Probability Theory and Related Fields (2019) 174:1091–1132
https://doi.org/10.1007/s00440-018-0877-2

CrossMark

Bootstrap confidence sets for spectral projectors of sample covariance

Alexey Naumov[1,2] · **Vladimir Spokoiny**[1,2,3] · **Vladimir Ulyanov**[1,4]

Received: 4 July 2017 / Revised: 25 July 2018 / Published online: 26 October 2018

Abstract
Let $X_1, \dots, X_n$ be i.i.d. sample in $\mathbb{R}^p$ with zero mean and the covariance matrix $\mathbf{\Sigma}$. The problem of recovering the projector onto an eigenspace of $\mathbf{\Sigma}$ from these observations naturally arises in many applications. Recent technique from Koltchinskii and Lounici (Ann Stat 45(1):121–157, 2017) helps to study the asymptotic distribution of the distance in the Frobenius norm $\left\|\mathbf{P}_r - \widehat{\mathbf{P}}_r\right\|_2$ between the true projector $\mathbf{P}_r$ on the subspace of the rth eigenvalue and its empirical counterpart $\widehat{\mathbf{P}}_r$ in terms of the effective rank of $\mathbf{\Sigma}$. This paper offers a bootstrap procedure for building sharp confidence sets for the true projector $\mathbf{P}_r$ from the given data. This procedure does not rely on the asymptotic distribution of $\left\|\mathbf{P}_r - \widehat{\mathbf{P}}_r\right\|_2$ and its moments. It could be applied for small or moderate sample size n and large dimension p. The main result states the validity of the proposed procedure for Gaussian samples with an explicit error bound for the error of bootstrap approximation. This bound involves some new sharp results on Gaussian comparison and Gaussian anti-concentration in high-dimensional spaces. Numeric results confirm a good performance of the method in realistic examples.

Keywords Sample covariance matrices · Spectral projectors · Multiplier bootstrap · Gaussian comparison and anti-concentration inequalities · Effective rank

Mathematics Subject Classification 62E17 · 62G09 · 62H25

✉ Alexey Naumov
anaumov@hse.ru

Vladimir Spokoiny
spokoiny@wias-berlin.de

Vladimir Ulyanov
vulyanov@cs.msu.ru

1 National Research University Higher School of Economics, Moscow, Russia

2 IITP RAS, Moscow, Russia

3 Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Germany

4 Lomonosov Moscow State University, Moscow, Russia

1 Introduction

Let $X, X_1, \ldots, X_n$ be independent identically distributed (i.i.d.) random vectors taking values in $\mathbb{R}^p$ with mean zero and $\mathbb{E}\,\|X\|^2 < \infty$. Denote by $\mathbf{\Sigma}$ its $p \times p$ symmetric covariance matrix defined as

$$\mathbf{\Sigma} \stackrel{\text{def}}{=} \mathbb{E}\left(XX^{\mathsf{T}}\right).$$

We also consider the sample covariance matrix $\widehat{\mathbf{\Sigma}}$ of the observations $X_1, \ldots, X_n$ defined as the average of $X_j X_j^{\mathsf{T}}$: with $\mathbf{X} \stackrel{\text{def}}{=} [X_1, \ldots, X_n] \in \mathbb{R}^{p \times n}$,

$$\widehat{\mathbf{\Sigma}} \stackrel{\text{def}}{=} \frac{1}{n} \sum_{j=1}^{n} X_j X_j^{\mathsf{T}} = \frac{1}{n} \mathbf{X}\mathbf{X}^{\mathsf{T}}.$$

In statistical applications, the true covariance matrix $\mathbf{\Sigma}$ is typically unknown and one often uses the sample covariance matrix $\widehat{\mathbf{\Sigma}}$ as its estimator. The accuracy $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|$ of estimation of $\mathbf{\Sigma}$ by $\widehat{\mathbf{\Sigma}}$, in particular, for p much larger than n, has been actively studied in the literature. We refer to [20] for an overview of the recent results based on the matrix Bernstein inequality; see also [23]. A bound in term of the effective rank $\mathtt{r}(\mathbf{\Sigma}) \stackrel{\text{def}}{=} \operatorname{tr}(\mathbf{\Sigma})/\|\mathbf{\Sigma}\|$ can be found in [11,22]. This or similar bounds on the spectral norm $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|$ can be effectively applied to relate the eigenvalues of $\mathbf{\Sigma}$ and of $\widehat{\mathbf{\Sigma}}$ under the spectral gap condition. This paper focuses on a slightly different problem of recovering the spectral projectors on the eigen-subspaces of $\mathbf{\Sigma}$ for few significantly positive eigenvalues. Such tasks naturally arise in many dimensionality reduction techniques for large p. In particular, the famous principal component analysis (PCA) projects the vector X onto the subspace spanned by the eigenvectors for the first principal eigenvalues. A significant error in recovering these eigenvectors would lead to a substantial loss of information contained in the data by PCA projection. The popular sliced inverse regression (SIR) method under the assumption of elliptically contoured distributions for high dimensional or functional data leads back to recovering the eigen-subspace from a finite sample; see e.g. [13] and references therein. The use of dimension reduction methods in deep networking architecture is discussed in [2] among others. We also mention the use of dimension reduction technique in numerical integration with applications to finance and insurance; see e.g. [10]. Justification of the assumption of low effective dimension in financial problems can be found in [24] among many others.

Surprisingly, the problem of recovering the spectral projectors (eigenvectors or eigen-subspaces) of $\mathbf{\Sigma}$ from the sample $X_1, \ldots, X_n$ for significantly positive spectral values is much less studied than the problem of recovering the covariance matrix $\mathbf{\Sigma}$. Recently [12] established sharp non-asymptotic bounds on the Frobenius distance $\|\mathbf{P}_r - \widehat{\mathbf{P}}_r\|_2$ between the spectral projectors $\mathbf{P}_r$ and its empirical counterparts $\widehat{\mathbf{P}}_r$ for the rth eigenvalue, as well as its asymptotic behaviour for large samples. This enables to build some asymptotic confidence sets for the target projector $\mathbf{P}_r$ as a proper elliptic vicinity of $\widehat{\mathbf{P}}_r$. However, it is well known that such asymptotic results apply only for really large samples due to a slow convergence of the normalized U-statistics to the limiting normal law.

The aim of this paper is to develop and validate a bootstrap procedure for building a confidence set for $\mathbf{P}_r$ which applies for small or moderate samples and for large dimension p. Bootstrap methods belong nowadays to most popular ways for measuring the significance of a test or for building a confidence set. The existing theory based on the high order expansions of the related statistics states the bootstrap validity for various parametric methods. However, an extension to a non-classical situation with a limited sample size and/or high parameter dimension meets serious problems. We refer to series of works [4–6] which validate a bootstrap procedure for a test based on the maximum of huge number of statistics. Their study reveals a close relation between bootstrap validity results, Gaussian comparison and the so called "anti-concentration" bounds for rectangle sets. The paper [19] studies applicability of likelihood based statistics for finite samples and large parameter dimension under possible model misspecification. The important step in the proof of bootstrap validity was again based on the Gaussian comparison and anti-concentration bounds but now for spherical sets.

This paper makes a further step in understanding the range of applicability of a weighted bootstrap method in constructing a finite sample confidence set for a spectral projector. A proof of bootstrap validity in this setup is a challenging task. The spectral projector is a non-linear and non-regular function of the covariance matrix, which itself is a quadratic function of the underlying multivariate distribution. In situations with high-dimensional space and small or moderate sample size the classical asymptotic methods of bootstrap validation do not apply. It appears that even in a Gaussian case the proof of bootstrap consistency requires to develop new probabilistic tools for establishing some sharp bounds for Gaussian comparison and anti-concentration in high-dimensional or even infinite dimensional Hilbert spaces. One has also to account for randomness of the bootstrap measure and the related bootstrap quantiles. The main contributions of this paper are:

– we offer a new bootstrap procedure for recovering the spectral projector on a low dimensional eigen-subspace;
– under condition that $X_1, \ldots, X_n$ are i.i.d. Gaussian, we prove the validity of this procedure and provide an upper bound on the error of bootstrap approximation which is dimension free and holds even for the dimension p which is exponential in the sample size and for small or moderate samples;
– a numerical study illustrates a very good performance of the proposed procedure in realistic setups;
– we establish new sharp results on Gaussian comparison and anti-concentration which are heavily used for proving the validity of the bootstrap procedure but they are probably of independent interest; see Lemmas 2 and 3.

The paper is organized as follows. The next section contains the description of the bootstrap procedure and the main results about its validity. Numerical results of Sect. 3 illustrate the performance of the procedure for finite samples. Main proofs are collected in Sect. 4. The results on Gaussian comparison and Gaussian anti-concentration see in Sect. 5. Appendix A gathers some auxilary statements and existing results.

Throughout the paper the following notation are used. We write $a \lesssim b$ ($a \gtrsim b$) if there exists some absolute constant C such that $a \le Cb$ ($a \ge Cb$ resp.). Similarly, $a \asymp b$ means that there exist c, C such that $c\,a \le b \le C\,a$. $\mathbb{R}$ (resp. $\mathbb{C}$) denotes the set of all real (resp. complex) numbers. For a self-adjoint operator $\mathbf{A}$ with eigenvalues $\lambda_k(\mathbf{A}), k \ge 1$, let us denote by $\|\mathbf{A}\|$ and $\|\mathbf{A}\|_s, s \ge 1$ the operator and Schatten s norm by $\|\mathbf{A}\| \stackrel{\mathrm{def}}{=} \sup_{\|x\|=1} \|\mathbf{A}x\|$ and $\|\mathbf{A}\|_s^s \stackrel{\mathrm{def}}{=} \sum_{k=1}^{\infty} |\lambda_k(\mathbf{A})|^s$. In particular, $\|\mathbf{A}\|_2$ is the Hilbert–Schmidt (Frobenius) norm of $\mathbf{A}$. For a self-adjoint positive operator $\mathbf{A}$ its effective rank is given by $\mathtt{r}(\mathbf{A}) \stackrel{\mathrm{def}}{=} \operatorname{tr} \mathbf{A}/\|\mathbf{A}\|$. We assume that all random variables are defined on common probability space $(\Omega, \mathfrak{F}, \mathbb{P})$. Let $\mathbb{E}$ be the mathematical expectation with respect to $\mathbb{P}$. We also denote by $\mathfrak{B}(\mathbb{H})$ the Borel σ-algebra. For r.v. X and Y we write $X \stackrel{\mathrm{d}}{=} Y$ if they are equally distributed.

2 Procedure and main results

This section presents the bootstrap procedure for building a confidence set for the true projector $\mathbf{P}_r$ and states the result about its validity.

2.1 Setup and problem

Let $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_p$ be the eigenvalues of $\boldsymbol{\Sigma}$ and $\mathbf{u}_j, j = 1, \ldots, p$, be the corresponding orthonormal eigenvectors. Matrix $\boldsymbol{\Sigma}$ has the following spectral decomposition

$$\boldsymbol{\Sigma} = \sum_{j=1}^{p} \sigma_j \mathbf{u}_j \mathbf{u}_j^{\mathsf{T}}. \tag{1}$$

Let $\mu_1 > \mu_2 > \cdots > \mu_q > 0$ with some $1 \le q \le p$, be strictly distinct eigenvalues of $\boldsymbol{\Sigma}$ and $\mathbf{P}_r, r = 1, \ldots, q$, be the corresponding spectral projectors (orthogonal projectors in $\mathbb{R}^p$). Denote $m_r \stackrel{\mathrm{def}}{=} \operatorname{Rank}(\mathbf{P}_r)$. We may rewrite (1) in terms of distinct eigenvalues and corresponding spectral projectors, namely

$$\boldsymbol{\Sigma} = \sum_{r=1}^{q} \mu_r \mathbf{P}_r.$$

Denote by $\Delta_r \stackrel{\mathrm{def}}{=} \{j : \sigma_j = \mu_r\}$. Then $|\Delta_r| = m_r$. Define $g_r \stackrel{\mathrm{def}}{=} \mu_r - \mu_{r+1} > 0$ for $r \ge 1$. Let $\overline{g}_r \stackrel{\mathrm{def}}{=} \min(g_{r-1}, g_r)$ for $r \ge 2$ and $\overline{g}_1 \stackrel{\mathrm{def}}{=} g_1$. The quantity $\overline{g}_r$ is the r-th spectral gap of the eigenvalue μ_r.

Consider now the sample covariance matrix $\widehat{\boldsymbol{\Sigma}}$. Similarly to (1), it can be represented as

$$\widehat{\boldsymbol{\Sigma}} = \sum_{j=1}^{p} \widehat{\sigma}_j \widehat{\mathbf{u}}_j \widehat{\mathbf{u}}_j^{\mathsf{T}},$$

where $\widehat{\sigma}_1 \geq \widehat{\sigma}_2 \geq \cdots \geq \widehat{\sigma}_p, \widehat{\mathbf{u}}_1, \ldots, \widehat{\mathbf{u}}_p$ are the eigenvalues and the corresponding eigenvectors of $\widehat{\boldsymbol{\Sigma}}$. Following [12] we may define clusters of eigenvalues $\widehat{\sigma}_j,\ j \in \Delta_r$. Let $\widehat{\mathbf{E}} \stackrel{\text{def}}{=} \widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}$. One can show that

$$\inf_{j \notin \Delta_r} |\widehat{\sigma}_j - \mu_r| \geq \overline{g}_r - \|\widehat{\mathbf{E}}\|, \quad \sup_{j \in \Delta_r} |\widehat{\sigma}_j - \mu_r| \leq \|\widehat{\mathbf{E}}\|.$$

Assume that $\|\widehat{\mathbf{E}}\| \leq \overline{g}_r/2$. Then all $\widehat{\sigma}_j,\ j \in \Delta_r$ may be covered by an interval

$$\left(\mu_r - \|\widehat{\mathbf{E}}\|, \mu_r + \|\widehat{\mathbf{E}}\|\right) \subset \left(\mu_r - \overline{g}_r/2, \mu_r + \overline{g}_r/2\right).$$

The rest of the eigenvalues of $\widehat{\boldsymbol{\Sigma}}$ are outside of the interval

$$\left(\mu_r - \left(\overline{g}_r - \|\widehat{\mathbf{E}}\|\right), \mu_r + \left(\overline{g}_r - \|\widehat{\mathbf{E}}\|\right)\right) \supset \left[\mu_r - \overline{g}_r/2, \mu_r + \overline{g}_r/2\right].$$

Let $\|\widehat{\mathbf{E}}\| < \frac{1}{4} \min_{1 \leq s \leq r} \overline{g}_s =: \overline{\delta}_r$. The set $\{\widehat{\sigma}_j,\ j \in \cup_{s=1}^{r} \Delta_s\}$ consists of r clusters, the diameter of each cluster being strictly smaller than $2\overline{\delta}_r$ and the distance between any two clusters being larger than $2\overline{\delta}_r$. We denote by $\widehat{\mathbf{P}}_r$ the projector on subspace spanned by the direct sum of $\widehat{\mathbf{u}}_j,\ j \in \Delta_r$.

It follows from [12, Lemma 5] that $\|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2$ has nearly weighted χ^2 distribution; see also Theorem 4 below. Therefore, after centering and standardization, it can be approximated by the standard normal distribution under some conditions on the spectrum of $\boldsymbol{\Sigma}$:

$$\mathcal{L}\left(\frac{\|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2 - \mathbb{E}\,\|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2}{\operatorname{Var}^{1/2}\left(\|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2\right)}\right) \approx \mathcal{N}(0, 1), \tag{2}$$

see [12, Theorem 6]. This allows to build an asymptotic elliptic confidence set for $\mathbf{P}_r$ in the form

$$\left\{\mathbf{P}_r : \frac{\|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2 - \mathbb{E}\,\|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2}{\operatorname{Var}^{1/2}\left(\|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2\right)} \leq z_\alpha\right\},$$

where z_α is a proper quantile of the standard normal law. However, there are at least two drawbacks of this approach. First, weak approximation in (2) can be very poor in some cases, especially if the effective rank of $\boldsymbol{\Sigma}$ is not large. Figure 1 illustrates this issue on the artificial Example 1 from Sect. 3 below. Second, this construction requires to know or to estimate the values $\mathbb{E}\,\|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2$ and $\operatorname{Var}\left(\|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2\right)$ which depend on the unknown covariance operator $\boldsymbol{\Sigma}$. A partial solution of this problem is discussed in [12]. It involves splitting the sample into three subsamples, and pilot estimation of the mean and the variance of $\|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2$. The approach only applies in some special cases, in particular, if the covariance matrix has a nearly spike structure. The present paper proposes another procedure which

- does not rely on the asymptotic distribution of the error $\|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2$,
- does not require to know the moments of $\|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2$,

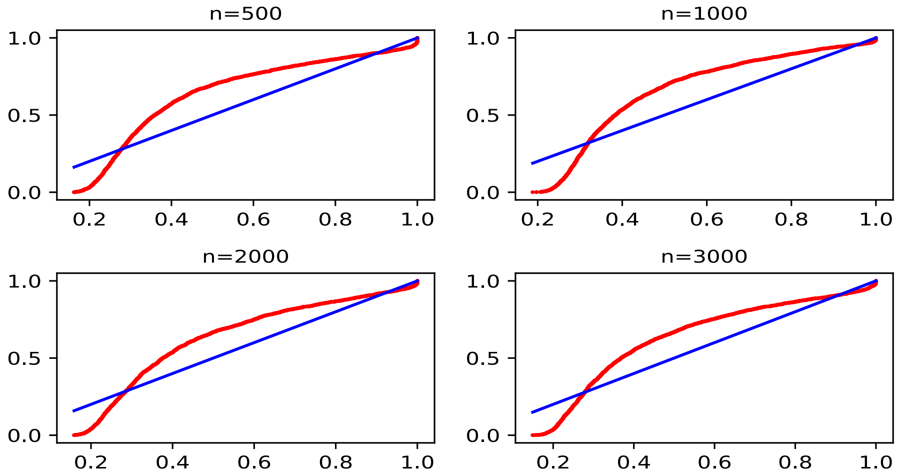

Fig. 1 PP-plot between the distribution from (2) and the standard Gaussian

- does not involve any data splitting,
- provides an explicit error bound for the bootstrap approximation in the case when sample comes from the Gaussian distribution.

The procedure is based on the resampling idea which allows to estimate directly the quantiles

$$\gamma_\alpha \stackrel{\text{def}}{=} \inf\left\{\gamma > 0\colon\ \mathbb{P}\left(n\|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2 > \gamma\right) \le \alpha\right\} \tag{3}$$

without estimating the covariance matrix $\boldsymbol{\Sigma}$. The introduced bootstrap procedure is described in the next section.

2.2 Bootstrap procedure

We introduce the following weighted version of $\widehat{\boldsymbol{\Sigma}}$:

$$\boldsymbol{\Sigma}^\circ \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n} w_i X_i X_i^\top, \tag{4}$$

where $w_1, \ldots, w_n$ are i.i.d. random variables, independent of $\mathbf{X} = (X_1, \ldots, X_n)$, with $\mathbb{E}\, w_1 = 1$, $\operatorname{Var} w_1 = 1$. A typical example used in this paper is to apply i.i.d. Gaussian weights $w_i \sim \mathcal{N}(1, 1)$. Denote by $\mathbb{P}^\circ(\cdot) \stackrel{\text{def}}{=} \mathbb{P}(\cdot \,|\, \mathbf{X})$ and $\mathbb{E}^\circ$ the corresponding conditional probability and expectation. It is obvious that

$$\mathbb{E}^\circ\, \boldsymbol{\Sigma}^\circ = \widehat{\boldsymbol{\Sigma}}. \tag{5}$$

In what follows we will often refer to "**X**-world" and "bootstrap world". In the **X**-world the sample **X** is random opposite to the bootstrap world, where **X** is fixed, but $w_1, \ldots, w_n$ are random. Then, Eq. (5) implies that in the bootstrap world we know precisely the expectation of $\mathbf{\Sigma}^\circ$ opposite to the **X**-world, where $\mathbf{\Sigma}$ is unknown. Similarly to (1) we may write

$$\mathbf{\Sigma}^\circ = \sum_{j=1}^{p} \sigma_j^\circ \mathbf{u}_j^\circ \mathbf{u}_j^{\circ\mathsf{T}}.$$

Let us denote by $\mathbf{P}_r^\circ$ a projector on the subspace spanned by the direct sum of $\mathbf{u}_j^\circ$, $j \in \Delta_r$. For a given α we define the quantile γ_α° as

$$\gamma_\alpha^\circ \overset{\text{def}}{=} \min\left\{\gamma > 0 \colon \mathbb{P}^\circ\left(n\|\mathbf{P}_r^\circ - \widehat{\mathbf{P}}_r\|_2^2 > \gamma\right) \le \alpha\right\}. \tag{6}$$

Note that this value γ_α° is defined w.r.t. the bootstrap measure, therefore, it depends on the data **X**. This bootstrap critical value γ_α° is applied in the **X**-world to build the confidence set

$$\mathcal{E}(\alpha) \overset{\text{def}}{=} \left\{\mathbf{P} \colon n\|\mathbf{P} - \widehat{\mathbf{P}}_r\|_2^2 \le \gamma_\alpha^\circ\right\}.$$

The main result given in the next section justifies this construction and evaluate the coverage probability of the true projector $\mathbf{P}_r$ by this set. It states that

$$\mathbb{P}\left(\mathbf{P}_r \notin \mathcal{E}(\alpha)\right) = \mathbb{P}\left(n\|\mathbf{P}_r - \widehat{\mathbf{P}}_r\|_2^2 > \gamma_\alpha^\circ\right) \approx \alpha.$$

2.3 Main results: bootstrap validity

To formulate the main result of this paper we introduce additional notation. Define the following block-matrix

$$\Gamma_r \overset{\text{def}}{=} \begin{pmatrix} \Gamma_{r1} & \mathbf{O} & \ldots & \mathbf{O} \\ \mathbf{O} & \Gamma_{r2} & \mathbf{O}\ldots & \mathbf{O} \\ \ldots & & & \\ \mathbf{O} & \ldots & \mathbf{O} & \Gamma_{rq} \end{pmatrix}, \tag{7}$$

where $\Gamma_{rs}, s \ne r$ are diagonal matrices of order $m_r m_s \times m_r m_s$ with values $2\mu_r\mu_s/(\mu_r - \mu_s)^2$ on the main diagonal. Let $\lambda_1(\Gamma_r) \ge \lambda_2(\Gamma_r) \ge \cdots$ be the eigenvalues of Γ_r.

The available bounds on the distance between the covariance matrix and its empirical counterpart claim that the eigenvalues of $\mathbf{\Sigma}$ can be recovered with accuracy $O(1/\sqrt{n})$; see e.g. [11,20,22,23]. Therefore, the part of the spectrum of $\mathbf{\Sigma}$ below a threshold of order $O(1/\sqrt{n})$ cannot be estimated. The same applies to the matrix Γ_r. Introduce the corresponding value $\mathfrak{m}$:

$$\lambda_{\mathfrak{m}}(\Gamma_r) \ge \operatorname{tr} \Gamma_r \left(\sqrt{\frac{\log n}{n}} + \sqrt{\frac{\log p}{n}} \right) > \lambda_{\mathfrak{m}+1}(\Gamma_r). \tag{8}$$

Denote by $\Pi_{\mathfrak{m}}$ a projector on the subspace spanned by the eigenvectors of Γ_r corresponding to its largest $\mathfrak{m}$ eigenvalues. Now we state our main result.

Theorem 1 *Let observations $X, X_1, \dots, X_n$ be i.i.d. Gaussian random vectors in $\mathbb{R}^p$ with $\mathbb{E}\, X = 0$ and $\mathbb{E}\, XX^T = \mathbf{\Sigma}$. Let γ_α° be defined by (6) for any $\alpha : 0 < \alpha < 1$, with i.i.d. Gaussian random weights $w_i \sim \mathcal{N}(1, 1)$ for $i = 1, \dots, n$. Then the following bound is fulfilled*

$$\left| \alpha - \mathbb{P}\left(n \|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2 > \gamma_\alpha^\circ \right) \right| \lesssim \diamondsuit, \tag{9}$$

where

$$\begin{aligned} \diamondsuit \overset{\text{def}}{=} & \frac{\mathfrak{m} \operatorname{tr} \Gamma_r}{\sqrt{\lambda_1(\Gamma_r)\lambda_2(\Gamma_r)}} \left(\sqrt{\frac{\log n}{n}} + \sqrt{\frac{\log p}{n}} \right) + \frac{\operatorname{tr}(\mathbf{I} - \Pi_{\mathfrak{m}})\Gamma_r}{\sqrt{\lambda_1(\Gamma_r)\lambda_2(\Gamma_r)}} \\ & + \frac{m_r \operatorname{tr}^3 \mathbf{\Sigma}}{\overline{g}_r^3 \sqrt{\lambda_1(\Gamma_r)\lambda_2(\Gamma_r)}} \left(\sqrt{\frac{\log^3 n}{n}} + \sqrt{\frac{\log^3 p}{n}} \right) \end{aligned} \tag{10}$$

and $\mathfrak{m}$ is defined by (8).

Remark 1 To replace the Gaussian assumption for $X_1, \dots, X_n$ by more wide setup, for example, sub-Gaussian or -exponential assumption is a challenge for a future research. Among other difficulties it will require, in particularly, a version of the central limit theorem for non i.i.d. random elements in high dimensional space with precise dependence of the rate of convergence on the dimension p. Some partial results are available, see e.g. [3], but they provide dependence on p, which is not sufficient for our purposes.

Remark 2 We choose w_j, $j = 1, \dots, n$ to be Gaussian r.v. variables. This choice may lead to the situation that the matrix (4) may have negative eigenvalues. This is not critical problem since if r-th gap $\overline{g}_r > 0$ we get that the r-th largest eigenvalue of $\mathbf{\Sigma}^\circ$ is positive and concentrated around μ_r with high probability.

Remark 3 The result (9) implicitly assumes that the error term $\diamondsuit$ is small. If $\diamondsuit \ge 1$ then (9) is meaningless. In particular, this implies that $p \lesssim \exp(n^{1/3})$.

Remark 4 The error term $\diamondsuit$ can be described in terms of $\mathbf{\Sigma}$. It is easy to check that for all r

$$\operatorname{tr} \Gamma_r \lesssim m_r \frac{\mu_r \operatorname{tr} \mathbf{\Sigma}}{\overline{g}_r^2} \le m_r \frac{\|\mathbf{\Sigma}\|^2 \mathtt{r}(\mathbf{\Sigma})}{\overline{g}_r^2}.$$

Let us consider, for example, the case $r = 2$ and $m_1 = m_2 = 1$. Introduce a function $f(x) = 2x\mu_2/(x - \mu_2)^2$ at the points $x = \mu_s$, $s \ne 2$. It is straightforward to check that the maximum of $f(x)$ is achieved at $x = \mu_1$ or μ_3. Moreover, assume that the largest values of $f(x)$ are $f(\mu_1)$ and $f(\mu_3)$. Then

$$\diamondsuit \lesssim \frac{\mathfrak{m}\ \mathrm{tr}\,\mathbf{\Sigma}}{\overline{g}_2}\sqrt{\frac{\mu_1}{\mu_3}}\left(\sqrt{\frac{\log n}{n}} + \sqrt{\frac{\log p}{n}}\right) + \sqrt{\frac{\mu_1}{\mu_3}}\frac{\mathrm{tr}(\mathbf{I} - \varPi_{\mathfrak{m}})\mathbf{\Sigma}}{\overline{g}_2} + \frac{\mathrm{tr}^3\,\mathbf{\Sigma}}{\overline{g}_2^2\,\mu_2}\sqrt{\frac{\mu_1}{\mu_3}}\left(\sqrt{\frac{\log^3 n}{n}} + \sqrt{\frac{\log^3 p}{n}}\right).$$

Although an analytic expression for the value γ_α° is not available, one can evaluate it from numerical simulations by generating a large number M of independent samples $\{w_1, \dots, w_n\}$ and computing from them the empirical distribution function of $n\|\mathbf{P}_r^\circ - \widehat{\mathbf{P}}_r\|_2^2$. In fact, standard arguments, see e.g. [18, Sect. 5.1], in combination with Theorem 5 suggest that the accuracy of Monte-Carlo approximation is of order $M^{-1/2}$. Theorem 1 justifies the use of this value γ_α° in place of γ_α defined in (3) provided that the error $\diamondsuit$ is sufficiently small.

3 Numerical results

This section illustrates the performance of the bootstrap procedure by means of few artificial examples. Namely, we check how well is the bootstrap approximation of the true quantiles. We use PP-plots to compare the distributions of $n\|\mathbf{P}_1^\circ - \widehat{\mathbf{P}}_1\|_2^2$ and $n\|\widehat{\mathbf{P}}_1 - \mathbf{P}_1\|_2^2$.

First we describe our setup. Let n be a sample size. We consider the different values of n, namely $n = 100, 300, 500, 1000, 2000, 3000$. Let $X_1, \dots, X_n$ have the normal distribution in $\mathbb{R}^p$, with zero mean and covariance matrix $\mathbf{\Sigma}$. The value of p and the choice of $\mathbf{\Sigma}$ will be described below. The distribution of $n\|\widehat{\mathbf{P}}_1 - \mathbf{P}_1\|_2^2$ is evaluated by using $M = 3000$ Monte-Carlo samples from the normal distribution with zero mean and covariance $\mathbf{\Sigma}$. The bootstrap distribution for a given realization X is evaluated by $M = 3000$ Monte-Carlo samples of bootstrap weights $\{w_1, \dots, w_n\}$. Since this distribution is random and depends on X, we finally use the median from 50 realizations of X for each quantile.

Example 1 In the first example we consider the following parameters:

- $p = 500$,
- $\mu_1 = 36, \mu_2 = 30, \mu_3 = 25, \mu_4 = 19$ and all other eigenvalues $\mu_s, s = 5, \dots, 500$ are uniformly distributed in $[1, 5]$.

Here we get $\overline{g}_1 = 6$ and $\mathtt{r}(\mathbf{\Sigma}) = 51.79$. Figure 2 shows the corresponding PP-plots for the empirical distribution of $n\|\widehat{\mathbf{P}}_1 - \mathbf{P}_1\|_2^2$ against its bootstrap counterpart. Table 1 shows the coverage probabilities of the quantiles estimated using the bootstrap.

Example 2 The second example parameters are:

- $p = 100$,

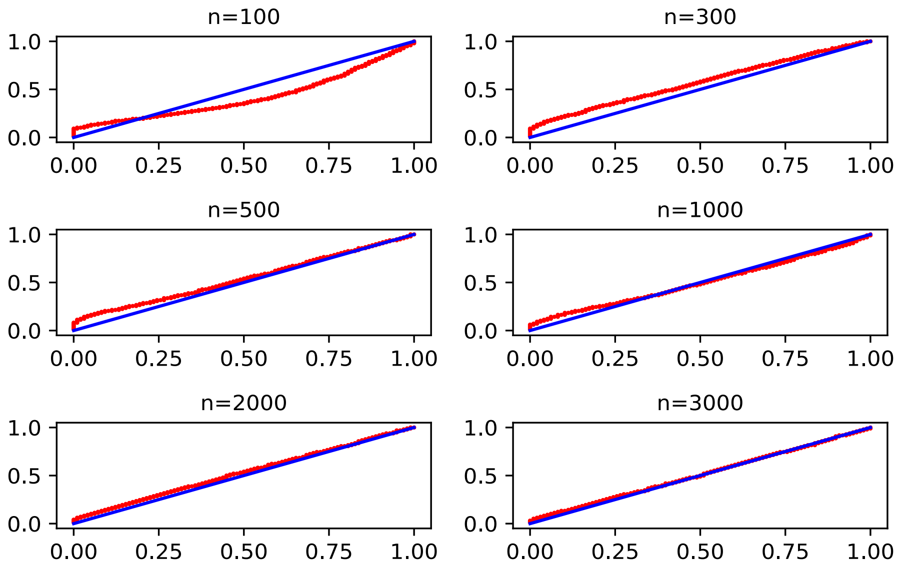

Fig. 2 PP-plot of the bootstrap procedure for Example 1

Table 1 Coverage probabilities for Example 1

n	Confidence levels					
	0.99	0.95	0.90	0.85	0.80	0.75
100	0.997	0.986	0.954	0.924	0.889	0.850
	0.004	0.026	0.052	0.074	0.091	0.104
300	0.992	0.937	0.873	0.812	0.754	0.692
	0.026	0.093	0.165	0.207	0.236	0.271
500	0.988	0.962	0.902	0.846	0.788	0.623
	0.054	0.139	0.227	0.264	0.323	0.174
1000	0.992	0.974	0.943	0.890	0.841	0.783
	0.021	0.062	0.114	0.066	0.153	0.170
2000	0.988	0.954	0.891	0.843	0.795	0.741
	0.021	0.059	0.081	0.098	0.126	0.142
3000	0.994	0.961	0.908	0.864	0.815	0.763
	0.016	0.053	0.073	0.081	0.092	0.101

For each n the first line corresponds to the median value of the coverage probability and the second line corresponds to the interquartile range

– $\mu_6, \ldots, \mu_{100}$ are distributed according to Marchenko–Pastur's density with the support on [0.71, 1.34], see [14],
– all other eigenvalues are $\mu_1 = 25.698$, $\mu_2 = 15.7688$, $\mu_3 = 10.0907$, $\mu_4 = 5.9214$, $\mu_5 = 3.4321$.

Here $\overline{g}_1 = 9.93$ and $\mathtt{r}(\boldsymbol{\Sigma}) = 6.12$. PP plots are presented on Fig. 3 and the coverage probabilities are collected in Table 2.

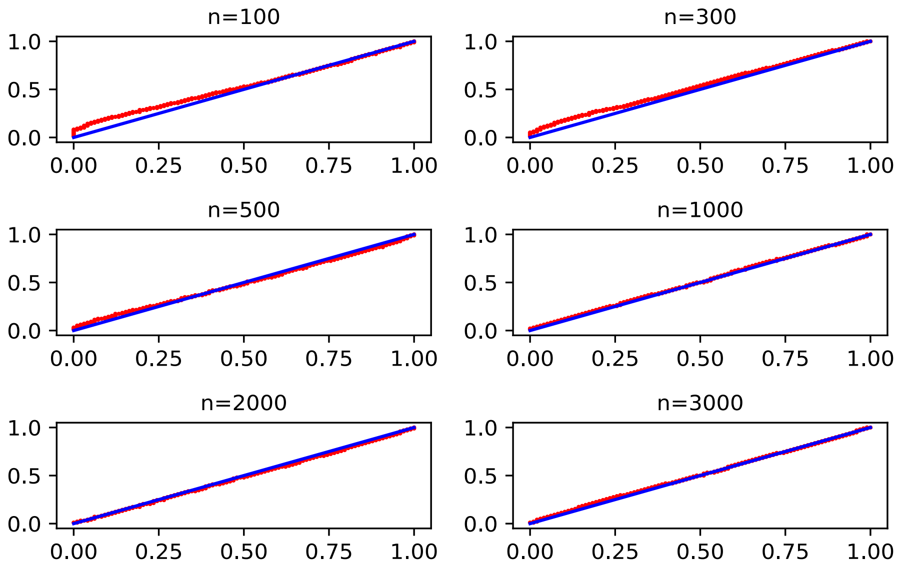

Fig. 3 PP-plot of the bootstrap procedure for Example 1

Table 2 Coverage probabilities for Example 2

n	Confidence levels					
	0.99	0.95	0.90	0.85	0.80	0.75
100	0.992	0.961	0.918	0.876	0.825	0.768
	0.027	0.091	0.146	0.197	0.231	0.257
300	0.988	0.942	0.886	0.832	0.784	0.735
	0.020	0.062	0.094	0.118	0.139	0.153
500	0.995	0.966	0.925	0.876	0.822	0.771
	0.013	0.035	0.072	0.104	0.120	0.122
1000	0.989	0.957	0.906	0.848	0.795	0.743
	0.012	0.038	0.062	0.086	0.093	0.098
2000	0.993	0.958	0.913	0.869	0.819	0.775
	0.011	0.028	0.053	0.065	0.076	0.083
3000	0.988	0.952	0.902	0.853	0.803	0.752
	0.006	0.021	0.047	0.053	0.062	0.070

For each n the first line corresponds to the median value of the coverage probability and the second line corresponds to the interquartile range

Example 3 The third example has the same setup as the previous one except $\mu_1 = \mu_2 = 25.698$. In that case $\mathbf{P}_1 = \mathbf{u}_1\mathbf{u}_1^\top + \mathbf{u}_2\mathbf{u}_2^\top$. Here $\overline{g}_1 = 9, 93$ and $\mathtt{r}(\mathbf{\Sigma}) = 6.51$. The result is on Fig. 4 and Table 3.

In all three examples we observe the same patterns. The bootstrap procedure mimics well the most of the underlying distribution of $n\|\widehat{\mathbf{P}}_1 - \mathbf{P}_1\|_2^2$. For a really small sample size $n = 100$, there is a problem of approximating the high quantiles, while for n

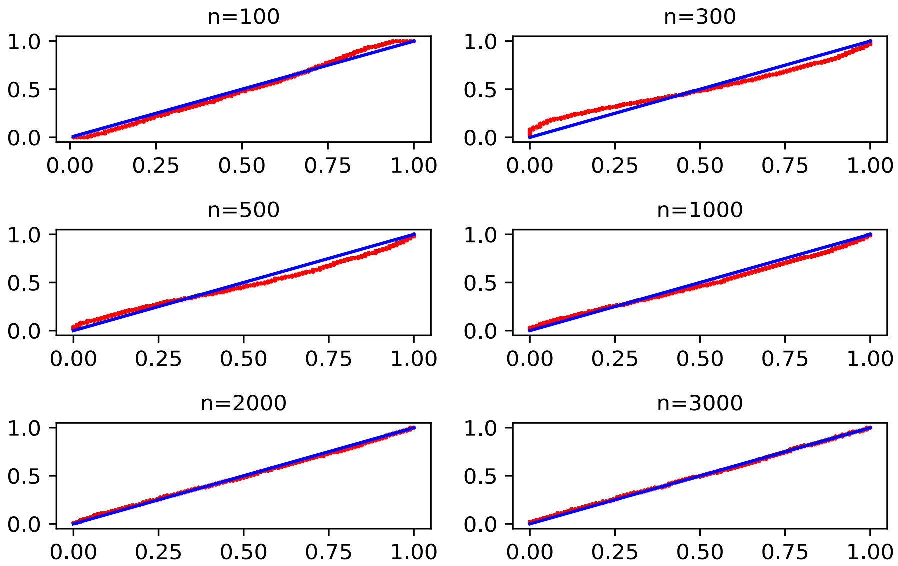

Fig. 4 PP-plot of the bootstrap procedure for Example 3

Table 3 Coverage probabilities for Example 3

n	Confidence levels					
	0.99	0.95	0.90	0.85	0.80	0.75
100	0.999	0.991	0.972	0.939	0.906	0.858
	0.003	0.015	0.035	0.059	0.089	0.114
300	0.999	0.981	0.950	0.919	0.873	0.816
	0.003	0.023	0.053	0.075	0.114	0.144
500	0.998	0.977	0.947	0.914	0.867	0.820
	0.005	0.020	0.041	0.057	0.087	0.106
1000	0.992	0.971	0.937	0.895	0.855	0.796
	0.010	0.031	0.061	0.073	0.105	0.129
2000	0.990	0.958	0.911	0.866	0.824	0.774
	0.006	0.016	0.024	0.034	0.052	0.055
3000	0.989	0.950	0.897	0.852	0.795	0.749
	0.004	0.022	0.034	0.049	0.061	0.064

For each n the first line corresponds to the median value of the coverage probability and the second line corresponds to the interquartile range

of order 300 or larger, it works surprisingly well in different setups and for different dimensions p including the case with $p > n$.

4 Proofs

This section presents the proof of the main theorem as well as some further statements. Before going to the proof we outline its main steps. In Sect. 4.2 we show that

$$\mathbf{X}\text{-world:} \qquad \mathcal{L}\big(n\|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2\big) \approx \mathcal{L}\big(\|\xi\|^2\big), \quad \xi \sim \mathcal{N}(0, \Gamma_r),$$

where Γ_r defined in (7). Further, in Sect. 4.3 we demonstrate that the similar relation holds in the bootstrap world, namely

$$\text{Bootstrap world:} \quad \mathcal{L}\big(n\|\mathbf{P}_r^\circ - \widehat{\mathbf{P}}_r\|_2^2\big) \approx \mathcal{L}\big(\|\xi^\circ\|^2\big), \quad \xi^\circ \sim \mathcal{N}(0, \Gamma_r^\circ),$$

where Γ_r° is defined below in (24). To compare ξ and ξ° we apply Gaussian comparison inequality, Lemma 2. The details are in Sect. 4.4. All necessary concentration inequalities for sample covariances in the $\mathbf{X}$-world and bootstrap world may be found in the Appendix A and Sect. 4.1 respectively.

In all our results, we implicitly assume

$$\frac{\operatorname{tr}\mathbf{\Sigma}}{\overline{g}_r}\left(\sqrt{\frac{\log n}{n}} + \sqrt{\frac{\log p}{n}}\right) \lesssim 1. \tag{11}$$

Otherwise, the main result becomes trivial.

4.1 Some concentration inequalities

Theorem 2 *Assume that the conditions of Theorem 1 hold. Then the following inequality holds with $\mathbb{P}$-probability at least $1 - \frac{1}{n}$*

$$\mathbb{P}^\circ\left(\|\mathbf{\Sigma}^\circ - \widehat{\mathbf{\Sigma}}\| \lesssim \operatorname{tr}\mathbf{\Sigma}\left[\sqrt{\frac{\log n}{n}} \bigvee \sqrt{\frac{\log p}{n}}\right]\right) \geq 1 - \frac{1}{n}.$$

Proof We prove this theorem applying a combination of matrix concentration inequalities. For simplicity we denote $\xi_i \overset{\text{def}}{=} w_i - 1$ and $\mathbf{A}_i \overset{\text{def}}{=} X_i X_i^\top$ for all $i = 1, \ldots, n$. It is easy to see that $\mathbf{\Sigma}^\circ - \widehat{\mathbf{\Sigma}}$ is a Gaussian matrix series. Indeed,

$$\mathbf{\Sigma}^\circ - \widehat{\mathbf{\Sigma}} = \frac{1}{n}\sum_{k=1}^{n} \xi_k \mathbf{A}_k.$$

Hence, to estimate $\|\mathbf{\Sigma}^\circ - \widehat{\mathbf{\Sigma}}\|$ we may directly apply Lemma 8, which gives us that

$$\mathbb{P}^\circ\left(\left\|\frac{1}{n}\sum_{k=1}^{n} \xi_k \mathbf{A}_k\right\| \lesssim \sigma\,\frac{\sqrt{\log n} + \sqrt{\log p}}{n}\right) \geq 1 - \frac{1}{n}, \tag{12}$$

where $\sigma^2 \overset{\text{def}}{=} \|\sum_{k=1}^n \mathbf{A}_k^2\|$. To finish the proof it remains to estimate with high $\mathbb{P}$-probability the variance parameter σ. This may be done by using the Bernstein matrix concentration inequality, Lemma 9. To proceed we need to check all assumptions of Lemma 9. Applying Lemma 5 with $p = 2$ we may show that

$$\mathbb{E}\,\|\mathbf{A}_k\|^2 \leq \mathbb{E}\,\|X_k\|^4 \lesssim \operatorname{tr}^2\mathbf{\Sigma}. \tag{13}$$

Moreover, application of the same lemma with $p \asymp \log^2 n$ gives us that

$$\mathbb{P}(\|\mathbf{A}_k\|^2 \lesssim \operatorname{tr}^2(\boldsymbol{\Sigma}) \log^2 n) \geq 1 - n^{-1}. \tag{14}$$

Let $\mathcal{E}_1 \stackrel{\text{def}}{=} \left\{\max_{1 \le k \le n} \|\mathbf{A}_k^2 - \mathbb{E}\,\mathbf{A}_k^2\| \lesssim \operatorname{tr}^2(\boldsymbol{\Sigma}) \log^2 n\right\}$. It follows from (13) and (14) and the union bound that $\mathbb{P}(\mathcal{E}_1^c) \le n^{-1}$. Introduce the variance parameter

$$\widetilde{\sigma}^2 \stackrel{\text{def}}{=} \Big\| \sum_{k=1}^{n} \mathbb{E}(\mathbf{A}_k^2 - \mathbb{E}\,\mathbf{A}_k^2)^2 \Big\|.$$

Analogously to (13) one may show that $\widetilde{\sigma}^2 \le n\,\mathbb{E}\,\|\mathbf{A}_1\|^4 \le n\,\mathbb{E}\,\|X\|^8 \lesssim n \operatorname{tr}^4 \boldsymbol{\Sigma}$. Applying Lemma 9 we get

$$\begin{aligned}
&\mathbb{P}\left(\Big\| \sum_{k=1}^{n} \left(\mathbf{A}_k^2 - \mathbb{E}\,\mathbf{A}_k^2\right) \Big\| \gtrsim \sqrt{n} \operatorname{tr}^2(\boldsymbol{\Sigma}) \left(\sqrt{\log n} + \sqrt{\log p}\right)\right) \\
&\quad \le \mathbb{P}\left(\Big\| \sum_{k=1}^{n} \left(\mathbf{A}_k^2 - \mathbb{E}\,\mathbf{A}_k^2\right) \Big\| \gtrsim \sqrt{n} \operatorname{tr}^2(\boldsymbol{\Sigma}) \left(\sqrt{\log n} + \sqrt{\log p}\right), \mathcal{E}_1\right) + \frac{1}{n} \le \frac{2}{n}.
\end{aligned} \tag{15}$$

Combining (13) and (15) we may write that with $\mathbb{P}$-probability at least $1 - \frac{1}{n}$

$$\sigma^2 \lesssim n \operatorname{tr}^2 \boldsymbol{\Sigma} + \sqrt{n}(\sqrt{\log n} + \sqrt{\log p}) \operatorname{tr}^2 \boldsymbol{\Sigma} \lesssim n \operatorname{tr}^2 \boldsymbol{\Sigma}.$$

Substituting the last inequality to (12) we finish the proof of this theorem. □

Let us introduce the following notations

$$\mathbf{E}^\circ \stackrel{\text{def}}{=} \boldsymbol{\Sigma}^\circ - \boldsymbol{\Sigma}, \quad \widehat{\mathbf{E}}^\circ \stackrel{\text{def}}{=} \boldsymbol{\Sigma}^\circ - \widehat{\boldsymbol{\Sigma}}, \quad \widehat{\mathbf{E}} = \widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}.$$

Denote $L_r(\widehat{\mathbf{E}}^\circ) \stackrel{\text{def}}{=} \mathbf{P}_r\left(\boldsymbol{\Sigma}^\circ - \widehat{\boldsymbol{\Sigma}}\right)\mathbf{C}_r + \mathbf{C}_r\left(\boldsymbol{\Sigma}^\circ - \widehat{\boldsymbol{\Sigma}}\right)\mathbf{P}_r$ where

$$\mathbf{C}_r \stackrel{\text{def}}{=} \sum_{s \ne r} \frac{1}{\mu_r - \mu_s} \mathbf{P}_s.$$

Theorem 3 (Concentration results in the bootstrap world) *Assume that the conditions of Theorem 1 hold. Then the following bound holds with $\mathbb{P}$-probability at least $1 - \frac{1}{n}$*

$$\mathbb{P}^\circ\left(\left|\|\mathbf{P}_r^\circ - \widehat{\mathbf{P}}_r\|_2^2 - \|L_r(\widehat{\mathbf{E}}^\circ)\|_2^2\right| \lesssim \Delta\right) \geq 1 - n^{-1},$$

where

$$\Delta \stackrel{\text{def}}{=} m_r \frac{\operatorname{tr}^3 \boldsymbol{\Sigma}}{\overline{g}_r^3} \left[\frac{\log n}{n} \bigvee \frac{\log p}{n}\right]^{3/2}.$$

Proof Applying Lemma 4 we may write

$$\mathbf{P}_r^\circ - \widehat{\mathbf{P}}_r = \mathbf{P}_r^\circ - \mathbf{P}_r - \left(\widehat{\mathbf{P}}_r - \mathbf{P}_r\right) = L_r\left(\mathbf{E}^\circ\right) - L_r\left(\widehat{\mathbf{E}}\right) + S_r\left(\mathbf{E}^\circ\right) + S_r\left(\widehat{\mathbf{E}}\right),$$

where

$$\|S_r\left(\widehat{\mathbf{E}}\right)\| \le 14\left(\|\widehat{\mathbf{E}}\|/\overline{g}_r\right)^2, \quad \|S_r\left(\mathbf{E}^\circ\right)\| \le 14\left(\|\mathbf{E}^\circ\|/\overline{g}_r\right)^2. \tag{16}$$

It is easy to see that $L_r\left(\mathbf{E}^\circ\right) - L_r(\mathbf{E}) = L_r(\widehat{\mathbf{E}}^\circ)$. We denote $S_r(\widehat{\mathbf{E}}^\circ) \stackrel{\text{def}}{=} S_r\left(\mathbf{E}^\circ\right) + S_r\left(\widehat{\mathbf{E}}\right)$. Then the difference $\|\mathbf{P}_r^\circ - \widehat{\mathbf{P}}_r\|_2^2 - \|L_r(\widehat{\mathbf{E}}^\circ)\|_2^2$ may be rewritten in the following way:

$$\|\mathbf{P}_r^\circ - \widehat{\mathbf{P}}_r\|_2^2 - \|L_r(\widehat{\mathbf{E}}^\circ)\|_2^2 = 2\left\langle L_r\left(\widehat{\mathbf{E}}^\circ\right), S_r\left(\widehat{\mathbf{E}}^\circ\right)\right\rangle + \|S_r(\widehat{\mathbf{E}}^\circ)\|_2^2.$$

Applying the Cauchy–Schwarz inequality we get

$$\left|\|\mathbf{P}_r^\circ - \widehat{\mathbf{P}}_r\|_2^2 - \|L_r(\widehat{\mathbf{E}}^\circ)\|_2^2\right| \le 2\|L_r\left(\widehat{\mathbf{E}}^\circ\right)\|_2 \|S_r(\widehat{\mathbf{E}}^\circ)\|_2 + \|S_r(\widehat{\mathbf{E}}^\circ)\|_2^2. \tag{17}$$

It follows from (16)

$$\|S_r\left(\mathbf{E}^\circ\right)\| \lesssim \left(\|\widehat{\mathbf{E}}\|/\overline{g}_r\right)^2 + \left(\|\mathbf{E}^\circ\|/\overline{g}_r\right)^2 \lesssim \left(\|\widehat{\mathbf{E}}\|/\overline{g}_r\right)^2 + \left(\|\widehat{\mathbf{E}}^\circ\|/\overline{g}_r\right)^2.$$

From Theorems 2 and 6, and condition (11) we may assume that without loss of generality that the following inequality holds

$$\max\left\{\|\mathbf{E}^\circ\|, \|\widehat{\mathbf{E}}\|\right\} \le \overline{g}_r/2.$$

This fact guarantees that $\operatorname{Rank}\mathbf{P}_r^\circ = \operatorname{Rank}\widehat{\mathbf{P}}_r = \operatorname{Rank}\mathbf{P}_r = m_r$. Applying (16) and Theorems 2, 6 we get that with $\mathbb{P}$-probability at least $1 - \frac{1}{n}$:

$$\mathbb{P}^\circ\left(\|S_r(\widehat{\mathbf{E}}^\circ)\|_2 \lesssim \sqrt{m_r}\frac{\operatorname{tr}^2\boldsymbol{\Sigma}}{\overline{g}_r^2}\left[\frac{\log n}{n} \bigvee \frac{\log p}{n}\right]\right) \ge 1 - \frac{1}{n}. \tag{18}$$

It remains to estimate $\|L_r(\widehat{\mathbf{E}}^\circ)\|_2$. We proceed similarly to the proof of Theorem 2. We get that with $\mathbb{P}$-probability at least $1 - \frac{1}{n}$:

$$\mathbb{P}^\circ\left(\|L_r(\widehat{\mathbf{E}}^\circ)\|_2 \lesssim \sqrt{m_r}\frac{\operatorname{tr}\boldsymbol{\Sigma}}{\overline{g}_r}\left[\sqrt{\frac{\log n}{n}} \bigvee \sqrt{\frac{\log p}{n}}\right]\right) \ge 1 - \frac{1}{n}.$$

From the last bound and inequalities (17) and (18) we conclude that with $\mathbb{P}$-probability at least $1 - \frac{1}{n}$:

$$\mathbb{P}^\circ\left(\|\mathbf{P}_r^\circ - \widehat{\mathbf{P}}_r\|_2^2 - \|L_r(\widehat{\mathbf{E}}^\circ)\|_2^2 \lesssim \Delta_1\right) \ge 1 - \frac{1}{n},$$

where

$$\Delta_1^* \stackrel{\text{def}}{=} m_r \frac{\operatorname{tr}^3 \mathbf{\Sigma}}{\overline{g}_r^3} \left[\frac{\log n}{n} \bigvee \frac{\log p}{n}\right] \left[\sqrt{\frac{\log n}{n}} \bigvee \sqrt{\frac{\log p}{n}}\right] + m_r \frac{\operatorname{tr}^4 \mathbf{\Sigma}}{\overline{g}_r^4} \left[\frac{\log n}{n} \bigvee \frac{\log p}{n}\right]^2 .$$

Applying condition (11) we get that

$$\Delta_1^* \le m_r \frac{\operatorname{tr}^3 \mathbf{\Sigma})}{\overline{g}_r^3} \left[\frac{\log n}{n} \bigvee \frac{\log p}{n}\right]^{3/2} .$$

□

4.2 Approximation in the X-world

The main result of this section is the following theorem.

Theorem 4 *Assume that the conditions of Theorem 1 hold. Let $\xi \sim \mathcal{N}(0, \Gamma_r)$, where Γ_r is defined in (7). Then for all $x : x > 0$ the following bounds hold*

$$\mathbb{P}\left(n\|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2 > x\right) \le \mathbb{P}\left(\|\xi\|^2 \ge x_-\right) + \diamondsuit_1,$$
$$\mathbb{P}\left(n\|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2 > x\right) \ge \mathbb{P}\left(\|\xi\|^2 \ge x_+\right) - \diamondsuit_1,$$

where $x_\pm \stackrel{\text{def}}{=} x \pm \diamondsuit_2$ and

$$\diamondsuit_1 \stackrel{\text{def}}{\asymp} m_r^{1/2} \frac{\operatorname{tr} \Gamma_r}{\sqrt{\lambda_1(\Gamma_r)\lambda_2(\Gamma_r)}} \left(\sqrt{\frac{\log n}{n}} + \sqrt{\frac{\log p}{n}}\right),$$
$$\diamondsuit_2 \stackrel{\text{def}}{\asymp} m_r \frac{\operatorname{tr}^3 \mathbf{\Sigma}}{\overline{g}_r^3} \sqrt{\frac{\log^3 n}{n}}.$$

Proof of Theorem 4 Let us fix an arbitrary $x \ge 0$. Without loss of generality we may assume that $\diamondsuit_1 \lesssim 1$. Otherwise the claim is trivial. This fact implies that the condition (46) holds.

Let us rewrite $\widehat{\mathbf{P}}_r - \mathbf{P}_r$ as follows

$$n\|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2 = 2n\|\mathbf{P}_r\widehat{\mathbf{E}}\mathbf{C}_r\|_2^2 + n\|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2 - 2n\|\mathbf{P}_r\widehat{\mathbf{E}}\mathbf{C}_r\|_2^2.$$

Theorem 7 implies that with probability at least $1 - \frac{1}{n}$

$$\left|n\|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2 - 2n\|\mathbf{P}_r\widehat{\mathbf{E}}\mathbf{C}_r\|_2^2\right| \le \Delta_1^* \stackrel{\text{def}}{\asymp} m_r \frac{\operatorname{tr}^3 \mathbf{\Sigma}}{\overline{g}_r^3} \sqrt{\frac{\log^3 n}{n}}.$$

Hence, we may write down the following two-sided inequalities

$$\mathbb{P}\left(2n\|\mathbf{P}_r\widehat{\mathbf{E}}\mathbf{C}_r\|_2^2 \geq x + \Delta_1^*\right) - \frac{1}{n} \leq \mathbb{P}\left(n\|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2 > x\right)$$
$$\leq \mathbb{P}\left(2n\|\mathbf{P}_r\widehat{\mathbf{E}}\mathbf{C}_r\|_2^2 \geq x - \Delta_1^*\right) + \frac{1}{n}.$$

For simplicity we denote $x_\pm \overset{\text{def}}{=} x \pm \Delta_1^*$. Without loss of generality, we consider the case of the upper bound only, i.e. we set $z \overset{\text{def}}{=} x_-$. Similar calculations are valid for x_+.

Let $\{\mathbf{e}_j\}_{j=1}^p$ be an arbitrary orthonormal basis in $\mathbb{R}^p$. Denote by $\Psi_{kl} \overset{\text{def}}{=} \mathbf{e}_k\mathbf{e}_l^\top, l, k = 1, \ldots, p$. Then $\{\Psi_{kl}\}_{k,l=1}^p$ is the orthonormal basis in $\mathbb{R}^{p\times p}$ with respect to the scalar product given by $\langle \mathbf{A}, \mathbf{B}\rangle \overset{\text{def}}{=} \operatorname{tr}\mathbf{A}\mathbf{B}^\top, \mathbf{A}, \mathbf{B} \in \mathbb{R}^{p\times p}$. By Parseval's identity

$$2n\|\mathbf{P}_r\widehat{\mathbf{E}}\mathbf{C}_r\|_2^2 = 2n\sum_{l,k=1}^p \langle \mathbf{P}_r\mathbf{E}\mathbf{C}_r, \Psi_{kl}\rangle^2 = 2n\sum_{l,k=1}^p \langle \mathbf{P}_r\mathbf{E}\mathbf{C}_r\mathbf{e}_l, \mathbf{e}_k\rangle^2.$$

We may set $\mathbf{e}_j \overset{\text{def}}{=} \mathbf{u}_j$. Taking into account definition of $\mathbf{P}_r$ and $\mathbf{C}_r$ the last equation may be rewritten as follows

$$2n\|\mathbf{P}_r\widehat{\mathbf{E}}\mathbf{C}_r\|_2^2 = 2n\sum_{k\in\Delta_r}\sum_{s\neq r}\sum_{l\in\Delta_s}\langle \mathbf{P}_r\widehat{\mathbf{E}}\mathbf{C}_r\mathbf{u}_l, \mathbf{u}_k\rangle^2.$$

Let us fix arbitrary $\mathbf{u}_k, k \in \Delta_r$ and $\mathbf{u}_l, l \in \Delta_s, s \neq r$. For simplicity we denote them by $\mathbf{u}$ and $\mathbf{v}$ respectively. Then

$$S(\mathbf{u}, \mathbf{v}) \overset{\text{def}}{=} \sqrt{2n}\langle \mathbf{P}_r\widehat{\mathbf{E}}\mathbf{C}_r\mathbf{v}, \mathbf{u}\rangle = \sqrt{\frac{2}{n}}\sum_{i=1}^n \langle \mathbf{u}, \mathbf{P}_r X_i\rangle\langle \mathbf{C}_r X_i, \mathbf{v}\rangle.$$

It is easy to see that $\langle \mathbf{u}, \mathbf{P}_r X_i\rangle$ is a Gaussian r.v. with zero mean and variance $\mathbb{E}\langle \mathbf{u}, \mathbf{P}_r X_i\rangle^2 = \langle \mathbf{u}, \mathbf{P}_r\boldsymbol{\Sigma}\mathbf{P}_r\mathbf{u}\rangle = \mu_r$. Then $\langle \mathbf{u}, \mathbf{P}_r X_i\rangle \overset{\text{d}}{=} \sqrt{\mu_r}\overline{\eta}_{\mathbf{u},i}$, where $\overline{\eta}_{\mathbf{u},i}, i = 1, \ldots, n$ are i.i.d. $\mathcal{N}(0, 1)$. Similarly we may write that $\langle \mathbf{C}_r X_i, \mathbf{v}\rangle \overset{\text{d}}{=} \sqrt{\frac{\mu_s}{(\mu_r-\mu_s)^2}}\eta_{\mathbf{v},i}$, where $\eta_{\mathbf{u},i}, i = 1, \ldots, n$ are i.i.d. $\mathcal{N}(0, 1)$. Hence, we obtain

$$S(\mathbf{u}, \mathbf{v}) \overset{\text{d}}{=} \frac{1}{\sqrt{n}}\sum_{i=1}^n \sqrt{\frac{2\mu_s\mu_r}{(\mu_r - \mu_s)^2}}\overline{\eta}_{\mathbf{u},i}\eta_{\mathbf{v},i}.$$

Let us fix another pair $\tilde{\mathbf{u}}, \tilde{\mathbf{v}}$ and investigate the covariance

$$\Gamma((\mathbf{u}, \mathbf{v}), (\tilde{\mathbf{u}}, \tilde{\mathbf{v}})) \overset{\text{def}}{=} \operatorname{Cov}(S(\mathbf{u}, \mathbf{v}), S(\tilde{\mathbf{u}}, \tilde{\mathbf{v}})).$$

It is straightforward to check that

$$\Gamma((\mathbf{u},\mathbf{v}),(\tilde{\mathbf{u}},\tilde{\mathbf{v}})) = 2\langle \boldsymbol{\Sigma}\mathbf{P}_r\mathbf{u}, \mathbf{P}_r\tilde{\mathbf{u}}\rangle\langle \boldsymbol{\Sigma}\mathbf{C}_r\mathbf{v}, \mathbf{C}_r\tilde{\mathbf{v}}\rangle = 2\Gamma_1(\mathbf{u},\tilde{\mathbf{u}})\Gamma_2(\mathbf{v},\tilde{\mathbf{v}}),$$

where for simplicity we denoted

$$\Gamma_1(\mathbf{u},\tilde{\mathbf{u}}) \overset{\text{def}}{=} \langle \boldsymbol{\Sigma}\mathbf{P}_r\mathbf{u}, \mathbf{P}_r\tilde{\mathbf{u}}\rangle, \quad \Gamma_2(\mathbf{v},\tilde{\mathbf{v}}) \overset{\text{def}}{=} \langle \mathbf{C}_r\boldsymbol{\Sigma}\mathbf{C}_r\mathbf{v}, \tilde{\mathbf{v}}\rangle.$$

Moreover, direct calculations yield that

$$\Gamma_1(\mathbf{u},\tilde{\mathbf{u}}) = \begin{cases} 0, & \text{if} \quad \mathbf{u} \neq \tilde{\mathbf{u}}, \\ \mu_r, & \text{if} \quad \mathbf{u} = \tilde{\mathbf{u}}, \end{cases} \quad \Gamma_2(\mathbf{v},\tilde{\mathbf{v}}) = \begin{cases} 0, & \text{if} \quad \mathbf{v} \neq \tilde{\mathbf{v}}, \\ \frac{\mu_s}{(\mu_r-\mu_s)^2}, & \text{if} \quad \mathbf{v} = \tilde{\mathbf{v}}. \end{cases} \tag{19}$$

We may think of $S_r \overset{\text{def}}{=} (S(\mathbf{u}_k,\mathbf{u}_l), k \in \Delta_r, s \neq r, l \in \Delta_s)$ as a random vector in the dimension $d \overset{\text{def}}{=} m_r \sum_{s\neq r} m_s$ (it is easy to see that $d \asymp p$) with the following covariance matrix Γ_r [compare with (7)]:

$$\Gamma_r \overset{\text{def}}{=} \begin{pmatrix} \Gamma_{r1} & \mathbf{O} & \dots & \mathbf{O} \\ \mathbf{O} & \Gamma_{r2} & \mathbf{O}\dots & \mathbf{O} \\ \dots & & & \\ \mathbf{O} & \dots & \mathbf{O} & \Gamma_{rq} \end{pmatrix},$$

where $\Gamma_{rs} = \frac{2\mu_r\mu_s}{(\mu_r-\mu_s)^2}\mathbf{I}_{m_r m_s}, s \neq r$, are diagonal matrices of order $m_r m_s \times m_r m_s$ with values $\frac{2\mu_r\mu_s}{(\mu_r-\mu_s)^2}$ on the main diagonal. In these notations we may write

$$\mathbb{P}\left(2n\|\mathbf{P}_r\widehat{\mathbf{E}}\mathbf{C}_r\|_2^2 \geq z\right) = \mathbb{P}\left(\|S_r\|^2 \geq z\right).$$

Since $\mathbf{P}_r X_i$ and $\mathbf{C}_r X_i$ are independent Gaussian vectors it is straightforward to check that the conditional distribution of S_r with respect to $\overline{Y} = (\mathbf{P}_r X_1, \dots, \mathbf{P}_r X_n)$ is Gaussian with zero mean and covariance matrix $\Gamma_r^{\overline{Y}} = \frac{1}{n}\sum_{i=1}^n \Gamma_{ri}^{\overline{Y}}$, where

$$\Gamma_{ri}^{\overline{Y}} \overset{\text{def}}{=} \left[\Gamma_{ri}^{\overline{Y}}\left(\left(\mathbf{u}_{k_1},\mathbf{u}_{l_1}\right),\left(\mathbf{u}_{k_2},\mathbf{u}_{l_2}\right)\right), k_1,k_2 \in \Delta_r, l_1 \in \Delta_{s_1}, l_2 \in \Delta_{s_2}, s_1,s_2 \neq r\right]$$

and

$$\Gamma_{ri}^{\overline{Y}}\left(\left(\mathbf{u}_{k_1},\mathbf{u}_{l_1}\right),\left(\mathbf{u}_{k_2},\mathbf{u}_{l_2}\right)\right) = 2\mu_r\overline{\eta}_{\mathbf{u}_{k_1},i}\overline{\eta}_{\mathbf{u}_{k_2},i}\Gamma_2\left(\mathbf{u}_{l_1},\mathbf{u}_{l_2}\right)$$

Due to (19) we conclude that $\Gamma_r^{\overline{Y}}((\mathbf{u}_{k_1},\mathbf{u}_{l_1}),(\mathbf{u}_{k_2},\mathbf{u}_{l_2})) = 0$ if $l_1 \neq l_2$. Let $\mathbb{P}(\cdot \,|\, \overline{Y})$ be the conditional probability w.r.t. $\overline{Y}$. We show that $\mathbb{P}(\|S_r\|^2 \geq z \,|\, \overline{Y})$ may be approximated by $\mathbb{P}(\|\xi\|^2 \geq z)$, where $\xi \sim \mathcal{N}(0,\Gamma_r)$. For this aim we may apply Corollary 1. Hence, we need to check that $\|\Gamma_r^{-\frac{1}{2}}\Gamma_r^{\overline{Y}}\Gamma_r^{-\frac{1}{2}} - \mathbf{I}\|$ is small. Let us denote by $\mathbb{P}_{\overline{Y}}(\cdot)$

the distribution of $\overline{Y}$, i.e. $\mathbb{P}_{\overline{Y}}(A) = \mathbb{P}(\overline{Y} \in A)$, $A \in \mathfrak{B}(\mathbb{R}^p)$. We also introduce the following event

$$\mathcal{E}_1(\delta) \stackrel{\text{def}}{=} \left\{\overline{Y}\colon \|\varGamma_r^{-\frac{1}{2}} \varGamma_r^{\overline{Y}} \varGamma_r^{-\frac{1}{2}} - \mathbf{I}\| \le \delta\right\}, \quad \delta > 0.$$

If $\max_{1\le k\le n} \|\varGamma_r^{-\frac{1}{2}} \varGamma_{rk}^{\overline{Y}} \varGamma_r^{-\frac{1}{2}} - \mathbf{I}\| \le R$ for some $R = R(n, \varGamma_r)$, then it follows from Lemma 9 that

$$\mathbb{P}_{\overline{Y}}\left(\left\| \frac{1}{n} \sum_{i=1}^{n} \big(\varGamma_r^{-\frac{1}{2}} \varGamma_{rk}^{\overline{Y}} \varGamma_r^{-\frac{1}{2}} - \mathbf{I}\big) \right\| \lesssim \frac{s}{n}\right) \ge 1 - d \cdot \exp\left(-\frac{s^2}{\sigma^2}\right), \tag{20}$$

provided that $Rs \lesssim \sigma^2$, where

$$\sigma^2 \stackrel{\text{def}}{=} \left\| \sum_{i=1}^{n} \mathbb{E}_{\overline{Y}} \left(\varGamma_r^{-\frac{1}{2}} \varGamma_{ri}^{\overline{Y}} \varGamma_r^{-\frac{1}{2}} - \mathbf{I}\right)^2 \right\|.$$

It is straightforward to check that $\varGamma_r^{-\frac{1}{2}} \varGamma_{ri}^{\overline{Y}} \varGamma_r^{-\frac{1}{2}}$ is a block-diagonal matrix. The number of blocks equals $\sum_{s\ne r} \varDelta_s$, all of them are the same and have the following structure

$$\begin{pmatrix} \overline{\eta}^2_{\mathbf{u}_{k_1},i} & \overline{\eta}_{\mathbf{u}_{k_1},i}\overline{\eta}_{\mathbf{u}_{k_2},i} & \cdots & \overline{\eta}_{\mathbf{u}_{k_1},i}\overline{\eta}_{\mathbf{u}_{k_r},i} \\ \overline{\eta}_{\mathbf{u}_{k_1},i}\overline{\eta}_{\mathbf{u}_{k_2},i} & \overline{\eta}^2_{\mathbf{u}_{k_2},i} & \cdots & \overline{\eta}_{\mathbf{u}_{k_2},i}\overline{\eta}_{\mathbf{u}_{k_r},i} \\ & & \cdots & \\ \overline{\eta}_{\mathbf{u}_{k_1},i}\overline{\eta}_{\mathbf{u}_{k_r},i} & \overline{\eta}_{\mathbf{u}_{k_2},i}\overline{\eta}_{\mathbf{u}_{k_r},i} & \cdots & \overline{\eta}^2_{\mathbf{u}_{k_r},i} \end{pmatrix},$$

where $k_j \in \varDelta_r$, $j = 1, \ldots, r$. Hence,

$$\begin{aligned} \|\varGamma_r^{-\frac{1}{2}} \varGamma_{ri}^{\overline{Y}} \varGamma_r^{-\frac{1}{2}} - \mathbf{I}\| \le \|\varGamma_r^{-\frac{1}{2}} \varGamma_{ri}^{\overline{Y}} \varGamma_r^{-\frac{1}{2}}\|_2 + 1 &= \left(\sum_{k_1,k_2\in\varDelta_r} \overline{\eta}^2_{\mathbf{u}_{k_1},i}\overline{\eta}^2_{\mathbf{u}_{k_2},i}\right)^{\frac{1}{2}} + 1 \\ &= \sum_{k\in\varDelta_r} \overline{\eta}^2_{\mathbf{u}_k,i} + 1. \end{aligned}$$

Applying Lemma 5 we obtain that

$$\mathbb{P}_{\overline{Y}}\left(\|\varGamma_r^{-\frac{1}{2}} \varGamma_{ri}^{\overline{Y}} \varGamma_r^{-\frac{1}{2}} - \mathbf{I}\| \lesssim m_r \log n\right) \ge 1 - \frac{1}{n}.$$

Moreover, let $R \stackrel{\text{def}}{\asymp} m_r \log n$. Denote $\mathcal{E}_2 \stackrel{\text{def}}{=} \left\{\max_{1\le i\le n} \|\varGamma_r^{-\frac{1}{2}} \varGamma_{ri}^{\overline{Y}} \varGamma_r^{-\frac{1}{2}} - \mathbf{I}\| \le R\right\}$. Then, $\mathbb{P}(\mathcal{E}_2) \ge 1 - \frac{1}{n}$.

Let us estimate σ^2. We fix $k_1, k_2 \in \Delta_r, l_1 \in \Delta_{s_1}, l_2 \in \Delta_{s_2}, s_1, s_2 \neq r$. Direct calculation gives us that

$$\begin{aligned}
&\mathbb{E}_{\overline{Y}}\left[\Gamma_r^{-1/2}\Gamma_{ri}^{\overline{Y}}\Gamma_r^{-1/2}\right]^2((\mathbf{u}_{k_1}, \mathbf{u}_{l_1}), (\mathbf{u}_{k_2}, \mathbf{u}_{l_2}))\\
&= \sum_{s\neq r}\sum_{k\in\Delta_r}\sum_{l\in\Delta_s}\mathbb{E}_{\overline{Y}}\,\overline{\eta}_{\mathbf{u}_{k_1},i}\overline{\eta}^2_{\mathbf{u}_{k,i}}\overline{\eta}_{\mathbf{u}_{k_2},i}\,\Gamma_2(\mathbf{u}_{l_1}, \mathbf{u}_l)\Gamma_2(\mathbf{u}_l, \mathbf{u}_{l_2})\\
&\quad\times\sqrt{\frac{(\mu_r-\mu_{s_1})^2(\mu_r-\mu_{s_2})^2}{\mu_{s_1}\mu_{s_2}}}\,\frac{(\mu_r-\mu_s)^2}{\mu_s}.
\end{aligned}$$

Let $(\mathbf{u}_{k_1}, \mathbf{u}_{l_1}) \neq (\mathbf{u}_{k_2}, \mathbf{u}_{l_2}))$. Then it is easy to check that

$$\mathbb{E}_{\overline{Y}}\left[\Gamma_r^{-1/2}\Gamma_{ri}^{\overline{Y}}\Gamma_r^{-1/2}\right]^2((\mathbf{u}_{k_1}, \mathbf{u}_{l_1}), (\mathbf{u}_{k_2}, \mathbf{u}_{l_2})) = 0.$$

This means that it is a diagonal matrix. Assume now that $(\mathbf{u}_{k_1}, \mathbf{u}_{l_1}) = (\mathbf{u}_{k_2}, \mathbf{u}_{l_2}))$. Then

$$\mathbb{E}_{\overline{Y}}\left[\Gamma_r^{-1/2}\Gamma_{ri}^{\overline{Y}}\Gamma_r^{-1/2}\right]^2((\mathbf{u}_{k_1}, \mathbf{u}_{l_1}), (\mathbf{u}_{k_1}, \mathbf{u}_{l_1})) = \sum_{k\in\Delta_r}\mathbb{E}_{\overline{Y}}\,\overline{\eta}_{\mathbf{u}_{k_1},i}\overline{\eta}^2_{\mathbf{u}_{k,i}}\overline{\eta}_{\mathbf{u}_{k_2},i} = m_r + 2.$$

Hence,

$$\|\,\mathbb{E}_{\overline{Y}}\big(\Gamma_r^{-\frac{1}{2}}\Gamma_{ri}^{\overline{Y}}\Gamma_r^{-\frac{1}{2}} - \mathbf{I}\big)^2\| \asymp m_r \text{ and } \sigma^2 \asymp nm_r. \tag{21}$$

Let us denote

$$\Delta_2^* \overset{\text{def}}{\asymp} \sqrt{m_r}\left(\sqrt{\frac{\log n}{n}} + \sqrt{\frac{\log p}{n}}\right).$$

It follows from (20) and (21) that

$$\mathbb{P}_{\overline{Y}}\Big(\mathcal{E}_1^c(\Delta_2^*)) \le \mathbb{P}_{\overline{Y}}(\mathcal{E}_1^c(\Delta_2^*) \cap \mathcal{E}_2\Big) + n^{-1} \le 2n^{-1}.$$

Similarly to the previous calculations we may also estimate the probability of the following event

$$\mathcal{E}_3(\delta) \overset{\text{def}}{=} \{\overline{Y} : \|\Gamma_r^{\overline{Y}} - \Gamma_r\| \le \delta\}, \quad \delta > 0.$$

It follows from Lemma 9 that

$$\mathbb{P}_{\overline{Y}}\left(\left\|\frac{1}{n}\sum_{i=1}^{n}\big(\Gamma_{ri}^{\overline{Y}} - \Gamma_r\big)\right\| \lesssim \frac{s}{n}\right) \ge 1 - p\cdot\exp\left(-\frac{s^2}{\tilde{\sigma}^2}\right), \tag{22}$$

where

$$\tilde{\sigma}^2 \stackrel{\text{def}}{=} \left\| \sum_{i=1}^{n} \mathbb{E}_{\overline{Y}} \big(\Gamma_{ri}^{\overline{Y}} - \Gamma_r \big)^2 \right\| \asymp n \max_{s \neq r} \frac{4\mu_r^2 \mu_s^2 (m_r + 2)}{(\mu_r - \mu_s)^4} \asymp n m_r \|\Gamma_r\|^2. \tag{23}$$

Here we applied the same arguments as above. Introduce the following quantity

$$\Delta_3^* \stackrel{\text{def}}{\asymp} \sqrt{m_r} \|\Gamma_r\| \left(\sqrt{\frac{\log n}{n}} + \sqrt{\frac{\log p}{n}} \right).$$

It follows from (22) and (23) that $\mathbb{P}(\mathcal{E}_3(\Delta_3^*)) \geq 1 - n^{-1}$. Let us denote $\mathcal{E} \stackrel{\text{def}}{=} \mathcal{E}_1(\Delta_2^*) \cap \mathcal{E}_3(\Delta_3^*)$. Without loss of generality we may assume that $\mathbb{P}(\mathcal{E}) \geq 1 - n^{-1}$. To finish the proof we apply Corollary 1 to obtain

$$\begin{aligned}
\mathbb{P}\left(\|S_r\|^2 \geq z \right) &= \int \mathbb{P}\left(\|S_r\|^2 \geq z \,\middle|\, \overline{Y} = y \right) d\, \mathbb{P}_{\overline{Y}}(y) \\
&= \int_{\mathcal{E}} \mathbb{P}\left(\|S_r\|^2 \geq z \,\middle|\, \overline{Y} = y \right) d\, \mathbb{P}_{\overline{Y}}(y) \\
&\quad + \int_{\mathcal{E}^c} \mathbb{P}\left(\|S_r\|^2 \geq z \,\middle|\, \overline{Y} = y \right) d\, \mathbb{P}_{\overline{Y}}(y) \\
&= \mathbb{P}(\|\xi\|^2 \geq z) + \mathcal{R}_n,
\end{aligned}$$

where

$$|\mathcal{R}_n| \leq \Delta_4^* \stackrel{\text{def}}{\asymp} \frac{\sqrt{m_r}\, \operatorname{tr} \Gamma_r}{\sqrt{\lambda_1(\Gamma_r)\lambda_2(\Gamma_r)}} \left(\sqrt{\frac{\log n}{n}} + \sqrt{\frac{\log p}{n}} \right).$$

Hence, we proved the following bound

$$\mathbb{P}\big(n \|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2 > x \big) \leq \mathbb{P}\big(\|\xi\|^2 \geq x_- \big) + \Delta_4^*.$$

Comparing definition of Δ_4 and Δ_1 with $\diamondsuit_1$ and $\diamondsuit_2$ resp. we get the claim of the theorem. □

4.3 Approximation in the bootstrap world

The main result of this section is the following theorem.

Theorem 5 *Assume that the conditions of Theorem 1 hold. Let $\xi^\circ \sim \mathcal{N}(0, \Gamma_r^\circ)$, where Γ_r° is defined below in (24). For all $x : x > 0$ the following bounds hold with $\mathbb{P}$-probability at least $1 - n^{-1}$:*

$$\begin{aligned}
\mathbb{P}^\circ \left(n \|\mathbf{P}_r^\circ - \widehat{\mathbf{P}}_r\|_2^2 > x \right) &\leq \mathbb{P}^\circ \left(\|\xi^\circ\|^2 \geq x_- \right) + n^{-1}, \\
\mathbb{P}^\circ \left(n \|\mathbf{P}_r^\circ - \widehat{\mathbf{P}}_r\|_2^2 > x \right) &\geq \mathbb{P}^\circ \left(\|\xi^\circ\|^2 \geq x_+ \right) - n^{-1}.
\end{aligned}$$

Here, $x_{\pm} \stackrel{\text{def}}{=} x \pm \diamondsuit_3$ *and*

$$\diamondsuit_3 \stackrel{\text{def}}{\asymp} m_r \frac{\operatorname{tr}^3 \boldsymbol{\Sigma}}{\overline{g}_r^3} \sqrt{\frac{\log^3 n}{n} + \frac{\log^3 p}{n}}.$$

Proof Let us fix an arbitrary $x \geq 0$. We introduce the following notations

$$\mathbf{E}^{\circ} \stackrel{\text{def}}{=} \boldsymbol{\Sigma}^{\circ} - \boldsymbol{\Sigma}, \quad \widehat{\mathbf{E}}^{\circ} \stackrel{\text{def}}{=} \boldsymbol{\Sigma}^{\circ} - \widehat{\boldsymbol{\Sigma}}.$$

and remind $\widehat{\mathbf{E}} = \widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}$. Applying Lemma 4 we may write

$$\mathbf{P}_r^{\circ} - \widehat{\mathbf{P}}_r = \mathbf{P}_r^{\circ} - \mathbf{P}_r - \left(\widehat{\mathbf{P}}_r - \mathbf{P}_r\right) = L_r\left(\mathbf{E}^{\circ}\right) - L_r\left(\widehat{\mathbf{E}}\right) + S_r\left(\mathbf{E}^{\circ}\right) + S_r\left(\widehat{\mathbf{E}}\right).$$

It is easy to see that

$$L_r\left(\mathbf{E}^{\circ}\right) - L_r\left(\widehat{\mathbf{E}}\right) = \mathbf{P}_r\left(\boldsymbol{\Sigma}^{\circ} - \widehat{\boldsymbol{\Sigma}}\right)\mathbf{C}_r + \mathbf{C}_r\left(\boldsymbol{\Sigma}^{\circ} - \widehat{\boldsymbol{\Sigma}}\right)\mathbf{P}_r \stackrel{\text{def}}{=} L_r\left(\hat{\mathbf{E}}^{\circ}\right).$$

Then

$$n\|\mathbf{P}_r^{\circ} - \widehat{\mathbf{P}}_r\|_2^2 = n\|L_r\left(\widehat{\mathbf{E}}^{\circ}\right)\|_2^2 + n\|\mathbf{P}_r^{\circ} - \widehat{\mathbf{P}}_r\|_2^2 - n\|L_r\left(\widehat{\mathbf{E}}^{\circ}\right)\|_2^2.$$

It follows from Theorem 3 that with $\mathbb{P}$ - probability at least $1 - n^{-1}$

$$\mathbb{P}^{\circ}\left(\left|n\|\mathbf{P}_r^{\circ} - \widehat{\mathbf{P}}_r\|_2^2 - n\|L_r\left(\widehat{\mathbf{E}}^{\circ}\right)\|_2^2\right| \leq \Delta_1^*\right) \geq 1 - n^{-1},$$

where

$$\Delta_1^* \stackrel{\text{def}}{\asymp} m_r \frac{\operatorname{tr}^3 \boldsymbol{\Sigma}}{\overline{g}_r^3} \sqrt{\frac{\log^3 n}{n} + \frac{\log^3 p}{n}}.$$

Introduce the notation $x_{\pm} \stackrel{\text{def}}{=} x \pm \Delta_1^*$. From the previous inequality we may conclude the following two-sided inequalities

$$\begin{aligned}\mathbb{P}^{\circ}\left(2n\|\mathbf{P}_r\widehat{\mathbf{E}}^{\circ}\mathbf{C}_r\|_2^2 \geq x_+\right) - n^{-1} &\leq \mathbb{P}^{\circ}\left(n\|\mathbf{P}_r^{\circ} - \widehat{\mathbf{P}}_r\|_2^2 > x\right) \\ &\leq \mathbb{P}^{\circ}\left(2n\|\mathbf{P}_r\widehat{\mathbf{E}}^{\circ}\mathbf{C}_r\|_2^2 \geq x_-\right) + n^{-1}.\end{aligned}$$

It follows that we need to estimate the term $2n\|\mathbf{P}_r\widehat{\mathbf{E}}^{\circ}\mathbf{C}_r\|_2^2$. Without loss of generality, we consider the case of the upper bound only, i.e. we set $z \stackrel{\text{def}}{=} x_+$. Similar calculations are valid for x_-. Analogously to the approximation in the $\mathbf{X}$-world we choose $\{\mathbf{u}_j\}_{j=1}^p$

as an orthonormal basis in $\mathbb{R}^p$. By Parseval's identity,

$$2n\|\mathbf{P}_r\widehat{\mathbf{E}}^{\circ}\mathbf{C}_r\|_2^2 = 2n\sum_{l,k=1}^{p}\left\langle\mathbf{P}_r\widehat{\mathbf{E}}^{\circ}\mathbf{C}_r\mathbf{u}_l,\mathbf{u}_k\right\rangle^2.$$

Applying the orthogonality of $\mathbf{P}_r$ and $\mathbf{C}_r$ we obtain

$$2n\|\mathbf{P}_r\widehat{\mathbf{E}}^{\circ}\mathbf{C}_r\|_2^2 = 2n\sum_{k\in\Delta_r}\sum_{s\neq r}\sum_{l\in\Delta_s}\langle\mathbf{P}_r\widehat{\mathbf{E}}^{\circ}\mathbf{C}_r\mathbf{u}_l,\mathbf{u}_k\rangle^2.$$

Let us fix arbitrary $\mathbf{u}_k$, $k\in\Delta_r$ and $\mathbf{u}_l$, $l\in\Delta_s$, $s\neq r$. For simplicity we denote them by $\mathbf{u}$ and $\mathbf{v}$ respectively. We may write

$$S^{\circ(\mathbf{u},\mathbf{v})}\stackrel{\text{def}}{=}\sqrt{2n}\langle\mathbf{P}_r\widehat{\mathbf{E}}^{\circ}\mathbf{C}_r\mathbf{v},\mathbf{u}\rangle=\sqrt{\frac{2}{n}}\sum_{i=1}^{n}\eta_i\langle\mathbf{u},\overline{Y}_i\rangle\langle\mathbf{v},Y_i\rangle,$$

where we denoted $\eta_i\stackrel{\text{def}}{=}w_i-1$, $\overline{Y}_i\stackrel{\text{def}}{=}\mathbf{P}_rX_i$ and $Y_i\stackrel{\text{def}}{=}\mathbf{C}_rX_i$. Since $\eta_i\sim\mathcal{N}(0,1)$, then

$$S^{\circ(\mathbf{u},\mathbf{v})}\stackrel{\text{d}}{=}\xi^{\circ}(\mathbf{u},\mathbf{v})\sim\mathcal{N}\left(0,\operatorname{Var}^{\circ}\left(\xi^{\circ}(\mathbf{u},\mathbf{v})\right)\right),\ \operatorname{Var}^{\circ}(\xi^{\circ}(\mathbf{u},\mathbf{v}))=\frac{2}{n}\sum_{i=1}^{n}\langle\mathbf{u},\overline{Y}_i)^2\langle\mathbf{v},Y_i\rangle^2.$$

Let us fix another pair $\tilde{\mathbf{u}}$, $\tilde{\mathbf{v}}$ and investigate the covariance

$$\Gamma_r^{\circ}\left((\mathbf{u},\mathbf{v}),(\tilde{\mathbf{u}},\tilde{\mathbf{v}})\right)\stackrel{\text{def}}{=}\operatorname{Cov}^{\circ}\left(\xi^{\circ}(\mathbf{u},\mathbf{v}),\xi^{\circ}(\tilde{\mathbf{u}},\tilde{\mathbf{v}})\right).$$

Direct calculations show that

$$\Gamma_r^{\circ}\left((\mathbf{u},\mathbf{v}),(\tilde{\mathbf{u}},\tilde{\mathbf{v}})\right)=\frac{2}{n}\sum_{i=1}^{n}\langle\mathbf{u},\overline{Y}_i\rangle\langle\tilde{\mathbf{u}},\overline{Y}_i\rangle\langle\mathbf{v},Y_i\rangle\langle\tilde{\mathbf{v}},Y_i\rangle.$$

We form the following covariance matrix

$$\Gamma_r^{\circ}\stackrel{\text{def}}{=}\left[\Gamma^{\circ}\left((\mathbf{u},\mathbf{v}),(\tilde{\mathbf{u}},\tilde{\mathbf{v}})\right)\right]_{((\mathbf{u},\mathbf{v}),(\tilde{\mathbf{u}},\tilde{\mathbf{v}}))}.\tag{24}$$

Denote $\xi^{\circ}\stackrel{\text{def}}{=}(\xi^{\circ}(\mathbf{u}_k,\mathbf{u}_l),k\in\Delta_r,s\neq r,l\in\Delta_s)$. Then

$$\mathbb{P}^{\circ}\left(2n\|\mathbf{P}_r\widehat{\mathbf{E}}^{\circ}\mathbf{C}_r\|_2^2\geq z\right)=\mathbb{P}^{\circ}\left(\|\xi^{\circ}\|^2\geq z\right).$$

Comparing definition of Δ_1 and $\diamondsuit_3$ we conclude the claim of the theorem. □

4.4 Gaussian comparison

In this section we prove the following Lemma.

Lemma 1 *Let $\xi \sim \mathcal{N}(0, \Gamma_r)$ and $\xi^\circ \sim \mathcal{N}(0, \Gamma_r^\circ)$, where Γ_r and Γ_r° are defined in (7), (24) respectively. Let $\mathfrak{m}$ be defined by the relations (8).*

Then the following holds with $\mathbb{P}$-probability al least $1 - n^{-1}$:

$$\sup_{x \geq 0} | \mathbb{P}(\|\xi\|^2 \geq x) - \mathbb{P}^\circ(\|\xi^\circ\|^2 \geq x)| \leq \diamondsuit_4,$$

where

$$\diamondsuit_4 \stackrel{\text{def}}{\asymp} \frac{\mathfrak{m}\,\operatorname{tr}\Gamma_r}{\sqrt{\lambda_1(\Gamma_r)\lambda_2(\Gamma_r)}}\left(\sqrt{\frac{\log n}{n}} + \sqrt{\frac{\log p}{n}}\right) + \frac{\operatorname{tr}(\mathbf{I} - \Pi_{\mathfrak{m}})\Gamma_r}{\sqrt{\lambda_1(\Gamma_r)\lambda_2(\Gamma_r)}}.$$

Here $\Pi_{\mathfrak{m}}$ is a projector on the subspace spanned by the eigenvectors of Γ_r corresponding to its largest $\mathfrak{m}$ eigenvalues.

Proof Without loss of generality we may assume that $\diamondsuit_4 \lesssim 1$. The proof is based on the application of Corollary 2. First we estimate $\|\Gamma_r^\circ - \Gamma_r\|$. Introduce the following notations

$$\Gamma_{ri}^\circ \stackrel{\text{def}}{=} [\Gamma_{ri}^\circ((\mathbf{u}_{k_1}, \mathbf{u}_{l_1}), (\mathbf{u}_{k_2}, \mathbf{u}_{l_2})), k_1, k_2 \in \Delta_r, l_1 \in \Delta_{s_1}, l_2 \in \Delta_{s_2}, s_1, s_2 \neq r],$$

where

$$\Gamma_{ri}^\circ((\mathbf{u}_{k_1}, \mathbf{u}_{l_1}), (\mathbf{u}_{k_2}, \mathbf{u}_{l_2})) \stackrel{\text{def}}{=} 2\sqrt{\frac{\mu_r^2}{(\mu_r - \mu_{s_1})^2(\mu_r - \mu_{s_2})^2}}\overline{\eta}_{\mathbf{u}_{k_1},i}\overline{\eta}_{\mathbf{u}_{k_2},i}\eta_{\mathbf{u}_{l_1},i}\eta_{\mathbf{u}_{l_2},i}.$$

In these notations we may rewrite Γ_r° as follows

$$\Gamma_r^\circ((\mathbf{u}_{k_1}, \mathbf{u}_{l_1}), (\mathbf{u}_{k_2}, \mathbf{u}_{l_2})) = \frac{1}{n}\sum_{i=1}^n \Gamma_{ri}^\circ((\mathbf{u}_{k_1}, \mathbf{u}_{l_1}), (\mathbf{u}_{k_2}, \mathbf{u}_{l_2})).$$

Due to Lemma 9 we need to show that there exists $R = R(n, \Gamma_r)$ such that

$$\max_{1 \leq k \leq n} \|\Gamma_{rk}^\circ - \Gamma_r\| \lesssim R,$$

and estimate

$$\tilde{\sigma}^2 = \left\|\sum_{k=1}^n \mathbb{E}(\Gamma_{rk}^\circ - \Gamma_r)^2\right\| = n\left\|\mathbb{E}(\Gamma_{r1}^\circ - \Gamma_r)^2\right\|.$$

It is obvious that $\|\Gamma_{ri}^{\circ} - \Gamma_r\| \le \|\Gamma_{ri}^{\circ}\| + \|\Gamma_r\|$. Let $Z_{rj} \stackrel{\text{def}}{=} (\overline{\eta}_{\mathbf{u}_k,j}\eta_{\mathbf{u}_l,j}, s \ne r, l \in \Delta_s, k \in \Delta_r)^{\mathsf{T}}$, $j = 1, \ldots, n$. Since $\Gamma_{r1}^{\circ} = Z_{r1}Z_{r1}^{\mathsf{T}}$ we obtain

$$\|\Gamma_{r1}^{\circ}\| = \|Z_{r1}\|^2 = 2\sum_{s\ne r}\sum_{k\in\Delta_r}\sum_{l\in\Delta_{s_1}} \frac{\mu_s\mu_r}{(\mu_s-\mu_r)^2}\overline{\eta}_{\mathbf{u}_k,1}^2\eta_{\mathbf{u}_l,1}^2.$$

Applying Lemma 5 we get $\mathbb{P}\left(\|\Gamma_{r1}^{\circ}\| \lesssim \log^2 n \ \operatorname{tr}\Gamma_r\right) \ge 1-n^{-1}$. Moreover, it is obvious that $\|\Gamma_r\| \le \operatorname{tr}\Gamma_r$. To bound $\max_{1\le i\le n}\|\Gamma_{ri}^{\circ} - \Gamma_r\|$ we introduce the following event

$$\mathcal{E}_1 \stackrel{\text{def}}{=} \left\{\max_{1\le i\le n}\|\Gamma_{ri}^{\circ} - \Gamma_r\| \lesssim \log^2 n \ \operatorname{tr}\Gamma_r.\right\}.$$

Using the union bound we may show that $\mathbb{P}(\mathcal{E}_1^c) \le n^{-1}$. It remains to estimate $\tilde{\sigma}^2$. Since $\tilde{\sigma}^2 = n\|\,\mathbb{E}(\Gamma_{r1}^{\circ})^2 - \Gamma_r^2\|$ we first calculate $\mathbb{E}(\Gamma_{r1}^{\circ})^2$. It follows that

$$\mathbb{E}(\Gamma_{r1}^{\circ})^2 = \mathbb{E}\,Z_{r1}Z_{r1}^{\mathsf{T}}Z_{r1}Z_{r1}^{\mathsf{T}} = \mathbb{E}\,\|Z_{r1}\|^2 Z_{r1}Z_{r1}^{\mathsf{T}}.$$

Let us fix some $s_1, s_2 \ne r, k_1, k_2 \in \Delta_r, l_1 \in \Delta_{s_1}, l_2 \in \Delta_{s_2}$. Then the entry of $\mathbb{E}(\Gamma_{r1}^{\circ})^2$ in the position $((\mathbf{u}_{k_1}, \mathbf{u}_{l_1}), (\mathbf{u}_{k_2}, \mathbf{u}_{l_2}))$ has the following form

$$\mathbb{E}\,\frac{4\mu_r\sqrt{\mu_{s_1}\mu_{s_2}}}{|\mu_{s_1}-\mu_r||\mu_{s_2}-\mu_r|}\overline{\eta}_{\mathbf{u}_{k_1}}\overline{\eta}_{\mathbf{u}_{k_2}}\eta_{\mathbf{u}_{l_1}}\eta_{\mathbf{u}_{l_2}}\sum_{s\ne r}\sum_{k\in\Delta_r}\sum_{l\in\Delta_s}\frac{\mu_s\mu_r}{(\mu_s-\mu_r)^2}\overline{\eta}_{\mathbf{u}_k}^2\eta_{\mathbf{u}_l}^2,$$

where $\overline{\eta}_{\mathbf{u}_k}, \eta_{\mathbf{u}_l}, k \in \Delta_r, l \in \Delta_s, s \ne r$, are i.i.d. $\mathcal{N}(0, 1)$ r.v. It is easy to check that all off-diagonal entries are equal zero and it remains to estimate the diagonal entries only. We obtain

$$\mathbb{E}\,\frac{4\mu_r\mu_{s_1}}{(\mu_{s_1}-\mu_r)^2}\overline{\eta}_{\mathbf{u}_{k_1}}^2\eta_{\mathbf{u}_{l_1}}^2\sum_{s\ne r}\sum_{k\in\Delta_r}\sum_{l\in\Delta_s}\frac{\mu_s\mu_r}{(\mu_s-\mu_r)^2}\overline{\eta}_{\mathbf{u}_k}^2\eta_{\mathbf{u}_l}^2 = \mathbb{E}\,S_1\,\mathbb{E}\,S_2,$$

where

$$\begin{aligned} S_1 &\stackrel{\text{def}}{=} \mu_r^2\sum_{k\in\Delta_r}\overline{\eta}_{\mathbf{u}_k}^2\overline{\eta}_{\mathbf{u}_{k_1}}^2, \\ S_2 &\stackrel{\text{def}}{=} 4\sum_{s\ne r}\sum_{l\in\Delta_s}\frac{\mu_s\mu_{s_1}}{(\mu_s-\mu_r)^2(\mu_{s_1}-\mu_r)^2}\eta_{\mathbf{u}_l}^2\eta_{\mathbf{u}_{l_1}}^2. \end{aligned}$$

We get that $\mathbb{E}\,S_1 \asymp \mu_r^2 m_r$ and

$$\mathbb{E}\,S_2 \asymp \sum_{s\ne r}\sum_{l\in\Delta_s}\frac{\mu_s\mu_{s_1}}{(\mu_s-\mu_r)^2(\mu_{s_1}-\mu_r)^2}.$$

Hence, $\tilde{\sigma}^2 \asymp n\|\varGamma_r\| \operatorname{tr} \varGamma_r = n\|\varGamma_r\|^2 \mathtt{r}(\varGamma_r)$. Let us introduce the following quantity

$$\Delta_1^* \overset{\text{def}}{\asymp} \|\varGamma_r\| \mathtt{r}^{\frac{1}{2}}(\varGamma_r) \left(\sqrt{\frac{\log n}{n}} + \sqrt{\frac{\log p}{n}} \right)$$

and the event $\mathcal{E}_2 \overset{\text{def}}{=} \{\|\varGamma_r^\circ - \varGamma_r\| \le \Delta_1^*\}$. Applying Lemma 9 we get

$$\mathbb{P}(\mathcal{E}_2^c) \le \mathbb{P}(\mathcal{E}_2^c \cap \mathcal{E}_1) + n^{-1} \le 2n^{-1}.$$

To apply Corollary 2 we also need to show that the remaining part of the trace of $\varGamma_r^\circ$ concentrates around its non-random counterpart. We take $\mathfrak{m}$ and $\varPi_{\mathfrak{m}}$ as stated in the lemma and let $\varPi_{\mathfrak{m}}^\circ$ be a projector on the subspace spanned by the eigenvectors of $\varGamma_r^\circ$ corresponding to its largest $\mathfrak{m}$ eigenvalues. It is easy to check that $\operatorname{tr}(\mathbf{I} - \varPi_{\mathfrak{m}}^\circ)\varGamma_r^\circ \le \operatorname{tr}(\mathbf{I} - \varPi_{\mathfrak{m}})\varGamma_r^\circ$. Denote $\overline{\varPi}_{\mathfrak{m}} \overset{\text{def}}{=} \mathbf{I} - \varPi_{\mathfrak{m}}$ and

$$\mathcal{E}_3 \overset{\text{def}}{=} \left\{ \left| \operatorname{tr} \overline{\varPi}_{\mathfrak{m}} \varGamma_r^\circ - \operatorname{tr} \overline{\varPi}_{\mathfrak{m}} \varGamma_r \right| \lesssim \operatorname{tr} \overline{\varPi}_{\mathfrak{m}} \varGamma_r \frac{\log^3 n}{\sqrt{n}} \right\}.$$

It is easy to check that

$$\operatorname{tr} \overline{\varPi}_{\mathfrak{m}} \varGamma_r^\circ = \frac{1}{n} \sum_{j=1}^n \operatorname{tr}(\overline{\varPi}_{\mathfrak{m}} Z_{rj})(\overline{\varPi}_{\mathfrak{m}} Z_{rj})^\mathsf{T} = \frac{2}{n} \sum_{j=1}^n \sum_{s \in \mathcal{T}_r} \sum_{k \in \Delta_r} \sum_{l \in \Delta_s} \frac{\mu_s \mu_r}{(\mu_s - \mu_r)^2} \overline{\eta}_{\mathbf{u}_k, j}^2 \eta_{\mathbf{u}_l, j}^2,$$

where $\mathcal{T}_r$ is the set of indices corresponding to the smallest $d - \mathfrak{m}$ eigenvalues of $\varGamma_r$. Moreover, simple calculations show that $\mathbb{E} \operatorname{tr} \overline{\varPi}_{\mathfrak{m}} \varGamma_r^\circ = \operatorname{tr} \overline{\varPi}_{\mathfrak{m}} \varGamma_r$. We introduce additional notations. Denote $\gamma_{sr} \overset{\text{def}}{=} \frac{\mu_s \mu_r}{(\mu_s - \mu_r)^2}$ and

$$Q_j \overset{\text{def}}{=} \sum_{s \in \mathcal{T}_r} \sum_{k \in \Delta_r} \sum_{l \in \Delta_s} \gamma_{sr} [\overline{\eta}_{\mathbf{u}_k, j}^2 \eta_{\mathbf{u}_l, j}^2 - 1], \quad j = 1, \dots, n.$$

It is obvious that Q_j are i.i.d. r.v. We estimate

$$\mathbb{E} \left| \operatorname{tr} \overline{\varPi}_{\mathfrak{m}} \varGamma_r^\circ - \operatorname{tr} \overline{\varPi}_{\mathfrak{m}} \varGamma_r \right|^m = \frac{2^m}{n^m} \mathbb{E} \left| \sum_{j=1}^n Q_j \right|^m.$$

Applying Rosenthal's inequality (see e.g. [17]) we obtain

$$\mathbb{E} \left| \sum_{j=1}^n Q_j \right|^m \le C^m \left(m^{\frac{m}{2}} n^{\frac{m}{2}} \mathbb{E}^{\frac{m}{2}} Q_1^2 + m^m n \, \mathbb{E} |Q_1|^m \right).$$

It remains to estimate $\mathbb{E} |Q_1|^m$. We may rewrite Q_1 as follows

$$Q_1 = A Q_{11} + Q_{11} Q_{12} + m_r Q_{12},$$

where

$$A \stackrel{\text{def}}{=} \sum_{s \in \mathcal{I}_r} \sum_{l \in \Delta_s} \gamma_{sr}, \qquad Q_{11} \stackrel{\text{def}}{=} \sum_{k \in \Delta_r} [\overline{\eta}^2_{\mathbf{u}_k} - 1],$$
$$Q_{12} \stackrel{\text{def}}{=} \sum_{s \in \mathcal{I}_r} \sum_{l \in \Delta_s} \gamma_{sr} [\eta^2_{\mathbf{u}_l} - 1].$$

Applying Lemma 5 we estimate each term separately and show that

$$\mathbb{E}\,|Q_1|^m \lesssim C^m m^{2m} \operatorname{tr}^m \overline{\varPi}_{\mathfrak{m}} \varGamma_r.$$

Hence,

$$\mathbb{E} \left| \operatorname{tr} \overline{\varPi}_{\mathfrak{m}} \varGamma_r^{\circ} - \operatorname{tr} \overline{\varPi}_{\mathfrak{m}} \varGamma_r \right|^m \le \frac{C^m m^{3m}}{n^{\frac{m}{2}}} \operatorname{tr}^m \overline{\varPi}_{\mathfrak{m}} \varGamma_r.$$

Choosing $m \asymp \log n$ and applying Markov's inequality we get $\mathbb{P}(\mathcal{E}_3) \ge 1 - n^{-1}$. Denote now $\mathcal{E} = \mathcal{E}_2 \cap \mathcal{E}_3$. It follows that $\mathbb{P}(\mathcal{E}) \ge 1 - \frac{1}{n}$. Applying Corollary 2 we get that for all $w \in \mathcal{E}$

$$\sup_{x \ge 0} |\, \mathbb{P}(\|\xi\|^2 \ge x) - \mathbb{P}^{\circ}(\|\xi^{\circ}\|^2 \ge x)| \lesssim \Delta_2^*,$$

where

$$\Delta_2^* \stackrel{\text{def}}{=} \frac{\|\varGamma_r\|\, \mathfrak{m}\, \mathtt{r}^{\frac{1}{2}}(\varGamma_r)}{\sqrt{\lambda_1(\varGamma_r)\lambda_2(\varGamma_r)}} \left(\sqrt{\frac{\log n}{n}} + \sqrt{\frac{\log p}{n}} \right) + \frac{\operatorname{tr}(\mathbf{I} - \varPi_{\mathfrak{m}})\varGamma_r}{\sqrt{\lambda_1(\varGamma_r)\lambda_2(\varGamma_r)}}.$$

Comparing Δ_2^* with $\diamondsuit_4$ we finish the proof of this lemma. □

4.5 Proof of the main result

This section collects the results of the previous sections and provides a proof of Theorem 1.

Proof of Theorem 1 Let us fix an event $\mathcal{E} \subset \varOmega$ which holds with $\mathbb{P}$ - probability at least $1 - \frac{1}{n}$. Suppose that for all $\omega \in \mathcal{E}$ the statements of Theorems 4, 5 and Lemma 1 hold. First we show that for all $x > 0$

$$\left| \mathbb{P}^{\circ} \left(n \|\mathbf{P}_r^{\circ} - \widehat{\mathbf{P}}_r\|_2^2 > x \right) - \mathbb{P} \left(n \|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2 > x \right) \right| \lesssim \diamondsuit, \tag{25}$$

where $\diamondsuit$ is defined in (10). Applying Theorem 5 we may show that

$$\mathbb{P}^{\circ} \left(n \|\mathbf{P}_r^{\circ} - \widehat{\mathbf{P}}_r\|_2^2 > x \right) \ge \mathbb{P}^{\circ} \left(\|\xi^{\circ}\|^2 \ge x + \diamondsuit_3 \right) - n^{-1},$$

where we recall that

$$\diamondsuit_3 \asymp m_r \frac{\operatorname{tr}^3 \boldsymbol{\Sigma}}{\overline{g}_r^3} \sqrt{\frac{\log^3 n}{n} + \frac{\log^3 p}{n}}.$$

Lemma 1 implies

$$\mathbb{P}^{\circ}\left(n\|\mathbf{P}_r^{\circ} - \widehat{\mathbf{P}}_r\|_2^2 > x\right) \geq \mathbb{P}\left(\|\xi\|^2 \geq x + \diamondsuit_3\right) - \diamondsuit_4 - n^{-1}, \tag{26}$$

where

$$\diamondsuit_4 \asymp \frac{\|\Gamma_r\|\, \mathfrak{m}\, \mathtt{r}^{\frac{1}{2}}(\Gamma_r)}{\sqrt{\lambda_1(\Gamma_r)\lambda_2(\Gamma_r)}} \left(\sqrt{\frac{\log n}{n}} + \sqrt{\frac{\log p}{n}}\right) + \frac{\operatorname{tr}\left(\mathbf{I} - \Pi_{\mathfrak{m}}\right)\Gamma_r}{\sqrt{\lambda_1(\Gamma_r)\lambda_2(\Gamma_r)}}.$$

As it is clear from (26) we need to get bounds for $\diamondsuit_3$-band of the squared norm of the Gaussian element ξ. For this purpose one can use Lemma 3. Then we get from (26)

$$\mathbb{P}^{\circ}(n\|\mathbf{P}_r^{\circ} - \widehat{\mathbf{P}}_r\|_2^2 > x) \geq \mathbb{P}(\|\xi\|^2 \geq x) - \overline{\diamondsuit}_3 - \diamondsuit_4,$$

where

$$\overline{\diamondsuit}_3 \asymp \frac{m_r \operatorname{tr}^3 \boldsymbol{\Sigma}}{\overline{g}_r^3 \sqrt{\lambda_1(\Gamma_r)\lambda_2(\Gamma_r)}} \sqrt{\frac{\log^3 n}{n} + \frac{\log^3 p}{n}}.$$

Finally, applying Theorem 4 and Lemma 3 we get

$$\mathbb{P}^{\circ}\left(n\|\mathbf{P}_r^{\circ} - \widehat{\mathbf{P}}_r\|_2^2 > x\right) \geq \mathbb{P}\left(n\|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2 > x\right) - \diamondsuit_1 - \diamondsuit_2 - \overline{\diamondsuit}_3 - \diamondsuit_4,$$

where

$$\diamondsuit_1 \asymp \frac{m_r^{1/2} \operatorname{tr} \Gamma_r}{\sqrt{\lambda_1(\Gamma_r)\lambda_2(\Gamma_r)}} \left(\sqrt{\frac{\log n}{n}} + \sqrt{\frac{\log p}{n}}\right),$$

$$\diamondsuit_2 \asymp \frac{m_r \operatorname{tr}^3 \boldsymbol{\Sigma}}{\overline{g}_r^3 \sqrt{\lambda_1(\Gamma_r)\lambda_2(\Gamma_r)}} \sqrt{\frac{\log^3 n}{n}}.$$

Similarly we may write down all inequalities in the opposite direction. It is easy to see that $\diamondsuit_1 + \diamondsuit_2 + \overline{\diamondsuit}_3 + \diamondsuit_4 \leq \diamondsuit$, where

$$\diamondsuit \asymp \frac{\mathfrak{m} \operatorname{tr} \Gamma_r}{\sqrt{\lambda_1(\Gamma_r)\lambda_2(\Gamma_r)}} \left(\sqrt{\frac{\log n}{n}} + \sqrt{\frac{\log p}{n}}\right) + \frac{\operatorname{tr}(\mathbf{I} - \Pi_{\mathfrak{m}})\Gamma_r}{\sqrt{\lambda_1(\Gamma_r)\lambda_2(\Gamma_r)}}$$
$$+ \frac{m_r \operatorname{tr}^3 \boldsymbol{\Sigma}}{\overline{g}_r^3 \sqrt{\lambda_1(\Gamma_r)\lambda_2(\Gamma_r)}} \left(\sqrt{\frac{\log^3 n}{n}} + \sqrt{\frac{\log^3 p}{n}}\right).$$

Hence, we finish the proof of (25). Now we show that for all $w \in \mathcal{E}$

$$\gamma_{\alpha+\varepsilon_1} \le \gamma_\alpha^\circ \le \gamma_{\alpha-\varepsilon_2} \tag{27}$$

with $\varepsilon_1 \stackrel{\text{def}}{=} 2\diamondsuit$, $\varepsilon_2 \stackrel{\text{def}}{=} \diamondsuit$. It follows from Theorem 4, Lemma 3 and definition of $\diamondsuit$ that

$$\mathbb{P}\left(n\|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2 > \gamma_\alpha\right) \ge \alpha - \diamondsuit.$$

Moreover, by definition (3) of γ_α we write $\mathbb{P}(n\|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2 > \gamma_\alpha) \le \alpha$. Both inequalities imply that

$$\alpha - \diamondsuit \le \mathbb{P}\left(n\|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2 > \gamma_\alpha\right) \le \alpha. \tag{28}$$

The proof of (27) follows from this inequality and (25):

$$\begin{aligned}
\mathbb{P}^\circ\left(n\|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2 > \gamma_{\alpha-\varepsilon_2}\right) &\le \mathbb{P}\left(n\|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2 > \gamma_{\alpha-\varepsilon_2}\right) + \diamondsuit \le \alpha,\\
\mathbb{P}^\circ\left(n\|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2 > \gamma_{\alpha+\varepsilon_1}\right) &\ge \mathbb{P}\left(n\|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2 > \gamma_{\alpha+\varepsilon_1}\right) - \diamondsuit \ge \alpha.
\end{aligned}$$

Hence, applying (27) and (28) we write

$$\begin{aligned}
\mathbb{P}\left(n\|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2 > \gamma_\alpha^\circ\right) - \alpha &\le \mathbb{P}\left(n\|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2 > \gamma_{\alpha+\varepsilon_1}\right) - \alpha \le 2\diamondsuit,\\
\mathbb{P}\left(n\|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2 > \gamma_\alpha^\circ\right) - \alpha &\ge \mathbb{P}\left(n\|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2 > \gamma_{\alpha-\varepsilon_2}\right) - \alpha \ge -2\diamondsuit.
\end{aligned}$$

The last two inequalities conclude the claim of the theorem. □

5 Gaussian comparison and anti-concentration inequalities

In this section we obtain bounds for the Kolmogorov distance between the probabilities of two Gaussian elements to hit a ball in a Hilbert space. The key property of these bounds is that they are dimension-free and depend on the nuclear (Schatten-one) norm of the difference between the covariance operators of the elements. We start from the discussion of the Gaussian comparison inequality.

Due to the Pinsker inequality the total variation distance between any probability measures $\mathbb{P}_1$ and $\mathbb{P}_2$ on $(\Omega, \mathfrak{F})$ may be bounded as follows

$$\sup_{A\in\mathfrak{F}} |\mathbb{P}_1(A) - \mathbb{P}_2(A)| \le \sqrt{\mathrm{KL}(\mathbb{P}_1, \mathbb{P}_2)/2}, \tag{29}$$

where

$$\mathrm{KL}(\mathbb{P}_1, \mathbb{P}_2) \stackrel{\text{def}}{=} \int \log \frac{d\,\mathbb{P}_1}{d\,\mathbb{P}_2} d\,\mathbb{P}_1$$

is the Kullback–Leibler divergence between $\mathbb{P}_1, \mathbb{P}_2$; see e.g. [21, pp. 88–132]. Let ξ and η be Gaussian elements in a real separable Hilbert space $\mathbb{H}$ with zero mean and covariance matrices $\mathbf{\Sigma}_\xi, \mathbf{\Sigma}_\eta$ resp. Assume that $\|\mathbf{\Sigma}_\xi^{-\frac{1}{2}}\mathbf{\Sigma}_\eta\mathbf{\Sigma}_\xi^{-\frac{1}{2}} - \mathbf{I}\| \le 1/2$. Taking $\mathbb{P}_1 \stackrel{\text{def}}{=} \mathcal{N}(0, \mathbf{\Sigma}_\xi)$ and $\mathbb{P}_2 \stackrel{\text{def}}{=} \mathcal{N}(0, \mathbf{\Sigma}_\eta)$ one may check (see e.g. [19]) that

$$\mathrm{KL}(\mathbb{P}_1, \mathbb{P}_2) \le \|\mathbf{\Sigma}_\xi^{-\frac{1}{2}}\mathbf{\Sigma}_\eta\mathbf{\Sigma}_\xi^{-\frac{1}{2}} - \mathbf{I}\|_2^2/2.$$

In particular, the last inequality and (29) imply that

$$\sup_{x>0} |\,\mathbb{P}(\|\xi\| \le x) - \mathbb{P}(\|\eta\| \le x)| \le \|\mathbf{\Sigma}_\xi^{-\frac{1}{2}}\mathbf{\Sigma}_\eta\mathbf{\Sigma}_\xi^{-\frac{1}{2}} - \mathbf{I}\|_2/2. \tag{30}$$

To apply this inequality one have to estimate the Hilbert–Schmidt norm in the r.h.s. of (30). Below we will show that using Bernstein's matrix inequality we may control the operator norm $\|\mathbf{\Sigma}_\xi^{-\frac{1}{2}}\mathbf{\Sigma}_\eta\mathbf{\Sigma}_\xi^{-\frac{1}{2}} - \mathbf{I}\|$ and the r.h.s. of (30) may be bounded up to some constant by $\sqrt{p}\|\mathbf{\Sigma}_\xi^{-\frac{1}{2}}\mathbf{\Sigma}_\eta\mathbf{\Sigma}_\xi^{-\frac{1}{2}} - \mathbf{I}\|$. The following lemma shows that it is possible to derive a dimensional free bound if we limit ourselves to the centred balls from the beginning.

We recall notation of the nuclear (Schatten-one) norm. For a self-adjoint operator $\mathbf{A}$ with eigenvalues $\lambda_k(\mathbf{A})$, $k \ge 1$, let us denote by $\|\mathbf{A}\|_1$ the nuclear norm by

$$\|\mathbf{A}\|_1 \stackrel{\text{def}}{=} \operatorname{tr}|\mathbf{A}| = \sum_{k=1}^{\infty} |\lambda_k(\mathbf{A})|.$$

We suppose below that $\mathbf{A}$ is a nuclear operator and $\|\mathbf{A}\|_1 < \infty$. Let $\mathbf{\Sigma}_\xi$ be a covariance operator of an arbitrary Gaussian random element in $\mathbb{H}$. By $\{\lambda_{k\xi}\}_{k\ge 1}$ we denote the set of its eigenvalues arranged in the non-increasing order, i.e. $\lambda_{1\xi} \ge \lambda_{1\xi} \ge \ldots$, and let $\mathbf{\Lambda}_\xi \stackrel{\text{def}}{=} \operatorname{diag}(\lambda_{j\xi})_{j=1}^{\infty}$. Note that $\sum_{j=1}^{\infty} \lambda_{j\xi} < \infty$. The following lemma is the main result of this section.

Lemma 2 *Let ξ and η be Gaussian elements in $\mathbb{H}$ with zero mean and covariance operators $\mathbf{\Sigma}_\xi$ and $\mathbf{\Sigma}_\eta$ respectively. The following inequality holds*

$$\sup_{x\ge 0}\left|\mathbb{P}\left(\|\xi\|^2 \le x\right) - \mathbb{P}\left(\|\eta\|^2 \le x\right)\right| \lesssim \left(\frac{1}{\sqrt{\lambda_{1\eta}\lambda_{2\eta}}} + \frac{1}{\sqrt{\lambda_{1\xi}\lambda_{2\xi}}}\right)\diamondsuit_0,$$

where $\diamondsuit_0 \stackrel{\text{def}}{=} \|\mathbf{\Lambda}_\xi - \mathbf{\Lambda}_\eta\|_1$.

More general problem to obtain the upper bounds for the closeness of two Gaussian measures with different means and covariance operators in the class of balls [9].

We complement the result of Lemma 2 with several remarks. The first remark is that by the Weilandt–Hoffman inequality, $\|\mathbf{\Lambda}_\xi - \mathbf{\Lambda}_\eta\|_1 \le \|\mathbf{\Sigma}_\xi - \mathbf{\Sigma}_\eta\|_1$, see e.g. [15]. This yields the bound in terms of the nuclear norm of the difference $\mathbf{\Sigma}_\xi - \mathbf{\Sigma}_\eta$, which may be more useful in a number of applications. The result of Lemma 2 may be also rewritten in terms of the operator norm $\|\boldsymbol{\Sigma}_{\xi}^{-\frac{1}{2}}\boldsymbol{\Sigma}_{\eta}\boldsymbol{\Sigma}_{\xi}^{-\frac{1}{2}} - \mathbf{I}\|$.

Corollary 1 *Under assumptions of Lemma 2 the following bound for $\diamondsuit_0$ holds*

$$\diamondsuit_0 \le \|\boldsymbol{\Sigma}_{\xi}^{-\frac{1}{2}}\boldsymbol{\Sigma}_{\eta}\boldsymbol{\Sigma}_{\xi}^{-\frac{1}{2}} - \mathbf{I}\| \operatorname{tr} \boldsymbol{\Sigma}_{\xi}.$$

Proof The proof follows directly from the following well known inequality $\|\mathbf{AB}\|_1 \le \|\mathbf{A}\|_1 \|\mathbf{B}\|$. □

In the current paper we will also use the following corollary of Lemma 2. Denote by $\mathbf{e}_{j\xi}$, $j \ge 1$—orthonormal eigenvectors of $\boldsymbol{\Sigma}_{\xi}$ and let $\Pi_{\xi,m} \stackrel{\text{def}}{=} \sum_{k=1}^{m} \mathbf{e}_{j\xi}\mathbf{e}_{j\xi}^{\mathsf{T}}$

Corollary 2 *Let $m : 1 \le m < \infty$. Under assumptions of Lemma 2 the following bound for $\diamondsuit_0$ holds*

$$\diamondsuit_0 \le m\|\boldsymbol{\Sigma}_{\xi} - \boldsymbol{\Sigma}_{\eta}\| + \operatorname{tr}(\mathbf{I} - \Pi_{\xi,m})\boldsymbol{\Sigma}_{\xi} + \operatorname{tr}(\mathbf{I} - \Pi_{\eta,m})\boldsymbol{\Sigma}_{\eta}. \tag{31}$$

Proof The proof is obvious. □

Remark 5 It is easy to see that we may assume without loss of generality that $\boldsymbol{\Sigma}_{\xi}$ and $\boldsymbol{\Sigma}_{\eta}$ are diagonal matrices. Then the last two terms in (31) are the sums of eigenvalues $\lambda_{j\xi}, \lambda_{j\eta}$, $j \ge m+1$.

In the next lemma we show that one may obtain dimensional free anti-concentration inequality for the squared norm of a Gaussian element with dependence on the first two largest eigenvalues of $\boldsymbol{\Sigma}$ only.

Lemma 3 (Δ-band of the squared norm of a Gaussian element) *Let ξ be a Gaussian element in $\mathbb{H}$ with zero mean and covariance operator $\boldsymbol{\Sigma}_{\xi}$. Then for arbitrary $\Delta > 0$ and any $\lambda > \lambda_1$*

$$\mathbb{P}(x < \|\xi\|^2 < x + \Delta) \le \left[\frac{e^{-x/(2\lambda)}}{2\sqrt{\lambda_{1\xi}\lambda_{2\xi}}} \prod_{j=3}^{\infty} (1 - \lambda_{j\xi}/\lambda)^{-1/2}\right] \Delta, \tag{32}$$

where $\lambda_{1\xi} \ge \lambda_{2\xi} \ge \cdots$ are the eigenvalues of $\boldsymbol{\Sigma}_{\xi}$. In particular, one has

$$\sup_{x>0} \mathbb{P}(x < \|\xi\|^2 < x + \Delta) \le \frac{\Delta}{\sqrt{\lambda_{1\xi}\lambda_{2\xi}}}. \tag{33}$$

Remark 6 The infinite product in the r.h.s. of (32) is convergent. Indeed, taking logarithm and using $\log(1+x) \ge x/(x+1)$ for $x > -1$ we obtain

$$0 < -\frac{1}{2} \log \prod_{j=3}^{\infty} (1 - \lambda_{j\xi}/\lambda) \le \frac{1}{2(\lambda - \lambda_{1\xi})} \sum_{j=3}^{\infty} \lambda_{j\xi} < \infty,$$

where we also used the fact that $\boldsymbol{\Sigma}_{\xi}$ is a nuclear operator and $\|\boldsymbol{\Sigma}_{\xi}\|_1 < \infty$. Taking $\lambda = \|\boldsymbol{\Sigma}_{\xi}\|_1$ we get $\prod_{j=3}^{\infty} (1 - \lambda_{j\xi}/\lambda)^{-1/2} \le \sqrt{e}$.

Remark 7 The right-hand sides of (32) and (33) depend on first two eigenvalues of $\mathbf{\Sigma}_\xi$. In general it is impossible to get similar bounds of order $O(\Delta)$ with dependence on $\lambda_{1\xi}$ only. It is easy to get in one dimensional case, i.e. when $\lambda_{1_\xi} = 1$ and $\lambda_{2\xi} = 0$, that for all positive $\Delta \le \log 2$ one has

$$\sup_{x>0} \mathbb{P}(x < \|\xi\|^2 < x + \Delta) \ge \Delta^{1/2}/(2\sqrt{\pi}).$$

Proof of Lemma 2 Without loss of generality we may assume that $\mathbf{\Sigma}_\xi = \mathbf{\Lambda}_\xi$, $\mathbf{\Sigma}_\eta = \mathbf{\Lambda}_\eta$, where $\mathbf{\Lambda}_\xi \overset{\text{def}}{=} \operatorname{diag}(\lambda_{1\xi}, \lambda_{2\xi}, \dots)$, $\mathbf{\Lambda}_\eta \overset{\text{def}}{=} \operatorname{diag}(\lambda_{1\eta}, \lambda_{2\eta}, \dots)$ and $\lambda_{1\xi} \ge \lambda_{1\xi} \ge \cdots$ and similarly in decreasing order for $\lambda_{i\eta}$'s.

Fix any $s : 0 \le s \le 1$. Let $Z(s)$ be a Gaussian random element in $\mathbb{H}$ with zero mean and diagonal covariance operator $\mathbf{\Lambda}(s)$:

$$\mathbf{\Lambda}(s) \overset{\text{def}}{=} s\mathbf{\Lambda}_\xi + (1-s)\mathbf{\Lambda}_\eta.$$

Denote by $\lambda_1(s) \ge \lambda_2(s) \ge \cdots$-eigenvalues of $\mathbf{\Lambda}(s)$. It is straightforward to check that the characteristic function $f(t,s)$ of $\|Z(s)\|_2^2$ can be written as

$$\begin{aligned} f(t,s) = \mathbb{E}\exp\{it\|Z(s)\|^2\} &= \prod_{j=1}^{\infty}(1-2it\lambda_j(s))^{-1/2} \\ &= \exp\Big\{-\frac{1}{2}\sum_{j=1}^{\infty}\log(1-2it\lambda_j(s))\Big\}. \end{aligned} \tag{34}$$

Indeed, one may use the following representation

$$Z(s) \overset{\text{d}}{=} \sum_{j=1}^{\infty} \sqrt{\lambda_j(s)}\, Z_j\, \mathbf{e}_j, \tag{35}$$

where Z_j, $j \ge 1$, are i.i.d. $\mathcal{N}(0,1)$ r.v. and $\mathbf{e}_j$, $j \ge 1$, be the standard orthonormal basis in $\mathbb{H}$. Then it is sufficient to apply an expression for the characteristic function of Z_j^2. We rewrite $f(t,s)$ in terms of trace-class operators

$$f(t,s) = \exp\big\{-\operatorname{tr}\log\big(\mathbf{I} - 2it\mathbf{\Lambda}(s)\big)/2\big\}.$$

It is well known, see e.g. [7, Sect. 6.2, p. 168], that for a continues d.f. $F(x)$ with c.f. $f(t)$ we may write

$$F(x) = \frac{1}{2} + \frac{i}{2\pi}\lim_{T\to\infty} \text{V.P.}\int_{-T}^{T} e^{-itx} f(t)\frac{dt}{t},$$

where V.P. stands for the principal value of the integral. Let us fix an arbitrary $x > 0$ and denote $\Delta(x) \overset{\text{def}}{=} \mathbb{P}(\|\xi\|^2 \le x) - \mathbb{P}(\|\eta\|^2 \le x)$. Then

$$\Delta(x) = \frac{i}{2\pi} \lim_{T\to\infty} \text{V.P.} \int_{-T}^{T} \frac{f(t,1) - f(t,0)}{t} e^{-itx}\, dt.$$

Since

$$f(t,1) - f(t,0) = \int_0^1 \frac{\partial f(t,s)}{\partial s}\, ds,$$

changing the order of integration we get

$$\Delta(x) = \frac{i}{2\pi} \lim_{T\to\infty} \text{V.P.} \int_0^1 \int_{-T}^{T} \frac{\partial f(t,s)/\partial s}{t} e^{-itx}\, dt\, ds$$

It is easy to check that

$$\partial f(t,s)/\partial s = i\, t\, f(t,s)\ \mathrm{tr}\big\{(\mathbf{\Lambda}_\xi - \mathbf{\Lambda}_\eta)\mathbf{G}(t,s)\big\},$$

where $\mathbf{G}(t,s) \overset{\text{def}}{=} (\mathbf{I} - 2it\,\mathbf{\Lambda}(s))^{-1}$. Hence,

$$\Delta(x) = -\frac{1}{2\pi} \lim_{T\to\infty} \int_0^1 \mathrm{tr}\left\{(\mathbf{\Lambda}_\xi - \mathbf{\Lambda}_\eta)\widehat{\mathbf{G}}(T,s)\right\} ds, \tag{36}$$

where

$$\widehat{\mathbf{G}}(T,s) \overset{\text{def}}{=} \int_{-T}^{T} f(t,s)\mathbf{G}(t,s)e^{-itx}\, dt,\ s \in [0,1],\ T > 0.$$

We show that for any $T > 0$ and $s \in [0,1]$ one has

$$\|\widehat{\mathbf{G}}(T,s)\| \le \frac{c}{\sqrt{\lambda_1(s)\lambda_2(s)}}. \tag{37}$$

For this aim we denote the eigenvalues of $\mathbf{G}(t,s)$ by $\mu_j(t,s) \overset{\text{def}}{=} (1 - 2it\lambda_j(s))^{-1}$. Let $\overline{Z}_j(s)$, $j \ge 1$ be independent exponentially distributed r.v. with mean $2\lambda_j(s)$ (we write $Exp(2\lambda_j(s))$), which are also independent of Z_k, $k \ge 1$. Then

$$\mathbb{E}\, e^{it\overline{Z}_j} = \mu_j(t,s). \tag{38}$$

Applying (38) we obtain

$$
\begin{aligned}
f(t,s)\mu_j(t,s) &= \mathbb{E}\exp\Big(it\big[\sum_{k\ge 1}\lambda_k(s)Z_k^2+\overline{Z}_j\big]\Big)\\
&= \mathbb{E}\,e^{ita_j^2}\cdot\mathbb{E}\Big(\exp\big(it\big[\lambda_1(s)Z_1^2+\lambda_2(s)Z_2^2\big]\big)\Big), \qquad (39)
\end{aligned}
$$

where $a_j^2 \stackrel{\text{def}}{=} \overline{Z}_j+\sum_{k\ge 3}\lambda_k(s)Z_k^2$. We fix j and get a bound for

$$
I \stackrel{\text{def}}{=} \left|\int_{-T}^{T} f(t,s)\mu_j(t,s)e^{-itx}\,dt\right|.
$$

Using (39) we obtain

$$
I \le \mathbb{E}\left|\int_{-T}^{T} e^{it(a_j^2-x)}\,\mathbb{E}\exp\Big(it\big[\lambda_1(s)Z_1^2+\lambda_2(s)Z_2^2\big]\Big)\,dt\right|.
$$

It follows from [8, Lemma 2.2] (see also [16, p. 242]) that there exists an absolute constant c such that

$$
\left|\int_{-T}^{T} e^{it(a_j^2-x)}\,\mathbb{E}\exp\Big(it\big[\lambda_1(s)Z_1^2+\lambda_2(s)Z_2^2\big]\Big)dt\right| \le \frac{c}{\sqrt{\lambda_1(s)\lambda_2(s)}}. \qquad (40)
$$

For readers convenience we repeat the proof of this inequality below in Lemma 10. Applying (40) we get that the absolute values of all eigenvalues of $\widehat{\mathbf{G}}(T,s)$ are bounded by $c(\lambda_1(s)\lambda_2(s))^{-1/2}$ and, therefore, we obtain (37). Hence

$$
\left|\operatorname{tr}\left\{(\mathbf{\Lambda}_\xi-\mathbf{\Lambda}_\eta)\widehat{\mathbf{G}}(T,s)\right\}\right| \le \|\mathbf{\Lambda}_\xi-\mathbf{\Lambda}_\eta\|_1\|\widehat{\mathbf{G}}(T,s)\| \le \frac{c\|\mathbf{\Lambda}_\xi-\mathbf{\Lambda}_\eta\|_1}{\sqrt{\lambda_1(s)\lambda_2(s)}}.
$$

The last inequality and (36) imply the claim of the lemma. □

Proof of Lemma 3 To simplify all notations we will omit index ξ.

The inequality (33) follows immediately from (32) if we take $\lambda=\|\mathbf{\Sigma}\|_1$ and use Remark 6.

In order to prove (32) it is sufficient to show that for a density function $g(u)$ of $\|\xi\|_2^2$ one has

$$
g(u) \le \frac{e^{-u/(2\lambda)}}{2\sqrt{\lambda_1\lambda_2}}\prod_{j=3}^{\infty}(1-\lambda_j/\lambda)^{-1/2}. \qquad (41)
$$

According to representation (35) $\|\xi\|^2 \stackrel{\text{d}}{=} \sum_{j=1}^{\infty}\lambda_j Z_j^2$, where $Z_1, Z_2, \ldots$ are i.i.d. $\mathcal{N}(0,1)$ r.v. We denote by $g(m,u)$, $m=1,2,\ldots$ (resp. $g_j(u)$, $j=1,2,\ldots$) the density function of $\sum_{j=1}^{m} \lambda_j Z_j^2$ (resp. $\lambda_j Z_j^2$). We have for all $j = 1, 2, \dots$ and any $\lambda > \lambda_1$

$$
\begin{aligned}
g_j(u) &= (2\pi u \lambda_j)^{-1/2} d_j(u) \\
&\le (2\pi u \lambda_j)^{-1/2} \exp\{-u/(2\lambda)\} d_j(\lambda\lambda_j/(\lambda-\lambda_j), u),
\end{aligned} \tag{42}
$$

where $d_j(u) = d(\lambda_j, u) = \exp\{-u/(2\lambda_j)\}$. Moreover,

$$
(2\pi u)^{-1/2} (\lambda-\lambda_j)^{1/2} / (\lambda\lambda_j)^{1/2}\, d_j(\lambda\lambda_j/(\lambda-\lambda_j), u)
$$

is the density function of $Z_j^2 \sqrt{\lambda\lambda_j}/(\lambda-\lambda_j)^{1/2}$. First consider $g(2, u)$:

$$
\begin{aligned}
g(2,u) &= \int_0^u g_1(u-v) g_2(v) dv \\
&\le \frac{\exp\{-u/(2\lambda)\}}{2\pi\sqrt{\lambda_1\lambda_2}} \int_0^1 \frac{dz}{\sqrt{(1-z)z}} = \frac{\exp\{-u/(2\lambda_1)\}}{2\sqrt{\lambda_1\lambda_2}}.
\end{aligned} \tag{43}
$$

Therefore, due to (42) and (43) we obtain

$$
\begin{aligned}
g(3,u) &= \int_0^u g_2(u-v) g_3(v) dv \\
&\le \frac{\exp\{-u/(2\lambda)\}}{2\sqrt{\lambda_1\lambda_2}\sqrt{2\pi\lambda_3}} \int_0^u \frac{d_3(\lambda\lambda_3/(\lambda-\lambda_3), v)}{\sqrt{v}} dv \\
&\le \frac{\exp\{-u/(2\lambda)\}}{2\sqrt{\lambda_1\lambda_2}} (1-\lambda_3/\lambda)^{-1/2}.
\end{aligned}
$$

In a similar way by induction we can get for any $m > 3$ that

$$
g(m,u) \le \frac{e^{-u/(2\lambda)}}{2\sqrt{\lambda_1\lambda_2}} \prod_{j=3}^{m} (1-\lambda_j/\lambda)^{-1/2}. \tag{44}
$$

Now take an arbitrary $\varepsilon > 0$ and any integer $m > 0$. Let $0 < \mu < 1/(2\lambda_j)$ for all $j \ge m+1$. By Markovs inequality and using the moment generating function of χ^2 we obtain

$$
\mathbb{P}\left(\sum_{j=m+1}^{\infty} \lambda_j Z_j^2 \ge \varepsilon^2 \right) \le e^{-\mu\varepsilon^2} \prod_{j=m+1}^{\infty} \mathbb{E}\, e^{\mu\lambda_j Z_j^2} = e^{-\mu\varepsilon^2} \prod_{j=m+1}^{\infty} \frac{1}{\sqrt{1-2\mu\lambda_j}}.
$$

Let us choose $\mu \overset{\text{def}}{=} 1/(4\sum_{j=m+1}^{\infty} \lambda_j)$. Taking logarithm and using $\log(1+x) \ge x/(x+1)$ for $x > -1$ we obtain

$$
0 < -\frac{1}{2} \log \prod_{j=m+1}^{\infty} (1-2\mu\lambda_j) \le \frac{\mu}{1-2\mu\lambda_{m+1}} \sum_{j=m+1}^{\infty} \lambda_j < \frac{1}{4(1-1/2)} = 1/2,
$$

It follows now that $\prod_{j=m+1}^{\infty}(1-2\mu\lambda_j)^{-1/2} \le \sqrt{e} < 2$. Hence,

$$\mathbb{P}\left(\sum_{j=m+1}^{\infty} \lambda_j Z_j^2 \ge \varepsilon^2\right) \le 2\exp\left\{-\varepsilon^2\left(4\sum_{j=m+1}^{\infty}\lambda_j\right)^{-1}\right\}.$$

Since $\|\mathbf{\Sigma}\|_1 < \infty$ it follows that $\sum_{j=m+1}^{\infty}\lambda_j$ tends to zero as m goes to infinity. Hence, there exists $M \stackrel{\text{def}}{=} M(\varepsilon)$ such that for all $m \ge M$

$$\mathbb{P}\left(\sum_{j=m+1}^{\infty} \lambda_j Z_j^2 \ge \varepsilon^2\right) \le \varepsilon^2.$$

Therefore, for any $m \ge M$

$$\mathbb{P}(x-\varepsilon < \|\xi\|_2^2 < x+\varepsilon) \le \varepsilon^2 + 2(\varepsilon+\varepsilon^2)\sup_{y\in T(\varepsilon,x)} g(m,y), \tag{45}$$

where $T(\varepsilon,x) = \{y \in \mathbb{R}^1 : x-\varepsilon-\varepsilon^2 \le y \le x+\varepsilon+\varepsilon^2\}$. Dividing the right-hand side of (45) by 2ε we obtain (41) from (44) as ε tends to 0. $\square$

Acknowledgements The authors are grateful to the Associate Editor and the Reviewers for the careful reading of the manuscript and pertinent comments. Their constructive feedback helped to improve the quality of this work and shape its final form. This work has been funded by the Russian Academic Excellence Project '5-100'. Results of Section 5 have been obtained under support of the RSF Grant No. 18-11-00132.

A Auxiliary results

A.1 Concentration inequalities for sample covariances and spectral projectors in X-world

In this section we present concentration inequalities for sample covariance matrices and spectral projectors in **X**-world.

Theorem 6 *Let $X, X_1, \ldots, X_n$ be i.i.d. centered Gaussian random vectors in $\mathbb{R}^p$ with covariance $\mathbf{\Sigma} = \mathbb{E}(XX^T)$. Then*

$$\mathbb{E}\,\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\| \lesssim \|\mathbf{\Sigma}\|\left(\sqrt{\frac{\mathtt{r}(\mathbf{\Sigma})}{n}} + \frac{\mathtt{r}(\mathbf{\Sigma})}{n}\right).$$

Moreover, for all $t \ge 1$ with probability $1-e^{-t}$

$$\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\| \lesssim \|\mathbf{\Sigma}\|\left[\sqrt{\frac{\mathtt{r}(\mathbf{\Sigma})}{n}} \bigvee \frac{\mathtt{r}(\mathbf{\Sigma})}{n} \bigvee \sqrt{\frac{t}{n}} \bigvee \frac{t}{n}\right].$$

Proof See [11, Theorem 6, Corollary 2]. $\square$

To deal with spectral projectors we need the following result which was proved in [12]. Let us introduce additional notations. We denote by $\widetilde{\boldsymbol{\Sigma}}$ an arbitrary perturbation of $\boldsymbol{\Sigma}$ and $\widetilde{\mathbf{E}} \stackrel{\text{def}}{=} \widetilde{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}$. Recall that

$$\mathbf{C}_r \stackrel{\text{def}}{=} \sum_{s \neq r} \frac{1}{\mu_r - \mu_s} \mathbf{P}_s.$$

Lemma 4 *Let $\widetilde{\boldsymbol{\Sigma}}$ be an arbitrary perturbation of $\boldsymbol{\Sigma}$ and let $\widetilde{\mathbf{P}}_r$ be the corresponding projector. The following bound holds:*

$$\|\widetilde{\mathbf{P}}_r - \mathbf{P}_r\| \le 4 \frac{\|\widetilde{\mathbf{E}}\|}{\overline{g}_r}.$$

Moreover, $\widetilde{\mathbf{P}}_r - \mathbf{P}_r = L_r(\widetilde{\mathbf{E}}) + S_r(\widetilde{\mathbf{E}})$, where $L_r(\widetilde{\mathbf{E}}) \stackrel{\text{def}}{=} \mathbf{C}_r \widetilde{\mathbf{E}} \mathbf{P}_r + \mathbf{P}_r \widetilde{\mathbf{E}} \mathbf{C}_r$ and

$$\|S_r(\widetilde{\mathbf{E}})\| \le 14 \left(\frac{\|\widetilde{\mathbf{E}}\|}{\overline{g}_r} \right)^2.$$

Proof See [12, Lemma 1]. □

Theorem 7 (Concentration results in $\mathbf{X}$-world) *Assume that the conditions of Theorem 1 hold. Then for all $t : 1 \le t \le n^{1/4}$ and*

$$\frac{\operatorname{tr} \boldsymbol{\Sigma}}{\overline{g}_r} \left(\sqrt{\frac{t}{n}} + \sqrt{\frac{\log p}{n}} \right) \lesssim 1, \tag{46}$$

the following bound holds with probability at least $1 - e^{-t}$

$$\left| \|\widehat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2 - \|L_r(\mathbf{E})\|_2^2 \right| \lesssim m_r \frac{\|\boldsymbol{\Sigma}\|^3 \mathtt{r}^3(\boldsymbol{\Sigma})}{\overline{g}_r^3} \left(\frac{t}{n} \right)^{3/2}.$$

Proof The proof follows from [12, Theorems 3, 5]. □

A.2 Concentration inequalities for sums of random variables and random matrices

In what follows for a vector $a = (a_1, \ldots, a_n)$ we denote $\|a\|_s \stackrel{\text{def}}{=} \left(\sum_{k=1}^n |a_k|^s \right)^{1/s}$. For a random variable X and $r > 0$ we define the ψ_r-norm by

$$\|X\|_{\psi_r} \stackrel{\text{def}}{=} \inf\{C > 0 : \mathbb{E} \exp(X/C)^r \le 2\}.$$

If a random variable X is such that for any $p \ge 1$, $\mathbb{E}^{1/p} |X|^p \le p^{1/r} K$, for some $K > 0$, then $\|X\|_{\psi_r} \le cK$ where $c > 0$ is a numerical constant.

Lemma 5 *Let $X, X_i, i = 1, \dots, n$ be i.i.d. random variables with $\mathbb{E}\, X = 0$ and $\|X\|_{\psi_r} \le 1$, $1 \le r \le 2$. Then there exists some absolute constant $C > 0$ such that for all $p \ge 1$*

$$\mathbb{E}\left|\sum_{k=1}^{n} a_k X_k\right|^p \le (Cp)^{p/2}\|a\|_2^p + (Cp)^p\|a\|_{r_*}^p,$$

where $a = (a_1, \dots, a_n)$ and $1/r + 1/r_* = 1$.

Proof See [1, Lemma 3.6]. □

Lemma 6 *If $0 < s < 1$ and $X_1, \dots, X_n$ are independent random variables satisfying $\|X\|_{\psi_s} \le 1$, then for all $a = (a_1, \dots, a_n) \in \mathbb{R}^n$ and $p \ge 2$*

$$\mathbb{E}\left|\sum_{k=1}^{n} a_k X_k\right|^p \le (Cp)^{p/2}\|a\|_2^p + C_s p^{p/s}\|a\|_p^p.$$

Moreover, for $s \ge 1/2$, C_s is bounded by some absolute constant.

Proof See [1, Lemma 3.7]. □

Lemma 7 *Let $\eta_1, \dots, \eta_n$ be i.i.d. standard normal random variables. For all $t \ge 1$*

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} a_i(\eta_i^4 - 3)\right| \gtrsim t^2\|a\|_2\right) \le e^{-t}. \tag{47}$$

Moreover, if $\overline{\eta}_1, \dots, \overline{\eta}_n$ are i.i.d. standard normal random variables and independent of $\eta_1, \dots, \eta_n$ then

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} a_i(\eta_i^2\overline{\eta}_i^2 - 1)\right| \gtrsim t^2\|a\|_2\right) \le e^{-t}. \tag{48}$$

Proof We prove (48) only. The proof of (47) is similar. Let $\epsilon_i, i = 1, \dots, n$, be i.i.d. Rademacher r.v. Denote $\xi_i \overset{\text{def}}{=} \eta_i^2\overline{\eta}_i^2 - 1, i = 1, \dots, n$. Applying Lemma 6 with $s = 1/2$ we write

$$\mathbb{E}\,|\sum_{i=1}^{n} a_i\xi_i|^p \le 2^p\,\mathbb{E}\,|\sum_{i=1}^{n} a_i\epsilon_i\xi_i|^p \le C^p p^{p/2}\|a\|_2^p + C^p p^{2p}\|a\|_p^p \le C^p p^{2p}\|a\|_2^p.$$

From Markov's inequality

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} a_i(\eta_i^2\overline{\eta}_i^2 - 1)\right| \ge t^2\|a\|_2\right) \le \frac{C^p p^{2p}}{t^{2p}}.$$

Taking $p = t/(Ce)^{1/2}$ we finish the proof of the lemma. □

Lemma 8 (Matrix Gaussian series) *Consider a finite sequence $\{\mathbf{A}_k\}$ of fixed, self-adjoint matrices with dimension d, and let $\{\xi_k\}$ be a finite sequence of independent standard normal random variables. Compute the variance parameter*

$$\sigma^2 \stackrel{\text{def}}{=} \left\| \sum_{k=1}^{n} \mathbf{A}_k^2 \right\|.$$

Then, for all $t \geq 0$,

$$\mathbb{P}\left(\left\| \sum_{k=1}^{n} \xi_k \mathbf{A}_k \right\| \geq t \right) \leq 2d \, \exp(-t^2/2\sigma^2).$$

Proof See in [20, Theorem 4.1]. □

Lemma 9 (Matrix Bernstein inequality) *Consider a finite sequence $\mathbf{X}_k$ of independent, random, self-adjoint matrices with dimension d. Assume that $\mathbb{E}\,\mathbf{X}_k = 0$ and $\lambda_{\max}(\mathbf{X}_k) \leq R$ almost surely. Compute the norm of the total variance,*

$$\sigma^2 \stackrel{\text{def}}{=} \left\| \sum_{k=1}^{n} \mathbb{E}\,\mathbf{X}_k^2 \right\|.$$

Then the following inequalities hold for all $t \geq 0$:

$$\mathbb{P}\left(\lambda_{\max}\left(\sum_{k=1}^{n} \mathbf{X}_k \right) \geq t \right) \leq d \, \exp\left(-\frac{t^2/2}{\sigma^2 + Rt/3} \right).$$

Moreover, if $\mathbb{E}\,\mathbf{X}_k = 0$ and $\mathbb{E}\,\mathbf{X}_k^p \preceq \frac{p!}{2} R^{p-2} \mathbf{A}_k^2$ then the following inequalities hold for all $t \geq 0$:

$$\mathbb{P}\left(\lambda_{\max}\left(\sum_{k=1}^{n} \mathbf{X}_k \right) \geq t \right) \leq d \, \exp\left(-\frac{t^2/2}{\tilde{\sigma}^2 + Rt} \right),$$

where

$$\tilde{\sigma}^2 \stackrel{\text{def}}{=} \left\| \sum_{k=1}^{n} \mathbf{A}_k^2 \right\|.$$

Proof See in [20, Theorem 6.1]. □

A.3 Auxiliary lemma

Lemma 10 *Assume that Z_1, Z_2 be i.i.d. and $\mathcal{N}(0, 1)$. Let λ_1, λ_2 be any positive numbers and $b \neq 0$. There exists an absolute constant c such that*

$$\left| \int_{-T}^{T} e^{itb}\, \mathbb{E} \exp\Big(it\big[\lambda_1 Z_1^2 + \lambda_2 Z_2^2\big] \Big) dt \right| \le \frac{c}{\sqrt{\lambda_1 \lambda_2}}. \tag{49}$$

Proof Denote the l.h.s. of (49) by I'. Using Euler's formula for complex exponential function we get for positive g and any $d \in \mathbb{R}$

$$g + id = \sqrt{g^2 + d^2} e^{i\zeta}, \quad \zeta = \arcsin \frac{d}{\sqrt{g^2 + d^2}}.$$

Hence, by (34) we get

$$I' = \left| \int_{-T}^{T} \exp\left(itb + \sum_{k=1}^{2} \frac{i\phi_k}{2} \right) \prod_{k=1}^{2} \left(1 + 4t^2 \lambda_k^2\right)^{-1/4} \right|,$$

where $\phi_k \overset{\text{def}}{=} \phi_k(t) \overset{\text{def}}{=} \arcsin\left(2\lambda_k t/(1 + 4t^2\lambda_k^2)^{\frac{1}{2}}\right)$. Since $\prod_{k=1}^{2} \left(1 + 4t^2\lambda_k^2\right)^{-1/4}$ is even function and $\phi_k(t)$, $k = 1, 2$, is odd function of t, we may rewrite I' as follows

$$I' = \frac{2}{\sqrt{\lambda_1 \lambda_2}} \left| \int_{0}^{T} \frac{1}{t} \sin\left(tb + \sum_{k=1}^{2} \frac{1}{2}\left(\phi_k - \frac{\pi}{2}\right)\right) \prod_{k=1}^{2} \left(\frac{t^2\lambda_k^2}{1 + 4t^2\lambda_k^2} \right)^{1/4} dt \right|.$$

We note that

$$\prod_{k=1}^{2} \left(\frac{t^2\lambda_k^2}{1 + 4t^2\lambda_k^2} \right)^{1/4} \le \sqrt{|t|\lambda_2}$$

Hence, to prove (49) it is enough to show that

$$I'' \overset{\text{def}}{=} \left| \int_{1/\lambda_2}^{T} \frac{1}{t} \sin\left(tb + \sum_{k=1}^{2} \frac{1}{2}\left(\phi_k - \frac{\pi}{2}\right)\right) \prod_{k=1}^{2} \left(\frac{t^2\lambda_k^2}{1 + 4t^2\lambda_k^2} \right)^{1/4} dt \right| \le c.$$

We may rewrite I'' as follows

$$I'' \le I_1'' + \cdots + I_4'',$$

where

$$
\begin{aligned}
I_1'' &\stackrel{\text{def}}{=} \left| \int_{1/\lambda_2}^{T} \frac{1}{t} \sin(tb)\, dt \right|, \\
I_2'' &\stackrel{\text{def}}{=} \left| \int_{1/\lambda_2}^{T} \frac{1}{t} \left[\sin\left(tb + \sum_{k=1}^{2} \frac{1}{2}\left(\phi_k - \frac{\pi}{2}\right)\right) - \sin(tb) \right] dt \right|, \\
I_3'' &\stackrel{\text{def}}{=} \left| \int_{1/\lambda_2}^{T} \frac{1}{t} \sin\left(tb + \sum_{k=1}^{2} \frac{1}{2}\left(\phi_k - \frac{\pi}{2}\right)\right) \left[1 - \left(\frac{t^2\lambda_1^2}{1+4t^2\lambda_1^2} \right)^{1/4} \right] dt \right|, \\
I_4'' &\stackrel{\text{def}}{=} \left| \int_{1/\lambda_2}^{T} \frac{1}{t} \sin\left(tb + \sum_{k=1}^{2} \frac{1}{2}\left(\phi_k - \frac{\pi}{2}\right)\right) \right. \\
&\qquad \left. \times \left[1 - \left(\frac{t^2\lambda_2^2}{1+4t^2\lambda_2^2} \right)^{1/4} \right] \left(\frac{t^2\lambda_1^2}{1+4t^2\lambda_1^2} \right)^{1/4} dt \right|.
\end{aligned}
$$

The bound $I_1'' \le c$ is true since for any positive A and B we have

$$
\left| \int_A^B \frac{\sin t}{t}\, dt \right| \le 2 \int_0^{\pi} \frac{\sin t}{t}\, dt.
$$

To estimate I_2'' we shall use the following inequalities

$$
\begin{aligned}
&|\sin(x+y) - \sin(x)| \le |y| \quad \text{for all } x, y \in \mathbb{R}, \\
&0 \le \frac{\pi}{2} - \arcsin(1-z) \le 2^{\frac{3}{2}} z^{\frac{1}{2}} \quad \text{for } 0 \le z \le 1.
\end{aligned}
$$

Applying these inequalities we get that

$$
\left| \sin\left(tb + \sum_{k=1}^{2} \frac{1}{2}\left(\phi_k - \frac{\pi}{2}\right)\right) - \sin(tb) \right| \le \frac{c'}{\lambda_2^2 t^2},
$$

where c' is some absolute constant. Hence,

$$
I_3'' \le \frac{c'}{\lambda_2^2} \int_{1/\lambda_2}^{\infty} \frac{1}{t^3}\, dt \le c.
$$

The estimates for I_3'' and I_4'' are similar. For simplicity we estimate I_3'' only. Applying the following inequality

$$
0 \le 1 - \left(\frac{t^2\lambda_k^2}{1+4t^2\lambda_k^2} \right)^{1/4} \le \frac{1}{4t^2\lambda_2^2}, \quad k = 1, 2,
$$

we obtain that

$$I_3'' \le \frac{c''}{\lambda_2^2} \int_{1/\lambda_2}^{\infty} \frac{1}{t^3}\, dt \le c,$$

where c'' is some absolute constant. □

References

1. Adamczak, R., Litvak, A., Pajor, A., Tomczak-Jaegermann, N.: Restricted isometry property of matrices with independent columns and neighborly polytopes by random sampling. Constr. Approx. **34**(1), 61–88 (2011)
2. Bengio, I., Courville, A.: Deep Learning. MIT Press, Cambridge (2016)
3. Bentkus, V.: A Lyapunov-type bound in R^d. Probab. Theory Appl. **49**(2), 311–323 (2005)
4. Chernozhukov, V., Chetverikov, D., Kato, K.: Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. Ann. Stat. **41**(6), 2786–2819 (2013)
5. Chernozhukov, V., Chetverikov, D., Kato, K.: Comparison and anti-concentration bounds for maxima of Gaussian random vectors. Probab. Theory Relat. Fields **162**(1–2), 47–70 (2015)
6. Chernozhukov, V., Chetverikov, D., Kato, K.: Central limit theorems and bootstrap in high dimensions. Ann. Probab. **45**(4), 2309–2352 (2017)
7. Chung, K.: A Course in Probability Theory, 3rd edn. Academic Press Inc., San Diego (2001)
8. Götze, F., Ulyanov, V.: Uniform approximation in the CLT for balls in Euclidean spaces. Preprint 00-0034, SFB 343, Univ. Bielefeld (2000)
9. Götze F, Naumov A, Spokoiny V, Ulyanov V (2019) Large ball probabilities, Gaussian comparison and anti-concentration. Bernoulli 25. arXiv:1708.08663v2
10. Holtz, M.: Sparse Grid Quadrature in High Dimensions with Applications in Finance and Insurance, Lecture Notes in Computational Science and Engineering, vol. 77. Springer, Berlin (2010)
11. Koltchinskii, V., Lounici, K.: Concentration inequalities and moment bounds for sample covariance operators. Bernoulli **23**(1), 110–133 (2017)
12. Koltchinskii, V., Lounici, K.: Normal approximation and concentration of spectral projectors of sample covariance. Ann. Stat. **45**(1), 121–157 (2017)
13. Li, Y., Hsing, T.: Deciding the dimension of effective dimension reduction space for functional and high-dimensional data. Ann. Stat. **38**(5), 3028–3062 (2010)
14. Marčenko, V., Pastur, L.: Distribution of eigenvalues in certain sets of random matrices. Mat. Sb. (N.S.) **72**(114), 507–536 (1967)
15. Markus, A.: The eigen- and singular values of the sum and product of linear operators. Russ. Math. Surv. **19**(4), 91–120 (1964)
16. Prokhorov, Y., Ulyanov, V.: Some approximation problems in statistics and probability. In: Limit Theorems in Probability, Statistics and Number Theory, Springer Proc. Math. Stat., vol. 42, pp. 235–249. Springer, Heidelberg (2013)
17. Rosenthal, H.: On the subspaces of L^p $(p>2)$ spanned by sequences of independent random variables. Isr. J. Math. **8**, 273–303 (1970)
18. Shao, J.: Mathematical Statistics. Springer, Berlin (2003)
19. Spokoiny, V., Zhilova, M.: Bootstrap confidence sets under model misspecification. Ann. Stat. **43**(6), 2653–2675 (2015)
20. Tropp, J.: User-friendly tail bounds for sums of random matrices. Found. Comput. Math. **12**(4), 389–434 (2012)
21. Tsybakov, A.: Introduction to Nonparametric Estimation. Springer, New York (2008)
22. van Handel, R.: Structured random matrices. In: Convexity and Concentration, vol. 161, pp. 107–165, IMA. Springer (2017)
23. Vershynin, R.: Introduction to the non-asymptotic analysis of random matrices. In: Eldar, Y.C., Kutyniok, G. (eds.) Compressed Sensing, pp. 210–268. Cambridge University Press, Cambridge (2012)
24. Wang, X., Sloan, I.: Why are high-dimensional finance problems often of low effective dimension? SIAM J. Sci. Comput. **27**(1), 159–183 (2005)