# Theory of functional principal components analysis for noisy and discretely observed data

Hang Zhou [1]

Department of Statistics
University of California, Davis

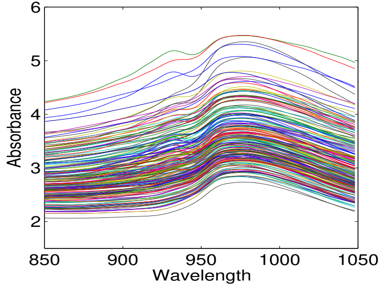
[1] jointly with Fang Yao and Dongyi Wei

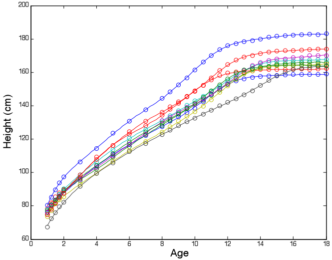
# Functional data

- A sample of subjects or experimental unit: one or more functions $X(t), t \in \mathcal{T}$, for each subject. WLOG, let $\mathcal{T} = [0, 1]$.
- Commonly adopted perspectives in FDA:
  - stochastic processes with smooth trajectories;
  - random element in a Hilbert space.
- Infinite dimensionality and smoothness
  - slowly diverging ranks, structural information ;
  - "bless of dimensionality": more measurements help, in contrast to high-dimensional data.

# Designs of functional data

- **Fully observed (ideal):** $X_i(t)$ available for all $t \in \mathcal{T}$.
- **Discretely observed (realistic):** *measurements are taken at discrete time points with noise:* $X_{ij} = X_i(t_{ij}) + \varepsilon_{ij}, i = 1, \ldots, n; j = 1 \ldots, N$.




- Q: How the discrete observations affect the estimation and convergence rate?

# Estimation of mean and covariance

- Mean function $\mu(t) = E\{X(t)\}$, $t \in \mathcal{T}$
- Covariance function $C(s,t) = \mathrm{cov}\{X(s), X(t)\}$, $s, t \in \mathcal{T}$
- Smoothing methods & strategies: kernel, spline, wavelets, ect.
  - pre-smoothing each curve before further analysis
  - pooling observations from all subjects
- Phase transition ($\sqrt{n}$-consistency) for mean and covariance estimation
  - pre-smoothing: $N \gtrsim n^{5/4}$
  - pooling: $N \gtrsim n^{1/4}$
- Related to the smoothness nature of functional data,
  no regularization is considered.

# Literature: smoothing mean and covariance

- **2005** — Pre-smoothing each curves before subsequent analysis (Ramsay and Silverman, 2005).
- **2005** — Pooling method for sparsely observed functional data (kernel) (Yao, Müller and Wang , 2005).
- **2010** — Uniform convergence rates for mean & covariance estimation (kernel) (Li and Hsing, 2010).
- **2011** — Phase transition for mean & covariance estimation (spline, RKHS) (Cai and Yuan, 2010 & 2011).
- **2016** — Unified theory for mean & covariance estimation (kernel) (Zhang and Wang, 2016).

Q: How the discretization and noise contamination affect the dimension reduction via FPCA?

# Representation of functional data

Assume that $X$ is a random process in $L^2(T)$ with covariance function

$$C(s,t) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s)\phi_k(t), \quad s,t \in \mathcal{T}$$

with the ordered eigenvalues $\lambda_1 > \lambda_2 > \ldots > 0$, $\lambda_k \asymp k^{-a}$, and orthonormal eigenfunctions $\phi_1, \phi_2, \ldots$.

- Karhunen-Loève expansion

$$X(t) = \mu(t) + \sum_{k=1}^{\infty} \xi_k \phi_k(t), \quad t \in \mathcal{T}$$

where $E(\xi_k) = 0,\ E(\xi_k^2) = \lambda_k,\ E(\xi_k \xi_\ell) = 0 \text{ for } k \neq \ell.$

# Infinite-dimensionality and regularization

- Infinite dimensionality:
  - The eigenvalues of $C(s,t)$ tend to zero and do not have a positive lower bound
  - The compact covariance operator $C(f) = \int_0^1 C(s,t)f(s)ds$ is non-invertible.
- Linear regression in $\mathbb{R}^d$: $Y_i = \langle X_i, \beta \rangle + \varepsilon_i, X_i, \beta \in \mathbb{R}^d$
  - Normal equation: $\beta = (\mathbb{E}X^T X)^{-1}\mathbb{E}X^T Y$
- Regularization is needed in models involving inverse issue
- Typical inverse problems: functional linear regression (FLR), generalized functional linear model (fGLM), functional Cox model, ...

# Eigenfunction with diverging index

- E.g., functional linear regression: $Y_i = \langle X_i, \beta \rangle + \varepsilon_i, X_i, \beta \in \mathcal{L}^2$
  - classical plug-in method (Hall and Horowitz, 2007):

$$\hat{\beta}(u) = \sum_{j=1}^{m} \hat{\lambda}_j^{-1} \left\langle n^{-1} \sum_{i=1}^{n} Y_i X_i, \hat{\phi}_j \right\rangle \hat{\phi}_j(u)$$

  - regularization: truncation on the number of FPCs
  - necessary to suppress approximation bias: $m = n^{\frac{1}{a+2b}} \to \infty$
- Highly depend on the convergence rate of a diverging number (with $n$) of eigenfunction estimates.
- Vanishing eigen-gap makes difficult for high-order estimates.
- The impact of discretely observed data is unknown.

# Timeline: theory of FPCA

**2006** — $\mathcal{L}^2$ bounds for a fixed number of eigenfunctions (kernel, $n^{-4/5}$) (Hall, Müller and Wang, 2006).

**2007** — Optimal rate $\|\hat{\phi}_k - \phi_k\|^2 = k^2/n$ obtained for fully observed functions (Hall and Horowitz, 2007).

**2009** — FPCA for reduced rank model (REML, spline, finite non-zero eigenvalues) (Paul and Peng, 2009).

**2010** — $\mathcal{L}^2$ bounds for a fixed number of eigenfunctions (RKHS, $\log n/n^{4/5}$, tensor space, trivial perturbation bound) (Cai and Yuan, 2010).

Optimal convergence for eigenfunctions with diverging index remains an **open problem** over a decade!

# Background

- Denote $\eta_k = |\lambda_k - \lambda_{k+1}|$ and $\Delta = \hat{C} - C$. Denote

$$m_n = \max\{k : \|\hat{C} - C\| < |\lambda_k - \lambda_{k+1}|/2\},$$

the first order expansion holds

$$\hat{\phi}_k - \phi_k \asymp \sum_{j \neq k} \frac{\langle (\hat{C} - C)\phi_k, \phi_j \rangle}{(\lambda_j - \lambda_k)} \phi_j \quad \text{for all } k \leq m_n \tag{1}$$

  - $m := m_n$ is the diverging number of eigenfunctions that can be well estimated based on the observed data
  - $\|\hat{C} - C\| \xrightarrow{p} 0$ as $n \to \infty$, thus $\lambda_{m_n} \to 0$ and $m_n \to \infty$.

- A classic bound can be derived from (1) and Bessel's equality directly:

$$\|\hat{\phi}_k - \phi_k\|^2 \leq \|\hat{C} - C\|^2/\eta_k^2, \qquad k \leq m. \tag{2}$$

- The bound $\|\hat{\phi}_k - \phi_k\|^2 \le \|\hat{C} - C\|^2/\eta_k^2$ is clearly not optimal.
- For finite $k$, this gives $\|\hat{\phi}_k - \phi_k\|$ a 2-d smoothing rate, however,
  - $\phi_k = \lambda_k^{-1} \int C(\cdot, t)\phi_k(t)\mathrm{d}t$, integration brings extra smoothness
- For diverging $k$, this gives $k^{2a+2}/n$
  - differ from the "fully observed" optimal rate by $k^{2a}/n$
- We shall resort to the original perturbation series rather than its approximation bound.

$$\|\hat{\phi}_k - \phi_k\|^2 \asymp \sum_{j \neq k} \frac{\langle (\hat{C} - C)\phi_k, \phi_j \rangle^2}{(\lambda_j - \lambda_k)^2}, \quad \text{for all } k \le m_n \tag{3}$$

# Sharp bound: fully observed

- For all $k \le m := m_n = \max\{l : \|\hat{C} - C\| < |\lambda_l - \lambda_{l+1}|/2\}$,

$$\|\hat{\phi}_k - \phi_k\|^2 \asymp \sum_{j \neq k} \frac{\langle (\hat{C} - C)\phi_k, \phi_j \rangle^2}{(\lambda_j - \lambda_k)^2}, \qquad k \le m.$$

- Fully observed: $\hat{C} = n^{-1} \sum_{i=1}^n X_i \otimes X_i$ (centered for simplicity),

$$\begin{aligned} \mathbb{E}\langle (\hat{C} - C)\phi_k, \phi_j \rangle^2 =& \mathbb{E}\left\{ \frac{1}{n} \sum_{i=1}^n (\xi_{i,k} - \bar{\xi}_k)(\xi_{i,j} - \bar{\xi}_j) \right\}^2 \\ =& \lambda_j \lambda_k (1 - n^{-1})^2 / n. \end{aligned}$$

- Reduction to $\lambda_j \lambda_k$ makes the summation converge w.r.t. $j$.
- $\|\hat{\phi}_k - \phi_k\|^2 = \sum_{j \neq k} \lambda_j \lambda_k / (\lambda_j - \lambda_k)^2 = O_p(k^2/n)$ is minimax optimal.

# Discretely observed

- Estimate $C$ by pooling kernel smoothing (Yao, Müller and Wang, 2005; Zhang and Wang, 2016).
- Without the covariance of fully observed curves, $\mathbb{E}\langle(\hat{C} - C)\phi_k, \phi_j\rangle^2$ is no longer the principal scores, but a kernel smoothing rate with bias $h^4$ and variance $n^{-1}$ (faster due to double integral).
- By Bessel's equality,

$$\mathbb{E}\|\hat{C} - C\|^2 = \sum_{k=1}^{\infty}\sum_{j=1}^{\infty}\mathbb{E}\langle(\hat{C} - C)\phi_k, \phi_j\rangle^2.$$

- One cannot sum up $\mathbb{E}\langle(\hat{C} - C)\phi_k, \phi_j\rangle^2/(\lambda_j - \lambda_k)^2$ w.r.t. all $j \neq k$ in the perturbation series directly.

# Key assumption

- Hall, Müller and Wang (2006) assumes that

$$\max_{1\le j\le r}\max_{s=0,1,2}\sup_{t\in[0,1]}|\phi_j^{(s)}(t)| \le \mathrm{Const}.$$

  which is only valid for a fixed $r$.

- Fourier basis: $\phi_j(x) = \cos(j\pi x), \phi_j^{(1)}(x) = -j\sin(j\pi x)$.
- Generalize this for diverging $j$

$$\|\phi_j^{(s)}(t)\|_\infty \lesssim j^{c/2}\|\phi_j^{(s-1)}\|_\infty \text{ for } s = 1, 2,$$

  - Higher frequency of $\phi_j$ for larger $j$, characterized by the amplitudes of its derivatives.
  - E.g., Fourier basis: $c = 2$.

# Building block

- $\sum_{\{j:j\neq k\}} \langle(\hat{C}-C)\phi_k,\phi_j\rangle^2/(\lambda_j-\lambda_k)^2$ is dominated by summation on $\{j: k/2 < j \neq k \leq 2k\}$ for each $k$.
- This inspires us to treat $\sum_{\{j:j>2k\}} E\langle(\hat{C}-C)\phi_k,\phi_j\rangle^2$ as a whole.

**Theorem 1**

*Denote* $\Delta = \hat{C} - C$, *if* $h^4 k^{2a+2c} = O(1)$

$$\mathbb{E}(\langle\Delta\phi_j,\phi_k\rangle^2) \lesssim \frac{1}{n}\left(j^{-a}k^{-a} + \frac{j^{-a}+k^{-a}}{N} + \frac{1}{N^2}\right) + h^4 k^{2c-2a}, \quad \textit{for } 1 \leq j \leq 2k$$

$$\sum_{j=2k+1}^{\infty} \mathbb{E}(\langle\Delta\phi_j,\phi_k\rangle^2) \lesssim \frac{1}{n}\left(k^{1-2a} + \frac{h^{-1}k^{-a}+k^{1-a}}{N} + \frac{1}{hN^2}\right) + h^4 k^{1+2c-2a}.$$

# Main results

## Theorem 2

*Let* $\Omega(n, N, h) := \{k : \|\Delta\| \le \eta_k/2, h^4 k^{2a+2c} \le C\}$*, denote* $m := \max\{k : k \in \Omega(n, N, h)\}$*. If* $h^4 m^{2a+2c} = O(1)$*,* $\frac{m^{2a+2}}{n} \to 0$*,* $\frac{m^{2a+2}}{nN^2h^2} \to 0$*,* $h^4 m^{2a+2} \to 0$*, then* $\mathbb{P}(\Omega) \to 1$*, for all* $k \le m$*,*

$$\mathbb{E}\|\hat{\phi}_k - \phi_k\|^2 \lesssim \frac{k^2}{n}\left\{1 + \left(\frac{k^a}{N}\right)^2\right\} + \frac{k^a}{nNh}\left(1 + \frac{k^a}{N}\right) + h^4 k^{2c+2}.$$

- *optimal rate in fully observed case*
- *variance term caused by kernel smoother*
- *error caused by decaying eigengaps*
- *bias term caused by kernel smoother*

# Phase transition

**Corollary 2.1**

*Let* $m \in \mathbb{N}_+$ *satisfies (M.1). For all* $k \leq m$ *and let* $h_{opt}(k) = (nN)^{-1/5} k^{(a-2c-2)/5} (1 + k^a/N)^{1/5}$,

1. *If* $N \gtrsim k^a$,

$$\mathbb{E}(\|\hat{\phi}_k - \phi_k\|^2) \lesssim \frac{k^2}{n} + \frac{k^{(4a+2c+2)/5}}{(nN)^{4/5}}.$$

*In addition, if* $N \geq n^{1/4} k^{a+c/2-2}$, $\mathbb{E}(\|\hat{\phi}_k - \phi_k\|^2) \lesssim k^2/n$.

2. *If* $N = o(k^a)$,

$$\mathbb{E}(\|\hat{\phi}_k - \phi_k\|^2) \lesssim \frac{k^{2a+2}}{nN^2} + \frac{k^{(8a+2c+2)/5}}{(nN^2)^{4/5}}.$$

# Phase transition

- Minimax optimality
  - For fixed $k$, this becomes the optimal 1-d rate $n^{-1}(1+(Nh)^{-1})+h^4$;
  - For $k\to\infty$, if $N\gtrsim\max\{k^a, n^{1/4}k^{a+c/2-2}\}$, $\mathbb{E}(\|\hat{\phi}_k-\phi_k\|^2)\lesssim k^2/n$.
- For the Fourier basis with $c=2$
  - Fixed $k$, phase transition occurs at $n^{1/4}$ (the same as mean and cov)
  - Diverging $k$, phase transition occurs at $n^{1/4}k^{a-1}$
  - $n^{1/4}k^{a-1}$ is slightly larger than $n^{1/4}$, reflecting efficiency of pooling smoothing with evaluated difficulty in FPCA.
- Fundamental result for inverse models (e.g., FLR), where the optimal bandwidth can be chosen as $h_{opt}(k_{\max})$.

# Relative asymptotic normality

## Theorem 3

*For* $m \in \mathbb{N}_+$ *satisfying* $h(m^{2c} + m^a) = o(1)$ *and*

$$\sqrt{n}(m^a + N)[m^2 n^{-1}\{1 + (j^a/N)^2\} + m^a (nNh)^{-1}(1 + m^a/N) + h^4 m^{2c+2}] = o(1).$$

$$\Sigma_n^{-\frac{1}{2}} \left( \frac{\hat{\lambda}_j - \lambda_j}{\lambda_j} - 2\lambda_j \sigma_K^2 h^2 \int_h^{1-h} \phi_j^{(2)}(u)\phi_j(u)\mathrm{d}u \right) \xrightarrow{d} \mathcal{N}(0,1), \; j \leq m,$$

*where*

$$\Sigma_n = \frac{1}{n}\left[ \frac{(N-2)(N-3)}{N(N-1)} \frac{\mathbb{E}(\xi_j^4)}{\lambda_j^2} + \frac{4(N-2)}{N(N-1)} \frac{\mathbb{E}\{\xi_j^2(\|X\phi_j\|^2 + \sigma_X^2)\}}{\lambda_j^2} + \frac{2}{N(N-1)} \frac{\mathbb{E}\{(\|X\phi_j\|^2 + \sigma_X^2)^2\}}{\lambda_j^2} - 1 \right].$$

# Phase transition

**Corollary** 4.1

1. *If* $N\lambda_j \to \infty$, $\sqrt{n}\lambda_j h^2 \int_h^{1-h} \phi_j^{(2)}(u)\phi_j(u)\mathrm{d}u \to 0$,
$$\sqrt{n}\left(\frac{\hat{\lambda}_j - \lambda_j}{\lambda_j}\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\mathbb{E}(\xi_j^4) - \lambda_j^2}{\lambda_j^2}\right).$$
2. *If* $N\lambda_j \to C_1$,
$$\sqrt{n}\left(\frac{\hat{\lambda}_j - \lambda_j}{\lambda_j} - 2\lambda_j\sigma_K^2 h^2 \int_h^{1-h} \phi_j^{(2)}(u)\phi_j(u)\mathrm{d}u\right)$$
$$\xrightarrow{d} \mathcal{N}\left(0, \frac{\mathbb{E}(\xi_j^4) - \lambda_j^2}{\lambda_j^2} + \frac{4\mathbb{E}\{\xi_j^2(\|X\phi_j\|^2 + \sigma_X^2)\}}{C_1\lambda_j} + \frac{2\mathbb{E}\{(\|X\phi_j\|^2 + \sigma_X^2)^2\}}{C_1^2}\right).$$
3. *If* $N\lambda_j \to 0$,
$$\sqrt{n}N\lambda_j\left(\frac{\hat{\lambda}_j - \lambda_j}{\lambda_j} - 2\lambda_j\sigma_K^2 h^2 \int_h^{1-h} \phi_j^{(2)}(u)\phi_j(u)\mathrm{d}u\right)$$
$$\xrightarrow{d} \mathcal{N}\left(0, \frac{2N}{N-1}\mathbb{E}\{(\|X\phi_j\|^2 + \sigma_X^2)^2\}\right).$$

# Uniform convergence of eigenfunctions

**Theorem 4**

*Let* $k_{\max} \in \mathbb{N}_+$ *satisfy* $hk_{\max} \le 1$, $h^4 k_{\max}^{2a+2c} \le 1$ *and* $\mathbb{P}(\Omega_u) \to 1$ *with* $\Omega_u := \{\|\Delta\|_{HS} \le \eta_{k_{\max}}/2,\ \|\Delta\|_\infty k_{\max}^a \le 1,\ \|\Delta\|_{HS} k_{\max}^{a+1} \le 1\}$, *for* $k \le k_{\max}$,

$$\mathbb{E}(\|\hat{\phi}_k - \phi_k\|_\infty) \lesssim \frac{k}{\sqrt{n}}(\sqrt{\ln n} + \ln k)\left\{1 + \frac{k^a}{N} + \sqrt{\frac{k^{a-1}}{Nh}}\left(1 + \sqrt{\frac{k^a}{N}}\right)\right\} + k^a \left|\frac{\ln n}{n}\right|^{1-\frac{1}{\beta}} \left|k^{1/2} + \frac{\ln n}{Nh}\right|^{1-\frac{1}{\beta}} h^{-\frac{1}{\beta}} + h^2 k^{c+1} \log k. \tag{4}$$

- When $\beta > 5/2$, the truncation bias is dominated by the other terms.
- When $N$ is sufficiently large, we may obtain the optimal uniform rate.

# Functional linear regression

- $Y_i$: scalar response
$$Y_i = \int_{\mathcal{T}} X_i(t)\beta(t)\mathrm{d}t + \varepsilon_i$$
$\beta \in \mathcal{L}^2[0,1]$, $\mathbb{E}\varepsilon_i = 0$ and $\mathbb{E}\varepsilon_i^2 < \infty$.
- $\beta = \sum_{k=1}^{\infty} b_k \phi_k$, $|b_k| \le Ck^{-b}$, $b > a/2 + 1$
- Plug-in method (Hall and Horowitz, 2007):
$$\sum_{j=1}^{m} \hat{\lambda}_j^{-1} \left\langle \frac{1}{n} \sum_{i=1}^{n} Y_i X_i, \hat{\phi}_j \right\rangle \hat{\phi}_j(u)$$
- To suppress the truncation bias: $m \asymp n^{1/(a+2b)} \to \infty$.

# Timeline: theory of FLR

**2005** Function on function regression based on sparse FPCA (Yao, Müller and Wang, 2005).

**2007** The optimal rate of FLM for fully observed data (Hall and Horowitz, 2007).

**2010** The optimal estimation and prediction of FLM for fully observed data by RKHS (Cai and Yuan, 2010 and 2012).

**2012** The optimal rate of fGLM for fully observed data by FPCA and change of measure (Dou, Pollard and Zhou, 2012).

**2022** Phase transition in FLM for discrete observed data by improved FPCA results (Zhou, Yao and Zhang, 2022).

# Application in FLR

- Apply our results in the Plug-in estimator,

| $b \geq a - c$ | | $b < a - c$ | |
|---|---|---|---|
| Sampling rate | $\|\hat{\beta} - \beta\|^2$ | Sampling rate | $\|\hat{\beta} - \beta\|^2$ |
| $N > n^{\frac{1}{4}+\frac{a+2c}{4(a+2b)}}$ | $n^{\frac{1-2b}{a+2b}}$ | $N > n^{\frac{1}{4}+\frac{3a-2b}{4(a+2b)}}$ | $n^{\frac{1-2b}{a+2b}}$ |
| $n^{\frac{2a}{a+5b+c}} \leq N < n^{\frac{a+b+c}{2a+4b}}$ | $(nN)^{\frac{2(1-2b)}{3a+5b+c}}$ | $n^{\frac{a+b+c}{3a+3b-c}} \leq N < n^{\frac{a}{a+2b}}$ | $(nN)^{\frac{(1-2b)}{2a+2b}}$ |
| $N < n^{\frac{2a}{a+5b+c}}$ | $(nN^2)^{\frac{2(1-2b)}{5a+5b+c}}$ | $N < n^{\frac{a+b+c}{3a+3b-c}}$ | $(nN^2)^{\frac{2(1-2b)}{5a+5b+c}}$ |

- The transition point is slightly larger than $n^{1/4}$ but smaller than $n^{1/2}$, reflecting the elevated difficulty due to FPC regularization.
- Result improved compared to Zhou, Yao and Zhang (2022).

# References


- Bhatia, R., Davis, C., & McIntosh, A. (1983). Perturbation of spectral subspaces and solution of linear operator equations. *Linear Algebra Appl.*, **52**, 45-67.
- Yao, F., Müller, H. G., & Wang, J. L. (2005). Functional data analysis for sparse longitudinal data. *JASA*, **100(470)**, 577-590.
- Yao, F., Müller, H. G., & Wang, J. L. (2005). Functional linear regression analysis for longitudinal data. *Ann. Statist.*, 2873-2903.
- Ramsay, J. & Silverman, B. (2005). Functional Data Analysis. Springer Science & Business Media.
- Hall, P., Müller, H. G., & Wang, J. L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.*, **34**, 1493-1517.
- Hall, P. & Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.*, **35**, 70-91.
- Paul, D., & Peng, J. (2009). Consistency of restricted maximum likelihood estimators of principal components. *Ann. Statist.*, 1229-1271.

# References


- Cai, T., & Yuan, M. (2010). Nonparametric covariance function estimation for functional and longitudinal data. Technical report, UPenn and Georgia Tech.
- Cai, T. T., & Yuan, M. (2011). Optimal estimation of the mean function based on discretely sampled functional data: Phase transition. *Ann. Statist.*, 2330-2355.
- Cai, T. T., & Yuan, M. (2012). Minimax and adaptive prediction for functional linear regression. *JASA*, **107(499)**, 1201-1216..
- Dou, W. W., Pollard, D., & Zhou, H. H. (2012). Estimation in functional regression for general exponential families *Ann. Statist.*, **40**, 2421-2451.
- Li, Y. & Hsing, T. (2016). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Ann. Statist.*, **38**, 3321-3351.
- Zhang, X. & Wang, J.-L. (2016). From sparse to dense functional data and beyond. *Ann. Statist.*, **44**, 2281-2321.
- Zhou, H., Yao, F., & Zhang, H. (2022). Functional linear regression for discretely observed data: from ideal to reality. *Biometrika*, doi.org/10.1093/biomet/asac053.

Thank You!

# Uniform convergence of covariance

- Uniform convergence rate of
  $L(s,t) = \sum_{i=1}^{n} v_i \sum_{1\le j\ne l\le N_i} K\left(\frac{T_{ij}-s}{h_\gamma}\right) K\left(\frac{T_{il}-t}{h_\gamma}\right) X_{ij} X_{il}$,
- Discretization: Let $\chi(\rho)$ be an mesh grid on $[0,1]^2$ with grid size $n^{-\rho}$

$$\sup_{s,t\in[0,1]} |L(s,t) - \mathbb{E}L(s,t)| \le \sup_{s,t\in\chi(\gamma)} |L(s,t) - \mathbb{E}L(s,t)| + o_p$$

- Truncation:
  $L^*(s,t) = \sum_{i=1}^{n} v_i \sum_{1\le j\ne l\le N_i} K\left(\frac{T_{ij}-s}{h_\gamma}\right) K\left(\frac{T_{il}-t}{h_\gamma}\right) X_{ij} X_{il} \boxed{\mathbf{1}_{\{|X_{ij}X_{il}|\le B_n\}}}$,
- Bernstein inequality:

$$P(\sup_{s,t} |L^*(s,t) - \mathbb{E}L^*(s,t)| > Mb_n) \le 2n^\rho \exp\left(-\frac{M^2 b_n^2}{M_U' b_n^2/\log n + 2B' M_K^2 B_n/n}\right)$$

with $b_n = \left\{\log(n)\left[\sum_{i=1}^{n} N_i(N_i-1) v_i^2 h_\gamma^2 + \sum_{i=1}^{n} N_i(N_i-1)(N_i-2) v_i^2 h_\gamma^3 + \sum_{i=1}^{n} N_i(N_i-1)(N_i-2)(N_i-3) v_i^2 h_\gamma^4\right]\right\}^{1/2}$ and $B_n = b_n[n/\log(n)]$.

# Issue with uniform convergence

- Assume $\mathbb{E}\|X\|_\infty^{2\beta} < \infty$, the following is needed to suppress the truncation bias (Li and Hsing, 2010, Zhang and Wang (2016)),

$$\left(N^{-2}h^2 + N^{-1}h^3 + h^4\right)\left(\frac{\log n}{n}\right)^{2/\beta-1} \longrightarrow \infty.$$

- For sparse case $N/(n/\log n)^{1/4} \to 0$ and $h \asymp (nN^2/\log n)^{-1/6}$, let $N \asymp n^{\tau/4}$

$$\frac{h^2}{N^2}\left(\frac{\log n}{n}\right)^{2/\beta-1} \to \infty \Longrightarrow \beta > \frac{3}{1-\tau}.$$

- For dense case $N/(n/\log n)^{1/4} \gtrsim C$ and $h \asymp (n/\log n)^{-1/4}$, this leads to a contradiction

$$\frac{h^2}{N^2}\left(\frac{\log n}{n}\right)^{2/\beta-1} \leq \left(\frac{\log n}{n}\right)^{2/\beta} \longrightarrow 0$$

# Double truncation

- The bound $L_i^*(s,t) \le B_n/n$ is too loose.
- Denote $J_i(s) = h^{-1}\sum_{l=1}^N K\left(\frac{T_{il}-s}{h}\right)$, given $M > 0$

$$\mathbb{P}\left(J_i(s) > M\right) \le \exp(-MNh/3)$$

- $$\begin{aligned} & h^{-2}L_i^*(s,t) \\ =& \frac{1}{n}\frac{h^{-2}}{N(N-1)}\sum_{1\le l_1\ne l_2\le N} \mathrm{K}\left(\frac{T_{il_1}-s}{h}\right)\mathrm{K}\left(\frac{T_{il_2}-t}{h}\right)X_{il_1}X_{il_2}\mathbf{1}_{(|X_{il_1}X_{il_2}|\le B_n)} \\ \le& B_n\frac{1}{n}\frac{N}{N-1}J_i(s)J_i(t). \end{aligned}$$

- $\mathbb{P}\left(|L_i^*(s,t)| > B_n h^2 \frac{1}{n}\frac{NM^2}{N-1}\right) \le 2\exp(-MNh/3)$.
- Double truncation: $\widetilde{L}_i^*(s,t) = L_i^*(s,t)\,\boxed{\mathbf{1}_{(|L_i^*(s,t)|\le B_n h^2 \frac{1}{n}\frac{NM^2}{N-1})}}$

- After twice truncation $\tilde{C}_*(s,t) = h^{-2}\sum_{i=1}^{n}\widetilde{L}_i^*(s,t)$.

$$\mathbb{E}\left[\sup_{s,t\in[0,1]}|\hat{C}(s,t)-\mathbb{E}\hat{C}(s,t)|\right] \le \mathbb{E}\left[\sup_{(s,t)\in\chi(\rho)}|\tilde{C}_*(s,t)-\mathbb{E}\tilde{C}_*(s,t)|\right]$$

$$+ \boxed{\mathbb{E}|D_1+D_2|} + \boxed{\mathbb{E}|E_1+E_2|} + \boxed{\mathbb{E}|F_1+F_2|}$$

$$\le \text{Const.}\left[\sqrt{\frac{\rho\ln n}{n}}\left(1+\frac{1}{Nh}\right) + B_n\frac{1}{n}M^2\rho\ln n\right] + \boxed{\text{Const.}n^{-\rho}h^{-3}}$$

$$+ \boxed{\text{Const.}B_n^{1-\beta}h^{-2}} + \boxed{\text{Const.}n^{2\rho}\left(1+\frac{1}{Nh}\right)\exp(-MNh/6)}.$$

- We can choose larger $B_n$ to eliminate the model bias. Go Back!

# Uniform convergence of covariance

- We propose a double truncation technique, and derive a sharper bound for the truncated bias (details).

**Theorem 5**

$$\mathbb{E}\left\{\sup_{s,t\in[0,1]}\left|\hat{C}(s,t)-\mathbb{E}\hat{C}(s,t)\right|\right\}$$
$$\lesssim \boxed{\sqrt{\frac{\ln n}{n}}\left(1+\frac{1}{Nh}\right)} + \boxed{\left|\frac{\ln n}{n}\right|^{1-\frac{1}{\beta}}\left|1+\frac{\ln n}{Nh}\right|^{2-\frac{2}{\beta}} h^{-\frac{2}{\beta}}}.$$

- When $\beta > 3$, the truncation bias (yellow) is dominated by main term (green) for all $N$.
- $\sqrt{\log n/n}$-convergence can be obtained for dense data, where the phase transition occurs when $N \gtrsim (n/\log n)^{1/4}$.

# Uniform convergence of eigenfunctions

**Theorem 6**

*Let* $k_{\max} \in \mathbb{N}_+$ *satisfy* $hk_{\max} \leq 1$, $h^4 k_{\max}^{2a+2c} \leq 1$ *and* $\mathbb{P}(\Omega_u) \to 1$ *with* $\Omega_u := \{\|\Delta\|_{HS} \leq \eta_{k_{\max}}/2,\ \|\Delta\|_\infty k_{\max}^a \leq 1,\ \|\Delta\|_{HS} k_{\max}^{a+1} \leq 1\}$, *for* $k \leq k_{\max}$,

$$\mathbb{E}(\|\hat{\phi}_k - \phi_k\|_\infty) \lesssim \frac{k}{\sqrt{n}}(\sqrt{\ln n} + \ln k)\left\{1 + \frac{k^a}{N} + \sqrt{\frac{k^{a-1}}{Nh}}\left(1 + \sqrt{\frac{k^a}{N}}\right)\right\} + k^a \left|\frac{\ln n}{n}\right|^{1-\frac{1}{\beta}} \left|k^{1/2} + \frac{\ln n}{Nh}\right|^{1-\frac{1}{\beta}} h^{-\frac{1}{\beta}} + h^2 k^{c+1} \log k. \tag{5}$$

- When $\beta > 5/2$, the truncation bias is dominated by the other terms.
- When $N$ is sufficiently large, we may obtain the optimal uniform rate.