

MINIMAX AND ADAPTIVE COVARIANCE FUNCTION ESTIMATION IN FUNCTIONAL DATA ANALYSIS

T. Tony Cai, Yicheng Li, and Ming Yuan

University of Pennsylvania and East China Normal University

Columbia University

Supplementary Material

This supplementary material contains technical proofs and additional numerical results for “Minimax and Adaptive Covariance Function Estimation in Functional Data Analysis.”

S1. Proofs of Upper Bounds

We prove the main results on the upper bounds in this section. For brevity, we shall consider here only the case when $\mu_0(\cdot)$ is known and (2.5) is employed in defining $\hat{C}_\lambda$, and we may further assume that $\mu_0(\cdot) = 0$ without loss of generality. The effect of this assumption in establishing the convergence rate of $\hat{C}_\lambda$ is negligible because of the condition (3.8) or (3.10), which essentially controls the error term that may come with using an estimated mean function instead of the true mean function. The proofs follow from

the same line of argument when using $\hat{\mu}$ but get considerably more tedious. For simplicity of notation, we denote $\mathcal{W}_2^{\alpha,\alpha} = \mathcal{W}_2^\alpha(\mathcal{T}) \otimes \mathcal{W}_2^\alpha(\mathcal{T})$ throughout this section. Write $C, c, \dots$ for generic positive constants whose values may change from line to line.

S1.1 Basis Expansion

We first introduce some notation and preliminary results that will be used in the proofs. It is well known that the Sobolev space $\mathcal{W}_2^\alpha(\mathcal{T})$ is a reproducing kernel Hilbert space (RKHS) with some reproducing kernel $K(\cdot, \cdot)$. Then, Mercer's theorem states that K admits the following eigenvalue decomposition:

$$K(s, t) = \sum_{k \geq 1} \rho_k \varphi_k(s) \varphi_k(t) \quad (\text{S1.1})$$

where the nonnegative scalars $\rho_1 \geq \rho_2 \geq \dots$ are the eigenvalues and the eigenfunctions $\{\varphi_k(\cdot) : k \geq 1\}$ form an orthonormal basis of $L_2(\mathcal{T})$, i.e.,

$$\langle \varphi_j, \varphi_k \rangle := \int_{\mathcal{T}} \varphi_j(t) \varphi_k(t) dt = \delta_{jk}, \quad (\text{S1.2})$$

where δ is Kronecker's delta. For the Sobolev space $\mathcal{W}_2^\alpha(\mathcal{T})$, these eigenvalues satisfy $\rho_k \asymp k^{-2\alpha}$.

Since the sample path of X belongs to the RKHS $\mathcal{W}_2^\alpha(\mathcal{T})$ almost surely,

we can write

$$X(\cdot) = \sum_{k \geq 1} x_k \varphi_k(\cdot) \quad (\text{S1.3})$$

where $x_k = \langle X, \varphi_k \rangle_{L_2}$ is the Fourier coefficient of X with respect to $\{\varphi_k(\cdot) : k \geq 1\}$. It is worth pointing out that, despite the resemblance, the expression given by (S1.3) differs from the usual Karhunen-Loève expansion. In (S1.3), the basis functions $\{\varphi_k : k \geq 1\}$ are identified with the reproducing kernel K and the function space $\mathcal{W}_2^\alpha(\mathcal{T})$, whereas the basis functions used in the Karhunen-Loève expansion are the eigenfunctions of the covariance function C_0 and are not known a priori. With the expansion (S1.3), the first two moments of X can be given in terms of those for the Fourier coefficients:

$$\begin{aligned} \mu_0(t) &:= \mathbb{E}X(t) = \sum_{k \geq 1} \mathbb{E}(x_k \varphi_k(t)); \\ C_0(s, t) &= \sum_{j, k \geq 1} \text{Cov}(x_k, x_j) \varphi_j(s) \varphi_k(t). \end{aligned}$$

The above expression naturally identifies C_0 with the tensor product space $\mathcal{W}_2^{\alpha, \alpha} := \mathcal{W}_2^\alpha(\mathcal{T}) \otimes \mathcal{W}_2^\alpha(\mathcal{T})$, an RKHS associated with reproducing kernel

$$K \otimes K((s_1, t_1), (s_2, t_2)) = K(s_1, s_2) K(t_1, t_2). \quad (\text{S1.4})$$

With slight abuse of notation, let the operator $\otimes$ also denote the tensor

product of real-valued functions on $\mathcal{T}$, i.e.,

$$f_1 \otimes f_2(s, t) = f_1(s)f_2(t). \quad (\text{S1.5})$$

Then the tensor product space $\mathcal{W}_2^{\alpha, \alpha}$ can be characterized as follows. If $f_1, f_2 \in \mathcal{W}_2^\alpha(\mathcal{T})$, then $f_1 \otimes f_2 \in \mathcal{W}_2^{\alpha, \alpha}$ with

$$\|f_1 \otimes f_2\|_{\mathcal{W}_2^{\alpha, \alpha}} = \|f_1\|_{\mathcal{W}_2^\alpha(\mathcal{T})} \|f_2\|_{\mathcal{W}_2^\alpha(\mathcal{T})}. \quad (\text{S1.6})$$

Linear combinations of such tensor products are dense in $\mathcal{W}_2^{\alpha, \alpha}$. The norm of a linear combination is given by

$$\left\| \sum_{j=1}^k f_{j_1} \otimes f_{j_2} \right\|_{\mathcal{W}_2^{\alpha, \alpha}}^2 = \sum_{i,j=1}^k \langle f_{i_1}, f_{j_1} \rangle_{\mathcal{W}_2^\alpha(\mathcal{T})} \langle f_{i_2}, f_{j_2} \rangle_{\mathcal{W}_2^\alpha(\mathcal{T})}. \quad (\text{S1.7})$$

Since $\{\varphi_k : k \geq 1\}$ is an orthonormal basis of $L_2(\mathcal{T})$, it follows that $\{\varphi_j(s)\varphi_k(t) : j, k \geq 1\}$ forms an orthonormal basis of $L_2(\mathcal{T} \times \mathcal{T})$. For a function $f \in L_2(\mathcal{T} \times \mathcal{T})$, write

$$f(s, t) = \sum_{j,k \geq 1} f_{jk} \varphi_j(s) \varphi_k(t). \quad (\text{S1.8})$$

Recall that $\{\rho_k : k \geq 1\}$ is the set of eigenvalues of K . It is clear (see, e.g., Wahba (1990)) that the tensor product space $\mathcal{W}_2^{\alpha, \alpha}$ contains all functions f such that

$$\|f\|_{\mathcal{W}_2^{\alpha, \alpha}}^2 := \sum_{j,k \geq 1} \rho_j^{-1} \rho_k^{-1} f_{jk}^2 < \infty. \quad (\text{S1.9})$$

For technical purposes, we further introduce a class of norms on $\mathcal{W}_2^{\alpha, \alpha}$.

To this end, recall that any $f \in \mathcal{W}_2^{\alpha, \alpha}$ can be written as

$$f(s, t) = \sum_{j, k \geq 1} f_{jk} \varphi_j(s) \varphi_k(t), \quad f_{jk} = \langle f, \varphi_j(\cdot) \varphi_k(\cdot) \rangle_{L_2}. \quad (\text{S1.10})$$

By the construction of $\{\varphi_k : k \geq 1\}$, we have

$$\|f\|_{L_2}^2 = \sum_{j, k \geq 1} f_{jk}^2, \quad \text{and} \quad \|f\|_{\mathcal{W}_2^{\alpha, \alpha}}^2 = \sum_{j, k \geq 1} \gamma_{jk}^{-1} f_{jk}^2, \quad (\text{S1.11})$$

where $\gamma_{jk} = \rho_j \rho_k$. Define, for $0 \leq a \leq 1$,

$$\|f\|_a^2 = \sum_{j, k \geq 1} \gamma_{jk}^{-a} f_{jk}^2. \quad (\text{S1.12})$$

It is clear that $\|\cdot\|_0 = \|\cdot\|_{L_2}$ and $\|\cdot\|_1 = \|\cdot\|_{\mathcal{W}_2^{\alpha, \alpha}}$.

S1.2 Proof of Proposition 1

By Jensen's inequality and the Hilbert-space variance identity, $\mu_0 = \mathbb{E}X$

belongs to $\mathcal{W}_2^\alpha(\mathcal{T})$ and

$$\mathbb{E}\|X - \mu_0\|_{\mathcal{W}_2^\alpha}^2 \leq \mathbb{E}\|X\|_{\mathcal{W}_2^\alpha}^2.$$

Let $Z = X - \mu_0$ and write

$$Z(\cdot) = \sum_{k \geq 1} z_k \varphi_k(\cdot).$$

Then the covariance function admits the expansion

$$C_0(s, t) = \sum_{j, k \geq 1} \mathbb{E}(z_j z_k) \varphi_j(s) \varphi_k(t).$$

It suffices to verify that

$$\|C_0\|_{\mathcal{W}_2^{\alpha,\alpha}}^2 = \sum_{j,k \geq 1} \rho_j^{-1} \rho_k^{-1} \{\mathbb{E}(z_j z_k)\}^2 < \infty. \quad (\text{S1.13})$$

By the Cauchy-Schwarz inequality, $\{\mathbb{E}(z_j z_k)\}^2 \leq \mathbb{E}(z_j^2) \mathbb{E}(z_k^2)$. Therefore,

$$\begin{aligned} \|C_0\|_{\mathcal{W}_2^{\alpha,\alpha}}^2 &\leq \sum_{j,k \geq 1} \rho_j^{-1} \rho_k^{-1} \mathbb{E}(z_j^2) \mathbb{E}(z_k^2) \\ &= \left(\sum_{k \geq 1} \rho_k^{-1} \mathbb{E}(z_k^2) \right)^2 \\ &= \left(\mathbb{E} \|X - \mu_0\|_{\mathcal{W}_2^{\alpha}}^2 \right)^2 \leq \left(\mathbb{E} \|X\|_{\mathcal{W}_2^{\alpha}}^2 \right)^2 < \infty. \end{aligned}$$

This completes the proof of Proposition 1. $\square$

S1.3 Proof of Theorem 2

For brevity, we shall assume that T follows a uniform distribution without loss of generality. Recall that

$$\hat{C}_\lambda = \arg \min_{C \in \mathcal{W}_2^{\alpha,\alpha}} \left\{ \mathcal{L}_{n,m}(C) + \lambda \|C\|_{\mathcal{W}_2^{\alpha,\alpha}}^2 \right\} \quad (\text{S1.14})$$

where

$$\mathcal{L}_{n,m}(C) = \frac{1}{nm(m-1)} \sum_{i=1}^n \sum_{1 \leq j \neq k \leq m} (Y_{ij} Y_{ik} - C(T_{ij}, T_{ik}))^2. \quad (\text{S1.15})$$

Observe that

$$\mathcal{L}_\infty(C) := \mathbb{E} \mathcal{L}_{n,m}(C)$$

$$\begin{aligned}
 &= \frac{1}{m(m-1)} \mathbb{E} \left(\sum_{1 \leq j \neq k \leq m} [Y_{1j} Y_{1k} - C(T_{1j}, T_{1k})]^2 \right) \\
 &= \mathbb{E} ([Y_{11} Y_{12} - C_0(T_{11}, T_{12})]^2) + \mathbb{E} ([C(T_{11}, T_{12}) - C_0(T_{11}, T_{12})]^2).
 \end{aligned}$$

Write

$$\bar{C}_\lambda = \arg \min_{C \in \mathcal{W}_2^{\alpha, \alpha}} \left\{ \mathcal{L}_\infty(C) + \lambda \|C\|_{\mathcal{W}_2^{\alpha, \alpha}}^2 \right\}. \quad (\text{S1.16})$$

Then

$$\hat{C}_\lambda - C_0 = (\bar{C}_\lambda - C_0) + (\hat{C}_\lambda - \bar{C}_\lambda). \quad (\text{S1.17})$$

The two terms can be regarded as deterministic and stochastic error, respectively. Write

$$\begin{aligned}
 \mathcal{L}_{\infty, \lambda}(C) &= \mathcal{L}_\infty(C) + \lambda \|C\|_{\mathcal{W}_2^{\alpha, \alpha}}^2; \\
 \mathcal{L}_{n, m, \lambda}(C) &= \mathcal{L}_{n, m}(C) + \lambda \|C\|_{\mathcal{W}_2^{\alpha, \alpha}}^2.
 \end{aligned}$$

Denote $G_\lambda = D^2 \mathcal{L}_{\infty, \lambda}(\bar{C}_\lambda)$ and

$$\tilde{C}_\lambda = \bar{C}_\lambda - G_\lambda^{-1} D \mathcal{L}_{n, m, \lambda}(\bar{C}_\lambda), \quad (\text{S1.18})$$

where D stands for the Fréchet derivatives. It is clear that the stochastic error can be further decomposed as

$$\hat{C}_\lambda - \bar{C}_\lambda = (\hat{C}_\lambda - \tilde{C}_\lambda) + (\tilde{C}_\lambda - \bar{C}_\lambda). \quad (\text{S1.19})$$

We now bound $\bar{C}_\lambda - C_0$, $\hat{C}_\lambda - \tilde{C}_\lambda$ and $\tilde{C}_\lambda - \bar{C}_\lambda$ separately.

We first consider the deterministic error $\bar{C}_\lambda - C_0$.

Lemma 1. *There exists a constant $c > 0$ such that*

$$\|\bar{C}_\lambda - C_0\|_a^2 \leq c\lambda^{1-a}\|C_0\|_{\mathcal{W}_2^{\alpha,\alpha}}^2.$$

Proof. Write

$$C_0(s, t) = \sum_{k_1, k_2=1}^{\infty} a_{k_1 k_2} \varphi_{k_1}(s) \varphi_{k_2}(t), \quad \bar{C}_\lambda(s, t) = \sum_{k_1, k_2=1}^{\infty} \bar{b}_{k_1 k_2} \varphi_{k_1}(s) \varphi_{k_2}(t).$$

Then

$$\bar{b}_{k_1, k_2} = (1 + \lambda \gamma_{k_1 k_2}^{-1})^{-1} a_{k_1 k_2}. \quad (\text{S1.20})$$

It follows that

$$\begin{aligned} \|\bar{C}_\lambda - C_0\|_a^2 &= \sum_{k_1, k_2=1}^{\infty} \gamma_{k_1 k_2}^{-a} (\bar{b}_{k_1 k_2} - a_{k_1 k_2})^2 \\ &= \sum_{k_1, k_2=1}^{\infty} \gamma_{k_1 k_2}^{-a} \left(\frac{\lambda \gamma_{k_1 k_2}^{-1}}{1 + \lambda \gamma_{k_1 k_2}^{-1}} \right)^2 a_{k_1 k_2}^2 \\ &\leq c\lambda^2 \sup_{k_1, k_2} \frac{\gamma_{k_1 k_2}^{-(1+a)}}{(1 + \lambda \gamma_{k_1 k_2}^{-1})^2} \sum_{k_1, k_2=1}^{\infty} \gamma_{k_1 k_2}^{-1} a_{k_1 k_2}^2 \\ &\leq c\lambda^{1-a} \|C_0\|_{\mathcal{W}_2^{\alpha,\alpha}}^2. \end{aligned}$$

□

Next, we consider $\tilde{C}_\lambda - \bar{C}_\lambda$.

Lemma 2. *There exists a constant $c > 0$ such that*

$$\begin{aligned} &\mathbb{E} \left\| \tilde{C}_\lambda - \bar{C}_\lambda \right\|_a^2 \\ &\leq c \left(n^{-1} + n^{-1} m^{-1} \lambda^{-(a+1/(2\alpha))} \log(1/\lambda) \right. \\ &\quad \left. + n^{-1} \lambda^{1-(a+1/(2\alpha))} \log(1/\lambda) \right). \end{aligned} \quad (\text{S1.21})$$

Proof. By definition

$$D\mathcal{L}_{n,m,\lambda}(\bar{C}_\lambda) + D^2\mathcal{L}_{\infty,\lambda}(\bar{C}_\lambda)(\tilde{C}_\lambda - \bar{C}_\lambda) = 0. \quad (\text{S1.22})$$

Notice that

$$\begin{aligned} D\mathcal{L}_{n,m,\lambda}(\bar{C}_\lambda) &= D\mathcal{L}_{n,m,\lambda}(\bar{C}_\lambda) - D\mathcal{L}_{\infty,\lambda}(\bar{C}_\lambda) \\ &= D\mathcal{L}_{n,m}(\bar{C}_\lambda) - D\mathcal{L}_\infty(\bar{C}_\lambda). \end{aligned} \quad (\text{S1.23})$$

Therefore

$$\begin{aligned} &\mathbb{E} [D\mathcal{L}_{n,m,\lambda}(\bar{C}_\lambda)f]^2 \\ &= \mathbb{E} [D\mathcal{L}_{n,m}(\bar{C}_\lambda)f - D\mathcal{L}_\infty(\bar{C}_\lambda)f]^2 \\ &= \frac{1}{n^2m^2(m-1)^2} \text{Var} \left[\sum_{i=1}^n \sum_{1 \leq j \neq k \leq m} ([Y_{ij}Y_{ik} - \bar{C}_\lambda(T_{ij}, T_{ik})] f(T_{ij}, T_{ik})) \right] \\ &= \frac{1}{nm^2(m-1)^2} \text{Var} \left[\sum_{1 \leq j \neq k \leq m} ([Y_{1j}Y_{1k} - \bar{C}_\lambda(T_{1j}, T_{1k})] f(T_{1j}, T_{1k})) \right] \\ &= \frac{1}{nm^2(m-1)^2} \text{Var} \left[\sum_{1 \leq j \neq k \leq m} ([C_0(T_{1j}, T_{1k}) - \bar{C}_\lambda(T_{1j}, T_{1k})] f(T_{1j}, T_{1k})) \right] \\ &\quad + \frac{1}{nm^2(m-1)^2} \mathbb{E} \left[\text{Var} \left(\sum_{1 \leq j \neq k \leq m} Y_{1j}Y_{1k} f(T_{1j}, T_{1k}) \middle| T \right) \right] \end{aligned}$$

We now bound the two terms separately. The second term can be bounded using the following technical lemma which is proved in Subsection S3.2.

Lemma 3. *Let $Y_j = X(T_j) + \epsilon_j$ where X is a mean-zero second-order stochastic process whose probability law belongs to $\mathcal{P}(\alpha; M_0, c_0)$, T_j are independent uniform random variables defined on $\mathcal{T}$, and ϵ_j are indepen-*

dent measurement errors with mean zero and variance σ_0^2 . Denote, for $f, g \in L_2(\mathcal{T})$,

$$U = \sum_{1 \leq j \neq k \leq m} Y_j Y_k f(T_j) g(T_k).$$

Then

$$\mathbb{E}[\text{Var}(U|T)] \leq 4c_0 m^4 \mathbb{E} \left(\int X(s) f(s) ds \right)^2 \mathbb{E} \left(\int X(s) g(s) ds \right)^2 + O(m^3).$$

In particular, if $f = \varphi_{k_1}$ and $g = \varphi_{k_2}$, then

$$\mathbb{E}[\text{Var}(U|T)] \leq 4c_0 m^4 a_{k_1 k_1} a_{k_2 k_2} + O(m^3),$$

where $\{a_{jk} : j, k \geq 1\}$ are the Fourier coefficients of C_0 with respect to the orthonormal basis $\{\varphi_j \otimes \varphi_k : j, k \geq 1\}$.

The second term can now be bounded using Lemma 3. By taking the two univariate functions in Lemma 3 to be φ_{k_1} and φ_{k_2} , we have

$$\mathbb{E} \left[\text{Var} \left(\sum_{1 \leq j \neq k \leq m} Y_{1j} Y_{1k} f(T_{1j}, T_{1k}) \middle| T \right) \right] \leq c m^4 a_{k_1 k_1} a_{k_2 k_2} + O(m^3).$$

To bound the first term, denote by

$$V_m = \sum_{1 \leq j \neq k \leq m} ([C_0(T_{1j}, T_{1k}) - \bar{C}_\lambda(T_{1j}, T_{1k})] f(T_{1j}, T_{1k})).$$

Observe that

$$\begin{aligned} \text{Var}(V_m) &\leq \mathbb{E}(V_m^2) \\ &= \sum_{\substack{1 \leq j \neq k \leq m \\ 1 \leq j' \neq k' \leq m}} \mathbb{E} \left([C_0(T_{1j}, T_{1k}) - \bar{C}_\lambda(T_{1j}, T_{1k})] f(T_{1j}, T_{1k}) \right. \end{aligned}$$

$$\begin{aligned}
 & \times [C_0(T_{1j'}, T_{1k'}) - \bar{C}_\lambda(T_{1j'}, T_{1k'})] f(T_{1j'}, T_{1k'}) \Big) \\
 & \leq \frac{m!}{(m-4)!} \left(\int_{\mathcal{T} \times \mathcal{T}} [C_0(s, t) - \bar{C}_\lambda(s, t)] f(s, t) ds dt \right)^2 + cm^3 \\
 & \leq m^4 \int_{\mathcal{T} \times \mathcal{T}} [C_0(s, t) - \bar{C}_\lambda(s, t)]^2 ds dt \int_{\mathcal{T} \times \mathcal{T}} f^2(s, t) ds dt + cm^3 \\
 & \leq c \left(m^4 \lambda \int_{\mathcal{T} \times \mathcal{T}} f^2(s, t) ds dt + m^3 \right).
 \end{aligned}$$

Take $f = \varphi_{k_1} \otimes \varphi_{k_2}$, i.e., $f(s, t) = \varphi_{k_1}(s)\varphi_{k_2}(t)$. Then the first term can be bounded by $c_0 n^{-1}(\lambda + m^{-1})$.

In summary, we have

$$\mathbb{E} [D\mathcal{L}_{n,m,\lambda}(\bar{C}_\lambda)\varphi_{k_1} \otimes \varphi_{k_2}]^2 \leq cn^{-1} (a_{k_1 k_1} a_{k_2 k_2} + \lambda + m^{-1}).$$

Therefore,

$$\begin{aligned}
 \mathbb{E} \left\| \tilde{C}_\lambda - \bar{C}_\lambda \right\|_a^2 &= \mathbb{E} \left\| G_\lambda^{-1} D\mathcal{L}_{n,m,\lambda}(\bar{C}_\lambda) \right\|_a^2 \\
 &= \mathbb{E} \left[\sum_{k_1, k_2 \geq 1} \gamma_{k_1 k_2}^{-a} (1 + \lambda \gamma_{k_1 k_2}^{-1})^{-2} (D\mathcal{L}_{n,m,\lambda}(\bar{C}_\lambda)\varphi_{k_1} \otimes \varphi_{k_2})^2 \right] \\
 &\leq cn^{-1} \left[\sum_{k_1, k_2 \geq 1} \gamma_{k_1 k_2}^{-a} (1 + \lambda \gamma_{k_1 k_2}^{-1})^{-2} a_{k_1 k_1} a_{k_2 k_2} \right] \\
 &\quad + cn^{-1}(\lambda + m^{-1}) \sum_{k_1, k_2 \geq 1} \gamma_{k_1 k_2}^{-a} (1 + \lambda \gamma_{k_1 k_2}^{-1})^{-2}
 \end{aligned}$$

Observe that

$$\begin{aligned}
 & \left[\sum_{k_1, k_2 \geq 1} \gamma_{k_1 k_2}^{-a} (1 + \lambda \gamma_{k_1 k_2}^{-1})^{-2} a_{k_1 k_1} a_{k_2 k_2} \right] \\
 & \leq \sum_{k_1, k_2 \geq 1} \rho_{k_1}^{-a} \rho_{k_2}^{-a} a_{k_1 k_1} a_{k_2 k_2}
 \end{aligned}$$

$$\begin{aligned}
 &= \left(\sum_{k_1 \geq 1} \rho_{k_1}^{-a} a_{k_1 k_1} \right)^2 \\
 &\leq \left(\sum_{k_1 \geq 1} \rho_{k_1}^{-1} a_{k_1 k_1} \right)^2.
 \end{aligned}$$

Recall that

$$\begin{aligned}
 a_{k_1 k_1} &= \int_{\mathcal{T} \times \mathcal{T}} \varphi_{k_1}(s) \varphi_{k_1}(t) C_0(s, t) ds dt \\
 &= \mathbb{E} \left(\int_{\mathcal{T}} X(t) \varphi_{k_1}(t) dt \right)^2.
 \end{aligned}$$

We get

$$\sum_{k_1 \geq 1} \rho_{k_1}^{-1} a_{k_1 k_1} = \mathbb{E} \sum_{k_1=1}^{\infty} \rho_{k_1}^{-1} \left(\int_{\mathcal{T}} X(t) \varphi_{k_1}(t) dt \right)^2 = \mathbb{E} \|X\|_{\mathcal{W}_2^\alpha}^2.$$

Hence, the first term can be bounded by cn^{-1} . Together with the fact that

(see, e.g., Lin, 2000)

$$\sum_{k_1, k_2=1}^{\infty} \gamma_{k_1 k_2}^{-a} (1 + \lambda \gamma_{k_1 k_2}^{-1})^{-2} \sim \lambda^{-(a+1/(2\alpha))} \log(1/\lambda),$$

the proof can now be completed by collecting all these inequalities. $\square$

It now remains to bound $\hat{C}_\lambda - \tilde{C}_\lambda$.

Lemma 4. *If there exists some $1/(2\alpha) < b < 1/2$ such that*

$$\frac{\log(1/\lambda)}{nm\lambda^{2b+1/(2\alpha)}} \rightarrow 0,$$

then for any $0 \leq a \leq b$,

$$\left\| \hat{C}_\lambda - \tilde{C}_\lambda \right\|_a^2 = o_p \left(n^{-1} + n^{-1} \lambda^{1-a-\frac{1}{2\alpha}} \log(1/\lambda) + (nm)^{-1} \lambda^{-a-\frac{1}{2\alpha}} \log(1/\lambda) \right).$$

Before proving Lemma 4, we state the following technical lemma, which is proved in Subsection S3.3.

Lemma 5. *Under the conditions of Theorem 2, if $1/(2\alpha) < b < 1/2$, then for any $0 \leq a \leq b$*

$$\left\| \hat{C}_\lambda - \tilde{C}_\lambda \right\|_a^2 = O_p \left(\frac{\log(1/\lambda)}{nm\lambda^{a+b+1/(2\alpha)}} \right) \left\| \hat{C}_\lambda - \bar{C}_\lambda \right\|_b^2.$$

Proof of Lemma 4. By Lemma 5,

$$\left\| \hat{C}_\lambda - \tilde{C}_\lambda \right\|_a^2 = O_p \left(\frac{\log(1/\lambda)}{nm\lambda^{a+b+1/(2\alpha)}} \right) \left\| \hat{C}_\lambda - \bar{C}_\lambda \right\|_b^2.$$

If $a > 1/(2\alpha)$, then taking $b = a$ yields

$$\left\| \hat{C}_\lambda - \tilde{C}_\lambda \right\|_a^2 = O_p \left(\frac{\log(1/\lambda)}{nm\lambda^{2a+1/(2\alpha)}} \right) \left\| \hat{C}_\lambda - \bar{C}_\lambda \right\|_a^2.$$

Since $a \leq b$, the assumed condition implies

$$\frac{\log(1/\lambda)}{nm\lambda^{2a+1/(2\alpha)}} \rightarrow 0,$$

and hence

$$\left\| \hat{C}_\lambda - \tilde{C}_\lambda \right\|_a = o_p \left(\left\| \hat{C}_\lambda - \bar{C}_\lambda \right\|_a \right).$$

Together with the triangular inequality

$$\left\| \tilde{C}_\lambda - \bar{C}_\lambda \right\|_a \geq \left\| \hat{C}_\lambda - \bar{C}_\lambda \right\|_a - \left\| \hat{C}_\lambda - \tilde{C}_\lambda \right\|_a = (1 - o_p(1)) \left\| \hat{C}_\lambda - \bar{C}_\lambda \right\|_a.$$

Therefore,

$$\left\| \hat{C}_\lambda - \bar{C}_\lambda \right\|_a^2 = O_p \left(\left\| \tilde{C}_\lambda - \bar{C}_\lambda \right\|_a^2 \right)$$

$$= O_p \left(n^{-1} + n^{-1} \lambda^{1-a-\frac{1}{2\alpha}} \log(1/\lambda) + (nm)^{-1} \lambda^{-a-\frac{1}{2\alpha}} \log(1/\lambda) \right).$$

Now consider the case when $a \leq 1/(2\alpha)$. From the previous discussion, we know that for any $1/(2\alpha) < b < 1/2$ such that

$$\frac{\log(1/\lambda)}{nm\lambda^{2b+1/(2\alpha)}} \rightarrow 0,$$

we have

$$\begin{aligned} \left\| \hat{C}_\lambda - \bar{C}_\lambda \right\|_b^2 &= O_p \left(n^{-1} + n^{-1} \lambda^{1-b-\frac{1}{2\alpha}} \log(1/\lambda) \right. \\ &\quad \left. + (nm)^{-1} \lambda^{-b-\frac{1}{2\alpha}} \log(1/\lambda) \right) \end{aligned}$$

Hence,

$$\begin{aligned} \left\| \hat{C}_\lambda - \tilde{C}_\lambda \right\|_a^2 &= O_p \left(\frac{\log(1/\lambda)}{nm\lambda^{a+b+1/(2\alpha)}} \right) \\ &\quad \times O_p \left(n^{-1} + n^{-1} \lambda^{1-b-\frac{1}{2\alpha}} \log(1/\lambda) \right. \\ &\quad \left. + (nm)^{-1} \lambda^{-b-\frac{1}{2\alpha}} \log(1/\lambda) \right) \\ &= o_p \left(n^{-1} \lambda^{b-a} + n^{-1} \lambda^{1-a-\frac{1}{2\alpha}} \log(1/\lambda) \right. \\ &\quad \left. + (nm)^{-1} \lambda^{-a-\frac{1}{2\alpha}} \log(1/\lambda) \right). \end{aligned}$$

To sum up, if there exists some $1/(2\alpha) < b < 1/2$ such that

$$\frac{\log(1/\lambda)}{nm\lambda^{2b+1/(2\alpha)}} \rightarrow 0,$$

then for any $0 \leq a \leq b$,

$$\left\| \hat{C}_\lambda - \tilde{C}_\lambda \right\|_a^2 = o_p \left(n^{-1} + n^{-1} \lambda^{1-a-\frac{1}{2\alpha}} \log(1/\lambda) + (nm)^{-1} \lambda^{-a-\frac{1}{2\alpha}} \log(1/\lambda) \right).$$

□

Since $\alpha > 1$, we may choose $1/(2\alpha) < b < 1/2$. For the smoothing parameter

$$\lambda \asymp (nm/\log n)^{-2\alpha/(2\alpha+1)},$$

this choice gives

$$\frac{\log(1/\lambda)}{nm\lambda^{2b+1/(2\alpha)}} \asymp \frac{\log(nm/\log n)}{(nm)^{(2\alpha-4\alpha b)/(2\alpha+1)}(\log n)^{(4\alpha b+1)/(2\alpha+1)}} \longrightarrow 0.$$

Combining the bounds on $\bar{C}_\lambda - C_0$, $\tilde{C}_\lambda - \bar{C}_\lambda$ and $\hat{C}_\lambda - \tilde{C}_\lambda$, we have

$$\|\hat{C}_\lambda - C_0\|_{L_2}^2 = O_p\left(\lambda + n^{-1} + n^{-1}\lambda^{1-\frac{1}{2\alpha}}\log(1/\lambda) + (nm)^{-1}\lambda^{-\frac{1}{2\alpha}}\log(1/\lambda)\right).$$

Taking

$$\lambda \sim (nm/\log n)^{-\frac{2\alpha}{2\alpha+1}}$$

yields that

$$\|\hat{C}_\lambda - C_0\|_{L_2}^2 = O_p\left((nm/\log n)^{-\frac{2\alpha}{2\alpha+1}} + n^{-1}\right),$$

which completes the proof of Theorem 2.

S1.4 Proof of Theorem 1

Under the common design, it is easy to see that the estimator $\hat{C}_\lambda$ defined in (2.4) can be rewritten as

$$\hat{C}_\lambda = \arg \min_{C \in \mathcal{W}_2^\alpha \otimes \mathcal{W}_2^\alpha} \left[\frac{1}{m(m-1)} \sum_{1 \leq j \neq k \leq m} (\bar{Z}_{jk} - C(T_j, T_k))^2 + \lambda \|C\|_{\mathcal{W}_2^\alpha \otimes \mathcal{W}_2^\alpha}^2 \right], \quad (\text{S1.24})$$

where we introduce

$$\bar{Z}_{jk} = \frac{1}{n} \sum_{i=1}^n Y_{ij} Y_{ik}, \quad 1 \leq j \neq k \leq m.$$

Let us denote by Q_T the spline interpolation operator on the tensor product space $\mathcal{W}_2^\alpha \otimes \mathcal{W}_2^\alpha$ over the grid points

$$\Omega = \{(T_j, T_k) : 1 \leq j \neq k \leq m\},$$

namely, $Q_T(f)$ is the unique solution in $\mathcal{W}_2^\alpha \otimes \mathcal{W}_2^\alpha$ of

$$\min_{g \in \mathcal{W}_2^\alpha \otimes \mathcal{W}_2^\alpha} \|g\|_{\mathcal{W}_2^\alpha \otimes \mathcal{W}_2^\alpha} \quad \text{subject to} \quad g(T_j, T_k) = f(T_j, T_k), \quad 1 \leq j \neq k \leq m. \quad (\text{S1.25})$$

It is known that Q_T is a linear operator. Moreover, by the definition (2.4) of $\hat{C}_\lambda$, we have $\hat{C}_\lambda = Q_T(\hat{C}_\lambda)$. Thus, we decompose the estimation error as

$$\begin{aligned} \|\hat{C}_\lambda - C_0\|_{L_2}^2 &\lesssim \|\hat{C}_\lambda - Q_T(C_0)\|_{L_2}^2 + \|Q_T(C_0) - C_0\|_{L_2}^2 \\ &= \|Q_T(\hat{C}_\lambda - C_0)\|_{L_2}^2 + \|C_0 - Q_T(C_0)\|_{L_2}^2. \end{aligned}$$

Under Assumption 1, the approximation error of the tensor product spline interpolation satisfies (Schumaker, 2007):

$$\|C_0 - Q_T(C_0)\|_{L_2}^2 \leq C m^{-2\alpha} \|C_0\|_{\mathcal{W}_2^\alpha \otimes \mathcal{W}_2^\alpha}^2 \lesssim m^{-2\alpha}.$$

For the estimation error, we have to use the stability property of the spline interpolation operator (see, e.g., Schumaker (2007) and the proof of

Theorem 2.2 in Cai and Yuan (2011)): there exists a constant $C > 0$ such that for any $f \in \mathcal{W}_2^\alpha \otimes \mathcal{W}_2^\alpha$,

$$\|Q_T(f)\|_{L_2}^2 \leq C \frac{1}{m(m-1)} \sum_{1 \leq j \neq k \leq m} f^2(T_j, T_k).$$

Thus,

$$\begin{aligned} & \left\| Q_T(\hat{C}_\lambda - C_0) \right\|_{L_2}^2 \\ & \lesssim \frac{1}{m(m-1)} \sum_{1 \leq j \neq k \leq m} \left[\hat{C}_\lambda(T_j, T_k) - C_0(T_j, T_k) \right]^2 \\ & \lesssim \frac{1}{m(m-1)} \sum_{1 \leq j \neq k \leq m} \left[\begin{aligned} & \left(\hat{C}_\lambda(T_j, T_k) - \bar{Z}_{jk} \right)^2 \\ & + \left(\bar{Z}_{jk} - C_0(T_j, T_k) \right)^2 \end{aligned} \right]. \end{aligned}$$

It is easy to see that $\mathbb{E}\bar{Z}_{jk} = C_0(T_j, T_k)$ and also

$$\begin{aligned} \text{Var}(\bar{Z}_{jk}) &= \frac{1}{n} \text{Var}(Y_{1j}Y_{1k}) \leq \frac{1}{n} \mathbb{E}Y_{1j}^2 Y_{1k}^2 \\ &= \frac{1}{n} \left[\mathbb{E}X_1(T_j)^2 X_1(T_k)^2 + \sigma_0^2 (C_0(T_j, T_j) + C_0(T_k, T_k)) + \sigma_0^4 \right] \\ &\leq \frac{C}{n}, \end{aligned}$$

where we bound the first term in the last inequality using

$$\mathbb{E}X_1(T_j)^2 X_1(T_k)^2 \leq \sqrt{\mathbb{E}X_1(T_j)^4 \mathbb{E}X_1(T_k)^4},$$

and

$$\sqrt{\mathbb{E}X_1(T_j)^4} \lesssim \mathbb{E}X_1(T_j)^2 \lesssim \mathbb{E}\|X_1\|_{\mathcal{W}_2^\alpha}^2 \lesssim 1.$$

Hence, it is shown that

$$\frac{1}{m(m-1)} \sum_{1 \leq j \neq k \leq m} (\bar{Z}_{jk} - C_0(T_j, T_k))^2 = O_p\left(\frac{1}{n}\right).$$

For the other term, by the definition of $\hat{C}_\lambda$, we have

$$\begin{aligned} & \frac{1}{m(m-1)} \sum_{1 \leq j \neq k \leq m} \left(\hat{C}_\lambda(T_j, T_k) - \bar{Z}_{jk} \right)^2 \\ & \leq \frac{1}{m(m-1)} \sum_{1 \leq j \neq k \leq m} \left(\hat{C}_\lambda(T_j, T_k) - \bar{Z}_{jk} \right)^2 + \lambda \left\| \hat{C}_\lambda \right\|_{\mathcal{W}_2^\alpha \otimes \mathcal{W}_2^\alpha}^2 \\ & \leq \frac{1}{m(m-1)} \sum_{1 \leq j \neq k \leq m} \left(C_0(T_j, T_k) - \bar{Z}_{jk} \right)^2 + \lambda \|C_0\|_{\mathcal{W}_2^\alpha \otimes \mathcal{W}_2^\alpha}^2 \\ & \lesssim O_p\left(\frac{1}{n}\right) + \lambda. \end{aligned}$$

Consequently, if we take $\lambda \lesssim n^{-1} + m^{-2\alpha}$, then this term is of order $O_p(n^{-1} + m^{-2\alpha})$. Combining the bounds on the two terms, we conclude the proof.

S1.5 Proof of Theorem 5

We write n for the training sample size and $\tilde{n} \asymp n$ for the validation sample size. Let $\delta \in (0, 1)$ be arbitrary. Let us first introduce the following concentration inequality (Caponnetto and Yao, 2010), which will be useful in our cross-validation analysis.

Lemma 6. *Let $\xi_1, \dots, \xi_n$ be n i.i.d. bounded random variables such that $|\xi_i| \leq B$ almost surely, $\mathbb{E}(\xi_i) = \mu$, and $\text{Var}(\xi_i) \leq \sigma^2$. Then for any $\alpha > 0$,*

any $\delta \in (0, 1)$, we have

$$\left| \frac{1}{n} \sum_{i=1}^n \xi_i - \mu \right| \leq \alpha \sigma^2 + \frac{3 + 4\alpha B}{6\alpha n} \ln \frac{2}{\delta} \quad (\text{S1.26})$$

holds with probability at least $1 - \delta$.

We use the known-mean formulation introduced at the beginning of this section and take $\mu_0 = 0$. Condition on the training samples, so that the candidates $\hat{f}_\lambda = \Pi_B \hat{C}_\lambda$, $\lambda \in \Lambda$, are fixed and the validation curves remain independent. Choose $B \geq \|C_0\|_\infty$, which is possible uniformly over the stated class by the Sobolev embedding and Proposition 1. Let p_T be the design density and define

$$R(f) = \int_{\mathcal{T}^2} [f(s, t) - C_0(s, t)]^2 p_T(s) p_T(t) ds dt.$$

Since p_T is bounded above and below, $R(f) \asymp \|f - C_0\|_{L_2}^2$. For $\|f\|_\infty \leq B$, write

$$\begin{aligned} H_f(s, t, y, y') &= (yy' - f(s, t))^2 - (yy' - C_0(s, t))^2 \\ &= [C_0(s, t) - f(s, t)][2yy' - f(s, t) - C_0(s, t)]. \end{aligned}$$

For two distinct observations from the same curve, the model gives

$$\mathbb{E}(Y_{ij} Y_{ik} \mid T_{ij}, T_{ik}) = C_0(T_{ij}, T_{ik}), \quad j \neq k.$$

Consequently,

$$\mathbb{E} H_f(T_{ij}, T_{ik}, Y_{ij}, Y_{ik}) = R(f).$$

The responses in this expectation belong to the same curve and need not be independent.

Take $M = (c_M \log n)^{1/2}$ for a sufficiently large constant c_M and set

$$\begin{aligned} A_{ijk} &= \mathbf{1}\{|Y_{ij}| \leq M, |Y_{ik}| \leq M\}, \\ \xi_i(f) &= \frac{1}{m(m-1)} \sum_{j \neq k} H_f(T_{ij}, T_{ik}, Y_{ij}, Y_{ik}) A_{ijk}, \\ \Delta_M(f) &= \frac{1}{\tilde{n}} \sum_{i=1}^{\tilde{n}} \xi_i(f), \\ \widehat{\Delta}(f) &= \mathcal{L}_{\text{val}}(f) - \mathcal{L}_{\text{val}}(C_0). \end{aligned}$$

Since $|H_f(s, t, y, y')| \leq C(1 + |yy'|)$ uniformly over $\|f\|_\infty \leq B$, we have

$$\sup_{\|f\|_\infty \leq B} |\widehat{\Delta}(f) - \Delta_M(f)| \leq \mathcal{R}_M,$$

where

$$\mathcal{R}_M = \frac{C}{\tilde{n}m(m-1)} \sum_{i=1}^{\tilde{n}} \sum_{j \neq k} (1 + |Y_{ij}Y_{ik}|)(1 - A_{ijk}).$$

The marginal sub-Gaussian tail bound and the Cauchy–Schwarz inequality give

$$\begin{aligned} \mathbb{E}\mathcal{R}_M &\leq C\mathbb{E}[(1 + |Y_{11}Y_{12}|)(1 - A_{112})] \\ &\leq C\{\mathbb{E}(1 + |Y_{11}Y_{12}|)^2\}^{1/2}\{\mathbb{P}(A_{112} = 0)\}^{1/2} \\ &\leq Ce^{-cM^2} \leq n^{-6}, \end{aligned}$$

for all sufficiently large n , after choosing c_M large enough. These bounds hold uniformly in m and do not require independence of Y_{11} and Y_{12} . They

also imply

$$\sup_{\|f\|_\infty \leq B} |\mathbb{E}\xi_i(f) - R(f)| \leq n^{-6}.$$

By Markov's inequality, $\mathcal{R}_M \leq n^{-2}$ with probability at least $1 - n^{-4}$.

For each fixed candidate, the variables $\xi_i(f)$ are independent across curves and satisfy $|\xi_i(f)| \leq CM^2$. Jensen's inequality, applied to the average over pairs within one curve, gives

$$\begin{aligned} \text{Var}(\xi_i(f)) &\leq \mathbb{E}\xi_i(f)^2 \\ &\leq \frac{1}{m(m-1)} \sum_{j \neq k} \mathbb{E}[H_f(T_{ij}, T_{ik}, Y_{ij}, Y_{ik})^2 A_{ijk}] \\ &\leq CM^4 R(f). \end{aligned}$$

Applying Lemma 6 to the $\tilde{n}$ validation curves with its parameter $\alpha = (2CM^4)^{-1}$, using $\tilde{n} \asymp n$ and a union bound over Λ , we obtain

$$|\Delta_M(\hat{f}_\lambda) - \mathbb{E}\xi_i(\hat{f}_\lambda)| \leq \frac{1}{2}R(\hat{f}_\lambda) + \frac{CM^4}{n} \log \frac{4|\Lambda|}{\delta},$$

simultaneously for all $\lambda \in \Lambda$ with conditional probability at least $1 - \delta/2$.

Combining this event with the truncation bounds, for each fixed $\delta \in (0, 1)$

and all sufficiently large n , we have

$$\frac{1}{2}R(\hat{f}_\lambda) - \tau \leq \hat{\Delta}(\hat{f}_\lambda) \leq \frac{3}{2}R(\hat{f}_\lambda) + \tau$$

simultaneously for all candidates with probability at least $1 - \delta$, where

$$\tau = \frac{C(\log n)^2}{n} \log \frac{4|\Lambda|}{\delta} + 2n^{-2}.$$

Since $\hat{\lambda} = \hat{\lambda}_{\text{cv}}$ minimizes the validation loss of the clipped candidates, these inequalities yield

$$R(\hat{f}_{\hat{\lambda}}) \leq 3R(\hat{f}_{\lambda_*}) + 4\tau, \quad \lambda_* \in \Lambda. \quad (\text{S1.27})$$

The probability bound is uniform in the training samples and hence also holds unconditionally.

It remains to identify an oracle candidate in the stated grid. Write $q = 1/(2\alpha)$ and

$$\lambda_0 = \left(\frac{\log n}{nm} \right)^{1/(1+q)}, \quad \lambda_{\dagger} = \max(\lambda_0, n^{-1}).$$

For all sufficiently large n , the dyadic grid contains λ_* with $\lambda_{\dagger} \leq \lambda_* \leq 2\lambda_{\dagger}$. Choose $q < b < 1/2$. Since $\lambda_* \geq \lambda_0$, the condition of Lemma 4 holds for λ_* by the calculation in the proof of Theorem 2. The bounds established there therefore give

$$\begin{aligned} \|\hat{C}_{\lambda_*} - C_0\|_{L_2}^2 &= O_p(\lambda_* + n^{-1} + n^{-1}\lambda_*^{1-q} \log(1/\lambda_*)) \\ &\quad + (nm)^{-1}\lambda_*^{-q} \log(1/\lambda_*)). \end{aligned}$$

Here $\lambda_* \leq 2(\lambda_0 + n^{-1})$, and $\lambda_* \geq n^{-1}$ implies $\log(1/\lambda_*) \leq \log n$. Moreover,

$$(nm)^{-1}\lambda_*^{-q} \log(1/\lambda_*) \leq (nm)^{-1}\lambda_0^{-q} \log n = \lambda_0.$$

The remaining stochastic term is $O(n^{-1})$ because $q < 1$. Thus $\|\hat{C}_{\lambda_*} - C_0\|_{L_2}^2 = O_p(\lambda_0 + n^{-1})$. Clipping cannot increase the pointwise error when

$|C_0| \leq B$, so $R(\hat{f}_{\lambda_*}) = O_p(\lambda_0 + n^{-1})$ as well. Finally, $|\Lambda| = O(\log n)$ implies $\tau = O_\delta((\log n)^3/n)$. Combining these bounds with (S1.27) and the equivalence of $R(f)$ and squared L_2 loss proves Theorem 5.

S1.6 Proof of Corollary 1

Let $\mathcal{C}_0$ and $\hat{\mathcal{C}}_\lambda$ denote the integral operators with kernels C_0 and $\hat{C}_\lambda$, respectively, and write $E = \hat{\mathcal{C}}_\lambda - \mathcal{C}_0$. For a fixed simple eigenvalue θ_k , define its local spectral gap by

$$\delta_k = \begin{cases} \theta_1 - \theta_2, & k = 1, \\ \min(\theta_{k-1} - \theta_k, \theta_k - \theta_{k+1}), & k \geq 2. \end{cases}$$

Then $\delta_k > 0$. The eigenvalue perturbation bound gives $|\hat{\theta}_k - \theta_k| \leq \|E\|_{\text{op}}$.

On the event $\|E\|_{\text{op}} < \delta_k/2$, let P_k be the orthogonal projection onto the span of ψ_k . The eigenfunction equation implies

$$(\mathcal{C}_0 - \hat{\theta}_k I)(I - P_k)\hat{\psi}_k = -(I - P_k)E\hat{\psi}_k.$$

The spectrum of $\mathcal{C}_0$ on the orthogonal complement of ψ_k is at distance at least $\delta_k/2$ from $\hat{\theta}_k$. Consequently, $\|(I - P_k)\hat{\psi}_k\|_{L_2} \leq 2\|E\|_{\text{op}}/\delta_k$. Using the sign convention $\langle \hat{\psi}_k, \psi_k \rangle \geq 0$, we obtain

$$\|\hat{\psi}_k - \psi_k\|_{L_2}^2 \leq 2\|(I - P_k)\hat{\psi}_k\|_{L_2}^2 \leq \frac{8}{\delta_k^2}\|E\|_{\text{op}}^2 \leq \frac{8}{\delta_k^2}\|\hat{C}_\lambda - C_0\|_{L_2}^2.$$

This is the usual spectral perturbation argument; see also Bhatia, Davis and McIntosh (1983). On the complementary event, the same bound follows from $\|\hat{\psi}_k - \psi_k\|_{L_2}^2 \leq 2 \leq 8\|E\|_{\text{op}}^2/\delta_k^2$. Invoking Theorem 1 or Theorem 2 completes the proof.

S2. Proofs of Minimax Lower Bounds

We first establish a parametric lower bound that is common to the two sampling designs.

Lemma 7. *Under either the common or independent design, there exist constants $c_1, p_1 > 0$ such that, for every covariance estimator $\tilde{C}$ and all sufficiently large n ,*

$$\sup_{L(X) \in \mathcal{P}'(\alpha; M_0)} \mathbb{P} \left\{ \|\tilde{C} - C_0\|_{L_2}^2 > c_1 n^{-1} \right\} \geq p_1. \quad (\text{S2.28})$$

Proof. Let $\mathbf{1}_m$ denote the m -vector of ones. Fix $\theta_0 > 0$ sufficiently small that $2\theta_0\|1\|_{\mathcal{W}_2^{\mathcal{S}}(\mathcal{T})}^2 \leq M_0$, and set

$$\theta_1 = \theta_0 + \eta n^{-1/2}, \quad \psi_\ell(t) = \sqrt{\theta_\ell}, \quad \ell = 0, 1,$$

where $\eta > 0$ is a sufficiently small constant. For all sufficiently large n , the processes $X(t) = Z\psi_\ell(t)$, $Z \sim N(0, 1)$, belong to $\mathcal{P}'(\alpha; M_0)$ and have covariance functions $C^{(\ell)}(s, t) = \theta_\ell$.

Under either sampling design, the distribution of the observations from one curve is centered Gaussian with covariance matrix

$$\Sigma_\ell = \theta_\ell \mathbf{1}_m \mathbf{1}_m^\top + \sigma_0^2 I_m.$$

The design locations carry no information about θ_ℓ because the functions ψ_ℓ are constant. Writing

$$r_n = \frac{\sigma_0^2 + m\theta_1}{\sigma_0^2 + m\theta_0},$$

the Kullback–Leibler divergence between the two n -curve experiments satisfies

$$\begin{aligned} \text{KL}(P_1^{(n)}, P_0^{(n)}) &= \frac{n}{2} \{r_n - 1 - \log r_n\} \\ &\leq \frac{n}{4} (r_n - 1)^2 \leq \frac{\eta^2}{4\theta_0^2}. \end{aligned}$$

Choose η so that the last quantity is at most $1/8$. Since

$$\|C^{(1)} - C^{(0)}\|_{L_2}^2 = (\theta_1 - \theta_0)^2 = \eta^2 n^{-1},$$

Pinsker's inequality gives $\|P_1^{(n)} - P_0^{(n)}\|_{\text{TV}} \leq 1/4$. Hence Le Cam's two-point argument yields

$$\inf_{\tilde{C}} \max_{\ell=0,1} P_\ell^{(n)} \left\{ \|\tilde{C} - C^{(\ell)}\|_{L_2}^2 \geq \frac{\eta^2}{4n} \right\} \geq \frac{1 - \|P_1^{(n)} - P_0^{(n)}\|_{\text{TV}}}{2} \geq \frac{3}{8},$$

which proves the lemma. $\square$

S2.1 Proof of Theorem 4

It remains to establish the nonparametric term. Let p_T be the density of the sampling locations, which is bounded above under Assumption 2. Set

$$e_0 = \rho_1^{1/2} \varphi_1, \quad v_b = M^{-1/2} \sum_{k=M+1}^{2M} b_{k-M} \rho_k^{1/2} \varphi_k, \quad b \in \{-1, 1\}^M,$$

and define

$$\psi_b = a(e_0 + v_b), \quad \psi_0 = ae_0, \quad a = (M_0/2)^{1/2}.$$

By orthogonality in $\mathcal{W}_2^\alpha(\mathcal{T})$,

$$\|\psi_b\|_{\mathcal{W}_2^\alpha(\mathcal{T})}^2 = a^2(1 + M^{-1} \sum_{k=1}^M b_k^2) = M_0, \quad \|\psi_0\|_{\mathcal{W}_2^\alpha(\mathcal{T})}^2 = M_0/2.$$

Thus, the processes $X = Z\psi_b$ and $X = Z\psi_0$, with $Z \sim N(0, 1)$, all belong to $\mathcal{P}'(\alpha; M_0)$.

For $b, b' \in \{-1, 1\}^M$, the eigenvalue decay $\rho_k \asymp k^{-2\alpha}$ gives

$$\|\psi_b - \psi_{b'}\|_{L_2}^2 \geq cM^{-(2\alpha+1)}H(b, b'), \quad (\text{S2.29})$$

where H denotes Hamming distance. Moreover, the low-frequency component e_0 is L_2 -orthogonal to every v_b , and hence

$$\|\psi_b + \psi_{b'}\|_{L_2}^2 \geq 4a^2\rho_1.$$

The identity

$$\|f \otimes f - g \otimes g\|_{L_2}^2 = \frac{1}{2}\|f + g\|_{L_2}^2\|f - g\|_{L_2}^2 + \frac{1}{2}(\|f\|_{L_2}^2 - \|g\|_{L_2}^2)^2 \quad (\text{S2.30})$$

therefore implies

$$\|\psi_b \otimes \psi_b - \psi_{b'} \otimes \psi_{b'}\|_{L_2}^2 \geq cM^{-(2\alpha+1)}H(b, b'). \quad (\text{S2.31})$$

By the Varshamov–Gilbert bound, there exist $b^{(1)}, \dots, b^{(N)} \in \{-1, 1\}^M$ such that

$$N \geq 2^{M/8}, \quad H(b^{(j)}, b^{(k)}) \geq M/8, \quad j \neq k.$$

The covariance functions associated with this packing are consequently separated in squared L_2 distance by at least $cM^{-2\alpha}$. In addition, $\|v_b\|_{L_2}^2 \asymp M^{-2\alpha}$ and $e_0 \perp v_b$ in L_2 , so (S2.30) also gives

$$\|\psi_b \otimes \psi_b - \psi_0 \otimes \psi_0\|_{L_2}^2 \geq cM^{-2\alpha}$$

for every packed covariance.

Let P_b and P_0 denote the distributions of the observed independent-design data under ψ_b and ψ_0 , respectively. Introduce the augmented distributions Q_b and Q_0 that additionally reveal the common Gaussian score Z_i for each curve. By the data-processing inequality and the conditional Gaussian likelihood,

$$\begin{aligned} \text{KL}(P_b, P_0) &\leq \text{KL}(Q_b, Q_0) \\ &= \frac{n}{2\sigma_0^2} \mathbb{E} \left[\sum_{j=1}^m Z_i^2 \{\psi_b(T_{ij}) - \psi_0(T_{ij})\}^2 \right] \\ &= \frac{nm}{2\sigma_0^2} \|\psi_b - \psi_0\|_{L_2(p_T)}^2 \leq CnmM^{-2\alpha}. \end{aligned} \quad (\text{S2.32})$$

Take

$$M = \lfloor c_M (nm)^{1/(2\alpha+1)} \rfloor.$$

Since $\log N \geq cM$, choosing c_M sufficiently large makes the right-hand side of (S2.32) at most $\kappa \log N$ for some fixed $\kappa < 1/8$. Given an estimator $\tilde{C}$, define a decoder by selecting the covariance in the packing, including the baseline, that is closest to $\tilde{C}$ in L_2 distance. Whenever the estimation error is smaller than half the minimum packing separation, this decoder identifies the correct covariance. Fano's lemma (Tsybakov, 2009) therefore yields constants $c_2, p_2 > 0$ such that

$$\sup_{L(X) \in \mathcal{P}'(\alpha; M_0)} \mathbb{P} \left\{ \|\tilde{C} - C_0\|_{L_2}^2 > c_2 (nm)^{-2\alpha/(2\alpha+1)} \right\} \geq p_2 \quad (\text{S2.33})$$

for all sufficiently large n .

Finally, combine Lemma 7 and (S2.33). For each (n, m) , use the former when n^{-1} is the larger term and the latter otherwise, decreasing the constant if necessary. This proves Theorem 4.

S2.2 Proof of Theorem 3

It remains to establish the discretization term. Choose a nonzero compactly supported smooth function φ such that

$$\text{supp } \varphi \subseteq (-1, 1), \quad \int \varphi(t) dt = 0. \quad (\text{S2.34})$$

Take $h = 1/(4m)$ and define the functions

$$\varphi_k(t) = h^\alpha \varphi\left(\frac{t - t_k^*}{h}\right), \quad t_k^* = (2k - 1)h, \quad k = 1, \dots, 2m.$$

We have

$$\|\varphi_k\|_{\mathcal{W}_2^\alpha(\mathcal{T})}^2 \leq C_\varphi h, \quad \|\varphi_k\|_{L^2}^2 = h^{2\alpha+1} \|\varphi\|_{L^2}^2 = (4m)^{-(2\alpha+1)} \|\varphi\|_{L^2}^2.$$

Moreover, φ_k is supported on the interval $(t_k^* - h, t_k^* + h)$ for $k = 1, \dots, 2m$.

Since we only have m design points, we can take a subset S of $\{1, \dots, 2m\}$ with cardinality m such that $(t_k^* - h, t_k^* + h)$ contains no design points for any $k \in S$. The Fourier characterization of $\mathcal{W}_2^\alpha(\mathcal{T})$, together with the Bessel bound for the equally spaced translates of φ , gives the uniform localization estimate

$$\left\| \sum_{k=1}^{2m} d_k \varphi_k \right\|_{\mathcal{W}_2^\alpha(\mathcal{T})}^2 \leq C_\varphi h \sum_{k=1}^{2m} d_k^2, \quad d_1, \dots, d_{2m} \in \mathbb{R}.$$

Taking $d_k = \mathbf{1}\{k \in S\}$ and using the disjoint supports, we obtain

$$\left\| \sum_{k \in S} \varphi_k \right\|_{\mathcal{W}_2^\alpha(\mathcal{T})}^2 \leq C_\varphi m h \leq C_\varphi, \quad \left\| \sum_{k \in S} \varphi_k \right\|_{L^2}^2 = m h^{2\alpha+1} \|\varphi\|_{L^2}^2 \asymp m^{-2\alpha}.$$

Write $u = \sum_{k \in S} \varphi_k$ and choose a fixed $a > 0$, depending only on M_0 , α , and φ , such that

$$\psi_+ = a(1 + u), \quad \psi_- = a(1 - u)$$

both have squared $\mathcal{W}_2^\alpha(\mathcal{T})$ norm at most M_0 . Thus, the Gaussian rank-one processes $X = Z\psi_+$ and $X = Z\psi_-$ belong to $\mathcal{P}'(\alpha; M_0)$.

Since $u(T_j) = 0$ at every common design point, the two processes induce exactly the same distribution of the observed data. On the other hand, $\int u = 0$, so (S2.30) yields

$$\|\psi_+ \otimes \psi_+ - \psi_- \otimes \psi_-\|_{L_2}^2 = 8a^4 \|u\|_{L_2}^2 \geq cm^{-2\alpha}.$$

The two L_2 balls centered at these covariance functions with radius equal to half their separation are disjoint. Because the observation laws are identical, any estimator must therefore have squared L_2 error at least $cm^{-2\alpha}$ with probability at least $1/2$ under one of the two processes. Consequently, for some $c_3 > 0$ and all sufficiently large n ,

$$\sup_{L(X) \in \mathcal{P}'(\alpha; M_0)} \mathbb{P} \left\{ \|\tilde{C} - C_0\|_{L_2}^2 > c_3 m^{-2\alpha} \right\} \geq \frac{1}{2}. \quad (\text{S2.35})$$

Combining Lemma 7 and (S2.35) proves Theorem 3.

S3. Proof of Auxiliary Results

S3.1 Eigenfunction Computation

The finite-dimensional representation in Lemma 1 also yields the functional principal components without discretizing the estimated covariance function. By Mercer's theorem, write

$$C_0(s, t) = \sum_{k \geq 1} \theta_k \psi_k(s) \psi_k(t), \quad (\text{S3.36})$$

where $\{\psi_k : k \geq 1\}$ is an orthonormal basis of L_2 and $\theta_1 \geq \theta_2 \geq \dots$ are the associated eigenvalues. Let

$$A = \begin{bmatrix} A_1 & 0 & 0 & \cdots & 0 \\ 0 & A_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & A_n \end{bmatrix}, \quad Q = \begin{bmatrix} Q_{11} & Q_{12} & Q_{13} & \cdots & Q_{1n} \\ Q_{21} & Q_{22} & Q_{23} & \cdots & Q_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ Q_{n1} & Q_{n2} & Q_{n3} & \cdots & Q_{nn} \end{bmatrix}, \quad (\text{S3.37})$$

where

$$Q_{i_1 i_2} = \left[\int_{\mathcal{T}} K(s, T_{i_1 j_1}) K(s, T_{i_2 j_2}) ds \right]_{1 \leq j_1, j_2 \leq m}, \quad 1 \leq i_1, i_2 \leq n. \quad (\text{S3.38})$$

Lemma 8. *Let $r = \text{rank}(Q)$ and write $Q = RDR^\top$, where R is an $nm \times r$ matrix with orthonormal columns and D is a positive diagonal $r \times r$ matrix. Let V be an orthogonal matrix of eigenvectors of $D^{1/2}R^\top ARD^{1/2}$, and set $U = RD^{-1/2}V$. An orthonormal set of eigenfunctions of $\hat{C}_\lambda$ spanning the space generated by the observed kernel sections is given by*

$$\hat{\psi}_k(\cdot) = U_k^\top \mathbf{g}(\cdot), \quad k = 1, \dots, r, \quad (\text{S3.39})$$

where U_k is the k th column of U and

$$\mathbf{g}(\cdot) = (K(\cdot, T_{11}), \dots, K(\cdot, T_{1m}), K(\cdot, T_{21}), \dots, K(\cdot, T_{nm}))^\top. \quad (\text{S3.40})$$

The corresponding eigenvalues are those of $D^{1/2}R^\top ARD^{1/2}$. Every eigenfunction with a nonzero eigenvalue belongs to this span.

Proof. We first show that the functions $\hat{\psi}_k$ are orthonormal. Observe that

$$\int_{\mathcal{T}} \mathbf{g}(s) \mathbf{g}'(s) ds = Q. \quad (\text{S3.41})$$

Therefore,

$$\int_{\mathcal{T}} \hat{\psi}_{k_1}(s) \hat{\psi}_{k_2}(s) ds = U'_{k_1} Q U_{k_2} = V'_{k_1} D^{-1/2} R' Q R D^{-1/2} V_{k_2} = \delta_{k_1 k_2}, \quad (\text{S3.42})$$

where δ is the Kronecker delta.

If $z \in \ker(Q)$, then $\int_{\mathcal{T}} [z' \mathbf{g}(s)]^2 ds = z' Q z = 0$. Since the kernel sections are continuous, $z' \mathbf{g}(s) = 0$ for every $s \in \mathcal{T}$. Thus $RR' \mathbf{g}(s) = \mathbf{g}(s)$. Denote by $\hat{\theta}_k$ the eigenvalues of $D^{1/2} R' A R D^{1/2}$. Then

$$\begin{aligned} \sum_k \hat{\theta}_k \hat{\psi}_k(s) \hat{\psi}_k(t) &= \mathbf{g}'(s) \left(\sum_k \hat{\theta}_k U_k U'_k \right) \mathbf{g}(t) \\ &= \mathbf{g}'(s) \left(R D^{-1/2} V \text{diag}(\hat{\theta}_1, \dots, \hat{\theta}_r) V' D^{-1/2} R' \right) \mathbf{g}(t) \\ &= \mathbf{g}'(s) R R' A R R' \mathbf{g}(t) \\ &= \hat{C}_{\lambda}(s, t), \end{aligned}$$

which verifies the eigenfunction equations. The range of the integral operator with kernel $\hat{C}_{\lambda}$ is contained in the span of the observed kernel sections, so every eigenfunction with a nonzero eigenvalue lies in this span. Under the common design, $r \leq m$ because the kernel sections repeat across curves.

This completes the proof. $\square$

S3.2 Proof of Lemma 3

Observe that

$$\begin{aligned}
U &= \sum_{1 \leq j \neq k \leq m} X(T_j)X(T_k)f(T_j)g(T_k) + \sum_{1 \leq j \neq k \leq m} \epsilon_j X(T_k)f(T_j)g(T_k) \\
&\quad + \sum_{1 \leq j \neq k \leq m} \epsilon_k X(T_j)f(T_j)g(T_k) + \sum_{1 \leq j \neq k \leq m} \epsilon_j \epsilon_k f(T_j)g(T_k) \\
&:= U_1 + U_2 + U_3 + U_4.
\end{aligned}$$

By the Cauchy-Schwarz inequality,

$$\mathbb{E}[\text{Var}(U|T)] \leq \mathbb{E}[\mathbb{E}(U^2|T)] \leq 4 \left(\mathbb{E}(U_1^2) + \mathbb{E}(U_2^2) + \mathbb{E}(U_3^2) + \mathbb{E}(U_4^2) \right).$$

We now bound the four terms on the right-hand side separately.

We begin with $\mathbb{E}(U_1^2)$.

$$\begin{aligned}
&\mathbb{E}(U_1^2) \\
&= \mathbb{E} \left(\sum_{1 \leq j_1 \neq j_2 \neq j_3 \neq j_4 \leq m} X(T_{j_1})X(T_{j_2})X(T_{j_3})X(T_{j_4}) \right. \\
&\quad \left. \times f(T_{j_1})g(T_{j_2})f(T_{j_3})g(T_{j_4}) \right) \\
&\quad + \mathbb{E} \left(\sum_{1 \leq j_1 \neq j_2 \neq j_3 \leq m} X(T_{j_1})X^2(T_{j_2})X(T_{j_3}) \right. \\
&\quad \left. \times f(T_{j_1})g(T_{j_2})f(T_{j_2})g(T_{j_3}) \right) \\
&\quad + \dots \\
&\quad + \mathbb{E} \left(\sum_{1 \leq j_1 \neq j_2 \leq m} X^2(T_{j_1})X^2(T_{j_2})f^2(T_{j_1})g^2(T_{j_2}) \right)
\end{aligned}$$

$$\begin{aligned}
 &= m(m-1)(m-2)(m-3) \\
 &\quad \times \mathbb{E} \left[\left(\int X(s)f(s)ds \right)^2 \left(\int X(s)g(s)ds \right)^2 \right] \\
 &+ m(m-1)(m-2) \\
 &\quad \times \mathbb{E} \left[\left(\int X(s)f(s)ds \right) \left(\int X(s)g(s)ds \right) \left(\int X^2(s)f(s)g(s)ds \right) \right] \\
 &+ m(m-1)(m-2) \\
 &\quad \times \mathbb{E} \left[\left(\int X(s)f(s)ds \right)^2 \left(\int X^2(s)g^2(s)ds \right) \right] \\
 &+ m(m-1)(m-2) \\
 &\quad \times \mathbb{E} \left[\left(\int X(s)g(s)ds \right)^2 \left(\int X^2(s)f^2(s)ds \right) \right] \\
 &+ m(m-1) \\
 &\quad \times \mathbb{E} \left[\left(\int X^2(s)f^2(s)ds \right) \left(\int X^2(s)g^2(s)ds \right) \right]
 \end{aligned}$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned}
 &\mathbb{E} \left[\left(\int X(s)f(s)ds \right)^2 \left(\int X(s)g(s)ds \right)^2 \right] \\
 &\leq \left[\mathbb{E} \left(\int X(s)f(s)ds \right)^4 \right]^{1/2} \left[\mathbb{E} \left(\int X(s)g(s)ds \right)^4 \right]^{1/2} \\
 &\leq c_0 \mathbb{E} \left(\int X(s)f(s)ds \right)^2 \mathbb{E} \left(\int X(s)g(s)ds \right)^2,
 \end{aligned}$$

where the last inequality holds because of (3.6). Next, note that

$$\mathbb{E} \left[\left(\int X(s)f(s)ds \right) \left(\int X(s)g(s)ds \right) \left(\int X^2(s)f(s)g(s)ds \right) \right]$$

$$\begin{aligned} &\leq \left(\mathbb{E} \left[\left(\int X(s)f(s)ds \right)^2 \left(\int X(s)g(s)ds \right)^2 \right] \right)^{1/2} \\ &\quad \times \left(\mathbb{E} \left[\left(\int X^2(s)f(s)g(s)ds \right)^2 \right] \right)^{1/2} \end{aligned}$$

Observe that

$$\begin{aligned} &\mathbb{E} \left[\left(\int X^2(s)f(s)g(s)ds \right)^2 \right] \\ &= \int_{\mathcal{T}^2} \mathbb{E}[X^2(s)X^2(t)]f(s)g(s)f(t)g(t)dsdt \\ &\leq \int_{\mathcal{T}^2} (\mathbb{E}[X^4(s)]\mathbb{E}[X^4(t)])^{1/2} |f(s)g(s)f(t)g(t)| dsdt \\ &\leq c_0 \left(\int_{\mathcal{T}} C_0(s,s) |f(s)g(s)| ds \right)^2. \end{aligned}$$

We have

$$\begin{aligned} &\mathbb{E} \left[\left(\int X(s)f(s)ds \right) \left(\int X(s)g(s)ds \right) \left(\int X^2(s)f(s)g(s)ds \right) \right] \\ &\leq c_0 \left[\mathbb{E} \left(\int X(s)f(s)ds \right)^2 \mathbb{E} \left(\int X(s)g(s)ds \right)^2 \right]^{1/2} \\ &\quad \times \int_{\mathcal{T}} C_0(s,s) |f(s)g(s)| ds. \end{aligned}$$

Similarly,

$$\begin{aligned} &\mathbb{E} \left[\left(\int X(s)f(s)ds \right)^2 \left(\int X^2(s)g^2(s)ds \right) \right] \\ &\leq c_0 \mathbb{E} \left(\int X(s)f(s)ds \right)^2 \left(\int_{\mathcal{T}} C_0(s,s)g^2(s)ds \right), \\ &\mathbb{E} \left[\left(\int X(s)g(s)ds \right)^2 \left(\int X^2(s)f^2(s)ds \right) \right] \end{aligned}$$

$$\begin{aligned}
 &\leq c_0 \mathbb{E} \left(\int X(s)g(s)ds \right)^2 \left(\int_{\mathcal{T}} C_0(s, s)f^2(s)ds \right), \\
 &\mathbb{E} \left[\left(\int X^2(s)f^2(s)ds \right) \left(\int X^2(s)g^2(s)ds \right) \right] \\
 &\leq c_0 \left(\int_{\mathcal{T}} C_0(s, s)f^2(s)ds \right) \left(\int_{\mathcal{T}} C_0(s, s)g^2(s)ds \right).
 \end{aligned}$$

Collecting all these inequalities, we conclude that

$$\mathbb{E}(U_1^2) \leq c_0 m^4 \mathbb{E} \left(\int X(s)f(s)ds \right)^2 \mathbb{E} \left(\int X(s)g(s)ds \right)^2 + O(m^3).$$

Now consider $\mathbb{E}(U_2^2) = \mathbb{E}(U_3^2)$.

$$\begin{aligned}
 \mathbb{E}(U_2^2) &= \mathbb{E} \left(\sum_{1 \leq j_1 \neq j_2 \neq j_3 \leq m} \epsilon_{j_1}^2 X(T_{j_2})X(T_{j_3})f^2(T_{j_1})g(T_{j_2})g(T_{j_3}) \right) \\
 &\quad + \mathbb{E} \left(\sum_{1 \leq j_1 \neq j_2 \leq m} \epsilon_{j_1}^2 X^2(T_{j_2})f^2(T_{j_1})g^2(T_{j_2}) \right) \\
 &= m(m-1)(m-2)\sigma_0^2 \left(\int_{\mathcal{T}} f^2(s)ds \right) \left(\int_{\mathcal{T}^2} C_0(s, t)g(s)g(t)dsdt \right) \\
 &\quad + m(m-1)\sigma_0^2 \left(\int_{\mathcal{T}} f^2(s)ds \right) \left(\int_{\mathcal{T}} C_0(s, s)g^2(s)ds \right).
 \end{aligned}$$

At last, we look at $\mathbb{E}(U_4^2)$.

$$\begin{aligned}
 \mathbb{E}(U_4^2) &= \mathbb{E} \left(\sum_{1 \leq j_1 \neq j_2 \leq m} \epsilon_{j_1}^2 \epsilon_{j_2}^2 f^2(T_{j_1})g^2(T_{j_2}) \right) \\
 &= m(m-1)\sigma_0^4 \left(\int_{\mathcal{T}} f^2(s)ds \right) \left(\int_{\mathcal{T}} g^2(s)ds \right).
 \end{aligned}$$

In summary, we have

$$\mathbb{E}(U^2) \leq 4c_0 m^4 \mathbb{E} \left(\int X(s)f(s)ds \right)^2 \mathbb{E} \left(\int X(s)g(s)ds \right)^2 + O(m^3).$$

The second statement of the lemma follows immediately from the fact that

$$a_{kk} = \int_{\mathcal{T} \times \mathcal{T}} \varphi_k(s) \varphi_k(t) C_0(s, t) ds dt = \mathbb{E} \left(\int_{\mathcal{T}} X(t) \varphi_k(t) dt \right)^2. \quad \square$$

S3.3 Proof of Lemma 5

By definition

$$G_\lambda(\hat{C}_\lambda - \tilde{C}_\lambda) = D^2 \mathcal{L}_{\infty, \lambda}(\bar{C}_\lambda)(\hat{C}_\lambda - \tilde{C}_\lambda).$$

The first-order condition implies that

$$D \mathcal{L}_{n, m, \lambda}(\hat{C}_\lambda) = D \mathcal{L}_{n, m, \lambda}(\bar{C}_\lambda) + D^2 \mathcal{L}_{n, m, \lambda}(\bar{C}_\lambda)(\hat{C}_\lambda - \bar{C}_\lambda) = 0,$$

where we used the fact that $\mathcal{L}_{n, m, \lambda}$ is quadratic. Together with (S1.22), we have

$$\begin{aligned} D^2 \mathcal{L}_{\infty, \lambda}(\bar{C}_\lambda)(\hat{C}_\lambda - \tilde{C}_\lambda) &= D^2 \mathcal{L}_{\infty, \lambda}(\bar{C}_\lambda)(\hat{C}_\lambda - \bar{C}_\lambda) + D^2 \mathcal{L}_{\infty, \lambda}(\bar{C}_\lambda)(\bar{C}_\lambda - \tilde{C}_\lambda) \\ &= D^2 \mathcal{L}_{\infty, \lambda}(\bar{C}_\lambda)(\hat{C}_\lambda - \bar{C}_\lambda) - D^2 \mathcal{L}_{n, m, \lambda}(\bar{C}_\lambda)(\hat{C}_\lambda - \bar{C}_\lambda) \\ &= D^2 \mathcal{L}_{\infty}(\bar{C}_\lambda)(\hat{C}_\lambda - \bar{C}_\lambda) - D^2 \mathcal{L}_{n, m}(\bar{C}_\lambda)(\hat{C}_\lambda - \bar{C}_\lambda). \end{aligned}$$

Therefore,

$$\hat{C}_\lambda - \tilde{C}_\lambda = G_\lambda^{-1} \left[D^2 \mathcal{L}_{\infty}(\bar{C}_\lambda)(\hat{C}_\lambda - \bar{C}_\lambda) - D^2 \mathcal{L}_{n, m}(\bar{C}_\lambda)(\hat{C}_\lambda - \bar{C}_\lambda) \right].$$

Then

$$\left\| \hat{C}_\lambda - \tilde{C}_\lambda \right\|_a^2 = \sum_{k_1, k_2=1}^{\infty} \gamma_{k_1 k_2}^{-a} (1 + \lambda \gamma_{k_1 k_2}^{-1})^{-2}$$

$$\times \left[\frac{1}{nm(m-1)} \sum_{i=1}^n \sum_{1 \leq j \neq k \leq m} h(T_{ij}, T_{ik}) - \int_{\mathcal{T} \times \mathcal{T}} h(s, t) ds dt \right]^2,$$

where

$$h(s, t) = (\hat{C}_\lambda(s, t) - \bar{C}_\lambda(s, t)) \varphi_{k_1}(s) \varphi_{k_2}(t) \in \mathcal{W}_2^{\alpha, \alpha}.$$

Write

$$h = \sum_{j_1, j_2 \geq 1} h_{j_1 j_2} \varphi_{j_1} \otimes \varphi_{j_2}.$$

Then

$$\begin{aligned} & \left[\frac{1}{nm(m-1)} \sum_{i=1}^n \sum_{1 \leq j \neq k \leq m} h(T_{ij}, T_{ik}) - \int_{\mathcal{T} \times \mathcal{T}} h(s, t) ds dt \right]^2 \\ &= \left[\sum_{j_1, j_2 \geq 1} h_{j_1 j_2} \left(\frac{1}{nm(m-1)} \sum_{i=1}^n \sum_{1 \leq j \neq k \leq m} \varphi_{j_1}(T_{ij}) \varphi_{j_2}(T_{ik}) \right. \right. \\ & \quad \left. \left. - \int_{\mathcal{T} \times \mathcal{T}} \varphi_{j_1}(s) \varphi_{j_2}(t) ds dt \right) \right]^2 \\ &\leq \sum_{j_1, j_2 \geq 1} \gamma_{j_1 j_2}^{-b} h_{j_1 j_2}^2 \\ &\quad \times \sum_{j_1, j_2 \geq 1} \gamma_{j_1 j_2}^b \left(\frac{1}{nm(m-1)} \sum_{i=1}^n \sum_{1 \leq j \neq k \leq m} \varphi_{j_1}(T_{ij}) \varphi_{j_2}(T_{ik}) \right. \\ & \quad \left. - \int_{\mathcal{T} \times \mathcal{T}} \varphi_{j_1}(s) \varphi_{j_2}(t) ds dt \right)^2. \end{aligned}$$

For each curve, write $F_{jk} = \varphi_{j_1}(T_{ij}) \varphi_{j_2}(T_{ik})$ for $j \neq k$, and let

$$U_i = \frac{1}{m(m-1)} \sum_{1 \leq j \neq k \leq m} F_{jk}.$$

Independence of the design points implies that $\text{Cov}(F_{jk}, F_{j'k'}) = 0$ whenever

the two index pairs are disjoint. Moreover, $\mathbb{E} F_{jk}^2 = 1$, so $|\text{Cov}(F_{jk}, F_{j'k'})| \leq$

1 by the Cauchy–Schwarz inequality. There are at most cm^3 ordered pairs of index pairs with a common index. Therefore,

$$\begin{aligned} \mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n U_i - \mathbb{E} U_1 \right)^2 &= \frac{1}{n} \text{Var}(U_1) \\ &\leq \frac{cm^3}{nm^2(m-1)^2} \leq \frac{c}{nm}. \end{aligned}$$

Since $\alpha b > 1/2$, multiplication is bounded in the tensor product Sobolev space corresponding to $\|\cdot\|_b$. In particular, $\|uv\|_b \leq c\|u\|_b\|v\|_b$. Since $1/(2\alpha) < b < 1/2$ and $0 \leq a \leq b$, we have $a + b + 1/(2\alpha) < 2$. This ensures convergence of the spectral sum below. Therefore,

$$\begin{aligned} \left\| \hat{C}_\lambda - \tilde{C}_\lambda \right\|_a^2 &\leq O_p \left(\frac{1}{nm} \right) \sum_{k_1, k_2 \geq 1} \gamma_{k_1 k_2}^{-a} (1 + \lambda \gamma_{k_1 k_2}^{-1})^{-2} \|h\|_b^2 \\ &\leq O_p \left(\frac{1}{nm} \right) \sum_{k_1, k_2 \geq 1} \gamma_{k_1 k_2}^{-a} (1 + \lambda \gamma_{k_1 k_2}^{-1})^{-2} \\ &\quad \times \left\| \hat{C}_\lambda - \bar{C}_\lambda \right\|_b^2 \|\varphi_{k_1} \otimes \varphi_{k_2}\|_b^2 \\ &= O_p \left(\frac{1}{nm} \right) \left\| \hat{C}_\lambda - \bar{C}_\lambda \right\|_b^2 \sum_{k_1, k_2 \geq 1} (1 + \lambda \gamma_{k_1 k_2}^{-1})^{-2} \gamma_{k_1 k_2}^{-(a+b)} \\ &= O_p \left(\frac{\log(1/\lambda)}{nm \lambda^{a+b+1/(2\alpha)}} \right) \left\| \hat{C}_\lambda - \bar{C}_\lambda \right\|_b^2. \quad \square \end{aligned}$$

S4. Additional Numerical Results

S4.1 Simulation Study

The following figures provide representative covariance and eigenfunction estimates under the independent and common designs. They complement the replicated-error summaries and the comparison with PACE reported in the main paper.

Even with $m = 5$ observations per curve, the proposed estimator captures the main covariance structure. The estimates improve as m increases under both designs.

S4.2 Electricity Load Diagrams Dataset

We further apply our method to the Electricity Load Diagrams Dataset (Trindade, 2015), which records electricity consumption at 15-minute intervals for 370 clients over multiple years. Daily load curves are treated as functional observations on a 24-hour domain with regular sampling, corresponding to a common design. Estimating the covariance function provides critical insights into the stochastic structure and temporal dependencies underlying electricity consumption patterns of different clients throughout the day.

Figure S3 on page 43 presents covariance estimates and leading eigen-

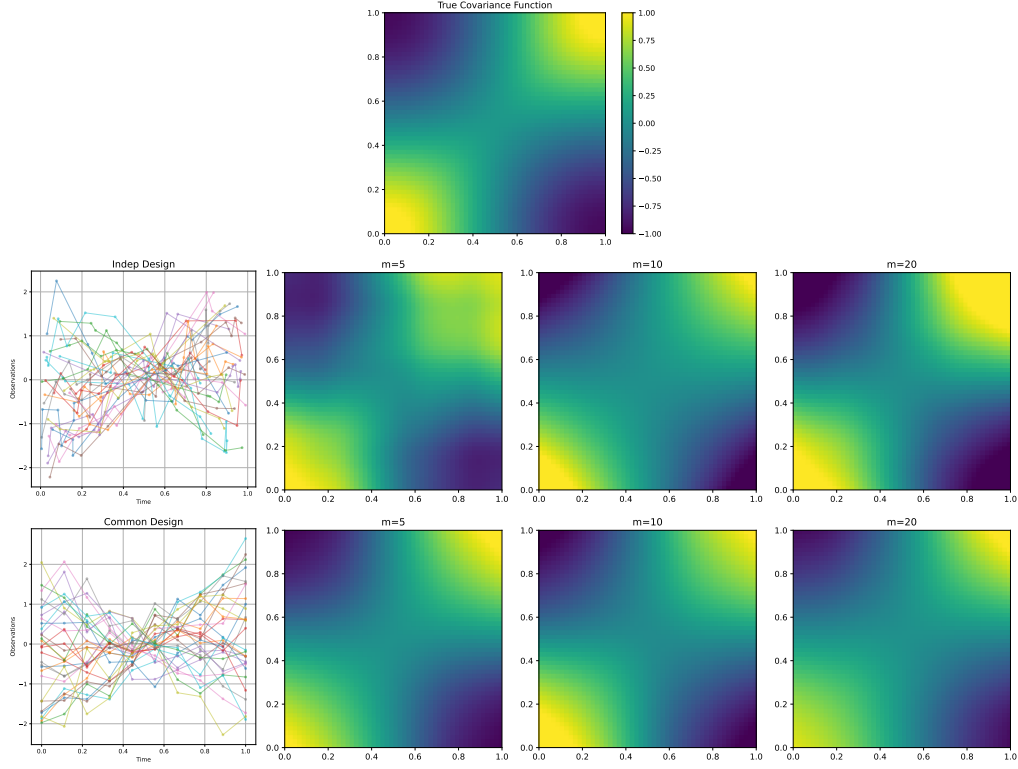


Figure S1: Covariance function estimation with different sampling designs and frequencies. The true covariance function is shown at the top. The second and third rows correspond to the independent and common designs, respectively, with $m = 5, 10, 20$ from left to right.

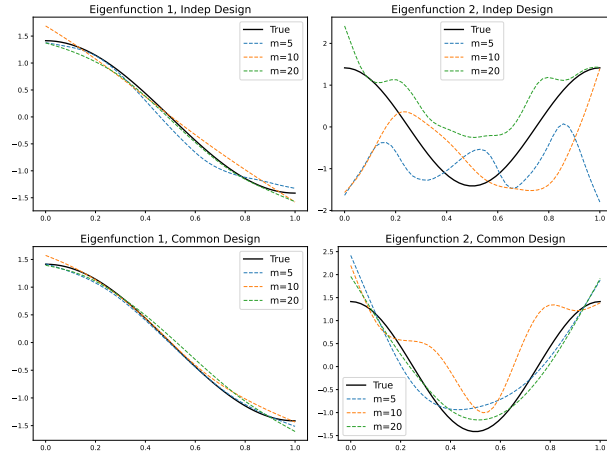


Figure S2: Estimated first two eigenfunctions under the independent design (top) and common design (bottom) with different sampling frequencies.

functions for two representative clients. The two clients have quite different covariance structures and eigenfunctions, reflecting distinct consumption behaviors. For the first client (MT_006), the covariance function shows large positive values during evening hours (6 PM to 10 PM), indicating consistent usage patterns likely driven by residential activities. The first eigenfunction captures this trend, exhibiting a pronounced peak during these hours. For the second client (MT_009), the covariance function exhibits large positive values during working hours (9 AM to 9 PM), indicating consistent usage patterns likely driven by business operations. Notably, a valley appears around noon, suggesting a dip in consumption possibly due to lunch breaks or reduced activity. The first eigenfunction reflects the overall consumption

S4. ADDITIONAL NUMERICAL RESULTS

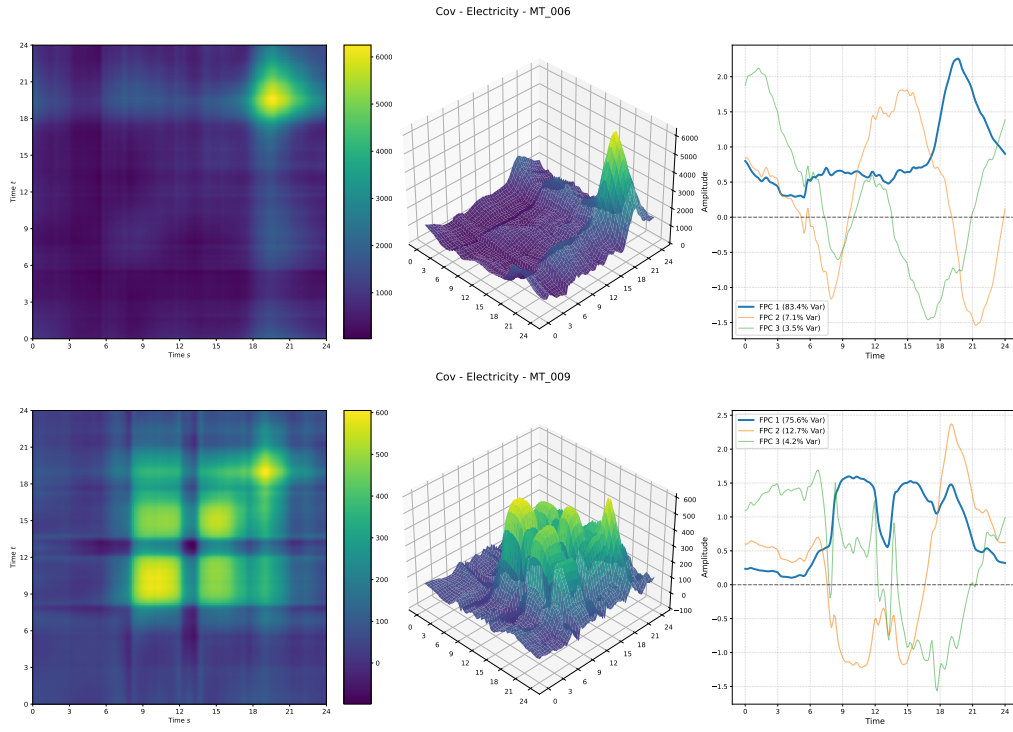


Figure S3: Estimated covariance functions and eigenfunctions for electricity consumption of two clients.

trend, while the second eigenfunction captures the daytime dip and the evening workload, indicating potential overtime activities.

These results demonstrate the effectiveness of our covariance function estimation method in capturing the underlying temporal dependencies and variability in electricity consumption patterns. The estimated covariance functions and eigenfunctions provide valuable insights into client-specific consumption behaviors, which can inform demand response strategies, energy management practices, and capacity planning that would benefit both utilities and consumers.

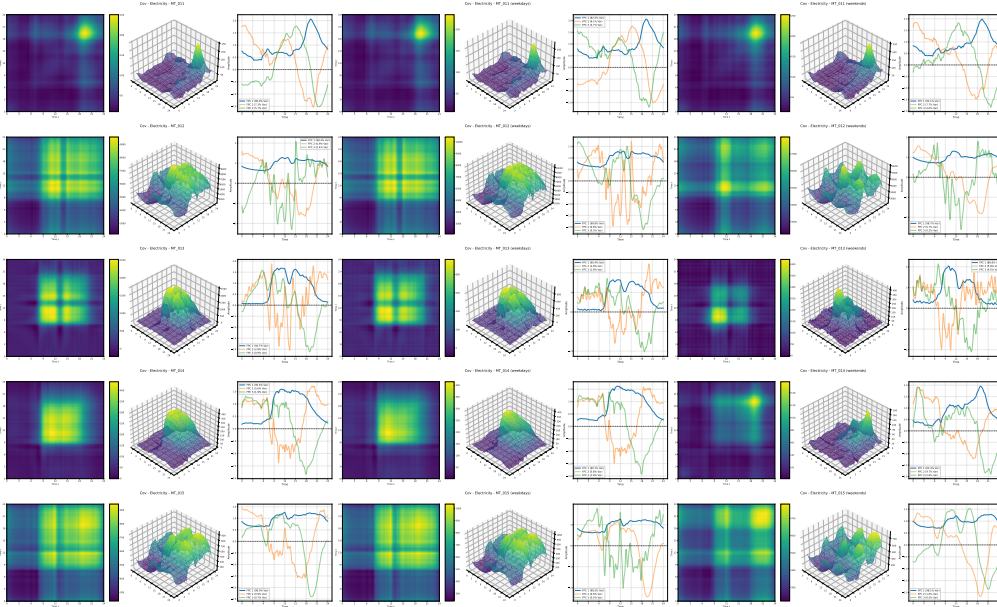


Figure S4: Additional estimated covariance functions for the Electricity Load Diagrams Dataset, corresponding to meters 11 to 15. Each row displays the overall estimate, weekday estimate, and weekend estimate for a specific meter.

S4.3 Air Quality Dataset

The Air Quality Dataset (Vito, 2008) from the UCI Machine Learning Repository records hourly averaged measurements from a multisensor air quality chemical device deployed at road level in a significantly polluted urban area of an Italian city, spanning March 2004 to February 2005 (one full year). It contains 9,358 instances, each corresponding to an hourly observation, with 15 attributes including timestamp (date and time), ground-truth concentrations of key pollutants measured by certified reference analyzers, responses from five embedded metal oxide chemical sensors, and meteorological variables (temperature in °C, relative humidity in %, absolute humidity).

This dataset is well-suited for functional data analysis due to its high-resolution time-series structure, where measurements can be viewed as multivariate functional curves over a daily domain (24 hourly points per day) or longer horizons (e.g., weekly or seasonal patterns). In our application, we focus on one pollutant concentration or sensor response as the response variable, treating each day’s profile as a functional observation over the 24-hour interval with fixed, regularly spaced sampling points ($m = 24$).

Results are reported in Figure S5 on page 46. For temperature (T), we observe large positive covariance during the day, while for absolute humidity

(AH), the covariance is larger at night. For benzene (C_6H_6) and carbon monoxide (CO) concentrations, we observe large positive covariance during working hours, particularly during rush hours, indicating the influence of traffic emissions. For nitrogen oxides (NO_x) and ozone sensor response (O₃), the covariance patterns are more spread out throughout the day, reflecting their complex formation processes influenced by various factors, including sunlight and precursor emissions.

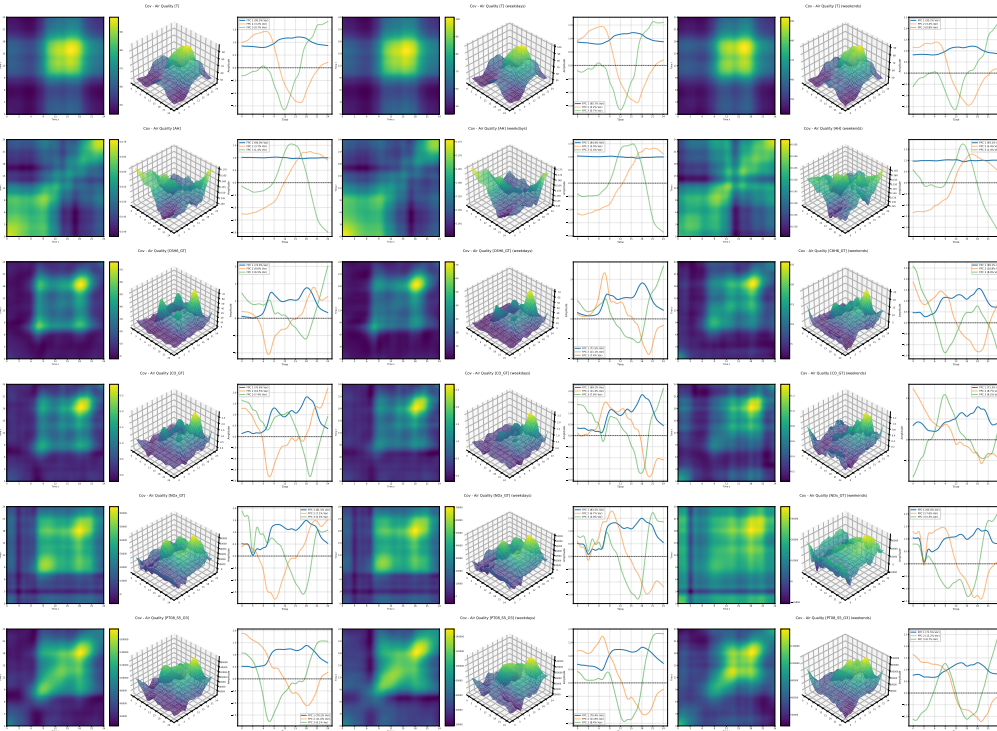


Figure S5: Estimated covariance functions for the Air Quality Dataset. Each row displays the overall estimate, weekday estimate, and weekend estimate for a specific variable.

S4.4 Longitudinal CD4 Count

CD4+ T cell counts are a critical biomarker for monitoring the immune function of individuals infected with HIV (Kaslow et al., 1987; Diggle et al., 2002). We analyze the longitudinal CD4 T cell count data from the AIDS Clinical Trials Group (ACTG) 388 study (Fischl et al., 2003), a multicenter randomized clinical trial sponsored by the National Institute of Allergy and Infectious Diseases (NIAID/NIH). The full ACTG 388 trial enrolled 517 HIV-1-infected patients and compared three antiretroviral treatment regimens, including highly active antiretroviral therapy (HAART) combinations.

One of the main objectives of the data analysis is to characterize the time course of the cell counts. This can be accomplished by examining the covariance function. The results are reported in Figure S6 on page 49. From a statistical modeling perspective, it is of particular interest to determine whether the time course can be modeled by a stationary process. From the covariance function estimate, however, this does not appear to be the case.

Additionally, we stratify the data by treatment groups (trt1, trt2, trt3) and estimate the covariance functions for each group separately. The results, also shown in Figure S6 on page 49, indicate that the covariance structures differ across treatment groups, suggesting that the treatment

regimens may have distinct effects on the variability and covariance patterns of CD4 counts over time.

S4.5 NREL Solar Integration National Dataset

The National Renewable Energy Laboratory (NREL) Solar Power Data for Integration Studies provides high-resolution (5-minute) synthetic photovoltaic (PV) power production time-series data for approximately 6,000 simulated PV plants across the United States, based on weather conditions for the year 2006. This dataset supports grid integration studies, capacity planning, and renewable energy modeling by offering high-resolution estimates of solar power output from hypothetical utility-scale PV (UPV, with single-axis tracking) and distributed PV (DPV, fixed tilt) plants. In our application, we treat the power output profiles as functional curves over a daily domain and consider only the daytime hours (6 AM to 6 PM).

We first investigate how the power output covariance structure varies across different geographic locations. We select two typical DPV sites from the dataset: one in Philadelphia, PA, and the other in San Francisco, CA. The estimated covariance functions for these two sites are presented in Figure S7 on page 51. The covariance structures are very similar between the two locations, both exhibiting large values along the diagonal (indicat-

Data are publicly available at <https://www.nrel.gov/grid/solar-power-data>

S4. ADDITIONAL NUMERICAL RESULTS

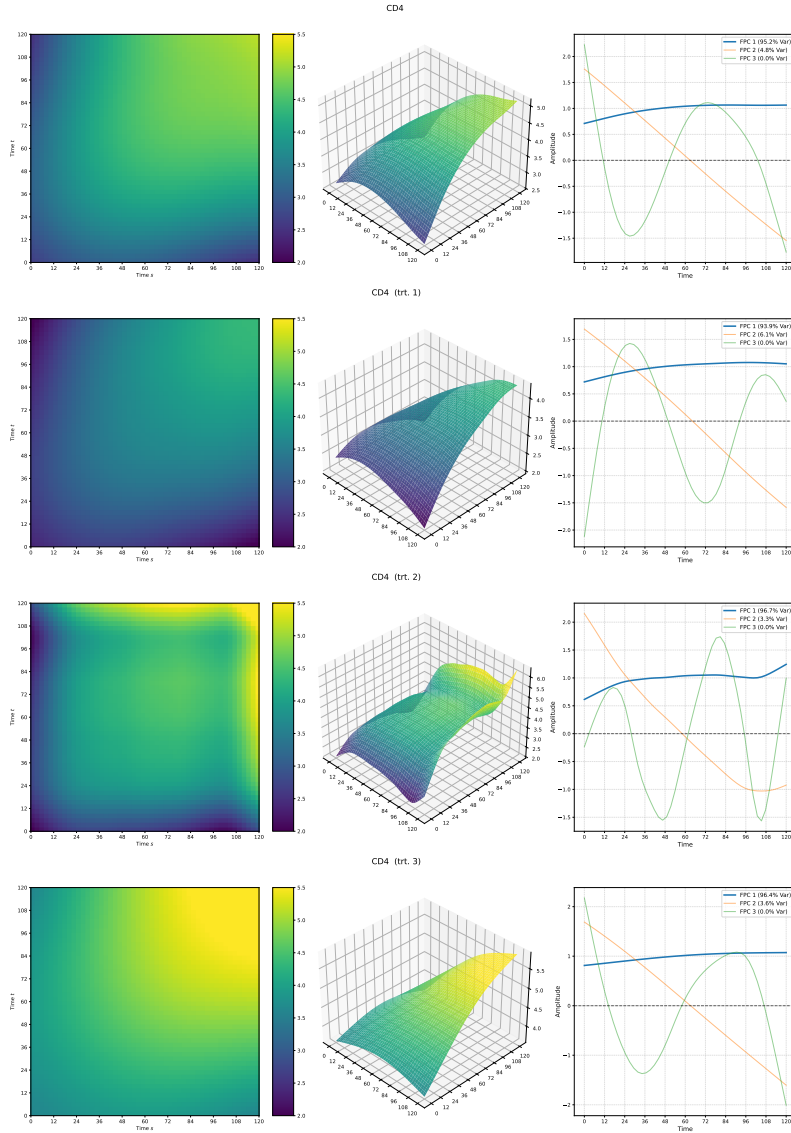


Figure S6: Estimated covariance functions for the longitudinal CD4 count data. The top panel shows the overall estimate, while the bottom three panels display estimates stratified by treatment groups (trt1, trt2, trt3). The color scale is shared across all plots for comparability.

ing strong co-variation in power output at similar times of the day) and a decrease in covariance as the time difference increases. The covariance function also appears somewhat flatter in San Francisco, CA, which may be attributed to the more consistent solar irradiance patterns in that region compared to Philadelphia, PA.

Now we investigate the impact of different PV system types on the covariance structure. UPV represents large, centralized utility-scale PV plants modeled with single-axis tracking, which follows the sun to maximize daily energy capture, leading to higher output, a more extended midday peak, and smoother profiles due to spatial averaging over larger areas. In contrast, DPV models smaller, distributed systems (e.g., rooftop or community solar) with fixed-tilt panels and additional low-pass filtering to simulate the aggregated smoothing effect of many dispersed installations, resulting in lower capacity factors, sharper ramps, and slightly more pronounced short-term variability despite the filtering. In Figure S8 on page 52, we can clearly observe distinct diurnal shapes and covariance structures: UPV has much smaller variability but relatively stronger midday covariance, while DPV curves show more variability and weaker long-range temporal dependence.

S4. ADDITIONAL NUMERICAL RESULTS

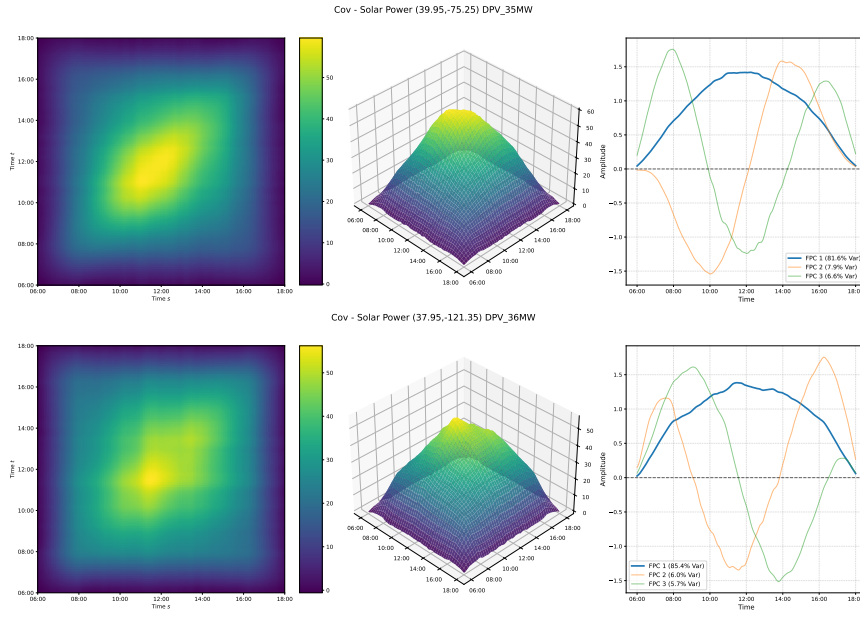


Figure S7: Estimated covariance functions for the NREL Solar Integration National Dataset. The top panel corresponds to the site in Philadelphia, PA, while the bottom panel corresponds to the site in San Francisco, CA.

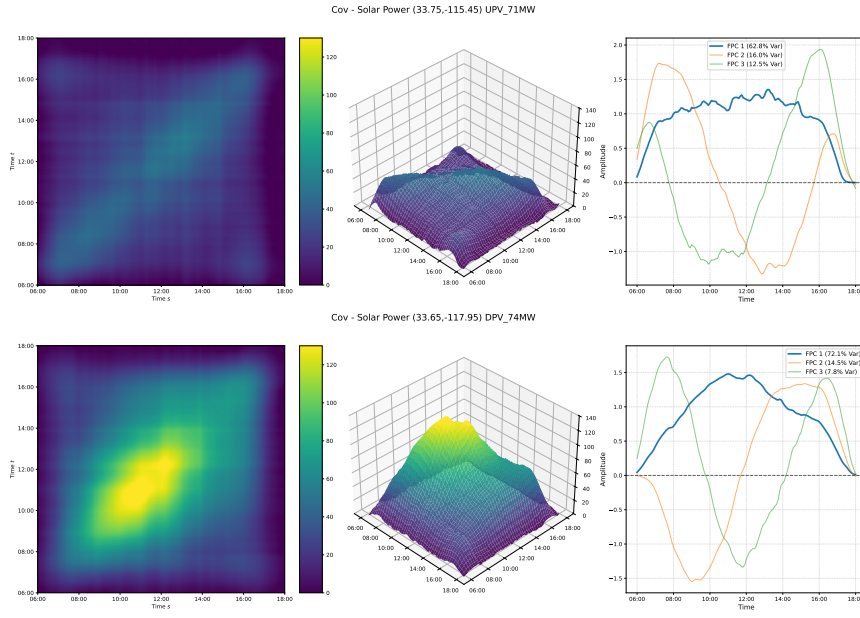


Figure S8: Comparison of estimated covariance functions for different PV system types in Southern California. The top panel corresponds to a utility-scale PV (UPV) system, while the bottom panel corresponds to a distributed PV (DPV) system.

Bibliography

- Bhatia, R., Davis, C. and McIntosh, A. (1983). Perturbation of spectral subspaces and solution of linear operator equations. *Linear Algebra and its Applications* **52–53**, 45–67.
- Cai, T. T. and Yuan, M. (2011). Optimal estimation of the mean function based on discretely sampled functional data: Phase transition. *The Annals of Statistics* **39**, 2330–2355.
- Caponnetto, A. and Yao, Y. (2010). Cross-validation based adaptation for regularization operators in learning theory. *Analysis and Applications* **8**, 161–183.
- Diggle, P. J., Heagerty, P. J., Liang, K.-Y. and Zeger, S. L. (2002). *Analysis of Longitudinal Data* (2nd ed.), Volume 25 of *Oxford Statistical Science Series*. Oxford: Oxford University Press.
- Fischl, M. A., Ribaud, H. J., Collier, A. C., Erice, A., Giuliano, M., Dehlenger, M. et al. (2003). A randomized trial of 2 different 4-drug antiretroviral regimens versus a 3-drug regimen, in advanced human immunodeficiency virus disease. *The Journal of Infectious Diseases* **188**, 625–634.
- Kaslow, R. A., Ostrow, D. G., Detels, R., Phair, J. P., Polk, B. F., Rinaldo,

- Jr., C. R. et al. (1987). The multicenter AIDS cohort study: Rationale, organization, and selected characteristics of the participants. *American Journal of Epidemiology* **126**, 310–318.
- Schumaker, L. L. (2007). *Spline Functions: Basic Theory* (3rd ed.). Cambridge Mathematical Library. Cambridge: Cambridge University Press.
- Trindade, A. (2015). ElectricityLoadDiagrams20112014. UCI Machine Learning Repository. <https://doi.org/10.24432/C58C86>.
- Tsybakov, A. B. (2009). *Introduction to Nonparametric Estimation* (1st ed.). Springer Series in Statistics. New York, NY: Springer.
- Vito, S. (2008). Air quality. UCI Machine Learning Repository. <https://doi.org/10.24432/C59K5F>.
- Wahba, G. (1990). *Spline Models for Observational Data*. Philadelphia, PA: Society for Industrial and Applied Mathematics.