Doubly Functional Graphical Models in High Dimensions

Xinghao Qiao[1], Cheng Qian[1], and Gareth M. James[2]

[1]*Department of Statistics, London School of Economics, U.K.*

[2]*Department of Data Sciences and Operations, University of Southern California, U.S.A.*

Abstract

We consider the problem of estimating a functional graphical model from a data set consisting of 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 it is often the case that the functions have 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 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 also demonstrate that the proposed method significantly outperforms possible competitors through an extensive set of simulation studies. Finally, our proposed method is applied to a brain imaging dataset, revealing some interesting findings.

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, $\mathbf{X} = (X_1, \ldots, X_p)^{\mathrm{T}}$, 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 the graph G is to recover the edge set E. In particular it is well known that, for a multivariate Gaussian distributed $\mathbf{X}$, recovering the structure of an undirected graph is equivalent to locating the non-zero components in the precision matrix (i.e. the inverse covariance matrix) of $\mathbf{X}$ (Lauritzen, 1996).

The past several years have witnessed the development of Gaussian graphical models in large p, small n, settings. One popular class of approaches, the graphical lasso, estimates the graphical model by optimizing a criterion involving the Gaussian log likelihood with a lasso type penalty on the entries of the precision matrix (Yuan and 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); Danaher et al. (2014). Another popular class of neighborhood based approaches, first proposed by Meinshausen and Buhlmann (2006), considers recovering the support of G by solving p lasso problems in a column-by-column manner. Cai et al. (2011) proposed a Dantzig-type variant of this approach, named constrained ℓ_1-minimization for inverse matrix estimation. Some recent works along this line of research include 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 Gaussian graphical model corresponds to the standard setting involving high dimensional, but static, data from which we estimate a single (static) 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 and Xing, 2011).

The setting we are interested in corresponds to the last row of Table 1, where the data can be considered functional. To illustrate the data structure and underlying network pattern Figure 1 provides a simple example. Specifically, the left hand side of Figure 1 plots $n = 100$

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

realizations of $p = 10$ random curves in $\mathcal{U} = [0, 1]$ (red lines), each of which corresponds to one underlying node. In practice, functions can be observed at either a dense grid of points (black dots) or a small subset of possible points (green squares), contaminated by measurement error. Qiao et al. (2017) propose a static functional graphical model, where a single network is constructed to encode the global conditional dependence relationship among large-scale Gaussian random functions. Li et al. (2017) relax the Gaussian assumption and explore the additive conditional dependence structure by treating p as fixed. By comparison 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).

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, e.g. 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

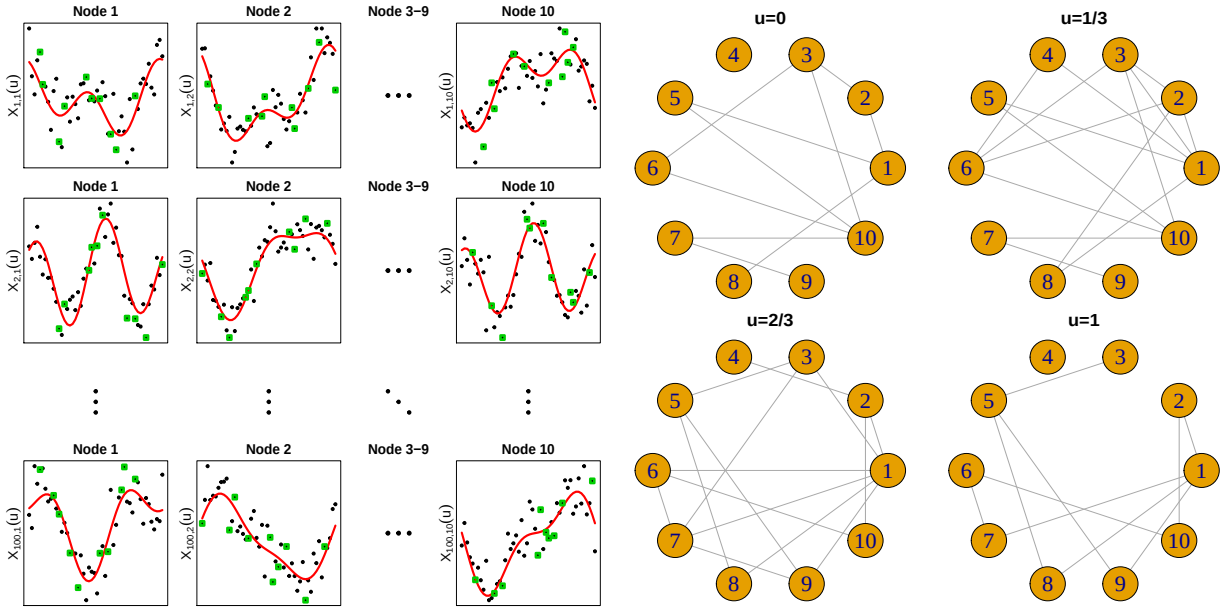


Figure 1: *Data generated from the simulation setting in Section 4. Left: The data matrix consists of* 100 *random functional realizations (red line), their noisy observations at either 50 evenly spaced points (black dots) or 10 randomly selected points (green squares), for* $j = 1, \ldots, 10$ *nodes. Right: Visualization of true functional network at 4 selected time points.*

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, but in practice curves are often observed at different sets of points. Another approach is to use nonparametric smoothers to estimate the cross-covariance operator 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 high dimensional large p, small n, scenario.

The main purpose of this paper is to propose a doubly functional graphical model that can be used to estimate the functional network for multivariate sparsely or densely sampled functional data. Our proposed method is powerful but can be implemented in three relatively simple steps. First, we apply a nonparametric approach to smooth p covariance operators 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 by p covariance matrix as it varies over $u \in \mathcal{U}$. Finally, we estimate the functional network by computing the functional sparse precision matrix at each of a set of points. This final step

can be easily implemented through existing approaches for estimating the sparse precision matrix. Our theoretical results make use of 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 nonparametric smoothing of $p(p+1)/2$ cross-covariance operators or "joint" methods, since we only need to smooth p covariance operators 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 uniform convergence rates for the estimated functional precision matrix even in the more challenging large p, small n, setting, for both sparsely and densely sampled functional data. Finally, empirically we demonstrate the superiority of our proposed method relative to its natural competitors.

2 Methodology

We begin by introducing some notation. For a vector $\mathbf{a} = (a_1, \ldots, a_p)^{\mathrm{T}}$, its ℓ_r norm is given by $|\mathbf{a}|_r = \{\sum_{i=1}^p |a_i|^r\}^{1/r}$. For a matrix $\mathbf{A} = (A_{ij}) \in \mathbb{R}^{p\times q}$, we define the element-wise ℓ_r norm by $|\mathbf{A}|_r = \{\sum_{i=1}^p \sum_{j=1}^q |A_{ij}|^r\}^{1/r}$, in particular $r = 2$ corresponds to the matrix Frobenius norm, $||\mathbf{A}||_F = |\mathbf{A}|_2$. We denote the matrices ℓ_1, ℓ_∞, ℓ_2 (operator) norms by $||\mathbf{A}||_1 = \max_{1\leqslant j\leqslant q} \sum_{i=1}^p |A_{ij}|$, $||\mathbf{A}||_\infty = \max_{1\leqslant i\leqslant p} \sum_{j=1}^q |A_{ij}|$, $||\mathbf{A}|| = \sup_{|\mathbf{x}|_2\leqslant 1} |\mathbf{A}\mathbf{x}|_2$, respectively. We use $x \wedge y = \min(x, y)$ and $x \vee y = \max(x, y)$. For a bivariate function $\psi(u, v), (u, v) \in \mathcal{U}^2$, define the Hilbert-Schmidt norm by $||\psi||_{\mathcal{S}} = \left(\int\int \psi(u, v)^2 dudv\right)^{1/2}$. 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 $g(n) = O\{f(n)\}$.

2.1 Doubly functional graphical models

Let $\mathbf{X}(u) = \big(X_1(u), \ldots, X_p(u)\big)^{\mathrm{T}}, u \in \mathcal{U}$ denote a p-dimensional vector of Gaussian random functions, with each X_j 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 $\mathbf{X}(u)$ has been centered to have mean zero. Let $\mathbf{C}(u,v) = \left\{C_{jk}(u,v)\right\}_{1\leqslant j,k\leqslant p}$ be the p by p matrix whose (j,k)-th element is $C_{jk}(u,v) = \mathrm{Cov}\big(X_j(u), X_k(v)\big)$, the cross-covariance between $X_j(u)$ and $X_k(v)$.

Let $G(u) = \big(V, E(u)\big)$ denote an undirected functional graph depending on $u \in \mathcal{U}$, with a vertex set $V = \{1, \ldots, p\}$ and corresponding functional edge set

$$E(u) = \left\{(j,k) : \mathrm{Cov}\big(X_j(u), X_k(u) | \{X_l(u), l \neq j,k\}\big) \neq 0, (j,k) \in V^2, j \neq k\right\}, u \in \mathcal{U}.$$

Standard results show that, for each $u \in \mathcal{U}$, $\mathbf{X}(u) \sim N(\mathbf{0}, \mathbf{\Sigma}(u))$, where $\mathbf{\Sigma}(u) = \mathbf{C}(u,u) \in \mathbb{R}^{p\times p}$ and $\mathbf{\Theta}(u) = \mathbf{\Sigma}(u)^{-1}$, then $\mathrm{Cov}\big(X_j(u), X_k(u) | \{X_l(u), l \neq j,k\}\big) = 0$ if and only if $\Theta_{jk}(u) = 0$. Hence $E(u)$ can be equivalently represented by

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

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

Step 1. For each $j \in V$, we adopt a data-driven basis expansion approach through functional principal component analysis. To be specific, the covariance function $C_{jj}(u,u)$ for $X_j(u)$ satisfies the eigen-decomposition, $\int_{v\in\mathcal{U}} C_{jj}(u,v)\phi_{jl}(v)dv = \omega_{jl}\phi_{jl}(u), l = 1,2,\ldots$. Here $\omega_l \geqslant 0$ is the l-th eigenvalue in non-increasing order, and its corresponding eigenfunction $\phi_{jl}(u)$ satisfies $\int_{u\in\mathcal{U}} \phi_{jl}(u)\phi_{jl'}(u)du = I(l = l')$, where $I(\cdot)$ is the indicator function. The Karhunen-Loève expansion allows us to expand each $X_j(u)$ as $X_j(u) = \sum_{l=1}^{\infty} \xi_{jl}\phi_{jl}(u)$, 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 $X_j(u)$ using the leading M principal components, i.e. $X_{j,M}(u) = \sum_{l=1}^{M} \xi_{jl}\phi_{jl}(u)$, where M is chosen large enough to provide a reasonable approximation to the trajectory $X_j(u)$. 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 2.3 for details.

Step 2. Once Step 1 has been performed for each $X_j(u)$ the M-dimensional functional representation leads to a natural approximation for the p by p functional covariance matrix $\mathbf{\Sigma}_M(u)$, with (j,k)-th entry given by:

$$\Sigma_{jk,M}(u) = \sum_{l=1}^{M}\sum_{m=1}^{M} \mathrm{Cov}(\xi_{jl}, \xi_{km})\phi_{jl}(u)\phi_{km}(u). \tag{2}$$

Step 3. Our final step involves computing a functional sparse precision matrix $\mathbf{\Theta}_M(u) = \mathbf{\Sigma}_M(u)^{-1}$. We implement this step by estimating $\mathbf{\Theta}_M(u)$ separately at each of a set of points in $\mathcal{U}$.

2.2 Estimation

Let $\mathbf{X}_i(u) = \big(X_{i1}(u), \ldots, X_{ip}(u)\big)^{\mathrm{T}}, i = 1, \ldots, n$ be i.i.d. copies of $\mathbf{X}(u)$. We assume that $X_{ij}(u)$ 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

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

where the e_{ijt}'s are i.i.d. with $E(e_{ijt}) = 0$ and $\mathrm{Var}(e_{ijt}) = \sigma^2$, independent of X_{ij}, and the U_{ijt}'s are sampled from some specific density f_U. We provide estimation details to implement our three-step approach from Section 2.1 as follows.

Step 1. To perform functional principal components analysis based on realizations $\mathbf{Y}_{ij} = (Y_{ij1}, \ldots, Y_{ijT_{ij}})^{\mathrm{T}}, i = 1, \ldots, n$, for each $j \in V$, we first compute the estimator for $C_{jj}(u, v)$. Let $\mathbf{\Sigma}_{\mathbf{Y}_{ij}}$ be the covariance matrix for $\mathbf{Y}_{ij}$ with (t, t')-th element $\big(\mathbf{\Sigma}_{\mathbf{Y}_{ij}}\big)_{tt'} = \mathrm{Cov}(Y_{ijt}, Y_{ijt'}) = C_{jj}(U_{ijt}, U_{ijt'}) + \sigma^2 I(t = t')$. A local linear surface smoother is applied to the off-diagonals of the "raw covariances", $Y_{ijt}Y_{ijt'}, t \neq t'$. 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 ith subject and the jth variable 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 and 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 eigen-pairs $(\widehat{\omega}_{jl}, \widehat{\phi}_{jl}), l = 1, \ldots, M$. The estimated principal component scores are $\widehat{\xi}_{ijl} = \int_{\mathcal{U}} \widehat{X}_{ij}(u)\widehat{\phi}_{jl}(u)du$. However, this approach requires the estimated trajectories, $\widehat{X}_{ij}(u)$, which are unavailable, especially for sparse designs. Instead, we propose to use the best linear unbiased predictors $\widetilde{\xi}_{ijl} = \boldsymbol{\zeta}_{ijl}^{\mathrm{T}}\mathbf{\Sigma}_{\mathbf{Y}_{ij}}^{-1}\mathbf{Y}_{ij}$ (Rice and Wu, 2001), where $\boldsymbol{\zeta}_{ijl}$ is a T_{ij}-dimensional vector with t-th

component

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

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

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

where $\widehat{\zeta}_{ijlt} = \int \widehat{C}_{jj}(U_{ijt}, v)\widehat{\phi}_{jl}(v)dv$, and $\big(\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}_{ij}}\big)_{tt'} = \widehat{C}_{jj}(U_{ijt}, U_{ijt'}) + \widehat{\sigma}^2 I(t = t')$. 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 $\widehat{\boldsymbol{\Sigma}}(u)$ 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

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

where $\mathbf{I} \in \mathbb{R}^{p\times p}$ is the identity matrix and $\lambda_n(u) \geqslant 0$ is a tuning parameter which controls the sparsity level of $\check{\boldsymbol{\Theta}}(u)$. The convex problem (6) can be further decomposed into p separate optimization problems. For $j = 1, \ldots, p$, we solve

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

where $\mathbf{e}_j \in \mathbb{R}^p$ is the unit vector with j-th coordinate 1 and $\widehat{\boldsymbol{\beta}}_j(u)$ corresponds to the j-th column of $\check{\boldsymbol{\Theta}}(u)$.

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

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

This symmetrization procedure guarantees that our estimator $\widehat{\boldsymbol{\Theta}}(u)$ achieves the same elementwise ℓ_∞ estimation error rate as $\check{\boldsymbol{\Theta}}(u)$. 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.

2.3 Selection of tuning parameters

To fit our proposed method, we must choose optimal values for three tuning parameters, h_j (the bandwidth in the surface smoothing step for $j \in V$), M_j (the number of selected principal components for $j \in V$) and $\lambda_n(u)$ (the regularization parameter to control the sparsity level of $\mathbf{\Theta}(u)$).

We adopt leave-one-curve-out cross validation (Rice and Silverman, 1991) to select the optimal smoothing parameters in (4). See Zhang and Wang (2016) for the discussion on 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 and Lin, 2007; Cai et al., 2011). We adopt the computationally more efficient approach based on the Bayesian Information Criterion to choose an optimal $\lambda_n(u)$ by minimizing

$$n\text{trace}\left(\widehat{\mathbf{\Theta}}_{\lambda_n(u)}(u)\widehat{\mathbf{\Sigma}}(u)\right) - n\log\det\left(\widehat{\mathbf{\Theta}}_{\lambda_n(u)}(u)\right) + \log n|\widehat{E}(u)|,$$

over a series of $\lambda_n(u)$ values, where $\widehat{\mathbf{\Theta}}_{\lambda_n(u)}(u)$ is the regularized estimator corresponding to $\lambda_n(u)$ and $|\widehat{E}(u)|$ is the number of non-zero components in $\widehat{\mathbf{\Theta}}_{\lambda_n(u)}(u)$. It is worth noting that the aforementioned approaches may tend to choose networks too dense to be interpretable. Stability analysis based graph selection (Meinshausen and Buhlmann, 2010) can be used to estimate edge sets with low false discovery rates.

2.4 Relationship to relevant work

We compare the doubly functional graphical model with the static functional graphical model of Qiao et al. (2017), where a single graph is constructed to depict the entire functional con-

ditional dependence structure. To illustrate the difference, we consider a simplified setting, where, for each $j \in V$, $X_j(u) = \boldsymbol{\xi}_j^{\mathrm{T}} \boldsymbol{\phi}_j(u)$ belongs to an M-dimensional Gaussian process. The static functional graphical model generates a single network by recovering the block sparsity pattern in $\boldsymbol{\Omega}^{-1} \in \mathbb{R}^{Mp \times Mp}$ whose (j,k)-th block is given by $\boldsymbol{\Omega}_{jk} = \mathrm{Cov}(\boldsymbol{\xi}_j, \boldsymbol{\xi}_k)$. By contrast, the doubly functional graphical model constructs a separate network for each value of u by estimating the sparsity structure in $\boldsymbol{\Theta}(u) = \left\{\boldsymbol{\Phi}(u)^{\mathrm{T}} \boldsymbol{\Omega} \boldsymbol{\Phi}(u)\right\}^{-1}$, where $\boldsymbol{\Phi}(u) \in \mathbb{R}^{Mp \times p}$ is block diagonal with j-th block given by $\boldsymbol{\phi}_j(u) \in \mathbb{R}^{M \times 1}, j \in V$. 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. Thus the static functional graphical model may end up with an overly dense set of edges that risks miss-characterizing the true structure. 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 $\mathbf{X}_i(u_t) = (X_1(u_t), \ldots, X_p(u_t))^{\mathrm{T}}$ for $i = 1, \ldots, n$ and $t = 1, \ldots, T$ with $u_t \in \mathcal{U}$, and both approaches fit a separate graphical model at a given set of values for u. However, Zhou et al. (2010) assumes only one observation at each u_t, i.e. $n = 1$, and models the $\mathbf{X}_i(u_t)$'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 $\{\mathbf{X}_i(u_t)\}_{i=1}^n$ as following a lag one stationary vector autoregressive model, i.e. $\mathbf{X}_i(u_t)$ is correlated with $\mathbf{X}_{i-1}(u_t)$. By contrast, we treat $\{\mathbf{X}_i(u)\}_{i=1}^n$ as independent realizations of an underlying multivariate functional 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

$$\mathcal{C}(q, s_0(p), K; \mathcal{U}) = \Big\{ \{\mathbf{\Theta}(u), u \in \mathcal{U}\} \big| \sup_{u\in\mathcal{U}} ||\mathbf{\Theta}(u)||_1 < K, \operatorname*{supmax}_{u\in\mathcal{U}} \sum_{j\in V}^{p} |\Theta_{jk}(u)|^q \leqslant s_0(p) \Big\}, \quad (10)$$

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 truly functional sparse situation, where even the densest $\mathbf{\Theta}(u)$ over $u \in \mathcal{U}$ has at most $s_0(p)$ non-zero entries on each row. Similar classes were used in estimating both static covariance models (Bickel and Levina, 2008) and its generalization to the dynamic setting (Chen and Leng, 2016). We extend the class of static "approximately sparse" precision matrices (Cai et al., 2011) to the functional version via (10), uniformly over which Theorems 1-2 hold.

To present the main theorems, we need the following regularity conditions.

Condition 1 *(i) Let* $\{U_{ijt} : i = 1, \ldots, n, j \in V, t = 1, \ldots, T_{ij}\}$ *be i.i.d. 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* $0 < m_f \leqslant \inf_{u\in\mathcal{U}} f_U(u) \leqslant \sup_{u\in\mathcal{U}} f_U(u) \leqslant M_f < \infty$*; (ii) For all* $i = 1, \ldots, n, j \in V$*, in the sparse measurement design,* $T_{ij} \leqslant T_0 < \infty$*, and in the dense design,* $T_{ij} \asymp T \to \infty$*.*

Condition 2 *For each* $j \in V$*, there exists a sequence* $n^{-\gamma}$ *for some* $0 < \gamma \leqslant 1/2$ *and some positive constants* C_1, C_2, C_3 *such that for any* $0 < \delta \leqslant C_1$*,*

$$P\left(\left\| \widehat{C}_{jj}(u, v) - E\{\widehat{C}_{jj}(u, v)\} \right\|_{\mathcal{S}} \geqslant \delta \right) \leqslant C_2 \exp(-C_1 n^{2\gamma} \delta^2), \quad (11)$$

$$P\left(\sup_{(u,v)\in\mathcal{U}^2} \left| \widehat{C}_{jj}(u, v) - E\{\widehat{C}_{jj}(u, v)\} \right| \geqslant \delta \right) \leqslant C_2 n^{C_3} \exp(-C_1 n^{2\gamma} \delta^2). \quad (12)$$

The concentration inequality in (11) guarantees that the L_2 convergence rate of $\widehat{C}_{jj}(u, v)$ to its expected value will be $n^{-\gamma}$, while (12) provides a uniform convergence rate of $(\log n)^{1/2} n^{-\gamma}$. To simplify notation, for each $j \in V$, we assume the same value of γ, which depends on the weighting scheme (w_{ij}), bandwidth choice (h_j) and the number of observed points (T_{ij}). See Theorems 4.1 and 5.1 of Zhang and Wang (2016) for details. A larger value of γ corresponds to a more frequent measurement schedule. Li et al. (2017) classified measurement schedules

as "non-dense" and "dense" according to whether $\widehat{C}_{jj}(u,v)$ has an $n^{-1/2}$ convergence rate. In particular, the sparse design assumed in Condition 1 belongs to Li et al.'s "non-dense" case and, provided T grows fast enough, the dense design in Condition 1 corresponds to Li et al.'s "dense" case. Under the Gaussian assumption for $X_{ij}(u)$ with exponential bounds on the tail probabilities, we assume that the L_2 and uniform convergence rates are equipped with the corresponding concentration inequalities in (11) and (12), respectively. Specially, we have proved that, for fully observed functional data, (11) and (12) hold with $\gamma = 1/2$ and $C_3 = 1$. See Lemmas 20 and 21 in the supplementary material for details.

Condition 3 *(i) The truncated dimension of the functional data, M, satisfies $M \asymp n^{\alpha}$ with some constant $\alpha > 0$; (ii) The principal component functions are continuous on the compact set $\mathcal{U}$ and satisfy $\max_{j\in V}\sup_{u\in\mathcal{U}}\sup_{l\geqslant 1}|\phi_{jl}(u)| = O(1)$; (iii) The eigenvalues satisfy $\omega_{j1} > \omega_{j2} > \cdots > \omega_{jM} > \omega_{j(M+1)} \geqslant \cdots$ 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 $l \geqslant 2$ and $d_{j1} = (\omega_{j1} - \omega_{j2})^{-1}$; (iv) There exists some constant $\nu > 0$ such that $\max_{j\in V}\sum_{l=M+1}^{\infty}\omega_{jl}^{1/2} \leqslant O(M^{-\nu})$.*

The parameter α in Condition 3 (i) determines the number of leading principal components that are necessary to provide a reasonable approximation to the infinite dimensional process. Condition 3 (iii) provides decay rates for the strictly decreasing sequence of $\omega_{j1},\ldots,\omega_{jM}$ and gaps between adjacent eigen-values, i.e. d_{jk}^{-1}'s, both of which are used to derive the convergence rates of estimated eigen-pairs (Bosq, 2000; Qiao et al., 2017). For the dense case, $\mathbf{\Delta}(u) = \mathbf{\Sigma}(u) - \mathbf{\Sigma}_M(u)$ measures the discrepancy between $\mathbf{\Sigma}(u)$ and its truncated approximation. Then Condition 3 (iv) can be used to show that the entries in $\mathbf{\Delta}(u)$ uniformly converge to zero at a certain rate. We finally remark that Conditions 3 (iii) and (iv) hold if, for instance, for each $j \in V$, $C_{jj}(u,v)$ satisfies Sacks-Ylvisaker conditions of order $r \in \mathbb{N}_0$, with $w_{jl} \asymp l^{-2r-2}$ (Ritter et al., 1995) and $\sum_{l=M+1}^{\infty}\omega_{jl}^{1/2} \leqslant O(M^{-r})$.

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. Define $\widetilde{\mathbf{\Sigma}}(u)$ with its (j,k)-th entry given by $\widetilde{\Sigma}_{jk}(u) = \sum_{l=1}^{\infty}\sum_{m=1}^{\infty}\mathrm{Cov}(\widetilde{\xi}_{ijl}, \widetilde{\xi}_{ikm})\phi_{jl}(u)\phi_{km}(u)$. Note we can prove that $\widehat{\mathbf{\Sigma}}(u)$ is a consistent estimator for $\widetilde{\mathbf{\Sigma}}(u)$, but $\widetilde{\mathbf{\Sigma}}(u)$ fails to converge to $\mathbf{\Sigma}(u)$ unless the T_{ij}'s diverge to ∞. Therefore, for the sparse design, we denote the population functional precision matrix by $\widetilde{\mathbf{\Theta}}(u) = \widetilde{\mathbf{\Sigma}}(u)^{-1}$ 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 $\boldsymbol{\Theta}(u)$ and $E(u)$ as the true functional precision matrix and edge set, respectively.

Theorem 1 *Suppose that Conditions 1–3 hold.*

(i) For the sparse design with $T_{ij} \leqslant T_0 < \infty$, suppose that $\{\widetilde{\boldsymbol{\Theta}}(u), u \in \mathcal{U}\}$ belongs to $\mathcal{C}(q, s_0(p), K; \mathcal{U})$. If $\lambda_n(u) = CK'(u)\big\{(\log p/n^{2\gamma-\alpha(2\beta+4)})^{1/2} + (\log p/n^{4\alpha\nu})^{1/2}\big\}$ with C sufficiently large, $K'(u)$ satisfying $\sup_{u\in\mathcal{U}} K'(u) \leqslant K$, $\log p/n^{2\gamma-\alpha(2\beta+4)} \to 0$ and $\log p/n^{4\alpha\nu} \to 0$, then we have

$$\sup_{u\in\mathcal{U}} \big\|\widehat{\boldsymbol{\Theta}}(u) - \widetilde{\boldsymbol{\Theta}}(u)\big\| = O_p\left\{K^{2(1-q)} s_0(p) \left(\sqrt{\frac{\log p}{n^{2\gamma-\alpha(2\beta+4)}}} + \sqrt{\frac{\log p}{n^{4\alpha\nu}}}\right)^{1-q}\right\}. \tag{13}$$

(ii) For the dense design with $T_{ij} \asymp T \to \infty$, suppose that $\{\boldsymbol{\Theta}(u), u \in \mathcal{U}\}$ 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 $\lambda_n(u) = CK'(u)\big\{(\log p/\kappa_{n,T}^2)^{1/2} + (\log p/n^{2\alpha\nu})^{1/2}\big\}$ with C sufficiently large, $K'(u)$ satisfying $\sup_{u\in\mathcal{U}} K'(u) \leqslant K$, $\log p/\kappa_{n,T}^2 \to 0$ and $\log p/n^{2\alpha\nu} \to 0$, then we have

$$\sup_{u\in\mathcal{U}} \big\|\widehat{\boldsymbol{\Theta}}(u) - \boldsymbol{\Theta}(u)\big\| = O_p\left\{K^{2(1-q)} s_0(p) \left(\sqrt{\frac{\log p}{\kappa_{n,T}^2}} + \sqrt{\frac{\log p}{n^{2\alpha\nu}}}\right)^{1-q}\right\}. \tag{14}$$

We observe that the uniform convergence rate in (14) for the dense design is governed by two sets of parameters: (1) dimensionality parameters, sample size (n), number of functional variables (p), number of observed time points (T); (2) internal parameters, the truncated dimension ($M \asymp n^{\alpha}$), decay rate of L_2 convergence for $\widehat{C}_{jj}(u, v)$ (γ), decay rate of eigenvalues (β), decay rate of truncated errors for sum of root-eigenvalues (ν), strength of sparsity (q), uniform upper bound of matrix magnitude (via (10)) for the uniformity class of functional precision matrices (K and $s_0(p)$). When the smoothness across p covariance operators differs, the convergence rate will be determined by the least smooth component (the largest M_j and the smallest h_j).

The uniform convergence rates in (13) and (14) consist of two terms, which reflect our familiar bias-variance tradeoff as commonly considered in the nonparametric setting. (For instance, in the sparse design with $q = 0$, corresponding to the functional graphical modelling setting, the bias is uniformly bounded by $O\big\{K^2 s_0(p)(\log p/n^{4\alpha\nu})^{1/2}\big\}$ and the variance is of the order $O_p\big\{K^2 s_0(p)(\log p/n^{2\gamma-\alpha(2\beta+4)})^{1/2}\big\}$.) To balance both terms in the sparse design, we choose $\alpha = \gamma/(\beta + 2\nu + 2)$, which provides a truncated dimension of $M \asymp n^{\gamma/(\beta+2\nu+2)}$ and the optimal uniform rate of convergence in (13) becomes

$O_p\left\{K^{2(1-q)}s_0(p)\left(\log p/n^{4\gamma\nu/(\beta+2\nu+2)}\right)^{(1-q)/2}\right\}$. When ν diverges to infinity, M approaches a fixed dimension and the optimal rate goes to $O_p\left\{K^{2(1-q)}s_0(p)\left(\log p/n^{2\gamma}\right)^{(1-q)/2}\right\}$. 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.

Condition 4 *(i) For the sparse design, let* $\widetilde{S}(u) = \left\{\widetilde{E}(u) \cup (1,1), \ldots, (p,p)\right\}$ *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* $S(u) = \left\{E(u) \cup (1,1), \ldots, (p,p)\right\}$ *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 and is crucial to develop the graph selection consistency result in Theorem 2. See Chen and Leng (2016) and Qiu et al. (2016) for analogous functional minimum signal strength conditions.

Theorem 2 *(i) For the sparse design, suppose that the conditions in Theorem 1(i) hold. If it is further assumed that* $\{\widetilde{\mathbf{\Theta}}(u), u \in \mathcal{U}\}$ *belongs to* $\mathcal{C}(0, s_0(p), K; \mathcal{U})$ *and* $\tau_n(u) = 4K'(u)\lambda_n(u)$, *then the event* $\left\{\widehat{E}(u) = \widetilde{E}(u)\right\}$ *holds with probability tending to 1 uniformly for* $u \in \mathcal{U}$ *satisfying Condition 4(i);*

(ii) For the dense design, suppose that the conditions in Theorem 1(ii) hold. If it is further assumed that $\{\mathbf{\Theta}(u), u \in \mathcal{U}\}$ *belongs to* $\mathcal{C}(0, s_0(p), K; \mathcal{U})$ *and* $\tau_n(u) = 4K'(u)\lambda_n(u)$, *then the event* $\left\{\widehat{E}(u) = E(u)\right\}$ *holds with probability tending to 1 uniformly for* $u \in \mathcal{U}$ *satisfying Condition 4(ii).*

When $q = 0$, we focus on the uniformity class of truly functional sparse precision matrices, which sheds light on the functional graph structure recovery. If we further assume that Condition 4 holds, then we are able to obtain the graph selection consistency result as stated in Theorem 2. Note one can understand Condition 4 as the minimum strength of the strong signal when $\mathbf{\Theta}(u)$ varies smoothly. For the non-smooth situation, Theorem 2 still holds uniformly in $u \in \mathcal{U}$. To summarize, to the best of our knowledge, Theorems 1 and 2 provide the first uniform consistency results for the functional sparse precision matrix estimation and its support recovery in the high-dimensional large p, small n, setting with both sparse and dense measurement schedules.

4 Simulations

To assess the finite sample performance of our method, we test the proposed doubly functional graphical model, using both the constrained ℓ_1-minimization and graphical lasso approaches. Sections 4.1 and 4.2 respectively consider scenarios where functions are observed at a common, or an irregular, sets of points.

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

$$Y_{ijt} = \mathbf{s}(u_{ijt})^{\mathrm{T}}\boldsymbol{\delta}_{ij} + \varepsilon_{ijt}, \quad i = 1, \ldots, 200, j = 1, \ldots, p, t = 1, \ldots, T_{ij}, \tag{15}$$

where u_{ijt}'s and ε_{ijt}'s are randomly sampled from Unif$[0,1]$ and $N(0, 0.5)$, respectively.

Since functional conditional dependence relationships are fully characterized by the corresponding functional sparsity pattern in $\boldsymbol{\Theta}(u) = \boldsymbol{\Sigma}(u)^{-1}$ we consider a setting, generalized from Zhou et al. (2010), to simulate a $\boldsymbol{\Theta}(u)$ corresponding to a slow graph evolution over $[0,1]$. When $u = 0$, the initial diagonal elements in $\boldsymbol{\Theta}(0)$ are set to 0.25. For $p = 50$, we randomly select $n_{\text{initial}} = 40$ out of $50(50-1)/2$ potential edges, with edge strengths generated from a Unif$[-0.3, -0.1]$ distribution. 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 Unif$[-0.3, -0.1]$ 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 $\boldsymbol{\Theta}(u)$. The animated heat map of absolute off-diagonal elements in $\boldsymbol{\Theta}(u)$ 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. 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.

Note that generating functions corresponding to a given functional covariance matrix, $\boldsymbol{\Sigma}(u) = \boldsymbol{\Theta}(u)^{-1}$, is itself a non-trivial problem. We develop an approach using ideas from

linear models of coregionalization (Genton and Kleiber, 2015) to generate our data. See Section A.1 in the appendix for details.

4.1 Common set of time points

When curves are measured at a common set of points $u_1, \ldots, u_T$, i.e. $u_{ijt} = u_t$ with $T_{ij} = T$ for all i, j, t in (15), we compare two versions of our method (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 $\mathbf{S}(u_t)$ of $(Y_{i1t}, \ldots, Y_{ipt})^{\mathrm{T}}$ for $i = 1, \ldots, n$, that is $\mathbf{S}_h(u) = \{\sum_{t=1}^T K_h(u_t - u)\mathbf{S}(u_t)\}\{\sum_{t=1}^T K_h(u_t - u)\}^{-1}, u \in [0, 1]$. We use a Gaussian kernel with the optimal bandwidth $h_{\mathrm{opt}} \propto \{\log p/(nT)\}^{1/3} \vee T^{-4/5}$ (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 minimization or graphical lasso on $\mathbf{S}(u_t)$ for $t = 1, \ldots, T$. To ensure these competitors would 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 these seven approaches based on both the estimation accuracy and graph recovery consistency. In terms of the estimation accuracy, for each candidate method, we calculate the mean of the ℓ_1, ℓ_2 and ℓ_F losses for the estimated precision matrices, respectively defined as $||\widehat{\mathbf{\Theta}}(u) - \mathbf{\Theta}(u)||_1$, $||\widehat{\mathbf{\Theta}}(u) - \mathbf{\Theta}(u)||$, and $||\widehat{\mathbf{\Theta}}(u) - \mathbf{\Theta}(u)||_F$, 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 $\frac{\#\{(j,k):\widehat{\Theta}_{jk}^{(\lambda_n)}(u)\neq 0 \text{ and } \Theta_{jk}(u)\neq 0\}}{\#\{(j,k):\Theta_{jk}(u)\neq 0\}}$ and $\frac{\#\{(j,k):\widehat{\Theta}_{jk}^{(\lambda_n)}(u)\neq 0 \text{ and } \Theta_{jk}(u)= 0\}}{\#\{(j,k):\Theta_{jk}(u)= 0\}}$, over a grid of $\lambda_n(u)$ values to produce the underlying 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: Average (standard error) means of matrices ℓ_1, ℓ_2, ℓ_F losses and 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. The best values are in bold font.

		ℓ_1		ℓ_2		ℓ_F		AUROC	
T	Method	$p = 50$	$p = 100$	$p = 50$	$p = 100$	$p = 50$	$p = 100$	$p = 50$	$p = 100$
10	DFGM-C	**16.6(0.04)**	23.3(0.36)	**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	16.7(0.06)	**23.3(0.36)**	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	21.1(0.06)	27.1(0.15)	18.2(0.03)	21.9(0.03)	86.7(0.14)	141.0(0.21)	8.1(0.02)	6.5(0.01)
	DGM-G	21.2(0.08)	26.9(0.14)	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	21.5(0.11)	40.0(0.58)	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	22.0(0.07)	41.0(0.11)	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	30.9(0.12)	41.8(0.11)	19.1(0.08)	34.8(0.08)	95.6(0.21)	194.7(0.85)	7.5(0.02)	6.3(0.01)
25	DFGM-C	**16.4(0.06)**	22.6(0.34)	**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	16.4(0.06)	**22.5(0.34)**	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	20.6(0.08)	26.4(0.15)	17.8(0.04)	21.5(0.03)	83.4(0.17)	138.9(0.20)	8.1(0.02)	6.8(0.01)
	DGM-G	20.6(0.07)	26.5(0.14)	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	22.7(0.15)	38.4(0.43)	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	25.9(0.17)	44.0(0.22)	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	27.7(0.36)	45.8(0.51)	23.6(0.33)	37.5(0.50)	103.1(0.85)	184.8(0.98)	8.0(0.02)	6.2(0.01)
50	DFGM-C	**15.9(0.06)**	**22.3(0.40)**	**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	16.0(0.06)	22.4(0.40)	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	19.8(0.08)	25.3(0.23)	16.9(0.04)	20.1(0.03)	78.9(0.15)	127.0(0.27)	8.5(0.02)	7.2(0.01)
	DGM-G	20.0(0.08)	25.3(0.20)	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	22.4(0.13)	37.7(0.62)	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	26.4(0.17)	39.1(0.21)	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	30.4(0.51)	40.4(0.55)	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.

Table 2 reports numerical summaries, averaged over 100 simulation runs, to compare different approaches over six simulation settings, corresponding to two different numbers of functional variables ($p = 50$ and $p = 100$) and three different levels of density for measurement schedules ($T = 10$, $T = 25$ and $T = 50$). 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 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, as one would expect, the best results are obtained for the more densely sampled case, with a smaller number of functional variables.

4.2 Irregular set of time points

When functions are observed at irregular sets of points, a common situation in practice, none of the three types of competitors described in Section 4.1 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 different p's ($p = 50$ and $p = 100$) and different numbers of observations for each trajectory (T_{ij}'s are generated from the discrete uniform distribution with sets $\{6, \ldots, 9\}$, $\{20, \ldots, 30\}$ and $\{50\}$). The measurement times are randomly sampled from Unif$[0, 1]$ for each observation and each functional variable. Thus, the observed points for (i, j) differ from those for (i', j') when $i \neq i'$ or $j \neq j'$. We average matrices ℓ_1, ℓ_2, ℓ_F 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 as in the previous table 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 in terms of estimation accuracy, but the methods are comparable in terms of the graph selection consistency.

Table 3: Average (standard error) means of matrices ℓ_1, ℓ_2, ℓ_F losses and AUROCs at $v_1, \ldots, v_{21}$ over 100 simulation runs. All entries have been multiplied by 10 for formatting reasons.

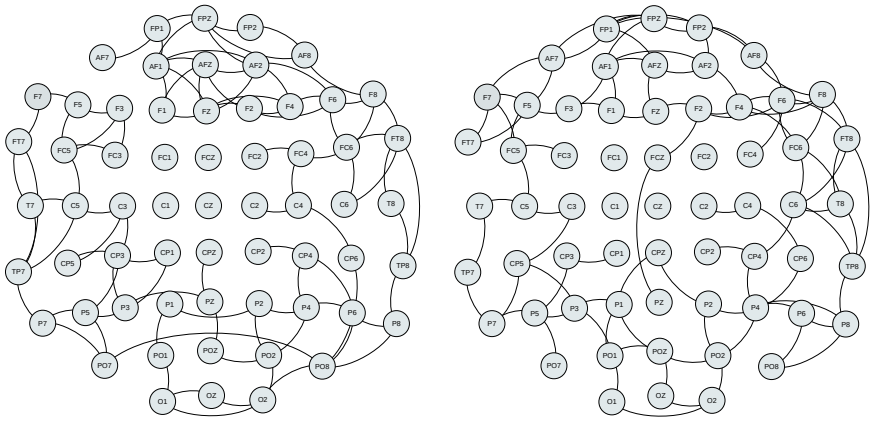
		ℓ_1		ℓ_2		ℓ_F		AUROC	
T_{ij}	Method	$p = 50$	$p = 100$	$p = 50$	$p = 100$	$p = 50$	$p = 100$	$p = 50$	$p = 100$
6-9	DFGM-C	32.1(0.18)	46.2(0.13)	25.9(0.15)	38.8(0.11)	122.4(0.59)	192.3(0.73)	7.6(0.02)	6.3(0.01)
	DFGM-G	32.3(0.21)	49.2(0.16)	26.0(0.16)	39.5(0.10)	128.5(0.49)	199.4(0.61)	7.6(0.02)	6.2(0.01)
20-30	DFGM-C	22.6(0.14)	27.3(0.22)	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	23.5(0.27)	28.5(0.28)	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	20.8(0.08)	26.1(0.15)	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	20.7(0.11)	26.6(0.12)	16.9(0.08)	21.9(0.10)	74.8(0.21)	138.6(0.32)	8.2(0.02)	6.8(0.01)

5 Real Data

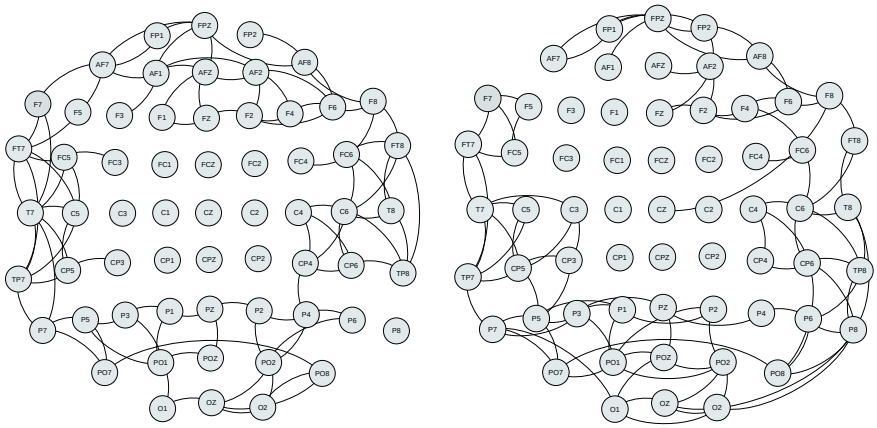
In this section, we apply our proposed approach to the EEG data set from an alcoholism study (Zhang et al., 1995), which is available at https://archive.ics.uci.edu/ml/datasets/EEG+Database. The data consists of 77 alcoholic and 45 control subjects. Each subject, exposed to either a single stimulus or 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 (Standard Electrode Position Nomenclature, American Electroencephalographic Association, 1990). Following the approach taken in Zhu et al. (2016) and Qiao et al. (2017), 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. We used the joint constrained ℓ_1-minimization approach (Cai et al., 2016) to simultaneously estimate two functional precision matrices. 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 for 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.

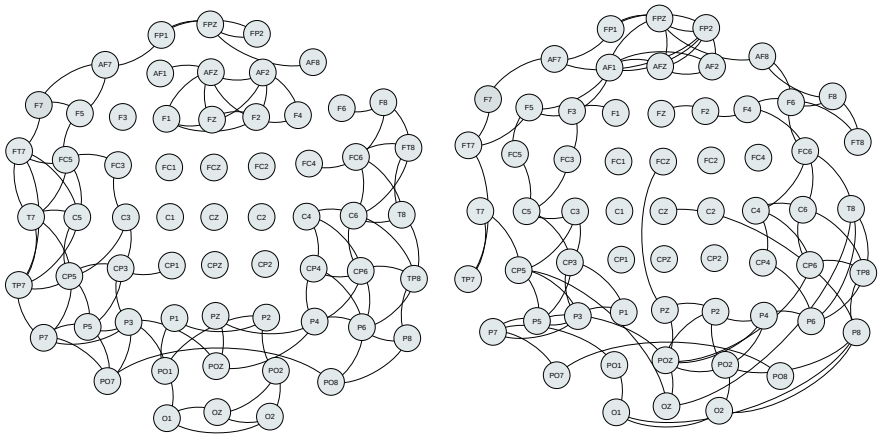
Figure 2 plots the estimated graphs for alcoholic and control groups at approximately $u = 0.2$, $u = 0.5$ and $u = 0.8$ seconds respectively. To visualize and interpret the functional network, we set the functional sparsity to 5% and only displayed the top 101 most important edges in Figure 2, where three anatomical landmarked electrodes, "X", "Y" and "nd", were removed. The animated adjacency matrices for $\widehat{\mathbf{\Theta}}(u)$ at 16 evenly spaced time points is available from http://personal.lse.ac.uk/qiaox/eeg_net.gif, where node names are provided in Table 4 in the Supplementary Material. We observe a few interesting patterns.



$u = 0.2$



$u = 0.5$



$u = 0.8$

Figure 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.*

First, the alcoholic and control groups share very similar block pattern, which reveals the existence of some regional effects for brain connectivity. Second, our estimated networks indicate clear dynamic structure. Those edge values within each block gradually change with some being bound to emerge and vanish over the evolution, e.g. 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, e.g. in Figure 2, electrodes from the left part of the brain are more connected in the alcoholic group.

Acknowledgements.

We are grateful to the Editor, Associate Editor and two referees for their useful comments and suggestions.

Supplementary material

Supplementary material available online includes all the technical proofs.

A Appendix

Appendix A.1 contains the procedure to generate simulated multivariate functional data. All the technical proofs and additional simulation results are provided in the Supplementary Material.

A.1 Procedure to generate simulated functional data

By the definition of $C_{jk}(u,v) = \mathrm{Cov}\big(X_{ij}(u), X_{ik}(v)\big) = \mathbf{s}(u)^{\mathrm{T}}\mathbf{\Omega}_{jk}\mathbf{s}(v)$ and orthonormality of $\mathbf{s}(u)$, we can easily show that

$$\mathbf{\Omega}_{jk} = \int_{(u,v)\in\mathcal{U}^2} \mathbf{s}(u)C_{jk}(u,v)\mathbf{s}(v)^{\mathrm{T}}dudv. \tag{16}$$

To generate multivariate functional observations, we first need to construct a valid matrix of cross-covariance operators $\big\{C_{jk}(u,v)\big\}_{1\leqslant j,k\leqslant p}$ (Guhaniyogi et al., 2013) and then obtain $\mathbf{\Omega}_{jk}$ from (16), by approximating the integral using the discretized sums, We take the idea of linear models of coregionalization (Genton and Kleiber, 2015) to represent the p-dimensional

multivariate random field as a linear combination of p independent univariate random fields by

$$C_{jk}(u,v) = \sum_{l=1}^{p} \rho(u-v) A_{jl}(u) A_{kl}(v), \tag{17}$$

where the correlation function $\rho(u-v)$ is independent of $l \in V$. Specially, for $u = v$, $\mathbf{\Sigma}(u) = \mathbf{C}(u,u) = \rho(0)\mathbf{A}(u)\mathbf{A}(u)^{\mathrm{T}}$, where we can set $\mathbf{A}(u) = \left\{A_{jk}(u)\right\}_{1 \leqslant j,k \leqslant p}$ to be the Cholesky factor of $\mathbf{\Sigma}(u)/\rho(0)$ and hence $\left\{C_{jk}(u,v)\right\}$ can be generated from (17) by letting $\rho(u-v) = \exp\{-(u-v)^2/2\sigma_\rho^2\}$, a univariate Gaussian kernel with $\sigma_\rho = 1$.

References

Bickel, P. and Levina, L. (2008). Covariance regularization by thresholding, *The Annals of Statistics* **136**: 2577–2604.

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

Cabral, J., Kringelbach, M. L. and Deco, G. (2014). Exploring the network dynamic underlying brain activity during rest, *Progress in Neurobiology* **114**: 102–131.

Cai, T., Li, H., Liu, W. and Xie, J. (2016). Joint estimation of multiple high-dimensional precision matrices, *Statistical Sinica* **26**: 445–464.

Cai, T., Liu, W. and 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. and Leng, C. (2016). Dynamic covariance models, *Journal of the American Statistical Association* **111**: 1196–1208.

Chun, H., Chen, M., Li, B. and Zhao, H. (2013). Joint conditional gaussian graphical models with multiple sources of genomic data, *Frontiers in Genetics* **4**: 294.

Dai, X., Muller, H. and Tao, W. (2017). Derivative principal component analysis for representing the time dynamics of longitudinal and functional data, *Statistica Sinica* .

Danaher, P., Wang, P. and Witten, D. (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., Hastie, T. and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso, *Biostatistics* **9**: 432–441.

Genton, M. G. and Kleiber, W. (2015). Cross-covariance functions for multivariate geostatistics, *Statistical Science* **30**: 147–163.

Guhaniyogi, R., Finley, A. O., Banerjee, S. and Kobe, R. (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. and Begleiter, H. (2006). Patterns of regional brain activity in alcohol-dependent subjects, *Alcoholism: Clinical and Experimental Research* **30**: 1986–1991.

Kolar, M. and 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. Oxford Statistical Science Series*, Oxford Univ. Press, New York.

Li, B., and Solea, E. (2017). A nonparametric graphical model for functional data with application to brain networks based on fmri, *Journal of the American Statistical Association* .

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

Meinshausen, N. and Buhlmann, P. (2010). Stability selection, *Journal of the Royal Statistical Society: Series B* **72**: 417–473.

Qiao, X., Guo, S. and James, G. (2017). Functional graphical models, *Journal of the American Statistical Association* .

Qiu, H., Han, F., Liu, H. and 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., Raskutti, G. and Yu, B. (2011). High-dimensional covariance estimation by minimizing l_1-penalized log-determinant divergence, *Electronic Journal of Statistics* **5**: 935–980.

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

Rice, J. and Wu, C. O. (2001). Nonparametric mixed effect models for unequally sampled noisy curves, *Biometrics* **57**: 253–259.

Ritter, K. G., Wasilkowski, W. and Wozniakowski, H. (1995). Multivariate integration and approximation for random fields satisfying sacks-ylvisaker conditions, *The Annals of Applied Probability* **5**: 518–540.

Storey, J. D., Xiao, W., Leek, T. J., Tompkins, R. G. and Davis, R. W. (2005). Significance analysis of time course microarray experiments, *Proceedings of the National Academy of Sciences* **102**: 12837–12842.

Witten, D., Friedman, J. and 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. and Wang, J. L. (2005). Functional data analysis for sparse longitudinal data, *Journal of the American Statistical Association* **100**: 577–590.

Yuan, M. and Lin, Y. (2007). Model selection and estimation in the gaussian graphical model, *Biometrika* **94**: 19–35.

Zhang, X., Begleiter, B., Porjesz, B., Wang, W. and Litke, A. (1995). Event related potentials during object recognition tasks, *Brain Research Bulletin* **38**: 531–538.

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

Zhou, S., Lafferty, J. and Wasserman, L. (2010). Time varying undirected graphs, *Machine Learning* **80**: 295–319.

Zhu, H., Stawn, N. and Dunson, D. (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, and Gareth M. James

This supplementary material contains the technical proofs supporting Section 3 in Appendix B and Table 4 supporting Section 5 in Appendix C.

B Technical proofs

In Sections B.1–B.2, we provide proofs of Theorems 1–2, respectively. In Section B.3, we provide proofs that, for fully observed functional data, (11) and (12) in Condition 2 hold.

B.1 Proof of Theorem 1

To prove Theorem 1, we will use Lemmas 1-19 with proofs as follows.

Lemma 1 *$\{\widehat{\mathbf{\Theta}}_1(u)\}$ be the solution to (6) and $\{\widehat{\mathbf{B}}(u)\} = \{(\widehat{\boldsymbol{\beta}}_1(u), \ldots, \widehat{\boldsymbol{\beta}}_p(u))\}$ with $\widehat{\boldsymbol{\beta}}_j(u)$'s being solutions to (7), then we have $\{\widehat{\mathbf{\Theta}}_1(u)\} = \{\widehat{\mathbf{B}}(u)\}$.*

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

Lemma 2 *Suppose that Condition 2 holds and $\delta \gg h^2$. Then there exists some positive constants C_1, C_2, C_3 such that for any $0 < \delta \leqslant C_1$, and for each $j = 1, \ldots, p$,*

$$P\Big(\big\|\widehat{C}_{jj}(u,v) - C_{jj}(u,v)\big\|_{\mathcal{S}} \geqslant \delta\Big) \leqslant C_2 \exp(-C_1 n^{2\gamma}\delta^2), \tag{B.1}$$

$$P\Big(\sup_{(u,v)\in\mathcal{U}^2} \big|\widehat{C}_{jj}(u,v) - C_{jj}(u,v)\big| \geqslant \delta\Big) \leqslant C_2 n^{C_3} \exp(-C_1 n^{2\gamma}\delta^2). \tag{B.2}$$

Proof. It follows from Theorem 4.2 of Zhang and Wang (2016) that, under some mild conditions, $\big\|E\big\{\widehat{C}_{jj}(u,v)\big\} - C_{jj}(u,v)\big\|_{\mathcal{S}} = O(h^2)$. By

$$\big\|\widehat{C}_{jj}(u,v) - C_{jj}(u,v)\big\|_{\mathcal{S}} \leqslant \big\|\widehat{C}_{jj}(u,v) - E\big\{\widehat{C}_{jj}(u,v)\big\}\big\|_{\mathcal{S}} + \big\|E\big\{\widehat{C}_{jj}(u,v)\big\} - C_{jj}(u,v)\big\|_{\mathcal{S}},$$

$\delta \gg h^2$ and Condition 2, we have $P\Big(\big\|\widehat{C}_{jj}(u,v) - C_{jj}(u,v)\big\|_{\mathcal{S}} \geqslant \delta\Big) \leqslant P\Big(\big\|\widehat{C}_{jj}(u,v) - E\big\{\widehat{C}_{jj}(u,v)\big\}\big\|_{\mathcal{S}} \geqslant \delta/2\Big) \leqslant C_2 \exp(-C_4 n^{2\gamma}\delta^2/4)$ with $C_1 = C_4/4$.

It follows from Theorem 5.2 of Zhang and Wang (2016) that $\sup_{(u,v)\in\mathcal{U}^2}\big|E\big\{\widehat{C}_{jj}(u,v)\big\} - C_{jj}(u,v)\big| = O(h^2)$. Following the same procedure, we can prove the concentration inequality in (B.2). For fully observed functional data, (B.1) reduce to Lemma 20 with $\gamma = 1/2$ and (B.2) reduces to Lemma 21 with $\gamma = 1/2, C_3 = 1$.

Lemma 3 *Suppose that Lemma 2 holds, there exists two positive constants C_1 and C_2 such that for any $0 < \delta \leqslant C_1$ and each $j = 1, \ldots, p$, $l = 1, \ldots, M$,*

$$P\left(|\widehat{\omega}_{jl} - \omega_{jl}| \geqslant \delta\right) \leqslant C_2 \exp(-C_1 n^{2\gamma}\delta^2).$$

Proof. By (4.43) of Bosq (2000), we have $\sup_{l\geqslant 1}|\widehat{\omega}_{jl} - \omega_{jl}| \leqslant ||\widehat{C}_{jj} - C_{jj}||_{\mathcal{S}}$. It then follows from (B.1) that for each $l = 1, \ldots, M$,

$$P\left(|\widehat{\omega}_{jl} - \omega_{jl}| \geqslant \delta\right) \leqslant P\left(||\widehat{C}_{jj} - C_{jj}||_{\mathcal{S}} \geqslant \delta\right) \leqslant C_2 \exp(-C_1 n^{2\gamma}\delta^2),$$

which completes the our proof.

Lemma 4 *Denote $\breve{\phi}_{jl} = sign < \widehat{\phi}_{jl}, \phi_{jl} > \phi_{jl}$. 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 2 holds, there exist two positive constants C_1 and C_2 such that for any $0 < \delta \leqslant C_1$ and each $j = 1, \ldots, p$, $l = 1, \ldots, M$,

$$P\left(||\widehat{\phi}_{jl} - \phi_{jl}|| \geqslant \delta\right) \leqslant C_2 \exp\left(-C_1 n^{2\gamma} l^{-2(\beta+1)}\delta^2\right).$$

Proof. The first part can be found in Lemma 4.3 of Bosq (2000). Since $\omega_{jl} \asymp l^{-\beta}$ and $d_{jl}\omega_{jl} = O(l)$, we have $d_{jl}^{-1} \geqslant d_0 l^{-(1+\beta)}$, where d_0 is some positive constant. It then follows from (B.1) that for each $l = 1, \ldots, M$,

$$P\left(||\widehat{\phi}_{jl} - \phi_{jl}|| \geqslant \delta\right) \leqslant P\left(||\widehat{C}_{jj} - C_{jj}||_{\mathcal{S}} \geqslant d_{jl}^{-1}\delta\right) \leqslant C_2 \exp(-C_3 n^{2\gamma} d_{jl}^{-2}\delta^2),$$

which completes the proof by setting $C_1 = C_3 d_0^2$.

Note that the concentration inequalities in Lemmas 3–4, which describe the exponential tail behaviours of estimated eigen-pairs, are derived based on the probability measure equipped from the exponential concentration rates in (B.1) and (B.2). To simplify our further derivation and exposition, we will derive the relevant convergence rates under the same probability measure as considered in Lemma 2.

Lemma 5 *Suppose that Condition 3 (ii) holds, then* $\max_{j \in V} \sup_{(u,v) \in \mathcal{U}^2} |C_{jj}(u,v)| = O(1)$.

Proof. Given the spectral decomposition $C_{jj}(u,v) = \sum_{k=1}^{\infty} \omega_{jk} \phi_{jk}(u) \phi_{jk}(v)$ for each $j \in V$, we have $\max_{j \in V} \sup_{(u,v) \in \mathcal{U}^2} |C_{jj}(u,v)| \leqslant \max_{j \in V} \sup_{k \geqslant 1, (u,v) \in \mathcal{U}^2} |\phi_{jk}(u)||\phi_{jk}(v)| \max_{j \in V} \sum_{k=1}^{\infty} \omega_{jk} = O(1)$.

Lemma 6 *Suppose that Lemma 2 and Condition 3 hold. Then there exist two positive constants* C_1 *and* C_2 *such that for any* $0 < \delta \leqslant C_1$ *and each* $j = 1, \ldots, p$, $l = 1, \ldots, M$,

$$\sup_{u \in \mathcal{U}} |\widehat{\phi}_{jl}(u) - \phi_{jl}(u)| = O_p\big\{ l^{\beta} (\log n)^{1/2} n^{-\gamma} + l^{2\beta+1} n^{-\gamma} \big\}.$$

Proof. For each $j = 1, \ldots, p$ and $l = 1, \ldots, M$, it follows from the corresponding 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

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

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

$$\sup_u |S_1(u)| \leqslant \omega_{jl}^{-1} |\widehat{\omega}_{jl} - \omega_{jl}| \sup_u |\phi_{jl}(u)| = O_p\big(l^{\beta} n^{-\gamma}\big).$$

(ii) It follows from Condition 3 and Lemma 2 that

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

(iii) It follows from Condition 3, Cauchy-Schwartz inequality and Lemmas 4–5 that

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

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

Lemma 7 *Suppose that Lemma 2 and Condition 3 hold. Then for each* $i = 1, \ldots, n$ *and* $j = 1, \ldots, p$,

$$||\widehat{\boldsymbol{\zeta}}_{ijl} - \boldsymbol{\zeta}_{ijl}|| = O_p\big(T_{ij}^{1/2} l n^{-\gamma}\big), \tag{B.3}$$

$$||\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}_{ij}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{Y}_{ij}}^{-1}|| = O_p\big\{T_{ij} (\log n)^{1/2} n^{-\gamma}\big\}. \tag{B.4}$$

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

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

where the fourth line follows from Cauchy-Schwartz inequality and the last line follows from Condition 3 and Lemmas 3–4. This completes our proof for (B.3).

From the definition of $\boldsymbol{\Sigma}_{\mathbf{Y}_{ij}} = \boldsymbol{\Sigma}_{\mathbf{X}_{ij}} + \sigma^2 \mathbf{I}_{T_{ij}}$, where $\mathbf{X}_{ij} = (X_{ij}(U_{ijT_1}), \ldots, X_{ij}(U_{ijT_{ij}}))^{\mathrm{T}}$, we have $||\boldsymbol{\Sigma}_{\mathbf{Y}_{ij}}^{-1}|| \leqslant \sigma^{-2}$. Applying Lemma 1 of Dai et al. (2017), we have $||\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}_{ij}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{Y}_{ij}}^{-1}|| \leqslant c\sigma^{-4} ||\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}_{ij}} - \boldsymbol{\Sigma}_{\mathbf{Y}_{ij}}|| \lesssim c\sigma^{-4} T_{ij} \sup_{t,t'} |(\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}_{ij}})_{tt'} - (\boldsymbol{\Sigma}_{\mathbf{Y}_{ij}})_{tt'}|$. Moreover, $\sup_{t,t'} |(\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}_{ij}})_{tt'} - (\boldsymbol{\Sigma}_{\mathbf{Y}_{ij}})_{tt'}| \leqslant |\widehat{\sigma}^2 - \sigma^2| + \sup_{u,v} |\widehat{C}_{jj}(u,v) - C_{jj}(u,v)|$, where the first term is dominated by the second term, see Corollary 1 of Yao et al. (2005) for details. We apply (B.2) in Lemma 2 and hence can obtain the convergence rate in (B.4).

Lemma 8 *Suppose that Conditions1 3 hold. Then for each* $i = 1, \ldots, n$, $j, k = 1, \ldots, p$ *and* $l, m = 1, \ldots, M$,

$$|\widehat{\xi}_{ijl} \widehat{\xi}_{ikm} - \widetilde{\xi}_{ijl} \widetilde{\xi}_{ikm}| = O_p\Big\{T_{ij} T_{ik} (l m^{-\beta} + l^{-\beta} m) n^{-\gamma} + (T_{ij}^2 T_{ik} + T_{ij} T_{ik}^2) l^{-\beta} m^{-\beta} (\log n)^{1/2} n^{-\gamma}\Big\}.$$

Furthermore, for sparse functional design with $T_{ij} \leqslant T_0 < \infty$, we have

$$|\widehat{\xi}_{ijl}\widehat{\xi}_{ikm} - \widetilde{\xi}_{ijl}\widetilde{\xi}_{ikm}| = O_p\Big\{lm^{-\beta}n^{-\gamma} + l^{-\beta}mn^{-\gamma} + l^{-\beta}m^{-\beta}(\log n)^{1/2}n^{-\gamma}\Big\}, \tag{B.5}$$

and for dense case with $T_{ij} \asymp T$, we have

$$|\widehat{\xi}_{ijl}\widehat{\xi}_{ikm} - \widetilde{\xi}_{ijl}\widetilde{\xi}_{ikm}| = O_p\Big\{T^2lm^{-\beta}n^{-\gamma} + T^2l^{-\beta}mn^{-\gamma} + T^3l^{-\beta}m^{-\beta}(\log n)^{1/2}n^{-\gamma}\Big\}. \tag{B.6}$$

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

$$\begin{aligned}
= \;& (\widehat{\boldsymbol{\zeta}}_{ijl} - \boldsymbol{\zeta}_{ijl})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{Y}_{ij}}^{-1}\mathbf{Y}_{ij}\boldsymbol{\zeta}_{ikm}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{Y}_{ik}}^{-1}\mathbf{Y}_{ik} + (\widehat{\boldsymbol{\zeta}}_{ijl} - \boldsymbol{\zeta}_{ijl})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{Y}_{ij}}^{-1}\mathbf{Y}_{ij}\boldsymbol{\zeta}_{ikm}^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}_{ik}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{Y}_{ik}}^{-1})\mathbf{Y}_{ik} \\
& + (\widehat{\boldsymbol{\zeta}}_{ijl} - \boldsymbol{\zeta}_{ijl})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{Y}_{ij}}^{-1}\mathbf{Y}_{ij}(\widehat{\boldsymbol{\zeta}}_{ikm} - \boldsymbol{\zeta}_{ikm})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{Y}_{ik}}^{-1}\mathbf{Y}_{ik} \\
& + (\widehat{\boldsymbol{\zeta}}_{ijl} - \boldsymbol{\zeta}_{ijl})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{Y}_{ij}}^{-1}\mathbf{Y}_{ij}(\widehat{\boldsymbol{\zeta}}_{ikm} - \boldsymbol{\zeta}_{ikm})^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}_{ik}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{Y}_{ik}}^{-1})\mathbf{Y}_{ik} \\
& + (\widehat{\boldsymbol{\zeta}}_{ijl} - \boldsymbol{\zeta}_{ijl})^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}_{ij}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{Y}_{ij}}^{-1})\mathbf{Y}_{ij}\boldsymbol{\zeta}_{ikm}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{Y}_{ik}}^{-1}\mathbf{Y}_{ik} \\
& + (\widehat{\boldsymbol{\zeta}}_{ijl} - \boldsymbol{\zeta}_{ijl})^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}_{ij}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{Y}_{ij}}^{-1})\mathbf{Y}_{ij}\boldsymbol{\zeta}_{ikm}^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}_{ik}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{Y}_{ik}}^{-1})\mathbf{Y}_{ik} \\
& + (\widehat{\boldsymbol{\zeta}}_{ijl} - \boldsymbol{\zeta}_{ijl})^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}_{ij}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{Y}_{ij}}^{-1})\mathbf{Y}_{ij}(\widehat{\boldsymbol{\zeta}}_{ikm} - \boldsymbol{\zeta}_{ikm})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{Y}_{ik}}^{-1}\mathbf{Y}_{ik} \\
& + (\widehat{\boldsymbol{\zeta}}_{ijl} - \boldsymbol{\zeta}_{ijl})^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}_{ij}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{Y}_{ij}}^{-1})\mathbf{Y}_{ij}(\widehat{\boldsymbol{\zeta}}_{ikm} - \boldsymbol{\zeta}_{ikm})^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}_{ik}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{Y}_{ik}}^{-1})\mathbf{Y}_{ik} \\
& + \boldsymbol{\zeta}_{ijl}^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}_{ij}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{Y}_{ij}}^{-1})\mathbf{Y}_{ij}\boldsymbol{\zeta}_{ikm}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{Y}_{ik}}^{-1}\mathbf{Y}_{ik} + \boldsymbol{\zeta}_{ijl}^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}_{ij}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{Y}_{ij}}^{-1})\mathbf{Y}_{ij}\boldsymbol{\zeta}_{ikm}^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}_{ik}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{Y}_{ik}}^{-1})\mathbf{Y}_{ik} \\
& + \boldsymbol{\zeta}_{ijl}^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}_{ij}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{Y}_{ij}}^{-1})\mathbf{Y}_{ij}(\widehat{\boldsymbol{\zeta}}_{ikm} - \boldsymbol{\zeta}_{ikm})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{Y}_{ik}}^{-1}\mathbf{Y}_{ik} \\
& + \boldsymbol{\zeta}_{ijl}^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}_{ij}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{Y}_{ij}}^{-1})\mathbf{Y}_{ij}(\widehat{\boldsymbol{\zeta}}_{ikm} - \boldsymbol{\zeta}_{ikm})^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}_{ik}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{Y}_{ik}}^{-1})\mathbf{Y}_{ik} \\
& + \boldsymbol{\zeta}_{ijl}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{Y}_{ij}}^{-1}\mathbf{Y}_{ij}(\widehat{\boldsymbol{\zeta}}_{ikm} - \boldsymbol{\zeta}_{ikm})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{Y}_{ik}}^{-1}\mathbf{Y}_{ik} + \boldsymbol{\zeta}_{ijl}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{Y}_{ij}}^{-1}\mathbf{Y}_{ij}(\widehat{\boldsymbol{\zeta}}_{ikm} - \boldsymbol{\zeta}_{ikm})^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}_{ik}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{Y}_{ik}}^{-1})\mathbf{Y}_{ik} \\
& + \boldsymbol{\zeta}_{ijl}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{Y}_{ij}}^{-1}\mathbf{Y}_{ij}\boldsymbol{\zeta}_{ikm}^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}_{ik}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{Y}_{ik}}^{-1})\mathbf{Y}_{ik} \\
= \;& I_1 + I_2 + \cdots + I_{15}.
\end{aligned}$$

We first list several results that can be used in the proof for the convergence rate. (i) $||\boldsymbol{\zeta}_{ijl}|| \leqslant T_{ij}^{1/2}\sup_t|\zeta_{ijlt}| = T_{ij}^{1/2}\sup_t|\int C_{jj}(U_{ijt}, v)\phi_{jl}(v)dv| \lesssim T_{ij}^{1/2}\sup_{u\in\mathcal{U}}|\int C_{jj}(u, v)\phi_{jl}(v)dv| = T_{ij}^{1/2}\sup_{u\in\mathcal{U}}|\int \omega_{jl}\phi_{jl}(u)| = O_p(T_{ij}^{1/2}l^{-\beta})$, where the last equality follows from Condition 3; (ii) $||\boldsymbol{\Sigma}_{\mathbf{Y}_{ij}}^{-1}|| \leqslant \sigma^{-2}$; (iii) $||\mathbf{Y}_{ij}|| \leqslant ||\mathbf{e}_{ij}|| + ||\mathbf{X}_{ij}|| = O_p(T_{ij}^{1/2})$, where $\mathbf{e}_{ij} = \big(e_{ij}(U_{ij1}), \ldots, e_{ij}(U_{ijT_{ij}})\big)^{\mathrm{T}}$.

Applying the above results in (i), (ii), (iii) and (B.3), (B.4) in Lemma 7, we have

$$||I_1|| \leqslant ||\widehat{\boldsymbol{\zeta}}_{ijl} - \boldsymbol{\zeta}_{ijl}||\;||\boldsymbol{\Sigma}_{\mathbf{Y}_{ij}}^{-1}||\;||\mathbf{Y}_{ij}||\;||\boldsymbol{\zeta}_{ikm}||\;||\boldsymbol{\Sigma}_{\mathbf{Y}_{ik}}^{-1}||\;||\mathbf{Y}_{ik}|| = O_p(T_{ij}T_{ik}lm^{-\beta}n^{-\gamma}),$$

$$|I_9| \leqslant ||\boldsymbol{\zeta}_{ijl}||\ ||\widehat{\boldsymbol{\Sigma}}^{-1}_{\mathbf{Y}_{ij}} - \boldsymbol{\Sigma}^{-1}_{\mathbf{Y}_{ij}}||\ ||\mathbf{Y}_{ij}||\ ||\boldsymbol{\zeta}_{ikm}||\ ||\boldsymbol{\Sigma}^{-1}_{\mathbf{Y}_{ik}}||\ ||\mathbf{Y}_{ik}|| = O_p\{T_{ij}^2 T_{ik} l^{-\beta} m^{-\beta} (\log n)^{1/2} n^{-\gamma}\},$$

$$||I_{13}|| \leqslant ||\boldsymbol{\zeta}_{ijl}||\ ||\boldsymbol{\Sigma}^{-1}_{\mathbf{Y}_{ij}}||\ ||\mathbf{Y}_{ij}||\ ||\widehat{\boldsymbol{\zeta}}_{ikm} - \boldsymbol{\zeta}_{ikm}||\ ||\boldsymbol{\Sigma}^{-1}_{\mathbf{Y}_{ik}}||\ ||\mathbf{Y}_{ik}|| = O_p(T_{ij} T_{ik} l^{-\beta} m n^{-\gamma}),$$

$$||I_{15}|| \leqslant ||\boldsymbol{\zeta}_{ijl}||\ ||\boldsymbol{\Sigma}^{-1}_{\mathbf{Y}_{ij}}||\ ||\mathbf{Y}_{ij}||\ ||\boldsymbol{\zeta}_{ikm}||\ ||\widehat{\boldsymbol{\Sigma}}^{-1}_{\mathbf{Y}_{ik}} - \boldsymbol{\Sigma}^{-1}_{\mathbf{Y}_{ik}}||\ ||\mathbf{Y}_{ik}|| = O_p\{T_{ij} T_{ik}^2 l^{-\beta} m^{-\beta} (\log n)^{1/2} n^{-\gamma}\}.$$

Since these four terms can dominate the other $||I_i||$ terms, we combine the derived convergence rates and thus complete our proof. For the sparse case with $T \leqslant T_0 < \infty$, we obtain (B.5) and for the dense case with $T_{ij} \asymp T$, we obtain (B.6).

Lemma 9 *Suppose that the dense case in Condition 1 and Condition 3 hold. Then for each $j, k = 1, \ldots, p$ and $l, m = 1, \ldots, M$,*

$$|\widetilde{\xi}_{ijl}\widetilde{\xi}_{ikm} - \xi_{ijl}\xi_{ikm}| = O_p\big(T^{-1/2} l^{-\beta/2} + T^{-1/2} m^{-\beta/2}\big).$$

Proof. Under joint Gaussian assumption for ξ_{ijl} and e_{ijt}, it follows from Section 2.4 of Yao et al. (2005) that $\xi_{ijl} - \widetilde{\xi}_{ijl} \sim N\big(0, \mathrm{Var}(\widetilde{\xi}_{ijl}) - \mathrm{Var}(\xi_{ijl})\big)$. Given that $\widetilde{\xi}_{ijl} = \boldsymbol{\zeta}_{ijl}^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}_{\mathbf{Y}_{ij}} \mathbf{Y}_{ij} = \omega_{jl} \boldsymbol{\phi}_{jl}^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}_{\mathbf{Y}_{ij}} \mathbf{Y}_{ij}$, we obtain $\mathrm{Var}(\xi_{ijl}) - \mathrm{Var}(\widetilde{\xi}_{ijl}) = \omega_{jl} - \omega_{jl}^2 \boldsymbol{\phi}_{jl}^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}_{\mathbf{Y}_{ij}} \boldsymbol{\phi}_{jl}$.

Note that $\int_{\mathcal{U}} \phi_{jl}(u)\phi_{jl'}(u)du = I(l = l')$, we use the discretized sum to approximate the integral and hence obtain $\boldsymbol{\phi}_{jl}^{\mathrm{T}} \boldsymbol{\phi}_{jl'}/T \asymp I(l = l')$. Moreover, given that $\boldsymbol{\Sigma}_{\mathbf{Y}_{ij}} = \boldsymbol{\Sigma}_{\mathbf{X}_{ij}} + \sigma^2 \mathbf{I}_{T_{ij}} \asymp \sum_{l=1}^{\infty} \omega_{jl} \boldsymbol{\phi}_{jl} \boldsymbol{\phi}_{jl}^{\mathrm{T}} + \sum_{l=1}^{\infty} \sigma^2 \boldsymbol{\phi}_{jl} \boldsymbol{\phi}_{jl}^{\mathrm{T}}/T$, we have $\boldsymbol{\Sigma}^{-1}_{\mathbf{Y}_{ij}} \asymp \sum_{l=1}^{\infty} (\omega_{jl} + \sigma^2/T)^{-1} \boldsymbol{\phi}_{jl} \boldsymbol{\phi}_{jl}^{\mathrm{T}}/T^2$. Using these results above, we have $\mathrm{Var}(\xi_{ijl}) - \mathrm{Var}(\widetilde{\xi}_{ijl}) = \omega_{jl} - \omega_{jl}^2 \boldsymbol{\phi}_{jl}^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}_{\mathbf{Y}_{ij}} \boldsymbol{\phi}_{jl} \asymp \omega_{jl} - \omega_{jl}^2 \boldsymbol{\phi}_{jl}^{\mathrm{T}} \big(\sum_{l'=1}^{\infty} (\omega_{jl'} + \sigma^2/T)^{-1} \boldsymbol{\phi}_{jl'} \boldsymbol{\phi}_{jl'}^{\mathrm{T}} \big) \boldsymbol{\phi}_{jl}/T^2 \asymp \omega_{jl} - \omega_{jl}^2 (\omega_{jl} + \sigma^2/T)^{-1} = \omega_{jl} \sigma^2 T^{-1} (\omega_{jl} + \sigma^2/T)^{-1} < T^{-1} \sigma^2$. By the tail probability bound for a Gaussian random variable, for $0 < \delta \leqslant C_1$, we have $P(|\widetilde{\xi}_{ijl} - \xi_{ijl}| \geqslant \delta) \leqslant C_2 \exp(-C_1 T \delta^2)$, where C_1 and C_2 are two positive constants. Equivalently, $|\widetilde{\xi}_{ijl} - \xi_{ijl}| = O_p(T^{-1/2})$. Applying the similar technique leads to $\mathrm{Var}(\xi_{ijl}) \asymp \omega_{jl}^2 (\omega_{jl} + \sigma^2/T)^{-1} < \omega_{jl} \asymp l^{-\beta}$ and $\xi_{ijl} = O_p(l^{-\beta/2})$ follows.

By the above results with the expansion $\widetilde{\xi}_{ijl}\widetilde{\xi}_{ikm} - \xi_{ijl}\xi_{ikm} = (\widetilde{\xi}_{ijl} - \xi_{ijl})\widetilde{\xi}_{ikm} + \xi_{ijl}(\widetilde{\xi}_{ikm} - \xi_{ikm})$, we have $|\widetilde{\xi}_{ijl}\widetilde{\xi}_{ikm} - \xi_{ijl}\xi_{ikm}| = O_p(T^{-1/2} l^{-\beta/2} + T^{-1/2} m^{-\beta/2})$, which completes our proof.

Lemma 10 *Suppose that Condition 3 holds. Then for each $j, k = 1, \ldots, p$ and $l, m = 1, \ldots, M$,*

$$\Big| \frac{1}{n} \sum_{i=1}^{n} \xi_{ijl}\xi_{ikm} - E(\xi_{ijl}\xi_{ikm}) \Big| = O_p(l^{-\beta/2} m^{-\beta/2} n^{-1/2}).$$

Proof. Let $\xi_{ijl} = \omega_{jl}^{1/2} a_{ijl}$, where $a_{ijl} \sim N(0,1)$. It follows from the proof for Theorem 1 in Qiao et al. (2017) that $|\frac{1}{n}\sum_{i=1}^{n} \xi_{ijl}\xi_{ikm} - E(\xi_{ijl}\xi_{ikm})| = \omega_{jl}^{1/2}\omega_{km}^{1/2}|\sum_{i=1}^{n} a_{ijl}a_{ikm} - E(a_{ijl}a_{ikm})| = l^{-\beta/2}m^{-\beta/2}O_p(n^{-1/2})$, which completes our proof.

Lemma 11 *Suppose that the dense case in Condition 1 and Conditions 2–3 hold. Then for each $j,k = 1,\ldots,p$ and $l,m = 1,\ldots,M$,*

$$\begin{aligned} & \Big|\frac{1}{n}\sum_{i=1}^{n} \widehat{\xi}_{ijl}\widehat{\xi}_{ikm} - E(\xi_{ijl}\xi_{ikm})\Big| \\ = \; & O_p\big\{T^2 l m^{-\beta} n^{-\gamma} + T^2 l^{-\beta} m n^{-\gamma} + T^3 l^{-\beta} m^{-\beta} (\log n)^{1/2} n^{-\gamma} \\ & + T^{-1/2} l^{-\beta/2} + T^{-1/2} m^{-\beta/2} + l^{-\beta/2} m^{-\beta/2} n^{-1/2}\big\}. \end{aligned}$$

Proof. By the expansion of $\frac{1}{n}\sum_{i=1}^{n} \widehat{\xi}_{ijl}\widehat{\xi}_{ikm} - E(\xi_{ijl}\xi_{ikm}) = \frac{1}{n}\sum_{i=1}^{n}\big\{(\widehat{\xi}_{ijl}\widehat{\xi}_{ikm} - \widetilde{\xi}_{ijl}\widetilde{\xi}_{ikm}) + (\widetilde{\xi}_{ijl}\widetilde{\xi}_{ikm} - \xi_{ijl}\xi_{ikm}) + (\xi_{ijl}\xi_{ikm} - E(\xi_{ijl}\xi_{ikm}))\big\}$ and the results in Lemmas 8–10, we can immediately obtain the convergence rate in Lemma 11.

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

Lemma 12 *Suppose that the dense case in Condition 1 and Condition 3 hold. Then for each $j,k = 1,\ldots,p$,*

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

Proof. It follows from Cauchy-Schwartz inequality $|\sigma_{jlkm}| = |E[\xi_{jl}\xi_{km}]| \leqslant \sqrt{E(\xi_{jl}^2)E(\xi_{km}^2)} = w_{jl}^{1/2} w_{km}^{1/2}$ and Condition 3 that the truncation error can be bounded as

$$\begin{aligned} \sup_u |\Sigma_{jk}(u) - \Sigma_{jk,M}(u)| &\leqslant \Big(\sum_{l=1}^{\infty}\sum_{m=1}^{\infty} - \sum_{l=1}^{M}\sum_{m=1}^{M}\Big)|\sigma_{jlkm}| \sup_u |\phi_{jl}(u)| \sup_u |\phi_{km}(u)|, \\ &\leqslant \Big(\sum_{l=1}^{\infty}\sum_{m=1}^{\infty} - \sum_{l=1}^{M}\sum_{m=1}^{M}\Big)\omega_{jl}^{1/2}\omega_{km}^{1/2} O(1) \\ &= O\Big(\sum_{l=1}^{M}\sum_{m=M+1}^{\infty} l^{-\beta/2}\omega_{km}^{1/2}\Big) \leqslant O(M^{-\nu}) \asymp n^{-\alpha\nu}, \end{aligned}$$

where the last line comes from Condition 3 and the fact that

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

Lemma 13 *Suppose that the dense case in Condition 1 and Conditions 2–3 hold. Then there exist some positive constants C_1, C_2, C_3 such that for any δ with $0 < \delta \leqslant C_1$ and each $j, k = 1, \ldots, p$,*

$$\begin{aligned} & P\big(\sup_{u\in\mathcal{U}}|\widehat{\Sigma}_{jk}(u) - \Sigma_{jk,M}(u)| > \delta\big) \\ \leqslant\ & C_2 n^{C_3}\exp\big\{-C_1 n^{2\gamma-\alpha(\beta+2)}\delta^2\big\} + C_2\exp\big\{-C_1 n^{2\gamma-(3\beta+4)}\delta^2\big\} \\ & + C_2\exp\big\{-C_1 T^{-4} n^{2\gamma-4\alpha}\delta^2\big\} + C_2 n^{C_3}\exp\big\{-C_1 T^{-6} n^{2\gamma}\delta^2\big\} + C_2\exp\big\{-C_1 T n^{-2\alpha}\delta^2\big\}. \end{aligned}$$

Proof. Note

$$\widehat{\Sigma}_{jk}(u) - \Sigma_{jk,M}(u) = \sum_{l=1}^{M}\sum_{m=1}^{M}\big\{\widehat{\phi}_{jl}(u)\widehat{\sigma}_{jlkm}\widehat{\phi}_{km}(u) - \phi_{jl}(u)\sigma_{jlkm}\phi_{km}(u)\big\}. \tag{B.7}$$

We can write

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

We first bound $I_1(u)$. It follows from Cauchy-Schwartz inequality on $|\sigma_{jlkm}|$, Condition 3 and Lemma 6 that

$$\begin{aligned} \sup_u|I_1(u)| \ &\leqslant\ \big\{\sup_u|\phi_{jl}(u)| + \sup_u|\widehat{\phi}_{jl}(u) - \phi_{jl}(u)|\big\}\big\{|\sigma_{jlkm}| + |\widehat{\sigma}_{jlkm} - \sigma_{jlkm}|\big\}\sup_u|\widehat{\phi}_{km}(u) - \phi_{km}(u)| \\ &= O_p\big[(l^{-\beta/2}m^{-\beta/2})\big\{m^{\beta}(\log n)^{1/2}n^{-\gamma} + m^{2\beta+1}n^{-\gamma}\big\}\big] \\ &= O_p\big\{l^{-\beta/2}m^{\beta/2}(\log n)^{1/2}n^{-\gamma} + l^{-\beta/2}m^{3\beta/2+1}n^{-\gamma}\big\}. \end{aligned}$$

We next bound $I_2(u)$. It follows from Condition 3 and Lemma 10 that

$$\begin{aligned} \sup_u|I_2(u)| \ &\leqslant\ \big\{\sup_u|\phi_{jl}(u)| + \sup_u|\widehat{\phi}_{jl}(u) - \phi_{jl}(u)|\big\}|\widehat{\sigma}_{jlkm} - \sigma_{jlkm}|\sup_u|\phi_{km}(u)| \\ &= O_p\big\{T^2 l m^{-\beta}n^{-\gamma} + T^2 l^{-\beta}mn^{-\gamma} + T^3 l^{-\beta}m^{-\beta}(\log n)^{1/2}n^{-\gamma} \\ &\quad + T^{-1/2}l^{-\beta/2} + T^{-1/2}m^{-\beta/2} + l^{-\beta/2}m^{-\beta/2}n^{-1/2}\big\}. \end{aligned}$$

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

$$\sup_u|I_3(u)| = O_p\big\{l^{\beta/2}m^{-\beta/2}(\log n)^{1/2}n^{-\gamma} + l^{3\beta/2+1}m^{-\beta/2}n^{-\gamma}\big\}.$$

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

$$\begin{aligned}
& \sup_u |\widehat{\Sigma}_{jk}(u) - \Sigma_{jk,M}(u)| \\
\leqslant\ & \sum_{l=1}^{M}\sum_{m=1}^{M} \sup_u |\widehat{\phi}_{jl}(u)\widehat{\sigma}_{jlkm}\widehat{\phi}_{km}(u) - \phi_{jl}(u)\sigma_{jlkm}\phi_{km}(u)| \\
\leqslant\ & O_p\Big[\sum_{l=1}^{M}\sum_{m=1}^{M} \big\{l^{-\beta/2}m^{\beta/2}(\log n)^{1/2}n^{-\gamma} + l^{-\beta/2}m^{3\beta/2+1}n^{-\gamma} + T^2 l^{-\beta}mn^{-\gamma} \\
& + T^3 l^{-\beta}m^{-\beta}(\log n)^{1/2}n^{-\gamma} + l^{-\beta/2}T^{-1/2} + l^{-\beta/2}m^{-\beta/2}n^{-1/2}\big\}\Big] \\
=\ & O_p\big\{M^{\beta/2+1}(\log n)^{1/2}n^{-\gamma} + M^{3\beta/2+2}n^{-\gamma} + T^2M^2n^{-\gamma} + T^3(\log n)^{1/2}n^{-\gamma} + MT^{-1/2}\big\}.
\end{aligned}$$

where the last line comes from

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

and other similar inequalities. By $M \asymp n^{\alpha}$ and the probability measures equipped with exponential decay tail bounds used in Lemmas 2 and 9, we can obtain the concentration inequality with suitable choices of C_1, C_2, C_3.

Lemma 14 *Suppose that the dense case in Condition 1 and Conditions 2–3 hold. Let* $\kappa_{n,T} = n^{\gamma-\alpha(3\beta/2+2)} \wedge T^{-3}n^{\gamma} \wedge T^{1/2}n^{-\alpha}$. *For sufficiently large* M', *if* $\delta = M'\big\{(\log p/\kappa_{n,T}^2)^{1/2} + (\log p/n^{2\alpha\nu})^{1/2}\big\}$, $\log p/\kappa_{n,T}^2 \to 0$ *and* $\log p/n^{2\alpha\nu} \to 0$, *then under high dimensional setting with* $p \gtrsim n$, *we have*

$$\operatorname*{supmax}_{u\in\mathcal{U}\ j,k} |\widehat{\Sigma}_{jk}(u) - \Sigma_{jk}(u)| = O_p\left\{\left(\frac{\log p}{\kappa_{n,T}^2}\right)^{1/2} + \left(\frac{\log p}{n^{2\alpha\nu}}\right)^{1/2}\right\}. \tag{B.8}$$

Proof. It follows from the triangular inequality, $\sup_{u\in\mathcal{U}}|\widehat{\Sigma}_{jk}(u) - \Sigma_{jk}(u)| \leqslant \sup_{u\in\mathcal{U}}|\widehat{\Sigma}_{jk}(u) - \Sigma_{jk,M}(u)| + \sup_{u\in\mathcal{U}}|\Sigma_{jk,M}(u) - \Sigma_{jk}(u)|$, $\delta \gg n^{-\alpha\nu}$, Lemmas 12 and 13 that there exist some positive constants C_1, C_2, C_3 such that

$$\begin{aligned}
& P\big(\sup_{u\in\mathcal{U}}|\widehat{\Sigma}_{jk}(u) - \Sigma_{jk}(u)| > \delta\big) \\
\leqslant\ & P\big(\sup_{u\in\mathcal{U}}|\widehat{\Sigma}_{jk}(u) - \Sigma_{jk,M}(u)| > \delta/2\big) \\
\leqslant\ & C_2n^{C_3}\exp\big\{-C_1n^{2\gamma-\alpha(\beta+2)}\delta^2\big\} + C_2\exp\big\{-C_1n^{2\gamma-\alpha(3\beta+4)}\delta^2\big\} \\
& + C_2\exp\big\{-C_1T^{-4}n^{2\gamma-4\alpha}\delta^2\big\} + C_2n^{C_3}\exp\big\{-C_1T^{-6}n^{2\gamma}\delta^2\big\} + C_2\exp\big\{-C_1Tn^{-2\alpha}\delta^2\big\}
\end{aligned}$$

Applying the union bound of probability, we obtain

$$\begin{aligned} & P\big(\sup_{u\in\mathcal{U}}\max_{j,k}|\widehat{\Sigma}_{jk}(u)-\Sigma_{jk}(u)|>\delta\big) \\ \leqslant\ & C_2p^2n^{C_3}\exp\big\{-C_1n^{2\gamma-\alpha(\beta+2)}\delta^2\big\}+C_2p^2\exp\big\{-C_1n^{2\gamma-\alpha(3\beta+4)}\delta^2\big\} \\ & +C_2p^2\exp\big\{-C_1T^{-4}n^{2\gamma-4\alpha}\delta^2\big\}+C_2p^2n^{C_3}\exp\big\{-C_1T^{-6}n^{2\gamma}\delta^2\big\}+C_2p^2\exp\big\{-C_1Tn^{-2\alpha}\delta^2\big\}. \end{aligned}$$

This exponential concentration inequality under $p\gtrsim n$ setting leads to the $\log p$ term in the uniform convergence rate in (B.8). Since $T\gtrsim n^{2\alpha}$ and $n^{2\gamma}\gtrsim T^6$ can imply $n^{2\gamma-4\alpha}\gtrsim T^4$, the third term on the right hand side is dominated by the last two terms. Furthermore, the second term dominates the first term. The uniform convergence rate in (B.8) follows with the choice of $\kappa_{n,T}$ and δ as stated in the lemma.

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 15–17 as follows.

Lemma 15 *Suppose that the sparse case in Condition 1 and Condition 3 hold. Then for each $j,k=1,\ldots,p$,*

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

Proof. By the joint Gaussian assumption, $\widetilde{\xi}_{ijl}=\omega_{jl}\boldsymbol{\phi}_{jl}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{Y}_{ij}}^{-1}\mathbf{Y}_{ij}$, Condition 3 and $T_{ij}\leqslant T_0$, we have $\widetilde{\xi}_{ijl}\sim N(0,\omega_{jl}^2)$. It follows from Cauchy-Schwartz inequality and $|\widetilde{\sigma}_{jlkm}|=|E\left[\widetilde{\xi}_{jl}\widetilde{\xi}_{km}\right]|\leqslant\sqrt{E(\widetilde{\xi}_{jl}^2)E(\widetilde{\xi}_{km}^2)}=w_{jl}w_{km}$ and Condition 3 that the truncation error can be bounded as

$$\begin{aligned} \sup_u|\widetilde{\Sigma}_{jk}(u)-\widetilde{\Sigma}_{jk,M}(u)| &\leqslant \Big(\sum_{l=1}^{\infty}\sum_{m=1}^{\infty}-\sum_{l=1}^{M}\sum_{m=1}^{M}\Big)|\widetilde{\sigma}_{jlkm}|\sup_u|\phi_{jl}(u)|\sup_u|\phi_{km}(u)|, \\ &\leqslant \Big(\sum_{l=1}^{\infty}\sum_{m=1}^{\infty}-\sum_{l=1}^{M}\sum_{m=1}^{M}\Big)\omega_{jl}\omega_{km}O(1) \\ &= O\Big(\sum_{l=1}^{M}\sum_{m=M+1}^{\infty}l^{-\beta}\omega_{km}\Big)\leqslant O(M^{-2\nu})\asymp n^{-2\alpha\nu}, \end{aligned}$$

where the last line comes from $\sum_{m=M+1}^{\infty}\omega_{km}\leqslant(\sum_{m=M+1}^{\infty}\omega_{km}^{1/2})^2$, Condition 3 and

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

Lemma 16 *Suppose that the sparse case in Condition 1 and Conditions 2–3 hold. Then for each* $j, k = 1, \ldots, p$ *and* $l, m = 1, \ldots, M$,

$$\Big|\frac{1}{n}\sum_{i=1}^{n}\widehat{\xi}_{ijl}\widehat{\xi}_{ikm} - E(\widetilde{\xi}_{ijl}\widetilde{\xi}_{ikm})\Big| = O_p\Big\{lm^{-\beta}n^{-\gamma} + l^{-\beta}mn^{-\gamma} + l^{-\beta}m^{-\beta}(\log n)^{1/2}n^{-\gamma}\Big\}.$$

Proof. Since $\widetilde{\xi}_{ijl} \sim N(0, \omega_{jl}^2)$, we let $\widetilde{\xi}_{ijl} = \omega_{jl}a_{ij}$ with $a_{ij} \sim N(0,1)$. It follows from the same technique used in the proof for Lemma 10 that

$$\Big|\frac{1}{n}\sum_{i=1}^{n}\widetilde{\xi}_{ijl}\widetilde{\xi}_{ikm} - E(\widetilde{\xi}_{ijl}\widetilde{\xi}_{ikm})\Big| = O_p(l^{-\beta}m^{-\beta}n^{-1/2}). \tag{B.9}$$

By the expansion of $\frac{1}{n}\sum_{i=1}^{n}\widehat{\xi}_{ijl}\widehat{\xi}_{ikm} - E(\widetilde{\xi}_{ijl}\widetilde{\xi}_{ikm}) = \frac{1}{n}\sum_{i=1}^{n}\big\{(\widehat{\xi}_{ijl}\widehat{\xi}_{ikm} - \widetilde{\xi}_{ijl}\widetilde{\xi}_{ikm}) + (\widetilde{\xi}_{ijl}\widetilde{\xi}_{ikm} - E(\widetilde{\xi}_{ijl}\widetilde{\xi}_{ikm}))\big\}$ and the results in Lemma 8 and (B.9), we can immediately obtain the convergence rate as stated in the lemma.

Lemma 17 *Suppose that the sparse case in Condition 1 and Conditions 2–3 hold. Then there exist some positive constants* C_1, C_2, C_3 *such that for any* δ *with* $0 < \delta \leqslant C_1$ *and each* $j, k = 1, \ldots, p$,

$$P\big(\sup_{u\in\mathcal{U}}|\widehat{\Sigma}_{jk}(u) - \widetilde{\Sigma}_{jk,M}(u)| > \delta\big) \leqslant C_2 n^{C_3}\exp\big\{-C_1 n^{2\gamma-2\alpha}\delta^2\big\} + C_2\exp\big\{-C_1 n^{2\gamma-\alpha(2\beta+4)}\delta^2\big\}$$

Proof. Note

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

We can write

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

We first bound $J_1(u)$. It follows from Cauchy-Schwartz inequality on $|\widetilde{\sigma}_{jlkm}|$, Condition 3 and Lemma 6 that

$$\begin{aligned}
\sup_u|J_1(u)| &\leqslant \big\{\sup_u|\phi_{jl}(u)| + \sup_u|\widehat{\phi}_{jl}(u) - \phi_{jl}(u)|\big\}\big\{|\widetilde{\sigma}_{jlkm}| + |\widehat{\sigma}_{jlkm} - \widetilde{\sigma}_{jlkm}|\big\}\sup_u|\widehat{\phi}_{km}(u) - \phi_{km}(u)| \\
&= O_p\big[(l^{-\beta}m^{-\beta})\big\{m^{\beta}(\log n)^{1/2}n^{-\gamma} + m^{2\beta+1}n^{-\gamma}\big\}\big] \\
&= O_p\big\{l^{-\beta}(\log n)^{1/2}n^{-\gamma} + l^{-\beta}m^{\beta+1}n^{-\gamma}\big\}.
\end{aligned}$$

We next bound $J_2(u)$. It follows from Condition 3 and Lemma 16 that

$$\begin{aligned}\sup_u|I_2(u)| &\leqslant \big\{\sup_u|\phi_{jl}(u)|+\sup_u|\widehat{\phi}_{jl}(u)-\phi_{jl}(u)|\big\}|\widehat{\sigma}_{jlkm}-\sigma_{jlkm}|\sup_u|\phi_{km}(u)| \\ &= O_p\big\{lm^{-\beta}n^{-\gamma}+l^{-\beta}mn^{-\gamma}+l^{-\beta}m^{-\beta}(\log n)^{1/2}n^{-\gamma}\big\}.\end{aligned}$$

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

$$\sup_u|J_3(u)| = O_p\big\{m^{-\beta}(\log n)^{1/2}n^{-\gamma}+l^{\beta+1}m^{-\beta}n^{-\gamma}\big\}.$$

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

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

where the last line comes from

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

and other similar inequalities. By $M\asymp n^{\alpha}$ and the probability measure considered in Lemma 2, we can obtain the concentration inequality with suitable choices of C_1, C_2, C_3.

Lemma 18 *Suppose that the sparse case in Condition 1 and Conditions 2–3 hold. For sufficiently large M', if $\delta = M'\big\{(\log p/n^{2\gamma-\alpha(2\beta+4)})^{1/2}+(\log p/n^{4\alpha\nu})^{1/2}\big\}$, $\log p/n^{2\gamma-\alpha(2\beta+4)}\to 0$ and $\log p/n^{4\alpha\nu}\to 0$, then under high dimensional setting with $p\gtrsim n$, we have*

$$\sup_{u\in\mathcal{U}}\max_{j,k}|\widehat{\Sigma}_{jk}(u)-\widetilde{\Sigma}_{jk}(u)| = O_p\left\{\left(\frac{\log p}{n^{2\gamma-\alpha(2\beta+4)}}\right)^{1/2}+\left(\frac{\log p}{n^{4\alpha\nu}}\right)^{1/2}\right\}. \tag{B.10}$$

Proof. It follows from $\sup_{u\in\mathcal{U}}|\widehat{\Sigma}_{jk}(u)-\widetilde{\Sigma}_{jk}(u)|\leqslant\sup_{u\in\mathcal{U}}|\widehat{\Sigma}_{jk}(u)-\widetilde{\Sigma}_{jk,M}(u)|+\sup_{u\in\mathcal{U}}|\widetilde{\Sigma}_{jk,M}(u)-\widetilde{\Sigma}_{jk}(u)|$, $\delta\gg n^{-2\alpha\nu}$, Lemmas 15–17 that there exist some positive constants C_1, C_2, C_3 such

that

$$\begin{aligned} & P\big(\sup_{u\in\mathcal{U}}|\widehat{\Sigma}_{jk}(u)-\widetilde{\Sigma}_{jk}(u)|>\delta\big) \\ \leqslant\ & P\big(\sup_{u\in\mathcal{U}}|\widehat{\Sigma}_{jk}(u)-\widetilde{\Sigma}_{jk,M}(u)|>\delta/2\big) \\ \leqslant\ & C_2n^{C_3}\exp\big\{-C_1n^{2\gamma-2\alpha}\delta^2\big\}+C_2\exp\big\{-C_1n^{2\gamma-\alpha(2\beta+4)}\delta^2\big\}. \end{aligned}$$

Applying the union bound of probability, we obtain

$$P\big(\sup_{u\in\mathcal{U}}\max_{j,k}|\widehat{\Sigma}_{jk}(u)-\widetilde{\Sigma}_{jk}(u)|>\delta\big)\leqslant C_2p^2n^{C_3}\exp\big\{-C_1n^{2\gamma-2\alpha}\delta^2\big\}+C_2p^2\exp\big\{-C_1n^{2\gamma-\alpha(2\beta+4)}\delta^2\big\}$$

and hence the uniform convergence rate in (B.10) follows when $p\gtrsim n$.

Lemma 19 *If $\lambda_n(u)\geqslant||\mathbf{\Theta}(u)||_1|\widehat{\mathbf{\Sigma}}(u)-\mathbf{\Sigma}(u)|_\infty$ for each $u\in\mathcal{U}$, then we have*

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

Proof. We will use the following property that, for two matrices $\mathbf{A}$ and $\mathbf{B}$

$$|\mathbf{AB}|_\infty\leqslant|\mathbf{A}|_\infty||\mathbf{B}||_1 \tag{B.11}$$

in our later proofs. For each $u\in\mathcal{U}$, by (B.11) and bound condition for $\lambda_n(u)$ we have

$$|\mathbf{I}-\widehat{\mathbf{\Sigma}}(u)\mathbf{\Theta}(u)|_\infty=|\big\{\mathbf{\Sigma}(u)-\widehat{\mathbf{\Sigma}}(u)\big\}\mathbf{\Theta}(u)|_\infty\leqslant|\mathbf{\Sigma}(u)-\mathbf{\Sigma}(u)|_\infty||\mathbf{\Theta}(u)||_1\leqslant\lambda_n(u). \tag{B.12}$$

By (B.12) and the optimization problem considered in (7), we obtain

$$|\widehat{\mathbf{\Sigma}}(u)\big\{\widehat{\mathbf{\Theta}}_1(u)-\mathbf{\Theta}(u)\big\}|_\infty\leqslant|\widehat{\mathbf{\Sigma}}(u)\widehat{\mathbf{\Theta}}_1(u)-\mathbf{I}|_\infty+|\mathbf{I}-\widehat{\mathbf{\Sigma}}(u)\mathbf{\Theta}(u)|_\infty\leqslant 2\lambda_n(u). \tag{B.13}$$

By (B.12) and the definition of $\widehat{\beta}_j(u)$, $j\in V$, we have $|\widehat{\beta}_j(u)|_1\leqslant||\mathbf{\Theta}(u)||_1$. By Lemma 1 we have $||\widehat{\mathbf{\Theta}}_1(u)||_1\leqslant||\mathbf{\Theta}(u)||_1$. This result together with (B.11), (B.13) and the lower bound condition for $\lambda_n(u)$ yield

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

which completes the proof.

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

$$||\mathbf{A}|| \leqslant ||\mathbf{A}||_1 = \max_j \sum_{k=1}^{p} |A_{jk}|. \tag{B.14}$$

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 (B.14), we can bound

$$\begin{aligned}
&\sup_u ||\widehat{\mathbf{\Theta}}(u) - \mathbf{\Theta}(u)|| \\
\leqslant\ & \sup_u \max_j \sum_{k=1}^{p} |\widehat{\Theta}_{jk}(u) - \Theta_{jk}(u)| \\
\leqslant\ & \sup_u \max_j \sum_{k=1}^{p} |\widehat{\Theta}_{jk}(u)I\{|\widehat{\Theta}_{jk}(u)| \geqslant 2\tau_n(u)\} - \Theta_{jk}(u)| \\
& + \sup_u \max_j \sum_{k=1}^{p} |\widehat{\Theta}_{jk}(u)I\{|\widehat{\Theta}_{jk}(u)| < 2\tau_n(u)\}| \\
\leqslant\ & 2\sup_u \max_j \sum_{k=1}^{p} |\widehat{\Theta}_{jk}(u)I\{|\widehat{\Theta}_{jk}(u)| \geqslant 2\tau_n(u)\} - \Theta_{jk}(u)| \\
\leqslant\ & 2\sup_u \max_j \sum_{k=1}^{p} |\Theta_{jk}(u)I\{|\Theta_{jk}(u)| < 2\tau_n(u)\}| \\
& + 2\sup_u \max_j \sum_{k=1}^{p} |\widehat{\Theta}_{jk}(u)I\{|\widehat{\Theta}_{jk}(u)| \geqslant 2\tau_n(u)\} - \Theta_{jk}(u)I\{|\Theta_{jk}(u)| \geqslant 2\tau_n(u)\}| \\
\leqslant\ & 2\{2\sup_u \tau_n(u)\}^{1-q} \sup_u \max_j \sum_{k=1}^{p} |\Theta_{jk}(u)|^q + 2\sup_u \tau_n(u) \sup_u \max_j \sum_{k=1}^{p} I\{|\widehat{\Theta}_{jk}(u)| \geqslant 2\tau_n(u)\} \\
& + 2\sup_u \max_j \sum_{k=1}^{p} |\Theta_{jk}(u)| \cdot \left|I\{|\widehat{\Theta}_{jk}(u)| \geqslant 2\tau_n(u)\} - I\{|\Theta_{jk}(u)| \geqslant 2\tau_n(u)\}\right|
\end{aligned}$$

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

$$
\begin{aligned}
&\leqslant 2\big\{2\sup_u \tau_n(u)\big\}^{1-q} s_0(p) + 2\sup_u \tau_n(u) \sup_u \max_j \sum_{k=1}^{p} I\big\{|\Theta_{jk}(u)| \geqslant 2\tau_n(u)\big\} \\
&\quad +2\sup_u \max_j \sum_{k=1}^{p} |\Theta_{jk}(u)| I\big\{\big||\Theta_{jk}(u)| - 2\tau_n(u)\big| \leqslant |\widehat{\Theta}_{jk}(u) - \Theta_{jk}(u)|\big\} \\
&\leqslant 2\big\{2\sup_u \tau_n(u)\big\}^{1-q} s_0(p) + 2\big\{\sup_u \tau_n(u)\big\}^{1-q} \sup_u \max_j \sum_{k=1}^{p} |\Theta_{jk}(u)|^q \\
&\quad +2\sup_u \max_j \sum_{k=1}^{p} |\Theta_{jk}(u)| I\big\{|\Theta_{jk}(u)| \leqslant 3\tau_n(u)\big\} \\
&\leqslant 2(1 + 2^{1-q} + 3^{1-q})\big\{\sup_u \tau_n(u)\big\}^{1-q} s_0(p).
\end{aligned}
$$

Then we can use (B.8) in Lemma 13, $K'(u) = ||\mathbf{\Theta}(u)||_1$ and Lemma 19 with the choice of $\lambda_n(u) = CK'(u)\left\{\left(\log p/\kappa_{n,T}^2\right)^{1/2} + \left(\log p/n^{2\alpha\nu}\right)^{1/2}\right\}$ to obtain the uniform convergence rate in (14), which completes the proof for the dense design.

For the sparse case, we substitute $\mathbf{\Theta}(u)$ and $\mathbf{\Sigma}(u)$ by $\widetilde{\mathbf{\Theta}}(u)$ and $\widetilde{\mathbf{\Sigma}}(u)$, respectively, and similarly can use (B.10) in Lemma 17, $K'(u) = ||\widetilde{\mathbf{\Theta}}(u)||_1$ and Lemma 19 with the choice of $\lambda_n(u) = CK'(u)\left\{\left(\log p/n^{2\gamma-\alpha(2\beta+4)}\right)^{1/2} + \left(\log p/n^{4\alpha\nu}\right)^{1/2}\right\}$ to obtain the uniform convergence rate in (13), which completes the proof for the sparse design.

B.2 Proof of Theorem 2

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

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

Hence

$$
P\Big(\sum_{j,k} I\big\{|\widehat{\Theta}_{jk}(u)| \geqslant \tau_n(u), \Theta_{jk}(u) = 0\big\}\Big) \leqslant P\Big(\max_{j,k} \sup_u |\widehat{\Theta}_{jk}(u) - \Theta_{jk}(u)| \geqslant \inf_u \tau_n(u)\Big), \quad \text{(B.15)}
$$

We can set $\tau_n(u) = 4K'(u)\lambda_n(u)$. This together with Lemmas 14, 19 and the choice of $\lambda_n(u) = CK'(u)\left\{\left(\log p/\kappa_{n,T}^2\right)^{1/2} + \left(\log p/n^{2\alpha\nu}\right)^{1/2}\right\}$ imply that the probability in (B.15) is bounded by $C_2 \exp\{2\log p - C_1\kappa_{n,T}^2 \inf_u \tau_n(u)^2\} = C_2 \exp\{2\log p - C_1 \inf_u K'(u)^2 \log p\}$. Hence

we can choose $\inf_u K'(u)$ 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

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

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 $\widehat{\Theta}_{jk}(u) I\big\{|\widehat{\Theta}_{jk}(u)| \geqslant \tau_n(u)\big\}$ 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 $\boldsymbol{\Theta}(u)$ by $\widetilde{\boldsymbol{\Theta}}(u)$ in (B.15). This fact together with Lemmas 18, 19 and the choice of $\lambda_n(u) = CK'(u)\Big\{\big(\log p/n^{2\gamma-\alpha(2\beta+4)}\big)^{1/2} + (\log p/n^{4\alpha\nu})^{1/2}\Big\}$ imply that the probability in (B.15) is bounded by $C_2 \exp\{2\log p - C_1 n^{2\gamma-\alpha(2\beta+4)} \inf_u \tau_n(u)^2\} = C_2 \exp\{2\log p - C_1 \inf_u K'(u)^2 \log p\}$. 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.

B.3 Concentration Inequalities for Fully Observed Functional Data

Lemma 20 *Suppose that Condition 3 (ii) holds. Then there exists some positive constants C_1, C_2 such that for any $0 < \delta \leqslant C_1$ and each $j = 1, \ldots, p$,*

$$P\big(\|\widehat{C}_{jj}(u,v) - C_{jj}(u,v)\|_{\mathcal{S}} \geqslant \delta\big) \leqslant C_2 \exp(-C_1 n\delta^2),$$

Proof. This lemma can be found in Lemma 6 in the Supplementary Material of Qiao et al. (2017) and hence the proof is omitted.

Lemma 21 *Suppose that Condition 3 (ii) holds and $\phi_{jl}(u)$'s are Lipschitz-continuous. Then there exists some positive constants C_1, C_2 such that for any $0 < \delta \leqslant C_1$ and each $j =$*

$1, \ldots, p$,

$$P\big(\sup_{(u,v)\in\mathcal{U}^2} \big|\widehat{C}_{jj}(u,v) - C_{jj}(u,v)\big| \geqslant \delta \big) \leqslant C_2 n \exp(-C_1 n\delta^2).$$

Proof. By the spectral decomposition $C_{jj}(u,v) = \sum_{k=1}^{\infty} \omega_{jk}\phi_{jk}(u)\phi_{jk}(v)$, we have

$$\begin{aligned}
\max_{j\in V}|C_{jj}(u,v) - C_{jj}(u',v)| &\leqslant \max_{j\in V}\sum_{l=1}^{\infty}\omega_{jl}|\phi_{jl}(u) - \phi_{jl}(u')||\phi_{jl}(v)| \\
&\leqslant \max_{j\in V}\sup_{l\geqslant 1, v\in\mathcal{U}}|\phi_{jl}(v)|\max_{j\in V}\sum_{l=1}^{\infty}\omega_{jl}\sup_{j\in V, l\geqslant 1} c_{jl}|u-u'| \leqslant C|u-u'|,
\end{aligned}$$

where c_{jl} and C are some positive constants and the last line follows from $\max_{j\in V}\sum_{k=1}^{\infty}\omega_{jk} < \infty$, $\max_{j\in V}\sup_{v\in\mathcal{U}}\sup_{l\geqslant 1}|\phi_{jl}(v)| = O(1)$ and the Lipschitz-continuity of $\phi_{jl}(v)$'s. So $C_{jj}(u,\cdot)$ and $C_{jj}(\cdot,v)$ are Lipschitz-continuous for each $u \in \mathcal{U}$ and $v \in \mathcal{U}$, respectively.

Let $\mathcal{U} = [a,b]$. We first reduce the problem from supremum over product of interval $[a,b]^2$ to the maximum over a grid of pairs on the product interval. We partition the interval $[a,b]$ into N subintervals B_k of equal length. Let u_k be the centers of B_k. By the Lipschitz-continuity of $C_{jj}(u,v)$, there exists some positive constant C_1

$$|C_{jj}(u,v) - C_{jj}(u',v')| \leqslant |C_{jj}(u,v) - C_{jj}(u',v)| + |C_{jj}(u',v) - C_{jj}(u',v')| \leqslant C_1(|v-v'| + |u-u'|).$$

For each (u,v) and $(u',v') \in \mathcal{U}^2$, it follows from the central limit theorem that with probability tending to 1, $|\widehat{C}_{jj}(u,v) - C_{jj}(u,v)| \leqslant C_2 n^{-1/2}$, $|\widehat{C}_{jj}(u',v') - C_{jj}(u',v')| \leqslant C_2 n^{-1/2}$, where C_2 is some positive constant. Using these facts, for each $(u,v) \in B_k \times B_{k'}$, there exist some positive constant C_3

$$\begin{aligned}
|\widehat{C}_{jj}(u,v) - C_{jj}(u,v)| &\leqslant |\widehat{C}_{jj}(u,v) - \widehat{C}_{jj}(u_k,v_{k'})| + |\widehat{C}_{jj}(u_k,v_{k'}) - C_{jj}(u_k,v_{k'})| \\
&\quad + |C_{jj}(u_k,v_{k'}) - C_{jj}(u,v)| \\
&\leqslant |\widehat{C}_{jj}(u_k,v_{k'}) - C_{jj}(u_k,v_{k'})| + C_3(N^{-1} + n^{-1/2}).
\end{aligned}$$

Thus $\sup_{(u,v)\in B_k\times B_{k'}} |\widehat{C}_{jj}(u,v) - C_{jj}(u,v)| \leqslant |\widehat{C}_{jj}(u_k,v_{k'}) - C_{jj}(u_k,v_{k'})| + C(N^{-1} + n^{-1/2})$. By taking $N = n^{1/2}$, we have

$$\sup_{(u,v)\in[a,b]^2} |\widehat{C}_{jj}(u,v) - C_{jj}(u,v)| \leqslant \max_{1\leqslant k,k'\leqslant N} |\widehat{C}_{jj}(u_k,v_{k'}) - C_{jj}(u_k,v_{k'})| + C_4 n^{-1/2}, \qquad \text{(B.16)}$$

where C_4 is a positive constant. Under the Gaussian assumption for $X_{ij}(u)$, it follows from the tail probability bound for $|\widehat{C}_{jj}(u,v)-C_{jj}(u,v)|$ that there exist positive constants C_5, C_6 such that $P\big(|\widehat{C}_{jj}(u_k,v_{k'})-C_{jj}(u_k,v_{k'})|\geqslant\delta\big)\leqslant C_6\exp(-C_5n\delta^2)$. Then $P\big(\max\limits_{1\leqslant k,k'\leqslant N}|\widehat{C}_{jj}(u_k,v_{k'})-C_{jj}(u_k,v_{k'})|\geqslant\delta\big)\leqslant nC_6\exp(-C_5n\delta^2)$. This together with (B.16) yield that

$$P\Big(\sup_{(u,v)\in[a,b]^2}|\widehat{C}_{jj}(u,v)-C_{jj}(u,v)|\geqslant\delta\Big)\leqslant nC_6\exp(-C_5n\delta^2),$$

which completes our proof for the lemma.

C Additional empirical results

We provide electrode/node names for $j=1,2,\ldots,64$ in Table 4.

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	