Electronic Journal of Statistics
Vol. 16 (2022) 527–591
ISSN: 1935-7524
https://doi.org/10.1214/21-EJS1960

Finite sample theory for high-dimensional functional/scalar time series with applications*

Qin Fang[1] **Shaojun Guo**[2] **and Xinghao Qiao**[1]

[1] *Department of Statistics, London School of Economics, London, WC2A 2AE, U.K.*
e-mail: q.fang3@lse.ac.uk; x.qiao@lse.ac.uk

[2] *Institute of Statistics and Big Data, Renmin University of China, Beijing, 100872, P.R. China. e-mail:* sjguo@ruc.edu.cn

Abstract: Statistical analysis of high-dimensional functional times series arises in various applications. Under this scenario, in addition to the intrinsic infinite-dimensionality of functional data, the number of functional variables can grow with the number of serially dependent observations. In this paper, 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.

MSC2020 subject classifications: Primary 62M10, 62R10; secondary 62J07.
Keywords and phrases: Cross-spectral stability measure, functional linear regression, functional principal component analysis, non-asymptotics, sub-Gaussian functional linear process, sparsity.

Received August 2020.

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

*Shaojun Guo was partially supported by the National Natural Science Foundation of China (No. 11771447).

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

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

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, Norinho and Hörmann, 2015) over a large number of sites and cumulative intraday return curves (Horváth, Kokoszka and Rice, 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, 2021) and, under a special independent setting, functional graphical models (Qiao, Guo and James, 2019) and functional additive regressions (Fan et al., 2014; Fan, James and Radchenko, 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, Li and Tewari, 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 (2021) 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, Kidziński and Kokoszka, 2015) in the additive form:

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

where p-dimensional functional covariates $\{\mathbf{X}_t(\cdot)\}$ and functional errors $\{\epsilon_t(\cdot)\}$ are generated from independent, centered, stationary functional processes, and $\{\beta_{hj}(\cdot,\cdot) : h = 0,\dots,L, j = 1,\dots,p\}$ 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 $\mathbf{Z}_t = (Z_{t1},\dots,Z_{td})^{\mathrm{T}}$ for $t = 1,\dots,n$, both of which are independent of errors $\{\epsilon_t\}$, as follows:

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

where $\{\beta_j(\cdot) : j = 1\dots,p\}$ are sparse functional coefficients and $\{\gamma_k : k = 1,\dots,d\}$ 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, James and Radchenko (2015); Xue and Yao (2021), respectively.

In addition to existing non-asymptotic results in Guo and Qiao (2021), 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 $h = 0,\dots,L$ in model (1.1) and estimates of $\mathrm{Cov}(\mathbf{X}_t(\cdot),\mathbf{Z}_t)$, $\mathrm{Cov}(\mathbf{Z}_t,\epsilon_t)$ and $\mathrm{Cov}(\mathbf{X}_t(\cdot),\epsilon_t)$ 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 our paper 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. $\{\mathbf{Y}_t(\cdot) = (Y_{t1}(\cdot),\dots,Y_{td}(\cdot))^{\mathrm{T}}\}$, 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, Chen, Qiao and Yao, 2021) and the functional factor model (Guo and Qiao, 2021).
- 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, Yao and Ziegelmann (2010); Hörmann and Kokoszka (2010); Panaretos and Tavakoli (2013); Aue, Norinho and Hörmann (2015); Hörmann, Kidziński and Kokoszka (2015); Pham and Panaretos (2018); Li, Robinson and Shang (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, Wang and Yao, 2016; Lin and Michailidis, 2017; Gao et al., 2019; Ghosh, Khare and Michailidis, 2019; Zhou and Raskutti, 2019; Wong, Li and Tewari, 2020; Lin and

Michailidis, 2020).

Outline. The remainder of the paper is organized as follows. In Section 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 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 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 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 $\mathbf{K}(\boldsymbol{f})(u) = \big(\sum_{j=1}^p \langle K_{1j}(u, \cdot), f_j(\cdot)\rangle, \ldots, \sum_{j=1}^p \langle K_{qj}(u, \cdot), f_j(\cdot)\rangle\big)^{\mathrm{T}} \in \mathbb{H}^q$, 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 $\|K\|_{\mathcal{S}} = \big(\int\int K(u, v)^2 dudv\big)^{1/2}$. For general $\mathbf{K}$, we define functional versions of Frobenius, elementwise ℓ_∞, matrix ℓ_1 and matrix ℓ_∞ norms by $\|\mathbf{K}\|_{\mathrm{F}} = \big(\sum_{i,j} \|K_{ij}\|_{\mathcal{S}}^2\big)^{1/2}$, $\|\mathbf{K}\|_{\max} = \max_{i,j} \|K_{ij}\|_{\mathcal{S}}$, $\|\mathbf{K}\|_1 = \max_j \sum_i \|K_{ij}\|_{\mathcal{S}}$ and $\|\mathbf{K}\|_\infty = \max_i \sum_j \|K_{ij}\|_{\mathcal{S}}$, respectively.

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.

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,

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

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

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

According to Guo and Qiao (2021), the functional stability measure of $\{\mathbf{X}_t(\cdot)\}$ is defined based on the functional Rayleigh quotients of $\boldsymbol{f}_\theta^X$ relative to $\mathbf{\Sigma}_0^X$,

$$\mathcal{M}^X = 2\pi \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi],\mathbf{\Phi}\in\mathbb{H}_0^p} \frac{\langle \mathbf{\Phi}, \boldsymbol{f}_\theta^X(\mathbf{\Phi})\rangle}{\langle \mathbf{\Phi}, \mathbf{\Sigma}_0^X(\mathbf{\Phi})\rangle}, \tag{2.1}$$

where $\mathbb{H}_0^p = \{\mathbf{\Phi} \in \mathbb{H}^p : \langle \mathbf{\Phi}, \mathbf{\Sigma}_0^X(\mathbf{\Phi})\rangle \in (0,\infty)\}$. To handle truly infinite-dimensional objects $\{\mathbf{X}_t(\cdot)\}$ with infinite, summable and decaying eigenvalues of $\mathbf{\Sigma}_0^X$, such stability measure $\mathcal{M}^X$ can more precisely capture the effect of small eigenvalues of $\mathbf{\Sigma}_0^X$ on the numerator in (2.1).

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. (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 $|J| \leqslant k$, the functional stability measure of $\{(X_{tj}(\cdot)) : j \in J\}_{t\in\mathbb{Z}}$ is defined by

$$\mathcal{M}_k^X = 2\pi \cdot \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi],\|\mathbf{\Phi}\|_0\leqslant k,\mathbf{\Phi}\in\mathbb{H}_0^p} \frac{\langle \mathbf{\Phi}, \boldsymbol{f}_\theta^X(\mathbf{\Phi})\rangle}{\langle \mathbf{\Phi}, \mathbf{\Sigma}_0^X(\mathbf{\Phi})\rangle},\ k = 1,\dots,p. \tag{2.2}$$

Under Condition 1, we have $\mathcal{M}_k^X \leqslant \mathcal{M}^X < \infty$.

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

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

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 2.4.

Definition 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

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

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

Condition 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 (2.3) is finite, i.e.

$$\mathcal{M}^{X,Y} = 2\pi \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi],\mathbf{\Phi}_1\in\mathbb{H}_0^p,\mathbf{\Phi}_2\in\mathbb{H}_0^d} \frac{\left|\langle\mathbf{\Phi}_1, \boldsymbol{f}_\theta^{X,Y}(\mathbf{\Phi}_2)\rangle\right|}{\sqrt{\langle\mathbf{\Phi}_1,\mathbf{\Sigma}_0^X(\mathbf{\Phi}_1)\rangle}\sqrt{\langle\mathbf{\Phi}_2,\mathbf{\Sigma}_0^Y(\mathbf{\Phi}_2)\rangle}} < \infty, \quad (2.3)$$

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

Remark 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 (2.1).*

(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.

$$\widetilde{\mathcal{M}}^{X,Y} = \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi],\boldsymbol{\nu}\in\widetilde{\mathbb{R}}_0^d} \sqrt{\frac{\boldsymbol{\nu}^{\mathrm{T}}\{\boldsymbol{f}_\theta^{X,Y}\}^*\boldsymbol{f}_\theta^{X,Y}\boldsymbol{\nu}}{\boldsymbol{\nu}^{\mathrm{T}}\boldsymbol{\nu}}} < \infty, \quad (2.4)$$

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 (2.3) is equivalent to*

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

whose upper bound is $\widetilde{\mathcal{M}}^{X,Y}$ *as justified in Lemma 1 in Appendix B.3. This demonstrates that, compared with (2.4), 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 2.4 and 2.5 for details.*

(d) *We can generalize (2.3) 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 similarly define* $\boldsymbol{f}_\theta^{X,Z} = \frac{1}{2\pi}\sum_{h\in\mathbb{Z}}\boldsymbol{\Sigma}_h^{X,Z}\exp(-ih\theta)$ *with* $\boldsymbol{\Sigma}_h^{X,Z}(\cdot) = Cov(\mathbf{X}_t(\cdot),\mathbf{Z}_{t+h})$. *According to (2.3), the mixed cross-spectral stability measure of* $\{\mathbf{X}_t(\cdot)\}$ *and* $\{\mathbf{Z}_t\}$ *can be defined by*

$$\mathcal{M}^{X,Z} = 2\pi \underset{\theta\in[-\pi,\pi],\boldsymbol{\Phi}\in\mathbb{H}_0^p,\boldsymbol{\nu}\in\mathbb{R}_0^d}{\text{ess sup}} \frac{\left|\langle\boldsymbol{\Phi},\boldsymbol{f}_\theta^{X,Z}\boldsymbol{\nu}\rangle\right|}{\sqrt{\langle\boldsymbol{\Phi},\boldsymbol{\Sigma}_0^X(\boldsymbol{\Phi})\rangle}\sqrt{\boldsymbol{\nu}^{\mathrm{T}}\boldsymbol{\Sigma}_0^Z\boldsymbol{\nu}}} \tag{2.5}$$

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

$$\mathcal{M}^{Z} = 2\pi\cdot \underset{\theta\in[-\pi,\pi],\boldsymbol{\nu}\in\mathbb{R}_0^d}{\text{ess sup}} \frac{\boldsymbol{\nu}^{\mathrm{T}}\boldsymbol{f}_\theta^{Z}\boldsymbol{\nu}}{\boldsymbol{\nu}^{\mathrm{T}}\boldsymbol{\Sigma}_0^Z\boldsymbol{\nu}}, \tag{2.6}$$

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 (2.5) and (2.6) play an essential role in the convergence analysis of the regularized estimates for model (1.2). See Section 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 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

$$\mathcal{M}_{k_1,k_2}^{X,Y} = 2\pi \underset{\substack{\theta\in[-\pi,\pi],\boldsymbol{\Phi}_1\in\mathbb{H}_0^p,\boldsymbol{\Phi}_2\in\mathbb{H}_0^d \\ \|\boldsymbol{\Phi}_1\|_0\leqslant k_1,\|\boldsymbol{\Phi}_2\|_0\leqslant k_2}}{\text{ess sup}} \frac{\left|\langle \boldsymbol{\Phi}_1, \boldsymbol{f}_\theta^{X,Y}(\boldsymbol{\Phi}_2)\rangle\right|}{\sqrt{\langle \boldsymbol{\Phi}_1, \boldsymbol{\Sigma}_0^X(\boldsymbol{\Phi}_1)\rangle}\sqrt{\langle \boldsymbol{\Phi}_2, \boldsymbol{\Sigma}_0^Y(\boldsymbol{\Phi}_2)\rangle}}. \tag{2.7}$$

Under Condition 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 (2.2), (2.5), (2.6) and (2.7), 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.

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 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 $x \in \mathbb{H}$,

$$\mathbb{E}\{e^{\langle x,X\rangle}\} \leqslant e^{\alpha^2\langle x,\Sigma_0(x)\rangle/2}. \tag{2.8}$$

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 2.4 and 2.5. We focus on multivariate functional linear processes with sub-Gaussian errors, namely sub-Gaussian functional linear processes:

$$\mathbf{X}_t(\cdot) = \sum_{l=0}^{\infty} \mathbf{A}_l(\boldsymbol{\varepsilon}_{t-l}), \quad t \in \mathbb{Z}, \tag{2.9}$$

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 3. It is worth noting that (2.9) 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 (2.9), we can derive the spectral density matrix function as

$$\boldsymbol{f}_\theta^X(u,v) = \frac{1}{2\pi} \int\int \mathcal{B}\left(e^{-i\theta}\right)(u,u')\boldsymbol{\Sigma}_0^\varepsilon(u',v')\mathcal{B}\left(e^{-i\theta}\right)^*(v,v')du'dv' \tag{2.10}$$

and the covariance matrix function as

$$\boldsymbol{\Sigma}_0^X(u,v) = \sum_{l=0}^{\infty} \int\int \mathbf{A}_l(u,u')\boldsymbol{\Sigma}_0^\varepsilon(u',v')\mathbf{A}_l^*(v,v')du'dv'. \tag{2.11}$$

Then we can express the functional stability measure $\mathcal{M}^X$ in (2.1) based on (2.10) and (2.11). The cross-spectral stability measure $\mathcal{M}^{X,Y}$ in (2.3) or $\mathcal{M}^{X,Z}$ in (2.5) can be expressed in a similar fashion.

Condition 3. The coefficient functions satisfy $\sum_{l=0}^{\infty} \|\mathbf{A}_l\|_\infty = O(1)$.

Condition 4. (i) 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 $j \in \{1,\ldots,p\}$; (ii) $\omega_0^\varepsilon = \max_j \int_{\mathcal{U}} \Sigma_{0,jj}^\varepsilon(u,u)du = O(1)$.

Condition 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 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 2 in Appendix B.3 for details. In general, we can relax Conditions 3 and 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.

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

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

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

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

Theorem 1. *Suppose that Conditions 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* $\eta > 0$*,*

$$P\left\{\left|\frac{\langle \mathbf{\Phi}_1, (\widehat{\mathbf{\Sigma}}_0^X - \mathbf{\Sigma}_0^X)(\mathbf{\Phi}_1)\rangle}{\langle \mathbf{\Phi}_1, \mathbf{\Sigma}_0^X(\mathbf{\Phi}_1)\rangle}\right| > \mathcal{M}_{k_1}^X \eta\right\} \leqslant 2\exp\left\{-cn\min\left(\eta^2, \eta\right)\right\}, \tag{2.12}$$

and

$$P\left\{\left|\frac{\langle \mathbf{\Phi}_1, (\widehat{\mathbf{\Sigma}}_h^{X,Y} - \mathbf{\Sigma}_h^{X,Y})(\mathbf{\Phi}_2)\rangle}{\langle \mathbf{\Phi}_1, \mathbf{\Sigma}_0^X(\mathbf{\Phi}_1)\rangle + \langle \mathbf{\Phi}_2, \mathbf{\Sigma}_0^Y(\mathbf{\Phi}_2)\rangle}\right| > \left(\mathcal{M}_{k_1}^X + \mathcal{M}_{k_2}^Y + \mathcal{M}_{k_1,k_2}^{X,Y}\right)\eta\right\} \tag{2.13}$$
$$\leqslant c_1\exp\{-c_2 n\min(\eta^2, \eta)\}.$$

Remark 2. *(2.12) extends the concentration inequality for normalized quadratic form of* $\widehat{\mathbf{\Sigma}}_0^X$ *in Theorem 1 of Guo and Qiao (2021) 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 (2021), so we present some results used in our subsequent analysis in Appendix E. The concentration inequality in (2.13) 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. It is also crucial in deriving subsequent concentration results, e.g. with suitable choices of* $\mathbf{\Phi}_1$ *and* $\mathbf{\Phi}_2$*, it facilitates the elementwise concentration bounds on* $\widehat{\mathbf{\Sigma}}_h^{X,Y}$ *in the following theorem.*

Theorem 2. *Suppose that conditions in Theorem 1 hold. Then there exists some constants* $c_1, c_3 > 0$ *such that for any* $\eta > 0$ *and each* $j = 1, \ldots, p$*,* $k = 1, \ldots, d$*,*

$$P\left\{\|\widehat{\Sigma}_{h,jk}^{X,Y} - \Sigma_{h,jk}^{X,Y}\|_{\mathcal{S}} > (\omega_0^X + \omega_0^Y)\mathcal{M}_{X,Y}\eta\right\} \leqslant c_1\exp\left\{-c_3 n\min(\eta^2, \eta)\right\}, \tag{2.14}$$

where $\mathcal{M}_{X,Y} = \mathcal{M}_1^X + \mathcal{M}_1^Y + \mathcal{M}_{1,1}^{X,Y}, \omega_0^X = \max_j \int_{\mathcal{U}} \Sigma_{0,jj}^X(u,u)du$ *and* $\omega_0^Y = \max_k \int_{\mathcal{U}} \Sigma_{0,kk}^Y(u,u)du$*. 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*

$$\|\widehat{\mathbf{\Sigma}}_h^{X,Y} - \mathbf{\Sigma}_h^{X,Y}\|_{\max} \lesssim \mathcal{M}_{X,Y}\sqrt{\frac{\log(pd)}{n}}. \tag{2.15}$$

Remark 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 (2.15). 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 (2.15) as*

$$\max_{1\leqslant j\leqslant p, 1\leqslant k\leqslant d} \|\widehat{\Sigma}_{h,jk}^{X,Z} - \Sigma_{h,jk}^{X,Z}\| \lesssim \mathcal{M}_{X,Z}\sqrt{\frac{\log(pd)}{n}}, \tag{2.16}$$

where $\mathcal{M}_{X,Z} = \mathcal{M}_1^X + \mathcal{M}_1^Z + \mathcal{M}_{1,1}^{X,Z}$. (2.16) can be justified in the proof of Proposition 1 in Appendix B.2.

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 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 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, 2021), so we only present them in Lemmas 25 and 27 in Appendix 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 $\sigma_{h,jklm}^{X,Y} = \text{Cov}(\zeta_{tjl}, \xi_{(t+h)km})$.

Theorem 3. *Suppose that Conditions 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 $\{\widehat{\sigma}_{h,jklm}^{X,Y}\}$ satisfy*

$$\max_{\substack{1\leqslant j\leqslant p, 1\leqslant k\leqslant d\\ 1\leqslant l\leqslant M_1, 1\leqslant m\leqslant M_2}} \frac{\left|\widehat{\sigma}_{h,jklm}^{X,Y} - \sigma_{h,jklm}^{X,Y}\right|}{(l^{\alpha_1+1} \vee m^{\alpha_2+1})\sqrt{\omega_{jl}^X\omega_{km}^Y}} \lesssim \mathcal{M}_{X,Y}\sqrt{\frac{\log(pdM_1M_2)}{n}}. \tag{2.17}$$

In the special case that $\{\mathbf{X}_t(\cdot)\}$ and $\{\mathbf{Y}_t(\cdot)\}$ are identical, (2.17) degenerates to the deviation bound on $\widehat{\sigma}_{h,jklm}^{X}$ under the Gaussianity assumption (Guo and Qiao, 2021). 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 $\widehat{\varrho}_{h,jkl}^{X,Z} = (n-h)^{-1}\sum_{t=1}^{n-h}\widehat{\zeta}_{tjl}Z_{(t+h)k}$ and $\varrho_{h,jkl}^{X,Z} = \text{Cov}(\zeta_{tjl}, Z_{(t+h)k})$. We are ready to extend (2.17) to the following mixed-process scenario.

Proposition 1. *Suppose that Conditions 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 $n \gtrsim \log(pdM_1)M_1^{3\alpha_1+2}\mathcal{M}_{X,Z}^2$, 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 $\{\widehat{\varrho}_{h,jkl}^{X,Z}\}$ satisfy*

$$\max_{\substack{1\leqslant j\leqslant p, 1\leqslant k\leqslant d\\ 1\leqslant l\leqslant M_1}} \frac{\left|\widehat{\varrho}_{h,jkl}^{X,Z} - \varrho_{h,jkl}^{X,Z}\right|}{l^{\alpha_1+1}\sqrt{\omega_{jl}^X}} \lesssim \mathcal{M}_{X,Z}\sqrt{\frac{\log(pdM_1)}{n}}. \tag{2.18}$$

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 $\Sigma_{h,j}^{X,\epsilon}(u,v) = \text{Cov}\{X_{tj}(u), \epsilon_{t+h}(v)\}$ and its estimate $\widehat{\Sigma}_{h,j}^{X,\epsilon}(u,v) = (n-h)^{-1}\sum_{t=1}^{n-h}X_{tj}(u)\epsilon_{t+h}(v)$. To provide theoretical analysis of the estimates for model (1.1), the FPCA-based representation in Appendix F suggests to investigate the consistency properties of the estimated cross terms, i.e. $\widehat{\sigma}_{h,jlm}^{X,\epsilon} = \langle\widehat{\psi}_{jl}, \langle\widehat{\Sigma}_{h,j}^{X,\epsilon}, \widehat{\phi}_m\rangle\rangle$ or $\widehat{\sigma}_{h,jlm}^{X,Y} = (n-h)^{-1}\sum_{t=1}^{n-h}\widehat{\zeta}_{tjl}\widehat{\xi}_{(t+h)m} = \langle\widehat{\psi}_{jl}, \langle\widehat{\Sigma}_{h,j}^{X,Y}, \widehat{\phi}_m\rangle\rangle$. As $\{\mathbf{X}_{t-h}(\cdot) : h = 0, \dots, L\}$ 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 2. *Suppose that Conditions 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* $n \gtrsim \log(pM_1M_2)(M_1^{4\alpha_1+4} \vee M_2^{4\alpha_2+4})(\mathcal{M}_1^X + \mathcal{M}^Y)^2$, *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*

$$\max_{\substack{1\leqslant j\leqslant p\\ 1\leqslant l\leqslant M_1, 1\leqslant m\leqslant M_2}} \frac{\left|\widehat{\sigma}^{X,\epsilon}_{h,jlm}\right|}{(l^{\alpha_1} \vee m^{\alpha_2})\sqrt{\omega^X_{jl}\omega^Y_m}} \lesssim (\mathcal{M}_1^X + \mathcal{M}^\epsilon)\sqrt{\frac{\log(pM_1M_2)}{n}}. \tag{2.19}$$

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 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 3. *Suppose that Conditions 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* $n \gtrsim \log(pM_1)M_1^{3\alpha_1+2}(\mathcal{M}_1^X)^2$, *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*

$$\max_{1\leqslant j\leqslant p, 1\leqslant l\leqslant M_1} \frac{\left|\widehat{\varrho}^{X,\epsilon}_{h,jl}\right|}{\sqrt{\omega^X_{jl}}} \lesssim (\mathcal{M}_1^X + \mathcal{M}^\epsilon)\sqrt{\frac{\log(pM_1)}{n}}. \tag{2.20}$$

Remark 4. *Benefiting from the independence assumption between* $\{\mathbf{X}_t(\cdot)\}$ *and* $\{\epsilon_t(\cdot)\}$, *Proposition 2 leads to a faster rate of convergence in (2.19) compared with (2.17) with* $d = 1$. *Proposition 2 also plays a crucial rule in the proof of Proposition 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 3 results in a faster rate in (2.20) than that in (2.18) with* $d = 1$. *In the proof of Proposition 8, we will apply Proposition 3 to verify that, with high probability, model (1.2) satisfies the corresponding deviation condition.*

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 2.5 to investigate the convergence properties of the estimates under high-dimensional scaling.

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 F, we can represent model (1.1) in the following matrix form

$$\mathbf{U} = \sum_{h=0}^{L}\sum_{j=1}^{p} \mathbf{V}_{hj}\boldsymbol{\Psi}_{hj} + \mathbf{R} + \mathbf{E}, \tag{3.1}$$

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 $\{\boldsymbol{\Psi}_{hj} : h = 0, \ldots, L, j = 1, \ldots, p\}$:

$$\frac{1}{2}\|\widehat{\mathbf{U}} - \sum_{h=0}^{L}\sum_{j=1}^{p} \widehat{\mathbf{V}}_{hj}\boldsymbol{\Psi}_{hj}\|_{\mathrm{F}}^2 + \lambda_n \sum_{h=0}^{L}\sum_{j=1}^{p} \|\widehat{\mathbf{V}}_{hj}\boldsymbol{\Psi}_{hj}\|_{\mathrm{F}}, \tag{3.2}$$

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 (3.2).

Finally, we estimate functional coefficients by

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

3.2. Theoretical properties

We begin with some notation that will be used in this section. For a block matrix $\mathbf{B} = (\mathbf{B}_{jk})_{1\leqslant j\leqslant p_1, 1\leqslant k\leqslant p_2} \in \mathbb{R}^{p_1q_1\times p_2q_2}$ 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 (3.2) over $\{\mathbf{\Psi}_{hj}\}$ is equivalent to the following optimization task:

$$\widehat{\mathbf{B}} = \underset{\mathbf{B}\in\mathbb{R}^{(L+1)pq_1\times q_2}}{\arg\min}\left\{\frac{1}{2(n-L)}\|\widehat{\mathbf{U}} - \widehat{\mathbf{Z}}\widehat{\mathbf{D}}^{-1}\mathbf{B}\|_{\mathrm{F}}^2 + \lambda_n\|\mathbf{B}\|_1^{(q_1,q_2)}\right\}. \tag{3.3}$$

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 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 $\{\psi_{jl}(u)\}_{l\geqslant1}$ and $\{\phi_m(v)\}_{m\geqslant1}$, 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 $\{a_{hjlm}\}_{l,m\geqslant1}$ 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 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 $\widetilde{\mathbf{D}}_0^X = \mathbf{I}_{L+1} \otimes \text{diag}(\Sigma_{0,11}^X, \ldots, \Sigma_{0,pp}^X)$. The infimum $\underline{\mu}$ of the functional Rayleigh quotient of $\widetilde{\boldsymbol{\Sigma}}^X$ relative to $\widetilde{\mathbf{D}}_0^X$ is bounded below by zero, i.e.

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

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

Condition 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 (2021).

Before presenting the consistency analysis of $\widehat{\boldsymbol{\beta}}$ in Theorem 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 4 below and Propositions 6–7 in Appendix A, respectively.

Proposition 4. *Suppose Conditions 1–5 and 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 $\widehat{\boldsymbol{\Gamma}} = (n-L)^{-1}\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}\widehat{\mathbf{Z}}\widehat{\mathbf{D}}^{-1} \in \mathbb{R}^{(L+1)pq_1 \times (L+1)pq_1}$ satisfies, with probability greater than $1 - c_1^*(pq_1)^{-c_2^*}$,*

$$\boldsymbol{\theta}^{\mathrm{T}}\widehat{\boldsymbol{\Gamma}}\boldsymbol{\theta} \geqslant \tau_2\|\boldsymbol{\theta}\|^2 - \tau_1\|\boldsymbol{\theta}\|_1^2 \quad \forall \boldsymbol{\theta} \in \mathbb{R}^{(L+1)pq_1}, \tag{3.4}$$

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

(3.4) 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 4 formalize this intuition by showing (3.4) holds with high probability. Furthermore, Propositions 6 and 7 verify the essential deviation bounds for model (1.1), where further discussions can be found in Appendix A.

Now we are ready to present the main convergence result.

Theorem 4. *Suppose that Conditions 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, $\lambda_n \geqslant 2C_0 s q_1^{1/2}\{(\mathcal{M}_1^X + \mathcal{M}^\epsilon) \vee \mathcal{M}^Y\}\{(q_1^{\alpha_1+3/2} \vee q_2^{\alpha_2+3/2})\sqrt{\frac{\log(pq_1q_2)}{n}} + q_1^{-\kappa+1/2}\}$ 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*

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_1 \lesssim \frac{q_1^{\alpha_1/2}s\lambda_n}{\underline{\mu}}, \tag{3.5}$$

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

Remark 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 (3.5) 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 5 are small and decay rate of the upper bounds for basis coefficients (κ) in Condition 6 and curvature $(\underline{\mu})$ in (3.4) 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 2 under the independence assumption between $\{\mathbf{X}_t(\cdot)\}$ and $\{\epsilon_t(\cdot)\}$, which is used to verify the deviation bound in Proposition 7. Moreover, provided that our estimation is based on the representation in (3.1), 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 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 (3.5) 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 (3.5) is consistent to that of the VFAR estimate in Guo and Qiao (2021).

4. High-dimensional partially functional linear regression

This section is organized in the same manner as Section 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.

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 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,

$$\mathcal{Y} = \sum_{j=1}^{p} \mathcal{X}_j \Psi_j + \mathcal{Z}\gamma + R + E, \tag{4.1}$$

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 (4.1), 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 :

$$\frac{1}{2}\|\mathcal{Y} - \sum_{j=1}^{p} \widehat{\mathcal{X}}_j \Psi_j - \mathcal{Z}\gamma\|^2 + \lambda_{n1} \sum_{j=1}^{p} \|\widehat{\mathcal{X}}_j \Psi_j\| + \widetilde{\lambda}_{n2}\|\gamma\|_1, \tag{4.2}$$

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 (4.2) 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 $j = 1, \ldots, p$.

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 5 by α and assume the same truncated dimension across $j = 1, \ldots, p$, denoted by q. Let $\widehat{\mathcal{X}} = (\widehat{\mathcal{X}}_1, \ldots, \widehat{\mathcal{X}}_p) \in \mathbb{R}^{n \times pq}$, $\Psi = (\Psi_1^{\mathrm{T}}, \ldots, \Psi_p^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{pq}$, $\widehat{D} =$

$\text{diag}(\widehat{D}_1, \dots, \widehat{D}_p) \in \mathbb{R}^{pq \times pq}$, 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 $j = 1, \dots, p$. Then our minimizing task in (4.2) is equivalent to

$$(\widehat{B}, \widehat{\gamma}) = \underset{B \in \mathbb{R}^{pq}, \gamma \in \mathbb{R}^d}{\arg\min} \left\{ \frac{1}{2n} \|\mathcal{Y} - \widehat{\Omega}B - \mathcal{Z}\gamma\|^2 + \lambda_{n1}\|B\|_1^{(q)} + \lambda_{n2}\|\gamma\|_1 \right\}, \tag{4.3}$$

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 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 8 controls the level of smoothness for non-zero coefficient functions in $\mathcal{B}(\cdot)$. See also Condition 6 for model (1.1) and its subsequent discussion.

Condition 9. For the mixed process $\{\mathbf{X}_t(\cdot), \mathbf{Z}_t\}_{t \in \mathbb{Z}}$, we denote a diagonal matrix function by $\mathbf{D}_0^X = \text{diag}(\Sigma_{0,11}^X, \dots, \Sigma_{0,pp}^X)$. The infimum $\underline{\mu}^*$ is bounded below by zero, i.e.

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

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 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 $\Big(\frac{\xi_{t11}}{\sqrt{\omega_{11}^X}}, \dots, \frac{\xi_{t1b_1}}{\sqrt{\omega_{1b_1}^X}}, \dots, \frac{\xi_{tp1}}{\sqrt{\omega_{p1}^X}}, \dots, \frac{\xi_{tpb_p}}{\sqrt{\omega_{pb_p}^X}}, Z_{t1}, \dots, Z_{td}\Big)^{\mathrm{T}} \in \mathbb{R}^{\sum_{j=1}^p b_j + d}$.

We next present Proposition 5 below and Propositions 8–9 in Appendix A to respectively show that the RE and deviation conditions are satisfied with high probability. These results together with Proposition 6(i) lead to theoretical guarantees for regularized estimates of model (1.2).

Proposition 5. *Suppose Conditions 1–5 and 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* $1 - c_1^*(pq+d)^{-c_2^*}$,

$$\frac{1}{n}\boldsymbol{\theta}^{\mathrm{T}}\mathcal{S}^{\mathrm{T}}\mathcal{S}\boldsymbol{\theta} \geqslant \tau_2^*\|\boldsymbol{\theta}\|^2 - \tau_1^*\|\boldsymbol{\theta}\|_1^2, \quad \forall \boldsymbol{\theta} \in \mathbb{R}^{pq+d}, \tag{4.4}$$

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 (4.4), 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 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 5 below. Similar to Proposition 7, Propositions 8 and 9 in Appendix 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 5. *Suppose that Conditions 1–5, 8 and 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*

$$\|\widehat{\mathcal{B}} - \mathcal{B}\|_1 + q^{\alpha/2}\|\widehat{\gamma} - \gamma\|_1 \lesssim \frac{q^{\alpha/2}\lambda_n(s_1 + s_2)}{\underline{\mu}^*}, \tag{4.5}$$

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

Remark 6. *(a) The error bound in (4.5) 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 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 (A.5) in Proposition 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 4 under the special case where the non-functional response results in the absence of* $\mathcal{M}^Y$ *and* q_2 *in the rate.*

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 5.1 and 5.2, respectively.

5.1. High-dimensional functional linear lagged regression

We consider model (1.1) with $L = 1$, where functional covariates $\{\mathbf{X}_t(\cdot)\}_{t=1,\dots,n}$ are generated from a sparse VFAR model (Guo and Qiao, 2021). Specifically, we generate $X_{tj}(u) = \boldsymbol{\zeta}_{tj}^{\mathrm{T}}\boldsymbol{\psi}(u)$ for $j = 1, \dots, p$ and $u \in \mathcal{U} = [0, 1]$, where $\boldsymbol{\psi}(\cdot) = (\psi_1(\cdot), \dots, \psi_5(\cdot))^{\mathrm{T}}$ is a 5-dimensional Fourier basis function and $\boldsymbol{\zeta}_t = (\boldsymbol{\zeta}_{t1}^{\mathrm{T}}, \dots, \boldsymbol{\zeta}_{tp}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{5p}$ 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 $\iota \sim$ 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 Unif$(0, 0.4)$ for $h = 0$ and 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, 2021) to solve the optimization problem (3.2) 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

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

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 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 (3.2) with $\lambda_n = 0$ still require to estimate $10 \times 5^2 = 250$ parameters, which is intrinsically a high-dimensional estimation problem.

TABLE 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 AUROC	ℓ_1/ℓ_2-LS Estimation error	OLS-O 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)

5.2. High-dimensional partially functional linear regression

We now consider model (1.2) with p-dimensional vector of functional covariates $\{\mathbf{X}_t(\cdot)\}_{t=1,\dots,n}$ and d-dimensional scalar covariates $\{\mathbf{Z}_t\}_{t=1,\dots,n}$, which are jointly generated in a similar procedure as in Section 5.1. Let $X_{tj}(u) = \boldsymbol{\zeta}_{tj}^{\mathrm{T}}\boldsymbol{\psi}(u)$ for $j = 1,\dots,p$ 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 $j = 1,\dots,p$, $\mathbf{W}_{jk}^* \in \mathbb{R}^{5\times 5}$ $(k = 1,\dots,p)$ and $\mathbf{W}_{jk}^* \in \mathbb{R}^5$ $(k = p+1,\dots,p+d)$ 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 $j = p+1,\dots,p+d$, $(\mathbf{W}_{jk}^*)^{\mathrm{T}} \in \mathbb{R}^5$ $(k = 1,\dots,p)$ and $\mathbf{W}_{jk}^* \in \mathbb{R}$ $(k = p+1,\dots,p+d)$ 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 $j \in S_1 = \{1,\dots,5\}$, 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 $k \in S_2 = \{1,\dots,10\}$, the non-zero scalar coefficients γ_k's are uniformly sampled from $[0.5, 1]$. Finally, we generate responses $\{Y_t\}_{t=1,\dots,n}$ 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 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 (4.2) on the training data and obtain $\{\widehat{\beta}_j^{(\lambda_{n1},\lambda_{n2})}(\cdot)\}_{j=1,\dots,p}$ and $\{\widehat{\gamma}_k^{(\lambda_{n1},\lambda_{n2})}\}_{k=1,\dots,d}$. 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. $\sum_{t=1}^n \big\{Y_t - \sum_{j=1}^p \int_{\mathcal{U}} X_{tj}(u)\widehat{\beta}_j^{(\lambda_{n1},\lambda_{n2})}(u)du - \sum_{k=1}^d Z_{tk}\widehat{\gamma}_k^{(\lambda_{n1},\lambda_{n2})}\big\}^2$.

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 2, where the relative estimation errors for functional and scalar coefficients are $\|\widehat{\mathcal{B}} - \mathcal{B}\|/\|\mathcal{B}\|$ and $\|\widehat{\gamma} - \gamma\|/\|\gamma\|$,

TABLE 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)

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.

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 paper and will be pursued elsewhere.

Appendix A: Additional theoretical results

We first present the following Propositions 6 and 7, in which we show that the essential deviation bounds for model (1.1) are satisfied with high probability.

Proposition 6. *Suppose that Conditions 1–5 hold. Then there exist some positive constants C_ψ, C_ω, C_ϕ, c_1^* and c_2^* such that (i) for $n \gtrsim \log(pq_1) q_1^{4\alpha_1+2}(\mathcal{M}_1^X)^2$,*

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

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

$$\max_{1\leqslant m\leqslant q_2} \|\widehat{\phi}_m - \phi_m\| \leqslant C_\phi \mathcal{M}^Y q_2^{\alpha_2+1} \sqrt{\frac{\log(q_2)}{n}}, \tag{A.2}$$

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

Proposition 7. *Suppose that Conditions 1–6 hold. Then there exist some positive constants C_0, c_1^* and c_2^* such that, for $n \gtrsim \log(pq_1q_2)(q_1^{4\alpha_1+4} \vee q_2^{4\alpha_2+4})(\mathcal{M}_1^X + \mathcal{M}^Y)^2$,*

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

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

(A.1) and (A.2) in Proposition 6 control deviation bounds for relevant estimated eigenpairs of $X_{tj}(\cdot)$ and $Y_t(\cdot)$ under the FPCA framework. (A.3) in Proposition 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 (3.1), are nicely concentrated around zero.

We next provide Propositions 8 and 9, where the essential deviation bounds for model (1.2) hold with high probability.

Proposition 8. *Suppose Conditions 1–5 and 8 hold. Then there exist some positive constants C_0^*, c_1^* and c_2^* such that, for $n \gtrsim \log(pq) q^{4\alpha+2}(\mathcal{M}_1^X)^2$,*

$$\frac{1}{n}\|\widehat{\Omega}^{\mathrm{T}}(\mathcal{Y} - \widehat{\Omega}B - \mathcal{Z}\gamma)\|_{\max}^{(q)} \leqslant C_0^* s_1 (\mathcal{M}_1^X + \mathcal{M}^\epsilon)\{q^{\alpha+2}\sqrt{\frac{\log(pq)}{n}} + q^{-\kappa+1}\}. \tag{A.4}$$

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

Proposition 9. *Suppose Conditions 1–5 and 8 hold. Then there exist some positive constants C_0^*, c_1^* and c_2^* such that, for $n \gtrsim \log(pqd) q^{3\alpha+2}\mathcal{M}_{X,Z}^2$,*

$$\frac{1}{n}\|\mathcal{Z}^{\mathrm{T}}(\mathcal{Y} - \widehat{\Omega}B - \mathcal{Z}\gamma)\|_{\max} \leqslant C_0^* s_1 (\mathcal{M}_{X,Z} + \mathcal{M}^\epsilon)\{q^{\alpha+1}\sqrt{\frac{\log(pq+d)}{n}} + q^{-\kappa+1/2}\}, \tag{A.5}$$

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

Intuitively, (A.4) in Proposition 8 (or (A.5) in Proposition 9) indicates the sample cross-covariance between estimated normalized FPC scores (or scalar covariates) and estimated errors is nicely concentrated around zero.

Appendix B: Proofs of theoretical results in Section 2

We provide proofs of theorems and propositions stated in Section 2 in Appendices B.1–B.2, followed by the supporting technical lemmas and their proofs in Appendix 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 $\mathcal{M}(\boldsymbol{f}^X, \mathbf{\Phi}) = 2\pi \cdot \operatorname{ess\,sup}_{\theta\in[-\pi,\pi]} |\langle \mathbf{\Phi}, \boldsymbol{f}_\theta^X(\mathbf{\Phi})\rangle|$.

B.1. Proofs of theorems

Proof of Theorem 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 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 $\|\mathbf{\Pi}_{M,L}\| \leqslant \mathcal{M}(\boldsymbol{f}_{M,L}^X, \mathbf{\Phi}_1)$ and $\lim_{M\to\infty} \mathcal{M}(\boldsymbol{f}_{M,L}^X, \mathbf{\Phi}_1) = \mathcal{M}(\boldsymbol{f}_L^X, \mathbf{\Phi}_1)$. 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 $\mathbf{Y}_{M,L}^{\mathrm{T}}\mathbf{Y}_{M,L}$ can be represented in the same form as $\langle \mathbf{e}_M, \mathbf{K}(\mathbf{e}_M)\rangle$ in Lemma 5, where $\mathbf{e}_M = (\boldsymbol{\varepsilon}_{M,n}^{\mathrm{T}}, \ldots, \boldsymbol{\varepsilon}_{M,1-L}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{H}^{(n+L)p}$. 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

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

$\mathbf{\Theta}_M(u) = \mathbf{I}_{n+L} \otimes \operatorname{diag}(\boldsymbol{\varphi}_{M,1}^{\mathrm{T}}, \ldots, \boldsymbol{\varphi}_{M,p}^{\mathrm{T}})$ with $\boldsymbol{\varphi}_{M,i} = \left(\sqrt{\omega_{i1}^e}\phi_{i1}, \ldots, \sqrt{\omega_{iM}^e}\phi_{iM}\right)^{\mathrm{T}}$ 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 $\mathbf{Y}_{M,L}^{\mathrm{T}}\mathbf{Y}_{M,L} = \mathbf{a}_{M,L}^{\mathrm{T}}\mathbf{\Pi}_{M,L}\mathbf{a}_{M,L}$ with $\mathbf{\Pi}_{M,L} = \mathbf{\Gamma}_{M,L}{}^{\mathrm{T}}\mathbf{\Gamma}_{M,L}$. Lemma 8 implies that $\|\operatorname{Var}(\mathbf{Y}_{M,L})\| = \|\mathbf{\Gamma}_{M,L}\mathbf{\Gamma}_{M,L}^{\mathrm{T}}\| \leqslant \mathcal{M}(\boldsymbol{f}_{M,L}^X, \mathbf{\Phi}_1)$, where $\mathcal{M}(\boldsymbol{f}_{M,L}^X, \mathbf{\Phi}_1) = 2\pi \cdot \operatorname{ess\,sup}_{\theta\in[-\pi,\pi]} \langle \mathbf{\Phi}_1, \boldsymbol{f}_{M,L,\theta}^X(\mathbf{\Phi}_1)\rangle$ and $\boldsymbol{f}_{M,L,\theta}^X(\cdot)$ 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 7, (B.8) in Lemma 5 and $\mathrm{rank}(\mathbf{\Gamma}_{\infty,L}^{\mathrm{T}}\mathbf{\Gamma}_{\infty,L}) = n$, we obtain

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

where $\mathcal{M}(f_L^X, \mathbf{\Phi}_1) = 2\pi\cdot \mathrm{ess\,sup}_{\theta\in[-\pi,\pi]}\langle \mathbf{\Phi}_1, \boldsymbol{f}_{L,\theta}^X(\mathbf{\Phi}_1)\rangle$ and $\boldsymbol{f}_{L,\theta}^X(\cdot)$ 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 9 and 10 imply that $\lim_{L\to\infty}\mathbb{E}\left\{\left|\langle \mathbf{\Phi}_1, (\widehat{\mathbf{\Sigma}}_{L,0}^X - \widehat{\mathbf{\Sigma}}_0^X)(\mathbf{\Phi}_1)\rangle\right|\right\} = 0$, $\lim_{L\to\infty}\langle \mathbf{\Phi}_1, \mathbf{\Sigma}_{L,0}^X(\mathbf{\Phi}_1)\rangle = \langle \mathbf{\Phi}_1, \mathbf{\Sigma}_0^X(\mathbf{\Phi}_1)\rangle$ and $\lim_{L\to\infty}\mathcal{M}(f_L^X, \mathbf{\Phi}_1) = \mathcal{M}(\boldsymbol{f}^X, \mathbf{\Phi}_1)$. Combining the above results and following the similar argument in the proof of Lemma 5, we obtain

$$\begin{aligned} P\left\{\left|\langle \mathbf{\Phi}_1, (\widehat{\mathbf{\Sigma}}_0^X - \mathbf{\Sigma}_0^X)(\mathbf{\Phi}_1)\rangle\right| > \mathcal{M}(\boldsymbol{f}^X, \mathbf{\Phi}_1)\eta\right\} \\ \leqslant 2\exp\left\{-cn\min\left(\eta^2,\eta\right)\right\}. \end{aligned} \tag{B.1}$$

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

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

which completes the proof of (2.12). Part (ii): For fixed vectors $\mathbf{\Phi}_1 \in \mathbb{H}^p$ and $\mathbf{\Phi}_2 \in \mathbb{H}^d$, we denote $\mathcal{M}(\boldsymbol{f}^{X,Y}, \mathbf{\Phi}_1, \mathbf{\Phi}_2) = 2\pi\cdot \mathrm{ess\,sup}_{\theta\in[-\pi,\pi]}|\langle \mathbf{\Phi}_1, \boldsymbol{f}_\theta^{X,Y}(\mathbf{\Phi}_2)\rangle|$. Define $\mathbf{M}_t(\cdot) = [(\mathbf{X}_t(\cdot))^{\mathrm{T}}, (\mathbf{Y}_t(\cdot))^{\mathrm{T}}]^{\mathrm{T}}$. Letting $\mathbf{\Phi} = (\mathbf{\Phi}_1^{\mathrm{T}}, \mathbf{\Phi}_2^{\mathrm{T}})^{\mathrm{T}}$, we have

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

Applying (B.1) on $\{\mathbf{X}_t(\cdot)\}$ and $\{\mathbf{Y}_t(\cdot)\}$, we obtain that

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

For $\{\mathbf{M}_t(\cdot)\}$, $\mathcal{M}(f^M, \mathbf{\Phi}) \leqslant \mathcal{M}(\boldsymbol{f}^X, \mathbf{\Phi}_1) + \mathcal{M}(\boldsymbol{f}^Y, \mathbf{\Phi}_2) + 2\mathcal{M}(\boldsymbol{f}^{X,Y}, \mathbf{\Phi}_1, \mathbf{\Phi}_2)$. This, together with (B.1) implies that

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

Combining the above results, we obtain

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

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

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

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 (B.2) 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

$$\begin{aligned}&P\left\{\left|\frac{\langle\boldsymbol{\Phi}_1,(\widehat{\boldsymbol{\Sigma}}_0^{X,Y}-\boldsymbol{\Sigma}_0^{X,Y})(\boldsymbol{\Phi}_2)\rangle}{\langle\boldsymbol{\Phi}_1,\boldsymbol{\Sigma}_0^X(\boldsymbol{\Phi}_1)\rangle+\langle\boldsymbol{\Phi}_2,\boldsymbol{\Sigma}_0^Y(\boldsymbol{\Phi}_2)\rangle}\right|>\left(\mathcal{M}_{k_1}^X+\mathcal{M}_{k_2}^Y+\mathcal{M}_{k_1,k_2}^{X,Y}\right)\eta\right\}\\ &\qquad\leqslant 6\exp\left\{-cn\min(\eta^2,\eta)\right\},\\ &P\left\{\left|\frac{\langle\boldsymbol{\Phi}_1,(\widehat{\boldsymbol{\Sigma}}_h^{X,Y}-\boldsymbol{\Sigma}_h^{X,Y})(\boldsymbol{\Phi}_2)\rangle}{\langle\boldsymbol{\Phi}_1,\boldsymbol{\Sigma}_0^X(\boldsymbol{\Phi}_1)\rangle+\langle\boldsymbol{\Phi}_2,\boldsymbol{\Sigma}_0^Y(\boldsymbol{\Phi}_2)\rangle}\right|>2\left(\mathcal{M}_{k_1}^X+\mathcal{M}_{k_2}^Y+\mathcal{M}_{k_1,k_2}^{X,Y}\right)\eta\right\}\\ &\qquad\leqslant 12\exp\{-cn\min(\eta^2,\eta)\}.\end{aligned}$$

Letting $c_2=c/4$, we complete the proof of (2.13). □

Proof of Theorem 2 Under FPCA framework, for each $k = 1, \dots, d$, we have $Y_{tk}(\cdot) = \sum_{m=1}^{\infty} \xi_{tkm}\phi_{km}(\cdot)$ with eigenpairs $(\omega_{km}^Y, \phi_{km})$, and for each $j = 1, \dots, p$, 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 $\mathbf{\Phi}_1 = (0, \dots, 0, \{\omega_{jl}^X\}^{-\frac{1}{2}}\psi_{jl}, 0, \dots, 0)^{\mathrm{T}}$ and $\mathbf{\Phi}_2 = (0, \dots, 0, \{\omega_{km}^Y\}^{-\frac{1}{2}}\phi_{km}, 0, \dots, 0)^{\mathrm{T}}$. Following the similar argument in the proof of Theorem 2 in Guo and Qiao (2021) with $2\sqrt{\omega_0^X\omega_0^Y} \leqslant \omega_0^X + \omega_0^Y$ and Theorem 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 (2.15). □

Before presenting the proof of Theorem 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 $j = 1, \dots, p$ and $l = 1, 2, \dots$. It follows from (4.43) and Lemma 4.3 of Bosq (2000) that

$$\sup_{l\geqslant 1}|\widehat{\omega}_{jl}^X - \omega_{jl}^X| \leqslant \|\widehat{\Delta}_{jj}^X\|_{\mathcal{S}} \quad \text{and} \quad \sup_{l\geqslant 1}\delta_{jl}^X\|\widehat{\psi}_{jl} - \psi_{jl}\| \leqslant 2\sqrt{2}\|\widehat{\Delta}_{jj}^X\|_{\mathcal{S}}. \tag{B.3}$$

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 $k = 1, \dots, d$ and $m = 1, 2, \dots$, we have

$$\sup_{m\geqslant 1}|\widehat{\omega}_{km}^Y - \omega_{km}^Y| \leqslant \|\widehat{\Delta}_{kk}^Y\|_{\mathcal{S}} \quad \text{and} \quad \sup_{m\geqslant 1}\delta_{km}^Y\|\widehat{\phi}_{km} - \phi_{km}\| \leqslant 2\sqrt{2}\|\widehat{\Delta}_{kk}^Y\|_{\mathcal{S}}. \tag{B.4}$$

Proof of Theorem 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

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

Let $\Omega^{X,Y}_{jk,\eta} = \Big\{\|\widehat{\Delta}^{X,Y}_{h,jk}\|_{\mathcal{S}} \leqslant (\omega^X_0 + \omega^Y_0)\mathcal{M}_{X,Y}\eta\Big\}$, $\Omega^X_{jj,\eta} = \Big\{\|\widehat{\Delta}^X_{jj}\|_{\mathcal{S}} \leqslant 2\mathcal{M}^X_1\omega^X_0\eta\Big\}$, $\Omega^Y_{kk,\eta} = \Big\{\|\widehat{\Delta}^Y_{kk}\|_{\mathcal{S}} \leqslant 2\mathcal{M}^Y_1\omega^Y_0\eta\Big\}$ and $\Omega_1 = \Big\{\|\widehat{\Delta}^{X,Y}_{h,jk}\|_{\mathcal{S}} \leqslant (\omega^X_0 + \omega^Y_0)\Big\}$. By Theorem 2 and Lemma 24, we have

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

On the event of $\Omega_1\cap\Omega^{X,Y}_{\eta,jk}\cap\Omega^X_{jj,\eta}\cap\Omega^Y_{kk,\eta}$, by Condition 5, (B.3), (B.4), Lemma 2 and the fact that $(\omega^X_0\omega^Y_0)^{1/2} \leqslant 1/2(\omega^X_0+\omega^Y_0)$, we obtain that

$$
\begin{aligned}
\left|\frac{I_1}{\sqrt{\omega^X_{jl}\omega^Y_{km}}}\right| &\leqslant c_0^{-1}(\alpha_1\alpha_2)^{1/2} l^{\alpha_1/2} m^{\alpha_2/2}\|\widehat{r}_{jl}\|(\|\widehat{\Delta}^{X,Y}_{h,jk}\|_{\mathcal{S}} + \|\Sigma^{X,Y}_{h,jk}\|_{\mathcal{S}})\|\widehat{w}_{km}\|\\
&\lesssim l^{3\alpha_1/2+1} m^{3\alpha_2/2+1}\|\widehat{\Delta}^X_{jj}\|_{\mathcal{S}}\|\widehat{\Delta}^Y_{kk}\|_{\mathcal{S}}(\|\widehat{\Delta}^{X,Y}_{h,jk}\|_{\mathcal{S}} + (\omega^X_0\omega^Y_0)^{1/2})\\
&\lesssim (l^{3\alpha_1+2}\vee m^{3\alpha_2+2})\mathcal{M}^X_1\mathcal{M}^Y_1\eta^2,\\
&\lesssim (l^{3\alpha_1+2}\vee m^{3\alpha_2+2})(\mathcal{M}^X_1+\mathcal{M}^Y_1)^2\eta^2,\\
\left|\frac{I_2}{\sqrt{\omega^X_{jl}\omega^Y_{km}}}\right| &\leqslant c_0^{-1}(\alpha_1\alpha_2)^{1/2} l^{\alpha_1/2} m^{\alpha_2/2}\|\widehat{\Delta}^{X,Y}_{h,jk}\|_{\mathcal{S}}(\|\widehat{r}_{jl}\| + \|\widehat{w}_{km}\|)\\
&\lesssim l^{\alpha_1/2} m^{\alpha_2/2}\|\widehat{\Delta}^{X,Y}_{h,jk}\|_{\mathcal{S}}(l^{\alpha_1+1}\|\widehat{\Delta}^X_{jj}\|_{\mathcal{S}} + m^{\alpha_2+1}\|\widehat{\Delta}^Y_{kk}\|_{\mathcal{S}})\\
&\lesssim (l^{2\alpha_1+1}\vee m^{2\alpha_2+1})\mathcal{M}_{X,Y}(\mathcal{M}_X\vee\mathcal{M}_Y)\eta^2,\\
&\lesssim (l^{2\alpha_1+1}\vee m^{2\alpha_2+1})\mathcal{M}^2_{X,Y}\eta^2,
\end{aligned}
$$

By Theorem 1,

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

Next, we consider the term $I_3 = \langle\widehat{r}_{jl}, \langle\Sigma^{X,Y}_{h,jk},\phi_{km}\rangle\rangle + \langle\psi_{jl},\langle\Sigma^{X,Y}_{h,jk},\widehat{w}_{km}\rangle\rangle$. By Condition 5, Lemmas 14 and 26 for $\{\mathbf{X}_t\}_{t\in\mathbb{Z}}$ and $\{\mathbf{Y}_t\}_{t\in\mathbb{Z}}$, we obtain that

$$
\begin{aligned}
&\left|\frac{I_3}{\sqrt{\omega^X_{jl}\omega^Y_{km}}}\right|\\
\lesssim&\mathcal{M}^X_1 l^{\alpha_1+1}\eta + (M^X_1)^2 l^{(5\alpha_1+4)/2}\eta^2 + \mathcal{M}^Y_1 m^{\alpha_2+1}\eta + (\mathcal{M}^Y_1)^2 m^{(5\alpha_2+4)/2}\eta^2\\
\lesssim&(l^{\alpha_1+1}\vee m^{\alpha_2+1})(\mathcal{M}^X_1+\mathcal{M}^Y_1)\eta + (l^{(5\alpha_1+4)/2}\vee m^{(5\alpha_2+4)/2})(\mathcal{M}^X_1+\mathcal{M}^Y_1)^2\eta^2
\end{aligned}
$$

holds with probability greater than $1 - 16\exp\{-\tilde{c}_4 n\min(\eta^2,\eta)\} - 8\exp\{-\tilde{c}_4 n(\{\mathcal{M}^X_1\}^2 l^{2(\alpha_1+1)}\vee\{\mathcal{M}^Y_1\}^2 m^{2(\alpha_2+1)})^{-1}\}$.

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

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

where $\mathcal{M}_{X,Y}=\mathcal{M}^X_1+\mathcal{M}^Y_1+\mathcal{M}^{X,Y}_{1,1}$. Applying the Boole's inequality, we obtain that

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

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

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

B.2. Proofs of propositions

Proof of Proposition 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 $\widehat{\mathbf{\Sigma}}^{X,Z}_h$,

$$P\left\{\left|\frac{\langle\mathbf{\Phi}_1,(\widehat{\mathbf{\Sigma}}^{X,Z}_h-\mathbf{\Sigma}^{X,Z}_h)\boldsymbol{\nu}\rangle}{\langle\mathbf{\Phi}_1,\mathbf{\Sigma}^X_0(\mathbf{\Phi}_1)\rangle+\boldsymbol{\nu}^{\mathrm{T}}\mathbf{\Sigma}^Z_0\boldsymbol{\nu}}\right|>\left(\mathcal{M}^X_{k_1}+\mathcal{M}^Z_{k_2}+\mathcal{M}^{X,Z}_{k_1,k_2}\right)\eta\right\}$$
$$\leqslant c_1\exp\{-c_2n\min(\eta^2,\eta)\}. \tag{B.5}$$

Provided with Lemma 28, the above result can be proved in similar way to (2.13) in Theorem 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

$\|\Sigma_{h,jk}^{X,Z} - \Sigma_{h,jk}^{X,Z}\|^2 = \sum_{l=1}^{\infty} \omega_{jl}^X (\sigma_{0,kk}^Z)^2 \Delta_{h,jkl}^2$. By Jensen's inequality, we have that

$$\begin{aligned}\mathbb{E}\Big\{\|\widehat{\Sigma}_{h,jk}^{X,Z} - \Sigma_{h,jk}^{X,Z}\|_{\mathcal{S}}^{2q}\Big\} &\leqslant (\sigma_{0,kk}^Z)^{2q}\Big(\sum_{l=1}^{\infty}\omega_{jl}^X\Big)^{q-1}\sum_{l=1}^{\infty}\omega_{jl}^X \mathbb{E}\big|\Delta_{h,jkl}\big|^{2q} \\ &\leqslant \{\sigma_0^Z\}^{2q}\{\omega_0^X\}^q \sup_l \mathrm{E}\big|\Delta_{h,jkl}\big|^{2q}.\end{aligned}$$

By (B.5), we obtain that

$$P\left\{|\Delta_{h,jkl}| > 2\mathcal{M}_{X,Z}\eta\right\} \leqslant c_1 \exp\{-c_2 n \min(\eta^2,\eta)\}.$$

Combining the above results and following the similar argument in the proof of Theorem 2 in Guo and Qiao (2021) yields

$$P\left\{\|\widehat{\Sigma}_{h,jk}^{X,Z} - \Sigma_{h,jk}^{X,Z}\| > 2\mathcal{M}_{X,Z}\sigma_0^Z\sqrt{\omega_0^X}\eta\right\} \leqslant c_1\exp\{-c_3 n\min(\eta^2,\eta)\}.$$

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

$$P\left\{\|\widehat{\Sigma}_{h,jk}^{X,Z} - \Sigma_{h,jk}^{X,Z}\| > ((\sigma_0^Z)^2 + \omega_0^X)\mathcal{M}_{X,Z}\eta\right\} \leqslant c_1\exp\{-c_3 n\min(\eta^2,\eta)\}. \quad \text{(B.6)}$$

This also implies (2.16).

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 $\Omega_{jk,\eta}^{X,Z} = \left\{\|\widehat{\Delta}_{h,jk}^{X,Z}\| \leqslant (\omega_0^X + (\sigma_0^Z)^2)\mathcal{M}_{X,Z}\eta\right\}$, $\Omega_{jj,\eta}^X = \left\{\|\widehat{\Delta}_{jj}^X\|_{\mathcal{S}} \leqslant 2\mathcal{M}_1^X\omega_0^X\eta\right\}$ and $\Omega_1 = \left\{\|\widehat{\Delta}_{h,jk}^{X,Z}\| \leqslant (\omega_0^X + (\sigma_0^Z)^2)\right\}$. By (B.6) and Lemma 24, we have

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

On the event of $\Omega_1 \cap \Omega_{\eta,jk}^{X,Z} \cap \Omega_{jj,\eta}^X$, by Condition 5, (B.3), Lemma 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 5, Lemma 26 and $\|\Sigma_{h,jk}^{XZ}\| \leqslant \omega_0^{1/2}\sigma_{0,kk}^{Z}$, we obtain that

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

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 (B.5) 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

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

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

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

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

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

which implies (2.18). □

Proof of Proposition 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 $\widehat{\Delta}_{h,j}^{X,\epsilon}=\widehat{\Sigma}_{h,j}^{X,\epsilon}-\Sigma_{h,j}^{X,\epsilon}$.

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

Denote $\Omega_{j,\eta}^{X,\epsilon}=\Big\{\|\widehat{\Delta}_{h,j}^{X,\epsilon}\|_{\mathcal{S}}\leqslant(\omega_0^X+\omega_0^\epsilon)\mathcal{M}_{X,\epsilon}\eta\Big\}$, $\Omega_{jj,\eta}^X=\Big\{\|\widehat{\Delta}_{jj}^X\|_{\mathcal{S}}\leqslant 2\mathcal{M}_1^X\omega_0^X\eta\Big\}$, $\Omega_\eta^Y=\Big\{\|\widehat{\Delta}^Y\|_{\mathcal{S}}\leqslant 2\mathcal{M}^Y\omega_0^Y\eta\Big\}$ and $\Omega_1=\Big\{\|\widehat{\Delta}_{h,j}^{X,\epsilon}\|_{\mathcal{S}}\leqslant(\omega_0^X+\omega_0^\epsilon)\Big\}$. By Theorem 2 and Lemma 24, we have

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

On the event of $\Omega_1 \cap \Omega_{j,\eta}^{X,\epsilon} \cap \Omega_{jj,\eta}^{X} \cap \Omega_{\eta}^{Y}$, by Condition 5, (B.3), (B.4) and Lemma 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

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

Letting $\eta = \rho_{11}\sqrt{\frac{\log(pM_1M_2)}{n}} < 1$ and $\rho_9 + \rho_{10}\rho_{11}\{\mathcal{M}_1^X + \mathcal{M}^Y\}(M_1^{2\alpha_1+2} \vee M_2^{2\alpha_2+2})\eta \leqslant \rho_{12}$, there exist some constants $c_9, c_{10} > 0$ such that

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

which completes the proof. □

Proof of Proposition 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 $\Omega_{j,\eta}^{X,\epsilon} = \Big\{ \|\widehat{\Delta}_{h,j}^{X,\epsilon}\| \leqslant (\omega_0^X + (\sigma_0^\epsilon)^2)\mathcal{M}_{X,\epsilon}\eta \Big\}$ and $\Omega_{jj,\eta}^{X} = \Big\{ \|\widehat{\Delta}_{jj}^{X}\|_{\mathcal{S}} \leqslant 2\mathcal{M}_1^X \omega_0^X \eta \Big\}$. By (B.6) and Lemma 24, we have

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

On the event of $\Omega_{\eta,j}^{X,\epsilon} \cap \Omega_{jj,\eta}^{X}$, by Condition 5, (B.3) and Lemma 2, we obtain that

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

By (B.5) 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

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

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 (2.20). □

B.3. Technical lemmas and their proofs

Lemma 1. *The non-functional version of our proposed cross-spectral stability measure satisfies*

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

where $\widetilde{\mathcal{M}}^{X,Y}$ is defined in (2.4).

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

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

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

Lemma 2. *Suppose that Conditions 3 and 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

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

which completes our proof. □

Before presenting Lemma 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 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 $\eta > 0$,*

$$P\left(|\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} - \mathbb{E}\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}| \geqslant \|\mathbf{A}\|\eta\right) \leqslant 2\exp\left\{-c\min\left(\frac{\eta^2}{\mathit{rank}(\mathbf{A})}, \eta\right)\right\}. \tag{B.7}$$

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 (B.7). □

Lemma 4. *Suppose that sub-Gaussian process $\{\varepsilon_{tj}(\cdot)\}_{t\in\mathbb{Z}}$ follows Definition 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 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

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

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

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

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 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 $\boldsymbol{\varphi}_{M,i} = \left(\sqrt{\omega_{i1}^e}\phi_{i1}, \dots, \sqrt{\omega_{iM}^e}\phi_{iM}\right)^{\mathrm{T}}$.
(ii) Suppose $\mathbf{K} = (K_{ij})_{N\times N}$ with each $K_{ij} \in \mathbb{S}$. For any nonempty subset $G \subset \mathbb{Z}_+ = \{1,2,\dots\}$ with $|G| < \infty$, write $G = \{g_1, \dots, g_{|G|}\}$ with $g_1 < \cdots < g_{|G|}$ and $\boldsymbol{\phi}_{G,i} = (\phi_{ig_1}, \dots, \phi_{ig_{|G|}})^{\mathrm{T}}$ for each $i = 1, \dots, N$. Let $\boldsymbol{\Phi}_G = \mathrm{diag}(\boldsymbol{\phi}_{G,1}^{\mathrm{T}}, \dots, \boldsymbol{\phi}_{G,N}^{\mathrm{T}})$, then we define

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

Condition 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 $\boldsymbol{\Theta}_M = \mathrm{diag}(\boldsymbol{\varphi}_{M,1}^{\mathrm{T}}, \dots, \boldsymbol{\varphi}_{M,N}^{\mathrm{T}})$ 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 5. *Suppose that* $\max_{1\leqslant i\leqslant N} \int_{\mathcal{U}} \Sigma_{ii}^e(u,u)du < \infty$ *and* $\mathbf{K}$ *satisfies Condition 10. Then, there exists some universal constant* $c > 0$ *such that for any given* $\eta > 0$,

$$P\left(|\langle \mathbf{e}, \mathbf{K}(\mathbf{e})\rangle - \mathbb{E}\langle \mathbf{e}, \mathbf{K}(\mathbf{e})\rangle| \geqslant b\eta\right) \leqslant 2\exp\left\{-c\min\left(\frac{\eta^2}{rank(\mathbf{K})}, \eta\right)\right\}. \tag{B.8}$$

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 3 and finally show that the inequality still hold under the infinite-dimensional setting.

Rewrite $\mathbf{e}_M = (e_{M,1}, \dots, e_{M,N})^{\mathrm{T}}$ with $e_{M,i} = \mathbf{a}_{M,i}^{\mathrm{T}}\boldsymbol{\varphi}_{M,i}$ and $\mathbf{a}_{M,i} = (a_{i1}, \dots, a_{iM})^{\mathrm{T}}$. Let $\mathbf{a}_M = (\mathbf{a}_{M,1}^{\mathrm{T}}, \dots, \mathbf{a}_{M,N}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{NM}$, 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 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 3 yields

$$\begin{aligned} &P\left(|\langle \mathbf{e}_M, \mathbf{K}(\mathbf{e}_M)\rangle - \mathbb{E}\langle \mathbf{e}_M, \mathbf{K}(\mathbf{e}_M)\rangle| \geqslant b_M\eta\right) \\ &\leqslant P\left(|\mathbf{a}_M^{\mathrm{T}}\boldsymbol{\Pi}_M\mathbf{a}_M - \mathbb{E}\mathbf{a}_M^{\mathrm{T}}\boldsymbol{\Pi}_M\mathbf{a}_M| \geqslant \|\boldsymbol{\Pi}_M\|\eta\right) \\ &\leqslant 2\exp\left\{-c\min\left(\frac{\eta^2}{\mathrm{rank}(\boldsymbol{\Pi}_M)}, \eta\right)\right\}. \end{aligned} \tag{B.9}$$

It follows from Lemma 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 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 $\mathrm{rank}(\boldsymbol{\Pi}_M) \leqslant \mathrm{rank}(\mathbf{K})$ and taking $M \to \infty$ on both sides of (B.9), we obtain (B.8), which completes the proof. $\square$

Lemma 6. *Under the same assumption and notation in Lemma 5 and its proof, we have*

$$\lim_{M\to\infty} \mathbb{E}\left\{\|\mathbf{e}_M - \mathbf{e}\|^2\right\} = 0 \tag{B.10}$$

and

$$\lim_{M\to\infty} \mathbb{E}\langle \mathbf{e}_M, \mathbf{K}(\mathbf{e}_M)\rangle = \mathbb{E}\langle \mathbf{e}, \mathbf{K}(\mathbf{e})\rangle. \tag{B.11}$$

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 4 implies that above goes to zero as $M \to \infty$, which completes the proof of (B.10).

By triangle inequality, we have

$$|\mathbb{E}\langle \mathbf{e}_M, \mathbf{K}(\mathbf{e}_M)\rangle - \mathbb{E}\langle \mathbf{e}, \mathbf{K}(\mathbf{e})\rangle| \leqslant |\mathbb{E}\langle \mathbf{e}_M, \mathbf{K}(\mathbf{e}_M - \mathbf{e})\rangle| + |\mathbb{E}\langle (\mathbf{e}_M - \mathbf{e}), \mathbf{K}(\mathbf{e})\rangle|. \tag{B.12}$$

By Jensen's inequality and Lemma 11, we have

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

From (B.10), we have $\lim_{M\to\infty} \mathbb{E}\left\{\|\mathbf{e}_M - \mathbf{e}\|^2\right\} = 0$ 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 (B.12) goes to zero when $M \to \infty$, which completes the proof of (B.11). □

Lemma 7. *Suppose Conditions 1, 3 and 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* $k = 1, \dots, p$,

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

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 in Appendix B.1, we have

$$
\begin{aligned}
&\lim_{M\to\infty} |\mathcal{M}(\boldsymbol{f}^X_{M,L}, \boldsymbol{\Phi}_1) - \mathcal{M}(\boldsymbol{f}^X_L, \boldsymbol{\Phi}_1)| \\
&= 2\pi \lim_{M\to\infty} \left| \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi]} |\langle \boldsymbol{\Phi}_1, \boldsymbol{f}^X_{M,L,\theta}(\boldsymbol{\Phi}_1)\rangle| - \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi]} |\langle \boldsymbol{\Phi}_1, \boldsymbol{f}^X_{L,\theta}(\boldsymbol{\Phi}_1)\rangle| \right| \\
&\leqslant 2\pi \lim_{M\to\infty} \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi]} \left| |\langle \boldsymbol{\Phi}_1, \boldsymbol{f}^X_{M,L,\theta}(\boldsymbol{\Phi}_1)\rangle| - |\langle \boldsymbol{\Phi}_1, \boldsymbol{f}^X_{L,\theta}(\boldsymbol{\Phi}_1)\rangle| \right| \\
&\leqslant \|\boldsymbol{\Phi}_1\|^2 \lim_{M\to\infty} \left\| \sum_{h\in\mathbb{Z}} (\boldsymbol{\Sigma}^X_{M,L,h} - \boldsymbol{\Sigma}^X_{L,h}) \right\|_{\mathrm{F}} \qquad \text{(by Lemma 11 and } |\exp(-ih\theta)| = 1) \\
&\leqslant \|\boldsymbol{\Phi}_1\|^2 \lim_{M\to\infty} \sum_{h\in\mathbb{Z}} \left\| \boldsymbol{\Sigma}^X_{M,L,h} - \boldsymbol{\Sigma}^X_{L,h} \right\|_{\mathrm{F}}.
\end{aligned}
$$

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

By triangle inequality and Lemma 12, we obtain that

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

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

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

Then,

$$
\begin{aligned}
&\lim_{M\to\infty}\left\|\boldsymbol{\Sigma}^X_{M,L,h}-\boldsymbol{\Sigma}^X_{L,h}\right\|_{\mathrm{F}}\\
=&\lim_{M\to\infty}\left\|\sum_{l=0}^{L-h}\int \mathbf{A}_{l+h}(u,u')\{\boldsymbol{\Sigma}_0^{\varepsilon_M}(u',v')-\boldsymbol{\Sigma}_0^{\varepsilon}(u',v')\}\{\mathbf{A}_l(v,v')\}^{\mathrm{T}}du'dv'\right\|_{\mathrm{F}}\\
\leqslant&\sum_{l=0}^{L-h}\|\mathbf{A}_l\|_{\mathrm{F}}\|\mathbf{A}_{l+h}\|_{\mathrm{F}}\lim_{M\to\infty}\|\boldsymbol{\Sigma}_0^{\varepsilon_M}-\boldsymbol{\Sigma}_0^{\varepsilon}\|_{\mathrm{F}}\quad\text{(by Lemma 11)}\\
\leqslant&\sum_{l=0}^{L-h}\|\mathbf{A}_l\|_{\mathrm{F}}\|\mathbf{A}_{l+h}\|_{\mathrm{F}}\lim_{M\to\infty}\left\{\sum_{j,k}\|\Sigma_{h,jk}^{\varepsilon_M}-\Sigma_{h,jk}^{\varepsilon}\|_{\mathcal{S}}^2\right\}^{1/2}\\
\leqslant&\sum_{l=0}^{L-h}\|\mathbf{A}_l\|_{\mathrm{F}}\|\mathbf{A}_{l+h}\|_{\mathrm{F}}\lim_{M\to\infty}\sum_{j,k}\|\Sigma_{0,jk}^{\varepsilon_M}-\Sigma_{0,jk}^{\varepsilon}\|_{\mathcal{S}}\\
=&0\quad\text{(by Lemmas 12 and 13)}
\end{aligned}
$$

which completes the proof. □

Lemma 8. *Suppose that conditions in Lemma 7 hold. For any $\boldsymbol{\Phi}_1\in\mathbb{H}_0^p$ with $\|\boldsymbol{\Phi}_1\|_0\leqslant k$ and $k=1,\dots,p$, define $\mathbf{Y}=(\langle\boldsymbol{\Phi}_1,\mathbf{X}_1\rangle,\dots,\langle\boldsymbol{\Phi}_1,\mathbf{X}_n\rangle)^{\mathrm{T}}$. Then*

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

Proof. The proof follows from the proof of Theorem 1 in Guo and Qiao (2021) and hence the proof is omitted here. □

Lemma 9. *Suppose that conditions in Lemma 7 hold. Let $\mathbf{X}_{L,t}(u)=\sum_{l=0}^{L}\mathbf{A}_l(\boldsymbol{\varepsilon}_{t-l})$. For any $\boldsymbol{\Phi}_1\in\mathbb{H}_0^p$ with $\|\boldsymbol{\Phi}_1\|_0\leqslant k$ $(k=1,\dots,p)$, define $\mathbf{Y}_L=(\langle\boldsymbol{\Phi}_1,\mathbf{X}_{L,1}\rangle,\dots,$ $\langle\boldsymbol{\Phi}_1,\mathbf{X}_{L,n}\rangle)^{\mathrm{T}}$ and $\mathbf{Y}=(\langle\boldsymbol{\Phi}_1,\mathbf{X}_1\rangle,\dots,\langle\boldsymbol{\Phi}_1,\mathbf{X}_n\rangle)^{\mathrm{T}}$, then*

$$\lim_{L\to\infty}\mathbb{E}\left\{\|\mathbf{Y}_L-\mathbf{Y}\|^2\right\}=0 \tag{B.13}$$

and

$$\lim_{L\to\infty}\mathbb{E}\left[\mathbf{Y}_L^{\mathrm{T}}\mathbf{Y}_L\right]=\mathbb{E}\left[\mathbf{Y}^{\mathrm{T}}\mathbf{Y}\right]. \tag{B.14}$$

Proof of (B.13). By definitions of $\mathbf{Y}_L$ and $\mathbf{Y}$, we have that

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

By Lemma 11, we have $\mathbb{E}\left\{|\langle\boldsymbol{\Phi}_1,\mathbf{X}_{L,t}-\mathbf{X}_t\rangle|^2\right\}\leqslant\|\boldsymbol{\Phi}_1\|^2\mathbb{E}\{\|\mathbf{X}_{L,t}-\mathbf{X}_t\|^2\}$. With the fact $\|\boldsymbol{\Phi}_1\|^2<\infty$, 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 13, we have $\mathbb{E}(\|\boldsymbol{\varepsilon}_{t-l}\|) \leqslant \sqrt{p\omega_0^{\boldsymbol{\varepsilon}}}$. This together with Lemma 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 12, we have $\sum_{l=0}^{\infty}\|\mathbf{A}_l\|_{\mathrm{F}} < \infty$. This together with the above yields

$$\lim_{L\to\infty}\mathbb{E}\{\|\mathbf{X}_{L,t} - \mathbf{X}_t\|^2\} = 0, \tag{B.15}$$

which completes the proof of (B.13).

Proof of (B.14). 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

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

By Jensen's inequality and Lemma 11, we have

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

Combining the above results with (B.15), we complete the proof of (B.14). ☐

Lemma 10. *Suppose conditions in Lemma 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

$$
\begin{aligned}
&\lim_{L\to\infty} |\mathcal{M}(\boldsymbol{f}_L^X, \boldsymbol{\Phi}_1) - \mathcal{M}(\boldsymbol{f}^X, \boldsymbol{\Phi}_1)| \\
&= 2\pi \lim_{L\to\infty} \left| \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi]} |\langle \boldsymbol{\Phi}_1, \boldsymbol{f}_{L,\theta}^X(\boldsymbol{\Phi}_1)\rangle| - \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi]} |\langle \boldsymbol{\Phi}_1, \boldsymbol{f}_\theta^X(\boldsymbol{\Phi}_1)\rangle| \right| \\
&\leqslant 2\pi \lim_{L\to\infty} \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi]} \left| |\langle \boldsymbol{\Phi}_1, \boldsymbol{f}_{L,\theta}^X(\boldsymbol{\Phi}_1)\rangle| - |\langle \boldsymbol{\Phi}_1, \boldsymbol{f}_\theta^X(\boldsymbol{\Phi}_1)\rangle| \right| \\
&\leqslant \|\boldsymbol{\Phi}_1\|^2 \lim_{L\to\infty} \left\| \sum_{h\in\mathbb{Z}} (\boldsymbol{\Sigma}_{L,h}^X - \boldsymbol{\Sigma}_h^X) \right\|_{\mathrm{F}} \quad \text{(by Lemma 11 and } |\exp(-ih\theta)| = 1) \\
&\leqslant \|\boldsymbol{\Phi}_1\|^2 \lim_{L\to\infty} \sum_{h\in\mathbb{Z}} \left\| \boldsymbol{\Sigma}_{L,h}^X - \boldsymbol{\Sigma}_h^X \right\|_{\mathrm{F}}.
\end{aligned}
$$

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 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

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

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

Then,

$$
\begin{aligned}
&\lim_{L\to\infty} \left\| \boldsymbol{\Sigma}_{L,h}^X - \boldsymbol{\Sigma}_h^X \right\|_{\mathrm{F}} \\
&= \lim_{L\to\infty} \left\| \sum_{l=L-h+1}^{\infty} \int \mathbf{A}_{l+h}(u,u')\boldsymbol{\Sigma}_0^{\varepsilon}(u',v')\{\mathbf{A}_l(v,v')\}^{\mathrm{T}} du'dv' \right\|_{\mathrm{F}} \\
&\leqslant p\omega_0^{\varepsilon} \lim_{L\to\infty} \sum_{l=L-h+1}^{\infty} \|\mathbf{A}_l\|_{\mathrm{F}} \|\mathbf{A}_{l+h}\|_{\mathrm{F}} \quad \text{(by Lemmas 11 and 13)} \\
&\leqslant p\omega_0^{\varepsilon} \lim_{L\to\infty} \sum_{l=L-h+1}^{\infty} \|\mathbf{A}_l\|_{\mathrm{F}} \sum_{l=L-h+1}^{\infty} \|\mathbf{A}_{l+h}\|_{\mathrm{F}} \\
&= 0 \quad \text{(by Lemma 12)},
\end{aligned}
$$

which completes the proof. ∎

Lemma 11. *(i) Let* $\mathbf{A} = (A_{ij})_{p\times q}$ *with each* $A_{ij} \in \mathbb{S}$ *and* $\mathbf{B} = (B_1, \ldots, B_q)^{\mathrm{T}} \in \mathbb{H}^q$.

$$\left\| \int\int \mathbf{A}(u,v)\mathbf{B}(v)dudv \right\| \leqslant \|\mathbf{A}\|_F \|\mathbf{B}\|. \tag{B.16}$$

Similarly, we have

$$\left\| \int \mathbf{A}(u,v)\mathbf{B}(v)dv \right\| \leqslant \|\mathbf{A}\|_F \|\mathbf{B}\|, \tag{B.17}$$

(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*

$$\left\| \int \mathbf{A}(u,z)\mathbf{B}(z,v)dz \right\|_F \leqslant \|\mathbf{A}\|_F \|\mathbf{B}\|_F. \tag{B.18}$$

Proof of (B.16). Let $C = \int\int \mathbf{A}(u,v)\mathbf{B}(v)dudv$, then we have that $|C_i| = |\sum_k \int\int A_{ik}(u,v)B_k(v)dudv| \leqslant \sum_k \|A_{ik}\|_{\mathcal{S}}\|B_k\|$.

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

Proof of (B.17). Let $C(u) = \int \mathbf{A}(u,v)\mathbf{B}(v)dv$, then we have that $C_i(u) = \sum_k \int A_{ik}(u,v)B_k(v)dv$.

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

Proof of (B.18). 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 (B.17), we obtain

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

Lemma 12. *Suppose that conditions in Lemma 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 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

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

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

$$
\begin{aligned}
\sum_{h\in\mathbb{Z}} \|\boldsymbol{\Sigma}_h^X(u,v)\|_{\mathrm{F}} &\leqslant 2\sum_{h=0}^{\infty} \|\boldsymbol{\Sigma}_h^X(u,v)\|_{\mathrm{F}} \\
&= 2\sum_{h=0}^{\infty} \|\sum_{l=0}^{\infty} \int \mathbf{A}_{l+h}(u,u')\boldsymbol{\Sigma}_0^{\varepsilon}(u',v')\{\mathbf{A}_l(v,v')\}^{\mathrm{T}} du'dv'\|_{\mathrm{F}} \\
&\leqslant 2\sum_{h=0}^{\infty}\sum_{l=0}^{\infty} \|\int \mathbf{A}_{l+h}(u,u')\boldsymbol{\Sigma}_0^{\varepsilon}(u',v')\{\mathbf{A}_l(v,v')\}^{\mathrm{T}} du'dv'\|_{\mathrm{F}} \\
&\leqslant 2\sum_{h=0}^{\infty}\sum_{l=0}^{\infty} \|\mathbf{A}_l\|_{\mathrm{F}}\|\mathbf{A}_{l+h}\|_{\mathrm{F}}\|\boldsymbol{\Sigma}_0^{\varepsilon}\|_{\mathrm{F}} \quad \text{(by Lemma 11)} \\
&\leqslant 2p\omega_0^{\varepsilon}\left\{\sum_{l=0}^{\infty}\|\mathbf{A}_l\|_{\mathrm{F}}\right\}^2 < \infty \quad \text{(by Lemme 13)},
\end{aligned}
$$

which completes the proof. □

Lemma 13. *For a p-dimensional vector process $\{\mathbf{X}_t(\cdot)\}_{t\in\mathbb{Z}}$, whose lag-h autocovariance 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*

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

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

$$
\lim_{M\to\infty}\|\Sigma_{h,jk}^{X_M} - \Sigma_{h,jk}^{X}\|_{\mathcal{S}} = 0. \tag{B.19}
$$

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 $\|\boldsymbol{\Sigma}_h\|_{\mathrm{F}}^2 = \sum_{j,k}\|\Sigma_{h,jk}\|_{\mathcal{S}}^2 \leqslant p^2\omega_0^2$. 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 (B.19).

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 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

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

Proof. This lemma can be proved in similar way to Lemma 8 of Guo and Qiao (2021) and hence the proof is omitted here. □

Appendix C: Proofs of theoretical results in Section 3

We present the proof of Theorem 4 in Appendix C.1 and proofs of Propositions 4–7 in Appendix C.2, followed by the supporting technical lemmas and their proofs in Appendix 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 $\mathbf{B} = (\mathbf{B}_{jk})_{1\leqslant j\leqslant p_1, 1\leqslant k\leqslant p_2} \in \mathbb{R}^{p_1q\times p_2q}$ 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.

C.1. Proof of Theorem 4

Denote the minimizer of (3.3) 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 $\mathbf{\Delta} = \widehat{\mathbf{B}} - \mathbf{B}$ 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 $\langle\langle \mathbf{A}, \mathbf{B} \rangle\rangle = \text{trace}(\mathbf{A}^{\mathrm{T}}\mathbf{B})$. Then we write

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

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 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 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

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

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,

$$\begin{aligned}\|\widehat{\boldsymbol\beta}-\boldsymbol\beta\|_1 \leqslant &\sum_{h,j}\|(\widehat{\boldsymbol\psi}_j(u)-\boldsymbol\psi_j(u))^{\mathrm T}\widehat{\boldsymbol\Psi}_{hj}\widehat{\boldsymbol\phi}(v)\|_{\mathcal S} + \sum_{h,j}\|\boldsymbol\psi_j(u)^{\mathrm T}\widehat{\boldsymbol\Psi}_{hj}(\widehat{\boldsymbol\phi}(v)-\boldsymbol\phi(v))\|_{\mathcal S}\\ &+\sum_{h,j}\|\boldsymbol\psi_j(u)^{\mathrm T}(\widehat{\boldsymbol\Psi}_{hj}-\boldsymbol\Psi_{hj})\boldsymbol\phi(v)\|_{\mathcal S} + \sum_{h,j}\|R_{hj}(u,v)\|_{\mathcal S}.\end{aligned} \tag{C.2}$$

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 (C.2), 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 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,

$$\|\boldsymbol\Psi\|_1^{(q_1,q_2)} = \sum_{h,j}\|\boldsymbol\Psi_{hj}\|_{\mathrm F} = O(s). \tag{C.3}$$

By the definition of ω_0^X, Condition 5 and Proposition 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 $\widehat{\boldsymbol\Psi}-\boldsymbol\Psi = \widehat{\mathbf D}^{-1}\widehat{\mathbf B}-\mathbf D^{-1}\mathbf B = \mathbf D^{-1}(\widehat{\mathbf B}-\mathbf B)+(\widehat{\mathbf D}^{-1}-\mathbf D^{-1})\widehat{\mathbf B}$. Then

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

This, together with (C.1) implies that,

$$\|\mathbf B\|_1^{(q_1,q_2)} = O(\sqrt{\omega_0^X}s), \tag{C.4}$$

and

$$\|\widehat{\boldsymbol\Psi}-\boldsymbol\Psi\|_1^{(q_1,q_2)} \leqslant \frac{96\alpha_1^{1/2}q_1^{\alpha_1/2}s\lambda_n}{c_0^{1/2}\tau_2}\{1+o(1)\}. \tag{C.5}$$

Combining (C.3) and (C.5), we have

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

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

Combining all the results with Proposition 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. □

C.2. Proofs of propositions

Proof of Proposition 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 7, $\omega_{\min}(\mathbf{\Gamma}) \geqslant \underline{\mu}$, where $\omega_{\min}(\mathbf{\Gamma})$ denotes the minimum eigenvalue of $\mathbf{\Gamma}$. This, together with Lemma 16, completes our proof. □

Proof of Proposition 6 This proposition can be proved in similar way to Proposition 3 of Guo and Qiao (2021) and hence the proof is omitted here. □

Proof of Proposition 7 Notice that $\widehat{\mathbf{U}} = \mathbf{Z}\mathbf{D}^{-1}\widetilde{\mathbf{B}} + \widehat{\mathbf{R}} + \widehat{\mathbf{E}}$, where $\widetilde{\mathbf{B}} = \mathbf{D}\widetilde{\mathbf{\Psi}}$ and $\{(h+1)j\}$-th row block of $\widetilde{\mathbf{\Psi}}$, $\widetilde{\mathbf{\Psi}}_{hj} = \int_{\mathcal{V}}\int_{\mathcal{U}}\boldsymbol{\psi}_j(u)\beta_{hj}(u,v)\widehat{\boldsymbol{\phi}}(v)^{\mathrm{T}}dudv$. 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

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

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

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

where $\widehat{\boldsymbol{\Gamma}}=(n-L)^{-1}\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}\widehat{\mathbf{Z}}\widehat{\mathbf{D}}^{-1}$. By Lemmas 15, 17, 18 and (C.4) in Appendix C.1, there exist some positive constants C_1^*, c_1^* and c_2^* such that

$$
\|I_1\|_{\max}^{(q_1,q_2)} \leqslant C_1^* s q_1^{1/2}(\mathcal{M}_1^X q_1^{\alpha_1+3/2} \vee \mathcal{M}_1^Y q_2^{\alpha_2+3/2})\sqrt{\frac{\log(pq_1 \vee q_2)}{n}} \tag{C.6}
$$

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

By Lemma 19, we obtain that there exist some positive constants C_2^*, c_1^* and c_2^* such that

$$
\|I_2\|_{\max}^{(q_1,q_2)} \leqslant C_2^* s q_1^{-\kappa+1} \tag{C.7}
$$

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 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

$$
\begin{aligned}
\|I_3\|_{\max}^{(q_1,q_2)} &\leqslant q_1^{1/2}\|\widehat{\mathbf{D}}^{-1}\mathbf{D}\|_{\max}\|(n-L)^{-1}\mathbf{D}^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}\widehat{\mathbf{E}}\mathbf{Q}^{-1}\|_{\max}\|\mathbf{Q}\|_{\mathrm{F}} \\
&\leqslant C_3^* q_1^{1/2}(\mathcal{M}_1^X+\mathcal{M}^{\epsilon})(q_1^{\alpha_1} \vee q_2^{\alpha_2})\sqrt{\frac{\log(pq_1q_2)}{n}}
\end{aligned} \tag{C.8}
$$

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

It follows from (C.6)–(C.8) that there exist some positive constants C_0, c_1^* and c_2^* such that

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

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

C.3. Technical lemmas and their proofs

Lemma 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 16. *Suppose that Conditions 1–5 hold. Then there exist some positive constants C_Γ, c_1^* and c_2^* such that*

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

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

Proof. The proof follows from Lemma 5 in Guo and Qiao (2021). □

Lemma 17. *Suppose that Conditions 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 26, Proposition 6 and following the similar argument in the proof of Lemma 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 16, shows that

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

Lemma 18. *Suppose that Conditions 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 $\|\widetilde{\boldsymbol{\Psi}} - \boldsymbol{\Psi}\|_1^{(q_1,q_2)}$. 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 $|a_{hjlm}| \leqslant u_{hj}(l+m)^{-\kappa-1/2}$.

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

It follows from Lemma 25, for $(h,j) \in S$,

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

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

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

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

Lemma 19. *Suppose that Conditions 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 $(h,j') \in S$,

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

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 24 and 25 and following the similar argument in the proof of the upper bound of $I_{1,tmhj'}$, we obtain that, for $(h,j') \in S$,

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

Then

$$\begin{aligned}
\Big\|\frac{1}{n}\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}\widehat{\mathbf{R}}\Big\|_{\max}^{(q_1,q_2)} &\lesssim s \max_{1\leqslant j\leqslant p} \sqrt{\sum_{l=1}^{q_1}\sum_{m=1}^{q_2} \{(q_1+m)^{-2\kappa} + \|\widehat{\phi}_m - \phi_m\|^2 q_1^{-2\kappa+2}\}} \\
&\lesssim s \max_{1\leqslant j\leqslant p} \sqrt{q_1^{-2\kappa+2} + q_1^{-2\kappa+3} q_2 \max_{1\leqslant m\leqslant q_2} \|\widehat{\phi}_m - \phi_m\|^2} \\
&= O_P\{s q_1^{-\kappa+1}\}. \quad \square
\end{aligned} \tag{C.9}$$

Appendix D: Proofs of theoretical results in Section 4

This section is organized in the same manner as Appendix C. The proofs of Theorem 5 and Propositions 5–9 are presented in Appendices D.1 and D.2, respectively, followed by supporting technical lemmas and their proofs in Appendix D.3.

D.1. Proof of Theorem 5

Here $\widehat{B} \in \mathbb{R}^{pq}$ and $\widehat{\gamma} \in \mathbb{R}^d$ are the minimizer of (4.3). 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 8, 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 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

$$\begin{aligned}\|\Delta\| + \|\delta\| &\lesssim \frac{\lambda_n\sqrt{s_1+s_2}}{\tau_2^*}, \\ \|\Delta\|_1^{(q)} + \|\delta\|_1 &\lesssim \frac{\lambda_n(s_1+s_2)}{\tau_2^*}.\end{aligned}$$

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 4, which shows

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

D.2. Proofs of propositions

Proof of Proposition 5 By Lemmas 16, 20 and 28, we obtain

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

Combining this with Condition 9 and following the similar argument in the proof of Proposition 4 implies Proposition 5. $\square$

Proof of Proposition 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 17 when $L = 0$ that there exist some positive constants C_{11}^*, c_1^* and c_2^* such that

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

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

Second, it follows from Lemma 22 that there exist some positive constants C_{12}^*, c_1^* and c_2^* such that

$$\|\frac{1}{n}\widehat{\Omega}^{\mathrm{T}}R\|_{\max}^{(q)} \leqslant C_{12}^* s_1 q^{-\kappa+1},$$

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

Third, it follows from Proposition 3 that there exist some positive constants C_{13}^*, c_1^* and c_2^* such that

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

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

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

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

Proof of Proposition 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 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

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

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

Second, it follows from Lemma 23 that there exist some positive constants C_{22}^*, c_1^* and c_2^* such that

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

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

Third, it follows from Lemma 28 that there exist some positive constants C_{23}^*, c_1^* and c_2^* such that

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

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

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

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

D.3. Technical lemmas and their proofs

Lemma 20. *Suppose that Conditions 1–5 hold. Then there exist some positive constants $\widetilde{C}_{1,Z\Gamma}$, c_1^* and c_2^* such that*

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

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

Proof. Note that

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

Let $\widehat{s}_{jkl}=\Big\{\omega_{jl}^X/\widehat{\omega}_{jl}^X\Big\}^{1/2}$, 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, 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

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

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

Lemma 21. *Suppose that Conditions 1–5 hold. Then there exist some positive constants $\widetilde{C}_{2,Z\Gamma}$, c_1^* and c_2^* such that*

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

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 (B.5) in Appendix 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 20, implies that

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

Lemma 22. *Suppose that Conditions 1–5 and 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 19 and hence the proof is omitted here. □

Lemma 23. *Suppose that Conditions 1–5 and 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 19 and hence the proof is omitted here. □

Appendix E: Existing results for sub-Gaussian (functional) linear processes

For ease of reference, we present some useful existing results in Guo and Qiao (2021), including non-asymptotic error bounds on estimated covariance matrix function, estimated eigenpairs and estimated (auto)covariance between estimated FPC scores. By Theorem 1, we can easily extend these results from Gaussian functional time series to accommodate sub-Gaussian functional linear processes in Lemmas 24–27. 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 28.

Lemma 24. *Suppose that Conditions 1, 3 and 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 $j, k = 1, \ldots, p$,*

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

Proof. This lemma follows directly from Theorem 1 and Theorem 2 of Guo and Qiao (2021) and hence the proof is omitted here. □

Lemma 25. *Suppose that Conditions 1, 3, 4 and 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 $\{\widehat{\omega}_{jl}^X\}$ and $\{\widehat{\psi}_{jl}\}$ satisfy*

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

Proof. This lemma follows directly from Theorem 1 and Theorem 3 of Guo and Qiao (2021) and hence the proof is omitted here. □

Lemma 26. *Suppose that conditions in Lemma 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 $\eta > 0$,*

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

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 and Lemma 3 of Guo and Qiao (2021) and hence the proof is omitted here. □

Lemma 27. *Suppose that conditions in Lemma 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*

$$\max_{\substack{1\leqslant j,k\leqslant p\\ 1\leqslant l,m\leqslant M}}\frac{\left|\widehat{\sigma}_{h,jklm}^X - \sigma_{h,jklm}^X\right|}{(l\vee m)^{\alpha+1}\sqrt{\omega_{jl}^X\omega_{km}^X}} \lesssim \mathcal{M}_1^X\sqrt{\frac{\log(pM)}{n}}. \tag{E.2}$$

Proof. This lemma follows directly from Theorem 1 and Theorem 4 of Guo and Qiao (2021) and hence the proof is omitted here. □

Lemma 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 $1 - \tilde{c}_{16}(d)^{-\tilde{c}_{17}}$,

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

(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. □

Appendix 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 (3.1).

Acknowledgments

We are grateful to the editor, the associate editor and two referees for their insightful comments, which have led to significant improvement of our paper.

References

Aue, A., Norinho, D. D. and Hörmann, S. (2015). On the prediction of stationary functional time series. *J. Amer. Statist. Assoc.* **110** 378–392. MR3338510

Basu, S. and Michailidis, G. (2015). Regularized estimation in sparse high-dimensional time series models. *Ann. Statist.* **43** 1535–1567. MR3357870

Bathia, N., Yao, Q. and Ziegelmann, F. (2010). Identifying the finite dimensionality of curve time series. *Ann. Statist.* **38** 3352–3386. MR2766855

Bosq, D. (2000). *Linear processes in function spaces. Lecture Notes in Statistics* **149**. Springer-Verlag, New York Theory and applications. MR1783138

Chang, J., Chen, C., Qiao, X., and Yao, Q. (2021). An autocovariance-based learning framework for high-dimensional functional time series. *arXiv:2008.12885v2*.

Cho, H., Goude, Y., Brossat, X. and Yao, Q. (2013). Modeling and forecasting daily electricity load curves: a hybrid approach. *J. Amer. Statist. Assoc.* **108** 7–21. MR3174599

Fan, Y., James, G. M. and Radchenko, P. (2015). Functional additive regression. *Ann. Statist.* **43** 2296–2325. MR3396986

Fan, Y., Foutz, N., James, G. M. and Jank, W. (2014). Functional response additive model estimation with online virtual stock markets. *Ann. Appl. Stat.* **8** 2435–2460. MR3292504

Gao, Z., Ma, Y., Wang, H. and Yao, Q. (2019). Banded spatio-temporal autoregressions. *J. Econometrics* **208** 211–230. MR3906968

Ghosh, S., Khare, K. and Michailidis, G. (2019). High-dimensional posterior consistency in Bayesian vector autoregressive models. *J. Amer. Statist. Assoc.* **114** 735–748. MR3963176

Guo, S. and Qiao, X. (2021). On consistency and sparsity for high-dimensional functional time series with application to autoregressions. *arXiv:2003.11462v2*.

Guo, S., Wang, Y. and Yao, Q. (2016). High-dimensional and banded vector autoregressions. *Biometrika* **103** 889–903. MR3620446

Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.* **35** 70–91. MR2332269

Hamilton, J. D. (1994). *Time series analysis.* Princeton University Press, Princeton, NJ. MR1278033

Han, Y. and Tsay, R. S. (2020). High-dimensional linear regression for dependent data with applications to nowcasting. *Statist. Sinica* **30** 1797–1827. MR4260745

Hörmann, S., Kidziński, L. u. and Kokoszka, P. (2015). Estimation in functional lagged regression. *J. Time Series Anal.* **36** 541–561. MR3356268

Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *Ann. Statist.* **38** 1845–1884. MR2662361

Horn, R. A. and Mathias, R. (1990). Cauchy-Schwarz inequalities associated with positive semidefinite matrices. *Linear Algebra Appl.* **142** 63–82. MR1077974

Horváth, L., Kokoszka, P. and Rice, G. (2014). Testing stationarity of functional time series. *J. Econometrics* **179** 66–82. MR3153649

Kong, D., Xue, K., Yao, F. and Zhang, H. H. (2016). Partially functional linear regression in high dimensions. *Biometrika* **103** 147–159. MR3465827

Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: inference for the number of factors. *Ann. Statist.* **40** 694–726. MR2933663

Li, D., Robinson, P. M. and Shang, H. L. (2020). Long-range dependent curve time series. *J. Amer. Statist. Assoc.* **115** 957–971. MR4107692

Li, Z., Lam, C., Yao, J. and Yao, Q. (2019). On testing for high-dimensional white noise. *Ann. Statist.* **47** 3382–3412. MR4025746

Lin, J. and Michailidis, G. (2017). Regularized estimation and testing for high-dimensional multi-block vector-autoregressive models. *J. Mach. Learn. Res.* **18** Paper No. 117, 49. MR3725456

Lin, J. and Michailidis, G. (2020). Regularized estimation of high-dimensional factor-augmented vector autoregressive (FAVAR) models. *J. Mach. Learn. Res.* **21** Paper No. 117, 51. MR4119185

Loh, P.-L. and Wainwright, M. J. (2012). High-dimensional regression with noisy and missing data: provable guarantees with nonconvexity. *Ann. Statist.* **40** 1637–1664. MR3015038

Luo, R. and Qi, X. (2017). Function-on-function linear regression by signal compression. *J. Amer. Statist. Assoc.* **112** 690–705. MR3671763

Panaretos, V. M. and Tavakoli, S. (2013). Fourier analysis of stationary time series in function space. *Ann. Statist.* **41** 568–603. MR3099114

Pham, T. and Panaretos, V. M. (2018). Methodology and convergence rates for functional time series regression. *Statist. Sinica* **28** 2521–2539. MR3839872

Qiao, X., Guo, S. and James, G. M. (2019). Functional graphical models. *J. Amer. Statist. Assoc.* **114** 211–222. MR3941249

Qiao, X., Qian, C., James, G. M. and Guo, S. (2020). Doubly functional graphical models in high dimensions. *Biometrika* **107** 415–431. MR4108937

Rudelson, M. and Vershynin, R. (2013). Hanson-Wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab.* **18** no. 82, 9. MR3125258

Shu, H. and Nan, B. (2019). Estimation of large covariance and precision matrices from temporally dependent observations. *Ann. Statist.* **47** 1321–1350. MR3911114

Simon, N. and Tibshirani, R. (2012). Standardization and the group Lasso penalty. *Statist. Sinica* **22** 983–1001. MR2987480

Sun, Y., Li, Y., Kuceyeski, A. and Basu, S. (2018). Large spectral density matrix estimation by thresholding. *arXiv:1812.00532*.

Wong, K. C., Li, Z. and Tewari, A. (2020). Lasso guarantees for β-mixing heavy-tailed time series. *Ann. Statist.* **48** 1124–1142. MR4102690

Wu, W.-B. and Wu, Y. N. (2016). Performance bounds for parameter estimates of high-dimensional linear models with correlated errors. *Electron. J. Stat.* **10** 352–379. MR3466186

Xue, K. and Yao, F. (2021). Hypothesis testing in large-scale functional linear regression. *Statist. Sinica* **31** 1101–1123. MR4286208

Zhou, H. H. and Raskutti, G. (2019). Non-parametric sparse additive auto-regressive network models. *IEEE Trans. Inform. Theory* **65** 1473–1492. MR3923181