Nonparametric and high-dimensional functional graphical models

Eftychia Solea and Holger Dette
Crest-Ensai and Ruhr-Universität Bochum

Abstract

We consider the problem of constructing nonparametric undirected graphical models for high-dimensional functional data. Most existing statistical methods in this context assume either a Gaussian distribution on the vertices or linear conditional means. In this article we provide a more flexible model which relaxes the linearity assumption by replacing it by an arbitrary additive form. The use of functional principal components offers an estimation strategy that uses a group lasso penalty to estimate the relevant edges of the graph. We establish statistical guarantees for the resulting estimators, which can be used to prove consistency if the dimension and the number of functional principal components diverge to infinity with the sample size. We also investigate the empirical performance of our method through simulation studies and a real data application.

Keywords: undirected graphical models; functional data; additive models; lasso; EEG data; brain networks.

1 Introduction

In recent years, there has been a large amount of work on estimating undirected graphical models that describe the conditional dependencies among the components of a p-dimensional random vector $X = (X^1, \dots, X^p)^\intercal$. Let $\mathsf{V} = \{1, \dots, \mathsf{p}\}$, and E denote a subset of $\{(i,j) \in \mathsf{V} \times \mathsf{V} : \mathsf{i} \neq \mathsf{j}\}$, which satisfies $(i,j) \in \mathsf{E}$ if and only if $(j,i) \in \mathsf{E}$. The pair $\mathsf{G} = (\mathsf{V}, \mathsf{E})$ constitutes an undirected graph, with V representing the set of vertices and E the set of edges. The vector X follows a graphical model if

$$(i,j) \notin \mathsf{E} \quad \Leftrightarrow \quad X^i \perp\!\!\!\perp X^j | X^{-(i,j)}, \tag{1.1}$$

where $X^{-(i,j)}$ represents the vector X with its ith and jth components removed, and for random elements A, B, and C, $A \perp\!\!\!\perp B|C$ means that A and B are conditionally independent given C. The goal is to estimate the edge set E based on a random sample from X.

If X is assumed to follow a p-dimensional Gaussian distribution with expectation $0 \in \mathbb{R}^d$ and positive definite covariance matrix $\Sigma \in \mathbb{R}^{p \times p}$, the model is called *Gaussian Graphical Model (GGM)* and has become very popular. For a Gaussian random vector $X = (X^1, \dots, X^p)^\intercal$, the structure of the precision matrix $\Theta = \Sigma^{-1}$ characterizes the conditional independence relationships among the variables $X^1, \dots, X^p$ (Lauritzen, 1996). Specifically,

$$X^i \perp\!\!\!\perp X^j | X^{-(i,j)} \quad \Leftrightarrow \quad \theta_{ij} = 0, \tag{1.2}$$

where θ_{ij} is the (i, j)th entry of the precision matrix Θ. Because of the relation (1.2), the estimation of the edge set E reduces to estimating the sparsity pattern of the precision matrix Θ. Hence, there exists a large amount of literature, which has its focus on estimating high-dimensional Gaussian graphical models. For example, Meinshausen and Bühlmann (2006) introduced a neighbourhood-based approach by solving p lasso linear regression problems for each node of the graph. Yuan and Lin (2007) and Friedman et al. (2008) considered a penalized maximum likelihood approach with the lasso penalty imposed on the off-diagonal entries of the precision matrix Θ. Based on a relation between partial correlation and regression coefficient, Peng et al. (2009) proposed to estimate a sparse GGM by imposing the lasso penalty on the partial correlations. Other developments on GGM include the SCAD and the adaptive lasso penalty (Lam and Fan, 2009), the Dantzig selector (Cai et al., 2011) and hard-thresholding (Bickel and Levina, 2008).

Despite of its simplicity, the Gaussian assumption can be very restrictive in practice and statistical inference based on the Gaussian distribution might be misleading if this assumption is violated. Therefore, more recent work has its focus on considering graphical models under less restrictive assumptions. For example, Liu et al. (2009, 2012) and Xue et al. (2012) relaxed the marginal Gaussian assumption on the vertices of the graph using copula transformations, and Voorman et al. (2013) allowed the conditional means of the variables to take an additive form. Li et al. (2014) and Lee et al. (2016a) proposed a non-Gaussian graphical model based on additive conditional independence (ACI), a three-way statistical relation that captures the spirit of conditional independence.

Most of the literature discussed so far has its focus on graphical models for finite dimensional data. However, many recent applications involve functional data, such as electroencephalogram (EEG) and functional magnetic resonance imaging (fMRI) data, where each sampling unit is modelled as a realization of a stochastic process varying over a time interval. In this paper, we are interested in estimating a nonparametric and high-dimensional undirected graphical model for multivariate functional data.

In contrast to the finite dimensional case, less literature can be found on graphical models for multivariate functional data. Qiao et al. (2018) proposed the Functional Gaussian Graphical Model (FGGM) assuming that X is a multivariate Gaussian random process. Roughly speaking, they used a truncated Karhunen-Loève expansion, say of order m_n, to reduce the infinite dimensional problem to a pm_n-dimensional problem for the principal component scores. The conditional independencies of the graph define a block sparsity structure, such that the properties of the precision matrix of the scores can be used to identify the edge set using a group lasso penalty. They called this method functional glasso, or simply fglasso, and the authors showed that, when m_n approaches infinity, consistent estimation of the edge set is possible. Zhu et al. (2016) proposed a Bayesian framework under the Gaussian assumption on the random functions for the analysis of functional graphical models, while Li and Solea (2018) relaxed this assumption by extending the concept of ACI to the functional setting.

In this paper, we introduce an alternative approach to relax the Gaussian assumption in the functional graphical model. Our research is motivated by the fact that in many applications the relation between the functional principal scores is rarely linear as implied by the assumption of the FGGM. To illustrate this observation, we consider an electroencephalography (EEG) dataset that

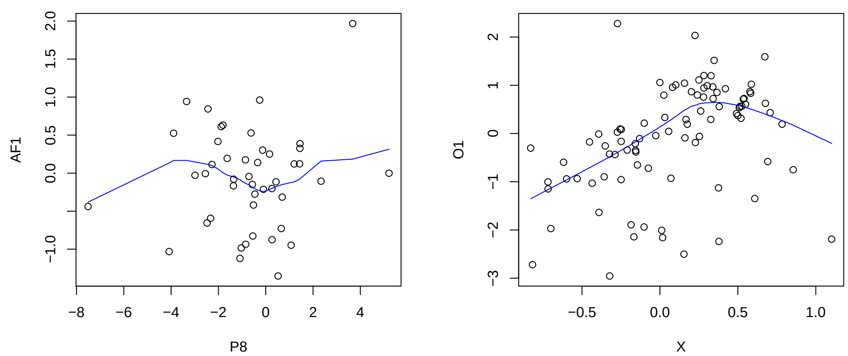

Figure 1: *Pairwise scatterplots for the control group between channels AF1 and P8 (left) and channels O1 and X (right).*

consists of two groups of subjects: 77 subjects in the alcoholic group, and 45 in the control group (Zhang et al., 1995; Ingber, 1997). For each subject, an EEG activity was recorded at 256 time points over a one second time interval using 64 electrodes placed on the subject's scalp. The goal is to construct a functional graphical model to characterize brain network connectivity for the two groups of subjects based on the functional data collected by the electrodes. In Figure 1, we display the pairwise scatterplots between channels using the first two principal components for the random functions of the control group. Clearly, this figure indicates that the conditional relationships among the scores corresponding to different vertices of the graph are nonlinear. Therefore, the Gaussian assumption of the FGGM is difficult to justify for the analysis of this type of data.

As an alternative, we propose a new nonparametric functional graphical model that allows the conditional relationships among the principal scores to take an additive structure. Our approach uses the traditional probabilistic concept of conditional independence, and then applies the additive structure to the scores of the Karhunen-Loéve expansion of each random function. We approximate each nonparametric additive component by a linear combination of B-splines basis functions. This enables us to estimate the edge set of the graph by imposing the group lasso penalty on a matrix formed by the coefficients in the spline approximation. We derive statistical guarantees for the resulting estimates, which can be used to prove consistency if the dimension p and the number of scores diverge to infinity with the sample size. This provides a useful methodology for general nonparametric analysis of high-dimensional functional graphical models. Our approach differs from related work on nonparametric functional graphical models by Lee et al. (2020) and Li and Solea (2018), who used additive conditional independence and reproducing kernel Hilbert spaces for neighbourhood selection via a functional additive regression operator.

The remainder of the article is organized as follows. Section 2 describes the methodology and proposes the nonparametric functional graphical model. Section 3 presents the estimation procedure. In Section 4 we study the theoretical properties of the resulting estimator. In Section 5 we conduct simulation studies to evaluate the finite sample properties of the proposed methodology, and in Section 6 we apply the new model to the motivating EGG dataset. We conclude with some final remarks in Section 7, while all proofs of the theoretical results are deferred to the Appendix.

2 Additive functional graphical models

We first provide a formal definition of an *additive function-on-function regression model* which will be used to define the functional graphical model considered in this paper. We begin by introducing some basic concepts from functional data analysis.

Throughout this paper $\mathcal{L}^2([0,1])$ denotes the space of all square-integrable functions defined on the interval $[0,1] \subset \mathbb{R}$. We denote by $\langle f, g\rangle = \int_{[0,1]} f(t)g(t)dt$ the common inner product in $\mathcal{L}^2([0,1])$ and by $\|f\| = \langle f, f\rangle^{1/2}$ the corresponding norm. Let $X = (X^1, \ldots, X^p)^\intercal$ denote a p-dimensional random element with mean 0 whose ith component X^i is an element of $\mathcal{L}^2([0,1])$ such that $E\|X^i\|^2 < \infty$. For each X^i, we define the corresponding covariance operator

$$\Sigma_{X^iX^i}(f)(t) = \int_T f(s)\sigma_{X^iX^i}(s,t)ds, \quad f \in \mathcal{L}^2([0,1]), \tag{2.1}$$

where $\sigma_{X^iX^i}(s,t) = \mathrm{cov}(X^i(s), X^i(t)) = E(X^i(s)X^i(t))$ is the covariance function of the random element X^i. The operator $\Sigma_{X^iX^i}$ is a compact Hilbert-Schmidt operator (see, for example, Hsing and Eubank, 2015), and there exists a spectral decomposition of the covariance function of the form

$$\sigma_{X^iX^i}(s,t) = \sum_{r=1}^{\infty} \lambda_r^i \phi_r^i(s)\phi_r^i(t), \tag{2.2}$$

where $\lambda_1^i \geq \lambda_2^i \geq \ldots$ are the eigenvalues and $\{\phi_k^i\}_{k\in\mathbb{N}}$ are orthonormal eigenfunctions satisfying

$$\int_T \sigma_{X^iX^i}(s,t)\phi_r^i(s)ds = \lambda_r^i\phi_r^i(t).$$

Consequently, each $X^i \in \mathcal{L}^2([0,1])$ can be represented by its Karhunen-Loéve expansion

$$X^i = \sum_{r=1}^{\infty} \sqrt{\lambda_r^i}\xi_r^i\,\phi_r^i \quad i = 1, \ldots, p, \tag{2.3}$$

where the random variables $\xi_r^i = \langle X^i, \phi_r^i\rangle / \sqrt{\lambda_r^i}$ are called the *functional principal component scores* and satisfy $E(\xi_r^i) = 0$, $\mathrm{var}(\xi_r^i) = 1$, $E(\xi_q^i\xi_r^i) = 0$ for $q \neq r$.

We next give our formal definition of the functional graphical model. Suppose $\mathsf{G} = (\mathsf{V}, \mathsf{E})$ is an undirected graph, where V denotes the finite set $\{1, \ldots, p\}$, and E denotes a subset of $\{(i,j) \in \mathsf{V} \times \mathsf{V} : \mathsf{i} \neq \mathsf{j}\}$, which satisfies $(i,j) \in \mathsf{E}$ if and only if $(j,i) \in \mathsf{E}$.

Definition 2.1 *A vector of random functions $X = (X^1, \ldots X^p)^\top \in \mathcal{L}^2([0,1]) \times \ldots, \times \mathcal{L}^2([0,1])$ is said to follow a functional graphical model with respect to an undirected graph $\mathsf{G} = (\mathsf{V}, \mathsf{E})$ if and only if*

$$X^i \perp\!\!\!\perp X^j | X^{-(i,j)}, \quad \forall\, (i,j) \notin \mathsf{E}.$$

Example 2.1 Qiao et al. (2018) assumed that $X = (X^1, \dots, X^p)^\top$ is a Gaussian process on $\mathcal{L}^2([0,1]) \times \dots \times \mathcal{L}^2([0,1])$ and define a Functional Gaussian Graphical model (FGGM) by the condition

$$(i,j) \notin \mathsf{E} \quad \Leftrightarrow \quad \mathrm{cov}[X^i(s), X^j(t) | X^{-(i,j)}] = 0 \quad \forall\, s,t \in [0,1]. \tag{2.4}$$

They proposed to approximate each X^i by the first m_n coefficients from the Karhunen-Loéve expansion (2.3). Thus for each X^i, one obtains a pm_n-dimensional Gaussian random vector $\xi^\top = ((\xi^1)^\top, \dots, (\xi^p)^\top)$ of scores, where $\xi^i = (\xi^i_1, \dots, \xi^i_{m_n})^\top$ is the vector of the first m_n functional principal component scores in the Karhunen-Loève expansion (2.3) of each X^i. Using the m_n-truncation Qiao et al. (2018) also showed that the FGGM can be represented as a conditional multivariate linear regression model with respect to the scores. Indeed, each ξ^i_q can be expressed as

$$\xi^i_q = \sum_{j \neq i}^{p} \sum_{r=1}^{m_n} B^{ij}_{qr} \xi^j_r + \epsilon^i_q, \quad i \in \mathsf{V}, q = 1, \dots, m_n, \tag{2.5}$$

such that $(\epsilon^i_q)_{1 \le q \le m_n}$ is uncorrelated with $(\xi^j_r)_{1 \le r \le m_n}, i \neq j$ if and only if

$$B^{ij}_n = (B^{ij}_{qr})_{1 \le q,r \le m_n} = -(\Theta^{ii}_n)^{-1} \Theta^{ij}_n, \quad (i,j) \in \mathsf{V} \times \mathsf{V}, i \neq j,$$

where $\Theta^{ij}_n \in \mathbb{R}^{m_n \times m_n}$ is the (i,j)th element of the block precision matrix $\Theta_n = (\Theta^{ij}_n)_{1 \le i,j \le p} \in \mathbb{R}^{pm_n \times pm_n}$ of the pm_n-dimensional vector ξ. Hence, under the Gaussian assumption the conditional relationships between nodes i and j are linear, and the network structure of the FGGM can also be recovered by the sparsity structure of the regression coefficient matrix B^{ij}_n. They used group-lasso penalized maximum likelihood estimation to address the blockwise sparsity of the precision matrix and showed that the precision matrix is a consistent estimate of the set E, when p and m_n approach infinity with increasing sample size. Note that the FGGM is the extension of the Gaussian graphical model of Yuan and Lin (2007) to the functional setting.

We use a generalization of the representation (2.5) to give a formal definition of the additive function-on-function model for multivariate functional data.

Definition 2.2 *Consider a vector of random functions* $X = (X^1, \dots X^p)^\top \in \mathcal{L}^2([0,1]) \times \dots, \times \mathcal{L}^2([0,1])$ *and suppose that each* X^i *has a Karhunen-Loéve expansion of the form* (2.3)*. The vector* X *follows the function-on-function additive model if for each pair* $(i,j) \in \mathsf{V} \times \mathsf{V}$ *there exists a sequence of smooth functions* $f^{ij} = \{f^{ij}_{qr} : q, r \in \mathbb{N}\}$ *defined on* $\mathbb{R}$ *with* $E[f^{ij}_{qr}(\xi^j_r)] = 0, q, r \in \mathbb{N}$*, such that*

$$E[\xi^i_q | \{\xi^j_r, j \neq i\}] = \sum_{j \neq i}^{p} \sum_{r=1}^{\infty} f^{ij}_{qr}(\xi^j_r) \tag{2.6}$$

Similar to the functional additive regression model of Han et al. (2018), model (2.6) relaxes the linearity assumption in FGGM by imposing an additive structure on the scores in the Karhunen-Loéve expansion, giving rise to a more flexible model than the FGGM. By definition, the scores ξ^i_r are uncorrelated, but we also require them to be independent in the following discussion as also postulated in Han et al. (2018). Furthermore, we assume that they take values in a closed and bounded interval $[-1, 1]$. For example, this can be achieved by taking a monotone transformation $\Psi : \mathbb{R} \rightarrow [-1, 1]$ (see Zhu et al., 2014; Wong et al., 2019).

We now define a new nonparametric functional graphical model which we call the Additive Functional Graphical Model (AFGM).

Definition 2.3 *Suppose $X = (X^1, \ldots X^p)^\top \in \mathcal{L}^2([0,1]) \times \ldots, \times \mathcal{L}^2([0,1])$ is associated with a functional graphical model $\mathsf{G} = (\mathsf{V}, \mathsf{E})$. If X is additionally a function-on-function additive model of the form* (2.6)*, then we say that X follows an additive functional graphical model, and write this statement as $X \sim \mathrm{AFGM}(\mathsf{G})$.*

The definition implies that the independence structure of X can be recovered by the sparse structure of the additive components f_{qr}^{ij} in the representation (2.6). Since each random function is infinite-dimensional, some type of regularization is needed by truncating the Karhunen-Loéve expansion (2.3) at a finite number of principal components, say m_n, where $(m_n)_{n\in\mathbb{N}}$ is a sequence converging to infinity with increasing sample size. Thus, we obtain a truncated version of model (2.6), that is

$$E[\xi_q^i | \{\xi_r^j, j \neq i; r = 1, \ldots, m_n\}] = \sum_{j\neq i}^{p} \sum_{r=1}^{m_n} f_{qr}^{ij}(\xi_r^j), \quad q = 1, \ldots, m_n, i \in \mathsf{V}. \tag{2.7}$$

Then, our goal is to estimate the "truncated" edge set

$$\mathsf{E_n} = \{(i,j) \in \mathsf{V} \times \mathsf{V} : i \neq j, f_{qr}^{ij} \neq 0 \ \text{ for some } \ q, r = 1, \ldots, m_n\} \tag{2.8}$$

Note that we aim to recover the edge set when each X^i is approximated by a finite sum of m_n terms rather than an infinite sum. Our theoretical results in Section 4 show that the edge set can be identified with probability converging to 1 as $m_n \to \infty$, $p \to \infty$ and $n \to \infty$.

Remark 2.1

(1) Note that it is not necessary to fix the sign of the eigenfunctions in the definition of scores $\xi_r^i = \langle X^i, \phi_r^i \rangle / \sqrt{\lambda_r}$ used in the representation (2.6) or (2.7), because a sign change can always be compensated by choosing the function $f_{qr}^{ij}(-x)$ instead of $f_{qr}^{ij}(x)$.

(2) Model (2.7) can be regarded as the nonparametric and additive version of the FGGM (2.5), and the generalization of the model of Voorman et al. (2013) to the functional setting where they propose a semi-parametric method which allows the conditional means of the random variables to take on an arbitrary additive structure.

3 Estimation and computation

In this section, we develop an estimation procedure for fully observed functional data to estimate the scores ξ^i for each X^i, which is used afterwards for the estimation of the edge set $\mathsf{E_n}$.

To be precise, let $X_1, \ldots, X_n$ be an independent sample from X, such that for each $u = 1, \ldots, n$, $X_u = (X_u^1, \ldots, X_u^p)^\top$ is a vector in $\mathcal{L}^2([0,1]) \times \ldots \times \mathcal{L}^2([0,1])$. Then, for each $i = 1, \ldots, p$, the covariance operator $\Sigma_{X^iX^i}$ can be estimated by

$$\hat{\Sigma}_{X^iX^i}(f)(t) = \int_{[0,1]} f(s)\hat{\sigma}_{X^iX^i}(s,t)ds, \quad f \in \mathcal{L}^2([0,1]),$$

where

$$\hat{\sigma}_{X^iX^i}(s,t) = \frac{1}{n}\sum_{u=1}^{n} X_u^i(s)X_u^i(t)$$

is the common estimator of the covariance function (note that the X^i are centered). Let $\hat{\lambda}_r^i$ and $\hat{\phi}_r^i$ be the sample eigenvalues and eigenfunctions obtained by solving the equation

$$\int_0^1 \hat{\sigma}_{X^iX^i}(s,t)\phi_r^i(s)ds = \lambda_r^i\phi_r^i(t), \quad r = 1, \ldots, m_n,$$

subject to the constraints $\langle \phi_q^i, \phi_r^i \rangle = 0$, for $q \neq r, q, r = 1, \ldots, m_n$ and $\|\phi_r^i\| = 1$. Then, the estimated scores $\hat{\xi}_{ur}^i$ are given by

$$\hat{\xi}_{ur}^i = (\hat{\lambda}_r^i)^{-1/2}\langle X_u^i, \hat{\phi}_r^i \rangle, \quad u = 1, \ldots, n, r = 1, \ldots, m_n, i \in \mathsf{V}.$$

For each $i \in \mathsf{V}$, let $\hat{\xi}_u^i = (\hat{\xi}_{u1}^i, \ldots, \hat{\xi}_{um_n}^i)^\top$ be the m_n-dimensional vector of the estimated scaled scores corresponding to the observation X_u, $u = 1, \ldots, n$. Following Huang et al. (2010) we use B-spline functions to approximate the additive components f_{qr}^{ij} in model (2.7). To be precise, let $-1 = \tau_0 < \tau_1 < \ldots < \tau_{L_n} < \tau_{L_n+1} = 1$ be an equidistant partition of the interval $[-1,1]$ into $L_n + 1$ subintervals $I_b = [\tau_b, \tau_{b+1}), b = 0, \ldots, L_n - 1$, and $I_{L_n} = [\tau_{L_n}, \tau_{L_n+1}]$.

For the number of knots we make the following assumption. Define $S_{\ell L_n}$ as the space of polynomial splines of degree $\ell \geq 1$ consisting of functions s satisfying: (i) the restriction of s to the interval I_b is a polynomial of degree ℓ for $1 \leq b \leq L_n$; (ii) for $\ell \geq 2$ and $1 \leq \ell' \leq \ell - 2$, s is a ℓ' times continuously differentiable on the interval $[-1,1]$. Then, there exists a basis of normalized B-splines functions $(h_k)_{1 \leq k \leq k_n}$ for the space $S_{\ell L_n}$, where $k_n = L_n + \ell + 1$, such that every function $s \in S_{\ell L_n}$ can be represented as

$$s(x) = \sum_{k=1}^{k_n} \beta_k h_k(x)$$

(see Schumaker (2007)). Under some smoothness conditions, the additive functions f_{qr}^{ij} can be represented by linear combinations of B-splines functions

$$f_{qr}^{ij}(x) = \sum_{k=1}^{\infty} \beta_{qrk}^{ij} h_k(x) \quad q, r = 1, 2, \ldots, \tag{3.1}$$

where the sum of squared coefficients is summable, that is

$$\sum_{k=1}^{\infty} (\beta_{qrk}^{ij})^2 < \infty. \tag{3.2}$$

By truncation of this series, we obtain the following approximation

$$f_{qr}^{ij}(x) \approx \sum_{k=1}^{k_n} h_k(x)\, \beta_{qrk}^{ij}, \quad q, r = 1, 2, \ldots, \tag{3.3}$$

where the sequence $(k_n)_{n\in\mathbb{N}}$ diverges to infinity as $n \to \infty$ (note that this can always be achieved by increasing the number of knots in the partition). Hence, the corresponding function f_{qr}^{ij} will be zero approximately if and only if $\|\beta_{qr}^{ij}\|_2^2 = 0$, where $\|\cdot\|_2$ denotes the Euclidean norm of the k_n-dimensional vector $\beta_{qr}^{ij} = (\beta_{qr1}^{ij}, \ldots, \beta_{qrk_n}^{ij})^\top$, $q, r = 1, \ldots, m_n$. Thus, to encourage sparsity we propose to minimize the criterion

$$PL_i(\beta, \hat{\xi}) = \frac{1}{2n} \sum_{q=1}^{m_n} \sum_{u=1}^{n} \Big(\hat{\xi}_{ur}^{i} - \sum_{j\neq i}^{p} \sum_{r=1}^{m_n} h^\intercal(\hat{\xi}_{ur}^{j}) \beta_{qr}^{ij} \Big)^2 + \lambda_n \sum_{j\neq i}^{p} \Big\{ \sum_{q=1}^{m_n} \sum_{r=1}^{m_n} \|\beta_{qr}^{ij}\|_2^2 \Big\}^{1/2},$$

subject to the constraint

$$\sum_{u=1}^{n} h^\intercal(\hat{\xi}_{ur}^{j}) \beta_{qr}^{ij} = 0, \quad q, r = 1, \ldots, m_n, j \in \mathsf{V}, \tag{3.4}$$

where $h^\intercal(x) = (h_1(x), \ldots, h_{k_n}(x))$ is the k_n-dimensional vector of the B-splines basis functions and λ_n is a tuning parameter. The group lasso penalty $\sum_{j\neq i}^{p} \{\sum_{q=1}^{m_n} \sum_{r=1}^{m_n} \|\beta_{qr}^{ij}\|_2^2\}^{1/2}$ enforces all regression coefficients $\beta_{qr1}^{ij}, \ldots, \beta_{qrk_n}^{ij}$ to either be all 0 or all nonzero, $q, r = 1, \ldots, m_n$. Note that the centering constraint (3.4) accounts for the fact that the function f_{qr}^{ij} in model (2.7) satisfies $E(f_{qr}^{ij}(\xi_r^j)) = 0$, for $q, r = 1, \ldots, m_n$.

This problem can be converted to an unconstrained optimisation problem by centering the basis functions. More precisely, defining

$$\tilde{h}_{nk}(\hat{\xi}_{ur}^{j}) = h_k(\hat{\xi}_{ur}^{j}) - \frac{1}{n} \sum_{u=1}^{n} h_k(\hat{\xi}_{ur}^{j}), \ k = 1, \ldots, k_n, r = 1, \ldots, m_n, j \in \mathsf{V}, \tag{3.5}$$

we consider the unconstrained optimization problem

$$\widehat{PL}_i(\beta, \hat{\xi}) = \frac{1}{2n} \sum_{q=1}^{m_n} \sum_{u=1}^{n} \Big(\hat{\xi}_{ur}^{i} - \sum_{j\neq i}^{p} \sum_{r=1}^{m_n} \tilde{h}_n^\intercal(\hat{\xi}_{ur}^{j}) \beta_{qr}^{ij} \Big)^2 + \lambda_n \sum_{j\neq i}^{p} \Big\{ \sum_{q=1}^{m_n} \sum_{r=1}^{m_n} \|\beta_{qr}^{ij}\|_2^2 \Big\}^{1/2}, \tag{3.6}$$

where

$$\tilde{h}_n(\hat{\xi}_{ur}^{j}) = (\tilde{h}_{n1}(\hat{\xi}_{ur}^{j}), \ldots, \tilde{h}_{nk_n}(\hat{\xi}_{ur}^{j}))^\intercal, \tag{3.7}$$

is the k_n-dimensional vector of the centered B-splines evaluated at the estimated scores.

Now let $\hat{\xi}^i = (\hat{\xi}_{ur}^{i})_{1\le u\le n, 1\le r\le m_n}$ be the $n \times m_n$ matrix of the estimated scores, and define

$$\tilde{\mathbf{H}}_{\mathbf{n}}^\intercal(\hat{\xi}^{-i}) = (\tilde{H}_n(\hat{\xi}^1), \ldots, \tilde{H}_n(\hat{\xi}^{i-1}), \tilde{H}_n(\hat{\xi}^{i+1}), \ldots, \tilde{H}_n(\hat{\xi}^p)) \in \mathbb{R}^{n\times(p-1)k_n m_n} \tag{3.8}$$

as the vector of matrices $\tilde{H}_n(\hat{\xi}^j) = (\tilde{h}_n^\intercal(\hat{\xi}_{ur}^{j}))_{1\le u\le n, 1\le r\le m_n} \in \mathbb{R}^{n\times k_n m_n}$. Similarly, let

$$B^i = (B^{ij}, j \in \mathsf{V} \setminus \{\mathsf{i}\}) \in \mathbb{R}^{(p-1)k_n m_n \times m_n},$$

be the vector of matrices $B^{ij} = (\beta^{ij}_{qr})_{1\le q\le m_n, 1\le r\le m_n} \in \mathbb{R}^{k_n m_n \times m_n}$, $j \neq i$. Then, following some algebraic manipulations, the objective function in (3.6) can be rewritten as

$$\widehat{PL}_i(B, \hat{\xi}) = \frac{1}{2n}\|\hat{\xi}^i - \tilde{\mathbf{H}}_{\mathbf{n}}^{\intercal}(\hat{\xi}^{-i})B^i\|_F^2 + \lambda_n \sum_{j\neq i}^{p} \|B^{ij}\|_F, \tag{3.9}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. Finally, we define $\hat{B}^i_n$ as the solution of

$$\hat{B}^i_n = \operatorname{argmin}\{\widehat{PL}_i(B, \hat{\xi}) : B \in \mathbb{R}^{(p-1)k_n m_n \times m_n}\},$$

and propose to estimate the set E_n in (2.8) by

$$\hat{\mathsf{E}}_n = \{(i,j) \in \mathsf{V}\times\mathsf{V} : i \neq j, \|\hat{B}^{ij}_n\|_F > 0 \text{ or } \|\hat{B}^{ji}_n\|_F > 0\}.$$

We summarize the algorithm below

(1) Implement FPCA to obtain the estimated scores $\hat{\xi}^i_{ur}$ of each observation X^i_u and then transform the scores into the range $[-1, 1]$ using a monotone transformation. Choose m_n such that 90% of the total variation is explained.

(2) For a given λ_n and for each $i \in \mathsf{V}$ solve the optimisation problem (3.9) using, for example, distance convex programming techniques, to find a sparse estimate of B^i.

(3) Declare that there is an edge between node i and node j if and only if either $\|\hat{B}^{ij}_n\|_F^2$ or $\|\hat{B}^{ji}_n\|_F^2$ are not zero.

4 Statistical guarantees

In this section we study the theoretical properties of the proposed estimator of the graph structure of the AFGM where we allow the number of nodes p to diverge to infinity with increasing sample size. A particular technical challenge in deriving the asymptotic theory consists in the fact that the additive structure is applied to the unobserved variables ξ^i_{ur}, and the estimator $\hat{B}^i_n$ obtained from minimizing (3.9) is based on the estimated scores. Thus, the error in these estimated coefficients must be taken into account for the analysis of the procedure.

We begin by introducing some notation. For any two positive sequences of real numbers $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$, we write $a_n \lesssim b_n$ if $a_n \le K_1 b_n$ for some constant $0 < K_1 < \infty$ which does not depend on n. We use the notation $a_n \asymp b_n$ representing the property $A \le \inf_n|\frac{a_n}{b_n}| \le \sup_n|\frac{a_n}{b_n}| \le B$, for positive constants A and B. Moreover, given a matrix $A = (a_{ij})_{1\le i\le M_1, 1\le j\le M_2} \in \mathbb{R}^{M_1\times M_2}$, we use $\|A\|_F$ for the Frobenious norm and $\|A\|_2$ for the operator norm. Finally, for any two symmetric matrices A and B, we use the notation $A \preceq B$ to denote the property that the matrix $B - A$ is nonnegative definite.

Let

$$B^{*i}_{m_n} = (B^{*ij}_{m_n}, j \in \mathsf{V}\setminus\{\mathsf{i}\}), \tag{4.1}$$

with $B^{*ij}_{m_n} = \{\beta^{*ij}_{qrk} : 1 \le q, r \le m_n, k \in \mathbb{N}\}$ be the true population matrix of parameters for the optimal prediction, defined by

$$B^{*i}_{m_n} = \text{argmin}_{\beta^{ij}_{qrk}} \sum_{q=1}^{m_n} E\Big[\xi^i_q - \sum_{j\neq i}^{p}\sum_{r=1}^{m_n}\sum_{k=1}^{\infty}(h_k(\xi^j_r) - E(h_k(\xi^j_r)))\,\beta^{ij}_{qrk}\Big]^2,$$

where $h_k(\cdot)_{k\ge 1}$ are the B-splines functions used in the representation (3.1). We define the truncated neighbourhood N^i_n of each node $i \in \mathsf{V}$ by

$$\mathsf{N}^i_\mathsf{n} = \{j \in \mathsf{V} \setminus \{i\} : \|B^{*ij}_{m_n}\|_F > 0\}$$

(note that $\|B^{*ij}_{m_n}\|_F < \infty$ by assumption (3.2)). Using this representation and observing the expansion (3.1) of f^{ij}_{qr}, the edge set E_n defined in (2.8) can be rewritten as

$$\mathsf{E}_\mathsf{n} = \{(i,j) \in \mathsf{V}\times\mathsf{V} : i \neq j, i \in \mathsf{N}^j_\mathsf{n} \text{ or } j \in \mathsf{N}^i_\mathsf{n}\}. \tag{4.2}$$

Let

$$f^{ij}_{qr}(\xi^j_r) = \sum_{k=1}^{\infty}\beta^{*ij}_{qrk}h_k(\xi^j_r) = \sum_{k=1}^{\infty}\beta^{*ij}_{qrk}\tilde{h}_k(\xi^j_r)$$

(for the second equality we use the fact that $E[f^{*ij}_{qr}(\xi^j_r)] = 0$), where the functions $\tilde{h}_k$ are defined by

$$\tilde{h}_k(\xi^j_r) = h_k(\xi^j_r) - E(h_k(\xi^j_r)). \tag{4.3}$$

We obtain from (2.7) the representation

$$\xi^i_q = \sum_{j\in\mathsf{N}^i_\mathsf{n}}\sum_{r=1}^{m_n} f^{ij}_{qr}(\xi^j_r) + \epsilon^i_q, \quad q = 1,\dots,m_n, i = 1,\dots,p, \tag{4.4}$$

where $\epsilon_q = \xi^i_q - E[\xi^i_q|\{\xi^j_r, j \neq i; r = 1,\dots,m_n\}]$. Thus, the best predictor of ξ^i_q is an additive function of the scores in the set of neighbours N^i_n of the node i only.

Let $\tilde{h}(\xi^j_r) = (\tilde{h}_1(\xi^j_r),\dots,\tilde{h}_{k_n}(\xi^j_r))^\intercal \in \mathbb{R}^{k_n}$ be the vector of the centered k_n B-splines evaluated at the unobserved scaled scores ξ^j_r, $r = 1,\dots,m_n$, $j \in \mathsf{V}$, and define the $1\times k_n m_n$ and $1 \times n^i k_n m_n$ vectors

$$\begin{aligned}\tilde{H}(\xi^j) &= (\tilde{h}^\intercal(\xi^j_r))_{1\le r\le m_n},\\ \tilde{\mathbf{H}}^\intercal(\xi^{\mathsf{N}^i_\mathsf{n}}) &= (\tilde{H}(\xi^j), j \in \mathsf{N}^i_\mathsf{n}),\end{aligned} \tag{4.5}$$

where $1 \le n^i \le p$ is the cardinality of the set N^i_n. Finally, we introduce the matrices

$$\Sigma^*_{\mathsf{N}^i_\mathsf{n}\mathsf{N}^i_\mathsf{n}} = E\Big(\tilde{\mathbf{H}}(\xi^{\mathsf{N}^i_\mathsf{n}})\tilde{\mathbf{H}}^\intercal(\xi^{\mathsf{N}^i_\mathsf{n}})\Big) \in \mathbb{R}^{n^i k_n m_n \times n^i k_n m_n} \tag{4.6}$$

and

$$\Sigma^*_{\xi^j\mathsf{N}^i_\mathsf{n}} = E\Big(\tilde{H}^\intercal(\xi^j)\tilde{\mathbf{H}}^\intercal(\xi^{\mathsf{N}^i_\mathsf{n}})\Big) \in \mathbb{R}^{k_n m_n \times n^i k_n m_n}. \tag{4.7}$$

Let

$$B^{*i}_n = (B^{*ij}_{m_n k_n}, j \in \mathsf{V}\setminus\{\mathsf{i}\}) \in \mathbb{R}^{(p-1)k_n m_n \times m_n}$$

denote the truncated version of the population matrix defined in (4.1), where

$$B^{*ij}_{m_nk_n} = \{\beta^{*ij}_{qrk} : 1 \le q, r \le m_n, 1 \le k \le k_n\}, \tag{4.8}$$

and define

$$B^{*\mathsf{N}^i_\mathsf{n}}_n = (B^{*ij}_{m_nk_n}, j \in \mathsf{N}^i_\mathsf{n}) \in \mathbb{R}^{n^i k_n m_n \times m_n}. \tag{4.9}$$

Recalling that $\hat{B}^i_n = (\hat{B}^{ij}, j \in \mathsf{V} \setminus \{\mathsf{i}\}) \in \mathbb{R}^{(p-1)k_nm_n\times m_n}$ is the solution to the minimization problem (3.9). The estimated neighbourhood for each node $i \in \mathsf{V}$ is defined

$$\hat{\mathsf{N}}^i_\mathsf{n} = \{j \in \mathsf{V} \setminus \{i\} : \|\hat{B}^{ij}_n\|_F > 0\},$$

which yields an alternative representation of the estimated edge set

$$\hat{\mathsf{E}}_\mathsf{n} = \{(i, j) \in \mathsf{V} \times \mathsf{V} : \mathsf{i} \in \hat{\mathsf{N}}^j_\mathsf{n} \text{ or } \mathsf{j} \in \hat{\mathsf{N}}^i_\mathsf{n}\}. \tag{4.10}$$

For the statement of our theoretical results we require several assumptions. Assumption 4.1 is a similar assumption as made by Qiao et al. (2018) and refers to the eigensystem of the covariance operator defined in (2.1).

Assumption 4.1 .
(i) There exist positive constants d_0, d_1 and d_2 such that

$$d_0 r^{-\beta} \le \lambda^i_r \le d_1 r^{-\beta}, \quad \lambda^i_r - \lambda^i_{r+1} \ge d_2^{-1} r^{-1-\beta} \quad \text{for } r \ge 1,$$

and for some $\beta > 1$.
(ii) The number of principal component scores m_n satisfies $m_n \asymp n^\alpha$ for some constant $\alpha \in [0, \frac{1}{2+3\beta})$.
(iii) The eigenfunctions ϕ^i_r of the covariance operator defined in (2.2) are continuous and satisfy

$$\max_{j\in\mathsf{V}} \sup_{s\in[0,1]} \sup_{r\in\mathbb{N}} |\phi^j_r(s)| \le C < \infty.$$

The next two conditions refer to the smoothness of the functions f^{ij}_{qr} in model (2.6). To be precise, let κ be a nonnegative integer and let $\rho \in (0, 1]$. We define $\mathscr{F}_{\kappa,\rho}$ as the Hölder space of functions $f : [0, 1] \to \mathbb{R}$ whose κth derivative exists and satisfies a Lipschitz condition of order ρ, and additionally satisfy the condition

$$\|f\|_\infty = \sup_{x\in[0,1]}|f(x)| \le F \tag{4.11}$$

for some $F > 0$.

Assumption 4.2 *Let $d = \kappa + \rho > 0.5$ and assume $f^{ij}_{qr} \in \mathscr{F}_{\kappa,\rho}$ and $E[f^{ij}_{qr}(\xi^j_{ur})] = 0$, for all $q, r = 1, \dots, m_n$ and $(i, j) \in \mathsf{V} \times \mathsf{V}$.*

Assumption 4.3 *The joint density function, say p^j, of the random vector $\xi^j = (\xi^j_1, \dots, \xi^j_{m_n})^\top$ is bounded away from zero and infinity on $[0, 1]^{m_n}$ for every $j = 1, \dots, p$.*

In order to derive graph estimation consistency, we make the following assumption about the errors $\epsilon_1^i, \ldots, \epsilon_{m_n}^i$ in model (2.7). A similar condition was also postulated by Voorman et al. (2013) for joint additive models in the multivariate setting.

Assumption 4.4 *There exists a constant $C > 0$ such that $P(|\epsilon_q^i| > x) \leq 2\exp(-Cx^2)$ for all $x \geq 0$ and $q = 1, \ldots, m_n, i \in \mathsf{V}$.*

Assumption 4.5 *$k_n = O(n^\nu)$ for some $\nu > 0$, where $\nu \leq \frac{\alpha(2+3\beta)}{2d-4}$ if $d \geq 2$.*

Assumption 4.6 *(Sparsity) There exists a constant $\theta > 0$ such that for all $i \in \mathsf{V}$*

$$\sum_{j \in \mathsf{N}_\mathsf{n}^i} \|B_{m_n k_n}^{*ij}\|_F < \theta.$$

Assumption 4.7 *(Bounded eigenspectrum) The minimum eigenvalue $\Lambda_{\min}(\Sigma^*_{\mathsf{N}_\mathsf{n}^i \mathsf{N}_\mathsf{n}^i})$ of the matrix $\Sigma^*_{\mathsf{N}^i \mathsf{N}^i}$ defined in (4.6) satisfies*

$$\Lambda_{min}(\Sigma^*_{\mathsf{N}_\mathsf{n}^i \mathsf{N}_\mathsf{n}^i}) > C_{\min}$$

for some constant $C_{\min} > 0$.

Assumption 4.8 *(Irrepresentable condition) There exists a constant $0 < \eta \leq 1$ such that*

$$\max_{j \notin \mathsf{N}_\mathsf{n}^i} \|\Sigma^*_{\xi^j \mathsf{N}_\mathsf{n}^i} (\Sigma^*_{\mathsf{N}_\mathsf{n}^i \mathsf{N}_\mathsf{n}^i})^{-1}\|_F \leq \frac{1-\eta}{\sqrt{n^i}}. \tag{4.12}$$

Assumption 4.7 states that the minimum eigenvalue of the population matrix $\Sigma_{\mathsf{N}_\mathsf{n}^i \mathsf{N}_\mathsf{n}^i}$ is bounded away from 0. Assumption 4.8 is the classical irrepresentable condition which is necessary and sufficient to show model selection consistency of the group lasso (Yuan and Lin, 2006; Bach, 2008). According to Meinshausen et al. (2009) if the irrepresentable condition is relaxed, the lasso selects the correct non-zero coefficients but it may select some additional zero components. Ravikumar et al. (2009) and Obozinski et al. (2011) considered similar assumptions for sparse additive and for high-dimensional models, respectively. We now state our main theoretical result for the estimator $\hat{\mathsf{N}}_n^i$ of the neighbourhood corresponding to the node $i \in \mathsf{V}$.

Theorem 4.1 *Suppose that Assumptions 4.1 - 4.8 are satisfied and the regularization parameter λ_n satisfies for all i*

$$\frac{n^i m_n^{3/2}}{k_n^d \sum_{j \in \mathsf{N}_\mathsf{n}^i} \|B_{m_n k_n}^{*ij}\|_F} \lesssim \lambda_n \lesssim (n^i)^{-3/2} (b_n^{*i})^3 (\sum_{j \in \mathsf{N}_\mathsf{n}^i} \|B_{m_n k_n}^{*ij}\|_F)^{-2}, \tag{4.13}$$

where $b_n^{*i} = \min_{j \in \mathsf{N}_\mathsf{n}^i} \|B_{m_n k_n}^{*ij}\|_F$. Then,

$$P\Big(\hat{\mathsf{N}}_n^i \neq \mathsf{N}_n^i\Big) \lesssim \exp\Big(-C_1 \frac{n^{1-\alpha(2+3\beta)} (\lambda_n \sum_{j \in \mathsf{N}_\mathsf{n}^i} \|B_{m_n k_n}^{*ij}\|_F)^2}{n^i m_n^2 k_n^4} + 2\log(p m_n k_n)\Big),$$

where $C_1 > 0$.

The proof of Theorem 4.1 is complicated and given in the Appendix. A major difficulty consists in the fact that the objective function (3.9) is based on the estimated scores and one has to establish concentration bounds in the estimation of the sample design matrix $\Sigma^{n}_{\mathsf{N}^i_n \mathsf{N}^i_n}$ using the estimated scores, rather than the true scores (stated as Theorem 8.1 in the Appendix).

Recalling the representations (4.2) and (4.10) for the edge set E_n and its estimate $\hat{\mathsf{E}}_n$ respectively. Using the union bound of probability and Theorem 4.1 we obtain the following result.

Corollary 4.1 *If the assumptions of Theorem 4.1 are satisfied, we have for a positive constant* $C_1 > 0$

$$P(\hat{\mathsf{E}}_n \neq \mathsf{E}_n) \lesssim \exp\Big(- C_1 \frac{n^{1-\alpha(2+3\beta)} (\lambda_n \min_{i=1}^{p} \sum_{j \in \mathsf{N}^i_n} \|B^{*ij}_{m_n k_n}\|_F)^2}{p m_n^2 k_n^4} + 2 \log(p m_n k_n) \Big).$$

Remark 4.1

(a) Under the assumptions of Theorem 4.1 it follows that

$$P(\hat{\mathsf{E}}_n \neq \mathsf{E}_n) \to 0,$$

if $n \to \infty, p \to \infty$ and

$$\frac{n^{1-\alpha(2+3\beta)} (\lambda_n \min_{i=1}^{p} b_n^{*i})}{n^i m_n^2 k_n^4} \to \infty \quad \text{and} \quad \frac{n^i m_n^2 k_n^4 \log(p m_n k_n)}{\lambda_n^2} = o\big(\min_{i=1}^{p} b_n^{*i}\big). \tag{4.14}$$

For example, if $m_n = O(n^{\alpha})$ with $\alpha \in [0, \frac{1}{2+3\beta})$, $k_n = O(n^{\nu})$ and $\max_{i \in \mathsf{V}} n^i = O(n^{\theta})$ with $0 \leq \theta < 1$, then, (4.14) reduces to

$$\frac{\log(p m_n k_n)}{n^{1-(\alpha(4+3\beta)+\theta+4\nu)} \lambda_n} = o\big(\min_{i=1}^{p} b_n^{*i}\big).$$

(b) In the case of scalar data ($\alpha = 0$), the conditions of Theorem 4.1 will be implied by

$$\frac{n^i}{k_n^d} \lesssim \lambda_n b_n^{*i}, \quad \lambda_n (n^i)^{3/2} \lesssim (b_n^{*i})^3 \quad \text{and} \quad \sqrt{\frac{n^i k_n^4 \log(p k_n)}{n \lambda_n^2}} = o(b_n^{*i}) \ ,$$

which are similar to the assumptions made in Ravikumar et al. (2009) and Voorman et al. (2013) for the analysis of scalar data by sparse additive models.

5 Finite sample properties

5.1 Simulated data

In this section we investigate the finite sample performance of the proposed model (AFGM) by means of a simulation study. We also compare the new methodology with the functional additive precision operator (FAPO) of Li and Solea (2018) and the FGGM of Qiao et al. (2018), where we consider two scenarios: nonlinear dependence and linear dependence.

Given an edge set E of a directed acyclic graph, we generate functional data by the model

$$X_u^i(t_s) = \sum_{(i,j)\in\mathsf{E}} \sum_{q=1}^{5} \sum_{r=1}^{5} f_{qr}^{ij}(\xi_{ur}^j)\phi_q(t_s) + \epsilon_{us}^i, \quad u = 1,\dots,n, \tag{5.1}$$

where $\phi_1^i,\dots,\phi_5^i$ are the first 5 functions of the orthonormal Fourier basis, and the errors ϵ_{us}^i form an i.i.d. sample from a $\mathcal{N}(0, 0.5^2)$ distribution. In all simulation experiments in this section, this data is smoothed to obtain continuous functions X_u^i using 10 B-spline basis functions of order 4; that is, piecewise polynomials of degree 3. As a consequence the scores satisfy a structural equation of the form

$$\xi_{uq}^i = \sum_{(i,j)\in\mathsf{E}} \sum_{r=1}^{5} f_{qr}^{ij}(\xi_{ur}^j) + \tilde{\epsilon}_{uq}^i, \quad u = 1,\dots,n, q = 1,\dots,5 \tag{5.2}$$

(see Pearl, 2002), where the errors $\tilde{\epsilon}_{uq}^i$ form an i.i.d. sample a centred normal distribution. For simplicity we assume $f_{qr}^{ij}(x) = f(x)$ for all $q, r = 1,\dots$ and for all $(i,j) \in E$. In all examples, we center $f(\xi_{ur}^j)$ to have 0 mean, and we generated $n = 100$ functions observed at 100 equally spaced time points $0 = t_1,\dots,t_{100} = 1$.

We consider directed acyclic graphs with $p = 100$ nodes so that 1% of pairs of vertices are randomly selected as edges. Then, we moralized the directed graph in order to obtain the undirected graph. We choose $m_n = 5$ functional principal components scores so that at least 90% of the total variation is explained. Furthermore, we approximate each additive function using B-splines of order 4. For simplicity we choose the same spline functions for all $j = 1,\dots,p$ and for all $r = 1,\dots,m_n$. For the choice of k_n, we follow Meier et al. (2009) and take $k_n = 4 + \lceil\sqrt{n}\rceil$.

For each scenario, we produce the average ROC curves (over 50 replications) for a range of 50 tuning parameters for the 3 functional graphical models estimators. To draw the curves, we compute for different regularization parameters λ the positive rate (sensitivity) and false positive rate (1-specificity) which are defined as

$$\text{TP} = \frac{\sum_{1\le j<i\le p} I\{(i,j) \in \mathsf{E}_\mathsf{n}, (\mathsf{i},\mathsf{j}) \in \hat{\mathsf{E}}_\mathsf{n}\}}{\sum_{1\le j<i\le p} I\{(i,j) \in \mathsf{E}_\mathsf{n}\}}, \text{FP} = \frac{\sum_{1\le j<i\le p} I\{(i,j) \notin \mathsf{E}_\mathsf{n}, (\mathsf{i},\mathsf{j}) \in \hat{\mathsf{E}}_\mathsf{n}\}}{\sum_{1\le j<i\le p} I\{(i,j) \notin \mathsf{E}_\mathsf{n}\}}.$$

5.1.1 Scenario 1: nonlinear models

We use the following nonlinear models, where the linearity assumption (2.5) does not hold. We first consider the following model, Model I, used in Zhu et al. (2014)

$$\text{Model I:} \qquad f(x) = 1.4 + 3x - \frac{1}{2} + \sin(2\pi(x - \frac{1}{2})) + 8(x - \frac{1}{3})^2 - \frac{8}{9}.$$

For the choice of scores in (5.2), we simulate ξ_{ur}^i independently from the uniform distribution $U[-1, 1]$ for all $r = 1,\dots,m_n, i \in \mathsf{V}, u = 1,\dots,n$. Furthermore, the errors ϵ_{uq}^i in (5.2) form an i.i.d. sample from the normal distribution with variance 0.1, that is $\mathcal{N}(0, 0.5^2)$.

The second example was considered in Meier et al. (2009)

$$\text{Model II:} \qquad f(x) = -\sin(2x) + x^2 - 25/12 + x + \exp(-x) - 2/5 \cdot \sinh(5/2).$$

The scores ξ^i_{ur} were simulated independently from the uniform distribution $U[-2.5, 2.5]$ for all $r = 1, \dots, m_n, i \in \mathsf{V}, u = 1, \dots, n$, and the errors ϵ^i_{uq} in (5.2) were simulated from the normal distribution $\mathcal{N}(0, 1)$.

The left and middle panel of Figure 1 show the averaged ROC curves over 40 replications corresponding to the two models. In first two lines Table 1, we report the means and standard deviations (in parentheses) of the associated area-under-curve (AUC) values. An AUC close to 1 means a better performance for the estimator. We observe from the plots in Figure 1 and from Table 1, that for the AFGM estimator the areas under the ROC are substantially larger than for the FGGM, indicating that our new method AFGM dominates the FGGM. Similarly, the AFGM performs better than FAPO, indicating the benefit of a sparse and high-dimensional scheme.

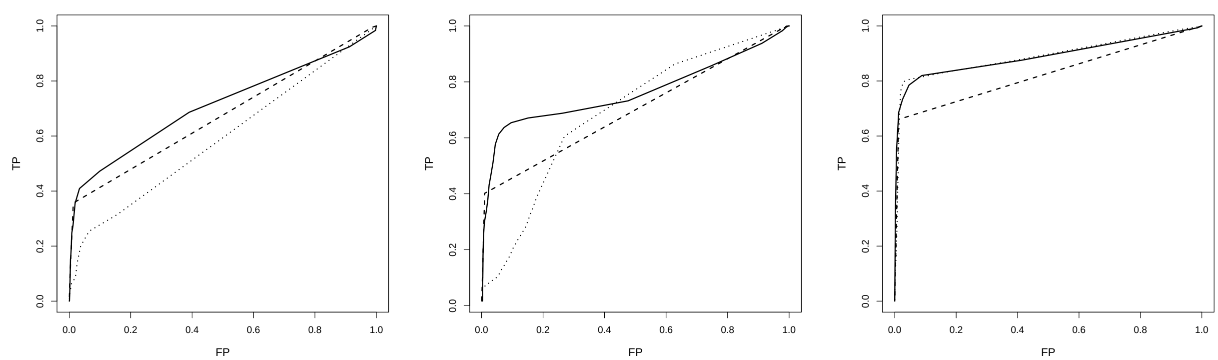

Figure 1. ROC curves ((AFGM (−), FAPO (− − −), FGGM (· · ·)) for Model I (left) and Model II (middle) and Model III (right).

p	Models	Methods		
		AFGM	FAPO	FGGM
100	I	0.73 (0.02)	0.67 (0.02)	0.59 (0.01)
	II	0.76 (0.01)	0.70 (0.02)	0.69 (0.02)
	III	0.89 (0.01)	0.82 (0.01)	0.90 (0.01)

Table 1. Means and standard errors (in parentheses) for AUC for models I and II.

5.1.2 Scenario 2: linear model.

Next, we consider a model, where the linearity assumption is satisfied, to see how much efficiency might be lost by employing a nonparametric model under the Gaussian assumption. The model is generated by (5.2) and (5.1) with

$$\text{Model III:} \qquad f(x) = x$$

The scores ξ_{ur}^i were simulated independently from the standard Gaussian distribution. To implement the AFGM we truncate the scores such that they are located in the interval $[-1, 1]$. The right panel in Figure 1 presents the averaged ROC curves for Model III. The lower part of Table 1 reports the means and standard deviations of AUC. We can see that under the linearity assumption the AFGM is comparable with the FGGM and both methods show an improvement compared to FAPO.

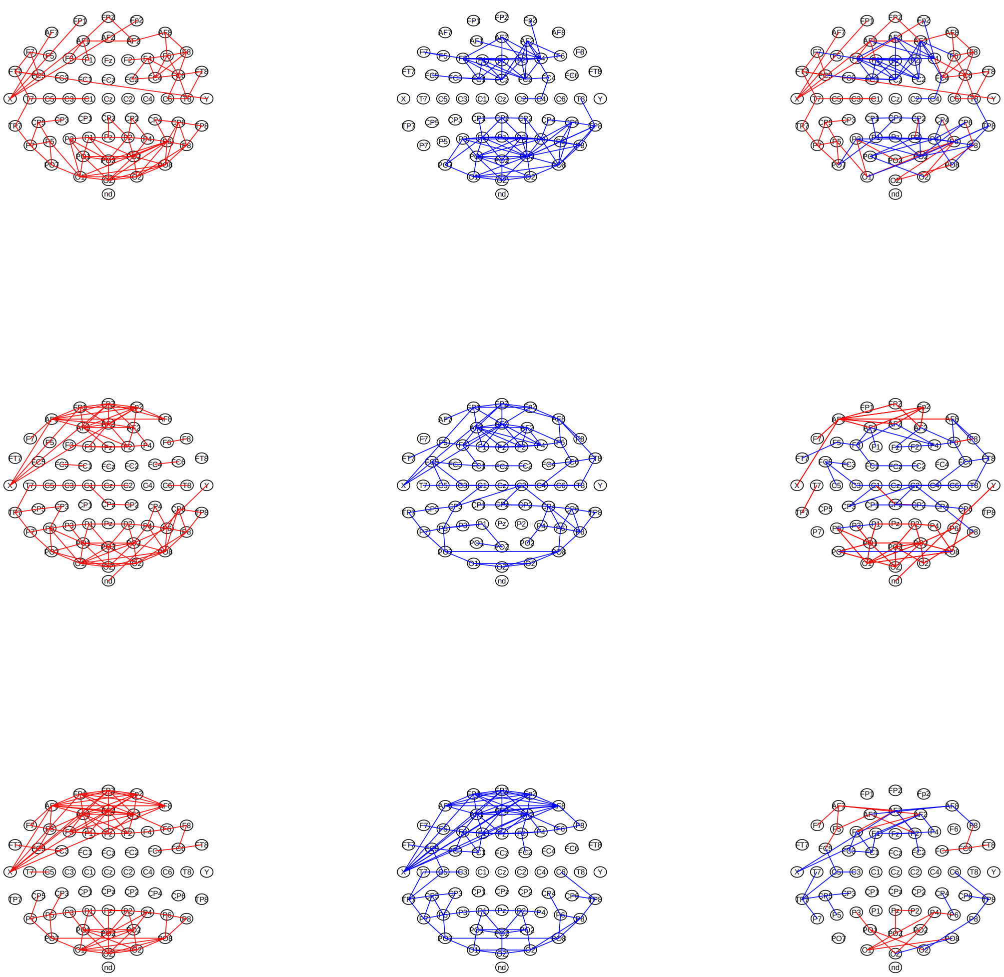

Figure 2. Estimated brain networks by AFGM (upper panels), FAPO (middle panels) and FGGM (lower panels) for the alcoholic group (left), the control group (middle) and differential brain networks (right).

6 Real data application

In this section we apply the new method to the EEG data set available at UCI Machine Learning Repository. The data involve 77 subjects in the alcoholic group and 45 subjects in the control group.

Each subject was exposed to a stimulus while brain activities were recorded from the 64 electrodes placed on the subject's scalp, over a one-second period in which 256 time points were sampled. See Zhang et al. (1995) and Ingber (1997) for more backgrounds of this data. The goal is to characterize functional connectivity among the 64 nodes for the two groups, based on the functional data collected from the electrodes.

We choose $k_n = 4 + \lceil \sqrt{n} \rceil$ B-spline functions of order 4 and number of scores equal to $m_n = 5$. Since our goal is to capture outstanding differences in brain connectivity between the alcoholic and control groups, we take the tuning constant λ_n to be such that 5% of the $\binom{64}{2}$ pairs of vertices are retained as edges.

Figure 2 shows the estimated brain networks constructed by the three methods for the alcoholic group (left), control group (middle). The right plots in Figure 2 represent the differential brain networks, where the red lines indicate the edges that are in the alcoholic network but not in the control network, and the blue lines indicate the edges that are in the control network but not in the alcoholic network.

We observe that the brain networks have different patterns for the two groups. For example, we observe for all methods, that there is increased functional connectivity in the left frontal area for the alcoholic group relative to the control.

7 Conclusions

In this paper, we utilise the idea of generalized additive models to develop a new nonparametric graphical model for multivariate functional data which does not require the assumption of a Gaussian distribution. The conditional relationships among the principal scores in the Karhunen-Loéve expansion of a random function are allowed to take an arbitrary additive rather than a linear form as imposed by the assumption of Gaussianity. The additive functions are then approximated by linear combinations of B-splines. This approximation allows us to develop a group lasso algorithm to estimate the graph that encourages blockwise sparsity to a matrix formed by the coefficients in the spline approximation. We have established consistency of the procedure while both the number of principal components and the number of nodes diverge to infinity with increasing sample size. By simulation study and an analysis of a data example we demonstrate the applicability of the new methodology.

The proposed model and methodology suggests many directions for future research. First, the asymptotic results here are developed under the framework where the order m_n in the expansion (2.7) tends to infinity with the sample size, and it is of interest if similar statistical guarantees can be obtained in model (2.6) with the infinite representation. Second, the theory is developed under the assumption that the random functions are fully observed. Therefore, an interesting and important question for future research is the extension of the methodology to smooth functions that are observed on a dense time grid such that the covariance operators and the functions are consistently estimated. Another important direction is the consideration of the sparse setting, where the functions are observed on a relatively small number of time points and are contaminated with noise. In this case, alternative approaches such as Yao et al. (2005), Xiao et al. (2018) and Petrovich et al. (2018) might be useful and will be further investigated in the future.

Acknowledgements This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Teilprojekt A1,C1) of the German Research Foundation (DFG). The authors would also like to thank Martina Stein who typed parts of this manuscript with considerable technical expertise.

References

Bach, F. R. (2008), "Consistency of the group lasso and multiple kernel learning," *Journal of Machine Learning Research*, 9, 1179–1225.

Bickel, P. J. and Levina, E. (2008), "Covariance regularization by thresholding," *Annals of Statistics*, 36, 2577–2604.

Bosq, D. (2012), *Linear Processes in Function Spaces: Theory and Applications*, vol. 149, Springer Science & Business Media.

Boucheron, S., Lugosi, G., and Massart, P. (2013), *Concentration Inequalities: A Nonasymptotic Theory of Independence*, Oxford University Press.

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

De Boor, C. (1978), *A Practical Guide to Splines*, vol. 27, springer-verlag New York.

Friedman, J., Hastie, T., and Tibshirani, R. (2008), "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, 9, 432–441.

Han, K., Müller, H.-G., Park, B. U., et al. (2018), "Smooth backfitting for additive modeling with small errors-in-variables, with an application to additive functional regression for multiple predictor functions," *Bernoulli*, 24, 1233–1265.

Hoeffding, W. (1963), "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, 58, 13–30.

Hsing, T. and Eubank, R. (2015), *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*, John Wiley & Sons.

Huang, J., Horowitz, J. L., and Wei, F. (2010), "Variable selection in nonparametric additive models," *Annals of Statistics*, 38, 2282–2313.

Ingber, L. (1997), "Statistical mechanics of neocortical interactions: Canonical momenta indicators of electroencephalography," *Physical Review E*, 55, 4578.

Lam, C. and Fan, J. (2009), "Sparsistency and rates of convergence in large covariance matrix estimation," *Annals of Statistics*, 37, 4254.

Lauritzen, S. L. (1996), *Graphical Models*, Oxford University Press.

Lee, K.-Y., Li, B., and Zhao, H. (2016a), "On an additive partial correlation operator and nonparametric estimation of graphical models," *Biometrika*, 103, 513–530.

— (2016b), "Variable selection via additive conditional independence," *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 78, 1037–1055.

Lee, K.-Y., Li, L., Li, B., and Zhao, H. (2020), "Nonparametric Functional Graphical Modeling through Functional Additive Regression Operator," *Working paper*.

Li, B., Chun, H., and Zhao, H. (2014), "On an Additive Semigraphoid Model for Statistical Networks With Application to Pathway Analysis," *Journal of the American Statistical Association*, 109, 1188–1204.

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

Liu, H., Han, F., Yuan, M., Lafferty, J., Wasserman, L., et al. (2012), "High-dimensional semiparametric Gaussian copula graphical models," *The Annals of Statistics*, 40, 2293–2326.

Liu, H., Lafferty, J., and Wasserman, L. (2009), "The nonparanormal: Semiparametric estimation of high dimensional undirected graphs," *Journal of Machine Learning Research*, 10, 2295–2328.

Meier, L., Van de Geer, S., Bühlmann, P., et al. (2009), "High-dimensional additive modeling," *Annals of Statistics*, 37, 3779–3821.

Meinshausen, N. and Bühlmann, P. (2006), "High-dimensional graphs and variable selection with the lasso," *Annals of Statistics*, 1436–1462.

Meinshausen, N., Yu, B., et al. (2009), "Lasso-type recovery of sparse representations for high-dimensional data," *Annals of Statistics*, 37, 246–270.

Obozinski, G., Wainwright, M. J., Jordan, M. I., et al. (2011), "Support union recovery in high-dimensional multivariate regression," *Annals of Statistics*, 39, 1–47.

Pearl, J. (2002), "Causality: models, reasoning, and inference," *IIE Transactions*, 34, 583–589.

Peng, J., Wang, P., Zhou, N., and Zhu, J. (2009), "Partial correlation estimation by joint sparse regression models," *Journal of the American Statistical Association*, 104, 735–746.

Petrovich, J., Reimherr, M., and Daymont, C. (2018), "Functional regression models with highly irregular designs," *arXiv preprint arXiv:1805.08518*.

Qiao, X., Guo, S., and James, G. M. (2018), "Functional graphical models," *Journal of the American Statistical Association*, 114, 1–12.

Ravikumar, P., Lafferty, J., Liu, H., and Wasserman, L. (2009), "Sparse additive models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71, 1009–1030.

Ravikumar, P., Wainwright, M. J., Raskutti, G., Yu, B., et al. (2011), "High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence," *Electronic Journal of Statistics*, 5, 935–980.

Schumaker, L. (2007), *Spline Functions: Basic Theory*, Cambridge University Press.

Stone, C. J. et al. (1985), "Additive regression and other nonparametric models," *Annals of Statistics*, 13, 689–705.

Voorman, A., Shojaie, A., and Witten, D. (2013), "Graph estimation with joint additive models," *Biometrika*, 101, 85–101.

Wong, R. K., Li, Y., and Zhu, Z. (2019), "Partially linear functional additive models for multivariate functional data," *Journal of the American Statistical Association*, 114, 406–418.

Xiao, L., Li, C., Checkley, W., and Crainiceanu, C. (2018), "Fast covariance estimation for sparse functional data," *Statistics and computing*, 28, 511–522.

Xue, L., Zou, H., et al. (2012), "Regularized rank-based estimation of high-dimensional nonparanormal graphical models," *Annals of Statistics*, 40, 2541–2571.

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

Yuan, M. and Lin, Y. (2006), "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68, 49–67.

— (2007), "Model selection and estimation in Gaussian graphical model," *Biometrika*, 94, 19–35.

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

Zhou, S., Shen, X., Wolfe, D., et al. (1998), "Local asymptotics for regression splines and confidence regions," *Annals of Statistics*, 26, 1760–1782.

Zhu, H., Strawn, N., and Dunson, D. B. (2016), "Bayesian graphical models for multivariate functional data," *Journal of Machine Learning Research*, 17, 7157–7183.

Zhu, H., Yao, F., and Zhang, H. H. (2014), "Structured functional additive regression in reproducing kernel Hilbert spaces," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76, 581–603.

8 Appendix: Proofs

8.1 Auxiliary results

In this section we state some auxiliary results, which will be used in the proof of the Theorem 4.1. The next Lemma provides a concentration inequality for the norm $\|\hat{\Sigma}_{X^iX^i} - \Sigma_{X^iX^i}\|_{HS}$. It can be proved by similar arguments as given in the proof of Lemma 6 in Qiao et al. (2018) observing the independence of the random variables ξ^i_{ur}. The details are omitted for the sake of brevity.

Lemma 8.1 *Suppose that Assumption 4.1 is satisfied. Then, there exists a constant C_1 such that for all $0 < \epsilon \leq C_1$ and for each $i = 1, \dots, p$*

$$P\Big(\|\hat{\Sigma}_{X^iX^i} - \Sigma_{X^iX^i}\|_{HS} \geq \epsilon\Big) \lesssim \exp(-C_1 n\epsilon^2).$$

Let $\xi^i = (\xi^i_{ur})^{1\leq r\leq m_n}_{1\leq u\leq n} \in \mathbb{R}^{n\times m_n}$ be the matrix of unobserved scores, and define

$$\tilde{h}_n(\xi^j_{ur}) = (\tilde{h}_{n1}(\xi^j_{ur}), \dots, \tilde{h}_{nk_n}(\xi^j_{ur}))^\intercal$$

as the vector of the centered k_n B-splines functions evaluated at the score ξ^j_{ur}, where

$$\tilde{h}_{nk}(\xi^j_{ur}) = h_k(\xi^j_{ur}) - \frac{1}{n}\sum_{u=1}^{n} h_k(\xi^j_{ur}), \quad k = 1, \dots k_n. \tag{8.1}$$

Note that this definition corresponds to (4.3), where the expectation has been replaced by its empirical counterpart, and to (3.5), where the estimated scores $\hat{\xi}^j_{ur}$ have been replaced by the unobserved scores ξ^j_{ur}. Let

$$\tilde{H}_n(\xi^j) = (\tilde{h}^\intercal_n(\xi^j_{ur}))_{1\leq u\leq n, 1\leq r\leq m_n} \in \mathbb{R}^{n\times k_n m_n}, \tag{8.2}$$

$$\tilde{\mathbf{H}}^\intercal_{\mathbf{n}}(\xi^{\mathsf{N}^i_\mathsf{n}}) = (\tilde{H}_n(\xi^j), j \in \mathsf{N}^i_\mathsf{n}) \in \mathbb{R}^{n\times n^i k_n m_n}, \tag{8.3}$$

and define

$$\Sigma^n_{\mathsf{N}^i_\mathsf{n}\mathsf{N}^i_\mathsf{n}} = \frac{1}{n}\tilde{\mathbf{H}}_{\mathbf{n}}(\xi^{\mathsf{N}^i_\mathsf{n}})\tilde{\mathbf{H}}^\intercal_{\mathbf{n}}(\xi^{\mathsf{N}^i_\mathsf{n}}) \;\in \mathbb{R}^{n^i k_n m_n \times n^i k_n m_n}, \tag{8.4}$$

which is the sample analog of the matrix $\Sigma^*_{\mathsf{N}^i_\mathsf{n}\mathsf{N}^i_\mathsf{n}}$ defined in (4.6). Similarly, let $\hat{\xi}^i = (\hat{\xi}^i_{ur})_{1\leq u\leq n, 1\leq r\leq m_n}$ be the $n \times m_n$ matrix of the estimated scores, and

$$\tilde{\mathbf{H}}^\intercal_{\mathbf{n}}(\hat{\xi}^{\mathsf{N}^i_\mathsf{n}}) = (\tilde{H}_n(\hat{\xi}^j), j \in \mathsf{N}^i_\mathsf{n}) \in \mathbb{R}^{n\times n^i k_n m_n}, \tag{8.5}$$

where

$$\tilde{H}_n(\hat{\xi}^j) = (\tilde{h}^\intercal_n(\hat{\xi}^j_{ur}))_{1\leq u\leq n, 1\leq r\leq m_n} \in \mathbb{R}^{n\times k_n m_n}, \tag{8.6}$$

and $\tilde{h}_n(\hat{\xi}^j_{ur})$ is defined in (3.7). Then,

$$\hat{\Sigma}^n_{\mathsf{N}^i_{\mathsf{n}}\mathsf{N}^i_{\mathsf{n}}} = \frac{1}{n}\tilde{\mathbf{H}}_{\mathbf{n}}(\hat{\xi}^{\mathsf{N}^i_{\mathsf{n}}})\tilde{\mathbf{H}}^{\mathsf{T}}_{\mathbf{n}}(\hat{\xi}^{\mathsf{N}^i_{\mathsf{n}}}) \in \mathbb{R}^{n^i k_n m_n \times n^i k_n m_n} \tag{8.7}$$

is the estimated version of the sample design matrix $\Sigma^n_{\mathsf{N}^i_{\mathsf{n}}\mathsf{N}^i_{\mathsf{n}}}$ in (8.4). The next result provides tail bounds for all entries of the matrix $\hat{\Sigma}^n_{\mathsf{N}^i_{\mathsf{n}}\mathsf{N}^i_{\mathsf{n}}} - \Sigma^n_{\mathsf{N}^i_{\mathsf{n}}\mathsf{N}^i_{\mathsf{n}}}$.

Theorem 8.1 *Suppose that Assumption 4.1 holds. Then, there exists a positive constant C_1 such that for any $\delta > 0$ satisfying $0 < \delta \leq C_1$ and for all $(i, j) \in \mathsf{V} \times \mathsf{V}$, $i \neq j$, $r, q = 1, \ldots, m_n$ and $k, \ell = 1, \ldots, k_n$, we have*

$$P\Big(\Big|\frac{1}{n}\sum_{u=1}^{n}\Big(\tilde{h}_{nk}(\hat{\xi}^i_{ur})\tilde{h}_{n\ell}(\hat{\xi}^j_{uq}) - \tilde{h}_{nk}(\xi^i_{ur})\tilde{h}_{n\ell}(\xi^j_{uq})\Big)\Big| \geq \delta\Big) \lesssim \exp\Big(-C_1 n^{1-\alpha(2+3\beta)} k_n^{-2}\delta^2\Big).$$

Proof. First, we have

$$\Big|\sum_{u=1}^{n}\big(\tilde{h}_{nk}(\hat{\xi}^i_{ur})\tilde{h}_{n\ell}(\hat{\xi}^j_{uq}) - \tilde{h}_{nk}(\xi^i_{ur})\tilde{h}_{n\ell}(\xi^j_{uq})\big)\Big| \leq T_1 + T_2,$$

where the terms T_1 and T_2 are defined as

$$T_1 = \Big|\sum_{u=1}^{n}\tilde{h}_{n\ell}(\xi^j_{uq})\big(\tilde{h}_{nk}(\hat{\xi}^i_{ur}) - \tilde{h}_{nk}(\xi^i_{ur})\big)\Big|, \quad T_2 = \Big|\sum_{u=1}^{n}\tilde{h}_{nk}(\hat{\xi}^i_{ur})\Big(\tilde{h}_{n\ell}(\hat{\xi}^j_{uq}) - \tilde{h}_{n\ell}(\xi^j_{uq})\Big)\Big|.$$

Consequently, for any $\delta > 0$,

$$P\Big(\Big|\sum_{u=1}^{n}\Big(\tilde{h}_{nk}(\hat{\xi}^i_{ur})\tilde{h}_{n\ell}(\hat{\xi}^j_{uq}) - \tilde{h}_{nk}(\xi^i_{ur})\tilde{h}_{n\ell}(\xi^j_{uq})\Big)\Big| \geq 2n\delta\Big) \leq P(T_1 \geq n\delta) + P(T_2 \geq n\delta), \tag{8.8}$$

and therefore it is sufficient to derive inequalities for the two probabilities on the right-hand side of (8.8).

(a) We start with the probability $P(T_1 \geq n\delta)$. By the definition of $\tilde{h}_{nk}(\hat{\xi}^i_{ur})$ and $\tilde{h}_{nk}(\xi^i_{ur})$ in (3.5) and (8.1), respectively, and some elementary calculations, we obtain for any $\delta > 0$

$$P(T_1 \geq n\delta) \leq P\Big(T_{11} \geq \frac{n\delta}{2}\Big) + P\Big(T_{12} \geq \frac{n\delta}{2}\Big),$$

where

$$T_{11} = \Big|\sum_{u=1}^{n} h_\ell(\xi^j_{uq})(h_k(\hat{\xi}^i_{ur}) - h_k(\xi^i_{ur}))\Big|,$$

$$T_{12} = \Big|n^{-1}\sum_{v=1}^{n} h_\ell(\xi^j_{vq})\sum_{u=1}^{n}(h_k(\hat{\xi}^i_{ur}) - h_k(\xi^i_{ur}))\Big|.$$

We now derive a concentration inequality for T_{11}. By Cauchy-Schwarz inequality and using the fact that $|h_\ell(\xi^j_{uq})| \leq 1$ we have

$$P\Big(T_{11} \geq \frac{n\delta}{2}\Big) \leq P\Big(\sum_{u=1}^{n} |h_k(\hat{\xi}^i_{ur}) - h_k(\xi^i_{ur})|^2 \geq \frac{n\delta^2}{4}\Big),$$

and Taylor's expansion gives

$$\Big|h_k(\hat{\xi}^i_{ur}) - h_k(\xi^i_{ur})\Big| = \Big|\frac{\partial h_k(\xi^*)}{\partial \xi^i_{ur}}\left(\hat{\xi}^i_{ur} - \xi^i_{ur}\right)\Big|, \tag{8.9}$$

where ξ^* lies in the line segment between $\hat{\xi}^i_{ur}$ and ξ^i_{ur}. From the derivative formula of the B-splines (see De Boor (1978), Ch.10) there exists a constant $M > 0$, independent of n, such that for all $k = 1, \ldots, k_n$ and $x \in [-1, 1]$,

$$\Big|\frac{\partial h_k(x)}{\partial x}\Big| \leq M L_n, \tag{8.10}$$

where L_n is the number of knots. As a result, using (8.9) and (8.10) we obtain

$$P(T_{11} \geq n\delta) \leq P\Big(\sum_{u=1}^{n} |h_k(\hat{\xi}^i_{ur}) - h_k(\xi^i_{ur})|^2 \geq \frac{n\delta^2}{4}\Big) \leq P\Big(\sum_{u=1}^{n} |\hat{\xi}^i_{ur} - \xi^i_{ur}|^2 \geq \frac{n\delta^2}{4M^2L_n^2}\Big).$$

Recall that $\hat{\xi}^i_{ur} = (\hat{\lambda}^i_r)^{-1/2}\langle X^i_u, \hat{\phi}^i_r\rangle$ and $\xi^i_{ur} = (\lambda^i_r)^{-1/2}\langle X^i_u, \phi^i_r\rangle$. Then,

$$\begin{aligned}\hat{\xi}^i_{ur} - \xi^i_{ur} =& (\hat{\lambda}^i_r)^{-1/2}\langle X^i_u, \hat{\phi}^i_r\rangle - (\lambda^i_r)^{-1/2}\langle X^i_u, \phi^i_r\rangle \\ \leq& ((\hat{\lambda}^i_r)^{-1/2} - (\lambda^i_r)^{-1/2})\langle X^i_u, \hat{\phi}^i_r\rangle + (\lambda^i_r)^{-1/2}\langle X^i_u, \hat{\phi}^i_r - \phi^i_r\rangle.\end{aligned}$$

By the Cauchy-Schwarz inequality and using the fact that $\|\hat{\phi}^i_r\| = 1$ we obtain,

$$\begin{aligned}|\hat{\xi}^i_{ur} - \xi^i_{ur}| &\leq |(\hat{\lambda}^i_r)^{-1/2} - (\lambda^i_r)^{-1/2}|\|X^i_u\| + (\lambda^i_r)^{-1/2}\|X^i_u\|\|\hat{\phi}^i_r - \phi^i_r\| \\ &\leq |(\hat{\lambda}^i_r)^{-1/2} - (\lambda^i_r)^{-1/2}|\|X^i_u\| + (\lambda^i_r)^{-1/2} d^i_r \|X^i_u\|\|\hat{\Sigma}_{X^iX^i} - \Sigma_{X^iX^i}\|_{HS},\end{aligned}$$

where we have used the inequality $\|\hat{\phi}^i_r - \phi^i_r\| \leq d^i_r\|\hat{\Sigma}_{X^iX^i} - \Sigma_{X^iX^i}\|_{HS}$ (see Lemma 4.3 in Bosq, 2012) and assume w.l.o.g. that $\hat{\phi}^i_r$ can be chosen to satisfy $\operatorname{sgn}\langle\hat{\phi}^i_r, \phi^i_r\rangle = 1$ (see the discussion in Remark 2.1). Using the inequality $(a+b)^2 \leq 2(a^2+b^2), a, b \in \mathbb{R}$, this implies

$$\begin{aligned}P\left(T_{11} \geq n\delta\right) \leq& P\Big(|(\hat{\lambda}^i_r)^{-1/2} - (\lambda^i_r)^{-1/2}|^2 \sum_{u=1}^{n} \|X^i_u\|^2 \geq \frac{n\delta^2}{16M^2L_n^2}\Big) \\ &+ P\Big((\lambda^i_r)^{-1}(d^i_r)^2\|\hat{\Sigma}_{X^iX^i} - \Sigma_{X^iX^i}\|^2_{HS} \sum_{u=1}^{n} \|X^i_u\|^2 \geq \frac{n\delta^2}{16M^2L_n^2}\Big).\end{aligned} \tag{8.11}$$

We now consider the first term at the right-hand side of (8.11). Observe that $\sum_{u=1}^n \|X^i_u\|^2 = \sum_{r=1}^\infty \lambda^i_r \sum_{u=1}^n (\xi^i_{ur})^2$, and recall that $\xi^i_{ur} \in [-1, 1]$ with $E(\xi^i_{ur}) = 1$, $E((\xi^i_{ur})^2) = 1$. Thus $|(\xi^i_{ur})^2 - 1| \leq 2$, which implies that for each $r = 1, \ldots, m_n$, $\sum_{u=1}^{n}((\xi_{ur}^i)^2 - 1)$ is a sub-Gaussian random variable with parameter proxy $\sigma^2 = 4n$. Consequently, we obtain by Theorem 2.1 of Boucheron et al. (2013)

$$E\Big\{\sum_{u=1}^{n}((\xi_{um}^i)^2 - 1)\Big\}^{2k} \leq k!(16n)^k, \quad k \geq 1.$$

Using the convexity of the function $x \mapsto x^{2k}$ and Jensen's inequality, it follows

$$\begin{aligned} E\Big|\sum_{u=1}^{n}(\|X_u^i\|^2 - E\|X_u^i\|^2)\Big|^{2k} &= E\Big\{\sum_{r=1}^{\infty}\lambda_r^i\sum_{u=1}^{n}((\xi_{um}^i)^2 - 1)\Big\}^{2k} \\ &\leq \sum_{r=1}^{\infty}\lambda_r^i E\Big\{\sum_{u=1}^{n}((\xi_{um}^i)^2 - 1)\Big\}^{2k}(\sum_{r=1}^{\infty}\lambda_r^i)^{2k-1} \\ &\leq k!(16\lambda_0^2 n)^k, \quad k \geq 1, \end{aligned}$$

where $\lambda_0 = \sup_{i \leq p}\sum_{r=1}^{\infty}\lambda_r^i < \infty$ (due to Assumption 4.1). Hence, from Theorem 2.1 of Boucheron et al. (2013), we obtain for all $\epsilon > 0$,

$$P\Big(\sum_{u=1}^{n}(\|X_u^i\|^2 - E\|X_u^i\|^2) \geq \epsilon\Big) \leq \exp\Big(-\frac{\epsilon^2}{128\lambda_0^2 n}\Big).$$

Furthermore, $E(\sum_{u=1}^{n}\|X_u^i\|^2) \leq n\lambda_0$. Thus, for all $\epsilon/2 \geq n\lambda_0$,

$$P\Big(\sum_{u=1}^{n}\|X_u^i\|^2 \geq \epsilon\Big) \leq P\Big(\sum_{u=1}^{n}(\|X_u^i\|^2 - E\|X_u^i\|^2) \geq \epsilon\Big) \leq \exp\Big(-\frac{\epsilon^2}{128\lambda_0^2 n}\Big). \tag{8.12}$$

Now, we obtain for the first probability on the right-hand side of (8.11)

$$\begin{aligned} &P\Big(|(\hat{\lambda}_r^i)^{-1/2} - (\lambda_r^i)^{-1/2}|^2\sum_{u=1}^{n}\|X_u^i\|^2 \geq \frac{n\delta^2}{16M^2L_n^2}\Big) \\ &\quad\leq P\Big(|(\hat{\lambda}_r^i)^{-1/2} - (\lambda_r^i)^{-1/2}| \geq \frac{\delta}{2M^{1/2}L_n}\Big) + P\Big(\sum_{u=1}^{n}\|X_u^i\|^2 \geq \frac{n}{4M}\Big). \end{aligned} \tag{8.13}$$

Define the event $\Omega_{m_n}^i = \{\|\hat{\Sigma}_{X^iX^i} - \Sigma_{X^iX^i}\|_{HS} < 2^{-1}\delta_{m_n}^i\}$, where $\delta_{m_n}^i = \min_{1 \leq r \leq m_n}\{\lambda_r^i - \lambda_{r+1}^i\}$. Assumption 4.1(i) implies $\delta_{m_n}^i \geq d_2^{-1}m_n^{-(1+\beta)}$ leading to

$$P((\Omega_{m_n}^i)^{\complement}) \leq P(\|\hat{\Sigma}_{X^iX^i} - \Sigma_{X^iX^i}\|_{HS} \geq 2^{-1}d_2^{-1}m_n^{-(1+\beta)}) \lesssim \exp(-C_1 n m_n^{-2(1+\beta)}), \tag{8.14}$$

for some $C_1 > 0$, where we have used Lemma 8.1 with $\delta = 2^{-1}d_2^{-1}m_n^{-(1+\beta)}$. Furthermore, from Lemma 4.43 of Bosq (2012) we have on the event $\Omega_{m_n}^i$

$$\sup_{r \geq 1}|\hat{\lambda}_r^i - \lambda_r^i| \leq \|\hat{\Sigma}_{X^iX^i} - \Sigma_{X^iX^i}\|_{HS} \leq 2^{-1}\delta_{m_n}^i \leq 2^{-1}\lambda_{m_n}^i.$$

This implies $\hat{\lambda}_r^i \geq \frac{\lambda_r^i}{2}$, $\hat{\lambda}_r^i \leq 2\lambda_r^i$ and

$$|(\hat{\lambda}_r^i)^{-1/2} - (\lambda_r^i)^{-1/2}| \leq \frac{(\hat{\lambda}_r^i)^{-1}|\hat{\lambda}_r^i - \lambda_r^i|(\lambda_r^i)^{-1}}{(\hat{\lambda}_r^i)^{-1/2} + (\lambda_r^i)^{-1/2}} \leq 2(\lambda_r^i)^{-3/2}|\hat{\lambda}_r^i - \lambda_r^i|.$$

This together with (8.14) imply that

$$P\Big(|(\hat{\lambda}_r^i)^{-1/2}-(\lambda_r^i)^{-1/2}|\geq\frac{\delta}{2M^{1/2}L_n}\Big)\leq P\Big(\Big(|(\hat{\lambda}_r^i)^{-1/2}-(\lambda_r^i)^{-1/2}|\geq\frac{\delta}{2M^{1/2}L_n}\Big)\cap\Omega_{m_n}^i\Big)+P((\Omega_{m_n}^i)^{\complement})$$
$$\leq P\Big(\Big(|\hat{\lambda}_r^i-\lambda_r^i|\geq\frac{\delta(\lambda_r^i)^{3/2}}{4M^{1/2}L_n}\Big)+P((\Omega_{m_n}^i)^{\complement})$$
$$\leq P\Big((\|\hat{\Sigma}_{X^iX^i}-\Sigma_{X^iX^i}\|_{HS}\geq\frac{\delta(\lambda_r^i)^{3/2}}{4M^{1/2}L_n}\Big)+P((\Omega_{m_n}^i)^{\complement}),$$

where we used Lemma 4.43 of Bosq (2012) for the third inequality. Therefore, from this, Lemma 8.1, (8.12) with $\epsilon=\frac{n}{4M}$ and the fact that $d_0r^{-\beta}\leq\lambda_r^i$, the right-hand side of (8.13) can be upper-bounded by

$$P\Big((\|\hat{\Sigma}_{X^iX^i}-\Sigma_{X^iX^i}\|_{HS}\geq\frac{\delta(\lambda_r^i)^{3/2}}{4M^{1/2}L_n}\Big)+P\Big(\sum_{u=1}^n\|X_u^i\|^2\geq\frac{n}{4M}\Big)+P((\Omega_{m_n}^i)^{\complement}) \tag{8.15}$$
$$\lesssim\exp(-C_1nL_n^{-2}m_n^{-3\beta}\delta^2)+\exp(-C_2n)+\exp(-C_3nm_n^{-2(\beta+1)}),$$

for all $0<\delta L_n^{-1}m_n^{-3\beta/2}\leq C_1$ and some positive constants C_1,C_2 and C_3.

For the second term on the right-hand side of the inequality (8.11) we use Lemma 8.1, (8.12) and the fact that $(d_r^i)^{-1}\geq\frac{d_2}{2\sqrt{2}}m_n^{-1-\beta}$, to obtain,

$$P\Big((\lambda_r^i)^{-1}(d_r^i)^2\|\hat{\Sigma}_{X^iX^i}-\Sigma_{X^iX^i}\|_{HS}^2\sum_{u=1}^n\|X_u^i\|^2\geq\frac{n\delta^2}{16M^2L_n^2}\Big)$$
$$\leq P\Big(\|\hat{\Sigma}_{X^iX^i}-\Sigma_{X^iX^i}\|_{HS}\geq\frac{\delta(\lambda_r^i)^{1/2}(d_r^i)^{-1}}{2M^{1/2}L_n}\Big)+P\Big(\sum_{u=1}^n\|X_u^i\|^2\geq\frac{n}{4M}\Big) \tag{8.16}$$
$$\lesssim\exp(-C_4n)+\exp(-C_5nL_n^{-2}m_n^{-(2+3\beta)}\delta^2),$$

for $0<\delta L_n^{-1}m_n^{-(2+3\beta)/2}\leq C_5$ and $C_4>0$, $C_5>0$.

Combining (8.15) and (8.16) we obtain for all $0<\delta L_n^{-1}m_n^{-3\beta/2}\leq C_3$ the inequality

$$P\left(T_{11}\geq n\delta\right)\lesssim\exp(-C_1n)+\exp(-C_2nm_n^{-2(\beta+1)})+\exp(-C_3nL_n^{-2}m_n^{-(2+3\beta)}\delta^2). \tag{8.17}$$

We now consider the probability $P\left(T_{12}\geq n\delta\right)$. Using the fact that $|h_\ell(\xi_{uq}^j)|\leq 1$ and Taylor's expansion (8.9) yield

$$P\left(T_{12}\geq n\delta\right)\ \leq\ P\Big(\Big|\sum_{u=1}^n(\hat{\xi}_{ur}^i-\xi_{ur}^i)\Big|\geq\frac{n\delta}{2ML_n}\Big)\leq P\Big(\sum_{u=1}^n|\hat{\xi}_{ur}^i-\xi_{ur}^i|^2\geq\frac{n\delta^2}{4M^2L_n^2}\Big),$$

where we have used the Cauchy-Schwarz inequality. Therefore, by similar arguments as used in the derivation of the bound for $P\left(T_{11}\geq n\delta\right)$, there exist positive constants C_4, C_5 and C_6 such that for all $0<\delta L_n^{-1}m_n^{-3\beta/2}\leq C_6$

$$P\left(T_{12}\geq n\delta\right)\lesssim\exp(-C_4n)+\exp(-C_5nm_n^{-2(\beta+1)})+\exp(-C_6nL_n^{-2}m_n^{-(2+3\beta)}\delta^2). \tag{8.18}$$

Combining (8.17) and (8.18) and choosing suitable constants, we obtain for $0 < \delta L_n^{-1} m_n^{-3\beta/2} \le C_1$,

$$P(T_1 \ge n\delta) \lesssim \exp(-C_1 n L_n^{-2} m_n^{-(2+3\beta)} \delta^2). \tag{8.19}$$

(b) To derive a bound for the term $P\left(T_2 \ge n\delta\right)$ in (8.8) we use the decomposition

$$\begin{aligned} P\left(T_2 \ge n\delta\right) \le & P\Big(\Big|\sum_{u=1}^n \Big(\tilde h_{nk}(\hat\xi_{ur}^i) - \tilde h_{nk}(\xi_{ur}^i)\Big)\Big(\tilde h_{n\ell}(\hat\xi_{uq}^j) - \tilde h_{n\ell}(\xi_{uq}^j)\Big)\Big| \ge \frac{n\delta}{2}\Big) \\ & + P\Big(\Big|\sum_{u=1}^n \tilde h_k(\xi_{ur}^i)\Big(\tilde h_{n\ell}(\hat\xi_{uq}^j) - \tilde h_\ell(\xi_{uq}^j)\Big)\Big| \ge \frac{n\delta}{2}\Big) \\ = & P\Big(T_{21} \ge \frac{n\delta}{2}\Big) + P\Big(T_{22} \ge \frac{n\delta}{2}\Big), \end{aligned}$$

where the last inequality defines the terms T_{21} and T_{22} in an obvious manner. For the second term, we obtain by the same arguments as used to estimate $P\left(T_1 \ge n\delta\right)$ for any $0 < \delta L_n^{-1} m_n^{-3\beta/2} \le C_2$,

$$P\Big(T_{22} \ge \frac{n\delta}{2}\Big) \lesssim \exp(-C_2 n L_n^{-2} m_n^{-(2+3\beta)} \delta^2). \tag{8.20}$$

For the first term, we use the definition of the centred B-splines in (3.5) and (8.1) to obtain for any $\delta > 0$

$$P\Big(T_{21} \ge \frac{n\delta}{2}\Big) \le P\Big(T_{211} \ge \frac{n\delta}{4}\Big) + P\Big(T_{212} \ge \frac{n\delta}{4}\Big) \ ,$$

where

$$\begin{aligned} T_{211} &= \Big|\sum_{u=1}^n \Big(h_k(\hat\xi_{ur}^i) - h_k(\xi_{ur}^i)\Big)\Big(h_\ell(\hat\xi_{uq}^j) - h_\ell(\xi_{uq}^j)\Big)\Big|, \\ T_{212} &= n^{-1}\Big|\sum_{u=1}^n \Big(h_k(\hat\xi_{ur}^i) - h_k(\xi_{ur}^i)\Big)\Big|\Big|\sum_{u=1}^n \Big(h_\ell(\hat\xi_{uq}^j) - h_\ell(\xi_{uq}^j)\Big)\Big|. \end{aligned}$$

To derive a concentration bound for the first term, we use (8.9), (8.10) and the Cauchy-Schwarz inequality to obtain

$$\begin{aligned} P\Big(T_{211} \ge \frac{n\delta}{4}\Big) &\le P\Big(\sum_{u=1}^n |\hat\xi_{ur}^i - \xi_{ur}^i||\hat\xi_{uq}^j - \xi_{uq}^j| \ge \frac{n\delta}{4M^2 L_n^2}\Big) \\ &\le 2P\Big(\sum_{u=1}^n |\hat\xi_{ur}^i - \xi_{ur}^i|^2 \ge \frac{n\delta}{4M^2 L_n^2}\Big) \lesssim \exp(-C_3 n L_n^{-2} m_n^{-(2+3\beta)} \delta), \end{aligned}$$

for a positive constant C_3 such that for $0 < \delta^{1/2} L_n^{-1} m_n^{-3\beta/2} \le C_3$. Here the last inequality follows by the same arguments as used for the bound of $P\left(T_{11} \ge n\delta\right)$. Finally, for the term $P\left(T_{212} \ge \frac{n\delta}{4}\right)$ we have

$$P\Big(T_{212} \ge \frac{n\delta}{4}\Big) = P\Big(\Big|\sum_{u=1}^n \Big(h_k(\hat\xi_{ur}^i) - h_k(\xi_{ur}^i)\Big)\Big|\Big|\sum_{u=1}^n \Big(h_\ell(\hat\xi_{uq}^j) - h_\ell(\xi_{uq}^j)\Big)\Big| \ge \frac{n^2\delta}{4}\Big).$$

Thus, we can apply similar techniques as used in the derivation of the bound (8.17) with $\frac{n\delta}{2}$ replaced by $\frac{n\delta^{1/2}}{2}$ leading to

$$P\Big(T_{212} \geq \frac{n\delta}{4}\Big) \lesssim \exp(-C_4 n L_n^{-2} m_n^{-(2+3\beta)}\delta), \tag{8.21}$$

for $0 < \delta^{1/2} L_n^{-1} m_n^{-3\beta/2} \leq C_4$. Combining these results with (8.8) and (8.19) and using the fact that $m_n \asymp n^{\alpha}$ and $k_n > L_n$, we obtain the assertion of the Theorem. □

Theorem 8.1 states that the elements of the matrix $\hat{\Sigma}_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i} - \Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}$ exhibit exponential-type probability tails. We also observe that the decay rate β of the eigenvalues appears in the tail behaviour. A similar condition of exponential tails is imposed on the elements of the sample covariance matrix of scalar and functional Gaussian data for the analysis of high-dimensional Gaussian graphical models (see, for example, Ravikumar et al., 2011; Qiao et al., 2018).

Proposition 8.1 *Suppose that Assumptions 4.2, 4.3 and condition* (4.11) *are satisfied. Then, there exist functions* $\tilde{f}^{ij}_{nqr} = \sum_{k=1}^{k_n} \beta^{ij}_{qrk}\tilde{h}_{nk}$ *and positive constants* c_1, C_1*, such that*

$$P\big(\Omega^{\complement}\big) \leq 2\exp\Big(-C_1\frac{nk_n^{-2d}}{n^i m_n^2} + \log(n^i m_n^2)\Big),$$

where

$$\Omega = \Big\{\max_{j\in\mathsf{N}_\mathsf{n}^i}\max_{1\leq q,r\leq m_n}\frac{1}{\sqrt{n}}\|\mathbf{f}^{ij}_{\mathbf{qr}} - \tilde{\mathbf{f}}^{ij}_{\mathbf{qr}}\|_2 < c_1 k_n^{-d}\Big\}\ , \tag{8.22}$$

and $\mathbf{f}^{ij}_{\mathbf{qr}} = \big(f^{ij}_{qr}(\xi^j_{1r}),\ldots,f^{ij}_{qr}(\xi^j_{nr})\big)^\top$, $\tilde{\mathbf{f}}^{ij}_{\mathbf{qr}} = \big(\tilde{f}^{ij}_{qr}(\xi^j_{1r}),\ldots,\tilde{f}^{ij}_{qr}(\xi^j_{nr})\big)^\top$.

Proof. By Assumptions 4.2 and 4.3 for any $f^{ij}_{qr} \in \mathscr{F}_{\kappa,\rho}$ there exists a B-spline $g^{ij}_{qr} = \sum_{k=1}^{k_n}\beta^{ij}_{qrk}h_k \in \mathcal{S}_{\ell L_n}$ and a positive constant c_1 such that

$$\|f^{ij}_{qr} - g^{ij}_{qr}\|_\infty \leq c_1 k_n^{-d},$$

(see Lemma 5 in Stone et al. (1985)). Let $\tilde{f}^{ij}_{qr}(\xi^j_{ur}) = g^{ij}_{qr}(\xi^j_{ur}) - \frac{1}{n}\sum_{u=1}^n g^{ij}_{qr}(\xi^j_{ur})$. Then recalling the notation (8.14) we have $\tilde{f}^{ij}_{qr}(\xi^j_{ur}) = \sum_{k=1}^{k_n}\beta^{ij}_{qrk}\tilde{h}_{nk}(\xi^j_{ur})$, and we obtain

$$\begin{aligned}\frac{1}{n}\|\mathbf{f}^{ij}_{\mathbf{qr}} - \tilde{\mathbf{f}}^{ij}_{\mathbf{qr}}\|_2^2 &\leq 2c_1^2k_n^{-2d} + 2\Big(\frac{1}{n}\sum_{u=1}^n g^{ij}_{qr}(\xi^j_{ur})\Big)^2\\ &\leq 2c_1^2k_n^{-2d} + 4\Big(\frac{1}{n}\sum_{u=1}^n(g^{ij}_{qr}(\xi^j_{ur}) - f^{ij}_{qr}(\xi^j_{ur}))\Big)^2 + 4\Big(\frac{1}{n}\sum_{u=1}^n f^{ij}_{qr}(\xi^j_{ur})\Big)^2\\ &\leq 6c_1^2k_n^{-2d} + 4\Big(\frac{1}{n}\sum_{u=1}^n f^{ij}_{qr}(\xi^j_{ur})\Big)^2.\end{aligned}$$

From this, condition (4.11), Hoeffding's inequality and the union bound, it follows with an appropriate constant $c_2 > 0$

$$P\Big(\max_{j\in \mathsf{N}_\mathsf{n}^i}\max_{1\le q,r\le m_n}\frac{1}{n}\|f_{qr}^{ij}(\xi_r^j)-\tilde{f}_{nqr}^{ij}(\xi_r^j)\|_2^2\ge c_1^2k_n^{-2d}\Big)\le P\Big(\max_{j\in \mathsf{N}_\mathsf{n}^i}\max_{1\le q,r\le m_n}\Big|\frac{1}{n}\sum_{u=1}^n f_{qr}^{ij}(\xi_{ur}^{ij})\Big|\ge c_2k_n^{-d}\Big)$$
$$\le 2\exp\Big(-\frac{nk_n^{-2d}}{2M^2n^im_n^2}+\log(n^im_n^2)\Big)\ ,$$

which completes the proof. $\square$

8.2 Rates of convergence for sample design matrices

In this section we show that if Assumptions 4.7 and 4.8 hold, then with high probability, the assumptions hold also for the corresponding *sample matrices*

$$\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}=\frac{1}{n}\tilde{\mathbf{H}}_\mathbf{n}(\xi^{\mathsf{N}_\mathsf{n}^i})\tilde{\mathbf{H}}_\mathbf{n}^\intercal(\xi^{\mathsf{N}_\mathsf{n}^i})\in\mathbb{R}^{n^ik_nm_n\times n^ik_nm_n},\ \Sigma^n_{\xi^j\mathsf{N}_\mathsf{n}^i}=\frac{1}{n}\tilde{H}_n^\intercal(\xi^j)\tilde{\mathbf{H}}_\mathbf{n}^\intercal(\xi^{\mathsf{N}_\mathsf{n}^i})\in\mathbb{R}^{k_nm_n\times k_nm_nn^i},\quad (8.23)$$

where $\tilde{H}_n(\xi^j)$ and $\tilde{\mathbf{H}}_\mathbf{n}(\xi^{\mathsf{N}_\mathsf{n}^i})$ are defined in (8.2) and (8.3), respectively. Note that the matrices in (8.23) are based on the unobserved scores and are the sample analogs of the matrices $\Sigma^*_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}$ and $\Sigma^*_{\xi^j\mathsf{N}_\mathsf{n}^i}$ in (4.6) and (4.7), respectively.

Lemma 8.2 *Suppose that Assumption 4.7 holds. Then, there exists a constant $C_1 > 0$ such that for any $\delta > 0$,*

$$P\left(\|\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}-\Sigma^*_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}\|_F\ge\delta\right)\le 2\exp\Big(-C_1\frac{n\delta^2}{(n^im_nk_n)^2}+2\log(n^im_nk_n)\Big).\quad (8.24)$$

$$P\left(\Lambda_{\min}(\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i})\le C_{\min}-\delta\right)\le 2\exp\Big(-C_1\frac{n\delta^2}{(n^im_nk_n)^2}+2\log(n^im_nk_n)\Big).\quad (8.25)$$

Proof. Weyl's Lemma yields

$$\Lambda_{\min}(\Sigma^*_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i})-\Lambda_{\min}(\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i})\le\|\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}-\Sigma^*_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}\|_2\le\|\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}-\Sigma^*_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}\|_F,$$

and by Assumption 4.7 we have,

$$P\left(\Lambda_{\min}(\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i})\le C_{\min}-\delta\right)\le P\left(\|\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}-\Sigma^*_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}\|_F\ge\delta\right).\quad (8.26)$$

By definition the $n^im_nk_n\times n^im_nk_n$ matrix $\Sigma^n_{\mathsf{N}_n^i\mathsf{N}_\mathsf{n}^i}-\Sigma^*_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}$ contains elements of the form

$$W^{j_1,j_2}_{k\ell,rq}=\frac{1}{n}\sum_{u=1}^n\tilde{h}_{nk}(\xi^{j_1}_{uq})\tilde{h}_{n\ell}(\xi^{j_2}_{ur})-E(\tilde{h_k}(\xi^{j_1}_q)\tilde{h_\ell}(\xi^{j_2}_r)),$$

which can be rewritten (recalling the notation (8.1)) as $A_1 - A_2$, where

$$A_1 = \frac{1}{n}\sum_{u=1}^{n} h_k(\xi_{uq}^{j_1})h_\ell(\xi_{ur}^{j_2}) - E(h_k(\xi_q^{j_1})h_\ell(\xi_r^{j_2})),$$

$$A_2 = \frac{1}{n^2}\sum_{u_1=1}^{n}\sum_{u_2=1}^{n} h_k(\xi_{u_1q}^{j_1})h_\ell(\xi_{u_2r}^{j_2}) - E(h_k(\xi_q^{j_1}))E(h_\ell(\xi_r^{j_2})).$$

Next, observe that the summands of A_1 have expectation 0 and are bounded in absolute value by 2. Therefore, by Hoeffding's inequality, we have $P(|A_1| \geq \epsilon) \leq 2\exp(-\frac{n\epsilon^2}{128})$ for any $\epsilon > 0$. Moreover, the term A_2 can be written as $\frac{n-1}{n}A_{21} + A_{22}$, where

$$A_{21} = \frac{1}{n(n-1)}\sum_{u_1 \neq u_2}^{n} h_k(\xi_{u_1q}^{j_1})h_\ell(\xi_{u_2r}^{j_2}) - E(h_k(\xi_{u_1q}^{j_1})h_\ell(\xi_{u_2r}^{j_2}),$$

is a U-statistic and $A_{22} = \frac{1}{n^2}\sum_{u_1=1}^{n} h_k(\xi_{u_1q}^{j_1})h_\ell(\xi_{u_1r}^{j_2}) - E(h_k(\xi_{u_1q}^{j_1})h_\ell(\xi_{u_1r}^{j_2}))$. Consequently, by Hoeffding's inequality for U-statistics (Hoeffding, 1963) $P(|A_{21}| \geq \epsilon) \leq 2\exp(-\frac{n\epsilon^2}{128})$ for any $\epsilon > 0$, and it is easy to see (due to the additional factor $1/n$) that A_{22} satisfies an even stronger concentration inequality. Therefore, it follows that for any $\epsilon > 0$

$$P\left(|W_{k\ell,rq}^{j_1,j_2}| \geq \epsilon\right) \leq 2\exp\left(-C_1 n\epsilon^2\right), \tag{8.27}$$

for some constant $C_1 > 0$. Thus, the union bound over the $(n^i m_n k_n)^2$ indices and the choice of $\epsilon = \frac{\delta}{n^i m_n k_n}$ in (8.27) yields (8.24). Finally, the assertion (8.25) follows from relation (8.26) at the beginning of the proof. □

The next Lemma guarantees that the matrices defined in (8.23) satisfy the irrepresentable condition in Assumption 4.8 with high probability.

Lemma 8.3 *If Assumption 4.7 and 4.8 are satisfied for some $0 < \eta \leq 1$, then*

$$P\Big(\max_{j \notin \mathsf{N}_\mathsf{n}^i} \|\Sigma^n_{\xi^j \mathsf{N}_\mathsf{n}^i}(\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i})^{-1}\|_F \geq \frac{1-\frac{\eta}{2}}{\sqrt{n^i}}\Big) \lesssim \exp\Big(-C_1\frac{n}{((n^i)^{5/4}m_n k_n)^2} + 2\log(pm_n k_n)\Big),$$

where C_1 is a positive constant that depends only on $C_{\min}$ and η.

Proof. First, we decompose

$$\max_{j \notin \mathsf{N}_\mathsf{n}^i} \|\Sigma^n_{\xi^j \mathsf{N}_\mathsf{n}^i}(\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i})^{-1}\|_F \leq \max_{j \notin \mathsf{N}_\mathsf{n}^i} \|\Sigma^n_{\xi^j \mathsf{N}_\mathsf{n}^i}(\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i})^{-1} - \Sigma^*_{\xi^j \mathsf{N}_\mathsf{n}^i}(\Sigma^*_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i})^{-1}\|_F + \max_{j \notin \mathsf{N}_\mathsf{n}^i} \|\Sigma^*_{\xi^j \mathsf{N}_\mathsf{n}^i}(\Sigma^*_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i})^{-1}\|_F.$$

By Assumption 4.8 we have $\max_{j \notin \mathsf{N}_\mathsf{n}^i} \|\Sigma^*_{\xi^j \mathsf{N}_\mathsf{n}^i}(\Sigma^*_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i})^{-1}\|_F \leq \frac{1-\eta}{\sqrt{n^i}}$ and it suffices to consider

$$P\Big(\max_{j \notin \mathsf{N}_\mathsf{n}^i} \|\Sigma^n_{\xi^j \mathsf{N}_\mathsf{n}^i}(\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i})^{-1} - \Sigma^*_{\xi^j \mathsf{N}_\mathsf{n}^i}(\Sigma^*_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i})^{-1}\|_F \geq \frac{\eta}{2\sqrt{n^i}}\Big).$$

For this purpose we use the decomposition $\Sigma^n_{\xi^j \mathsf{N}^i_n}(\Sigma^n_{\mathsf{N}^i_n\mathsf{N}^i_n})^{-1} - \Sigma^*_{\xi^j \mathsf{N}^i_n}(\Sigma^*_{\mathsf{N}^i_n\mathsf{N}^i_n})^{-1} = T^j_1 + T^j_2 + T^j_3$ where

$$T^j_1 = \Sigma^*_{\xi^j \mathsf{N}^i_n}\left((\Sigma^n_{\mathsf{N}^i_n\mathsf{N}^i_n})^{-1} - (\Sigma^*_{\mathsf{N}^i_n\mathsf{N}^i_n})^{-1}\right) \ , \ \ T^j_2 = \left(\Sigma^n_{\xi^j \mathsf{N}^i_n} - \Sigma^*_{\xi^j \mathsf{N}^i_n}\right)(\Sigma^*_{\mathsf{N}^i_n\mathsf{N}^i_n})^{-1},$$
$$T^j_3 = \left(\Sigma^n_{\xi^j \mathsf{N}^i_n} - \Sigma^*_{\xi^j \mathsf{N}^i_n}\right)\left((\Sigma^n_{\mathsf{N}^i_n\mathsf{N}^i_n})^{-1} - (\Sigma^*_{\mathsf{N}^i_n\mathsf{N}^i_n})^{-1}\right),$$

and control the probabilities $P\big(\max_{j\notin \mathsf{N}^i_n} \|T^j_h\|_F \geq \frac{\eta}{6\sqrt{n^i}}\big)$ separately.

(a) For the first term T^j_1, we use the identity $A^{-1} - B^{-1} = A^{-1}(B-A)B^{-1}$ and obtain from Assumption 4.8

$$\begin{aligned}
\max_{j\notin \mathsf{N}^i_n} \|T^j_1\|_F &\leq \max_{j\notin \mathsf{N}^i_n} \|\Sigma^*_{\xi^j \mathsf{N}^i_n}(\Sigma^*_{\mathsf{N}^i_n\mathsf{N}^i_n})^{-1}\|_F \|(\Sigma^n_{\mathsf{N}^i_n\mathsf{N}^i_n} - \Sigma^*_{\mathsf{N}^i_n\mathsf{N}^i_n})(\Sigma^n_{\mathsf{N}^i_n\mathsf{N}^i_n})^{-1}\|_F \\
&\leq \frac{(1-\eta)}{\sqrt{n^i}} \|\Sigma^n_{\mathsf{N}^i_n\mathsf{N}^i_n} - \Sigma^*_{\mathsf{N}^i_n\mathsf{N}^i_n}\|_F \|(\Sigma^n_{\mathsf{N}^i_n\mathsf{N}^i_n})^{-1}\|_2.
\end{aligned}$$

Thus, defining the event $\mathcal{T} = \{\|(\Sigma^n_{\mathsf{N}^i_n\mathsf{N}^i_n})^{-1}\|_2 \leq \frac{2}{C_{\min}}\}$ we obtain

$$\begin{aligned}
P\Big(\max_{j\notin \mathsf{N}^i_n} \|T^j_1\|_F \geq \frac{\eta}{6\sqrt{n^i}}\Big) &\leq P\Big(\|\Sigma^n_{\mathsf{N}^i_n\mathsf{N}^i_n} - \Sigma^*_{\mathsf{N}^i_n\mathsf{N}^i_n}\|_F \|(\Sigma^n_{\mathsf{N}^i_n\mathsf{N}^i_n})^{-1}\|_2 \geq \frac{\eta}{6(1-\eta)}\Big) \\
&\leq P\Big(\|\Sigma^n_{\mathsf{N}^i_n\mathsf{N}^i_n} - \Sigma^*_{\mathsf{N}^i_n\mathsf{N}^i_n}\|_F \geq \frac{\eta C_{\min}}{12(1-\eta)}\Big) + P\left(\mathcal{T}^\complement\right) \qquad (8.28) \\
&\leq 4\exp\Big(-C_1 \frac{n}{(n^i m_n k_n)^2} + 2\log(n^i m_n k_n)\Big),
\end{aligned}$$

where we used Lemma 8.2 with $\delta = \frac{\eta C_{\min}}{12(1-\eta)}$ and $\delta = \frac{C_{\min}}{2}$ for the last inequality.

(b) For the second term T^j_2, we have

$$\max_{j\notin \mathsf{N}^i_n} \|T^j_2\|_F \leq \|(\Sigma^*_{\mathsf{N}^i_n\mathsf{N}^i_n})^{-1}\|_2 \max_{j\notin \mathsf{N}^i_n} \|\Sigma^n_{\xi^j \mathsf{N}^i_n} - \Sigma^*_{\xi^j \mathsf{N}^i_n}\|_F \leq C^{-1}_{\min} \max_{j\notin \mathsf{N}^i_n} \|\Sigma^n_{\xi^j \mathsf{N}^i_n} - \Sigma^*_{\xi^j \mathsf{N}^i_n}\|_F,$$

where we used Assumption 4.7 in the second inequality. Thus,

$$P\Big(\max_{j\notin \mathsf{N}^i_n} \|T^j_2\|_F \geq \frac{\eta}{6\sqrt{n^i}}\Big) \leq P\Big(\max_{j\notin \mathsf{N}^i_n} \|\Sigma^n_{\xi^j \mathsf{N}^i_n} - \Sigma^*_{\xi^j \mathsf{N}^i_n}\|_F \geq \frac{\eta C_{\min}}{6\sqrt{n^i}}\Big).$$

Now, using similar arguments as in the proof of (8.24) in Lemma 8.2, we can show

$$P\left(\|\Sigma^n_{\xi^j \mathsf{N}^i_n} - \Sigma^*_{\xi^j \mathsf{N}^i_n}\|_F \geq \delta\right) \leq 2\exp\Big(-C_1 \frac{n\delta^2}{n^i m_n^2 k_n^2} + \log(n^i m_n^2 k_n^2)\Big), \qquad (8.29)$$

for some positive constant $C_1 > 0$. This bound with $\delta = \frac{\eta C_{\min}}{6\sqrt{n^i}}$ and the union bound yield

$$\begin{aligned}
P\Big(\max_{j\notin \mathsf{N}^i_n} \|T^j_2\|_F \geq \frac{\eta}{6\sqrt{n^i}}\Big) &\leq (p-n^i) P\Big(\|\Sigma^n_{\xi^j \mathsf{N}^i_n} - \Sigma^*_{\xi^j \mathsf{N}^i_n}\|_F \geq \frac{\eta C_{\min}}{6\sqrt{n^i}}\Big) \\
&\leq 2\exp\Big(-C_1 \frac{n}{(n^i m_n k_n)^2} + \log(n^i m_n^2 k_n^2) + \log(p-n^i)\Big). \qquad (8.30)
\end{aligned}$$

(c) For the third term T_3^j, we have

$$P\Big(\max_{j\notin \mathsf{N}_\mathsf{n}^i} \|T_3^j\|_F \geq \frac{\eta}{6\sqrt{n^i}}\Big) \leq P\Big(\max_{j\notin \mathsf{N}_\mathsf{n}^i} \|\Sigma^n_{\xi^j \mathsf{N}_\mathsf{n}^i} - \Sigma^*_{\xi^j \mathsf{N}_\mathsf{n}^i}\|_F \geq \sqrt{\frac{\eta}{6\sqrt{n^i}}}\Big) + P\Big(\|(\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i})^{-1} - (\Sigma^*_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i})^{-1}\|_F \geq \sqrt{\frac{\eta}{6\sqrt{n^i}}}\Big). \tag{8.31}$$

Using (8.29) with $\delta = \sqrt{\frac{\eta}{6\sqrt{n^i}}}$ we obtain for the first term on the right-hand side of (8.31)

$$P\Big(\max_{j\notin \mathsf{N}_\mathsf{n}^i} \|\Sigma^n_{\xi^j \mathsf{N}_\mathsf{n}^i} - \Sigma^*_{\xi^j \mathsf{N}_\mathsf{n}^i}\|_F \geq \sqrt{\frac{\eta}{6\sqrt{n^i}}}\Big) \lesssim \exp\Big(-C_1\frac{n}{(n^i)^{3/2} m_n^2 k_n^2} + \log(n^i m_n^2 k_n^2) + \log((p-n^i) m_n^2 k_n^2)\Big). \tag{8.32}$$

To derive a bound for the second term in (8.31) we apply the same arguments as used for the term T_1^j

$$P\Big(\|(\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i})^{-1} - (\Sigma^*_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i})^{-1}\|_F \geq \sqrt{\frac{\eta}{6\sqrt{n^i}}}\Big) \lesssim \exp\Big(-C_1\frac{n}{(n^i)^{5/2} m_n^2 k_n^2} 2\log(n^i m_n k_n)\Big) + \exp\Big(-C_1\frac{n}{(n^i m_n k_n)^2} + 2\log(n^i m_n k_n)\Big). \tag{8.33}$$

Thus, from (8.31), (8.32) and (8.33) we obtain

$$\begin{aligned} P\Big(\max_{j\notin \mathsf{N}_\mathsf{n}^i} \|T_3^j\|_F \geq \frac{\eta}{6\sqrt{n^i}}\Big) \lesssim{}& \exp\Big(-C_1\frac{n}{(n^i)^{3/2}(m_n k_n)^2} + \log(n^i m_n^2 k_n^2) + \log((p-n^i) m_n^2 k_n^2)\Big) \\ &+ 2\exp\Big(-C_1\frac{n}{(n^i)^{5/2} m_n^2 k_n^2} + 2\log(n^i m_n k_n)\Big) \\ &+ 2\exp\Big(-C_1\frac{n}{(n^i m_n k_n)^2} + 2\log(n^i m_n k_n)\Big). \end{aligned} \tag{8.34}$$

Putting together (8.28), (8.30) and (8.34) and using the fact $\log n^i \leq \log(p-n^i) \leq \log p$ (since $n^i \leq p$) we conclude

$$P\Big(\max_{j\notin \mathsf{N}_\mathsf{n}^i} \|\Sigma^n_{\xi^j \mathsf{N}_\mathsf{n}^i}(\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i})^{-1}\|_F \geq \frac{1-\frac{\eta}{2}}{\sqrt{n^i}}\Big) \leq 2\exp\Big(-C_1\frac{n}{(n^i)^{5/2} m_n^2 k_n^2} + 2\log(p m_n k_n)\Big),$$

for some positive constant C_1 that depends on $C_{\min}$ and η. This completes the proof. □

8.3 Proof of Theorem 4.1

We begin establishing the model selection consistency given that Assumptions 4.7 and 4.8 are satisfied by the sample matrices defined in (8.23). In particular we define the event

$$\mathcal{N} = \{\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}, \Sigma^n_{\xi^j \mathsf{N}_\mathsf{n}^i} \text{ satisfy Assumptions 4.7 and 4.8}\} \tag{8.35}$$

and state the following result, which is the essential step in the proof of Theorem 4.1 and will be proved in Section 8.4 below.

Proposition 8.2 *If the assumptions of Theorem 4.1 are satisfied. Then,*

$$P(\hat{\mathsf{N}}_n^i \neq \mathsf{N}_n^i \cap \mathcal{N}) \lesssim \exp\Big(-C_1 \frac{n^{1-\alpha(2+3\beta)}(\lambda_n \sum_{j\in\mathsf{N}_\mathsf{n}^i} \|B^{*ij}_{m_n k_n}\|_F)^2}{n^i m_n^2 k_n^4} + 2\log(n^i m_n k_n)\Big), \tag{8.36}$$

where C_1 is a positive constant.

We have

$$P(\hat{\mathsf{N}}_\mathsf{n}^i \neq \mathsf{N}_\mathsf{n}^i) \leq \mathsf{P}(\hat{\mathsf{N}}_\mathsf{n}^i \neq \mathsf{N}_\mathsf{n}^i \text{ and } \mathcal{N}) + \mathsf{P}(\mathcal{N}^\complement)\ ,$$

where the first probability on the right hand side can be estimated by (8.36). Moreover, by Lemmas 8.2 and 8.3,

$$P(\mathcal{N}^\complement) \lesssim \exp\Big(-C_1 \frac{n}{((n^i)^{5/4} m_n k_n)^2} + 2\log(p m_n k_n)\Big) + \exp\Big(-C_1 \frac{n}{(n^i)^2 m_n k_n)^2} + 2\log(p m_n k_n)\Big),$$

and this proves Theorem 4.1.

8.4 Proof of Proposition 8.2

We follow a similar strategy as in Bach (2008) and Lee et al. (2016b), who showed consistency of the group lasso in a reproducing kernel Hilbert space framework. First, we consider the following alternative form of the group lasso problem (3.9)

$$\widehat{PL}_i(B^i, \hat{\xi}) = \frac{1}{2n}\|\hat{\xi}^i - \tilde{\mathbf{H}}_\mathbf{n}^\intercal(\hat{\xi}^{-i})B^i\|_F^2 + \frac{\lambda_n}{2}(\sum_{j\neq i}^{p} \|B^{ij}\|_F)^2. \tag{8.37}$$

Because the function $x \to x^2$, $x \geq 0$ is monotone, problem (8.37) leads to the same regularisation paths as problem (3.9) (see Bach, 2008, , page 1187 for more details). To derive the Karush-Kuhn-Tucker (KKT) conditions, we recall the notations (8.3), (8.4) and define the matrices

$$\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i \xi^j} = \frac{1}{n}\tilde{\mathbf{H}}_\mathbf{n}(\hat{\xi}^{\mathsf{N}_\mathsf{n}^i})\tilde{H}_n(\hat{\xi}^j) \ \in \mathbb{R}^{n^i k_n m_n \times k_n m_n}, \tag{8.38}$$

$$\hat{\Sigma}^n_{\xi^j \xi^i} = \frac{1}{n}\tilde{H}_n^\intercal(\hat{\xi}^j)\hat{\xi}^i \ \in \mathbb{R}^{k_n m_n \times m_n} \tag{8.39}$$

when $j \neq i$ and

$$\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i \xi^i} = \frac{1}{n}\tilde{\mathbf{H}}_\mathbf{n}(\hat{\xi}^{\mathsf{N}_\mathsf{n}^i})\hat{\xi}^i \ \in \mathbb{R}^{n^i k_n m_n \times m_n}, \tag{8.40}$$

where the matrices $\tilde{H}_n(\hat{\xi}^j)$ and $\tilde{\mathbf{H}}_\mathbf{n}(\hat{\xi}^{\mathsf{N}_\mathsf{n}^i})$ have been defined in (8.6) and (8.5) respectively. We also denote by $\Sigma^n_{\mathsf{N}_\mathsf{n}^i \xi^j}$, $\Sigma^n_{\xi^j \xi^i}$, $\Sigma^n_{\mathsf{N}_\mathsf{n}^i \xi^i}$ the versions of (8.38), (8.39), (8.40) that use the true scores ξ^i_{ur} instead of the estimated $\hat{\xi}^i_{ur}$ (see also equation (8.7)).

Lemma 8.4 *(KKT conditions) A matrix* $B^i = (B^{ij}, j \in \mathsf{V} \setminus \{\mathsf{i}\}) \in \mathbb{R}^{(p-1)k_n m_n \times m_n}$ *with support* N^i_n *is optimal for problem* (8.37) *if and only if*

$$(\hat{\Sigma}^n_{\mathsf{N}^i_\mathsf{n}\mathsf{N}^i_\mathsf{n}} + \lambda_n \hat{D}_{\mathsf{N}^i_\mathsf{n}}) B^{\mathsf{N}^i_\mathsf{n}} - \hat{\Sigma}^n_{\mathsf{N}^i_\mathsf{n}\xi^i} = 0, \quad \textit{for all } j \in \mathsf{N}^i_\mathsf{n}, \tag{8.41a}$$

$$\|\hat{\Sigma}^n_{\xi^j \mathsf{N}^i_\mathsf{n}} B^{\mathsf{N}^i_\mathsf{n}} - \hat{\Sigma}^n_{\xi^j \xi^i}\|_F \le \lambda_n \sum_{j \neq i}^p \|B^{ij}\|_F, \quad \textit{for all } j \notin \mathsf{N}^i_\mathsf{n} \tag{8.41b}$$

where $\hat{\Sigma}^n_{\mathsf{N}^i_\mathsf{n}\mathsf{N}^i_\mathsf{n}}$ *is defined in* (8.7)*,* $B^{\mathsf{N}^i_\mathsf{n}} = (B^{ij}, j \in \mathsf{N}^i_\mathsf{n}) \in \mathbb{R}^{n^i k_n m_n \times m_n}$*,* $B = (\beta^{ij}_{qrk} : 1 \le q, r \le m_n, 1 \le k \le k_n)$ *and*

$$\hat{D}_{\mathsf{N}^i_\mathsf{n}} = \operatorname{diag}\big((\hat{D}_{\mathsf{N}^i_\mathsf{n}})_{jj} : j \in \hat{\mathsf{N}}^i_\mathsf{n}\big)\big)$$

is a block diagonal matrix with n^i *elements* $(\hat{D}_{\mathsf{N}^i_\mathsf{n}})_{jj} = \frac{\sum_{\ell \neq i}^p \|\hat{B}^{i\ell}\|_F}{\|\hat{B}^{ij}\|_F} I_{k_n m_n} \in \mathbb{R}^{k_n m_n \times k_n m_n}$.

The idea of the proof is to first construct an estimator $\hat{B}^{\mathsf{N}^i_\mathsf{n}}_n$ by minimizing the following restricted problem given the true support N^i_n. That is,

$$\hat{B}^{\mathsf{N}^i_\mathsf{n}}_n = \operatorname{argmin}\big\{\widehat{PL}_{\mathsf{N}^i_\mathsf{n}}(B, \hat{\xi}) : B \in \mathbb{R}^{n^i k_n m_n \times m_n}\big\}, \tag{8.42}$$

where

$$\widehat{PL}_{\mathsf{N}^i_\mathsf{n}}(B, \hat{\xi}) = \frac{1}{2n}\|\hat{\xi}^i - \tilde{\mathbf{H}}^\intercal_\mathbf{n}(\hat{\xi}^{\mathsf{N}^i_\mathsf{n}})B\|^2_F + \frac{\lambda_n}{2}\Big(\sum_{j \in \mathsf{N}^i_\mathsf{n}}^p \|B^{ij}\|_F\Big)^2, \tag{8.43}$$

(note that $\widehat{PL}_{\mathsf{N}^i_\mathsf{n}}(B, \hat{\xi})$ corresponds to the function (8.37), where we put $B^{ij} = 0$ whenever $j \notin \mathsf{N}^i$). and to show that the minimizer in (8.42) is "close" to the true matrix $B^{*\mathsf{N}^i_\mathsf{n}}_n$ defined in (4.9). To achieve this we use similar arguments as in Bach (2008) and construct another auxiliary estimator $\tilde{B}^{\mathsf{N}^i_\mathsf{n}}_n$ that minimizes the restricted penalized function, where the group lasso penalty in (8.42) is replaced by an ℓ_2-type penalty. More precisely, $\tilde{B}^{\mathsf{N}^i_\mathsf{n}}_n$ is defined by

$$\tilde{B}^{\mathsf{N}^i_\mathsf{n}}_n = \operatorname{argmin}\big\{\widetilde{PL}_{\mathsf{N}^i_\mathsf{n}}(B, \hat{\xi}) : B \in \mathbb{R}^{n^i k_n m_n \times m_n}\big\}, \tag{8.44}$$

where

$$\widetilde{PL}_{\mathsf{N}^i_\mathsf{n}}(B, \hat{\xi}) = \frac{1}{2n}\|\hat{\xi}^i - \tilde{\mathbf{H}}^\intercal_\mathbf{n}(\hat{\xi}^{-i})B\|^2_F + \frac{\lambda_n}{2}\Big(\sum_{\ell \in \mathsf{N}^i_\mathsf{n}} \|B^{*i\ell}_{m_n k_n}\|_F\Big)\Big(\sum_{j \in \mathsf{N}^i_\mathsf{n}} \frac{\|B^{ij}\|^2_F}{\|B^{*ij}_{m_n k_n}\|_F}\Big).$$

We now proceed in the following steps:

(1) In Proposition 8.3 we show that the distance $\|\tilde{B}^{\mathsf{N}^i_\mathsf{n}}_n - B^{*\mathsf{N}^i_\mathsf{n}}_n\|_F$ is small with high probability.

(2) In Proposition 8.4 we show that $\hat{B}^{\mathsf{N}^i_\mathsf{n}}_n$ is close to $\tilde{B}^{\mathsf{N}^i_\mathsf{n}}_n$ with high probability.

(3) In Proposition 8.5 we use this result to derive a concentration bound for $\|\hat{B}_n^{\mathsf{N}_n^i} - B^{*\mathsf{N}_n^i}\|_F$.

(4) We then construct the oracle minimiser $(\hat{B}^{\mathsf{N}_n^i}, \mathbf{0})$, where $\hat{B}^{\mathsf{N}_n^i}$ is the minimiser of (8.42) and $\mathbf{0}$ consists of $(p-1-n^i)$ zero $k_n m_n \times m_n$ matrices.

(5) Finally, in Proposition 8.6 we show that the oracle minimiser is optimal for the restricted problem (8.42) given the true support N_n^i; that is, it satisfies (8.41b).

The minimisation problem (8.42) is convex; however, for $p > n$, it need not to be strictly convex, so that there may not be a unique solution. Nevertheless, the next lemma shows that the matrix $\hat{\Sigma}^n_{\mathsf{N}_n^i\mathsf{N}_n^i}$ defined in (8.7) is strictly positive definite with high probability, and hence the objective function (8.42) is strictly convex, and thus $\hat{B}^{\mathsf{N}_n^i}$ is the unique optimal solution.

Lemma 8.5 *There exists a constant $C_1 > 0$ such that*

$$P\Big(\Lambda_{\min}(\hat{\Sigma}^n_{\mathsf{N}_n^i\mathsf{N}_n^i}) \geq \frac{C_{\min}}{4}\Big) \gtrsim 1 - \exp\Big(-C_1 \frac{n^{1-\alpha(2+3\beta)}}{(n^i)^2 m_n^2 k_n^4} + 2\log(n^i m_n k_n)\Big).$$

Proof. By Weyl's Lemma, we have $\Lambda_{\min}(\Sigma^n_{\mathsf{N}_n^i\mathsf{N}_n^i}) \leq \Lambda_{\min}(\hat{\Sigma}^n_{\mathsf{N}_n^i\mathsf{N}_n^i}) + \|\hat{\Sigma}^n_{\mathsf{N}_n^i\mathsf{N}_n^i} - \Sigma^n_{\mathsf{N}_n^i\mathsf{N}_n^i}\|_2$, and we get

$$P\Big(\Lambda_{\min}(\hat{\Sigma}^n_{\mathsf{N}_n^i\mathsf{N}_n^i}) \leq \frac{C_{\min}}{4} \text{ and } \Lambda_{\min}(\Sigma^n_{\mathsf{N}_n^i\mathsf{N}_n^i}) > \frac{C_{\min}}{2}\Big) \leq P\Big(\|\hat{\Sigma}^n_{\mathsf{N}_n^i\mathsf{N}_n^i} - \Sigma^n_{\mathsf{N}_n^i\mathsf{N}_n^i}\|_2 \geq \frac{C_{\min}}{4}\Big).$$

Furthermore, using $\delta^2 = \frac{1}{(n^i m_n k_n)^2}\frac{C_{\min}}{4}$ in Theorem 8.1 with the union bound over the $(n^i m_n k_n)^2$ index pairs of the matrix $\hat{\Sigma}^n_{\mathsf{N}_n^i\mathsf{N}_n^i} - \Sigma^n_{\mathsf{N}_n^i\mathsf{N}_n^i}$, yields for some positive constant C_1

$$P\Big(\|\hat{\Sigma}^n_{\mathsf{N}_n^i\mathsf{N}_n^i} - \Sigma^n_{\mathsf{N}_n^i\mathsf{N}_n^i}\|_F \geq \frac{C_{\min}}{4}\Big) \lesssim \exp\Big(-C_1 \frac{n^{1-\alpha(2+3\beta)}}{(n^i)^2 m_n^2 k_n^4} + 2\log(n^i m_n k_n)\Big).$$

The assertion now follows by the same arguments as given in the proof of Lemma 8.2. □

Proposition 8.3 *Suppose Assumptions 4.1-4.7 hold and the regularization parameter λ_n satisfies*

$$\frac{n^i m_n^{3/2}}{k_n^d} \lesssim \sqrt{\frac{2}{C_{\min}}}\lambda_n \sum_{j \in \mathsf{N}_n^i} \|B^{*ij}_{m_n k_n}\|_F. \tag{8.45}$$

Then, there exists a constant $c_2 \in (0, 1/2)$ such that, for any $\delta > 0$ satisfying

$$\frac{2}{C_{\min}}\sqrt{n^i}\lambda_n \sum_{j \in \mathsf{N}_n^i} \|B^{*ij}_{m_n k_n}\|_F \leq c_2\delta, \tag{8.46}$$

we have for the minimizer of (8.44)

$$P\big(\|\tilde{B}_n^{\mathsf{N}_n^i} - B_n^{*\mathsf{N}_n^i}\|_F \geq \delta\big) \lesssim \exp\Big(-C_1 \frac{n^{1-\alpha(2+3\beta)}\delta^2}{(n^i)^2 m_n^2 k_n^4} + 2\log(n^i m_n k_n)\Big),$$

where $B_n^{*\mathsf{N}_n^i}$ is defined in (4.9) *and the constant C_1 satisfies* $0 < \delta \leq C_1$.

Proof. Before we start with the proof we note that condition (8.45) refers to the spline approximation error from including only k_n terms and the second condition (8.46) represents the bias due to ridge penalisation.

For the proof we use similar arguments as given in the proof of Proposition 2 of Lee et al. (2020). The main change that we need to consider is the approximation error of the additive regression functions by splines. First, the minimizer $\tilde{B}_n^{\mathsf{N}^i}$ defined in (8.44) is of the form

$$\tilde{B}_n^{\mathsf{N}_\mathsf{n}^i} = \left(\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i} + \lambda_n D^*_{\mathsf{N}_\mathsf{n}^i}\right)^{-1}\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i\xi^i}.$$

where $D^*_{\mathsf{N}_\mathsf{n}^i}$ is a block diagonal matrix with $(D^*_{\mathsf{N}_\mathsf{n}^i})_{jj} = \sum_{\ell\neq i}^p \|B^{*i\ell}_{m_nk_n}\|_F/\|B^{*ij}_{m_nk_n}\|_F I_{k_nm_n}$, $j \in \mathsf{N}_\mathsf{n}^i$ as diagonal blocks, and the matrices $\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}$ and $\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i\xi^i}$ are defined in (8.7) and (8.38), respectively.

A simple calculation shows that

$$\|\tilde{B}_n^{\mathsf{N}_\mathsf{n}^i} - B_n^{*\mathsf{N}_\mathsf{n}^i}\|_F \leq T_1 + T_2 + T_3,$$

where the terms T_1, T_2 and T_3 are defined by

$$\begin{aligned}
T_1 &= \big\|(\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i} + \lambda_n D^*_{\mathsf{N}_\mathsf{n}^i})^{-1}(\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i\xi^i} - \Sigma^n_{\mathsf{N}_\mathsf{n}^i\xi^i})\big\|_F,\\
T_2 &= \big\|\{(\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i} + \lambda_n D^*_{\mathsf{N}_\mathsf{n}^i})^{-1} - (\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i} + \lambda_n D^*_{\mathsf{N}_\mathsf{n}^i})^{-1}\}\Sigma^n_{\mathsf{N}_\mathsf{n}^i\xi^i}\big\|_F,\\
T_3 &= \big\|(\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i} + \lambda_n D^*_{\mathsf{N}_\mathsf{n}^i})^{-1}\Sigma^n_{\mathsf{N}_\mathsf{n}^i\xi^i} - B_n^{*\mathsf{N}_\mathsf{n}^i}\big\|_F.
\end{aligned}$$

Thus, $P\big(\|\tilde{B}_n^{\mathsf{N}_\mathsf{n}^i} - B_n^{*\mathsf{N}_\mathsf{n}^i}\|_F \geq 3\delta\big) \leq \sum_{i=1}^3 P(T_i \geq \delta)$, and it is sufficient to derive bounds for the three probabilities corresponding to the random variables T_1, T_2 and T_3. Starting with T_1 we have

$$T_1 \leq \|(\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i} + \lambda_n D^*_{\mathsf{N}_\mathsf{n}^i})^{-1}\|_2\|\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i\xi^i} - \Sigma^n_{\mathsf{N}_\mathsf{n}^i\xi^i}\|_F \leq \frac{2}{C_{\min}}\|\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i\xi^i} - \Sigma^n_{\mathsf{N}_\mathsf{n}^i\xi^i}\|_F,$$

where we use the fact that

$$\|\hat{\Sigma}_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}\|_2 \geq \frac{C_{\min}}{2}$$

on the event $\mathcal{N}$ and that $(\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i} + \lambda_n D^*_{\mathsf{N}_\mathsf{n}^i})^{-1} \preceq \big(\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}\big)^{-1}$. Therefore, using Lemma 8.2, similar arguments as given in the proof of Theorem 8.1 and applying the union bound over the $n^i m_n^2 k_n$ pairs, we obtain

$$P\left(T_1 \geq \delta\right) \leq P\Big(\|\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i\xi^i} - \Sigma^n_{\mathsf{N}_\mathsf{n}^i\xi^i}\|_F \geq \frac{C_{\min}\delta}{2}\Big) \lesssim \exp\Big(-C_1\frac{n^{1-\alpha(2+3\beta)}\delta^2}{n^i m_n^2 k_n^3} + \log(n^i m_n^2 k_n)\Big), \tag{8.47}$$

for $0 < \frac{\delta}{n^i m_n^2 k_n} \leq C_1$ with $C_1 > 0$ depending on $C_{\min}$. To derive the bound for the probability $P\left(T_2 \geq \delta\right)$ we use the identity $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$ to obtain (on the event $\mathcal{N}$)

$$\begin{aligned}
T_2 &\leq \|(\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i} + \lambda_n D^*_{\mathsf{N}_\mathsf{n}^i})^{-1}\|_2\|(\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i} - \hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i})(\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i} + \lambda_n D^*_{\mathsf{N}_\mathsf{n}^i})^{-1}\Sigma^n_{\mathsf{N}_\mathsf{n}^i\xi^i}\|_F\\
&\leq \frac{2}{C_{\min}}\|\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i} - \hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}\|_F\|(\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i} + \lambda_n D^*_{\mathsf{N}_\mathsf{n}^i})^{-1}\Sigma^n_{\mathsf{N}_\mathsf{n}^i\xi^i}\|_2.
\end{aligned} \tag{8.48}$$

Recall the relation (4.4), the notation $f_{qr}^{ij} = \sum_{k=1}^{\infty} \beta_{qrk}^{*ij} h_k \in \mathscr{F}_{\kappa,\rho}$, and let

$$w_{uq}^{i} = \sum_{j \in \mathsf{N}_{\mathsf{n}}^{i}} \sum_{r=1}^{m_n} (f_{qr}^{ij}(\xi_{ur}^{j}) - \tilde{f}_{nqr}^{ij}(\xi_{ur}^{j})), \quad q = 1, \ldots, m_n, u = 1, \ldots, n, \tag{8.49}$$

where $\tilde{f}_{qr}^{ij}$ denotes the function from Proposition 8.1. Then, we can rewrite relation (4.4) in the form

$$\xi^{i} = \tilde{\mathbf{H}}_{\mathbf{n}}^{\intercal}(\xi^{\mathsf{N}_{\mathsf{n}}^{i}}) B_{n}^{*\mathsf{N}_{\mathsf{n}}^{i}} + w^{i} + \epsilon^{i} \in \mathbb{R}^{n \times m_n},$$

where $w^{i} = (w_{uq}^{i})_{1 \le u \le n, 1 \le q \le m_n}$, $\epsilon^{i} = (\epsilon_{uq}^{i})_{1 \le u \le n, 1 \le q \le m_n} \in \mathbb{R}^{n \times m_n}$ and $\tilde{\mathbf{H}}_{\mathbf{n}}^{\top}(\xi^{\mathsf{N}_{\mathsf{n}}^{i}})$ is defined in (8.3). Furthermore, by multiplying from the left the above equation with $\frac{\tilde{\mathbf{H}}_{\mathbf{n}}(\xi^{\mathsf{N}^{i}})}{n}$ we obtain

$$\Sigma_{\mathsf{N}_{\mathsf{n}}^{i}\xi^{i}}^{n} = \Sigma_{\mathsf{N}_{\mathsf{n}}^{i}\mathsf{N}_{\mathsf{n}}^{i}}^{n} B_{n}^{*\mathsf{N}_{\mathsf{n}}^{i}} + \frac{\tilde{\mathbf{H}}_{\mathbf{n}}(\xi^{\mathsf{N}_{\mathsf{n}}^{i}})}{n} w^{i} + \frac{\tilde{\mathbf{H}}_{\mathbf{n}}(\xi^{\mathsf{N}_{\mathsf{n}}^{i}})}{n} \epsilon^{i}, \tag{8.50}$$

$$\Sigma_{\xi^{j}\xi^{i}}^{n} = \Sigma_{\xi^{j}\mathsf{N}_{\mathsf{n}}^{i}}^{n} B_{n}^{*\mathsf{N}_{\mathsf{n}}^{i}} - \frac{\tilde{H}_{n}^{\intercal}(\xi^{j})}{n} w^{i} - \frac{\tilde{H}_{n}^{\intercal}(\xi^{j})}{n} \epsilon^{i}. \tag{8.51}$$

where the matrix $\Sigma_{\mathsf{N}_{\mathsf{n}}^{i}\xi^{i}}^{n}$ is defined in Section 8.4 and $\Sigma_{\xi^{j}\mathsf{N}_{\mathsf{n}}^{i}}^{n} = (\Sigma_{\mathsf{N}_{\mathsf{n}}^{i}\xi^{j}}^{n})^{\intercal}$. Using this representation and the triangle inequality we get

$$\begin{aligned}
\|(\Sigma_{\mathsf{N}_{\mathsf{n}}^{i}\mathsf{N}_{\mathsf{n}}^{i}}^{n} + \lambda_n D_{\mathsf{N}_{\mathsf{n}}^{i}}^{*})^{-1} \Sigma_{\mathsf{N}_{\mathsf{n}}^{i}\xi^{i}}^{n}\|_2 \le & \|(\Sigma_{\mathsf{N}_{\mathsf{n}}^{i}\mathsf{N}_{\mathsf{n}}^{i}}^{n} + \lambda_n D_{\mathsf{N}_{\mathsf{n}}^{i}}^{*})^{-1} \Sigma_{\mathsf{N}_{\mathsf{n}}^{i}\mathsf{N}_{\mathsf{n}}^{i}}^{n} B_{n}^{*\mathsf{N}_{\mathsf{n}}^{i}}\|_2 + \|(\Sigma_{\mathsf{N}_{\mathsf{n}}^{i}\mathsf{N}_{\mathsf{n}}^{i}}^{n} + \lambda_n D_{\mathsf{N}_{\mathsf{n}}^{i}}^{*})^{-1} \frac{\tilde{\mathbf{H}}_{\mathbf{n}}(\xi^{\mathsf{N}_{\mathsf{n}}^{i}})}{n} w^{i}\|_2 \\
& + \|(\Sigma_{\mathsf{N}_{\mathsf{n}}^{i}\mathsf{N}_{\mathsf{n}}^{i}}^{n} + \lambda_n D_{\mathsf{N}_{\mathsf{n}}^{i}}^{*})^{-1} \frac{\tilde{\mathbf{H}}_{\mathbf{n}}(\xi^{\mathsf{N}_{\mathsf{n}}^{i}})}{n} \epsilon^{i}\|_2.
\end{aligned}$$

As a result from this and (8.48), it follows that for all $\delta > 0$ (on the event $\mathcal{N}$)

$$P(T_2 \ge \delta) \le P\Big(T_{21} \ge \frac{\delta}{3}\Big) + P\Big(T_{22} \ge \frac{\delta}{3}\Big) + P\Big(T_{23} \ge \frac{\delta}{3}\Big), \tag{8.52}$$

where

$$\begin{aligned}
T_{21} =& \frac{2}{C_{\min}} \|\hat{\Sigma}_{\mathsf{N}_{\mathsf{n}}^{i}\mathsf{N}_{\mathsf{n}}^{i}}^{n} - \Sigma_{\mathsf{N}_{\mathsf{n}}^{i}\mathsf{N}_{\mathsf{n}}^{i}}^{n}\|_F \|(\Sigma_{\mathsf{N}_{\mathsf{n}}^{i}\mathsf{N}_{\mathsf{n}}^{i}}^{n} + \lambda_n D_{\mathsf{N}_{\mathsf{n}}^{i}}^{*})^{-1} \Sigma_{\mathsf{N}_{\mathsf{n}}^{i}\mathsf{N}_{\mathsf{n}}^{i}}^{n} B_{n}^{*\mathsf{N}_{\mathsf{n}}^{i}}\|_2, \\
T_{22} =& \frac{2}{C_{\min}} \|\hat{\Sigma}_{\mathsf{N}_{\mathsf{n}}^{i}\mathsf{N}_{\mathsf{n}}^{i}}^{n} - \Sigma_{\mathsf{N}_{\mathsf{n}}^{i}\mathsf{N}_{\mathsf{n}}^{i}}^{n}\|_F \|(\Sigma_{\mathsf{N}_{\mathsf{n}}^{i}\mathsf{N}_{\mathsf{n}}^{i}}^{n} + \lambda_n D_{\mathsf{N}_{\mathsf{n}}^{i}}^{*})^{-1} \frac{\tilde{\mathbf{H}}_{\mathbf{n}}(\xi^{\mathsf{N}_{\mathsf{n}}^{i}})}{n} w^{i}\|_2, \\
T_{23} =& \frac{2}{C_{\min}} \|\hat{\Sigma}_{\mathsf{N}_{\mathsf{n}}^{i}\mathsf{N}_{\mathsf{n}}^{i}}^{n} - \Sigma_{\mathsf{N}_{\mathsf{n}}^{i}\mathsf{N}_{\mathsf{n}}^{i}}^{n}\|_F \|(\Sigma_{\mathsf{N}_{\mathsf{n}}^{i}\mathsf{N}_{\mathsf{n}}^{i}}^{n} + \lambda_n D_{\mathsf{N}_{\mathsf{n}}^{i}}^{*})^{-1} \frac{\tilde{\mathbf{H}}_{\mathbf{n}}(\xi^{\mathsf{N}_{\mathsf{n}}^{i}})}{n} \epsilon^{i}\|_2.
\end{aligned}$$

Next we derive upper bounds for the probabilities in (8.52). For T_{21} observe that

$$T_{21} \le \frac{2}{C_{\min}} \|\hat{\Sigma}_{\mathsf{N}_{\mathsf{n}}^{i}\mathsf{N}_{\mathsf{n}}^{i}}^{n} - \Sigma_{\mathsf{N}_{\mathsf{n}}^{i}\mathsf{N}_{\mathsf{n}}^{i}}^{n}\|_F \|\|(\Sigma_{\mathsf{N}_{\mathsf{n}}^{i}\mathsf{N}_{\mathsf{n}}^{i}}^{n} + \lambda_n D_{\mathsf{N}_{\mathsf{n}}^{i}}^{*})^{-1} \Sigma_{\mathsf{N}_{\mathsf{n}}^{i}\mathsf{N}_{\mathsf{n}}^{i}}^{n}\|_2 \|B_{n}^{*\mathsf{N}_{\mathsf{n}}^{i}}\|_F \le \frac{2}{C_{\min}} \|\hat{\Sigma}_{\mathsf{N}_{\mathsf{n}}^{i}\mathsf{N}_{\mathsf{n}}^{i}}^{n} - \Sigma_{\mathsf{N}_{\mathsf{n}}^{i}\mathsf{N}_{\mathsf{n}}^{i}}^{n}\|_F,$$

where the second inequality uses the fact that $\|B_{n}^{*\mathsf{N}_{\mathsf{n}}^{i}}\|_F < \infty$ by assumption (3.2) and that the norm $\|(\Sigma_{\mathsf{N}_{\mathsf{n}}^{i}\mathsf{N}_{\mathsf{n}}^{i}}^{n} + \lambda_n D_{\mathsf{N}_{\mathsf{n}}^{i}}^{*})^{-1} \Sigma_{\mathsf{N}_{\mathsf{n}}^{i}\mathsf{N}_{\mathsf{n}}^{i}}^{n}\|_2$ is bounded by one. Therefore, it follows from Theorem 8.1 with δ replaced by $\frac{C_{\min}\delta}{6n^{i} m_n k_n}$ and the union bound over the $(n^{i} m_n k_n)^2$ pairs that

$$P\Big(T_{21} \ge \frac{\delta}{3}\Big) \le P\Big(\|\hat{\Sigma}_{\mathsf{N}_{\mathsf{n}}^{i}\mathsf{N}_{\mathsf{n}}^{i}}^{n} - \Sigma_{\mathsf{N}_{\mathsf{n}}^{i}\mathsf{N}_{\mathsf{n}}^{i}}^{n}\|_F \ge \frac{\delta C_{\min}}{6}\Big) \lesssim \exp\Big(- C_2 \frac{n^{1-\alpha(2+3\beta)} \delta^2}{(n^{i})^2 m_n^2 k_n^4} + 2\log(n^{i} m_n k_n)\Big), \tag{8.53}$$

for $0 < \frac{\delta}{n^i m_n k_n} \leq C_2$.

For the term T_{22} note that

$$\|(\Sigma^n_{\mathsf{N}^i_n\mathsf{N}^i_n} + \lambda_n D^*_{\mathsf{N}^i_n})^{-1} \frac{\tilde{\mathbf{H}}_\mathbf{n}(\xi^{\mathsf{N}^i_n})}{\sqrt{n}}\|_2 \leq \|(\Sigma^n_{\mathsf{N}^i_n\mathsf{N}^i_n})^{-1} \frac{\tilde{\mathbf{H}}_\mathbf{n}(\xi^{\mathsf{N}^i_n})}{\sqrt{n}}\|_2 = \Lambda_{\min}(\Sigma^n_{\mathsf{N}^i_n\mathsf{N}^i_n})^{-1/2}, \tag{8.54}$$

where we used Lemma 8.2 for the last inequality with $\delta = \frac{C_{\min}}{2}$. Thus, (on the event $\mathcal{N}$) the term T_{22} can be bounded by

$$T_{22} \leq \Big(\frac{2}{C_{\min}}\Big)^{3/2} \|\hat{\Sigma}^n_{\mathsf{N}^i_n\mathsf{N}^i_n} - \Sigma^n_{\mathsf{N}^i_n\mathsf{N}^i_n}\|_F \|\frac{w^i}{\sqrt{n}}\|_F.$$

Recall the notation of w^i_{uq} in (8.49) and the definition of the event Ω in Proposition 8.1. Then, if the event Ω holds, we have,

$$\begin{aligned}
\|\frac{1}{\sqrt{n}} w^i\|^2_F = \frac{1}{n}\sum_{u=1}^{n}\sum_{q=1}^{m_n}(w^i_{uq})^2 &= \frac{1}{n}\sum_{u=1}^{n}\sum_{q=1}^{m_n}\Big(\sum_{j\in\mathsf{N}^i}\sum_{r=1}^{m_n}(f^{ij}_{qr}(\xi^j_{ur}) - \tilde{f}^{ij}_{nqr}(\xi^j_{ur}))\Big)^2, \\
&\leq \frac{n^i m_n}{n}\sum_{j\in\mathsf{N}^i}\sum_{q=1}^{m_n}\sum_{r=1}^{m_n}\sum_{u=1}^{n}(f^{ij}_{qr}(\xi^j_{ur}) - \tilde{f}^{ij}_{nqr}(\xi^j_{ur}))^2 \\
&\leq \frac{n^i m_n}{n}\sum_{j\in\mathsf{N}^i}\sum_{q=1}^{m_n}\sum_{r=1}^{m_n}\max_{j\in\mathsf{N}^i}\max_{1\leq q,r\leq m_n}\|\mathbf{f}^{ij}_{\mathbf{qr}} - \tilde{\mathbf{f}}^{ij}_{\mathbf{qr}}\|_2 \;\leq\; c_1(n^i)^2 m_n^3 k_n^{-2d},
\end{aligned} \tag{8.55}$$

and by assumption (8.45) it follows that on the event Ω

$$\|\frac{1}{\sqrt{n}} w^i\|_F \lesssim \sqrt{\frac{2}{C_{\min}}}\lambda_n \sum_{j\in\mathsf{N}^i_n}\|B^{*ij}_{m_n k_n}\|_F. \tag{8.56}$$

As a result,

$$\begin{aligned}
P\Big(T_{22} \geq \frac{\delta}{3}\Big) &\leq P\Big(\Big(\frac{2}{C_{\min}}\Big)^2 c_1\lambda_n\sum_{j\in\mathsf{N}^i_n}\|B^{*ij}_{m_n k_n}\|_F\|\hat{\Sigma}^n_{\mathsf{N}^i_n\mathsf{N}^i_n} - \Sigma^n_{\mathsf{N}^i_n\mathsf{N}^i_n}\|_F \geq \frac{\delta}{3}\Big) + P(\Omega^\complement) \\
&\lesssim \exp\Big(-C_3\frac{n^{1-\alpha(2+3\beta)}(\lambda_n\sum_{j\in\mathsf{N}^i_n}\|B^{*ij}_{m_n k_n}\|_F)^{-2}\delta^2}{(n^i)^2 m_n^2 k_n^4} + 2\log(n^i m_n k_n)\Big) \\
&\quad + \exp\Big(-C_3\frac{n}{n^i m_n^2 k_n^{2d}} + \log(n^i m_n^2)\Big)
\end{aligned} \tag{8.57}$$

for $0 < \delta(\lambda_n\sum_{j\in\mathsf{N}^i_n}\|B^{*ij}_n\|_F)^{-1} \leq C_3$, where we have used Theorem 8.1 and Proposition 8.1.

We next derive an upper bound for the probability corresponding to the term T_{23} in (8.52) noting that (on the event $\mathcal{N}$)

$$T_{23} \leq \Big(\frac{2}{C_{\min}}\Big)^2 \|\hat{\Sigma}^n_{\mathsf{N}^i_n\mathsf{N}^i_n} - \Sigma^n_{\mathsf{N}^i_n\mathsf{N}^i_n}\|_F \|\frac{\tilde{\mathbf{H}}_\mathbf{n}(\xi^{\mathsf{N}^i_n})}{n}\epsilon^i\|_F. \tag{8.58}$$

The (i,j) element of the matrix $\frac{\tilde{\mathbf{H}}_{\mathbf{n}}(\xi^{\mathsf{N}_n^i})}{n}\epsilon^i$ can be written as an i.i.d sum of the form $\frac{1}{n}\sum_{u=1}^n \tilde{h}_{nk}(\xi^j_{ur})\epsilon^i_{uq}$. Thus, by Assumption 4.4 it follows that

$$P\Big(\Big|\frac{1}{n}\sum_{u=1}^n \tilde{h}_{nk}(\xi^j_{ur})\epsilon^i_{uq}\Big| \geq \epsilon\Big) \leq 2\exp(-C_5 n\epsilon^2),$$

for any $\epsilon > 0$. Therefore, by applying the union bound over the $n^i m_n^2 k_n$ gives

$$P\Big(\|\frac{\tilde{\mathbf{H}}_{\mathbf{n}}(\xi^{\mathsf{N}_n^i})}{n}\epsilon^i\|_F \geq \epsilon\Big) \leq 2\exp\Big(-C_5\frac{n\epsilon^2}{n^i m_n^2 k_n} + \log(n^i m_n^2 k_n)\Big). \tag{8.59}$$

Using this inequality with $\epsilon = C_{\min}/6$, (8.53) and (8.58) gives

$$P\Big(T_{23} \geq \frac{\delta}{3}\Big) \lesssim \exp\Big(-C_4\frac{n^{1-\alpha(2+3\beta)}\delta^2}{(n^i)^2 m_n^2 k_n^4} + 2\log(n^i m_n k_n)\Big) + \exp\Big(-C_5\frac{n}{n^i m_n^2 k_n} + \log(n^i m_n^2 k_n)\Big), \tag{8.60}$$

for $0 < \frac{\delta}{n^i m_n k_n} \leq C_4$. Therefore from (8.53), (8.57) and (8.60) it follows that

$$P(T_2 \geq \delta) \lesssim \exp\Big(-C_2\frac{n^{1-\alpha(2+3\beta)}\delta^2}{(n^i)^2 m_n^2 k_n^4} + 2\log(n^i m_n k_n)\Big) + 2\exp\Big(-C_3\frac{n}{n^i m_n^2 k_n^{2d}} + \log(n^i m_n^2)\Big), \tag{8.61}$$

for $0 < \frac{\delta}{n^i m_n k_n} \leq C_2$, where the first term dominates the second one because of Assumption 4.5.

Finally, we derive an upper bound for the probability involving T_3. Using representation (8.50) we obtain

$$T_3 \leq T_{31} + T_{32} + T_{33}\ , \tag{8.62}$$

where

$$\begin{aligned}
T_{31} =& \|(\Sigma^n_{\mathsf{N}_n^i\mathsf{N}_n^i} + \lambda_n D^*_{\mathsf{N}_n^i})^{-1}\Sigma^n_{\mathsf{N}_n^i\mathsf{N}_n^i} B_n^{*\mathsf{N}_n^i} - B_n^{*\mathsf{N}_n^i}\|_F,\\
T_{32} =& \|(\Sigma^n_{\mathsf{N}_n^i\mathsf{N}_n^i} + \lambda_n D^*_{\mathsf{N}_n^i})^{-1}\frac{\tilde{\mathbf{H}}_{\mathbf{n}}(\xi^{\mathsf{N}_n^i})}{n} w^i\|_F,\\
T_{33} =& \|(\Sigma^n_{\mathsf{N}_n^i\mathsf{N}_n^i} + \lambda_n D^*_{\mathsf{N}_n^i})^{-1}\frac{\tilde{\mathbf{H}}_{\mathbf{n}}(\xi^{\mathsf{N}_n^i})}{n} \epsilon^i\|_F.
\end{aligned}$$

For the first term on the right-hand side of the above inequality we have (on the event $\mathcal{N}$)

$$T_{31} = \lambda_n\|(\Sigma^n_{\mathsf{N}_n^i\mathsf{N}_n^i} + \lambda_n D^*_{\mathsf{N}_n^i})^{-1} D^*_{\mathsf{N}_n^i} B_n^{*\mathsf{N}_n^i}\|_F \leq \frac{2}{C_{\min}}\lambda_n (n^i)^{1/2}\sum_{j\in\mathsf{N}_n^i}\|B^{*ij}_{m_n k_n}\|_F \leq \frac{\delta}{2(1+c_1)},$$

where we used condition (8.46) with $c_1 = \frac{1}{2c_2} - 1$ and the fact that $\|\mathrm{diag}(\frac{B^{*ij}_{m_n k_n}}{\|B^{*ij}_{m_n k_n}\|_F} : j \in \mathsf{N}^i_n)\|_F = (n^i)^{1/2}$. Moreover, by applying the same arguments for deriving the bound of T_{22} and by using (8.54), conditions (8.45) and (8.46) it follows that on the event Ω

$$T_{32} \leq c_1\left(\frac{2}{C_{\min}}\right)^{1/2}\frac{n^i m_n^{3/2}}{k_n^d} \leq c_1\frac{2}{C_{\min}}\lambda_n\sum_{j\in\mathsf{N}_n^i}\|B^{*ij}_{m_n k_n}\|_F \leq \frac{c_1}{2(1+c_1)}\delta.$$

Therefore, inequalities (8.62) and (8.59) imply that for all $\delta > 0$

$$
\begin{aligned}
P\left(T_3 \geq \delta\right) &\leq P\Big(\frac{2}{C_{\min}}\|\frac{\tilde{\mathbf{H}}_{\mathbf{n}}(\xi^{\mathsf{N}_{\mathsf{n}}^i})\epsilon^i}{n}\|_F \geq \frac{\delta}{2}\Big) + P(\Omega^{\complement}) \\
&\leq 2\exp\Big(- C_3\frac{n\delta^2}{n^i m_n^2 k_n} + \log(n^i m_n^2 k_n)\Big) + 2\exp\Big(- C_3\frac{n}{n^i m_n^2 k_n^{2d}} + \log(n^i m_n^2)\Big) \qquad (8.63)
\end{aligned}
$$

Thus, by (8.47), (8.61) and (8.63), we have shown, for any $\delta >$ such that $0 < \delta \leq C_1$ and $0 < \delta \leq C_2$

$$
\begin{aligned}
P(\|\tilde{B}_n^{\mathsf{N}_{\mathsf{n}}^i} - B_n^{*\mathsf{N}_{\mathsf{n}}^i}\|_F \geq \delta) \lesssim &\exp\Big(- C_1\frac{n^{1-\alpha(2+3\beta)}\delta^2}{n^i m_n^2 k_n^3} + \log(n^i m_n^2 k_n)\Big) \\
&+ \exp\Big(- C_2\frac{n^{1-\alpha(2+3\beta)}\delta^2}{(n^i)^2 m_n^2 k_n^4} + 2\log(n^i m_n k_n)\Big) \\
&+ \exp\Big(- C_3\frac{n\delta^2}{n^i m_n^2 k_n} + \log(n^i m_n^2 k_n)\Big).
\end{aligned}
$$

Since the second term dominates the first and the third, the assertion in Proposition 8.3 follows. □

The next proposition brings $\hat{B}_n^{\mathsf{N}_{\mathsf{n}}^i} = (\hat{B}_n^{ij}, j \in \mathsf{N}_{\mathsf{n}}^i)$ close to $\tilde{B}_n^{\mathsf{N}_{\mathsf{n}}^i} = (\tilde{B}_n^{ij}, j \in \mathsf{N}_{\mathsf{n}}^i)$, from which we can establish the concentration inequality for $\hat{B}_n^{\mathsf{N}_{\mathsf{n}}^i}$.

Proposition 8.4 *Let* $\hat{B}_n^{\mathsf{N}_{\mathsf{n}}^i}$ *be the minimiser of* (8.42) *and* $\tilde{B}_n^{\mathsf{N}_{\mathsf{n}}^i}$ *be the minimiser of* (8.44). *If* $\Lambda_{\min}(\hat{\Sigma}^n_{\mathsf{N}_{\mathsf{n}}^i\mathsf{N}_{\mathsf{n}}^i}) \geq \frac{C_{\min}}{4}$ *then,*

$$
\|\hat{B}_n^{\mathsf{N}_{\mathsf{n}}^i} - \tilde{B}_n^{\mathsf{N}_{\mathsf{n}}^i}\|_F \leq \frac{10}{C_{\min}}\lambda_n n^i (b_n^{*i})^{-1}\|\tilde{B}_n^{\mathsf{N}_{\mathsf{n}}^i} - B_n^{*\mathsf{N}_{\mathsf{n}}^i}\|_F \sum_{j\in\mathsf{N}_{\mathsf{n}}^i}\|B^{*ij}_{m_n k_n}\|_F,
$$

where $b_n^{*i} = \min_{j\in\mathsf{N}_{\mathsf{n}}^i}\|B^{*ij}_{m_n k_n}\|_F$.

Proof. The idea of the proof is similar as in the proof of Proposition 3 in Lee et al. (2020). Consider the sphere $\mathcal{S}_n(\delta_n) = \{B \in \mathbb{R}^{n^i k_n m_n \times m_n} : \|B - \tilde{B}_n^{\mathsf{N}_{\mathsf{n}}^i}\|_F = \delta_n\}$, where $(\delta_n)_{n\in\mathbb{N}}$ is a positive sequence of real numbers. For $\epsilon \in [0, 1]$ let

$$
f(\epsilon) = \widehat{PL}_{\mathsf{N}_{\mathsf{n}}^i}(\tilde{B}_n^{\mathsf{N}_{\mathsf{n}}^i} + \epsilon A, \hat{\xi}^i),
$$

where the function $\widehat{PL}_{\mathsf{N}_{\mathsf{n}}^i}$ is defined in (8.43) and $A = B - \tilde{B}_n^{\mathsf{N}_{\mathsf{n}}^i}$. A straightforward calculation gives for the first and the second derivatives of the function $f(\epsilon)$

$$\begin{aligned}\dot f(\epsilon) &= -\langle \hat\Sigma^n_{\mathsf{N}^i_n \xi^i}, A\rangle_F + \langle A, \hat\Sigma^n_{\mathsf{N}^i_n\mathsf{N}^i_n}(\tilde B^{\mathsf{N}^i_n}_n + \epsilon A)\rangle_F \\ &\quad + \lambda_n \sum_{j\in \mathsf{N}^i_n} \|\tilde B^{ij}_n + \epsilon A^j\|_F \sum_{k\in\mathsf{N}^i_n}\Big(\langle \tilde B^{ik}_n + \epsilon A^k, A^k\rangle_F \|\tilde B^{ik}_n + \epsilon A^k\|_F^{-1}\Big). \\ \ddot f(\epsilon) &= \langle A, \hat\Sigma^n_{\mathsf{N}^i_n\mathsf{N}^i_n} A\rangle_F + \lambda_n \Big(\sum_{j\in\mathsf{N}^i_n} \|\tilde B^{ij}_n + \epsilon A^j\|_F^{-1}\langle \tilde B^{ij}_n + \epsilon A^j, A^j\rangle_F\Big)^2 \\ &\quad + \lambda_n \sum_{j\in\mathsf{N}^i_n} \|\tilde B^{ij}_n + \epsilon A^j\|_F \sum_{k\in\mathsf{N}^i_n} \|A^k\|_F^2 \|\tilde B^{ik}_n + \epsilon A^k\|_F^{-1} \\ &\quad - \lambda_n \sum_{j\in\mathsf{N}^i_n} \|\tilde B^{ij}_n + \epsilon A^j\|_F \sum_{k\in\mathsf{N}^i_n}\Big(\langle \tilde B^{ik}_n + \epsilon A^k, A^k\rangle_F^2 \|\tilde B^{ik}_n + \epsilon A^k\|_F^{-3}\Big).\end{aligned}$$

where $\langle\cdot,\cdot\rangle_F$ denotes the Frobenious inner product and $\hat\Sigma^n_{\mathsf{N}^i_n\xi^i}$, $\hat\Sigma^n_{\mathsf{N}^i_n\mathsf{N}^i_n}$ are defined in (8.40) and (8.7) respectively. By construction, $f(0) = \widehat{PL}_{\mathsf{N}^i_n}(\tilde B^{\mathsf{N}^i_n}_n, \hat\xi^i)$, $f(1) = \widehat{PL}_{\mathsf{N}^i_n}(B, \hat\xi^i)$ and by Taylor's theorem, we have for some $\epsilon \in (0,1)$

$$\widehat{PL}_{\mathsf{N}^i_n}(B, \hat\xi^i) - \widehat{PL}_{\mathsf{N}^i_n}(\tilde B^{\mathsf{N}^i_n}_n, \hat\xi^i) = f(1) - f(0) = \dot f(0) + \frac{\ddot f(\epsilon)}{2}. \tag{8.64}$$

The Cauchy-Schwarz inequality yields for any $B \in \mathcal{S}_n(\delta_n)$ and $\epsilon\in[0,1]$

$$\ddot f(\epsilon) \geq \langle A, \hat\Sigma^n_{\mathsf{N}^i_n\mathsf{N}^i_n} A\rangle_F + \lambda_n\Big(\sum\nolimits_{j\in\mathsf{N}^i_n} \|\tilde B^{ij}_n + \epsilon A^j\|_F^{-1}\langle \tilde B^{ij}_n + \epsilon A^j, A^j\rangle_F\Big)^2 \geq \frac{C_{\min}}{2}\delta_n^2. \tag{8.65}$$

On the other hand, by Lemma A7 in Lee et al. (2016b) it follows that

$$\begin{aligned}|\dot f(0)| &\leq \lambda_n \sum_{j\in\mathsf{N}^i_n}\sum_{k\in\mathsf{N}^i_n}\Big[\|\tilde B^{ij}_n - B^{*ij}_{m_nk_n}\|_F \|\tilde B^{ik}_n - B^k\|_F \\ &\quad + \|B^{*ij}_{m_nk_n}\|_F \|B^{*ik}_{m_nk_n}\|_F^{-1} \|\tilde B^{ij}_n - B^{*ij}_{m_nk_n}\|_F \|\tilde B^{ik}_n - B^k\|_F\Big].\end{aligned}$$

A further application of the Cauchy-Schwarz inequality gives

$$\begin{aligned}|\dot f(0)| &\leq \lambda_n \sqrt{n^i \sum_{j\in\mathsf{N}^i_n} \|\tilde B^{ij}_n - B^{*ij}_{m_nk_n}\|_F^2 n^i \sum_{k\in\mathsf{N}^i_n} \|\tilde B^{ik}_n - B^k\|_F^2} \\ &\quad + \lambda_n \sum_{j\in\mathsf{N}^i_n} \|B^{*ij}_{m_nk_n}\|_F \|\tilde B^{ij}_n - B^{*ij}_{m_nk_n}\|_F (b^{*i}_n)^{-1} \sum_{k\in\mathsf{N}^i_n} \|\tilde B^{ik}_n - B^k\|_F \\ &= \lambda_n n^i \|\tilde B^{\mathsf{N}^i_n}_n - B^{*\mathsf{N}^i_n}_n\|_F \|\tilde B^{\mathsf{N}^i_n}_n - B\|_F \\ &\quad + \lambda_n (b^{*i}_n)^{-1} \sum_{j\in\mathsf{N}^i_n} \|B^{*ij}_{m_nk_n}\|_F \|\tilde B^{ij}_n - B^{*ij}_{m_nk_n}\|_F \sum_{k\in\mathsf{N}^i_n} \|\tilde B^{ik}_n - B^k\|_F \\ &\leq \lambda_n n^i \|\tilde B^{\mathsf{N}^i_n}_n - B^{*\mathsf{N}^i_n}_n\|_F \|\tilde B^{\mathsf{N}^i_n}_n - B\|_F \\ &\quad + \lambda_n \sqrt{n^i} (b^{*i}_n)^{-1} \|\tilde B^{\mathsf{N}^i_n}_n - B^{*\mathsf{N}^i_n}_n\|_F \|\tilde B^{\mathsf{N}^i_n}_n - B\|_F \|B^{*\mathsf{N}^i_n}_n\|_F.\end{aligned}$$

Using the fact $b_n^{*i} \le \|B_n^{*\mathsf{N}_n^i}\|_F \le \sum_{j\in\mathsf{N}_n^i} \|B_{m_nk_n}^{*ij}\|_F$ we obtain

$$|\dot{f}(0)| \le 2\lambda_n n^i (b_n^{*i})^{-1} \sum_{j\in\mathsf{N}_n^i} \|B_{m_nk_n}^{*ij}\|_F \|\tilde{B}_n^{\mathsf{N}_n^i} - B_n^{*\mathsf{N}_n^i}\|_F \|\tilde{B}_n^{\mathsf{N}_n^i} - B\|_F. \tag{8.66}$$

Hence, combining (8.64), (8.65) and (8.66), we obtain

$$\widehat{PL}_{\mathsf{N}_n^i}(B,\hat{\xi}^i) - \widehat{PL}_{\mathsf{N}_n^i}(\tilde{B}_n^{\mathsf{N}_n^i},\hat{\xi}^i) \ge -2\lambda_n n^i (b_n^{*i})^{-1} \sum_{j\in\mathsf{N}_n^i} \|B_{m_nk_n}^{*ij}\|_F \|\tilde{B}_n^{\mathsf{N}_n^i} - B_n^{*\mathsf{N}_n^i}\|_F + \frac{C_{\min}}{4}\delta_n^2$$

If we choose $\delta_n^2 = \frac{10}{C_{\min}}\lambda_n n^i (b_n^{*i})^{-1} \sum_{j\in\mathsf{N}_n^i} \|B_{m_nk_n}^{*ij}\|_F \|\tilde{B}_n^{\mathsf{N}_n^i} - B_n^{*\mathsf{N}_n^i}\|_F$, it follows that

$$\widehat{PL}_{\mathsf{N}_n^i}(B,\hat{\xi}^i) - \widehat{PL}_{\mathsf{N}_n^i}(\tilde{B}_n^{\mathsf{N}_n^i},\hat{\xi}^i) > 0.$$

Since the function $\widehat{PL}_{\mathsf{N}_n^i}(B,\hat{\xi}^i)$ is convex, the minimizer $\hat{B}_n^{\mathsf{N}_n^i}$ of $\widehat{PL}_{\mathsf{N}_n^i}(B,\hat{\xi}^i)$ is going be inside the sphere defined by $\mathcal{S}_n(\delta_n)$, that is,

$$\|\hat{B}_n^{\mathsf{N}_n^i} - \tilde{B}_n^{\mathsf{N}_n^i}\|_F \le \frac{10}{C_{\min}}\lambda_n n^i (b_n^{*i})^{-1} \sum_{j\in\mathsf{N}_n^i} \|B_{m_nk_n}^{*ij}\|_F \|\tilde{B}_n^{\mathsf{N}_n^i} - B_n^{*\mathsf{N}_n^i}\|_F.$$

□

Using the Propositions 8.3 and 8.4, we now can establish the concentration bounds for $\|\hat{B}_n^{\mathsf{N}_n^i} - B_n^{*\mathsf{N}_n^i}\|_F$.

Proposition 8.5 *Suppose Assumptions of Proposition 8.3 are satisfied and that δ satisfies*

$$\frac{2}{C_{\min}}\lambda_n (n^i)^{3/2} \Big(\sum_{j\in\mathsf{N}_n^i} \|B_{m_nk_n}^{*ij}\|_F\Big)^2 \le c_2 b_n^{*i}\delta \tag{8.67}$$

for some constant $c_2 > 0$. Then,

$$P\Big(\|\hat{B}_n^{\mathsf{N}_n^i} - B_n^{*\mathsf{N}_n^i}\|_F \ge \delta\Big) \lesssim \exp\Big(-C_1 \frac{n^{1-\alpha(2+3\beta)}(b_n^{*i})^2\delta^2}{(n^i)^4 m_n^2 k_n^4 (\sum_{j\in\mathsf{N}_n^i} \|B_{m_nk_n}^{*ij}\|_F)^2} + 2\log(n^i m_n k_n)\Big),$$

where $C_1 > 0$ such that $0 < \delta \le C_1$.

Proof. By Proposition 8.4 and the triangle inequality,

$$\begin{aligned}
\|\hat{B}_n^{\mathsf{N}_n^i} - B_n^{*\mathsf{N}_n^i}\|_F &\le \|\hat{B}_n^{\mathsf{N}_n^i} - \tilde{B}_n^{\mathsf{N}_n^i}\|_F + \|B_n^{*\mathsf{N}_n^i} - \tilde{B}_n^{\mathsf{N}_n^i}\|_F \\
&\le (b_n^{*i})^{-1}\|B_n^{*\mathsf{N}_n^i} - \tilde{B}_n^{\mathsf{N}_n^i}\|_F \Big(\frac{10}{C_{\min}}\lambda_n n^i \sum_{j\in\mathsf{N}_n^i} \|B_{m_nk_n}^{*ij}\|_F + b_n^{*i}\Big) \\
&\lesssim n^i (b_n^{*i})^{-1}\|B_n^{*\mathsf{N}_n^i} - \tilde{B}_n^{\mathsf{N}_n^i}\|_F \sum_{j\in\mathsf{N}_n^i} \|B_{m_nk_n}^{*ij}\|_F,
\end{aligned}$$

where we have used the fact that $b_n^{*i} \leq \sum_{j\in \mathsf{N}_\mathsf{n}^i} \|B^{*ij}_{m_n k_n}\|_F$ and $\lambda_n \lesssim 1$. The assertion now follows from Proposition 8.3 with δ replaced by $b_n^{*i}\delta\big(n^i \sum_{j\in \mathsf{N}_\mathsf{n}^i} \|B^{*ij}_{m_n k_n}\|_F\big)^{-1}$. □

Let $\hat{B}_n^{\mathsf{N}_\mathsf{n}^i}$ be the minimizer of the restricted problem (8.42). By construction, the estimator $(\hat{B}_n^{\mathsf{N}_\mathsf{n}^i}, \mathbf{0})$ obtained from $\hat{B}_n^{\mathsf{N}_\mathsf{n}^i}$ by adding blocks with 0 elements whenever $j \notin \mathsf{N}_\mathsf{n}^i$, satisfies the first KKT-condition (8.41a). To prove that $(\hat{B}_n^{\mathsf{N}_\mathsf{n}^i}, \mathbf{0})$ is, with high probability, optimal for problem (8.42), it is therefore sufficient to show that the second KKT-condition (8.41b) is satisfied. This is the statement of the following proposition.

Proposition 8.6 *The matrix* $(\hat{B}_n^{\mathsf{N}_\mathsf{n}^i}, \mathbf{0})$ *satisfies* (8.41b) *with high probability, in the sense that*

$$P(\max_{j\notin \mathsf{N}_\mathsf{n}^i} \|\hat{\Sigma}^n_{\xi^j \mathsf{N}_\mathsf{n}^i} \hat{B}_n^{\mathsf{N}_\mathsf{n}^i} - \hat{\Sigma}^n_{\xi^j \xi^i}\|_F \geq \lambda_n \sum_{j\neq i}^{p} \|\hat{B}_n^{ij}\|_F)$$
$$\lesssim \exp\Big(-C_1 \frac{n^{1-\alpha(2+3\beta)}(\lambda_n \sum_{j\in \mathsf{N}_\mathsf{n}^i} \|B^{*ij}_{m_n k_n}\|_F)^2}{n^i m_n^2 k_n^4} + 2\log(n^i m_n k_n)\Big),$$

where C_1 is a positive constant.

Proof. The idea of the proof is similar as in the proof of Proposition 4 in Lee et al. (2020). By the first optimality condition (8.41a), we have for all $j \in \mathsf{N}_\mathsf{n}^i$,

$$\hat{B}_n^{\mathsf{N}_\mathsf{n}^i} = (\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i \mathsf{N}_\mathsf{n}^i} + \lambda_n \hat{D}_{\mathsf{N}_\mathsf{n}^i})^{-1} \hat{\Sigma}^n_{\mathsf{N}^i \xi^i}, \tag{8.68}$$

where $\hat{D}_{\mathsf{N}_\mathsf{n}^i}$ is defined in Lemma 8.4. Using (8.68) in the expression at the left-hand side of condition (8.41b) gives

$$\hat{\Sigma}^n_{\xi^j \mathsf{N}_\mathsf{n}^i} \hat{B}_n^{\mathsf{N}_\mathsf{n}^i} - \hat{\Sigma}^n_{\xi^j \xi^i} = \hat{\Sigma}^n_{\xi^j \mathsf{N}_\mathsf{n}^i} (\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i \mathsf{N}_\mathsf{n}^i} + \lambda_n \hat{D}_{\mathsf{N}_\mathsf{n}^i})^{-1} \hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i \xi^i} - \hat{\Sigma}^n_{\xi^j \xi^i} = R_1^j + \ldots + R_7^j,$$

where

$$\begin{aligned}
R_1^j &= (\hat{\Sigma}^n_{\xi^j \mathsf{N}_\mathsf{n}^i} - \Sigma^n_{\xi^j \mathsf{N}_\mathsf{n}^i})(\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i \mathsf{N}_\mathsf{n}^i} + \lambda_n \hat{D}_{\mathsf{N}_\mathsf{n}^i})^{-1} \Sigma^n_{\mathsf{N}_\mathsf{n}^i \xi^i} \\
R_2^j &= \Sigma^n_{\xi^j \mathsf{N}_\mathsf{n}^i} (\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i \mathsf{N}_\mathsf{n}^i} + \lambda_n \hat{D}_{\mathsf{N}_\mathsf{n}^i})^{-1} (\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i \xi^i} - \Sigma^n_{\mathsf{N}_\mathsf{n}^i \xi^i}) \\
R_3^j &= \Sigma^n_{\xi^j \mathsf{N}_\mathsf{n}^i} \{(\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i \mathsf{N}_\mathsf{n}^i} + \lambda_n \hat{D}_{\mathsf{N}_\mathsf{n}^i})^{-1} - (\Sigma^n_{\mathsf{N}_\mathsf{n}^i \mathsf{N}_\mathsf{n}^i} + \lambda_n \hat{D}_{\mathsf{N}_\mathsf{n}^i})^{-1}\} \Sigma^n_{\mathsf{N}_\mathsf{n}^i \xi^i} \\
R_4^j &= (\hat{\Sigma}^n_{\xi^j \mathsf{N}_\mathsf{n}^i} - \Sigma^n_{\xi^j \mathsf{N}_\mathsf{n}^i})(\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i \mathsf{N}_\mathsf{n}^i} + \lambda_n \hat{D}_{\mathsf{N}_\mathsf{n}^i})^{-1} (\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i \xi^i} - \Sigma^n_{\mathsf{N}_\mathsf{n}^i \xi^i}) \\
R_5^j &= \Sigma^n_{\xi^j \xi^i} - \hat{\Sigma}^n_{\xi^j \xi^i} \\
R_6^j &= \Sigma^n_{\xi^j \mathsf{N}_\mathsf{n}^i} \{(\Sigma_{\mathsf{N}_\mathsf{n}^i \mathsf{N}_\mathsf{n}^i} + \lambda_n \hat{D}_{\mathsf{N}_\mathsf{n}^i})^{-1} - (\Sigma^n_{\mathsf{N}_\mathsf{n}^i \mathsf{N}_\mathsf{n}^i} + \lambda_n D^*_{\mathsf{N}_\mathsf{n}^i})^{-1}\} \Sigma^n_{\mathsf{N}_\mathsf{n}^i \xi^i} \\
R_7^j &= \Sigma^n_{\xi^j \mathsf{N}_\mathsf{n}^i} (\Sigma^n_{\mathsf{N}_\mathsf{n}^i \mathsf{N}_\mathsf{n}^i} + \lambda_n D^*_{\mathsf{N}_\mathsf{n}^i})^{-1} \Sigma^n_{\mathsf{N}_\mathsf{n}^i \xi^i} - \Sigma^n_{\xi^j \xi^i}.
\end{aligned}$$

In the following we derive bounds for the probabilities

$$P\Big(\max_{j\notin \mathsf{N}_\mathsf{n}^i}\|R_r^j\|_F \geq \frac{\lambda_n}{7}\sum_{j\in \mathsf{N}_\mathsf{n}^i}\|\hat{B}_n^{ij}\|_F\Big) \tag{8.69}$$

$(r = 1, \dots, 7)$. For this purpose we proceed in two steps.

Step 1: First, we define the event, that there exists a constant $0 < c_0 < 1$, such that

$$\mathcal{A}_0 = \Big\{\|\hat{B}_n^{\mathsf{N}_\mathsf{n}^i} - B_n^{*\mathsf{N}_\mathsf{n}^i}\|_F \leq c_0(n^i)^{-1/2}\sum_{j\in \mathsf{N}_\mathsf{n}^i}\|B_{m_nk_n}^{*ij}\|_F\Big\}.$$

Then, by Proposition 8.5 with $\delta = c_0(n^i)^{-1/2}\sum_{j\in \mathsf{N}^i}\|B_{m_nk_n}^{*}\|_F$ we have

$$\begin{aligned} P(\mathcal{A}_0^\complement) &\lesssim \exp\Big(-C_1\frac{n^{1-\alpha(2+3\beta)}(b_n^{*i})^2}{(n^i)^5m_n^2k_n^4} + 2\log(n^im_nk_n)\Big) \\ &\lesssim \exp\Big(-C_1\frac{n^{1-\alpha(2+3\beta)}(\lambda_n\sum_{j\in \mathsf{N}_\mathsf{n}^i}\|B_{m_nk_n}^{*ij}\|_F)^2}{n^im_n^2k_n^4} + 2\log(n^im_nk_n)\Big), \end{aligned} \tag{8.70}$$

for some constant $C_1 > 0$.

Now, on the event $\mathcal{A}_0$, it follows by the Cauchy-Schwarz inequality that

$$\Big|\sum_{j\in \mathsf{N}_\mathsf{n}^i}\|\hat{B}_n^{ij}\|_F - \sum_{j\in \mathsf{N}_\mathsf{n}^i}\|B_{m_nk_n}^{*ij}\|_F\Big| \leq c_0\sum_{j\in \mathsf{N}_\mathsf{n}^i}\|B_{m_nk_n}^{*ij}\|_F.$$

Therefore, we have for $r = 1, \dots, 7$

$$P\Big(\max_{j\notin \mathsf{N}_\mathsf{n}^i}\|R_r^j\|_F \geq \frac{\lambda_n}{7}\sum_{j\in \mathsf{N}_\mathsf{n}^i}\|\hat{B}_n^{ij}\|_F\Big) \leq P\Big(\max_{j\notin \mathsf{N}_\mathsf{n}^i}\|R_r^j\|_F \geq \frac{\lambda_n(1-c_0)}{7}\sum_{j\in \mathsf{N}_\mathsf{n}^i}\|B_{m_nk_n}^{*ij}\|_F\Big) + P(\mathcal{A}_0^c),$$

where the second probability can be bounded by (8.70). Thus, in order to control the probabilities of the terms $\max_{j\notin \mathsf{N}_\mathsf{n}^i}\|R_r^j\|_F \geq \frac{\lambda_n}{7}\sum_{j\in \mathsf{N}_\mathsf{n}^i}\|\hat{B}_n^{ij}\|_F$ it suffices to derive the probabilities $P(\max_{j\notin \mathsf{N}_\mathsf{n}^i}\|R_r^j\|_F \geq c_1\lambda_n\sum_{j\in \mathsf{N}_\mathsf{n}^i}\|B_{m_nk_n}^{*ij}\|_F)$ for all $r = 1, \dots, 7$, where $c_1 = \frac{1-c_0}{7}$.

Step 2: *Term* R_1^j. Substituting the representation (8.50) used in the proof of Proposition 8.3 we can rewrite R_1^j as

$$R_1^j =\leq (\hat{\Sigma}^n_{\xi^j\mathsf{N}_\mathsf{n}^i} - \Sigma^n_{\xi^j\mathsf{N}_\mathsf{n}^i})(\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i} + \lambda_n\hat{D}_{\mathsf{N}^i})^{-1}\Big(\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}B_n^{*\mathsf{N}_\mathsf{n}^i} + \frac{\tilde{\mathbf{H}}_\mathbf{n}(\xi^{\mathsf{N}_\mathsf{n}^i})}{n}w^i + \frac{\tilde{\mathbf{H}}_\mathbf{n}(\xi^{\mathsf{N}_\mathsf{n}^i})}{n}\epsilon^i\Big),$$

and by the union bound it follows that

$$P\Big(\max_{j\notin \mathsf{N}_\mathsf{n}^i}\|R_1^j\|_F \geq c_1\lambda_n\sum_{j\in \mathsf{N}_\mathsf{n}^i}\|B_{m_nk_n}^{*ij}\|_F\Big) \leq \sum_{r=1}^3 P\Big(\max_{j\notin \mathsf{N}_\mathsf{n}^i}\|R_{1r}^j\|_F \geq \frac{c_1\lambda_n}{3}\sum_{j\in \mathsf{N}_\mathsf{n}^i}\|B_{m_nk_n}^{*ij}\|_F\Big), \tag{8.71}$$

where R^j_{11}, R^j_{12}, R^j_{13} are defined in an obvious manner satisfying

$$\begin{aligned}
\|R^j_{11}\|_F &\leq \|\hat{\Sigma}^n_{\xi^j \mathsf{N}^i_n} - \Sigma^n_{\xi^j \mathsf{N}^i_n}\|_F \|(\hat{\Sigma}^n_{\mathsf{N}^i_n\mathsf{N}^i_n} + \lambda_n \hat{D}_{\mathsf{N}^i_n})^{-1} \Sigma^n_{\mathsf{N}^i_n\mathsf{N}^i_n} B^{*\mathsf{N}^i_n}_n\|_2. \\
\|R^j_{12}\|_F &\leq \|\hat{\Sigma}^n_{\xi^j \mathsf{N}^i_n} - \Sigma^n_{\xi^j \mathsf{N}^i_n}\|_F \|(\hat{\Sigma}^n_{\mathsf{N}^i_n\mathsf{N}^i_n} + \lambda_n \hat{D}_{\mathsf{N}^i_n})^{-1} \frac{\tilde{\mathbf{H}}_{\mathbf{n}}(\xi^{\mathsf{N}^i_n})}{n} w^i\|_2. \\
\|R^j_{13}\|_F &\leq \|\hat{\Sigma}^n_{\xi^j \mathsf{N}^i_n} - \Sigma^n_{\xi^j \mathsf{N}^i_n}\|_F \|(\hat{\Sigma}^n_{\mathsf{N}^i_n\mathsf{N}^i_n} + \lambda_n \hat{D}_{\mathsf{N}^i_n})^{-1} \frac{\tilde{\mathbf{H}}_{\mathbf{n}}(\xi^{\mathsf{N}^i_n})}{n} \epsilon^i\|_2.
\end{aligned}$$

Next we derive bounds for the probabilities on the right hand side of (8.71) starting with the term $\|R^j_{11}\|_F$. Observing that

$$(\hat{\Sigma}^n_{\mathsf{N}^i_n\mathsf{N}^i_n} - \lambda_n \hat{D}_{\mathsf{N}^i_n})^{-1} \Sigma^n_{\mathsf{N}^i_n\mathsf{N}^i_n} = (\hat{\Sigma}^n_{\mathsf{N}^i_n\mathsf{N}^i_n} + \lambda_n \hat{D}_{\mathsf{N}^i_n})^{-1} (\Sigma^n_{\mathsf{N}^i_n\mathsf{N}^i_n} - \hat{\Sigma}^n_{\mathsf{N}^i_n\mathsf{N}^i_n}) + (\hat{\Sigma}^n_{\mathsf{N}^i_n\mathsf{N}^i_n} + \lambda_n \hat{D}_{\mathsf{N}^i_n})^{-1} \hat{\Sigma}^n_{\mathsf{N}^i_n\mathsf{N}^i_n}.$$

we have

$$\begin{aligned}
\|(\hat{\Sigma}^n_{\mathsf{N}^i_n\mathsf{N}^i_n} + \lambda_n \hat{D}_{\mathsf{N}^i_n})^{-1} \Sigma^n_{\mathsf{N}^i_n\mathsf{N}^i_n}\|_2 &\leq \|(\hat{\Sigma}^n_{\mathsf{N}^i_n\mathsf{N}^i_n} + \lambda_n \hat{D}_{\mathsf{N}^i_n})^{-1} (\hat{\Sigma}^n_{\mathsf{N}^i_n\mathsf{N}^i_n} - \Sigma^n_{\mathsf{N}^i_n\mathsf{N}^i_n})\|_2 + \|(\hat{\Sigma}^n_{\mathsf{N}^i_n\mathsf{N}^i_n} + \lambda_n \hat{D}_{\mathsf{N}^i_n})^{-1} \hat{\Sigma}^n_{\mathsf{N}^i_n\mathsf{N}^i_n}\|_2 \\
&\leq \frac{4}{C_{\min}} \|\hat{\Sigma}^n_{\mathsf{N}^i_n\mathsf{N}^i_n} - \Sigma^n_{\mathsf{N}^i_n\mathsf{N}^i_n}\|_F + 1, \qquad (8.72)
\end{aligned}$$

where we used the fact that $\hat{\Sigma}^n_{\mathsf{N}^i_n\mathsf{N}^i_n} \preceq \hat{\Sigma}^n_{\mathsf{N}^i_n\mathsf{N}^i_n} + \lambda_n I_{\mathsf{N}^i_n} \preceq \hat{\Sigma}^n_{\mathsf{N}^i_n\mathsf{N}^i_n} + \lambda_n \hat{D}_{\mathsf{N}^i_n}$. From this, it follows that

$$\max_{j \notin \mathsf{N}^i_n} \|R^j_{11}\|_F \lesssim \max_{j \notin \mathsf{N}^i_n} \{\|\hat{\Sigma}^n_{\xi^j \mathsf{N}^i_n} - \Sigma^n_{\xi^j \mathsf{N}^i_n}\|_F\} \Big(\frac{4}{C_{\min}} \|\hat{\Sigma}^n_{\mathsf{N}^i_n\mathsf{N}^i_n} - \Sigma^n_{\mathsf{N}^i_n\mathsf{N}^i_n}\|_F + 1 \Big),$$

where we use the fact that $\|B^{*\mathsf{N}^i_n}_n\|_2$ is bounded. Consider the event

$$\mathcal{A}_1 = \{ \max_{j \notin \mathsf{N}^i_n} \{\|\hat{\Sigma}^n_{\xi^j \mathsf{N}^i_n} - \Sigma^n_{\xi^j \mathsf{N}^i_n}\|_F\} \leq \frac{c_1 \lambda_n \sum_{l \in \mathsf{N}^i_n} \|B^{*il}_{m_n k_n}\|_F}{6} \}.$$

Then,

$$\begin{aligned}
P\Big(\max_{j \notin \mathsf{N}^i_n} \|R^j_{11}\|_F \geq \frac{c_1 \lambda_n}{3} \sum_{j \in \mathsf{N}^i_n} \|B^{*ij}_{m_n k_n}\|_F \Big) &\leq P\Big(\|\hat{\Sigma}^n_{\mathsf{N}^i_n\mathsf{N}^i_n} - \Sigma^n_{\mathsf{N}^i_n\mathsf{N}^i_n}\|_F \geq \frac{C_{\min}}{4} \Big) + P\Big(\mathcal{A}_1^{\complement} \Big), \\
&\lesssim \exp\Big(-C_2 \frac{n^{1-\alpha(2+3\beta)}}{(n^i)^2 m_n^2 k_n^4} + 2\log(n^i m_n k_n) \Big) \\
&+ \exp\Big(-C_2 \frac{n^{1-\alpha(2+3\beta)} (\lambda_n \sum_{l \in \mathsf{N}^i_n} \|B^{*il}_{m_n k_n}\|_F)^2}{n^i m_n^2 k_n^4} + \log(n^i m_n^2 k_n^2) + \log((p - n^i) m_n^2 k_n^2) \Big),
\end{aligned}$$

where we have used Theorem 8.1 and the union bound.

For the term R^j_{12}, we use the same arguments as for the term T_2 in the proof of Proposition 8.3. Specifically, recall the definition of the event Ω in (8.22) and the calculation in (8.49) to obtain on the event Ω

$$\begin{aligned}
\max_{j \notin \mathsf{N}^i_n} \|R^j_{12}\|_F &\leq \max_{j \notin \mathsf{N}^i_n} \|\hat{\Sigma}^n_{\xi^j \mathsf{N}^i_n} - \Sigma^n_{\xi^j \mathsf{N}^i_n}\|_F \Big(\frac{2}{C_{\min}} \|\hat{\Sigma}^n_{\mathsf{N}^i_n\mathsf{N}^i_n} - \Sigma^n_{\mathsf{N}^i_n\mathsf{N}^i_n}\|_F + 1 \Big) \Big(\frac{C_{\min}}{2} \Big)^{-1/2} c_2 \frac{n^i m_n^{3/2}}{k_n^d} \\
&\leq \max_{j \notin \mathsf{N}^i_n} \|\hat{\Sigma}^n_{\xi^j \mathsf{N}^i_n} - \Sigma^n_{\xi^j \mathsf{N}^i_n}\|_F \Big(\frac{4}{C_{\min}} \|\hat{\Sigma}^n_{\mathsf{N}^i_n\mathsf{N}^i_n} - \Sigma^n_{\mathsf{N}^i_n\mathsf{N}^i_n}\|_F + 1 \Big) \Big(\frac{4}{C_{\min}} \Big) c_2 \lambda_n \sum_{j \in \mathsf{N}^i_n} \|B^{*ij}_{m_n k_n}\|_F
\end{aligned}$$

for some constant $c_2 > 0$, where we used condition (4.13) for the last inequality.

Then, conditioning on the event $\mathcal{A}_2 = \{\max_{j \notin \mathsf{N}_n^i} \|\hat{\Sigma}^n_{\xi^j \mathsf{N}_n^i} - \Sigma^n_{\xi^j \mathsf{N}_n^i}\|_F \geq \frac{c_1 C_{\min}}{c_2 12}\}$, we have by Proposition 8.1

$$
\begin{aligned}
&P\Big(\max_{j \notin \mathsf{N}_n^i} \|R_{12}\| \geq \frac{c_1}{3} \lambda_n \sum_{j \in \mathsf{N}_n^i} \|B^{*ij}_{m_n k_n}\|_F\Big) \\
&\quad \leq P\Big(\|\hat{\Sigma}^n_{\mathsf{N}_n^i \mathsf{N}_n^i} - \Sigma^n_{\mathsf{N}_n^i \mathsf{N}_n^i}\|_F \geq \frac{C_{\min}}{4}\Big) + P\Big(\max_{j \notin \mathsf{N}_n^i} \|\hat{\Sigma}^n_{\xi^j \mathsf{N}_n^i} - \Sigma^n_{\xi^j \mathsf{N}_n^i}\|_F \geq \frac{c_1 C_{\min}}{c_2 12}\Big) + P(\Omega^{\complement}) \\
&\quad \lesssim \exp\Big(- C_3 \frac{n^{1-\alpha(2+3\beta)}}{(n^i)^2 m_n^2 k_n^4} + 2\log(n^i m_n k_n)\Big) \\
&\qquad + \exp\Big(- C_3 \frac{n^{1-\alpha(2+3\beta)}}{n^i m_n^2 k_n^4} + \log(n^i m_n^2 k_n^2) + \log((p - n^i) m_n^2 k_n^2)\Big) \\
&\qquad + \exp(-C_3 \frac{n}{n^i m_n^2 k_n^{2d}} + \log(n^i m_n^2)).
\end{aligned}
$$

For the term R_{13}^j, we have

$$
\max_{j \notin \mathsf{N}_n^i} \|R_{13}^j\|_F \leq \frac{2}{C_{\min}} \max_{j \notin \mathsf{N}_n^i} \|\hat{\Sigma}^n_{\xi^j \mathsf{N}_n^i} - \Sigma^n_{\xi^j \mathsf{N}_n^i}\|_F \|\frac{\tilde{\mathbf{H}}_n(\xi^{\mathsf{N}_n^i})}{n} \epsilon^i\|_F,
$$

which yields

$$
\begin{aligned}
&P\Big(\max_{j \notin \mathsf{N}_n^i} \|R_{13}^j\|_F \geq \frac{c_1}{3} \lambda_n \sum_{j \in \mathsf{N}_n^i} \|B^{*ij}_{m_n k_n}\|_F\Big) \\
&\quad \leq P\Big(\max_{j \notin \mathsf{N}_n^i} \|\hat{\Sigma}^n_{\xi^j \mathsf{N}_n^i} - \Sigma^n_{\xi^j \mathsf{N}_n^i}\|_F \geq \frac{c_1 \lambda_n \sum_{j \in \mathsf{N}_n^i} \|B^{*ij}_{m_n k_n}\|_F}{3}\Big) + P\Big(\|\frac{\tilde{\mathbf{H}}_n(\xi^{\mathsf{N}_n^i})}{n} \epsilon^i\|_F \geq \frac{C_{\min}}{2}\Big) \\
&\quad \lesssim \exp\Big(- C_4 \frac{n^{1-\alpha(2+3\beta)} (\lambda_n \sum_{j \in \mathsf{N}_n^i} \|B^{*ij}_{m_n k_n}\|_F)^2}{n^i m_n^2 k_n^4} + \log(n^i m_n^2 k_n^2) + \log((p - n^i) m_n^2 k_n^2)\Big) \\
&\qquad + \exp\Big(- C_4 \frac{n}{n^i m_n^2 k_n} + \log(n^i m_n^2 k_n)\Big),
\end{aligned}
$$

where we used Theorem 8.1 and (8.59). Combining together the results for the terms $\|R_{13}^j\|$, $\|R_{12}^j\|$ and $\|R_{13}^j\|$ we conclude that

$$
P\Big(\max_{j \notin \mathsf{N}_n^i} \|R_1^j\|_F \geq c_1 \lambda_n \sum_{j \in \mathsf{N}_n^i} \|B^{*ij}_{m_n k_n}\|_F\Big) \lesssim \exp\Big(- C_2 \frac{n^{1-\alpha(2+3\beta)} (\lambda_n \sum_{l \in \mathsf{N}_n^i} \|B^{*il}_{m_n k_n}\|_F)^2}{n^i m_n^2 k_n^4} + \log(p m_n^2 k_n^2)\Big).
$$

Term R_2^j. First we write

$$
\begin{aligned}
\max_{j \notin \mathsf{N}_n^i} \|R_2^j\|_F &\leq \max_{j \notin \mathsf{N}_n^i} \big\{\|\Sigma^n_{\xi^j \mathsf{N}_n^i} (\Sigma^n_{\mathsf{N}_n^i \mathsf{N}_n^i})^{-1}\|_F\big\} \|\Sigma^n_{\mathsf{N}_n^i \mathsf{N}_n^i} (\hat{\Sigma}^n_{\mathsf{N}_n^i \mathsf{N}_n^i} + \lambda_n \hat{D}_{\mathsf{N}_n^i})^{-1}\|_2 \|\hat{\Sigma}^n_{\mathsf{N}_n^i \xi^i} - \Sigma^n_{\mathsf{N}_n^i \xi^i}\|_F \\
&\leq \frac{1-\eta}{\sqrt{n^i}} \|\Sigma^n_{\mathsf{N}_n^i \mathsf{N}_n^i} (\hat{\Sigma}^n_{\mathsf{N}_n^i \mathsf{N}_n^i} + \lambda_n \hat{D}_{\mathsf{N}_n^i})^{-1}\|_2 \|\hat{\Sigma}^n_{\mathsf{N}_n^i \xi^i} - \Sigma^n_{\mathsf{N}_n^i \xi^i}\|_F \\
&\leq \frac{1-\eta}{\sqrt{n^i}} (\frac{4}{C_{\min}} \|\hat{\Sigma}^n_{\mathsf{N}_n^i \mathsf{N}_n^i} - \Sigma^n_{\mathsf{N}_n^i \mathsf{N}_n^i}\|_F + 1) \|\hat{\Sigma}^n_{\mathsf{N}_n^i \xi^i} - \Sigma^n_{\mathsf{N}_n^i \xi^i}\|_F,
\end{aligned}
$$

where we used (8.72) and the second inequality holds with high probability by Lemma 8.3. As a result, it follows that

$$
\begin{aligned}
&P\Big(\max_{j\notin \mathsf{N}_\mathsf{n}^i}\{\|R_2\|_F \geq c_1\lambda_n \sum_{j\in\mathsf{N}_\mathsf{n}^i}\|B^{*ij}_{m_nk_n}\|_F\}\Big)\\
&\qquad\leq P\Big(\frac{1-\eta}{\sqrt{n^i}}(\frac{4}{C_{\min}}\|\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}-\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}\|_F+1)\|\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i\xi^i}-\Sigma^n_{\mathsf{N}_\mathsf{n}^i\xi^i}\|_F \geq c_1\lambda_n\sum_{j\in\mathsf{N}_\mathsf{n}^i}\|B^{*ij}_{m_nk_n}\|_F\Big)\\
&\qquad\leq P\Big(\|\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}-\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}\|_F\geq\frac{C_{\min}}{4}\Big)+P(\mathcal{A}_2^{\complement})\\
&\qquad\lesssim \exp\left(-C_3\frac{n^{1-\alpha(2+3\beta)}}{(n^i)^2m_n^2k_n^4}+2\log(n^im_nk_n)\right)\\
&\qquad\qquad+\exp\Big(-C_3\frac{n^{1-\alpha(2+3\beta)}(\lambda_n\sum_{j\in\mathsf{N}_\mathsf{n}^i}\|B^{*ij}_{m_nk_n}\|_F)^2}{m_n^2k_n^3}+\log(n^im_n^2k_n)\Big),
\end{aligned}
$$

where

$$
\mathcal{A}_2=\Big\{\max_{j\notin\mathsf{N}_\mathsf{n}^i}\|\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i\xi^i}-\Sigma^n_{\mathsf{N}_\mathsf{n}^i\xi^i}\|_F\leq\frac{c_1}{2}\frac{\sqrt{n^i}}{1-\eta}\lambda_n\sum_{j\in\mathsf{N}_\mathsf{n}^i}\|B^{*ij}_{m_nk_n}\|_F\Big\},
$$

and we used Theorem 8.1 and the union bound for the last inequality.

Term R_3^j. Using the identity $A^{-1}-B^{-1}=A^{-1}(B-A)B^{-1}$ and following similar arguments used to obtain bounds for the terms R_1 and R_2, we get (note that we are working on the event $\mathcal{N}$)

$$
P\Big(\max_{j\notin\mathsf{N}_\mathsf{n}^i}\|R_3^j\|_F\geq c_1\lambda_n\sum_{j\in\mathsf{N}_\mathsf{n}^i}\|B^{*ij}_{m_nk_n}\|_F\Big)\leq\sum_{r=1}^3P\Big(\max_{j\notin\mathsf{N}^i}\|R^j_{3r}\|_F\geq\frac{c_1}{3}\lambda_n\sum_{j\in\mathsf{N}_\mathsf{n}^i}\|B^{*ij}_{m_nk_n}\|_F\Big),
$$

where,

$$
\begin{aligned}
\|R^j_{31}\|_F&\leq\frac{1-\eta}{\sqrt{n^i}}\Big(1+\frac{4}{C_{\min}}\|\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}-\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}\|_F\Big)\|\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}-\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}\|_F\|(\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}+\lambda_n\hat{D}_{\mathsf{N}_\mathsf{n}^i})^{-1}\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}B_n^{*\mathsf{N}_\mathsf{n}^i}\|_F,\\
\|R^j_{32}\|_F&\leq\frac{1-\eta}{\sqrt{n^i}}\Big(1+\frac{4}{C_{\min}}\|\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}-\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}\|_F\Big)\|\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}-\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}\|_F\|(\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}+\lambda_n\hat{D}_{\mathsf{N}_\mathsf{n}^i})^{-1}\frac{\tilde{\mathbf{H}}_n(\xi^{\mathsf{N}_\mathsf{n}^i})}{n}w^i\|_F,\\
\|R^j_{33}\|_F&\leq\frac{1-\eta}{\sqrt{n^i}}\Big(1+\frac{4}{C_{\min}}\|\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}-\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}\|_F\Big)\|\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}-\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}\|_F\|(\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}+\lambda_n\hat{D}_{\mathsf{N}_\mathsf{n}^i})^{-1}\frac{\tilde{\mathbf{H}}_n(\xi^{\mathsf{N}_\mathsf{n}^i})}{n}\epsilon^i\|_F,
\end{aligned}
$$

(note that the terms on the right-hand side are independent of j). We next derive bounds for the probabilities $P\left(\|R^j_{3r}\|_F\geq\frac{c_1}{3}\lambda_n\sum_{j\in\mathsf{N}_\mathsf{n}^i}\|B^{*ij}_{m_nk_n}\|_F\right)$ for $r=1,2,3$. For the term $\|R_{31}\|_F$, note that

$$
\|R^j_{31}\|_F\lesssim\frac{(1-\eta)}{\sqrt{n^i}}\|\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}-\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}\|_F\Big(\frac{4}{C_{\min}}\|\hat{\Sigma}^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}-\Sigma^n_{\mathsf{N}_\mathsf{n}^i\mathsf{N}_\mathsf{n}^i}\|_F+1\Big)\ ,
$$

which implies (using the same arguments as before)

$$P\Big(\max_{j\notin \mathsf{N}^i}\|R_{31}^j\|_F\geq \frac{c_1}{3}\lambda_n\sum_{j\in\mathsf{N}_n^i}\|B_{m_nk_n}^{*ij}\|_F\Big)$$
$$\leq P\Big(\|\Sigma^n_{\mathsf{N}_n^i\mathsf{N}_n^i}-\hat{\Sigma}^n_{\mathsf{N}_n^i\mathsf{N}_n^i}\|_F\geq \frac{\sqrt{n^i}\lambda_n\sum_{j\in\mathsf{N}_n^i}\|B_{m_nk_n}^{*ij}\|_F}{1-\eta}\frac{c_1}{6}\Big)+P\Big(\|\Sigma^n_{\mathsf{N}_n^i\mathsf{N}_n^i}-\hat{\Sigma}^n_{\mathsf{N}_n^i\mathsf{N}_n^i}\|_F\geq\frac{C_{\min}}{4}\Big)$$
$$\lesssim \exp\Big(-C_4\frac{n^{1-\alpha(2+3\beta)}(\lambda_n\sum_{j\in\mathsf{N}_n^i}\|B_{m_nk_n}^{*ij}\|_F)^2}{n^im_n^2k_n^4}+2\log(n^im_nk_n)\Big)$$
$$+\exp\Big(-C_4\frac{n^{1-\alpha(2+3\beta)}}{(n^i)^2m_n^2k_n^4}+2\log(n^im_nk_n)\Big),$$

for some positive constant C_4 that depends on η and $C_{\min}$. Similarly, for the term $\|R_{32}^j\|_F$ have

$$\|R_{32}^j\|_F\leq\frac{1-\eta}{\sqrt{n^i}}\|\Sigma^n_{\mathsf{N}_n^i\mathsf{N}_n^i}-\hat{\Sigma}^n_{\mathsf{N}_n^i\mathsf{N}_n^i}\|_F(\frac{4}{C_{\min}}\|\hat{\Sigma}^n_{\mathsf{N}_n^i\mathsf{N}_n^i}-\Sigma^n_{\mathsf{N}_n^i\mathsf{N}_n^i}\|_F+1)\|\frac{w^i}{\sqrt{n}}\|_F.$$

Hence, recalling the definition of the set Ω in (8.22) and using conditions (4.13) and (8.56), Proposition 8.1 and Theorem 8.1, we obtain for some positive constant

$$P\Big(\max_{j\notin \mathsf{N}^i}\|R_{32}^j\|_F\geq \frac{c_1}{3}\lambda_n\sum_{j\in\mathsf{N}_n^i}\|B_{m_nk_n}^{*ij}\|_F\Big)$$
$$\leq P\Big(\|\Sigma^n_{\mathsf{N}_n^i\mathsf{N}_n^i}-\hat{\Sigma}^n_{\mathsf{N}_n^i\mathsf{N}_n^i}\|_F\geq\frac{\sqrt{n^i}c_1}{12c_2(1-\eta)}\Big)+P\Big(\|\Sigma^n_{\mathsf{N}_n^i\mathsf{N}_n^i}-\hat{\Sigma}^n_{\mathsf{N}_n^i\mathsf{N}_n^i}\|_F\geq\frac{C_{\min}}{4}\Big)+P(\Omega^{\complement})$$
$$\lesssim \exp\Big(-C_5\frac{n^{1-\alpha(2+3\beta)}}{n^im_n^2k_n^4}+2\log(n^im_nk_n)\Big)+\exp\Big(-C_5\frac{n^{1-\alpha(2+3\beta)}}{(n^i)^2m_n^2k_n^4}+2\log(n^im_nk_n)\Big)$$
$$+\exp(-C_3\frac{n}{n^im_n^2k_n^{2d}}+\log(n^im_n^2)),$$

for some positive constant C_5 that depends on η and $C_{\min}$. For the term $\|R_{33}^j\|_F$, we write

$$\|R_{33}^j\|_F\leq\frac{(1-\eta)}{\sqrt{n^i}}(\frac{2}{C_{\min}})^{1/2}\Big(1+\frac{4}{C_{\min}}\|\Sigma^n_{\mathsf{N}_n^i\mathsf{N}_n^i}-\hat{\Sigma}^n_{\mathsf{N}_n^i\mathsf{N}_n^i}\|_F\Big)\|\Sigma^n_{\mathsf{N}_n^i\mathsf{N}_n^i}-\hat{\Sigma}^n_{\mathsf{N}_n^i\mathsf{N}_n^i}\|_F\|\frac{\tilde{\mathbf{H}}_{\mathbf{n}}(\xi^{\mathsf{N}_n^i})}{n}\epsilon^i\|_F,$$

and consider the event $\mathcal{A}_3=\{\|\Sigma^n_{\mathsf{N}_n^i\mathsf{N}_n^i}-\hat{\Sigma}^n_{\mathsf{N}_n^i\mathsf{N}_n^i}\|_F\leq\frac{c_1\sqrt{n^i}}{1-\eta}\lambda_n\sum_{j\in\mathsf{N}_n^i}\|B_{m_nk_n}^{*ij}\|_F\}$. Then

$$P\Big(\max_{j\notin \mathsf{N}^i}\|R_{33}^j\|_F\geq \frac{c_1}{3}\lambda_n\sum_{j\in\mathsf{N}_n^i}\|B_{m_nk_n}^{*ij}\|_F\Big)$$
$$\leq P\Big(\max_{j\notin \mathsf{N}^i}\|R_{33}^j\|_F\geq \frac{c_1}{3}\lambda_n\sum_{j\in\mathsf{N}_n^i}\|B_{m_nk_n}^{*ij}\|_F\text{ and }\mathcal{A}_3\Big)+P(\mathcal{A}_3^{\complement})$$
$$\leq P\Big(\|\frac{\tilde{\mathbf{H}}_{\mathbf{n}}(\xi^{\mathsf{N}_n^i})}{n}\epsilon^i\|_F\geq\frac{\sqrt{C_{\min}}}{6\sqrt{2}}\Big)+P\Big(\|\Sigma^n_{\mathsf{N}_n^i\mathsf{N}_n^i}-\hat{\Sigma}^n_{\mathsf{N}_n^i\mathsf{N}_n^i}\|_F\geq\frac{C_{\min}}{4}\Big)+P(\mathcal{A}_3^{\complement})$$
$$\lesssim \exp\Big(-C_6\frac{n^{1-\alpha(2+3\beta)}(\lambda_n\sum_{j\in\mathsf{N}_n^i}\|B_{m_nk_n}^{*ij}\|_F)^2}{n^im_n^2k_n^4}+2\log(n^im_nk_n)\Big)$$
$$+\exp\Big(-C_6\frac{n^{1-\alpha(2+3\beta)}}{(n^i)^2m_n^2k_n^4}+2\log(n^im_nk_n)\Big)+\exp\Big(-C_6\frac{n}{n^im_n^2k_n}+\log(n^im_n^2k_n)\Big),$$

where we used (8.59) and Theorem 8.1. Combining the results of $\|R^j_{31}\|_F$, $\|R^j_{32}\|_F$ and $\|R^j_{33}\|_F$ we conclude,

$$P\Big(\max_{j\notin \mathsf{N}^i_\mathsf{n}} \|R^j_3\|_F \geq c_1\lambda_n \sum_{j\in \mathsf{N}^i_\mathsf{n}} \|B^{*ij}_{m_nk_n}\|_F\Big) \lesssim \exp\Big(-C_4\frac{n^{1-\alpha(2+3\beta)}(\lambda_n\sum_{j\in\mathsf{N}^i_\mathsf{n}}\|B^{*ij}_{m_nk_n}\|_F)^2}{n^i m_n^2 k_n^4} + 2\log(n^i m_n k_n)\Big).$$

Term R^j_4. We have

$$\begin{aligned}\max_{j\notin \mathsf{N}^i_\mathsf{n}} \|R^j_4\|_F &\leq \max_{j\notin \mathsf{N}^i_\mathsf{n}} \|\hat{\Sigma}^n_{\xi^j\mathsf{N}^i_\mathsf{n}} - \Sigma^n_{\xi^j\mathsf{N}^i_\mathsf{n}}\|_F\|(\hat{\Sigma}^n_{\mathsf{N}^i_\mathsf{n}\mathsf{N}^i_\mathsf{n}} + \lambda_n\hat{D}_{\mathsf{N}^i_\mathsf{n}})^{-1}\|_2\|\hat{\Sigma}^n_{\mathsf{N}^i_\mathsf{n}\xi^i} - \Sigma_{\mathsf{N}^i_\mathsf{n}\xi^i}\|_F\\ &\leq \frac{2}{C_{\min}}\max_{j\notin \mathsf{N}^i_\mathsf{n}}\|\hat{\Sigma}_{\xi^j\mathsf{N}^i_\mathsf{n}} - \Sigma^n_{\xi^j\mathsf{N}^i_\mathsf{n}}\|_F\|\hat{\Sigma}_{\mathsf{N}^i_\mathsf{n}\xi^i} - \Sigma^n_{\mathsf{N}^i_\mathsf{n}\xi^i}\|_F.\end{aligned}$$

Recall the definition of event $\mathcal{A}_1 = \{\max_{j\notin \mathsf{N}^i_\mathsf{n}}\|\hat{\Sigma}^n_{\xi^j\mathsf{N}^i_\mathsf{n}} - \Sigma^n_{\xi^j\mathsf{N}^i_\mathsf{n}}\|_F \leq \lambda_n\sum_{j\in\mathsf{N}^i_\mathsf{n}}\|B^{*ij}_{m_nk_n}\|_F\}$, then we have

$$\begin{aligned}P\Big(\max_{j\notin \mathsf{N}^i_\mathsf{n}} \|R^j_4\|_F \geq c_1\lambda_n \sum_{j\in \mathsf{N}^i_\mathsf{n}} \|B^{*ij}_{m_nk_n}\|_F\Big) &\leq P\Big(\|\hat{\Sigma}^n_{\mathsf{N}^i_\mathsf{n}\xi^i} - \Sigma^n_{\mathsf{N}^i_\mathsf{n}\xi^i}\|_F \geq c_1\frac{C_{\min}}{2}\Big) + P(\mathcal{A}_1^\complement)\\ &\lesssim \exp\Big(-C_5\frac{n^{1-\alpha(2+3\beta)}}{n^i m_n{}^2 k_n^3} + \log(n^i m_n^2 k_n)\Big)\\ &+ \exp\Big(-C_5\frac{n^{1-\alpha(2+3\beta)}(\lambda_n\sum_{j\in\mathsf{N}^i_\mathsf{n}}\|B^{*ij}_{m_nk_n}\|_F)^2}{n^i m_n^2 k_n^4} + \log\big(pm_n^2k_n^2\big)\Big),\end{aligned}$$

for some positive constant C_5, where the estimates follow from Theorem 8.1.

Term R^j_5. For the term R^j_5 we obtain by the same argument for some positive constant C_6

$$\begin{aligned}P\Big(\max_{j\notin \mathsf{N}^i_\mathsf{n}} \|R^j_5\|_F \geq c_1\lambda_n \sum_{j\in \mathsf{N}^i_\mathsf{n}} \|B^{*ij}_{m_nk_n}\|_F\Big) &\leq (p-n^i)P\Big(\|\Sigma^n_{\xi^j\xi^i} - \hat{\Sigma}^n_{\xi^j\xi^i}\|_F \geq c_1\lambda_n\sum_{j\in \mathsf{N}^i_\mathsf{n}} \|B^{*ij}_{m_nk_n}\|_F\Big)\\ &\lesssim \exp\Big(-C_6\frac{n^{1-\alpha(2+3\beta)}(\lambda_n\sum_{j\in\mathsf{N}^i_\mathsf{n}}\|B^{*ij}_{m_nk_n}\|_F)^2}{m_n^2k_n^3} + \log(pm_n^2k_n)\Big).\end{aligned}$$

Term R^j_6. Using again the identity $A^{-1} - B^{-1} = A^{-1}(B-A)B^{-1}$ and (8.50), we have

$$R^j_6 = \lambda_n\Sigma^n_{\xi^j\mathsf{N}^i_\mathsf{n}}(\Sigma^n_{\mathsf{N}^i_\mathsf{n}\mathsf{N}^i_\mathsf{n}} + \lambda_nD^*_{\mathsf{N}^i_\mathsf{n}})^{-1}(D^*_{\mathsf{N}^i_\mathsf{n}} - \hat{D}_{\mathsf{N}^i_\mathsf{n}})(\Sigma^n_{\mathsf{N}^i_\mathsf{n}\mathsf{N}^i_\mathsf{n}} + \lambda_n\hat{D}_{\mathsf{N}^i_\mathsf{n}})^{-1}\Big(\Sigma^n_{\mathsf{N}^i_\mathsf{n}\mathsf{N}^i_\mathsf{n}}B_n^{*\mathsf{N}^i_\mathsf{n}} + \frac{\tilde{\mathbf{H}}_\mathbf{n}(\xi^{\mathsf{N}^i_\mathsf{n}})}{n}w^i + \frac{\tilde{\mathbf{H}}_\mathbf{n}(\xi^{\mathsf{N}^i_\mathsf{n}})}{n}\epsilon^i\Big).$$

Obviously, on the event $\mathcal{N}$ we have

$$P\Big(\max_{j\notin \mathsf{N}^i_\mathsf{n}} \|R^j_6\|_F \geq c_1\lambda_n \sum_{j\in \mathsf{N}^i_\mathsf{n}} \|B^{*ij}_{m_nk_n}\|_F\Big) \leq \sum_{r=1}^3 P\Big(\|R_{3r}\|_F \geq \frac{c_1}{3}\lambda_n\sum_{j\in \mathsf{N}^i_\mathsf{n}} \|B^{*ij}_{m_nk_n}\|_F\Big),$$

where

$$\begin{aligned}\|R^j_{61}\|_F &\leq \frac{1-\eta}{\sqrt{n^i}}\lambda_n\|D^*_{\mathsf{N}^i_\mathsf{n}} - \hat{D}_{\mathsf{N}^i_\mathsf{n}}\|_2,\\ \|R^j_{62}\|_F &\leq \frac{1-\eta}{\sqrt{n^i}}\lambda_n\|D^*_{\mathsf{N}^i_\mathsf{n}} - \hat{D}_{\mathsf{N}^i_\mathsf{n}}\|_2\|(\Sigma^n_{\mathsf{N}^i_\mathsf{n}\mathsf{N}^i_\mathsf{n}} + \lambda_n\hat{D}_{\mathsf{N}^i_\mathsf{n}})^{-1}\frac{\tilde{\mathbf{H}}_\mathbf{n}(\xi^{\mathsf{N}^i_\mathsf{n}})}{n}w^i\|_F,\\ \|R^j_{63}\|_F &\leq \frac{1-\eta}{\sqrt{n^i}}\lambda_n\|D^*_{\mathsf{N}^i_\mathsf{n}} - \hat{D}_{\mathsf{N}^i_\mathsf{n}}\|_2\|(\Sigma^n_{\mathsf{N}^i_\mathsf{n}\mathsf{N}^i_\mathsf{n}} + \lambda_n\hat{D}_{\mathsf{N}^i_\mathsf{n}})^{-1}\frac{\tilde{\mathbf{H}}_\mathbf{n}(\xi^{\mathsf{N}^i_\mathsf{n}})}{n}\epsilon^i\|_F.\end{aligned}$$

Now, using the same computations as in the proof of Proposition 4 in Lee et al. (2020), we obtain for the operator norm of the matrix $(D^*_{\mathsf{N}^i_n})_{jj} - (\hat{D}_{\mathsf{N}^i_n})_{jj}$

$$
\begin{aligned}
\|(D^*_{\mathsf{N}^i_n})_{jj} - (\hat{D}_{\mathsf{N}^i_n})_{jj}\|_2 &= \Big| \frac{\sum_{\ell \in \mathsf{N}^i_n} \|\hat{B}^{i\ell}_n\|_F}{\|\hat{B}^{ij}_n\|_F} - \frac{\sum_{\ell \in \mathsf{N}^i_n} \|B^{*i\ell}_{m_n k_n}\|_F}{\|B^{*ij}_{m_n k_n}\|_F} \Big| \\
&\leq \frac{\|B^{*ij}_{m_n k_n}\|_F \sum_{\ell \in \mathsf{N}^i_n} \Big| \|\hat{B}^{i\ell}_n\|_F - \|B^{*i\ell}_{m_n k_n}\|_F \Big| + \Big| \|B^{*ij}_{m_n k_n}\|_F - \|\hat{B}^{ij}_n\|_F \Big| \sum_{\ell \in \mathsf{N}^i_n} \|B^{*i\ell}_{m_n k_n}\|_F}{\|\hat{B}^{ij}_n\|_F \|B^{*ij}_{m_n k_n}\|_F} \\
&\leq \frac{\|B^{*ij}_{m_n k_n}\|_F \sum_{\ell \in \mathsf{N}^i_n} \|\hat{B}^{i\ell}_n - B^{*i\ell}_n\|_F + \|B^{*ij}_{m_n k_n} - \hat{B}^{ij}_n\|_F \sum_{\ell \in \mathsf{N}^i_n} \|B^{*i\ell}_{m_n k_n}\|_F}{\|\hat{B}^{ij}_n\|_F \|B^{*ij}_{m_n k_n}\|_F} \\
&\leq \frac{\sum_{\ell \in \mathsf{N}^i_n} \|\hat{B}^{i\ell}_n - B^{*i\ell}_{m_n k_n}\|_F \sum_{\ell \in \mathsf{N}^i_n} \|B^{*i\ell}_{m_n k_n}\|_F}{\|B^{*ij}_{m_n k_n}\|_F (\|B^{*ij}_{m_n k_n}\|_F - \|\hat{B}^{ij}_n - B^{*ij}_{m_n k_n}\|_F)}.
\end{aligned}
$$

On the event $\mathcal{A}_4 = \{\|\hat{B}^{\mathsf{N}^i_n}_n - B^{*\mathsf{N}^i_n}_n\|_F \leq \frac{b^{*i}_n}{2}\}$, we obtain

$$
\begin{aligned}
\|(D^*_{\mathsf{N}^i_n})_{jj} - (\hat{D}_{\mathsf{N}^i_n})_{jj}\|_2 &\leq 2(b^{*i}_n)^{-2} \sum_{\ell \in \mathsf{N}^i_n} \|\hat{B}^{i\ell}_n - B^{*i\ell}_{m_n k_n}\|_F \sum_{\ell \in \mathsf{N}^i_n} \|B^{*i\ell}_{m_n k_n}\|_F \\
&\leq 2(b^{*i}_n)^{-2} (n^i)^{1/2} \|\hat{B}^{\mathsf{N}^i_n}_n - B^{*\mathsf{N}^i_n}_n\|_F \sum_{\ell \in \mathsf{N}^i_n} \|B^{*i\ell}_{m_n k_n}\|_F,
\end{aligned}
$$

where we used the Cauchy-Schwarz inequality. Consequently,

$$
P\Big(\max_{j \notin \mathsf{N}^i} \|R^j_{61}\|_F \geq \frac{c_1}{3} \lambda_n \sum_{j \in \mathsf{N}^i_n} \|B^{*ij}_{m_n k_n}\|_F \Big) \leq P(\|\hat{B}^{\mathsf{N}^i_n}_n - B^{*\mathsf{N}^i_n}_n\|_F \geq \frac{c_1}{6} \frac{1}{1-\eta} (b^{*i}_n)^2) + P(\mathcal{A}^{\mathsf{C}}_4).
$$

Now, we apply Proposition 8.5 with $\delta = \frac{c_1}{6} \frac{1}{1-\eta} (b^{*i}_n)^2$ and $\delta = \frac{b^{*i}_n}{2}$ to obtain

$$
\begin{aligned}
&P\Big(\max_{j \notin \mathsf{N}^i} \|R^j_{61}\|_F \geq \frac{c_1}{3} \lambda_n \sum_{j \in \mathsf{N}^i_n} \|B^{*ij}_{m_n k_n}\|_F \Big) \\
&\quad \lesssim \exp\Big(- C_7 \frac{n^{1-\alpha(2+3\beta)} (b^{*i}_n)^6}{(n^i)^4 m_n^2 k_n^4 (\sum_{j \in \mathsf{N}^i_n} \|B^{*ij}_{m_n k_n}\|_F)^2} + 2\log(n^i m_n k_n) \Big) \\
&\quad\quad + \exp\Big(- C_7 \frac{n^{1-\alpha(2+3\beta)} (b^{*i}_n)^4}{(n^i)^4 m_n^2 k_n^4 (\sum_{j \in \mathsf{N}^i_n} \|B^{*ij}_{m_n k_n}\|_F)^2} + 2\log(n^i m_n k_n) \Big) \\
&\quad \lesssim \exp\Big(- C_7 \frac{n^{1-\alpha(2+3\beta)} (\lambda_n \sum_{j \in \mathsf{N}^i_n} \|B^{*ij}_{m_n k_n}\|_F)^2}{n^i m_n^2 k_n^4} + 2\log(n^i m_n k_n) \Big),
\end{aligned}
$$

where we used $\frac{2}{C_{\min}} \lambda_n (n^i)^{3/2} (\sum_{j \in \mathsf{N}^i_n} \|B^{*ij}_{m_n k_n}\|_F)^2 \leq c_2 (b^{*i}_n)^s$ $(s = 2, 3)$ for the last inequality (see condition (8.67)).

By the same arguments and using (8.54), (8.55) and Assumption (4.13), we can show the existence of a constant $C_8 > 0$ such that

$$
\begin{aligned}
P\Big(\max_{j\notin \mathsf{N}^i} \|R^j_{62}\|_F \ge \frac{c_1}{3}\lambda_n \sum_{j\in \mathsf{N}^i_{\mathsf{n}}} \|B^{*ij}_{m_nk_n}\|_F\Big) \\
&\lesssim \exp\Big(- C_8 \frac{n^{1-\alpha(2+3\beta)}\lambda_n^{-2}(b_n^{*i})^6}{(n^i)^4 m_n^2 k_n^4 (\sum_{j\in \mathsf{N}^i_{\mathsf{n}}} \|B^{*ij}_{m_nk_n}\|_F)^4} + 2\log(n^i m_n k_n)\Big) \\
&\quad + \exp\Big(- C_8 \frac{n^{1-\alpha(2+3\beta)}(b_n^{*i})^4}{(n^i)^4 m_n^2 k_n^4 (\sum_{j\in \mathsf{N}^i_{\mathsf{n}}} \|B^{*ij}_{m_nk_n}\|_F)^2} + 2\log(n^i m_n k_n)\Big) + P(\Omega^{\complement}) \\
&\lesssim \exp\Big(- C_7 \frac{n^{1-\alpha(2+3\beta)}(\lambda_n \sum_{j\in \mathsf{N}^i_{\mathsf{n}}} \|B^{*ij}_{m_nk_n}\|_F)^2}{n^i m_n^2 k_n^4} + 2\log(n^i m_n k_n)\Big) + P(\Omega^{\complement}) \ ,
\end{aligned}
$$

where the probability $P(\Omega^{\complement})$ can be estimated by Proposition 8.1 and is dominated by the first term because of Assumption4.5. Similarly, using (8.59), we obtain

$$
\begin{aligned}
P\Big(\max_{j\notin \mathsf{N}^i} \|R^j_{63}\|_F \ge \frac{c_1}{3}\lambda_n \sum_{j\in \mathsf{N}^i_{\mathsf{n}}} \|B^{*ij}_{m_nk_n}\|_F\Big) &\lesssim \exp\Big(- C_9 \frac{n^{1-\alpha(2+3\beta)}(b_n^{*i})^6}{(n^i)^4 m_n^2 k_n^4 (\sum_{j\in \mathsf{N}^i_{\mathsf{n}}} \|B^{*ij}_{m_nk_n}\|_F)^2} + 2\log(n^i m_n k_n)\Big) \\
&\quad + \exp\Big(- C_9 \frac{n^{1-\alpha(2+3\beta)}(b_n^{*i})^4}{(n^i)^4 m_n^2 k_n^4 (\sum_{j\in \mathsf{N}^i_{\mathsf{n}}} \|B^{*ij}_{m_nk_n}\|_F)^2} + 2\log(n^i m_n k_n)\Big) \\
&\quad + \exp\Big(- C_9 \frac{n}{n^i m_n^2 k_n} + \log(n^i m_n^2 k_n)\Big) \\
&\lesssim \exp\Big(- C_7 \frac{n^{1-\alpha(2+3\beta)}(\lambda_n \sum_{j\in \mathsf{N}^i_{\mathsf{n}}} \|B^{*ij}_{m_nk_n}\|_F)^2}{n^i m_n^2 k_n^4} + 2\log(n^i m_n k_n)\Big) \\
&\quad + \exp\Big(- C_9 \frac{n}{n^i m_n^2 k_n} + \log(n^i m_n^2 k_n)\Big).
\end{aligned}
$$

Combining these results, we can conclude that

$$
P\Big(\max_{j\notin \mathsf{N}^i} \|R^j_{6}\|_F \ge \frac{c_1}{3}\lambda_n \sum_{j\in \mathsf{N}^i_{\mathsf{n}}} \|B^{*ij}_{m_nk_n}\|_F\Big) \lesssim \exp\Big(- C_7 \frac{n^{1-\alpha(2+3\beta)}(\lambda_n \sum_{j\in \mathsf{N}^i_{\mathsf{n}}} \|B^{*ij}_{m_nk_n}\|_F)^2}{n^i m_n^2 k_n^4} + 2\log(n^i m_n k_n)\Big),
$$

and (8.69) holds for $r = 6$ (note that this argument requires condition (4.13)).

Term R^j_7. Observing (8.50) and (8.51), the term R^j_7 can be further decomposed as

$$
\begin{aligned}
R^j_7 =& \Sigma^n_{\xi^j \mathsf{N}^i_{\mathsf{n}}}(\Sigma^n_{\mathsf{N}^i_{\mathsf{n}}\mathsf{N}^i_{\mathsf{n}}} + \lambda_n D^*_{\mathsf{N}^i_{\mathsf{n}}})^{-1}\Big(\Sigma^n_{\mathsf{N}^i_{\mathsf{n}}\mathsf{N}^i_{\mathsf{n}}} B_n^{*\mathsf{N}^i_{\mathsf{n}}} + \frac{\tilde{\mathbf{H}}_{\mathbf{n}}(\xi^{\mathsf{N}^i_{\mathsf{n}}})}{n} w^i + \frac{\tilde{\mathbf{H}}_{\mathbf{n}}(\xi^{\mathsf{N}^i_{\mathsf{n}}})}{n}\epsilon^i\Big) \\
&- \Sigma^n_{\xi^j \mathsf{N}^i_{\mathsf{n}}} B_n^{*\mathsf{N}^i_{\mathsf{n}}} - \frac{\tilde{H}^{\mathsf{T}}_n(\xi^j)}{n} w^i - \frac{\tilde{H}^{\mathsf{T}}_n(\xi^j)}{n}\epsilon^i \\
=& R^j_{71} + R^j_{72} + R^j_{73},
\end{aligned}
$$

where

$$
\begin{aligned}
R^j_{71} &= \Sigma^n_{\xi^j \mathsf{N}^i_{\mathsf{n}}}(\Sigma^n_{\mathsf{N}^i_{\mathsf{n}}\mathsf{N}^i_{\mathsf{n}}} + \lambda_n D^*_{\mathsf{N}^i_{\mathsf{n}}})^{-1}\Sigma^n_{\mathsf{N}^i_{\mathsf{n}}\mathsf{N}^i_{\mathsf{n}}} B_n^{*\mathsf{N}^i_{\mathsf{n}}} - \Sigma^n_{\xi^j \mathsf{N}^i_{\mathsf{n}}} B_n^{*\mathsf{N}^i_{\mathsf{n}}} \\
R^j_{72} &= \Sigma^n_{\xi^j \mathsf{N}^i_{\mathsf{n}}}(\Sigma^n_{\mathsf{N}^i_{\mathsf{n}}\mathsf{N}^i_{\mathsf{n}}} + \lambda_n D^*_{\mathsf{N}^i_{\mathsf{n}}})^{-1}\frac{\tilde{\mathbf{H}}_{\mathbf{n}}(\xi^{\mathsf{N}^i_{\mathsf{n}}})}{n} w^i - \frac{\tilde{H}^{\mathsf{T}}_n(\xi^j)}{n} w^i \\
R^j_{73} &= \Sigma^n_{\xi^j \mathsf{N}^i_{\mathsf{n}}}(\Sigma^n_{\mathsf{N}^i_{\mathsf{n}}\mathsf{N}^i_{\mathsf{n}}} + \lambda_n D^*_{\mathsf{N}^i_{\mathsf{n}}})^{-1}\frac{\tilde{\mathbf{H}}_{\mathbf{n}}(\xi^{\mathsf{N}^i_{\mathsf{n}}})}{n}\epsilon^i - \frac{\tilde{H}^{\mathsf{T}}_n(\xi^j)}{n}\epsilon^i.
\end{aligned}
$$

Turning to $\|R^j_{71}\|_F$, we have

$$\begin{aligned}\max_{j\notin \mathsf{N}^i_{\mathsf{n}}}\|R^j_{71}\|_F &\le \max_{j\notin \mathsf{N}^i_{\mathsf{n}}}\|\Sigma^n_{\xi^j \mathsf{N}^i_{\mathsf{n}}}(\Sigma^n_{\mathsf{N}^i_{\mathsf{n}}\mathsf{N}^i_{\mathsf{n}}}+\lambda_n D^*_{\mathsf{N}^i_{\mathsf{n}}})^{-1}\Big(\Sigma^n_{\mathsf{N}^i_{\mathsf{n}}\mathsf{N}^i_{\mathsf{n}}}-(\Sigma^n_{\mathsf{N}^i_{\mathsf{n}}\mathsf{N}^i_{\mathsf{n}}}+\lambda_n D^*_{\mathsf{N}^i_{\mathsf{n}}})\Big)B^{*\mathsf{N}^i_{\mathsf{n}}}_n\|_F \\ &= \lambda_n \max_{j\notin \mathsf{N}^i_{\mathsf{n}}}\|\Sigma^n_{\xi^j \mathsf{N}^i_{\mathsf{n}}}(\Sigma^n_{\mathsf{N}^i_{\mathsf{n}}\mathsf{N}^i_{\mathsf{n}}})^{-1}D^*_{\mathsf{N}^i_{\mathsf{n}}}B^{*\mathsf{N}^i_{\mathsf{n}}}_n\|_F \\ &\le (1-\eta)\lambda_n\sum_{j\in \mathsf{N}^i_{\mathsf{n}}}\|B^{*ij}_{m_nk_n}\|_F, \qquad (8.73)\end{aligned}$$

where we have used $\Sigma^n_{\mathsf{N}^i_{\mathsf{n}}\mathsf{N}^i_{\mathsf{n}}} \preccurlyeq \Sigma^n_{\mathsf{N}^i_{\mathsf{n}}\mathsf{N}^i_{\mathsf{n}}}+\lambda_n D^*_{\mathsf{N}^i_{\mathsf{n}}}$ and Assumption 4.8.

For the term R^j_{72}, we first write

$$\begin{aligned}\max_{j\notin \mathsf{N}^i_{\mathsf{n}}}\|R^j_{72}\|_F &\le \max_{j\notin \mathsf{N}^i_{\mathsf{n}}}\|\frac{\tilde{H}^\intercal_n(\xi^j)}{\sqrt{n}}\Big(I-\frac{\tilde{\mathbf{H}}_{\mathbf{n}}(\xi^{\mathsf{N}^i_{\mathsf{n}}})}{\sqrt{n}}(\Sigma^n_{\mathsf{N}^i_{\mathsf{n}}\mathsf{N}^i_{\mathsf{n}}}+\lambda_n D^*_{\mathsf{N}^i_{\mathsf{n}}})^{-1}\frac{\tilde{\mathbf{H}}_{\mathbf{n}}(\xi^{\mathsf{N}^i_{\mathsf{n}}})}{\sqrt{n}}\Big)\frac{w^i}{\sqrt{n}}\|_F \\ &\le \max_{j\notin \mathsf{N}^i_{\mathsf{n}}}\|\frac{\tilde{H}^\intercal_n(\xi^j)}{\sqrt{n}}\|_2\|\frac{w^i}{\sqrt{n}}\|_F\|I-\frac{\tilde{\mathbf{H}}_{\mathbf{n}}(\xi^{\mathsf{N}^i_{\mathsf{n}}})}{\sqrt{n}}(\Sigma^n_{\mathsf{N}^i_{\mathsf{n}}\mathsf{N}^i_{\mathsf{n}}})^{-1}\frac{\tilde{\mathbf{H}}_{\mathbf{n}}(\xi^{\mathsf{N}^i_{\mathsf{n}}})}{\sqrt{n}}\|_2.\end{aligned}$$

The matrix $I-\frac{\tilde{\mathbf{H}}_{\mathbf{n}}(\xi^{\mathsf{N}^i_{\mathsf{n}}})}{\sqrt{n}}(\Sigma^n_{\mathsf{N}^i_{\mathsf{n}}\mathsf{N}^i_{\mathsf{n}}})^{-1}\frac{\tilde{\mathbf{H}}_{\mathbf{n}}(\xi^{\mathsf{N}^i_n})}{\sqrt{n}}$ defines a projection and therefore its operator norm is 1. Moreover, by the Lemma 6.2 in Zhou et al. (1998) we have $\|\frac{\tilde{H}^\intercal_n(\xi^j)}{\sqrt{n}}\frac{\tilde{H}_n(\xi^j)}{\sqrt{n}})\|_2 \le \frac{1}{k_n}$ for all $j\in\mathsf{V}$. Thus, observing (8.55), it follows that on the event Ω defined in(8.22)

$$\|\frac{\tilde{H}^\intercal_n(\xi^j)}{\sqrt{n}}\|_2\|\frac{w^i}{\sqrt{n}}\|_F \lesssim \frac{n^i m_n^{3/2}}{k_n^{d+1/2}}.$$

By condition (4.13), we can choose a constant $c_3>0$ such that

$$\max_{j\notin \mathsf{N}^i_{\mathsf{n}}}\|R^j_{72}\|_F \le \frac{n^i m_n^{3/2}}{k_n^{d+1/2}} \le c_3\sqrt{\frac{2}{C_{\min}}}\lambda_n\sum_{j\in \mathsf{N}^i_{\mathsf{n}}}\|B^{*ij}_{m_nk_n}\|_F.$$

Combining this inequality with the constant $c_3=\sqrt{\frac{C_{\min}}{2}}(1-\eta-\frac{c_1}{2})$ and (8.73) now yields

$$\begin{aligned}&P\Big(\max_{j\notin \mathsf{N}^i_{\mathsf{n}}}\|R^j_7\|_F \ge c_1\lambda_n\sum_{j\in \mathsf{N}^i_{\mathsf{n}}}\|B^{*ij}_{m_nk_n}\|_F\Big) \le P\Big(\max_{j\notin \mathsf{N}^i_{\mathsf{n}}}\|R^j_{73}\|_F \ge \frac{c_1}{2}\lambda_n\sum_{j\in \mathsf{N}^i_{\mathsf{n}}}\|B^{*ij}_{m_nk_n}\|_F\Big) \\ &\le P\Big(\|\frac{\tilde{\mathbf{H}}^\intercal_{\mathbf{n}}(\xi^{\mathsf{N}^i_{\mathsf{n}}})}{n}\epsilon^i\|_F \ge \frac{c_1}{4}\frac{\sqrt{n^i}}{(1-\eta)}\lambda_n\sum_{j\in \mathsf{N}^i_{\mathsf{n}}}\|B^{*ij}_{m_nk_n}\|_F\Big)+(p-n^i)P\Big(\|\frac{\tilde{H}^\intercal_n(\xi^j)}{n}\epsilon^i\|_F \ge \frac{c_1}{4}\lambda_n\sum_{j\in \mathsf{N}^i_{\mathsf{n}}}\|B^{*ij}_{m_nk_n}\|_F\Big) \\ &\lesssim \exp\Big(-C_8\frac{n\Big(\lambda_n\sum_{j\in \mathsf{N}^i_{\mathsf{n}}}\|B^{*ij}_{m_nk_n}\|_F\Big)^2}{n^i m_n^2 k_n}+\log(n^i m_n^2 k_n)\Big) \\ &\quad +\exp\Big(-C_8\frac{n\Big(\lambda_n\sum_{j\in \mathsf{N}^i_{\mathsf{n}}}\|B^{*ij}_{m_nk_n}\|_F\Big)^2}{m_n^2 k_n}+\log((p-n^i)m_n^2 k_n)\Big)+P(\Omega^\complement).\end{aligned}$$

Finally, combining the result (8.70) from Step 1 with the estimates for $R_1,\dots,R_7$, we conclude that

$$\begin{aligned}&P\Big(\max_{j\notin \mathsf{N}^i_{\mathsf{n}}}\|\hat{\Sigma}^n_{\xi^j \mathsf{N}^i_{\mathsf{n}}}\hat{B}^{\mathsf{N}^i_{\mathsf{n}}}_n-\hat{\Sigma}^n_{\xi^j\xi^i}\|_F \ge \lambda_n\sum_{j\neq i}^p\|\hat{B}^{ij}_n\|_F\Big) \\ &\qquad\lesssim \exp\Big(-C_1\frac{n^{1-\alpha(2+3\beta)}(\lambda_n\sum_{j\in \mathsf{N}^i_{\mathsf{n}}}\|B^{*ij}_{m_nk_n}\|_F)^2}{n^i m_n^2 k_n^4}+2\log(pm_nk_n)\Big),\end{aligned}$$

such that $\lambda_n (n^i)^{3/2} \lesssim (b_n^{*i})^3 (\sum_{j\in \mathsf{N}_\mathsf{n}^i} \|B_{m_n k_n}^{*ij}\|_F)^{-2}$. □

Proof of Proposition 8.2 First, note that the event $\hat{\mathsf{N}}_\mathsf{n}^i = \mathsf{N}_\mathsf{n}^i$ holds if and only if

$$\|\hat{B}_n^{ij}\|_F \neq 0 \quad \forall j \in \mathsf{N}_\mathsf{n}^i \quad \text{and} \quad \|\hat{\mathsf{B}}_\mathsf{n}^{ij}\|_\mathsf{F} = 0 \quad \forall \mathsf{j} \notin \mathsf{N}_\mathsf{n}^i,$$

which is implied by the conditions

$$\|\hat{B}_n^{\mathsf{N}_\mathsf{n}^i} - B_n^{*\mathsf{N}_\mathsf{n}^i}\|_F < b_n^{*i} = \min_{j\in \mathsf{N}^i} \|B_{m_n k_n}^{*ij}\|_F,$$
$$\max_{j\notin \mathsf{N}_\mathsf{n}^i} \|\hat{\Sigma}_{\xi^j \mathsf{N}^i}^n \hat{B}_n^{\mathsf{N}_\mathsf{n}^i} - \hat{\Sigma}_{\xi^j \xi^i}^n\|_F \leq \lambda_n \sum_{j\in \mathsf{N}^i} \|\hat{B}_n^{ij}\|_F.$$

Thus, the Proposition 8.5 (with $\delta = b_n^{*i}$, an obvious estimate of the probability using condition (8.67)) and Proposition 8.6 we can conclude that

$$P(\hat{\mathsf{N}}_\mathsf{n}^i \neq \mathsf{N}_\mathsf{n}^i \text{ and } \mathcal{N}) \lesssim \exp\Big(- C_1 \frac{n^{1-\alpha(2+3\beta)}(\lambda_n \sum_{j\in \mathsf{N}_\mathsf{n}^i} \|B_{m_n k_n}^{*ij}\|_F)^2}{n^i m_n^2 k_n^4} + 2\log(p m_n k_n)\Big)$$

and this completes the proof of Proposition 8.2. □