

# Bayesian inference for spectral projectors of the covariance matrix

Igor Silin<sup>1,3,4,5</sup> and Vladimir Spokoiny<sup>2,3,4,5</sup>

<sup>1</sup>*Moscow Institute of Physics and Technology,  
141701, Dolgoprudny, RF.  
e-mail: [siliniv@gmail.com](mailto:siliniv@gmail.com)*

<sup>2</sup>*Weierstrass Institute and Humboldt University,  
Mohrenstr. 39, 10117 Berlin, Germany.  
e-mail: [spokoiny@wias-berlin.de](mailto:spokoiny@wias-berlin.de)*

<sup>3</sup>*National Research University Higher School of Economics,  
20 Myasnitskaya ulitsa, 101000, Moscow, RF.*

<sup>4</sup>*Skolkovo Institute of Science and Technology (Skoltech),  
143026, Moscow, RF.*

<sup>5</sup>*Institute for Information Transmission Problems RAS,  
Bolshoy Karetny per. 19, 127051, Moscow, RF.*

**Abstract:** Let  $X_1, \dots, X_n$  be an i.i.d. sample in  $\mathbb{R}^p$  with zero mean and the covariance matrix  $\Sigma^*$ . The classical PCA approach recovers the projector  $\mathbf{P}_{\mathcal{J}}^*$  onto the principal eigenspace of  $\Sigma^*$  by its empirical counterpart  $\hat{\mathbf{P}}_{\mathcal{J}}$ . Recent paper [24] investigated the asymptotic distribution of the Frobenius distance between the projectors  $\|\hat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2$ , while [27] offered a bootstrap procedure to measure uncertainty in recovering this subspace  $\mathbf{P}_{\mathcal{J}}^*$  even in a finite sample setup. The present paper considers this problem from a Bayesian perspective and suggests to use the credible sets of the pseudo-posterior distribution on the space of covariance matrices induced by the conjugated Inverse Wishart prior as sharp confidence sets. This yields a numerically efficient procedure. Moreover, we theoretically justify this method and derive finite sample bounds on the corresponding coverage probability. Contrary to [24, 27], the obtained results are valid for non-Gaussian data: the main assumption that we impose is the concentration of the sample covariance  $\hat{\Sigma}$  in a vicinity of  $\Sigma^*$ . Numerical simulations illustrate good performance of the proposed procedure even on non-Gaussian data in a rather challenging regime.

**MSC 2010 subject classifications:** Primary 62F15, 62H25, 62G20; secondary 62F25.

**Keywords and phrases:** Covariance matrix, spectral projector, principal component analysis, Bernstein–von Mises theorem.

Received March 2018.

## 1. Introduction

Let the observed data  $\mathbf{X}^n = (X_1, \dots, X_n)$  be a collection of independent identically distributed zero-mean random vectors in  $\mathbb{R}^p$  and let  $X$  be a generic

random vector with the same distribution. Denote by  $\Sigma^*$  its covariance matrix:

$$\Sigma^* \stackrel{\text{def}}{=} \mathbb{E}(XX^\top).$$

Usually one estimates the true unknown covariance by the sample covariance matrix, given by

$$\widehat{\Sigma} \stackrel{\text{def}}{=} \frac{1}{n} \sum_{j=1}^n X_j X_j^\top.$$

Quantifying the quality of approximation of  $\Sigma^*$  by  $\widehat{\Sigma}$  is one of the most classical problems in statistics. Surprisingly, a number of deep and strong results in this area appeared quite recently. The progress is mainly due to Bernstein type results on the spectral norm  $\|\widehat{\Sigma} - \Sigma^*\|_\infty$  in the random matrix theory, see, for instance, [22, 29, 31, 33, 1]. It appears that the quality of approximation is of order  $n^{-1/2}$  while the dimensionality  $p$  only enters logarithmically in the error bound. This allows to apply the results even in the cases of very high data dimension.

Functionals of the covariance matrix also arise in applications frequently. For instance, eigenvalues are well-studied in different regimes, see [26, 9, 19, 34] and many more references therein. The Frobenius norm and other  $l_r$ -norms of covariance matrix are of great interest in financial applications; see, e.g. [10].

Much less is known about the quality of estimation of a spectral projector which is a nonlinear functional of the covariance matrix. However, such objects arise in dimension reduction methods, manifold learning and spectral methods in community detection, see [11] and references therein for an overview of problems where spectral projectors play crucial role. Special attention should be focused on the Principal Component Analysis (PCA), probably the most famous dimension reduction method. Nowadays PCA-based methods are actively used in deep networking architecture [17] and finance [12], along with other applications. Over the past decade huge progress was achieved in theoretical guarantees for sparse PCA in high dimensions, see [18, 5, 3, 6, 13].

Suppose we fix some set  $\mathcal{J}$  of eigenspaces of  $\Sigma^*$  and consider a direct sum of these eigenspaces and the associated true projector  $P_{\mathcal{J}}^*$ . Its empirical counterpart is given by  $\widehat{P}_{\mathcal{J}}$  computed from the sample covariance  $\widehat{\Sigma}$ . The recent paper [28] presents new bounds on so called excess risk of PCA defined as  $\text{Tr}[\Sigma^*(P_{\mathcal{J}}^* - \widehat{P}_{\mathcal{J}})]$ .

This paper focuses on quantification of uncertainty in recovering the spectral projector  $P_{\mathcal{J}}^*$  from its empirical counterpart  $\widehat{P}_{\mathcal{J}}$ . More precisely, the random quantity of our interest is the squared Frobenius distance between the true projector and the sample one  $\|\widehat{P}_{\mathcal{J}} - P_{\mathcal{J}}^*\|_2^2$ . Even though the projector  $P_{\mathcal{J}}^*$  is a complex non-linear mapping of  $\Sigma^*$ , a recent technique from [21] allows to approximate  $(\widehat{P}_{\mathcal{J}} - P_{\mathcal{J}}^*)$  by a linear functional of  $(\widehat{\Sigma} - \Sigma^*)$  with root-n accuracy. Several results about the distribution of this random variable are

available for Gaussian observations:  $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \Sigma^*)$ . For the case when  $\mathcal{J}$  corresponds to a single eigenvalue in the spectrum and consists of one eigenspace, the normal approximation of  $n\|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2$  was shown in [24] with a tight bound on

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P} \left\{ \frac{n\|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2 - \mathbb{E} \left( n\|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2 \right)}{\text{Var}^{1/2} \left( n\|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2 \right)} \leq x \right\} - \Phi(x) \right|,$$

where  $\Phi(x)$  is the standard normal distribution function. However, the distribution of  $n\|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2$  depends on the unknown covariance matrix via the first two moments of  $n\|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2$ . Due to this fact, it is difficult to use this result directly for constructing the confidence sets for the true projector  $\mathbf{P}_{\mathcal{J}}^*$ . The paper [23] demonstrates convergence to a Cauchy-type limit independent from the true covariance operator in a “high-complexity” setting where  $\frac{\text{Tr}(\Sigma^*)}{\|\Sigma^*\|_{\infty}} \rightarrow \infty$ . Besides, a bootstrap approach can be used to overcome the described problem; see [27]. The bootstrap validity result is based on the approximation of the distribution of  $n\|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2$  by the distribution of a Gaussian quadratic form  $\|\xi\|^2$ . Namely, for the Gaussian data, Theorem 4.3 of [27] provides the following statement:

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}(n\|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2 \leq x) - \mathbb{P}(\|\xi\|^2 \leq x) \right| \leq \overline{\diamond}, \quad (1.1)$$

where  $\xi$  is a zero mean Gaussian vector with a specific covariance structure and  $\overline{\diamond}$  is an explicit error term. The similar approximation is obtained in the bootstrap world, this reduces the original problem to the question about Gaussian comparison and Gaussian anti-concentration for large balls.

This paper suggests to look at this problem from a Bayesian point of view. The standard approach for a nonparametric analysis of the posterior distribution is based on the prominent Bernstein – von Mises (BvM) phenomenon. The BvM result states some pivotal (Gaussian) behavior of the posterior. The paper [8] developed a general framework for functional BvM theorem, while [35] used similar ideas to demonstrate asymptotic normality of approximately linear functionals of covariance and precision matrices. In particular, it can be used to justify the use of Bayesian credible sets as frequentist confidence sets for the target parameter; see [25, 32, 15, 20, 4, 7] among others. In this work, we aim to address a similar question specifically for spectral projectors of the covariance matrix. It appears that the general BvM technique can be significantly improved and refined for the problem at hand. The use of the classical conjugated Wishart prior helps not only to build a numerically efficient procedure but also to establish precise finite sample results for the posterior credible sets under mild and general assumptions on the data distribution. The key observation here is that, similarly to the bootstrap approach of [27], the credible level sets for the posterior are nearly elliptic, and the corresponding posterior probability can be approximated by a generalized chi-squared-type distribution. This

allows to apply the recent “large ball probability” bounds on Gaussian comparison and Gaussian anti-concentration from [16]. Moreover, in the contrary to the latter paper [27], we do not require a Gaussian distribution of the data. Our results claim that the posterior credible sets can be used as frequentist confidence regions even under a possible model misspecification when the true data generation measure is not Gaussian. We still work with the Gaussian likelihood, in this sense our procedure is *pseudo-Bayesian* and the constructed credible sets should be referred to as *pseudo-posterior*; see [14]. In our study we allow the dimension  $p$  to grow with the sample size  $n$ , however, we need to assume “ $p^3/n$  is small” in order to make our results meaningful.

The main contributions of this paper are as follows.

- We offer a new procedure for building elliptic confidence sets for the true projector based on Bayesian simulation from the Inverse Wishart prior. The procedure is fully data-driven and numerically efficient, its complexity is proportional to the squared dimension and independent of sample size. Numerical simulations confirm good performance of the proposed method for artificial data: both Gaussian and non-Gaussian (not even sub-Gaussian).
- We establish novel results on the coverage properties of pseudo-posterior credible sets for a complicated non-linear problem of recovering the eigenspace of the sample covariance matrix. The results apply under mild conditions on the data distribution. In particular, we do not require Gaussianity of the observations.

The rest of the paper is structured as follows. Some notations are introduced in Section 2.1. Section 2.2 discusses the model. Our pseudo-Bayesian framework and the main result of the paper about the pseudo-posterior credible sets are described in Section 2.3. The use of such sets as frequentist confidence sets is discussed in Section 2.4. Some numerical results on simulated data are demonstrated in Section 3. Section 4 contains the proofs of the main theorems. Some auxiliary results from the literature and the rest of the proofs are collected in Appendix A and Appendix B, respectively.

## 2. Problem and main results

This section explains our setup and states the main results.

### 2.1. Notations

We will use the following notations throughout the paper. The space of real-valued  $p \times p$  matrices is denoted by  $\mathbb{R}^{p \times p}$ , while  $\mathbb{S}_+^p$  means the set of positive-semidefinite matrices. We write  $\mathbf{I}_d$  for the identity matrix of size  $d \times d$ ,  $\text{rank}(A)$  and  $\text{Tr}(B)$  stand for the *rank* of a matrix  $A$  and the *trace* of a square matrix  $B$ . Further,  $\|A\|_\infty$  stands for the *spectral norm* of a matrix  $A$ , while  $\|A\|_1$  means the *nuclear norm*. The *Frobenius scalar product* of two matrices  $A$  and  $B$  of

the same size is  $\langle A, B \rangle_2 \stackrel{\text{def}}{=} \text{Tr}(A^\top B)$ , while the *Frobenius norm* is denoted by  $\|A\|_2$ . When applied to a vector,  $\|\cdot\|$  means just its *Euclidean norm*. The *effective rank* of a square matrix  $B$  is defined by  $r(B) \stackrel{\text{def}}{=} \frac{\text{Tr}(B)}{\|B\|_\infty}$ . The relation  $a \lesssim b$  means that there exists an absolute constant  $C$ , different from line to line, such that  $a \leq Cb$ , while  $a \asymp b$  means that  $a \lesssim b$  and  $b \lesssim a$ . By  $a \vee b$  and  $a \wedge b$  we mean maximum and minimum of  $a$  and  $b$ , respectively. In the sequel we will often be considering intersections of events of probability greater than  $1 - 1/n$ . Without loss of generality, we will write that the probability measure of such an intersection is  $1 - 1/n$ , since it can be easily achieved by adjusting constants. Throughout the paper we assume that  $p < n$ .

## 2.2. Setup and problem

Let  $X_1, \dots, X_n$  be i.i.d. zero mean with  $\text{Var}(X) = \Sigma^*$ . Without loss of generality, we can assume that  $\Sigma^* \in \mathbb{S}_+^p$  is invertible, otherwise one can easily transform the data in such a way that the covariance matrix for the transformed data will be invertible. Let  $\sigma_1^* \geq \dots \geq \sigma_p^*$  be the ordered eigenvalues of  $\Sigma^*$ . Suppose that among them there are  $q$  distinct eigenvalues  $\mu_1^* > \dots > \mu_q^*$ . Introduce groups of indices  $\Delta_r^* = \{j : \mu_r^* = \sigma_j^*\}$  and denote by  $m_r^*$  the multiplicity factor (dimension)  $|\Delta_r^*|$  for all  $r \in \{1, \dots, q\}$ . The corresponding eigenvectors are denoted as  $u_1^*, \dots, u_p^*$ . We will use the projector on the  $r$ -th eigenspace of dimension  $m_r^*$ :

$$P_r^* = \sum_{j \in \Delta_r^*} u_j^* u_j^{*\top}$$

and the eigendecomposition

$$\Sigma^* = \sum_{j=1}^p \sigma_j^* u_j^* u_j^{*\top} = \sum_{r=1}^q \mu_r^* \left( \sum_{j \in \Delta_r^*} u_j^* u_j^{*\top} \right) = \sum_{r=1}^q \mu_r^* P_r^*.$$

We also introduce the spectral gaps  $g_r^*$ :

$$g_r^* = \begin{cases} \mu_1^* - \mu_2^*, & r = 1, \\ (\mu_{r-1}^* - \mu_r^*) \wedge (\mu_r^* - \mu_{r+1}^*), & r \in \{2, \dots, q-1\}, \\ \mu_{q-1}^* - \mu_q^*, & r = q. \end{cases}$$

Suppose that  $\widehat{\Sigma}$  has  $p$  eigenvalues  $\widehat{\sigma}_1 > \dots > \widehat{\sigma}_p$  (distinct with probability one). The corresponding eigenvectors are denoted as  $\widehat{u}_1, \dots, \widehat{u}_p$ . Suppose that  $\|\widehat{\Sigma} - \Sigma^*\|_\infty \leq \frac{1}{4} \min_{r \in \{1, \dots, q\}} g_r^*$ . Then, as shown in [21], we can identify clusters of the eigenvalues of  $\widehat{\Sigma}$  corresponding to each eigenvalue of  $\Sigma^*$  and therefore

determine  $\Delta_r^*$  and  $m_r^*$  for all  $r \in \{1, \dots, q\}$ . Then we can define the sample projector on the  $r$ -th eigenspace of dimension  $m_r^*$ :

$$\widehat{\mathbf{P}}_r = \sum_{j \in \Delta_r^*} \widehat{u}_j \widehat{u}_j^\top.$$

Under the condition that the spectral gap is sufficiently large, [27] approximated the distribution of  $n\|\widehat{\mathbf{P}}_r - \mathbf{P}_r^*\|_2^2$  by the distribution of a Gaussian quadratic form  $\|\xi\|^2$  with  $\xi \sim \mathcal{N}(0, \Gamma_r^*)$  and  $\Gamma_r^*$  is a block-matrix of the form

$$\Gamma_r^* \stackrel{\text{def}}{=} \begin{bmatrix} \Gamma_{r1}^* & O & \cdots & O \\ O & \Gamma_{r2}^* & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & \Gamma_{rq}^* \end{bmatrix} \tag{2.1}$$

with  $(q - 1)$  diagonal blocks of sizes  $m_r^* m_s^* \times m_r^* m_s^*$ :

$$\Gamma_{rs}^* \stackrel{\text{def}}{=} \frac{2\mu_r^* \mu_s^*}{(\mu_r^* - \mu_s^*)^2} \cdot \mathbf{I}_{m_r^* m_s^*}, \quad s \neq r.$$

Below we extend these result in two aspects. First, our approach allows to pick a block of eigenspaces corresponding to an interval  $\mathcal{J}$  in  $\{1, \dots, q\}$  from  $r^-$  to  $r^+$ . Second, we relax the assumption on Gaussianity of the data.

Let

$$\mathcal{J} = \{r^-, r^- + 1, \dots, r^+\}.$$

Define also the subset of indices

$$\mathcal{I}_{\mathcal{J}} \stackrel{\text{def}}{=} \{k: k \in \Delta_r^*, r \in \mathcal{J}\},$$

and introduce the projector onto the direct sum of the eigenspaces associated with  $\mathbf{P}_r^*$  for all  $r \in \mathcal{J}$ :

$$\mathbf{P}_{\mathcal{J}}^* \stackrel{\text{def}}{=} \sum_{r \in \mathcal{J}} \mathbf{P}_r^* = \sum_{k \in \mathcal{I}_{\mathcal{J}}} u_k^* u_k^{*\top}.$$

Its empirical counterpart is given by

$$\widehat{\mathbf{P}}_{\mathcal{J}} \stackrel{\text{def}}{=} \sum_{r \in \mathcal{J}} \widehat{\mathbf{P}}_r = \sum_{k \in \mathcal{I}_{\mathcal{J}}} \widehat{u}_k \widehat{u}_k^\top.$$

For instance, when  $\mathcal{J} = \{1, \dots, q_{\text{eff}}\}$  for some  $q_{\text{eff}} < q$ , then  $\widehat{\mathbf{P}}_{\mathcal{J}}$  is exactly what is recovered by PCA. Below we focus on  $n\|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2$  rather than  $n\|\widehat{\mathbf{P}}_r - \mathbf{P}_r^*\|_2^2$ .

The projector dimension for  $\mathcal{J}$  is given by  $m_{\mathcal{J}}^* = \sum_{r \in \mathcal{J}} m_r^*$ . Its *spectral gap* can be defined as

$$g_{\mathcal{J}}^* \stackrel{\text{def}}{=} \begin{cases} \mu_{r^+}^* - \mu_{r^++1}^*, & \text{if } r^- = 1; \\ \mu_{r^- - 1}^* - \mu_{r^-}^*, & \text{if } r^+ = q; \\ (\mu_{r^- - 1}^* - \mu_{r^-}^*) \wedge (\mu_{r^+}^* - \mu_{r^++1}^*), & \text{otherwise.} \end{cases}$$

Define also for  $\mathcal{J} = \{r^-, r^- + 1, \dots, r^+\}$

$$l_{\mathcal{J}}^* = \mu_{r^-}^* - \mu_{r^+}^*.$$

To describe the distribution of the projector  $\widehat{\mathbf{P}}_{\mathcal{J}}$ , introduce the following matrix  $\Gamma_{\mathcal{J}}^*$  of size  $m_{\mathcal{J}}^*(p - m_{\mathcal{J}}^*) \times m_{\mathcal{J}}^*(p - m_{\mathcal{J}}^*)$ :

$$\Gamma_{\mathcal{J}}^* \stackrel{\text{def}}{=} \text{diag}(\Gamma_{\mathcal{J}}^r)_{r \in \mathcal{J}}, \tag{2.2}$$

$$\Gamma_{\mathcal{J}}^r \stackrel{\text{def}}{=} \text{diag}(\Gamma^{r,s})_{s \notin \mathcal{J}},$$

$$\Gamma^{r,s} \stackrel{\text{def}}{=} \frac{2\mu_r^* \mu_s^*}{(\mu_r^* - \mu_s^*)^2} \cdot \mathbf{I}_{m_r^* m_s^*}, \quad r \in \mathcal{J}, s \notin \mathcal{J}.$$

It is easy to notice that when  $\mathcal{J} = \{r\}$  then this definition coincides with (2.1).

Our results apply under one rather mild and natural condition on the distribution of  $\mathbf{X}^n = (X_1, \dots, X_n)$  that the sample covariance matrix  $\widehat{\Sigma}$  concentrates around the true covariance  $\Sigma^*$ :

$$\|\widehat{\Sigma} - \Sigma^*\|_{\infty} \leq \widehat{\delta}_n \|\Sigma^*\|_{\infty} \tag{2.3}$$

with probability  $1 - 1/n$ . In this result we do not use, say, independence of the  $X_i$ 's or zero mean property, everything is done conditioned on the data  $\mathbf{X}^n$ . The value  $\widehat{\delta}_n$  clearly depends on the underlying data distributions, but it allows to work with much wider classes of probability measures rather than just Gaussian or sub-Gaussian. For the Gaussian case one may take

$$\widehat{\delta}_n \asymp \sqrt{\frac{r(\Sigma^*) + \log(n)}{n}}.$$

Several more examples of possible distributions and the corresponding  $\widehat{\delta}_n$  for them are provided in Appendix A, see Theorem A.1. So, throughout the rest of the paper we assume that the data satisfy condition (2.3).

### 2.3. Pseudo-Bayesian framework and credible level sets

Let  $\Pi$  be a prior distribution on the set of considered covariance matrices  $\Sigma$ . Even though our data are not necessary Gaussian, we can consider the Gaussian

log-likelihood:

$$\ell_n(\boldsymbol{\Sigma}) = -\frac{n}{2} \log \det(\boldsymbol{\Sigma}) - \frac{n}{2} \text{Tr}(\boldsymbol{\Sigma}^{-1} \widehat{\boldsymbol{\Sigma}}) - \frac{np}{2} \log(2\pi).$$

In case of Gaussian data, the posterior measure of a set  $\mathcal{A} \subset \mathbb{S}_+^p$  can be expressed as

$$\Pi(\mathcal{A} \mid \mathbf{X}^n) = \frac{\int_{\mathcal{A}} \exp(\ell_n(\boldsymbol{\Sigma})) d\Pi(\boldsymbol{\Sigma})}{\int_{\mathbb{S}_+^p} \exp(\ell_n(\boldsymbol{\Sigma})) d\Pi(\boldsymbol{\Sigma})}.$$

However, we can study this random measure for non-Gaussian data as well. As the Gaussian log-likelihood  $\ell_n(\boldsymbol{\Sigma})$  does not necessarily correspond to the true distribution of our data, we call the random measure  $\Pi(\cdot \mid \mathbf{X}^n)$  a *pseudo-posterior*, [14]. Once a prior is fixed, we can easily sample matrices  $\boldsymbol{\Sigma}$  from this pseudo-posterior distribution. Denote eigenvalues of  $\boldsymbol{\Sigma}$  as  $\sigma_1 > \dots > \sigma_p$  (assume they are distinct with probability one) and eigenvectors as  $u_1, \dots, u_p$ . The corresponding projector onto the  $r$ -th eigenspace of dimension  $m_r^*$  is

$$\mathbf{P}_r = \sum_{k \in \Delta_r^*} u_k u_k^\top.$$

and the projector on the direct sum of eigenspaces associated with  $\mathbf{P}_r$  for  $r \in \mathcal{J}$  is

$$\mathbf{P}_{\mathcal{J}} = \sum_{r \in \mathcal{J}} \mathbf{P}_r = \sum_{k \in \mathcal{I}_{\mathcal{J}}} u_k u_k^\top.$$

In this work we focus on the conjugate prior to the multivariate Gaussian distribution, that is, the Inverse Wishart distribution  $\mathcal{IW}_p(\mathbf{G}, p + b - 1)$  with  $\mathbf{G} \in \mathbb{S}_+^p$ ,  $0 < b \lesssim p$ . Its density is given by

$$\frac{d\Pi(\boldsymbol{\Sigma})}{d\boldsymbol{\Sigma}} \propto \exp\left(-\frac{2p + b}{2} \log \det(\boldsymbol{\Sigma}) - \frac{1}{2} \text{Tr}(\mathbf{G}\boldsymbol{\Sigma}^{-1})\right).$$

Some nice properties of the Inverse Wishart prior distribution allow us to obtain the following result which we will use for uncertainty quantification statements in the next section instead of the Bernstein–von Mises Theorem.

**Theorem 2.1.** *Assume that the distribution of the data  $\mathbf{X}^n = (X_1, \dots, X_n)$  fulfills the sample covariance concentration property (2.3) with  $\widehat{\delta}_n$  satisfying*

$$\widehat{\delta}_n \leq \frac{g_{\mathcal{J}}^*}{4\|\boldsymbol{\Sigma}^*\|_{\infty}} \wedge \frac{r(\boldsymbol{\Sigma}^*)}{p}.$$

*Consider the prior  $\Pi(\boldsymbol{\Sigma})$  given by the Inverse Wishart distribution  $\mathcal{IW}_p(\mathbf{G}, p + b - 1)$ . Let  $\xi \sim \mathcal{N}(0, \Gamma_{\mathcal{J}}^*)$  with  $\Gamma_{\mathcal{J}}^*$  defined by (2.2).*

Then with probability  $1 - \frac{1}{n}$

$$\sup_{x \in \mathbb{R}} \left| \Pi \left( n \|\mathbf{P}_{\mathcal{J}} - \widehat{\mathbf{P}}_{\mathcal{J}}\|_2^2 \leq x \mid \mathbf{X}^n \right) - \mathbb{P}(\|\xi\|^2 \leq x) \right| \lesssim \diamond,$$

where

$$\diamond = \diamond(n, p, \boldsymbol{\Sigma}^*) \stackrel{\text{def}}{=} \frac{\diamond_1 + \diamond_2 + \diamond_3}{\|\Gamma_{\mathcal{J}}^*\|_2^{1/2} (\|\Gamma_{\mathcal{J}}^*\|_2^2 - \|\Gamma_{\mathcal{J}}^*\|_{\infty}^2)^{1/4}} + \frac{1}{n}. \quad (2.4)$$

The terms  $\diamond_1$  through  $\diamond_3$  can be described as

$$\begin{aligned} \diamond_1 &\asymp \left\{ (\log(n) + p) \left( \left( 1 + \frac{l_{\mathcal{J}}^*}{g_{\mathcal{J}}^*} \right) \frac{\sqrt{m_{\mathcal{J}}^*} \|\boldsymbol{\Sigma}^*\|_{\infty}}{g_{\mathcal{J}}^*} + m_{\mathcal{J}}^* \right) \|\boldsymbol{\Sigma}^*\|_{\infty} + m_{\mathcal{J}}^* \|\mathbf{G}\|_{\infty} \right\} \times \\ &\quad \times \frac{m_{\mathcal{J}}^* \|\boldsymbol{\Sigma}^*\|_{\infty}}{g_{\mathcal{J}}^{*2}} \sqrt{\frac{\log(n) + p}{n}}, \\ \diamond_2 &\asymp \frac{\|\boldsymbol{\Sigma}^*\|_{\infty} \left( m_{\mathcal{J}}^* \|\boldsymbol{\Sigma}^*\|_{\infty}^2 \wedge \text{Tr}(\boldsymbol{\Sigma}^{*2}) \right)}{g_{\mathcal{J}}^{*3}} p \left( \widehat{\delta}_n + \frac{p}{n} \right), \\ \diamond_3 &\asymp \frac{(m_{\mathcal{J}}^*)^{3/2} \|\boldsymbol{\Sigma}^*\|_{\infty} \text{Tr}(\boldsymbol{\Sigma}^*)}{g_{\mathcal{J}}^{*2}} \sqrt{\frac{\log(n)}{n}}. \end{aligned}$$

**Remark 2.1.** The bound (2.4) can be made more transparent if we fix  $\boldsymbol{\Sigma}^*$  and focus on the dependence on  $p, n, \widehat{\delta}_n$  and the desired subspace dimension  $m_{\mathcal{J}}^*$  only (freezing the eigenvalues, the spectral gaps and multiplicities of the eigenvalues):

$$\diamond \asymp \sqrt{\frac{(m_{\mathcal{J}}^*)^4 (p^3 + \log^3(n))}{n}} \vee m_{\mathcal{J}}^* p \widehat{\delta}_n,$$

or, in the sub-Gaussian case,

$$\diamond \asymp \sqrt{\frac{(m_{\mathcal{J}}^*)^4 (p^3 + \log^3(n))}{n}}.$$

Moreover, in the case of spiked covariance model we expect  $\|\Gamma_{\mathcal{J}}^*\|_2$  to behave as  $\sqrt{p m_{\mathcal{J}}^*}$  which improves the previous bounds to

$$\diamond \asymp \sqrt{\frac{(m_{\mathcal{J}}^*)^3 (p^2 + \log^3(n)/p)}{n}} \vee \sqrt{m_{\mathcal{J}}^* p \widehat{\delta}_n},$$

and

$$\diamond \asymp \sqrt{\frac{(m_{\mathcal{J}}^*)^3 (p^2 + \log^3(n)/p)}{n}},$$

respectively.

**2.4. Gaussian approximation and frequentist uncertainty quantification for spectral projectors**

For the Gaussian data, Theorem 4.3 of [27] provides the explicit error bound (1.1) with the error term  $\bar{\diamond}$  of the following form:

$$\begin{aligned} \sup_{x \in \mathbb{R}} \left| \mathbb{P}(n \|\widehat{\mathbf{P}}_r - \mathbf{P}_r^*\|_2^2 \leq x) - \mathbb{P}(\|\xi\|^2 \leq x) \right| &\lesssim \bar{\diamond}, \\ \bar{\diamond} = \bar{\diamond}(n, p, \boldsymbol{\Sigma}^*) &\stackrel{\text{def}}{=} \frac{\sqrt{m_r^*} \text{Tr}(\Gamma_r^*)}{\sqrt{\lambda_1(\Gamma_r^*) \lambda_2(\Gamma_r^*)}} \left( \sqrt{\frac{\log(n)}{n}} + \sqrt{\frac{\log(p)}{n}} \right) \\ &\quad + \frac{m_r^*}{g_r^{*3}} \frac{\text{Tr}^3(\boldsymbol{\Sigma}^*)}{\sqrt{\lambda_1(\Gamma_r^*) \lambda_2(\Gamma_r^*)}} \sqrt{\frac{\log^3(n)}{n}}. \end{aligned} \tag{2.5}$$

The goal of this section is to extend this result to include the case of a generalized spectral cluster and of non-Gaussian data. Before formulating the result, let us introduce the following auxiliary matrices

$$\begin{aligned} \mathbf{U}_{\mathcal{J}}^* &\stackrel{\text{def}}{=} \left\{ \frac{u_k^{*\top}}{\sqrt{\mu_r^*}} \right\}_{\substack{r \in \mathcal{J} \\ k \in \Delta_r^*}} \in \mathbb{R}^{m_{\mathcal{J}}^* \times p}, \\ \mathbf{V}_{\mathcal{J}}^* &\stackrel{\text{def}}{=} \left\{ \frac{u_l^{*\top}}{\sqrt{\mu_s^*}} \right\}_{\substack{s \notin \mathcal{J} \\ l \in \Delta_s^*}} \in \mathbb{R}^{(p - m_{\mathcal{J}}^*) \times p}. \end{aligned}$$

Then the following theorem holds.

**Theorem 2.2.** *Assume the distribution of the data  $\mathbf{X}^n = (X_1, \dots, X_n)$  fulfills the sample covariance concentration property (2.3) with some  $\widehat{\delta}_n$ . Suppose additionally that the projections  $\mathbf{P}_{\mathcal{J}}^* X$  and  $(\mathbf{I}_p - \mathbf{P}_{\mathcal{J}}^*) X$  are independent and the following third moments are finite:*

$$\mathbb{E} \|\mathbf{U}_{\mathcal{J}}^* X\|^3 \leq +\infty, \quad \mathbb{E} \|\mathbf{V}_{\mathcal{J}}^* X\|^3 \leq +\infty.$$

Let  $\xi \sim \mathcal{N}(0, \Gamma_{\mathcal{J}}^*)$  with  $\Gamma_{\mathcal{J}}^*$  defined by (2.2). Then

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P} \left( n \|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2 \leq x \right) - \mathbb{P}(\|\xi\|^2 \leq x) \right| \lesssim \bar{\diamond},$$

where

$$\begin{aligned} \bar{\diamond} = \bar{\diamond}(n, p, \boldsymbol{\Sigma}^*) &\stackrel{\text{def}}{=} \mathbb{E}\|\mathbf{U}_{\mathcal{J}}^* X\|^3 \mathbb{E}\|\mathbf{V}_{\mathcal{J}}^* X\|^3 \frac{p^{1/4}}{\sqrt{n}} \\ &+ \frac{\bar{\Delta}}{\|\Gamma_{\mathcal{J}}^*\|_2^{1/2} (\|\Gamma_{\mathcal{J}}^*\|_2^2 - \|\Gamma_{\mathcal{J}}^*\|_{\infty}^2)^{1/4}}, \\ \bar{\Delta} &\stackrel{\text{def}}{=} nm_{\mathcal{J}}^* \left(1 + \frac{l_{\mathcal{J}}^*}{g_{\mathcal{J}}^*}\right) \left\{ \left(1 + \frac{l_{\mathcal{J}}^*}{g_{\mathcal{J}}^*}\right) \frac{\widehat{\delta}_n^4}{g_{\mathcal{J}}^{*4}} \vee \frac{|\mathcal{J}| \widehat{\delta}_n^3}{g_{\mathcal{J}}^{*3}} \right\}. \end{aligned} \quad (2.6)$$

**Remark 2.2.** The condition on independence of  $\mathbf{P}_{\mathcal{J}}^* X$  and  $(\mathbf{I}_p - \mathbf{P}_{\mathcal{J}}^*)X$  is not very restrictive, in fact, it has a natural interpretation: while we are interested in the “signal”  $\mathbf{P}_{\mathcal{J}}^* X$ , the orthogonal part  $(\mathbf{I}_p - \mathbf{P}_{\mathcal{J}}^*)X$  can be considered as “noise”, and it is plausible to assume that the “noise” is independent from the “signal”. It also worth mentioning that this condition was not required in our main result above about the behavior of the pseudo-posterior.

**Remark 2.3.** The components of  $\mathbf{U}_{\mathcal{J}}^* X$  are the  $m_{\mathcal{J}}^*$ -dimensional coordinates of  $X$  after projecting onto the eigenspace of interest and proper scaling. Similarly, the components of  $\mathbf{V}_{\mathcal{J}}^* X$  are the  $(p - m_{\mathcal{J}}^*)$ -dimensional coordinates of  $X$  after projecting onto the orthogonal complement and proper scaling. In general, the factors  $\mathbb{E}\|\mathbf{U}_{\mathcal{J}}^* X\|^3$  and  $\mathbb{E}\|\mathbf{V}_{\mathcal{J}}^* X\|^3$  from the error bound  $\bar{\diamond}$  depend on how heavy the tails of the distribution of  $X$  are. However, it is easy to show that in case of a sub-Gaussian random vector  $X$  the behaviour is as follows:

$$\begin{aligned} \mathbb{E}\|\mathbf{U}_{\mathcal{J}}^* X\|^3 &\lesssim (m_{\mathcal{J}}^*)^{3/2}, \\ \mathbb{E}\|\mathbf{V}_{\mathcal{J}}^* X\|^3 &\lesssim (p - m_{\mathcal{J}}^*)^{3/2}. \end{aligned}$$

Coupled with Theorem A.1, (ii), this allows to bound the error term  $\bar{\diamond}$  in the sub-Gaussian case as

$$\bar{\diamond} \lesssim \sqrt{\frac{(m_{\mathcal{J}}^*)^3 p^{3.5}}{n}} + m_{\mathcal{J}}^* |\mathcal{J}| \sqrt{\frac{p^3 + \log^3(n)}{n}},$$

where, for simplicity, the characteristics of  $\boldsymbol{\Sigma}^*$  are hidden in the constants.

The proof of this result is presented in Appendix B. The obtained bound is worse than (2.5) when the full dimension  $p$  is large. This is the payment for the Gaussian approximation which appears for non-Gaussian data. Our result makes use of the Gaussian approximation technique from [2]. Some recent developments in Gaussian approximation for a probability of a ball indicate that the bound (2.6) can be improved even further; see [30]. Comparison of the results of Theorem 2.1 and Theorem 2.2 reveals that the pseudo-posterior distribution of  $n\|\mathbf{P}_{\mathcal{J}} - \widehat{\mathbf{P}}_{\mathcal{J}}\|_2^2$  given the data perfectly mimics the distribution of  $n\|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2$ , and, therefore, can be applied to building of elliptic confidence

sets for the true projector. Specifically, for any significance level  $\alpha \in (0; 1)$  (or confidence level  $1 - \alpha$ ) we can estimate the true quantile

$$\gamma_\alpha \stackrel{\text{def}}{=} \inf \left\{ \gamma > 0 : \mathbb{P} \left( n \|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2 > \gamma \right) \leq \alpha \right\}$$

by the following counterpart which can be numerically assessed using Bayesian credible sets:

$$\gamma_\alpha^\circ \stackrel{\text{def}}{=} \inf \left\{ \gamma > 0 : \Pi \left( n \|\mathbf{P}_{\mathcal{J}} - \widehat{\mathbf{P}}_{\mathcal{J}}\|_2^2 > \gamma \mid \mathbf{X}^n \right) \leq \alpha \right\}.$$

Then, the main results presented above imply the following corollary.

**Corollary 2.3.** *Assume that all conditions of Theorem 2.1 and Theorem 2.2 are fulfilled. Then*

$$\sup_{\alpha \in (0; 1)} \left| \alpha - \mathbb{P} \left( n \|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2 > \gamma_\alpha^\circ \right) \right| \lesssim \diamond + \overline{\diamond},$$

where  $\diamond = \diamond(n, p, \boldsymbol{\Sigma}^*)$ ,  $\overline{\diamond} = \overline{\diamond}(n, p, \boldsymbol{\Sigma}^*)$  are defined by (2.4), (2.6), respectively.

### 3. Numerical experiments

This section shows by mean of artificial data that the proposed pseudo-Bayesian approach works quite well even for large data dimension and limited sample size. We also want to track how the quality depends on the sample size  $n$  and the dimension  $p$ . In our experiments we first fix some true covariance matrix  $\boldsymbol{\Sigma}^*$  of size  $p \times p$ . Without loss of generality we consider only diagonal matrices  $\boldsymbol{\Sigma}^*$ , so  $\boldsymbol{\Sigma}^*$  is defined by the distinct eigenvalues  $\mu_r^*$  and the multiplicities  $m_r^*$ . We also specify the desired subspace that we want to investigate by fixing  $\mathcal{J}$ . Further, for different sample sizes  $n$  we repeat the following two-step procedure. The first step is to determine the quantiles of  $n \|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2$ . For that we generate 3000 samples  $\mathbf{X}^n$ , compute the corresponding  $\widehat{\mathbf{P}}_{\mathcal{J}}$  and then just take  $\alpha$ -quantiles of the obtained realizations  $n \|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2$  for  $\alpha$  from 0.001 to 0.999 with step 0.001. The second step is to estimate the quantiles of the pseudo-posterior distribution of  $n \|\mathbf{P}_{\mathcal{J}} - \widehat{\mathbf{P}}_{\mathcal{J}}\|_2^2$ . We generate 50 samples  $\mathbf{X}^n$  and for each realization we generate 3000 pseudo-posterior covariance matrices  $\boldsymbol{\Sigma}$  from the Inverse Wishart distribution with  $G = \mathbf{I}_p$ ,  $b = 1$ . Then we compute the corresponding  $\mathbf{P}_{\mathcal{J}}$  and take the  $\alpha$ -quantiles of  $n \|\mathbf{P}_{\mathcal{J}} - \widehat{\mathbf{P}}_{\mathcal{J}}\|_2^2$  just as in the first step. Namely, for each  $\alpha$  we get 50 quantile estimates  $\gamma_\alpha^{\circ(j)}$ ,  $j \in \{1, \dots, 50\}$  (suppose we order them in ascending order) and take median of them. For the true quantiles from the first step and the medians of the quantile estimates from the second step we build a QQ-plot, which consists of points with coordinates  $(\gamma_\alpha, \gamma_\alpha^{\circ(25)})$  for various  $\alpha$ . We expect that the constructed

QQ-plot is close to the identity line indicating that these two distributions are close to each other. Also we present a table with median coverage probabilities

$$\mathbb{P}\left(n\|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2 \leq \gamma_{\alpha}^{\circ(25)}\right)$$

and interquartile ranges

$$\mathbb{P}\left(n\|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2 \leq \gamma_{\alpha}^{\circ(38)}\right) - \mathbb{P}\left(n\|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2 \leq \gamma_{\alpha}^{\circ(12)}\right)$$

of this coverage probability for the desired confidence levels  $1 - \alpha$  from the list

$$\{0.99, 0.95, 0.90, 0.85, 0.80, 0.75\}.$$

In the first experiment we work with Gaussian data. The parameters of the experiment are as follows:

- $p = 100$ ,  $m_r^* = 1$  for all  $r \in \{1, \dots, 100\}$ .
- $\mu_1^* = 25.698$ ,  $\mu_2^* = 15.7688$ ,  $\mu_3^* = 10.0907$ ,  $\mu_4^* = 5.9214$ ,  $\mu_5^* = 3.4321$  and the rest of the eigenvalues  $\mu_6^*, \dots, \mu_{100}^*$  are from the Marchenko – Pastur law with support  $[0.71; 1.34]$ .
- $\mathcal{J} = \{1\}$ , so we investigate the one-dimensional principal subspace given by  $\mathbf{P}_1^*$ .

The QQ-plots are depicted on Figure 1 while the coverage probabilities and the interquartile ranges are presented in Table 1.

The setup of this experiment is exactly the same as the second example of [27], so the performance of our pseudo-Bayesian method can be directly compared with the performance of the Bootstrap approach (cf. Figure 2 and Table 2 of [27]). The accuracy of these two procedures is approximately the same.

In the second experiment we check how our method performs on non-Gaussian data. We generate each component of the vectors  $X_j$  independently yielding diagonal covariance matrix. In addition to Gaussian distribution, we consider also the following three options: the uniform distribution on the interval  $[-a; a]$ , the Laplace distribution with scaling parameter  $a$  and the discrete uniform distribution with three values  $\{-a, 0, a\}$ . In each case the parameter  $a$  is chosen in such a way that ensures the variance located on the diagonal of the covariance matrix fixed earlier. So, the parameters of the experiment are as follows:

- $p = 100$ ,  $m_1^* = 3$ ,  $m_2^* = 3$ ,  $m_3^* = 3$  and the rest of the multiplicities  $m_4^*, \dots, m_{91}^*$  are one.
- $\mu_1^* = 25$ ,  $\mu_2^* = 20$ ,  $\mu_3^* = 15$ ,  $\mu_4^* = 10$ ,  $\mu_5^* = 7.5$ ,  $\mu_6^* = 5$  and the rest of the eigenvalues  $\mu_7^*, \dots, \mu_{100}^*$  are from the uniform distribution on  $[0; 3]$ .
- The first nine components were generated according to: uniform, Laplace, discrete, Gaussian, Laplace, discrete, Laplace, Laplace, uniform distributions, respectively. The rest of the components are Gaussian.
- $\mathcal{J} = \{1, 2, 3\}$ , so we investigate nine-dimensional subspace given by  $\mathbf{P}_1^* + \mathbf{P}_2^* + \mathbf{P}_3^*$ .

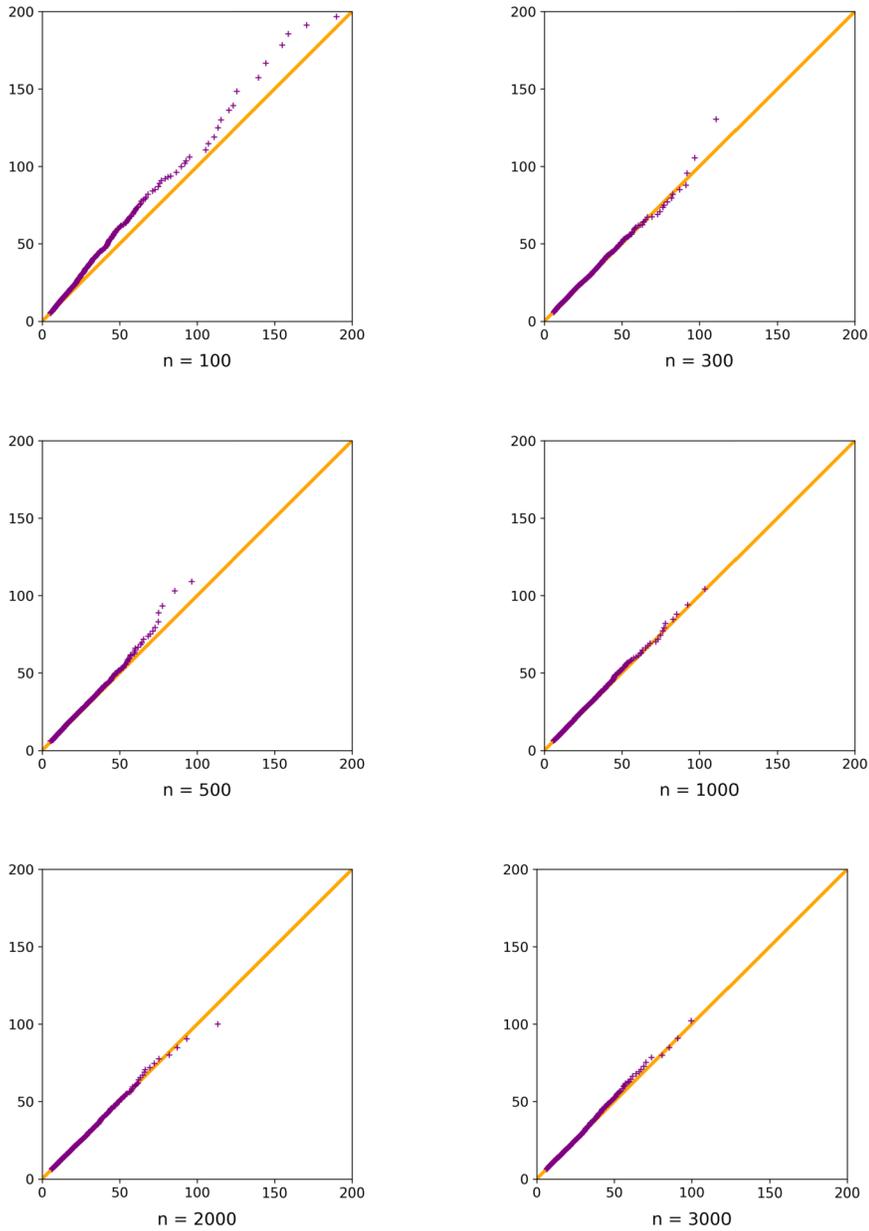


FIG 1. QQ-plots of the proposed pseudo-Bayesian procedure for the first experiment (Gaussian data).

The QQ-plots are depicted on Figure 2 while the coverage probabilities and the interquartile ranges are presented in Table 2.

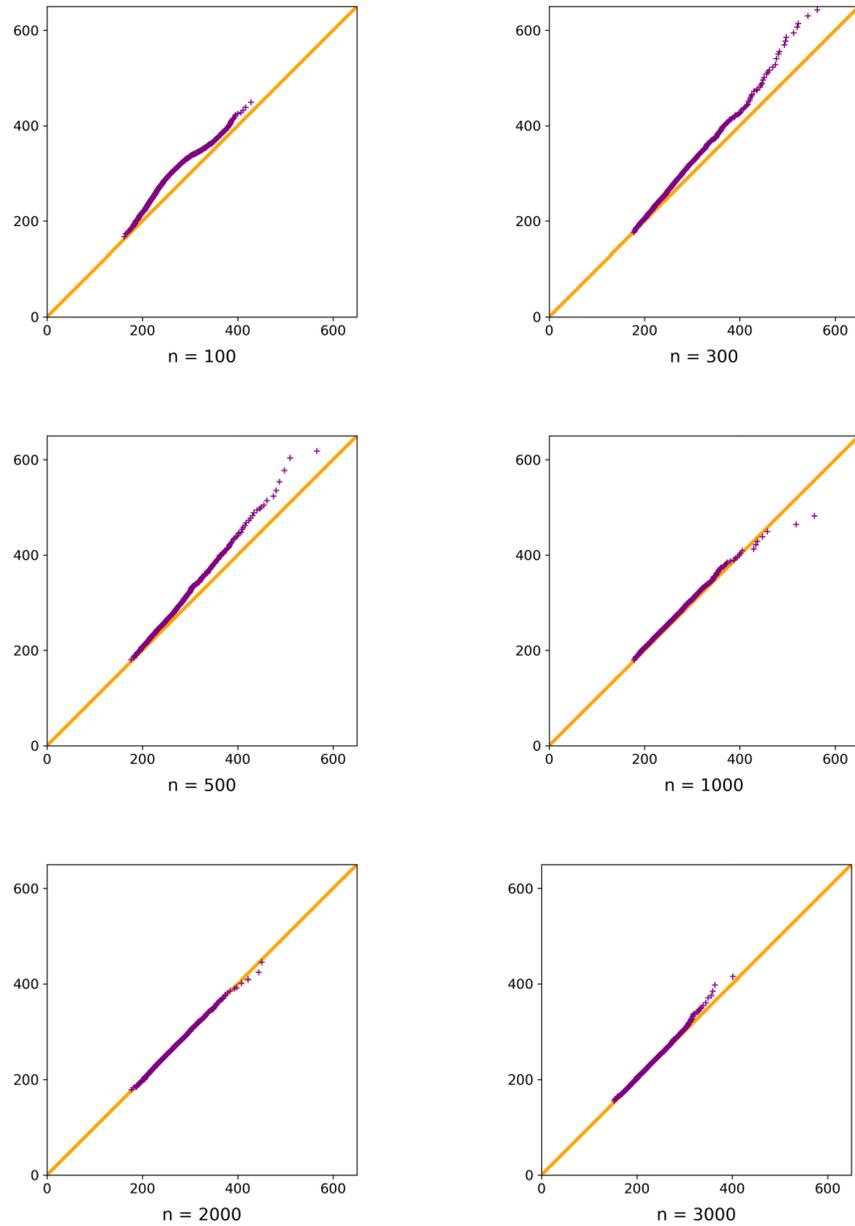


FIG 2. QQ-plots of the proposed pseudo-Bayesian procedure for the second experiment (non-Gaussian data).

The performance of the proposed procedure is very good except the case when the sample size is of the same order as the dimension. However, this

TABLE 1  
 Coverage probabilities and interquartile ranges of the proposed pseudo-Bayesian procedure for the first experiment (Gaussian data).

$n$	Confidence levels ( $1 - \alpha$ )					
	0.99	0.95	0.90	0.85	0.80	0.75
100	0.993	0.968	0.929	0.893	0.854	0.809
	0.023	0.061	0.103	0.145	0.176	0.204
300	0.988	0.952	0.906	0.851	0.805	0.762
	0.026	0.085	0.143	0.184	0.199	0.216
500	0.993	0.955	0.909	0.865	0.812	0.771
	0.022	0.072	0.099	0.108	0.123	0.126
1000	0.990	0.956	0.908	0.859	0.817	0.767
	0.014	0.049	0.066	0.067	0.091	0.104
2000	0.992	0.952	0.898	0.847	0.793	0.747
	0.009	0.030	0.058	0.064	0.067	0.065
3000	0.992	0.959	0.908	0.849	0.802	0.750
	0.005	0.024	0.036	0.054	0.046	0.063

TABLE 2  
 Coverage probabilities and interquartile ranges of the proposed pseudo-Bayesian procedure for the second experiment (non-Gaussian data)

$n$	Confidence levels ( $1 - \alpha$ )					
	0.99	0.95	0.90	0.85	0.80	0.75
100	0.997	0.979	0.954	0.927	0.895	0.870
	0.009	0.044	0.070	0.085	0.094	0.099
300	0.993	0.964	0.935	0.903	0.868	0.836
	0.008	0.047	0.077	0.108	0.144	0.176
500	0.996	0.972	0.944	0.917	0.874	0.832
	0.014	0.053	0.099	0.139	0.166	0.194
1000	0.990	0.957	0.920	0.882	0.841	0.796
	0.011	0.050	0.098	0.131	0.171	0.188
2000	0.991	0.951	0.904	0.850	0.803	0.755
	0.019	0.048	0.088	0.113	0.124	0.146
3000	0.994	0.958	0.913	0.863	0.813	0.771
	0.007	0.033	0.054	0.069	0.079	0.092

regime lies beyond the scope of our results. If we have enough data, the method demonstrates very good results even in such challenging situations as recovering a direct sum of several subspaces from non-Gaussian (even not sub-Gaussian) data.

#### 4. Main proofs

This section collects the proofs of the main results. Some additional technical statements are postponed to the Appendix.

##### 4.1. Proof of Theorem 2.1

The Inverse Wishart prior  $\mathcal{IW}_p(\mathbf{G}, p + b - 1)$  is conjugate to the multivariate Gaussian distribution, so our pseudo-posterior  $\Pi(\boldsymbol{\Sigma} \mid \mathbf{X}^n)$  is

$\mathcal{IW}_p(\mathbf{G} + n\widehat{\Sigma}, n + p + b - 1)$ . We will actively use the following well-known property of the Wishart distribution:

$$\Sigma^{-1} \mid \mathbf{X}^n \stackrel{d}{=} \sum_{j=1}^{n+p+b-1} W_j W_j^\top \mid \mathbf{X}^n,$$

where  $W_j \mid \mathbf{X}^n \stackrel{i.i.d.}{\sim} \mathcal{N}(0, (\mathbf{G} + n\widehat{\Sigma})^{-1})$ .

For shortness in this section we will use the notation  $n_p \stackrel{\text{def}}{=} n + p + b - 1$  and we assume that  $b \lesssim p$ . As we will see, this assumption will help us to simplify the bounds, while the case  $b \gtrsim p$  does not bring any gain. Moreover, define

$$\Sigma_{n,p} \stackrel{\text{def}}{=} \frac{1}{n_p} \mathbf{G} + \frac{n}{n_p} \widehat{\Sigma}$$

and

$$\mathbf{E}_{n,p} \stackrel{\text{def}}{=} \frac{1}{n_p} \sum_{j=1}^{n_p} Z_j Z_j^\top - \mathbf{I}_p,$$

where  $Z_j \mid \mathbf{X}^n \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \mathbf{I}_p)$ . Then  $\Sigma^{-1} \mid \mathbf{X}^n$  can be represented as

$$\Sigma^{-1} \mid \mathbf{X}^n \stackrel{d}{=} \Sigma_{n,p}^{-1/2} (\mathbf{E}_{n,p} + \mathbf{I}_p) \Sigma_{n,p}^{-1/2}.$$

We may think that in the ‘‘posterior’’ world all randomness comes from  $\mathbf{E}_{n,p}$ . Moreover, due to Theorem A.1, (i), there is a random set  $\mathcal{Y}$  such that on this set

$$\|\mathbf{E}_{n,p}\|_\infty \lesssim \sqrt{\frac{\log(n_p) + p}{n_p}} \leq \sqrt{\frac{\log(n) + p}{n}},$$

and its pseudo-posterior measure

$$\Pi(\mathcal{Y} \mid \mathbf{X}^n) \geq 1 - \frac{1}{n}.$$

**Step 1** First, we will need the following lemma.

**Lemma 4.1.** *The following holds on the random set  $\mathcal{Y}$ :*

$$\|\Sigma - \widehat{\Sigma}\|_\infty \lesssim \|\widehat{\Sigma}\|_\infty \sqrt{\frac{\log(n) + p}{n}} + \frac{\|\mathbf{G}\|_\infty}{n}. \quad (4.1)$$

*Proof.* Since  $\Sigma^{-1} | \mathbf{X}^n \stackrel{d}{=} \Sigma_{n,p}^{-1/2} (\mathbf{E}_{n,p} + \mathbf{I}_p) \Sigma_{n,p}^{-1/2}$ , we have

$$\begin{aligned} \Sigma - \widehat{\Sigma} &= \Sigma_{n,p}^{1/2} (\mathbf{E}_{n,p} + \mathbf{I}_p)^{-1} \Sigma_{n,p}^{1/2} - \widehat{\Sigma} \\ &= \Sigma_{n,p}^{1/2} [(\mathbf{E}_{n,p} + \mathbf{I}_p)^{-1} - \mathbf{I}_p] \Sigma_{n,p}^{1/2} + \Sigma_{n,p} - \widehat{\Sigma}. \end{aligned}$$

Note that

$$\begin{aligned} \|(\mathbf{E}_{n,p} + \mathbf{I}_p)^{-1} - \mathbf{I}_p\|_\infty &= \left\| \sum_{s=1}^{\infty} (-\mathbf{E}_{n,p})^s \right\|_\infty \\ &\leq \sum_{s=1}^{\infty} \|\mathbf{E}_{n,p}\|_\infty^s = \frac{\|\mathbf{E}_{n,p}\|_\infty}{1 - \|\mathbf{E}_{n,p}\|_\infty} \lesssim \|\mathbf{E}_{n,p}\|_\infty. \end{aligned}$$

Hence,

$$\|\Sigma - \widehat{\Sigma}\|_\infty \lesssim \|\Sigma_{n,p}\|_\infty \|\mathbf{E}_{n,p}\|_\infty + \|\Sigma_{n,p} - \widehat{\Sigma}\|_\infty.$$

Finally, the observations that

$$\begin{aligned} \|\Sigma_{n,p}\|_\infty &\leq \frac{\|\mathbf{G}\|_\infty}{n} + \|\widehat{\Sigma}\|_\infty, \\ \|\Sigma_{n,p} - \widehat{\Sigma}\|_\infty &\leq \frac{\|\mathbf{G}\|_\infty}{n} + \frac{n_p - n}{n} \|\widehat{\Sigma}\|_\infty, \end{aligned}$$

finish the proof. □

The condition on the significant spectral gap for  $\Sigma^*$  and the bound (2.3) on the operator norm  $\|\widehat{\Sigma} - \Sigma^*\|$  imply a significant spectral gap for the empirical covariance  $\widehat{\Sigma}$ . The crucial Lemma A.2 applied with the central projector  $\widehat{\mathbf{P}}_{\mathcal{J}}$  in place of  $\mathbf{P}_{\mathcal{J}}^*$  allows to obtain the bound on how close the linear operator

$$\widehat{L}_{\mathcal{J}}(\Sigma - \widehat{\Sigma}) \stackrel{\text{def}}{=} \sum_{k \in \mathcal{I}_{\mathcal{J}}} \sum_{l \notin \mathcal{I}_{\mathcal{J}}} \frac{\widehat{u}_k \widehat{u}_k^\top (\Sigma - \widehat{\Sigma}) \widehat{u}_l \widehat{u}_l^\top + \widehat{u}_l \widehat{u}_l^\top (\Sigma - \widehat{\Sigma}) \widehat{u}_k \widehat{u}_k^\top}{\widehat{\sigma}_k - \widehat{\sigma}_l}$$

is to  $\mathbf{P}_{\mathcal{J}} - \widehat{\mathbf{P}}_{\mathcal{J}}$ .

**Lemma 4.2.** *The following holds on the random set  $\Upsilon$ :*

$$\sqrt{n} \|\mathbf{P}_{\mathcal{J}} - \widehat{\mathbf{P}}_{\mathcal{J}} - \widehat{L}_{\mathcal{J}}(\Sigma - \widehat{\Sigma})\|_2 \lesssim \widehat{\Delta}_0,$$

where

$$\widehat{\Delta}_0 \stackrel{\text{def}}{=} \sqrt{\frac{m_{\mathcal{J}}^*}{n}} \left( 1 + \frac{\widehat{l}_{\mathcal{J}}}{\widehat{g}_{\mathcal{J}}} \right) \frac{(\log(n) + p) \|\widehat{\Sigma}\|_\infty^2 + \|\mathbf{G}\|_\infty^2/n}{\widehat{g}_{\mathcal{J}}^2},$$

and  $\widehat{l}_{\mathcal{J}}, \widehat{g}_{\mathcal{J}}$  are empirical versions of  $l_{\mathcal{J}}^*, g_{\mathcal{J}}^*$ .

*Proof.* It follows from (A.2) from Lemma A.2 that

$$\|\mathbf{P}_{\mathcal{J}} - \widehat{\mathbf{P}}_{\mathcal{J}} - \widehat{\mathbf{L}}_{\mathcal{J}}(\boldsymbol{\Sigma} - \widehat{\boldsymbol{\Sigma}})\|_{\infty} \lesssim \left(1 + \frac{\widehat{l}_{\mathcal{J}}}{\widehat{g}_{\mathcal{J}}}\right) \frac{\|\boldsymbol{\Sigma} - \widehat{\boldsymbol{\Sigma}}\|_{\infty}^2}{\widehat{g}_{\mathcal{J}}^2}.$$

It is easy to see that the rank of  $\widehat{\mathbf{L}}_{\mathcal{J}}(\boldsymbol{\Sigma} - \widehat{\boldsymbol{\Sigma}})$  is at most  $2m_{\mathcal{J}}^*$ , and thus the rank of  $\mathbf{P}_{\mathcal{J}} - \widehat{\mathbf{P}}_{\mathcal{J}} - \widehat{\mathbf{L}}_{\mathcal{J}}(\boldsymbol{\Sigma} - \widehat{\boldsymbol{\Sigma}})$  is at most  $4m_{\mathcal{J}}^*$ . Hence, taking into account the relation between the Frobenius and the spectral norm of a matrix via rank and (4.1) from Lemma 4.1, we obtain the desired statement.  $\square$

The representation

$$\boldsymbol{\Sigma}^{-1} | \mathbf{X}^n \stackrel{d}{=} \boldsymbol{\Sigma}_{n,p}^{-1/2} (\mathbf{E}_{n,p} + \mathbf{I}_p) \boldsymbol{\Sigma}_{n,p}^{-1/2}.$$

helps to obtain the next result showing that  $\widehat{\mathbf{L}}_{\mathcal{J}}(\boldsymbol{\Sigma} - \widehat{\boldsymbol{\Sigma}})$  can be approximated by  $\widehat{\mathbf{S}}_{\mathcal{J}} = \widehat{\mathbf{L}}_{\mathcal{J}} \left( -\widehat{\boldsymbol{\Sigma}}^{1/2} \mathbf{E}_{n,p} \widehat{\boldsymbol{\Sigma}}^{1/2} \right)$ .

**Lemma 4.3.** *It holds*

$$\widehat{\mathbf{L}}_{\mathcal{J}}(\boldsymbol{\Sigma}) = \widehat{\mathbf{L}}_{\mathcal{J}} \left( -\widehat{\boldsymbol{\Sigma}}^{1/2} \mathbf{E}_{n,p} \widehat{\boldsymbol{\Sigma}}^{1/2} \right) + \mathcal{R}_{\mathcal{J}} = \widehat{\mathbf{S}}_{\mathcal{J}} + \mathcal{R}_{\mathcal{J}},$$

where the remainder  $\mathcal{R}_{\mathcal{J}}$  fulfills on the random set  $\Upsilon$

$$\sqrt{n} \|\mathcal{R}_{\mathcal{J}}\|_2 \lesssim \widehat{\Delta}_1 \stackrel{\text{def}}{=} \frac{m_{\mathcal{J}}^*}{\sqrt{n}} \cdot \frac{(\log(n) + p) \|\widehat{\boldsymbol{\Sigma}}\|_{\infty} + \|\mathbf{G}\|_{\infty}}{\widehat{g}_{\mathcal{J}}}.$$

*Proof.* Define  $\mathbf{R}_{n,p}$  by

$$\mathbf{R}_{n,p} \stackrel{\text{def}}{=} (\mathbf{I}_p + \mathbf{E}_{n,p})^{-1} - \mathbf{I}_p + \mathbf{E}_{n,p}.$$

Its spectral norm can be bounded as

$$\|\mathbf{R}_{n,p}\|_{\infty} \lesssim \left\| \sum_{s=2}^{\infty} (-\mathbf{E}_{n,p})^s \right\|_{\infty} \leq \sum_{s=2}^{\infty} \|\mathbf{E}_{n,p}\|_{\infty}^s = \frac{\|\mathbf{E}_{n,p}\|_{\infty}^2}{1 - \|\mathbf{E}_{n,p}\|_{\infty}} \lesssim \|\mathbf{E}_{n,p}\|_{\infty}^2.$$

So

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{n,p}^{1/2} (\mathbf{E}_{n,p} + \mathbf{I}_p)^{-1} \boldsymbol{\Sigma}_{n,p}^{1/2} = \boldsymbol{\Sigma}_{n,p}^{1/2} (\mathbf{I}_p - \mathbf{E}_{n,p} + \mathbf{R}_{n,p}) \boldsymbol{\Sigma}_{n,p}^{1/2}.$$

Therefore for  $\boldsymbol{\Sigma} - \widehat{\boldsymbol{\Sigma}}$  we have

$$\begin{aligned} \boldsymbol{\Sigma} - \widehat{\boldsymbol{\Sigma}} &= \boldsymbol{\Sigma}_{n,p}^{1/2} (\mathbf{I}_p - \mathbf{E}_{n,p} + \mathbf{R}_{n,p}) \boldsymbol{\Sigma}_{n,p}^{1/2} - \widehat{\boldsymbol{\Sigma}} \\ &= -\boldsymbol{\Sigma}_{n,p}^{1/2} \mathbf{E}_{n,p} \boldsymbol{\Sigma}_{n,p}^{1/2} + \boldsymbol{\Sigma}_{n,p}^{1/2} \mathbf{R}_{n,p} \boldsymbol{\Sigma}_{n,p}^{1/2} + \boldsymbol{\Sigma}_{n,p} - \widehat{\boldsymbol{\Sigma}}. \end{aligned}$$

From  $\Sigma_{n,p}^{1/2} \mathbf{E}_{n,p} \Sigma_{n,p}^{1/2}$  we pass to  $\widehat{\Sigma}^{1/2} \mathbf{E}_{n,p} \widehat{\Sigma}^{1/2}$ :

$$\begin{aligned} \Sigma - \widehat{\Sigma} &= -\widehat{\Sigma}^{1/2} \mathbf{E}_{n,p} \widehat{\Sigma}^{1/2} + (\widehat{\Sigma}^{1/2} \mathbf{E}_{n,p} \widehat{\Sigma}^{1/2} - \Sigma_{n,p}^{1/2} \mathbf{E}_{n,p} \Sigma_{n,p}^{1/2}) \\ &\quad + \Sigma_{n,p}^{1/2} \mathbf{R}_{n,p} \Sigma_{n,p}^{1/2} + \Sigma_{n,p} - \widehat{\Sigma} \\ &= -\widehat{\Sigma}^{1/2} \mathbf{E}_{n,p} \widehat{\Sigma}^{1/2} + \mathbf{R}_1 + \mathbf{R}_2 + \mathbf{R}_3, \end{aligned}$$

where we introduce the remainder terms

$$\mathbf{R}_1 \stackrel{\text{def}}{=} \widehat{\Sigma}^{1/2} \mathbf{E}_{n,p} \widehat{\Sigma}^{1/2} - \Sigma_{n,p}^{1/2} \mathbf{E}_{n,p} \Sigma_{n,p}^{1/2},$$

$$\mathbf{R}_2 \stackrel{\text{def}}{=} \Sigma_{n,p}^{1/2} \mathbf{R}_{n,p} \Sigma_{n,p}^{1/2},$$

$$\mathbf{R}_3 \stackrel{\text{def}}{=} \Sigma_{n,p} - \widehat{\Sigma}.$$

They can be bounded as

$$\begin{aligned} \|\mathbf{R}_1\|_\infty &\leq \|\mathbf{E}_{n,p}\|_\infty \|\widehat{\Sigma} - \Sigma_{n,p}\|_\infty^{1/2} \left( \|\Sigma_{n,p}\|_\infty^{1/2} + \|\widehat{\Sigma}\|_\infty^{1/2} \right), \\ \|\mathbf{R}_2\|_\infty &\leq \|\mathbf{R}_{n,p}\|_\infty \|\Sigma_{n,p}\|_\infty \lesssim \|\mathbf{E}_{n,p}\|_\infty^2 \|\Sigma_{n,p}\|_\infty, \\ \|\mathbf{R}_3\|_\infty &\lesssim \frac{\|\mathbf{G}\|_\infty + (n_p - n) \|\widehat{\Sigma}\|_\infty}{n_p}. \end{aligned}$$

Hence, omitting higher order terms, on  $\Upsilon$  we have

$$\begin{aligned} \|\mathbf{R}_1\|_\infty &\lesssim \|\widehat{\Sigma}\|_\infty^{1/2} \left( \|\mathbf{G}\|_\infty + p \|\widehat{\Sigma}\|_\infty \right)^{1/2} \frac{\sqrt{\log(n) + p}}{n}, \\ \|\mathbf{R}_2\|_\infty &\lesssim \|\widehat{\Sigma}\|_\infty \frac{\log(n) + p}{n}, \\ \|\mathbf{R}_3\|_\infty &\lesssim \frac{\|\mathbf{G}\|_\infty + p \|\widehat{\Sigma}\|_\infty}{n}. \end{aligned}$$

Now we summarize

$$\widehat{L}_{\mathcal{J}}(\Sigma - \widehat{\Sigma}) = \widehat{S}_{\mathcal{J}} + \mathcal{R}_{\mathcal{J}}$$

with

$$\begin{aligned} \widehat{S}_{\mathcal{J}} &\stackrel{\text{def}}{=} - \sum_{k \in \mathcal{I}_{\mathcal{J}}} \sum_{l \notin \mathcal{I}_{\mathcal{J}}} \frac{\widehat{\sigma}_k^{1/2} \widehat{\sigma}_l^{1/2} (\widehat{u}_k \widehat{u}_k^\top \mathbf{E}_{n,p} \widehat{u}_l \widehat{u}_l^\top + \widehat{u}_l \widehat{u}_l^\top \mathbf{E}_{n,p} \widehat{u}_k \widehat{u}_k^\top)}{\widehat{\sigma}_k - \widehat{\sigma}_l}, \\ \mathcal{R}_{\mathcal{J}} &\stackrel{\text{def}}{=} \sum_{k \in \mathcal{I}_{\mathcal{J}}} \sum_{l \notin \mathcal{I}_{\mathcal{J}}} \frac{(\widehat{u}_k \widehat{u}_k^\top (\mathbf{R}_1 + \mathbf{R}_2 + \mathbf{R}_3) \widehat{u}_l \widehat{u}_l^\top + \widehat{u}_l \widehat{u}_l^\top (\mathbf{R}_1 + \mathbf{R}_2 + \mathbf{R}_3) \widehat{u}_k \widehat{u}_k^\top)}{\widehat{\sigma}_k - \widehat{\sigma}_l}. \end{aligned}$$

Moreover,

$$\begin{aligned}
\|\mathcal{R}_{\mathcal{J}}\|_2 &\leq 2 \left\| \sum_{k \in \mathcal{I}_{\mathcal{J}}} \hat{u}_k \hat{u}_k^\top \sum_{l \notin \mathcal{I}_{\mathcal{J}}} \frac{(\mathbf{R}_1 + \mathbf{R}_2 + \mathbf{R}_3) \hat{u}_l \hat{u}_l^\top}{\hat{\sigma}_k - \hat{\sigma}_l} \right\|_2 \\
&\leq 2 \sum_{k \in \mathcal{I}_{\mathcal{J}}} \|\hat{u}_k \hat{u}_k^\top\|_2 \left\| \sum_{l \notin \mathcal{I}_{\mathcal{J}}} \frac{(\mathbf{R}_1 + \mathbf{R}_2 + \mathbf{R}_3) \hat{u}_l \hat{u}_l^\top}{\hat{\sigma}_k - \hat{\sigma}_l} \right\|_\infty \\
&\leq 2 \sum_{k \in \mathcal{I}_{\mathcal{J}}} \|\mathbf{R}_1 + \mathbf{R}_2 + \mathbf{R}_3\|_\infty \left\| \sum_{l \notin \mathcal{I}_{\mathcal{J}}} \frac{\hat{u}_l \hat{u}_l^\top}{\hat{\sigma}_k - \hat{\sigma}_l} \right\|_\infty \\
&\leq \frac{2m_{\mathcal{J}}^*}{\hat{g}_{\mathcal{J}}} (\|\mathbf{R}_1\|_\infty + \|\mathbf{R}_2\|_\infty + \|\mathbf{R}_3\|_\infty),
\end{aligned}$$

which provides the desired bound. Similarly, we have

$$\|\hat{\mathcal{S}}_{\mathcal{J}}\|_2 \leq \frac{2m_{\mathcal{J}}^* \|\hat{\Sigma}\|_\infty}{\hat{g}_{\mathcal{J}}} \|\mathbf{E}_{n,p}\|_\infty \lesssim \frac{m_{\mathcal{J}}^* \|\hat{\Sigma}\|_\infty}{\hat{g}_{\mathcal{J}}} \sqrt{\frac{\log(n) + p}{n}},$$

where the last inequality holds on  $\mathcal{Y}$ . □

The results of Lemmas 4.2 and 4.3 yield on the random set  $\mathcal{Y}$

$$\sqrt{n} \|\hat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}} - \hat{\mathcal{S}}_{\mathcal{J}}\|_2 \lesssim \hat{\Delta}_0 + \hat{\Delta}_1.$$

In addition,

$$\begin{aligned}
&\left| n \|\mathbf{P}_{\mathcal{J}} - \hat{\mathbf{P}}_{\mathcal{J}}\|_2^2 - n \|\hat{\mathcal{S}}_{\mathcal{J}}\|_2^2 \right| \\
&= n \|\mathbf{P}_{\mathcal{J}} - \hat{\mathbf{P}}_{\mathcal{J}} - \hat{\mathcal{S}}_{\mathcal{J}}\|_2^2 + 2 \left\langle \sqrt{n}(\mathbf{P}_{\mathcal{J}} - \hat{\mathbf{P}}_{\mathcal{J}} - \hat{\mathcal{S}}_{\mathcal{J}}), \sqrt{n} \hat{\mathcal{S}}_{\mathcal{J}} \right\rangle_2 \\
&\leq n \|\mathbf{P}_{\mathcal{J}} - \hat{\mathbf{P}}_{\mathcal{J}} - \hat{\mathcal{S}}_{\mathcal{J}}\|_2^2 + 2\sqrt{n} \|\mathbf{P}_{\mathcal{J}} - \hat{\mathbf{P}}_{\mathcal{J}} - \hat{\mathcal{S}}_{\mathcal{J}}\|_2 \cdot \sqrt{n} \|\hat{\mathcal{S}}_{\mathcal{J}}\|_2.
\end{aligned}$$

Thus, taking into account the bound for  $\|\hat{\mathcal{S}}_{\mathcal{J}}\|_2$  and neglecting higher order terms, on  $\mathcal{Y}$  we obtain

$$\left| n \|\mathbf{P}_{\mathcal{J}} - \hat{\mathbf{P}}_{\mathcal{J}}\|_2^2 - n \|\hat{\mathcal{S}}_{\mathcal{J}}\|_2^2 \right| \lesssim \hat{\Delta}_2, \quad (4.2)$$

where

$$\begin{aligned}
\hat{\Delta}_2 \stackrel{\text{def}}{=} &\left\{ (\log(n) + p) \left( \left( 1 + \frac{\hat{l}_{\mathcal{J}}}{\hat{g}_{\mathcal{J}}} \right) \frac{\sqrt{m_{\mathcal{J}}^*} \|\hat{\Sigma}\|_\infty}{\hat{g}_{\mathcal{J}}} + m_{\mathcal{J}}^* \right) \|\hat{\Sigma}\|_\infty + m_{\mathcal{J}}^* \|\mathbf{G}\|_\infty \right\} \times \\
&\times \frac{m_{\mathcal{J}}^* \|\hat{\Sigma}\|_\infty}{\hat{g}_{\mathcal{J}}^2} \sqrt{\frac{\log(n) + p}{n}}.
\end{aligned}$$

**Step 2** The norm  $n\|\widehat{S}_{\mathcal{J}}\|_2^2$  can be decomposed as follows:

$$\begin{aligned} n\|\widehat{S}_{\mathcal{J}}\|_2^2 &= 2n \sum_{k'=1}^p \sum_{l'=1}^p \sum_{k \in \mathcal{I}_{\mathcal{J}}} \sum_{l \notin \mathcal{I}_{\mathcal{J}}} \frac{\widehat{\sigma}_k \widehat{\sigma}_l}{(\widehat{\sigma}_k - \widehat{\sigma}_l)^2} (\widehat{u}_{k'}^\top \widehat{u}_k \widehat{u}_k^\top \mathbf{E}_{n,p} \widehat{u}_l \widehat{u}_l^\top \widehat{u}_{l'})^2 = \\ &= 2n \sum_{k \in \mathcal{I}_{\mathcal{J}}} \sum_{l \notin \mathcal{I}_{\mathcal{J}}} \frac{\widehat{\sigma}_k \widehat{\sigma}_l}{(\widehat{\sigma}_k - \widehat{\sigma}_l)^2} (\widehat{u}_k^\top \mathbf{E}_{n,p} \widehat{u}_l)^2. \end{aligned}$$

Introduce a vector  $\widehat{\xi}_{\mathcal{J}} \in \mathbb{R}^{m_{\mathcal{J}}^*(p-m_{\mathcal{J}}^*)}$  with components

$$\widehat{\xi}_{k,l} = \sqrt{2n} \frac{\widehat{\sigma}_k^{1/2} \widehat{\sigma}_l^{1/2}}{\widehat{\sigma}_k - \widehat{\sigma}_l} \widehat{u}_k^\top \mathbf{E}_{n,p} \widehat{u}_l,$$

for  $k \in \mathcal{I}_{\mathcal{J}}, l \notin \mathcal{I}_{\mathcal{J}}$ , ordered in some particular way that will become clear later. Note that  $n\|\widehat{S}_{\mathcal{J}}\|_2^2 = \|\widehat{\xi}_{\mathcal{J}}\|^2$ . Clearly, for each  $k \leq p$  and  $j \leq n_p$

$$\eta_{k,j} \stackrel{\text{def}}{=} \widehat{u}_k^\top Z_j \mid \mathbf{X}^n \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1).$$

Then the components can be rewritten as

$$\widehat{\xi}_{k,l} = \frac{\sqrt{2} \widehat{\sigma}_k^{1/2} \widehat{\sigma}_l^{1/2}}{\widehat{\sigma}_k - \widehat{\sigma}_l} \frac{\sqrt{n}}{n_p} \sum_{j=1}^{n_p} \eta_{k,j} \eta_{l,j},$$

for  $k \in \mathcal{I}_{\mathcal{J}}, l \notin \mathcal{I}_{\mathcal{J}}$ . To understand the covariance structure of  $\widehat{\xi}_{\mathcal{J}}$ , consider one more pair  $(k', l')$  and investigate the covariance:

$$\begin{aligned} \widehat{\Gamma}_{(k,l),(k',l')} &\stackrel{\text{def}}{=} \text{Cov}(\widehat{\xi}_{k,l}, \widehat{\xi}_{k',l'} \mid \mathbf{X}^n) \\ &= \frac{2n}{n_p^2} \sum_{j,j'=1}^{n_p} \frac{\widehat{\sigma}_k^{1/2} \widehat{\sigma}_l^{1/2} \widehat{\sigma}_{k'}^{1/2} \widehat{\sigma}_{l'}^{1/2}}{(\widehat{\sigma}_k - \widehat{\sigma}_l)(\widehat{\sigma}_{k'} - \widehat{\sigma}_{l'})} \mathbb{E}(\eta_{k,j} \eta_{l,j} \eta_{k',j'} \eta_{l',j'} \mid \mathbf{X}^n) \\ &= \frac{2n}{n_p} \delta_{k,k'} \delta_{l,l'} \frac{\widehat{\sigma}_k \widehat{\sigma}_l}{(\widehat{\sigma}_k - \widehat{\sigma}_l)^2} \end{aligned}$$

with  $\delta_{k,k'} = \mathbb{1}(k = k')$ . Therefore, the covariance matrix of  $\widehat{\xi}_{\mathcal{J}}$  is diagonal:

$$\widehat{\Gamma}_{\mathcal{J}} \stackrel{\text{def}}{=} \frac{2n}{n_p} \cdot \text{diag} \left( \frac{2 \widehat{\sigma}_k \widehat{\sigma}_l}{(\widehat{\sigma}_k - \widehat{\sigma}_l)^2} \right)_{k \in \mathcal{I}_{\mathcal{J}}, l \notin \mathcal{I}_{\mathcal{J}}}.$$

This matrix  $\widehat{\Gamma}_{\mathcal{J}}$  can be compared with the matrix  $\Gamma_{\mathcal{J}}^*$  defined in (2.2).

**Lemma 4.4.** *On the event where  $\|\widehat{\Sigma} - \Sigma^*\|_{\infty} \leq g_{\mathcal{J}}^*/4$  it holds*

$$\|\widehat{\Gamma}_{\mathcal{J}} - \Gamma_{\mathcal{J}}^*\|_1 \lesssim \widehat{\Delta}_3 \tag{4.3}$$

with

$$\widehat{\Delta}_3 \stackrel{\text{def}}{=} \frac{p \left( m_{\mathcal{J}}^* \|\boldsymbol{\Sigma}^*\|_{\infty}^2 \wedge \text{Tr}(\boldsymbol{\Sigma}^{*2}) \right)}{g_{\mathcal{J}}^{*3}} \left( \|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*\|_{\infty} + \frac{p}{n} \|\boldsymbol{\Sigma}^*\|_{\infty} \right).$$

*Proof.* As both matrices  $\widehat{\Gamma}_{\mathcal{J}}$  and  $\Gamma_{\mathcal{J}}^*$  are diagonal, it holds

$$\|\widehat{\Gamma}_{\mathcal{J}} - \Gamma_{\mathcal{J}}^*\|_1 \leq 2 \sum_{k \in \mathcal{I}_{\mathcal{J}}} \sum_{l \notin \mathcal{I}_{\mathcal{J}}} \left| \frac{n}{n_p} \frac{\widehat{\sigma}_k \widehat{\sigma}_l}{(\widehat{\sigma}_k - \widehat{\sigma}_l)^2} - \frac{\sigma_k^* \sigma_l^*}{(\sigma_k^* - \sigma_l^*)^2} \right|.$$

Let us fix arbitrary  $k \in \mathcal{I}_{\mathcal{J}}, l \notin \mathcal{I}_{\mathcal{J}}$  and upperbound the corresponding term of the sum. We will extensively use  $|\widehat{\sigma}_k - \sigma_k^*| \leq \|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*\|_{\infty}$  and  $|\widehat{\sigma}_l - \sigma_l^*| \leq \|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*\|_{\infty}$  which holds due to the Weyl's inequality.

So, we have

$$\begin{aligned} \left| \frac{n}{n_p} \frac{\widehat{\sigma}_k \widehat{\sigma}_l}{(\widehat{\sigma}_k - \widehat{\sigma}_l)^2} - \frac{\sigma_k^* \sigma_l^*}{(\sigma_k^* - \sigma_l^*)^2} \right| &\leq \left| \frac{n}{n_p} \frac{\widehat{\sigma}_k \widehat{\sigma}_l}{(\widehat{\sigma}_k - \widehat{\sigma}_l)^2} - \frac{n}{n_p} \frac{\sigma_k^* \sigma_l^*}{(\sigma_k^* - \sigma_l^*)^2} \right| \\ &\quad + \left| \left( \frac{n}{n_p} - 1 \right) \frac{\sigma_k^* \sigma_l^*}{(\sigma_k^* - \sigma_l^*)^2} \right|. \end{aligned}$$

Since  $n/n_p \leq 1$ , the first term is controlled by

$$\begin{aligned} \left| \frac{\widehat{\sigma}_k \widehat{\sigma}_l}{(\widehat{\sigma}_k - \widehat{\sigma}_l)^2} - \frac{\sigma_k^* \sigma_l^*}{(\sigma_k^* - \sigma_l^*)^2} \right| &\leq \left| \frac{\widehat{\sigma}_k \widehat{\sigma}_l (\sigma_k^* - \sigma_l^*)^2 - \sigma_k^* \sigma_l^* (\widehat{\sigma}_k - \widehat{\sigma}_l)^2}{(\sigma_k^* - \sigma_l^*)^2 (\widehat{\sigma}_k - \widehat{\sigma}_l)^2} \right| \\ &= \left| \frac{(\sigma_k^* + \varepsilon_k) (\sigma_l^* + \varepsilon_l) (\sigma_k^* - \sigma_l^*)^2 - \sigma_k^* \sigma_l^* (\sigma_k^* - \sigma_l^* + \varepsilon_k - \varepsilon_l)^2}{(\sigma_k^* - \sigma_l^*)^2 (\widehat{\sigma}_k - \widehat{\sigma}_l)^2} \right|, \end{aligned}$$

where we introduced  $\varepsilon_k = \widehat{\sigma}_k - \sigma_k^*$  and  $\varepsilon_l = \widehat{\sigma}_l - \sigma_l^*$ . Then, crossing out the term  $\sigma_k^* \sigma_l^* (\sigma_k^* - \sigma_l^*)^2$  in the numerator, we obtain

$$\begin{aligned} \left| \frac{\widehat{\sigma}_k \widehat{\sigma}_l}{(\widehat{\sigma}_k - \widehat{\sigma}_l)^2} - \frac{\sigma_k^* \sigma_l^*}{(\sigma_k^* - \sigma_l^*)^2} \right| &\leq \\ &\leq \left| \frac{(\sigma_k^* \varepsilon_l + \sigma_l^* \varepsilon_k + \varepsilon_k \varepsilon_l) (\sigma_k^* - \sigma_l^*)^2 - \sigma_k^* \sigma_l^* (2(\sigma_k^* - \sigma_l^*) (\varepsilon_k - \varepsilon_l) + (\varepsilon_k - \varepsilon_l)^2)}{(\sigma_k^* - \sigma_l^*)^2 (\widehat{\sigma}_k - \widehat{\sigma}_l)^2} \right|. \end{aligned}$$

On the event where  $\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*\|_{\infty} \leq g_{\mathcal{J}}^*/4$  we have  $|\varepsilon_k|, |\varepsilon_l| \leq |\sigma_k^* - \sigma_l^*|/4$ , therefore we can omit the terms  $\varepsilon_k \varepsilon_l$  and  $(\varepsilon_k - \varepsilon_l)^2$  paying a constant factor

for that:

$$\begin{aligned}
 & \left| \frac{\widehat{\sigma}_k \widehat{\sigma}_l}{(\widehat{\sigma}_k - \widehat{\sigma}_l)^2} - \frac{\sigma_k^* \sigma_l^*}{(\sigma_k^* - \sigma_l^*)^2} \right| \\
 & \lesssim \left| \frac{(\sigma_k^* \varepsilon_l + \sigma_l^* \varepsilon_k) (\sigma_k^* - \sigma_l^*)^2 - 2 \sigma_k^* \sigma_l^* (\sigma_k^* - \sigma_l^*) (\varepsilon_k - \varepsilon_l)}{(\sigma_k^* - \sigma_l^*)^2 (\widehat{\sigma}_k - \widehat{\sigma}_l)^2} \right| \\
 & = \left| \frac{(\sigma_k^* \varepsilon_l + \sigma_l^* \varepsilon_k) (\sigma_k^* - \sigma_l^*) - 2 \sigma_k^* \sigma_l^* (\varepsilon_k - \varepsilon_l)}{(\sigma_k^* - \sigma_l^*) (\widehat{\sigma}_k - \widehat{\sigma}_l)^2} \right| \\
 & = \left| \frac{\varepsilon_k (\sigma_k^{*2} + \sigma_k^* \sigma_l^*) - \varepsilon_l (\sigma_l^{*2} + \sigma_k^* \sigma_l^*)}{(\sigma_k^* - \sigma_l^*) (\widehat{\sigma}_k - \widehat{\sigma}_l)^2} \right|.
 \end{aligned}$$

Since  $\sigma_k^* \sigma_l^* \leq (\sigma_k^{*2} + \sigma_l^{*2})/2$ , we get

$$\left| \frac{\widehat{\sigma}_k \widehat{\sigma}_l}{(\widehat{\sigma}_k - \widehat{\sigma}_l)^2} - \frac{\sigma_k^* \sigma_l^*}{(\sigma_k^* - \sigma_l^*)^2} \right| \lesssim \frac{\sigma_k^{*2} + \sigma_l^{*2}}{(\sigma_k^* - \sigma_l^*) (\widehat{\sigma}_k - \widehat{\sigma}_l)^2} \|\widehat{\Sigma} - \Sigma^*\|_\infty.$$

As to the denominator, again considering the event where  $\|\widehat{\Sigma} - \Sigma^*\|_\infty \leq g_{\mathcal{J}}^*/4$ , we have  $|\varepsilon_k|, |\varepsilon_l| \leq |\sigma_k^* - \sigma_l^*|/4$  that results in

$$|\widehat{\sigma}_k - \widehat{\sigma}_l| = |\sigma_k^* - \sigma_l^* + \varepsilon_k - \varepsilon_l| \geq |\sigma_k^* - \sigma_l^*|/2.$$

Hence,

$$\left| \frac{\widehat{\sigma}_k \widehat{\sigma}_l}{(\widehat{\sigma}_k - \widehat{\sigma}_l)^2} - \frac{\sigma_k^* \sigma_l^*}{(\sigma_k^* - \sigma_l^*)^2} \right| \lesssim \frac{\sigma_k^{*2} + \sigma_l^{*2}}{(\sigma_k^* - \sigma_l^*)^3} \|\widehat{\Sigma} - \Sigma^*\|_\infty.$$

As to the term  $\left| \left( \frac{n}{n_p} - 1 \right) \frac{\sigma_k^* \sigma_l^*}{(\sigma_k^* - \sigma_l^*)^2} \right|$ , it is simply bounded as

$$\left| \left( \frac{n}{n_p} - 1 \right) \frac{\sigma_k^* \sigma_l^*}{(\sigma_k^* - \sigma_l^*)^2} \right| \lesssim \frac{p}{n} \cdot \frac{\sigma_k^{*2} + \sigma_l^{*2}}{(\sigma_k^* - \sigma_l^*)^2} \leq \frac{p}{n} \cdot \frac{\sigma_k^{*2} + \sigma_l^{*2}}{(\sigma_k^* - \sigma_l^*)^3} \|\Sigma^*\|_\infty.$$

The last inequality uses the fact that  $|\sigma_k^* - \sigma_l^*| \leq \|\Sigma^*\|_\infty$  which is rather useless, however it allows to write the final bound in a convenient form and doesn't worsen the result.

Putting this all together, we get

$$\|\widehat{\Gamma}_{\mathcal{J}} - \Gamma_{\mathcal{J}}^*\|_1 \lesssim \sum_{k \in \mathcal{I}_{\mathcal{J}}} \sum_{l \notin \mathcal{I}_{\mathcal{J}}} \frac{\sigma_k^{*2} + \sigma_l^{*2}}{(\sigma_k^* - \sigma_l^*)^3} \left( \|\widehat{\Sigma} - \Sigma^*\|_\infty + \frac{p}{n} \|\Sigma^*\|_\infty \right),$$

which provides the desired result once we notice that  $\sum_{k \in \mathcal{I}_{\mathcal{J}}} \sum_{l \notin \mathcal{I}_{\mathcal{J}}} (\sigma_k^{*2} + \sigma_l^{*2})$  can be bounded by both  $2p \operatorname{Tr}(\Sigma^{*2})$  and  $2pm_{\mathcal{J}}^* \|\Sigma^*\|_\infty$ .  $\square$

Unfortunately, the entries  $\widehat{\xi}_{k,l}$  of  $\widehat{\xi}_{\mathcal{J}}$  are not Gaussian because of the product  $\eta_{k,j} \eta_{l,j}$ . This does not allow to apply the Gaussian comparison Lemma A.4. To get rid of this issue, we condition on  $\widehat{\mathbf{P}}_{\mathcal{J}} Z$ . Namely, in the ‘‘posterior’’ world random vectors  $\widehat{\mathbf{P}}_{\mathcal{J}} Z_j$  and  $(\mathbf{I}_p - \widehat{\mathbf{P}}_{\mathcal{J}}) Z_j$  are Gaussian and uncorrelated, therefore, independent, so we can condition on  $Z_{\mathcal{J}} \stackrel{\text{def}}{=} (\widehat{\mathbf{P}}_{\mathcal{J}} Z_1, \dots, \widehat{\mathbf{P}}_{\mathcal{J}} Z_{n_p})$  to get that  $\widehat{S}_{\mathcal{J}}$  is conditionally on  $\mathbf{X}^n, Z_{\mathcal{J}}$  Gaussian random vector with the covariance matrix

$$\widetilde{\Gamma}_{\mathcal{J}} \stackrel{\text{def}}{=} \text{Cov}(\widehat{\xi}_{\mathcal{J}} \mid \mathbf{X}^n, Z_{\mathcal{J}}).$$

It holds similarly to the above

$$\begin{aligned} \widetilde{\Gamma}_{(k,l),(k',l')} &\stackrel{\text{def}}{=} \text{Cov}(\widehat{\xi}_{k,l}, \widehat{\xi}_{k',l'} \mid \mathbf{X}^n, Z_{\mathcal{J}}) \\ &= \frac{2n}{n_p^2} \sum_{j,j'=1}^{n_p} \frac{\widehat{\sigma}_k^{1/2} \widehat{\sigma}_l^{1/2} \widehat{\sigma}_{k'}^{1/2} \widehat{\sigma}_{l'}^{1/2}}{(\widehat{\sigma}_k - \widehat{\sigma}_l)(\widehat{\sigma}_{k'} - \widehat{\sigma}_{l'})} \mathbb{E}(\eta_{k,j} \eta_{l,j} \eta_{k',j'} \eta_{l',j'} \mid \mathbf{X}^n, Z_{\mathcal{J}}) \\ &= \frac{2n}{n_p} \widetilde{\delta}_{k,k'} \delta_{l,l'} \frac{\widehat{\sigma}_k^{1/2} \widehat{\sigma}_{k'}^{1/2} \widehat{\sigma}_l}{(\widehat{\sigma}_k - \widehat{\sigma}_l)(\widehat{\sigma}_{k'} - \widehat{\sigma}_l)} \end{aligned}$$

with

$$\widetilde{\delta}_{k,k'} \stackrel{\text{def}}{=} \frac{1}{n_p} \sum_{j=1}^{n_p} \eta_{k,j} \eta_{k',j}.$$

**Lemma 4.5.** *It holds on a random set of pseudo-posterior measure  $1 - \frac{1}{n}$*

$$\max_{k,k' \in \mathcal{I}_{\mathcal{J}}} |\widetilde{\delta}_{k,k'} - \delta_{k,k'}| \lesssim \sqrt{\frac{\log(n_p + m_{\mathcal{J}}^*)}{n_p}},$$

and on this set

$$\|\widetilde{\Gamma}_{\mathcal{J}} - \widehat{\Gamma}_{\mathcal{J}}\|_1 \lesssim \widehat{\Delta}_4 \stackrel{\text{def}}{=} \frac{(m_{\mathcal{J}}^*)^{3/2} \|\widehat{\Sigma}\|_{\infty} \text{Tr}(\widehat{\Sigma})}{\widehat{g}_{\mathcal{J}}^2} \sqrt{\frac{\log(n_p + m_{\mathcal{J}}^*)}{n_p}}. \quad (4.4)$$

*Proof.* The first result of the lemma follows easily from usual concentration inequalities for sub-exponential random variables and union bound for at most  $|\mathcal{I}_{\mathcal{J}}|^2 = (m_{\mathcal{J}}^*)^2$  pairs of  $k, k'$ .

To obtain the second inequality we represent  $\widetilde{\Gamma}_{\mathcal{J}}$  and  $\widehat{\Gamma}_{\mathcal{J}}$  as

$$\begin{aligned} \widetilde{\Gamma}_{\mathcal{J}} &= \text{diag} \left( \widetilde{\Gamma}_{\mathcal{J}}^{(l)} \right)_{l \notin \mathcal{I}_{\mathcal{J}}}, \\ \widehat{\Gamma}_{\mathcal{J}} &= \text{diag} \left( \widehat{\Gamma}_{\mathcal{J}}^{(l)} \right)_{l \notin \mathcal{I}_{\mathcal{J}}}. \end{aligned}$$

Due to this block structure we have

$$\|\tilde{\Gamma}_{\mathcal{J}} - \hat{\Gamma}_{\mathcal{J}}\|_1 = \sum_{l \notin \mathcal{I}_{\mathcal{J}}} \|\tilde{\Gamma}_{\mathcal{J}}^{(l)} - \hat{\Gamma}_{\mathcal{J}}^{(l)}\|_1.$$

Let us fix  $l \notin \mathcal{I}_{\mathcal{J}}$  and focus on the corresponding block with size  $m_{\mathcal{J}}^* \times m_{\mathcal{J}}^*$ . It's easy to observe that for each  $k, k' \in \mathcal{I}_{\mathcal{J}}$

$$\tilde{\Gamma}_{(k,l),(k',l)} - \hat{\Gamma}_{(k,l),(k',l)} = \frac{2n}{n_p} \frac{\hat{\sigma}_k^{1/2} \hat{\sigma}_{k'}^{1/2} \hat{\sigma}_l}{(\hat{\sigma}_k - \hat{\sigma}_l)(\hat{\sigma}_{k'} - \hat{\sigma}_l)} \cdot (\tilde{\delta}_{k,k'} - \delta_{k,k'})$$

and, therefore,

$$\max_{k, k' \in \mathcal{I}_{\mathcal{J}}} \left| \tilde{\Gamma}_{(k,l),(k',l)} - \hat{\Gamma}_{(k,l),(k',l)} \right| \leq \frac{2\|\hat{\Sigma}\|_{\infty} \hat{\sigma}_l}{\hat{g}_{\mathcal{J}}^2} \max_{k, k' \in \mathcal{I}_{\mathcal{J}}} |\tilde{\delta}_{k,k'} - \delta_{k,k'}|.$$

Finally, since

$$\begin{aligned} \|\tilde{\Gamma}_{\mathcal{J}}^{(l)} - \hat{\Gamma}_{\mathcal{J}}^{(l)}\|_1 &\leq \sqrt{m_{\mathcal{J}}^*} \|\tilde{\Gamma}_{\mathcal{J}}^{(l)} - \hat{\Gamma}_{\mathcal{J}}^{(l)}\|_2 \\ &\leq (m_{\mathcal{J}}^*)^{3/2} \max_{k, k' \in \mathcal{I}_{\mathcal{J}}} \left| \tilde{\Gamma}_{(k,l),(k',l)} - \hat{\Gamma}_{(k,l),(k',l)} \right|, \end{aligned}$$

the obtained inequalities provide the result of the lemma.  $\square$

Putting together (4.3) and (4.4) yields the bound

$$\|\tilde{\Gamma}_{\mathcal{J}} - \Gamma_{\mathcal{J}}^*\|_1 \lesssim \hat{\Delta}_3 + \hat{\Delta}_4.$$

The Gaussian comparison Lemma A.4 can be used to compare the conditional distribution of  $\|\hat{\xi}_{\mathcal{J}}\|$  given  $\mathbf{X}^n, \hat{\mathbf{P}}_{\mathcal{J}}Z$  and the unconditional distribution of  $\|\xi_{\mathcal{J}}\|$ : on a random set of pseudo-posterior measure  $1 - \frac{1}{n}$

$$\begin{aligned} &\sup_{x \in \mathbb{R}} \left| \mathbb{P} \left( \|\hat{\xi}_{\mathcal{J}}\|^2 \leq x \mid \mathbf{X}^n, Z_{\mathcal{J}} \right) - \mathbb{P} \left( \|\xi_{\mathcal{J}}\|^2 \leq x \right) \right| \\ &\lesssim \frac{\hat{\Delta}_3 + \hat{\Delta}_4}{\|\Gamma_{\mathcal{J}}^*\|_2^{1/2} (\|\Gamma_{\mathcal{J}}^*\|_2^2 - \|\Gamma_{\mathcal{J}}^*\|_{\infty}^2)^{1/4}}. \end{aligned}$$

Of course, integrating w.r.t.  $\hat{\mathbf{P}}_{\mathcal{J}}Z$  ensures similar result when conditioning on the data  $\mathbf{X}^n$  only:

$$\begin{aligned} &\sup_{x \in \mathbb{R}} \left| \mathbb{P} \left( \|\hat{\xi}_{\mathcal{J}}\|^2 \leq x \mid \mathbf{X}^n \right) - \mathbb{P} \left( \|\xi_{\mathcal{J}}\|^2 \leq x \right) \right| \\ &\lesssim \frac{\hat{\Delta}_3 + \hat{\Delta}_4}{\|\Gamma_{\mathcal{J}}^*\|_2^{1/2} (\|\Gamma_{\mathcal{J}}^*\|_2^2 - \|\Gamma_{\mathcal{J}}^*\|_{\infty}^2)^{1/4}} + \frac{1}{n} \end{aligned} \tag{4.5}$$

with probability one.

**Step 3** So far our bounds  $\widehat{\Delta}_2, \widehat{\Delta}_3, \widehat{\Delta}_4$  were obtained in the “posterior” world, and they are random in the  $\mathbf{X}^n$ -world since they depend on  $\widehat{\Sigma}$ . We want to bound them by deterministic counterparts  $\Delta_1, \Delta_2, \Delta_3$  with probability  $1 - 1/n$ . To do so, we basically need to upperbound  $\|\widehat{\Sigma}\|_\infty$ ,  $\text{Tr}(\widehat{\Sigma})$ ,  $\widehat{l}_{\mathcal{J}}$  and to lowerbound  $\widehat{g}_{\mathcal{J}}$ . This can be done as follows:

$$\begin{aligned} \|\widehat{\Sigma}\|_\infty &\leq \|\Sigma^*\|_\infty + \|\widehat{\Sigma} - \Sigma^*\|_\infty \leq \|\Sigma^*\|_\infty(1 + \widehat{\delta}_n), \\ \text{Tr}(\widehat{\Sigma}) &\leq \text{Tr}(\Sigma^*) + p\|\widehat{\Sigma} - \Sigma^*\|_\infty \leq \text{Tr}(\Sigma^*) \left(1 + \frac{p\widehat{\delta}_n \|\Sigma^*\|_\infty}{\text{Tr}(\Sigma^*)}\right). \end{aligned}$$

with probability  $1 - 1/n$ . Due to the definition of the spectral gap, there is an index  $j$  such that  $\widehat{g}_{\mathcal{J}} = \widehat{\sigma}_j - \widehat{\sigma}_{j+1}$ . Then due to the Weyl’s inequality we have

$$\begin{aligned} \widehat{g}_{\mathcal{J}} &= \widehat{\sigma}_j - \widehat{\sigma}_{j+1} \geq \sigma_j^* - \sigma_{j+1}^* - |\widehat{\sigma}_j - \sigma_j^*| - |\widehat{\sigma}_{j+1} - \sigma_{j+1}^*| \\ &\geq g_{\mathcal{J}}^* - 2\|\widehat{\Sigma} - \Sigma^*\|_\infty \geq g_{\mathcal{J}}^* - 2\widehat{\delta}_n \|\Sigma^*\|_\infty \end{aligned}$$

with probability  $1 - 1/n$ . Similarly we can upperbound  $\widehat{l}_{\mathcal{J}}$ . Further we can plug the obtained inequalities directly in  $\widehat{\Delta}_2, \widehat{\Delta}_3, \widehat{\Delta}_4$  in order to get  $\Delta_1, \Delta_2, \Delta_3$  without any assumptions on  $\widehat{\delta}_n$ . Another option is to use our assumption

$$\widehat{\delta}_n \leq \frac{g_{\mathcal{J}}^*}{4\|\Sigma^*\|_\infty} \wedge \frac{r(\Sigma^*)}{p},$$

that ensures  $\|\widehat{\Sigma}\|_\infty \lesssim \|\Sigma^*\|_\infty$ ,  $\text{Tr}(\widehat{\Sigma}) \lesssim \text{Tr}(\Sigma^*)$ ,  $\widehat{g}_{\mathcal{J}} \gtrsim g_{\mathcal{J}}^*$ ,  $\widehat{l}_{\mathcal{J}} \lesssim l_{\mathcal{J}}^*$ . This allows to obtain more transparent bounds on  $\Delta_1, \Delta_2, \Delta_3$ . Also note that this assumption guarantees that the event from Lemma 4.4 is of probability at least  $1 - 1/n$ . We conclude that

$$\begin{aligned} \widehat{\Delta}_2 &\lesssim \Delta_2 \stackrel{\text{def}}{=} \left\{ (\log(n) + p) \left( \left(1 + \frac{l_{\mathcal{J}}^*}{g_{\mathcal{J}}^*}\right) \frac{\sqrt{m_{\mathcal{J}}^*} \|\Sigma^*\|_\infty}{g_{\mathcal{J}}^*} + 1 \right) \|\Sigma^*\|_\infty + \|\mathbf{G}\|_\infty \right\} \times \\ &\quad \times \frac{m_{\mathcal{J}}^* \|\Sigma^*\|_\infty}{g_{\mathcal{J}}^{*2}} \sqrt{\frac{\log(n) + p}{n}}, \\ \widehat{\Delta}_3 &\lesssim \Delta_3 \stackrel{\text{def}}{=} \frac{\|\Sigma^*\|_\infty \left( m_{\mathcal{J}}^* \|\Sigma^*\|_\infty^2 \wedge \text{Tr}(\Sigma^{*2}) \right)}{g_{\mathcal{J}}^{*3}} p \left( \widehat{\delta}_n + \frac{p}{n} \right), \\ \widehat{\Delta}_4 &\lesssim \Delta_4 \stackrel{\text{def}}{=} \frac{(m_{\mathcal{J}}^*)^{3/2} \|\Sigma^*\|_\infty \text{Tr}(\Sigma^*)}{g_{\mathcal{J}}^{*2}} \sqrt{\frac{\log(n)}{n}} \end{aligned}$$

with probability  $1 - 1/n$  in the  $\mathbf{X}^n$ -world.

Now we combine the obtained bounds. For  $\Delta_2$  defined above and arbitrary  $x \in \mathbb{R}$  it holds

$$\begin{aligned} & \Pi \left( n \|\mathbf{P}_{\mathcal{J}} - \widehat{\mathbf{P}}_{\mathcal{J}}\|_2^2 \leq x \mid \mathbf{X}^n \right) \\ & \leq \Pi \left( n \|\widehat{\mathcal{S}}_{\mathcal{J}}\|_2^2 \leq x + \Delta_2 \mid \mathbf{X}^n \right) \\ & \quad + \Pi \left( n \|\mathbf{P}_{\mathcal{J}} - \widehat{\mathbf{P}}_{\mathcal{J}}\|_2^2 - n \|\widehat{\mathcal{S}}_{\mathcal{J}}\|_2^2 \leq -\Delta_2 \mid \mathbf{X}^n \right). \end{aligned}$$

Since  $n \|\widehat{\mathcal{S}}_{\mathcal{J}}\|_2^2 \mid \mathbf{X}^n \stackrel{d}{=} \|\widehat{\xi}_{\mathcal{J}}\|^2 \mid \mathbf{X}^n$ ,  $\widehat{\Delta}_2 \lesssim \Delta_2$  with probability  $1 - \frac{1}{n}$ , and taking (4.2) into account, we deduce

$$\Pi \left( n \|\mathbf{P}_{\mathcal{J}} - \widehat{\mathbf{P}}_{\mathcal{J}}\|_2^2 \leq x \mid \mathbf{X}^n \right) \leq \Pi \left( \|\xi_{\mathcal{J}}\|^2 \leq x + \Delta_2 \mid \mathbf{X}^n \right) + \Pi \left( \mathcal{Y}^c \mid \mathbf{X}^n \right)$$

with probability  $1 - \frac{1}{n}$ . Subtracting  $\mathbb{P} \left( \|\xi_{\mathcal{J}}\|^2 \leq x \right)$  and taking supremum of both sides, we get

$$\begin{aligned} & \sup_{x \in \mathbb{R}} \left\{ \Pi \left( n \|\mathbf{P}_{\mathcal{J}} - \widehat{\mathbf{P}}_{\mathcal{J}}\|_2^2 \leq x \mid \mathbf{X}^n \right) - \mathbb{P} \left( \|\xi_{\mathcal{J}}\|^2 \leq x \right) \right\} \\ & \leq \sup_{x \in \mathbb{R}} \left\{ \Pi \left( \|\widehat{\xi}_{\mathcal{J}}\|^2 \leq x + \Delta_2 \mid \mathbf{X}^n \right) - \mathbb{P} \left( \|\xi_{\mathcal{J}}\|^2 \leq x + \Delta_2 \right) \right\} \\ & \quad + \sup_{x \in \mathbb{R}} \left\{ \mathbb{P} \left( \|\xi_{\mathcal{J}}\|^2 \leq x + \Delta_2 \right) - \mathbb{P} \left( \|\xi_{\mathcal{J}}\|^2 \leq x \right) \right\} + \Pi \left( \mathcal{Y}^c \mid \mathbf{X}^n \right). \end{aligned}$$

The first term in the right-hand side is bounded by  $\frac{\Delta_3 + \Delta_4}{\|\Gamma_{\mathcal{J}}^*\|_2^{1/2} (\|\Gamma_{\mathcal{J}}^*\|_2^2 - \|\Gamma_{\mathcal{J}}^*\|_{\infty}^2)^{1/4}} + \frac{1}{n}$  with probability  $1 - \frac{1}{n}$  due to (4.5). The second term does not exceed  $\frac{\Delta_2}{\|\Gamma_{\mathcal{J}}^*\|_2^{1/2} (\|\Gamma_{\mathcal{J}}^*\|_2^2 - \|\Gamma_{\mathcal{J}}^*\|_{\infty}^2)^{1/4}}$  according to the Gaussian anti-concentration Lemma A.3. The last term is at most  $\frac{1}{n}$  by definition of  $\mathcal{Y}$ . Therefore,

$$\begin{aligned} & \sup_{x \in \mathbb{R}} \left\{ \Pi \left( n \|\mathbf{P}_{\mathcal{J}} - \widehat{\mathbf{P}}_{\mathcal{J}}\|_2^2 \leq x \mid \mathbf{X}^n \right) - \mathbb{P} \left( \|\xi_{\mathcal{J}}\|^2 \leq x \right) \right\} \\ & \lesssim \frac{\Delta_2 + \Delta_3 + \Delta_4}{\|\Gamma_{\mathcal{J}}^*\|_2^{1/2} (\|\Gamma_{\mathcal{J}}^*\|_2^2 - \|\Gamma_{\mathcal{J}}^*\|_{\infty}^2)^{1/4}} + \frac{1}{n} \end{aligned}$$

with probability  $1 - \frac{1}{n}$ . Similarly, one derives

$$\begin{aligned} & \sup_{x \in \mathbb{R}} \left\{ \mathbb{P} \left( \|\xi_{\mathcal{J}}\|^2 \leq x \right) - \Pi \left( n \|\mathbf{P}_{\mathcal{J}} - \widehat{\mathbf{P}}_{\mathcal{J}}\|_2^2 \leq x \mid \mathbf{X}^n \right) \right\} \\ & \lesssim \frac{\Delta_2 + \Delta_3 + \Delta_4}{\|\Gamma_{\mathcal{J}}^*\|_2^{1/2} (\|\Gamma_{\mathcal{J}}^*\|_2^2 - \|\Gamma_{\mathcal{J}}^*\|_{\infty}^2)^{1/4}} + \frac{1}{n} \end{aligned}$$

with probability  $1 - \frac{1}{n}$ . The previous two inequalities yield the desired result.

#### 4.2. Proof of Corollary 2.3

Let  $\xi_{\mathcal{J}} \sim \mathcal{N}(0, \Gamma^*)$ . Due to Theorem 2.2 we have

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P} \left( n \|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2 > x \right) - \mathbb{P}(\|\xi_{\mathcal{J}}\|^2 > x) \right| \lesssim \bar{\diamond}.$$

Fix arbitrary significance level  $\alpha \in (0; 1)$  (or confidence level  $1 - \alpha$ ). Recall that by  $\gamma_{\alpha}$  we denote  $\alpha$ -quantile of  $n\|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2$ . Let us fix an event  $\Theta$  such that

$$\sup_{x \in \mathbb{R}} \left| \Pi \left( n \|\mathbf{P}_{\mathcal{J}} - \widehat{\mathbf{P}}_{\mathcal{J}}\|_2^2 > x \mid \mathbf{X}^n \right) - \mathbb{P}(\|\xi_{\mathcal{J}}\|^2 > x) \right| \lesssim \diamond.$$

According to Theorem 2.1 its probability is at least  $1 - 1/n$ . Hence, by the triangle inequality it holds on  $\Theta$

$$\sup_{x \in \mathbb{R}} \left| \Pi \left( n \|\mathbf{P}_{\mathcal{J}} - \widehat{\mathbf{P}}_{\mathcal{J}}\|_2^2 > x \mid \mathbf{X}^n \right) - \mathbb{P} \left( n \|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2 > x \right) \right| \leq \diamond' \asymp \bar{\diamond} + \diamond.$$

Therefore, taking  $x = \gamma_{\alpha - \diamond'}$  and  $x = \gamma_{\alpha + \diamond'}$ , we get on  $\Theta$

$$\begin{aligned} \left| \Pi \left( n \|\mathbf{P}_{\mathcal{J}} - \widehat{\mathbf{P}}_{\mathcal{J}}\|_2^2 > \gamma_{\alpha - \diamond'} \mid \mathbf{X}^n \right) - (\alpha - \diamond') \right| &\leq \diamond', \\ \left| \Pi \left( n \|\mathbf{P}_{\mathcal{J}} - \widehat{\mathbf{P}}_{\mathcal{J}}\|_2^2 > \gamma_{\alpha + \diamond'} \mid \mathbf{X}^n \right) - (\alpha + \diamond') \right| &\leq \diamond'. \end{aligned}$$

Thus,

$$\begin{aligned} \Pi \left( n \|\mathbf{P}_{\mathcal{J}} - \widehat{\mathbf{P}}_{\mathcal{J}}\|_2^2 > \gamma_{\alpha - \diamond'} \mid \mathbf{X}^n \right) &\leq (\alpha - \diamond') + \diamond' = \alpha, \\ \Pi \left( n \|\mathbf{P}_{\mathcal{J}} - \widehat{\mathbf{P}}_{\mathcal{J}}\|_2^2 > \gamma_{\alpha + \diamond'} \mid \mathbf{X}^n \right) &\geq (\alpha + \diamond') - \diamond' = \alpha. \end{aligned}$$

By definition of  $\gamma_{\alpha}^{\circ}$  the previous two inequalities yield

$$\gamma_{\alpha + \diamond'} \leq \gamma_{\alpha}^{\circ} \leq \gamma_{\alpha - \diamond'} \quad \text{on } \Theta.$$

Hence,

$$\begin{aligned} \mathbb{P}(\gamma_{\alpha}^{\circ} < \gamma_{\alpha + \diamond'}) &\leq \mathbb{P}(\Theta^c) \leq \frac{1}{n}, \\ \mathbb{P}(\gamma_{\alpha}^{\circ} > \gamma_{\alpha - \diamond'}) &\leq \mathbb{P}(\Theta^c) \leq \frac{1}{n}. \end{aligned}$$

Now we can write the following chain of inequalities:

$$\begin{aligned} & \mathbb{P}\left(n\|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2 > \gamma_{\alpha}^{\circ}\right) \\ & \leq \mathbb{P}\left(\left\{n\|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2 > \gamma_{\alpha+\diamond'}\right\} \cup \{\gamma_{\alpha}^{\circ} < \gamma_{\alpha+\diamond'}\}\right) \\ & \leq \mathbb{P}\left(n\|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2 > \gamma_{\alpha+\diamond'}\right) + \mathbb{P}\left(\gamma_{\alpha}^{\circ} < \gamma_{\alpha+\diamond'}\right) \leq \alpha + \diamond' + \frac{1}{n} \end{aligned}$$

and

$$\begin{aligned} & \mathbb{P}\left(n\|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2 > \gamma_{\alpha}^{\circ}\right) = 1 - \mathbb{P}\left(n\|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2 \leq \gamma_{\alpha}^{\circ}\right) \\ & \geq 1 - \mathbb{P}\left(\left\{n\|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2 \leq \gamma_{\alpha-\diamond'}\right\} \cup \{\gamma_{\alpha}^{\circ} > \gamma_{\alpha-\diamond'}\}\right) \\ & \geq 1 - \mathbb{P}\left(n\|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2 \leq \gamma_{\alpha-\diamond'}\right) - \mathbb{P}\left(\gamma_{\alpha}^{\circ} > \gamma_{\alpha-\diamond'}\right) \\ & = \mathbb{P}\left(n\|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2 > \gamma_{\alpha-\diamond'}\right) - \mathbb{P}\left(\gamma_{\alpha}^{\circ} > \gamma_{\alpha-\diamond'}\right) \geq \alpha - \diamond' - \frac{1}{n}. \end{aligned}$$

Finally, these inequalities imply the following bound

$$\left|\alpha - \mathbb{P}\left(n\|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2 > \gamma_{\alpha}^{\circ}\right)\right| \leq \diamond' + \frac{1}{n},$$

which concludes the proof.

### Acknowledgements

We thank an anonymous Referee for very valuable comments and suggestions which led to a significant improvement of the paper.

This work has been funded by the Russian Academic Excellence Project ‘5-100’. Results of Section 2 has been obtained under support of the RSF grant No. 18-11-00132. Financial support by the German Research Foundation (DFG) through the Collaborative Research Center 1294 is gratefully acknowledged.

### Appendix A: Auxiliary results

Here we formulate some well-known results that were used throughout the paper.

The following theorem gathers several crucial results on concentration of sample covariance.

**Theorem A.1.** *Let  $X_1, \dots, X_n$  be i.i.d. zero-mean random vectors in  $\mathbb{R}^p$ . Denote the true covariance matrix as  $\Sigma^* \stackrel{\text{def}}{=} \mathbb{E}(X_i X_i^\top)$  and the sample covariance as  $\widehat{\Sigma} \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^n X_i X_i^\top$ . Suppose the data are obtained from:*

(i) Gaussian distribution  $\mathcal{N}(0, \Sigma^*)$ . In this case, define  $\widehat{\delta}_n$  as

$$\widehat{\delta}_n \asymp \sqrt{\frac{r(\Sigma^*) + \log(n)}{n}};$$

(ii) Sub-Gaussian distribution. In this case, define  $\widehat{\delta}_n$  as

$$\widehat{\delta}_n \asymp \sqrt{\frac{p + \log(n)}{n}};$$

(iii) a distribution supported in some centered Euclidean ball of radius  $R$ . In this case, define  $\widehat{\delta}_n$  as

$$\widehat{\delta}_n \asymp \frac{R}{\sqrt{\|\Sigma^*\|}} \sqrt{\frac{\log(n)}{n}};$$

(iv) log-concave probability measure. In this case, define  $\widehat{\delta}_n$  as

$$\widehat{\delta}_n \asymp \sqrt{\frac{\log^6(n)}{np}}.$$

Then in all the cases above the following concentration result for  $\widehat{\Sigma}$  holds with the corresponding  $\widehat{\delta}_n$ :

$$\|\widehat{\Sigma} - \Sigma^*\|_\infty \leq \widehat{\delta}_n \|\Sigma^*\|_\infty$$

with probability at least  $1 - \frac{1}{n}$ .

*Proof.* (i) See [22], Corollary 2. (ii) This is a well-known simple result presented in a range of papers and lecture notes. See, e.g. [29], Theorem 4.6. (iii) See [33], Corollary 5.52. Usually the radius  $R$  is taken such that  $\frac{R}{\sqrt{\|\Sigma^*\|}} \asymp \frac{\sqrt{\text{Tr}(\Sigma^*)}}{\sqrt{\|\Sigma^*\|}} = \sqrt{r(\Sigma^*)}$ . (iv) See [1], Theorem 4.1.  $\square$

The following lemma is a crucial tool when working with spectral projectors.

**Lemma A.2.** *The following bound holds for all  $\mathcal{J} = \{r^-, r^- + 1, \dots, r^+\}$  with  $1 \leq r^- \leq r^+ \leq q$ :*

$$\|\widetilde{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_\infty \leq 4 \left(1 + \frac{2 l_{\mathcal{J}}^*}{\pi g_{\mathcal{J}}^*}\right) \frac{\|\widetilde{\Sigma} - \Sigma^*\|_\infty}{g_{\mathcal{J}}^*}.$$

Moreover, the following representation holds:

$$\widetilde{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^* = L_{\mathcal{J}}(\widetilde{\Sigma} - \Sigma^*) + R_{\mathcal{J}}(\widetilde{\Sigma} - \Sigma^*), \quad (\text{A.1})$$

where

$$L_{\mathcal{J}}(\tilde{\Sigma} - \Sigma^*) \stackrel{\text{def}}{=} \sum_{r \in \mathcal{J}} \sum_{s \notin \mathcal{J}} \frac{\mathbf{P}_r^*(\tilde{\Sigma} - \Sigma^*)\mathbf{P}_s^* + \mathbf{P}_s^*(\tilde{\Sigma} - \Sigma^*)\mathbf{P}_r^*}{\mu_r^* - \mu_s^*}$$

and

$$\|R_{\mathcal{J}}(\tilde{\Sigma} - \Sigma^*)\|_{\infty} \leq 15 \left(1 + \frac{2 l_{\mathcal{J}}^*}{\pi g_{\mathcal{J}}^*}\right) \left(\frac{\|\tilde{\Sigma} - \Sigma^*\|_{\infty}}{g_{\mathcal{J}}^*}\right)^2. \quad (\text{A.2})$$

*Proof.* Apply Lemma 2 from [21]. □

This lemma shows that  $\tilde{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*$  can be approximated by the linear operator  $L_{\mathcal{J}}(\tilde{\Sigma} - \Sigma^*)$ .

The next lemma from [16] provides upper bound for  $\Delta$ -band of the squared norm of a Gaussian element.

**Lemma A.3** (Gaussian anti-concentration). *Let  $\xi$  be a Gaussian element in Hilbert space  $\mathbb{H}$  with zero mean and covariance operator  $\Sigma_{\xi}$ . Then for arbitrary  $\Delta > 0$  one has*

$$\mathbb{P}(x < \|\xi\|^2 < x + \Delta) \leq \frac{\Delta}{\|\Sigma_{\xi}\|_2^{1/2} (\|\Sigma_{\xi}\|_2^2 - \|\Sigma_{\xi}\|_{\infty}^2)^{1/4}}.$$

*Proof.* See [16], Theorem 2.7. □

One more lemma from [16] describes how close are the distributions of the norms of two Gaussian elements in terms of their covariance operators. Note that the bound is dimension free.

**Lemma A.4** (Gaussian comparison). *Let  $\xi$  and  $\eta$  be Gaussian elements in Hilbert space  $\mathbb{H}$  with zero mean and covariance operators  $\Sigma_{\xi}$  and  $\Sigma_{\eta}$ , respectively. The following inequality holds*

$$\sup_{x \in \mathbb{R}} |\mathbb{P}(\|\xi\|^2 \geq x) - \mathbb{P}(\|\eta\|^2 \geq x)| \lesssim \|\Sigma_{\xi} - \Sigma_{\eta}\|_1 \times \left( \frac{1}{\|\Sigma_{\xi}\|_2^{1/2} (\|\Sigma_{\xi}\|_2^2 - \|\Sigma_{\xi}\|_{\infty}^2)^{1/4}} + \frac{1}{\|\Sigma_{\eta}\|_2^{1/2} (\|\Sigma_{\eta}\|_2^2 - \|\Sigma_{\eta}\|_{\infty}^2)^{1/4}} \right).$$

*Proof.* See [16], Theorem 2.1. □

## Appendix B: Auxiliary proofs

### B.1. Proof of Theorem 2.2

The proof consists of three steps.

**Step 1** Apply the representation (A.1) from Lemma A.2 to  $\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*$ :

$$\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^* = L_{\mathcal{J}}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*) + R_{\mathcal{J}}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*).$$

Then, for  $n\|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2$  one has

$$\begin{aligned} n\|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2 &= n\|L_{\mathcal{J}}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\|_2^2 + n\|R_{\mathcal{J}}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\|_2^2 \\ &\quad + 2n\langle L_{\mathcal{J}}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*), R_{\mathcal{J}}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*) \rangle_2. \end{aligned}$$

Let us estimate how good  $n\|L_{\mathcal{J}}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\|_2^2$  approximates  $n\|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2$ : clearly, we have

$$\begin{aligned} &\left| n\|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2 - n\|L_{\mathcal{J}}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\|_2^2 \right| \\ &\leq n\|R_{\mathcal{J}}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\|_2^2 + 2n\|L_{\mathcal{J}}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\|_2 \|R_{\mathcal{J}}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\|_2. \end{aligned}$$

Let us elaborate on the right-hand side. First, since

$$R_{\mathcal{J}}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*) = \widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^* - \sum_{r \in \mathcal{J}} \sum_{s \notin \mathcal{J}} \frac{\mathbf{P}_r^*(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\mathbf{P}_s^* + \mathbf{P}_s^*(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\mathbf{P}_r^*}{\mu_r^* - \mu_s^*}$$

and  $\widehat{\mathbf{P}}_{\mathcal{J}}, \mathbf{P}_{\mathcal{J}}^*, \sum_{r \in \mathcal{J}} \sum_{s \notin \mathcal{J}} \mathbf{P}_r^*(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\mathbf{P}_s^*$  have rank at most  $m_{\mathcal{J}}^*$ , then the rank of  $R_{\mathcal{J}}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)$  is at most  $4m_{\mathcal{J}}^*$ . Hence, due to the relation between the Frobenius and the operator norms via rank, we have

$$\|R_{\mathcal{J}}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\|_2 \leq \sqrt{4m_{\mathcal{J}}^*} \|R_{\mathcal{J}}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\|_{\infty}.$$

The bound (A.2) from Lemma A.2 gives

$$\|R_{\mathcal{J}}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\|_2 \leq \sqrt{4m_{\mathcal{J}}^*} \cdot 15 \left( 1 + \frac{2l_{\mathcal{J}}^*}{\pi g_{\mathcal{J}}^*} \right) \frac{\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*\|_{\infty}^2}{g_{\mathcal{J}}^{*2}}.$$

Now let us bound  $\|L_{\mathcal{J}}(\widehat{\Sigma} - \Sigma^*)\|_{\infty}$ :

$$\begin{aligned}
 \|L_{\mathcal{J}}(\widehat{\Sigma} - \Sigma^*)\|_{\infty} &= \left\| \sum_{r \in \mathcal{J}} \sum_{s \notin \mathcal{J}} \frac{\mathbf{P}_r^*(\widehat{\Sigma} - \Sigma^*)\mathbf{P}_s^* + \mathbf{P}_s^*(\widehat{\Sigma} - \Sigma^*)\mathbf{P}_r^*}{\mu_r^* - \mu_s^*} \right\|_{\infty} \\
 &\leq 2 \left\| \sum_{r \in \mathcal{J}} \sum_{s \notin \mathcal{J}} \frac{\mathbf{P}_r^*(\widehat{\Sigma} - \Sigma^*)\mathbf{P}_s^*}{\mu_r^* - \mu_s^*} \right\|_{\infty} = 2 \left\| \sum_{r \in \mathcal{J}} \mathbf{P}_r^* \sum_{s \notin \mathcal{J}} \frac{(\widehat{\Sigma} - \Sigma^*)\mathbf{P}_s^*}{\mu_r^* - \mu_s^*} \right\|_{\infty} \\
 &\leq 2 \sum_{r \in \mathcal{J}} \|\mathbf{P}_r^*\| \left\| \sum_{s \notin \mathcal{J}} \frac{\mathbf{P}_s^*}{\mu_r^* - \mu_s^*} \right\|_{\infty} \|\widehat{\Sigma} - \Sigma^*\|_{\infty} \\
 &\leq \frac{2|\mathcal{J}|\|\widehat{\Sigma} - \Sigma^*\|_{\infty}}{\min_{r \in \mathcal{J}, s \notin \mathcal{J}} |\mu_r^* - \mu_s^*|} \leq 2|\mathcal{J}| \frac{\|\widehat{\Sigma} - \Sigma^*\|_{\infty}}{g_{\mathcal{J}}^*}.
 \end{aligned}$$

Then, for  $\|L_{\mathcal{J}}(\widehat{\Sigma} - \Sigma^*)\|_2$  we have

$$\|L_{\mathcal{J}}(\widehat{\Sigma} - \Sigma^*)\|_2 = \sqrt{2m_{\mathcal{J}}^*} \|L_{\mathcal{J}}(\widehat{\Sigma} - \Sigma^*)\|_{\infty} \leq \sqrt{2m_{\mathcal{J}}^*} 2|\mathcal{J}| \frac{\|\widehat{\Sigma} - \Sigma^*\|_{\infty}}{g_{\mathcal{J}}^*}.$$

Putting this all together, we obtain

$$\begin{aligned}
 &\left| n\|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2 - n\|L_{\mathcal{J}}(\widehat{\Sigma} - \Sigma^*)\|_2^2 \right| \\
 &\lesssim nm_{\mathcal{J}}^* \left( 1 + \frac{l_{\mathcal{J}}^*}{g_{\mathcal{J}}^*} \right)^2 \frac{\|\widehat{\Sigma} - \Sigma^*\|_{\infty}^4}{g_{\mathcal{J}}^{*4}} + nm_{\mathcal{J}}^* |\mathcal{J}| \left( 1 + \frac{l_{\mathcal{J}}^*}{g_{\mathcal{J}}^*} \right) \frac{\|\widehat{\Sigma} - \Sigma^*\|_{\infty}^3}{g_{\mathcal{J}}^{*3}}.
 \end{aligned}$$

The concentration condition for the sample covariance (2.3) provides

$$\begin{aligned}
 &\left| n\|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2 - n\|L_{\mathcal{J}}(\widehat{\Sigma} - \Sigma^*)\|_2^2 \right| \lesssim \overline{\Delta}, \\
 &\overline{\Delta} = nm_{\mathcal{J}}^* \left( 1 + \frac{l_{\mathcal{J}}^*}{g_{\mathcal{J}}^*} \right) \left( \left( 1 + \frac{l_{\mathcal{J}}^*}{g_{\mathcal{J}}^*} \right) \frac{\widehat{\delta}_n^4}{g_{\mathcal{J}}^{*4}} \vee \frac{|\mathcal{J}| \widehat{\delta}_n^3}{g_{\mathcal{J}}^{*3}} \right) \quad (\text{B.1})
 \end{aligned}$$

with probability  $1 - \frac{1}{n}$ .

**Step 2** Following [27], we can choose  $\{u_j^*\}_{j=1}^p$  as an orthonormal basis in  $\mathbb{R}^p$  and represent  $n\|L_{\mathcal{J}}(\widehat{\Sigma} - \Sigma^*)\|_2^2$  as

$$\begin{aligned}
n\|L_{\mathcal{J}}(\widehat{\Sigma} - \Sigma^*)\|_2^2 &= n \left\| \sum_{r \in \mathcal{J}} \sum_{s \notin \mathcal{J}} \frac{P_r^*(\widehat{\Sigma} - \Sigma^*)P_s^* + P_s^*(\widehat{\Sigma} - \Sigma^*)P_r^*}{\mu_r^* - \mu_s^*} \right\|_2^2 \\
&= n \sum_{l,k=1}^p \left( u_k^{*\top} \sum_{r \in \mathcal{J}} \sum_{s \notin \mathcal{J}} \frac{P_r^*(\widehat{\Sigma} - \Sigma^*)P_s^* + P_s^*(\widehat{\Sigma} - \Sigma^*)P_r^*}{\mu_r^* - \mu_s^*} u_l^* \right)^2 \\
&= n \sum_{l,k=1}^p \left( u_k^{*\top} \sum_{r_1 \in \mathcal{J}} \sum_{s_1 \notin \mathcal{J}} \frac{P_{r_1}^*(\widehat{\Sigma} - \Sigma^*)P_{s_1}^* + P_{s_1}^*(\widehat{\Sigma} - \Sigma^*)P_{r_1}^*}{\mu_{r_1}^* - \mu_{s_1}^*} u_l^* \right) \\
&\quad \times \left( u_k^{*\top} \sum_{r_2 \in \mathcal{J}} \sum_{s_2 \notin \mathcal{J}} \frac{P_{r_2}^*(\widehat{\Sigma} - \Sigma^*)P_{s_2}^* + P_{s_2}^*(\widehat{\Sigma} - \Sigma^*)P_{r_2}^*}{\mu_{r_2}^* - \mu_{s_2}^*} u_l^* \right) \\
&= n \sum_{l,k=1}^p \sum_{\substack{r_1 \in \mathcal{J} \\ s_1 \notin \mathcal{J}}} \sum_{\substack{r_2 \in \mathcal{J} \\ s_2 \notin \mathcal{J}}} \left( u_k^{*\top} \frac{P_{r_1}^*(\widehat{\Sigma} - \Sigma^*)P_{s_1}^* + P_{s_1}^*(\widehat{\Sigma} - \Sigma^*)P_{r_1}^*}{\mu_{r_1}^* - \mu_{s_1}^*} u_l^* \right) \\
&\quad \times \left( u_k^{*\top} \frac{P_{r_2}^*(\widehat{\Sigma} - \Sigma^*)P_{s_2}^* + P_{s_2}^*(\widehat{\Sigma} - \Sigma^*)P_{r_2}^*}{\mu_{r_2}^* - \mu_{s_2}^*} u_l^* \right).
\end{aligned}$$

As we can see, the only terms that survive in this sum are the terms with  $r_1 = r_2 = r \in \mathcal{J}$ ,  $s_1 = s_2 = s \notin \mathcal{J}$ ,  $k \in \Delta_r^*$ ,  $l \in \Delta_s^*$ , and due to the symmetry the factor 2 appears. So, we derive

$$\begin{aligned}
n\|L_{\mathcal{J}}(\widehat{\Sigma} - \Sigma^*)\|_2^2 &= 2n \sum_{\substack{k \in \Delta_r^* \\ r \in \mathcal{J}}} \sum_{\substack{l \in \Delta_s^* \\ s \notin \mathcal{J}}} \left( u_k^{*\top} \frac{P_r^*(\widehat{\Sigma} - \Sigma^*)P_s^*}{\mu_r^* - \mu_s^*} u_l^* \right)^2 \\
&= 2n \sum_{\substack{k \in \Delta_r^* \\ r \in \mathcal{J}}} \sum_{\substack{l \in \Delta_s^* \\ s \notin \mathcal{J}}} \left( \frac{u_k^{*\top} (\widehat{\Sigma} - \Sigma^*) u_l^*}{\mu_r^* - \mu_s^*} \right)^2.
\end{aligned}$$

Now let us define for all  $k \in \mathcal{I}_{\mathcal{J}}$  and  $l \notin \mathcal{I}_{\mathcal{J}}$

$$S_{\mathcal{J}}(u_k^*, u_l^*) = \sqrt{2n} \frac{u_k^{*\top} (\widehat{\Sigma} - \Sigma^*) u_l^*}{\mu_r^* - \mu_s^*}.$$

This set of quantities can be considered as matrix

$$\{S_{\mathcal{J}}(u_k^*, u_l^*)\}_{\substack{k \in \mathcal{I}_{\mathcal{J}} \\ l \notin \mathcal{I}_{\mathcal{J}}}} \in \mathbb{R}^{m_{\mathcal{J}}^* \times (p - m_{\mathcal{J}}^*)},$$

or, we can arrange a vector  $S_{\mathcal{J}} \in \mathbb{R}^{m_{\mathcal{J}}^*(p-m_{\mathcal{J}}^*)}$  with components  $S_{\mathcal{J}}(u_k^*, u_l^*)$  ordered in some particular way. Let us notice that

$$n\|L_{\mathcal{J}}(\widehat{\Sigma} - \Sigma^*)\|_2^2 = \|S_{\mathcal{J}}\|^2.$$

**Step 3** Now our goal is to show that  $S_{\mathcal{J}}$  is approximately  $\mathcal{N}(0, \Gamma_{\mathcal{J}}^*)$  using a version of Berry-Esseen theorem given by [2]. Represent  $S_{\mathcal{J}}$  as

$$S_{\mathcal{J}} = \frac{1}{\sqrt{n}} \sum_{j=1}^n S^{(j)},$$

where  $S^{(j)}$  is a random vector with components

$$S^{(j)}(u_k^*, u_l^*) = \frac{\sqrt{2}}{\mu_r^* - \mu_s^*} (u_k^{*\top} X_j) \cdot (u_l^{*\top} X_j)$$

for all  $k \in \mathcal{I}_{\mathcal{J}}$  and  $l \notin \mathcal{I}_{\mathcal{J}}$ .

It is straightforward to verify that the covariance matrix of  $S^{(j)}$  (and hence of  $S_{\mathcal{J}}$ ) is  $\Gamma_{\mathcal{J}}^*$  from (2.2) under the condition that  $\mathbf{P}_{\mathcal{J}}^* X_j$  and  $(\mathbf{I}_p - \mathbf{P}_{\mathcal{J}}^*) X_j$  are independent. Consider an entry of the covariance matrix of  $S^{(j)}$  indexed by  $(k, l)$  and  $(k', l')$ , where  $k \in \Delta_r^*, k' \in \Delta_{r'}^*, r, r' \in \mathcal{J}$  and  $l \in \Delta_s^*, l' \in \Delta_{s'}^*, s, s' \notin \mathcal{J}$ :

$$\begin{aligned} \text{Cov} \left( S^{(j)} \right)_{\substack{(k,l) \\ (k',l')}} &= \mathbb{E} \left( S^{(j)}(u_k^*, u_l^*) \cdot S^{(j)}(u_{k'}^*, u_{l'}^*) \right) \\ &= \frac{2 \mathbb{E} \left[ (u_k^{*\top} X_j) \cdot (u_l^{*\top} X_j) \cdot (u_{k'}^{*\top} X_j) \cdot (u_{l'}^{*\top} X_j) \right]}{(\mu_r^* - \mu_s^*)(\mu_{r'}^* - \mu_{s'}^*)}. \end{aligned}$$

Now, the independence of  $\mathbf{P}_{\mathcal{J}}^* X_j$  and  $(\mathbf{I}_p - \mathbf{P}_{\mathcal{J}}^*) X_j$  implies the independence of  $(u_k^*, u_{k'}^*)^{\top} \mathbf{P}_{\mathcal{J}}^* X_j$  and  $(u_l^*, u_{l'}^*)^{\top} (\mathbf{I}_p - \mathbf{P}_{\mathcal{J}}^*) X_j$ , which can be rewritten as independence of  $(u_k^{*\top} X_j, u_{k'}^{*\top} X_j)^{\top}$  and  $(u_l^{*\top} X_j, u_{l'}^{*\top} X_j)^{\top}$ . This means that the expectation in the expression for the covariance entry can be splitted as

$$\text{Cov} \left( S^{(j)} \right)_{\substack{(k,l) \\ (k',l')}} = \frac{2 \mathbb{E} \left[ (u_k^{*\top} X_j)(u_{k'}^{*\top} X_j) \right] \cdot \mathbb{E} \left[ (u_l^{*\top} X_j)(u_{l'}^{*\top} X_j) \right]}{(\mu_r^* - \mu_s^*)(\mu_{r'}^* - \mu_{s'}^*)}.$$

The observation that  $u_k^{*\top} \Sigma^* u_{k'}^* = \mu_r^* \cdot \mathbb{1}\{k = k'\}$  and  $u_l^{*\top} \Sigma^* u_{l'}^* = \mu_s^* \cdot \mathbb{1}\{l = l'\}$  establishes the fact that  $\text{Cov}(S^{(j)}) = \Gamma_{\mathcal{J}}^*$ .

To apply Theorem 1.1 from [2], we need to bound  $\mathbb{E}\|\Gamma_{\mathcal{J}}^{*-1/2} S^{(j)}\|^3$ . First, let us notice that

$$\left[ \Gamma_{\mathcal{J}}^{*-1/2} S^{(j)} \right] (u_k^*, u_l^*) = \frac{u_k^{*\top} X_j}{\sqrt{\mu_r^*}} \cdot \frac{u_l^{*\top} X_j}{\sqrt{\mu_s^*}}.$$

Further, recalling the auxiliary matrices

$$\mathbf{U}_{\mathcal{J}}^* \stackrel{\text{def}}{=} \left\{ \frac{u_k^{*\top}}{\sqrt{\mu_r^*}} \right\}_{k \in \mathcal{I}_{\mathcal{J}}} \in \mathbb{R}^{m_{\mathcal{J}}^* \times p},$$

$$\mathbf{V}_{\mathcal{J}}^* \stackrel{\text{def}}{=} \left\{ \frac{u_l^{*\top}}{\sqrt{\mu_s^*}} \right\}_{l \notin \mathcal{I}_{\mathcal{J}}} \in \mathbb{R}^{(p-m_{\mathcal{J}}^*) \times p},$$

we have

$$\begin{aligned} \|\Gamma_{\mathcal{J}}^{*-1/2} S^{(j)}\|^2 &= \sum_{k \in \mathcal{I}_{\mathcal{J}}} \sum_{l \notin \mathcal{I}_{\mathcal{J}}} \frac{(u_k^{*\top} X_j)^2}{\mu_r^*} \cdot \frac{(u_l^{*\top} X_j)^2}{\mu_s^*} \\ &= \left\{ \sum_{k \in \mathcal{I}_{\mathcal{J}}} \frac{(u_k^{*\top} X_j)^2}{\mu_r^*} \right\} \cdot \left\{ \sum_{l \notin \mathcal{I}_{\mathcal{J}}} \frac{(u_l^{*\top} X_j)^2}{\mu_s^*} \right\} = \|\mathbf{U}_{\mathcal{J}}^* X_j\|^2 \|\mathbf{V}_{\mathcal{J}}^* X_j\|^2. \end{aligned}$$

Then,

$$\mathbb{E} \|\Gamma_{\mathcal{J}}^{*-1/2} S^{(j)}\|^3 = \mathbb{E} (\|\mathbf{U}_{\mathcal{J}}^* X_j\|^3 \|\mathbf{V}_{\mathcal{J}}^* X_j\|^3) = \mathbb{E} \|\mathbf{U}_{\mathcal{J}}^* X_j\|^3 \cdot \mathbb{E} \|\mathbf{V}_{\mathcal{J}}^* X_j\|^3,$$

where we again used the fact that the independence of  $\mathbf{P}_{\mathcal{J}}^* X_j$  and  $(\mathbf{I}_p - \mathbf{P}_{\mathcal{J}}^*) X_j$  implies the independence of  $\mathbf{U}_{\mathcal{J}}^* \mathbf{P}_{\mathcal{J}}^* X_j = \mathbf{U}_{\mathcal{J}}^* X_j$  and  $\mathbf{V}_{\mathcal{J}}^* (\mathbf{I}_p - \mathbf{P}_{\mathcal{J}}^*) X_j = \mathbf{V}_{\mathcal{J}}^* X_j$ .

Therefore, Theorem 1.1 from [2] yields

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P} (\|S_{\mathcal{J}}\|^2 \leq x) - \mathbb{P} (\|\xi\|^2 \leq x) \right| \lesssim \mathbb{E} \|\mathbf{U}_{\mathcal{J}}^* X\|^3 \cdot \mathbb{E} \|\mathbf{V}_{\mathcal{J}}^* X\|^3 \cdot \frac{p^{1/4}}{\sqrt{n}},$$

or, recalling that  $\|S_{\mathcal{J}}\|^2 = n \|L_{\mathcal{J}}(\widehat{\Sigma} - \Sigma^*)\|_2^2$ ,

$$\begin{aligned} \sup_{x \in \mathbb{R}} \left| \mathbb{P} \left( n \|L_{\mathcal{J}}(\widehat{\Sigma} - \Sigma^*)\|_2^2 \leq x \right) - \mathbb{P} (\|\xi\|^2 \leq x) \right| \\ \lesssim \mathbb{E} \|\mathbf{U}_{\mathcal{J}}^* X\|^3 \cdot \mathbb{E} \|\mathbf{V}_{\mathcal{J}}^* X\|^3 \cdot \frac{p^{1/4}}{\sqrt{n}}. \end{aligned}$$

**Step 4** Next, for  $\overline{\Delta}$  defined by (B.1) from Step 1 we may write for any  $x \in \mathbb{R}$

$$\begin{aligned} \mathbb{P} \left( n \|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2 \geq x \right) &\leq \mathbb{P} \left( n \|L_{\mathcal{J}}(\widehat{\Sigma} - \Sigma^*)\|_2^2 \geq x - \overline{\Delta} \right) \\ &\quad + \mathbb{P} \left( n \|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2 - n \|L_{\mathcal{J}}(\widehat{\Sigma} - \Sigma^*)\|_2^2 \geq \overline{\Delta} \right). \end{aligned}$$

Hence,

$$\begin{aligned} & \sup_{x \in \mathbb{R}} \left\{ \mathbb{P} \left( n \|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2 \geq x \right) - \mathbb{P} \left( \|\xi\|^2 \geq x \right) \right\} \\ & \leq \sup_{x \in \mathbb{R}} \left\{ \mathbb{P} \left( \|L_{\mathcal{J}}(\widehat{\Sigma} - \Sigma^*)\|^2 \geq x - \overline{\Delta} \right) - \mathbb{P} \left( \|\xi\|^2 \geq x - \overline{\Delta} \right) \right\} \\ & \quad + \sup_{x \in \mathbb{R}} \left\{ \mathbb{P} \left( \|\xi\|^2 \geq x - \overline{\Delta} \right) - \mathbb{P} \left( \|\xi\|^2 \geq x \right) \right\} \\ & \quad + \mathbb{P} \left( n \|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2 - n \|L_{\mathcal{J}}(\widehat{\Sigma} - \Sigma^*)\|_2^2 \geq \overline{\Delta} \right). \end{aligned}$$

The first term in the right-hand side was bounded in Step 3 by

$$\mathbb{E} \|\mathbf{U}_{\mathcal{J}}^* X\|^3 \cdot \mathbb{E} \|\mathbf{V}_{\mathcal{J}}^* X\|^3 \cdot \frac{p^{1/4}}{\sqrt{n}}.$$

The second term is bounded by  $\frac{\overline{\Delta}}{\|I_{\mathcal{J}}^*\|_2^{1/2} (\|I_{\mathcal{J}}^*\|_2^2 - \|I_{\mathcal{J}}^*\|_{\infty}^2)^{1/4}}$  according to the Anti-concentration Lemma A.3. The last term is less than  $1/n$  in view of (B.1) from Step 1. Therefore,

$$\begin{aligned} & \sup_{x \in \mathbb{R}} \left\{ \mathbb{P} \left( n \|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2 \geq x \right) - \mathbb{P} \left( \|\xi\|^2 \geq x \right) \right\} \\ & \lesssim \mathbb{E} \|\mathbf{U}_{\mathcal{J}}^* X\|^3 \cdot \mathbb{E} \|\mathbf{V}_{\mathcal{J}}^* X\|^3 \cdot \frac{p^{1/4}}{\sqrt{n}} + \frac{\overline{\Delta}}{\|I_{\mathcal{J}}^*\|_2^{1/2} (\|I_{\mathcal{J}}^*\|_2^2 - \|I_{\mathcal{J}}^*\|_{\infty}^2)^{1/4}} + \frac{1}{n}. \end{aligned}$$

Similarly, one can verify that

$$\begin{aligned} & \sup_{x \in \mathbb{R}} \left\{ \mathbb{P} \left( \|\xi\|^2 \geq x \right) - \mathbb{P} \left( n \|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2 \geq x \right) \right\} \\ & \lesssim \mathbb{E} \|\mathbf{U}_{\mathcal{J}}^* X\|^3 \cdot \mathbb{E} \|\mathbf{V}_{\mathcal{J}}^* X\|^3 \cdot \frac{p^{1/4}}{\sqrt{n}} + \frac{\overline{\Delta}}{\|I_{\mathcal{J}}^*\|_2^{1/2} (\|I_{\mathcal{J}}^*\|_2^2 - \|I_{\mathcal{J}}^*\|_{\infty}^2)^{1/4}} + \frac{1}{n}. \end{aligned}$$

Putting together the previous two bounds, we derive the final result:

$$\begin{aligned} & \sup_{x \in \mathbb{R}} \left| \mathbb{P} \left( n \|\widehat{\mathbf{P}}_{\mathcal{J}} - \mathbf{P}_{\mathcal{J}}^*\|_2^2 \geq x \right) - \mathbb{P} \left( \|\xi\|^2 \geq x \right) \right| \\ & \lesssim \mathbb{E} \|\mathbf{U}_{\mathcal{J}}^* X\|^3 \cdot \mathbb{E} \|\mathbf{V}_{\mathcal{J}}^* X\|^3 \cdot \frac{p^{1/4}}{\sqrt{n}} + \frac{\overline{\Delta}}{\|I_{\mathcal{J}}^*\|_2^{1/2} (\|I_{\mathcal{J}}^*\|_2^2 - \|I_{\mathcal{J}}^*\|_{\infty}^2)^{1/4}} + \frac{1}{n}. \end{aligned}$$

## References

- [1] ADAMCZAK, R., LITVAK, A. E., PAJOR, A. and TOMCZAK-JAEGERMANN, N. (2010). Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles. *J. Amer. Math. Soc.*, **23**, 535–561. [MR2601042](#)

- [2] BENTKUS, V. (2005). A Lyapunov-type bound in  $\mathbb{R}^d$ . *Theory Probab. Appl.*, **49**, 2, 311–323. [MR2144310](#)
- [3] BERTHET, Q. and RIGOLLET, P. (2013). Optimal detection of sparse principal components in high dimension. *Ann. Statist.*, **41**, 4, 1780–1815. [MR3127849](#)
- [4] BICKEL, P. J. and KLEIJN, B. J. K. (2012). The semiparametric Bernstein – von Mises theorem. *Ann. Statist.*, **40**, 206–237. [MR3013185](#)
- [5] BIRNBAUM, A., JOHNSTONE, I. M., NADLER, B. and PAUL, D. (2013). Minimax bounds for sparse PCA with noisy high-dimensional data. *Ann. Statist.*, **41**, 3, 1055–1084. [MR3113803](#)
- [6] CAI, T. T., MA, Z. and WU, Y. (2013). Sparse PCA: optimal rates and adaptive estimation. *Ann. Statist.*, **41**, 6, 3074–3110. [MR3161458](#)
- [7] CASTILLO, I. and NICKL, R. (2013). Nonparametric Bernstein – von Mises theorems in Gaussian white noise. *Ann. Statist.*, **41**, 4, 1999–2028. [MR3127856](#)
- [8] CASTILLO, I. and ROUSSEAU, J. (2015). A Bernstein – von Mises theorem for smooth functionals in semiparametric models. *Ann. Statist.*, **43**, 6, 2353–2383. [MR3405597](#)
- [9] EL KAROUI, N. (2007). Tracy – Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Ann. Probab.*, **35**, 2, 663–714. [MR2308592](#)
- [10] FAN, J., RIGOLLET, P. and WANG, W. (2015). Estimation of functionals of sparse covariance matrices. *Ann. Statist.*, **43**, 6, 2706–2737. [MR3405609](#)
- [11] FAN, J., SUN, Q., ZHOU, W-X. and ZHU, Z. (2018). Principal component analysis for big data. *arXiv:1801.01602*.
- [12] HOLTZ, M. (2010). *Sparse grid quadrature in high dimensions with applications in finance and insurance*. Lecture notes in computational science and engineering, **77**, Springer, Berlin. [MR2743492](#)
- [13] GAO, C. and ZHOU, H. H. (2015). Rate-optimal posterior contraction for sparse PCA. *Ann. Statist.*, **43**, 2, 785–818. [MR3325710](#)
- [14] GHOSH, J. K. and BASU, A. (2017) General Robust Bayes Pseudo-Posterior: Exponential Convergence results with Applications, *arXiv:1708.09692*.
- [15] GHOSH, J. K. and RAMAMOORTHY, R. V. (2003). Introduction to the non-asymptotic analysis of random matrices. In *Bayesian nonparametrics*, Springer Verlag, New York. [MR1992245](#)
- [16] GOETZE, F., NAUMOV, A., SPOKOINY, V. and ULYANOV, V. (2018). Large ball probabilities, Gaussian comparison and anti-concentration. *arXiv:1708.08663*.
- [17] GOODFELLOW, I., BENGIO, Y. and COURVILLE, A. (2016). *Deep learning*, MIT Press. [MR3617773](#)
- [18] JOHNSTONE, I. M. and LU, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *J. Amer. Statist. Assoc.*, **104**, 682–693. [MR2751448](#)
- [19] JOHNSTONE, I. M. (2010). High dimensional statistical inference and ran-

- dom matrices. In *International Congress of Mathematicians*, **I**, 307–333, Eur. Math. Soc., Zurich. [MR2334195](#)
- [20] JOHNSTONE, I. M. (2007). High dimensional Bernstein–von Mises: simple examples *Inst. Math. Stat. Collect.*, **6**, 87–98. [MR2798513](#)
- [21] KOLTCHINSKII, V. and LOUNICI, K. (2016). Asymptotics and concentration bounds for bilinear forms of spectral projectors of sample covariance. *Ann. Inst. H. Poincaré Probab. Statist.*, **52**, 4, 1976–2013. [MR3573302](#)
- [22] KOLTCHINSKII, V. and LOUNICI, K. (2017). Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli*, **23**, 1, 110–133. [MR3556768](#)
- [23] KOLTCHINSKII, V. and LOUNICI, K. (2017). New asymptotic results in Principal Component Analysis. *Sankhya A.*, **79**, 2, 254–297. [MR3707422](#)
- [24] KOLTCHINSKII, V. and LOUNICI, K. (2017). Normal approximation and concentration of spectral projectors of sample covariance. *Ann. Statist.*, **45**, 1, 121–157. [MR3611488](#)
- [25] LE CAM, L. and YANG, G. L. (1990). *Asymptotics in statistics: some basic concepts*. Springer, New York. [MR1066869](#)
- [26] MARCHENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)*, **72** (114), 4, 507–536. [MR0208649](#)
- [27] NAUMOV, A., SPOKOINY, V. and ULYANOV, V. (2017). Bootstrap confidence sets for spectral projectors of sample covariance. *arXiv:1703.00871*.
- [28] REISS, M. and WAHL, M. (2018). Non-asymptotic upper bounds for the reconstruction error of PCA. *arXiv:1609.03779*.
- [29] RIGOLLET, P. (2015). *Lecture notes on High-dimensional statistics*.
- [30] SPOKOINY, V. Gaussian approximation for a large ball probability. Manuscript.
- [31] TROPP, J. (2012). User-Friendly Tail Bounds for Sums of Random Matrices. *Found. Comput. Math.*, **12**, 4, 389–434. [MR2946459](#)
- [32] VAN DER VAART, A. W. (2000). *Asymptotic statistics. Cambridge series in statistical and probabilistic mathematics 3*, Cambridge University Press, Cambridge. [MR1652247](#)
- [33] VERSHYNIN, R. (2016). Introduction to the non-asymptotic analysis of random matrices. In *Compressed sensing*, 210–268, Cambridge University Press, Cambridge. [MR2963170](#)
- [34] WANG, W. and FAN, J. (2017). Asymptotics of empirical eigenstructure for high dimensional spiked covariance *Ann. Statist.*, **45**, 3, 1342–1374. [MR3662457](#)
- [35] ZHOU, H. H. and GAO, C. (2016). Bernstein-von Mises theorems for functionals of the covariance matrix. *Electron. J. Statist.*, **10**, 2, 1751–1806. [MR3522660](#)