Journal of Machine Learning Research 1 (2023) 1-33 Submitted */*; Revised */*; Published */*

From sparse to dense functional data in high dimensions: Revisiting phase transitions from a non-asymptotic perspective

Shaojun Guo SJGUO@RUC.EDU.CN
Institute of Statistics and Big Data
Renmin University of China
Beijing 100872, China

Dong Li MALIDONG@TSINGHUA.EDU.CN
Center for Statistical Science
Department of Industrial Engineering
Tsinghua University
Beijing 100084, China

Xinghao Qiao X.QIAO@LSE.AC.UK
Department of Statistics
London School of Economics
London, WC2A 2AE, U.K.

Yizhu Wang WANGYIZH20@MAILS.TSINGHUA.EDU.CN
Department of Mathematical Sciences
Tsinghua University
Beijing 100084, China

Editor:

Abstract

Nonparametric estimation of the mean and covariance functions is ubiquitous in functional data analysis and local linear smoothing techniques are most frequently used. Zhang and Wang (2016) explored different types of asymptotic properties of the estimation, which reveal interesting phase transition phenomena based on the relative order of the average sampling frequency per subject T to the number of subjects n, partitioning the data into three categories: "sparse", "semi-dense", and "ultra-dense". In an increasingly available high-dimensional scenario, where the number of functional variables p is large in relation to n, we revisit this open problem from a non-asymptotic perspective by deriving comprehensive concentration inequalities for the local linear smoothers. Besides being of interest by themselves, our non-asymptotic results lead to elementwise maximum rates of L_2 convergence and uniform convergence serving as a fundamentally important tool for further convergence analysis when p grows exponentially with n and possibly T. With the presence of extra $\log p$ terms to account for the high-dimensional effect, we then investigate the scaled phase transitions and the corresponding elementwise maximum rates from sparse to semi-dense to ultra-dense functional data in high dimensions. Finally, numerical studies are carried out to confirm our established theoretical properties.

Keywords: Concentration inequalities, high-dimensional partially observed functional data, elementwise maximum rates, local linear smoother, mean and covariance functions.

1 Introduction

A fundamental issue in functional data analysis is the nonparametric estimation of the mean and covariance functions based on discretely sampled and noisy curves. The estimated quantities are not only of interest by themselves but also serve as building blocks for dimension reduction and subsequent modeling of functional data, such as functional principal component analysis (FPCA) (Yao et al., 2005a; Li and Hsing, 2010) and functional linear regression (Yao et al., 2005b). Among candidate nonparametric smoothers, we focus on the most commonly-adopted local linear smoothing method due to its simplicity and attractive local and boundary correction properties.

In a typical functional data setting, we have n random curves, representing n subjects, observed with errors, at T_i randomly sampled time points for the ith subject. The sampling frequency T_i plays a pivotal role in the estimation, as it may affect the choice of the estimation procedure. The literature can be loosely divided into two categories. The first category corresponds to dense functional data, where T_i's are larger than some order of n. A conventional approach to handle such data implements nonparametric smoothing to the observations from each subject to eliminate the noise, thus reconstructing each individual curve before subsequent analysis (Zhang and Chen, 2007). The second category referred to as sparse functional data, accords with bounded T_i's. Under such a scenario, the pre-smoothing step is no longer applicable, an alternative pooling strategy considers pooling the data from all subjects to build strength across all observations (Yao et al., 2005a; Li and Hsing, 2010). More recently, Zhang and Wang (2016) provided a comprehensive analysis of phase transitions and the associated rates of convergence for three types of asymptotic properties: local asymptotic normality, L_2 convergence, and uniform convergence. They proposed to further partition dense functional data into new categories: "semi-dense" and "ultra-dense", depending on whether the root-n rate is achieved with negligible asymptotic bias or not. However, these aforementioned asymptotic results are only suitable for handling univariate or low-dimensional multivariate functional data.

With recent advances in data collection technology, high-dimensional functional datasets become increasingly available. Examples include time-course gene expression data, and electroencephalography and functional magnetic resonance imaging data, where signals are measured over time at a large number of regions of interest (Zhu et al., 2016; Li and Solea, 2018; Zapata et al., 2022; Fang et al., 2023). Those data can be represented as a p-vector of random functions $\mathbf{X}_i(\cdot) = \{X_{i1}(\cdot), \ldots, X_{ip}(\cdot)\}^{\mathrm{T}}$ for $i = 1, \ldots, n$ defined on a compact set $\mathcal{U}$, with p-vector of mean functions $\boldsymbol{\mu}(\cdot) = \{\mu_1(\cdot), \ldots, \mu_p(\cdot)\}^{\mathrm{T}} = \mathbb{E}\{\mathbf{X}(\cdot)\}$ and $(p \times p)$-matrix of marginal- and cross-covariance functions

$$\boldsymbol{\Sigma}(u, v) = \{\Sigma_{jk}(u, v)\}_{p \times p}, \quad \Sigma_{jk}(u, v) = \mathrm{cov}\{X_{ij}(u), X_{ik}(v)\}.$$

In a high-dimensional regime, the dimension p can be diverging with, or even larger than, the number of subjects n. In practice, each $X_{ij}(\cdot)$ is observed subject to error contamination at T_{ij} random time points; see (3) below.

Within the high-dimensional statistical learning framework, it is essential to conduct non-asymptotic analysis of the estimators by developing concentration inequalities under a given performance metric, which can lead to probabilistic error bounds in the elementwise maximum norm as a function of n, p, and possibly T_{ij}'s under our setup. Existing literature has mainly focused on fully observed functional data, based on which concentration inequalities for the estimated covariance functions were established in Qiao et al. (2019) and Zapata et al. (2022). In practical scenarios where curves are partially observed with errors, addressing dense functional data is achievable by applying the pre-smoothing technique to observations from each i, j (Kong et al., 2016). Alternatively, a unified pooling-type local linear smoothing approach can be employed for estimating the mean functions $\mu_j(\cdot)$'s and marginal- and (or) cross-covariance functions $\Sigma_{jk}(\cdot,\cdot)$'s across j, k to handle both sparsely and densely observed functional data (Li and Solea, 2018; Qiao et al., 2020; Lee et al., 2023; Fang et al., 2023). Under certain lower-dimensional structural assumptions, one can possibly develop the nonparametric smoothing method for the joint estimation of elements in $\mathbf{\Sigma}(\cdot,\cdot)$, which, however, becomes challenging due to the observation of each $X_{ij}(\cdot)$ at different sets of points. The elementwise approach is computationally feasible as it can be easily parallelized and those FPCA-based methods only necessitate the estimation of marginal- instead of cross-covariance functions (Qiao et al., 2020; Solea and Li, 2022). Moreover, such approach can be largely accelerated in a common scenario where each $X_{ij}(\cdot)$ is observed at the same set of points across j especially with the aid of linear binning (Fan and Marron, 1994), resulting in an efficient estimation procedure. See Remark 1.

On the theory side, this approach entails dealing with the second-order U-statistics with complex dependence structures, posing a technically challenging task. Qiao et al. (2020) made the first attempt to derive some sub-optimal concentration inequalities for local linear smoothers of marginal-covariance functions $\widehat{\Sigma}_{jj}(\cdot,\cdot)$'s, albeit under a restrictive finite-dimensional setting. Lee et al. (2023) established the convergence of their proposed estimation of conditional functional graphical models under the assumption of elementwise maximum rate for the covariance smoothers:

$$\max_{1\leqslant j,k\leqslant p} \|\widehat{\Sigma}_{jk} - \Sigma_{jk}\|_{\mathcal{S}} = O_P(\log pn^{-\tau}), \tag{1}$$

where $\|\cdot\|_{\mathcal{S}}$ denotes the Hilbert–Schmidt norm, and the parameter $\tau \in (0, 1/2]$ reflects the average sampling frequency, with larger values yielding denser observational points. Fang et al. (2023) developed the functional covariance estimation with theoretical guarantees by assuming generalized sub-Gaussian-type concentration inequalities for local linear smoothers $\widehat{\Sigma}_{jk}(\cdot,\cdot)$'s, resulting in an improved elementwise maximum rate:

$$\max_{1\leqslant j,k\leqslant p} \|\widehat{\Sigma}_{jk} - \Sigma_{jk}\|_{\mathcal{S}} = O_P\{(\log p)^{1/2} n^{-\tau} + h^2\}, \tag{2}$$

where $h > 0$ is the bandwidth parameter. However, it remains of theoretical interest to ask:

- What are the exact forms of such rates as functions of n, p, T_{ij}'s, and associated bandwidth parameters under cases with different sampling frequencies?
- Are these rates well-established in the sense of specifying the largest values of τ and, compared to Zhang and Wang (2016), exhibiting any corresponding phase transition phenomena in the high-dimensional setting?

This paper aims to fill crucial theoretical gaps related to local linear smoothers frequently adopted in existing literature. Specifically, we present a systematic and unified non-asymptotic analysis of local linear smoothers for the mean and covariance functions to accommodate both sparsely and densely observed functional data in high dimensions. While our focus is not to introduce new methodologies for handling high-dimensional partially observed functional data, we make three new contributions as follows.

- First, we develop generalized sub-Gaussian-type concentration inequalities for each functional element of the mean and covariance estimators in both L_2 norm and supremum norm. Compared to the asymptotic results in Zhang and Wang (2016), our non-asymptotic error bounds lead to the same rates of L_2 convergence and uniform convergence, and reveal the same phase transition phenomena depending on the relative order of the average sampling frequency per subject to $n^{1/4}$ for dense functional data. See Remarks 4 and 5.

- Second, we derive elementwise maximum rates of both L_2 and uniform convergence for the mean and covariance estimators. Notably, we fundamentally improve the rates (1) and (2) assumed in existing literature in the sense of precisely specifying the largest values of τ under cases with different sampling frequencies. These established rates in Theorems 6 and 7 serve as a foundational tool to provide theoretical guarantees for a set of models that can handle high-dimensional partially observed functional data, such as functional graphical models (Li and Solea, 2018; Qiao et al., 2019; Zhao et al., 2022; Solea and Li, 2022; Zapata et al., 2022; Lee et al., 2023; Tsai et al., 2023), functional additive regressions (Fan et al., 2014, 2015; Kong et al., 2016; Luo and Qi, 2017; Wang et al., 2022), and functional covariance estimation (Fang et al., 2023).

- Third, with the presence of additional $\log p$ terms to account for the high-dimensional effect in our established elementwise maximum rates, the scaled phase transitions for high-dimensional dense functional data occur based on the ratios of the average sampling frequency per subject to $n^{1/4}(\log p)^{-1/4}$. This leads to a further partition of dense functional data into categories of "semi-dense" and "ultra-dense", depending on whether the parametric rate $(\log p)^{1/2}n^{-1/2}$ can be attained or not. With suitable choices of optimal bandwidths, we also present the optimal elementwise maximum rates from sparse to semi-dense to ultra-dense functional data, which correspondingly extend the optimal rates in Zhang and Wang (2016) to the high-dimensional setting. See Remarks 8 and 9.

Outline of the paper. In Section 2, we present the nonparametric smoothing approach to estimate the mean and covariance functions. In Section 3, we investigate the non-asymptotic and convergence properties of the proposed local linear smoothers and discuss the associated phase transition phenomena. The established theoretical results are validated through simulations in Section 4. All technical proofs are relegated to the Appendix.

Notation. We summarize here some notation to be used throughout the paper. For a positive integer q, we write $[q] = \{1, \dots, q\}$. For $x, y \in \mathbb{R}$, we write $x \vee y = \max(x, y)$ and $x \wedge y = \min(x, y)$. We use $I(\cdot)$ to denote an indicator function. Let $L_2(\mathcal{U})$ be a Hilbert space of square-integrable functions on a compact interval $\mathcal{U}$ equipped with the inner product $\langle f, g\rangle = \int f(u)g(u)\mathrm{d}u$ for $f(\cdot)$, $g(\cdot) \in L_2(\mathcal{U})$ and the induced L_2 norm $\|\cdot\| = \langle \cdot, \cdot\rangle^{1/2}$. For any bivariate function $\Phi(\cdot, \cdot)$ in the tensor product space $L_2(\mathcal{U}) \otimes L_2(\mathcal{U})$, we also use Φ to denote the linear operator induced from the kernel function $\Phi(\cdot, \cdot)$, that is, for any

$f(\cdot) \in L_2(\mathcal{U})$, $\Phi(f)(\cdot) = \int \Phi(\cdot, v) f(v) \mathrm{d}v \in L_2(\mathcal{U})$, and denote its Hilbert–Schmidt norm by $\{\int\int \Phi(u,v)^2 \mathrm{d}u\mathrm{d}v\}^{1/2}$. For two positive sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n \lesssim b_n$ or $b_n \gtrsim a_n$ if there exist a positive constant c such that $\limsup_{n\to\infty} a_n/b_n \leqslant c$. We write $a_n \asymp b_n$ if and only if $a_n \lesssim b_n$ and $b_n \lesssim a_n$ hold simultaneously.

2 Methodology

Let $\mathbf{X}_i(\cdot) = \{X_{i1}(\cdot), \ldots, X_{ip}(\cdot)\}^{\mathrm{T}}$ for $i \in [n]$ be independently and identically distributed copies of $\mathbf{X}(\cdot)$ defined on $\mathcal{U}$ with mean $\boldsymbol{\mu}(\cdot)$ and covariance $\mathbf{\Sigma}(\cdot, \cdot)$. For any $i \in [n]$ and $j \in [p]$, $X_{ij}(\cdot)$ is not directly observable in practice. Instead it is observed, with random errors, at T_{ij} random time points, $U_{i1}, \ldots, U_{iT_{ij}} \in \mathcal{U}$. Let Y_{ijt} be the observed value of $X_{ij}(U_{ijt})$ satisfying

$$Y_{ijt} = X_{ij}(U_{ijt}) + \varepsilon_{ijt}, \tag{3}$$

where the errors ε_{ijt}'s, independent of X_{ij}'s, are independently and identically distributed copies of ε_j with $\mathbb{E}(\varepsilon_j) = 0$ and $\mathrm{var}(\varepsilon_j) = \sigma_j^2 < \infty$.

Based on the observed data $\{(U_{ijt}, Y_{ijt}) : i \in [n], j \in [p], t \in [T_{ij}]\}$, we present a unified procedure to estimate the mean functions $\mu_j(\cdot)$'s and the marginal- and cross-covariance functions $\Sigma_{jk}(\cdot, \cdot)$'s for both sparsely and densely observed functional data. In what follows, denote $K_h(\cdot) = h^{-1}K(\cdot/h)$ for a univariate kernel K with bandwidth $h > 0$. For each j, a local linear smoother is firstly applied to $\{(U_{ijt}, Y_{ijt}) : i \in [n], t \in [T_{ij}]\}$, and hence the estimated mean function is attained via $\hat{\mu}_j(u) = \hat{b}_0$, where

$$(\hat{b}_0, \hat{b}_1) = \arg\min_{b_0, b_1} \sum_{i=1}^{n} v_{ij} \sum_{t=1}^{T_{ij}} \Big\{ Y_{ijt} - b_0 - b_1(U_{ijt} - u) \Big\}^2 K_{h_{\mu,j}}(U_{ijt} - u). \tag{4}$$

The weight v_{ij} is attached to each observation for the ith subject and the jth functional variable such that $\sum_{i=1}^{n} T_{ij} v_{ij} = 1$ (Zhang and Wang, 2016).

For each $i \in [n], j, k \in [p], t \in [T_{ij}]$ and $s \in [T_{ik}]$, once the mean functions are estimated, let $\Theta_{ijkts} = \{Y_{ijt} - \hat{\mu}_j(U_{ijt})\}\{Y_{iks} - \hat{\mu}_k(U_{iks})\}$ be the "raw covariance" between Y_{ijt} and Y_{iks}. Notice that $\mathrm{cov}(Y_{ijt}, Y_{iks}) = \Sigma_{jk}(U_{ijt}, U_{iks}) + \sigma_j^2 I(j = k) I(t = s)$. To estimate the marginal-covariance function $\Sigma_{jj}(\cdot, \cdot)$ for each j or the cross-covariance function $\Sigma_{jk}(\cdot, \cdot)$ for each $j \neq k$, we employ local linear surface smoothers to the off-diagonals of the raw marginal-covariances $(\Theta_{ijjts})_{1 \leqslant t \neq s \leqslant T_{ij}}$ or to the raw cross-covariances $(\Theta_{ijkts})_{t \in [T_{ij}], s \in [T_{ik}]}$. Specifically, we minimize

$$\sum_{i=1}^{n} w_{ijk} \sum_{(t,s) \in \mathcal{T}} \Big\{ \Theta_{ijkts} - \beta_0 - \beta_1(U_{ijt} - u) - \beta_2(U_{iks} - v) \Big\}^2 K_{h_{\Sigma,jk}}(U_{ijt} - u) K_{h_{\Sigma,jk}}(U_{iks} - v). \tag{5}$$

with respect to $(\beta_0, \beta_1, \beta_2)$, where the set $\mathcal{T}$ equals to $\{(t,s) : t \in [T_{ij}], s \in [T_{ij}], t \neq s\}$ if $j = k$ or $\{(t,s) : t \in [T_{ij}], s \in [T_{ik}]\}$ if $j \neq k$, and the weight w_{ijk} is assigned to each triplet (i, j, k) such that $\sum_{i=1}^{n} T_{ij}\{T_{ik} - I(j = k)\} w_{ijk} = 1$. See the weights to estimate marginal-covariance functions in Zhang and Wang (2016). The resulting marginal- or cross-covariance estimator is $\widehat{\Sigma}_{jk}(u, v) = \hat{\beta}_0$. For ease of presentation, we assume that the mean functions $\mu_j(\cdot)$'s are known in advance when discussing the concentration and convergence results related to the covariance estimators $\widehat{\Sigma}_{jk}(u, v)$'s in Section 3 below. However, it is noteworthy that all our discussions remain valid even when $\mu_j(\cdot)$'s are unknown as long as a few additional technical assumptions are imposed.

Our estimation procedure allows general weighting schemes for $\{v_i\}_{i\in[n]}$, $\{w_{ijk}\}_{i\in[n],j,k\in[p]}$ such that two types of frequently-used schemes in existing literature are special cases of them. One type assigns the same weights to each observation (Yao et al., 2005a) with $v_{ij} = (\sum_{i=1}^n T_{ij})^{-1}$ and $w_{ijk} = [\sum_{i=1}^n T_{ij}\{T_{ik} - I(j = k)\}]^{-1}$, so a subject with a larger number of observations receives more weights in total. The other type assigns the same weights to each subject (Li and Hsing, 2010), thus leading to $v_{ij} = (nT_{ij})^{-1}$ and $w_{ijk} = [nT_{ij}\{T_{ik} - I(j = k)\}]^{-1}$.

Remark 1 *(i) Suppose that the estimated mean and covariance functions are evaluated at a grid of $R \times R$ locations over $\mathcal{U}^2$. Under high-dimensional settings, it is apparent that our nonparametric smoothing approach suffers from high computational cost in kernel evaluations, particularly when estimating $p(p+1)/2$ marginal- and cross-covariance functions. In a common practical scenario, where each $X_{ij}(\cdot)$ is observed at the same set of time points $U_1, \dots, U_{T_i} \in \mathcal{U}$ across $j \in [p]$, the number of kernel evaluations is reduced from $O(\sum_{i=1}^n \sum_{j=1}^p T_{ij}R)$ to $O(\sum_{i=1}^n T_i R)$. To further speed up the computation, we can adopt the linear binning technique (Fan and Marron, 1994) to approximate the mean and covariance estimation. This would largely reduce the number of kernel evaluations to $O(R)$ while requiring only $O(\sum_{i=1}^n T_i)$ additional operations. See the detailed implementation of binning in Fang et al. (2023). Our conducted numerical experiments show that such binned implementation offers substantially improved computational efficiency without sacrificing any estimation accuracy.*
(ii) In a general scenario when $X_{ij}(\cdot)$'s are observed at different sets of time points, parallel estimation for $j, k \in [p]$ can be employed, resulting in an efficient procedure. In contrast, under certain lower-dimensional structural assumptions, the possible development of nonparametric smoothing method for the joint estimation of components in $\mathbf{\Sigma}(\cdot,\cdot)$ becomes challenging in this general scenario, and is thus left for future research.
(iii) Due to the infinite-dimensional nature of functional data, it is standard practice to employ FPCA as a dimension reduction technique before subsequent modelling, which is evident in commonly used high-dimensional functional models, as exemplified in Kong et al. (2016); Qiao et al. (2019); Solea and Li (2022); Zapata et al. (2022). When dealing with partially observed functional data, this necessitates the estimation of only marginal-covariance functions $\Sigma_{jj}(\cdot,\cdot)$ across $j \in [p]$, which can be easily paralleled for fast computation.

3 Theory

Before presenting the concentration and convergence results, we impose the following regularity assumptions.

Assumption 1 *For each $i \in [n]$ and $j \in [p]$, $X_{ij}(\cdot)$ is a sub-Gaussian random process and ε_{ij} is a sub-Gaussian random variable, that is, there exists some positive constant c such that $\mathbb{E}\{\exp(\langle x, X_{ij} - \mu_j\rangle)\} \leqslant \exp\{2^{-1}c^2\langle x, \Sigma_{jj}(x)\rangle\}$ for all $x(\cdot) \in L^2(\mathcal{U})$ and $\mathbb{E}\{\exp(\varepsilon_{ij}z)\} \leqslant \exp(c^2\sigma_j^2 z^2/2)$ for all $z \in \mathbb{R}$.*

Assumption 2 *For each $i \in [n]$ and $j \in [p]$, under the sparse design, $T_{ij} \leqslant T_0 < \infty$, and, under the dense design, $T_{ij} \to \infty$ and there exists some positive constant $\bar{c}$ such that $\max_{i,j} T_{ij}(\min_{i,j} T_{ij})^{-1} \leqslant \bar{c}$.*

Assumption 3 *Under the dense design, there exists some positive constant c_0 such that $\max_{i,j} v_{ij}(\min_{i,j} v_{ij})^{-1} \leqslant c_0$ and $\max_{i,j,k} w_{ijk}(\min_{i,j,k} w_{ijk})^{-1} \leqslant c_0$.*

Assumption 4 *(i) Let $\{U_{ijt} : i \in [n], t \in [T_{ij}]\}$ be independently and identically distributed copies of a random variable U defined on $\mathcal{U}$. The density $f_U(\cdot)$ of U satisfies $0 < m_f \leqslant \inf_{\mathcal{U}} f_U(u) \leqslant \sup_{\mathcal{U}} f_U(u) \leqslant M_f < \infty$ for some positive constants m_f and M_f; (ii) $\mathbf{X}(\cdot)$, U and $\{\varepsilon_j\}_{j\in[p]}$ are mutually independent.*

Assumption 5 *Let $B_{jk} = [(k-1)\widetilde{T}_j^{-1}, k\widetilde{T}_j^{-1}]$ for $k \in [\widetilde{T}_j]$ with $\widetilde{T}_j = \max_i T_{ij}$, there exists some constant $C > 0$ such that the cardinality $\#\{U_{ijt} : U_{ijt} \in B_{jk}, t \in [T_{ij}]\} \leqslant C$ for each $i \in [n]$, $j \in [p]$ and $k \in [\widetilde{T}_j]$.*

Assumption 6 *(i) $K(\cdot)$ is symmetric probability density function on $[-1, 1]$ with $\int u^2 K(u)du < \infty$ and $\int K(u)^2 du < \infty$. (ii) $K(\cdot)$ is Lipschitz continuous: there exists some positive constant L such that $|K(u) - K(v)| \leqslant L|u - v|$ for any $u, v \in [-1, 1]$.*

Assumption 7 *(i) $\partial^2\mu_j(u)/\partial u^2$ is uniformly bounded over $u \in \mathcal{U}$ and $j \in [p]$; (ii) $\partial^2\Sigma_{jk}(u,v)/\partial u^2$, $\partial^2\Sigma_{jk}(u,v)/\partial u\partial v$, and $\partial^2\Sigma_{jk}(u,v)/\partial v^2$ are uniformly bounded over $(u,v) \in \mathcal{U}^2$ and $(j,k) \in [p]^2$.*

The sub-Gaussianities in Assumption 1 for both Hilbert space-valued random elements $X_{ij}(\cdot)$'s and random errors ε_{ij}'s together imply that the observations Y_{ijt}'s in (3) are sub-Gaussian, which plays a crucial role in deriving our subsequent concentration inequalities. The dense case in Assumption 2 corresponds to a common practical scenario, where the sampling frequencies T_{ij}'s are of the same order across $i \in [n]$ and $j \in [p]$. Under such a scenario, Assumption 3 is automatically satisfied by two frequently-used weighting schemes including "equal weight per observation" and "equal weight per subject". Assumption 5 means that all observational time points are distributed in the sense of "uniformly on $\mathcal{U}$". This prevents the occurrence of an extreme case where a large number of time points are concentrated in some small areas while leaving too few points in other regions. Assumptions 4, 6 and 7 are standard in the literature of local linear smoothing for functional data (Yao et al., 2005a; Zhang and Wang, 2016) adaptable to the multivariate setting.

Theorem 2 *Suppose that Assumptions 1-6 hold. For each $j \in [p]$, let $\gamma_{n,T,h,j} = n(1 \wedge \bar{T}_{\mu,j}h_{\mu,j})$ with the corresponding average sampling frequency per subject $\bar{T}_{\mu,j} = n^{-1}\sum_{i=1}^n T_{ij}$, then there exist some positive constants c_1, c_2 (independent of $n, p, \bar{T}_{\mu,j}$'s) and arbitrarily small $\epsilon_1 > 0$ such that for any $\delta \in (0, 1]$,*

$$\mathbb{P}\big(\|\hat{\mu}_j - \tilde{\mu}_j\| \geqslant \delta\big) \leqslant c_2 \exp\big(-c_1\gamma_{n,T,h,j}\delta^2\big), \tag{6}$$

$$\mathbb{P}\Big\{\sup_{u\in\mathcal{U}} \big|\hat{\mu}_j(u) - \tilde{\mu}_j(u)\big| \geqslant \delta\Big\} \leqslant \frac{c_2(n^{\epsilon_1}\gamma_{n,T,h,j})^{1/2}}{h_{\mu,j}^2} \exp\big(-c_1\gamma_{n,T,h,j}\delta^2\big), \tag{7}$$

where $\tilde{\mu}_j(u)$ is a deterministic univariate function that converges to $\mu_j(u)$ as $h_{\mu,j} \to 0$. See (A.1) in Appendix A for the exact form of $\tilde{\mu}_j(u)$.

Theorem 3 *Suppose that Assumptions 1-6 hold. For each $j,k \in [p]$, let $\nu_{n,T,h,jk} = n(1 \wedge \bar{T}_{\Sigma,jk}^2 h_{\Sigma,jk}^2)$ with the corresponding average sampling frequency per subject being $\bar{T}_{\Sigma,jk} = [n^{-1}\sum_{i=1}^n T_{ij}\{T_{ik} - I(j=k)\}]^{1/2}$, then there exist some positive constants c_3, c_4 (independent of $n, p, \bar{T}_{\Sigma,jk}$'s) and arbitrarily small $\epsilon_2 > 0$ such that for any $\delta \in (0,1]$,*

$$\mathbb{P}\Big(\|\widehat{\Sigma}_{jk} - \widetilde{\Sigma}_{jk}\|_{\mathcal{S}} \geqslant \delta\Big) \leqslant c_4 \exp\big(-c_3\nu_{n,T,h,jk}\delta^2\big), \tag{8}$$

$$\mathbb{P}\Big\{\sup_{(u,v)\in\mathcal{U}^2}\big|\widehat{\Sigma}_{jk}(u,v) - \widetilde{\Sigma}_{jk}(u,v)\big| \geqslant \delta\Big\} \leqslant \frac{c_4 n^{\epsilon_2}\nu_{n,T,h,jk}}{h_{\Sigma,jk}^6}\exp\big(-c_3\nu_{n,T,h,jk}\delta^2\big), \tag{9}$$

where $\widetilde{\Sigma}_{jk}(u,v)$ is a deterministic bivariate function that converges to $\Sigma_{jk}(u,v)$ as $h_{\Sigma,jk} \to 0$. See (A.19) in Appendix A for the exact form of $\widetilde{\Sigma}_{jk}(u,v)$.

Remark 4 *The concentration inequalities in Theorems 2 and 3 imply that $\hat{\mu}_j$ and $\widehat{\Sigma}_{jk}$ are nicely concentrated around $\tilde{\mu}_j$ and $\widetilde{\Sigma}_{jk}$, respectively, in both L_2 norm and supremum norm with generalized sub-Gaussian-type tail behaviors. It is worth mentioning that such L_2 and uniform concentration results are derived based on the local concentration inequalities of $\hat{\mu}_j(u)$ and $\widehat{\Sigma}_{j,k}(u,v)$ for fixed interior points $u, v \in \mathcal{U}$, which enjoy the same tail behaviors as (6) and (8). Besides being fundamental to derive elementwise maximum error bounds that are essential for further convergence analysis under high-dimensional settings, these non-asymptotic results lead to the same rates of L_2 convergence and uniform convergence compared to those in Zhang and Wang (2016). Specifically, under extra Assumption 7, it holds that*

$$\|\hat{\mu}_j - \mu_j\| = O_P\big\{n^{-1/2} + (n\bar{T}_{\mu,j}h_{\mu,j})^{-1/2} + h_{\mu,j}^2\big\},$$

$$\sup_{u\in\mathcal{U}}\big|\hat{\mu}_j(u) - \mu_j(u)\big| = O_P\big[(\log n)^{1/2}n^{-1/2}\{1 + (\bar{T}_{\mu,j}h_{\mu,j})^{-1/2}\} + h_{\mu,j}^2\big],$$

$$\|\widehat{\Sigma}_{jk} - \Sigma_{jk}\|_{\mathcal{S}} = O_P\big\{n^{-1/2} + (n\bar{T}_{\Sigma,jk}^2 h_{\Sigma,jk}^2)^{-1/2} + h_{\Sigma,jk}^2\big\},$$

$$\sup_{(u,v)\in\mathcal{U}^2}\big|\widehat{\Sigma}_{jk}(u,v) - \Sigma_{jk}(u,v)\big| = O_P\big[(\log n)^{1/2}n^{-1/2}\{1 + (\bar{T}_{\Sigma,jk}^2 h_{\Sigma,jk}^2)^{-1/2}\} + h_{\Sigma,jk}^2\big].$$

Remark 5 *The above rates of convergence reveal interesting phase transition phenomena depending on the ratio of the average sampling frequency per subject $\bar{T}_{\mu,j}$ (or $\bar{T}_{\Sigma,jk}$) to $n^{1/4}$. In the following, we use different rates of L_2 convergence for $\hat{\mu}_j$ and $\widehat{\Sigma}_{jk}$ to illustrate a systematic partition of partially observed functional data into three categories:*

1. *Under the sparse design, when $h_{\mu,j} \asymp n^{-1/5}$,*

$$\|\hat{\mu}_j - \mu_j\| = O_P(n^{-1/2}h_{\mu,j}^{-1/2} + h_{\mu,j}^2) = O_P(n^{-2/5});$$

when $h_{\Sigma,jk} \asymp n^{-1/6}$,

$$\|\widehat{\Sigma}_{jk} - \Sigma_{jk}\|_{\mathcal{S}} = O_P(n^{-1/2}h_{\Sigma,jk}^{-1} + h_{\Sigma,jk}^2) = O_P(n^{-1/3}).$$

2. *Under the dense design, when* $\bar{T}_{\mu,j} n^{-1/4} \to 0$ *with* $h_{\mu,j} \asymp (n\bar{T}_{\mu,j})^{-1/5}$,

$$\|\hat{\mu}_j - \mu_j\| = O_P(n^{-1/2}\bar{T}_{\mu,j}^{-1/2} h_{\mu,j}^{-1/2} + h_{\mu,j}^2) = O_P\{(n\bar{T}_{\mu,j})^{-2/5}\};$$

when $\bar{T}_{\Sigma,jk} n^{-1/4} \to 0$ *with* $h_{\Sigma,j} \asymp (n\bar{T}_{\Sigma,jk}^2)^{-1/6}$,

$$\|\widehat{\Sigma}_{jk} - \Sigma_{jk}\|_{\mathcal{S}} = O_P(n^{-1/2}\bar{T}_{\Sigma,jk}^{-1} h_{\Sigma,jk}^{-1} + h_{\Sigma,jk}^2) = O_P\{(n\bar{T}_{\Sigma,jk}^2)^{-1/3}\}.$$

3. *Under the dense design, when* $\bar{T}_{\mu,j} n^{-1/4} \to \tilde{c}$ *(some positive constant) with* $h_{\mu,j} \asymp n^{-1/4}$ *or* $\bar{T}_{\mu,j} n^{-1/4} \to \infty$ *with* $h_{\mu,j} = o(n^{-1/4})$ *and* $\bar{T}_{\mu,j} h_{\mu,j} \to \infty$,

$$\|\hat{\mu}_j - \mu_j\| = O_P(n^{-1/2});$$

when $\bar{T}_{\Sigma,jk} n^{-1/4} \to \tilde{c}$ *with* $h_{\Sigma,jk} \asymp n^{-1/4}$ *or* $\bar{T}_{\Sigma,jk} n^{-1/4} \to \infty$ *with* $h_{\Sigma,jk} = o(n^{-1/4})$ *and* $\bar{T}_{\Sigma,jk} h_{\Sigma,jk} \to \infty$,

$$\|\widehat{\Sigma}_{jk} - \Sigma_{jk}\|_{\mathcal{S}} = O_P(n^{-1/2}).$$

As $\bar{T}_{\mu,j}$ *and* $\bar{T}_{\Sigma,jk}$ *grow very fast, case 3 results in the root-n rate complying with the parametric rate for fully observed functional data. As* $\bar{T}_{\mu,j}$ *and* $\bar{T}_{\Sigma,jk}$ *grow moderately fast, case 2 corresponds to the optimal minimax rates (Zhang and Wang, 2016), which are slower than root-n but faster than the counterparts for sparsely observed functional data. Our established convergence rates in cases 1, 2 and 3 allow free choices of* (j,k)*, and are respectively consistent to those of the mean and covariance estimators under categories of "sparse", "semi-dense" and "ultra-dense" univariate functional data introduced in Zhang and Wang (2016).*

Theorem 6 *Suppose that the assumptions in Theorem 2 and Assumption 7(i) hold, and* $(\min_j \gamma_{n,T,h,j})^{-1} \log p \to 0$, $\max_j h_{\mu,j} \to 0$ *as* $n, p \to \infty$*. It then holds that*

$$\max_{j\in[p]} \|\hat{\mu}_j - \mu_j\| = O_P\left\{\left(\frac{\log p}{\min_j \gamma_{n,T,h,j}}\right)^{1/2} + \max_j h_{\mu,j}^2\right\}, \tag{10}$$

and, if $\min_j h_{\mu,j} \asymp \{\log(p \vee n)/n\}^{\kappa_1}$ *for some* $\kappa_1 \in (0, 1/2]$,

$$\max_{j\in[p]} \sup_{u\in\mathcal{U}} \left|\hat{\mu}_j(u) - \mu_j(u)\right| = O_P\left[\left\{\frac{\log(p\vee n)}{\min_j \gamma_{n,T,h,j}}\right\}^{1/2} + \max_j h_{\mu,j}^2\right]. \tag{11}$$

Theorem 7 *Suppose that the assumptions in Theorem 3 and Assumption 7(ii) hold, and* $(\min_{j,k} \nu_{n,T,h,jk})^{-1} \log p \to 0$, $\max_{j,k} h_{\Sigma,jk} \to 0$ *as* $n, p \to \infty$*. It then holds that*

$$\max_{j,k\in[p]} \|\widehat{\Sigma}_{jk} - \Sigma_{jk}\|_{\mathcal{S}} = O_P\left\{\left(\frac{\log p}{\min_{j,k} \nu_{n,T,h,jk}}\right)^{1/2} + \max_{j,k} h_{\Sigma,jk}^2\right\}, \tag{12}$$

and, if $\min_{j,k} h_{\Sigma,jk} \asymp \{\log(p \vee n)/n\}^{\kappa_2}$ *for some* $\kappa_2 \in (0, 1/2]$,

$$\max_{j,k\in[p]} \sup_{(u,v)\in\mathcal{U}^2} \left|\widehat{\Sigma}_{jk}(u,v) - \Sigma_{jk}(u,v)\right| = O_P\left[\left\{\frac{\log(p\vee n)}{\min_{j,k} \nu_{n,T,h,jk}}\right\}^{1/2} + \max_{j,k} h_{\Sigma,jk}^2\right]. \tag{13}$$

We observe that the elementwise maximum rates of L_2 convergence and uniform convergence are governed by both dimensionality parameters $(n, p, \{\bar{T}_{\mu,j}\}_{j\in[p]}, \{\bar{T}_{\Sigma,jk}\}_{j,k\in[p]})$ and internal parameters $(\{h_{\mu,j}\}_{j\in[p]}, \{h_{\Sigma,jk}\}_{j,k\in[p]})$. Each convergence rate is composed of two terms reflecting our familiar variance-bias tradeoff in nonparametric statistics. It is easy to see that the variance terms are determined by the least frequently sampled and smoothed components, that is the smallest $\bar{T}_{\mu,j}$ (or $\bar{T}_{\Sigma,jk}$) and $h_{\mu,j}$ (or $h_{\Sigma,jk}$) across j, k, whereas the highest level of smoothness with the largest $h_{\mu,j}$ (or $h_{\Sigma,jk}$) controls the bias terms.

Remark 8 *To facilitate further discussion, we consider the simplified setting where $\bar{T}_{\mu,j} \asymp \bar{T}_\mu, h_{\mu,j} \asymp h_\mu$ and $\bar{T}_{\Sigma,jk} \asymp \bar{T}_\Sigma, h_{\Sigma,jk} \asymp h_\Sigma$ for each j, k. Compared to cases 1–3 above, the corresponding elementwise maximum rates of convergence for $\{\hat{\mu}_j\}_{j\in[p]}$ (or $\{\widehat{\Sigma}_{jk}\}_{j,k\in[p]}$) in Theorem 6 (or Theorem 7) reveal scaled phase transitions for dense functional data depending on the relative order of $\bar{T}_\mu$ (or $\bar{T}_\Sigma$) to $n^{1/4}(\log p)^{-1/4}$ instead of $n^{1/4}$. In the following, we use elementwise maximum rates of L_2 convergence to illustrate the phase transition phenomena and the optimal estimation from sparse to dense functional data in high dimensions. In terms of uniform convergence, the same phenomena occur as long as $p \gtrsim n$.*

(i) Under the sparse design, when $h_\mu \asymp (\log p)^{1/5} n^{-1/5}$,

$$\max_j \|\hat{\mu}_j - \mu_j\| = O_P\left\{\left(\frac{\log p}{n h_\mu}\right)^{1/2} + h_\mu^2\right\} = O_P\left\{\left(\frac{\log p}{n}\right)^{2/5}\right\};$$

when $h_\Sigma \asymp (\log p)^{1/6} n^{-1/6}$,

$$\max_{j,k} \|\widehat{\Sigma}_{jk} - \Sigma_{jk}\|_{\mathcal{S}} = O_P\left\{\left(\frac{\log p}{n h_\Sigma^2}\right)^{1/2} + h_\Sigma^2\right\} = O_P\left\{\left(\frac{\log p}{n}\right)^{1/3}\right\}.$$

(ii) Under the dense design, when $\bar{T}_\mu (\log p)^{1/4} n^{-1/4} \to 0$ with $h_\mu \asymp (\log p)^{1/5} (n\bar{T}_\mu)^{-1/5}$,

$$\max_j \|\hat{\mu}_j - \mu_j\| = O_P\left\{\left(\frac{\log p}{n\bar{T}_\mu h_\mu}\right)^{1/2} + h_\mu^2\right\} = O_P\left\{\left(\frac{\log p}{n\bar{T}_\mu}\right)^{2/5}\right\}; \quad (14)$$

when $\bar{T}_\Sigma (\log p)^{1/4} n^{-1/4} \to 0$ with $h_\Sigma \asymp (\log p)^{1/6} (n\bar{T}_\Sigma^2)^{-1/6}$,

$$\max_{j,k} \|\widehat{\Sigma}_{jk} - \Sigma_{jk}\|_{\mathcal{S}} = O_P\left\{\left(\frac{\log p}{n\bar{T}_\Sigma^2 h_\Sigma^2}\right)^{1/2} + h_\Sigma^2\right\} = O_P\left\{\left(\frac{\log p}{n\bar{T}_\Sigma^2}\right)^{1/3}\right\}. \quad (15)$$

(iii) Under the dense design, when $\bar{T}_\mu (\log p)^{1/4} n^{-1/4} \to \tilde{c}$ with $h_\mu \asymp (\log p)^{1/4} n^{-1/4}$ or $\bar{T}_\mu (\log p)^{1/4} n^{-1/4} \to \infty$ with $h_\mu = o\{(\log p)^{1/4} n^{-1/4}\}$ and $\bar{T}_\mu h_\mu \to \infty$,

$$\max_j \|\hat{\mu}_j - \mu_j\| = O_P\left\{\left(\frac{\log p}{n}\right)^{1/2}\right\};$$

when $\bar{T}_\Sigma (\log p)^{1/4} n^{-1/4} \to \tilde{c}$ with $h_\Sigma \asymp (\log p)^{1/4} n^{-1/4}$ or $\bar{T}_\Sigma (\log p)^{1/4} n^{-1/4} \to \infty$ with $h_\Sigma = o\{(\log p)^{1/4} n^{-1/4}\}$ and $\bar{T}_\Sigma h_\Sigma \to \infty$,

$$\max_{j,k} \|\widehat{\Sigma}_{jk} - \Sigma_{jk}\|_{\mathcal{S}} = O_P\left\{\left(\frac{\log p}{n}\right)^{1/2}\right\}.$$

Remark 9 *In a similar spirit to the partitioned three categories for univariate functional data (see cases 1, 2 and 3 above), we can also term the high-dimensional partially observed functional data in cases (i), (ii), and (iii) as "sparse", "semi-dense", and "ultra-dense", respectively. The main difference lies in the presence of additional* $\log p$ *terms to account for the high-dimensional effect.*

- *As* $\bar{T}_\mu$ *and* $\bar{T}_\Sigma$ *grow at least in the order of* $n^{1/4}(\log p)^{-1/4}$, *the attained optimal rate* $(\log p)^{1/2}n^{-1/2}$ *is identical to that for the fully observed functional data (Zapata et al., 2022), presenting that the theory for high-dimensional ultra-dense functional data falls in the parametric paradigm.*

- *As* $\bar{T}_\mu$ *and* $\bar{T}_\Sigma$ *diverge slower than* $n^{1/4}(\log p)^{-1/4}$, *if we let* $h_\mu \asymp (\log p)^{1/5}(n\bar{T}_\mu)^{-1/5}$ *and* $h_\Sigma \asymp (\log p)^{1/6}(n\bar{T}_\Sigma^2)^{-1/6}$ *to balance the corresponding variance and bias terms, the optimal rates for high-dimensional semi-dense functional data are respectively achieved in (14) and (15). These rates degenerate to the minimax rates in case 2 when p is fixed. With the choice of elementwise optimal bandwidths* $h_\mu \asymp (n\bar{T}_\mu)^{-1/5}$ *and* $h_\Sigma \asymp (n\bar{T}_\Sigma^2)^{-1/6}$, *we obtain*

  $$\max_j \|\hat{\mu}_j - \mu_j\| = O_P\{(\log p)^{1/2}(n\bar{T}_\mu)^{-2/5}\}, \ \max_{j,k} \|\widehat{\Sigma}_{jk} - \Sigma_{jk}\| = O_P\{(\log p)^{1/2}(n\bar{T}_\Sigma)^{-1/3}\},$$

  which are respectively slower than the optimal rates in (14) and (15). Such discussion applies analogously to the sparse functional setting; see cases 1 and (i).

- *Compared to the asymptotic results for cases 1, 2 and 3 under a fixed p scenario, the high-dimensionality in cases (i), (ii) and (iii) leads to the scaled phase transitions, optimal selected bandwidths, and corresponding optimal rates, each of which is up to a factor of* $\log p$ *at some polynomial order.*

4 Simulations

In this section, we examine the finite-sample performance of the local linear smoothers for the mean and covariance function estimation in high dimensions.

We generalize the simulated example for univariate functional data in Zhang and Wang (2016) to the multivariate setting by generating

$$X_{ij}(u) = \mu_j(u) + \boldsymbol{\phi}(u)^{\mathrm{T}}\boldsymbol{\theta}_{ij}, \quad i \in [n], j \in [p], u \in \mathcal{U} = [0,1],$$

where the true mean function $\mu_j(u) = 1.5\sin\{3\pi(u+0.5)\} + 2u^3$, the basis function $\boldsymbol{\phi}(u) = \{\sqrt{2}\cos(2\pi u), \sqrt{2}\sin(2\pi u), \sqrt{2}\cos(4\pi u), \sqrt{2}\sin(4\pi u)\}^{\mathrm{T}}$ and the basis coefficient vector $\boldsymbol{\theta}_i = (\boldsymbol{\theta}_{i1}^{\mathrm{T}}, \ldots, \boldsymbol{\theta}_{ip}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{4p}$ is sampled independently from a mean zero multivariate Gaussian distribution with block covariance matrix $\boldsymbol{\Lambda} \in \mathbb{R}^{4p\times 4p}$ whose (j,k)th block is given by $\boldsymbol{\Lambda}_{jk} = \rho^{|j-k|}\mathrm{diag}\{2^{-2}, \ldots, 5^{-2}\} \in \mathbb{R}^{4\times 4}$ for $j,k \in [p]$. Hence the (j,k)th entry of the true covariance functions $\boldsymbol{\Sigma}(\cdot,\cdot) = \{\Sigma_{jk}(\cdot,\cdot)\}_{p\times p}$ is $\Sigma_{jk}(u,v) = \boldsymbol{\phi}(u)^{\mathrm{T}}\boldsymbol{\Lambda}_{jk}\boldsymbol{\phi}(v)$. We then generate the observed values $Y_{ijt} = X_{ij}(U_{ijt}) + \varepsilon_{ijt}$ for $t = 1, \ldots, T_{ij} = T$, where the time points U_{ijt}'s and errors ε_{ijt}'s are sampled independently from Uniform$[0,1]$ and $\mathcal{N}(0, 0.5^2)$, respectively.

We use the Epanechnikov kernel with bandwidth values varying on a dense grid. To evaluate the performance of $\hat{\mu}_j(\cdot)$ for $j \in [p]$ and $\widehat{\Sigma}_{jk}(\cdot,\cdot)$ for $(j,k) \in [p]^2$ given specific bandwidth $h_{\mu,j}$ and $h_{\Sigma,jk}$, we define the corresponding mean integrated squared errors (MISE) as $\text{MISE}(\widehat{\mu}_j, h_{\mu,j}) = \int_{\mathcal{U}}\{\widehat{\mu}_j(u) - \mu_j(u)\}^2 \text{d}u$ and $\text{MISE}(\widehat{\Sigma}_{jk}, h_{\Sigma,jk}) = \int_{\mathcal{U}}\int_{\mathcal{U}}\{\widehat{\Sigma}_{jk}(u,v) - \Sigma_{jk}(u,v)\}^2 \text{d}u\text{d}v$. We first calculate the elementwise minimal MISEs for the mean and covariance estimators over the grids of candidate bandwidths in prespecified sets $\mathcal{H}_\mu$ and $\mathcal{H}_\Sigma$, respectively. We then compute their averages and maximums over $j \in [p]$ for the mean functions, that is,

$$\text{AveMISE}(\mu) = \frac{1}{p}\sum_j \min_{h_{\mu,j}} \text{MISE}(\widehat{\mu}_j, h_{\mu,j}),$$

$$\text{MaxMISE}(\mu) = \max_j \min_{h_{\mu,j}} \text{MISE}(\widehat{\mu}_j, h_{\mu,j}),$$

and over $(j,k) \in [p]^2$ for the covariance functions, that is,

$$\text{AveMISE}(\Sigma) = \frac{1}{p^2}\sum_j\sum_k \min_{h_{\Sigma,jk}} \text{MISE}(\widehat{\Sigma}_{jk}, h_{\Sigma,jk}),$$

$$\text{MaxMISE}(\Sigma) = \max_{j,k} \min_{h_{\Sigma,jk}} \text{MISE}(\widehat{\Sigma}_{jk}, h_{\Sigma,jk}).$$

We next use the example of estimating the mean functions to illustrate the rationale of the above measures. While $\text{AveMISE}(\mu)$ presents the averaged elementwise minimal MISEs across $j \in [p]$, some simple calculations in Appendix B show that the attainable quantity of the minimal elementwise maximum of MISEs, $\min_{(h_{\mu,1},\ldots,h_{\mu,p}) \in \mathcal{H}_\mu^p} \max_{j\in[p]} \text{MISE}(\widehat{\mu}_j, h_{\mu,j})$, is equal to $\text{MaxMISE}(\mu)$.

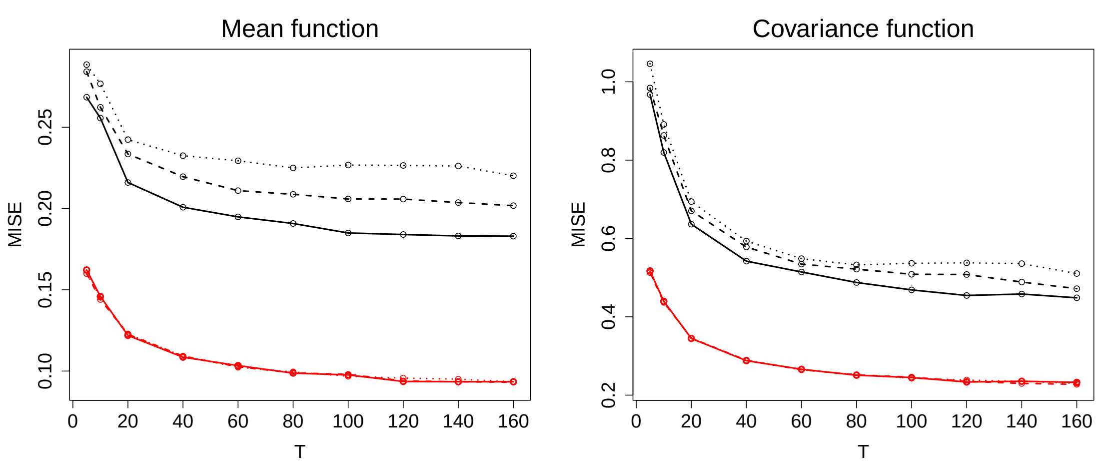

Figure 1: Plots of the average MaxMISEs (black) and AveMISEs (red) against T with $p = 50$ (solid), 100 (dashed) and 150 (dotted) for the mean estimators (left) and the covariance estimators (right).

We consider settings of $n = 100$, $p = 50, 100, 150$, and $T = 5$, 10, 20, 40, 60, 80, 100, 120, 140, 160, varying from sparse to semi-dense to ultra-dense measurement schedules. We ran each simulation 100 times. Figure 1 plots the average AveMISEs and MaxMISEs as functions of the sampling frequency T for the estimated mean and covariance functions. A few apparent patterns are observable from Figure 1. First, both MaxMISEs and AveMISEs display a similar trend as T increases from 5 to 160, with a steep decline followed by a slight decrease and then a period of stability. Such a trend roughly corresponds to the three categories of "sparse", "semi-dense", and "ultra-dense", respectively. Second, while AveMISEs reflect the performance for univariate functional data, MaxMISEs gradually enlarge as p increases from 50 to 150, providing empirical evidence to support that the associated rates in high-dimensional settings all depend on $\log p$-based multiplicative factors. Additionally, it is observable that the high-dimensionality causes the transition phase between semi-dense and ultra-dense functional data to slightly shift to the left. Third, compared to the results for the estimated mean functions, an increase in T for sparse and semi-dense functional data leads to an enhanced reduction in relative MISEs for the covariance estimators. These observations further validate our established theoretical results in Section 3.

Appendix A. Technical proofs

The appendix contains proofs of all theorems. Throughout, we use $c, c_1, c_2, \ldots$ to denote generic positive finite constants that may be different in different uses.

A.1 Proof of Theorem 2

We organize the proof in four steps. First, we will define $\widehat{\mu}(\cdot)$, $\widetilde{\mu}(\cdot)$ and obtain the decomposition of $\widehat{\mu}(\cdot) - \widetilde{\mu}(\cdot)$. Second, we will prove the local concentration inequality for fixed interior point $u \in \mathcal{U}$. Third, we will prove the concentration inequality in L_2 norm. Finally, we will prove the concentration inequality in supremum norm.

A.1.1 Definition and decomposition

Without loss of generality, we let $h_{\mu,j} = h$ for $j \in [p]$ and denote $\mathbf{e}_0 = (1,0)^{\mathrm{T}}$, $\widetilde{\mathbf{U}}_{ijt} = \big\{1, (U_{ijt} - u)/h\big\}^{\mathrm{T}}$,

$$\widehat{\mathbf{S}}_j(u) = \sum_{i=1}^{n} v_{ij} \sum_{t=1}^{T_{ij}} \widetilde{\mathbf{U}}_{ijt} \widetilde{\mathbf{U}}_{ijt}^{\mathrm{T}} K_h(U_{ijt} - u),$$
$$\widehat{\mathbf{R}}_j(u) = \sum_{i=1}^{n} v_{ij} \sum_{t=1}^{T_{ij}} \widetilde{\mathbf{U}}_{ijt} K_h(U_{ijt} - u) Y_{ijk}.$$

A simple calculation yields $\widehat{\mu}_j(u) = \mathbf{e}_0^{\mathrm{T}} \{\widehat{\mathbf{S}}_j(u)\}^{-1} \widehat{\mathbf{R}}_j(u)$. Let

$$\widetilde{\mu}_j(u) = \mathbf{e}_0^{\mathrm{T}} \big[\mathbb{E}\{\widehat{\mathbf{S}}_j(u)\}\big]^{-1} \mathbb{E}\{\widehat{\mathbf{R}}_j(u)\}. \tag{A.1}$$

We can decompose $\widehat{\mu}_j(u) - \widetilde{\mu}_j(u)$ as

$$\begin{aligned}\widehat{\mu}_j(u) - \widetilde{\mu}_j(u) =& \mathbf{e}_0^{\mathrm{T}} [\mathbb{E}\{\widehat{\mathbf{S}}_j(u)\}]^{-1} [\widehat{\mathbf{R}}_j(u) - \mathbb{E}\{\widehat{\mathbf{R}}_j(u)\}] \\ & - \mathbf{e}_0^{\mathrm{T}} \{\widehat{\mathbf{S}}_j(u)\}^{-1} [\widehat{\mathbf{S}}_j(u) - \mathbb{E}\{\widehat{\mathbf{S}}_j(u)\}] [\mathbb{E}\{\widehat{\mathbf{S}}_j(u)\}]^{-1} \widehat{\mathbf{R}}_j(u),\end{aligned}$$

which then implies that

$$\begin{aligned}|\widehat{\mu}_j(u) - \widetilde{\mu}_j(u)| \leqslant\ & \|\mathbb{E}\{\widehat{\mathbf{S}}_j(u)\}\|_{\min}^{-1}\|\widehat{\mathbf{R}}_j(u) - \mathbb{E}\{\widehat{\mathbf{R}}_j(u)\}\| \\ & + \|\widehat{\mathbf{S}}_j(u)\|_{\min}^{-1}\|\mathbb{E}\{\widehat{\mathbf{S}}_j(u)\}\|_{\min}^{-1}\|\widehat{\mathbf{R}}_j(u)\|\|\widehat{\mathbf{S}}_j(u) - \mathbb{E}\{\widehat{\mathbf{S}}_j(u)\}\|_{\mathrm{F}},\end{aligned} \tag{A.2}$$

where, for any vector $\mathbf{b} = (b_1, \ldots, b_p)^{\mathrm{T}}$, we write $\|\mathbf{b}\| = (\sum_i b_i^2)^{1/2}$ and, for any matrix $\mathbf{B} = (B_{ij})_{p\times q}$, we write $\|\mathbf{B}\|_{\min} = \{\lambda_{\min}(\mathbf{B}^{\mathrm{T}}\mathbf{B})\}^{1/2}$ and $\|\mathbf{B}\|_{\mathrm{F}} = (\sum_{i,j} B_{ij}^2)^{1/2}$ to denote its Frobenius norm.

A.1.2 Local concentration inequality

We will firstly show that there exists some positive constant c (independent of u) such that for any $\delta > 0$ and $u \in \mathcal{U}$,

$$\mathbb{P}\Big\{\big\|\widehat{\mathbf{S}}_j(u) - \mathbb{E}\{\widehat{\mathbf{S}}_j(u)\}\big\|_{\mathrm{F}} \geqslant \delta\Big\} \leqslant 8\exp\Big(-\frac{cn\bar{T}_{\mu,j}h\delta^2}{1+\delta}\Big). \tag{A.3}$$

For $k, l = 1, 2$, let $\widehat{S}_{jkl}(u)$ be the (k,l)th entry of $\widehat{\mathbf{S}}_j(u)$. Under Assumptions 4 and 6, we obtain that for any integer $q = 2, 3, \ldots$ and $s = 0, 1, 2$,

$$\mathbb{E}\left\{\Big|\Big(\frac{U_{ijt}-u}{h}\Big)^s K_h(U_{ijt}-u)\Big|^q\right\} \leqslant \int h^{-q}K^q\Big(\frac{t-u}{h}\Big)\Big|\frac{t-u}{h}\Big|^{sq} f_U(t)dt \leqslant ch^{1-q}. \tag{A.4}$$

Note that Assumption 3 implies that the weights v_{ij}'s are of the same order $v_{ij} \asymp (n\bar{T}_{\mu,j})^{-1}$. By (A.4), it holds that

$$\sum_{i=1}^n\sum_{t=1}^{T_{ij}}\mathbb{E}\left\{\Big|\Big(\frac{U_{ijt}-u}{h}\Big)^s K_h(U_{ijt}-u)\Big|^2\right\} \leqslant cn\bar{T}_{\mu,j}h^{-1},$$

$$\sum_{i=1}^n\sum_{t=1}^{T_{ij}}\mathbb{E}\left\{\Big|\Big(\frac{U_{ijt}-u}{h}\Big)^s K_h(U_{ijt}-u)\Big|^q\right\} \leqslant 2^{-1}q!cn\bar{T}_{\mu,j}h^{-1}h^{2-q} \quad \text{for } q \geqslant 3.$$

By the Bernstein inequality (see Theorem 2.10 and Corollary 2.11 of Boucheron et al. (2014)), we obtain that there exists some positive constant c (independent of u) such that for any $\delta > 0$ and $u \in \mathcal{U}$,

$$\mathbb{P}\Big\{\big|\widehat{S}_{jkl}(u) - \mathbb{E}\{\widehat{S}_{jkl}(u)\}\big| \geqslant \delta\Big\} \leqslant 2\exp\Big(-\frac{cn\bar{T}_{\mu,j}h\delta^2}{1+\delta}\Big),$$

for $k, l = 1, 2$, which, by the union bound of probability, implies that (A.3) holds.

For $k = 1, 2$, let $\widehat{R}_{jk}(u)$ be the kth element of $\widehat{\mathbf{R}}_j(u)$. We will next show that there exists some positive constant c (independent of u) such that for any $\delta > 0$ and $u \in \mathcal{U}$,

$$\mathbb{P}\Big\{\big|\widehat{R}_{jk}(u) - \mathbb{E}\{\widehat{R}_{jk}(u)\}\big| \geqslant \delta\Big\} \leqslant c\exp\Big(-\frac{c\gamma_{n,T,h,j}\delta^2}{1+\delta}\Big), \tag{A.5}$$

where $\gamma_{n,T,h,j} = n(1 \wedge \bar{T}_{\mu,j}h)$. We only need to consider the case $k = 1$, while the case $k = 2$ can be demonstrated in a similar manner. Denote that

$$\begin{aligned}\xi_{ijt} &= Y_{ijt} - \mu_j(U_{ijt}),\\ \widehat{R}_{j3}(u) &= \sum_{i=1}^{n} v_{ij} \sum_{t=1}^{T_{ij}} K_h(U_{ijt} - u)\mu_j(U_{ijt}),\\ \widehat{R}_{j4}(u) &= \sum_{i=1}^{n} v_{ij} \sum_{t=1}^{T_{ij}} K_h(U_{ijt} - u)\xi_{ijt}.\end{aligned}$$

Then $\widehat{R}_{j1}(u) - \mathbb{E}\{\widehat{R}_{j1}(u)\}$ can be rewritten as

$$\widehat{R}_{j1}(u) - \mathbb{E}\{\widehat{R}_{j1}(u)\} = \widehat{R}_{j3}(u) - \mathbb{E}\{\widehat{R}_{j3}(u)\} + \widehat{R}_{j4}(u). \tag{A.6}$$

Following the same procedure to prove (A.3) and using the Bernstein inequality, we can obtain that there exists some positive constant c such that for any $\delta > 0$ and $u \in \mathcal{U}$,

$$\mathbb{P}\Big\{\big|\widehat{R}_{j3}(u) - \mathbb{E}\{\widehat{R}_{j3}(u)\}\big| \geqslant \delta\Big\} \leqslant 2\exp\Big(-\frac{cn\bar{T}_{\mu,j}h\delta^2}{1+\delta}\Big). \tag{A.7}$$

Now we consider the tail behavior of $\widehat{R}_{j4}(u)$. Define the event $V_j = \big\{U_{ijt}, t \in [T_{ij}], i \in [n]\big\}$. It follows from sub-Gaussianities in Assumption 1 and (3) that $\mathbb{E}\{\exp(\lambda\xi_{ijt})\big|V_j\} \leqslant \exp(\lambda^2c^2)$ for any $\lambda \in \mathbb{R}$. Rewrite $\widehat{R}_{j4}(u) = \sum_{i=1}^{n} v_{ij}\psi_{ij1}(u)$ with $\psi_{ij1}(u) = \sum_{t=1}^{T_{ij}} K_h(U_{ijt} - u)\xi_{ijt}$. Note that for each i Assumption 3 implies that $v_{ij} \asymp (n\bar{T}_{\mu,j})^{-1}$. If $\sum_{t=1}^{T_{ij}} K_h(U_{ijt}-u) > 0$, by Jensen's inequality, we have

$$\begin{aligned}&\mathbb{E}\Big[\exp\{\lambda v_{ij}\psi_{ij1}(u)\}\Big|V_j\Big]\\ \leqslant&\frac{1}{\sum_{t=1}^{T_{ij}} K_h(U_{ijt}-u)} \sum_{t=1}^{T_{ij}} K_h(U_{ijt}-u)\mathbb{E}\left[\exp\Big\{\lambda v_{ij}\xi_{ijt}\sum_{t=1}^{T_{ij}} K_h(U_{ijt}-u)\Big\}\Big|V_j\right]\\ \leqslant&\exp\left[\lambda^2c^2(n\bar{T}_{\mu,j})^{-2}\Big\{\sum_{t=1}^{T_{ij}} K_h(U_{ijt}-u)\Big\}^2\right].\end{aligned}$$

Clearly, the above inequality still holds even if $\sum_{t=1}^{T_{ij}} K_h(U_{ijt}-u) = 0$. Assumption 5 implies that the number of nonzero terms in $\sum_{t=1}^{T_{ij}} K_h(U_{ijt}-u)$ has an upper bound $c(1 \vee \bar{T}_{\mu,j}h)$, which yields that $\sum_{t=1}^{T_{ij}} K_h(U_{ijt}-u) \leqslant ch^{-1}(1 \vee \bar{T}_{\mu,j}h)$. Therefore, for any $\lambda \in \mathbb{R}$, we obtain that

$$\begin{aligned}\mathbb{E}\left[\exp\Big\{\lambda\sum_{i=1}^{n} v_{ij}\psi_{ij1}(u)\Big\}\Big|V_j\right] &\leqslant \exp\left[\lambda^2c^2\sum_{i=1}^{n} v_{ij}^2\Big\{\sum_{t=1}^{T_{ij}} K_h(U_{ijt}-u)\Big\}^2\right]\\ &\leqslant \exp\Big\{\lambda^2c^2(n\bar{T}_{\mu,j}h)^{-2}h(1 \vee \bar{T}_{\mu,j}h)\sum_{i=1}^{n}\sum_{t=1}^{T_{ij}} K_h(U_{ijt}-u)\Big\}.\end{aligned}$$

For any $\delta > 0$, define the event $\Omega_{j,1}(\delta) = \big\{ \sum_{i=1}^n \sum_{t=1}^{T_{ij}} K_h(U_{ijt} - u) \leqslant c(1+\delta) n \bar{T}_{\mu,j} \big\}$. We have

$$\mathbb{E}\left[\exp\left\{ \lambda \sum_{i=1}^n v_{ij} \psi_{ij1}(u) \right\} \middle| \Omega_{j,1}(\delta) \right] \leqslant \exp\left\{ \lambda^2 c^2 (1+\delta)(n\bar{T}_{\mu,j}h)^{-2} n\bar{T}_{\mu,j} h (1 \vee \bar{T}_{\mu,j}h) \right\}$$
$$= \exp\left\{ (1+\delta)\lambda^2 c^2 (n\bar{T}_{\mu,j}h)^{-1} (1 \vee \bar{T}_{\mu,j}h) \right\}.$$

As a consequence, we obtain that

$$\mathbb{P}\Big\{ \sum_{i=1}^n v_{ij}\psi_{ij1}(u) \geqslant \delta \Big| \Omega_{j,1}(\delta) \Big\} \leqslant \exp\Big\{ -\lambda\delta + (1+\delta)\lambda^2 c^2 (n\bar{T}_{\mu,j}h)^{-1}(1 \vee \bar{T}_{\mu,j}h) \Big\}. \quad \text{(A.8)}$$

With the choice of $\lambda = n\bar{T}_{\mu,j}h\delta / \{2(1+\delta)c^2(1 \vee \bar{T}_{\mu,j}h)\}$, (A.8) degenerates to

$$\mathbb{P}\Big\{ \sum_{i=1}^n v_{ij}\psi_{ij1}(u) \geqslant \delta \Big| \Omega_{j,1}(\delta) \Big\} \leqslant \exp\Big\{ -\frac{cn(1 \wedge \bar{T}_{\mu,j}h)\delta^2}{1+\delta} \Big\}. \quad \text{(A.9)}$$

Note that $\sum_{i=1}^n \sum_{t=1}^{T_{ij}} \mathbb{E}\{K_h(U_{ijt} - u)\} \leqslant cn\bar{T}_{\mu,j}$. By the Bernstein inequality, we obtain that there exists some positive constant c such that for any $\delta > 0$

$$\mathbb{P}\left(\sum_{i=1}^n \sum_{t=1}^{T_{ij}} \Big[K_h(U_{ijt} - u) - \mathbb{E}\big\{K_h(U_{ijt} - u)\big\} \Big] \geqslant n\bar{T}_{\mu,j}\delta \right) \leqslant \exp\left(-\frac{cn\bar{T}_{\mu,j}h\delta^2}{1+\delta} \right),$$

which implies that

$$1 - \mathbb{P}\{\Omega_{j,1}(\delta)\} \leqslant \exp\Big(-\frac{cn\bar{T}_{\mu,j}h\delta^2}{1+\delta} \Big). \quad \text{(A.10)}$$

Combining (A.9) and (A.10), we obtain that there exists some constant $c > 0$ such that for any $\delta > 0$

$$\mathbb{P}\Big\{ \widehat{R}_{j4}(u) \geqslant \delta \Big\} \leqslant 2\exp\Big\{ -\frac{cn(1 \wedge \bar{T}_{\mu,j}h)\delta^2}{1+\delta} \Big\}.$$

and, consequently,

$$\mathbb{P}\Big\{ \big|\widehat{R}_{j4}(u)\big| \geqslant \delta \Big\} \leqslant 4\exp\Big\{ -\frac{cn(1 \wedge \bar{T}_{\mu,j}h)\delta^2}{1+\delta} \Big\}.$$

This together with (A.6) and (A.7) yields that, there exists some constant c such that

$$\mathbb{P}\Big\{ \big|\widehat{R}_{j1}(u) - \mathbb{E}\{\widehat{R}_{j1}(u)\}\big| \geqslant \delta \Big\} \leqslant 6\exp\Big\{ -\frac{cn(1 \wedge \bar{T}_{\mu,j}h)\delta^2}{1+\delta} \Big\}.$$

Define the event $\Omega_{j,2}(\delta) = \{\|\widehat{\mathbf{S}}_j(u) - \mathbb{E}\{\widehat{\mathbf{S}}_j(u)\}\|_{\mathrm{F}} \leqslant \delta/2\}$. Note that $\mathbb{E}\{\widehat{\mathbf{S}}_j(u)\}$ is positive definite. On the event $\Omega_{j,2}(\delta)$ with $\delta \in (0, 1]$, we obtain that

$$\|\widehat{\mathbf{S}}_j(u)\|_{\min} \geqslant c(1 - \delta/2). \quad \text{(A.11)}$$

By (A.3), we have

$$1-\mathbb{P}\big\{\Omega_{j,2}(\delta)\big\}\leqslant 8\exp\Big(-\frac{cn\bar{T}_{\mu,j}h\delta^2}{1+\delta}\Big). \tag{A.12}$$

Define the event $\Omega_{j,3}(\delta)=\big\{\|\widehat{\mathbf{R}}_j(u)-\mathbb{E}\{\widehat{\mathbf{R}}_j(u)\}\|\leqslant\delta\big\}$. Note that, under Assumption 3 with $v_{ij}\asymp(n\bar{T}_{\mu,j})^{-1}$, $\sum_{i=1}^n v_{ij}\sum_{t=1}^{T_{ij}}E\{K_h(U_{ijt}-u)\}\leqslant c$ and $\mu_j(\cdot)$ is uniformly bounded over $\mathcal{U}$, hence $\|\mathbb{E}\{\widehat{\mathbf{R}}_j(u)\}\|$ is uniformly bounded over $\mathcal{U}$. On the event $\Omega_{j,3}(\delta)$, we have

$$\big\|\widehat{\mathbf{R}}_j(u)\big\|\leqslant c(1+\delta). \tag{A.13}$$

On the event $\Omega_{j,2}(\delta)\cap\Omega_{j,3}(\delta)$ with $\delta\in(0,1]$, it follows from (A.2), (A.11) and (A.13) that

$$|\widehat{\mu}_j(u)-\widetilde{\mu}_j(u)|\leqslant c\delta+c(1-\delta/2)^{-1}(1+\delta)\delta\leqslant c_3\delta.$$

This together with concentration inequalities in (A.5) and (A.12) implies that there exist some positive universal constants c_1 and c_2 such that for any $\delta\in(0,1]$ and $u\in\mathcal{U}$,

$$\mathbb{P}\Big\{|\widehat{\mu}_j(u)-\widetilde{\mu}_j(u)|\geqslant\delta\Big\}\leqslant c_2\exp\left(-c_1\gamma_{n,T,h,j}\delta^2\right),$$

which completes the proof of local concentration inequality for the mean estimator.

A.1.3 Concentration inequality in L_2 norm

In the proof, we need the following lemma.

Lemma 10 *Let X be a random variable. If for some constants $c_1,c_2>0$, $\mathbb{P}(|X|>\delta)\leqslant c_1\exp\{-c_2^{-1}\min(\delta^2,\delta)\}$ for any $\delta>0$, then for any integer $q\geqslant 1$,*

$$\mathbb{E}(X^{2q})\leqslant q!c_1(4c_2)^q+(2q)!c_1(4c_2)^{2q}.$$

Conversely, if for some positive constants a_1,a_2, $\mathbb{E}(X^{2q})\leqslant q!a_1a_2^q+(2q)!a_1a_2^{2q}$ for any integer $q\geqslant 1$, then by letting $c_1^*=a_1$ and $c_2^*=32(a_2+a_2^2)$, we have that

$$\mathbb{P}(|X|>\delta)\leqslant c_1^*\exp\{-c_2^{*-1}\min(\delta^2,\delta)\}$$

for any $\delta>0$.

Proof This lemma can be proved in a similar way to Theorem 2.3 of Boucheron et al. (2014) and hence the proof is omitted here. In the proof, the following two inequalities are used, i.e. for any $c,\delta>0$,

$$\frac{1}{2}\min(\delta^2,\delta)\leqslant\frac{\delta^2}{1+\delta}\leqslant\min(\delta^2,\delta)$$

and

$$\sqrt{\frac{c\delta}{2}}+\frac{c\delta}{2}\leqslant\frac{c(\delta+\sqrt{\delta^2+4\delta/c})}{2}\leqslant\sqrt{c\delta}+c\delta.$$

■

We are now ready to derive the L_2 concentration inequality of $\|\widehat{\mu}_j - \widetilde{\mu}_j\|$. Let

$$\widetilde{\mathbf{S}}_j(u) = (n\bar{T}_{\mu,j})^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T_{ij}} \widetilde{\mathbf{U}}_{ijt} \widetilde{\mathbf{U}}_{ijt}^{\mathrm{T}} K_h(U_{ijt} - u). \tag{A.14}$$

Then we have that $\|\widehat{\mathbf{S}}_j(u)\|_{\min} \geqslant c\|\widetilde{\mathbf{S}}_j(u)\|_{\min}$. We now give a lower bound on $\|\widetilde{\mathbf{S}}_j(u)\|_{\min}$. Denote $W = \sup_{u\in\mathcal{U}} \big\|\widetilde{\mathbf{S}}_j(u) - \mathbb{E}\{\widetilde{\mathbf{S}}_j(u)\}\big\|_{\mathrm{F}}$. Let $\widetilde{S}_{jkl}(u)$ be the (k,l)th entry of $\widetilde{\mathbf{S}}_j(u)$ for $k,l = 1,2$. Note that $\mathbb{E}\{|(U_{ijt}-u)^a h^{-a} K_h(U_{ijt}-u)|\} \leqslant c$ and $\mathbb{E}(W) \leqslant 4\max_{k,l} \mathbb{E}\{\sup_{u\in\mathcal{U}} |\widetilde{S}_{jkl}(u)|\}$ for $a = 0,1,2$. In an analogy to Lemma 13.5 of Boucheron et al. (2014), we can show that $\mathbb{E}(W) \leqslant c(n\bar{T}_{\mu,j})^{-1/2}$. Note that Lemma 13.5 of Boucheron et al. (2014) relies on the results presented in Lemma 13.1 of Boucheron et al. (2014). Consequently, Lemma 13.5 assumes that the corresponding index set is countable in order to apply Lemma 13.1, as the supremums of the summation of indicator functions may not be measurable. However, in our specific case, each component of $\widetilde{\mathbf{S}}_j(u) - \mathbb{E}\{\widetilde{\mathbf{S}}_j(u)\}$ is the sum of continuous functions, for which the supremums over $\mathcal{U}$ are measurable. Therefore, when we extend the index set from countable to the uncountable set $\mathcal{U}$, this lemma, as well as Theorems 11.10 and 12.5 of Boucheron et al. (2014), still hold true and can be applicable to our situation. Moreover, it follows from the facts $\mathrm{var}(W) \leqslant \mathbb{E}(W^2) \leqslant 4\max_{k,l} \mathrm{var}\{\sup_{u\in\mathcal{U}} \widetilde{S}_{jkl}(u)\}$, $|(U_{ijt}-u)^a h^{-a} K_h(U_{ijt}-u)| \leqslant ch^{-1}$, $\mathbb{E}\{(U_{ijt}-u)^{2a} h^{-2a} K_h^2(U_{ijt}-u)\} \leqslant ch^{-1}$ for $a = 0,1,2$, and Theorem 11.10 of Boucheron et al. (2014) that $\mathrm{var}(n\bar{T}_{\mu,j}hW) \leqslant 2\mathbb{E}(n\bar{T}_{\mu,j}hW) + \sum_{i=1}^{n}\sum_{t=1}^{T_{ij}} ch^{-1}h^2 \leqslant c(n\bar{T}_{\mu,j})^{1/2}h + cn\bar{T}_{\mu,j}h$, which implies that the variance of W is bounded by $c(n\bar{T}_{\mu,j}h)^{-1} \leqslant c\gamma_{n,T,h,j}^{-1}$. Applying Theorem 12.5 of Boucheron et al. (2014) yields that there exists some positive constant c such that, for any $\delta > 0$,

$$\mathbb{P}\big\{W - \mathbb{E}(W) > \delta\big\} \leqslant \exp\Big(-\frac{c\gamma_{n,T,h,j}\delta^2}{1+\delta}\Big). \tag{A.15}$$

Define the event $\Omega_{j,4}(\delta) = \big\{\sup_{u\in\mathcal{U}} \big\|\widetilde{\mathbf{S}}_j(u) - \mathbb{E}\{\widetilde{\mathbf{S}}_j(u)\}\big\|_{\mathrm{F}} \leqslant \delta/2\big\}$ with $\delta \in (0,1]$. By (A.15), we obtain that there exists some constant $c > 0$ such that, for any $\delta \in (0,1]$,

$$1 - \mathbb{P}\left\{\Omega_{j,4}(\delta)\right\} \leqslant 2\exp\left(-c\gamma_{n,T,h,j}\delta^2\right). \tag{A.16}$$

On the event $\Omega_{j,4} = \Omega_{j,4}(\delta_1)$ with $c\gamma_{n,T,h,j}^{-1/2} < \delta_1 \leqslant 1$, $\|\widehat{\mathbf{S}}_j(u)\|_{\min} \geqslant c\|\widetilde{\mathbf{S}}_j(u)\|_{\min} \geqslant c(1 - \delta_1/2) \geqslant c/2$. Note that $\mathbb{E}\{\widehat{\mathbf{S}}_j(u)\}$ is positive definite and $\|\mathbb{E}\{\widehat{\mathbf{R}}_j(u)\}\|$ is uniformly bounded over $\mathcal{U}$. On the event $\Omega_{j,4}$, it thus follows from (A.2) and $\|\widehat{\mathbf{R}}_j(u)\| \leqslant \|\widehat{\mathbf{R}}_j(u) - \mathbb{E}\{\widehat{\mathbf{R}}_j(u)\}\| + \|\mathbb{E}\{\widehat{\mathbf{R}}_j(u)\}\|$ that

$$|\widehat{\mu}_j(u) - \widetilde{\mu}_j(u)| \leqslant c\|\widehat{\mathbf{R}}_j(u) - \mathbb{E}\{\widehat{\mathbf{R}}_j(u)\}\| + c\|\widehat{\mathbf{S}}_j(u) - \mathbb{E}\{\widehat{\mathbf{S}}_j(u)\}\|_{\mathrm{F}}, \tag{A.17}$$

where the positive constant c does not depend on $u \in \mathcal{U}$.

Combining (A.17) with (A.3), (A.5) and applying the first part of Lemma 10 yields that, for any $u \in \mathcal{U}$ and integer $q \geqslant 1$,

$$\mathbb{E}\Big\{|\widehat{\mu}_j(u) - \widetilde{\mu}_j(u)|^{2q}\Big|\Omega_{j,4}\Big\} \leqslant q!c\Big(\frac{4}{c\gamma_{n,T,h,j}}\Big)^q + (2q)!c\Big(\frac{4}{c\gamma_{n,T,h,j}}\Big)^{2q}.$$

Applying the second part of Lemma 10 and (A.16), we can show that, for each $\delta \in (0,1]$,

$$\mathbb{P}\big(\|\hat{\mu}_j - \tilde{\mu}_j\| \geqslant \delta\big) \leqslant \mathbb{P}\big(\|\hat{\mu}_j - \tilde{\mu}_j\| \geqslant \delta \big| \Omega_{j,4}\big) + \mathbb{P}\big(\Omega_{j,4}^c\big) \leqslant c_2 \exp\big(-c_1 \gamma_{n,T,h,j}\delta^2\big),$$

which means (6) in Theorem 2 holds and completes the proof of concentration inequality for the mean estimator in L_2 norm.

A.1.4 Concentration inequality in supremum norm

We will derive the uniform concentration bound of $\sup_{u\in\mathcal{U}}|\widehat{\mu}_j(u) - \widetilde{\mu}_j(u)|$. We partition the interval $\mathcal{U} = [0,1]$ into N subintervals I_s for $s \in [N]$ of equal length. Let u_s be the center of I_s, then we have

$$\sup_{u\in\mathcal{U}}\big|\widehat{\mu}_j(u) - \widetilde{\mu}_j(u)\big| \leqslant \max_{s\in[N]}\Big[\big|\widehat{\mu}_j(u_s) - \widetilde{\mu}_j(u_s)\big| + \big|\{\widehat{\mu}_j(u_s) - \widehat{\mu}_j(u)\} - \{\widetilde{\mu}_j(u_s) - \widetilde{\mu}_j(u)\}\big|\Big].$$

We need to bound the second term. By some calculations, it suffices to bound $\Big|\{\widehat{R}_{jk}(u) - \widehat{R}_{jk}(u_s)\} - \big[\mathbb{E}\{\widehat{R}_{jk}(u)\} - \mathbb{E}\{\widehat{R}_{jk}(u_s)\}\big]\Big|$ and $\Big|\{\widehat{S}_{jkl}(u) - \widehat{S}_{jkl}(u_s)\} - \big[\mathbb{E}\{\widehat{S}_{jkl}(u)\} - \mathbb{E}\{\widehat{S}_{jkl}(u_s)\}\big]\Big|$ for $k,l = 1,2$, which means that we need to bound $\big|\widehat{R}_{jk}(u) - \widehat{R}_{jk}(u_s)\big|$ and $\big|\widehat{S}_{jkl}(u) - \widehat{S}_{jkl}(u_s)\big|$. Let $u \in I_s$ and consider $|\widehat{R}_{j1}(u) - \widehat{R}_{j1}(u_s)|$ first. Define the event $\Omega_{R,j1} = \big\{\sum_{i=1}^n v_{ij}\sum_{t=1}^{T_{ij}}|Y_{ijt}| \leqslant \mathbb{E}(\sum_{i=1}^n v_{ij}\sum_{t=1}^{T_{ij}}|Y_{ijt}|) + 1\big\}$. On this event, it follows from Assumption 6(ii) that

$$\begin{aligned}\big|\widehat{R}_{j1}(u) - \widehat{R}_{j1}(u_s)\big| &\leqslant \Big|\sum_{i=1}^n v_{ij}\sum_{t=1}^{T_{ij}} Y_{ijt}\{K_h(U_{ijt}-u) - K_h(U_{ijt}-u_s)\}\Big| \\ &\leqslant \frac{c|u-u_s|}{h^2}\sum_{i=1}^n v_{ij}\sum_{t=1}^{T_{ij}}\Big|Y_{ijt}\Big| \leqslant \frac{c}{Nh^2}\left\{\mathbb{E}\Big(\sum_{i=1}^n v_{ij}\sum_{t=1}^{T_{ij}}\Big|Y_{ijt}\Big|\Big) + 1\right\} \leqslant \frac{c}{Nh^2}.\end{aligned}$$

Applying similar techniques as above, we can define events $\Omega_{R,jk}$ and $\Omega_{S,jkl}$ for $k,l = 1,2$. On the intersection of these events, we can obtain that $|\widehat{R}_{jk}(u) - \widehat{R}_{jk}(u_s)| \leqslant c(Nh^2)^{-1}$ and $|\widehat{S}_{jkl}(u) - \widehat{S}_{jkl}(u_s)| \leqslant c(Nh^2)^{-1}$. Combing the above results, we have

$$\sup_{u\in\mathcal{U}}\big|\widehat{\mu}_j(u) - \widetilde{\mu}_j(u)\big| \leqslant \max_{s\in[N]}\big|\widehat{\mu}_j(u_s) - \widetilde{\mu}_j(u_s)\big| + \frac{c}{Nh^2}.$$

Applying Hoeffding's inequality, we obtain that $\mathbb{P}(\Omega_{R,jk}^c) \leqslant \exp\{-(2\sum_{i=1}^n c^2 v_{ij}^2 T_{ij}^2)^{-1}\} = \exp(-cn) \leqslant \exp(-c\gamma_{n,T,h,j})$ and $\mathbb{P}(\Omega_{S,jkl}^c) \leqslant \exp\{-(2\sum_{i=1}^n c^2 v_{ij}^2 T_{ij}^2)^{-1}\} = \exp(-cn) \leqslant \exp(-c\gamma_{n,T,h,j})$ for $k,l = 1,2$. It follows from the above results and the union bound of probability with the choice of $N = c(h^2\delta)^{-1}$ that there exist some positive constants c_1 and c_2 such that, for any $\delta \in (0,1]$,

$$\mathbb{P}\Big\{\sup_{u\in\mathcal{U}}|\widehat{\mu}_j(u) - \widetilde{\mu}_j(u)| \geqslant \delta\Big\} \leqslant \frac{c_2}{h^2\delta}\exp(-c_1\gamma_{n,T,h,j}\delta^2). \tag{A.18}$$

Take arbitrarily small $\epsilon_1 > 0$. If $n^{\epsilon_1}\gamma_{n,T,h,j}\delta^2 \geqslant 1$, then the right side of (A.18) reduces to $c_2\{n^{\epsilon_1}\gamma_{n,T,h,j}\}^{1/2}h^{-2}\exp(-c_1\gamma_{n,T,h,j}\delta^2)$. If $n^{\epsilon_1}\gamma_{n,T,h,j}\delta^2 \leqslant 1$, we can choose c_2 and $n^{\epsilon_1} > c$

such that $c_2 \exp(-c_1 c^{-1}) \geqslant 1$ and the same bound $c_1\{n^{\epsilon_1}\gamma_{n,T,h,j}\}^{1/2}h^{-2}\exp(-c_1\gamma_{n,T,h,j}\delta^2)$ can therefore still be used. Hence (7) in Theorem 2 holds, which completes the proof of concentration inequality for the mean estimator in supremum norm.

A.2 Proofs of Theorem 3

We organize the proof in four steps. First, we will define $\widehat{\Sigma}(\cdot,\cdot)$, $\widetilde{\Sigma}(\cdot,\cdot)$ and obtain the decomposition of $\widehat{\Sigma}(\cdot,\cdot) - \widetilde{\Sigma}(\cdot,\cdot)$. Second, we will prove the local concentration inequality for fixed $(u,v) \in \mathcal{U}^2$. Third, we will prove the concentration inequality in Hilbert–Schmidt norm. Finally, we will prove the concentration inequality in the supremum norm.

A.2.1 Definition and decomposition

Without loss of generality, let $h_{\Sigma,jk} = h$ for $(j,k) \in [p]^2$ and denote $\tilde{\mathbf{e}}_0 = (1,0,0)^{\mathrm{T}}$, $\widetilde{\mathbf{U}}_{ijkts}(u,v) = \big\{1, (U_{ijt}-u)/h, (U_{iks}-v)/h\big\}^{\mathrm{T}}$. For $j=k$, let

$$\widehat{\boldsymbol{\Xi}}_{jj}(u,v) = \sum_{i=1}^n w_{ijj} \sum_{1\leqslant t\neq s\leqslant T_{ij}} \widetilde{\mathbf{U}}_{ijjts}(u,v)\widetilde{\mathbf{U}}_{ijjts}^{\mathrm{T}}(u,v)K_h(U_{ijt}-u)K_h(U_{ijs}-v),$$

$$\widehat{\mathbf{Z}}_{jj}(u,v) = \sum_{i=1}^n w_{ijj} \sum_{1\leqslant t\neq s\leqslant T_{ij}} \widetilde{\mathbf{U}}_{ijjts}(u,v)\Theta_{ijjts}K_h(U_{ijt}-u)K_h(U_{ijs}-v).$$

For $j \neq k$, let

$$\widehat{\boldsymbol{\Xi}}_{jk}(u,v) = \sum_{i=1}^n w_{ijk} \sum_{t=1}^{T_{ij}}\sum_{s=1}^{T_{ik}} \widetilde{\mathbf{U}}_{ijkts}(u,v)\widetilde{\mathbf{U}}_{ijkts}^{\mathrm{T}}(u,v)K_h(U_{ijt}-u)K_h(U_{iks}-v),$$

$$\widehat{\mathbf{Z}}_{jk}(u,v) = \sum_{i=1}^n w_{ijk} \sum_{t=1}^{T_{ij}}\sum_{s=1}^{T_{ik}} \widetilde{\mathbf{U}}_{ijkts}(u,v)\Theta_{ijkts}K_h(U_{ijt}-u)K_h(U_{iks}-v).$$

A simple calculation yields $\widehat{\Sigma}_{jk}(u,v) = \tilde{\mathbf{e}}_0^{\mathrm{T}}\{\widehat{\boldsymbol{\Xi}}_{jk}(u,v)\}^{-1}\widehat{\mathbf{Z}}_{jk}(u,v)$. Let

$$\widetilde{\Sigma}_{jk}(u,v) = \tilde{\mathbf{e}}_0^{\mathrm{T}}\big[\mathbb{E}\{\widehat{\boldsymbol{\Xi}}_{jk}(u,v)\}\big]^{-1}\mathbb{E}\{\widehat{\mathbf{Z}}_{jk}(u,v)\}. \tag{A.19}$$

We can decompose $\widehat{\Sigma}_{jk}(u,v) - \widetilde{\Sigma}_{jk}(u,v)$ as

$$\begin{aligned}\widehat{\Sigma}_{jk}(u,v) - \widetilde{\Sigma}_{jk}(u,v) =& \tilde{\mathbf{e}}_0^{\mathrm{T}}(\{\widehat{\boldsymbol{\Xi}}_{jk}(u,v)\}^{-1} - [\mathbb{E}\{\widehat{\boldsymbol{\Xi}}_{jk}(u,v)\}]^{-1})\widehat{\mathbf{Z}}_{jk}(u,v)\\ &+ \tilde{\mathbf{e}}_0^{\mathrm{T}}[\mathbb{E}\{\widehat{\boldsymbol{\Xi}}_{jk}(u,v)\}]^{-1}[\widehat{\mathbf{Z}}_{jk}(u,v) - \mathbb{E}\{\widehat{\mathbf{Z}}_{jk}(u,v)\}],\end{aligned}$$

which further implies that

$$\begin{aligned}&\big|\widehat{\Sigma}_{jk}(u,v) - \widetilde{\Sigma}_{jk}(u,v)\big|\\ \leqslant& \|\mathbb{E}\{\widehat{\boldsymbol{\Xi}}_{jk}(u,v)\}\|_{\min}^{-1}\|\widehat{\mathbf{Z}}_{jk}(u,v) - \mathbb{E}\{\widehat{\mathbf{Z}}_{jk}(u,v)\}\|\\ &+ \|\mathbb{E}\{\widehat{\boldsymbol{\Xi}}_{jk}(u,v)\}\|_{\min}^{-1}\|\widehat{\boldsymbol{\Xi}}_{jk}(u,v)\|_{\min}^{-1}\|\widehat{\mathbf{Z}}_{jk}(u,v)\|\|\widehat{\boldsymbol{\Xi}}_{jk}(u,v) - \mathbb{E}\{\widehat{\boldsymbol{\Xi}}_{jk}(u,v)\}\|_{\mathrm{F}}.\end{aligned} \tag{A.20}$$

In the following, we will prove the concentration results for case $j \neq k$, and the results for the case $j = k$ can be proved in a similar manner.

A.2.2 Local concentration inequality

We will firstly show that there exists some positive constant c (independent of u,v) such that for any $\delta>0$ and $(u,v)\in\mathcal{U}^2$,

$$\mathbb{P}\Big\{\big\|\widehat{\boldsymbol{\Xi}}_{jk}(u,v)-\mathbb{E}\{\widehat{\boldsymbol{\Xi}}_{jk}(u,v)\}\big\|_{\mathrm{F}}\geqslant\delta\Big\}\leqslant 18\exp\Big(-\frac{c\nu_{n,T,h,jk}\delta^2}{1+\delta}\Big). \tag{A.21}$$

For $m,l=1,2,3$, let $\widehat{\Xi}_{jkml}(u,v)$ be the (m,l)th entry of $\widehat{\boldsymbol{\Xi}}_{jk}(u,v)$. It follows from Assumptions 4 and 6 that for any integer $q=2,3,\dots$ and $s,s'=0,1,2$,

$$\begin{aligned}&\mathbb{E}\left\{\Big|\Big(\frac{U_{ijt}-u}{h}\Big)^s\Big(\frac{U_{ikt'}-v}{h}\Big)^{s'}K_h(U_{ijt}-u)K_h(U_{ikt'}-v)\Big|^q\right\}\\&\leqslant\int h^{-2q}K^q\Big(\frac{t-u}{h}\Big)K^q\Big(\frac{t'-v}{h}\Big)\Big|\frac{t-u}{h}\Big|^{sq}\Big|\frac{t'-v}{h}\Big|^{sq}f_U(t)f_U(t')dtdt'\leqslant ch^{2-2q}.\end{aligned} \tag{A.22}$$

Note that Assumption 3 implies that the weights are of the same order $w_{ijk}\asymp(n\bar{T}^2_{\Sigma,jk})^{-1}$. By (A.22),

$$\sum_{i=1}^n\sum_{t=1}^{T_{ij}}\sum_{t'=1}^{T_{ik}}\mathbb{E}\left\{|(\frac{U_{ijt}-u}{h})^s(\frac{U_{ikt'}-v}{h})^{s'}K_h(U_{ijt}-u)K_h(U_{ikt'}-v)|^2\right\}\leqslant cn\bar{T}^2_{\Sigma,jk}h^{-2},$$

$$\sum_{i=1}^n\sum_{t=1}^{T_{ij}}\sum_{t'=1}^{T_{ik}}\mathbb{E}\left\{|(\frac{U_{ijt}-u}{h})^s(\frac{U_{ikt'}-v}{h})^{s'}K_h(U_{ijt}-u)K_h(U_{ikt'}-v)|^q\right\}\leqslant 2^{-1}q!cn\bar{T}^2_{\Sigma,jk}h^{2-2q}$$

for $q\geqslant 3$. Applying the Bernstein inequality yields that there exists some positive constant c (independent of u,v) such that for any $\delta>0$ and $(u,v)\in\mathcal{U}^2$,

$$\mathbb{P}\Big\{\big|\widehat{\Xi}_{jkml}(u,v)-\mathbb{E}\{\widehat{\Xi}_{jkml}(u,v)\}\big|\geqslant\delta\Big\}\leqslant 2\exp\Big(-\frac{c\nu_{n,T,h,jk}\delta^2}{1+\delta}\Big),$$

for $m,l=1,2,3$, which, by the union bound of probability, implies that (A.21) holds.

For $m=1,2,3$, let $\widehat{Z}_{jkm}(u,v)$ be the mth element of $\widehat{\mathbf{Z}}_{jk}(u,v)$. We will next show that, there exits some positive constant c (independent of u,v) such that for any $\delta>0$ and $(u,v)\in\mathcal{U}^2$,

$$\mathbb{P}\Big\{\big|\widehat{Z}_{jkm}(u,v)-\mathbb{E}\{\widehat{Z}_{jkm}(u,v)\}\big|\geqslant\delta\Big\}\leqslant c\exp\Big(-\frac{c\nu_{n,T,h,jk}\delta^2}{1+\delta}\Big), \tag{A.23}$$

where $\nu_{n,T,h,jk}=n(1\wedge\bar{T}^2_{\Sigma,jk}h^2)$. We only need to consider the case $m=1$, while the results for cases $m=2,3$ can be proved in a similar way. Denote that

$$\zeta_{ijkts}=\{Y_{ijt}-\mu_j(U_{ijt})\}\{Y_{iks}-\mu_k(U_{iks})\}-\Sigma_{jk}(U_{ijt},U_{iks}),$$

$$\widehat{Z}_{jk4}(u,v)=\sum_{i=1}^n w_{ijk}\sum_{t=1}^{T_{ij}}\sum_{s=1}^{T_{ik}}K_h(U_{ijt}-u)K_h(U_{ikt'}-v)\Sigma_{jk}(U_{ijt},U_{iks}),$$

$$\widehat{Z}_{jk5}(u,v)=\sum_{i=1}^n w_{ij}\sum_{t=1}^{T_{ij}}\sum_{s=1}^{T_{ik}}K_h(U_{ijt}-u)K_h(U_{iks}-v)\zeta_{ijkts}.$$

Then we rewrite $\widehat{Z}_{jk1}(u,v)-\mathbb{E}\{\widehat{Z}_{jk1}(u,v)\}$ as

$$\widehat{Z}_{jk1}(u,v)-\mathbb{E}\{\widehat{Z}_{jk1}(u,v)\}=\widehat{Z}_{jk4}(u,v)-\mathbb{E}\{\widehat{Z}_{jk4}(u,v)\}+\widehat{Z}_{jk5}(u,v). \tag{A.24}$$

Following the same procedure to prove (A.21) with the aid of the Bernstein inequality, we can obtain that there exists some positive constant c such that for any $\delta>0$ and $(u,v)\in\mathcal{U}^2$,

$$\mathbb{P}\Big\{\big|\widehat{Z}_{jk4}(u,v)-\mathbb{E}\{\widehat{Z}_{jk4}(u,v)\}\big|\geqslant\delta\Big\}\leqslant 2\exp\Big(-\frac{c\nu_{n,T,h,jk}\delta^2}{1+\delta}\Big). \tag{A.25}$$

Now we consider the tail behavior of $\widehat{Z}_{jk5}(u,v)$. Define the event $\widetilde{V}_{jk}=\{(U_{ijt},U_{iks}),t\in[T_{ij}],s\in[T_{ik}],i\in[n]\}$. The sub-Gaussianities under Assumption 1 implies that, conditional on the event $\widetilde{V}_{jk}$, $Y_{ijt}-\mu_j(U_{ijt})$ and $Y_{iks}-\mu_k(U_{iks})$ are sub-Gaussian random variables, then $\{Y_{ijt}-\mu_j(U_{ijt})\}\{Y_{iks}-\mu_k(U_{iks})\}$ is a sub-exponential random variable, and hence we have $\mathbb{E}\{\exp(\lambda\zeta_{ijkts})|\widetilde{V}_{jk}\}\leqslant\exp\big\{(1-c\lambda)^{-1}c\lambda^2\big\}$ for any $\lambda\in(0,c^{-1})$. Rewrite $\widehat{Z}_{jk5}(u,v)=\sum_{i=1}^n w_{ijk}\phi_{ijk1}(u,v)$ with $\phi_{ijk1}(u,v)=\sum_{t=1}^{T_{ij}}\sum_{s=1}^{T_{ik}}K_h(U_{ijt}-u)K_h(U_{iks}-v)\zeta_{ijkts}$. Note that for each i Assumption 3 implies that $w_{ijk}\asymp(n\bar{T}^2_{\Sigma,jk})^{-1}$. If $\sum_{t=1}^{T_{ij}}\sum_{s=1}^{T_{ik}}K_h(U_{ijt}-u)K_h(U_{iks}-v)>0$ holds, it follows from Jensen's inequality that

$$\begin{aligned}
&\mathbb{E}\Big[\exp\{\lambda w_{ijk}\phi_{ijk1}(u,v)\}\Big|\widetilde{V}_{jk}\Big]\\
\leqslant&\frac{1}{\sum_{t=1}^{T_{ij}}\sum_{s=1}^{T_{ik}}K_h(U_{ijt}-u)K_h(U_{iks}-v)}\sum_{t=1}^{T_{ij}}\sum_{s=1}^{T_{ik}}\Big(K_h(U_{ijt}-u)K_h(U_{iks}-v)\\
&\times\mathbb{E}\Big[\exp\Big\{\lambda w_{ijk}\zeta_{ijkts}\sum_{t=1}^{T_{ij}}\sum_{s=1}^{T_{ik}}K_h(U_{ijt}-u)K_h(U_{iks}-v)\Big\}\Big|\widetilde{V}_{jk}\Big]\Big)\\
\leqslant&\exp\left[\frac{c\lambda^2(n\bar{T}^2_{\Sigma,jk})^{-2}\Big\{\sum_{t,s}K_h(U_{ijt}-u)K_h(U_{iks}-v)\Big\}^2}{1-c\lambda(n\bar{T}^2_{\Sigma,jk})^{-1}\sum_{t,s}K_h(U_{ijt}-u)K_h(U_{iks}-v)}\right].
\end{aligned}$$

where $0<\lambda(n\bar{T}^2_{\Sigma,jk})^{-1}\sum_{t=1}^{T_{ij}}\sum_{s=1}^{T_{ik}}K_h(U_{ijt}-u)K_h(U_{iks}-v)<c^{-1}$. It is obvious that the above inequality still holds even if $\sum_{t=1}^{T_{ij}}\sum_{s=1}^{T_{ik}}K_h(U_{ijt}-u)K_h(U_{iks}-v)=0$. Assumption 5 implies that the number of nonzero terms in $\sum_{t=1}^{T_{ij}}\sum_{s=1}^{T_{ik}}K_h(U_{ijt}-u)K_h(U_{iks}-v)$ has an upper bound $c(1\vee\bar{T}^2_{\Sigma,jk}h^2)$, which yields that $\sum_{t=1}^{T_{ij}}\sum_{s=1}^{T_{ik}}K_h(U_{ijt}-u)K_h(U_{iks}-v)\leqslant ch^{-2}(1\vee\bar{T}^2_{\Sigma,jk}h^2)$. Therefore, for any λ satisfying $0<\lambda(n\bar{T}^2_{\Sigma,jk}h^2)^{-1}(1\vee\bar{T}^2_{\Sigma,jk}h^2)<c^{-1}$ for some constant $c>0$, we obtain that

$$\begin{aligned}
&\mathbb{E}\left[\exp\Big\{\lambda\sum_{i=1}^n w_{ijk}\phi_{ijk1}(u,v)\Big\}\Big|\widetilde{V}_{jk}\right]\\
&\leqslant\exp\left\{\frac{c\lambda^2(n\bar{T}^2_{\Sigma,jk})^{-2}h^{-2}(1\vee\bar{T}^2_{\Sigma,jk}h^2)\sum_{i=1}^n\sum_{t=1}^{T_{ij}}\sum_{s=1}^{T_{ik}}K_h(U_{ijt}-u)K_h(U_{iks}-v)}{1-c\lambda(n\bar{T}^2_{\Sigma,jk}h^2)^{-1}(1\vee\bar{T}^2_{\Sigma,jk}h^2)}\right\}.
\end{aligned}$$

For any $\delta > 0$, define the event

$$\Lambda_{jk,1}(\delta) = \left\{ \sum_{i=1}^{n} \sum_{t=1}^{T_{ij}} \sum_{s=1}^{T_{ik}} K_h(U_{ijt} - u) K_h(U_{iks} - v) \leqslant c(1+\delta) n \bar{T}^2_{\Sigma,jk} \right\}.$$

We have

$$\mathbb{E}\left[\exp\left\{ \lambda \sum_{i=1}^{n} w_{ijk} \phi_{ijk1}(u,v) \right\} \Big| \Lambda_{jk,1}(\delta) \right] \leqslant \exp\left\{ \frac{c\lambda^2 (1+\delta)(n\bar{T}^2_{\Sigma,jk} h^2)^{-1} (1 \vee \bar{T}^2_{\Sigma,jk} h^2)}{1 - c\lambda (n\bar{T}^2_{\Sigma,jk} h^2)^{-1}(1 \vee \bar{T}^2_{\Sigma,jk} h^2)} \right\}.$$

Consequently, we obtain that

$$\mathbb{P}\Big\{ \sum_{i=1}^{n} w_{ijk} \phi_{ijk1}(u) \geqslant \delta \Big| \Lambda_{jk,1}(\delta) \Big\} \leqslant \exp\left\{ -\lambda\delta + \frac{c\lambda^2 (1+\delta)(n\bar{T}^2_{\Sigma,jk} h^2)^{-1} (1 \vee \bar{T}^2_{\Sigma,jk} h^2)}{1 - c\lambda (n\bar{T}^2_{\Sigma,jk} h^2)^{-1}(1 \vee \bar{T}^2_{\Sigma,jk} h^2)} \right\}. \tag{A.26}$$

With the choice of $\lambda = n\bar{T}^2_{\Sigma,jk} h^2 \delta \big\{ 2c(1+\delta)(1 \vee \bar{T}^2_{\Sigma,jk} h^2) + c\delta (1 \vee \bar{T}^2_{\Sigma,jk} h^2) \big\}^{-1}$, (A.26) reduces to

$$\mathbb{P}\Big\{ \sum_{i=1}^{n} w_{ijk} \phi_{ijk1}(u,v) \geqslant \delta \Big| \Lambda_{jk,1}(\delta) \Big\} \leqslant \exp\Big\{ -\frac{c\nu_{n,T,h,jk}\delta^2}{1+\delta} \Big\}, \tag{A.27}$$

where the constant c is chosen to satisfy $c\lambda (n\bar{T}^2_{\Sigma,jk} h^2)^{-1} (1 \vee \bar{T}^2_{\Sigma,jk} h^2) \leqslant 1/2$. Note that

$$\sum_{i=1}^{n} \sum_{t=1}^{T_{ij}} \sum_{s=1}^{T_{ik}} \mathbb{E}\big\{ K_h(U_{ijt} - u) K_h(U_{iks} - v) \big\} \leqslant cn\bar{T}^2_{\Sigma,jk}.$$

By the Bernstein inequality, we obtain that there exists some positive constant c such that for any $\delta > 0$

$$\begin{aligned} &\mathbb{P}\left(\sum_{i=1}^{n} \sum_{t=1}^{T_{ij}} \sum_{s=1}^{T_{ik}} \Big[K_h(U_{ijt} - u) K_h(U_{iks} - v) - \mathbb{E}\big\{ K_h(U_{ijt} - u) K_h(U_{iks} - v) \big\} \Big] \geqslant n\bar{T}^2_{\Sigma,jk} \delta \right) \\ \leqslant \quad &\exp\left(\frac{-c\nu_{n,T,h,jk}\delta^2}{1+\delta} \right) \end{aligned}$$

which implies that

$$1 - \mathbb{P}\{\Lambda_{jk,1}(\delta)\} \leqslant \exp\Big(-\frac{c\nu_{n,T,h,jk}\delta^2}{1+\delta} \Big). \tag{A.28}$$

Combining (A.27) and (A.28), we obtain that there exists some constant $c > 0$ such that for any $\delta > 0$,

$$\mathbb{P}\Big\{ \widehat{Z}_{jk5}(u,v) \geqslant \delta \Big\} \leqslant 2\exp\Big\{ -\frac{c\nu_{n,T,h,jk}\delta^2}{1+\delta} \Big\},$$

which leads to

$$\mathbb{P}\Big\{ \big|\widehat{Z}_{jk5}(u,v)\big| \geqslant \delta \Big\} \leqslant 4\exp\Big\{ -\frac{c\nu_{n,T,h,jk}\delta^2}{1+\delta} \Big\}.$$

It follows from the above, (A.24) and (A.25) that for each $\delta > 0$ and $(u,v) \in \mathcal{U}^2$, there exists some positive constant c such that

$$\mathbb{P}\Big\{\big|\widehat{Z}_{jk1}(u,v) - \mathbb{E}\{\widehat{Z}_{jk1}(u,v)\}\big| \geqslant \delta\Big\} \leqslant 6\exp\Big\{-\frac{c\nu_{n,T,h,jk}\delta^2}{1+\delta}\Big\}.$$

Define the event $\Lambda_{jk,2}(\delta) = \big\{\|\widehat{\boldsymbol{\Xi}}_{jk}(u,v) - \mathbb{E}\{\widehat{\boldsymbol{\Xi}}_{jk}(u,v)\}\|_{\mathrm{F}} \leqslant \delta/2\big\}$. Note that $\mathbb{E}\{\widehat{\boldsymbol{\Xi}}_{jk}(u,v)\}$ is positive definite. On the event $\Lambda_{jk,2}(\delta)$ with $\delta \in (0,1]$, we obtain that

$$\|\widehat{\boldsymbol{\Xi}}_{jk}(u,v)\|_{\min} \geqslant c(1-\delta/2). \tag{A.29}$$

By (A.21), we have

$$1 - \mathbb{P}\big\{\Lambda_{jk,2}(\delta)\big\} \leqslant 18\exp\Big(-\frac{c\nu_{n,T,h,jk}\delta^2}{1+\delta}\Big). \tag{A.30}$$

Define the event $\Lambda_{jk,3}(\delta) = \big\{\big\|\widehat{\mathbf{Z}}_{jk}(u,v) - \mathbb{E}\{\widehat{\mathbf{Z}}_{jk}(u,v)\}\big\| \leqslant \delta\big\}$. Note that, under Assumption 3 with $w_{ijk} \asymp (n\bar{T}^2_{\Sigma,jk})^{-1}$, $\sum_{i=1}^n w_{ijk}\sum_{t=1}^{T_{ij}}\sum_{s=1}^{T_{ik}}\mathbb{E}\{K_h(U_{ijt}-u)K_h(U_{iks}-v)\} \leqslant c$, hence $\|\mathbb{E}\{\widehat{\mathbf{Z}}_{jk}(u,v)\}\|$ is uniformly bounded over $\mathcal{U}^2$. On the event $\Omega_{jk,3}(\delta)$, we have

$$\big\|\widehat{\mathbf{Z}}_{jk}(u,v)\big\| \leqslant c(1+\delta). \tag{A.31}$$

On the event $\Lambda_{jk,2}(\delta) \cap \Lambda_{jk,3}(\delta)$ with $\delta \in (0,1]$, it follows from (A.20), (A.29) and (A.31) that

$$\big|\widehat{\Sigma}_{jk}(u,v) - \widetilde{\Sigma}_{jk}(u,v)\big| \leqslant c\delta + c(1-\delta/2)^{-1}(1+\delta)\delta \leqslant c_3\delta.$$

This together with concentration inequalities in (A.23) and (A.30) implies that there exist some positive universal constants c_1 and c_2 such that for any $\delta \in (0,1]$ and $(u,v) \in \mathcal{U}^2$,

$$\mathbb{P}\Big\{\big|\widehat{\Sigma}_{jk}(u,v) - \widetilde{\Sigma}_{jk}(u,v)\big| \geqslant \delta\Big\} \leqslant c_2\exp\left(-c_1\nu_{n,T,h,jk}\delta^2\right),$$

which completes the proof of local concentration inequality for the covariance estimator.

A.2.3 Concentration inequality in Hilbert–Schmidt norm

We will derive the L_2 concentration inequality of $\|\widehat{\Sigma}_{jk} - \widetilde{\Sigma}_{jk}\|_{\mathcal{S}}$. Let

$$\widetilde{\boldsymbol{\Xi}}_{jk}(u,v) = (n\bar{T}^2_{\Sigma,jk})^{-1}\sum_{i=1}^n\sum_{t=1}^{T_{ij}}\sum_{s=1}^{T_{ik}}\widetilde{\mathbf{U}}_{ijkts}\widetilde{\mathbf{U}}_{ijkts}^{\mathrm{T}}K_h(U_{ijt}-u)K_h(U_{iks}-v).$$

Then we have that $\|\widehat{\boldsymbol{\Xi}}_{jk}(u,v)\|_{\min} \geqslant c\|\widetilde{\boldsymbol{\Xi}}_{jk}(u,v)\|_{\min}$. Similar to Appendix A.1.3, we will give a lower bound on $\|\widetilde{\boldsymbol{\Xi}}_{jk}(u,v)\|_{\min}$. Denote $\widetilde{W} = \sup_{(u,v)\in\mathcal{U}^2}\big\|\widetilde{\boldsymbol{\Xi}}_{jk}(u,v) - \mathbb{E}\{\widetilde{\boldsymbol{\Xi}}_{jk}(u,v)\}\big\|_{\mathrm{F}}$. For $t,s = 1,2,3$, let $\widetilde{\boldsymbol{\Xi}}_{jkts}(u,v)$ be the (t,s)th entry of $\widetilde{\boldsymbol{\Xi}}_{jk}(u,v)$. Note that $\mathbb{E}\{|(U_{ijt}-u)^a h^{-a}(U_{iks}-v)^b h^{-b}K_h(U_{ijt}-u)K_h(U_{iks}-v)|\} \leqslant c$ for $a,b = 0,1,2$, and, moreover, $\mathbb{E}(\widetilde{W}) \leqslant 6\max_{t,s}\mathbb{E}\{\sup_{(u,v)\in\mathcal{U}^2}|\widetilde{\boldsymbol{\Xi}}_{jkts}(u,v)|\}$. In an analogy to Lemma 13.5 of Boucheron et al. (2014) and by the similar arguments below (A.14) in Appendix A.1.3, we can show that $\mathbb{E}(\widetilde{W}) \leqslant c(n\bar{T}^2_{\Sigma,jk})^{-1/2}$. In addition, it follows from the facts $\mathrm{var}(W) \leqslant \mathbb{E}(W^2) \leqslant 9\max_{t,s}\mathrm{var}\{\sup_{(u,v)\in\mathcal{U}^2}\widetilde{\boldsymbol{\Xi}}_{jkts}(u,v)\}$, $|(U_{ijt}-u)^a h^{-a}(U_{iks}-v)^b h^{-b}K_h(U_{ijt}-u)K_h(U_{iks}-v)| \leqslant$

ch^{-2}, $\mathbb{E}\{(U_{ijt}-u)^{2a}h^{-2a}(U_{iks}-v)^{2b}h^{-2b}K_h^2(U_{ijt}-u)K_h^2(U_{iks}-v)\} \leqslant ch^{-2}$ for $a,b=0,1,2$, and Theorem 11.10 of Boucheron et al. (2014) that $\text{var}(n\bar{T}^2_{\Sigma,jk}h^2\widetilde{W}) \leqslant 2\mathbb{E}(n\bar{T}^2_{\Sigma,jk}h^2\widetilde{W}) + \sum_{i=1}^n\sum_{t=1}^{T_{ij}}\sum_{s=1}^{T_{ik}} ch^{-2}h^4 \leqslant c(n\bar{T}^2_{\Sigma,jk})^{1/2}h^2 + cn\bar{T}^2_{\Sigma,jk}h^2$, which implies that the variance of $\widetilde{W}$ is bounded by $c(n\bar{T}^2_{\Sigma,jk}h^2)^{-1} \leqslant c(\nu_{n,T,h,jk})^{-1}$. Noting the similar arguments below (A.14) and applying Theorem 12.5 of Boucheron et al. (2014), we obtain that there exists some positive constant c such that, for any $\delta > 0$,

$$\mathbb{P}\Big\{\widetilde{W} - \mathbb{E}(\widetilde{W}) > \delta\Big\} \leqslant \exp\Big(-\frac{c\nu_{n,T,h,jk}\delta^2}{1+\delta}\Big). \tag{A.32}$$

Define the event $\Lambda_{jk,4}(\delta) = \big\{\sup_{(u,v)\in\mathcal{U}^2}\big\|\widetilde{\boldsymbol{\Xi}}_{jk}(u,v) - \mathbb{E}\{\widetilde{\boldsymbol{\Xi}}_{jk}(u,v)\}\big\|_{\text{F}} \leqslant \delta/2\big\}$ with $\delta \in (0,1]$. By (A.32), we obtain that there exists some constant $c > 0$ such that, for any $\delta \in (0,1]$,

$$1 - \mathbb{P}\left\{\Lambda_{jk,4}(\delta)\right\} \leqslant 2\exp\left(-c\nu_{n,T,h,jk}\delta^2\right).$$

On the event $\Lambda_{jk,4} = \Lambda_{jk,4}(\tilde{\delta}_1)$ with $c(\nu_{n,T,h,jk})^{-1/2} < \tilde{\delta}_1 \leqslant 1$, we have $\|\widehat{\boldsymbol{\Xi}}_{jk}(u,v)\|_{\min} \geqslant c\|\widetilde{\boldsymbol{\Xi}}_{jk}(u,v)\|_{\min} \geqslant c(1-\tilde{\delta}_1/2) \geqslant c/2$. Notice that $\mathbb{E}\{\widehat{\boldsymbol{\Xi}}_{jk}(u,v)\}$ is positive definite and $\|\mathbb{E}\{\widehat{\mathbf{Z}}_{jk}(u,v)\}\|$ is uniformly bounded over $\mathcal{U}^2$. On the event $\Lambda_{jk,4}$, it thus follows from (A.20) and $\|\widehat{\mathbf{Z}}_{jk}(u,v)\| \leqslant \|\widehat{\mathbf{Z}}_{jk}(u,v) - \mathbb{E}\{\widehat{\mathbf{Z}}_{jk}(u,v)\}\| + \|\mathbb{E}\{\widehat{\mathbf{Z}}_{jk}(u,v)\}\|$ that

$$|\widehat{\Sigma}_{jk}(u,v) - \widetilde{\Sigma}_{jk}(u,v)| \leqslant c\|\widehat{\mathbf{Z}}_{jk}(u,v) - \mathbb{E}\{\widehat{\mathbf{Z}}_{jk}(u,v)\}\| + c\|\widehat{\boldsymbol{\Xi}}_{jk}(u,v) - \mathbb{E}\{\widehat{\boldsymbol{\Xi}}_{jk}(u,v)\}\|_{\text{F}} \tag{A.33}$$

and the positive constant c does not depend on $(u,v) \in \mathcal{U}^2$.

Combining (A.33) with (A.21), (A.23) and applying the first part of Lemma 10 yields that, for any $(u,v) \in \mathcal{U}^2$ and integer $q \geqslant 1$,

$$\mathbb{E}\Big\{|\widehat{\Sigma}_{jk}(u,v) - \widetilde{\Sigma}_{jk}(u,v)|^{2q}\Big|\Lambda_{jk,4}\Big\} \leqslant q!c\Big(\frac{4}{c\nu_{n,T,h,jk}}\Big)^q + (2q)!c\Big(\frac{4}{c\nu_{n,T,h,jk}}\Big)^{2q}.$$

Applying the second part of Lemma 10, we can show that, for each $\delta \in (0,1]$,

$$\mathbb{P}\Big(\|\widehat{\Sigma}_{jk} - \widetilde{\Sigma}_{jk}\|_{\mathcal{S}} \geqslant \delta\Big) \leqslant \mathbb{P}\Big(\|\widehat{\Sigma}_{jk} - \widetilde{\Sigma}_{jk}\|_{\mathcal{S}} \geqslant \delta\big|\Lambda_{jk,4}\Big) + \mathbb{P}\big(\Lambda^c_{jk,4}\big) \leqslant c_4\exp\big(-c_3\nu_{n,T,h,jk}\delta^2\big),$$

which means (8) in Theorem 3 holds and completes the proof of concentration inequality for the covariance estimator in Hilbert–Schmidt norm.

A.2.4 Concentration inequality in supremum norm

We will derive the uniform concentration bound of $\sup_{(u,v)\in\mathcal{U}^2}|\widehat{\Sigma}_{jk}(u,v) - \widetilde{\Sigma}_{jk}(u,v)|$. We partition the interval $\mathcal{U} = [0,1]$ into N subintervals I_s for $s \in [N]$ of equal length. Let u_s and $v_{s'}$ be the centers of I_s and $I_{s'}$, respectively, then we have

$$\begin{aligned}\sup_{(u,v)\in\mathcal{U}^2}|\widehat{\Sigma}_{jk}(u,v) - \widetilde{\Sigma}_{jk}(u,v)| \leqslant \max_{s,s'\in[N]}\Big[&\big|\widehat{\Sigma}_{jk}(u_s,v_{s'}) - \widetilde{\Sigma}_{jk}(u_s,v_{s'})\big| \\ &+ \big|\{\widehat{\Sigma}_{jk}(u,v) - \widehat{\Sigma}_{jk}(u_s,v_{s'})\} - \{\widetilde{\Sigma}_{jk}(u,v) - \widetilde{\Sigma}_{jk}(u_s,v_{s'})\}\big|\Big].\end{aligned}$$

We need to bound the second term. By some calculations, it suffices to bound $\Big|\{\widehat{Z}_{jkm}(u,v)-\widehat{Z}_{jkm}(u_s,v_{s'})\}-\big[\mathbb{E}\{\widehat{Z}_{jkm}(u,v)\}-\mathbb{E}\{\widehat{Z}_{jkm}(u_s,v_{s'})\}\big]\Big|$ and $\Big|\{\widehat{\Xi}_{jkml}(u,v)-\widehat{\Xi}_{jkml}(u_s,v_{s'})\}-\big[\mathbb{E}\{\widehat{\Xi}_{jkml}(u,v)\}-\mathbb{E}\{\widehat{\Xi}_{jkml}(u_s,v_{s'})\}\big]\Big|$ for $m,l=1,2,3$, which means that we need to bound $\big|\widehat{Z}_{jkm}(u,v)-\widehat{Z}_{jkm}(u_s,v_{s'})\big|$ and $\big|\widehat{\Xi}_{jkml}(u,v)-\widehat{\Xi}_{jkml}(u_s,v_{s'})\big|$. Let $(u,v)\in I_s\times I_{s'}$ and consider $|\widehat{Z}_{jk1}(u,v)-\widehat{Z}_{jk1}(u_s,v_{s'})|$ for the case of $j\neq k$ first. The results for the case of $j=k$ can be proved in a similar fashion. Define the event $\Lambda_{Z,jk1}=\big\{\sum_{i=1}^n w_{ijk}\sum_{t=1}^{T_{ij}}\sum_{t'=1}^{T_{ik}}|\Theta_{ijktt'}|\leqslant \mathbb{E}(\sum_{i=1}^n w_{ijk}\sum_{t=1}^{T_{ij}}\sum_{t'=1}^{T_{ik}}|\Theta_{ijktt'}|)+1\big\}$. On this event, it follows from Assumption 6(ii) that

$$
\begin{aligned}
&\Big|\widehat{Z}_{jk1}(u,v)-\widehat{Z}_{jk1}(u_s,v_{s'})\Big|\\
\leqslant&\Big|\sum_{i=1}^n w_{ijk}\sum_{t=1}^{T_{ij}}\sum_{t'=1}^{T_{ik}}\Theta_{ijktt'}\Big[\{K_h(U_{ijt}-u)-K_h(U_{ijt}-u_s)\}K_h(U_{ikt'}-v)\\
&+\big\{K_h(U_{ikt'}-v)-K_h(U_{ikt'}-v_{s'})\big\}K_h(U_{ijt}-u_s)\Big]\Big|\\
\leqslant&\frac{c(|u-u_s|\vee|v-v_{s'}|)}{h^2}\sum_{i=1}^n w_{ijk}\sum_{t=1}^{T_{ij}}\sum_{t'=1}^{T_{ik}}\Big|\Theta_{ijktt'}\Big|\big\{K_h(U_{ikt'}-v)+K_h(U_{ijt}-u_s)\big\}\\
\leqslant&\frac{c}{Nh^3}\left\{\mathbb{E}\Big(\sum_{i=1}^n w_{ijk}\sum_{t=1}^{T_{ij}}\sum_{t'=1}^{T_{ik}}\Big|\Theta_{ijktt'}\Big|\Big)+1\right\}\leqslant\frac{c}{Nh^3}.
\end{aligned}
$$

Applying similar techniques as above, we can define events $\Lambda_{Z,jkm}$ and $\Lambda_{\Xi,jkml}$ for $m,l=1,2,3$. On the intersection of these events, we can obtain that $|\widehat{Z}_{jkm}(u,v)-\widehat{Z}_{jkm}(u_s,v_{s'})|\leqslant c(Nh^3)^{-1}$ and $|\widehat{\Xi}_{jkml}(u,v)-\widehat{\Xi}_{jkml}(u_s,v_{s'})|\leqslant c(Nh^3)^{-1}$. Combing the above results, we have

$$
\sup_{(u,v)\in\mathcal{U}^2}|\widehat{\Sigma}_{jk}(u,v)-\widetilde{\Sigma}_{jk}(u,v)|\leqslant\max_{s,s'\in[N]}\big|\widehat{\Sigma}_{jk}(u_s,v_{s'})-\widetilde{\Sigma}_{jk}(u_s,v_{s'})\big|+\frac{c}{Nh^3}.
$$

By the Bernstein inequality, we have $\mathbb{P}(\Lambda_{Z,jkm}^{\mathrm{c}})\leqslant\exp\{-\lambda+c_1^2\lambda^2(n-c_2\lambda)^{-1}\}$ for $\lambda\in(0,nc^{-1}]$. With $\lambda=n(2c_1^2+c_2)^{-1}$, the right-side reduces to $\exp(-cn)\leqslant\exp(-c\nu_{n,T,h,jk})$ for $m,l=1,2,3$. Similarly, $\mathbb{P}(\Lambda_{\Xi,jkml}^{\mathrm{c}})\leqslant\exp(-cn)\leqslant\exp(-c\nu_{n,T,h,jk})$ for $m,l=1,2,3$. It follows from the above results and the union bound of probability with the choice of $N=c(h^3\delta)^{-1}$ that there exists some positive constants c_3 and c_4 such that, for any $\delta\in(0,1]$,

$$
\mathbb{P}\Big\{\sup_{(u,v)\in\mathcal{U}^2}|\widehat{\Sigma}_{jk}(u,v)-\widetilde{\Sigma}_{jk}(u,v)|\geqslant\delta\Big\}\leqslant\frac{c_4}{h^6\delta^2}\exp(-c_3\nu_{n,T,h,jk}\delta^2).\tag{A.34}
$$

Take arbitrarily small $\epsilon_2>0$. If $n^{\epsilon_2}\nu_{n,T,h,jk}\delta^2\geqslant1$, the right side of (A.34) reduces to $c_4n^{\epsilon_2}\nu_{n,T,h,jk}h^{-6}\exp(-c_3\nu_{n,T,h,jk}\delta^2)$. If $n^{\epsilon_2}\nu_{n,T,h,jk}\delta^2\leqslant1$, we can choose c_4 and $n^{\epsilon_2}>c$ such that $c_4\exp(-c_3c^{-1})\geqslant1$ and hence the same bound $c_4n^{\epsilon_2}\nu_{n,T,h,jk}h^{-6}\exp(-c_3\nu_{n,T,h,jk}\delta^2)$ can still be used. We complete the proof of (9) in Theorem 3, the concentration inequality for the covariance estimator in supremum norm.

A.3 Proof of Theorem 6

Note that $\|\widehat{\mu}_j - \mu_j\| \leqslant \|\widehat{\mu}_j - \widetilde{\mu}_j\| + \|\widetilde{\mu}_j - \mu_j\|$, it suffices to bound $\|\widetilde{\mu}_j - \mu_j\|$. By (A.1), for any $u \in \mathcal{U}$,

$$\widetilde{\mu}_j(u) - \mu_j(u) = \mathbf{e}_0^{\mathrm{T}}\Big[\mathbb{E}\{\widehat{\mathbf{S}}_j(u)\}\Big]^{-1}\mathbb{E}\Big[\widehat{\mathbf{R}}_j(u) - \widehat{\mathbf{S}}_j(u)\{\mu_j(u), 0\}^{\mathrm{T}}\Big].$$

By the Taylor expansion, we have

$$\mathbb{E}_\varepsilon\Big[\widehat{\mathbf{R}}_j(u) - \widehat{\mathbf{S}}_j(u)\{\mu_j(u), 0\}^{\mathrm{T}}\Big|V_j\Big] = \sum_{i=1}^{n} v_{ij}\sum_{t=1}^{T_{ij}} \widetilde{\mathbf{U}}_{ijt}K_{h_{\mu,j}}(U_{ijt}-u)\big\{\mu_j(U_{ijt}) - \mu_j(u)\big\} := \mathbf{J}_1 + \mathbf{J}_2,$$

with

$$\mathbf{J}_1 = \sum_{i=1}^{n} v_{ij}\sum_{t=1}^{T_{ij}} \widetilde{\mathbf{U}}_{ijt}K_{h_{\mu,j}}(U_{ijt}-u)\frac{U_{ijt}-u}{h_{\mu,j}}h_{\mu,j}\frac{\partial \mu_j(u)}{\partial u},$$
$$\mathbf{J}_2 = \frac{1}{2}\sum_{i=1}^{n} v_{ij}\sum_{t=1}^{T_{ij}} \widetilde{\mathbf{U}}_{ijt}K_{h_{\mu,j}}(U_{ijt}-u)\Big(\frac{U_{ijt}-u}{h_{\mu,j}}\Big)^2 h_{\mu,j}^2\frac{\partial \mu_j^2(\delta_{ijt})}{\partial u^2},$$

where $\delta_{ijt} \in [u - h_{\mu,j}, u + h_{\mu,j}]$ and $\mathbb{E}_\varepsilon$ denotes the expectation over $\{Y_{ijt}\}$ in (3) conditional on the event $V_j = \{U_{ijt}, t \in [T_{ij}], i \in [n]\}$. First consider $\mathbf{J}_1$, which equals to the second column of $\widehat{\mathbf{S}}_j(u)$ multiplied by $h_{\mu,j}\partial\mu_j(u)/\partial u$, hence

$$\mathbb{E}_U\big[\mathbf{e}_0^{\mathrm{T}}[\mathbb{E}\{\widehat{\mathbf{S}}_j(u)\}]^{-1}\mathbf{J}_1\big] = h_{\mu,j}\frac{\partial\mu_j(u)}{\partial u}\mathbf{e}_0^{\mathrm{T}}[\mathbb{E}\{\widehat{\mathbf{S}}_j(u)\}]^{-1}\mathbb{E}\{\widehat{\mathbf{S}}_j(u)\}(0,1)^{\mathrm{T}} = 0,$$

where $\mathbb{E}_U$ denotes the expectation over V_j. Consider $\mathbf{J}_2$ next. Under Assumption 7, we have $K_1 = \sup_{j\in[p],\xi\in\mathcal{U}}|\partial\mu_j^2(\xi)/\partial u^2| < \infty$. Each entry of $|\mathbf{J}_2|$ is bounded by the $(1,1)$th entry of $\widehat{\mathbf{S}}_j(u)$ multiplied by $K_1h_{\mu,j}^2/2$, and by $\mathbb{E}\{K_h(U_{ijt}-u)\} \leqslant 1$ we have that the $(1,1)$th entry of $\mathbb{E}\{\widehat{\mathbf{S}}_j(u)\}$ is bounded by 1. Note that $\mathbb{E}\{\widehat{\mathbf{S}}_j(u)\}$ is positive definite. These results together yield that

$$\big|\mathbb{E}_U(\mathbf{e}_0^{\mathrm{T}}[\mathbb{E}\{\widehat{\mathbf{S}}_j(u)\}]^{-1}\mathbf{J}_2)\big| \leqslant \|\mathbb{E}\{\widehat{\mathbf{S}}_j(u)\}\|_{\min}^{-1}\|\mathbb{E}_U(|\mathbf{J}_2|)\| \leqslant cK_1h_{\mu,j}^2,$$

which implies that $|\widetilde{\mu}_j(u) - \mu_j(u)| \leqslant cK_1h_{\mu,j}^2$ for any $u \in \mathcal{U}$. Hence

$$\sup_{u\in\mathcal{U}}\big|\widetilde{\mu}_j(u) - \mu_j(u)\big| = O(h_{\mu,j}^2) \;\text{ and }\; \big\|\widetilde{\mu}_j - \mu_j\big\| = O(h_{\mu,j}^2). \tag{A.35}$$

For $M > 0$ with the choice of $\delta = (\log p)^{1/2}(\min_j \gamma_{n,T,h,j})^{-1/2}M \leqslant 1$, it follows from the union bound of probability and (6) in Theorem 2 that

$$\begin{aligned}&\mathbb{P}\Big\{\frac{\max_{j\in[p]}\|\widehat{\mu}_j - \widetilde{\mu}_j\|}{(\log p)^{1/2}(\min_j\gamma_{n,T,h,j})^{-1/2}} \geqslant M\Big\}\\ \leqslant\;& \sum_{j=1}^{p}\mathbb{P}\Big\{\|\widehat{\mu}_j - \widetilde{\mu}_j\| \geqslant M(\log p)^{1/2}(\min_j\gamma_{n,T,h,j})^{-1/2}\Big\}\\ \leqslant\;& \sum_{j=1}^{p}c\exp\Big(-c\gamma_{n,T,h,j}M^2\frac{\log p}{\min_j\gamma_{n,T,h,j}}\Big) \leqslant c\exp\{(1-cM^2)\log p\}.\end{aligned} \tag{A.36}$$

We can choose a large M such that $1 - cM^2 < 0$, the right-side of (A.36) tends to 0. Hence $\max_{j\in[p]} \|\widehat{\mu}_j - \widetilde{\mu}_j\| = O_P\{(\log p)^{1/2}(\min_j \gamma_{n,T,h,j})^{-1/2}\}$. Combing this with (A.35) yields that

$$\max_{j\in[p]} \|\hat{\mu}_j - \mu_j\| = O_P\left\{\Big(\frac{\log p}{\min_j \gamma_{n,T,h,j}}\Big)^{1/2} + \max_j h^2_{\mu,j}\right\},$$

which completes the proof of (10).

The rate of convergence in (11) can be proved following a similar procedure. Let $h_{\mu,\min} = \min_j h_{\mu,j}$, we assume that $h_{\mu,\min} \asymp \{\log(p \vee n)/n\}^{\kappa_1}$ where $\kappa_1 \in (0, 1/2]$. Consider $h_{\mu,\min}$ with some $\kappa_1^* > 1/2$, the corresponding rate is not faster than that with $\kappa_1 \in (0, 1/2]$. To be specific, under the sparse design, the rate of $\max_{j\in[p]} \sup_{u\in\mathcal{U}} |\hat{\mu}_j(u) - \mu_j(u)|$ is $\{\log(p \vee n)/n\}^{1/2-\kappa_1^*/2}$, which is slower than $\{\log(p \vee n)/n\}^{2/5}$ with $\kappa_1 = 1/5$. Under the dense design with $\bar{T}_\mu\{\log(p \vee n)/n\}^{\kappa_1^*} \to 0$ and $\bar{T}_\mu\{\log(p \vee n)/n\}^{3/2} \to 0$, the rate is $\{\log(p \vee n)/n\}^{1/2-\kappa_1^*/2}\bar{T}_\mu^{-1/2}$, which is slower than $\{\log(p\vee n)/n\}^{2/5}\bar{T}_\mu^{-2/5}$ with $\kappa_1 \in (1/5, 1/2]$. Under the dense design with $\bar{T}_\mu\{\log(p \vee n)/n\}^{\kappa_1^*} \to 0$ and $\bar{T}_\mu\{\log(p \vee n)/n\}^{3/2} \to \tilde{c}$ or ∞, the rate is $\{\log(p\vee n)/n\}^{1/2-\kappa_1^*/2}\bar{T}_\mu^{-1/2}$, which is slower than $\{\log(p\vee n)/n\}^{1/2}$ with $\kappa_1 = 1/4$. Under the dense design with $\bar{T}_\mu\{\log(p \vee n)/n\}^{\kappa_1^*} \to \tilde{c}$ or ∞, the rate is $\{\log(p \vee n)/n\}^{1/2}$, which is the same as $\{\log(p \vee n)/n\}^{1/2}$ with $\kappa_1 = 1/4$. Based on the above four cases, if $\kappa_1^* > 1/2$, the corresponding rate is not faster than that with some $\kappa_1 \in (0, 1/2]$ and hence the κ_1 that corresponds with the optimal bandwidth under sparse or dense design is in $(0, 1/2]$. For $M > 0$ with the choice of $\delta = \{\log(p \vee n)\}^{1/2}(\min_j \gamma_{n,T,h,j})^{-1/2}M \leqslant 1$, by the union bound of probability and (7) in Theorem 2, we have

$$\begin{aligned}
&\mathbb{P}\Big\{\frac{\max_{j\in[p]}\sup_{u\in\mathcal{U}} |\widehat{\mu}_j(u) - \widetilde{\mu}_j(u)|}{\{\log(p \vee n)\}^{1/2}(\min_j \gamma_{n,T,h,j})^{-1/2}} \geqslant M\Big\}\\
\leqslant\ & \sum_{j=1}^{p} \mathbb{P}\Big\{\sup_{u\in\mathcal{U}} |\widehat{\mu}_j(u) - \widetilde{\mu}_j(u)| \geqslant M\{\log(p \vee n)\}^{1/2}(\min_j \gamma_{n,T,h,j})^{-1/2}\Big\}\\
\leqslant\ & \sum_{j=1}^{p} \frac{c(n^{\epsilon_1}\gamma_{n,T,h,j})^{1/2}}{h^2_{\mu,\min}} \exp\Big\{-c\gamma_{n,T,h,j}M^2\frac{\log(p \vee n)}{\min_j \gamma_{n,T,h,j}}\Big\}\\
\leqslant\ & c\exp\Big\{\Big(\frac{3+\epsilon_1}{2} + 2\kappa_1 - cM^2\Big)\log(p \vee n)\Big\}. \qquad (A.37)
\end{aligned}$$

We can choose a large M such that $3 + \epsilon_1 + 4\kappa_1 - 2cM^2 < 0$, the right-side of (A.37) tends to 0. Combing this with (A.35) yields that

$$\max_{j\in[p]}\sup_{u\in\mathcal{U}} \big|\hat{\mu}_j(u) - \mu_j(u)\big| = O_P\left[\Big\{\frac{\log(p \vee n)}{\min_j \gamma_{n,T,h,j}}\Big\}^{1/2} + \max_j h^2_{\mu,j}\right],$$

which completes the proof of (11).

A.4 Proof of Theorem 7

Note $\|\widehat{\Sigma}_{jk} - \Sigma_{jk}\|_{\mathcal{S}} \leqslant \|\widehat{\Sigma}_{jk} - \widetilde{\Sigma}_{jk}\|_{\mathcal{S}} + \|\widetilde{\Sigma}_{jk} - \Sigma_{jk}\|_{\mathcal{S}}$, it suffices to bound $\|\widetilde{\Sigma}_{jk} - \Sigma_{jk}\|_{\mathcal{S}}$. By (A.19), for any $(u, v) \in \mathcal{U}^2$,

$$\widetilde{\Sigma}_{jk}(u, v) - \Sigma_{jk}(u, v) = \tilde{\mathbf{e}}_0^{\mathrm{T}}\Big[\mathbb{E}\big\{\widehat{\boldsymbol{\Xi}}_{jk}(u, v)\big\}\Big]^{-1}\mathbb{E}\Big[\widehat{\mathbf{Z}}_{jk}(u, v) - \widehat{\boldsymbol{\Xi}}_{jk}(u, v)\big\{\Sigma_{jk}(u, v), 0, 0\big\}^{\mathrm{T}}\Big].$$

By the Taylor expansion, we have

$$\mathbb{E}_\varepsilon\Big[\widehat{\mathbf{Z}}_{jk}(u,v)-\widehat{\boldsymbol{\Xi}}_{jk}(u,v)\big\{\Sigma_{jk}(u,v),0,0\big\}^{\mathrm{T}}|\widetilde{V}_{jk}\Big]$$
$$=\sum_{i=1}^{n}w_{ijk}\sum_{t=1}^{T_{ij}}\sum_{s=1}^{T_{ik}}\widetilde{\mathbf{U}}_{ijt}K_{h_{\Sigma,jk}}(U_{ijt}-u)\big\{\Sigma_{jk}(U_{ijt},U_{iks})-\Sigma_{jk}(u,v)\big\}:=\mathbf{L}_1+\mathbf{L}_2+\mathbf{L}_3,$$

where

$$\mathbf{L}_1=\sum_{i=1}^{n}w_{ijk}\sum_{t=1}^{T_{ij}}\sum_{s=1}^{T_{ik}}\widetilde{\mathbf{U}}_{ijt}K_{h_{\Sigma,jk}}(U_{ijt}-u)\frac{U_{ijt}-u}{h_{\Sigma,jk}}h_{\Sigma,jk}\frac{\partial\Sigma_{jk}}{\partial u}(u,v),$$
$$\mathbf{L}_2=\sum_{i=1}^{n}w_{ijk}\sum_{t=1}^{T_{ij}}\sum_{s=1}^{T_{ik}}\widetilde{\mathbf{U}}_{ijt}K_{h_{\Sigma,jk}}(U_{ijt}-u)\frac{U_{iks}-v}{h_{\Sigma,jk}}h_{\Sigma,jk}\frac{\partial\Sigma_{jk}}{\partial v}(u,v),$$
$$\mathbf{L}_3=\sum_{i=1}^{n}w_{ijk}\sum_{t=1}^{T_{ij}}\sum_{s=1}^{T_{ik}}\widetilde{\mathbf{U}}_{ijt}K_{h_{\Sigma,jk}}(U_{ijt}-u)\widetilde{L}_{ijk},$$
$$\widetilde{L}_{ijk}=\frac{1}{2}h_{\Sigma,jk}^2\Big(\frac{U_{ijt}-u}{h_{\Sigma,jk}}\Big)^2\frac{\partial^2\Sigma_{jk}}{\partial u^2}(\delta_{ijkts1},\delta_{ijkts2})+\frac{1}{2}h_{\Sigma,jk}^2\Big(\frac{U_{iks}-v}{h_{\Sigma,jk}}\Big)^2\frac{\partial^2\Sigma_{jk}}{\partial v^2}(\delta_{ijkts1},\delta_{ijkts2})$$
$$+h_{\Sigma,jk}^2\frac{U_{ijt}-u}{h_{\Sigma,jk}}\frac{U_{iks}-v}{h_{\Sigma,jk}}\frac{\partial^2\Sigma_{jk}}{\partial u\partial v}(\delta_{ijkts1},\delta_{ijkts2}),$$

$(\delta_{ijkts1},\delta_{ijkts2})\in[u-h_{\Sigma,jk},u+h_{\Sigma,jk}]\times[v-h_{\Sigma,jk},v+h_{\Sigma,jk}]$ and the event $\widetilde{V}_{jk}=\{(U_{ijt},U_{iks}),t\in[T_{ij}],s\in[T_{ik}],i\in[n]\}$. First consider L_1, which equals to the second column of $\widehat{\boldsymbol{\Xi}}_{jk}(u,v)$ multiplied by $h_{\Sigma,jk}\partial\Sigma_{jk}(u,v)/\partial u$, hence $\mathbb{E}_U\Big(\tilde{\mathbf{e}}_0^{\mathrm{T}}\Big[\mathbb{E}\big\{\widehat{\boldsymbol{\Xi}}_{jk}(u,v)\big\}\Big]^{-1}\mathbf{L}_1\Big)$ equals to

$$h_{\Sigma,jk}\frac{\partial\Sigma_{jk}}{\partial u}(u,v)\tilde{\mathbf{e}}_0^{\mathrm{T}}\Big[\mathbb{E}\big\{\widehat{\boldsymbol{\Xi}}_{jk}(u,v)\big\}\Big]^{-1}\mathbb{E}\Big\{\widehat{\boldsymbol{\Xi}}_{jk}(u,v)\Big\}(0,1,0)^{\mathrm{T}}=0,$$

where $\mathbb{E}_U$ denotes the expectation over $\widetilde{V}_{jk}$. Following the similar procedure, we can show that

$$\mathbb{E}_U\Big(\tilde{\mathbf{e}}_0^{\mathrm{T}}\Big[\mathbb{E}\big\{\widehat{\boldsymbol{\Xi}}_{jk}(u,v)\big\}\Big]^{-1}\mathbf{L}_2\Big)=0.$$

Then consider $\mathbf{L}_3$. By Assumption 7, we have

$$K_2=\sup_{(j,k)\in[p]^2,(u,v)\in\mathcal{U}^2}\{\partial^2\Sigma_{jk}(u,v)/\partial v^2,\partial^2\Sigma_{jk}(u,v)/\partial u^2,\partial^2\Sigma_{jk}(u,v)/\partial u\partial v\}<\infty.$$

Each entry of $|\mathbf{L}_3|$ is bounded by the $(1,1)$th entry of $\widehat{\boldsymbol{\Xi}}_{jk}(u,v)$ multiplied by $2K_2h_{\Sigma,jk}^2$, and by $\mathbb{E}\{K_h(U_{ijt}-u)K_h(U_{iks}-v)\}\leqslant 1$, the $(1,1)$th entry of $\mathbb{E}\{\widehat{\boldsymbol{\Xi}}_{jk}(u,v)\}$ is bounded by 1. Note that $\mathbb{E}\{\widehat{\boldsymbol{\Xi}}_{jk}(u,v)\}$ is positive definite. Combining these results, we have

$$\Big|\mathbb{E}_U\Big(\tilde{\mathbf{e}}_0^{\mathrm{T}}\Big[\mathbb{E}\big\{\widehat{\boldsymbol{\Xi}}_{jk}(u,v)\big\}\Big]^{-1}\mathbf{L}_3\Big)\Big|\leqslant\|\mathbb{E}\{\widehat{\boldsymbol{\Xi}}_{jk}(u,v)\}\|_{\min}^{-1}\|\mathbb{E}_U(|\mathbf{L}_3|)\|\leqslant cK_2h_{\Sigma,jk}^2,$$

which implies that $|\widetilde{\Sigma}_{jk}(u,v) - \Sigma_{jk}(u,v)| \leqslant cK_2 h^2_{\Sigma,jk}$ for any $(u,v) \in \mathcal{U}^2$. Hence

$$\sup_{(u,v)\in\mathcal{U}^2} \big|\widetilde{\Sigma}_{jk}(u,v) - \Sigma_{jk}(u,v)\big| = O(h^2_{\Sigma,jk}) \text{ and } \|\widetilde{\Sigma}_{jk} - \Sigma_{jk}\| = O(h^2_{\Sigma,jk}). \tag{A.38}$$

For $M > 0$ with the choice of $\delta = (\log p)^{1/2}(\min_{j,k} \nu_{n,T,h,jk})^{-1/2} M \leqslant 1$, it follows from the union bound of probability and (8) in Theorem 3 that

$$\begin{aligned} &\mathbb{P}\Big\{ \frac{\max_{j,k\in[p]} \|\widehat{\Sigma}_{jk} - \widetilde{\Sigma}_{jk}\|_{\mathcal{S}}}{(\log p)^{1/2}(\min_{j,k}\nu_{n,T,h,jk})^{-1/2}} \geqslant M \Big\} \\ &\leqslant \sum_{j=1}^{p}\sum_{k=1}^{p} \mathbb{P}\Big\{ \|\widehat{\Sigma}_{jk} - \widetilde{\Sigma}_{jk}\|_{\mathcal{S}} \geqslant M(\log p)^{1/2}(\min_{j,k}\nu_{n,T,h,jk})^{-1/2} \Big\} \\ &\leqslant \sum_{j=1}^{p}\sum_{k=1}^{p} c\exp\Big(-c\nu_{n,T,h,jk} M^2 \frac{\log p}{\min_{j,k}\nu_{n,T,h,jk}} \Big) \leqslant c\exp\{(2 - cM^2)\log p\}. \end{aligned} \tag{A.39}$$

We can choose a large M such that $2 - cM^2 < 0$, the right-side of (A.39) tends to 0. Hence $\max_{j,k\in[p]} \|\widehat{\Sigma}_{jk} - \widetilde{\Sigma}_{jk}\|_{\mathcal{S}} = O_P\{(\log p)^{1/2}(\min_{j,k}\nu_{n,T,h,jk})^{-1/2}\}$. Combing this with (A.38) yields that

$$\max_{j,k\in[p]} \|\widehat{\Sigma}_{jk} - \Sigma_{jk}\|_{\mathcal{S}} = O_P\left\{ \Big(\frac{\log p}{\min_{j,k}\nu_{n,T,h,jk}} \Big)^{1/2} + \max_{j,k} h^2_{\Sigma,jk} \right\}.$$

which completes the proof of (12).

The rate of convergence in (13) can be proved following a similar procedure. Let $h_{\Sigma,\min} = \min_{j,k} h_{\Sigma,jk}$, we can assume that $h_{\Sigma,\min} \asymp \{\log(p\vee n)/n\}^{\kappa_2}$ where $\kappa_2 \in (0, 1/2]$. Consider $h_{\Sigma,\min}$ with some $\kappa_2^* > 1/2$, the corresponding rate is not faster than that with some $\kappa_2 \in (0, 1/2]$. Specifically, under the sparse design, the rate of $\max_{j,k\in[p]} \sup_{(u,v)\in\mathcal{U}^2} |\widehat{\Sigma}_{jk}(u,v) - \Sigma_{jk}(u,v)|$ is $\{\log(p \vee n)/n\}^{1/2-\kappa_2^*}$, which is slower than $\{\log(p \vee n)/n\}^{1/3}$ with $\kappa_2 = 1/6$. Under the dense design with $\bar{T}_\Sigma\{\log(p \vee n)/n\}^{\kappa_2^*} \to 0$ and $\bar{T}_\Sigma\{\log(p \vee n)/n\} \to 0$, the rate is $\{\log(p \vee n)/n\}^{1/2-\kappa_2^*}\bar{T}_\Sigma^{-1}$, which is slower than $\{\log(p \vee n)/n\}^{1/3}\bar{T}_\Sigma^{-2/3}$ with $\kappa_2 \in (1/6, 1/2]$. Under the dense design with $\bar{T}_\Sigma\{\log(p \vee n)/n\}^{\kappa_2^*} \to 0$ and $\bar{T}_\Sigma\{\log(p \vee n)/n\} \to \widetilde{c}$ or ∞, the rate is $\{\log(p \vee n)/n\}^{1/2-\kappa_2^*}\bar{T}_\Sigma^{-1}$, which is slower than $\{\log(p \vee n)/n\}^{1/2}$ with $\kappa_2 = 1/4$. Under the dense design with $\bar{T}_\Sigma\{\log(p \vee n)/n\}^{\kappa_2^*} \to \widetilde{c}$ or ∞, the rate is $\{\log(p \vee n)/n\}^{1/2}$, which is the same as $\{\log(p \vee n)/n\}^{1/2}$ with $\kappa_2 = 1/4$. Based on the above four cases, if $\kappa_2^* > 1/2$, the corresponding rate is not faster than that with some $\kappa_2 \in (0, 1/2]$ and hence κ_2 that corresponds with the optimal bandwidth under sparse or dense design is in $(0, 1/2]$. For $M > 0$ with the choice of $\delta = \{\log(p \vee n)\}^{1/2}(\min_j \gamma_{n,T,h,j})^{-1/2} M \leqslant 1$, by the union bound of probability and (9) in Theorem 3, we have

$$\begin{aligned}
&\mathbb{P}\Big\{\frac{\max_{j,k\in[p]}\sup_{(u,v)\in\mathcal{U}^2}|\widehat{\Sigma}_{jk}(u,v)-\widetilde{\Sigma}_{jk}(u,v)|}{\log(p\vee n)^{1/2}(\min_{j,k}\nu_{n,T,h,jk})^{-1/2}}\geqslant M\Big\}\\
\leqslant\;& \sum_{j=1}^{p}\sum_{k=1}^{p}\mathbb{P}\Big\{\sup_{(u,v)\in\mathcal{U}^2}|\widehat{\Sigma}_{jk}(u,v)-\widetilde{\Sigma}_{jk}(u,v)|\geqslant M\{\log(p\vee n)\}^{1/2}(\min_{j,k}\nu_{n,T,h,jk})^{-1/2}\Big\}\\
\leqslant\;& \sum_{j=1}^{p}\sum_{k=1}^{p}\frac{cn^{\epsilon_2}\nu_{n,T,h,jk}}{h_{\Sigma,\min}^{6}}\exp\Big\{-c\nu_{n,T,h,jk}M^2\frac{\log(p\vee n)}{\min_{j,k}\nu_{n,T,h,jk}}\Big\}\\
\leqslant\;& c\exp\{(3+\epsilon_2+6\kappa_2-cM^2)\log(p\vee n)\}.
\end{aligned}$$

We can choose a large M such that $3+\epsilon_2+6\kappa_2-cM^2<0$, the right-side of the above inequality tends to 0. Hence

$$\max_{j,k\in[p]}\sup_{(u,v)\in\mathcal{U}^2}|\widehat{\Sigma}_{jk}(u,v)-\widetilde{\Sigma}_{jk}(u,v)|=O_P\{\log(p\vee n)^{1/2}(\min_{j,k}\nu_{n,T,h,jk})^{-1/2}\}.$$

Combing this with (A.38) yields that

$$\max_{j,k\in[p]}\sup_{(u,v)\in\mathcal{U}^2}\big|\widehat{\Sigma}_{jk}(u,v)-\Sigma_{jk}(u,v)\big|=O_P\left[\Big\{\frac{\log(p\vee n)}{\min_{j,k}\nu_{n,T,h,jk}}\Big\}^{1/2}+\max_{j,k}h_{\Sigma,jk}^2\right].$$

which completes the proof of (13).

Appendix B. Verification of the claim in Section 4

For the mean estimator, define the set of r candidate bandwidths $\mathcal{H}_\mu=\{h_\mu^{(1)},\ldots,h_\mu^{(r)}\}$. In our simulations, the bandwidth for each dimension can be chosen from $\mathcal{H}_\mu$ freely, and hence there are r^p possible outcomes. The targeted evaluation metric is $\text{global}_{\text{opt}}=\min_{(m_1,\ldots,m_p)\in[r]^p}\max_{j\in[p]}\text{MISE}(\widehat{\mu}_j,h_\mu^{(m_j)})$. We will show that $\text{global}_{\text{opt}}=\text{MaxMISE}(\mu)$, the right side of which is much easier to calculate as it only takes into account pr cases. On one hand, it is obvious that $\text{global}_{\text{opt}}\leqslant\text{MaxMISE}(\mu)$. On the other hand, for fixed $(j,m_j)\in[p]\times[r]$, $\text{MISE}(\widehat{\mu}_j,h_\mu^{(m_j)})\geqslant\min_{m\in[r]}\text{MISE}(\widehat{\mu}_j,h_\mu^{(m)})$, and hence $\text{global}_{\text{opt}}\geqslant$

$$\min_{(m_1,\ldots,m_p)\in[r]^p}\max_{j\in[p]}\min_{m\in[r]}\text{MISE}(\widehat{\mu}_j,h_\mu^{(m)})=\max_{j\in[p]}\min_{m\in[r]}\text{MISE}(\widehat{\mu}_j,h_\mu^{(m)})=\text{MaxMISE}(\mu).$$

Combining the above results yields $\text{global}_{\text{opt}}=\text{MaxMISE}(\mu)$. The corresponding claim for the covariance estimator can be verified in the same way.

References

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence.* Oxford University Press, 2014.

Jianqing Fan and James S. Marron. Fast implementations of nonparametric curve estimators. *Journal of Computational and Graphical Statistics*, 3:35–56, 1994.

Yingying Fan, Natasha Foutz, Gareth M. James, and Wolfgang Jank. Functional response additive model estimation with online virtual stock markets. *The Annals of Applied Statistics*, 8:2435–2460, 2014.

Yingying Fan, Gareth M. James, and Peter Radchenko. Functional additive regression. *The Annals of Statistics*, 43:2296–2325, 2015.

Qin Fang, Shaojun Guo, and Xinghao Qiao. Adaptive functional thresholding for sparse covariance function estimation in high dimensions. *Journal of the American Statistical Association, in press*, 2023.

Dehan Kong, Kaijie Xue, Fang Yao, and Hao Zhang. Partially functional linear regression in high dimensions. *Biometrika*, 103:147–159, 2016.

Kuang-Yao Lee, Dingjue Ji, Lexin Li, Todd Constable, and Hongyu Zhao. Conditional functional graphical models. *Journal of the American Statistical Association*, 118:257–271, 2023.

Bing Li and Eftychia Solea. A nonparametric graphical model for functional data with application to brain networks based on fMRI. *Journal of the American Statistical Association*, 113:1637–1655, 2018.

Yehua Li and Tailen Hsing. Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics*, 38:3321–3351, 2010.

Ruiyan Luo and Xin Qi. Function-on-function linear regression by signal compression. *Journal of the American Statistical Association*, 112:690–705, 2017.

Xinghao Qiao, Shaojun Guo, and Gareth M. James. Functional graphical models. *Journal of the American Statistical Association*, 114:211–222, 2019.

Xinghao Qiao, Cheng Qian, Gareth M. James, and Shaojun Guo. Doubly functional graphical models in high dimensions. *Biometrika*, 107:415–431, 2020.

Eftychia Solea and Bing Li. Copula Gaussian graphical models for functional data. *Journal of the American Statistical Association*, 117:781–793, 2022.

Katherine Tsai, Boxin Zhao, Sanmi Koyejo, and Mladen Kolar. Latent multimodal functional graphical model estimation. *Journal of the American Statistical Association, in press*, 2023.

Daren Wang, Zifeng Zhao, Yi Yu, and Rebecca Willett. Functional linear regression with mixed predictors. *Journal of Machine Learning Research*, 23:1–82, 2022.

Fang Yao, Hans-Georg Müller, and Jane-Ling Wang. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100:577–590, 2005a.

Fang Yao, Hans-Georg Müller, and Jane-Ling Wang. Functional linear regression analysis for longitudinal data. *The Annals of Statistics*, 33:2873–2903, 2005b.

Javier Zapata, Sang-Yun Oh, and Alexander Petersen. Partial separability and functional graphical models for multivariate gaussian processes. *Biometrika*, 109:665–681, 2022.

Jinting Zhang and Jianwei Chen. Statistical inferences for functional data. *The Annals of Statistics*, 35:1052–1079, 2007.

Xiaoke Zhang and Jane-Ling Wang. From sparse to dense functional data and beyond. *The Annals of Statistics*, 44:2281–2321, 2016.

Boxin Zhao, Y. Samuel Wang, and Mladen Kolar. FuDGE: A method to estimate a functional differential graph in a high-dimensional setting. *Journal of Machine Learning Research*, 23:1–82, 2022.

Hongxiao Zhu, Nate Strawn, and David B. Dunson. Bayesian graphical models for multivariate functional data. *Journal of Machine Learning Research*, 17:1–27, 2016.