Biometrika (2015), **102**, 2, *pp.* 421–437
Printed in Great Britain
doi: 10.1093/biomet/asv006
Advance Access publication 2 April 2015

Effective dimension reduction for sparse functional data

By F. YAO, E. LEI

Department of Statistical Sciences, University of Toronto, Toronto, Ontario M5S 3G3, Canada
fyao@utstat.toronto.edu edwin@utstat.toronto.edu

and Y. WU

Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695, U.S.A.
wu@stat.ncsu.edu

Summary

We propose a method of effective dimension reduction for functional data, emphasizing the sparse design where one observes only a few noisy and irregular measurements for some or all of the subjects. The proposed method borrows strength across the entire sample and provides a way to characterize the effective dimension reduction space, via functional cumulative slicing. Our theoretical study reveals a bias-variance trade-off associated with the regularizing truncation and decaying structures of the predictor process and the effective dimension reduction space. A simulation study and an application illustrate the superior finite-sample performance of the method.

Some key words: Cumulative slicing; Effective dimension reduction; Inverse regression; Sparse functional data.

1. Introduction

In functional data analysis, one is often interested in how a scalar response $Y \in \mathbb{R}$ varies with a smooth trajectory $X(t)$, where t is an index variable defined on a closed interval $\mathcal{T}$; see Ramsay & Silverman (2005). To be specific, one seeks to model the relationship $Y = M(X; \epsilon)$, where M is a smooth functional and the error process ϵ has zero mean and finite variance σ^2 and is independent of X. Although modelling M parametrically can be restrictive in many applications, modelling M nonparametrically is infeasible in practice due to slow convergence rates associated with the curse of dimensionality. Therefore a class of semiparametric index models has been proposed to approximate $M(X; \epsilon)$ with an unknown link function $g : \mathbb{R}^{K+1} \to \mathbb{R}$; that is,

$$Y = g\left(\langle \beta_1, X\rangle, \ldots, \langle \beta_K, X\rangle; \epsilon\right), \tag{1}$$

where K is the reduced dimension of the model, $\beta_1, \ldots, \beta_K$ are linearly independent index functions, and $\langle u, v\rangle = \int u(t)v(t)\,\mathrm{d}t$ is the usual L^2 inner product. The functional linear model $Y = \beta_0 + \int \beta_1(t)X(t)\,\mathrm{d}t + \epsilon$ is a special case and has been studied extensively (Cardot et al., 1999; Müller & Stadtmüller, 2005; Yao et al., 2005b; Cai & Hall, 2006; Hall & Horowitz, 2007; Yuan & Cai, 2010).

In this article, we tackle the index model (1) from the perspective of effective dimension reduction, in the sense that the K linear projections $\langle \beta_1, X\rangle, \ldots, \langle \beta_K, X\rangle$ form a sufficient statistic. This is particularly useful when the process X is infinite-dimensional. Our primary goal

is to discuss dimension reduction for functional data, especially when the trajectories are corrupted by noise and are sparsely observed with only a few observations for some, or even all, of the subjects. Pioneered by Li (1991) for multivariate data, effective dimension reduction methods are typically link-free, requiring neither specification nor estimation of the link function (Duan & Li, 1991), and aim to characterize the K-dimensional effective dimension reduction space $S_{Y|X} = \text{span}(\beta_1, \ldots, \beta_K)$ onto which X is projected. Such index functions β_k are called effective dimension reduction directions, K is the structural dimension, and $S_{Y|X}$ is also known as the central subspace (Cook, 1998). Li (1991) characterized $S_{Y|X}$ via the inverse mean $E(X \mid Y)$ by sliced inverse regression, which has motivated much work for multivariate data. For instance, Cook & Weisberg (1991) estimated $\text{var}(X \mid Y)$, Li (1992) dealt with the Hessian matrix of the regression curve, Xia et al. (2002) proposed minimum average variance estimation as an adaptive approach based on kernel methods, Chiaromonte et al. (2002) modified sliced inverse regression for categorical predictors, Li & Wang (2007) worked with empirical directions, and Zhu et al. (2010) proposed cumulative slicing estimation to improve on sliced inverse regression.

The literature on effective dimension reduction for functional data is relatively sparse. Ferré & Yao (2003) proposed functional sliced inverse regression for completely observed functional data, and Li & Hsing (2010) developed sequential χ^2 testing procedures to decide the structural dimension of functional sliced inverse regression. Apart from effective dimension reduction approaches, James & Silverman (2005) estimated the index and link functions jointly for an additive form $g(\langle\beta_1, X\rangle, \ldots, \langle\beta_K, X\rangle; \epsilon) = \beta_0 + \sum_{k=1}^{K} g_k(\langle\beta_k, X\rangle) + \epsilon$, assuming that the trajectories are densely or completely observed and that the index and link functions are elements of a finite-dimensional spline space. Chen et al. (2011) estimated the index and additive link functions nonparametrically and relaxed the finite-dimensional assumption for theoretical analysis but retained the dense design.

Jiang et al. (2014) proposed an inverse regression method for sparse functional data by estimating the conditional mean $E\{X(t) \mid Y = \tilde{y}\}$ with a two-dimensional smoother applied to pooled observed values of X in a local neighbourhood of $(t, \tilde{y})$. The computation associated with a two-dimensional smoother is considerable and further increased by the need to select two different bandwidths. In contrast, we aim to estimate the effective dimension reduction space by drawing inspiration from cumulative slicing for multivariate data (Zhu et al., 2010). When adapted to the functional setting, cumulative slicing offers a novel and computationally simple way of borrowing strength across subjects to handle sparsely observed trajectories. This advantage has not been exploited elsewhere. As we will demonstrate later, although extending cumulative slicing to completely observed functional data is straightforward, it adopts a different strategy for the sparse design via a one-dimensional smoother, with potentially effective usage of the data.

2. METHODOLOGY

2·1. *Dimension reduction for functional data*

Let $\mathcal{T}$ be a compact interval, and let X be a random variable defined on the real separable Hilbert space $H \equiv L^2(\mathcal{T})$ endowed with inner product $\langle f, g\rangle = \int_{\mathcal{T}} f(t)g(t)\,\mathrm{d}t$ and norm $\|f\| = \langle f, f\rangle^{1/2}$. We assume that:

Assumption 1. X is centred and has a finite fourth moment, $\int_{\tau} E\{X^4(t)\}\,\mathrm{d}t < \infty$.

Under Assumption 1, the covariance surface of X is $\Sigma(s, t) = E\{X(s)X(t)\}$, which generates a Hilbert–Schmidt operator Σ on H that maps f to $(\Sigma f)(t) = \int_{\tau} \Sigma(s, t) f(s)\,\mathrm{d}s$. This operator can be written succinctly as $\Sigma = E(X \otimes X)$, where the tensor product $u \otimes v$ denotes

the rank-one operator on H that maps w to $(u \otimes v)w = \langle u, w\rangle v$. By Mercer's theorem, Σ admits a spectral decomposition $\Sigma = \sum_{j=1}^{\infty} \alpha_j \phi_j \otimes \phi_j$, where the eigenfunctions $\{\phi_j\}_{j=1,2,\ldots}$ form a complete orthonormal system in H and the eigenvalues $\{\alpha_j\}_{j=1,2,\ldots}$ are strictly decreasing and positive such that $\sum_{j=1}^{\infty} \alpha_j < \infty$. Finally, recall that the effective dimension reduction directions $\beta_1, \ldots, \beta_K$ in model (1) are linearly independent functions in H, and the response $Y \in \mathbb{R}$ is assumed to be conditionally independent of X given the K projections $\langle \beta_1, X\rangle, \ldots, \langle \beta_K, X\rangle$.

Zhu et al. (2010) observed that for a fixed $\tilde{y} \in \mathbb{R}$, using two slices $I_1 = (-\infty, \tilde{y}]$ and $I_2 = (\tilde{y}, +\infty)$ would maximize the use of data and minimize the variability in each slice. The kernel of the sliced inverse regression operator $\mathrm{var}\{E(X \mid Y)\}$ is estimated by the two-slice version $\Lambda_0(s, t; \tilde{y}) \propto m(s, \tilde{y})m(t, \tilde{y})$, where $m(t, \tilde{y}) = E\{X(t)1(Y \leqslant \tilde{y})\}$ is an unconditional expectation, in contrast to the conditional expectation $E\{X(t) \mid Y\}$ of functional sliced inverse regression. Since Λ_0 with a fixed $\tilde{y}$ spans at most one direction of $S_{Y|X}$, it is necessary to combine all possible estimates of $m(t, \tilde{y})$ by letting $\tilde{y}$ run across the support of $\tilde{Y}$, an independent copy of Y. Therefore, the kernel of the proposed functional cumulative slicing is

$$\Lambda(s, t) = E\{m(s, \tilde{Y})m(t, \tilde{Y})w(\tilde{Y})\}, \tag{2}$$

where $w(\tilde{y})$ is a known nonnegative weight function. Denote the corresponding integral operator of $\Lambda(s, t)$ by Λ also. The following theorem establishes the validity of our proposal. Analogous to the multivariate case, a linearity assumption is needed.

Assumption 2. For any function $b \in H$, there exist constants $c_0, \ldots, c_K \in \mathbb{R}$ such that

$$E(\langle b, X\rangle \mid \langle \beta_1, X\rangle, \ldots, \langle \beta_K, X\rangle) = c_0 + \sum_{k=1}^{K} c_k \langle \beta_k, X\rangle.$$

This assumption is satisfied when X has an elliptically contoured distribution, which is more general than, but has a close connection to, a Gaussian process (Cambanis et al., 1981; Li & Hsing, 2010).

THEOREM 1. *If Assumptions* 1 *and* 2 *hold for model* (1)*, then the linear space spanned by* $\{m(t, \tilde{y}) : \tilde{y} \in \mathbb{R}\}$ *is contained in the linear space spanned by* $\{\Sigma\beta_1, \ldots, \Sigma\beta_K\}$*, i.e.,* $\mathrm{span}(\{m(t, \tilde{y}) : \tilde{y} \in \mathbb{R}\}) \subseteq \mathrm{span}(\Sigma\beta_1, \ldots, \Sigma\beta_K)$.

An important observation from Theorem 1 is that for any $b \in H$ orthogonal to the space spanned by $\{\Sigma\beta_1, \ldots, \Sigma\beta_K\}$ and for any $x \in H$, we have $\langle b, \Lambda x\rangle = 0$, implying that $\mathrm{range}(\Lambda) \subseteq \mathrm{span}(\Sigma\beta_1, \ldots, \Sigma\beta_K)$. If Λ has K nonzero eigenvalues, the space spanned by its eigenfunctions is precisely $\mathrm{span}(\Sigma\beta_1, \ldots, \Sigma\beta_K)$. Recall that our goal is to estimate the central subspace $S_{Y|X}$, even though the effective dimension reduction directions themselves are not identifiable. For specificity, we regard these eigenfunctions of $\Sigma^{-1}\Lambda$ associated with the K largest nonzero eigenvalues as the index functions $\beta_1, \ldots, \beta_K$, unless stated otherwise.

As the covariance operator Σ is Hilbert–Schmidt, it is not invertible when defined from H to H. Similarly to Ferré & Yao (2005), let R_Σ denote the range of Σ, and let $R_\Sigma^{-1} = \{b \in H : \sum_{j=1}^{\infty} \alpha_j^{-1} \langle b, \phi_j\rangle \phi_j, b \in R_\Sigma\}$. Then Σ is a one-to-one mapping from $R_\Sigma^{-1} \subset H$ onto R_Σ, with inverse $\Sigma^{-1} = \sum_{j=1}^{\infty} \alpha_j^{-1} \phi_j \otimes \phi_j$. This is reminiscent of finding a generalized inverse of a matrix. Let $\xi_j = \langle X, \phi_j\rangle$ denote the jth principal component, or generalized Fourier coefficient, of X, and assume that:

Assumption 3. $\sum_{j=1}^{\infty} \sum_{l=1}^{\infty} \alpha_j^{-2} \alpha_l^{-1} E^2[E\{\xi_j 1(Y \leqslant \tilde{Y}) \mid \tilde{Y}\} E\{\xi_l 1(Y \leqslant \tilde{Y}) \mid \tilde{Y}\}] < \infty$.

PROPOSITION 1. *Under Assumptions* 1–3*, the eigenspace associated with the K nonnull eigenvalues of* $\Sigma^{-1}\Lambda$ *is well-defined in H.*

This is a direct analogue of Theorem 4.8 in He et al. (2003) and Theorem 2.1 in Ferré & Yao (2005).

2·2. *Functional cumulative slicing for sparse functional data*

For the data $\{(X_i, Y_i) : i = 1, \ldots, n\}$, independent and identically distributed as (X, Y), the predictor trajectories X_i are observed intermittently, contaminated with noise, and collected in the form of repeated measurements $\{(T_{ij}, U_{ij}) : i = 1, \ldots, n;\ j = 1, \ldots, N_i\}$, where $U_{ij} = X_i(T_{ij}) + \varepsilon_{ij}$ with measurement error ε_{ij} that is independent and identically distributed as ε with zero mean and constant variance σ_x^2, and independent of all other random variables. When only a few observations are available for some or even all subjects, individual smoothing to recover X_i is infeasible and one must pool data across subjects for consistent estimation.

To estimate the functional cumulative slicing kernel Λ in (2), the key quantity is the unconditional mean $m(t, \tilde{y}) = E\{X(t)1(Y \leqslant \tilde{y})\}$. For sparsely and irregularly observed X_i, cross-sectional estimation as used in multivariate cumulative slicing is inapplicable. To maximize the use of available data, we propose to pool the repeated measurements across subjects via a scatterplot smoother, which works in conjunction with the strategy of cumulative slicing. We use a local linear estimator $\hat{m}(t, \tilde{y}) = \hat{a}_0$ (Fan & Gijbels, 1996), solving

$$\min_{(a_0,a_1)} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left\{U_{ij}1(Y_i \leqslant \tilde{y}) - a_0 - a_1(T_{ij} - t)\right\}^2 K_1\left(\frac{T_{ij} - t}{h_1}\right), \tag{3}$$

where K_1 is a nonnegative and symmetric univariate kernel density and $h_1 = h_1(n)$ is the bandwidth to control the amount of smoothing. We ignore the dependence among data from the same individual (Lin & Carroll, 2000) and use leave-one-curve-out crossvalidation to select h_1 (Rice & Silverman, 1991). Then an estimator of the kernel function $\Lambda(s, t)$ is its sample moment

$$\hat{\Lambda}(s, t) = \frac{1}{n}\sum_{i=1}^{n} \hat{m}(s, Y_i)\hat{m}(t, Y_i)w(Y_i). \tag{4}$$

The distinction between our method and that of Jiang et al. (2014) lies in the inverse function $m(t, y)$ which forms the effective dimension reduction space. It is notable that (4) is a univariate smoother that includes the effective data satisfying $\{T_{ij} \in (t - h_1, t + h_1),\ Y_i \leqslant y\}$, roughly at an order of $(nh_1)^{1/2}$ for estimating $m(t, y) = E\{X(t)1(Y \leqslant y)\}$ for a sparse design with $E(N_n) < \infty$, where N_n is the expected number of repeated observations per subject. By contrast, equation (2·4) in Jiang et al. (2014) uses the data satisfying $\{T_{ij} \in (t - h_t, t + h_t),\ Y_i \in (y - h_y, y + h_y)\}$ for estimating $m(t, y) = E\{X(t) \mid Y = y\}$, roughly at an order of $(nh_th_y)^{1/2}$. This is reflected in the faster convergence of the estimated operator $\hat{\Lambda}$ compared with $\hat{\Gamma}_e$ in Jiang et al. (2014), indicating potentially effective usage of the data based on univariate smoothing. The computation associated with a two-dimensional smoother is considerable and further exacerbated by the need to select different bandwidths h_t and h_y.

For the covariance operator Σ, following Yao et al. (2005a), denote the observed raw covariances by $G_i(T_{ij}, T_{il}) = U_{ij}U_{il}$. Since $E\{G_i(T_{ij}, T_{il}) \mid T_{ij}, T_{il}\} = \mathrm{cov}\{X(T_{ij}), X(T_{il})\} + \sigma^2\delta_{jl}$, where δ_{jl} is 1 if $j = l$ and 0 otherwise, the diagonal of the raw covariances should be removed.

Solving

$$\min_{(b_0,b_1,b_2)} \sum_{i=1}^{n} \sum_{j \neq l}^{N_i} \left\{ G_i(T_{ij}, T_{il}) - b_0 - b_1(T_{ij} - s) - b_2(T_{il} - t) \right\}^2 K_2 \left(\frac{T_{ij} - s}{h_2}, \frac{T_{il} - t}{h_2} \right) \tag{5}$$

yields $\hat{\Sigma}(s, t) = \hat{b}_0$, where K_2 is a nonnegative bivariate kernel density and $h_2 = h_2(n)$ is the bandwidth chosen by leave-one-curve-out crossvalidation; see Yao et al. (2005a) for details on the implementation. Since the inverse operator Σ^{-1} is unbounded, we regularize by projection onto a truncated subspace. To be precise, let s_n be a possibly divergent sequence and let $\Pi_{s_n} = \sum_{j=1}^{s_n} \phi_j \otimes \phi_j$ and $\hat{\Pi}_{s_n} = \sum_{j=1}^{s_n} \hat{\phi}_j \otimes \hat{\phi}_j$ denote the orthogonal projectors onto the eigensubspaces associated with the s_n largest eigenvalues of Σ and $\hat{\Sigma}$, respectively. Then $\Sigma_{s_n} = \Pi_{s_n} \Sigma \Pi_{s_n}$ and $\hat{\Sigma}_{s_n} = \hat{\Pi}_{s_n} \hat{\Sigma} \hat{\Pi}_{s_n}$ are two sequences of finite-rank operators converging to Σ and $\hat{\Sigma}$ as $n \to \infty$, with bounded inverses $\Sigma_{s_n}^{-1} = \sum_{j=1}^{s_n} \alpha_j^{-1} \phi_j \otimes \phi_j$ and $\hat{\Sigma}_{s_n}^{-1} = \sum_{j=1}^{s_n} \hat{\alpha}_j^{-1} \hat{\phi}_j \otimes \hat{\phi}_j$, respectively. Finally, we obtain the eigenfunctions associated with the K largest nonzero eigenvalues of $\hat{\Sigma}_{s_n}^{-1} \hat{\Lambda}$ as the estimates of the effective dimension reduction directions $\{\hat{\beta}_{k,s_n}\}_{k=1,\ldots,K}$.

The situation for completely observed X_i is similar to the multivariate case and considerably simpler. The quantities $m(t, \tilde{y})$ and $\Sigma(s, t)$ are easily estimated by their respective sample moments $\hat{m}(t, \tilde{y}) = n^{-1} \sum_{i=1}^{n} X_i(t) 1(Y_i \leqslant \tilde{y})$ and $\hat{\Sigma}(s, t) = n^{-1} \sum_{i=1}^{n} X_i(s) X_i(t)$, while the estimate of Λ remains the same as (4). For densely observed X_i, individual smoothing can be used as a pre-processing step to recover smooth trajectories, and the estimation error introduced in this step can be shown to be asymptotically negligible under certain design conditions, i.e., it is equivalent to the ideal situation of the completely observed X_i (Hall et al., 2006).

For small values of Y_i, $\hat{m}(t, Y_i)$ obtained by (3) may be unstable due to the smaller number of pooled observations in the slice. A suitable weight function w may be used to refine the estimator $\hat{\Lambda}(s, t)$. In our numerical studies, the naive choice of $w \equiv 1$ performed fairly well compared to other methods. Analogous to the multivariate case, choosing an optimal w remains an open question.

Ferré & Yao (2005) avoided inverting Σ with the claim that for a finite-rank operator Λ, $\text{range}(\Lambda^{-1} \Sigma) = \text{range}(\Sigma^{-1} \Lambda)$; however, Cook et al. (2010) showed that this requires more stringent conditions that are not easily fulfilled.

The selection of K_n and s_n deserves further study. For selecting the structural dimension K, the only relevant work to date is Li & Hsing (2010), where sequential χ^2 tests are used to determine K for the method of Ferré & Yao (2003). How to extend such tests to sparse functional data, if feasible at all, is worthy of further exploration. It is also important to tune the truncation parameter s_n that contributes to the variance-bias trade-off of the resulting estimator, although analytical guidance for this is not yet available.

3. Asymptotic properties

In this section we present asymptotic properties of the functional cumulative slicing kernel operator and the effective dimension reduction directions for sparse functional data. The numbers of measurements N_i and the observation times T_{ij} are considered to be random, to reflect a sparse and irregular design. Specifically, we make the following assumption.

Assumption 4. The N_i are independent and identically distributed as a positive discrete random variable N_n, where $E(N_n) < \infty$, $\text{pr}(N_n \geqslant 2) > 0$ and $\text{pr}(N_n \leqslant M_n) = 1$ for some constant

sequence M_n that is allowed to diverge, i.e., $M_n \to \infty$ as $n \to \infty$. Moreover, $(\{T_{ij},\ j \in J_i\}, \{U_{ij},\ j \in J_i\})$ are independent of N_i for $J_i \subseteq \{1, \ldots, N_i\}$.

Writing $T_i = (T_{i1}, \ldots, T_{iN_i})^{\mathrm{T}}$ and $U_i = (U_{i1}, \ldots, U_{iN_i})^{\mathrm{T}}$, the data quadruplets $Z_i = \{T_i, U_i, Y_i, N_i\}$ are thus independent and identically distributed. Extremely sparse designs are also covered, with only a few measurements for each subject. Other regularity conditions are standard and listed in the Appendix, including assumptions on the smoothness of the mean and covariance functions of X, the distributions of the observation times, and the bandwidths and kernel functions used in the smoothing steps. Write $\|A\|_{\mathrm{H}}^2 = \int_{\mathcal{T}} \int_{\mathcal{T}} A^2(s,t)\,\mathrm{d}s\,\mathrm{d}t$ for $A \in L^2(\mathcal{T} \times \mathcal{T})$.

THEOREM 2. *Under Assumptions* 1, 4 *and* A1–A4 *in the Appendix, we have*

$$\left\|\hat{\Lambda} - \Lambda\right\|_{\mathrm{H}} = O_{\mathrm{p}}\left(n^{-1/2}h_1^{-1/2} + h_1^2\right), \quad \left\|\hat{\Sigma} - \Sigma\right\|_{\mathrm{H}} = O_{\mathrm{p}}\left(n^{-1/2}h_2^{-1} + h_2^2\right).$$

The key result here is the L^2 convergence of the estimated operator $\hat{\Lambda}$, in which we exploit the projections of nonparametric U-statistics together with a decomposition of $\hat{m}(t, \tilde{y})$ to overcome the difficulty caused by the dependence among irregularly spaced measurements. The estimator $\hat{\Lambda}$ is obtained by averaging the smoothers $\hat{m}(t, Y_i)$ over Y_i, which is crucial in order to achieve the univariate convergence rate for this bivariate estimator. The convergence of the covariance operator Σ is presented for completeness, given in Theorem 2 of Yao & Müller (2010).

We are now ready to characterize the estimation of the central subspace $S_{Y|X} = \mathrm{span}(\beta_1, \ldots, \beta_K)$. Unlike the multivariate or finite-dimensional case, where the convergence of $\hat{S}_{Y|X}$ follows immediately from the convergence of $\hat{\Sigma}$ and $\hat{\Lambda}$ given a bounded Σ^{-1}, we have to approximate Σ^{-1} with a sequence of truncated estimates $\hat{\Sigma}_{s_n}^{-1}$, which introduces additional variability and bias inherent in a functional inverse problem. Since we specifically regarded the index functions $\{\beta_1, \ldots, \beta_K\}$ as the eigenfunctions associated with the K largest eigenvalues of $\Sigma^{-1}\Lambda$, their estimates are thus equivalent to $\hat{S}_{Y|X}$. For some constant $C > 0$, we require the eigenvalues of Σ to satisfy the following condition:

Assumption 5. $\alpha_j > \alpha_{j+1} > 0$, $E(\xi_j^4) \leqslant C\alpha_j^2$, and $\alpha_j - \alpha_{j+1} \geqslant C^{-1}j^{-a-1}$ for $j \geqslant 1$.

This condition on the decaying speed of the eigenvalues α_j prevents the spacings between consecutive eigenvalues from being too small, and also implies that $\alpha_j \geqslant Cj^{-a}$ with $a > 1$ given the boundedness of Σ. Expressing the index functions as $\beta_k = \sum_{j=1}^{\infty} b_{kj}\phi_j$ $(k = 1, \ldots, K)$, we impose a decaying structure on the generalized Fourier coefficients $b_{kj} = \langle \beta_k, \phi_j \rangle$:

Assumption 6. $|b_{kj}| \leqslant Cj^{-b}$ for $j \geqslant 1$ and $k = 1, \ldots, K$, where $b > 1/2$.

In order to accurately estimate the eigenfunctions ϕ_j from $\hat{\Sigma}$, one requires $j \leqslant \sup\{\ell : \alpha_\ell - \alpha_{\ell+1} > 2\|\hat{\Sigma} - \Sigma\|_{\mathrm{H}}\}$, i.e., that the distance to α_j from the nearest eigenvalue does not fall below $2\|\hat{\Sigma} - \Sigma\|_{\mathrm{H}}$ (Hall & Hosseini-Nasab, 2006); this implicitly places an upper bound on the truncation parameter s_n. Given Assumption 5 and Theorem 2, we provide a sufficient condition on s_n. Here we write $c_{1n} \asymp c_{2n}$ when $c_{1n} = O(c_{2n})$ and $c_{2n} = O(c_{1n})$.

Assumption 7. As $n \to \infty$, $s_n^{a+1}(n^{-1/2}h_2^{-1} + h_2^2) \to 0$; moreover, if $h_2 \asymp n^{-1/6}$, $s_n = o\{n^{1/(3a+3)}\}$.

THEOREM 3. *Under Assumptions* 1–7 *and* A1–A4 *in the Appendix, for all* $k = 1, \ldots, K$,

$$\|\hat{\beta}_{k,s_n} - \beta_k\| = O_{\mathrm{p}}\left\{s_n^{3a/2+1}\left(n^{-1/2}h_1^{-1/2} + h_1^2\right) + s_n^{2a+3/2}\left(n^{-1/2}h_2^{-1} + h_2^2\right) + s_n^{-b+1/2}\right\}. \tag{6}$$

This result associates the convergence of $\hat{\beta}_{k,s_n}$ with the truncation parameter s_n and the decay rates of α_j and b_{kj}, indicating a bias-variance trade-off with respect to s_n. One can view s_n as a tuning parameter that is allowed to diverge slowly and which controls the resolution of the covariance estimation. Specifically, the first two terms on the right-hand side of (6) are attributed to the variability of estimating $\Sigma_{s_n}^{-1}\Lambda$ with $\hat{\Sigma}_{s_n}^{-1}\hat{\Lambda}$, and the last term corresponds to the approximation bias of $\Sigma_{s_n}^{-1}\Lambda$. The first term of the variance is due to $\|\hat{\Sigma}_{s_n}^{-1}\hat{\Lambda}\hat{\Sigma}_{s_n}^{-1/2} - \hat{\Sigma}_{s_n}^{-1}\Lambda\hat{\Sigma}_{s_n}^{-1/2}\|_{\mathrm{H}}$ and becomes increasingly unstable with a larger truncation. The second part of the variance is due to $\|(\Sigma_{s_n}^{-1} - \hat{\Sigma}_{s_n}^{-1})\Lambda\Sigma_{s_n}^{-1/2}\|_{\mathrm{H}}$, and the approximation bias is determined by the smoothness of β_k; for instance, a smoother β_k with a larger b leads to a smaller bias.

4. SIMULATIONS

In this section we illustrate the performance of the proposed functional cumulative slicing method in terms of estimation and prediction. Although our proposal is link-free for estimating index functions β_k, a general index model (1) may lead to model predictions with high variability, especially given the relatively small sample sizes frequently encountered in functional data analysis. Thus we follow Chen et al. (2011) in assuming an additive structure for the link function g in (1), i.e., $Y = \beta_0 + \sum_{k=1}^{K} g_k(\langle\beta_k, X\rangle) + \epsilon$. In each Monte Carlo run, a sample of $n = 200$ functional trajectories is generated from the process $X_i(t) = \sum_{j=1}^{50}\xi_{ij}\phi_j(t)$, where $\phi_j(t) = \sin(\pi tj/5)/\sqrt{5}$ for j even and $\phi_j(t) = \cos(\pi tj/5)/\sqrt{5}$ for j odd, the functional principal component scores ξ_{ij} are independent and identically distributed as $N(0, j^{-1\cdot5})$, and $\mathcal{T} = [0, 10]$. For the setting of sparsely observed functional data, the number of observations per subject, N_i, is chosen uniformly from $\{5, \ldots, 10\}$, the observational times T_{ij} are independent and identically distributed as Un[0, 10], and the measurement error ε_{ij} is independent and identically distributed as $N(0, 0{\cdot}1)$. The effective dimension reduction directions are generated by $\beta_1(t) = \sum_{j=1}^{50} b_j\phi_j(t)$, where $b_j = 1$ for $j = 1, 2, 3$ and $b_j = 4(j-2)^{-3}$ for $j = 4, \ldots, 50$, and $\beta_2(t) = 0{\cdot}3^{1/2}(t/5 - 1)$, which cannot be represented with finite Fourier terms. The following single- and multiple-index models are considered:

$$\begin{aligned}
&\text{Model I:} && Y = \sin(\pi\langle\beta_1, X\rangle/4) + \epsilon,\\
&\text{Model II:} && Y = \arctan(\pi\langle\beta_1, X\rangle/2) + \epsilon,\\
&\text{Model III:} && Y = \sin(\pi\langle\beta_1, X\rangle/3) + \exp(\langle\beta_2, X\rangle/3) + \epsilon,\\
&\text{Model IV:} && Y = \arctan(\pi\langle\beta_1, X\rangle) + \sin(\pi\langle\beta_2, X\rangle/6)/2 + \epsilon,
\end{aligned}$$

where the regression error ϵ is independent and identically distributed as $N(0, 1)$ for all models.

We compare our method with that of Jiang et al. (2014) for sparse functional data in terms of estimation and prediction. Denote the true structural dimension by K_0. Due to the nonidentifiability of the β_k, we examine the projection operator of the effective dimension space, i.e., $P = \sum_{k=1}^{K_0}\beta_k \otimes \beta_k$ and $\hat{P}_{K,s_n} = \sum_{k=1}^{K}\hat{\beta}_{K,s_n} \otimes \hat{\beta}_{K,s_n}$. To assess the estimation of the effective dimension reduction space, we calculate $\|\hat{P}_{K,s_n} - P\|_{\mathrm{H}}$ as the estimation error. To assess model

Table 1. *Estimation error and relative prediction error, multiplied by* 100*, obtained from* 100 *Monte Carlo repetitions (with standard errors in parentheses) for sparse functional data*

Model	Metric	FCS	IRLD	Metric	FCS	IRLD
I		61·1 (1·1)	61·3 (1·1)		17·7 (0·6)	17·9 (0·5)
II	Estimation	59·3 (1·0)	59·5 (1·0)	Prediction	19·6 (0·6)	19·4 (0·5)
III	error	63·7 (0·8)	63·9 (0·9)	error	18·8 (0·5)	19·5 (0·4)
IV		63·8 (0·8)	63·9 (0·9)		45·2 (1·1)	45·4 (1·1)

FCS, functional cumulative slicing; IRLD, the method of Jiang et al. (2014), where (K, s_n) is selected by minimizing the estimation and prediction errors.

Table 2. *Estimation error and relative prediction error, multiplied by* 100*, obtained from* 100 *Monte Carlo repetitions (with standard errors in parentheses) for dense functional data*

Metric	Model	FCS	IRLD	FSIR5	FSIR10	FIND
	I	39·2 (1·6)	45·5 (1·5)	59·4 (2·1)	61·7 (2·2)	47·1 (1·6)
Estimation	II	35·5 (1·4)	38·1 (1·3)	56·1 (1·8)	57·8 (1·9)	44·5 (1·5)
error	III	59·6 (0·8)	63·1 (0·8)	72·6 (1·1)	74·1 (1·3)	63·6 (0·9)
	IV	57·2 (0·6)	59·0 (0·6)	69·3 (1·0)	68·9 (0·9)	61·0 (0·8)
	I	11·1 (0·6)	12·7 (0·5)	17·1 (0·7)	16·7 (0·6)	16·1 (1·1)
Prediction	II	9·8 (0·5)	10·5 (0·4)	15·5 (0·7)	16·9 (1·0)	14·9 (0·8)
error	III	13·5 (0·5)	15·2 (0·5)	15·8 (0·6)	16·6 (0·5)	14·7 (0·6)
	IV	19·9 (0·7)	21·9 (0·7)	31·1 (1·4)	32·2 (1·4)	24·2 (1·2)

FCS, functional cumulative slicing; IRLD, inverse regression for longitudinal data (Jiang et al., 2014); FSIR5, functional sliced inverse regression (Ferré & Yao, 2003) with five slices; FSIR10, functional sliced inverse regression (Ferré & Yao, 2003) with ten slices; FIND, functional index model (Chen et al., 2011).

prediction, we estimate the link functions g_k nonparametrically by fitting a generalized additive model $Y_i = \beta_0 + \sum_{k=1}^{K} g_k(Z_{ik}) + \epsilon_i$ (Hastie & Tibshirani, 1990), where $Z_{ik} = \langle \hat{\beta}_{k,s_n}, \tilde{X}_i \rangle$ with $\tilde{X}_i$ being the best linear unbiased predictor of X_i (Yao et al., 2005a). We generate a validation sample of size 500 in each Monte Carlo run and calculate the average of the relative prediction errors, $500^{-1} \sum_{i=1}^{500} (\hat{Y}_i^* - Y_i^*)^2/\sigma^2$, over different values of (K, s_n), where $\sigma^2 = 1$ and $\hat{Y}_i^* = \hat{\beta}_0 + \sum_{k=1}^{K} \hat{g}_k(Z_{ik}^*)$ with $Z_{ik}^* = \langle \hat{\beta}_{k,s_n}, X_i^* \rangle$, the X_i^* being the underlying trajectories in the testing sample. We report in Table 1 the average estimation and prediction errors, minimized over (K, s_n), along with their standard errors over 100 Monte Carlo repetitions. For estimation and prediction, both methods selected $(K, s_n) = (1, 3)$ for the single-index models I and II, and selected $(K, s_n) = (2, 2)$ for the multiple-index models III and IV. The two approaches perform comparably in this sparse setting, which could be due to the inverse covariance estimation that dominates the overall performance. Our method takes one-third of the computation time of the method of Jiang et al. (2014) for this sparse design.

We also present simulation results for dense functional data, where $N_i = 50$ and the T_{ij} are sampled independently and identically from Un[0, 10]. With (K, s_n) selected so as to minimize the estimation and prediction errors, we compare our proposal with the method of Jiang et al. (2014), functional sliced inverse regression (Ferré & Yao, 2003) using five or ten slices, and the functional index model of Chen et al. (2011). Table 2 indicates that our method slightly outperforms the method of Jiang et al. (2014), followed by the method of Chen et al. (2011), while functional sliced inverse regression (Ferré & Yao, 2003) is seen to be suboptimal. Our method takes only one-sixth of the time required by Jiang et al. (2014) for this setting.

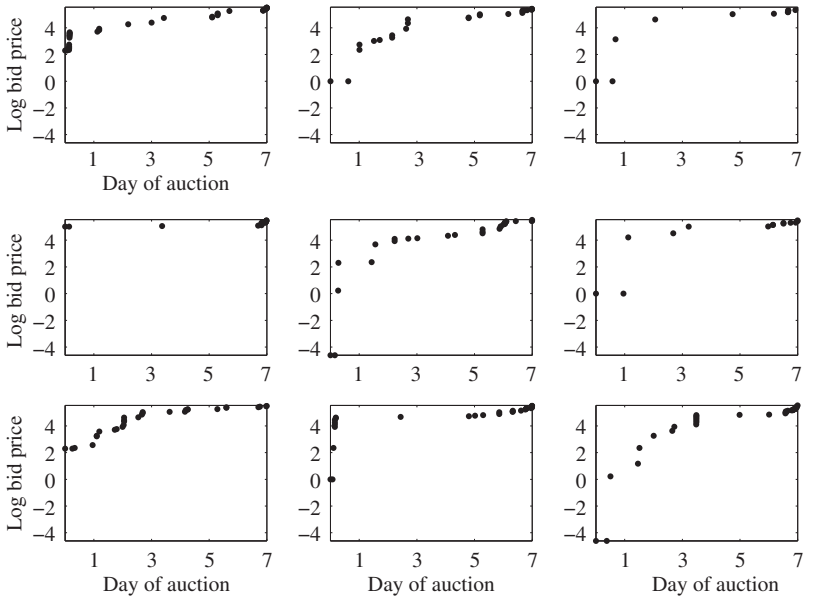

Fig. 1. Observed bid prices over the seven-day auction period of nine randomly selected auctions, after log-transform.

5. Data application

In this application, we study the relationship between the winning bid price of 156 Palm M515 PDA devices auctioned on eBay between March and May of 2003 and the bidding history over the seven-day period of each auction. Each observation from a bidding history represents a live bid, the actual price a winning bidder would pay for the device, known as the willingness-to-pay price. Further details on the bidding mechanism can be found in Liu & Müller (2009). We adopt the view that the bidding histories are independent and identically distributed realizations of a smooth underlying price process. Due to the nature of online auctions, the jth bid of the ith auction usually arrives irregularly at time T_{ij}, and the number of bids N_i can vary widely, from nine to 52 for this dataset. As is usual in modelling prices, we take the log-transform of the bid prices. Figure 1 shows a sample of nine randomly selected bid histories over the seven-day period of the respective auction. Typically, the bid histories are sparse until the final hours of each auction, when bid sniping occurs. At this point, snipers place their bids at the last possible moment to try to deny competing bidders the chance of placing a higher bid.

Since our main interest is in the predictive power of price histories up to time T for the winning bid prices, we consider the regression of the winning price on the history trajectory $X(t)$ ($t \in [0, T]$), and set $T = 4{\cdot}5, 4{\cdot}6, 4{\cdot}7, \ldots, 6{\cdot}8$ days. For each analysis on the domain $[0, T]$, we select the optimal structural dimension K and the truncation parameter s_n by minimizing the average five-fold crossvalidated prediction error over 20 random partitions. Figure 2(a) shows the minimized average crossvalidated prediction errors, compared with those obtained using the method of Jiang et al. (2014). With the increasing prediction power as the bidding histories encompass more data, the proposed method appears to yield more favourable prediction across different time domains.

As an illustration, we present the analysis for $T = 6$. The estimated model components using the proposed method are shown in Fig. 2(b), with the parameters chosen as $K = 2$ and $s_n = 2$. The first index function assigns contrasting weights to bids made before and after the first day, indicating that some bidders tend to underbid at the beginning only to quickly overbid

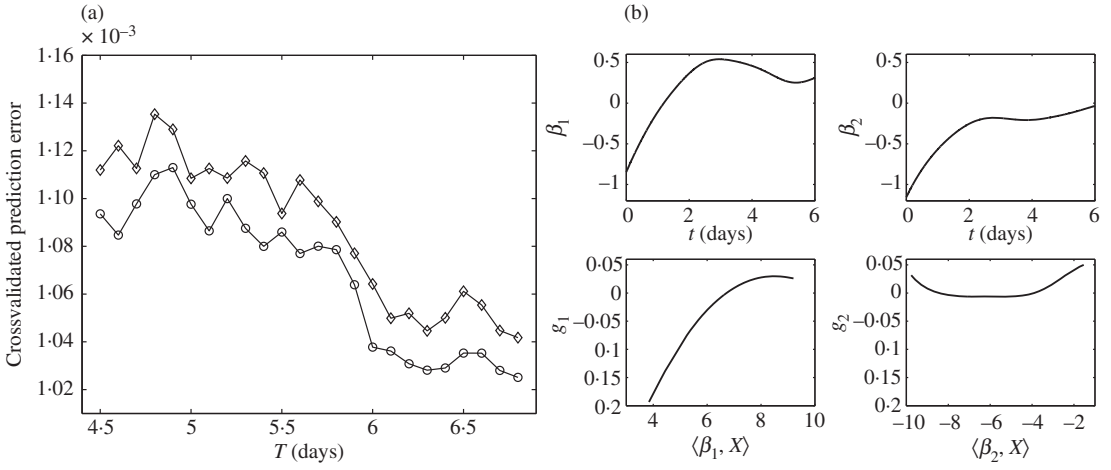

Fig. 2. (a) Average five-fold crossvalidated prediction errors for functional cumulative slicing (circles) and the method of Jiang et al. (2014) (diamonds) over 20 random partitions across different time domains $[0, T]$, for sparse eBay auction data. (b) Estimated model components for eBay auction data using functional cumulative slicing with $K = 2$ and $s_n = 2$; the upper panels show the estimated index functions, i.e., the effective dimension reduction directions, and the lower panels show the additive link functions.

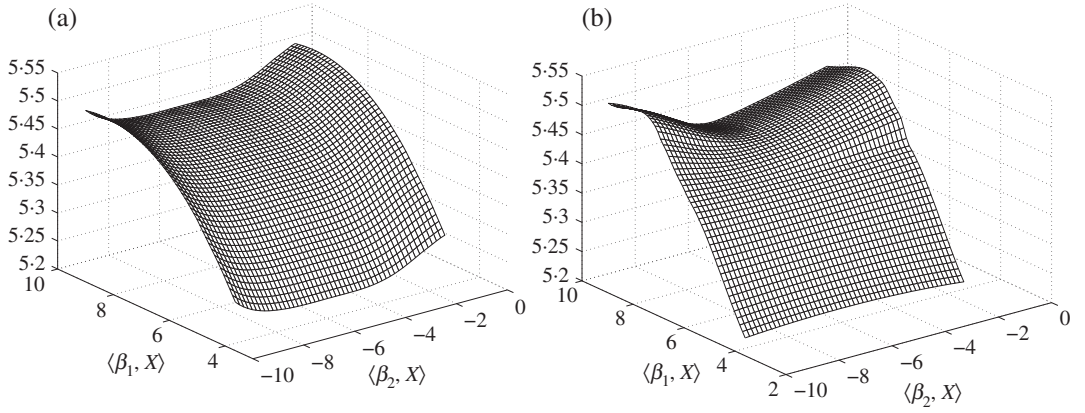

Fig. 3. Fitted regression surfaces for the eBay data: (a) additive; (b) unstructured.

relative to the mean. The second index represents a cautious type of bidding behaviour, entering at a lower price and slowly increasing towards the average level. These features contribute most towards the prediction of the winning bid prices. Also seen are the nonlinear patterns in the estimated additive link functions. Using these estimated model components, we display in Fig. 3(a) the additive surface $\hat{\beta}_0 + \hat{g}_1(\langle\hat{\beta}_1, X\rangle) + \hat{g}_2(\langle\hat{\beta}_2, X\rangle)$. We also fit an unstructured index model $g(\langle\beta_1, X\rangle, \langle\beta_2, X\rangle)$, where g is nonparametrically estimated using a bivariate local linear smoother; this is shown in Fig. 3(b), and is seen to agree reasonably well with the additive regression surface.

ACKNOWLEDGEMENT

We thank two reviewers, an associate editor, and the editor for their helpful comments. This research was partially supported by the U.S. National Institutes of Health and National Science Foundation, and the Natural Sciences and Engineering Research Council of Canada.

Appendix

Regularity conditions and auxiliary lemmas

Without loss of generality, we assume that the known weight function is $w(\cdot) = 1$. Write $\mathcal{T} = [a, b]$ and $\mathcal{T}^\delta = [a - \delta, b + \delta]$ for some $\delta > 0$; denote a single observation time by T and a pair of observation times by $(T_1, T_2)^{\mathrm{T}}$, with densities $f(t)$ and $f_2(s, t)$, respectively. Recall the unconditional mean function $m(t, y) = E\{X(t)1(Y \leqslant y)\}$. The regularity conditions for the underlying moment functions and design densities are as follows, where ℓ_1 and ℓ_2 are nonnegative integers. We assume that:

Assumption A1. $\partial^2\Sigma/(\partial s^{\ell_1}\partial t^{\ell_2})$ is continuous on $\mathcal{T}^\delta \times \mathcal{T}^\delta$ for $\ell_1 + \ell_2 = 2$, and $\partial^2 m/\partial t^2$ is bounded and continuous with respect to $t \in \mathcal{T}$ for all $y \in \mathbb{R}$.

Assumption A2. $f_1^{(1)}(t)$ is continuous on $\mathcal{T}^\delta$ with $f_1(t) > 0$, and $\partial f_2/(\partial s^{\ell_1}\partial t^{\ell_2}) f_2$ is continuous on $\mathcal{T}^\delta \times \mathcal{T}^\delta$ for $\ell_1 + \ell_2 = 1$ with $f_2(s, t) > 0$.

Assumption A1 can be guaranteed by a twice-differentiable process, and Assumption A2 is standard and implies the boundedness and Lipschitz continuity of f. Recall the bandwidths h_1 and h_2 used in the smoothing steps for $\hat{m}$ in (3) and $\hat{\Sigma}$ in (5), respectively; we assume that:

Assumption A3. $h_1 \to 0$, $h_2 \to 0$, $nh_1^3/\log n \to \infty$, and $nh_2^2 \to \infty$.

We say that a bivariate kernel function K_2 is of order (ν, ℓ), where ν is a multi-index $\nu = (\nu_1, \nu_2)^{\mathrm{T}}$, if

$$\iint u^{\ell_1} v^{\ell_2} K_2(u, v)\,\mathrm{d}u\,\mathrm{d}v \begin{cases} = 0, & 0 \leqslant \ell_1 + \ell_2 < \ell,\ \ell_1 \neq \nu_1,\ \ell_2 \neq \nu_2, \\ = (-1)^{|\nu|}\nu_1!\,\nu_2!, & \ell_1 = \nu_1,\ \ell_2 = \nu_2, \\ \neq 0, & \ell_1 + \ell_2 = \ell, \end{cases}$$

where $|\nu| = \nu_1 + \nu_2 < \ell$. The univariate kernel K is said to be of order (ν, ℓ) for a univariate $\nu = \nu_1$ if this definition holds with $\ell_2 = 0$ on the right-hand side, integrating only over the argument u on the left-hand side. The following standard conditions on the kernel densities are required.

Assumption A4. The kernel functions K_1 and K_2 are nonnegative with compact supports, bounded, and of order $(0, 2)$ and $\{(0, 0)^{\mathrm{T}}, 2\}$, respectively.

Lemma A1 is a mean-squared version of Theorem 1 in Martins-Filho & Yao (2006), which asserts the asymptotic equivalence of a nonparametric V-statistic to the projection of the corresponding U-statistic. Lemma A2 is a restatement of Lemma 1(b) of Martins-Filho & Yao (2007) adapted to sparse functional data.

Lemma A1. *Let $\{Z_i\}_{i=1}^n$ be a sequence of independent and identically distributed random variables, and let u_n and v_n be U- and V-statistics with kernel function $\psi_n(Z_1, \ldots, Z_k)$. In addition, let $\hat{u}_n = n^{-1}k\sum_{i=1}^n\{\psi_{1n}(Z_i) - \phi_n\} + \phi_n$, where $\psi_{1n}(Z_i) = E\{\psi_n(Z_{i1}, \ldots, Z_{ik}) \mid Z_i\}$ for $i \in \{i_1, \ldots, i_k\}$ and $\phi_n = E\{\psi_n(Z_1, \ldots, Z_k)\}$. If $E\{\psi_n^2(Z_1, \ldots, Z_k)\} = o(n)$, then $nE\{(v_n - \hat{u}_n)^2\} = o(1)$.*

Lemma A2. *Given Assumptions* 1–4 *and* A1–A4, *let*

$$s_k(t) = \sum_{i=1}^n \sum_{j=1}^{M_n} \frac{1(N_i \geqslant j)}{nh_1} K_1\left(\frac{T_{ij} - t}{h_1}\right)\left(\frac{T_{ij} - t}{h_1}\right)^k.$$

Then $\sup_{t\in\mathcal{T}} h_1^{-1}|s_k(t) - E\{s_k(t)\}| = O_{\mathrm{p}}(1)$ *for* $k = 0, 1, 2$.

Proofs of the theorems

Proof of Theorem 1. This theorem is an analogue of Theorem 1 in Zhu et al. (2010); thus its proof is omitted. □

Proof of Theorem 2. For brevity, we write M_n and N_n as M and N, respectively. Let

$$S_n(t) = \sum_{i=1}^{n}\sum_{j=1}^{M} \frac{1(N_i \geqslant j)}{nh_1E(N)} K_1\left(\frac{T_{ij}-t}{h_1}\right)\begin{pmatrix} 1 & (T_{ij}-t)/h_1 \\ (T_{ij}-t)/h_1 & \{(T_{ij}-t)/h_1\}^2 \end{pmatrix},$$

$$S(t) = \begin{pmatrix} f_T(t) & 0 \\ 0 & f_T(t)\sigma_K^2 \end{pmatrix},$$

where $\sigma_K^2 = \int u^2K(u)\,\mathrm{d}u$. The local linear estimator of $m(t,\tilde{y})$ with kernel K_1 is

$$\hat{m}(t,\tilde{y}) = (1,0)S_n^{-1}(t)\begin{pmatrix} \sum_i\sum_j 1(N_i \geqslant j)\{nh_1E(N)\}^{-1}K_1\left\{(T_{ij}-t)/h_1\right\} U_{ij}1(Y_i \leqslant \tilde{y}) \\ \sum_i\sum_j 1(N_i \geqslant j)\{nh_1E(N)\}^{-1}K_1\left\{(T_{ij}-t)/h_1\right\}\left\{(T_{ij}-t)/h_1\right\} U_{ij}1(Y_i \leqslant \tilde{y}) \end{pmatrix}.$$

Let $U_{ij}^*(t,\tilde{y}) = U_{ij}1(Y_i \leqslant \tilde{y}) - m(t,\tilde{y}) - m^{(1)}(t,\tilde{y})(T_{ij}-t)$, $W_n(z,t) = (1,0)S_n^{-1}(t)(1,z)^{\mathrm{T}}K_1(z)$.

Then

$$\hat{m}(t,\tilde{y}) - m(t,\tilde{y}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{M}\frac{1(N_i \geqslant j)}{h_1E(N)}\,W_n\left(\frac{T_{ij}-t}{h_1}, t\right)U_{ij}^*(t,\tilde{y}).$$

Denote a point between T_{ij} and t by t_{ij}^*; by Taylor expansion, $U_{ij}^*(\tilde{y}) = U_{ij}1(Y_i \leqslant \tilde{y}) - m(T_{ij},\tilde{y}) + m^{(2)}(t_{ij}^*,\tilde{y})(T_{ij}-t)^2/2$. Finally, let $e_{ij}(\tilde{y}) = U_{ij}1(Y_i \leqslant \tilde{y}) - m(T_{ij},\tilde{y})$. Then

$$\begin{aligned}\hat{m}(t,\tilde{y}) - m(t,\tilde{y}) = {}& \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{M}\frac{1(N_i \geqslant j)}{h_1E(N)f_T(t)}K_1\left(\frac{T_{ij}-t}{h_1}\right)e_{ij}(\tilde{y}) \\ & + \frac{1}{2n}\sum_{i=1}^{n}\sum_{j=1}^{M}\frac{h_1 1(N_i \geqslant j)}{E(N)f_T(t)}K_1\left(\frac{T_{ij}-t}{h_1}\right)\left(\frac{T_{ij}-t}{h_1}\right)^2 m^{(2)}(t_{ij}^*,\tilde{y}) + A_n(t,\tilde{y}),\end{aligned}$$

where

$$A_n(t,\tilde{y}) = \hat{m}(t,\tilde{y}) - m(t,\tilde{y}) - \{nh_1E(N)f_T(t)\}^{-1}\sum_i\sum_j 1(N_i \geqslant j)K_1\{(T_{ij}-t)/h_1\}U_{ij}^*(t,\tilde{y}).$$

This allows us to write $\hat{\Lambda}(s,t) - \Lambda(s,t) = I_{1n}(s,t) + I_{2n}(s,t) + I_{3n}(s,t)$, where

$$I_{1n}(s,t) = \frac{1}{n}\sum_{k=1}^{n}\left[m(s,Y_k)\{\hat{m}(t,Y_k) - m(t,Y_k)\} + m(t,Y_k)\{\hat{m}(s,Y_k) - m(s,Y_k)\}\right],$$

$$I_{2n}(s,t) = \frac{1}{n}\sum_{k=1}^{n}\left\{\hat{m}(s,Y_k) - m(s,Y_k)\right\}\left\{\hat{m}(t,Y_k) - m(t,Y_k)\right\},$$

$$I_{3n}(s,t) = \frac{1}{n}\sum_{k=1}^{n}m(s,Y_k)m(t,Y_k) - \Lambda(s,t),$$

which implies, by the Cauchy–Schwarz inequality, that $\|\hat{\Lambda} - \Lambda\|_{\mathrm{H}}^2 = O_{\mathrm{p}}(\|I_{1n}\|_{\mathrm{H}}^2 + \|I_{2n}\|_{\mathrm{H}}^2 + \|I_{3n}\|_{\mathrm{H}}^2)$. In the rest of the proofs, we drop the subscript H and the dummy variable in integrals for brevity. Recall that we defined Z_i as the underlying data quadruplet (T_i, U_i, Y_i, N_i). Further, let $\sum_{(p)} h_{i_1,\ldots,i_p}$ denote the sum of $h_{i_1,\ldots,i_p}$ over the permutations of $i_1,\ldots,i_p$. Finally, by Assumptions A1, A2 and A4, write $0 < B_T^{\mathrm{L}} \leqslant f(t) \leqslant B_T^{\mathrm{U}} < \infty$ for the lower and upper bounds of the density function of T, $|K_1(x)| \leqslant B_K < \infty$ for the bound on the kernel function K_1, and $|\partial^2 m/\partial t^2| \leqslant B_{2m} < \infty$ for the bound on the second partial derivative of $m(t,\tilde{y})$ with respect to t.

(a) We further decompose $I_{1n}(s,t)$ into $I_{1n}(s,t) = I_{11n}(s,t) + I_{12n}(s,t) + I_{13n}(s,t)$, where

$$I_{11n}(s,t) = \frac{1}{n^2}\sum_{k=1}^{n}\sum_{i=1}^{n}\sum_{j=1}^{M}\left\{\frac{1(N_i \geqslant j)}{h_1 E(N) f_T(t)} K_1\left(\frac{T_{ij}-t}{h_1}\right) e_{ij}(Y_k) m(s, Y_k) + \frac{1(N_i \geqslant j)}{h_1 E(N) f_T(s)} K_1\left(\frac{T_{ij}-s}{h_1}\right) e_{ij}(Y_k) m(t, Y_k)\right\}$$

$$I_{12n}(s,t) = \frac{1}{2n^2}\sum_{k=1}^{n}\sum_{i=1}^{n}\sum_{j=1}^{M}\left\{\frac{h_1 1(N_i \geqslant j)}{E(N) f_T(t)} K_1\left(\frac{T_{ij}-t}{h_1}\right)\left(\frac{T_{ij}-t}{h_1}\right)^2 m^{(2)}(t_{ij}^*, Y_k) m(s, Y_k) + \frac{h_1 1(N_i \geqslant j)}{E(N) f_T(s)} K_1\left(\frac{T_{ij}-s}{h_1}\right)\left(\frac{T_{ij}-s}{h_1}\right)^2 m^{(2)}(t_{ij}^*, Y_k) m(t, Y_k)\right\}$$

$$I_{13n}(s,t) = \frac{1}{n}\sum_{k=1}^{n}\left\{m(s, Y_k) A_n(t, Y_k) + m(t, Y_k) A_n(s, Y_k)\right\},$$

which we analyse individually below.

We first show that $E\|I_{11n}\|^2 = O(n^{-1}h_1^{-1})$. We write $I_{11n}(s,t)$ as

$$I_{11n}(s,t) = \frac{1}{2n^2}\sum_{k=1}^{n}\sum_{i=1}^{n}\sum_{(2)}\left\{h_{ik}(s,t) + h_{ik}(t,s)\right\} = \frac{1}{2n^2}\sum_{k=1}^{n}\sum_{i=1}^{n}\psi_n(Z_i, Z_k; s, t) = \frac{1}{2}v_n(s,t),$$

where $v_n(s,t)$ is a V-statistic with symmetric kernel $\psi_n(Z_i, Z_k; s, t)$ and

$$h_{ik}(s,t) = \sum_{j=1}^{M}\frac{1(N_i \geqslant j)}{h_1 E(N) f_T(t)} K_1\left(\frac{T_{ij}-t}{h_1}\right) e_{ij}(Y_k) m(s, Y_k).$$

Since $E\{e_{ij}(Y_k) \mid T_{ij}, Y_k\} = 0$, it is easy to show that $E\{h_{ik}(s,t)\} = E\{h_{ik}(t,s)\} = E\{h_{ki}(s,t)\} = E\{h_{ki}(t,s)\} = 0$. Thus $\theta_n(s,t) = E\{\psi_n(Z_i, Z_k; s, t)\} = 0$. Additionally,

$$\begin{aligned}\psi_{1n}(Z_i; s, t) &= E\{\psi_n(Z_i, Z_k; s, t) \mid Z_i\} \\ &= \sum_{j=1}^{M}\frac{1(N_i \geqslant j)}{h_1 E(N) f_T(t)} K_1\left(\frac{T_{ij}-t}{h_1}\right) E\{e_{ij}(Y_k) m(s, Y_k) \mid Z_i\} \\ &\quad + \sum_{j=1}^{M}\frac{1(N_i \geqslant j)}{h_1 E(N) f_T(s)} K_1\left(\frac{T_{ij}-s}{h_1}\right) E\{e_{ij}(Y_k) m(t, Y_k) \mid Z_i\}.\end{aligned}$$

If $E\{\psi_n^2(Z_i, Z_k; s, t)\} = o(n)$, Lemma A1 gives $nE\{v_n(s,t) - \tilde{u}_n(s,t)\}^2 = o(1)$, where $\tilde{u}_n(s,t) = 2n^{-1}\sum_{i=1}^{n}\psi_{1n}(Z_i; s, t)$ is the projection of the corresponding U-statistic. Since the projection of a U-statistic is a sum of independent and identically distributed random variables $\psi_{1n}(Z_i; s, t)$, $E\|I_{11n}\|^2 \leqslant 2n^{-1}\iint \mathrm{var}[E\{h_{ik}(s,t) \mid Z_i\}] + 2n^{-1}\iint \mathrm{var}[E\{h_{ik}(t,s) \mid Z_i\}] + o(n^{-1})$, where

$$\begin{aligned}&\frac{2}{n}\iint \mathrm{var}\left[E\{h_{ik}(s,t) \mid Z_i\}\right] \mathrm{d}s\, \mathrm{d}t \\ &\quad \leqslant \sum_{j=1}^{M}\frac{2P(N_i \geqslant j)}{nh_1^2 E(N)}\iint f_T^{-2}(t) E\left[K_1^2\left(\frac{T_{ij}-t}{h_1}\right) E^2\{e_{ij}(Y_k) m(s, Y_k) \mid Z_i\}\right] \mathrm{d}s\, \mathrm{d}t \\ &\quad = \sum_{j=1}^{M}\frac{2P(N_i \geqslant j)}{nh_1 E(N)}\iiint f_T^{-2}(t) K_1^2(u)\end{aligned}$$

$$\times E_{X_i,Y_i,\varepsilon_i}\left[E^2_{Y_k}\{e_{ij}(Y_k)m(s,Y_k)\mid T_{ij}=t+uh_1\}\right] f_T(t+uh_1)\,\mathrm{d}u\,\mathrm{d}s\,\mathrm{d}t$$

$$\to \sum_{j=1}^{M}\frac{2\|K_1\|^2P(N_i\geqslant j)}{nh_1E(N)}\iint f_T^{-1}(t)E_{X_i,Y_i,\varepsilon_i}\left[E^2_{Y_k}\{e_{ij}(Y_k)m(s,Y_k)\mid T_{ij}=t\}\right]$$

$$\leqslant \frac{8\|K_1\|^2}{nh_1B_T^{\mathrm{L}}E(N)}E\|X^4\|+\frac{4\|K_1\|^2\sigma^2}{nh_1B_T^{\mathrm{L}}E(N)}E\|X^2\|=O\left(\frac{1}{nh_1}\right),$$

where the first line follows from the Cauchy–Schwarz inequality, the second line is obtained by letting $u=h_1^{-1}(T_{ij}-t)$ and observing that T_{ij} is independent of X_i, Y_i and ε_i, and the third line follows from a variant of the dominated convergence theorem (Prakasa Rao, 1983, p. 35) that allows us to derive rates of convergence for nonparametric regression estimators. Thus $E\|I_{11n}\|^2=O(n^{-1}h_1^{-1})$, provided that $E\{\psi_n^2(Z_i,Z_k;s,t)\}=o(n)$ for all i and k, which we will show below. For $i\neq k$,

$$E\{\psi_n^2(Z_i,Z_k;s,t)\}=2E\{h_{ik}^2(s,t)\}+2E\{h_{ik}^2(t,s)\}+4E\{h_{ik}(s,t)h_{ik}(t,s)\}$$
$$+4E\{h_{ik}(s,t)h_{ki}(s,t)\}+4E\{h_{ik}(s,t)h_{ki}(t,s)\}.$$

Observe that

$$\frac{1}{n}E\{h_{ik}^2(s,t)\}=\sum_{j=l}^{M}\sum_{l=1}^{M}\frac{P\{N_i\geqslant\max(j,l)\}}{E^2(N)f_T^2(t)}$$
$$\times E\left\{(nh_1^2)^{-1}K_1\left(\frac{T_{ij}-t}{h_1}\right)K_1\left(\frac{T_{il}-t}{h_1}\right)e_{ij}(Y_k)e_{il}(Y_k)m^2(s,Y_k)\right\}.$$

For $j=l$, applying the dominated convergence theorem to the expectation on the right-hand side gives $n^{-1}h_1^{-1}\|K_1\|^2f_T(t)E\{e_{ij}^2(Y_k)m^2(s,Y_k)\mid T_{ij}=t\}$, and hence $n^{-1}E\{h_{ik}^2(s,t)\}=o(1)$ by Assumption A3. For $j\neq l$, a similar argument gives $n^{-1}f_T^2(t)E\{e_{ij}(Y_k)e_{il}(Y_k)m^2(s,Y_k)\mid T_{ij}=T_{il}=t\}$. The next two terms, $E\{h_{ik}^2(t,s)\}$ and $E\{h_{ik}(s,t)h_{ik}(t,s)\}$, can be handled similarly, as well as $E\{h_{ik}(s,t)h_{ki}(s,t)\}=o(n)$ and the case of $i=k$. Thus $E\{\psi_n^2(Z_i,Z_k;s,t)\}=o(n)$.

Using similar derivations, one can show that $E\|I_{12n}\|^2=O(h_1^4)+o(n^{-1})$.

We next show that $\|I_{13n}\|^2=O_{\mathrm{p}}(n^{-1}h_1+h_1^6)$. Following Lemma 2 of Martins-Filho & Yao (2007),

$$|A_n(t,Y_k)|=\left|\sum_{j=1}^{M}\sum_{i=1}^{n}\frac{1(N_i\geqslant j)}{nh_1E(N)}\left\{W_n\left(\frac{T_{ij}-t}{h_1},t\right)-f_T^{-1}(t)\,K_1\left(\frac{T_{ij}-t}{h_1}\right)\right\}U_{ij}^*(t,Y_k)\right|$$
$$\leqslant h_1^{-1}\left[(1,0)\left\{S_n^{-1}(t)-S^{-1}(t)\right\}^2(1,0)^{\mathrm{T}}\right]^{1/2}$$
$$\times\left\{\left|\sum_j\sum_i\frac{1(N_i\geqslant j)}{nE(N)}K_1\left(\frac{T_{ij}-t}{h_1}\right)U_{ij}^*(t,Y_k)\right|\right.$$
$$\left.+\left|\sum_j\sum_i\frac{1(N_i\geqslant j)}{nE(N)}K_1\left(\frac{T_{ij}-t}{h_1}\right)\left(\frac{T_{ij}-t}{h_1}\right)U_{ij}^*(t,Y_k)\right|\right\}$$
$$=h_1^{-1}\left[(1,0)\left\{S_n^{-1}(t)-S^{-1}(t)\right\}^2(1,0)^{\mathrm{T}}\right]^{1/2}R_n(t,Y_k).$$

Lemma A2 gives $\sup_{t\in\mathcal{T}}h_1^{-1}|[(1,0)\{S_n^{-1}(t)-S^{-1}(t)\}^2(1,0)^{\mathrm{T}}]^{1/2}|=O_{\mathrm{p}}(1)$. Next, $R_n(t,Y_k)\leqslant|R_{n1}(t,Y_k)|+|R_{n2}(t,Y_k)|+|R_{n3}(t,Y_k)|+|R_{n4}(t,Y_k)|$, where

$$R_{n1}(t,Y_k)=\sum_{j=1}^{M}\sum_{i=1}^{n}\frac{1(N_i\geqslant j)}{nE(N)}K_1\left(\frac{T_{ij}-t}{h_1}\right)e_{ij}(Y_k),$$

$$R_{n2}(t, Y_k) = \sum_{j=1}^{M} \sum_{i=1}^{n} \frac{h_1^2 1(N_i \geqslant j)}{2nE(N)} K_1\left(\frac{T_{ij} - t}{h_1}\right)\left(\frac{T_{ij} - t}{h_1}\right)^2 m^{(2)}(t_{ij}^*, Y_k),$$

$$R_{n3}(t, Y_k) = \sum_{j=1}^{M} \sum_{i=1}^{n} \frac{1(N_i \geqslant j)}{nE(N)} K_1\left(\frac{T_{ij} - t}{h_1}\right)\left(\frac{T_{ij} - t}{h_1}\right) e_{ij}(Y_k),$$

$$R_{n4}(t, Y_k) = \sum_{j=1}^{M} \sum_{i=1}^{n} \frac{h_1^2 1(N_i \geqslant j)}{2nE(N)} K_1\left(\frac{T_{ij} - t}{h_1}\right)\left(\frac{T_{ij} - t}{h_1}\right)^3 m^{(2)}(t_{ij}^*, Y_k).$$

Thus $n^{-1}\sum_k m(s, Y_k) R_{n1}(t, Y_k) = h_1 f_T(t) I_{11n}(s, t)$ leads to $\|h_1 f_T I_{11n}\|^2 = O_{\mathrm{p}}(n^{-1}h_1)$, and $n^{-1}\sum_k m(s, Y_k) R_{n2}(t, Y_k) = h_1 f_T(t) I_{12n}(s, t)$ leads to $\|h_1 f_T I_{12n}\|^2 = O_{\mathrm{p}}(h_1^6)$. It follows similarly that the third and fourth terms are $O_{\mathrm{p}}(n^{-1}h_1)$ and $O_{\mathrm{p}}(h_1^6)$, respectively. Hence, $\|I_{13n}\|^2 = O_{\mathrm{p}}(n^{-1}h_1 + h_1^6)$. Combining the previous results gives $\|I_{1n}\|^2 = O_{\mathrm{p}}\{(nh_1)^{-1} + h_1^4\}$.

(b) These terms are of higher order and are omitted for brevity.

(c) By the law of large numbers, $\|n^{-1}\sum_{i=1}^{n} m(\cdot, Y_i) m(\cdot, Y_i) - \Lambda\|^2 = O_{\mathrm{p}}(n^{-1})$.

Combining the above results leads to $\|\hat{\Lambda} - \Lambda\|^2 = O_{\mathrm{p}}(n^{-1}h_1^{-1} + h_1^4)$. □

Proof of Theorem 3. To facilitate the theoretical derivation, for each $k = 1, \ldots, K$ let $\eta_k = \Sigma^{1/2}\beta_k$ and $\hat{\eta}_{k,s_n} = \hat{\Sigma}_{s_n}^{1/2}\hat{\beta}_{k,s_n}$ be, respectively, the normalized eigenvectors of the equations $\Sigma^{-1}\Lambda\Sigma^{-1/2}\eta_k = \lambda_k\beta_k$ and $\hat{\Sigma}_{s_n}^{-1}\hat{\Lambda}\hat{\Sigma}_{s_n}^{-1/2}\hat{\eta}_{k,s_n} = \hat{\lambda}_{k,s_n}\hat{\beta}_{k,s_n}$. Then

$$\begin{aligned}\|\hat{\beta}_{k,s_n} - \beta_k\| &\leqslant \|\hat{\lambda}_{k,s_n}^{-1}\hat{\Sigma}_{s_n}^{-1}\hat{\Lambda}\hat{\Sigma}_{s_n}^{-1/2} - \lambda_k^{-1}\Sigma^{-1}\Lambda\Sigma^{-1/2}\| + \lambda_k^{-1}\|\Sigma^{-1}\Lambda\Sigma^{-1/2}\|\,\|\hat{\eta}_{k,s_n} - \eta_k\| \\ &\leqslant \hat{\lambda}_{k,s_n}^{-1}\|\hat{\Sigma}_{s_n}^{-1}\hat{\Lambda}\hat{\Sigma}_{s_n}^{-1/2} - \Sigma^{-1}\Lambda\Sigma^{-1/2}\| \\ &\quad + \|\Sigma^{-1}\Lambda\Sigma^{-1/2}\|\left(|\hat{\lambda}_{k,s_n}^{-1} - \lambda_k^{-1}| + \lambda_k^{-1}\|\hat{\eta}_{k,s_n} - \eta_k\|\right),\end{aligned}$$

using the fact that $\hat{\lambda}_{k,s_n}^{-1} \leqslant \lambda_k^{-1} + |\hat{\lambda}_{k,s_n}^{-1} - \lambda_k^{-1}|$. Applying standard theory for self-adjoint compact operators (Bosq, 2000) gives

$$\begin{aligned}|\hat{\lambda}_{k,s_n} - \lambda_k| &\leqslant \|\hat{\Sigma}_{s_n}^{-1/2}\hat{\Lambda}\hat{\Sigma}_{s_n}^{-1/2} - \Sigma^{-1/2}\Lambda\Sigma^{-1/2}\|, \\ \|\hat{\eta}_{k,s_n} - \eta_k\| &\leqslant C\|\hat{\Sigma}_{s_n}^{-1/2}\hat{\Lambda}\hat{\Sigma}_{s_n}^{-1/2} - \Sigma^{-1/2}\Lambda\Sigma^{-1/2}\| \quad (k = 1, \ldots, K),\end{aligned}$$

where $C > 0$ is a generic positive constant. Thus $\|\hat{\beta}_{k,s_n} - \beta_k\|^2 = O_{\mathrm{p}}(I_{1n} + I_{2n})$, where

$$I_{1n} = \left\|\hat{\Sigma}_{s_n}^{-1}\hat{\Lambda}\hat{\Sigma}_{s_n}^{-1/2} - \Sigma^{-1}\Lambda\Sigma^{-1/2}\right\|^2, \quad I_{2n} = \left\|\hat{\Sigma}_{s_n}^{-1/2}\hat{\Lambda}\hat{\Sigma}_{s_n}^{-1/2} - \Sigma^{-1/2}\Lambda\Sigma^{-1/2}\right\|^2.$$

It suffices to show that $I_{1n} = O_{\mathrm{p}}\{s_n^{3a+2}(n^{-1/2}h_1^{-1/2} + h_1^2) + s_n^{(4a+3)}(n^{-1/2}h_2^{-1} + h_2^2) + s_n^{-2b+1}\}$. The calculations for I_{2n} are similar and yield that $I_{2n} = o_{\mathrm{p}}(I_{1n})$.

Observe that $I_{1n} \leqslant 3I_{11n} + 3I_{12n} + 3I_{13n}$, where $I_{11n} = \|\Sigma_{s_n}^{-1}\Lambda\Sigma_{s_n}^{-1/2} - \Sigma^{-1}\Lambda\Sigma^{-1/2}\|^2$, $I_{12n} = \|\hat{\Sigma}_{s_n}^{-1}\Lambda\hat{\Sigma}_{s_n}^{-1/2} - \Sigma_{s_n}^{-1}\Lambda\Sigma_{s_n}^{-1/2}\|^2$ and $I_{13n} = \|\hat{\Sigma}_{s_n}^{-1}\hat{\Lambda}\hat{\Sigma}_{s_n}^{-1/2} - \hat{\Sigma}_{s_n}^{-1}\Lambda\hat{\Sigma}_{s_n}^{-1/2}\|^2$. Recall that $\Pi_{s_n} = \sum_{j=1}^{s_n}\phi_j \otimes \phi_j$ is the orthogonal projector onto the eigenspace associated with the s_n largest eigenvalues of Σ. Let I denote the identity operator and $\Pi_{s_n}^{\perp} = I - \Pi_{s_n}$ the operator perpendicular to Π_{s_n}, i.e., $\Pi_{s_n}^{\perp}$ is the orthogonal projector onto the eigenspace associated with eigenvalues of Σ that are less than α_{s_n}. Thus $\Sigma_{s_n}^{-1}\Lambda\Sigma_{s_n}^{-1/2} = \Pi_{s_n}\Sigma^{-1}\Lambda\Sigma^{-1/2}\Pi_{s_n}$ allows us to write $I_{11n} \leqslant \|\Pi_{s_n}^{\perp}\Sigma^{-1}\Lambda\Sigma^{-1/2}\| + \|\Sigma^{-1}\Lambda\Sigma^{-1/2}\Pi_{s_n}^{\perp}\|$. Since $\Sigma^{-1}\Lambda\Sigma^{-1/2}\eta_k = \lambda_k\beta_k$,

$$\|\Pi_{s_n}^{\perp}\Sigma^{-1}\Lambda\Sigma^{-1/2}\|^2 \leqslant \sum_{k=1}^{K}\lambda_k^2\left\|\sum_{i>s_n}\sum_{j=1}^{\infty} b_{kj}\langle\phi_i, \phi_j\rangle\phi_i\right\|^2 \leqslant \sum_{k=1}^{K}\lambda_k^2\sum_{i>s_n} b_{ki}^2$$

$$\leqslant C \sum_{k=1}^{K} \lambda_k^2 \sum_{i>s_n} i^{-2b} = O(s_n^{-2b+1});$$

similarly, $\|\Sigma^{-1}\Lambda\Sigma^{-1/2}\Pi_{s_n}^{\perp}\|^2 = O(s_n^{-2b+1})$.

We decompose I_{12n} as $I_{12n} \leqslant 3I_{121n} + 3I_{122n} + 3I_{123n}$, where $I_{121n} = \|(\Sigma_{s_n}^{-1} - \hat{\Sigma}_{s_n}^{-1})\Lambda\Sigma_{s_n}^{-1/2}\|^2$, $I_{122n} = \|\Sigma_{s_n}^{-1}\Lambda(\Sigma_{s_n}^{-1/2} - \hat{\Sigma}_{s_n}^{-1/2})\|^2$ and $I_{123n} = \|(\Sigma_{s_n}^{-1} - \hat{\Sigma}_{s_n}^{-1})\Lambda(\Sigma_{s_n}^{-1/2} - \hat{\Sigma}_{s_n}^{-1/2})\|^2$. Note that $I_{121n} \leqslant 6\|\Lambda\Sigma^{-1/2}\Pi_{s_n}\|^2(I_{1211n} + I_{1212n})$, where

$$I_{1211n} = \left\| \sum_{j=1}^{s_n} (\alpha_j^{-1} - \hat{\alpha}_j^{-1}) \hat{\phi}_j \otimes \hat{\phi}_j \right\|^2, \quad I_{1212n} = \left\| \sum_{j=1}^{s_n} \alpha_j^{-1} (\hat{\phi}_j \otimes \hat{\phi}_j - \phi_j \otimes \phi_j) \right\|^2.$$

Under Assumption 7, for all $1 \leqslant j \leqslant s_n$, $|\hat{\alpha}_j - \alpha_j| \leqslant \|\hat{\Sigma} - \Sigma\| \leqslant 2^{-1}(\alpha_j - \alpha_{j+1})$ implies that $\hat{\alpha}_j \geqslant 2^{-1}(\alpha_j + \alpha_{j+1}) \geqslant C^{-1} j^{-a}$, i.e., $\hat{\alpha}_j^{-1} \leqslant C j^a$ for some $C > 0$. Thus

$$I_{1211n} \leqslant \sum_{j=1}^{s_n} (\hat{\alpha}_j - \alpha_j)^2 (\alpha_j \hat{\alpha}_j)^{-2} \leqslant C\|\hat{\Sigma} - \Sigma\|^2 \sum_{j=1}^{s_n} j^{4a} = O_{\mathrm{p}}\left\{ s_n^{4a+1}(n^{-1}h_2^{-2} + h_2^4) \right\}.$$

For I_{1212n}, using the fact that $\|\hat{\phi}_j - \phi_j\| \leqslant 2\sqrt{2}\delta_j^{-1}\|\hat{\Sigma} - \Sigma\|$ (Bosq, 2000), where $\delta_1 = \alpha_1 - \alpha_2$ and $\delta_j = \min_{2 \leqslant \ell \leqslant j}(\alpha_{\ell-1} - \alpha_\ell, \alpha_\ell - \alpha_{\ell+1})$ for $j > 1$, we have that $\delta_j^{-1} \leqslant j^{a+1}$ and

$$I_{1212n} \leqslant 2 \sum_{j=1}^{s_n} \alpha_j^{-2} \|\hat{\phi}_j - \phi_j\|^2 \leqslant C\|\hat{\Sigma} - \Sigma\|^2 \sum_{j=1}^{s_n} j^{4a+2} = O_{\mathrm{p}}\left\{ s_n^{4a+3}(n^{-1}h_2^{-2} + h_2^4) \right\}.$$

Using $\Lambda\Sigma^{-1/2}\eta_k = \lambda_k \Sigma\beta_k$, we obtain $\|\Lambda\Sigma^{-1/2}\Pi_{s_n}\|^2 \leqslant \sum_{k=1}^{K} \lambda_k^2 \| \sum_{i=1}^{s_n} \alpha_i \sum_{j=1}^{\infty} b_{kj} \langle \phi_i, \phi_j \rangle \ \phi_i \|^2 \leqslant \sum_{k=1}^{K} \lambda_k^2 \sum_{i=1}^{s_n} \alpha_i^2 b_{ki}^2 < \infty$. Thus $I_{121n} = O_{\mathrm{p}}\{s_n^{4a+3}(n^{-1}h_2^{-2} + h_2^4)\}$. Using decompositions similar to the one for I_{121n}, both I_{122n} and I_{123n} can be shown to be $o_{\mathrm{p}}\{s_n^{4a+3}(n^{-1}h_2^{-2} + h_2^4)\}$. This leads to $I_{12n} = O_{\mathrm{p}}\{s_n^{4a+3}(n^{-1}h_2^{-2} + h_2^4)\}$.

Note that $I_{13n} \leqslant \|\hat{\Sigma}_{s_n}^{-1}\|^2 \|\hat{\Lambda} - \Lambda\|^2 \|\hat{\Sigma}_{s_n}^{-1/2}\|^2$, where $\|\hat{\Sigma}_{s_n}^{-1}\|^2 \leqslant \sum_{j=1}^{s_n} \hat{\alpha}_j^{-2} \leqslant C \sum_{j=1}^{s_n} j^{2a} = O_{\mathrm{p}}(s_n^{2a+1})$ and, similarly, $\|\hat{\Sigma}_{s_n}^{-1/2}\|^2 = O_{\mathrm{p}}(s_n^{a+1})$. From Theorem 2, we have $I_{13n} = O_{\mathrm{p}}\{s_n^{3a+2}(n^{-1}h_1^{-1} + h_1^4)\}$. Combining the above results leads to (6). □

REFERENCES

BOSQ, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*. New York: Springer.

CAI, T. T. & HALL, P. (2006). Prediction in functional linear regression. *Ann. Statist.* **34**, 2159–79.

CAMBANIS, S., HUANG, S. & SIMONS, G. (1981). On the theory of elliptically contoured distributions. *J. Mult. Anal.* **11**, 368–85.

CARDOT, H., FERRATY, F. & SARDA, P. (1999). Functional linear model. *Statist. Prob. Lett.* **45**, 11–22.

CHEN, D., HALL, P. & MÜLLER, H.-G. (2011). Single and multiple index functional regression models with nonparametric link. *Ann. Statist.* **39**, 1720–47.

CHIAROMONTE, F., COOK, D. R. & LI, B. (2002). Sufficient dimension reduction in regressions with categorical predictors. *Ann. Statist.* **30**, 475–97.

COOK, D. R. (1998). *Regression Graphics: Ideas for Studying Regressions through Graphics*. New York: Wiley.

COOK, D. R. & WEISBERG, S. (1991). Comment on "Sliced inverse regression for dimension reduction". *J. Am. Statist. Assoc.* **86**, 328–32.

COOK, D. R., FORZANI, L. & YAO, A.-F. (2010). Necessary and sufficient conditions for consistency of a method for smoothed functional inverse regression. *Statist. Sinica* **20**, 235–8.

DUAN, N. & LI, K.-C. (1991). Slicing regression: A link-free regression method. *Ann. Statist.* **19**, 505–30.

FAN, J. & GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. London: Chapman & Hall.

FERRÉ, L. & YAO, A.-F. (2003). Functional sliced inverse regression analysis. *Statistics* **37**, 475–88.

FERRÉ, L. & YAO, A.-F. (2005). Smoothed functional inverse regression. *Statist. Sinica* **15**, 665–83.

HALL, P. & HOROWITZ, J. L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.* **35**, 70–91.
HALL, P. & HOSSEINI-NASAB, M. (2006). On properties of functional principal components analysis. *J. R. Statist. Soc.* B **68**, 109–26.
HALL, P., MÜLLER, H.-G. & WANG, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.* **34**, 1493–517.
HASTIE, T. J. & TIBSHIRANI, R. J. (1990). *Generalized Additive Models*. London: Chapman & Hall.
HE, G., MÜLLER, H.-G. & WANG, J.-L. (2003). Functional canonical analysis for square integrable stochastic processes. *J. Mult. Anal.* **85**, 54–77.
JAMES, G. M. & SILVERMAN, B. W. (2005). Functional adaptive model estimation. *J. Am. Statist. Assoc.* **100**, 565–76.
JIANG, C.-R., YU, W. & WANG, J.-L. (2014). Inverse regression for longitudinal data. *Ann. Statist.* **42**, 563–91.
LI, B. & WANG, S. (2007). On directional regression for dimension reduction. *J. Am. Statist. Assoc.* **102**, 997–1008.
LI, K.-C. (1991). Sliced inverse regression for dimension reduction. *J. Am. Statist. Assoc.* **86**, 316–27.
LI, K.-C. (1992). On principal Hessian directions for data visualization and dimension reduction: Another application of Stein's lemma. *J. Am. Statist. Assoc.* **87**, 1025–39.
LI, Y. & HSING, T. (2010). Deciding the dimension of effective dimension reduction space for functional and high-dimensional data. *Ann. Statist.* **38**, 3028–62.
LIN, X. & CARROLL, R. J. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *J. Am. Statist. Assoc.* **95**, 520–34.
LIU, B. & MÜLLER, H.-G. (2009). Estimating derivatives for samples of sparsely observed functions, with application to on-line auction dynamics. *J. Am. Statist. Assoc.* **104**, 704–14.
MARTINS-FILHO, C. & YAO, F. (2006). A note on the use of V and U statistics in nonparametric models of regression. *Ann. Inst. Statist. Math.* **58**, 389–406.
MARTINS-FILHO, C. & YAO, F. (2007). Nonparametric frontier estimation via local linear regression. *J. Economet.* **141**, 283–319.
MÜLLER, H.-G. & STADTMÜLLER, U. (2005). Generalized functional linear models. *Ann. Statist.* **33**, 774–805.
PRAKASA RAO, B. L. S. (1983). *Nonparametric Functional Estimation*. Orlando, Florida: Academic Press.
RAMSAY, J. O. & SILVERMAN, B. W. (2005). *Functional Data Analysis*. New York: Springer, 2nd ed.
RICE, J. A. & SILVERMAN, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. R. Statist. Soc.* B **53**, 233–43.
XIA, Y., TONG, H., LI, W. & ZHU, L.-X. (2002). An adaptive estimation of dimension reduction space. *J. R. Statist. Soc.* B **64**, 363–410.
YAO, F. & MÜLLER, H.-G. (2010). Empirical dynamics for longitudinal data. *Ann. Statist.* **38**, 3458–86.
YAO, F., MÜLLER, H.-G. & WANG, J.-L. (2005a). Functional data analysis for sparse longitudinal data. *J. Am. Statist. Assoc.* **100**, 577–90.
YAO, F., MÜLLER, H.-G. & WANG, J.-L. (2005b). Functional linear regression analysis for longitudinal data. *Ann. Statist.* **33**, 2873–903.
YUAN, M. & CAI, T. T. (2010). A reproducing kernel Hilbert space approach to functional linear regression. *Ann. Statist.* **38**, 3412–44.
ZHU, L.-P., ZHU, L.-X. & FENG, Z.-H. (2010). Dimension reduction in regressions through cumulative slicing estimation. *J. Am. Statist. Assoc.* **105**, 1455–66.

[*Received May* 2014. *Revised January* 2015]