\dagger}\widehat{\Gamma}_e^{d}\widehat{\Gamma}_e^{d}\widehat{\Gamma}_m^{\dagger}$ where $\widehat{\Gamma}_e^{d}$ is defined in (D.1). Define

$$
\mathtt{F}':=\left\{\frac{\lambda_d(T)}{2}\leqslant\lambda_d(\widehat{T}_m)\leqslant\frac{3\lambda_d(T)}{2}\right\}=\left\{|\lambda_d(\widehat{T}_m)-\lambda_d(T)|\leqslant\frac{\lambda_d(T)}{2}\right\}. \tag{E.8}
$$

To prove Theorem 2, we only need to prove $\mathbb{E}\left[\left\|P_{\widehat{\mathcal{S}}_{Y|\boldsymbol{x}}}-P_{\mathcal{S}_{Y|\boldsymbol{x}}}\right\|^{1/2}1_{\mathtt{E}\cap\mathtt{F}\cap\mathtt{F}'}\right]\lesssim n^{\frac{-(2\beta-1)}{4(\alpha+2\beta)}}$ and $\mathbb{P}(\mathtt{E}\cap\mathtt{F}\cap\mathtt{F}')\xrightarrow{n\to\infty}1$.

Before we delve into the prove, we introduce some convenient notation. For any $\boldsymbol{u}\in\mathcal{H}$, define $\boldsymbol{u}^*:\mathcal{H}\to\mathbb{R},\boldsymbol{v}\mapsto\langle\boldsymbol{u},\boldsymbol{v}\rangle$. Then $\boldsymbol{u}^*$ is the adjoint operator of $\boldsymbol{u}:\mathbb{R}\to\mathcal{H},\lambda\mapsto\lambda\boldsymbol{u},(\forall\lambda\in\mathbb{R})$ since $\boldsymbol{u}^*\boldsymbol{v}=\langle\boldsymbol{u}^*\boldsymbol{v},1\rangle=\langle\boldsymbol{v},\boldsymbol{u}\rangle$.

Similarly, for any d elements in $\mathcal{H}$, say $\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_d$, we can define $\boldsymbol{B}:=(\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_d):\mathbb{R}^d\to L^2[0,1]$ and its adjoint $\boldsymbol{B}^*$.

We also define the 'truncated central space'

$$\mathcal{S}_{Y|\boldsymbol{X}}^{(m)} = \Pi_m \mathcal{S}_{Y|\boldsymbol{X}} = \operatorname{span}\{\boldsymbol{\beta}_1^{(m)}, \ldots, \boldsymbol{\beta}_d^{(m)}\}, \tag{E.9}$$

where $\boldsymbol{\beta}_k^{(m)} := \Pi_m(\boldsymbol{\beta}_k), k \in [d]$. For such a truncated central space, we have the following proposition, whose proof is deferred to the end.

Proposition 3. *Under Assumption* 4*, if m is sufficiently large, we have*

$$\left\| P_{\mathcal{S}_{Y|\boldsymbol{X}}} - P_{\mathcal{S}_{Y|\boldsymbol{X}}^{(m)}} \right\| \lesssim m^{-\frac{2\beta-1}{2}}. \tag{E.10}$$

Then we only need to show that

$$\mathbb{E}\left[\left\| P_{\widehat{\mathcal{S}}_{Y|\boldsymbol{X}}} - P_{\mathcal{S}_{Y|\boldsymbol{X}}^{(m)}} \right\|^{1/2} 1_{\mathrm{E}\cap\mathrm{F}\cap\mathrm{F}'} \right] \lesssim n^{\frac{-(2\beta-1)}{4(\alpha+2\beta)}}.$$

Recall that $T := \Gamma^{-1}\widetilde{\Gamma}_e \left(\Gamma^{-1}\widetilde{\Gamma}_e\right)^*$ and $\widehat{T}_m = \widehat{\Gamma}_m^{\dagger}\widehat{\Gamma}_e^d\widehat{\Gamma}_e^d\widehat{\Gamma}_m^{\dagger}$. Let $T_m := \Pi_m T \Pi_m$. We have $P_{\mathcal{S}_{Y|\boldsymbol{X}}^{(m)}} = P_{T_m}$ and $T_m = \Gamma_m^{\dagger}\widetilde{\Gamma}_e\widetilde{\Gamma}_e\Gamma_m^{\dagger}$. Furthermore, $P_{\widehat{\mathcal{S}}_{Y|\boldsymbol{X}}} = P_{\widehat{\Gamma}_m^{\dagger}\widehat{\Gamma}_e^d} = P_{\widehat{T}_m}$.

By $\sin\Theta$ theorem (Lemma 20), we have

$$\begin{aligned} \left\| P_{\widehat{\mathcal{S}}_{Y|\boldsymbol{X}}} - P_{\mathcal{S}_{Y|\boldsymbol{X}}^{(m)}} \right\| = \left\| P_{T_m} - P_{\widehat{T}_m} \right\| &\leqslant \frac{\pi}{2} \frac{\left\| T_m - \widehat{T}_m \right\|}{\min\{\lambda_{\min}^{+}(T_m), \lambda_{\min}^{+}(\widehat{T}_m)\}} \\ &\leqslant \frac{\pi}{2} \frac{\left(\left\| \widehat{\Gamma}_m^{\dagger}\widehat{\Gamma}_e^d \right\| + \left\| \Gamma_m^{\dagger}\widetilde{\Gamma}_e \right\| \right) \left\| \widehat{\Gamma}_m^{\dagger}\widehat{\Gamma}_e^d - \Gamma_m^{\dagger}\widetilde{\Gamma}_e \right\|}{\min\{\lambda_{\min}^{+}(T_m), \lambda_{\min}^{+}(\widehat{T}_m)\}}. \end{aligned} \tag{E.11}$$

We first provide an upper bound on the numerator of the right hand side of (E.11).

Similar to the argument of Lemma 7, we find that

$$\mathbb{E}\left[\left\|\Gamma_m^{\dagger}\widetilde{\Gamma}_e\right\|^2 1_{\mathsf{E}\cap\mathsf{F}}\right] \lesssim 1 \quad \text{and} \quad \mathbb{E}\left[\left\|\widehat{\Gamma}_m^{\dagger}\widehat{\Gamma}_e^d\right\|^2 1_{\mathsf{E}\cap\mathsf{F}}\right] \lesssim 1. \tag{E.12}$$

We move on to analyze $\left\|\widehat{\Gamma}_m^{\dagger}\widehat{\Gamma}_e^d - \Gamma_m^{\dagger}\widetilde{\Gamma}_e\right\|$.

Firstly, for any $\boldsymbol{u} \in \widehat{\mathcal{S}}_e$ with $\|\boldsymbol{u}\| = 1$, we have $\widehat{\Gamma}_e^d\boldsymbol{u} = \widehat{\Gamma}_e\boldsymbol{u}$. Then by Theorem 4, we have

$$\begin{aligned}
&\mathbb{E}\left[1_{\mathsf{E}\cap\mathsf{F}}\left\|\left(\widehat{\Gamma}_m^{\dagger}\widehat{\Gamma}_e^d - \Gamma_m^{\dagger}\widetilde{\Gamma}_e\right)(\boldsymbol{u})\right\|\right] = \mathbb{E}\left[1_{\mathsf{E}\cap\mathsf{F}}\left\|\left(\widehat{\Gamma}_m^{\dagger}\widehat{\Gamma}_e - \Gamma_m^{\dagger}\widetilde{\Gamma}_e\right)(\boldsymbol{u})\right\|\right] \\
\leqslant&\mathbb{E}\left[1_{\mathsf{E}\cap\mathsf{F}}\left\|\left(\widehat{\Gamma}_m^{\dagger}\widehat{\Gamma}_e - \Gamma_m^{\dagger}\widetilde{\Gamma}_e\right)\right\|\right] \lesssim Hn^{\frac{-(2\beta-1)}{2(\alpha+2\beta)}}.
\end{aligned} \tag{E.13}$$

This shows that $\mathbb{E}\left[\left\|\left(\widehat{\Gamma}_m^{\dagger}\widehat{\Gamma}_e^d - \Gamma_m^{\dagger}\widetilde{\Gamma}_e\right)|_{\widehat{\mathcal{S}}_e}\right\| 1_{\mathsf{E}\cap\mathsf{F}}\right] \lesssim Hn^{\frac{-(2\beta-1)}{2(\alpha+2\beta)}}$.

Secondly, for any $\boldsymbol{u} \in \widehat{\mathcal{S}}_e^{\perp}$ with $\|\boldsymbol{u}\| = 1$,

$$\begin{aligned}
\left\|\left(\widehat{\Gamma}_m^{\dagger}\widehat{\Gamma}_e^d - \Gamma_m^{\dagger}\widetilde{\Gamma}_e\right)(\boldsymbol{u})\right\| &= \left\|\Gamma_m^{\dagger}\widetilde{\Gamma}_e(\boldsymbol{u})\right\| = \left\|\Gamma_m\Gamma^{-1}\widetilde{\Gamma}_e P_{\mathcal{S}_e}(\boldsymbol{u})\right\| \\
&= \left\|\Gamma_m\Gamma^{-1}\widetilde{\Gamma}_e(P_{\mathcal{S}_e} - P_{\widehat{\mathcal{S}}_e})\boldsymbol{u}\right\| \leqslant \left\|\Gamma_m\Gamma^{-1}\widetilde{\Gamma}_e(P_{\mathcal{S}_e} - P_{\widehat{\mathcal{S}}_e})\right\| \\
&\leqslant \|\Gamma_m\| \left\|\Gamma^{-1}\widetilde{\Gamma}_e\right\| \left\|P_{\mathcal{S}_e} - P_{\widehat{\mathcal{S}}_e}\right\|.
\end{aligned} \tag{E.14}$$

Then by Lemma 7 and Theorem 1, we have

$$\mathbb{E}\left[\left\|\left(\widehat{\Gamma}_m^{\dagger}\widehat{\Gamma}_e^d - \Gamma_m^{\dagger}\widetilde{\Gamma}_e\right)(\boldsymbol{u})\right\| 1_{\mathsf{E}\cap\mathsf{F}}\right] \lesssim \sqrt{\frac{H^2}{n}}.$$

It implies that $\mathbb{E}\left[1_{\mathsf{E}\cap\mathsf{F}}\left\|\left(\widehat{\Gamma}_m^{\dagger}\widehat{\Gamma}_e^d - \Gamma_m^{\dagger}\widetilde{\Gamma}_e\right)|_{\widehat{\mathcal{S}}_e^{\perp}}\right\|\right] \lesssim \sqrt{\frac{H^2}{n}}$. Combing Equations (E.13) and (E.14), we obtain

$$\mathbb{E}\left[1_{\mathsf{E}\cap\mathsf{F}}\left\|\widehat{\Gamma}_m^{\dagger}\widehat{\Gamma}_e^d - \Gamma_m^{\dagger}\widetilde{\Gamma}_e\right\|\right] \lesssim \left(Hn^{\frac{-(2\beta-1)}{2(\alpha+2\beta)}}\right). \tag{E.15}$$

Lastly, we provide a lower bound on the denominator of the right hand side of (E.11).

If $\tau > \frac{6\|\Gamma_e\|}{\lambda^+_{\min}(\Gamma_e)}$, then $\mathrm{Im}(\widetilde{\Gamma}_e) = \mathrm{Im}(\Gamma_e)$ on the event $\mathtt{E}$. By Lemma 16, one has

$$\|\Gamma^{-1}\widetilde{\Gamma}_e - \Gamma^\dagger_m\widetilde{\Gamma}_e\| \xrightarrow{m\to\infty} 0. \tag{E.16}$$

By Lemma 21, one has $\sigma^+_{\min}(\Gamma^\dagger_m\widetilde{\Gamma}_e) \geqslant \frac{\sigma^+_{\min}(\Gamma^{-1}\Gamma_e)}{2}$ for sufficiently large m, where $\sigma^+_{\min}$ denotes the infimum of the positive singular values.

By Markov inequality, we have

$$\begin{aligned}
&\mathbb{P}(1_{\mathtt{E}\cap\mathtt{F}}|\lambda_d(\widehat{T}_m) - \lambda_d(T)| \geqslant \frac{\lambda_d(T)}{2}) \leqslant \frac{\mathbb{E}[1_{\mathtt{E}\cap\mathtt{F}}|\lambda_d(\widehat{T}_m) - \lambda_d(T)|]}{\lambda_d(T)/2} \leqslant \frac{\mathbb{E}[1_{\mathtt{E}\cap\mathtt{F}}\|\widehat{T}_m - T\|]}{\lambda_d(T)/2} \\
&\lesssim \frac{\mathbb{E}[1_{\mathtt{E}\cap\mathtt{F}}(\|\widehat{\Gamma}^\dagger_m\widehat{\Gamma}^d_e - \Gamma^{-1}\widetilde{\Gamma}_e\|^2 + 2\|\widehat{\Gamma}^\dagger_m\widehat{\Gamma}^d_e - \Gamma^{-1}\widetilde{\Gamma}_e\|\|\Gamma^{-1}\widetilde{\Gamma}_e\|)]}{\lambda_d(T)/2} \\
&\lesssim \frac{\mathbb{E}[1_{\mathtt{E}\cap\mathtt{F}}\|\widehat{\Gamma}^\dagger_m\widehat{\Gamma}^d_e - \Gamma^{-1}\widetilde{\Gamma}_e\|]}{\lambda_d(T)/2} \xrightarrow{m\to\infty} 0.
\end{aligned} \tag{E.17}$$

where the second inequality comes from Lemma 21, the fourth comes from Lemma 7 and the fifth comes from (E.15)-(E.16).

Recall that

$$\mathtt{F}' := \left\{\frac{\lambda_d(T)}{2} \leqslant \lambda_d(\widehat{T}_m) \leqslant \frac{3\lambda_d(T)}{2}\right\} = \left\{|\lambda_d(\widehat{T}_m) - \lambda_d(T)| \leqslant \frac{\lambda_d(T)}{2}\right\}.$$

Thus (E.17) implies

$$\mathbb{P}((\mathtt{F}')^c \cap (\mathtt{E} \cap \mathtt{F})) \xrightarrow{m\to\infty} 0. \tag{E.18}$$

Furthermore,

$$\mathbb{P}(\mathtt{E}\cap\mathtt{F}\cap\mathtt{F}') = \mathbb{P}(\mathtt{E}\cap\mathtt{F}) - \mathbb{P}(\mathtt{E}\cap\mathtt{F}\cap(\mathtt{F}')^c) \xrightarrow{m\to\infty} 1.$$

On the event $\mathtt{E}\cap\mathtt{F}\cap\mathtt{F}'$, it holds that $\sigma^{+}_{\min}(\widehat{\Gamma}^{\dagger}_{m}\widehat{\Gamma}^{d}_{e}) \geqslant \frac{\sigma^{+}_{\min}(\Gamma^{-1}\Gamma_e)}{4}$, which implies

$$\min\{\lambda^{+}_{\min}(T_m), \lambda^{+}_{\min}(\widehat{T}_m)\} \geqslant C \tag{E.19}$$

for some constant $C > 0$.

Inserting (E.12), (E.15) and (E.19) into (E.11), we have

$$\begin{aligned}
&\mathbb{E}^2\left[\left\|P_{\widehat{\mathcal{S}}_{Y|\boldsymbol{X}}} - P_{\mathcal{S}^{(m)}_{Y|\boldsymbol{X}}}\right\|^{1/2} 1_{\mathtt{E}\cap\mathtt{F}\cap\mathtt{F}'}\right] = \mathbb{E}^2\left[\left\|P_{T_m} - P_{\widehat{T}_m}\right\|^{1/2} 1_{\mathtt{E}\cap\mathtt{F}\cap\mathtt{F}'}\right] \\
\lesssim& \frac{\mathbb{E}^2\left[\left(\left\|\widehat{\Gamma}^{\dagger}_{m}\widehat{\Gamma}^{d}_{e}\right\| + \left\|\Gamma^{\dagger}_{m}\widetilde{\Gamma}_{e}\right\|\right)^{1/2}\left\|\widehat{\Gamma}^{\dagger}_{m}\widehat{\Gamma}^{d}_{e} - \Gamma^{\dagger}_{m}\widetilde{\Gamma}_{e}\right\|^{1/2} 1_{\mathtt{E}\cap\mathtt{F}\cap\mathtt{F}'}\right]}{\min\{\lambda^{+}_{\min}(T_m), \lambda^{+}_{\min}(\widehat{T}_m)\}1_{\mathtt{E}\cap\mathtt{F}\cap\mathtt{F}'}} \\
\lesssim& \mathbb{E}\left[\left(\left\|\widehat{\Gamma}^{\dagger}_{m}\widehat{\Gamma}^{d}_{e}\right\| + \left\|\Gamma^{\dagger}_{m}\widetilde{\Gamma}_{e}\right\|\right) 1_{\mathtt{E}\cap\mathtt{F}\cap\mathtt{F}'}\right]\mathbb{E}\left[\left\|\widehat{\Gamma}^{\dagger}_{m}\widehat{\Gamma}^{d}_{e} - \Gamma^{\dagger}_{m}\widetilde{\Gamma}_{e}\right\| 1_{\mathtt{E}\cap\mathtt{F}\cap\mathtt{F}'}\right] \lesssim Hn^{\frac{-(2\beta-1)}{2(\alpha+2\beta)}}.
\end{aligned}$$

This completes the proof of Theorem 2.

Proof of Proposition 3. Let $\mathcal{B} := \sum_{i=1}^{d}\boldsymbol{\beta}_i\otimes\boldsymbol{\beta}_i$ and $\mathcal{B}^{(m)} := \sum_{i=1}^{d}\boldsymbol{\beta}^{(m)}_i\otimes\boldsymbol{\beta}^{(m)}_i$. Note that $\mathrm{Im}(\mathcal{B}) = \mathrm{span}\{\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_d\} = \mathcal{S}_{Y|\boldsymbol{X}}$. Similarly, $\mathrm{Im}(\mathcal{B}^{(m)}) = \mathrm{span}\{\boldsymbol{\beta}^{(m)}_1,\ldots,\boldsymbol{\beta}^{(m)}_d\} = \mathcal{S}^{(m)}_{Y|\boldsymbol{X}}$. Thus $\left\|P_{\mathcal{S}_{Y|\boldsymbol{X}}} - P_{\mathcal{S}^{(m)}_{Y|\boldsymbol{X}}}\right\| = \|P_{\mathcal{B}} - P_{\mathcal{B}^{(m)}}\|$.

By Lemma 20, we have

$$\|P_{\mathcal{B}} - P_{\mathcal{B}^{(m)}}\| \leqslant \frac{\pi}{2}\frac{\|\mathcal{B}-\mathcal{B}^{(m)}\|}{\min\{\lambda^{+}_{\min}(\mathcal{B}), \lambda^{+}_{\min}(\mathcal{B}^{(m)})\}}. \tag{E.20}$$

Note that $\mathcal{B}-\mathcal{B}^{(m)}$ is self-adjoint, then

$$\begin{aligned}\left\|\mathcal{B}-\mathcal{B}^{(m)}\right\| &= \sup_{\boldsymbol{u}\in\mathbb{S}_{\mathcal{H}}}|\langle(\mathcal{B}-\mathcal{B}^{(m)})(\boldsymbol{u}),\boldsymbol{u}\rangle| = \sup_{\boldsymbol{u}\in\mathbb{S}_{\mathcal{H}}}|\langle\mathcal{B}\boldsymbol{u},\boldsymbol{u}\rangle-\langle\mathcal{B}^{(m)}\boldsymbol{u},\boldsymbol{u}\rangle| \\ &= \sup_{\boldsymbol{u}\in\mathbb{S}_{\mathcal{H}}}\left|\sum_{i=1}^{d}\left[\langle\boldsymbol{\beta}_i,\boldsymbol{u}\rangle^2-\langle\boldsymbol{\beta}_i^{(m)},\boldsymbol{u}\rangle^2\right]\right| = \sup_{\boldsymbol{u}\in\mathbb{S}_{\mathcal{H}}}\left|\sum_{i=1}^{d}\langle\boldsymbol{\beta}_i-\boldsymbol{\beta}_i^{(m)},\boldsymbol{u}\rangle\langle\boldsymbol{\beta}_i+\boldsymbol{\beta}_i^{(m)},\boldsymbol{u}\rangle\right| \\ &\leqslant \sup_{\boldsymbol{u}\in\mathbb{S}_{\mathcal{H}}}\sum_{i=1}^{d}\left|\langle\boldsymbol{\beta}_i-\boldsymbol{\beta}_i^{(m)},\boldsymbol{u}\rangle\langle\boldsymbol{\beta}_i+\boldsymbol{\beta}_i^{(m)},\boldsymbol{u}\rangle\right| \leqslant \sum_{i=1}^{d}\left\|\boldsymbol{\beta}_i-\boldsymbol{\beta}_i^{(m)}\right\|\left\|\boldsymbol{\beta}_i+\boldsymbol{\beta}_i^{(m)}\right\|,\end{aligned}$$

where the first inequality comes from the triangle inequality, and the second inequality comes from the Cauchy-Schwarz inequality and $\|\boldsymbol{u}\|=1$. According to Assumption 4, one can get

$$\begin{aligned}\left\|\boldsymbol{\beta}_i-\boldsymbol{\beta}_i^{(m)}\right\| &= \left\|\sum_{j=m+1}^{\infty}b_{ij}\phi_j\right\| = \sqrt{\sum_{j=m+1}^{\infty}b_{ij}^2} \lesssim \sqrt{\sum_{j=m+1}^{\infty}j^{-2\beta}}; \\ \left\|\boldsymbol{\beta}_i+\boldsymbol{\beta}_i^{(m)}\right\| &\leqslant \|\boldsymbol{\beta}_i\|+\left\|\boldsymbol{\beta}_i^{(m)}\right\| \leqslant 2\|\boldsymbol{\beta}_i\| = 2\sqrt{\sum_{j=1}^{\infty}b_{ij}^2} \lesssim \sqrt{\sum_{j=1}^{\infty}j^{-2\beta}}.\end{aligned}$$

Because $\beta>1/2$, one has

$$\sum_{j=m+1}^{\infty}\frac{1}{j^{2\beta}} \lesssim \frac{1}{m^{2\beta-1}}; \qquad \sum_{j=1}^{\infty}\frac{1}{j^{2\beta}} < \infty.$$

Thus, one can get

$$\left\|\mathcal{B}-\mathcal{B}^{(m)}\right\| \lesssim m^{-\frac{2\beta-1}{2}}. \tag{E.21}$$

Then we show that $\min\{\lambda_{\min}^{+}(\mathcal{B}),\lambda_{\min}^{+}(\mathcal{B}^{(m)})\} \geqslant C$ for some constant $C>0$. Since $\mathrm{rank}(\mathcal{B})=d$, one can get that $\lambda_{\min}^{+}(\mathcal{B})=\lambda_d(\mathcal{B})$. It is easy to see $\mathrm{rank}(\mathcal{B}^{(m)})\leqslant d$ by $\mathcal{B}^{(m)}=\Pi_m\mathcal{B}\Pi_m$, thus one can assume that $\lambda_{\min}^{+}(\mathcal{B}^{(m)})=\lambda_j(\mathcal{B}^{(m)})$ for some $j\leqslant d$. By

Lemma 21 and (E.21), one has:

$$|\lambda_j(\mathcal{B}^{(m)}) - \lambda_j(\mathcal{B})| \leqslant \left\|\mathcal{B} - \mathcal{B}^{(m)}\right\| \lesssim m^{-\frac{2\beta-1}{2}}.$$

Thus for sufficiently large m, one has

$$\lambda_j\left(\mathcal{B}^{(m)}\right) \geqslant \frac{\lambda_d(\mathcal{B})}{2} \Longrightarrow \min\{\lambda^+_{\min}(\mathcal{B}), \lambda^+_{\min}(\mathcal{B}^{(m)})\} \geqslant \frac{\lambda_d(\mathcal{B})}{2}. \tag{E.22}$$

Inserting (E.21) and (E.22) into (E.20) leads to

$$\left\|P_{\mathcal{S}_{Y|\boldsymbol{X}}} - P_{\mathcal{S}^{(m)}_{Y|\boldsymbol{X}}}\right\| \lesssim m^{-\frac{2\beta-1}{2}}.$$

Thus we complete the proof of Proposition 3. □

F. Proof of Theorem 3

F.1 Proof outline

We sketch the proof outline in this subsection and defer the proof details to the subsequent subsections.

We follow the the standard procedure of applying Fano's inequality to obtain the minimax lower bound. The following lemma is one version of the generalized Fano method.

Lemma 8 (Yu (1997))**.** *Let $N \geqslant 2$ be an integer and $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N\} \subset \Theta_0$ index a collection of probability measures $\mathbb{P}_{\boldsymbol{\theta}_i}$ on a measurable space $(\mathcal{X}, \mathcal{A})$. Let ρ be a pseudometric on Θ_0*

and suppose that for all $i \neq j$

$$\rho(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j) \geqslant \alpha_N, \quad \textit{and} \quad KL(\mathbb{P}_{\boldsymbol{\theta}_i}, \mathbb{P}_{\boldsymbol{\theta}_j}) \leqslant \beta_N.$$

Then every $\mathcal{A}$*-measurable estimator* $\hat{\boldsymbol{\theta}}$ *satisfies*

$$\max_i \mathbb{P}\left(\rho(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}_i) \geqslant \frac{\alpha_N}{2}\right) \geqslant 1 - \frac{\beta_N + \log 2}{\log N}.$$

To apply Lemma 8, we need to construct a family of distributions that are separated from each other in the parameter space but close to each other in terms of the KL-divergence.

Let us first recall the following Varshamov–Gilbert bound (Tsybakov, 2009, Lemma 2.9).

Lemma 9. *For any* $m > 8$*, there exists a set* $\Theta := \{\boldsymbol{\theta}^{(0)}, \ldots, \boldsymbol{\theta}^{(N)}\} \subset \{-1, 1\}^m$*, such that*

1). $\boldsymbol{\theta}^{(0)} = (-1, \ldots, -1)$*;*

2). for any $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta$ *and* $\boldsymbol{\theta} \neq \boldsymbol{\theta}'$, $\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|^2 \geqslant m/2$*;*

3). $N \geqslant 2^{m/8}$.

Let $\phi_1(t) = 1, \phi_{j+1}(t) = \sqrt{2}\cos(j\pi t), j \geqslant 1$. For any $\boldsymbol{\theta} = (\theta_i)_{i \in [m]} \in \Theta$ in Lemma 9 and any $\beta > 3/2$, let us define the central space $\mathcal{S}(\boldsymbol{\theta}) := \operatorname{span}\{\boldsymbol{\beta}_1^{\boldsymbol{\theta}}, \ldots, \boldsymbol{\beta}_d^{\boldsymbol{\theta}}\}$ as follows:

$$\boldsymbol{\beta}_i^{\boldsymbol{\theta}} := \sum_{k=im+1}^{(i+1)m} \theta_{k-im} k^{-\beta} \phi_k + \phi_i$$

where $m = \widetilde{C} n^{\frac{1}{\alpha+2\beta}}$ for some $\widetilde{C} = \widetilde{C}(\alpha, \beta)$ to be determined. We assume d is fixed and $m > d$. The collection $\{\mathcal{S}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$ satisfies a nice property stated in the following lemma.

Lemma 10. *For any different $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ in Θ, we have*

$$\|P_{\mathcal{S}(\boldsymbol{\theta})} - P_{\mathcal{S}(\boldsymbol{\theta}')}\|^2 \geqslant 4\vartheta n^{-\frac{2\beta-1}{\alpha+2\beta}}$$

for some constant $\vartheta > 0$ that depends on α and β.

We next move on to construct a population corresponds to each $\mathcal{S}(\boldsymbol{\theta})$. For this purpose, we will make use of a construction in Lin et al. (2021b), described as follows. For any $x \in \mathbb{R}$, let $\phi(x)$ be a smooth function which maps $(-\infty, 0]$ to 0 and $[1, \infty)$ to 1 and has a positive first derivative over $(0, 1)$. For any $\boldsymbol{z} = (z_i)_{i=1}^d \in \mathbb{R}^d$, let $f(\boldsymbol{z}) := \sum_{i \leqslant d} 2^{i-1}\phi(z_i/\zeta)$ where ζ is sufficiently small such that for $\boldsymbol{Z} = (Z_i)_{i\in[d]} \sim N(0, \boldsymbol{I}_d)$, the probability $\mathbb{P}(\exists i, 0 < Z_i < \zeta) \leqslant d\zeta/(\sqrt{2\pi}) < 2^{-d}$. Let $Y = Af(\boldsymbol{Z}) + \varepsilon$ for some positive constant A and $\varepsilon \sim N(0, 1)$. If A is sufficiently large, the distribution of $(\boldsymbol{Z}, Y)$ satisfies the coverage condition (Lin et al., 2021b, Lemma 15). Note that the joint distribution of $(\boldsymbol{Z}, Y)$ is Lebesgue continuous and it is easy to check that $\mathbb{E}[\boldsymbol{Z}|Y = y]$ is a continuous function with respect to y using the formula for conditional expectation. Thus by Lemma 1, we know that $(\boldsymbol{Z}, Y)$ satisfied WSSC.

Now we describe how to construct a distribution $\mathbb{P}_{\boldsymbol{\theta}}$ of $(\boldsymbol{X}, Y)$ for any given $\boldsymbol{\theta} \in \Theta$. For some α satisfying $\alpha > 1$ and $\frac{1}{2}\alpha + 1 < \beta$, let $\boldsymbol{X} = \sum_{j=1}^\infty j^{-\alpha/2} X_j \phi_j$ such that $X_j \overset{iid}{\sim} N(0, 1), j \geqslant 1$. Then $\Gamma = \sum_{j=1}^\infty j^{-\alpha}\phi_j \otimes \phi_j$. For any $\boldsymbol{\beta}_i (i \in [d])$, we construct the joint distribution of $(\boldsymbol{X}, Y)$ as follows:

$$Y = Af((\boldsymbol{B}^*\Gamma\boldsymbol{B})^{-1/2}\boldsymbol{B}^*\boldsymbol{X}) + \varepsilon, \quad \boldsymbol{X} = \sum_{j=1}^\infty j^{-\alpha/2} X_j \phi_j, \quad \epsilon \sim N(0, 1), \tag{F.1}$$

where $\boldsymbol{B} := (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_d) : \mathbb{R}^d \to L^2[0, 1]$.

For this distribution of $(\boldsymbol{X}, Y)$, we can prove that it belongs to the distribution class

$\mathfrak{M}(\alpha, \beta, \tau)$.

Lemma 11. *For* $(\boldsymbol{X}, Y)$ *constructed in* (F.1)*, it holds that*

i) $\|\Gamma\| \leqslant C$, $\lambda_{\min}(\Gamma|_{\mathcal{S}_e}) \geqslant c$ *and* $c \leqslant \lambda_d(\Gamma_e) \leqslant \lambda_1(\Gamma_e) \leqslant C$ *for two positive constants* c *and* C *that do not depend on* $\boldsymbol{\theta}$, n, *and* m;

ii) *the central curve* $\boldsymbol{m}(y) = \mathbb{E}[\boldsymbol{X}|Y = y]$ *is weak sliced stable with respect to* Y;

iii) $(\boldsymbol{X}, Y) \in \mathcal{F}(\alpha, \beta, c_1, c_2)$.

Furthermore, we have the following upper bound on the pairwise KL-divergence:

Lemma 12. *Let* $\mathbb{P}_{\boldsymbol{\theta}}$ *and* $\mathbb{P}_{\boldsymbol{\theta}'}$ *be the joint distributions of* $(\boldsymbol{X}, Y)$ *induced by* $\boldsymbol{\beta}_i^{\boldsymbol{\theta}}(i \in [d])$ *and* $\boldsymbol{\beta}_i^{\boldsymbol{\theta}'}(i \in [d])$ *respectively. Then we have*

$$\mathrm{KL}(\mathbb{P}_{\boldsymbol{\theta}}, \mathbb{P}_{\boldsymbol{\theta}'}) \lesssim m^{-\alpha} \sum_{i=1}^{d} \|\boldsymbol{\beta}_i^{\theta} - \boldsymbol{\beta}_i^{\theta'}\|^2.$$

Since

$$\sum_{i=1}^{d} \|\boldsymbol{\beta}_i^{\theta} - \boldsymbol{\beta}_i^{\theta'}\|^2 \lesssim m^{-2\beta} \|\boldsymbol{\theta} - \boldsymbol{\theta}'\|^2,$$

we have

$$\mathrm{KL}(\mathbb{P}_{\boldsymbol{\theta}}, \mathbb{P}_{\boldsymbol{\theta}'}) \leqslant C(d) m^{-(\alpha+2\beta)} \sum_{i=1}^{d} \|\boldsymbol{\beta}_i^{\theta} - \boldsymbol{\beta}_i^{\theta'}\|^2 \leqslant C(d) m^{-(\alpha+2\beta-1)}.$$

We can now prove the minimax lower bound using Lemma 8:

$$
\begin{aligned}
&\inf_{\widehat{\mathcal{S}}_{Y|\boldsymbol{X}}} \sup_{\mathcal{M}\in\mathfrak{M}(\alpha,\beta,\tau)} \mathbb{P}_{\mathcal{M}}\left(\left\|P_{\widehat{\mathcal{S}}_{Y|\boldsymbol{X}}} - P_{\mathcal{S}_{Y|\boldsymbol{X}}}\right\|^2 \geqslant \vartheta n^{-\frac{2\beta-1}{\alpha+2\beta}}\right) \\
\geqslant & \inf_{\widehat{\mathcal{S}}_{Y|\boldsymbol{X}}} \sup_{\boldsymbol{\theta}\in\Theta} \mathbb{P}_{\boldsymbol{\theta}}\left(\|P_{\widehat{\mathcal{S}}_{Y|\boldsymbol{X}}} - P_{\mathcal{S}(\boldsymbol{\theta})}\|^2 \geqslant \vartheta n^{\frac{-2\beta+1}{\alpha+2\beta}}\right) \\
\geqslant & 1 - \frac{\max \mathrm{KL}(\mathbb{P}^n_{\boldsymbol{\theta}}, \mathbb{P}^n_{\boldsymbol{\theta}'}) + \log(2)}{\log(|\Theta|)} \\
\geqslant & 1 - \frac{C(d) n m^{-(\alpha+2\beta-1)} + \log 2}{\frac{m}{2}\log(2)} \\
= & 1 - \frac{C(d)\tilde{C}^{-(\alpha+2\beta)} m + \log(2)}{\frac{m}{2}\log(2)} \geqslant 0.9
\end{aligned}
$$

where the constant ϑ comes from Lemma 10 and in the last equation we have chosen $m = \widetilde{C} n^{\frac{1}{\alpha+2\beta}}$ for some $\widetilde{C} \geqslant \left(\frac{\log(2)}{30C(d)}\right)^{-\frac{1}{\alpha+2\beta}}$. This finishes the proof of Theorem 3.

F.2 Proof of Lemma 10

Proof. Note that $P_{\mathcal{S}(\boldsymbol{\theta})} = \sum_{i=1}^d \widetilde{\boldsymbol{\beta}}_i^{\boldsymbol{\theta}} \otimes \widetilde{\boldsymbol{\beta}}_i^{\boldsymbol{\theta}}$, where $\widetilde{\boldsymbol{\beta}}_i^{\boldsymbol{\theta}}$ is the normalization of $\boldsymbol{\beta}_i^{\boldsymbol{\theta}}$. By the definition of $\boldsymbol{\beta}_i^{\boldsymbol{\theta}}$, for any $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$, $\langle \boldsymbol{\beta}_i^{\boldsymbol{\theta}}, \boldsymbol{\beta}_j^{\boldsymbol{\theta}'}\rangle = 0$ if $i \neq j$. We have

$$\left\|P_{\mathcal{S}(\boldsymbol{\theta})} - P_{\mathcal{S}(\boldsymbol{\theta}')}\right\|^2 \geqslant \left\|(P_{\mathcal{S}(\boldsymbol{\theta})} - P_{\mathcal{S}(\boldsymbol{\theta}')})(\widetilde{\boldsymbol{\beta}}_1^{\boldsymbol{\theta}})\right\|^2 = \left\|\widetilde{\boldsymbol{\beta}}_1^{\boldsymbol{\theta}} - \langle \widetilde{\boldsymbol{\beta}}_1^{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\beta}}_1^{\boldsymbol{\theta}'}\rangle \widetilde{\boldsymbol{\beta}}_1^{\boldsymbol{\theta}'}\right\|^2$$

$$= \frac{1}{\|\boldsymbol{\beta}_1^{\boldsymbol{\theta}}\|^2} \left\| \sum_{k=m+1}^{2m} (\theta_{k-m} - \langle \widetilde{\boldsymbol{\beta}}_1^{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\beta}}_1^{\boldsymbol{\theta}'}\rangle \theta'_{k-m}) k^{-\beta} \phi_k \right\|^2$$

$$= \frac{1}{\|\boldsymbol{\beta}_1^{\boldsymbol{\theta}}\|^2} \left\| \sum_{\substack{k\in\{m+1,\dots,2m\} \\ \theta_{k-m}\neq\theta'_{k-m}}} (\theta_{k-m} - \langle \widetilde{\boldsymbol{\beta}}_1^{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\beta}}_1^{\boldsymbol{\theta}'}\rangle \theta'_{k-m}) k^{-\beta} \phi_k + \sum_{\substack{k\in\{m+1,\dots,2m\} \\ \theta_{k-m}=\theta'_{k-m}}} (\theta_{k-m} - \langle \widetilde{\boldsymbol{\beta}}_1^{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\beta}}_1^{\boldsymbol{\theta}'}\rangle \theta'_{k-m}) k^{-\beta} \phi_k \right\|^2$$

$$\geqslant \frac{1}{\|\boldsymbol{\beta}_1^{\boldsymbol{\theta}}\|^2} \left\| \sum_{\substack{k\in\{m+1,\dots,2m\} \\ \theta_{k-m}\neq\theta'_{k-m}}} (\theta_{k-m} - \langle \widetilde{\boldsymbol{\beta}}_1^{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\beta}}_1^{\boldsymbol{\theta}'}\rangle \theta'_{k-m}) k^{-\beta} \phi_k \right\|^2$$

$$\geqslant \frac{1}{\|\boldsymbol{\beta}_1^{\boldsymbol{\theta}}\|^2} \left\| \sum_{\substack{k\in\{m+1,\dots,2m\} \\ \theta_{k-m}\neq\theta'_{k-m}}} k^{-\beta} \phi_k \right\|^2 = \frac{1}{\|\boldsymbol{\beta}_1^{\boldsymbol{\theta}}\|^2} \sum_{\substack{k\in\{m+1,\dots,2m\} \\ \theta_{k-m}\neq\theta'_{k-m}}} k^{-2\beta}.$$

From the property (2) of Lemma 9, we know that there are at least $m/8$ k's satisfying $k \in \{m+1, \dots, 2m\}, \theta_{k-m} \neq \theta'_{k-m}$. Thus

$$\frac{1}{\|\boldsymbol{\beta}_1^{\boldsymbol{\theta}}\|^2} \sum_{\substack{k\in\{m+1,\dots,2m\} \\ \theta_{k-m}\neq\theta'_{k-m}}} k^{-2\beta} \geqslant \frac{1}{(1 + \sum_{k=m+1}^{2m} k^{-2\beta})^2} \sum_{k=15m/8}^{2m} k^{-2\beta} \asymp m^{-2\beta+1}.$$

□

F.3 Proof of Lemma 11

Proof. **Proof of i)**

It is easy to check that $\|\Gamma\| = 1$. Now we give a lower bound of $\lambda_{\min}(\Gamma|_{\mathcal{S}_e})$. For any unit function $\boldsymbol{u} \subseteq \operatorname{Im}(\Gamma_e) = \Gamma \mathcal{S}_{Y|\boldsymbol{X}}$, let $\boldsymbol{u} = \sum_{i=1}^{d} a_i \Gamma \boldsymbol{\beta}_i$, then

$$\begin{aligned} \Gamma \boldsymbol{\beta}_i &= \sum_{k=im+1}^{(i+1)m} \theta_{k-im} k^{-(\alpha+\beta)} \phi_k + i^{-\alpha} \phi_i \\ \boldsymbol{u} &= \sum_{i=1}^{d} a_i \Gamma \boldsymbol{\beta}_i = \sum_{i=1}^{d} (\sum_{k=im+1}^{(i+1)m} a_i \theta_{k-im} k^{-(\alpha+\beta)} \phi_k + a_i i^{-\alpha} \phi_i) \end{aligned}$$

where a_i satisfies

$$\sum_{i=1}^{d} a_i^2 (\sum_{k=im+1}^{(i+1)m} k^{-(2\alpha+2\beta)} + i^{-2\alpha}) = 1.$$

This implies that $\sum_{i=1}^{d} a_i^2 \geqslant 1/4$.

Then

$$\langle \Gamma(\boldsymbol{u}), \boldsymbol{u} \rangle = \sum_{i=1}^{d} a_i^2 [\sum_{k=im+1}^{(i+1)m} k^{-(3\alpha+2\beta)} + i^{-3\alpha}] \geqslant \sum_{i=1}^{d} a_i^2 i^{-3\alpha} \geqslant d^{-3\alpha} \sum_{i=1}^{d} a_i^2 \geqslant d^{-3\alpha}/4$$

which means

$$\lambda_{\min}(\Gamma|_{\mathcal{S}_e}) \geqslant d^{-3\alpha}/4.$$

Next we show that there exist positive constants λ_- and λ_+ such that $\lambda_- \leqslant \lambda_d(\Gamma_e) \leqslant \lambda_1(\Gamma_e) \leqslant \lambda_+$.

On the one hand, since $\boldsymbol{X}$ is a Gaussian process, the linearity condition holds and $\operatorname{Im}(\Gamma_e) \subseteq \Gamma \mathcal{S}_{Y|\boldsymbol{X}}$ (Lian and Li, 2014). Thus, $\operatorname{rank}(\Gamma_e) \leqslant d$. On the other hand, let $\boldsymbol{Z} = (\boldsymbol{B}^* \Gamma \boldsymbol{B})^{-1/2} \boldsymbol{B}^* \boldsymbol{X}$ and $\Gamma_{ez} = \operatorname{Cov}(\mathbb{E}[\boldsymbol{Z}|Y])$. Then by Lemma 15, since $\boldsymbol{B}^* \Gamma \boldsymbol{B} = \operatorname{diag}\{\sum_{k=im+1}^{(i+1)m} k^{-(\alpha+2\beta)} + i^{-\alpha}\}_{i=1}^{d}$ is positive definite and invertible, we have

$$\boldsymbol{Z} \sim N(0, \boldsymbol{I}_d) \quad \text{and} \quad \Gamma_{ez} = (\boldsymbol{B}^* \Gamma \boldsymbol{B})^{-1/2} \boldsymbol{B}^* \Gamma_e \boldsymbol{B} (\boldsymbol{B}^* \Gamma \boldsymbol{B})^{-1/2}.$$

Thus, $\mathrm{rank}(\Gamma_e) \geqslant \mathrm{rank}(\Gamma_{ez}) = d$. We conclude that $\mathrm{rank}(\Gamma_e) = \mathrm{rank}(\Gamma_{ez}) = d$, i.e., the coverage condition holds and $\mathrm{Im}(\Gamma_e) = \Gamma\mathcal{S}_{Y|\boldsymbol{X}}$.

For the unit function $\boldsymbol{u} \subseteq \mathrm{Im}(\Gamma_e) = \Gamma\mathcal{S}_{Y|\boldsymbol{X}}$, we have $\langle \Gamma_e(\boldsymbol{u}), \boldsymbol{u}\rangle = \mathrm{var}(\mathbb{E}[\langle \boldsymbol{X}, \boldsymbol{u}\rangle | Y])$. Then we know that $\mathbb{E}[\langle \boldsymbol{X}, \boldsymbol{u}\rangle | \boldsymbol{B}^*\boldsymbol{X}] = \boldsymbol{u}^*\Gamma\boldsymbol{B}(\boldsymbol{B}^*\Gamma\boldsymbol{B})^{-1}\boldsymbol{B}^*\boldsymbol{X}$ by Lemma 14.

By the law of total expectation,

$$\mathbb{E}[\langle \boldsymbol{X}, \boldsymbol{u}\rangle | Y] = \mathbb{E}[\mathbb{E}[\langle \boldsymbol{X}, \boldsymbol{u}\rangle | \boldsymbol{B}^*\boldsymbol{X}] | Y] = \mathbb{E}[\boldsymbol{u}^*\Gamma\boldsymbol{B}(\boldsymbol{B}^*\Gamma\boldsymbol{B})^{-1}\boldsymbol{B}^*\boldsymbol{X} | Y] = \boldsymbol{u}^*\Gamma\boldsymbol{B}(\boldsymbol{B}^*\Gamma\boldsymbol{B})^{-1/2}\mathbb{E}[\boldsymbol{Z}|Y],$$

thus $\langle \Gamma_e(\boldsymbol{u}), \boldsymbol{u}\rangle = \boldsymbol{u}^*\Gamma\boldsymbol{B}(\boldsymbol{B}^*\Gamma\boldsymbol{B})^{-1/2}\Gamma_{ez}(\boldsymbol{B}^*\Gamma\boldsymbol{B})^{-1/2}\boldsymbol{B}^*\Gamma\boldsymbol{u}$ where $\Gamma_{ez} := \mathrm{var}(\mathbb{E}[\boldsymbol{Z}|Y])$.

Upper bound on $\lambda_1(\Gamma_e)$. We immediately have

$$\begin{aligned}
\|\Gamma_e\| =& \|\Gamma\boldsymbol{B}(\boldsymbol{B}^*\Gamma\boldsymbol{B})^{-1/2}\Gamma_{ez}(\boldsymbol{B}^*\Gamma\boldsymbol{B})^{-1/2}\boldsymbol{B}^*\Gamma\| \\
\leqslant& \|\Gamma^{1/2}\boldsymbol{B}(\boldsymbol{B}^*\Gamma\boldsymbol{B})^{-1/2}\Gamma_{ez}(\boldsymbol{B}^*\Gamma\boldsymbol{B})^{-1/2}\boldsymbol{B}^*\Gamma^{1/2}\|\|\Gamma^{1/2}\|^2 \leqslant \|\Gamma\|\|\Gamma_{ez}\|
\end{aligned}$$

where $\Gamma^{1/2}$ is the unique positive operator such that $\Gamma^{1/2}\Gamma^{1/2} = \Gamma$, i.e., $\Gamma^{1/2} := \sum_{j=1}^{\infty} j^{-\alpha/2}\phi_j \otimes \phi_j$.

Lower bound on $\lambda_d(\Gamma_e)$. By Lemma 13 (min-max theorem), we have

$$\begin{aligned}
\lambda_i(\Gamma_e) =& \lambda_i(\Gamma\boldsymbol{B}(\boldsymbol{B}^*\Gamma\boldsymbol{B})^{-1/2}\Gamma_{ez}(\boldsymbol{B}^*\Gamma\boldsymbol{B})^{-1/2}\boldsymbol{B}^*\Gamma) \\
=& \lambda_i(\Gamma^{1/2}\boldsymbol{B}(\boldsymbol{B}^*\Gamma\boldsymbol{B})^{-1/2}\Gamma_{ez}(\boldsymbol{B}^*\Gamma\boldsymbol{B})^{-1/2}\boldsymbol{B}^*\Gamma^{1/2}\Gamma) \\
\geqslant& \lambda_i(\Gamma^{1/2}\boldsymbol{B}(\boldsymbol{B}^*\Gamma\boldsymbol{B})^{-1/2}\Gamma_{ez}(\boldsymbol{B}^*\Gamma\boldsymbol{B})^{-1/2}\boldsymbol{B}^*\Gamma^{1/2})\lambda_{\min}(\Gamma|_{\mathcal{S}_e}) \geqslant \lambda_{\min}(\Gamma|_{\mathcal{S}_e})\lambda_i(\Gamma_{ez}).
\end{aligned}$$

Since $\|\Gamma\|$, $d^{-3\alpha}/4$, and the matrix Γ_{ez} do not depend on n, we conclude the existence of the constants λ_- and λ_+.

Proof of ii)

Next we show that the central curve $\boldsymbol{m}(y) = \mathbb{E}[\boldsymbol{X}|Y = y]$ is weak sliced stable with respect to Y. The WSSC for $\mathbb{E}[\boldsymbol{Z}|Y = y]$ implies WSSC for $\mathbb{E}[\boldsymbol{X}|Y = y]$ since

$$\begin{aligned}
&\frac{1}{H}\sum_{h=0}^{H-1} \operatorname{var}\left(\langle \boldsymbol{u}, \mathbb{E}[\boldsymbol{X}|Y]\rangle \mid a_h \leqslant Y \leqslant a_{h+1}\right)\\
=&\frac{1}{H}\sum_{h=0}^{H-1} \operatorname{var}\left(\langle \boldsymbol{u}, \Gamma\boldsymbol{B}(\boldsymbol{B}^*\Gamma\boldsymbol{B})^{-1/2}\mathbb{E}[\boldsymbol{Z}|Y]\rangle \mid a_h \leqslant Y \leqslant a_{h+1}\right)\\
=&\frac{1}{H}\sum_{h=0}^{H-1} \operatorname{var}\left(\langle (\boldsymbol{B}^*\Gamma\boldsymbol{B})^{-1/2}\boldsymbol{B}^*\Gamma\boldsymbol{u}, \mathbb{E}[\boldsymbol{Z}|Y]\rangle \mid a_h \leqslant Y \leqslant a_{h+1}\right)\\
\leqslant&\frac{1}{\tau}\operatorname{var}\left(\langle (\boldsymbol{B}^*\Gamma\boldsymbol{B})^{-1/2}\boldsymbol{B}^*\Gamma\boldsymbol{u}, \mathbb{E}[\boldsymbol{Z}|Y]\rangle\right)\\
=&\frac{1}{\tau}\operatorname{var}\left(\langle \boldsymbol{u}, \mathbb{E}[\boldsymbol{X}|Y]\rangle\right) \quad (\forall \boldsymbol{u} \in \mathbb{S}_{\mathcal{H}})
\end{aligned}$$

where the inequality comes from the WSSC of $\mathbb{E}[\boldsymbol{Z}|Y = y]$ and the fact that $(\boldsymbol{B}^*\Gamma\boldsymbol{B})^{-1/2}\boldsymbol{B}^*\Gamma\boldsymbol{u} \in \mathbb{R}^d$.

Proof of iii) Since $\mathbb{E}[X_i^4] = 3$, we know that Assumption 3 holds by taking c_1 to be 3. Since $\alpha > 1, \frac{1}{2}\alpha + 1 < \beta, \lambda_j = j^{-\alpha}$ and $|b_{ij}| \lesssim j^{-\beta}$ by the definition of $\boldsymbol{\beta}_i^{\boldsymbol{\theta}}$, we know that Assumption 4 holds. Combining these two results with **i)** and **ii)**, we know that **iii)** holds. □

F.4 Proof of Lemma 12

Proof. For simplicity of notation, we define $\boldsymbol{B} := (\boldsymbol{\beta}_1^{\boldsymbol{\theta}}, \ldots, \boldsymbol{\beta}_d^{\boldsymbol{\theta}}) : \mathbb{R}^d \to L^2[0,1]$ and $\boldsymbol{B}' := (\boldsymbol{\beta}_1^{\boldsymbol{\theta}'}, \ldots, \boldsymbol{\beta}_d^{\boldsymbol{\theta}'}) : \mathbb{R}^d \to L^2[0,1]$. Let $\mathbb{E}_{\boldsymbol{\theta}}$ denotes the expectation with respective to $\mathbb{P}_{\boldsymbol{B}}$ and ϕ_d

be the density function for $N(0, \boldsymbol{I}_d)$. Then we have

$$\begin{aligned}
\mathrm{KL}(\mathbb{P}_{\boldsymbol{\theta}}, \mathbb{P}_{\boldsymbol{\theta}'}) &\leqslant \mathrm{KL}(\mathbb{P}_{\boldsymbol{\theta}}, \mathbb{P}_{\boldsymbol{\theta}'}) + \mathbb{E}_{(\boldsymbol{X}, Y)\sim \mathbb{P}_{\boldsymbol{\theta}}} \left(\mathrm{KL}(\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{Z} \mid \boldsymbol{X}, Y), \mathbb{P}_{\boldsymbol{\theta}'}(\boldsymbol{Z} \mid \boldsymbol{X}, Y)\right) \\
&= \mathrm{KL}(\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{X}, \boldsymbol{Z}, Y), \mathbb{P}_{\boldsymbol{\theta}'}(\boldsymbol{X}, \boldsymbol{Z}, Y)) \\
&= \mathbb{E}_{\boldsymbol{\theta}}[\log\left(\frac{\phi_d(\boldsymbol{Z} - \boldsymbol{B}^* \boldsymbol{X})}{\phi_d(\boldsymbol{Z} - \boldsymbol{B}'^* \boldsymbol{X})}\right)] \\
&= \mathbb{E}_{\boldsymbol{\theta}}\left(-\frac{1}{2}\|\boldsymbol{Z} - \boldsymbol{B}^*\boldsymbol{X}\|^2 + \frac{1}{2}\|\boldsymbol{Z} - \boldsymbol{B}^{'*}\boldsymbol{X}\|^2\right) \\
&= \frac{1}{2}\mathbb{E}[\|(\boldsymbol{B} - \boldsymbol{B}')^*\boldsymbol{X}\|^2] \\
&= \frac{1}{2}\mathbb{E}[\sum_{i=1}^{d}\langle \boldsymbol{\beta}_i^{\theta} - \boldsymbol{\beta}_i^{\theta'}, \boldsymbol{X}\rangle^2] \\
&\lesssim m^{-\alpha}\sum_{i=1}^{d}\|\boldsymbol{\beta}_i^{\theta} - \boldsymbol{\beta}_i^{\theta'}\|^2.
\end{aligned}$$

□

G. Assisting Lemmas

Lemma 13 (Minimax theorem)**.** *Assume that A is a positive semi-definite and compact operator with its eigenvalues $\{\widetilde{\lambda}_i\}$ ordered as $\widetilde{\lambda}_1 \geqslant \cdots \geqslant \widetilde{\lambda}_n \geqslant \cdots \geqslant 0$, then*

$$\widetilde{\lambda}_n = \inf_{E_{n-1}} \sup_{x\in E_{n-1}^{\perp}, \|x\|=1} \langle Ax, x\rangle$$

where E_{n-1} with dimension $n-1$ is a closed linear subspace of an Hilbert space $\widetilde{\mathcal{H}}$.

It is a classic result in standard functional analysis textbook.

Lemma 14. *Assume $\boldsymbol{X} = \sum_{i=1}^{\infty} a_i X_i \phi_i \in \mathcal{H}, \boldsymbol{u} = \sum_{i=1}^{N} b_i\phi_i \in \mathcal{H}, \boldsymbol{\beta}_j = \sum_{i=1}^{N} c_{ij}\phi_i \in$*

$\mathcal{H}, \quad \forall j \in [d]$, *then we have*

$$\mathbb{E}[\langle \boldsymbol{X}, \boldsymbol{u} \rangle | \boldsymbol{B}^* \boldsymbol{X}] = \boldsymbol{u}^* \Gamma \boldsymbol{B} (\boldsymbol{B}^* \Gamma \boldsymbol{B})^{-1} \boldsymbol{B}^* \boldsymbol{X}$$

where $\Gamma = \mathbb{E}[\boldsymbol{X} \otimes \boldsymbol{X}], \boldsymbol{B} := (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_d) : \mathbb{R}^d \to L^2[0,1]$.

Proof. Define $\boldsymbol{X'} = \{a_i X_i\}_{i=1}^N \in \mathbb{R}^N, \boldsymbol{u'} = \{b_i\}_{i=1}^N, \boldsymbol{\beta}'_j = \{c_{ij}\}_{i=1}^N, \boldsymbol{B'} = (\boldsymbol{\beta}'_1, \ldots, \boldsymbol{\beta}'_d) \in \mathbb{R}^{p \times d}$. Then using results from multivariate normal distribution, we have

$$\mathbb{E}[\langle \boldsymbol{X}, \boldsymbol{u} \rangle | \boldsymbol{B}^* \boldsymbol{X}] = \mathbb{E}[\langle \boldsymbol{X}', \boldsymbol{u}' \rangle | \boldsymbol{B}^{'\top} \boldsymbol{X}'] = \boldsymbol{u}^{'\top} \boldsymbol{\Sigma} \boldsymbol{B}' (\boldsymbol{B}^{'\top} \boldsymbol{\Sigma} \boldsymbol{B}')^{-1} \boldsymbol{B}^{'\top} \boldsymbol{X}'$$

where Σ is the covariance matrix of $\boldsymbol{X'}$.

We then complete the proof by using the following relationships:

$$\boldsymbol{u}^{'\top} \boldsymbol{\Sigma} \boldsymbol{B}' = \boldsymbol{u}^* \Gamma \boldsymbol{B}, (\boldsymbol{B}^{'\top} \boldsymbol{\Sigma} \boldsymbol{B}')^{-1} = (\boldsymbol{B}^* \Gamma \boldsymbol{B})^{-1}, \boldsymbol{B}^{'\top} \boldsymbol{X}' = \boldsymbol{B}^* \boldsymbol{X}.$$

□

Lemma 15. *If* T *is an operator defined on* $\mathcal{H}_1 \to \mathcal{H}_2$ *where* $\mathcal{H}_i, i = 1, 2$ *is a Hilbert space.* $\boldsymbol{X} \in \mathcal{H}_1$ *is a random element satisfying* $\mathbb{E}[\boldsymbol{X}] = 0$ *. Then we have* $\text{var}(T\boldsymbol{X}) = T\text{var}(\boldsymbol{X})T^*$.

Proof. For any $\boldsymbol{u}_1, \boldsymbol{u}_2 \in \mathcal{H}_2$, we have

$$\langle T\text{var}(\boldsymbol{X})T^* \boldsymbol{u}_1, \boldsymbol{u}_2 \rangle = \langle T\mathbb{E}[\boldsymbol{X} \otimes \boldsymbol{X}]T^* \boldsymbol{u}_1, \boldsymbol{u}_2 \rangle = \langle \mathbb{E}[\boldsymbol{X} \otimes \boldsymbol{X}]T^* \boldsymbol{u}_1, T^* \boldsymbol{u}_2 \rangle$$

since $\mathbb{E}[\boldsymbol{X}] = 0$. By the definition of convariance operator and expectation, we have

$$\langle \mathbb{E}[\boldsymbol{X} \otimes \boldsymbol{X}]T^*\boldsymbol{u}_1, T^*\boldsymbol{u}_2\rangle = \langle \mathbb{E}[\langle \boldsymbol{X}, T^*\boldsymbol{u}_1\rangle \boldsymbol{X}], T^*\boldsymbol{u}_2\rangle = \mathbb{E}[\langle \boldsymbol{X}, T^*\boldsymbol{u}_1\rangle \langle \boldsymbol{X}, T^*\boldsymbol{u}_2\rangle].$$

Similarly, we have

$$\langle \mathrm{var}(T\boldsymbol{X})\boldsymbol{u}_1, \boldsymbol{u}_2\rangle = \langle \mathbb{E}[T\boldsymbol{X} \otimes T\boldsymbol{X}]\boldsymbol{u}_1, \boldsymbol{u}_2\rangle = \mathbb{E}[\langle T\boldsymbol{X}, \boldsymbol{u}_1\rangle \langle T\boldsymbol{X}, \boldsymbol{u}_2\rangle].$$

Then the proof is completed by noticing the following

$$\mathbb{E}[\langle T\boldsymbol{X}, \boldsymbol{u}_1\rangle \langle T\boldsymbol{X}, \boldsymbol{u}_2\rangle] = \mathbb{E}[\langle \boldsymbol{X}, T^*\boldsymbol{u}_1\rangle \langle \boldsymbol{X}, T^*\boldsymbol{u}_2\rangle].$$

□

Lemma 16. *If T is of finite rank, then we have $\lim_{m\to\infty} \|\Pi_m T - T\| = 0$.*

Proof. By the triangle inequality and compatibility of operator norm, one has

$$\|\Pi_m T - T\| \leqslant \|(\Pi_m - I)T\|$$

where $I = \sum_{i=1}^{\infty} \phi_i \otimes \phi_i$ for $\{\phi_i\}_{i\in\mathbb{Z}_{\geqslant 1}}$ being an orthonormal basis of $\mathcal{H}$.

Since T is of finite rank, let us assume that $\{e_i\}_{i=1}^k$ is an orthonormal basis of $\mathrm{Im}(T)$ where $k = \mathrm{rank}(T)$. For any $\beta \in \mathcal{H}$ such that $\|\beta\| = 1$, one has $\|T\beta\| \leqslant \|T\|\|\beta\| = \|T\|$, so one can assume that $T\beta \in \mathrm{Im}(T)$ admits the following expansion under basis $\{e_i\}_{i=1}^k$:

$$T\beta = \sum_{i=1}^{k} b_i e_i, \quad \sum_{i=1}^{k} b_i^2 \leqslant \|T\|^2 < \infty.$$

Thus

$$\|(I-\Pi_m)T\beta\| = \left\| \sum_{i=1}^{k} (I-\Pi_m) b_i e_i \right\| \leqslant \sum_{i=1}^{k} |b_i| \cdot \|(I-\Pi_m)e_i\|.$$

Clearly, $\|(\Pi_m - I)\alpha\|$ $(\forall \alpha \in \mathcal{H})$ tends to 0 as $m \to \infty$ since

$$(I-\Pi_m)\alpha = \left(\sum_{i=m+1}^{\infty} \phi_i \otimes \phi_i \right) \left(\sum_{i=1}^{\infty} c_i \phi_i \right) = \sum_{i=m+1}^{\infty} c_i \phi_i \xrightarrow{m\to\infty} 0$$

where we have assumed that $\alpha = \sum\limits_{i=1}^{\infty} c_i \phi_i$.

Thus $\forall \varepsilon > 0$, there exists some $N_i > 0$ such that $\forall m > N_i$ one has $\|(\Pi_m - I)e_i\| < \varepsilon$, $(\forall i = 1, ..., k)$. Let $N = \max\{N_1, \cdots, N_k\}$, then $\forall m > N$ one has

$$\|(I-\Pi_m)T\beta\| \leqslant \sum_{i=1}^{k} |b_i| \cdot \|(I-\Pi_m)e_i\| \leqslant \sum_{i=1}^{k} |b_i| \varepsilon \leqslant k\varepsilon \|T\|,$$

which means that $\forall m > N$, one has

$$\|(\Pi_m - I)T\| = \sup_{\|\beta\|=1} \|(\Pi_m - I)T\beta\| \leqslant k\varepsilon\|T\|.$$

Thus $\lim\limits_{m\to\infty} \|(\Pi_m - I)T\| = 0$. Then the proof of Lemma 16 is completed. □

G.1 Properties of sliced partition

Lemma 17 (Corollary 1 in Lin et al. (2018b)). *In the slicing inverse regression contexts, recall that S_h denotes the h-th interval $(y_{h-1,c}, y_{h,c}]$ for $2 \leqslant h \leqslant H-1$ and $S_1 = (-\infty, y_{1,c}]$, $S_H = (y_{H-1,c}, \infty)$. We have that $x_{h,i}, i = 1, \cdots, c-1$ can be treated as $c-1$ random samples of $x\Big|(y \in S_h)$ for $h = 1, ..., H-1$ and $x_{H,1}, ..., x_{H,c}$ can be treated as c random samples of $x\Big|(y \in S_H)$.*

Lemma 18 (Lemma 11 in Lin et al. (2018b)). *For any sufficiently large H, c and $n > \frac{4H}{\gamma}+1$, the sliced partition $\mathfrak{S}_H(n)$ is a γ-partition with probability at least*

$$1 - CH^2\sqrt{n+1}\exp\left(-\frac{\gamma^2(n+1)}{32H^2}\right)$$

for some absolute constant C.

Lemma 19 (Lemma 10 in Lin et al. (2018b)). *Suppose that (x, y) are defined over σ-finite space $\mathcal{X} \times \mathcal{Y}$ and g is a non-negative function such that $\mathbb{E}[g(x)]$ exists. For any fixed positive constants $C_1 < 1 < C_2$, there exists a constant C which only depends on C_1, C_2 such that for any partition $\mathbb{R} = \bigcup_{h=1}^{H} S_h$ where S_h are intervals satisfying*

$$\frac{C_1}{H} \leqslant \mathbb{P}(y \in S_h) \leqslant \frac{C_2}{H}, \forall h, \tag{G.1}$$

we have

$$\sup_h \mathbb{E}(g(x)\Big|y \in S'_h) \leqslant CH\mathbb{E}[g(x)].$$

G.2 Sin Theta Theorem

Lemma 20 (Proposition 2.3 in Seelmann (2014))**.** *Let B be a self-adjoint operator on a separable Hilbert space $\widetilde{\mathcal{H}}$, and let $V \in \mathcal{L}(\widetilde{\mathcal{H}})$ be another self-adjoint operator where $\mathcal{L}\left(\widetilde{\mathcal{H}}\right)$ stands for the space of bounded linear operators from a Hilbert space $\widetilde{\mathcal{H}}$ to $\widetilde{\mathcal{H}}$. Write the spectra of B and $B+V$ as*

$$\operatorname{spec}(B) = \sigma \cup \Sigma \quad \textit{and} \quad \operatorname{spec}(B+V) = \omega \cup \Omega$$

with $\sigma \cap \Sigma = \varnothing = \omega \cap \Omega$, and suppose that there is $\widehat{d} > 0$ such that

$$\operatorname{dist}(\sigma, \Omega) \geqslant \widehat{d} \quad \textit{and} \quad \operatorname{dist}(\Sigma, \omega) \geqslant \widehat{d}$$

where $\mathrm{d}ist(\sigma, \Sigma) := \min\{|a-b| : a \in \sigma, b \in \Omega\}$. Then, it holds that

$$\|P_B(\sigma) - P_{B+V}(\omega)\| \leqslant \frac{\pi}{2} \frac{\|V\|}{\widehat{d}}$$

where $P_B(\sigma)$ denotes the spectral projection for B associated with σ, i.e.,

$$P_B(\sigma) := \frac{1}{2\pi \mathrm{i}} \oint_\gamma \frac{\mathrm{d}z}{z-B},$$

where γ is a contour on $\mathbb{C}$ that encloses σ but no other elements of $\operatorname{spec}(B)$.

Remark 2 (Spectral projection)**.** We note that, if further B is compact, the spectral projection $P_B(\sigma)$ coincide with the projection operator onto the closure of the space spanned by the eigenfunctions associated with the eigenvalues in σ.

Specifically, if B is compact, by the spectral decomposition theorem one has

$$B = \sum_{i=1}^{\infty} \mu_i e_i \otimes e_i \quad \text{and} \quad (z - B)^{-1} = \sum_{i=1}^{\infty} (z - \mu_i)^{-1} e_i \otimes e_i,$$

where $\mathrm{spec}(B) := \{\mu_i\}_{i=1}^{\infty}$ satisfies $|\mu_i| \xrightarrow{i \to \infty} 0$. Then $\forall v \in \mathcal{H}$, it holds that

$$\begin{aligned} P_B(\sigma)v &= \frac{1}{2\pi \mathrm{i}} \oint_\gamma (z - B)^{-1} v \, \mathrm{d}z = \frac{1}{2\pi \mathrm{i}} \oint_\gamma \sum_{i=1}^{\infty} (z - \mu_i)^{-1} \langle e_i, v \rangle e_i \, \mathrm{d}z \\ &= \sum_{i=1}^{\infty} \left[\left(\frac{1}{2\pi \mathrm{i}} \oint_\gamma (z - \mu_i)^{-1} \, \mathrm{d}z \right) \langle e_i, v \rangle e_i \right] = \sum_{i \in \{i : \mu_i \in \sigma\}} \langle e_i, v \rangle e_i. \end{aligned}$$

In particular, if $\sigma = \mathrm{spec}(B) \backslash \{0\}$, then $P_B(\sigma)$ is the projection operator onto the $\overline{\mathrm{Im}}(B)$.

G.3 Wely inequality for a self-adjoint and compact operator

Lemma 21. *Let M and N be two self-adjoint, positive semi-definite and compact operators defined on a Hilbert space $\widetilde{\mathcal{H}}$ with their respective eigenvalues $\{\mu_i\}, \{\nu_i\}$ ordered as follows*

$$M : \mu_1 \geqslant \cdots \geqslant \mu_n \geqslant \cdots \geqslant 0 \quad \textit{and} \quad N : \nu_1 \geqslant \cdots \geqslant \nu_n \geqslant \cdots \geqslant 0.$$

Then the following inequalities hold: $|\mu_k - \nu_k| \leqslant \|M - N\|$, $k \geqslant 1$.

H. Additional Simulation Results

This section contains additional simulation results of Section 4. Specifically, in Section H.1, we give an explanation of why $\boldsymbol{X}$ in Model I to III is equivalent to a construction that satisfies the assumption that Γ is non-singular. In Section H.2, we provide guidelines for the choice of H in practice. Sections H.3 and H.4 contain additional simulation results of

Sections 4.2 and 4.3 respectively when (i): $\varepsilon \sim N(0,1), H = 10$; (ii):$\varepsilon \sim N(0,0.25), H = 10$; (iii):$\varepsilon \sim N(0,2), H = 10$. Lastly, we compare FSIR with PCA on selecting optimal m in Section H.5.

H.1 The explanation on difference between simulation models and the theoretical assumptions

Here, we explain why the construction of $\boldsymbol{X}$ in simulation models do not contradict the assumption that Γ is non-singular.

Assume $\boldsymbol{X} = \sum_{i=1}^{\infty} a_i X_i \phi_i \in \mathcal{H}$ and $\boldsymbol{\beta}_j = \sum_{i=1}^{N} c_{ij}\phi_i \in \mathcal{H},\ \forall j \in [d]$. Then, we have

$$
\begin{aligned}
\mathbb{E}[\boldsymbol{X}|f(\boldsymbol{B}^*\boldsymbol{X},\epsilon)] =& \mathbb{E}[\sum_{i=1}^{N} a_i X_i \phi_i | f(\boldsymbol{B}^*\boldsymbol{X},\epsilon)] + \mathbb{E}[\sum_{i=N+1}^{\infty} a_i X_i \phi_i | f(\boldsymbol{B}^*\boldsymbol{X},\epsilon)] \\
=& \mathbb{E}[\sum_{i=1}^{N} a_i X_i \phi_i | f(\boldsymbol{B}^*\boldsymbol{X},\epsilon)] + \mathbb{E}[\sum_{i=N+1}^{\infty} a_i X_i \phi_i] \\
=& \mathbb{E}[\sum_{i=1}^{N} a_i X_i \phi_i | f(\boldsymbol{B}^*\boldsymbol{X},\epsilon)].
\end{aligned}
$$

Thus, in terms of Γ_e, the truncation on $\boldsymbol{\beta}_j$ can be transferred to the truncation on $\boldsymbol{X}$. Note that the SIR estimate only involves Γ_m and Γ_e. This suggests that when $\boldsymbol{X}$ has infinitely many terms, the SIR estimate remains the same before and after we do truncation on $\boldsymbol{X}$ as long as m is smaller than N. Therefore, we directly simulate the truncated version of $\boldsymbol{X}$ in Model I to III.

H.2 Guidelines on selecting H

In practical scenarios, the selection of the slices number H can be influential. Our simulation studies suggest that selecting $H \geq \ln(n)$ is often sufficient to achieve desirable numerical results and this selection meets the theoretical requirement that $H > H_0$ since $\ln(n)$ will eventually exceed the constant H_0 defined in Theorem 2. Furthermore, we recommend selecting H within the range $[10, 35]$ to accommodate finite-sample scenarios in practice.

We conducted a series of experiments to substantiate the practicality of this guideline. Specifically, we simulated the FSIR process as described in Section 4.3 and determined the minimum average subspace estimation error across 100 repetitions for various values of m in the set $\{2, 3, \ldots, 13, 14, 20, 30, 40\}$. We first set $n = 20,000$. The value of H is initiated at a baseline value $\ln(n) = \ln(20000) \approx 10$ and increased in increments of 5. The results are presented in Table 2, which illustrates that the subspace estimation error is relatively insensitive to variations in H as long as $H \geq \ln(n)$. For comparative purposes, at an exceptionally low H value, such as $H = 2$, the errors for the three models are recorded at 0.077, 0.291, and 0.02, respectively, which are significantly higher than those obtained for $H \geq 10$. This comparison highlights the necessity of adhering to the guideline of $H \geq \ln(n)$ for robust model performances. We then expanded the dataset sizes to $n = 50,000$ and $n = 200,000$ respectively and conducted the experiments with H chosen in the same way as before; the results are consistent with earlier findings. These empirical results substantiate the practicality of our proposed guideline.

FSIR-OT	H	2	10	15	20	25	30
$n = 20000$	Model I	**0.077**	**0.067**	**0.06**	**0.066**	**0.066**	**0.065**
	Model II	**0.291**	**0.024**	**0.03**	**0.026**	**0.028**	**0.026**
	Model III	**0.02**	**0.01**	**0.01**	**0.01**	**0.01**	**0.01**
	H	2	11	16	21	26	31
$n = 50000$	Model I	**0.060**	**0.051**	**0.052**	**0.053**	**0.052**	**0.050**
	Model II	**0.464**	**0.016**	**0.015**	**0.015**	**0.016**	**0.016**
	Model III	**0.013**	**0.010**	**0.009**	**0.009**	**0.010**	**0.009**
	H	2	13	18	23	28	33
$n = 200000$	Model I	**0.041**	**0.036**	**0.035**	**0.036**	**0.036**	**0.036**
	Model II	**0.338**	**0.008**	**0.007**	**0.008**	**0.008**	**0.008**
	Model III	**0.007**	**0.004**	**0.004**	**0.005**	**0.004**	**0.005**

Table 2: The minimum average subspace estimation error over 100 repeated experiments with respect to different m of FSIR-OT for various models in the case of $\varepsilon \sim N(0, 2)$.

H.3 Additional simulation results of Section 4.2

The left panels of Figures 4 - 6 are the average subspace estimation error under Model (I) where n ranges in $\{2\times10^3, 2\times10^4, 5\times10^4, 2\times10^5, 5\times10^5, 10^6\}$, m ranges in $\{3, 4, \ldots, 25\}$. The optimal value of m (denoted by m^*) for each n is marked with a red circle. The shaded areas represent the standard error bands associated with these estimates (all smaller than 0.011). The right panel of Figures 4 - 6 illustrate the linear dependence of $\log(m^*)$ on $\log(n)$. The solid line characterizes the linear trend of $\log(m^*)$ against $\log(n)$. The dotted lines are their least-squares fittings, with their slopes estimated as 0.183, 0.2 and 0.196 respectively, which are close to the theoretical value of 2/11. These results are consistent with the theoretically optimal choice of m in FSIR-OT.

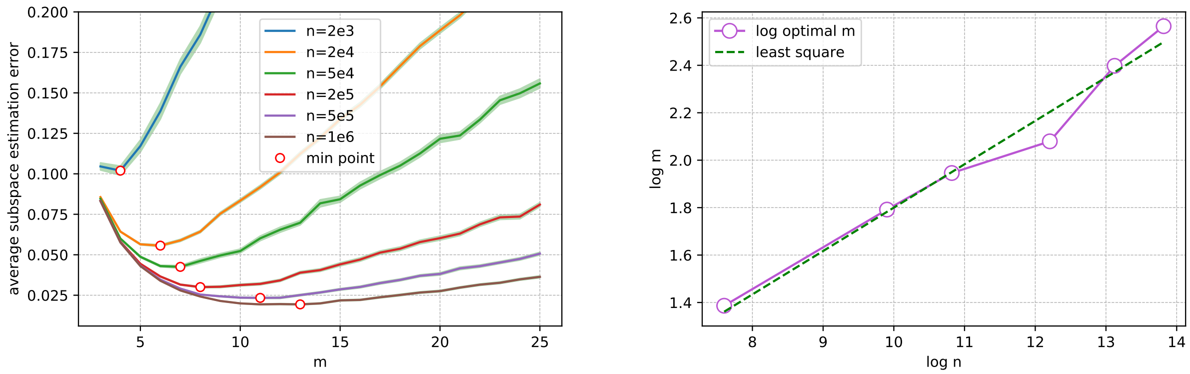

Figure 4: Experiments for the optimal choice of truncation parameter m with $\varepsilon \sim N(0,1)$ and $H = 10$. Left: average subspace estimation error with increasing m for different n. Right: linear trend of $\log(m^*)$ against $\log(n)$, with a slope of 0.183 and $R^2 > 0.98$.

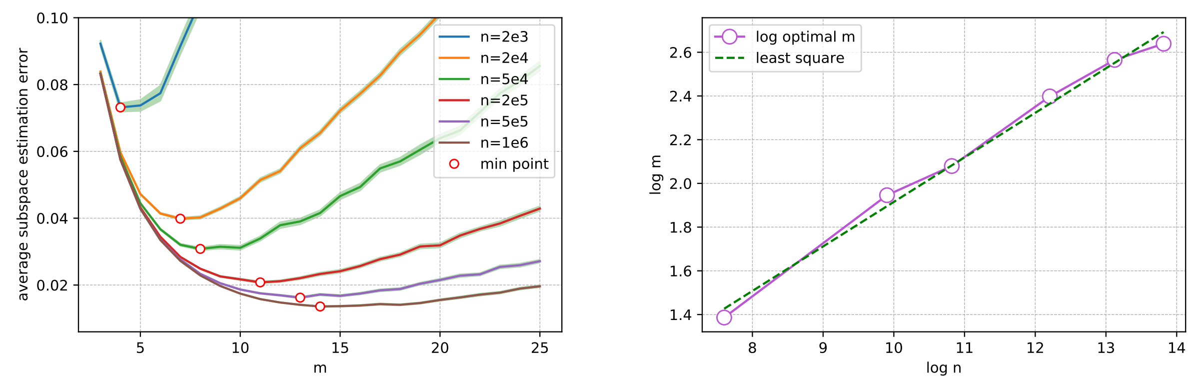

Figure 5: Experiments for the optimal choice of truncation parameter m with $\varepsilon \sim N(0,0.25)$ and $H = 10$. Left: average subspace estimation error with increasing m for different n. Right: linear trend of $\log(m^*)$ against $\log(n)$, with a slope of 0.2 and $R^2 > 0.99$.

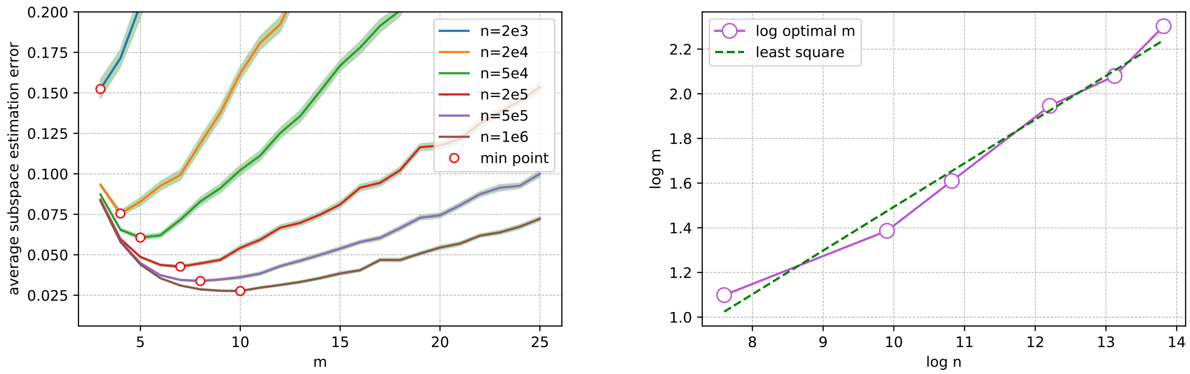

Figure 6: Experiments for the optimal choice of truncation parameter m with $\varepsilon \sim N(0, 2)$ and $H = 10$. Left: average subspace estimation error with increasing m for different n. Right: linear trend of $\log(m^*)$ against $\log(n)$, with a slope of 0.196 and $R^2 > 0.98$.

H.4 Additional simulation results of Section 4.3

For each model of Models (I) - (III), we calculate the average subspace estimation error of FSIR-OT and RFSIR based on 100 replications, where $n = 20000$, the truncation parameter of FSIR-OT m ranges in $\{2, 3, \ldots, 13, 14, 20, 30, 40\}$, and the regularization parameter in RFSIR ρ ranges in $0.01 \times \{1, 2, \cdots, 9, 10, 15, 20, 25, 30, 40, \cdots, 140, 150\}$. Detailed results are presented in Figures 7 - 9, where we mark the minimal error in each model with red '$\times$'. The shaded areas represent the corresponding standard errors, all of which are less than 0.01. When $\varepsilon \sim N(0, 1), H = 10$, for FSIR-OT, the minimal errors for $\mathcal{M}_1$, $\mathcal{M}_2$, and $\mathcal{M}_3$ are $0.06, 0.02$, and 0.01 respectively. For RFSIR, the corresponding minimal errors are $0.08, 0.06$, and 0.01. When $\varepsilon \sim N(0, 0.25), H = 10$, for FSIR-OT, the minimal errors for $\mathcal{M}_1$, $\mathcal{M}_2$, and $\mathcal{M}_3$ are $0.04, 0.02$, and 0.01 respectively. For RFSIR, the corresponding minimal errors are $0.06, 0.04$, and 0.01. When $\varepsilon \sim N(0, 2), H = 10$, for FSIR-OT, the minimal errors for $\mathcal{M}_1$, $\mathcal{M}_2$, and $\mathcal{M}_3$ are $0.078, 0.032$, and 0.017 respectively. For RFSIR, the corresponding minimal errors are $0.109, 0.105$, and 0.014. The results here suggest that the performance of

FSIR-OT is generally superior to, or at the very least equivalent to, that of the RFSIR.

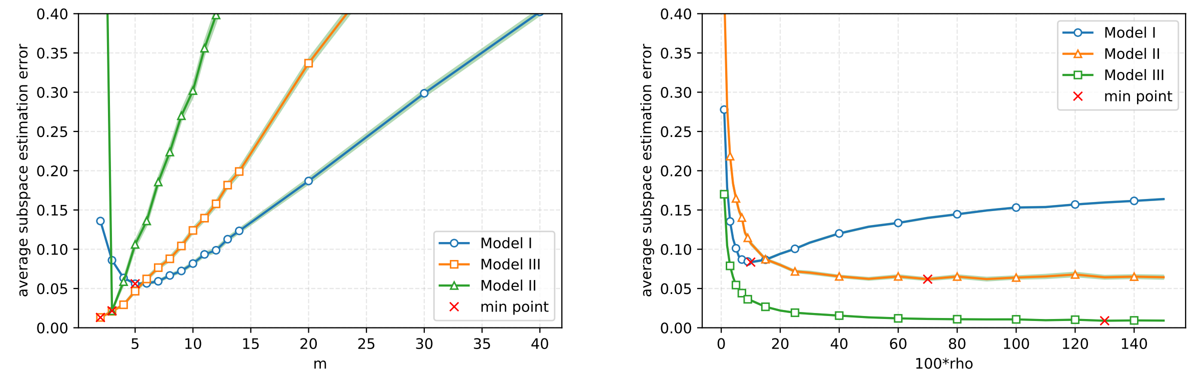

Figure 7: Average subspace estimation error of FSIR-OT and RFSIR for various models in the case of $\varepsilon \sim N(0,1)$ and $H = 10$. The standard errors are all below 0.01. Left: FSIR-OT with different truncation parameter m; Right: RFSIR with different values of the regularization parameter ρ.

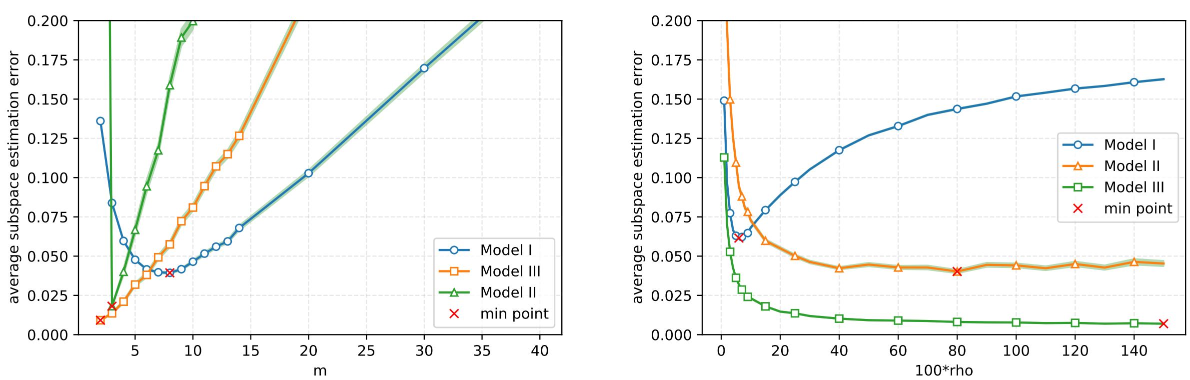

Figure 8: Average subspace estimation error of FSIR-OT and RFSIR for various models in the case of $\varepsilon \sim N(0, 0.25)$ and $H = 10$. Average subspace estimation error of FSIR-OT and RFSIR for various models. The standard errors are all below 0.008. Left: FSIR-OT with different truncation parameter m; Right: RFSIR with different values of the regularization parameter ρ.

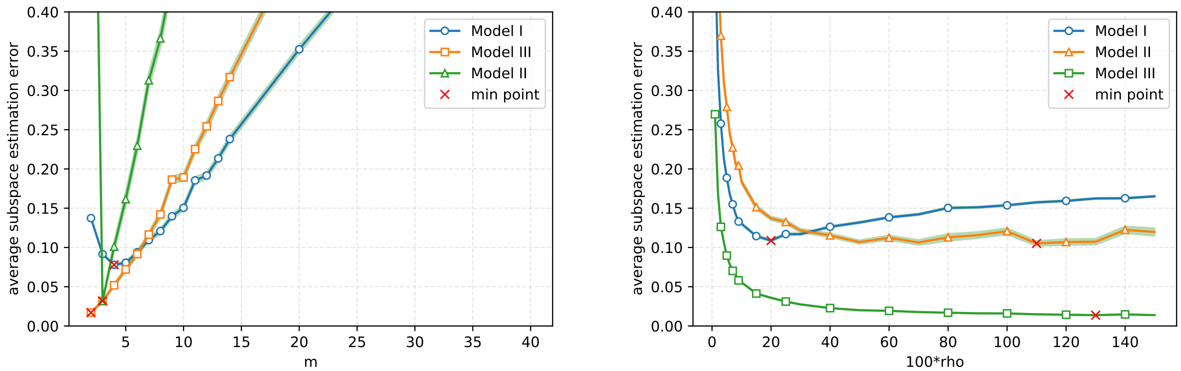

Figure 9: Average subspace estimation error of FSIR-OT and RFSIR for various models in the case of $\varepsilon \sim N(0,2)$ and $H = 10$. Average subspace estimation error of FSIR-OT and RFSIR for various models. The standard errors are all below 0.012. Left: FSIR-OT with different truncation parameter m; Right: RFSIR with different values of the regularization parameter ρ.

H.5 Comparison with PCA on selecting optimal m

SIR is a supervised learning method, whereas PCA is an unsupervised learning method. Consequently, using PCA to select m without any information from the response variable is intuitively incorrect.

Specifically, we attempt to use PCA to select m and highlight the drawbacks of this approach in the following. Here we consider the bike sharing data set studied in Section 4.4. First, we calculate the proportion of the total eigenvalue sum explained by the first i ($i \leq 14$) eigenvalues of $\widehat{\Gamma}$. We find that the first eigenvalue alone accounts for over $99.4\% \geq 99\%$ of the total, suggesting that choosing m as 1 is a good option, and increasing m further adds little value. Under this parameter selection (with d only taking 1), the Gaussian process regression error after SIR dimension reduction is 0.201, significantly higher than the optimal result 0.188 in Table 1, corresponding to $d = 2$ and $m = 6$. Therefore, using the unsupervised learning method PCA does not provide a reasonable way to select m.