Doubly Functional Graphical Models in High Dimensions

BY XINGHAO QIAO, CHENG QIAN

Department of Statistics, London School of Economics, London WC2A 2AE, U.K.

x.qiao@lse.ac.uk   c.qian2@lse.ac.uk

GARETH M. JAMES

*Department of Data Sciences and Operations, University of Southern California, Los Angeles,
California 90089, U.S.A.*

gareth@usc.edu

AND SHAOJUN GUO

Institute of Statistics and Big Data, Renmin University of China, Beijing, 100872, P. R. China

sjguo@ruc.edu.cn

SUMMARY

We consider estimating a functional graphical model from multivariate functional observations. In functional data analysis, the classical assumption is that each function has been measured over a densely sampled grid. However, in practice the functions have often been observed, with measurement error, at a relatively small number of points. In this paper, we propose a class of doubly functional graphical models to capture the evolving conditional dependence relationship among a large number of sparsely or densely sampled functions. Our approach first implements a nonparametric smoother to perform functional principal components analysis for each curve, then estimates a functional covariance matrix and finally computes sparse precision matrices, which in turn provide the doubly functional graphical model. We derive some novel concentration bounds, uniform convergence rates and model selection properties of our estimator for both sparsely and densely sampled functional data in the high-dimensional large p, small n, regime. We demonstrate via simulations that the proposed method significantly outperforms possible competitors. Our proposed method is also applied to a brain imaging dataset.

Some key words: Constrained ℓ_1-minimization; Functional principal component; Functional precision matrix; Graphical model; High-dimensional data; Sparsely sampled functional data.

1. INTRODUCTION

Undirected graphical models depicting conditional dependence relationships among p random variables, $X = (X_1, \ldots, X_p)^{\mathrm{T}}$, have attracted considerable attention in recent years. Let $G = (V, E)$ be an undirected graph characterized by the vertex set $V = \{1, \ldots, p\}$ and the edge set E, which consists of all pairs (j, k) such that X_j and X_k are conditionally dependent given the remaining $p - 2$ variables. A central question in understanding the structure of G is to recover the edge set E. In particular for a multivariate Gaussian distributed X, recovering the structure of an undirected graph is equivalent to locating the non-zero components in the precision matrix, that is, the inverse covariance matrix, of X (Lauritzen, 1996).

Table 1: Graphical models for different types of data and corresponding graph.

		Graphical Model	
		Static	Functional
Data	Static	Gaussian graphical model	Dynamic graphical model
	Functional	Static functional graphical model	Doubly functional graphical model

The past several years have witnessed the development of Gaussian graphical models in large p, small n, settings. One popular class of estimation approaches, the graphical lasso, considers optimizing a criterion involving the Gaussian log-likelihood with a lasso-type penalty on the entries of the precision matrix (Yuan & Lin, 2007; Friedman et al., 2008). For examples of recent developments, see Zhou et al. (2010), Ravikumar et al. (2011), Witten et al. (2011), Chun et al. (2013) and Danaher et al. (2014). Another popular class of neighborhood-based estimation approaches, first proposed by Meinshausen & Buhlmann (2006), considers recovering the support of G by solving p lasso problems columnwise. Cai et al. (2011) proposed a Dantzig-type variant of this approach, named constrained ℓ_1-minimization for inverse matrix estimation. Some recent work along this line of research includes Cai et al. (2016) and Qiu et al. (2016).

In this paper, we consider estimating functional graphical models based on multivariate functional data. Table 1 illustrates the distinction by dividing the data and corresponding network into static vs functional categories. The first entry in the table, Gaussian graphical models, corresponds to the standard setting involving high dimensional, but static, data from which we estimate a single graphical model. One may also observe multiple samples of independent but non-identically distributed static data, where distributions evolve over time, and wish to compute graphical models for each sample. These dynamic graphical models often adopt a nonparametric approach (Zhou et al., 2010; Kolar & Xing, 2011).

Our setting corresponds to the last row of Table 1, where the data can be considered functional. Figure 1 illustrates the data structure and underlying network pattern using a simple example. Its left-hand side plots $n = 100$ realizations of $p = 10$ random curves in $\mathcal{U} = [0, 1]$, each of which corresponds to one underlying node. In practice, functions can be observed at either a dense grid of points or a small subset of possible points, and may also be contaminated by measurement error. Qiao et al. (2019) model such data using a static functional graphical model, where a single network is constructed to encode the global conditional dependence relationship among high-dimensional Gaussian random functions. Li et al. (2018) relax the Gaussian assumption and explore the additive conditional dependence structure by treating p as fixed. Our goal is to present a doubly functional graphical model where both the data and the network are functional in nature. The right-hand side of Figure 1 provides a visualization of our model, where the network edges evolve over $\mathcal{U}$. We aim to estimate the functional network in the right-hand panel based on either sparsely or densely observed functions in the left-hand panel.

Our motivating example is an electroencephalography, EEG, dataset, which measures signals from 64 electrodes placed at standard brain locations over 256 time points for subjects from an alcoholic group and from a non-alcoholic group. When the function at each location is specified over a period of time, existing work has shown that edges will disappear and emerge over time (Cabral et al., 2014). The objective is thus to investigate differences between the alcoholic and control group networks in order to understand how the two populations differ. Other important examples include different types of medical imaging data and gene expression data measured over time (Storey et al., 2005).
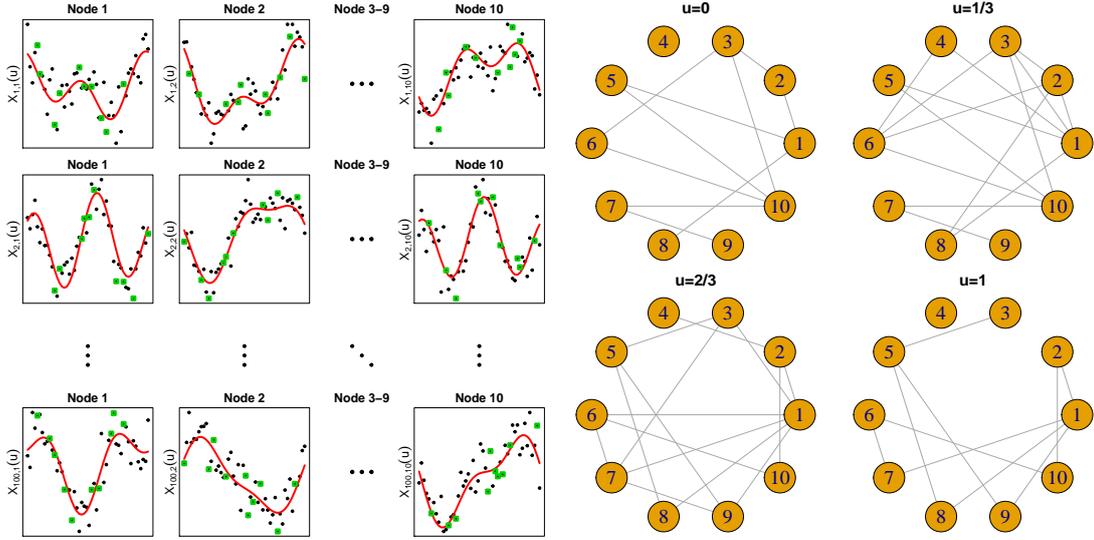
Fig. 1: *Data generated from the simulation setting in Section* 4. *Left: The data matrix consists of* 100 *random functional realizations (red line), their noisy observations at either 50 evenly spaced points (black dots) or 10 randomly selected points (green squares), for* $j = 1, \ldots, 10$ *nodes. Right: Visualization of true functional network at 4 selected time points.*

One approach to address this sort of functional data would be to first sample each function at a grid of points, $u_1, \ldots, u_T$, and then estimate T graphs. This could be achieved by separately estimating T networks using a standard method, for example, the graphical lasso or constrained ℓ_1-minimization, by jointly estimating T graphs that share certain characteristics (Chun et al., 2013; Danaher et al., 2014; Cai et al., 2016), or by estimating the functional graph based on the smoothed sample covariance matrix estimator (Qiu et al., 2016). However these approaches all share one major deficiency, that is they will only work if all random functions are sampled at a common set of grid points, whereas in practice curves are often observed at different sets of points. Another approach is to use nonparametric smoothers to estimate the cross-covariance function between the jth and kth functions for all $j, k = 1, \ldots, p$, and to use these to compute the functional network. However, this would involve computing $p(p + 1)/2$ pairwise terms, which is not computationally scalable, especially under the large p, small n, scenario.

Our proposed method consists of three steps. First, we apply a nonparametric approach to smooth p covariance functions and represent each curve using the first M functional principal components, with the functional principal component scores framed as conditional expectations. Second, the finite-dimensional representations of the curves lead to the functional estimate for the $p \times p$ covariance matrix as it varies over $u \in \mathcal{U}$. Finally, we estimate the functional network by computing the functional sparse precision matrix on a grid of points. This final step can be easily implemented through existing approaches for estimating the sparse precision matrix. Our theoretical results use the constrained ℓ_1-minimization method, because we have found that it provides somewhat superior results in our empirical studies, but other methods, such as the graphical lasso, could easily be applied.

Our approach has six key advantages. First, it is simple to understand and implement, making use of existing statistical software packages. Second, it can handle noisy curves observed at an irregular set of points. Third, it is computationally efficient relative to approaches such as non-

parametric smoothing of $p(p + 1)/2$ cross-covariance functions or jointly estimating T networks, since we only need to smooth p covariance functions and the networks can be computed separately once the functional covariance matrix has been estimated. Fourth, the functional nature of our covariance matrix tends to ensure similar graphical models for neighboring grid points, even though the networks are fit separately. Fifth, the method enjoys desirable consistency properties. Theoretically, we establish some novel concentration bounds and uniform convergence rates of the estimated functional precision matrix in the large p, small n, setting, for both sparsely and densely observed functional data. Finally, empirically we demonstrate the superiority of our proposed method relative to its natural competitors.

2. Methodology

2·1. *Notation*

We begin by introducing some notation. For a vector $a = (a_1, \ldots, a_p)^{\mathrm{T}}$, its ℓ_r norm is $|a|_r = (\sum_i |a_i|^r)^{1/r}$. For a matrix $A = (A_{ij}) \in \mathbb{R}^{p \times q}$, we define the elementwise ℓ_r norm by $|A|_r = (\sum_{i,j} |A_{ij}|^r)^{1/r}$, in particular $r = 2$ corresponds to the Frobenius norm, $||A||_F = |A|_2$. We denote the matrix operator norm by $||A|| = \sup_{|x|_2 \leqslant 1} |Ax|_2$. We use $x \wedge y = \min(x, y)$ and $x \vee y = \max(x, y)$. For a bivariate function $\psi(\cdot, \cdot)$, defined on $\mathcal{U}^2$, we denote the Hilbert–Schmidt norm by $||\psi||_{\mathcal{S}} = \{\iint \psi(u, v)^2 du dv\}^{1/2}$. We write $f(n) = O\{g(n)\}$ if $f(n) \leqslant cg(n)$ for some positive constant $c < \infty$. The notation $f(n) \asymp g(n)$ means that $f(n) = O\{g(n)\}$ and $g(n) = O\{f(n)\}$.

2·2. *Doubly functional graphical models*

Let $X(\cdot) = \{X_1(\cdot), \ldots, X_p(\cdot)\}^{\mathrm{T}}$ denote a p-dimensional vector of Gaussian random functions, with each $X_j(\cdot)$ in $\mathcal{L}_2(\mathcal{U})$, a Hilbert space of square integrable functions on $\mathcal{U}$, a compact subset of the real line. Without loss of generality, we assume that $X(\cdot)$ has been centered to have mean zero. Let $C(u, v) = \{C_{jk}(u, v)\}_{1 \leqslant j, k \leqslant p}$ be the $p \times p$ matrix whose (j, k)th element is $C_{jk}(u, v) = \mathrm{cov}\{X_j(u), X_k(v)\}$ for $(u, v) \in \mathcal{U}^2$.

Let $G(\cdot) = \{V, E(\cdot)\}$ denote a functional undirected graph with a vertex set $V = \{1, \ldots, p\}$ and corresponding functional edge set

$$E(u) = \{(j, k) : \mathrm{cov}[X_j(u), X_k(u) \mid \{X_l(u), l \neq j, k\}] \neq 0, (j, k) \in V^2, j \neq k\}, u \in \mathcal{U}.$$

Standard results show that, for each $u \in \mathcal{U}$, $X(u)$ follows a multivariate Gaussian distribution with covariance matrix $\Sigma(u) = C(u, u) \in \mathbb{R}^{p \times p}$ and $\Theta(u) = \Sigma(u)^{-1}$. Hence, $\mathrm{cov}[X_j(u), X_k(u) | \{X_l(u), l \neq j, k\}] = 0$ if and only if $\Theta_{jk}(u) = 0$ and $E(u)$ can be equivalently represented by

$$E(u) = \{(j, k) : \Theta_{jk}(u) \neq 0, (j, k) \in V^2, j \neq k\}, u \in \mathcal{U}. \tag{1}$$

We use a three-step approach to recover $E(u)$, that is, to identify the locations of the non-zero entries of $\Theta(u)$ in a functional fashion.

Step 1. For each $j \in V$, we adopt a data-driven basis expansion approach through functional principal component analysis. Specifically, the covariance function $C_{jj}(u, v)$ satisfies $\int_{\mathcal{U}} C_{jj}(u, v)\phi_{jl}(v)dv = \omega_{jl}\phi_{jl}(u)$ $(l = 1, 2, \ldots)$, where the eigenpairs $\{\omega_{jl}, \phi_{jl}(\cdot)\}_{l \geqslant 1}$ satisfy $\omega_{j1} \geqslant \omega_{j2} \geqslant \cdots \geqslant 0$ and $\int_{\mathcal{U}} \phi_{jl}(u)\phi_{jl'}(u)du = I(l = l')$ with $I(\cdot)$ being the indicator function. The Karhunen-Loève expansion allows us to expand each $X_j(\cdot)$ as $X_j(\cdot) = \sum_{l=1}^{\infty} \xi_{jl}\phi_{jl}(\cdot)$, where $\xi_{jl} = \int_{\mathcal{U}} X_j(u)\phi_{jl}(u)du \sim N(0, \omega_{jl})$ are the principal component scores, with ξ_{jl} being independent of $\xi_{jl'}$ for $l \neq l'$. Due to the infinite-dimensional nature of functional data, a

standard approach is to approximate $X_j(\cdot)$ using the leading M principal components, that is, $X_{j,M}(\cdot) = \sum_{l=1}^{M} \xi_{jl}\phi_{jl}(\cdot)$, where M is chosen large enough to provide a reasonable approxima- 140 tion to $X_j(\cdot)$. Potentially one could use a separate M_j for each $j \in V$. To simplify our notation we focus on the setting where the M_j's are the same across $j \in V$. However, our theoretical results in Section 3 extend naturally to the more general setting. In our empirical studies, we select different M_j's, see Section 2·4 for details.

Step 2. Once Step 1 has been performed for each $X_j(\cdot)$ the M-dimensional functional rep- 145 resentation leads to a natural approximation for the $p \times p$ functional covariance matrix $\Sigma_M(u)$, with (j, k)th entry given by:

$$\Sigma_{jk,M}(u) = \sum_{l=1}^{M} \sum_{m=1}^{M} \text{cov}(\xi_{jl}, \xi_{km})\phi_{jl}(u)\phi_{km}(u), u \in \mathcal{U}. \tag{2}$$

Step 3. Our final step involves computing a functional sparse precision matrix $\Theta_M(u) = \Sigma_M(u)^{-1}$ at a set of points in $\mathcal{U}$.

Let $X_i(\cdot) = \big\{X_{i1}(\cdot), \ldots, X_{ip}(\cdot)\big\}^{\mathrm{T}}$ $(i = 1, \ldots, n)$ be independent and identically distributed copies of $X(\cdot)$. We assume that $X_{ij}(\cdot)$ is observed, with measurement error, at random time points, $U_{ijt} \in \mathcal{U}$ for $t = 1, \ldots, T_{ij}$, where for dense measurement schedules all T_{ij} are larger than some order of n, and for sparse designs all T_{ij} are bounded. Let Y_{ijt} represent the observed value of $X_{ij}(U_{ijt})$. Then 155

$$Y_{ijt} = X_{ij}(U_{ijt}) + e_{ijt} = \sum_{l=1}^{\infty} \xi_{ijl}\phi_{jl}(U_{ijt}) + e_{ijt}, \tag{3}$$

where the e_{ijt}'s are independent and identically distributed with $E(e_{ijt}) = 0$ and $\text{var}(e_{ijt}) = \sigma^2$, independent of $X_{ij}(\cdot)$. We provide estimation details to implement our three-step approach from Section 2·2 as follows.

Step 1. To perform functional principal component analysis based on realizations $Y_{ij} = (Y_{ij1}, \ldots, Y_{ijT_{ij}})^{\mathrm{T}}$ $(i = 1, \ldots, n)$ for each $j \in V$, we first compute the estimator for $C_{jj}(u, v)$. 160 Let $\Sigma_{Y_{ij}}$ be the covariance matrix for Y_{ij} with (t, t')th element, $\big(\Sigma_{Y_{ij}}\big)_{tt'} = \text{cov}(Y_{ijt}, Y_{ijt'}) = C_{jj}(U_{ijt}, U_{ijt'}) + \sigma^2 I(t = t')$. A local linear surface smoother is applied to the off-diagonals of the raw covariances, $\{Y_{ijt}Y_{ijt'}\}_{1 \leqslant t \neq t' \leqslant T_{ij}}$. Denote $K_h(\cdot) = h^{-1}K(\cdot/h)$ for a univariate kernel function K with a positive bandwidth h. We consider minimizing

$$\sum_{i=1}^{n} w_{ij} \sum_{1 \leqslant t \neq t' \leqslant T_{ij}} \Big\{Y_{ijt}Y_{ijt'} - \beta_0 - \beta_1(U_{ijt} - u) - \beta_2(U_{ijt'} - v)\Big\}^2 K_{h_j}(U_{ijt} - u)K_{h_j}(U_{ijt'} - v),$$
$$\tag{4}$$

with respect to $(\beta_0, \beta_1, \beta_2)$, where the weight w_{ij} is chosen for the ith subject and jth vari- 165 able such that $\sum_{i=1}^{n} T_{ij}(T_{ij} - 1)w_{ij} = 1$. For details on the choices of w_{ij} under different measurement schedules, we refer to Zhang & Wang (2016). The resulting covariance estimator is obtained as $\widehat{C}_{jj}(u, v) = \widehat{\beta}_0$.

We next perform eigen-decomposition on $\widehat{C}_{jj}(u, v)$ and obtain the estimated eigenpairs $\{\widehat{\omega}_{jl}, \widehat{\phi}_{jl}(\cdot)\}_{1 \leqslant l \leqslant M}$. A natural estimate for the principal component score, ξ_{ijl}, is $\int_{\mathcal{U}} \widehat{X}_{ij}(u)\widehat{\phi}_{jl}(u)du$, which, for very dense data, can be well approximated by numerical integration based on observations $\{U_{ijt}, Y_{ijt}, \widehat{\phi}_{jl}(U_{ijt})\}_{1 \leqslant t \leqslant T_{ij}}$. However, this numerical integration approach fails in settings with sparse designs or dense designs with missing values. Instead, we

propose to use the best linear unbiased predictors $\widetilde{\xi}_{ijl} = \zeta_{ijl}^{\mathrm{T}} \Sigma_{Y_{ij}}^{-1} Y_{ij}$ (Rice & Wu, 2001), where ζ_{ijl} is a T_{ij}-dimensional vector with tth component

$$\zeta_{ijlt} = \mathrm{cov}(\xi_{ijl}, Y_{ijt}) = E\left\{\int X_{ij}(v)\phi_{jl}(v)dv X_{ij}(U_{ijt})\right\} = \int C_{jj}(U_{ijt}, v)\phi_{jl}(v)dv.$$

Although we do not place any distributional assumptions on the errors here, when e_{ijt} and ξ_{ijl} are jointly Gaussian, $\widetilde{\xi}_{ijl}$ reduces to the conditional expectation of ξ_{ijl} given Y_{ij} (Yao et al., 2005). We then obtain the estimator for $\widetilde{\xi}_{ijl}$ as

$$\widehat{\xi}_{ijl} = \widehat{\zeta}_{ijl}^{\mathrm{T}} \widehat{\Sigma}_{Y_{ij}}^{-1} Y_{ij}, \tag{5}$$

where $\widehat{\zeta}_{ijlt} = \int \widehat{C}_{jj}(U_{ijt}, v)\widehat{\phi}_{jl}(v)dv$ and $\left(\widehat{\Sigma}_{Y_{ij}}\right)_{tt'} = \widehat{C}_{jj}(U_{ijt}, U_{ijt'}) + \widehat{\sigma}^2 I(t = t')$. See Yao et al. (2005) for details on the estimate $\widehat{\sigma}^2$ of σ^2.

Step 2. Once the functional principal components analysis has been performed, we substitute the terms in (2) by their estimated values and thus obtain $\widehat{\Sigma}(u)$ with its (j, k)th entry given by $\widehat{\Sigma}_{jk}(u) = n^{-1} \sum_{i=1}^{n} \sum_{l=1}^{M} \sum_{m=1}^{M} \widehat{\xi}_{ijl}\widehat{\xi}_{ikm}\widehat{\phi}_{jl}(u)\widehat{\phi}_{km}(u)$.

Step 3. Finally, for a set of points $u \in \mathcal{U}$, we estimate $\Theta_{jk}(u)$. One of the advantages of our approach is that a variety of standard sparse precision matrix methods can be used to implement this step. Our empirical results suggest that the constrained ℓ_1-minimization (Cai et al., 2011) provides the most accurate results so we use that approach here. To be specific, we solve the following constrained optimization problem

$$\widecheck{\Theta}(u) = \underset{\Theta \in \mathbb{R}^{p \times p}}{\arg\min} |\Theta|_1 \text{ subject to } |\widehat{\Sigma}(u)\Theta - I|_\infty \leqslant \lambda_n(u), \tag{6}$$

where $I \in \mathbb{R}^{p \times p}$ is the identity matrix and $\lambda_n(u) \geqslant 0$ is a tuning parameter which controls the sparsity level of $\widecheck{\Theta}(u)$. The convex problem (6) can be further decomposed into p separate optimization problems. For each $j \in V$, we solve

$$\widehat{\beta}_j(u) = \underset{\beta \in \mathbb{R}^p}{\arg\min} |\beta|_1 \text{ subject to } |\widehat{\Sigma}(u)\beta - e_j|_\infty \leqslant \lambda_n(u), \tag{7}$$

where $e_j \in \mathbb{R}^p$ is the unit vector with jth coordinate being 1 and $\widehat{\beta}_j(u)$ corresponds to the jth column of $\widecheck{\Theta}(u)$.

Our target estimator $\widehat{\Theta}(u)$ is attained by the final step of symmetrizing $\widecheck{\Theta}(u)$ whose (j, k) and (k, j)-th entries are obtained by

$$\widehat{\Theta}_{jk}(u) = \widehat{\Theta}_{kj}(u) = \widecheck{\Theta}_{jk}(u)I\{|\widecheck{\Theta}_{jk}(u)| \leqslant |\widecheck{\Theta}_{kj}(u)|\} + \widecheck{\Theta}_{kj}(u)I\{|\widecheck{\Theta}_{jk}(u)| > |\widecheck{\Theta}_{kj}(u)|\}. \tag{8}$$

This symmetrization procedure guarantees that $\widehat{\Theta}(u)$ achieves the same elementwise ℓ_∞ estimation error rate as $\widecheck{\Theta}(u)$. We obtain the final estimated functional edge set as

$$\widehat{E}(u) = \left\{(j, k) : |\widehat{\Theta}_{jk}(u)| > \tau_n(u), (j, k) \in V^2, j \neq k\right\}, u \in \mathcal{U}, \tag{9}$$

where $\tau_n(u) > 0$ is a threshold parameter. Empirical results suggest that $\tau_n(u)$ can be set to zero or a very small value, so we include this term merely for establishing the graph support recovery consistency in Section 3.

2·4.   *Selection of tuning parameters*

To fit our proposed method, we must choose optimal values for the tuning parameters, h_j, M_j ($j = 1, \ldots, p$) and $\lambda_n(u)$.

We adopt leave-one-curve-out cross validation (Rice & Silverman, 1991) to select optimal values for h_j in (4). See Zhang & Wang (2016) for a discussion of two advantages of using this method. Typical approaches to choose the M_j's include leave-one-curve-out cross validation and the Akaike Information Criterion (Yao et al., 2005). We take the later approach since it is computationally less intensive while numerical performance is similar to that obtained from cross-validation.

Popular approaches, such as cross-validation and the information criterion, for the selection of $\lambda_n(u)$ have been broadly studied in the static graphical models literature (Yuan & Lin, 2007; Cai et al., 2011). We adopt the more computationally efficient Bayesian Information Criterion approach, which chooses an optimal $\lambda_n(u)$ by minimizing

$$\mathrm{BIC}\{\lambda_n(u)\} = n\mathrm{tr}\{\widehat{\Theta}_{\lambda_n(u)}(u)\widehat{\Sigma}(u)\} - n\log\det\{\widehat{\Theta}_{\lambda_n(u)}(u)\} + \log(n)|\widehat{E}(u)|, \qquad (10)$$

over a series of $\lambda_n(u)$ values, where $\widehat{\Theta}_{\lambda_n(u)}(u)$ is the regularized estimator corresponding to $\lambda_n(u)$ and $|\widehat{E}(u)|$ is the number of non-zero components in $\widehat{\Theta}_{\lambda_n(u)}(u)$. It is worth noting that $\widehat{\Sigma}(u)$ is obtained in Step 2 of the estimation, so is a fixed quantity in terms of (10). Hence, the effective sample size in the BIC is n, which is independent of the T_{ij}'s.

2·5. *Relationship to relevant work*

We compare the doubly functional graphical model with the static functional graphical model of Qiao et al. (2019), To illustrate the difference, we consider a simplified setting, where, for each $j \in V$, $X_j(\cdot) = \xi_j^{\mathrm{T}}\phi_j(\cdot)$ belongs to an M-dimensional Gaussian process. The static functional graphical model generates a single network by recovering the block sparsity pattern in $\Omega^{-1} \in \mathbb{R}^{Mp \times Mp}$ whose (j, k)th block is $\Omega_{jk} = \mathrm{cov}(\xi_j, \xi_k)$. By contrast, the doubly functional graphical model constructs a separate network for each value of u by estimating the sparsity structure in $\Theta(u) = \{\Phi(u)^{\mathrm{T}}\Omega\Phi(u)\}^{-1}$, where $\Phi(u) \in \mathbb{R}^{Mp \times p}$ is block-diagonal with jth block given by $\phi_j(u) \in \mathbb{R}^{M \times 1}$. Each approach has different pros and cons. The static functional graphical model provides a single global network, an advantage which aids interpretation. However, the network will exhibit an edge if two functions are conditionally related at even very distinct time points, so may end up with an overly dense set of edges. By comparison, the doubly functional graphical model provides a cross-sectional view of the graphical model which has the potential to illustrate structural changes in the network as a function of u, a detail that the static model may miss.

Two other papers with similarities to our approach are Zhou et al. (2010) and Qiu et al. (2016), which both fit dynamic graphical models. As with our work, the data in these papers consists of $X_i(u_t) = (X_1(u_t), \dots, X_p(u_t))^{\mathrm{T}}$ $(i = 1, \dots, n; t = 1, \dots, T)$ and both approaches fit a separate graphical model at a given set of values for $u \in \mathcal{U}$. However, Zhou et al. (2010) assumes only one observation at each u_t, that is, $n = 1$, and models the $X_i(u_t)$'s as independent over u_t, so their data structure is a special case of that in our work and Qiu et al. (2016). Alternatively, Qiu et al. (2016) models $\{X_i(u_t)\}_{i=1}^n$ as following a lag-one stationary vector autoregressive model, that is $X_i(u_t)$ is correlated with $X_{i-1}(u_t)$. By contrast, we treat $\{X_i(u)\}_{i=1}^n$ as independent realizations of an underlying multivariate Gaussian process, with each $X_{ij}(u)$ observed, with error, at an irregular set of points, as described in (3). All three methods generate graphical models at a specified set of values for u, but are designed to tackle rather different situations. In addition, as mentioned previously, both Zhou et al. (2010) and Qiu et al. (2016) require that the data be sampled on a common grid of values for u so can not be implemented in the more realistic setting we consider where functions are observed at different points.

3. THEORY

In this section, we investigate the theoretical properties of our proposed approach for both the sparse and dense measurement schedules. We begin by introducing parameter spaces of functional approximately sparse precision matrices

$$\mathcal{C}\{q, s_0(p), K; \mathcal{U}\} = \Big\{\{\Theta(u), u \in \mathcal{U}\} \,\Big|\, \sup_{u\in\mathcal{U}}||\Theta(u)||_1 < K, \sup_{u\in\mathcal{U}}\max_{j\in V} \sum_{k=1}^{p} |\Theta_{jk}(u)|^q \leqslant s_0(p)\Big\}, \tag{11}$$

for $0 \leqslant q < 1$. In the special case of $q = 0$, then $\mathcal{C}(0, s_0(p), K; \mathcal{U})$ corresponds to the functional truly sparse situation, where even the densest $\Theta(u)$ over $u \in \mathcal{U}$ has at most $s_0(p)$ non-zero entries on each row. Similar classes were used in estimating static covariance models (Bickel & Levina, 2008) and its generalization to the dynamic setting (Chen & Leng, 2016). We extend the class of static approximately sparse precision matrices (Cai et al., 2011) to the functional version via (11), uniformly over which Theorems 1-2 hold.

To present the main theorems, we need the following regularity conditions. We relegate some standard conditions to the Supplementary Material.

Condition 1. In the sparse measurement design $T_{ij} \leqslant T_0 < \infty$ for all $i = 1, \ldots, n, j \in V$, and in the dense design $T_{ij} = T \to \infty$ with the U_{ijt}'s independent of i.

Condition 2. For each $j \in V$, there exist some positive constants c_1, c_2, c_3 and $\gamma \leqslant 1/2$ such that for any $0 < \delta \leqslant 1$,

$$\mathrm{pr}\Big(\big\|\widehat{C}_{jj} - \widetilde{C}_{jj}\big\|_{\mathcal{S}} \geqslant \delta\Big) \leqslant c_2 \exp(-c_1 n^{2\gamma}\delta^2), \tag{12}$$

$$\mathrm{pr}\Big\{\sup_{(u,v)\in\mathcal{U}^2}\big|\widehat{C}_{jj}(u,v) - \widetilde{C}_{jj}(u,v)\big| \geqslant \delta\Big\} \leqslant c_2 n^{c_3} \exp(-c_1 n^{2\gamma}\delta^2), \tag{13}$$

where $\widetilde{C}_{jj}(u,v)$ is a deterministic covariance function which converges to $C_{jj}(u,v)$ as $h_j = h \to 0$. See (B.9) in the Supplementary Material for the exact form of $\widetilde{C}_{jj}(u,v)$.

To simplify notation, we assume $T_{ij} = T$ for the dense case in Condition 1 and $h_j = h$ in Condition 2. We also assume in Condition 2, for each $j \in V$, a single value of γ, which depends on h and possibly T for the dense design. Condition 2 is satisfied by a large class of measurement designs with larger values of γ corresponding to a more frequent measurement schedule. For sparsely sampled functional data, we have proved in Lemma 4 in the Supplementary Material that (12)–(13) hold by choosing $\gamma = 1/2 - a$ and $c_3 = 1 + 2a$ with $h \asymp n^{-a}$ for some positive constant $a < 1/2$. Lemma 4 also results in L_2 and uniform convergence rates of $n^{-1/2}h^{-1}$ and $(\log n)^{1/2}n^{-1/2}h^{-1}$, respectively for $\widehat{C}_{jj}(u,v)$ to $\widetilde{C}_{jj}(u,v)$, which are consistent with results for the sparse case in Zhang & Wang (2016). Under dense measurement designs, we have proved in Lemma 5 in the Supplementary Material that Condition 2 holds with $\gamma = 1/2 \wedge (1/3 - \epsilon/2 + b/6 - 2a/3) > 0$ for fixed small constant $\epsilon > 0$ as long as $h \asymp n^{-a}$ and $T \asymp n^b$ for some positive constants a, b. Provided that T grows fast enough, the resulting L_2 and uniform convergence rates become $n^{-1/2}$ and $(\log n)^{1/2}n^{-1/2}$, respectively, belonging to the ultra-dense class in Zhang & Wang (2016). We also have proved that, for fully observed functional data with $\widetilde{C}_{jj}(u,v) = C_{jj}(u,v)$, Condition 2 holds with $\gamma = 1/2$ and $c_3 = 1$. See Lemmas 1–2 in the Supplementary Material for details.

Condition 3. (i) The truncated dimension of the functional data, M, satisfies $M \asymp n^\alpha$ for some constant $\alpha > 0$; (ii) The principal component functions are continuous on the compact

set $\mathcal{U}$ and satisfy $\max_{j \in V} \sup_{u \in \mathcal{U}} \sup_{l \geq 1} |\phi_{jl}(u)| = O(1)$; (iii) The eigenvalues satisfy $\omega_{j1} > \omega_{j2} > \cdots > \omega_{jM} > \omega_{j(M+1)} \geq \cdots$ and there exists some constant $\beta > 2$ with $\alpha(2\beta + 1) < 1/2$, such that, for each $l = 1, \ldots, M$, $\omega_{jl} \asymp l^{-\beta}$, $d_{jl}\omega_{jl} = O(l)$ uniformly in $j \in V$, where $d_{jl} = \max\left\{(\omega_{j(l-1)} - \omega_{jl})^{-1}, (\omega_{jl} - \omega_{j(l+1)})^{-1}\right\}$ if $l \geq 2$ and $d_{j1} = (\omega_{j1} - \omega_{j2})^{-1}$; (iv) There exists some constant $\nu > 0$ such that $\max_{j \in V} \sum_{l=M+1}^{\infty} \omega_{jl} \leq O(M^{-\nu})$.

The parameter α in Condition 3 (i) determines the number of leading principal components used to approximate the infinite dimensional process, with larger values providing better approximations. Condition 3 (iii) provides decay rates for the strictly decreasing sequence of $\omega_{j1}, \ldots, \omega_{jM}$ and gaps between adjacent eigenvalues, d_{jl}^{-1}'s, both of which are used to derive the convergence rates of estimated eigenpairs (Qiao et al., 2019). Condition 3 (iv) guarantees that, for each $j \in V$, $C_{jj}(u, v)$ has finite trace, with ν controlling the rate at which the approximation error decreases with M.

Now we are ready to establish the uniform convergence rate and graph recovery consistency of the proposed estimator as stated in Theorems 1 and 2, respectively. Define $\widetilde{\Sigma}(u)$ with its (j, k)th entry given by $\widetilde{\Sigma}_{jk}(u) = \sum_{l=1}^{\infty} \sum_{m=1}^{\infty} \mathrm{cov}(\widetilde{\xi}_{ijl}, \widetilde{\xi}_{ikm})\phi_{jl}(u)\phi_{km}(u)$. We can prove that $\widehat{\Sigma}(u)$ is a consistent estimator for $\widetilde{\Sigma}(u)$, but $\widetilde{\Sigma}(u)$ fails to converge to $\Sigma(u)$ unless T diverges to ∞. Therefore, for the sparse design, we denote the population functional precision matrix by $\widetilde{\Theta}(u) = \widetilde{\Sigma}(u)^{-1}$ and the corresponding edge set by $\widetilde{E}(u) = \left\{(j, k) : \widetilde{\Theta}_{jk}(u) \neq 0, (j, k) \in V^2, j \neq k\right\}$, both of which are conditional on the random locations $\{U_{ijt} : i = 1, \ldots, n, j \in V, t = 1, \ldots, T_{ij}\}$. For the dense design, we use $\Theta(u)$ and $E(u)$ as the true functional precision matrix and edge set, respectively.

THEOREM 1. *Suppose that Conditions 1–3 and 5–6 in the Supplementary Material hold.*

(i) For the sparse design, suppose that $\{\widetilde{\Theta}(u), u \in \mathcal{U}\}$ belongs to $\mathcal{C}(q, s_0(p), K; \mathcal{U})$. If $\lambda_n(u) = cK'(u)\{(\log p/n^{2\gamma - 3\alpha\beta - 4\alpha})^{1/2} + (\log p/n^{\alpha\nu})^{1/2}\}$ with c sufficiently large, $K'(u)$ satisfying $\sup_{u \in \mathcal{U}} K'(u) \leq K$, $\log p/n^{2\gamma - 3\alpha\beta - 4\alpha} \to 0$, $\log p/n^{\alpha\nu} \to 0$ and $h^2 n^{\gamma} \to 0$, then we have

$$\sup_{u \in \mathcal{U}} \left\| \widehat{\Theta}(u) - \widetilde{\Theta}(u) \right\| = O_p \left[K^{2(1-q)} s_0(p) \left\{ \left(\frac{\log p}{n^{2\gamma - 3\alpha\beta - 4\alpha}} \right)^{1/2} + \left(\frac{\log p}{n^{\alpha\nu}} \right)^{1/2} \right\}^{1-q} \right]. \quad (14)$$

(ii) For the dense design, suppose that $\{\Theta(u), u \in \mathcal{U}\}$ belongs to $\mathcal{C}(q, s_0(p), K; \mathcal{U})$ and let $\kappa_{n,T} \asymp n^{\gamma - \alpha(3\beta/2 + 2)} \wedge T^{-3}n^{\gamma} \wedge T^{1/2}n^{-\alpha}$. Furthermore, if the (X_{ij}, e_{ij})'s are jointly Gaussian and $\lambda_n(u) = cK'(u)\{(\log p/\kappa_{n,T}^2)^{1/2} + (\log p/n^{\alpha\nu})^{1/2}\}$ with c sufficiently large, $K'(u)$ satisfying $\sup_{u \in \mathcal{U}} K'(u) \leq K$, $\log p/\kappa_{n,T}^2 \to 0$, $\log p/n^{\alpha\nu} \to 0$ and $h^2 n^{\gamma} \to 0$, then we have

$$\sup_{u \in \mathcal{U}} \left\| \widehat{\Theta}(u) - \Theta(u) \right\| = O_p \left[K^{2(1-q)} s_0(p) \left\{ \left(\frac{\log p}{\kappa_{n,T}^2} \right)^{1/2} + \left(\frac{\log p}{n^{\alpha\nu}} \right)^{1/2} \right\}^{1-q} \right]. \quad (15)$$

We observe that the uniform convergence rates in (14) and (15) are governed by two sets of parameters: (i) dimensionality parameters: n, p and T (ii) internal parameters: $\alpha, \beta, \gamma, \nu, q, K$ and $s_0(p)$. When T is bounded, the rate in (15) reduces to that in (14). We provide two remarks for the sparse case. First, the rate in (14) consists of two terms, which reflect our familiar bias-variance tradeoff as commonly considered in the nonparametric setting. Under the functional graphical model setting with $q = 0$, the bias term is bounded by $O\{K^2 s_0(p)(\log p/n^{\alpha\nu})^{1/2}\}$ and the variance is of the order $O_p\{K^2 s_0(p)(\log p/n^{2\gamma - 3\alpha\beta - 4\alpha})^{1/2}\}$. It is easy to see that larger values of α or β or smaller values of γ yield a larger variance, while enlarging α or ν results in a smaller bias. To balance both terms, we choose $\alpha = 2\gamma/(3\beta + \nu + 4)$,

which provides a truncated dimension of $M \asymp n^{2\gamma/(3\beta+\nu+4)}$ and the optimal rate in (14) becomes $O_p\big\{K^{2(1-q)}s_0(p)\big(\log p/n^{2\gamma\nu/(3\beta+\nu+4)}\big)^{(1-q)/2}\big\}$. When ν diverges to infinity with $n^{-\gamma} = n^{-1/2}h^{-1}$ implied from Lemma 4, M approaches a fixed dimension and the optimal rate goes to $O_p\big[K^{2(1-q)}s_0(p)\{\log p/(nh^2)\}^{(1-q)/2}\big]$. In functional data analysis, one often only needs to consider the first several principal components as it is usually the case that the truncation errors decay to zero fast, so assuming a very small α and a large ν is generally appropriate. Second, we can relax Conditions 2 and 3 by allowing parameters $\gamma, \alpha, \beta, \nu$ to depend on $j \in V$. Then, for example, the variance term in (14) will be determined by the least smooth component with the tightest eigen-gaps, that is the smallest h_j and the largest M_j, β_j.

We provide two remarks for the dense case. First, in comparison with (14), the variance in (15) is additionally determined by the second and third terms of $\kappa_{n,T}$, which are due to the convergence of $n^{-1}\sum_{i=1}^n \widehat{\xi}_{ijl}\widehat{\xi}_{ikm}$ and $n^{-1}\sum_{i=1}^n \widetilde{\xi}_{ijl}\widetilde{\xi}_{ikm}$ ($j, k \in V, l, m = 1, \ldots, M$), respectively. See Lemmas 14–15 in the Supplementary Material for details. Second, to discuss how the convergence rate depends on T, we focus on a practical scenario where $q = 0$ and M approaches a fixed dimension. (i) When T grows at a relatively slow speed with $T = o(n^{2\gamma/7})$, then $\kappa_{n,T} \asymp T^{1/2}$ so the variance becomes $O_p\big\{K^2 s_0(p)(\log p/T)^{1/2}\big\}$; (ii) When T grows moderately fast with $T^{-1} = o(n^{-2\gamma/7})$, then $\kappa_{n,T} \asymp T^{-3}n^\gamma$ so the variance is of the order $O_p\big\{K^2 s_0(p)(T^6 \log p/n^{2\gamma})^{1/2}\big\}$. In this sense, our proposed method can not effectively handle very dense measurement schedules. This is because (5) requires multiplying a T-dimensional vector by the inverse of a T by T matrix, which provides a poor estimate of the conditional expectation, $\widetilde{\xi}_{ijl}$, when T is large. However, the dense setting is actually relatively easy to handle because in this setting we can calculate a direct estimate of ξ_{ijl} via approximate numerical integration, resulting in improved convergence rates and empirical performance. In all other respects our basic methodology follows through. See Section C in the Supplementary Material for further details.

Next we introduce Condition 4, which is crucial to develop the graph selection consistency result in Theorem 2.

Condition 4. (i) For the sparse design, let $\widetilde{S}(u) = \big\{\widetilde{E}(u) \cup (1,1), \ldots, (p,p)\big\}$ be the augmented set for $u \in \mathcal{U}$, then $\min_{(j,k)\in\widetilde{S}(u)}|\widetilde{\Theta}_{jk}(u)| > 2\tau_n(u)$. (ii) For the dense design, let $S(u) = \big\{E(u) \cup (1,1), \ldots, (p,p)\big\}$ for $u \in \mathcal{U}$, then $\min_{(j,k)\in S(u)}|\Theta_{jk}(u)| > 2\tau_n(u)$.

Condition 4 requires the minimum signal strength on the augmented set be large enough to ensure that non-zero components are correctly retained. We can understand this condition as bounding the minimum strength of the strong signal when $\Theta(u)$ varies smoothly. See Chen & Leng (2016) and Qiu et al. (2016) for analogous functional minimum signal strength conditions.

THEOREM 2. *Suppose that Conditions* 1–6 *hold.*
(i) For the sparse design, if it is further assumed that $\{\widetilde{\Theta}(u), u \in \mathcal{U}\}$ *belongs to* $\mathcal{C}(0, s_0(p), K; \mathcal{U})$ *and* $\tau_n(u) = 4K'(u)\lambda_n(u)$, *then the event* $\big\{\widehat{E}(u) = \widetilde{E}(u)\big\}$ *holds with probability tending to 1 uniformly for* $u \in \mathcal{U}$;
(ii) For the dense design, if it is further assumed that $\{\Theta(u), u \in \mathcal{U}\}$ *belongs to* $\mathcal{C}(0, s_0(p), K; \mathcal{U})$ *and* $\tau_n(u) = 4K'(u)\lambda_n(u)$, *then the event* $\big\{\widehat{E}(u) = E(u)\big\}$ *holds with probability tending to 1 uniformly for* $u \in \mathcal{U}$.

4. SIMULATIONS

4·1. *Setup*

We assess the finite-sample performance of the doubly functional graphical model, using both the constrained ℓ_1-minimization and graphical lasso approaches. Sections 4·2 and 4·3 respectively consider scenarios where functions are observed at common, or irregular, sets of points.

In each simulated scenario, we generate functional variables using $X_{ij}(\cdot) = s(\cdot)^{\mathrm{T}}\delta_{ij}$ ($j = 1, \ldots, p$), where $s(\cdot)$ is a 5-dimensional Fourier basis function and $\delta_i = (\delta_{i1}^{\mathrm{T}}, \ldots, \delta_{ip}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{5p}$ follows a multivariate Gaussian distribution with mean zero and covariance matrix $\Omega \in \mathbb{R}^{5p \times 5p}$, whose (j, k)th block is Ω_{jk} ($j, k = 1, \ldots, p$). The observed values, Y_{ijt}, are then generated, with error, from

$$Y_{ijt} = s(u_{ijt})^{\mathrm{T}}\delta_{ij} + \varepsilon_{ijt} \quad (i = 1, \ldots, 200;\ j = 1, \ldots, p;\ t = 1, \ldots, T_{ij}), \qquad (16)$$

where $p = 50$ or 100, and the u_{ijt}'s and ε_{ijt}'s are randomly sampled from Uniform$[0, 1]$ and $N(0, 0.5)$, respectively.

Since functional conditional dependence relationships are fully characterized by the corresponding functional sparsity pattern in $\Theta(u) = \Sigma(u)^{-1}$ we consider a setting, generalized from Zhou et al. (2010), to simulate a $\Theta(u)$ corresponding to a slow graph evolution over $[0, 1]$. When $u = 0$, the initial diagonal elements in $\Theta(0)$ are set to 0.25. For $p = 50$, we randomly select $n_{\text{initial}} = 40$ out of $50 \times (50 - 1)/2$ potential edges, with edge strengths generated from Uniform$[-0.3, -0.1]$. To create dynamic graphs, we choose $u = 0, 0.1, \ldots, 0.9$ as change points and at each point randomly choose $n_{\text{grow}} = n_{\text{decay}} = 10$ edges, which will simultaneously appear and vanish, respectively, over $[u, u + 0.5)$. For n_{grow} edges, we set the strengths to be 0 at u and the underlying components grow linearly to values generated from Uniform$[-0.3, -0.1]$ at $u + 0.5$. Analogously, among the non-zero entries at u, each decaying edge linearly decays to 0 in $[u, u + 0.5)$. Over the evolution where edges emerge and disappear, when we subtract a value from $\Theta_{jk}(u)$ and $\Theta_{kj}(u)$ for $j \neq k$, we can always add the same value on $\Theta_{jj}(u)$ and $\Theta_{kk}(u)$ to guarantee positive definiteness of $\Theta(u)$. For $p = 100$, we set $n_{\text{initial}} = 160$, $n_{\text{grow}} = n_{\text{decay}} = 40$ and functional precision matrices are generated in the same manor.

We develop an approach using ideas from linear models of coregionalization (Genton & Kleiber, 2015) to generate our data from $\Sigma(u) = \Theta(u)^{-1}$. See the appendix for details.

4·2. *Common set of time points*

When curves are measured at a common set of points $u_1, \ldots, u_T$, that is, $u_{ijt} = u_t$ with $T_{ij} = T$ for all i, j, t in (16), we compare two versions of our method by computing the precision matrix using either constrained ℓ_1-minimization or the graphical lasso with three other types of competitors. The first type, dynamic graphical models, is based on applying the constrained ℓ_1-minimization or the graphical lasso on the smoothed estimate of the sample covariance matrix $S(u_t)$ of $(Y_{i1t}, \ldots, Y_{ipt})^{\mathrm{T}}$ ($i = 1, \ldots, n$), that is, $S_h(u) = \left\{ \sum_{t=1}^{T} K_h(u_t - u)S(u_t) \right\} \left\{ \sum_{t=1}^{T} K_h(u_t - u) \right\}^{-1}, u \in [0, 1]$. We use a Gaussian kernel with the optimal bandwidth proportional to $\{\log p/(nT)\}^{1/3} \vee T^{-4/5}$ (Qiu et al., 2016), so for the empirical work in this paper we choose the proportionality constant in the range $(0, 3]$, which gives good results in all the settings we considered. The second joint type of method, can simultaneously estimate T precision matrices that share similar sparsity patterns or edge values. The group graphical lasso (Danaher et al., 2014) is implemented in our numerical comparison. We also attempted to fit the fused graphical lasso (Danaher et al., 2014) and joint constrained ℓ_1-minimization (Cai et al., 2016). However, neither approach is scalable especially when T is large, so we do not report their results here. The third type is the naive approach which simply applies the constrained ℓ_1

Table 2: Average (standard error) means of operator, Frobenius losses and areas under the receiver operating characteristic curves (AUROC) at $u_1, \ldots, u_T$ over 100 simulation runs. All entries have been multiplied by 10 for formatting reasons.

T	Method	Operator norm		Frobenius norm		AUROC	
		$p = 50$	$p = 100$	$p = 50$	$p = 100$	$p = 50$	$p = 100$
	DFGM-C	14.1(0.02)	17.6(0.02)	65.9(0.09)	107.5(0.11)	8.2(0.02)	7.0(0.01)
	DFGM-G	14.2(0.03)	18.0(0.01)	68.1(0.09)	114.6(0.11)	8.1(0.02)	6.9(0.02)
	DGM-C	18.2(0.03)	21.9(0.03)	86.7(0.14)	141.0(0.21)	8.1(0.02)	6.5(0.01)
10	DGM-G	18.4(0.03)	22.7(0.02)	90.9(0.17)	153.6(0.26)	8.1(0.02)	6.4(0.02)
	GGL	19.0(0.14)	28.3(0.27)	95.6(0.14)	148.7(0.27)	7.9(0.03)	6.3(0.01)
	Naive-C	18.9(0.04)	35.0(0.05)	88.5(0.10)	187.9(0.77)	7.9(0.02)	6.0(0.01)
	Naive-G	19.1(0.08)	34.8(0.08)	95.6(0.21)	194.7(0.85)	7.5(0.02)	6.3(0.01)
	DFGM-C	14.0(0.03)	17.3(0.02)	64.0(0.12)	105.7(0.12)	8.4(0.02)	7.5(0.01)
	DFGM-G	14.1(0.03)	17.3(0.02)	64.2(0.12)	106.5(0.12)	8.4(0.02)	7.3(0.01)
	DGM-C	17.8(0.04)	21.5(0.03)	83.4(0.17)	138.9(0.20)	8.1(0.02)	6.8(0.01)
25	DGM-G	17.8(0.04)	21.7(0.03)	83.9(0.17)	141.1(0.20)	8.2(0.02)	6.7(0.01)
	GGL	18.3(0.16)	27.3(0.21)	94.7(0.16)	142.4(0.89)	8.0(0.02)	6.7(0.01)
	Naive-C	22.4(0.13)	36.5(0.21)	93.1(0.42)	180.8(0.96)	7.8(0.02)	6.2(0.01)
	Naive-G	23.6(0.33)	37.5(0.50)	103.1(0.85)	184.8(0.98)	8.0(0.02)	6.2(0.01)
	DFGM-C	13.3(0.03)	16.5(0.02)	60.8(0.11)	98.2(0.16)	8.8(0.02)	7.7(0.01)
	DFGM-G	13.4(0.03)	16.7(0.02)	62.4(0.12)	102.6(0.17)	8.8(0.02)	7.6(0.01)
	DGM-C	16.9(0.04)	20.1(0.03)	78.9(0.15)	127.0(0.27)	8.5(0.02)	7.2(0.01)
50	DGM-G	17.1(0.04)	20.7(0.03)	80.2(0.17)	135.4(0.30)	8.7(0.02)	7.1(0.01)
	GGL	18.2(0.16)	24.9(0.28)	93.7(0.14)	137.0(0.65)	8.4(0.03)	7.1(0.01)
	Naive-C	22.8(0.15)	31.8(0.21)	97.8(0.41)	186.0(1.00)	8.1(0.02)	6.7(0.01)
	Naive-G	26.1(0.49)	40.0(0.57)	105.7(1.06)	189.4(1.19)	8.3(0.02)	6.5(0.01)

DFGM, doubly functional graphical model; DGM, dynamic graphical model; GGL, group graphical lasso; Naive, naive approach; C, constrained ℓ_1-minimization; G, graphical lasso.

minimization or graphical lasso on $S(u_t)$ for $t = 1, \ldots, T$. To ensure these competitors work for sparse designs we split $[0, 1]$ into five equal subintervals, with $[T/5]$ points randomly sampled from each interval.

We examine the performance of seven approaches based on estimation accuracy and graph recovery consistency. In terms of the estimation accuracy, we calculate the mean of the operator and Frobenius losses for the estimated precision matrices, respectively defined as $||\widehat{\Theta}(u) - \Theta(u)||$ and $||\widehat{\Theta}(u) - \Theta(u)||_F$, at $u_1, \ldots, u_T$. In terms of the model selection consistency, we plot the true positive rates against false positive rates, respectively defined as $\#\{(j, k) : \widehat{\Theta}_{jk}^{(\lambda_n)}(u) \neq 0 \text{ and } \Theta_{jk}(u) \neq 0\}/\#\{(j, k) : \Theta_{jk}(u) \neq 0\}$ and $\#\{(j, k) : \widehat{\Theta}_{jk}^{(\lambda_n)}(u) \neq 0 \text{ and } \Theta_{jk}(u) = 0\}/\#\{(j, k) : \Theta_{jk}(u) = 0\}$, over a grid of $\lambda_n(u)$ values to produce the receiver operating characteristic curve at each u_t. For each comparison approach, we compute the average area under the curve at $u_1, \ldots, u_T$, with values closer to one indicating better performance in recovering the graph support.

Table 2 reports numerical summaries to compare different approaches over six simulation settings, corresponding to $p = 50, 100$ and $T = 10, 25, 50$. Several conclusions can be drawn from Table 2. First, in all scenarios, our proposed approach is superior to the competing methods in both estimation accuracy and model selection consistency, and in many cases the improvements are highly statistically significant. Among the others, dynamic-graphical-model-based approaches perform better than the remaining methods and the naive methods, which do not borrow

Table 3: Average (standard error) means of operator, Frobenius losses and AUROCs at $v_1, \ldots, v_{21}$ over 100 simulation runs. All entries have been multiplied by 10 for formatting reasons.

T_{ij}	Method	Operator norm		Frobenius norm		AUROC	
		$p = 50$	$p = 100$	$p = 50$	$p = 100$	$p = 50$	$p = 100$
6-9	DFGM-C	25.9(0.15)	38.8(0.11)	122.4(0.59)	192.3(0.73)	7.6(0.02)	6.3(0.01)
	DFGM-G	26.0(0.16)	39.5(0.10)	128.5(0.49)	199.4(0.61)	7.6(0.02)	6.2(0.01)
20-30	DFGM-C	20.1(0.11)	22.6(0.03)	94.0(0.21)	152.8(0.29)	8.0(0.02)	6.6(0.01)
	DFGM-G	21.1(0.21)	22.1(0.08)	95.9(0.35)	152.9(0.25)	8.0(0.02)	6.5(0.01)
50	DFGM-C	17.0(0.05)	21.3(0.06)	74.3(0.13)	137.2(0.20)	8.3(0.02)	6.9(0.01)
	DFGM-G	16.9(0.08)	21.9(0.10)	74.8(0.21)	138.6(0.32)	8.2(0.02)	6.8(0.01)

strength across adjacent points, provide the worst performance. Second, we observe that implementing the constrained ℓ_1-minimization and the graphical lasso give comparable results in many scenarios with the former type providing large improvements in a couple of cases. Third, the best results are obtained for the more densely sampled case, with a smaller number of functional variables.

4·3. *Irregular set of time points*

When functions are observed at irregular sets of points, none of the three types of competitors described in Section 4·2 are applicable. Hence, in this section we compare the sample performance of our doubly-functional-graphical-model based constrained ℓ_1 minimization and graphical lasso methods to each other. We consider six scenarios, corresponding to $p = 50, 100$ and different T_{ij}'s generated from the discrete uniform distribution with sets $\{6, \ldots, 9\}$, $\{20, \ldots, 30\}$ and $\{50\}$. The measurement times are randomly sampled from Uniform$[0, 1]$ for each pair of (i, j). We average the operator, Frobenius losses and the areas under the curves at 21 evenly-spaced time points, $0 = v_1, \ldots, v_{21} = 1$. Table 3 presents numerical results for all six simulations. We observe similar trends to those in Table 2 with results deteriorating somewhat for smaller values of T_{ij} and larger values of p. In general the constrained ℓ_1 minimization approach outperforms the graphical lasso on the estimation accuracy, but the methods are comparable in terms of graph selection consistency.

5. REAL DATA

In this section, we apply our proposed approach to the EEG data set, available at `https://archive.ics.uci.edu/ml/datasets/EEG+Database`, from an alcoholism study (Zhang et al., 1995). The data consists of measurements on 77 alcoholic and 45 control subjects. Each subject, exposed to either a single stimulus or two stimuli, completed 120 trials. EEG signals were measured at 256 time points over a one second time interval at 64 electrodes/nodes, placed at standard locations. Following the approach taken in Zhu et al. (2016) and Qiao et al. (2019), we averaged EEG signals, filtered at α-band (Hayden et al., 2006), across all trials under the single stimulus. The α-band filtering was performed using the eegfilt function in MATLAB. Existing research has shown that the networks embedded in EEG data evolve over time, where edges are bound to emerge and disappear (Cabral et al., 2014). In this study, our target is to estimate functional networks involving $p = 64$ nodes based on $n_a = 77$ and $n_c = 45$ functional observations for alcoholic and control groups respectively and to explore the differences in their brain connectivity patterns.

Since the graphical structures for alcoholic and non-alcoholic groups share some common edges, it is advantageous to jointly estimate two networks. Hence, we used the joint constrained ℓ_1-minimization approach (Cai et al., 2016) to simultaneously estimate two functional precision matrices. In addition, to stabilize the functional graph selection, at each time point, we bootstrapped each group by randomly selecting n_a and n_c samples with replacement from the alcoholic and control groups respectively, performed functional principal components analysis, implemented joint constrained ℓ_1-minimization to obtain two estimated networks and repeated the above procedure 100 times. Those edges, which were chosen more than 50 times out of 100 bootstrap samples, were finally selected as important edges. See Cai et al. (2016) for details on the selection of relevant regularization parameters.

Figure 2 plots the estimated graphs for the alcoholic and control groups at approximately $u = 0.2$, $u = 0.5$ and $u = 0.8$ seconds respectively. To visualize and interpret the functional network we set the functional sparsity to 5% and only displayed the top 101 most important edges in Figure 2, where three anatomical landmarked electrodes, X, Y and nd, were removed. The node names are provided in Table 4 in the Supplementary Material. We observe a few interesting patterns. First, the alcoholic and control groups share very similar block patterns, which reveals the existence of some regional effects for brain connectivity. Second, our estimated networks indicate clear dynamic structure. The edge values within each block gradually change with edges emerging and vanishing over the evolution, for example in Figure 2, electrodes FC6 and T8 in the control group are connected at $u = 0.2$ and 0.8, but disconnected at $u = 0.5$. Third, the dynamic networks differ between the two groups especially in certain regions, for example in Figure 2, electrodes from the left part of the brain are more connected in the alcoholic group.

SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online includes additional regularity conditions, all the technical proofs, further discussion and additional empirical results.

A. APPENDIX

A·1. *Procedure to generate simulated functional data*

By the definition of $C_{jk}(u, v) = \text{cov}\{X_{ij}(u), X_{ik}(v)\} = s(u)^{\mathrm{T}}\Omega_{jk}s(v)$ and orthonormality of $s(u)$, we can easily show that

$$\Omega_{jk} = \int_{(u,v)\in\mathcal{U}^2} s(u)C_{jk}(u,v)s(v)^{\mathrm{T}}dudv. \tag{A1}$$

To generate multivariate functional observations, we first need to construct a valid matrix of cross-covariance operators $\{C_{jk}(u,v)\}_{1\leqslant j,k\leqslant p}$ (Guhaniyogi et al., 2013) and then obtain Ω_{jk} from (A1), by approximating the integral using the discretized sums, We take the idea of linear models of coregionalization (Genton & Kleiber, 2015) to represent the p-dimensional multivariate random field as a linear combination of p independent univariate random fields by

$$C_{jk}(u,v) = \sum_{l=1}^{p} \rho(u-v)A_{jl}(u)A_{kl}(v), \tag{A2}$$

$u = 0.2$

$u = 0.5$

$u = 0.8$
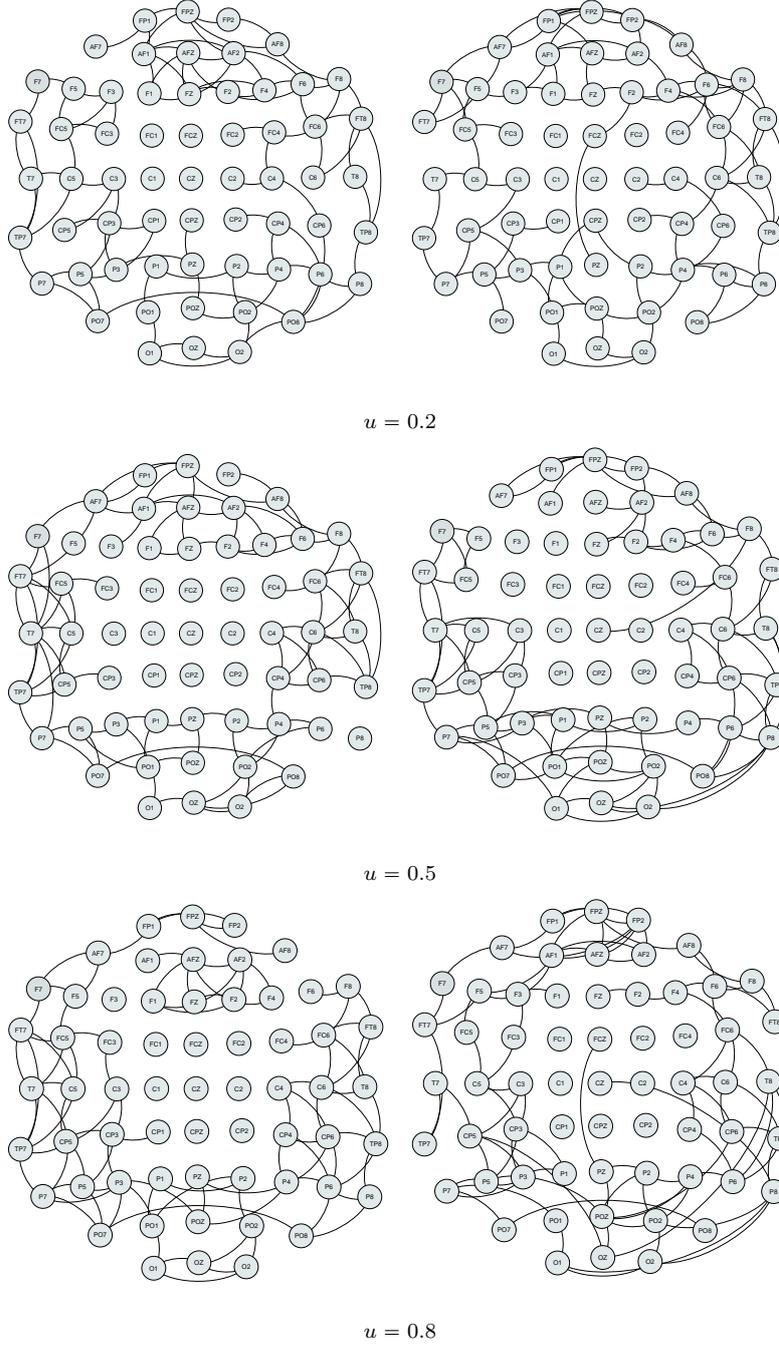
Fig. 2: *Left and right graphs plot the estimated dynamic networks at approximately $u = 0.2$, $u = 0.5$ and $u = 0.8$, for the alcoholic and control groups, respectively.*

where the correlation function $\rho(u - v)$ is independent of $l \in V$. Specially, for $u = v$, $\Sigma(u) = C(u, u) = \rho(0)A(u)A(u)^{\mathsf{T}}$, where we can set $A(u) = \{A_{jk}(u)\}_{1 \leqslant j,k \leqslant p}$ to be the Cholesky factor of $\Sigma(u)/\rho(0)$ and

490 hence $\{C_{jk}(u,v)\}$ can be generated from (A2) by letting $\rho(u-v) = \exp\{-(u-v)^2/2\sigma_\rho^2\}$, a univariate Gaussian kernel with $\sigma_\rho = 1$.

REFERENCES

BICKEL, P. J. & LEVINA, L. (2008). Covariance regularization by thresholding. *The Annals of Statistics* **136**, 2577–2604.

495 CABRAL, J., KRINGELBACH, M. L. & DECO, G. (2014). Exploring the network dynamic underlying brain activity during rest. *Progress in Neurobiology* **114**, 102–131.

CAI, T., LI, H., LIU, W. & XIE, J. (2016). Joint estimation of multiple high-dimensional precision matrices. *Statistical Sinica* **26**, 445–464.

CAI, T., LIU, W. & LUO, X. (2011). A constrained l_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association* **106**, 594–607.

CHEN, Z. & LENG, C. (2016). Dynamic covariance models. *Journal of the American Statistical Association* **111**, 1196–1208.

CHUN, H., CHEN, M., LI, B. & ZHAO, H. (2013). Joint conditional gaussian graphical models with multiple sources of genomic data. *Frontiers in Genetics* **4**, 294.

505 DANAHER, P., WANG, P. & WITTEN, D. M. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society: Series B* **76**, 373–397.

FRIEDMAN, J. H., HASTIE, T. J. & TIBSHIRANI, R. J. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9**, 432–441.

GENTON, M. G. & KLEIBER, W. (2015). Cross-covariance functions for multivariate geostatistics. *Statistical Science* **30**, 147–163.

GUHANIYOGI, R., FINLEY, A. O., BANERJEE, S. & KOBE, R. K. (2013). Modeling complex spatial dependencies: low rank spatially varying cross-covariances with application to soil nutrient data. *Journal of Agricultural, Biological and Environmental Statistics* **18**, 274–298.

HAYDEN, E. P., WIEGAND, R. E., MEYER, E. T., BAUER, L. O., O'CONNOR, S. J., NURNBERGER JR., J. I., CHORLIAN, D. B., PROJESZ, B. & BEGLEITER, H. (2006). Patterns of regional brain activity in alcohol-dependent subjects. *Alcoholism: Clinical and Experimental Research* **30**, 1986–1991.

KOLAR, M. & XING, E. (2011). On time varying undirected graphs. *Journal of Machine Learning Research:Workshop and Conference Proceedings* **15**, 407–415.

LAURITZEN, S. L. (1996). *Graphical Models*. New York: Oxford Univ. Press.

520 LI, B., & SOLEA, E. (2018). A nonparametric graphical model for functional data with application to brain networks based on fmri. *Journal of the American Statistical Association* **113**, 1637–1655.

MEINSHAUSEN, N. & BUHLMANN, P. (2006). High dimensional graphs and variable selection with the lasso. *The Annals of Statistics* **34**, 1436–1462.

QIAO, X., GUO, S. & JAMES, G. M. (2019). Functional graphical models. *Journal of the American Statistical Association* **114**, 211–222.

525 QIU, H., HAN, F., LIU, H. & CAFFO, B. (2016). Joint estimation of multiple graphical models from high dimensional time series. *Journal of the Royal Statistical Society: Series B* **78**, 487–504.

RAVIKUMAR, P., WAINWRIGHT, M. J., RASKUTTI, G. & YU, B. (2011). High-dimensional covariance estimation by minimizing l_1-penalized log-determinant divergence. *Electronic Journal of Statistics* **5**, 935–980.

530 RICE, J. & WU, C. O. (2001). Nonparametric mixed effect models for unequally sampled noisy curves. *Biometrics* **57**, 253–259.

RICE, J. A. & SILVERMAN, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society: Series B* **53**, 233–243.

STOREY, J. D., XIAO, W., LEEK, T. J., TOMPKINS, R. G. & DAVIS, R. W. (2005). Significance analysis of time course microarray experiments. *Proceedings of the National Academy of Sciences* **102**, 12837–12842.

WITTEN, D. M., FRIEDMAN, J. H. & SIMON, N. (2011). New insights and faster computations for the graphical lasso. *Journal of Computational and Graphical Statistics* **20**, 892–900.

YAO, F., MULLER, H. G. & WANG, J. L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100**, 577–590.

540 YUAN, M. & LIN, Y. (2007). Model selection and estimation in the gaussian graphical model. *Biometrika* **94**, 19–35.

ZHANG, X. & WANG, J. L. (2016). From sparse to dense functional data and beyond. *The Annals of Statistics* **5**, 2281–2321.

ZHANG, X. L., BEGLEITER, B., PORJESZ, B., WANG, W. & LITKE, A. (1995). Event related potentials during object recognition tasks. *Brain Research Bulletin* **38**, 531–538.

545 ZHOU, S., LAFFERTY, J. & WASSERMAN, L. (2010). Time varying undirected graphs. *Machine Learning* **80**, 295–319.

ZHU, H., STAWN, N. & DUNSON, D. B. (2016). Bayesian graphical models for multivariate functional data. *Journal of Machine Learning Research* **17**, 1–27.

Supplementary Material to "Doubly Functional Graphical Models in High Dimensions"

Xinghao Qiao, Cheng Qian, Gareth M. James and Shaojun Guo

This supplementary material contains some standard regularity conditions and the technical proofs supporting Section 3 in Appendix B, discussion of the approximate numerical integration approach under very dense measurement schedules in Appendix C and additional empirical results in Appendix D.

B. Technical proofs

We present several regularity conditions in Section B·1. In Sections B·2, B·3 and B·4, we prove that the concentration bounds in (12)–(13) of Condition 2 hold for fully observed functional data, sparsely observed functional data and densely observed functional data, respectively. In Sections B·5 and B·6, we provide proofs of Theorems 1 and 2, respectively. For convenient presentation in our proofs, we will use $c_1, c_2, \ldots$ as positive constants.

B·1. *Regularity conditions*

To investigate the theoretical properties of the proposed method in Theorems 1 and 2, we also need Conditions 5 and 6 below. They are standard in functional data analysis literature.

Condition 5. (i) Let $\{U_{ijt} : i = 1, \ldots, n, j \in V, t = 1, \ldots, T_{ij}\}$ be independent and identically distributed copies of a random variable U with density $f_U(\cdot)$ defined on the compact set $\mathcal{U}$, with the T_{ij}'s fixed. There exist some constants m_f, M_f, such that $0 < m_f \leqslant \inf_{\mathcal{U}} f_U(u) \leqslant \sup_{\mathcal{U}} f_U(u) \leqslant M_f < \infty$; (ii) X, e and U are independent.

Condition 6. For each $j \in V$, $\partial^2 C_{jj}(u,v)/\partial u^2$, $\partial^2 C_{jj}(u,v)/\partial u \partial v$ and $\partial^2 C_{jj}(u,v)/\partial v^2$ are bounded on $\mathcal{U}^2$.

To prove Condition 2, we need Conditions 8–9 below. We first assume the one-dimensional kernel K to satisfy the following condition.

Condition 7. (i) $K(\cdot)$ is symmetric probability density function on $[-1, 1]$ with $\int u^2 K(u) < \infty$ and $\int K(u)^2 du < \infty$. (ii) $K(\cdot)$ is Lipschitz continuous: there exists some positive constant L such that

$$|K(u) - K(v)| \leqslant L|u - v|, \quad \text{for any } u, v \in [0, 1].$$

Condition 8. For each $j \in V$, $C_{jj}(u,v)$ is in the trace class with $\max_{j \in V} \sum_{l=1}^{\infty} \omega_{jl} < \infty$.

Condition 9. For each $j \in V$, the eigenfunctions, $\phi_{j1}(\cdot), \phi_{j2}(\cdot), \ldots$ are Lipschitz-continuous, that is, there exists some positive constant L such that

$$|\phi_{jl}(u) - \phi_{jl}(v)| \leqslant L|u - v|, \quad \text{for any } (u, v) \in \mathcal{U}^2 \text{ and } l = 1, 2, \ldots.$$

B·2. *Concentration bounds for fully observed functional data*

LEMMA 1. *Suppose that Conditions 3 (ii) and 8 hold, then there exists some positive constants c_1 and c_2 such that for any $0 < \delta \leqslant 1$ and each $j \in V$, we have*

$$\text{pr}\left(\left\|\widehat{C}_{jj} - C_{jj}\right\|_{\mathcal{S}} \geqslant \delta\right) \leqslant c_2 \exp(-c_1 n \delta^2),$$

Proof. This lemma can be found in Lemma 6 in the supplementary material of Qiao et al. (2018) and hence the proof is omitted. □

LEMMA 2. *Suppose Conditions 3 (ii), 8 and 9 hold, then there exists some positive constants c_1 and c_2 such that for any $0 < \delta \leqslant 1$ and each $j \in V$, we have*

$$\mathrm{pr}\left\{\sup_{(u,v)\in\mathcal{U}^2}|\widehat{C}_{jj}(u,v) - C_{jj}(u,v)| \geqslant \delta\right\} \leqslant c_2 n \exp(-c_1 n \delta^2).$$

Proof. Without loss of generality, let $\mathcal{U} = [0, 1]$. We first reduce the problem from supremum over product of interval $[0, 1]^2$ to the maximum over a grid of pairs on the product interval. We partition the interval $[0, 1]$ into N subintervals B_k $(k = 1, \dots, N)$ of equal length. Let $u_k, v_{k'}$ be the centers of B_k and $B_{k'}$, respectively. For each $(u, v) \in B_k \times B_{k'}$, by the Lipschitz-continuity of $\phi_{jl}(\cdot)$'s and the expansion of $X_{ij}(u) = \sum_{l=1}^{\infty} \xi_{ijl}\phi_{jl}(u)$, we have

$$|\widehat{C}_{jj}(u,v) - \widehat{C}_{jj}(u_k, v_{k'})|$$

$$= |n^{-1}\sum_{i=1}^{n}\{X_{ij}(u)X_{ij}(v) - X_{ij}(u_k)X_{ij}(v_{k'})\}|$$

$$\leqslant \left|n^{-1}\sum_{i=1}^{n}[\{X_{ij}(u) - X_{ij}(u_k)\}X_{ij}(v) + X_{ij}(u_k)\{X_{ij}(v) - X_{ij}(v_{k'})\}]\right|$$

$$\leqslant n^{-1}\sum_{i=1}^{n}\{|\sum_{l=1}^{\infty}\xi_{ijl}X_{ij}(v)|\max_l|\phi_{jl}(u) - \phi_{jl}(u_k)| + |\sum_{l=1}^{\infty}\xi_{ijl}X_{ij}(u_k)|\max_l|\phi_{jl}(v) - \phi_{jl}(v_k)|\}$$

$$\leqslant n^{-1}\sum_{i=1}^{n}\{|\sum_{l=1}^{\infty}\xi_{ijl}X_{ij}(v)|c_1|u - u_k| + |\sum_{l=1}^{\infty}\xi_{ijl}X_{ij}(u_k)|c_1|v - v_{k'}|\}$$

$$\leqslant c_2\big(E\{|\sum_{l=1}^{\infty}\xi_{ijl}X_{ij}(u)|\} + 1\big)N^{-1}$$

on the event $\Omega = \left\{n^{-1}\sum_{i=1}^{n}|\sum_{l=1}^{\infty}\xi_{ijl}X_{ij}(u)| \leqslant E\{|\sum_{l=1}^{\infty}\xi_{ijl}X_{ij}(u)|\} + 1\right\}$. By the Cauchy–Schwarz inequality, Conditions 3(ii) and 8, we have $\left[E\{|\sum_{l=1}^{\infty}\xi_{ijl}X_{ij}(u)|\}\right]^2 \leqslant E\{(\sum_{l=1}^{\infty}\xi_{ijl})^2\}E\{X_{ij}(u)^2\} \leqslant (\sum_{l=1}^{\infty}\omega_{jl})(\sum_{l=1}^{\infty}\omega_{jl})O(1) < \infty$. Combing the above results and some specific calculations yields

$$\left||\widehat{C}_{jj}(u,v) - C_{jj}(u,v)| - |\widehat{C}_{jj}(u_k,v_{k'}) - C_{jj}(u_k,v_{k'})|\right|$$

$$\leqslant \left|\{\widehat{C}_{jj}(u,v) - \widehat{C}_{jj}(u_k,v_{k'})\} - \{C_{jj}(u,v) - C_{jj}(u_k,v_{k'})\}\right|$$

$$\leqslant 2c_2(E\{|\sum_{l=1}^{\infty}\xi_{ijl}X_{ij}(u)|\} + 1)N^{-1} < c_3N^{-1},$$

which implies that

$$\sup_{(u,v)\in B_k\times B_{k'}}|\widehat{C}_{jj}(u,v) - C_{jj}(u,v)| \leqslant \max_{1\leqslant k\leqslant N}|\widehat{C}_{jj}(u_k,v_{k'}) - C_{jj}(u_k,v_{k'})| + C_3N^{-1}.$$

Let $N = c_3\delta^{-1}$. Under the Gaussian assumption for $X_{ij}(u)$ and $\sum_{l=1}^{\infty}\xi_{ijl}$ with finite variance by Condition 8, it follows from the Bernstein inequality (Boucheron et al., 2014) that there exist positive constants c_4 and c_5 such that for any $0 < \delta \leqslant 1$, $\mathrm{pr}\{|\widehat{C}_{jj}(u_k,v_{k'}) - C_{jj}(u_k,v_{k'})| \geqslant \delta\} \leqslant c_5\exp(-c_4n\delta^2)$ and $\mathrm{pr}(\Omega^C) \leqslant c_5\exp(-c_4n)$. Combing the above results and applying the union bound of probability, we obtain that

$$\mathrm{pr}\left\{\sup_{(u,v)\in[1,1]^2}|\widehat{C}_{jj}(u,v) - C_{jj}(u,v)| \geqslant \delta\right\} \leqslant c_6\delta^{-1}\exp(-c_4n\delta^2).$$

If $\delta \geqslant n^{-1}$, the right hand side of the above inequality is reduced to $c_6 n \exp(-c_4 n \delta^2)$. If $\delta < n^{-1}$, we can choose c_6 and $n > c_7$, such that $c_6 \exp(-c_4 c_7^{-1}) \geqslant 1$ and hence the bound $c_6 n \exp(-c_4 n \delta^2)$ can still be used. We complete the proof for this lemma. □.

B·3. *Concentration bounds for sparsely observed functional data*

In Lemma 4, we prove the concentration bounds for sparsely observed functional data. The proof of Lemma 4 replies on the concentration result for the second-order U-statistics in Lemma 3.

LEMMA 3. *Let $Z_{ij} = (X_{ij}, Y_{ij})^{\mathrm{T}}$ $(i = 1, \ldots, n, j = 1, \ldots, m)$. Suppose that $\{Z_{ij}\}$ are independent and identically distributed random vectors. Consider the U-statistics $S_n = \frac{1}{nm(m-1)} \sum_{i=1}^{n} \sum_{j \neq k}^{m} X_{ij} Y_{ik}$. If there exist two positive constants v and c such that for each $q = 2, 3, \ldots,$*

$$\sup_{1 \leqslant i \leqslant n} \sup_{1 \leqslant j \neq k \leqslant m} E \Big| X_{ij} Y_{ik} - E\big(X_{ij} Y_{ik}\big) \Big|^q \leqslant \frac{q!}{2} v c^{q-2}, \tag{B.1}$$

then for any $\delta > 0$, we have

$$\mathrm{pr}\left(\Big| S_n - E S_n \Big| \geqslant \delta \right) \leqslant 2 \exp \left\{ -\frac{n(m-1)\delta^2}{4(v + c\delta)} \right\}. \tag{B.2}$$

Proof. Let $J = [m/2]$, the largest integer smaller than or equal to $m/2$ and define

$$W(Z_{i1}, \ldots, Z_{im}) = \frac{1}{J} \Big\{ X_{i1} Y_{i2} + \cdots + X_{i(2J-1)} Y_{i(2J)} \Big\}.$$

Let $\sum_{(p)}$ denote the sum of all $m!$ permutations $(j_1, \ldots, j_m)$ of $(1, \ldots, m)$. We can show that S_n can be expressed as

$$S_n = \frac{1}{m!} \sum_{(p)} \frac{1}{n} \sum_{i=1}^{n} W(Z_{ij_1}, \ldots, Z_{ij_m}) = \frac{1}{m!} \sum_{(p)} \frac{1}{nJ} \sum_{i=1}^{n} \sum_{k=1}^{J} X^*_{ij_{(2k-1)} j_{2k}},$$

where $X^*_{ij_1 j_2} = X_{ij_1} Y_{ij_2}$ for $1 \leqslant i \leqslant n$ and $1 \leqslant j_1 \neq j_2 \leqslant m$. The above expression represents S_n as an average of $m!$ terms, each of which itself is an average of nJ independent random variables.

Let $\phi(x) = \exp(x) - x - 1$. Hence, for $\lambda > 0$, by monotone convergence theorem and (B.1), we have that

$$\sup_{i, j_1, j_2} E \Big\{ \phi\big(\lambda X^*_{ij_1 j_2} - \lambda E X^*_{ij_1 j_2} \big) \Big\} \leqslant \sum_{q=2}^{\infty} \frac{\lambda^q E\big\{ |X^*_{ij_1 j_2} - E X^*_{ij_1 j_2})|^q \big\}}{q!} \leqslant \frac{v}{2} \sum_{q=2}^{\infty} \lambda^q c^{q-2} \leqslant \frac{v \lambda^2}{2(1 - c\lambda)}$$

for $0 < \lambda < c^{-1}$. Using the above result and the fact that $\log(x) \leqslant x - 1$ for $x > 0$, we have that for $\lambda > 0$,

$$\log \Big[E \exp \big\{ \lambda(S_n - E S_n) \big\} \Big]$$

$$\leqslant \sum_{i=1}^{n} \log \left(\frac{1}{m!} \sum_{(p)} \exp \left[\sum_{k=1}^{J} \log E \exp \left\{ \frac{\lambda}{nJ} (X^*_{ij_{2k-1} j_{2k}} - E X^*_{ij_{2k-1} j_{2k}}) \right\} \right] \right)$$

$$\leqslant \sum_{i=1}^{n} \log \left(\frac{1}{m!} \sum_{(p)} \exp \left[\sum_{k=1}^{J} \log \left\{ 1 + \frac{n^{-2} J^{-2} v \lambda^2}{2(1 - cn^{-1} J^{-1} \lambda)} \right\} \right] \right)$$

$$= \frac{n^{-1} J^{-1} v \lambda^2}{2(1 - cn^{-1} J^{-1} \lambda)}, \ 0 < \lambda < nJc^{-1}.$$

Applying Bernstein inequality in Corollary 2.11 of Boucheron et al. (2014), we obtain that for any $\delta > 0$,

$$\mathrm{pr}\left(\Big| S_n - E S_n \Big| \geqslant \delta \right) \leqslant 2 \exp \left\{ -\frac{nJ\delta^2}{2(v + c\delta)} \right\}.$$

Note that $J \geqslant (m-1)/2$ and hence the concentration inequality in (B.2) follows. □

LEMMA 4. *Suppose that Conditions 3(ii), 5, 7 and 8 hold, $h_j = h$ and $T_{ij} \leqslant T_0 < \infty$ for all $i = 1, \ldots, n$ and $j \in V$. Then there exist a deterministic covariance function $\widetilde{C}_{jj}(u, v)$ defined in (B.9) and some positive constants c_1, c_2 such that for any $\delta \in (0, 1]$ and each $j \in V$, we have*

$$\mathrm{pr}\Big\{\big\|\widehat{C}_{jj} - \widetilde{C}_{jj}\big\|_{\mathcal{S}} \geqslant \delta\Big\} \leqslant c_2 \exp\left(-c_1 n h^2 \delta^2\right) \tag{B.3}$$

and

$$\mathrm{pr}\left\{\sup_{(u,v)\in\mathcal{U}^2}\Big|\widehat{C}_{jj}(u,v) - \widetilde{C}_{jj}(u,v)\Big| \geqslant \delta\right\} \leqslant c_2 n h^{-2} \exp\left(-c_1 n h^2 \delta^2\right). \tag{B.4}$$

If $h \asymp n^{-a}$ for some positive constant $a < 1/2$, there exist some positive constants c_3 and c_4 such that (B.3) and (B.4) reduce to

$$\mathrm{pr}\Big\{\big\|\widehat{C}_{jj} - \widetilde{C}_{jj}\big\|_{\mathcal{S}} \geqslant \delta\Big\} \leqslant c_2 \exp\left(-c_3 n^{1-2a}\delta^2\right) \tag{B.5}$$

and

$$\mathrm{pr}\left\{\sup_{(u,v)\in\mathcal{U}^2}\Big|\widehat{C}_{jj}(u,v) - \widetilde{C}_{jj}(u,v)\Big| \geqslant \delta\right\} \leqslant c_4 n^{1+2a} \exp\left(-c_3 n^{1-2a}\delta^2\right), \tag{B.6}$$

respectively.

Proof. Denote $\widetilde{U}_{ijtt'}(u, v) = \big\{1, (U_{ijt} - u)/h, (U_{ijt'} - v)/h\big\}^{\mathrm{T}}$, $e_0 = (1, 0, 0)^{\mathrm{T}}$,

$$\begin{aligned}
\widehat{\Xi}_j(u,v) &= n^{-1}\sum_{i=1}^n w_{ij}\sum_{1\leqslant t\neq t'\leqslant T_{ij}} \widetilde{U}_{ijtt'}(u,v)\widetilde{U}_{ijtt'}^{\mathrm{T}}(u,v)K_h(U_{ijt}-u)K_h(U_{ijt'}-v),\\
\widehat{Z}_j(u,v) &= n^{-1}\sum_{i=1}^n w_{ij}\sum_{1\leqslant t\neq t'\leqslant T_{ij}} \widetilde{U}_{ijtt'}(u,v)Y_{ijt}Y_{ijt'}K_h(U_{ijt}-u)K_h(U_{ijt'}-v).
\end{aligned} \tag{B.7}$$

Then we express $\widehat{C}_{jj}(u, v)$ as

$$\widehat{C}_{jj}(u,v) = e_0^{\mathrm{T}}\big\{\widehat{\Xi}_j(u,v)\big\}^{-1}\widehat{Z}_j(u,v). \tag{B.8}$$

For $k, \ell = 1, 2, 3$, denote $\widehat{\Xi}_{jk\ell}(u,v)$ and $\widehat{Z}_{jk}(u,v)$ be the (k, ℓ)th entry of $\widehat{\Xi}_j(u,v)$ and the kth element of $\widehat{Z}_j(u,v)$, respectively. We define a deterministic covariance function by

$$\widetilde{C}_{jj}(u,v) = e_0^{\mathrm{T}}\Big[E\big\{\widehat{\Xi}_j(u,v)\big\}\Big]^{-1}\Big[E\big\{\widehat{Z}_j(u,v)\big\}\Big]. \tag{B.9}$$

We next prove that there exists two positive constants c_1 and c_2 such that for any $\delta > 0$ and each $j \in V$, $m, \ell = 1, 2, 3$ and $(u, v) \in \mathcal{U}^2$,

$$\mathrm{pr}\left\{\Big|\widehat{\Xi}_{jm\ell}(u,v) - E\big\{\widehat{\Xi}_{jm\ell}(u,v)\big\}\Big| \geqslant \delta\right\} \leqslant 2\exp\left(-\frac{c_1 n h^2 \delta^2}{1+\delta}\right) \tag{B.10}$$

and

$$\mathrm{pr}\left\{\Big|\widehat{Z}_{jm}(u,v) - E\big\{\widehat{Z}_{jm}(u,v)\big\}\Big| \geqslant \delta\right\} \leqslant c_2 \exp\left(-\frac{c_1 n h^2 \delta^2}{1+\delta}\right). \tag{B.11}$$

It follows from Conditions 5 and 7 that for integer $q = 2, 3, \ldots$ and $s, s' = 0, 1, 2$, we have

$$
E\Big| K_h(U_{ijt} - u)K_h(U_{ijt'} - v)\Big(\frac{U_{ijt} - u}{h}\Big)^s \Big(\frac{U_{ijt} - u}{h}\Big)^{s'}\Big|^q
$$

$$
\leqslant \int\int \frac{1}{h^{2q}} K^q\Big(\frac{t-u}{h}\Big) K^q\Big(\frac{t'-v}{h}\Big)\Big|\frac{t-u}{h}\Big|^{sq}\Big|\frac{t'-v}{h}\Big|^{s'q} f_U(t)f_U(t')dtdt'
$$

$$
\leqslant \frac{1}{h^{2q-2}} \int_{-1}^{1}\int_{-1}^{1} K^q(x)K^q(x')|x|^{sq}|x|^{s'q} f_U(hx+u)f_U(hx'+v)dxdx'
$$

$$
\leqslant c_3(c_4 h^{-2})(c_4 h^{-2})^{q-2}. \tag{B.12}
$$

By (B.12), the concentration result for the U-statistics in Lemma 3, $T_{ij} \leqslant T_0 < \infty$ and the first line of (B.7), we can prove that the concentration bound in (B.10) holds for $m, \ell = 1, 2, 3$.

Then we turn to prove (B.11). Consider $m = 1$ first. Define

$$
\chi_{ij1}(u, v) = \sum_{1 = t \neq t' \leqslant T_{ij}} Y_{ijt} Y_{ijt'} K_h(U_{ijt} - u)K_h(U_{ijt'} - v).
$$

Then $\widehat{Z}_{j1}(u, v) = n^{-1}\sum_{i=1}^{n} w_{ij}\chi_{ij1}(u, v)$. If follows from the fact $\sum_{l=1}^{\infty} \xi_{ijl}$ and e_{ijt} are Gaussian with finite variance, $Y_{ijt} = \sum_{l=1}^{\infty} \xi_{ijl}\phi_{jl}(U_{ijt}) + e_{ijt}$, Conditions 3 (ii) and 8 that for $q = 2, 3, \ldots$,

$$
E(|Y_{ijt}|^{2q}) \leqslant 2^{2q-1}\left\{ E\Big|\sum_{l=1}^{\infty} \xi_{ijl}\phi_{jl}(U_{ijt})\Big|^{2q} + E|e_{ijt}|^{2q}\right\}
$$

$$
\leqslant 2^{2q-1}\left\{ E\Big(\sum_{l=1}^{\infty} \xi_{ijl}\Big)^{2q} \sup_{i,j,t} E|\phi_{jl}(U_{ijt})|^{2q} + E|e_{ijt}|^{2q}\right\} \leqslant \frac{1}{2}q!c_5^q.
$$

Combing the above results implies that for $q = 2, 3, \ldots$,

$$
E|\chi_{ij1}(u, v)|^q \leqslant c_6 h^{-2q+2} \sum_{1 \leqslant t \neq t' \leqslant T_{ij}} E|Y_{ijt}Y_{ijt'}|^q \leqslant \frac{1}{2}q!c_7^q h^{-2q+2}. \tag{B.13}
$$

It then follows from the Bernstein inequality that (B.11) holds. Similar techniques used in proving (B.12) and (B.13) can be directly applied to other terms, $\widehat{Z}_{jm}(u, v)$ for $m = 2, 3$ and hence (B.11) holds.

We next decompose $\widehat{C}_{jj}(u, v) - \widetilde{C}_{jj}(u, v)$ as

$$
\widehat{C}_{jj}(u, v) - \widetilde{C}_{jj}(u, v) = e_0^{\mathrm{T}}\left(\{\widehat{\Xi}_j(u, v)\}^{-1} - \Big[E\{\widehat{\Xi}_j(u, v)\}\Big]^{-1}\right)\widehat{Z}_j(u, v)
$$
$$
+ e_0^{\mathrm{T}}\Big[E\{\widehat{\Xi}_j(u, v)\}\Big]^{-1}\Big[\widehat{Z}_j(u, v) - E\{\widehat{Z}_j(u, v)\}\Big]. \tag{B.14}
$$

By Conditions 3(ii), 5 and 7, some calculations and similar developments used in Zhang & Wang (2016) that $E\{\widehat{\Xi}_j(u, v)\}$ is positive definite and $E\{\widehat{Z}_j(u, v)\}$ is bounded. These results together with (B.14), concentration inequalities in (B.10)–(B.11) and some calculations imply that there exist some positive constants c_8 and c_9 such that for any $\delta > 0$ and each $j \in V$, $(u, v) \in \mathcal{U}^2$,

$$
\mathrm{pr}\Big\{|\widehat{C}_{jj}(u, v) - \widetilde{C}_{jj}(u, v)| \geqslant \delta\Big\} \leqslant c_9 \exp\left(-\frac{c_8 nh^2\delta^2}{1+\delta}\right). \tag{B.15}
$$

We then derive the concentration bound for $\|\widehat{C}_{jj} - \widetilde{C}_{jj}\|_{\mathcal{S}}$. It follows from (B.15) and the first part of Lemma 2 in Guo & Qiao (2018) that for any $(u, v) \in \mathcal{U}^2$ and integer $q \geqslant 1$,

$$
E\big|\widehat{C}_{jj}(u, v) - \widetilde{C}_{jj}(u, v)\big|^{2q} \leqslant q!c_{10}^q(n^{-1}h^{-2})^q + (2q)!c_{10}^q(n^{-1}h^{-2})^{2q}.
$$

This result together with the fact $E\|\widehat{C}_{jj} - \widetilde{C}_{jj}\|_{\mathcal{S}}^{2q} \leqslant |\mathcal{U}|^{2q} E|\widehat{C}_{jj}(u,v) - \widetilde{C}_{jj}(u,v)|^{2q}$ implies that

$$E\|\widehat{C}_{jj} - \widetilde{C}_{jj}\|_{\mathcal{S}}^{2q} \leqslant |\mathcal{U}|^{2q} c_{10}^q \Big\{ (n^{-1}h^{-2})^q + (2q)!(n^{-1}h^{-2})^{2q} \Big\}.$$

Applying the second part of Lemma 2 in Guo & Qiao (2018), we can show that, for each $\delta \in (0,1)$, the L_2 concentration bound for $\|\widehat{C}_{jj} - \widetilde{C}_{jj}\|_{\mathcal{S}}$ in (B.3) holds.

We finally derive the uniform concentration bound for $\sup_{(u,v)\in\mathcal{U}^2} |\widehat{C}_{jj}(u,v) - \widetilde{C}_{jj}(u,v)|$. Let $\mathcal{U} = [0,1]$. We partition the interval $\mathcal{U}$ into N subintervals B_k of equal length. Let $u_k, v_{k'}$ be the centers of B_k and $B_{k'}$, respectively. Then we have

$$\sup_{(u,v)\in B_k \times B_{k'}} |\widehat{C}_{jj}(u,v) - \widetilde{C}_{jj}(u,v)| \leqslant \max_{1\leqslant k\leqslant N} |\widehat{C}_{jj}(u_k, v_{k'}) - \widetilde{C}_{jj}(u_k, v_{k'})|$$
$$+ \Big| \big\{ \widehat{C}_{jj}(u,v) - \widehat{C}_{jj}(u_k, v_{k'}) \big\} - \big\{ \widetilde{C}_{jj}(u,v) - \widetilde{C}_{jj}(u_k, v_{k'}) \big\} \Big|.$$

We then need to bound $\Big| \big\{ \widehat{C}_{jj}(u,v) - \widehat{C}_{jj}(u_k, v_{k'}) \big\} - \big\{ \widetilde{C}_{jj}(u,v) - \widetilde{C}_{jj}(u_k, v_{k'}) \big\} \Big|$. By (B.8), (B.9) (B.14) and some calculations, it suffices to bound $\Big| \big\{ \widehat{Z}_{jm}(u,v) - \widehat{Z}_{jm}(u_k, v_{k'}) \big\} - \big[E\{\widehat{Z}_{jm}(u,v)\} - E\{\widehat{Z}_{jm}(u_k, v_{k'})\} \big] \Big|$ and $\Big| \big\{ \widehat{\Xi}_{jm\ell}(u,v) - \widehat{\Xi}_{jm\ell}(u_k, v_{k'}) \big\} - \big[E\{\widehat{\Xi}_{jm\ell}(u,v)\} - E\{\widehat{\Xi}_{jm\ell}(u_k, v_{k'})\} \big] \Big|$ for $m, \ell = 1, 2, 3$, which means we need to bound $|\{\widehat{Z}_{jm}(u,v) - \widehat{Z}_{jm}(u_k, v_{k'})\}|$ and $|\{\widehat{\Xi}_{jm\ell}(u,v) - \widehat{\Xi}_{jm\ell}(u_k, v_{k'})\}|$ for $m, \ell = 1, 2, 3$.

Let $(u,v) \in B_k \times B_{k'}$. Consider $m = 1$ first, it follows from Condition 7 that

$$\big| \big\{ \widehat{Z}_{j1}(u,v) - \widehat{Z}_{j1}(u_k, v_{k'}) \big\} \big|$$
$$\leqslant \Big| n^{-1} \sum_{i=1}^n w_{ij} \sum_{1\leqslant t\neq t'\leqslant T_{ij}} Y_{ijt} Y_{ijt'} \Big[\big\{ K_h(U_{ijt} - u) - K_h(U_{ijt} - u_k) \big\} K_h(U_{ijt'} - v) +$$
$$+ K_h(U_{ijt} - u_k) \big\{ K_h(U_{ijt'} - v) - K_h(U_{ijt'} - v_{k'}) \big\} \Big] \Big|$$
$$\leqslant n^{-1} \sum_{i=1}^n w_{ij} \Big| \sum_{1\leqslant t\neq t'\leqslant T_{ij}} Y_{ijt} Y_{ijt'} \Big| c_{11} h^{-2} \big\{ K_h(U_{ijt'} - v)|u - u_k| + K_h(U_{ijt} - u_k)|v - v_{k'}| \big\}$$
$$\leqslant c_{12} \Big[E\Big(w_{ij}| \sum_{1\leqslant t\neq t'\leqslant T_{ij}} Y_{ijt} Y_{ijt'} \big\{ K_h(U_{ijt'} - v) + K_h(U_{ijt} - u_k)| \big\} \Big) + 1 \Big] h^{-2} N^{-1}$$
$$= c_{12} \{ E(\widetilde{\chi}_{ij1}) + 1 \}(Nh^2)^{-1} \leqslant c_{13}(Nh^3)^{-1}, \tag{B.16}$$

on the event $\Omega_{Z,j1} = \big\{ n^{-1} \sum_{i=1}^n \widetilde{\chi}_{ij1} \leqslant E(\widetilde{\chi}_{ij1}) + 1 \big\}$. The last inequality follows from the Cauchy–Schwartz inequality and $E\{Y_{ijt}^2 K_h^2(U_{ijt} - u_k)\} = \sum_{l=1}^\infty E\{\xi_{ijl}^2 \phi_{jl}^2(U_{ijt}) K_h^2(U_{ijt} - u_k)\} + E\{e_{ijt}^2 K_h^2(U_{ijt} - u_k)\} \leqslant C_9 h^{-1}$, implied from Conditions 3(ii), 5, 7, 8 and some calculations similar to the proof of (B.12). Applying similar techniques to obtain (B.16), we can define events $\Omega_{Z,jm}, \Omega_{\Xi,jm\ell}$ for $m, \ell = 1, 2, 3$, respectively, and show that, on the above events, $|\{\widehat{Z}_{j1}(u,v) - \widehat{Z}_{j1}(u_k, v_{k'})\}| \leqslant c_{14}(Nh^3)^{-1}$ for $m = 2, 3$ and $|\{\widehat{\Xi}_{jm\ell}(u,v) - \widehat{\Xi}_{jm\ell}(u_k, v_{k'})\}| \leqslant c_{14}(Nh^3)^{-1}$ for $m, \ell = 1, 2, 3$. Combing the above results, we have

$$\sup_{(u,v)\in B_k \times B_{k'}} |\widehat{C}_{jj}(u,v) - \widetilde{C}_{jj}(u,v)| \leqslant \max_{1\leqslant k\leqslant N} |\widehat{C}_{jj}(u_k, v_{k'}) - \widetilde{C}_{jj}(u_k, v_{k'})| + C_{13}(Nh^3)^{-1}.$$

Let $N = c_{15}(h^3\delta)^{-1}$. It follows from (B.15), techniques used in the proof of (B.12), moment inequalities similar to (B.13) and the Bernstein inequality that there exists positive constants c_{16} and c_{17} such that for any $\delta \in (0,1]$, $\mathrm{pr}\big\{ |\widehat{C}_{jj}(u_k, v_{k'}) - \widetilde{C}_{jj}(u_k, v_{k'})| \geqslant \delta \big\} \leqslant c_{17} \exp(-c_{16} n h^2 \delta^2)$, $\mathrm{pr}(\Omega_{Z,jm}^C) \leqslant c_{17} \exp(-c_{16} n h^2)$ and $\mathrm{pr}(\Omega_{\Xi,jm\ell}^C) \leqslant c_{17} \exp(-c_{16} n h^2)$ for $m, \ell = 1, 2, 3$. Combing the above results

and by the union bound of probability, we have that for each $\delta \in (0, 1]$,

$$\text{pr}\left\{ \sup_{(u,v)\in[1,1]^2} |\widehat{C}_{jj}(u,v) - \widetilde{C}_{jj}(u,v)| \geqslant \delta \right\} \leqslant c_{19}(h^3\delta)^{-1} \exp\left(-c_{18}nh^2\delta^2 \right).$$

If $\delta \geqslant (nh)^{-1}$, the right side of the above inequality is reduced to $c_{19}nh^{-2}\exp\left(-c_{18}nh^2\delta^2 \right)$. If $\delta < (nh)^{-1}$, we can choose c_{19} and $n > c_{20}$ such that $c_{19}\exp(-c_{18}c_{20}^{-1}) \geqslant 1$ and hence the bound $c_{19}nh^{-2}\exp\left(-c_{18}nh^2\delta^2 \right)$ can still be used. We complete our proof for concentration inequalities in (B.3) and (B.4). □

B·4. *Concentration bounds for densely observed functional data* 645

LEMMA 5. *Suppose that Conditions* 1, 3 *(ii)*, 5, 7 *and* 8 *hold. Let* d_j *be dimensionality of* $X_j(\cdot)$ *such that* $\omega_{jd_j} > 0$ *and* $\omega_{j(d_j+1)} = 0$. *Suppose that* $d_j = d < \infty$, $h_j = h$ *for* $j \in V$. *Then there exist a deterministic covariance function* $\widetilde{C}_{jj}(u,v)$ *defined in (B.9), some positive constants* c_1 *and* c_2, *a fixed small constant* $\epsilon > 0$ *such that for any* $\delta \in (0, 1]$ *and each* $j \in V$,

$$\text{pr}\left\{ \|\widehat{C}_{jj} - \widetilde{C}_{jj}\|_{\mathcal{S}} \geqslant \delta \right\} \leqslant c_2 \exp\left\{ -c_1 \min(n, n^{2/3-\epsilon}T^{1/3}h^{4/3})\delta^2 \right\} \tag{B.17}$$

and 650

$$\begin{aligned}
&\text{pr}\left\{ \sup_{(u,v)\in\mathcal{U}^2} \left|\widehat{C}_{jj}(u,v) - \widetilde{C}_{jj}(u,v)\right| \geqslant \delta \right\} \\
&\leqslant c_2 \max(n, n^{2/3-\epsilon}T^{-1/3}h^{-4/3}) \exp\left\{ -c_1 \min(n, n^{2/3-\epsilon}T^{1/3}h^{4/3})\delta^2 \right\}.
\end{aligned} \tag{B.18}$$

If $h \asymp n^{-a}$ *and* $T \asymp n^b$ *for some positive constants* a, b *with* $4a/3 - b/3 < 2/3 - \epsilon$, *there exist some positive constants* c_3 *and* c_4 *such that (B.17) and (B.18) reduce to*

$$\text{pr}\left\{ \|\widehat{C}_{jj} - \widetilde{C}_{jj}\|_{\mathcal{S}} \geqslant \delta \right\} \leqslant c_2 \exp\left\{ -c_3 n^{\min\{1,(2/3-\epsilon+b/3-4a/3)\}}\delta^2 \right\} \tag{B.19}$$

and

$$\begin{aligned}
&\text{pr}\left\{ \sup_{(u,v)\in\mathcal{U}^2} \left|\widehat{C}_{jj}(u,v) - \widetilde{C}_{jj}(u,v)\right| \geqslant \delta \right\} \\
&\leqslant c_4 n^{\max\{1,(2/3-\epsilon-b/3+4a/3)\}} \exp\left\{ -c_3 n^{\min\{1,(2/3-\epsilon+b/3-4a/3)\}}\delta^2 \right\},
\end{aligned} \tag{B.20}$$

respectively.

Proof. We will show that there exist two positive constants c_1 and c_2 such that for any $0 < \delta \leqslant 1$ and each 655 $j \in V, m, \ell = 1, 2, 3$ and $(u, v) \in \mathcal{U}^2$,

$$\text{pr}\left\{ \left|\widehat{\Xi}_{jm\ell}(u,v) - E\{\widehat{\Xi}_{jm\ell}(u,v)\}\right| \geqslant \delta \right\} \leqslant c_2 \exp\left(-c_1 nTh^2\delta^2 \right) \tag{B.21}$$

and

$$\text{pr}\left\{ \left|\widehat{Z}_{jm}(u,v) - E\{\widehat{Z}_{jm}(u,v)\}\right| \geqslant \delta \right\} \leqslant c_2 \exp\left\{ -c_1 \min(n, n^{2/3-\epsilon}T^{1/3}h^{4/3})\delta^2 \right\}. \tag{B.22}$$

Then combing results in (B.14), (B.21), (B.22) and following the same developments to prove (B.15), we can obtain that there exist two positive constants c_3 and c_4 such that for any $0 < \delta \leqslant 1$, and each $j \in V$ and $(u, v) \in \mathcal{U}^2$, 660

$$\text{pr}\left\{ \left|\widehat{C}_{jj}(u,v) - \widetilde{C}_{jj}(u,v)\right| \geqslant \delta \right\} \leqslant c_4 \exp\left\{ -c_3 \min(n, n^{2/3-\epsilon}T^{1/3}h^{4/3})\delta^2 \right\}. \tag{B.23}$$

It follows from (B.7), Condition 1, (B.12) and Lemma 3 that the concentration bound in (B.21) holds for $m, \ell = 1, 2, 3$.

Now we turn to prove (B.22). Similar to the proof of (B.11), it suffices to derive the concentration bound for $\widehat{Z}_{j1}(u, v)$. For each pair (i, j) $(i = 1, \ldots, n, j = 1, \ldots, p)$, denote

$$\chi_{ij}(u, v) = \sum_{1 = t \neq t' \leqslant T} Y_{ijt} Y_{ijt'} K_h(U_{ijt} - u) K_h(U_{ijt'} - v),$$

which can be further decomposed as

$$\begin{aligned}
\chi_{ij}(u, v) &= \sum_{1 \leqslant t \neq t' \leqslant T} e_{ijt} e_{ijt'} K_h(U_{ijt} - u) K_h(U_{ijt'} - v) \\
&\quad + \sum_{l=1}^{d} \xi_{ijl} \sum_{1 \leqslant t \neq t' \leqslant T} e_{ijt'} \phi_{jl}(U_{ijt}) K_h(U_{ijt} - u) K_h(U_{ijt'} - v) \\
&\quad + \sum_{l=1}^{d} \xi_{ijl} \sum_{1 \leqslant t \neq t' \leqslant T} e_{ijt} \phi_{jl}(U_{ijt'}) K_h(U_{ijt} - u) K_h(U_{ijt'} - v) \\
&\quad + \sum_{l,l'=1}^{d} \xi_{ijl} \xi_{ijl'} \sum_{1 \leqslant t \neq t' \leqslant T} \phi_{jl}(U_{ijt}) \phi_{jl'}(U_{ijt'}) K_h(U_{ijt} - u) K_h(U_{ijt'} - v) \\
&= \chi_{ij}^{(1)}(u, v) + \chi_{ij}^{(2)}(u, v) + \chi_{ij}^{(3)}(u, v) + \chi_{ij}^{(4)}(u, v).
\end{aligned}$$

Therefore, we have

$$\widehat{Z}_{j1}(u, v) = \sum_{k=1}^{4} n^{-1} \sum_{i=1}^{n} w_{ij} \chi_{ij}^{(k)}(u, v).$$

In the following, we will prove the concentration bound for each of the four terms.

(i). Consider the first term. Note that $\chi_{ij}^{(1)}(u, v)$ is a U-statistics. For each integer $q = 2, 3, \ldots$, by the fact that e_{ijt}'s are Gaussian and similar techniques used to prove (B.12), we have

$$E \left| e_{ijt} e_{ijt'} K_h(U_{ijt} - u) K_h(U_{ijt'} - v) \right|^q \leqslant \frac{q!}{2} (c_5 h^{-2})(c_5 h^{-2})^{q-2}.$$

It then follows from Lemma 3 that for each $\delta > 0$,

$$\mathrm{pr} \left[\left| n^{-1} \sum_{i=1}^{n} w_{ij} \left\{ \chi_{ij}^{(1)}(u, v) - E \chi_{ij}^{(1)}(u, v) \right\} \right| \geqslant \delta \right] \leqslant 2 \exp \left(-\frac{c_6 n T h^2 \delta^2}{1 + \delta} \right). \tag{B.24}$$

(ii). Consider the second term. To derive its concentration bound, we implement a truncation technique. Let $a_{ijl} = \omega_{jl}^{-1/2} \xi_{ijl} \sim N(0, 1)$, $a_{ijl}^{(1)} = a_{ijl} I(|a_{ijl}| \leqslant n^\theta)$ and $a_{ijl}^{(2)} = a_{ijl} I(|a_{ijl}| > n^\theta)$ for some $\theta > 0$. Then we have $\chi_{ij}^{(2)}(u, v) = \chi_{ij}^{(2,1)}(u, v) + \chi_{ij}^{(2,2)}(u, v)$, where

$$\chi_{ij}^{(2,m)}(u, v) = \sum_{l=1}^{d} \omega_{jl}^{1/2} a_{ijl}^{(k)} \sum_{1 \leqslant t \neq t' \leqslant T} e_{ijt'} \phi_{jl}(U_{ijt}) K_h(U_{ijt} - u) K_h(U_{ijt'} - v), \quad m = 1, 2.$$

Note that, similar to the proof for Lemma 3, $w_{ij} \sum_{t \neq t'}^{T} e_{ijt'} \phi_{j\ell}(U_{ijt}) K_h(U_{ijt} - u) K_h(U_{ijt'} - v)$ can be expressed as an average of $T!$ terms, each of which is itself an average of $[T/2]$ independent random variables. For each integer $q = 2, 3, \ldots$, by the facts e_{ijt}'s are Gaussian, $T \geqslant 2\sqrt{T}$ for $T \geqslant 4$, Condition 3(ii) and techniques used to prove (B.12), we have

$$E \left| w_{ij} \sum_{1 \leqslant t \neq t' \leqslant T} e_{ijt'} \phi_{jl}(U_{ijt}) K_h(U_{ijt} - u) K_h(U_{ijt'} - v) \right|^q \leqslant q! c_7^q T^{-q/2} h^{-2(q-1)}.$$

Combing the above results and some specific calculations yield that

$$E\big|w_{ij}\chi_{ij}^{(2,1)}(u,v)\big|^q \leqslant \left(\sum_{\ell=1}^{d}\omega_{jl}^{1/2}\right)^q n^{\theta(q-2)}c_8^q T^{-q/2}h^{-2(q-1)}$$

$$\leqslant \frac{1}{2}q!(c_9^2 h^{-2}T^{-1})(c_9 n^\theta h^{-2}T^{-1/2})^{q-2}. \tag{B.25}$$

Applying the Bernstein inequality, we obtain that for each $\delta > 0$,

$$\mathrm{pr}\left\{\left|n^{-1}\sum_{i=1}^{n}w_{ij}\chi_{ij}^{(2,1)}(u,v)\right| \geqslant \delta\right\} \leqslant 2\exp\left(-\frac{c_{10}nh^2 T\delta^2}{1+n^\theta T^{1/2}\delta}\right). \tag{B.26}$$

Note that $\big\{\sup_{1\leqslant i\leqslant n,1\leqslant l\leqslant d}|a_{ijl}| \leqslant n^\theta\big\} \subset \{\chi_{ij}^{(2,2)}(u,v)=0\}$. Applying the Bernstein inequality for a_{ijl}'s and the union bound of probability, we obtain that for each $\delta > 0$,

$$\mathrm{pr}\left\{\left|n^{-1}\sum_{i=1}^{n}w_{ij}\chi_{ij}^{(2,2)}(u,v)\right| \neq 0\right\} \leqslant 2nd\exp\left(-c_{11}n^{2\theta}\right). \tag{B.27}$$

(iii) Consider the third term. Similar to the development in (ii), we can write $\chi_{ij}^{(3)}(u,v) = \chi_{ij}^{(3,1)}(u,v) + \chi_{ij}^{(3,2)}(u,v)$, where

$$\chi_{ij}^{(3,k)}(u,v) = \sum_{l=1}^{d}\omega_{jl}^{1/2}a_{ijl}^{(k)}\sum_{1\leqslant t\neq t'\leqslant T}e_{ijt'}\phi_{jl}(U_{ijt})K_h(U_{ijt}-v)K_h(U_{ijt'}-u),\quad k=1,2.$$

Following the same developments in proving (B.28) and (B.27), we can obtain that for each $\delta > 0$,

$$\mathrm{pr}\left\{\left|n^{-1}\sum_{i=1}^{n}w_{ij}\chi_{ij}^{(3,1)}(u,v)\right| \geqslant \delta\right\} \leqslant 2\exp\left(-\frac{c_{12}nTh^2\delta^2}{1+n^\theta T^{1/2}\delta}\right). \tag{B.28}$$

and

$$\mathrm{pr}\left\{\left|n^{-1}\sum_{i=1}^{n}w_{ij}\chi_{ij}^{(3,2)}(u,v)\right| \neq 0\right\} \leqslant 2nd\exp\left(-c_{13}n^{2\theta}\right). \tag{B.29}$$

(iv). Consider the fourth term. For any (i,j,l,l') for $i=1,\ldots,n$, $j \in V$, $l,l'=1,\ldots,d$, denote

$$\psi_{ij,ll'}(u,v) = \sum_{1\leqslant t\neq t'\leqslant T}\phi_{jl}(U_{ijt})\phi_{jl'}(U_{ijt'})K_h(U_{ijt}-u)K_h(U_{ijt'}-v).$$

We can further decompose $\chi_{ij}^{(4)}(u,v)$ as

$$\chi_{ij}^{(4)}(u,v) = \sum_{l,l'=1}^{d}\xi_{ijl}\xi_{ijl'}E\psi_{ij,ll'}(u,v) + \sum_{l,l'=1}^{d}\xi_{ijl}\xi_{ijl'}\big\{\psi_{ij,ll'}(u,v)-E\psi_{ij,ll'}(u,v)\big\}$$

$$= \chi_{ij}^{(5)}(u,v) + \chi_{ij}^{(6)}(u,v).$$

Applying the fact that $\sum_l \xi_{ijl}$ is Gaussian, Conditions 3(ii), 8, techniques used to prove (B.12) and some specific calculations yields that for $q=2,3,\ldots,$

$$\left|E\big\{w_{ij}\chi_{ij}^{(5)}(u,v)\big\}^q\right| \leqslant \left|E\left(\sum_l \xi_{ijl}\right)^{2q}\right|\sup_{i,j,l,l'}\left|E\big\{w_{ij}\psi_{ll'}(u,v)\big\}\right|^q \leqslant c_{14}^q q!.$$

Applying the Bernstein inequality, we obtain that for each $\delta > 0$,

$$\mathrm{pr}\left[\left|n^{-1}\sum_{i=1}^{n}w_{ij}\left\{\chi_{ij}^{(5)}(u,v)-E\chi_{ij}^{(5)}(u,v)\right\}\right|\geqslant \delta\right]\leqslant 2\exp\left(-\frac{c_{15}n\delta^2}{1+\delta}\right). \qquad \text{(B.30)}$$

Now we turn to the term $\sum_{i=1}^{n} n^{-1}w_{ij}\chi_{ij}^{(6)}(u,v)$. To derive its concentration bound, we use the truncation technique again. Observe that $a_{ijl}a_{ijl'} = a_{ijl}^{(1)}a_{ijl'}^{(1)} + a_{ijl}^{(2)}a_{ijl'}^{(1)} + a_{ijl}^{(1)}a_{ijl'}^{(2)}$. Similar to the procedure in (ii), we can write $\chi_{ij}^{(6)}(u,v) = \chi_{ij}^{(6,1)}(u,v) + \chi_{ij}^{(6,2)}(u,v)$, where

$$\chi_{ij}^{(6,1)}(u,v) = \sum_{l,l'=1}^{d}\omega_{jl}^{1/2}\omega_{jl'}^{1/2}a_{ijl'}^{(1)}a_{ijl'}^{(1)}\left\{\psi_{ij,ll'}(u,v)-E\psi_{ij,ll'}(u,v)\right\}.$$

It follows from the proof of Lemma 3 that $w_{ij}\psi_{ij,\ell\ell'}(u,v)$ can be expressed as an average of $T!$ terms, each of which is itself an average of $[T/2]$ independent random variables. Following the similar developments to prove (B.25), we can show that

$$E\left|w_{ij}a_{ijl}^{(1)}a_{ijl'}^{(1)}\left\{\psi_{ij,ll'}(u,v)-E\psi_{ij,ll'}(u,v)\right\}\right|^{q}\leqslant \frac{1}{2}q!(c_{16}^2 T^{-1}h^{-2})(c_{16}n^{\theta}T^{-1/2}h^{-2})^{q-2}.$$

Applying the Bernstein inequality again, we obtain that for each $\delta > 0$,

$$\mathrm{pr}\left[\left|n^{-1}\sum_{i=1}^{n}w_{ij}\chi_{ij}^{(6,1)}(u,v)\right|\geqslant \delta\right]\leqslant 2\exp\left(-\frac{c_{17}nTh^2\delta^2}{1+n^{\theta}T^{1/2}\delta}\right).$$

On the other hand, we can also show that

$$\mathrm{pr}\left\{\left|n^{-1}\sum_{i=1}^{n}w_{ij}\chi_{ij}^{(6,2)}(u,v)\right|\neq 0\right\}\leqslant 2nd\exp\left(-c_{18}n^{2\theta}\right). \qquad \text{(B.31)}$$

Choosing $n^{\theta} = (nT^{1/2}h^2)^{1/3}$ and combing the concentration inequalities in (B.24), (B.26)–(B.31), we obtain that there exist two positive constants c_{19} and c_{20} such that for any $\delta \in (0,1]$,

$$\mathrm{pr}\left\{\left|\widehat{Z}_{j1}(u,v)-E\{\widehat{Z}_{j1}(u,v)\}\right|\geqslant \delta\right\}\leqslant c_{20}\exp\left\{-c_{19}\min(n,n^{2/3-\epsilon}T^{1/3}h^{4/3})\delta^2\right\}, \qquad \text{(B.32)}$$

where ϵ is obtained from $\log(n)\leqslant n^{\epsilon}$ for any small constant $\epsilon > 0$. We can apply the same procedure and prove that the concentration bound in (B.22) also holds for $m = 2,3$.

Next, given the pointwise concentration bound in (B.23), it follows from the same procedure to prove (B.15) that we can derive the L_2 concentration bound for $\|\widehat{C}_{jj} - \widetilde{C}_{jj}\|_{\mathcal{S}}$ in (B.17). Finally, applying the partition technique and then following the similar developments to prove Lemma 2 and (B.4), we can derive the uniform concentration bound for $\sup_{(u,v)\in\mathcal{U}^2}\left|\widehat{C}_{jj}(u,v)-\widetilde{C}_{jj}(u,v)\right|$ in (B.18). We complete the proof for this lemma. □

In Lemma 5, it is worthy noting that we assume that $d_j < \infty$ for each $j \in V$. We leave the development of concentration results for $\widehat{C}_{jj}(u,v)$ under the infinite dimensional setting as our future work.

B·5.   *Proof of Theorem* 1

To prove Theorem 1, we will use Lemmas 6-25 with proofs as follows.

LEMMA 6. $\{\widehat{\Theta}_1(u)\}$ *be the solution to* (6) *and* $\widehat{B}(u) = \{\widehat{\beta}_1(u),\ldots,\widehat{\beta}_p(u)\}$ *with* $\widehat{\beta}_j(u)$*'s being solutions to* (7), *then we have* $\{\widehat{\Theta}_1(u)\} = \{\widehat{B}(u)\}$.

Proof. We can follow exactly the same steps in Lemma 1 of Cai et al. (2011), thus the proof here is omitted.□

LEMMA 7. *Suppose that Conditions 2 and 6 hold. Then there exists some positive constants c_1, c_2, c_3 such that for any δ with $h^2 \ll \delta \leqslant 1$ and each $j \in V$, we have*

$$\mathrm{pr}\left(\|\widehat{C}_{jj} - C_{jj}\|_{\mathcal{S}} \geqslant \delta\right) \leqslant c_2 \exp(-c_1 n^{2\gamma}\delta^2), \tag{B.33}$$

$$\mathrm{pr}\left\{\sup_{(u,v)\in\mathcal{U}^2} |\widehat{C}_{jj}(u,v) - C_{jj}(u,v)| \geqslant \delta\right\} \leqslant c_2 n^{c_3} \exp(-c_1 n^{2\gamma}\delta^2). \tag{B.34}$$

Proof. It follows from the proofs for Theorems 3.2 and 4.2 of Zhang & Wang (2016) that, under Condition 6, $\|\widetilde{C}_{jj} - C_{jj}\|_{\mathcal{S}} = O(h^2)$. By the triangular inequality with

$$\|\widehat{C}_{jj} - C_{jj}\|_{\mathcal{S}} \leqslant \|\widehat{C}_{jj} - \widetilde{C}_{jj}\|_{\mathcal{S}} + \|\widetilde{C}_{jj} - C_{jj}\|_{\mathcal{S}},$$

$\delta \gg h^2$ and (12) in Condition 2, we have $P\left(\|\widehat{C}_{jj} - C_{jj}\|_{\mathcal{S}} \geqslant \delta\right) \leqslant P\left(\|\widehat{C}_{jj} - \widetilde{C}_{jj}\|_{\mathcal{S}} \geqslant \delta/2\right) \leqslant c_2 \exp(-c_4 n^{2\gamma}\delta^2/4)$ with $c_1 = c_4/4$. Finally, it follows from (13), Condition 6, Theorem 5.2 of of Zhang & Wang (2016) with $\sup_{(u,v)\in\mathcal{U}^2}|\widetilde{C}_{jj}(u,v) - C_{jj}(u,v)| = O(h^2)$ and similar developments that (B.34) follows. □

LEMMA 8. *Suppose that Lemma 7 holds, then we have*

$$\max_{j\in V}\|\widehat{C}_{jj} - C_{jj}\|_{\mathcal{S}} = O_p\left(\sqrt{\frac{\log p}{n^{2\gamma}}}\right), \tag{B.35}$$

$$\max_{j\in V}\sup_{(u,v)\in\mathcal{U}^2}|\widehat{C}_{jj}(u,v) - C_{jj}(u,v)| = O_p\left(\sqrt{\frac{\log p}{n^{2\gamma}}}\right), \tag{B.36}$$

Proof. Applying the union bound of probability on (B.33),

$$\max_{j\in V}\mathrm{pr}\left(\|\widehat{C}_{jj} - C_{jj}\|_{\mathcal{S}} \geqslant \delta\right) \leqslant p c_2 \exp(-c_1 n^{2\gamma}\delta^2),$$

then (B.35) follows. Similarly, applying the union bound of probability on (B.34) under $p \gtrsim n$, (B.36) follows. □

LEMMA 9. *Suppose that Lemma 8 holds, then we have* $\max_{j\in v} |\widehat{\omega}_{jl} - \omega_{jl}| = O_p\left(\sqrt{\frac{\log p}{n^{2\gamma}}}\right)$.

Proof. By (4.43) of Bosq (2000) and Lemma 8, we have $\max_{j\in V}\sup_{l\geqslant 1}|\widehat{\omega}_{jl} - \omega_{jl}| \leqslant \max_j \|\widehat{C}_{jj} - C_{jj}\|_{\mathcal{S}} = O_p\left(\sqrt{\frac{\log p}{n^{2\gamma}}}\right)$, which completes the our proof. □

LEMMA 10. *Denote $\breve{\phi}_{jl} = sign\langle\widehat{\phi}_{jl}, \phi_{jl}\rangle\phi_{jl}$. Then*

$$\|\widehat{\phi}_{jl} - \breve{\phi}_{jl}\| \leqslant d_{jl}\|\widehat{C}_{jj} - C_{jj}\|_{\mathcal{S}},$$

where $d_{jl} = 2\sqrt{2}\max\left(\omega_{j(l-1)} - \omega_{jl}\right)^{-1}, \left(\omega_{jl} - \omega_{j(l+1)}\right)^{-1}$ if $l \geqslant 2$ and $d_{j1} = 2\sqrt{2}(\omega_{j1} - \omega_{j2})^{-1}$. Moreover, suppose that Lemma 8 and Condition 3 hold, then for each $l = 1, 2, \ldots$, we have $\max_{j\in V}\|\widehat{\phi}_{jl} - \phi_{jl}\| = O_p\left(l^{\beta+1}\sqrt{\frac{\log p}{n^{2\gamma}}}\right)$.

Proof. The first part can be found in Lemma 4.3 of Bosq (2000). By Condition 3, $\omega_{jl} \asymp l^{-\beta}$, $d_{jl}\omega_{jl} = O(l)$ and hence $d_{jl} \leqslant c_1 l^{1+\beta}$, which completes our proof. □

To simplify notation, we will use ϕ_{jl} rather than $\breve{\phi}$ in our proofs.

LEMMA 11. *Suppose that Condition 3 holds, then $\max_{(j,k)\in V^2}\sup_{(u,v)\in\mathcal{U}^2}|C_{jk}(u,v)| = O(1)$.*

Proof. Observe that $X_j(u)X_k(v) = \sum_{l=1}^{\infty}\sum_{m=1}^{\infty}\xi_{jl}\xi_{km}\phi_{jl}(u)\phi_{km}(v)$. By the Cauchy-Schwartz inequality, $E(\xi_{jl}\xi_{jl'}) = 0$ for $l \neq l'$ and Condition 3, we have

$$
\begin{aligned}
\max_{(j,k)\in V^2}\sup_{(u,v)\in\mathcal{U}^2}|C_{jk}(u,v)|^2 &\leqslant \max_{(j,k)\in V^2}\Big\{\sum_{l=1}^{\infty}\sum_{m=1}^{\infty}E(\xi_{jl}\xi_{km})\sup_{\mathcal{U}}\phi_{jl}(u)\sup_{\mathcal{U}}\phi_{km}(v)\Big\}^2 \\
&\leqslant \max_{j\in V}\sum_{l=1}^{\infty}\big\{E(\xi_{jl}^2)\sup_{\mathcal{U}}\phi_{jl}(u)^2\big\}\max_{k\in V}\sum_{m=1}^{\infty}\big\{E(\xi_{km}^2)\sup_{\mathcal{U}}\phi_{km}(v)^2\big\} \\
&= O\Big(\max_{j\in V}\sum_{l=1}^{\infty}\omega_{jl}\max_{k\in V}\sum_{m=1}^{\infty}\omega_{km}\Big) = O(1).
\end{aligned}
$$

This completes our proof for this lemma. □

LEMMA 12. *Suppose that Lemma 7 and Condition 3 hold, then for each $l = 1, 2, \ldots$,*

$$
\max_{j\in V}\sup_{u\in\mathcal{U}}|\widehat{\phi}_{jl}(u) - \phi_{jl}(u)| = O_p\Big(l^{2\beta+1}\sqrt{\frac{\log p}{n^{2\gamma}}}\Big).
$$

Proof. For each (j, l), it follows from eigen-decompositions $\omega_{jl}\phi_{jl}(u) = \int_{\mathcal{U}}C_{jj}(u,v)\phi_{jl}(v)dv$ and $\widehat{\omega}_{jl}\widehat{\phi}_{jl}(u) = \int_{\mathcal{U}}\widehat{C}_{jj}(u,v)\widehat{\phi}_{jl}(v)dv$ that we can decompose $\widehat{\phi}_{jl}(u) - \phi_{jl}(u)$ as

$$
\begin{aligned}
&= -\omega_{jl}^{-1}\big(\widehat{\omega}_{jl} - \omega_{jl}\big)\phi_{jl}(u) - \omega_{jl}^{-1}\big(\widehat{\omega}_{jl} - \omega_{jl}\big)\big(\widehat{\phi}_{jl}(u) - \phi_{jl}(u)\big) \\
&\quad + \omega_{jl}^{-1}\int\big\{\widehat{C}_{jj}(u,v) - C_{jj}(u,v)\big\}\phi_{jl}(v)dv + \omega_{jl}^{-1}\int\big\{\widehat{C}_{jj}(u,v) - C_{jj}(u,v)\big\}\big(\widehat{\phi}_{jl}(v) - \phi_{jl}(v)\big)dv \\
&\quad + \omega_{jl}^{-1}\int C_{jj}(u,v)\big\{\widehat{\phi}_{jl}(v) - \phi_{jl}(v)\big\}dv \\
&= S_1(u) + S_2(u) + S_3(u) + S_4(u) + S_5(u).
\end{aligned}
$$

(i) It follows from Condition 3 and Lemma 9 that

$$
\max_j\sup_u|S_1(u)| \leqslant \max_j\omega_{jl}^{-1}\max_j|\widehat{\omega}_{jl} - \omega_{jl}|\sup_u|\phi_{jl}(u)| = O_p\Big(l^{\beta}\sqrt{\frac{\log p}{n^{2\gamma}}}\Big).
$$

(ii) It follows from Condition 3 and Lemma 8 that

$$
\max_j\sup_u|S_3(u)| \leqslant \max_j\omega_{jl}^{-1}\max_j\sup_{u,v}|\widehat{C}_{jj}(u,v) - C_{jj}(u,v)|\max_j\int|\phi_{jl}(v)|dv = O_p\Big(l^{\beta}\sqrt{\frac{\log p}{n^{2\gamma}}}\Big).
$$

(iii) It follows from Condition 3, the Cauchy–Schwartz inequality and Lemmas 10–11 that

$$
\max_j\sup_u|S_5(u)| \leqslant \max_j\omega_{jl}^{-1}\max_j\sup_u\|C_{jj}(u,\cdot)\|\cdot\max_j\|\widehat{\phi}_{jl} - \phi_{jl}\| = O_p\Big(l^{\beta}l^{\beta+1}\sqrt{\frac{\log p}{n^{2\gamma}}}\Big).
$$

Since the above three terms can dominate the others, we combine convergence rates in (i), (ii), (iii) and obtain the rate as stated in the lemma. □

LEMMA 13. *Suppose that Lemma 8 and Condition 3 hold. Then for each $i = 1, \ldots, n$, we have*

$$
\max_{j\in V}\|\widehat{\zeta}_{ijl} - \zeta_{ijl}\| = O_p\Big(\max_j T_{ij}^{1/2}l\sqrt{\frac{\log p}{n^{2\gamma}}}\Big), \tag{B.37}
$$

$$
\max_{j\in V}\|\widehat{\Sigma}_{Y_{ij}}^{-1} - \Sigma_{Y_{ij}}^{-1}\| = O_p\Big(\max_j T_{ij}\sqrt{\frac{\log p}{n^{2\gamma}}}\Big). \tag{B.38}
$$

Proof. We have $\max_j ||\widehat{\zeta}_{ijl} - \zeta_{ijl}|| \leqslant \max_j T_{ij}^{1/2} \sup_t |\widehat{\zeta}_{ijlt} - \zeta_{ijlt}|$ and

$$
\begin{aligned}
&\max_j \sup_t |\widehat{\zeta}_{ijlt} - \zeta_{ijlt}| \\
&= \max_j \sup_t |\int \widehat{C}_{jj}(U_{ijt}, v)\widehat{\phi}_{jl}(v)dv - \int C_{jj}(U_{ijt}, v)\phi_{jl}(v)dv| \\
&= \max_j \sup_t |\widehat{\omega}_{jl}\widehat{\phi}_{jl}(U_{ijt}) - \omega_{jl}\phi_{jl}(U_{ijt})| \\
&\leqslant \max_j \sup_{u\in\mathcal{U}} |\int (\widehat{\omega}_{jl} - \omega_{jl})\widehat{\phi}_{jl}(u)| + \max_j \sup_{u\in\mathcal{U}} |\int \omega_{jl}(\widehat{\phi}_{jl}(u) - \phi_{jl}(u))| \\
&\leqslant \max_j |\widehat{\omega}_{jl} - \omega_{jl}| \max_j \sup_u (|\phi_{jl}(u)| + |\widehat{\phi}_{jl}(u) - \phi_{jl}(u)|) + \max_j \omega_{jl}\sup_u ||\widehat{\phi}_{jl}(u) - \phi_{jl}(u)|| \\
&= O_p\Big(\sqrt{\frac{\log p}{n^{2\gamma}}} + l^{-\beta}l^{\beta+1}\sqrt{\frac{\log p}{n^{2\gamma}}}\Big),
\end{aligned}
$$

where the last line follows from Condition 3 and Lemmas 9–10. This completes our proof for (B.37).

From the definition of $\Sigma_{Y_{ij}} = \Sigma_{X_{ij}} + \sigma^2 I_{T_{ij}}$, where $X_{ij} = \{X_{ij}(U_{ijT_1}), \ldots, X_{ij}(U_{ijT_{ij}})\}^{\mathrm{T}}$, we have $\|\Sigma_{Y_{ij}}^{-1}\| \leqslant \sigma^{-2}$. Applying Lemma 1 of Dai et al. (2018), we have $\|\widehat{\Sigma}_{Y_{ij}}^{-1} - \Sigma_{Y_{ij}}^{-1}\| \leqslant c_1\sigma^{-4}\|\widehat{\Sigma}_{Y_{ij}} - \Sigma_{Y_{ij}}\| \leqslant c_2\sigma^{-4}T_{ij}\sup_{t,t'} |(\widehat{\Sigma}_{Y_{ij}})_{tt'} - (\Sigma_{Y_{ij}})_{tt'}|$. Moreover, $\sup_{t,t'} |(\widehat{\Sigma}_{Y_{ij}})_{tt'} - (\Sigma_{Y_{ij}})_{tt'}| \leqslant |\widehat{\sigma}^2 - \sigma^2| + \sup_{u,v} |\widehat{C}_{jj}(u,v) - C_{jj}(u,v)|$, where the first term is dominated by the second term, see Corollary 1 of Yao et al. (2015) for details. We take $\max_{j\in V}$ on both sides, apply (B.36) in Lemma 8 and hence can obtain the convergence rate in (B.38). □

LEMMA 14. *Suppose that Conditions* 1, 3 *and* 5 *hold. First consider the dense measurement design. For each* $l, m = 1, \ldots, M$, *we have*

$$
\max_{(j,k)\in V^2} \Big|\frac{1}{n}\sum_{i=1}^n (\widehat{\xi}_{ijl}\widehat{\xi}_{ikm} - \widetilde{\xi}_{ijl}\widetilde{\xi}_{ikm})\Big| = O_p\Big\{T^2(lm^{-\beta} + l^{-\beta}m)\sqrt{\frac{\log p}{n^{2\gamma}}} + T^3 l^{-\beta}m^{-\beta}\sqrt{\frac{\log p}{n^{2\gamma}}}\Big\}.
$$
(B.39)

Furthermore, for sparse measurement design with time points U_{ijt}*'s possibly depending on* i *and* $T \leqslant$ $T_0 < \infty$, *we have*

$$
\max_{(j,k)\in V^2} \Big|\frac{1}{n}\sum_{i=1}^n (\widehat{\xi}_{ijl}\widehat{\xi}_{ikm} - \widetilde{\xi}_{ijl}\widetilde{\xi}_{ikm})\Big| = O_p\Big\{(lm^{-\beta} + l^{-\beta}m)\sqrt{\frac{\log p}{n^{2\gamma}}}\Big\}.
$$
(B.40)

Proof. We write $\widehat{\xi}_{ijl}\widehat{\xi}_{ikm} - \widetilde{\xi}_{ijl}\widetilde{\xi}_{ikm} = \widehat{\zeta}_{ijl}^{\mathrm{T}}\widehat{\Sigma}_{Y_{ij}}^{-1}Y_{ij}\widehat{\zeta}_{ikm}^{\mathrm{T}}\widehat{\Sigma}_{Y_{ik}}^{-1}Y_{ik} - \zeta_{ijl}^{\mathrm{T}}\Sigma_{Y_{ij}}^{-1}Y_{ij}\zeta_{ikm}^{\mathrm{T}}\Sigma_{Y_{ik}}^{-1}Y_{ik}$, which can further be decomposed as

$$
\begin{aligned}
&= (\widehat{\zeta}_{ijl} - \zeta_{ijl})^{\mathrm{T}}\Sigma_{Y_{ij}}^{-1}Y_{ij}Y_{ik}^{\mathrm{T}}\Sigma_{Y_{ik}}^{-1}\zeta_{ikm} + (\widehat{\zeta}_{ijl} - \zeta_{ijl})^{\mathrm{T}}\Sigma_{Y_{ij}}^{-1}Y_{ij}Y_{ik}^{\mathrm{T}}(\widehat{\Sigma}_{Y_{ik}}^{-1} - \Sigma_{Y_{ik}}^{-1})\zeta_{ikm} \\
&\quad + (\widehat{\zeta}_{ijl} - \zeta_{ijl})^{\mathrm{T}}\Sigma_{Y_{ij}}^{-1}Y_{ij}Y_{ik}^{\mathrm{T}}\Sigma_{Y_{ik}}^{-1}(\widehat{\zeta}_{ikm} - \zeta_{ikm}) \\
&\quad + (\widehat{\zeta}_{ijl} - \zeta_{ijl})^{\mathrm{T}}\Sigma_{Y_{ij}}^{-1}Y_{ij}Y_{ik}^{\mathrm{T}}(\widehat{\Sigma}_{Y_{ik}}^{-1} - \Sigma_{Y_{ik}}^{-1})(\widehat{\zeta}_{ikm} - \zeta_{ikm}) \\
&\quad + (\widehat{\zeta}_{ijl} - \zeta_{ijl})^{\mathrm{T}}(\widehat{\Sigma}_{Y_{ij}}^{-1} - \Sigma_{Y_{ij}}^{-1})Y_{ij}Y_{ik}^{\mathrm{T}}\Sigma_{Y_{ik}}^{-1}\zeta_{ikm} \\
&\quad + (\widehat{\zeta}_{ijl} - \zeta_{ijl})^{\mathrm{T}}(\widehat{\Sigma}_{Y_{ij}}^{-1} - \Sigma_{Y_{ij}}^{-1})Y_{ij}Y_{ik}^{\mathrm{T}}(\widehat{\Sigma}_{Y_{ik}}^{-1} - \Sigma_{Y_{ik}}^{-1})\zeta_{ikm} \\
&\quad + (\widehat{\zeta}_{ijl} - \zeta_{ijl})^{\mathrm{T}}(\widehat{\Sigma}_{Y_{ij}}^{-1} - \Sigma_{Y_{ij}}^{-1})Y_{ij}Y_{ik}^{\mathrm{T}}\Sigma_{Y_{ik}}^{-1}(\widehat{\zeta}_{ikm} - \zeta_{ikm}) \\
&\quad + (\widehat{\zeta}_{ijl} - \zeta_{ijl})^{\mathrm{T}}(\widehat{\Sigma}_{Y_{ij}}^{-1} - \Sigma_{Y_{ij}}^{-1})Y_{ij}Y_{ik}^{\mathrm{T}}(\widehat{\Sigma}_{Y_{ik}}^{-1} - \Sigma_{Y_{ik}}^{-1})(\widehat{\zeta}_{ikm} - \zeta_{ikm}) \\
&\quad + \zeta_{ijl}^{\mathrm{T}}(\widehat{\Sigma}_{Y_{ij}}^{-1} - \Sigma_{Y_{ij}}^{-1})Y_{ij}Y_{ik}^{\mathrm{T}}\Sigma_{Y_{ik}}^{-1}\zeta_{ikm} + \zeta_{ijl}^{\mathrm{T}}(\widehat{\Sigma}_{Y_{ij}}^{-1} - \Sigma_{Y_{ij}}^{-1})Y_{ij}Y_{ik}^{\mathrm{T}}(\widehat{\Sigma}_{Y_{ik}}^{-1} - \Sigma_{Y_{ik}}^{-1})\zeta_{ikm} \\
&\quad + \zeta_{ijl}^{\mathrm{T}}(\widehat{\Sigma}_{Y_{ij}}^{-1} - \Sigma_{Y_{ij}}^{-1})Y_{ij}Y_{ik}^{\mathrm{T}}\Sigma_{Y_{ik}}^{-1}(\widehat{\zeta}_{ikm} - \zeta_{ikm}) \\
&\quad + \zeta_{ijl}^{\mathrm{T}}(\widehat{\Sigma}_{Y_{ij}}^{-1} - \Sigma_{Y_{ij}}^{-1})Y_{ij}Y_{ik}^{\mathrm{T}}(\widehat{\Sigma}_{Y_{ik}}^{-1} - \Sigma_{Y_{ik}}^{-1})(\widehat{\zeta}_{ikm} - \zeta_{ikm}) \\
&\quad + \zeta_{ijl}\Sigma_{Y_{ij}}^{-1}Y_{ij}Y_{ik}^{\mathrm{T}}\Sigma_{Y_{ik}}^{-1}(\widehat{\zeta}_{ikm} - \zeta_{ikm}) + \zeta_{ijl}^{\mathrm{T}}\Sigma_{Y_{ij}}^{-1}Y_{ij}Y_{ik}^{\mathrm{T}}(\widehat{\Sigma}_{Y_{ik}}^{-1} - \Sigma_{Y_{ik}}^{-1})(\widehat{\zeta}_{ikm} - \zeta_{ikm}) \\
&\quad + \zeta_{ijl}^{\mathrm{T}}\Sigma_{Y_{ij}}^{-1}Y_{ij}Y_{ik}^{\mathrm{T}}(\widehat{\Sigma}_{Y_{ik}}^{-1} - \Sigma_{Y_{ik}}^{-1})\zeta_{ikm} \\
&= I_1 + I_2 + \cdots + I_{15}.
\end{aligned}
$$

We first list several results that can be used in our proof. Note that the time points U_{ijt}'s do not depend on i and $T_{ij} = T \to \infty$. (i) $\max_j \|\zeta_{ijl}\| \leqslant T^{1/2}\max_j\sup_t|\zeta_{ijlt}| = T^{1/2}\max_j\sup_t|\int C_{jj}(U_{ijt}, v)\phi_{jl}(v)dv| \leqslant T^{1/2}\max_j\sup_{u \in \mathcal{U}}|\int C_{jj}(u, v)\phi_{jl}(v)dv| = T^{1/2}\max_j\sup_{\mathcal{U}}|\int \omega_{jl}\phi_{jl}(u)| = O(T^{1/2}l^{-\beta})$, where the last equality follows from Condition 3; (ii) From the definition of $\Sigma_{Y_{ij}}$, we have $\|\Sigma_{Y_{ij}}^{-1}\| \leqslant \sigma^{-2}$; (iii) By Lemma 11, $\max_{j,k}\|n^{-1}\sum_{i=1}^n Y_{ij}Y_{ik}^{\mathrm{T}}\| \leqslant T\max_{j,k}\sup_{t,t'}|(n^{-1}\sum_{i=1}^n Y_{ij}Y_{ik}^{\mathrm{T}})_{tt'}| \leqslant \max_{j,k}T\sup_{(u,v)\in\mathcal{U}^2}|\widehat{C}_{jk}(u, v)| = O_p(T)$.

Applying the above results in (i), (ii), (iii) and (B.37), (B.38) in Lemma 13, we have

$$
\max_{j,k}\|n^{-1}\sum_i I_1\| \leqslant \max_j\|\widehat{\zeta}_{ijl} - \zeta_{ijl}\| \max_j\|\Sigma_{Y_{ij}}^{-1}\| \max_{j,k}\|n^{-1}\sum_i Y_{ij}Y_{ik}\| \max_k\|\zeta_{ikm}\| \max_k\|\Sigma_{Y_{ik}}^{-1}\|
$$
$$
= O_p\Big(T^2 l m^{-\beta}\sqrt{\frac{\log p}{n^{2\gamma}}}\Big),
$$

$$
\max_{j,k}|n^{-1}\sum_i I_9| \leqslant \max_j\|\zeta_{ijl}\| \max_j\|\widehat{\Sigma}_{Y_{ij}}^{-1} - \Sigma_{Y_{ij}}^{-1}\| \max_{j,k}\|n^{-1}\sum_i Y_{ij}Y_{ik}\| \max_k\|\zeta_{ikm}\| \max_k\|\Sigma_{Y_{ik}}^{-1}\|
$$
$$
= O_p\Big(T^3 l^{-\beta} m^{-\beta}\sqrt{\frac{\log p}{n^{2\gamma}}}\Big),
$$

$$
\max_{j,k}\|n^{-1}\sum_i I_{13}\| \leqslant \max_j\|\zeta_{ijl}\| \max_j\|\Sigma_{Y_{ij}}^{-1}\| \max_{j,k}\|n^{-1}\sum_i Y_{ij}Y_{ik}\| \max_k\|\widehat{\zeta}_{ikm} - \zeta_{ikm}\| \max_k\|\Sigma_{Y_{ik}}^{-1}\|
$$
$$
= O_p\Big(T^2 l^{-\beta} m\sqrt{\frac{\log p}{n^{2\gamma}}}\Big),
$$

$$
\max_{j,k}\|n^{-1}\sum_i I_{15}\| \leqslant \max_j\|\zeta_{ijl}\| \max_j\|\Sigma_{Y_{ij}}^{-1}\| \max_{j,k}\|n^{-1}\sum_i Y_{ij}Y_{ik}\| \max_k\|\zeta_{ikm}\| \max_k\|\widehat{\Sigma}_{Y_{ik}}^{-1} - \Sigma_{Y_{ik}}^{-1}\|
$$
$$
= O_p\Big(T^3 l^{-\beta} m^{-\beta}\sqrt{\frac{\log p}{n^{2\gamma}}}\Big).
$$

Since these four terms can dominate the other $\max_{j,k} \|n^{-1} \sum_i I_l\|$ terms, we combine the above convergence rates and obtain (B.39). For the sparse case with $T_{ij} \leqslant T_0 < \infty$, following the same developments to derive (B.39) without requiring U_{ijt}'s depending on i, we can obtain (B.40). We complete our proof for this lemma. $\square$

LEMMA 15. *Suppose that conditions for the dense case in Lemma* 14 *hold, then for each* $l, m = 1, \ldots, M$, *we have*

$$\max_{(j,k) \in V^2} \left| \frac{1}{n} \sum_{i=1}^{n} (\widetilde{\xi}_{ijl} \widetilde{\xi}_{ikm} - \xi_{ijl} \xi_{ikm}) \right| = O_p \{ (\log p)^{1/2} T^{-1/2} (l^{-\beta/2} + m^{-\beta/2}) \}.$$

Proof. We can write

$$
\begin{aligned}
\widetilde{\xi}_{ijl} \widetilde{\xi}_{ikm} - \xi_{ijl} \xi_{ikm} &= \{ (\widetilde{\xi}_{ijl} - \xi_{ijl}) \widetilde{\xi}_{ikm} \} - E\{ (\widetilde{\xi}_{ijl} - \xi_{ijl}) \widetilde{\xi}_{ikm} \} + E\{ (\widetilde{\xi}_{ijl} - \xi_{ijl}) \widetilde{\xi}_{ikm} \} \\
&\quad + \{ \xi_{ijl} (\widetilde{\xi}_{ikm} - \xi_{ikm}) \} - E\{ \xi_{ijl} (\widetilde{\xi}_{ikm} - \xi_{ikm}) \} + E\{ \xi_{ijl} (\widetilde{\xi}_{ikm} - \xi_{ikm}) \} \\
&= I_1 + I_2 + I_3 + I_4.
\end{aligned}
$$

Consider I_1 first. Under joint Gaussian assumption for ξ_{ijl} and e_{ijt}, it follows from Section 2.4 of Yao et al. (2005) that $\xi_{ijl} - \widetilde{\xi}_{ijl} \sim N(0, \mathrm{var}(\xi_{ijl}) - \mathrm{var}(\widetilde{\xi}_{ijl}))$. For each integer $q = 2, 3, \ldots$, by Condition 3, (33) of Dai et al. (2018) and the fact that both $\widetilde{\xi}_{ijl} - \xi_{ijl}$ and $\widetilde{\xi}_{ikm}$ are Gaussian with variances $O(T^{-1})$ and $O(m^{-\beta})$, respectively, we have

$$E|(\widetilde{\xi}_{ijl} - \xi_{ijl}) \widetilde{\xi}_{ikm}|^q \leqslant c_1 q! (c_2 T^{-1} m^{-\beta})(c_2 T^{-1/2} m^{-\beta/2})^{q-2}.$$

Applying the Bernstein inequality, we can obtain that for each $0 < \delta \leqslant c_3 T^{-1/2} m^{-\beta/2}$,

$$\mathrm{pr}(|I_1| \geqslant \delta) \leqslant c_5 \exp\left(-c_4 T m^\beta \delta^2 \right).$$

Similar developments are applied to I_3, we have $\mathrm{pr}(|I_3| \geqslant \delta) \leqslant c_7 \exp\left(-c_6 T l^\beta \delta^2 \right)$. By the Cauchy-Schwartz inequality, I_2 and I_4 are of the order $O(T^{-1/2} m^{-\beta/2} + T^{-1/2} l^{-\beta/2})$. Combing the above results and the techniques used in the proof of Lemma 7 with $\delta \gg T^{-1/2} m^{-\beta/2} + T^{-1/2} l^{-\beta/2}$, we have

$$\mathrm{pr}\left(\left| \frac{1}{n} \sum_{i=1}^{n} (\widetilde{\xi}_{ijl} \widetilde{\xi}_{ikm} - \xi_{ijl} \xi_{ikm}) \right| \geqslant \delta \right) \leqslant c_9 n \exp\left(-c_8 T m^\beta \delta^2 \right) + c_9 n \exp\left(-c_8 T l^\beta \delta^2 \right).$$

Taking $\max_{(j,k) \in V^2}$ on the left side above and again applying the union bound of probability under $p \gtrsim n$, we then complete our proof for this lemma. $\square$

LEMMA 16. *Suppose that Condition* 3 *holds, then for each* $l, m = 1, \ldots, M$, *we have*

$$\max_{(j,k) \in V^2} \left| \frac{1}{n} \sum_{i=1}^{n} \xi_{ijl} \xi_{ikm} - E(\xi_{ijl} \xi_{ikm}) \right| = O_p \left(l^{-\beta/2} m^{-\beta/2} \sqrt{\frac{\log p}{n}} \right).$$

Proof. Let $\xi_{ijl} = \omega_{jl}^{1/2} a_{ijl}$, where $a_{ijl} \sim N(0,1)$. Then $n^{-1} \sum_{i=1}^{n} \xi_{ijl} \xi_{ikm} - E(\xi_{ijl} \xi_{ikm}) = \omega_{jl}^{1/2} \omega_{km}^{1/2} \{ n^{-1} \sum_{i=1}^{n} a_{ijl} a_{ikm} - E(a_{ijl} a_{ikm}) \}$. Again, applying the Bernstein inequality and taking the union bound of probability, we obtain $\max_{j,k} |\{ n^{-1} \sum_{i=1}^{n} a_{ijl} a_{ikm} - E(a_{ijl} a_{ikm})| = O_p \{ (\log p / n)^{1/2} \}$. By Condition 3, we complete our proof for this lemma. $\square$

LEMMA 17. *Suppose that conditions for the dense case in Lemma* 14 *hold, then for each* $l, m = 1, \ldots, M$, *we have*

$$
\max_{(j,k) \in V^2} \left| \frac{1}{n} \sum_{i=1}^n \widehat{\xi}_{ijl} \widehat{\xi}_{ikm} - E(\xi_{ijl} \xi_{ikm}) \right|
$$

$$
= O_p \left\{ T^2 l m^{-\beta} \sqrt{\frac{\log p}{n^{2\gamma}}} + T^2 l^{-\beta} m \sqrt{\frac{\log p}{n^{2\gamma}}} + T^3 l^{-\beta} m^{-\beta} (\log n)^{1/2} \sqrt{\frac{\log p}{n^{2\gamma}}} \right.
$$

$$
\left. + T^{-1/2} l^{-\beta/2} (\log p)^{1/2} + T^{-1/2} m^{-\beta/2} (\log p)^{1/2} + l^{-\beta/2} m^{-\beta/2} \sqrt{\frac{\log p}{n}} \right\}.
$$

Proof. By the expansion of $n^{-1} \sum_{i=1}^n \widehat{\xi}_{ijl} \widehat{\xi}_{ikm} - E(\xi_{ijl} \xi_{ikm}) = n^{-1} \{ (\widehat{\xi}_{ijl} \widehat{\xi}_{ikm} - \widetilde{\xi}_{ijl} \widetilde{\xi}_{ikm}) + (\widetilde{\xi}_{ijl} \widetilde{\xi}_{ikm} - \xi_{ijl} \xi_{ikm}) + (\xi_{ijl} \xi_{ikm} - E(\xi_{ijl} \xi_{ikm}) \}$ and the results in Lemmas 14–16, we can immediately obtain the convergence rate, which completes our proof for this lemma. □

Our next lemma presents the uniform bound for the bias term due to M-dimensional truncated approximation.

LEMMA 18. *Suppose that the dense case in Condition* 5 *and Condition* 3 *hold, then for each* $j, k = 1, \ldots, p$, *we have*

$$
\sup_u |\Sigma_{jk}(u) - \Sigma_{jk,M}(u)| = O(n^{-\alpha\nu/2}).
$$

Proof. By the Cauchy-Schwartz inequality, the uncorrelatedness between ξ_{jl} and $\xi_{jl'}$ for $l \neq l'$ and Condition 3, we have

$$
\sup_u |\Sigma_{jk}(u) - \Sigma_{jk,M}(u)|^2 = O \left[\left\{ \sum_{l=1}^M \sum_{m=M+1}^\infty E(\xi_{jl} \xi_{km}) \sup_{\mathcal{U}} \phi_{jl}(u) \sup_{\mathcal{U}} \phi_{km}(u) \right\}^2 \right]
$$

$$
= O \left[\sum_{l=1}^M \left\{ E(\xi_{jl}^2) \sup_{\mathcal{U}} \phi_{jl}(u)^2 \right\} \sum_{m=M+1}^\infty \left\{ E(\xi_{km}^2) \sup_{\mathcal{U}} \phi_{km}(u)^2 \right\} \right]
$$

$$
= O \left(\sum_{l=1}^M \sum_{m=M+1}^\infty l^{-\beta} \omega_{km} \right)
$$

$$
\leqslant O(M^{-\nu}) \asymp n^{-\alpha\nu},
$$

where the last line comes from the fact that

$$
\sum_{l=2}^M l^{-\beta} \leqslant \sum_{l=2}^M \int_{l-1}^l x^{-\beta} dx = \int_1^M x^{-\beta} dx = \frac{1}{\beta - 1} \left(1 - M^{-(\beta-1)} \right).
$$

This completes our proof for this lemma. □

Let $\sigma_{jlkm} = E(\xi_{jl} \xi_{km})$, $\widetilde{\sigma}_{jlkm} = E(\widetilde{\xi}_{jl} \widetilde{\xi}_{km})$ and $\widehat{\sigma}_{jlkm} = n^{-1} \sum_{i=1}^n \widehat{\xi}_{ijl} \widehat{\xi}_{ikm}$.

LEMMA 19. *Suppose that conditions for the dense case in Lemma* 14 *hold, then we have*

$$
\max_{(j,k) \in V^2} \sup_{u \in \mathcal{U}} |\widehat{\Sigma}_{jk}(u) - \Sigma_{jk,M}(u)| = O_p \left\{ n^{\alpha(3\beta/2+2)} \sqrt{\frac{\log p}{n^{2\gamma}}} + T^3 \sqrt{\frac{\log p}{n^{2\gamma}}} + n^\alpha T^{-1/2} (\log p)^{1/2} \right\}
$$

Proof. Observe that

$$
\widehat{\Sigma}_{jk}(u) - \Sigma_{jk,M}(u) = \sum_{l=1}^M \sum_{m=1}^M \left\{ \widehat{\phi}_{jl}(u) \widehat{\sigma}_{jlkm} \widehat{\phi}_{km}(u) - \phi_{jl}(u) \sigma_{jlkm} \phi_{km}(u) \right\}. \tag{B.41}
$$

We can express

$$\widehat{\phi}_{jl}(u)\widehat{\sigma}_{jlkm}\widehat{\phi}_{km}(u) - \phi_{jl}(u)\sigma_{jlkm}\phi_{km}(u)$$
$$= \widehat{\phi}_{jl}(u)\widehat{\sigma}_{jlkm}\{\widehat{\phi}_{km}(u) - \phi_{km}(u)\} + \widehat{\phi}_{jl}(u)(\widehat{\sigma}_{jlkm} - \sigma_{jlkm})\phi_{km}(u)$$
$$+ \{\widehat{\phi}_{jl}(u) - \phi_{jl}(u)\}\sigma_{jlkm}\phi_{km}(u)$$
$$= I_1(u) + I_2(u) + I_3(u).$$

We first bound $I_1(u)$. It follows from Cauchy-Schwartz the inequality on $|\sigma_{jlkm}|$, Condition 3 and Lemma 12 that

$$\max_{j,k}\sup_u|I_1(u)| \leqslant \max_j\left\{\sup_u|\phi_{jl}(u)| + \sup_u|\widehat{\phi}_{jl}(u) - \phi_{jl}(u)|\right\}$$
$$\cdot \max_{j,k}\left\{|\sigma_{jlkm}| + |\widehat{\sigma}_{jlkm} - \sigma_{jlkm}|\right\}\max_k\sup_u|\widehat{\phi}_{km}(u) - \phi_{km}(u)|$$
$$= O_p\left(l^{-\beta/2}m^{-\beta/2}m^{2\beta+1}\sqrt{\frac{\log p}{n^{2\gamma}}}\right)$$
$$= O_p\left(l^{-\beta/2}m^{3\beta/2+1}\sqrt{\frac{\log p}{n^{2\gamma}}}\right).$$

We next bound $I_2(u)$. It follows from Condition 3 and Lemma 16 that

$$\max_{j,k}\sup_u|I_2(u)| \leqslant \max_j\left\{\sup_u|\phi_{jl}(u)| + \sup_u|\widehat{\phi}_{jl}(u) - \phi_{jl}(u)|\right\}\max_{j,k}|\widehat{\sigma}_{jlkm} - \sigma_{jlkm}|\max_k\sup_u|\phi_{km}(u)|$$
$$= O_p\left\{T^2 lm^{-\beta}\sqrt{\frac{\log p}{n^{2\gamma}}} + T^2 l^{-\beta}m\sqrt{\frac{\log p}{n^{2\gamma}}} + T^3 l^{-\beta}m^{-\beta}\sqrt{\frac{\log p}{n^{2\gamma}}}\right.$$
$$\left. + T^{-1/2}l^{-\beta/2}(\log p)^{1/2} + T^{-1/2}m^{-\beta/2}(\log p)^{1/2} + l^{-\beta/2}m^{-\beta/2}\sqrt{\frac{\log p}{n}}\right\}.$$

Applying the similar technique used to bound $I_1(u)$, we obtain

$$\max_{j,k}\sup_u|I_3(u)| = O_p\left(l^{3\beta/2+1}m^{-\beta/2}\sqrt{\frac{\log p}{n^{2\gamma}}}\right).$$

Combing bound results for $I_1(u)$, $I_2(u)$ and $I_3(u)$, we have

$$\max_{j,k}\sup_u|\widehat{\Sigma}_{jk}(u) - \Sigma_{jk,M}(u)|$$
$$\leqslant \sum_{l=1}^M\sum_{m=1}^M\max_{j,k}\sup_u|\widehat{\phi}_{jl}(u)\widehat{\sigma}_{jlkm}\widehat{\phi}_{km}(u) - \phi_{jl}(u)\sigma_{jlkm}\phi_{km}(u)|$$
$$\leqslant O_p\left\{\sum_{l=1}^M\sum_{m=1}^M\left(l^{-\beta/2}m^{3\beta/2+1}\sqrt{\frac{\log p}{n^{2\gamma}}} + T^2 l^{-\beta}m\sqrt{\frac{\log p}{n^{2\gamma}}}\right.\right.$$
$$\left.\left. + T^3 l^{-\beta}m^{-\beta}\sqrt{\frac{\log p}{n^{2\gamma}}} + l^{-\beta/2}T^{-1/2}(\log p)^{1/2} + l^{-\beta/2}m^{-\beta/2}\sqrt{\frac{\log p}{n}}\right)\right\}$$
$$= O_p\left\{M^{3\beta/2+2}\sqrt{\frac{\log p}{n^{2\gamma}}} + T^2 M^2\sqrt{\frac{\log p}{n^{2\gamma}}} + T^3\sqrt{\frac{\log p}{n^{2\gamma}}} + MT^{-1/2}(\log p)^{1/2}\right\}. \tag{B.42}$$

where the last line comes from

$$\sum_{l=2}^M\sum_{m=1}^M l^{-\beta/2}m^{3\beta/2+1} \leqslant \int_1^M x^{-\beta/2}dx\int_1^{M+1}y^{3\beta/2+1}dy = O(M^{3\beta/2+2})$$

and other similar inequalities. It is worth noting that actually there are terms $\{\log(Mp)\}^{1/2}$ instead of the current terms $(\log p)^{1/2}$ in (B.42) due to the union bound of probability and the equipped sub-Gaussian type of concentration bounds from Lemma 7 for each (j, k, l, m), however provided that $M \asymp n^\alpha$ and $p \gtrsim n$, $O[\{\log(Mp)\}^{1/2}]$ reduces to $O\{(\log p)^{1/2}\}$. Moreover, since $\frac{T^2 M^2}{n^\gamma} = \frac{T^3}{n^\gamma}(MT^{-1/2})^2$, the second term in (B.42) is dominated by the third and fourth terms. With $M \asymp n^\alpha$, we complete our proof for this lemma. □

LEMMA 20. *Suppose that the conditions for the dense case in Lemma* 14 *hold. Let* $\kappa_{n,T} = n^{\gamma - \alpha(3\beta/2 + 2)} \wedge T^{-3}n^\gamma \wedge T^{1/2}n^{-\alpha}$. *If* $\log p/\kappa_{n,T}^2 \to 0$, $\log p/n^{\alpha\nu} \to 0$ *and* $h^2 n^\gamma \to 0$, *then under high dimensional setting with* $p \gtrsim n$, *we have*

$$\max_{(j,k) \in V^2} \sup_{u \in \mathcal{U}} |\widehat{\Sigma}_{jk}(u) - \Sigma_{jk}(u)| = O_p\left\{ \left(\frac{\log p}{\kappa_{n,T}^2}\right)^{1/2} + \left(\frac{\log p}{n^{\alpha\nu}}\right)^{1/2} \right\}. \tag{B.43}$$

Proof. It follows from the triangular inequality, $\sup_{u \in \mathcal{U}} |\widehat{\Sigma}_{jk}(u) - \Sigma_{jk}(u)| \leqslant \sup_{u \in \mathcal{U}} |\widehat{\Sigma}_{jk}(u) - \Sigma_{jk,M}(u)| + \sup_{u \in \mathcal{U}} |\Sigma_{jk,M}(u) - \Sigma_{jk}(u)|$, Lemmas 18 and 19 that (B.43) can be obtained. □

We next turn to the sparse situation with $T_{ij} \leqslant T_0 < \infty$ uniformly in $i = 1, \ldots, n, j \in V$, and prove Lemmas 21–23 as follows.

LEMMA 21. *Suppose the sparse case in Condition* 1 *and Condition* 3 *hold, then for each* $j, k = 1, \ldots, p$, *we have*

$$\sup_u |\widetilde{\Sigma}_{jk}(u) - \widetilde{\Sigma}_{jk,M}(u)| = O(n^{-\alpha\nu/2}).$$

Proof. By the joint Gaussian property of $\xi_{ijl} - \widetilde{\xi}_{ijl}$ with variance $E(\xi_{ijl}^2) - E(\widetilde{\xi}_{ijl}^2)$, the Cauchy-Schwartz inequality, $E(\widetilde{\xi}_{jl}\widetilde{\xi}_{jl'}) = E(\xi_{jl}\xi_{jl'}) - E\{E(\xi_{jl}\xi_{jl'})|Y_{ij}\} = 0$ for $l \neq l'$, Condition 3, and the same developments in the proof of Lemma 18, we can bound the truncation error by

$$\sup_u |\widetilde{\Sigma}_{jk}(u) - \widetilde{\Sigma}_{jk,M}(u)|^2 = O\Big[\Big\{ \sum_{l=1}^M \sum_{m=M+1}^\infty E(\widetilde{\xi}_{jl}\widetilde{\xi}_{km}) \sup_{\mathcal{U}} \phi_{jl}(u) \sup_{\mathcal{U}} \phi_{km}(u) \Big\}^2 \Big]$$

$$= O\Big[\sum_{l=1}^M \big\{ E(\widetilde{\xi}_{jl}^2) \sup_{\mathcal{U}} \phi_{jl}(u)^2 \big\} \sum_{m=M+1}^\infty \big\{ E(\widetilde{\xi}_{km}^2) \sup_{\mathcal{U}} \phi_{km}(u)^2 \big\} \Big]$$

$$= O\Big(\sum_{l=1}^M \sum_{m=M+1}^\infty l^{-\beta} \omega_{km} \Big) \leqslant O(M^{-\nu}) \asymp n^{-\alpha\nu},$$

which completes our proof. □

LEMMA 22. *Suppose the sparse case in Condition* 1 *and Conditions* 2–3 *hold. Then for each* $l, m = 1, \ldots, M$,

$$\max_{(j,v) \in V^2} \Big| \frac{1}{n} \sum_{i=1}^n \widehat{\xi}_{ijl}\widehat{\xi}_{ikm} - E(\widetilde{\xi}_{ijl}\widetilde{\xi}_{ikm}) \Big| = O_p\left(lm^{-\beta}\sqrt{\frac{\log p}{n^{2\gamma}}} + l^{-\beta}m\sqrt{\frac{\log p}{n^{2\gamma}}} + l^{-\beta/2}m^{-\beta/2}\sqrt{\frac{\log p}{n}} \right).$$

Proof. Observe that $\widetilde{\xi}_{ijl}$ is Gaussian with $\mathrm{var}(\widetilde{\xi}_{ijl}) < \mathrm{var}(\xi_{ijl}) = \omega_{jl}$. Following the same techniques used in the proofs of Lemmas 15–16 and applying the Bernstein inequality again, we have

$$\max_{j,k} \Big| \frac{1}{n} \sum_{i=1}^n \widetilde{\xi}_{ijl}\widetilde{\xi}_{ikm} - E(\widetilde{\xi}_{ijl}\widetilde{\xi}_{ikm}) \Big| = \widetilde{O}_p(l^{-\beta/2}m^{-\beta/2}\sqrt{\frac{\log p}{n}}). \tag{B.44}$$

By the expansion of $n^{-1}\sum_{i=1}^n \{\widehat{\xi}_{ijl}\widehat{\xi}_{ikm} - E(\widetilde{\xi}_{ijl}\widetilde{\xi}_{ikm})\} = n^{-1}\sum_{i=1}^n \big\{ (\widehat{\xi}_{ijl}\widehat{\xi}_{ikm} - \widetilde{\xi}_{ijl}\widetilde{\xi}_{ikm}) + (\widetilde{\xi}_{ijl}\widetilde{\xi}_{ikm} - E(\widetilde{\xi}_{ijl}\widetilde{\xi}_{ikm})) \big\}$, (B.40) in Lemma 14 and (B.44), we can obtain the convergence rate as stated in the lemma. □

LEMMA 23. *Suppose that the sparse case in Condition* 1 *and Conditions* 2–3 *hold. Then we have*

$$\max_{(j,k)\in V^2} \sup_{u\in\mathcal{U}} |\widehat{\Sigma}_{jk}(u) - \widetilde{\Sigma}_{jk,M}(u)| = O_p\Big\{ n^{\alpha(3\beta/2+2)}\sqrt{\frac{\log p}{n^{2\gamma}}} \Big\}$$

Proof. Observe that

$$\widehat{\Sigma}_{jk}(u) - \widetilde{\Sigma}_{jk,M}(u) = \sum_{l=1}^M \sum_{m=1}^M \big\{ \widehat{\phi}_{jl}(u)\widehat{\sigma}_{jlkm}\widehat{\phi}_{km}(u) - \phi_{jl}(u)\widetilde{\sigma}_{jlkm}\phi_{km}(u) \big\}.$$

We can write

$$\begin{aligned}
&\widehat{\phi}_{jl}(u)\widehat{\sigma}_{jlkm}\widehat{\phi}_{km}(u) - \phi_{jl}(u)\widetilde{\sigma}_{jlkm}\phi_{km}(u)\\
&= \widehat{\phi}_{jl}(u)\widehat{\sigma}_{jlkm}\big\{\widehat{\phi}_{km}(u) - \phi_{km}(u)\big\} + \widehat{\phi}_{jl}(u)(\widehat{\sigma}_{jlkm} - \widetilde{\sigma}_{jlkm})\phi_{km}(u)\\
&\quad + \big\{\widehat{\phi}_{jl}(u) - \phi_{jl}(u)\big\}\widetilde{\sigma}_{jlkm}\phi_{km}(u)\\
&= J_1(u) + J_2(u) + J_3(u).
\end{aligned}$$

We first bound $J_1(u)$. It follows from the Cauchy-Schwartz inequality on $|\widetilde{\sigma}_{jlkm}|$, $\mathrm{var}(\widetilde{\xi}_{ijl}) \leqslant \mathrm{var}(\xi_{ijl})$, Condition 3 and Lemma 12 that

$$\begin{aligned}
\max_{j,k}\sup_u |J_1(u)| &\leqslant \max_j \big\{ \sup_u|\phi_{jl}(u)| + \sup_u|\widehat{\phi}_{jl}(u) - \phi_{jl}(u)| \big\}\\
&\quad \cdot \max_{j,k}\big\{ |\widetilde{\sigma}_{jlkm}| + |\widehat{\sigma}_{jlkm} - \widetilde{\sigma}_{jlkm}| \big\} \max_k \sup_u |\widehat{\phi}_{km}(u) - \phi_{km}(u)|\\
&= O_p\left(l^{-\beta/2} m^{-\beta/2} m^{2\beta+1}\sqrt{\frac{\log p}{n^{2\gamma}}} \right) = O_p\left(l^{-\beta/2} m^{3\beta/2+1}\sqrt{\frac{\log p}{n^{2\gamma}}} \right)
\end{aligned}$$

We next bound $J_2(u)$. It follows from Condition 3 and Lemma 22 that

$$\begin{aligned}
\max_{j,k}\sup_u |I_2(u)| &\leqslant \max_j \big\{ \sup_u|\phi_{jl}(u)| + \sup_u|\widehat{\phi}_{jl}(u) - \phi_{jl}(u)| \big\} \max_{j,k}|\widehat{\sigma}_{jlkm} - \widetilde{\sigma}_{jlkm}| \max_k \sup_u|\phi_{km}(u)|\\
&= O_p\Big(lm^{-\beta}\sqrt{\frac{\log p}{n^{2\gamma}}} + l^{-\beta}m\sqrt{\frac{\log p}{n^{2\gamma}}} + l^{-\beta/2}m^{-\beta/2}\sqrt{\frac{\log p}{n}} \Big).
\end{aligned}$$

Applying the similar technique used to bound $J_1(u)$, we obtain

$$\max_{j,k}\sup_u |J_3(u)| = O_p\left(l^{3\beta/2+1} m^{-\beta/2}\sqrt{\frac{\log p}{n^{2\gamma}}} \right).$$

Combing bound results for $J_1(u)$, $J_2(u)$ and $J_3(u)$, we have

$$\begin{aligned}
&\sup_u |\widehat{\Sigma}_{jk}(u) - \widetilde{\Sigma}_{jk,M}(u)|\\
&\leqslant \sum_{l=1}^M \sum_{m=1}^M \sup_u |\widehat{\phi}_{jl}(u)\widehat{\sigma}_{jlkm}\widehat{\phi}_{km}(u) - \phi_{jl}(u)\widetilde{\sigma}_{jlkm}\phi_{km}(u)|\\
&\leqslant O_p\left\{ \sum_{l=1}^M \sum_{m=1}^M \Big(l^{-\beta/2}m^{3\beta/2+1}\sqrt{\frac{\log p}{n^{2\gamma}}} + lm^{-\beta}\sqrt{\frac{\log p}{n^{2\gamma}}} + l^{-\beta/2}m^{-\beta/2}\sqrt{\frac{\log p}{n}} \Big) \right\}\\
&= O_p\big\{ M^{3\beta/2+2}\sqrt{\frac{\log p}{n^{2\gamma}}} \big\},
\end{aligned}$$

where the last line is due to the similar developments in the proof of Lemma 19. By $M \asymp n^\alpha$, we complete our proof for this lemma. □

810

LEMMA 24. *Suppose that the sparse case in Condition* 1 *and Conditions* 2–3 *hold. If* $\log p/n^{2\gamma-\alpha(3\beta+4)} \to 0$, $\log p/n^{\alpha\nu} \to 0$ *and* $h^2 n^\gamma \to 0$, *then under high dimensional setting with* $p \gtrsim n$, *we have*

$$\sup_{u\in\mathcal{U}}\max_{j,k}|\widehat{\Sigma}_{jk}(u) - \widetilde{\Sigma}_{jk}(u)| = O_p\left\{\left(\frac{\log p}{n^{2\gamma-3\alpha\beta-4\alpha}}\right)^{1/2} + \left(\frac{\log p}{n^{\alpha\nu}}\right)^{1/2}\right\}. \tag{B.45}$$

Proof. It follows from the triangular inequality with $\sup_{u\in\mathcal{U}}|\widehat{\Sigma}_{jk}(u) - \widetilde{\Sigma}_{jk}(u)| \leqslant \sup_{u\in\mathcal{U}}|\widehat{\Sigma}_{jk}(u) - \widetilde{\Sigma}_{jk,M}(u)| + \sup_{u\in\mathcal{U}}|\widetilde{\Sigma}_{jk,M}(u) - \widetilde{\Sigma}_{jk}(u)|$, Lemmas 21 and 23 that (B.45) can be obtained. □

LEMMA 25. *If* $\lambda_n(u) \geqslant \|\Theta(u)\|_1|\widehat{\Sigma}(u) - \Sigma(u)|_\infty$ *for each* $u \in \mathcal{U}$, *then we have*

$$|\widehat{\Theta}_1(u) - \Theta(u)|_\infty \leqslant 4\|\Theta(u)\|_1 \lambda_n(u).$$

Proof. We will use the following property that, for two matrices A and B

$$|AB|_\infty \leqslant |A|_\infty \|B\|_1 \tag{B.46}$$

in our proofs. For each $u \in \mathcal{U}$, by (B.46) and bound condition for $\lambda_n(u)$ we have

$$|I - \widehat{\Sigma}(u)\Theta(u)|_\infty = |\{\Sigma(u) - \widehat{\Sigma}(u)\}\Theta(u)|_\infty \leqslant |\Sigma(u) - \Sigma(u)|_\infty \|\Theta(u)\|_1 \leqslant \lambda_n(u). \tag{B.47}$$

By (B.47) and the optimization problem considered in (7), we obtain

$$|\widehat{\Sigma}(u)\{\widehat{\Theta}_1(u) - \Theta(u)\}|_\infty \leqslant |\widehat{\Sigma}(u)\widehat{\Theta}_1(u) - I|_\infty + |I - \widehat{\Sigma}(u)\Theta(u)|_\infty \leqslant 2\lambda_n(u). \tag{B.48}$$

By (B.47) and the definition of $\widehat{\beta}_j(u)$, $j \in V$, we have $|\widehat{\beta}_j(u)|_1 \leqslant \|\Theta(u)\|_1$. By Lemma 6 we have $\|\widehat{\Theta}_1(u)\|_1 \leqslant \|\Theta(u)\|_1$. This result together with (B.46), (B.48) and the lower bound condition for $\lambda_n(u)$ yield

$$\begin{aligned}
|\widehat{\Theta}_1(u) - \Theta(u)|_\infty &\leqslant \|\Theta(u)\|_1|\Sigma(u)\big(\widehat{\Theta}_1(u) - \Theta(u)\big)|_\infty \\
&\leqslant \|\Theta(u)\|_1\left[|\widehat{\Sigma}(u)\{\widehat{\Theta}_1(u) - \Theta(u)\}|_\infty + |\{\Sigma(u) - \widehat{\Sigma}(u)\}\{\widehat{\Theta}_1(u) - \Theta(u)\}|_\infty\right] \\
&\leqslant \|\Theta(u)\|_1\left\{2\lambda_n(u) + |\Sigma(u) - \widehat{\Sigma}(u)|_\infty\|\widehat{\Theta}_1(u) - \Theta(u)\|_1\right\} \\
&\leqslant \|\Theta(u)\|_1\{2\lambda_n(u) + |\Sigma(u) - \widehat{\Sigma}(u)|_\infty 2\|\Theta(u)\|_1\} = 4\|\Theta(u)\|_1\lambda_n(u),
\end{aligned}$$

which completes our proof for this lemma. □

Proof of Theorem 1: In our following proof, we will use the following property that, for any symmetric matrix $A \in \mathbb{R}^{p\times p}$

$$\|A\| \leqslant \|A\|_1 = \max_j \sum_{k=1}^p |A_{jk}|. \tag{B.49}$$

Let $\max_{1\leqslant j,k\leqslant p}|\widehat{\Theta}_{jk}(u) - \Theta_{jk}(u)| = \tau_n(u)$. From (6) and (8), we have $|\widehat{\Theta}_{jk}(u)| \leqslant |\widehat{\Theta}_{1jk}(u)| \leqslant |\Theta_{jk}(u)|$. This together with the fact that

$$|\widehat{\Theta}_{jk}(u)| \geqslant |\Theta_{jk}(u)| - |\widehat{\Theta}_{jk}(u)I\{|\widehat{\Theta}_{jk}(u)| \geqslant 2\tau_n(u)\} - \Theta_{jk}(u)| + |\widehat{\Theta}_{jk}(u)I\{|\widehat{\Theta}_{jk}(u)| < 2\tau_n(u)\}|.$$

leads to $|\widehat{\Theta}_{jk}I\{|\widehat{\Theta}_{jk}(u)| < 2\tau_n(u)\}| \leqslant |\widehat{\Theta}_{jk}(u)I\{|\widehat{\Theta}_{jk}(u)| \geqslant 2\tau_n(u)\} - \Theta_{jk}(u)|$. Using this result and 830
(B.49), we can bound

$$
\sup_u ||\widehat{\Theta}(u) - \Theta(u)||
$$

$$
\leqslant \sup_u \max_j \sum_{k=1}^p |\widehat{\Theta}_{jk}(u) - \Theta_{jk}(u)|
$$

$$
\leqslant \sup_u \max_j \sum_{k=1}^p |\widehat{\Theta}_{jk}(u)I\{|\widehat{\Theta}_{jk}(u)| \geqslant 2\tau_n(u)\} - \Theta_{jk}(u)|
$$

$$
+ \sup_u \max_j \sum_{k=1}^p |\widehat{\Theta}_{jk}(u)I\{|\widehat{\Theta}_{jk}(u)| < 2\tau_n(u)\}|
$$

$$
\leqslant 2\sup_u \max_j \sum_{k=1}^p |\widehat{\Theta}_{jk}(u)I\{|\widehat{\Theta}_{jk}(u)| \geqslant 2\tau_n(u)\} - \Theta_{jk}(u)|
$$

$$
\leqslant 2\sup_u \max_j \sum_{k=1}^p |\Theta_{jk}(u)I\{|\Theta_{jk}(u)| < 2\tau_n(u)\}|
$$

$$
+ 2\sup_u \max_j \sum_{k=1}^p |\widehat{\Theta}_{jk}(u)I\{|\widehat{\Theta}_{jk}(u)| \geqslant 2\tau_n(u)\} - \Theta_{jk}(u)I\{|\Theta_{jk}(u)| \geqslant 2\tau_n(u)\}|
$$

$$
\leqslant 2\{2\sup_u \tau_n(u)\}^{1-q} \sup_u \max_j \sum_{k=1}^p |\Theta_{jk}(u)|^q + 2\sup_u \tau_n(u)\sup_u \max_j \sum_{k=1}^p I\{|\widehat{\Theta}_{jk}(u)| \geqslant 2\tau_n(u)\}
$$

$$
+ 2\sup_u \max_j \sum_{k=1}^p |\Theta_{jk}(u)| \cdot |I\{|\widehat{\Theta}_{jk}(u)| \geqslant 2\tau_n(u)\} - I\{|\Theta_{jk}(u)| \geqslant 2\tau_n(u)\}|
$$

It follows from the assumption $\{\Theta(u), u \in \mathcal{U}\} \in \mathcal{C}(q, s_0(p), K; \mathcal{U})$ that the expression above can be further bounded by

$$
\leqslant 2\{2\sup_u \tau_n(u)\}^{1-q} s_0(p) + 2\sup_u \tau_n(u)\sup_u \max_j \sum_{k=1}^p I\{|\Theta_{jk}(u)| \geqslant 2\tau_n(u)\}
$$

$$
+ 2\sup_u \max_j \sum_{k=1}^p |\Theta_{jk}(u)|I\{||\Theta_{jk}(u)| - 2\tau_n(u)| \leqslant |\widehat{\Theta}_{jk}(u) - \Theta_{jk}(u)|\}
$$

$$
\leqslant 2\{2\sup_u \tau_n(u)\}^{1-q} s_0(p) + 2\{\sup_u \tau_n(u)\}^{1-q} \sup_u \max_j \sum_{k=1}^p |\Theta_{jk}(u)|^q
$$

$$
+ 2\sup_u \max_j \sum_{k=1}^p |\Theta_{jk}(u)|I\{|\Theta_{jk}(u)| \leqslant 3\tau_n(u)\}
$$

$$
\leqslant 2(1 + 2^{1-q} + 3^{1-q})\{\sup_u \tau_n(u)\}^{1-q} s_0(p).
$$

Then we can use (B.43) in Lemma 19, $K'(u) = \|\Theta(u)\|_1$ and Lemma 25 with the choice of $\lambda_n(u) = cK'(u)\{(\log p/\kappa_{n,T}^2)^{1/2} + (\log p/n^{\alpha\nu})^{1/2}\}$ to obtain the uniform convergence rate in (15), which com- 835
pletes the proof for the dense design.

For the sparse case, we substitute $\Theta(u)$ and $\Sigma(u)$ by $\widetilde{\Theta}(u)$ and $\widetilde{\Sigma}(u)$, respectively, and similarly can use (B.45) in Lemma 23, $K'(u) = \|\widetilde{\Theta}(u)\|_1$ and Lemma 25 with the choice of $\lambda_n(u) = cK'(u)\{(\log p/n^{2\gamma - \alpha(2\beta+4)})^{1/2} + (\log p/n^{4\alpha\nu})^{1/2}\}$ to obtain the uniform convergence rate in (14), which completes the proof for the sparse design. □ 840

B·6. *Proof of Theorem* 2

For each $u \in \mathcal{U}$, by (9), we have

$$\{(j,k) : (j,k) \in \widehat{E}(u), \Theta_{jk}(u) = 0\} = \{(j,k) : |\widehat{\Theta}_{jk}(u)| > \tau_n(u), \Theta_{jk}(u) = 0\}$$
$$\subseteq \{(j,k) : |\widehat{\Theta}_{jk}(u) - \Theta_{jk}(u)| \geqslant \tau_n(u)\}.$$

Hence

$$\text{pr}\Big[\sum_{j,k} I\{|\widehat{\Theta}_{jk}(u)| \geqslant \tau_n(u), \Theta_{jk}(u) = 0\}\Big] \leqslant \text{pr}\Big[\max_{j,k}\sup_u|\widehat{\Theta}_{jk}(u) - \Theta_{jk}(u)| \geqslant \inf_u\tau_n(u)\Big], \quad \text{(B.50)}$$

We can set $\tau_n(u) = 4K'(u)\lambda_n(u)$. This together with Lemmas 20, 25 and the choice of $\lambda_n(u) = cK'(u)\big\{\big(\log p/\kappa_{n,T}^2\big)^{1/2} + \big(\log p/n^{\alpha\nu}\big)^{1/2}\big\}$ imply that the probability in (B.50) is bounded by $c_2\exp\{2\log p - c_1\kappa_{n,T}^2\inf_u\tau_n(u)^2\} = c_2\exp\{2\log p - c_1\inf_uK'(u)^2\log p\}$. Hence we can choose $\inf_uK'(u)$ sufficiently large such that the probability bound goes to zero and hence $\widehat{E}(u)$ is a subset of the true edge set $E(u)$ with probability tending to 1.

Moreover, it follows from Condition 4 (ii) and (9) that for each $u \in \mathcal{U}$, the event

$$\{(j,k) : \widehat{\Theta}_{jk}(u) \leqslant \tau_n(u), \Theta_{jk}(u) > 0 \text{ or } \widehat{\Theta}_{jk}(u) \geqslant -\tau_n(u), \Theta_{jk}(u) < 0\}$$
$$\subseteq \{(j,k) : |\widehat{\Theta}_{jk}(u) - \Theta_{jk}(u)| \geqslant 2\tau_n(u) - \tau_n(u)\}.$$

Then using the above argument again, we obtain the same probability bound tending to zero, which implies that $\{E(u) \subseteq \widehat{E}(u)\}$ holds with probability tending to 1 . We can see that the thresholded estimator $\widehat{\Theta}_{jk}(u)I\{|\widehat{\Theta}_{jk}(u)| \geqslant \tau_n(u)\}$ recovers not only the true sparsity pattern, but also the signs of nonzero elements (sign consistency). Hence for each $u \in \mathcal{U}$ we have that $P\big(E(u) = \widehat{E}(u)\big) = 1 - o(1)$, which completes our proof for the dense design.

For the sparse case, we organize our proof in a similar way to the dense case. We first replace $\Theta(u)$ by $\widetilde{\Theta}(u)$ in (B.50). This fact together with Lemmas 24, 25 and the choice of $\lambda_n(u) = cK'(u)\big\{\big(\log p/n^{2\gamma - \alpha(3\beta+4)}\big)^{1/2} + \big(\log p/n^{\alpha\nu}\big)^{1/2}\big\}$ imply that the probability in (B.50) is bounded by $c_2\exp\{2\log p - c_1n^{2\gamma - \alpha(3\beta+4)}\inf_u\tau_n(u)^2\} = c_2\exp\{2\log p - c_1\inf_uK'(u)^2\log p\}$. Then by the similar argument to the dense case using Condition 4 (i), we can show that both $\{\widehat{E}(u) \subseteq \widetilde{E}(u)\}$ and $\{\widetilde{E}(u) \subseteq \widehat{E}(u)\}$ hold with probability tending to 1, which completes our proof for the sparse design.

C.   ESTIMATION OF ξ_{ijl} UNDER VERY DENSE MEASUREMENT DESIGNS

In Step 1 of the estimation, for each $i = 1, \ldots, n$, and $j \in V$, a natural estimate for ξ_{ijl} is $\int_{\mathcal{U}} \widehat{X}_{ij}(u)\widehat{\phi}_{jl}(u)du$. This approach requires the estimated trajectories $\widehat{X}_{ij}(\cdot)$, which are unavailable for the sparse case with $T_{ij} \leqslant T_0 < \infty$. As discussed previously, in this setting we estimate conditional expectations, $\widetilde{\xi}_{ijl}$, by $\widehat{\xi}_{ijl}^{(1)} = \widehat{\zeta}_{ijl}^{\text{T}}\widehat{\Sigma}_{Y_{ij}}^{-1}Y_{ij}$. However, for the dense case with $T_{ij} \to \infty$, $\int_{\mathcal{U}} \widehat{X}_{ij}(u)\widehat{\phi}_{jl}(u)du$ can be well approximated via numerical integration based on observations $\{U_{ijt}, Y_{ijt}\widehat{\phi}_{jl}(U_{ijt})\}_{1 \leqslant t \leqslant T_{ij}}$. Specifically, we implement a Trapezoid rule-based numerical integration with the non-uniform grid as follows.

$$\widehat{\xi}_{ijl}^{(2)} = \sum_{t=2}^{T_{ij}} \frac{Y_{ij(t-1)}\widehat{\phi}_{jl}(U_{ij(t-1)}) + Y_{ijt}\widehat{\phi}_{jl}(U_{ijt})}{2}|U_{ijt} - U_{ij(t-1)}|. \quad \text{(C.1)}$$

This numerical integration approach is also used to estimate functional principal component scores under dense measurement schedules in the R package fdapace (Dai et al., 2019).

Let $\widehat{\Theta}^{(1)}(u)$ and $\widehat{\Theta}^{(2)}(u)$ be the estimators for $\Theta(u)$ formed by $\widehat{\xi}_{ijl}^{(1)}$ and $\widehat{\xi}_{ijl}^{(2)}$, respectively. Intuitively, both $\widehat{\xi}_{ijl}^{(2)} - \xi_{ijl}$ and $\widehat{\Theta}^{(2)}(u) - \Theta(u)$ converge in probability to 0 under certain norms as $T_{ij} \to \infty$. For the very dense case with the T_{ij}'s growing fast enough, $\widehat{\Theta}^{(2)}(u)$ converges to $\Theta(u)$ at a faster rate than

$\widehat{\Theta}^{(1)}(u)$, so we can rely on (C.1) to calculate $\widehat{\xi}^{(2)}_{ijl}$ and $\widehat{\Theta}^{(2)}(u)$. For the sparse case or the slightly dense case with the T_{ij}'s growing slowly, the numerical integration approach does not work well, so we implement the conditional-expectation-based approach to obtain $\widehat{\xi}^{(1)}_{ijl}$ and $\widehat{\Theta}^{(1)}(u)$, the theoretical properties of which are presented in Theorem 1 (ii). We leave the theoretical investigations of $\widehat{\Theta}^{(2)}(u)$ and the phase transition phenomena from sparse to dense functional data in $\widehat{\Theta}^{(1)}(u)$ and $\widehat{\Theta}^{(2)}(u)$ under high dimensional scaling for future work.

D. Additional empirical results

D·1. *Simulations*

In Section 4·1, the animated heat map of absolute off-diagonal elements in $\Theta(u)$ at, for example 50 equally spaced points, is available from `http://personal.lse.ac.uk/qiaox/sim.eg.gif`, where the darker color corresponds to the stronger conditional dependence relationship.

D·2. *Real data*

We provide electrode/node names for $j = 1, \ldots, 64$ in Table 4. The animated adjacency matrices for $\widehat{\Theta}(u)$ at 16 evenly spaced time points is available from `http://personal.lse.ac.uk/qiaox/eeg_net.gif`.

Table 4: Electrode/node names for $j = 1, 2, \ldots, 64$.

Index	1	2	3	4	5	6	7	8	9	10	11	12	13
Name	FP1	FP2	F7	F8	AF1	AF2	FZ	F4	F3	FC6	FC5	FC2	FC1
Index	14	15	16	17	18	19	20	21	22	23	24	25	26
Name	T8	T7	CZ	C3	C4	CP5	CP6	CP1	CP2	P3	P4	PZ	P8
Index	27	28	29	30	31	32	33	34	35	36	37	38	39
Name	P7	PO2	PO1	O2	O1	X	AF7	AF8	F5	F6	FT7	FT8	FPZ
Index	40	41	42	43	44	45	46	47	48	49	50	51	52
Name	FC4	FC3	C6	C4	F2	F1	TP8	TP7	AFZ	CP3	CP4	P5	P6
Index	53	54	55	56	57	58	59	60	61	62	63	64	
Name	C1	C2	PO7	PO8	FCZ	POZ	OZ	P2	P1	CPZ	nd	Y	

References

Bosq, D. (2000). *Linear Process in Function Spaces*. New York: Springer.

Boucheron, S., Lugosi, G. & Massart, P. (2014). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press.

Cai, T., Liu, W. & Luo, X. (2011). A constrained l_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, **106**, 594-607.

Dai, X., Hadjipantelis, P. Z., Han, K., Ji, H., Lin, S. C., Muller, H. G. & Wang, J. L. (2019). fdapace: Functional data analysis and empirical dynamics. URL `https://cran.r-project.org/web/packages/fdapace/index.html`. R package version 0.4.1.

Dai, X., Muller, H. G. & Tao, W. (2018). Derivative principal component analysis for representing the time dynamics of longitudinal and functional data. *Statistical Sinica*, , **28**, 1583-1609.

Guo, S. & Qiao, X. (2018). A general theory for large-scale curve time series via functional stability measure. *Preprint*.

Qiao, X., Guo, S. & James, G. M. (2018). Functional graphical models. *Journal of the American Statistical Association*, **114**, 211-222.

Yao, F., Muller, H. G. & Wang, J. L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, **100**, 577-590.

Zhang, X. & Wang, J. L. (2016). From sparse to dense functional data and beyond. *The Annals of Statistics*, **5**, 2281-2321.

905