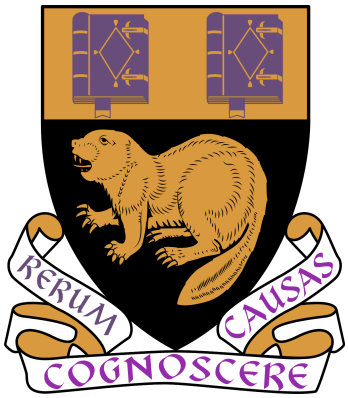
High-dimensional Functional Data/Time Series Analysis: Finite-sample Theory, Adaptive Functional Thresholding and Prediction



Qin Fang

The Department of Statistics

London School of Economics and Political Science

A thesis submitted for the degree of

Doctor of Philosophy

October 2022

To my loving parents,
乐君 & 晓东

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I confirm that a version of Chapter 1 was adapted into a paper entitled "Finite Sample Theory for High-Dimensional Functional/Scalar Time Series with Applications," co-authored with my supervisor, Dr. Xinghao Qiao, and Prof. Shaojun Guo from Renmin University of China. This work was published in *Electronic Journal of Statistics*, see Fang et al. (2022).

I further confirm that a version of Chapter 2 was adapted into a paper entitled "Adaptive Functional Thresholding for Sparse Covariance Function Estimation in High Dimensions," co-authored with Dr. Xinghao Qiao and Prof. Shaojun Guo. This work is under revision at *Journal of the American Statistical Association*.

Finally, I confirm that Chapter 3 was jointly co-authored with Dr. Xinghao Qiao, Prof. Jinyuan Chang from Southwestern University of Finance and Economics, and Prof. Qiwei Yao from the London School of Economics and Political Science. We plan to submit this chapter for publication soon.

Acknowledgements

First of all, I would like to express my deepest gratitude to my supervisor, Dr. Xinghao Qiao, for his incredible patience, continuous encouragement and strong guidance. Had it not been for his dedicated involvement and continued advice, this work would not have taken shape.

I would like to extend my sincere thanks to Prof. Shaojun Guo for his valuable research discussions and helpful suggestions, and Prof. Qiwei Yao for his generous support and trust during my PhD studies. Their wisdom, personal integrity and undying curiosity continue to be an inspiration to me.

I gratefully acknowledge the support of the London School of Economics and Political Science and the Department of Statistics for providing the studentship that allowed me to pursue my PhD degree. Moreover, I express my deepest appreciation to the professional services team in our department, in particular Penny Montague and Imelda Noble, for their outstanding administrative support.

Last but not least, I would like to thank my parents for their unconditional love, support and ongoing encouragement. Some special words of gratitude go to my amazing friends, with whom I have shared moments of deep anxiety but also of big excitement over the years: Xiaoyan Xie, Han Yan, Nina Wang, Fengzhi Zhao, Shuqing Si and my group JWMN. Thanks guys for always being there for me and inspiring me in all kinds of ways. My sincere thanks also go to all my family members, my dear friends, and to whom that have walked with me in this once in a lifetime journey.

Abstract

Statistical analysis of high-dimensional functional data/times series arises in various applications. Examples include different types of brain imaging data in neuroscience (Zhu et al., 2016; Li and Solea, 2018), age-specific mortality rates for different prefectures (Gao et al., 2019a) and intraday energy consumption trajectories (Cho et al., 2013) for thousands of households, to list a few. Under this scenario, in addition to the intrinsic infinite-dimensionality of functional data, the number of functional variables can grow with the number of independent or serially dependent observations, posing new challenges to existing work. In this thesis, we consider three fundamental tasks in high-dimensional functional data/times series analysis: finite sample theory, covariance function estimation (with a new class of adaptive functional thresholding operators) and modelling/prediction.

In the first chapter, we focus on the theoretical analysis of relevant estimated cross-(auto)covariance terms between two multivariate functional time series or a mixture of multivariate functional and scalar time series beyond the Gaussianity assumption. We introduce a new perspective on dependence by proposing functional cross-spectral stability measure to characterize the effect of dependence on these estimated cross terms, which are essential in the estimates for additive functional linear regressions. With the proposed functional cross-spectral stability measure, we develop useful concentration inequalities for estimated cross-(auto)covariance matrix functions to accommodate more general sub-Gaussian functional linear processes and, furthermore, establish finite sample theory for relevant estimated terms under a commonly adopted functional principal component analysis framework. Using our derived non-asymptotic results, we investigate the convergence properties of the regularized estimates for two additive functional linear regression applications under sparsity assumptions including functional linear lagged regression and partially functional linear regression in the context of high-dimensional functional/scalar time series.

In the second chapter, we consider estimating sparse covariance functions for high-dimensional functional data, where the number of random functions p is comparable to, or even larger than the sample size n. Aided by the Hilbert–Schmidt norm of functions, we introduce a new class of functional thresholding operators that combine functional versions of thresholding and shrinkage, and propose the adaptive functional thresholding of the sample covariance function capturing the variability of individual functional entries. We investigate the convergence and support recovery properties of our proposed estimator under a high-dimensional regime where p can grow exponentially with n. Our simulations demonstrate that the adaptive functional thresholding estimators significantly outperform the competing estimators. Finally, we illustrate the proposed method by the analysis of brain functional connectivity using two neuroimaging datasets.

The third chapter proposes a two-step procedure to model and predict high-dimensional functional time series, where the number p of function-valued variables is large in relation to the number n of serially dependent observations. Our first step uses the eigenanalysis of a positive definite matrix to look for linear transformation of original high-dimensional functional time series such that the transformed curve series can be segmented into multiple groups of low-dimensional subseries, and the subseries in different groups are uncorrelated both contemporaneously and serially. Modelling each low-dimensional subseries separately will not lose the overall linear dynamical information and at the same time avoid the overparametrization issue arisen from directly modelling original high-dimensional functional time series. Our second step estimates the finite-dimensional dynamical structure for each group of the transformed curve series that converts the problem of modelling low-dimensional functional time series to that of modelling vector time series. Efficient strategies can be implemented to predict vector time series groupwisely, which can then be converted back to predict groups of transformed curve subseries and finally original functional time series. We investigate the theoretical properties of our proposal when p diverges at an exponential rate of n. The superior finite-sample performance of the proposed methods is illustrated through both extensive simulations and three real datasets.

Contents

Chapter 1

Finite Sample Theory for High-Dimensional Functional/Scalar Time Series with Applications

1.1 Introduction

Functional time series have received a great deal of attention in the last decade in order to provide methodology for functional data objects that are observed sequentially over time. Despite progress being made in this area, existing literature has focused on the statistical analysis of a single or small number of random functions. The increasing availability of large dataset with multiple functional features corresponds to the data structure of

$$\mathbf{X}_t(u) = \big\{X_{t1}(u), \ldots, X_{tp}(u)\big\}^{\mathrm{T}}, \quad t = 1, \ldots, n, \ u \in \mathcal{U},$$

with covariance matrix function $\mathbf{\Sigma}_0^X(u, v) = \mathrm{Cov}\{\mathbf{X}_t(u), \mathbf{X}_t(v)\}$, where, under the high-dimensional and dependent setting, the number of functional variables (p) can be comparable to, or even larger than, the number of serially dependent observations (n), posing new challenges to existing work.

Examples of high-dimensional functional time series include daily electricity consumption curves (Cho et al., 2013) for a large collection of households, half-hourly measured PM10 curves (Aue et al., 2015) over a large number of sites and cumulative intraday return curves (Horváth et al., 2014) for hundreds of stocks. These applications require developing learning techniques to handle such type of data. One large class considers imposing various functional sparsity assumptions on the model parameter space, e.g. *vector functional autoregressions* (VFAR) (Guo and Qiao, 2022) and, under a special independent setting, functional graphical models (Qiao et al., 2019) and functional additive regressions (Fan et al., 2014, 2015; Kong et al., 2016; Luo and Qi, 2017; Xue and Yao, 2021), where the corresponding regularized estimates are proposed.

Within the high-dimensional time series framework, it is essential to establish necessary concentration inequalities for dependent data and assess how the presence of serial dependence affects non-asymptotic error bounds. See relevant concentration results for Gaussian process (Basu and Michailidis, 2015), linear process or linear spatio-temporal model with more general noise distributions (Sun et al., 2018; Shu and Nan, 2019) and heavy tailed time series (Wong et al., 2020). Compared with theoretical analysis of scalar time series, the added technical challenges that arise to handle functional time series involve developing non-asymptotic results for dependent processes within an abstract Hilbert space and characterizing the effect of serial dependence in $\{\mathbf{X}_t(\cdot)\}$ with infinite, summable and decaying eigenvalues of $\mathbf{\Sigma}_0^X$.

Theoretical investigation of high-dimensional functional time series is rather incomplete. Guo and Qiao (2022) proposed a functional stability measure for Gaussian functional time series by controlling the functional Rayleigh quotients of spectral density matrix functions relative to $\mathbf{\Sigma}_0^X$ and hence can precisely capture the effect of small eigenvalues. Moreover, they relied on it to establish concentration bounds on sample (auto)covariance matrix function of $\mathbf{X}_t(\cdot)$, serving as a fundamental tool to provide theoretical guarantees for the proposed three-step procedure and the regularized VFAR estimate, in a high dimensional regime. However, their proposed stability measure only facilitates finite sample theory to accommodate Gaussian functional time series and is not sufficient to evaluate the effect of serial dependence on the estimated cross-(auto)covariance terms in a non-asymptotic way, which plays a crucial role in the theoretical analysis of a wide class of additive functional linear regressions under the high-dimensional regime when the serial dependence exists.

To illustrate, we consider two important examples of additive functional linear regressions in the context of high-dimensional functional/scalar time series. The first

example considers the high-dimensional extension of functional linear lagged regression (Hörmann et al., 2015b) in the additive form:

$$Y_t(v) = \sum_{h=0}^{L} \sum_{j=1}^{p} \int_{\mathcal{U}} X_{(t-h)j}(u)\beta_{hj}(u,v)du + \epsilon_t(v), \quad t = L+1, \ldots, n, (u,v) \in \mathcal{U} \times \mathcal{V}, \tag{1.1}$$

where p-dimensional functional covariates $\{\mathbf{X}_t(\cdot)\}$ and functional errors $\{\epsilon_t(\cdot)\}$ are generated from independent, centered, stationary functional processes, and $\{\beta_{hj}(\cdot,\cdot) : h = 0, \ldots, L, j = 1, \ldots, p\}$ are sparse functional coefficients to be estimated. Under an independent setting without lagged functional covariates, model (1.1) reduces to the additive function-on-function linear regression (Luo and Qi, 2017).

The second example studies partially functional linear regression (Kong et al., 2016) consisting of a mixture of p-dimensional functional time series $\{\mathbf{X}_t(\cdot)\}$ and d-dimensional scalar time series $\mathbf{Z}_t = (Z_{t1}, \ldots, Z_{td})^{\mathrm{T}}$ for $t = 1, \ldots, n$, both of which are independent of errors $\{\epsilon_t\}$, as follows:

$$Y_t = \sum_{j=1}^{p} \int_{\mathcal{U}} X_{tj}(u)\beta_j(u)du + \sum_{k=1}^{d} Z_{tk}\gamma_k + \epsilon_t, \quad t = 1, \ldots, n, u \in \mathcal{U}, \tag{1.2}$$

where $\{\beta_j(\cdot) : j = 1 \ldots, p\}$ are sparse functional coefficients and $\{\gamma_k : k = 1, \ldots, d\}$ are sparse scalar coefficients. Whereas Kong et al. (2016) focused on an independent scenario and treated p as fixed, we allow both p and d to be diverging with n under a more general dependence structure. See also special cases of model (1.2) without functional covariates or scalar covariates in Basu and Michailidis (2015); Wu and Wu (2016) or Fan et al. (2015); Xue and Yao (2021), respectively.

There are many modern applications of the proposed two additive functional linear regression models. In environmental studies, for example, pollutant episodes often exhibit a strong correlation with unfavorable meteorological conditions, resulting in the diminished ability of the atmosphere to disperse the pollutants (Ziomas et al., 1995; Bai et al., 2018). Model (1.1) thus can be applied to forecast the daily pollution curves with meteorological variables, e.g. historic daily weather data, as the functional covariates $\{\mathbf{X}_t(\cdot)\}$. See also Chang et al. (2022) for an example of the application of model (1.2), which aims to forecast the daily intraday return of S&P 100 based on observed cumulative intraday return trajectories of component stocks.

In addition to existing non-asymptotic results in Guo and Qiao (2022), the central challenge to provide theoretical supports for the regularized estimates for models (1.1) and (1.2) is: (i) to characterize how the underlying dependence structure affects the non-asymptotic error bounds on those essential estimated cross-

(auto)covariance terms, e.g. estimated cross-covariance functions between $\mathbf{X}_t(\cdot)$ and $Y_{t+h}(\cdot)$ (or $\epsilon_{t+h}(\cdot)$) for $h = 0, \ldots, L$ in model (1.1) and estimates of $\text{Cov}\{\mathbf{X}_t(\cdot), \mathbf{Z}_t\}$, $\text{Cov}(\mathbf{Z}_t, \epsilon_t)$ and $\text{Cov}\{\mathbf{X}_t(\cdot), \epsilon_t\}$ in model (1.2); (ii) to develop useful non-asymptotic results beyond Gaussian functional/scalar time series.

To address such challenges, the main contribution of this chapter is threefold.

- First, we propose a novel functional cross-spectral stability measure between $\{\mathbf{X}_t(\cdot)\}$ and d-dimensional functional (or scalar) time series, i.e. $\{\mathbf{Y}_t(\cdot) = (Y_{t1}(\cdot), \ldots, Y_{td}(\cdot))^{\mathrm{T}}\}$, defined on $\mathcal{V}$ or $\{\mathbf{Z}_t\}$, based on their cross-spectral density properties. Compared with the direct functional extension of the cross-stability measure in Basu and Michailidis (2015), our functional cross-spectral stability measure can more precisely capture the effect of small eigenvalues to handle truly infinite-dimensional functional objects. It also facilitates the development of non-asymptotic results for $\widehat{\mathbf{\Sigma}}_h^{X,Y}$ and $\widehat{\mathbf{\Sigma}}_h^{X,Z}$, which respectively are estimates of cross-(auto)covariance terms, $\mathbf{\Sigma}_h^{X,Y}(u, v) = \text{Cov}(\mathbf{X}_t(u), \mathbf{Y}_{t+h}(v))$ and $\mathbf{\Sigma}_h^{X,Z} = \text{Cov}(\mathbf{X}_t(u), \mathbf{Z}_{t+h})$ for all integer h. Moreover, it provides insights into how $\widehat{\mathbf{\Sigma}}_h^{X,Y}$ and $\widehat{\mathbf{\Sigma}}_h^{X,Z}$ are affected by the presence of serial dependence.

- Second, we establish finite sample theory in a non-asymptotic way for relevant estimated (cross)-(auto)covariance terms beyond Gaussian functional (or scalar) time series to accommodate more general multivariate functional linear processes with sub-Gaussian functional errors. Our finite sample results and adopted techniques are general, and can be applied broadly to provide theoretical guarantees for regularized estimates of other high-dimensional functional time series models, e.g., the autocovariance-based estimates of sparse functional linear regressions (Chang et al., 2022) and the functional factor model (Guo and Qiao, 2022).

- Third, due to the infinite dimensionality of the functional covariates, dimension reduction is necessary in the estimation. One common approach is *functional principal component analysis* (FPCA). We hence establish useful deviation bounds on relevant estimated terms under a FPCA framework. To illustrate using models (1.1) and (1.2), we implement FPCA-based three-step procedures to estimate unknown parameters under sparsity constraints. With derived non-asymptotic results, we verify functional analogs of routinely used restricted eigenvalue and deviation conditions in the lasso literature (Loh and Wainwright, 2012; Basu and Michailidis, 2015) and, furthermore, investigate the convergence properties of regularized estimates under a high-dimensional and serially dependent setting.

Literature review. Our work lies in the intersection of two strands of literature: functional time series and high-dimensional time series. In the context of functional time series, many standard univariate or low-dimensional time series methods have been recently adapted to the functional domain with theoretical properties explored from a standard asymptotic perspective, see, e.g., Bosq (2000); Bathia et al. (2010); Hörmann and Kokoszka (2010); Panaretos and Tavakoli (2013); Aue et al. (2015); Hörmann et al. (2015b); Pham and Panaretos (2018); Li et al. (2020) and reference therein. In the context of high-dimensional time series, some lower-dimensional structural assumptions are often incorporated on the model parameter space and different regularized estimation procedures have been developed for the respective learning tasks including, e.g., high-dimensional sparse linear regression (Basu and Michailidis, 2015; Wu and Wu, 2016; Han and Tsay, 2020) and high-dimensional sparse vector autoregression (Guo et al., 2016; Lin and Michailidis, 2017; Gao et al., 2019b; Ghosh et al., 2019; Zhou and Raskutti, 2019; Wong et al., 2020; Lin and Michailidis, 2020).

Outline. The remainder of the chapter is organized as follows. In Section 1.2, we propose cross-stability measures under functional and mixed-process scenarios, define sub-Gaussian functional linear processes and rely on them to present finite sample theory for estimated (cross-)terms used in subsequent analyses. In Section 1.3, we consider sparse high-dimensional functional linear lagged model in (1.1), develop the penalized least squares estimation procedure and apply our derived non-asymptotic results to provide theoretical guarantees for the estimates. Section 1.4 is devoted to the modelling, regularized estimation and application of established deviation bounds on the theoretical analysis of sparse high-dimensional partially functional linear model in (1.2). Finally, we examine the finite-sample performance of the proposed methods for both models (1.1) and (1.2) through simulation studies in Section 1.5. All technical proofs are relegated to the appendix.

Notation. Let $\mathbb{Z}$ and $\mathbb{R}$ denote the sets of integers and real numbers, respectively. For $x, y \in \mathbb{R}$, we use $x \vee y = \max(x, y)$. For two positive sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n \lesssim b_n$ or $a_n = O(b_n)$ or $b_n \gtrsim a_n$ if there exists a positive constant c independent of n such that $a_n/b_n \leqslant c$. We write $a_n \asymp b_n$ if $a_n \lesssim b_n$ and $a_n \gtrsim b_n$. For a vector $\mathbf{x} \in \mathbb{R}^p$, we denote its ℓ_1, ℓ_2 and maximum norms by $\|\mathbf{x}\|_1 = \sum_{j=1}^p |x_j|$, $\|\mathbf{x}\| = (\sum_{j=1}^p |x_j|^2)^{1/2}$ and $\|\mathbf{x}\|_{\max} = \max_j |x_j|$, respectively. For a matrix $\mathbf{B} \in \mathbb{R}^{p\times q}$, we denote its Frobenius norm by $\|\mathbf{B}\|_{\mathrm{F}} = \big(\sum_{i,j} \mathrm{B}_{ij}^2\big)^{1/2}$. Let $L_2(\mathcal{U})$ be a Hilbert space of square integrable functions on a compact interval $\mathcal{U}$. For $f, g \in L_2(\mathcal{U})$, we denote the inner product by $\langle f, g\rangle = \int_{\mathcal{U}} f(u)g(u)du$ for $f, g \in L_2(\mathcal{U})$ with the norm $\|\cdot\| = \langle\cdot,\cdot\rangle^{1/2}$. For a Hilbert space $\mathbb{H} \subseteq L_2(\mathcal{U})$, we denote the p-fold Cartesian product by $\mathbb{H}^p = \mathbb{H} \times \cdots \times \mathbb{H}$ and the tensor product

$\mathbb{S} = \mathbb{H} \otimes \mathbb{H}$. For $\boldsymbol{f} = (f_1, \ldots, f_p)^{\mathrm{T}}$ and $\mathbf{g} = (g_1, \ldots, g_p)^{\mathrm{T}}$ in $\mathbb{H}^p$, we denote the inner product by $\langle \boldsymbol{f}, \mathbf{g} \rangle = \sum_{i=1}^{p} \langle f_i, g_i \rangle$ with induced norm of $\boldsymbol{f}$ by $\|\boldsymbol{f}\| = \langle \boldsymbol{f}, \boldsymbol{f} \rangle^{1/2}$, ℓ_1 norm by $\|\boldsymbol{f}\|_1 = \sum_{i=1}^{p} \|f_i\|$, and ℓ_0 norm by $\|\boldsymbol{f}\|_0 = \sum_{i=1}^{p} I(\|f_i\| \neq 0)$, where $I(\cdot)$ is the indicator function. For an integral matrix operator $\mathbf{K} : \mathbb{H}^p \to \mathbb{H}^q$ induced from the kernel matrix function $\mathbf{K} = (K_{ij})_{q \times p}$ with each $K_{ij} \in \mathbb{S}$ through $\mathbf{K}(\boldsymbol{f})(u) = \left\{ \sum_{j=1}^{p} \langle K_{1j}(u, \cdot), f_j(\cdot) \rangle, \ldots, \sum_{j=1}^{p} \langle K_{qj}(u, \cdot), f_j(\cdot) \rangle \right\}^{\mathrm{T}} \in \mathbb{H}^q$, for any given $\boldsymbol{f} \in \mathbb{H}^p$. To simplify notation, we will use $\mathbf{K}$ to denote both the kernel function and the operator. When $p = q = 1$, $\mathbf{K}$ degenerates to K and we denote its Hilbert–Schmidt norm by $\|K\|_{\mathcal{S}} = \left\{ \int \int K(u, v)^2 du dv \right\}^{1/2}$. For general $\mathbf{K}$, we define functional versions of Frobenius, elementwise ℓ_∞, matrix ℓ_1 and matrix ℓ_∞ norms by $\|\mathbf{K}\|_{\mathrm{F}} = \left(\sum_{i,j} \|K_{ij}\|_{\mathcal{S}}^2 \right)^{1/2}$, $\|\mathbf{K}\|_{\max} = \max_{i,j} \|K_{ij}\|_{\mathcal{S}}$, $\|\mathbf{K}\|_1 = \max_j \sum_i \|K_{ij}\|_{\mathcal{S}}$ and $\|\mathbf{K}\|_\infty = \max_i \sum_j \|K_{ij}\|_{\mathcal{S}}$, respectively.

1.2 Finite sample theory

In this section, we first review functional stability measure and propose functional cross-spectral stability measure. We then introduce the definitions of sub-Gaussian process and multivariate functional linear process. Finally, we rely on our proposed stability measures to develop finite sample theory for useful estimated terms to accommodate sub-Gaussian functional linear processes.

1.2.1 Functional stability measure

Consider a p-dimensional vector of weakly stationary functional time series $\{\mathbf{X}_t(\cdot)\}_{t \in \mathbb{Z}}$ defined on $\mathcal{U}$, with mean zero and $p \times p$ autocovariance matrix functions,

$$\boldsymbol{\Sigma}_h^X(u, v) = \mathrm{Cov}\{\mathbf{X}_t(u), \mathbf{X}_{t+h}(v)\} = \{\Sigma_{h,jk}^X(u, v)\}_{1 \leqslant j,k \leqslant p}, \quad t, h \in \mathbb{Z}, (u, v) \in \mathcal{U}^2.$$

These autocovariance matrix functions (or operators) encode the second-order dynamical properties of $\{\mathbf{X}_t(\cdot)\}$ and typically serve as the main focus of functional time series analysis. From a frequency domain analysis prospective, spectral density matrix function (or operator) aggregates autocovariance information at different lag orders $h \in \mathbb{Z}$ at a frequency $\theta \in [-\pi, \pi]$ as

$$\boldsymbol{f}_\theta^X = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \boldsymbol{\Sigma}_h^X \exp(-ih\theta).$$

According to Guo and Qiao (2022), the functional stability measure of $\{\mathbf{X}_t(\cdot)\}$ is defined based on the functional Rayleigh quotients of $\boldsymbol{f}_\theta^X$ relative to $\boldsymbol{\Sigma}_0^X$,

$$\mathcal{M}^X = 2\pi \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi],\boldsymbol{\Phi}\in\mathbb{H}_0^p} \frac{\langle \boldsymbol{\Phi}, \boldsymbol{f}_\theta^X(\boldsymbol{\Phi})\rangle}{\langle \boldsymbol{\Phi}, \boldsymbol{\Sigma}_0^X(\boldsymbol{\Phi})\rangle}, \tag{1.3}$$

where $\mathbb{H}_0^p = \{\boldsymbol{\Phi} \in \mathbb{H}^p : \langle \boldsymbol{\Phi}, \boldsymbol{\Sigma}_0^X(\boldsymbol{\Phi})\rangle \in (0,\infty)\}$ and ess sup denotes the essential supremum, that is for a measurable real-valued function $m : M \to \mathbb{R}$ defined on M, $\operatorname{ess\,sup}(m) = \inf\{\nu \in \mathbb{R} : m(\varpi) \leqslant \nu \text{ for almost all } \varpi \in M\}$. To handle truly infinite-dimensional objects $\{\mathbf{X}_t(\cdot)\}$ with infinite, summable and decaying eigenvalues of $\boldsymbol{\Sigma}_0^X$, such stability measure $\mathcal{M}^X$ can more precisely capture the effect of small eigenvalues of $\boldsymbol{\Sigma}_0^X$ on the numerator in (1.3).

We next impose a condition on $\mathcal{M}^X$ and introduce the functional stability measure of subprocesses of $\{\mathbf{X}_t(\cdot)\}$, which will be used in our subsequent analysis.

Condition 1.1. *(i) The spectral density matrix operator* $\boldsymbol{f}_\theta^X, \theta \in [-\pi,\pi]$ *exists; (ii)* $\mathcal{M}^X < \infty$.

For any k-dimensional subset $J \subseteq \{1,\dots,p\}$ with its cardinality $|J| \leqslant k$, we can measure the stability of the subprocess $\big\{\big(X_{tj}(\cdot)\big) : j \in J\big\}_{t\in\mathbb{Z}}$ in a similar fashion. The functional stability measure of all k-dimensional subprocesses of $\{\mathbf{X}_t(\cdot)\}$ is thus defined by

$$\mathcal{M}_k^X = 2\pi \cdot \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi],\|\boldsymbol{\Phi}\|_0\leqslant k,\boldsymbol{\Phi}\in\mathbb{H}_0^p} \frac{\langle \boldsymbol{\Phi}, \boldsymbol{f}_\theta^X(\boldsymbol{\Phi})\rangle}{\langle \boldsymbol{\Phi}, \boldsymbol{\Sigma}_0^X(\boldsymbol{\Phi})\rangle}, \ k = 1,\dots,p. \tag{1.4}$$

Under Condition 1.1, we have $\mathcal{M}_1^X \leqslant \mathcal{M}_2^X \leqslant \cdots \leqslant \mathcal{M}_p^X = \mathcal{M}^X < \infty$.

1.2.2 Functional cross-spectral stability measure

Consider $\{\mathbf{X}_t(\cdot)\}$ and $\{\mathbf{Y}_t(\cdot)\}$, where $\{\mathbf{Y}_t(\cdot)\}_{t\in\mathbb{Z}}$ is a d-dimensional vector of centered and weakly stationary functional time series, defined on $\mathcal{V}$, with lag-h autocovariance matrix function given by

$$\boldsymbol{\Sigma}_h^Y(u,v) = \operatorname{Cov}\{\mathbf{Y}_t(u), \mathbf{Y}_{t+h}(v)\} = \{\Sigma_{h,jk}^Y(u,v)\}_{1\leqslant j,k\leqslant d}, \quad t,h \in \mathbb{Z}, (u,v) \in \mathcal{V}^2.$$

To characterize the effect of dependence on the cross-covariance between two sequences of joint stationary multivariate functional time series, we can correspondingly define the cross-spectral density matrix function (or operator) and functional

cross-spectral stability measure. The proposed cross-spectral stability measure plays a crucial role in the non-asymptotic analysis of relevant estimated cross terms, e.g., estimated cross-(auto)covariance matrix functions in Section 1.2.4.

Definition 1.1. *The cross-spectral density matrix function between* $\{\mathbf{X}_t(\cdot)\}_{t\in\mathbb{Z}}$ *and* $\{\mathbf{Y}_t(\cdot)\}_{t\in\mathbb{Z}}$ *is defined by*

$$\boldsymbol{f}_\theta^{X,Y} = \frac{1}{2\pi}\sum_{h\in\mathbb{Z}} \boldsymbol{\Sigma}_h^{X,Y}\exp(-ih\theta), \quad \theta\in[-\pi,\pi],$$

where $\boldsymbol{\Sigma}_h^{X,Y}(u,v) = \mathrm{Cov}\{\mathbf{X}_t(u), \mathbf{Y}_{t+h}(v)\} = \{\Sigma_{h,jk}^{X,Y}(u,v)\}_{1\leqslant j\leqslant p, 1\leqslant k\leqslant d}, t, h\in\mathbb{Z}, (u,v)\in \mathcal{U}\times\mathcal{V}$.

Condition 1.2. *For* $\{\mathbf{X}_t(\cdot)\}_{t\in\mathbb{Z}}$ *and* $\{\mathbf{Y}_t(\cdot)\}_{t\in\mathbb{Z}}$, $\boldsymbol{f}_\theta^{X,Y}, \theta\in[-\pi,\pi]$ *exists and the functional cross-spectral stability measure defined in (1.5) is finite, i.e.*

$$\mathcal{M}^{X,Y} = 2\pi \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi],\boldsymbol{\Phi}_1\in\mathbb{H}_0^p,\boldsymbol{\Phi}_2\in\mathbb{H}_0^d} \frac{\left|\left\langle \boldsymbol{\Phi}_1, \boldsymbol{f}_\theta^{X,Y}(\boldsymbol{\Phi}_2)\right\rangle\right|}{\sqrt{\left\langle \boldsymbol{\Phi}_1, \boldsymbol{\Sigma}_0^X(\boldsymbol{\Phi}_1)\right\rangle}\sqrt{\left\langle \boldsymbol{\Phi}_2, \boldsymbol{\Sigma}_0^Y(\boldsymbol{\Phi}_2)\right\rangle}} < \infty, \tag{1.5}$$

where $\mathbb{H}_0^p = \{\boldsymbol{\Phi}\in\mathbb{H}^p : \langle\boldsymbol{\Phi}, \boldsymbol{\Sigma}_0^X(\boldsymbol{\Phi})\rangle\in(0,\infty)\}$ *and* $\mathbb{H}_0^d = \{\boldsymbol{\Phi}\in\mathbb{H}^d : \langle\boldsymbol{\Phi}, \boldsymbol{\Sigma}_0^Y(\boldsymbol{\Phi})\rangle\in(0,\infty)\}$.

Remark 1.1. *(a) If* $\{\mathbf{X}_t(\cdot)\}$ *are independent of* $\{\mathbf{Y}_t(\cdot)\}$, *then* $\mathcal{M}^{X,Y} = 0$. *Moreover, in the special case that* $\{\mathbf{X}_t(\cdot)\}$ *and* $\{\mathbf{Y}_t(\cdot)\}$ *are identical,* $\mathcal{M}^{X,Y}$ *degenerates to* $\mathcal{M}^X$ *in (1.3).*

(b) Under the non-functional setting where $\mathbf{X}_t\in\mathbb{R}^p$ *and* $\mathbf{Y}_t\in\mathbb{R}^d$, *Basu and Michailidis (2015) introduced an upper bound condition for their proposed cross-spectral stability measure with* $p = d$, *i.e.*

$$\widetilde{\mathcal{M}}^{X,Y} = \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi],\boldsymbol{\nu}\in\widetilde{\mathbb{R}}_0^d} \sqrt{\frac{\boldsymbol{\nu}^{\mathrm{T}}\{\boldsymbol{f}_\theta^{X,Y}\}^* \boldsymbol{f}_\theta^{X,Y}\boldsymbol{\nu}}{\boldsymbol{\nu}^{\mathrm{T}}\boldsymbol{\nu}}} < \infty, \tag{1.6}$$

where $\widetilde{\mathbb{R}}_0^d = \{\boldsymbol{\nu}\in\mathbb{R}^d : \boldsymbol{\nu}^{\mathrm{T}}\boldsymbol{\nu}\in(0,\infty)\}$ *and* $*$ *denotes the conjugate. This measure relates the cross-stability condition to the largest singular value of the cross-spectral density matrix* $\boldsymbol{f}_\theta^{X,Y}$. *On the other hand, the non-functional analog of (1.5) is equivalent to*

$$\operatorname*{ess\,sup}_{\theta\in[-\pi,\pi],\boldsymbol{\nu}_1\in\widetilde{\mathbb{R}}_0^p,\boldsymbol{\nu}_2\in\widetilde{\mathbb{R}}_0^d} \frac{\left|\boldsymbol{\nu}_1^{\mathrm{T}}\boldsymbol{f}_\theta^{X,Y}\boldsymbol{\nu}_2\right|}{\sqrt{\boldsymbol{\nu}_1^{\mathrm{T}}\boldsymbol{\nu}_1}\sqrt{\boldsymbol{\nu}_2^{\mathrm{T}}\boldsymbol{\nu}_2}} < \infty,$$

whose upper bound is $\widetilde{\mathcal{M}}^{X,Y}$ *as justified in Lemma 1.1 in Appendix 1.B.3. This demonstrates that, compared with (1.6), our proposed cross-stability measure corresponds to a milder condition.*

(c) For two truly infinite-dimensional functional objects, one limitation of the functional analog of $\widetilde{\mathcal{M}}^{X,Y}$ *is that it only controls the largest singular value of* $\boldsymbol{f}_\theta^{X,Y}$*. By contrast, our proposed* $\mathcal{M}^{X,Y}$ *can more precisely characterize the effect of singular values of* $\boldsymbol{f}_\theta^{X,Y}$ *relative to small eigenvalues of* $\boldsymbol{\Sigma}_0^X$ *and* $\boldsymbol{\Sigma}_0^Y$*. Furthermore, it facilitates the development of finite sample theory for normalized versions of relevant estimated cross terms, where the normalization is formed by the corresponding eigenvalues in the denominator of* $\mathcal{M}^{X,Y}$*. See Sections 1.2.4 and 1.2.5 for details.*

(d) We can generalize (1.5) to measure the serial and cross dependence structure between a mixture of multivariate functional and scalar time series. Specifically, consider $\{\mathbf{X}_t(\cdot)\}_{t\in\mathbb{Z}}$ *and* d*-dimensional vector time series* $\{\mathbf{Z}_t\}_{t\in\mathbb{Z}}$ *with autocovariance matrices* $\boldsymbol{\Sigma}_h^Z$ *for* $h \in \mathbb{Z}$*. We can also define* $\boldsymbol{f}_\theta^{X,Z} = \frac{1}{2\pi}\sum_{h\in\mathbb{Z}} \boldsymbol{\Sigma}_h^{X,Z}\exp(-ih\theta)$ *with* $\boldsymbol{\Sigma}_h^{X,Z}(\cdot) = \mathrm{Cov}(\mathbf{X}_t(\cdot), \mathbf{Z}_{t+h})$*. According to (1.5), the mixed cross-spectral stability measure of* $\{\mathbf{X}_t(\cdot)\}$ *and* $\{\mathbf{Z}_t\}$ *can be defined by*

$$\mathcal{M}^{X,Z} = 2\pi \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi],\boldsymbol{\Phi}\in\mathbb{H}_0^p,\boldsymbol{\nu}\in\mathbb{R}_0^d} \frac{\left|\langle\boldsymbol{\Phi}, \boldsymbol{f}_\theta^{X,Z}\boldsymbol{\nu}\rangle\right|}{\sqrt{\langle\boldsymbol{\Phi}, \boldsymbol{\Sigma}_0^X(\boldsymbol{\Phi})\rangle}\sqrt{\boldsymbol{\nu}^{\mathrm{T}}\boldsymbol{\Sigma}_0^Z\boldsymbol{\nu}}} \tag{1.7}$$

and the non-functional stability measure of $\{\mathbf{Z}_t\}$ *reduces to*

$$\mathcal{M}^{Z} = 2\pi \cdot \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi],\boldsymbol{\nu}\in\mathbb{R}_0^d} \frac{\boldsymbol{\nu}^{\mathrm{T}}\boldsymbol{f}_\theta^Z\boldsymbol{\nu}}{\boldsymbol{\nu}^{\mathrm{T}}\boldsymbol{\Sigma}_0^Z\boldsymbol{\nu}}, \tag{1.8}$$

where $\mathbb{R}_0^d = \{\boldsymbol{\nu}\in\mathbb{R}^d : \boldsymbol{\nu}^{\mathrm{T}}\boldsymbol{\Sigma}_0^Z\boldsymbol{\nu} \in (0,\infty)\}$*. The proposed stability measures in (1.7) and (1.8) play an essential role in the convergence analysis of the regularized estimates for model (1.2). See Section 1.4 for details.*

For any k_1-dimensional subset J of $\{1,\ldots,p\}$ and k_2-dimensional subset K of $\{1,\ldots,d\}$, we can accordingly define the functional cross-stability measure of two subprocesses.

Definition 1.2. *Consider subprocesses* $\{(X_{tj}(\cdot)) : j\in J\}_{t\in\mathbb{Z}}$ *for* $J \subseteq \{1,\ldots,p\}$ *with* $|J| \leqslant k_1$ $(k_1 = 1,\ldots,p)$ *and* $\{(Y_{tk}(\cdot)) : k\in K\}_{t\in\mathbb{Z}}$ *for* $K \subseteq \{1,\ldots,d\}$ *with* $|K| \leqslant k_2$

$(k_2 = 1, \ldots, d)$, *their functional cross-spectral stability measure is defined by*

$$\mathcal{M}_{k_1,k_2}^{X,Y} = 2\pi \operatorname*{ess\,sup}_{\substack{\theta\in[-\pi,\pi],\boldsymbol{\Phi}_1\in\mathbb{H}_0^p,\boldsymbol{\Phi}_2\in\mathbb{H}_0^d \\ \|\boldsymbol{\Phi}_1\|_0\leqslant k_1,\|\boldsymbol{\Phi}_2\|_0\leqslant k_2}} \frac{\left|\langle \boldsymbol{\Phi}_1, \boldsymbol{f}_\theta^{X,Y}(\boldsymbol{\Phi}_2)\rangle\right|}{\sqrt{\langle \boldsymbol{\Phi}_1, \boldsymbol{\Sigma}_0^X(\boldsymbol{\Phi}_1)\rangle}\sqrt{\langle \boldsymbol{\Phi}_2, \boldsymbol{\Sigma}_0^Y(\boldsymbol{\Phi}_2)\rangle}}. \tag{1.9}$$

Under Condition 1.2, it is easy to verify that,

$$\mathcal{M}_{k_1,k_2}^{X,Y} \leqslant \mathcal{M}_{k_1',k_2'}^{X,Y} \leqslant \mathcal{M}^{X,Y} < \infty \quad \text{for } k_1 \leqslant k_1' \text{ and } k_2 \leqslant k_2'.$$

According to (1.4), (1.7), (1.8) and (1.9), we can similarly define $\mathcal{M}_{k_1,k_2}^{X,Z}$ and $\mathcal{M}_{k_2}^{Z}$ for $k_1 = 1, \ldots, p$ and $k_2 = 1, \ldots, d$, which will be used in our subsequent analysis.

1.2.3 Sub-Gaussian functional linear process

Before presenting relevant non-asymptotic results beyond Gaussian functional time series, we introduce the definitions of sub-Gaussian process and multivariate functional linear process in this section.

Provided that our non-asymptotic analysis is based on the infinite-dimensional analog of Hanson–Wright inequality (Rudelson and Vershynin, 2013) for sub-Gaussian random variables taking values within a Hilbert space, we first define sub-Gaussian process as follows.

Definition 1.3. *Let $X_t(\cdot)$ be a mean zero random variable in $\mathbb{H}$ and $\Sigma_0 : \mathbb{H} \to \mathbb{H}$ be a covariance operator. Then $X_t(\cdot)$ is a sub-Gaussian process if there exists an $\alpha \geqslant 0$ such that for all $x \in \mathbb{H}$,*

$$\mathbb{E}\{e^{\langle x, X\rangle}\} \leqslant e^{\alpha^2\langle x, \Sigma_0(x)\rangle/2}. \tag{1.10}$$

The proof of Hanson–Wright inequality for serially dependent random functions relies on the fact that uncorrelated Gaussian random functions are also independent, which does not apply for non-Gaussian random functions. However, we show that, for a larger class of non-Gaussian functional time series, it is possible to develop finite sample theory for useful estimated terms in Sections 1.2.4 and 1.2.5. We focus on multivariate functional linear processes with sub-Gaussian errors, namely sub-Gaussian functional linear processes:

$$\mathbf{X}_t(\cdot) = \sum_{l=0}^{\infty} \mathbf{A}_l(\boldsymbol{\varepsilon}_{t-l}), \quad t \in \mathbb{Z}, \tag{1.11}$$

where $\mathbf{A}_l = (A_{l,jk})_{p\times p}$ with each $A_{l,jk} \in \mathbb{S}$ and $\boldsymbol{\varepsilon}_t(\cdot) = \{\varepsilon_{t1}(\cdot), \ldots, \varepsilon_{tp}(\cdot)\}^{\mathrm{T}} \in \mathbb{H}^p$. $\{\boldsymbol{\varepsilon}_t(\cdot)\}_{t\in\mathbb{Z}}$ denotes a sequence of p-dimensional vector of random functions, whose components are independent sub-Gaussian processes satisfying Definition 1.3. It is worth noting that (1.11) not only extends the functional linear processes (Bosq, 2000) to the multivariate setting but also can be seen as a generalization of p-dimensional linear processes (Li et al., 2019) to the functional domain.

Denote the polynomial $\mathcal{B}(z)(u,v) = \sum_{l=0}^{\infty} \mathbf{A}_l(u,v)z^l$ for $u, v \in \mathcal{U}$. Under (1.11), we can derive the spectral density matrix function as

$$\boldsymbol{f}_\theta^X(u,v) = \frac{1}{2\pi}\int\int \mathcal{B}\left(e^{-i\theta}\right)(u,u')\boldsymbol{\Sigma}_0^{\varepsilon}(u',v')\mathcal{B}\left(e^{-i\theta}\right)^*(v,v')du'dv' \tag{1.12}$$

and the covariance matrix function as

$$\boldsymbol{\Sigma}_0^X(u,v) = \sum_{l=0}^{\infty}\int\int \mathbf{A}_l(u,u')\boldsymbol{\Sigma}_0^{\varepsilon}(u',v')\mathbf{A}_l^*(v,v')du'dv'. \tag{1.13}$$

Then we can express the functional stability measure $\mathcal{M}^X$ in (1.3) based on (1.12) and (1.13). The cross-spectral stability measure $\mathcal{M}^{X,Y}$ in (1.5) or $\mathcal{M}^{X,Z}$ in (1.7) can be expressed in a similar fashion.

Condition 1.3. *The coefficient functions satisfy* $\sum_{l=0}^{\infty}\|\mathbf{A}_l\|_\infty = O(1)$.

Condition 1.4. *(i)* $\omega_0^{\varepsilon} = \max_j \int_{\mathcal{U}} \Sigma_{0,jj}^{\varepsilon}(u,u)du = O(1)$; *(ii) The marginal-covariance functions of* $\{\boldsymbol{\varepsilon}_t(\cdot)\}$, $\Sigma_{0,jj}^{\varepsilon}(u,v)$*'s, are continuous on* $\mathcal{U}^2$ *and uniformly bounded over* $j \in \{1,\ldots,p\}$.

Condition 1.3 ensures functional analog of standard condition of elementwise absolute summability of moving average coefficients for multivariate linear processes (Hamilton, 1994) under Hilbert–Schmidt norm. It also guarantees the stationarity of $\{\mathbf{X}_t(\cdot)\}$ and, furthermore together with Condition 1.4, implies that $\omega_0^X = \max_j \int_{\mathcal{U}} \Sigma_{0,jj}^X(u,u)du = O(1)$, both of which are essential in our subsequent analysis. See Lemma 1.2 in Appendix 1.B.3 for details. In general, we can relax Conditions 1.3 and 1.4 by allowing $\sum_{l=0}^{\infty}\|\mathbf{A}_l\|_\infty$ and ω_0^{ε} to grow at very slow rates as p increases, then our subsequent non-asymptotic bounds will depend on ω_0^X, or, more precisely, these two terms, which complicate the presentation of theoretical results.

1.2.4 Concentration bounds on sample (cross-)(auto)covariance matrix function

We construct estimated (auto)covariance of $\{\mathbf{X}_t(\cdot)\}_{t=1}^n$ by

$$\widehat{\mathbf{\Sigma}}_h^X(u,v) = \frac{1}{n-h}\sum_{t=1}^{n-h}\mathbf{X}_t(u)\mathbf{X}_{t+h}(v)^{\mathrm{T}}, \quad h=0,1,\ldots, \ (u,v)\in\mathcal{U}^2,$$

and estimated cross-(auto)covariance matrix functions between $\{\mathbf{X}_t(\cdot)\}$ and $\{\mathbf{Y}_t(\cdot)\}$ by

$$\widehat{\mathbf{\Sigma}}_h^{X,Y}(u,v) = \frac{1}{n-h}\sum_{t=1}^{n-h}\mathbf{X}_t(u)\mathbf{Y}_{t+h}(v)^{\mathrm{T}}, \quad h=0,1,\ldots, \ (u,v)\in\mathcal{U}\times\mathcal{V}.$$

Theorem 1.1. *Suppose that Conditions 1.1–1.4 hold for sub-Gaussian functional linear processes, $\{\mathbf{X}_t(\cdot)\}$, $\{\mathbf{Y}_t(\cdot)\}$ and h is fixed. Then for any given vectors $\mathbf{\Phi}_1\in\mathbb{H}_0^p$ and $\mathbf{\Phi}_2\in\mathbb{H}_0^d$ with $\|\mathbf{\Phi}_1\|_0\leqslant k_1, \|\mathbf{\Phi}_2\|_0\leqslant k_2$ $(k_1=1,\ldots,p, k_2=1,\ldots,d)$, there exists some constants $c, c_1, c_2>0$ such that for any $\eta>0$,*

$$P\left\{\left|\frac{\langle\mathbf{\Phi}_1,(\widehat{\mathbf{\Sigma}}_0^X-\mathbf{\Sigma}_0^X)(\mathbf{\Phi}_1)\rangle}{\langle\mathbf{\Phi}_1,\mathbf{\Sigma}_0^X(\mathbf{\Phi}_1)\rangle}\right|>\mathcal{M}_{k_1}^X\eta\right\}\leqslant 2\exp\left\{-cn\min\left(\eta^2,\eta\right)\right\}, \tag{1.14}$$

and

$$P\left\{\left|\frac{\langle\mathbf{\Phi}_1,(\widehat{\mathbf{\Sigma}}_h^{X,Y}-\mathbf{\Sigma}_h^{X,Y})(\mathbf{\Phi}_2)\rangle}{\langle\mathbf{\Phi}_1,\mathbf{\Sigma}_0^X(\mathbf{\Phi}_1)\rangle+\langle\mathbf{\Phi}_2,\mathbf{\Sigma}_0^Y(\mathbf{\Phi}_2)\rangle}\right|>\left(\mathcal{M}_{k_1}^X+\mathcal{M}_{k_2}^Y+\mathcal{M}_{k_1,k_2}^{X,Y}\right)\eta\right\} \tag{1.15}$$
$$\leqslant c_1\exp\{-c_2n\min(\eta^2,\eta)\}.$$

Remark 1.2. *(1.14) extends the concentration inequality for normalized quadratic form of $\widehat{\mathbf{\Sigma}}_0^X$ in Theorem 1 of Guo and Qiao (2022) under the Gaussianity assumption to accommodate a larger class of sub-Gaussian functional linear processes and serves as a starting point to establish further useful non-asymptotic results, e.g. those listed in Theorems 1–4 and Proposition 1 of Guo and Qiao (2022), so we present some results used in our subsequent analysis in Appendix 1.E. The concentration inequality in (1.15) illustrates that the tail for normalized bilinear form of $\widehat{\mathbf{\Sigma}}_h^{X,Y}-\mathbf{\Sigma}_h^{X,Y}$ behaves in a sub-Gaussian or sub-exponential way depending on which term in the tail bound is dominant. Note that the sub-Gaussian condition is imposed to facilitate the use of Hanson–Wright-type inequality in our non-asymptotic analysis. We believe a Nagaev-type concentration bound can be derived under a weaker finite polynomial moments condition, in which case heavy-tailed functional errors are allowed. It is*

also interesting to develop non-asymptotic results for more general non-Gaussian functional time series under other commonly adopted dependence frameworks.

With suitable choices of $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$, Theorem 1.1 facilitates the elementwise concentration bounds on $\widehat{\mathbf{\Sigma}}_h^{X,Y}$ as follows.

Theorem 1.2. *Suppose that conditions in Theorem 1.1 hold. Then there exists some constants $c_1, c_3 > 0$ such that for any $\eta > 0$ and each $j = 1, \ldots, p$, $k = 1, \ldots, d$,*

$$P\left\{\|\widehat{\Sigma}_{h,jk}^{X,Y} - \Sigma_{h,jk}^{X,Y}\|_{\mathcal{S}} > (\omega_0^X + \omega_0^Y)\mathcal{M}_{X,Y}\eta\right\} \leqslant c_1 \exp\left\{-c_3 n \min(\eta^2, \eta)\right\}, \tag{1.16}$$

where $\omega_0^X = \max_j \int_{\mathcal{U}} \Sigma_{0,jj}^X(u,u)du, \omega_0^Y = \max_k \int_{\mathcal{U}} \Sigma_{0,kk}^Y(u,u)du$ and $\mathcal{M}_{X,Y} = \mathcal{M}_1^X + \mathcal{M}_1^Y + \mathcal{M}_{1,1}^{X,Y}$. In particular, there exists some constant $c_4 > 0$ such that, for sample size $n \gtrsim \log(pd)$, with probability greater than $1 - c_1(pd)^{-c_4}$, the estimate $\widehat{\mathbf{\Sigma}}_h^{X,Y}$ satisfies the bound

$$\|\widehat{\mathbf{\Sigma}}_h^{X,Y} - \mathbf{\Sigma}_h^{X,Y}\|_{\max} \lesssim \mathcal{M}_{X,Y}\sqrt{\frac{\log(pd)}{n}}. \tag{1.17}$$

Remark 1.3. *In the deviation bounds established above, the effects of dependence are commonly captured by the sum of marginal-spectral and cross-spectral stability measures, $\mathcal{M}_{X,Y} = \mathcal{M}_1^X + \mathcal{M}_1^Y + \mathcal{M}_{1,1}^{X,Y}$, with larger values yielding a slower elementwise ℓ_∞ rate in (1.17). Under a mixed-process scenario consisting of $\{\mathbf{X}_t(\cdot)\}$ and d-dimensional time series $\{\mathbf{Z}_t\}$ belonging to multivariate linear processes with sub-Gaussian errors (Sun et al., 2018), namely sub-Gaussian linear processes, it is easy to extend (1.17) as*

$$\max_{1\leqslant j\leqslant p, 1\leqslant k\leqslant d} \|\widehat{\Sigma}_{h,jk}^{X,Z} - \Sigma_{h,jk}^{X,Z}\| \lesssim \mathcal{M}_{X,Z}\sqrt{\frac{\log(pd)}{n}}, \tag{1.18}$$

where $\mathcal{M}_{X,Z} = \mathcal{M}_1^X + \mathcal{M}_1^Z + \mathcal{M}_{1,1}^{X,Z}$. (1.18) can be justified in the proof of Proposition 1.1 in Appendix 1.B.2.

1.2.5 Rates in elementwise ℓ_∞ norm under a FPCA framework

For each $j = 1, \ldots, p$, suppose that $X_{1j}(\cdot), \ldots, X_{nj}(\cdot)$ are n serially dependent observations of $X_j(\cdot)$. The Karhunen–Loève theorem (Bosq, 2000) serving as the theoretical basis of FPCA allows us to represent each functional observation in the form

of $X_{tj}(\cdot) = \sum_{l=1}^{\infty} \zeta_{tjl}\psi_{jl}(\cdot)$. Here the coefficients $\zeta_{tjl} = \langle X_{tj}, \psi_{jl}\rangle$, namely FPC scores, are uncorrelated random variables with mean zero and $\mathrm{Cov}(\zeta_{tjl}, \zeta_{tjl'}) = \omega_{jl}^X I(l = l')$. In this formulation, $\{(\omega_{jl}^X, \psi_{jl})\}_{l=1}^{\infty}$ are eigenpairs satisfying $\langle \Sigma_{0,jj}^X(u, \cdot), \psi_{jl}(\cdot)\rangle = \omega_{jl}^X \psi_{jl}(u)$. Similarly, for each $k = 1, \ldots, d$, we represent $Y_{tk}(\cdot) = \sum_{m=1}^{\infty} \xi_{tkm}\phi_{km}(\cdot)$ with eigenpairs $\{(\omega_{km}^Y, \phi_{km})\}_{m=1}^{\infty}$.

To estimate relevant terms under a FPCA framework, for each j, we perform an eigenanalysis on $\widehat{\Sigma}_{0,jj}^X(u, v) = n^{-1}\sum_{t=1}^{n} X_{tj}(u)X_{tj}(v)$, i.e. $\langle \widehat{\Sigma}_{0,jj}^X(u, \cdot), \widehat{\psi}_{jl}(\cdot)\rangle = \widehat{\omega}_{jl}^X \widehat{\psi}_{jl}(u)$, where $\{(\widehat{\omega}_{jl}^X, \widehat{\psi}_{jl})\}_{l=1}^{\infty}$ denote the estimated eigenpairs. The corresponding estimated FPC scores are given by $\widehat{\zeta}_{tjl} = \langle X_{tj}, \widehat{\psi}_{jl}\rangle$. Furthermore, relevant estimated terms for $\{Y_{tk}(\cdot)\}$, i.e. $\widehat{\omega}_{km}^Y, \widehat{\phi}_{km}(\cdot), \widehat{\xi}_{tkm}$, can be obtained in the same manner.

Before presenting relevant deviation bounds in elementwise ℓ_∞ norm, which are essential under high-dimensional regime, $(\log p \vee \log d)/n \to 0$, we impose the following lower bound condition on the eigengaps.

Condition 1.5. *For each $j = 1, \ldots, p$ and $k = 1, \ldots, d$, $\omega_{j1}^X > \omega_{j2}^X > \cdots > 0$ and $\omega_{k1}^Y > \omega_{k2}^Y > \cdots > 0$. There exist some positive constants c_0 and $\alpha_1, \alpha_2 > 1$ such that $\omega_{jl}^X - \omega_{j(l+1)}^X \geqslant c_0 l^{-\alpha_1 - 1}$ for $l = 1, \ldots, \infty$ and $\omega_{km}^Y - \omega_{k(m+1)}^Y \geqslant c_0 m^{-\alpha_2 - 1}$ for $m = 1, \ldots, \infty$.*

Condition 1.5 implies the lower bounds on eigenvalues, i.e. $\omega_{jl}^X \geqslant c_0\alpha_1^{-1} l^{-\alpha_1}$ and $\omega_{km}^Y \geqslant c_0\alpha_2^{-1} m^{-\alpha_2}$. See also Kong et al. (2016) and Qiao et al. (2020) for similar conditions.

In practice, the infinite series in the Karhunen–Loève expansions of $X_{tj}(\cdot)$ and $Y_{tm}(\cdot)$ are truncated at M_1 and M_2, chosen data-adaptively, which transforms the infinite-dimensional learning task into the modelling of multivariate time series. Given sub-Gaussian functional linear process $\{\mathbf{X}_t(\cdot)\}$, to aid convergence analysis under high-dimensional scaling, we establish elementwise concentration inequalities and, furthermore, elementwise ℓ_∞ error bounds on relevant estimated terms, i.e. estimated eigenpairs and sample (auto)covariance between estimated FPC scores. These results are of the same forms as those under the Gaussianity assumption (Guo and Qiao, 2022), so we only present them in Lemmas 1.25 and 1.27 in Appendix 1.E.

In the following, we focus on sample cross-(auto)covariance between estimated FPC scores, $\widehat{\sigma}_{h,jklm}^{X,Y} = (n-h)^{-1}\sum_{t=1}^{n-h} \widehat{\zeta}_{tjl}\widehat{\xi}_{(t+h)km}$, and establish a normalized deviation bound in elementwise ℓ_∞ norm on how $\widehat{\sigma}_{h,jklm}^{X,Y}$ concentrates around $\sigma_{h,jklm}^{X,Y} = \mathrm{Cov}(\zeta_{tjl}, \xi_{(t+h)km})$.

Theorem 1.3. *Suppose that Conditions 1.1–1.5 hold for sub-Gaussian functional linear processes, $\{\mathbf{X}_t(\cdot)\}$, $\{\mathbf{Y}_t(\cdot)\}$, and h is fixed. Let M_1 and M_2 be positive integers possibly depending on (n, p, d). If $n \gtrsim \log(pdM_1M_2)(M_1^{4\alpha_1+2} \vee M_2^{4\alpha_2+2})\mathcal{M}_{X,Y}^2$, then*

there exist some positive constants c_5 *and* c_6 *such that, with probability greater than* $1 - c_5(pdM_1M_2)^{-c_6}$, *the estimates* $\{\widehat{\sigma}^{X,Y}_{h,jklm}\}$ *satisfy*

$$\max_{\substack{1\leqslant j\leqslant p, 1\leqslant k\leqslant d\\ 1\leqslant l\leqslant M_1, 1\leqslant m\leqslant M_2}} \frac{\left|\widehat{\sigma}^{X,Y}_{h,jklm} - \sigma^{X,Y}_{h,jklm}\right|}{(l^{\alpha_1+1}\vee m^{\alpha_2+1})\sqrt{\omega^X_{jl}\omega^Y_{km}}} \lesssim \mathcal{M}_{X,Y}\sqrt{\frac{\log(pdM_1M_2)}{n}}. \tag{1.19}$$

In the special case that $\{\mathbf{X}_t(\cdot)\}$ and $\{\mathbf{Y}_t(\cdot)\}$ are identical, (1.19) degenerates to the deviation bound on $\widehat{\sigma}^X_{h,jklm}$ under the Gaussianity assumption (Guo and Qiao, 2022). We next consider a mixed process scenario consisting of $\{\mathbf{X}_t(\cdot)\}$ and $\{\mathbf{Z}_t\}$ and establish a normalized deviation bound in elementwise ℓ_∞ norm on sample cross-(auto)covariance between estimated FPC scores of $\{X_{tj}(\cdot)\}$ and $Z_{(t+h)k}$. Define $\widehat{\varrho}^{X,Z}_{h,jkl} = (n-h)^{-1}\sum_{t=1}^{n-h}\widehat{\zeta}_{tjl}Z_{(t+h)k}$ and $\varrho^{X,Z}_{h,jkl} = \mathrm{Cov}(\zeta_{tjl}, Z_{(t+h)k})$. We are ready to extend (1.19) to the following mixed-process scenario.

Proposition 1.1. *Suppose that Conditions 1.1–1.5 hold for sub-Gaussian functional linear process* $\{\mathbf{X}_t(\cdot)\}$, $\{\mathbf{Z}_t\}$ *follows sub-Gaussian linear process and* h *is fixed. Let* M_1 *be a positive integer possibly depending on* (n,p,d). *If sample size* $n \gtrsim \log(pdM_1)M_1^{3\alpha_1+2}\mathcal{M}^2_{X,Z}$, *then there exist some constants* $c_7, c_8 > 0$ *such that, with probability greater than* $1 - c_7(pdM_1)^{-c_8}$, *the estimates* $\{\widehat{\varrho}^{X,Z}_{h,jkl}\}$ *satisfy*

$$\max_{\substack{1\leqslant j\leqslant p, 1\leqslant k\leqslant d\\ 1\leqslant l\leqslant M_1}} \frac{\left|\widehat{\varrho}^{X,Z}_{h,jkl} - \varrho^{X,Z}_{h,jkl}\right|}{l^{\alpha_1+1}\sqrt{\omega^X_{jl}}} \lesssim \mathcal{M}_{X,Z}\sqrt{\frac{\log(pdM_1)}{n}}. \tag{1.20}$$

We next consider $\{\epsilon_t(\cdot)\}_{t=1}^n$, defined on $\mathcal{V}$, which can be seen as the error term in model (1.1) being independent of $\{\mathbf{X}_t(\cdot)\}$. Define $\Sigma^{X,\epsilon}_{h,j}(u,v) = \mathrm{Cov}\{X_{tj}(u), \epsilon_{t+h}(v)\}$ and its estimate $\widehat{\Sigma}^{X,\epsilon}_{h,j}(u,v) = (n-h)^{-1}\sum_{t=1}^{n-h}X_{tj}(u)\epsilon_{t+h}(v)$. To provide theoretical analysis of the estimates for model (1.1), the FPCA-based representation in Appendix 1.F suggests to investigate the consistency properties of the estimated cross terms, i.e. $\widehat{\sigma}^{X,\epsilon}_{h,jlm} = \langle\widehat{\psi}_{jl}, \langle\widehat{\Sigma}^{X,\epsilon}_{h,j}, \widehat{\phi}_m\rangle\rangle$ or $\widehat{\sigma}^{X,Y}_{h,jlm} = (n-h)^{-1}\sum_{t=1}^{n-h}\widehat{\zeta}_{tjl}\widehat{\xi}_{(t+h)m} = \langle\widehat{\psi}_{jl}, \langle\widehat{\Sigma}^{X,Y}_{h,j}, \widehat{\phi}_m\rangle\rangle$. As $\{\mathbf{X}_{t-h}(\cdot) : h = 0,\ldots,L\}$ and $\{\epsilon_t(\cdot)\}$ are independent and can together determine the response $\{Y_t(\cdot)\}$ via (1.1), it is more sensible to study the former term, i.e. how $\widehat{\sigma}^{X,\epsilon}_{h,jlm}$ deviates from $\sigma^{X,\epsilon}_{h,jlm} = 0$ in the following proposition.

Proposition 1.2. *Suppose that Conditions 1.1–1.5 hold for sub-Gaussian functional linear processes* $\{\mathbf{X}_t(\cdot)\}$, $\{\epsilon_t(\cdot)\}$ *and* h *is fixed. Let* M_1, M_2 *be positive integers possibly depending on* (n,p). *If* $n \gtrsim \log(pM_1M_2)(M_1^{4\alpha_1+4}\vee M_2^{4\alpha_2+4})(\mathcal{M}^X_1 + \mathcal{M}^Y)^2$, *then there exist some constants* $c_9, c_{10} > 0$ *such that, with probability greater than*

$1 - c_9(pM_1M_2)^{-c_{10}}$, *the estimates* $\{\widehat{\sigma}^{X,\epsilon}_{h,jlm}\}$ *satisfy*

$$\max_{\substack{1\leqslant j\leqslant p\\ 1\leqslant l\leqslant M_1, 1\leqslant m\leqslant M_2}} \frac{\left|\widehat{\sigma}^{X,\epsilon}_{h,jlm}\right|}{(l^{\alpha_1}\vee m^{\alpha_2})\sqrt{\omega^X_{jl}\omega^Y_m}} \lesssim (\mathcal{M}^X_1 + \mathcal{M}^\epsilon)\sqrt{\frac{\log(pM_1M_2)}{n}}. \tag{1.21}$$

Finally, we consider a mixed-process scenario in model (1.2), where $\{\epsilon_t\}_{t=1}^n$ are scalar errors, independent of both $\{\mathbf{X}_t(\cdot)\}$ and $\{\mathbf{Z}_t\}$. In addition to Proposition 1.1 above, the following proposition demonstrates how $\widehat{\varrho}^{X,\epsilon}_{h,jl} = (n-h)^{-1}\sum_{t=1}^{n-h}\widehat{\zeta}_{tjl}\epsilon_{t+h}$ converges to $\varrho^{X,\epsilon}_{h,jl} = \mathrm{Cov}(\zeta_{tjl}, \epsilon_{t+h}) = 0$.

Proposition 1.3. *Suppose that Conditions 1.1–1.5 hold for sub-Gaussian functional linear process* $\{\mathbf{X}_t(\cdot)\}$, $\{\epsilon_t\}$ *is sub-Gaussian linear process and* h *is fixed. Let* M_1 *be positive integer possibly depending on* (n,p). *If* $n \gtrsim \log(pM_1)M_1^{3\alpha_1+2}(\mathcal{M}^X_1)^2$, *then there exist some constants* $c_{11}, c_{12} > 0$ *such that, with probability greater than* $1 - c_{11}(pM_1)^{-c_{12}}$, *the estimates* $\{\widehat{\varrho}^{X,\epsilon}_{h,jl}\}$ *satisfy*

$$\max_{1\leqslant j\leqslant p, 1\leqslant l\leqslant M_1} \frac{\left|\widehat{\varrho}^{X,\epsilon}_{h,jl}\right|}{\sqrt{\omega^X_{jl}}} \lesssim (\mathcal{M}^X_1 + \mathcal{M}^\epsilon)\sqrt{\frac{\log(pM_1)}{n}}. \tag{1.22}$$

Remark 1.4. *Benefiting from the independence assumption between* $\{\mathbf{X}_t(\cdot)\}$ *and* $\{\epsilon_t(\cdot)\}$, *Proposition 1.2 leads to a faster rate of convergence in (1.21) compared with (1.19) with* $d = 1$. *Proposition 1.2 also plays a crucial rule in the proof of Proposition 1.7 to demonstrate that, with high probability, model (1.1) satisfies the routinely used deviation condition. Analogously, taking an advantage of the independence assumption between* $\{\mathbf{X}_t(\cdot)\}$ *and* $\{\epsilon_t\}$, *Proposition 1.3 results in a faster rate in (1.22) than that in (1.20) with* $d = 1$. *In the proof of Proposition 1.8, we will apply Proposition 1.3 to verify that, with high probability, model (1.2) satisfies the corresponding deviation condition.*

1.3 High-dimensional functional linear lagged regression

In this section, we first develop a three-step procedure to estimate sparse functional coefficients in model (1.1) and then apply our derived finite sample results in Section 1.2.5 to investigate the convergence properties of the estimates under high-dimensional scaling.

1.3.1 Estimation procedure

Consider functional linear lagged regression model in (1.1), where $\{\beta_{hj} \in \mathbb{S} : h = 0, \ldots, L, j = 1, \ldots, p\}$ are unknown functional coefficients and $\{\epsilon_t(\cdot)\}_{t=1}^n$ are mean-zero errors from sub-Gaussian functional linear process, independent of $\{\mathbf{X}_t(\cdot)\}_{t=1}^n$ from sub-Gaussian functional linear process. Given observed data $\{Y_t, \mathbf{X}_t\}_{t=1}^n$, our goal is to estimate a vector of functional coefficients, $\boldsymbol{\beta} = (\beta_{01}, \ldots, \beta_{0p}, \ldots, \beta_{L1}, \ldots, \beta_{Lp})^{\mathrm{T}}$ with each $\beta_{hj} \in \mathbb{S}$. To assure a feasible solution under a high-dimensional regime, we impose a sparsity assumption on $\boldsymbol{\beta}$. To be specific, we assume that $\boldsymbol{\beta}$ is functional s-sparse with support set $S = \big\{(h, j) \in \{0, \ldots, L\} \times \{1, \ldots, p\} : \|\beta_{hj}\|_{\mathcal{S}} \neq 0\big\}$ and its cardinality $|S| = s$, much smaller than the dimensionality, $p(L+1)$.

Due to the infinite dimensional nature of functional data, we approximate each $X_{tj}(\cdot)$ and $Y_t(\cdot)$ under the Karhunen–Loève expansion truncated at q_{1j} and q_2, respectively, i.e.

$$X_{tj}(\cdot) \approx \sum_{l=1}^{q_{1j}} \zeta_{tjl}\psi_{jl}(\cdot) = \boldsymbol{\zeta}_{tj}^{\mathrm{T}}\boldsymbol{\psi}_j(\cdot), \quad Y_t(\cdot) \approx \sum_{m=1}^{q_2} \xi_{tm}\phi_m(\cdot) = \boldsymbol{\xi}_t^{\mathrm{T}}\boldsymbol{\phi}(\cdot),$$

where $\boldsymbol{\zeta}_{tj} = (\zeta_{tj1}, \ldots, \zeta_{tjq_{1j}})^{\mathrm{T}}$, $\boldsymbol{\psi}_j(\cdot) = \{\psi_{j1}(\cdot), \ldots, \psi_{jq_{1j}}(\cdot)\}^{\mathrm{T}}$, $\boldsymbol{\xi}_t = (\xi_{t1}, \ldots, \xi_{tq_2})^{\mathrm{T}}$ and $\boldsymbol{\phi}(\cdot) = \{\phi_1(\cdot), \ldots, \phi_{q_2}(\cdot)\}^{\mathrm{T}}$. The truncation levels q_{1j} and q_2 are carefully chosen so as to provide reasonable approximations to each $X_{tj}(\cdot)$ and $Y_t(\cdot)$. See Kong et al. (2016) for the selection of the truncated dimension in practice.

According to Appendix 1.F, we can represent model (1.1) in the following matrix form

$$\mathbf{U} = \sum_{h=0}^{L}\sum_{j=1}^{p} \mathbf{V}_{hj}\boldsymbol{\Psi}_{hj} + \mathbf{R} + \mathbf{E}, \tag{1.23}$$

where $\boldsymbol{\Psi}_{hj} = \int_{\mathcal{V}}\int_{\mathcal{U}} \boldsymbol{\psi}_j(u)\beta_{hj}(u, v)\boldsymbol{\phi}(v)^{\mathrm{T}}dudv \in \mathbb{R}^{q_{1j}\times q_2}$, $\mathbf{U} \in \mathbb{R}^{(n-L)\times q_2}$ with its row vectors given by $\boldsymbol{\xi}_{L+1}, \ldots, \boldsymbol{\xi}_n$ and $\mathbf{V}_{hj} \in \mathbb{R}^{(n-L)\times q_{1j}}$ with its row vectors given by $\boldsymbol{\zeta}_{(L+1-h)j}, \ldots, \boldsymbol{\zeta}_{(n-h)j}$. Note $\mathbf{R}$ and $\mathbf{E}$ are $(n-L) \times q_2$ matrices whose row vectors are formed by truncation errors $\{\mathbf{r}_t \in \mathbb{R}^{q_2} : t = L+1, \ldots, n\}$ and random errors $\{\boldsymbol{\epsilon}_t \in \mathbb{R}^{q_2} : t = L+1, \ldots, n\}$ respectively.

We develop the following three-step estimation procedure.

First, we perform FPCA on $\{X_{tj}(\cdot)\}_{t=1}^n$ for each $j = 1, \ldots, p$ and $\{Y_t(\cdot)\}_{t=1}^n$, thus obtaining estimated FPC scores and eigenfunctions, i.e. $\widehat{\zeta}_{tjl}, \widehat{\psi}_{jl}(\cdot)$ for $l \geqslant 1$ and $\widehat{\xi}_{tm}, \widehat{\phi}_{tm}(\cdot)$ for $m \geqslant 1$, respectively.

Second, it is worth noting that the problem of recovering functional sparsity structure in $\boldsymbol{\beta}$ is equivalent to estimating the block sparsity pattern in $\{\boldsymbol{\Psi}_{hj} : h = 0, \ldots, L, j = 1, \ldots, p\}$. Specifically, if $\beta_{hj}(\cdot, \cdot)$ is zero, all entries in $\boldsymbol{\Psi}_{hj}$ will be zero. This motivates us to incorporate a standardized group lasso penalty (Simon and

Tibshirani, 2012) by minimizing the following penalized regression criterion over $\{\mathbf{\Psi}_{hj} : h = 0, \ldots, L, j = 1, \ldots, p\}$:

$$\frac{1}{2}\|\widehat{\mathbf{U}} - \sum_{h=0}^{L}\sum_{j=1}^{p}\widehat{\mathbf{V}}_{hj}\mathbf{\Psi}_{hj}\|_{\mathrm{F}}^2 + \lambda_n \sum_{h=0}^{L}\sum_{j=1}^{p}\|\widehat{\mathbf{V}}_{hj}\mathbf{\Psi}_{hj}\|_{\mathrm{F}}, \tag{1.24}$$

where $\widehat{\mathbf{U}}$ and $\widehat{\mathbf{V}}_{hj}$ are the estimates of $\mathbf{U}$ and $\mathbf{V}_{hj}$, respectively, and λ_n is a non-negative regularization parameter. Let $\{\widehat{\mathbf{\Psi}}_{hj}\}$ be the minimizer of (1.24).

Finally, we estimate functional coefficients by

$$\widehat{\beta}_{hj}(u, v) = \widehat{\boldsymbol{\psi}}_j(u)^{\mathrm{T}}\widehat{\mathbf{\Psi}}_{hj}\widehat{\boldsymbol{\phi}}(v), \quad (u, v) \in \mathcal{U} \times \mathcal{V}, h = 0, \ldots, L, j = 1, \ldots, p.$$

1.3.2 Theoretical properties

We begin with some notation that will be used in this section. For a block matrix $\mathbf{B} = (\mathbf{B}_{jk})_{1\leqslant j\leqslant p_1, 1\leqslant k\leqslant p_2} \in \mathbb{R}^{p_1 q_1 \times p_2 q_2}$ with the (j, k)-th block $\mathbf{B}_{jk} \in \mathbb{R}^{q_1\times q_2}$, we define its (q_1, q_2)-block versions of elementwise ℓ_∞ and matrix ℓ_1 norms by $\|\mathbf{B}\|_{\max}^{(q_1,q_2)} = \max_{j,k}\|\mathbf{B}_{jk}\|_{\mathrm{F}}$ and $\|\mathbf{B}\|_1^{(q_1,q_2)} = \max_k \sum_j \|\mathbf{B}_{jk}\|_{\mathrm{F}}$, respectively. To simplify notation, we will assume the same q_{1j} across $j = 1, \ldots, p$, but our theoretical results extend naturally to the more general setting where q_{1j}'s are different.

Let $\widehat{\mathbf{Z}} = (\widehat{\mathbf{V}}_{01}, \ldots, \widehat{\mathbf{V}}_{0p}, \ldots, \widehat{\mathbf{V}}_{L1}, \ldots, \widehat{\mathbf{V}}_{Lp}) \in \mathbb{R}^{(n-L)\times(L+1)pq_1}$, $\mathbf{\Psi} = (\mathbf{\Psi}_{01}^{\mathrm{T}}, \ldots, \mathbf{\Psi}_{0p}^{\mathrm{T}}, \ldots, \mathbf{\Psi}_{L1}^{\mathrm{T}}, \ldots, \mathbf{\Psi}_{Lp}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{(L+1)pq_1\times q_2}$ and $\widehat{\mathbf{D}} = \mathrm{diag}(\widehat{\mathbf{D}}_{01}, \ldots, \widehat{\mathbf{D}}_{0p}, \ldots, \widehat{\mathbf{D}}_{L1}, \ldots, \widehat{\mathbf{D}}_{Lp}) \in \mathbb{R}^{(L+1)pq_1\times(L+1)pq_1}$ with $\widehat{\mathbf{D}}_{hj} = \{(n-L)^{-1}\widehat{\mathbf{V}}_{hj}^{\mathrm{T}}\widehat{\mathbf{V}}_{hj}\}^{1/2} \in \mathbb{R}^{q_1\times q_1}$ for $h = 0, \ldots, L$ and $j = 1, \ldots, p$. Then minimizing (1.24) over $\{\mathbf{\Psi}_{hj}\}$ is equivalent to the following optimization task:

$$\widehat{\mathbf{B}} = \underset{\mathbf{B}\in\mathbb{R}^{(L+1)pq_1\times q_2}}{\arg\min}\left\{\frac{1}{2(n-L)}\|\widehat{\mathbf{U}} - \widehat{\mathbf{Z}}\widehat{\mathbf{D}}^{-1}\mathbf{B}\|_{\mathrm{F}}^2 + \lambda_n\|\mathbf{B}\|_1^{(q_1,q_2)}\right\}. \tag{1.25}$$

Then we have $\widehat{\mathbf{\Psi}} = \widehat{\mathbf{D}}^{-1}\widehat{\mathbf{B}}$ with its $\{(h+1)j\}$-th row block given by $\widehat{\mathbf{\Psi}}_{hj}$.

Before our convergence analysis, we present the following regularity conditions.

Condition 1.6. *For each $(h, j) \in S$, $\beta_{hj}(u, v) = \sum_{l,m=1}^{\infty} a_{hjlm}\psi_{jl}(u)\phi_m(v)$ and there exist some positive constants $\kappa > (\alpha_1 \vee \alpha_2)/2 + 1$ and μ_{hj} such that $|a_{hjlm}| \leqslant \mu_{hj}(l + m)^{-\kappa-1/2}$ for $l, m \geqslant 1$.*

We expand each non-zero functional coefficient $\beta_{hj}(u, v)$ using principal component functions $\{\psi_{jl}(u)\}_{l\geqslant 1}$ and $\{\phi_m(v)\}_{m\geqslant 1}$, which respectively provide the most rapidly

convergent representation of $\{X_{tj}(u)\}$ and $\{Y_t(v)\}$ in the L_2 sense. Such condition prevents the coefficients $\{a_{hjlm}\}_{l,m\geqslant 1}$ from decreasing too slowly with parameter κ controlling the level of smoothness in non-zero components of $\{\beta_{hj}(\cdot,\cdot)\}$. See similar smoothness conditions in functional linear regression literature (Hall and Horowitz, 2007; Kong et al., 2016).

Condition 1.7. *Denote the covariance matrix function by*

$$\widetilde{\mathbf{\Sigma}}^X = \begin{pmatrix} \mathbf{\Sigma}_0^X & \mathbf{\Sigma}_1^X & \cdots & \mathbf{\Sigma}_L^X \\ \mathbf{\Sigma}_1^X & \mathbf{\Sigma}_0^X & \cdots & \mathbf{\Sigma}_{L-1}^X \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Sigma}_L^X & \mathbf{\Sigma}_{L-1}^X & \cdots & \mathbf{\Sigma}_0^X \end{pmatrix}$$

and the diagonal matrix function by $\widetilde{\mathbf{D}}_0^X = \mathbf{I}_{L+1} \otimes diag(\Sigma_{0,11}^X, \ldots, \Sigma_{0,pp}^X)$. *The infimum* $\underline{\mu}$ *of the functional Rayleigh quotient of* $\widetilde{\mathbf{\Sigma}}^X$ *relative to* $\widetilde{\mathbf{D}}_0^X$ *is bounded below by zero, i.e.*

$$\underline{\mu} = \inf_{\mathbf{\Phi} \in \bar{\mathbb{H}}_0^{(L+1)p}} \frac{\langle \mathbf{\Phi}, \widetilde{\mathbf{\Sigma}}^X(\mathbf{\Phi}) \rangle}{\langle \mathbf{\Phi}, \widetilde{\mathbf{D}}_0^X(\mathbf{\Phi}) \rangle} > 0,$$

where $\mathbf{\Phi} \in \bar{\mathbb{H}}_0^{(L+1)p} = \{\mathbf{\Phi} \in \mathbb{H}^{(L+1)p} : \langle \mathbf{\Phi}, \widetilde{\mathbf{D}}_0^X(\mathbf{\Phi}) \rangle \in (0, \infty)\}$.

Condition 1.7 can be interpreted as requiring the minimum eigenvalue of the correlation matrix function for $(\mathbf{X}_{t-L}^{\mathrm{T}}, \ldots, \mathbf{X}_t^{\mathrm{T}})^{\mathrm{T}}$ to be bounded below by zero. See also a similar condition in Guo and Qiao (2022).

Before presenting the consistency analysis of $\widehat{\boldsymbol{\beta}}$ in Theorem 1.4, we show that the functional analogs of the restricted eigenvalue (RE) condition and the deviation condition in the lasso literature (Loh and Wainwright, 2012) are satisfied with high probability in Proposition 1.4 below and Propositions 1.6–1.7 in Appendix 1.A, respectively.

Proposition 1.4. *Suppose Conditions 1.1–1.5 and 1.7 hold. Then there exist some positive constants* C_Γ, c_1^* *and* c_2^* *such that, for* $n \gtrsim \log(pq_1) q_1^{4\alpha_1+2} (\mathcal{M}_1^X)^2$, *the matrix* $\widehat{\mathbf{\Gamma}} = (n-L)^{-1} \widehat{\mathbf{D}}^{-1} \widehat{\mathbf{Z}}^{\mathrm{T}} \widehat{\mathbf{Z}} \widehat{\mathbf{D}}^{-1} \in \mathbb{R}^{(L+1)pq_1 \times (L+1)pq_1}$ *satisfies, with probability greater than* $1 - c_1^*(pq_1)^{-c_2^*}$,

$$\boldsymbol{\theta}^{\mathrm{T}} \widehat{\mathbf{\Gamma}} \boldsymbol{\theta} \geqslant \tau_2 \|\boldsymbol{\theta}\|^2 - \tau_1 \|\boldsymbol{\theta}\|_1^2 \quad \forall \boldsymbol{\theta} \in \mathbb{R}^{(L+1)pq_1}, \tag{1.26}$$

where $\tau_1 = C_\Gamma \mathcal{M}_1^X q_1^{\alpha_1+1} \sqrt{\log(pq_1)/n}$ *and* $\tau_2 = \underline{\mu}$.

(1.26) can be viewed as the functional extension of RE condition under the FPCA framework. Intuitively, it provides some insight into the eigenstructure of the sample correlation matrix of a vector formed by estimated lagged FPC scores of $\{X_{tj}(\cdot)\}_{j=1}^{p}$. In particular, for any $\boldsymbol{\theta} \in \mathbb{R}^{(L+1)pq_1}$ such that $\tau_1\|\boldsymbol{\theta}\|_1^2/\tau_2\|\boldsymbol{\theta}\|^2$ is relatively small, $\boldsymbol{\theta}^{\mathrm{T}}\widehat{\boldsymbol{\Gamma}}\boldsymbol{\theta}/\|\boldsymbol{\theta}\|^2$ is bounded away from 0. Proposition 1.4 formalize this intuition by showing (1.26) holds with high probability. Furthermore, Propositions 1.6 and 1.7 verify the essential deviation bounds for model (1.1), where further discussions can be found in Appendix 1.A.

Now we are ready to present the main convergence result.

Theorem 1.4. *Suppose that Conditions 1.1–1.7 hold with* $\tau_2 \geqslant 32\tau_1 q_1 q_2 s$. *If* $n \gtrsim \log(pq_1q_2)(q_1^{4\alpha_1+4} \vee q_2^{4\alpha_2+4})(\mathcal{M}_1^X + \mathcal{M}^Y)^2$, *then there exist some positive constants* c_1^* *and* c_2^* *such that, for any regularization parameter,* $\lambda_n \geqslant 2C_0 s q_1^{1/2}\big\{(\mathcal{M}_1^X + \mathcal{M}^\epsilon) \vee \mathcal{M}^Y\big\}\big\{(q_1^{\alpha_1+3/2} \vee q_2^{\alpha_2+3/2})\sqrt{\frac{\log(pq_1q_2)}{n}} + q_1^{-\kappa+1/2}\big\}$ *and* $q_1^{\alpha_1/2} s\lambda_n \to 0$ *as* $n, p, q_1, q_2 \to \infty$, *the estimate* $\widehat{\boldsymbol{\beta}}$ *satisfies*

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_1 \lesssim \frac{q_1^{\alpha_1/2} s\lambda_n}{\underline{\mu}}, \tag{1.27}$$

with probability greater than $1 - c_1^*(pq_1q_2)^{-c_2^*}$.

Remark 1.5. *(a) The error bound of* $\widehat{\boldsymbol{\beta}}$ *under functional* ℓ_1 *norm is determined by sample size* (n), *number of functional variables* (p), *functional sparsity level* (s) *as well as internal parameters, e.g., the convergence rate in (1.27) is better when truncated dimensions* (q_1, q_2), *functional stability measures* $(\mathcal{M}_1^X, \mathcal{M}^\epsilon, \mathcal{M}^Y)$, *decay rates of the lower bounds for eigenvalues* (α_1, α_2) *in Condition 1.5 are small and decay rate of the upper bounds for basis coefficients* (κ) *in Condition 1.6 and curvature* $(\underline{\mu})$ *in (1.26) are large.*

(b) The serial dependence contributes the additional term $(\mathcal{M}_1^X + \mathcal{M}^\epsilon) \vee \mathcal{M}^Y$ *in the error bound. Specifically, the presence of* $\mathcal{M}_1^X + \mathcal{M}^\epsilon$ *is due to Proposition 1.2 under the independence assumption between* $\{\mathbf{X}_t(\cdot)\}$ *and* $\{\epsilon_t(\cdot)\}$, *which is used to verify the deviation bound in Proposition 1.7. Moreover, provided that our estimation is based on the representation in (1.23), formed by eigenfunctions* $\{\phi_m(\cdot)\}$ *of* Σ_0^Y, *the term* $\mathcal{M}^Y$ *comes from the consistency analysis of* $\{\widehat{\phi}_m\}$ *in Proposition 1.6.*

(c) Note that the VFAR model can be rowwisely viewed as a special case of model (1.1). The serial dependence in the error bound of the VFAR estimate is captured by $\mathcal{M}_1^X$ *partially due to its presence in the deviation bounds on estimated cross-covariance between response* $\{\mathbf{X}_t(\cdot)\}$ *and covariates* $\{\mathbf{X}_{t-h}(\cdot) : 1 \leqslant h \leqslant L\}$. *By*

contrast, the serial dependence effect in (1.27) partially comes from estimated cross-covariance between covariates $\{\mathbf{X}_{t-h}(\cdot) : 0 \leqslant h \leqslant L\}$ *and error* $\{\epsilon_t(\cdot)\}$ *instead of that between* $\{\mathbf{X}_{t-h}(\cdot) : 0 \leqslant h \leqslant L\}$ *and response* $\{Y_t(\cdot)\}$ *due to the fact that* $\{Y_t(\cdot)\}$ *is completely determined by* $\{\mathbf{X}_{t-h}(\cdot) : 0 \leqslant h \leqslant L\}$ *and* $\{\epsilon_t(\cdot)\}$ *via (1.1) given* $\boldsymbol{\beta}$*. Specially, if* $\mathcal{M}^{\epsilon} \vee \mathcal{M}^{Y} \lesssim \mathcal{M}_1^{X}$, $q_1 \asymp q_2$ *and* $\alpha_1 = \alpha_2$*, the rate in (1.27) is consistent to that of the VFAR estimate in Guo and Qiao (2022).*

1.4 High-dimensional partially functional linear regression

This section is organized in the same manner as Section 1.3. We first present the three-step procedure to estimate sparse functional and scalar coefficients in model (1.2) and then study the estimation consistency in the high-dimensional regime.

1.4.1 Estimation procedure

Consider partially functional linear regression model in (1.2), where $\mathcal{B}(\cdot) = \{\beta_1(\cdot), \ldots, \beta_p(\cdot)\}^{\mathrm{T}}$ are functional coefficients of functional covariates $\{\mathbf{X}_t(\cdot)\}_{t=1}^n$ and $\gamma = (\gamma_1, \ldots, \gamma_d)^{\mathrm{T}}$ are regression coefficients of scalar covariates $\{\mathbf{Z}_t\}_{t=1}^n$. $\{\epsilon_t\}_{t=1}^n$ are mean-zero errors from sub-Gaussian linear process, independent of $\{\mathbf{Z}_t\}$ from sub-Gaussian linear process and $\{\mathbf{X}_t(\cdot)\}$ from sub-Gaussian functional linear process. To estimate $\mathcal{B}(\cdot)$ and γ under large p and d scenario, we assume some sparsity patterns in model (1.2), i.e. $\mathcal{B}(\cdot)$ is functional s_1-sparse, with support $S_1 = \{j \in \{1, \ldots, p\} : \|\beta_j\| \neq 0\}$ and cardinality $s_1 = |S_1|$, and γ is s_2-sparse, with support $S_2 = \{j \in \{1, \ldots, d\} : \gamma_j \neq 0\}$ and cardinality $s_2 = |S_2|$. Here s_1 and s_2 are much smaller than dimension parameters, p and d, respectively.

Under the Karhunen–Loève expansion of each $X_{tj}(\cdot)$ as described in Section 1.3.1, model (1.2) can be rewritten as

$$Y_t = \sum_{j=1}^{p} \sum_{l=1}^{q_j} \zeta_{tjl} \langle \psi_{jl}, \beta_j \rangle + \sum_{j=1}^{d} Z_{tj}\gamma_j + r_t + \epsilon_t,$$

where $r_t = \sum_{j=1}^{p} \sum_{l=q_j+1}^{\infty} \zeta_{tjl} \langle \psi_{jl}, \beta_j \rangle$. Let $\mathcal{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}} \in \mathbb{R}^n$, $\mathcal{Z} = (\mathcal{Z}_1, \ldots, \mathcal{Z}_d) \in \mathbb{R}^{n \times d}$, $\mathcal{Z}_j = (Z_{1j}, \ldots, Z_{nj})^{\mathrm{T}} \in \mathbb{R}^n$, $\gamma = (\gamma_1, \ldots, \gamma_d)^{\mathrm{T}} \in \mathbb{R}^d$, $\mathcal{X}_j \in \mathbb{R}^{n \times q_j}$ with its row

vectors given by $\boldsymbol{\zeta}_{1j}, \ldots, \boldsymbol{\zeta}_{nj}$ and $\Psi_j = \int_{\mathcal{U}} \boldsymbol{\psi}_j(u)\beta_j(u)du \in \mathbb{R}^{q_j}$. Then we can represent model (1.2) in the following matrix form,

$$\mathcal{Y} = \sum_{j=1}^{p} \mathcal{X}_j \Psi_j + \mathcal{Z}\gamma + R + E, \tag{1.28}$$

where $R = (r_1, \ldots, r_n)^{\mathrm{T}} \in \mathbb{R}^n$ and $E = (\epsilon_1, \ldots, \epsilon_n)^{\mathrm{T}} \in \mathbb{R}^n$ correspond to the truncation and random errors, respectively.

Our proposed three-step estimation procedure proceeds as follows. We start with performing FPCA on each $\{X_{tj}(\cdot)\}_{t=1}^n$, and hence obtain estimated FPC scores $\{\widehat{\zeta}_{tjl}\}$ and eigenfunctions $\{\widehat{\psi}_{jl}(\cdot)\}$. Motivated from (1.28), we then develop a regularized least square approach by incorporating a standardized group lasso penalty for $\{\Psi_j\}_{j=1}^p$ and the lasso penalty for γ, aimed to shrink all elements in Ψ_j of unimportant functional covariates and coefficients of unimportant scalar covariates to be exactly zero. Specifically, we consider minimizing the following criterion over $\Psi_1, \ldots, \Psi_p$ and γ :

$$\frac{1}{2}\|\mathcal{Y} - \sum_{j=1}^{p} \widehat{\mathcal{X}}_j \Psi_j - \mathcal{Z}\gamma\|^2 + \lambda_{n1} \sum_{j=1}^{p} \|\widehat{\mathcal{X}}_j \Psi_j\| + \widetilde{\lambda}_{n2}\|\gamma\|_1, \tag{1.29}$$

where $\widehat{\mathcal{X}}_j$ is the estimate of $\mathcal{X}_j$, and $\lambda_{n1}, \widetilde{\lambda}_{n2}$ are non-negative regularization parameters. Let the minimizers of (1.29) be $\widehat{\Psi}_1, \ldots, \widehat{\Psi}_p$ and $\widehat{\gamma}$. Finally, our estimated functional coefficients are given by $\widehat{\beta}_j(\cdot) = \widehat{\boldsymbol{\psi}}_j(\cdot)^{\mathrm{T}}\widehat{\Psi}_j$ for $j = 1, \ldots, p$.

1.4.2 Theoretical properties

We start with some notation that will be used in this section. For a block vector $B = (b_1^{\mathrm{T}}, \ldots, b_p^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{pq}$ with the j-th block $b_j \in \mathbb{R}^q$, we define its q-block versions of ℓ_1 and elementwise ℓ_∞ norms by $\|B\|_1^{(q)} = \sum_j \|b_j\|$ and $\|B\|_{\max}^{(q)} = \max_j \|b_j\|$, respectively. To simplify our notation, we denote α_1 in Condition 1.5 by α and assume the same truncated dimension across $j = 1, \ldots, p$, denoted by q. Let $\widehat{\mathcal{X}} = (\widehat{\mathcal{X}}_1, \ldots, \widehat{\mathcal{X}}_p) \in \mathbb{R}^{n \times pq}$, $\Psi = (\Psi_1^{\mathrm{T}}, \ldots, \Psi_p^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{pq}$, $\widehat{D} = \mathrm{diag}(\widehat{D}_1, \ldots, \widehat{D}_p) \in \mathbb{R}^{pq \times pq}$, where $\widehat{D}_j = \{n^{-1}\widehat{\mathcal{X}}_j^{\mathrm{T}}\widehat{\mathcal{X}}_j\}^{1/2} \in \mathbb{R}^{q \times q}$ for $j = 1, \ldots, p$. Then our minimizing task in (1.29) is equivalent to

$$(\widehat{B}, \widehat{\gamma}) = \underset{B \in \mathbb{R}^{pq}, \gamma \in \mathbb{R}^d}{\arg\min} \left\{ \frac{1}{2n}\|\mathcal{Y} - \widehat{\Omega}B - \mathcal{Z}\gamma\|^2 + \lambda_{n1}\|B\|_1^{(q)} + \lambda_{n2}\|\gamma\|_1 \right\}, \tag{1.30}$$

where $\widehat{\Omega} = \widehat{\mathcal{X}}\widehat{D}^{-1}$ and $\lambda_{n2} = \widetilde{\lambda}_{n2}/n$. Then $\widehat{\Psi} = \widehat{D}^{-1}\widehat{B}$ with its j-th row block given by $\widehat{\Psi}_j$.

Condition 1.8. *For $j \in S_1$, $\beta_j(u) = \sum_{l=1}^{\infty} a_{jl}\psi_{jl}(u)$ and there exist some positive constants $\kappa > \alpha/2 + 1$ and μ_j such that $|a_{jl}| \leqslant \mu_j l^{-\kappa}$ for $l \geqslant 1$.*

Condition 1.8 controls the level of smoothness for non-zero coefficient functions in $\mathcal{B}(\cdot)$. See also Condition 1.6 for model (1.1) and its subsequent discussion.

Condition 1.9. *For the mixed process $\{\mathbf{X}_t(\cdot), \mathbf{Z}_t\}_{t\in\mathbb{Z}}$, we denote a diagonal matrix function by $\mathbf{D}_0^X = diag(\Sigma_{0,11}^X, \ldots, \Sigma_{0,pp}^X)$. The infimum $\underline{\mu}^*$ is bounded below by zero, i.e.*

$$\underline{\mu}^* = \inf_{\mathbf{\Phi}\in\bar{\mathbb{H}}_0^p, \boldsymbol{\nu}\in\tilde{\mathbb{R}}_0^d} \frac{\langle \mathbf{\Phi}, \mathbf{\Sigma}_0^X(\mathbf{\Phi})\rangle + \langle \mathbf{\Phi}, \mathbf{\Sigma}_0^{X,Z}\boldsymbol{\nu}\rangle + \boldsymbol{\nu}^{\mathrm{T}}\mathbf{\Sigma}_0^{Z,X}(\mathbf{\Phi}) + \boldsymbol{\nu}^{\mathrm{T}}\mathbf{\Sigma}_0^Z\boldsymbol{\nu}}{\langle \mathbf{\Phi}, \mathbf{D}_0^X(\mathbf{\Phi})\rangle + \boldsymbol{\nu}^{\mathrm{T}}\boldsymbol{\nu}} > 0,$$

where $\bar{\mathbb{H}}_0^p = \{\mathbf{\Phi} \in \mathbb{H}^p : \langle \mathbf{\Phi}, \mathbf{D}_0^X(\mathbf{\Phi})\rangle \in (0, \infty)\}$.

This condition is similar to Condition 1.7. In the special case where each $X_{tj}(\cdot)$ is b_j-dimensional, $\underline{\mu}^*$ reduces to the minimum eigenvalue of the covariance matrix of $\big(\frac{\xi_{t11}}{\sqrt{\omega_{11}^X}}, \ldots, \frac{\xi_{t1b_1}}{\sqrt{\omega_{1b_1}^X}}, \ldots, \frac{\xi_{tp1}}{\sqrt{\omega_{p1}^X}}, \ldots, \frac{\xi_{tpb_p}}{\sqrt{\omega_{pb_p}^X}}, Z_{t1}, \ldots, Z_{td}\big)^{\mathrm{T}} \in \mathbb{R}^{\sum_{j=1}^p b_j + d}$.
We next present Proposition 1.5 below and Propositions 1.8–1.9 in Appendix 1.A to respectively show that the RE and deviation conditions are satisfied with high probability. These results together with Proposition 1.6(i) lead to theoretical guarantees for regularized estimates of model (1.2).

Proposition 1.5. *Suppose Conditions 1.1–1.5 and 1.9 hold. Let $\mathcal{S} = (\widehat{\Omega}, \mathcal{Z}) \in \mathbb{R}^{n\times(pq+d)}$, then there exist some positive constants $C_{Z\Gamma}, c_1^*$ and c_2^* such that, for $n \gtrsim \log(pqd)q^{4\alpha+2}\mathcal{M}_{X,Z}^2$, with probability greater than $1 - c_1^*(pq+d)^{-c_2^*}$,*

$$\frac{1}{n}\boldsymbol{\theta}^{\mathrm{T}}\mathcal{S}^{\mathrm{T}}\mathcal{S}\boldsymbol{\theta} \geqslant \tau_2^*\|\boldsymbol{\theta}\|^2 - \tau_1^*\|\boldsymbol{\theta}\|_1^2, \quad \forall \boldsymbol{\theta} \in \mathbb{R}^{pq+d}, \tag{1.31}$$

where $\tau_1^* = C_{Z\Gamma}\mathcal{M}_{X,Z}q^{\alpha+1}\sqrt{\frac{\log(pq+d)}{n}}$ and $\tau_2^* = \underline{\mu}^*$.

Instead of verifying RE conditions on $n^{-1}\widehat{\Omega}^{\mathrm{T}}\widehat{\Omega}$ and $n^{-1}\mathcal{Z}^{\mathrm{T}}\mathcal{Z}$ separately, since $\widehat{\Omega}$ is correlated with $\mathcal{Z}$, we define $\mathcal{S} = (\widehat{\Omega}, \mathcal{Z})$ and verify (1.31), which requires $n^{-1}\boldsymbol{\theta}^{\mathrm{T}}\mathcal{S}^{\mathrm{T}}\mathcal{S}\boldsymbol{\theta}$ to be strictly positive as long as $\tau_1^*\|\boldsymbol{\theta}\|_1^2/\tau_2^*\|\boldsymbol{\theta}\|^2$ is relatively small. Let $\boldsymbol{\theta} = (\Delta^{\mathrm{T}}, \delta^{\mathrm{T}})^{\mathrm{T}}$ with $\Delta = \widehat{B} - B$ and $\delta = \widehat{\gamma} - \gamma$, applying Proposition 1.5 with suitable choice of τ_2^* yields that, with high probability, $n^{-1}(\widehat{\Omega}\Delta + \mathcal{Z}\delta)^{\mathrm{T}}(\widehat{\Omega}\Delta + \mathcal{Z}\delta) \geqslant \frac{\tau_2^*}{4}(\|\Delta\| + \|\delta\|)^2$, which plays a crucial role in the proof of Theorem 1.5 below. Similar to Proposition 1.7,

Propositions 1.8 and 1.9 in Appendix 1.A verify that, with high probability, the essential deviation bounds hold for model (1.2).

Now we are ready to present the main theorem about the error bound for $\widehat{B}$ and $\widehat{\gamma}$.

Theorem 1.5. *Suppose that Conditions 1.1–1.5, 1.8 and 1.9 hold with* $\tau_2^* \geqslant 64\tau_1^* q(s_1+s_2)$. *If* $n \gtrsim \log(pqd)q^{4\alpha+2}\mathcal{M}_{X,Z}^2$, *then, for any regularization parameters,* $\lambda_n \asymp \lambda_{n1} \asymp \lambda_{n2} \geqslant 2C_0^* s_1(\mathcal{M}_{X,Z}+\mathcal{M}^\epsilon)[q^{\alpha+2}\{\log(pq+d)/n\}^{1/2}+q^{-\kappa+1}]$ *with* $q^{\alpha/2}\lambda_n(s_1+s_2)\to 0$ *as* $n,p,q,d\to\infty$, *the estimates* $\widehat{\mathcal{B}}$ *and* $\widehat{\gamma}$ *satisfy*

$$\|\widehat{\mathcal{B}}-\mathcal{B}\|_1 + q^{\alpha/2}\|\widehat{\gamma}-\gamma\|_1 \lesssim \frac{q^{\alpha/2}\lambda_n(s_1+s_2)}{\underline{\mu}^*}, \tag{1.32}$$

with probability greater than $1-c_1^*(pq+d)^{-c_2^*}$.

Remark 1.6. *(a) The error bound in (1.32) is governed by both dimensionality parameters* (n,p,d,s_1,s_2) *and internal parameters* $(\mathcal{M}^X,\mathcal{M}^Z,\mathcal{M}^{X,Z},\mathcal{M}^\epsilon,q,\alpha,\kappa,\underline{\mu}^*)$. *See also similar Remark 1.5 (a) for model (1.1).*

(b) Note that the sparse stochastic regression (Basu and Michailidis, 2015; Wu and Wu, 2016) can be viewed as a special case of model (1.2) without the functional part. Under such scenario, the absence of $\{\mathbf{X}_t(\cdot)\}$ *degenerates (1.37) in Proposition 1.9 to* $n^{-1}\|\mathcal{Z}^{\mathrm{T}}(\mathcal{Y}-\mathcal{Z}\gamma)\|_{\max} \leqslant \widetilde{C}_0(\mathcal{M}_1^Z+\mathcal{M}^\epsilon)(\log d/n)^{1/2}$ *and simplifies the error bound to* $\|\widehat{\gamma}-\gamma\|_1 \lesssim \lambda_{n2}s_2/\tau_2^*$ *with* $\lambda_{n2} \geqslant 2\widetilde{C}_0(\mathcal{M}_1^Z+\mathcal{M}^\epsilon)(\log d/n)^{1/2}$ *for some positive constant* $\widetilde{C}_0$, *which is of the same order as the rate in Basu and Michailidis (2015).*

(c) In another special scenario where scalar covariates are not included in (1.2), the error bound reduces to $\|\widehat{\mathcal{B}}-\mathcal{B}\|_1 \lesssim q^{\alpha/2}\lambda_{n1}s_1/\tau_2^*$ *with* $\lambda_{n1} \geqslant 2C_0^* s_1(\mathcal{M}_1^X+\mathcal{M}^\epsilon)\{q^{\alpha+2}\sqrt{\frac{\log(pq)}{n}}+q^{-\kappa+1}\}$. *Interestingly, this rate is consistent to that of* $\widehat{\boldsymbol{\beta}}$ *in Theorem 1.4 under the special case where the non-functional response results in the absence of* $\mathcal{M}^Y$ *and* q_2 *in the rate.*

1.5 Simulation studies

We conduct a number of simulations to evaluate the finite-sample performance of our proposed ℓ_1/ℓ_2-penalized least squares estimators (ℓ_1/ℓ_2-LS) for models (1.1) and (1.2) in Sections 1.5.1 and 1.5.2, respectively.

1.5.1 High-dimensional functional linear lagged regression

We consider model (1.1) with $L = 1$, where functional covariates $\{\mathbf{X}_t(\cdot)\}_{t=1,\ldots,n}$ are generated from a sparse VFAR model (Guo and Qiao, 2022). Specifically, we generate $X_{tj}(u) = \boldsymbol{\zeta}_{tj}^{\mathrm{T}}\boldsymbol{\psi}(u)$ for $j = 1,\ldots,p$ and $u \in \mathcal{U} = [0,1]$, where $\boldsymbol{\psi}(\cdot) = \{\psi_1(\cdot),\ldots,\psi_5(\cdot)\}^{\mathrm{T}}$ is a 5-dimensional Fourier basis function and $\boldsymbol{\zeta}_t = (\boldsymbol{\zeta}_{t1}^{\mathrm{T}},\ldots,\boldsymbol{\zeta}_{tp}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{5p}$ are generated from a stationary block sparse vector autoregressive (VAR) model, $\boldsymbol{\zeta}_t = \mathbf{W}\boldsymbol{\zeta}_{t-1} + \boldsymbol{\eta}_t$. The transition matrix $\mathbf{W} = (\mathbf{W}_{jk})_{p\times p} \in \mathbb{R}^{5p\times 5p}$ is block sparse such that $\sum_{k=1}^{p} I(\|\mathbf{W}_{jk}\|_{\mathrm{F}} \neq 0) = 5$ for each j, and $\boldsymbol{\eta}_t$ are sampled independently from $N(\mathbf{0}, \mathbf{I}_{5p})$. The nonzero elements in $\mathbf{W}$ are sampled from $N(0,1)$ and we rescale $\mathbf{W}$ by $\iota\mathbf{W}/\rho(\mathbf{W})$ with $\iota \sim \mathrm{Unif}[0.5,1]$ to guarantee the stationarity of $\{\boldsymbol{\zeta}_t\}$. For each $(h,j) \in S = \{0,1\} \times \{1,\ldots,5\}$, we generate non-zero functional coefficients $\beta_{hj}(u,v) = \sum_{l,m=1}^{5} b_{hjlm}\psi_l(u)\psi_m(v)$, where b_{hjlm}'s are sampled from $\mathrm{Unif}(0,0.4)$ for $h = 0$ and $\mathrm{Unif}(0,0.15)$ for $h = 1$. The functional responses $\{Y_t(v) : v \in \mathcal{V}\}_{t=1,\ldots,n}$ with $\mathcal{V} = [0,1]$ are then generated from model (1.1), where $\epsilon_t(v) = \sum_{m=1}^{5} e_{tm}\psi_m(v)$ with e_{tm}'s being independent $N(0,1)$ variables.

In our simulations, we consider $n = 75, 100, 150$ dependent observations for $p = 40, 80$ and replicate each simulation 100 times. The truncated dimensions q_{1j} for $j = 1,\ldots p$ and q_2 are selected by the ratio-based method (Lam and Yao, 2012). To select the regularization parameter λ_n, there exists several possible methods such as AIC/BIC and cross-validation. The AIC/BIC requires to specify the effective degrees of freedom, which poses a challenging task for functional data under the high-dimensional setting and is left for future study. In this example, we generate two separate training and validation samples of the same size n. For a sequence of λ_n values, we implement the block fast iterative shrinkage-thresholding (FISTA) algorithm (Guo and Qiao, 2022) to solve the optimization problem (1.24) on the training data, obtain $\{\widehat{\beta}_{hj}^{(\lambda_n)}(\cdot,\cdot)\}_{h=0,1,j=1,\ldots,p}$ as a function of λ_n, calculate the squared error between observed and fitted responses on the validation set, i.e. $\sum_{t=1}^{n}\|Y_t(\cdot) - \sum_{h=0}^{L}\sum_{j=1}^{p}\int_{\mathcal{U}} X_{(t-h)j}(u)\widehat{\beta}_{hj}^{(\lambda_n)}(u,\cdot)du\|^2$ and choose the optimal $\widehat{\lambda}_n$ with the smallest error.

We evaluate the performance of ℓ_1/ℓ_2-LS in terms of both model selection consistency and estimation accuracy. For model selection consistency, we plot the true positive rates against false positive rates, respectively defined as

$$\frac{\#\{(h,j) : \|\widehat{\beta}_{hj}^{(\lambda_n)}\|_{\mathcal{S}} \neq 0 \text{ and } \|\beta_{hj}\|_{\mathcal{S}} \neq 0\}}{\#\{(h,j) : \|\beta_{hj}\|_{\mathcal{S}} \neq 0\}},$$

$$\frac{\#\{(h,j) : \|\widehat{\beta}_{hj}^{(\lambda_n)}\|_{\mathcal{S}} \neq 0 \text{ and } \|\beta_{hj}\|_{\mathcal{S}} = 0\}}{\#\{(h,j) : \|\beta_{hj}\|_{\mathcal{S}} = 0\}}$$

Table 1.1: The mean and standard error (in parentheses) of AUROCs and estimation errors for model (1.1) over 100 simulation runs.

n	p	ℓ_1/ℓ_2-LS		OLS-O
		AUROC	Estimation error	Estimation error
75	40	0.849(0.006)	0.727(0.005)	1.116(0.011)
	80	0.834(0.007)	0.768(0.005)	1.121(0.012)
100	40	0.898(0.005)	0.648(0.005)	0.777(0.006)
	80	0.879(0.007)	0.684(0.005)	0.787(0.006)
150	40	0.953(0.004)	0.544(0.004)	0.550(0.004)
	80	0.942(0.004)	0.576(0.004)	0.547(0.004)

over a grid of values of λ_n to produce a ROC curve, and then calculate the *area under the ROC curve* (AUROC) with values closer to 1 indicating better performance in support recovery. The estimation accuracy is measured by the relative estimation error $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_{\mathrm{F}}/\|\boldsymbol{\beta}\|_{\mathrm{F}}$. For comparison, we also implement the ordinary least squares in the oracle case (OLS-O), which uses the true sparsity structure in the estimates and does not perform variable selection. Table 1.1 gives some numerical summaries. Several conclusions can be drawn. First, the model selection consistency and estimation accuracy are improved as n increases or p decreases. Second, ℓ_1/ℓ_2-LS provides substantially improved estimation accuracy over OLS-O especially in the "large p, small n" scenario. This is not surprising, since implementing OLS-O in the sense of (1.24) with $\lambda_n = 0$ still require to estimate $10 \times 5^2 = 250$ parameters, which is intrinsically a high-dimensional estimation problem.

1.5.2 High-dimensional partially functional linear regression

We now consider model (1.2) with p-dimensional vector of functional covariates $\{\mathbf{X}_t(\cdot)\}_{t=1,\ldots,n}$ and d-dimensional scalar covariates $\{\mathbf{Z}_t\}_{t=1,\ldots,n}$, which are jointly generated in a similar procedure as in Section 1.5.1. Let $X_{tj}(u) = \boldsymbol{\zeta}_{tj}^{\mathrm{T}}\boldsymbol{\psi}(u)$ for $j = 1,\ldots,p$ and $u \in [0,1]$, and $(\boldsymbol{\zeta}_t^{\mathrm{T}}, \mathbf{Z}_t^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{5p+d}$ are jointly generated from a stationary VAR(1) process with a block sparse transition matrix $\mathbf{W}^* \in \mathbb{R}^{(5p+d)\times(5p+d)}$, whose (j,k)-th block is $\mathbf{W}_{jk}^*$. In particular, for each $j = 1,\ldots,p$, $\mathbf{W}_{jk}^* \in \mathbb{R}^{5\times 5}$ $(k = 1,\ldots,p)$ and $\mathbf{W}_{jk}^* \in \mathbb{R}^5$ $(k = p+1,\ldots,p+d)$ such that $\sum_{k=1}^{p} I(\|\mathbf{W}_{jk}^*\|_{\mathrm{F}} \neq 0) = \sum_{k=p+1}^{p+d} I(\|\mathbf{W}_{jk}^*\| \neq 0) = 5$. For each $j = p+1,\ldots,p+d$, $(\mathbf{W}_{jk}^*)^{\mathrm{T}} \in \mathbb{R}^5$ $(k = 1,\ldots,p)$ and $\mathbf{W}_{jk}^* \in \mathbb{R}$ $(k = p+1,\ldots,p+d)$ such that $\sum_{k=1}^{p} I\big(\|(\mathbf{W}_{jk}^*)^{\mathrm{T}}\| \neq 0\big) = \sum_{k=p+1}^{p+d} I(|\mathbf{W}_{jk}^*| \neq 0) = 5$. For each $j \in S_1 = \{1,\ldots,5\}$, the non-zero functional coefficients are generated by $\beta_j(u) = \sum_{l=1}^{5} b_{jl}\psi_l(u)$, where b_{jl}'s are uniformly sampled from $[0, 0.15]$. For each $k \in S_2 = \{1,\ldots,10\}$, the non-zero scalar coefficients

Table 1.2: The mean and standard error (in parentheses) of AUROCs and estimation errors for model (1.2) over 100 simulation runs.

n	$p=d$	ℓ_1/ℓ_2-LS			OLS-O	
		AUROC	$\|\widehat{\mathcal{B}}-\mathcal{B}\|/\|\mathcal{B}\|$	$\|\widehat{\gamma}-\gamma\|/\|\gamma\|$	$\|\widehat{\mathcal{B}}-\mathcal{B}\|/\|\mathcal{B}\|$	$\|\widehat{\gamma}-\gamma\|/\|\gamma\|$
75	40	0.901(0.004)	1.034(0.013)	0.283(0.005)	1.741(0.034)	0.196(0.005)
	80	0.868(0.004)	1.051(0.012)	0.363(0.008)	1.750(0.039)	0.198(0.005)
100	40	0.919(0.003)	0.999(0.007)	0.235(0.005)	1.376(0.024)	0.151(0.004)
	80	0.902(0.004)	1.025(0.008)	0.283(0.005)	1.417(0.025)	0.151(0.004)
150	40	0.945(0.003)	0.938(0.008)	0.185(0.004)	1.006(0.018)	0.113(0.003)
	80	0.937(0.004)	0.972(0.009)	0.216(0.004)	1.061(0.018)	0.113(0.003)

γ_k's are uniformly sampled from $[0.5, 1]$. Finally, we generate responses $\{Y_t\}_{t=1,\ldots,n}$ from model (1.2), where ϵ_t's are sampled from $N(0,1)$.

We simulate the data under six different settings, where $n \in \{75, 100, 150\}$ and $p = d \in \{40, 80\}$, and replicate each simulation 100 times. For a sequence of pairs of $(\lambda_{n1}, \lambda_{n2})$, following the procedure in Section 1.4.1, we truncate each functional covariate with q_j chosen by the ratio-based method, apply the block FISTA algorithm to minimize the criterion (1.29) on the training data and obtain $\{\widehat{\beta}_j^{(\lambda_{n1},\lambda_{n2})}(\cdot)\}_{j=1,\ldots,p}$ and $\{\widehat{\gamma}_k^{(\lambda_{n1},\lambda_{n2})}\}_{k=1,\ldots,d}$. The optimal regularization parameters $(\widehat{\lambda}_{n1}, \widehat{\lambda}_{n2})$ are selected by minimizing the prediction error on the validation data with size n, i.e. $\sum_{t=1}^{n}\{Y_t - \sum_{j=1}^{p}\int_{\mathcal{U}} X_{tj}(u)\widehat{\beta}_j^{(\lambda_{n1},\lambda_{n2})}(u)du - \sum_{k=1}^{d} Z_{tk}\widehat{\gamma}_k^{(\lambda_{n1},\lambda_{n2})}\}^2$.

We examine the performance of ℓ_1/ℓ_2-LS based on AUROCs and estimation errors, and compare it with the performance of OLS-O, where the sparsity structures in the estimates are determined by the true model in advance. The numerical results are summarized in Table 1.2, where the relative estimation errors for functional and scalar coefficients are $\|\widehat{\mathcal{B}}-\mathcal{B}\|/\|\mathcal{B}\|$ and $\|\widehat{\gamma}-\gamma\|/\|\gamma\|$, respectively. A few trends are apparent. First, as expected, we obtain improved overall support recovery and estimation accuracies as n increases or p and d decrease. Second, although ℓ_1/ℓ_2-LS is outperformed by OLS-O with lower estimation errors for scalar coefficients, it provides more accurate estimates of functional coefficients relative to OLS-O, since, in the oracle case, the number of unknown parameters is still relatively large especially when n is small.

1.6 Discussion

We identify several directions for future study. First, it is possible to extend our established finite sample theory for stationary functional linear processes with sub-Gaussian errors to that with more general noise distributions, e.g. generalized sub-exponential process, or even non-stationary functional processes. Second, it is of

interest to develop useful non-asymptotic results under other commonly adopted dependence framework, e.g. moment-based dependence measure (Hörmann and Kokoszka, 2010) and different types of mixing conditions (Bosq, 2000). However, moving from standard asymptotic analysis to non-asymptotic analysis would pose complicated theoretical challenges. Third, from a frequency domain perspective, it is interesting to study the non-asymptotic behaviour of smoothed periodogram estimators (Panaretos and Tavakoli, 2013) for spectral density matrix function, served as the frequency domain analog of the sample covariance matrix function. Under a high-dimensional regime, it is also interesting to develop the functional thresholding strategy to estimate sparse spectral density matrix functions. These topics are beyond the scope of the current chapter and will be pursued elsewhere.

1.A Additional theoretical results

We first present the following Propositions 1.6 and 1.7, in which we show that the essential deviation bounds for model (1.1) are satisfied with high probability.

Proposition 1.6. *Suppose that Conditions 1.1–1.5 hold. Then there exist some positive constants C_ψ, C_ω, C_ϕ, c_1^* and c_2^* such that (i) for $n \gtrsim \log(pq_1)q_1^{4\alpha_1+2}(\mathcal{M}_1^X)^2$,*

$$\begin{aligned} \max_{1\leqslant j\leqslant p, 1\leqslant l\leqslant q_1} \left| \frac{\{\widehat{\omega}_{jl}^X\}^{-1/2} - \{\omega_{jl}^X\}^{-1/2}}{\{\omega_{jl}^X\}^{-1/2}} \right| &\leqslant C_\omega \mathcal{M}_1^X \sqrt{\frac{\log(pq_1)}{n}}, \\ \max_{1\leqslant j\leqslant p, 1\leqslant l\leqslant q_1} \|\widehat{\psi}_{jl} - \psi_{jl}\| &\leqslant C_\psi \mathcal{M}_1^X q_1^{\alpha_1+1} \sqrt{\frac{\log(pq_1)}{n}}, \end{aligned} \tag{1.33}$$

with probability greater than $1 - c_1^*\{pq_1\}^{-c_2^*}$; (ii) for $n \gtrsim \log(q_2)q_2^{4\alpha_2+2}(\mathcal{M}^Y)^2$,

$$\max_{1\leqslant m\leqslant q_2} \|\widehat{\phi}_m - \phi_m\| \leqslant C_\phi \mathcal{M}^Y q_2^{\alpha_2+1} \sqrt{\frac{\log(q_2)}{n}}, \tag{1.34}$$

with probability greater than $1 - c_1^*\{q_2\}^{-c_2^*}$.

Proposition 1.7. *Suppose that Conditions 1.1–1.6 hold. Then there exist some positive constants C_0, c_1^* and c_2^* such that, for $n \gtrsim \log(pq_1q_2)(q_1^{4\alpha_1+4} \vee q_2^{4\alpha_2+4})(\mathcal{M}_1^X + \mathcal{M}^Y)^2$,*

$$\begin{aligned} &(n-L)^{-1}\|\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}(\widehat{\mathbf{U}} - \widehat{\mathbf{Z}}\widehat{\mathbf{D}}^{-1}\mathbf{B})\|_{\max}^{(q_1,q_2)} \\ \leqslant &C_0 s q_1^{1/2}\{(\mathcal{M}_1^X + \mathcal{M}^\epsilon) \vee \mathcal{M}^Y\} \left\{ (q_1^{\alpha_1+3/2} \vee q_2^{\alpha_2+3/2}) \sqrt{\frac{\log(pq_1q_2)}{n}} + q_1^{-\kappa+1/2} \right\}, \end{aligned} \tag{1.35}$$

with probability greater than $1 - c_1^*(pq_1q_2)^{-c_2^*}$.

(1.33) and (1.34) in Proposition 1.6 control deviation bounds for relevant estimated eigenpairs of $X_{tj}(\cdot)$ and $Y_t(\cdot)$ under the FPCA framework. (1.35) in Proposition 1.7 ensures that the sample cross-covariance between estimated lagged-and-normalized FPC scores and estimated errors consisting of truncated and random errors due to (1.23), are nicely concentrated around zero.

We next provide Propositions 1.8 and 1.9, where the essential deviation bounds for model (1.2) hold with high probability.

Proposition 1.8. *Suppose Conditions 1.1–1.5 and 1.8 hold. Then there exist some positive constants C_0^*, c_1^* and c_2^* such that, for $n \gtrsim \log(pq)q^{4\alpha+2}(\mathcal{M}_1^X)^2$,*

$$\frac{1}{n}\|\widehat{\Omega}^{\mathrm{T}}(\mathcal{Y} - \widehat{\Omega}B - \mathcal{Z}\gamma)\|_{\max}^{(q)} \leqslant C_0^* s_1(\mathcal{M}_1^X + \mathcal{M}^\epsilon)\left\{q^{\alpha+2}\sqrt{\frac{\log(pq)}{n}} + q^{-\kappa+1}\right\}. \quad (1.36)$$

with probability greater than $1 - c_1^*(pq)^{-c_2^*}$.

Proposition 1.9. *Suppose Conditions 1.1–1.5 and 1.8 hold. Then there exist some positive constants C_0^*, c_1^* and c_2^* such that, for $n \gtrsim \log(pqd)q^{3\alpha+2}\mathcal{M}_{X,Z}^2$,*

$$\frac{1}{n}\|\mathcal{Z}^{\mathrm{T}}(\mathcal{Y} - \widehat{\Omega}B - \mathcal{Z}\gamma)\|_{\max} \leqslant C_0^* s_1(\mathcal{M}_{X,Z} + \mathcal{M}^\epsilon)\left\{q^{\alpha+1}\sqrt{\frac{\log(pq+d)}{n}} + q^{-\kappa+1/2}\right\}, \quad (1.37)$$

with probability greater than $1 - c_1^*(pq+d)^{-c_2^*}$.

Intuitively, (1.36) in Proposition 1.8 (or (1.37) in Proposition 1.9) indicates the sample cross-covariance between estimated normalized FPC scores (or scalar covariates) and estimated errors is nicely concentrated around zero.

1.B Proofs of theoretical results in Section 1.2

We provide proofs of theorems and propositions stated in Section 1.2 in Appendices 1.B.1–1.B.2, followed by the supporting technical lemmas and their proofs in Appendix 1.B.3. Throughout, we use $C_0, C_1, \ldots$, $c, c_1, \ldots$, $\tilde{c}_1, \tilde{c}_2, \ldots$, $\rho, \rho_1, \rho_2, \ldots$ to denote positive constants. For a matrix $\mathbf{B} \in \mathbb{R}^{p\times q}$, we denote its operator norm by $\|\mathbf{B}\| = \sup_{\|\mathbf{x}\|_2\leqslant 1}\|\mathbf{B}\mathbf{x}\|_2$. For $\phi_1, \phi_2 \in \mathbb{H}$ and $K \in \mathbb{S}$, we respectively denote $\int_{\mathcal{U}} K(u,v)\phi_1(u)du$, $\int_{\mathcal{V}} K(u,v)\phi_2(v)dv$ and $\int_{\mathcal{U}}\int_{\mathcal{V}} K(u,v)\phi_1(u)\phi_2(v)dudv$ by $\langle\phi_1, K\rangle$, $\langle K, \phi_2\rangle$ and $\langle\phi_1, \langle K, \phi_2\rangle\rangle$. For a fixed $\mathbf{\Phi} \in \mathbb{H}^p$, we denote $\mathcal{M}(\boldsymbol{f}^X, \mathbf{\Phi}) = 2\pi \cdot \operatorname{ess\,sup}_{\theta\in[-\pi,\pi]} |\langle \mathbf{\Phi}, \boldsymbol{f}_\theta^X(\mathbf{\Phi})\rangle|$.

1.B.1 Proofs of theorems

Proof of Theorem 1.1 Part (i): Define $\mathbf{Y} = (\langle\mathbf{\Phi}_1, \mathbf{X}_1\rangle, \ldots, \langle\mathbf{\Phi}_1, \mathbf{X}_n\rangle)^{\mathrm{T}}$, then we obtain $|\langle\mathbf{\Phi}_1, (\widehat{\mathbf{\Sigma}}_0^X - \mathbf{\Sigma}_0^X)(\mathbf{\Phi}_1)\rangle| = \frac{1}{n}|\mathbf{Y}^{\mathrm{T}}\mathbf{Y} - \mathbb{E}(\mathbf{Y}^{\mathrm{T}}\mathbf{Y})|$. Our proof is organised as follows: We first introduce the M-truncated sub-Gaussian process $\mathbf{X}_{M,L,t}(u) = \sum_{l=0}^L \mathbf{A}_l(\boldsymbol{\varepsilon}_{M,t-l})$, where $\varepsilon_{M,tj}(\cdot) = \sum_{l=1}^M \sqrt{\omega_{jl}^\varepsilon}a_{tjl}\phi_{jl}(\cdot)$ for $j = 1, \ldots, p$. We then apply the inequality in Lemma 1.5 on $\mathbf{X}_{\infty,L,t} = \mathbf{X}_{L,t}(u) = \sum_{l=0}^L \mathbf{A}_l(\boldsymbol{\varepsilon}_{t-l})$ by proving

$\|\mathbf{\Pi}_{M,L}\| \leqslant \mathcal{M}(\boldsymbol{f}^X_{M,L}, \mathbf{\Phi}_1)$ and $\lim_{M\to\infty} \mathcal{M}(\boldsymbol{f}^X_{M,L}, \mathbf{\Phi}_1) = \mathcal{M}(\boldsymbol{f}^X_{L}, \mathbf{\Phi}_1)$. Finally, we will show that such inequality still holds as $L \to \infty$.

When L and M are both fixed, we first define $\mathbf{Y}_{M,L} = (\langle \mathbf{\Phi}_1, \mathbf{X}_{M,L,1}\rangle, \ldots, \langle \mathbf{\Phi}_1, \mathbf{X}_{M,L,n}\rangle)^{\mathrm{T}}$. Then $\mathbf{Y}^{\mathrm{T}}_{M,L}\mathbf{Y}_{M,L}$ can be represented in the same form as $\langle \mathbf{e}_M, \mathbf{K}(\mathbf{e}_M)\rangle$ in Lemma 1.5, where $\mathbf{e}_M = (\boldsymbol{\varepsilon}^{\mathrm{T}}_{M,n}, \ldots, \boldsymbol{\varepsilon}^{\mathrm{T}}_{M,1-L})^{\mathrm{T}} \in \mathbb{H}^{(n+L)p}$. We rewrite $\mathbf{Y}_{M,L}$ as

$$\mathbf{Y}_{M,L} = \int\int (\mathbf{I}_n \otimes \mathbf{\Phi}_1(u)^{\mathrm{T}})\mathbf{W}_L(u,v)\mathbf{\Theta}_M(v)\mathbf{a}_{M,L}dudv = \mathbf{\Gamma}_{M,L}\mathbf{a}_{M,L},$$

where

$$\mathbf{W}_L = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{A}_0 & \cdots & \mathbf{A}_{L-1} & \mathbf{A}_L \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_0 & \mathbf{A}_1 & \cdots & \mathbf{A}_L & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{A}_0 & \mathbf{A}_1 & \cdots & \cdots & \cdots & \mathbf{A}_L & \cdots & \mathbf{0} \end{pmatrix},$$

$\mathbf{\Theta}_M(u) = \mathbf{I}_{n+L} \otimes \mathrm{diag}(\boldsymbol{\varphi}^{\mathrm{T}}_{M,1}, \ldots, \boldsymbol{\varphi}^{\mathrm{T}}_{M,p})$ with $\boldsymbol{\varphi}_{M,i} = \left(\sqrt{\omega^e_{i1}}\phi_{i1}, \ldots, \sqrt{\omega^e_{iM}}\phi_{iM}\right)^{\mathrm{T}}$ and $\mathbf{a}_{M,L} = (a_{n11}, \ldots, a_{n1M}, \ldots, a_{np1}, \ldots, a_{npM}, \ldots, a_{(1-L)p1}, \ldots, a_{(1-L)pM})^{\mathrm{T}} \in \mathbb{R}^{(n+L)pM}$. Then we can write $\mathbf{Y}^{\mathrm{T}}_{M,L}\mathbf{Y}_{M,L} = \mathbf{a}^{\mathrm{T}}_{M,L}\mathbf{\Pi}_{M,L}\mathbf{a}_{M,L}$ with $\mathbf{\Pi}_{M,L} = \mathbf{\Gamma}_{M,L}{}^{\mathrm{T}}\mathbf{\Gamma}_{M,L}$. Lemma 1.8 implies that $\|\mathrm{Var}(\mathbf{Y}_{M,L})\| = \|\mathbf{\Gamma}_{M,L}\mathbf{\Gamma}^{\mathrm{T}}_{M,L}\| \leqslant \mathcal{M}(\boldsymbol{f}^X_{M,L}, \mathbf{\Phi}_1)$, where $\mathcal{M}(\boldsymbol{f}^X_{M,L}, \mathbf{\Phi}_1) = 2\pi \cdot \mathrm{ess\,sup}_{\theta\in[-\pi,\pi]}\langle \mathbf{\Phi}_1, \boldsymbol{f}^X_{M,L,\theta}(\mathbf{\Phi}_1)\rangle$ and $\boldsymbol{f}^X_{M,L,\theta}(\cdot)$ is the spectral density matrix operator of process $\{\mathbf{X}_{M,L,t}(\cdot)\}_{t\in\mathbb{Z}}$.

Define $\mathbf{Y}_L = \mathbf{Y}_{\infty,L} = (\langle \mathbf{\Phi}_1, \mathbf{X}_{L,1}\rangle, \ldots, \langle \mathbf{\Phi}_1, \mathbf{X}_{L,n}\rangle)^{\mathrm{T}}$. By Lemma 1.7, (1.45) in Lemma 1.5 and $\mathrm{rank}(\mathbf{\Gamma}^{\mathrm{T}}_{\infty,L}\mathbf{\Gamma}_{\infty,L}) = n$, we obtain

$$\begin{aligned} &P\{|\langle \mathbf{\Phi}_1, (\widehat{\mathbf{\Sigma}}^X_{L,0} - \mathbf{\Sigma}^X_{L,0})(\mathbf{\Phi}_1)| > \mathcal{M}(\boldsymbol{f}^X_L, \mathbf{\Phi}_1)\eta\} \\ =&P\{|\mathbf{Y}^{\mathrm{T}}_L\mathbf{Y}_L - \mathbb{E}\mathbf{Y}^{\mathrm{T}}_L\mathbf{Y}_L| > n\mathcal{M}(\boldsymbol{f}^X_L, \mathbf{\Phi}_1)\eta\} \leqslant 2\exp\left\{-cn\min\left(\eta^2, \eta\right)\right\}, \end{aligned}$$

where $\mathcal{M}(f^X_L, \mathbf{\Phi}_1) = 2\pi \cdot \mathrm{ess\,sup}_{\theta\in[-\pi,\pi]}\langle \mathbf{\Phi}_1, \boldsymbol{f}^X_{L,\theta}(\mathbf{\Phi}_1)\rangle$ and $\boldsymbol{f}^X_{L,\theta}(\cdot)$ is the spectral density matrix operator of $\{\mathbf{X}_{L,t}(\cdot)\}_{t\in\mathbb{Z}}$.

Next, we need to show that this result still holds as $L \to \infty$. Lemmas 1.9 and 1.10 imply that $\lim_{L\to\infty}\mathbb{E}\left\{\left|\langle \mathbf{\Phi}_1, (\widehat{\mathbf{\Sigma}}^X_{L,0} - \widehat{\mathbf{\Sigma}}^X_0)(\mathbf{\Phi}_1)\rangle\right|\right\} = 0$, $\lim_{L\to\infty}\langle \mathbf{\Phi}_1, \mathbf{\Sigma}^X_{L,0}(\mathbf{\Phi}_1)\rangle = \langle \mathbf{\Phi}_1, \mathbf{\Sigma}^X_0(\mathbf{\Phi}_1)\rangle$ and $\lim_{L\to\infty}\mathcal{M}(f^X_L, \mathbf{\Phi}_1) = \mathcal{M}(\boldsymbol{f}^X, \mathbf{\Phi}_1)$. Combining the above results and following the similar argument in the proof of Lemma 1.5, we obtain

$$\begin{aligned} P\left\{\left|\langle \mathbf{\Phi}_1, (\widehat{\mathbf{\Sigma}}^X_0 - \mathbf{\Sigma}^X_0)(\mathbf{\Phi}_1)\rangle\right| > \mathcal{M}(\boldsymbol{f}^X, \mathbf{\Phi}_1)\eta\right\} \\ \leqslant 2\exp\left\{-cn\min\left(\eta^2, \eta\right)\right\}. \end{aligned} \tag{1.38}$$

Provided that $\mathcal{M}(\boldsymbol{f}^X, \boldsymbol{\Phi}_1) \leqslant \mathcal{M}_{k_1}^X \langle \boldsymbol{\Phi}_1, \boldsymbol{\Sigma}_0^X(\boldsymbol{\Phi}_1)\rangle$, we obtain

$$P\left\{\left|\frac{\langle \boldsymbol{\Phi}_1, (\widehat{\boldsymbol{\Sigma}}_0^X - \boldsymbol{\Sigma}_0^X)(\boldsymbol{\Phi}_1)\rangle}{\langle \boldsymbol{\Phi}_1, \boldsymbol{\Sigma}_0^X(\boldsymbol{\Phi}_1)\rangle}\right| > \mathcal{M}_{k_1}^X \eta\right\} \leqslant 2\exp\left\{-cn\min\left(\eta^2, \eta\right)\right\},$$

which completes the proof of (1.14). Part (ii): For fixed vectors $\boldsymbol{\Phi}_1 \in \mathbb{H}^p$ and $\boldsymbol{\Phi}_2 \in \mathbb{H}^d$, we denote $\mathcal{M}(\boldsymbol{f}^{X,Y}, \boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2) = 2\pi \cdot \operatorname{ess\,sup}_{\theta\in[-\pi,\pi]} |\langle \boldsymbol{\Phi}_1, \boldsymbol{f}_\theta^{X,Y}(\boldsymbol{\Phi}_2)\rangle|$. Define $\mathbf{M}_t(\cdot) = [(\mathbf{X}_t(\cdot))^{\mathrm{T}}, (\mathbf{Y}_t(\cdot))^{\mathrm{T}}]^{\mathrm{T}}$. Letting $\boldsymbol{\Phi} = (\boldsymbol{\Phi}_1^{\mathrm{T}}, \boldsymbol{\Phi}_2^{\mathrm{T}})^{\mathrm{T}}$, we have

$$\begin{aligned}\langle \boldsymbol{\Phi}_1, (\widehat{\boldsymbol{\Sigma}}_0^{X,Y} - \boldsymbol{\Sigma}_0^{X,Y})(\boldsymbol{\Phi}_2)\rangle =& \frac{1}{2}[\langle \boldsymbol{\Phi}, (\widehat{\boldsymbol{\Sigma}}_0^M - \boldsymbol{\Sigma}_0^M)(\boldsymbol{\Phi})\rangle - \langle \boldsymbol{\Phi}_1, (\widehat{\boldsymbol{\Sigma}}_0^X - \boldsymbol{\Sigma}_0^X)(\boldsymbol{\Phi}_1)\rangle \\ & - \langle \boldsymbol{\Phi}_2, (\widehat{\boldsymbol{\Sigma}}_0^Y - \boldsymbol{\Sigma}_0^Y)(\boldsymbol{\Phi}_2)\rangle].\end{aligned}$$

Applying (1.38) on $\{\mathbf{X}_t(\cdot)\}$ and $\{\mathbf{Y}_t(\cdot)\}$, we obtain that

$$\begin{aligned}P\left\{\left|\langle \boldsymbol{\Phi}_1, (\widehat{\boldsymbol{\Sigma}}_0^X - \boldsymbol{\Sigma}_0^X)(\boldsymbol{\Phi}_1)\rangle\right| > \mathcal{M}(\boldsymbol{f}^X, \boldsymbol{\Phi}_1)\eta\right\} &\leqslant 2\exp\{-cn\min(\eta^2, \eta)\}, \\ P\left\{\left|\langle \boldsymbol{\Phi}_2, (\widehat{\boldsymbol{\Sigma}}_0^Y - \boldsymbol{\Sigma}_0^Y)(\boldsymbol{\Phi}_2)\rangle\right| > \mathcal{M}(\boldsymbol{f}^Y, \boldsymbol{\Phi}_2)\eta\right\} &\leqslant 2\exp\{-cn\min(\eta^2, \eta)\}.\end{aligned}$$

For $\{\mathbf{M}_t(\cdot)\}$, $\mathcal{M}(f^M, \boldsymbol{\Phi}) \leqslant \mathcal{M}(\boldsymbol{f}^X, \boldsymbol{\Phi}_1) + \mathcal{M}(\boldsymbol{f}^Y, \boldsymbol{\Phi}_2) + 2\mathcal{M}(\boldsymbol{f}^{X,Y}, \boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2)$. This, together with (1.38) implies that

$$\begin{aligned}P\left\{\left|\langle \boldsymbol{\Phi}, (\widehat{\boldsymbol{\Sigma}}_0^M - \boldsymbol{\Sigma}_0^M)(\boldsymbol{\Phi})\rangle\right| > \{\mathcal{M}(\boldsymbol{f}^X, \boldsymbol{\Phi}_1) + \mathcal{M}(\boldsymbol{f}^Y, \boldsymbol{\Phi}_2) + 2\mathcal{M}(\boldsymbol{f}^{X,Y}, \boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2)\}\eta\right\} \\ \leqslant 2\exp\{-cn\min(\eta^2, \eta)\}.\end{aligned}$$

Combining the above results, we obtain

$$\begin{aligned}P\left\{\left|\langle \boldsymbol{\Phi}_1, (\widehat{\boldsymbol{\Sigma}}_0^{X,Y} - \boldsymbol{\Sigma}_0^{X,Y})(\boldsymbol{\Phi}_2)\rangle\right| > \{\mathcal{M}(\boldsymbol{f}^X, \boldsymbol{\Phi}_1) + \mathcal{M}(\boldsymbol{f}^Y, \boldsymbol{\Phi}_2) + \mathcal{M}(\boldsymbol{f}^{X,Y}, \boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2)\}\eta\right\} \\ \leqslant 6\exp\{-cn\min(\eta^2, \eta)\}.\end{aligned} \tag{1.39}$$

For $h > 0$, let $\mathbf{U}_{1,t} = \mathbf{X}_t + \mathbf{X}_{t+h}$, $\mathbf{U}_{2,t} = \mathbf{X}_t - \mathbf{X}_{t+h}$, $\mathbf{V}_{1,t} = \mathbf{Y}_t + \mathbf{Y}_{t+h}$ and $\mathbf{V}_{2,t} = \mathbf{Y}_t - \mathbf{Y}_{t+h}$. Accordingly, we have that

$$\begin{aligned}\langle \boldsymbol{\Phi}_1, \boldsymbol{\Sigma}_l^{U_1,V_1}(\boldsymbol{\Phi}_2)\rangle &= 2\langle \boldsymbol{\Phi}_1, \boldsymbol{\Sigma}_l^{X,Y}(\boldsymbol{\Phi}_2)\rangle + \langle \boldsymbol{\Phi}_1, \boldsymbol{\Sigma}_{l-h}^{X,Y}(\boldsymbol{\Phi}_2)\rangle + \langle \boldsymbol{\Phi}_1, \boldsymbol{\Sigma}_{l+h}^{X,Y}(\boldsymbol{\Phi}_2)\rangle, \\ \langle \boldsymbol{\Phi}_1, \boldsymbol{\Sigma}_l^{U_2,V_2}(\boldsymbol{\Phi}_2)\rangle &= 2\langle \boldsymbol{\Phi}_1, \boldsymbol{\Sigma}_l^{X,Y}(\boldsymbol{\Phi}_2)\rangle - \langle \boldsymbol{\Phi}_1, \boldsymbol{\Sigma}_{l-h}^{X,Y}(\boldsymbol{\Phi}_2)\rangle - \langle \boldsymbol{\Phi}_1, \boldsymbol{\Sigma}_{l+h}^{X,Y}(\boldsymbol{\Phi}_2)\rangle,\end{aligned}$$

and

$$\boldsymbol{f}_\theta^{U_1,V_1} = (2 + \exp(-ih\theta) + \exp(ih\theta))\boldsymbol{f}_\theta^{X,Y},$$
$$\boldsymbol{f}_\theta^{U_2,V_2} = (2 - \exp(-ih\theta) - \exp(ih\theta))\boldsymbol{f}_\theta^{X,Y}.$$

Combining these with the definition of $\mathcal{M}(\boldsymbol{f}^{X,Y}, \boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2)$ yields

$$\begin{aligned}
&4\langle \boldsymbol{\Phi}_1, (\widehat{\boldsymbol{\Sigma}}_h^{X,Y} - \boldsymbol{\Sigma}_h^{X,Y})(\boldsymbol{\Phi}_2)\rangle \\
=&\langle \boldsymbol{\Phi}_1, (\widehat{\boldsymbol{\Sigma}}_0^{U_1,V_1} - \boldsymbol{\Sigma}_0^{U_1,V_1})(\boldsymbol{\Phi}_2)\rangle - \langle \boldsymbol{\Phi}_1, (\widehat{\boldsymbol{\Sigma}}_0^{U_2,V_2} - \boldsymbol{\Sigma}_0^{U_2,V_2})(\boldsymbol{\Phi}_2)\rangle,
\end{aligned}$$

and

$$\mathcal{M}(\boldsymbol{f}^{U_1,V_1}, \boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2) \leqslant 4\mathcal{M}(\boldsymbol{f}^{X,Y}, \boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2).$$

By similar arguments, we obtain $\mathcal{M}(\boldsymbol{f}^{U_i}, \boldsymbol{\Phi}_1) \leqslant 4\mathcal{M}(\boldsymbol{f}^{X}, \boldsymbol{\Phi}_1)$ and $\mathcal{M}(\boldsymbol{f}^{V_i}, \boldsymbol{\Phi}_2) \leqslant 4\mathcal{M}(\boldsymbol{f}^{Y}, \boldsymbol{\Phi}_2)$, for $i = 1, 2$. Then it follows from (1.39) that

$$\begin{aligned}
&P\left\{\left|\langle \boldsymbol{\Phi}_1, (\widehat{\boldsymbol{\Sigma}}_h^{X,Y} - \boldsymbol{\Sigma}_h^{X,Y})(\boldsymbol{\Phi}_2)\rangle\right| > 2\{\mathcal{M}(\boldsymbol{f}^{X}, \boldsymbol{\Phi}_1) + \mathcal{M}(\boldsymbol{f}^{Y}, \boldsymbol{\Phi}_2) + \mathcal{M}(\boldsymbol{f}^{X,Y}, \boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2)\}\eta\right\} \\
\leqslant& \sum_{i=1}^{2} P\left\{\left|\langle \boldsymbol{\Phi}_1, (\widehat{\boldsymbol{\Sigma}}_0^{U_i,V_i} - \boldsymbol{\Sigma}_0^{U_i,V_i})(\boldsymbol{\Phi}_2)\rangle\right| > \{\mathcal{M}(\boldsymbol{f}^{U_i}, \boldsymbol{\Phi}_1) + \mathcal{M}(\boldsymbol{f}^{V_i}, \boldsymbol{\Phi}_2) + \mathcal{M}(\boldsymbol{f}^{U_i,V_i}, \boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2)\}\eta\right\} \\
\leqslant& 12\exp\{-cn\min(\eta^2, \eta)\}.
\end{aligned}$$

Provided that $\mathcal{M}(\boldsymbol{f}^{X,Y}, \boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2) \leqslant \mathcal{M}_{k_1,k_2}^{X,Y}(\langle \boldsymbol{\Phi}_1, \boldsymbol{\Sigma}_0^X(\boldsymbol{\Phi}_1)\rangle + \langle \boldsymbol{\Phi}_2, \boldsymbol{\Sigma}_0^Y(\boldsymbol{\Phi}_2)\rangle)$ and $\mathcal{M}(\boldsymbol{f}^{X}, \boldsymbol{\Phi}_1) \leqslant \mathcal{M}_{k_1}^{X}\langle \boldsymbol{\Phi}_1, \boldsymbol{\Sigma}_0^X(\boldsymbol{\Phi}_1)\rangle$, we obtain

$$\begin{aligned}
P\left\{\left|\frac{\langle \boldsymbol{\Phi}_1, (\widehat{\boldsymbol{\Sigma}}_0^{X,Y} - \boldsymbol{\Sigma}_0^{X,Y})(\boldsymbol{\Phi}_2)\rangle}{\langle \boldsymbol{\Phi}_1, \boldsymbol{\Sigma}_0^X(\boldsymbol{\Phi}_1)\rangle + \langle \boldsymbol{\Phi}_2, \boldsymbol{\Sigma}_0^Y(\boldsymbol{\Phi}_2)\rangle}\right| > \left(\mathcal{M}_{k_1}^X + \mathcal{M}_{k_2}^Y + \mathcal{M}_{k_1,k_2}^{X,Y}\right)\eta\right\} \\
\leqslant 6\exp\left\{-cn\min(\eta^2, \eta)\right\},
\end{aligned}$$

$$\begin{aligned}
P\left\{\left|\frac{\langle \boldsymbol{\Phi}_1, (\widehat{\boldsymbol{\Sigma}}_h^{X,Y} - \boldsymbol{\Sigma}_h^{X,Y})(\boldsymbol{\Phi}_2)\rangle}{\langle \boldsymbol{\Phi}_1, \boldsymbol{\Sigma}_0^X(\boldsymbol{\Phi}_1)\rangle + \langle \boldsymbol{\Phi}_2, \boldsymbol{\Sigma}_0^Y(\boldsymbol{\Phi}_2)\rangle}\right| > 2\left(\mathcal{M}_{k_1}^X + \mathcal{M}_{k_2}^Y + \mathcal{M}_{k_1,k_2}^{X,Y}\right)\eta\right\} \\
\leqslant 12\exp\{-cn\min(\eta^2, \eta)\}.
\end{aligned}$$

Letting $c_2 = c/4$, we complete the proof of (1.15). □

Proof of Theorem 1.2 Under FPCA framework, for each $k = 1, \ldots, d$, we have $Y_{tk}(\cdot) = \sum_{m=1}^{\infty} \xi_{tkm}\phi_{km}(\cdot)$ with eigenpairs $(\omega_{km}^Y, \phi_{km})$, and for each $j = 1, \ldots, p$, we have $X_{tj}(\cdot) = \sum_{l=1}^{\infty} \zeta_{tjl}\psi_{jl}(\cdot)$ with eigenpairs $(\omega_{jl}^X, \psi_{jl})$. Denote $\mathcal{M}_{X,Y} = \mathcal{M}_1^X + \mathcal{M}_1^Y + \mathcal{M}_{1,1}^{X,Y}$. Let $\mathbf{\Phi}_1 = (0, \ldots, 0, \{\omega_{jl}^X\}^{-\frac{1}{2}}\psi_{jl}, 0, \ldots, 0)^{\mathrm{T}}$ and $\mathbf{\Phi}_2 = (0, \ldots, 0, \{\omega_{km}^Y\}^{-\frac{1}{2}}\phi_{km}, 0, \ldots, 0)^{\mathrm{T}}$. Following the similar argument in the proof of Theorem 2 in Guo and Qiao (2022) with $2\sqrt{\omega_0^X\omega_0^Y} \leqslant \omega_0^X + \omega_0^Y$ and Theorem 1.1, we can prove

$$P\left\{\|\widehat{\Sigma}_{h,jk}^{X,Y} - \Sigma_{h,jk}^{X,Y}\|_{\mathcal{S}} > (\omega_0^X + \omega_0^Y)\mathcal{M}_{X,Y}\eta\right\} \leqslant c_1 \exp\{-c_3 n \min(\eta^2, \eta)\}.$$

By the definition of $\|\widehat{\mathbf{\Sigma}}_h^{X,Y} - \mathbf{\Sigma}_h^{X,Y}\|_{\max} = \max_{1\leqslant j\leqslant p, 1\leqslant k\leqslant d} \|\widehat{\Sigma}_{h,jk}^{X,Y} - \Sigma_{h,jk}^{X,Y}\|_{\mathcal{S}}$, we have that

$$P\left\{\|\widehat{\mathbf{\Sigma}}_h^{X,Y} - \mathbf{\Sigma}_h^{X,Y}\|_{\max} > (\omega_0^X + \omega_0^Y)\mathcal{M}_{X,Y}\eta\right\} \leqslant c_1 pd \exp\{-c_3 n \min(\eta^2, \eta)\}.$$

Let $\eta = \rho\sqrt{\log(pd)/n} \leqslant 1$ and $\rho^2 c_3 > 1$, which can be achieved for sufficiently large n. We obtain that

$$P\left\{\|\widehat{\mathbf{\Sigma}}_h^{X,Y} - \mathbf{\Sigma}_h^{X,Y}\|_{\max} > (\omega_0^X + \omega_0^Y)\mathcal{M}_{X,Y}\rho\sqrt{\frac{\log(pd)}{n}}\right\} \leqslant c_1 (pd)^{1-c\rho^2},$$

which implies (1.17). □

Before presenting the proof of Theorem 1.3, we provide some useful inequalities for estimated eigenpairs under the FPCA framework. For $\{\mathbf{X}_t(\cdot)\}_{t\in\mathbb{Z}}$, let $\delta_{jl}^X = \min_{1\leqslant l'\leqslant l}\{\omega_{jl'}^X - \omega_{j(l'+1)}^X\}$ and $\widehat{\Delta}_{jl}^X = \widehat{\Sigma}_{0,jl}^X - \Sigma_{0,jl}^X$ for $j = 1, \ldots, p$ and $l = 1, 2, \ldots$. It follows from (4.43) and Lemma 4.3 of Bosq (2000) that

$$\sup_{l\geqslant 1} |\widehat{\omega}_{jl}^X - \omega_{jl}^X| \leqslant \|\widehat{\Delta}_{jj}^X\|_{\mathcal{S}} \quad \text{and} \quad \sup_{l\geqslant 1} \delta_{jl}^X \|\widehat{\psi}_{jl} - \psi_{jl}\| \leqslant 2\sqrt{2}\|\widehat{\Delta}_{jj}^X\|_{\mathcal{S}}. \tag{1.40}$$

Similarly, for process $\{\mathbf{Y}_t(\cdot)\}_{t\in\mathbb{Z}}$, let $\delta_{km}^Y = \min_{1\leqslant m'\leqslant m}\{\omega_{km'}^Y - \omega_{k(m'+1)}^Y\}$ and $\widehat{\Delta}_{km}^Y = \widehat{\Sigma}_{0,km}^Y - \Sigma_{0,km}^Y$ for $k = 1, \ldots, d$ and $m = 1, 2, \ldots$, we have

$$\sup_{m\geqslant 1} |\widehat{\omega}_{km}^Y - \omega_{km}^Y| \leqslant \|\widehat{\Delta}_{kk}^Y\|_{\mathcal{S}} \quad \text{and} \quad \sup_{m\geqslant 1} \delta_{km}^Y \|\widehat{\phi}_{km} - \phi_{km}\| \leqslant 2\sqrt{2}\|\widehat{\Delta}_{kk}^Y\|_{\mathcal{S}}. \tag{1.41}$$

Proof of Theorem 1.3 Recall $\widehat{\sigma}_{h,jklm}^{X,Y} = \frac{1}{n-h}\sum_{t=1}^{n-h} \widehat{\zeta}_{tjl}\widehat{\xi}_{(t+h)km}$ and $\sigma_{h,jklm}^{X,Y} = \mathrm{Cov}(\zeta_{tjl}, \xi_{(t+h)km}) = \langle\psi_{jl}, \langle\Sigma_{h,jk}^{X,Y}, \phi_{km}\rangle\rangle$. Let $\widehat{r}_{jl} = \widehat{\psi}_{jl} - \psi_{jl}$, $\widehat{w}_{km} = \widehat{\phi}_{km} - \phi_{km}$ and

$\widehat{\Delta}_{h,jk}^{X,Y} = \widehat{\Sigma}_{h,jk}^{X,Y} - \Sigma_{h,jk}^{X,Y}$, then

$$
\begin{aligned}
&\widehat{\sigma}_{h,jklm}^{X,Y} - \sigma_{h,jklm}^{X,Y} \\
=&\langle \widehat{r}_{jl}, \langle \widehat{\Sigma}_{h,jk}^{X,Y}, \widehat{w}_{km} \rangle \rangle + \left(\langle \widehat{r}_{jl}, \langle \widehat{\Delta}_{h,jk}^{X,Y}, \phi_{km} \rangle \rangle + \langle \psi_{jl}, \langle \widehat{\Delta}_{h,jk}^{X,Y}, \widehat{w}_{km} \rangle \rangle \right) \\
&+ \left(\langle \widehat{r}_{jl}, \langle \Sigma_{h,jk}^{X,Y}, \phi_{km} \rangle \rangle + \langle \psi_{jl}, \langle \Sigma_{h,jk}^{X,Y}, \widehat{w}_{km} \rangle \rangle \right) + \langle \psi_{jl}, \langle \widehat{\Delta}_{h,jk}^{X,Y}, \phi_{km} \rangle \rangle \\
=&I_1 + I_2 + I_3 + I_4.
\end{aligned}
$$

Let $\Omega_{jk,\eta}^{X,Y} = \left\{ \|\widehat{\Delta}_{h,jk}^{X,Y}\|_{\mathcal{S}} \leqslant (\omega_0^X + \omega_0^Y)\mathcal{M}_{X,Y}\eta \right\}$, $\Omega_{jj,\eta}^{X} = \left\{ \|\widehat{\Delta}_{jj}^{X}\|_{\mathcal{S}} \leqslant 2\mathcal{M}_1^X \omega_0^X \eta \right\}$, $\Omega_{kk,\eta}^{Y} = \left\{ \|\widehat{\Delta}_{kk}^{Y}\|_{\mathcal{S}} \leqslant 2\mathcal{M}_1^Y \omega_0^Y \eta \right\}$ and $\Omega_1 = \left\{ \|\widehat{\Delta}_{h,jk}^{X,Y}\|_{\mathcal{S}} \leqslant (\omega_0^X + \omega_0^Y) \right\}$. By Theorem 1.2 and Lemma 1.24, we have

$$
\begin{aligned}
P\left((\Omega_{jk,\eta}^{X,Y})^C \right) &\leqslant c_1 \exp\{-c_3 n \min(\eta^2, \eta)\}, \\
P\left((\Omega_{jj,\eta}^{X})^C \right) &\leqslant 4 \exp\{-\tilde{c}_1 n \min(\eta^2, \eta)\}, \\
P\left((\Omega_{kk,\eta}^{Y})^C \right) &\leqslant 4 \exp\{-\tilde{c}_1 n \min(\eta^2, \eta)\}, \\
P\left((\Omega_1)^C \right) &\leqslant c_1 \exp\{-c_3 n (\mathcal{M}_{X,Y})^{-2}\}.
\end{aligned}
$$

On the event of $\Omega_1 \cap \Omega_{\eta,jk}^{X,Y} \cap \Omega_{jj,\eta}^{X} \cap \Omega_{kk,\eta}^{Y}$, by Condition 1.5, (1.40), (1.41), Lemma 1.2 and the fact that $(\omega_0^X \omega_0^Y)^{1/2} \leqslant 1/2(\omega_0^X + \omega_0^Y)$, we obtain that

$$
\begin{aligned}
\left| \frac{I_1}{\sqrt{\omega_{jl}^X \omega_{km}^Y}} \right| &\leqslant c_0^{-1} (\alpha_1 \alpha_2)^{1/2} l^{\alpha_1/2} m^{\alpha_2/2} \|\widehat{r}_{jl}\| (\|\widehat{\Delta}_{h,jk}^{X,Y}\|_{\mathcal{S}} + \|\Sigma_{h,jk}^{X,Y}\|_{\mathcal{S}}) \|\widehat{w}_{km}\| \\
&\lesssim l^{3\alpha_1/2+1} m^{3\alpha_2/2+1} \|\widehat{\Delta}_{jj}^{X}\|_{\mathcal{S}} \|\widehat{\Delta}_{kk}^{Y}\|_{\mathcal{S}} (\|\widehat{\Delta}_{h,jk}^{X,Y}\|_{\mathcal{S}} + (\omega_0^X \omega_0^Y)^{1/2}) \\
&\lesssim (l^{3\alpha_1+2} \vee m^{3\alpha_2+2}) \mathcal{M}_1^X \mathcal{M}_1^Y \eta^2, \\
&\lesssim (l^{3\alpha_1+2} \vee m^{3\alpha_2+2}) (\mathcal{M}_1^X + \mathcal{M}_1^Y)^2 \eta^2,
\end{aligned}
$$

$$
\begin{aligned}
\left| \frac{I_2}{\sqrt{\omega_{jl}^X \omega_{km}^Y}} \right| &\leqslant c_0^{-1} (\alpha_1 \alpha_2)^{1/2} l^{\alpha_1/2} m^{\alpha_2/2} \|\widehat{\Delta}_{h,jk}^{X,Y}\|_{\mathcal{S}} (\|\widehat{r}_{jl}\| + \|\widehat{w}_{km}\|) \\
&\lesssim l^{\alpha_1/2} m^{\alpha_2/2} \|\widehat{\Delta}_{h,jk}^{X,Y}\|_{\mathcal{S}} (l^{\alpha_1+1} \|\widehat{\Delta}_{jj}^{X}\|_{\mathcal{S}} + m^{\alpha_2+1} \|\widehat{\Delta}_{kk}^{Y}\|_{\mathcal{S}}) \\
&\lesssim (l^{2\alpha_1+1} \vee m^{2\alpha_2+1}) \mathcal{M}_{X,Y} (\mathcal{M}_X \vee \mathcal{M}_Y) \eta^2, \\
&\lesssim (l^{2\alpha_1+1} \vee m^{2\alpha_2+1}) \mathcal{M}_{X,Y}^2 \eta^2,
\end{aligned}
$$

By Theorem 1.1,

$$P\left\{\left|\frac{I_4}{\sqrt{\omega_{jl}^X\omega_{km}^Y}}\right| \geqslant 2\mathcal{M}_{XY}\eta\right\} \leqslant c_1\exp\{-c_2 n\min(\eta^2,\eta)\}.$$

Next, we consider the term $I_3 = \langle\widehat{r}_{jl},\langle\Sigma_{h,jk}^{X,Y},\phi_{km}\rangle\rangle + \langle\psi_{jl},\langle\Sigma_{h,jk}^{X,Y},\widehat{w}_{km}\rangle\rangle$. By Condition 1.5, Lemmas 1.14 and 1.26 for $\{\mathbf{X}_t\}_{t\in\mathbb{Z}}$ and $\{\mathbf{Y}_t\}_{t\in\mathbb{Z}}$, we obtain that

$$\begin{aligned}
&\left|\frac{I_3}{\sqrt{\omega_{jl}^X\omega_{km}^Y}}\right| \\
\lesssim&\mathcal{M}_1^X l^{\alpha_1+1}\eta + (M_1^X)^2 l^{(5\alpha_1+4)/2}\eta^2 + \mathcal{M}_1^Y m^{\alpha_2+1}\eta + (\mathcal{M}_1^Y)^2 m^{(5\alpha_2+4)/2}\eta^2 \\
\lesssim&(l^{\alpha_1+1}\vee m^{\alpha_2+1})(\mathcal{M}_1^X+\mathcal{M}_1^Y)\eta + (l^{(5\alpha_1+4)/2}\vee m^{(5\alpha_2+4)/2})(\mathcal{M}_1^X+\mathcal{M}_1^Y)^2\eta^2
\end{aligned}$$

holds with probability greater than $1-16\exp\{-\tilde{c}_4 n\min(\eta^2,\eta)\} - 8\exp\{-\tilde{c}_4 n(\{\mathcal{M}_1^X\}^2 l^{2(\alpha_1+1)}\vee\{\mathcal{M}_1^Y\}^2 m^{2(\alpha_2+1)})^{-1}\}$.

Combining the above results, we obtain that there exists positive constants ρ_1, ρ_2, $\tilde{c}_7$ and $\tilde{c}_8$ such that

$$\begin{aligned}
&P\left\{\left|\frac{\widehat{\sigma}_{h,jklm}^{X,Y}-\sigma_{h,jklm}^{X,Y}}{\sqrt{\omega_{jl}^X\omega_{km}^Y}}\right| \geqslant \rho_1\mathcal{M}_{X,Y}(l^{\alpha_1+1}\vee m^{\alpha_2+1})\eta + \rho_2\mathcal{M}_{X,Y}^2(l^{3\alpha_1+2}\vee m^{3\alpha_2+2})\eta^2\right\} \\
&\leqslant \tilde{c}_8\exp\{-\tilde{c}_7 n\min(\eta^2,\eta)\} + \tilde{c}_8\exp\{-\tilde{c}_7\mathcal{M}_{X,Y}^{-2}n(l^{2(\alpha_1+1)}\vee m^{2(\alpha_2+1)})^{-1}\},
\end{aligned}$$

where $\mathcal{M}_{X,Y} = \mathcal{M}_1^X + \mathcal{M}_1^Y + \mathcal{M}_{1,1}^{X,Y}$. Applying the Boole's inequality, we obtain that

$$\begin{aligned}
&P\left\{\max_{\substack{1\leqslant j\leqslant p\\1\leqslant k\leqslant d\\1\leqslant l\leqslant M_1\\1\leqslant m\leqslant M_2}}\left|\frac{\widehat{\sigma}_{h,jklm}^{X,Y}-\sigma_{h,jklm}^{X,Y}}{\sqrt{\omega_{jl}^X\omega_{km}^Y}}\right| \geqslant \rho_1\mathcal{M}_{X,Y}(l^{\alpha_1+1}\vee m^{\alpha_2+1})\eta + \rho_2\mathcal{M}_{X,Y}^2(l^{3\alpha_1+2}\vee m^{3\alpha_2+2})\eta^2\right\} \\
&\leqslant pdM_1M_2\{\tilde{c}_8\exp\{-\tilde{c}_7 n\min(\eta^2,\eta)\} + \tilde{c}_8\exp\{-\tilde{c}_7\mathcal{M}_{X,Y}^{-2}n(l^{2(\alpha_1+1)}\vee m^{2(\alpha_2+1)^{-1}})\}.
\end{aligned}$$

Letting $\eta = \rho_3\sqrt{\frac{\log(pdM_1M_2)}{n}} < 1$ and $\rho_1+\rho_2\rho_3\mathcal{M}_{X,Y}(M_1^{2\alpha_1+1}\vee M_2^{2\alpha_2+1})\eta \leqslant \rho_4$, there

exist some constants $c_5, c_6 > 0$ such that

$$P\left\{\max_{\substack{1\leqslant j\leqslant p,1\leqslant k\leqslant d\\ 1\leqslant l\leqslant M_1,1\leqslant m\leqslant M_2}}\left|\frac{\widehat{\sigma}^{X,Y}_{h,jklm}-\sigma^{X,Y}_{h,jklm}}{\sqrt{\omega^X_{jl}\omega^Y_{km}}}\right|\geqslant \rho_3\rho_4\mathcal{M}_{X,Y}(M_1^{\alpha_1+1}\vee M_2^{\alpha_2+1})\sqrt{\frac{\log(pdM_1M_2)}{n}}\right\}$$
$$\leqslant c_5(pdM_1M_2)^{c_6}.$$

□

1.B.2 Proofs of propositions

Proof of Proposition 1.1 Under a mixed-process scenario consisting of $\{\mathbf{X}_t(\cdot)\}$ and d-dimensional time series $\{\mathbf{Z}_t\}$, we obtain the concentration bound on $\widehat{\mathbf{\Sigma}}^{X,Z}_h$,

$$P\left\{\left|\frac{\langle\mathbf{\Phi}_1,(\widehat{\mathbf{\Sigma}}^{X,Z}_h-\mathbf{\Sigma}^{X,Z}_h)\boldsymbol{\nu}\rangle}{\langle\mathbf{\Phi}_1,\mathbf{\Sigma}^X_0(\mathbf{\Phi}_1)\rangle+\boldsymbol{\nu}^{\mathrm{T}}\mathbf{\Sigma}^Z_0\boldsymbol{\nu}}\right|>\left(\mathcal{M}^X_{k_1}+\mathcal{M}^Z_{k_2}+\mathcal{M}^{X,Z}_{k_1,k_2}\right)\eta\right\} \tag{1.42}$$
$$\leqslant c_1\exp\{-c_2n\min(\eta^2,\eta)\}.$$

Provided with Lemma 1.28, the above result can be proved in similar way to (1.15) in Theorem 1.1, hence we omit it here.

Denote $\sigma^Z_{0,kk}=\sqrt{\mathrm{Var}(Z_k)}$, $(\sigma^Z_0)^2=\max_{1\leqslant k\leqslant d}\mathrm{Var}(Z_k)<\infty$ and $\mathcal{M}_{X,Z}=\mathcal{M}^X_1+\mathcal{M}^Z_1+\mathcal{M}^{X,Z}_{1,1}$. Letting $\mathbf{\Phi}_1=(0,\ldots,0,\{\omega^X_{jl}\}^{-\frac{1}{2}}\psi_{jl},0,\ldots,0)^{\mathrm{T}}$ and $\boldsymbol{\nu}=(0,\ldots,0,\{\sigma^Z_{0,kk}\}^{-1},0,\ldots,0)^{\mathrm{T}}$, we obtain that $\Delta_{h,jkl}=\langle\mathbf{\Phi}_1,(\widehat{\mathbf{\Sigma}}^{X,Z}_h-\mathbf{\Sigma}^{X,Z}_h)\boldsymbol{\nu}\rangle=(\omega^X_{jl})^{-1/2}(\sigma^Z_{0,kk})^{-1}\langle\psi_{jl},\widehat{\Sigma}^{X,Z}_{h,jk}-\Sigma^{X,Z}_{h,jk}\rangle$ and $\langle\mathbf{\Phi}_1,\mathbf{\Sigma}^X_0(\mathbf{\Phi}_1)\rangle=\boldsymbol{\nu}^{\mathrm{T}}\mathbf{\Sigma}^Z_0\boldsymbol{\nu}=1$. Then $\|\Sigma^{X,Z}_{h,jk}-\Sigma^{X,Z}_{h,jk}\|^2=\sum_{l=1}^{\infty}\omega^X_{jl}(\sigma^Z_{0,kk})^2\Delta^2_{h,jkl}$. By Jensen's inequality, we have that

$$\mathbb{E}\left\{\|\widehat{\Sigma}^{X,Z}_{h,jk}-\Sigma^{X,Z}_{h,jk}\|^{2q}_{\mathcal{S}}\right\}\leqslant(\sigma^Z_{0,kk})^{2q}\left(\sum_{l=1}^{\infty}\omega^X_{jl}\right)^{q-1}\sum_{l=1}^{\infty}\omega^X_{jl}\mathbb{E}\left|\Delta_{h,jkl}\right|^{2q}$$
$$\leqslant\{\sigma^Z_0\}^{2q}\{\omega^X_0\}^q\sup_l\mathrm{E}\left|\Delta_{h,jkl}\right|^{2q}.$$

By (1.42), we obtain that

$$P\left\{|\Delta_{h,jkl}|>2\mathcal{M}_{X,Z}\eta\right\}\leqslant c_1\exp\{-c_2n\min(\eta^2,\eta)\}.$$

Combining the above results and following the similar argument in the proof of

Theorem 2 in Guo and Qiao (2022) yields

$$P\left\{\|\widehat{\Sigma}_{h,jk}^{X,Z}-\Sigma_{h,jk}^{X,Z}\|>2\mathcal{M}_{X,Z}\sigma_0^Z\sqrt{\omega_0^X}\eta\right\}\leqslant c_1\exp\{-c_3n\min(\eta^2,\eta)\}.$$

Then with the fact that $2\sqrt{(\sigma_0^Z)^2\omega_0^X}\leqslant(\sigma_0^Z)^2+\omega_0^X$, we obtain

$$P\left\{\|\widehat{\Sigma}_{h,jk}^{X,Z}-\Sigma_{h,jk}^{X,Z}\|>((\sigma_0^Z)^2+\omega_0^X)\mathcal{M}_{X,Z}\eta\right\}\leqslant c_1\exp\{-c_3n\min(\eta^2,\eta)\}. \tag{1.43}$$

This also implies (1.18).

Recall that $\widehat{\varrho}_{h,jkl}^{X,Z}=\frac{1}{n-h}\sum_{t=1}^{n-h}\widehat{\zeta}_{tjl}Z_{(t+h)k}$ and $\varrho_{h,jkl}^{X,Z}=\mathrm{Cov}(\zeta_{tjl},Z_{(t+h)k})$. Let $\widehat{r}_{jl}=\widehat{\psi}_{jl}-\psi_{jl}$ and $\widehat{\Delta}_{h,jk}^{X,Z}=\widehat{\Sigma}_{h,jk}^{X,Z}-\Sigma_{h,jk}^{X,Z}$. We have

$$\begin{aligned}\widehat{\varrho}_{h,jkl}^{X,Z}-\varrho_{h,jkl}^{X,Z}&=\langle\widehat{r}_{jl},\widehat{\Delta}_{h,jk}^{X,Z}\rangle+\langle\widehat{r}_{jl},\Sigma_{h,jk}^{X,Z}\rangle+\langle\psi_{jl},\widehat{\Delta}_{h,jk}^{X,Z}\rangle\\&=I_1+I_2+I_3.\end{aligned}$$

Let $\Omega_{jk,\eta}^{X,Z}=\left\{\|\widehat{\Delta}_{h,jk}^{X,Z}\|\leqslant(\omega_0^X+(\sigma_0^Z)^2)\mathcal{M}_{X,Z}\eta\right\}$, $\Omega_{jj,\eta}^{X}=\left\{\|\widehat{\Delta}_{jj}^{X}\|_{\mathcal{S}}\leqslant2\mathcal{M}_1^X\omega_0^X\eta\right\}$ and $\Omega_1=\left\{\|\widehat{\Delta}_{h,jk}^{X,Z}\|\leqslant(\omega_0^X+(\sigma_0^Z)^2)\right\}$. By (1.43) and Lemma 1.24, we have

$$\begin{aligned}P\left((\Omega_{jk,\eta}^{X,Y})^C\right)&\leqslant c_1\exp\{-c_3n\min(\eta^2,\eta)\},\\P\left((\Omega_{jj,\eta}^{X})^C\right)&\leqslant 4\exp\{-\tilde{c}_1n\min(\eta^2,\eta)\},\\P\left((\Omega_1)^C\right)&\leqslant c_1\exp\{-c_3n(\mathcal{M}_{X,Z})^{-2}\}.\end{aligned}$$

On the event of $\Omega_1\cap\Omega_{\eta,jk}^{X,Z}\cap\Omega_{jj,\eta}^X$, by Condition 1.5, (1.40), Lemma 1.2 and $(\sigma_0^Z)^2<\infty$, we obtain that

$$\begin{aligned}\left|\frac{I_1}{\sqrt{\omega_{jl}^X}}\right|&\lesssim l^{\alpha_1/2}\|\widehat{\Delta}_{h,jk}^{X,Z}\|\|\widehat{r}_{jl}\|\lesssim l^{3\alpha_1/2+1}\|\widehat{\Delta}_{h,jk}^{X,Z}\|\|\widehat{\Delta}_{jj}^{X}\|_{\mathcal{S}}\\&\lesssim l^{3\alpha_1/2+1}\mathcal{M}_{X,Z}\mathcal{M}_1^X\eta^2.\end{aligned}$$

By Condition 1.5, Lemma 1.26 and $\|\Sigma_{h,jk}^{XZ}\|\leqslant\omega_0^{1/2}\sigma_{0,kk}^Z$, we obtain that

$$\left|\frac{I_2}{\sqrt{\omega_{jl}^X}}\right|\lesssim\mathcal{M}_1^Xl^{\alpha_1+1}\eta+(M_1^X)^2l^{(5\alpha_1+4)/2}\eta^2$$

holds with probability greater than $1-8\exp\{-\tilde{c}_4 n\min(\eta^2,\eta)\}-4\exp\{-\tilde{c}_4 n$ $(\{\mathcal{M}_1^X\}^{-2}l^{-2(\alpha_1+1)})\}$. By (1.42) and the fact that $\sqrt{(\sigma_0^Z)^2\omega_0^X}\leqslant 1/2\{(\sigma_0^Z)^2+\omega_0^X\}$, we obtain that

$$P\left\{\left|\frac{I_3}{\sqrt{\omega_{jl}^X}}\right|\geqslant 2\mathcal{M}_{X,Z}\sigma_0^Z\eta\right\}\leqslant c_1\exp\{-c_2 n\min(\eta^2,\eta)\}.$$

Combining the above results, we obtain that there exists positive constants ρ_5, ρ_6, $\tilde{c}_9$ and $\tilde{c}_{10}$ such that

$$\begin{aligned}P&\left\{\left|\frac{\widehat{\varrho}_{h,jkl}^{X,Z}-\varrho_{h,jkl}^{X,Z}}{\sqrt{\omega_{jl}^X}}\right|\geqslant\rho_5\mathcal{M}_{X,Z}l^{\alpha_1+1}\eta+\rho_6\mathcal{M}_{X,Z}^2 l^{(5\alpha_1+4)/2}\eta^2\right\}\\&\leqslant\tilde{c}_{10}\exp\{-\tilde{c}_9 n\min(\eta^2,\eta)\}+\tilde{c}_{10}\exp\{-\tilde{c}_9\mathcal{M}_{X,Z}^{-2}nl^{-2(\alpha_1+1)}\}.\end{aligned}$$

Letting $\eta=\rho_7\sqrt{\frac{\log(pdM_1)}{n}}<1$ and $\rho_5+\rho_6\rho_7\mathcal{M}_{X,Z}M_1^{1.5\alpha_1+1}\eta\leqslant\rho_8$, there exist some constants $c_7,c_8>0$ such that

$$P\left\{\max_{\substack{1\leqslant j\leqslant p,1\leqslant k\leqslant d\\1\leqslant l\leqslant M_1}}\left|\frac{\widehat{\varrho}_{h,jkl}^{X,Z}-\varrho_{h,jkl}^{X,Z}}{\sqrt{\omega_{jl}^X}}\right|\geqslant\rho_7\rho_8\mathcal{M}_{X,Z}M_1^{\alpha_1+1}\sqrt{\frac{\log(pdM_1)}{n}}\right\}\leqslant c_7(pdM_1)^{c_8},$$

which implies (1.20). □

Proof of Proposition 1.2 To simplify our notation, we will denote $\widehat{\sigma}_{h,jlm}^{X,\epsilon}$ and $\sigma_{h,jlm}^{X,\epsilon}$ by $\widehat{\sigma}_{h,jlm}$ and $\sigma_{h,jlm}$ in subsequent proofs. Recall that $\widehat{\sigma}_{h,jlm}=\langle\widehat{\psi}_{jl},\langle\widehat{\Sigma}_{h,j}^{X,\epsilon},\widehat{\phi}_m\rangle\rangle$ and $\sigma_{h,jlm}=\langle\psi_{jl},\langle\Sigma_{h,j}^{X,\epsilon},\phi_m\rangle$. Since we assume $\{\mathbf{X}_t(\cdot)\}$ and $\{\epsilon_t(\cdot)\}$ are independent processes, $\sigma_{h,jlm}=0$.

Let $\widehat{r}_{jl}=\widehat{\psi}_{jl}-\psi_{jl}$, $\widehat{w}_m=\widehat{\phi}_m-\phi_m$ and $\widehat{\Delta}_{h,j}^{X,\epsilon}=\widehat{\Sigma}_{h,j}^{X,\epsilon}-\Sigma_{h,j}^{X,\epsilon}$.

$$\begin{aligned}\widehat{\sigma}_{h,jlm}&=\langle\widehat{r}_{jl},\langle\widehat{\Sigma}_{h,j}^{X,\epsilon},\widehat{w}_m\rangle\rangle+\left(\langle\widehat{r}_{jl},\langle\widehat{\Delta}_{h,j}^{X,\epsilon},\phi_m\rangle\rangle+\langle\psi_{jl},\langle\widehat{\Delta}_{h,j}^{X,\epsilon},\widehat{w}_m\rangle\rangle\right)\\&\quad+\langle\psi_{jl},\langle\widehat{\Delta}_{h,j}^{X,\epsilon},\phi_m\rangle\rangle\\&=I_1+I_2+I_3.\end{aligned}$$

Denote $\Omega_{j,\eta}^{X,\epsilon}=\left\{\|\widehat{\Delta}_{h,j}^{X,\epsilon}\|_{\mathcal{S}}\leqslant(\omega_0^X+\omega_0^\epsilon)\mathcal{M}_{X,\epsilon}\eta\right\}$, $\Omega_{jj,\eta}^X=\left\{\|\widehat{\Delta}_{jj}^X\|_{\mathcal{S}}\leqslant 2\mathcal{M}_1^X\omega_0^X\eta\right\}$, $\Omega_\eta^Y=$

$\left\{\|\widehat{\Delta}^Y\|_{\mathcal{S}} \leqslant 2\mathcal{M}^Y\omega_0^Y\eta\right\}$ and $\Omega_1 = \left\{\|\widehat{\Delta}_{h,j}^{X,\epsilon}\|_{\mathcal{S}} \leqslant (\omega_0^X + \omega_0^\epsilon)\right\}$. By Theorem 1.2 and Lemma 1.24, we have

$$
\begin{aligned}
P\left((\Omega_{j,\eta}^{X,\epsilon})^C\right) &\leqslant c_1\exp\{-c_3 n\min(\eta^2,\eta)\},\\
P\left((\Omega_{jj,\eta}^{X})^C\right) &\leqslant 4\exp\{-\tilde{c}_1 n\min(\eta^2,\eta)\},\\
P\left((\Omega_{\eta}^{Y})^C\right) &\leqslant 4\exp\{-\tilde{c}_1 n\min(\eta^2,\eta)\},\\
P\left((\Omega_1)^C\right) &\leqslant c_1\exp\{-c_3 n(\mathcal{M}_{X,\epsilon})^{-2}\}.
\end{aligned}
$$

On the event of $\Omega_1\cap\Omega_{j,\eta}^{X,\epsilon}\cap\Omega_{jj,\eta}^{X}\cap\Omega_{\eta}^{Y}$, by Condition 1.5, (1.40), (1.41) and Lemma 1.2, we obtain that

$$
\begin{aligned}
\left|\frac{I_1}{\sqrt{\omega_{jl}^X\omega_m^Y}}\right| &\leqslant c_0^{-1}(\alpha_1\alpha_2)^{1/2}l^{\alpha_1/2}m^{\alpha_2/2}\|\widehat{r}_{jl}\|(\|\widehat{\Delta}_{h,j}^{X,\epsilon}\|_{\mathcal{S}} + \|\Sigma_{h,j}^{X,\epsilon}\|_{\mathcal{S}})\|\widehat{w}_m\|\\
&\lesssim (l^{3\alpha_1+2}\vee m^{3\alpha_2+2})\mathcal{M}_1^X\mathcal{M}^Y\eta^2,\\
\left|\frac{I_2}{\sqrt{\omega_{jl}^X\omega_m^Y}}\right| &\lesssim l^{\alpha_1/2}m^{\alpha_2/2}\|\widehat{\Delta}_{h,j}^{X,\epsilon}\|_{\mathcal{S}}(l^{\alpha_1+1}\|\widehat{\Delta}_{jj}^{X}\|_{\mathcal{S}} + m^{\alpha_2+1}\|\widehat{\Delta}^{Y}\|_{\mathcal{S}})\\
&\lesssim (l^{2\alpha_1+1}\vee m^{2\alpha_2+1})\mathcal{M}_{X,\epsilon}\{\mathcal{M}_1^X + \mathcal{M}^Y\}\eta^2,\\
\left|\frac{I_3}{\sqrt{\omega_{jl}^X\omega_m^Y}}\right| &\leqslant c_0^{-1}(\alpha_1\alpha_2)^{1/2}l^{\alpha_1/2}m^{\alpha_2/2}\|\widehat{\Delta}_{h,j}^{X,\epsilon}\|_{\mathcal{S}} \lesssim (l^{\alpha_1}\vee m^{\alpha_2})\mathcal{M}_{X,\epsilon}\eta.
\end{aligned}
$$

Combining the above results, we obtain that there exists positive constants ρ_9, ρ_{10}, $\tilde{c}_{11}$ and $\tilde{c}_{12}$ such that

$$
\begin{aligned}
&P\left\{\left|\frac{\widehat{\sigma}_{h,jlm}}{\sqrt{\omega_{jl}^X\omega_m^Y}}\right| \geqslant \rho_9\mathcal{M}_{X,\epsilon}(l^{\alpha_1}\vee m^{\alpha_2})\eta + \rho_{10}\mathcal{M}_{X,\epsilon}(\mathcal{M}_1^X+\mathcal{M}^Y)(l^{3\alpha_1+2}\vee m^{3\alpha_2+2})\eta^2\right\}\\
&\leqslant \tilde{c}_{12}\exp\{-\tilde{c}_{11}n\min(\eta^2,\eta)\} + \tilde{c}_{12}\exp\{-\tilde{c}_{11}\mathcal{M}_{X,\epsilon}^{-2}n\}.
\end{aligned}
$$

Letting $\eta = \rho_{11}\sqrt{\frac{\log(pM_1M_2)}{n}} < 1$ and $\rho_9 + \rho_{10}\rho_{11}\{\mathcal{M}_1^X+\mathcal{M}^Y\}(M_1^{2\alpha_1+2}\vee M_2^{2\alpha_2+2})\eta \leqslant$

ρ_{12}, there exist some constants $c_9, c_{10} > 0$ such that

$$P\left\{\max_{\substack{1\leqslant j\leqslant p\\ 1\leqslant l\leqslant M_1, 1\leqslant m\leqslant M_2}}\left|\frac{\widehat{\sigma}_{h,jlm}-\sigma_{h,jlm}}{\sqrt{\omega_{jl}^X\omega_m^Y}}\right| \geqslant \rho_{11}\rho_{12}(\mathcal{M}_1^X+\mathcal{M}^\epsilon)(M_1^{\alpha_1}\vee M_2^{\alpha_2})\sqrt{\frac{\log(pM_1M_2)}{n}}\right\}$$
$$\leqslant c_9(pM_1M_2)^{c_{10}},$$

which completes the proof. □

Proof of Proposition 1.3 Recall that $\widehat{\varrho}_{h,jl}^{X,\epsilon} = \frac{1}{n-h}\sum_{t=1}^{n-h}\widehat{\zeta}_{tjl}\epsilon_{t+h}$ and $\varrho_{h,jl}^{X,\epsilon} = \mathrm{Cov}(\zeta_{tjl},\epsilon_{t+h})$. Let $\widehat{r}_{jl} = \widehat{\psi}_{jl} - \psi_{jl}$ and $\widehat{\Delta}_{h,j}^{X,\epsilon} = \widehat{\Sigma}_{h,j}^{X,\epsilon} - \Sigma_{h,j}^{X,\epsilon}$. We have

$$\begin{aligned}\widehat{\varrho}_{h,jl}^{X,\epsilon} - \varrho_{h,jl}^{X,\epsilon} &= \langle\widehat{r}_{jl}, \widehat{\Delta}_{h,j}^{X,\epsilon}\rangle + \langle\psi_{jl}, \widehat{\Delta}_{h,j}^{X,\epsilon}\rangle\\ &= I_1 + I_2.\end{aligned}$$

Let $\Omega_{j,\eta}^{X,\epsilon} = \left\{\|\widehat{\Delta}_{h,j}^{X,\epsilon}\| \leqslant (\omega_0^X + (\sigma_0^\epsilon)^2)\mathcal{M}_{X,\epsilon}\eta\right\}$and $\Omega_{jj,\eta}^X = \left\{\|\widehat{\Delta}_{jj}^X\|_{\mathcal{S}} \leqslant 2\mathcal{M}_1^X\omega_0^X\eta\right\}$. By (1.43) and Lemma 1.24, we have

$$\begin{aligned}P\left((\Omega_{j,\eta}^{X,\epsilon})^C\right) &\leqslant c_1\exp\{-c_3 n\min(\eta^2,\eta)\},\\ P\left((\Omega_{jj,\eta}^X)^C\right) &\leqslant 4\exp\{-\tilde{c}_1 n\min(\eta^2,\eta)\}.\end{aligned}$$

On the event of $\Omega_{\eta,j}^{X,\epsilon}\cap\Omega_{jj,\eta}^X$, by Condition 1.5, (1.40) and Lemma 1.2, we obtain that

$$\begin{aligned}\left|\frac{I_1}{\sqrt{\omega_{jl}^X}}\right| &\lesssim l^{\alpha_1/2}\|\widehat{\Delta}_{h,j}^{X,\epsilon}\|\|\widehat{r}_{jl}\| \lesssim l^{3\alpha_1/2+1}\|\widehat{\Delta}_{h,j}^{X,\epsilon}\|\|\widehat{\Delta}_{jj}^X\|_{\mathcal{S}}\\ &\lesssim l^{3\alpha_1/2+1}\mathcal{M}_{X,\epsilon}\mathcal{M}_1^X\eta^2.\end{aligned}$$

By (1.42) and the fact that $\sqrt{(\sigma_0^\epsilon)^2\omega_0^X} \leqslant 1/2\{(\sigma_0^\epsilon)^2+\omega_0^X\}$, we obtain that

$$P\left\{\left|\frac{I_2}{\sqrt{\omega_{jl}^X}}\right| \geqslant 2\mathcal{M}_{X,\epsilon}\sigma_0^\epsilon\eta\right\} \leqslant c_1\exp\{-c_2 n\min(\eta^2,\eta)\}.$$

Combining the above results, we obtain that there exists positive constants ρ_{13}, ρ_{14},

$\tilde{c}_{13}$ and $\tilde{c}_{14}$ such that

$$P\left\{\left|\frac{\widehat{\varrho}_{h,jl}^{X,\epsilon}-\varrho_{h,jl}^{X,\epsilon}}{\sqrt{\omega_{jl}^{X}}}\right|\geqslant\rho_{13}\mathcal{M}_{X,\epsilon}\eta+\rho_{14}l^{3\alpha_1/2+1}\mathcal{M}_{X,\epsilon}\mathcal{M}_1^X\eta^2\right\}\leqslant\tilde{c}_{14}\exp\{-\tilde{c}_{13}n\min(\eta^2,\eta)\}.$$

Letting $\eta=\rho_{15}\sqrt{\frac{\log(pM_1)}{n}}<1$ and $\rho_{13}+\rho_{14}\rho_{15}\mathcal{M}_1^X M_1^{3\alpha_1/2+1}\eta\leqslant\rho_{16}$, there exist some constants $c_{11},c_{12}>0$ such that

$$P\left\{\max_{\substack{1\leqslant j\leqslant p\\1\leqslant l\leqslant M_1}}\left|\frac{\widehat{\varrho}_{h,jl}^{X,\epsilon}-\varrho_{h,jl}^{X,\epsilon}}{\sqrt{\omega_{jl}^{X}}}\right|\geqslant\rho_{15}\rho_{16}\mathcal{M}_{X,\epsilon}\sqrt{\frac{\log(pM_1)}{n}}\right\}\leqslant c_{11}(pM_1)^{c_{12}},$$

which implies (1.22). □

1.B.3 Technical lemmas and their proofs

Lemma 1.1. *The non-functional version of our proposed cross-spectral stability measure satisfies*

$$\underset{\theta\in[-\pi,\pi],\boldsymbol{\nu}_1\in\mathbb{R}_0^p,\boldsymbol{\nu}_2\in\mathbb{R}_0^d}{ess\ sup}\frac{\left|\boldsymbol{\nu}_1^{\mathrm{T}}\boldsymbol{f}_\theta^{X,Y}\boldsymbol{\nu}_2\right|}{\sqrt{\boldsymbol{\nu}_1^{\mathrm{T}}\boldsymbol{\nu}_1}\sqrt{\boldsymbol{\nu}_2^{\mathrm{T}}\boldsymbol{\nu}_2}}\leqslant\widetilde{\mathcal{M}}^{X,Y},$$

where $\widetilde{\mathcal{M}}^{X,Y}$ is defined in (1.6).

Proof. For any fixed $\theta\in[-\pi,\pi]$, we perform singular value decomposition on $\boldsymbol{f}_\theta^{X,Y}=\mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{T}}$, where $\mathbf{D}$ is a diagonal matrix with singular values $\{\sigma_i\}$ of $\boldsymbol{f}_\theta^{X,Y}$ on the diagonal. Then

$$\begin{aligned}
&\max_{\boldsymbol{\nu}_1\in\widetilde{\mathbb{R}}_0^p,\boldsymbol{\nu}_2\in\widetilde{\mathbb{R}}_0^d}\frac{\left|\boldsymbol{\nu}_1^{\mathrm{T}}\boldsymbol{f}_\theta^{X,Y}\boldsymbol{\nu}_2\right|}{\sqrt{\boldsymbol{\nu}_1^{\mathrm{T}}\boldsymbol{\nu}_1}\sqrt{\boldsymbol{\nu}_2^{\mathrm{T}}\boldsymbol{\nu}_2}}\\
=&\max_{\mathbf{x}\in\widetilde{\mathbb{R}}_0^p,\mathbf{y}\in\widetilde{\mathbb{R}}_0^d}\frac{|\mathbf{x}^{\mathrm{T}}\mathbf{D}\mathbf{y}|}{\sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{x}}\sqrt{\mathbf{y}^{\mathrm{T}}\mathbf{y}}}\quad(\mathbf{x}=\mathbf{U}^{\mathrm{T}}\boldsymbol{\nu}_1,\mathbf{y}=\mathbf{V}^{\mathrm{T}}\boldsymbol{\nu}_2)\\
=&\max_{\mathbf{x}\in\widetilde{\mathbb{R}}_0^p,\mathbf{y}\in\widetilde{\mathbb{R}}_0^d}\frac{\sum x_iy_i\sigma_i}{\sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{x}}\sqrt{\mathbf{y}^{\mathrm{T}}\mathbf{y}}}\leqslant\max_{\mathbf{x}\in\widetilde{\mathbb{R}}_0^p,\mathbf{y}\in\widetilde{\mathbb{R}}_0^d}\frac{\sqrt{\sum x_i^2\sum(y_i\sigma_i)^2}}{\sqrt{\sum x_i^2}\sqrt{\sum y_i^2}}\\
\leqslant&\max_{\mathbf{y}\in\widetilde{\mathbb{R}}_0^d}\sqrt{\frac{\sum(y_i\sigma_i)^2}{\sum y_i^2}}\leqslant\max(\sigma_i)\\
\leqslant&\max_{\boldsymbol{\nu}\in\widetilde{\mathbb{R}}_0^d}\sqrt{\frac{\boldsymbol{\nu}^{\mathrm{T}}\{\boldsymbol{f}_\theta^{X,Y}\}^*\boldsymbol{f}_\theta^{X,Y}\boldsymbol{\nu}}{\boldsymbol{\nu}^{\mathrm{T}}\boldsymbol{\nu}}}.
\end{aligned}$$

This holds almost everywhere for $\theta \in [-\pi, \pi]$, which completes our proof. □

Lemma 1.2. *Suppose that Conditions 1.3 and 1.4 hold, then* $\omega_0^X = O(1)$.

Proof. Recall that $\mathbf{X}_t(u) = \sum_{l=0}^{\infty} \int \mathbf{A}_l(u,v)\boldsymbol{\varepsilon}_{t-l}(v)dv$ and $\boldsymbol{\varepsilon}_t(\cdot)$'s are i.i.d. mean-zero functional processes. Let $\mathbf{A}_{l,j}$ denote the j-th row of $\mathbf{A}_l$. Then

$$
\begin{aligned}
&\max_{1\leqslant j\leqslant p} \int \Sigma_{0,jj}^X(u,u)du \\
&= \max_j \int \mathbb{E}\left\{X_{tj}(u)X_{tj}(u)\right\} du \\
&= \max_j \int \mathbb{E}\Big[\Big\{\sum_{l=0}^{\infty}\sum_{k=1}^{p}\int A_{l,jk}(u,v)\varepsilon_{t-l,k}(v)dv\Big\}^2\Big]du \\
&\leqslant \max_j \int \mathbb{E}\Big[\Big\{\sum_{l=0}^{\infty}\sum_{k=1}^{p}\sqrt{\int (A_{l,jk}(u,v))^2dv}\sqrt{\int(\varepsilon_{t-l,k}(v))^2dv}\Big\}^2\Big]du \\
&\leqslant \max_j \sum_{l=0}^{\infty}\sum_{k=1}^{p}\int\int (A_{l,jk}(u,v))^2dudv \max_{l,k}\mathbb{E}\Big\{\int(\varepsilon_{t-l,k}(v))^2dv\Big\} \\
&\leqslant \omega_0^{\varepsilon}\max_j \sum_{l=0}^{\infty}\sum_{k=1}^{p}\|A_{l,jk}\|_{\mathcal{S}}^2 \leqslant \omega_0^{\varepsilon}\max_j\sum_{l=0}^{\infty}(\sum_{k=1}^{p}\|A_{l,jk}\|_{\mathcal{S}})^2 \\
&= \omega_0^{\varepsilon}\sum_{l=0}\|\mathbf{A}_l\|_{\infty}^2 \\
&\leqslant \omega_0^{\varepsilon}\{\sum_{l=0}\|\mathbf{A}_l\|_{\infty}\}^2 = O(1),
\end{aligned}
$$

which completes our proof. □

Before presenting Lemma 1.3, we define sub-Gaussian distribution and sub-Gaussian norm as follows. A centered random variable x with variance proxy σ^2 is sub-Gaussian if for any $t > 0$, $P(|x| > t) \leqslant 2\exp(-t^2/(2\sigma^2))$. The sub-Gaussian norm of x is defined by $\|x\|_{\psi_2} = \inf\{K > 0 : \mathbb{E}\exp(x^2/K^2) \leqslant 2\}$.

Lemma 1.3. *Let* $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ *be a random vector with independent mean zero sub-Gaussian coordinates. Without loss of generality, we assume that* $\mathbb{E}x_i^2 = 1$ *for* $i = 1, \ldots, n$. *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix. Then there exists some universal constant* $c > 0$ *such that for any given* $\eta > 0$,

$$
P\left(|\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} - \mathbb{E}\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}| \geqslant \|\mathbf{A}\|\eta\right) \leqslant 2\exp\left\{-c\min\left(\frac{\eta^2}{rank(\mathbf{A})}, \eta\right)\right\}. \tag{1.44}
$$

Proof. It follows from Theorem 1.1 of Rudelson and Vershynin (2013) and $\|x_i\|_{\psi_2} =$

1 for $i = 1, \ldots, n$, that there exists a constant $c > 0$ such that

$$P\left(|\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} - \mathbb{E}\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}| \geqslant t\right) \leqslant 2\exp\left\{-c\min\left(\frac{t^2}{\|\mathbf{A}\|_{\mathrm{F}}^2}, \frac{t}{\|\mathbf{A}\|}\right)\right\}.$$

By $\|\mathbf{A}\|_{\mathrm{F}} \leqslant \sqrt{\mathrm{rank}(\mathbf{A})}\|\mathbf{A}\|$ and letting $t = \eta\|\mathbf{A}\|$, we obtain (1.44). □

Lemma 1.4. *Suppose that sub-Gaussian process $\{\varepsilon_{tj}(\cdot)\}_{t\in\mathbb{Z}}$ follows Definition 1.3. Under Karhunen-Loève expansion $\varepsilon_{tj}(\cdot) = \sum_{l=1}^{\infty}\xi_{tjl}\phi_{jl}(\cdot) = \sum_{l=1}^{\infty}\sqrt{\omega_{jl}^{\varepsilon}}a_{tjl}\phi_{jl}(\cdot)$ with $\mathbb{E}(a_{tjl}) = 0$ and $\mathbb{E}(a_{tjl}^2) = 1$ for $t \in \mathbb{Z}$ and $j = 1\ldots, p$, a_{tjl} follows sub-Gaussian distribution with $\|a_{tjl}\|_{\psi_2} = 1$, that is for all $\eta > 0$, $t \in \mathbb{Z}$, $j = 1, \ldots, p$ and $l \geqslant 1$,*

$$P[|a_{tjl}| > \eta] \leqslant 2\exp(-\eta^2/2).$$

Proof. By Definition 1.3, for all $x \in \mathbb{H}$, $\mathbb{E}\{e^{\langle x, X\rangle}\} \leqslant e^{\alpha^2\langle x, \Sigma_0(x)\rangle/2}$. Combining with the choice of $x = c\phi_{jl}(\cdot)$ for $c > 0$ and orthonormality of $\{\phi_{jl}(\cdot)\}$ yields

$$\mathbb{E}\left(e^{c\sqrt{\omega_{jl}^{\varepsilon}}a_{tjl}}\right) \leqslant e^{\alpha^2c^2\omega_{jl}^{\boldsymbol{\varepsilon}}/2}.$$

Without loss of generality, we assume $\alpha = 1$. By Markov's inequality and the above result, we have that for all $c > 0$,

$$P(a_{tjl} > \eta) \leqslant P\left(e^{c\sqrt{\omega_{jl}^{\varepsilon}}a_{tjl}} > e^{c\sqrt{\omega_{jl}^{\boldsymbol{\varepsilon}}}\eta}\right) \leqslant \frac{\mathbb{E}\left(e^{c\sqrt{\omega_{jl}^{\varepsilon}}a_{tjl}}\right)}{e^{c\sqrt{\omega_{jl}^{\varepsilon}}\eta}} \leqslant e^{c^2\omega_{jl}^{\varepsilon}/2 - c\sqrt{\omega_{jl}^{\varepsilon}}\eta}.$$

Choosing $c = \eta/\sqrt{\omega_{jl}^{\varepsilon}}$, we have $P(a_{tjl} > \eta) \leqslant e^{-\frac{\eta^2}{2}}$. In the same manner with the choice of $x = -c\phi_{jl}(\cdot)$ for $c > 0$, we can prove $P(a_{tjl} < -\eta) \leqslant e^{-\frac{\eta^2}{2}}$. Combining the above results, $P[|a_{tjl}| > \eta] = P(a_{tjl} > \eta) + P(a_{tjl} < -\eta) \leqslant 2e^{-\frac{\eta^2}{2}}$ which completes the proof. □

Before presenting Lemma 1.5 below, we give some definitions:
(i) Suppose that $\mathbf{e} = (e_1, \ldots, e_N)^{\mathrm{T}} \in \mathbb{H}^N$ is formed by N independent mean zero sub-Gaussian processes with $e_i(\cdot) = \sum_{l=1}^{\infty}\sqrt{\omega_{il}^{e}}a_{il}\phi_{il}(\cdot)$ under the Karhunen-Loève expansion. Define $\boldsymbol{\varphi}_{M,i} = \left(\sqrt{\omega_{i1}^{e}}\phi_{i1}, \ldots, \sqrt{\omega_{iM}^{e}}\phi_{iM}\right)^{\mathrm{T}}$.
(ii) Suppose $\mathbf{K} = (K_{ij})_{N\times N}$ with each $K_{ij} \in \mathbb{S}$. For any nonempty subset $G \subset \mathbb{Z}_+ = \{1,2,\ldots\}$ with $|G| < \infty$, write $G = \{g_1, \ldots, g_{|G|}\}$ with $g_1 < \cdots < g_{|G|}$ and $\boldsymbol{\phi}_{G,i} = (\phi_{ig_1}, \ldots, \phi_{ig_{|G|}})^{\mathrm{T}}$ for each $i = 1, \ldots, N$. Let $\boldsymbol{\Phi}_G = \mathrm{diag}(\boldsymbol{\phi}_{G,1}^{\mathrm{T}}, \ldots, \boldsymbol{\phi}_{G,N}^{\mathrm{T}})$, then

we define

$$\operatorname{rank}(\mathbf{K}) = \sup_{G\subset\mathbb{Z}_+,|G|<\infty} \operatorname{rank}\left(\int\int \boldsymbol{\Phi}_G^{\mathrm{T}}(u)\mathbf{K}(u,v)\boldsymbol{\Phi}_G(v)dudv\right).$$

Condition 1.10. *Let* $\boldsymbol{\Pi}_M = \int\int \boldsymbol{\Theta}_M^{\mathrm{T}}(u)\mathbf{K}(u,v)\boldsymbol{\Theta}_M(v)dudv$ *with* $\boldsymbol{\Theta}_M$ *taking the form* $\boldsymbol{\Theta}_M = diag(\boldsymbol{\varphi}_{M,1}^{\mathrm{T}},\ldots,\boldsymbol{\varphi}_{M,N}^{\mathrm{T}})$ *and* $\mathbf{K} = (K_{ij})_{N\times N}$ *with each* $K_{ij} \in \mathbb{S}$. *It satisfies that* $\|\boldsymbol{\Pi}_M\| \leqslant b_M$ *and* $\lim_{M\to\infty} b_M = b$.

Lemma 1.5. *Suppose that* $\max_{1\leqslant i\leqslant N}\int_{\mathcal{U}} \Sigma_{ii}^e(u,u)du < \infty$ *and* $\mathbf{K}$ *satisfies Condition 1.10. Then, there exists some universal constant* $c > 0$ *such that for any given* $\eta > 0$,

$$P\left(|\langle \mathbf{e},\mathbf{K}(\mathbf{e})\rangle - \mathbb{E}\langle \mathbf{e},\mathbf{K}(\mathbf{e})\rangle| \geqslant b\eta\right) \leqslant 2\exp\left\{-c\min\left(\frac{\eta^2}{rank(\mathbf{K})},\eta\right)\right\}. \tag{1.45}$$

Proof. We organize our proof as follows: First, we truncate $e_i(\cdot)$ to M-dimensional process $e_{M,i}(\cdot) = \sum_{l=1}^M \sqrt{\omega_{il}^e}a_{il}\phi_{il}(\cdot)$, then apply Hanson-Wright inequality in Lemma 1.3 and finally show that the inequality still hold under the infinite-dimensional setting.

Rewrite $\mathbf{e}_M = (e_{M,1},\ldots,e_{M,N})^{\mathrm{T}}$ with $e_{M,i} = \mathbf{a}_{M,i}^{\mathrm{T}}\boldsymbol{\varphi}_{M,i}$ and $\mathbf{a}_{M,i} = (a_{i1},\ldots,a_{iM})^{\mathrm{T}}$. Let $\mathbf{a}_M = (\mathbf{a}_{M,1}^{\mathrm{T}},\ldots,\mathbf{a}_{M,N}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{NM}$, then we have $\langle \mathbf{e}_M,\mathbf{K}(\mathbf{e}_M)\rangle = \mathbf{a}_M^{\mathrm{T}}\boldsymbol{\Pi}_M\mathbf{a}_M$. By Lemma 1.4, elements in $\mathbf{a}_M \in \mathbb{R}^{NM}$ are i.i.d. sub-Gaussian with $\mathbb{E}(a_{il}) = 0$ and $\mathbb{E}(a_{il}^2) = 1$. Combining this with Lemma 1.3 yields

$$\begin{aligned}
&P\left(|\langle \mathbf{e}_M,\mathbf{K}(\mathbf{e}_M)\rangle - \mathbb{E}\langle \mathbf{e}_M,\mathbf{K}(\mathbf{e}_M)\rangle| \geqslant b_M\eta\right)\\
\leqslant\ &P\left(|\mathbf{a}_M^{\mathrm{T}}\boldsymbol{\Pi}_M\mathbf{a}_M - \mathbb{E}\mathbf{a}_M^{\mathrm{T}}\boldsymbol{\Pi}_M\mathbf{a}_M| \geqslant \|\boldsymbol{\Pi}_M\|\eta\right)\\
\leqslant\ &2\exp\left\{-c\min\left(\frac{\eta^2}{\operatorname{rank}(\boldsymbol{\Pi}_M)},\eta\right)\right\}.
\end{aligned} \tag{1.46}$$

It follows from Lemma 1.6 that $\langle \mathbf{e}_M,\mathbf{K}(\mathbf{e}_M)\rangle$ converges in probability to $\langle \mathbf{e},\mathbf{K}(\mathbf{e})\rangle$ and $\lim_{M\to\infty}\mathbb{E}\langle \mathbf{e}_M,\mathbf{K}(\mathbf{e}_M)\rangle = \mathbb{E}\langle \mathbf{e},\mathbf{K}(\mathbf{e})\rangle$. These results together with Condition 1.10 imply that

$$\langle \mathbf{e}_M,\mathbf{K}(\mathbf{e}_M)\rangle - \mathbb{E}\langle \mathbf{e}_M,\mathbf{K}(\mathbf{e}_M)\rangle - b_M\eta$$

converges in distribution to

$$\langle \mathbf{e},\mathbf{K}(\mathbf{e})\rangle - \mathbb{E}\langle \mathbf{e},\mathbf{K}(\mathbf{e})\rangle - b\eta.$$

Finally, by the fact that $\operatorname{rank}(\boldsymbol{\Pi}_M) \leqslant \operatorname{rank}(\mathbf{K})$ and taking $M \to \infty$ on both sides of

(1.46), we obtain (1.45), which completes the proof. ▫

Lemma 1.6. *Under the same assumption and notation in Lemma 1.5 and its proof, we have*

$$\lim_{M\to\infty} \mathbb{E}\left\{\|\mathbf{e}_M - \mathbf{e}\|^2\right\} = 0 \tag{1.47}$$

and

$$\lim_{M\to\infty} \mathbb{E}\langle \mathbf{e}_M, \mathbf{K}(\mathbf{e}_M)\rangle = \mathbb{E}\langle \mathbf{e}, \mathbf{K}(\mathbf{e})\rangle. \tag{1.48}$$

Proof. Since $\|\mathbf{e}_M - \mathbf{e}\|^2 = \sum_{i=1}^N \|e_{M,i} - e_i\|^2 = \sum_{i=1}^N \|\sum_{l=M+1}^\infty \sqrt{\omega_{il}^e} a_{il}\phi_{il}\|^2$, it suffices to show $\lim_{M\to\infty} \mathbb{E}\left\{\|\sum_{l=M+1}^\infty \sqrt{\omega_{il}^e} a_{il}\phi_{il}\|^2\right\} = 0$. By $\mathbb{E}(a_{il}a_{il'}) = 1\{l = l'\}$ and the orthonormality of $\{\phi_{il}\}$, we have

$$\mathbb{E}\left\{\int \left(\sum_{l=M+1}^\infty \sqrt{\omega_{il}^\varepsilon} a_{il}\phi_{il}(u)\right)^2 du\right\} = \sum_{l=M+1}^\infty \omega_{il}^\varepsilon.$$

This together with Condition 1.4 implies that above goes to zero as $M \to \infty$, which completes the proof of (1.47).

By triangle inequality, we have

$$|\mathbb{E}\langle \mathbf{e}_M, \mathbf{K}(\mathbf{e}_M)\rangle - \mathbb{E}\langle \mathbf{e}, \mathbf{K}(\mathbf{e})\rangle| \leqslant |\mathbb{E}\langle \mathbf{e}_M, \mathbf{K}(\mathbf{e}_M - \mathbf{e})\rangle| + |\mathbb{E}\langle (\mathbf{e}_M - \mathbf{e}), \mathbf{K}(\mathbf{e})\rangle|. \tag{1.49}$$

By Jensen's inequality and Lemma 1.11, we have

$$|\mathbb{E}\langle \mathbf{e}_M, \mathbf{K}(\mathbf{e}_M - \mathbf{e})\rangle|^2 \leqslant \|\mathbf{K}\|_{\mathrm{F}}^2 \mathbb{E}(\|\mathbf{e}_M\|^2)\mathbb{E}(\|\mathbf{e}_M - \mathbf{e}\|^2),$$

$$|\mathbb{E}\langle (\mathbf{e}_M - \mathbf{e}), \mathbf{K}(\mathbf{e})\rangle|^2 \leqslant \|\mathbf{K}\|_{\mathrm{F}}^2 \mathbb{E}(\|\mathbf{e}\|^2)\mathbb{E}(\|\mathbf{e}_M - \mathbf{e}\|^2).$$

From (1.47), we have $\lim_{M\to\infty} \mathbb{E}\{\|\mathbf{e}_M - \mathbf{e}\|^2\} = 0$ and $\lim_{M\to\infty} \mathbb{E}\{\|\mathbf{e}_M\|^2\} = \mathbb{E}\{\|\mathbf{e}\|^2\}$. Combining these with $\mathbb{E}(\|\mathbf{e}\|^2) \leqslant N \max_{1\leqslant i\leqslant N} \int_{\mathcal{U}} \Sigma_{ii}^e(u,u)du < \infty$ and $\|\mathbf{K}\|_{\mathrm{F}} < \infty$ implies the right side of (1.49) goes to zero when $M \to \infty$, which completes the proof of (1.48). ▫

Lemma 1.7. *Suppose Conditions 1.1, 1.3 and 1.4 hold for stationary sub-Gaussian process* $\{\mathbf{X}_t(\cdot)\}_{t\in\mathbb{Z}}$. *Let* $\mathbf{X}_{M,L,t}(u) = \sum_{l=0}^L \mathbf{A}_l(\boldsymbol{\varepsilon}_{M,t-l})$. *Then, for any* $\boldsymbol{\Phi}_1 \in \mathbb{H}_0^p$ *with* $\|\boldsymbol{\Phi}_1\|_0 \leqslant k$ *and* $k = 1, \ldots, p$,

$$\lim_{M\to\infty} \mathcal{M}(\boldsymbol{f}_{M,L}^X, \boldsymbol{\Phi}_1) = \mathcal{M}(\boldsymbol{f}_L^X, \boldsymbol{\Phi}_1).$$

Proof. By the definitions of $\mathcal{M}(\boldsymbol{f}_{M,L}^X, \boldsymbol{\Phi})$ and $\boldsymbol{f}_{M,L,\theta}^X(\boldsymbol{\Phi})$ in the proof of Theorem 1.1 in Appendix 1.B.1, we have

$$
\begin{aligned}
&\lim_{M\to\infty} |\mathcal{M}(\boldsymbol{f}_{M,L}^X, \boldsymbol{\Phi}_1) - \mathcal{M}(\boldsymbol{f}_L^X, \boldsymbol{\Phi}_1)| \\
&= 2\pi \lim_{M\to\infty} \left| \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi]} |\langle \boldsymbol{\Phi}_1, \boldsymbol{f}_{M,L,\theta}^X(\boldsymbol{\Phi}_1)\rangle| - \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi]} |\langle \boldsymbol{\Phi}_1, \boldsymbol{f}_{L,\theta}^X(\boldsymbol{\Phi}_1)\rangle| \right| \\
&\leqslant 2\pi \lim_{M\to\infty} \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi]} \left| |\langle \boldsymbol{\Phi}_1, \boldsymbol{f}_{M,L,\theta}^X(\boldsymbol{\Phi}_1)\rangle| - |\langle \boldsymbol{\Phi}_1, \boldsymbol{f}_{L,\theta}^X(\boldsymbol{\Phi}_1)\rangle| \right| \\
&\leqslant \|\boldsymbol{\Phi}_1\|^2 \lim_{M\to\infty} \left\| \sum_{h\in\mathbb{Z}} (\boldsymbol{\Sigma}_{M,L,h}^X - \boldsymbol{\Sigma}_{L,h}^X) \right\|_{\mathrm{F}} \quad \text{(by Lemma 1.11 and } |\exp(-ih\theta)| = 1) \\
&\leqslant \|\boldsymbol{\Phi}_1\|^2 \lim_{M\to\infty} \sum_{h\in\mathbb{Z}} \left\| \boldsymbol{\Sigma}_{M,L,h}^X - \boldsymbol{\Sigma}_{L,h}^X \right\|_{\mathrm{F}}.
\end{aligned}
$$

Provided that $\|\boldsymbol{\Phi}_1\|^2 < \infty$, it suffices to prove that $\sum_{h=-\infty}^{\infty} \left\| \boldsymbol{\Sigma}_{M,L,h}^X - \boldsymbol{\Sigma}_{L,h}^X \right\|_{\mathrm{F}} < \infty$ and $\lim_{M\to\infty} \left\| \boldsymbol{\Sigma}_{M,L,h}^X - \boldsymbol{\Sigma}_{L,h}^X \right\|_{\mathrm{F}} = 0$.

By triangle inequality and Lemma 1.12, we obtain that

$$
\sum_{h=-\infty}^{\infty} \left\| \boldsymbol{\Sigma}_{M,L,h}^X - \boldsymbol{\Sigma}_{L,h}^X \right\|_{\mathrm{F}} \leqslant \sum_{h=-\infty}^{\infty} \|\boldsymbol{\Sigma}_{M,L,h}^X\|_{\mathrm{F}} + \sum_{h=-\infty}^{\infty} \|\boldsymbol{\Sigma}_{L,h}^X\|_{\mathrm{F}} < \infty.
$$

We next prove $\lim_{M\to\infty} \left\| \boldsymbol{\Sigma}_{M,L,h}^X - \boldsymbol{\Sigma}_{L,h}^X \right\|_{\mathrm{F}} = 0$. Write

$$
\begin{aligned}
\boldsymbol{\Sigma}_{M,L,h}^X(u,v) &= \mathbb{E}\left\{ \mathbf{X}_{M,L,t-h}(u) \mathbf{X}_{M,L,t}^{\mathrm{T}}(v) \right\} \\
&= \sum_{l=0}^{L-h} \int \mathbf{A}_{l+h}(u,u') \boldsymbol{\Sigma}_0^{\varepsilon_M}(u',v') \{\mathbf{A}_l(v,v')\}^{\mathrm{T}} du' dv', \\
\boldsymbol{\Sigma}_{L,h}^X(u,v) &= \mathbb{E}\left\{ \mathbf{X}_{L,t-h}(u) \mathbf{X}_{L,t}^{\mathrm{T}}(v) \right\} \\
&= \sum_{l=0}^{L-h} \int \mathbf{A}_{l+h}(u,u') \boldsymbol{\Sigma}_0^{\varepsilon}(u',v') \{\mathbf{A}_l(v,v')\}^{\mathrm{T}} du' dv'.
\end{aligned}
$$

Then,

$$
\begin{aligned}
&\lim_{M\to\infty} \left\|\mathbf{\Sigma}^X_{M,L,h} - \mathbf{\Sigma}^X_{L,h}\right\|_{\mathrm{F}} \\
=&\lim_{M\to\infty} \left\| \sum_{l=0}^{L-h} \int \mathbf{A}_{l+h}(u,u')\{\mathbf{\Sigma}_0^{\varepsilon_M}(u',v') - \mathbf{\Sigma}_0^{\varepsilon}(u',v')\}\{\mathbf{A}_l(v,v')\}^{\mathrm{T}} du'dv' \right\|_{\mathrm{F}} \\
\leqslant& \sum_{l=0}^{L-h} \|\mathbf{A}_l\|_{\mathrm{F}} \|\mathbf{A}_{l+h}\|_{\mathrm{F}} \lim_{M\to\infty} \|\mathbf{\Sigma}_0^{\varepsilon_M} - \mathbf{\Sigma}_0^{\varepsilon}\|_{\mathrm{F}} \quad \text{(by Lemma 1.11)} \\
\leqslant& \sum_{l=0}^{L-h} \|\mathbf{A}_l\|_{\mathrm{F}} \|\mathbf{A}_{l+h}\|_{\mathrm{F}} \lim_{M\to\infty} \left\{ \sum_{j,k} \|\Sigma_{h,jk}^{\varepsilon_M} - \Sigma_{h,jk}^{\varepsilon}\|_{\mathcal{S}}^2 \right\}^{1/2} \\
\leqslant& \sum_{l=0}^{L-h} \|\mathbf{A}_l\|_{\mathrm{F}} \|\mathbf{A}_{l+h}\|_{\mathrm{F}} \lim_{M\to\infty} \sum_{j,k} \|\Sigma_{0,jk}^{\varepsilon_M} - \Sigma_{0,jk}^{\varepsilon}\|_{\mathcal{S}} \\
=&0 \quad \text{(by Lemmas 1.12 and 1.13)}
\end{aligned}
$$

which completes the proof. ▫

Lemma 1.8. *Suppose that conditions in Lemma 1.7 hold. For any* $\mathbf{\Phi}_1 \in \mathbb{H}_0^p$ *with* $\|\mathbf{\Phi}_1\|_0 \leqslant k$ *and* $k = 1, \ldots, p$, *define* $\mathbf{Y} = (\langle \mathbf{\Phi}_1, \mathbf{X}_1 \rangle, \ldots, \langle \mathbf{\Phi}_1, \mathbf{X}_n \rangle)^{\mathrm{T}}$. *Then*

$$\|\mathit{Var}(\mathbf{Y})\| \leqslant \mathcal{M}(\boldsymbol{f}^X, \mathbf{\Phi}_1) \leqslant \mathcal{M}_k^X \langle \mathbf{\Phi}_1, \mathbf{\Sigma}_0^X(\mathbf{\Phi}_1) \rangle.$$

Proof. The proof follows from the proof of Theorem 1 in Guo and Qiao (2022) and hence the proof is omitted here. ▫

Lemma 1.9. *Suppose that conditions in Lemma 1.7 hold. Let* $\mathbf{X}_{L,t}(u) = \sum_{l=0}^{L} \mathbf{A}_l(\boldsymbol{\varepsilon}_{t-l})$. *For any* $\mathbf{\Phi}_1 \in \mathbb{H}_0^p$ *with* $\|\mathbf{\Phi}_1\|_0 \leqslant k$ $(k = 1, \ldots, p)$, *define* $\mathbf{Y}_L = (\langle \mathbf{\Phi}_1, \mathbf{X}_{L,1} \rangle, \ldots, \langle \mathbf{\Phi}_1, \mathbf{X}_{L,n} \rangle)^{\mathrm{T}}$ *and* $\mathbf{Y} = (\langle \mathbf{\Phi}_1, \mathbf{X}_1 \rangle, \ldots, \langle \mathbf{\Phi}_1, \mathbf{X}_n \rangle)^{\mathrm{T}}$, *then*

$$\lim_{L\to\infty} \mathbb{E}\left\{\|\mathbf{Y}_L - \mathbf{Y}\|^2\right\} = 0 \tag{1.50}$$

and

$$\lim_{L\to\infty} \mathbb{E}\left[\mathbf{Y}_L^{\mathrm{T}} \mathbf{Y}_L\right] = \mathbb{E}\left[\mathbf{Y}^{\mathrm{T}} \mathbf{Y}\right]. \tag{1.51}$$

Proof of (1.50). By definitions of $\mathbf{Y}_L$ and $\mathbf{Y}$, we have that

$$\mathbb{E}\left\{\|\mathbf{Y}_L - \mathbf{Y}\|^2\right\} = \sum_{t=1}^{n} \mathbb{E}\left\{|\langle \mathbf{\Phi}_1, \mathbf{X}_{L,t} - \mathbf{X}_t \rangle|^2\right\}$$

By Lemma 1.11, we have $\mathbb{E}\left\{|\langle \mathbf{\Phi}_1, \mathbf{X}_{L,t} - \mathbf{X}_t\rangle|^2\right\} \leqslant \|\mathbf{\Phi}_1\|^2 \mathbb{E}\{\|\mathbf{X}_{L,t} - \mathbf{X}_t\|^2\}$. With the fact $\|\mathbf{\Phi}_1\|^2 < \infty$, it suffices to prove that $\lim_{L\to\infty} \mathbb{E}\{\|\mathbf{X}_{L,t} - \mathbf{X}_t\|^2\} = 0$ for $t = 1, \ldots, n$. By Lemma 1.13, we have $\mathbb{E}(\|\boldsymbol{\varepsilon}_{t-l}\|) \leqslant \sqrt{p\omega_0^{\boldsymbol{\varepsilon}}}$. This together with Lemma 1.11 implies that

$$\begin{aligned}
\mathbb{E}(\|\mathbf{X}_{L,t} - \mathbf{X}_t\|^2) &= \mathbb{E}\left\{\left\|\sum_{l=L+1}^{\infty} \int \mathbf{A}_l(u,v)\boldsymbol{\varepsilon}_{t-l}(v)dv\right\|^2\right\} \\
&\leqslant \mathbb{E}\left(\sum_{l_1=L+1}^{\infty}\sum_{l_2=L+1}^{\infty} \|\mathbf{A}_{l_1}\|_{\mathrm{F}}\|\mathbf{A}_{l_2}\|_{\mathrm{F}}\|\boldsymbol{\varepsilon}_{t-l_1}\|\|\boldsymbol{\varepsilon}_{t-l_2}\|\right) \\
&\leqslant p\omega_0^{\boldsymbol{\varepsilon}}\left(\sum_{l=L+1}^{\infty}\|\mathbf{A}_l\|_{\mathrm{F}}\right)^2.
\end{aligned}$$

By Lemma 1.12, we have $\sum_{l=0}^{\infty}\|\mathbf{A}_l\|_{\mathrm{F}} < \infty$. This together with the above yields

$$\lim_{L\to\infty} \mathbb{E}\{\|\mathbf{X}_{L,t} - \mathbf{X}_t\|^2\} = 0, \tag{1.52}$$

which completes the proof of (1.50).

Proof of (1.51). Next we show that $\lim_{L\to\infty} \mathbb{E}\left[\mathbf{Y}_L^{\mathrm{T}}\mathbf{Y}_L\right] - \mathbb{E}\left[\mathbf{Y}^{\mathrm{T}}\mathbf{Y}\right] = 0$. Write

$$\begin{aligned}
&\left|\mathbb{E}\left[\mathbf{Y}_L^{\mathrm{T}}\mathbf{Y}_L\right] - \mathbb{E}\left[\mathbf{Y}^{\mathrm{T}}\mathbf{Y}\right]\right| \\
&= n\left|\langle \mathbf{\Phi}_1, (\mathbf{\Sigma}_{L,0}^X - \mathbf{\Sigma}_0^X)(\mathbf{\Phi}_1)\rangle\right| \\
&= n\left|\int \mathbf{\Phi}_1^{\mathrm{T}}(u)\mathbb{E}\left(\mathbf{X}_{L,t}(u)\mathbf{X}_{L,t}^{\mathrm{T}}(v) - \mathbf{X}_t(u)\mathbf{X}_t^{\mathrm{T}}(v)\right)\mathbf{\Phi}_1(v)dudv\right| \\
&\leqslant n\left|\int \mathbf{\Phi}_1^{\mathrm{T}}\mathbb{E}\left(\mathbf{X}_{L,t}(\mathbf{X}_{L,t} - \mathbf{X}_t)^{\mathrm{T}}\right)\mathbf{\Phi}_1 dudv\right| \\
&\quad + n\left|\int \mathbf{\Phi}_1^{\mathrm{T}}\mathbb{E}\left((\mathbf{X}_{L,t} - \mathbf{X}_t)\mathbf{X}_t^{\mathrm{T}}\right)\mathbf{\Phi}_1 dudv\right|.
\end{aligned}$$

By Jensen's inequality and Lemma 1.11, we have

$$\begin{aligned}
\left|\int \mathbf{\Phi}_1^{\mathrm{T}}\mathbb{E}\left(\mathbf{X}_{L,t}(\mathbf{X}_{L,t} - \mathbf{X}_t)^{\mathrm{T}}\right)\mathbf{\Phi}_1 dudv\right|^2 &\leqslant \|\mathbf{\Phi}_1\|^4\mathbb{E}\{\|\mathbf{X}_{L,t}\|^2\}\mathbb{E}\{\|\mathbf{X}_{L,t} - \mathbf{X}_t\|^2\}, \\
\left|\int \mathbf{\Phi}_1^{\mathrm{T}}\mathbb{E}\left((\mathbf{X}_{L,t} - \mathbf{X}_t)\mathbf{X}_t^{\mathrm{T}}\right)\mathbf{\Phi}_1 dudv\right|^2 &\leqslant \|\mathbf{\Phi}_1\|^4\mathbb{E}\{\|\mathbf{X}_t\|^2\}\mathbb{E}\{\|\mathbf{X}_{L,t} - \mathbf{X}_t\|^2\}.
\end{aligned}$$

Combining the above results with (1.52), we complete the proof of (1.51). □

Lemma 1.10. *Suppose conditions in Lemma 1.7 hold. Let* $\mathbf{X}_{L,t}(u) = \sum_{l=0}^{L} \mathbf{A}_l(\boldsymbol{\varepsilon}_{t-l})$. *Then, for any* $\boldsymbol{\Phi}_1 \in \mathbb{H}_0^p$ *with* $\|\boldsymbol{\Phi}_1\|_0 \leqslant k$ *and* $k = 1, \ldots, p$,

$$\lim_{L\to\infty} \mathcal{M}(\boldsymbol{f}_L^X, \boldsymbol{\Phi}_1) = \mathcal{M}(\boldsymbol{f}^X, \boldsymbol{\Phi}_1).$$

Proof. By definitions of $\mathcal{M}(\boldsymbol{f}^X, \boldsymbol{\Phi})$ and $\boldsymbol{f}_\theta^X(\boldsymbol{\Phi})$, we have

$$\begin{aligned}
&\lim_{L\to\infty} |\mathcal{M}(\boldsymbol{f}_L^X, \boldsymbol{\Phi}_1) - \mathcal{M}(\boldsymbol{f}^X, \boldsymbol{\Phi}_1)| \\
&= 2\pi \lim_{L\to\infty} \left| \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi]} |\langle \boldsymbol{\Phi}_1, \boldsymbol{f}_{L,\theta}^X(\boldsymbol{\Phi}_1)\rangle| - \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi]} |\langle \boldsymbol{\Phi}_1, \boldsymbol{f}_\theta^X(\boldsymbol{\Phi}_1)\rangle| \right| \\
&\leqslant 2\pi \lim_{L\to\infty} \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi]} \left| |\langle \boldsymbol{\Phi}_1, \boldsymbol{f}_{L,\theta}^X(\boldsymbol{\Phi}_1)\rangle| - |\langle \boldsymbol{\Phi}_1, \boldsymbol{f}_\theta^X(\boldsymbol{\Phi}_1)\rangle| \right| \\
&\leqslant \|\boldsymbol{\Phi}_1\|^2 \lim_{L\to\infty} \left\| \sum_{h\in\mathbb{Z}} (\boldsymbol{\Sigma}_{L,h}^X - \boldsymbol{\Sigma}_h^X) \right\|_{\mathrm{F}} \quad \text{(by Lemma 1.11 and } |\exp(-ih\theta)| = 1) \\
&\leqslant \|\boldsymbol{\Phi}_1\|^2 \lim_{L\to\infty} \sum_{h\in\mathbb{Z}} \left\| \boldsymbol{\Sigma}_{L,h}^X - \boldsymbol{\Sigma}_h^X \right\|_{\mathrm{F}}.
\end{aligned}$$

With $\|\boldsymbol{\Phi}_1\|^2 < \infty$, it suffices to prove $\sum_{h=-\infty}^{\infty} \left\| \boldsymbol{\Sigma}_{L,h}^X - \boldsymbol{\Sigma}_h^X \right\|_{\mathrm{F}} < \infty$ and $\lim_{L\to\infty} \left\| \boldsymbol{\Sigma}_{L,h}^X - \boldsymbol{\Sigma}_h^X \right\|_{\mathrm{F}} = 0$.

By triangle inequality and Lemma 1.12, we obtain that

$$\sum_{h=-\infty}^{\infty} \left\| \boldsymbol{\Sigma}_{L,h}^X - \boldsymbol{\Sigma}_h^X \right\|_{\mathrm{F}} \leqslant \sum_{h=-\infty}^{\infty} \|\boldsymbol{\Sigma}_{L,h}^X\|_{\mathrm{F}} + \sum_{h=-\infty}^{\infty} \|\boldsymbol{\Sigma}_h^X\|_{\mathrm{F}} < \infty.$$

We next prove $\lim_{L\to\infty} \left\| \boldsymbol{\Sigma}_{L,h}^X - \boldsymbol{\Sigma}_h^X \right\|_{\mathrm{F}} = 0$. Write

$$\begin{aligned}
\boldsymbol{\Sigma}_h^X(u,v) &= \mathbb{E}\left(\mathbf{X}_{t-h}(u)\mathbf{X}_t^{\mathrm{T}}(v)\right) = \sum_{l=0}^{\infty} \int \mathbf{A}_{l+h}(u,u')\boldsymbol{\Sigma}_0^{\varepsilon}(u',v')\{\mathbf{A}_l(v,v')\}^{\mathrm{T}} du'dv', \\
\boldsymbol{\Sigma}_{L,h}^X(u,v) &= \mathbb{E}\left(\mathbf{X}_{L,t-h}(u)\mathbf{X}_{L,t}^{\mathrm{T}}(v)\right) = \sum_{l=0}^{L-h} \int \mathbf{A}_{l+h}(u,u')\boldsymbol{\Sigma}_0^{\varepsilon}(u',v')\{\mathbf{A}_l(v,v')\}^{\mathrm{T}} du'dv'.
\end{aligned}$$

Then,

$$
\begin{aligned}
&\lim_{L\to\infty} \left\|\mathbf{\Sigma}^X_{L,h} - \mathbf{\Sigma}^X_h\right\|_{\mathrm{F}} \\
&= \lim_{L\to\infty} \left\| \sum_{l=L-h+1}^{\infty} \int \mathbf{A}_{l+h}(u,u')\mathbf{\Sigma}^{\varepsilon}_0(u',v')\{\mathbf{A}_l(v,v')\}^{\mathrm{T}} du'dv' \right\|_{\mathrm{F}} \\
&\leqslant p\omega_0^{\varepsilon} \lim_{L\to\infty} \sum_{l=L-h+1}^{\infty} \|\mathbf{A}_l\|_{\mathrm{F}}\|\mathbf{A}_{l+h}\|_{\mathrm{F}} \quad \text{(by Lemmas 1.11 and 1.13)} \\
&\leqslant p\omega_0^{\varepsilon} \lim_{L\to\infty} \sum_{l=L-h+1}^{\infty} \|\mathbf{A}_l\|_{\mathrm{F}} \sum_{l=L-h+1}^{\infty} \|\mathbf{A}_{l+h}\|_{\mathrm{F}} \\
&= 0 \quad \text{(by Lemma 1.12)},
\end{aligned}
$$

which completes the proof. □

Lemma 1.11. *(i)Let* $\mathbf{A} = (A_{ij})_{p\times q}$ *with each* $A_{ij} \in \mathbb{S}$ *and* $\mathbf{B} = (B_1, \ldots, B_q)^{\mathrm{T}} \in \mathbb{H}^q$.

$$
\left\| \int\int \mathbf{A}(u,v)\mathbf{B}(v)dudv \right\| \leqslant \|\mathbf{A}\|_F \|\mathbf{B}\|. \tag{1.53}
$$

Similarly, we have

$$
\left\| \int \mathbf{A}(u,v)\mathbf{B}(v)dv \right\| \leqslant \|\mathbf{A}\|_F \|\mathbf{B}\|, \tag{1.54}
$$

(ii)Let $\mathbf{A} = (A_{ij})_{p\times q}$ *with each* $A_{ij} \in \mathbb{S}$ *and* $\mathbf{B} = (B_{jk})_{q\times r}$ *with each* $B_{jk} \in \mathbb{S}$. *Then we have*

$$
\left\| \int \mathbf{A}(u,z)\mathbf{B}(z,v)dz \right\|_F \leqslant \|\mathbf{A}\|_F \|\mathbf{B}\|_F. \tag{1.55}
$$

Proof of (1.53). Let $C = \int\int \mathbf{A}(u,v)\mathbf{B}(v)dudv$. Thus $|C_i| = |\sum_k \int\int A_{ik}(u,v)B_k(v)dudv| \leqslant \sum_k \|A_{ik}\|_{\mathcal{S}}\|B_k\|$.

$$
\begin{aligned}
\|C\|^2 = \sum_i |C_i|^2 &\leqslant \sum_i \left(\sum_k \|A_{ik}\|_{\mathcal{S}}\|B_k\| \right)^2 \\
&\leqslant \sum_i \left(\sum_k \|A_{ik}\|^2_{\mathcal{S}} \right) \left(\sum_k \|B_k\|^2 \right) \quad \text{(by Cauchy-Schwarz inequality)} \\
&\leqslant \sum_{i,k} \|A_{ik}\|^2_{\mathcal{S}} \sum_k \|B_k\|^2 = \|\mathbf{A}\|^2_{\mathrm{F}}\|\mathbf{B}\|^2.
\end{aligned}
$$

Proof of (1.54). Let $C(u) = \int \mathbf{A}(u,v)\mathbf{B}(v)dv$, then we have that $C_i(u) = \sum_k \int A_{ik}(u,v)B_k(v)dv$.

$$\begin{aligned}
\|C\|^2 &= \sum_i \int C_i(u)^2 du = \sum_i \int \left\{ \sum_k \int A_{ik}(u,v)B_k(v)dv \right\}^2 du \\
&\leqslant \sum_i \int \left\{ \sum_k \sqrt{\int A_{ik}^2(u,v)dv \int B_k^2(v)dv} \right\}^2 du \\
&\leqslant \sum_i \int \left\{ \sum_k \int A_{ik}^2(u,v)dv \sum_k \int B_k^2(v)dv \right\} du \\
&= \sum_{i,k} \|A_{ik}\|^2 \sum_k \|B_k\|^2 = \|\mathbf{A}\|_{\mathrm{F}}^2 \|\mathbf{B}\|^2.
\end{aligned}$$

Proof of (1.55). Let $\mathbf{C}(u,v) = \int \mathbf{A}(u,z)\mathbf{B}(z,v)dz$, then we have $C_{ij}(u,v) = \sum_k \int A_{ik}(u,z)B_{kj}(z,v)dz$. Following the similar argument in the proof of (1.54), we obtain

$$\begin{aligned}
\|\mathbf{C}\|_{\mathrm{F}}^2 &= \sum_{i,j} \int\int C_{ij}(u,v)^2 dudv = \sum_{i,j} \int\int \left\{ \sum_k \int A_{ik}(u,z)B_{kj}(z,v)dz \right\}^2 dudv \\
&\leqslant \sum_{i,j} \int\int \left\{ \sum_k \int A_{ik}^2(u,z)dz \sum_k \int B_{kj}^2(z,v)dz \right\} dudv \\
&= \|\mathbf{A}\|_{\mathrm{F}}^2 \|\mathbf{B}\|_{\mathrm{F}}^2.
\end{aligned}$$

□

Lemma 1.12. *Suppose that conditions in Lemma 1.7 hold. Then we have*

$$\sum_{l=0}^{\infty} \|\mathbf{A}_l\|_F < \infty$$

and

$$\sum_{h\in\mathbb{Z}} \|\mathbf{\Sigma}_h^X\|_F \leqslant 2p\omega_0^{\varepsilon} \left\{ \sum_{l=0}^{\infty} \|\mathbf{A}_l\|_F \right\}^2 < \infty.$$

Proof. It follows from Condition 1.3 that

$$\sum_{l=0}^{\infty}\|\mathbf{A}_l\|_{\mathrm{F}} = \sum_{l=0}^{\infty}\left\{\sum_{j,k}\|A_{l,jk}\|_{\mathcal{S}}^2\right\}^{1/2} \leqslant \sum_{l=0}^{\infty}\sum_{j}\|\mathbf{A}_l\|_{\infty} < \infty.$$

Provided that $\mathbf{X}_t(u) = \sum_{l=0}^{\infty}\int \mathbf{A}_l(u,v)\boldsymbol{\varepsilon}_{t-l}(v)dv$ and $\boldsymbol{\varepsilon}_t(\cdot)$'s are i.i.d. mean zero sub-Gaussian processes, we have

$$\begin{aligned}\boldsymbol{\Sigma}_h^X(u,v) &= \mathbb{E}\left\{\mathbf{X}_{t-h}(u)\mathbf{X}_t^{\mathrm{T}}(v)\right\} \\ &= \sum_{l=0}^{\infty}\int \mathbf{A}_{l+h}(u,u')\mathbb{E}\left\{\boldsymbol{\varepsilon}_{t-l}(u')\boldsymbol{\varepsilon}_{t-l}^{\mathrm{T}}(v')\right\}\left\{\mathbf{A}_l(v,v')\right\}^{\mathrm{T}}du'dv' \\ &= \sum_{l=0}^{\infty}\int \mathbf{A}_{l+h}(u,u')\boldsymbol{\Sigma}_0^{\varepsilon}(u',v')\{\mathbf{A}_l(v,v')\}^{\mathrm{T}}du'dv'.\end{aligned}$$

This together with the fact that $\boldsymbol{\Sigma}_{-h}^X(u,v) = \left\{\boldsymbol{\Sigma}_h^X(v,u)\right\}^{\mathrm{T}}$ implies that

$$\begin{aligned}\sum_{h\in\mathbb{Z}}\|\boldsymbol{\Sigma}_h^X(u,v)\|_{\mathrm{F}} &\leqslant 2\sum_{h=0}^{\infty}\|\boldsymbol{\Sigma}_h^X(u,v)\|_{\mathrm{F}} \\ &= 2\sum_{h=0}^{\infty}\|\sum_{l=0}^{\infty}\int \mathbf{A}_{l+h}(u,u')\boldsymbol{\Sigma}_0^{\varepsilon}(u',v')\{\mathbf{A}_l(v,v')\}^{\mathrm{T}}du'dv'\|_{\mathrm{F}} \\ &\leqslant 2\sum_{h=0}^{\infty}\sum_{l=0}^{\infty}\|\int \mathbf{A}_{l+h}(u,u')\boldsymbol{\Sigma}_0^{\varepsilon}(u',v')\{\mathbf{A}_l(v,v')\}^{\mathrm{T}}du'dv'\|_{\mathrm{F}} \\ &\leqslant 2\sum_{h=0}^{\infty}\sum_{l=0}^{\infty}\|\mathbf{A}_l\|_{\mathrm{F}}\|\mathbf{A}_{l+h}\|_{\mathrm{F}}\|\boldsymbol{\Sigma}_0^{\varepsilon}\|_{\mathrm{F}} \quad \text{(by Lemma 1.11)} \\ &\leqslant 2p\omega_0^{\varepsilon}\left\{\sum_{l=0}^{\infty}\|\mathbf{A}_l\|_{\mathrm{F}}\right\}^2 < \infty \quad \text{(by Lemme 1.13)},\end{aligned}$$

which completes the proof. □

Lemma 1.13. *For a p-dimensional vector process* $\{\mathbf{X}_t(\cdot)\}_{t\in\mathbb{Z}}$, *whose lag-h auto-covariance matrix function is* $\boldsymbol{\Sigma}_h = (\Sigma_{h,jk})_{1\leqslant j,k\leqslant p}$ *with each* $\Sigma_{h,jk} \in \mathbb{S}$ *and* $\omega_0 = \max_{1\leqslant j\leqslant p}\int \Sigma_{0,jj}(u,u)du < \infty$, *we have*

$$\|\Sigma_{h,jk}\|_{\mathcal{S}} \leqslant \omega_0, \quad \|\boldsymbol{\Sigma}_h\|_F \leqslant p\omega_0, \quad \mathbb{E}(\|\mathbf{X}_t\|) \leqslant \sqrt{p\omega_0} \quad \textit{and} \quad \mathbb{E}(\|\mathbf{X}_t\|^2) \leqslant p\omega_0.$$

Let $X_{M,tj}(\cdot) = \sum_{l=1}^{M} \xi_{tjl}\phi_{jl}(\cdot)$ *be the* M*-truncated process, we have*

$$\lim_{M\to\infty} \|\Sigma_{h,jk}^{X_M} - \Sigma_{h,jk}^{X}\|_{\mathcal{S}} = 0. \tag{1.56}$$

Proof. By $\Sigma_{h,jk} = \sum_{l,m=1}^{\infty} \mathbb{E}(\xi_{tjl}\xi_{(t+h)km})\phi_{jl}(u)\phi_{km}(v)$, orthonormality of $\{\phi_{jl}\}$ and Cauchy–Schwarz inequality, we obtain

$$\begin{aligned}\|\Sigma_{h,jk}\|_{\mathcal{S}}^2 &= \int \left\{ \sum_{l,m=1}^{\infty} \mathbb{E}(\xi_{tjl}\xi_{(t+h)km})\phi_{jl}(u)\phi_{km}(v) \right\}^2 dudv \\ &= \sum_{l,m=1}^{\infty} \mathbb{E}(\xi_{tjl}\xi_{(t+h)km})^2 \leqslant \sum_{l,m=1}^{\infty} \mathbb{E}(\xi_{tjl}^2)\mathbb{E}(\xi_{(t+h)km}^2) \leqslant \omega_0^2.\end{aligned}$$

This implies that $\|\mathbf{\Sigma}_h\|_{\mathrm{F}}^2 = \sum_{j,k} \|\Sigma_{h,jk}\|_{\mathcal{S}}^2 \leqslant p^2\omega_0^2$. By the similar arguments, we have

$$\begin{aligned}\|\Sigma_{h,jk}^{X_M} - \Sigma_{h,jk}^{X}\|_{\mathcal{S}}^2 &= \int \left\{ \sum_{l,m=M+1}^{\infty} \mathbb{E}(\xi_{tjl}\xi_{(t+h)km})\phi_{jl}(u)\phi_{km}(v) \right\}^2 dudv \\ &= \sum_{l,m=M+1}^{\infty} \mathbb{E}(\xi_{tjl}\xi_{(t+h)km})^2 \leqslant \sum_{l,m=M+1}^{\infty} \mathbb{E}(\xi_{tjl}^2)\mathbb{E}(\xi_{(t+h)km}^2).\end{aligned}$$

Since $\sum_{l=0}^{\infty} \mathbb{E}(\xi_{tjl}^2) \leqslant \omega_0 < \infty$, the above goes to zero when $M \to \infty$, completing the proof of (1.56).

Provided that $X_{tj}(\cdot) = \sum_{l=1}^{\infty} \xi_{tjl}\phi_{jl}(\cdot)$, orthonormality of $\{\phi_{jl}\}$ and Jensen's inequality, we have

$$\begin{aligned}\mathbb{E}(\|\mathbf{X}_t\|) &= \mathbb{E}\left\{ \sqrt{\sum_{j=1}^{p} \int X_{tj}^2(u)du} \right\} \leqslant \sqrt{\sum_{j=1}^{p} \mathbb{E}\left\{ \int X_{tj}^2(u)du \right\}} \\ &\leqslant \sqrt{\sum_{j=1}^{p}\sum_{l=0}^{\infty} \mathbb{E}(\xi_{tjl}^2)} \leqslant \sqrt{p\omega_0}.\end{aligned}$$

Similarly, we obtain that $\mathbb{E}(\|\mathbf{X}_t\|^2) = \mathbb{E}\left\{\sum_{j=1}^{p} \int X_{tj}^2(u)du\right\} = \sum_j \sum_l \mathbb{E}(\xi_{tjl}^2) \leqslant p\omega_0$. □

Lemma 1.14. *For process* $\{\mathbf{X}_t(\cdot)\}_{t\in\mathbb{Z}}$ *and* $\{\mathbf{Y}_t(\cdot)\}_{t\in\mathbb{Z}}$, *we have that*

$$\|\Sigma_{h,jk}^{X,Y}\|_{\mathcal{S}} \leqslant \sqrt{\omega_0^X \omega_0^Y},$$

and

$$\|\langle \Sigma_{h,jk}^{X,Y}, \phi_{km}\rangle\| \leqslant \sqrt{\omega_0^X \omega_{km}^Y} \quad \textit{and} \quad \|\langle \Sigma_{h,jk}^{X,Y}, \psi_{jl}\rangle\| \leqslant \sqrt{\omega_{jl}^X \omega_0^Y}.$$

Proof. This lemma can be proved in similar way to Lemma 8 of Guo and Qiao (2022) and hence the proof is omitted here. □

1.C Proofs of theoretical results in Section 1.3

We present the proof of Theorem 1.4 in Appendix 1.C.1 and proofs of Propositions 1.4–1.7 in Appendix 1.C.2, followed by the supporting technical lemmas and their proofs in Appendix 1.C.3. For a matrix $\mathbf{A} \in \mathbb{R}^{p\times q}$, we denote its elementwise maximum norm by $\|\mathbf{A}\|_{\max} = \max_{i,j}|A_{ij}|$. To simplify our notation, for a square-block matrix $\mathbf{B} = (\mathbf{B}_{jk})_{1\leqslant j\leqslant p_1, 1\leqslant k\leqslant p_2} \in \mathbb{R}^{p_1 q\times p_2 q}$ with the (j,k)-th block $\mathbf{B}_{jk} \in \mathbb{R}^{q\times q}$, we use $\|\mathbf{B}\|_{\max}^{(q)}$ and $\|\mathbf{B}\|_1^{(q)}$ to denote its block versions of elementwise ℓ_∞ and matrix ℓ_1 norms.

1.C.1 Proof of Theorem 1.4

Denote the minimizer of (1.25) by $\widehat{\mathbf{B}} \in \mathbb{R}^{(L+1)pq_1\times q_2}$. Then

$$\frac{1}{2(n-L)}\|\widehat{\mathbf{U}} - \widehat{\mathbf{Z}}\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{B}}\|_{\mathrm{F}}^2 + \lambda_n\|\widehat{\mathbf{B}}\|_1^{(q_1,q_2)} \leqslant \frac{1}{2(n-L)}\|\widehat{\mathbf{U}} - \widehat{\mathbf{Z}}\widehat{\mathbf{D}}^{-1}\mathbf{B}\|_{\mathrm{F}}^2 + \lambda_n\|\mathbf{B}\|_1^{(q_1,q_2)}$$

Let $\boldsymbol{\Delta} = \widehat{\mathbf{B}} - \mathbf{B}$ and S^c be the complement of S in the set $\{0,\ldots,L\}\times\{1,\ldots,p\}$, For matrices $\mathbf{A},\mathbf{B}\in\mathbb{R}^{p_1\times p_2}$, we let $\langle\langle \mathbf{A},\mathbf{B}\rangle\rangle = \mathrm{trace}(\mathbf{A}^{\mathrm{T}}\mathbf{B})$. Then we write

$$\begin{aligned}
&\frac{1}{2}\langle\langle \boldsymbol{\Delta}, \widehat{\boldsymbol{\Gamma}}\boldsymbol{\Delta}\rangle\rangle \\
\leqslant&\frac{1}{n-L}\langle\langle \boldsymbol{\Delta}, \widehat{\mathbf{D}}^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}(\widehat{\mathbf{U}} - \widehat{\mathbf{Z}}\widehat{\mathbf{D}}^{-1}\mathbf{B})\rangle\rangle + \lambda_n(\|\mathbf{B}\|_1^{(q_1,q_2)} - \|\mathbf{B}+\boldsymbol{\Delta}\|_1^{(q_1,q_2)}) \\
\leqslant&\frac{1}{n-L}\langle\langle \boldsymbol{\Delta}, \widehat{\mathbf{D}}^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}(\widehat{\mathbf{U}} - \widehat{\mathbf{Z}}\widehat{\mathbf{D}}^{-1}\mathbf{B})\rangle\rangle + \lambda_n(\|\boldsymbol{\Delta}_S\|_1^{(q_1,q_2)} - \|\boldsymbol{\Delta}_{S^c}\|_1^{(q_1,q_2)}),
\end{aligned}$$

where $\widehat{\mathbf{\Gamma}} = (n-L)^{-1}\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}\widehat{\mathbf{Z}}\widehat{\mathbf{D}}^{-1}$. By Proposition 1.7 and $\lambda_n \geqslant 2C_0 s q_1^{1/2}((\mathcal{M}_1^X + \mathcal{M}^{\epsilon}) \vee \mathcal{M}_1^Y)\{(q_1^{\alpha_1+3/2} \vee q_2^{\alpha_2+3/2})\sqrt{\frac{\log(pq_1q_2)}{n}} + q_1^{-\kappa+1/2}\}$, we have

$$\begin{aligned}
&\frac{1}{n-L}|\langle\langle \mathbf{\Delta}, \widehat{\mathbf{D}}^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}(\widehat{\mathbf{U}} - \widehat{\mathbf{Z}}\widehat{\mathbf{D}}^{-1}\mathbf{B})\rangle\rangle| \\
\leqslant &\frac{1}{n-L}\|\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}(\widehat{\mathbf{U}} - \widehat{\mathbf{Z}}\widehat{\mathbf{D}}^{-1}\mathbf{B})\|_{\max}^{(q_1,q_2)}\|\mathbf{\Delta}\|_1^{(q_1,q_2)} \\
\leqslant &\frac{\lambda_n}{2}(\|\mathbf{\Delta}_S\|_1^{(q_1,q_2)} + \|\mathbf{\Delta}_{S^c}\|_1^{(q_1,q_2)}).
\end{aligned}$$

This implies that

$$0 \leqslant \frac{1}{2}\langle\langle \mathbf{\Delta}, \widehat{\mathbf{\Gamma}}\mathbf{\Delta}\rangle\rangle \leqslant \frac{3\lambda_n}{2}\|\mathbf{\Delta}_S\|_1^{(q_1,q_2)} - \frac{\lambda_n}{2}\|\mathbf{\Delta}_{S^c}\|_1^{(q_1,q_2)} \leqslant \frac{3}{2}\lambda_n\|\mathbf{\Delta}\|_1^{(q_1,q_2)}.$$

Therefore $\|\mathbf{\Delta}\|_1^{(q_1,q_2)} \leqslant 4\|\mathbf{\Delta}_S\|_1^{(q_1,q_2)} \leqslant 4\sqrt{s}\|\mathbf{\Delta}\|_{\mathrm{F}}$. By Proposition 1.4 and $\tau_2 \geqslant 32\tau_1 q_1 q_2 s$, we obtain

$$\langle\langle \mathbf{\Delta}, \widehat{\mathbf{\Gamma}}\mathbf{\Delta}\rangle\rangle \geqslant \tau_2\|\mathbf{\Delta}\|_{\mathrm{F}}^2 - \tau_1 q_1 q_2\{\|\mathbf{\Delta}\|_1^{(q_1,q_2)}\}^2 \geqslant (\tau_2 - 16\tau_1 q_1 q_2 s)\|\mathbf{\Delta}\|_{\mathrm{F}}^2 \geqslant \frac{\tau_2}{2}\|\mathbf{\Delta}\|_{\mathrm{F}}^2.$$

Therefore,

$$\frac{\tau_2}{4}\|\mathbf{\Delta}\|_{\mathrm{F}}^2 \leqslant \frac{3}{2}\lambda_n\|\mathbf{\Delta}\|_1^{(q_1,q_2)} \leqslant 6\lambda_n s^{1/2}\|\mathbf{\Delta}\|_{\mathrm{F}},$$

which implies that

$$\|\mathbf{\Delta}\|_{\mathrm{F}} \leqslant \frac{24 s^{1/2}\lambda_n}{\tau_2} \text{ and } \|\mathbf{\Delta}\|_1^{(q_1,q_2)} \leqslant \frac{96 s\lambda_n}{\tau_2}. \tag{1.57}$$

Here, we aim to prove the upper bound of $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_1$. For each $(h,j) \in S$ we have,

$$\begin{aligned}
\widehat{\beta}_{hj} - \beta_{hj} &= \widehat{\boldsymbol{\psi}}_j(u)^{\mathrm{T}}\widehat{\mathbf{\Psi}}_{hj}\widehat{\boldsymbol{\phi}}(v) - \boldsymbol{\psi}_j(u)^{\mathrm{T}}\mathbf{\Psi}_{hj}\boldsymbol{\phi}(v) + R_{hj}(u,v) \\
&= (\widehat{\boldsymbol{\psi}}_j(u) - \boldsymbol{\psi}_j(u))^{\mathrm{T}}\widehat{\mathbf{\Psi}}_{hj}\widehat{\boldsymbol{\phi}}(v) + \boldsymbol{\psi}_j(u)^{\mathrm{T}}\widehat{\mathbf{\Psi}}_{hj}(\widehat{\boldsymbol{\phi}}(v) - \boldsymbol{\phi}(v)) \\
&\quad + \boldsymbol{\psi}_j(u)^{\mathrm{T}}(\widehat{\mathbf{\Psi}}_{hj} - \mathbf{\Psi}_{hj})\boldsymbol{\phi}(v) + R_{hj}(u,v),
\end{aligned}$$

where $R_{hj}(u,v) = -\sum_{l=q_1+1}^{\infty}\sum_{m=q_2+1}^{\infty} a_{hjlm}\psi_{jl}(u)\phi_m(v)$. Therefore,

$$\begin{aligned}\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|_1 \leqslant & \sum_{h,j}\|(\widehat{\boldsymbol{\psi}}_j(u)-\boldsymbol{\psi}_j(u))^{\mathrm{T}}\widehat{\boldsymbol{\Psi}}_{hj}\widehat{\boldsymbol{\phi}}(v)\|_{\mathcal{S}} + \sum_{h,j}\|\boldsymbol{\psi}_j(u)^{\mathrm{T}}\widehat{\boldsymbol{\Psi}}_{hj}(\widehat{\boldsymbol{\phi}}(v)-\boldsymbol{\phi}(v))\|_{\mathcal{S}} \\ & + \sum_{h,j}\|\boldsymbol{\psi}_j(u)^{\mathrm{T}}(\widehat{\boldsymbol{\Psi}}_{hj}-\boldsymbol{\Psi}_{hj})\boldsymbol{\phi}(v)\|_{\mathcal{S}} + \sum_{h,j}\|R_{hj}(u,v)\|_{\mathcal{S}}.\end{aligned} \tag{1.58}$$

Due to the orthonormality of $\{\psi_{jl}(\cdot)\}$ and $\{\phi_m(\cdot)\}$ and the estimated eigenfunctions $\{\widehat{\psi}_{jl}(\cdot)\}$ and $\{\widehat{\phi}_m(\cdot)\}$,

$$\begin{aligned}&\|(\widehat{\boldsymbol{\psi}}_j(u)-\boldsymbol{\psi}_j(u))^{\mathrm{T}}\widehat{\boldsymbol{\Psi}}_{hj}\widehat{\boldsymbol{\phi}}(v)\|_{\mathcal{S}} \leqslant q_1^{1/2}\|\widehat{\boldsymbol{\Psi}}_{hj}\|_{\mathrm{F}}\max_l\|\widehat{\psi}_{jl}-\psi_{jl}\|,\\ &\|\boldsymbol{\psi}_j(u)^{\mathrm{T}}\widehat{\boldsymbol{\Psi}}_{hj}(\widehat{\boldsymbol{\phi}}(v)-\boldsymbol{\phi}(v))\|_{\mathcal{S}} \leqslant q_2^{1/2}\|\widehat{\boldsymbol{\Psi}}_{hj}\|_{\mathrm{F}}\max_m\|\widehat{\phi}_m-\psi_m\|,\\ &\|\boldsymbol{\psi}_j(u)^{\mathrm{T}}(\widehat{\boldsymbol{\Psi}}_{hj}-\boldsymbol{\Psi}_{hj})\boldsymbol{\phi}(v)\|_{\mathcal{S}} = \|\widehat{\boldsymbol{\Psi}}_{hj}-\boldsymbol{\Psi}_{hj}\|_{\mathrm{F}}.\end{aligned}$$

To bound the first three terms of (1.58), we start with the upper bound of $\sum_{h,j}\|\widehat{\boldsymbol{\Psi}}_{hj}-\boldsymbol{\Psi}_{hj}\|_{\mathrm{F}} = \|\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi}\|_1^{(q_1,q_2)}$ and $\sum_{h,j}\|\widehat{\boldsymbol{\Psi}}_{hj}\|_{\mathrm{F}} = \|\widehat{\boldsymbol{\Psi}}\|_1^{(q_1,q_2)}$. From Condition 1.6, for $(h,j)\in S$, $\|\boldsymbol{\Psi}_{hj}\|_{\mathrm{F}} = \{\sum_{l=1}^{q_1}\sum_{m=1}^{q_2}\mu_{hj}^2(l+m)^{-2\kappa-1}\}^{1/2} \leqslant \{\mu_{hj}^2\int_1^{q_2}\int_1^{q_1}(x+y)^{-2\kappa-1}dxdy\}^{1/2} = O(\mu_{hj})$. For $(h,j)\in S^c$, $\boldsymbol{\Psi}_{hj}=0$. Hence,

$$\|\boldsymbol{\Psi}\|_1^{(q_1,q_2)} = \sum_{h,j}\|\boldsymbol{\Psi}_{hj}\|_{\mathrm{F}} = O(s). \tag{1.59}$$

By the definition of ω_0^X, Condition 1.5 and Proposition 1.6, we have $\|\mathbf{D}\|_{\max}\leqslant\sqrt{\omega_0^X}$, $\|\mathbf{D}^{-1}\|_{\max}\leqslant\alpha_1^{1/2}c_0^{-1/2}q_1^{\alpha_1/2}$ and

$$\|\widehat{\mathbf{D}}^{-1}-\mathbf{D}^{-1}\|_{\max}\leqslant\alpha_1^{1/2}c_0^{-1/2}q_1^{\alpha_1/2}C_\omega\mathcal{M}_1^X\sqrt{\frac{\log(pq_1)}{n}}.$$

Recall that $\widehat{\mathbf{\Psi}} - \mathbf{\Psi} = \widehat{\mathbf{D}}^{-1}\widehat{\mathbf{B}} - \mathbf{D}^{-1}\mathbf{B} = \mathbf{D}^{-1}(\widehat{\mathbf{B}} - \mathbf{B}) + (\widehat{\mathbf{D}}^{-1} - \mathbf{D}^{-1})\widehat{\mathbf{B}}$. Then

$$\begin{aligned}
\|\widehat{\mathbf{\Psi}} - \mathbf{\Psi}\|_1^{(q_1,q_2)} &\leqslant \|\mathbf{D}^{-1}\|_{\max}\|\widehat{\mathbf{B}} - \mathbf{B}\|_1^{(q_1,q_2)} + \|\widehat{\mathbf{D}}^{-1} - \mathbf{D}^{-1}\|_{\max}\|\widehat{\mathbf{B}}\|_1^{(q_1,q_2)} \\
&\leqslant \|\mathbf{D}^{-1}\|_{\max}\|\widehat{\mathbf{B}} - \mathbf{B}\|_1^{(q_1,q_2)} + \|\widehat{\mathbf{D}}^{-1} - \mathbf{D}^{-1}\|_{\max}\|\widehat{\mathbf{B}} - \mathbf{B}\|_1^{(q_1,q_2)} \\
&\quad + \|\widehat{\mathbf{D}}^{-1} - \mathbf{D}^{-1}\|_{\max}\|\mathbf{B}\|_1^{(q_1,q_2)} \\
&\leqslant \|\mathbf{D}^{-1}\|_{\max}\|\widehat{\mathbf{B}} - \mathbf{B}\|_1^{(q_1,q_2)} + \|\widehat{\mathbf{D}}^{-1} - \mathbf{D}^{-1}\|_{\max}\|\widehat{\mathbf{B}} - \mathbf{B}\|_1^{(q_1,q_2)} \\
&\quad + \|\widehat{\mathbf{D}}^{-1} - \mathbf{D}^{-1}\|_{\max}\|\mathbf{D}\|_{\max}\|\mathbf{\Psi}\|_1^{(q_1,q_2)}.
\end{aligned}$$

This, together with (1.57) implies that,

$$\|\mathbf{B}\|_1^{(q_1,q_2)} = O(\sqrt{\omega_0^X}s), \tag{1.60}$$

and

$$\|\widehat{\mathbf{\Psi}} - \mathbf{\Psi}\|_1^{(q_1,q_2)} \leqslant \frac{96\alpha_1^{1/2}q_1^{\alpha_1/2}s\lambda_n}{c_0^{1/2}\tau_2}\{1 + o(1)\}. \tag{1.61}$$

Combining (1.59) and (1.61), we have

$$\|\widehat{\mathbf{\Psi}}\|_1^{(q_1,q_2)} = O(s).$$

To bound the fourth term of (1.58), $\|R_{hj}\|_{\mathcal{S}} = O(\|\sum_{l=1}^{q_1}\sum_{m=q_2+1}^{\infty}a_{hjlm}\psi_{jl}\phi_m\|_{\mathcal{S}} \vee \|\sum_{l=q_1+1}^{\infty}\sum_{m=1}^{q_2}a_{hjlm}\psi_{jl}\phi_m\|_{\mathcal{S}}) = O(\mu_{hj}\min(q_1,q_2)^{-\kappa+1/2})$, for each $(h,j) \in S$. For $(h,j) \in S^c$, $\|R_{hj}\|_{\mathcal{S}} = 0$. Hence, $\sum\limits_{h,j}\|R_{hj}\|_{\mathcal{S}} = O(s\min(q_1,q_2)^{-\kappa+1/2})$.

Combining all the results with Proposition 1.6, we obtain

$$\begin{aligned}
\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_1 &\leqslant \|\widehat{\mathbf{\Psi}}\|_1^{(q_1,q_2)}\left\{q_1^{1/2}\max_{j,l}\|\widehat{\psi}_{jl} - \psi_{jl}\|_{\mathcal{S}} + q_2^{1/2}\max_{m}\|\widehat{\phi}_m - \phi_m\|_{\mathcal{S}}\right\} \\
&\quad + \|\widehat{\mathbf{\Psi}} - \mathbf{\Psi}\|_1^{(q_1,q_2)} + \sum_{h,j}\|R_{hj}\|_{\mathcal{S}} \\
&\leqslant \frac{96\alpha_1^{1/2}q_1^{\alpha_1/2}s\lambda_n}{c_0^{1/2}\tau_2}\{1 + o(1)\},
\end{aligned}$$

which completes the proof. □

1.C.2 Proofs of propositions

Proof of Proposition 1.4 Define $\mathbf{\Gamma} = (n-L)^{-1}\mathbf{D}^{-1}\mathbb{E}\{\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\}\mathbf{D}^{-1}$. Note that $\boldsymbol{\theta}^{\mathrm{T}}\widehat{\mathbf{\Gamma}}\boldsymbol{\theta} = \boldsymbol{\theta}^{\mathrm{T}}\mathbf{\Gamma}\boldsymbol{\theta} + \boldsymbol{\theta}^{\mathrm{T}}(\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma})\boldsymbol{\theta}$. Hence we have

$$\boldsymbol{\theta}^{\mathrm{T}}\widehat{\mathbf{\Gamma}}\boldsymbol{\theta} \geqslant \boldsymbol{\theta}^{\mathrm{T}}\mathbf{\Gamma}\boldsymbol{\theta} - \|\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma}\|_{\max}\|\boldsymbol{\theta}\|_1^2.$$

By Condition 1.7, $\omega_{\min}(\mathbf{\Gamma}) \geqslant \underline{\mu}$, where $\omega_{\min}(\mathbf{\Gamma})$ denotes the minimum eigenvalue of $\mathbf{\Gamma}$. This, together with Lemma 1.16, completes our proof. □

Proof of Proposition 1.6 This proposition can be proved in similar way to Proposition 3 of Guo and Qiao (2022) and hence the proof is omitted here. □

Proof of Proposition 1.7 Notice that $\widehat{\mathbf{U}} = \mathbf{Z}\mathbf{D}^{-1}\tilde{\mathbf{B}} + \widehat{\mathbf{R}} + \widehat{\mathbf{E}}$, where $\tilde{\mathbf{B}} = \mathbf{D}\tilde{\mathbf{\Psi}}$ and $\{(h+1)j\}$-th row block of $\tilde{\mathbf{\Psi}}$, $\tilde{\mathbf{\Psi}}_{hj} = \int_{\mathcal{V}}\int_{\mathcal{U}}\boldsymbol{\psi}_j(u)\beta_{hj}(u,v)\widehat{\boldsymbol{\phi}}(v)^{\mathrm{T}}dudv$. The matrix $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{E}}$ are both $(n-L)\times q_2$ matrices whose row vectors are formed by $\{\widehat{\mathbf{r}}_t = (\widehat{r}_{t1},\ldots,\widehat{r}_{tq_2})^{\mathrm{T}}\}_{L+1}^{n}$ and $\{\widehat{\boldsymbol{\epsilon}}_t = (\widehat{\epsilon}_{t1},\ldots,\widehat{\epsilon}_{tq_2})^{\mathrm{T}}\}_{L+1}^{n}$ respectively, where $\widehat{r}_{tm} = \sum_{h=0}^{L}\sum_{j=1}^{p}\sum_{l=q_1+1}^{\infty}\langle\langle\psi_{jl},\beta_{hj}\rangle,\widehat{\phi}_m\rangle\zeta_{tjl}$ and $\widehat{\epsilon}_{tm} = \langle\epsilon_t,\widehat{\phi}_m\rangle$. Then we rewrite

$$\begin{aligned}
&\frac{1}{n-L}\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}(\widehat{\mathbf{U}} - \widehat{\mathbf{Z}}\widehat{\mathbf{D}}^{-1}\mathbf{B})\\
&= \frac{1}{n-L}\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}(\mathbf{Z}\mathbf{D}^{-1}\tilde{\mathbf{B}} - \widehat{\mathbf{Z}}\widehat{\mathbf{D}}^{-1}\mathbf{B}) + \frac{1}{n-L}\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}\widehat{\mathbf{R}} + \frac{1}{n-L}\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}\widehat{\mathbf{E}}\\
&= I_1 + I_2 + I_3.
\end{aligned}$$

Next, we show the deviation bounds of the above three parts.

$$\begin{aligned}
&\|I_1\|_{\max}^{(q_1,q_2)}\\
=&\|\frac{1}{n-L}\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}(\mathbf{Z}\mathbf{D}^{-1} - \widehat{\mathbf{Z}}\widehat{\mathbf{D}}^{-1})\mathbf{B}\|_{\max}^{(q_1,q_2)} + \|\frac{1}{n-L}\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}\mathbf{Z}\mathbf{D}^{-1}(\tilde{\mathbf{B}} - \mathbf{B})\|_{\max}^{(q_1,q_2)}\\
\leqslant&\|\frac{1}{n-L}\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}(\mathbf{Z}\mathbf{D}^{-1} - \widehat{\mathbf{Z}}\widehat{\mathbf{D}}^{-1})\|_{\max}^{(q_1)}\|\mathbf{B}\|_1^{(q_1,q_2)}\\
&+ \|\frac{1}{n-L}\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}\mathbf{Z}\mathbf{D}^{-1}\|_{\max}^{(q_1)}\|\tilde{\mathbf{B}} - \mathbf{B}\|_1^{(q_1,q_2)}\\
\leqslant&\|\frac{1}{n-L}\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}(\mathbf{Z}\mathbf{D}^{-1} - \widehat{\mathbf{Z}}\widehat{\mathbf{D}}^{-1})\|_{\max}^{(q_1)}\|\mathbf{B}\|_1^{(q_1,q_2)} + \|\widehat{\mathbf{\Gamma}}\|_{\max}^{(q_1)}\|\tilde{\mathbf{B}} - \mathbf{B}\|_1^{(q_1,q_2)}\\
&+ \|\frac{1}{n-L}\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}(\mathbf{Z}\mathbf{D}^{-1} - \widehat{\mathbf{Z}}\widehat{\mathbf{D}}^{-1})\|_{\max}^{(q_1)}\|\tilde{\mathbf{B}} - \mathbf{B}\|_1^{(q_1,q_2)},
\end{aligned}$$

where $\widehat{\mathbf{\Gamma}} = (n-L)^{-1}\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}\widehat{\mathbf{Z}}\widehat{\mathbf{D}}^{-1}$. By Lemmas 1.15, 1.17, 1.18 and (1.60) in Appendix 1.C.1, there exist some positive constants C_1^*, c_1^* and c_2^* such that

$$\|I_1\|_{\max}^{(q_1,q_2)} \leqslant C_1^* s q_1^{1/2}(\mathcal{M}_1^X q_1^{\alpha_1+3/2} \vee \mathcal{M}_1^Y q_2^{\alpha_2+3/2})\sqrt{\frac{\log(pq_1 \vee q_2)}{n}} \tag{1.62}$$

with probability greater than $1 - c_1^*(pq_1 \vee q_2)^{-c_2^*}$.

By Lemma 1.19, we obtain that there exist some positive constants C_2^*, c_1^* and c_2^* such that

$$\|I_2\|_{\max}^{(q_1,q_2)} \leqslant C_2^* s q_1^{-\kappa+1} \tag{1.63}$$

with probability greater than $1 - c_1^*(pq_1q_2)^{-c_2^*}$.

Let $\mathbf{Q} = ((n-L)^{-1}\mathbf{U}^{\mathrm{T}}\mathbf{U})^{1/2} = \mathrm{diag}(\{\omega_1^Y\}^{1/2}, \ldots, \{\omega_q^Y\}^{1/2})$. It follows from Proposition 1.2 and $\|\mathbf{Q}\|_{\mathrm{F}} \leqslant \sqrt{\omega_0^Y}$ that there exist some positive constants C_3^*, c_1^* and c_2^* such that

$$\begin{aligned}\|I_3\|_{\max}^{(q_1,q_2)} &\leqslant q_1^{1/2}\|\widehat{\mathbf{D}}^{-1}\mathbf{D}\|_{\max}\|(n-L)^{-1}\mathbf{D}^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}\widehat{\mathbf{E}}\mathbf{Q}^{-1}\|_{\max}\|\mathbf{Q}\|_{\mathrm{F}} \\ &\leqslant C_3^* q_1^{1/2}(\mathcal{M}_1^X + \mathcal{M}^\epsilon)(q_1^{\alpha_1} \vee q_2^{\alpha_2})\sqrt{\frac{\log(pq_1q_2)}{n}}\end{aligned} \tag{1.64}$$

with probability greater than $1 - c_1^*(pq_1q_2)^{-c_2^*}$.

It follows from (1.62)–(1.64) that there exist some positive constants C_0, c_1^* and c_2^* such that

$$\begin{aligned}&\frac{1}{n-L}\|\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}(\widehat{\mathbf{U}} - \widehat{\mathbf{Z}}\widehat{\mathbf{D}}^{-1}\mathbf{B})\|_{\max}^{(q_1,q_2)} \\ \leqslant\ & C_0 s q_1^{1/2}\left((\mathcal{M}_1^X + \mathcal{M}^\epsilon) \vee \mathcal{M}_1^Y\right)\left\{\left(q_1^{\alpha_1+3/2} \vee q_2^{\alpha_2+3/2}\right)\sqrt{\frac{\log(pq_1q_2)}{n}} + q_1^{-\kappa+1/2}\right\}\end{aligned}$$

with probability greater than $1 - c_1^*(pq_1q_2)^{-c_2^*}$, which completes the proof. □

1.C.3 Technical lemmas and their proofs

Lemma 1.15. $\|\widehat{\mathbf{\Gamma}}\|_{\max}^{(q_1)} = O(q_1^{1/2})$.

Proof. For a semi-positive definite block matrix

$$\mathbf{A} = \begin{pmatrix} \boldsymbol{L} & \boldsymbol{X} \\ \boldsymbol{X}^{\mathrm{T}} & \boldsymbol{M} \end{pmatrix},$$

we have that $\|\mathbf{X}\|_{\mathrm{F}}^2 \leqslant \|\boldsymbol{L}\|_{\mathrm{F}}\|\mathbf{M}\|_{\mathrm{F}}$. This can be seen as a special case of $p = 1$ in Theorem 4.2 of Horn and Mathias (1990). Without loss of generality, we take $L = 0$ as an example. Let $\widehat{\mathbf{\Gamma}}_{jk} = (\widehat{\Gamma}_{jl,km})_{1\leqslant l,m\leqslant q_1}$. Then for $j = k$, by the diagonal structure of $\widehat{\mathbf{\Gamma}}_{jj}$, we have $\|\widehat{\mathbf{\Gamma}}_{jj}\|_{\mathrm{F}} = O(q_1^{1/2})$. Applying the above inequality, we obtain $\|\widehat{\mathbf{\Gamma}}_{jk}\|_{\mathrm{F}} \leqslant \sqrt{\|\widehat{\mathbf{\Gamma}}_{jj}\|_{\mathrm{F}}\|\widehat{\mathbf{\Gamma}}_{kk}\|_{\mathrm{F}}} = O(q_1^{1/2})$. □

Lemma 1.16. *Suppose that Conditions 1.1–1.5 hold. Then there exist some positive constants C_Γ, c_1^* and c_2^* such that*

$$\left\|\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma}\right\|_{\max} \leqslant C_\Gamma \mathcal{M}_1^X q_1^{\alpha_1+1}\sqrt{\frac{\log(pq_1)}{n}}$$

with probability greater than $1 - c_1^*(pq_1)^{-c_2^*}$.

Proof. The proof follows from Lemma 5 in Guo and Qiao (2022). □

Lemma 1.17. *Suppose that Conditions 1.1–1.5 hold. Then there exist some positive constants $\tilde{C}_\Gamma$, c_1^* and c_2^* such that*

$$\|\frac{1}{n-L}\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}(\mathbf{Z}\mathbf{D}^{-1} - \widehat{\mathbf{Z}}\widehat{\mathbf{D}}^{-1})\|_{\max} \leqslant \tilde{C}_\Gamma \mathcal{M}_1^X q_1^{\alpha_1+1}\sqrt{\frac{\log(pq_1)}{n}}$$

with probability greater than $1 - c_1^*(pq_1)^{-c_2^*}$.

Proof. We first consider $\|\frac{1}{n-L}\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}\mathbf{Z}\mathbf{D}^{-1} - \mathbf{\Gamma}\|_{\max}$. By Lemma 1.26, Proposition 1.6

and following the similar argument in the proof of Lemma 1.27, we obtain that

$$\begin{aligned}
&\max_{j,k,l,m} \frac{(n-L)^{-1}\sum_{t=L+1}^{n}\widehat{\zeta}_{(t-h)jl}\zeta_{tkm}}{\sqrt{\widehat{\omega}_{jl}^{X}\omega_{km}^{X}}} - \frac{\mathbb{E}(\zeta_{(t-h)jl}\zeta_{tkm})}{\sqrt{\omega_{jl}^{X}\omega_{km}^{X}}} \\
&\lesssim \max_{j,k,l,m} \frac{\langle\widehat{\psi}_{jl},\langle\widehat{\Sigma}_{h,jk}^{X},\phi_{km}\rangle\rangle - \langle\psi_{jl},\langle\Sigma_{h,jk}^{X},\phi_{km}\rangle\rangle}{\sqrt{\omega_{jl}^{X}\omega_{km}^{X}}} \\
&\lesssim \max_{j,k,l,m} \frac{\langle\widehat{\psi}_{jl}-\psi_{jl},\langle\Sigma_{h,jk}^{X},\phi_{km}\rangle\rangle + \langle\widehat{\psi}_{jl},\langle\widehat{\Sigma}_{h,jk}^{X}-\Sigma_{h,jk}^{X},\phi_{km}\rangle\rangle}{\sqrt{\omega_{jl}^{X}\omega_{km}^{X}}} \\
&\lesssim \mathcal{M}_1^X q_1^{\alpha_1+1}\sqrt{\frac{\log(pq_1)}{n}}
\end{aligned}$$

holds with probability greater than $1-c_1^*(pq_1)^{-c_2^*}$. This, together with Lemma 1.16, shows that

$$\begin{aligned}
&\|\frac{1}{n-L}\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}(\mathbf{Z}\mathbf{D}^{-1}-\widehat{\mathbf{Z}}\widehat{\mathbf{D}}^{-1})\|_{\max} \\
\leqslant &\|\frac{1}{n-L}\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}\mathbf{Z}\mathbf{D}^{-1}-\mathbf{\Gamma}\|_{\max} + \|\widehat{\mathbf{\Gamma}}-\mathbf{\Gamma}\|_{\max} \\
= &O_P\{\mathcal{M}_1^X q_1^{\alpha_1+1}\sqrt{\frac{\log(pq_1)}{n}}\}.
\end{aligned}$$

□

Lemma 1.18. *Suppose that Conditions 1.1–1.6 hold. Then there exist some positive constants C_B, c_1^* and c_2^* such that*

$$\|\widetilde{\mathbf{B}}-\mathbf{B}\|_1^{(q_1,q_2)} \leqslant C_B s\mathcal{M}_1^Y q_2^{\alpha_2+3/2}\sqrt{\frac{\log(q_2)}{n}}$$

with probability greater than $1-c_1^*(q_2)^{-c_2^*}$.

Proof. We start with the convergence rate of $\|\widetilde{\mathbf{\Psi}}-\mathbf{\Psi}\|_1^{(q_1,q_2)}$. Elementwisely, for fixed h, j and $l=1,\ldots,q_1, m=1,\ldots,q_2$, we have that

$$\langle\langle\psi_{jl},\beta_{hj}\rangle,\widehat{\phi}_m\rangle - \langle\langle\psi_{jl},\beta_{hj}\rangle,\phi_m\rangle = \langle\langle\psi_{jl},\beta_{hj}\rangle,\widehat{\phi}_m-\phi_m\rangle = I_1$$

Recall that $\beta_{hj} = \sum_{l,m=1}^{\infty} a_{hjlm}\psi_{jl}(u)\phi_m(v)$ and $|a_{hjlm}| \leqslant u_{hj}(l+m)^{-\kappa-1/2}$.

$$I_1 = \langle\langle \psi_{jl}, \sum_{l',m'=1}^{\infty} a_{hjl'm'}\psi_{jl'}\phi_{m'}\rangle, \widehat{\phi}_m - \phi_m\rangle = \sum_{m'=1}^{\infty} a_{hjlm'}\langle \phi_{m'}, \widehat{\phi}_m - \phi_m\rangle$$
$$\lesssim \|\widehat{\phi}_m - \phi_m\| u_{hj} l^{-\kappa+1/2}.$$

It follows from Lemma 1.25, for $(h,j) \in S$,

$$\|\widetilde{\mathbf{\Psi}}_{hj} - \mathbf{\Psi}_{hj}\|_{\mathrm{F}} = \sqrt{\sum_{l=1}^{q_1}\sum_{m=1}^{q_2} I_1^2} \lesssim u_{hj} q_2^{1/2} \max_{1\leqslant m\leqslant q_2} \|\widehat{\phi}_m - \phi_m\|$$
$$= O_P\left\{ u_{hj}\mathcal{M}_1^Y q_2^{\alpha_2+3/2}\sqrt{\frac{\log(q_2)}{n}} \right\}.$$

Then $\|\widetilde{\mathbf{\Psi}} - \mathbf{\Psi}\|_1^{(q_1,q_2)} = \sum_{h=0}^{L}\sum_{j=1}^{p} \|\widetilde{\mathbf{\Psi}}_{hj} - \mathbf{\Psi}_{hj}\|_{\mathrm{F}} = O_P\left\{ s\mathcal{M}_1^Y q_2^{\alpha_2+3/2}\sqrt{\frac{\log(q_2)}{n}} \right\}$. This result, together with $\|\mathbf{D}\|_{\max} \leqslant \{\omega_0^X\}^{1/2}$, implies that there exists C_B such that

$$\begin{aligned} \|\tilde{\mathbf{B}} - \mathbf{B}\|_1^{(q_1,q_2)} &= \|\mathbf{D}(\widetilde{\mathbf{\Psi}} - \mathbf{\Psi})\|_1^{(q_1,q_2)} \leqslant \|\mathbf{D}\|_{\max}\|\widetilde{\mathbf{\Psi}} - \mathbf{\Psi}\|_1^{(q_1,q_2)} \\ &\leqslant C_B s\mathcal{M}_1^Y q_2^{\alpha_2+3/2}\sqrt{\frac{\log(q_2)}{n}}, \end{aligned}$$

with probability greater than $1 - c_1^*(q_2)^{-c_2^*}$. □

Lemma 1.19. *Suppose that Conditions 1.1–1.6 hold. Then there exist some positive constants C_R, c_1^* and c_2^* such that*

$$\|(n-L)^{-1}\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}\widehat{\mathbf{R}}\|_{\max}^{(q_1,q_2)} \leqslant C_R s q_1^{-\kappa+1}$$

with probability greater than $1 - c_1^*(pq_1q_2)^{-c_2^*}$.

Proof. Recall that we have $\widehat{r}_{tm} = \sum_{h=0}^{L}\sum_{j=1}^{p}\sum_{l=q_1+1}^{\infty}\langle\langle\psi_{jl}, \beta_{hj}\rangle, \widehat{\phi}_m\rangle\zeta_{tjl} = \sum_{h=0}^{L}\sum_{j=1}^{p}\tilde{r}_{tmhj}$. The matrix $\widehat{\mathbf{R}}$ are $(n-L)\times q_2$ matrices whose row vectors are formed by $\{\widehat{\mathbf{r}}_t = (\widehat{r}_{t1},\ldots,\widehat{r}_{tq_2})^{\mathrm{T}}, t = L+1,\ldots,n\}$. By Cauchy-Schwarz inequality and the definition

of $\widehat{\omega}_{jl}^X$, we obtain

$$\begin{aligned}
&\frac{(n-L)^{-1}\sum_{t=L+1}^{n}\widehat{\zeta}_{(t-h)jl}\sum_{h=0}^{L}\sum_{j'=1}^{p}\tilde{r}_{tmhj'}}{\{\widehat{\omega}_{jl}^X\}^{1/2}} \\
&\leqslant \sum_{h=0}^{L}\sum_{j'=1}^{p}\sqrt{(n-L)^{-1}\sum_{t=L+1}^{n}\tilde{r}_{tmhj'}^2} \\
&= \sum_{h=0}^{L}\sum_{j'=1}^{p}\sqrt{\mathbb{E}(\tilde{r}_{tmhj'}^2)+(n-L)^{-1}\sum_{t=L+1}^{n}\{\tilde{r}_{tmhj'}^2-\mathbb{E}(\tilde{r}_{tmhj'}^2)\}} \\
&= \sum_{h=0}^{L}\sum_{j'=1}^{p}\sqrt{I_{1,tmhj'}+I_{2,tmhj'}}.
\end{aligned}$$

Recall that $\mathrm{Cov}(\zeta_{tjl},\zeta_{tjl'})=\omega_{jl}^X I(l=l')$, $\beta_{hj}(u,v)=\sum_{l,m=1}^{\infty}a_{hjlm}\psi_{jl}(u)\phi_m(v)$ and $|a_{hjlm}|\leqslant u_{hj}(l+m)^{-\kappa-1/2}$. Then for $(h,j')\in S$,

$$\begin{aligned}
I_{1,tmhj'} &= \mathbb{E}\left[\left(\sum_{l'=q_1+1}^{\infty}\langle\psi_{j'l'},\langle\beta_{hj'},\widehat{\phi}_m\rangle\rangle\zeta_{tj'l'}\right)^2\right]=\sum_{l'=q_1+1}^{\infty}\langle\psi_{j'l'},\langle\beta_{hj'},\widehat{\phi}_m\rangle\rangle^2\omega_{j'l'} \\
&\lesssim \sum_{l'=q_1+1}^{\infty}\langle\psi_{j'l'},\langle\sum_{l'',m''=1}^{\infty}a_{hj'l''m''}\psi_{j'l''}\phi_{m''}\rangle,\phi_m+(\widehat{\phi}_m-\phi_m)\rangle\rangle^2 \\
&\lesssim \sum_{l'=q_1+1}^{\infty}a_{hj'l'm}^2+\|\widehat{\phi}_m-\phi_m\|^2\sum_{l'=q_1+1}^{\infty}\left(\sum_{m''=1}^{\infty}a_{hj'l'm''}\right)^2 \\
&\lesssim u_{hj'}^2(q_1+m)^{-2\kappa}+u_{hj'}^2\|\widehat{\phi}_m-\phi_m\|^2q_1^{-2\kappa+2}.
\end{aligned}$$

To provide the upper bound of $I_{2,tmhj'}$, we start with

$$\begin{aligned}
&\frac{\sum_{t=L+1}^{n}[\zeta_{tj'l_1}\zeta_{tj'l_2}-\mathbb{E}(\zeta_{tj'l_1}\zeta_{tj'l_2})]}{n-L} \\
&=\langle\psi_{j'l_1},\langle\widehat{\Sigma}_{0,j'j'}^X-\Sigma_{0,j'j'}^X,\psi_{j'l_2}\rangle\rangle\leqslant\|\widehat{\Sigma}_{0,j'j'}^X-\Sigma_{0,j'j'}^X\|_{\mathcal{S}}=O_P\{\mathcal{M}_1^X n^{-1/2}\}.
\end{aligned}$$

Combining this result with Lemmas 1.24 and 1.25 and following the similar argument

in the proof of the upper bound of $I_{1,tmhj'}$, we obtain that, for $(h, j') \in S$,

$$
\begin{aligned}
& I_{2,tmhj'} \\
&= \sum_{l_1,l_2=q_1+1}^{\infty} \langle \psi_{j'l_1}, \langle \beta_{hj'}, \widehat{\phi}_m \rangle \rangle \langle \psi_{j'l_2}, \langle \beta_{hj'}, \widehat{\phi}_m \rangle \rangle \frac{\sum_{t=L+1}^{n} [\zeta_{tj'l_1} \zeta_{tj'l_2} - \mathbb{E}(\zeta_{tj'l_1} \zeta_{tj'l_2})]}{n-L} \\
&\leqslant \|\widehat{\Sigma}^X_{0,j'j'} - \Sigma^X_{0,j'j'}\|_{\mathcal{S}} \left\{ \sum_{l'=q_1+1}^{\infty} \langle \psi_{j'l'}, \langle \beta_{hj'}, \widehat{\phi}_m \rangle \rangle \right\}^2 = o_P(I_{1,tmhj'}).
\end{aligned}
$$

Then

$$
\begin{aligned}
\|\frac{1}{n} \widehat{\mathbf{D}}^{-1} \widehat{\mathbf{Z}}^{\mathrm{T}} \widehat{\mathbf{R}}\|_{\max}^{(q_1,q_2)} &\lesssim s \max_{1 \leqslant j \leqslant p} \sqrt{\sum_{l=1}^{q_1} \sum_{m=1}^{q_2} \{(q_1+m)^{-2\kappa} + \|\widehat{\phi}_m - \phi_m\|^2 q_1^{-2\kappa+2}\}} \\
&\lesssim s \max_{1 \leqslant j \leqslant p} \sqrt{q_1^{-2\kappa+2} + q_1^{-2\kappa+3} q_2 \max_{1 \leqslant m \leqslant q_2} \|\widehat{\phi}_m - \phi_m\|^2} \\
&= O_P\{s q_1^{-\kappa+1}\}.
\end{aligned} \tag{1.65}
$$

□

1.D Proofs of theoretical results in Section 1.4

This section is organized in the same manner as Appendix 1.C. The proofs of Theorem 1.5 and Propositions 1.5–1.9 are presented in Appendices 1.D.1 and 1.D.2, respectively, followed by supporting technical lemmas and their proofs in Appendix 1.D.3.

1.D.1 Proof of Theorem 1.5

Here $\widehat{B} \in \mathbb{R}^{pq}$ and $\widehat{\gamma} \in \mathbb{R}^d$ are the minimizer of (1.30). Then

$$
\begin{aligned}
& \frac{1}{2n} \|\mathcal{Y} - \widehat{\mathcal{X}} \widehat{D}^{-1} \widehat{B} - \mathcal{Z} \widehat{\gamma}\|^2 + \lambda_{n1} \|\widehat{B}\|_1^{(q)} + \lambda_{n2} \|\widehat{\gamma}\|_1 \\
\leqslant & \frac{1}{2n} \|\mathcal{Y} - \widehat{\mathcal{X}} \widehat{D}^{-1} B - \mathcal{Z} \gamma\|^2 + \lambda_{n1} \|B\|_1^{(q)} + \lambda_{n2} \|\gamma\|_1.
\end{aligned}
$$

Letting $\Delta = \widehat{B} - B$, $\delta = \widehat{\gamma} - \gamma$, S_1^c be the complement of S_1 in the set $\{1, \ldots, p\}$ and S_2^c be the complement of S_2 in the set $\{1, \ldots, d\}$, we have

$$\begin{aligned}
& \frac{1}{2n}\{\Delta^{\mathrm{T}}\widehat{\Omega}^{\mathrm{T}}\widehat{\Omega}\Delta + 2\Delta^{\mathrm{T}}\widehat{\Omega}^{\mathrm{T}}\mathcal{Z}\delta + \delta^{\mathrm{T}}\mathcal{Z}^{\mathrm{T}}\mathcal{Z}\delta\} \\
\leqslant\ & \frac{1}{n}(\Delta^{\mathrm{T}}\widehat{\Omega}^{\mathrm{T}} + \delta^{\mathrm{T}}\mathcal{Z}^{\mathrm{T}})(\mathcal{Y} - \widehat{\Omega}B - \mathcal{Z}\gamma) + \lambda_{n1}(\|B\|_1^{(q)} - \|B + \Delta\|_1^{(q)}) \\
& + \lambda_{n2}(\|\gamma\|_1 - \|\gamma + \delta\|_1) \\
\leqslant\ & \frac{1}{n}\Delta^{\mathrm{T}}\widehat{\Omega}^{\mathrm{T}}(\mathcal{Y} - \widehat{\Omega}B - \mathcal{Z}\gamma) + \frac{1}{n}\delta^{\mathrm{T}}\mathcal{Z}^{\mathrm{T}}(\mathcal{Y} - \widehat{\Omega}B - \mathcal{Z}\gamma) \\
& + \lambda_{n1}(\|\Delta_{S_1}\|_1^{(q)} - \|\Delta_{S_1^c}\|_1^{(q)}) + \lambda_{n2}(\|\delta_{S_2}\|_1 - \|\delta_{S_2^c}\|_1),
\end{aligned}$$

where $\widehat{\Omega} = \widehat{\mathcal{X}}\widehat{D}^{-1}$. By Propositions 1.8, 1.9 and the choice of $\lambda_n \asymp \lambda_{n1} \asymp \lambda_{n2} \geqslant 2C_0^* s_1(\mathcal{M}_{X,Z} + \mathcal{M}^{\epsilon})[q^{\alpha+2}\{\log(pq+d)/n\}^{1/2} + q^{-\kappa+1}]$, we obtain that

$$\begin{aligned}
\frac{1}{n}\Delta^{\mathrm{T}}\widehat{\Omega}^{\mathrm{T}}(\mathcal{Y} - \widehat{\Omega}B - \mathcal{Z}\gamma) &\leqslant \frac{1}{n}\|\Delta\|_1^{(q)}\|\widehat{\Omega}^{\mathrm{T}}(\mathcal{Y} - \widehat{\Omega}B - \mathcal{Z}\gamma)\|_{\max}^{(q)} \\
&\leqslant \frac{\lambda_n}{2}(\|\Delta_{S_1}\|_1^{(q)} + \|\Delta_{S_1^c}\|_1^{(q)}), \\
\frac{1}{n}\delta^{\mathrm{T}}\mathcal{Z}^{\mathrm{T}}(\mathcal{Y} - \widehat{\Omega}B - \mathcal{Z}\gamma) &\leqslant \frac{1}{n}\|\delta\|_1\|\mathcal{Z}^{\mathrm{T}}(\mathcal{Y} - \widehat{\Omega}B - \mathcal{Z}\gamma)\|_{\max} \\
&\leqslant \frac{\lambda_n}{2}(\|\delta_{S_2}\|_1 + \|\delta_{S_2^c}\|_1).
\end{aligned}$$

Combining the above results, we have

$$0 \leqslant \frac{3}{2}(\|\Delta_{S_1}\|_1^{(q)} + \|\delta_{S_2}\|_1) - \frac{1}{2}(\ \|\Delta_{S_1^c}\|_1^{(q)} + \|\delta_{S_2^c}\|_1).$$

This ensures $\|\Delta_{S_1^c}\|_1^{(q)} + \|\delta_{S_2^c}\|_1 \leqslant 3(\|\Delta_{S_1}\|_1^{(q)} + \|\delta_{S_2}\|_1)$. Then we have that

$$\|\Delta\|_1^{(q)} + \|\delta\|_1 \leqslant 4(\|\Delta_{S_1}\|_1^{(q)} + \|\delta_{S_2}\|_1) \leqslant 4(\sqrt{s_1}\|\Delta\| + \sqrt{s_2}\|\delta\|) \leqslant 4\sqrt{s_1 + s_2}(\|\Delta\| + \|\delta\|).$$

This, together with Proposition 1.5, $\|\Delta\|_1 \leqslant \sqrt{q}\|\Delta\|_1^{(q)}$ and $\tau_2^* \geqslant 64\tau_1^* q(s_1+s_2)$ implies

$$\begin{aligned}
&\frac{1}{n}\{\Delta^{\mathrm{T}}\widehat{\Omega}^{\mathrm{T}}\widehat{\Omega}\Delta + 2\Delta^{\mathrm{T}}\widehat{\Omega}^{\mathrm{T}}\mathcal{Z}\delta + \delta^{\mathrm{T}}\mathcal{Z}^{\mathrm{T}}\mathcal{Z}\delta\} \\
\geqslant&\tau_2^*(\|\Delta\|^2 + \|\delta\|^2) - \tau_1^*(\sqrt{q}\|\Delta\|_1^{(q)} + \|\delta\|_1)^2 \\
\geqslant&\frac{\tau_2^*}{2}(\|\Delta\| + \|\delta\|)^2 - \tau_1^* q(\|\Delta\|_1^{(q)} + \|\delta\|_1)^2 \\
\geqslant&\{\frac{\tau_2^*}{2} - 16\tau_1^* q(s_1+s_2)\}(\|\Delta\| + \|\delta\|)^2 \\
\geqslant&\frac{\tau_2^*}{4}(\|\Delta\| + \|\delta\|)^2.
\end{aligned}$$

This implies

$$\frac{\tau_2^*}{8}(\|\Delta\| + \|\delta\|)^2 \leqslant \frac{3\lambda_n}{2}(\|\Delta\|_1^{(q)} + \|\delta\|_1) \leqslant 6\lambda_n\sqrt{s_1+s_2}(\|\Delta\| + \|\delta\|).$$

Therefore, we obtain that

$$\|\Delta\| + \|\delta\| \lesssim \frac{\lambda_n\sqrt{s_1+s_2}}{\tau_2^*},$$

$$\|\Delta\|_1^{(q)} + \|\delta\|_1 \lesssim \frac{\lambda_n(s_1+s_2)}{\tau_2^*}.$$

Provided that $\|D^{-1}\|_{\max} \leqslant \alpha^{1/2}c_0^{-1/2}q^{\alpha/2}$, the rest can be proved in a similar way to the proof of Theorem 1.4, which shows

$$\begin{aligned}
\|\widehat{\mathcal{B}} - \mathcal{B}\|_1 + q^{\alpha/2}\|\widehat{\gamma} - \gamma\|_1 &\leqslant \|\widehat{\Psi} - \Psi\|_1^{(q)} + q^{\alpha/2}\|\widehat{\gamma} - \gamma\|_1 + o(1) \\
&\leqslant \|D^{-1}\|_{\max}\|\widehat{B} - B\|_1 + q^{\alpha/2}\|\widehat{\gamma} - \gamma\|_1 + o(1) \\
&\lesssim \frac{q^{\alpha/2}\lambda_n(s_1+s_2)}{\tau_2^*}\{1 + o(1)\}.
\end{aligned}$$

□

1.D.2 Proofs of propositions

Proof of Proposition 1.5 By Lemmas 1.16, 1.20 and 1.28, we obtain

$$\begin{aligned}
&\left\|\frac{1}{n}\mathcal{S}^{\mathrm{T}}\mathcal{S}-\frac{1}{n}\mathbb{E}\{\mathcal{S}^{\mathrm{T}}\mathcal{S}\}\right\|_{\max}\\
&=\max\left(\left\|\frac{1}{n}\mathcal{Z}^{\mathrm{T}}\mathcal{Z}-\frac{1}{n}\mathbb{E}\{\mathcal{Z}^{\mathrm{T}}\mathcal{Z}\}\right\|_{\max},\left\|\frac{1}{n}\mathcal{Z}^{\mathrm{T}}\widehat{\Omega}-\frac{1}{n}\mathbb{E}\{\mathcal{Z}^{\mathrm{T}}\Omega\}\right\|_{\max},\|\widehat{\mathbf{\Gamma}}-\mathbf{\Gamma}\|_{\max}\right)\\
&=O_P\left\{\max\left(\mathcal{M}_1^Z\sqrt{\frac{\log(d)}{n}},\mathcal{M}_1^X q^{\alpha+1}\sqrt{\frac{\log(pq)}{n}},\mathcal{M}_{X,Z}q^{\alpha+1}\sqrt{\frac{\log(pqd)}{n}}\right)\right\}\\
&=O_P\left\{\mathcal{M}_{X,Z}q^{\alpha+1}\sqrt{\frac{\log(pq+d)}{n}}\right\}.
\end{aligned}$$

Combining this with Condition 1.9 and following the similar argument in the proof of Proposition 1.4 implies Proposition 1.5. □

Proof of Proposition 1.8 Notice that

$$\frac{1}{n}\widehat{\Omega}^{\mathrm{T}}(\mathcal{Y}-\widehat{\Omega}B-\mathcal{Z}\gamma)=\frac{1}{n}\widehat{\Omega}^{\mathrm{T}}((\Omega-\widehat{\Omega})B+R+E)$$

where $\widehat{\Omega}=\widehat{\mathcal{X}}\widehat{D}^{-1}$, $B=D\Psi$ and j-th row of Ψ takes the form that $\Psi_j=\int_{\mathcal{U}}\boldsymbol{\psi}_j(u)\beta_j(u)du$. Recall that $r_t=\sum_{j=1}^{p}\sum_{l=q+1}^{\infty}\zeta_{tjl}\langle\psi_{jl},\beta_j\rangle$. Then it follows from Lemma 1.17 when $L=0$ that there exist some positive constants C_{11}^*, c_1^* and c_2^* such that

$$\begin{aligned}
\left\|\frac{1}{n}\widehat{\Omega}^{\mathrm{T}}(\Omega-\widehat{\Omega})B\right\|_{\max}^{(q)}&\leqslant\left\|\frac{1}{n}\widehat{\Omega}^{\mathrm{T}}(\Omega-\widehat{\Omega})\right\|_{\max}^{(q)}\|B\|_1^{(q)}\\
&\leqslant C_{11}^*s_1\mathcal{M}_1^X q^{\alpha+2}\sqrt{\frac{\log(pq)}{n}},
\end{aligned}$$

with probability greater than $1-c_1^*(pq)^{-c_2^*}$.

Second, it follows from Lemma 1.22 that there exist some positive constants C_{12}^*, c_1^* and c_2^* such that

$$\left\|\frac{1}{n}\widehat{\Omega}^{\mathrm{T}}R\right\|_{\max}^{(q)}\leqslant C_{12}^*s_1q^{-\kappa+1},$$

with probability greater than $1-c_1^*(pq)^{-c_2^*}$.

Third, it follows from Proposition 1.3 that there exist some positive constants C_{13}^*, c_1^*

and c_2^* such that

$$\left\|\frac{1}{n}\widehat{\Omega}^{\mathrm{T}}E\right\|_{\max}^{(q)} = \left\|\frac{1}{n}\widehat{D}^{-1}DD^{-1}\widehat{\mathcal{X}}^{\mathrm{T}}E\right\|_{\max}^{(q)} \leqslant C_{13}^*\{\mathcal{M}_1^X + \mathcal{M}^\epsilon\}q^{1/2}\sqrt{\frac{\log(pq)}{n}},$$

with probability greater than $1 - c_1^*(pq)^{-c_2^*}$.

Combining the above results, we obtain that there exist some positive constants C_{01}, c_1^* and c_2^* such that

$$\frac{1}{n}\|\widehat{\Omega}^{\mathrm{T}}(\mathcal{Y} - \widehat{\Omega}B - \mathcal{Z}\gamma)\|_{\max}^{(q)} \leqslant C_{01}s_1(\mathcal{M}_1^X + \mathcal{M}^\epsilon)\left\{q^{\alpha+2}\sqrt{\frac{\log(pq)}{n}} + q^{-\kappa+1}\right\}$$

with probability greater than $1 - c_1^*(pq)^{-c_2^*}$. □

Proof of Proposition 1.9 Notice that

$$\frac{1}{n}\mathcal{Z}^{\mathrm{T}}(\mathcal{Y} - \widehat{\Omega}B - \mathcal{Z}\gamma) = \frac{1}{n}\mathcal{Z}^{\mathrm{T}}((\Omega - \widehat{\Omega})B + R + E).$$

First, we show the deviation bound of $\frac{1}{n}\mathcal{Z}^{\mathrm{T}}(\Omega - \widehat{\Omega})B$. It follows from Lemma 1.21 and the fact that $\|\Psi_j\|_1 = \sum_{j=1}^q u_j l^{-\kappa} = O(u_j)$, for $j \in S_1$, that there exist some positive constants C_{21}^*, c_1^* and c_2^* such that

$$\begin{aligned}\left\|\frac{1}{n}\mathcal{Z}^{\mathrm{T}}(\Omega - \widehat{\Omega})B\right\|_{\max} &\leqslant \left\|\frac{1}{n}\mathcal{Z}^{\mathrm{T}}(\Omega - \widehat{\Omega})\right\|_{\max}\|B\|_1 \\ &\leqslant \left\|\frac{1}{n}\mathcal{Z}^{\mathrm{T}}(\Omega - \widehat{\Omega})\right\|_{\max}\|D\|_{\max}\|\Psi\|_1 \\ &\leqslant C_{21}^* s_1 \mathcal{M}_{X,Z} q^{\alpha+1}\sqrt{\frac{\log(pqd)}{n}},\end{aligned}$$

with probability greater than $1 - c_1^*(pqd)^{-c_2^*}$.

Second, it follows from Lemma 1.23 that there exist some positive constants C_{22}^*, c_1^* and c_2^* such that

$$\left\|\frac{1}{n}\mathcal{Z}^{\mathrm{T}}R\right\|_{\max} \leqslant C_{22}^* s_1 q^{-\kappa+1/2},$$

with probability greater than $1 - c_1^*(pqd)^{-c_2^*}$.

Third, it follows from Lemma 1.28 that there exist some positive constants C_{23}^*, c_1^* and c_2^* such that

$$\left\|\frac{1}{n}\mathcal{Z}^{\mathrm{T}}E\right\|_{\max} \leqslant C_{23}^*\{\mathcal{M}_1^Z + \mathcal{M}^\epsilon\}\sqrt{\frac{\log(d)}{n}},$$

with probability greater than $1 - c_1^*(d)^{-c_2^*}$.

Combining the above results, we obtain that there exist some positive constants C_{02}, c_1^* and c_2^* such that

$$\frac{1}{n}\|\mathcal{Z}^{\mathrm{T}}(\mathcal{Y} - \widehat{\Omega}B - \mathcal{Z}\gamma)\|_{\max} \leqslant C_{02}s_1\{\mathcal{M}_{X,Z} + \mathcal{M}^{\epsilon})\}\left\{q^{\alpha+1}\sqrt{\frac{\log(pq+d)}{n}} + q^{-\kappa+1/2}\right\}$$

with probability greater than $1 - c_1^*(pq+d)^{-c_2^*}$. □

1.D.3 Technical lemmas and their proofs

Lemma 1.20. *Suppose that Conditions 1.1–1.5 hold. Then there exist some positive constants $\widetilde{C}_{1,Z\Gamma}$, c_1^* and c_2^* such that*

$$\left\|\frac{1}{n}\mathcal{Z}^{\mathrm{T}}\widehat{\Omega} - \frac{1}{n}\mathbb{E}\{\mathcal{Z}^{\mathrm{T}}\Omega\}\right\|_{\max} \leqslant \widetilde{C}_{1,Z\Gamma}\mathcal{M}_{X,Z}q^{\alpha+1}\sqrt{\frac{\log(pqd)}{n}}$$

with probability greater than $1 - c_1^*(pqd)^{-c_2^*}$.

Proof. Note that

$$\left\|\frac{1}{n}\mathcal{Z}^{\mathrm{T}}\widehat{\Omega} - \frac{1}{n}\mathbb{E}\{\mathcal{Z}^{\mathrm{T}}\Omega\}\right\|_{\max} = \max_{\substack{1\leqslant j\leqslant p, 1\leqslant k\leqslant d\\ 1\leqslant l\leqslant q}}\left|\{\widehat{\omega}_{jl}^X\}^{-1/2}\widehat{\varrho}_{h,jkl}^{X,Z} - \{\omega_{jl}^X\}^{-1/2}\varrho_{h,jkl}^{X,Z}\right|.$$

Let $\widehat{s}_{jkl} = \left\{\omega_{jl}^X/\widehat{\omega}_{jl}^X\right\}^{1/2}$, then we obtain that

$$\begin{aligned}
&\{\widehat{\omega}_{jl}^X\}^{-1/2}\widehat{\varrho}_{h,jkl}^{X,Z} - \{\omega_{jl}^X\}^{-1/2}\varrho_{h,jkl}^{X,Z}\\
=&\widehat{s}_{jkl}\frac{\widehat{\varrho}_{h,jkl}^{X,Z} - \varrho_{h,jkl}^{X,Z}}{\{\omega_{jl}^X\}^{1/2}} + \frac{\{\omega_{jl}^X\}^{1/2} - \{\widehat{\omega}_{jl}^X\}^{1/2}}{\{\widehat{\omega}_{jl}^X\}^{1/2}}\frac{\varrho_{h,jkl}^{X,Z}}{\{\omega_{jl}^X\}^{1/2}}.
\end{aligned}$$

It follows Propositions 1.1, 1.6 and the fact $\mathbb{E}(\zeta_{tjl}Z_{tk}) \leqslant \sigma_k^Z\{\omega_{jl}^X\}^{1/2}$ that there exist some positive constants $\widetilde{C}_{1,Z\Gamma}, c_1^*$ and c_2^* such that

$$\left\|\frac{1}{n}\mathcal{Z}^{\mathrm{T}}\widehat{\Omega} - \frac{1}{n}\mathbb{E}\{\mathcal{Z}^{\mathrm{T}}\Omega\}\right\|_{\max} \leqslant \widetilde{C}_{1,Z\Gamma}\mathcal{M}_{X,Z}q^{\alpha+1}\sqrt{\frac{\log(pqd)}{n}}$$

with probability greater than $1 - c_1^*(pqd)^{-c_2^*}$. □

Lemma 1.21. *Suppose that Conditions 1.1–1.5 hold. Then there exist some positive constants* $\widetilde{C}_{2,Z\Gamma}$, c_1^* *and* c_2^* *such that*

$$\left\|\frac{1}{n}\mathcal{Z}^{\mathrm{T}}(\Omega-\widehat{\Omega})\right\|_{\max} \leqslant \widetilde{C}_{2,Z\Gamma}\mathcal{M}_{X,Z}q^{\alpha+1}\sqrt{\frac{\log(pqd)}{n}}$$

with probability greater than $1-c_1^*(pqd)^{-c_2^*}$.

Proof. We first consider $\|\frac{1}{n}\mathcal{Z}^{\mathrm{T}}\Omega-\frac{1}{n}\mathbb{E}\{\mathcal{Z}^{\mathrm{T}}\Omega\}\|_{\max}$. By (1.42) in Appendix 1.B.2, we obtain that

$$\begin{aligned}&\max_{j,k,m}\frac{(n-L)^{-1}\sum_{t=L+1}^{n}Z_{(t-h)j}\zeta_{tkm}}{\sqrt{\omega_{km}^X}}-\frac{\mathbb{E}(Z_{(t-h)j}\zeta_{tkm})}{\sqrt{\omega_{km}^X}}\\&=\max_{j,k,m}\frac{\langle\widehat{\Sigma}_{h,jk}^{Z,X},\psi_{km}\rangle-\langle\Sigma_{h,jk}^{Z,X},\psi_{km}\rangle}{\sqrt{\omega_{km}^X}}=O_P\{\mathcal{M}_{X,Z}\sqrt{\frac{\log(pqd)}{n}}\}.\end{aligned}$$

This, together with Lemma 1.20, implies that

$$\begin{aligned}\left\|\frac{1}{n}\mathcal{Z}^{\mathrm{T}}(\Omega-\widehat{\Omega})\right\|_{\max}&\leqslant\left\|\frac{1}{n}\mathcal{Z}^{\mathrm{T}}\Omega-\frac{1}{n}\mathbb{E}\{\mathcal{Z}^{\mathrm{T}}\Omega\}\right\|_{\max}+\left\|\frac{1}{n}\mathcal{Z}^{\mathrm{T}}\widehat{\Omega}-\frac{1}{n}\mathbb{E}\{\mathcal{Z}^{\mathrm{T}}\Omega\}\right\|_{\max}\\&=O_P\left\{\mathcal{M}_{X,Z}q^{\alpha+1}\sqrt{\frac{\log(pqd)}{n}}\right\}.\end{aligned}$$

□

Lemma 1.22. *Suppose that Conditions 1.1–1.5 and 1.8 hold. Then there exist some positive constants* C_{R1}, c_1^* *and* c_2^* *such that*

$$\|n^{-1}\widehat{\Omega}^{\mathrm{T}}R\|_{\max}^{(q)}\leqslant C_{R1}s_1q^{-\kappa+1}$$

with probability greater than $1-c_1^*(pq)^{-c_2^*}$.

Proof. This lemma can be proved in a similar way to Lemma 1.19 and hence the proof is omitted here. □

Lemma 1.23. *Suppose that Conditions 1.1–1.5 and 1.8 hold. Then there exist some positive constants* C_{R2}, c_1^* *and* c_2^* *such that*

$$\|n^{-1}\mathcal{Z}^{\mathrm{T}}R\|_{\max}\leqslant C_{R2}s_1q^{-\kappa+1/2}$$

with probability greater than $1 - c_1^*(pqd)^{-c_2^*}$.

Proof. This lemma can be proved in a similar way to Lemma 1.19 and hence the proof is omitted here. □

1.E Existing results for sub-Gaussian (functional) linear processes

For ease of reference, we present some useful existing results in Guo and Qiao (2022), including non-asymptotic error bounds on estimated covariance matrix function, estimated eigenpairs and estimated (auto)covariance between estimated FPC scores. By Theorem 1.1, we can easily extend these results from Gaussian functional time series to accommodate sub-Gaussian functional linear processes in Lemmas 1.24–1.27. Moreover, we also present non-asymptotic error bounds on estimated (cross-)covariance matrix in Basu and Michailidis (2015) to accommodate sub-Gaussian linear processes in Lemma 1.28.

Lemma 1.24. *Suppose that Conditions 1.1, 1.3 and 1.4 hold for sub-Gaussian linear process* $\{\mathbf{X}_t(\cdot)\}_{t\in\mathbb{Z}}$. *Then there exists some universal constant* $\tilde{c}_1 > 0$ *such that for any* $\eta > 0$ *and each* $j, k = 1, \ldots, p$,

$$P\left\{\|\widehat{\Sigma}^X_{0,jk} - \Sigma^X_{0,jk}\|_{\mathcal{S}} > 2\omega_0^X \mathcal{M}_1^X \eta\right\} \leqslant 4\exp\{-\tilde{c}_1 n \min(\eta^2, \eta)\}.$$

Proof. This lemma follows directly from Theorem 1.1 and Theorem 2 of Guo and Qiao (2022) and hence the proof is omitted here. □

Lemma 1.25. *Suppose that Conditions 1.1, 1.3, 1.4 and 1.5 hold for sub-Gaussian linear process* $\{\mathbf{X}_t(\cdot)\}_{t\in\mathbb{Z}}$. *Let* M *be a positive integer possibly depending on* (n, p). *If* $n \gtrsim \log(pM)M^{4\alpha+2}(M_1^X)^2$, *then there exist some constants* $\tilde{c}_2, \tilde{c}_3 > 0$ *such that, with probability greater than* $1 - \tilde{c}_2(pM)^{-\tilde{c}_3}$, *the estimates* $\{\widehat{\omega}^X_{jl}\}$ *and* $\{\widehat{\psi}_{jl}\}$ *satisfy*

$$\max_{1\leqslant j\leqslant p, 1\leqslant l\leqslant M}\left\{\left|\frac{\widehat{\omega}^X_{jl} - \omega^X_{jl}}{\omega^X_{jl}}\right| + \left\|\frac{\widehat{\psi}_{jl} - \psi_{jl}}{l^{\alpha+1}}\right\|\right\} \lesssim \mathcal{M}_1^X\sqrt{\frac{\log(pM)}{n}}. \tag{1.66}$$

Proof. This lemma follows directly from Theorem 1.1 and Theorem 3 of Guo and Qiao (2022) and hence the proof is omitted here. □

Lemma 1.26. *Suppose that conditions in Lemma 1.25 hold. Then there exists some universal constant $\tilde{c}_4 > 0$ such that for each $j = 1, \ldots, p, l = 1, \ldots, d_j$, any given function $g \in \mathbb{H}$ and $\eta > 0$,*

$$P\left\{\left|\langle\widehat{\psi}_{jl} - \psi_{jl}, g\rangle\right| \geqslant \tilde{\rho}_1 \|g^{-jl}\|_\omega \mathcal{M}_1^X \{\omega_{jl}^X\}^{1/2} l^{\alpha+1}\eta + \tilde{\rho}_2 \|g\| \{\mathcal{M}_1^X\}^2 l^{2(\alpha+1)}\eta^2\right\}$$
$$\leqslant 8\exp\left\{-\tilde{c}_4 n \min(\eta^2, \eta)\right\} + 4\exp\left\{-\tilde{c}_4 \{\mathcal{M}_1^X\}^{-2} n l^{-2(\alpha+1)}\right\},$$

where $g(\cdot) = \sum_{l=1}^{\infty} g_{jl}\psi_{jl}(\cdot)$, $\|g^{-jl}\|_\omega = \left(\sum_{l':l'\neq l} \omega_{jl'} g_{jl'}^2\right)^{1/2}$, $\tilde{\rho}_1 = 2c_0^{-1}\omega_0^X$ and $\tilde{\rho}_2 = 4(6+2\sqrt{2})c_0^{-2}\{\omega_0^X\}^2$ with $c_0 \leqslant 4\mathcal{M}_1^X \omega_0^X l^{\alpha+1}$.

Proof. This lemma follows directly from Theorem 1.1 and Lemma 3 of Guo and Qiao (2022) and hence the proof is omitted here. □

Lemma 1.27. *Suppose that conditions in Lemma 1.25 hold. Let M be a positive integer possibly depending on (n, p). If $n \gtrsim \log(pM)M^{4\alpha+2}(\mathcal{M}_1^X)^2$, then there exist some constants $\tilde{c}_5, \tilde{c}_6 > 0$ such that, with probability greater than $1 - \tilde{c}_5(pM)^{-\tilde{c}_6}$, the estimates $\{\widehat{\sigma}_{h,jklm}^X\}$ satisfies*

$$\max_{\substack{1\leqslant j,k\leqslant p \\ 1\leqslant l,m\leqslant M}} \frac{\left|\widehat{\sigma}_{h,jklm}^X - \sigma_{h,jklm}^X\right|}{(l\vee m)^{\alpha+1}\sqrt{\omega_{jl}^X \omega_{km}^X}} \lesssim \mathcal{M}_1^X \sqrt{\frac{\log(pM)}{n}}. \tag{1.67}$$

Proof. This lemma follows directly from Theorem 1.1 and Theorem 4 of Guo and Qiao (2022) and hence the proof is omitted here. □

Lemma 1.28. *(i)Suppose $\{\mathbf{Z}_t\}$ is from d-dimensional sub-Gaussian linear process with absolute summable coefficients and bounded $\mathcal{M}^Z$. For any given vector $\boldsymbol{\nu} \in \mathbb{R}_0^d$ with $\|\boldsymbol{\nu}\|_0 \leqslant k$ $(k = 1, \ldots, d)$, denote $\mathcal{M}(\boldsymbol{f}_Z, \boldsymbol{\nu}) = 2\pi \cdot ess\sup_{\theta\in[-\pi,\pi]} \boldsymbol{\nu}^{\mathrm{T}} \boldsymbol{f}_Z \boldsymbol{\nu}$. Then there exists some constants $c, \tilde{c}_{16}, \tilde{c}_{17} > 0$ such that for any $\eta > 0$,*

$$P\left\{\left|\boldsymbol{\nu}^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_0^Z - \boldsymbol{\Sigma}_0^Z)\boldsymbol{\nu}\right| > \mathcal{M}(\boldsymbol{f}_Z, \boldsymbol{\nu})\eta\right\} \leqslant 2\exp\left\{-cn\min\left(\eta^2, \eta\right)\right\},$$

and

$$P\left\{\left|\frac{\boldsymbol{\nu}^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_0^Z - \boldsymbol{\Sigma}_0^Z)\boldsymbol{\nu}}{\boldsymbol{\nu}^{\mathrm{T}}\boldsymbol{\Sigma}_0^Z\boldsymbol{\nu}}\right| > \mathcal{M}_k^Z\eta\right\} \leqslant 2\exp\left\{-cn\min\left(\eta^2, \eta\right)\right\}.$$

In particular, with probability greater than $1-\tilde{c}_{16}(d)^{-\tilde{c}_{17}}$,

$$\max_{1\leqslant j,k\leqslant d}|\widehat{\Sigma}^{Z}_{0,jk}-\Sigma^{Z}_{0,jk}|\lesssim \mathcal{M}_1^{Z}\sqrt{\frac{\log(d)}{n}}.$$

(ii)Suppose $\{\epsilon_t\}$ *is from sub-Gaussian linear process with absolute summable coefficients, bounded* $\mathcal{M}^{\epsilon}$ *and independent of* $\{\mathbf{Z}_t\}$*. Then there exist some positive constants* $\tilde{c}_{18},\tilde{c}_{19}$ *such that with probability greater than* $1-\tilde{c}_{18}(d)^{-\tilde{c}_{19}}$,

$$\max_{1\leqslant j\leqslant d}\left|\sum_{t=1}^{n}Z_{tj}\epsilon_t/n\right|\lesssim(\mathcal{M}_1^{Z}+\mathcal{M}^{\epsilon})\sqrt{\frac{\log(d)}{n}}.$$

Proof. This lemma can be proved in similar way to Proposition 2.4 of Basu and Michailidis (2015) and be extended to sub-Gaussian linear process setting following the similar techniques used in the proof of Theorem 1.1. □

1.F Matrix representation of model (1.1)

It follows from the Karhunen-Loève expansion that model (1.1) can be rewritten as

$$\sum_{m=1}^{\infty}\xi_{tm}\phi_m(v)=\sum_{h=0}^{L}\sum_{j=1}^{p}\sum_{l=1}^{\infty}\langle\psi_{jl}(u),\beta_{hj}(u,v)\rangle\zeta_{(t-h)jl}+\epsilon_t(v),$$

This, together with orthonormality of $\{\phi_m(\cdot)\}_{m\geqslant 1}$, implies that

$$\xi_{tm}=\sum_{h=0}^{L}\sum_{j=1}^{p}\sum_{l=1}^{q_{1j}}\langle\langle\psi_{jl}(u),\beta_{hj}(u,v)\rangle,\phi_m(v)\rangle\zeta_{(t-h)jl}+r_{tm}+\epsilon_{tm},$$

where $r_{tm}=\sum_{h=0}^{L}\sum_{j=1}^{p}\sum_{l=q_{1j}+1}^{\infty}\langle\langle\psi_{jl}(u),\beta_{hj}(u,v)\rangle,\phi_m(v)\rangle\zeta_{(t-h)jl}$ and $\epsilon_{tm}=\langle\phi_m,\epsilon_t\rangle$ for $m=1,\ldots,q_2$, represent the approximation and random errors, respectively. Let $\mathbf{r}_t=(r_{t1},\ldots,r_{tq_2})^{\mathrm{T}}$ and $\boldsymbol{\epsilon}_t=(\epsilon_{t1},\ldots,\epsilon_{tq_2})^{\mathrm{T}}$. Let $\mathbf{R}$ and $\mathbf{E}$ be $(n-L)\times q_2$ matrices whose row vectors are formed by $\{\mathbf{r}_t,t=L+1,\ldots,n\}$ and $\{\boldsymbol{\epsilon}_t,t=L+1,\ldots,n\}$ respectively. Then (1.1) can be represented in the matrix form of (1.23).

Chapter 2

Adaptive Functional Thresholding for Sparse Covariance Function Estimation in High Dimensions

2.1 Introduction

The covariance function estimation plays an important role in functional data analysis, while existing methods are restricted to data with a single or small number of random functions. Recent advances in technology have made multivariate or even high-dimensional functional datasets increasingly common in various applications: for example, time-course gene expression data in genomics (Storey et al., 2005), air pollution data in environmental studies (Kong et al., 2016) and different types of brain imaging data in neuroscience (Zhu et al., 2016; Li and Solea, 2018). Under such scenario, suppose we observe n independent samples $\mathbf{X}_i(\cdot) = \{X_{i1}(\cdot), \ldots, X_{ip}(\cdot)\}^{\mathrm{T}}$ $(i = 1, \ldots, n)$ defined on a compact interval $\mathcal{U}$ with covariance function $\mathbf{\Sigma}(u, v) = \{\Sigma_{jk}(u, v)\}_{p\times p} = \mathrm{Cov}\{\mathbf{X}_i(u), \mathbf{X}_i(v)\}$[1] for $u, v \in \mathcal{U}$, which can also been seen as a matrix of marginal- and cross-covariance functions. Besides being of interest in itself, an estimation of $\mathbf{\Sigma}(\cdot, \cdot)$ is useful for numerous procedures including multivariate functional principal components analysis (Happ and Greven, 2018), classification by functional linear discriminant analysis (Park et al., 2022) and recovering functional graphical models (Qiao et al., 2019).

[1]To clarify, we reuse the symbols that are presented in the previous chapter and will redefine the notation in Chapters 2 and 3, respectively.

This chapter focuses on estimating $\mathbf{\Sigma}$ under high-dimensional scaling, where p can be comparable to, or even larger than n. In this setting, the sample covariance function

$$\widehat{\mathbf{\Sigma}}(u,v) = \{\widehat{\Sigma}_{jk}(u,v)\}_{p\times p} = \frac{1}{n-1}\sum_{i=1}^{n}\{\mathbf{X}_i(u) - \bar{\mathbf{X}}(u)\}\{\mathbf{X}_i(v) - \bar{\mathbf{X}}(v)\}^{\mathrm{T}}, \quad u,v \in \mathcal{U},$$

where $\bar{\mathbf{X}}(\cdot) = n^{-1}\sum_{i=1}^{n}\mathbf{X}_i(\cdot)$, performs poorly, and some lower-dimensional structural assumptions need to be imposed to estimate $\mathbf{\Sigma}(u,v)$ consistently. In contrast to extensive work on estimating high-dimensional sparse covariance matrices, for example, Bickel and Levina (2008); Rothman et al. (2009); Cai and Liu (2011); Chen and Leng (2016); Avella-Medina et al. (2018) and Wang et al. (2021), research on sparse covariance function estimation in high dimensions remains largely unaddressed in literature.

In this chapter, we consider estimating sparse covariance functions via adaptive functional thresholding. Note that the words "sparse" and "sparsity" are used to describe the non-zero structure of the high-dimensional covariance function. To achieve this, we introduce a new class of functional thresholding operators that combine functional versions of thresholding and shrinkage based on the Hilbert–Schmidt norm of functions, and develop an adaptive functional thresholding procedure on $\widehat{\mathbf{\Sigma}}(\cdot,\cdot)$ using entry-dependent functional thresholds that automatically adapt to the variability of $\widehat{\Sigma}_{jk}(\cdot,\cdot)$'s. To provide theoretical guarantees of our method under high-dimensional scaling, it is essential to develop standardized concentration results taking into account the variability adjustment. Compared with the adaptive thresholding for non-functional data (Cai and Liu, 2011), the intrinsic infinite-dimensionality of each $X_{ij}(\cdot)$ leads to a substantial rise in the complexity of sparsity modeling and theoretical analysis, as one needs to rely on some functional norm of standardized $\widehat{\Sigma}_{jk}$'s, for example, the Hilbert–Schmidt norm, to enforce the functional sparsity in $\widehat{\mathbf{\Sigma}}$ and tackle more technical challenges for standardized processes within an abstract Hilbert space.

There are many applications of our proposed sparse covariance function estimation method in neuroimaging analysis, where brain signals are measured over time at a large number of regions of interest (ROIs) for individuals. Examples include the brain-computer interface classification (Lotte et al., 2018) and the brain functional connectivity identification (Rogers et al., 2007). Traditional neuroimaging analysis models brain signals for each subject as multivariate random variables, where each ROI is represented by a random variable, and hence the covariance/correlation matrices of interest are estimated by treating the time-course data of each ROI as repeated observations. However, due to the non-stationary and dynamic features of

signals (Chang and Glover, 2010), the strategy of averaging over time fails to characterize the time-varying structure leading to the loss of information in the original space. To overcome these drawbacks, we follow recent proposals to model signals directly as multivariate random functions with each ROI represented by a random function (Li and Solea, 2018; Qiao et al., 2019; Solea and Li, 2022; Hu and Yao, 2021). The identified functional sparsity pattern in our estimate of $\mathbf{\Sigma}$ can be used to recover the functional connectivity network among different ROIs, which is illustrated using examples of functional magnetic resonance imaging (fMRI) datasets in Section 2.5 and Section 2.C.2 of the Appendix.

This chapter makes useful contributions at multiple fronts. On the method side, it generalizes the thresholding/sparsity concept in multivariate statistics to functional settings and offers a novel adaptive functional thresholding proposal to handle the heteroscedastic problem of the sparse covariance function estimation by incorporating variance effects of individual entries of the sample covariance function into functional thresholding. Such procedure also provides an alternative of identifying correlation-based functional connectivity with no need to specify the correlation function, the estimation of which poses additional challenges as the inverses of marginal-covariance functions are unbounded. On the theory side, we show that adaptive functional thresholding estimators enjoy the convergence and support recovery properties under a high-dimensional regime. The proof relies on tools from empirical process theory due to the infinite-dimensional nature of functional data and some novel standardized concentration bounds in the Hilbert–Schmidt norm to deal with issues of high-dimensionality and variance adjustment. Our theoretical results and adopted techniques are general, and can be applied to other settings in high-dimensional functional data analysis. Empirically, we demonstrate the uniform superiority of adaptive functional thresholding estimators over the universal functional thresholding estimators via both simulation studies and the functional connectivity analysis of two neuroimaging datasets.

The remainder of this chapter is organized as follows. In Section 2.2, we introduce a class of functional thresholding operators, based on which the adaptive functional thresholding of the sample covariance function is proposed. Section 2.3 presents convergence and support recovery analysis of the adaptive functional thresholding estimator. In Sections 2.4 and 2.5, we examine the finite-sample performance of the proposed method through simulations and the functional connectivity analysis of a neuroimaging dataset, respectively. Section 2.6 concludes this chapter by discussing three potential extensions. All technical proofs are relegated to the Appendix.

2.2 Methodology

We begin by introducing some notation. Let $L_2(\mathcal{U})$ denotes a Hilbert space of square integrable functions defined on $\mathcal{U}$ and $\mathbb{S} = L_2(\mathcal{U}) \otimes L_2(\mathcal{U})$, where $\otimes$ is the Kronecker product. For any $K \in \mathbb{S}$, we denote its Hilbert–Schmidt norm by $\|K\|_{\mathcal{S}} = \{\int\int K(u,v)^2 du dv\}^{1/2}$. With the aid of Hilbert–Schmidt norm, for any regularization parameter $\lambda \geqslant 0$, we first define a class of functional thresholding operators $s_\lambda : \mathbb{S} \to \mathbb{S}$ that satisfy the following conditions:

(i) $\|s_\lambda(Z)\|_{\mathcal{S}} \leqslant c\|Y\|_{\mathcal{S}}$ for all Z and $Y \in \mathbb{S}$ that satisfy $\|Z - Y\|_{\mathcal{S}} \leqslant \lambda$ and some $c > 0$;

(ii) $\|s_\lambda(Z)\|_{\mathcal{S}} = 0$ for $\|Z\|_{\mathcal{S}} \leqslant \lambda$;

(iii) $\|s_\lambda(Z) - Z\|_{\mathcal{S}} \leqslant \lambda$ for all $Z \in \mathbb{S}$.

Our proposed functional thresholding operators can be viewed as the functional generalization of thresholding operators (Cai and Liu, 2011). Instead of a simple pointwise extension of such thresholding operators under functional domain, we advocate a global thresholding rule based on the Hilbert–Schmidt norm of functions that encourages the functional sparsity, in the sense that $s_\lambda(Z)(u,v) = 0$, for all $u, v \in \mathcal{U}$, if $\|Z\|_{\mathcal{S}} \leqslant \lambda$ under condition (ii). Condition (iii) limits the amount of (global) functional shrinkage in the Hilbert–Schmidt norm to be no more than λ.

Conditions (i)–(iii) are satisfied by functional versions of some commonly adopted thresholding rules, which are introduced as solutions to the following penalized quadratic loss problem with various penalties:

$$s_\lambda(Z) = \arg\min_{\theta \in \mathbb{S}} \left\{ \frac{1}{2}\|\theta - Z\|_{\mathcal{S}}^2 + p_\lambda(\theta) \right\} \tag{2.1}$$

with $p_\lambda(\theta) = \tilde{p}_\lambda(\|\theta\|_{\mathcal{S}})$ being a penalty function of $\|\theta\|_{\mathcal{S}}$ to enforce the functional sparsity.

The soft functional thresholding rule results from solving (2.1) with an ℓ_1/ℓ_2 type of penalty, $p_\lambda(\theta) = \lambda\|\theta\|_{\mathcal{S}}$, and takes the form of $s_\lambda^{\mathrm{s}}(Z) = Z(1 - \lambda/\|Z\|_{\mathcal{S}})_+$, where $(x)_+ = \max(x, 0)$ for $x \in \mathbb{R}$. This rule can be viewed as a functional generalization of the group lasso solution under the multivariate setting (Yuan and Lin, 2006). To solve (2.1) with an ℓ_0/ℓ_2 type of penalty, $p_\lambda(\theta) = 2^{-1}\lambda^2 I(\|\theta\|_{\mathcal{S}} \neq 0)$, we obtain hard functional threhsolding rule as $ZI(\|Z\|_{\mathcal{S}} \geqslant \lambda)$, where $I(\cdot)$ is an indicator function. As a comparison, soft functional thresholding corresponds to the maximum

amount of functional shrinkage allowed by condition (iii), whereas no shrinkage results from hard functional thresholding. Taking the compromise between soft and hard functional thresholding, we next propose functional versions of SCAD (Fan and Li, 2001) and adaptive lasso (Zou, 2006) thresholding rules. With a SCAD penalty (Fan and Li, 2001) operating on $\|\cdot\|_{\mathcal{S}}$ instead of $|\cdot|$ for the univariate scalar case, SCAD functional thresholding $s_{\lambda}^{\mathrm{SC}}(Z)$ is the same as soft functional thresholding if $\|Z\|_{\mathcal{S}} < 2\lambda$, and equals $Z\{(a-1) - a\lambda/\|Z\|_{\mathcal{S}}\}/(a-2)$ for $\|Z\|_{\mathcal{S}} \in [2\lambda, a\lambda]$ and Z if $\|Z\|_{\mathcal{S}} > a\lambda$, where $a > 2$. Analogously, adaptive lasso functional thresholding rule is $s_{\lambda}^{\mathrm{AL}}(Z) = Z(1 - \lambda^{\eta+1}/\|Z\|_{\mathcal{S}}^{\eta+1})_+$ with $\eta \geqslant 0$.

Our proposed functional generalizations of soft, SCAD and adaptive lasso thresholding rules can be checked to satisfy conditions (i)–(iii), see Appendix 2.B for details. To present a unified theoretical analysis, we focus on functional thresholding operators $s_{\lambda}(Z)$ satisfying conditions (i)–(iii). It is worth noting that, although the hard functional thresholding does not satisfy condition (i), theoretical results in Section 2.3 still hold for hard functional thresholding estimators under similar conditions with corresponding proofs differing slightly.

In general, conditions (i)–(iii) are satisfied by a number of solutions to (2.1), where the presence of $\|\cdot\|_{\mathcal{S}}$ in both the loss and various penalty functions leads to functional thresholding rules as functions of $\|Z\|_{\mathcal{S}}$. Such connection demonstrates the rationality of imposing Hilbert–Schmidt-norm based conditions (i)–(iii). For examples of functional data with some local spikes, one may suggest another class of functional thresholding operators $\tilde{s}_{\lambda}(Z)$ satisfying three supremum-norm based conditions analogous to conditions (i)–(iii), where, for any $K \in \mathbb{S}$, we denote its supremum norm by $\|K\|_{\infty} = \sup_{u,v\in\mathcal{U}} |K(u,v)|$. In this case, $\tilde{s}_{\lambda}(Z)$ can not be directly derived as the solution to (2.1) with $p_{\lambda}(\theta) = \tilde{p}_{\lambda}(\|\theta\|_{\infty})$. However, by substituting $\|\cdot\|_{\mathcal{S}}$ in $s_{\lambda}^{\mathrm{S}}(Z)$, $s_{\lambda}^{\mathrm{SC}}(Z)$ and $s_{\lambda}^{\mathrm{AL}}(Z)$ with $\|\cdot\|_{\infty}$, the corresponding supremum-norm based functional thresholding rules can be presented and checked to satisfy three conditions for $\tilde{s}_{\lambda}(Z)$ in a similar fashion. To study theoretical properties analogous to Theorems 2.1 and 2.2 in Section 2.3, the main challenge is to establish concentration bounds on some standardized processes in the supremum norm, where our tools and results in Appendix 2.A can be applied accordingly. In this regard, the $\|\cdot\|_{\mathcal{S}}$ that we adopt in $s_{\lambda}(Z)$ is not necessarily the unique choice, but serves as the building block for the sparse covariance function estimation problem.

We now discuss our estimation procedure based on $s_{\lambda}(Z)$. As the variance of $\widehat{\Sigma}_{jk}(u,v)$ depends on the distribution of $\{X_{ij}(u), X_{ik}(v)\}$ through higher-order moments, which is intrinsically a heteroscedastic problem, it is more desirable to use entry-dependent functional thresholds that automatically takes into account the variability of $\widehat{\Sigma}_{jk}$'s.

To achieve this, define the variance factors $\Theta_{jk}(u,v) = \mathrm{Var}\big([X_{ij}(u)-E\{X_{ij}(u)\}][X_{ik}(v)-E\{X_{ik}(v)\}]\big)$ with corresponding estimators

$$\widehat{\Theta}_{jk}(u,v) = \frac{1}{n}\sum_{i=1}^{n}\Big[\{X_{ij}(u)-\bar{X}_j(u)\}\{X_{ik}(v)-\bar{X}_k(v)\}-\widehat{\Sigma}_{jk}(u,v)\Big]^2 \quad (j,k=1,\ldots,p).$$

Then the adaptive functional thresholding estimator $\widehat{\mathbf{\Sigma}}_{\mathrm{A}} = \{\widehat{\Sigma}^{\mathrm{A}}_{jk}(\cdot,\cdot)\}_{p\times p}$ is defined by

$$\widehat{\Sigma}^{\mathrm{A}}_{jk} = \widehat{\Theta}^{1/2}_{jk} \times s_\lambda\left(\frac{\widehat{\Sigma}_{jk}}{\widehat{\Theta}^{1/2}_{jk}}\right), \tag{2.2}$$

which uses a single threshold level to functionally threshold standardized entries, $\widehat{\Sigma}_{jk}/\widehat{\Theta}^{1/2}_{jk}$ for all j,k, resulting in entry-dependent functional thresholds for $\widehat{\Sigma}_{jk}$'s. The selection of optimal λ is of practical importance and will be discussed in details in Section 2.4.

An alternative approach to estimate $\mathbf{\Sigma}$ is the universal functional thresholding estimator

$$\widehat{\mathbf{\Sigma}}_{\mathrm{U}} = \{\widehat{\Sigma}^{\mathrm{U}}_{jk}(\cdot,\cdot)\}_{p\times p} \quad \text{with} \quad \widehat{\Sigma}^{\mathrm{U}}_{jk} = s_\lambda\big(\widehat{\Sigma}_{jk}\big),$$

where a universal threshold level is used for all entries. In a similar spirit to Rothman et al. (2009), the consistency of $\widehat{\mathbf{\Sigma}}_{\mathrm{U}}$ requires the assumption that marginal-covariance functions are uniformly bounded in the nuclear norm, that is $\max_j \|\Sigma_{jj}\|_{\mathcal{N}} \leqslant M$, where $\|\Sigma_{jj}\|_{\mathcal{N}} = \int_{\mathcal{U}} \Sigma_{jj}(u,u)du$. However, intuitively speaking, such universal method does not perform well when nuclear norms vary over a wide range, or even fails when the uniform boundedness assumption is violated. Section 2.4 provides some empirical evidence to support this intuition.

2.3 Theoretical properties

We begin with some notation. For a random variable W, define the Orlicz norm $\|W\|_\psi = \inf\big\{c > 0 : E[\psi(|W|/c)] \leqslant 1\big\}$, where $\psi : [0,\infty) \to [0,\infty)$ is a non-decreasing, non-zero convex function with $\psi(0) = 0$ and the norm takes the value ∞ if no finite c exists for which $E[\psi(|W|/c)] \leqslant 1$. Denote $\psi_k(x) = \exp(x^k) - 1$ for $k \geqslant 1$. Let the packing number $D(\epsilon, d)$ be the maximal number of points that can fit in the compact interval $\mathcal{U}$ while maintaining a distance greater than ϵ between all points with respect to the semimetric d. We refer to Chapter 8 of Kosorok (2008) for further explanations. For $\{X_{ij}(u) : u \in \mathcal{U}, i = 1,\ldots,n, j = 1,\ldots,p\}$, define the standardized processes by $Y_{ij}(u) = [X_{ij}(u) - E\{X_{ij}(u)\}]/\sigma_j(u)^{1/2}$, where $\sigma_j(u) = \Sigma_{jj}(u,u)$.

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

Condition 2.1. *(i) For each i and j, $Y_{ij}(\cdot)$ is a separable stochastic process with the semimetric $d_j(u,v) = \|Y_{1j}(u) - Y_{1j}(v)\|_{\psi_2}$ for $u, v \in \mathcal{U}$; (ii) For some $u_0 \in \mathcal{U}$, $\max_{1\leqslant j\leqslant p} \|Y_{1j}(u_0)\|_{\psi_2}$ is bounded.*

Condition 2.2. *The packing numbers $D(\epsilon, d_j)$'s satisfy $\max_{1\leqslant j\leqslant p} D(\epsilon, d_j) \leqslant C\epsilon^{-r}$ for some constants $C, r > 0$ and $\epsilon \in (0, 1]$.*

Condition 2.3. *For some constant $\tau > 0$, $\min_{j,k} \inf_{u,v\in\mathcal{U}} Var\{Y_{1j}(u)Y_{1k}(v)\} \geqslant \tau$.*

Condition 2.4. *The pair (n, p) satisfies $\log p/n^{1/4} \to 0$ as n and $p \to \infty$.*

Conditions 2.1 and 2.2 are standard to characterize the modulus of continuity of sub-Gaussian processes $Y_{ij}(\cdot)$'s, as described in Chapter 8 of Kosorok (2008). These conditions also imply that there exist some positive constants C_0 and η such that $E[\exp(t\|Y_{1j}\|^2)] \leqslant C_0$ for all $|t| \leqslant \eta$ and j with $\|Y_{1j}\| = \{\int_{\mathcal{U}} Y_{1j}(u)^2 du\}^{1/2}$, which plays a crucial role in our proof when applying concentration inequalities within Hilbert space. Condition 2.3 restricts the variances of $Y_{ij}(u)Y_{ik}(v)$'s to be uniformly bounded away from zero so that they can be well estimated. It also facilitates the development of some standardized concentration results. This condition excludes the case of a Brownian motion $X_{ij}(\cdot)$ starting at 0 for some j. However, replacing $X_{ij}(\cdot)$ with a contaminated process $X_{ij}(\cdot) + \xi_{ij}$, where ξ_{ij}'s are independent from a normal distribution with zero mean and a small variance and are independent of $X_{ij}(\cdot)$'s, Condition 2.3 is fulfilled while the cross-covariance structure in Σ remains the same in the sense that $\mathrm{Cov}\{X_{ij}(u) + \xi_{ij}, X_{ik}(v)\} = \mathrm{Cov}\{X_{ij}(u), X_{ik}(v)\}$ for $k \neq j$ and $u, v \in \mathcal{U}$. Condition 2.4 allows the high-dimensional case, where p can diverge at some exponential rate as n increases.

In the following, we establish the convergence rate of the adaptive functional thresholding estimator $\widehat{\mathbf{\Sigma}}_{\mathrm{A}}$ over a large class of "approximately sparse" covariance functions defined by

$$\mathcal{C}(q, s_0(p); \mathcal{U}) = \Big\{\mathbf{\Sigma} : \mathbf{\Sigma} \succeq 0, \max_{1\leqslant j\leqslant p} \sum_{k=1}^{p} \|\sigma_j\|_\infty^{(1-q)/2} \|\sigma_k\|_\infty^{(1-q)/2} \|\Sigma_{jk}\|_{\mathcal{S}}^q \leqslant s_0(p)\Big\},$$

for some $0 \leqslant q < 1$, where $\|\sigma_j\|_\infty = \sup_{u\in\mathcal{U}} \sigma_j(u)$ and $\mathbf{\Sigma} \succeq 0$ means that $\mathbf{\Sigma} = \{\Sigma_{jk}(\cdot,\cdot)\}_{p\times p}$ is positive semidefinite, that is $\sum_{j,k} \int\int \Sigma_{jk}(u,v) a_j(u) a_k(v) du dv \geqslant 0$ for any $a_j(\cdot) \in L^2(\mathcal{U})$ $(j = 1, \ldots, p)$. See Cai and Liu (2011) for a similar class of covariance matrices for non-functional data. Compared with the class

$$\mathcal{C}^*(q, s_0(p); \mathcal{U}) = \big\{\mathbf{\Sigma} : \mathbf{\Sigma} \succeq 0, \max_j \|\sigma_j\|_{\mathcal{N}} \leqslant M, \max_j \sum_{k=1}^{p} \|\Sigma_{jk}\|_{\mathcal{S}}^q \leqslant s_0(p)\big\},$$

over which the universal functional thresholding estimator $\widehat{\mathbf{\Sigma}}_{\mathrm{U}}$ can be shown to be consistent, the columns of a covariance function in $\mathcal{C}(q, s_0(p); \mathcal{U})$ are required to be within a weighted ℓ_q/ℓ_2 ball, instead of a standard ℓ_q/ℓ_2 ball, where the weights are determined by $\|\sigma_j\|_\infty$'s. Unlike $\mathcal{C}^*(q, s_0(p); \mathcal{U})$, $\mathcal{C}(q, s_0(p); \mathcal{U})$ no longer requires the uniform boundedness assumption on $\|\sigma_j\|_{\mathcal{N}}$'s and allows $\max_j \|\sigma_j\|_{\mathcal{N}} \to \infty$. In the special case of $q = 0$, $\mathcal{C}(q, s_0(p); \mathcal{U})$ corresponds to a class of truly sparse covariance functions. Note that the constant $s_0(p)$ is allowed to depend on p and can be regarded implicitly as the restriction on functional sparsity.

Theorem 2.1. *Suppose that Conditions 2.1-2.4 hold. Then there exists some constant $\delta > 0$ such that, uniformly on $\mathcal{C}(q, s_0(p); \mathcal{U})$, if $\lambda = \delta(\log p/n)^{1/2}$,*

$$\|\widehat{\mathbf{\Sigma}}_{\mathrm{A}} - \mathbf{\Sigma}\|_1 = \max_{1 \leqslant k \leqslant p} \sum_{j=1}^{p} \|\widehat{\Sigma}_{jk}^{\mathrm{A}} - \Sigma_{jk}\|_{\mathcal{S}} = O_P\left\{ s_0(p) \left(\frac{\log p}{n} \right)^{\frac{1-q}{2}} \right\}. \tag{2.3}$$

Theorem 2.1 presents the convergence result in a functional version of matrix ℓ_1 norm. The rate in (2.3) is consistent to those of sparse covariance matrix estimates (Rothman et al., 2009; Cai and Liu, 2011).

We finally turn to investigate the support recovery consistency of $\widehat{\mathbf{\Sigma}}_{\mathrm{A}}$ over the parameter space of truly sparse covariance functions defined by

$$\mathcal{C}_0(s_0(p); \mathcal{U}) = \big\{ \mathbf{\Sigma} : \mathbf{\Sigma} \succeq 0, \max_{1 \leqslant j \leqslant p} \sum_{k=1}^{p} I(\|\Sigma_{jk}\|_{\mathcal{S}} \neq 0) \leqslant s_0(p) \big\},$$

which assumes that $\{\Sigma_{jk}(\cdot, \cdot)\}_{p \times p}$ has at most $s_0(p)$ non-zero entries on each row. The theorem below shows that, with the choice of $\lambda = \delta(\log p/n)^{1/2}$ for some constant $\delta > 0$, $\widehat{\mathbf{\Sigma}}_{\mathrm{A}}$ exactly recovers the support of $\mathbf{\Sigma}$, $\mathrm{supp}(\mathbf{\Sigma}) = \{(j, k) : \|\Sigma_{jk}\|_{\mathcal{S}} \neq 0\}$, with high probability.

Theorem 2.2. *Suppose that Conditions 2.1-2.4 hold and $\big\|\Sigma_{jk}/\Theta_{jk}^{1/2}\big\|_{\mathcal{S}} > (2\delta + \gamma)(\log p/n)^{1/2}$ for all $(j, k) \in supp(\mathbf{\Sigma})$ and some $\gamma > 0$, where δ is stated in Theorem 2.1. Then we have that*

$$\inf_{\Sigma \in \mathcal{C}_0} \mathrm{pr}\big\{ supp(\widehat{\mathbf{\Sigma}}_A) = supp(\mathbf{\Sigma}) \big\} \to 1 \text{ as } n \to \infty.$$

Theorem 2.2 guarantees that $\widehat{\mathbf{\Sigma}}_{\mathrm{A}}$ achieves the exact recovery of functional sparsity structure in $\mathbf{\Sigma}$, that is the graph support in functional connectivity analysis, with probability tending to 1. This theorem holds under the condition that the Hilbert–Schmidt norms of non-zero standardized functional entries exceed a certain

threshold, which ensures that non-zero components are correctly retained. See an analogous minimum signal strength condition for sparse covariance matrices in Cai and Liu (2011).

2.4 Simulations

We conduct a number of simulations to compare adaptive functional thresholding estimators to universal functional thresholding estimators. In each scenario, to mimic the infinite-dimensionality of random curves, we generate functional variables by $X_{ij}(u) = \mathbf{s}(u)^{\mathrm{T}}\boldsymbol{\theta}_{ij}$ for $i = 1, \ldots, n, j = 1, \ldots, p$ and $u \in \mathcal{U} = [0, 1]$, where $\mathbf{s}(u)$ is a 50-dimensional Fourier basis function and $\boldsymbol{\theta}_i = (\boldsymbol{\theta}_{i1}^{\mathrm{T}}, \ldots, \boldsymbol{\theta}_{ip}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{50p}$ is generated from a mean zero multivariate Gaussian distribution with covariance matrix $\boldsymbol{\Omega} = (\boldsymbol{\Omega}_{jk})_{p\times p}$. The functional sparsity pattern in $\boldsymbol{\Sigma} = \{\Sigma_{jk}(\cdot, \cdot)\}_{p\times p}$ with its (j, k)th entry $\Sigma_{jk}(u, v) = \mathbf{s}(u)^{\mathrm{T}}\boldsymbol{\Omega}_{jk}\mathbf{s}(v)$ can be characterized by the block sparsity structure in $\boldsymbol{\Omega}$. Define $\boldsymbol{\Omega}_{jk} = \omega_{jk}\mathbf{D}$ with $\mathbf{D} = \mathrm{diag}(1^{-2}, \ldots, 50^{-2})$ and hence $\mathrm{Cov}(\theta_{ijk}, \theta_{ijk'}) \sim k^{-2}I(k = k')$ for $k, k' = 1, \ldots, 50$. Then we generate $\boldsymbol{\Omega}$ with different block sparsity patterns as follows.

- Model 1 (block banded). For $j, k = 1, \ldots, p/2$, $\omega_{jk} = (1 - |j - k|/10)_+$. For $j, k = p/2 + 1, \ldots, p$, $\omega_{jk} = 4I(j = k)$.
- Model 2 (block sparse without any special structure). For $j, k = p/2+1, \ldots, p$, $\omega_{jk} = 4I(j = k)$. For $j, k = 1, \ldots, p/2$, we generate $\boldsymbol{\omega} = (\omega_{jk})_{p/2\times p/2} = \mathbf{B} + \delta\mathbf{I}_{p/2}$, where elements of $\mathbf{B}$ are sampled independently from Uniform$[0.3, 0.8]$ with probability 0.2 or 0 with probability 0.8, and $\delta = \{-\lambda_{\min}(\mathbf{B}), 0\} + 0.01$ to guarantee the positive definiteness of $\boldsymbol{\Omega}$.

We implement a cross-validation approach (Bickel and Levina, 2008) for choosing the optimal thresholding parameter in $\widehat{\boldsymbol{\Sigma}}_{\mathrm{A}}$. Specifically, we randomly divide the sample $\{\mathbf{X}_i : i = 1, \ldots, n\}$ into two subsamples of size n_1 and n_2, where $n_1 = n(1 - 1/\log n)$ and $n_2 = n/\log n$ and repeat this N times. Let $\widehat{\boldsymbol{\Sigma}}_{\mathrm{A},1}^{(\nu)}(\lambda)$ and $\widehat{\boldsymbol{\Sigma}}_{\mathrm{S},2}^{(\nu)}$ be the adaptive functional thresholding estimator with thresholding parameter λ and the sample covariance function based on n_1 and n_2 observations, respectively, from the νth split. We select the optimal $\hat{\lambda}$ by minimizing

$$\widehat{R}(\lambda) = N^{-1}\sum_{\nu=1}^{N} \|\widehat{\boldsymbol{\Sigma}}_{\mathrm{A},1}^{(\nu)}(\lambda) - \widehat{\boldsymbol{\Sigma}}_{\mathrm{S},2}^{(\nu)}\|_{\mathrm{F}}^2,$$

where $\|\cdot\|_{\mathrm{F}}$ denotes the functional version of Frobenius norm, that is for any $\mathbf{K} = \{K_{jk}(\cdot,\cdot)\}_{p\times p}$ with each $K_{jk} \in \mathbb{S}$, $\|\mathbf{K}\|_{\mathrm{F}} = (\sum_{j,k}\|K_{jk}\|_{\mathcal{S}}^2)^{1/2}$. The optimal thresholding parameter in $\widehat{\mathbf{\Sigma}}_{\mathrm{U}}$ can be selected in a similar fashion.

We compare the adaptive functional thresholding estimator $\widehat{\mathbf{\Sigma}}_{\mathrm{A}}$ to the universal functional thresholding estimator $\widehat{\mathbf{\Sigma}}_{\mathrm{U}}$ under hard, soft, SCAD (with $a = 3.7$) and adaptive lasso (with $\eta = 3$) functional thresholding rules. Here $\hat{\lambda}$ is selected by the cross-validation procedure with $N = 5$. We also obtain the sample covariance function $\widehat{\mathbf{\Sigma}}_{\mathrm{S}}$, the results of which deteriorate severely compared with the competitors, so we do not report their results here. We generate $n = 100$ observations for $p = 50, 100, 150$ and replicate each simulation 100 times. We examine the performance of nine approaches by both estimation and support recovery accuracies. In terms of the estimation accuracy, Table 2.1 reports numerical summaries of losses measured by functional versions of Frobenius and matrix ℓ_1 norms. To assess the support recovery consistency, we present in Table 2.2 the average of true positive rates (TPRs) and false positive rates (FPRs), defined as $\mathrm{TPR} = \#\{(j,k) : \|\widehat{\Sigma}_{jk}\|_{\mathcal{S}} \neq 0 \text{ and } \|\Sigma_{jk}\|_{\mathcal{S}} \neq 0\}/\#\{(j,k) : \|\Sigma_{jk}\|_{\mathcal{S}} \neq 0\}$ and $\mathrm{FPR} = \#\{(j,k) : \|\widehat{\Sigma}_{jk}\|_{\mathcal{S}} \neq 0 \text{ and } \|\Sigma_{jk}\|_{\mathcal{S}} = 0\}/\#\{(j,k) : \|\Sigma_{jk}\|_{\mathcal{S}} = 0\}$. For comparison, we also present the support recovery accuracy of the pairwise testing for uncorrelatedness (Zhang, 2013) with multiple testing adjustments in Table 2.3 of Appendix 2.C. The direct implementation of such inference procedure involves the eigen-decomposition of four-way tensors, which results in higher computational cost especially for large p.

Several conclusions can be drawn from Tables 2.1 and 2.2. First, in all scenarios, the adaptive functional thresholding estimator $\widehat{\mathbf{\Sigma}}_{\mathrm{A}}$ provides substantially improved accuracy over the universal functional thresholding estimator $\widehat{\mathbf{\Sigma}}_{\mathrm{U}}$ regardless of the thresholding rule or the loss used. Second, for support recovery, again $\widehat{\mathbf{\Sigma}}_{\mathrm{A}}$ uniformly outperforms $\widehat{\mathbf{\Sigma}}_{\mathrm{U}}$, which fails to recover the functional sparsity pattern especially when p is large. Third, the adaptive functional thresholding approach using the hard and the adaptive lasso functional thresholding rules tends to have lower losses and lower TPRs/FPRs than that using the soft and the SCAD functional thresholding rules.

2.5 Real Data

In this section, we aim to investigate the association between the brain functional connectivity and fluid intelligence (gF), the capacity to solve problems independently of acquired knowledge (Cattell, 1987). The dataset contains subjects of

Table 2.1: The average (standard error) functional matrix losses over 100 simulation runs.

Model	Method	$p = 50$		$p = 100$		$p = 150$	
		$\widehat{\mathbf{\Sigma}}_{\text{A}}$	$\widehat{\mathbf{\Sigma}}_{\text{U}}$	$\widehat{\mathbf{\Sigma}}_{\text{A}}$	$\widehat{\mathbf{\Sigma}}_{\text{U}}$	$\widehat{\mathbf{\Sigma}}_{\text{A}}$	$\widehat{\mathbf{\Sigma}}_{\text{U}}$
1	Functional Frobenius norm						
	Hard	5.51(0.04)	12.16(0.02)	8.08(0.04)	17.65(0.01)	10.18(0.04)	21.84(0.01)
	Soft	6.41(0.06)	10.58(0.08)	9.60(0.05)	16.81(0.07)	12.12(0.06)	21.60(0.05)
	SCAD	5.79(0.05)	10.73(0.08)	8.69(0.05)	16.92(0.07)	11.03(0.06)	21.63(0.04)
	Adap. lasso	5.39(0.04)	11.66(0.08)	7.92(0.04)	17.64(0.01)	9.94(0.05)	21.83(0.01)
	Functional matrix ℓ_1 norm						
	Hard	4.05(0.06)	9.44(0.01)	4.62(0.05)	9.52(0.01)	4.90(0.05)	9.55(0.01)
	Soft	5.16(0.07)	8.29(0.08)	6.03(0.05)	9.33(0.02)	6.38(0.05)	9.51(0.01)
	SCAD	4.49(0.08)	8.46(0.07)	5.47(0.06)	9.38(0.02)	5.90(0.06)	9.52(0.01)
	Adap.lasso	3.94(0.07)	9.11(0.07)	4.64(0.06)	9.51(0.01)	4.98(0.06)	9.55(0.01)
2	Functional Frobenius norm						
	Hard	5.78(0.03)	9.59(0.02)	9.72(0.04)	16.14(0.01)	14.38(0.06)	22.75(0.01)
	Soft	6.27(0.03)	8.73(0.04)	10.50(0.05)	15.29(0.05)	15.10(0.05)	22.35(0.05)
	SCAD	6.06(0.03)	8.76(0.04)	10.17(0.05)	15.33(0.05)	14.79(0.06)	22.36(0.04)
	Adap. lasso	5.57(0.03)	9.29(0.04)	9.22(0.04)	16.06(0.02)	13.33(0.06)	22.74(0.01)
	Functional matrix ℓ_1 norm						
	Hard	2.94(0.03)	4.85(0.01)	4.94(0.05)	7.27(0.01)	7.86(0.07)	10.54(0.01)
	Soft	3.39(0.03)	4.61(0.04)	5.51(0.04)	7.06(0.02)	8.42(0.05)	10.43(0.01)
	SCAD	3.31(0.03)	4.59(0.03)	5.43(0.04)	7.06(0.02)	8.35(0.05)	10.44(0.01)
	Adap. lasso	2.85(0.03)	4.76(0.02)	4.77(0.05)	7.23(0.01)	7.57(0.07)	10.54(0.01)

resting-state fMRI scans and the corresponding gF scores, measured by the 24-item Raven's Progressive Matrices, from the Human Connectome Project (HCP). We follow many recent proposals based on HCP by modelling signals as multivariate random functions with each region of interest (ROI) representing one random function (Zapata et al., 2022; Lee et al., 2021; Miao et al., 2022). We focus our analysis on $n_{\text{low}} = 73$ subjects with intelligence scores $gF \leqslant 8$ and $n_{\text{high}} = 85$ subjects with $gF \geqslant 23$, and consider $p = 83$ ROIs of three generally acknowledged modules in neuroscience study (Finn et al., 2015): the medial frontal (29 ROIs), frontoparietal (34 ROIs) and default mode modules (20 ROIs). For each subject, the BOLD signals at each ROI are collected every 0.72 seconds for a total of $L = 1200$ measurement locations (14.4 minutes). We first implement the ICA-FIX preprocessed pipeline (Glasser et al., 2013) and a standard band-pass filter at $[0.01, 0.08]$ Hz to exclude frequency bands not implicated in resting state functional connectivity (Biswal et al., 1995). Figure 2.8 in Appendix 2.C.3 displays exemplified trajectories of pre-smoothed data. The adaptive functional thresholding method is then adopted to estimate the sparse covariance function and therefore the brain networks.

The sparsity structures in $\widehat{\mathbf{\Sigma}}_{\text{A}}$ for both groups are displayed in Figure 2.1. With $\widehat{\lambda}$ selected by the cross-validation, the network associated with $\widehat{\mathbf{\Sigma}}_{\text{A}}$ for subjects with $gF \geqslant 23$ is more densely connected than that with $gF \leqslant 8$, as evident from Fig. 2.1(a)–(b). We further set the sparsity level to 70% and 85%, and present the corresponding sparsity patterns in Fig. 2.1(c)–(f). The results clearly indicate the

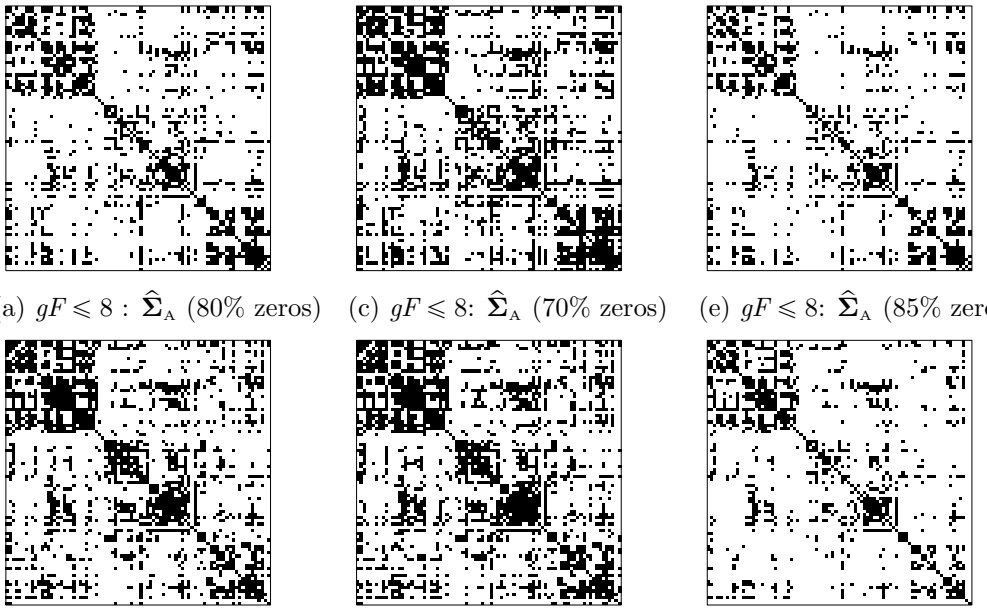
Table 2.2: The average TPRs/ FPRs over 100 simulation runs.

Model	Method	$p = 50$		$p = 100$		$p = 150$	
		$\widehat{\mathbf{\Sigma}}_{\text{A}}$	$\widehat{\mathbf{\Sigma}}_{\text{U}}$	$\widehat{\mathbf{\Sigma}}_{\text{A}}$	$\widehat{\mathbf{\Sigma}}_{\text{U}}$	$\widehat{\mathbf{\Sigma}}_{\text{A}}$	$\widehat{\mathbf{\Sigma}}_{\text{U}}$
1	Hard	0.70/0.00	0.00/0.00	0.66/0.00	0.00/0.00	0.63/0.00	0.00/0.00
	Soft	0.89/0.07	0.48/0.17	0.85/0.04	0.24/0.05	0.83/0.03	0.06/0.01
	SCAD	0.89/0.07	0.44/0.14	0.85/0.04	0.21/0.04	0.84/0.03	0.06/0.01
	Adap. lasso	0.78/0.00	0.11/0.02	0.74/0.00	0.00/0.00	0.72/0.00	0.00/0.00
2	Hard	0.77/0.00	0.00/0.00	0.68/0.00	0.00/0.00	0.63/0.00	0.00/0.00
	Soft	0.99/0.06	0.50/0.07	0.97/0.04	0.31/0.04	0.96/0.04	0.12/0.02
	SCAD	0.99/0.06	0.48/0.06	0.97/0.05	0.29/0.04	0.96/0.05	0.11/0.02
	Adap. lasso	0.91/0.00	0.11/0.01	0.85/0.00	0.02/0.00	0.83/0.00	0.00/0.00

existence of three diagonal blocks under all sparsity levels, complying with the identification of the medial frontal, frontoparietal and default mode modules in Finn et al. (2015). We also implement the universal functional thresholding method. However, compared with $\widehat{\mathbf{\Sigma}}_{\text{A}}$, the results of $\widehat{\mathbf{\Sigma}}_{\text{U}}$ suffer from the heteroscedasticity, as demonstrated in Section 1.5 and Section 2.C.2 of the Appendix, and fail to detect any noticeable block structure, hence we choose not to report them here. To explore the impact of gF on the functional connectivity, we compute the connectivity strength using the standardized form $\|\widehat{\Sigma}^{\text{A}}_{jk}\|_{\mathcal{S}}/\{\|\widehat{\Sigma}^{\text{A}}_{jj}\|_{\mathcal{S}}\|\widehat{\Sigma}^{\text{A}}_{kk}\|_{\mathcal{S}}\}^{1/2}$ for $j, k = 1 \ldots, p$. Interestingly, we observe from Figure 2.2 that subjects with $gF \geqslant 23$ tend to have enhanced brain connectivity in the medial frontal and frontoparietal modules, while the connectivity strength in the default mode module declines. This agrees with existing neuroscience literature reporting a strong positive association between intelligence score and the medial frontal/frontoparietal functional connectivity in the resting state (Van Den Heuvel et al., 2009; Finn et al., 2015), and lends support to the conclusion that lower default mode module activity is associated with better cognitive performance (Anticevic et al., 2012). See also Section 2.C.2 of the Appendix, in which we illustrate our adaptive functional thresholding estimation using another ADHD dataset.

2.6 Discussion

We conclude this chapter by discussing three directions for future study. The first extension considers estimating functional Gaussian graphical models targeting at recovering the conditional dependence structure among p random functions. Qiao et al. (2019) proposed to estimate a block sparse inverse covariance matrix by treating dimensions of random functions as approaching infinity. However, to deal with truly infinite-dimensional Gaussian processes, it is desirable to avoid the estimation



(a) $gF \leqslant 8$: $\widehat{\mathbf{\Sigma}}_{\text{A}}$ (80% zeros) (c) $gF \leqslant 8$: $\widehat{\mathbf{\Sigma}}_{\text{A}}$ (70% zeros) (e) $gF \leqslant 8$: $\widehat{\mathbf{\Sigma}}_{\text{A}}$ (85% zeros)

(b) $gF \geqslant 23$: $\widehat{\mathbf{\Sigma}}_{\text{A}}$ (72% zeros) (d) $gF \geqslant 23$: $\widehat{\mathbf{\Sigma}}_{\text{A}}$ (70% zeros) (f) $gF \geqslant 23$: $\widehat{\mathbf{\Sigma}}_{\text{A}}$ (85% zeros)

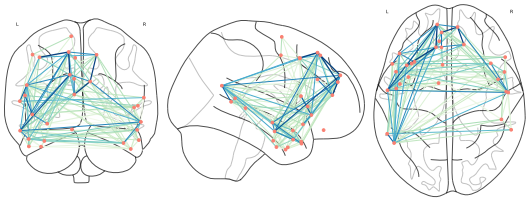
Figure 2.1: Estimated sparsity structures in $\widehat{\mathbf{\Sigma}}_{\text{A}}$ using soft functional thresholding rule at fluid intelligence $gF \leqslant 8$ and $gF \geqslant 23$: (a)–(b) with the corresponding $\hat{\lambda}$ selected by fivefold cross-validation; (c)–(f) with the estimated functional sparsity levels set at 70% and 85%.

of sparse inverse covariance function due to its unboundedness. For non-functional Gaussian graphical models, an innovative transformation (Fan and Lv, 2016) converts the problem of estimating sparse inverse covariance matrix to that of sparse covariance matrix estimation. It is thus of great interest to generalize this transformation strategy to the functional domain and hence our proposed sparse covariance function estimation approach can be applied.

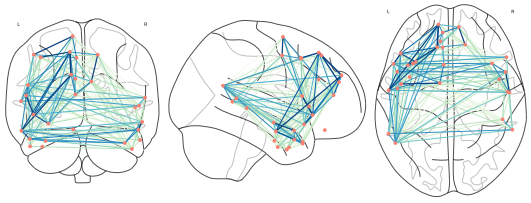
The second topic is about the classification for multivariate functional data, where estimating the covariance function plays a crucial role. Existing literature has focused on univariate or low-dimensional functional data, while our proposal of estimating sparse covariance function can be possibly incorporated into the development of functional classification under high-dimensional settings.

The third potential extension involves developing adaptive functional thresholding strategy for a practical scenario where functions are sparsely or densely observed with errors. This extension could be achieved using a nonparametric smoothing approach (Yao et al., 2005). It is interesting to develop standardized concentration results under different measurement schedules, which would pose complicated theoretical challenges.

These topics are beyond the scope of the current chapter and will be pursued elsewhere.

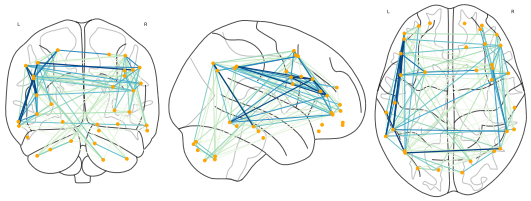


(a) $gF \leqslant 8$: the medial frontal module in Fig. 2.1(e)

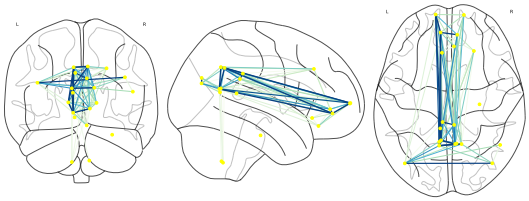


(d) $gF \geqslant 23$: the medial frontal module in Fig. 2.1(f)

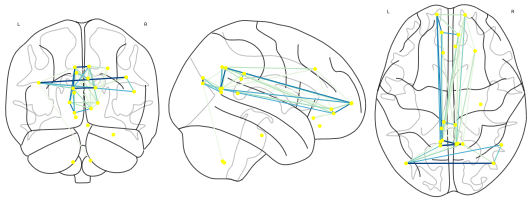
(b) $gF \leqslant 8$: the frontoparietal module in Fig. 2.1(e)



(e) $gF \geqslant 23$: the frontoparietal module in Fig. 2.1(f)



(c) $gF \leqslant 8$: the default mode module in Fig. 2.1(e)



(f) $gF \geqslant 23$: the default mode module in Fig. 2.1(f)

Figure 2.2: The connectivity strengths in Fig. 2.1(e)–(f) at fluid intelligence $gF \leqslant 8$ and $gF \geqslant 23$. Salmon, orange and yellow nodes represent the ROIs in the medial frontal, frontoparietal and default mode modules, respectively. The edge color from cyan to blue corresponds to the value of $\|\widehat{\Sigma}^{\mathrm{A}}_{jk}\|_{\mathcal{S}}/\{\|\widehat{\Sigma}^{\mathrm{A}}_{jj}\|_{\mathcal{S}}\|\widehat{\Sigma}^{\mathrm{A}}_{kk}\|_{\mathcal{S}}\}^{1/2}$ from small to large.

2.A Technical proofs

Before stating the regularity conditions, we make some notation. For a function $Z \in \mathbb{S}$, define $\|Z\|_\infty = \sup_{u,v\in\mathcal{U}}|Z(u,v)|$. For two sequences of real numbers $\{a_n\}$ and $\{b_n\}$, write $a_n \lesssim b_n$ if there exists some constant C such that $|a_n| \leqslant C|b_n|$ holds for all n, and similarly, for two sequences of real processes $\{a_n(u), u \in \mathcal{U}\}$ and $\{b_n(u), u \in \mathcal{U}\}$, write $a_n(u) \lesssim b_n(u)$ if there exists some constant C such that $|a_n(u)| \leqslant C|b_n(u)|$ holds for all n and $u \in \mathcal{U}$. Without loss of generality, in the following we assume that $E\{X_{ij}(u)\} \equiv 0$ and both estimators $\widehat{\Sigma}_{jk}(u,v)$ and $\widehat{\Theta}_{jk}(u,v)$ are defined as

$$\widehat{\Sigma}_{jk}(u,v) = \frac{1}{n}\sum_{i=1}^{n} X_{ij}(u)X_{ik}(v) \text{ and } \widehat{\Theta}_{jk}(u,v) = \frac{1}{n}\sum_{i=1}^{n} X_{ij}(u)^2X_{ik}(v)^2 - \widehat{\Sigma}_{jk}(u,v)^2,$$

respectively.

Lemma 2.1. *Suppose that Conditions 2.1–2.4 hold. Then for any $M > 0$, there exists some constant $\rho_1 > 0$ such that*

$$\mathrm{pr}\left\{\max_{j,k}\left\|\frac{\widehat{\Theta}_{jk} - \Theta_{jk}}{\Theta_{jk}}\right\|_\infty \geqslant \rho_1\frac{\log^2 p}{n^{1/2}}\right\} = O(p^{-M}).$$

Proof. Denote $\widetilde{\Theta}_{jk}(u,v) = E\{X_{ij}(u)^2X_{ik}(v)^2\}$. We decompose $\widehat{\Theta}_{jk}(u,v) - \Theta_{jk}(u,v)$ as

$$\begin{aligned}
&\widehat{\Theta}_{jk}(u,v) - \Theta_{jk}(u,v)\\
&= \Sigma_{jk}(u,v)^2 - \widehat{\Sigma}_{jk}(u,v)^2 + \frac{1}{n}\sum_{i=1}^{n}\Big\{X_{ij}(u)^2X_{ik}(v)^2 - \widetilde{\Theta}_{jk}(u,v)\Big\}.
\end{aligned}$$

By Condition 2.3, $\Theta_{jk}(u,v) \geqslant \tau\sigma_j(u)\sigma_k(v)$ for each $j,k = 1,\ldots,p$. Hence,

$$\begin{aligned}
&\left|\frac{\widehat{\Theta}_{jk}(u,v) - \Theta_{jk}(u,v)}{\Theta_{jk}(u,v)}\right|\\
&\leqslant \left|\frac{\Sigma_{jk}(u,v)^2 - \widehat{\Sigma}_{jk}(u,v)^2}{\tau\sigma_j(u)\sigma_k(v)}\right| + \left|\frac{1}{n}\sum_{i=1}^{n}\frac{X_{ij}(u)^2X_{ik}(v)^2 - \widetilde{\Theta}_{jk}(u,v)}{\tau\sigma_j(u)\sigma_k(v)}\right|\\
&= I_{jk}^{(1)}(u,v) + I_{jk}^{(2)}(u,v).
\end{aligned}$$

First, consider the concentration bound for $\|I_{jk}^{(1)}\|_\infty$. Denote $\widetilde{Y}_{ijk}(u,v) = Y_{ij}(u)Y_{ik}(v) - \Sigma_{jk}(u,v)/\{\sigma_j(u)^{1/2}\sigma_k(v)^{1/2}\}$ and let $d_{jk}((u,v),(u',v')) = d_j(u,u') + d_k(v,v')$. Ap-

plying Theorem 8.4 in Kosorok (2008) under Conditions 2.1 and 2.2, we obtain that, there exists some constant $C_1 > 0$ such that $\left\| \sup_{u \in \mathcal{U}} |Y_{1j}(u)| \right\|_{\psi_2} \leqslant C_1$ for all $j = 1, \ldots, p$. By the property of ψ_1-norm, we have that

$$
\begin{aligned}
&\|Y_{ij}(u)Y_{ik}(v) - Y_{ij}(u')Y_{ik}(v')\|_{\psi_1} \\
&\leqslant \|Y_{ij}(u)\{Y_{ik}(v) - Y_{ik}(v')\}\|_{\psi_1} + \|\{Y_{ij}(u) - Y_{ij}(u')\}Y_{ik}(v')\|_{\psi_1} \\
&\leqslant \|Y_{ij}(u)\|_{\psi_2} \|Y_{ik}(v) - Y_{ik}(v')\|_{\psi_2} + \|Y_{ik}(v')\|_{\psi_2} \|Y_{ij}(u) - Y_{ij}(u')\|_{\psi_2} \\
&\lesssim \{d_j(u, u') + d_k(v, v')\} = d_{jk}((u, v), (u', v')),
\end{aligned}
$$

which implies that

$$
\left\| \widetilde{Y}_{ijk}(u, v) - \widetilde{Y}_{ijk}(u', v') \right\|_{\psi_1} \lesssim d_{jk}((u, v), (u', v')). \tag{2.4}
$$

Note that

$$
\bar{Z}_{jk}(u, v) = \frac{\widehat{\Sigma}_{jk}(u, v) - \Sigma_{jk}(u, v)}{\sigma_j(u)^{1/2}\sigma_k(v)^{1/2}} = \frac{1}{n}\sum_{i=1}^{n} \left\{ Y_{ij}(u)Y_{ik}(v) - \frac{\Sigma_{jk}(u, v)}{\sigma_j(u)^{1/2}\sigma_k(v)^{1/2}} \right\},
$$

and for a random variable X and any integer $m \geqslant 1$, $E\|X\|^m \leqslant m!\|X\|_{\psi_1}^m$. By Bernstein's inequality and Lemma 8.3 of Kosorok (2008), we have that for $u, v, u', v' \in \mathcal{U}$,

$$
\left\| n^{1/2} \left\{ \bar{Z}_{jk}(u, v) - \bar{Z}_{jk}(u', v') \right\} \right\|_{\psi_1} \lesssim d_{jk}((u, v), (u', v')).
$$

For the semimetric d_{jk}, $D(\epsilon, d_{jk}) \leqslant D(\epsilon/2, d_j)D(\epsilon/2, d_k) \lesssim \epsilon^{-2r}$. Applying Theorem 8.4 in Kosorok (2008) with Conditions 2.1 and 2.2 again, we obtain that, there exists some constant $C_2 > 0$ such that

$$
\max_{1 \leqslant j,k \leqslant p} \left\| \sup_{u,v \in \mathcal{U}} |n^{1/2}\bar{Z}_{jk}(u, v)| \right\|_{\psi_1} \leqslant C_2.
$$

This immediately implies that there exist some universal constant $C_3 > 0$ such that for any $x > 0$,

$$
\mathrm{pr}\left\{ \max_{j,k} \sup_{u,v \in \mathcal{U}} \left| \frac{\widehat{\Sigma}_{jk}(u, v) - \Sigma_{jk}(u, v)}{\sigma_j(u)^{1/2}\sigma_k(v)^{1/2}} \right| > x \right\} \lesssim p^2 \exp\{-C_3 n^{1/2} x\}.
$$

As a result, for any $M > 0$, there exists some constant $\tilde{\rho}_1 > 0$ such that

$$\mathrm{pr}\left\{\max_{j,k}\sup_{u,v\in\mathcal{U}}\left|\frac{\widehat{\Sigma}_{jk}(u,v)-\Sigma_{jk}(u,v)}{\sigma_j(u)^{1/2}\sigma_k(v)^{1/2}}\right| > \tilde{\rho}_1\frac{\log p}{n^{1/2}}\right\} \lesssim p^{-M}. \tag{2.5}$$

Observe that

$$\left|\frac{\widehat{\Sigma}_{jk}(u,v)^2-\Sigma_{jk}(u,v)^2}{\sigma_j(u)\sigma_k(v)}\right| \leqslant \left|\frac{\widehat{\Sigma}_{jk}(u,v)-\Sigma_{jk}(u,v)}{\sigma_j(u)^{1/2}\sigma_k(v)^{1/2}}\right|^2 + 2\left|\frac{\widehat{\Sigma}_{jk}(u,v)-\Sigma_{jk}(u,v)}{\sigma_j(u)^{1/2}\sigma_k(v)^{1/2}}\right|,$$

since $|\Sigma_{jk}(u,v)| \leqslant \sigma_j(u)^{1/2}\sigma_k(v)^{1/2}$. By the inequality (2.5), we have that

$$\mathrm{pr}\left\{\max_{j,k}\|I_{jk}^{(1)}\|_\infty > 2\tilde{\rho}_1\frac{\log p}{n^{1/2}} + \tilde{\rho}_1^2\frac{\log^2 p}{n}\right\} \lesssim p^{-M}. \tag{2.6}$$

We next control the bound for $\|I_{jk}^{(2)}\|_\infty$ through the truncation technique. Note that

$$\frac{X_{ij}(u)^2X_{ik}(v)^2-\widetilde{\Theta}_{jk}(u,v)}{\sigma_j(u)\sigma_k(v)} = Y_{ij}(u)^2Y_{ik}(v)^2 - \frac{\widetilde{\Theta}_{jk}(u,v)}{\sigma_j(u)\sigma_k(v)}.$$

Define that $Y_{ij}^*(u) = Y_{ij}(u)I\left\{\|Y_{ij}\|_\infty \leqslant C_4\log^{1/2}(p\vee n)\right\}$ and

$$Z_{ijk}^*(u,v) = Y_{ij}^*(u)^2Y_{ik}^*(v)^2 - E\{Y_{ij}^*(u)^2Y_{ik}^*(v)^2\}.$$

By the property of ψ_1-norm and $|Y_{ij}^*(u)^2 - Y_{ij}^*(u')^2| \leqslant 2C_4\log^{1/2}(p\vee n)|Y_{ij}^*(u) - Y_{ij}^*(u')|$, we have that

$$\begin{aligned}
&\left\|Y_{ij}^*(u)^2Y_{ik}^*(v)^2 - Y_{ij}^*(u')^2Y_{ik}^*(v')^2\right\|_{\psi_1}\\
&\leqslant \left\|Y_{ij}^*(u)^2\{Y_{ik}^*(v)^2 - Y_{ik}^*(v')^2\}\right\|_{\psi_1} + \left\|\{Y_{ij}^*(u)^2 - Y_{ij}^*(u')^2\}Y_{ik}^*(v')^2\right\|_{\psi_1}\\
&\lesssim \log(p\vee n)\Big\{\|Y_{ij}^*(u)\|_{\psi_2}\left\|Y_{ik}^*(v) - Y_{ik}^*(v')\right\|_{\psi_2} + \|Y_{ik}^*(v')\|_{\psi_2}\left\|Y_{ij}^*(u) - Y_{ij}^*(u')\right\|_{\psi_2}\Big\}\\
&\lesssim \log(p\vee n)\{d_j(u,u') + d_k(v,v')\} \lesssim \log(p\vee n)d_{jk}((u,v),(u',v')),
\end{aligned}$$

which implies that, similar to (2.4),

$$\left\|Z_{ijk}^*(u,v) - Z_{ijk}^*(u',v')\right\|_{\psi_1} \lesssim \log(p\vee n)d_{jk}((u,v),(u',v')).$$

Let $\bar{Z}^*_{jk}(u,v) = n^{-1}\sum_{i=1}^n Z^*_{ijk}(u,v)$. We apply the similar technique of $\bar{Z}_{jk}$ above to the term $\bar{Z}^*_{jk}$ and obtain that there exists some universal constant $C_5 > 0$ such that for any $x > 0$,

$$\mathrm{pr}\left\{\max_{j,k}\sup_{u,v\in\mathcal{U}}\left|\frac{\bar{Z}^*_{jk}(u,v)}{\log(p\vee n)}\right| > x\right\} \lesssim p^2\exp(-C_5 n^{1/2}x).$$

As a result, for any $M > 0$, there exists some constant $\tilde{\rho}_2 > 0$ such that

$$\mathrm{pr}\left\{\max_{j,k}\sup_{u,v\in\mathcal{U}}\left|\bar{Z}^*_{jk}(u,v)\right| > \tilde{\rho}_2\frac{\log^2(p\vee n)}{n^{1/2}}\right\} \lesssim p^{-M}.$$

Now we consider the bound of the term $\|Y_{ij}\|_\infty$. By Conditions 2.1-2.2 and Theorem 8.4 of Kosorok (2008), we immediately have that there exists some constant $C_6 > 0$

$$\max_{1\leqslant i\leqslant n, 1\leqslant j\leqslant p}\left\|\sup_{u\in\mathcal{U}}|Y_{ij}(u)|\right\|_{\psi_2} \leqslant C_6,$$

which also implies that there exists some constant $C_7 > 0$ such that for any $x > 0$,

$$\mathrm{pr}\left\{\max_{1\leqslant i\leqslant n, 1\leqslant j\leqslant p}\|Y_{ij}(u)\|_\infty > x\right\} \lesssim np\exp(-C_7x^2).$$

Hence we obtain that for any $M > 0$, there exists some constant $C_4 > 0$ such that

$$\mathrm{pr}\left\{\max_{1\leqslant i\leqslant n, 1\leqslant j\leqslant p}\|Y_{ij}\|_\infty > C_4\log^{1/2}(p\vee n)\right\} \lesssim (p\vee n)^{-M}. \tag{2.7}$$

On the event

$$\Omega_{n0} = \left\{\max_{1\leqslant i\leqslant n, 1\leqslant j\leqslant p}\|Y_{ij}\|_\infty \leqslant C_4\log^{1/2}(p\vee n)\right\},$$

we find that

$$\begin{aligned} Y_{ij}(u)^2Y_{ik}(v)^2 - \frac{\widetilde{\Theta}_{jk}(u,v)}{\sigma_j(u)\sigma_k(v)} &= Y^*_{ij}(u)^2Y^*_{ik}(v)^2 - E\left\{Y^*_{ij}(u)^2Y^*_{ik}(v)^2\right\} \\ &\quad + E\left\{Y^*_{ij}(u)^2Y^*_{ik}(v)^2 - Y_{ij}(u)^2Y_{ik}(v)^2\right\}. \end{aligned}$$

Note that $Y^*_{ij}(u)^2 - Y_{ij}(u)^2 = Y_{ij}(u)^2 I\{\|Y_{ij}\|_\infty > C_4\log^{1/2}(p\vee n)\}$. By the inequality (2.7), we can obtain that

$$\left|E\left\{Y^*_{ij}(u)^2Y^*_{ik}(v)^2 - Y_{ij}(u)^2Y_{ik}(v)^2\right\}\right| \lesssim (p\vee n)^{-M}.$$

Therefore, for any $M > 0$, there exist some constant $\tilde{\rho}_3 > 0$ such that

$$\mathrm{pr}\left\{\max_{1\leqslant j\leqslant p}\|I_{jk}^{(2)}\|_\infty > \tilde{\rho}_3\frac{\log^2(p\vee n)}{n^{1/2}}\right\} \lesssim p^{-M}. \tag{2.8}$$

Combining (2.6) and (2.8), we obtain that for any $M > 0$, there exists some constant $\rho_1 > 0$ such that

$$\mathrm{pr}\left\{\max_{j,k}\left\|\frac{\widehat{\Theta}_{jk}-\Theta_{jk}}{\Theta_{jk}}\right\|_\infty \geqslant \rho_1\frac{\log^2(p\vee n)}{n^{1/2}}\right\} \lesssim p^{-M}.$$

The proof is complete. ▫

Lemma 2.2. *Suppose that Conditions 2.1–2.4 hold. Then for any $M > 0$, there exist some constant $\rho_2 > 0$ such that*

$$\max_{j,k}\left\|\frac{\Theta_{jk}^{1/2}-\widehat{\Theta}_{jk}^{1/2}}{\widehat{\Theta}_{jk}^{1/2}}\right\|_\infty \leqslant \rho_2\frac{\log^2 p}{n^{1/2}} \tag{2.9}$$

with probability greater than $1 - O(p^{-M})$.

Proof. Let the event $\Omega_n(s) = \{\|(\widehat{\Theta}_{jk}-\Theta_{jk})/\Theta_{jk}\|_\infty \leqslant s\log^2 p/n^{1/2} \leqslant 1/2\}$. For any $M > 0$, it follows from Lemma 2.1 that there exists some constant $\rho_1 > 0$ such that $\mathrm{pr}\{\Omega_n(\rho_1)\} \geqslant 1 - O(p^{-M})$. Since

$$\left\|\frac{\Theta_{jk}}{\widehat{\Theta}_{jk}}\right\|_\infty = \left\|\frac{\Theta_{jk}-\widehat{\Theta}_{jk}}{\widehat{\Theta}_{jk}}+1\right\|_\infty \leqslant \left\|\frac{\Theta_{jk}-\widehat{\Theta}_{jk}}{\Theta_{jk}}\right\|_\infty\left\|\frac{\Theta_{jk}}{\widehat{\Theta}_{jk}}\right\|_\infty + 1,$$

hence, on the event $\Omega_n(\rho_1)$, we have that $\|\Theta_{jk}/\widehat{\Theta}_{jk}\|_\infty \leqslant 2$. As a result, on the event $\Omega_n(\rho_1)$, it follows that

$$\left\|\frac{\Theta_{jk}^{1/2}-\widehat{\Theta}_{jk}^{1/2}}{\widehat{\Theta}_{jk}^{1/2}}\right\|_\infty = \left\|\frac{\Theta_{jk}-\widehat{\Theta}_{jk}}{\widehat{\Theta}_{jk}+\widehat{\Theta}_{jk}^{1/2}\Theta_{jk}^{1/2}}\right\|_\infty \leqslant \left\|\frac{\Theta_{jk}-\widehat{\Theta}_{jk}}{\Theta_{jk}}\right\|_\infty\left\|\frac{\Theta_{jk}}{\widehat{\Theta}_{jk}}\right\|_\infty \leqslant 2\rho_1\frac{\log^2 p}{n^{1/2}}.$$

Take $\rho_2 = 2\rho_1$ and the proof is complete. ▫

Lemma 2.3. *Suppose that Conditions 2.1–2.4 holds. Then for any $M > 0$, there exist some positive constant $\rho_3 > 0$ such that*

$$\max_{j,k}\left\|\frac{\widehat{\Sigma}_{jk}-\Sigma_{jk}}{\widehat{\Theta}_{jk}^{1/2}}\right\|_{\mathcal{S}} \leqslant \rho_3\left(\frac{\log p}{n}\right)^{1/2}$$

with probability greater than $1 - O(p^{-M})$.

Proof. Let $\widetilde{Y}_{ijk}(u,v) = Y_{ij}(u)Y_{ik}(v) - \Sigma_{jk}(u,v)/\{\sigma_j(u)^{1/2}\sigma_k(v)^{1/2}\}$ and

$$\bar{Z}_{jk}(u,v) = \frac{\widehat{\Sigma}_{jk}(u,v) - \Sigma_{jk}(u,v)}{\sigma_j(u)^{1/2}\sigma_k(v)^{1/2}} = \frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}_{ijk}(u,v).$$

We first derive the concentration bound of $\|\bar{Z}_{jk}\|_{\mathcal{S}}$. It follows from the proof of Lemma 2.1 that there exists some constant $C_8 > 0$ such that

$$\max_{j,k}\left\|\sup_{u,v\in\mathcal{U}}\left|\widetilde{Y}_{1jk}(u,v)\right|\right\|_{\psi_1} \leqslant C_8.$$

which further implies that $\max_{j,k}\left\|\|\widetilde{Y}_{1jk}\|_{\mathcal{S}}\right\|_{\psi_1} \leqslant C_8$. As a result, it follows from Theorem 2.5 of Bosq (2000) that there exists some universal constant $C_9 > 0$ such that for any $x > 0$

$$\mathrm{pr}\left(\|\bar{Z}_{jk}\|_{\mathcal{S}} \geqslant x\right) \leqslant 2\exp\{-C_9 n\min(x^2, x)\}.$$

For any $M > 0$, there exists some constant $\tilde{\rho} > 0$ that

$$\|\bar{Z}_{jk}\|_{\mathcal{S}} \leqslant \tilde{\rho}\left(\frac{\log p}{n}\right)^{1/2} \tag{2.10}$$

with probability greater than $1 - O(p^{-M})$.

Now we derive the bound of $\left\|(\widehat{\Sigma}_{jk} - \Sigma_{jk})/\widehat{\Theta}_{jk}^{1/2}\right\|_{\mathcal{S}}$. Note that Condition 2.3 implies that $\Theta_{jk}(u,v) \geqslant \tau\sigma_j(u)\sigma_k(v)$. We obtain that

$$\left\|\frac{\widehat{\Sigma}_{jk} - \Sigma_{jk}}{\widehat{\Theta}_{jk}^{1/2}}\right\|_{\mathcal{S}} \leqslant \left\|\frac{\widehat{\Sigma}_{jk} - \Sigma_{jk}}{\Theta_{jk}^{1/2}}\right\|_{\mathcal{S}}\left\|\frac{\Theta_{jk}^{1/2}}{\widehat{\Theta}_{jk}^{1/2}}\right\|_{\infty} \leqslant \|\tau^{-1/2}\bar{Z}_{jk}\|_{\mathcal{S}}\left(\left\|\frac{\Theta_{jk}^{1/2} - \widehat{\Theta}_{jk}^{1/2}}{\widehat{\Theta}_{jk}^{1/2}}\right\|_{\infty} + 1\right).$$

Hence, together with (2.10) and Lemma 2.2, the lemma follows. The proof is complete. □

Proof of Theorem 2.1. For easy representation, define

$$\widehat{\Phi}_{jk}(u,v) = \frac{\widehat{\Sigma}_{jk}(u,v)}{\widehat{\Theta}_{jk}(u,v)^{1/2}}, \quad \widetilde{\Phi}_{jk}(u,v) = \frac{\Sigma_{jk}(u,v)}{\widehat{\Theta}_{jk}(u,v)^{1/2}} \quad \text{and} \quad \Phi_{jk}(u,v) = \frac{\Sigma_{jk}(u,v)}{\Theta_{jk}(u,v)^{1/2}}.$$

Let

$$\Omega_{n1} = \Big\{ \max_{j,k} \|\widehat{\Phi}_{jk} - \widetilde{\Phi}_{jk}\|_{\mathcal{S}} \leqslant \lambda \Big\}, \Omega_{n2} = \left\{ \max_{j,k} \left\| \frac{\widehat{\Theta}_{jk} - \Theta_{jk}}{\Theta_{jk}} \right\|_{\infty} \leqslant \frac{1}{2} \right\}.$$

It is immediate to see that under the event Ω_{n2}, $2^{-1}\|\Theta_{jk}\|_\infty \leqslant \|\widehat{\Theta}_{jk}\|_\infty \leqslant 2\|\Theta_{jk}\|_\infty$ for all j and k. By Conditions 2.1–2.3, we have $\Theta_{jk}(u,v) \leqslant C'\sigma_j(u)\sigma_k(v)$ and $\Theta_{jk}(u,v) \geqslant \tau\sigma_j(u)\sigma_k(v)$ Then under the event $\Omega_{n1} \cap \Omega_{n2}$ and Conditions (i)-(iii) on $S_\lambda(Z)$, we obtain that

$$\begin{aligned}
&\sum_{k=1}^{p} \|\widehat{\Sigma}^{\mathrm{A}}_{jk} - \Sigma_{jk}\|_{\mathcal{S}} \\
&= \sum_{k=1}^{p} \|\widehat{\Sigma}^{\mathrm{A}}_{jk} - \Sigma_{jk}\|_{\mathcal{S}} I\{\|\widehat{\Phi}_{jk}\|_{\mathcal{S}} \geqslant \lambda\} + \sum_{k=1}^{p} \|\Sigma_{jk}\|_{\mathcal{S}} I\{\|\widehat{\Phi}_{jk}\|_{\mathcal{S}} < \lambda\} \\
&\leqslant \sum_{k=1}^{p} \Big\{ \|s_\lambda(\widehat{\Phi}_{jk}) - \widehat{\Phi}_{jk}\|_{\mathcal{S}} + \|\widehat{\Phi}_{jk} - \widetilde{\Phi}_{jk}\|_{\mathcal{S}} \Big\} \|\widehat{\Theta}^{1/2}_{jk}\|_\infty I\{\|\widehat{\Phi}_{jk}\|_{\mathcal{S}} \geqslant \lambda, \|\widetilde{\Phi}_{jk}\|_{\mathcal{S}} \geqslant \lambda\} \\
&\quad + \sum_{k=1}^{p} \Big\| \big[s_\lambda(\widehat{\Phi}_{jk}) - \widetilde{\Phi}_{jk}\big] \widehat{\Theta}^{1/2}_{jk} \Big\|_{\mathcal{S}} I\{\|\widehat{\Phi}_{jk}\|_{\mathcal{S}} \geqslant \lambda, \|\widetilde{\Phi}_{jk}\|_{\mathcal{S}} < \lambda\} + \sum_{k=1}^{p} \|\Sigma_{jk}\|_{\mathcal{S}} I\{\|\widetilde{\Phi}_{jk}\|_{\mathcal{S}} < 2\lambda\} \\
&\leqslant \sum_{k=1}^{p} 2\lambda \|\widehat{\Theta}^{1/2}_{jk}\|_\infty I\{\|\widetilde{\Phi}_{jk}\|_{\mathcal{S}} \geqslant \lambda\} + \sum_{k=1}^{p} (1+c)\|\widetilde{\Phi}_{jk}\|_{\mathcal{S}} \|\widehat{\Theta}^{1/2}_{jk}\|_\infty I\{\|\widetilde{\Phi}_{jk}\|_{\mathcal{S}} < \lambda\} \\
&\quad + \sum_{k=1}^{p} \|\widetilde{\Phi}_{jk}\|_{\mathcal{S}} \|\widehat{\Theta}^{1/2}_{jk}\|_\infty I\{\|\widetilde{\Phi}_{jk}\|_{\mathcal{S}} < 2\lambda\} \\
&\lesssim \lambda^{1-q} \sum_{k=1}^{p} \|\widehat{\Theta}_{jk}\|^{1/2}_\infty \|\widetilde{\Phi}_{jk}\|^q_{\mathcal{S}} \lesssim \lambda^{1-q} \sum_{k=1}^{p} \|\sigma_j\|^{(1-q)/2}_\infty \|\sigma_k\|^{(1-q)/2}_\infty \|\Sigma_{jk}\|^q_{\mathcal{S}} \lesssim s_0(p) \left(\frac{\log p}{n} \right)^{\frac{1-q}{2}}.
\end{aligned}$$

Since there exists some constant $\delta > 0$ such that $\mathrm{pr}\{\Omega^C_{n1}\} + \mathrm{pr}\{\Omega^C_{n2}\} \lesssim p^{-M}$, the theorem follows. ▫

Proof of Theorem 2.2. We consider two sets: $S_{n1} = \{(j,k) : \|\widehat{\Sigma}^{\mathrm{A}}_{jk}\|_{\mathcal{S}} \neq 0 \text{ and } \|\Sigma_{jk}\|_{\mathcal{S}} = 0\}$ and $S_{n2} = \{(j,k) : \|\widehat{\Sigma}^{\mathrm{A}}_{jk}\|_{\mathcal{S}} = 0 \text{ and } \|\Sigma_{jk}\|_{\mathcal{S}} \neq 0\}$. It suffices to prove that

$$\mathrm{pr}\big(|S_{n1}| > 0\big) + \mathrm{pr}\big(|S_{n2}| > 0\big) \to 0,$$

as $n, p \to \infty$. By Conditions (i)-(iii) on $S_\lambda(Z)$,

$$S_{n1} = \left\{ (j,k) : \left\| \frac{\widehat{\Sigma}_{jk}}{\widehat{\Theta}^{1/2}_{jk}} \right\|_{\mathcal{S}} > \lambda \text{ and } \|\Sigma_{jk}\|_{\mathcal{S}} = 0 \right\} \subset \left\{ (j,k) : \left\| \frac{\widehat{\Sigma}_{jk} - \Sigma_{jk}}{\widehat{\Theta}^{1/2}_{jk}} \right\|_{\mathcal{S}} > \lambda \right\}$$

Therefore, with the choice $\lambda = \delta(\log p/n)^{1/2}$, we obtain

$$P(|S_{n1}| > 0) \leqslant P\left\{\max_{j,k}\left\|\frac{\widehat{\Sigma}_{jk} - \Sigma_{jk}}{\widehat{\Theta}_{jk}^{1/2}}\right\|_{\mathcal{S}} > \lambda\right\} \lesssim p^{-M}. \tag{2.11}$$

for some prespecified $M > 0$. Similarly, we have

$$S_{n2} = \left\{(j,k) : \left\|\frac{\widehat{\Sigma}_{jk}}{\widehat{\Theta}_{jk}^{1/2}}\right\|_{\mathcal{S}} \leqslant \lambda \text{ and } \|\Sigma_{jk}\|_{\mathcal{S}} \neq 0\right\}.$$

Note that $\|\Sigma_{jk}\|_{\mathcal{S}} \neq 0$ implies that

$$(2\delta + \gamma)\left(\frac{\log p}{n}\right)^{1/2} < \left\|\frac{\Sigma_{jk}}{\Theta_{jk}^{1/2}}\right\|_{\mathcal{S}} \leqslant \left[\left\|\frac{\Sigma_{jk} - \widehat{\Sigma}_{jk}}{\widehat{\Theta}_{jk}^{1/2}}\right\|_{\mathcal{S}} + \left\|\frac{\widehat{\Sigma}_{jk}}{\widehat{\Theta}_{jk}^{1/2}}\right\|_{\mathcal{S}}\right]\left\|\frac{\widehat{\Theta}_{jk}^{1/2}}{\Theta_{jk}^{1/2}}\right\|_{\infty}. \tag{2.12}$$

Let $\Omega_{n3} = \left\{\|(\widehat{\Theta}_{jk}^{1/2} - \Theta_{jk}^{1/2})/\widehat{\Theta}_{jk}^{1/2}\|_{\infty} \leqslant \epsilon\right\}$ for some small constant $0 < \epsilon < \gamma/(4\delta + 2\gamma)$. Conditioned on the event of Ω_{n3}, the inequality

$$\left\|\frac{\widehat{\Theta}_{jk}^{1/2}}{\Theta_{jk}^{1/2}}\right\|_{\infty} \leqslant \left\|\frac{\widehat{\Theta}_{jk}^{1/2} - \Theta_{jk}^{1/2}}{\widehat{\Theta}_{jk}^{1/2}}\right\|_{\infty}\left\|\frac{\widehat{\Theta}_{jk}^{1/2}}{\Theta_{jk}^{1/2}}\right\|_{\infty} + 1$$

implies that $\|\widehat{\Theta}_{jk}^{1/2}/\Theta_{jk}^{1/2}\|_{\infty} \leqslant 1/(1-\epsilon)$. This together with (2.12) shows that

$$S_{n2} \cap \Omega_{n3} \subset \left\{(j,k) : \left\|\frac{\widehat{\Sigma}_{jk} - \Sigma_{jk}}{\widehat{\Theta}_{jk}^{1/2}}\right\|_{\mathcal{S}} > \delta\left(\frac{\log p}{n}\right)^{1/2}\right\}.$$

As a result,

$$\mathrm{pr}(|S_{n2}| > 0) \leqslant \mathrm{pr}(\Omega_{n3}^C) + \mathrm{pr}\left\{\max_{j,k}\left\|\frac{\widehat{\Sigma}_{jk} - \Sigma_{jk}}{\widehat{\Theta}_{jk}^{1/2}}\right\|_{\mathcal{S}} > \delta\left(\frac{\log p}{n}\right)^{1/2}\right\} \lesssim p^{-M}. \tag{2.13}$$

Combining (2.11) and (2.13), we complete our proof. ▫

2.B Examples of functional thresholding operators

In Section 2.B.1, we verify that our proposed soft, SCAD and adaptive lasso functional thresholding rules satisfy conditions (i)–(iii) in Section 2.2. We then present the derivations of these three functional thresholding rules in Section 2.B.2.

2.B.1 Condition verification

It is directly implied from the thresholding rules that the soft, SCAD and adaptive lasso functional methods satisfy condition (ii). Since the soft functional thresholding has the largest amount of functional shrinkage in the Hilbert–Schmidt norm compared with SCAD and adaptive lasso methods, it suffices to show that the soft functional thresholding satisfies condition (iii). For $\|Z\|_{\mathcal{S}} \leqslant \lambda$, the thresholding effect leads to $\|0 - Z\|_{\mathcal{S}} \leqslant \lambda$. When $\|Z\|_{\mathcal{S}} > \lambda$, we obtain that $\|Z\lambda/\|Z\|_{\mathcal{S}}\|_{\mathcal{S}} = \lambda$.

We next show that the above three thresholding methods satisfy condition (i). By the triangle inequality, $\|Z - Y\|_{\mathcal{S}} \leqslant \lambda$ in condition (i) implies that $\big|\|Z\|_{\mathcal{S}} - \lambda\big| \leqslant \|Y\|_{\mathcal{S}}$.

- Soft functional thresholding: If $\|Z\|_{\mathcal{S}} \leqslant \lambda$, $0 \leqslant c\|Y\|_{\mathcal{S}}$ directly holds for all $Y \in \mathbb{S}$ and $c > 0$. When $\|Z\|_{\mathcal{S}} > \lambda$, we have $\|s^{\mathrm{S}}_{\lambda}(Z)\|_{\mathcal{S}} = \|Z\|_{\mathcal{S}} - \lambda \leqslant \|Y\|_{\mathcal{S}}$ with the choice of $c = 1$.

- SCAD functional thresholding: When $\|Z\|_{\mathcal{S}} \leqslant 2\lambda$, $s^{\mathrm{SC}}_{\lambda}(Z)$ is the same as the soft functional thresholding rule. For $\|Z\|_{\mathcal{S}} > 2\lambda$, we have $\|s^{\mathrm{SC}}_{\lambda}(Z)\|_{\mathcal{S}} \leqslant \|Z\|_{\mathcal{S}} \leqslant \|Y\|_{\mathcal{S}} + \lambda \leqslant \|Y\|_{\mathcal{S}} + \|Z\|_{\mathcal{S}}/2$ and hence $\|s^{\mathrm{SC}}_{\lambda}(Z)\|_{\mathcal{S}} \leqslant \|Z\|_{\mathcal{S}} \leqslant 2\|Y\|_{\mathcal{S}}$. Combining the above results, we take $c = 2$.

- Adaptive lasso functional thresholding: Let $\lceil \eta \rceil$ denote the smallest integer greater than or equal to η. For $\|Z\|_{\mathcal{S}} \leqslant \lambda$, this condition holds for all $Y \in \mathbb{S}$ and $c > 0$. For $\|Z\|_{\mathcal{S}} > \lambda$, we have that $\|s^{\mathrm{AL}}_{\lambda}(Z)\|_{\mathcal{S}} = \|Z(1 - \lambda^{\eta+1}/\|Z\|^{\eta+1}_{\mathcal{S}})\|_{\mathcal{S}} = (\|Z\|^{\eta+1}_{\mathcal{S}} - \lambda^{\eta+1})/\|Z\|^{\eta}_{\mathcal{S}} \leqslant (\|Z\|^{\lceil\eta\rceil+1}_{\mathcal{S}} - \lambda^{\lceil\eta\rceil+1})/\|Z\|^{\lceil\eta\rceil}_{\mathcal{S}} = (\|Z\|_{\mathcal{S}} - \lambda)(\|Z\|^{\lceil\eta\rceil}_{\mathcal{S}} + \|Z\|^{\lceil\eta\rceil-1}_{\mathcal{S}}\lambda + \cdots + \lambda^{\lceil\eta\rceil})/\|Z\|^{\lceil\eta\rceil}_{\mathcal{S}} \leqslant (\lceil\eta\rceil + 1)\|Y\|_{\mathcal{S}}$. Hence, for any $\eta \geqslant 0$, we can find $c = \lceil\eta\rceil + 1$. In the special case of $\eta = 0$, $s^{\mathrm{AL}}_{\lambda}(Z)$ degenerates to the soft functional thresholding rule with $c = 1$, which is consistent with our finding for the soft functional thresholding.

2.B.2 Derivations of the functional thresholding rules from various penalty functions

Soft functional thresholding can be obtained via

$$s_\lambda^{\mathrm{S}}(Z) = \arg\min_{\theta \in \mathbb{S}} \left\{ \frac{1}{2}\|\theta - Z\|_{\mathcal{S}}^2 + \lambda \|\theta\|_{\mathcal{S}} \right\}. \tag{2.14}$$

First, we show that if $\|Z\|_{\mathcal{S}} \leqslant \lambda$, then $\|s_\lambda^{\mathrm{S}}(Z)\|_{\mathcal{S}} = 0$ and hence $s_\lambda^{\mathrm{S}}(Z) = 0$. This results from the fact that, for any θ,

$$\begin{aligned} \frac{1}{2}\|\theta - Z\|_{\mathcal{S}}^2 + \lambda\|\theta\|_{\mathcal{S}} &\geqslant \frac{1}{2}\big(\|\theta\|_{\mathcal{S}} - \|Z\|_{\mathcal{S}}\big)^2 + \lambda\|\theta\|_{\mathcal{S}} \\ &= \frac{1}{2}\|\theta\|_{\mathcal{S}}^2 + (\lambda - \|Z\|_{\mathcal{S}})\|\theta\|_{\mathcal{S}} + \frac{1}{2}\|Z\|_{\mathcal{S}}^2 \geqslant \frac{1}{2}\|Z\|_{\mathcal{S}}^2. \end{aligned}$$

Second, we show that if $\|Z\|_{\mathcal{S}} > \lambda$, then $\|s_\lambda^{\mathrm{S}}(Z)\|_{\mathcal{S}} \neq 0$. In fact, we can find $\theta_c = cZ$ with $c = 1 - \lambda/\|Z\|_{\mathcal{S}} > 0$ such that

$$\frac{1}{2}\|\theta_c - Z\|_{\mathcal{S}}^2 + \lambda\|\theta_c\|_{\mathcal{S}} = \frac{1}{2}(1-c)^2\|Z\|_{\mathcal{S}}^2 + \lambda c\|Z\|_{\mathcal{S}} < \frac{1}{2}\|Z\|_{\mathcal{S}}^2.$$

As a result, we are able to take the first derivative of (2.14) with respect to θ and set $p'_\lambda(\theta) = \theta - Z + \lambda\theta/\|\theta\|_{\mathcal{S}} = 0$. Thus, $\widehat{\theta} = Z\|\widehat{\theta}\|_{\mathcal{S}}/\big(\|\widehat{\theta}\|_{\mathcal{S}} + \lambda\big)$, which implies that $\|\widehat{\theta}\|_{\mathcal{S}} = \|Z\|_{\mathcal{S}} - \lambda$. Combining the above results, we have that $\widehat{\theta} = Z(1 - \lambda/\|Z\|_{\mathcal{S}})_+$.

The SCAD and adaptive lasso functional thresholding rules can be derived in a similar fashion. Hence, we only present their penalty functions here. The functional version of SCAD penalty takes the form of

$$\begin{aligned} p_\lambda(\theta) =& \lambda\|\theta\|_{\mathcal{S}} I(\|\theta\|_{\mathcal{S}} \leqslant \lambda) + \frac{2a\lambda\|\theta\|_{\mathcal{S}} - \|\theta\|_{\mathcal{S}}^2 - \lambda^2}{2(a-1)} I(\lambda < \|\theta\|_{\mathcal{S}} \leqslant a\lambda) \\ &+ \frac{\lambda^2(a+1)}{2} I(\|\theta\|_{\mathcal{S}} > a\lambda), \end{aligned}$$

for $a > 2$. For the functional version of adaptive lasso penalty, we use $p_\lambda(\theta) = \lambda^{\eta+1}\|Z\|_{\mathcal{S}}^{-\eta}\|\theta\|_{\mathcal{S}}$, for $\eta \geqslant 0$. A similar adaptive lasso penalty function operating on $|\cdot|$ for the univariate scalar case can be found in Rothman et al. (2009).

2.C Additional empirical results

2.C.1 Simulation studies

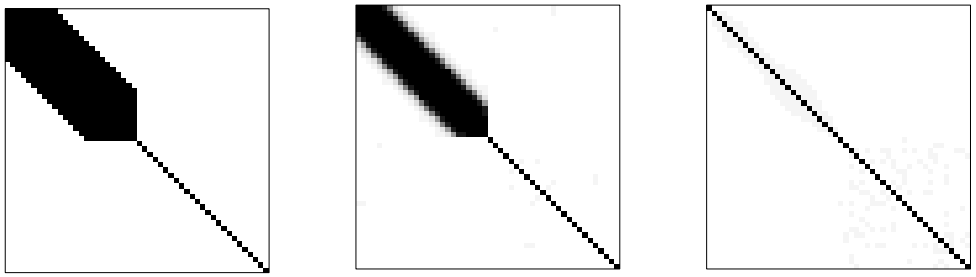
Table 2.3 presents the average TPRs/FPRs of pairwise testing over 100 simulation runs with two multiple testing adjustments: Benjamini-Hochberg Procedure (B-H) and Bonferroni correction. The B-H procedure attains similar FPRs/FPRs with the adaptive functional thresholding estimator $\widehat{\mathbf{\Sigma}}_{\text{A}}$ under the adaptive lasso functional thresholding rule, which also demonstrate the effectiveness of the proposed adaptive thresholding idea. As expected, the Bonferroni correction procedure provides much lower TPRs due to its conservative nature. Figures 2.3 and 2.4 plot the heat maps of the frequency of the zeros identified for the Hilbert–Schimidt norm of each entry of the estimated covariance function, when $p = 50$, out of 100 simulation runs. The true nonzero patterns of Model 1 and 2 are presented in Figures 2.3(a) and 2.4(a), respectively. Figure 2.5 displays the average receiver operating characteristic (ROC) curves (plots of true positive rates versus false positive rates over a sequence of λ values) for both the adaptive functional thresholding and universal functional thresholding methods. These results again demonstrate the uniform superiority of the adaptive functional thresholding method in terms of graph selection consistency.

Table 2.3: The average TPRs/ FPRs of pairwise testing over 100 simulation runs.

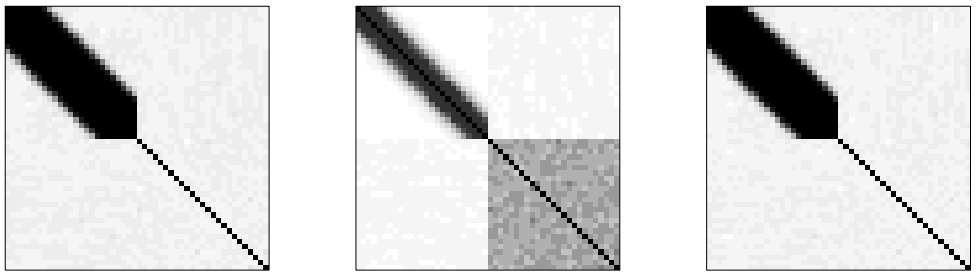
Model	Method	$p = 50$	$p = 100$	$p = 150$
1	B-H	0.80/0.01	0.75/0.00	0.73/0.00
	Bonferroni	0.65/0.00	0.58/0.00	0.54/0.00
2	B-H	0.87/0.00	0.79/0.00	0.75/0.00
	Bonferroni	0.60/0.00	0.43/0.00	0.30/0.00

2.C.2 ADHD dataset

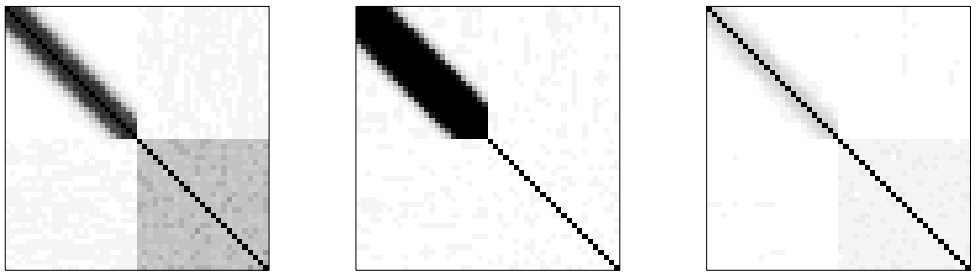
In this section, we illustrate our adaptive functional thresholding estimation using the ADHD-200 Sample, collected by New York University Medical Center. This dataset consists of resting-state fMRI scans with Blood Oxygenation Level-Dependent (BOLD) signals recorded every 2 seconds in the whole brain with $L = 172$ locations in total, for $n_{\text{ADHD}} = 90$ patients diagnosed with attention-deficit/hyperactivity disorder (ADHD) and $n_{\text{TDC}} = 87$ typically-developing controls (TDC). The preprocessing of the raw fMRI data is performed by Neuro Bureau using the Athena



(a) True (b) Hard $\widehat{\mathbf{\Sigma}}_{\text{A}}$ (c) Hard $\widehat{\mathbf{\Sigma}}_{\text{U}}$

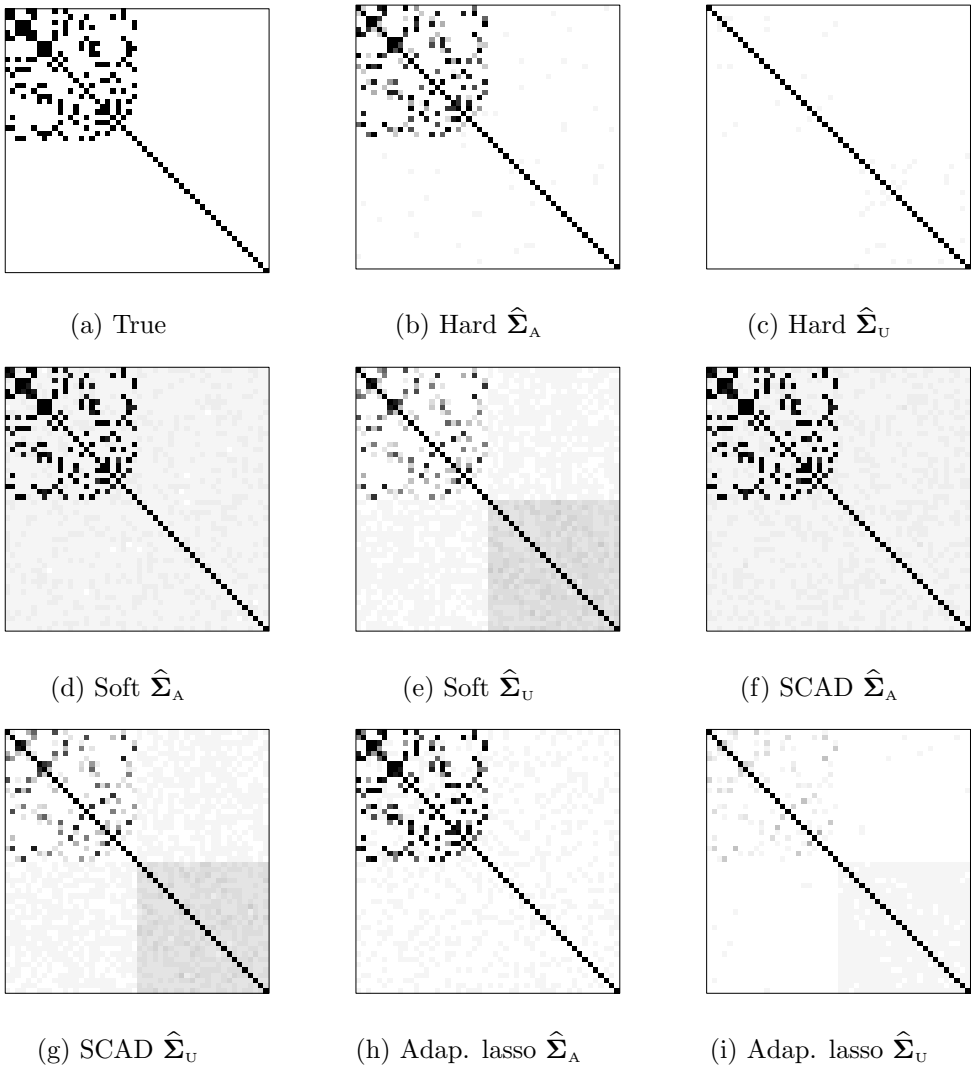


(d) Soft $\widehat{\mathbf{\Sigma}}_{\text{A}}$ (e) Soft $\widehat{\mathbf{\Sigma}}_{\text{U}}$ (f) SCAD $\widehat{\mathbf{\Sigma}}_{\text{A}}$



(g) SCAD $\widehat{\mathbf{\Sigma}}_{\text{U}}$ (h) Adap. lasso $\widehat{\mathbf{\Sigma}}_{\text{A}}$ (i) Adap. lasso $\widehat{\mathbf{\Sigma}}_{\text{U}}$

Figure 2.3: Heat maps of the frequency of the zeros identified for the Hilbert–Schimidt norm of each entry of the estimated covariance function (when $p = 50$) for Model 1 out of 100 simulation runs. White and black correspond to 100/100 and 0/100 zeros identified, respectively.



(a) True (b) Hard $\widehat{\mathbf{\Sigma}}_{\text{A}}$ (c) Hard $\widehat{\mathbf{\Sigma}}_{\text{U}}$

(d) Soft $\widehat{\mathbf{\Sigma}}_{\text{A}}$ (e) Soft $\widehat{\mathbf{\Sigma}}_{\text{U}}$ (f) SCAD $\widehat{\mathbf{\Sigma}}_{\text{A}}$

(g) SCAD $\widehat{\mathbf{\Sigma}}_{\text{U}}$ (h) Adap. lasso $\widehat{\mathbf{\Sigma}}_{\text{A}}$ (i) Adap. lasso $\widehat{\mathbf{\Sigma}}_{\text{U}}$

Figure 2.4: Heat maps of the frequency of the zeros identified for the Hilbert–Schimidt norm of each entry of the estimated covariance function (when $p = 50$) for Model 2 out of 100 simulation runs. White and black correspond to 100/100 and 0/100 zeros identified, respectively.

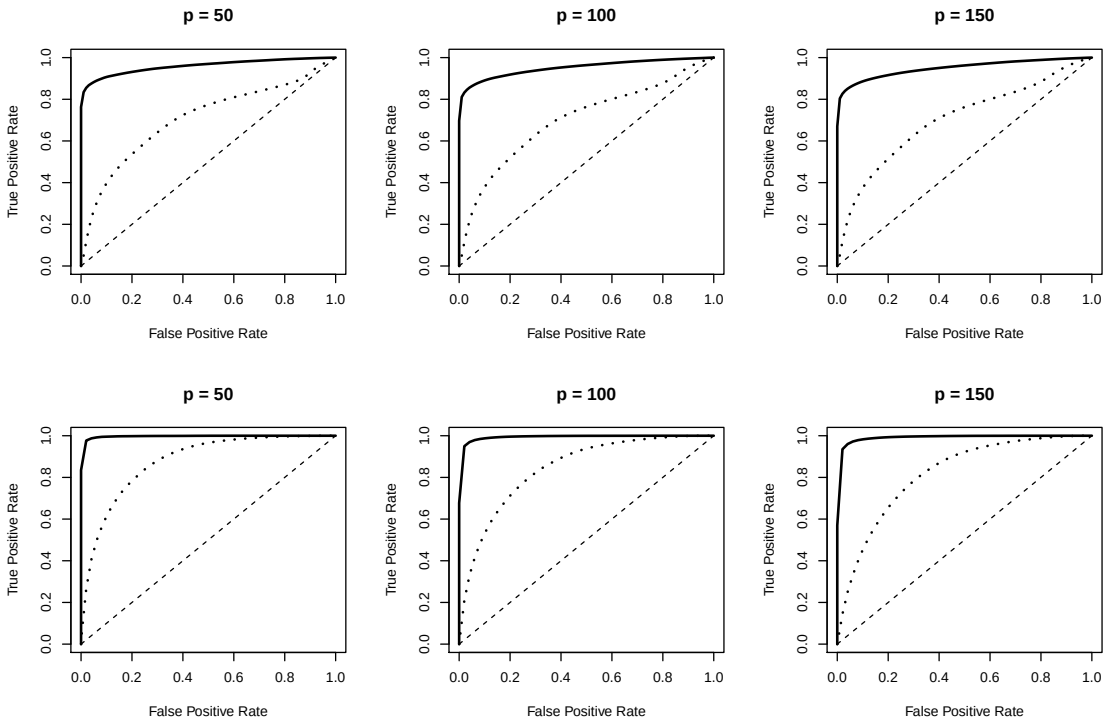
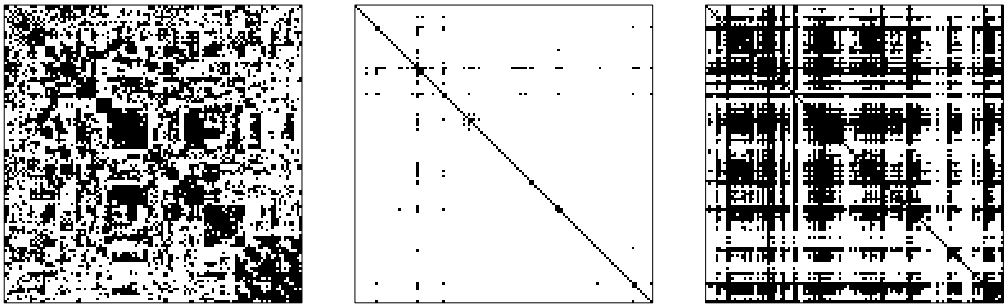


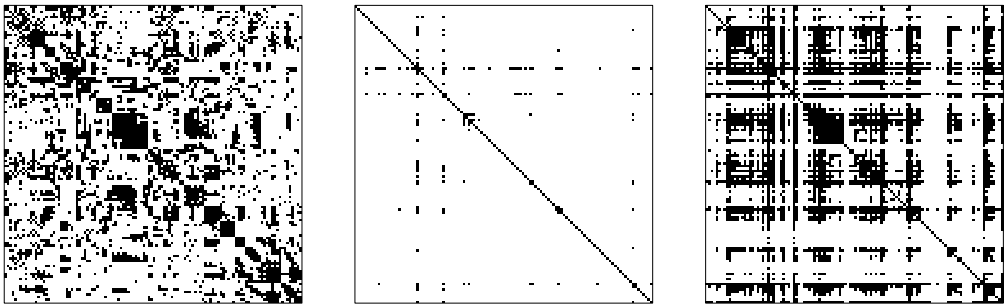
Figure 2.5: Model 1 (top row) and Model 2 (bottom row) for $p = 50, 100, 150$: Comparison of the average ROC curves for adaptive functional thresholding (solid line) and universal functional thresholding (dotted line) over 100 simulation runs.

pipeline (Bellec et al., 2017). See Figure 2.7 in Section 2.C.3 for plots of pre-smoothed BOLD signals at a selection of ROIs. Following Li and Solea (2018) based on the same dataset, we treat the signals at different ROIs as multivariate functional data. Our goal is to construct resting state functional connectivity networks among $p = 116$ ROIs (Tzourio-Mazoyer et al., 2002), with the first 90 ROIs from the cerebrum and the last 26 ROIs from the cerebellum, for ADHD and TDC groups, respectively. To this end, we implement adaptive and universal functional thresholding methods to discover the networks for two groups.

Figure 2.6 plots the sparsity patterns in estimated covariance functions corresponding to identified functional connectivity networks. We observe several interesting patterns. First, with $\hat{\lambda}$ selected by the cross-validation, $\widehat{\mathbf{\Sigma}}_{\text{A}}$ in Fig. 2.6(a)–(b) reveal clear blockwise connectivity structures with two blocks coinciding with the regions of the cerebrum and the cerebellum, while $\widehat{\mathbf{\Sigma}}_{\text{U}}$ in Fig. 2.6(c)–(d) result in very sparse networks. Second, under the same sparsity levels as those of $\widehat{\mathbf{\Sigma}}_{\text{A}}$ in Fig. 2.6(a)–(b), $\widehat{\mathbf{\Sigma}}_{\text{U}}$ in Fig. 2.6(e)–(f) only retain edges related to large marginal-covariance functions but fail to identify some essential within-network connections, e.g., those of the cerebellar region (Dobromyslin et al., 2012) on the bottom right corner. Third, the ADHD group has increased connections relative to the TDC group, which is in line with the finding in Konrad and Eickhof (2010) that ADHD patients tend to



(a) ADHD: $\widehat{\mathbf{\Sigma}}_{\text{A}}$ (57.50% zeros) (c) ADHD: $\widehat{\mathbf{\Sigma}}_{\text{U}}$ (98.94% zeros) (e) ADHD: $\widehat{\mathbf{\Sigma}}_{\text{U}}$ (57.50% zeros)



(b) TDC: $\widehat{\mathbf{\Sigma}}_{\text{A}}$ (71.24% zeros) (d) TDC: $\widehat{\mathbf{\Sigma}}_{\text{U}}$ (98.85% zeros) (f) TDC: $\widehat{\mathbf{\Sigma}}_{\text{U}}$ (71.24% zeros)

Figure 2.6: The sparsity structures in $\widehat{\mathbf{\Sigma}}_{\text{A}}$ and $\widehat{\mathbf{\Sigma}}_{\text{U}}$ for ADHD and TDC groups: (a)–(d) with the corresponding $\hat{\lambda}$ selected by fivefold cross-validation using soft functional threosholding rule; (e)–(f) with the same sparsity levels as those in (a)–(b). Black corresponds to nonzero entries of $\widehat{\mathbf{\Sigma}}_{\text{A}}$ and $\widehat{\mathbf{\Sigma}}_{\text{U}}$ (identified edges connecting a subset of ROIs).

exhibit abnormal spontaneous functional connectivity patterns.

2.C.3 Additional real data results

Figures 2.7 and 2.8 display the pre-smoothed BOLD signal trajectories at a selection of ROIs of subjects from the ADHD and HCP datesets, respectively. Figures 2.9 and 2.10 plot the connectivity strengths at fluid intelligence $gF \leqslant 8$ and $gF \geqslant 23$ in Fig. 2.1(a)–(b) and Fig. 2.1(c)–(d), respectively. We observe that as gF increases, the connectivity strengths in the medial frontal and frontoparietal modules tend to increase while those in the default mode module decrease, which is consistent with our finding in Section 2.5.

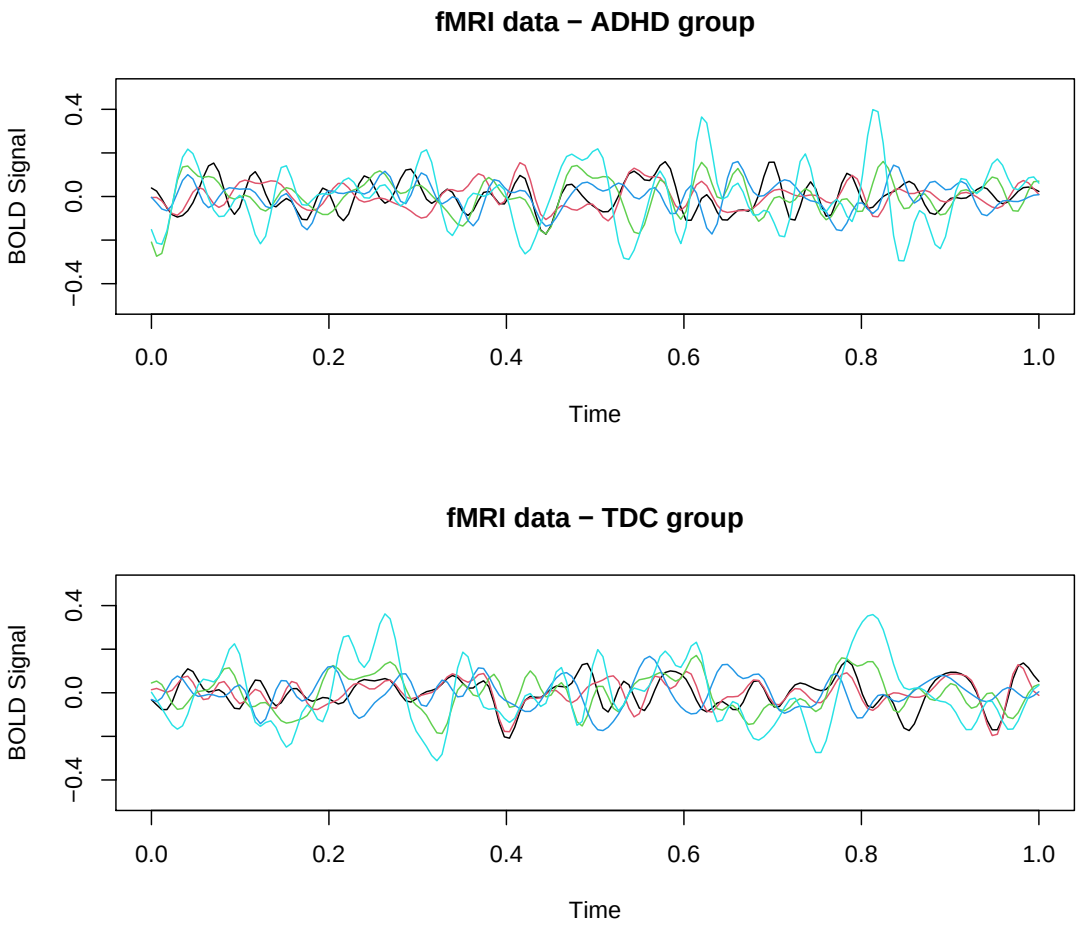


Figure 2.7: ADHD dataset: the smoothed BOLD signals at the first 5 ROIs of two subjects in ADHD and TDC groups respectively. The 5.73-minute interval with 172 scanning points is rescaled to $[0, 1]$.

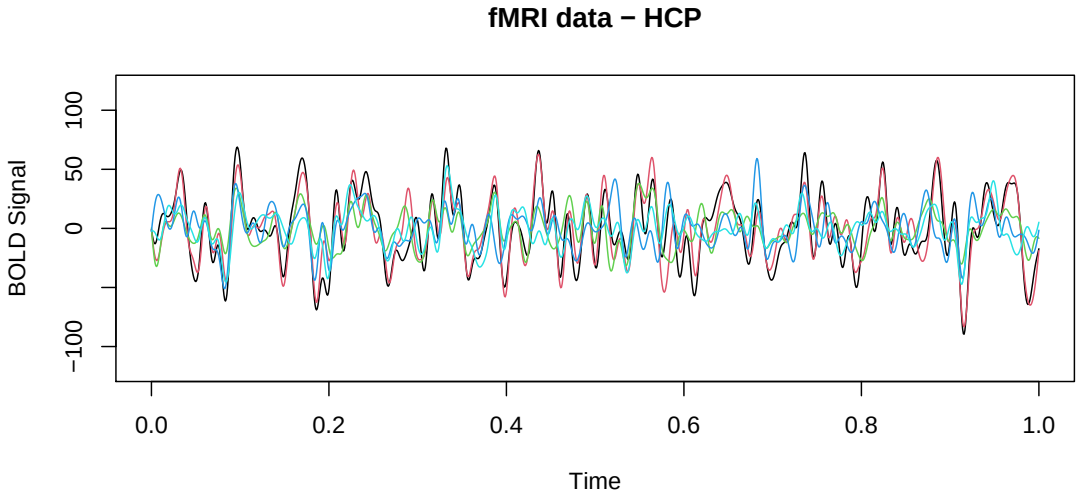
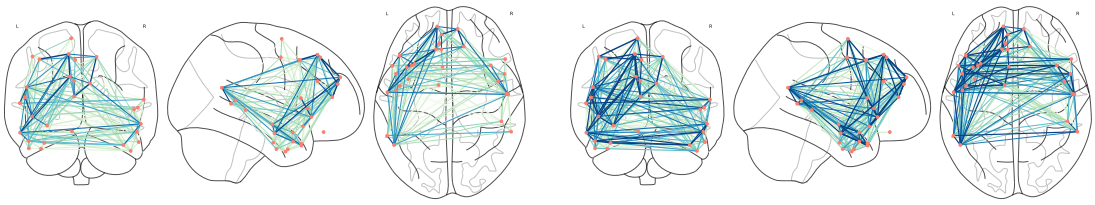
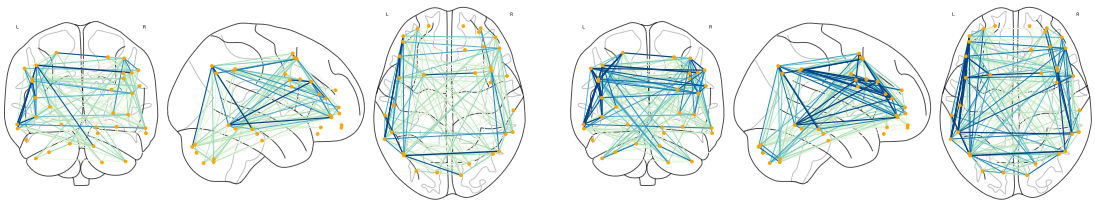


Figure 2.8: HCP dataset: the smoothed BOLD signals at the first 5 ROIs of one subject. The 14.40-minute interval with 1200 scanning points (14.40 mins) is rescaled to $[0, 1]$.



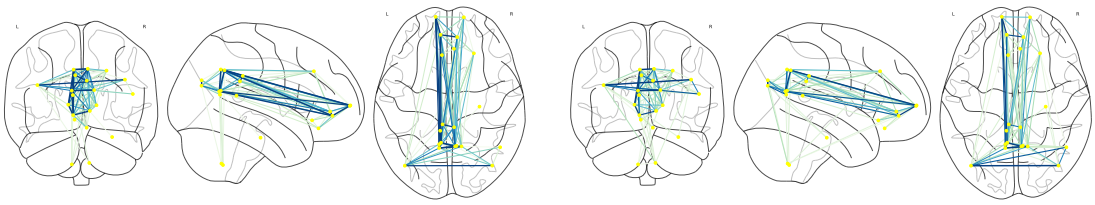
(a) $gF \leqslant 8$: the medial frontal module in Fig. 2.1(a)

(d) $gF \geqslant 23$: the medial frontal module in Fig. 2.1(b)



(b) $gF \leqslant 8$: the frontoparietal module in Fig. 2.1(a)

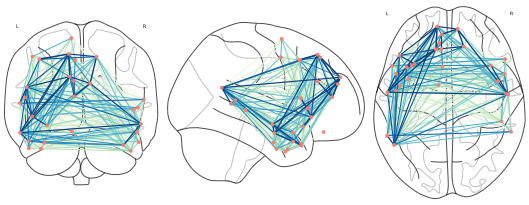
(e) $gF \geqslant 23$: the frontoparietal module in Fig. 2.1(b)



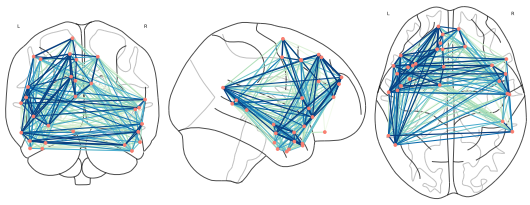
(c) $gF \leqslant 8$: the default mode module in Fig. 2.1(a)

(f) $gF \geqslant 23$: the default mode module in Fig. 2.1(b)

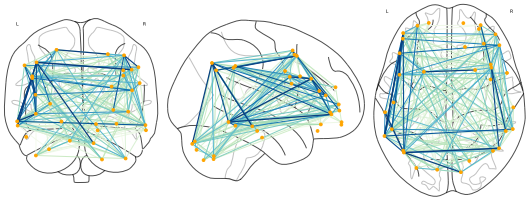
Figure 2.9: The connectivity strengths in Fig. 2.1(a)–(b) at fluid intelligence $gF \leqslant 8$ and $gF \geqslant 23$. Salmon, orange and yellow nodes represent the ROIs in the medial frontal, frontoparietal and default mode modules, respectively. The edge color from cyan to blue corresponds to the value of $\|\widehat{\Sigma}^{\mathrm{A}}_{jk}\|_{\mathcal{S}} / \{\|\widehat{\Sigma}^{\mathrm{A}}_{jj}\|_{\mathcal{S}} \|\widehat{\Sigma}^{\mathrm{A}}_{kk}\|_{\mathcal{S}}\}^{1/2}$ from small to large.



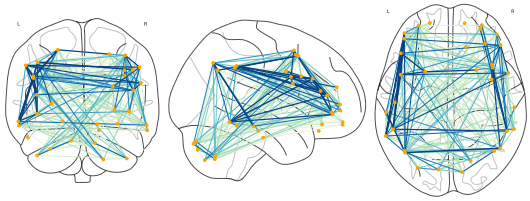
(a) $gF \leqslant 8$: the medial frontal module in Fig. 2.1(c)



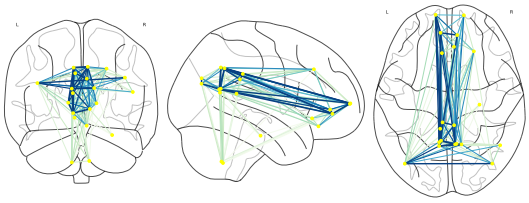
(d) $gF \geqslant 23$: the medial frontal module in Fig. 2.1(d)



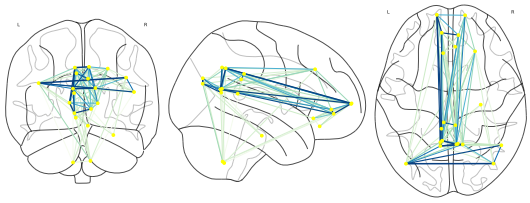
(b) $gF \leqslant 8$: the frontoparietal module in Fig. 2.1(c)



(e) $gF \geqslant 23$: the frontoparietal module in Fig. 2.1(d)



(c) $gF \leqslant 8$: the default mode module in Fig. 2.1(c)



(f) $gF \geqslant 23$: the default mode module in Fig. 2.1(d)

Figure 2.10: The connectivity strengths in Fig. 2.1(c)–(d) at fluid intelligence $gF \leqslant 8$ and $gF \geqslant 23$. Salmon, orange and yellow nodes represent the ROIs in the medial frontal, frontoparietal and default mode modules, respectively. The edge color from cyan to blue corresponds to the value of $\|\widehat{\Sigma}^{\mathrm{A}}_{jk}\|_{\mathcal{S}} / \{\|\widehat{\Sigma}^{\mathrm{A}}_{jj}\|_{\mathcal{S}} \|\widehat{\Sigma}^{\mathrm{A}}_{kk}\|_{\mathcal{S}}\}^{1/2}$ from small to large.

Chapter 3

On the modelling and prediction of high-dimensional functional time series

3.1 Introduction

Functional time series typically refers to continuous-time records that are naturally divided into consecutive time intervals, such as days, months or years, over which the observed curves are treated as serially dependent realizations of an underlying stochastic process. With recent advances in data collection technology, multivariate or even high-dimensional functional time series are rising ubiquitously in many applications. Typical examples include daily pollution concentration curves (Hörmann et al., 2015a) and annual temperature curves (Aue et al., 2018) collected at a number of stations, annual age-specific mortality rates for different prefectures (Gao et al., 2019b) and intraday energy consumption trajectories (Cho et al., 2013) for thousands of households, to list a few. These modern applications call for new methods to tackle problems involving high-dimensional functional time series.

We consider p-dimensional functional time series $\mathbf{Y}_t(u) = \{Y_{t1}(u), \ldots, Y_{tp}(u)\}^{\mathrm{T}}$ for $i = 1, \ldots, n$ defined on a compact set $\mathcal{U}$. Under a high-dimensional regime, not only $Y_{tj}(\cdot)$'s are infinite-dimensional objects exhibiting the serial dependence across observations, the dimension p is comparable to, or even larger than, the sample size n, posing a challenging learning task that is largely unexplored in the literature.

Hence, it is of fundamental importance to develop an effective procedure to model and predict $\mathbf{Y}_t(u)$ with large p.

A standard procedure in the existing literature is to firstly extract features by performing dimension reduction for each component series $Y_{tj}(\cdot)$ separately via, e.g., functional principal component analysis (FPCA) or dynamic FPCA (Hörmann et al., 2015a) or the method of Bathia et al. (2010), and then considers modelling p vector time series separately or applying high-dimensional techniques to model p vector time series jointly under some lower-dimensional structural assumptions, e.g., regularized vector autoregressions (Guo and Qiao, 2022; Chang et al., 2022) and factor model (Gao et al., 2019b). However, modelling the p vector time series separately fails to account for the cross serial dependence among different component series of $\mathbf{Y}_t(\cdot)$ that is essential in predicting future values of $\mathbf{Y}_t(\cdot)$, while the gain from incorporating those cross-(auto)covariance in a high-dimensional joint model is typically not enough to compensate the errors accumulated from estimating a large number of parameters. Alternatively, one can concatenate multiple functions to perform dimension reduction for $\mathbf{Y}_t(\cdot)$ directly via, e.g., multivariate FPCA (Chiou et al., 2014; Happ and Greven, 2018) when p is fixed or sparse FPCA (Hu and Yao, 2021) when p is large. However, these methods do not consider the serial dependence information and the subsequent modelling of extracted features under high-dimensional scaling remains untrapped.

In this chapter, we decompose $\mathbf{Y}_t(\cdot)$ as follows:

$$\mathbf{Y}_t(u) = \mathbf{X}_t(u) + \boldsymbol{\varepsilon}_t(u), \quad u \in \mathcal{U}, \tag{3.1}$$

where $\mathbf{X}_t(u) = \{X_{t1}(u), \ldots, X_{tp}(u)\}^{\mathrm{T}}$ is the finite-dimensional curve dynamics, $\boldsymbol{\varepsilon}_t(u) = \{\varepsilon_{t1}(u), \ldots, \varepsilon_{tp}(u)\}^{\mathrm{T}}$ is white noise in the sense that $\mathbb{E}\{\boldsymbol{\varepsilon}_t(u)\} = \mathbf{0}$ and $\mathbb{E}\{\boldsymbol{\varepsilon}_t(u)\boldsymbol{\varepsilon}_s(v)^{\mathrm{T}}\} = \mathbf{0}$ for any $u, v \in \mathcal{U}$ and $t \neq s$. Note $\{\mathbf{X}_t(\cdot)\}_{t=1}^n$ and $\{\boldsymbol{\varepsilon}_t(\cdot)\}_{t=1}^n$ are uncorrelated and unobservable. Under the decomposition (3.1), the linear dynamic structure of $\mathbf{Y}_t(\cdot)$ is entirely driven by a finite number of scalar coefficients under suitable basis expansion of $\mathbf{X}_t(\cdot)$, while no parts of $\mathbf{X}_t(\cdot)$ are white noise since those parts are absorbed into $\boldsymbol{\varepsilon}_t(\cdot)$. Therefore, the problem of modelling $\mathbf{Y}_t(\cdot)$ is reduced to that of modelling the associated finite-dimensional vector time series of $\mathbf{X}_t(\cdot)$. Efficient strategies can be implemented to predict vector time series, which can be further re-transformed to predict multivariate functional time series.

When the dimension p is moderate or large, the intrinsic dimensionality of $\mathbf{X}_t(\cdot)$ is large, thus leading to less efficient predictions of the associated high-dimensional vector time series of $\mathbf{X}_t(\cdot)$. Our first segmentation step proposes to transform $\mathbf{Y}_t(\cdot)$

into a new p-dimensional functional times series by a contemporaneous linear transformation. The estimation of the proposed transformation boils down to the eigenanalysis of a positive definite matrix, formed by the double integral and sum of quadratic forms in (auto)covariance functions of $\mathbf{Y}_t(\cdot)$ from different time lags. A maximum-pairwise-covariance-based permutation is further developed to segment the p transformed curve series into multiple groups, where curves from different groups are uncorrelated across all time lags. Hence, the overall linear dynamics is effectively converted into the cross serial dependence among subseries from the same group, and an initial effective dimension reduction is achieved. Within each group, the transformed curves can also be decomposed in the same form of (3.1), i.e. the sum of two uncorrelated and latent components, one finite-dimensional dynamic and one white noise. Inspired by the fact that the autocovariance function automatically filters out the noise, the proposed second step is applied to each group of transformed curve subseries based on the eigenanalysis of a positive-definite operator defined in terms of their autocovariance functions. Such proposal is targeted to estimate the finite-dimensional dynamic structure and based on which predict low-dimensional transformed curve subseries in a groupwise fashion. Finally, owing to the one-to-one linear transformation in the segmentation step, the good predictive performance of transformed curve series can be groupwisely transformed back to the prediction of the original high-dimensional functional time series.

This chapter makes useful contributions at multiple fronts. First, the segmentation transformation ensures zero cross serial correlations among curve subseries from different groups while maintaining the useful cross dynamical information within each group. Hence our proposal is more advantageous than predicting each $Y_{tj}(\cdot)$ separately. It also avoids predicting $\mathbf{Y}_t(\cdot)$ directly based on a joint model, thus not suffering from the 'curse of dimensionality'. Second, despite the basic idea of the proposed transformation being similar to the so-called PCA for scalar time series of Chang et al. (2018), our proposal relies on the double integral to take full advantage of the functional nature of the data by gathering the (auto)covariance information at each $(u, v) \in \mathcal{U}^2$ and then integrating over $\mathcal{U}^2$. Third, aided by the enforced sparsity, we propose a novel functional thresholding procedure, which guarantees the consistent estimation under a high-dimensional regime. Fourth, the autocovariance-based dimension reduction approach makes the good use of the serial dependence information in our estimation, which is most relevant in the context of time series modelling and prediction. Our proposal extends the univariate approach of Bathia et al. (2010) by taking into account the cross-autocovariance to accommodate multivariate functional time series and by allowing different component series with different domains.

Literature review. Our work lies in the intersection of two strands of literature: functional time series and high-dimensional time series. In the context of functional time series, many standard univariate or vector time series theory and methods have been adapted to the functional domain, see, e.g., Bathia et al. (2010); Hörmann and Kokoszka (2010); Cho et al. (2013); Panaretos and Tavakoli (2013); Aue et al. (2015); Hörmann et al. (2015a); Aue et al. (2018); Li et al. (2020); Chen et al. (2022); Jiao et al. (2021), among many others. In the context of high-dimensional scalar time series, the available methods to reduce the number of parameters can be loosely divided into two categories: (i) regularization (Han et al., 2015; Basu and Michailidis, 2015; Guo et al., 2016; Lin and Michailidis, 2017; Ghosh et al., 2019; Wilms et al., 2021) and (ii) dimension reduction via factor model (Pena and Box, 1987; Bai and Ng, 2002; Forni et al., 2005; Pan and Yao, 2008; Lam et al., 2011; Lam and Yao, 2012; Stock and Watson, 2012; Fan et al., 2016) or independent component analysis (Tiao and Tsay, 1989; Back and Weigend, 1997; Matteson and Tsay, 2011; Chang et al., 2018), each of which corresponds to a large body of literature and hence an incomplete list of the relevant references is presented here.

The remainder of the paper is organized as follows. In Section 3.2, we develop an estimation procedure for the first step of segmentation transformation with the help of permutation and functional thresholding. Section 3.3 specifies the methodology for the second dimension reduction step that can be applied to estimate the finite-dimensional dynamical structure within each segmented group. We investigate the associated theoretical properties of the proposed two-step procedure in Section 3.4. The finite-sample performance of our methods are examined through extensive simulations in Section 3.5. Section 3.6 applies our proposal to three real datasets, revealing its superior predictive performance over the competitors. All technical proofs are relegated to Section 3.A of the Appendix.

3.2 Segmentation transformation

3.2.1 Model setting

In this section, we focus on the case where p is large or moderately large. Our first segmentation step transforms linearly observed curves $\mathbf{Y}_t(\cdot)$ in (3.1) into p-vector of new functional time series $\mathbf{Z}_t(\cdot) = \{Z_{t1}(\cdot), \ldots, Z_{tp}(\cdot)\}^{\mathrm{T}}$ in the form of (3.2) below such that p new curve series can be divided into q groups of sizes $p_1, \ldots, p_q$ respectively, and the curves from different groups are uncorrelated across all time

lags, where $q \leqslant p$, $p_l \geqslant 1$ and $\sum_{l=1}^{q} p_l = p$. To be specific, we consider that $\mathbf{Y}_t(u)$ admits a latent segmentation structure:

$$\mathbf{Y}_t(u) = \mathbf{A}\mathbf{Z}_t(u) = \mathbf{A}\big\{\mathbf{Z}_{t,1}(u)^{\mathrm{T}}, \ldots, \mathbf{Z}_{t,q}(u)^{\mathrm{T}}\big\}^{\mathrm{T}}, \quad u \in \mathcal{U}, \tag{3.2}$$

where $\mathbf{A}$ is a $p \times p$ unknown constant matrix, the l-th group is formed by p_l-vector of functional time series $\mathbf{Z}_{t,l}(u)$, and $\mathrm{Cov}\{\mathbf{Z}_{t,l}(u), \mathbf{Z}_{s,l'}(v)\} = \mathbf{0}$ for all t, s, $l \neq l'$ and $u, v \in \mathcal{U}$. Write $\mathbf{A} = (\mathbf{A}_1, \ldots, \mathbf{A}_q)$, where $\mathbf{A}_l$ has p_l columns. To simplify the matter concerned, we restrict to orthogonal transformation only, i.e., $\mathbf{A}^{\mathrm{T}}\mathbf{A} = \mathbf{A}\mathbf{A}^{\mathrm{T}} = \mathbf{I}_p$ (an $p \times p$ identity matrix), which together with (3.2) implies that

$$\mathbf{Z}_{t,l}(u) = \mathbf{A}_l^{\mathrm{T}}\mathbf{Y}_t(u), \quad l = 1, \ldots, q. \tag{3.3}$$

Hence, it holds that

$$\begin{aligned} \boldsymbol{\Sigma}_{z,k}(u,v) &\equiv \mathrm{Cov}\{\mathbf{Z}_t(u), \mathbf{Z}_{t+k}(v)\} \\ &= \mathbf{A}^{\mathrm{T}}\mathrm{Cov}\{\mathbf{Y}_t(u), \mathbf{Y}_{t+k}(v)\}\mathbf{A} \equiv \mathbf{A}^{\mathrm{T}}\boldsymbol{\Sigma}_{y,k}(u,v)\mathbf{A} \ \text{ for any } \ u, v \in \mathcal{U}_0, \end{aligned} \tag{3.4}$$

where $\boldsymbol{\Sigma}_{z,k}(u,v)$ is block-diagonal with the main block sizes $p_1, \ldots, p_q$.

Remark 3.1. *(i) There is an identifiable issue among* $\mathbf{A}$ *and* $\mathbf{Z}_t(\cdot)$*, since (3.2) and the block-diagonal structure of* $\boldsymbol{\Sigma}_{z,k}(u,v)$ *(i.e., q subseries* $\mathbf{Z}_{t,l}(\cdot)$ *are uncorrelated with each other both contemporaneously and serially) remain unchanged if* $\{\mathbf{A}, \mathbf{Z}_t(\cdot)\}$ *is replaced by* $\{\mathbf{A}\boldsymbol{\Gamma}, \boldsymbol{\Gamma}^{-1}\mathbf{Z}_t(\cdot)\}$ *for any invertible matrix* $\boldsymbol{\Gamma} = diag(\boldsymbol{\Gamma}, \ldots, \boldsymbol{\Gamma}_q)$*, where* $\boldsymbol{\Gamma}_l$ *is a* $p_l \times p_l$ *submatrix. With the additional constraint that* $\mathbf{A}$ *is an orthogonal matrix,* $\mathbf{A}$ *and* $\mathbf{Z}_t(\cdot)$ *still can not be determined uniquely, but the linear spaces spanned by the columns of* $\mathbf{A}_l$*, denoted by* $\mathcal{C}(\mathbf{A}_l)$ *for* $l = 1 \ldots, q$*, and the latent segmentation structure of* $\mathbf{Z}_t(\cdot)$ *can.*
(ii) To guarantee the orthogonality of $\mathbf{A}$ *while maintaining the block-diagonal structure, we can adopt a normalization step to* $\mathbf{Y}_t(\cdot)$*. Define* $\mathbf{V}_y = \int_{\mathcal{U}_0} \boldsymbol{\Sigma}_{y,0}(u,u)\,\mathrm{d}u$ *and* $\mathbf{V}_z = \int_{\mathcal{U}_0} \boldsymbol{\Sigma}_{z,0}(u,u)\,\mathrm{d}u$*. We then replace* $\mathbf{Y}_t(\cdot)$ *and* $\mathbf{Z}_t(\cdot)$ *by* $\mathbf{V}_y^{-1/2}\mathbf{Y}_t(\cdot)$ *and* $\mathbf{V}_z^{-1/2}\mathbf{Z}_t(\cdot)$*, respectively. It follows from (3.2) that* $\mathbf{V}_y^{-1/2}\mathbf{Y}_t(\cdot) = \mathbf{A}^*\{\mathbf{V}_z^{-1/2}\mathbf{Z}_t(\cdot)\}$ *with the normalized transformation matrix* $\mathbf{A}^* = \mathbf{V}_y^{-1/2}\mathbf{A}\mathbf{V}_z^{1/2}$ *and*

$$\mathbf{I}_p = \int_{\mathcal{U}_0} Var\{\mathbf{V}_y^{-1/2}\mathbf{Y}_t(u)\}\,\mathrm{d}u = \mathbf{A}^* \int_{\mathcal{U}_0} Var\{\mathbf{V}_z^{-1/2}\mathbf{Z}_t(u)\}\mathbf{A}^{*\mathrm{T}}\,\mathrm{d}u = \mathbf{A}^*\{\mathbf{A}^*\}^{\mathrm{T}},$$

which implies that $\mathbf{A}^*$ *is an orthogonal matrix. Moreover, the block diagonal structure of the autocovariance matrices of* $\mathbf{V}_z^{-1/2}\mathbf{Z}_t(\cdot)$ *is the same as that of* $\mathbf{Z}_t(\cdot)$*, there-*

fore $\mathbf{V}_z^{-1/2}\mathbf{Z}_t(\cdot)$ *maintains the latent segmentation structure of* $\mathbf{Z}_t(\cdot)$. *In practice, we apply such normalization idea as a preliminary step by substituting* $\mathbf{Y}_t(\cdot)$ *with* $\widehat{\mathbf{V}}_y^{-1/2}\mathbf{Y}_t(\cdot)$, *where* $\widehat{\mathbf{V}}_y = n^{-1}\sum_{t=1}^{n}\int_{\mathcal{U}_0}\{\mathbf{Y}_t(u)-\bar{\mathbf{Y}}(u)\}\{\mathbf{Y}_t(u)-\bar{\mathbf{Y}}(u)\}^{\mathrm{T}}\,\mathrm{d}u$ *is the sample estimator for* $\mathbf{V}_y$. *It is noteworthy that the above normalization relies on* $\mathbf{V}_y$ *instead of the double integral via* $\widetilde{\mathbf{V}}_y = \int_{\mathcal{U}_0}\int_{\mathcal{U}_0}\mathbf{\Sigma}_{y,0}(u,v)\,\mathrm{d}u\mathrm{d}v$, *since the positive-definiteness no longer holds for* $\widetilde{\mathbf{V}}_y$.

Remark 3.2. *Combining (3.1) and (3.3) implies that the transformed curves subseries* $\mathbf{Z}_{t,l}(u)$ *arises as the sum of one dynamical component* $\widetilde{\mathbf{X}}_{t,l}(u)$ *and one white noise component* $\widetilde{\boldsymbol{\varepsilon}}_{t,l}(u)$:

$$\mathbf{Z}_{t,l}(u) = \mathbf{A}_l^{\mathrm{T}}\mathbf{X}_t(u) + \mathbf{A}_l^{\mathrm{T}}\boldsymbol{\varepsilon}_t(u) \equiv \widetilde{\mathbf{X}}_{t,l}(u) + \widetilde{\boldsymbol{\varepsilon}}_{t,l}(u), \tag{3.5}$$

which takes same form of (3.1) with $\widetilde{\mathbf{X}}_{t,l}(\cdot)$ *and* $\widetilde{\boldsymbol{\varepsilon}}_{t,l}(\cdot)$ *being uncorrelated and latent. Within each of the* q *groups, our second dimension reduction step applies techniques in Section 3.3 to estimated curve series of* $\mathbf{Z}_{t,l}(\cdot)$, *thus identifying the finite-dimensional structure of* $\widetilde{\mathbf{X}}_{tl}(\cdot)$ *and based on which predicting future values of* $\mathbf{Z}_{t,l}(\cdot)$. *According to the transformation in (3.2), we finally make predictions for original curve series* $\mathbf{Y}_t(\cdot)$. *See also Remark 3.6 and Section 3.3.4.*

3.2.2 Estimation procedure

To identify the latent segmentation structure in (3.2), we need to estimate $\mathbf{A} = (\mathbf{A}_1, \ldots, \mathbf{A}_q)$, or more precisely, the spanned linear spaces, $\mathcal{C}(\mathbf{A}_1), \ldots, \mathcal{C}(\mathbf{A}_q)$. Given some prespecified positive integer κ_0, let

$$\begin{aligned}\mathbf{W}_z &= \sum_{k=0}^{\kappa_0}\int_{\mathcal{U}_0}\int_{\mathcal{U}_0}\mathbf{\Sigma}_{z,k}(u,v)\{\mathbf{\Sigma}_{z,k}(u,v)\}^{\mathrm{T}}\,\mathrm{d}u\mathrm{d}v,\\ \mathbf{W}_y &= \sum_{k=0}^{\kappa_0}\int_{\mathcal{U}_0}\int_{\mathcal{U}_0}\mathbf{\Sigma}_{y,k}(u,v)\{\mathbf{\Sigma}_{y,k}(u,v)\}^{\mathrm{T}}\,\mathrm{d}u\mathrm{d}v.\end{aligned} \tag{3.6}$$

It follows from (3.4), (3.6) and the orthogonal constraint $\mathbf{A}\mathbf{A}^{\mathrm{T}} = \mathbf{I}_p$ that

$$\mathbf{W}_z = \mathbf{A}^{\mathrm{T}}\mathbf{W}_y\mathbf{A}. \tag{3.7}$$

To highlight the key idea, we first consider the case with $q = p$, i.e., $Z_{t1}(\cdot), \ldots, Z_{tp}(\cdot)$ are p uncorrelated functional time series across all time lags. Then $\mathbf{\Sigma}_{z,k}(u,v)$ in (3.4) reduces to a diagonal matrix, and so does $\mathbf{W}_z$. Hence the columns of $\mathbf{A}$ are

the eigenvectors of matrix $\mathbf{W}_y$; see (3.7). We show below that this is still the case when $q < p$. Noting that all $\mathbf{\Sigma}_{z,k}(u, v)$ for $k \geqslant 0$ and $u, v \in \mathcal{U}_0$, are block diagonal matrices with the main block sizes $p_1, \ldots, p_q$, it is easy to see from (3.6) that $\mathbf{W}_z$, is also a block diagonal matrix of the same sizes. Let $\mathbf{\Gamma}_z$ be the orthogonal matrix consisting of orthonormal eigenvectors of $\mathbf{W}_z$, i.e.,

$$\mathbf{W}_z \mathbf{\Gamma}_z = \mathbf{\Gamma}_z \mathbf{D}, \tag{3.8}$$

where $\mathbf{D}$ is a diagonal matrix consisting of p eigenvalues. Arranging the diagonal elements of $\mathbf{D}$ based on the order of the blocks in $\mathbf{W}_z$, $\mathbf{\Gamma}_z$ is also a block diagonal matrix of the same type as $\mathbf{W}_z$. Then (3.7) implies that

$$\mathbf{W}_y \mathbf{A}\mathbf{\Gamma}_z = \mathbf{A}\mathbf{W}_z\mathbf{\Gamma}_z = \mathbf{A}\mathbf{\Gamma}_z\mathbf{D}.$$

Thus the columns of $\mathbf{\Gamma}_y \equiv \mathbf{A}\mathbf{\Gamma}_z$ are the orthonormal eigenvectors of $\mathbf{W}_y$. Combining this with (3.2) yields that

$$\mathbf{\Gamma}_y^{\mathrm{T}} \mathbf{Y}_t(\cdot) = \mathbf{\Gamma}_z^{\mathrm{T}} \mathbf{A}^{\mathrm{T}} \mathbf{Y}_t(\cdot) = \mathbf{\Gamma}_z^{\mathrm{T}} \mathbf{Z}_t(\cdot). \tag{3.9}$$

Since $\mathbf{\Gamma}_z$ is a block diagonal orthogonal matrix with q blocks, $\mathbf{\Gamma}_z^{\mathrm{T}}\mathbf{Z}_t(\cdot)$ effectively applies orthogonal transformation within each of the q groups of $\mathbf{Z}_t(\cdot)$ such that neither contemporaneous nor serial correlation exists across the q groups in $\mathbf{\Gamma}_z^{\mathrm{T}}\mathbf{Z}_t(\cdot)$. This implies that knowing $\mathbf{\Gamma}_z^{\mathrm{T}}\mathbf{Z}_t(\cdot)$ is as good as knowing the latent segmentation $\mathbf{Z}_t(\cdot)$. Hence $\mathbf{\Gamma}_y$ can be seen as a column-permutation of the latent matrix $\mathbf{A}$. We summarize the above finding in a proposition below. Note that the order of eigenvectors in $\mathbf{\Gamma}_z$ is unknown.

Proposition 3.1. (i) *The orthogonal matrix $\mathbf{\Gamma}_z$ in (3.8) can be taken as a block-diagonal orthogonal matrix with the same block structure as $\mathbf{W}_z$.*

(ii) *An orthogonal matrix $\mathbf{\Gamma}_z$ satisfies (3.8) if and only if its columns are a permutation of the columns of a block-diagonal orthogonal matrix described in* (i)*, provided that any two different main diagonal blacks in $\mathbf{W}_z$ do not share the same eigenvalues.*

The above proposition appears to be the same as Proposition 1 of Chang et al. (2018), though it deals with a different segmentation problem. Its proof is therefore omitted.

To discover the latent segmentation $\mathbf{Z}_t$, we perform the following two steps of operations:

1. Let $\widehat{\mathbf{\Sigma}}_{y,k}(u,v)$ be the consistent estimator of $\mathbf{\Sigma}_{y,k}(u,v)$ for $k=1,\ldots,\kappa_0$. Replacing $\mathbf{\Sigma}_{y,k}(u,v)$ in (3.6) by $\widehat{\mathbf{\Sigma}}_{y,k}(u,v)$, we obtain an estimator of $\mathbf{W}_y$ via

$$\widehat{\mathbf{W}}_y = \sum_{k=0}^{\kappa_0} \int_{\mathcal{U}_0} \int_{\mathcal{U}_0} \widehat{\mathbf{\Sigma}}_{y,k}(u,v)\{\widehat{\mathbf{\Sigma}}_{y,k}(u,v)\}^{\mathrm{T}} \,\mathrm{d}u\mathrm{d}v, \tag{3.10}$$

   and calculate its orthonormal eigenvectors $(\widehat{\boldsymbol{\eta}}_1,\ldots,\widehat{\boldsymbol{\eta}}_p)$ corresponding to the ordered eigenvalues $\lambda_1(\widehat{\mathbf{W}}_y) \geqslant \cdots \geqslant \lambda_p(\widehat{\mathbf{W}}_y)$.

2. We construct $\widehat{\mathbf{A}} = (\widehat{\mathbf{A}}_1,\ldots,\widehat{\mathbf{A}}_q)$ with the corresponding columns being a permutation of $(\widehat{\boldsymbol{\eta}}_1,\ldots,\widehat{\boldsymbol{\eta}}_p)$ such that $\widehat{\mathbf{Z}}_t(\cdot) \equiv \widehat{\mathbf{A}}^{\mathrm{T}}\mathbf{Y}_t(\cdot)$ can be divided into q uncorrelated groups $\widehat{\mathbf{A}}_1^{\mathrm{T}}\mathbf{Y}_t(\cdot),\ldots,\widehat{\mathbf{A}}_q^{\mathrm{T}}\mathbf{Y}_t(\cdot)$.

Remark 3.3. *(i) We have developed the estimation procedure assuming that the number of groups q is known or can be identified correctly. In practice, q is unknown. We will see in Sections 3.2.3 and 3.4 that q as well as the segmentation structure of $\mathbf{Z}_t(\cdot)$ can be well estimated.*
(ii) To ensure $\widehat{\mathbf{A}}$ in Step 2 a valid estimator in the sense that $\mathcal{C}(\widehat{\mathbf{A}}_j)$ is consistent to $\mathcal{C}(\mathbf{A}_j)$ for $j=1,\ldots,q$, it is essential to make use of the consistent estimators for $\mathbf{\Sigma}_{y,k}(u,v)$ in Step 1 under different asymptotic scenarios. When p is fixed or $p=o(n^{-1/2})$, the sample versions of $\mathbf{\Sigma}_{y,k}(u,v)$ for $k=1,\ldots,\kappa_0$,

$$\widehat{\mathbf{\Sigma}}^{\mathrm{S}}_{y,k}(u,v) = \frac{1}{n-k}\sum_{t=1}^{n-k}\{\mathbf{Y}_t(u)-\bar{\mathbf{Y}}(u)\}\{\mathbf{Y}_{t+k}(v)-\bar{\mathbf{Y}}(v)\}^{\mathrm{T}}, \tag{3.11}$$

are consistent and hence can be plugged in (3.10) to obtain $\widehat{\mathbf{W}}_y$. When p grows faster than $n^{1/2}$, some sparsity assumptions on $\mathbf{A}$ can be imposed, which facilitates the development of a thresholded estimator for $\mathbf{\Sigma}_{y,k}(u,v)$ to retain the consistency. See details in Section 3.2.4.
(iii) Note that $\mathbf{W}_y$ in (3.6) relies on the double integral and the sum to accumulate the (auto)covariance information as much as possible from each $(u,v) \in \mathcal{U}_0^2$ and from lags $k=0$ to κ_0, whereas fixing at certain (u,v) or time lag may lead to spurious estimation results. One can define an alternative positive definite matrix by integrating along the diagonal path $u=v \in \mathcal{U}_0$,

$$\breve{\mathbf{W}}_y = \sum_{k=0}^{\kappa_0} \int_{\mathcal{U}_0} \mathbf{\Sigma}_{y,k}(u,u)\{\mathbf{\Sigma}_{y,k}(u,u)\}^{\mathrm{T}} \,\mathrm{d}u.$$

However $\breve{\mathbf{W}}_y$ suffers from the loss of (auto)covariance information for $u \neq v$ incurred by the single integral. Note that all integrated terms in $\mathbf{W}_y$ are non-negative

definite. Hence there is no information cancellation over different lags, leading to the results insensitive to the choice of κ_0*. In practice a small* κ_0 *is often sufficient, provided that the first* κ_0 *lags carry sufficient information on the block diagonal structure, while enlarging* κ_0 *will add more noise to estimate* $\mathbf{W}_y$*.*

3.2.3 Permutation

In this section, we adopt a maximum cross-(auto)covariance method to divide the components of $\widehat{\mathbf{Z}}_t(\cdot)$ into q uncorrelated groups, where q and the group sizes $p_1, \ldots, p_q$ are unknown. Recall that for curves $Z_{ti}(\cdot)$ and $Z_{tj}(\cdot)$ within the same latent group, one would expect that the lag-k cross-(auto)covariance function, that is $\Sigma^{(k)}_{z,ij}(u,v) = \mathrm{Cov}\{Z_{ti}(u), Z_{(t+k)j}(v)\}$ to be significantly different from zero for some k, and $u, v \in \mathcal{U}_0$, thus leading to at least one large $\|\Sigma^{(k)}_{z,ij}\|_{\mathcal{S}}$, where $\|\cdot\|_{\mathcal{S}}$ denotes the Hilbert–Schmidt norm, i.e., for any $\mathcal{B} \in \mathbb{S} = L^2(\mathcal{U}_0) \otimes L^2(\mathcal{U}_0)$, $\|\mathcal{B}\|_{\mathcal{S}} = \{\int_{\mathcal{U}_0}\int_{\mathcal{U}_0} \mathcal{B}(u,v)^2 \mathrm{d}u\mathrm{d}v\}^{1/2}$. Let $\widehat{Z}_{tj}(\cdot) = \widehat{\boldsymbol{\eta}}_j^{\mathrm{T}} \mathbf{Y}_t(\cdot)$ for $j = 1, \cdots, p$. Inspired by the above fact, we define the maximum cross-(auto)covariance over the lags between prespecified $-m$ and m for any pair $i \neq j$ under the Hilbert–Schmidt norm as

$$\widehat{T}_{ij} = \max_{|k| \leqslant m} \|\widehat{\Sigma}^{(k)}_{z,ij}\|_{\mathcal{S}} = \max_{0 \leqslant k \leqslant m} \max\left\{\|\widehat{\boldsymbol{\eta}}_i^{\mathrm{T}} \widehat{\boldsymbol{\Sigma}}_{y,k} \widehat{\boldsymbol{\eta}}_j\|_{\mathcal{S}}, \|\widehat{\boldsymbol{\eta}}_i^{\mathrm{T}} (\widehat{\boldsymbol{\Sigma}}_{y,k})^{\mathrm{T}} \widehat{\boldsymbol{\eta}}_j\|_{\mathcal{S}}\right\}, \tag{3.12}$$

and take $\widehat{Z}_{ti}(\cdot)$ and $\widehat{Z}_{tj}(\cdot)$ as a pair of significantly cross-correlated curves if $\widehat{T}_{ij}$ is greater than an appropriate threshold level $\tau > 0$. To be specific, we rearrange $p_0 = p(p-1)/2$ values of $\widehat{T}_{ij}$ $(1 \leqslant i < j \leqslant p)$ in the descending order, $\widehat{T}_{(1)} \geqslant \cdots \geqslant \widehat{T}_{(p_0)}$ and compute

$$\hat{\varrho} = \underset{1 \leqslant j \leqslant c_\varrho p_0}{\mathrm{argmax}} \widehat{T}_{(j)} / \widehat{T}_{(j+1)}, \tag{3.13}$$

where $c_\varrho \in (0,1)$ is a prescribed constant. Corresponding to $\widehat{T}_{(1)}, \ldots, \widehat{T}_{(\hat{\varrho})}$, we identify $\hat{\varrho}$ pairs of cross-correlated curves.

Remark 3.4. *The intuition behind (3.13) is clear as follows. Let* $T_{(1)} \geqslant \cdots \geqslant T_{(p_0)}$ *be the ordered true cross-(auto)covariances under the Hilbert–Schmidt norm. Suppose there are only* ϱ *cross-correlated pairs among total* p_0 *pairs, i.e.,* $T_{(\varrho)} > 0$ *and* $T_{(\varrho+1)} = 0$*, then the ratio* $T_{(j)}/T_{(j+1)} = \infty$ *for* $j = \varrho$*. To avoid the case of arbitrary large ratio* $\widehat{T}_{(j)}/\widehat{T}_{(j+1)}$ *for* $j > \varrho$*, we consider* $c_\varrho \in (0,1)$*. See other applications of such ratio-based estimator in Lam and Yao (2012); Ahn and Horenstein (2013) and Chang et al. (2018).*

The permutation in Step 2 can then be performed as follows. We start with p groups with each $\widehat{Z}_{tj}(\cdot)$ in one group and then repeatedly merge two groups if two cross-correlated curves are split over the two groups. The iteration is terminated until all the cross-correlated pairs are within one group. Hence we obtain the estimated group structure of $\widehat{\mathbf{Z}}_t(\cdot)$ with the number of the final groups $\hat{q}$ being the estimated value for q. It is worth mentioning that in practice, one could repeat the permutation step multiple times within each of the discovered groups to enhance the accuracy of segmentations when p is large. See Section 3.5.2 for more technical details and empirical evidence to support this proposal.

3.2.4 Functional thresholding

Our problem of interest now becomes how to estimate $\mathbf{\Sigma}_{y,k}$ for $k = 0, \ldots, \kappa_0$ and $\mathbf{W}_y$ in (3.6) consistently. When p diverges faster than $n^{1/2}$, the sample (auto)covariance function $\widehat{\mathbf{\Sigma}}^{\mathrm{S}}_{y,k}$ in (3.11) is no longer a consistent estimator for $\mathbf{\Sigma}_{y,k}$. In such high-dimensional case, we impose the sparsity condition below on the transformation matrix $\mathbf{A}$.

Condition 3.1. *For* $\mathbf{A} = (A_{ij})_{p\times p}$, *there has some constant* $\alpha \in [0, 1)$, *such that* $\max_{1\leqslant i\leqslant p} \sum_{j=1}^{p} |A_{ij}|^{\alpha} \leqslant s_1$ *and* $\max_{1\leqslant j\leqslant p} \sum_{i=1}^{p} |A_{ij}|^{\alpha} \leqslant s_2$.

The parameters s_1 and s_2 determine the row and column sparsity levels of $\mathbf{A}$, respectively. We may allow s_1 and s_2 to grow at slow rates as p increases. The row sparsity with small s_1 entails that each component of $\mathbf{Y}_t(\cdot)$ is a linear combination of a small number of components in $\mathbf{Z}_t(\cdot)$, while the column sparsity with small s_2 corresponds to the case that each $Z_{tj}(\cdot)$ has impact on only a few components of $\mathbf{Y}_t(\cdot)$. The parameter α also controls the sparsity level of $\mathbf{A}$ with a smaller value yielding a sparser $\mathbf{A}$. It can be inferred from the fact $\mathbf{\Sigma}_{y,k}(u, v) = \mathbf{A}\mathbf{\Sigma}_{z,k}(u, v)\mathbf{A}^{\mathrm{T}}$ that our imposed sparsity constraint in $\mathbf{A}$ is inherited by the functional sparsity structure in $\mathbf{\Sigma}_{y,k}$ as justified in the following lemma.

Lemma 3.1. *Let* $\mathbf{\Sigma}_{y,k}(u, v) = \{\Sigma^{(k)}_{y,ij}(u, v)\}_{p\times p}$. *Under Condition 3.1,* $\sum_{i=1}^{p} \|\Sigma^{(k)}_{y,ij}\|^{\alpha}_{\mathcal{S}} \leqslant \Xi$ *and* $\sum_{j=1}^{p} \|\Sigma^{(k)}_{y,ij}\|^{\alpha}_{\mathcal{S}} \leqslant \Xi$ *for* $k = 1, \ldots, \kappa_0$, *where* $\Xi = s_1 s_2\big(2\max_{1\leqslant l\leqslant q} p_l + 1\big)$.

Lemma 3.1 relies on the Hilbert–Schmidt norm to encourage the functional sparsity pattern in $\mathbf{\Sigma}_{y,k}$ in the sense that, for some i, j, k, $\|\Sigma^{(k)}_{y,ij}\|_{\mathcal{S}} = 0$ if and only if $\Sigma^{(k)}_{y,ij}(u, v) = 0$ for all $u, v \in \mathcal{U}_0$. This lemma reveals that the functional sparsity structures in columns/rows of $\mathbf{\Sigma}_{y,k}$ are determined by s_1, s_2 and α with smaller values yielding functional sparser $\mathbf{\Sigma}_{y,k}$.

Inspired by the spirit of thresholded estimator for large covariance matrix (Bickel and Levina, 2008), we apply the functional thresholding rule, that combines the functional generalizations of hard thresholding and shrinkage with the aid of the Hilbert–Schmidt norm of functions, on entries of the sample autocovariance function $\widehat{\mathbf{\Sigma}}^{\mathrm{S}}_{y,k}(u,v) = \{\widehat{\Sigma}^{\mathrm{S}}_{y,k,ij}(u,v)\}_{p\times p}$ in (3.11). Hence we construct the functional thresholding estimator by

$$\mathcal{T}_{\omega_k}(\widehat{\mathbf{\Sigma}}^{\mathrm{S}}_{y,k})(u,v) = \left[\widehat{\Sigma}^{\mathrm{S}}_{y,k,ij}(u,v) I\{\|\widehat{\Sigma}^{\mathrm{S}}_{y,k,ij}\|_{\mathcal{S}} \geqslant \omega_k\}\right]_{p\times p}, \quad u,v \in \mathcal{U}_0, \tag{3.14}$$

where $I(\cdot)$ is the indicator function and $\omega_k \geqslant 0$ is the thresholding parameter at lag k. Taking $\widehat{\mathbf{\Sigma}}_{y,k}$ in (3.10) as $\mathcal{T}_{\omega_k}(\widehat{\mathbf{\Sigma}}^{\mathrm{S}}_{y,k})$ yields

$$\widehat{\mathbf{W}}_y = \sum_{k=1}^{\kappa_0} \int_{\mathcal{U}_0}\int_{\mathcal{U}_0} \mathcal{T}_{\omega_k}(\widehat{\mathbf{\Sigma}}^{\mathrm{S}}_{y,k})(u,v)\{\mathcal{T}_{\omega_k}(\widehat{\mathbf{\Sigma}}^{\mathrm{S}}_{y,k})(u,v)\}^{\mathrm{T}}\,\mathrm{d}u\mathrm{d}v. \tag{3.15}$$

Remark 3.5. *The thresholding parameter ω_k for each $k = 1,\ldots,\kappa_0$ can be selected using a K-fold cross-validation approach. Specifically, we sequentially divide the set $\{1,\ldots,n\}$ into K validation sets $V_1,\ldots,V_K$ of approximately equal size. For each $j = 1,\ldots,K$, let $\widehat{\mathbf{\Sigma}}^{\mathrm{S},(j)}_{y,k}(u,v)$ and $\widehat{\mathbf{\Sigma}}^{\mathrm{S},(-j)}_{y,k}(u,v)$ be the sample lag-k autocovaraince functions based on the j-th validation set $\{\mathbf{Y}_t(\cdot) : t \in V_j\}$ and the remaining $K-1$ sets $\{\mathbf{Y}_t(\cdot) : t \in V_i, i \neq j\}$, respectively. We select the optimal $\hat{\omega}_k$ by minimizing*

$$R(\omega_k) = \frac{1}{K}\sum_{j=1}^{K} \|\mathcal{T}_{\omega_k}(\widehat{\mathbf{\Sigma}}^{\mathrm{S},(j)}_{y,k}) - \widehat{\mathbf{\Sigma}}^{\mathrm{S},(-j)}_{y,k}\|^2_{\mathcal{S},F},$$

where $\|\cdot\|_{\mathcal{S},F}$ denotes the functional version of Frobenius norm, i.e, for any $\mathcal{B} = (\mathcal{B}_{ij})_{p\times p}$ with each $\mathcal{B}_{ij} \in \mathbb{S}$, $\|\mathcal{B}\|_{\mathcal{S},F} = (\sum_{i,j}\|\mathcal{B}_{ij}\|^2_{\mathcal{S}})^{1/2}$. Given the time break from the leave-out validation set, the autocovariance estimation based on the remaining $K-1$ groups is affected by κ_0 misutilized lagged terms. However, this effect is negligible especially for sufficiently large n.

3.3 Estimate finite-dimensional structure

3.3.1 Model setting

Recall that the transformed curve subseries $\mathbf{Z}_{t,l}(u)$ is expressed as the sum of two uncorrelated and latent components, one finite-dimensional dynamic and one

white noise, in the sense of (3.5). The goal of our second dimension reduction step is to identify the finite-dimensional structure of $\widetilde{\mathbf{X}}_{t,l}(u)$ and based on which to predict future values of $\mathbf{Z}_{t,l}(u)$. Observe that both $\mathbf{Z}_{t,l}(u)$ in (3.5) and $\mathbf{Y}_t(u)$ in (3.1) are decomposed in the same form. To present the methodology, we focus on a general model setup in this section for p-vector of functional time series $\mathbf{Y}_t(\mathbf{u}) = \{Y_{t1}(u_1), \ldots, Y_{tp}(u_p)\}^{\mathrm{T}}$ satisfying

$$\mathbf{Y}_t(\mathbf{u}) = \mathbf{X}_t(\mathbf{u}) + \boldsymbol{\varepsilon}_t(\mathbf{u}), \quad \mathbf{u} = (u_1, \ldots, u_p) \in \mathcal{U} = \mathcal{U}_1 \times \cdots \times \mathcal{U}_p, \tag{3.16}$$

where $\mathbf{X}_t(\mathbf{u}) = \{X_{t1}(u_1), \ldots, X_{tp}(u_p)\}^{\mathrm{T}}$ is the latent and finite-dimensional dynamical component, and is uncorrelated with the white noise component $\boldsymbol{\varepsilon}_t(\mathbf{u}) = \{\varepsilon_{t1}(u_1), \ldots, \varepsilon_{tp}(u_p)\}^{\mathrm{T}}$. Note that (3.16) takes the same form of (3.1) but allows different component curves to be defined on the different sets $\mathcal{U}_1, \ldots, \mathcal{U}_p$, which are all subintervals of the real line.

Remark 3.6. *Under this general setting when p is small, say $p \leqslant 3$, we can directly perform the dimension reduction approach developed in Section 3.3 on observed $\mathbf{Y}_t(\mathbf{u})$ to estimate the finite-dimensional structure of $\mathbf{X}_t(\mathbf{u})$ and based on which to predict future values of $\mathbf{Y}_t(\mathbf{u})$. The appealing of this approach is that the dimensionality of $\mathbf{X}_t(\mathbf{u})$ is small. When p is large or moderately large, it is unrealistic to expect this dimensionality to be still small. Therefore, under a common setting in practice where all component series of $\mathbf{Y}_t(\cdot)$ share the same support $\mathcal{U}_0$, we first adopt the segmentation step in Section 3.2 on $\mathbf{Y}_t(\cdot)$ such that the transformed curve series are segmented into q uncorrelated groups. Then, within each group $l = 1, \ldots, q$, the second dimension reduction step applies the methodology in this section to estimated transformed curve subseries*

$$\widehat{\mathbf{Z}}_{t,l}(u) = \widehat{\mathbf{A}}_l^{\mathrm{T}} \mathbf{Y}_t(u), \quad u \in \mathcal{U}_0, \tag{3.17}$$

See details in Section 3.3.4 below.

Under (3.16), we also assume that both $\boldsymbol{\mu}(\mathbf{u}) \equiv \mathbb{E}\{\mathbf{X}_t(\mathbf{u})\}$ and

$$\mathbf{M}_k(\mathbf{u}, \mathbf{v}) \equiv \mathbb{E}[\{\mathbf{X}_t(\mathbf{u}) - \boldsymbol{\mu}(\mathbf{u})\}\{\mathbf{X}_{t+k}(\mathbf{v}) - \boldsymbol{\mu}(\mathbf{v})\}^{\mathrm{T}}] \quad (k = 0, 1, \ldots), \tag{3.18}$$

do not depend on t and $\int_{\mathcal{U}} \mathbb{E}\big[\{\mathbf{X}_t(\mathbf{u})\}^{\mathrm{T}} \mathbf{X}_t(\mathbf{u}) + \{\boldsymbol{\varepsilon}_t(\mathbf{u})\}^{\mathrm{T}} \boldsymbol{\varepsilon}_t(\mathbf{u})\big] \, \mathrm{d}\mathbf{u} < \infty$. Then $\mathbf{X}_t(\mathbf{u})$ admits the Karhunen–Loéve expansion:

$$\mathbf{X}_t(\mathbf{u}) - \boldsymbol{\mu}(\mathbf{u}) = \sum_{j=1}^{\infty} \xi_{tj} \boldsymbol{\varphi}_j(\mathbf{u}), \quad \mathbf{u} \in \mathcal{U}, \tag{3.19}$$

where $\boldsymbol{\varphi}_j(\mathbf{u}) = \{\varphi_{j1}(u_1), \ldots, \varphi_{jp}(u_p)\}^{\mathrm{T}}$ for $j \geqslant 1$ are a sequence of deterministic orthonormal functions in $L^2(\mathcal{U})$ satisfying $\int_{\mathcal{U}} \{\boldsymbol{\varphi}_i(\mathbf{u})\}^{\mathrm{T}} \boldsymbol{\varphi}_j(\mathbf{u})\, \mathrm{d}\mathbf{u} = 1$ if $i = j$ and 0 otherwise, and ξ_{tj} are scalar random variables defined as $\xi_{tj} = \int_{\mathcal{U}} \{\mathbf{X}_t(\mathbf{u}) - \boldsymbol{\mu}(\mathbf{u})\}^{\mathrm{T}} \boldsymbol{\varphi}_j(\mathbf{u})\, \mathrm{d}\mathbf{u}$ for $j \geqslant 1$. Furthermore it holds that $\mathbb{E}(\xi_{tj}) = 0$, $\mathrm{Var}(\xi_{tj}) = \lambda_j$ and $\mathrm{Cov}(\xi_{ti}, \xi_{tj}) = 0$ for $i \neq j$. We rank $\{\xi_{tj}\}_{j \geqslant 1}$ such that $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant 0$. We refer to Chiou et al. (2014) and Happ and Greven (2018) for further details on the Karhunen–Loéve expansion for multivariate functional data. Note that $\{\boldsymbol{\varphi}_j(\mathbf{u})\}_{j \geqslant 1}$ are the eigenfunctions of the covariance function $\mathbf{M}_0(\mathbf{u}, \mathbf{v})$ defined in (3.18). They do not reflect the serial correlations of the curves across different times.

Let $L^2(\mathcal{U}_j)$ be a Hilbert space of squared integrable functions defined on $\mathcal{U}_j$ for $j = 1, \ldots, p$. We denote the p-fold Cartesian product by $L^2(\mathcal{U}) = L^2(\mathcal{U}_1) \times \cdots \times L^2(\mathcal{U}_p)$. For $\boldsymbol{\mathcal{F}}, \boldsymbol{\mathcal{G}} \in L^2(\mathcal{U})$, we denote the inner product by

$$\langle \boldsymbol{\mathcal{F}}, \boldsymbol{\mathcal{G}} \rangle = \int_{\mathcal{U}} \{\boldsymbol{\mathcal{F}}(\mathbf{u})\}^{\mathrm{T}} \boldsymbol{\mathcal{G}}(\mathbf{u})\, \mathrm{d}\mathbf{u}, \tag{3.20}$$

with induced norm $\|\cdot\| = \langle \cdot, \cdot \rangle^{1/2}$. Note $\mathbf{M}_k$ in (3.18) can be regarded as the kernel of an induced linear operator acting on $L^2(\mathcal{U})$ in the sense that it maps $\boldsymbol{\mathcal{G}}(\cdot) \in L^2(\mathcal{U})$ to $\mathbf{M}_k(\boldsymbol{\mathcal{G}})(\cdot) = \int_{\mathcal{U}} \mathbf{M}_k(\cdot\,, \mathbf{v}) \boldsymbol{\mathcal{G}}(\mathbf{v})\, \mathrm{d}\mathbf{v} \in L^2(\mathcal{U})$. To abuse the notation a bit, we use $\mathbf{M}_k$ to denote both the kernel function and the induced operator. Provided that $\{(\lambda_j, \boldsymbol{\varphi}_j(\mathbf{u}))\}_{j \geqslant 1}$ are the eigenpairs of $\mathbf{M}_0(\mathbf{u}, \mathbf{v})$, it then holds that

$$\int_{\mathcal{U}} \mathbf{M}_0(\mathbf{u}, \mathbf{v}) \boldsymbol{\varphi}_j(\mathbf{v})\, \mathrm{d}\mathbf{v} = \lambda_j \boldsymbol{\varphi}_j(\mathbf{u}), \quad \mathbf{u} \in \mathcal{U} \ \text{ and } \ j \geqslant 1. \tag{3.21}$$

Remark 3.7. *Now we give an heuristic interpretation of inner product (3.20) and eigen-equation (3.21), which lead to a simple and direct way to calculate eigenvalues and eigenfunctions in Section 3.3.3 below. If we view each element in $L^2(\mathcal{U})$ as a matrix with p columns and each column being a curve, (3.20) can be viewed as the 'standard' inner product for vectors applying to the long vectors obtained by stacking the p 'columns' together for each element of $L^2(\mathcal{U})$. On the other hand, $\mathbf{M}_0(\mathbf{u}, \mathbf{v}) = \big[\mathrm{Cov}\{X_{ti}(u_i), X_{tj}(v_j)\}\big]_{p \times p}$, see (3.18). Now we view each curve $X_{ti}(\cdot)$ as an infinitely long vector, and we inflate $\mathbf{M}_0$ above by replacing its (i, j)-th element by $\mathrm{Cov}\{X_{ti}(\cdot), X_{tj}(\cdot)\}$, a covariance matrix of two infinitely long vectors. Then eigen-equation (3.21) may be conceptually 'recasted' as $\mathbf{M}_0 \mathbf{b}_j = \lambda_j \mathbf{b}_j$ $(j = 1, \ldots, r)$, where $\mathbf{M}_0$ is a block matrix with infinite sizes and its (i, j)-th block is $\mathrm{Cov}\{X_{ti}(\cdot), X_{tj}(\cdot)\}$, and $\mathbf{b}_j$ is a long vector obtained by stacking all the component curves of $\boldsymbol{\varphi}_j(\cdot)$ together. This transforms the eigen-problem in a space of curve bundles into an eigen-problem of a matrix (of infinite sizes).*

The goal of the dimension reduction approach is to model and predict $\mathbf{Y}_t(\mathbf{u})$ based on some latent low-dimensional structure when p is small. We assume that $\mathbf{X}_t(\mathbf{u})$ is r-dimensional in the sense that $\lambda_r > 0$ and $\lambda_{r+1} = 0$ (Bathia et al., 2010). We then estimate this low-dimensional structure based on autocovariance of the curve series. Specifically, this requires to identify r and to estimate the dynamic space, $\mathcal{C}(\boldsymbol{\varphi})$, spanned by the orthonormal functions $\boldsymbol{\varphi}_1(\mathbf{u}), \ldots, \boldsymbol{\varphi}_r(\mathbf{u})$. See Sections 3.3.2 and 3.3.3 below.

3.3.2 Methodology

When $\mathbf{X}_t(\cdot)$ is r-dimensional, it follows from (3.16) and (3.19) that

$$\mathbf{Y}_t(\mathbf{u}) = \boldsymbol{\mu}(\mathbf{u}) + \sum_{j=1}^{r} \xi_{tj}\, \boldsymbol{\varphi}_j(\mathbf{u}) + \boldsymbol{\varepsilon}_t(\mathbf{u}). \tag{3.22}$$

Given some prescribed positive integer k_0, put

$$\widehat{\mathbf{M}}_k(\mathbf{u}, \mathbf{v}) = \frac{1}{n - k_0} \sum_{t=1}^{n-k_0} \{\mathbf{Y}_t(\mathbf{u}) - \bar{\mathbf{Y}}(\mathbf{u})\}\{\mathbf{Y}_{t+k}(\mathbf{v}) - \bar{\mathbf{Y}}(\mathbf{v})\}^{\mathrm{T}} \quad \text{for } k = 0, 1, \ldots, k_0, \tag{3.23}$$

where $\bar{\mathbf{Y}}(\cdot) = n^{-1} \sum_{t=1}^{n} \mathbf{Y}_t(\cdot)$. Based on (3.21), a natural way to estimate the finite-dimensional structure (3.22) is to perform an eigenanalysis for the operator $\widehat{\mathbf{M}}_0$. Unfortunately, $\widehat{\mathbf{M}}_0$ is not a consistent estimator for $\mathbf{M}_0$, as $\mathrm{Cov}\{\mathbf{Y}_t(\mathbf{u}), \mathbf{Y}_t(\mathbf{v})\} = \mathbf{M}_0(\mathbf{u}, \mathbf{v}) + \mathrm{Cov}\{\boldsymbol{\varepsilon}_t(\mathbf{u}), \boldsymbol{\varepsilon}_t(\mathbf{v})\}$. Motivated from the fact that $\mathrm{Cov}\{\mathbf{Y}_t(\mathbf{u}), \mathbf{Y}_{t+k}(\mathbf{v})\} = \mathbf{M}_k(\mathbf{u}, \mathbf{v})$ for any $k \geqslant 1$, which ensures that $\widehat{\mathbf{M}}_k$ is a legitimate estimator for $\mathbf{M}_k$, we proceed to estimate (3.22) based on $\widehat{\mathbf{M}}_k$ for $k \geqslant 1$ instead.

Define a nonnegative operator to pull together the autocovariance information at different lags:

$$\mathbf{K}(\mathbf{u}, \mathbf{v}) = \sum_{k=1}^{k_0} \int_{\mathcal{I}} \mathbf{M}_k(\mathbf{u}, \mathbf{w})\{\mathbf{M}_k(\mathbf{v}, \mathbf{w})\}^{\mathrm{T}}\, \mathrm{d}\mathbf{w}, \quad \mathbf{u}, \mathbf{v} \in \mathcal{U}, \tag{3.24}$$

whose sample counterpart is

$$\begin{aligned}\widehat{\mathbf{K}}(\mathbf{u},\mathbf{v}) &= \sum_{k=1}^{k_0}\int_{\mathcal{I}} \widehat{\mathbf{M}}_k(\mathbf{u},\mathbf{w})\{\widehat{\mathbf{M}}_k(\mathbf{v},\mathbf{w})\}^{\mathrm{T}}\,\mathrm{d}\mathbf{w} \\ &= \frac{1}{(n-k_0)^2}\sum_{t,s=1}^{n-k_0}\sum_{k=1}^{k_0}\{\mathbf{Y}_t(\mathbf{u})-\bar{\mathbf{Y}}(\mathbf{u})\}\{\mathbf{Y}_s(\mathbf{v})-\bar{\mathbf{Y}}(\mathbf{v})\}^{\mathrm{T}}\langle \mathbf{Y}_{t+k}-\bar{\mathbf{Y}}, \mathbf{Y}_{s+k}-\bar{\mathbf{Y}}\rangle,\end{aligned} \tag{3.25}$$

see (3.23). Note the non-negativity of $\mathbf{K}(\mathbf{u},\mathbf{v})$ ensures no cancellation of the information accumulated from lags 1 to k_0, thus making the estimation not sensitive to the choice of k_0. In practice, we tend to select small k_0, as the strongest autocorrections usually appears at the small time lags.

Let $\boldsymbol{\xi}_t = (\xi_{t1},\ldots,\xi_{tr})^{\mathrm{T}}$ and $\boldsymbol{\Omega}_k = \mathbb{E}(\boldsymbol{\xi}_t\boldsymbol{\xi}_{t+k}^{\mathrm{T}})$. Denote by $\boldsymbol{\psi}_j(\mathbf{u}) = \{\psi_{j1}(u_1),\ldots,\psi_{jp}(u_p)\}^{\mathrm{T}}$ for $j=1,\ldots,r$ the orthonormal eigenfunctions of $\mathbf{K}(\mathbf{u},\mathbf{v})$ corresponding to the r largest eigenvalues $\theta_1 \geqslant \cdots \geqslant \theta_r > 0$ of $\mathbf{K}(\mathbf{u},\mathbf{v})$. Let $\mathcal{C}(\boldsymbol{\psi}) = \mathrm{span}\{\boldsymbol{\psi}_1(\mathbf{u}),\ldots,\boldsymbol{\psi}_r(\mathbf{u})\}$. Then it follows from Proposition 1 of Bathia et al. (2010) that operator $\mathbf{K}$ only has r positive eigenvalues with $\theta_i = 0$ for $i \geqslant r+1$ under model (3.22), and $\mathcal{C}(\boldsymbol{\psi}) = \mathcal{C}(\boldsymbol{\varphi})$, provided that $\boldsymbol{\Omega}_k$ is a full-ranked matrix for some $k \leqslant k_0$. Therefore, $\mathbf{X}_t(\cdot)$ can be expanded using r basis functions $\boldsymbol{\psi}_1(\cdot),\ldots,\boldsymbol{\psi}_r(\cdot)$, i.e.,

$$\mathbf{X}_t(\mathbf{u}) - \boldsymbol{\mu}(\mathbf{u}) = \sum_{j=1}^{r}\zeta_{tj}\boldsymbol{\psi}_j(\mathbf{u}), \quad \mathbf{u}\in\mathcal{U},$$

where $\zeta_{tj} = \langle \mathbf{X}_t - \boldsymbol{\mu}, \boldsymbol{\psi}_j\rangle$. As a result, the linear dynamic structure of $\mathbf{Y}_t(\cdot)$ is fully captured by that of the r-dimensional vector process $\boldsymbol{\zeta}_t = (\zeta_{t1},\ldots,\zeta_{tr})^{\mathrm{T}}$.

Based on the above findings, we propose the following procedure to predict $\mathbf{Y}_t(\cdot)$ consisting of three steps:

a. Carry out an eigenanalysis on $\widehat{\mathbf{K}}(\mathbf{u},\mathbf{v})$ to obtain $\hat{r}$ non-zero estimated eigenvalues and estimated eigenfunctions $\widehat{\boldsymbol{\psi}}_1(\mathbf{u}),\ldots,\widehat{\boldsymbol{\psi}}_{\hat{r}}(\mathbf{u})$, see details in Section 3.3.3. The corresponding estimated coefficients are $\hat{\zeta}_{tj} = \langle \mathbf{Y}_t - \bar{\mathbf{Y}}, \boldsymbol{\psi}_j\rangle$ for $j=1,\ldots,\hat{r}$.

b. Model the $\hat{r}$-dimensional vector process $\hat{\boldsymbol{\zeta}}_t = (\hat{\zeta}_{t1},\ldots,\hat{\zeta}_{t\hat{r}})^{\mathrm{T}}$ based on VAR or other vector time series models and obtain h-step ahead prediction as $\hat{\boldsymbol{\zeta}}_{t+h}$.

c. Recover h-step ahead functional prediction as

$$\widehat{\mathbf{Y}}_{t+h}(\mathbf{u}) = \bar{\mathbf{Y}}(\mathbf{u}) + \sum_{j=1}^{\hat{r}}\hat{\zeta}_{(t+h)j}\hat{\boldsymbol{\psi}}_j(\mathbf{u}), \quad h\geqslant 0. \tag{3.26}$$

3.3.3 Eigenanalysis and estimation of r

Performing an eigenanalysis in a Hilbert space consisting of multiple curves is not a trivial matter. To overcome the difficulties due to multiple curves, Happ and Greven (2018) first calculate each of p component eigenfunctions separately using the existing methods for univariate curves and then construct the required multiple eigenfunctions from those univariate ones based on the theory of integral equations (Zemyan, 2012).

Drawing the inspiration from Remark 3.7 above, we propose a simple and direct method for estimating the eigenfunctions of operator $\mathbf{K}$. We transform the eigenanalysis for $\mathbf{K}$ to that for a matrix (of finite sizes) based on a well-known matrix duality property that $\mathbf{AB}^{\mathrm{T}}$ and $\mathbf{B}^{\mathrm{T}}\mathbf{A}$ share the same non-zero eigenvalues for any matrices $\mathbf{A}$ and $\mathbf{B}$ of the same sizes, which also holds for operators in a Hilbert space. When $d = 1$, the proposal reduces to the method of Bathia et al. (2010) for univariate functional time series.

We present a heuristic argument first. To view the operator $\widehat{\mathbf{K}}(\cdot,\cdot)$ defined in (3.25) as $\mathbf{AB}^{\mathrm{T}}$, denote by $\mathbf{y}_t$ the vector of infinite length obtained by stacking p curves $Y_{t1}(\cdot) - \bar{Y}_1(\cdot), \ldots, Y_{tp}(\cdot) - \bar{Y}_p(\cdot)$ together. Here we view each curve $Y_{tj}(\cdot)$ as an infinitely long vector, and thus $\mathbf{y}_t$ is viewed as consisting of p infinitely long vectors. Naturally we view $\mathbf{y}_t^{\mathrm{T}}\mathbf{y}_s$ as $\langle \mathbf{Y}_t - \bar{\mathbf{Y}}, \mathbf{Y}_s - \bar{\mathbf{Y}} \rangle$. Put $\mathcal{Y}_k = (\mathbf{y}_{1+k}, \ldots, \mathbf{y}_{n-k_0+k})$ for $k = 0, 1, \ldots, k_0$. Then $\widehat{\mathbf{K}}$ may be represented by an $\infty \times \infty$ matrix

$$\widehat{\mathbf{K}} = \frac{1}{(n-k_0)^2} \mathcal{Y}_0 \sum_{k=1}^{k_0} \mathcal{Y}_k^{\mathrm{T}} \mathcal{Y}_k \mathcal{Y}_0^{\mathrm{T}}.$$

Furthermore $\widehat{\boldsymbol{\psi}}(\mathbf{u}) = \{\hat{\psi}_1(u_1), \cdots, \hat{\psi}_p(u_p)\}^{\mathrm{T}}$ is an eigenfunction of $\widehat{\mathbf{K}}$ if and only if the $\infty \times 1$ vector obtained by stacking p curves $\hat{\psi}_1(\cdot), \cdots, \hat{\psi}_p(\cdot)$ together is the eigenvector of $\widehat{\mathbf{K}}$. Now applying the aforementioned duality, $\widehat{\mathbf{K}}$ shares the same non-zero eigenvalues with $(n-k_0) \times (n-k_0)$ matrix

$$\check{\mathbf{K}} = \frac{1}{(n-k_0)^2} \sum_{k=1}^{k_0} \mathcal{Y}_k^{\mathrm{T}} \mathcal{Y}_k \mathcal{Y}_0^{\mathrm{T}} \mathcal{Y}_0. \tag{3.27}$$

For each $j = 1, \ldots, r$, let $\boldsymbol{\gamma}_j = (\gamma_{1j}, \ldots, \gamma_{nj})^{\mathrm{T}}$ be the eigenvector of $\check{\mathbf{K}}$ corresponding to its j-th largest eigenvalue. The duality also implies that $\widehat{\boldsymbol{\psi}}_j(\mathbf{u}) = \{\widehat{\psi}_{j1}(u_1), \cdots, \widehat{\psi}_{jp}(u_p)\}^{\mathrm{T}}$ is the eigenfunction of $\widehat{\mathbf{K}}$ corresponding to its j-th largest

eigenvalue, where

$$\widehat{\psi}_{ji}(u_i) = \sum_{t=1}^{n-k_0} \gamma_{tj}\{Y_{ti}(u_i) - \bar{Y}_i(u_i)\}, \quad i = 1, \ldots, p, \quad j = 1, \ldots, r. \tag{3.28}$$

Note that the eigenfunctions $\widehat{\boldsymbol{\psi}}_j(\cdot), \ldots, \widehat{\boldsymbol{\psi}}_r(\cdot)$ obtained above may not be orthonormal. But they may be made orthonormal by applying a Gram-Schmidt algorithm.

The heuristic argument presented above is justified by Proposition 1 below. Its proof is similar to that of Proposition 2 of Bathia et al. (2010), and is therefore omitted.

Proposition 3.2. *The operator* $\widehat{\mathbf{K}}$ *shares the same non-zero eigenvalues with matrix* $\check{\mathbf{K}}$ *defined in* (3.27) *with the corresponding eigenfunctions given in* (3.28).

In practice we need to estimate r (i.e. the number of non-zero eigenvalues). Let $\lambda_1(\check{\mathbf{K}}) \geqslant \cdots \geqslant \lambda_{n-k_0}(\check{\mathbf{K}}) \geqslant 0$ be the eigenvalues of $\check{\mathbf{K}}$. We take the commonly-adopted ratio-based estimator for r as:

$$\widehat{r} = \underset{1 \leqslant j \leqslant c_r(n-k_0)}{\operatorname{argmax}} \frac{\lambda_j(\check{\mathbf{K}})}{\lambda_{j+1}(\check{\mathbf{K}})}, \tag{3.29}$$

where $c_r \in (0, 1)$ is a prescribed constant. In empirical studies, we take $c_r = 0.75$ to avoid the fluctuations due to the ratios of extreme small values. See further discussion in Lam and Yao (2012). The above method also applies to the cases where $\mathcal{U}_j$, for different j, have different dimensions, as in Happ and Greven (2018).

3.3.4 Dimension reduction and prediction for moderate and large p

When p is moderate or large, after the first segmentation step in Section 3.2, our second dimension reduction step applies the techniques in Sections 3.3.2 and 3.3.3 to each estimated transformed curve subseries $\widehat{\mathbf{Z}}_{t,l}(\cdot)$ in (3.17) instead of $\mathbf{Z}_{t,l}(\cdot)$ for $l = 1, \ldots, q$. Specifically, following the same spirit as (3.24), we define a nonnegative operator,

$$\mathbf{K}_l(u, v) = \sum_{k=1}^{k_0} \int_{\mathcal{U}_0} \mathbf{M}_{k,l}(u, w)\{\mathbf{M}_{k,l}(v, w)\}^{\mathrm{T}} \, \mathrm{d}w, \quad u, v \in \mathcal{U}_0, \tag{3.30}$$

for the l-th group, where $\mathbf{M}_{k,l}(u,v) = \mathrm{Cov}\{\mathbf{Z}_{t,l}(u), \mathbf{Z}_{t+k,l}(v)\}$. With the aid of (3.17), the estimators of $\mathbf{M}_{k,l}(u,v)$ and $\mathbf{K}_l(u,v)$ are respectively defined by

$$\widehat{\mathbf{M}}_{k,l}(u,v) = \widehat{\mathbf{A}}_l^{\mathrm{T}} \widehat{\mathbf{\Sigma}}_{y,k}(u,v) \widehat{\mathbf{A}}_l, \tag{3.31}$$

$$\widehat{\mathbf{K}}_l(u,v) = \sum_{k=1}^{k_0} \int_{\mathcal{U}_0} \widehat{\mathbf{M}}_{k,l}(u,w) \{\widehat{\mathbf{M}}_{k,l}(v,w)\}^{\mathrm{T}} \, \mathrm{d}w. \tag{3.32}$$

Implementing the three-step procedure in Section 3.3.2 on $\widehat{\mathbf{K}}_l(u,v)$ for each l, we obtain the h-step ahead prediction for transformed curve subseries $\widehat{\mathbf{Z}}_{t+h,l}(\cdot)$ and hence the h-step ahead prediction for original curve subseries $\widehat{\mathbf{Y}}_{t+h,l}(\cdot) = \widehat{\mathbf{A}}_l \widehat{\mathbf{Z}}_{t+h,l}(\cdot)$. It is noteworthy that (3.31) requires the consistent estimators of $\mathbf{\Sigma}_{y,k}(u,v)$ for $k = 1, \ldots, k_0$. Its implementation under the high-dimensional setting can thus be done by setting $\widehat{\mathbf{\Sigma}}_{y,k} = \mathcal{T}_{\omega_k}(\widehat{\mathbf{\Sigma}}^{\mathrm{S}}_{y,k})$.

3.4 Theoretical properties

In this section, we present theoretical analysis of our estimation procedure consisting of the segmentation step followed by the dimension reduction step.

Before imposing the regularity conditions, we solidify some notation and definition. For any $\boldsymbol{\mathcal{B}} = (\mathcal{B}_{ij})_{p\times q}$ with each $\mathcal{B}_{ij} \in \mathbb{S}$, we denote its functional version of matrix l_∞ norm by $\|\boldsymbol{\mathcal{B}}\|_{\mathcal{S},\infty} = \max_i \sum_j \|\mathcal{B}_{ij}\|_{\mathcal{S}}$. Denote the p-fold Cartesian product defined on $\mathcal{U}_0$ by $\mathbb{H} = L^2(\mathcal{U}_0) \times \cdots \times L^2(\mathcal{U}_0)$. We define the functional version of sub-Gaussianity that facilities the development of non-asymptotic results for Hilbert space-valued random elements.

Definition 3.1. *Let $X_t(\cdot)$ be a mean zero random variable in $L^2(\mathcal{U}_0)$ and $\Sigma_0 : L^2(\mathcal{U}_0) \to L^2(\mathcal{U}_0)$ be a covariance operator. Then $X_t(\cdot)$ is a sub-Gaussian process if there exists a constant $c > 0$ such that $\mathbb{E}[\exp\{\langle x, X_t - \mathbb{E}(X_t)\rangle\}] \leqslant \exp\{2^{-1}c^2\langle x, \Sigma_0(x)\rangle\}$ for all $x \in L^2(\mathcal{U}_0)$.*

Condition 3.2. *(i) $\mathbf{Y}_t(\cdot)$ is a sequence of multivariate functional linear processes with sub-Gaussian errors, i.e., $\mathbf{Y}_t(\cdot) = \sum_{l=0}^{\infty} \boldsymbol{\mathcal{D}}_l(\boldsymbol{\epsilon}_{t-l})$, where $\boldsymbol{\mathcal{D}}_l = (\mathcal{D}_{l,ij})_{p\times p}$ with each $\mathcal{D}_{l,ij} \in \mathbb{S}$ and $\boldsymbol{\epsilon}_t(\cdot) = \{\epsilon_{t1}(\cdot), \ldots, \epsilon_{tp}(\cdot)\}^{\mathrm{T}}$ with independent components of mean-zero sub-Gaussian processes satisfying Definition 3.1; (ii) The coefficient functions satisfy $\sum_{l=0}^{\infty} \|\boldsymbol{\mathcal{D}}_l\|_{\mathcal{S},\infty} = O(1)$; (iii) $\max_j \int_{\mathcal{U}_0} \mathrm{Cov}\{\epsilon_{tj}(u), \epsilon_{tj}(u)\} \, \mathrm{d}u = O(1)$.*

Condition 3.3. *The spectral density operator* $\boldsymbol{f}_{y,\theta} = 2\pi^{-1}\sum_{k\in\mathbb{Z}}\boldsymbol{\Sigma}_{y,k}\exp(-ik\theta)$ *for* $\theta \in [-\pi,\pi]$ *exists and the functional stability measure defined in (3.33) is finite, i.e.,*

$$\mathcal{M}_y = 2\pi \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi],\boldsymbol{\Phi}\in\mathbb{H}_0} \frac{\langle \boldsymbol{\Phi}, \boldsymbol{f}_{y,\theta}(\boldsymbol{\Phi})\rangle}{\langle \boldsymbol{\Phi}, \boldsymbol{\Sigma}_{y,0}(\boldsymbol{\Phi})\rangle} < \infty, \tag{3.33}$$

where $\mathbb{H}_0 = \{\boldsymbol{\Phi} \in \mathbb{H} : \langle \boldsymbol{\Phi}, \boldsymbol{\Sigma}_{y,0}(\boldsymbol{\Phi})\rangle \in (0,\infty)\}$.

Condition 3.4. κ_0, m *and* k_0 *are fixed positive integers.*

Condition 3.3 places a finite upper bound on the functional stability measure, which characterizes the effect of small decaying eigenvalues of $\boldsymbol{\Sigma}_{y,0}$ on the numerator of (3.33), thus being able to handle infinite-dimensional functional objects $Y_{tj}(\cdot)$. See its detailed discussion in Guo and Qiao (2022). Condition 3.2 (i) can be viewed as the functional (or multivariate) generalization of the multivariate (or functional) linear process. Condition 3.2 (ii) and (iii) guarantees the covariance-stationarity of $\{\mathbf{Y}_t(\cdot)\}$ and implies that $\max_j \int_{\mathcal{U}_0} \Sigma^{(0)}_{y,jj}(u,u)\mathrm{d}u = O(1)$ (Fang et al., 2022). Both conditions are essential to derive the convergence rate for $\widehat{\boldsymbol{\Sigma}}_{y,k}$ under the functional version of ℓ_∞ norm, $\max_{i,j}\|\widehat{\Sigma}^{(k)}_{y,ij} - \Sigma^{(k)}_{y,ij}\|_{\mathcal{S}} = O_p\{\mathcal{M}_y(\log p/n)^{1/2}\}$, which plays a crucial rule in our theoretical analysis. In general, we can relax Conditions 3.2(ii) and (iii) by allowing $\sum_{l=0}^{\infty}\|\boldsymbol{\mathcal{D}}_l\|_{\mathcal{S},\infty}$ and $\max_j \int_{\mathcal{U}_0} \mathrm{Cov}\{\epsilon_{tj}(u), \epsilon_{tj}(u)\}\,\mathrm{d}u$ to diverge slowly with p, then our established rates below will depend on these two terms.

We first establish the group recovery consistency of the segmentation step. We reformulate the permutation step in Section 3.2.3 in an equivalent graph representation way. With an appropriate level $\tau_n > 0$, we build an estimated graph $(G, \widehat{E})$ with vertex set $G = \{1,\ldots,p\}$ and edge set

$$\widehat{E} = \big\{(i,j) : \widehat{T}_{ij} > \tau_n\big\}, \tag{3.34}$$

and split it into multiple connected subgraphs $(\widehat{G}_{l'}, \widehat{E}_{l'})$ for $l' = 1,\ldots,\hat{q}$. Note that p vertexes in G corresponds to the ordered eigenvectors $(\widehat{\boldsymbol{\eta}}_1,\ldots,\widehat{\boldsymbol{\eta}}_p)$. Denote by $\{(\lambda_j(\mathbf{W}_y), \boldsymbol{\eta}_j)\}_{j=1}^p$ the (eigenvalue, eigenvector) pairs of $\mathbf{W}_y$ with $\lambda_1(\mathbf{W}_y) \geqslant \cdots \geqslant \lambda_p(\mathbf{W}_y)$. The true segmented groups specified as $G_1,\ldots,G_q$ forms a partition of G such that

$$\boldsymbol{\Gamma}_{G_l} = (\boldsymbol{\eta}_j)_{j\in G_l} \in \mathbb{R}^{p\times p_l} \text{ with } |G_l| = p_l \text{ and } \mathcal{C}(\boldsymbol{\Gamma}_{G_l}) = \mathcal{C}(\mathbf{A}_l) \text{ for } l = 1,\ldots,q. \tag{3.35}$$

Condition 3.5. *For each* l *and every possible* $\mathbf{H}_l = (\mathbf{h}_{l,1},\ldots,\mathbf{h}_{l,p_l}) \in \mathbb{R}^{p\times p_l}$ *with* $\mathcal{C}(\mathbf{H}_l) = \mathcal{C}(\boldsymbol{\Gamma}_{G_l}) = \mathcal{C}(\mathbf{A}_l)$, *there exists a connected graph* $\big(\{1,\ldots,p_l\}, E_l\big)$, *some*

$\varsigma_l > 0$ *and fixed* m *such that*

$$\max_{0\leqslant k\leqslant m} \max\left\{\|\mathbf{h}_{l,i}^{\mathrm{T}}\boldsymbol{\Sigma}_{y,k}\mathbf{h}_{l,j}\|_{\mathcal{S}}, \|\mathbf{h}_{l,i}^{\mathrm{T}}(\boldsymbol{\Sigma}_{y,k})^{\mathrm{T}}\mathbf{h}_{l,j}\|_{\mathcal{S}}\right\} \geqslant \varsigma_l, \quad \forall (i,j) \in E_l.$$

Condition 3.5 formalizes the supporting intuition of the permutation step in Section 3.2.3. To be specific, due to the fact that $\mathbf{A} = (\mathbf{A}_1, \ldots, \mathbf{A}_q)$ in (3.2) is not uniquely defined, this condition ensures that the group G_l is inseparable at the minimal signal level $\varsigma_l > 0$ given any legitimate transformation $\mathbf{H}_l$ for each l. Recall that $\mathbf{W}_z = \mathrm{diag}(\mathbf{W}_{z,1}, \ldots, \mathbf{W}_{z,q})$ in (3.7) is a block diagonal matrix, where $\mathbf{W}_{z,l}$ is a $p_l \times p_l$ matrix. For each $l = 1, \ldots, q$, we further define the minimum difference between eigenvalues of $\mathbf{W}_{z,l}$ and those of other $\mathbf{W}_{z,j}$'s as

$$\rho_l = \min_{j\neq l} \min_{\tilde{\lambda}_l\in\lambda(\mathbf{W}_{z,l}),\tilde{\lambda}_j\in\lambda(\mathbf{W}_{z,j})} |\tilde{\lambda}_l - \tilde{\lambda}_j|,$$

where $\lambda(\mathbf{B})$ denotes the set of eigenvalues of the matrix $\mathbf{B}$. Let $\rho = \min_{1\leqslant l\leqslant q} \rho_l > 0$ and $\varsigma = \min_{1\leqslant l\leqslant q} \varsigma_l > 0$.

Theorem 3.1. *Let Conditions 3.1–3.5 hold and* $\nu_n = \Xi^2 \mathcal{M}_y^{1-\alpha} (\log p/n)^{(1-\alpha)/2}$, *where* Ξ *is specified in Lemma 3.1. There exits some constant* $c > 0$ *such that* $c\rho^{-1}\Xi\nu_n < \tau_n < \varsigma - c\rho^{-1}\Xi\nu_n$, *then* G_l *for each* l *in (3.35) is detected by (3.34) in the sense of* $\max_{1\leqslant l\leqslant q} P(\widehat{G}_{l'} \neq G_l) \to 0$ *for some* $l' \in \{1, \ldots, q\}$.

Theorem 3.1 guarantees the group recovery consistency of our segmentation step, which further implies that $\hat{q} = q$ and $\hat{p}_l = p_l$ for $l = 1, \ldots, q$ hold with high probability. Supported by Theorem 3.1, our subsequent theoretical results are developed by assuming that the group structure of $\mathbf{Z}_t(\cdot)$, (i.e., $\{G_l, l = 1, \ldots, q\}$) is correctly identified or known.

To evaluate the errors in estimating $\mathcal{C}(\mathbf{A}_l) = \mathcal{C}(\boldsymbol{\Gamma}_{G_l})$ for $l = 1, \ldots, q$, we use a discrepancy measure (Chang et al., 2018) of two linear spaces spanned by the columns of $\mathbf{B}_i \in \mathbb{R}^{p\times\tilde{p}}$ with $\mathbf{B}_i^{\mathrm{T}}\mathbf{B}_i = \mathbf{I}_{\tilde{p}}$ for $i = 1, 2$ as

$$D\{\mathcal{C}(\mathbf{B}_1), \mathcal{C}(\mathbf{B}_2)\} = \sqrt{1 - \tilde{p}^{-1}\mathrm{trace}(\mathbf{B}_1\mathbf{B}_1^{\mathrm{T}}\mathbf{B}_2\mathbf{B}_2^{\mathrm{T}})} \in [0, 1]. \tag{3.36}$$

Then $D\{\mathcal{C}(\mathbf{B}_1), \mathcal{C}(\mathbf{B}_2)\}$ is equal to 0 if and only if $\mathcal{C}(\mathbf{B}_1) = \mathcal{C}(\mathbf{B}_2)$, and to 1 if and only if the two spaces are orthogonal.

Theorem 3.2. *Let Conditions 3.1–3.4 hold and* $\widehat{\boldsymbol{\Gamma}}_{G_l} = (\hat{\boldsymbol{\eta}}_j)_{j\in G_l}$. *There exists some constants* $c_k > 0$ *for* $k = 1, \ldots, \kappa_0$ *such that, with the choice of threshold levels*

$\omega_k = c_k \mathcal{M}_y (\log p/n)^{1/2}$,

$$\max_{1 \leqslant l \leqslant q} \rho_l D\big\{\mathcal{C}(\widehat{\mathbf{\Gamma}}_{G_l}), \mathcal{C}(\mathbf{\Gamma}_{G_l})\big\} = O_p\big(\nu_n\big), \tag{3.37}$$

where ν_n is specified in Theorem 3.1.

Remark 3.8. *Theorem 3.2 presents the uniform convergence rate over $l = 1, \ldots, q$ for $\rho_l D\big\{\mathcal{C}(\widehat{\mathbf{\Gamma}}_{G_l}), \mathcal{C}(\mathbf{\Gamma}_{G_l})\big\}$, which is determined by both dimensionality parameters (n, p, s_1, s_2) and internal parameters (M_y, α). It is easy to see that the rate is faster for smaller values of $\{s_1, s_2, \mathcal{M}_y, \alpha\}$, while enlarging the minimum eigen-gap between $\mathbf{W}_{z,l}$ and other blocks (i.e., larger ρ_l) reduces the difficulty of estimating $\mathcal{C}(\mathbf{\Gamma}_{G_l})$.*

We now turn to investigate the theoretical properties of the second dimension reduction step. Inherited from the segmentation step, $\{\mathbf{Z}_{t,l}(\cdot), l = 1, \ldots, q\}$ in (3.6) relies on the specific form of $\mathbf{A} = (\mathbf{A}_1, \ldots, \mathbf{A}_q)$, and thus is not uniquely defined. Yet intuitively, we only require a certain transformation matrix to make our subsequent analysis related to the set $\{\hat{\boldsymbol{\eta}}_j, j = 1, \ldots, p\}$ mathematically tractable.

Denote by $\{\mathbf{\Pi}_{G_l}, l = 1, \ldots, q\}$ the particular set of legitimate transformation we are interested in. To save space and avoid confusion, we defer the construction of $\mathbf{\Pi}_{G_l}$ to (3.44) in Section 3.A of the Appendix. Note that $\mathcal{C}(\mathbf{\Pi}_{G_l}) = \mathcal{C}(\mathbf{A}_l)$ for each l. Let $\mathbf{Z}_{t,l}(\cdot) = \mathbf{\Pi}_{G_l}^{\mathrm{T}} \mathbf{Y}_t(\cdot)$. Recall that the primary goal of this step is to identify r_l and to estimate the dynamic space spanned by $\boldsymbol{\psi}_{l,1}(\cdot), \ldots, \boldsymbol{\psi}_{l,r_l}(\cdot)$, denoted by $\mathcal{C}(\boldsymbol{\psi}_l) = \mathrm{span}\{\boldsymbol{\psi}_{l,1}(\cdot), \ldots, \boldsymbol{\psi}_{l,r_l}(\cdot)\}$, for each l, where $\boldsymbol{\psi}_{l,j} = (\psi_{l,j1}, \ldots, \psi_{l,jp_l})^{\mathrm{T}}$ for $j = 1, \ldots, r_l$ are the orthonormal eigenfunctions of $\mathbf{K}_l$ in (3.30) corresponding to non-zero eigenvalues $\theta_{l,1}, \ldots, \theta_{l,r_l}$. Also denote by $\{\hat{\theta}_{l,j}, \widehat{\boldsymbol{\psi}}_{l,j}(\cdot)\}_{j \geqslant 1}$ the (eigenvalue, eigenfunction) pairs of $\widehat{\mathbf{K}}_l$ in (3.32). Note we always arrange the eigenvalues in descending order. Our asymptotic results are based on the following regularity conditions:

Condition 3.6. $\max_{1 \leqslant l \leqslant q} \mathbb{E}\big(\|\mathbf{Z}_{t,l}\|^2\big) = O(1)$ *and* $\max_{1 \leqslant l \leqslant q} p_l = O(1)$.

Condition 3.7. *For each l, all r_l non-zero eigenvalues of $\mathbf{K}_l$ are different, i.e., $\theta_{l,1} > \cdots > \theta_{l,r_l} > 0 = \theta_{l,r_l+1} = \ldots$.*

In the spirit of (3.29), we can obtain ratio-based estimator $\hat{r}_l$ for r_l based on the eigenanalysis of $\widehat{\mathbf{K}}_l$ for each l. To facilitate the consistency analysis of $\hat{r}_l$ and to avoid cases of '0/0', we propose a modified estimator

$$\hat{r}_l = \underset{1 \leqslant j \leqslant n - k_0}{\mathrm{argmax}} \frac{\lambda_j(\widehat{\mathbf{K}}_l) + \delta_n}{\lambda_{j+1}(\widehat{\mathbf{K}}_l) + \delta_n}, \tag{3.38}$$

where $\delta_n > 0$ provides an upper bound correction to $\lambda_j(\widehat{\mathbf{K}}_l)$ for $j > r_l$ and all l.

Theorem 3.3. *Suppose Conditions 3.1–3.4 and 3.6–3.7 hold,* $\delta_n = \rho^{-1}\Xi\nu_n$ *and* $\delta_n \max_l \theta_{l,1} / \min_l \theta_{l,r_l}^2 \to 0$. *Then for* $\hat{r}_l$ *defined in (3.38), we have that* $\min_{1\leqslant l\leqslant q} P(\hat{r}_l = r_l) \to 1$.

Theorem 3.3 shows that r_l can be correctly identified via (3.38) with probability tending to one uniformly over $\{1, \ldots, q\}$. In practice, provided that δ_n is usually hard to be specified, we instead adopt (3.29) operating on $\widehat{\mathbf{K}}_l$ to estimate r_l for each group. This theorem also provides support for the assumption that $r_l = \hat{r}_l$ or is known for all l to facilitate further convergence results.

Let $\mathcal{C}(\widehat{\boldsymbol{\psi}}_l) = \operatorname{span}\{\widehat{\boldsymbol{\psi}}_{l,1}(\cdot), \ldots, \widehat{\boldsymbol{\psi}}_{l,r_l}(\cdot)\}$ be the dynamic space spanned by r_l estimated eigenfunctions. To measure the discrepancy between $\mathcal{C}(\boldsymbol{\psi}_l)$ and $\mathcal{C}(\widehat{\boldsymbol{\psi}}_l)$, we introduce the following metric. For two r-dimensional subspaces $\mathcal{C}(\mathbf{b}_1) = \operatorname{span}\{\mathbf{b}_{11}(\cdot), \ldots, \mathbf{b}_{1r}(\cdot)\}$ and $\mathcal{C}(\mathbf{b}_2) = \{\mathbf{b}_{21}(\cdot), \ldots, \mathbf{b}_{2r}(\cdot)\}$ of $\mathbb{H}$ satisfying $\langle \mathbf{b}_{ij}, \mathbf{b}_{ik} \rangle = 1$ if $j = k$ and 0 otherwise for $i = 1, 2$, the discrepancy measure between $\mathcal{C}(\mathbf{b}_1)$ and $\mathcal{C}(\mathbf{b}_2)$ is defined as

$$D\{\mathcal{C}(\mathbf{b}_1), \mathcal{C}(\mathbf{b}_2)\} = \sqrt{1 - r^{-1} \sum_{j,k=1}^{r} \left(\langle \mathbf{b}_{1j}, \mathbf{b}_{2k} \rangle\right)^2} \quad \in [0, 1].$$

This measure equals 0 if and only if $\mathcal{C}(\mathbf{b}_1) = \mathcal{C}(\mathbf{b}_2)$ and 1 if and only if two spaces are orthogonal. It is worth noting that this can be seen as a multivariate generalization of the discrepancy measure used in Bathia et al. (2010) and also the generalization of (3.36) to the functional domain.

Theorem 3.4. *Let Conditions 3.1–3.4 and 3.6–3.7 hold. Then we have that*

$$\max_{1\leqslant l\leqslant q} \rho_l D\{\mathcal{C}(\hat{\boldsymbol{\psi}}_l\}, \mathcal{C}(\boldsymbol{\psi}_l)\} = O_p(\Xi\nu_n). \tag{3.39}$$

Comparing with (3.37), the uniform convergence rate in (3.39) is slower by a multiplicative factor Ξ. This comes from controlling the additional term $\boldsymbol{\Sigma}_{y,k}$ (recall (3.31)) in the sense of $\|\boldsymbol{\Sigma}_{y,k}\|_{\mathcal{S},\infty} \leqslant \Xi \max_{i,j} \|\Sigma_{y,ij}^{(k)}\|_{\mathcal{S}}^{1-\alpha} = O(\Xi)$ as required to bound $\|\widehat{\mathbf{M}}_{k,l} - \mathbf{M}_{k,l}\|_{\mathcal{S}}$ for each l.

3.5 Simulation studies

We conduct a series of simulations to illustrate the finite sample performance of the proposed methods for cases when p is moderate and large in Sections 3.5.1 and 3.5.2, respectively.

3.5.1 Moderate p

In each simulated scenario, we generate p-vector of observed functional time series $\mathbf{Y}_t(u)$ for $t = 1, \ldots, n$ and $u \in \mathcal{U}_0 = [0, 1]$ by (3.2), where the entries of $\mathbf{A}$ are sampled from Uniform$[-3, 3]$ and p-vector of transformed functional time series $\mathbf{Z}_t(u)$ is decomposed as the sum of a dynamic element $\mathbf{X}_t(u) = \{X_{t1}(u), \ldots, X_{tp}(u)\}^{\mathrm{T}}$ and a white noise element $\boldsymbol{\varepsilon}_t(u) = \{\varepsilon_{t1}(u), \ldots, \varepsilon_{tp}(u)\}^{\mathrm{T}}$ according to (3.1). Each curve component of $\boldsymbol{\varepsilon}_t(\cdot)$ is generated by $\varepsilon_{tj}(u) = \sum_{l=1}^{10} 2^{-(l-1)} e_{tjl}\psi_l(u)$ for $j = 1, \ldots, p$, where e_{tjl} are independent standard normal and $\{\psi_l(\cdot)\}_{l=1}^{10}$ is a 10-dimensional Fourier basis function. To generate $\mathbf{Z}_t(\cdot)$ with predesigned group structure, we need to generate its finite-dimensional dynamic element $\mathbf{X}_t(\cdot)$ with the same group structure. Specifically, let $\vartheta_{tg}(u) = \sum_{l=1}^{5} \zeta_{tgl}\psi_l(u)$ be 5-dimensional curve dynamics for $g = 1, \ldots, 30$. The basis coefficients $\boldsymbol{\zeta}_{tg} = (\zeta_{tg1}, \ldots, \zeta_{tg5})^{\mathrm{T}}$ are generated from a stationary VAR model $\boldsymbol{\zeta}_{tg} = \mathbf{U}_g\boldsymbol{\zeta}_{(t-1)g} + \mathbf{e}_t$ for each g. The entries of $\mathbf{U}_g \in \mathbb{R}^{5\times 5}$ are sampled from Uniform$[-3, 3]$ and rescaled by $\iota/\rho(\mathbf{U}_g)$ with $\rho(\mathbf{U}_g)$ being the spectral radius of $\mathbf{U}_g$ and $\iota \sim$ Uniform$[0.5, 1]$ to guarantee the stationary of $\boldsymbol{\zeta}_{tg}$. The components of the innovation $\boldsymbol{e}_t$ are sampled independently from $\mathcal{N}(0, 1)$. We consider the following three examples to generate $\mathbf{X}_t(\cdot)$ with different group structures for $p = 6, 10, 15$ based on $\vartheta_{t1}(\cdot), \ldots, \vartheta_{t5}(\cdot)$.

EXAMPLE 1. $X_{t1}(\cdot) = \vartheta_{t1}(\cdot)$, $X_{tj}(\cdot) = \vartheta_{(t+j-2)2}(\cdot)$ for $j = 2, 3$ and $X_{tj}(\cdot) = \vartheta_{(t+j-4)3}(\cdot)$ for $j = 4, 5, 6$.

EXAMPLE 2. $X_{tj}(\cdot)$ for $j = 1, \ldots, 6$ are the same as those in Example 1 and $X_{tj}(\cdot) = \vartheta_{(t+j-7)4}(\cdot)$ for $j = 7, \ldots, 10$.

EXAMPLE 3. $X_{tj}(\cdot)$ for $j = 1, \ldots, 10$ are the same as those in Example 2 and $X_{tj}(\cdot) = \vartheta_{(t+j-11)5}(\cdot)$ for $j = 11, \ldots, 15$.

Therefore, $\mathbf{X}_t(\cdot)$ consists of $q = 3, 4$ and 5 uncorrelated groups of curve subseries in Examples 1, 2 and 3, respectively, where the number of component curves per group is $p_l = l$ for $l = 1, \ldots, q$. The white noise sequence $\boldsymbol{\varepsilon}_t(\cdot)$ ensures that $\mathbf{Z}_t(\cdot)$ shares the same group structure as $\mathbf{X}_t(\cdot)$. Unless otherwise stated, we set $k_0 = \kappa_0 = m = 5$ and $c_r = c_\varrho = 0.75$ in our simulations, as our simulation results suggest that they are insensitive to the maximum lag orders and the prescribed constants used in ratio-based estimators.

The performance of our proposed procedure is examined in terms of linear space specification, group identification and post-sample prediction. We start with the definition of an 'effective' specification of the q linear spanned spaces $\mathcal{C}(\mathbf{A}_l)$ for $l = 1, \ldots, q$. Since the ranks of $\mathbf{A}_i$ and $\widehat{\mathbf{A}}_i$ are not necessarily the same, we use a

general discrepancy measure (Chang et al., 2018) of two linear spaces spanned by the columns of $\mathbf{B}_i \in \mathbb{R}^{p\times\tilde{p}_i}$ with $\text{rank}(\mathbf{B}_i) = \tilde{p}_i$ for $i = 1, 2$ as

$$\widetilde{D}^2\big(\mathcal{C}(\mathbf{B}_1), \mathcal{C}(\mathbf{B}_2)\big) = 1 - \frac{1}{\min(\tilde{p}_1, \tilde{p}_2)}\text{trace}(\mathbf{Q}_1\mathbf{Q}_2) \in [0, 1],$$

where $\mathbf{Q}_i = \mathbf{B}_i(\mathbf{B}_i^{\mathrm{T}}\mathbf{B}_i)^{-1}\mathbf{B}_i^{\mathrm{T}}$, for $i = 1, 2$. Then $\widetilde{D}^2(\mathcal{C}(\mathbf{B}_1), \mathcal{C}(\mathbf{B}_2))$ is equal to 0 if and only if $\mathcal{C}(\mathbf{B}_1) \subset \mathcal{C}(\mathbf{B}_2)$ or $\mathcal{C}(\mathbf{B}_2) \subset \mathcal{C}(\mathbf{B}_1)$, and to 1 if and only if the two spaces are orthogonal. We call $(\widehat{\mathbf{A}}_1, \ldots, \widehat{\mathbf{A}}_{\hat{q}})$ an effective specification for $\mathbf{A} = (\mathbf{A}_1, \ldots, \mathbf{A}_q)$ if (i) $1 < \hat{q} \leqslant q$; (ii) After pairing each $\mathbf{A}_l$ $(l = 1, \ldots, q)$ with $\widehat{\mathbf{A}}_j$ for which

$$l' = f(l) := \underset{j\in(1,\ldots,\hat{q})}{\text{argmin}}\ \widetilde{D}^2\big(\mathcal{C}(\mathbf{A}_l), \mathcal{C}(\widehat{\mathbf{A}}_j)\big),$$

the ranks of $\mathbf{A}_l$ paired with that of the $\widehat{\mathbf{A}}_{l'}$ satisfy $\sum_{l:f(l)=l'}\text{rank}(\mathbf{A}_l) = \text{rank}(\widehat{\mathbf{A}}_{l'})$ for each $l' = 1, \ldots, \hat{q}$. Intuitively, such specification for $\mathbf{A}$ leads to an effective segmentation for $\mathbf{Z}_t(\cdot)$ in the sense that each identified group in $\widehat{\mathbf{Z}}_t(\cdot)$ contains at least one, but not all, groups in $\mathbf{Z}_t(\cdot)$. To ease reference, we call the above situation 'effective segmentation' hereafter. For the special case of complete segmentation $(\hat{q} = q)$, we compute the maximum and averaged estimation errors for $(\widehat{\mathbf{A}}_1, \ldots, \widehat{\mathbf{A}}_{\hat{q}})$, respectively defined as

$$\max D^2(\widehat{\mathbf{A}}, \mathbf{A}) = \max_{1\leqslant l\leqslant q}\widetilde{D}^2(\mathcal{C}(\mathbf{A}_l), \mathcal{C}(\widehat{\mathbf{A}}_{f(l)})) \text{ and } \bar{D}^2(\widehat{\mathbf{A}}, \mathbf{A}) = \frac{1}{q}\sum_{l=1}^{q}\widetilde{D}^2(\mathcal{C}(\mathbf{A}_l), \mathcal{C}(\widehat{\mathbf{A}}_{f(l)})),$$

respectively, to assess the ability of our method in fully recovering the spanned spaces, $\mathcal{C}(\mathbf{A}_1), \ldots, \mathcal{C}(\mathbf{A}_q)$.

To evaluate the post-sample predictive accuracy, we integrate **seg**mentation transformation and dimension reduction in Sections 3.2 and 3.3 into the **V**AR estimation (denoted as SegV) to obtain h-step ahead prediction consisting of three steps below.

i. Treat the first $n - h$ observations as training data, adopt the normalization step in Section 3.2.1 to obtain $\widetilde{\mathbf{Y}}_t(\cdot) = \widehat{\mathbf{V}}_y^{-1/2}\mathbf{Y}_t(\cdot)$, implement the segmentation transformation step in Section 3.2.2 on $\{\widetilde{\mathbf{Y}}_t(\cdot)\}_{t=1}^{n-h}$ and the permutation step in Section 3.2.3 to calculate $(\widehat{\mathbf{A}}_1, \ldots, \widehat{\mathbf{A}}_{\hat{q}})$ and to identify $\hat{q}$ uncorrelated groups, $\widehat{\mathbf{Z}}_{t,1}(\cdot), \ldots, \widehat{\mathbf{Z}}_{t,\hat{q}}(\cdot)$.

ii. Following Section 3.3.4, within each identified group, apply the three-step approach in Section 3.3.2 on $\{\widehat{\mathbf{Z}}_{t,l}(\cdot)\}_{t=1}^{n-h}$ to achieve the h-step ahead prediction $\widehat{\mathbf{Z}}_{n,l}(\cdot)$ for $l = 1, \ldots, \hat{q}$. In particular, for each l, select the best VAR that best fits each lower-dimensional vector process $\{\hat{\boldsymbol{\zeta}}_{t,l}\}_{t=1}^{n-h}$ according to the AIC criterion.

iii. Obtain the h-step ahead prediction for normalized curves $\widetilde{\mathbf{Y}}_n = (\mathbf{Y}_{n,1}^{\mathrm{T}}, \ldots, \mathbf{Y}_{n,\hat{q}}^{\mathrm{T}})^{\mathrm{T}}$ with each $\widetilde{\mathbf{Y}}_{n,l}(\cdot) = \widehat{\mathbf{A}}_l \widehat{\mathbf{Z}}_{n,l}(\cdot)$ and hence for original curves $\widehat{\mathbf{Y}}_n(\cdot) = \widehat{\mathbf{V}}_y^{1/2} \widetilde{\mathbf{Y}}_n(\cdot)$. Compute the mean squared prediction error (MSPE), defined as

$$\mathrm{MSPE} = \frac{1}{pN} \sum_{j=1}^{p} \sum_{i=1}^{N} \{\hat{Y}_{nj}(v_i) - Y_{nj}(v_i)\}^2, \tag{3.40}$$

where $v_1, \ldots, v_N$ are equally spaced time points in $[0, 1]$.

We compute the relative prediction error as the ratio of MSPE in (3.40) to that under the 'oracle' case, which uses the true $\mathbf{A}$ and the embedded group structure in the estimation. For comparison, we also implement an **uni**variate functional prediction method on each $Y_{tj}(\cdot)$ separately by performing univariate dimension reduction (Bathia et al., 2010), then predicting vector time series based on the best fitted **V**AR model and finally recovering functional prediction (denoted as UniV).

We generate $n = 200, 400, 800, 1600, 3200$ observations for each example and replicate each simulation 500 times. Table 3.1 provide numerical summaries, including the relative frequencies of the effective segmentation with respect to $\hat{q} = q$ and $\hat{q} \geqslant q - 1$, and the estimation errors for $\widehat{\mathbf{A}}$ under the complete segmentation case. Note that due to the normalized model assumption, we shall use a transformed version of $\mathbf{A}$ in computing the estimation errors. Let $\mathbf{A}^* = \mathbf{V}_y^{-1/2} \mathbf{A} \mathbf{V}_z^{1/2} = (\mathbf{A}_1^*, \ldots, \mathbf{A}_q^*)$ and $\widetilde{\mathbf{A}} = \mathbf{V}_y^{-1/2} \mathbf{A} = (\widetilde{\mathbf{A}}_1, \ldots, \widetilde{\mathbf{A}}_q)$. Since $\mathbf{V}_z$ is a block-diagonal matrix, it holds that $\mathcal{C}(\widetilde{\mathbf{A}}_l) = \mathcal{C}(\mathbf{A}_l^*)$ for $l = 1, \ldots, q$. Hence, $\widetilde{D}^2(\mathcal{C}(\widehat{\mathbf{A}}), \mathcal{C}(\mathbf{A}))$ can be calculated by replacing $\mathbf{A}$ by $\mathbf{V}_y^{-1/2} \mathbf{A}$. As one would expected, the proposed method provides higher proportions of effective segmentation and lower estimation errors as n increases, and performs fairly well for reasonably large n as p increases. For $(p, n) = (6, 200)$, we observe 62.6% complete segmentation with $\bar{D}^2(\widehat{\mathbf{A}}, \mathbf{A})$ as low as 0.079. Furthermore, the proportions of effective segmentation with $\hat{q} \geqslant q - 1$ are above 93% for $n \geqslant 200$. Similar results can be found for cases of $(p, n) = (10, 800+)$ and $(15, 1600+)$, whose proportions of effective segmentation with $\hat{q} \geqslant q - 1$ remain higher than 87.4% and 83%, respectively. Table 3.1 also reports the relative h-step ahead prediction errors. It is evident that SegV significantly outperforms UniV in all settings, demonstrating the effectiveness of our proposed segmentation transformation and dimension reduction in predicting future values. Although the proportions of complete segmentation are not high especially when $p = 15$, the corresponding proportions of $\hat{q} \geqslant q - 1$ become substantially higher, and SegV performs very similarly to the oracle case with its relative prediction errors being close to 1.

Table 3.1: Examples 1, 2 and 3: The relative frequencies of effective segmentation with respect to $\hat{q} = q$ and $\hat{q} \geqslant q - 1$, and the means (standard deviations) of $\max D^2(\widehat{\mathbf{A}}, \mathbf{A})$, $\bar{D}^2(\widehat{\mathbf{A}}, \mathbf{A})$, and relative MSPEs over 500 simulation runs.

p		$n = 200$	$n = 400$	$n = 800$	$n = 1600$	$n = 3200$
6	$\hat{q} = q$	0.626	0.722	0.772	0.88	0.972
	$\hat{q} \geqslant q - 1$	0.93	0.988	0.998	1	1
	$\max D^2(\widehat{\mathbf{A}}, \mathbf{A})$	0.128(0.088)	0.089(0.066)	0.053(0.048)	0.035(0.037)	0.023(0.027)
	$\bar{D}^2(\widehat{\mathbf{A}}, \mathbf{A})$	0.079(0.052)	0.053(0.038)	0.030(0.025)	0.019(0.019)	0.012(0.014)
	SegV	1.081(0.172)	1.048(0.105)	1.026(0.065)	1.014(0.048)	1.010(0.036)
	UniV	1.584(0.453)	1.598(0.423)	1.596(0.379)	1.651(0.443)	1.623(0.430)
10	$\hat{q} = q$	0.324	0.444	0.644	0.806	0.898
	$\hat{q} \geqslant q - 1$	0.490	0.688	0.874	0.972	0.994
	$\max D^2(\widehat{\mathbf{A}}, \mathbf{A})$	0.301(0.108)	0.193(0.09)	0.117(0.064)	0.072(0.049)	0.035(0.025)
	$\bar{D}^2(\widehat{\mathbf{A}}, \mathbf{A})$	0.183(0.059)	0.115(0.047)	0.069(0.035)	0.041(0.024)	0.019(0.013)
	SegV	1.291(0.271)	1.174(0.215)	1.089(0.143)	1.059(0.091)	1.037(0.070)
	UniV	1.708(0.404)	1.836(0.410)	1.841(0.436)	1.862(0.392)	1.863(0.397)
15	$\hat{q} = q$	0.032	0.178	0.410	0.622	0.790
	$\hat{q} \geqslant q - 1$	0.086	0.344	0.616	0.832	0.948
	$\max D^2(\widehat{\mathbf{A}}, \mathbf{A})$	0.426(0.091)	0.347(0.121)	0.241(0.113)	0.157(0.091)	0.090(0.059)
	$\bar{D}^2(\widehat{\mathbf{A}}, \mathbf{A})$	0.273(0.054)	0.195(0.05)	0.128(0.042)	0.077(0.033)	0.041(0.019)
	SegV	1.477(0.313)	1.363(0.277)	1.166(0.156)	1.091(0.098)	1.056(0.069)
	UniV	1.805(0.370)	1.967(0.394)	2.033(0.394)	2.001(0.384)	2.064(0.413)

3.5.2 Large p

Under a high-dimensional large p scenario, a natural question to ask is whether the segmentation method still perform well, and if not, whether a satisfactory improvement is attainable via the functional-thresholding developed in Section 3.2.4. To this end, we generate $\mathbf{Y}_t(\cdot)$ for $p = 30, 60$ and $n = 200, 400$ by the same procedure as in Section 3.5.1. Specifically, we let $X_{t(3l-2)}(\cdot) = \vartheta_{tl}(\cdot)$, $X_{t(3l-1)}(\cdot) = \vartheta_{(t+1)l}(\cdot)$, $X_{t(3l)}(\cdot) = \vartheta_{(t+2)l}(\cdot)$ for $l = 1, \ldots, q$. This setting ensures q uncorrelated groups of curve subseries in $\mathbf{X}_t(\cdot)$ with $p_l = 3$ component curves per group and hence $q = 10$ and 20 correspond to $p = 30$ and 60, respectively. Let the $p \times p$ transformation matrix $\mathbf{A} = \mathbf{\Delta}_1 + \delta \mathbf{\Delta}_2$. Here $\mathbf{\Delta}_1 = \text{diag}(\mathbf{\Delta}_{11}, \ldots, \mathbf{\Delta}_{1(p/6)})$ with elements of each $\mathbf{\Delta}_{1i} \in \mathbb{R}^{6\times 6}$ being sampled from Uniform$[-3, 3]$ for $i = 1, \ldots, p/6$, and $\mathbf{\Delta}_2$ is a matrix with two randomly selected non-zero elements from Uniform$[-1, 1]$ each row. We set $\delta = 0.1$ and 0.5. It is notable that our setting results in a very high-dimensional learning task in the sense that the intrinsic dimension $30 \times 5 = 150$ or $60 \times 5 = 300$ is large relative to the sample size $n = 200$ or 400..

We assess the performance of ordinary segmentation (Seg) with functional-thresholding-based segmentation (FTSeg) in discovering the group structure. The optimal thresholding parameters $\hat{\omega}_k$ in FTSeg are selected by the five-fold cross validation as discussed in Remark 3.5(i). Tables 3.2 and 3.3 report the relative frequencies of the effective segmentation with respect to different levels of group identification $\hat{q} \geqslant q/2$ and $\hat{q} \geqslant 3q/5$. To enhance the accuracy of identified group structure, we propose to

Table 3.2: The relative frequencies of effective segmentation with respect to $\hat{q} \geqslant q/2$ over 500 simulation runs.

(p, ω)	Seg		FTSeg	
	$n = 200$	$n = 400$	$n = 200$	$n = 400$
(30,0.1)	0	0	0.704	0.990
(30,0.5)	0	0	0.564	0.820
(60,0.1)	0	0	0.296	0.978
(60,0.5)	0	0	0.192	0.712

refine the estimated groups by repeating FTSeg R ($\geqslant 2$) times. To be precise, the i-th round of refinement via FTSeg is performed within each group discovered in the $(i-1)$-th round with $c_\varrho = 1$ in (3.13) for $i = 1, \ldots, R$, and hence $(\widehat{\mathbf{A}}_1, \ldots, \widehat{\mathbf{A}}_{\hat{q}})$ is updated after each iteration. The segmentation results for $R = 5$ and 10 are also provided in Table 3.3. Finally, we compare the predictive performance of UniV, SegV with FTSegV and its refined versions, which substitute Seg in Step i with FTSeg and its R-round refinements, respectively, before Steps ii and iii. Table 3.4 presents the relative prediction errors for all five comparison methods.

Several conclusions can be drawn from Tables 3.2, 3.3 and 3.4. First, the performance of Seg severely deteriorates under the high-dimensional setting. Specifically, this procedure fails to detect any effective segmentation, thus leading to the elevated prediction errors. By comparison, FTSeg does a reasonably good job in recovering the group structure of $\mathbf{Z}_t(\cdot)$ as evident from Table 3.2, and FTSegV exhibits superior predictive performance over SegV and UniV in all scenarios. Second, comparing the results among different R, we observe that repeating the segmentation step can largely refine the identified group structure, whereas its influence on improving the predictive accuracy is limited. For example, under the setting $(p, \omega, n) = (60, 0.1, 400)$, the proportion of effective segmentation with $\hat{q} \geqslant 3q/5$ jumps to 0.772 and 0.9 from the initial 0.012 after 5 and 10 iterations, respectively, implying that $\hat{q}$ tends to q as R increases, while the relative prediction error only decreases slightly from 1.11 to 1.099. This phenomenon highlights the success of FTSegV when p is large in the sense that it performs comparably well to the oracle method especially for sufficiently large n. Although FTSeg fails to efficiently recover the group structure in $\mathbf{Z}_t(\cdot)$ under this case, it achieves an effective dimension reduction to provide significant improvement in high-dimensional functional prediction. If the main purpose lies in the group identification with enhanced accuracy, one can apply FTSeg to $\widehat{\mathbf{Z}}_t(\cdot)$ repeatedly with a moderate value of R.

Table 3.3: The relative frequencies of effective segmentation with respect to $\hat{q} \geqslant 3q/5$ over 500 simulation runs. The highest values are in bold font.

(p,ω)	Seg		FTSeg					
					$R=5$		$R=10$	
	$n=200$	$n=400$	$n=200$	$n=400$	$n=200$	$n=400$	$n=200$	$n=400$
(30,0.1)	0	0	0.080	0.002	0.468	0.630	**0.492**	**0.772**
(30,0.5)	0	0	0.076	0.002	0.368	0.510	**0.378**	**0.660**
(60,0.1)	0	0	0.084	0.012	0.144	0.772	**0.144**	**0.900**
(60,0.5)	0	0	0.066	0.008	0.078	0.650	**0.072**	**0.786**

Table 3.4: Means (standard deviations) of relative MSPEs over 500 simulation runs. The lowest values are in bold font.

Method	(p,ω)	n =200	n =400	(p,ω)	n =200	n =400
FTSegV		1.243(0.162)	1.095(0.105)		1.249(0.122)	1.110(0.073)
FTSegV ($R=5$)		1.225(0.153)	1.091(0.101)		1.250(0.123)	1.104(0.071)
FTSegV ($R=10$)	(30,0.1)	**1.222(0.151)**	**1.087(0.099)**	(60,0.1)	**1.249(0.122)**	**1.099(0.071)**
SegV		1.814(0.376)	1.901(0.368)		1.813(0.271)	1.907(0.265)
UniV		1.631(0.313)	1.735(0.317)		1.599(0.214)	1.682(0.210)
FTSegV		1.268(0.176)	1.134(0.134)		1.285(0.134)	1.149(0.101)
FTSegV ($R=5$)		1.255(0.171)	1.128(0.130)		1.282(0.136)	1.142(0.098)
FTSegV ($R=10$)	(30,0.5)	**1.250(0.168)**	**1.128(0.127)**	(60,0.5)	**1.281(0.136)**	**1.141(0.099)**
SegV		1.815(0.377)	1.903(0.369)		1.813(0.271)	1.905(0.264)
UniV		1.635(0.315)	1.740(0.317)		1.603(0.215)	1.684(0.209)

3.6 Real data analysis

In this section, we apply our proposed SegV and FTSegV to three real data examples arising from different fields. Our main goal is to evaluate the post-sample predictive accuracy of both methods. By comparison, we also implement componentwise univariate prediction method (UniV) and the multivariate prediction method of Gao et al. (2019b) (denoted as GSY) to jointly predict p component series by fitting a factor model to estimated scores obtained via eigenanalysis of the long-run covariance function (Hörmann et al., 2015a). To evaluate the effectiveness of the segmentation step, we also consider cases of under-segmentation and over-segmentation for both SegV and FTSegV. In particular, after Step i, the under-segmentation updates the identified groups by merging two groups with the largest T_{ij} for i and j from two groups before subsequent analysis, while the over-segmentation regards each component of the transformed curve series as an individual group and then applies UniV componentwisely. For a fair comparison, the orders of VAR models adopted in SegV, FTSegV and UniV are determined by the AIC criterion without any fine-tuning being applied, whereas GSY is implemented using the R package `ftsa`.

To examine the predictive performance, we apply an expanding window approach

to the observed data $Y_{tj}(v_i)$ for $t = 1, \ldots, n, j = 1, \ldots, p, i = 1, \ldots, N$. We split the dataset into a training set and a test set respectively consisting of the first n_1 and the remaining n_2 observations. For any positive integer h, we implement each comparison method on the training set, obtain h-step ahead prediction on the test data based on the fitted model, increase the training size by one and repeat the above procedure n_2+1-h times and finally compute the h-step ahead mean absolute prediction error (MAPE) and mean squared prediction error (MSPE), respectively, defined as

$$\begin{aligned} \text{MAPE}(h) =& \frac{1}{(n_2+1-h)pN} \sum_{t=n_1+h}^{n} \sum_{j=1}^{p} \sum_{i=1}^{N} \left|\hat{Y}_{tj}(v_i) - Y_{tj}(v_i)\right|, \\ \text{MSPE}(h) =& \frac{1}{(n_2+1-h)pN} \sum_{t=n_1+h}^{n} \sum_{j=1}^{p} \sum_{i=1}^{N} \left\{\hat{Y}_{tj}(v_i) - Y_{tj}(v_i)\right\}^2. \end{aligned} \tag{3.41}$$

3.6.1 UK annual temperature data

The first dataset, which is available at https://www.metoffice.gov.uk/research/climate/maps-and-data/historic-station-data, consists of monthly mean temperature collected at $p = 22$ measuring stations across Britain from 1959 to 2020 ($n = 62$). Let $Y_{tj}(v_i)$ ($t = 1, \ldots, 62$, $j = 1, \ldots, 22$, $i = 1, \ldots, 12$) be the mean temperature during month $v_i = i$ of year $1958 + t$ measured at the j-th station. The observed temperature curves are smoothed using a 10-dimensional Fourier basis that characterizes the periodic pattern over the annual cycle. See Figure 3.1 in Appendix 3.B for plots of smoothed annual temperature curves. We divide the smoothed dataset into the training set of size $n_1 = 41$ and the test set of size $n_2 = 21$. Since the smoothed curve series exhibit very weak autocorrelations beyond $k = 3$ and the training size is relatively small, we use $k_0 = \kappa_0 = m = 3$ in this example.

The values of MAPE and MSPE for $h = 1, 2, 3$ defined in (3.41) are summarized in Table 3.5. Several obvious patterns are observable. First, our proposed SegV and FTSegV perform similarly well and both provide the highest predictive accuracies among all comparison methods for all h. This demonstrates the effectiveness of reducing the number of parameters via the segmentation in predicting high-dimensional functional time series, while the latent transformation matrix may not be approximately sparse in practice. Second, although the cases of under- and over-segmentation are slightly inferior to the correct-segmentation case, they significantly outperform UniV and GSY in one- and two-step-ahead predictions. It is worth noting that the over segmentation ignores all the correlations among different components of transformed curves, whereas UniV neglects those of original

curves. This observation reveals that the transformation step can also improve the prediction efficiently.

Table 3.5: Comparison of MAFEs and MSFEs for three versions of SegV, FTSegV, and two competitors on the UK temperature curves for $h = 1, 2, 3$. The lowest values are in bold font.

Method	MAFE			MSFE		
	$h=1$	$h=2$	$h=3$	$h=1$	$h=2$	$h=3$
SegV	**0.786**	0.806	**0.827**	**1.073**	1.075	**1.155**
Under.SegV	0.805	0.826	0.883	1.152	1.135	1.266
Over.SegV	0.797	0.821	0.845	1.101	1.126	1.174
FTSegV	0.789	**0.806**	0.828	1.077	**1.073**	1.158
Under.FTSegV	0.791	0.820	0.872	1.105	1.112	1.250
Over.FTSegV	0.797	0.821	0.845	1.101	1.126	1.174
UniV	0.936	0.951	0.976	1.450	1.450	1.458
GSY	0.894	0.884	0.854	1.346	1.338	1.219

3.6.2 Japanese mortality data

The second dataset, which can be downloaded from https://www.ipss.go.jp/p-toukei/JMD/index-en.html, contains age-specific and gender-specific mortality rates for $p = 47$ prefectures in Japan during 1975 to 2017 ($n = 43$). Following the recent proposal of Gao et al. (2019b), we model the log transformation of the mortality rate of people aged $v_i = i - 1$ living in the j-th prefecture during year $1974 + t$ as a random curve $Y_{tj}(v_i)$ ($t = 1, \ldots, 43$, $j = 1, \ldots, 47$, $i = 1, \ldots, 96$) and perform smoothing for observed mortality curves via smoothing splines. Figure 3.2 in Appendix 3.B displays exemplified trajectories of smoothed mortality curves. The post-sample prediction are carried out in an identical way to Section 3.6.1. We choose $k_0 = \kappa_0 = m = 3$ in our estimation and treat the smoothed curves in the first $n_1 = 33$ years and the last $n_2 = 10$ years as the training sample and the test sample, respectively.

Table 3.6 reports the MAPEs and MSPEs for Japanese females and males. Again it is obvious that SegV and FTSegV provide the best predictive performance uniformly for both females and males, and all h. One may also notice that, compared with SegV and Under.SegV, Over.SegV does not perform well for males. In most cases, the decorrelated curve series for males admits $\hat{q} = 44$ groups with 43 groups of size 1 and one large group of size 4. However, Over.SegV fails to account for the cross serial dependence within such large group, thus leading to less accurate predictions. On the other hand, the transformed curves for females reveal a common structure with one group of size 2 and the remaining groups of size 1. As expected, Over.SegV

performs slightly better in this case. This finding again confirms the effectiveness of our procedure, in particular, the within group cross dependence information is also valuable in the post-sample prediction.

Table 3.6: MAFEs and MSFEs for eight competing methods on the Japanese female and male mortality curves for $h = 1, 2, 3$. All numbers are multiplied by 10. The lowest values are in bold font.

	Method	MAFE			MSFE		
		$h = 1$	$h = 2$	$h = 3$	$h = 1$	$h = 2$	$h = 3$
Female	SegV	1.393	**1.414**	1.468	**0.482**	**0.470**	0.486
	Under.SegV	1.537	1.661	1.853	0.528	0.560	0.642
	Over.SegV	1.427	1.610	1.814	0.482	0.520	0.588
	FTSegV	**1.392**	1.417	**1.468**	0.484	0.471	**0.484**
	Under.FTSegV	1.542	1.661	1.846	0.533	0.560	0.638
	Over.FTSegV	1.433	1.617	1.816	0.484	0.523	0.588
	UniV	1.602	1.858	2.136	0.523	0.618	0.737
	GSY	1.618	1.691	1.682	0.678	0.733	0.706
Male	SegV	**1.374**	1.461	1.543	0.436	0.453	0.481
	Under.SegV	1.394	1.491	1.608	0.443	0.464	0.503
	Over.SegV	1.506	1.678	1.897	0.468	0.514	0.603
	FTSegV	1.376	**1.444**	**1.521**	**0.435**	**0.446**	**0.473**
	Under.FTSegV	1.389	1.482	1.596	0.440	0.460	0.499
	Over.FTSegV	1.512	1.673	1.894	0.470	0.512	0.604
	UniV	1.568	1.855	2.167	0.485	0.595	0.743
	GSY	1.550	1.581	1.576	0.669	0.663	0.628

3.6.3 Energy consumption data

Our third dataset contains energy consumption readings (in kWh) taken at half hourly intervals for thousands of London households, and is available at https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households. In our study, we select households with flat energy prices during the period between December 2012 and May 2013 ($n = 182$) after removing samples with too many missing records, and hence construct 4000 samples of daily energy consumption curves observed at $T = 48$ equally spaced time points. To alleviate the impact of randomness from individual curves, we randomly split the data into p groups of equal size, then take the sample average of curves within each group and finally smooth the averaged curves based on a 15-dimensional Fourier basis. See Figure 3.3 in Appendix 3.B for some examples of the smoothed intraday consumption curves. We target to evaluate the h-day ahead predictive accuracy for intraday energy consumption curves in May 2013 based on the training data from December 2012 to the previous day. The eight comparison methods are built in the same manner as Section 3.6.1 with $k_0 = \kappa_0 = m = 5$.

Table 3.7 presents the mean prediction errors for $h = 1, 2, 3$ and $p = 40, 80$. A few trends are apparent. First, the prediction errors for $p = 80$ are higher than those for $p = 40$ as higher dimensionality poses more challenges in prediction. Second, likewise in previous examples, SegV and FTSegV attain the lowest prediction errors in comparison to five competing methods under all scenarios. All segmentation-based methods consistently outperform UniV and GSY by a large margin. Third, despite being developed for high-dimensional functional time series prediction, GSY provides the worst result in this example.

Table 3.7: MAFEs and MSFEs for eight competing methods on the energy consumption curves for $h = 1, 2, 3$ and $p = 40, 80$. All numbers are multiplied by 10^2. The lowest values are in bold font.

	Method	MAFE			MSFE		
		$h = 1$	$h = 2$	$h = 3$	$h = 1$	$h = 2$	$h = 3$
$p = 40$	SegV	1.639	1.748	1.793	0.047	0.053	0.054
	Under.SegV	1.669	1.766	1.794	0.048	0.054	0.054
	Over.SegV	1.709	1.873	1.964	0.049	0.058	0.062
	FTSegV	**1.637**	**1.747**	**1.791**	**0.047**	**0.053**	**0.054**
	Under.FTSegV	1.669	1.766	1.793	0.048	0.054	0.054
	Over.FTSegV	1.708	1.872	1.963	0.049	0.058	0.062
	UniV	1.867	2.009	2.109	0.058	0.067	0.072
	GSY	2.142	2.264	2.32	0.099	0.110	0.119
$p = 80$	SegV	**1.996**	2.058	2.071	**0.070**	0.075	0.075
	Under.SegV	2.025	2.092	2.104	0.072	0.077	0.077
	Over.SegV	2.022	2.132	2.187	0.070	0.078	0.081
	FTSegV	2.012	**2.055**	**2.070**	0.071	**0.074**	**0.074**
	Under.FTSegV	2.040	2.087	2.104	0.073	0.076	0.077
	Over.FTSegV	2.045	2.138	2.190	0.072	0.078	0.081
	UniV	2.221	2.362	2.463	0.083	0.093	0.100
	GSY	2.833	2.826	2.781	0.159	0.159	0.159

3.A Additional Results and Proofs

We begin by introducing some notation. For a vector $\mathbf{b} \in \mathbb{R}^p$, we denote its ℓ_2 norm by $\|\mathbf{b}\|_2 = (\sum_{j=1}^p |b_j|^2)^{1/2}$. For a matrix $\mathbf{B} \in \mathbb{R}^{p\times q}$, we let $\|\mathbf{B}\|_2 = \lambda_{\max}^{1/2}(\mathbf{B}^{\mathrm{T}}\mathbf{B})$, where $\lambda_{\max}(\mathbf{M})$ denotes the largest eigenvalue of the matrix $\mathbf{M}$. For $\boldsymbol{\mathcal{B}} = (\mathcal{B}_{ij})_{p\times p}$ with its (i,j)-th component $\mathcal{B}_{ij} \in \mathbb{S}$, we define the functional version of matrix ℓ_1 norm by $\|\boldsymbol{\mathcal{B}}\|_{\mathcal{S},1} = \max_j \sum_i \|\mathcal{B}_{ij}\|_{\mathcal{S}}$. We use $\otimes$ to denote the Kronecker product. For two positive sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n \lesssim b_n$ or $b_n \gtrsim a_n$ if there exist a positive constant c such that $a_n/b_n \leqslant c$. Throughout, we use c, c_0 to denote generic positive finite constants that may be different in different uses.

3.A.1 Proofs of main theorems

Proof of Theorem 3.1. Recall $\mathbf{W}_y$ in (3.6) and $\widehat{\mathbf{W}}_y$ in (3.15). It follows from Lemma 3.4 and fixed κ_0 under Condition 3.4 that

$$\begin{aligned}
&\|\widehat{\mathbf{W}}_y - \mathbf{W}_y\|_2 \\
\leqslant& \sum_{k=0}^{\kappa_0} \left\| \int\int \mathcal{T}_{\omega_k}(\widehat{\boldsymbol{\Sigma}}^{\mathrm{S}}_{y,k})(u,v)\{\mathcal{T}_{\omega_k}(\widehat{\boldsymbol{\Sigma}}^{\mathrm{S}}_{y,k})(u,v)\}^{\mathrm{T}} - \boldsymbol{\Sigma}_{y,k}(u,v)\{\boldsymbol{\Sigma}_{y,k}(u,v)\}^{\mathrm{T}}\, dudv \right\|_2 \\
=& O_p\left\{ \Xi^2 \mathcal{M}^{1-\alpha} \left(\frac{\log p}{n}\right)^{\frac{1-\alpha}{2}} \right\}.
\end{aligned} \tag{3.42}$$

Due to the fact that $\max_{1\leqslant j\leqslant p} |\lambda_j(\widehat{\mathbf{W}}_y) - \lambda_j(\mathbf{W}_y)| \leqslant \|\widehat{\mathbf{W}}_y - \mathbf{W}_y\|_2$, we obtain $\max_{1\leqslant j\leqslant p} |\lambda_j(\widehat{\mathbf{W}}_y) - \lambda_j(\mathbf{W}_y)| \to_p 0$ given $\Xi(p)^2\mathcal{M}^{1-\alpha}\,(\log p/n)^{(1-\alpha)/2} = o(1)$. Recall $\mathbf{W}_y$ and $\mathbf{W}_z = \mathrm{diag}(\mathbf{W}_{z,1}, \ldots, \mathbf{W}_{z,q})$ share the same eigenvalues, and $\mathbf{W}_{z,l_1}$ and $\mathbf{W}_{z,l_2}$ do not share same eigenvalues if $l_1 \neq l_2$. Therefore, there exists a map $\pi : \{1,\ldots,p\} \to \{1,\ldots,q\}$ such that each $\lambda_j(\widehat{\mathbf{W}}_y)$ converges to some eigenvalue of $\mathbf{W}_{z,\pi(j)}$. Recall $\widehat{\boldsymbol{\eta}}_j$ is the eigenvector of $\widehat{\mathbf{W}}_y$ associated with $\lambda_j(\widehat{\mathbf{W}}_y)$. For each $l = 1,\ldots,q$, we let $\widehat{\mathbf{A}}_l$ be a $p \times p_l$ matrix whose columns are $\widehat{\boldsymbol{\eta}}_j$ with $\pi(j) = l$. If we view $\widehat{\mathbf{A}}_l$, $\widehat{\mathbf{W}}_y$ and $\mathbf{W}_y$ as $\mathbf{Q}_1$, $\mathbf{B}$ and $\mathbf{B}+\mathbf{E}$ in Lemma 3.7 respectively, then the corresponding $\mathbf{Q}_1^\star$, denoted by $\mathbf{H}_l$, provides an orthonormal basis of $\mathcal{C}(\mathbf{A}_l)$. Therefore, applying Lemma 3.7 yields that, for each l,

$$\|\widehat{\mathbf{A}}_l - \mathbf{H}_l\|_2 \leqslant\ 8\rho_l^{-1}\|\widehat{\mathbf{W}}_y - \mathbf{W}_y\|_2. \tag{3.43}$$

Write $\widehat{\boldsymbol{\Gamma}}_{G_l} = (\widehat{\boldsymbol{\eta}}_j)_{j\in G_l} \in \mathbb{R}^{p\times p_l}$. Note that the columns of $\widehat{\mathbf{A}}_l$ are also permutation of

$\{\hat{\boldsymbol{\eta}}_j, j \in G_l\}$. By (3.43), we have that

$$\begin{aligned}\|\widehat{\boldsymbol{\Gamma}}_{G_l}\widehat{\boldsymbol{\Gamma}}_{G_l}^{\mathrm{T}} - \boldsymbol{\Gamma}_{G_l}\boldsymbol{\Gamma}_{G_l}^{\mathrm{T}}\|_2 &= \|\widehat{\mathbf{A}}_l\widehat{\mathbf{A}}_l^{\mathrm{T}} - \mathbf{H}_l\mathbf{H}_l^{\mathrm{T}}\|_2 \\ &\leqslant \|\widehat{\mathbf{A}}_l - \mathbf{H}_l\|_2^2 + 2\|\mathbf{H}_l\|_2\|\widehat{\mathbf{A}}_l - \mathbf{H}_l\|_2 \\ &\lesssim \rho_l^{-1}\|\widehat{\mathbf{W}}_y - \mathbf{W}_y\|_2.\end{aligned}$$

Let $\mathbf{U}\mathrm{diag}(d_1, \ldots, d_{p_l})\widehat{\mathbf{U}}^{\mathrm{T}}$ be the SVD of $\boldsymbol{\Gamma}_{G_l}\boldsymbol{\Gamma}_{G_l}^{\mathrm{T}}\widehat{\boldsymbol{\Gamma}}_{G_l}\widehat{\boldsymbol{\Gamma}}_{G_l}^{\mathrm{T}}$ where $d_1 \geqslant \cdots \geqslant d_{p_l}$. Define

$$\boldsymbol{\Pi}_{G_l} = \mathbf{U}\widehat{\mathbf{U}}^{\mathrm{T}}\widehat{\boldsymbol{\Gamma}}_{G_l}, \quad l = 1, \ldots, q. \tag{3.44}$$

Hence, there exists a $p \times p$ matrix $\boldsymbol{\Pi} \equiv (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_p)$ such that $\boldsymbol{\Pi}_{G_l} = (\boldsymbol{\gamma}_j)_{j \in G_l}$. Following the same techniques as in the proof of Theorem 2 in Han et al. (2021), we have that $\boldsymbol{\Pi}_{G_l}\boldsymbol{\Pi}_{G_l}^{\mathrm{T}} = \boldsymbol{\Gamma}_{G_l}\boldsymbol{\Gamma}_{G_l}^{\mathrm{T}}$ and

$$\begin{aligned}\max_{1\leqslant j\leqslant p}\rho_l\|\hat{\boldsymbol{\eta}}_j - \boldsymbol{\gamma}_j\|_2 \leqslant \max_{1\leqslant l\leqslant q}\rho_l\|\widehat{\boldsymbol{\Gamma}}_{G_l} - \boldsymbol{\Pi}_{G_l}\|_2 &\leqslant \max_{1\leqslant l\leqslant q}\sqrt{2}\rho_l\|\widehat{\boldsymbol{\Gamma}}_{G_l}\widehat{\boldsymbol{\Gamma}}_{G_l}^{\mathrm{T}} - \boldsymbol{\Gamma}_{G_l}\boldsymbol{\Gamma}_{G_l}^{\mathrm{T}}\|_2 \\ &\lesssim \|\widehat{\mathbf{W}}_y - \mathbf{W}_y\|_2,\end{aligned} \tag{3.45}$$

Denote by $E = \Big\{(i,j) : \max_{0\leqslant k\leqslant m}\max\big[\|\boldsymbol{\gamma}_i^{\mathrm{T}}\boldsymbol{\Sigma}_{y,k}\boldsymbol{\gamma}_j\|_{\mathcal{S}}, \|\boldsymbol{\gamma}_i^{\mathrm{T}}\{\boldsymbol{\Sigma}_{y,k}\}^{\mathrm{T}}\boldsymbol{\gamma}_j\|_{\mathcal{S}}\big] > 0\Big\}$ the edge set of $G = \{1, \ldots, p\}$ under $\boldsymbol{\Pi}$. The true segmented groups in (3.35), defined according to the ordered eigenvectors of $\mathbf{W}_y$, can also be found via splitting $\{G, E\}$ into multiple connected subgraphs (G_l, E_l), for $l = 1, \ldots, q$.

Define $\varpi_n = \Xi\mathcal{M}^{1-\alpha}(\log p/n)^{(1-\alpha)/2}$. Consider the event $\Omega_n = \big\{\max_{0\leqslant k\leqslant m}\|\mathcal{T}_{\omega_k}(\widehat{\boldsymbol{\Sigma}}_{y,k}^{\mathrm{S}}) - \boldsymbol{\Sigma}_{y,k}\|_{\mathcal{S},1} \leqslant c_0\varpi_n\big\}$. By (3.42) and (3.45), it is immediate to see that there exists some constant c such that $\max_{1\leqslant j\leqslant p}\|\hat{\boldsymbol{\eta}}_j - \boldsymbol{\gamma}_j\|_2 \leqslant c\rho^{-1}\nu_n$, where $\nu_n = \Xi^2\mathcal{M}_y^{1-\alpha}(\log p/n)^{(1-\alpha)/2}$. By Condition 3.5 and Lemma 3.6, for each $(i,j) \in E$,

$$\begin{aligned}\widehat{T}_{ij} &= \max_{0\leqslant k\leqslant m}\max\Big\{\|\hat{\boldsymbol{\eta}}_i^{\mathrm{T}}\mathcal{T}_{\omega_k}(\widehat{\boldsymbol{\Sigma}}_{y,k}^{\mathrm{S}})\hat{\boldsymbol{\eta}}_j\|_{\mathcal{S}}, \|\hat{\boldsymbol{\eta}}_i^{\mathrm{T}}\{\mathcal{T}_{\omega_k}(\widehat{\boldsymbol{\Sigma}}_{y,k}^{\mathrm{S}})\}^{\mathrm{T}}\hat{\boldsymbol{\eta}}_j\|_{\mathcal{S}}\Big\}, \\ &\geqslant \varsigma - \max_{1\leqslant j\leqslant p}\|\hat{\boldsymbol{\eta}}_j - \boldsymbol{\gamma}_j\|_2\max_{0\leqslant k\leqslant m}\|\boldsymbol{\Sigma}_{y,k}\|_{\mathcal{S},1} - \max_{0\leqslant k\leqslant m}\|\mathcal{T}_{\omega_k}(\widehat{\boldsymbol{\Sigma}}_{y,k}^{\mathrm{S}}) - \boldsymbol{\Sigma}_{y,k}\|_{\mathcal{S},1},\end{aligned}$$

where $\varsigma = \min_{1\leqslant l\leqslant q}\varsigma_l$. Combining the above results with (3.49) (in the proof of Lemma 3.4) under the event Ω_n yields that

$$\min_{(i,j)\in E}\widehat{T}_{ij} \geqslant \varsigma - c\rho^{-1}\Xi\nu_n,$$

where $\rho = \min_{1\leqslant l\leqslant q}\rho_l$. Similarly, we have that

$$\max_{(i,j)\notin E}\hat{T}_{ij} \leqslant c\rho^{-1}\Xi\nu_n.$$

Letting $c\rho^{-1}\Xi\nu_n < \tau_n < \varsigma - c\rho^{-1}\Xi\nu_n$, we obtain the exact recovery of the true segmentation $\{G_1,\ldots,G_q\}$ in the sense of $\hat{E} = E$. Note it follows from Lemma 3.3 that $P(\Omega_n^C) = o(1)$. Hence we complete the proof of Theorem 3.1. □

Proof of Theorem 3.2. By (3.42), (3.43) and the remark for Lemma 1 of Chang et al. (2018), we obtain that

$$\max_{1\leqslant l\leqslant q}\rho_l D(\mathcal{C}(\widehat{\mathbf{A}}_l),\mathcal{C}(\mathbf{A}_l)) \lesssim \|\widehat{\mathbf{W}}_y - \mathbf{W}_y\|_2 = O_p\left\{\Xi^2\mathcal{M}^{1-\alpha}\left(\frac{\log p}{n}\right)^{\frac{1-\alpha}{2}}\right\},$$

which completes the proof. □

Proof of Theorem 3.3. We follow the same notation as in the proofs of Theorems 3.1 and 3.2. Recall that $\widehat{\mathbf{\Gamma}}_{G_l} = (\hat{\boldsymbol{\eta}}_j)_{j\in G_l} \equiv (\hat{\boldsymbol{\eta}}_{l,1},\ldots,\hat{\boldsymbol{\eta}}_{l,p_l})$ and $\mathbf{\Pi}_{G_l} = (\boldsymbol{\gamma}_j)_{j\in G_l} \equiv (\boldsymbol{\gamma}_{l,1}\ldots,\boldsymbol{\gamma}_{l,p_l})$. We now focus our analysis based on $\mathbf{\Pi}$ in the sense that $\mathbf{M}_{k,l}(u,v) = \mathbf{\Pi}_{G_l}^{\mathrm{T}}\mathbf{\Sigma}_{y,k}(u,v)\mathbf{\Pi}_{G_l} \equiv \{M_{ij}^{(k,l)}(u,v)\}_{p_l\times p_l}$ for each true group $l = 1,\ldots,q$. Recall that

$$\widehat{\mathbf{M}}_{k,l}(u,v) = \widehat{\mathbf{A}}_l^{\mathrm{T}}\mathcal{T}_{\omega_k}(\widehat{\mathbf{\Sigma}}_{y,k}^{\mathrm{S}})(u,v)\widehat{\mathbf{A}}_l = \widehat{\mathbf{\Gamma}}_{G_l}^{\mathrm{T}}\mathcal{T}_{\omega_k}(\widehat{\mathbf{\Sigma}}_{y,k}^{\mathrm{S}})(u,v)\widehat{\mathbf{\Gamma}}_{G_l} \equiv \{\widehat{M}_{ij}^{(k,l)}(u,v)\}_{p_l\times p_l}.$$

Notice that

$$\widehat{M}_{ij}^{(k,l)} - M_{ij}^{(k,l)} = \hat{\boldsymbol{\eta}}_{l,i}^{\mathrm{T}}\mathcal{T}_{\omega_k}(\widehat{\mathbf{\Sigma}}_{y,k}^{\mathrm{S}})\hat{\boldsymbol{\eta}}_{l,j} - \boldsymbol{\gamma}_{l,i}^{\mathrm{T}}\mathbf{\Sigma}_{y,k}\boldsymbol{\gamma}_{l,j} = I_1 + I_2 + I_3 + I_4 + I_5,$$

where $I_1 = (\hat{\boldsymbol{\eta}}_{l,i} - \boldsymbol{\gamma}_{l,i})^{\mathrm{T}}\{\mathcal{T}_{\omega_k}(\widehat{\mathbf{\Sigma}}_{y,k}^{\mathrm{S}}) - \mathbf{\Sigma}_{y,k}\}\hat{\boldsymbol{\eta}}_{l,j}$, $I_2 = (\hat{\boldsymbol{\eta}}_{l,i} - \boldsymbol{\gamma}_{l,i})^{\mathrm{T}}\mathbf{\Sigma}_{y,k}(\hat{\boldsymbol{\eta}}_{l,j} - \boldsymbol{\gamma}_{l,j})$, $I_3 = (\hat{\boldsymbol{\eta}}_{l,i} - \boldsymbol{\gamma}_{l,i})^{\mathrm{T}}\mathbf{\Sigma}_{y,k}\boldsymbol{\gamma}_{l,j}$, $I_4 = \boldsymbol{\gamma}_{l,i}^{\mathrm{T}}\{\mathcal{T}_{\omega_k}(\widehat{\mathbf{\Sigma}}_{y,k}^{\mathrm{S}}) - \mathbf{\Sigma}_{y,k}\}\hat{\boldsymbol{\eta}}_{l,j}$ and $I_5 = \boldsymbol{\gamma}_{l,i}^{\mathrm{T}}\mathbf{\Sigma}_{y,k}(\hat{\boldsymbol{\eta}}_{l,j} - \boldsymbol{\gamma}_{l,j})$. By (3.45), (3.49), the orthonormality of $\hat{\boldsymbol{\eta}}_{l,j}$, $\boldsymbol{\gamma}_{l,j}$ and Lemmas 3.3, 3.6, we obtain that

$$\max_{i,j,l}\rho_l|I_1| = O_p(\Xi^{-1}\nu_n^2), \quad \max_{i,j,l}\rho_l^2|I_2| = O_p(\Xi\nu_n^2),$$

$$\max_{i,j,l}|I_4| = O_p(\Xi^{-1}\nu_n), \quad \max_{i,j,l}\rho_l(|I_3| + |I_5|) = O_p(\Xi\nu_n).$$

The above results lead to

$$\max_{i,j,l} \rho_l \|\widehat{M}_{ij}^{(k,l)} - M_{ij}^{(k,l)}\|_{\mathcal{S}} = O_p(\Xi\nu_n),$$

which together with Condition 3.6 implies that

$$\max_{1\leqslant l\leqslant q} \rho_l \|\widehat{\mathbf{M}}_{k,l} - \mathbf{M}_{k,l}\|_{\mathcal{S},\mathrm{F}} = O_p(\Xi\nu_n).$$

Write $\mathbf{Z}_{t,l}(\cdot) = (Z_{t,1}^{(l)}(\cdot), \ldots, Z_{t,p_l}^{(l)}(\cdot))^{\mathrm{T}}$. It follows from Cauchy–Schwartz inequality and Condition 3.6 that

$$\begin{aligned}
&\max_l \|\mathbf{M}_{k,l}\|_{\mathcal{S},\mathrm{F}} = \max_l \sqrt{\sum_{i,j} \int\int \{M_{ij}^{(k,l)}(u,v)\}^2 \mathrm{d}u\mathrm{d}v} \\
\leqslant &\max_l \sqrt{\sum_i \int \mathbb{E}[\{Z_{t,i}^{(l)}(u)\}^2]du} \sqrt{\sum_j \int \mathbb{E}[\{Z_{t+k,j}^{(l)}(u)\}^2]du} \leqslant \max_l \mathbb{E}\big(\|\mathbf{Z}_{t,l}\|^2\big) = O(1).
\end{aligned}$$

Combining the above results, we have

$$\begin{aligned}
&\max_{1\leqslant l\leqslant q} \rho_l \|\widehat{\mathbf{K}}_l - \mathbf{K}_l\|_{\mathcal{S},\mathrm{F}} \\
\leqslant &\max_l \sum_{k=1}^{k_0} \rho_l \|\widehat{\mathbf{M}}_{k,l} - \mathbf{M}_{k,l}\|_{\mathcal{S},\mathrm{F}}^2 + 2\max_l \rho_l \sum_{k=1}^{k_0} \|\mathbf{M}_{k,l}\|_{\mathcal{S},\mathrm{F}} \|\widehat{\mathbf{M}}_{k,l} - \mathbf{M}_{k,l}\|_{\mathcal{S},\mathrm{F}} \qquad (3.46) \\
= &O_p(\Xi\nu_n).
\end{aligned}$$

This, together with Theorem 2 of Chiou et al. (2014), implies that

$$\max_{1\leqslant l\leqslant q} \rho_l |\hat{\theta}_{l,j} - \theta_{l,j}| = O_p(\Xi\nu_n), \quad \max_{1\leqslant l\leqslant q} \rho_l \|\hat{\boldsymbol{\psi}}_{j,l} - \boldsymbol{\psi}_{j,l}\| = O_p(\Xi\nu_n). \qquad (3.47)$$

Recall that

$$\hat{r}_l = \operatorname*{argmax}_{1\leqslant j\leqslant n-k_0} \frac{\hat{\theta}_{l,j} + \delta_n}{\hat{\theta}_{l,j+1} + \delta_n}.$$

The condition $\delta_n \max_l \theta_{l,1} / \min_l \theta_{l,r_l}^2 \to 0$ with $\delta_n = \rho^{-1}\Xi\nu_n$, implies that $\delta_n = o(\min_l \theta_{l,r_l})$. Combining this with (3.47), we obtain for $j < r_l$,

$$\max_{1\leqslant l\leqslant q} \frac{\hat{\theta}_{l,j} + \delta_n}{\hat{\theta}_{l,j+1} + \delta_n} = \max_{1\leqslant l\leqslant q} \frac{\hat{\theta}_{l,j} - \theta_{l,j} + \theta_{l,j} + \delta_n}{\hat{\theta}_{l,j+1} - \theta_{l,j+1} + \theta_{l,j+1} + \delta_n} \to_p \frac{\max_{1\leqslant l\leqslant q} \theta_{l,1}}{\min_{1\leqslant l\leqslant q} \theta_{l,r_l}}.$$

For $j = r_l$,

$$\min_{1\leqslant l\leqslant q}\frac{\hat{\theta}_{l,r_l}+\delta_n}{\hat{\theta}_{l,r_l+1}+\delta_n}=\min_{1\leqslant l\leqslant q}\frac{\hat{\theta}_{l,r_l}-\theta_{l,r_l}+\theta_{l,r_l}+\delta_n}{\hat{\theta}_{l,r_l+1}+\delta_n}\to_p\frac{\min_{1\leqslant l\leqslant q}\theta_{l,r_l}}{\delta_n}.$$

For $j > r_l$, we have

$$\max_{1\leqslant l\leqslant q}\frac{\hat{\theta}_{l,j}+\delta_n}{\hat{\theta}_{l,j+1}+\delta_n}\to_p C.$$

Hence, under the condition of $\delta_n \max_l \theta_{l,1}/\min_l \theta^2_{l,r_l} \to 0$, r_l is correctly identified via $\hat{r}_l$ uniformly over $\{1,\ldots,q\}$. □

Proof of Theorem 3.4. By (3.47), the definition of $D\{\mathcal{C}(\hat{\boldsymbol{\psi}}_l),\mathcal{C}(\boldsymbol{\psi}_l)\}$ and the orthonormality of $\boldsymbol{\psi}_{j,l}$, we obtain that

$$\begin{aligned}\max_{1\leqslant l\leqslant q}\rho_l\sqrt{2r_l}D\{\mathcal{C}(\hat{\boldsymbol{\psi}}_l),\mathcal{C}(\boldsymbol{\psi}_l)\}&=\max_{1\leqslant l\leqslant q}\rho_l\left\|\sum_{j=1}^{r_l}\left(\hat{\boldsymbol{\psi}}_{l,j}\otimes\hat{\boldsymbol{\psi}}_{l,j}^{\mathrm{T}}-\boldsymbol{\psi}_{l,j}\otimes\boldsymbol{\psi}_{l,j}^{\mathrm{T}}\right)\right\|_{\mathcal{S},\mathrm{F}}\\&\leqslant\max_{1\leqslant l\leqslant q}\rho_l\sum_{j=1}^{r_l}\|\hat{\boldsymbol{\psi}}_{j,l}-\boldsymbol{\psi}_{j,l}\|^2+2\max_{1\leqslant l\leqslant q}\rho_l\sum_{j=1}^{r_l}\|\boldsymbol{\psi}_{j,l}\|\|\hat{\boldsymbol{\psi}}_{j,l}-\boldsymbol{\psi}_{j,l}\|\\&=O_p(\Xi\nu_n).\square\end{aligned}$$

3.A.2 Technical lemmas and their proofs

Proof of Lemma 3.1. Recall that $\int\mathbb{E}[\{Z_{tj}(u)\}^2]\mathrm{d}u = 1$ as discussed in Section 3.2.1. Hence,

$$\begin{aligned}\|\Sigma^{(k)}_{z,ij}\|^2_{\mathcal{S}}&=\int\int[\mathbb{E}\{Z_{ti}(u)Z_{(t+k)j}(v)\}]^2\mathrm{d}u\mathrm{d}v\\&\leqslant\int\mathbb{E}[\{Z_{ti}(u)\}^2]\mathrm{d}u\int\mathbb{E}[\{Z_{(t+k)j}(v)\}^2]\mathrm{d}v\leqslant 1.\end{aligned}$$

By the inequality $(a+b)^\alpha \leqslant a^\alpha + b^\alpha$ for $x, y \geqslant 0$ and $\alpha \in [0,1)$, we obtain that

$$\begin{aligned}
\sum_{i=1}^{p} \|\Sigma_{y,ij}^{(k)}\|_{\mathcal{S}}^{\alpha} &\leqslant \sum_{i=1}^{p}\sum_{l,m=1}^{p} |A_{il}|^{\alpha}|A_{jm}|^{\alpha}\|\Sigma_{z,lm}^{(k)}\|_{\mathcal{S}}^{\alpha} \\
&\leqslant \max_{l,m,k}\|\Sigma_{z,lm}^{(k)}\|_{\mathcal{S}}^{\alpha} \max_{l}\sum_{i=1}^{p}|A_{il}|^{\alpha} \sum_{|l-m|\leqslant \max_l p_l} |A_{jm}|^{\alpha} \\
&\leqslant s_2 \max_{l,m,k}\|\Sigma_{z,lm}^{(k)}\|_{\mathcal{S}}^{\alpha}(2\max_{l} p_l + 1)\sum_{m=1}^{p}|A_{jm}|^{\alpha} \\
&\leqslant s_1 s_2 (2\max_{l} p_l + 1)
\end{aligned}$$

In the same manner, we can prove the second result in this lemma. ▫

Lemma 3.2. *Suppose that Conditions 3.2 and 3.3 hold for sub-Gaussian linear process* $\{\mathbf{Y}_t(\cdot)\}_{t\in\mathbb{Z}}$. *Then, for* $k = 1, \ldots, \kappa_0$, *there exists some universal constant* $\tilde{c} > 0$ *such that for any* $\eta > 0$ *and each* $i, j = 1, \ldots, p$,

$$P\left\{\|\widehat{\Sigma}_{y,ij}^{(k)} - \Sigma_{y,ij}^{(k)}\|_{\mathcal{S}} > \mathcal{M}\eta\right\} \leqslant 8\exp\{-\tilde{c}n\min(\eta^2, \eta)\}.$$

In particular, if sample size $n \gtrsim \log(p)$, *then for any* $M > 0$, *there exist some positive constant* $\tilde{c}_1 > 0$ *such that*

$$\max_{i,j}\left\|\widehat{\Sigma}_{y,ij}^{(k)} - \Sigma_{y,ij}^{(k)}\right\|_{\mathcal{S}} \leqslant \tilde{c}_1\mathcal{M}\sqrt{\frac{\log p}{n}}$$

with probability greater than $1 - O(p^{-M})$.

Proof. This lemma follows directly from Theorem 1 of Fang et al. (2022) and Theorem 2 of Guo and Qiao (2022) and hence the proof is omitted here. ▫

Lemma 3.3. *Suppose that Conditions 3.1–3.3 hold. Then there exists some constant* $\delta > 0$ *such that for* $k \leqslant \kappa_0$, *if* $\omega_k = \delta\sqrt{\log p/n}$, *it holds that*

$$\|\mathcal{T}_{\omega_k}(\widehat{\mathbf{\Sigma}}_{y,k}^{\mathrm{S}}) - \mathbf{\Sigma}_{y,k}\|_{\mathcal{S},1} = O_p\left\{\Xi\mathcal{M}^{1-\alpha}\left(\frac{\log p}{n}\right)^{\frac{1-\alpha}{2}}\right\},$$

$$\|\mathcal{T}_{\omega_k}(\widehat{\mathbf{\Sigma}}_{y,k}^{\mathrm{S}}) - \mathbf{\Sigma}_{y,k}\|_{\mathcal{S},\infty} = O_p\left\{\Xi\mathcal{M}^{1-\alpha}\left(\frac{\log p}{n}\right)^{\frac{1-\alpha}{2}}\right\}.$$

Proof. Denote by $\mathcal{T}_{\omega_k}(\widehat{\Sigma}_{y,ij}^{(k)})$ the (i,j)-th component of $\mathcal{T}_{\omega_k}(\widehat{\mathbf{\Sigma}}_{y,k}^{\mathrm{S}})$. Under the event

of $\max_{i,j}\left\|\widehat{\Sigma}_{y,ij}^{(k)}-\Sigma_{y,ij}^{(k)}\right\|_{\mathcal{S}}\leqslant\omega_k$, we have that

$$
\begin{aligned}
&\max_j\sum_{i=1}^p\|\mathcal{T}_{\omega_k}(\widehat{\Sigma}_{y,ij}^{(k)})-\Sigma_{y,ij}^{(k)}\|_{\mathcal{S}}\\
=&\max_j\sum_{i=1}^p\|\mathcal{T}_{\omega_k}(\widehat{\Sigma}_{y,ij}^{(k)})-\Sigma_{y,ij}^{(k)}\|_{\mathcal{S}}I\{\|\widehat{\Sigma}_{y,ij}^{(k)}\|_{\mathcal{S}}\geqslant\omega_k\}\\
&+\max_j\sum_{i=1}^p\|\mathcal{T}_{\omega_k}(\widehat{\Sigma}_{y,ij}^{(k)})-\Sigma_{y,ij}^{(k)}\|_{\mathcal{S}}I\{\|\widehat{\Sigma}_{y,ij}^{(k)}\|_{\mathcal{S}}<\omega_k\}\\
\leqslant&\max_j\sum_{i=1}^p\left\{\|\mathcal{T}_{\omega_k}(\widehat{\Sigma}_{y,ij}^{(k)})-\widehat{\Sigma}_{y,ij}^{(k)}\|_{\mathcal{S}}+\|\widehat{\Sigma}_{y,ij}^{(k)}-\Sigma_{y,ij}^{(k)}\|_{\mathcal{S}}\right\}I\{\|\widehat{\Sigma}_{y,ij}^{(k)}\|_{\mathcal{S}}\geqslant\omega_k,\|\Sigma_{y,ij}^{(k)}\|_{\mathcal{S}}\geqslant\omega_k\}\\
&+\max_j\sum_{i=1}^p\|\mathcal{T}_{\omega_k}(\widehat{\Sigma}_{y,ij}^{(k)})-\Sigma_{y,ij}^{(k)}\|_{\mathcal{S}}I\{\|\widehat{\Sigma}_{y,ij}^{(k)}\|_{\mathcal{S}}\geqslant\omega_k,\|\Sigma_{y,ij}^{(k)}\|_{\mathcal{S}}<\omega_k\}\\
&+\max_j\sum_{i=1}^p\|\Sigma_{y,ij}^{(k)}\|_{\mathcal{S}}I\{\|\widehat{\Sigma}_{y,ij}^{(k)}\|_{\mathcal{S}}<\omega_k\}\\
\leqslant&2\omega_k\sum_{i=1}^p I\{\|\Sigma_{y,ij}^{(k)}\|_{\mathcal{S}}\geqslant\omega_k\}+\max_j\sum_{i=1}^p\|\widehat{\Sigma}_{y,ij}^{(k)}-\Sigma_{y,ij}^{(k)}\|_{\mathcal{S}}I\{\|\widehat{\Sigma}_{y,ij}^{(k)}\|_{\mathcal{S}}\geqslant\omega_k,\|\Sigma_{y,ij}^{(k)}\|_{\mathcal{S}}<\omega_k\}\\
&+\max_j\sum_{i=1}^p\|\Sigma_{y,ij}^{(k)}\|_{\mathcal{S}}I\{\|\Sigma_{y,ij}^{(k)}\|_{\mathcal{S}}<2\omega_k\}\\
=&:Q_1+Q_2+Q_3
\end{aligned}
$$

By Lemma 3.1, we have that

$$
Q_1+Q_3\leqslant C_\alpha\omega_k^{1-\alpha}\sum_{i=1}^p\|\Sigma_{y,ij}^{(k)}\|_{\mathcal{S}}^\alpha\lesssim\omega_k^{1-\alpha}\Xi. \tag{3.48}
$$

To bound Q_2, for $\tilde{\theta}\in(0,1)$, we write

$$
\begin{aligned}
Q_2\leqslant&\max_j\sum_{i=1}^p\|\widehat{\Sigma}_{y,ij}^{(k)}-\Sigma_{y,ij}^{(k)}\|_{\mathcal{S}}I\{\|\widehat{\Sigma}_{y,ij}^{(k)}\|_{\mathcal{S}}\geqslant\omega_k,\|\Sigma_{y,ij}^{(k)}\|_{\mathcal{S}}\leqslant\tilde{\theta}\omega_k\}\\
&+\max_j\sum_{i=1}^p\|\widehat{\Sigma}_{y,ij}^{(k)}-\Sigma_{y,ij}^{(k)}\|_{\mathcal{S}}I\{\|\widehat{\Sigma}_{y,ij}^{(k)}\|_{\mathcal{S}}\geqslant\omega_k,\tilde{\theta}\omega_k<\|\Sigma_{y,ij}^{(k)}\|_{\mathcal{S}}<\omega_k\}\\
\leqslant&\omega_k\max_j\sum_{i=1}^p I\{\|\widehat{\Sigma}_{y,ij}^{(k)}-\Sigma_{y,ij}^{(k)}\|_{\mathcal{S}}>(1-\tilde{\theta})\omega_k\}+\omega_k^{1-\alpha}\tilde{\theta}^{-\alpha}\Xi
\end{aligned}
$$

By Lemma 3.2, we obtain that, for $n \gtrsim \log(p)$,

$$P\Big[\sum_{i=1}^{p} I\{\|\widehat{\Sigma}_{y,ij}^{(k)} - \Sigma_{y,ij}^{(k)}\|_{\mathcal{S}} > (1-\tilde{\theta})\omega_k\}\Big] = P\Big\{\max_{1\leqslant i,j\leqslant p} \|\widehat{\Sigma}_{y,ij}^{(k)} - \Sigma_{y,ij}^{(k)}\|_{\mathcal{S}} \geqslant (1-\tilde{\theta})\omega_k\Big\}$$
$$\leqslant 8p^2 \exp\{-\tilde{c}(1-\tilde{\theta})^2 \log p\} \to 0,$$

Hence, $Q_2 \lesssim \omega_k^{1-\alpha}\Xi$. This, together with (3.48), implies that

$$\max_j \sum_{i=1}^{p} \|\mathcal{T}_{\omega_k}(\widehat{\Sigma}_{y,ij}^{(k)}) - \Sigma_{y,ij}^{(k)}\|_{\mathcal{S}} = O_p\left\{\Xi\mathcal{M}^{1-\alpha}\left(\frac{\log p}{n}\right)^{\frac{1-\alpha}{2}}\right\}.$$

The second result can be proved in the similar manner. Hence, the proof is complete. □

Lemma 3.4. *Suppose Conditions 3.1–3.3 hold. Then we have that*

$$\Big\|\int\int\Big[\mathcal{T}_{\omega_k}(\widehat{\mathbf{\Sigma}}_{y,k}^{\mathrm{S}})(u,v)\{\mathcal{T}_{\omega_k}(\widehat{\mathbf{\Sigma}}_{y,k}^{\mathrm{S}})(u,v)\}^{\mathrm{T}} - \mathbf{\Sigma}_{y,k}(u,v)\{\mathbf{\Sigma}_{y,k}(u,v)\}^{\mathrm{T}}\Big]\mathrm{d}u\mathrm{d}v\Big\|_2$$
$$= O_p\left\{\Xi^2\mathcal{M}^{1-\alpha}\left(\frac{\log p}{n}\right)^{\frac{1-\alpha}{2}}\right\}$$

Proof. It follows from Lemma 3.5 that

$$\Big\|\int\int\Big[\mathcal{T}_{\omega_k}(\widehat{\mathbf{\Sigma}}_{y,k}^{\mathrm{S}})(u,v)\{\mathcal{T}_{\omega_k}(\widehat{\mathbf{\Sigma}}_{y,k}^{\mathrm{S}})(u,v)\}^{\mathrm{T}} - \mathbf{\Sigma}_{y,k}(u,v)\{\mathbf{\Sigma}_{y,k}(u,v)\}^{\mathrm{T}}\Big]\mathrm{d}u\mathrm{d}v\Big\|_2$$
$$\leqslant 2\sqrt{\|\mathbf{\Sigma}_{y,k}\|_{\mathcal{S},1}\|\mathbf{\Sigma}_{y,k}\|_{\mathcal{S},\infty}\|\mathcal{T}_{\omega_k}(\widehat{\mathbf{\Sigma}}_{y,k}^{\mathrm{S}}) - \mathbf{\Sigma}_{y,k}\|_{\mathcal{S},1}\|\mathcal{T}_{\omega_k}(\widehat{\mathbf{\Sigma}}_{y,k}^{\mathrm{S}}) - \mathbf{\Sigma}_{y,k}\|_{\mathcal{S},\infty}}$$
$$+\|\mathcal{T}_{\omega_k}(\widehat{\mathbf{\Sigma}}_{y,k}^{\mathrm{S}}) - \mathbf{\Sigma}_{y,k}\|_{\mathcal{S},1}\|\mathcal{T}_{\omega_k}(\widehat{\mathbf{\Sigma}}_{y,k}^{\mathrm{S}}) - \mathbf{\Sigma}_{y,k}\|_{\mathcal{S},\infty}.$$

By Lemma 3.1, we have that

$$\begin{aligned}\|\mathbf{\Sigma}_{y,k}\|_{\mathcal{S},1} &= \max_j \sum_{i=1}^{p} \|\Sigma_{y,ij}^{(k)}\|_{\mathcal{S}} \leqslant \Xi \max_{i,j} \|\Sigma_{y,ij}^{(k)}\|_{\mathcal{S}}^{1-\alpha} \lesssim \Xi, \\ \|\mathbf{\Sigma}_{y,k}\|_{\mathcal{S},\infty} &= \max_i \sum_{j=1}^{p} \|\Sigma_{y,ij}^{(k)}\|_{\mathcal{S}} \leqslant \Xi \max_{i,j} \|\Sigma_{y,ij}^{(k)}\|_{\mathcal{S}}^{1-\alpha} \lesssim \Xi.\end{aligned} \tag{3.49}$$

Combining the above results with Lemma 3.3, we complete the proof. □

Lemma 3.5. *Let* $\mathbf{B}_1 = (B_{1,ij})_{p\times p}$ *with each* $B_{1,ij} \in \mathbb{S}$ *and* $\mathbf{B}_2 = (B_{2,ij})_{p\times p}$ *with each*

$B_{2,ij} \in \mathbb{S}$. *Then*

$$\left\| \int\int \mathbf{B}_1(u,v)\{\mathbf{B}_2(u,v)\}^{\mathrm{T}} \mathrm{d}u\mathrm{d}v \right\|_2 \leqslant \sqrt{\|\mathbf{B}_1\|_{\mathcal{S},\infty}\|\mathbf{B}_1\|_{\mathcal{S},1}}\sqrt{\|\mathbf{B}_2\|_{\mathcal{S},\infty}\|\mathbf{B}_2\|_{\mathcal{S},1}}.$$

Proof. Notice that

$$\begin{aligned}
\left\| \int\int \mathbf{B}_1(u,v)\{\mathbf{B}_2(u,v)\}^{\mathrm{T}} \mathrm{d}u\mathrm{d}v \right\|_1 &= \max_{1\leqslant j\leqslant p} \sum_{i=1}^{p} \left| \int\int \sum_k B_{1,ik}(u,v) B_{2,jk}(u,v) \mathrm{d}u\mathrm{d}v \right| \\
&\leqslant \max_{1\leqslant j\leqslant p} \sum_{i=1}^{p}\sum_{k=1}^{p} \|B_{1,ik}\|_{\mathcal{S}} \|B_{2,jk}\|_{\mathcal{S}} \\
&\leqslant \max_{1\leqslant k\leqslant p} \sum_{i=1}^{p} \|B_{1,ik}\|_{\mathcal{S}} \max_{1\leqslant j\leqslant p} \sum_{k=1}^{p} \|B_{2,jk}\|_{\mathcal{S}} \\
&= \|\mathbf{B}_1\|_{\mathcal{S},1}\|\mathbf{B}_2\|_{\mathcal{S},\infty}.
\end{aligned} \tag{3.50}$$

By similar argument, we obtain that

$$\left\| \int\int \mathbf{B}_1(u,v)\{\mathbf{B}_2(u,v)\}^{\mathrm{T}} \mathrm{d}u\mathrm{d}v \right\|_\infty \leqslant \|\mathbf{B}_1\|_{\mathcal{S},\infty}\|\mathbf{B}_2\|_{\mathcal{S},1}. \tag{3.51}$$

Combining (3.50), (3.51) and the matrix norm inequality $\|\mathbf{E}\|^2 \leqslant \|\mathbf{E}\|_\infty \|\mathbf{E}\|_1$ for any matrix $\mathbf{E} \in \mathbb{R}^{p\times p}$, we complete the proof. □

Lemma 3.6. *Let* $\mathbf{B} = (B_{ij})_{p\times p}$ *with each* $B_{ij} \in \mathbb{S}$, $\mathbf{b}_1 \in \mathbb{R}^p$ *and* $\mathbf{b}_2 \in \mathbb{R}^p$. *Then*

$$\|\mathbf{b}_1^{\mathrm{T}}\mathbf{B}\mathbf{b}_2\|_{\mathcal{S}} \leqslant \|\mathbf{b}_1\|_2\|\mathbf{b}_2\|_2\sqrt{\|\mathbf{B}\|_{\mathcal{S},\infty}\|\mathbf{B}\|_{\mathcal{S},1}}.$$

Proof. By elementary calculation and Lemma 3.5, we obtain that

$$\begin{aligned}
&\|\mathbf{b}_1^{\mathrm{T}}\mathbf{B}\mathbf{b}_2\|_{\mathcal{S}}^2 = \int\int \mathbf{b}_1^{\mathrm{T}}\mathbf{B}(u,v)\mathbf{b}_2\mathbf{b}_2^{\mathrm{T}}\{\mathbf{B}(u,v)\}^{\mathrm{T}}\mathbf{b}_1 \mathrm{d}u\mathrm{d}v \\
\leqslant &\int\int \|\mathbf{b}_2\mathbf{b}_2^{\mathrm{T}}\|_2 \|\mathbf{b}_1^{\mathrm{T}}\mathbf{B}(u,v)\|_2^2 \mathrm{d}u\mathrm{d}v \leqslant \|\mathbf{b}_2\|_2^2 \int\int \mathbf{b}_1^{\mathrm{T}}\mathbf{B}(u,v)\{\mathbf{B}(u,v)\}^{\mathrm{T}}\mathbf{b}_1 \mathrm{d}u\mathrm{d}v \\
\leqslant &\|\mathbf{b}_1\|_2^2\|\mathbf{b}_2\|_2^2\|\mathbf{B}\|_{\mathcal{S},\infty}\|\mathbf{B}\|_{\mathcal{S},1},
\end{aligned}$$

which completes our proof. □

Lemma 3.7. *Suppose* $\mathbf{B}$ *and* $\mathbf{B}+\mathbf{E}$ *are* $m \times m$ *symmetric matrices and that* $\mathbf{Q} = (\mathbf{Q}_1, \mathbf{Q}_2)$, *where* $\mathbf{Q}_1$ *is an* $m \times l$ *matrix and* $\mathbf{Q}_2$ *is an* $m \times (m-l)$ *matrix, is an orthogonal matrix such that* $\mathcal{C}(\mathbf{Q}_1)$ *is an invariant subspace for* $\mathbf{B}$ *(that is,*

$\mathbf{B} \cdot \mathcal{C}(\mathbf{Q}_1) \subset \mathcal{C}(\mathbf{Q}_1)$). *Partition the matrices* $\mathbf{Q}^{\mathrm{T}}\mathbf{B}\mathbf{Q}$ *and* $\mathbf{Q}^{\mathrm{T}}\mathbf{E}\mathbf{Q}$ *as follows:*

$$\mathbf{Q}^{\mathrm{T}}\mathbf{B}\mathbf{Q} = \begin{pmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 \end{pmatrix} \quad \textit{and} \quad \mathbf{Q}^{\mathrm{T}}\mathbf{E}\mathbf{Q} = \begin{pmatrix} \mathbf{E}_{11} & \mathbf{E}_{21}^{\mathrm{T}} \\ \mathbf{E}_{21} & \mathbf{E}_{22} \end{pmatrix}.$$

If $\mathrm{sep}(\mathbf{D}_1, \mathbf{D}_2) = \min_{\mu_1 \in \lambda(\mathbf{D}_1), \mu_2 \in \lambda(\mathbf{D}_2)} |\mu_1 - \mu_2| > 0$, *where* $\lambda(\mathbf{M})$ *denotes the set of eigenvalues of the matrix* $\mathbf{M}$, *and* $\|\mathbf{E}\|_2 \leqslant \mathrm{sep}(\mathbf{D}_1, \mathbf{D}_2)/5$, *then there exists a matrix* $\mathbf{P} \in \mathbb{R}^{(m-l)\times l}$ *with* $\|\mathbf{P}\|_2 \leqslant 4\|\mathbf{E}_{21}\|_2/\mathrm{sep}(\mathbf{D}_1, \mathbf{D}_2)$ *such that the columns of* $\mathbf{Q}_1^\star = (\mathbf{Q}_1 + \mathbf{Q}_2\mathbf{P})(\mathbf{I} + \mathbf{P}^{\mathrm{T}}\mathbf{P})^{-1/2}$ *define an orthonormal basis for a subspace that is invariant for* $\mathbf{B} + \mathbf{E}$.

This is Theorem 8.1.10 of Golub and Van Loan (1996). From Lemma 3.7, we have

$$\begin{aligned} \|\mathbf{Q}_1^\star - \mathbf{Q}_1\|_2 &= \|\{\mathbf{Q}_1 + \mathbf{Q}_2\mathbf{P} - \mathbf{Q}_1(\mathbf{I} + \mathbf{P}^{\mathrm{T}}\mathbf{P})^{1/2}\}(\mathbf{I} + \mathbf{P}^{\mathrm{T}}\mathbf{P})^{-1/2}\|_2 \\ &\leqslant \|\mathbf{Q}_1\{\mathbf{I} - (\mathbf{I} + \mathbf{P}^{\mathrm{T}}\mathbf{P})^{1/2}\}\|_2 + \|\mathbf{Q}_2\mathbf{P}\|_2 \\ &\leqslant 2\|\mathbf{P}\|_2 \leqslant \frac{8}{\mathrm{sep}(\mathbf{D}_1, \mathbf{D}_2)}\|\mathbf{E}_{21}\|_2 \leqslant \frac{8}{\mathrm{sep}(\mathbf{D}_1, \mathbf{D}_2)}\|\mathbf{E}\|_2. \end{aligned}$$

3.B Additional real data results

Figures 3.1–3.3 display exemplified trajectories of smoothed curves from the UK annual temperature data, Japanese mortality data and energy consumption data, respectively.

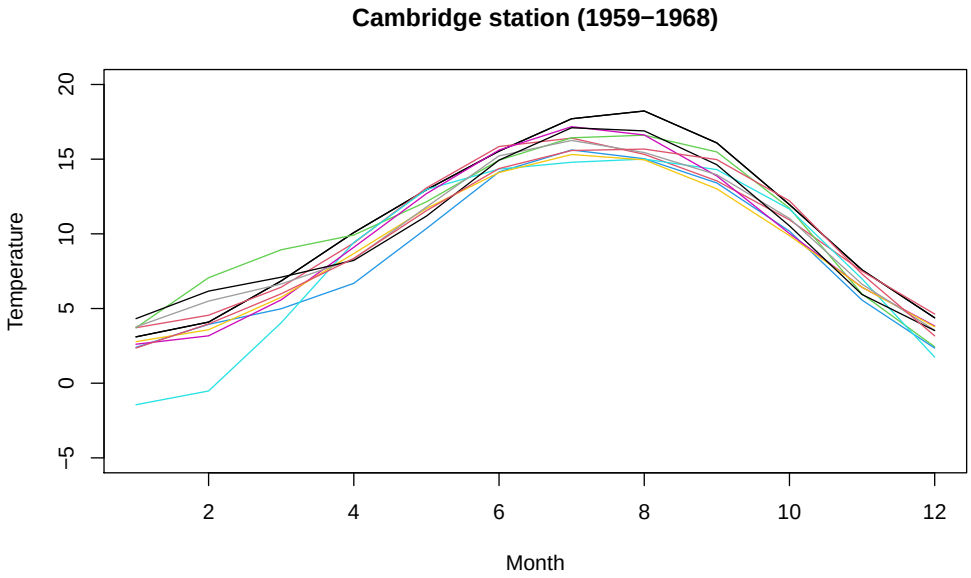


Figure 3.1: UK annual temperature data: the smoothed annual temperature data measured at the Cambridge station from 1959 to 1968.

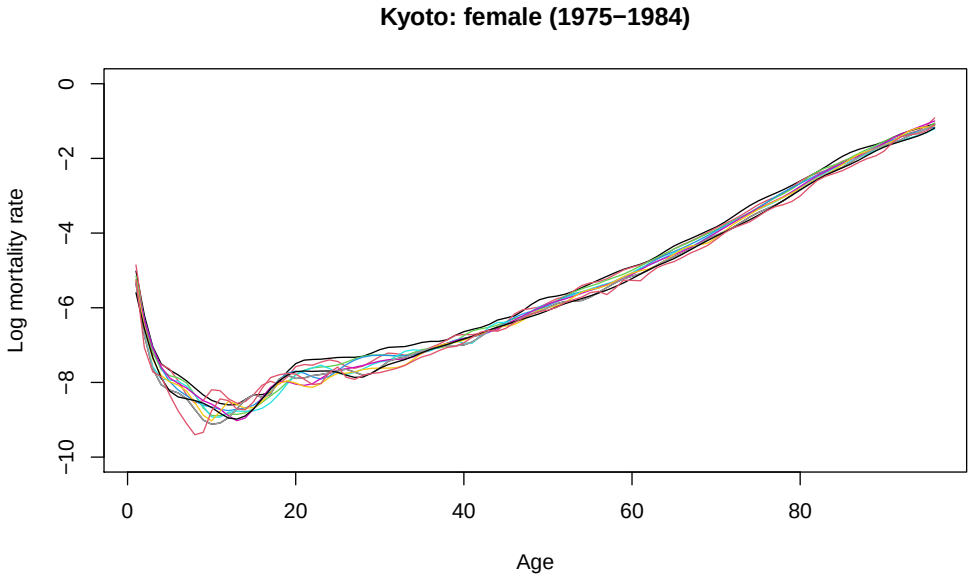


Figure 3.2: Japanese mortality data: Log smoothed female mortality rate in the Kyoto prefecture from 1975 to 1984.

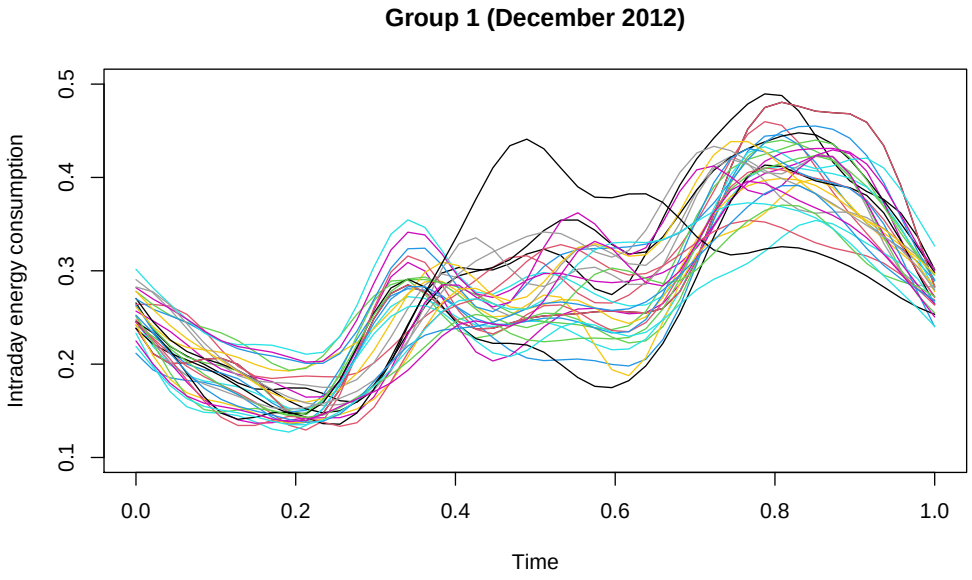


Figure 3.3: Energy consumption data: the smoothed intraday group-averaged consumption curves of Group 1 in December 2012.

Bibliography

Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, 81(3):1203–1227.

Anticevic, A., Cole, M. W., Murray, J. D., Corlett, P. R., Wang, X.-J., and Krystal, J. H. (2012). The role of default network deactivation in cognition and disease. *Trends in cognitive sciences*, 16(12):584–592.

Aue, A., Norinho, D. D., and Hörmann, S. (2015). On the prediction of stationary functional time series. *Journal of the American Statistical Association*, 110(509):378–392.

Aue, A., Rice, G., and Sonmez, O. (2018). Detecting and dating structural breaks in functional data without dimension reduction. *Journal of the Royal Statistical Society: Series B*, 80(3):509–529.

Avella-Medina, M., Battey, H. S., Fan, J., and Li, Q. (2018). Robust estimation of high-dimensional covariance and precision matrices. *Biometrika*, 105(2):271–284.

Back, A. D. and Weigend, A. S. (1997). A first application of independent component analysis to extracting structure from stock returns. *International Journal of Neural Systems*, 8(04):473–484.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bai, L., Wang, J., Ma, X., and Lu, H. (2018). Air pollution forecasts: An overview. *International Journal of Environmental Research and Public Health*, 15(4):780.

Basu, S. and Michailidis, G. (2015). Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics*, 43(4):1535–1567.

Bathia, N., Yao, Q., and Ziegelmann, F. (2010). Identifying the finite dimensionality of curve time series. *The Annals of Statistics*, 38(6):3352–3386.

Bickel, P. J. and Levina, E. (2008). Covariance regularization by thresholding. *The Annals of Statistics*, 36(6):2577–2604.

Biswal, B., Zerrin Yetkin, F., Haughton, V. M., and Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine*, 34(4):537–541.

Bosq, D. (2000). *Linear processes in function spaces*, volume 149 of *Lecture Notes in Statistics*. Springer-Verlag, New York. Theory and applications.

Cai, T. and Liu, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association*, 106(494):672–684.

Cattell, R. B. (1987). *Intelligence: Its Structure, Growth and Action*. Elsevier.

Chang, C. and Glover, G. H. (2010). Time–frequency dynamics of resting-state brain connectivity measured with fMRI. *Neuroimage*, 50(1):81–98.

Chang, J., Chen, C., Qiao, X., and Yao, Q. (2022). An autocovariance-based learning framework for high-dimensional functional time series. *Journal of Econometrics*, in press.

Chang, J., Guo, B., and Yao, Q. (2018). Principal component analysis for second-order stationary vector time series. *The Annals of Statistics*, 46(5):2094–2124.

Chen, C., Guo, S., and Qiao, X. (2022). Functional linear regression: dependence and error contamination. *Journal of Business & Economic Statistics*, 40(1):444–457.

Chen, Z. and Leng, C. (2016). Dynamic covariance models. *Journal of the American Statistical Association*, 111(515):1196–1207.

Chiou, J.-M., Chen, Y.-T., and Yang, Y.-F. (2014). Multivariate functional principal component analysis: a normalization approach. *Statistica Sinica*, 24(4):1571–1596.

Cho, H., Goude, Y., Brossat, X., and Yao, Q. (2013). Modeling and forecasting daily electricity load curves: a hybrid approach. *Journal of the American Statistical Association*, 108(501):7–21.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360.

Fan, J., Liao, Y., and Wang, W. (2016). Projected principal component analysis in factor models. *The Annals of Statistics*, 44(1):219–254.

Fan, Y., Foutz, N., James, G. M., and Jank, W. (2014). Functional response additive model estimation with online virtual stock markets. *The Annals of Applied Statistics*, 8(4):2435–2460.

Fan, Y., James, G. M., and Radchenko, P. (2015). Functional additive regression. *The Annals of Statistics*, 43(5):2296–2325.

Fan, Y. and Lv, J. (2016). Innovated scalable efficient estimation in ultra-large Gaussian graphical models. *The Annals of Statistics*, 44(5):2098–2126.

Fang, Q., Guo, S., and Qiao, X. (2022). Finite sample theory for high-dimensional functional/scalar time series with applications. *Electronic Journal of Statistics*, 16(1):527–591.

Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., Papademetris, X., and Constable, R. T. (2015). Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, 18(11):1664–1671.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2005). The generalized dynamic factor model: one-sided estimation and forecasting. *Journal of the American Statistical Association*, 100(471):830–840.

Gao, Y., Shang, H. L., and Yang, Y. (2019a). High-dimensional functional time series forecasting: an application to age-specific mortality rates. *Journal of Multivariate Analysis*, 170:232–243.

Gao, Z., Ma, Y., Wang, H., and Yao, Q. (2019b). Banded spatio-temporal autoregressions. *Journal of Econometrics*, 208(1):211–230.

Ghosh, S., Khare, K., and Michailidis, G. (2019). High-dimensional posterior consistency in Bayesian vector autoregressive models. *Journal of the American Statistical Association*, 114(526):735–748.

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R., et al. (2013). The minimal preprocessing pipelines for the human connectome project. *Neuroimage*, 80:105–124.

Golub, G. H. and Van Loan, C. F. (1996). *Matrix Computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, fourth edition.

Guo, S. and Qiao, X. (2022). On consistency and sparsity for high-dimensional functional time series with application to autoregressions. *Bernoulli*, in press.

Guo, S., Wang, Y., and Yao, Q. (2016). High-dimensional and banded vector autoregressions. *Biometrika*, 103(4):889–903.

Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics*, 35(1):70–91.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, Princeton, NJ.

Han, F., Lu, H., and Liu, H. (2015). A direct estimation of high dimensional stationary vector autoregressions. *Journal of Machine Learning Research*, 16:3115–3150.

Han, Y., Chen, R., Zhang, C.-H., and Yao, Q. (2021). Simultaneous decorrelation of matrix time series. *arXiv preprint arXiv:2103.09411*.

Han, Y. and Tsay, R. S. (2020). High-dimensional linear regression for dependent data with applications to nowcasting. *Statistica Sinica*, 30(4):1797–1827.

Happ, C. and Greven, S. (2018). Multivariate functional principal component analysis for data observed on different (dimensional) domains. *Journal of the American Statistical Association*, 113(522):649–659.

Hörmann, S., Kidziński, Ł., and Hallin, M. (2015a). Dynamic functional principal components. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 77(2):319–348.

Hörmann, S., Kidziński, Ł., and Kokoszka, P. (2015b). Estimation in functional lagged regression. *Journal of Time Series Analysis*, 36(4):541–561.

Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *The Annals of Statistics*, 38(3):1845–1884.

Horn, R. A. and Mathias, R. (1990). Cauchy-Schwarz inequalities associated with positive semidefinite matrices. *Linear Algebra and its Applications*, 142:63–82.

Horváth, L., Kokoszka, P., and Rice, G. (2014). Testing stationarity of functional time series. *Journal of Econometrics*, 179(1):66–82.

Hu, X. and Yao, F. (2021). Sparse functional principal component analysis in high dimensions. *arXiv:2011.00959v2*.

Jiao, S., Aue, A., and Ombao, H. (2021). Functional time series prediction under partial observation of the future curve. *Journal of the American Statistical Association*, in press.

Kong, D., Xue, K., Yao, F., and Zhang, H. H. (2016). Partially functional linear regression in high dimensions. *Biometrika*, 103(1):147–159.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer Series in Statistics. Springer, New York.

Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: inference for the number of factors. *The Annals of Statistics*, 40(2):694–726.

Lam, C., Yao, Q., and Bathia, N. (2011). Estimation of latent factors for high-dimensional time series. *Biometrika*, 98(4):901–918.

Lee, K.-Y., Ji, D., Li, L., Constable, T., and Zhao, H. (2021). Conditional functional graphical models. *Journal of the American Statistical Association*, in press.

Li, B. and Solea, E. (2018). A nonparametric graphical model for functional data with application to brain networks based on fMRI. *Journal of the American Statistical Association*, 113(524):1637–1655.

Li, D., Robinson, P. M., and Shang, H. L. (2020). Long-range dependent curve time series. *Journal of the American Statistical Association*, 115(530):957–971.

Li, Z., Lam, C., Yao, J., and Yao, Q. (2019). On testing for high-dimensional white noise. *The Annals of Statistics*, 47(6):3382–3412.

Lin, J. and Michailidis, G. (2017). Regularized estimation and testing for high-dimensional multi-block vector-autoregressive models. *Journal of Machine Learning Research*, 18:Paper No. 117, 49.

Lin, J. and Michailidis, G. (2020). Regularized estimation of high-dimensional factor-augmented vector autoregressive (FAVAR) models. *Journal of Machine Learning Research*, 21:Paper No. 117, 51.

Loh, P.-L. and Wainwright, M. J. (2012). High-dimensional regression with noisy and missing data: provable guarantees with nonconvexity. *The Annals of Statistics*, 40(3):1637–1664.

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., and Yger, F. (2018). A review of classification algorithms for eeg-based brain–computer interfaces: a 10 year update. *Journal of Neural Engineering*, 15(3):031005.

Luo, R. and Qi, X. (2017). Function-on-function linear regression by signal compression. *Journal of the American Statistical Association*, 112(518):690–705.

Matteson, D. S. and Tsay, R. S. (2011). Dynamic orthogonal components for multivariate time series. *Journal of the American Statistical Association*, 106(496):1450–1463.

Miao, R., Zhang, X., and Wong, R. K. (2022). A wavelet-based independence test for functional data with an application to meg functional connectivity. *Journal of the American Statistical Association*, in press.

Pan, J. and Yao, Q. (2008). Modelling multiple time series via common factors. *Biometrika*, 95(2):365–379.

Panaretos, V. M. and Tavakoli, S. (2013). Fourier analysis of stationary time series in function space. *The Annals of Statistics*, 41(2):568–603.

Park, J., Ahn, J., and Jeon, Y. (2022). Sparse functional linear discriminant analysis. *Biometrika*, 109(1):209–226.

Pena, D. and Box, G. E. (1987). Identifying a simplifying structure in time series. *Journal of the American statistical Association*, 82(399):836–843.

Pham, T. and Panaretos, V. M. (2018). Methodology and convergence rates for functional time series regression. *Statistica Sinica*, 28(4):2521–2539.

Qiao, X., Guo, S., and James, G. M. (2019). Functional graphical models. *Journal of the American Statistical Association*, 114(525):211–222.

Qiao, X., Qian, C., James, G. M., and Guo, S. (2020). Doubly functional graphical models in high dimensions. *Biometrika*, 107(2):415–431.

Rogers, B. P., Morgan, V. L., Newton, A. T., and Gore, J. C. (2007). Assessing functional connectivity in the human brain by fMRI. *Magnetic Resonance Imaging*, 25(10):1347–1357.

Rothman, A. J., Levina, E., and Zhu, J. (2009). Generalized thresholding of large covariance matrices. *Journal of the American Statistical Association*, 104(485):177–186.

Rudelson, M. and Vershynin, R. (2013). Hanson-Wright inequality and sub-Gaussian concentration. *Electronic Communications in Probability*, 18:1–9.

Shu, H. and Nan, B. (2019). Estimation of large covariance and precision matrices from temporally dependent observations. *The Annals of Statistics*, 47(3):1321–1350.

Simon, N. and Tibshirani, R. (2012). Standardization and the group Lasso penalty. *Statistica Sinica*, 22(3):983–1001.

Solea, E. and Li, B. (2022). Copula Gaussian graphical models for functional data. *Journal of the American Statistical Association*, 117(538):781–793.

Stock, J. H. and Watson, M. W. (2012). Generalized shrinkage methods for forecasting using many predictors. *Journal of Business & Economic Statistics*, 30(4):481–493.

Storey, J. D., Xiao, W., Leek, J. T., Tompkins, R. G., and Davis, R. W. (2005). Significance analysis of time course microarray experiments. *Proceedings of the National Academy of Sciences*, 102(36):12837–12842.

Sun, Y., Li, Y., Kuceyeski, A., and Basu, S. (2018). Large spectral density matrix estimation by thresholding. *arXiv:1812.00532*.

Tiao, G. C. and Tsay, R. S. (1989). Model specification in multivariate time series. *Journal of the Royal Statistical Society: Series B (Methodological)*, 51(2):157–195.

Van Den Heuvel, M. P., Stam, C. J., Kahn, R. S., and Pol, H. E. H. (2009). Efficiency of functional brain networks and intellectual performance. *Journal of Neuroscience*, 29(23):7619–7624.

Wang, H., Peng, B., Li, D., and Leng, C. (2021). Nonparametric estimation of large covariance matrices with conditional sparsity. *Journal of Econometrics*, 223(1):53–72.

Wilms, I., Basu, S., Bien, J., and Matteson, D. S. (2021). Sparse identification and estimation of large-scale vector autoregressive moving averages. *Journal of the American Statistical Association*, in press.

Wong, K. C., Li, Z., and Tewari, A. (2020). Lasso guarantees for β-mixing heavy-tailed time series. *The Annals of Statistics*, 48(2):1124–1142.

Wu, W.-B. and Wu, Y. N. (2016). Performance bounds for parameter estimates of high-dimensional linear models with correlated errors. *Electronic Journal of Statistics*, 10(1):352–379.

Xue, K. and Yao, F. (2021). Hypothesis testing in large-scale functional linear regression. *Statistica Sinica*, 31(2):1101–1123.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B*, 68(1):49–67.

Zapata, J., Oh, S.-Y., and Petersen, A. (2022). Partial separability and functional graphical models for multivariate Gaussian processes. *Biometrika*, 109(3):665–681.

Zemyan, S. M. (2012). *The Classical Theory of Integral Equations: A Concise Treatment*. Springer Science & Business Media.

Zhang, J.-T. (2013). *Analysis of Variance for Functional Data*. CRC Press.

Zhou, H. H. and Raskutti, G. (2019). Non-parametric sparse additive auto-regressive network models. *Institute of Electrical and Electronics Engineers. Transactions on Information Theory*, 65(3):1473–1492.

Zhu, H., Strawn, N., and Dunson, D. B. (2016). Bayesian graphical models for multivariate functional data. *Journal of Machine Learning Research*, 17:Paper No. 204, 27.

Ziomas, I. C., Melas, D., Zerefos, C. S., Bais, A. F., and Paliatsos, A. G. (1995). Forecasting peak pollutant levels from meteorological variables. *Atmospheric Environment*, 29(24):3703–3711.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429.