

Principal components analysis for sparsely observed correlated functional data using a kernel smoothing approach

Debashis Paul and Jie Peng

Department of Statistics

University of California

Davis, CA 95616

e-mail: debpaul@ucdavis.edu; jiepeng@ucdavis.edu

Abstract: We consider the problem of functional principal component analysis for correlated functional data. In particular, we focus on a separable covariance structure and consider irregularly and possibly sparsely observed sample trajectories. By observing that under the sparse measurements setting, the empirical covariance of pre-smoothed sample trajectories is a highly biased estimator along the diagonal, we propose to modify the empirical covariance by estimating the diagonal and off-diagonal parts of the covariance kernel separately. We prove that under a separable covariance structure, this method can consistently estimate the eigenfunctions of the covariance kernel. We also quantify the role of the correlation in the L^2 risk of the estimator, and show that under a weak correlation regime, the risk achieves the optimal nonparametric rate when the number of measurements per curve is bounded.

AMS 2000 subject classifications: Primary 62G20; secondary 62H25.

Keywords and phrases: Functional data analysis, principal component analysis, kernel smoothing, consistency.

Received July 2010.

1. Introduction

For various scientific studies, the data consist of measurements corresponding to one subject or experimental unit being recorded in time or space, which can often be viewed as functional data (e.g., [10, 26]). If the goal of the analysis is either data compression, model building or studying covariate effects, one popular approach is to use the *functional principal components* (i.e., the eigenvalues and eigenfunctions of the covariance kernel), for example, in functional linear regression [4, 5, 12, 31–33], functional discriminant analysis [16] and functional clustering [8, 17, 25, 29]. Common approaches to functional principal components analysis (henceforth, FPCA) include local polynomial smoothing of empirical covariances [13, 31, 32], and representing the eigenfunctions in a known basis of smooth functions [3, 6, 15, 24, 27].

FPCA has been studied primarily under the assumption that the observed trajectories are independent realizations from an underlying stochastic process. In particular, almost all theoretical developments for FPCA make this assumption, e.g., [6, 13, 31, 35] and [23]. However, in practice, FPCA is often applied to correlated functional data. For instance, [26] describe an application of FPCA to capture variations in the yearly temperature cycle across several cities in Canada. Since one expects spatial correlation in the weather pattern, the observed trajectories are probably correlated. Another scenario for correlated functional data is when the different subjects are sampled from a population divided into various strata, for example in genetic studies where the strata could correspond to different racial groups or families [36]. Correlated functional data also arise in time-course gene expression studies since it is expected that the expression profiles of genes involved in the same biological processes are correlated [28]. Such data have been analyzed by FPCA, e.g., in [30]. Atmospheric studies also often produce correlated functional data. For example, in a study of atmospheric radiations [14], vertical profiles of atmospheric radiation are recorded at different times. It is reasonable to assume that, after removing a possible long term time-trend, the radiation intensity is correlated across time as well as altitude. These examples illustrate the need for studying the effects of correlation across the sample trajectories on the estimates of the functional principal components. In this paper, we study the asymptotic behavior of FPCA under the “separable covariance structure” described below which is often used to model spatio-temporal processes. Specifically, the theoretical developments are carried out for a method proposed in Section 2 of this paper, which is based on the empirical covariance of pre-smoothed sample trajectories, with suitable modifications along the diagonal. We conjecture that similar asymptotic results hold for other FPCA procedures for dealing with possibly irregularly observed functional data, e.g., the PACE method in [31, 32].

We first give a brief overview of the functional data model and the separable correlation structure. Suppose that we observe n realizations of an L^2 -stochastic process $\{X(t) : t \in [0, 1]\}$ at a sequence of points on the interval $[0, 1]$, with additive measurement noise. That is, the observed data $\{Y_{ij} : 1 \leq j \leq m_i; 1 \leq i \leq n\}$ can be modeled as:

$$Y_{ij} = X_i(T_{ij}) + \sigma\varepsilon_{ij}, \quad (1.1)$$

where $\{\varepsilon_{ij}\}$ are i.i.d. with mean 0 and variance 1. Since $X(t)$ is an L^2 stochastic process, by *Mercer’s Theorem* [1] there exists a positive semi-definite kernel $C(\cdot, \cdot)$ such that $Cov(X(s), X(t)) = C(s, t)$ and each $X_i(t)$ has the following a.s. representation in terms of the eigenfunctions of the kernel $C(\cdot, \cdot)$:

$$X_i(t) = \mu(t) + \sum_{\nu=1}^{\infty} \sqrt{\lambda_{\nu}} \psi_{\nu}(t) \xi_{i\nu}, \quad (1.2)$$

where $\mu(\cdot) = \mathbb{E}(X(\cdot))$ is the mean function; $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ are the eigenvalues of $C(\cdot, \cdot)$; $\psi_{\nu}(\cdot)$ are the corresponding orthonormal eigenfunctions, i.e.,

$C(s, t) = \sum_{\nu} \lambda_{\nu} \psi_{\nu}(s) \psi_{\nu}(t)$; and the random variables $\{\xi_{i\nu} : \nu \geq 1\}$, for each i , are uncorrelated with zero mean and unit variance. Furthermore, we assume that for each pair (i, j) with $1 \leq i, j \leq n$,

$$\mathbb{E}(\xi_{i\nu} \xi_{j\nu'}) = \delta_{\nu\nu'} \rho_{ij}, \quad (1.3)$$

for $\nu, \nu' \geq 1$, and $-1 \leq \rho_{ij} \leq 1$, with $\rho_{ii} \equiv 1$, where $\delta_{\nu\nu'} = 1$ if $\nu = \nu'$ and 0 if $\nu \neq \nu'$. (1.3) describes a *separable covariance structure* for the noiseless processes. That is, the processes $\{X_i(\cdot)\}_{i=1}^n$ satisfy, $Cov(X_i(s), X_j(t)) = \rho_{ij} C(s, t)$.

A separable covariance structure assumes that the correlation across the trajectories and across time are independent of each other. An immediate extension is the coregionalization model for modeling multivariate spatial processes [2, 11, 34]. Here, the principal component scores $\{\xi_{i\nu}\}$ are modeled as mean zero random variables with

$$\mathbb{E}(\xi_{i\nu}, \xi_{j\nu'}) = \rho_{ij}^{(\nu\nu')}, \quad (1.4)$$

where the $n \times n$ matrices $\mathbf{R}_{\nu\nu'} = ((\rho_{ij}^{(\nu\nu')}))$ satisfy certain asymptotic identifiability conditions. Note that, the case $\mathbf{R}_{\nu\nu'} = \delta_{\nu\nu'} \mathbf{R}$ corresponds to the separable covariance model described in (1.3). One example of this type of processes is the weed growth data studied in [2], where the trajectories are weed growth profiles for different locations in the agricultural field.

We now give a brief description of the estimation procedure proposed in this paper. Our method is motivated by the setting where the observation times for each curve are sparse and irregular, even though it can be applied when the observations are dense and/or on a regular grid. We first pre-smooth each sample curve by a kernel smoother and then obtain the “empirical covariance kernel” of the pre-smoothed curves. However, this naive estimator gives a highly biased estimate along the diagonal of the covariance kernel when the number of measurements per curve is small (see Proposition 2.1). We mitigate this problem by estimating the diagonal and the off-diagonal parts of the covariance kernel separately, and then merging them together using a smooth weight kernel. Moreover, we use a linearized kernel smoothing which helps in reducing bias. As already indicated, the pre-smoothed sample trajectories are poor estimates for the true trajectories when the measurements are sparse, and so the motivation of pre-smoothing is not to get good estimates of the trajectories. Rather, pre-smoothing should be viewed as a mathematical step in the development of the estimator of the covariance kernel. This allows us to combine the data in a way that is convenient from both computational and analytical points of view. This pooling of information across trajectories in the proposed procedure leads to a consistent estimator of the covariance kernel and its eigenfunctions. The use of the term “pre-smoothing” or “pre-smoothed sample trajectories” throughout this paper is for the ease of reference.

Now we summarize the main results of this paper. We obtain explicit expressions for the integrated mean squared error of the covariance kernel and its estimated eigenfunctions under the separable covariance model (Theorems 3.1 and 3.2). The quantification of the role of correlation in the risk behavior for the sep-

arable covariance model (Theorem 3.2) and the coregionalization model (Theorem 3.4) is seemingly new in the literature. We also derive a lower bound on the rate of convergence of the first eigenfunction (Theorem 3.3). This lower bound and the matching upper bound on the rate of convergence for the i.i.d. case shows that the proposed estimator attains the optimal nonparametric rate when $\max_{1 \leq i \leq n} m_i$ is either bounded or increases to infinity slowly with n . This result parallels a similar result derived in [13] for the local polynomial estimator proposed in [31]. Moreover, we show that if the correlation between sample curves is “weak” in a suitable sense, then the optimal rate of convergence in the correlated and i.i.d. cases are the same. The latter is a principal finding of this paper, and justifies the use of FPCA even when the sample curves are “weakly” correlated.

The rest of the paper is organized as follows. In Section 2, we propose the estimation procedure. In Section 3, we state the main results about the consistency and rate of convergence of the estimators of the covariance kernel and its eigenfunctions. In Section 4, we give an outline of the proof of the main results and discuss their implications. In Section 5, we conduct a simulation study to illustrate the finite sample performance of the proposed estimation procedure and the effect of correlation. We also compare this method with the PACE method proposed in [31]. In Section 6, we apply the proposed method to a time-course gene expression data set. In Section 7, we conclude the paper by a summary. Technical details are provided in the Appendix.

2. Method

Throughout this paper, we assume for simplicity that the mean curve is known and has been subtracted from the data. When the mean $\mu(\cdot)$ is estimated by averaging the pre-smoothed sample curves, it can be shown that, subtracting the estimated mean $\hat{\mu}(\cdot)$ from individual pre-smoothed sample curves, results in no change in the rate of convergence of the estimated eigenfunctions by the proposed procedure. Moreover, for theoretical analysis, we restrict ourselves to the setting where in the expansion (1.2) the number of nonzero eigenvalues λ_ν , is finite, i.e., $\lambda_M > 0 = \lambda_{M+1}$ for some $M \geq 1$ (for a brief discussion for the case of infinite number of nonzero eigenvalues, see Section 3.6). We also assume that in the observed data model (1.1), the vector of observation times $\mathbf{T}_i = \{T_{ij} : j = 1, \dots, m_i\}$ are randomly sampled from a continuous distribution with a density g supported on $[0, 1]$. We further assume that there are constants $0 < c_0 \leq c_1 < \infty$ such that $c_0 \leq g(\cdot) \leq c_1$. Throughout the paper, we assume that g is known, even though it can be replaced by a kernel density estimator. It can be shown that under appropriate regularity conditions on g , and under an optimal choice of bandwidth, the results on the rates of convergence of the estimators continue to hold (see Remark 3.1 for more discussions).

2.1. Empirical covariance estimator

A popular method in nonparametric function estimation is kernel smoothing. In principle, one can adopt a similar idea for FPCA by first smoothing individual

sample curves, and then computing the covariance of the “pre-smoothed” sample curves, followed by an eigen-analysis of this “empirical covariance kernel”. In the following, we first describe briefly such an approach, and then show that even in the case of i.i.d. and noiseless (i.e., $\sigma^2 = 0$) realizations, the estimator thus obtained has an intrinsic bias while estimating the diagonal of the covariance kernel, unless the number of measurements per curve is large.

Let $K(x)$ be a kernel with an adequate degree of smoothness, and satisfying the following conditions:

- B1** (i) $\text{supp}(K) = [-B_K, B_K]$ for some $B_K > 0$; (ii) K is symmetric about 0; (iii) $\int K(x)dx = 1$; (iv) $\int xK(x)dx = 0$; (v) $\int K'(x)dx = 0$; (vi) $\int xK'(x)dx = 1$.

We then define the presmoothed sample curves as follows:

$$\tilde{X}_i(t) = \frac{1}{m_i} \sum_{j=1}^{m_i} Y_{ij} K_{h_{n,i}}(t - T_{ij}), \quad i = 1, \dots, n, \quad (2.1)$$

where $K_h(x) := h^{-1}K(h^{-1}x)$ for $h > 0$ and $h_{n,i}$ is the bandwidth for the i -th curve. Then the empirical covariance based on the presmoothed curves is simply

$$\tilde{C}(s, t) = \frac{1}{g(s)g(t)} \frac{1}{n} \sum_{i=1}^n \tilde{X}_i(t) \tilde{X}_i(s). \quad (2.2)$$

In the following, we first derive the expectation of $\tilde{C}(s, t)$ in order to quantify the bias, when $h_{n,i} = h_n$ for all i , under the assumption that $C(\cdot, \cdot)$ is twice continuously differentiable. For simplicity of exposition, in the following proposition we assume that the density of the design points $\{T_{ij}\}_{j=1}^{m_i}$, for each subject, is uniform on $[0, 1]$, i.e., $g(t) \equiv 1$. Define $\bar{C}(t) = C(t, t)$ for $t \in [0, 1]$, and $K^{(2)}(x) := \int K(x-u)K(-u)du$.

Proposition 2.1. *Suppose that $h_n \rightarrow 0$ as $n \rightarrow \infty$. When $s \neq t$,*

$$\begin{aligned} & \mathbb{E}[\tilde{X}_i(s)\tilde{X}_i(t)] \\ &= \frac{1}{m_i h_n} K^{(2)}\left(\frac{s-t}{h_n}\right) (\bar{C}(t) + \sigma^2) + \frac{1}{m_i} \bar{C}'(t) \int u K(-u) K\left(\frac{s-t}{h_n} - u\right) du \\ & \quad + \left(1 - \frac{1}{m_i}\right) C(s, t) + \frac{1}{m_i} O(h_n) + O(h_n^2). \end{aligned} \quad (2.3)$$

And, if $s = t$, then

$$\mathbb{E}[\tilde{X}_i(t)^2] = \frac{1}{m_i h_n} K^{(2)}(0) (\bar{C}(t) + \sigma^2) + \left(1 - \frac{1}{m_i}\right) \bar{C}(t) + \frac{1}{m_i} O(h_n) + O(h_n^2). \quad (2.4)$$

The $O(\cdot)$ terms involve $\sup_{t \in [0, 1]} |\bar{C}''(t)|$, $\sup_{s, t \in [0, 1]} \|\mathcal{D}^2 C(s, t)\|$ and $\int u^2 K(u) du$, where \mathcal{D}^2 is the Hessian operator.

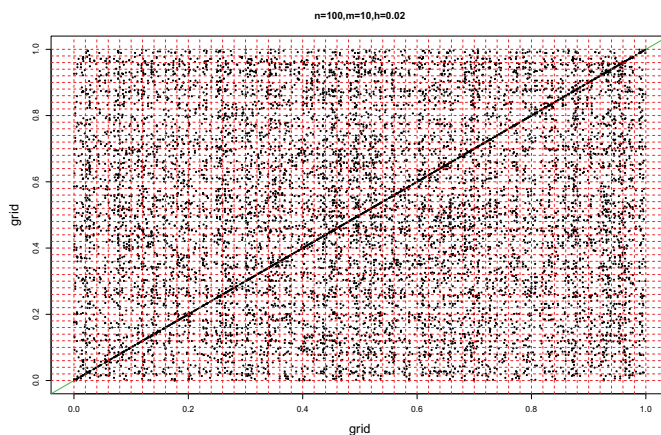


FIG 1. Points $\{T_{ij} : j = 1, \dots, 10\}_{i=1}^{100}$ are randomly drawn from the interval $[0, 1]$ and the pairs $(T_{ij}, T_{ij'})$ for $1 \leq j, j' \leq 10$ are plotted for each i . The grid has spacing $h = 0.02$.

By Proposition 2.1, it is easy to see that $\mathbb{E}[\tilde{X}_i(s)\tilde{X}_i(t)] = (1 - m_i^{-1})C(s, t) + O(h_n^2)$ if $|s - t| > 2B_K h_n$, since then the first two terms in (2.3) as well as the $O(h_n)$ term all vanish (see the proof in Appendix B for more details). This shows that $\tilde{C}(s, t)$ should be multiplied by $m_i/(m_i - 1)$ to get rid of the trivial bias. However, (2.3) and (2.4) also show that $\tilde{C}(s, t)$ is a highly biased estimate of $C(s, t)$ near the diagonal even after this modification, unless $h_n \underline{m}_n \rightarrow \infty$ where $\underline{m}_n := \min_{1 \leq i \leq n} m_i$. This is because the first terms in (2.3) and (2.4) are always positive along the diagonal (i.e., when $|s - t| < 2B_K h_n$), which result in an overestimation. Note that, this is the case even if $\sigma^2 = 0$, i.e., the observations are “noiseless”. In fact the degree of overestimation gets big by a scale factor of h_n as soon as $|s - t| < 2B_K h_n$. This observation motivates us to modify the empirical covariance by estimating the diagonal and off diagonal separately.

To help understand the reason for this bias, let us consider the case when m_i 's are bounded and T_{ij} 's are sampled independently from $\text{Uniform}[0, 1]$. The number of pairs of the form $\{(T_{ij}, T_{ij'}) : 1 \leq j, j' \leq m_i; i = 1, \dots, n\}$ falling into a given off-diagonal square of side-length h_n is $NO(h_n^2)$, where N is the total number of such pairs. In contrast, the number of such pairs falling into a square of side-length h_n along the diagonal is $NO(h_n)$. Therefore, in terms of covariance estimation, the measurements are much denser along the diagonal and this explains the difference in rates shown by Proposition 2.1. Figure 1 gives a visual demonstration of this point.

2.2. Modification to empirical covariance kernel

In this section, we propose a modification to deal with the bias in empirical covariance kernel described in Section 2.1. We propose to remedy the effect of

unequal scale along the diagonal of the covariance kernel (and the resulting bias) by estimating the diagonal and the off-diagonal parts separately. We then use a suitable (smooth) weight kernel to combine those two estimates together.

We also propose to use a linearized version of the kernel smoothing to reduce the bias while controlling the variance. For this purpose, define $\overline{Q}(s, t)$ to be a *tensor-product kernel* (that is a kernel of the form $Q(s, t) = \overline{Q}(s)\overline{Q}(t)$ for some smooth function \overline{Q}) with the following properties, together referred to as condition **B2**:

- (i) \overline{Q} is supported on $[-C_Q, C_Q]$, for some $C_Q > 0$, and $\overline{Q}(\cdot) \geq 0$;
- (ii) $\|\overline{Q}\|_\infty < \infty$, where $\|\overline{Q}\|_\infty := \sup_s |\overline{Q}(s)|$;
- (iii) $\sum_{k \in \mathbb{Z}} \overline{Q}(x - k) = 1$.
- (iv) \overline{Q} is symmetric about 0.

Property (iii) can be rephrased as saying that integer translates of \overline{Q} form a *partition of unity*. As an example, the B-spline basis functions [7] satisfy all four properties. Let $Q_h(s, t)$ denote the kernel $Q(h^{-1}s, h^{-1}t)$.

Recall that $\overline{C}(t) = C(t, t)$. Observe that, the conditional expectation of Y_{ij}^2 given \mathbf{T}_i is $C(T_{ij}, T_{ij}) + \sigma^2$. Thus, $\overline{C}(t) + \sigma^2$ can be estimated by a linearized kernel smoothing of the terms $\{m_i^{-1}Y_{ij}^2 : j = 1, \dots, m_i; i = 1, \dots, n\}$. Define a grid on $[0, 1]$ with spacings h_n and denote the grid points by $\{s_l : l = 1, \dots, L_n\}$ where $L_n = c_L/h_n$ for an appropriately chosen $c_L \approx 1$. Then define,

$$\widehat{C}_{*,h_n}(t) = \frac{1}{g(t)} \frac{1}{n} \sum_{i=1}^n \sum_{l=1}^{L_n} [S_i(s_l) + (t - s_l)S'_i(s_l)] \overline{Q}_{h_n}(t - s_l), \tag{2.5}$$

where $\overline{Q}_h(s) = \overline{Q}(h^{-1}s)$, with

$$S_i(s) := \frac{1}{m_i} \sum_{j=1}^{m_i} Y_{ij}^2 K_{h_n}(s - T_{ij}). \tag{2.6}$$

Finally the diagonal $\overline{C}(t)$ can be estimated by $\widehat{\overline{C}}(t) := \widehat{C}_*(t) - \widehat{\sigma}^2$, where $\widehat{\sigma}^2$ is an estimator of σ^2 (discussed in Section 2.3). Note that, (2.5) is a linearized version of the conventional kernel smoothing, which can be interpreted as a local linear smoothing of the empirical variances. A similar principle is applied to construct an estimator for the off-diagonal part (see (2.7) below). As mentioned earlier, the linearized estimator helps to reduce bias in the estimate as compared to the usual kernel smoother.

Let $\widetilde{X}_i(t)$ be the i -th smoothed sample curve as defined in (2.1), and $\widetilde{X}'_i(t)$ be the derivative of $\widetilde{X}_i(t)$. Then define the estimate of the off-diagonal part as (with a slight abuse of notation)

$$\begin{aligned} \widetilde{C}_{h_n}(s, t) &= \frac{1}{g(s)g(t)} \frac{1}{n} \sum_{i=1}^n w(m_i) \sum_{l,l'=1}^{L_n} \left[(\widetilde{X}_i(s_l) + (s - s_l)\widetilde{X}'_i(s_l)) \right. \\ &\quad \left. \cdot (\widetilde{X}_i(s_{l'}) + (t - s_{l'})\widetilde{X}'_i(s_{l'})) Q_{h_n}(s - s_l, t - s_{l'}) \right]. \end{aligned} \tag{2.7}$$

Here $w(m_i) = m_i / (m_i - 1)$ are weights determined through an asymptotic bias analysis (Proposition 2.1). Note that, as long as $|s - t| \geq Ah_n$ for some constant A depending on B_K and C_Q , in the inner sum in definition (2.7), the terms for which $l = l'$ are absent. Therefore, according to our analysis in the previous section, they do not contribute anything by way of bias.

Now let $W(\cdot, \cdot)$ be a kernel on the domain $[0, 1]^2$, defined as

$$W(s, t) := W(s - t) = \begin{cases} 0 & \text{if } |s - t| > \frac{1}{2} \\ 1 & \text{if } |s - t| \leq \frac{1}{2} \end{cases} \quad (2.8)$$

Define $W_{\tilde{h}_n}(s, t) = W((s - t) / \tilde{h}_n)$ and $\overline{W}_{\tilde{h}_n}(s, t) = 1 - W_{\tilde{h}_n}(s, t)$, where $\tilde{h}_n = Ah_n$ for the constant A mentioned in the previous paragraph. We then smooth the kernels $W_{\tilde{h}_n}$ and $\overline{W}_{\tilde{h}_n}$ by convolving them with a Gaussian kernel with a small bandwidth $\tau_n = o(h_n)$, and denote the resulting kernels still by $W_{\tilde{h}_n}$ and $\overline{W}_{\tilde{h}_n}$, respectively. Finally, the proposed estimator of $C(s, t)$ is defined as as

$$\hat{C}_{h_n}(s, t) = \overline{W}_{\tilde{h}_n}(s, t) \tilde{C}_{h_n}(s, t) + W_{\tilde{h}_n}(s, t) \max \left\{ \hat{C}_{h_n} \left(\frac{s+t}{2}, h_n^2 \right) \right\}. \quad (2.9)$$

The use of maximum in the second term guarantees that the estimator of the diagonal is positive and the bias is of the order $O(h_n^2)$.

One important feature of the proposed estimator is that, in (2.5) and (2.7), the function g is explicitly utilized to adjust for the sampling variability, whereas in local linear smoothing, e.g. in [31], g is implicitly utilized, and is essentially estimated by using the same bandwidth for estimating the covariance. In practice, we replace g in (2.5) and (2.7) by an estimator \hat{g} using an optimal bandwidth, which is different from the bandwidth used to estimate the covariance kernel. This is likely to result in a slightly less variable estimator of the covariance kernel especially when the measurements per curve are sparse. Similarly as in (2.7), where the sample trajectories are pre-smoothed and the density of the observation times are estimated separately, such an approach may be adopted, as an alternative to the local polynomial regression method, to estimate moments of the processes other than the mean and the covariance. For example, this can be used in functional linear regression.

2.3. Estimation of σ^2

Here we briefly outline a method for estimating the error variance σ^2 . The method is similar to the approach taken in [32], and hence we omit the details.

First, for a given bandwidth h_n , we estimate the function $C(s, t)$ for $|s - t| > Ah_n$, for some A depending on B_K and C_K , using (2.7). Then, as in [32], we estimate the diagonal $\{\overline{C}(t) : t \in [0, 1]\}$, using an *oblique* linear interpolation, by

$$\hat{\overline{C}}_{0, h_n}(t) = \int_{A_1}^{A_2} \frac{1}{2} \left(\tilde{C}_{h_n}(t - uh_n, t + uh_n) + \tilde{C}_{h_n}(t + uh_n, t - uh_n) \right) d\tilde{G}(u), \quad (2.10)$$

for some probability distribution function \tilde{G} supported on $[A_1, A_2]$ where $A_1 > A$. On the other hand, we estimate the curve $\{\bar{C}(t) + \sigma^2 : t \in [0, 1]\}$ by $\hat{C}_{*,h_n}(t)$ defined in (2.5). Now, we estimate σ^2 by

$$\hat{\sigma}^2 = \frac{1}{T_1 - T_0} \int_{T_0}^{T_1} \left(\hat{C}_{*,h_n}(t) - \hat{C}_{0,h_n}(t) \right) dt, \quad (2.11)$$

where $0 < T_0 < T_1 < 1$. It can be shown that (Corollary 3.1 in Section 3) the estimator $\hat{\sigma}^2$ thus obtained is consistent for an appropriate choice of h_n .

3. Asymptotic properties

In this section, we present the asymptotic properties of the proposed estimator. Our main interest is in the estimation accuracy of the covariance kernel and its eigenfunctions. The statements of the results and the associated regularity conditions are given below. In Sections 3.1 to 3.3, we assume that the data follow the separable covariance model (1.3). In Section 3.4, we discuss the connection between the “dense” design and the purely functional data case. In Section 3.5, we present results under co-regionalization model for the covariance. Finally, in Section 3.6, we briefly discuss the situation for infinite dimensional processes. Throughout, we assume that the data follow the model (1.1) where $X_i(\cdot)$ are Gaussian processes with mean 0 and the noise ε_{ij} are also Gaussian.

3.1. Asymptotics under the separable covariance model

We first state the following assumptions on g , the density of the design points; C , the covariance kernel; and $\{\psi_k\}_{k=1}^M$, the eigenfunctions.

- A1** g is twice continuously differentiable and the second derivative is Hölder(α), for some $\alpha \in (0, 1)$. Also, the same holds for the covariance kernel C .
- A2** There are constants $0 < c_0 \leq c_1 < \infty$ such that $c_0 \leq g(\cdot) \leq c_1$.
- A3** $\max_k \{\|\psi_k\|_\infty, \|\psi'_k\|_\infty, \|\psi''_k\|_\infty\}$ is bounded.
- A4** In the definition (2.9), $\tilde{h}_n = Ah_n$ for some constant $A \geq 4(B_K + C_Q)$.

We also assume that the kernels $K(\cdot)$ and $\bar{Q}(\cdot)$ satisfy conditions **B1** and **B2**, respectively. We need to make further assumptions about the covariance kernel C and the correlations among the sample curves. Let \mathbf{R} denote an $n \times n$ matrix with (i, j) -th entry ρ_{ij} . Assume:

- C1** $\lambda_1 > \lambda_2 > \dots > \lambda_M > 0$ and $\lambda_{M+1} = \dots = 0$. That is, the nonzero eigenvalues are all distinct and the covariance kernel is of finite dimension.
- C2** $\max_{1 \leq \nu \leq M} (\lambda_\nu - \lambda_{\nu+1})^{-1}$ is bounded above.
- C3** $\|\mathbf{R}\| \leq \kappa_n$ for some $\kappa_n > 0$ where $\|\cdot\|$ denotes the operator norm.

Note that, **C3** imposes a stability condition on the correlation matrix \mathbf{R} . In other words, the sample curves are “weakly correlated” as $\|\mathbf{R}\|$ is bounded by κ_n . Define $\underline{m}_n := \min_{1 \leq i \leq n} m_i$ and $\overline{m}_n := \max_{1 \leq i \leq n} m_i$. We further assume that

- C4** $\overline{m}_n / \underline{m}_n$ is bounded above as $n \rightarrow \infty$.

We now give the bias and variance of the proposed estimator of the covariance kernel $C(\cdot, \cdot)$.

Theorem 3.1. *Suppose that conditions **A1-A4**, **B1-B2** and **C4** hold. Assume that σ^2 is known and $\widehat{C}(\cdot) = \widehat{C}_*(\cdot) - \sigma^2$ where $\widehat{C}_*(\cdot)$ is defined through (2.5). Then, with $h_n = o(1)$ and $nh_n^2 \rightarrow \infty$, the estimator \widehat{C} satisfies:*

$$\mathbb{E}[\widehat{C}(s, t)] = C(s, t) + O(h_n^2), \tag{3.1}$$

$$\begin{aligned} \text{Var}[\widehat{C}(s, t)] &= O\left(\frac{1}{n}\right) + O\left(\max\left\{\frac{1}{nh_n^2 \underline{m}_n^2}, \frac{1}{nh_n \underline{m}_n}\right\}\right) \\ &\quad + \left(\frac{1}{n^2} \sum_{i \neq j}^n \rho_{ij}^2\right) O(1), \end{aligned} \tag{3.2}$$

where the $O(\cdot)$ terms are uniform in $s, t \in [0, 1]$.

As can be seen from Corollary 3.2 (Section 3.2) there is no change in the bounds even if σ^2 is replaced by $\widehat{\sigma}^2$ defined in (2.11).

Next, we state the result about the asymptotic behavior of the estimated eigenfunctions. Let the loss function for ψ_ν be the modified L^2 -loss:

$$L(\widehat{\psi}_\nu, \psi_\nu) := \|\widehat{\psi}_\nu - \text{sign}(\langle \widehat{\psi}_\nu, \psi_\nu \rangle) \psi_\nu\|_2^2, \tag{3.3}$$

where $\|\cdot\|_2$ denotes the L^2 norm, and $\langle \widehat{\psi}_\nu, \psi_\nu \rangle = \int_0^1 \widehat{\psi}_\nu(x) \psi_\nu(x) dx$. For Theorem 3.2, we only need to assume that the estimator $\widehat{\sigma}^2$ of σ^2 satisfies $\mathbb{E}(\widehat{\sigma}^2 - \sigma^2)^2 = o(1)$. See Corollary 3.1 for the asymptotic behavior of $\widehat{\sigma}^2$.

Theorem 3.2. *Suppose that conditions **A1-A4**, **B1-B2** and **C1-C4** hold. Let h_n be such that*

$$\overline{m}_n h_n = o(1), \quad \overline{m}_n^2 = o(nh_n^2 / \log n), \quad \text{and} \quad \kappa_n (\log n)^{1/2} n^{-1/2} (\overline{m}_n h_n)^{-1} = o(1) \tag{3.4}$$

as $n \rightarrow \infty$. Then the estimator $\widehat{\psi}_\nu$, which is the eigenfunction corresponding to the ν -th largest eigenvalue of \widehat{C} , satisfies: for any arbitrary but fixed $\epsilon > 0$,

$$\sup_{(C, g) \in \Theta} \mathbb{E}L(\widehat{\psi}_\nu, \psi_\nu) \tag{3.5}$$

$$\leq (1 + \epsilon) \frac{1}{n} \left(\sum_{1 \leq k \neq \nu \leq M} \frac{\lambda_k \lambda_\nu}{(\lambda_k - \lambda_\nu)^2} \right) \tag{3.6}$$

$$+ (1 + \epsilon) \left(\frac{1}{n^2} \sum_{i \neq j}^n \rho_{ij}^2 \right) \left(\sum_{1 \leq k \neq \nu \leq M} \frac{\lambda_k \lambda_\nu}{(\lambda_k - \lambda_\nu)^2} + O(h_n) \right) \tag{3.7}$$

$$+ O(h_n^4) + O\left(\frac{1}{nh_n \underline{m}_n}\right), \tag{3.8}$$

where Θ denotes the class of covariance-density pairs (C, g) satisfying the conditions **A1-A3** and **C1-C3**.

Remark 3.1. In Theorem 3.1, we have assumed that g is known. However, if a kernel-based estimate of g is used instead, then under assumptions **A1** and **A2**, it is straightforward to show that, if an optimal bandwidth is used to estimate g , then the rates of convergence of the resulting eigenfunction estimators, in terms of convergence in probability, are the same as those in Theorem 3.1. This follows from the fact that, under the stated assumptions, we have $\|\widehat{g} - g\|_2^2 = O_P((n\underline{m}_n)^{-4/5})$ and $\|\widehat{g} - g\|_\infty^2 = O_P(\log n(n\underline{m}_n)^{-4/5})$. These together imply that the contribution of the estimated g in the L^2 -risk of $\widehat{\psi}_\nu$ is no bigger than the best rate of convergence under the “ g known” case, which can be formally shown by a simple perturbation argument.

3.2. Asymptotics of σ^2

We can obtain the rate of convergence of the estimator $\widehat{\sigma}^2$ defined in (2.11) from Theorem 3.1, as illustrated in the following corollary.

Corollary 3.1. *Suppose that conditions **A1-A4**, **B1-B2** and **C4** hold. Then, with $h_n = o(1)$ and $nh_n^2 \rightarrow \infty$,*

$$\begin{aligned} \mathbb{E}(\widehat{\sigma}^2 - \sigma^2)^2 &= O\left(\frac{1}{n}\right) + O\left(\max\left\{\frac{1}{nh_n^2\underline{m}_n^2}, \frac{1}{nh_n\underline{m}_n}\right\}\right) \\ &\quad + \left(\frac{1}{n^2} \sum_{i \neq j}^n \rho_{ij}^2\right) O(1) + O(h_n^4), \end{aligned} \quad (3.9)$$

where the $O(\cdot)$ terms are uniform in $s, t \in [0, 1]$.

Using Corollary 3.1 and Theorem 3.1, we get a bound on the variance of the proposed estimator of the covariance kernel when σ^2 is estimated by $\widehat{\sigma}^2$.

Corollary 3.2. *Suppose that conditions **A1-A4**, **B1-B2** and **C4** hold. Then, with $h_n = o(1)$ and $nh_n^2 \rightarrow \infty$,*

$$\begin{aligned} \text{Var}[\widehat{C}(s, t)] &= O\left(\frac{1}{n}\right) + O\left(\max\left\{\frac{1}{nh_n^2\underline{m}_n^2}, \frac{1}{nh_n\underline{m}_n}\right\}\right) \\ &\quad + \left(\frac{1}{n^2} \sum_{i \neq j}^n \rho_{ij}^2\right) O(1) + O(h_n^4), \end{aligned} \quad (3.10)$$

where the $O(\cdot)$ terms are uniform in $s, t \in [0, 1]$.

3.3. Rates of convergence

One implication of Theorems 3.1 and 3.2 is that, if the correlation between sample curves is “weak” in a suitable sense, then the best upper bound on the L^2 risk for the correlated and i.i.d. cases are the same. Comparing with the i.i.d. case, we immediately see that, in order for this to hold for the eigenfunctions,

under the conditions of Theorem 3.2, we need

$$\frac{1}{n^2} \sum_{i \neq j} \rho_{ij}^2 = o\left(\frac{1}{nh_{n,e}\underline{m}_n}\right), \tag{3.11}$$

where $h_{n,e}$ stands for the optimal bandwidth for estimating eigenfunctions for the i.i.d. case. Moreover by Corollary 3.2, for the covariance kernel, it is sufficient that,

$$\frac{1}{n^2} \sum_{i \neq j} \rho_{ij}^2 = o\left(\max\left\{\frac{1}{nh_{n,c}\underline{m}_n}, \frac{1}{nh_{n,c}^2\underline{m}_n^2}\right\}\right), \tag{3.12}$$

where $h_{n,c}$ stands for the optimal bandwidth for estimating the covariance kernel for the i.i.d. case. Specifically, $h_{n,c} = (n\underline{m}_n^2)^{-1/6}$ (by Corollary 3.2 and under the setting $\overline{m}_n h_{n,c} = o(1)$), and $h_{n,e} = (n\underline{m}_n)^{-1/5}$ (by Theorem 3.2). Observe that, the optimal bandwidth for estimating the covariance and its eigenfunctions are different for the scenarios considered under Theorems 3.1 and 3.2. Specifically, (3.4) implies in particular that $\overline{m}_n = o((n/\log n)^{1/4})$, which essentially implies a regime of “sparse” observations.

As an example, (3.11) is satisfied if \mathbf{R} has the first order autoregressive structure, i.e., $\rho_{ij} = \rho^{|i-j|}$ for some $\rho \in (-1, 1)$. This will not hold if \mathbf{R} has the intra-class correlation structure, i.e., if $\rho_{ij} = \rho$ for all $i \neq j$.

We now claim that, the proposed estimator of the eigenfunctions achieves the optimal nonparametric rate when the trajectories are “weakly” correlated, in the sense that (3.11) holds, and \overline{m}_n is either bounded or increases slowly with n , as implied by (3.4). This claim follows from Theorem 3.2 and the following result (Theorem 3.3) which gives a lower bound on the rate of convergence of the first eigenfunction under the i.i.d. setting, when either \overline{m}_n is bounded or approaches infinity sufficiently slowly. This bound is a refinement over an analogous result (Theorem 2) in [13] in the sense that, for the latter, the number of measurements per curve was assumed to be bounded. It can be shown, by essentially using the same arguments, that as long as the covariance kernel C is of finite rank, and the distance between successive eigenvalues of C is bounded from below, the same asymptotic lower bound holds for estimators of any eigenfunction.

Theorem 3.3. *Let \mathcal{C} denote the class of covariance kernels $C(\cdot, \cdot)$ on $[0, 1]^2$ with rank ≥ 1 , and nonzero eigenvalues $\{\lambda_j\}_{j \geq 1}$ satisfying $k_0 \geq \lambda_1 > \lambda_2 \geq 0$ with $\lambda_1 - \lambda_2 \geq k_1$, and the first eigenfunction ψ_1 being twice differentiable and satisfying $\|\psi_1''\|_\infty \leq k_2$, for some constants $k_0, k_1, k_2 > 0$. Also, let \mathcal{G} denote the class of continuous densities g on $[0, 1]$ such that $c_1 \leq g \leq c_2$ for some $0 < c_1 \leq 1 \leq c_2 < \infty$. Also suppose that the number of measurements m_i 's satisfy $\underline{m}_n \leq m_i \leq \overline{m}_n$, for $\overline{m}_n \geq \underline{m}_n \geq 4$, such that $\overline{m}_n/\underline{m}_n \leq k_3$ for some $k_3 < \infty$, and $\overline{m}_n = o(n^{2/3})$. Let \mathcal{D} denote the space of such designs $D = \{m_i\}_{i=1}^n$. Then for sufficiently large n , for any estimator $\widehat{\psi}_1$ with l_2 norm one, the following holds:*

$$\sup_{D \in \mathcal{D}} \sup_{g \in \mathcal{G}} \sup_{C \in \mathcal{C}} \mathbb{E} \|\widehat{\psi}_1 - \psi_1\|_{l_2}^2 \geq k_4(n\underline{m}_n)^{-4/5}. \tag{3.13}$$

Proof of Theorem 3.3 is given in Appendix G of [22].

3.4. Connection between the dense design and the purely functional data case

We can regard the terms given by (3.6) and (3.7) as the parametric components of the risk, and the term in (3.8) as the nonparametric component, since the latter depends on the bandwidth h_n while the former do not (up to the leading order). Moreover, the term $O(h_n^4)$ in (3.8) can be seen as the integrated squared bias of $\widehat{\psi}_\nu$, while $O(1/(nh_n\bar{m}_n))$ is the nonparametric component of the integrated variance of $\widehat{\psi}_\nu$. As shown in Section 3.3, if we take $h_n \sim n^{-1/5}$, then for bounded \bar{m}_n , we get the optimal nonparametric rate if (3.11) also holds (note that $h_{n,e} \sim n^{-1/5}$).

In contrast, we can consider the case when $\bar{m}_n \gg n^{1/4} \log n$ which may be referred to as the “dense” design. This setting is clearly different from the one described by (3.4), referred to as the “sparse” design. Even though a detailed analysis of this setting is beyond the scope of this paper, here we discuss the key aspects of the asymptotic behavior of the L^2 -risk of the eigenfunction estimator under the “dense” design. It can be shown that, with appropriate choice of the bandwidth, the risks of the estimated eigenfunctions have a parametric rate of convergence. The main technical difference from the derivations under the “sparse” design is that a modification to the proof of Proposition 4.2 is needed in order to obtain an analogous result on the fluctuation of the estimated covariance kernel from its expectation. Specifically, suppose that $\bar{m}_n \gg n^{1/4} \log n$ and $\kappa_n = o(\sqrt{n/\log n})$. If h_n is chosen such that $h_n = o(n^{-1/4})$ and $\bar{m}_n h_n \gg \sqrt{\log n}$, then the rate of convergence of the estimated eigenfunctions coincides with that of the “purely functional” setting, i.e., when the individual sample curves are observed completely and without noise [9]. Indeed, the difference from the “sparse” case lies only in the maximal fluctuation of the estimated covariance kernel from its expectation. Here, we continue to have that the asymptotic integrated squared bias for the estimated eigenfunctions is $O(h_n^4)$ and the nonparametric component of the integrated variance is $O((nh_n\bar{m}_n)^{-1})$. However, now both these terms are $o(n^{-1})$ since $h_n = o(n^{-1/4})$ and $\bar{m}_n h_n \gg \sqrt{\log n}$. Hence, both are negligible compared to the parametric component of the asymptotic variance which is of the order $O(\max\{n^{-1}, n^{-2} \sum_{i \neq j} \rho_{ij}^2\})$. This phenomenon is also indicated in [13] when the trajectories are i.i.d. In other words, asymptotically there is no difference between the risk of the estimated eigenfunctions from densely sampled, noisy data, and that from data measured on the continuum and without noise. Thus, for the “sparse” design, smoothing is effectively performed only on the covariance kernel, whereas for the “dense” design, smoothing has the effect of denoising individual curves as well.

3.5. Asymptotics under functional coregionalization model

The above asymptotic analysis can be carried out for the more general class of models, the “functional coregionalization models”. Here, we only present a result on the asymptotic risk of estimated eigenfunctions when the principal

component scores follow (1.4). For simplicity of exposition, we assume that $\rho_{ij}^{(\nu\nu')} = \delta_{\nu\nu'}\rho_{ij}^{(\nu)}$ for $1 \leq \nu, \nu' \leq M$ and $i, j = 1, \dots, n$ with $\rho_{ii}^{(\nu)} = 1$ for all i, ν . In other words, the principal component scores $(\xi_{i\nu})_{i=1}^n$ are uncorrelated for different ν . Under this setting, we have

$$\text{Cov}(X_i(s), X_j(t)) = \sum_{\nu=1}^M \lambda_\nu \rho_{ij}^{(\nu)} \psi_\nu(s) \psi_\nu(t), \quad s, t \in [0, 1].$$

Under an asymptotic identifiability condition (C6), we obtain the following generalization (Theorem 3.4) of Theorem 3.2. This result follows essentially by using the same arguments in proving Theorems 3.1 and 3.2. It has an analog when the data are correlated high-dimensional random vectors instead of correlated functional data [9].

Define $\mathbf{R}_\nu = ((\rho_{ij}^{(\nu)}))$, $\nu = 1, \dots, M$. We introduce conditions on \mathbf{R}_ν .

C5 $\max_{1 \leq \nu \leq M} \|\mathbf{R}_\nu\| \leq \kappa_n$ for some $\kappa_n > 0$.

C6 $\max_{1 \leq \nu \leq M} \frac{1}{n^2} \text{tr}[(\mathbf{R}_\nu - I_n)^2] \rightarrow 0$ as $n \rightarrow \infty$.

Note that, C5 is an analog of C3. Also, C6 provides the asymptotic identifiability of the eigenfunctions $\{\psi_\nu\}_{\nu=1}^M$ of the kernel $C(s, t) = \sum_{\nu=1}^M \lambda_\nu \psi_\nu(s) \psi_\nu(t)$. This is because, C6 implies in particular that $|\frac{1}{n} \text{tr}(\mathbf{R}_\nu) - 1| \rightarrow 0$ as $n \rightarrow \infty$ for all ν , and therefore,

$$\frac{1}{n} \sum_{i=1}^n \text{Cov}(X_i(s), X_i(t)) = \sum_{\nu=1}^M \lambda_\nu \left(\frac{1}{n} \sum_{i=1}^n \rho_{ii}^{(\nu)} \right) \psi_\nu(s) \psi_\nu(t) \rightarrow \sum_{\nu=1}^M \lambda_\nu \psi_\nu(s) \psi_\nu(t)$$

as $n \rightarrow \infty$.

Theorem 3.4. *Suppose that conditions A1-A4, B1-B2, C1-C2 and C4-C6 hold. Let h_n satisfy (3.4). Then the estimator $\hat{\psi}_\nu$, which is the eigenfunction corresponding to the ν -th largest eigenvalue of \hat{C} , satisfies: for any arbitrary but fixed $\epsilon > 0$,*

$$\begin{aligned} & \sup_{(C, g) \in \Theta} \mathbb{E}L(\hat{\psi}_\nu, \psi_\nu) \\ & \leq (1 + \epsilon) \frac{1}{n} \left(\sum_{1 \leq k \neq \nu \leq M} \frac{\lambda_k \lambda_\nu}{(\lambda_k - \lambda_\nu)^2} \right) \\ & \quad + (1 + \epsilon) \left(\sum_{1 \leq k \neq \nu \leq M} \left(\frac{1}{n^2} \text{tr}(\mathbf{R}_k \mathbf{R}_\nu - I_n) \right) \left(\frac{\lambda_k \lambda_\nu}{(\lambda_k - \lambda_\nu)^2} + O(h_n) \right) \right) \\ & \quad + O(h_n^4) + O\left(\frac{1}{nh_n \underline{m}_n}\right), \end{aligned} \tag{3.14}$$

where Θ denotes the class of covariance-density pairs (C, g) satisfying the conditions A1-A3, C1-C2 and C5-C6.

3.6. Infinite dimensional covariance kernel

Our theoretical analysis assumes that only a finite number of eigenvalues of the covariance kernel are nonzero. There are two issues while extending the results to the setting where the number of nonzero eigenvalues is infinite. The first one relates to the model for correlation across the trajectories. Notice that “functional coregionalization” model discussed in Section 3.5 assumes a different correlation structure among the trajectories for each principal component score (or eigenvalue). In this case, unless fairly elaborate structural assumptions are made about these component-specific correlation matrices, it is not feasible to establish the rates of convergence results. Thus, we only focus on the case when the covariance structure is “separable”, i.e., satisfying equation (1.3). The second issue relates to the regularity of the covariance kernel of the trajectories. In order to establish consistency of the leading eigenfunctions, we require a certain rate of decay of the eigenvalues and a certain degree of regularity of the eigenfunctions. Specifically, we assume that ψ_k is twice continuously differentiable for all k , and

$$\sum_{k=1}^{\infty} k^{\beta} \lambda_k \|\psi_k\|_{\infty}^2 < \infty \quad \text{for some } \beta > 1, \quad (3.15)$$

where $\|\psi_k\|_{\infty} = \sup_{s \in [0,1]} |\psi_k(s)|$. (3.15) is a mild condition. For example, if $\psi_k(t) = \sqrt{2} \cos(2k\pi t)$ for k even and $\psi_k(t) = \sqrt{2} \sin(2k\pi t)$ for k odd, then under the assumption that $C(s, t)$ is twice continuously differentiable, $\sum_{k=1}^{\infty} k^2 \lambda_k \|\psi_k\|_{\infty}^2 = \sum_{k=1}^{\infty} k^2 \lambda_k < \infty$.

We present the following result about the rate of convergence of the estimated eigenfunctions when the covariance structure of the principal component scores is separable and the covariance kernel C is twice continuously differentiable and satisfies (3.15).

Theorem 3.5. *Assume that the observations satisfy (1.1) and (1.2) with the separable correlation structure (1.3) and with $\mu \equiv 0$, $\sigma = 1$. Suppose that the covariance-density pair (C, g) satisfies conditions **A1**, **A2**, **A4**, **B1**, **B2**, **C3** and **C4**, that the eigenvalues of C are all distinct and the set $\{(\lambda_k, \psi_k) : k = 1, 2, \dots\}$ of eigenvalue-eigenfunction pairs satisfy (3.15), and that the kernel C is twice continuously differentiable with bounded second derivative. Moreover, assume that the bandwidth h_n satisfies (3.4). Let ν be any fixed index. Then, for some constant $c > 1$ (not depending on C and g),*

$$\begin{aligned} \mathbb{E}L(\widehat{\psi}_{\nu}, \psi_{\nu}) &\leq c \frac{1}{n} \left(\sum_{k \neq \nu}^{\infty} \frac{\lambda_k \lambda_{\nu}}{(\lambda_k - \lambda_{\nu})^2} \right) \\ &\quad + c \left(\frac{1}{n^2} \sum_{i \neq j}^n \rho_{ij}^2 \right) \left(\sum_{k \neq \nu}^{\infty} \frac{\lambda_k \lambda_{\nu}}{(\lambda_k - \lambda_{\nu})^2} + O(h_n) \right) \\ &\quad + O(h_n^4) + O\left(\frac{1}{nh_n \underline{m}_n}\right), \end{aligned}$$

where $\widehat{\psi}_{\nu}$ denotes the eigenfunction corresponding to the ν -th largest eigenvalue of \widehat{C} , the estimated covariance kernel defined in (2.9).

The proof of Theorem 3.5 follows a similar line of arguments as that of Theorem 3.1. In order to handle the possibility of infinite number of nonzero eigenvalues for the covariance kernel C , we extend our method of proof through a truncation technique. In essence, now M represents the number of “significant eigenvalues”, and we let M grow with n at an appropriate rate, so that the projection of $C(s, t)$ onto the span of $\{\psi_k : k > M\}$ is negligible. The details are omitted.

4. Outline of the Proof of Theorems 3.1 and 3.2

In this section, we briefly describe the main ideas leading to the proofs of Theorems 3.1 and 3.2, which require repeated computation of mixed moments of correlated Gaussian random variables. The details are given in the appendices. The basic idea in the computation of the moments is to treat the diagonal and the off-diagonal parts of $\widehat{C}(\cdot, \cdot)$ separately. The proof of Theorem 3.2 heavily relies on an application of Lemma A.1, which leads to a first order expansion of the eigenfunctions of the estimated covariance kernel around the corresponding eigenfunctions of the true covariance kernel, and provides a bound on the approximation error. In view of this lemma, the key quantity in the derivation of asymptotic risk is the computation of $\mathbb{E} \| H_\nu \widehat{C} \psi_\nu \|_2^2$ (Proposition 4.1), where the operator H_ν is defined in (4.1). Once we obtain an expression for this (as given in Section 4.1), we use a probabilistic bound on the operator norm of the difference between estimated and true covariance kernels (Proposition 4.2), to complete the proof.

Since the covariance kernel and its estimate are continuous functions, we shall treat H_ν as an operator on $C([0, 1])$ and extensively use the following representation of its kernel:

$$\begin{aligned}
 H_\nu(x, y) &:= \overline{H}_\nu(x, y) - \frac{1}{\lambda_\nu} \delta(x, y), \quad \text{with} \\
 \overline{H}_\nu(x, y) &= \sum_{1 \leq k \neq \nu \leq M} \frac{\lambda_k}{\lambda_\nu(\lambda_k - \lambda_\nu)} \psi_k(x) \psi_k(y) + \frac{1}{\lambda_\nu} \psi_\nu(x) \psi_\nu(y), \quad (4.1)
 \end{aligned}$$

where $\delta(x, y)$ denotes the Dirac’s δ operator, i.e., $\int \delta(x, y) f(y) dy = f(x)$ for all $f \in C([0, 1])$. Note that we treat H_ν interchangeably as an integral operator on $C([0, 1])$ and its kernel. Thus, H_ν operates on $f \in C([0, 1])$ as

$$(H_\nu f)(x) = \int_0^1 H_\nu(x, y) f(y) dy = \int_0^1 \overline{H}_\nu(x, y) f(y) dy - \frac{1}{\lambda_\nu} f(x), \quad x \in [0, 1].$$

The key properties of the kernel $H_\nu(\cdot)$ that we repeatedly use (and which are valid due to the orthonormality of the functions ψ_1, \dots, ψ_M) are

$$\begin{aligned}
 (H_\nu \psi_\nu)(x) &= 0 \\
 (H_\nu \psi_k)(x) &= \frac{1}{\lambda_k - \lambda_\nu} \psi_k(x), \quad \text{for } k \neq \nu \\
 \text{and } (H_\nu f)(x) &= -\frac{1}{\lambda_\nu} f(x)
 \end{aligned}$$

for $f \in C([0, 1])$ such that $\int_0^1 f(x) \psi_k(x) dx = 0$ for all $k = 1, \dots, M$.

We use the operator H_ν to produce a first order expansion of $\widehat{\psi}_\nu$ around ψ_ν by a perturbation analysis of the estimated covariance kernel treated as an integral operator. Lemma A.1 in Appendix A indicates the role of H_ν in this analysis. Note that, we have the following correspondence with the objects in Lemma A.1: $A \equiv C$, $B \equiv \widehat{C} - C$, $H_r(A) \equiv H_\nu$, $p_r(A) \equiv \psi_\nu$ and $p_r(A + B) \equiv \widehat{\psi}_\nu$. The main advantage of this expansion and the form of the bound on the remainder term is that they reduce the problem of bounding the L^2 -risk of $\widehat{\psi}_\nu$ essentially to that of computing $\mathbb{E} \| H_\nu \widehat{C} \psi_\nu \|_2^2$, i.e., the expected value of the squared integral of the first term on the RHS of (A.1). This is because, by virtue of Lemma A.1, we have $L(\widehat{\psi}_\nu, \psi_\nu) \leq \| H_\nu \widehat{C} \psi_\nu \|_2^2 (1 + c\epsilon)$ for some constant $c > 0$ if $\| \widehat{C} - C \| < \epsilon$ for any $\epsilon > 0$ suitably small.

4.1. Asymptotic risk for estimating ψ_ν

The key results in this section are Propositions 4.1 and 4.2, which are proved in the Appendix C.

Proposition 4.1. *Under the assumptions of Theorem 3.2, we have*

$$\begin{aligned} & \mathbb{E} \| H_\nu \widehat{C} \psi_\nu \|_2^2 \\ & \leq (1 + \epsilon) \frac{1}{n} \left(\sum_{1 \leq k \neq \nu \leq M} \frac{\lambda_k \lambda_\nu}{(\lambda_k - \lambda_\nu)^2} \right) \\ & \quad + (1 + \epsilon) \left(\frac{1}{n^2} \sum_{i \neq j} \rho_{ij}^2 \right) \left(\sum_{1 \leq k \neq \nu \leq M} \frac{\lambda_k \lambda_\nu}{(\lambda_k - \lambda_\nu)^2} + O(h_n) \right) \\ & \quad + O(h_n^4) + O\left(\frac{1}{nh_n \underline{m}_n}\right) \end{aligned} \tag{4.2}$$

for any arbitrary but fixed $\epsilon > 0$.

Here we briefly describe the main idea of the proof. For convenience of exposition, throughout we replace $\max\{\widehat{C}(\frac{s+t}{2}), h_n^2\}$ in the definition (2.9) by $\widetilde{C}(\frac{s+t}{2})$. Also, for convenience, we consider the unsmoothed version (2.8) of the kernel W , and take $\widetilde{h}_n = Ah_n$, where $A \geq 4(B_K + C_Q)$. The purpose is to separate the contributions from the diagonal and off-diagonal parts of the estimator. The main decompositions that facilitate the computations are given by (4.1), and (C.1) and (C.2) in Appendix C. As indicated in the previous subsection, (4.1) is used in isolating the contributions of the terms involving the eigenfunctions $\{\psi_k\}$ in the expansion of $\widehat{\psi}_\nu$ around ψ_ν , and (C.2) reduces bounding the term $\mathbb{E} \| H_\nu \widehat{C} \psi_\nu \|_2^2$ to that of bounding $\mathbb{E} \| H_\nu \widetilde{C}_c \psi_\nu \|_2^2$, where $\widetilde{C}_c(s, t)$ is simply replacing $\widehat{\sigma}^2$ by σ^2 in the definition of $\widehat{C}(s, t)$ (see (C.1)). The crucial computational advantage is that, now the bounds can be obtained simply by analyzing the bias and covariances of the kernel $\widetilde{C}_c(s, t)$. Note also that, if σ^2 is assumed to be known, then the decomposition (C.2) is not required, and we can get rid of the multiplicative factor $(1 + \epsilon)$ in the expression (3.5) for the risk in Theorem 3.2.

4.2. Norm bound on $\widehat{C} - \mathbb{E}\widehat{C}$

To complete the proof of Theorem 3.2, we need to find a probabilistic bound for $\|\widehat{C} - \mathbb{E}\widehat{C}\|$, where $\|\cdot\|$ denotes the operator norm. We first find a bound on the supremum norm $\|\widehat{C} - \mathbb{E}\widehat{C}\|_\infty := \sup_{x,y \in [0,1]} |\widehat{C}(x,y) - \mathbb{E}\widehat{C}(x,y)|$. We then use the fact that, $\|\widehat{C} - \mathbb{E}\widehat{C}\| \leq \|\widehat{C} - \mathbb{E}\widehat{C}\|_F$, where $\|\cdot\|_F$ denotes the Hilbert-Schmidt norm, and $\|\widehat{C} - \mathbb{E}\widehat{C}\|_F \leq \|\widehat{C} - \mathbb{E}\widehat{C}\|_\infty$. Note that, by piecewise differentiability of the estimate \widehat{C} , in order to provide exponential bounds for the variation of $\|\widehat{C} - \mathbb{E}\widehat{C}\|_\infty$, it is enough to derive exponential bounds for the fluctuations of $|\widehat{C}(s,t) - \mathbb{E}[\widehat{C}(s,t)]|$ for a finite (but polynomially growing with n) number of points $(s,t) \in [0,1]$. In the following proposition, without loss of generality, we assume that $g \equiv 1$.

Proposition 4.2. *Suppose that conditions A1-A4, B1-B2 and C1-C4 hold, and let h_n satisfy (3.4). Then, given $\eta > 0$, there is a $c_\eta > 0$ such that for every fixed $s, t \in (0, 1)$,*

$$\mathbb{P}\left(|\widehat{C}(s,t) - \mathbb{E}(\widehat{C}(s,t))| > c_\eta \kappa_n \sqrt{\frac{\log n}{n\overline{m}_n^2 h_n^2}}\right) \leq n^{-\eta}. \tag{4.3}$$

By Lemma A.1 and the fact that $\|\widehat{\psi}_\nu\|_2 = \|\psi_\nu\|_2 = 1$,

$$\begin{aligned} \mathbb{E}L(\widehat{\psi}_\nu, \psi_\nu) &\leq \mathbb{E}\|H_\nu \widehat{C} \psi_\nu\|_2^2 (1 + \delta_{n,\eta}) \\ &\quad + 2\mathbb{P}\left(\|\widehat{C} - \mathbb{E}(\widehat{C})\| > c'_\eta \kappa_n \sqrt{\frac{\log n}{n\overline{m}_n^2 h_n^2}}\right) \end{aligned}$$

for some $\eta > 0$, $c'_\eta > 0$ and $\delta_{n,\eta} \rightarrow 0$ appropriately chosen. Then the proof of Theorem 3.2 follows by using Propositions 4.1 and 4.2.

5. Simulation

We carry out a simulation study in which the true trajectories follow the model (1.1):

$$X_i(t) = \sum_{\nu=1}^3 \sqrt{\lambda_\nu} \xi_{i\nu} \psi_\nu(t), \quad t \in [0, 1], \tag{5.1}$$

where $\lambda_\nu = \nu^{-0.6}$, i.e., $\lambda_1 = 1$, $\lambda_2 \approx 0.66$ and $\lambda_3 \approx 0.52$; $\psi_\nu(t) = \sqrt{2} \sin(\nu\pi t)$; $\xi_{i\nu}$ are i.i.d. $N(0, 1)$ and $\text{Cov}(\xi_{i\nu}, \xi_{j\nu'}) = \delta_{\nu\nu'} \rho^{|i-j|}$ where, by convention, $0^0 = 1$. We consider three different values of ρ : 0, 0.5 and 0.9. The observations are generated by the model (1.2):

$$Y_{ij} = X_i(T_{ij}) + \sigma \varepsilon_{ij}, \quad j = 1, \dots, m_i; \quad i = 1, \dots, n$$

where m_i 's are i.i.d. $\{10, \dots, 20\}$, T_{ij} 's are i.i.d. $\text{Uniform}(0, 1)$ and ε_{ij} 's are i.i.d. $N(0, 1)$. Throughout this section, $n = 250$ and $\sigma = 0.4$. The choice of

the eigenfunctions is motivated by the real data application where the observed trajectories are nearly periodic (see Section 6).

We use the following “bi-weight” kernel for pre-smoothing the sample curves

$$K(x) = \frac{15}{16}(1 - x^2)^2 \mathbf{1}(|x| \leq 1). \quad (5.2)$$

We estimate the mean function $\mu(\cdot)$ and the density $g(\cdot)$ of the (pooled) sampling points using a local linear regression estimator and a kernel density estimator, respectively, where the bandwidths are chosen by cross-validation. We subtract the estimated mean from the observations to get the centered observations $\widehat{Y}_{ij} = Y_{ij} - \widehat{\mu}(T_{ij})$ which are used as inputs for the proposed procedure. We estimate the noise variance σ^2 using the method described in Section 2.3. We also choose $T_0 = 0.2$ and $T_1 = 0.8$ in (2.10).

We propose to use cross-validation to choose the following criterion for determination of the optimal bandwidth h as well as the number of nonzero eigenvalues. We randomly select 20% of the sample trajectories as “test data”, denoted by \mathcal{T} , and set them aside. We then estimate the covariance kernel from the rest of the data using a bandwidth \widetilde{h} and denote the corresponding noise variance, eigenvalues and eigenfunctions by $\widetilde{\sigma}^2$, $\widetilde{\lambda}_\nu$ and $\widetilde{\psi}_\nu$, $\nu = 1, 2, \dots$, respectively. Then, for each $\widetilde{M} \geq 1$, we compute the following cross-validated negative pseudo-log-likelihood (pretending that the sample trajectories are independent Gaussians, also referred to as the *empirical Kullback-Leibler loss*):

$$\text{CV}(\widetilde{h}, \widetilde{M}) = \frac{1}{2} \sum_{i \in \mathcal{T}} \widehat{\mathbf{Y}}_i^T \widetilde{\Sigma}_i^{-1} \widehat{\mathbf{Y}}_i + \frac{1}{2} \sum_{i \in \mathcal{T}} \log |\widetilde{\Sigma}_i| \quad (5.3)$$

where $\widehat{\mathbf{Y}}_i = (\widehat{Y}_{i1}, \dots, \widehat{Y}_{im_i})^T$ and

$$\widetilde{\Sigma}_i = \sum_{\nu=1}^{\widetilde{M}} \widetilde{\lambda}_\nu \widetilde{\psi}_{i\nu} \widetilde{\psi}_{i\nu}^T + \widetilde{\sigma}^2 I_{m_i}$$

where $\widetilde{\psi}_{i\nu} = (\widetilde{\psi}_\nu(T_{i1}), \dots, \widetilde{\psi}_\nu(T_{im_i}))^T$. We choose the combination $(\widehat{h}, \widehat{M})$ that minimizes the cross-validation score (5.3). Finally, the estimates $\widehat{\lambda}_\nu$ and $\widehat{\psi}_\nu$, $\nu = 1, \dots, \widehat{M}$ are computed from the complete data set using the bandwidth \widehat{h} .

For comparison, we carry out the PACE procedure in [31] based on local polynomial smoothing of the empirical covariances, where the bandwidths are selected by leave-one-curve-out cross-validation.

We report the estimation errors for the eigenfunctions and eigenvalues, averaged over 50 independent replicates, in Tables 1 and 2, respectively. The results show that for both estimation procedures, compared to the i.i.d. case (i.e., $\rho = 0$), the estimation accuracy only mildly deteriorates when the correlation is moderate (i.e., $\rho = 0.5$). However, when the correlation is high (i.e., $\rho = 0.9$), the performance deteriorates considerably. Also, for all ρ , the proposed method performs better than PACE in terms of estimating the eigenfunctions. PACE is better for estimating the first eigenvalues while the two methods perform similar

TABLE 1
Estimation errors for the eigenfunctions: the errors are measured by the integrated squared deviations (ISE) from the truth

		Kernel smoothing			PACE		
		ψ_1	ψ_2	ψ_3	ψ_1	ψ_2	ψ_3
$\rho = 0$	mean(ISE)	0.0418	0.1285	0.1263	0.0768	0.2234	0.2010
	median(ISE)	0.0322	0.0643	0.0591	0.0559	0.1349	0.0877
	s.d.(ISE)	0.0406	0.2170	0.2133	0.0761	0.2500	0.2384
$\rho = 0.5$	mean(ISE)	0.0486	0.1308	0.1315	0.0817	0.2772	0.2601
	median(ISE)	0.0337	0.0681	0.0731	0.0599	0.1795	0.1487
	s.d.(ISE)	0.0421	0.1762	0.1708	0.0769	0.3604	0.3585
$\rho = 0.9$	mean(ISE)	0.2975	0.5095	0.3953	0.3478	0.6136	0.5017
	median(ISE)	0.1401	0.3097	0.2171	0.1567	0.4409	0.3060
	s.d.(ISE)	0.4031	0.4982	0.4564	0.4467	0.5426	0.5109

TABLE 2
Estimation errors for the eigenvalues: the errors are measured by the squared deviations (SE) from the truth

		Kernel smoothing			PACE		
		ψ_1	ψ_2	ψ_3	ψ_1	ψ_2	ψ_3
$\rho = 0$	mean(SE)	0.0223	0.0054	0.0043	0.0110	0.0046	0.0053
	median(SE)	0.0113	0.0035	0.0025	0.0051	0.0030	0.0039
	s.d.(SE)	0.0277	0.0067	0.0063	0.0173	0.0048	0.0057
$\rho = 0.5$	mean(SE)	0.0273	0.0057	0.0057	0.0171	0.0067	0.0088
	median(SE)	0.0149	0.0025	0.0032	0.0084	0.0038	0.0070
	s.d.(SE)	0.0333	0.0066	0.0073	0.0238	0.0081	0.0094
$\rho = 0.9$	mean(SE)	0.0718	0.0250	0.0298	0.0574	0.0275	0.0350
	median(SE)	0.0353	0.0158	0.0212	0.0300	0.0180	0.0317
	s.d.(SE)	0.1107	0.0321	0.0247	0.0731	0.0278	0.0259

in estimating the other two eigenvalues. Finally, the model selection criterion (5.3) selects the true model dimension $M(= 3)$ in 90%, 88% and 92% replicates for $\rho = 0, 0.5$ and 0.9 , respectively.

6. Application

We apply the proposed procedure to a time course gene expression data that measure the expression profiles of cell-cycle regulated genes [28]. The experiment is concerned with a collection of yeast cells whose cycles were synchronized according to the α -factor. There are 6178 genes in total, and the expression level of each gene was, measured every 7 minutes between 0 and 119 minutes, covering two cell cycles. This data set has been analyzed by a number of researchers, including [30] who use it as an illustration of their method for functional principal component analysis. In this paper, we focus on the ‘‘G1 cluster’’ which, after the elimination of genes with very little or no measurements, consists of 297 genes. It is expected that, the expression profiles of these genes are correlated. Among these 297 genes, 74 genes have one or more observations are missing. The expression trajectories for all 297 genes, and the mean function estimated by local linear regression are plotted in Figure 2.

We apply the proposed method and find four significant principal components. These four eigenfunctions are plotted in Figure 3. To check the prediction

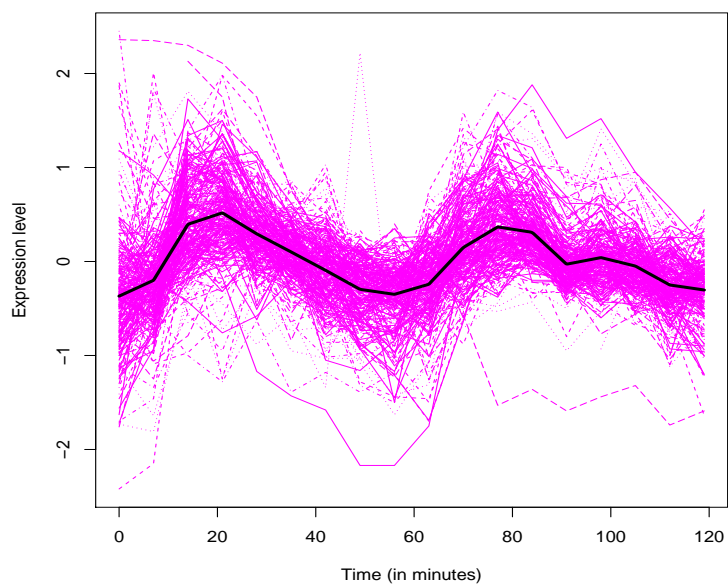


FIG 2. Gene expression profiles for the 297 genes in the 'G1 cluster' of the yeast cell cycle data (purple trajectories), together with the estimated mean function (solid black line).

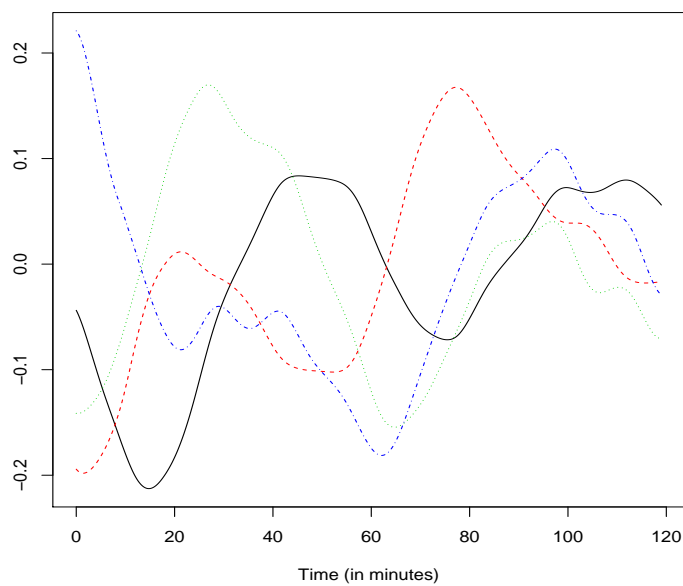


FIG 3. Four leading eigenfunctions for the "G1 cluster" of the yeast cell cycle data. First: '—' (black), second: '- -' (red), third: '...' (green), fourth: '- . -' (blue).

performance, we fitted the expression profiles for these genes by

$$\hat{X}_i(t) = \hat{\mu}(t) + \sum_{\nu=1}^{\hat{M}} \sqrt{\hat{\lambda}_\nu} \hat{\xi}_{i\nu} \hat{\psi}_\nu(t).$$

Here, the principal components scores are estimated by the best linear unbiased predictor, i.e., $\hat{\xi}_{i\nu} = \sqrt{\hat{\lambda}_\nu} \hat{\psi}_{i\nu}^T \hat{\Sigma}_i^{-1} \hat{\mathbf{Y}}_i$, where $\hat{\mathbf{Y}}_i = (Y_{ij} - \hat{\mu}(T_{ij}))_{j=1}^{m_i}$ and $\hat{\Sigma}_i = \sum_{\nu=1}^{\hat{M}} \hat{\lambda}_\nu \hat{\psi}_{i\nu} \hat{\psi}_{i\nu}^T + \hat{\sigma}^2 I_{m_i}$ with $\hat{\psi}_{i\nu} = (\hat{\psi}_\nu(T_{ij}))_{j=1}^{m_i}$. The fitted trajectories of four genes along with their observed expression levels are depicted in Figure 4. As can be seen from this figure, the fitted trajectories are able to capture the main features of the observed expression profiles.

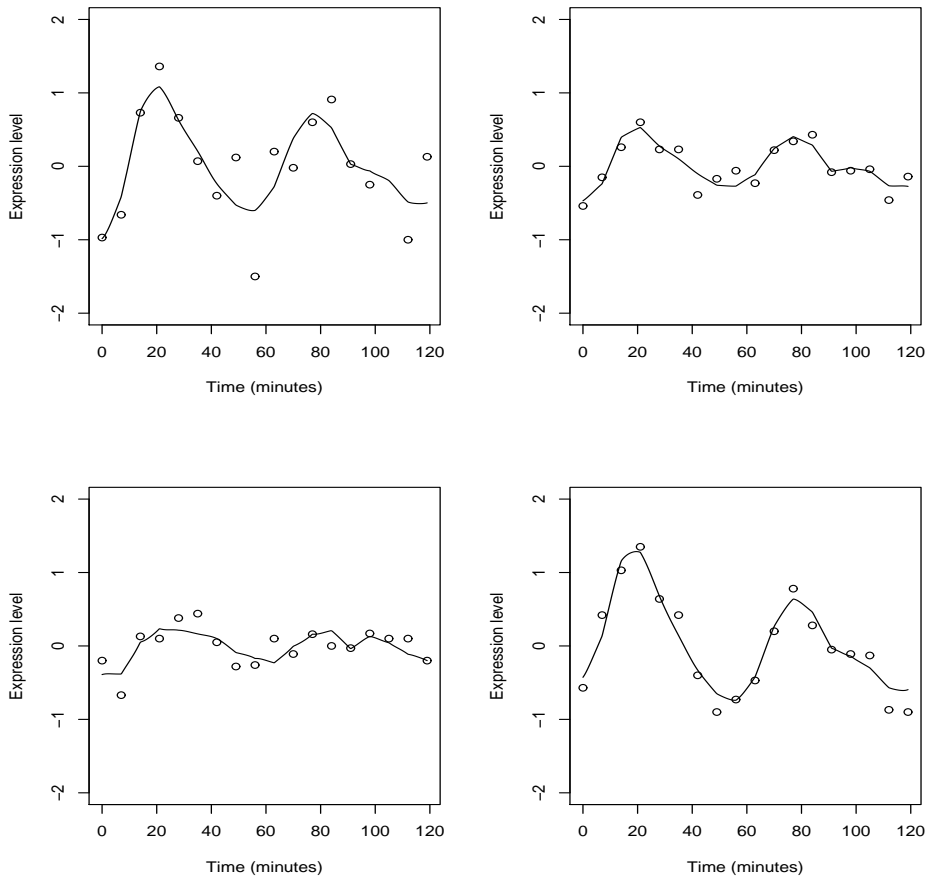


FIG 4. True (dots) and predicted (solid lines) expression profiles for four genes in the “G1 cluster” of the yeast cell cycle data.

7. Summary

In this paper, we study FPCA for correlated functional data. In particular, we propose and analyze an estimator for the covariance kernel which is based on merging two separate estimators: (i) the estimator of the off-diagonal part based on computing linearized empirical covariances of pre-smoothed sample curves; (ii) the estimator of the diagonal part based on linearized kernel smoothing of the empirical variances. This estimator is motivated by the observation that the empirical covariance kernel of the pre-smoothed sample curves is a highly biased estimator along the diagonal under the sparse measurements setting. Asymptotic risk behavior of the proposed estimator is studied under the assumption that the covariance of the noiseless processes is separable (Section 3.1) or follows a functional coregionalization model (section 3.5). Exact quantification of the correlation effect on the asymptotic risk for the eigenfunctions is obtained (Theorems 3.2 and 3.4). It is also shown that the L^2 -risk for the eigenfunctions achieves the optimal rate when the correlation is weak and the number of measurements per curve is bounded. We also conduct simulation studies to illustrate the effect of correlation on the proposed method. As expected, for moderate levels of correlation, there is little degradation in the performance of the estimators compared to the i.i.d. case. However, when the correlation is strong, there is a considerable degradation.

Appendix A

Perturbation of eigen-structure

The following lemma about the perturbation of the eigenfunctions of an integral operator of Hilbert-Schmidt class is a modified version of a similar result in [21]. Several variants of this lemma appear in the literature (see, e.g., [4, 19]), and most of them implicitly use the approach taken in [18]. In the following we use $\|A\|$ to denote the operator norm of an integral operator A , i.e., the largest singular value of A , acting on $L^2([0, 1])$.

Lemma A.1. *Let A and B be two integral operators acting on $L^2([0, 1])$ with symmetric, continuous, bounded kernels on $[0, 1]$, and let A be of finite rank. Let the unique nonzero eigenvalues of operator A be denoted by $\lambda_1(A), \lambda_2(A), \dots, \lambda_m(A)$. For any $r \geq 1$, if $\lambda_r(A)$ is of multiplicity 1, then denote the eigenfunction associated with the r -th eigenvalue by $p_r(A)$. Use analogous notations for the eigenvalues and eigenfunctions of the operator $A + B$. Then*

$$p_r(A + B) - \text{sign}\langle p_r(A + B), p_r(A) \rangle p_r(A) = -H_r(A) B p_r(A) + R_r \quad (\text{A.1})$$

where

$$H_r(A) := \sum_{1 \leq s \neq r \leq m} \frac{1}{\lambda_s(A) - \lambda_r(A)} P_{\mathcal{E}_s}(A) - \frac{1}{\lambda_r(A)} \left(\delta - \sum_{1 \leq s \leq m} P_{\mathcal{E}_s}(A) \right)$$

where $P_{\mathcal{E}_s}(A)$ denotes the orthogonal projection operator onto the eigen-subspace \mathcal{E}_s corresponding to eigenvalue $\lambda_s(A)$ (possibly multi-dimensional) and δ denotes the Dirac's delta function. Define Δ_r

$$\Delta_r := \frac{\|B\|}{\min\{\min_{1 \leq s \neq r \leq m} |\lambda_s(A) - \lambda_r(A)|, |\lambda_r(A)|\}}.$$

Then, the residual term R_r in (A.1) can be bounded as

$$\|R_r\| \leq \min \left\{ 10\Delta_r^2, \Delta_r \|H_r(A)Bp_r(A)\| \cdot \left(\frac{2(1+2\Delta_r)}{1-2\Delta_r(1+2\Delta_r)} + \frac{1}{(1-2\Delta_r(1+2\Delta_r))^2} \right) \right\}$$

where the second bound holds only if $\Delta_r < (\sqrt{5} - 1)/4$.

Deviations of quadratic forms

In the following, suppose that $\Phi : \mathcal{Z} \rightarrow \mathbb{R}^{n \times n}$ is a measurable function. Let Z be a random variable taking values in \mathcal{Z} . Let $C(\kappa, A, B) := \{z \in \mathcal{Z} : \|\Phi(z)\| \leq \kappa A \text{ and } \text{tr}(\Phi(z)\Phi(z)^T) \leq \kappa^2 B^2\}$ for constants $\kappa, A, B > 0$.

Lemma A.2. *Suppose that X and Y are i.i.d. $N_n(0, I)$ and are independent of Z . Then for all L such that $L\kappa A < 1/\sqrt{2}$ and for all $0 < t < L\kappa^2 B^2(1+L^2\kappa^2 A^2)$,*

$$\begin{aligned} & \mathbb{P}(|X^T \Phi(Z)Y| > t, Z \in C(\kappa, A, B)) \\ & \leq 2 \exp\left(-\frac{t^2}{2\kappa^2 B^2(1+L^2\kappa^2 A^2)}\right) \\ & \leq 2 \exp\left(-\frac{t^2}{3\kappa^2 B^2}\right). \end{aligned}$$

Lemma A.3. *Suppose that X is distributed as $N_n(0, I)$ and is independent of Z . Assume that $\Phi(z) = (\Phi(z))^T$ for all $z \in \mathcal{Z}$. Then for all L such that $L\kappa A < 1/4$ and for all $0 < t < 2L\kappa^2 B^2(1+4L\kappa A)$,*

$$\begin{aligned} & \mathbb{P}(|X^T \Phi(Z)X - \text{tr}(\Phi(Z))| > t, Z \in C(\kappa, A, B)) \\ & \leq 2 \exp\left(-\frac{t^2}{4\kappa^2 B^2(1+4L\kappa A)}\right) \\ & \leq 2 \exp\left(-\frac{t^2}{8\kappa^2 B^2}\right). \end{aligned}$$

Proof of Lemma A.2. Let us denote the eigenvalues of $\Phi(z)\Phi(z)^T$ by $a_1 \geq \dots \geq a_n \geq 0$. Then, a_j 's are functions of z , but we do not make this explicit for notational simplicity. Note that, for $z \in C(\kappa, a, b)$, we have

$$\|a\|_\infty := \max_{1 \leq j \leq n} a_j \leq \kappa^2 A^2 \quad \text{and} \quad \sum_{j=1}^n a_j \leq \kappa^2 B^2. \tag{A.2}$$

By Chebyshev’s inequality, for any $\lambda > 0$, $t > 0$, and for all $z \in \mathcal{Z}$ we have

$$\mathbb{P}(X^T \Phi(z) Y > t) \leq e^{-\lambda t} \mathbb{E} \left(e^{\lambda X^T \Phi(z) Y} \right) = e^{-\lambda t} \mathbb{E} \left(e^{\frac{\lambda^2}{2} X^T \Phi(z) (\Phi(z))^T X} \right), \tag{A.3}$$

where the second equality follows from the fact that $Y \sim N(0, I_n)$ and is independent of X . Now suppose that $\lambda < L$ with L satisfying $L\kappa A < 1/\sqrt{2}$, and $z \in C(\kappa, A, B)$. Then, using the fact that X has i.i.d. $N(0, 1)$ coordinates, we can express the RHS of (A.3) as

$$\begin{aligned} & \exp \left(-\lambda t - \frac{1}{2} \sum_{j=1}^n \log(1 - \lambda^2 a_j) \right) \\ = & \exp \left(-\lambda t + \frac{\lambda^2}{2} \sum_{j=1}^n a_j + \frac{1}{2} \sum_{k=2}^{\infty} \frac{1}{k} \lambda^{2k} \sum_{j=1}^n a_j^k \right) \\ \leq & \exp \left(-\lambda t + \frac{1}{2} \left(\lambda^2 \sum_{j=1}^n a_j \right) \left(1 + \frac{1}{2} \lambda^2 \|a\|_{\infty} \sum_{k=2}^{\infty} \lambda^{2(k-2)} \|a\|_{\infty}^{k-2} \right) \right) \\ \leq & \exp \left(-\lambda t + \frac{\lambda^2}{2} \kappa^2 B^2 (1 + L^2 \kappa^2 A^2) \right) \end{aligned} \tag{A.4}$$

by (A.2) and the fact that $L^2 \kappa^2 A^2 < 1/2$. Treating the last term as a function of λ , we notice that it is maximized at

$$\lambda_*(t) = \frac{t}{\kappa^2 B^2 (1 + L^2 \kappa^2 A^2)}.$$

Notice that, for all $t \in (0, L\kappa^2 B^2 (1 + L^2 \kappa^2 A^2))$ we have $\lambda_*(t) < L$. This shows that for all $t \in (0, L\kappa^2 B^2 (1 + L^2 \kappa^2 A^2))$, we have

$$\begin{aligned} & \exp \left(-\lambda t - \frac{1}{2} \sum_{j=1}^n \log(1 - \lambda^2 a_j) \right) \\ \leq & \exp \left(-\lambda_*(t) t + \frac{1}{2} (\lambda_*(t))^2 \kappa^2 B^2 (1 + L^2 \kappa^2 A^2) \right) \\ = & \exp \left(-\frac{t^2}{2\kappa^2 B^2 (1 + L^2 \kappa^2 A^2)} \right) \leq \exp \left(-\frac{t^2}{3\kappa^2 B^2} \right). \end{aligned} \tag{A.5}$$

Thus, we conclude the proof of Lemma A.2 by combining (A.3), (A.4) and (A.5), and using symmetry, and the independence of (X, Y) and Z . \square

Proof of Lemma A.3. Let us denote the eigenvalues of $\Phi(z) = \Phi(z)^T$ by $b_1 \geq \dots \geq b_n$. For $z \in C(\kappa, A, B)$, we have,

$$\|b\|_{\infty} := \max_{1 \leq j \leq n} |b_j| \leq \kappa A \quad \text{and} \quad \sum_{j=1}^n b_j^2 \leq \kappa^2 B^2. \tag{A.6}$$

By Chebyshev’s inequality, for any $\lambda > 0$, $t > 0$, and for all $z \in \mathcal{Z}$ we have

$$\mathbb{P}(X^T \Phi(z) X - \text{tr}(\Phi(z)) > t) \leq e^{-\lambda(t + \text{tr}(\Phi(z)))} \mathbb{E} \left(e^{\lambda X^T \Phi(z) X} \right). \tag{A.7}$$

Suppose that $\lambda < L$ with L satisfying $L\kappa A < 1/\sqrt{2}$, and $z \in C(\kappa, A, B)$. Then, using the fact that X has i.i.d. $N(0, 1)$ coordinates, we can express the RHS of (A.7) as

$$\begin{aligned} & \exp \left(-\lambda \left(t + \sum_{j=1}^n b_j \right) - \frac{1}{2} \sum_{j=1}^n \log(1 - 2\lambda b_j) \right) \\ = & \exp \left(-\lambda \left(t + \sum_{j=1}^n b_j \right) + \lambda \sum_{j=1}^n b_j + \lambda^2 \sum_{j=1}^n b_j^2 + \frac{1}{2} \sum_{k=3}^{\infty} \frac{1}{k} (2\lambda)^k \sum_{j=1}^n b_j^k \right) \\ \leq & \exp \left(-\lambda t + \left(\lambda^2 \sum_{j=1}^n b_j^2 \right) \left(1 + \sum_{k=3}^{\infty} \frac{2}{k} (2L)^{k-2} \|b\|_{\infty}^{k-2} \right) \right) \\ \leq & \exp(-\lambda t + \lambda^2 \kappa^2 B^2 (1 + 4L\kappa A)) \end{aligned} \tag{A.8}$$

by (A.6) and the fact that $2L\kappa A \leq 1/2$. Treating the last term as a function of λ , we notice that it is maximized at

$$\lambda_*(t) = \frac{t}{2\kappa^2 B^2 (1 + 4L\kappa A)}.$$

Since for all $t \in (0, 2L\kappa^2 B^2 (1 + 4L\kappa A))$ we have $\lambda_*(t) < L$, it implies that for all $t \in (0, 2L\kappa^2 B^2 (1 + 4L\kappa A))$, we have

$$\begin{aligned} & \exp \left(-\lambda \left(t + \sum_{j=1}^n b_j \right) - \frac{1}{2} \sum_{j=1}^n \log(1 - 2\lambda b_j) \right) \\ \leq & \exp(-\lambda_*(t)t + \lambda_*(t)\kappa^2 B^2 (1 + 4L\kappa A)) \\ = & \exp \left(-\frac{t^2}{4\kappa^2 B^2 (1 + 4L\kappa A)} \right) \leq \exp \left(-\frac{t^2}{8\kappa^2 B^2} \right). \end{aligned} \tag{A.9}$$

Combining (A.7), (A.8) and (A.9) we get a bound for $\mathbb{P}(X^T \Phi(z) X - \text{tr}(\Phi(z)) > t)$ of the form given by the second line of (A.9). The proof of Lemma A.3 is completed by following analogous steps to get an identical bound for $\mathbb{P}(X^T \Phi(z) X - \text{tr}(\Phi(z)) < -t)$. \square

Computation of conditional mixed moments

In order to calculate the bias and variance of the proposed estimator, we need to compute the conditional expectations $\mathbb{E}(Y_{i_1 j_1} Y_{i_1 j'_1} Y_{i_2 j_2} Y_{i_2 j'_2} | \mathbf{T}_{i_1}, \mathbf{T}_{i_2})$ for various choices of $i_1, i_2, j_1, j'_1, j_2, j'_2$. We shall use the following well-known result, which is a special case of *Wick formula* ([20], p. 129), for computation of mixed moments of a Gaussian random vector.

Lemma A.4. *If W_1, W_2, W_3 and W_4 are jointly Gaussian with mean zero and covariance matrix Σ , then*

$$\mathbb{E}(W_1 W_2 W_3 W_4) = \Sigma_{12} \Sigma_{34} + \Sigma_{13} \Sigma_{24} + \Sigma_{14} \Sigma_{23}. \quad (\text{A.10})$$

We shall use the formula to compute the above mixed moments with the observation that $\text{Cov}(X_{i_1 j_1}, X_{i_2 j_2} | \mathbf{T}_{i_1}, \mathbf{T}_{i_2}) = \rho_{i_1 i_2} C(T_{i_1 j_1}, T_{i_2 j_2})$. The details of this computation in various generic cases are given in Appendix F of [22].

Appendix B

In the following, we shall often write h and \tilde{h} to denote h_n and \tilde{h}_n , respectively, and we shall drop the subscript h_n from the covariance estimates. For example, \hat{C} will be used to denote \hat{C}_{h_n} .

Proof of Proposition 2.1

By elementary calculations, and supposing that $m_i \geq 2$ for each $1 \leq i \leq n$, we have

$$\begin{aligned} & \mathbb{E}[\tilde{X}_i(s) \tilde{X}_i(t)] \\ &= \frac{1}{m_i^2} \sum_{j, j'=1}^{m_i} \mathbb{E} \left[Y_{ij} Y_{ij'} \frac{1}{h_n^2} K \left(\frac{s - T_{ij}}{h_n} \right) K \left(\frac{s - T_{ij'}}{h_n} \right) \right] \\ &= \frac{m_i}{m_i^2} \frac{1}{h_n^2} \int (C(u, u) + \sigma^2) K \left(\frac{s - u}{h_n} \right) K \left(\frac{t - u}{h_n} \right) du + \\ & \quad \frac{m_i(m_i - 1)}{m_i^2} \frac{1}{h_n^2} \int \int C(u, v) K \left(\frac{s - u}{h_n} \right) K \left(\frac{t - u}{h_n} \right) dudv \\ &= \frac{1}{m_i} \frac{1}{h_n} \int (C(t + h_n u, t + h_n u) + \sigma^2) K(-u) K \left(\frac{s - t}{h_n} - u \right) du \\ & \quad + \frac{m_i - 1}{m_i} \int \int C(s + h_n u, t + h_n v) K(-u) K(-v) dudv \\ &= \frac{1}{m_i h_n} \left[(\bar{C}(t) + \sigma^2) \int K(-u) K \left(\frac{s - t}{h_n} - u \right) du \right. \\ & \quad \left. + h_n \bar{C}'(t) \int u K(-u) K \left(\frac{s - t}{h_n} - u \right) du + O(h_n^2) \right] \\ & \quad + \left(1 - \frac{1}{m_i} \right) C(s, t) \int \int K(-u) K(-v) dudv \\ & \quad + \left(1 - \frac{1}{m_i} \right) h_n \int \int [C_s(s, t)u + C_t(s, t)v] K(-u) K(-v) dudv + O(h_n^2), \end{aligned} \quad (\text{B.1})$$

where the last step is by Taylor series expansions. Now, noticing that K is symmetric about 0, $\int K(x) dx = 1$ and $\int xK(x) dx = 0$, (2.3) and (2.4) follow from (B.1) after simplifications.

Asymptotic pointwise bias (3.1)

We first compute the expected value of the estimate described by (2.9). For simplicity of notations, we express $\tilde{X}_i(s_l) + (s - s_l)\tilde{X}'_i(s_l)$ by $\tilde{X}_{i,l}(s)$. Observe that

$$\tilde{X}_{i,l}(s) = \frac{1}{m_i} \sum_{j=1}^{m_i} Y_{ij} \frac{1}{h_n} \left[K \left(\frac{s_l - T_{ij}}{h_n} \right) + \frac{s - s_l}{h_n} K' \left(\frac{s_l - T_{ij}}{h_n} \right) \right]$$

Let the support of kernel $K(\cdot)$ be denoted by $[-B_K, B_K]$. Then, for each fixed $j = 1, \dots, m_i$, and $i = 1, \dots, n$,

$$\begin{aligned} & \mathbb{E} \left[Y_{ij}^2 \left\{ \left(K \left(\frac{s_l - T_{ij}}{h_n} \right) + \frac{s - s_l}{h_n} K' \left(\frac{s_l - T_{ij}}{h_n} \right) \right) \right. \right. \\ & \quad \cdot \left. \left. \left(K \left(\frac{s_{l'} - T_{ij}}{h_n} \right) + \frac{s - s_{l'}}{h_n} K' \left(\frac{s_{l'} - T_{ij}}{h_n} \right) \right) \right\} \right] \\ &= \int [C(u, u) + \sigma^2] g^2(u) \left[\left(K \left(\frac{s_l - u}{h_n} \right) + \frac{s - s_l}{h_n} K' \left(\frac{s_l - u}{h_n} \right) \right) \right. \\ & \quad \cdot \left. \left(K \left(\frac{s_{l'} - u}{h_n} \right) + \frac{s - s_{l'}}{h_n} K' \left(\frac{s_{l'} - u}{h_n} \right) \right) \right] du, \end{aligned} \tag{B.2}$$

which is 0, if $|s_l - s_{l'}| > 2B_K h_n$, since this implies that $K(\frac{s_l - u}{h_n})K(\frac{s_{l'} - u}{h_n}) = 0$ for all $u \in \mathbb{R}$. If $|s_l - s_{l'}| \leq 2B_K h_n$, there is nonzero contribution of the term (B.2) in $\mathbb{E}[\tilde{X}_{i,l}(s)\tilde{X}_{i,l'}(t)]\bar{Q}_{h_n}(s - s_l)\bar{Q}_{h_n}(t - s_{l'})$ only if $|s - t| \leq 2(B_K + C_Q)h_n$, where $\text{supp}(\bar{Q}) = [-C_Q, C_Q]$. Thus, if $A > 4(B_K + C_Q)$, then for $|s - t| > Ah_n/2$, we have

$$\begin{aligned} & w(m_i)\mathbb{E}(\tilde{X}_{i,l}(s)\tilde{X}_{i,l'}(t)) \\ &= \frac{1}{h_n^2} \int \int C(u, v)g(u)g(v) \left[\left(K \left(\frac{s_l - u}{h_n} \right) + \frac{s - s_l}{h_n} K' \left(\frac{s_l - u}{h_n} \right) \right) \right. \\ & \quad \left. \left(K \left(\frac{s_{l'} - v}{h_n} \right) + \frac{s - s_{l'}}{h_n} K' \left(\frac{s_{l'} - v}{h_n} \right) \right) \right] dudv \\ &= \int \int C(s_l + xh_n, s_{l'} + yh_n)g(s_l + xh_n)g(s_{l'} + yh_n) \\ & \quad \cdot \left[\left(K(x) + \frac{s - s_l}{h_n} K'(-x) \right) \left(K(y) + \frac{s - s_{l'}}{h_n} K'(-y) \right) \right] dx dy. \end{aligned} \tag{B.3}$$

We assume that the conditions in Section 3 hold. Then using the representation (B.3), and the calculations done in Appendix F of [22], we get an expression for the asymptotic bias in estimating $C(s, t)$ as a function of the bandwidth h_n . These results are summarized in the following lemmas, where C_s, C_{ss} and C_t, C_{tt} denote the first and second partial derivatives of $C(s, t)$ with respect to s and t , respectively.

Lemma B.1 (Expectation of $\tilde{C}(s, t)$). *Let $K_2 = \int x^2 K(x) dx$,*

$$\mathbf{Q}_{h_n}(s) = \sum_{l=1}^{L_n} \bar{Q}_{h_n}(s - s_l), \quad \text{and} \quad \mathbf{Q}_{h_n}^{(2)}(s) = \sum_{l=1}^{L_n} \left(\frac{s - s_l}{h_n}\right)^2 \bar{Q}_{h_n}(s - s_l).$$

Then, for $|s - t| > 2Ah_n$,

$$\begin{aligned} \mathbb{E}\tilde{C}(s, t) &= C(s, t)\mathbf{Q}_{h_n}(s)\mathbf{Q}_{h_n}(t) \\ &+ \frac{h_n^2}{2}C(s, t) \left[\frac{g''(s)}{g(s)}(K_2\mathbf{Q}_{h_n}(s) - \mathbf{Q}_{h_n}^{(2)}(s))\mathbf{Q}_{h_n}(t) \right. \\ &\quad \left. + \frac{g''(t)}{g(t)}(K_2\mathbf{Q}_{h_n}(t) - \mathbf{Q}_{h_n}^{(2)}(t))\mathbf{Q}_{h_n}(s) \right] \\ &+ h_n^2 C_s \frac{g'(s)}{g(s)}(K_2\mathbf{Q}_{h_n}(s) - \mathbf{Q}_{h_n}^{(2)}(s))\mathbf{Q}_{h_n}(t) \\ &+ h_n^2 C_t \frac{g'(t)}{g(t)}(K_2\mathbf{Q}_{h_n}(t) - \mathbf{Q}_{h_n}^{(2)}(t))\mathbf{Q}_{h_n}(s) \\ &+ \frac{h_n^2}{2} \frac{1}{g(s)g(t)} \left[C_{ss}(K_2\mathbf{Q}_{h_n}(s) - \mathbf{Q}_{h_n}^{(2)}(s))\mathbf{Q}_{h_n}(t) \right. \\ &\quad \left. + C_{tt}(K_2\mathbf{Q}_{h_n}(t) - \mathbf{Q}_{h_n}^{(2)}(t))\mathbf{Q}_{h_n}(s) \right] \\ &+ O(h_n^{2+\alpha}). \end{aligned} \tag{B.4}$$

Note that because of property (iii) of the kernel \bar{Q} , and the fact that $s_l = (l+a)h_n$ for $l = 1, \dots, L_n$, for some constant $a \in [-3, 3]$, we have for $s \in (c, 1-c)$, for some $c \in (0, 1)$,

$$\mathbf{Q}_{h_n}(s) = \sum_{l=1}^{L_n} \bar{Q}\left(\frac{s}{h_n} - a - l\right) = 1.$$

Therefore, we can choose L_n and the sequence of points $\{s_l\}_{l=1}^{L_n}$ so that $L_n \approx h_n^{-1}$, and $\mathbf{Q}_h(s) \equiv 1$ for all $s \in [0, 1]$. That is, from Lemma B.1, we have $\mathbb{E}\tilde{C}(s, t) = C(s, t) + O(h_n^2)$.

Lemma B.2 (Expectation of $\widehat{C}_*(t)$). *Let $\bar{C}'(t)$ and $\bar{C}''(t)$ denote the first and second derivative of the function $\bar{C}(t) := C(t, t)$. Then, uniformly in t ,*

$$\begin{aligned} \mathbb{E}\widehat{C}_*(t) &= (\bar{C}(t) + \sigma^2)\mathbf{Q}_{h_n}(t) \\ &+ \frac{h_n^2}{2}(\bar{C}(t) + \sigma^2) \left(\frac{g''(t)}{g(t)}\right) (K_2\mathbf{Q}_{h_n}(t) - \mathbf{Q}_{h_n}^{(2)}(t)) \\ &+ h_n^2 \bar{C}'(t) \left(\frac{g'(t)}{g(t)}\right) (K_2\mathbf{Q}_{h_n}(t) - \mathbf{Q}_{h_n}^{(2)}(t)) \\ &+ \frac{h_n^2}{2} \frac{\bar{C}''(t)}{g(t)} (K_2\mathbf{Q}_{h_n}(t) - \mathbf{Q}_{h_n}^{(2)}(t)) + O(h_n^{2+\alpha}). \end{aligned} \tag{B.5}$$

Proof of Lemma B.2 follows along the lines of Lemma B.1. Furthermore, if an estimator $\hat{\sigma}^2$ is such that $\mathbb{E}\hat{\sigma}^2 = \sigma^2 + O(h_n^2)$, then it follows from Lemma B.2 that the estimator $\widehat{C}(t) := \widehat{C}_*(t) - \hat{\sigma}^2$ satisfies

$$\mathbb{E}\widehat{C}(t) = C(t) + O(h_n^2), \tag{B.6}$$

uniformly on $t \in [0, 1]$, since $\mathbf{Q}_{h_n}(t) \equiv 1$ on $t \in [0, 1]$. Next, since $C(s, t) = C(t, s)$ and $C(s, t)$ is smooth, it follows that $C_s - C_t \equiv 0$. Consequently, using a Taylor series expansion, it follows that, for any $A > 0$,

$$C(s, t) = \overline{C}\left(\frac{s+t}{2}\right) + O(h_n^2), \quad \text{for } |s-t| \leq \frac{Ah_n}{2}. \tag{B.7}$$

Combining (B.6) and (B.7) we get,

$$\mathbb{E}\widehat{C}\left(\frac{s+t}{2}\right) = C(s, t) + O(h_n^2), \quad \text{for } |s-t| \leq \frac{Ah_n}{2}, \quad s, t \in [0, 1]. \tag{B.8}$$

Appendix C

Proof of Proposition 4.1

We start by recalling the definition of the operator kernel H_ν . The first step is to express $\widehat{C}(s, t)$ as $\widetilde{C}_c(s, t) - W_{\widetilde{h}_n}(s, t)(\hat{\sigma}^2 - \sigma^2)$, where

$$\widetilde{C}_c(s, t) = \overline{W}_{\widetilde{h}_n}(s, t)\widetilde{C}(s, t) + W_{\widetilde{h}_n}(s, t)\left(\widehat{C}_*\left(\frac{s+t}{2}\right) - \sigma^2\right). \tag{C.1}$$

Therefore, in order to separate the effect of estimating σ^2 , use the fact that for any fixed $\epsilon > 0$,

$$\begin{aligned} & \|H_\nu \widehat{C} \psi_\nu\|_2^2 \\ & \leq (1 + \epsilon) \|H_\nu \widetilde{C}_c \psi_\nu\|_2^2 + \left(1 + \frac{1}{\epsilon}\right) (\hat{\sigma}^2 - \sigma^2)^2 \|H_\nu W_{\widetilde{h}_n} \psi_\nu\|_2^2 \\ & = (1 + \epsilon) \|H_\nu \widetilde{C}_c \psi_\nu\|_2^2 + \left(1 + \frac{1}{\epsilon}\right) (\hat{\sigma}^2 - \sigma^2)^2 O(h_n^4). \end{aligned} \tag{C.2}$$

The equality follows since using $H_\nu \psi_\nu = 0$, the definition of $W_{\widetilde{h}_n}$, and the *Mean Value Theorem*, we have

$$\begin{aligned} |(H_\nu W_{\widetilde{h}_n} \psi_\nu)(x)| & = \left| \int H_\nu(x, s) \int_{(s+Ah_n/2) \vee 0}^{(s-Ah_n/2) \wedge 1} (\psi_\nu(t) - \psi_\nu(s)) dt ds \right| \\ & \leq \frac{A^2 h_n^2}{2} \|\psi'_\nu\|_\infty \left[\int |\overline{H}_\nu(x, s)| ds + \frac{1}{\lambda_\nu} \right]. \end{aligned}$$

Since $\mathbb{E}(\hat{\sigma}^2 - \sigma^2)^2 = o(1)$, it is enough to show that $\mathbb{E} \| H_\nu \tilde{C}_c \psi_\nu \|_2^2$ has the bound given by the RHS of (4.2), without the multiplicative factor $(1 + \epsilon)$. With a slight abuse of notation, we write $\widehat{C}(s)$ to indicate $\widehat{C}_*(s) - \sigma^2$. Then, since

$$\begin{aligned} (H_\nu \tilde{C}_c \psi_\nu)(x) &= \int \int H_\nu(x, s) \overline{W}_{\tilde{h}_n}(s, t) \tilde{C}(s, t) \psi_\nu(t) ds dt \\ &\quad + \int \int H_\nu(x, s) W_{\tilde{h}_n}(s, t) \widehat{C}\left(\frac{s+t}{2}\right) \psi_\nu(t) ds dt, \end{aligned}$$

it follows that, $\| H_\nu \tilde{C}_c \psi_\nu \|_2^2$ equals

$$\begin{aligned} &\int \int \int \int \int H_\nu(x, s_1) H_\nu(x, s_2) \overline{W}_{\tilde{h}_n}(s_1, t_1) \overline{W}_{\tilde{h}_n}(s_2, t_2) \\ &\quad \cdot \tilde{C}(s_1, t_1) \tilde{C}(s_2, t_2) \psi_\nu(t_1) \psi_\nu(t_2) ds_1 ds_2 dt_1 dt_2 dx \\ &+ \int \int \int \int \int H_\nu(x, s_1) H_\nu(x, s_2) W_{\tilde{h}_n}(s_1, t_1) W_{\tilde{h}_n}(s_2, t_2) \\ &\quad \cdot \widehat{C}\left(\frac{s_1+t_1}{2}\right) \widehat{C}\left(\frac{s_2+t_2}{2}\right) \psi_\nu(t_1) \psi_\nu(t_2) ds_1 ds_2 dt_1 dt_2 dx \\ &+ \int \int \int \int \int H_\nu(x, s_1) H_\nu(x, s_2) W_{\tilde{h}_n}(s_1, t_1) \overline{W}_{\tilde{h}_n}(s_2, t_2) \\ &\quad \cdot \tilde{C}(s_1, t_1) \widehat{C}\left(\frac{s_2+t_2}{2}\right) \psi_\nu(t_1) \psi_\nu(t_2) ds_1 ds_2 dt_1 dt_2 dx. \end{aligned} \quad (\text{C.3})$$

Thus, in order to obtain $\mathbb{E} \| H_\nu \tilde{C}_c \psi_\nu \|_2^2$, we need to evaluate $\mathbb{E}[\tilde{C}(s_1, t_1) \tilde{C}(s_2, t_2)]$, $\mathbb{E}[\widehat{C}(\frac{s_1+t_1}{2}) \widehat{C}(\frac{s_2+t_2}{2})]$, and $\mathbb{E}[\tilde{C}(s_1, t_1) \widehat{C}(\frac{s_2+t_2}{2})]$.

Let

$$U_i(s, t) = \sum_{l, l'=1}^{L_n} \frac{1}{m_i^2} \sum_{j, j'=1}^{m_i} Y_{ij} Y_{ij'} \tilde{K}_{s,l}(T_{ij}) \tilde{K}_{s,l}(T_{ij'}) \overline{Q}_h(s - s_l) \overline{Q}_h(t - s_{l'}), \quad (\text{C.4})$$

where $\tilde{K}_{s,l}(\cdot)$ is defined by

$$\tilde{K}_{s,l}(u) = \frac{1}{h} \left[K\left(\frac{s_l - u}{h}\right) + \left(\frac{s - s_l}{h}\right) K'\left(\frac{s_l - u}{h_n}\right) \right], \quad (\text{C.5})$$

for $s \in [0, 1]$ and $l = 1, \dots, L_n$. Then we can express the expectation of the first term on the RHS of (C.3) as

$$\begin{aligned} &\frac{1}{n^2} \sum_{i=1}^n w^2(m_i) \int \int \int \int \int H_\nu(x, s_1) H_\nu(x, s_2) \overline{W}_{\tilde{h}_n}(s_1, t_1) \overline{W}_{\tilde{h}_n}(s_2, t_2) \\ &\quad \cdot [g(s_1)g(s_2)g(t_1)g(t_2)]^{-1} \mathbb{E}[U_i(s_1, t_1) U_i(s_2, t_2)] \psi_\nu(t_1) \psi_\nu(t_2) ds_1 ds_2 dt_1 dt_2 dx \\ &+ \frac{1}{n^2} \sum_{i_1 \neq i_2}^n w(m_{i_1}) w(m_{i_2}) \int \int \int \int \int H_\nu(x, s_1) H_\nu(x, s_2) \overline{W}_{\tilde{h}_n}(s_1, t_1) \overline{W}_{\tilde{h}_n}(s_2, t_2) \\ &\quad \cdot [g(s_1)g(s_2)g(t_1)g(t_2)]^{-1} \mathbb{E}[U_{i_1}(s_1, t_1) U_{i_2}(s_2, t_2)] \psi_\nu(t_1) \psi_\nu(t_2) ds_1 ds_2 dt_1 dt_2 dx. \end{aligned} \quad (\text{C.6})$$

The following proposition is the key to get a simplified bound on (C.6). It is proved using a lengthy, but fairly straightforward calculation, the details of which are given in Appendix F of [22].

Proposition C.1. *Suppose that $A > 4(B_K + C_Q)$. Then for $|s_k - t_k| > \frac{1}{2}Ah_n$ ($k = 1, 2$), we have*

$$\begin{aligned} & \frac{1}{n^2} \sum_{i=1}^n w^2(m_i) \frac{\mathbb{E}[U_i(s_1, t_1)U_i(s_2, t_2)]}{g(s_1)g(s_2)g(t_1)g(t_2)} \\ = & \frac{1}{n^2} \sum_{i=1}^n \frac{(m_i - 2)(m_i - 3)}{m_i(m_i - 1)} [(C(s_1, t_1) + O(h_n^2))(C(s_2, t_2) + O(h_n^2)) \\ & + (C(s_1, s_2) + O(h_n^2))(C(t_1, t_2) + O(h_n^2)) \\ & + (C(s_1, t_2) + O(h_n^2))(C(s_2, t_1) + O(h_n^2))] \\ & + Z_1 + Z_2 + Z_3 + Z_4 + Z_5 + Z_6, \end{aligned} \tag{C.7}$$

where the quantities $Z_j := Z_j(s_1, s_2, t_1, t_2)$, $j = 1, \dots, 6$. Here, Z_1, \dots, Z_4 are asymptotically equivalent to $Z(s_1, s_2)$, $Z(s_1, t_2)$, $Z(t_1, s_2)$ and $Z(t_1, t_2)$, respectively; and Z_5 , and Z_6 are asymptotically equivalent to $\tilde{Z}(s_1, s_2, t_1, t_2)$ and $\tilde{Z}(s_1, t_2, t_1, s_2)$, respectively, where

$$Z(s, t) = \begin{cases} O\left(\frac{1}{nh_n \underline{m}_n}\right) & \text{if } |s - t| \leq \frac{Ah_n}{2} \\ 0 & \text{otherwise;} \end{cases}$$

and

$$\tilde{Z}(s_1, s_2, t_1, t_2) = \begin{cases} O\left(\frac{1}{nh_n^2 \underline{m}_n^2}\right) & \text{if } \max\{|s_1 - s_2|, |t_1 - t_2|\} \leq \frac{Ah_n}{2} \\ O\left(\frac{1}{nh_n \underline{m}_n^2}\right) & \text{if } |s_1 - s_2| \leq \frac{Ah_n}{2} \text{ and } |t_1 - t_2| > \frac{Ah_n}{2} \\ O\left(\frac{1}{nh_n \underline{m}_n^2}\right) & \text{if } |s_1 - s_2| > \frac{Ah_n}{2} \text{ and } |t_1 - t_2| \leq \frac{Ah_n}{2} \\ 0 & \text{otherwise.} \end{cases}$$

Also,

$$\begin{aligned} & \frac{1}{n^2} \sum_{i_1 \neq i_2}^n w(m_{i_1})w(m_{i_2}) \frac{\mathbb{E}[U_{i_1}(s_1, t_1)U_{i_2}(s_2, t_2)]}{g(s_1)g(s_2)g(t_1)g(t_2)} \\ = & \frac{n-1}{n} (C(s_1, t_1) + O(h_n^2))(C(s_2, t_2) + O(h_n^2)) \\ & + \frac{1}{n^2} \left(\sum_{i_1 \neq i_2}^n \rho_{i_1 i_2}^2 \right) [(C(s_1, s_2) + O(h_n^2))(C(t_1, t_2) + O(h_n^2)) \\ & + (C(s_1, t_2) + O(h_n^2))(C(s_2, t_1) + O(h_n^2))]. \end{aligned} \tag{C.8}$$

In all of above the $O(\cdot)$ terms are uniform in s_1, s_2, t_1, t_2 in their respective domains.

Now we deal with the last two terms on the RHS of (C.3). Let

$$V_i(s) = \sum_{l=1}^{L_n} \frac{1}{m_i} \sum_{j=1}^{m_i} Y_{ij}^2 \tilde{K}_{s,l}(T_{ij}). \quad (\text{C.9})$$

Then,

$$\widehat{C}_*(s) = \frac{1}{n} \sum_{i=1}^n [g(s)]^{-1} V_i(s) \overline{Q}_{h_n}(s - s_i).$$

For convenience, in the rest of this subsection we shall use z_k to denote $(s_k + t_k)/2$, for $k = 1, 2$. Then the following proposition describes the contribution of the quantities of the type $\mathbb{E}[V_{i_1}(z_1)V_{i_2}(z_2)]$ and $\mathbb{E}[U_{i_1}(s_1, t_1)V_{i_2}(z_2)]$.

Proposition C.2. *Suppose that $A > 4(B_K + C_Q)$. Then for (i) $|s_k - t_k| \leq \frac{Ah_n}{2}$, $k = 1, 2$,*

$$\begin{aligned} & \frac{1}{n^2} \sum_{i=1}^n \frac{\mathbb{E}(V_i(z_1)V_i(z_2))}{g(z_1)g(z_2)} \\ & + \frac{1}{n^2} \sum_{i_1 \neq i_2}^n \frac{\mathbb{E}(V_{i_1}(z_1)V_{i_2}(z_2))}{g(z_1)g(z_2)} - \sigma^2 [\mathbb{E}(\widehat{C}_*(z_1)) + \mathbb{E}(\widehat{C}_*(z_2))] + \sigma^4 \\ = & C(s_1, t_1)C(s_2, t_2) - \left(\frac{1}{n^2} \sum_{i=1}^n \frac{1}{m_i} \right) (C(s_1, t_1) + \sigma^2)(C(s_2, t_2) + \sigma^2) + O(h_n^2) \\ & + \left(\frac{1}{n} \left(1 - \frac{1}{n} \sum_{i=1}^n \frac{1}{m_i} \right) + \frac{1}{n^2} \sum_{i_1 \neq i_2}^n \rho_{i_1 i_2}^2 \right) \\ & \cdot (C(s_1, s_2)C(t_1, t_2) + C(s_1, t_2)C(s_2, t_1) + O(h_n)) + Z_7, \end{aligned} \quad (\text{C.10})$$

where $Z_7 := Z_7(z_1, z_2)$ is asymptotically equivalent to $Z(z_1, z_2)$. Next, if (ii) $|s_1 - t_1| > \frac{Ah_n}{2}$ and $|s_2 - t_2| \leq \frac{Ah_n}{2}$, then

$$\begin{aligned} & \frac{1}{n^2} \sum_{i=1}^n w(m_i) \mathbb{E}(U_i(s_1, t_1)V_i(z_2)) \\ & + \frac{1}{n^2} \sum_{i_1 \neq i_2} w(m_{i_1}) \mathbb{E}(U_{i_1}(s_1, t_1)V_{i_2}(z_2)) - \sigma^2 \mathbb{E}\tilde{C}(s_1, t_1) \\ = & (C(s_1, t_1) + O(h_n^2))(C(s_2, t_2) + O(h_n^2)) \\ & - \left(\frac{1}{n^2} \sum_{i=1}^n \frac{2}{m_i} \right) (C(s_1, t_1) + O(h_n^2))(C(s_2, t_2) + \sigma^2 + O(h_n^2)) \\ & + \left(\frac{1}{n} \left(1 - \frac{1}{n} \sum_{i=1}^n \frac{2}{m_i} \right) + \frac{1}{n^2} \sum_{i_1 \neq i_2}^n \rho_{i_1 i_2}^2 \right) \\ & \cdot (C(s_1, s_2)C(t_1, t_2) + C(s_1, t_2)C(s_2, t_1) + O(h_n)) + Z_8 + Z_9, \end{aligned} \quad (\text{C.11})$$

where the $O(h_n^2)$ terms within brackets in the first term on the RHS depend on (s_1, t_1) and (s_2, t_2) respectively, and $Z_j := Z_j(s_1, t_1, z_2)$, $j = 8, 9$ satisfy

$$Z_8 = \begin{cases} O\left(\frac{1}{nh_n^2 \underline{m}_n}\right) & \text{if } |s_1 - s_2| \leq \frac{Ah_n}{2} \\ 0 & \text{otherwise;} \end{cases}$$

$$Z_9 = \begin{cases} O\left(\frac{1}{nh_n^2 \underline{m}_n}\right) & \text{if } |t_1 - s_2| \leq \frac{Ah_n}{2} \\ 0 & \text{otherwise.} \end{cases}$$

The proof of Proposition 4.1 is now finished by using the definitions of $\mathbb{E}[\tilde{C}(s_1, t_1)\tilde{C}(s_2, t_2)]$, $\mathbb{E}[\widehat{C}(\frac{s_1+t_1}{2})\widehat{C}(\frac{s_2+t_2}{2})]$, and $\mathbb{E}[\tilde{C}(s_1, t_1)\widehat{C}(\frac{s_2+t_2}{2})]$; using the properties of the kernel $H_\nu(x, y)$; and the bounds in Propositions C.1 and C.2 and plugging everything back into the expectation of (C.3). For details, see Appendix E.

Appendix D

Asymptotic pointwise variance (3.2)

In this section, we prove (3.2), (3.9) and (3.10). Most of the derivations are similar to that of Proposition 4.1. Thus we simply give a brief outline.

First, using the fact that $\overline{W}_{\tilde{h}_n}(s, t)W_{\tilde{h}_n}(s, t) = 0$, we obtain

$$\begin{aligned} & \text{Var}(\widehat{C}(s, t)) \\ &= \overline{W}_{\tilde{h}_n}(s, t)\text{Var}(\tilde{C}(s, t)) + W_{\tilde{h}_n}(s, t)\text{Var}\left(\widehat{C}_*\left(\frac{s+t}{2}\right) - \widehat{\sigma}^2\right) \\ &\leq \overline{W}_{\tilde{h}_n}(s, t)\text{Var}(\tilde{C}(s, t)) + 2W_{\tilde{h}_n}(s, t)\left[\text{Var}\left(\widehat{C}_*\left(\frac{s+t}{2}\right)\right) + \text{Var}(\widehat{\sigma}^2)\right]. \end{aligned}$$

Since $\mathbb{E}(\widehat{\sigma}^2 - \sigma^2)^2$ has the rate given by (3.9) (Corollary 3.1), we only need to provide bounds for $\overline{W}_{\tilde{h}_n}(s, t)\text{Var}(\tilde{C}(s, t))$ and $W_{\tilde{h}_n}(s, t)\text{Var}(\widehat{C}_*(\frac{s+t}{2}))$. We state these in the following propositions which are proved in Appendix F of [22].

Proposition D.1.

$$\begin{aligned} & \overline{W}_{\tilde{h}_n}(s, t)\text{Var}(\tilde{C}(s, t)) \\ &= O\left(\frac{1}{n}\right) + \left(\frac{1}{n^2} \sum_{i \neq j}^n \rho_{ij}^2\right) O(1) + O\left(\max\left\{\frac{1}{nh_n^2 \underline{m}_n^2}, \frac{1}{nh_n \underline{m}_n}\right\}\right). \end{aligned} \tag{D.1}$$

Proposition D.2.

$$W_{\tilde{h}_n}(s, t)\text{Var}\left(\widehat{C}_*\left(\frac{s+t}{2}\right)\right) = O\left(\frac{1}{n}\right) + \left(\frac{1}{n^2} \sum_{i \neq j}^n \rho_{ij}^2\right) O(1) + O\left(\frac{1}{nh_n \underline{m}_n}\right). \tag{D.2}$$

The proof of (3.10) is finished by combining Propositions D.1 and D.2 and Corollary 3.1.

Proof of Corollary 3.1

First observe that,

$$\begin{aligned}
 & \mathbb{E}(\widehat{\sigma}^2 - \sigma^2)^2 \\
 &= \frac{1}{(T_1 - T_0)^2} \int_{T_0}^{T_1} \int_{T_0}^{T_1} \mathbb{E} \left[\left(\widehat{C}_*(t) - \sigma^2 - \widehat{C}_0(t) \right) \left(\widehat{C}_*(s) - \sigma^2 - \widehat{C}_0(s) \right) \right] ds dt \\
 &\leq \sup_{t \in [T_0, T_1]} \mathbb{E}(\widehat{C}_*(t) - \sigma^2 - \widehat{C}_0(t))^2 \quad (\text{by Cauchy-Schwarz inequality}) \\
 &\leq 2 \sup_{t \in [T_0, T_1]} \text{Var}(\widehat{C}_*(t)) + 2 \sup_{t \in [T_0, T_1]} \text{Var}(\widehat{C}_0(t)) \\
 &\quad + \sup_{t \in [T_0, T_1]} \left(\mathbb{E}(\widehat{C}_*(t)) - \sigma^2 - \mathbb{E}(\widehat{C}_0(t)) \right)^2. \tag{D.3}
 \end{aligned}$$

By Propositions D.1 and D.2, and the definition (2.10) of \widehat{C}_0 , the sum of the first two term on the RHS on (D.3) is bounded by

$$O\left(\frac{1}{n}\right) + \left(\frac{1}{n^2} \sum_{i \neq j}^n \rho_{ij}^2 \right) O(1) + O\left(\max \left\{ \frac{1}{nh_n^2 \underline{m}_n^2}, \frac{1}{nh_n \underline{m}_n} \right\} \right).$$

On the other hand, since for any bounded $u \in [A_1, A_2]$,

$$\left| \frac{1}{2} (C(t - h_n u, t + h_n u) + C(t + h_n u, t - h_n u)) - C(t, t) \right| = O(h_n^2),$$

uniformly in $t \in [T_0, T_1]$, it follows from Lemmas B.1 and B.2 (Appendix B) that the last term on the RHS of (D.3) is $O(h_n^4)$.

Proof of Proposition 4.2

Without loss of generality we assume g to be the uniform density on $[0, 1]$. For convenience we drop subscript n from h_n , \underline{m}_n and \overline{m}_n . We need to consider two cases separately: (i) $|s - t| > Ah/2$ and (ii) $|s - t| \leq Ah/2$.

First consider the case $|s - t| > Ah/2$. Then, we have $\widehat{C}(s, t) - \mathbb{E}[\widehat{C}(s, t)] = \overline{W}_{Ah}(s, t)(\widetilde{C}(s, t) - \mathbb{E}[\widetilde{C}(s, t)])$. Let

$$B(s, T_{ij}) = \sum_{l=1}^{L_n} \widetilde{K}_{s,l}(T_{ij}) \overline{Q}_h(s - s_l), \quad 1 \leq j \leq m_i, 1 \leq i \leq n.$$

Since $|\widetilde{K}_{s,l}(T_{ij})| = O(h^{-1})$ and the summands are nonzero for a bounded number of indices l , there exists a constant $C_3 > 0$ such that

$$\sup_{s \in [0, 1]} \max_{1 \leq i \leq n} \max_{1 \leq j \leq m_i} |B(s, T_{ij})| \leq C_3 h^{-1}. \tag{D.4}$$

Note further that $B(s, T_{ij}) = 0$ if $|s - T_{ij}| > 2(B_K + C_Q)h$. Next,

$$\begin{aligned} \sum_{l=1}^{L_n} \tilde{X}_{i,l}(s) &= \frac{1}{m_i} \sum_{j=1}^{m_i} (X_i(T_{ij}) + \sigma \varepsilon_{ij}) B(s, T_{ij}) \\ &= \sum_{k=1}^M \sqrt{\lambda_k} \xi_{ik} \left(\frac{1}{m_i} \sum_{j=1}^{m_i} \psi_k(T_{ij}) B(s, T_{ij}) \right) + \sigma \frac{1}{m_i} \sum_{j=1}^{m_i} \varepsilon_{ij} B(s, T_{ij}) \\ &= \sum_{k=1}^M \sqrt{\lambda_k} \xi_{ik} B_{i,k}(s) + \sigma \frac{1}{m_i} \sum_{j=1}^{m_i} \varepsilon_{ij} B(s, T_{ij}), \end{aligned}$$

where $B_{i,k}(s) := \frac{1}{m_i} \sum_{j=1}^{m_i} \psi_k(T_{ij}) B(s, T_{ij})$. By (D.4), there exists $C_4 > 0$ such that

$$\sup_{s \in [0,1]} \max_{1 \leq k \leq M} \max_{1 \leq i \leq n} |B_{i,k}(s)| \leq C_4 h^{-1}. \tag{D.5}$$

Also, since $A > 4(B_K + C_Q)h$ and since $|s - t| \geq \frac{Ah}{2}$, it follows that $B(s, T_{ij}) \times B(t, T_{ij}) = 0$. Furthermore, for each $k = 1, \dots, M$, $\{B_{i,k}(s)\}_{i=1}^n$ are independent, and these random variables are independent of $\{\xi_{ik} : 1 \leq k \leq M\}_{i=1}^n$ and $\{\varepsilon_i\}_{i=1}^n$ where $\varepsilon_i = (\varepsilon_{ij})_{j=1}^{m_i}$. Let $\mathbf{B}_i(s) = (B(s, T_{ij}))_{j=1}^{m_i}$. Then, we can express $\tilde{C}(s, t) - \mathbb{E}[\tilde{C}(s, t)]$ as,

$$\begin{aligned} &\tilde{C}(s, t) - \mathbb{E}[\tilde{C}(s, t)] \\ &= \sum_{1 \leq k \neq k' \leq M} \sqrt{\lambda_k \lambda_{k'}} \frac{1}{n} \sum_{i=1}^n \xi_{ik} \xi_{ik'} w(m_i) B_{i,k}(s) B_{i,k'}(t) \\ &\quad + \sum_{k=1}^M \lambda_k \frac{1}{n} \sum_{i=1}^n (\xi_{ik}^2 - 1) w(m_i) B_{i,k}(s) B_{i,k}(t) \\ &\quad + \sum_{k=1}^M \lambda_k \frac{1}{n} \sum_{i=1}^n w(m_i) (B_{i,k}(s) B_{i,k}(t) - \mathbb{E}(B_{i,k}(s) B_{i,k}(t))) \\ &\quad + \sigma \sum_{k=1}^M \sqrt{\lambda_k} \frac{1}{n} \sum_{i=1}^n \frac{w(m_i)}{m_i} \xi_{ik} (\varepsilon_i^T \mathbf{B}_i(t) B_{i,k}(s) + \varepsilon_i^T \mathbf{B}_i(s) B_{i,k}(t)) \\ &\quad + \sigma^2 \frac{1}{n} \sum_{i=1}^n \frac{w(m_i)}{m_i^2} [\varepsilon_i^T \mathbf{B}_i(s) (\mathbf{B}_i(t))^T \varepsilon_i - (\mathbf{B}_i(t))^T \mathbf{B}_i(s)] \\ &\quad + \sigma^2 \frac{1}{n} \sum_{i=1}^n \frac{w(m_i)}{m_i^2} [(\mathbf{B}_i(t))^T \mathbf{B}_i(s) - \mathbb{E}((\mathbf{B}_i(t))^T \mathbf{B}_i(s))]. \end{aligned}$$

The last term in the above expression vanishes since $|s - t| > 4(B_K + C_Q)h$ so that, in particular, $(\mathbf{B}_i(t))^T \mathbf{B}_i(s) = 0$. Note that, $\max_{1 \leq i \leq n} w(m_i)$ is bounded. By (D.5), $\max_{1 \leq k, k' \leq M} |B_{i,k}(s) B_{i,k'}(t)| \leq C_4^2 h^{-2}$ and by straightforward com-

putations

$$\max_{1 \leq k, k' \leq M} \max_{1 \leq i \leq n} \text{Var}(B_{i,k}(s)B_{i,k'}(t)) \leq C_5 \max\{(\underline{m}h)^{-2}, (\underline{m}h)^{-1}\} \quad \text{for } C_5 > 0. \quad (\text{D.6})$$

Since $\bar{m}_n h = o(1)$, for sufficiently large n , the bound in (D.6) is of the form $C_5(\bar{m}_n h)^{-2}$. Thus, using the condition that $\bar{m}_n^2 = o(nh^2/\log n)$, by *Bernstein's inequality*, we conclude that given $\eta > 0$, there exists $c_{1,\eta} > 0$ such that for sufficiently large n ,

$$\begin{aligned} & \mathbb{P} \left(\max_{1 \leq k \leq M} \left| \frac{1}{n} \sum_{i=1}^n w(m_i) (B_{i,k}(s)B_{i,k}(t) - \mathbb{E}(B_{i,k}(s)B_{i,k}(t))) \right| > c_{1,\eta} \sqrt{\frac{\log n}{n\bar{m}_n^2 h^2}} \right) \\ & \leq n^{-\eta}. \end{aligned} \quad (\text{D.7})$$

Again, since $\bar{m}h = o(1)$, from direct calculations, for sufficiently large n

$$\max_{1 \leq k, k' \leq M} \text{Var}((B_{i,k}(s)B_{i,k'}(t))^2) \leq C_6(\bar{m}h)^{-6}.$$

From this, (D.5) and *Bernstein's inequality*, we have, for any $x > 0$, and for any $1 \leq k, k' \leq M$,

$$\begin{aligned} & \mathbb{P} \left(\frac{1}{n} \sum_{i=1}^n (w(m_i))^2 [B_{i,k}(s)B_{i,k'}(t) - \mathbb{E}(B_{i,k}(s)B_{i,k'}(t))]^2 \right. \\ & \quad \left. - \frac{1}{n} \sum_{i=1}^n (w(m_i))^2 \text{Var}(B_{i,k}(s)B_{i,k'}(t)) > x \right) \\ & \leq \exp \left(-\frac{nx^2}{2C_6(\bar{m}h)^{-6} + xC_7h^{-4}} \right) \end{aligned} \quad (\text{D.8})$$

for some constant $C_7 > 0$. We take

$$x \sim \begin{cases} \left(\frac{\log n}{n(\bar{m}h)^6} \right)^{1/2} & \text{if } \bar{m}^6 = O\left(\frac{nh^2}{\log n}\right) \\ \frac{\log n}{nh^4} & \text{if } \bar{m}^6 \gg \left(\frac{nh^2}{\log n}\right) \end{cases}$$

and recall that by (3.4), $\bar{m}^2 = o(nh^2/\log n)$ and $n(\bar{m}h)^2/\log n \gg \kappa_n^2 \geq 1$, so that

$$\frac{1}{(\bar{m}h)^2} \gg \max \left\{ \left(\frac{\log n}{n(\bar{m}h)^6} \right)^{1/2}, \frac{\log n}{nh^4} \right\},$$

to conclude from (D.8) that there exists $c_{3,\eta} > 0$ such that for sufficiently large n ,

$$\begin{aligned} & \mathbb{P} \left(\frac{1}{n} \sum_{i=1}^n (w(m_i))^2 [B_{i,k}(s)B_{i,k'}(t) - \mathbb{E}(B_{i,k}(s)B_{i,k'}(t))]^2 > (C_5 + c_{3,\eta}) \frac{1}{\bar{m}^2 h^2} \right) \\ & \leq n^{-\eta}. \end{aligned} \quad (\text{D.9})$$

Since $\max_{1 \leq k, k' \leq M} |\mathbb{E}(B_{i,k}(s)B_{i,k'}(t))| = O(1)$, a bound similar to (D.9) holds even when the summands are not centered.

Next, we deal with the first term on the RHS of (D.6). Conditioning on \mathbf{T} , we can express $\xi_k = (\xi_{ik})_{i=1}^n$ as $\xi_k := \mathbf{R}^{1/2} \tilde{\xi}_k$, where the random vectors $\tilde{\xi}_k$ are independently distributed as $N_n(0, I)$ for different k 's, and are independent of $\{\varepsilon_i\}_{i=1}^n$. Then we can write (conditionally on \mathbf{T})

$$\frac{1}{n} \sum_{i=1}^n \xi_{ik} \xi_{ik'} w(m_i) B_{i,k}(s) B_{i,k'}(t) = \tilde{\xi}_k^T \Phi(\mathbf{T}) \tilde{\xi}_{k'},$$

where $\Phi(\mathbf{T}) = \mathbf{R}^{1/2} \text{diag}((\frac{w(m_i)}{n} B_{i,k}(s) B_{i,k'}(t))_{i=1}^n) \mathbf{R}^{1/2}$. Observe that by (D.5) and condition C3, we have $\|\Phi(\mathbf{T})\| \leq C_4 \kappa_n / (nh^2)$. Moreover, by the non-centered version of (D.9), there is a set D_n in the sigma-field generated by \mathbf{T} , such that, $P(\mathbf{T} \in D_n) \geq 1 - n^{-\eta}$, and for $\mathbf{T} \in D_n$, we have $\text{tr}(\Phi(\mathbf{T})^2) \leq c_{4,\eta} \kappa_n^2 / (n\bar{m}^2 h^2)$. Therefore, by an application of Lemma A.2 (with $A \sim (nh^2)^{-1}$, $B \sim (n\bar{m}^2 h^2)^{-1/2}$ and $t \sim \kappa_n (\log n / (n\bar{m}^2 h^2))^{1/2}$), we have, for some $c_{5,\eta} > 0$,

$$\max_{1 \leq k \neq k' \leq M} \mathbb{P} \left(\left| \frac{1}{n} \sum_{i=1}^n \xi_{ik} \xi_{ik'} w(m_i) B_{i,k}(s) B_{i,k'}(t) \right| > c_{5,\eta} \kappa_n \sqrt{\frac{\log n}{n\bar{m}^2 h^2}}, \mathbf{T} \in D_n \right) \leq n^{-\eta}. \quad (\text{D.10})$$

Very similar arguments (using Lemma A.3 in addition to Lemma A.2) can be used to obtain bounds of order $\kappa_n (\log n / (n\bar{m}^2 h^2))^{1/2}$ (that hold with probability at least $1 - O(n^{-\eta})$, for any given $\eta > 0$) for the second, fourth and fifth terms on the RHS of (D.6). Thus, combining these bounds, we have, for some constant $c_{6,\eta} > 0$,

$$\mathbb{P} \left(\left| \overline{W}_{Ah}(s, t) \left(\widehat{C}(s, t) - \mathbb{E}(\widehat{C}(s, t)) \right) \right| > c_{6,\eta} \kappa_n \sqrt{\frac{\log n}{n\bar{m}^2 h^2}} \right) \leq n^{-\eta}. \quad (\text{D.11})$$

Finally, we deal with the case $|s - t| \leq Ah/2$. In this case, we have $\widehat{C}(s, t) - \mathbb{E}[\widehat{C}(s, t)] = W_{Ah}(s, t) (\widehat{C}_*(\frac{s+t}{2}) - \mathbb{E}[\widehat{C}_*(\frac{s+t}{2})])$ (ignoring the maximum over h^2 in the definition). Then similar (but somewhat simpler) arguments, now involving Lemma A.3, show that for some $c_{7,\eta} > 0$,

$$\mathbb{P} \left(\left| W_{Ah}(s, t) \left(\widehat{C}_* \left(\frac{s+t}{2} \right) - \mathbb{E} \left[\widehat{C}_* \left(\frac{s+t}{2} \right) \right] \right) \right| > c_{7,\eta} \kappa_n \sqrt{\frac{\log n}{n\bar{m}^2 h^2}} \right) \leq n^{-\eta}. \quad (\text{D.12})$$

Combining (D.11) and (D.12) we obtain the result.

Appendix E

Computation of $\mathbb{E} \|H_\nu \widetilde{C}_c \psi_\nu\|_2^2$

We put the different pieces derived in Appendix C together to obtain the bound for $\mathbb{E} \|H_\nu \widetilde{C}_c \psi_\nu\|_2^2$. For ease of notation, we denote by $\mathcal{H}_\nu \equiv \mathcal{H}_\nu(x, s_1, s_2, t_1, t_2)$

the *integral operator* with kernel $H_\nu(x, s_1)H_\nu(x, s_2)\psi_\nu(t_1)\psi_\nu(t_2)$. Then, with r_1, r_2 taking values 0 or 1, not both equal to 0,

$$\begin{aligned} & \int \int \int \int \int \mathcal{H}_\nu(x, s_1, s_2, t_1, t_2)(C(s_1, t_1))^{r_1}(C(s_2, t_2))^{r_2} ds_1 ds_2 dt_1 dt_2 dx \\ = & 0; \end{aligned} \tag{E.1}$$

$$\begin{aligned} & \int \int \int \int \int \mathcal{H}_\nu(x, s_1, s_2, t_1, t_2)(C(s_1, t_2))^{r_1}(C(s_2, t_1))^{r_2} ds_1 ds_2 dt_1 dt_2 dx \\ = & 0; \text{ and} \end{aligned} \tag{E.2}$$

$$\begin{aligned} & \int \int \int \int \int \mathcal{H}_\nu(x, s_1, s_2, t_1, t_2)C(s_1, s_2)C(t_1, t_2) ds_1 ds_2 dt_1 dt_2 dx \\ = & \sum_{1 \leq k \neq \nu \leq M} \frac{\lambda_\nu \lambda_k}{(\lambda_k - \lambda_\nu)^2}. \end{aligned} \tag{E.3}$$

Implicitly using (E.7)–(E.9), we also have the bound

$$\left| \int \int \int \int \int \mathcal{H}_\nu(x, s_1, s_2, t_1, t_2)R(s_1, s_2, t_1, t_2) ds_1 ds_2 dt_1 dt_2 dx \right| = O(\| R \|_\infty). \tag{E.4}$$

From Proposition C.1, the total contribution in (C.6) of (C.8) and the first term on the RHS of (C.7), becomes

$$\begin{aligned} & \left(\frac{1}{n} \left(1 - \frac{1}{n} \sum_{i=1}^n \frac{4m_i - 6}{m_i(m_i - 1)} \right) + \frac{n - 1}{n} \right) \\ & \cdot \int \left[\int \int H_\nu(x, s) \overline{W}_{\tilde{h}_n}(s, t) [C(s, t) + O(h_n^2)] \psi_\nu(t) ds dt \right]^2 dx \\ & + \left(\frac{1}{n} \left(1 - \frac{1}{n} \sum_{i=1}^n \frac{4m_i - 6}{m_i(m_i - 1)} \right) + \frac{1}{n^2} \sum_{i_1 \neq i_2} \rho_{i_1 i_2}^2 \right) \\ & \cdot \int \int \int \int \int H_\nu(x, s_1) H_\nu(x, s_2) \overline{W}_{\tilde{h}_n}(s_1, t_1) \overline{W}_{\tilde{h}_n}(s_2, t_2) \\ & \quad (C(s_1, s_2) + O(h_n^2)) (C(t_1, t_2) + O(h_n^2)) \psi_\nu(t_1) \psi_\nu(t_2) ds_1 ds_2 dt_1 dt_2 dx \\ & + \left(\frac{1}{n} \left(1 - \frac{1}{n} \sum_{i=1}^n \frac{4m_i - 6}{m_i(m_i - 1)} \right) + \frac{1}{n^2} \sum_{i_1 \neq i_2} \rho_{i_1 i_2}^2 \right) \\ & \cdot \int \int \int \int \int H_\nu(x, s_1) H_\nu(x, s_2) \overline{W}_{\tilde{h}_n}(s_1, t_1) \overline{W}_{\tilde{h}_n}(s_2, t_2) \\ & \quad (C(s_1, t_2) + O(h_n^2)) (C(s_2, t_1) + O(h_n^2)) \psi_\nu(t_1) \psi_\nu(t_2) ds_1 ds_2 dt_1 dt_2 dx. \end{aligned} \tag{E.5}$$

Since $H_\nu C \psi_\nu \equiv 0$, it can be checked that the first integral in (E.5) is $O(h_n^2)$. On the other hand, from the definition of $\overline{W}_{\tilde{h}_n}(s, t)$ and the fact that $H_\nu C \psi_\nu \equiv 0$, it follows that the last integral term is $O(\tilde{h}_n)$.

Next, apply \mathcal{H}_ν to the following functions: $W_{\tilde{h}_n}(s_1, t_1)W_{\tilde{h}_n}(s_2, t_2)D_2(s_1, s_2, t_1, t_2)$ and $2\overline{W}_{\tilde{h}_n}(s_1, t_1)W_{\tilde{h}_n}(s_2, t_2)D_3(s_1, s_2, t_1, t_2)$, where $D_2(s_1, s_2, t_1, t_2)$ and $D_3(s_1, s_2, t_1, t_2)$ are the terms given by the sum of the first three terms on the RHS of (C.10) (including the isolated $O(h_n^2)$ term), and the sum of the first three terms on the RHS of (C.11), respectively. Then, adding these terms to (E.5), we have, by (E.1)–(E.4), (E.9) (for dealing with the isolated $O(h_n^2)$ term in (C.10)), and the comment following (E.5), that this sum equals

$$\begin{aligned}
 R_1 &= \frac{1}{n} \left(\sum_{1 \leq k \neq \nu \leq M} \frac{\lambda_k \lambda_\nu}{(\lambda_k - \lambda_\nu)^2} \right) \\
 &+ \left(\frac{1}{n^2} \sum_{i_1 \neq i_2}^n \rho_{i_1 i_2}^2 \right) \left(\sum_{1 \leq k \neq \nu \leq M} \frac{\lambda_k \lambda_\nu}{(\lambda_k - \lambda_\nu)^2} + O(h_n) \right) \\
 &+ O(h_n^4) + O\left(\frac{1}{nm_n}\right) + O\left(\frac{h_n}{n}\right). \tag{E.6}
 \end{aligned}$$

Next, for notational convenience, denote the integral operator \mathcal{H}_ν applied to Z_j multiplied by $\overline{W}_{\tilde{h}_n}(s_1, t_1)\overline{W}_{\tilde{h}_n}(s_2, t_2)W_{\tilde{h}_n}(s_1, t_2)$ (where Z_j are as in Propositions C.1–C.2) by $\mathcal{H}_\nu \overline{W}^{s_1, t_1} \overline{W}^{s_2, t_2} W^{s_1, t_2} Z_j$. Using (E.7)–(E.12), and the bounds in Proposition C.1 for Z_j , $j = 1, \dots, 4$, we have,

$$\begin{aligned}
 R_2 &:= \mathcal{H}_\nu \overline{W}^{s_1, t_1} \overline{W}^{s_2, t_2} Z_1 = \mathcal{H}_\nu \overline{W}^{s_1, t_1} \overline{W}^{s_2, t_2} W^{s_1, s_2} Z_1 = O\left(\frac{1}{nh_n \underline{m}_n}\right), \\
 R_3 &:= \mathcal{H}_\nu \overline{W}^{s_1, t_1} \overline{W}^{s_2, t_2} Z_2 = \mathcal{H}_\nu \overline{W}^{s_1, t_1} \overline{W}^{s_2, t_2} W^{s_1, t_2} Z_2 = O\left(\frac{1}{nm_n}\right), \\
 R_4 &:= \mathcal{H}_\nu \overline{W}^{s_1, t_1} \overline{W}^{s_2, t_2} Z_3 = \mathcal{H}_\nu \overline{W}^{s_1, t_1} \overline{W}^{s_2, t_2} W^{s_2, t_1} Z_3 = O\left(\frac{1}{nm_n}\right), \\
 R_5 &:= \mathcal{H}_\nu \overline{W}^{s_1, t_1} \overline{W}^{s_2, t_2} Z_4 = \mathcal{H}_\nu \overline{W}^{s_1, t_1} \overline{W}^{s_2, t_2} W^{t_1, t_2} Z_4 = O\left(\frac{1}{nm_n}\right).
 \end{aligned}$$

Using analogous reasoning, from Propositions C.1 and C.2 we also have

$$\begin{aligned}
 R_6 &:= \mathcal{H}_\nu \overline{W}^{s_1, t_1} \overline{W}^{s_2, t_2} Z_5 = O\left(\frac{1}{nh_n \underline{m}_n^2}\right), \\
 R_7 &:= \mathcal{H}_\nu \overline{W}^{s_1, t_1} \overline{W}^{s_2, t_2} Z_6 = O\left(\frac{1}{nm_n^2}\right) \\
 R_8 &:= \mathcal{H}_\nu W^{s_1, t_1} W^{s_2, t_2} Z_7 = O\left(\frac{h_n}{nm_n}\right), \\
 R_9 &:= \mathcal{H}_\nu \overline{W}^{s_1, t_1} W^{s_2, t_2} Z_8 = O\left(\frac{1}{nh_n \underline{m}_n}\right), \\
 R_{10} &:= \mathcal{H}_\nu \overline{W}^{s_1, t_1} W^{s_2, t_2} Z_9 = O\left(\frac{1}{nm_n}\right).
 \end{aligned}$$

Hence, combining (E.6) with the bounds for R_2 to R_{10} , using the definitions of $\mathbb{E}[\tilde{C}(s_1, t_1)\tilde{C}(s_2, t_2)]$, $\mathbb{E}[\widehat{C}(\frac{s_1+t_1}{2})\widehat{C}(\frac{s_2+t_2}{2})]$, and $\mathbb{E}[\tilde{C}(s_1, t_1)\widehat{C}(\frac{s_2+t_2}{2})]$, and plugging everything back into (C.3), we complete the proof of Proposition 4.1.

Some error bounds involving Dirac's δ

Here, we provide some key estimates that are crucial to obtaining the overall risk bound. They all involve the operator H_ν . Due to the decomposition (4.1) we can reduce the computations of these bounds to integrals involving $\{\psi_k(\cdot)\}_{k=1}^M$ and $\delta(\cdot, \cdot)$. Throughout we assume that $R(s_1, s_2, t_1, t_2)$ is a “nice” function satisfying certain boundedness conditions. Then the following bounds hold.

$$\begin{aligned} & \left| \int \int \int \int \delta(x, s_1)\delta(x, s_2)R(s_1, s_2, t_1, t_2)\psi_\nu(t_1)\psi_\nu(t_2)ds_1ds_2dt_1dt_2 \right| \\ & \leq \|R\|_\infty \|\psi_\nu\|_\infty^2. \end{aligned} \quad (\text{E.7})$$

$$\begin{aligned} & \left| \int \int \int \int \delta(x, s_1)\delta(x, s_2)W_{Ah_n}(s_1, t_1) \right. \\ & \quad \left. \cdot R(s_1, s_2, t_1, t_2)\psi_\nu(t_1)\psi_\nu(t_2)ds_1ds_2dt_1dt_2 \right| \\ & \leq Ah_n \|R\|_\infty \|\psi_\nu\|_\infty^2. \end{aligned} \quad (\text{E.8})$$

$$\begin{aligned} & \left| \int \int \int \int \delta(x, s_1)\delta(x, s_2)W_{Ah_n}(s_1, t_1)W_{Ah_n}(s_2, t_2) \right. \\ & \quad \left. \cdot R(s_1, s_2, t_1, t_2)\psi_\nu(t_1)\psi_\nu(t_2)ds_1ds_2dt_1dt_2 \right| \\ & \leq (Ah_n)^2 \|R\|_\infty \|\psi_\nu\|_\infty^2. \end{aligned} \quad (\text{E.9})$$

$$\begin{aligned} & \left| \int \int \int \int \delta(x, s_1)\delta(x, s_2)W_{Ah_n}(t_1, t_2) \right. \\ & \quad \left. \cdot R(s_1, s_2, t_1, t_2)\psi_\nu(t_1)\psi_\nu(t_2)ds_1ds_2dt_1dt_2 \right| \\ & \leq Ah_n \|R\|_\infty \|\psi_\nu\|_\infty^2. \end{aligned} \quad (\text{E.10})$$

$$\begin{aligned} & \left| \int \int \int \int \delta(x, s_1)\delta(x, s_2)W_{Ah_n}(t_1, t_2)W_{Ah_n}(s_2, t_2) \right. \\ & \quad \left. \cdot R(s_1, s_2, t_1, t_2)\psi_\nu(t_1)\psi_\nu(t_2)ds_1ds_2dt_1dt_2 \right| \\ & \leq (Ah_n)^2 \|R\|_\infty \|\psi_\nu\|_\infty^2. \end{aligned} \quad (\text{E.11})$$

$$\begin{aligned} & \left| \int \int \int \int \delta(x, s_1)\delta(x, s_2)W_{Ah_n}(t_1, s_2)W_{Ah_n}(s_2, t_2) \right. \\ & \quad \left. \cdot R(s_1, s_2, t_1, t_2)\psi_\nu(t_1)\psi_\nu(t_2)ds_1ds_2dt_1dt_2 \right| \\ & \leq (Ah_n)^2 \|R\|_\infty \|\psi_\nu\|_\infty^2. \end{aligned} \quad (\text{E.12})$$

Acknowledgement

The authors thank the editor, the associate editor and the referee for their helpful comments and suggestions. Paul's research is partially supported by the NSF grants DMS-0806128, DMR-1035468 and DMS-1106690. Peng's research is partially supported by the NSF grants DMS-0806128 and DMS-1007583.

References

- [1] ASH, R. B. (1972). *Real Analysis and Probability*, Academic Press. [MR0435320](#)
- [2] BANERJEE, S. and JOHNSON, G. A. (2006). Coregionalized single- and multiresolution spatially varying growth curve modeling with application to weed growth. *Biometrics* **61**, 617-625. [MR2247216](#)
- [3] BESSE, P., CARDOT, H. and FERRATY, F. (1997). Simultaneous nonparametric regression of unbalanced longitudinal data. *Computational Statistics and Data Analysis* **24**, 255-270. [MR1455696](#)
- [4] CAI, T. and HALL, P. (2006). Prediction in functional linear regression. *Annals of Statistics* **34**, 2159-2179. [MR2291496](#)
- [5] CARDOT, H., FERRATY F. and SARDA P. (1999). Functional Linear Model. *Statistics and Probability Letters* **45**, 11-22. [MR1718346](#)
- [6] CARDOT, H. (2000). Nonparametric estimation of smoothed principal components analysis of sampled noisy functions. *Journal of Nonparametric Statistics* **12**, 503-538. [MR1785396](#)
- [7] CHUI, C. (1987). *Multivariate Splines*. SIAM. [MR1033490](#)
- [8] CHIOU, J.-M. and LI, P.-L. (2007). Functional clustering and identifying substructures of longitudinal data. *Journal of the Royal Statistical Society, Series B*, **69**, 679-699. [MR2370075](#)
- [9] CHEN, K., PAUL, D. and WANG, J.-L. (2009). Properties of principal component analysis for correlated data. *Technical Report*, University of California, Davis.
- [10] FERRATY, F. and VIEU, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer. [MR2229687](#)
- [11] GELFAND, A. E., SCHMIDT, A., BANERJEE, S. and SIRMANS, C. F. (2004). Nonstationary multivariate process modeling through spatially varying coregionalization (with discussion). *Test* **13**, 1-50. [MR2154003](#)
- [12] HALL, P. and HOROWITZ, J. L. (2007). Methodology and convergence rates for functional linear regression. *Annals of Statistics* **35**, 70-91. [MR2332269](#)
- [13] HALL, P., MÜLLER, H.-G. and WANG, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Annals of Statistics* **34**, 1493-1517. [MR2278365](#)
- [14] HLUBINKA, D. and PRCHAL, L. (2007). Changes in atmospheric radiation from the statistical point of view. *Computational Statistics and Data Analysis* **51**, 4926-4941. [MR2364550](#)

- [15] JAMES, G. M., HASTIE, T. J. and SUGAR, C. A. (2000). Principal component models for sparse functional data. *Biometrika*, **87**, 587-602. [MR1789811](#)
- [16] JAMES, G. M. and HASTIE, T. (2001). Functional linear discriminant analysis for irregularly sampled curves. *Journal of the Royal Statistical Society, Series B*, **64**, 411-432. [MR1858401](#)
- [17] JAMES, G. M. and SUGAR, C. A. (2003). Clustering for sparsely sampled functional data. *Journal of the American Statistical Association*, **98**, 397-408. [MR1995716](#)
- [18] KATO, T. (1980). *Perturbation Theory of Linear Operators*. Springer-Verlag.
- [19] KNEIP, A. and UTIKAL, K. J. (2001). Inference for density families using functional principal component analysis, *Journal of the American Statistical Association*, **96**, 519-542. [MR1946423](#)
- [20] NICA, A. and SPEICHER, R. (2006). *Lectures on the Combinatorics of Free Probability*. Cambridge University Press. [MR2266879](#)
- [21] PAUL, D. and JOHNSTONE, I. M. (2007). Augmented sparse principal component analysis for high dimensional data. *Technical Report*. (<http://anson.ucdavis.edu/~debashis/techrep/augmented-spca.pdf>)
- [22] PAUL, D. and PENG, J. (2008). Principal components analysis for sparsely observed correlated functional data using a kernel smoothing approach. *Technical report*. [arXiv:0807.1106v1](#).
- [23] PAUL, D. and PENG, J. (2009). Consistency of restricted maximum likelihood estimators of principal components. *Annals of Statistics*, **37**, 1229-1271. [MR2509073](#)
- [24] PENG, J. and PAUL, D. (2009). A geometric approach to maximum likelihood estimation of the functional principal components from sparse longitudinal data. *Journal of Computational and Graphical Statistics*, **18**, 995-1015. [MR2598035](#)
- [25] PENG, J. and MÜLLER, H.-G. (2008). Distance-based clustering of sparsely observed stochastic processes, with applications to online auctions. To appear in *Annals of Applied Statistics*. [MR2516804](#)
- [26] RAMSAY, J. and SILVERMAN, B. W. (2005). *Functional Data Analysis, 2nd Edition*. Springer. [MR2168993](#)
- [27] RICE, J. A. and WU, C. O. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics*, **57**, 253-259. [MR1833314](#)
- [28] SPELLMAN, P.T., SHERLOCK, G., ZHANG, M. Q., IYER, V. R., ANDERS, K., EISEN, M. B., BROWN, P. O., BOTSTEIN, D. and FUTCHER, B. (1998). Comprehensive identification of cell cycle-regulated genes of the yeast *saccharomyces cerevisiae* by microarray hybridization. *Molecular Biology of the Cell*, **9**, 3273-3297.
- [29] TANG, R. and MÜLLER, H.-G. (2008). Time-synchronized clustering of gene expression trajectories. *Biostatistics*, **10**, 32-45
- [30] YAO, F. and LEE, T. C. M. (2006). Penalized spline models for functional principal component analysis. *Journal of the Royal Statistical Society, Series B* **68**, 3-25. [MR2212572](#)

- [31] YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100**, 577-590. [MR2160561](#)
- [32] YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005). Functional linear regression for longitudinal data. *Annals of Statistics* **33**, 2873-2903. [MR2253106](#)
- [33] YUAN, M. and CAI, T. T. (2010). A reproducing kernel Hilbert space approach to functional linear regression. *Annals of Statistics* **38**, 3412-3444. [MR2766857](#)
- [34] WACKERNAGEL, H. (2003). *Multivariate Geostatistics, 3rd Edition*. Springer.
- [35] ZHANG, J. T. and Chen, J. (2007). Statistical inferences for functional data. *Annals of Statistics*, **35**, 1052-1079. [MR2341698](#)
- [36] ZHOU, L., HUANG, J., MARTINEZ, J. G., MAITY, A., BALADANDAYUTHAPANI, V. and CARROLL, R. C. (2010). Reduced rank mixed effects models for spatially correlated hierarchical functional data. *Journal of the American Statistical Association* **105**, 390-400. [MR2757206](#)