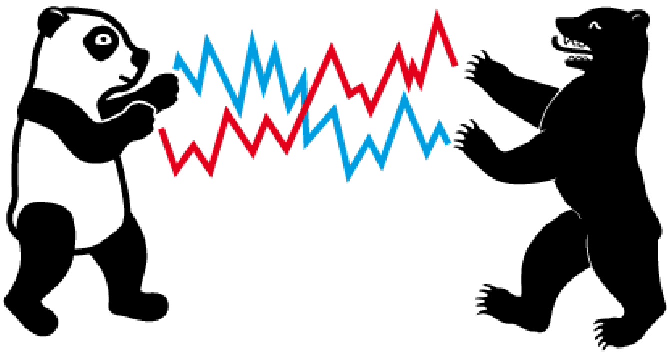
IRTG 1792 Discussion Paper 2018-027

Bayesian inference for spectral projectors of covariance matrix

Igor Silin *
Vladimir Spokoiny *[2]



* Skolkovo Institute of Science and Technology, Russia
*[2] WIAS, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the International Research Training Group 1792 "High Dimensional Nonstationary Time Series".

http://irtg1792.hu-berlin.de
ISSN 2568-5619

International Research Training Group 1792

Bayesian inference for spectral projectors of covariance matrix

Igor Silin*

Moscow Institute of Physics and Technology,
Skolkovo Institute of Science and Technology,
Institute for Information Transmission Problems RAS,
Moscow, Russia
e-mail: siliniv@gmail.com

and

Vladimir Spokoiny*

Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Germany
Higher School of Economics, Skoltech, Moscow, Russia
e-mail: spokoiny@wias-berlin.de

Abstract: Let $X_1, \ldots, X_n$ be i.i.d. sample in $\mathbb{R}^p$ with zero mean and the covariance matrix $\boldsymbol{\Sigma}^*$. The classic principal component analysis estimates the projector $\boldsymbol{P}^*_{\mathcal{J}}$ onto the direct sum of some eigenspaces of $\boldsymbol{\Sigma}^*$ by its empirical counterpart $\widehat{\boldsymbol{P}}_{\mathcal{J}}$. Recent papers [20, 23] investigate the asymptotic distribution of the Frobenius distance between the projectors $\|\widehat{\boldsymbol{P}}_{\mathcal{J}} - \boldsymbol{P}^*_{\mathcal{J}}\|_2$. The problem arises when one tries to build a confidence set for the true projector effectively. We consider the problem from Bayesian perspective and derive an approximation for the posterior distribution of the Frobenius distance between projectors. The derived theorems hold true for non-Gaussian data: the only assumption that we impose is the concentration of the sample covariance $\widehat{\boldsymbol{\Sigma}}$ in a vicinity of $\boldsymbol{\Sigma}^*$. The obtained results are applied to construction of sharp confidence sets for the true projector. Numerical simulations illustrate good performance of the proposed procedure even on non-Gaussian data in quite 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.

*Supported by the Russian Science Foundation grant (project 14-50-00150)

Contents

1. Introduction

Suppose the data $\boldsymbol{X}^n = (X_1, \ldots, X_n)$ are independent identically distributed zero-mean random vectors in $\mathbb{R}^p$. Denote by $\boldsymbol{\Sigma}^*$ its covariance matrix:

$$\boldsymbol{\Sigma}^* \stackrel{\text{def}}{=} \mathbb{E}\left(X_j X_j^\top\right).$$

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

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

Quantifying the quality of approximation of $\boldsymbol{\Sigma}^*$ by $\widehat{\boldsymbol{\Sigma}}$ is one of the most classical problems in statistics. Surprisingly, there is 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 $\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*\|_\infty$ in the random matrix theory, see, for instance, [19, 24, 25, 27, 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 on very high data dimension.

Functionals of covariance matrix also arise in applications frequently. For instance, eigenvalues are well-studied in different regimes, see [22, 9, 16] 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. Suppose we fix some set $\mathcal{J}$ of eigenspaces of $\boldsymbol{\Sigma}^*$ and consider direct sum of these eigenspaces and the associated true projector $\boldsymbol{P}^*_{\mathcal{J}}$. Its empirical counterpart is given by $\widehat{\boldsymbol{P}}_{\mathcal{J}}$ computed from the sample covariance $\widehat{\boldsymbol{\Sigma}}$. These objects are closely related to the Principal Component Analysis (PCA), probably the most famous dimension reduction method. Nowadays PCA-based methods are actively used in deep networking architecture [14] and finance [11], along with other applications. Over the past decade huge progress was achieved in theoretical guarantees for sparse PCA in high dimensions, see [15, 5, 3, 6, 12].

The random quantity of our interest is the squared Frobenius distance between the true projector and the sample one $\|\widehat{\boldsymbol{P}}_{\mathcal{J}} - \boldsymbol{P}^*_{\mathcal{J}}\|_2^2$. Even though this is a complex non-linear object, recent technique from [18] allows to approximate $(\widehat{\boldsymbol{P}}_{\mathcal{J}} - \boldsymbol{P}^*_{\mathcal{J}})$ by a linear functional of $(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)$ with root-n accuracy. Several results about the distribution of this random variable are available for the case when the observations are Gaussian: $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(0, \boldsymbol{\Sigma}^*)$. The normal approximation of $n\|\widehat{\boldsymbol{P}}_r - \boldsymbol{P}^*_r\|_2^2$ was shown in [20] with a tight bound on

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}\left\{ \frac{n\|\widehat{\boldsymbol{P}}_r - \boldsymbol{P}^*_r\|_2^2 - \mathbb{E}\left(n\|\widehat{\boldsymbol{P}}_r - \boldsymbol{P}^*_r\|_2^2\right)}{\mathrm{Var}^{1/2}\left(n\|\widehat{\boldsymbol{P}}_r - \boldsymbol{P}^*_r\|_2^2\right)} \le x \right\} - \Phi(x) \right|,$$

where $\Phi(x)$ is the standard normal distribution function. However, the distribution of $n\|\widehat{\boldsymbol{P}}_r - \boldsymbol{P}^*_r\|_2^2$ depends on the unknown covariance matrix which makes difficult to use this result for constructing the confidence sets for the true projector $\boldsymbol{P}^*_{\mathcal{J}}$. A bootstrap approach can be used to overcome this difficulty; see [23]. The bootstrap validity result is based on the approximation of the distribution of $n\|\widehat{\boldsymbol{P}}_r - \boldsymbol{P}^*_r\|_2^2$ by the distribution of a Gaussian quadratic form $\|\xi\|^2$. Namely, for the Gaussian data, Theorem 4.3 of [23] provides the following statement:

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}(n\|\widehat{\boldsymbol{P}}_r-\boldsymbol{P}_r^*\|_2^2\le x)-\mathbb{P}(\|\xi\|^2\le x)\right|\le\overline{\Diamond}, \tag{1.1}$$

where ξ 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 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. BvM result states some pivotal (Gaussian) behavior of the posterior. The paper [8] developed a general framework for functional BvM theorem, while [28] 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 [21, 26, 13, 17, 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 to establish precise finite sample results for the posterior credible sets under mild and general assumptions of the data distribution. The key observation here is that, similarly to the bootstrap approach of [23], the credible level sets for the posterior are nearly elliptic, and the corresponding posterior probability can be approximated by a special chi-squared-type distribution. This allows to apply the recent "large ball" results on Gaussian comparison and Gaussian anti-concentration from [23]. Moreover, in the contrary to the latter paper [23], we do not require Gaussian distribution of the data. We also provide explicit bounds on the approximation error in terms of p, n and $\boldsymbol{\Sigma}^*$. Finally, we justify the use of the Bayesian credible level sets as frequentist confidence sets.

The main contributions of this paper are as follows.

- We establish novel results on the coverage properties of 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.

- We offer a new procedure for building sharp 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).

The rest of the paper is structured as follows. Some notations are introduced in Section 2.1. Section 2.2 discusses the model. Bayesian framework and the main result of the paper about the 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 $\boldsymbol{I}_d$ for the identity matrix of size $d\times d$, $\operatorname{rank}(A)$ and $\operatorname{Tr}(B)$ 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 $\langle A, B\rangle_2 \stackrel{\text{def}}{=} \operatorname{Tr}(A^\top B)$, 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 $r(B) \stackrel{\text{def}}{=} \frac{\operatorname{Tr}(B)}{\|B\|_\infty}$. The relation $a \lesssim b$ means that there exists an absolute constant C, different from line to line, such that $a \le Cb$, 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 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

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

$$\boldsymbol{P}_r^* = \sum_{j \in \Delta_r^*} u_j^* {u_j^*}^\top$$

and the eigendecomposition

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

We also introduce the spectral gaps g_r^*:

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

Suppose that $\widehat{\boldsymbol{\Sigma}}$ has p eigenvalues $\widehat{\sigma}_1 > \ldots > \widehat{\sigma}_p$ (distinct with probability one). The corresponding eigenvectors are denoted as $\widehat{u}_1, \ldots, \widehat{u}_p$. Suppose that $\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*\|_\infty \leq \frac{1}{4} \min_{r \in \overline{1,q}} g_r^*$. Then, as shown in [18], we can identify clusters of the eigenvalues of $\widehat{\boldsymbol{\Sigma}}$ corresponding to each eigenvalue of $\boldsymbol{\Sigma}^*$ and therefore determine Δ_r^* and m_r^* for all $r \in \overline{1,q}$. Then we can define the sample projector on the r-th eigenspace of dimension m_r^*:

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

Under the condition that the spectral gap is sufficiently large, [23] approximated the distribution of $n\|\widehat{\boldsymbol{P}}_r - \boldsymbol{P}_r^*\|_2^2$ by the distribution of a Gaussian quadratic form $\|\xi\|^2$ with $\xi \sim \mathcal{N}(0, \mathit{\Gamma}_r^*)$ and $\mathit{\Gamma}_r^*$ is a block-matrix of the form

$$\mathit{\Gamma}_r^* \stackrel{\text{def}}{=} \begin{bmatrix} \mathit{\Gamma}_{r1}^* & O & \dots & O \\ O & \mathit{\Gamma}_{r2}^* & \dots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \dots & \mathit{\Gamma}_{rq}^* \end{bmatrix} \tag{2.1}$$

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

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

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^+ :

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

Define also the subset of indices

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

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

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

Its empirical counterpart is given by

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

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

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

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

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

$$l^*_{\mathcal{J}} = \mu^*_{r^-} - \mu^*_{r^+}.$$

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

$$\begin{aligned} \varGamma^*_{\mathcal{J}} &\stackrel{\text{def}}{=} \operatorname{diag}\left(\varGamma^r_{\mathcal{J}}\right)_{r\in\mathcal{J}}, \\ \varGamma^r_{\mathcal{J}} &\stackrel{\text{def}}{=} \operatorname{diag}\left(\varGamma^{r,s}\right)_{s\notin\mathcal{J}}, \\ \varGamma^{r,s} &\stackrel{\text{def}}{=} \frac{2\mu^*_r\mu^*_s}{(\mu^*_r - \mu^*_s)^2} \cdot \boldsymbol{I}_{m^*_r m^*_s}, \quad r \in \mathcal{J}, s \notin \mathcal{J}. \end{aligned} \tag{2.2}$$

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

Second, we relax the assumption on Gaussianity of the data. The only condition that our main result require from the underlying distribution of the independent random vectors $\boldsymbol{X}^n = (X_1, \ldots, X_n)$ is the concentration of the sample covariance matrix $\widehat{\boldsymbol{\varSigma}}$ around the true covariance $\boldsymbol{\varSigma}^*$:

$$\|\widehat{\boldsymbol{\varSigma}} - \boldsymbol{\varSigma}^*\|_\infty \le \widehat{\delta}_n \|\boldsymbol{\varSigma}^*\|_\infty \tag{2.3}$$

with probability $1-1/n$. Clearly, the bound $\widehat{\delta}_n$ from the condition can vary for different distributions of the data, but it allows to work with much wider classes of probability measures rather than just Gaussian or sub-Gaussian. While for the Gaussian case one may take

$$\widehat{\delta}_n \asymp \sqrt{\frac{r(\boldsymbol{\varSigma}^*)}{n}} \ \vee \ \sqrt{\frac{\log(n)}{n}},$$

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. Bayesian framework and credible level sets

In Bayesian framework one imposes a prior distribution Π on the set of considered covariance matrices $\boldsymbol{\Sigma}$. Even though our data are not necessary Gaussian, we can consider the Gaussian log-likelihood:

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

The posterior measure of a set $B \subset \mathbb{S}_+^p$ can be expressed as

$$\Pi\left(B \,\middle|\, \boldsymbol{X}^n\right) = \frac{\int_B \exp\left(l_n(\boldsymbol{\Sigma})\right) d\Pi(\boldsymbol{\Sigma})}{\int_{\mathbb{S}_+^p} \exp\left(l_n(\boldsymbol{\Sigma})\right) d\Pi(\boldsymbol{\Sigma})}.$$

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

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

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

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

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

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

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* $\boldsymbol{X}^n = (X_1, \ldots, X_n)$ *fulfills the sample covariance concentration property (2.3). Consider the prior* $\Pi(\boldsymbol{\Sigma})$ *given by the Inverse Wishart distribution* $\mathcal{IW}_p(\boldsymbol{G}, p+b-1)$. *Let* $\xi \sim \mathcal{N}(0, \Gamma^*_{\mathcal{J}})$ *with* $\Gamma^*_{\mathcal{J}}$ *defined by (2.2). Then with probability* $1 - \frac{1}{n}$

$$\sup_{x\in\mathbb{R}} \left| \Pi\left(n \|\boldsymbol{P}_{\mathcal{J}} - \widehat{\boldsymbol{P}}_{\mathcal{J}}\|_2^2 \le x \,\middle|\, \boldsymbol{X}^n \right) - \mathbb{P}(\|\xi\|^2 \le x) \right| \lesssim \Diamond,$$

where

$$\Diamond \stackrel{\text{def}}{=} \frac{\Diamond_1 + \Diamond_2 + \Diamond_3}{\sqrt{\lambda_1(\Gamma^*_{\mathcal{J}})\lambda_2(\Gamma^*_{\mathcal{J}})}} + \frac{1}{n}\,. \tag{2.4}$$

The terms $\Diamond_1$ *through* $\Diamond_3$ *can be described as*

$$\begin{aligned}
\Diamond_1 &\asymp \left\{ (\log(n)+p)\left(\left(1 + \frac{l^*_{\mathcal{J}}}{g^*_{\mathcal{J}}}\right) \frac{\sqrt{m^*_{\mathcal{J}}}\|\boldsymbol{\Sigma}^*\|^2_\infty}{g^*_{\mathcal{J}}} \vee \|\boldsymbol{\Sigma}^*\|_2 \right) + \|\boldsymbol{G}\|_2 \right\} \\
&\qquad\qquad\qquad\qquad\qquad\qquad \times \frac{\operatorname{Tr}(\boldsymbol{\Sigma}^*)}{{g^*_{\mathcal{J}}}^2} \sqrt{\frac{\log(n)+p}{n}}, \\
\Diamond_2 &\asymp \frac{\|\boldsymbol{\Sigma}^*\|_\infty \left(m^*_{\mathcal{J}} \|\boldsymbol{\Sigma}^*\|^2_\infty \wedge \operatorname{Tr}\left({\boldsymbol{\Sigma}^*}^2\right)\right)}{{g^*_{\mathcal{J}}}^3}\, p\left(\widehat{\delta}_n + \frac{p}{n}\right), \\
\Diamond_3 &\asymp \frac{{m^*_{\mathcal{J}}}^{3/2}\|\boldsymbol{\Sigma}^*\|_\infty \operatorname{Tr}(\boldsymbol{\Sigma}^*)}{{g^*_{\mathcal{J}}}^2} \sqrt{\frac{\log(n)}{n}}
\end{aligned}$$

with $\widehat{\delta}_n$ *from (2.3).*

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

$$\Diamond \asymp \sqrt{\frac{p^3}{n}} \vee \sqrt{\frac{\log^3(n)}{n}} \vee p\,\widehat{\delta}_n,$$

or, in the Gaussian case,

$$\Diamond \asymp \sqrt{\frac{p^3}{n}} \vee \sqrt{\frac{\log^3(n)}{n}}.$$

2.4. Gaussian approximation and frequentist uncertainty quantification for spectral projectors

For the Gaussian data, Theorem 4.3 of [23] provides the explicit error bound (1.1) with the error term $\overline{\diamondsuit}$ of the following form:

$$\sup_{x\in\mathbb{R}} \left|\mathbb{P}(n\|\widehat{\boldsymbol{P}}_r - \boldsymbol{P}_r^*\|_2^2 \le x) - \mathbb{P}(\|\xi\|^2 \le x)\right| \le \overline{\diamondsuit},$$

$$\overline{\diamondsuit} \asymp \frac{\sqrt{m_r^*}\,\mathrm{Tr}(\varGamma_r^*)}{\sqrt{\lambda_1(\varGamma_r^*)\lambda_2(\varGamma_r^*)}} \left(\sqrt{\frac{\log(n)}{n}} + \sqrt{\frac{\log(p)}{n}}\right) + \frac{m_r^*}{{g_r^*}^3}\frac{\mathrm{Tr}^3(\boldsymbol{\Sigma}^*)}{\sqrt{\lambda_1(\varGamma_r^*)\lambda_2(\varGamma_r^*)}}\sqrt{\frac{\log^3(n)}{n}}\,. \tag{2.5}$$

The next theorem extends this result to include the case of a generalized spectral cluster and of non-Gaussian data.

Theorem 2.2. *Assume the distribution of the data $\boldsymbol{X}^n = (X_1, \ldots, X_n)$ fulfills the sample covariance concentration property (2.3). Suppose additionally that the projections $\boldsymbol{P}^*_{\mathcal{J}} X_j$ and $(\boldsymbol{I}_p - \boldsymbol{P}^*_{\mathcal{J}})X_j$ are independent and the sixth moment of rescaled random vector ${\boldsymbol{\Sigma}^*}^{-1/2}X_j$ is bounded: $\mathbb{E}\|{\boldsymbol{\Sigma}^*}^{-1/2}X_j\|^6 \le C$. Let $\xi \sim \mathcal{N}(0, \varGamma^*_{\mathcal{J}})$ with $\varGamma^*_{\mathcal{J}}$ defined by (2.2). Then*

$$\sup_{x\in\mathbb{R}} \left|\mathbb{P}\left(n\|\widehat{\boldsymbol{P}}_{\mathcal{J}} - \boldsymbol{P}^*_{\mathcal{J}}\|_2^2 \le x\right) - \mathbb{P}(\|\xi\|^2 \le x)\right| \lesssim \overline{\diamondsuit},$$

where

$$\overline{\diamondsuit} \stackrel{\mathrm{def}}{=} \frac{p^{1/4}}{\sqrt{n}} + \frac{\overline{\Delta}}{\sqrt{\lambda_1(\varGamma^*_{\mathcal{J}})\lambda_2(\varGamma^*_{\mathcal{J}})}}, \tag{2.6}$$

$$\overline{\Delta} \stackrel{\mathrm{def}}{=} n m^*_{\mathcal{J}}\left(1 + \frac{l^*_{\mathcal{J}}}{g^*_{\mathcal{J}}}\right)\left\{\left(1 + \frac{l^*_{\mathcal{J}}}{g^*_{\mathcal{J}}}\right)\frac{\widehat{\delta}_n^4}{{g^*_{\mathcal{J}}}^4} \vee \frac{|\mathcal{J}|\,\widehat{\delta}_n^3}{{g^*_{\mathcal{J}}}^3}\right\}.$$

Remark 2.2. *The condition on independence of $\boldsymbol{P}^*_{\mathcal{J}} X_j$ and $(\boldsymbol{I}_p - \boldsymbol{P}^*_{\mathcal{J}})X_j$ may seem rather restrictive, however it has a natural interpretation: while we are interested in the "signal" $\boldsymbol{P}^*_{\mathcal{J}} X_j$, the orthogonal part $(\boldsymbol{I}_p - \boldsymbol{P}^*_{\mathcal{J}})X_j$ can be considered as "noise", and it is not too restrictive 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 posterior.*

The proof of this result is presented in Appendix B. The obtained bound is worse than (2.5) because we cannot utilize Gaussianity of the data anymore, and the result makes use of Gaussian approximation technique from [2]. However, recent developments in Gaussian approximation for a probability of a ball indicate that the bound (2.6) can be significantly improved.

Comparison of the results of Theorem 2.1 and Theorem 2.2 reveals that the posterior distribution of $n\|\boldsymbol{P}_{\mathcal{J}} - \widehat{\boldsymbol{P}}_{\mathcal{J}}\|_2^2$ given the data perfectly mimics the distribution of $n\|\widehat{\boldsymbol{P}}_{\mathcal{J}} - \boldsymbol{P}^*_{\mathcal{J}}\|_2^2$, 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

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

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

$$\gamma^\circ_\alpha \stackrel{\text{def}}{=} \inf\left\{\gamma > 0 : \varPi\left(n\|\boldsymbol{P}_{\mathcal{J}} - \widehat{\boldsymbol{P}}_{\mathcal{J}}\|_2^2 > \gamma\ \big|\, \boldsymbol{X}^n\right) \le \alpha\right\}.$$

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*

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

where $\diamondsuit$ *,* $\overline{\diamondsuit}$ *are defined by (2.4), (2.6), respectively.*

3. Numerical experiments

This section shows by mean of artificial data that the proposed 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. The organization of the experiments is the following. Let us fix some true covariance matrix $\boldsymbol{\Sigma}^*$ of size $p\times p$. Without loss of generality we consider only diagonal $\boldsymbol{\Sigma}^*$ in all our experiments, so $\boldsymbol{\Sigma}^*$ is defined by the distinct eigenvalues μ^*_r and the multiplicities m^*_r. We also specify the desired subspace that we want to investigate by fixing $\mathcal{J}$. After that, for different sample sizes

n we repeat the following two-step procedure. The first step is to determine the quantiles of $n\|\widehat{\boldsymbol{P}}_{\mathcal{J}} - \boldsymbol{P}^*_{\mathcal{J}}\|_2^2$. For that we generate 3000 samples $\boldsymbol{X}^n$, compute the corresponding $\widehat{\boldsymbol{P}}_{\mathcal{J}}$ and then just take $\alpha-$ quantiles of the obtained realizations $n\|\widehat{\boldsymbol{P}}_{\mathcal{J}} - \boldsymbol{P}^*_{\mathcal{J}}\|_2^2$ for α from 0.001 to 0.999 with step 0.001 . The second step is to estimate the quantiles of the pseudo-posterior distribution of $n\|\boldsymbol{P}_{\mathcal{J}} - \widehat{\boldsymbol{P}}_{\mathcal{J}}\|_2^2$. We generate 50 samples $\boldsymbol{X}^n$ and for each realization we generate 3000 pseudo-posterior covariance matrices $\boldsymbol{\Sigma}$ from the Inverse Wishart distribution with $G = \boldsymbol{I}_p,\ b = 1$. Then we compute the corresponding $\boldsymbol{P}_{\mathcal{J}}$ and take the $\alpha-$ quantiles of $n\|\boldsymbol{P}_{\mathcal{J}} - \widehat{\boldsymbol{P}}_{\mathcal{J}}\|_2^2$ just as in the first step. So, for each α we get 50 quantile estimates γ°_α and take median of them. For the quantiles from the first and the second step we build the QQ-plot. Also we present a table with coverage probabilities $\mathbb{P}(n\|\widehat{\boldsymbol{P}}_{\mathcal{J}} - \boldsymbol{P}^*_{\mathcal{J}}\|_2^2 \le \gamma^\circ_\alpha)$ and interquartile ranges for $1-\alpha$ from $\{0.99, 0.95, 0.90, 0.85, 0.80, 0.75\}$.

Let us look at the examples of conducted simulations. In the first experiment we work with Gaussian data. The parameters of the experiment are as follows:

- $p = 100$.
- $m^*_r = 1$ for all $r = \overline{1, 100}$.
- $\mu^*_1 = 26,\ \mu^*_2 = 16,\ \mu^*_3 = 10,\ \mu^*_4 = 6,\ \mu^*_5 = 3.5$ and the rest of the eigenvalues $\mu^*_6, \ldots, \mu^*_{100}$ are from the uniform distribution on $[0.7;\ 1.3]$.
- $\mathcal{J} = \{1\}$, so we investigate one-dimensional principal subspace given by $\boldsymbol{P}^*_1$.

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

In the second experiment we check how our method performs on non-Gaussian data. We generate non-Gaussian data in the following way. Since we consider only diagonal matrices, we can generate components of the vectors X_j independently. Except 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 $m^*_4, \ldots, m^*_{91}$

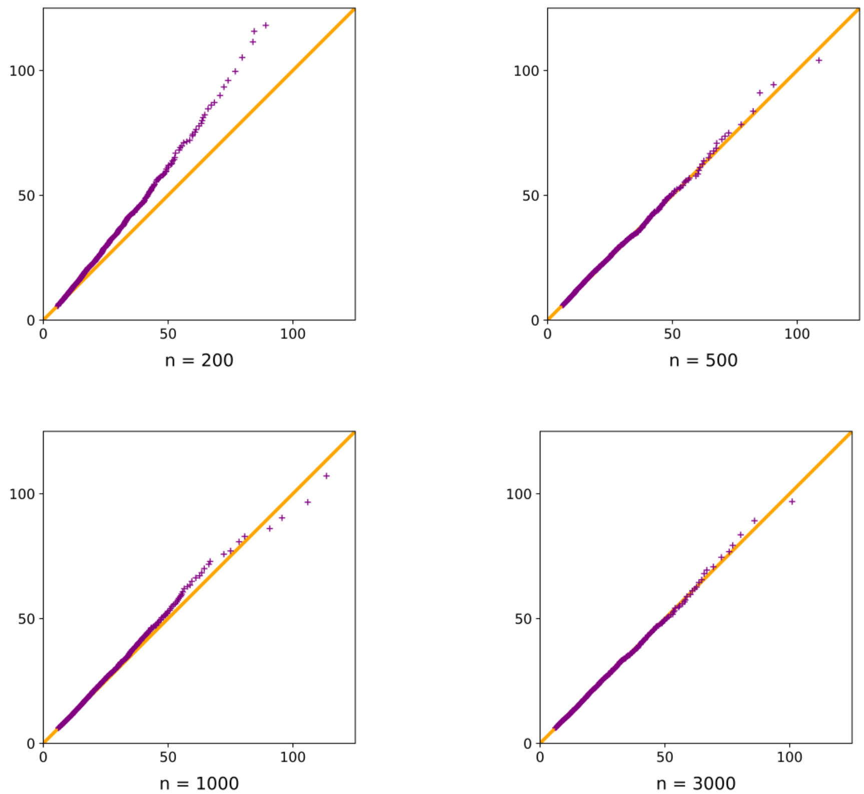


Fig 1. *QQ-plots of the proposed Bayesian procedure for the first experiment (Gaussian data).*

Table 1
Coverage probabilities of the proposed Bayesian procedure for the first experiment (Gaussian data).

	Confidence levels $(1-\alpha)$					
n	0.99	0.95	0.90	0.85	0.80	0.75
200	0.997	0.972	0.935	0.900	0.859	0.815
	0.008	0.042	0.066	0.104	0.127	0.147
500	0.992	0.950	0.898	0.856	0.805	0.758
	0.023	0.076	0.101	0.123	0.142	0.153
1000	0.992	0.958	0.910	0.859	0.816	0.768
	0.012	0.043	0.076	0.080	0.083	0.094
3000	0.991	0.950	0.902	0.854	0.806	0.753
	0.009	0.025	0.035	0.039	0.043	0.056

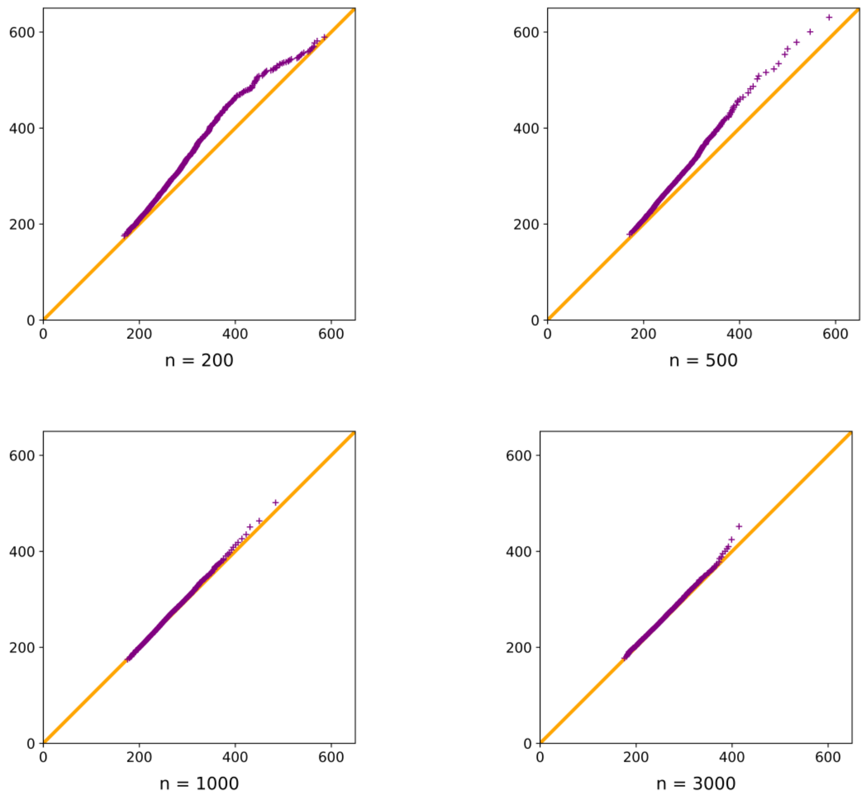


Fig 2. *QQ-plots of the proposed Bayesian procedure for the second experiment (non-Gaussian data).*

Table 2
Coverage probabilities of the proposed Bayesian procedure for the second experiment (non-Gaussian data)

	Confidence levels $(1-\alpha)$					
n	0.99	0.95	0.90	0.85	0.80	0.75
200	0.993	0.972	0.946	0.919	0.886	0.858
	0.019	0.071	0.117	0.158	0.185	0.233
500	0.995	0.978	0.952	0.925	0.889	0.852
	0.015	0.045	0.087	0.120	0.145	0.175
1000	0.992	0.959	0.909	0.864	0.818	0.774
	0.015	0.049	0.080	0.106	0.126	0.141
3000	0.992	0.958	0.914	0.867	0.821	0.776
	0.014	0.033	0.055	0.062	0.083	0.082

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 $\mu_7^*, \ldots, \mu_{100}^*$ 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 $\boldsymbol{P}_1^* + \boldsymbol{P}_2^* + \boldsymbol{P}_3^*$.

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

As we can see from the experiment, the performance of the proposed procedure is rather poor when the sample size is of the same order as the dimension. However, this regime lies beyond the scope of our results. If we have enough data, the methods demonstrates high quality even in such challenging situation as recovering a direct sum of three 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 $\mathcal{IW}_p(\boldsymbol{G}, p + b - 1)$ is conjugate to the multivariate Gaussian distribution, so our pseudo-posterior $\Pi\left(\boldsymbol{\Sigma} \,\middle|\, \boldsymbol{X}^n\right)$ is $\mathcal{IW}_p(\boldsymbol{G} + n\widehat{\boldsymbol{\Sigma}}, n + p + b - 1)$. We will actively use the following well-known property of the Wishart distribution:

$$\boldsymbol{\Sigma}^{-1} \,\big|\, \boldsymbol{X}^n \overset{d}{=} \sum_{j=1}^{n+p+b-1} W_j W_j^\top \,\big|\, \boldsymbol{X}^n,$$

where $W_j \,\big|\, \boldsymbol{X}^n \overset{i.i.d.}{\sim} \mathcal{N}(0, (\boldsymbol{G} + n\widehat{\boldsymbol{\Sigma}})^{-1})$.

For shortness in this section we will use the notation $n_p \overset{\text{def}}{=} n + p + b - 1$ 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

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

and

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

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

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

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

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

and its posterior measure

$$\Pi\left(\Upsilon \,\big|\, \boldsymbol{X}^n\right) \ge 1 - \frac{1}{n}\,.$$

Step 1 First, we will need the following lemma.

Lemma 4.1. *The following holds on the random set* Υ *:*

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

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

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

Note that

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

Hence,

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

Finally, the observations that

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

finish the proof. □

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

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

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

Lemma 4.2. *The following holds on the random set* Υ *:*

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

where

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

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

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

It is easy to see that the rank of $\widehat{L}_{\mathcal{J}}(\boldsymbol{\Sigma} - \widehat{\boldsymbol{\Sigma}})$ is at most $2m^*_{\mathcal{J}}$, and thus the rank of $\boldsymbol{P}_{\mathcal{J}} - \widehat{\boldsymbol{P}}_{\mathcal{J}} - \widehat{L}_{\mathcal{J}}(\boldsymbol{\Sigma} - \widehat{\boldsymbol{\Sigma}})$ 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. □

The representation

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

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

Lemma 4.3. *It holds*

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

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

$$\sqrt{n}\|\mathcal{R}_{\mathcal{J}}\|_2 \lesssim \widehat{\Delta}_1 \overset{\text{def}}{=} \frac{1}{\sqrt{n}} \cdot \frac{(\log(n)+p)\|\widehat{\boldsymbol{\Sigma}}\|_2 + \|\boldsymbol{G}\|_2}{\widehat{g}_{\mathcal{J}}}.$$

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

$$\boldsymbol{R}_{n,p} \overset{\text{def}}{=} \left(\boldsymbol{I}_p + \boldsymbol{E}_{n,p}\right)^{-1} - \boldsymbol{I}_p + \boldsymbol{E}_{n,p}.$$

Its spectral norm can be bounded as

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

So

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

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

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

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

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

where we introduce the remainder terms

$$
\begin{aligned}
\boldsymbol{R}_1 &\stackrel{\text{def}}{=} \widehat{\boldsymbol{\Sigma}}^{1/2}\,\boldsymbol{E}_{n,p}\,\widehat{\boldsymbol{\Sigma}}^{1/2} - \boldsymbol{\Sigma}_{n,p}^{1/2}\,\boldsymbol{E}_{n,p}\,\boldsymbol{\Sigma}_{n,p}^{1/2},\\
\boldsymbol{R}_2 &\stackrel{\text{def}}{=} \boldsymbol{\Sigma}_{n,p}^{1/2}\,\boldsymbol{R}_{n,p}\,\boldsymbol{\Sigma}_{n,p}^{1/2},\\
\boldsymbol{R}_3 &\stackrel{\text{def}}{=} \boldsymbol{\Sigma}_{n,p} - \widehat{\boldsymbol{\Sigma}}.
\end{aligned}
$$

They can be bounded in Frobenius norm:

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

Hence, omitting higher order terms, on Υ we have

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

Now we summarize

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

with

$$
\widehat{S}_{\mathcal{J}} \stackrel{\text{def}}{=} -\sum_{k\in\mathcal{I}_{\mathcal{J}}} \sum_{l\notin\mathcal{I}_{\mathcal{J}}} \frac{\widehat{\sigma}_k^{1/2}\,\widehat{\sigma}_l^{1/2}\,(\widehat{u}_k\widehat{u}_k^\top \boldsymbol{E}_{n,p}\,\widehat{u}_l\widehat{u}_l^\top + \widehat{u}_l\widehat{u}_l^\top \boldsymbol{E}_{n,p}\,\widehat{u}_k\widehat{u}_k^\top)}{\widehat{\sigma}_k - \widehat{\sigma}_l},
$$

$$
\mathcal{R}_{\mathcal{J}} \stackrel{\text{def}}{=} \sum_{k\in\mathcal{I}_{\mathcal{J}}} \sum_{l\notin\mathcal{I}_{\mathcal{J}}} \frac{(\widehat{u}_k\widehat{u}_k^\top (\boldsymbol{R}_1+\boldsymbol{R}_2+\boldsymbol{R}_3)\,\widehat{u}_l\widehat{u}_l^\top + \widehat{u}_l\widehat{u}_l^\top (\boldsymbol{R}_1+\boldsymbol{R}_2+\boldsymbol{R}_3)\,\widehat{u}_k\widehat{u}_k^\top)}{\widehat{\sigma}_k - \widehat{\sigma}_l}.
$$

Moreover,

$$
\begin{aligned}
\left\|\mathcal{R}_{\mathcal{J}}\right\|_2^2 &= \operatorname{Tr}\bigl(\mathcal{R}_{\mathcal{J}}\mathcal{R}_{\mathcal{J}}^\top\bigr) = 2\sum_{k\in\mathcal{I}_{\mathcal{J}}} \sum_{l\notin\mathcal{I}_{\mathcal{J}}} \frac{\left(\widehat{u}_k^\top (\boldsymbol{R}_1+\boldsymbol{R}_2+\boldsymbol{R}_3)\,\widehat{u}_l\right)^2}{(\widehat{\sigma}_k - \widehat{\sigma}_l)^2}\\
&\le \frac{2\|\boldsymbol{R}_1+\boldsymbol{R}_2+\boldsymbol{R}_3\|_2^2}{\widehat{g}_{\mathcal{J}}^2},
\end{aligned}
$$

and the inequality

$$\left\|\mathcal{R}_{\mathcal{J}}\right\|_2 \le \frac{\sqrt{2}}{\widehat{g}_{\mathcal{J}}}\left(\|\boldsymbol{R}_1\|_2 + \|\boldsymbol{R}_2\|_2 + \|\boldsymbol{R}_3\|_2\right)$$

provides the desired bound. Similarly, we have

$$\left\|\widehat{S}_{\mathcal{J}}\right\|_2 \le \frac{\sqrt{2}\|\boldsymbol{E}_{n,p}\|_\infty}{\widehat{g}_{\mathcal{J}}}\left(\sum_{k\in\mathcal{I}_{\mathcal{J}}}\widehat{\sigma}_k\right)^{1/2}\left(\sum_{l\notin\mathcal{I}_{\mathcal{J}}}\widehat{\sigma}_l\right)^{1/2} \lesssim \frac{\operatorname{Tr}(\widehat{\boldsymbol{\Sigma}})}{\widehat{g}_{\mathcal{J}}}\sqrt{\frac{\log(n)+p}{n}},$$

where the last inequality holds on $\varUpsilon$. □

The results of Lemmas 4.2 and 4.3 yield on the random set $\varUpsilon$

$$\sqrt{n}\|\widehat{\boldsymbol{P}}_{\mathcal{J}} - \boldsymbol{P}_{\mathcal{J}} - \widehat{S}_{\mathcal{J}}\|_2 \lesssim \widehat{\varDelta}_0 + \widehat{\varDelta}_1.$$

In addition,

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

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

$$\left|n\|\boldsymbol{P}_{\mathcal{J}} - \widehat{\boldsymbol{P}}_{\mathcal{J}}\|_2^2 - n\|\widehat{S}_{\mathcal{J}}\|_2^2\right| \lesssim \widehat{\varDelta}_2, \tag{4.2}$$

where

$$\begin{aligned}
\widehat{\varDelta}_2 \stackrel{\text{def}}{=} &\left\{(\log(n)+p)\left(\left(1+\frac{\widehat{l}_{\mathcal{J}}}{\widehat{g}_{\mathcal{J}}}\right)\frac{\sqrt{m^*_{\mathcal{J}}}\|\widehat{\boldsymbol{\Sigma}}\|_\infty^2}{\widehat{g}_{\mathcal{J}}} \vee \|\widehat{\boldsymbol{\Sigma}}\|_2\right) + \|\boldsymbol{G}\|_2\right\} \times \\
&\times \frac{\operatorname{Tr}(\widehat{\boldsymbol{\Sigma}})}{\widehat{g}_{\mathcal{J}}^2}\sqrt{\frac{\log(n)+p}{n}}.
\end{aligned}$$

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

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

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

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

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 $k \le p$ and $j \le n_p$

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

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:

$$\begin{aligned}
\widehat{\varGamma}_{(k,l),(k',l')} &\stackrel{\mathrm{def}}{=} \operatorname{Cov}(\widehat{\xi}_{k,l}, \widehat{\xi}_{k',l'} \,\big|\, \boldsymbol{X}^n) \\
&= \frac{2n}{n_p^2}\sum_{j,j'=1}^{n_p}\frac{\widehat{\sigma}_k^{1/2}\,\widehat{\sigma}_l^{1/2}\,\widehat{\sigma}_{k'}^{1/2}\,\widehat{\sigma}_{l'}^{1/2}}{(\widehat{\sigma}_k - \widehat{\sigma}_l)(\widehat{\sigma}_{k'} - \widehat{\sigma}_{l'})}\mathbb{E}\big(\eta_{k,j}\,\eta_{l,j}\,\eta_{k',j'}\,\eta_{l',j'} \,\big|\, \boldsymbol{X}^n\big) \\
&= \frac{2n}{n_p}\delta_{k,k'}\,\delta_{l,l'}\,\frac{\widehat{\sigma}_k\,\widehat{\sigma}_l}{(\widehat{\sigma}_k - \widehat{\sigma}_l)^2}
\end{aligned}$$

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

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

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

Lemma 4.4. *It holds*

$$\big\|\widehat{\varGamma}_{\mathcal{J}} - \varGamma^*_{\mathcal{J}}\big\|_1 \lesssim \widehat{\varDelta}_3 \tag{4.3}$$

with

$$\widehat{\varDelta}_3 \stackrel{\mathrm{def}}{=} \frac{p\,\Big(m^*_{\mathcal{J}}\,\|\boldsymbol{\Sigma}^*\|_\infty^2 \wedge \operatorname{Tr}\Big({\boldsymbol{\Sigma}^*}^2\Big)\Big)}{{g^*_{\mathcal{J}}}^3}\Big(\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*\|_\infty + \frac{p}{n}\|\boldsymbol{\Sigma}^*\|_\infty\Big).$$

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

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

Due to the Weyl's inequality, $|\widehat{\sigma}_k - \sigma^*_k| \le \|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*\|_\infty$ for any k. Then, after some technical calculations, we get

$$\big\|\widehat{\Gamma}_{\mathcal{J}} - \Gamma^*_{\mathcal{J}}\big\|_1 \lesssim \sum_{k \in \mathcal{I}_{\mathcal{J}}} \sum_{l \notin \mathcal{I}_{\mathcal{J}}} \frac{{\sigma^*_k}^2 + {\sigma^*_l}^2}{(\sigma^*_k - \sigma^*_l)^3} \Big(\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*\|_\infty + \frac{p}{n} \|\boldsymbol{\Sigma}^*\|_\infty) \Big),$$

which provides the desired result. □

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 $\widehat{\boldsymbol{P}}_{\mathcal{J}} Z$. Namely, in the "posterior" world random vectors $\widehat{\boldsymbol{P}}_{\mathcal{J}} Z_j$ and $(\boldsymbol{I}_p - \widehat{\boldsymbol{P}}_{\mathcal{J}}) Z_j$ are Gaussian and uncorrelated, therefore, independent, so we can condition on $Z_{\mathcal{J}} \stackrel{\mathrm{def}}{=} (\widehat{\boldsymbol{P}}_{\mathcal{J}} Z_1, \ldots, \widehat{\boldsymbol{P}}_{\mathcal{J}} Z_{n_p})$ to get that $\widehat{S}_{\mathcal{J}}$ is conditionally on $\boldsymbol{X}^n, Z_{\mathcal{J}}$ Gaussian random vector with the covariance matrix

$$\widetilde{\Gamma}_{\mathcal{J}} \stackrel{\mathrm{def}}{=} \mathrm{Cov}\big(\widehat{\xi}_{\mathcal{J}} \,\big|\, \boldsymbol{X}^n, Z_{\mathcal{J}}\big).$$

It holds similarly to the above

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

with

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

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

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

and on this set

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

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

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

Due to this block structure we have

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

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 $k, k' \in \mathcal{I}_{\mathcal{J}}$

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

and, therefore,

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

Finally, since

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

the obtained inequalities provide the result of the lemma. □

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

$$\left\|\widetilde{\Gamma}_{\mathcal{J}} - \Gamma_{\mathcal{J}}^*\right\|_1 \lesssim \widehat{\Delta}_3 + \widehat{\Delta}_4.$$

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

$$\sup_{x \in \mathbb{R}} \left| \varPi \left(\|\widehat{\xi}_{\mathcal{J}}\|^2 \leq x \,\middle|\, \boldsymbol{X}^n, Z_{\mathcal{J}} \right) - \mathbb{P}\left(\|\xi_{\mathcal{J}}\|^2 \leq x \right) \right| \lesssim \frac{\widehat{\varDelta}_3 + \widehat{\varDelta}_4}{\sqrt{\lambda_1(\varGamma^*_{\mathcal{J}})\lambda_2(\varGamma^*_{\mathcal{J}})}} .$$

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

$$\sup_{x \in \mathbb{R}} \left| \varPi \left(\|\widehat{\xi}_{\mathcal{J}}\|^2 \leq x \,\middle|\, \boldsymbol{X}^n \right) - \mathbb{P}\left(\|\xi_{\mathcal{J}}\|^2 \leq x \right) \right| \lesssim \frac{\widehat{\varDelta}_3 + \widehat{\varDelta}_4}{\sqrt{\lambda_1(\varGamma^*_{\mathcal{J}})\lambda_2(\varGamma^*_{\mathcal{J}})}} + \frac{1}{n} \tag{4.5}$$

with probability one.

Step 3 So far we worked in the "posterior world" and our bounds $\widehat{\varDelta}_2, \widehat{\varDelta}_3, \widehat{\varDelta}_4$ are random, since they depend on the data $\boldsymbol{X}^n$. Clearly, one can verify that due to the Weyl's inequality and the condition (2.3) the empirical objects in the random bounds can be replaced by the true ones with high probability: the only payment for this is a multiplicative constant factor, if we assume that $\widehat{\delta}_n$ is small enough and neglect higher order terms. So, we get

$$\begin{aligned}
\widehat{\varDelta}_2 \lesssim \varDelta_2 &\stackrel{\text{def}}{=} \left\{ (\log(n) + p) \left(\left(1 + \frac{l^*_{\mathcal{J}}}{g^*_{\mathcal{J}}} \right) \frac{\sqrt{m^*_{\mathcal{J}}} \|\boldsymbol{\Sigma}^*\|^2_\infty}{g^*_{\mathcal{J}}} \vee \|\boldsymbol{\Sigma}^*\|_2 \right) + \|\boldsymbol{G}\|_2 \right\} \times \\
&\qquad \times \frac{\operatorname{Tr}(\boldsymbol{\Sigma}^*)}{{g^*_{\mathcal{J}}}^2} \sqrt{\frac{\log(n) + p}{n}}, \\
\widehat{\varDelta}_3 \lesssim \varDelta_3 &\stackrel{\text{def}}{=} \frac{\|\boldsymbol{\Sigma}^*\|_\infty \left(m^*_{\mathcal{J}} \|\boldsymbol{\Sigma}^*\|^2_\infty \wedge \operatorname{Tr}\left({\boldsymbol{\Sigma}^*}^2 \right) \right)}{{g^*_{\mathcal{J}}}^3} \, p \left(\widehat{\delta}_n + \frac{p}{n} \right), \\
\widehat{\varDelta}_4 \lesssim \varDelta_4 &\stackrel{\text{def}}{=} \frac{{m^*_{\mathcal{J}}}^{3/2} \|\boldsymbol{\Sigma}^*\|_\infty \operatorname{Tr}(\boldsymbol{\Sigma}^*)}{{g^*_{\mathcal{J}}}^2} \sqrt{\frac{\log(n)}{n}}
\end{aligned}$$

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

Now we combine the obtained bounds. For $\varDelta_2$ defined above and arbitrary

$x \in \mathbb{R}$ it holds

$$\begin{aligned} &\Pi\left(n\|\boldsymbol{P}_{\mathcal{J}} - \widehat{\boldsymbol{P}}_{\mathcal{J}}\|_2^2 \le x \,\middle|\, \boldsymbol{X}^n\right) \\ &\qquad \le \Pi\left(n\|\widehat{S}_{\mathcal{J}}\|_2^2 \le x + \Delta_2 \,\middle|\, \boldsymbol{X}^n\right) \\ &\qquad\qquad + \Pi\left(n\|\boldsymbol{P}_{\mathcal{J}} - \widehat{\boldsymbol{P}}_{\mathcal{J}}\|_2^2 - n\|\widehat{S}_{\mathcal{J}}\|_2^2 \le -\Delta_2 \,\middle|\, \boldsymbol{X}^n\right). \end{aligned}$$

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

$$\Pi\left(n\|\boldsymbol{P}_{\mathcal{J}} - \widehat{\boldsymbol{P}}_{\mathcal{J}}\|_2^2 \le x \,\middle|\, \boldsymbol{X}^n\right) \le \Pi\left(\|\xi_{\mathcal{J}}\|^2 \le x + \Delta_2 \,\middle|\, \boldsymbol{X}^n\right) + \Pi\left(\Upsilon^c \,\middle|\, \boldsymbol{X}^n\right)$$

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

$$\begin{aligned} &\sup_{x\in\mathbb{R}}\left\{\Pi\left(n\|\boldsymbol{P}_{\mathcal{J}} - \widehat{\boldsymbol{P}}_{\mathcal{J}}\|_2^2 \le x \,\middle|\, \boldsymbol{X}^n\right) - \mathbb{P}\left(\|\xi_{\mathcal{J}}\|^2 \le x\right)\right\} \\ &\qquad \le \sup_{x\in\mathbb{R}}\left\{\Pi\left(\|\widehat{\xi}_{\mathcal{J}}\|^2 \le x + \Delta_2 \,\middle|\, \boldsymbol{X}^n\right) - \mathbb{P}\left(\|\xi_{\mathcal{J}}\|^2 \le x + \Delta_2\right)\right\} \\ &\qquad\qquad + \sup_{x\in\mathbb{R}}\left\{\mathbb{P}\left(\|\xi_{\mathcal{J}}\|^2 \le x + \Delta_2\right) - \mathbb{P}\left(\|\xi_{\mathcal{J}}\|^2 \le x\right)\right\} + \Pi\left(\Upsilon^c \,\middle|\, \boldsymbol{X}^n\right). \end{aligned}$$

The first term in the right-hand side is bounded by $\frac{\Delta_3 + \Delta_4}{\sqrt{\lambda_1(\Gamma^*_{\mathcal{J}})\lambda_2(\Gamma^*_{\mathcal{J}})}} + \frac{1}{n}$ with probability $1 - \frac{1}{n}$ due to (4.5). The second term does not exceed $\frac{\Delta_2}{\sqrt{\lambda_1(\Gamma^*_{\mathcal{J}})\lambda_2(\Gamma^*_{\mathcal{J}})}}$ according to the Gaussian anti-concentration Lemma A.3. The last term is at most $\frac{1}{n}$ by definition of Υ. Therefore,

$$\sup_{x\in\mathbb{R}}\left\{\Pi\left(n\|\boldsymbol{P}_{\mathcal{J}} - \widehat{\boldsymbol{P}}_{\mathcal{J}}\|_2^2 \le x \,\middle|\, \boldsymbol{X}^n\right) - \mathbb{P}\left(\|\xi_{\mathcal{J}}\|^2 \le x\right)\right\} \lesssim \frac{\Delta_2 + \Delta_3 + \Delta_4}{\sqrt{\lambda_1(\Gamma^*_{\mathcal{J}})\lambda_2(\Gamma^*_{\mathcal{J}})}} + \frac{1}{n}$$

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

$$\sup_{x\in\mathbb{R}}\left\{\mathbb{P}\left(\|\xi_{\mathcal{J}}\|^2 \le x\right) - \Pi\left(n\|\boldsymbol{P}_{\mathcal{J}} - \widehat{\boldsymbol{P}}_{\mathcal{J}}\|_2^2 \le x \,\middle|\, \boldsymbol{X}^n\right)\right\} \lesssim \frac{\Delta_2 + \Delta_3 + \Delta_4}{\sqrt{\lambda_1(\Gamma^*_{\mathcal{J}})\lambda_2(\Gamma^*_{\mathcal{J}})}} + \frac{1}{n}$$

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

4.2. Proof of Corollary 2.3

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

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

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

$$\sup_{x\in\mathbb{R}} \left| \varPi\left(n\|\boldsymbol{P}_{\mathcal{J}} - \widehat{\boldsymbol{P}}_{\mathcal{J}}\|_2^2 > x \,\middle|\, \boldsymbol{X}^n \right) - \mathbb{P}(\|\xi_{\mathcal{J}}\|^2 > x) \right| \lesssim \lozenge.$$

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

$$\sup_{x\in\mathbb{R}} \left| \varPi\left(n\|\boldsymbol{P}_{\mathcal{J}} - \widehat{\boldsymbol{P}}_{\mathcal{J}}\|_2^2 > x \,\middle|\, \boldsymbol{X}^n \right) - \mathbb{P}\left(n\|\widehat{\boldsymbol{P}}_{\mathcal{J}} - \boldsymbol{P}^*_{\mathcal{J}}\|_2^2 > x \right) \right| \le \lozenge' \asymp \overline{\lozenge} + \lozenge.$$

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

$$\begin{aligned}
\left| \varPi\left(n\|\boldsymbol{P}_{\mathcal{J}} - \widehat{\boldsymbol{P}}_{\mathcal{J}}\|_2^2 > \gamma_{\alpha-\lozenge'} \,\middle|\, \boldsymbol{X}^n \right) - (\alpha - \lozenge') \right| &\le \lozenge', \\
\left| \varPi\left(n\|\boldsymbol{P}_{\mathcal{J}} - \widehat{\boldsymbol{P}}_{\mathcal{J}}\|_2^2 > \gamma_{\alpha+\lozenge'} \,\middle|\, \boldsymbol{X}^n \right) - (\alpha + \lozenge') \right| &\le \lozenge'.
\end{aligned}$$

Thus,

$$\begin{aligned}
\varPi\left(n\|\boldsymbol{P}_{\mathcal{J}} - \widehat{\boldsymbol{P}}_{\mathcal{J}}\|_2^2 > \gamma_{\alpha-\lozenge'} \,\middle|\, \boldsymbol{X}^n \right) &\le (\alpha - \lozenge') + \lozenge' = \alpha, \\
\varPi\left(n\|\boldsymbol{P}_{\mathcal{J}} - \widehat{\boldsymbol{P}}_{\mathcal{J}}\|_2^2 > \gamma_{\alpha+\lozenge'} \,\middle|\, \boldsymbol{X}^n \right) &\ge (\alpha + \lozenge') - \lozenge' = \alpha.
\end{aligned}$$

By definition of γ°_α the previous two inequalities yield

$$\gamma_{\alpha+\lozenge'} \le \gamma^\circ_\alpha \le \gamma_{\alpha-\lozenge'} \quad \text{on } \varTheta.$$

Hence,

$$\begin{aligned}
\mathbb{P}\left(\gamma^\circ_\alpha < \gamma_{\alpha+\lozenge'} \right) &\le \mathbb{P}\left(\varTheta^c \right) \le \frac{1}{n}, \\
\mathbb{P}\left(\gamma^\circ_\alpha > \gamma_{\alpha-\lozenge'} \right) &\le \mathbb{P}\left(\varTheta^c \right) \le \frac{1}{n}.
\end{aligned}$$

Now we can write the following chain of inequalities:

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

and

$$
\begin{aligned}
&\mathbb{P}\left(n\|\widehat{\boldsymbol{P}}_{\mathcal{J}}-\boldsymbol{P}^{*}_{\mathcal{J}}\|_2^2>\gamma^{\circ}_{\alpha}\right)=1-\mathbb{P}\left(n\|\widehat{\boldsymbol{P}}_{\mathcal{J}}-\boldsymbol{P}^{*}_{\mathcal{J}}\|_2^2\le\gamma^{\circ}_{\alpha}\right)\\
&\quad\ge 1-\mathbb{P}\left(\left\{n\|\widehat{\boldsymbol{P}}_{\mathcal{J}}-\boldsymbol{P}^{*}_{\mathcal{J}}\|_2^2\le\gamma_{\alpha-\lozenge'}\right\}\cup\{\gamma^{\circ}_{\alpha}>\gamma_{\alpha-\lozenge'}\}\right)\\
&\quad\ge 1-\mathbb{P}\left(n\|\widehat{\boldsymbol{P}}_{\mathcal{J}}-\boldsymbol{P}^{*}_{\mathcal{J}}\|_2^2\le\gamma_{\alpha-\lozenge'}\right)-\mathbb{P}\left(\gamma^{\circ}_{\alpha}>\gamma_{\alpha-\lozenge'}\right)\\
&\quad=\mathbb{P}\left(n\|\widehat{\boldsymbol{P}}_{\mathcal{J}}-\boldsymbol{P}^{*}_{\mathcal{J}}\|_2^2>\gamma_{\alpha-\lozenge'}\right)-\mathbb{P}\left(\gamma^{\circ}_{\alpha}>\gamma_{\alpha-\lozenge'}\right)\ge\alpha-\lozenge'-\frac{1}{n}.
\end{aligned}
$$

Finally, these inequalities imply the following bound

$$
\left|\alpha-\mathbb{P}\left(n\|\widehat{\boldsymbol{P}}_{\mathcal{J}}-\boldsymbol{P}^{*}_{\mathcal{J}}\|_2^2>\gamma^{\circ}_{\alpha}\right)\right|\le\lozenge'+\frac{1}{n},
$$

which concludes the proof.

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 $X_1,\ldots,X_n$ be i.i.d. zero-mean random vectors in $\mathbb{R}^p$. Denote the true covariance matrix as $\boldsymbol{\Sigma}^{*}\overset{\text{def}}{=}\mathbb{E}\left(X_iX_i^{\top}\right)$ and the sample covariance as $\widehat{\boldsymbol{\Sigma}}\overset{\text{def}}{=}\frac{1}{n}\sum_{i=1}^{n}X_iX_i^{\top}$. Suppose the data are obtained from:*
(i) Gaussian distribution $\mathcal{N}(0,\boldsymbol{\Sigma}^{*})$. In this case, define $\widehat{\delta}_n$ as

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

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

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

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

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

(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 $\widehat{\boldsymbol{\Sigma}}$ *holds with the corresponding* $\widehat{\delta}_n$*:*

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

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

Proof. (i) See [19], Corollary 2. (ii) This is a well-known simple result presented in a range of papers and lecture notes. See, e.g. [24], Theorem 4.6. (iii) See [27], Corollary 5.52. Usually the radius R is taken such that $\frac{R}{\sqrt{\|\boldsymbol{\Sigma}^*\|}} \asymp \frac{\sqrt{\operatorname{Tr}(\boldsymbol{\Sigma}^*)}}{\sqrt{\|\boldsymbol{\Sigma}^*\|}} = \sqrt{r(\boldsymbol{\Sigma}^*)}$. (iv) See [1], Theorem 4.1. □

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

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

$$\|\widetilde{\boldsymbol{P}}_{\mathcal{J}} - \boldsymbol{P}^*_{\mathcal{J}}\|_\infty \le 4\left(1 + \frac{2}{\pi}\frac{l^*_{\mathcal{J}}}{g^*_{\mathcal{J}}}\right)\frac{\|\widetilde{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*\|_\infty}{g^*_{\mathcal{J}}}.$$

Moreover, the following representation holds:

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

where

$$L_{\mathcal{J}}(\widetilde{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*) \stackrel{\text{def}}{=} \sum_{r\in\mathcal{J}}\sum_{s\notin\mathcal{J}} \frac{\boldsymbol{P}^*_r(\widetilde{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\boldsymbol{P}^*_s + \boldsymbol{P}^*_s(\widetilde{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\boldsymbol{P}^*_r}{\mu^*_r - \mu^*_s}$$

and

$$\|R_{\mathcal{J}}(\widetilde{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\|_\infty \le 15\left(1 + \frac{2}{\pi}\frac{l^*_{\mathcal{J}}}{g^*_{\mathcal{J}}}\right)\left(\frac{\|\widetilde{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*\|_\infty}{g^*_{\mathcal{J}}}\right)^2. \tag{A.2}$$

Proof. Apply Lemma 2 from [18]. □

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

The next Lemma from [23] provides upper bound for $\Delta-$ band of the squared norm of a Gaussian element.

Lemma A.3 (Gaussian anti-concentration)**.** *Let ξ be a Gaussian element in Hilbert space $\mathbb{H}$ with zero mean and covariance operator Γ. Then for arbitrary $\Delta > 0$ and any $\lambda > \lambda_1$*

$$\mathbb{P}(x < \|\xi\|^2 < x + \Delta) \le C_1 \Delta,$$

where

$$C_1 \overset{\text{def}}{=} \frac{e^{-x/(2\lambda)}}{\sqrt{\lambda_1 \lambda_2}} \prod_{j=3}^{\infty} (1 - \lambda_j/\lambda)^{-1/2}$$

and $\lambda_1 \ge \lambda_2 \ge \ldots$ are the eigenvalues of Γ. In particular, one has

$$\mathbb{P}(x < \|\xi\|^2 < x + \Delta) \le \frac{\Delta}{\sqrt{\lambda_1 \lambda_2}}.$$

Proof. See [23], Lemma 5.4. □

One more Lemma from [23] 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 ξ and η be Gaussian elements in Hilbert space $\mathbb{H}$ with zero mean and covariance operators $\boldsymbol{\Sigma}_\xi$ and $\boldsymbol{\Sigma}_\eta$, respectively. The following inequality holds*

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}(\|\xi\|^2 \ge x) - \mathbb{P}(\|\eta\|^2 \ge x) \right| \lesssim \left(\frac{1}{\sqrt{\lambda_{1\xi} \lambda_{2\xi}}} + \frac{1}{\sqrt{\lambda_{1\eta} \lambda_{2\eta}}} \right) \Diamond_0,$$

where $\lambda_{1\xi}, \lambda_{2\xi}$ are two largest eigenvalues of $\boldsymbol{\Sigma}_\xi$, $\lambda_{1\eta}, \lambda_{2\eta}$ are two largest eigenvalues of $\boldsymbol{\Sigma}_\eta$ and

$$\Diamond_0 \overset{\text{def}}{=} \|\boldsymbol{\Sigma}_\xi - \boldsymbol{\Sigma}_\eta\|_1 \le \|\boldsymbol{\Sigma}_\xi^{-\frac{1}{2}} \boldsymbol{\Sigma}_\eta \boldsymbol{\Sigma}_\xi^{-\frac{1}{2}} - \boldsymbol{I}\|_\infty \operatorname{Tr}(\boldsymbol{\Sigma}_\xi).$$

Proof. See [23], Lemma 5.1 and Corollary 5.2. □

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 $\widehat{\boldsymbol{P}}_{\mathcal{J}} - \boldsymbol{P}^*_{\mathcal{J}}$:

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

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

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

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

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

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

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

and $\widehat{\boldsymbol{P}}_{\mathcal{J}}$, $\boldsymbol{P}^*_{\mathcal{J}}$, $\sum_{r\in\mathcal{J}}\sum_{s\notin\mathcal{J}} \boldsymbol{P}^*_r(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\boldsymbol{P}^*_s$ 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

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

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

$$\|R_{\mathcal{J}}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\|_2 \le \sqrt{4m^*_{\mathcal{J}}} \cdot 15 \left(1 + \frac{2}{\pi}\frac{l^*_{\mathcal{J}}}{g^*_{\mathcal{J}}}\right) \frac{\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*\|_\infty^2}{{g^*_{\mathcal{J}}}^2}.$$

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

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

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

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

Putting this all together, we obtain

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

The concentration condition for the sample covariance (2.3) provides

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

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

Step 2 Following [23], we can choose $\{u_j^*\}_{j=1}^p$ as an orthonormal basis in $\mathbb{R}^p$ and represent $n\|L_{\mathcal{J}}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\|_2^2$ as

$$
\begin{aligned}
n\|L_{\mathcal{J}}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\|_2^2 &= n \left\| \sum_{r\in\mathcal{J}} \sum_{s\notin\mathcal{J}} \frac{\boldsymbol{P}_r^*(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\boldsymbol{P}_s^* + \boldsymbol{P}_s^*(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\boldsymbol{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{\boldsymbol{P}_r^*(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\boldsymbol{P}_s^* + \boldsymbol{P}_s^*(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\boldsymbol{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{\boldsymbol{P}_{r_1}^*(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\boldsymbol{P}_{s_1}^* + \boldsymbol{P}_{s_1}^*(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\boldsymbol{P}_{r_1}^*}{\mu_{r_1}^* - \mu_{s_1}^*} \, u_l^* \right) \\
&\qquad \times \left(u_k^{*\top} \sum_{r_2\in\mathcal{J}} \sum_{s_2\notin\mathcal{J}} \frac{\boldsymbol{P}_{r_2}^*(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\boldsymbol{P}_{s_2}^* + \boldsymbol{P}_{s_2}^*(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\boldsymbol{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{\boldsymbol{P}_{r_1}^*(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\boldsymbol{P}_{s_1}^* + \boldsymbol{P}_{s_1}^*(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\boldsymbol{P}_{r_1}^*}{\mu_{r_1}^* - \mu_{s_1}^*} \, u_l^* \right) \\
&\qquad\qquad \times \left(u_k^{*\top} \frac{\boldsymbol{P}_{r_2}^*(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\boldsymbol{P}_{s_2}^* + \boldsymbol{P}_{s_2}^*(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\boldsymbol{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

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

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

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

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

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

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 $\boldsymbol{P}^*_{\mathcal{J}} X_j$ and $(\boldsymbol{I}_p - \boldsymbol{P}^*_{\mathcal{J}})X_j$ 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 $l \in \Delta^*_s, l' \in \Delta^*_{s'}$, $s, s' \notin \mathcal{J}$:

$$\begin{aligned} Cov\left(S^{(j)}\right)_{\substack{(k,l)\\(k',l')}} &= \mathbb{E}\left(S^{(j)}(u^*_k, u^*_l) \cdot S^{(j)}(u^*_{k'}, u^*_{l'})\right) \\ &= \frac{2\,\mathbb{E}\left[({u^*_k}^\top X_j) \cdot ({u^*_l}^\top X_j) \cdot ({u^*_{k'}}^\top X_j) \cdot ({u^*_{l'}}^\top X_j)\right]}{(\mu^*_r - \mu^*_s)(\mu^*_{r'} - \mu^*_{s'})}. \end{aligned}$$

Now, the independence of $\boldsymbol{P}^*_{\mathcal{J}} X_j$ and $(\boldsymbol{I}_p - \boldsymbol{P}^*_{\mathcal{J}})X_j$ implies the independence of $(u^*_k, u^*_{k'})^\top \boldsymbol{P}^*_{\mathcal{J}} X_j$ and $(u^*_l, u^*_{l'})^\top (\boldsymbol{I}_p - \boldsymbol{P}^*_{\mathcal{J}})X_j$, 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

$$Cov\left(S^{(j)}\right)_{\substack{(k,l)\\(k',l')}} = \frac{2\,\mathbb{E}\left[({u^*_k}^\top X_j)({u^*_{k'}}^\top X_j)\right] \cdot \mathbb{E}\left[({u^*_l}^\top X_j)({u^*_{l'}}^\top X_j)\right]}{(\mu^*_r - \mu^*_s)(\mu^*_{r'} - \mu^*_{s'})}.$$

The observation that ${u^*_k}^\top \boldsymbol{\Sigma}^* u^*_{k'} = \mu^*_r \cdot \mathbb{1}\{k = k'\}$ and ${u^*_l}^\top \boldsymbol{\Sigma}^* u^*_{l'} = \mu^*_s \cdot \mathbb{1}\{l = l'\}$ establishes the fact that $Cov(S^{(j)}) = \Gamma^*_{\mathcal{J}}$.

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

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

Further, introducing the following auxiliary matrices

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

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

we have

$$\|\varGamma_{\mathcal{J}}^{*\,-1/2} S^{(j)}\|^2 = \sum_{k \in \mathcal{I}_{\mathcal{J}}} \sum_{l \notin \mathcal{I}_{\mathcal{J}}} \frac{(u_k^{*\top} X_j)^2}{\mu_r^*} \cdot \frac{(u_l^{*\top} X_j)^2}{\mu_s^*}$$

$$= \left\{ \sum_{k \in \mathcal{I}_{\mathcal{J}}} \frac{(u_k^{*\top} X_j)^2}{\mu_r^*} \right\} \cdot \left\{ \sum_{l \notin \mathcal{I}_{\mathcal{J}}} \frac{(u_l^{*\top} X_j)^2}{\mu_s^*} \right\} = \|\boldsymbol{U}_{\mathcal{J}}^* X_j\|^2 \, \|\boldsymbol{V}_{\mathcal{J}}^* X_j\|^2.$$

Then,

$$\begin{aligned}
\mathbb{E}\|\varGamma_{\mathcal{J}}^{*\,-1/2} S^{(j)}\|^3 &= \mathbb{E}\left(\|\boldsymbol{U}_{\mathcal{J}}^* X_j\|^3 \, \|\boldsymbol{V}_{\mathcal{J}}^* X_j\|^3\right) \\
&= \mathbb{E}\left(\|\boldsymbol{U}_{\mathcal{J}}^* \boldsymbol{\Sigma}^{*1/2} \boldsymbol{\Sigma}^{*\,-1/2} X_j\|^3 \, \|\boldsymbol{V}_{\mathcal{J}}^* \boldsymbol{\Sigma}^{*1/2} \boldsymbol{\Sigma}^{*\,-1/2} X_j\|^3\right) \\
&\le \mathbb{E}\left(\|\boldsymbol{U}_{\mathcal{J}}^* \boldsymbol{\Sigma}^{*1/2}\|_\infty^3 \|\boldsymbol{\Sigma}^{*\,-1/2} X_j\|^3 \, \|\boldsymbol{V}_{\mathcal{J}}^* \boldsymbol{\Sigma}^{*1/2}\|_\infty^3 \|\boldsymbol{\Sigma}^{*\,-1/2} X_j\|^3\right).
\end{aligned}$$

Observing that $\|\boldsymbol{U}_{\mathcal{J}}^* \boldsymbol{\Sigma}^{*1/2}\|_\infty = \|\boldsymbol{V}_{\mathcal{J}}^* \boldsymbol{\Sigma}^{*1/2}\|_\infty = 1$, we deduce

$$\mathbb{E}\|\varGamma_{\mathcal{J}}^{*\,-1/2} S^{(j)}\|^3 \le \mathbb{E}\|\boldsymbol{\Sigma}^{*\,-1/2} X_j\|^6 \,\le\, C.$$

Therefore, Theorem 1.1 from [2] yields

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}\left(\|S_{\mathcal{J}}\|^2 \le x\right) - \mathbb{P}(\|\xi\|^2 \le x) \right| \,\lesssim\, \frac{p^{1/4}}{\sqrt{n}},$$

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

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}\left(n\|L_{\mathcal{J}}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*)\|_2^2 \le x\right) - \mathbb{P}(\|\xi\|^2 \le x) \right| \,\lesssim\, \frac{p^{1/4}}{\sqrt{n}}.$$

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

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

Hence,

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

The first term in the right-hand side was bounded in Step 3 by $\frac{p^{1/4}}{\sqrt{n}}$. The second term is bounded by $\frac{\overline{\Delta}}{\sqrt{\lambda_1(\Gamma^*_{\mathcal{J}})\lambda_2(\Gamma^*_{\mathcal{J}})}}$ 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,

$$\begin{aligned}
&\sup_{x\in\mathbb{R}}\left\{\mathbb{P}\left(n\|\widehat{\boldsymbol{P}}_{\mathcal{J}} - \boldsymbol{P}^*_{\mathcal{J}}\|_2^2 \geq x\right) - \mathbb{P}\left(\|\xi\|^2 \geq x\right)\right\} \\
&\quad\lesssim \frac{p^{1/4}}{\sqrt{n}} + \frac{\overline{\Delta}}{\sqrt{\lambda_1(\Gamma^*_{\mathcal{J}})\lambda_2(\Gamma^*_{\mathcal{J}})}} + \frac{1}{n}.
\end{aligned}$$

Similarly, one can verify that

$$\begin{aligned}
&\sup_{x\in\mathbb{R}}\left\{\mathbb{P}\left(\|\xi\|^2 \geq x\right) - \mathbb{P}\left(n\|\widehat{\boldsymbol{P}}_{\mathcal{J}} - \boldsymbol{P}^*_{\mathcal{J}}\|_2^2 \geq x\right)\right\} \\
&\quad\lesssim \frac{p^{1/4}}{\sqrt{n}} + \frac{\overline{\Delta}}{\sqrt{\lambda_1(\Gamma^*_{\mathcal{J}})\lambda_2(\Gamma^*_{\mathcal{J}})}} + \frac{1}{n}.
\end{aligned}$$

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

$$\begin{aligned}
&\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(n\|\widehat{\boldsymbol{P}}_{\mathcal{J}} - \boldsymbol{P}^*_{\mathcal{J}}\|_2^2 \geq x\right) - \mathbb{P}\left(\|\xi\|^2 \geq x\right)\right| \\
&\quad\lesssim \frac{p^{1/4}}{\sqrt{n}} + \frac{\overline{\Delta}}{\sqrt{\lambda_1(\Gamma^*_{\mathcal{J}})\lambda_2(\Gamma^*_{\mathcal{J}})}} + \frac{1}{n}.
\end{aligned}$$

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.

[2] Bentkus, V. (2005). A Lyapunov-type bound in $\mathbb{R}^d$. *Theory Probab. Appl.*, **49**, 2, 311–323.

[3] Berthet, Q. and Rigollet, P. (2013). Optimal detection of sparse principal components in high dimension. *Ann. Statist.*, **41**, 4, 1780–1815.

[4] Bickel, P. J. and Kleijn, B. J. K. (2012). The semiparametric Bernstein - von Mises theorem. *Ann. Statist.*, **40**, 206-237.

[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.

[6] Cai, T. T., Ma, Z. and Wu, Y. (2013). Sparse PCA: optimal rates and adaptive estimation. *Ann. Statist.*, **41**, 6, 3074–3110.

[7] Castillo, I. and Nickl, R. (2013). Nonparametric Bernstein – von Mises theorems in Gaussian white noise. *Ann. Statist.*, **41**, 4, 1999–2028.

[8] Castillo, I. and Rousseau, J. (2015). A Bernstein – von Mises theorem for smooth functionals in semiparametric models. *Ann. Statist.*, **43**, 6, 2353–2383.

[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.

[10] Fan, J., Rigollet, P. and Wang, W. (2015). Estimation of functionals of sparse covariance matrices. *Ann. Statist.*, **43**, 6, 2706–2737.

[11] 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.

[12] Gao, C. and Zhou, H. H. (2015). Rate-optimal posterior contraction for sparse PCA. *Ann. Statist.*, **43**, 2, 785–818.

[13] Ghosh, J. K. and Ramamoorthi, R. V. (2003). Introduction to the non-asymptotic analysis of random matrices. In *Bayesian nonparametrics*, Springer Verlag, New York.

[14] Goodfellow, I., Bengio, Y. and Courville, A. (2016). *Deep learning*,

MIT Press.

[15] 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.

[16] JOHNSTONE, I. M. (2010). High dimensional statistical inference and random matrices. In *International Congress of Mathematicians*, **I**, 307–333, Eur. Math. Soc., Zurich

[17] JOHNSTONE, I. M. (2007). High dimensional Bernstein–von Mises: simple examples *Inst. Math. Stat. Collect.*, **6**, 87–98.

[18] KOLTCHINSKII, V. and LOUNICI, K. (2016). Asymptotics and concentration bounds for bilinear forms of spectral projectors of sample covariance. *Ann. Inst. H. Poincar Probab. Statist.*, **52**, 4, 1976–2013.

[19] KOLTCHINSKII, V. and LOUNICI, K. (2017). Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli*, **23**, 1, 110–133.

[20] KOLTCHINSKII, V. and LOUNICI, K. (2017). Normal approximation and concentration of spectral projectors of sample covariance. *Ann. Statist.*, **45**, 1, 121–157.

[21] LE CAM, L. and YANG, G. L. (1990). *Asymptotics in statistics: some basic concepts*. Springer, New York.

[22] 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.

[23] NAUMOV, A., SPOKOINY, V. and ULYANOV, V. (2013). Bootstrap confidence sets for spectral projectors of sample covariance. *arXiv:1703.00871*.

[24] RIGOLLET, P. (2015). *Lecture notes on High-dimensional statistics*.

[25] TROPP, J. (2012). User-Friendly Tail Bounds for Sums of Random Matrices. *Found. Comput. Math.*, **12**, 4, 389–434.

[26] VAN DER VAART, A. W. (2000). *Asymptotic statistics. Cambridge series in statistical and probabilistic mathematics 3*, Cambridge University Press, Cambridge.

[27] VERSHYNIN, R. (2016). Introduction to the non-asymptotic analysis of random matrices. In *Compressed sensing*, 210–268, Cambridge University Press, Cambridge.

[28] ZHOU, H. H. and GAO, C. (2016). Bernstein-von Mises theorems for functionals of the covariance matrix. *Electron. J. Statist.*, **10**, 2, 1751–1806.

IRTG 1792 Discussion Paper Series 2018

For a complete list of Discussion Papers published, please visit irtg1792.hu-berlin.de.

001 "Data Driven Value-at-Risk Forecasting using a SVR-GARCH-KDE Hybrid" by Marius Lux, Wolfgang Karl Härdle and Stefan Lessmann, January 2018.
002 "Nonparametric Variable Selection and Its Application to Additive Models" by Zheng-Hui Feng, Lu Lin, Ruo-Qing Zhu asnd Li-Xing Zhu, January 2018.
003 "Systemic Risk in Global Volatility Spillover Networks: Evidence from Option-implied Volatility Indices " by Zihui Yang and Yinggang Zhou, January 2018.
004 "Pricing Cryptocurrency options: the case of CRIX and Bitcoin" by Cathy YH Chen, Wolfgang Karl Härdle, Ai Jun Hou and Weining Wang, January 2018.
005 "Testing for bubbles in cryptocurrencies with time-varying volatility" by Christian M. Hafner, January 2018.
006 "A Note on Cryptocurrencies and Currency Competition" by Anna Almosova, January 2018.
007 "Knowing me, knowing you: inventor mobility and the formation of technology-oriented alliances" by Stefan Wagner and Martin C. Goossen, February 2018.
008 "A Monetary Model of Blockchain" by Anna Almosova, February 2018.
009 "Deregulated day-ahead electricity markets in Southeast Europe: Price forecasting and comparative structural analysis" by Antanina Hryshchuk, Stefan Lessmann, February 2018.
010 "How Sensitive are Tail-related Risk Measures in a Contamination Neighbourhood?" by Wolfgang Karl Härdle, Chengxiu Ling, February 2018.
011 "How to Measure a Performance of a Collaborative Research Centre" by Alona Zharova, Janine Tellinger-Rice, Wolfgang Karl Härdle, February 2018.
012 "Targeting customers for profit: An ensemble learning framework to support marketing decision making" by Stefan Lessmann, Kristof Coussement, Koen W. De Bock, Johannes Haupt, February 2018.
013 "Improving Crime Count Forecasts Using Twitter and Taxi Data" by Lara Vomfell, Wolfgang Karl Härdle, Stefan Lessmann, February 2018.
014 "Price Discovery on Bitcoin Markets" by Paolo Pagnottoni, Dirk G. Baur, Thomas Dimpfl, March 2018.
015 "Bitcoin is not the New Gold - A Comparison of Volatility, Correlation, and Portfolio Performance" by Tony Klein, Hien Pham Thu, Thomas Walther, March 2018.
016 "Time-varying Limit Order Book Networks" by Wolfgang Karl Härdle, Shi Chen, Chong Liang, Melanie Schienle, April 2018.
017 "Regularization Approach for NetworkModeling of German EnergyMarket" by Shi Chen, Wolfgang Karl Härdle, Brenda López Cabrera, May 2018.
018 "Adaptive Nonparametric Clustering" by Kirill Efimov, Larisa Adamyan, Vladimir Spokoiny, May 2018.
019 "Lasso, knockoff and Gaussian covariates: a comparison" by Laurie Davies, May 2018.

IRTG 1792, Spandauer Straße 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.

IRTG 1792 Discussion Paper Series 2018

For a complete list of Discussion Papers published, please visit irtg1792.hu-berlin.de.

020 "A Regime Shift Model with Nonparametric Switching Mechanism" by Haiqiang Chen, Yingxing Li, Ming Lin and Yanli Zhu, May 2018.
021 "LASSO-Driven Inference in Time and Space" by Victor Chernozhukov, Wolfgang K. Härdle, Chen Huang, Weining Wang, June 2018.
022 " Learning from Errors: The case of monetary and fiscal policy regimes" by Andreas Tryphonides, June 2018.
023 "Textual Sentiment, Option Characteristics, and Stock Return Predictability" by Cathy Yi-Hsuan Chen, Matthias R. Fengler, Wolfgang Karl Härdle, Yanchu Liu, June 2018.
024 "Bootstrap Confidence Sets For Spectral Projectors Of Sample Covariance" by A. Naumov, V. Spokoiny, V. Ulyanov, June 2018.
025 "Construction of Non-asymptotic Confidence Sets in 2 -Wasserstein Space" by Johannes Ebert, Vladimir Spokoiny, Alexandra Suvorikova, June 2018.
026 "Large ball probabilities, Gaussian comparison and anti-concentration" by Friedrich Götze, Alexey Naumov, Vladimir Spokoiny, Vladimir Ulyanov, June 2018.
027 "Bayesian inference for spectral projectors of covariance matrix" by Igor Silin, Vladimir Spokoiny, June 2018.

IRTG 1792, Spandauer Straße 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.