# THE COST OF DISCRETIZATION IN FUNCTIONAL LINEAR REGRESSION: MINIMAX RATES AND ADAPTATION

BY T. TONY CAI[1,a] AND YICHENG LI[2,b]

[1]Department of Statistics and Data Science, The Wharton School, University of Pennsylvania
[a]tcai@wharton.upenn.edu

[2]KLATASDS-MOE, School of Statistics, East China Normal University, Shanghai, China
[b]li_yicheng@foxmail.com

We study scalar-on-function linear regression when each covariate curve is observed only through finitely many noisy point evaluations. Our goal is to characterize the minimax estimation and prediction risks as joint functions of the number of trajectories $n$ and the within-trajectory resolution $m$. Working in a fixed trigonometric eigenbasis, with covariance eigenvalues decaying at rate $\alpha$ and slope function of Sobolev smoothness $s$, we derive matching minimax upper and lower bounds under two canonical sampling schemes.

Under an independent random design, the minimax prediction rate is

$$n^{-\frac{2\alpha+2s}{2\alpha+2s+1}} + (nm)^{-\frac{2\alpha+2s}{4\alpha+2s+1}}.$$

The first term is the fully observed functional linear regression benchmark, while the second term captures the cost of noisy point evaluations after amplification by the inverse covariance operator. Under a common design on an equally spaced grid, the shared sampling geometry introduces additional obstructions, and the minimax prediction rate becomes

$$n^{-\frac{2\alpha+2s}{2\alpha+2s+1}} + (nm)^{-\frac{2\alpha+2s}{4\alpha+2s+1}} + m^{-(2\alpha+2s)} + m^{-4\alpha}.$$

Here the third term represents fixed-grid discretization error, whereas the fourth reflects the cost of identifying unknown eigenvalues from observations on a common grid.

We further construct data-driven adaptive estimators, based on covariance-scale screening and blockwise prediction-energy thresholding, that attain these rates without prior knowledge of the eigenvalue sequence or the smoothness indices. The results reveal a sharp resolution-dependent phase transition under independent design and a richer phase diagram under common design. Numerical simulations and a real-data example illustrate the theoretical findings.

**1. Introduction.** Functional linear regression (FLR) is a central model for scalar-on-function regression in functional data analysis (Ramsay and Silverman, 2005; Wang, Chiou and Müller, 2016; Hsing and Eubank, 2015). It provides a basic framework for studying how an entire trajectory, curve, image profile, or time-varying signal predicts a scalar outcome. When the functional covariate is fully observed, estimation and prediction are supported by a mature minimax theory. For example, Yuan and Cai (2010) provided an RKHS framework with smoothness regularization that yields minimax optimal rates for both slope estimation and prediction. Other related work covers functional PCA and regularization for slope estimation (Hall and Horowitz, 2007), adaptive estimation (Cai and Yuan, 2012; Cai, Zhang and Zhou, 2018), and recent general regularization methods (Fan, Guo and Shi, 2024; Gupta, Sivananthan and Sriperumbudur, 2025).

*MSC2020 subject classifications:* Primary 62R10; secondary 62G08, 62G20, 62C20.
*Keywords and phrases:* Adaptive estimation, Common design, Discrete noisy observations, Functional linear regression, Independent design, Minimax rate.

In realistic applications, however, the predictor trajectory is rarely observed as an ideal function on a continuum. Instead, each subject is typically measured only at finitely many locations, and each measurement may be contaminated by noise (Cai and Yuan, 2011; Zhang and Wang, 2016; Zhou, Yao and Zhang, 2023). This regime arises in longitudinal studies, wearable sensing, imaging, environmental monitoring, and other acquisition systems where functional predictors are available only through discrete noisy measurements. The distinction is not a technical nuisance. Discrete noisy observation changes the statistical experiment itself: the sampling frequency, measurement noise, and sampling geometry all affect how much information about the latent trajectory and the slope function is available. Consequently, procedures that are optimal for fully observed FLR need not remain optimal, or even rate-optimal, when the covariates are observed only through noisy point evaluations. A fundamental question is therefore how the minimax risks for slope estimation and prediction are affected by discretization, and how the within-trajectory resolution $m$ interacts with the number of trajectories $n$ to determine the effective information in the data.

This paper studies this "cost of discretization" for functional linear regression. Concretely, the data consist of scalar responses and noisy point evaluations

$$Y_i = \langle X_i, \beta \rangle_{L^2} + \varepsilon_i, \qquad Z_{ij} = X_i(t_{ij}) + \delta_{ij}, \quad j = 1, \ldots, m, \quad i = 1, \ldots, n.$$

Here $X_i$ are independent latent trajectories, $\beta$ is the unknown slope function, and $t_{ij}$ are the sampling locations. For estimation, performance is measured by the squared $L^2$ risk

$$\mathbb{E}\|\hat{\beta} - \beta\|_{L^2}^2$$

of the slope function itself. For prediction, performance is measured by the excess prediction risk

$$\mathbb{E}_\star[\langle X_\star, \hat{\beta} - \beta \rangle_{L^2}^2],$$

where $X_\star$ is an independent test trajectory drawn from the same distribution as the $X_i$'s. The estimation and prediction problems are closely related but not identical: estimation measures recovery of the slope function in $L^2$, whereas prediction weights errors through the covariance structure of the future trajectory. In this paper, the two analyses are parallel, so we focus on prediction risk in the main theorems for clarity, while the corresponding rates for estimation error are also discussed. The full setting is given in Section 2.

The sampling scheme itself is a central part of the statistical problem. We focus on two standard schemes: *independent design* and *common design* (Cai and Yuan, 2010). Under independent design, each subject is observed at its own random sampling locations; under common design, all trajectories share the same deterministic grid. These two designs represent common data-acquisition mechanisms and lead to fundamentally different forms of information loss. In independent design, random sampling locations can be pooled across subjects, so increasing $n$ also improves coverage of the domain. In common design, all subjects are observed through the same grid, so increasing $n$ reduces stochastic variation but cannot remove geometric aliasing or fixed-grid approximation errors. Thus the central question is not only how the minimax rates depend on $(n, m)$, but also how the answer changes with the sampling geometry. Identifying this dependence is essential for understanding when discretely observed FLR behaves like the fully observed problem, when the total number of scalar measurements $nm$ is the limiting resource, and when the grid itself creates an irreducible obstruction.

A related line of work studies the cost of discretization for functional mean and covariance estimation, where the sampling frequency already changes the statistical problem before the inverse structure of FLR enters. Sparse methods for functional data developed by Yao, Müller and Wang (2005a) and Hall, Müller and Wang (2006) provide foundational tools for

TABLE 1

*Rate terms appearing in the minimax prediction risk under the main observation settings. Define $\nu := 2\alpha + 2s$ and $\kappa := 4\alpha + 2s + 1$. The known $\lambda$ common design row treats the eigenvalue sequence as fixed and available, whereas the last row treats it as unknown.*

| Setting | Dense (I) | | Sampling (II) | | Grid (III) | | Eigenvalue identification (IV) |
|---|---|---|---|---|---|---|---|
| Fully observed | $n^{-\nu/(\nu+1)}$ | | | | | | |
| Independent design | $n^{-\nu/(\nu+1)}$ | $+$ | $(nm)^{-\nu/\kappa}$ | | | | |
| Common design, known $\lambda$ | $n^{-\nu/(\nu+1)}$ | $+$ | $(nm)^{-\nu/\kappa}$ | $+$ | $m^{-\nu}$ | | |
| Common design, unknown $\lambda$ | $n^{-\nu/(\nu+1)}$ | $+$ | $(nm)^{-\nu/\kappa}$ | $+$ | $m^{-\nu}$ | $+$ | $m^{-4\alpha}$ |

recovering mean and covariance structure from partial noisy trajectories. For mean estimation, Cai and Yuan (2011) made this cost explicit by deriving distinct minimax rates and phase transitions in $m$ under common and independent designs respectively. Analogous rates have been obtained for covariance estimation (Cai and Yuan, 2010) and high-dimensional functional data (Petersen, 2024). The sparse to dense viewpoint was further systematized by Zhang and Wang (2016) and has also been studied through predictive distributions (Gajardo, Dai and Müller, 2021). These works often identify a transition between two regimes: when $m$ is large, the fully observed rate is recovered, whereas when $m$ is small, the rate is dominated by a sampling term that depends on $m$.

The FLR problem considered here is nevertheless different in an essential way with an additional inverse-regression structure. One must recover not only the sample path $X$, but also the slope function $\beta$ and the interaction between the two. As a result, discretization terms can arise in FLR that have no analogue in mean or covariance estimation. For discretely observed scalar-on-function regression, the most relevant prior work is Zhou, Yao and Zhang (2023), which focuses on independent random design and gives sufficient conditions on $m$ under which the fully observed dense rate is recovered. However, it does not provide a minimax lower bound that depends on $m$, so it is unclear whether its threshold for the dense regime is sharp, and it also does not address sampling on a common grid. What left open by these works is a complete minimax theory for the FLR targets that treats $n$ and $m$ as joint statistical resources, separates the term due to noisy point evaluations from the fully observed benchmark, and distinguishes these statistical terms from losses created by a common grid.

We address this problem by establishing minimax rates with respect to both $n$ and $m$ for the two sampling schemes. To isolate the cost of discretization, we work in a fixed trigonometric basis and assume that the covariance operator of $X$ is diagonalized by this basis, which includes periodic stationary covariance kernels. This formulation sets aside the additional effects of covariance eigenbasis estimation and of the alignment between the covariance geometry and the regularity scale of $\beta$ (Cai and Yuan, 2012), which are present even in the fully observed FLR model and are not the focus of this paper. The resulting rates show that discretely observed FLR has its own phase transition structure, not merely the sparse to dense transition known from mean and covariance estimation. Under independent design, the cost of discretization appears through a sampling term with a distinct exponent from the dense FLR term due to amplification by covariance inversion. Under common design, the rate is richer: the same statistical branch remains, but the shared grid creates additional geometric obstructions through fixed grid discretization and the identification of unknown eigenvalues.

To illustrate our results in detail, let $\alpha$ be the decay rate of the covariance eigenvalues of $X$ and $s$ be the smoothness of the slope function $\beta$ (see Section 2 for precise definitions). In the case of independent design, the random sampling locations can be pooled across subjects, so the discretization cost appears as a single sampling term caused by noisy point evaluations. In the range $\alpha > 1/2$ and $s \geq 0$ of minimum regularity requirement, the minimax prediction risk is of order

$$n^{-\frac{2\alpha+2s}{2\alpha+2s+1}} + (nm)^{-\frac{2\alpha+2s}{4\alpha+2s+1}}.$$

The first term is the minimax rate for fully observed or dense FLR (Cai and Yuan, 2012; Zhou, Yao and Zhang, 2023). The exponent $2\alpha + 2s$ reflects the effective smoothness of the problem as the sum of the slope smoothness and the covariance decay. The second term does not appear in the fully observed theory and represents the sampling price paid for noisy point evaluations. Its exponent has a larger denominator because measurement noise is amplified by the inverse covariance eigenvalues before it enters prediction. The phase transition between the two terms occurs at $m \asymp n^{2\alpha/(2\alpha+2s+1)}$. Below this threshold, the effective information is governed by the total number of scalar measurements $nm$; above it, further refinement within each curve no longer improves the rate, and the effective information is governed by the number of curves $n$ alone.

Common design is different because all trajectories are viewed through the same grid. The two statistical terms above are still present, but a common grid leaves a geometric error that cannot be averaged away by increasing the number of curves. When the eigenvalue sequence $(\lambda_k)_{k\in\mathbb{Z}}$ is known, the oracle rate for an equal grid consists of the two statistical terms and an additional term $m^{-(2\alpha+2s)}$, representing the reconstruction error left by the fixed grid. When $(\lambda_k)_{k\in\mathbb{Z}}$ is unknown, estimation must also contend with an obstruction from eigenvalue identification, and the sharp rate has the four term decomposition

$$n^{-\frac{2\alpha+2s}{2\alpha+2s+1}} + (nm)^{-\frac{2\alpha+2s}{4\alpha+2s+1}} + m^{-(2\alpha+2s)} + m^{-4\alpha}.$$

The additional term $m^{-4\alpha}$ depends solely on the covariance decay and sampling frequency, and represents the cost of identifying the eigenvalues from the common grid. Our lower bound shows that this term is unavoidable when the eigenvalues are unknown and is not merely a byproduct of the adaptive construction. This contrasts with independent design, where estimating the eigenvalue sequence does not change the minimax rate. In this case, the phase diagram is more complex, with multiple transitions between the four terms as $m$ increases. See Figure 1 for an illustration.

Our analysis also requires several new technical ingredients beyond the fully observed FLR theory, as discretization cannot be analyzed as a separate error. On the lower bound side, we establish a delicate Fisher information control that accounts for noisy point evaluations. In addition, under common design, we develop two distinct indistinguishability constructions to capture the effects of fixed grid discretization and unknown eigenvalue identification. On the upper bound side, the interplay between the slope function, the covariance function, and sampling geometry creates a nontrivial tradeoff between bias and variance that cannot be resolved by a single truncation or thresholding step for adaptation. To this end, we construct adaptive procedures that combine eigenvalue screening with blockwise thresholding of prediction energy. We believe that these techniques will be of independent interest for other functional data problems with discretization and measurement noise.

1.1. *Related work.* Fully observed scalar-on-function regression has a well-developed minimax and regularization theory. Classical results develop generalized, principal component, smoothing, and prediction-oriented methods for the functional linear model; in particular, Müller and Stadtmüller (2005) studied generalized functional linear models, Cai and Hall (2006) separated prediction from slope recovery, and Hall and Horowitz (2007) established convergence rates for principal component estimators. The RKHS formulation of Yuan and Cai (2010) and the adaptive prediction theory of Cai and Yuan (2012); Cai, Zhang and Zhou (2018) provide benchmark rates when the covariate trajectories are observed as functions. Related inverse problem viewpoints include the Le Cam equivalence result of Meister (2011), and recent spectral, general regularization, or unified model analyses include Fan, Guo and Shi (2024); Gupta, Sivananthan and Sriperumbudur (2025); Balasubramanian, Müller and Sriperumbudur (2025).

For sparsely or discretely observed functional data, a central theme is that the sampling frequency within each curve changes the statistical experiment. The PACE methodology of Yao, Müller and Wang (2005a) and the principal component theory of Hall, Müller and Wang (2006) provide foundational tools for recovering latent functional structure from noisy partial trajectories. For mean and covariance structure, Cai and Yuan (2011, 2010) established sharp rates under discrete sampling, and Zhang and Wang (2016) clarified the transition from sparse to dense regimes. More recent work extends this regime analysis to predictive distributions and high-dimensional functional data (Gajardo, Dai and Müller, 2021; Petersen, 2024; Guo et al., 2025).

Several papers address functional regression with sparse, discrete, or noisy covariates directly. Yao, Müller and Wang (2005b) developed a functional regression procedure for sparse longitudinal predictor and response processes using conditional principal component scores. Li and Hsing (2007) and Cardot et al. (2007) studied slope estimation when functional predictors are observed on discrete grids with measurement error, and Ferraty et al. (2012) investigated presmoothing before functional linear regression. The closest minimax comparator is Zhou, Yao and Zhang (2023), which proves that the fully observed rate is attainable under independent random design once $m$ is sufficiently large. However, none of these papers gives a matching $m$-dependent minimax lower bound for the prediction risk.

1.2. *Organization of the paper.* The rest of the paper is organized as follows. Section 2 introduces the problem formulation, including the model, parameter spaces, and prediction and estimation risks. Section 3 studies independent design: it states the sampling assumption, gives the oracle upper bound, proves the lower bound, and constructs the adaptive estimator and upper bound. Section 4 follows the same structure for the common design problem on an equal grid, where the phase diagram includes the additional terms from fixed grid discretization and the identification of unknown eigenvalues. Section 5 presents simulation studies and a real data example. Section 6 returns to the main conclusions, provides the corresponding $L^2$ estimation rates, and discusses the current limitations and next open questions. All proofs and technical results are collected in the Supplementary Material (Cai and Li, 2026).

We use the following notation throughout the paper. Let $L^2([0,1])$ be the space of square integrable functions on $[0,1]$ with inner product $\langle\cdot,\cdot\rangle_{L^2}$ and norm $\|\cdot\|_{L^2}$. The expectation operator is denoted by $\mathbb{E}$, with subscripts indicating the variable of integration when needed. For real numbers $a$ and $b$, write $a\wedge b=\min\{a,b\}$ and $a\vee b=\max\{a,b\}$, and let $\lfloor x\rfloor$ be the integer part of $x$. For a finite set $A$, $|A|$ denotes its cardinality, and $\mathbf{1}\{\cdot\}$ denotes the indicator function. For two nonnegative quantities $a$ and $b$, the notation $a\lesssim b$ means that $a\le Cb$ for a constant $C$ independent of $n$, $m$, and the tuning parameters under discussion; $a\gtrsim b$ and $a\asymp b$ are defined analogously. The constants hidden in these comparisons may depend on fixed model parameters such as $c_\lambda, C_\lambda, R, \sigma_\varepsilon, \sigma_\delta$, and on fixed compact ranges of smoothness parameters when uniform adaptive statements are considered.

**2. Problem Formulation.** We use the following functional linear regression model and loss criteria. Let $X(\cdot)$ be a centered Gaussian process on the unit interval $[0,1]$ with continuous path. Let $K(s,t)=\mathbb{E}[X(s)X(t)]$ be its covariance function and $\Sigma$ be the covariance operator defined as $(\Sigma f)(t)=\int_0^1 K(s,t)f(s)\,\mathrm{d}s$ for any $f\in L^2[0,1]$. Let $\{X_i(\cdot)\}_{i=1}^n$ be independent copies of $X$. The response $Y_i$ follows the scalar-on-function regression model

$$Y_i=\int_0^1 X_i(t)\beta(t)\,\mathrm{d}t+\varepsilon_i=\langle X_i,\beta\rangle_{L^2}+\varepsilon_i, \tag{1}$$

where $\beta$ is an unknown slope function and $\varepsilon_i$ are independent $\mathcal{N}(0, \sigma_\varepsilon^2)$ errors.

The predictor trajectories $X_i$ are observed only through noisy point evaluations. Let $m$ be the number of sampling locations per trajectory and let $\{t_{ij} : j = 1, \ldots, m\}$ be the sampling locations for the $i$-th trajectory. The observed point evaluations are

$$Z_{ij} = X_i(t_{ij}) + \delta_{ij}, \tag{2}$$

where $\delta_{ij}$ are independent $\mathcal{N}(0, \sigma_\delta^2)$ measurement errors with $\sigma_\delta^2 > 0$. The families $\{\delta_{ij}\}$, $\{\varepsilon_i\}$, and $\{X_i\}$ are mutually independent. Throughout the paper, $\varepsilon_i$ denotes the scalar response error and $\delta_{ij}$ denotes the pointwise measurement error. The two sampling schemes for $t_{ij}$ are specified later in Assumption 1 and Assumption 2.

Given an estimator $\hat{\beta}$ of the slope function $\beta$, our primary performance criterion is the excess prediction risk

$$\mathbb{E}_\star \left[ \langle X_\star, \hat{\beta} - \beta \rangle_{L^2}^2 \right], \tag{3}$$

where $X_\star$ is an independent test trajectory drawn from the same distribution as the $X_i$'s and $\mathbb{E}_\star$ denotes the expectation with respect to $X_\star$. We also consider slope recovery under the squared $L^2$ estimation loss $\|\hat{\beta} - \beta\|_{L^2}^2$. The main theorems below are stated for prediction risk; the corresponding rates for estimation risk are discussed in Section 6.

To focus on the impact of discrete noisy sampling and avoid technicalities related to the choice of basis, we work with the fixed orthonormal trigonometric basis on $[0, 1]$

$$e_0(t) := 1, \qquad e_r(t) := \sqrt{2}\cos(2\pi r t), \qquad e_{-r}(t) := \sqrt{2}\sin(2\pi r t), \quad r \geq 1. \tag{4}$$

The choice of the index set $\mathbb{Z}$ is for notational convenience, with $r = 0$ corresponding to the constant function, positive indices labeling cosine functions, and negative indices labeling sine functions. Then, every square integrable function $f$ is represented as

$$f = \sum_{r \in \mathbb{Z}} a_r e_r, \qquad a_r = \langle f, e_r \rangle_{L^2} = \int_0^1 f(t) e_r(t)\, \mathrm{d}t, \qquad a_r \in \mathbb{R}.$$

Throughout the paper, we assume that $\Sigma$ is diagonalized by the trigonometric basis $(e_r)_{r \in \mathbb{Z}}$ with eigenvalues $(\lambda_r)_{r \in \mathbb{Z}}$. It is satisfied, for example, by periodic stationary covariance kernels $K(s, t) = \kappa(s - t \bmod 1)$, which arise naturally for many periodic data settings, such as daily or weekly trajectories, when covariance depends on circular lag. This assumption lets us focus on the statistical cost created by noisy point evaluations and by the sampling geometry, rather than on the additional impact of estimating the covariance eigenbasis or the alignment between the slope function and the covariance geometry. In a principal component formulation, the covariance operator is diagonal in its own population eigenbasis, but that basis is unknown and covariance dependent. The need of estimating the covariance eigenbasis from discretely observed data may introduce an additional source of error, which is not the main focus of this paper. Moreover, imposing coefficient decay or smoothness of $\beta$ in that basis ties the slope class to the covariance geometry, thereby introducing an alignment issue between covariance and slope regularity (Cai and Yuan, 2012; Gupta, Sivananthan and Sriperumbudur, 2025) that is already present in the fully observed case. Therefore, we instead use the fixed trigonometric basis both to diagonalize the covariance operator and to unify the smoothness classes for the predictor trajectories $X_i$ and for the slope function $\beta$.

Consequently, the predictor trajectories $X_i$ have the Karhunen–Loève expansion

$$X_i(t) = \sum_{r \in \mathbb{Z}} x_{ir} e_r(t), \qquad \mathbb{E} x_{ir}^2 = \lambda_r, \tag{5}$$

where the Fourier scores are independent centered Gaussian variables. We also assume that the unknown slope function $\beta$ has the Fourier expansion

$$\beta(t) = \sum_{r\in\mathbb{Z}} \theta_r e_r(t), \qquad \theta_r \in \mathbb{R}. \tag{6}$$

We introduce the cross-covariance function as

$$g(t) := \mathbb{E}[Y_1 X_1(t)] = (\Sigma\beta)(t) = \sum_{r\in\mathbb{Z}} \gamma_r e_r(t), \qquad \gamma_r := \lambda_r \theta_r. \tag{7}$$

With the fixed trigonometric basis, the excess prediction risk can be expressed as a weighted $\ell^2$ error of the Fourier coefficients. If $\hat\beta$ has the Fourier expansion $\hat\beta = \sum_{r\in\mathbb{Z}} \hat\theta_r e_r$, then

$$\mathcal{R}(\hat\beta;\theta,\lambda) := \mathbb{E}_\star\left[\langle X_\star, \hat\beta - \beta\rangle^2\right] = \sum_{r\in\mathbb{Z}} \lambda_r |\hat\theta_r - \theta_r|^2. \tag{8}$$

Since $\hat\beta$ depends on the training data, $\mathcal{R}(\hat\beta;\theta,\lambda)$ is the test risk conditional on that training sample. The minimax bounds below use the outer expectation $\mathbb{E}\mathcal{R}(\hat\beta;\theta,\lambda)$, taken over all training randomness, including the latent trajectories, design points, response errors, and measurement errors.

For $\alpha > 1/2$, we define the eigenvalue class

$$\mathcal{L}_\alpha(c_\lambda, C_\lambda) := \left\{(\lambda_r)_{r\in\mathbb{Z}} : c_\lambda(1+|r|)^{-2\alpha} \le \lambda_r \le C_\lambda(1+|r|)^{-2\alpha}\right\} \tag{9}$$

where $0 < c_\lambda \le C_\lambda < \infty$ are fixed constants. The requirement $\alpha > 1/2$ is necessary to ensure that the eigenvalues are summable and the predictor trajectories $X_i$ are continuous almost surely. For $s \ge 0$ and $R_0 > 0$, the slope smoothness is indexed by the Sobolev ball

$$\Theta_s(R_0) := \left\{\theta = (\theta_r)_{r\in\mathbb{Z}} : \sum_{r\in\mathbb{Z}} (1+|r|)^{2s}\theta_r^2 \le R_0^2\right\}. \tag{10}$$

The condition $\theta \in \Theta_s(R_0)$ is equivalent to assuming that the slope function $\beta$ belongs to the periodic Sobolev space $H^s([0,1])$. Throughout the paper, we assume that $\alpha > 1/2$ and $s \ge 0$ without further mention, which is the minimal regularity range for the problem to be well defined.

## 3. Independent Design.

Under independent design, each subject is observed on an independent random grid. Formally, we introduce the following assumption on the design points.

ASSUMPTION 1 (Independent design). The design points satisfy $t_{ij} \overset{\text{iid}}{\sim} \mathrm{Unif}[0,1]$, independently across $i$ and $j$, and independently of all other random quantities.

The results in this section show that discretization affects the risk only through an additional statistical measurement term.

### 3.1. *Oracle Estimator.*

We begin with an oracle estimator that depends on the eigenvalue sequence $(\lambda_r)_{r\in\mathbb{Z}}$ and uses a deterministic cutoff $d$, which may depend on the smoothness indices $(\alpha, s)$. Define the cross-covariance coefficient estimator for $\gamma_r$ by

$$\bar Z_{ir} := \frac{1}{m}\sum_{j=1}^m Z_{ij} e_r(t_{ij}), \qquad \hat\gamma_r := \frac{1}{n}\sum_{i=1}^n Y_i \bar Z_{ir}. \tag{11}$$

Under independent design, $\hat{\gamma}_r$ is an unbiased estimator of $\gamma_r$. Since $\gamma_r = \lambda_r\theta_r$, define the plug-in estimator

$$\tilde{\beta}(t) \coloneqq \sum_{|r|\le d} \lambda_r^{-1}\hat{\gamma}_r e_r(t), \tag{12}$$

where the cutoff $d$ is a tuning parameter. The oracle estimator satisfies the following upper bound on the excess prediction risk.

THEOREM 3.1 (Oracle upper bound under independent design). *Assume that $\lambda \in \mathcal{L}_\alpha(c_\lambda, C_\lambda)$ is known. Under Assumption 1, for every $d \ge 1$,*

$$\sup_{\lambda\in\mathcal{L}_\alpha(c_\lambda,C_\lambda)} \sup_{\theta\in\Theta_s(R_0)} \mathbb{E}\mathcal{R}(\tilde{\beta};\theta,\lambda) \lesssim \frac{d}{n} + \frac{d^{2\alpha+1}}{nm} + d^{-(2\alpha+2s)}. \tag{13}$$

*Consequently, take the optimized cutoff*

$$d^* \asymp n^{\frac{1}{2\alpha+2s+1}} \wedge (nm)^{\frac{1}{4\alpha+2s+1}} \tag{14}$$

*and let $\tilde{\beta}^*$ be the corresponding estimator. Then*

$$\sup_{\lambda\in\mathcal{L}_\alpha(c_\lambda,C_\lambda)} \sup_{\theta\in\Theta_s(R_0)} \mathbb{E}\mathcal{R}(\tilde{\beta}^*;\theta,\lambda) \lesssim n^{-\frac{2\alpha+2s}{2\alpha+2s+1}} + (nm)^{-\frac{2\alpha+2s}{4\alpha+2s+1}}. \tag{15}$$

The proof of Theorem 3.1 is deferred to Appendix B of the Supplementary Material (Cai and Li, 2026). We remark here that $d/n$ is the standard estimation contribution from estimating $d$ coordinates in fully observed FLR, $d^{-(2\alpha+2s)}$ is the prediction tail, and $d^{2\alpha+1}/(nm)$ is the cost of estimating $\gamma_r$ from noisy point evaluations before division by $\lambda_r$.

3.2. *Lower Bound.* We next provide a matching lower bound.

THEOREM 3.2 (Lower bound under independent design). *Under Assumption 1, we have*

$$\inf_{\hat{\beta}} \sup_{\lambda\in\mathcal{L}_\alpha(c_\lambda,C_\lambda)} \sup_{\theta\in\Theta_s(R_0)} \mathbb{E}\mathcal{R}(\hat{\beta};\theta,\lambda) \gtrsim n^{-\frac{2\alpha+2s}{2\alpha+2s+1}} + (nm)^{-\frac{2\alpha+2s}{4\alpha+2s+1}}. \tag{16}$$

The proof of Theorem 3.2 is based on a van Trees argument over Gaussian block submodels tailored to the independent random design. The main difficulty is to identify the impact of discrete noisy observations on the Fisher information. A straightforward reduction to ordinary nonparametric regression only recovers the dense observation contribution. To capture the cost of discretization, our refined analysis provides an alternative upper bound of the Fisher information that scales with $m$, leading to the second term in the lower bound. In addition, we have to choose an optimized block size $d$ to balance the two terms and prior information, which turns out to mirror the decomposition in the upper bound in (13). The detailed proof is given in Appendix D of the Supplementary Material (Cai and Li, 2026).

3.3. *Adaptive Estimator.* We now construct an adaptive estimator that attains the minimax rate without knowledge of $(\lambda_r)_{r\in\mathbb{Z}}$, $\alpha$, or $s$. The main difficulty is that adaptation cannot be separated from covariance estimation. The oracle risk balance is expressed in the prediction norm (8), so the scale at which a frequency component should be kept or discarded depends on the eigenvalue $\lambda_r$. The coefficient itself is also recovered through the inverse relation $\theta_r = \gamma_r/\lambda_r$. Thus the eigenvalues enter both the selection step and the final inversion step. A plug-in adaptive procedure must therefore estimate $(\lambda_r)_{r\in\mathbb{Z}}$ accurately enough to calibrate the selection rule, while preventing small or poorly estimated eigenvalues from producing unstable denominators. This differs from standard block thresholding in fully observed FLR because both the block energy and the local prediction metric are unknown. The same covariance pilot that stabilizes the inverse estimator $\hat{\gamma}_r/\check{\lambda}_r$ also determines whether the empirical prediction energy is meaningful on a block. To this end, we partition the subjects into two disjoint groups

$$\{1,\dots,n\} = I_\lambda \cup I_\gamma, \quad n_\lambda := |I_\lambda|, \quad n_\gamma := |I_\gamma|, \quad n_\lambda \asymp n_\gamma \asymp n, \tag{17}$$

where $I_\lambda$ is used for the covariance pilot and $I_\gamma$ is used for the cross-covariance coefficients. This sample splitting makes the covariance pilot independent of the cross-covariance coefficient estimator, simplifying the analysis of the expected prediction error. If one only aims for an upper bound in high probability, this independence is not essential and the same upper bound can be given without splitting the subjects. In practice, it is therefore preferable to use all subjects in both steps, which typically improves the constant factors.

We begin with the eigenvalue pilot. For each $i \in I_\lambda$, split the indices within each curve into two disjoint sets

$$J_i^{(1)} \cup J_i^{(2)} = \{1,\dots,m\}, \qquad m_i^{(a)} := |J_i^{(a)}| \asymp m, \quad a = 1,2.$$

For $r \in \mathbb{Z}$ and $a = 1,2$, define

$$U_{ir}^{(a)} := \frac{1}{m_i^{(a)}} \sum_{j \in J_i^{(a)}} Z_{ij} e_r(t_{ij}),$$

and set

$$\check{\lambda}_r := \frac{1}{n_\lambda} \sum_{i \in I_\lambda} U_{ir}^{(1)} U_{ir}^{(2)}. \tag{18}$$

The split within each curve makes the two averages conditionally independent given the latent curve, removing the leading bias from measurement noise; in particular, $\check{\lambda}_r$ is unbiased for $\lambda_r$.

On the regression group, estimate the cross-covariance coefficients similarly to the oracle case by

$$\hat{\gamma}_r := \frac{1}{n_\gamma} \sum_{i \in I_\gamma} Y_i \bar{Z}_{ir}. \tag{19}$$

The adaptive selection is carried out on a deterministic active frequency band. Let the largest active dyadic block index and the corresponding symmetric cutoff level be

$$L_n := \left\lfloor \log_2 \left( C_L \min\{n^{1/2}, (nm)^{1/3}\} \right) \right\rfloor, \qquad K_n := 2^{L_n}, \tag{20}$$

where $C_L > 0$ is a fixed numerical constant. Then $K_n \asymp n^{1/2} \wedge (nm)^{1/3}$, which is large enough for the oracle comparison in (14) uniformly on compact subsets of $\{\alpha > 1/2,\ s \geq 0\}$. Define the dyadic blocks by $B_0 = \{0, \pm 1\}$ and $B_\ell = \{r : 2^{\ell-1} < |r| \leq 2^\ell\}$, for $\ell \geq 1$. For each active block $B_\ell$, $0 \leq \ell \leq L_n$, define the empirical prediction energy and its variance proxy by

$$\hat{S}_\ell := \sum_{r \in B_\ell} \frac{|\hat{\gamma}_r|^2}{\check{\lambda}_r}, \qquad \hat{V}_\ell := \frac{C_V}{n} \sum_{r \in B_\ell} \left( 1 + \frac{1}{m \check{\lambda}_r} \right). \tag{21}$$

Here $C_V > 0$ is a fixed sufficiently large constant. The retention rule requires both a reliable covariance pilot on the block and empirical energy above the variance proxy. We define the thresholding level for the covariance pilot by

$$\zeta_{n,m} := M_0 m^{-1} \sqrt{\frac{\log(nm)}{n_\lambda}}, \tag{22}$$

where $M_0 > 0$ is a fixed sufficiently large constant.

The thresholded coefficient estimator is

$$\hat{\theta}_r := \frac{\hat{\gamma}_r}{\check{\lambda}_r} \mathbf{1}\left\{ \min_{u \in B_\ell} \check{\lambda}_u \geq 2\zeta_{n,m} \text{ and } \hat{S}_\ell \geq \hat{V}_\ell \right\}, \quad \text{for } r \in B_\ell,\ 0 \leq \ell \leq L_n. \tag{23}$$

The first thresholding is used to avoid unstable division by small or poorly estimated eigenvalues, so that $\check{\lambda}_r$ is positive on the retained block and the division is well-defined. The second thresholding keeps only the significant components in terms of the prediction norm for adaptivity. Set $\hat{\theta}_r = 0$ for $|r| > K_n$. Write $\hat{\theta}_{B_\ell} = (\hat{\theta}_r)_{r \in B_\ell}$ for the estimated coordinates in the block $B_\ell$.

Finally, we add a projection step for stability. Fix a sufficiently large constant $R \geq R_0$. Let $\Pi_R$ denote the projection operator $\Pi_R(u) := \min(1, R/\|u\|_2)u$. Define the projected coordinates by

$$\bar{\theta}_{B_\ell} := \Pi_R\left(\hat{\theta}_{B_\ell}\right), \qquad 0 \leq \ell \leq L_n.$$

Writing $\bar{\theta}_r$ for the resulting coordinates, define the final estimator

$$\hat{\beta}(t) := \sum_{|r| \leq K_n} \bar{\theta}_r e_r(t). \tag{24}$$

THEOREM 3.3 (Adaptive upper bound under independent design). *Under Assumption 1, uniformly over any compact subset of $\{(\alpha, s) : \alpha > 1/2,\ s \geq 0\}$, the estimator (24) satisfies*

$$\sup_{\lambda \in \mathcal{L}_\alpha(c_\lambda, C_\lambda)} \sup_{\theta \in \Theta_s(R_0)} \mathbb{E}\mathcal{R}(\hat{\beta}; \theta, \lambda) \lesssim n^{-\frac{2\alpha+2s}{2\alpha+2s+1}} + (nm)^{-\frac{2\alpha+2s}{4\alpha+2s+1}}. \tag{25}$$

Theorem 3.3 shows that the adaptive estimator under independent design attains the minimax rate without knowledge of $(\lambda_r)_{r \in \mathbb{Z}}$, $\alpha$, or $s$, with no additional cost in the rate. The proof controls the error from the covariance pilot, the cross-covariance estimator, and the blockwise selection rule in the prediction norm. The key point is that the eigenvalue threshold prevents unstable inversion, while the empirical prediction energy threshold retains only blocks whose signal is large relative to the dense variance and the variance due to noisy measurements. The proof is given in Appendix C of the Supplementary Material (Cai and Li, 2026).

3.4. *Discussion.* We now discuss the implications of the results under independent design. Let

$$\mathcal{R}_{n,m}^{\star,\mathrm{ind}} := \inf_{\hat{\beta}} \sup_{\lambda \in \mathcal{L}_\alpha(c_\lambda, C_\lambda)} \sup_{\theta \in \Theta_s(R_0)} \mathbb{E}\mathcal{R}(\hat{\beta}; \theta, \lambda),$$

where the infimum and risk are evaluated under Assumption 1. Combining the adaptive upper bound with the lower bound yields the following minimax rate:

COROLLARY 3.4 (Minimax rate under independent design). *Under Assumption 1, $\alpha > 1/2$, and $s \geq 0$,*

$$\mathcal{R}_{n,m}^{\star,\mathrm{ind}} \asymp \underbrace{n^{-\frac{2\alpha+2s}{2\alpha+2s+1}}}_{\mathrm{I}} + \underbrace{(nm)^{-\frac{2\alpha+2s}{4\alpha+2s+1}}}_{\mathrm{II}}. \tag{26}$$

Under independent design, the rate is the larger of the fully observed FLR rate (Yuan and Cai, 2010; Cai and Yuan, 2012) and a discretization term due to measurement noise. This yields a sharp phase transition in the resolution $m$ within each curve. When $m$ is small, the discretization term dominates; when $m$ is large, the fully observed term dominates. Let $\zeta_* = 2\alpha/(2\alpha+2s+1)$ be the critical exponent for $m$, $\nu = 2\alpha+2s$ and $\kappa = \nu+2\alpha+1$. The rate is summarized as

$$\mathcal{R}^{\star,\mathrm{ind}}_{n,m} \asymp \begin{cases} (nm)^{-\nu/\kappa}, & m \le n^{\zeta_*}, \\ n^{-\nu/(\nu+1)}, & m \ge n^{\zeta_*}. \end{cases}$$

In the sparse regime, the cost of discretization under independent design enters through the number $nm$ of effective scalar observations. The exponent $\frac{2\alpha+2s}{4\alpha+2s+1}$ comes from the interplay between the smoothness of the slope function and the eigenvalue decay, as well as the inverse-covariance amplification incurred when converting the cross-covariance coefficient estimation error to the slope coefficient estimation error. As a result, the exponent in the discretization term is smaller than that in the fully observed term.

This sparse to dense transition is similar to that of mean function estimation where the optimal rate is given by $n^{-1} + (mn)^{-(2\alpha-1)/(2\alpha)}$ under independent design (Cai and Yuan, 2011). For mean function estimation, the fully observed rate is parametric but the discretization term is nonparametric, whereas for FLR, both terms are nonparametric. In both settings, the exponents in the two terms differ, and the discretization term has the smaller exponent.

We also compare our result to the most closely related work of Zhou, Yao and Zhang (2023), who study the independent random design and provide sufficient conditions for attaining the fully observed benchmark rate in the dense regime with noisy discrete covariates. Translating their notation to ours, their sufficient sampling condition per curve is $m \gtrsim n^{\zeta_{\mathrm{ZYZ}}}$ with $\zeta_{\mathrm{ZYZ}} = \frac{4\alpha+2}{2\alpha+2s+1} \vee \frac{2\alpha+s+1/2}{2\alpha+2s+1}$. This exponent satisfies $\zeta_{\mathrm{ZYZ}} > \zeta_*$ for all $\alpha > 1/2$ and $s \ge 0$, and the gap can be arbitrarily close to 1 when $\alpha$ is large and $s$ is small, so the sufficient condition in Zhou, Yao and Zhang (2023) is not tight. In contrast, by showing a matching lower bound, particularly with respect to $m$, our result establishes the sharp phase transition in the resolution within each curve and provides a complete characterization of the cost of discretization under independent design.

Finally, we consider optimal sampling allocation under a fixed total sampling budget. This question is relevant when the total sampling budget is fixed by cost or other constraints, while the allocation between the number of subjects and the number of measurements per subject remains flexible. Under a fixed total scalar sampling budget $nm = N$, where $N$ denotes the total number of scalar point measurements and $\nu, \kappa$ are as above, the rate under independent design with fixed budget is

$$\inf_{nm=N} \mathcal{R}^{\star,\mathrm{ind}}_{n,m} \asymp N^{-\nu/\kappa},$$

which is attained by any allocation satisfying $m \lesssim N^{2\alpha/\kappa}$. Thus, under independent design with a fixed total budget, the rate favors more subjects with fewer measurements per subject over fewer subjects with more measurements per subject.

## 4. Common Design.

Under common design, all trajectories are observed on the same deterministic grid. This setting is natural for many regularly sampled data sets, for example when measurements are recorded every minute or every hour on a common schedule. We impose the following deterministic design assumption.

ASSUMPTION 2 (Common design). All trajectories share the deterministic equal grid

$$t_{ij} = t_j = \frac{j-1}{m}, \qquad j = 1, \ldots, m.$$

The shared grid changes the role of discretization. Unlike independent design, the sampling geometry is not averaged out across subjects, so frequencies with the same trace on the grid remain statistically coupled even as $n$ increases. The minimax rate under common design contains the same two statistical terms as in the case of independent design. The shared grid also creates a term from fixed grid discretization and, when the eigenvalue sequence is unknown, an additional term from eigenvalue identification.

4.1. *Oracle Estimator.* We first study an oracle estimator that knows the eigenvalue sequence $(\lambda_r)_{r\in\mathbb{Z}}$. The construction is the same as the oracle estimator under independent design in (12) with a cutoff level $d$. However, under common design, the cross-covariance estimator $\hat{\gamma}$ in (11) is no longer an unbiased estimator of the population coefficient $\gamma_r$, since bias can arise from aliasing of distinct frequencies on the shared grid. Here and below, "known eigenvalues" means that the estimator is allowed to use the particular eigenvalue sequence $(\lambda_r)_{r\in\mathbb{Z}}$. The following upper bound shows that an additional term from fixed grid discretization appears in the oracle risk.

THEOREM 4.1 (Oracle upper bound under common design). *Assume that $\lambda \in \mathcal{L}_\alpha(c_\lambda, C_\lambda)$ is known. Under Assumption 2, for every integer $1 \le d \le m/4$,*

$$\sup_{\theta\in\Theta_s(R_0)} \mathbb{E}\mathcal{R}(\tilde{\beta};\theta,\lambda) \lesssim \frac{d}{n} + \frac{d^{2\alpha+1}}{nm} + d^{-(2\alpha+2s)} + m^{-(2\alpha+2s)}. \tag{27}$$

*Take the optimized cutoff*

$$d^* \asymp m \wedge n^{\frac{1}{2\alpha+2s+1}} \wedge (nm)^{\frac{1}{4\alpha+2s+1}}. \tag{28}$$

*Let $\tilde{\beta}^*$ denote the estimator $\tilde{\beta}$ constructed with $d = d^*$. Then*

$$\sup_{\theta\in\Theta_s(R_0)} \mathbb{E}\mathcal{R}(\tilde{\beta}^*;\theta,\lambda) \lesssim n^{-\frac{2\alpha+2s}{2\alpha+2s+1}} + (nm)^{-\frac{2\alpha+2s}{4\alpha+2s+1}} + m^{-(2\alpha+2s)}. \tag{29}$$

The proof is given in Appendix E of the Supplementary Material (Cai and Li, 2026). The term $d/n$ is the usual finite-dimensional estimation contribution. The term $d^{2\alpha+1}/(nm)$ is the cost due to measurement noise after division by the low frequency eigenvalues, and $d^{-(2\alpha+2s)}$ is the bias of truncation. The additional term $m^{-(2\alpha+2s)}$ represents the reconstruction error left by the fixed grid. These terms are discussed further in Subsection 4.4.

4.2. *Lower Bound.* We next prove minimax lower bounds under common design. The first part is the oracle experiment with a fixed known eigenvalue sequence, and its lower bound contains the two statistical terms from independent design together with the fixed grid discretization term. The second part is the genuine unknown eigenvalue experiment; in that case an additional identification term from the common grid is unavoidable.

THEOREM 4.2 (Lower bound under common design). *Let Assumption 2 hold. If $\lambda \in \mathcal{L}_\alpha(c_\lambda, C_\lambda)$ is fixed and known, then*

$$\inf_{\hat{\beta}} \sup_{\theta\in\Theta_s(R_0)} \mathbb{E}\mathcal{R}(\hat{\beta};\theta,\lambda) \gtrsim n^{-\frac{2\alpha+2s}{2\alpha+2s+1}} + (nm)^{-\frac{2\alpha+2s}{4\alpha+2s+1}} + m^{-(2\alpha+2s)}. \tag{30}$$

*Moreover, in the case when $\lambda$ is unknown, we have*

$$\inf_{\hat{\beta}} \sup_{\lambda\in\mathcal{L}_\alpha(c_\lambda,C_\lambda)} \sup_{\theta\in\Theta_s(R_0)} \mathbb{E}\mathcal{R}(\hat{\beta};\theta,\lambda) \gtrsim n^{-\frac{2\alpha+2s}{2\alpha+2s+1}} + (nm)^{-\frac{2\alpha+2s}{4\alpha+2s+1}} + m^{-(2\alpha+2s)} + m^{-4\alpha}. \tag{31}$$

Theorem 4.2 is proved in Appendix G of the Supplementary Material (Cai and Li, 2026). The two statistical terms are obtained through the same van Trees strategy as in the case of independent design, with the Fisher information evaluated for the deterministic grid experiment. The term $m^{-(2\alpha+2s)}$ is proved by constructing alternatives whose induced cross-covariance functions vanish on the grid and are therefore indistinguishable no matter how large $n$ is. The $m^{-4\alpha}$ term enters only when the eigenvalues are unknown; it is obtained from alternatives with the same law on the grid but different covariance scales for inversion. These two lower bound constructions are qualitatively different. The first keeps the covariance scale fixed and hides regression signal between grid points through the cross-covariance $g = \Sigma\beta$. The second varies the covariance scale itself, showing that even when the observed grid distribution is unchanged, the inverse-regression target can differ enough to create the $m^{-4\alpha}$ term.

4.3. *Adaptive Estimator.* We now construct an adaptive estimator for common design with unknown eigenvalues and unknown smoothness indices. The construction follows the adaptive estimator under independent design in Subsection 3.3, with modifications in the covariance pilot to account for the shared grid. The main difficulty is that, under a common grid, the eigenvalues cannot be estimated in the same way as in the independent-design case. When $m$ is even, one may attempt to use an interlaced partition of the grid to construct an estimator for $\lambda_r$ similar to (18), but the same approach fails for odd $m$ because of nonnegligible bias. We therefore introduce a covariance pilot based on the difference between the active band and a fixed floor at high frequency; this pilot is accurate enough for blockwise adaptation. The role of this pilot is twofold: it removes the measurement-noise contribution from grid Fourier coefficients and provides a scale for the inverse operation in the blockwise selection rule.

We split the subjects into two groups as before, denoted by $I_\lambda \cup I_\gamma = \{1, \ldots, n\}$ with $n_\lambda \asymp n_\gamma \asymp n$. With $\bar{Z}_{ir}$ defined as in (11) (so here $t_{ij} = t_j$), we define the pilot estimator for $\lambda_r$ as the difference between the active band and a fixed floor at high frequency:

$$\check{\lambda}_r := \frac{1}{n_\lambda} \sum_{i \in I_\lambda} \left( \bar{Z}_{ir}^2 - \bar{Z}_{iH_m}^2 \right), \tag{32}$$

where we take $H_m = \lfloor (m-1)/2 \rfloor$. This subtraction removes the additional variance contribution in $\bar{Z}_{ir}^2$ due to noisy observations. At the same time, the frequency $H_m$ is chosen not too high, to avoid excessive variance, and not too low, so that $\lambda_{H_m}$ is negligible.

Under common design, the active band must be chosen more conservatively to ensure that the bias of $\hat{\gamma}_r$ does not dominate the variance. For a sufficiently small constant $c_L \in (0, 1/8)$, we define the active band cutoff as

$$L_m := \lfloor \log_2 (c_L m \wedge n) \rfloor \qquad K_m := 2^{L_m}.$$

Additionally, the thresholding level must also be modified to reflect the different balance between bias, variance and additional bias created by the shared grid under common design. Instead of (22), we use

$$\zeta_{n,m} := M_0 \left( \frac{1}{m} \sqrt{\frac{\ell_{n,m}^\lambda}{n_\lambda}} + \frac{\ell_{n,m}^\lambda}{n_\lambda} \right), \quad \ell_{n,m}^\lambda := \log\bigl(n(K_m + 1)\bigr), \tag{33}$$

where $M_0 > 0$ is sufficiently large. The adaptive estimator for common design, still denoted by $\hat{\beta}$, is then defined as in (24) with the above modifications. The following upper bound states that the adaptive estimator attains the minimax rate. The proof is given in Appendix F of the Supplementary Material (Cai and Li, 2026).

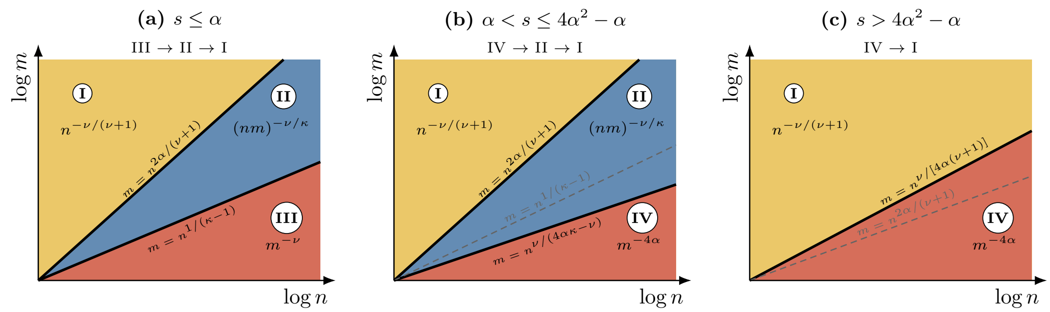


FIG 1. *Phase diagram for the minimax rate under common design. The dominant term in (35) changes with the grid size $m$ within each trajectory, and the transition points depend on the relative size of $s$ and $\alpha$.*

THEOREM 4.3 (Adaptive upper bound under common design). *Under Assumption 2, uniformly over any compact subset of $\{(\alpha, s) : \alpha > 1/2,\ s \ge 0\}$, we have*

$$\sup_{\lambda \in \mathcal{L}_\alpha(c_\lambda, C_\lambda)} \sup_{\theta \in \Theta_s(R_0)} \mathbb{E}\mathcal{R}(\hat{\beta}; \theta, \lambda) \lesssim n^{-\frac{2\alpha+2s}{2\alpha+2s+1}} + (nm)^{-\frac{2\alpha+2s}{4\alpha+2s+1}} + m^{-(2\alpha+2s)} + m^{-4\alpha}. \tag{34}$$

Theorem 4.3 shows that the adaptive procedure matches all four terms in the lower bound when the eigenvalues are unknown under common design. In particular, the fourth term in (34) should not be interpreted as a technical price of the adaptive construction. It is the minimax obstruction created by unknown covariance scales on a shared grid, as reflected by the lower bound in (31). The last two terms have different origins: $m^{-(2\alpha+2s)}$ is the term from fixed grid discretization already present with known eigenvalues, whereas $m^{-4\alpha}$ is the additional cost of estimating a covariance scale that can be aliased on the common grid. Keeping these terms separate is important because they dominate in different smoothness regimes and lead to different phase transitions.

4.4. *Discussion.* Denote the minimax risk under common design as

$$\mathcal{R}^{\star,\mathrm{cd}}_{n,m} := \inf_{\hat{\beta}} \sup_{\lambda \in \mathcal{L}_\alpha(c_\lambda, C_\lambda)} \sup_{\theta \in \Theta_s(R_0)} \mathbb{E}\mathcal{R}(\hat{\beta}; \theta, \lambda).$$

Combining Theorem 4.2 and Theorem 4.3 yields the following minimax rate.

COROLLARY 4.4 (Minimax rate under common design). *Under Assumption 2,*

$$\mathcal{R}^{\star,\mathrm{cd}}_{n,m} \asymp \underbrace{n^{-\frac{2\alpha+2s}{2\alpha+2s+1}}}_{\mathrm{I}} + \underbrace{(nm)^{-\frac{2\alpha+2s}{4\alpha+2s+1}}}_{\mathrm{II}} + \underbrace{m^{-(2\alpha+2s)}}_{\mathrm{III}} + \underbrace{m^{-4\alpha}}_{\mathrm{IV}}. \tag{35}$$

The minimax rate in (35) has four terms. The first two terms are the same as in the case of independent design: the term (I) is the fully observed FLR benchmark, and the term (II) is the contribution caused by noisy point evaluations. The last two terms are specific to the common grid. The term (III) is the approximation error from the fixed grid that is already present in the oracle analysis on an equal grid, whereas the term (IV) is the additional cost of identifying the covariance scale needed for inversion when the eigenvalues are unknown. We next discuss several implications of this minimax rate.

4.4.1. *Phase transitions.* We identify the phase transitions between the four rates across different regimes of $m$ and $(\alpha, s)$. To simplify the presentation, write

$$\nu := 2\alpha + 2s, \qquad \kappa := 4\alpha + 2s + 1.$$

Figure 1 displays the following regimes:

(a) If $s \le \alpha$, the impact of unknown eigenvalues is negligible compared to the approximation error from the fixed grid, so the rate coincides with the oracle rate and has the form

$$\mathcal{R}^{\star,\mathrm{cd}}_{n,m} \asymp \begin{cases} m^{-\nu}, & m \le n^{1/(\kappa-1)}, \\ (nm)^{-\nu/\kappa}, & n^{1/(\kappa-1)} \le m \le n^{2\alpha/(\nu+1)}, \\ n^{-\nu/(\nu+1)}, & m \ge n^{2\alpha/(\nu+1)}. \end{cases}$$

(b) If $\alpha < s \le 4\alpha^2 - \alpha$, then the term for identifying unknown eigenvalues dominates the approximation error from the fixed grid at low resolution, while the sampling term still determines an intermediate regime. We have

$$\mathcal{R}^{\star,\mathrm{cd}}_{n,m} \asymp \begin{cases} m^{-4\alpha}, & m \le n^{\nu/(4\alpha\kappa-\nu)}, \\ (nm)^{-\nu/\kappa}, & n^{\nu/(4\alpha\kappa-\nu)} \le m \le n^{2\alpha/(\nu+1)}, \\ n^{-\nu/(\nu+1)}, & m \ge n^{2\alpha/(\nu+1)}. \end{cases}$$

In this range $m^{-4\alpha}$ dominates $m^{-\nu}$ at low resolution, but the sampling term still forms a nonempty intermediate regime.

(c) If $s > 4\alpha^2 - \alpha$, then $m^{-4\alpha}$ is the dominant term when $m$ is small. The transition is instead between the covariance identification term and the fully observed benchmark. We have

$$\mathcal{R}^{\star,\mathrm{cd}}_{n,m} \asymp \begin{cases} m^{-4\alpha}, & m \le n^{\nu/[4\alpha(\nu+1)]}, \\ n^{-\nu/(\nu+1)}, & m \ge n^{\nu/[4\alpha(\nu+1)]}. \end{cases}$$

4.4.2. *Comparison with independent design.* We compare the minimax rate under common design with the rate under independent design in Corollary 3.4. The two statistical terms (I) and (II) coincide in both settings, showing the fundamental statistical difficulty of the problem. However, the sampling geometry changes the nature of discretization, so the discretization cost under common design contains two additional terms (III) and (IV). Under common design, all trajectories are observed through the same grid projection, so increasing $n$ only reduces the stochastic error on this projection, but it does not remove aliasing or errors in identifying the denominators created by the fixed grid.

The threshold for the dense regime is consequently different. For independent design, the fully observed benchmark is reached once $m \gtrsim n^{\zeta_{\mathrm{ind}}}$ with $\zeta_{\text{ind}} = 2\alpha/(\nu+1)$. For common design, the dense regime requires $m \gtrsim n^{\zeta_{\mathrm{ind}} \vee \zeta_{\mathrm{cd}}}$, where $\zeta_{\text{cd}} = \nu/[4\alpha(\nu+1)]$. The second condition is active precisely in the high smoothness range $s > 4\alpha^2 - \alpha$.

4.4.3. *Comparison with mean estimation.* It is instructive to compare (35) with the corresponding minimax theory for functional mean or covariance estimation under discrete noisy observations (Cai and Yuan, 2011, 2010). In mean estimation, the rates under independent and common design take the schematic forms

$$n^{-1} + (mn)^{-(2\alpha-1)/(2\alpha)} \qquad \text{and} \qquad n^{-1} + m^{-(2\alpha-1)},$$

respectively. Both rates enter the dense regime at the same resolution scale $m \asymp n^{1/(2\alpha-1)}$. Moreover, throughout the range $m \lesssim n^{1/(2\alpha-1)}$, the sampling term under independent design is no larger than the term under a common grid, while above this scale both sampling terms are dominated by the parametric term $n^{-1}$. Thus the phase transition for mean estimation remains essentially a transition between two regimes: a sampling-limited regime and a dense parametric regime.

The FLR rate under common design is more intricate because the regression problem is an inverse problem and the slope smoothness $s$ enters the balance. The sampling term is affected by inversion through $\lambda_r^{-1}$, whereas the terms from the fixed grid and from eigenvalue identification have different dependence on $s$. Due to the joint dependence on $(\alpha, s)$, the sampling term (II) is no longer dominated by the additional grid terms (III) and (IV) at low resolution. Consequently, the phase diagram can have up to four distinct regimes.

The term $m^{-4\alpha}$ is specific to this FLR setting and has no analogue in mean estimation. It depends only on the covariance decay and the grid resolution, not on the smoothness $s$ of the slope. Consequently, for sufficiently smooth slopes, the leading obstruction under common design is not the approximation of $\beta$ but the identification of the covariance scale required to transform the cross-covariance into the slope.

4.4.4. *Source of the two grid error terms.* We next give further insight into the two additional grid error terms. From the perspective of the upper bound, term (III) comes from the bias of the cross-covariance coefficient estimator $\hat{\gamma}_r$ due to aliasing on the common grid. Indeed, the shared grid makes $\hat{\gamma}_r$ a biased estimator of the population coefficient $\gamma_r$, and the bias can be large enough to dominate the variance at low frequencies. Term (IV) enters through the need to invert the covariance operator in the regression problem: under the fixed grid, eigenvalue estimation is biased, which leads to a nonnegligible error in the estimation of the slope function.

From the perspective of the lower bound, the two error terms correspond to two geometric obstructions. The term (III) reflects the approximation error created by the fixed grid, since small bumps lying between the grid points cannot be detected by the observations on the grid. When the eigenvalues are unknown, indistinguishability can also arise from the eigenvalues, yielding a separate $m^{-4\alpha}$ identification cost (IV).

4.4.5. *Optimal sampling allocation.* Finally, consider the optimal allocation of the sampling budget $nm = N$ under common design, where $N$ denotes the total number of scalar point measurements and $\nu, \kappa$ are as in the phase-transition display above. Minimizing (35) over $(n, m)$ gives the following regimes:

- If $s \le \alpha$, then
$$\inf_{nm=N} \mathcal{R}^{\star,\mathrm{cd}}_{n,m} \asymp N^{-\nu/\kappa}, \quad \text{for} \quad N^{1/\kappa} \lesssim m \lesssim N^{2\alpha/\kappa}.$$
- If $\alpha < s \le 4\alpha^2 - \alpha$, then
$$\inf_{nm=N} \mathcal{R}^{\star,\mathrm{cd}}_{n,m} \asymp N^{-\nu/\kappa}, \quad \text{for} \quad N^{\nu/(4\alpha\kappa)} \lesssim m \lesssim N^{2\alpha/\kappa}.$$
- If $s > 4\alpha^2 - \alpha$, then the optimal allocation is given by
$$m^\star \asymp N^{\nu/[4\alpha(\nu+1)+\nu]}, \quad n^\star \asymp N^{4\alpha(\nu+1)/[4\alpha(\nu+1)+\nu]},$$
and
$$\inf_{nm=N} \mathcal{R}^{\star,\mathrm{cd}}_{n,m} \asymp N^{-4\alpha\nu/[4\alpha(\nu+1)+\nu]}.$$
In this high smoothness range, the optimal choice balances the fully observed term with the eigenvalue identification term, so a nonnegligible part of the budget must be spent on increasing the grid resolution.

Thus, under common design, an optimal allocation must reserve enough budget for the grid resolution itself. Increasing the number of trajectories cannot compensate for a grid that is too coarse to control the $m$-only terms. This contrasts with independent design, where the optimal rate under a fixed budget can be attained by allocating the budget primarily to the number of trajectories.

**5. Numerical Experiments.** In this section, we present simulations that illustrate the qualitative implications of the theory and a real data example that compares the practical estimator with established benchmarks. In the numerical implementation, the estimators use all training subjects for eigenvalue and coefficient estimation for better finite-sample performance.

5.1. *Simulations.* In the simulations, we use the same trigonometric coordinates as in (4). The data-generating model is truncated to $|r| \le K_0$ with $K_0 = 64$, giving a total of 129 basis functions. For a smoothness pair $(\alpha, s)$, the true covariance and slope coefficients are

$$\lambda_r = (1+|r|)^{-2\alpha}, \qquad \theta_r = c_s(1+|r|)^{-(s+1/2)}, \qquad |r| \le K_0,$$

where the constant $c_s$ is chosen so that $\sum_{|r|\le K_0}(1+|r|)^{2s}\theta_r^2 = 1$. This choice of $\theta_r$ lies on the boundary of the Sobolev ellipsoid in (10) with radius one. For each subject, we generate independent scores $\xi_{ir} \sim N(0, \lambda_r)$, set $X_i(t) = \sum_{|r|\le K_0} \xi_{ir} e_r(t)$, and generate

$$Y_i = \sum_{|r|\le K_0} \xi_{ir}\theta_r + \varepsilon_i, \qquad \varepsilon_i \sim N(0, 0.5^2).$$

The observed functional data are $Z_{ij} = X_i(t_{ij}) + \delta_{ij}$, with independent measurement errors $\delta_{ij} \sim N(0, 0.1^2)$ in the qualitative phase panels. For common design, $t_{ij} = (j-1)/m$ is a common equispaced grid for all subjects; for independent design, $t_{ij}$ are independently sampled from the uniform distribution on $[0,1]$. The reported excess prediction risk is computed against the true coefficients as $\sum_{|r|\le K_0} \lambda_r(\hat{\theta}_r - \theta_r)^2$. The implementation under common design uses the high-frequency control pilot on the full grid in (32), with fixed tuning constants across all settings.

We report the results for three smoothness pairs $(\alpha, s) = (1, 0.5), (1, 1.5), (1, 4)$ under both independent and common design in Figure 2. For independent design, the three panels display a similar qualitative pattern. The risk decreases as one moves from the lower left corner to the upper right corner, reflecting the two-term structure of the rate under independent design in (26). The first term improves with the number of trajectories, while the second improves with the total number of noisy point evaluations. Changing $s$ modifies the theoretical exponents, but the overall geometry of the heatmaps remains comparable across the three smoothness pairs.

The common-design panels show a different behavior. When $m$ is small, increasing $n$ alone produces only limited improvement, because the shared grid leaves an approximation or identification error that depends on $m$ and cannot be averaged away over subjects. The three panels also have visibly different decay patterns, in agreement with the three regimes in Subsection 4.4.1. For $(\alpha, s) = (1, 0.5)$, the dominant effect at low resolution is the term $m^{-(2\alpha+2s)}$ from the fixed grid, so the risk decreases mainly as the grid resolution $m$ increases. For $(\alpha, s) = (1, 1.5)$ and $(1, 4)$, the heatmaps exhibit a more mixed dependence on $m$ and $n$, consistent with the interaction between the sampling term, the term for identifying unknown eigenvalues, and the fully observed benchmark.

Figure 3 compares the two observation schemes on the same grid of $(m, n)$. We choose relatively large $n$ to show the asymptotic behavior more clearly. The common-design estimator remains strongly constrained by the grid when $m = 6$ or 7: increasing $n$ alone gives

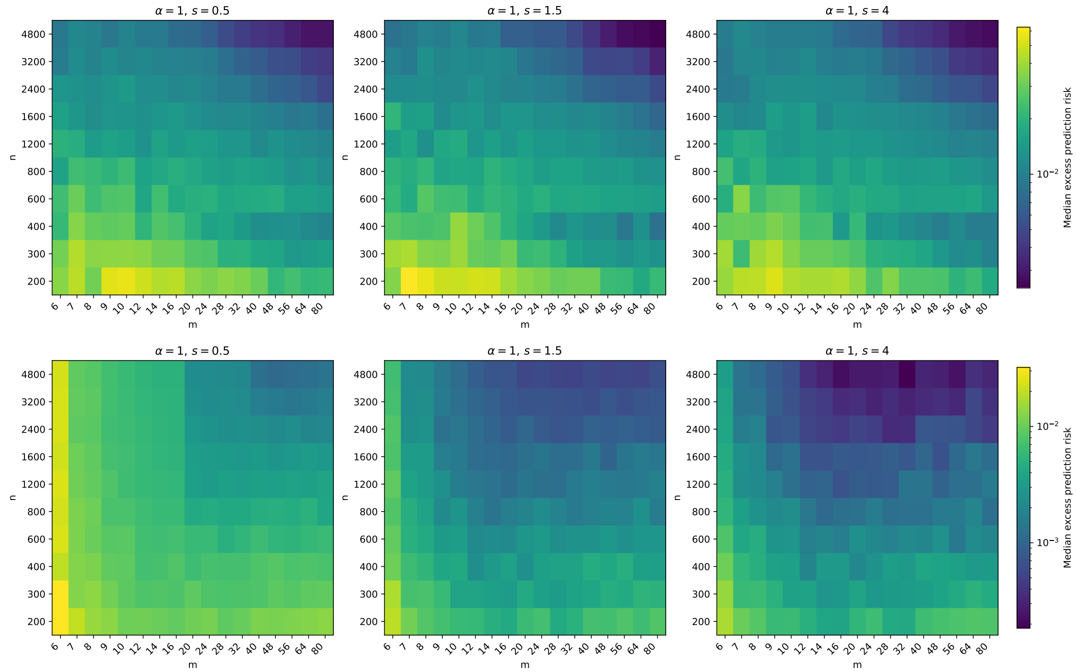


FIG 2. *Simulation heatmaps under independent design (top) and common design (bottom). Each cell plots the median excess prediction risk over* 50 *repetitions. For common design, the three panels correspond to the three regimes in Subsection 4.4.1.*

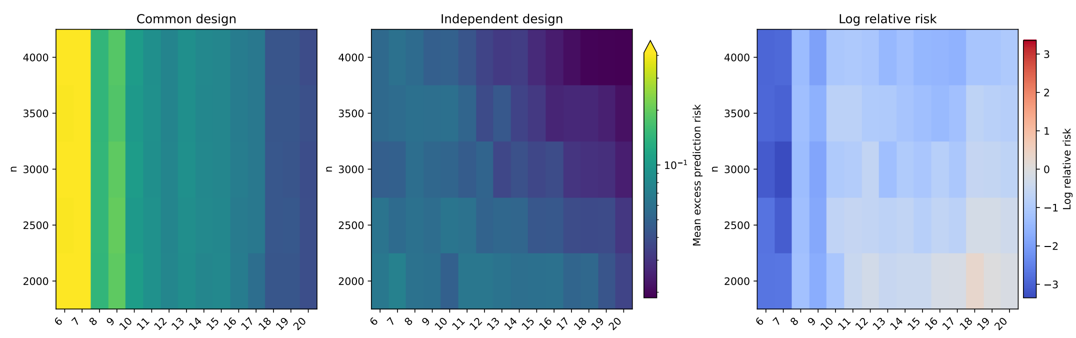


FIG 3. *Matched comparison between common and independent design for* $(\alpha, s) = (0.6, 0.1)$*. The two left panels show mean excess prediction risk, and the right panel shows* $\log_2(R_{\mathrm{IND}}/R_{\mathrm{CD}})$*, so negative values (blue) favor independent design.*

little improvement because the same low-resolution grid is shared by all subjects. As $m$ increases, the common-design risk decreases and the gap between the two designs narrows. The independent-design estimator is nevertheless smaller over most of the displayed grid, reflecting the additional fixed grid error terms under common design that cannot be averaged away across subjects.

To further investigate the rate behavior under common design, we compare the empirical rates with the theoretical prediction in more detail. To display the qualitative rate behavior more clearly and isolate the effect of biased coefficient and eigenvalue estimation, we use a version of the adaptive estimator with an oracle choice of the truncation level instead of the block thresholding procedure. We consider three smoothness pairs $(\alpha, s) =$

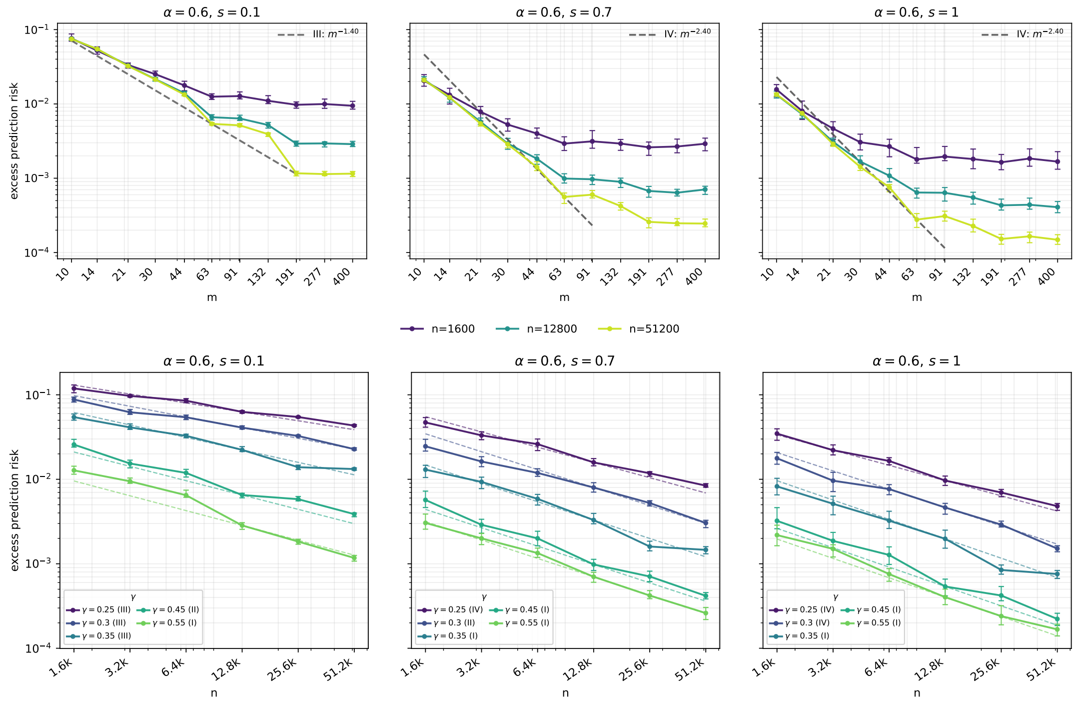


FIG 4. *Common design rate comparisons as $m$ varies (top) and along $m = n^{\gamma}$ (bottom). Each panel uses* 50 *repetitions and plots the median excess prediction risk with interquartile error bars. The dashed reference lines show the low-resolution term predicted by (35) in the top plots, and show the dominant rate in the bottom plots.*

$(0.6, 0.1), (0.6, 0.7), (0.6, 1)$, which lie in the three regimes in Subsection 4.4.1, and report the results in Figure 4. The top part of Figure 4 plots the risk against $m$ under various values of $n$. For each fixed $n$, the risk decreases as $m$ grows until the grid is fine enough that the part of the error depending on $n$ becomes dominant. After this point, the curves form a plateau, and the plateau level decreases as $n$ increases. This pattern matches the phase transition described by (35). The dashed reference lines show the low-resolution terms predicted by the theory: $m^{-(2\alpha+2s)}$ in the first panel and $m^{-4\alpha}$ in the second and third panels, and the empirical curves align closely with these theoretical guides. The bottom part of Figure 4 gives the corresponding comparison along paths $m = n^{\gamma}$. Different values of $\gamma$ correspond to different regimes in the phase diagram and hence to different dominant terms in the rate. As $\gamma$ increases, the curves move from grid-constrained behavior toward the statistical terms, and the transition occurs at different values of $\gamma$ across the three smoothness pairs. The empirical curves again align with the theoretical rates shown by the dashed reference lines, supporting the predicted rate behavior in the different regimes. Together, the two plots support the rate behavior predicted by the theory.

Figure 5 isolates the effect of estimating the eigenvalue sequence under common design. It compares the risk-oracle estimator that uses the true eigenvalues with the corresponding estimator that estimates the eigenvalues from common-grid data. The known eigenvalue estimator has smaller prediction error and reaches the dense plateau at a much smaller grid size $m$. The plateau levels are nevertheless comparable, as both estimators are then mainly governed by the dense FLR term, which does not depend on the grid resolution. This contrast is consistent with the additional unknown eigenvalue term in (35).

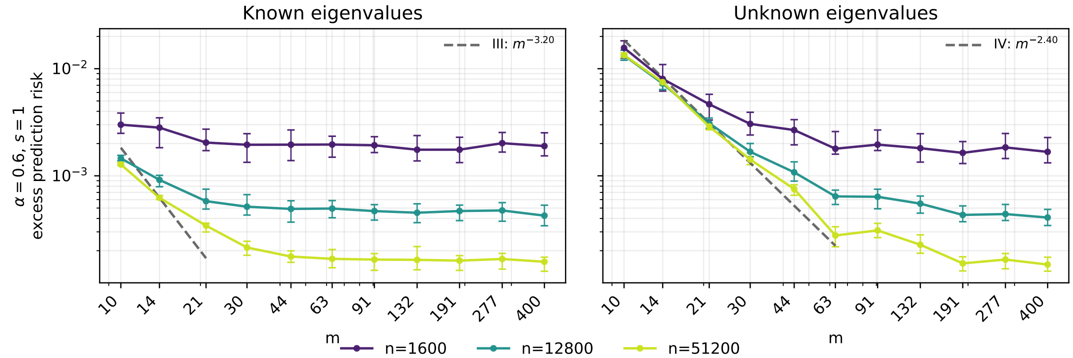


FIG 5. *Known versus estimated eigenvalues under common design for* $(\alpha, s) = (0.6, 1)$. *The two panels use the same vertical scale.*

5.2. *Real Data.* We next compare our estimators with other FLR methods on the wheat data set, a commonly used benchmark in the FLR literature. The data contain $100$ near-infrared spectra measured on a common grid of $701$ wavelengths from $1100$ nm to $2500$ nm, with wheat protein content as the scalar response. Following the experimental setup in Zhou, Yao and Zhang (2023), we use the first $80$ samples for training and the last $20$ samples for testing, and compare performance under sparse observation settings with $m = 10$ and $m = 30$ observed wavelengths per curve. For the estimator under common design, we use equispaced common subgrids with $m = 10, 30$, so no subsampling randomness is introduced. The estimator under independent design is evaluated under random subsampling within each curve with $m = 10, 30$ over $200$ repetitions. The results and the published benchmarks from Zhou, Yao and Zhang (2023) are reported in Table 2. In this evaluation, the estimator under common design has the smallest reported prediction error. While seemingly contrary to the theory, this finite-sample behavior may reflect smaller hidden constants for equispaced fixed grids than for random subsampling. The comparison should nevertheless be interpreted with the sampling protocol in mind: the common design estimator uses deterministic subgrids, whereas the independent design estimator and the published benchmarks are evaluated under random subsampling protocols.

TABLE 2

*Wheat data prediction error. CD for common design and IND for independent design are the two estimators from this paper. The benchmarks are quoted from Table 2 in Zhou, Yao and Zhang (2023): Plug-in (Hall and Horowitz, 2007); IN: approximated least square method with integrated scores; PACE (Yao, Müller and Wang, 2005b). Numbers in parentheses are standard deviations over repetitions; no standard deviation is reported for CD because the common subgrid is deterministic.*

| $m$ | CD | IND | Plug-in | IN | PACE | Zhou et al. |
|---|---|---|---|---|---|---|
| 10 | **0.294** | 0.473 (0.037) | 0.781 (0.840) | 0.499 (0.198) | 1.517 (9.112) | 0.345 (0.074) |
| 30 | **0.252** | 0.396 (0.044) | 0.513 (0.925) | 0.398 (0.183) | 0.363 (0.244) | 0.278 (0.066) |

**6. Discussion and Future Directions.** This paper establishes sharp minimax rates for functional linear regression with noisy discrete observations under two canonical sampling designs. In the independent design, each functional trajectory is observed on its own random grid; in the common design, all trajectories are observed on a shared fixed grid. These two designs represent substantially different statistical regimes. In the independent design, the randomness of the observation grids provides a form of averaging across trajectories, while in the common design the shared sampling geometry creates additional identifiability and approximation issues. By treating the two settings separately, our results clarify how the within-trajectory resolution $m$ interacts with the sample size $n$, the smoothness of the slope function, and the decay of the covariance spectrum.

A central contribution of the paper is to isolate the precise cost of discretization in functional linear regression. Both sampling designs inherit the fully observed benchmark rate, corresponding to the idealized setting in which the entire functional predictor is available without discretization error. In addition, both designs contain a term reflecting the statistical cost of recovering the relevant cross-covariance information from noisy point evaluations. This term captures the effective loss of information caused by observing only $m$ noisy measurements per trajectory rather than the full curve. Under the common design, however, two further effects appear. One is the deterministic approximation error induced by the fixed grid; the other is the cost of identifying or reliably estimating the unknown covariance eigenvalues on the grid. These additional terms are absent, or substantially attenuated, under the independent design because independent random grids provide more diversified sampling information across subjects.

Although the paper focuses on excess prediction risk, the same arguments extend naturally to the $L^2$ estimation risk

$$\mathbb{E}\|\hat{\beta} - \beta\|_{L^2}^2.$$

In Fourier coordinates, prediction risk weights the squared coordinate error by $\lambda_r$, while $L^2$ risk removes this weight. Thus high-frequency components become more difficult to estimate under $L^2$ loss. For $s > 0$, the independent-design estimation rate becomes

$$n^{-\frac{2s}{2\alpha+2s+1}} + (nm)^{-\frac{2s}{4\alpha+2s+1}}.$$

Under the common design with unknown eigenvalues, the corresponding rate is

$$n^{-\frac{2s}{2\alpha+2s+1}} + (nm)^{-\frac{2s}{4\alpha+2s+1}} + m^{-2s} + m^{-4\alpha}.$$

At the proof level, this change amounts to normalizing the cross-covariance estimation error by $\lambda_r^2$ rather than $\lambda_r$, and replacing the prediction-weighted block loss in the van Trees lower bounds by the unweighted block loss. The adaptive procedure also carries over: the eigenvalue screening step is unchanged, while the second thresholding step is based on the empirical $L^2$ block energy

$$\sum_{r \in B_\ell} \frac{|\hat{\gamma}_r|^2}{\check{\lambda}_r^2},$$

with the corresponding variance proxy adjusted accordingly.

Methodologically, the adaptive estimator is not specific to this FLR model. The main difficulty is that the loss depends jointly on the smoothness of the slope function and on the unknown covariance spectrum. The estimator must therefore identify frequency blocks containing estimable signal while also determining whether the covariance scale on those blocks can be stably inverted. This motivates the two-stage thresholding rule: an eigenvalue threshold first removes unstable blocks, and an empirical energy threshold then selects blocks whose signal exceeds the relevant noise level. This principle may be useful in other inverse or semiparametric problems where the effective loss is governed by an unknown nuisance operator, metric, or covariance structure.

The lower-bound arguments also provide tools beyond the present setting. Under the independent design, the discretization cost is characterized through Fisher information bounds for noisy point evaluations, showing that the loss from observing only finitely many measurements per trajectory is intrinsic rather than estimator-specific. Under the common design, the lower bounds exploit unidentifiability created by the interaction between the covariance function, the slope function, and the shared sampling grid. These mechanisms may be relevant for other functional problems in which discrete sampling and inverse structure interact.

Several limitations and open questions remain. The analysis considers two stylized acquisition schemes: independent random grids and a regular common grid. Extending sharp minimax theory to heterogeneous protocols, such as irregular, nonuniform, or partially deterministic grids, would broaden the scope of the results. Another direction is to relax the fixed trigonometric basis formulation. If the covariance eigenbasis must be estimated from discretely observed data, or if the smoothness scale of $\beta$ is not aligned with the covariance geometry, additional costs may appear. A complete theory would need to combine the present analysis of the dependence on $m$ with covariance eigenbasis recovery (Cai and Yuan, 2010) and the alignment effects between covariance geometry and slope regularity studied in fully observed FLR. Finally, it would be interesting to understand how these discretization phenomena affect downstream goals such as inference, testing, distributed learning, and privacy-constrained functional prediction.

## SUPPLEMENTARY MATERIAL

**Supplement to "The Cost of Discretization in Functional Linear Regression: Minimax Rates and Adaptation"**
This supplement contains additional numerical experiments, proofs, and auxiliary results omitted from the main text.

## REFERENCES


BALASUBRAMANIAN, K., MÜLLER, H.-G. and SRIPERUMBUDUR, B. K. (2025). Functional Linear and Single-Index Models: A Unified Approach via Gaussian Stein Identity. *Bernoulli* **31** 973–1006. https://doi.org/10.3150/24-BEJ1755

CAI, T. T. and HALL, P. (2006). Prediction in Functional Linear Regression. *The Annals of Statistics* **34** 2159–2179. https://doi.org/10.1214/009053606000000830

CAI, T. T. and LI, Y. (2026). Supplement to "The Cost of Discretization in Functional Linear Regression: Minimax Rates and Adaptation".

CAI, T. T. and YUAN, M. (2010). Nonparametric Covariance Function Estimation for Functional and Longitudinal Data Technical report, University of Pennsylvania and Georgia Institute of Technology.

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. https://doi.org/10.1214/11-AOS898

CAI, T. T. and YUAN, M. (2012). Minimax and Adaptive Prediction for Functional Linear Regression. *Journal of the American Statistical Association* **107** 1201–1216. https://doi.org/10.1080/01621459.2012.716337

CAI, T. T., ZHANG, L. and ZHOU, H. H. (2018). Adaptive Functional Linear Regression via Functional Principal Component Analysis and Block Thresholding. *Statistica Sinica* **28** 2455–2468. https://doi.org/10.5705/ss.202017.0099

CARDOT, H., CRAMBES, C., KNEIP, A. and SARDA, P. (2007). Smoothing Splines Estimators in Functional Linear Regression with Errors-in-Variables. *Computational Statistics & Data Analysis* **51** 4832–4848. https://doi.org/10.1016/j.csda.2006.07.029

FAN, J., GUO, Z.-C. and SHI, L. (2024). Spectral Algorithms for Functional Linear Regression. *Communications on Pure and Applied Analysis* **23** 895–915. https://doi.org/10.3934/cpaa.2024039

FERRATY, F., GONZALEZ-MANTEIGA, W., MARTINEZ-CALVO, A. and VIEU, P. (2012). Presmoothing in Functional Linear Regression. *Statistica Sinica* **22** 69–94. https://doi.org/10.5705/ss.2010.085

GAJARDO, Á., DAI, X. and MÜLLER, H.-G. (2021). Predictive Distributions and the Transition from Sparse to Dense Functional Data. https://doi.org/10.48550/arXiv.2109.02236

GUO, S., LI, D., QIAO, X. and WANG, Y. (2025). From Sparse to Dense Functional Data in High Dimensions: Revisiting Phase Transitions from a Non-Asymptotic Perspective. *Journal of Machine Learning Research* **26** 1–40.

GUPTA, N., SIVANANTHAN, S. and SRIPERUMBUDUR, B. K. (2025). Optimal Rates for Functional Linear Regression with General Regularization. *Applied and Computational Harmonic Analysis* **76** 101745. https://doi.org/10.1016/j.acha.2024.101745

HALL, P. and HOROWITZ, J. L. (2007). Methodology and Convergence Rates for Functional Linear Regression. *The Annals of Statistics* **35** 70–91. https://doi.org/10.1214/009053606000000957

HALL, P., MÜLLER, H.-G. and WANG, J.-L. (2006). Properties of Principal Component Methods for Functional and Longitudinal Data Analysis. *The Annals of Statistics* **34** 1493–1517. https://doi.org/10.1214/009053606000000272

HSING, T. and EUBANK, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. *Wiley Series in Probability and Statistics*. John Wiley & Sons, Chichester. https://doi.org/10.1002/9781118762547

LI, Y. and HSING, T. (2007). On Rates of Convergence in Functional Linear Regression. *Journal of Multivariate Analysis* **98** 1782–1804. https://doi.org/10.1016/j.jmva.2006.10.004

MEISTER, A. (2011). Asymptotic Equivalence of Functional Linear Regression and a White Noise Inverse Problem. *The Annals of Statistics* **39** 1471–1495. https://doi.org/10.1214/10-AOS872

MÜLLER, H.-G. and STADTMÜLLER, U. (2005). Generalized Functional Linear Models. *The Annals of Statistics* **33** 774–805. https://doi.org/10.1214/009053604000001156

PETERSEN, A. (2024). 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. https://doi.org/10.1016/j.jmva.2024.105355

RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2 ed. *Springer Series in Statistics*. Springer, New York.

WANG, J.-L., CHIOU, J.-M. and MÜLLER, H.-G. (2016). Functional Data Analysis. *Annual Review of Statistics and Its Application* **3** 257–295. https://doi.org/10.1146/annurev-statistics-041715-033624

YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005a). Functional Data Analysis for Sparse Longitudinal Data. *Journal of the American Statistical Association* **100** 577–590. https://doi.org/10.1198/016214504000001745

YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005b). Functional Linear Regression Analysis for Longitudinal Data. *The Annals of Statistics* **33** 2873–2903. https://doi.org/10.1214/009053605000000660

YUAN, M. and CAI, T. T. (2010). A Reproducing Kernel Hilbert Space Approach to Functional Linear Regression. *The Annals of Statistics* **38** 3412–3444. https://doi.org/10.1214/09-AOS772

ZHANG, X. and WANG, J.-L. (2016). From Sparse to Dense Functional Data and Beyond. *The Annals of Statistics* **44** 2281–2321. https://doi.org/10.1214/16-AOS1446

ZHOU, H., YAO, F. and ZHANG, H. (2023). Functional Linear Regression for Discretely Observed Data: From Ideal to Reality. *Biometrika* **110** 381–393. https://doi.org/10.1093/biomet/asac053