Statistica Sinica

MINIMAX AND ADAPTIVE COVARIANCE FUNCTION ESTIMATION IN FUNCTIONAL DATA ANALYSIS

T. Tony Cai, Yicheng Li, and Ming Yuan

University of Pennsylvania, East China Normal University and Columbia University

Abstract: The covariance function is a central object in functional and longitudinal data analysis, yet its optimal estimation remains challenging because of its bivariate structure and its sensitivity to the sampling design. We study covariance function estimation for functional data under two distinct designs: the *common design*, in which all curves are observed on a shared grid, and the *independent design*, in which the sampling locations vary across curves. By exploiting the tensor-product structure of the covariance function, we propose a unified smoothing-spline estimator that applies to both settings.

We establish minimax optimal convergence rates for the proposed estimator under both sampling designs and identify several phase transition phenomena. In particular, under the independent design with sparse sampling, the estimator attains a nearly "univariate" minimax rate of $(nm)^{-2\alpha/(2\alpha+1)}$, up to a logarithmic factor, despite the intrinsically bivariate nature of the covariance function. By contrast, under the common design, the sparse-regime rate is dominated by the discretization error $m^{-2\alpha}$. We further develop an adaptive implementation based on cross-validation for selecting the smoothing parameter, which does not require prior knowledge of the smoothness parameter α or the sampling-frequency regime. Numerical experiments on both simulated and real datasets support the theoretical findings and demonstrate the practical effectiveness of the proposed method.

Keywords and phrases: Covariance function, minimax rate, functional data analysis, method of regularization, phase transition, Sobolev space, tensor product space.

1. Introduction

The covariance function is a fundamental object in functional and longitudinal data analysis. Many statistical methodologies—including functional clustering, functional principal component analysis, functional linear discriminant analysis, and functional linear regression—depend critically on accurate covariance estimation; see, for example, Diggle et al. [2002], Ramsay and Silverman [2005], Ferraty et al. [2012], Koner and Staicu [2023]. The covariance function characterizes dependence across locations and encodes essential features of the underlying stochastic process, including the smoothness and regularity of its sample paths; see, e.g., Stein [1999]. Consequently, estimating the covariance function from random samples is a problem of fundamental importance.

In the idealized setting where each trajectory is observed continuously and without measurement error, the sample covariance achieves the usual parametric rate of convergence under mild regularity conditions, for example under integrated squared error; see, e.g., Bosq [2000]. In practice, however, such complete and noiseless observations are rarely available. In many applications, especially longitudinal studies, each curve is observed only at a finite number of discrete locations and is further contaminated by measurement error. This realistic observation regime fundamentally changes the statistical difficulty of covariance estimation. In this paper, we study covariance function estimation from such discrete and noisy functional observations.

Let $X(t)$ be a second-order (square-integrable) stochastic process on an interval $\mathcal{T} \subset \mathbb{R}$, e.g., $\mathcal{T} = [0, 1]$. Its covariance function is

$$C_0(s,t) = \mathbb{E}\big[\,(X(s) - \mu_0(s))\,(X(t) - \mu_0(t))\,\big], \qquad s, t \in \mathcal{T}, \tag{1.1}$$

where $\mu_0(t) = \mathbb{E}[X(t)]$. We observe n independent realizations $X_1, \ldots, X_n$ at subject-specific locations with measurement error:

$$Y_{ij} = X_i(T_{ij}) + \varepsilon_{ij}, \qquad j = 1, \ldots, m, \quad i = 1, \ldots, n, \tag{1.2}$$

where the measurement errors ε_{ij} are independent with mean zero and variance σ_0^2. We further assume that the process X, the sampling locations T_{ij}, and the errors ε_{ij} are mutually independent. Such a model is commonly adopted and suitable for a large number of applications in functional and longitudinal data analysis (see, e.g., Shi et al. [1996], Staniswalis and Lee [1998], James and Hastie [2001], Diggle et al. [2002], Ramsay and Silverman [2005], Hsing and Eubank [2015]). For simplicity we assume a common per-curve sample size m; this is not essential and is adopted only to highlight the interplay between the number of curves n and the sampling frequency m. More general sampling schemes are discussed in Section 5.

A substantial literature studies covariance function and FPCA estimation from noisy, discretely observed trajectories under (1.2); see, among many others, James et al. [2000], Rice and Wu [2001], Yao et al. [2005], Hall et al. [2006], Paul and Peng [2009], Li and Hsing [2010]. More recently, for covariance function estimation, Xiao et al. [2018] proposed fast covariance smoothing for sparse functional data, and Xiao [2020] studied the asymptotic properties of penalized spline estimators for functional data. Zhang and Wang [2016] developed a sparse-to-dense asymptotic theory for local linear mean and covariance estimators, while Petersen [2024], Guo et al. [2025] investigated related phase-transition phenomena for discretely observed high-dimensional functional data. For partially observed trajectories, Descary and Panaretos [2019] established nonparametric recovery under fragmented designs via a low-rank matrix-completion viewpoint, and

Delaigle et al. [2021], Lin and Wang [2022] developed covariance estimators for fragmented or snippet data with substantial off-diagonal missingness. For eigenfunction and FPCA estimation, Zhou et al. [2025] developed a theory for noisy discretely observed data that permits the eigenfunction index to diverge and identifies the sampling frequencies at which the fully observed rate is recovered. Adjacent work addresses different statistical targets: Tan et al. [2026] proposed functional SVD for dimension reduction of heterogeneous trajectories without requiring a common covariance structure, whereas Zhou et al. [2026] developed covariance testing procedures for noisy discretely observed functional data with a growing number of eigenprojections.

Among the various methods, a common strategy for estimating C_0 is a two-step approach: first smooth each curve, and then form the sample covariance of the reconstructed trajectories (e.g., Ramsay and Silverman [2005]). While intuitive, its theoretical justification is limited and often relies on restrictive sampling regimes. For instance, Hall et al. [2006] studied twice differentiable sample paths and showed that when trajectories are sufficiently dense (e.g., $m \gg n^{1/4+\delta}$ for some $\delta > 0$) the two-step estimator can achieve the parametric rate n^{-1} under integrated squared error, whereas when m is fixed, treating covariance estimation as a bivariate smoothing problem yields the usual nonparametric rate $n^{-2/3}$. More recently, Zhang and Wang [2016] developed a unified asymptotic theory for local linear smoothers under varying numbers of observations per subject, including general weighting schemes such as OBS (same weight to each observation) and SUBJ (same weight to each subject). Follow-up works include Lin and Wang [2022], Petersen [2024], Guo et al. [2025]. Nonetheless, it remains unclear how estimation accuracy depends jointly on (n, m) beyond twice differentiability, and whether existing upper bounds are minimax optimal. Addressing these questions is important for developing methods that are both broadly applicable and theoretically justified.

Another important aspect is the impact of the *sampling design* on estimation accuracy. Two designs are particularly common in applications [Cai and Yuan, 2011]:

1. **Common fixed design:** the sampling locations are fixed and identical across subjects, i.e., $T_{ij} = T_j$ for all i.

2. **Independent random design:** the locations T_{ij} are independent across i and j.

Cai and Yuan [2011] shows that the optimal rates for estimating the mean function differ sharply between these two designs and exhibit phase transitions as m varies. However, whether analogous phenomena occur for covariance function estimation is much less understood.

In this paper, we provide a general method for estimating the covariance function within a unified framework using tensor product smoothing splines in Sobolev spaces that is applicable to various sampling frequencies and designs. The proposed method is motivated by a careful examination of the function space in which the covariance function C_0 resides. By assuming that the sample path of X belongs to a Sobolev space, we show that C_0 necessarily belongs to a tensor product space naturally connected with the smoothness of the sample path of X. This distinguishes covariance estimation from conventional bivariate smoothing, and leads to a regularization scheme that exploits the tensor structure. Moreover, we propose a cross-validation procedure for selecting the smoothing parameter, which allows for adaptive estimation without prior knowledge of the smoothness parameter α or the sampling frequency regime.

In terms of theoretical properties, we aim to characterize the *optimal rates of convergence* for both common and independent designs and elucidate the *phase transition* phenomena that occur as the sampling frequency m varies relative to the sample size n. Under the independent design, the estimator converges to C_0 at the rate (omitting logarithmic factors)

$(nm)^{-2\alpha/(2\alpha+1)} + n^{-1}$ under integrated squared error, assuming that the sample paths of X are $\alpha(> 1/2)$ times differentiable; under the common design, the rate becomes $O_p(m^{-2\alpha} + n^{-1})$. Corresponding minimax lower bounds are also established to show that these rates are optimal. Interestingly, despite the bivariate nature of C_0, the tensor structure implies that the price for estimating C_0, relative to estimating a univariate mean function, is at most logarithmic.

The rate also exhibits an interesting phase transition effect similar to those observed when studying mean function estimation in functional data analysis [Cai and Yuan, 2011]. Namely, under the independent design, when the functions are sparsely sampled, i.e., $m \ll n^{1/(2\alpha)} \log n$, the convergence rates are determined by the total number of observations $N := nm$. On the other hand, when the functions are densely sampled, i.e., $m \gg n^{1/(2\alpha)} \log n$, the rates remain $1/n$ regardless of m. Under the common design, the situation is quite different: when $m \ll n^{1/(2\alpha)}$, the rate is dominated by the discretization error $m^{-2\alpha}$ and does not depend on n; when $m \gg n^{1/(2\alpha)}$, the rate becomes n^{-1}. Perhaps more surprisingly, our results suggest that the proposed method is to a certain degree immune to the "curse of dimensionality". Since covariance functions are bivariate functions, one would naturally anticipate nonparametric estimation of C_0 to be more difficult than estimating a univariate function such as the mean function. It is, however, somewhat unexpected to see that the only price we pay here is at most a logarithmic factor.

Despite the generality of the method, we show that the estimators can be efficiently computed thanks to a representer theorem, which makes our procedure easily implementable and enables us to take advantage of existing techniques and algorithms for smoothing splines. Moreover, we show that our estimate of the covariance function enables explicit calculation of the functional principal components and therefore avoids the usual numerical approximation (see, e.g., Ramsay and Silverman [2005]) where one has to trade numerical accuracy for computational efficiency. The numerical performance of the estimator is investigated using both simulated and real data.

The rest of the paper is organized as follows. Section 2 introduces the methodology for estimating the covariance function based on tensor product smoothing splines. Section 3 establishes the rates of convergence and discusses the phase transition. The implementation of the proposed method and numerical experiments are presented in Section 4.

2. Methodology

In this section, we introduce a regularization method for estimating the covariance function from discretely sampled noisy data. The key motivation is a structural observation: when the sample paths of the underlying process lie in a Sobolev space, the covariance function does not merely belong to a generic bivariate function class, but instead resides in a specific tensor product Sobolev space. This tensor structure fundamentally distinguishes covariance function estimation from conventional bivariate smoothing and motivates a regularization framework based on tensor product smoothing splines.

2.1 Covariance Function and Tensor Product Spaces

Let $W_2^{\alpha}(\mathcal{T})$ denote the Sobolev space of order α on $\mathcal{T} = [0, 1]$. When α is a positive integer, $W_2^{\alpha}(\mathcal{T})$ consists of functions whose derivatives up to order $\alpha - 1$ are absolutely continuous and whose α-th derivative belongs to $L_2(\mathcal{T})$. More generally, $W_2^{\alpha}(\mathcal{T})$ can be defined for non-integer α via standard interpolation. Throughout this paper, we assume that the sample paths of X belong to $W_2^{\alpha}(\mathcal{T})$ for some $\alpha > 1/2$, which in particular guarantees continuity.

Since the sample paths of X belong to $W_2^{\alpha}(\mathcal{T})$, the covariance function

$C_0(s,t)$ inherits a tensor product smoothness structure. This is formalized in the following proposition.

Proposition 1. *If the sample paths of X belong to $\mathcal{W}_2^\alpha(\mathcal{T})$ almost surely and $\mathbb{E}\,\|X\|^2_{\mathcal{W}_2^\alpha} < \infty$, then the covariance function C_0 belongs to $\mathcal{W}_2^\alpha(\mathcal{T}) \otimes \mathcal{W}_2^\alpha(\mathcal{T})$ and satisfies*

$$\|C_0\|_{\mathcal{W}_2^\alpha \otimes \mathcal{W}_2^\alpha} \le \mathbb{E}\,\|X\|^2_{\mathcal{W}_2^\alpha}. \tag{2.3}$$

2.2 Regularization Estimator

Proposition 1 reveals a key structural property of the covariance function C_0: it belongs to the tensor product Sobolev space $\mathcal{W}_2^\alpha \otimes \mathcal{W}_2^\alpha$. This observation motivates us to incorporate the tensor structure explicitly into the estimation procedure, leading to a regularization framework tailored to covariance function estimation.

It is easy to see that the random vector $Y_{i\cdot} := (Y_{i1}, \ldots, Y_{im})'$ has mean $\mu_{i\cdot} := (\mu_0(T_{i1}), \ldots, \mu_0(T_{im}))$ and covariance matrix $\Sigma_i := [C_0(T_{ij_1}, T_{ij_2})]_{1 \le j_1, j_2 \le m} + \sigma_0^2 I_m$. If the mean function μ_0 is known, one can regress $[Y_{ij_1} - \mu_0(T_{ij_1})][Y_{ij_2} - \mu_0(T_{ij_2})]$ on (T_{ij_1}, T_{ij_2}) to estimate C_0. If we have a good estimator $\hat{\mu}(\cdot)$ of the mean function μ_0, we can replace μ_0 by $\hat{\mu}$ in the regression. For simplicity, we assume that $\hat{\mu}$ is constructed from independent auxiliary samples. In the light of Proposition 1, we seek an estimator of C_0 within the tensor product Sobolev space $W_2^\alpha(\mathcal{T}) \otimes W_2^\alpha(\mathcal{T})$, as given by the following regularization problem:

$$\hat{C}_\lambda = \underset{C \in \mathcal{W}_2^\alpha(\mathcal{T}) \otimes \mathcal{W}_2^\alpha(\mathcal{T})}{\arg\min} \left[\mathcal{L}_{n,m}(C) + \lambda \|C\|^2_{\mathcal{W}_2^\alpha \otimes \mathcal{W}_2^\alpha} \right], \tag{2.4}$$

where

$$\mathcal{L}_{n,m}(C) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{m(m-1)} \sum_{1 \leq j \neq k \leq m} \left[(Y_{ij} - \hat{\mu}(T_{ij}))(Y_{ik} - \hat{\mu}(T_{ik})) - C(T_{ij}, T_{ik}) \right]^2, \tag{2.5}$$

and $\lambda \geq 0$ is a tuning parameter that balances the fidelity to the data measured by $\mathcal{L}_{n,m}$ and smoothness of the estimate measured by the squared norm. By excluding the diagonal terms, we avoid the bias that would otherwise be introduced by the noise variance σ_0^2.

The problem of estimating the mean function μ_0 has been extensively studied. In particular, Cai and Yuan [2011] investigated the optimal rates for estimating μ_0 under various settings. It was shown that with appropriate tuning parameters, the smoothing spline estimator obtained by pooling all observations and regressing Y_{ij} on T_{ij} achieves the optimal rate of convergence. We shall therefore adopt the proposed spline estimate. Interested readers are referred to Cai and Yuan [2011] for more detailed discussions on optimal estimation of the mean function.

Remark 1. Other loss functions may also be used in place of $\mathcal{L}_{n,m}$ to measure the goodness of the fit to the data in (2.4). One particular example occurs when the random function X follows a Gaussian process and the measurement errors ε follow a centered normal distribution. In this case, it is not hard to see that the random vector $Y_{i\cdot}$ follows a multivariate normal distribution. It is then natural to use the negative log-likelihood instead of $\mathcal{L}_{n,m}$ in (2.4). Here we opt for $\mathcal{L}_{n,m}$ because of its general applicability.

3. Rates of Convergence

In this section, we establish the theoretical properties of the proposed covariance estimator $\hat{C}_\lambda$. We first derive upper bounds on the convergence rates under the common and independent sampling designs, highlighting distinct phase transition phenomena as the sampling frequency m varies relative to the number of curves n. We then complement the upper bounds with matching minimax lower bounds, showing that the obtained rates are nearly optimal over appropriate Sobolev classes. Let $\mathcal{P}(\alpha; M_0, c_0)$ be the collection of probability measures of X such that

(a) the sample path of X belongs to $\mathcal{W}_2^\alpha(\mathcal{T})$ almost surely and $\mathbb{E}\|X\|^2_{\mathcal{W}_2^\alpha(\mathcal{T})} \leq M_0$;

(b) there exists a numerical constant $c_0 \geq 1$ such that $\mathbb{E}X^4(t) \leq c_0[\mathbb{E}X^2(t)]^2$

for any $t \in \mathcal{T}$, and

$$\mathbb{E}\left(\int_{\mathcal{T}} X(t)f(t)dt\right)^4 \leq c_0 \left(\mathbb{E}\left(\int_{\mathcal{T}} X(t)f(t)dt\right)^2\right)^2, \tag{3.6}$$

for any $f \in L_2(\mathcal{T})$.

The first condition essentially specifies the smoothness of X; the second condition concerns the fourth moment of X and is satisfied with $c_0 = 3$ when X follows a Gaussian process.

3.1 Upper Bounds

In the following, we establish upper bounds for the convergence rate of $\hat{C}_\lambda$ under both common and independent designs. For the common fixed design, we make the following assumption on the sampling locations. It is standard in the analysis of functional data under the common fixed design and ensures that the discretization error is well controlled; see, for example, Cai and Yuan [2011].

Assumption 1 (Common Design)**.** The sampling locations are identical across all subjects, i.e., $T_{ij} = T_j$ for all i, and the grid points $0 = T_0 \leq T_1 \leq T_2 \leq \cdots \leq T_m \leq T_{m+1} = 1$ are fixed and satisfy

$$\max_{0 \leq j \leq m} (T_{j+1} - T_j) \leq \frac{C_T}{m}, \tag{3.7}$$

for some constant $C_T > 0$.

The convergence rate of the covariance estimator depends on the accuracy of the mean estimator at the common design points. The following theorem shows that, provided $\hat{\mu}$ attains the usual optimal rate under the common design over the fixed grid, the proposed covariance estimator achieves the minimax optimal rate.

Theorem 1 (Upper Bound under Common Design)**.** *Suppose that $X(\cdot)$ follows a distribution in $\mathcal{P}(\alpha; M_0, c_0)$ and that Assumption 1 holds. Assume further that $\mathbb{E}\varepsilon_{ij}^4 < \infty$. Let $\hat{\mu}$ be an estimator of the mean function μ_0 such that*

$$\frac{1}{m}\sum_{j=1}^{m}\left(\hat{\mu}(T_j) - \mu_0(T_j)\right)^2 = O_p\left(m^{-2\alpha} + n^{-1}\right). \tag{3.8}$$

If $\lambda \lesssim m^{-2\alpha} + n^{-1}$, then

$$\left\|\hat{C}_\lambda - C_0\right\|_2^2 = O_p\left(m^{-2\alpha} + n^{-1}\right). \tag{3.9}$$

We now turn to the independent design. The following assumption is standard.

Assumption 2 (Independent Design)**.** The sampling locations T_{ij} are independent and identically distributed with a density bounded above and below on $\mathcal{T}$.

Theorem 2 (Upper Bound under Independent Design)**.** *Suppose that $X(\cdot)$ follows a distribution in $\mathcal{P}(\alpha; M_0, c_0)$ for $\alpha > 1$ and Assumption 2 holds.*

Assume $\mathbb{E}\varepsilon_{ij}^4 < \infty$. *Let* $\hat{\mu}$ *satisfy*

$$\|\hat{\mu} - \mu_0\|_2^2 = O_p\left(\left(\frac{\log n}{mn}\right)^{\frac{2\alpha}{2\alpha+1}} + \frac{1}{n}\right). \tag{3.10}$$

If $\lambda \asymp (\frac{\log n}{mn})^{\frac{2\alpha}{2\alpha+1}}$, *then*

$$\left\|\hat{C}_\lambda - C_0\right\|_2^2 = O_p\left(\left(\frac{\log n}{mn}\right)^{\frac{2\alpha}{2\alpha+1}} + \frac{1}{n}\right). \tag{3.11}$$

3.2 Minimax Lower Bounds

To assess the optimality of the proposed estimator, we show that the upper bounds derived in Section 3 are nearly rate-optimal by establishing the following minimax lower bounds. In fact, we will show that even if the covariance function C_0 is known a priori to be of rank one, i.e., X has a single principal component, the rates established in Theorem 1 and Theorem 2 cannot be much improved. To this end, let $\mathcal{P}'(\alpha; M_0)$ be the collection of probability measures of X such that $X(\cdot) = Z\psi(\cdot)$, where $Z \sim N(0,1)$ and $\|\psi\|^2_{\mathcal{W}_2^\alpha(\mathcal{T})} \le M_0$. Then $\mathcal{P}'(\alpha; M_0) \subset \mathcal{P}(\alpha; M_0, c_0)$ for every $c_0 \ge 3$.

Theorem 3 (Lower Bound under Common Design)**.** *There exists a constant* $c > 0$ *such that for any estimator* $\tilde{C}$ *based on common design data (Assumption* 1*):*

$$\liminf_{n\to\infty} \sup_{L(X)\in\mathcal{P}'(\alpha;M_0)} \mathbb{P}\left\{\|\tilde{C} - C_0\|_2^2 > c(m^{-2\alpha} + n^{-1})\right\} > 0. \tag{3.12}$$

Theorem 4 (Lower Bound under Independent Design). *There exists a constant $c > 0$ such that for any estimator $\tilde{C}$ based on independent design data (Assumption 2):*

$$\liminf_{n\to\infty} \sup_{L(X)\in\mathcal{P}'(\alpha;M_0)} \mathbb{P}\left\{\|\tilde{C} - C_0\|_2^2 > c((nm)^{-2\alpha/(2\alpha+1)} + n^{-1})\right\} > 0. \quad (3.13)$$

These theorems show that our tensor product smoothing spline estimator $\hat{C}_\lambda$ attains the fundamental limit under the common design and does so up to a logarithmic factor under the independent design. The phase transitions identified are therefore intrinsic properties of the statistical problem, not artifacts of the estimation method.

3.3 Discussion

Rates of Convergence. We begin by recalling the optimal rates for estimating the mean function μ_0. Cai and Yuan [2011] showed that a smoothing spline estimator attains the minimax-optimal rates under both common and independent designs. Specifically, under the independent design, the mean function can be estimated at the rate $O_p\left((mn)^{-\frac{2\alpha}{2\alpha+1}} + n^{-1}\right)$ in integrated squared error, whereas under the common design the optimal rate is $O_p\left(m^{-2\alpha} + n^{-1}\right)$.

A striking feature of our results is that, despite the covariance function being intrinsically bivariate, the convergence rates for covariance estimation differ from those for mean estimation by at most a logarithmic factor $(\log n)^{\frac{2\alpha}{2\alpha+1}}$. This is in sharp contrast with the conventional intuition that estimating a bivariate function should be substantially more difficult than estimating a univariate one.

To illustrate this point, consider the case where the sample paths of X belong to the Sobolev space $\mathcal{W}_2^2([0,1])$, corresponding to twice differentiable functions. A naive approach would treat the covariance function C_0 as an element of the second-order Sobolev space $\mathcal{W}_2^2([0,1]^2)$ on the unit square. As shown by Stone [1982], the minimax rates for estimating functions from $\mathcal{W}_2^2([0,1])$ and $\mathcal{W}_2^2([0,1]^2)$ are $M^{-4/5}$ and $M^{-2/3}$, respectively, when M independent observations are available. In covariance estimation, although one can form nm^2 pairwise observations $\{(Y_{ij_1} - \mu_0(T_{ij_1}))(Y_{ij_2} - \mu_0(T_{ij_2})) : 1 \leq i \leq n, 1 \leq j_1, j_2 \leq m\}$, these quantities are highly redundant, as they are constructed from only nm noisy observations and are strongly correlated due to the functional nature of the data.

To appreciate the importance of the association between C_0 and the tensor product space, we first consider the case when m is fixed. In this regime, the effective sample size is of order n for both mean and covariance estimation. While the mean function can be estimated at the optimal rate $n^{-4/5}$ [Cai and Yuan, 2011], existing approaches that rely solely on the twice differentiability of C_0, such as local polynomial smoothing [Hall et al., 2006], achieve at best the rate $n^{-2/3}$. In contrast, Proposition 1 shows that C_0 lies in the much smaller tensor product space $\mathcal{W}_2^2([0,1]) \otimes \mathcal{W}_2^2([0,1])$, rather than in the full bivariate Sobolev space. As a consequence, our method attains the substantially faster rate $(n/\log n)^{-4/5}$, which is unattainable by methods treating C_0 as a generic bivariate function.

Phase Transition. When the sampling frequency m is finite, the estimation problem becomes more delicate due to data redundancy and within-curve dependence. Theorem 1 and Theorem 2 reveal a curious phase transition behavior of the convergence rate. When the functions are densely sampled, the sampling frequency does not matter and the rate is determined solely by the number of curves:

$$\|\hat{C} - C_0\|_{L_2}^2 = O_p\left(n^{-1}\right), \tag{3.14}$$

which is the parametric rate necessary for estimating a constant function. This suggests that when sampled frequently enough, we can estimate the covariance function as well as if the entire functions are observable.

But when the functions are sparsely sampled, the sampling frequency plays a critical role in the estimation accuracy. It is interesting to note that for both designs, the boundary between the sparse and dense regimes occurs at $m \asymp n^{1/(2\alpha)}$, up to logarithmic factors. However, the two designs behave quite differently in the sparse regime. For the independent design, the rate is jointly determined by the number of curves and the number of observations on each curve through the total number of observations $N := nm$:

$$\|\hat{C} - C_0\|_{L_2}^2 = O_p\left((N/\log n)^{-\frac{2\alpha}{2\alpha+1}}\right), \tag{3.15}$$

reflecting efficient information pooling across subjects. Under the common design, by contrast, the rate is bottlenecked by discretization and depends only on m,

$$\|\hat{C} - C_0\|_{L_2}^2 = O_p\left(m^{-2\alpha}\right). \tag{3.16}$$

Thus, the independent design is strictly superior in the sparse regime, as random sampling locations mitigate the limitations imposed by a fixed grid. This phenomenon parallels the phase transition observed in mean function estimation [Cai and Yuan, 2011].

It is also interesting to compare our results with those obtained by Zhang and Wang [2016] and its subsequent developments [Lin and Wang, 2022, Petersen, 2024, Guo et al., 2025], which also study the phase transition phenomenon for covariance function estimation under the independent design. Assuming twice differentiability of the mean and covariance functions (corresponding to $\alpha = 2$), Zhang and Wang [2016] showed that the transition occurs at $m \asymp n^{1/4}$ under the independent design, which agrees with our results. However, their analysis does not address the common design case, nor does it establish minimax lower bounds, both of which are key components of the present work.

3.4 Choosing λ via Cross-Validation

The performance of the proposed estimator $\hat{C}_\lambda$ depends on the choice of the tuning parameter λ, which typically depends on the unknown smoothness parameter α. Because α also determines the boundary between sparse and dense sampling-frequency regimes, this tuning problem is precisely where adaptation to unknown smoothness and sampling-frequency structure enters. For the common fixed design, Theorem 1 suggests that $\lambda = O(m^{-2\alpha} + n^{-1})$ is appropriate, so it suffices to choose λ to be sufficiently small. In contrast, for the independent design, Theorem 2 indicates that λ should be chosen to be of the order $(\frac{\log n}{mn})^{\frac{2\alpha}{2\alpha+1}}$, while choosing λ too small or too large would lead to suboptimal performance. Thus, selecting an appropriate value of λ is crucial for the success of covariance function estimation under the independent design. While data-driven optimal choice of the tuning parameter is generally difficult, a common practice is to select λ via cross-validation. Below, we show that cross-validation can indeed yield a good choice of λ, leading to nearly the same rates of convergence as the optimal one given in Theorem 2.

Suppose we are additionally given $\tilde{n}$ validation samples $\{(\tilde{T}_{ij}, \tilde{Y}_{ij}) : 1 \leq i \leq \tilde{n}, 1 \leq j \leq m\}$ independently generated from the same model as $\{Y_{ij}\}$. Denote by Π_B the clipping operator into $[-B, B]$ for $B > 0$. We define the empirical loss on the validation data as

$$\mathcal{L}_{\text{val}}(\hat{C}) = \frac{1}{\tilde{n}} \sum_{i=1}^{\tilde{n}} \frac{1}{m(m-1)} \sum_{1 \leq j \neq k \leq m} \left[(\tilde{Y}_{ij} - \hat{\mu}(\tilde{T}_{ij}))(\tilde{Y}_{ik} - \hat{\mu}(\tilde{T}_{ik})) - \hat{C}(\tilde{T}_{ij}, \tilde{T}_{ik}) \right]^2, \tag{3.17}$$

where we are still given a good mean function estimator $\hat{\mu}$.

Let $\hat{C}_\lambda$, $\lambda \in \Lambda$, be the estimates obtained from (2.4) for a set Λ of candidate values of λ. We then choose $\hat{\lambda}$ by minimizing the validation loss:

$$\hat{\lambda}_{\text{cv}} = \arg\min_{\lambda \in \Lambda} \mathcal{L}_{\text{val}}(\Pi_B \hat{C}_\lambda). \tag{3.18}$$

The following theorem shows that the resulting estimator $\hat{C}_{\hat{\lambda}_{\text{cv}}}$ is almost as good as the oracle estimator that knows the best λ in advance.

Theorem 5. *Suppose that the conditions of Theorem 2 hold and Y_{ij} is sub-Gaussian. Suppose that the validation sample size $\tilde{n}$ satisfies $\tilde{n} \asymp n$. Take B_0 to be a large constant and*

$$\Lambda = \left\{ 2^{-k} : k = 0, 1, \ldots, \lceil \log_2(nm) \rceil \right\}.$$

Then, we have

$$\left\|\Pi_{B_0}\hat{C}_{\hat{\lambda}_{\mathrm{cv}}} - C_0\right\|_2^2 = O_p\left(\left(\frac{\log n}{mn}\right)^{\frac{2\alpha}{2\alpha+1}} + \frac{(\log n)^3}{n}\right). \tag{3.19}$$

Remark 2. The requirement that Y_{ij} is sub-Gaussian is mild and is satisfied when X is a Gaussian process and ε is sub-Gaussian. While our result is stated for covariance function estimation, similar results can be shown for mean function estimation and thus for the whole estimation procedure.

3.5 Estimating Principal Components

Although not our main focus, we note that convergence rates of the estimated functional principal components can be easily derived from Theorem 1 and Theorem 2. To this end, we consider the theoretical properties of $\hat{\psi}_k$ for $k \le K_0$ where K_0 is fixed. Without loss of generality, we assume that $\langle \psi_k, \hat{\psi}_k \rangle \ge 0$. Then we have the following result.

Corollary 1. *Under the conditions of Theorem* 1 *or Theorem* 2*, if* θ_k *is of multiplicity one, then the following rates hold under the common and independent designs, respectively,*

$$\left\|\hat{\psi}_k - \psi_k\right\|_{L_2}^2 = O_p\left(m^{-2\alpha} + \frac{1}{n}\right), \quad \textit{or} \quad \left\|\hat{\psi}_k - \psi_k\right\|_{L_2}^2 = O_p\left(\left(\frac{\log n}{mn}\right)^{\frac{2\alpha}{2\alpha+1}} + \frac{1}{n}\right). \tag{3.20}$$

The corollary shows that the covariance estimator can be directly plugged into FPCA, where the eigenfunctions can be computed explicitly from the finite-dimensional kernel representation described in the supplementary material, while still achieving good statistical properties. Despite this simple construction, the resulting rates are optimal up to logarithmic factors for any fixed component with a simple eigenvalue, matching the optimal rates established in Hall et al. [2006] and Zhou et al. [2025]. However, we refer to Zhou et al. [2025] for a systematic treatment and optimal rates for estimating eigenfunctions and eigenvalues, which also allows for diverging eigenfunction indices but requires additional spectral decay assumptions. Nevertheless, our plug-in approach is simple and practical, and achieves nearly optimal rates for fixed components without additional assumptions.

4. Numerical Experiments

This section discusses the implementation and numerical performance of the proposed covariance function estimator. Additional numerical results can be found in the supplementary material.

4.1 Computation

We now discuss the implementation of the proposed estimator and the computation of the associated functional principal components. It is well known that $\mathcal{W}_2^\alpha(\mathcal{T})$ is a reproducing kernel Hilbert space (RKHS), and we denote by K the reproducing kernel of $\mathcal{W}_2^\alpha(\mathcal{T})$. Then, the tensor product space $\mathcal{W}_2^\alpha(\mathcal{T}) \otimes \mathcal{W}_2^\alpha(\mathcal{T})$ is also an RKHS with the reproducing kernel given by $K \otimes K$ defined as $(K \otimes K)((s_1, t_1), (s_2, t_2)) = K(s_1, s_2)K(t_1, t_2)$. The following lemma shows that, although the optimization problem (2.4) is formulated over an infinite-dimensional function space, it admits a finite-dimensional representation in terms of the reproducing kernel K evaluated at the sampling locations.

Lemma 1. *There exist $m \times m$ symmetric matrices $A_1, \ldots, A_n$ whose diagonal entries are zeros such that*

$$\hat{C}_\lambda(s,t) = \sum_{i=1}^{n} \left\{ [K(s, T_{ij})]_{1 \le j \le m} A_i [K(t, T_{ij})]'_{1 \le j \le m} \right\}. \tag{4.21}$$

Lemma 1 can be proved by an argument similar to that of Theorem 1.3.1 in Wahba [1990], and we omit the proof here for brevity. Write

$$G_{i_1 i_2} = [K(T_{i_1 j_1}, T_{i_2 j_2})]_{1 \le j_1, j_2 \le m}, \qquad 1 \le i_1, i_2 \le n. \tag{4.22}$$

Observe that for any function $\hat{C}_\lambda$ that admits representation (4.21), its squared RKHS norm $\|\hat{C}_\lambda\|^2_{\mathcal{W}_2^\alpha(\mathcal{T}) \otimes \mathcal{W}_2^\alpha(\mathcal{T})}$ can be written as

$$\|\hat{C}_\lambda\|^2_{\mathcal{W}_2^\alpha(\mathcal{T}) \otimes \mathcal{W}_2^\alpha(\mathcal{T})} = \sum_{i_1, i_2 = 1}^{n} \mathrm{Tr}(G_{i_1 i_2} A_{i_2} G_{i_2 i_1} A_{i_1}), \tag{4.23}$$

which is a convex quadratic function of the entries of the A_i. Because $\mathcal{L}_{n,m}$ is also convex and quadratic, computing $\hat{C}_\lambda$, or equivalently solving for the

A_i, becomes a convex quadratic programming problem and can be solved fairly easily. Readers are also referred to Wahba [1990] and Gu [2002] for further discussions on algorithms and strategies to solve this type of problem efficiently. Alternatively, a closed-form solution can also be derived.

We briefly discuss the computational complexity of the proposed estimator. For independent design, let $d = m(m-1)/2$ and $D = nd$. The direct implementation forms a $D \times D$ kernel system, requiring $O(D^2) = O(n^2m^4)$ memory and $O(D^3) = O(n^3m^6)$ time for a fixed λ. For L candidate values, the kernel matrix is reused, so fitting costs $O(D^2 + LD^3)$. Under the common design, all curves share the same m sampling locations. After averaging the pairwise products in $O(nm^2)$ operations, only one $d \times d$ system remains. The direct cost is $O(nm^2 + m^6)$ time and $O(m^4)$ memory for a fixed λ, or $O(nm^2 + m^4 + Lm^6)$ time for L values.

The same kernel representation yields the functional principal components without discretizing $\hat{C}_\lambda$; details are provided in the supplementary material. Under the independent design, FPCA reduces to an $nm \times nm$ symmetric eigenproblem, requiring $O(n^2m^2)$ memory and $O(n^3m^3)$ time for a full eigendecomposition. Under the common design, it reduces to an $m \times m$ eigenproblem, with $O(m^2)$ memory and $O(m^3)$ time.

4.2 Simulation Studies

We conduct simulation studies to assess the empirical performance of the proposed method and to verify the theoretical convergence behavior under different sampling designs. In particular, we examine how the estimation accuracy depends on the number of curves n, the sampling frequency m, and whether the sampling locations follow a common or independent design.

In the simulations, random functions X are generated according to

$$X(t) = \sum_{k=1}^{50} \zeta_k Z_k \cos(k\pi t), \qquad t \in [0,1], \tag{4.24}$$

where Z_k are i.i.d. random variables uniformly distributed on $[-\sqrt{3}, \sqrt{3}]$ and $\zeta_k = (-1)^{k+1} k^{-\alpha}$. It is easy to see that X belongs to the Sobolev space $\mathcal{W}_2^{\alpha}([0,1])$ almost surely. In particular, we set $\alpha = 2$. The resulting covariance function is

$$C_0(s,t) = \sum_{k=1}^{50} k^{-2\alpha} \cos(k\pi s)\cos(k\pi t).$$

For the independent design, the sampling locations T_{ij} are drawn independently from the uniform distribution on $[0,1]$. For the common design, the sampling locations are taken as m equally spaced grid points. Fifty random functions and observations were then generated according to the model (1.2), where the measurement errors ε_{ij} were independently sampled from $N(0, \sigma_0^2)$ with $\sigma_0 = 0.368$.

For fitting the simulated data, we use the Sobolev space $W_2^2([0,1])$ as the univariate RKHS and its tensor product for the covariance function. Its reproducing kernel is the cubic smoothing spline kernel

$$K(s,t) = 1 + st + \frac{1}{6}\min(s,t)^2\left(3\max(s,t) - \min(s,t)\right).$$

We apply the proposed method to the simulated datasets to obtain covariance function estimates as well as estimates of the functional principal components. The smoothing parameter λ was selected by five-fold cross-validation.

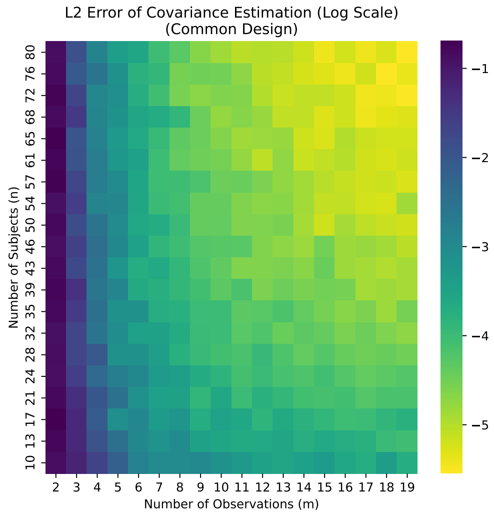
Representative covariance and eigenfunction estimates under the two designs are provided in the supplementary material. They show that the proposed estimator captures the main covariance structure even when $m = 5$, with accuracy improving as m increases.

To further illustrate how the estimation accuracy improves with increasing n and m, and how the sampling design affects the estimation accuracy, we repeated each design and (n, m) combination 50 times and computed the integrated squared error (ISE) of the covariance function estimate. Figure 1 reports the average log-ISE and its Monte Carlo standard error. The average log-ISE heatmaps reveal markedly different patterns under the two designs. Under the common design, estimation accuracy is primarily driven by the sampling frequency m when m is small, reflecting discretization effects. In contrast, under the independent design, both n and m contribute to accuracy through information pooling across curves, leading to a more uniform reduction in ISE. These empirical findings are consistent with the phase transition phenomena predicted by the theory.

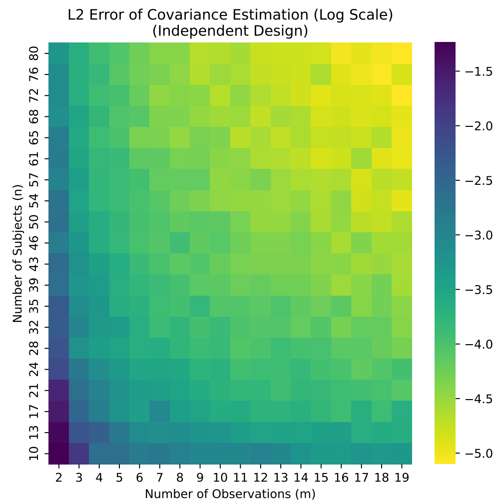
We also compare our regularization estimator (Reg) with the PACE covariance estimator of Yao et al. [2005]. For each design, we use $n = 50$ and $m = 3, 5, 10$. The smoothing parameter for the proposed estimator is selected by five-fold cross-validation, while PACE selects its covariance bandwidth by generalized cross-validation. Table 1 on page 31 reports the average ISE and its Monte Carlo standard error. Under the common design, the proposed estimator has lower average ISE at all three sampling frequencies. Under the independent design, the performance of the two methods is roughly comparable. Nevertheless, we emphasize that the main focus of this paper is the theoretical understanding of the phase transition phenomenon in covariance function estimation in terms of minimax rates up to constants.

4.3 Bike Sharing Dataset

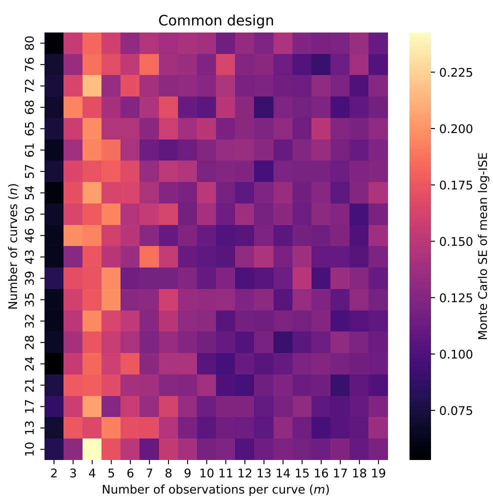
To further illustrate the usefulness of the proposed covariance function estimator, we apply the proposed method to the Bike Sharing Dataset from the



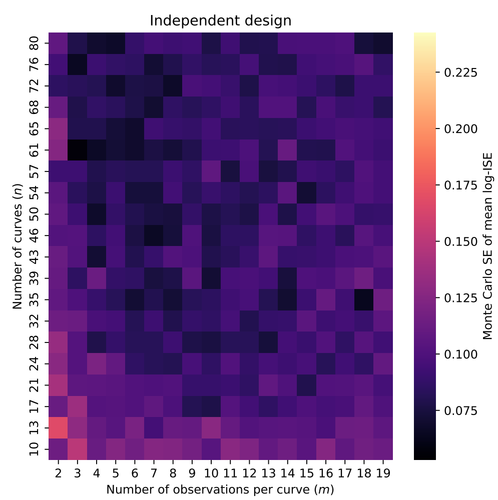
(a) Common design.



(b) Independent design.



(c) Common design.



(d) Independent design.

Figure 1: Simulation results across varying n and m, based on 50 independent replications per design and grid point. The upper panels show average log-ISE; the lower panels show the Monte Carlo standard errors of these averages.

Table 1: Comparison with PACE at $n = 50$. Entries are average ISE with standard errors in parentheses. Lower values are better.

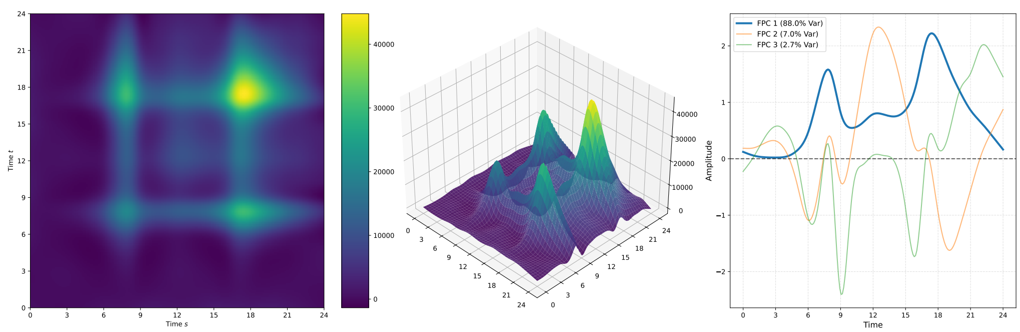
Design	Method	$m = 3$	$m = 5$	$m = 10$
Common	Reg	0.0781 (0.0037)	0.0106 (0.0013)	0.0113 (0.0016)
	PACE	0.1582 (0.0030)	0.0226 (0.0016)	0.0132 (0.0015)
Independent	Reg	0.0401 (0.0043)	0.0309 (0.0030)	0.0164 (0.0016)
	PACE	0.0372 (0.0029)	0.0284 (0.0032)	0.0150 (0.0015)

UCI Machine Learning Repository [Fanaee-T, 2013], which records hourly bike rental counts in Washington, D.C., from January 1, 2011 to December 31, 2012. We use the total rental count and treat each day as a curve observed at $m = 24$ equally spaced hourly time points. This yields 731 daily curves under a common sampling design: 521 weekdays and 210 weekend days. The hourly data matrix has 0.94% missing entries; fitting and cross-validation omit the affected observations and off-diagonal products. The mean and covariance smoothing parameters are selected separately by five-fold cross-validation over a logarithmic grid from 10^{-10} to 10^{-1}, using a fixed split with seed 1000. The selected (λ_μ, λ_C) pairs are $(10^{-10}, 10^{-10})$ for the full data and weekdays, and $(1.93 \times 10^{-9}, 10^{-10})$ for weekends. The small covariance smoothing parameters accord with the common-design guidance in Theorem 1, under which $\lambda_C \lesssim m^{-2\alpha} + n^{-1}$ suffices.

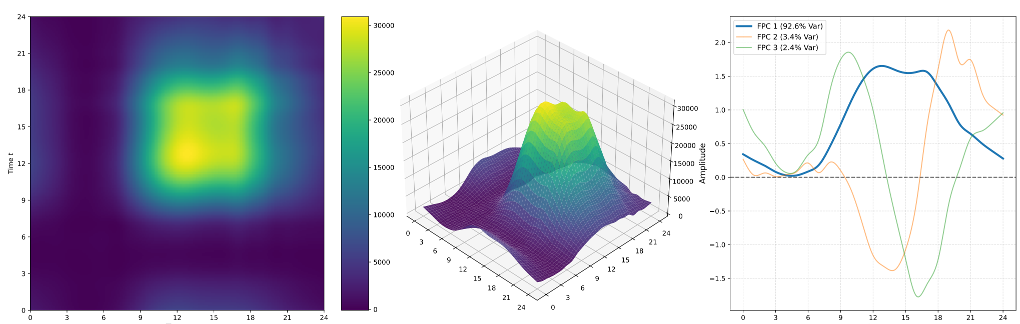
Estimating the covariance function for the dataset is particularly meaningful as it enables a deeper understanding of temporal dependencies and variability in bike rental patterns, revealing intraday structures such as peak commuting hours (e.g., mornings and evenings) and off-peak lulls. This estimation facilitates downstream analyses, including FPCA to identify dominant modes of variation—potentially reducing dimensionality for predictive modeling of future rentals or anomaly detection in urban mobility.

We performed separate covariance function estimation for weekday and weekend subsets, recognizing the potential heterogeneity in rental patterns driven by user behavior. FPCA was subsequently applied to the estimated covariance functions to decompose the variability in daily rental curves over the 24-hour domain. The results are presented in Figure 2 on page 33.

It is clear from the estimated covariance functions and eigenfunctions that distinct temporal structures exist between weekdays and weekends. For weekdays, the covariance function exhibits bimodal peaks corresponding to morning (around 7–9 AM) and evening (around 5–7 PM) rush hours, reflecting commuting demands. In contrast, weekend patterns show a unimodal structure with a broader peak in the afternoon (12–6 PM), indicative of recreational usage. This distinction can also be observed in the first eigenfunction in FPCA for each case, where around 90% of the variance is explained. These differences highlight differing user behaviors and temporal dependencies in bike rentals across the week, justifying stratified estimation to avoid biasing the overall covariance toward dominant weekday signals.



(a) Weekdays



(b) Weekends

Figure 2: Estimated covariance functions and the first three eigenfunctions for bike rental counts on weekdays (top) and weekends (bottom).

5. Conclusion and Discussion

We have developed a regularized covariance function estimator for discretely observed functional data with measurement error. The method exploits the tensor product structure induced by the underlying process, rather than treating the covariance as a generic bivariate function. The resulting upper and lower bounds characterize the distinct phase transitions under common and independent designs, and the same estimator yields nearly optimal estimates of fixed functional principal components.

The assumption of equal observation number under the independent design is mainly expositional. The arguments extend when the curve-specific sample sizes m_i are uniformly comparable to a common scale m, and to random m_i with mean m and uniformly bounded variance. The construction also extends to functional data on other compact domains, with rates adjusted to the effective dimension.

Although the explicit analysis is developed for Sobolev RKHSs, the tensor-product principle is more general. If the sample paths belong almost surely to an RKHS $\mathcal{H}(K)$ and have finite second moment in its RKHS norm, then the covariance function belongs to $\mathcal{H}(K) \otimes \mathcal{H}(K)$, leading to an analogous penalized estimator. Under the independent design, similar rates follow under suitable spectral decay conditions on K. The analysis under common design additionally requires approximation and stability conditions that are not automatic for a general kernel; establishing verifiable conditions for broad kernel classes remains an open problem.

Two further questions are especially relevant. The direct kernel implementation can be costly for large n or m, motivating scalable algorithms that preserve the statistical guarantees. Under the independent design, it also remains open whether the logarithmic gap between the upper and lower bounds is intrinsic or an artifact of the analysis.

Supplementary Material

Detailed proofs, computational details, and additional numerical results are provided in the supplementary material.

References

Denis Bosq. *Linear Processes in Function Spaces: Theory and Applications*, volume 149. Springer Science & Business Media, 2000.

T. Tony Cai and Ming Yuan. Optimal estimation of the mean function based on discretely sampled functional data: Phase transition. *The Annals of Statistics*, 39(5):2330–2355, October 2011. ISSN 0090-5364. doi: 10.1214/11-AOS898.

Aurore Delaigle, Peter Hall, Wei Huang, and Alois Kneip. Estimating the Covariance of Fragmented and Other Related Types of Functional Data. *Journal of the American Statistical Association*, 116(535):1383–1401, July 2021. ISSN 0162-1459. doi: 10.1080/01621459.2020.1723597.

M-H Descary and V M Panaretos. Recovering covariance from functional fragments. *Biometrika*, 106(1):145–160, March 2019. ISSN 0006-3444. doi: 10.1093/biomet/asy055.

Peter J. Diggle, Patrick J. Heagerty, Kung-Yee Liang, and Scott L. Zeger. *Analysis of Longitudinal Data*, volume 25 of *Oxford Statistical Science Series*. Oxford University Press, Oxford, 2 edition, 2002. ISBN 978-0-19-852484-7. doi: 10.1093/oso/9780198524847.001.0001.

Hadi Fanaee-T. Bike Sharing. UCI Machine Learning Repository, 2013. URL `https://archive.ics.uci.edu/dataset/275/bike+sharing+dataset`.

Frederic Ferraty, Wenceslao Gonzalez-Manteiga, Adela Martinez-Calvo, and Philippe Vieu. Presmoothing in functional linear regression. *Statistica Sinica*, 22(1):69–94, January 2012. ISSN 1017-0405. doi: 10.5705/ss.2010.085.

Chong Gu. *Smoothing Spline ANOVA Models*. Springer Series in Statistics. Springer New York, New York, NY, 2002. ISBN 978-1-4419-2966-2 978-1-4757-3683-0. doi: 10.1007/978-1-4757-3683-0.

Shaojun Guo, Dong Li, Xinghao Qiao, and Yizhu Wang. From Sparse to Dense Functional Data in High Dimensions: Revisiting Phase Transitions from a Non-Asymptotic Perspective. *Journal of Machine Learning Research*, 26(15):1–40, 2025. ISSN 1533-7928.

Peter Hall, Hans-Georg Müller, and Jane-Ling Wang. Properties of principal component methods for functional and longitudinal data analysis. *The Annals of Statistics*, 34(3):1493–1517, 2006. doi: 10.1214/009053606000000272.

Tailen Hsing and Randall Eubank. *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley Series in Probability and Statistics. Wiley, May 2015. ISBN 978-0-470-01691-6 978-1-118-76254-7. doi: 10.1002/9781118762547.

Gareth M. James and Trevor J. Hastie. Functional linear discriminant analysis for irregularly sampled curves. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 63(3):533–550, 2001.

Gareth M. James, Trevor J. Hastie, and Catherine A. Sugar. Principal component models for sparse functional data. *Biometrika*, 87(3):587–602, 2000.

Salil Koner and Ana-Maria Staicu. Second-Generation Functional Data. *Annual Review of Statistics and Its Application*, 10(1):547–572, March 2023. ISSN 2326-8298, 2326-831X. doi: 10.1146/annurev-statistics-032921-033726.

Yehua Li and Tailen Hsing. Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics*, 38(6):3321–3351, December 2010. ISSN 0090-5364, 2168-8966. doi: 10.1214/10-AOS813.

Zhenhua Lin and Jane-Ling Wang. Mean and Covariance Estimation for Functional Snippets. *Journal of the American Statistical Association*, 117(537):348–360, January 2022. ISSN 0162-1459. doi: 10.1080/01621459.2020.1777138.

Debashis Paul and Jie Peng. Consistency of restricted maximum likelihood estimators of principal components. *The Annals of Statistics*, 37(3):1229–1271, 2009. doi: 10.1214/08-AOS608.

Alexander Petersen. Mean and covariance estimation for discretely observed high-dimensional functional data: Rates of convergence and division of observational regimes. *Journal of Multivariate Analysis*, 204:105355, November 2024. ISSN 0047-259X. doi: 10.1016/j.jmva.2024.105355.

Jim Ramsay and B. W. Silverman. *Functional Data Analysis (Springer Series in Statistics)*. Springer, 2nd edition, 2005. ISBN 978-0-387-40080-8.

John A. Rice and Colin O. Wu. Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics*, 57(1):253–259, 2001.

Minggao Shi, Robert E. Weiss, and Jeremy M. G. Taylor. An Analysis of Paediatric CD4 Counts for Acquired Immune Deficiency Syndrome Using Flexible Random Curves. *Applied Statistics*, 45(2):151–163, 1996. ISSN 00359254. doi: 10.2307/2986151.

Joan G. Staniswalis and J. Jack Lee. Nonparametric Regression Analysis of Longitudinal Data. *Journal of the American Statistical Association*, 93(444):1403–1418, December 1998. ISSN 0162-1459, 1537-274X. doi: 10.1080/01621459.1998.10473801.

Michael L. Stein. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Science & Business Media, 1999.

Charles J. Stone. Optimal Global Rates of Convergence for Nonparametric Regression. *The Annals of Statistics*, 10(4):1040–1053, 1982. ISSN 00905364. doi: 10.1214/aos/1176345969.

Jianbin Tan, Pixu Shi, and Anru R. Zhang. Functional-SVD for Heterogeneous Trajectories: Case Studies in Health. *Journal of the American Statistical Association*, pages 1–16, June 2026. ISSN 0162-1459, 1537-274X. doi: 10.1080/01621459.2026.2625441.

Grace Wahba. *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics, January 1990. ISBN 978-0-89871-244-5 978-1-61197-012-8. doi: 10.1137/1.9781611970128.

Luo Xiao. Asymptotic properties of penalized splines for functional data. *Bernoulli*, 26(4): 2847–2875, November 2020. ISSN 1350-7265. doi: 10.3150/20-BEJ1209.

Luo Xiao, Cai Li, William Checkley, and Ciprian Crainiceanu. Fast covariance estimation for sparse functional data. *Statistics and Computing*, 28(3):511–522, May 2018. ISSN 1573-1375. doi: 10.1007/s11222-017-9744-8.

Fang Yao, Hans-Georg Müller, and Jane-Ling Wang. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590, June 2005. ISSN 0162-1459, 1537-274X. doi: 10.1198/016214504000001745.

Xiaoke Zhang and Jane-Ling Wang. From sparse to dense functional data and beyond. *The Annals of Statistics*, 44(5):2281–2321, October 2016. ISSN 0090-5364, 2168-8966. doi: 10.1214/16-AOS1446.

Hang Zhou, Dongyi Wei, and Fang Yao. Theory of functional principal component analysis for discretely observed data. *The Annals of Statistics*, 53(5):2103–2127, October 2025. ISSN 0090-5364. doi: 10.1214/25-AOS2531.

Yang Zhou, Jin Yang, and Fang Yao. Covariance Test for Discretely Observed Functional Data: When and How it Works? *Journal of the American Statistical Association*, pages 1–12, June 2026. ISSN 0162-1459, 1537-274X. doi: 10.1080/01621459.2026.2641240.

Department of Statistics and Data Science, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104.

E-mail: tcai@wharton.upenn.edu

KLATASDS-MOE, School of Statistics, East China Normal University, Shanghai 200062, China.

E-mail: ycli@sfs.ecnu.edu.cn

Department of Statistics, Columbia University.

E-mail: ming.yuan@columbia.edu