The Annals of Statistics
2017, Vol. 45, No. 1, 121–157
DOI: 10.1214/16-AOS1437

NORMAL APPROXIMATION AND CONCENTRATION OF SPECTRAL PROJECTORS OF SAMPLE COVARIANCE

BY VLADIMIR KOLTCHINSKII[1] AND KARIM LOUNICI[2]

Georgia Institute of Technology

Let $X, X_1, \dots, X_n$ be i.i.d. Gaussian random variables in a separable Hilbert space $\mathbb{H}$ with zero mean and covariance operator $\Sigma = \mathbb{E}(X \otimes X)$, and let $\hat{\Sigma} := n^{-1} \sum_{j=1}^n (X_j \otimes X_j)$ be the sample (empirical) covariance operator based on $(X_1, \dots, X_n)$. Denote by P_r the spectral projector of Σ corresponding to its rth eigenvalue μ_r and by $\hat{P}_r$ the empirical counterpart of P_r. The main goal of the paper is to obtain tight bounds on

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

where $\|\cdot\|_2$ denotes the Hilbert–Schmidt norm and Φ is the standard normal distribution function. Such accuracy of normal approximation of the distribution of squared Hilbert–Schmidt error is characterized in terms of so-called effective rank of Σ defined as $\mathbf{r}(\Sigma) = \frac{\mathrm{tr}(\Sigma)}{\|\Sigma\|_\infty}$, where $\mathrm{tr}(\Sigma)$ is the trace of Σ and $\|\Sigma\|_\infty$ is its operator norm, as well as another parameter characterizing the size of $\mathrm{Var}(\|\hat{P}_r - P_r\|_2^2)$. Other results include nonasymptotic bounds and asymptotic representations for the mean squared Hilbert–Schmidt norm error $\mathbb{E}\|\hat{P}_r - P_r\|_2^2$ and the variance $\mathrm{Var}(\|\hat{P}_r - P_r\|_2^2)$, and concentration inequalities for $\|\hat{P}_r - P_r\|_2^2$ around its expectation.

1. Introduction. Let X be a mean zero Gaussian random vector in a separable Hilbert space $\mathbb{H}$ with covariance operator $\Sigma = \mathbb{E}(X \otimes X)$ and let $X_1, \dots, X_n$ be a sample of n i.i.d. copies of X. The sample covariance operator $\hat{\Sigma} = \hat{\Sigma}_n$ is defined as follows: $\hat{\Sigma} := \hat{\Sigma}_n := n^{-1} \sum_{j=1}^n (X_j \otimes X_j)$. Denote by μ_r the rth eigenvalue of Σ (in a decreasing order) and by P_r the corresponding spectral projector of Σ (i.e., the orthogonal projector on the eigenspace of eigenvalue μ_r). Let $\hat{P}_r$ denote properly defined empirical counterpart of P_r (see Section 2.2 for a precise definition). The main goal of the paper is to obtain a tight bound on the accuracy of normal approximation of the distribution of the squared Hilbert–Schmidt norm error $\|\hat{P}_r - P_r\|_2^2$ of the estimator $\hat{P}_r$. Another goal is to provide bounds on the

Received September 2015; revised January 2016.
[1]Supported in part by NSF Grants DMS-15-09739, DMS-12-07808, CCF-1523768 and CCF-1415498.
[2]Supported in part by Simons Grant 315477 and NSF Career Grant DMS-14-54515.
MSC2010 subject classifications. 62H12.
Key words and phrases. Sample covariance, spectral projectors, effective rank, principal component analysis, concentration inequalities, normal approximation, perturbation theory.

risk $\mathbb{E}\|\hat{P}_r - P_r\|_2^2$ of this estimator as well as nonasymptotic bounds on concentration of random variables $\|\hat{P}_r - P_r\|_2^2$ around its expectation. These bounds will be expressed in terms of natural complexity parameters of the problem, the most important one being the so-called *effective rank* $\mathbf{r}(\Sigma)$ that has been recently used in the literature (see [2, 18, 20]).

DEFINITION 1. The following quantity $\mathbf{r}(\Sigma) := \frac{\operatorname{tr}(\Sigma)}{\|\Sigma\|_\infty}$ will be called the effective rank of Σ.

Here, $\operatorname{tr}(\Sigma)$ denotes the trace of Σ and $\|\Sigma\|_\infty$ denotes its operator norm. The above definition clearly implies that $\mathbf{r}(\Sigma) \le \operatorname{rank}(\Sigma)$. A recent result by Koltchinskii and Lounici (see [14]) shows that, in the Gaussian case, the size of the operator norm error $\|\hat{\Sigma} - \Sigma\|_\infty$ of sample covariance $\hat{\Sigma}$ is completely characterized by $\|\Sigma\|_\infty$ and $\mathbf{r}(\Sigma)$. This makes the effective rank $\mathbf{r}(\Sigma)$ the crucial complexity parameter of the problems of estimation of covariance and its spectral characteristics (its principal components) that allows one to study principal component analysis (PCA) problems in a unified dimension-free framework that includes their high-dimensional and infinite-dimensional versions (functional PCA, kernel PCA, etc.). Our goal is to study the problem in a "high-complexity setting", where both the sample size n and the effective rank $\mathbf{r}(\Sigma)$ are large, although our primary focus is on the case when $\mathbf{r}(\Sigma) = o(n)$ which implies operator norm consistency of both $\hat{\Sigma}$ and $\hat{P}_r$. This setting is much closer to high-dimensional covariance estimation and PCA problems than to standard results on PCA in Hilbert spaces with a fixed value of $\operatorname{tr}(\Sigma)$ (see, e.g., [4]) that are commonly used in the literature on functional PCA and kernel PCA. It includes, in particular, high-dimensional *spiked covariance models* (see [9, 10, 19]) in which

$$\Sigma = \sum_{j=1}^{m} s_j^2(\theta_j \otimes \theta_j) + \sigma^2 P_p, \tag{1.1}$$

where $\{\theta_j\}$ is an orthonormal basis of $\mathbb{H}$, $s_1^2 > s_2^2 > \cdots > s_m^2$ are the variances of m independent components of the "signal", σ^2 is the variance of the noise components and $P_p := \sum_{j=1}^{p}(\theta_j \otimes \theta_j)$ is the orthogonal projector on the linear span of the vectors $\theta_1, \dots, \theta_p$, where $p > m$. This models the covariance of a Gaussian signal with m independent components observed in an independent Gaussian white noise. It is usually assumed that the number of components m and the variances $s_1^2, \dots, s_m^2, \sigma^2$ are fixed, but the overall dimension of the problem $p = p_n \to \infty$ as $n \to \infty$ is large, implying that

$$\operatorname{tr}(\Sigma) = \sum_{j=1}^{m} s_j^2 + \sigma^2 p \sim \sigma^2 p \to \infty \qquad \text{as } n \to \infty$$

and $\mathbf{r}(\Sigma) \sim \frac{\sigma^2}{s_1^2 + \sigma^2} p$. Estimation of the components of the "signal" $\theta_1, \dots, \theta_m$ is viewed as PCA for unknown covariance Σ. It is common to consider a sequence of high-dimensional problems in spaces $\mathbb{R}^p$, $p = p_n$ (rather than explicitly embed the spaces $\mathbb{R}^p$ into an infinite dimensional Hilbert space $\mathbb{H}$). To assess the performance of the PCA, the loss function $L(a,b) := 2(1 - |\langle a, b\rangle|)$, where $a, b \in \mathbb{R}^p$ are unit vectors, was used in [1]. A closely related loss function is defined by $L'(a,b) := \|a \otimes a - b \otimes b\|_2^2 = 2(1 - \langle a, b\rangle^2)$; see, for instance, [3, 18, 21]. In the case of spiked covariance model with $\sigma^2 = 1$ and $\frac{p}{n} \to 0$ as $n \to \infty$, the following asymptotic representation of the risk holds [1]:

$$\mathbb{E}L(\hat{\theta}_j, \theta_j) = \biggl[\frac{(p-m)(1+s_j^2)}{n s_j^4} + \frac{1}{n}\sum_{k \neq j} \frac{(1+s_j^2)(1+s_k^2)}{(s_j^2 - s_k^2)^2}\biggr]\bigl(1 + o(1)\bigr), \quad j = 1, \dots, m. \tag{1.2}$$

In this paper, we are not making any structural assumptions on the covariance operator Σ, such as the spiked covariance model, sparsity, etc., but rather study the problem in terms of complexity parameter $\mathbf{r}(\Sigma)$. We derive representations of the Hilbert–Schmidt risk $\mathbb{E}\|\hat{P}_r - P_r\|_2^2$ of empirical spectral projectors in the case when $\mathbf{r}(\Sigma) = o(n)$ that imply representation (1.2) for spiked covariance model. Specifically, we prove that

$$\mathbb{E}\|\hat{P}_r - P_r\|_2^2 = \bigl(1 + o(1)\bigr)\frac{A_r(\Sigma)}{n}, \tag{1.3}$$

where $A_r(\Sigma) = 2\operatorname{tr}(P_r \Sigma P_r)\operatorname{tr}(C_r \Sigma C_r)$ and the operator C_r is defined as $C_r := \sum_{s \neq r} \frac{P_s}{\mu_r - \mu_s}$. In addition, we show that

$$\operatorname{Var}\bigl(\|\hat{P}_r - P_r\|_2^2\bigr) = \bigl(1 + o(1)\bigr)\frac{B_r^2(\Sigma)}{n^2}, \tag{1.4}$$

where $B_r(\Sigma) := 2\sqrt{2}\|P_r \Sigma P_r\|_2 \|C_r \Sigma C_r\|_2$, and derive concentration bounds for random variable $\|\hat{P}_r - P_r\|_2^2$ around its expectation. One of the main results of the paper is the following bound on the accuracy of normal approximation of random variable $\|\hat{P}_r - P_r\|_2^2$ that holds under rather mild assumptions:

$$\begin{aligned} &\sup_{x \in \mathbb{R}}\biggl|\mathbb{P}\biggl\{\frac{\|\hat{P}_r - P_r\|_2^2 - \mathbb{E}\|\hat{P}_r - P_r\|_2^2}{\operatorname{Var}^{1/2}(\|\hat{P}_r - P_r\|_2^2)} \le x\biggr\} - \Phi(x)\biggr| \\ &\quad \le C\biggl[\frac{1}{B_r(\Sigma)} + \frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)\sqrt{n}}\sqrt{\log\biggl(\frac{B_r(\Sigma)\sqrt{n}}{\mathbf{r}(\Sigma)} \vee 2\biggr)}\biggr], \end{aligned} \tag{1.5}$$

where $\Phi(x)$ denotes the standard normal distribution function. This bound implies that the distribution of random variable $\frac{\|\hat{P}_r - P_r\|_2^2 - \mathbb{E}\|\hat{P}_r - P_r\|_2^2}{\operatorname{Var}^{1/2}(\|\hat{P}_r - P_r\|_2^2)}$ is asymptotically standard normal as soon as $n \to \infty$, $B_r(\Sigma) \to \infty$ and $\frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)\sqrt{n}} \to 0$ which, in particular, implies that $\mathbf{r}(\Sigma) = o(n)$ [see (4.5), (4.6)].

In our paper [13], asymptotics and concentration bounds for bilinear forms $\langle \hat{P}_r u, v\rangle, u, v \in \mathbb{H}$ of empirical spectral projectors $\hat{P}_r$ were studied in a similar setting under the assumption $\mathbf{r}(\Sigma) = o(n)$.

Note that in the case of spiked covariance model, the classical PCA is known to yield inconsistent estimators of the eigenvectors when condition $p = o(n)$ [which is a special case of our condition $\mathbf{r}(\Sigma) = o(n)$] fails and $\frac{p}{n} \to c > 0$; see, for example, [10]. In [1], a thresholding procedure in spirit of diagonal thresholding of Johnstone and Lu [10] was proposed and it was proved that it achieves optimality in the minimax sense for the loss $L(\cdot, \cdot)$ under sparsity conditions on the eigenvectors of Σ.

Throughout the paper, for $A, B > 0$, the notation $A \lesssim B$ means that there exists an absolute constant $C > 0$ such that $A \le CB$. Similarly, $A \gtrsim B$ means that $A \ge CB$ for an absolute constant $C > 0$ and $A \asymp B$ means that $A \lesssim B$ and $A \gtrsim B$. In the cases when the constant C in the above bounds might depend on some parameter(s), say, γ, and we want to emphasize this dependence, we will write $A \lesssim_\gamma B$, $A \gtrsim_\gamma B$, or $A \asymp_\gamma B$. Also, throughout the paper (as it was already done in the Introduction), $\|\cdot\|_2$ denotes the Hilbert–Schmidt norm and $\|\cdot\|_\infty$ the operator norm of operators acting in $\mathbb{H}$. With a minor abuse of notation, $\langle\cdot,\cdot\rangle$ denotes both the inner product of $\mathbb{H}$ and the Hilbert–Schmidt inner product. We will also use the sign $\otimes$ to denote the tensor product. For instance, for $u, v \in \mathbb{H}$, $u \otimes v$ is a linear operator in $\mathbb{H}$ defined as follows: $(u \otimes v)x = u\langle v, x\rangle, x \in \mathbb{H}$.

In what follows, we will frequently prove exponential bounds for certain random variables, say, ξ, of the following type: for some constant $C > 0$ and for all $t \ge 1$, with probability at least $1 - e^{-t}$, $\xi \le C\sqrt{t}$. Often it will be proved instead that the inequality holds with probability, say, $1 - 2e^{-t}$. In such cases, it is easy to rewrite the probability bound in the initial form by changing the value of the constant C. For instance, replacing t by $t + \log 2$ allows one to claim that with probability $1 - e^{-t}$, $\xi \le C\sqrt{t + \log 2} \le C(1 + \log 2)^{1/2}\sqrt{t}$ that holds for all $t \ge 1$. In such cases, it will be said without further explanation that probability bound $1 - 2e^{-t}$ can be replaced by $1 - e^{-t}$ by adjusting the constants.

2. Preliminaries.

In this section, we discuss recent bounds on the operator norm $\|\hat{\Sigma}_n - \Sigma\|_\infty$ obtained in [14] and several well-known results of perturbation theory used throughout the paper (see also [13]).

2.1. *Bounds on the operator norm* $\|\hat{\Sigma}_n - \Sigma\|_\infty$. In [14], it was proved that, in the Gaussian case, moment bounds and concentration inequalities for the operator norm $\|\hat{\Sigma} - \Sigma\|_\infty$ are completely characterized by the operator norm $\|\Sigma\|_\infty$ and the effective rank $\mathbf{r}(\Sigma)$. More precisely, the following theorems hold.

THEOREM 1. *Let $X, X_1, \dots, X_n$ be i.i.d. centered Gaussian random vectors in $\mathbb{H}$ with covariance $\Sigma = \mathbb{E}(X \otimes X)$. Then, for all $p \ge 1$,*

$$\mathbb{E}^{1/p}\|\hat{\Sigma} - \Sigma\|_\infty^p \asymp_p \|\Sigma\|_\infty \max\left\{\sqrt{\frac{\mathbf{r}(\Sigma)}{n}}, \frac{\mathbf{r}(\Sigma)}{n}\right\}. \tag{2.1}$$

THEOREM 2. *Let $X, X_1, \dots, X_n$ be i.i.d. centered Gaussian random vectors in $\mathbb{H}$ with covariance $\Sigma = \mathbb{E}(X \otimes X)$. Then there exists a constant $C > 0$ such that for all $t \geq 1$ with probability at least $1 - e^{-t}$,*

$$\big| \|\hat{\Sigma} - \Sigma\|_\infty - \mathbb{E}\|\hat{\Sigma} - \Sigma\|_\infty \big| \leq C\|\Sigma\|_\infty \bigg[\bigg(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee 1 \bigg) \sqrt{\frac{t}{n}} \vee \frac{t}{n} \bigg]. \tag{2.2}$$

As a consequence of this bound and (2.1), *with some constant $C > 0$ and with the same probability*

$$\|\hat{\Sigma} - \Sigma\|_\infty \leq C\|\Sigma\|_\infty \bigg[\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \frac{\mathbf{r}(\Sigma)}{n} \vee \sqrt{\frac{t}{n}} \vee \frac{t}{n} \bigg]. \tag{2.3}$$

2.2. *Perturbation theory.* Several simple and well-known facts on perturbations of linear operators (see Kato [11]) will be stated in a form suitable for our purposes. The proofs of some of these facts that seem not to be readily available in the literature were given in [13] (see also Koltchinskii [15] and Kneip and Utikal [12] for some bounds in the same direction).

Let $\Sigma : \mathbb{H} \mapsto \mathbb{H}$ be a compact symmetric operator (in our case, the covariance operator of a random vector X in $\mathbb{H}$) with the spectrum $\sigma(\Sigma)$. The following spectral representation is well known to hold with the series converging in the operator norm: $\Sigma = \sum_{r \geq 1} \mu_r P_r$, where μ_r denotes distinct nonzero eigenvalues of Σ arranged in decreasing order and P_r the corresponding spectral projectors. Denote by $\sigma_i = \sigma_i(\Sigma)$ the eigenvalues of Σ arranged in nonincreasing order and repeated with their respective multiplicities. Let $\Delta_r = \{i : \sigma_i(\Sigma) = \mu_r\}$ and let $m_r := \operatorname{card}(\Delta_r)$ denote the multiplicity of μ_r. Define $g_r := g_r(\Sigma) := \mu_r - \mu_{r+1} > 0, r \geq 1$. Let $\bar{g}_r := \bar{g}_r(\Sigma) := \min(g_{r-1}, g_r)$ for $r \geq 2$ and $\bar{g}_1 := g_1$. The quantity $\bar{g}_r$ will be called *the rth spectral gap*, *or the spectral gap of eigenvalue μ_r.*

Let now $\tilde{\Sigma} := \Sigma + E$ be another compact symmetric operator in $\mathbb{H}$ with spectrum $\sigma(\tilde{\Sigma})$ and eigenvalues $\tilde{\sigma}_i = \sigma_i(\tilde{\Sigma}), i \geq 1$ (arranged in nonincreasing order and repeated with their multiplicities), where E is a perturbation of Σ. By Lidskii's inequality,

$$\sup_{j \geq 1} \big|\sigma_j(\Sigma) - \sigma_j(\tilde{\Sigma})\big| \leq \sup_{j \geq 1} \big|\sigma_j(E)\big| = \|E\|_\infty.$$

Thus, for all $r \geq 1$,

$$\inf_{j \notin \Delta_r} |\tilde{\sigma}_j - \mu_r| \geq \bar{g}_r - \sup_{j \geq 1} |\tilde{\sigma}_j - \sigma_j| \geq \bar{g}_r - \|E\|_\infty$$

and

$$\sup_{j \in \Delta_r} |\tilde{\sigma}_j - \mu_r| = \sup_{j \in \Delta_r} |\tilde{\sigma}_j - \sigma_j| \leq \|E\|_\infty.$$

Assuming that the perturbation E is small in the sense that $\|E\|_\infty < \frac{\bar{g}_r}{2}$, it is easy to conclude that all the eigenvalues $\tilde{\sigma}_j, j \in \Delta_r$ are covered by an interval

$$\big(\mu_r - \|E\|_\infty, \mu_r + \|E\|_\infty\big) \subset (\mu_r - \bar{g}_r/2, \mu_r + \bar{g}_r/2)$$

and the rest of the eigenvalues of $\tilde{\Sigma}$ are outside of the interval

$$\big(\mu_r - \big(\bar{g}_r - \|E\|_\infty\big), \mu_r + \big(\bar{g}_r - \|E\|_\infty\big)\big) \supset [\mu_r - \bar{g}_r/2, \mu_r + \bar{g}_r/2].$$

Moreover, under the assumption $\|E\|_\infty < \frac{1}{4}\min_{1\le s\le r} \bar{g}_s =: \bar{\delta}_r$, the set $\{\sigma_j(\tilde{\Sigma}) : j \in \bigcup_{s=1}^r \Delta_s\}$ of the largest eigenvalues of $\tilde{\Sigma}$ consists of r "clusters", the diameter of each cluster being strictly smaller than $2\bar{\delta}_r$ and the distance between any two clusters being larger than $2\bar{\delta}_r$. Thus, it is possible to identify clusters of eigenvalues of $\tilde{\Sigma}$ corresponding to each of the r largest distinct eigenvalues $\mu_s, s = 1, \dots, r$ of Σ. Let $\tilde{P}_r$ be the orthogonal projector on the direct sum of eigenspaces of $\tilde{\Sigma}$ corresponding to the eigenvalues $\tilde{\sigma}_j, j \in \Delta_r$ (to the rth cluster of eigenvalues of $\tilde{\Sigma}$). The following "partial resolvent" operator will be frequently used throughout the paper: $C_r := \sum_{s\ne r} \frac{1}{\mu_r - \mu_s} P_s$.

We will need a couple of lemmas proved in [13] (see Lemmas 1 and 4 therein).

LEMMA 1. *The following bound holds*:

$$\|\tilde{P}_r - P_r\|_\infty \le 4\frac{\|E\|_\infty}{\bar{g}_r}. \tag{2.4}$$

Moreover,

$$\tilde{P}_r - P_r = L_r(E) + S_r(E), \tag{2.5}$$

where

$$L_r(E) := C_r E P_r + P_r E C_r \tag{2.6}$$

and

$$\|S_r(E)\|_\infty \le 14\left(\frac{\|E\|_\infty}{\bar{g}_r}\right)^2. \tag{2.7}$$

LEMMA 2. *Let* $\gamma \in (0, 1)$ *and suppose that*

$$\delta \le \frac{1-\gamma}{1+\gamma}\frac{\bar{g}_r}{2}. \tag{2.8}$$

Suppose also that

$$\|E\|_\infty \le (1+\gamma)\delta \quad \text{and} \quad \|E'\|_\infty \le (1+\gamma)\delta. \tag{2.9}$$

Then there exists a constant $C_\gamma > 0$ *such that*

$$\|S_r(E) - S_r(E')\|_\infty \le C_\gamma \frac{\delta}{\bar{g}_r^2}\|E - E'\|_\infty. \tag{2.10}$$

3. Bounds on the risk of empirical spectral projectors. Let $\hat{P}_r$ be the orthogonal projector on the direct sum of eigenspaces of $\hat{\Sigma}$ corresponding to the eigenvalues $\{\sigma_j(\hat{\Sigma}), j \in \Delta_r\}$ (in other words, to the rth cluster of eigenvalues of $\hat{\Sigma}$, see Section 2.2).

We will state simple bounds for the bias $\mathbb{E}\hat{P}_r - P_r$ and the "variance" $\mathbb{E}\|\hat{P}_r - \mathbb{E}\hat{P}_r\|_2^2$ that immediately imply a representation of the risk $\mathbb{E}\|\hat{P}_r - P_r\|_2^2$.

Denote

$$A_r(\Sigma) := 2\,\mathrm{tr}(P_r \Sigma P_r)\,\mathrm{tr}(C_r \Sigma C_r). \tag{3.1}$$

It is easy to see that

$$A_r(\Sigma) \le 2\frac{m_r \mu_r}{\bar{g}_r^2}\|\Sigma\|_\infty \mathbf{r}(\Sigma) \tag{3.2}$$

and

$$A_r(\Sigma) \ge 2\left(\frac{m_r \mu_r}{\|\Sigma\|_\infty}\mathbf{r}(\Sigma) - \frac{m_r \mu_r^2}{\|\Sigma\|_\infty^2}\right), \tag{3.3}$$

which implies that

$$A_r(\Sigma) \asymp \mathbf{r}(\Sigma) \tag{3.4}$$

[assuming that $\|\Sigma\|_\infty$ and m_r are both bounded, $\bar{g}_r$ is bounded away from 0 and $\mathbf{r}(\Sigma) \to \infty$].

THEOREM 3. *The following bounds hold*:

1.

$$\|\mathbb{E}\hat{P}_r - P_r\|_\infty \lesssim \frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\left(\frac{\mathbf{r}(\Sigma)}{n} \vee \left(\frac{\mathbf{r}(\Sigma)}{n}\right)^2\right) \tag{3.5}$$

and

$$\|\mathbb{E}\hat{P}_r - P_r\|_2 \lesssim \sqrt{m_r}\frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\left(\frac{\mathbf{r}(\Sigma)}{n} \vee \left(\frac{\mathbf{r}(\Sigma)}{n}\right)^2\right). \tag{3.6}$$

2. *In addition*,

$$\mathbb{E}\|\hat{P}_r - \mathbb{E}\hat{P}_r\|_2^2 = \frac{A_r(\Sigma)}{n} + \rho_n, \tag{3.7}$$

where

$$|\rho_n| \le \frac{m_r\|\Sigma\|_\infty^4}{\bar{g}_r^4}\left(\left(\frac{\mathbf{r}(\Sigma)}{n}\right)^{3/2} \vee \left(\frac{\mathbf{r}(\Sigma)}{n}\right)^4\right). \tag{3.8}$$

3. *If* $\Sigma = \Sigma^{(n)}$, *the sequences* $\|\Sigma^{(n)}\|_\infty$ *and* $m_r = m_r^{(n)}$ *are both bounded*, $\bar{g}_r = \bar{g}_r^{(n)}$ *is bounded away from* 0, *and*

$$\mathbf{r}(\Sigma) \to \infty, \qquad \mathbf{r}(\Sigma) = o(n),$$

then the following representation holds:

$$\mathbb{E}\|\hat{P}_r - P_r\|_2^2 = \frac{A_r(\Sigma)}{n} + O\biggl(\biggl(\frac{\mathbf{r}(\Sigma)}{n}\biggr)^{3/2}\biggr) = \bigl(1 + o(1)\bigr)\frac{A_r(\Sigma)}{n}. \tag{3.9}$$

REMARK 1. In the case of spiked covariance model (1.1) for all $r = 1, \dots, m$,

$$A_r(\Sigma) = 2\biggl(\frac{(p-m)(s_r^2 + \sigma^2)}{s_r^4} + \sum_{1 \le j \le m, j \ne r} \frac{(s_j^2 + \sigma^2)(s_r^2 + \sigma^2)}{(s_r^2 - s_j^2)^2}\biggr).$$

Assuming that $m, s_1^2, \dots, s_m^2, \sigma^2$ are fixed, $p \to \infty$ and $p = o(n)$ as $n \to \infty$, it is easy to check that (3.9) implies bound (1.2) obtained in [1].

PROOF. Recall the following relationship (see Lemma 1):

$$\hat{P}_r - P_r = L_r(E) + S_r(E), \tag{3.10}$$

where $E := \hat{\Sigma} - \Sigma$, $L_r(E) := C_r E P_r + P_r E C_r$ and $S_r(E) := \hat{P}_r - P_r - L_r(E)$. Clearly, $C_r P_r = P_r C_r = 0$ (due to the orthogonality of P_r and $P_s, s \ne r$). Also, $P_r X$ and $C_r X$ are independent random variables (since, by the same orthogonality property, they are uncorrelated and X is Gaussian).

To prove Claim 1, note that, since $\mathbb{E}L_r(E) = 0$, we have $\mathbb{E}\hat{P}_r - P_r = \mathbb{E}S_r(E)$. Therefore, by bound (2.7) of Lemma 1, we get

$$\|\mathbb{E}\hat{P}_r - P_r\|_\infty \le \mathbb{E}\|S_r(E)\|_\infty \le 14\frac{\mathbb{E}\|E\|_\infty^2}{\bar{g}_r^2}. \tag{3.11}$$

Bound (3.5) now follows from Theorem 1. Bound (3.6) is also obvious since $\hat{P}_r, P_r$ are operators of rank m_r, $L_r(E)$ is of rank at most $2m_r$ and $S_r(E) = \hat{P}_r - P_r - L_r(E)$ is of rank at most $4m_r$. Thus, $\|S_r(E)\|_2 \lesssim \sqrt{m_r}\|S_r(E)\|_\infty$, and the result follows from the previous bounds.

To prove Claim 2, note that $\hat{P}_r - \mathbb{E}\hat{P}_r = L_r(E) + S_r(E) - \mathbb{E}S_r(E)$. Therefore,

$$\begin{aligned} \|\hat{P}_r - \mathbb{E}\hat{P}_r\|_2^2 &= \|L_r(E)\|_2^2 + \|S_r(E) - \mathbb{E}S_r(E)\|_2^2 \\ &\quad + 2\langle L_r(E), S_r(E) - \mathbb{E}S_r(E)\rangle. \end{aligned} \tag{3.12}$$

The following representations are obvious:

$$C_r E P_r = n^{-1}\sum_{j=1}^n C_r X_j \otimes P_r X_j, \qquad P_r E C_r = n^{-1}\sum_{j=1}^n P_r X_j \otimes C_r X_j. \tag{3.13}$$

Note that, by (3.13), due to orthogonality of $C_r E P_r$, $P_r E C_r$ and due to independence of $P_r X, C_r X$,

$$\begin{aligned}
\mathbb{E}\|L_r(E)\|_2^2 &= \mathbb{E}\|C_r E P_r + P_r E C_r\|_2^2 = \mathbb{E}(\|C_r E P_r\|_2^2 + \|P_r E C_r\|_2^2) \\
&= 2\mathbb{E}\|C_r E P_r\|_2^2 = 2\mathbb{E}\left\| n^{-1}\sum_{j=1}^n P_r X_j \otimes C_r X_j \right\|_2^2 \\
&= \frac{2\mathbb{E}\|P_r X \otimes C_r X\|_2^2}{n} = \frac{2\mathbb{E}\|P_r X\|^2\|C_r X\|^2}{n} \\
&= \frac{2\mathbb{E}\|P_r X\|^2\mathbb{E}\|C_r X\|^2}{n} = \frac{2\operatorname{tr}(P_r \Sigma P_r)\operatorname{tr}(C_r \Sigma C_r)}{n} = \frac{A_r(\Sigma)}{n}.
\end{aligned} \tag{3.14}$$

Next, note that $\mathbb{E}\|S_r(E) - \mathbb{E}S_r(E)\|_2^2 \leq \mathbb{E}\|S_r(E)\|_2^2$. Recall that $S_r(E)$ is of rank $\leq 4m_r$ and $\|S_r(E)\|_2^2 \leq 4m_r\|S_r(E)\|_\infty^2$. Quite similarly to (3.11), one can prove that $\mathbb{E}\|S_r(E)\|_\infty^2 \lesssim \frac{1}{\bar{g}_r^4}\mathbb{E}\|E\|_\infty^4$. Therefore, by Theorem 1, we get

$$\mathbb{E}\|S_r(E) - \mathbb{E}S_r(E)\|_2^2 \lesssim m_r \frac{\|\Sigma\|_\infty^4}{\bar{g}_r^4}\left(\left(\frac{\mathbf{r}(\Sigma)}{n}\right)^2 \vee \left(\frac{\mathbf{r}(\Sigma)}{n}\right)^4\right). \tag{3.15}$$

As a consequence of (3.2), (3.14) and (3.15), it easily follows that

$$\begin{aligned}
\mathbb{E}|\langle L_r(E), S_r(E) - \mathbb{E}S_r(E)\rangle| &\leq \mathbb{E}^{1/2}\|L_r(E)\|_2^2\mathbb{E}^{1/2}\|S_r(E) - \mathbb{E}S_r(E)\|_2^2 \\
&\lesssim \sqrt{\frac{A_r(\Sigma)}{n}}\sqrt{m_r}\frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\left(\frac{\mathbf{r}(\Sigma)}{n} \vee \left(\frac{\mathbf{r}(\Sigma)}{n}\right)^2\right) \\
&\lesssim m_r \frac{\|\Sigma\|_\infty^3}{\bar{g}_r^3}\left(\left(\frac{\mathbf{r}(\Sigma)}{n}\right)^{3/2} \vee \left(\frac{\mathbf{r}(\Sigma)}{n}\right)^{5/2}\right)
\end{aligned} \tag{3.16}$$

(3.7) and (3.8) now follow from (3.12), (3.14), (3.15) and (3.16).

We now prove Claim 3. Using (2.5), we easily get

$$\|\hat{P}_r - P_r\|_2^2 = \|L_r(E)\|_2^2 + \|S_r(E)\|_2^2 + 2\langle L_r(E), S_r(E)\rangle.$$

In view of (3.14), we have $\mathbb{E}\|L_r(E)\|_2^2 = \frac{A_r(\Sigma)}{n}$. By an argument similar to that of (3.15) and (3.16) and under the assumption of the claim, we obtain

$$\mathbb{E}\|S_r(E)\|_2^2 + 2\mathbb{E}|\langle L_r(E), S_r(E)\rangle| \lesssim \left(\frac{\mathbf{r}(\Sigma)}{n}\right)^{3/2}.$$

The result follows from the last two displays. □

4. Concentration inequalities. The main goal of this section is to derive a concentration bound for the squared Hilbert–Schmidt error $\|\hat{P}_r - P_r\|_2^2$ around its expectation. Denote

$$B_r(\Sigma) := 2\sqrt{2}\|P_r \Sigma P_r\|_2\|C_r \Sigma C_r\|_2. \tag{4.1}$$

THEOREM 4. *Suppose that, for some* $\gamma \in (0,1)$,

$$\mathbb{E}\|\hat{\Sigma} - \Sigma\|_\infty \le \frac{(1-\gamma)\bar{g}_r}{2}. \tag{4.2}$$

Moreover, let $t \ge 1$ *and suppose that*

$$m_r \lesssim 1, \qquad \frac{\|\Sigma\|_\infty}{\bar{g}_r}\sqrt{\frac{t}{n}} \lesssim 1. \tag{4.3}$$

Then, for some constant $D_\gamma > 0$ *with probability at least* $1 - e^{-t}$,

$$\begin{aligned} &\big|\|\hat{P}_r - P_r\|_2^2 - \mathbb{E}\|\hat{P}_r - P_r\|_2^2\big| \\ &\quad \le D_\gamma\left[\frac{B_r(\Sigma)}{n}\sqrt{t} \vee \frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\frac{t}{n} \vee \frac{\|\Sigma\|_\infty^3}{\bar{g}_r^3}\frac{\mathbf{r}(\Sigma)}{n}\sqrt{\frac{t}{n}}\right]. \end{aligned} \tag{4.4}$$

Note that the first term $\frac{B_r(\Sigma)}{n}\sqrt{t}$ in the right-hand side of (4.4) is dominant if $B_r(\Sigma) \to \infty$ and $\frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)\sqrt{n}} \to 0$ [provided also that $\frac{\|\Sigma\|_\infty}{\bar{g}_r} \lesssim 1$ and $B_r(\Sigma) \gg \sqrt{t}$]. In the next section, it will be shown that under the same assumptions the random variable $\frac{\|\hat{P}_r - P_r\|_2^2 - \mathbb{E}\|\hat{P}_r - P_r\|_2^2}{\operatorname{Var}^{1/2}(\|\hat{P}_r - P_r\|_2^2)}$ is close in distribution to the standard normal and, in addition, $\operatorname{Var}^{1/2}(\|\hat{P}_r - P_r\|_2^2) = (1 + o(1))\frac{B_r(\Sigma)}{n}$. Note also that the assumption $\frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)\sqrt{n}} \to 0$ imply that $\mathbf{r}(\Sigma) = o(n)$ provided that $\|\Sigma\|_\infty \lesssim \bar{g}_r$ and $m_r \lesssim 1$. Indeed,

$$\begin{aligned} B_r^2(\Sigma) &= 8\|P_r\Sigma P_r\|_2^2\|C_r\Sigma C_r\|_2^2 = 8\sum_{s\ne r}\frac{\mu_r^2\mu_s^2 m_r m_s}{(\mu_r - \mu_s)^4} \le 8\frac{m_r\|\Sigma\|_\infty^3}{\bar{g}_r^4}\sum_{s\ne r}\mu_s m_s \\ &\le 8\frac{m_r\|\Sigma\|_\infty^3}{\bar{g}_r^4}\operatorname{tr}(\Sigma) = 8\frac{m_r\|\Sigma\|_\infty^4}{\bar{g}_r^4}\mathbf{r}(\Sigma) \lesssim \mathbf{r}(\Sigma). \end{aligned} \tag{4.5}$$

Therefore,

$$\frac{\mathbf{r}(\Sigma)}{n} \lesssim \left(\frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)\sqrt{n}}\right)^2 \to 0. \tag{4.6}$$

The main ingredient in the proofs of these results is a concentration bound for the random variables $\|\hat{P}_r - P_r\|_2^2 - \|L_r(E)\|_2^2$ given below (recall that $E = \hat{\Sigma} - \Sigma$).

THEOREM 5. *Suppose that, for some* $\gamma \in (0,1)$, *condition* (4.2) *holds. Then there exists a constant* $L_\gamma > 0$ *such that for all* $t \ge 1$ *the following bound holds with probability at least* $1 - e^{-t}$:

$$\begin{aligned} &\big|\|\hat{P}_r - P_r\|_2^2 - \|L_r(E)\|_2^2 - \mathbb{E}(\|\hat{P}_r - P_r\|_2^2 - \|L_r(E)\|_2^2)\big| \\ &\quad \le L_\gamma m_r\frac{\|\Sigma\|_\infty^3}{\bar{g}_r^3}\left(\frac{\mathbf{r}(\Sigma)}{n} \vee \frac{t}{n} \vee \left(\frac{t}{n}\right)^2\right)\sqrt{\frac{t}{n}}. \end{aligned} \tag{4.7}$$

PROOF. It easily follows from Theorem 1 that under assumption (4.2)

$$\|\Sigma\|_\infty\biggl(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}}\vee\frac{\mathbf{r}(\Sigma)}{n}\biggr)\lesssim\frac{(1-\gamma)\bar{g}_r}{2}\le\|\Sigma\|_\infty,$$

which implies that $\mathbf{r}(\Sigma)\lesssim n$. Theorem 2 implies that for some constant $C'>0$ and for all $t\ge 1$ with probability at least $1-e^{-t}$

$$\|\hat{\Sigma}-\Sigma\|_\infty\le\mathbb{E}\|\hat{\Sigma}-\Sigma\|_\infty+C'\|\Sigma\|_\infty\biggl(\sqrt{\frac{t}{n}}\vee\frac{t}{n}\biggr).$$

We will first assume that

$$C\|\Sigma\|_\infty\sqrt{\frac{t}{n}}\le\frac{\gamma\bar{g}_r}{4} \tag{4.8}$$

with a sufficiently large constant $C\ge 1$ (the proof of the concentration bound in the opposite case will be much easier). This assumption easily implies that $t\le n$ and, if $C\ge C'$,

$$C'\|\Sigma\|_\infty\biggl(\sqrt{\frac{t}{n}}\vee\frac{t}{n}\biggr)\le C\|\Sigma\|_\infty\sqrt{\frac{t}{n}}.$$

Denote

$$\delta_n(t):=\mathbb{E}\|\hat{\Sigma}-\Sigma\|_\infty+C\|\Sigma\|_\infty\sqrt{\frac{t}{n}}.$$

Then $\mathbb{P}\{\|\hat{\Sigma}-\Sigma\|_\infty\ge\delta_n(t)\}\le e^{-t}$.

The main part of the proof is the derivation of a concentration inequality for the function

$$g(X_1,\dots,X_n)=\bigl(\|\hat{P}_r-P_r\|_2^2-\|L_r(E)\|_2^2\bigr)\varphi\biggl(\frac{\|E\|_\infty}{\delta}\biggr),$$

where, for some $\gamma\in(0,1)$, φ is a Lipschitz function on $\mathbb{R}_+$ with constant $\frac{1}{\gamma}$, $0\le\varphi(s)\le 1$, $\varphi(s)=1, s\le 1$, $\varphi(s)=0, s>1+\gamma$, and $\delta>0$ is such that $\|E\|_\infty\le\delta$ with a high probability. This inequality will be then used with $\delta=\delta_n(t)$. Together with Theorem 2, it will imply bound (4.7) under assumption (4.8).

Our main tool is the following concentration inequality that easily follows from Gaussian isoperimetric inequality (see, e.g., [17], Theorem 1.2).

LEMMA 3. *Let $X_1,\dots,X_n$ be i.i.d. centered Gaussian random variables in $\mathbb{H}$ with covariance operator Σ. Let $f:\mathbb{H}^n\mapsto\mathbb{R}$ be a function satisfying the following Lipschitz condition with some $L>0$:*

$$|f(x_1,\dots,x_n)-f(x_1',\dots,x_n')|\le L\biggl(\sum_{j=1}^n\|x_j-x_j'\|^2\biggr)^{1/2},$$
$$x_1,\dots,x_n,x_1',\dots,x_n'\in\mathbb{H}.$$

Suppose that, for a real number M,

$$\mathbb{P}\{f(X_1,\dots,X_n)\ge M\}\ge 1/4 \quad \textit{and} \quad \mathbb{P}\{f(X_1,\dots,X_n)\le M\}\ge 1/4.$$

Then there exists a constant $D>0$ *such that for all* $t\ge 1$,

$$\mathbb{P}\{|f(X_1,\dots,X_n)-M|\ge DL\|\Sigma\|_\infty^{1/2}\sqrt{t}\}\le e^{-t}.$$

The derivation of the inequality of Lemma 3 from the isoperimetric inequality is similar to the standard derivation when M is the median.

We have to check now that the function $g(X_1,\dots,X_n)$ satisfies the Lipschitz condition (with a minor abuse of notation we view $X_1,\dots,X_n$ here as nonrandom vectors in $\mathbb{H}$ rather than random variables).

LEMMA 4. *Suppose that, for some* $\gamma\in(0,1/2)$,

$$\delta\le\frac{1-2\gamma}{1+2\gamma}\frac{\bar{g}_r}{2}. \tag{4.9}$$

Then there exists a constant $D_\gamma>0$ *such that, for all* $X_1,\dots,X_n,X_1',\dots,X_n'\in\mathbb{H}$,

$$\begin{aligned}&|g(X_1,\dots,X_n)-g(X_1',\dots,X_n')|\\ &\quad\le D_\gamma m_r\frac{\delta^2}{\bar{g}_r^3}\frac{\|\Sigma\|_\infty^{1/2}+\sqrt{\delta}}{\sqrt{n}}\left(\sum_{j=1}^n\|X_j-X_j'\|^2\right)^{1/2}.\end{aligned} \tag{4.10}$$

PROOF. Observe that

$$\begin{aligned}\|\hat{P}_r-P_r\|_2^2-\|L_r(E)\|_2^2&=\|L_r(E)+S_r(E)\|_2^2-\|L_r(E)\|_2^2\\ &=2\langle L_r(E),S_r(E)\rangle+\|S_r(E)\|_2^2=:\tilde{g}(E).\end{aligned}$$

Also, note that $L_r(E)$ is an operator of rank at most $2m_r$ and $S_r(E)=\hat{P}_r-P_r-L_r(E)$ has rank at most $4m_r$ (under the assumption that $\|E\|_\infty<\bar{g}_r/2$ implying that $\hat{P}_r$ is of rank m_r). This allows us to bound the Hilbert–Schmidt norms of such operators in terms of their operator norms: $\|A\|_2^2\le\operatorname{rank}(A)\|A\|_\infty^2$. Thus, we get

$$|g(X_1,\dots,X_n)|\le 4\sqrt{2}m_r(\|L_r(E)\|_\infty\|S_r(E)\|_\infty+\|S_r(E)\|_\infty^2)\varphi\left(\frac{\|E\|_\infty}{\delta}\right).$$

Since $\varphi(\frac{\|E\|_\infty}{\delta})=0$ if $\|E\|_\infty\ge(1+\gamma)\delta$, claims (2.6), (2.7) of Lemma 1 imply that, under assumption (4.9)

$$|g(X_1,\dots,X_n)|\le c_\gamma m_r\left(\frac{\delta}{\bar{g}_r}\right)^3, \tag{4.11}$$

for some constant $c_\gamma>0$ depending only γ.

We will denote $\hat{\Sigma}' := n^{-1}\sum_{j=1}^n X_j' \otimes X_j'$ and $E' := \hat{\Sigma}' - \Sigma$. Using now (2.6), (2.7), (4.11) and the fact that φ is bounded by 1 and Lipschitz with constant $\frac{1}{\gamma}$, which implies that the function $t \mapsto \varphi(\frac{t}{\delta})$ is Lipschitz with constant $\frac{1}{\gamma\delta}$, we easily get that, under the assumptions

$$\|E\|_\infty \le (1+\gamma)\delta, \qquad \|E'\|_\infty \le (1+\gamma)\delta, \tag{4.12}$$

the following inequality holds:

$$\begin{aligned}
&\left|\tilde{g}(E)\varphi\left(\frac{\|E\|_\infty}{\delta}\right) - \tilde{g}(E')\varphi\left(\frac{\|E'\|_\infty}{\delta}\right)\right| \\
&\quad \le \left|\tilde{g}(E) - \tilde{g}(E')\right| + \frac{c_\gamma}{\gamma}\frac{\delta^2}{\bar{g}_r^3}\|E - E'\|_\infty \\
&\quad \le 2\left|\langle L_r(E - E'), S_r(E)\rangle\right| + 2\left|\langle L_r(E'), S_r(E) - S_r(E')\rangle\right| \\
&\qquad + \left|\langle S_r(E) - S_r(E'), S_r(E) + S_r(E')\rangle\right| + \frac{c_\gamma}{\gamma}\frac{\delta^2}{\bar{g}_r^3}\|E - E'\|_\infty.
\end{aligned} \tag{4.13}$$

Using the Lipschitz bound of Lemma 2 and (2.6), (2.7) of Lemma 1, we easily get that

$$\left|g(X_1,\dots,X_n) - g(X_1',\dots,X_n')\right| \le c_\gamma' m_r \frac{\delta^2}{\bar{g}_r^3}\|E - E'\|_\infty, \tag{4.14}$$

where $c_\gamma' > 0$ depends only on γ.

A similar bound holds in the case when

$$\|E\|_\infty \le (1+\gamma)\delta, \qquad \|E'\|_\infty > (1+\gamma)\delta$$

[when both norms are larger than $(1+\gamma)\delta$, the function φ is equal to zero and the bound is trivial]. Indeed, first consider the case when $\|E - E'\|_\infty \ge \gamma\delta$. Then, in view of (4.11), we have

$$\begin{aligned}
&\left|\tilde{g}(E)\varphi\left(\frac{\|E\|_\infty}{\delta}\right) - \tilde{g}(E')\varphi\left(\frac{\|E'\|_\infty}{\delta}\right)\right| \\
&\quad = \left|\tilde{g}(E)\varphi\left(\frac{\|E\|_\infty}{\delta}\right)\right| \le c_\gamma m_r \frac{\delta^3}{\bar{g}_r^3} \le \frac{c_\gamma}{\gamma} m_r \frac{\delta^2}{\bar{g}_r^3}\|E - E'\|_\infty.
\end{aligned}$$

On the other hand, if $\|E - E'\|_\infty < \gamma\delta$, we have that $\|E'\|_\infty \le (1+2\gamma)\delta$ and, taking into account assumption (4.9), we can repeat the argument in the case (4.12) ending up with the same bound as (4.14) with a positive constant (possibly different from c_γ', but still depending only on γ) in the right-hand side.

The following bound (see Lemma 5 in [13]) provides a control of $\|E - E'\|_\infty$:

$$\begin{aligned}
&\|E - E'\|_\infty \\
&\quad \le \frac{4\|\Sigma\|_\infty^{1/2} + 4\sqrt{2\delta}}{\sqrt{n}}\left(\sum_{j=1}^n \|X_j - X_j'\|^2\right)^{1/2} \vee \frac{4}{n}\sum_{j=1}^n \|X_j - X_j'\|^2.
\end{aligned} \tag{4.15}$$

Now substitute the last bound in the right-hand side of (4.14) and observe that, in view of (4.11), the left-hand side of (4.14) can be also upper bounded by $2c_\gamma m_r \frac{\delta^3}{\bar{g}_r^3}$. Therefore, we get that with some constant $L_\gamma > 0$,

$$
\begin{aligned}
&\bigl|g(X_1, \dots, X_n) - g\bigl(X_1', \dots, X_n'\bigr)\bigr| \\
&\quad \le 4c_\gamma' m_r \frac{\delta^2}{\bar{g}_r^3}\left[\frac{\|\Sigma\|_\infty^{1/2} + \sqrt{2\delta}}{\sqrt{n}}\left(\sum_{j=1}^n \|X_j - X_j'\|^2\right)^{1/2} \vee \frac{1}{n}\sum_{j=1}^n \|X_j - X_j'\|^2\right] \\
&\quad \wedge 2c_\gamma m_r \frac{\delta^3}{\bar{g}_r^3} \qquad (4.16)\\
&\quad \le L_\gamma m_r \frac{\delta^2}{\bar{g}_r^3}\left[\frac{\|\Sigma\|_\infty^{1/2} + \sqrt{2\delta}}{\sqrt{n}}\left(\sum_{j=1}^n \|X_j - X_j'\|^2\right)^{1/2}\right. \\
&\qquad \left.\vee \left(\frac{1}{n}\sum_{j=1}^n \|X_j - X_j'\|^2 \wedge \delta\right)\right].
\end{aligned}
$$

Using an elementary inequality $a \wedge b \le \sqrt{ab}$, $a, b \ge 0$, we get

$$
\frac{1}{n}\sum_{j=1}^n \|X_j - X_j'\|^2 \wedge \delta \le \sqrt{\frac{\delta}{n}}\left(\sum_{j=1}^n \|X_j - X_j'\|^2\right)^{1/2}.
$$

This allows us to drop the last term in the maximum in the right-hand side of (4.16) (since a similar expression is a part of the first term). This yields bound (4.10). □

Getting back to the proof of Theorem 5, it will be convenient to prove first a version of its concentration bound with a median instead of the mean. Denote by $\mathrm{Med}(\eta)$ a median of a random variable η and define $M := \mathrm{Med}(\|\hat{P}_r - P_r\|_2^2 - \|L_r(E)\|_2^2)$. Let $\delta := \delta_n(t)$ and suppose that $t \ge \log(4)$ (by adjusting the constants, one can replace this condition by $t \ge 1$ as it is done in the statement of the theorem). Under conditions (4.2) and (4.8), $\delta_n(t) \le (1 - \frac{\gamma}{2})\frac{\bar{g}_r}{2} = \frac{1-2\gamma'}{1+2\gamma'}\frac{\bar{g}_r}{2}$ for some $\gamma' \in (0, 1/2)$. Thus, the function $g(X_1, \dots, X_n)$ satisfies the Lipschitz condition (4.10) with some constant $D_\gamma' = D_{\gamma'}$. Also, we have $\mathbb{P}\{\|E\|_\infty \ge \delta\} \le e^{-t} \le 1/4$. Note that on the event $\{\|E\|_\infty < \delta\}$, $g(X_1, \dots, X_n) = \|\hat{P}_r - P_r\|_2^2 - \|L_r(E)\|_2^2$. Therefore,

$$
\begin{aligned}
\mathbb{P}\bigl\{g(X_1, \dots, X_n) \ge M\bigr\} &\ge \mathbb{P}\bigl\{g(X_1, \dots, X_n) \ge M, \|E\|_\infty < \delta\bigr\} \\
&\ge \mathbb{P}\bigl\{\|\hat{P}_r - P_r\|_2^2 - \bigl\|L_r(E)\bigr\|_2^2 \ge M\bigr\} - \mathbb{P}\bigl\{\|E\|_\infty \ge \delta\bigr\} \\
&\ge 1/4.
\end{aligned}
$$

Quite similarly, $\mathbb{P}\{g(X_1,\dots,X_n)\le M\}\ge 1/4$. It follows from Lemma 3 that with probability at least $1-e^{-t}$

$$\bigl|g(X_1,\dots,X_n)-M\bigr|\le L'_\gamma m_r\frac{\delta_n(t)^2}{\bar{g}_r^3}\|\Sigma\|_\infty^{1/2}\bigl(\|\Sigma\|_\infty^{1/2}+\sqrt{\delta_n(t)}\bigr)\sqrt{\frac{t}{n}}$$

with some constant $L'_\gamma>0$. Using the bound

$$\delta_n(t)\lesssim\|\Sigma\|_\infty\biggl(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}}\vee\sqrt{\frac{t}{n}}\biggr)$$

that easily follows from the definition of $\delta_n(t)$ and the bound of Theorem 1, we get that with some $L_\gamma>0$ and with the same probability

$$\bigl|g(X_1,\dots,X_n)-M\bigr|\le L_\gamma m_r\frac{\|\Sigma\|_\infty^3}{\bar{g}_r^3}\biggl(\frac{\mathbf{r}(\Sigma)}{n}\vee\frac{t}{n}\biggr)\sqrt{\frac{t}{n}}.$$

Since $\mathbb{P}\{\|E\|_\infty\ge\delta\}\le e^{-t}$ and $g(X_1,\dots,X_n)=\|\hat{P}_r-P_r\|_2^2-\|L_r(E)\|_2^2$ when $\|E\|_\infty<\delta$, we can conclude that with probability at least $1-2e^{-t}$

$$\begin{aligned}\bigl|\|\hat{P}_r-P_r\|_2^2-\|L_r(E)\|_2^2-M\bigr|&\le L_\gamma m_r\frac{\|\Sigma\|_\infty^3}{\bar{g}_r^3}\biggl(\frac{\mathbf{r}(\Sigma)}{n}\vee\frac{t}{n}\biggr)\sqrt{\frac{t}{n}}\\&\le L_\gamma m_r\frac{\|\Sigma\|_\infty^3}{\bar{g}_r^3}\biggl(\frac{\mathbf{r}(\Sigma)}{n}\vee\frac{t}{n}\vee\Bigl(\frac{t}{n}\Bigr)^2\biggr)\sqrt{\frac{t}{n}}.\end{aligned}$$

Adjusting the value of the constant L_γ one can replace the probability bound $1-2e^{-t}$ by $1-e^{-t}$.

We will now prove a similar bound in the case when condition (4.8) does not hold. Then

$$\frac{\|\Sigma\|_\infty}{\bar{g}_r}\sqrt{\frac{t}{n}}\ge\frac{\gamma}{4C}. \tag{4.17}$$

It follows from bound (2.4) and the definition of $L_r(E)$ that, for some constant $c>0$,

$$\bigl|\|\hat{P}_r-P_r\|_2^2-\|L_r(E)\|_2^2\bigr|\le cm_r\frac{\|E\|_\infty^2}{\bar{g}_r^2}.$$

We can now use the bounds of Theorems 1 and 2 to show that under condition (4.2) for some $C>0$ with probability at least $1-e^{-t}$,

$$\bigl|\|\hat{P}_r-P_r\|_2^2-\|L_r(E)\|_2^2\bigr|\le Cm_r\frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\biggl(\frac{\mathbf{r}(\Sigma)}{n}\vee\frac{t}{n}\vee\Bigl(\frac{t}{n}\Bigr)^2\biggr).$$

In view of condition (4.17), we get from the last bound that with some $L'_\gamma > 0$ with probability at least $1 - e^{-t}$,

$$\big|\|\hat{P}_r - P_r\|_2^2 - \|L_r(E)\|_2^2\big| \le L'_\gamma m_r \frac{\|\Sigma\|_\infty^3}{\bar{g}_r^3}\left(\frac{\mathbf{r}(\Sigma)}{n} \vee \frac{t}{n} \vee \left(\frac{t}{n}\right)^2\right)\sqrt{\frac{t}{n}}.$$

This easily implies the following bound on the median M:

$$M \le L'_\gamma m_r \frac{\|\Sigma\|_\infty^3}{\bar{g}_r^3}\left(\frac{\mathbf{r}(\Sigma)}{n} \vee \frac{\log 2}{n} \vee \left(\frac{\log 2}{n}\right)^2\right)\sqrt{\frac{\log 2}{n}}.$$

Therefore, for some $L_\gamma > 0$ and for all $t \ge 1$, with probability at least $1 - e^{-t}$

$$\begin{aligned} &\big|\|\hat{P}_r - P_r\|_2^2 - \|L_r(E)\|_2^2 - M\big| \\ &\quad\le L_\gamma m_r \frac{\|\Sigma\|_\infty^3}{\bar{g}_r^3}\left(\frac{\mathbf{r}(\Sigma)}{n} \vee \frac{t}{n} \vee \left(\frac{t}{n}\right)^2\right)\sqrt{\frac{t}{n}}, \end{aligned} \tag{4.18}$$

and the last bound was proved in both cases (4.8) and (4.17).

It remains to integrate out the tails of exponential bound (4.18) to get the inequality

$$\begin{aligned} \big|\mathbb{E}(\|\hat{P}_r - P_r\|_2^2 - \|L_r(E)\|_2^2) - M\big| &\le \mathbb{E}\big|\|\hat{P}_r - P_r\|_2^2 - \|L_r(E)\|_2^2 - M\big| \\ &\le \bar{L}_\gamma m_r \frac{\|\Sigma\|_\infty^3}{\bar{g}_r^3}\left(\frac{\mathbf{r}(\Sigma)}{n} \vee \frac{1}{n}\right)\sqrt{\frac{1}{n}} \end{aligned} \tag{4.19}$$

with some $\bar{L}_\gamma > 0$. Indeed, denote $\xi := |\|\hat{P}_r - P_r\|_2^2 - \|L_r(E)\|_2^2 - M|$ and

$$\Delta_n(t) := L_\gamma m_r \frac{\|\Sigma\|_\infty^3}{\bar{g}_r^3}\left(\frac{\mathbf{r}(\Sigma)}{n} \vee \frac{t}{n} \vee \left(\frac{t}{n}\right)^2\right)\sqrt{\frac{t}{n}}.$$

We have $\Delta_n(1) = L_\gamma m_r \frac{\|\Sigma\|_\infty^3}{\bar{g}_r^3}(\frac{\mathbf{r}(\Sigma)}{n} \vee \frac{1}{n})\sqrt{\frac{1}{n}}$. It is immediate to check that, for $t \ge 1$, $\Delta_n(t) \le \Delta_n(1)t^{5/2}$. Therefore, it follows from (4.18) that $\mathbb{P}\{\xi \ge \Delta_n(1)t^{5/2}\} \le e^{-t}$, $t \ge 1$ or, equivalently, $\mathbb{P}\{\xi \ge \Delta_n(1)t\} \le e^{-t^{2/5}}$, $t \ge 1$. Thus, we have

$$\mathbb{E}\frac{\xi}{\Delta_n(1)} = \int_0^\infty \mathbb{P}\left\{\frac{\xi}{\Delta_n(1)} \ge t\right\} dt \le 1 + \int_1^\infty e^{-t^{2/5}}\, dt := d,$$

where d is a numerical constant, which implies (4.19). Along with (4.18), this implies concentration inequality (4.7). □

We now turn to the proof of Theorem 4.

PROOF OF THEOREM 4. In view of Theorem 5, it is sufficient to obtain a concentration bound for $\|L_r(E)\|_2^2 - \mathbb{E}\|L_r(E)\|_2^2$. This could be done by rewriting $\|L_r(E)\|_2^2$ in terms of U-statistics and using the corresponding exponential bounds. However, we will follow a different (more elementary) path that directly utilizes the Gaussiness of random variables $\{X_j\}$. The key ingredient is the following simple representation lemma. In what follows, $\xi \overset{d}{=} \eta$ means that random variables ξ and η have the same distribution.

LEMMA 5. *The following representation holds*:

$$n\|L_r(E)\|_2^2 \overset{d}{=} 2 \sum_{k\in\Delta_r} \gamma_k \|C_r X^{(k)}\|^2, \tag{4.20}$$

where γ_k *are the eigenvalues of the random matrix* $\Gamma_r := \frac{1}{n}\sum_{i=1}^n P_r X_i \otimes P_r X_i$ *and* $X^{(k)}$, $k \in \Delta_r$ *are i.i.d. copies of* X *independent of* Γ_r.

PROOF. Note that $n\|L_r(E)\|_2^2 = n\|P_r E C_r + C_r E P_r\|_2^2$. Since the operators $P_r E C_r$ and $C_r E P_r$ are orthogonal with respect to the Hilbert–Schmidt inner product and

$$\|P_r E C_r\|_2^2 = \operatorname{tr}(P_r E C_r C_r E P_r) = \operatorname{tr}(C_r E P_r P_r E C_r) = \|C_r E P_r\|_2^2,$$

we have

$$\|P_r E C_r + C_r E P_r\|_2^2 = \|P_r E C_r\|_2^2 + \|C_r E P_r\|_2^2 = 2\|P_r E C_r\|_2^2.$$

Also, note that $P_r E C_r = \frac{1}{n}\sum_{j=1}^n P_r X_j \otimes C_r X_j$. Therefore,

$$n\|L_r(E)\|_2^2 = 2n\|P_r E C_r\|_2^2 = 2\left\|\frac{1}{\sqrt{n}}\sum_{j=1}^n P_r X_j \otimes C_r X_j\right\|_2^2. \tag{4.21}$$

Define the following mapping:

$$T(u_1 \otimes u_2 \otimes u_3 \otimes u_4) = (u_1 \otimes u_3 \otimes u_2 \otimes u_4), \qquad u_1, u_2, u_3, u_4 \in \mathbb{H}.$$

It can be extended in a unique way by linearity and continuity to a bounded linear operator $T : \mathbb{H} \otimes \mathbb{H} \otimes \mathbb{H} \otimes \mathbb{H} \mapsto \mathbb{H} \otimes \mathbb{H} \otimes \mathbb{H} \otimes \mathbb{H}$.

Recall that $P_r X_j$, $j = 1, \dots, n$ and $C_r X_j$, $j = 1, \dots, n$ are centered Gaussian random variables and they are uncorrelated (see the proof of Theorem 3). Therefore, they are also independent. Conditionally on $P_r X_j$, $j = 1, \dots, n$, the distribution of random operator $U := \frac{1}{\sqrt{n}}\sum_{j=1}^n P_r X_j \otimes C_r X_j$ is centered Gaussian with covariance

$$\begin{aligned}
&\mathbb{E}(U \otimes U | P_r X_j, j = 1, \dots, n) \\
&\quad = n^{-1}\sum_{j=1}^n \mathbb{E}(P_r X_j \otimes C_r X_j \otimes P_r X_j \otimes C_r X_j | P_r X_j, j = 1, \dots, n) \\
&\quad = T\big(\Gamma_r \otimes \mathbb{E}(C_r X \otimes C_r X)\big) = T\big(\Gamma_r \otimes (C_r \Sigma C_r)\big).
\end{aligned}$$

Note that Γ_r can be viewed as a symmetric operator acting in the eigenspace of eigenvalue μ_r, and it is nonnegatively definite. Thus, it has spectral representation $\Gamma_r = \sum_{k\in\Delta_r} \gamma_k \phi_k \otimes \phi_k$, where $\gamma_k \geq 0$ are its eigenvalues and ϕ_k are its orthonormal eigenvectors (that belong to the eigenspace of μ_r). It follows that

$$\mathbb{E}(U \otimes U | P_r X_j, j = 1, \dots, n) = T\Big(\sum_{k\in\Delta_r} \gamma_k\big(\phi_k \otimes \phi_k \otimes \mathbb{E}(C_r X \otimes C_r X)\big)\Big).$$

Let $X^{(k)}, k \in \Delta_r$ be independent copies of X (also independent of $X_1, \dots, X_n$). Denote $V := \sum_{k\in\Delta_r} \sqrt{\gamma_k}\phi_k \otimes C_r X^{(k)}$. It is now easy to check that

$$\mathbb{E}(V \otimes V | P_r X_j, j = 1, \dots, n) = T\Big(\sum_{k\in\Delta_r} \gamma_k\big(\phi_k \otimes \phi_k \otimes \mathbb{E}(C_r X \otimes C_r X)\big)\Big),$$

implying that conditional distributions of U and V given $P_r X_j, j = 1, \dots, n$ are the same. As a consequence, the distribution of $n\|L_r(E)\|_2^2 = 2\|U\|_2^2$ coincides with the distribution of random variable

$$2\|V\|_2^2 = 2\sum_{k\in\Delta_r} \gamma_k \|\phi_k \otimes C_r X^{(k)}\|_2^2 = 2\sum_{k\in\Delta_r} \gamma_k \|C_r X^{(k)}\|^2. \tag{4.22}$$

□

Note that

$$\|C_r X^{(k)}\|^2 = \sum_{s\neq r} \sum_{j\in\Delta_s} \frac{\mu_s}{(\mu_s - \mu_r)^2} \eta_{k,j}^2,$$

where $\eta_{k,j} := \mu_s^{-1/2}\langle X^{(k)}, \theta_j\rangle, j \in \Delta_s, k \in \Delta_r, s \neq r$ are i.i.d. standard normal random variables, $\{\theta_j : j \in \Delta_s\}$ being an orthonormal basis of the eigenspace corresponding to $\mu_s, s \geq 1$. In view of representation (4.22), we get

$$n\|L_r(E)\|_2^2 \overset{d}{=} 2\sum_{k\in\Delta_r} \sum_{s\neq r} \sum_{j\in\Delta_s} \frac{\gamma_k \mu_s}{(\mu_s - \mu_r)^2} \eta_{k,j}^2$$

and, since $\gamma_k, k \in \Delta_r$ and $\eta_{k,j}, j \in \Delta_s, k \in \Delta_r$ are independent,

$$\begin{aligned}
n\mathbb{E}\|L_r(E)\|_2^2 &= 2\sum_{k\in\Delta_r} \sum_{s\neq r} \sum_{j\in\Delta_s} \frac{\mathbb{E}\gamma_k \mu_s}{(\mu_s - \mu_r)^2} = 2\sum_{s\neq r} \frac{\mathbb{E}\operatorname{tr}(\Gamma_r) m_s \mu_s}{(\mu_s - \mu_r)^2} \\
&= 2\sum_{s\neq r} \frac{\operatorname{tr}(P_r \Sigma P_r) m_s \mu_s}{(\mu_s - \mu_r)^2} = 2\sum_{s\neq r} \frac{m_r \mu_r m_s \mu_s}{(\mu_s - \mu_r)^2} \\
&= 2\operatorname{tr}(P_r \Sigma P_r)\operatorname{tr}(C_r \Sigma C_r) = A_r(\Sigma).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\|L_r(E)\|_2^2 - \mathbb{E}\|L_r(E)\|_2^2 &\overset{d}{=} \frac{2}{n} \sum_{k\in\Delta_r} \sum_{s\neq r} \sum_{j\in\Delta_s} \frac{\mu_r \mu_s}{(\mu_s - \mu_r)^2} \frac{\gamma_k}{\mu_r} (\eta_{k,j}^2 - 1) \\
&\quad + \frac{2}{n} \sum_{k\in\Delta_r} \sum_{s\neq r} \frac{\mu_r m_s \mu_s}{(\mu_s - \mu_r)^2} \Big(\frac{\gamma_k}{\mu_r} - 1\Big).
\end{aligned} \tag{4.23}$$

In order to control the right-hand side in the above display, the following elementary lemma will be used (see, e.g., [20], Proposition 5.16).

LEMMA 6. *Let $\{\xi_k\}$ be i.i.d. standard normal random variables. There exists a numerical constant $c > 0$ such that for all $t > 0$*

$$\mathbb{P}\biggl(\biggl|\sum_k \lambda_k(\xi_k^2 - 1)\biggr| \ge t\biggr) \le 2\biggl(\exp\biggl\{-\frac{ct^2}{\sum_k \lambda_k^2}\biggr\} \vee \exp\biggl\{-\frac{ct}{\sup_k |\lambda_k|}\biggr\}\biggr).$$

Applying the bound of the lemma to the first term in the right-hand side of relationship (4.23) *conditionally on* $\gamma_k, k \in \Delta_r$, we get that with probability at least $1 - e^{-t}$

$$\biggl|\frac{2}{n}\sum_{k\in\Delta_r}\sum_{s\neq r}\sum_{j\in\Delta_s}\frac{\mu_r\mu_s}{(\mu_s-\mu_r)^2}\frac{\gamma_k}{\mu_r}(\eta_{k,j}^2 - 1)\biggr|$$
$$\lesssim \biggl(\sum_{s\neq r}\frac{\mu_r^2 m_s \mu_s^2}{(\mu_s-\mu_r)^4}\frac{\sum_{k\in\Delta_r}\gamma_k^2}{\mu_r^2}\biggr)^{1/2}\frac{\sqrt{t}}{n} \vee \sup_{s\neq r}\frac{\mu_r\mu_s}{(\mu_s-\mu_r)^2}\frac{\sup_{k\in\Delta_r}\gamma_k}{\mu_r}\frac{t}{n}.$$

Since $\sup_{k\in\Delta_r}\gamma_k = \|\Gamma_r\|_\infty$, $\sum_{k\in\Delta_r}\gamma_k^2 \le m_r\|\Gamma_r\|_\infty^2$ and

$$B_r^2(\Sigma) = 8\sum_{s\neq r}\frac{m_r\mu_r^2 m_s\mu_s^2}{(\mu_s-\mu_r)^4},$$

the last bound can be rewritten as

$$\biggl|\frac{2}{n}\sum_{k\in\Delta_r}\sum_{s\neq r}\sum_{j\in\Delta_s}\frac{\mu_r\mu_s}{(\mu_s-\mu_r)^2}\frac{\gamma_k}{\mu_r}(\eta_{k,j}^2 - 1)\biggr|$$
$$\lesssim B_r(\Sigma)\frac{\|\Gamma_r\|_\infty}{\mu_r}\frac{\sqrt{t}}{n} \vee \frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\frac{\|\Gamma_r\|_\infty}{\mu_r}\frac{t}{n}. \tag{4.24}$$

As to the second term in the right-hand side of (4.23), the following bound is straightforward:

$$\biggl|\frac{2}{n}\sum_{k\in\Delta_r}\sum_{s\neq r}\frac{\mu_r m_s\mu_s}{(\mu_s-\mu_r)^2}\biggl(\frac{\gamma_k}{\mu_r} - 1\biggr)\biggr|$$
$$\le \frac{2}{n}\sum_{s\neq r}\frac{m_r\mu_r m_s\mu_s}{(\mu_s-\mu_r)^2}\frac{\|\Gamma_r - P_r\Sigma P_r\|_\infty}{\mu_r} = \frac{A_r(\Sigma)}{n}\frac{\|\Gamma_r - P_r\Sigma P_r\|_\infty}{\mu_r}. \tag{4.25}$$

Theorems 1 and 2 easily imply that for all $t \ge 1$ with probability at least $1 - e^{-t}$

$$\|\Gamma_r - \mu_r P_r\|_\infty = \biggl\|n^{-1}\sum_{j=1}^n P_r X_j \otimes P_r X_j - \mathbb{E}(P_r X \otimes P_r X)\biggr\|_\infty$$
$$\lesssim \mu_r\biggl(\sqrt{\frac{m_r}{n}} \vee \frac{m_r}{n} \vee \sqrt{\frac{t}{n}} \vee \frac{t}{n}\biggr).$$

Under additional assumptions $m_r \lesssim n$, $t \lesssim n$, this bound could be simplified as

$$\|\Gamma_r - \mu_r P_r\|_\infty \lesssim \mu_r\biggl(\sqrt{\frac{m_r}{n}} \vee \sqrt{\frac{t}{n}}\biggr) \tag{4.26}$$

and it implies that $\frac{\|\Gamma_r\|_\infty}{\mu_r} \lesssim 1$.

Thus, representation (4.23) and bounds (4.24), (4.25) imply that with probability at least $1 - e^{-t}$

$$\begin{aligned} &\bigl|\|L_r(E)\|_2^2 - \mathbb{E}\|L_r(E)\|_2^2\bigr| \\ &\quad\lesssim B_r(\Sigma)\frac{\sqrt{t}}{n} \vee \frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\frac{t}{n} \vee \frac{A_r(\Sigma)}{n}\biggl(\sqrt{\frac{m_r}{n}} \vee \sqrt{\frac{t}{n}}\biggr). \end{aligned} \tag{4.27}$$

To complete the proof, it is enough to combine bound (4.27) with concentration inequality of Theorem 5, to use bound (3.2) to control $A_r(\Sigma)$ and to take into account conditions (4.3) to simplify the resulting bound. □

5. Normal approximation of squared Hilbert–Schmidt norm errors of empirical spectral projectors. The main result of this section is the following theorem.

THEOREM 6. *Suppose that, for some constants $c_1, c_2 > 0$, $m_r \le c_1$ and $\|\Sigma\|_\infty \le c_2\bar{g}_r$. Suppose also condition (4.2) holds with some $\gamma \in (0,1)$. Then the following bounds hold with some constant $C > 0$ depending only on γ, c_1, c_2:*

$$\begin{aligned} &\sup_{x\in\mathbb{R}}\biggl|\mathbb{P}\biggl\{\frac{n}{B_r(\Sigma)}\bigl(\|\hat{P}_r - P_r\|_2^2 - \mathbb{E}\|\hat{P}_r - P_r\|_2^2\bigr) \le x\biggr\} - \Phi(x)\biggr| \\ &\quad\le C\biggl[\frac{1}{B_r(\Sigma)} + \frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)\sqrt{n}}\sqrt{\log\biggl(\frac{B_r(\Sigma)\sqrt{n}}{\mathbf{r}(\Sigma)} \vee 2\biggr)}\biggr] \end{aligned} \tag{5.1}$$

and

$$\begin{aligned} &\sup_{x\in\mathbb{R}}\biggl|\mathbb{P}\biggl\{\frac{\|\hat{P}_r - P_r\|_2^2 - \mathbb{E}\|\hat{P}_r - P_r\|_2^2}{\mathrm{Var}^{1/2}(\|\hat{P}_r - P_r\|_2^2)} \le x\biggr\} - \Phi(x)\biggr| \\ &\quad\le C\biggl[\frac{1}{B_r(\Sigma)} + \frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)\sqrt{n}}\sqrt{\log\biggl(\frac{B_r(\Sigma)\sqrt{n}}{\mathbf{r}(\Sigma)} \vee 2\biggr)}\biggr], \end{aligned} \tag{5.2}$$

where $\Phi(x)$ denotes the distribution function of standard normal random variable.

This result essentially means that as soon as $B_r(\Sigma) \to \infty$ and $\frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)\sqrt{n}} \to 0$ as $n \to \infty$ (for $\Sigma = \Sigma^{(n)}$), the sequence of random variables $\frac{\|\hat{P}_r - P_r\|_2^2 - \mathbb{E}\|\hat{P}_r - P_r\|_2^2}{\mathrm{Var}^{1/2}(\|\hat{P}_r - P_r\|_2^2)}$ is asymptotically standard normal. As it could be seen from the proof given below, the first term in the normal approximation bound, $\frac{1}{B_r(\Sigma)}$, is related to a Berry–Esseen-type bound on normal approximation of $\|L_r(E)\|_2^2 - \mathbb{E}\|L_r(E)\|_2^2$, which is the quadratic function of the Gaussian data (based on the first-order term in the perturbation series for $\hat{P}_r$). The second term $\frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)\sqrt{n}}\sqrt{\log(\frac{B_r(\Sigma)\sqrt{n}}{\mathbf{r}(\Sigma)} \vee 2)}$ is related to approximation of $\|\hat{P}_r - P_r\|_2^2 - \mathbb{E}\|\hat{P}_r - P_r\|_2^2$ by the quadratic function $\|L_r(E)\|_2^2 - \mathbb{E}\|L_r(E)\|_2^2$ (see Theorem 5).

We will first establish the following fact that would allow us to replace $\frac{B_r(\Sigma)}{n}$ in bound (5.1) by a normalizing factor $\mathrm{Var}^{1/2}(\|\hat{P}_r - P_r\|_2^2)$ in bound (5.2).

THEOREM 7. *Suppose condition* (4.2) *holds for some* $\gamma \in (0,1)$. *Then the following bound holds with some constant* $C_\gamma > 0$:

$$\left|\frac{n}{B_r(\Sigma)}\mathrm{Var}^{1/2}\big(\|\hat{P}_r - P_r\|_2^2\big) - 1\right| \le C_\gamma m_r \frac{\|\Sigma\|_\infty^3}{\bar{g}_r^3}\frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)\sqrt{n}} + \frac{m_r+1}{n}. \tag{5.3}$$

Bound (5.3) shows that, under the assumptions $m_r \frac{\|\Sigma\|_\infty^3}{\bar{g}_r^3}\frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)\sqrt{n}} = o(1)$ and $m_r = o(n)$, we have

$$\mathrm{Var}^{1/2}\big(\|\hat{P}_r - P_r\|_2^2\big) = \big(1+o(1)\big)\frac{B_r(\Sigma)}{n}.$$

REMARK 2. Note that in the case of spiked covariance model (1.1), for $r = 1,\dots,m$,

$$\begin{aligned} B_r(\Sigma) := 2\sqrt{2}\Bigg(&\sum_{1\le j\le m, j\ne r}\frac{(s_r^2+\sigma^2)^2(s_j^2+\sigma^2)^2}{(s_r^2-s_j^2)^4} \\ &+ \frac{(s_r^2+\sigma^2)^2\sigma^4(p-m)}{s_r^8}\Bigg)^{1/2}, \end{aligned} \tag{5.4}$$

which, under the assumption that the parameters $m, s_1^2, \dots, s_m^2, \sigma^2$ are fixed, but $p = p_n \to \infty$ as $n \to \infty$ yields that

$$B_r(\Sigma) = \big(1+o(1)\big)\frac{2\sqrt{2}(s_r^2+\sigma^2)\sigma^2\sqrt{p}}{s_r^4} \qquad \text{as } p \to \infty. \tag{5.5}$$

Note also that $\mathbf{r}(\Sigma) \sim \frac{\sigma^2}{s_1^2+\sigma^2}p$. Thus, the condition $p = o(n)$ implies $\frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)\sqrt{n}} \to 0$ as $n \to \infty$. Therefore, Theorem 7 yields that

$$\mathrm{Var}^{1/2}\big(\|\hat{P}_r - P_r\|_2^2\big) = \big(1+o(1)\big)\frac{2\sqrt{2}(s_r^2+\sigma^2)\sigma^2}{s_r^4}\frac{\sqrt{p}}{n}.$$

Moreover, the bounds on the accuracy of normal approximation of Theorem 6 are of the order

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left\{\frac{\|\hat{P}_r-P_r\|_2^2-\mathbb{E}\|\hat{P}_r-P_r\|_2^2}{\mathrm{Var}^{1/2}(\|\hat{P}_r-P_r\|_2^2)}\le x\right\}-\Phi(x)\right|$$
$$\le C\left[\frac{1}{\sqrt{p}}+\sqrt{\frac{p}{n}\log\left(\frac{n}{p}\vee 2\right)}\right],$$

so, the asymptotic normality of $\|\hat{P}_r-P_r\|_2^2$ holds if $p=p_n\to\infty$ and $p=o(n)$ as $n\to\infty$. If $p\le\sqrt{\frac{n}{\log n}}$, the first term in the normal approximation bound becomes dominant, so, the error rate is $O(\frac{1}{\sqrt{p}})$. By the optimality of the normal approximation rate in the classical Berry–Esseen theorem, it could be seen from our proof below that the rate $O(\frac{1}{\sqrt{p}})$ is optimal in this range of the values of p.

We now provide the proof of Theorem 5.3.

PROOF OF THEOREM 5.3. In view of relationships $n\|L_r(E)\|_2^2\overset{d}{=}2\|V\|_2^2$ and (4.22) (see the proof of Lemma 5), we have

$$\mathrm{Var}(\|L_r(E)\|_2^2)=\frac{4}{n^2}\mathrm{Var}(\|V\|_2^2)=\frac{4}{n^2}\mathrm{Var}\left(\sum_{k\in\Delta_r}\gamma_k\|C_rX^{(k)}\|^2\right)$$
$$=\frac{4}{n^2}\mathbb{E}\left[\mathrm{Var}\left(\sum_{k\in\Delta_r}\gamma_k\|C_rX^{(k)}\|^2\Big|P_rX_1,\dots,P_rX_n\right)\right]\tag{5.6}$$
$$+\frac{4}{n^2}\mathrm{Var}\left(\mathbb{E}\left[\sum_{k\in\Delta_r}\gamma_k\|C_rX^{(k)}\|^2\Big|P_rX_1,\dots,P_rX_n\right]\right).$$

Recall that γ_k, $k\in\Delta_r$ depend only $P_rX_1,\dots,P_rX_n$ and that $X^{(k)}$, $k\in\Delta_r$ are independent of $X_1,\dots,X_n$. Thus, we get

$$\mathbb{E}\left[\mathrm{Var}\left(\sum_{k\in\Delta_r}\gamma_k\|C_rX^{(k)}\|^2\Big|P_rX_1,\dots,P_rX_n\right)\right]$$
$$=\mathbb{E}\left[\sum_{k\in\Delta_r}\gamma_k^2\,\mathrm{Var}(\|C_rX^{(k)}\|^2)\right]$$
$$=\sum_{k\in\Delta_r}\mathbb{E}[\gamma_k^2]\,\mathrm{Var}(\|C_rX^{(k)}\|^2)\tag{5.7}$$
$$=\mathbb{E}[\|\Gamma_r\|_2^2]\,\mathrm{Var}(\|C_rX\|^2).$$

By an easy computation,

$$\mathbb{E}\|\Gamma_r\|_2^2 = \mathbb{E}\left\|n^{-1}\sum_{j=1}^{n} P_r X_j \otimes P_r X_j\right\|_2^2 = m_r \mu_r^2\left(1+\frac{m_r+1}{n}\right)$$

and, for i.i.d. standard normal random variables $\{\eta_j\}$

$$\operatorname{Var}(\|C_r X\|^2) = \operatorname{Var}\left(\sum_{s\neq r}\sum_{j\in\Delta_s}\frac{\mu_s}{(\mu_s-\mu_r)^2}\eta_j^2\right) = \frac{1}{4m_r\mu_r^2}B_r^2(\Sigma).$$

Therefore,

$$\begin{aligned}&\mathbb{E}\left[\operatorname{Var}\left(\sum_{k\in\Delta_r}\gamma_k\|C_r X^{(k)}\|^2 \,\Big|\, P_r X_1,\dots,P_r X_n\right)\right]\\ &\quad = \frac{B_r^2(\Sigma)}{4}\left(1+\frac{m_r+1}{n}\right).\end{aligned} \tag{5.8}$$

Similarly, we have

$$\begin{aligned}&\operatorname{Var}\left(\mathbb{E}\left[\sum_{k\in\Delta_r}\gamma_k\|C_r X^{(k)}\|^2 \,\Big|\, P_r X_1,\dots,P_r X_n\right]\right)\\ &\quad = \operatorname{Var}\left(\sum_{k\in\Delta_r}\gamma_k\mathbb{E}[\|C_r X^{(k)}\|^2]\right)\\ &\quad = \operatorname{Var}(\operatorname{tr}(\Gamma_r))(\mathbb{E}[\|C_r X\|^2])^2\end{aligned} \tag{5.9}$$

and

$$\operatorname{Var}(\operatorname{tr}(\Gamma_r)) = \frac{2m_r\mu_r^2}{n}, \qquad (\mathbb{E}[\|C_r X\|^2])^2 = \frac{1}{4m_r^2\mu_r^2}A_r^2(\Sigma),$$

implying that

$$\operatorname{Var}\left(\mathbb{E}\left[\sum_{k\in\Delta_r}\gamma_k\|C_r X^{(k)}\|^2 \,\Big|\, P_r X_1,\dots,P_r X_n\right]\right) = \frac{A_r^2(\Sigma)}{2m_r n}. \tag{5.10}$$

It follows from (5.6), (5.8) and (5.10) that

$$\operatorname{Var}(\|L_r(E)\|_2^2) = \frac{B_r^2(\Sigma)}{n^2}\left(1+\frac{m_r+1}{n}\right) + \frac{2A_r^2(\Sigma)}{m_r n^3}. \tag{5.11}$$

Denote now

$$\begin{aligned}\xi &:= \|\hat{P}_r - P_r\|_2^2 - \mathbb{E}\|\hat{P}_r - P_r\|_2^2,\\ \eta &:= \|L_r(E)\|_2^2 - \mathbb{E}\|L_r(E)\|_2^2,\end{aligned} \tag{5.12}$$

and $\sigma_\xi^2 = \mathbb{E}\xi^2$, $\sigma_\eta^2 = \mathbb{E}\eta^2$. Combining concentration bound of Theorem 5 with the identity $\mathbb{E}|\xi - \eta|^2 = \int_0^\infty \mathbb{P}\{|\xi - \eta|^2 > t\}\,dt$, we obtain that

$$|\sigma_\xi - \sigma_\eta| \le \sqrt{\mathbb{E}|\xi - \eta|^2} \le C_\gamma m_r \frac{\|\Sigma\|_\infty^3}{\bar{g}_r^3} \frac{\mathbf{r}(\Sigma)}{n} \frac{1}{\sqrt{n}}, \tag{5.13}$$

for some $C_\gamma > 0$ depending only on γ.

To complete the proof, observe that identity (5.11) implies that

$$\begin{aligned}\left|\frac{n}{B_r(\Sigma)} \operatorname{Var}^{1/2}(\|L_r(E)\|_2^2) - 1\right| &\le \sqrt{1 + \frac{m_r + 1}{n} + \frac{2A_r^2(\Sigma)}{m_r B_r^2(\Sigma) n}} - 1 \\ &\le \sqrt{1 + \frac{m_r + 1}{n}} - 1 + \frac{\sqrt{2} A_r(\Sigma)}{\sqrt{m_r} B_r(\Sigma)\sqrt{n}} \\ &\le \frac{m_r + 1}{n} + \frac{\sqrt{2} A_r(\Sigma)}{\sqrt{m_r} B_r(\Sigma)\sqrt{n}},\end{aligned}$$

then bound $A_r(\Sigma)$ using (3.2) and combine the resulting bound with (5.13). □

We now return to the proof of Theorem 6.

PROOF OF THEOREM 6. Under notation (5.12), we will upper bound $\sup_{x\in\mathbb{R}} |\mathbb{P}\{\frac{n}{B_r(\Sigma)}\xi \le x\} - \Phi(x)|$. Theorem 7 will allow us to rewrite the normalizing factor in terms of the variance. First recall that by Theorem 5, with probability at least $1 - e^{-t}$,

$$|\xi - \eta| \le L_\gamma m_r \frac{\|\Sigma\|_\infty^3}{\bar{g}_r^3} \left(\frac{\mathbf{r}(\Sigma)}{n} \vee \frac{t}{n} \vee \left(\frac{t}{n}\right)^2\right) \sqrt{\frac{t}{n}}. \tag{5.14}$$

Also, by (4.23),

$$\begin{aligned}\eta &\stackrel{d}{=} \frac{2}{n} \sum_{k\in\Delta_r} \sum_{s\ne r} \sum_{j\in\Delta_s} \frac{\mu_r \mu_s}{(\mu_s - \mu_r)^2} (\eta_{k,j}^2 - 1) \\ &\quad + \frac{2}{n} \sum_{k\in\Delta_r} \sum_{s\ne r} \sum_{j\in\Delta_s} \frac{\mu_r \mu_s}{(\mu_s - \mu_r)^2} \left(\frac{\gamma_k}{\mu_r} - 1\right) (\eta_{k,j}^2 - 1) \\ &\quad + \frac{2}{n} \sum_{k\in\Delta_r} \sum_{s\ne r} \frac{\mu_r m_s \mu_s}{(\mu_s - \mu_r)^2} \left(\frac{\gamma_k}{\mu_r} - 1\right) \\ &=: \zeta_1 + \zeta_2 + \zeta_3.\end{aligned} \tag{5.15}$$

Similarly to bound (4.24), we get that with probability at least $1 - e^{-t}$

$$|\zeta_2| \lesssim B_r(\Sigma) \frac{\|\Gamma_r - \mu_r P_r\|_\infty}{\mu_r} \frac{\sqrt{t}}{n} \vee \frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2} \frac{\|\Gamma_r - \mu_r P_r\|_\infty}{\mu_r} \frac{t}{n}. \tag{5.16}$$

Assume that $1 \le t \lesssim n$ and $m_r \lesssim n$. It follows from (5.16), (4.25), (4.26) and also from bound (3.2) on $A_r(\Sigma)$ that

$$
\begin{aligned}
&\left|\frac{n}{B_r(\Sigma)}(\zeta_2+\zeta_3)\right| \\
&\quad\lesssim \left(\sqrt{\frac{m_r}{n}} \vee \sqrt{\frac{t}{n}}\right)\sqrt{t} \vee \frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\left(\sqrt{\frac{m_r}{n}} \vee \sqrt{\frac{t}{n}}\right)\frac{t}{B_r(\Sigma)} \\
&\quad\quad \vee \frac{A_r(\Sigma)}{B_r(\Sigma)}\left(\sqrt{\frac{m_r}{n}} \vee \sqrt{\frac{t}{n}}\right) \\
&\quad\lesssim \left(\sqrt{\frac{m_r}{n}} \vee \sqrt{\frac{t}{n}}\right)\sqrt{t} \vee \frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\left(\sqrt{\frac{m_r}{n}} \vee \sqrt{\frac{t}{n}}\right)\frac{t}{B_r(\Sigma)} \\
&\quad\quad \vee m_r\frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)}\left(\sqrt{\frac{m_r}{n}} \vee \sqrt{\frac{t}{n}}\right).
\end{aligned}
\tag{5.17}
$$

Under the assumptions of the theorem $m_r \lesssim 1$, $\|\Sigma\|_\infty \lesssim \bar{g}_r$, it is easy to get from (5.14), (5.15) and (5.17) that

$$
\frac{n}{B_r(\Sigma)}\xi \overset{d}{=} \tau + \zeta,
\tag{5.18}
$$

where

$$
\tau := \frac{2}{B_r(\Sigma)}\sum_{k\in\Delta_r}\sum_{s\neq r}\sum_{j\in\Delta_s}\frac{\mu_r\mu_s}{(\mu_s-\mu_r)^2}\left(\eta_{k,j}^2-1\right)
\tag{5.19}
$$

and the remainder ζ satisfies the following bound with probability at least $1-e^{-t}$:

$$
|\zeta| \lesssim \frac{t}{\sqrt{n}} \vee \frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)\sqrt{n}}\sqrt{t} \vee \frac{t^{3/2}}{B_r(\Sigma)\sqrt{n}}.
\tag{5.20}
$$

We now use the Berry–Esseen theorem and a simple limiting argument that allows one to apply it to a (possibly) infinite sum of independent random variables (5.19) to get the following bound:

$$
\begin{aligned}
\sup_{x\in\mathbb{R}}\left|\mathbb{P}\{\tau \le x\} - \Phi(x)\right| &\lesssim \frac{\sum_{s\neq r}\frac{m_r\mu_r^3 m_s\mu_s^3}{(\mu_s-\mu_r)^6}}{\left(\sum_{s\neq r}\frac{m_r\mu_r^2 m_s\mu_s^2}{(\mu_s-\mu_r)^4}\right)^{3/2}} \\
&\lesssim \frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\frac{1}{B_r(\Sigma)},
\end{aligned}
\tag{5.21}
$$

where we also used the fact that $B_r^2(\Sigma) = 8\sum_{s\neq r}\frac{m_r\mu_r^2 m_s\mu_s^2}{(\mu_s-\mu_r)^4}$.

It follows from (5.18), (5.20) and (5.21) that with some constants $c', c'' > 0$, for all $x \in \mathbb{R}$,

$$
\begin{aligned}
&\mathbb{P}\left\{\frac{n}{B_r(\Sigma)}\xi \le x\right\} \\
&\quad \le \mathbb{P}\left\{\tau \le x + c'\left(\frac{t}{\sqrt{n}} \vee \frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)\sqrt{n}}\sqrt{t} \vee \frac{t^{3/2}}{B_r(\Sigma)\sqrt{n}}\right)\right\} + e^{-t} \\
&\quad \le \Phi\left(x + c'\left(\frac{t}{\sqrt{n}} \vee \frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)\sqrt{n}}\sqrt{t} \vee \frac{t^{3/2}}{B_r(\Sigma)\sqrt{n}}\right)\right) + e^{-t} + \frac{c''}{B_r(\Sigma)} \\
&\quad \le \Phi(x) + c'\left(\frac{t}{\sqrt{n}} \vee \frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)\sqrt{n}}\sqrt{t} \vee \frac{t^{3/2}}{B_r(\Sigma)\sqrt{n}}\right) + e^{-t} + \frac{c''}{B_r(\Sigma)},
\end{aligned}
\tag{5.22}
$$

where we used the fact that Φ is a Lipschitz function with constant less than one. Quite similarly,

$$
\begin{aligned}
&\mathbb{P}\left\{\frac{n}{B_r(\Sigma)}\xi \le x\right\} \\
&\quad \ge \mathbb{P}\left\{\tau \le x - c'\left(\frac{t}{\sqrt{n}} \vee \frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)\sqrt{n}}\sqrt{t} \vee \frac{t^{3/2}}{B_r(\Sigma)\sqrt{n}}\right)\right\} - e^{-t} \\
&\quad \ge \Phi\left(x - c'\left(\frac{t}{\sqrt{n}} \vee \frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)\sqrt{n}}\sqrt{t} \vee \frac{t^{3/2}}{B_r(\Sigma)\sqrt{n}}\right)\right) - e^{-t} - \frac{c''}{B_r(\Sigma)} \\
&\quad \ge \Phi(x) - c'\left(\frac{t}{\sqrt{n}} \vee \frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)\sqrt{n}}\sqrt{t} \vee \frac{t^{3/2}}{B_r(\Sigma)\sqrt{n}}\right) - e^{-t} - \frac{c''}{B_r(\Sigma)}.
\end{aligned}
\tag{5.23}
$$

It follows from (5.22) and (5.23) that

$$
\begin{aligned}
&\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left\{\frac{n}{B_r(\Sigma)}\xi \le x\right\} - \Phi(x)\right| \\
&\qquad \le c'\left(\frac{t}{\sqrt{n}} \vee \frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)\sqrt{n}}\sqrt{t} \vee \frac{t^{3/2}}{B_r(\Sigma)\sqrt{n}}\right) + \frac{c''}{B_r(\Sigma)} + e^{-t}.
\end{aligned}
\tag{5.24}
$$

The last bound will be used with

$$
t = \log B_r(\Sigma) \wedge \log\left(\frac{B_r(\Sigma)\sqrt{n}}{\mathbf{r}(\Sigma)} \vee 2\right) \wedge \log n,
\tag{5.25}
$$

which implies that

$$
e^{-t} \lesssim \frac{t}{\sqrt{n}} \vee \frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)\sqrt{n}}\sqrt{t} \vee \frac{t^{3/2}}{B_r(\Sigma)\sqrt{n}} \vee \frac{1}{B_r(\Sigma)}
\tag{5.26}
$$

and we also have

$$
\frac{t^{3/2}}{B_r(\Sigma)\sqrt{n}} \le \frac{\log n}{\sqrt{n}} \frac{\sqrt{\log B_r(\Sigma)}}{B_r(\Sigma)}.
$$

Without loss of generality, we can assume that $B_r(\Sigma)$ is bounded away from 0 by a numerical constant so that $\frac{\sqrt{\log B_r(\Sigma)}}{B_r(\Sigma)} \le 1$ (otherwise, the bounds of the theorem trivially hold). This implies that $\frac{t^{3/2}}{B_r(\Sigma)\sqrt{n}} \le \frac{\log n}{\sqrt{n}}$ and (5.24) implies

$$
\begin{aligned}
&\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left\{\frac{n}{B_r(\Sigma)}\xi \le x\right\} - \Phi(x)\right| \\
&\quad \le C\left[\frac{1}{B_r(\Sigma)} + \frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)\sqrt{n}}\sqrt{\log\left(\frac{B_r(\Sigma)\sqrt{n}}{\mathbf{r}(\Sigma)} \vee 2\right)} + \frac{\log n}{\sqrt{n}}\right].
\end{aligned} \tag{5.27}
$$

We can now use Theorem 7 to replace the normalization with $\frac{n}{B_r(\Sigma)}$ by the normalization with the standard deviation of ξ. To this end, note that

$$
\frac{\xi}{\sigma_\xi} = \frac{n}{B_r(\Sigma)}\xi + \left(\frac{1}{\sigma_\xi} - \frac{n}{B_r(\Sigma)}\right)\xi. \tag{5.28}
$$

Under the assumptions $m_r \lesssim 1$ and $\|\Sigma\|_\infty \lesssim \bar{g}_r$, we get from Theorem 7 that

$$
\left|\frac{n}{B_r(\Sigma)}\sigma_\xi - 1\right| \lesssim \frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)\sqrt{n}} + \frac{1}{n}.
$$

Without loss of generality, we can and do assume that $\frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)\sqrt{n}} + \frac{1}{n} \le c$ for a small enough constant $c > 0$ so that $|\frac{n}{B_r(\Sigma)}\sigma_\xi - 1| \le 1/2$ (otherwise, the bound of the theorem is trivial). Then

$$
\left|\left(\frac{1}{\sigma_\xi} - \frac{n}{B_r(\Sigma)}\right)\xi\right| \le \left|\frac{n}{B_r(\Sigma)}\sigma_\xi - 1\right|\frac{|\xi|}{\sigma_\xi} \lesssim \left(\frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)\sqrt{n}} + \frac{1}{n}\right)\frac{n}{B_r(\Sigma)}|\xi|.
$$

Combining this with bound of Theorem 4, we get that with probability at least $1 - e^{-t}$

$$
\left|\left(\frac{1}{\sigma_\xi} - \frac{n}{B_r(\Sigma)}\right)\xi\right| \lesssim \left(\frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)\sqrt{n}} + \frac{1}{n}\right)\left(\sqrt{t} \vee \frac{t}{B_r(\Sigma)} \vee \frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)\sqrt{n}}\sqrt{t}\right).
$$

Using the last bound with t defined by (5.25), we easily get that

$$
\left|\left(\frac{1}{\sigma_\xi} - \frac{n}{B_r(\Sigma)}\right)\xi\right| \lesssim \left[\frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)\sqrt{n}}\sqrt{\log\left(\frac{B_r(\Sigma)\sqrt{n}}{\mathbf{r}(\Sigma)} \vee 2\right)} + \frac{\log n}{\sqrt{n}}\right]. \tag{5.29}
$$

By proving bounds on $\mathbb{P}\{\frac{\xi}{\sigma_\xi} \le x\}$ similar to (5.22), (5.23), it follows from (5.27), (5.28) and (5.29) that

$$
\begin{aligned}
&\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left\{\frac{\xi}{\sigma_\xi} \le x\right\} - \Phi(x)\right| \\
&\quad \le C\left[\frac{1}{B_r(\Sigma)} + \frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)\sqrt{n}}\sqrt{\log\left(\frac{B_r(\Sigma)\sqrt{n}}{\mathbf{r}(\Sigma)} \vee 2\right)} + \frac{\log n}{\sqrt{n}}\right].
\end{aligned} \tag{5.30}
$$

To complete the proof, it is enough to show that the term $\frac{\log n}{\sqrt{n}}$ in bounds (5.27) and (5.30) could be dropped. To this end, use bound (4.5) to get

$$\frac{\log n}{\sqrt{n}} \lesssim \frac{1}{\sqrt{B_r(\Sigma)}} \frac{\sqrt{B_r(\Sigma)}}{n^{1/4}} \lesssim \frac{1}{B_r(\Sigma)} + \frac{B_r^2(\Sigma)}{B_r(\Sigma)\sqrt{n}} \lesssim \frac{1}{B_r(\Sigma)} + \frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)\sqrt{n}}$$
$$\lesssim \frac{1}{B_r(\Sigma)} + \frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)\sqrt{n}} \sqrt{\log\left(\frac{B_r(\Sigma)\sqrt{n}}{\mathbf{r}(\Sigma)} \vee 2\right)}.$$

Thus, we indeed can drop the term $\frac{\log n}{\sqrt{n}}$ in bounds (5.27) and (5.30) simultaneously increasing the value of constant C. This yields bounds (5.1), (5.2) and completes the proof. □

6. Concluding remarks.

1. *Asymptotics of* $\|\hat{P}_r - P_r\|_2^2$. We start this section with deducing an asymptotic normality result from the nonasymptotic bound of Theorem 6. To this end, consider a sequence of problems in which the data is sampled from Gaussian distributions in $\mathbb{H}$ with mean zero and covariance $\Sigma = \Sigma^{(n)}$. Let $X = X^{(n)}$ be a centered Gaussian random vector in $\mathbb{H}$ with covariance operator $\Sigma = \Sigma^{(n)}$ and let $X_1 = X_1^{(n)}, \dots, X_n = X_n^{(n)}$ be i.i.d. copies of $X^{(n)}$. The sample covariance based on $(X_1^{(n)}, \dots, X_n^{(n)})$ is denoted by $\hat{\Sigma}_n$. Let $\sigma(\Sigma^{(n)})$ be the spectrum of $\Sigma^{(n)}$, $\mu_r^{(n)}, r \ge 1$ be distinct nonzero eigenvalues of $\Sigma^{(n)}$ arranged in decreasing order and $P_r^{(n)}, r \ge 1$ be the corresponding spectral projectors. As before, denote $\Delta_r^{(n)} := \{j : \sigma_j(\Sigma^{(n)}) = \mu_r^{(n)}\}$ and let $\hat{P}_r^{(n)}$ be the orthogonal projector on the direct sum of eigenspaces corresponding to the eigenvalues $\{\sigma_j(\hat{\Sigma}_n), j \in \Delta_r^{(n)}\}$.

Suppose that the spectral projector of $\Sigma^{(n)}$ to be estimated is $P^{(n)} = P_{r_n}^{(n)}$, the corresponding eigenvalue is $\mu^{(n)} = \mu_{r_n}^{(n)}$, its multiplicity is $m^{(n)} = m_{r_n}^{(n)}$ and its spectral gap is $\bar{g}^{(n)} = \bar{g}_{r_n}^{(n)}$. Denote

$$B_n := B_{r_n}(\Sigma^{(n)}) := 2\sqrt{2}\|C^{(n)}\Sigma^{(n)}C^{(n)}\|_2 \|P^{(n)}\Sigma^{(n)}P^{(n)}\|_2.$$

The following assumption on $\Sigma^{(n)}$ will be needed.

ASSUMPTION 1. Suppose the following conditions hold:

$$\sup_{n\ge 1} m^{(n)} < +\infty \quad \text{and} \quad \sup_{n\ge 1} \frac{\|\Sigma^{(n)}\|_\infty}{\bar{g}^{(n)}} < +\infty; \tag{6.1}$$

$$B_n \to \infty \quad \text{and} \quad \frac{\mathbf{r}(\Sigma^{(n)})}{B_n\sqrt{n}} \to 0 \qquad \text{as } n \to \infty. \tag{6.2}$$

Note that Assumption 1 implies that $\mathbf{r}(\Sigma^{(n)}) \to \infty$ and $\mathbf{r}(\Sigma^{(n)}) = o(n)$ as $n \to \infty$. This easily follows from

$$B_n \leq 2\sqrt{2}\sqrt{m^{(n)}}\left(\frac{\|\Sigma^{(n)}\|_\infty}{g^{(n)}}\right)^2 \sqrt{\mathbf{r}(\Sigma^{(n)})} = O\big(\sqrt{\mathbf{r}(\Sigma^{(n)})}\big)$$

and (6.2). It is also easy to see that, under mild further assumptions, $B_n \asymp \|\Sigma^{(n)}\|_2$.

COROLLARY 1. *Suppose Assumption 1 holds. Then*

$$\mathrm{Var}(\|\hat{P}^{(n)} - P^{(n)}\|_2^2) = \left(\frac{B_n}{n}\right)^2 (1 + o(1))$$

and the sequences of random variables

$$\left\{\frac{n(\|\hat{P}^{(n)} - P^{(n)}\|_2^2 - \mathbb{E}\|\hat{P}^{(n)} - P^{(n)}\|_2^2)}{B_n}\right\}_{n\geq 1} \tag{6.3}$$

and

$$\left\{\frac{(\|\hat{P}^{(n)} - P^{(n)}\|_2^2 - \mathbb{E}\|\hat{P}^{(n)} - P^{(n)}\|_2^2)}{\sqrt{\mathrm{Var}(\|\hat{P}^{(n)} - P^{(n)}\|_2^2)}}\right\}_{n\geq 1}$$

both converge in distribution to the standard normal random variable.

2. *Data-driven versions of asymptotic results*. Neither normal approximation bounds of Theorem 6, nor the asymptotic normality result of Corollary 1 could be directly used to construct confidence regions for spectral projectors of covariance operators or to develop hypotheses tests. The reason is that, in these results, the squared Hilbert–Schmidt norm $\|\hat{P}^{(n)} - P^{(n)}\|_2^2$ is centered with its expectation and normalized with its standard deviation [or, alternatively, with $\frac{n}{B_r(\Sigma)}$] that depend on unknown covariance operator Σ. It would be of interest to develop "data-driven" versions of these results, but this problem seems to be challenging and goes beyond the scope of the current paper. At the moment, we have only a partial solution (that is far from being perfect) of this problem in the case when the target spectral projector $P^{(n)}$ is one-dimensional (that is, the eigenvalue $\mu^{(n)}$ is of multiplicity one). We briefly outline such a result below. Assume that we are given a sample of size $3n$ of i.i.d. centered Gaussian vectors

$$\{X_1^{(n)}, \dots, X_n^{(n)}; \tilde{X}_1^{(n)}, \dots, \tilde{X}_n^{(n)}; \bar{X}_1^{(n)}, \dots, \bar{X}_n^{(n)}\},$$

with common covariance operator $\Sigma^{(n)}$. For each of the three subsamples of size n, define its sample covariance operator:

$$\hat{\Sigma}^{(n)} = \frac{1}{n}\sum_{i=1}^n X_i^{(n)} \otimes X_i^{(n)}, \qquad \tilde{\Sigma}^{(n)} = \frac{1}{n}\sum_{i=1}^n \tilde{X}_i^{(n)} \otimes \tilde{X}_i^{(n)},$$

$$\bar{\Sigma}^{(n)} = \frac{1}{n}\sum_{i=1}^n \bar{X}_i^{(n)} \otimes \bar{X}_i^{(n)}.$$

Let $\hat{P}^{(n)}$ be the orthogonal projector onto the eigenspace associated with the eigenvalue $\hat{\mu}^{(n)}$ of $\hat{\Sigma}^{(n)}$ (which is of multiplicity one with a high probability). Similarly, $\tilde{P}^{(n)}$ and $\bar{P}^{(n)}$ are the orthogonal projectors onto the eigenspaces associated with the eigenvalue $\tilde{\mu}^{(n)}$ of $\tilde{\Sigma}^{(n)}$ and the eigenvalue $\bar{\mu}^{(n)}$ of $\bar{\Sigma}^{(n)}$, respectively. Denote

$$\hat{b}^{(n)} = \sqrt{\langle \hat{P}^{(n)}, \tilde{P}^{(n)} \rangle} - 1 \quad \text{and} \quad \tilde{b}^{(n)} = \sqrt{\langle \tilde{P}^{(n)}, \bar{P}^{(n)} \rangle} - 1.$$

It turns out that the statistic $-2\hat{b}^{(n)}$ can be used as an estimator of the expectation $\mathbb{E}\|\hat{P}^{(n)} - P^{(n)}\|_2^2$ while the statistic $|(1+\hat{b}^{(n)})^2 - (1+\tilde{b}^{(n)})^2|$ can be used to estimate the standard deviation $\mathrm{Var}^{1/2}(\|\hat{P}^{(n)} - P^{(n)}\|_2^2)$ (note that $\hat{b}^{(n)}$ was introduced and studied in [13] as an estimator of a "bias parameter" of empirical spectral projectors and empirical eigenvectors). Moreover, it can be proved that, under Assumption 1, the sequence

$$\left\{ \frac{\|\hat{P}^{(n)} - P^{(n)}\|_2^2 + 2\hat{b}^{(n)}}{|(1+\hat{b}^{(n)})^2 - (1+\tilde{b}^{(n)})^2|} \right\}_{n\geq 1} \tag{6.4}$$

converges in distribution to a random variable with the following density:

$$\frac{1}{2}\left[\frac{1}{\beta} f\left(\frac{x-\alpha}{\beta}\right) + \frac{1}{\beta} f\left(\frac{x+\alpha}{\beta}\right)\right],$$

where $f(x) = \frac{1}{\pi(1+x^2)}$, $x \in \mathbb{R}$ is the standard Cauchy density, $\alpha = 1/6$ and $\beta = \frac{\sqrt{23}}{6}$ (in fact, the limit is distributed as the ratio $\frac{\xi}{|\eta|}$, where ξ, η are two correlated centered normal random variables and, as a result, the limit distribution is a mixture of two rescaled Cauchy distributions).

For the spiked covariance model (1.1) with $m, s_1^2, \ldots, s_m^2, \sigma^2$ being fixed and $p = p_n \to \infty$ as $n \to \infty$, it is easy to find a simpler version of data-driven normalization with the limit distribution being standard normal. For simplicity, assume that $m = 1$, so, the goal is to estimate the first principal components θ_1. Recall that in this case $B_n = B_1(\Sigma^{(n)}) = \frac{2\sqrt{2}(s_1^2+\sigma^2)\sigma^2\sqrt{p_n-1}}{s_1^4}$ [see (5.5)]. Based on the fact that $\|\Sigma^{(n)}\|_\infty = s_1^2 + \sigma^2$ and $\mathrm{tr}(\Sigma^{(n)}) = s_1^2 + p_n\sigma^2$, we suggest the following estimator of B_n: $\hat{B}_n = 2\sqrt{2}\frac{\hat{\mu}_1^{(n)}(\mathrm{tr}(\hat{\Sigma}^{(n)}) - \hat{\mu}_1^{(n)})}{(p_n\hat{\mu}_1^{(n)} - \mathrm{tr}(\Sigma^{(n)}))^2}(p_n - 1)^{3/2}$, where $\hat{\mu}_1^n$ is the largest eigenvalue of $\hat{\Sigma}^{(n)} = \frac{1}{n}\sum_{i=1}^n X_i^{(n)} \otimes X_i^{(n)}$. In the case of such a spiked covariance model, Assumption 1 is equivalent to $p = p_n \to \infty$ and $p = o(n)$. Under these assumptions, it is easy to prove that $\frac{\hat{B}_n}{n} = \frac{B_n}{n}(1 + o_{\mathbb{P}}(1))$. Let $P_1 = \theta_1 \otimes \theta_1$. Then it can be proved that the sequence

$$\left\{ \frac{n}{\hat{B}_n}\left(\|\hat{P}_1^{(n)} - P_1^{(n)}\|_2^2 + 2\hat{b}^{(n)}\right) \right\}_{n\geq 1} \tag{6.5}$$

converges in distribution to a standard normal random variable.

3. *Simulations*. To illustrate the asymptotic behaviour of standard PCA, we consider the following spiked covariance setting. Let $X_1, \dots, X_n$, $\tilde{X}_1, \dots, \tilde{X}_n$, $\bar{X}_1, \dots, \bar{X}_n$ be $3n$ i.i.d. random vectors in $\mathbb{R}^p$ with covariance $\Sigma = s_1^2(\theta_1 \otimes \theta_1) + \sigma^2 I_p$, $s_1^2 = 2$, $\sigma^2 = 1/10$, where θ_1 is an arbitrary unit vector in $\mathbb{R}^p$. For selected values of (n, p), we computed the statistic $\|\hat{P}_1^{(n)} - P_1\|_2^2$, $\hat{B}_n$ and the empirical bias estimators $\hat{b}_1^{(n)}$, $\tilde{b}_1^{(n)}$ as well as the statistics (6.4), (6.5) and

$$\frac{n}{B_n}(\|\hat{P}_1^{(n)} - P_1\|_2^2 + 2\hat{b}^{(n)}). \tag{6.6}$$

The last statistic is not completely data driven, it involves B_n that depends on unknown parameters of covariance. It is included in our simulations to study the impact of estimation of B_n on normal approximation. We performed 1000 replications of this experiment.

In Table 1, we compare the sample mean of the statistic $\|\hat{P}_1^{(n)} - P_1\|_2^2$ denoted by $\hat{m}_n$ (that provides an estimator of the risk $\mathbb{E}\|\hat{P}_1^{(n)} - P_1\|_2^2$ based on the repeated samples of size n) to the estimated risk $-2\hat{b}_1^{(n)}$ for each individual sample and the first-order approximation of the theoretical risk derived in (1.3) which can be computed easily in this model since $A_n := A_1(\Sigma) = 2\frac{(s_1^2+\sigma^2)\sigma^2}{s_1^4}(p-1)$. More precisely, in the second row of the table the sample means of $\frac{|2\hat{b}_1^{(n)}+\hat{m}_n|}{|\hat{m}_n|}$ over 1000 replications of the experiment are presented. The results show that $-2\hat{b}_1^{(n)}$ provides a somewhat better approximation of the risk $\mathbb{E}\|\hat{P}_1^{(n)} - P_1\|_2^2$ than the first-order approximation (1.3) for small sample size. For relatively large sample size, the first-order approximation (1.3) becomes more precise than the estimator $-2\hat{b}_1^{(n)}$.

In Table 2, we compare the sample variance of the statistic $\|\hat{P}_1^{(n)} - P_1\|_2^2$ denoted by $\hat{S}_n^2$ to the variance estimator $\tilde{V}_n := ((1 + \hat{b}_1^n)^2 - (1 + \tilde{b}_1^{(n)})^2)^2$ and also to the first-order approximation of the theoretical variance $\frac{B_n^2}{n^2}$ derived in (1.4) with $B_n = 2\sqrt{2}\frac{(s_1^2+\sigma^2)\sigma^2}{s_1^4}\sqrt{p-1}$. Again, in the second row of the table the sample means of

TABLE 1

Relative deviation of the risk approximation $\frac{A_n}{n}$ and the risk estimator $-2\hat{b}_1^{(n)}$ from the sample risk $\hat{m}_n$ for $p = 10^3$

n	**100**	**200**	**300**	**500**	**10^3**	**2.10^4**
$\frac{\|A_n/n-\hat{m}_n\|}{\|\hat{m}_n\|}$	0.49	0.24	0.15	0.1	0.049	0.008
$\frac{\|2\hat{b}_1^{(n)}+\hat{m}_n\|}{\|\hat{m}_n\|}$	0.07	0.06	0.054	0.052	0.045	0.036

TABLE 2

Relative deviation of the variance estimator $\tilde{V}_n$ and the variance approximation $\frac{B_n^2}{n^2}$ from the sample variance $\hat{S}_n^2$ for $p = 10^3$

n	100	200	300	500	10^3	2.10^4
$\frac{\lvert B_n^2/n^2 - \hat{S}_n^2\rvert}{\hat{S}_n^2}$	0.62	0.65	0.66	0.58	0.42	0.07
$\frac{\lvert \tilde{V}_n - \hat{S}_n^2\rvert}{\hat{S}_n^2}$	0.82	0.73	0.67	0.58	0.39	0.05

$\frac{|\tilde{V}_n - \hat{S}_n^2|}{\hat{S}_n^2}$ over 1000 replications of the experiment are presented. We observe that $\tilde{V}_n$ and $\frac{B_n^2}{n^2}$ provide reasonable approximation of the variance of $\|\hat{P}_1^{(n)} - P_1\|_2^2$ only for relatively large sample sizes.

Finally, we compute empirical densities of the statistics (6.4), (6.5) and (6.6) and compare them with their respective theoretical limiting distributions in Figure 1. For (6.5) and (6.6), we also provide the empirical mean and variance. This simulation study seems to confirm the theoretical limiting distributions we derived

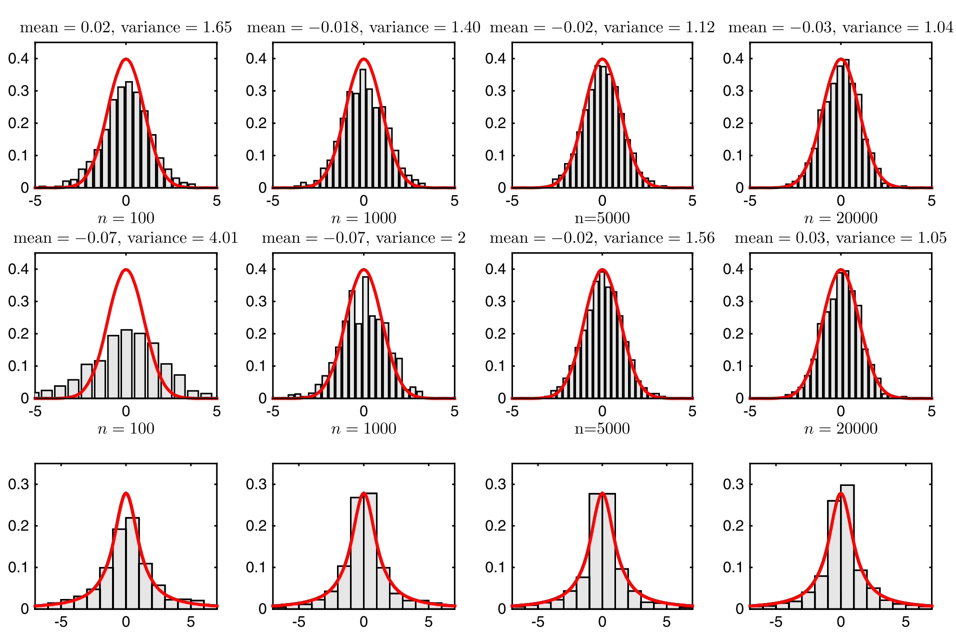

FIG. 1. *Top*: *empirical distribution of* (6.6) *and standard normal density for* $p = 1000$. *Middle*: *empirical distribution of* (6.5) *and standard normal density for* $p = 1000$. *Bottom*: *empirical distribution of* (6.4) *and density of the theoretical mixture of Cauchy distributions for* $p = 1000$.

TABLE 3
Ratio of $\hat{B}_n$ and B_n for $p = 10^3$

n	**100**	**200**	**500**	$\mathbf{10^3}$	$\mathbf{5.10^3}$
$\frac{\hat{B}_n}{B_n}$	0.65	0.78	0.90	0.94	0.99

above for the statistics (6.4), (6.5) and (6.6). In Table 3, we compare $\hat{B}_n$ to the theoretical value B_n in our experiment. It appears that the estimator $\hat{B}_n$ of B_n is reasonably precise for $n \geq 10^3$ in this spiked covariance model.

4. *Beyond the Gaussian case*. Extension of the results of the paper beyond the Gaussian case poses several challenging problems. An interesting class of models to consider would be log-concave distributions in high-dimensional spaces, but this seems to be a rather hard task. Even in a simpler case of sub-Gaussian random vectors, the difficulties are considerable. A major stumbling block is the extension of concentration inequality of Theorem 5 to the case of i.i.d. random vectors $X, X_1, \ldots, X_n$ that are not necessarily Gaussian. Standard concentration inequalities for product measures available in the literature do not seem to provide a straightforward solution of this problem. One exception is the model described by the following assumption.

ASSUMPTION 2. Let $X = \Sigma^{1/2} Z$, where Σ is a $d \times d$ covariance matrix and Z is a random vector in $\mathbb{H} = \mathbb{R}^d$, sampled from the density e^{-V}, $V : \mathbb{R}^d \mapsto \mathbb{R}$, with mean zero and identity covariance matrix I_d. In addition, assume that V is a smooth function with the Hessian V'' satisfying the condition $V''(x) \succcurlyeq c I_d, x \in \mathbb{R}^d$ for some constant $c > 0$.

This is a special class of log-concave distributions in $\mathbb{R}^d$ and it is known that linear forms $\langle X, u \rangle, u \in \mathbb{R}^d$ are sub-Gaussian random variables (for general log-concave distributions such linear forms are only sub-exponential). Moreover, concentration inequalities for Lipschitz functions similar to the Gaussian concentration inequality also hold under Assumption 2 (see [16], Theorems 2.7 and 2.18, Proposition 2.18). For this model, bound of Theorem 1 holds (as it does in the general sub-Gaussian case, see [14]) and so does bound of Theorem 2 implying that the effective rank $\mathbf{r}(\Sigma)$ plays the same role of the main complexity parameter as in the Gaussian case. Moreover, concentration bound of Theorem 5 could be also easily extended for the observations satisfying Assumption 2 with, essentially, identical proof. Thus, as in the Gaussian case, the study of normal approximation and concentration of $\|\hat{P}_r - P_r\|_2^2 - \mathbb{E}\|\hat{P}_r - P_r\|_2^2$ reduces to the study of $\|L_r(E)\|_2^2 - \mathbb{E}\|L_r(E)\|_2^2$. Unfortunately, such models (with an exception of the

Gaussian case) are of limited interest and extending concentration inequality of Theorem 5 to more general models looks rather challenging.

There are also serious difficulties with the development of normal approximation bounds for $\|L_r(E)\|_2^2 - \mathbb{E}\|L_r(E)\|_2^2$ in the non-Gaussian case. A possible approach is based on the following simple representation that is a consequence of (4.21):

$$n\big(\|L_r(E)\|_2^2 - \mathbb{E}\|L_r(E)\|_2^2\big) = \frac{2}{n}\sum_{i=1}^{n}\big(h_r(X_i, X_i) - \mathbb{E}h_r(X, X)\big) + \frac{2}{n}\sum_{i\neq j} h_r(X_i, X_j),$$

where

$$h_r(x, y) := \langle P_r x, P_r y\rangle\langle C_r x, C_r y\rangle, \qquad x, y \in \mathbb{H}. \tag{6.7}$$

The problem of normal approximation is then reduced to the study of degenerate U-statistic $U_{r,n} := \sum_{i<j} h_r(X_i, X_j)$ with symmetric kernel h_r. Such problems have been intensively studied in the literature, primarily, based on the martingale approach (see Hall [8], de Jong [5]). The following characteristics are involved in the conditions of asymptotic normality of $U_{r,n}$:

$$\begin{aligned} &\nu_{r,2} := \mathbb{E}^{1/2} h_r^2(X_1, X_2), \qquad \nu_{r,4} := \mathbb{E}^{1/4} h_r^4(X_1, X_2), \\ &\kappa_r := \operatorname{Var}\big(h_r(X, X)\big). \end{aligned} \tag{6.8}$$

We will also need

$$\begin{aligned} &\tilde{\nu}_{r,2} := \mathbb{E}^{1/2}\tilde{h}_r^2(X_1, X_2), \\ &\qquad \text{where } \tilde{h}_r(x, y) := \mathbb{E}h_r(X, x)h_r(X, y), x, y \in \mathbb{H}. \end{aligned} \tag{6.9}$$

Note that the kernel h_r depends only on the covariance Σ, but the corresponding characteristics $\nu_{r,2}, \nu_{r,4}, \tilde{\nu}_{r,2}$ depend on the whole distribution of X (that might not be completely characterized by Σ in the non-Gaussian case).

As in Corollary 1, let $X^{(n)}$ be a mean zero random vector in $\mathbb{H}$ with covariance $\Sigma^{(n)}$ (but not necessarily Gaussian) and let $X_1^{(n)}, \dots, X_n^{(n)}$ be its i.i.d. copies. Denote by $h_r^{(n)}$ the kernels generated by $\Sigma^{(n)}$ as in (6.7) and, similarly to (6.8), (6.9), define $\nu_{r,2}^{(n)}, \nu_{r,4}^{(n)}, \tilde{\nu}_{r,2}^{(n)}$. As before, denote $\mu^{(n)} = \mu_{r_n}^{(n)}$, $P^{(n)} = P_{r_n}^{(n)}$, $\hat{P}^{(n)} = \hat{P}_{r_n}^{(n)}$, $E^{(n)} := \hat{\Sigma}_n - \Sigma^{(n)}$, etc. Also denote

$$\begin{aligned} &h^{(n)} := h_{r_n}^{(n)}, \qquad \nu_2^{(n)} := \nu_{r_n,2}^{(n)}, \qquad \nu_4^{(n)} := \nu_{r_n,4}^{(n)}, \\ &\tilde{\nu}_2^{(n)} = \tilde{\nu}_{r_n,2}^{(n)}, \qquad \kappa^{(n)} := \kappa_{r_n}^{(n)}. \end{aligned}$$

Theorem 1 of Hall [8] yields the following condition of asymptotic normality of $\sum_{i<j} h^{(n)}(X_i^{(n)}, X_j^{(n)})$: if

$$\frac{(\tilde{\nu}_2^{(n)})^2 + n^{-1}(\nu_4^{(n)})^4}{(\nu_2^{(n)})^4} \to 0 \qquad \text{as } n \to \infty, \tag{6.10}$$

then the sequence of random variables

$$\left\{\frac{\sum_{i<j} h^{(n)}(X_i^{(n)}, X_j^{(n)})}{2^{-1/2} n \nu_2^{(n)}}\right\}_{n\geq 1}$$

is asymptotically standard normal. Under an additional assumption that

$$\frac{\kappa^{(n)}}{n(\nu_2^{(n)})^2} \to 0 \qquad \text{as } n \to \infty, \tag{6.11}$$

this result implies that the sequence

$$\left\{\frac{n(\|L_{r_n}(E^{(n)})\|_2^2 - \mathbb{E}\|L_{r_n}(E^{(n)})\|_2^2)}{2\sqrt{2}\nu_2^{(n)}}\right\}_{n\geq 1}$$

is also asymptotically standard normal. Together with concentration bound of Theorem 5 (that holds under Assumption 2), these considerations yield the following asymptotic normality result.

PROPOSITION 1. *Suppose, for all $n \geq 1$, $X^{(n)} \in \mathbb{R}^{d_n}$ satisfies Assumption 2 with some $V = V^{(n)}$ and with some constant $c > 0$ that does not depend on n. If, in addition, conditions (6.1), (6.10), (6.11) and the condition*

$$\frac{\mathbf{r}(\Sigma^{(n)})}{\nu_2^{(n)}\sqrt{n}} \to 0 \qquad \text{as } n \to \infty$$

hold, then the sequence

$$\left\{\frac{\|\hat{P}^{(n)} - P^{(n)}\|_2^2 - \mathbb{E}\|\hat{P}^{(n)} - P^{(n)}\|_2^2}{\sqrt{\operatorname{Var}(\|\hat{P}^{(n)} - P^{(n)}\|_2^2)}}\right\}_{n\geq 1}$$

is asymptotically standard normal.

In the Gaussian case, the conditions on the kernels $h_r^{(n)}$ could be simplified and expressed in terms of the quantity $B_r(\Sigma^{(n)})$ leading precisely to the conditions of asymptotic normality of Corollary 1 (and providing its alternative proof). However, in the general, not necessarily Gaussian case, the conditions of normal approximation would remain expressed in terms of the kernels $h_r^{(n)}$. Moreover, they depend not only on covariance matrices $\Sigma^{(n)}$, but on the actual distribution of the data, they are more complicated and much harder to interpret. In addition, the bounds on the accuracy of normal approximation in the non-Gaussian case that follow from Berry–Esseen-type bounds for U-statistics and for martingales (see, e.g., [7]) are significantly weaker than the bound of Theorem 6 proved in the previous section (but only in the Gaussian case). New Berry–Esseen type bounds for U-statistics by Eichelsbacher and Thäle [6], based on Stein method and Malliavin calculus in a Poisson space, look very promising, but they have been proved only for a Poissonized version of U-statistics.

Complete understanding of this and other aspects of the problem (such as concentration bounds of Theorem 5) in the non-Gaussian case is beyond the scope of this paper.

REFERENCES

[1] BIRNBAUM, A., JOHNSTONE, I. M., NADLER, B. and PAUL, D. (2013). Minimax bounds for sparse PCA with noisy high-dimensional data. *Ann. Statist.* **41** 1055–1084. MR3113803

[2] BUNEA, F. and XIAO, L. (2015). On the sample covariance matrix estimator of reduced effective rank population matrices, with applications to fPCA. *Bernoulli* **21** 1200–1230. MR3338661

[3] CAI, T. T., MA, Z. and WU, Y. (2013). Sparse PCA: Optimal rates and adaptive estimation. *Ann. Statist.* **41** 3074–3110. MR3161458

[4] DAUXOIS, J., POUSSE, A. and ROMAIN, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *J. Multivariate Anal.* **12** 136–154. MR0650934

[5] DE JONG, P. (1987). A central limit theorem for generalized quadratic forms. *Probab. Theory Related Fields* **75** 261–277. MR0885466

[6] EICHELSBACHER, P. and THÄLE, C. (2014). New Berry–Esseen bounds for non-linear functionals of Poisson random measures. *Electron. J. Probab.* **19** no. 102, 25. MR3275854

[7] HAEUSLER, E. (1988). On the rate of convergence in the central limit theorem for martingales with discrete and continuous time. *Ann. Probab.* **16** 275–299. MR0920271

[8] HALL, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivariate Anal.* **14** 1–16. MR0734096

[9] JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961

[10] JOHNSTONE, I. M. and LU, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *J. Amer. Statist. Assoc.* **104** 682–693. MR2751448

[11] KATO, T. (1980). *Perturbation Theory for Linear Operators*. Springer, New York.

[12] KNEIP, A. and UTIKAL, K. J. (2001). Inference for density families using functional principal component analysis. *J. Amer. Statist. Assoc.* **96** 519–542. MR1946423

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

[14] KOLTCHINSKII, V. and LOUNICI, K. (2016). Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli*. To appear. Available at arXiv:1405.2468.

[15] KOLTCHINSKII, V. I. (1998). Asymptotics of spectral projections of some random matrices approximating integral operators. In *High Dimensional Probability* (*Oberwolfach*, 1996). *Progress in Probability* **43** 191–227. Birkhäuser, Basel. MR1652327

[16] LEDOUX, M. (2001). *The Concentration of Measure Phenomenon*. *Mathematical Surveys and Monographs* **89**. Amer. Math. Soc., Providence, RI. MR1849347

[17] LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces*: *Isoperimetry and Processes*. Springer, Berlin. MR1102015

[18] LOUNICI, K. (2013). Sparse principal component analysis with missing observations. In *High Dimensional Probability VI*. *Progress in Probability* **66** 327–356. Birkhäuser, Basel. MR3443508

[19] PAUL, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist*. *Sinica* **17** 1617–1642. MR2399865

[20] VERSHYNIN, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing* 210–268. Cambridge Univ. Press, Cambridge. MR2963170

[21] VU, V. and LEI, J. (2012). Minimax rates of estimation for sparse PCA in high dimensions. *J*. *Mach*. *Learn*. *Res*. **22** 1278–1286.

SCHOOL OF MATHEMATICS
GEORGIA INSTITUTE OF TECHNOLOGY
ATLANTA, GEORGIA 30332-0160
USA
E-MAIL: vlad@math.gatech.edu
klounici@math.gatech.edu