SUPPLEMENT TO "FUNCTIONAL LINEAR REGRESSION FROM SPARSE TO DENSE DESIGNS: A POOLING-RIDGE METHOD AND MINIMAX OPTIMALITY"

BY SHUNXING YAN[1,a] AND FANG YAO[*2,b]

[1]*Department of Statistics and Data Science, Tsinghua University* [a]*sxyan@tsinghua.edu.cn*

[2]*School of Mathematical Sciences, Center for Statistical Science, Peking University* [b]*fyao@math.pku.edu.cn*

In this supplementary material, we provide the mathematical proofs, auxiliary lemmas, and additional numerical results for the main manuscript. Section S.1 contains the proofs of all propositions and theorems in Section 3, and the lemmas required are shown in S.2 together with their proofs. Finally, Section S.3 presents additional numerical results that were omitted from Section 4 due to space limitations. Throughout this supplement, all $\mathcal{L}^2$ spaces, norms, inner products, and kernel/covariance integrals on $\mathcal{T}_x$, $\mathcal{T}_y$, and $\mathcal{T}_y \times \mathcal{T}_x$ are understood with respect to the design measures ν_x, ν_y, and $\nu_y \otimes \nu_x$, respectively.

S.1. Proof of main results. This section presents the proofs of Proposition 3.1, Theorem 3.1, Theorem 3.2 and Remark 3.1 for scalar-on-function regression, as well as Proposition 3.2, Theorem 3.3, and Theorem 3.4 for function-on-function regression.

PROOF OF PROPOSITION 3.1. Without loss of generality, throughout the upper-bound proofs below, we will assume that both $\mathcal{H}(K)$ and $\mathcal{H}(\Pi)$ are dense in $\mathcal{L}^2$. To prove the operator $(\widehat{\Gamma} + \lambda I) : \mathcal{H}(K) \to \mathcal{H}(K)$ is positive definite, it suffices to verify the positive definiteness of $(\widehat{\Pi} + \lambda I) : \mathcal{L}^2(\mathcal{T}_x, \nu_x) \to \mathcal{L}^2(\mathcal{T}_x, \nu_x)$. We first study the operator norm of $(\Pi + \lambda I)^{-1/2}(\widehat{\Pi} - \Pi)(\Pi + \lambda I)^{-1/2}$, i.e.

$$\begin{aligned}
&\|(\Pi + \lambda I)^{-1/2}(\widehat{\Pi} - \Pi)(\Pi + \lambda I)^{-1/2}\|_{op} \\
&= \sup_{h_1, h_2 : \|h_1\|_{\mathcal{L}^2} = \|h_2\|_{\mathcal{L}^2} = 1} \langle h_1, (\Pi + \lambda I)^{-1/2}(\widehat{\Pi} - \Pi)(\Pi + \lambda I)^{-1/2} h_2 \rangle_{\mathcal{L}^2}
\end{aligned}$$

[*]Author to whom correspondence may be addressed.

Denote the basic expansion form that $h_1 = \sum_{k=1}^{\infty} a_k \varphi_k$ and $h_2 = \sum_{k=1}^{\infty} b_k \varphi_k$. Then the operator norm is given by

$$
\begin{aligned}
&\sup_{(a_k,b_k)_{k=1}^{\infty}} \left\langle \sum_{k=1}^{\infty} a_k \varphi_k, (\Pi + \lambda I)^{-1/2} (\widehat{\Pi} - \Pi)(\Pi + \lambda I)^{-1/2} \sum_{k=1}^{\infty} b_k \varphi_k \right\rangle_{\mathcal{L}^2} \\
&= \sup_{(a_k,b_k)_{k=1}^{\infty}} \left\langle \sum_{k=1}^{\infty} a_k (\Pi + \lambda I)^{-1/2} \varphi_k, (\widehat{\Pi} - \Pi) \sum_{k=1}^{\infty} (\Pi + \lambda I)^{-1/2} b_k \varphi_k \right\rangle_{\mathcal{L}^2} \\
&= \sup_{(a_k,b_k)_{k=1}^{\infty}} \sum_{k_1=1}^{\infty} \sum_{k_2=1}^{\infty} a_{k_1} b_{k_2} (\gamma_{k_1} + \lambda)^{-1/2} (\gamma_{k_2} + \lambda)^{-1/2} \left\langle \varphi_{k_1}, (\widehat{\Pi} - \Pi) \varphi_{k_2} \right\rangle_{\mathcal{L}^2} \\
&\leq \sup_{(a_k,b_k)_{k=1}^{\infty}} \left(\sum_{k_1=1}^{\infty} \sum_{k_2=1}^{\infty} a_{k_1}^2 b_{k_2}^2 \right)^{1/2} \left(\sum_{k_1=1}^{\infty} \sum_{k_2=1}^{\infty} \frac{1}{(\gamma_{k_1} + \lambda)(\gamma_{k_2} + \lambda)} \left\langle \varphi_{k_1}, (\widehat{\Pi} - \Pi) \varphi_{k_2} \right\rangle_{\mathcal{L}^2}^2 \right)^{1/2} \\
&= \left(\sum_{k_1=1}^{\infty} \sum_{k_2=1}^{\infty} (\gamma_{k_1} + \lambda)^{-1} (\gamma_{k_2} + \lambda)^{-1} \left\langle \varphi_{k_1}, (\widehat{\Pi} - \Pi) \varphi_{k_2} \right\rangle_{\mathcal{L}^2}^2 \right)^{1/2},
\end{aligned}
$$

where the supremum is taken over all $(a_k)_{k=1}^{\infty}$ and $(b_k)_{k=1}^{\infty}$ with $\sum_{k=1}^{\infty} a_k^2 = \sum_{k=1}^{\infty} b_k^2 = 1$ and the fourth line is given by Cauchy inequality. To obtain a probabilistic bound of the last line, we only need to consider its expectation. In the following proof, note that $m_{x,i} \asymp m_x$ for all $i : 1 \leq i \leq n$, for the brevity of notation, we treat all $m_{x,i}$ are the same and all $m_{y,i}$ are the same. All the arguments also apply to the situation when they are different. Specially

$$
\begin{aligned}
\left\langle \varphi_{k_1}, \widehat{\Pi} \varphi_{k_2} \right\rangle_{\mathcal{L}^2} &= \left\langle \varphi_{k_1}, L_K^{-1/2} S_K^* \widehat{\Gamma} S_K^{*-1} L_K^{1/2} \varphi_{k_2} \right\rangle_{\mathcal{L}^2} \\
&= \left\langle S_K L_K^{-1/2} \varphi_{k_1}, \widehat{\Gamma} S_K^{*-1} L_K^{1/2} \varphi_{k_2} \right\rangle_{\mathcal{H}(K)} \\
&= \left\langle S_K^{*-1} L_K^{1/2} \varphi_{k_1}, \widehat{\Gamma} S_K^{*-1} L_K^{1/2} \varphi_{k_2} \right\rangle_{\mathcal{H}(K)}
\end{aligned}
$$

where the last line uses the fact $L_K = S_K^* S_K$. Plugging the definition of $\widehat{\Pi}$ in the above formulation, we have

$$
\begin{aligned}
\left\langle \varphi_{k_1}, \widehat{\Pi} \varphi_{k_2} \right\rangle_{\mathcal{L}^2} &= \left\langle S_K^{*-1} L_K^{1/2} \varphi_{k_1}, \widehat{\Gamma} S_K^{*-1} L_K^{1/2} \varphi_{k_2} \right\rangle_{\mathcal{H}(K)} \\
&= \frac{1}{n} \sum_{i=1}^{n} \frac{1}{m_x(m_x - 1)} \sum_{j_1 \neq j_2} X_{ij_1} X_{ij_2} (L_K^{1/2} \varphi_{k_1})(T_{ij_1}) (L_K^{1/2} \varphi_{k_2})(T_{ij_2}).
\end{aligned}
$$

Note that $\mathbb{E}\left[\left\langle \varphi_{k_1}, (\widehat{\Pi}-\Pi)\varphi_{k_2}\right\rangle_{\mathcal{L}^2}\right]=0$, we have

$$\begin{aligned}
&\mathbb{E}\left[\left\langle \varphi_{k_1}, (\widehat{\Pi}-\Pi)\varphi_{k_2}\right\rangle_{\mathcal{L}^2}^2\right] = \mathrm{Var}\left(\left\langle \varphi_{k_1}, \widehat{\Pi}\varphi_{k_2}\right\rangle_{\mathcal{L}^2}\right)\\
&= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_x(m_x-1)}\sum_{j_1\neq j_2} X_{ij_1}X_{ij_2}(L_K^{1/2}\varphi_{k_1})(T_{ij_1})(L_K^{1/2}\varphi_{k_2})(T_{ij_2})\right)\\
&= \frac{1}{n}\mathrm{Var}\left(\frac{1}{m_x(m_x-1)}\sum_{j_1\neq j_2} X_{ij_1}X_{ij_2}(L_K^{1/2}\varphi_{k_1})(T_{ij_1})(L_K^{1/2}\varphi_{k_2})(T_{ij_2})\right) \qquad (1)\\
&\lesssim \frac{1}{n}\mathbb{E}\left[\left(\frac{1}{m_x^2}\sum_{j_1\neq j_2} X_{ij_1}X_{ij_2}(L_K^{1/2}\varphi_{k_1})(T_{ij_1})(L_K^{1/2}\varphi_{k_2})(T_{ij_2})\right)^2\right].
\end{aligned}$$

Expanding on the last line, it is equal to

(2)

$$\begin{aligned}
&\frac{1}{nm_x^4}\mathbb{E}\left[\sum_{j_1\neq j_2, j_3\neq j_4}\prod_{u=1}^{2} X_{1j_{2u-1}}X_{1j_{2u}}(L_K^{1/2}\varphi_{k_1})(T_{1j_{2u-1}})(L_K^{1/2}\varphi_{k_2})(T_{1j_{2u}})\right]\\
&\lesssim \frac{1}{nm_x^4}\mathbb{E}\left[\sum_{a=2}^{4}\sum_{\substack{j_1\neq j_2, j_3\neq j_4\\ |\{j_1,j_2,j_3,j_4\}|=a}}\prod_{u=1}^{2} X_{1j_{2u-1}}X_{1j_{2u}}(L_K^{1/2}\varphi_{k_1})(T_{1j_{2u-1}})(L_K^{1/2}\varphi_{k_2})(T_{1j_{2u}})\right].
\end{aligned}$$

Here, $|A|$ refers to the number of elements in the set A that are different from each other. Now we calculate each term with different $|\{j_1,j_2,j_3,j_4\}|$. For the case that $|\{j_1,j_2,j_3,j_4\}|=4$, we have

(3)

$$\begin{aligned}
&\frac{1}{nm_x^4}\mathbb{E}\left[\sum_{\substack{j_1\neq j_2, j_3\neq j_4\\ |\{j_1,j_2,j_3,j_4\}|=4}}\prod_{u=1}^{2} X_{1j_{2u-1}}X_{1j_{2u}}(L_K^{1/2}\varphi_{k_1})(T_{1j_{2u-1}})(L_K^{1/2}\varphi_{k_2})(T_{1j_{2u}})\right]\\
&\lesssim \frac{1}{n}\mathbb{E}\left[\left(\iint_{\mathcal{T}_x\times\mathcal{T}_x} X(t_1)X(t_2)(L_K^{1/2}\varphi_{k_1})(t_1)(L_K^{1/2}\varphi_{k_2})(t_2)\,d\nu_x(t_1)\,d\nu_x(t_2)\right)^2\right]\\
&\leq \frac{1}{n}\left(\mathbb{E}\left[\left(\int_{\mathcal{T}_x} X(t)(L_K^{1/2}\varphi_{k_1})(t)\,d\nu_x(t)\right)^4\right]\mathbb{E}\left[\left(\int_{\mathcal{T}_x} X(t)(L_K^{1/2}\varphi_{k_2})(t)\,d\nu_x(t)\right)^4\right]\right)^{1/2}\\
&\lesssim \frac{1}{n}\mathbb{E}\left[\left(\int_{\mathcal{T}_x} X(t)(L_K^{1/2}\varphi_{k_1})(t)\,d\nu_x(t)\right)^2\right]\mathbb{E}\left[\left(\int_{\mathcal{T}_x} X(t)(L_K^{1/2}\varphi_{k_2})(t)\,d\nu_x(t)\right)^2\right] = \frac{1}{n}\gamma_{k_1}\gamma_{k_2}.
\end{aligned}$$

where the second inequality is given Cauchy inequality and the third inequality uses the Assumption 1. Then, for the term that $|\{j_1,j_2,j_3,j_4\}|=3$, we have

$$\frac{1}{nm_x^4}\mathbb{E}\left[\sum_{\substack{j_1\neq j_2, j_3\neq j_4\\ |\{j_1,j_2,j_3,j_4\}|=3}}\prod_{u=1}^{2} X_{1j_{2u-1}}X_{1j_{2u}}(L_K^{1/2}\varphi_{k_1})(T_{1j_{2u-1}})(L_K^{1/2}\varphi_{k_2})(T_{1j_{2u}})\right]$$

$$\begin{aligned}\lesssim \frac{1}{nm_x}\mathbb{E}\Big[\iiint (X(t_1)+\varepsilon)^2 X(t_2)X(t_3)\, \Big(&(L_K^{1/2}\varphi_{k_1})^2(t_1)(L_K^{1/2}\varphi_{k_2})(t_2)(L_K^{1/2}\varphi_{k_2})(t_3)\\ &+(L_K^{1/2}\varphi_{k_1})(t_1)(L_K^{1/2}\varphi_{k_2})(t_1)(L_K^{1/2}\varphi_{k_1})(t_2)(L_K^{1/2}\varphi_{k_2})(t_3)\\ &+(L_K^{1/2}\varphi_{k_2})^2(t_1)(L_K^{1/2}\varphi_{k_1})(t_2)(L_K^{1/2}\varphi_{k_1})(t_3)\Big)\, d\nu_x(t_1)\, d\nu_x(t_2)\, d\nu_x(t_3)\Big]\end{aligned}$$

For the first term, it is equal to

$$\begin{aligned}&\frac{1}{nm_x}\mathbb{E}\left[\left(\int_{\mathcal{T}_x}(X(t_1)+\varepsilon)^2(L_K^{1/2}\varphi_{k_1})^2(t_1)\, d\nu_x(t_1)\right)\left(\int_{\mathcal{T}_x}X(t_2)(L_K^{1/2}\varphi_{k_2})(t_2)\, d\nu_x(t_2)\right)^2\right]\\ &\le \frac{1}{nm_x}\sqrt{\mathbb{E}\left[\left(\int_{\mathcal{T}_x}(X(t_1)+\varepsilon)^2(L_K^{1/2}\varphi_{k_1})^2(t_1)\, d\nu_x(t_1)\right)^2\right]\mathbb{E}\left[\left(\int_{\mathcal{T}_x}X(t_2)(L_K^{1/2}\varphi_{k_2})(t_2)\, d\nu_x(t_2)\right)^4\right]}\\ &\lesssim \frac{1}{nm_x}\int_{\mathcal{T}_x}(L_K^{1/2}\varphi_{k_1})^2(t_1)\, d\nu_x(t_1)\mathbb{E}\left[\left(\int_{\mathcal{T}_x}X(t_2)(L_K^{1/2}\varphi_{k_2})(t_2)\, d\nu_x(t_2)\right)^2\right]\lesssim \frac{\gamma_{k_2}}{nm_x}\|L_K^{1/2}\varphi_{k_1}\|_{\mathcal{L}^2}^2,\end{aligned}$$

where the first inequality applies Cauchy inequality and the second inequality uses the fact

$$\begin{aligned}&\mathbb{E}\left[\left(\int_{\mathcal{T}_x}(X(t_1)+\varepsilon)^2(L_K^{1/2}\varphi_{k_1})^2(t_1)\, d\nu_x(t_1)\right)^2\right]\\ &=\mathbb{E}\left[\iint_{\mathcal{T}_x\times\mathcal{T}_x}(X(t_1)+\varepsilon_1)^2(X(t_2)+\varepsilon_2)^2(L_K^{1/2}\varphi_{k_1})^2(t_1)(L_K^{1/2}\varphi_{k_1})^2(t_2)\, d\nu_x(t_1)\, d\nu_x(t_2)\right]\\ &\lesssim \iint_{\mathcal{T}_x\times\mathcal{T}_x}(L_K^{1/2}\varphi_{k_1})^2(t_1)(L_K^{1/2}\varphi_{k_1})^2(t_2)\, d\nu_x(t_1)\, d\nu_x(t_2)=\left(\int_{\mathcal{T}_x}(L_K^{1/2}\varphi_{k_1})^2(t_1)\, d\nu_x(t_1)\right)^2.\end{aligned}$$

Here, ε_1 and ε_2 are independent copies of ε. By the uniformly bounded fourth-order moments of $X(t)$ and ε and the Cauchy inequality,

$$\sup_{t_1,t_2\in\mathcal{T}_x}\mathbb{E}\left[(X(t_1)+\varepsilon_1)^2(X(t_2)+\varepsilon_2)^2\right]\lesssim 1.$$

Thus, the measurement-error moments enter only the implicit constant, while Assumption 1 controls the fourth moment of the linear functional of X. Similarly, we can bound the third term $\lesssim \gamma_{k_1}\|L_K^{1/2}\varphi_{k_2}\|_{\mathcal{L}^2}^2/nm_x$. Meanwhile, a similar calculation implies that the second term is bounded by the geometric mean of $\gamma_{k_2}\|L_K^{1/2}\varphi_{k_1}\|_{\mathcal{L}^2}^2/nm_x$ and $\gamma_{k_1}\|L_K^{1/2}\varphi_{k_2}\|_{\mathcal{L}^2}^2/nm_x$. Therefore, we have

$$\begin{aligned}&\frac{1}{nm_x^4}\mathbb{E}\left[\sum_{\substack{j_1\neq j_2, j_3\neq j_4\\ |\{j_1,j_2,j_3,j_4\}|=3}}\prod_{u=1}^{2} X_{1j_{2u-1}}X_{1j_{2u}}(L_K^{1/2}\varphi_{k_1})(T_{1j_{2u-1}})(L_K^{1/2}\varphi_{k_2})(T_{1j_{2u}})\right]\\ &\lesssim \frac{1}{nm_x}\left(\gamma_{k_1}\|L_K^{1/2}\varphi_{k_2}\|_{\mathcal{L}^2}^2+\gamma_{k_2}\|L_K^{1/2}\varphi_{k_1}\|_{\mathcal{L}^2}^2\right).\end{aligned}\tag{4}$$

Lastly, we consider the case that for the term that $|\{j_1,j_2,j_3,j_4\}|=2$, we have

(5)
$$\frac{1}{nm_x^4}\mathbb{E}\left[\sum_{\substack{j_1\neq j_2, j_3\neq j_4\\ |\{j_1,j_2,j_3,j_4\}|=2}}\prod_{u=1}^{2}X_{1j_{2u-1}}X_{1j_{2u}}(L_K^{1/2}\varphi_{k_1})(T_{1j_{2u-1}})(L_K^{1/2}\varphi_{k_2})(T_{1j_{2u}})\right]$$
$$\lesssim \frac{1}{nm_x^2}\mathbb{E}\Big[\iint (X(t_1)+\varepsilon_1)^2(X(t_2)+\varepsilon_2)^2\Big((L_K^{1/2}\varphi_{k_1})^2(t_1)(L_K^{1/2}\varphi_{k_2})^2(t_2)$$
$$+(L_K^{1/2}\varphi_{k_1})(t_1)(L_K^{1/2}\varphi_{k_2})(t_1)(L_K^{1/2}\varphi_{k_1})(t_2)(L_K^{1/2}\varphi_{k_2})(t_2)\Big)\,d\nu_x(t_1)\,d\nu_x(t_2)\Big]$$
$$\lesssim \frac{1}{nm_x^2}\Big(\int(L_K^{1/2}\varphi_{k_1})^2(t)\,d\nu_x(t)\int(L_K^{1/2}\varphi_{k_2})^2(t)\,d\nu_x(t)+(\int(L_K^{1/2}\varphi_{k_1})(t)(L_K^{1/2}\varphi_{k_2})(t)\,d\nu_x(t))^2\Big)$$
$$\lesssim \frac{1}{nm_x^2}\|L_K^{1/2}\varphi_{k_1}\|_{\mathcal{L}^2}^2\|L_K^{1/2}\varphi_{k_2}\|_{\mathcal{L}^2}^2.$$

Plugging (3),(4),(5) in (1) and (2), we have

$$\mathbb{E}\left[\left\langle\varphi_{k_1},(\widehat{\Pi}-\Pi)\varphi_{k_2}\right\rangle_{\mathcal{L}^2}^2\right]$$
$$\lesssim \frac{\gamma_{k_1}\gamma_{k_2}}{n}+\frac{\gamma_{k_1}\|L_K^{1/2}\varphi_{k_2}\|_{\mathcal{L}^2}^2+\gamma_{k_2}\|L_K^{1/2}\varphi_{k_1}\|_{\mathcal{L}^2}^2}{nm_x}+\frac{\|L_K^{1/2}\varphi_{k_1}\|_{\mathcal{L}^2}^2\|L_K^{1/2}\varphi_{k_2}\|_{\mathcal{L}^2}^2}{nm_x^2} \tag{6}$$
$$\lesssim \frac{1}{n}k_1^{-2r}k_2^{-2r}+\frac{1}{nm_x}(k_1^{2s-2r}k_2^{-2r}+k_1^{-2r}k_2^{2s-2r})+\frac{1}{nm_x^2}k_1^{2s-2r}k_2^{2s-2r}.$$

In (2), measurement-error terms vanish in the four-distinct-index case, while their expectations in the three- and two-distinct-index cases are uniformly controlled by the fourth-moment assumption. The corresponding numbers of index configurations are of orders m_x^4, m_x^3, and m_x^2, which yield n^{-1}, $(nm_x)^{-1}$, and $(nm_x^2)^{-1}$ after normalization. Hence measurement error affects only the implicit constants, not the rate orders. Therefore,

$$\mathbb{E}\left[\sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty}(\gamma_{k_1}+\lambda)^{-1}(\gamma_{k_2}+\lambda)^{-1}\left\langle\varphi_{k_1},(\widehat{\Pi}-\Pi)\varphi_{k_2}\right\rangle_{\mathcal{L}^2}^2\right]$$
$$\lesssim \sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty}(k_1^{-2r}+\lambda)^{-1}(k_2^{-2r}+\lambda)^{-1}\mathbb{E}\left[\left\langle\varphi_{k_1},(\widehat{\Pi}-\Pi)\varphi_{k_2}\right\rangle_{\mathcal{L}^2}^2\right]$$
$$\lesssim \frac{1}{n}\sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty}\frac{k_1^{-2r}k_2^{-2r}+\frac{1}{m_x}k_1^{2s-2r}k_2^{-2r}+\frac{1}{m_x^2}k_1^{2s-2r}k_2^{2s-2r}}{(k_1^{-2r}+\lambda)(k_2^{-2r}+\lambda)}.$$

Note that

$$\sum_{k=1}^{\infty}\frac{k^{-2r}}{k^{-2r}+\lambda}=\sum_{k=1}^{\infty}\frac{1}{1+\lambda k^{2r}}\lesssim\int_0^{\infty}\frac{1}{1+\lambda x^{2r}}dx$$
$$=\lambda^{-1/2r}\int_0^{\infty}\frac{1}{1+y^{2r}}dy\lesssim\lambda^{-1/2r}.$$

and

$$\sum_{k=1}^{\infty} \frac{k^{2s-2r}}{k^{-2r}+\lambda} = \sum_{k=1}^{\infty} \frac{k^{2s}}{1+\lambda k^{2r}} \lesssim \int_0^{\infty} \frac{x^{2s}}{1+\lambda x^{2r}} dx$$
$$= \lambda^{-(2s+1)/2r} \int_0^{\infty} \frac{y^{2s}}{1+y^{2r}} dy \lesssim \lambda^{-(2s+1)/2r},$$

we have

$$\mathbb{E}\left[\sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty} (\gamma_{k_1}+\lambda)^{-1}(\gamma_{k_2}+\lambda)^{-1}\left\langle \varphi_{k_1}, (\widehat{\Pi}-\Pi)\varphi_{k_2}\right\rangle_{\mathcal{L}^2}^2\right]$$
$$\lesssim \frac{1}{n}\left(\lambda^{-1/2r} + \frac{\lambda^{-(2s+1)/2r}}{m_x}\right)^2 = o(1),$$

where we use the fact $r > s + 1/2$ and $s > 1/2$. Hence

$$\|(\Pi+\lambda I)^{-1/2}(\widehat{\Pi}-\Pi)(\Pi+\lambda I)^{-1/2}\|_{op} = o_p(1). \tag{7}$$

This implies that, with probability tending to 1, the left side is less than $1/2$, which yields

$$-\frac{1}{2}I \preccurlyeq (\Pi+\lambda I)^{-1/2}(\widehat{\Pi}-\Pi)(\Pi+\lambda I)^{-1/2} \preccurlyeq \frac{1}{2}I.$$

Furthermore,

$$-\frac{1}{2}(\Pi+\lambda I) \preccurlyeq \widehat{\Pi}-\Pi \preccurlyeq \frac{1}{2}(\Pi+\lambda I).$$

Then

$$\frac{1}{2}(\Pi+\lambda I) \preccurlyeq \widehat{\Pi}+\lambda I \preccurlyeq \frac{3}{2}(\Pi+\lambda I).$$

Recalling that $\Pi + \lambda I$ is positive definite, the desired result follows.

□

PROOF OF THEOREM 3.1. In this proof, we aim to bound

$$\mathcal{E}_{\beta_0}(\hat{\beta}) = \|L_C^{1/2}(\hat{\beta}-\beta_0)\|_{\mathcal{L}^2}^2 = \|\Pi^{1/2}L_K^{-1/2}(\hat{\beta}-\beta_0)\|_{\mathcal{L}^2}^2.$$

For notational simplicity, we denote $\hat{f} = L_K^{-1/2}\hat{\beta}$ and $f_0 = L_K^{-1/2}\beta_0$. Because β_0 is not necessarily in $\mathcal{H}(K)$, f_0 is not necessary in $\mathcal{L}^2(\mathcal{T}_x, \nu_x)$. Denote $\bar{f} = (\Pi+\lambda I)^{-1}\Pi L_K^{-1/2}\beta_0$. Then

$$\mathcal{E}_{\beta_0}(\hat{\beta}) \lesssim \|\Pi^{1/2}(\hat{f}-\bar{f})\|_{\mathcal{L}^2}^2 + \|\Pi^{1/2}(\bar{f}-f_0)\|_{\mathcal{L}^2}^2. \tag{8}$$

Now we mainly consider the first term (8). Rearranging the formula of the estimator, we have

$$\hat{\beta} = (\widehat{\Gamma}+\lambda I)^{-1}\frac{1}{n}\sum_{i=1}^{n} Y_i \frac{1}{m_x}\sum_{j=1}^{m_x} X_{ij}K(T_{ij},\cdot),$$
$$= (\widehat{\Gamma}+\lambda I)^{-1}\frac{1}{n}\sum_{i=1}^{n}\int_{\mathcal{T}_x} X_i(t)\beta_0(t)\,d\nu_x(t)\frac{1}{m_x}\sum_{j=1}^{m_x} X_{ij}K(T_{ij},\cdot),$$
$$+ (\widehat{\Gamma}+\lambda I)^{-1}\frac{1}{n}\sum_{i=1}^{n} e_i \frac{1}{m_x}\sum_{j=1}^{m_x} X_{ij}K(T_{ij},\cdot).$$

Denote $\widetilde{\Pi}$ that

$$\widetilde{\Pi} f = \frac{1}{n}\sum_{i=1}^{n}\int_{\mathcal{T}_x} X_i(t)(L_K^{1/2} f)(t)\, d\nu_x(t)\frac{1}{m_x}\sum_{j=1}^{m_x} X_{ij} L_K^{-1/2} K(T_{ij},\cdot),$$

for f that $(L_K^{1/2} f) \in \mathcal{L}^2(\mathcal{T}_x, \nu_x)$, it easy to verify that $\mathbb{E}[\widetilde{\Pi} f] = \Pi f$. Additionally, we also denote that

$$g_{nm} = \frac{1}{n}\sum_{i=1}^{n} e_i \frac{1}{m_x}\sum_{j=1}^{m_x} X_{ij} L_K^{-1/2} K(T_{ij},\cdot).$$

Now, we have

$$\hat{f} = (\widehat{\Pi} + \lambda I)^{-1}\widetilde{\Pi} f_0 + (\widehat{\Pi} + \lambda I)^{-1} g_{nm}.$$

Then

$$\begin{aligned}(\widehat{\Pi} + \lambda I)(\hat{f} - \bar{f}) &= \widetilde{\Pi} f_0 + g_{nm} - (\widehat{\Pi} + \lambda I)\bar{f} \\ &= \widetilde{\Pi} f_0 - \Pi f_0 + (\Pi + \lambda I)\bar{f} - (\widehat{\Pi} + \lambda I)\bar{f} + g_{nm} \\ &= (\widetilde{\Pi} - \Pi) f_0 + (\Pi - \widehat{\Pi})\bar{f} + g_{nm}.\end{aligned}$$

Therefore

$$\begin{aligned}\hat{f} - \bar{f} &= \hat{f} - \bar{f} - (\Pi + \lambda I)^{-1}(\widehat{\Pi} + \lambda I)(\hat{f} - \bar{f}) + (\Pi + \lambda I)^{-1}(\widehat{\Pi} + \lambda I)(\hat{f} - \bar{f}) \\ &= (\Pi + \lambda I)^{-1}(\Pi - \widehat{\Pi})(\hat{f} - \bar{f}) + (\Pi + \lambda I)^{-1}(\widetilde{\Pi} - \Pi) f_0 \\ &\quad + (\Pi + \lambda I)^{-1}(\Pi - \widehat{\Pi})\bar{f} + (\Pi + \lambda I)^{-1} g_{nm}.\end{aligned} \tag{9}$$

Choose ν such that $\max\{0, \alpha/2\} < \nu < (2r - 2s - 1)/(4r)$. This interval is nonempty because $r - s \geq r_b > 1/2$ and $\alpha < (2r - 2s - 1)/(2r)$. This choice implies $0 \leq \nu$, $2\nu > \alpha$, $4\nu r - 2r + 2s < -1$, and $-4\nu r - 2r + 2s < -1$. It immediately implies

$$\begin{aligned}\|\Pi^{1/2}(\hat{f} - \bar{f})\|_{\mathcal{L}^2}^2 \lesssim &\|\Pi^{1/2}(\Pi + \lambda I)^{-1}(\Pi - \widehat{\Pi})\Pi^{-\nu}\|_{op}^2 \|\Pi^{\nu}(\hat{f} - \bar{f})\|_{\mathcal{L}^2}^2 \\ &+ \|\Pi^{1/2}(\Pi + \lambda I)^{-1}(\widetilde{\Pi} - \Pi)\Pi^{-\alpha/2}\|_{op}^2 \|\Pi^{\alpha/2} f_0\|_{\mathcal{L}^2}^2 \\ &+ \|\Pi^{1/2}(\Pi + \lambda I)^{-1}(\widehat{\Pi} - \Pi)\Pi^{-\alpha/2}\|_{op}^2 \|\Pi^{\alpha/2} \bar{f}\|_{\mathcal{L}^2}^2 \\ &+ \|\Pi^{1/2}(\Pi + \lambda I)^{-1} g_{nm}\|_{\mathcal{L}^2}^2.\end{aligned} \tag{10}$$

The first term on the right-hand side of (10) satisfies

$$\begin{aligned}&\|\Pi^{1/2}(\Pi + \lambda I)^{-1}(\Pi - \widehat{\Pi})\Pi^{-\nu}\|_{op}^2 \|\Pi^{\nu}(\hat{f} - \bar{f})\|_{\mathcal{L}^2}^2 \\ &= O_p\left(n^{-1}\lambda^{-1/(2r)} + (nm_x)^{-1}\lambda^{-(2s+1)/(2r)}\right) o_p(1) \\ &= o_p\left(n^{-1}\lambda^{-1/(2r)} + (nm_x)^{-1}\lambda^{-(2s+1)/(2r)}\right).\end{aligned}$$

Here the first factor is controlled by Lemma S.2.2, while Lemma S.2.4 gives

$$\|\Pi^{\nu}(\hat{f} - \bar{f})\|_{\mathcal{L}^2}^2 = O_p\left(n^{-1}\lambda^{-1+2\nu-1/(2r)} + (nm_x)^{-1}\lambda^{-1+2\nu-(2s+1)/(2r)}\right) = o_p(1),$$

where the last relation follows from the choice of λ and $2\nu > \alpha$. The second term and the third term can be bounded as

$$\begin{aligned}&\|\Pi^{1/2}(\Pi + \lambda I)^{-1}(\widetilde{\Pi} - \Pi)\Pi^{-\alpha/2}\|_{op}^2 \|\Pi^{\alpha/2} f_0\|_{\mathcal{L}^2}^2 \\ &= O_p\left(n^{-1}\lambda^{-1/2r} + (nm_x)^{-1}\lambda^{-(2s+1)/2r}\right),\end{aligned}$$

and

$$
\begin{aligned}
&\|\Pi^{1/2}(\Pi+\lambda I)^{-1}(\widehat{\Pi}-\Pi)\Pi^{-\alpha/2}\|_{op}^2\|\Pi^{\alpha/2}\bar{f}\|_{\mathcal{L}^2}^2 \\
&\lesssim \|\Pi^{1/2}(\Pi+\lambda I)^{-1}(\widehat{\Pi}-\Pi)\Pi^{-\alpha/2}\|_{op}^2\left(\|\Pi^{\alpha/2}(f_0-\bar{f})\|_{\mathcal{L}^2}^2+\|\Pi^{\alpha/2}f_0\|_{\mathcal{L}^2}^2\right) \\
&= O_p\left(n^{-1}\lambda^{-1/2r}+(nm_x)^{-1}\lambda^{-(2s+1)/2r}\right),
\end{aligned}
$$

where the last inequality is given by Lemma S.2.2 and Lemma S.2.1. Further, Lemma S.2.3 yields

$$
\|\Pi^{1/2}(\Pi+\lambda I)^{-1}g_{nm}\|_{\mathcal{L}^2}^2 = O_p\left(n^{-1}\lambda^{-1/2r}+(nm_x)^{-1}\lambda^{-(2s+1)/2r}\right).
$$

Plugging the above several results in (10), we have

$$
\|\Pi^{1/2}(\hat{f}-\bar{f})\|_{\mathcal{L}^2}^2 = O_p\left(n^{-1}\lambda^{-1/2r}+(nm_x)^{-1}\lambda^{-(2s+1)/2r}\right). \tag{11}
$$

Combining (8), (11) and Lemma S.2.1, we have

$$
\mathcal{E}_{\beta_0}(\hat{\beta}) = O_p\left(\lambda^{1-\alpha}+n^{-1}\lambda^{-1/2r}+(nm_x)^{-1}\lambda^{-(2s+1)/2r}\right).
$$

Then the desired result follows. □

PROOF OF THEOREM 3.2. In this part, we prove the lower bound. We only need to construct a specific case satisfying this lower bound rate. By Assumption 2(a), the Mercer decomposition of K is

$$
K(s,t)=\sum_{i=1}^{\infty}\lambda_i\phi_i^K(s)\phi_i^K(t),
$$

with $\lambda_i \asymp i^{-2r_b}$ and $r_b \le r-s$. Denote $[\cdot]$ be the rounding function. In this specific case, we set

$$
\varphi_k = \phi^K_{[k^{(r-s)/r_b}]}(t), \tag{12}
$$

and

$$
C(t_1,t_2)=\sum_{k=1}^{\infty}k^{-2s}\varphi_k(t_1)\varphi_k(t_2).
$$

Because $(r-s)/r_b \ge 1$, it ensures that $\varphi_{k_1}\neq\varphi_{k_2}$ if $k_1\neq k_2$. Then

$$
\Pi = L_{K^{1/2}CK^{1/2}} = \sum_{k=1}^{\infty}k^{-2s}\lambda_{[k^{(r-s)/r_b}]}\varphi_k\otimes\varphi_k := \sum_{k=1}^{\infty}\gamma_k\varphi_k\otimes\varphi_k.
$$

It is easy to verify that the eigenvalues $\{\gamma_k\}_{k=1}^{\infty}$ of Π satisfy $\gamma_k \asymp k^{-2r}$ and the eigenfunctions $\{\varphi_k\}_{k=1}^{\infty}$ satisfy $\|L_K^{1/2}\varphi_k\|_{\mathcal{L}^2}^2 \asymp k^{-2(r-s)}$.

Step 1. To illustrate the first term in the conclusion, we assume that $X_{ij}=X_i(T_{ij})$ is observed without noise. Theorem 1 in Cai and Yuan (2012) has proved that the convergence rate lower bound which takes infimum over all $\tilde{\beta}$ based on fully observed training data $X_i(t), Y_i$ is not faster than $n^{-2r/(2r+1)}$ when $\alpha=0$. Note that

$$
\{\tilde{\beta}: \text{estimator based on } X_{ij},T_{ij},Y_i\} \subset \{\tilde{\beta}: \text{estimator based on } X_i(t),Y_i\},
$$

when $\alpha=0$, the first term in Theorem 3.2 is deduced from Theorem 1 in Cai and Yuan (2012). When $\alpha\neq 0$, the proof is similar, thus we omit it.

Step 2. Assume that X is a Gaussian process that $X_i(t) = \sum_{k=1}^{\infty} k^{-s}\xi_{ik}\varphi_k(t)$, where ξ_{ik} are i.i.d. standard normal variables; T_{ij} are sampled from the probability measure ν_x on $\mathcal{T}_x$ and the measurement error ε_{ij}'s and e_i's are from normal distribution $N(0,\sigma^2)$. Denote $\mathcal{P}_{i,\beta_0}$ as the distribution of $(X_{i1}, X_{i2}, ..., X_{im_x}, Y_i)^T$ and $\mathcal{P}_{\beta_0}$ as the joint distribution of $\mathcal{P}_{i,\beta_0}$ for $1 \le i \le n$. Fix arbitrary i, define

$$\Sigma := \begin{pmatrix} C(T_{i\cdot}, T_{i\cdot}^T) + \sigma^2 I_{m_x} & \int C(T_{i\cdot}, t)\beta_0(t)\, d\nu_x \\ \int C(t, T_{i\cdot}^T)\beta_0(t)\, d\nu_x & \iint C(t_1,t_2)\beta_0(t_1)\beta_0(t_2)\, d\nu_x\, d\nu_x + \sigma^2 \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12}(\beta_0) \\ \Sigma_{21}(\beta_0) & \Sigma_{22}(\beta_0) \end{pmatrix}$$

where $C(T_{i\cdot}, T_{i\cdot}^T) = (C(T_{ij}, T_{il}))_{j,l} \in \mathbb{R}^{m_x \times m_x}$, $C(T_{i\cdot}, t) = (C(T_{i1}, t), C(T_{i2}, t), ..., C(T_{im_x}, t))^T$ and $C(t, T_{i\cdot}^T) = C(T_{i\cdot}, t)^T$. Conditional on the measurement points $T_{ij}'s$, the observations $(X_{i1}, X_{i2}, ..., X_{im_x}, Y_i)^T$ follows a joint normal distribution

$$(X_{i1}, X_{i2}, ..., X_{im_x}, Y_i)^T | \{T_{ij}\}_{j=1}^{m_x} \sim N(0, \Sigma),$$

By the properties of the normal distribution, conditional on $X_{ij}'s$ and $T_{ij}'s$, the variable Y_i follows a normal distribution

$$\begin{aligned} Y_i | \{T_{ij}, X_{ij}\}_{j=1}^{m_x} \sim & N\left(\Sigma_{21}(\beta_0)\Sigma_{11}^{-1}X_{i\cdot}, \Sigma_{22}(\beta_0) - \Sigma_{21}(\beta_0)\Sigma_{11}^{-1}\Sigma_{12}(\beta_0)\right) \\ & := N(\mu_{m_x}(\beta_0), \Sigma_{m_x}(\beta_0)). \end{aligned} \tag{13}$$

where $X_{i\cdot} = (X_{i1}, X_{i2}, ..., X_{im_x})^T$.

Note that the distribution in (13) depends on the true slope function β_0. We consider $\beta_0 = \beta$ and $\beta_0 = \beta'$ for different slope functions with $\|\beta\|_{\mathcal{L}^2}^2, \|\beta'\|_{\mathcal{L}^2}^2 \le C_1$ for some constant C_1. Then, conditional on $X_{ij}'s$ and $T_{ij}'s$, the Kullback–Leibler distance from $\mathcal{P}_{i,\beta}$ and $\mathcal{P}_{i,\beta'}$ is

$$\begin{aligned} \mathrm{D_{KL}}(\mathcal{P}_{i,\beta} || \mathcal{P}_{i,\beta'}) &= \int \log\left(\frac{d\mathcal{P}_{i,\beta}}{d\mathcal{P}_{i,\beta'}}\right) d\mathcal{P}_{i,\beta} \\ &= \mathbb{E}_{T,X}\left[\mathrm{D_{KL}}(\mathcal{P}_{i,\beta}(Y_i|T,X) || \mathcal{P}_{i,\beta'}(Y_i|T,X))\right] \\ &= \frac{1}{2}\mathbb{E}_{T,X}\left[\frac{(\mu_{m_x}(\beta') - \mu_{m_x}(\beta))^2}{\Sigma_{m_x}(\beta')}\right] + \frac{1}{2}\mathbb{E}_{T,X}\left[\frac{\Sigma_{m_x}(\beta)}{\Sigma_{m_x}(\beta')} - \log\frac{\Sigma_{m_x}(\beta)}{\Sigma_{m_x}(\beta')} - 1\right]. \end{aligned} \tag{14}$$

To bound the first term of (14),

$$\mathbb{E}_{T,X}\left[\frac{(\mu_{m_x}(\beta') - \mu_{m_x}(\beta))^2}{\Sigma_{m_x}(\beta')}\right] = \mathbb{E}_{T,X}\left[\frac{((\Sigma_{21}(\beta') - \Sigma_{21}(\beta))\Sigma_{11}^{-1}X_{i\cdot})^2}{\Sigma_{m_x}(\beta')}\right]$$

Note that $\Sigma_{m_x}(\beta') \ge \sigma^2$, it is less than

$$\begin{aligned} &\sigma^{-2}\mathbb{E}_{T,X}\left[((\Sigma_{21}(\beta') - \Sigma_{21}(\beta))\Sigma_{11}^{-1}X_{i\cdot})^2\right] \\ &= \sigma^{-2}\mathbb{E}_{T,X}\left[(\Sigma_{21}(\beta') - \Sigma_{21}(\beta))\Sigma_{11}^{-1}X_{i\cdot}X_{i\cdot}^T\Sigma_{11}^{-1}\left(\Sigma_{21}(\beta') - \Sigma_{21}(\beta)\right)^T\right] \\ &= \sigma^{-2}\mathbb{E}_{T}\left[(\Sigma_{21}(\beta') - \Sigma_{21}(\beta))\Sigma_{11}^{-1}\left(\Sigma_{21}(\beta') - \Sigma_{21}(\beta)\right)^T\right]. \end{aligned}$$

Recall that $\Sigma_{11} \succcurlyeq \sigma^2 I_{m_x}$, the above is also less than

$$\sigma^{-4}\mathbb{E}_T\left[\|(\Sigma_{21}(\beta') - \Sigma_{21}(\beta))\|_2^2\right] = m_x\sigma^{-4}\mathbb{E}_T\left[\left(\int C(T_{i1}, t)(\beta - \beta')(t)\, d\nu_x(t)\right)^2\right]$$

This implies

$$\frac{1}{2}\mathbb{E}_{T,X}\left[\frac{(\mu_{m_x}(\beta') - \mu_{m_x}(\beta))^2}{\Sigma_{m_x}(\beta')}\right] \lesssim m_x \|L_C(\beta - \beta')\|_{\mathcal{L}^2}^2. \tag{15}$$

Now we consider the second term of (14). Note that

$$\begin{pmatrix} C(T_{i\cdot}, T_{i\cdot}^T) + \sigma^2 I_{m_x} & \int C(T_{i\cdot}, t)\beta_0(t)\, d\nu_x(t) \\ \int C(t, T_{i\cdot}^T)\beta_0(t)\, d\nu_x(t) & \iint C(t_1, t_2)\beta_0(t_1)\beta_0(t_2)\, d\nu_x(t_1)\, d\nu_x(t_2) \end{pmatrix}$$

is positive definite and it is equal to

$$\begin{pmatrix} I_{m_x} & 0 \\ \Sigma_{21}\Sigma_{11}^{-1} & 1 \end{pmatrix} \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{m_x}(\beta_0) - \sigma^2 \end{pmatrix} \begin{pmatrix} I_{m_x} & \Sigma_{11}^{-1}\Sigma_{12} \\ 0 & 1 \end{pmatrix},$$

we have $\Sigma_{m_x}(\beta_0) - \sigma^2 \geq 0$ for $\beta_0 = \beta$ and β'. Moreover,

$$\Sigma_{m_x}(\beta_0) = \Sigma_{22}(\beta_0) - \Sigma_{21}(\beta_0)\Sigma_{11}^{-1}\Sigma_{12}(\beta_0) \leq \Sigma_{22}(\beta_0) \leq \sigma^2 + \|\beta_0\|_{\mathcal{L}^2}^2,$$

where the second applies the fact that the largest eigenvalue of the covariance function is 1. Because $\|\beta\|_{\mathcal{L}^2}^2, \|\beta'\|_{\mathcal{L}^2}^2 \leq C_1$ for some constant C_1, we know

$$\frac{\sigma^2}{\sigma^2 + C_1} \leq \frac{\Sigma_{m_x}(\beta)}{\Sigma_{m_x}(\beta')} \leq \frac{\sigma^2 + C_1}{\sigma^2},$$

i.e. $\Sigma_{m_x}(\beta')^{-1}\Sigma_{m_x}(\beta)$ is uniformly bounded away from 0 and ∞. Note that there is a constant C_2 that may depend on C_1 and σ^2 that for any $x : (\sigma^2 + C_1)^{-1}\sigma^2 \leq x \leq \sigma^{-2}(\sigma^2 + C_1)$, it holds $x - \log x - 1 \leq C_2(x-1)^2$. Therefore,

$$\begin{aligned} &\frac{\Sigma_{m_x}(\beta)}{\Sigma_{m_x}(\beta')} - \log\frac{\Sigma_{m_x}(\beta)}{\Sigma_{m_x}(\beta')} - 1 \leq C_2\left(\frac{\Sigma_{m_x}(\beta)}{\Sigma_{m_x}(\beta')} - 1\right)^2 \\ &= C_2\frac{\left(\Sigma_{m_x}(\beta) - \Sigma_{m_x}(\beta')\right)^2}{\Sigma_{m_x}(\beta')^2} \leq C_2\sigma^{-4}\left(\Sigma_{m_x}(\beta) - \Sigma_{m_x}(\beta')\right)^2. \end{aligned} \tag{16}$$

Note that

$$\begin{aligned} \left(\Sigma_{m_x}(\beta) - \Sigma_{m_x}(\beta')\right)^2 &\leq 2\left(\Sigma_{22}(\beta) - \Sigma_{22}(\beta')\right)^2 \\ &\quad + 2\left(\Sigma_{21}(\beta)\Sigma_{11}^{-1}\Sigma_{12}(\beta) - \Sigma_{21}(\beta')\Sigma_{11}^{-1}\Sigma_{12}(\beta')\right)^2. \end{aligned} \tag{17}$$

For the first term

$$\begin{aligned} \left(\Sigma_{22}(\beta) - \Sigma_{22}(\beta')\right)^2 &= \left(\langle L_C\beta, \beta\rangle_{\mathcal{L}^2} - \langle L_C\beta', \beta'\rangle_{\mathcal{L}^2}\right)^2 = \langle L_C(\beta - \beta'), \beta + \beta'\rangle_{\mathcal{L}^2}^2 \\ &\leq \|L_C(\beta - \beta')\|_{\mathcal{L}^2}^2\|\beta + \beta'\|_{\mathcal{L}^2}^2 \leq 4C_1\|L_C(\beta - \beta')\|_{\mathcal{L}^2}^2. \end{aligned} \tag{18}$$

Now, we bound the second term. Because $\Sigma_{m_x}(\beta_0) \geq \sigma^2$ for $\beta_0 = \beta$ and β', we have

$$\Sigma_{21}(\beta_0)\Sigma_{11}^{-1}\Sigma_{12}(\beta_0) \leq \Sigma_{22}(\beta_0) - \sigma^2 \leq \|\beta_0\|_{\mathcal{L}^2}^2 \leq C_1. \tag{19}$$

For the second term

$$\begin{aligned} &\left(\Sigma_{21}(\beta)\Sigma_{11}^{-1}\Sigma_{12}(\beta) - \Sigma_{21}(\beta')\Sigma_{11}^{-1}\Sigma_{12}(\beta')\right)^2 \\ &\quad = \left((\Sigma_{21}(\beta) - \Sigma_{21}(\beta'))\Sigma_{11}^{-1}\left(\Sigma_{21}(\beta) + \Sigma_{21}(\beta')\right)^T\right)^2 \\ &\quad \leq \left((\Sigma_{21}(\beta) - \Sigma_{21}(\beta'))\Sigma_{11}^{-1}\left(\Sigma_{21}(\beta) - \Sigma_{21}(\beta')\right)^T\right) \\ &\qquad \cdot \left((\Sigma_{21}(\beta) + \Sigma_{21}(\beta'))\Sigma_{11}^{-1}\left(\Sigma_{21}(\beta) + \Sigma_{21}(\beta')\right)^T\right) \\ &\quad \leq 4C_1(\Sigma_{21}(\beta) - \Sigma_{21}(\beta'))\Sigma_{11}^{-1}\left(\Sigma_{21}(\beta) - \Sigma_{21}(\beta')\right)^T \\ &\quad \leq 4C_1\sigma^{-2}\|\Sigma_{21}(\beta) - \Sigma_{21}(\beta')\|_2^2. \end{aligned} \tag{20}$$

where the first inequality is given by Cauchy inequality because Σ_{11} is positive definite, the second inequality use (19) and the last inequality is due to $\Sigma_{11} \succcurlyeq \sigma^2 I_{m_x}$. Plugging (18) and (20) in (17) and further (16), then taking expectation yields

$$\mathbb{E}\left[\frac{\Sigma_{m_x}(\beta)}{\Sigma_{m_x}(\beta')} - \log\frac{\Sigma_{m_x}(\beta)}{\Sigma_{m_x}(\beta')} - 1\right] \le C_3 m_x \|L_C(\beta-\beta')\|_{\mathcal{L}^2}^2, \tag{21}$$

where the constant C_3 depends on C_1 and σ^2. Combining (14), (15) and (21), we obtain

$$\mathrm{D_{KL}}(\mathcal{P}_{i,\beta}||\mathcal{P}_{i,\beta'}) \le C_4 m_x \|L_C(\beta-\beta')\|_{\mathcal{L}^2}^2$$

where C_4 is a constant depending on C_1 and σ^2. Recall the observations from different subjects are independent, we have

$$\mathrm{D_{KL}}(\mathcal{P}_{\beta}||\mathcal{P}_{\beta'}) \le C_4 n m_x \|L_C(\beta-\beta')\|_{\mathcal{L}^2}^2. \tag{22}$$

Step 3. Now we construct the packing net. Set $M = [C_5(nm_x)^{1/(2(1-\alpha)r+2s+1)}]$ with a large constant C_5 determined later. Let $\boldsymbol{a} = (a_1,\dots,a_M) \in \{0,1\}^M$, and define

$$\beta_{\boldsymbol{a}} = \sum_{k=M+1}^{2M} a_{k-M} M^{-1/2} k^{(\alpha-1)r+s}\varphi_k.$$

First, we verify that $\|\Pi^{\alpha/2}L_K^{-1/2}\beta_{\boldsymbol{a}}\|_{\mathcal{L}^2} \lesssim 1$. In fact,

$$\begin{aligned}\|\Pi^{\alpha/2}L_K^{-1/2}\beta_{\boldsymbol{a}}\|_{\mathcal{L}^2}^2 &= \left\|\sum_{k=M+1}^{2M} a_{k-M}M^{-1/2}\Pi^{\alpha/2}L_K^{-1/2}k^{(\alpha-1)r+s}\varphi_k\right\|_{\mathcal{L}^2}^2\\ &= \sum_{k=M+1}^{2M} a_{k-M}^2 M^{-1}k^{2(\alpha-1)r+2s}\left\|\Pi^{\alpha/2}L_K^{-1/2}\varphi_k\right\|_{\mathcal{L}^2}^2 \lesssim 1.\end{aligned}$$

Additionally, we have

$$\begin{aligned}\|\beta_{\boldsymbol{a}}\|_{\mathcal{L}^2}^2 &= \left\|\sum_{k=M+1}^{2M} a_{k-M}M^{-1/2}k^{(\alpha-1)r+s}\varphi_k\right\|_{\mathcal{L}^2}^2\\ &= \sum_{k=M+1}^{2M} a_{k-M}^2 M^{-1}k^{2(\alpha-1)r+2s} \lesssim 1.\end{aligned} \tag{23}$$

By applying the Varshamov-Gilbert lemma, there exists a subset $\mathcal{A} \subseteq \{0,1\}^M$ such that: $|\mathcal{A}| \ge \exp(M/8)$, and for any distinct vectors $\boldsymbol{a},\boldsymbol{a}' \in \mathcal{A}$, the Hamming distance satisfies $\mathrm{Ham}(\boldsymbol{a},\boldsymbol{a}') \ge M/8$. Recalling (22) and (23),

$$\begin{aligned}\mathrm{D_{KL}}(\mathcal{P}_{\beta_{\boldsymbol{a}}}||\mathcal{P}_{\beta_{\boldsymbol{a}'}}) &\lesssim nm_x\|L_C(\beta_{\boldsymbol{a}}-\beta_{\boldsymbol{a}'})\|_{\mathcal{L}^2}^2\\ &= nm_x\left\|\sum_{k=M+1}^{2M}(a_{k-M}-a'_{k-M})M^{-1/2}k^{(\alpha-1)r-s}\varphi_k\right\|_{\mathcal{L}^2}^2\\ &= nm_x\sum_{k=M+1}^{2M}(a_{k-M}-a'_{k-M})^2M^{-1}k^{2(\alpha-1)r-2s}\\ &\lesssim nm_x M^{2(\alpha-1)r-2s}.\end{aligned}$$

Take any $\boldsymbol{a}_0 \in \mathcal{A}$, we have

$$\frac{1}{|\mathcal{A}|}\sum_{\boldsymbol{a}\in\mathcal{A}} \mathrm{D_{KL}}(\mathcal{P}_{\beta_{\boldsymbol{a}}}||\mathcal{P}_{\beta_{\boldsymbol{a}_0}}) \leq C_6 n m_x M^{2(\alpha-1)r-2s} \leq bM/8$$

for any $b \in (0, 1/8)$, where the last inequality is given by taking the constant C_5 large enough. In addition, for any $\boldsymbol{a}, \boldsymbol{a}' \in \mathcal{A}$,

$$\begin{aligned}
\|L_C^{1/2}(\beta_{\boldsymbol{a}} - \beta_{\boldsymbol{a}'})\|_{\mathcal{L}^2}^2 &= \left\| \sum_{k=M+1}^{2M} (a_{k-M} - a'_{k-M}) M^{-1/2} k^{(\alpha-1)r} \varphi_k \right\|_{\mathcal{L}^2}^2 \\
&= \sum_{k=M+1}^{2M} (a_{k-M} - a'_{k-M})^2 M^{-1} k^{2(\alpha-1)r} \\
&\geq M^{-1}(2M)^{2(\alpha-1)r}\,\mathrm{Ham}(\boldsymbol{a}, \boldsymbol{a}') \\
&\geq 4C_6 (n m_x)^{-2(1-\alpha)r/(2(1-\alpha)r+2s+1)},
\end{aligned} \tag{24}$$

for some constant C_6 independent of n. An application of Theorem 2.5 in Tsybakov and Zaiats (2009) now implies

$$\begin{aligned}
&\inf_{\tilde{\beta}} \sup_{\mathcal{P}_{Z,\epsilon} \in \mathcal{Q}_{Z,\epsilon}} \mathbb{P}\left(\|L_C^{1/2}(\tilde{\beta} - \beta_0)\|_{\mathcal{L}^2}^2 > C_6 (n m_x)^{-2(1-\alpha)r/(2(1-\alpha)r+2s+1)} \right) \\
&\geq \frac{\sqrt{|\mathcal{A}|}}{1+\sqrt{|\mathcal{A}|}} \left(1 - 2b - \sqrt{\frac{2b}{\log|\mathcal{A}|}} \right) \to 1 - 2b.
\end{aligned} \tag{25}$$

This completes the proof. □

PROOF OF REMARK 3.1. We only need to verify the result on $\|\hat{\beta} - \beta_0\|_{\mathcal{L}^2}^2$ from the upper and lower bounds, which suggests modifying the proof of Theorem 3.1 and Theorem 3.2. In the modification, the fact that $r = r_b + r_c$, $s = r_c$ and $\alpha = 0$ should be noted.

We first check the upper bound. Because that the eigenfunctions of K and C are aligned and the eigendecay are k^{-2r_b} and k^{-2r_c}, respectively, we have

$$L_K \asymp \Pi^{\frac{r_b}{r_b+r_c}} \text{ and } L_C \asymp \Pi^{\frac{r_c}{r_b+r_c}}.$$

Therefore $\|\hat{\beta} - \beta_0\|_{\mathcal{L}^2}^2 \lesssim \|\Pi^{\frac{r_b}{2(r_b+r_c)}}(\hat{f} - \bar{f})\|_{\mathcal{L}^2}^2 + \|\Pi^{\frac{r_b}{2(r_b+r_c)}}(\bar{f} - f_0)\|_{\mathcal{L}^2}^2$. Taking a $\nu > 0$ satisfying $4\nu(r_b + r_c) - 2r_b < -1$, we have

$$\begin{aligned}
\|\Pi^{\frac{r_b}{2(r_b+r_c)}}(\hat{f} - \bar{f})\|_{\mathcal{L}^2}^2 \lesssim &\|\Pi^{\frac{r_b}{2(r_b+r_c)}}(\Pi + \lambda I)^{-1}(\Pi - \widehat{\Pi})\Pi^{-\nu}\|_{op}^2 \|\Pi^{\nu}(\hat{f} - \bar{f})\|_{\mathcal{L}^2}^2 \\
&+ \|\Pi^{\frac{r_b}{2(r_b+r_c)}}(\Pi + \lambda I)^{-1}(\widetilde{\Pi} - \Pi)\|_{op}^2 \|f_0\|_{\mathcal{L}^2}^2 \\
&+ \|\Pi^{\frac{r_b}{2(r_b+r_c)}}(\Pi + \lambda I)^{-1}(\widehat{\Pi} - \Pi)\|_{op}^2 \|\bar{f}\|_{\mathcal{L}^2}^2 \\
&+ \|\Pi^{\frac{r_b}{2(r_b+r_c)}}(\Pi + \lambda I)^{-1} g_{nm}\|_{\mathcal{L}^2}^2.
\end{aligned}$$

Then a similar analysis with (10) implies the desired upper bound.

For the lower bound, we also need to modify the proof of Theorem 3.2. It suffices to replace (24) with

$$\begin{aligned}\|\beta_{\boldsymbol{a}}-\beta_{\boldsymbol{a}'}\|_{\mathcal{L}^2}^2 &= \left\|\sum_{k=M+1}^{2M}(a_{k-M}-a'_{k-M})M^{-1/2}k^{-r_b}\varphi_k\right\|_{\mathcal{L}^2}^2 \\ &= \sum_{k=M+1}^{2M}(a_{k-M}-a'_{k-M})^2M^{-1}k^{-2r_b} \\ &\geq M^{-1}(2M)^{-2r_b}\,\mathrm{Ham}(\boldsymbol{a},\boldsymbol{a}') \\ &\geq 4C_6'(nm_x)^{-2r_b/(2r_b+4r_c+1)},\end{aligned}$$

for some constant C_6' independent of n. Then the remaining proof is all the same. $\square$

PROOF OF PROPOSITION 3.2. This part is slightly more complex than the proof of Proposition 3.1. We only need to show that

$$(\widehat{\Pi}+\lambda I):\mathcal{L}^2(\mathcal{T}_y\times\mathcal{T}_x,\nu_y\otimes\nu_x)\to\mathcal{L}^2(\mathcal{T}_y\times\mathcal{T}_x,\nu_y\otimes\nu_x)$$

is positive definite for λ that

$$\lambda^{-1}=o\left(n^{r_x}\wedge(nm_x^2)^{\frac{r_x}{2s+1}}\wedge(nm_y\log^{-2}n)^{r_y}\wedge(nm_y\log^{-2}n)^{r_x}\wedge(nm_x^2m_y\log^{-2}n)^{\frac{r_x}{2s+1}}\right).$$

Denote $\widehat{\Pi}_{x,i}:\mathcal{L}^2(\mathcal{T}_x,\nu_x)\to\mathcal{L}^2(\mathcal{T}_x,\nu_x)$ and $\widehat{\Pi}_{y,i}:\mathcal{L}^2(\mathcal{T}_y,\nu_y)\to\mathcal{L}^2(\mathcal{T}_y,\nu_y)$ such that

$$\widehat{\Pi}_{x,i}f=\frac{1}{m_x(m_x-1)}\sum_{j_1\neq j_2}X_{ij_1}X_{ij_2}(L_{K^x}^{1/2}f)(T_{ij_1})(K^x)^{1/2}(T_{ij_2},\cdot)$$

and

$$\widehat{\Pi}_{y,i}g=\frac{1}{m_y}\sum_{l=1}^{m_y}(L_{K^y}^{1/2}g)(S_{il})(K^y)^{1/2}(S_{il},\cdot),$$

where $K^{1/2}$ is the function that $L_{K^{1/2}}=L_K^{1/2}$. Therefore,

$$\begin{aligned}\widehat{\Pi}+\lambda I &= \frac{1}{n}\sum_{i=1}^{n}\widehat{\Pi}_{y,i}\otimes\widehat{\Pi}_{x,i}+\lambda I \\ &= \frac{1}{n}\sum_{i=1}^{n}\widehat{\Pi}_{y,i}\otimes(\widehat{\Pi}_{x,i}-\Pi_x)+(\frac{1}{n}\sum_{i=1}^{n}\widehat{\Pi}_{y,i})\otimes\Pi_x+\lambda I \\ &= \frac{1}{n}\sum_{i=1}^{n}(\widehat{\Pi}_{y,i}-\Pi_y)\otimes(\widehat{\Pi}_{x,i}-\Pi_x)+\Pi_y\otimes\frac{1}{n}\sum_{i=1}^{n}(\widehat{\Pi}_{x,i}-\Pi_x) \\ &\quad+\frac{1}{n}\sum_{i=1}^{n}(\widehat{\Pi}_{y,i}-\Pi_y)\otimes\Pi_x+\Pi+\lambda I.\end{aligned}\tag{26}$$

For the first term in the last line of (26), denote $\Delta\Pi:=\frac{1}{n}\sum_{i=1}^n\Delta\Pi_i:=\frac{1}{n}\sum_{i=1}^n(\widehat{\Pi}_{y,i}-\Pi_y)\otimes(\widehat{\Pi}_{x,i}-\Pi_x)$. We first analyze $(\Pi+\lambda I)^{-1/2}(\Delta\Pi)(\Pi+\lambda I)^{-1/2}$. Its operator norm is expressed as

$$\begin{aligned}&\|(\Pi+\lambda I)^{-1/2}(\Delta\Pi)(\Pi+\lambda I)^{-1/2}\|_{op} \\ &=\sup_{h_1,h_2:\|h_1\|_{\mathcal{L}^2}=\|h_2\|_{\mathcal{L}^2}=1}\langle h_1,(\Pi+\lambda I)^{-1/2}(\Delta\Pi)(\Pi+\lambda I)^{-1/2}h_2\rangle_{\mathcal{L}^2},\end{aligned}$$

where the supremum is taken over all h_1 and $h_2 \in \mathcal{L}^2(\mathcal{T}_y \times \mathcal{T}_x, \nu_y \otimes \nu_x)$ with $\|h_1\|_{\mathcal{L}^2} = \|h_2\|_{\mathcal{L}^2} = 1$. Assume that

$$h_1 = \sum_{u=1}^{\infty}\sum_{v=1}^{\infty} a_{uv}\varphi_{uv} \quad \text{and} \quad h_2 = \sum_{u=1}^{\infty}\sum_{v=1}^{\infty} b_{uv}\varphi_{uv}.$$

with $\sum_{u=1}^{\infty}\sum_{v=1}^{\infty} a_{uv}^2 = \sum_{u=1}^{\infty}\sum_{v=1}^{\infty} b_{uv}^2 = 1$. Then the operator norm can become

$$\begin{aligned}
&\sup_{(a_{uv},b_{uv})} \left\langle \sum_{u=1}^{\infty}\sum_{v=1}^{\infty} a_{uv}\varphi_{uv}, (\Pi+\lambda I)^{-1/2}(\Delta\Pi)(\Pi+\lambda I)^{-1/2}\sum_{u=1}^{\infty}\sum_{v=1}^{\infty} b_{uv}\varphi_{uv} \right\rangle_{\mathcal{L}^2} \\
&= \sup_{(a_{uv},b_{uv})} \sum_{u_1,v_1}\sum_{u_2,v_2} a_{u_1v_1} b_{u_2v_2} (\gamma_{u_1v_1}+\lambda)^{-1/2}(\gamma_{u_2v_2}+\lambda)^{-1/2} \langle \varphi_{u_1v_1}, (\Delta\Pi)\varphi_{u_2v_2}\rangle_{\mathcal{L}^2} \\
&\le \sup_{(a_{uv},b_{uv})} \left(\sum_{u_1,v_1}\sum_{u_2,v_2} a_{u_1v_1}^2 b_{u_2v_2}^2 \right)^{1/2} \left(\sum_{u_1,v_1}\sum_{u_2,v_2} \frac{\langle \varphi_{u_1v_1}, (\Delta\Pi)\varphi_{u_2v_2}\rangle_{\mathcal{L}^2}^2}{(\gamma_{u_1v_1}+\lambda)(\gamma_{u_2v_2}+\lambda)} \right)^{1/2} \\
&= \left(\sum_{u_1,v_1}\sum_{u_2,v_2} (\gamma_{u_1v_1}+\lambda)^{-1}(\gamma_{u_2v_2}+\lambda)^{-1} \langle \varphi_{u_1v_1}, (\Delta\Pi)\varphi_{u_2v_2}\rangle_{\mathcal{L}^2}^2 \right)^{1/2},
\end{aligned}$$

where the supremum is taken over all (a_{uv}, b_{uv}) with $\sum_{u=1}^{\infty}\sum_{v=1}^{\infty} a_{uv}^2 = \sum_{u=1}^{\infty}\sum_{v=1}^{\infty} b_{uv}^2 = 1$ and the inequality is implied by Cauchy inequality. Recalling the definition of operator $\Delta\Pi$, we have

$$\begin{aligned}
&\mathbb{E}\left[\langle \varphi_{u_1v_1}, (\Delta\Pi)\varphi_{u_2v_2}\rangle_{\mathcal{L}^2}^2\right] = \mathrm{Var}\left(\langle \varphi_{u_1v_1}, (\Delta\Pi)\varphi_{u_2v_2}\rangle_{\mathcal{L}^2}\right) \\
&= \frac{1}{n}\mathrm{Var}\left(\langle \varphi_{u_1v_1}, (\Delta\Pi_i)\varphi_{u_2v_2}\rangle_{\mathcal{L}^2}\right) \lesssim \frac{1}{n}\mathbb{E}\left[\langle \varphi_{u_1v_1}, (\Delta\Pi_i)\varphi_{u_2v_2}\rangle_{\mathcal{L}^2}^2\right] \\
&= \frac{1}{n}\mathbb{E}\left[\left\langle \varphi_{u_1}^x, (\widehat{\Pi}_{x,i}-\Pi_x)\varphi_{u_2}^x \right\rangle_{\mathcal{L}^2}^2\right] \mathbb{E}\left[\left\langle \varphi_{v_1}^y, (\widehat{\Pi}_{y,i}-\Pi_y)\varphi_{v_2}^y \right\rangle_{\mathcal{L}^2}^2\right],
\end{aligned} \tag{27}$$

where the last line uses the independence between S_{il}'s and T_{ij}'s, X_{ij}'s. Similar to the derivation of (6) in the proof of Proposition 3.1, with the measurement-error moments again absorbed only into the implicit constants,

$$\begin{aligned}
&\mathbb{E}\left[\left\langle \varphi_{u_1}^x, (\widehat{\Pi}_{x,i}-\Pi_x)\varphi_{u_2}^x \right\rangle_{\mathcal{L}^2}^2\right] \le \mathbb{E}\left[\left\langle \varphi_{u_1}^x, \widehat{\Pi}_{x,i}\varphi_{u_2}^x \right\rangle_{\mathcal{L}^2}^2\right] \\
&\lesssim u_1^{-2r_x}u_2^{-2r_x} + \frac{1}{m_x}(u_1^{2s-2r_x}u_2^{-2r_x} + u_1^{-2r_x}u_2^{2s-2r_x}) + \frac{1}{m_x^2}u_1^{2s-2r_x}u_2^{2s-2r_x}.
\end{aligned} \tag{28}$$

Additionally, when $v_1 \ne v_2$, we have

$$\begin{aligned}
&\mathbb{E}\left[\left\langle \varphi_{v_1}^y, (\widehat{\Pi}_{y,i}-\Pi_y)\varphi_{v_2}^y \right\rangle_{\mathcal{L}^2}^2\right] \le \mathbb{E}\left[\left\langle \varphi_{v_1}^y, \widehat{\Pi}_{y,i}\varphi_{v_2}^y \right\rangle_{\mathcal{L}^2}^2\right] \\
&= \frac{1}{m_y^2}\mathbb{E}\left[\left(\sum_{l=1}^{m_y} L_{K^y}^{1/2}\varphi_{v_1}^y(S_{il}) L_{K^y}^{1/2}\varphi_{v_2}^y(S_{il}) \right)^2\right] \\
&= \frac{1}{m_y^2}\gamma_{v_1}^y\gamma_{v_2}^y\mathbb{E}\left[\left(\sum_{l=1}^{m_y} \varphi_{v_1}^y(S_{il})\varphi_{v_2}^y(S_{il}) \right)^2\right] \lesssim \frac{1}{m_y}v_1^{-2r_y}v_2^{-2r_y}.
\end{aligned} \tag{29}$$

When $v_1 = v_2$, under more careful calculations

$$
\begin{aligned}
&\mathbb{E}\left[\left\langle \varphi^y_{v_1}, (\widehat{\Pi}_{y,i} - \Pi_y)\varphi^y_{v_2}\right\rangle^2_{\mathcal{L}^2}\right] \\
&= \frac{1}{m_y^2}\mathbb{E}\left[\left(\sum_{l=1}^{m_y} L^{1/2}_{K^y}\varphi^y_{v_1}(S_{il})L^{1/2}_{K^y}\varphi^y_{v_2}(S_{il})\right)^2\right] - \left(\mathbb{E}\left[L^{1/2}_{K^y}\varphi^y_{v_1}(S_{il})L^{1/2}_{K^y}\varphi^y_{v_2}(S_{il})\right]\right)^2 \\
&= \frac{m_y(m_y-1)(\gamma^y_{v_1})^2}{m_y^2}\|\varphi^y_{v_1}\|^4_{\mathcal{L}^2} + \frac{m_y(\gamma^y_{v_1})^2}{m_y^2}\|\varphi^y_{v_1}\|^4_{\mathcal{L}^4} - (\gamma^y_{v_1})^2\|\varphi^y_{v_1}\|^4_{\mathcal{L}^2} \\
&\lesssim \frac{1}{m_y} v_1^{-2r_y} v_2^{-2r_y},
\end{aligned}
\tag{30}
$$

due to the condition $\|\varphi^y_{v_1}\|_{\mathcal{L}^4} \lesssim 1$ uniformly. Plugging (28) and (29),(30) into (27) yields

$$
\begin{aligned}
&\mathbb{E}\left[\langle \varphi_{u_1v_1}, (\Delta\Pi)\varphi_{u_2v_2}\rangle^2_{\mathcal{L}^2}\right] \\
&\qquad \lesssim \frac{v_1^{-2r_y}v_2^{-2r_y}}{nm_y}\left(u_1^{-2r_x} + \frac{u_1^{2s-2r_x}}{m_x}\right)\left(u_2^{-2r_x} + \frac{u_2^{2s-2r_x}}{m_x}\right).
\end{aligned}
\tag{31}
$$

Therefore,

$$
\begin{aligned}
&\sum_{u_1,v_1}\sum_{u_2,v_2}(\gamma_{u_1v_1}+\lambda)^{-1}(\gamma_{u_2v_2}+\lambda)^{-1}\mathbb{E}\left[\langle \varphi_{u_1v_1}, (\Delta\Pi)\varphi_{u_2v_2}\rangle^2_{\mathcal{L}^2}\right] \\
&\lesssim \frac{1}{nm_y}\left(\sum_{u,v}\frac{m_x u^{-2r_x} + u^{2s-2r_x}}{m_x(u^{-2r_x}v^{-2r_y}+\lambda)v^{2r_y}}\right)^2.
\end{aligned}
\tag{32}
$$

For $x \in [u, u+1]$ and $y \in [v, v+1]$, each summand below is bounded by a constant multiple of the corresponding integrand, uniformly in $\lambda \in (0,1)$. Hence,

$$
\begin{aligned}
\sum_{u,v}\frac{u^{-2r_x}}{(u^{-2r_x}v^{-2r_y}+\lambda)v^{2r_y}} &= \sum_{u,v}\frac{1}{1+\lambda u^{2r_x}v^{2r_y}} \lesssim \int_1^\infty\int_1^\infty \frac{1}{1+\lambda x^{2r_x}y^{2r_y}}dxdy \\
&\lesssim \begin{cases} \lambda^{-1/2r_y} & \text{if } r_x > r_y, \\ \lambda^{-1/2r_y}\log(1/\lambda) & \text{if } r_x = r_y, \\ \lambda^{-1/2r_x} & \text{if } r_x < r_y, \end{cases}
\end{aligned}
$$

and

$$
\sum_{u,v}\frac{u^{2s-2r_x}}{(u^{-2r_x}v^{-2r_y}+\lambda)v^{2r_y}} = \sum_{u,v}\frac{u^{2s}}{1+\lambda u^{2r_x}v^{2r_y}} \lesssim \int_1^\infty\int_1^\infty \frac{x^{2s}}{1+\lambda x^{2r_x}y^{2r_y}}dxdy
$$

$$
\le \int_1^\infty\int_1^\infty \frac{1}{1+\lambda z^{2r_x/(2s+1)}y^{2r_y}}dzdy \lesssim \begin{cases} \lambda^{-1/2r_y} & \text{if } r_x > r_y(2s+1), \\ \lambda^{-1/2r_y}\log(1/\lambda) & \text{if } r_x = r_y(2s+1), \\ \lambda^{-(2s+1)/2r_x} & \text{if } r_x < r_y(2s+1), \end{cases}
$$

where the last inequality applies Lemma S.2.9. We have

$$\begin{aligned}\frac{1}{nm_y}&\left(\sum_{u,v}\frac{m_x u^{-2r_x}+u^{2s-2r_x}}{m_x(u^{-2r_x}v^{-2r_y}+\lambda)v^{2r_y}}\right)^2\\&\lesssim\left(\frac{\lambda^{-1/(r_x\wedge r_y)}}{nm_y}+\frac{\lambda^{-1/((r_x/(2s+1))\wedge r_y)}}{nm_x^2 m_y}\right)\log^2(n)\\&=o(1).\end{aligned}$$

Plugging the above conclusion into (32), we have

$$\sum_{u_1,v_1}\sum_{u_2,v_2}(\gamma_{u_1v_1}+\lambda)^{-1}(\gamma_{u_2v_2}+\lambda)^{-1}\mathbb{E}\left[\langle\varphi_{u_1v_1},(\Delta\Pi)\varphi_{u_2v_2}\rangle^2_{\mathcal{L}^2}\right]=o(1).$$

Therefore,

$$\|(\Pi+\lambda I)^{-1/2}(\Delta\Pi)(\Pi+\lambda I)^{-1/2}\|_{op}=o_p(1).$$

This suggests that

$$-o_p(1)I\preccurlyeq(\Pi+\lambda I)^{-1/2}(\Delta\Pi)(\Pi+\lambda I)^{-1/2}\preccurlyeq o_p(1)I,$$

where we denote that $o_p(1)$ is nonnegative. Moreover, we have:

$$-o_p(1)(\Pi+\lambda I)\preccurlyeq\Delta\Pi\preccurlyeq o_p(1)(\Pi+\lambda I). \tag{33}$$

Now we consider the second term in the last line of (26). Similar with (7), we can obtain

$$\|(\Pi_x+\lambda I)^{-1/2}\frac{1}{n}\sum_{i=1}^n(\widehat{\Pi}_{x,i}-\Pi_x)(\Pi_x+\lambda I)^{-1/2}\|_{op}=o_p(1).$$

This implies

$$-o_p(1)(\Pi_x+\lambda I)\preccurlyeq\frac{1}{n}\sum_{i=1}^n(\widehat{\Pi}_{x,i}-\Pi_x)\preccurlyeq o_p(1)(\Pi_x+\lambda I),$$

thus further

$$-o_p(1)(\Pi+\lambda I)\preccurlyeq\Pi_y\otimes\frac{1}{n}\sum_{i=1}^n(\widehat{\Pi}_{x,i}-\Pi_x)\preccurlyeq o_p(1)(\Pi+\lambda I), \tag{34}$$

where we use the fact that $\Pi_y\otimes(\Pi_x+\lambda I)=\Pi+\lambda\Pi_y\otimes I\preccurlyeq\Pi+C\lambda I$ for some constant C.

To study the third term in the last line of (26), we consider

$$\begin{aligned}&\|(\Pi_y+\lambda I)^{-1/2}\frac{1}{n}\sum_{i=1}^n(\widehat{\Pi}_{y,i}-\Pi_y)(\Pi_y+\lambda I)^{-1/2}\|_{op}\\&=\sup_{h_1,h_2:\|h_1\|_{\mathcal{L}^2}=\|h_2\|_{\mathcal{L}^2}=1}\langle h_1,(\Pi_y+\lambda I)^{-1/2}\frac{1}{n}\sum_{i=1}^n(\widehat{\Pi}_{y,i}-\Pi_y)(\Pi_y+\lambda I)^{-1/2}h_2\rangle_{\mathcal{L}^2}\\&\le\left(\sum_{v_1=1}^\infty\sum_{v_2=1}^\infty(\gamma^y_{v_1}+\lambda)^{-1}(\gamma^y_{v_2}+\lambda)^{-1}\left\langle\varphi^y_{v_1},\frac{1}{n}\sum_{i=1}^n(\widehat{\Pi}_{y,i}-\Pi_y)\varphi^y_{v_2}\right\rangle^2_{\mathcal{L}^2}\right)^{1/2},\end{aligned}$$

where the establishment of the last line is similar to the previous results. Because

$$\mathbb{E}\left[\left\langle\varphi^y_{v_1},\frac{1}{n}\sum_{i=1}^n(\widehat{\Pi}_{y,i}-\Pi_y)\varphi^y_{v_2}\right\rangle^2_{\mathcal{L}^2}\right]=\frac{1}{n}\mathbb{E}\left[\left\langle\varphi^y_{v_1},(\widehat{\Pi}_{y,i}-\Pi_y)\varphi^y_{v_2}\right\rangle^2_{\mathcal{L}^2}\right]\lesssim\frac{v_1^{-2r_y}v_2^{-2r_y}}{nm_y}.$$

where the inequality is given by (29) and (30). Now, we have

$$\sum_{v_1=1}^{\infty}\sum_{v_2=1}^{\infty}(\gamma^y_{v_1}+\lambda)^{-1}(\gamma^y_{v_2}+\lambda)^{-1}\mathbb{E}\left[\left\langle \varphi^y_{v_1}, \frac{1}{n}\sum_{i=1}^{n}(\widehat{\Pi}_{y,i}-\Pi_y)\varphi^y_{v_2}\right\rangle^2_{\mathcal{L}^2}\right]$$
$$\lesssim \frac{1}{nm_y}\left(\sum_{v=1}^{\infty}\frac{v^{-2r_y}}{v^{-2r_y}+\lambda}\right)^2 = \frac{1}{nm_y}\left(\sum_{v=1}^{\infty}\frac{1}{1+\lambda v^{2r_y}}\right)^2 \lesssim \frac{\lambda^{-1/r_y}}{nm_y} = o(1).$$

where the final $o(1)$ follows from the relaxed condition $\lambda^{-1} = o\big((nm_y \log^{-2} n)^{r_y}\big)$. Therefore,

$$\|(\Pi_y+\lambda I)^{-1/2}\frac{1}{n}\sum_{i=1}^{n}(\widehat{\Pi}_{y,i}-\Pi_y)(\Pi_y+\lambda I)^{-1/2}\|_{op} = o_p(1).$$

This yields

$$-o_p(1)(\Pi_y+\lambda I) \preccurlyeq \frac{1}{n}\sum_{i=1}^{n}(\widehat{\Pi}_{y,i}-\Pi_y) \preccurlyeq o_p(1)(\Pi_y+\lambda I),$$

which further implies

$$-o_p(1)(\Pi_y+\lambda I)\otimes\Pi_x \preccurlyeq \frac{1}{n}\sum_{i=1}^{n}(\widehat{\Pi}_{y,i}-\Pi_y)\otimes\Pi_x \preccurlyeq o_p(1)(\Pi_y+\lambda I)\otimes\Pi_x. \tag{35}$$

Plugging (33), (34) and (35) in (26), we have

$$\widehat{\Pi}+\lambda I \succcurlyeq (1-o_p(1))(\Pi+\lambda I).$$

This implies the desired result.

□

PROOF OF THEOREM 3.3. For notational convenience, we still define $\hat{f} = L_K^{-1/2}\hat{\beta}$ and $f_0 = L_K^{-1/2}\beta_0$. Note that β_0 is not necessarily an element of $\mathcal{H}(K)$, implying that f_0 may not belong to $\mathcal{L}^2(\mathcal{T}_y\times\mathcal{T}_x, \nu_y\otimes\nu_x)$. Let $\bar{f} = (\Pi+\lambda I)^{-1}\Pi L_K^{-1/2}\beta_0$. With these definitions in place, we can establish the following inequality for the error

$$\mathcal{E}_{\beta_0}(\hat{\beta}) \lesssim \|\Pi^{1/2}(\hat{f}-\bar{f})\|^2_{\mathcal{L}^2} + \|\Pi^{1/2}(\bar{f}-f_0)\|^2_{\mathcal{L}^2}. \tag{36}$$

To bound the first term, we denote $\widetilde{\Pi}$ as the operator

$$\widetilde{\Pi}f = \frac{1}{nm_y}\sum_{i=1}^{n}\sum_{l=1}^{m_y}\int_{\mathcal{T}_x}X_i(t)(L_K^{1/2}f)(S_{il},t)\,d\nu_x(t)\frac{1}{m_x}\sum_{j=1}^{m_x}X_{ij}L_K^{-1/2}K(S_{il},T_{ij};\cdot,\cdot),$$

for f that $(L_K^{1/2}f)\in\mathcal{L}^2(\mathcal{T}_y\times\mathcal{T}_x,\nu_y\otimes\nu_x)$, it easy to verify that $\mathbb{E}[\widetilde{\Pi}f] = \Pi f$. Furthermore, we also denote that

$$g_{nm} = \frac{1}{nm_y}\sum_{i=1}^{n}\sum_{l=1}^{m_y}(U_{il}+e_{il})\frac{1}{m_x}\sum_{j=1}^{m_x}X_{ij}L_K^{-1/2}K(S_{il},T_{ij};\cdot,\cdot).$$

Similar to (9) in the proof of Theorem 3.1, we have

$$\begin{aligned}\hat{f}-\bar{f} =& (\Pi+\lambda I)^{-1}(\Pi-\widehat{\Pi})(\hat{f}-\bar{f}) + (\Pi+\lambda I)^{-1}(\widetilde{\Pi}-\Pi)f_0 \\ &+ (\Pi+\lambda I)^{-1}(\Pi-\widehat{\Pi})\bar{f} + (\Pi+\lambda I)^{-1}g_{nm}.\end{aligned} \tag{37}$$

Let ν be a value satisfying the inequalities

$$0 \le \nu, \quad 4\nu r_x - 2r_x + 2s < -1, \quad 4\nu r_y - 2r_y < -1, \quad 2\nu > \alpha.$$

We need to verify that such a value for ν is non-empty. Since $r_x - s > 1/2$ and $r_y > 1/2$, this can be achieved by checking $\alpha < (2r_x - 2s - 1)/2r_x$ and $\alpha < (2r_y - 1)/2r_y$, which follows from Assumption 3. Then we can obtain

$$\begin{aligned} \|\Pi^{1/2}(\hat{f} - \bar{f})\|_{\mathcal{L}^2}^2 \lesssim & \|\Pi^{1/2}(\Pi + \lambda I)^{-1}(\Pi - \widehat{\Pi})\Pi^{-\nu}\|_{op}^2 \|\Pi^{\nu}(\hat{f} - \bar{f})\|_{\mathcal{L}^2}^2 \\ & + \|\Pi^{1/2}(\Pi + \lambda I)^{-1}(\widetilde{\Pi} - \Pi)\Pi^{-\alpha/2}\|_{op}^2 \|\Pi^{\alpha/2} f_0\|_{\mathcal{L}^2}^2 \\ & + \|\Pi^{1/2}(\Pi + \lambda I)^{-1}(\widehat{\Pi} - \Pi)\Pi^{-\alpha/2}\|_{op}^2 \|\Pi^{\alpha/2} \bar{f}\|_{\mathcal{L}^2}^2 \\ & + \|\Pi^{1/2}(\Pi + \lambda I)^{-1} g_{nm}\|_{\mathcal{L}^2}^2. \end{aligned} \tag{38}$$

For the first term, we have

$$\begin{aligned} & \|\Pi^{1/2}(\Pi + \lambda I)^{-1}(\Pi - \widehat{\Pi})\Pi^{-\nu}\|_{op}^2 \|\Pi^{\nu}(\hat{f} - \bar{f})\|_{\mathcal{L}^2}^2 \\ & \lesssim O_p\left(n^{-1}\lambda^{-1/2r_x} + (nm_x)^{-1}\lambda^{-(2s+1)/2r_x}\log(1/\lambda) + (nm_y)^{-1}\lambda^{-1/2r_y}\log(1/\lambda)\right) o_p(1) \\ & = o_p\left(n^{-1}\lambda^{-1/2r_x} + (nm_x)^{-1}\lambda^{-(2s+1)/2r_x}\log(1/\lambda) + (nm_y)^{-1}\lambda^{-1/2r_y}\log(1/\lambda)\right), \end{aligned}$$

where Lemmas S.2.6 and S.2.8 give

$$\begin{aligned} \|\Pi^{\nu}(\hat{f} - \bar{f})\|_{\mathcal{L}^2}^2 = O_p\Big(& n^{-1}\lambda^{-1+2\nu-1/(2r_x)} + (nm_x)^{-1}\lambda^{-1+2\nu-(2s+1)/(2r_x)}\log(1/\lambda) \\ & + (nm_y)^{-1}\lambda^{-1+2\nu-1/(2r_y)}\log(1/\lambda)\Big) = o_p(1), \end{aligned}$$

where the last relation follows from the choice of λ in Theorem 3.3 and $2\nu > \alpha$; the resulting polynomial margin absorbs the logarithmic factors. For the second and third terms, we obtain

$$\begin{aligned} & \|\Pi^{1/2}(\Pi + \lambda I)^{-1}(\widetilde{\Pi} - \Pi)\Pi^{-\alpha/2}\|_{op}^2 \|\Pi^{\alpha/2} f_0\|_{\mathcal{L}^2}^2 \\ & = O_p\left(n^{-1}\lambda^{-1/2r_x} + (nm_x)^{-1}\lambda^{-(2s+1)/2r_x}\log(1/\lambda) + (nm_y)^{-1}\lambda^{-1/2r_y}\log(1/\lambda)\right), \end{aligned}$$

and

$$\begin{aligned} & \|\Pi^{1/2}(\Pi + \lambda I)^{-1}(\widehat{\Pi} - \Pi)\Pi^{-\alpha/2}\|_{op}^2 \|\Pi^{\alpha/2} \bar{f}\|_{\mathcal{L}^2}^2 \\ & \lesssim \|\Pi^{1/2}(\Pi + \lambda I)^{-1}(\widehat{\Pi} - \Pi)\Pi^{-\alpha/2}\|_{op}^2 \left(\|\Pi^{\alpha/2}(f_0 - \bar{f})\|_{\mathcal{L}^2}^2 + \|\Pi^{\alpha/2} f_0\|_{\mathcal{L}^2}^2\right) \\ & = O_p\left(n^{-1}\lambda^{-1/2r_x} + (nm_x)^{-1}\lambda^{-(2s+1)/2r_x}\log(1/\lambda) + (nm_y)^{-1}\lambda^{-1/2r_y}\log(1/\lambda)\right), \end{aligned}$$

where the last inequality follows from Lemma S.2.6 and Lemma S.2.5. Furthermore, the last term in (38) is bounded by Lemma S.2.7. Substituting the above results into (38), we obtain the following bound

$$\begin{aligned} \|\Pi^{1/2}(\hat{f} - \bar{f})\|_{\mathcal{L}^2}^2 = O_p\big(& n^{-1}\lambda^{-1/2r_x} + (nm_x)^{-1}\lambda^{-(2s+1)/2r_x}\log(1/\lambda) \\ & + (nm_y)^{-1}\lambda^{-1/2r_y}\log(1/\lambda)\big), \end{aligned} \tag{39}$$

where the first log term can be omitted when $r_x \neq (2s+1)r_y$ and the second log term can be removed when $r_x \neq r_y$. By combining (36), (39), and Lemma S.2.5, we arrive at the desired result of $\mathcal{E}_{\beta_0}(\hat{\beta})$:

$$O_p\left(\lambda^{1-\alpha} + n^{-1}\lambda^{-1/2r_x} + (nm_x)^{-1}\lambda^{-(2s+1)/2r_x}\log(1/\lambda) + (nm_y)^{-1}\lambda^{-1/2r_y}\log(1/\lambda)\right),$$

where the first log term can be omitted when $r_x \neq (2s+1)r_y$ and the second log term can be removed when $r_x \neq r_y$. This concludes the proof.

$\square$

PROOF OF THEOREM 3.4. The proof of this theorem is divided into two parts. The first part establishes the first two terms of the conclusion, primarily leveraging the result in Theorem 3.2. The second part proves the third term, which is analogous to a classical regression problem. Without loss of generality, set $\mathcal{T}_x = \mathcal{T}_y = [0,1]$.

Step 1. Let $h = \varphi_1^y$ be an $\mathcal{L}^2(\mathcal{T}_y, \nu_y)$-normalized eigenfunction of $\Pi_y = L_{K^y}$ corresponding to $\gamma_1^y > 0$. For the first two terms in the lower bound, consider the scalar-on-function submodel in Theorem 3.2, with $r = r_x$ and slope function β_x, and set

$$\beta_0(s,t) = h(s)\beta_x(t), \qquad U_i(s) = h(s)u_i, \qquad e_{il} = 0,$$

where $u_i \sim N(0,1)$ is independent of all other random variables. Since $L_K = L_{K^y} \otimes L_{K^x}$ and $\Pi = \Pi_y \otimes \Pi_x$, we have

$$\left\| \Pi^{\alpha/2} L_K^{-1/2} \beta_0 \right\|_{\mathcal{L}^2}^2 = (\gamma_1^y)^{\alpha-1} \left\| \Pi_x^{\alpha/2} L_{K^x}^{-1/2} \beta_x \right\|_{\mathcal{L}^2}^2 \lesssim 1.$$

Thus, this construction satisfies Assumption 3(c) for every admissible α.

Writing

$$Z_i = \int_{\mathcal{T}_x} X_i(t)\beta_x(t)\, d\nu_x(t) + u_i,$$

the response observations satisfy $Y_{il} = h(S_{il})Z_i$. Hence the function-on-function experiment is obtained from the scalar-on-function experiment by the parameter-independent randomization that independently draws the S_{il}'s and sets $Y_{il} = h(S_{il})Z_i$; it is therefore no more informative than the scalar-on-function experiment.

For any estimator $\widetilde{\beta}$ based on the function-on-function observations, define

$$\widetilde{\beta}_h(t) = \int_{\mathcal{T}_y} h(s)\widetilde{\beta}(s,t)\, d\nu_y(s).$$

Because $\|h\|_{\mathcal{L}^2} = 1$, the contraction property of the orthogonal projection onto $\mathrm{span}\{h\}$ gives

$$\mathcal{E}_{\beta_0}(\widetilde{\beta}) \geq \left\| L_C^{1/2}(\widetilde{\beta}_h - \beta_x) \right\|_{\mathcal{L}^2(\mathcal{T}_x, \nu_x)}^2.$$

Composing any function-on-function estimator with the above parameter-independent randomization and then applying this projection produces a (possibly randomized) estimator in the scalar-on-function experiment. Consequently, Theorem 3.2 yields the first two terms in the asserted lower bound.

Step 2. It remains to prove the third term in the lower bound. For this lower-bound submodel, select distinct eigenfunctions $\{\varphi_u^x : u \geq 1\}$ such that

$$L_{K^x}\varphi_u^x = \lambda^x_{[u^{(r_x - s)/r_b}]}\varphi_u^x, \qquad u \geq 1.$$

Here the selection is possible because $r_x - s \geq r_b$. Set $\mu_u = u^{-2s}$ and

$$C(t_1,t_2) = \sum_{u=1}^{\infty} \mu_u \varphi_u^x(t_1)\varphi_u^x(t_2), \qquad X_i(t) = \sum_{u=1}^{\infty} \mu_u^{1/2}\xi_{iu}\varphi_u^x(t),$$

where the ξ_{iu}'s are independent Rademacher random variables. Then $L_C\varphi_u^x = \mu_u\varphi_u^x$, and

$$\Pi_x = \sum_{u=1}^{\infty} \mu_u \lambda^x_{[u^{(r_x - s)/r_b}]} \varphi_u^x \otimes \varphi_u^x = \sum_{u=1}^{\infty} \gamma_u^x \varphi_u^x \otimes \varphi_u^x,$$

and hence

$$\gamma_u^x \asymp u^{-2r_x}, \qquad (\gamma_u^x)^{-1}\|L_{K^x}^{1/2}\varphi_u^x\|_{\mathcal{L}^2}^2 = \mu_u^{-1} = u^{2s}.$$

Thus Assumption 3(a) is satisfied. Since the chosen eigenfunctions are uniformly bounded and $s > 1/2$, the covariance series converges uniformly, $\sup_t \mathbb{E}|X_i(t)|^4 < \infty$, and the Rademacher fourth-moment inequality verifies Assumption 1. We also set $\varepsilon_{ij} = 0$, $U_i = 0$, and let $e_{il} \sim N(0, \sigma^2)$ be independent of all other random variables. For any $g \in \mathcal{L}^2(\mathcal{T}_y)$, consider the slope function

$$\beta(s,t) = g(s)(L_{K^x}^{1/2}\varphi_1^x)(t).$$

Then

$$\int_{\mathcal{T}_x} X_i(t)\beta(s,t)\,d\nu_x(t) = g(s)\int_{\mathcal{T}_x} X_i(t)(L_{K^x}^{1/2}\varphi_1^x)(t)\,d\nu_x(t) = (\gamma_1^x)^{1/2}\xi_{i1}g(s).$$

Therefore the observations of the response satisfy

$$Y_{il} = (\gamma_1^x)^{1/2}\xi_{i1}g(S_{il}) + e_{il}.$$

Now give the variables ξ_{i1}'s to the estimator as additional information. This only makes the estimation problem easier, so any lower bound in this enlarged experiment is also a lower bound for the original experiment. Define $Z_{il} := (\gamma_1^x)^{-1/2}\xi_{i1}Y_{il}$. Then

$$Z_{il} = g(S_{il}) + \widetilde{e}_{il}, \qquad \widetilde{e}_{il} := (\gamma_1^x)^{-1/2}\xi_{i1}e_{il} \sim N(0, \sigma^2/\gamma_1^x).$$

Thus the problem contains the standard nonparametric regression subproblem of estimating g from nm_y observations. We consider $\beta_0 = g(s)(L_{K^x}^{1/2}\varphi_1^x)(t)$ and $\beta_0 = g'(s)(L_{K^x}^{1/2}\varphi_1^x)(t)$, then the Kullback–Leibler divergence

$$\mathrm{D}_{\mathrm{KL}}(\mathcal{P}_\beta\|\mathcal{P}_{\beta'}) \le C_1 nm_y\|g - g'\|_{\mathcal{L}^2}^2. \tag{40}$$

Let $M = \left\lceil C_2(nm_y)^{1/(2(1-\alpha)r_y+1)}\right\rceil$, where $C_2 > 0$ is a sufficiently large constant. For $\boldsymbol{a} = (a_1, \ldots, a_M) \in \{0,1\}^M$, define

$$g_{\boldsymbol{a}}(s) = \sum_{k=M+1}^{2M} a_{k-M}M^{-1/2}k^{(\alpha-1)r_y}\varphi_k^y(s),$$

and set $\beta_{\boldsymbol{a}}(s,t) = g_{\boldsymbol{a}}(s)(L_{K^x}^{1/2}\varphi_1^x)(t)$. We have

$$\begin{aligned}\left\|\Pi^{\alpha/2}L_K^{-1/2}\beta_{\boldsymbol{a}}\right\|_{\mathcal{L}^2(\mathcal{T}_y\times\mathcal{T}_x)}^2 &= \left\|\Pi_y^{\alpha/2}L_{K^y}^{-1/2}g_{\boldsymbol{a}}\right\|_{\mathcal{L}^2(\mathcal{T}_y)}^2\left\|\Pi_x^{\alpha/2}\varphi_1^x\right\|_{\mathcal{L}^2(\mathcal{T}_x)}^2\\ &= (\gamma_1^x)^\alpha\left\|\Pi_y^{(\alpha-1)/2}g_{\boldsymbol{a}}\right\|_{\mathcal{L}^2(\mathcal{T}_y)}^2\\ &= (\gamma_1^x)^\alpha\sum_{k=M+1}^{2M} a_{k-M}^2M^{-1}k^{2(\alpha-1)r_y}(\gamma_k^y)^{\alpha-1} \lesssim 1,\end{aligned}$$

where the last inequality follows from $\gamma_k^y \asymp k^{-2r_y}$.

Utilizing the Varshamov-Gilbert bound, one can identify a subset $\mathcal{A} \subseteq \{0,1\}^M$ that satisfies $|\mathcal{A}| \ge \exp(M/8)$, and for any distinct vectors $\boldsymbol{a}, \boldsymbol{a}' \in \mathcal{A}$, the Hamming distance between them is bounded below by $\mathrm{Ham}(\boldsymbol{a},\boldsymbol{a}') \ge M/8$. By (40),

$$\begin{aligned}
\mathrm{D}_{\mathrm{KL}}(\mathcal{P}_{\beta_{\boldsymbol{a}}}||\mathcal{P}_{\beta_{\boldsymbol{a}'}}) &\lesssim nm_y\|g_{\boldsymbol{a}} - g_{\boldsymbol{a}'}\|_{\mathcal{L}^2}^2 \\
&= nm_y\left\|\sum_{k=M+1}^{2M}(a_{k-M} - a'_{k-M})M^{-1/2}k^{(\alpha-1)r_y}\varphi_k^y\right\|_{\mathcal{L}^2}^2 \\
&= nm_y\sum_{k=M+1}^{2M}(a_{k-M} - a'_{k-M})^2M^{-1}k^{2(\alpha-1)r_y} \lesssim nm_yM^{2(\alpha-1)r_y}.
\end{aligned}$$

For arbitrary $\boldsymbol{a}_0 \in \mathcal{A}$, we have the following inequality

$$\frac{1}{|\mathcal{A}|}\sum_{\boldsymbol{a}\in\mathcal{A}}\mathrm{D}_{\mathrm{KL}}(\mathcal{P}_{\beta_{\boldsymbol{a}}}||\mathcal{P}_{\beta_{\boldsymbol{a}_0}}) \lesssim nm_yM^{2(\alpha-1)r_y} \le bM/8,$$

for any $b \in (0,1/8)$, where the second inequality follows by choosing the constant C_2 sufficiently large. Additionally, for any distinct vectors $\boldsymbol{a},\boldsymbol{a}' \in \mathcal{A}$, we have

$$\begin{aligned}
\|(I\otimes L_C)^{1/2}(\beta_{\boldsymbol{a}} - \beta_{\boldsymbol{a}'})\|_{\mathcal{L}^2}^2 &= \gamma_1^x\|g_{\boldsymbol{a}} - g_{\boldsymbol{a}'}\|_{\mathcal{L}^2(\mathcal{T}_y)}^2 \\
&= \gamma_1^x\sum_{k=M+1}^{2M}(a_{k-M} - a'_{k-M})^2M^{-1}k^{2(\alpha-1)r_y} \\
&\ge \gamma_1^xM^{-1}(2M)^{2(\alpha-1)r_y}\,\mathrm{Ham}(\boldsymbol{a},\boldsymbol{a}') \\
&\ge 4C_3(nm_y)^{-2(1-\alpha)r_y/(2(1-\alpha)r_y+1)}.
\end{aligned}$$

Here $C_3 > 0$ may depend on the fixed constant γ_1^x but is independent of n, m_x, m_y. Applying Theorem 2.5 of Tsybakov and Zaiats (2009) yields

$$\begin{aligned}
&\inf_{\tilde\beta}\sup_{\mathcal{P}\in\mathcal{Q}}\mathbb{P}\left(\mathcal{E}_{\beta_0}(\tilde\beta) > C_3(nm_y)^{-2(1-\alpha)r_y/(2(1-\alpha)r_y+1)}\right) \\
&\ge \frac{\sqrt{|\mathcal{A}|}}{1+\sqrt{|\mathcal{A}|}}\left(1 - 2b - \sqrt{\frac{2b}{\log|\mathcal{A}|}}\right) \to 1 - 2b \quad \text{as } n\to\infty.
\end{aligned}$$

This concludes the proof.

□

S.2. Lemmas and their proofs. In this section, we present the lemmas used in Section S.1 along with their proofs.

LEMMA S.2.1 (Deterministic error for scalar-on-function regression). Consider the scalar-on-function regression problem, for β_0 satisfying $\Pi^{\alpha/2}L_K^{-1/2}\beta_0 \in \mathcal{L}^2(\mathcal{T}_x,\nu_x)$ and ν satisfying $0 \le 2\nu - \alpha \le 2$, we have

$$\|\Pi^\nu\bar f - \Pi^\nu L_K^{-1/2}\beta_0\|_{\mathcal{L}^2}^2 \lesssim \lambda^{2\nu-\alpha},$$

where $\bar f = (\Pi + \lambda I)^{-1}\Pi L_K^{-1/2}\beta_0$.

PROOF OF LEMMA S.2.1. Because $\Pi^{\alpha/2}L_K^{-1/2}\beta_0 \in \mathcal{L}^2(\mathcal{T}_x,\nu_x)$, we assume that

$$\beta_0 = \sum_{k=1}^{\infty}a_kL_K^{1/2}\varphi_k \quad \text{where} \quad \sum_{k=1}^{\infty}\gamma_k^\alpha a_k^2 < \infty.$$

Then

$$\bar{f} = (\Pi + \lambda I)^{-1}\Pi L_K^{-1/2}\beta_0 = \sum_{k=1}^{\infty} \frac{a_k \gamma_k}{\gamma_k + \lambda}\varphi_k.$$

Therefore, we have

$$\|\Pi^\nu \bar{f} - \Pi^\nu L_K^{-1/2}\beta_0\|_{\mathcal{L}^2}^2 = \sum_{k=1}^{\infty} \gamma_k^{2\nu}\left(\frac{a_k\lambda}{\gamma_k + \lambda}\right)^2 \le \sup_{k\ge 1}\frac{\lambda^2\gamma_k^{2\nu-\alpha}}{(\gamma_k+\lambda)^2}\sum_{k=1}^{\infty}\gamma_k^{\alpha}a_k^2$$

Using Young's inequality, when $2\nu - \alpha \in [0, 2]$,

$$\gamma_k + \lambda \gtrsim \gamma_k^{(2\nu-\alpha)/2}\lambda^{1-(2\nu-\alpha)/2}.$$

Hence, we obtain

$$\|\Pi^\nu \bar{f} - \Pi^\nu L_K^{-1/2}\beta_0\|_{\mathcal{L}^2}^2 \lesssim \lambda^{2\nu-\alpha}.$$

The desired result follows. □

LEMMA S.2.2. Consider the scalar-on-function regression problem, if $a < 1/2 + 1/(4r)$, $-4ar - 2r + 2s < -1$, $4br - 2r + 2s < -1$ and Assumption 1 hold,

$$\|\Pi^a(\Pi + \lambda I)^{-1}(\widetilde{\Pi} - \Pi)\Pi^{-b}\|_{op}^2 = O_p(n^{-1}\lambda^{-1+2a-1/2r} + (nm_x)^{-1}\lambda^{-1+2a-(2s+1)/2r})$$

and

$$\|\Pi^a(\Pi + \lambda I)^{-1}(\widehat{\Pi} - \Pi)\Pi^{-b}\|_{op}^2 = O_p(n^{-1}\lambda^{-1+2a-1/2r} + (nm_x)^{-1}\lambda^{-1+2a-(2s+1)/2r}).$$

PROOF OF LEMMA S.2.2. We only prove the first result, while the second result follows a similar argument. Consider the operator norm of $\Pi^a(\Pi + \lambda I)^{-1}(\widetilde{\Pi} - \Pi)\Pi^{-b}$, which is defined by

$$\|\Pi^a(\Pi + \lambda I)^{-1}(\widetilde{\Pi} - \Pi)\Pi^{-b}\|_{op} = \sup_{h_1,h_2:\|h_1\|_{\mathcal{L}^2}=\|h_2\|_{\mathcal{L}^2}=1}\langle h_1, \Pi^a(\Pi + \lambda I)^{-1}(\widetilde{\Pi} - \Pi)\Pi^{-b}h_2\rangle_{\mathcal{L}^2},$$

where the supremum is taken over all function h_1 and h_2 with $\|h_1\|_{\mathcal{L}^2} = \|h_2\|_{\mathcal{L}^2} = 1$ in the space $\mathcal{L}^2(\mathcal{T}_x, \nu_x)$. Let $h_1 = \sum_{k=1}^\infty a_k\varphi_k$ and $h_2 = \sum_{k=1}^\infty b_k\varphi_k$ be the expansions of h_1 and h_2 in terms of the orthonormal basis $\{\varphi_k\}$.Then the operator norm can be expressed as

$$\begin{aligned}
&\sup_{(a_k,b_k)_{k=1}^\infty}\left\langle \sum_{k=1}^{\infty} a_k\varphi_k, \Pi^a(\Pi + \lambda I)^{-1}(\widetilde{\Pi} - \Pi)\Pi^{-b}\sum_{k=1}^{\infty} b_k\varphi_k\right\rangle_{\mathcal{L}^2} \\
&= \sup_{(a_k,b_k)_{k=1}^\infty}\sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty} a_{k_1}b_{k_2}\gamma_{k_1}^a(\gamma_{k_1}+\lambda)^{-1}\gamma_{k_2}^{-b}\left\langle\varphi_{k_1}, (\widetilde{\Pi} - \Pi)\varphi_{k_2}\right\rangle_{\mathcal{L}^2} \\
&\le \sup_{(a_k,b_k)_{k=1}^\infty}\left(\sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty} a_{k_1}^2 b_{k_2}^2\right)^{1/2}\left(\sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty}\frac{\gamma_{k_1}^{2a}\gamma_{k_2}^{-2b}}{(\gamma_{k_1}+\lambda)^2}\left\langle\varphi_{k_1}, (\widetilde{\Pi} - \Pi)\varphi_{k_2}\right\rangle_{\mathcal{L}^2}^2\right)^{1/2} \\
&= \left(\sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty}\frac{\gamma_{k_1}^{2a}\gamma_{k_2}^{-2b}}{(\gamma_{k_1}+\lambda)^2}\left\langle\varphi_{k_1}, (\widetilde{\Pi} - \Pi)\varphi_{k_2}\right\rangle_{\mathcal{L}^2}^2\right)^{1/2},
\end{aligned}$$

where the supremum is taken over all $(a_k)_{k=1}^\infty$ and $(b_k)_{k=1}^\infty$ with $\sum_{k=1}^\infty a_k^2 = \sum_{k=1}^\infty b_k^2 = 1$ and the third line is implied from Cauchy inequality.

Now, we need to bound $\mathbb{E}[\langle \varphi_{k_1}, (\widetilde{\Pi} - \Pi)\varphi_{k_2}\rangle^2_{\mathcal{L}^2}]$. Note that

$$\left\langle \varphi_{k_1}, \widetilde{\Pi}\varphi_{k_2} \right\rangle_{\mathcal{L}^2} = \frac{1}{n}\sum_{i=1}^{n}\int_{\mathcal{T}_x} X_i(t)(L_K^{1/2}\varphi_{k_2})(t)\, d\nu_x(t) \frac{1}{m_x}\sum_{j=1}^{m_x} X_{ij}(L_K^{1/2}\varphi_{k_1})(T_{ij}).$$

Then a similar analysis with (1)-(6) in the proof of Proposition 3.1 yields

$$\begin{aligned} &\mathbb{E}\left[\left\langle \varphi_{k_1}, (\widetilde{\Pi} - \Pi)\varphi_{k_2} \right\rangle^2_{\mathcal{L}^2}\right] \\ &\lesssim \frac{1}{n} k_1^{-2r} k_2^{-2r} + \frac{1}{nm_x}(k_1^{2s-2r}k_2^{-2r} + k_1^{-2r}k_2^{2s-2r}) + \frac{1}{nm_x^2} k_1^{2s-2r}k_2^{2s-2r}. \end{aligned} \tag{41}$$

Therefore, we have

$$\begin{aligned} &\mathbb{E}\left[\sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty} \frac{\gamma_{k_1}^{2a}\gamma_{k_2}^{-2b}}{(\gamma_{k_1}+\lambda)^2} \left\langle \varphi_{k_1}, (\widetilde{\Pi} - \Pi)\varphi_{k_2} \right\rangle^2_{\mathcal{L}^2}\right] \\ &\lesssim \frac{1}{n}\sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty} \frac{\gamma_{k_1}^{2a}\gamma_{k_2}^{-2b}}{(\gamma_{k_1}+\lambda)^2}\left(k_1^{-2r}k_2^{-2r} + \frac{1}{m_x}k_1^{2s-2r}k_2^{-2r} + \frac{1}{m_x}k_1^{-2r}k_2^{2s-2r} + \frac{1}{m_x^2}k_1^{2s-2r}k_2^{2s-2r}\right) \\ &\lesssim \frac{1}{n}\sum_{k_1=1}^{\infty} \frac{k_1^{-4ar-2r}}{(k_1^{-2r}+\lambda)^2}\left(1 + \frac{1}{m_x}k_1^{2s}\right)\sum_{k_2=1}^{\infty} k_2^{4br-2r}\left(1 + \frac{1}{m_x}k_2^{2s}\right) \end{aligned} \tag{42}$$

Note that

$$\begin{aligned} \sum_{k_1=1}^{\infty} \frac{k_1^{-4ar-2r}}{(k_1^{-2r}+\lambda)^2} &= \sum_{k_1=1}^{\infty} \frac{k_1^{-4ar+2r}}{(1+\lambda k_1^{2r})^2} \lesssim \int_0^{\infty} \frac{x^{-4ar+2r}}{(1+\lambda x^{2r})^2}dx \\ &= \lambda^{-1+2a-1/2r}\int_0^{\infty} \frac{y^{-4ar+2r}}{(1+y^{2r})^2}dy \asymp \lambda^{-1+2a-1/2r}, \end{aligned}$$

and

$$\begin{aligned} \sum_{k_1=1}^{\infty} \frac{k_1^{-4ar-2r+2s}}{(k_1^{-2r}+\lambda)^2} &= \sum_{k_1=1}^{\infty} \frac{k_1^{-4ar+2r+2s}}{(1+\lambda k_1^{2r})^2} \lesssim \int_0^{\infty} \frac{x^{-4ar+2r+2s}}{(1+\lambda x^{2r})^2}dx \\ &= \lambda^{-1+2a-(2s+1)/2r}\int_0^{\infty} \frac{y^{-4ar+2r+2s}}{(1+y^{2r})^2}dy \asymp \lambda^{-1+2a-(2s+1)/2r}, \end{aligned}$$

The condition $a < 1/2 + 1/(4r)$ guarantees integrability near zero in both preceding displays, whereas $-4ar - 2r + 2s < -1$ guarantees integrability at infinity. We have

$$\sum_{k_1=1}^{\infty} \frac{k_1^{-4ar-2r}}{(k_1^{-2r}+\lambda)^2}\left(1 + \frac{1}{m_x}k_1^{2s}\right) \asymp \lambda^{-1+2a-1/2r} + \frac{1}{m_x}\lambda^{-1+2a-(2s+1)/2r}.$$

Meanwhile, when $4br - 2r + 2s < -1$, the second factor in the last line (42)

$$\sum_{k_2=1}^{\infty} k_2^{4br-2r}\left(1 + \frac{1}{m_x}k_2^{2s}\right) \lesssim 1.$$

Plugging the above two results in (42), we have

$$\begin{aligned} &\mathbb{E}\left[\sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty} \frac{\gamma_{k_1}^{2a}\gamma_{k_2}^{-2b}}{(\gamma_{k_1}+\lambda)^2} \left\langle \varphi_{k_1}, (\widetilde{\Pi} - \Pi)\varphi_{k_2} \right\rangle^2_{\mathcal{L}^2}\right] \\ &\qquad \lesssim n^{-1}\lambda^{-1+2a-1/2r} + (nm_x)^{-1}\lambda^{-1+2a-(2s+1)/2r}. \end{aligned}$$

Then the desired result follows. □

LEMMA S.2.3. In the scalar-on-function regression problem, recalling that

$$g_{nm} = \frac{1}{n}\sum_{i=1}^{n} e_i \frac{1}{m_x}\sum_{j=1}^{m_x} X_{ij} L_K^{-1/2} K(T_{ij}, \cdot),$$

for any $\nu \in [0, 1/2]$, we have

$$\|\Pi^\nu(\Pi + \lambda I)^{-1} g_{nm}\|_{\mathcal{L}^2}^2 = O_p\left(n^{-1}\lambda^{-1+2\nu-1/2r} + (nm_x)^{-1}\lambda^{-1+2\nu-(2s+1)/2r}\right).$$

PROOF OF LEMMA S.2.3. Note that

$$\begin{aligned}
\|\Pi^\nu(\Pi + \lambda I)^{-1} g_{nm}\|_{\mathcal{L}^2}^2 &= \sum_{k=1}^{\infty} \langle \Pi^\nu(\Pi + \lambda I)^{-1} g_{nm}, \varphi_k\rangle_{\mathcal{L}^2}^2 \\
&= \sum_{k=1}^{\infty} \frac{\gamma_k^{2\nu}}{(\gamma_k + \lambda)^2} \langle g_{nm}, \varphi_k\rangle_{\mathcal{L}^2}^2.
\end{aligned} \tag{43}$$

Recalling that

$$g_{nm} = \frac{1}{n}\sum_{i=1}^{n} e_i \frac{1}{m_x}\sum_{j=1}^{m_x} X_{ij} L_K^{-1/2} K(T_{ij}, \cdot),$$

we have

$$\begin{aligned}
\langle g_{nm}, \varphi_k\rangle_{\mathcal{L}^2} &= \left\langle \frac{1}{n}\sum_{i=1}^{n} e_i \frac{1}{m_x}\sum_{j=1}^{m_x} X_{ij} L_K^{-1/2} K(T_{ij}, \cdot), \varphi_k \right\rangle_{\mathcal{L}^2} \\
&= \left\langle \frac{1}{n}\sum_{i=1}^{n} e_i \frac{1}{m_x}\sum_{j=1}^{m_x} X_{ij} K(T_{ij}, \cdot), L_K^{1/2}\varphi_k \right\rangle_{\mathcal{H}} \\
&= \frac{1}{n}\sum_{i=1}^{n} e_i \frac{1}{m_x}\sum_{j=1}^{m_x} X_{ij} (L_K^{1/2}\varphi_k)(T_{ij}).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{E}\left[\langle g_{nm}, \varphi_k\rangle_{\mathcal{L}^2}^2\right] &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} e_i \frac{1}{m_x}\sum_{j=1}^{m_x} X_{ij} (L_K^{1/2}\varphi_k)(T_{ij})\right) \\
&= \frac{1}{n}\mathrm{Var}\left(e_1 \frac{1}{m_x}\sum_{j=1}^{m_x} X_{1j} (L_K^{1/2}\varphi_k)(T_{1j})\right) \\
&\lesssim \frac{1}{n}\mathbb{E}\left[\left(\frac{1}{m_x}\sum_{j=1}^{m_x} X_{1j} (L_K^{1/2}\varphi_k)(T_{1j})\right)^2\right]
\end{aligned}$$

Expanding on the last line, we know it is less than a constant multiplied by

$$\begin{aligned}
&\frac{1}{n}\mathbb{E}\left[X_{11}(L_K^{1/2}\varphi_k)(T_{11}) X_{12}(L_K^{1/2}\varphi_k)(T_{12})\right] + \frac{1}{nm_x}\mathbb{E}\left[X_{11}^2 (L_K^{1/2}\varphi_k)(T_{11})^2\right] \\
&\lesssim \frac{1}{n}\gamma_k + \frac{1}{nm_x}\|L_K^{1/2}\varphi_k(t)\|_{\mathcal{L}^2}^2 \lesssim \frac{1}{n}k^{-2r} + \frac{1}{nm_x}k^{2s-2r}.
\end{aligned} \tag{44}$$

Plugging in (43), we have

$$\mathbb{E}\left[\|\Pi^\nu(\Pi+\lambda I)^{-1}g_{nm}\|_{\mathcal{L}^2}^2\right] = \sum_{k=1}^\infty \frac{\gamma_k^{2\nu}}{(\gamma_k+\lambda)^2}\mathbb{E}\left[\langle g_{nm},\varphi_k\rangle_{\mathcal{L}^2}^2\right]$$

$$\lesssim \frac{1}{n}\sum_{k=1}^\infty \frac{k^{-4\nu r-2r}}{(k^{-2r}+\lambda)^2} + \frac{1}{nm_x}\sum_{k=1}^\infty \frac{k^{-4\nu r+2s-2r}}{(k^{-2r}+\lambda)^2}$$

$$\lesssim \frac{1}{n}\lambda^{2\nu-1-1/2r} + \frac{1}{nm_x}\lambda^{2\nu-1-(2s+1)/2r}.$$

where the last line is due to

$$\sum_{k=1}^\infty \frac{k^{-4\nu r-2r}}{(k^{-2r}+\lambda)^2} = \sum_{k=1}^\infty \frac{k^{-4\nu r+2r}}{(1+\lambda k^{2r})^2} \lesssim \int_0^\infty \frac{x^{-4\nu r+2r}}{(1+\lambda x^{2r})^2}dx$$

$$= \lambda^{2\nu-1-1/2r}\int_0^\infty \frac{y^{-4\nu r+2r}}{(1+y^{2r})^2}dy \lesssim \lambda^{2\nu-1-1/2r}$$

and

$$\sum_{k=1}^\infty \frac{k^{-4\nu r+2s-2r}}{(k^{-2r}+\lambda)^2} = \sum_{k=1}^\infty \frac{k^{-4\nu r+2s+2r}}{(1+\lambda k^{2r})^2} \lesssim \int_0^\infty \frac{x^{-4\nu r+2s+2r}}{(1+\lambda x^{2r})^2}dx$$

$$= \lambda^{2\nu-1-(2s+1)/2r}\int_0^\infty \frac{y^{-4\nu r+2s+2r}}{(1+y^{2r})^2}dy \lesssim \lambda^{2\nu-1-(2s+1)/2r}.$$

Then the desired result follows from Markov inequality. □

LEMMA S.2.4. Under the same conditions of Theorem 3.1, for any $0 \le \nu$ satisfying $-4\nu r - 2r + 2s < -1$, $4\nu r - 2r + 2s < -1$ and $2\nu > \alpha$, we have

$$\|\Pi^\nu(\hat{f}-\bar{f})\|_{\mathcal{L}^2}^2 = O_p\left(n^{-1}\lambda^{-1+2\nu-1/2r} + (nm_x)^{-1}\lambda^{-1+2\nu-(2s+1)/2r}\right).$$

PROOF OF LEMMA S.2.4. Using (9), we have

$$\begin{aligned}\|\Pi^\nu(\hat{f}-\bar{f})\|_{\mathcal{L}^2}^2 \lesssim & \|\Pi^\nu(\Pi+\lambda I)^{-1}(\Pi-\widehat{\Pi})(\hat{f}-\bar{f})\|_{\mathcal{L}^2}^2 \\ & + \|\Pi^\nu(\Pi+\lambda I)^{-1}(\widetilde{\Pi}-\Pi)f_0\|_{\mathcal{L}^2}^2 \\ & + \|\Pi^\nu(\Pi+\lambda I)^{-1}(\widehat{\widetilde{\Pi}}-\Pi)\bar{f}\|_{\mathcal{L}^2}^2 \\ & + \|\Pi^\nu(\Pi+\lambda I)^{-1}g_{nm}\|_{\mathcal{L}^2}^2.\end{aligned} \tag{45}$$

We bound the first term by

$$\|\Pi^\nu(\Pi+\lambda I)^{-1}(\Pi-\widehat{\Pi})(\hat{f}-\bar{f})\|_{\mathcal{L}^2}^2$$

$$\lesssim \|\Pi^\nu(\Pi+\lambda I)^{-1}(\Pi-\widehat{\Pi})\Pi^{-\nu}\|_{op}^2\|\Pi^\nu(\hat{f}-\bar{f})\|_{\mathcal{L}^2}^2$$

$$\lesssim o_p(1)\|\Pi^\nu(\hat{f}-\bar{f})\|_{\mathcal{L}^2}^2,$$

where the last line applies Lemma S.2.2 with $a = b = \nu$ under the two exponent conditions in its statement. The second term and the third term are bounded by

$$\|\Pi^\nu(\Pi+\lambda I)^{-1}(\widetilde{\Pi}-\Pi)f_0\|_{\mathcal{L}^2}^2$$

$$\lesssim \|\Pi^\nu(\Pi+\lambda I)^{-1}(\widetilde{\Pi}-\Pi)\Pi^{-\alpha/2}\|_{op}^2\|\Pi^{\alpha/2}f_0\|_{\mathcal{L}^2}^2$$

$$= O_p(n^{-1}\lambda^{-1+2\nu-1/2r} + (nm_x)^{-1}\lambda^{-1+2\nu-(2s+1)/2r})$$

and

$$\begin{aligned}&\|\Pi^{\nu}(\Pi+\lambda I)^{-1}(\widehat{\Pi}-\Pi)\bar{f}\|_{\mathcal{L}^2}^2\\&\quad\lesssim \|\Pi^{\nu}(\Pi+\lambda I)^{-1}(\widehat{\Pi}-\Pi)\Pi^{-\alpha/2}\|_{op}^2(\|\Pi^{\alpha/2}(\bar{f}-f_0)\|_{\mathcal{L}^2}+\|\Pi^{\alpha/2}f_0\|_{\mathcal{L}^2})^2\\&\quad=O_p(n^{-1}\lambda^{-1+2\nu-1/2r}+(nm_x)^{-1}\lambda^{-1+2\nu-(2s+1)/2r})\end{aligned}$$

where we use Lemma S.2.2 and Lemma S.2.1. Lastly, Lemma S.2.3 implies

$$\|\Pi^{\nu}(\Pi+\lambda I)^{-1}g_{nm}\|_{\mathcal{L}^2}^2=O_p\left(n^{-1}\lambda^{-1+2\nu-1/2r}+(nm_x)^{-1}\lambda^{-1+2\nu-(2s+1)/2r}\right).$$

Combining the above results, we have

$$\|\Pi^{\nu}(\hat{f}-\bar{f})\|_{\mathcal{L}^2}^2=O_p\left(n^{-1}\lambda^{-1+2\nu-1/2r}+(nm_x)^{-1}\lambda^{-1+2\nu-(2s+1)/2r}\right).$$

Then the desired result follows. □

LEMMA S.2.5 (Deterministic error for function-on-function regression). Consider the function-on-function regression problem, for β_0 satisfying $\Pi^{\alpha/2}L_K^{-1/2}\beta_0\in\mathcal{L}^2(\mathcal{T}_y\times\mathcal{T}_x,\nu_y\otimes\nu_x)$ and ν satisfying $0\le 2\nu-\alpha\le 2$, we have

$$\|\Pi^{\nu}\bar{f}-\Pi^{\nu}L_K^{-1/2}\beta_0\|_{\mathcal{L}^2}^2\lesssim\lambda^{2\nu-\alpha},$$

where $\bar{f}=(\Pi+\lambda I)^{-1}\Pi L_K^{-1/2}\beta_0$.

PROOF OF LEMMA S.2.5. Since $\Pi^{\alpha/2}L_K^{-1/2}\beta_0\in\mathcal{L}^2(\mathcal{T}_y\times\mathcal{T}_x,\nu_y\otimes\nu_x)$, we assume that

$$\beta_0=\sum_{u,v}a_{uv}L_K^{1/2}\varphi_{uv}\quad\text{where}\quad\sum_{uv}\gamma_{uv}^{\alpha}a_{uv}^2<\infty.$$

Therefore, it follows that

$$\|\Pi^{\nu}\bar{f}-\Pi^{\nu}L_K^{-1/2}\beta_0\|_{\mathcal{L}^2}^2=\sum_{u,v}\gamma_{uv}^{2\nu}\left(\frac{a_{uv}\lambda}{\gamma_{uv}+\lambda}\right)^2\le\sup_{u,v}\frac{\lambda^2\gamma_{uv}^{2\nu-\alpha}}{(\gamma_{uv}+\lambda)^2}\sum_{u,v}\gamma_{uv}^{\alpha}a_{uv}^2\lesssim\lambda^{2\nu-\alpha}.$$

The desired result follows. □

LEMMA S.2.6. Consider the function-on-function regression problem, if $0\le a\le 1/2$, $-4ar_x-2r_x+2s<-1$, $4br_x-2r_x+2s<-1$, $4br_y-2r_y<-1$ and Assumption 1 hold,

$$\begin{aligned}&\|\Pi^{a}(\Pi+\lambda I)^{-1}(\widetilde{\Pi}-\Pi)\Pi^{-b}\|_{op}^2=O_p(n^{-1}\lambda^{-1+2a-1/2r_x}\\&\quad+(nm_x)^{-1}\lambda^{-1+2a-(2s+1)/2r_x}\log(1/\lambda)+(nm_y)^{-1}\lambda^{-1+2a-1/2r_y}\log(1/\lambda))\end{aligned}$$

and

$$\begin{aligned}&\|\Pi^{a}(\Pi+\lambda I)^{-1}(\widehat{\Pi}-\Pi)\Pi^{-b}\|_{op}^2=O_p(n^{-1}\lambda^{-1+2a-1/2r_x}\\&\quad+(nm_x)^{-1}\lambda^{-1+2a-(2s+1)/2r_x}\log(1/\lambda)+(nm_y)^{-1}\lambda^{-1+2a-1/2r_y}\log(1/\lambda)),\end{aligned}$$

where the first log term can be removed when $r_x\ne(2s+1)r_y$ and the second log term can be removed when $r_x\ne r_y$.

PROOF OF LEMMA S.2.6. We provide the proof of the second result, as the first one follows a similar argument. We examine the operator norm of $\Pi^a(\Pi+\lambda I)^{-1}(\widehat{\Pi}-\Pi)\Pi^{-b}$, which is defined as

$$\begin{aligned}&\|\Pi^a(\Pi+\lambda I)^{-1}(\widehat{\Pi}-\Pi)\Pi^{-b}\|_{\text{op}}\\&=\sup_{\substack{h_1,h_2\in\mathcal{L}^2(\mathcal{T}_y\times\mathcal{T}_x,\nu_y\otimes\nu_x)\\ \|h_1\|_{\mathcal{L}^2}=\|h_2\|_{\mathcal{L}^2}=1}}\left\langle h_1,\Pi^a(\Pi+\lambda I)^{-1}(\widehat{\Pi}-\Pi)\Pi^{-b}h_2\right\rangle_{\mathcal{L}^2},\end{aligned}$$

where the supremum is taken over all h_1 and $h_2\in\mathcal{L}^2(\mathcal{T}_y\times\mathcal{T}_x,\nu_y\otimes\nu_x)$ with $\|h_1\|_{\mathcal{L}^2}=\|h_2\|_{\mathcal{L}^2}=1$. Assume that

$$h_1=\sum_{u=1}^{\infty}\sum_{v=1}^{\infty}a_{uv}\varphi_{uv}\quad\text{and}\quad h_2=\sum_{u=1}^{\infty}\sum_{v=1}^{\infty}b_{uv}\varphi_{uv},$$

with the constraints $\sum_{u=1}^{\infty}\sum_{v=1}^{\infty}a_{uv}^2=\sum_{u=1}^{\infty}\sum_{v=1}^{\infty}b_{uv}^2=1$. The operator norm can then be expressed as:

$$\begin{aligned}&\sup_{(a_{uv},b_{uv})}\left\langle\sum_{u=1}^{\infty}\sum_{v=1}^{\infty}a_{uv}\varphi_{uv},\Pi^a(\Pi+\lambda I)^{-1}(\widehat{\Pi}-\Pi)\Pi^{-b}\sum_{u=1}^{\infty}\sum_{v=1}^{\infty}b_{uv}\varphi_{uv}\right\rangle_{\mathcal{L}^2}\\&=\sup_{(a_{uv},b_{uv})}\sum_{u_1,v_1}\sum_{u_2,v_2}a_{u_1v_1}b_{u_2v_2}\gamma_{u_1v_1}^a(\gamma_{u_1v_1}+\lambda)^{-1}\gamma_{u_2v_2}^{-b}\left\langle\varphi_{u_1v_1},(\widehat{\Pi}-\Pi)\varphi_{u_2v_2}\right\rangle_{\mathcal{L}^2}\\&\le\sup_{(a_{uv},b_{uv})}\left(\sum_{u_1,v_1}\sum_{u_2,v_2}a_{u_1v_1}^2b_{u_2v_2}^2\right)^{1/2}\left(\sum_{u_1,v_1}\sum_{u_2,v_2}\frac{\left\langle\varphi_{u_1v_1},(\widehat{\Pi}-\Pi)\varphi_{u_2v_2}\right\rangle_{\mathcal{L}^2}^2}{\gamma_{u_1v_1}^{-2a}(\gamma_{u_1v_1}+\lambda)^2\gamma_{u_2v_2}^{2b}}\right)^{1/2}\\&=\left(\sum_{u_1,v_1}\sum_{u_2,v_2}\gamma_{u_1v_1}^{2a}(\gamma_{u_1v_1}+\lambda)^{-2}\gamma_{u_2v_2}^{-2b}\left\langle\varphi_{u_1v_1},(\widehat{\Pi}-\Pi)\varphi_{u_2v_2}\right\rangle_{\mathcal{L}^2}^2\right)^{1/2},\end{aligned}$$

where the supremum is taken over all (a_{uv},b_{uv}) satisfying $\sum_{u=1}^{\infty}\sum_{v=1}^{\infty}a_{uv}^2=\sum_{u=1}^{\infty}\sum_{v=1}^{\infty}b_{uv}^2=1$ and the inequality follows directly from the Cauchy-Schwarz inequality.

Recalling the definition of the operator $\widehat{\Pi}-\Pi$, we have

$$\begin{aligned}&\mathbb{E}\left[\left\langle\varphi_{u_1v_1},(\widehat{\Pi}-\Pi)\varphi_{u_2v_2}\right\rangle_{\mathcal{L}^2}^2\right]=\operatorname{Var}\left(\left\langle\varphi_{u_1v_1},\widehat{\Pi}\varphi_{u_2v_2}\right\rangle_{\mathcal{L}^2}\right)\\&=\frac{1}{n}\operatorname{Var}\left(\frac{1}{m_ym_x(m_x-1)}\sum_{l=1}^{m_y}\sum_{j_1\neq j_2}X_{ij_1}X_{ij_2}L_K^{1/2}\varphi_{u_1v_1}(S_{il},T_{ij_1})L_K^{1/2}\varphi_{u_2v_2}(S_{il},T_{ij_2})\right)\\&\lesssim\frac{1}{nm_x^4m_y^2}\mathbb{E}\left[\left(\sum_{l=1}^{m_y}\sum_{j_1\neq j_2}X_{ij_1}X_{ij_2}L_K^{1/2}\varphi_{u_1v_1}(S_{il},T_{ij_1})L_K^{1/2}\varphi_{u_2v_2}(S_{il},T_{ij_2})\right)^2\right].\end{aligned}$$

Note that

$$L_K^{1/2}\varphi_{uv}(S_{il},T_{ij})=L_{K^y}^{1/2}\varphi_v^y(S_{il})L_{K^x}^{1/2}\varphi_u^x(T_{ij}),$$

and that S_{il} are independent with T_{ij}, X_{ij}, we have

$$\begin{aligned}\mathbb{E}\left[\left\langle\varphi_{u_1v_1},(\widehat{\Pi}-\Pi)\varphi_{u_2v_2}\right\rangle^2_{\mathcal{L}^2}\right] &\lesssim \frac{1}{nm_x^4m_y^2}\mathbb{E}\left[\left(\sum_{l=1}^{m_y}L_{K^y}^{1/2}\varphi^y_{v_1}(S_{il})L_{K^y}^{1/2}\varphi^y_{v_2}(S_{il})\right)^2\right] \\ &\cdot\mathbb{E}\left[\left(\sum_{j_1\neq j_2}X_{ij_1}X_{ij_2}L_{K^x}^{1/2}\varphi^x_{u_1}(T_{ij_1})L_{K^x}^{1/2}\varphi^x_{u_2}(T_{ij_2})\right)^2\right].\end{aligned} \tag{46}$$

Similar to the derivation of (6) in the proof of Proposition 3.1, with the measurement-error moments again absorbed only into the implicit constants,

$$\begin{aligned}&\frac{1}{m_x^2}\mathbb{E}\left[\left(\sum_{j_1\neq j_2}X_{ij_1}X_{ij_2}L_{K^x}^{1/2}\varphi^x_{u_1}(T_{ij_1})L_{K^x}^{1/2}\varphi^x_{u_2}(T_{ij_2})\right)^2\right] \\ &\lesssim m_x^2u_1^{-2r_x}u_2^{-2r_x}+m_x(u_1^{2s-2r_x}u_2^{-2r_x}+u_1^{-2r_x}u_2^{2s-2r_x})+u_1^{2s-2r_x}u_2^{2s-2r_x}.\end{aligned} \tag{47}$$

Additionally, when $v_1\neq v_2$,

$$\begin{aligned}&\mathbb{E}\left[\left(\sum_{l=1}^{m_y}L_{K^y}^{1/2}\varphi^y_{v_1}(S_{il})L_{K^y}^{1/2}\varphi^y_{v_2}(S_{il})\right)^2\right] \\ &=\gamma^y_{v_1}\gamma^y_{v_2}\mathbb{E}\left[\left(\sum_{l=1}^{m_y}\varphi^y_{v_1}(S_{il})\varphi^y_{v_2}(S_{il})\right)^2\right]\lesssim m_yv_1^{-2r_y}v_2^{-2r_y}.\end{aligned} \tag{48}$$

When $v_1=v_2$, we have

$$\mathbb{E}\left[\left(\sum_{l=1}^{m_y}L_{K^y}^{1/2}\varphi^y_{v_1}(S_{il})L_{K^y}^{1/2}\varphi^y_{v_2}(S_{il})\right)^2\right]\lesssim m_y^2\gamma^y_{v_1}\gamma^y_{v_2}\|\varphi^y_{v_1}\|^4_{\mathcal{L}^4}\lesssim m_y^2v_1^{-4r_y}. \tag{49}$$

Combining (46) and (47), (48), for $v_1\neq v_2$

$$\begin{aligned}&\mathbb{E}\left[\left\langle\varphi_{u_1v_1},(\widehat{\Pi}-\Pi)\varphi_{u_2v_2}\right\rangle^2_{\mathcal{L}^2}\right] \\ &\lesssim\frac{v_1^{-2r_y}v_2^{-2r_y}}{nm_x^2m_y}\left(m_x^2u_1^{-2r_x}u_2^{-2r_x}+m_x(u_1^{2s-2r_x}u_2^{-2r_x}+u_1^{-2r_x}u_2^{2s-2r_x})+u_1^{2s-2r_x}u_2^{2s-2r_x}\right).\end{aligned}$$

Using (49), we have, for $v_1=v_2$

$$\begin{aligned}&\mathbb{E}\left[\left\langle\varphi_{u_1v_1},(\widehat{\Pi}-\Pi)\varphi_{u_2v_2}\right\rangle^2_{\mathcal{L}^2}\right] \\ &\lesssim\frac{v_1^{-4r_y}}{nm_x^2}\left(m_x^2u_1^{-2r_x}u_2^{-2r_x}+m_x(u_1^{2s-2r_x}u_2^{-2r_x}+u_1^{-2r_x}u_2^{2s-2r_x})+u_1^{2s-2r_x}u_2^{2s-2r_x}\right).\end{aligned}$$

Hence, we have

$$\sum_{u_1,v_1}\sum_{u_2,v_2}\gamma_{u_1v_1}^{2a}(\gamma_{u_1v_1}+\lambda)^{-2}\gamma_{u_2v_2}^{-2b}\mathbb{E}\left[\left\langle\varphi_{u_1v_1},(\widehat{\Pi}-\Pi)\varphi_{u_2v_2}\right\rangle_{\mathcal{L}^2}^2\right]$$

$$\lesssim\sum_{u_1,v_1}\sum_{u_2,v_2}\frac{m_x^2u_1^{-2r_x}u_2^{-2r_x}+m_x(u_1^{2s-2r_x}u_2^{-2r_x}+u_1^{-2r_x}u_2^{2s-2r_x})+u_1^{2s-2r_x}u_2^{2s-2r_x}}{nm_x^2m_y(u_1^{-2r_x}v_1^{-2r_y}+\lambda)^2u_1^{4ar_x}v_1^{2r_y+4ar_y}u_2^{-4br_x}v_2^{2r_y-4br_y}}$$

$$+\sum_{u_1,u_2,v}\frac{m_x^2u_1^{-2r_x}u_2^{-2r_x}+m_x(u_1^{2s-2r_x}u_2^{-2r_x}+u_1^{-2r_x}u_2^{2s-2r_x})+u_1^{2s-2r_x}u_2^{2s-2r_x}}{nm_x^2(u_1^{-2r_x}v^{-2r_y}+\lambda)^2u_1^{4ar_x}u_2^{-4br_x}v^{4(a-b+1)r_y}} \tag{50}$$

$$\lesssim\frac{1}{nm_y}\sum_{u_1,v_1}\frac{m_xu_1^{-2r_x}+u_1^{2s-2r_x}}{m_x(u_1^{-2r_x}v_1^{-2r_y}+\lambda)^2u_1^{4ar_x}v_1^{2r_y+4ar_y}}\sum_{u_2,v_2}\frac{m_xu_2^{-2r_x}+u_2^{2s-2r_x}}{m_xu_2^{-4br_x}v_2^{2r_y-4br_y}}$$

$$+\sum_{u_1,u_2,v}\frac{m_x^2u_1^{-2r_x}u_2^{-2r_x}+m_x(u_1^{2s-2r_x}u_2^{-2r_x}+u_1^{-2r_x}u_2^{2s-2r_x})+u_1^{2s-2r_x}u_2^{2s-2r_x}}{nm_x^2(u_1^{-2r_x}v^{-2r_y}+\lambda)^2u_1^{4ar_x}u_2^{-4br_x}v^{4(a-b+1)r_y}}.$$

Because $-2r_y+4br_y<-1$ and $4br_x-2r_x+2s<-1$, we have

$$\sum_{u_2,v_2}\frac{m_xu_2^{-2r_x}+u_2^{2s-2r_x}}{m_xu_2^{-4br_x}v_2^{2r_y-4br_y}}\lesssim 1.$$

Under the standing conditions, $0\le a\le 1/2$, $r_x-s>1/2$, and $r_y>1/2$ ensure that the changes of variables below are valid and that the exponents to which Lemma S.2.10 is applied are greater than $1/2$. Additionally, the same unit-square comparison, followed by Lemma S.2.10 yields

$$\sum_{u_1,v_1}\frac{u_1^{-2r_x}}{(u_1^{-2r_x}v_1^{-2r_y}+\lambda)^2u_1^{4ar_x}v_1^{2r_y+4ar_y}}\lesssim\int_1^\infty\int_1^\infty\frac{x^{-4ar_x+2r_x}y^{-4ar_y+2r_y}}{(1+\lambda x^{2r_x}y^{2r_y})^2}dxdy$$

$$\lesssim\int_1^\infty\int_1^\infty\left(1+\lambda t_1^{2r_x/(1+2(1-2a)r_x)}t_2^{2r_y/(1+2(1-2a)r_y)}\right)^{-2}dt_1dt_2 \tag{51}$$

$$\lesssim\lambda^{-1+2a-(1/2r_x\vee 1/2r_y)}\log(1/\lambda),$$

where $\log(1/\lambda)$ is not necessary when $r_x\neq r_y$. We also have

$$\sum_{u_1,v_1}\frac{u_1^{2s-2r_x}}{(u_1^{-2r_x}v_1^{-2r_y}+\lambda)^2u_1^{4ar_x}v_1^{2r_y+4ar_y}}\lesssim\int_1^\infty\int_1^\infty\frac{x^{2s-4ar_x+2r_x}y^{-4ar_y+2r_y}}{(1+\lambda x^{2r_x}y^{2r_y})^2}dxdy$$

$$\lesssim\int_1^\infty\int_1^\infty\left(1+\lambda t_1^{2r_x/(1+2(1-2a)r_x+2s)}t_2^{2r_y/(1+2(1-2a)r_y)}\right)^{-2}dt_1dt_2 \tag{52}$$

$$\lesssim\lambda^{-1+2a-((2s+1)/2r_x\vee 1/2r_y)}\log(1/\lambda),$$

where $\log(1/\lambda)$ is not necessary when $r_x\neq(2s+1)r_y$. Combining the above three inequalities implies

$$\frac{1}{nm_y}\sum_{u_1,v_1}\frac{m_xu_1^{-2r_x}+u_1^{2s-2r_x}}{m_x(u_1^{-2r_x}v_1^{-2r_y}+\lambda)^2u_1^{4ar_x}v_1^{2r_y+4ar_y}}\sum_{u_2,v_2}\frac{m_xu_2^{-2r_x}+u_2^{2s-2r_x}}{m_xu_2^{-4br_x}v_2^{2r_y-4br_y}}$$

$$\lesssim\frac{\lambda^{-1+2a-(1/2r_x\vee 1/2r_y)}\log(1/\lambda)}{nm_y}+\frac{\lambda^{-1+2a-((2s+1)/2r_x\vee 1/2r_y)}\log(1/\lambda)}{nm_xm_y}, \tag{53}$$

where the first $\log(1/\lambda)$ is not necessary when $r_x \neq r_y$ and the second $\log(1/\lambda)$ is not necessary when $r_x \neq (2s+1)r_y$.

Now we bound the last line of (50). First, we have

$$\begin{aligned}
&\sum_{u_1,u_2,v} \frac{u_1^{-2r_x} u_2^{-2r_x}}{(u_1^{-2r_x} v^{-2r_y} + \lambda)^2 u_1^{4ar_x} u_2^{-4br_x} v^{4(a-b+1)r_y}} \lesssim \sum_{u_1,v} \frac{u_1^{-4ar_x+2r_x} v^{-4(a-b)r_y}}{(1+\lambda u_1^{2r_x} v^{2r_y})^2} \\
&\lesssim \int_1^\infty \int_1^\infty \frac{x^{-4ar_x+2r_x}}{(1+\lambda x^{2r_x} z^{2r_y})^2 z^{4(a-b)r_y}} dxdz \\
&\lesssim \lambda^{-1+2a-1/2r_x} \int_0^\infty \frac{t^{-2a+1/2r_x}}{(1+t)^2} dt \int_1^\infty z^{-r_y/r_x - 2(1-2b)r_y} dz \\
&\lesssim \lambda^{-1+2a-1/2r_x},
\end{aligned}$$

where the first inequality uses the condition $4br_x - 2r_x + 2s < -1$ which implies that the summation series on u_2 converges, and the third inequality sets $t = \lambda x^{2r_x} z^{2r_y}$. Similarly, we have

$$\begin{aligned}
&\sum_{u_1,u_2,v} \frac{u_1^{-2r_x} u_2^{2s-2r_x}}{(u_1^{-2r_x} v^{-2r_y} + \lambda)^2 u_1^{4ar_x} u_2^{-4br_x} v^{4(a-b+1)r_y}} \\
&\qquad \lesssim \sum_{u_1,v} \frac{u_1^{-4ar_x+2r_x} v^{-4(a-b)r_y}}{(1+\lambda u_1^{2r_x} v^{2r_y})^2} \lesssim \lambda^{-1+2a-1/2r_x}.
\end{aligned}$$

In the same fashion, we can establish that

$$\begin{aligned}
&\sum_{u_1,u_2,v} \frac{u_1^{2s-2r_x} u_2^{-2r_x}}{(u_1^{-2r_x} v^{-2r_y} + \lambda)^2 u_1^{4ar_x} u_2^{-4br_x} v^{4(a-b+1)r_y}} \lesssim \sum_{u_1,v} \frac{u_1^{2s-4ar_x+2r_x} v^{-4(a-b)r_y}}{(1+\lambda u_1^{2r_x} v^{2r_y})^2} \\
&\lesssim \iint \frac{x^{2s-4ar_x+2r_x}}{(1+\lambda x^{2r_x} z^{2r_y})^2 z^{4(a-b)r_y}} dxdz \lesssim \lambda^{-1+2a-(2s+1)/2r_x},
\end{aligned}$$

and

$$\sum_{u_1,u_2,v} \frac{u_1^{2s-2r_x} u_2^{2s-2r_x}}{(u_1^{-2r_x} v^{-2r_y} + \lambda)^2 u_1^{4ar_x} u_2^{-4br_x} v^{4(a-b+1)r_y}} \lesssim \lambda^{-1+2a-(2s+1)/2r_x}.$$

The above four results bound the last line of (50) by

$$\begin{aligned}
&\sum_{u_1,u_2,v} \frac{m_x^2 u_1^{-2r_x} u_2^{-2r_x} + m_x(u_1^{2s-2r_x} u_2^{-2r_x} + u_1^{-2r_x} u_2^{2s-2r_x}) + u_1^{2s-2r_x} u_2^{2s-2r_x}}{nm_x^2 (u_1^{-2r_x} v^{-2r_y} + \lambda)^2 u_1^{4ar_x} u_2^{-4br_x} v^{4(a-b+1)r_y}} \\
&\qquad \lesssim n^{-1}\lambda^{-1+2a-1/2r_x} + (nm_x)^{-1}\lambda^{-1+2a-(2s+1)/2r_x}.
\end{aligned} \tag{54}$$

Plugging the (53) and (54) into (50), we obtain

$$\begin{aligned}
&\sum_{u_1,v_1}\sum_{u_2,v_2} \gamma_{u_1v_1}^{2a} (\gamma_{u_1v_1} + \lambda)^{-2} \gamma_{u_2v_2}^{-2b} \mathbb{E}\left[\left\langle \varphi_{u_1v_1}, (\widehat{\Pi} - \Pi)\varphi_{u_2v_2} \right\rangle_{\mathcal{L}^2}^2\right] \\
&\lesssim n^{-1}\lambda^{-1+2a-1/2r_x} + (nm_x)^{-1}\lambda^{-1+2a-(2s+1)/2r_x} + (nm_y)^{-1}\lambda^{-1+2a-(1/2r_x \vee 1/2r_y)} \log(1/\lambda) \\
&\quad + (nm_x m_y)^{-1}\lambda^{-1+2a-((2s+1)/2r_x \vee 1/2r_y)} \log(1/\lambda) \\
&\lesssim n^{-1}\lambda^{-1+2a-1/2r_x} + (nm_x)^{-1}\lambda^{-1+2a-(2s+1)/2r_x} \log(1/\lambda) + (nm_y)^{-1}\lambda^{-1+2a-1/2r_y} \log(1/\lambda),
\end{aligned}$$

where the first $\log$ term in the last line can be removed when $r_x \neq (2s+1)r_y$ and the second $\log$ term can be removed when $r_x \neq r_y$. Then the desired result follows. $\square$

LEMMA S.2.7. In the function-on-function regression problem, recalling that

$$g_{nm} = \frac{1}{nm_y}\sum_{i=1}^{n}\sum_{l=1}^{m_y}(U_{il}+e_{il})\frac{1}{m_x}\sum_{j=1}^{m_x}X_{ij}L_K^{-1/2}K(S_{il},T_{ij};\cdot,\cdot).$$

for any $\nu \in [0,1/2]$, we have

$$\begin{aligned}\|\Pi^\nu(\Pi+\lambda I)^{-1}g_{nm}\|_{\mathcal{L}^2}^2 = O_p\big(&n^{-1}\lambda^{-1+2\nu-1/2r_x} + (nm_x)^{-1}\lambda^{-1+2\nu-(2s+1)/2r_x}\log(1/\lambda)\\ &+ (nm_y)^{-1}\lambda^{-1+2\nu-1/2r_y}\log(1/\lambda)\big),\end{aligned}$$

where the first $\log$ term can be removed when $r_x \neq (2s+1)r_y$ and the second $\log$ term can be removed when $r_x \neq r_y$.

PROOF OF LEMMA S.2.7. Note that

$$\begin{aligned}\|\Pi^\nu(\Pi+\lambda I)^{-1}g_{nm}\|_{\mathcal{L}^2}^2 &= \sum_{u,v}\langle\Pi^\nu(\Pi+\lambda I)^{-1}g_{nm},\varphi_{uv}\rangle_{\mathcal{L}^2}^2\\ &= \sum_{u,v}\frac{\gamma_{uv}^{2\nu}}{(\gamma_{uv}+\lambda)^2}\langle g_{nm},\varphi_{uv}\rangle_{\mathcal{L}^2}^2.\end{aligned}$$

Recalling the definition of g_{nm} in the function-on-function regression problem, we have

$$\langle g_{nm},\varphi_{uv}\rangle_{\mathcal{L}^2} = \frac{1}{nm_y}\sum_{i=1}^{n}\sum_{l=1}^{m_y}(U_{il}+e_{il})\frac{1}{m_x}\sum_{j=1}^{m_x}X_{ij}(L_K^{1/2}\varphi_{uv})(S_{il},T_{ij}).$$

Thus, we have

$$\begin{aligned}\mathbb{E}[\langle g_{nm},\varphi_{uv}\rangle_{\mathcal{L}^2}^2] &= \mathrm{Var}\left(\frac{1}{nm_y}\sum_{i=1}^{n}\sum_{l=1}^{m_y}(U_{il}+e_{il})\frac{1}{m_x}\sum_{j=1}^{m_x}X_{ij}(L_K^{1/2}\varphi_{uv})(S_{il},T_{ij})\right)\\ &= \frac{1}{n}\mathrm{Var}\left(\frac{1}{m_y}\sum_{l=1}^{m_y}(U_{il}+e_{il})\frac{1}{m_x}\sum_{j=1}^{m_x}X_{ij}(L_K^{1/2}\varphi_{uv})(S_{il},T_{ij})\right).\end{aligned}$$

Because $(L_K^{1/2}\varphi_{uv})(S_{il},T_{ij}) = (L_{K^y}^{1/2}\varphi_v^y)(S_{il})(L_{K^x}^{1/2}\varphi_u^x)(T_{ij}) = (\gamma_v^y)^{1/2}\varphi_v^y(S_{il})(L_{K^x}^{1/2}\varphi_u^x)(T_{ij})$, we have

$$\begin{aligned}\mathbb{E}[\langle g_{nm},\varphi_{uv}\rangle_{\mathcal{L}^2}^2] &= \frac{\gamma_v^y}{n}\mathrm{Var}\left(\frac{1}{m_y}\sum_{l=1}^{m_y}(U_{il}+e_{il})\frac{1}{m_x}\sum_{j=1}^{m_x}X_{ij}\varphi_v^y(S_{il})(L_{K^x}^{1/2}\varphi_u^x)(T_{ij})\right)\\ &\le \frac{\gamma_v^y}{n}\mathbb{E}\left[\left(\frac{1}{m_y}\sum_{l=1}^{m_y}(U_{il}+e_{il})\varphi_v^y(S_{il})\frac{1}{m_x}\sum_{j=1}^{m_x}X_{ij}(L_{K^x}^{1/2}\varphi_u^x)(T_{ij})\right)^2\right] \qquad (55)\\ &= \frac{\gamma_v^y}{n}\mathbb{E}\left[\left(\frac{1}{m_y}\sum_{l=1}^{m_y}(U_{il}+e_{il})\varphi_v^y(S_{il})\right)^2\right]\mathbb{E}\left[\left(\frac{1}{m_x}\sum_{j=1}^{m_x}X_{ij}(L_{K^x}^{1/2}\varphi_u^x)(T_{ij})\right)^2\right].\end{aligned}$$

For the first factor in the last line of (55),

$$\mathbb{E}\left[\left(\frac{1}{m_y}\sum_{l=1}^{m_y}(U_{il}+e_{il})\varphi_v^y(S_{il})\right)^2\right] \lesssim \frac{1}{m_y} + \langle \varphi_v^y, L_{C^U}\varphi_v^y\rangle_{\mathcal{L}^2}$$

where C^U is the covariance function of U, i.e. $C^U(s_1,s_2)=\mathbb{E}[U(s_1)U(s_2)]$. For the second term in the last line of (55), a similar analysis with (44) implies

$$\mathbb{E}\left[\left(\frac{1}{m_x}\sum_{j=1}^{m_x}X_{ij}(L_{K^x}^{1/2}\varphi_u^x)(T_{ij})\right)^2\right] \lesssim u^{-2r_x} + \frac{1}{m_x}u^{2s-2r_x}.$$

Then, plugging the above results into (55) yields

$$\mathbb{E}[\langle g_{nm},\varphi_{uv}\rangle_{\mathcal{L}^2}^2] \lesssim \frac{v^{-2r_y}}{n}(u^{-2r_x}+\frac{1}{m_x}u^{2s-2r_x})(\frac{1}{m_y}+\langle \varphi_v^y, L_{C^U}\varphi_v^y\rangle_{\mathcal{L}^2})$$

Therefore,

$$\mathbb{E}\left[\|\Pi^\nu(\Pi+\lambda I)^{-1}g_{nm}\|_{\mathcal{L}^2}^2\right] = \sum_{u,v}\frac{\gamma_{uv}^{2\nu}}{(\gamma_{uv}+\lambda)^2}\mathbb{E}[\langle g_{nm},\varphi_{uv}\rangle_{\mathcal{L}^2}^2]$$

$$\lesssim \frac{1}{n}\sum_{u,v}\frac{u^{-4\nu r_x}v^{-4\nu r_y-2r_y}}{(u^{-2r_x}v^{-2r_y}+\lambda)^2}(u^{-2r_x}+\frac{1}{m_x}u^{2s-2r_x})(\frac{1}{m_y}+\langle \varphi_v^y, L_{C^U}\varphi_v^y\rangle_{\mathcal{L}^2}) \tag{56}$$

$$\lesssim \frac{1}{n}\sum_{u,v}\frac{u^{-4\nu r_x+2r_x}v^{-4\nu r_y+2r_y}}{(1+\lambda u^{2r_x}v^{2r_y})^2}(1+\frac{1}{m_x}u^{2s})(\frac{1}{m_y}+\langle \varphi_v^y, L_{C^U}\varphi_v^y\rangle_{\mathcal{L}^2}).$$

It is easy to see that

$$\sum_{u,v}\frac{u^{-4\nu r_x+2r_x}v^{-4\nu r_y+2r_y}}{(1+\lambda u^{2r_x}v^{2r_y})^2}(1+\frac{1}{m_x}u^{2s})\langle \varphi_v^y, L_{C^U}\varphi_v^y\rangle_{\mathcal{L}^2}$$

$$\le \sum_{u,v}\frac{\lambda^{2\nu-1}}{(1+\lambda u^{2r_x}v^{2r_y})^{1+2\nu}}(1+\frac{1}{m_x}u^{2s})\langle \varphi_v^y, L_{C^U}\varphi_v^y\rangle_{\mathcal{L}^2}$$

$$\le \sum_{u}\frac{\lambda^{2\nu-1}}{(1+\lambda u^{2r_x})^{1+2\nu}}(1+\frac{1}{m_x}u^{2s})\sum_v\langle \varphi_v^y, L_{C^U}\varphi_v^y\rangle_{\mathcal{L}^2}$$

$$\lesssim \lambda^{2\nu-1-1/2r_x}+\frac{1}{m_x}\lambda^{2\nu-1-(2s+1)/2r_x},$$

where the first inequality is due to $1-2\nu\ge 0$, the second inequality uses the fact $v\ge 1$, the last inequality is because the trace $\sum_v\langle \varphi_v^y, L_{C^U}\varphi_v^y\rangle_{\mathcal{L}^2}<\infty$, and

$$\sum_u\frac{1}{(1+\lambda u^{2r_x})^{1+2\nu}} \lesssim \int_0^\infty\frac{1}{(1+\lambda x^{2r_x})^{1+2\nu}}dx = \int_0^\infty\frac{\lambda^{-1/2r_x}}{(1+y^{2r_x})^{1+2\nu}}dy \lesssim \lambda^{-1/2r_x}$$

and

$$\sum_u\frac{u^{2s}}{(1+\lambda u^{2r_x})^{1+2\nu}} \lesssim \int_0^\infty\frac{x^{2s}}{(1+\lambda x^{2r_x})^{1+2\nu}}dx = \int_0^\infty\frac{\lambda^{-(2s+1)/2r_x}y^{2s}}{(1+y^{2r_x})^{1+2\nu}}dy \lesssim \lambda^{-(2s+1)/2r_x}.$$

Additionally, (51) and (52) yield

$$\sum_{u,v}\frac{u^{-4\nu r_x+2r_x}v^{-4\nu r_y+2r_y}}{(1+\lambda u^{2r_x}v^{2r_y})^2} \lesssim \lambda^{-1+2\nu-(1/2r_x\vee 1/2r_y)}\log(1/\lambda),$$

and

$$\sum_{u,v} \frac{u^{2s-4\nu r_x+2r_x} v^{-4\nu r_y+2r_y}}{(1+\lambda u^{2r_x} v^{2r_y})^2} \lesssim \lambda^{-1+2\nu-((2s+1)/2r_x \vee 1/2r_y)} \log(1/\lambda).$$

where the first log term can be removed if $r_x \neq r_y$ and the second $\log(1/\lambda)$ is also not necessary when $r_x \neq (2s+1)r_y$. Plugging the above results into (56), we have

$$\begin{aligned}
&\mathbb{E}\left[\|\Pi^\nu(\Pi+\lambda I)^{-1} g_{nm}\|_{\mathcal{L}^2}^2\right] \\
&\lesssim n^{-1}\lambda^{-1+2\nu-1/2r_x} + (nm_x)^{-1}\lambda^{-1+2\nu-(2s+1)/2r_x}\log(1/\lambda) + (nm_y)^{-1}\lambda^{-1+2\nu-1/2r_y}\log(1/\lambda),
\end{aligned}$$

where the first log term can be removed when $r_x \neq (2s+1)r_y$ and the second log term can be removed when $r_x \neq r_y$. Then the desired result follows. □

LEMMA S.2.8. Under the same conditions of Theorem 3.3, for any $0 \le \nu$ satisfying $4\nu r_x - 2r_x + 2s < -1$, $4\nu r_y - 2r_y < -1$ and $2\nu > \alpha$, we have

$$\begin{aligned}
\|\Pi^\nu(\hat{f}-\bar{f})\|_{\mathcal{L}^2}^2 = O_p\big(&n^{-1}\lambda^{-1+2\nu-1/2r_x} + (nm_x)^{-1}\lambda^{-1+2\nu-(2s+1)/2r_x}\log(1/\lambda) \\
&+ (nm_y)^{-1}\lambda^{-1+2\nu-1/2r_y}\log(1/\lambda)\big),
\end{aligned}$$

where the first log term can be removed when $r_x \neq (2s+1)r_y$ and the second log term can be removed when $r_x \neq r_y$.

PROOF OF LEMMA S.2.8. From the same analysis as (9), we obtain

$$\begin{aligned}
\|\Pi^\nu(\hat{f}-\bar{f})\|_{\mathcal{L}^2}^2 \lesssim &\|\Pi^\nu(\Pi+\lambda I)^{-1}(\Pi-\widehat{\Pi})(\hat{f}-\bar{f})\|_{\mathcal{L}^2}^2 \\
&+ \|\Pi^\nu(\Pi+\lambda I)^{-1}(\widetilde{\Pi}-\Pi)f_0\|_{\mathcal{L}^2}^2 \\
&+ \|\Pi^\nu(\Pi+\lambda I)^{-1}(\widehat{\Pi}-\Pi)\bar{f}\|_{\mathcal{L}^2}^2 \\
&+ \|\Pi^\nu(\Pi+\lambda I)^{-1} g_{nm}\|_{\mathcal{L}^2}^2.
\end{aligned} \tag{57}$$

We can bound the first term as

$$\begin{aligned}
&\|\Pi^\nu(\Pi+\lambda I)^{-1}(\Pi-\widehat{\Pi})(\hat{f}-\bar{f})\|_{\mathcal{L}^2}^2 \\
&\quad\lesssim \|\Pi^\nu(\Pi+\lambda I)^{-1}(\Pi-\widehat{\Pi})\Pi^{-\nu}\|_{op}^2 \|\Pi^\nu(\hat{f}-\bar{f})\|_{\mathcal{L}^2}^2 \\
&\quad\lesssim o_p(1)\|\Pi^\nu(\hat{f}-\bar{f})\|_{\mathcal{L}^2}^2,
\end{aligned}$$

where the last inequality follows from Lemma S.2.6, given $0 \le \nu$, $4\nu r_x - 2r_x + 2s < -1$, $4\nu r_y - 2r_y < -1$. These conditions imply $0 \le \nu < 1/2$ and $-4\nu r_x - 2r_x + 2s < -1$. Assumption 3(c) supplies the conditions in Lemma S.2.6 for $b = \alpha/2$. Hence all subsequent applications of Lemmas S.2.6 and S.2.7 in this proof satisfy their stated parameter restrictions. Next, we bound the second and third terms

$$\begin{aligned}
&\|\Pi^\nu(\Pi+\lambda I)^{-1}(\widetilde{\Pi}-\Pi)f_0\|_{\mathcal{L}^2}^2 \\
&\lesssim \|\Pi^\nu(\Pi+\lambda I)^{-1}(\widetilde{\Pi}-\Pi)\Pi^{-\alpha/2}\|_{op}^2 \|\Pi^{\alpha/2}f_0\|_{\mathcal{L}^2}^2 \\
&= O_p(n^{-1}\lambda^{-1+2\nu-1/2r_x} + (nm_x)^{-1}\lambda^{-1+2\nu-(2s+1)/2r_x}\log(1/\lambda) + (nm_y)^{-1}\lambda^{-1+2\nu-1/2r_y}\log(1/\lambda))
\end{aligned}$$

and

$$\begin{aligned}
&\|\Pi^\nu(\Pi+\lambda I)^{-1}(\widehat{\Pi}-\Pi)\bar{f}\|_{\mathcal{L}^2}^2 \\
&\lesssim \|\Pi^\nu(\Pi+\lambda I)^{-1}(\widehat{\Pi}-\Pi)\Pi^{-\alpha/2}\|_{op}^2 \Big(\|\Pi^{\alpha/2}(\bar{f}-f_0)\|_{\mathcal{L}^2} + \|\Pi^{\alpha/2}f_0\|_{\mathcal{L}^2}\Big)^2 \\
&= O_p(n^{-1}\lambda^{-1+2\nu-1/2r_x} + (nm_x)^{-1}\lambda^{-1+2\nu-(2s+1)/2r_x}\log(1/\lambda) + (nm_y)^{-1}\lambda^{-1+2\nu-1/2r_y}\log(1/\lambda)),
\end{aligned}$$

using Lemmas S.2.6 and S.2.5. Finally, Lemma S.2.7 implies

$$\|\Pi^{\nu}(\Pi+\lambda I)^{-1}g_{nm}\|_{\mathcal{L}^2}^2$$
$$= O_p(n^{-1}\lambda^{-1+2\nu-1/2r_x} + (nm_x)^{-1}\lambda^{-1+2\nu-(2s+1)/2r_x}\log(1/\lambda) + (nm_y)^{-1}\lambda^{-1+2\nu-1/2r_y}\log(1/\lambda)).$$

Combining the above results, the desired conclusion follows. □

LEMMA S.2.9. Given $\lambda \in (0,1)$ and positive constants $a, b > 1$,

$$\int_1^{\infty}\int_1^{\infty}\frac{1}{1+\lambda x^a y^b}dxdy \le C\begin{cases}\lambda^{-1/b} & \text{if } a>b,\\ \lambda^{-1/b}\big(1+\log(1/\lambda)\big) & \text{if } a=b,\\ \lambda^{-1/a} & \text{if } a<b,\end{cases}$$

where $C = C(a,b)$ is a constant depending on a and b but independent of λ.

PROOF OF LEMMA S.2.9. When $a > b$, let $t = \lambda x^a y^b$, then

$$\int_1^{\infty}\int_1^{\infty}\frac{1}{1+\lambda x^a y^b}dxdy = b^{-1}\lambda^{-1/b}\int_{\lambda}^{\infty}\frac{t^{-(b-1)/b}}{1+t}dt\int_1^{(t/\lambda)^{1/a}}x^{-a/b}dx$$
$$\lesssim \lambda^{-1/b}\int_{\lambda}^{\infty}\frac{t^{-(b-1)/b}}{1+t}dt \lesssim \lambda^{-1/b}.$$

When $a = b$, we have

$$\int_1^{\infty}\int_1^{\infty}\frac{1}{1+\lambda x^a y^b}dxdy = b^{-1}\lambda^{-1/b}\int_{\lambda}^{\infty}\frac{t^{-(b-1)/b}}{1+t}dt\int_1^{(t/\lambda)^{1/a}}x^{-a/b}dx$$
$$\lesssim \lambda^{-1/b}\int_{\lambda}^{\infty}\frac{t^{-(b-1)/b}}{1+t}\log(t/\lambda)dt \lesssim \lambda^{-1/b}\big(1+\log(1/\lambda)\big).$$

When $a < b$, similar to the case when $a > b$, we have

$$\int_1^{\infty}\int_1^{\infty}\frac{1}{1+\lambda x^a y^b}dxdy \lesssim \lambda^{-1/a}.$$

The desired result follows. □

LEMMA S.2.10. Given $\lambda \in (0,1)$ and positive constants $a, b > 1/2$,

$$\int_1^{\infty}\int_1^{\infty}\frac{1}{(1+\lambda x^a y^b)^2}dxdy \le C\begin{cases}\lambda^{-1/b} & \text{if } a>b,\\ \lambda^{-1/b}\big(1+\log(1/\lambda)\big) & \text{if } a=b,\\ \lambda^{-1/a} & \text{if } a<b,\end{cases}$$

where $C = C(a,b)$ is a constant depending on a and b but independent of λ.

PROOF OF LEMMA S.2.10. When $a > b$, let $t = \lambda x^a y^b$, then

$$\int_1^{\infty}\int_1^{\infty}\frac{1}{(1+\lambda x^a y^b)^2}dxdy = b^{-1}\lambda^{-1/b}\int_{\lambda}^{\infty}\frac{t^{-(b-1)/b}}{(1+t)^2}dt\int_1^{(t/\lambda)^{1/a}}x^{-a/b}dx$$
$$\lesssim \lambda^{-1/b}\int_{\lambda}^{\infty}\frac{t^{-(b-1)/b}}{(1+t)^2}dt \lesssim \lambda^{-1/b}.$$

When $a = b$, we have

$$\int_1^\infty \int_1^\infty \frac{1}{(1+\lambda x^a y^b)^2} dxdy = b^{-1}\lambda^{-1/b} \int_\lambda^\infty \frac{t^{-(b-1)/b}}{(1+t)^2} dt \int_1^{(t/\lambda)^{1/a}} x^{-a/b} dx$$

$$\lesssim \lambda^{-1/b} \int_\lambda^\infty \frac{t^{-(b-1)/b}}{(1+t)^2} \log(t/\lambda) dt \lesssim \lambda^{-1/b}\big(1+\log(1/\lambda)\big).$$

When $a < b$, similar to the case when $a > b$, we have

$$\int_1^\infty \int_1^\infty \frac{1}{(1+\lambda x^a y^b)^2} dxdy \lesssim \lambda^{-1/a}.$$

The desired result follows. □

S.3. Additional numerical results. In this section, we present additional simulation results that were omitted in Section 4. Specifically, Tables S.1 and S.2 report the results under the Legendre-polynomial design for the scalar-on-function regression problem in Section 4.1 and the function-on-function regression problem in Section 4.2, respectively. The parameter setting is kept the same as in the main text, while the predictor process is generated from the shifted Legendre polynomial basis instead of the cosine basis.

TABLE S.1

The Monte Carlo averages and standard deviation (in parentheses) of the error $\mathcal{E}_{\beta_0}$ of the proposed method and the other five benchmark methods, for scalar-on-function regression with Legendre design.

n	m	Proposed	FullyRKHS	Plug-in	IN	PACE	MC
50	3	**0.5969** (0.3908)	1.1927 (0.6283)	0.9754 (0.9018)	2.2280 (1.9076)	1.8539 (3.0333)	1.4595 (0.8550)
50	6	**0.2932** (0.2061)	0.6151 (0.3357)	0.4728 (0.4107)	0.8277 (0.6169)	0.5080 (1.5598)	0.6865 (0.3770)
50	10	**0.1626** (0.1285)	0.2938 (0.1969)	0.3563 (0.2792)	0.4542 (0.2868)	0.3313 (0.9622)	0.3272 (0.2230)
50	20	0.1102 (0.0896)	0.1471 (0.0959)	0.1846 (0.1389)	0.1876 (0.1338)	**0.1067** (0.0768)	0.1441 (0.0895)
100	3	**0.4484** (0.3037)	1.0695 (0.4663)	0.5588 (0.4510)	1.8206 (1.4538)	1.6142 (3.3726)	1.1916 (0.4659)
100	6	**0.1705** (0.1287)	0.4623 (0.1873)	0.3291 (0.2174)	0.6313 (0.2970)	0.3981 (1.3964)	0.4817 (0.1801)
100	10	**0.1017** (0.0839)	0.2553 (0.1373)	0.2166 (0.1468)	0.3481 (0.1943)	0.1554 (0.1237)	0.2627 (0.1395)
100	20	**0.0592** (0.0486)	0.1198 (0.0577)	0.1180 (0.1052)	0.1629 (0.0992)	0.0838 (0.0740)	0.1254 (0.0589)
200	3	**0.3381** (0.2111)	0.9882 (0.2533)	0.4147 (0.2945)	1.8225 (1.1086)	0.6416 (2.0476)	1.0520 (0.2777)
200	6	**0.0949** (0.0732)	0.3921 (0.1147)	0.2626 (0.1755)	0.6724 (0.2279)	0.2022 (0.1409)	0.4039 (0.1147)
200	10	**0.0659** (0.0615)	0.2241 (0.0798)	0.1610 (0.1130)	0.3028 (0.1194)	0.1212 (0.0910)	0.2284 (0.0850)
200	20	**0.0330** (0.0204)	0.1021 (0.0390)	0.0681 (0.0632)	0.1307 (0.0734)	0.0481 (0.0500)	0.0989 (0.0378)

TABLE S.2

The Monte Carlo averages and standard deviation (in parentheses) of the error $\mathcal{E}_{\beta_0}$ of the proposed method and the other five benchmark methods, for function-on-function regression with Legendre design.

n	m_x	m_y	Proposed	OPFFR-S	IN	PACE	MC	RKHSPCA
50	3	3	**1.2003** (2.5519)	6.0646 (8.7640)	2.5135 (1.3869)	2.2826 (2.6002)	2.2597 (1.2515)	1.3592 (1.8245)
50	6	6	**0.5433** (0.4860)	9.3907 (25.9140)	1.1715 (0.5343)	2.7042 (9.2213)	1.1683 (0.5856)	0.7611 (0.5451)
50	10	10	**0.4137** (0.2068)	2.2702 (13.4549)	0.7226 (0.2842)	0.7723 (1.1082)	0.7987 (0.3691)	0.6978 (0.6102)
50	20	20	**0.2698** (0.1520)	0.3657 (0.1682)	0.4631 (0.1729)	0.2988 (0.1245)	0.5358 (0.2324)	0.5517 (0.2722)
100	3	3	**0.6563** (0.8342)	8.0053 (19.9908)	2.1049 (0.8571)	2.4663 (3.7791)	1.7488 (1.1667)	0.8208 (0.6584)
100	6	6	**0.3775** (0.1929)	3.0308 (5.8192)	0.9577 (0.3038)	1.1401 (1.9415)	0.8239 (0.2836)	0.4948 (0.2325)
100	10	10	**0.2552** (0.1012)	0.5046 (0.3055)	0.5533 (0.1736)	0.3901 (0.1719)	0.6004 (0.1870)	0.4882 (0.2154)
100	20	20	**0.1696** (0.0942)	0.2377 (0.1202)	0.3416 (0.1354)	0.1837 (0.0786)	0.3780 (0.1276)	0.4248 (0.1329)
200	3	3	**0.4490** (0.2876)	7.5340 (28.9551)	1.8989 (0.6774)	1.1957 (2.3275)	1.3112 (0.3870)	0.5562 (0.2709)
200	6	6	**0.3305** (0.5559)	1.6453 (2.6825)	0.8008 (0.2395)	0.5699 (0.5154)	0.6995 (0.2131)	0.4275 (0.1294)
200	10	10	**0.1437** (0.0786)	0.3183 (0.2444)	0.4693 (0.1466)	0.2592 (0.1423)	0.5038 (0.1827)	0.3925 (0.1400)
200	20	20	**0.0929** (0.0515)	0.1572 (0.0769)	0.2374 (0.0975)	0.1198 (0.0555)	0.2754 (0.1143)	0.2915 (0.0998)

REFERENCES

CAI, T. T. and YUAN, M. (2012). Minimax and adaptive prediction for functional linear regression. *Journal of the American Statistical Association* **107** 1201–1216.

TSYBAKOV, A. B. and ZAIATS, V. (2009). *Introduction to nonparametric estimation* **11**. Springer.