Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2016, Vol. 52, No. 4, 1976–2013
DOI: 10.1214/15-AIHP705

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Asymptotics and concentration bounds for bilinear forms of spectral projectors of sample covariance

Vladimir Koltchinskii[1] and Karim Lounici[2]

School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0160, USA.
E-mail: vlad@math.gatech.edu; klounici@math.gatech.edu

Received 29 August 2014; revised 5 July 2015; accepted 31 July 2015

Abstract. Let $X, X_1, \dots, X_n$ be i.i.d. Gaussian random variables with zero mean and covariance operator $\Sigma = \mathbb{E}(X \otimes X)$ taking values in a separable Hilbert space $\mathbb{H}$. Let

$$\mathbf{r}(\Sigma) := \frac{\operatorname{tr}(\Sigma)}{\|\Sigma\|_\infty}$$

be the effective rank of Σ, $\operatorname{tr}(\Sigma)$ being the trace of Σ and $\|\Sigma\|_\infty$ being its operator norm. Let

$$\hat{\Sigma}_n := n^{-1} \sum_{j=1}^{n} (X_j \otimes X_j)$$

be the sample (empirical) covariance operator based on $(X_1, \dots, X_n)$. The paper deals with a problem of estimation of spectral projectors of the covariance operator Σ by their empirical counterparts, the spectral projectors of $\hat{\Sigma}_n$ (empirical spectral projectors). The focus is on the problems where both the sample size n and the effective rank $\mathbf{r}(\Sigma)$ are large. This framework includes and generalizes well known high-dimensional spiked covariance models. Given a spectral projector P_r corresponding to an eigenvalue μ_r of covariance operator Σ and its empirical counterpart $\hat{P}_r$, we derive sharp concentration bounds for bilinear forms of empirical spectral projector $\hat{P}_r$ in terms of sample size n and effective dimension $\mathbf{r}(\Sigma)$. Building upon these concentration bounds, we prove the asymptotic normality of bilinear forms of random operators $\hat{P}_r - \mathbb{E}\hat{P}_r$ under the assumptions that $n \to \infty$ and $\mathbf{r}(\Sigma) = o(n)$. In a special case of eigenvalues of multiplicity one, these results are rephrased as concentration bounds and asymptotic normality for linear forms of empirical eigenvectors. Other results include bounds on the bias $\mathbb{E}\hat{P}_r - P_r$ and a method of bias reduction as well as a discussion of possible applications to statistical inference in high-dimensional Principal Component Analysis.

Résumé. Soient $X, X_1, \dots, X_n$ des vecteurs gaussiens à valeurs dans un espace de Hilbert séparable $\mathbb{H}$, i.i.d. et centrés. Nous définissons l'opérateur de covariance $\Sigma = \mathbb{E}(X \otimes X)$ et le rang effectif de Σ

$$\mathbf{r}(\Sigma) := \frac{\operatorname{tr}(\Sigma)}{\|\Sigma\|_\infty}$$

où $\operatorname{tr}(\Sigma)$ est la trace of Σ et $\|\Sigma\|_\infty$ est sa norme d'opérateur. Nous considérons

$$\hat{\Sigma}_n := n^{-1} \sum_{j=1}^{n} (X_j \otimes X_j)$$

[1] Supported in part by NSF Grants DMS-1207808, DMS-1509739 and CCF-1415498.
[2] Supported in part by NSF CAREER Grant DMS-1454515 and Simons Collaboration Grant 315477.

l'opérateur de covariance empirique construit à partir des observations $(X_1, \dots, X_n)$. Ce papier considère le problème d'estimation des projecteurs spectraux de l'opérateur de covariance Σ par les projecteurs spectraux empiriques, c'est-à-dire les projecteurs spectraux de $\hat{\Sigma}_n$. Nous nous concentrons sur les problèmes où le nombre d'observations n et le rang effectif $\mathbf{r}(\Sigma)$ sont grands. Ce cadre inclut et généralise les modèles de spiked covariance en grande dimension. Soient P_r un projecteur spectral correspondant à une valeur propre μ_r de l'opérateur de covariance Σ et $\hat{P}_r$ sa version empirique. Nous établissons des bornes de concentrations fines sur les formes bilinéaires du projecteur empirique $\hat{P}_r$, qui dépendent du nombre d'observations n et de la dimension effective $\mathbf{r}(\Sigma)$. Nous exploitons ensuite ces bornes de concentration pour établir la normalité asymptotique des formes bilinéaires des opérateurs aléatoires $\hat{P}_r - \mathbb{E}\hat{P}_r$ sous les hypothèses que $n \to \infty$ et $\mathbf{r}(\Sigma) = o(n)$. Dans le cas particulier des valeurs propres de multiplicité 1, ces résultats sont reformulés en terme de bornes de concentration et de normalité asymptotique pour les formes linéaires des vecteurs propres empiriques. Nous prouvons aussi de nouveaux résultats sur le biais $\mathbb{E}\hat{P}_r - P_r$ incluant notamment une méthode de réduction du biais. Finalement, nous discutons des applications possibles de ces résultats à l'inférence statistique en grande dimension pour l'analyse en composantes principales.

MSC: 62H12

Keywords: Sample covariance; Spectral projectors; Effective rank; Principal Component Analysis; Concentration inequalities; Asymptotic distribution; Perturbation theory

1. Introduction

Principal Component Analysis (*PCA*) is among the most popular methods of exploring the covariance structure of a random process in a wide array of applications. It is of a particular interest in high-dimensional statistics as a tool of dimension reduction and feature extraction.

Let X be a random vector in $\mathbb{R}^p$ with zero mean and covariance matrix Σ. The classical PCA is based on estimating the eigenvalues and the associated spectral projectors of Σ by the eigenvalues and the spectral projectors of the sample covariance matrix $\hat{\Sigma}_n$ based on n i.i.d. replications of X, that is, the sample (empirical) eigenvalues and the sample (empirical) spectral projectors. Assessing the performance of the standard PCA raises naturally a question of how the sample eigenvalues and sample spectral projectors deviate from their population counterparts. In the 'standard setting,' where $p \geq 1$ is fixed and $n \to \infty$, Anderson [2] established the limiting joint distribution of the sample eigenvalues and the associated sample eigenvectors (see also Theorem 13.5.1 in [3]). These results have been extended in [10] to the case of i.i.d. data in infinite-dimensional Hilbert spaces (they have been used and further developed in numerous papers that followed, see, e.g., [24]).

A number of authors considered a 'high-dimensional setting,' where the dimension $p = p_n$ is allowed to grow with the sample size n. Marchenko and Pastur [23] derived the "limiting density" of the spectrum of $\hat{\Sigma}_n$ in the case when $\Sigma = I_p$ is the identity matrix and $\frac{p}{n} \to c \in (0, 1]$ as $n \to \infty$ (more precisely, they obtained the a.s. limit of the empirical distribution of the eigenvalues). Under the same conditions, Johnstone [11] proved that the largest empirical eigenvalue (properly normalized) converges in distribution to the Tracy–Widom law. The accuracy of this approximation was studied in [13,21]. Assuming that the covariance matrix Σ is the sum of the identity matrix and a small finite rank symmetric positive semi-definite perturbation, Baik, Ben Arrous and Peche [4] discovered a phase transition effect where the sample versions of the non-unit eigenvalues satisfy different asymptotic properties that depend on how far from 1 the non-unit eigenvalues are. Another line of research is a non-asymptotic theory of sample covariance where the main goal is to obtain sharp non-asymptotic bounds on the operator norm $\|\hat{\Sigma}_n - \Sigma\|_\infty$; a review of these results can be found in [30].

Concerning the estimation of spectral projectors, Johnstone and Lu [12] proved that the classical PCA approach could fail to produce a consistent estimator when $\frac{p}{n} \to c > 0$ as $n \to \infty$. To overcome this difficulty, several authors proposed alternative estimators of the covariance matrix Σ and studied their performance under various sparsity assumptions on Σ. See, for instance, [19,22,26,31] and the references cited therein.

We turn now to formulating the PCA problem in a general separable Hilbert space $\mathbb{H}$. This framework includes not only the classical high-dimensional setting, but also PCA for functional data (FPCA), see Ramsay and Silverman [27], and kernel PCA (KPCA) in machine learning, see Schölkopf, Smola and Müller [29], Blanchard, Bousquet and Zwald [6].

It will be assumed that $\mathbb{H}$ is a real Hilbert space, but, in some cases (especially, when one has to deal with resolvents of operators in $\mathbb{H}$), it has to be extended to a complex Hilbert space $\mathbb{H}_{\mathbb{C}} := \{u + iv : u, v \in \mathbb{H}\}$ with a standard extension of the inner product. In what follows, $\langle \cdot, \cdot \rangle$ denotes the inner product of $\mathbb{H}$ with $\|\cdot\|$ being the corresponding norm. With a little abuse of notation, we also denote by $\langle \cdot, \cdot \rangle$ the standard inner product in the space of Hilbert–Schmidt operators acting in $\mathbb{H}$, the corresponding Hilbert–Schmidt norm being denoted by $\|\cdot\|_2$. The notation $\|\cdot\|_\infty$ will be used for the operator norm of linear operators:

$$\|A\|_\infty := \sup_{\|u\|\le 1} \|Au\|, \quad A : \mathbb{H} \mapsto \mathbb{H}.$$

More generally, $\|\cdot\|_p$, $p \in [0, +\infty]$ denotes the Schatten p-norm. Given vectors $u, v \in \mathbb{H}$, $u \otimes v$ is the tensor product of u and v (that is, $u \otimes v$ is an operator from $\mathbb{H}$ into $\mathbb{H}$ acting as follows: $(u \otimes v)x = \langle v, x\rangle u$, $x \in \mathbb{H}$). If P is the orthogonal projector on a subspace $L \subset \mathbb{H}$, then $P^\perp$ denotes the projector on the orthogonal complement $L^\perp$.

The following notations are used throughout the paper: for nonnegative B_1, B_2, $B_1 \lesssim B_2$ (equivalently, $B_2 \gtrsim B_1$) means that there exists an absolute constant $C > 0$ such that $B_1 \le CB_2$. If $B_1 \lesssim B_2$ and $B_1 \gtrsim B_2$, we will write $B_1 \asymp B_2$. Sometimes, the signs $\lesssim$, $\gtrsim$ and $\asymp$ will be provided with subscripts. For instance, $B_1 \lesssim_a B_2$ would mean that $B_1 \le CB_2$, where C is a constant that might depend on a.

Let $X, X_1, \dots, X_n$ be i.i.d. random vectors in $\mathbb{H}$ with mean zero and $\mathbb{E}\|X\|^2 < +\infty$. Denote by $\Sigma = \mathbb{E}(X \otimes X)$ the covariance matrix of X and let

$$\hat{\Sigma} := \hat{\Sigma}_n := n^{-1} \sum_{j=1}^n X_j \otimes X_j$$

be the sample covariance based on the observations $(X_1, \dots, X_n)$. Since Σ is a compact symmetric nonnegatively definite operator (in fact, a trace class operator), it has the following spectral decomposition $\Sigma = \sum_{r=1}^\infty \mu_r P_r$, where $\mu_r = \mu_r(\Sigma)$ are distinct strictly positive eigenvalues of Σ (to be specific, arranged in decreasing order) and P_r are the corresponding spectral projectors (orthogonal projectors in $\mathbb{H}$). Clearly, $m_r := \operatorname{rank}(P_r) < +\infty$ is the multiplicity of the eigenvalue μ_r in the spectrum $\sigma(\Sigma)$ of Σ (in other words, it is the dimension of the eigenspace of Σ that corresponds to μ_r). It will be convenient in what follows to denote by $\sigma_j = \sigma_j(\Sigma)$, $j \ge 1$ the eigenvalues of Σ arranged in a nonincreasing order and *repeated with their multiplicities*. Let $\Delta_r := \{j : \sigma_j = \mu_r\}$. Then $\operatorname{card}(\Delta_r) = m_r$. Of course, the sample covariance $\hat{\Sigma}$ admits a similar spectral representation. Note that since the rank of $\hat{\Sigma}$ is at most n, it has at most n non-zero eigenvalues. Denote by $\hat{P}_r$ the orthogonal projector on the direct sum of eigenspaces of $\hat{\Sigma}$ corresponding to the eigenvalues $\{\sigma_j(\hat{\Sigma}), j \in \Delta_r\}$. It is well known (and it will be discussed in detail in the next section) that as soon as $\hat{\Sigma}$ is close enough to Σ in the operator norm, the eigenvalues $\{\sigma_j(\hat{\Sigma}), j \in \Delta_r\}$ are in a small neighborhood of μ_r and all other eigenvalues of $\hat{\Sigma}$ are separated from this neighborhood. Thus, for each r, if n is sufficiently large, there is a cluster $\{\sigma_j(\hat{\Sigma}), j \in \Delta_r\}$ of eigenvalues of $\hat{\Sigma}$ and the corresponding spectral projector $\hat{P}_r$ is a natural estimator of P_r (note that, in this case, $\operatorname{rank}(\hat{P}_r) = \operatorname{rank}(P_r) = m_r$).

We will be interested in asymptotic properties of the "empirical" spectral projector $\hat{P}_r$ as an estimator of the true spectral projector P_r. The following assumption holds throughout the paper:

Assumption 1. *Assume that $X, X_1, \dots, X_n$ are i.i.d. random variables sampled from a Gaussian distribution in $\mathbb{H}$ with zero mean and covariance Σ.*

We are especially interested in the case when not only the sample size n is large, but also the trace of matrix Σ, $\operatorname{tr}(\Sigma)$, is large as well (formally, one has to deal with a sequence of problems with covariances $\Sigma^{(n)}$ such that $\operatorname{tr}(\Sigma^{(n)}) \to \infty$ as $n \to \infty$). This is a crucial difference with other literature on PCA in Hilbert spaces (such as [10]) where it is typically assumed that $\operatorname{tr}(\Sigma)$ is a constant. This is what makes our results closer to what has been studied in the literature on PCA in high dimensions. To simplify the matter, we will assume that the individual eigenvalues in the spectrum of Σ are not large, so, the operator norm $\|\Sigma\|_\infty$ will be bounded by a constant. In this case, it makes sense to characterize the dimensionality of the problem by the so called "*effective rank*" of Σ (which also tends to infinity).

Definition 1. *The following quantity*

$$\mathbf{r}(\Sigma) := \frac{\operatorname{tr}(\Sigma)}{\|\Sigma\|_\infty}$$

will be called the effective rank of Σ.

Clearly, $\mathbf{r}(\Sigma) \leq \mathrm{rank}(\Sigma)$. Our setting includes, in particular, a popular high-dimensional *spiked covariance model* (see [11,12,25]) described in the following example.

Example 1 (Spiked covariance model). *Suppose that $\{\theta_k\}$ is an orthonormal basis in $\mathbb{H}$ and let $S := \sum_{k=1}^m s_k \zeta_k \theta_k$ be a "signal," s_j, $j = 1, \dots, m$ being nonrandom positive real numbers and ζ_j, $j = 1, \dots, m$ being i.i.d. standard normal random variables. Let $\dot{W}$ be a Gaussian white noise (a centered Gaussian r.v. with mean zero and identity covariance operator) that could be informally written as $\dot{W} = \sum_{k\geq 1} \eta_k \theta_k$, where $\{\eta_k\}$ are i.i.d. standard normal random variables (independent also of $\{\zeta_k\}$). Note that $\dot{W}$ is not a random vector in $\mathbb{H}$, but the family of linear functionals $\langle \dot{W}, u\rangle$, $u \in \mathbb{H}$ is well defined as an isonormal Gaussian process indexed by $\mathbb{H}$, that is, a centered Gaussian process with covariance function*

$$\mathbb{E}\langle \dot{W}, u\rangle\langle \dot{W}, v\rangle = \langle u, v\rangle, \quad u, v \in \mathbb{H}.$$

Thus, $\dot{W}$ is defined in a "weak sense" and it is well known that it can be also formally described as a random variable in a proper extension $\mathbb{H}_- \supset \mathbb{H}$ (often defined as a space of linear functionals on a dense linear subspace of $\mathbb{H}$). Suppose that S is observed in additive "white noise," that is, the observation of S is $X = S + \sigma \dot{W}$. More precisely, we will assume that the data consists of i.i.d. copies $X_1^{(n)}, \dots, X_n^{(n)}$ of a random vector $X^{(n)} \in \mathbb{H}$, where

$$X^{(n)} = S + \sigma \dot{W}^{(n)}, \qquad \dot{W}^{(n)} = \sum_{k=1}^{p} \eta_k \theta_k, \quad p > m, p = p_n \to \infty \text{ as } n \to \infty.$$

It is easy to see that $X^{(n)}$ can be rewritten as

$$X^{(n)} = \sum_{j=1}^{m} \sqrt{s_j^2 + \sigma^2}\xi_j\theta_j + \sigma \sum_{j=m+1}^{p_n} \xi_j\theta_j,$$

where ξ_j are i.i.d. standard normal random variables. The covariance of $X^{(n)}$ is

$$\Sigma^{(n)} = \mathbb{E}\big(X^{(n)} \otimes X^{(n)}\big) = \sum_{j=1}^{m}\big(s_j^2 + \sigma^2\big)(\theta_j \otimes \theta_j) + \sigma^2 P_{m,p_n},$$

where P_{m,p_n} denotes the orthogonal projector on the linear span of vectors θ_j, $j = m+1, \dots, p_n$. Clearly, for a fixed m,

$$\mathrm{tr}\big(\Sigma^{(n)}\big) = \sum_{j=1}^{m} s_j^2 + \sigma^2 p_n \asymp p_n \to \infty \quad \text{as } n \to \infty.$$

Estimation of the vectors $\theta_1, \dots, \theta_m$ (the components of the "signal") can be now viewed as a PCA problem for unknown covariance $\Sigma^{(n)}$.

Obviously, as it is usually done in the literature, one can also phrase the model of the above example as a sequence of high-dimensional problems in spaces $\mathbb{R}^p$, $p = p_n$ (without an explicit embedding of $\mathbb{R}^p$ into an infinite dimensional Hilbert space $\mathbb{H}$). In such a high-dimensional setting, the performance of the PCA is usually assessed by measuring the "alignment" between the target eigenvector and its estimator. In [5], the authors considered the loss function $L(a, b) := 2(1 - |\langle a, b\rangle|)$, where $a, b \in \mathbb{R}^p$ are unit vectors. It is closely related to the loss function

$$L'(a, b) := \|a \otimes a - b \otimes b\|_2^2 = 2\big(1 - \langle a, b\rangle^2\big),$$

that is used, for instance, in [8,19,31]. For the spiked covariance model described above, where $s_1 > \cdots > s_m > 0$, $\sigma^2 = 1$ and $m \geq 1$ are fixed and $\frac{p}{n} \to 0$ as $n \to \infty$, the following asymptotic representation of the risk of classical

PCA was obtained in [5]:

$$\mathbb{E}L(\hat{\theta}_j,\theta_j)=\left[\frac{(p-m)(1+s_j^2)}{ns_j^4}+\frac{1}{n}\sum_{k\neq j}\frac{(1+s_j^2)(1+s_k^2)}{(s_j^2-s_k^2)^2}\right]\bigl(1+o(1)\bigr),\quad \forall 1\le j\le m. \tag{1.1}$$

In [5], the authors also considered the setting $\frac{p}{n}\to c>0$ as $n\to\infty$, where the classical PCA is known to produce inconsistent estimators of the eigenvectors (see, for instance, [12]), and proposed a thresholding procedure related to, but more refined than the diagonal thresholding of Johnstone and Lu [12] that achieves optimality in the minimax sense for the loss $L(\cdot,\cdot)$ under sparsity conditions on the eigenvectors of Σ.

The loss functions L and L' are not suitable for the *support recovery* problem, that is, the estimation of the set $\mathrm{supp}(\theta_r):=\{j:\theta_r^{(j)}\neq 0\}$ for an eigenvector θ_r. To the best of our knowledge, very few results on this problem are available in the high-dimensional setting and they are obtained under very restrictive conditions on the covariance structure. For instance, in [1], a spiked covariance model was considered, where $\Sigma=s_1^2\theta_1\otimes\theta_1+\left(\begin{array}{c|c} I_k & 0\\ \hline 0 & \Gamma_{p-k}\end{array}\right)$, the first k entries of $\theta_1\in S^{p-1}$ are equal to $\pm\frac{1}{\sqrt{k}}$ for some $k\ge 1$ and Γ_{p-k} is symmetric positive semi-definite with $\|\Gamma_{p-k}\|_\infty\le 1$. The authors established an asymptotic support recovery result for the SDP-relaxation methodology introduced in [9], assuming that $k=O(\log p)$ is known, that $n\ge C(\Sigma)k\log(p-k)$, where $C(\Sigma)>0$ depends only on Σ, and also assuming the existence of a rank one solution of the SDP optimization problem.

Asymptotics of eigenvectors of sample covariance in a high-dimensional spiked covariance model were studied by Paul [25]. Namely, he considered a problem, where $X\sim N_p(0,\Sigma)$ with a spiked covariance matrix

$$\Sigma=\mathrm{diag}\bigl(s_1^2,s_2^2,\dots,s_m^2,1,\dots,1\bigr)$$

and fixed $s_1>\dots>s_m>1$, $m\ge 1$. Let $\hat{\theta}_j$ be the jth sample eigenvector and let $\hat{\theta}_j=(\hat{\theta}_{A,j},\hat{\theta}_{B,j})$, where $\hat{\theta}_{A,j}$ is the subvector corresponding to the first m components and $\hat{\theta}_{B,j}$ contains the remaining $p-m$ components. Paul [25] established that $\frac{\hat{\theta}_{B,j}}{\|\hat{\theta}_{B,j}\|}$ is uniformly distributed in the unit sphere S^{p-m-1} and is independent of $\|\hat{\theta}_{B,j}\|$. In addition, if $\frac{p}{n}-c=o(\frac{1}{n^{1/2}})$ with $c\in(0,1)$ and $s_j^2>1+\sqrt{c}$, then also

$$\sqrt{n}\biggl(\frac{\hat{\theta}_{A,j}}{\|\hat{\theta}_{A,j}\|}-e_j\biggr)\to N\bigl(0,\Sigma_j(s_j)\bigr)\quad \text{as } n\to\infty,$$

where

$$\Sigma_j(s_j)=\biggl(\frac{1}{1-c/(s_j^2-1)^2}\biggr)\sum_{1\le k\neq j\le m}\frac{(s_ks_j)^2}{(s_k^2-s_j^2)^2}(e_k\otimes e_k),$$

and e_k is the kth vector of the canonical basis of $\mathbb{R}^p$.

The spiked covariance model is a special case of more general models discussed in the next example.

Example 2 (More general spiked models). *Let Σ be a symmetric nonnegatively definite bounded operator that admits the following representation*

$$\Sigma=\sum_{r=1}^m\mu_rP_r+\Upsilon,$$

where μ_r are distinct positive numbers, P_r are projectors on mutually orthogonal finite dimensional subspaces of $\mathbb{H}$ and $\Upsilon:\mathbb{H}\mapsto\mathbb{H}$ is a nonnegatively definite symmetric bounded operator such that $P_r\Upsilon=\Upsilon P_r=0, r=1,\dots,m$. Moreover, suppose that $\|\Upsilon\|_\infty<\min_{1\le r\le m}\mu_r$ (in which case the spectrum of Σ is the union of two separated sets, $\{\mu_1,\dots,\mu_r\}$ and the spectrum of the operator Υ). Note that since Υ is not necessarily of trace class, it might not be a covariance operator of a random vector in $\mathbb{H}$ with a bounded strong second moment, and the same applies to Σ. However, Σ and Υ can be always viewed as covariance operators of "generalized random elements" (linear functionals on dense linear subspaces of $\mathbb{H}$), the same way as the identity operator is the covariance operator of the white *noise $\dot{W}$. Let P_{L_n} be the orthogonal projector on a finite-dimensional subspace $L_n \subset \mathbb{H}$. Suppose that* $\dim(L_n) \to \infty$ *as $n \to \infty$, $\bigcup_{n\geq 1} L_n$ is dense in $\mathbb{H}$ and $P_r\mathbb{H} \subset L_n$, $r = 1, \dots, m$ for all large enough n. Let $X^{(n)}$ be a centered Gaussian vector in $\mathbb{H}$ with covariance operator $\Sigma^{(n)} = P_{L_n}\Sigma P_{L_n}$ and let $X_1^{(n)}, \dots, X_n^{(n)}$ be i.i.d. copies of $X^{(n)}$. Then the problem becomes to estimate the principal spectral projectors P_r, $r = 1, \dots, m$ based on the sample $(X_1^{(n)}, \dots, X_n^{(n)})$, which is again a PCA problem. If* $\operatorname{tr}(\Upsilon) = \infty$, *then also* $\operatorname{tr}(\Sigma) = \infty$ *and* $\operatorname{tr}(\Sigma^{(n)}) \to \infty$ *as $n \to \infty$. One can go even further and consider the case of more general covariance operators $\Sigma^{(n)}$ of the observations $X_1^{(n)}, \dots, X_n^{(n)}$ that converge in some sense (for instance, in the sense of strong convergence of operators) to a symmetric nonnegatively definite operator Σ.*

In this paper and in a related paper [17], we develop a general theory of the asymptotic behavior of spectral projectors of the sample covariance operators that encompasses the spike covariance models described above as well as more general models of covariance operators for observations in a separable Hilbert space. We are especially interested in the case when $\mathbf{r}(\Sigma^{(n)}) = o(n)$, which is a necessary and sufficient condition for convergence of the sample covariance $\hat{\Sigma}_n$ to the true covariance Σ in the operator norm (and which, essentially, implies consistency of eigenvalues and of spectral projectors of sample covariance as estimators of their population counterparts). More specifically, our contributions include the following:

- In Section 2, we review recent moment bounds and concentration inequalities (see [18]) for $\|\hat{\Sigma}_n - \Sigma\|_\infty$ showing that, in the Gaussian case, the size of this random variable is completely characterized by two parameters, the operator norm $\|\Sigma\|_\infty$ and the effective rank $\mathbf{r}(\Sigma)$. This implies that $\|\hat{\Sigma}_n - \Sigma\|_\infty \to 0$ (a.s. and in the mean) if and only if $\mathbf{r}(\Sigma) = o(n)$. In the same section, we discuss several results in perturbation theory used throughout the paper.
- In Section 3, we obtain basic concentration inequalities for bilinear forms of empirical spectral projectors $\hat{P}_r$. In particular, we show that the following representation holds:

$$\hat{P}_r - \mathbb{E}P_r = L_r + R_r,$$

  where the main term L_r is linear with respect to $\hat{\Sigma} - \Sigma$ and, thus, it can be represented as a sum of i.i.d. random variables. The bilinear forms of the remainder term R_r satisfy sharp Gaussian type concentration inequalities, implying, in particular, that

$$\big|\langle R_r u, v\rangle\big| = O_{\mathbb{P}}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}}\sqrt{\frac{1}{n}}\right).$$

  If $\mathbf{r}(\Sigma) = o(n)$, the bilinear forms $\langle R_r u, v\rangle$ are of the order $o_{\mathbb{P}}(n^{-1/2})$ and asymptotic normality of the bilinear forms $\langle(\hat{P}_r - \mathbb{E}P_r)u, v\rangle$ can be easily deduced from the central limit theorem applied to the linear term $\langle L_r u, v\rangle$.
- In Section 4, we derive an asymptotic representation for the bias $\mathbb{E}\hat{P}_r - P_r$ of the empirical spectral projector $\hat{P}_r$ showing that its main term is an operator of the form $P_r W_r P_r$, where $\|W_r\|_\infty = O(\frac{\mathbf{r}(\Sigma)}{n})$, and the remainder is of the order $O(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}}\sqrt{\frac{1}{n}})$. This implies, in particular, that, in the case when $m_r = 1$ (the case of simple eigenvalue) the bias is proportional to the one-dimensional true spectral projector P_r up to a higher order term (indicating that a multiplicative correction can lead to a bias reduction).
- In Section 5 we derive the asymptotic distributions of bilinear forms of the empirical spectral projectors. In particular, we show that, under the assumption $\mathbf{r}(\Sigma) = o(n)$, the finite dimensional distributions of

$$\sqrt{n}\big\langle(\hat{P}_r - \mathbb{E}\hat{P}_r)u, v\big\rangle, \quad u, v \in \mathbb{H}$$

  converge weakly to the finite dimensional distributions of a Gaussian process. Our results show that the "variance part" of the error $\langle(\hat{P}_r - P_r)u, v\rangle$ is relatively well-behaved and that its dominating part is "bias," which might require further attention in statistical applications.
- In Section 6, we study in more detail the case of spectral projectors corresponding to an isolated eigenvalue of multiplicity $m_r = 1$. In this case, we prove the asymptotic normality of properly centered and normalized linear

forms $\langle\hat{\theta}_r, u\rangle, u \in \mathbb{H}$ of the corresponding sample eigenvector $\hat{\theta}_r$. Namely, we prove the weak convergence of finite dimensional distributions of stochastic processes

$$n^{1/2}\langle\hat{\theta}_r - \sqrt{1+b_r}\theta_r, u\rangle, \quad u \in \mathbb{H}$$

to the finite dimensional distributions of a Gaussian process for properly chosen "bias parameters" b_r. We also obtain non-asymptotic concentration bounds for the l_∞-norm $\|\hat{\theta}_r - \sqrt{1+b_r}\theta_r\|_{\ell_\infty}$. In addition, we propose an estimator of the bias parameter b_r that converges to the true parameter at a rate faster than $n^{-1/2}$ and develop a bias reduction method based on this estimator. At the end of Section 6, we briefly discuss potential applications of these results, in particular, to the problem of support recovery of the eigenvector of interest as well as sparse PCA estimation.

In a related paper [17], we obtained an asymptotic formula for the Hilbert–Schmidt norm risk $\mathbb{E}\|\hat{P}_r - P_r\|_2^2$ of empirical spectral projectors under the assumption that $\mathbf{r}(\Sigma) = o(n)$. In a special case of spiked covariance model, it implies representation (1.1). We also proved in [17] the asymptotic normality of a properly normalized sequence

$$\big\{\|\hat{P}_r - P_r\|_2^2 - \mathbb{E}\|\hat{P}_r - P_r\|_2^2\big\}.$$

2. Preliminaries

In this section, we review bounds on the operator norm $\|\hat{\Sigma}_n - \Sigma\|_\infty$ and discuss several well known facts of perturbation theory that will be frequently used in what follows.

2.1. *Bounds on the operator norm* $\|\hat{\Sigma}_n - \Sigma\|_\infty$

It is well known (see [30]) that, for a sub–sub-Gaussian isotropic distribution (that is, in the case when $\Sigma = I_p$), with probability at least $1 - e^{-t}$

$$\|\hat{\Sigma}_n - \Sigma\|_\infty \le C\left(\sqrt{\frac{p}{n}} \vee \frac{p}{n} \vee \sqrt{\frac{t}{n}} \vee \frac{t}{n}\right), \tag{2.1}$$

for some numerical constant $C > 0$ (see Theorem 5.39 and the comments after this theorem). The proof is based on an ε-net argument that does not yield an optimal bound for general (nonisotropic) sub–sub-Gaussian distributions. In [7, 20], similar results were derived for sub–sub-Gaussian distributions and low-rank covariance matrices. However the bounds in the last two papers are suboptimal by a logarithmic factor (they are based on a noncommutative Bernstein inequality).

The following theorems (see Koltchinskii and Lounici [18]) could be viewed as an extension of bound (2.1) to the nonisotropic and infinite-dimensional case. These results show that in the Gaussian case, the size of the operator norm $\|\hat{\Sigma}_n - \Sigma\|_\infty$ is completely characterized by the operator norm $\|\Sigma\|_\infty$ and the effective rank $\mathbf{r}(\Sigma)$. In particular, if $\Sigma = \Sigma^{(n)}$ with $\|\Sigma^{(n)}\|_\infty$ uniformly bounded, then $\|\hat{\Sigma}_n - \Sigma^{(n)}\|_\infty \to 0$ a.s. as $n \to \infty$ if and only if $\mathbf{r}(\Sigma^{(n)}) = o(n)$.

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 $p \ge 1$,*

$$\mathbb{E}^{1/p}\|\hat{\Sigma}_n - \Sigma\|_\infty^p \asymp_p \|\Sigma\|_\infty \max\left\{\sqrt{\frac{\mathbf{r}(\Sigma)}{n}}, \frac{\mathbf{r}(\Sigma)}{n}\right\}. \tag{2.2}$$

We will also need a concentration inequality for $\|\hat{\Sigma}_n - \Sigma\|_\infty$.

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 exist a constant $C_1 > 0$ such that for all $t \ge 1$ with probability at least $1 - e^{-t}$,*

$$\big|\|\hat{\Sigma}_n - \Sigma\|_\infty - \mathbb{E}\|\hat{\Sigma}_n - \Sigma\|_\infty\big| \le C_1\|\Sigma\|_\infty\left[\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee 1\right)\sqrt{\frac{t}{n}} \vee \frac{t}{n}\right]. \tag{2.3}$$

As a consequence of this bound and (2.2), *with some constant* $C_2 > 0$ *and with the same probability*

$$\|\hat{\Sigma}_n - \Sigma\|_\infty \le C_2\|\Sigma\|_\infty\left[\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \frac{\mathbf{r}(\Sigma)}{n} \vee \sqrt{\frac{t}{n}} \vee \frac{t}{n}\right]. \tag{2.4}$$

Remark 1.

1. *The notion of effective rank* $\mathbf{r}(\Sigma)$ *and the results of Theorems* 1 *and* 2 *can be extended to the case of Gaussian random variables in separable Banach spaces*, *see* [18].
2. *The bound of Theorem* 1 *and bound* (2.4) *of Theorem* 2 *hold in a more general case*, *when* $X, X_1, \dots, X_n$ *are i.i.d. centered sub-Gaussian vectors in* $\mathbb{H}$, *that is*, *for some constant* $c > 0$,

$$\big\|\langle X, u\rangle\big\|_{\psi_2}^2 \le c\mathbb{E}\langle X, u\rangle^2, \quad u \in \mathbb{H}. \tag{2.5}$$

Here $\|\cdot\|_{\psi_2}$ *is the Orlicz norm for* $\psi_2(t) = e^{t^2} - 1, t \ge 0$ (*the Orlicz norm in the space of sub-Gaussian random variables*).

2.2. *Several facts on perturbation theory*

In this section, we discuss several useful results of perturbation theory (see Kato [14]) adapted for our purposes. Some facts in the same direction can be found in Koltchinskii [16] and Kneip and Utikal [15].

Let $\Sigma : \mathbb{H} \mapsto \mathbb{H}$ be a compact symmetric operator (in applications, it will be the covariance operator of a random vector X in $\mathbb{H}$). Let $\sigma(\Sigma)$ be the spectrum of Σ. It is well known that the following spectral representation holds

$$\Sigma = \sum_{r\ge 1} \mu_r P_r$$

with distinct non-zero eigenvalues μ_r and spectral projectors P_r and with the series converging in the operator norm. We will also use notations $\sigma_i = \sigma_i(\Sigma)$, Δ_r, m_r, etc., already introduced in Section 1.

Define

$$g_r := g_r(\Sigma) := \mu_r - \mu_{r+1} > 0, \quad r \ge 1.$$

Let $\bar{g}_r := \bar{g}_r(\Sigma) := \min(g_{r-1}, g_r)$ for $r \ge 2$ and $\bar{g}_1 := g_1$. In what follows, $\bar{g}_r$ will be called *the* r*th spectral gap*, or *the spectral gap of eigenvalue* μ_r.

Let now $\tilde{\Sigma}$ be another compact symmetric operator in $\mathbb{H}$ with spectrum $\sigma(\tilde{\Sigma})$ and eigenvalues $\tilde{\sigma}_i = \sigma_i(\tilde{\Sigma}), i \ge 1$ (arranged in nonincreasing order and repeated with their multiplicities). Denote $E := \tilde{\Sigma} - \Sigma$. According to well known Lidskii's inequality,

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

This implies that, for all $r \ge 1$,

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

and

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

Suppose that

$$\|E\|_\infty < \frac{\bar{g}_r}{2}. \tag{2.6}$$

Then, 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, if

$$\|E\|_\infty < \frac{1}{4} \min_{1 \le s \le r} \bar{g}_s =: \bar{\delta}_r,$$

then the set $\{\sigma_j(\tilde{\Sigma}) : j \in \bigcup_{s=1}^{r} \Delta_s\}$ of the largest eigenvalues of $\tilde{\Sigma}$ will be divided into r clusters, each of them being of diameter strictly smaller than $2\bar{\delta}_r$ and the distance between any two clusters being larger than $2\bar{\delta}_r$. In principle, this allows one 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.

Denote $\tilde{P}_r$ the orthogonal projector on the direct sum of eigenspaces of $\tilde{\Sigma}$ corresponding to the eigenvalues $\tilde{\sigma}_j$, $j \in \Delta_r$ (in other words, to the rth cluster of eigenvalues of Σ). Denote also

$$C_r := \sum_{s \ne r} \frac{1}{\mu_r - \mu_s} P_s.$$

Lemma 1. *The following bound holds*:

$$\|\tilde{P}_r - P_r\|_\infty \le 4 \frac{\|E\|_\infty}{\bar{g}_r}. \tag{2.7}$$

Moreover,

$$\tilde{P}_r - P_r = L_r(E) + S_r(E), \tag{2.8}$$

where

$$L_r(E) := C_r E P_r + P_r E C_r \tag{2.9}$$

and

$$\big\|S_r(E)\big\|_\infty \le 14 \left(\frac{\|E\|_\infty}{\bar{g}_r}\right)^2. \tag{2.10}$$

Proof. Assume first that $\|E\|_\infty \le \bar{g}_r/4$. Denote by γ_r the circle in $\mathbb{C}$ with center μ_r and radius $\frac{\bar{g}_r}{2}$. Note that the eigenvalues μ_r of Σ and $\tilde{\sigma}_j$, $j \in \Delta_r$ of $\tilde{\Sigma}$ are inside this circle while the rest of the eigenvalues of these operators are outside. Combining these facts with the Riesz formula for spectral projectors (see, for instance, [14], p. 39), we get that

$$\tilde{P}_r = -\frac{1}{2\pi i} \oint_{\gamma_r} R_{\tilde{\Sigma}}(\eta)\, d\eta,$$

where $R_A(\eta) = (A - \eta I)^{-1}$ is the resolvent of an operator A in $\mathbb{H}$.

The following computation is standard:

$$\begin{aligned} R_{\tilde{\Sigma}}(\eta) &= R_{\Sigma+E}(\eta) = (\Sigma + E - \eta I)^{-1} \\ &= \big[(\Sigma - \eta I)\big(I + (\Sigma - \eta I)^{-1} E\big)\big]^{-1} \\ &= \big(I + R_\Sigma(\eta) E\big)^{-1} R_\Sigma(\eta) \\ &= \sum_{k \ge 0} (-1)^k \big[R_\Sigma(\eta) E\big]^k R_\Sigma(\eta), \qquad \eta \in \gamma_r. \end{aligned} \tag{2.11}$$

The series in the right-hand side converges absolutely in the operator norm since

$$\big\|R_\Sigma(\eta)E\big\|_\infty \le \big\|R_\Sigma(\eta)\big\|_\infty\|E\|_\infty \le \frac{2}{\bar g_r}\|E\|_\infty \le \frac{1}{2} < 1, \quad \eta \in \gamma_r,$$

where we have used that $\|R_\Sigma(\eta)\|_\infty \le \frac{2}{\bar g_r}$ for any $\eta \in \gamma_r$. Next, we get from (2.11) that

$$\begin{aligned}\tilde P_r &= -\frac{1}{2\pi i}\oint_{\gamma_r} R_\Sigma(\eta)\,d\eta - \frac{1}{2\pi i}\oint_{\gamma_r}\sum_{k\ge 1}(-1)^k\big[R_\Sigma(\eta)E\big]^k R_\Sigma(\eta)\,d\eta \\ &= P_r - \frac{1}{2\pi i}\oint_{\gamma_r}\sum_{k\ge 1}(-1)^k\big[R_\Sigma(\eta)E\big]^k R_\Sigma(\eta)\,d\eta,\end{aligned}$$

where we again used the Riesz formula. Thus,

$$\begin{aligned}\|\tilde P_r - P_r\|_\infty &\le 2\pi\frac{\bar g_r}{2}\frac{1}{2\pi}\left(\frac{2}{\bar g_r}\right)^2\|E\|_\infty\sum_{k=0}^{\infty}\left(\frac{2}{\bar g_r}\|E\|_\infty\right)^k \\ &\le \frac{2\|E\|_\infty/\bar g_r}{1-2\|E\|_\infty/\bar g_r}.\end{aligned}$$

Under the assumption $\|E\|_\infty \le \bar g_r/4$, we get that

$$\|\tilde P_r - P_r\|_\infty \le \frac{4\|E\|_\infty}{\bar g_r},$$

so, (2.7) holds in this case. Since $\tilde P_r$, P_r are both orthogonal projectors, it is easy to see that $\|\tilde P_r - P_r\|_\infty \le 1$, implying that (2.7) also holds when $\|E\|_\infty > \bar g_r/4$.

We turn to the proof of the remaining bounds. It is easy to check (using the orthogonality of operators C_rEP_r, P_rEC_r) that

$$\big\|L_r(E)\big\|_\infty = \|C_rEP_r + P_rEC_r\|_\infty \le \sqrt{2}\|C_r\|_\infty\|E\|_\infty \le \frac{\sqrt{2}}{\bar g_r}\|E\|_\infty.$$

Therefore,

$$\big\|S_r(E)\big\|_\infty = \big\|\tilde P_r - P_r - L_r(E)\big\|_\infty \le \|\tilde P_r - P_r\|_\infty + \big\|L_r(E)\big\|_\infty \le 1 + \frac{\sqrt{2}}{\bar g_r}\|E\|_\infty. \tag{2.12}$$

Assuming that $\|E\|_\infty \le g_r/3$, we have the following representation:

$$\hat P_r - P_r = L'_r(E) + S'_r(E), \tag{2.13}$$

where

$$L'_r(E) = \frac{1}{2\pi i}\oint_{\gamma_r} R_\Sigma(\eta)ER_\Sigma(\eta)\,d\eta$$

and

$$S'_r(E) := -\frac{1}{2\pi i}\oint_{\gamma_r}\sum_{k\ge 2}(-1)^k\big[R_\Sigma(\eta)E\big]^k R_\Sigma(\eta)\,d\eta.$$

As for the first order linear term $L'_r(E)$, we use the spectral representation of the resolvent $R_\Sigma(\eta)$,

$$R_\Sigma(\eta) = \sum_{j\ge 1}\frac{1}{\mu_j - \eta}P_j$$

(with the series convergent in operator norm uniformly in $\eta \in \gamma_r$), to derive that

$$
\begin{aligned}
L'_r(E) &= \frac{1}{2\pi i}\oint_{\gamma_r}\sum_{j\geq 1}\frac{1}{\mu_j-\eta}P_j E\sum_{j\geq 1}\frac{1}{\mu_j-\eta}P_j\,d\eta \\
&= \sum_{j_1,j_2\geq 1}\frac{1}{2\pi i}\oint_{\gamma_r}\frac{d\eta}{(\mu_{j_1}-\eta)(\mu_{j_2}-\eta)}P_{j_1}EP_{j_2}.
\end{aligned}
$$

Note that, if $j_1 = r$, $j_2 = s \neq r$, then, by Cauchy formula,

$$
\frac{1}{2\pi i}\oint_{\gamma_r}\frac{d\eta}{(\mu_{j_1}-\eta)(\mu_{j_2}-\eta)} = \frac{1}{2\pi i}\oint_{\gamma_r}\frac{d\eta}{(\eta-\mu_r)(\eta-\mu_s)} = \frac{1}{\mu_r-\mu_s}.
$$

Similarly, if $j_2 = r$, $j_1 = s \neq r$, then

$$
\frac{1}{2\pi i}\oint_{\gamma_r}\frac{d\eta}{(\mu_{j_1}-\eta)(\mu_{j_2}-\eta)} = \frac{1}{\mu_r-\mu_s}.
$$

In all other cases,

$$
\frac{1}{2\pi i}\oint_{\gamma_r}\frac{d\eta}{(\mu_{j_1}-\eta)(\mu_{j_2}-\eta)} = 0.
$$

Therefore,

$$
L'_r(E) = \sum_{s\neq r}\frac{1}{\mu_r-\mu_s}P_sEP_r + \sum_{s\neq r}\frac{1}{\mu_r-\mu_s}P_rEP_s = C_rEP_r + P_rEC_r = L_r(E)
$$

and, as a consequence, $S'_r(E) = S_r(E)$. Similarly to (2.7), it can be proved that, under the assumption $\|E\|_\infty \leq \bar{g}_r/3$,

$$
\big\|S_r(E)\big\|_\infty \leq 12\left(\frac{\|E\|_\infty}{\bar{g}_r}\right)^2. \tag{2.14}
$$

Bound (2.10) now easily follows from (2.14) and (2.12). □

We will state below a simple generalization of Lemma 1. Given $I = \{r_1, r_1+1, \dots, r_2\} \subset \mathbb{N}$, $1 \leq r_1 \leq r_2$, denote $\Delta_I := \{j : \sigma_j = \mu_r, r \in I\}$ and let $P_I = \sum_{r\in I} P_r$ be the orthogonal projector on the direct sum of the eigenspaces of Σ corresponding to the eigenvalues $\mu_r, r \in I$. Denote $L_I := \mu_{r_1} - \mu_{r_2}$ and define

$$
\bar{g}_I := \min(\mu_{r_2}-\mu_{r_2+1}, \mu_{r_1-1}-\mu_{r_1}) \quad \text{if } r_1 > 1 \quad \text{and} \quad \bar{g}_I := \mu_{r_2}-\mu_{r_2+1} \quad \text{if } r_1 = 1.
$$

Finally, let $\tilde{P}_I$ be the orthogonal projector on the direct sum of the eigenspaces of $\tilde{\Sigma}$ corresponding to the eigenvalues $\tilde{\sigma}_j$, $j \in \Delta_I$. Note that, if $\|E\|_\infty < \bar{g}_I/2$, then the set of eigenvalues $\{\tilde{\sigma}_j : j \in \Delta_I\}$ is covered by the interval $(\mu_{r_2} - \bar{g}_I/2, \mu_{r_1} + \bar{g}_I/2)$ and the rest of the eigenvalues of $\tilde{\Sigma}$ are outside of the interval $[\mu_{r_2} - \bar{g}_I/2, \mu_{r_1} + \bar{g}_I/2]$. Denote

$$
\gamma_I := \big\{\eta \in \mathbb{C} : \operatorname{dist}\big(\eta; [\mu_{r_2}, \mu_{r_1}]\big) = \bar{g}_I/2\big\}.
$$

In what follows, γ_I will be viewed as a counter-clockwise contour and in (2.17) below it can be replaced by an arbitrary contour γ that separates the eigenvalues $\{\mu_r : r \in I\}$ from the rest of the spectrum of Σ.

Lemma 2. *The following bound holds*:

$$
\|\tilde{P}_I - P_I\|_\infty \leq 4\left(1 + \frac{2}{\pi}\frac{L_I}{\bar{g}_I}\right)\frac{\|E\|_\infty}{\bar{g}_I}. \tag{2.15}
$$

Moreover, the following representation holds

$$\tilde{P}_I - P_I = L_I(E) + S_I(E), \tag{2.16}$$

where the linear part $L_I(E)$ *is given by*

$$L_I(E) := \frac{1}{2\pi i} \oint_{\gamma_I} R_\Sigma(\eta) E R_\Sigma(\eta)\, d\eta \tag{2.17}$$

and the remainder $S_I(E)$ *satisfies the bound*

$$\big\| S_I(E) \big\|_\infty \le 15\left(1 + \frac{2}{\pi}\frac{L_I}{\bar{g}_I}\right)\left(\frac{\|E\|_\infty}{\bar{g}_I}\right)^2. \tag{2.18}$$

The proof of this lemma is quite similar to the proof of Lemma 1 and it will be skipped.

3. Concentration inequalities for bilinear forms 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).

The goal of this section is to derive useful representations and concentration bounds for the bilinear forms $\langle (\hat{P}_r - P_r)u, v\rangle, u, v \in \mathbb{H}$ of spectral projectors for a properly isolated eigenvalue μ_r. These results will be used in subsequent sections to show asymptotic normality of the bilinear forms $\langle (\hat{P}_r - P_r)u, v\rangle$ under the assumption that $\mathbf{r}(\Sigma) = o(n)$.

In the results below, it will be assumed that, for some $\gamma \in (0, 1)$,

$$\mathbb{E}\|\hat{\Sigma} - \Sigma\|_\infty \le \frac{(1-\gamma)\bar{g}_r}{2}. \tag{3.1}$$

In view of Theorem 1, this assumption implies that

$$\|\Sigma\|_\infty\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \frac{\mathbf{r}(\Sigma)}{n}\right) \lesssim \frac{\bar{g}_r}{2} \le \|\Sigma\|_\infty.$$

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

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

If

$$C'\|\Sigma\|_\infty\left(\sqrt{\frac{t}{n}} \vee \frac{t}{n}\right) \le \frac{\gamma \bar{g}_r}{2},$$

then $\mathbb{P}(\|\hat{\Sigma}_n - \Sigma\|_\infty < \frac{\bar{g}_r}{2}) \ge 1 - e^{-t}$. It was pointed out in Section 2.2 that, in this case, the cluster $\{\sigma_j(\hat{\Sigma}_n) : j \in \Delta_r\}$ of eigenvalues of $\hat{\Sigma}$ is well separated from the rest of the spectrum of $\hat{\Sigma}$ and the spectral projector $\hat{P}_r$ can be viewed as an estimator of the spectral projector P_r (in particular, these two projectors are of the same rank m_r). It will be shown below that, under assumption (3.1), the bilinear form $\langle (\hat{P}_r - P_r)u, v\rangle$ can be represented as a sum of a part that is linear in $\hat{\Sigma}_n - \Sigma$ and a remainder that is smaller than the linear part, provided that $\mathbf{r}(\Sigma) = o(n)$. The linear part is defined in terms of operator

$$L_r := C_r(\hat{\Sigma} - \Sigma)P_r + P_r(\hat{\Sigma} - \Sigma)C_r = n^{-1}\sum_{j=1}^n (C_r X_j \otimes P_r X_j + P_r X_j \otimes C_r X_j)$$

and the remainder in terms of operator

$$R_r := (\hat{P}_r - P_r) - \mathbb{E}(\hat{P}_r - P_r) - L_r = \hat{P}_r - \mathbb{E}\hat{P}_r - L_r.$$

Theorem 3. *Suppose that, for some $\gamma \in (0,1)$, (3.1) is satisfied. Then, there exists a constant $D_\gamma > 0$ such that, for all $u, v \in \mathbb{H}$, the following bound holds with probability at least $1 - e^{-t}$:*

$$\big|\langle R_r u, v\rangle\big| \le D_\gamma \frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{t}{n}} \vee \frac{t}{n}\right)\sqrt{\frac{t}{n}}\|u\|\|v\|. \tag{3.2}$$

Taking into account Theorem 2, note that, if $\Sigma = \Sigma^{(n)}$, $\|\Sigma^{(n)}\|_\infty = O(1)$, $\bar{g}_r = \bar{g}_r^{(n)}$ is bounded away from zero and $\mathbf{r}(\Sigma^{(n)}) \le cn$ for a sufficiently small c, then bound (3.2) implies that

$$\langle R_r u, v\rangle = O_{\mathbb{P}}\big(n^{-1/2}\big) \quad \text{as } n \to \infty, u, v \in \mathbb{H}.$$

Moreover, if $\mathbf{r}(\Sigma^{(n)}) = o(n)$, it follows from (3.2) that

$$\langle R_r u, v\rangle = o_{\mathbb{P}}\big(n^{-1/2}\big). \tag{3.3}$$

Let

$$\xi(u,v) := \langle X, P_r v\rangle\langle X, C_r u\rangle, \quad u, v \in \mathbb{H}$$

and let

$$\xi_j(u,v) := \langle X_j, P_r v\rangle\langle X_j, C_r u\rangle, \quad u, v \in \mathbb{H}, j = 1, \dots, n$$

be independent copies of ξ. Note that

$$\mathbb{E}\xi(u,v) = \mathbb{E}\langle X, P_r v\rangle \mathbb{E}\langle X, C_r u\rangle = 0$$

and

$$\begin{aligned}\mathbb{E}\xi(u,v)\xi\big(u',v'\big) &= \mathbb{E}\langle X, P_r v\rangle\big\langle X, P_r v'\big\rangle \mathbb{E}\langle X, C_r u\rangle\big\langle X, C_r u'\big\rangle \\ &= \big\langle P_r\Sigma P_r v, v'\big\rangle\big\langle C_r\Sigma C_r u, u'\big\rangle,\end{aligned}$$

where it was used that Gaussian random variables $\langle X, P_r v\rangle$, $\langle X, C_r u\rangle$ are uncorrelated and, hence, independent. This implies that the covariance function of the random field $\xi(u,v) + \xi(v,u)$, $u, v \in \mathbb{H}$ is given by

$$\begin{aligned}\tilde{\Gamma}\big(u,v;u',v'\big) &:= \mathbb{E}\big(\xi(u,v) + \xi(v,u)\big)\big(\xi\big(u',v'\big) + \xi\big(v',u'\big)\big) \\ &= \big\langle P_r\Sigma P_r v, v'\big\rangle\big\langle C_r\Sigma C_r u, u'\big\rangle + \big\langle P_r\Sigma P_r v, u'\big\rangle\big\langle C_r\Sigma C_r u, v'\big\rangle \\ &\quad + \big\langle P_r\Sigma P_r u, u'\big\rangle\big\langle C_r\Sigma C_r v, v'\big\rangle + \big\langle P_r\Sigma P_r u, v'\big\rangle\big\langle C_r\Sigma C_r v, u'\big\rangle.\end{aligned}$$

The bilinear forms

$$n^{1/2}\langle L_r u, v\rangle = n^{-1/2}\sum_{j=1}^n \big(\xi_j(u,v) + \xi_j(v,u)\big), \quad u, v \in \mathbb{H}$$

have the same covariance function $\tilde{\Gamma}$. Moreover, it is easy to see that, under proper assumptions, they are asymptotically normal. Thus, (3.3) implies the asymptotic normality of $\langle \hat{P}_r - \mathbb{E}\hat{P}_r u, v\rangle$, $u, v \in \mathbb{H}$. This result will be discussed in detail in the next section.

The next statement immediately follows from Theorem 3 and Bernstein inequality for sums of i.i.d. subexponential random variables $\xi_j(u,v)$, $j = 1, \dots, n$. In particular, it shows that, under the assumptions of Theorem 3,

$$\langle \hat{P}_r - \mathbb{E}\hat{P}_r u, v\rangle = O_{\mathbb{P}}\big(n^{-1/2}\big) \quad \text{as } n \to \infty, u, v \in \mathbb{H}.$$

Corollary 1. *Under the assumption of Theorem 3, with some constants $D, D_\gamma > 0$, for all $u, v \in \mathbb{H}$ and for all $t \geq 1$ with probability at least $1 - e^{-t}$,*

$$\begin{aligned}\left|\langle \hat{P}_r - \mathbb{E}\hat{P}_r u, v\rangle\right| &\leq D\frac{\|\Sigma\|_\infty}{\bar{g}_r}\sqrt{\frac{t}{n}}\|u\|\|v\| \\ &\quad + D_\gamma \frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{t}{n}} \vee \frac{t}{n}\right)\sqrt{\frac{t}{n}}\|u\|\|v\|. \end{aligned} \tag{3.4}$$

Remark 2. *Note that $\xi_j(u, v) = 0$, $j = 1, \dots, n$ in the case when both u and v belong to the eigenspace corresponding to the eigenvalue μ_r (since, in this case, $C_r u = C_r v = 0$), or in the case when both u and v are in the orthogonal complement of this space (since then $P_r u = P_r v = 0$). Therefore, for such u, v the first term in the righthand side of (3.4) could be dropped and the bound reduces only to the second term.*

We now turn to the proof of Theorem 3.

Proof of Theorem 3. Clearly, it will be enough to prove bound (3.2) for $\|u\| \leq 1$, $\|v\| \leq 1$. This will be assumed throughout the proof.

First note that $L_r = L_r(E)$, where $E := \hat{\Sigma} - \Sigma$. Since

$$\mathbb{E}L_r = \mathbb{E}L_r(E) = 0,$$

we get that

$$R_r = L_r(E) + S_r(E) - \mathbb{E}\big(L_r(E) + S_r(E)\big) - L_r(E) = S_r(E) - \mathbb{E}S_r(E)$$

(recall Lemma 1).

Under condition (3.1), we have $\mathbf{r}(\Sigma) \lesssim n$. Theorem 2 implies that with some constant $C' > 0$ and for all $t \geq 1$ with probability at least $1 - e^{-t}$,

$$\|\hat{\Sigma} - \Sigma\|_\infty \leq \mathbb{E}\|\hat{\Sigma} - \Sigma\|_\infty + C'\|\Sigma\|_\infty\left(\sqrt{\frac{t}{n}} \vee \frac{t}{n}\right).$$

If $C'\|\Sigma\|_\infty(\sqrt{\frac{t}{n}} \vee \frac{t}{n}) \leq \frac{\gamma \bar{g}_r}{4}$, then it is easy to see that $t \lesssim n$ and, for some $C > 0$,

$$C'\|\Sigma\|_\infty\left(\sqrt{\frac{t}{n}} \vee \frac{t}{n}\right) \leq C\|\Sigma\|_\infty\sqrt{\frac{t}{n}}.$$

We will first assume that

$$C\|\Sigma\|_\infty\sqrt{\frac{t}{n}} \leq \frac{\gamma \bar{g}_r}{4} \tag{3.5}$$

(the proof of the concentration bound in the opposite case will be much easier). Let

$$\delta_n(t) := \mathbb{E}\|\hat{\Sigma}_n - \Sigma\|_\infty + C\|\Sigma\|_\infty\sqrt{\frac{t}{n}}. \tag{3.6}$$

Clearly, $\mathbb{P}\{\|\hat{\Sigma} - \Sigma\|_\infty \geq \delta_n(t)\} \leq e^{-t}$.

The main part of the proof is the study of concentration of the random variable $\langle S_r(E)u, v\rangle$ around its expectation. To this end, we first study the concentration properties of "truncated" random variable

$$\big\langle S_r(E)u, v\big\rangle\varphi\left(\frac{\|E\|_\infty}{\delta}\right),$$

where, for some $\gamma \in (0,1)$, φ is a Lipschitz function with constant $\frac{1}{\gamma}$ on $\mathbb{R}_+$, $0 \le \varphi(s) \le 1$, $\varphi(s) = 1$, $s \le 1$, $\varphi(s) = 0$, $s > 1+\gamma$, and $\delta > 0$ is such that $\|E\|_\infty \le \delta$ with a high probability.

Our main tool is the following concentration inequality that easily follows from Gaussian isoperimetric inequality.

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 $L > 0$:*

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

Suppose that, for a real number M,

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

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

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

Lemma 3 will be applied to the function

$$f(X_1,\dots,X_n) := \bigl\langle S_r(E)u, v\bigr\rangle \varphi\biggl(\frac{\|E\|_\infty}{\delta}\biggr).$$

With a little abuse of notation, assume for now that $X_1,\dots,X_n$ are nonrandom vectors in $\mathbb{H}$. For $X_1',\dots,X_n' \in \mathbb{H}$, denote

$$E' = \hat{\Sigma}' - \Sigma, \quad \hat{\Sigma}' = n^{-1}\sum_{j=1}^n X_j' \otimes X_j'.$$

Let $\hat{P}_r'$ be the orthogonal projector on the direct sum of eigenspaces of $\hat{\Sigma}'$ corresponding to its eigenvalues $\{\sigma_j(\hat{\Sigma}') : j \in \Delta_r\}$.

We have to check the Lipschitz condition for the function f. We will start with the following simple fact based on perturbation theory bounds of Section 2.2.

Lemma 4. *Let $\gamma \in (0,1)$ and suppose that*

$$\delta \le \frac{1-\gamma}{1+\gamma}\frac{\bar{g}_r}{2}. \tag{3.7}$$

Suppose also that

$$\|E\|_\infty \le (1+\gamma)\delta \quad \textit{and} \quad \bigl\|E'\bigr\|_\infty \le (1+\gamma)\delta. \tag{3.8}$$

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

$$\bigl\|S_r(E) - S_r\bigl(E'\bigr)\bigr\|_\infty \le C_\gamma \frac{\delta}{\bar{g}_r^2}\bigl\|E - E'\bigr\|_\infty. \tag{3.9}$$

Proof. Note that, by the definition of $S_r(E)$,

$$S_r\bigl(E'\bigr) - S_r(E) = \hat{P}_r' - \hat{P}_r - L_r\bigl(E' - E\bigr). \tag{3.10}$$

For $\hat{P}_r' - \hat{P}_r$, we will use decomposition of Lemma 2 that yields:

$$\hat{P}_r' - \hat{P}_r = \hat{L}_r\bigl(E' - E\bigr) + \hat{S}_r\bigl(E' - E\bigr) \tag{3.11}$$

with

$$\hat{L}_r\big(E'-E\big)=\frac{1}{2\pi i}\oint_{\gamma_r} R_{\hat{\Sigma}}(\eta)\big(E'-E\big)R_{\hat{\Sigma}}(\eta)\,d\eta$$

and

$$\big\|\hat{S}_r\big(E'-E\big)\big\|_\infty\le 15\bigg(1+\frac{4}{\pi}\frac{\|E\|_\infty}{\bar{g}_r-2\|E\|_\infty}\bigg)\frac{\|E-E'\|_\infty^2}{(\bar{g}_r-2\|E\|_\infty)^2}. \tag{3.12}$$

More precisely, we used Lemma 2 with $\hat{\Sigma}$ instead of Σ and with $\hat{\Sigma}'$ instead of $\tilde{\Sigma}$. Observe that the set of eigenvalues $\{\sigma_j(\hat{\Sigma}):j\in\Delta_r\}$ can be written as $\{\mu_i(\hat{\Sigma}):i\in I\}$ for some $I\subset\mathbb{N}$. Also, we have $\Delta_I=\Delta_r$, $\hat{P}_I=\hat{P}_r$ and $\hat{P}'_I=\hat{P}'_r$. Finally, in our case $L_I\le 2\|E\|_\infty$ and

$$\bar{g}_I\ge\bar{g}_r-2\|E\|_\infty.$$

We could also replace the contour γ_I used in Lemma 2 by the circle γ_r since these two contours separate the same part of the spectrum of $\hat{\Sigma}$ from the rest of the spectrum.

Note now that

$$\begin{aligned}\hat{L}_r\big(E'-E\big)-L_r\big(E'-E\big)&=\frac{1}{2\pi i}\oint_{\gamma_r}\big(R_{\hat{\Sigma}}(\eta)-R_\Sigma(\eta)\big)\big(E'-E\big)R_{\hat{\Sigma}}(\eta)\,d\eta\\&\quad+\frac{1}{2\pi i}\oint_{\gamma_r}R_\Sigma(\eta)\big(E'-E\big)\big(R_{\hat{\Sigma}}(\eta)-R_\Sigma(\eta)\big)\,d\eta,\end{aligned}$$

which implies the bound

$$\begin{aligned}\big\|\hat{L}_r\big(E'-E\big)-L_r\big(E'-E\big)\big\|_\infty&\le\bigg[\frac{1}{2\pi}\oint_{\gamma_r}\big\|R_{\hat{\Sigma}}(\eta)-R_\Sigma(\eta)\big\|_\infty\big\|R_{\hat{\Sigma}}(\eta)\big\|_\infty\,d\eta\\&\quad+\frac{1}{2\pi}\oint_{\gamma_r}\big\|R_\Sigma(\eta)\big\|_\infty\big\|R_{\hat{\Sigma}}(\eta)-R_\Sigma(\eta)\big\|_\infty\,d\eta\bigg]\big\|E-E'\big\|_\infty.\end{aligned} \tag{3.13}$$

Since $\|E\|_\infty<\bar{g}_r/2$, we get that, for all $\eta\in\gamma_r$,

$$\big\|R_\Sigma(\eta)\big\|_\infty\le\frac{2}{\bar{g}_r},\qquad\big\|R_{\hat{\Sigma}}(\eta)\big\|_\infty\le\frac{2}{\bar{g}_r-2\|E\|_\infty}.$$

Using respresentation (2.11) (with $\hat{\Sigma}$ instead of $\tilde{\Sigma}$), we easily get that

$$\big\|R_{\hat{\Sigma}}(\eta)-R_\Sigma(\eta)\big\|_\infty\le\sum_{k\ge 1}\big\|R_\Sigma(\eta)\big\|_\infty^{k+1}\|E\|_\infty^k\le\frac{2}{\bar{g}_r}\frac{(2/\bar{g}_r)\|E\|_\infty}{1-(2/\bar{g}_r)\|E\|_\infty}=\frac{4\|E\|_\infty}{\bar{g}_r(\bar{g}_r-2\|E\|_\infty)}.$$

Due to these bounds, it follows from (3.13) that

$$\big\|\hat{L}_r\big(E'-E\big)-L_r\big(E'-E\big)\big\|_\infty\le\frac{8\|E\|_\infty}{(\bar{g}_r-2\|E\|_\infty)^2}\big\|E-E'\big\|_\infty. \tag{3.14}$$

We combine now (3.10), (3.11), (3.12) and (3.14) to get

$$\begin{aligned}&\big\|S_r(E)-S_r\big(E'\big)\big\|_\infty\\&\quad\le\frac{8\|E\|_\infty}{(\bar{g}_r-2\|E\|_\infty)^2}\big\|E-E'\big\|_\infty+15\bigg(1+\frac{4}{\pi}\frac{\|E\|_\infty}{\bar{g}_r-2\|E\|_\infty}\bigg)\frac{\|E-E'\|_\infty^2}{(\bar{g}_r-2\|E\|_\infty)^2}\\&\quad\le\frac{8\|E\|_\infty}{(\bar{g}_r-2\|E\|_\infty)^2}\big\|E-E'\big\|_\infty+30\bigg(1+\frac{4}{\pi}\frac{\|E\|_\infty}{\bar{g}_r-2\|E\|_\infty}\bigg)\frac{\|E\|_\infty\vee\|E'\|_\infty}{(\bar{g}_r-2\|E\|_\infty)^2}\big\|E-E'\big\|_\infty.\end{aligned}$$

To complete the proof, it is enough to use conditions (3.7), (3.8) that, in particular, imply

$$\bar{g}_r - 2\|E\|_\infty \ge \bar{g}_r - 2(1+\gamma)\delta \ge \gamma \bar{g}_r.$$

□

Lemma 5. *Suppose that, for some $\gamma \in (0, 1/2)$,*

$$\delta \le \frac{1-2\gamma}{1+2\gamma}\frac{\bar{g}_r}{2}. \tag{3.15}$$

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

$$\bigl|f(X_1, \dots, X_n) - f\bigl(X_1', \dots, X_n'\bigr)\bigr| \le D_\gamma \frac{\delta}{\bar{g}_r^2} \frac{\|\Sigma\|_\infty^{1/2} + \delta^{1/2}}{\sqrt{n}} \left(\sum_{j=1}^n \bigl\|X_j - X_j'\bigr\|^2\right)^{1/2}. \tag{3.16}$$

Proof. Since $\varphi(\frac{\|E\|_\infty}{\delta}) = 0$ if $\|E\|_\infty \ge (1+\gamma)\delta$, bound (2.10) of Lemma 1 implies that

$$\bigl|f(X_1, \dots, X_n)\bigr| = \left|\langle S_r(E)u, v\rangle \varphi\left(\frac{\|E\|_\infty}{\delta}\right)\right| \le 14(1+\gamma)^2 \frac{\delta^2}{\bar{g}_r^2}. \tag{3.17}$$

Using now bounds (3.9), (3.17) and the fact that φ 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

$$\|E\|_\infty \le (1+\gamma)\delta, \qquad \bigl\|E'\bigr\|_\infty \le (1+\gamma)\delta, \tag{3.18}$$

the following inequality holds:

$$\begin{aligned} &\left|\langle S_r(E)u, v\rangle \varphi\left(\frac{\|E\|_\infty}{\delta}\right) - \langle S_r\bigl(E'\bigr)u, v\rangle \varphi\left(\frac{\|E'\|_\infty}{\delta}\right)\right| \\ &\quad \le \bigl\|S_r(E) - S_r\bigl(E'\bigr)\bigr\|_\infty + \frac{14(1+\gamma)^2}{\gamma}\frac{\delta}{\bar{g}_r^2}\bigl\|E - E'\bigr\|_\infty \\ &\quad \le \left(C_\gamma + \frac{14(1+\gamma)^2}{\gamma}\right)\frac{\delta}{\bar{g}_r^2}\bigl\|E - E'\bigr\|_\infty. \end{aligned} \tag{3.19}$$

It remains to prove a similar bound in the case when

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

(when both norms are larger than $(1+\gamma)\delta$, the function φ is equal to zero and the bound is trivial). First consider the case when $\|E - E'\|_\infty \ge \gamma\delta$. Then, in view of (3.17), we have

$$\begin{aligned} &\left|\langle S_r(E)u, v\rangle \varphi\left(\frac{\|E\|_\infty}{\delta}\right) - \langle S_r\bigl(E'\bigr)u, v\rangle \varphi\left(\frac{\|E\|_\infty}{\delta}\right)\right| \\ &\quad = \left|\langle S_r(E)u, v\rangle \varphi\left(\frac{\|E\|_\infty}{\delta}\right)\right| \le 14(1+\gamma)^2\frac{\delta^2}{\bar{g}_r^2} \le \frac{14(1+\gamma)^2}{\gamma}\frac{\delta}{\bar{g}_r^2}\bigl\|E - E'\bigr\|_\infty. \end{aligned}$$

Finally, if $\|E - E'\|_\infty < \gamma\delta$, we have that $\|E'\|_\infty \le (1+2\gamma)\delta$ and, taking into account assumption (3.15), we can repeat the argument in the case (3.18) ending up with the same bound as (3.19) with constant $C_{2\gamma} + \frac{14(1+2\gamma)^2}{\gamma}$ instead of $C_\gamma + \frac{14(1+\gamma)^2}{\gamma}$ in the right-hand side. Thus, with some constant $L_\gamma > 0$,

$$\left|\langle S_r(E)u, v\rangle \varphi\left(\frac{\|E\|_\infty}{\delta}\right) - \langle S_r\bigl(E'\bigr)u, v\rangle \varphi\left(\frac{\|E\|_\infty}{\delta}\right)\right| \le L_\gamma \frac{\delta}{\bar{g}_r^2}\bigl\|E - E'\bigr\|_\infty. \tag{3.20}$$

We will now control $\|E - E'\|_\infty$. Note that

$$\begin{aligned}
\|E - E'\|_\infty &= \sup_{\|u\|\le 1, \|v\|\le 1} \big|\langle (E - E')u, v\rangle\big| \\
&= \sup_{\|u\|\le 1, \|v\|\le 1} \left| n^{-1}\sum_{j=1}^n \langle X_j, u\rangle\langle X_j, v\rangle - \langle X'_j, u\rangle\langle X'_j, v\rangle\right| \\
&\le \sup_{\|u\|\le 1, \|v\|\le 1} \left| n^{-1}\sum_{j=1}^n \langle X_j, u\rangle\langle X_j - X'_j, v\rangle\right| + \sup_{\|u\|\le 1, \|v\|\le 1} \left| n^{-1}\sum_{j=1}^n \langle X_j - X'_j, u\rangle\langle X'_j, v\rangle\right| \\
&\le \sup_{\|u\|\le 1}\left(n^{-1}\sum_{j=1}^n \langle X_j, u\rangle^2\right)^{1/2} \sup_{\|v\|\le 1}\left(n^{-1}\sum_{j=1}^n \langle X_j - X'_j, v\rangle^2\right)^{1/2} \\
&\quad + \sup_{\|u\|\le 1}\left(n^{-1}\sum_{j=1}^n \langle X_j - X'_j, u\rangle^2\right)^{1/2} \sup_{\|v\|\le 1}\left(n^{-1}\sum_{j=1}^n \langle X'_j, v\rangle^2\right)^{1/2} \\
&\le \frac{\|\hat\Sigma\|_\infty^{1/2} + \|\hat\Sigma'\|_\infty^{1/2}}{\sqrt{n}} \left(\sum_{j=1}^n \|X_j - X'_j\|^2\right)^{1/2}.
\end{aligned}$$

Clearly, it is enough to consider the case when at least one of the norms $\|E\|_\infty, \|E'\|_\infty$ is not larger than 2δ. To be specific, assume that $\|E\|_\infty \le 2\delta$. Then

$$\|\hat\Sigma\|_\infty^{1/2} + \|\hat\Sigma'\|_\infty^{1/2} \le 2\|\hat\Sigma\|_\infty^{1/2} + \|E - E'\|_\infty^{1/2} \le 2\|\Sigma\|_\infty^{1/2} + 2\sqrt{2\delta} + \|E - E'\|_\infty^{1/2}.$$

Therefore,

$$\|E - E'\|_\infty \le \frac{2\|\Sigma\|_\infty^{1/2} + 2\sqrt{2\delta}}{\sqrt{n}} \left(\sum_{j=1}^n \|X_j - X'_j\|^2\right)^{1/2} + \frac{\|E - E'\|_\infty^{1/2}}{\sqrt{n}} \left(\sum_{j=1}^n \|X_j - X'_j\|^2\right)^{1/2},$$

which easily implies

$$\|E - E'\|_\infty \le \frac{4\|\Sigma\|_\infty^{1/2} + 4\sqrt{2\delta}}{\sqrt{n}} \left(\sum_{j=1}^n \|X_j - X'_j\|^2\right)^{1/2} \vee \frac{4}{n}\sum_{j=1}^n \|X_j - X'_j\|^2. \tag{3.21}$$

Now substitute the last bound in the right-hand side of (3.20) and also observe that, in view of (3.17), the left-hand side of (3.20) can be also upper bounded by $28(1+\gamma)^2\frac{\delta^2}{\bar g_r^2}$. Therefore, we get that with some constant $L'_\gamma > 0$,

$$\begin{aligned}
&\left|\langle S_r(E)u, v\rangle \varphi\left(\frac{\|E\|_\infty}{\delta}\right) - \langle S_r(E')u, v\rangle \varphi\left(\frac{\|E\|_\infty}{\delta}\right)\right| \\
&\quad \le 4L_\gamma \frac{\delta}{\bar g_r^2}\left[\frac{\|\Sigma\|_\infty^{1/2} + \sqrt{2\delta}}{\sqrt{n}}\left(\sum_{j=1}^n \|X_j - X'_j\|^2\right)^{1/2} \vee \frac{1}{n}\sum_{j=1}^n \|X_j - X'_j\|^2\right] \wedge 28(1+\gamma)^2\frac{\delta^2}{\bar g_r^2} \\
&\quad \le L'_\gamma \frac{\delta}{\bar g_r^2}\left[\frac{\|\Sigma\|_\infty^{1/2} + \sqrt{2\delta}}{\sqrt{n}}\left(\sum_{j=1}^n \|X_j - X'_j\|^2\right)^{1/2} \vee \left(\frac{1}{n}\sum_{j=1}^n \|X_j - X'_j\|^2 \wedge \delta\right)\right].
\end{aligned} \tag{3.22}$$

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 (3.22) (since a similar expression is a part of the first term). This yields bound (3.16). □

We set $\delta := \delta_n(t)$, where $\delta_n(t)$ is defined by (3.6). Without loss of generality, we can assume that $t \ge \log 4$ and $e^{-t} \le 1/4$ (the result can be extended to all $t \ge 1$ by adjusting the constants). Recall that, by Theorem 2, $\mathbb{P}\{\|E\|_\infty \ge \delta\} \le \frac{1}{4}$. In addition, in view of (3.1) and (3.5), $\delta_n(t) \le (1 - \frac{\gamma}{2})\frac{\bar{g}_r}{2} = \frac{1-2\gamma'}{1+2\gamma'}\frac{\bar{g}_r}{2}$ for some $\gamma' \in (0, 1/2)$. Thus the function $f(X_1, \dots, X_n)$ satisfies the Lipschitz condition (3.16) with some constant $D_\gamma' = D_{\gamma'}$.

To complete the proof of Theorem 3, denote $\mathrm{Med}(\eta)$ a median of a random variable η, and let $M := \mathrm{Med}(\langle S_r(E)u, v\rangle)$. Since $f(X_1, \dots, X_n) = \langle S_r(E)u, v\rangle$ on the event $\{\|E\|_\infty < \delta\}$, we have

$$\begin{aligned}\mathbb{P}\bigl\{f(X_1, \dots, X_n) \ge M\bigr\} &\ge \mathbb{P}\bigl\{f(X_1, \dots, X_n) \ge M, \|E\|_\infty < \delta\bigr\} \\ &= \mathbb{P}\bigl\{\bigl\langle S_r(E)u, v\bigr\rangle \ge M, \|E\|_\infty < \delta\bigr\} \ge \mathbb{P}\bigl\{\bigl\langle S_r(E)u, v\bigr\rangle \ge M\bigr\} - \mathbb{P}\bigl\{\|E\|_\infty \ge \delta\bigr\} \ge 1/4\end{aligned}$$

and, similarly,

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

It follows from Lemma 3 and Lemma 5 that with some constant $D_\gamma > 0$, for all $t \ge 1$ with probability at least $1 - e^{-t}$,

$$\bigl|f(X_1, \dots, X_n) - M\bigr| \le D_\gamma \frac{\delta}{\bar{g}_r^2}\bigl(\|\Sigma\|_\infty^{1/2} + \delta^{1/2}\bigr)\|\Sigma\|_\infty^{1/2}\sqrt{\frac{t}{n}}.$$

Therefore, under condition (3.5), we get that for all $t \ge 1$, with probability at least $1 - 2e^{-t}$

$$\begin{aligned}\bigl|\bigl\langle S_r(E)u, v\bigr\rangle - M\bigr| &\le C_\gamma \frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{t}{n}}\right)\sqrt{\frac{t}{n}} \\ &\le C_\gamma \frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{t}{n}} \vee \frac{t}{n}\right)\sqrt{\frac{t}{n}}\end{aligned} \tag{3.23}$$

with some constant $C_\gamma > 0$.

We will now prove a similar bound in the case when (3.5) does not hold. Then,

$$\frac{\|\Sigma\|_\infty}{\bar{g}_r}\sqrt{\frac{t}{n}} \ge \frac{\gamma}{4C}. \tag{3.24}$$

By bound (2.7),

$$\bigl|\bigl\langle S_r(E)u, v\bigr\rangle\bigr| \le \bigl\|S_r(E)\bigr\|_\infty \le \|\hat{P}_r - P_r\|_\infty + \bigl\|L_r(E)\bigr\|_\infty \le \|\hat{P}_r - P_r\|_\infty + 2\frac{\|E\|_\infty}{\bar{g}_r} \le 6\frac{\|E\|_\infty}{\bar{g}_r}.$$

We can now use the bounds of Theorems 1 and 2 combined with (3.24) to get that for all $t \ge 1$, with probability at least $1 - e^{-t}$ that

$$\bigl|\bigl\langle S_r(E)u, v\bigr\rangle\bigr| \lesssim \frac{\|\Sigma\|_\infty}{\bar{g}_r}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{t}{n}} \vee \frac{t}{n}\right).$$

Note that the above display also holds for all $t \geq \log 2$ (possibly, with different constants in relationships $\lesssim, \lesssim_\gamma$). For $t = \log 2$, $1 - e^{-t} = 1/2$ and it follows that

$$M \lesssim_\gamma \frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{1}{n}} \vee \frac{1}{n}\right).$$

Combining the last two displays, we get for all $t \geq 1$, with probability at least $1 - e^{-t}$

$$\left|\langle S_r(E)u, v\rangle - M\right| \lesssim_\gamma \frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{t}{n}} \vee \frac{t}{n}\right),$$

and using the condition (3.24), we easily extend (3.23) to all values of $t \geq 1$.

Integrating the tails of this exponential bound it is easy to see that, with some $D_\gamma > 0$,

$$\left|\mathbb{E}\langle S_r(E)u, v\rangle - M\right| \leq \mathbb{E}\left|\langle S_r(E)u, v\rangle - M\right| \leq D_\gamma \frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{1}{n}} \vee \frac{1}{n}\right)\sqrt{\frac{1}{n}},$$

which, in turn, implies that one can replace M by the expectation $\mathbb{E}\langle S_r(E)u, v\rangle$ in the concentration bound and get that with some $D_\gamma > 0$ and with probability at least $1 - 2e^{-t}$

$$\left|\langle S_r(E)u, v\rangle - \mathbb{E}\langle S_r(E)u, v\rangle\right| \leq D_\gamma \frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{t}{n}} \vee \frac{t}{n}\right)\sqrt{\frac{t}{n}}.$$

By adjusting the constant D_γ, we can replace $1 - 2e^{-t}$ by $1 - e^{-t}$.

This completes the proof of the theorem. □

4. A representation of the bias $\mathbb{E}\hat{P}_r - P_r$

In this section, we study the bias $W_r := \mathbb{E}\hat{P}_r - P_r$ of the empirical spectral projector $\hat{P}_r$. Under mild assumptions, we show that

$$\mathbb{E}\hat{P}_r - P_r = P_r W_r P_r + T_r,$$

where the main term $P_r W_r P_r$ is a symmetric operator of rank m_r such that

$$\|P_r W_r P_r\|_\infty \leq \|W_r\|_\infty \lesssim \frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\frac{\mathbf{r}(\Sigma)}{n} \tag{4.1}$$

and the remainder term T_r satisfies the bound $\|T_r\|_\infty = O(n^{-1/2})$. Moreover, in the case when $\mathbf{r}(\Sigma) = o(n)$, we have $\|T_r\|_\infty = o(n^{-1/2})$.

Theorem 4. *Suppose that, for some $\gamma \in (0, 1)$, (3.1) is satisfied. Then, there exists a constant $D_\gamma > 0$ such that*

$$\|\mathbb{E}\hat{P}_r - P_r - P_r W_r P_r\|_\infty \leq D_\gamma \frac{m_r\|\Sigma\|_\infty^2}{\bar{g}_r^2}\sqrt{\frac{\mathbf{r}(\Sigma)}{n}}\frac{1}{\sqrt{n}}. \tag{4.2}$$

Remark 3. *Note that the operator $P_r W_r P_r$ does satisfy condition (4.1). This follows from bound (2.10) and Theorem 1.*

Proof of Theorem 4. Note that, under assumption (3.1), Theorem 1 implies that $\mathbf{r}(\Sigma) \lesssim n$. It is enough to prove bound (4.2) in the case when $\mathbf{r}(\Sigma) > r_\gamma$ for an arbitrary $r_\gamma \geq 1$ depending only on γ. Indeed, by bound (2.10) and

Theorem 1,

$$\|\mathbb{E}\hat{P}_r - P_r - P_r W_r P_r\|_\infty \le \|\mathbb{E}\hat{P}_r - P_r\|_\infty + \big\|P_r(\mathbb{E}\hat{P}_r - P_r)P_r\big\|_\infty$$
$$\le 2\|\mathbb{E}\hat{P}_r - P_r\|_\infty = 2\big\|\mathbb{E}S_r(E)\big\|_\infty \lesssim \frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\frac{\mathbf{r}(\Sigma)}{n}.$$

If $\mathbf{r}(\Sigma) \le r_\gamma$, this yields

$$\|\mathbb{E}\hat{P}_r - P_r - P_r W_r P_r\|_\infty \lesssim_\gamma \frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\frac{1}{n} \lesssim \frac{m_r\|\Sigma\|_\infty^2}{\bar{g}_r^2}\sqrt{\frac{\mathbf{r}(\Sigma)}{n}}\frac{1}{\sqrt{n}},$$

implying (4.2). In what follows, we assume that $\mathbf{r}(\Sigma) \ge r_\gamma$ with r_γ to be chosen later on. Let

$$\delta_n := \mathbb{E}\|E\|_\infty + C\|\Sigma\|_\infty\sqrt{\frac{\log \mathbf{r}(\Sigma)}{n}}$$

with a constant $C > 0$ chosen so that

$$\mathbb{P}\big\{\|E\|_\infty \ge \delta_n\big\} \le \exp\big\{-\log \mathbf{r}(\Sigma)\big\} = \frac{1}{\mathbf{r}(\Sigma)} \tag{4.3}$$

(such a choice is possible due to Theorem 2). Assume first that

$$C\|\Sigma\|_\infty\sqrt{\frac{\log \mathbf{r}(\Sigma)}{n}} > \frac{\gamma}{2}\frac{\bar{g}_r}{2}. \tag{4.4}$$

In view of Theorem 1 and condition (3.1), we also have that, for some constant $C_1 > 0$,

$$C_1\|\Sigma\|_\infty\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \le \mathbb{E}\|\hat{\Sigma} - \Sigma\| \le (1-\gamma)\frac{\bar{g}_r}{2}.$$

Therefore,

$$C_1\|\Sigma\|_\infty\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \le \mathbb{E}\|\hat{\Sigma} - \Sigma\| \le C\frac{2(1-\gamma)}{\gamma}\|\Sigma\|_\infty\sqrt{\frac{\log \mathbf{r}(\Sigma)}{n}},$$

implying that

$$\frac{\mathbf{r}(\Sigma)}{\log \mathbf{r}(\Sigma)} \le \frac{C^2}{C_1^2}\frac{4(1-\gamma^2)}{\gamma^2} := c_\gamma.$$

It is easy to see that the set $\{r \ge 1 : \frac{r}{\log r} \le c_\gamma\}$ is either empty (then set $r_\gamma := 1$), or it is an interval $[r_\gamma^-, r_\gamma^+]$ for some $1 \le r_\gamma^- \le r_\gamma^+ < \infty$. In this case, set $r_\gamma = r_\gamma^+$. Thus, if $\mathbf{r}(\Sigma) > r_\gamma$, condition (4.4) does not hold. In what follows, assume that $\mathbf{r}(\Sigma) > r_\gamma$ and we have

$$C\|\Sigma\|_\infty\sqrt{\frac{\log \mathbf{r}(\Sigma)}{n}} \le \frac{\gamma}{2}\frac{\bar{g}_r}{2}. \tag{4.5}$$

This implies that $\delta_n \le (1-\gamma/2)\frac{\bar{g}_r}{2}$.

Consider the following representation

$$\begin{aligned}\mathbb{E}\hat{P}_r - P_r &= \mathbb{E}\big(L_r(E) + S_r(E)\big) = \mathbb{E}S_r(E)\\ &= \mathbb{E}P_r S_r(E) P_r + \mathbb{E}\big(P_r^\perp S_r(E)P_r + P_r S_r(E)P_r^\perp + P_r^\perp S_r(E)P_r^\perp\big)I\big(\|E\|_\infty \le \delta_n\big)\\ &\quad + \mathbb{E}\big(P_r^\perp S_r(E)P_r + P_r S_r(E)P_r^\perp + P_r^\perp S_r(E)P_r^\perp\big)I\big(\|E\|_\infty > \delta_n\big)\end{aligned} \tag{4.6}$$

and provide bounds for its relevant terms.

Recall formula (2.11) and note that, under the assumption $\|E\|_\infty < \frac{\bar{g}_r}{2}$, the series in the right-hand side converges in the operator norm absolutely and uniformly in $\eta \in \gamma_r$. Under this assumption,

$$S_r(E) = -\sum_{k\geq 2} \frac{1}{2\pi i} \oint_{\gamma_r} (-1)^k \bigl[R_\Sigma(\eta)E\bigr]^k R_\Sigma(\eta)\, d\eta. \tag{4.7}$$

Denote

$$\tilde{R}_\Sigma(\eta) := \sum_{s\notin\Delta_r} \frac{1}{\mu_s - \eta} P_s.$$

Then

$$R_\Sigma(\eta) = \frac{1}{\mu_r - \eta} P_r + \tilde{R}_\Sigma(\eta).$$

It is easy to check that

$$\begin{aligned} P_r^\perp \bigl[R_\Sigma(\eta)E\bigr]^k R_\Sigma(\eta) P_r &= P_r^\perp \frac{1}{\mu_r - \eta} \bigl[R_\Sigma(\eta)E\bigr]^k P_r \\ &= \frac{1}{(\mu_r - \eta)^2} \sum_{s=2}^k \bigl(\tilde{R}_\Sigma(\eta)E\bigr)^{s-1} P_r E \bigl(R_\Sigma(\eta)E\bigr)^{k-s} P_r \\ &\quad + \frac{1}{\mu_r - \eta} \bigl(\tilde{R}_\Sigma(\eta)E\bigr)^k P_r. \end{aligned}$$

To understand the last equality, note that, in each bracket of the expression

$$\bigl[R_\Sigma(\eta)E\bigr]^k = \bigl[R_\Sigma(\eta)E\bigr] \cdots \bigl[R_\Sigma(\eta)E\bigr],$$

$R_\Sigma(\eta)$ can be replaced by the sum of two terms, $\frac{1}{\mu_r-\eta} P_r$ and $\tilde{R}_\Sigma(\eta)$. Index s in the sum is the number of the first bracket where $\frac{1}{\mu_r-\eta} P_r$ is chosen. If $s = 1$, the corresponding term is equal to 0 since $P_r^\perp P_r = 0$. The last term corresponds to the case when $\tilde{R}_\Sigma(\eta)$ is chosen from each of the brackets.

We can now write

$$\begin{aligned} P_r^\perp S_r(E) P_r = -\sum_{k\geq 2} (-1)^k \frac{1}{2\pi i} \oint_{\gamma_r} \Biggl[&\frac{1}{(\mu_r - \eta)^2} \sum_{s=2}^k \bigl(\tilde{R}_\Sigma(\eta)E\bigr)^{s-1} P_r E \bigl(R_\Sigma(\eta)E\bigr)^{k-s} P_r \\ &+ \frac{1}{\mu_r - \eta} \bigl(\tilde{R}_\Sigma(\eta)E\bigr)^k P_r \Biggr] d\eta. \end{aligned} \tag{4.8}$$

Since $P_r = \sum_{l\in\Delta_r} (\theta_l \otimes \theta_l)$, where $\{\theta_l : l \in \Delta_r\}$ is an arbitrary orthonormal basis of the eigenspace corresponding to the eigenvalue μ_r, we get that, for all $v \in \mathbb{H}$,

$$\bigl(\tilde{R}_\Sigma(\eta)E\bigr)^{s-1} P_r E \bigl(R_\Sigma(\eta)E\bigr)^{k-s} P_r v = \sum_{l\in\Delta_r} \bigl(\tilde{R}_\Sigma(\eta)E\bigr)^{s-1} (\theta_l \otimes \theta_l) E \bigl(R_\Sigma(\eta)E\bigr)^{k-s} P_r v \tag{4.9}$$

$$= \sum_{l\in\Delta_r} \bigl\langle E\bigl(R_\Sigma(\eta)E\bigr)^{k-s} P_r v, \theta_l \bigr\rangle \bigl(\tilde{R}_\Sigma(\eta)E\bigr)^{s-2} \tilde{R}_\Sigma(\eta) E \theta_l. \tag{4.10}$$

Clearly,

$$\bigl|\bigl\langle E\bigl(R_\Sigma(\eta)E\bigr)^{k-s} P_r v, \theta_l \bigr\rangle\bigr| \leq \bigl\|R_\Sigma(\eta)\bigr\|_\infty^{k-s} \|E\|_\infty^{k-s+1} \|v\|,$$

which implies that

$$\mathbb{E}\big|\big\langle E\big(R_\Sigma(\eta)E\big)^{k-s}P_r v,\theta_l\big\rangle\big|^2 I\big(\|E\|_\infty\le\delta_n\big)\le\bigg(\frac{2}{\bar g_r}\bigg)^{2(k-s)}\delta_n^{2(k-s+1)}\|v\|^2. \tag{4.11}$$

We also have

$$\begin{aligned}\big(\tilde R_\Sigma(\eta)E\big)^{s-2}\tilde R_\Sigma(\eta)E\theta_l &= \big(\tilde R_\Sigma(\eta)E\big)^{s-2}\tilde R_\Sigma(\eta)(\hat\Sigma-\Sigma)\theta_l\\ &= \big(\tilde R_\Sigma(\eta)E\big)^{s-2}\tilde R_\Sigma(\eta)\hat\Sigma\theta_l = n^{-1}\sum_{j=1}^n\langle X_j,\theta_l\rangle\big(\tilde R_\Sigma(\eta)E\big)^{s-2}\tilde R_\Sigma(\eta)X_j,\end{aligned}$$

where we used the fact that

$$\tilde R_\Sigma(\eta)\Sigma\theta_l=\mu_r\tilde R_\Sigma(\eta)\theta_l=0.$$

It is easy to check that the random variables $(\tilde R_\Sigma(\eta)E)^{s-2}\tilde R_\Sigma(\eta)X_j$, $j=1,\dots,n$ are functions of random variables $P_sX_j : s\ne r$, $j=1,\dots,n$ that are independent of $\langle X_j,\theta_l\rangle$, $l\in\Delta_r$, $j=1,\dots,n$ (recall that X_j, $j=1,\dots,n$ are i.i.d. Gaussian, and P_rX_j, $j=1,\dots,n$ and $P_sX_j : s\ne r$, $j=1,\dots,n$ are uncorrelated and, hence, independent). Given $u\in\mathbb{H}$, denote

$$\zeta_j(u)=\big\langle\big(\tilde R_\Sigma(\eta)E\big)^{s-2}\tilde R_\Sigma(\eta)X_j,u\big\rangle,\quad j=1,\dots,n$$

(which are complex valued random variables). Write $\zeta_j(u)=\zeta_j^{(1)}(u)+i\zeta_j^{(2)}(u)$, where $\zeta_j^{(1)}(u),\zeta_j^{(2)}(u)$ are real valued. Denote also

$$\alpha(u):=\alpha^{(1)}(u)+i\alpha^{(2)}(u):=\big\langle\big(\tilde R_\Sigma(\eta)E\big)^{s-2}\tilde R_\Sigma(\eta)E\theta_l,u\big\rangle.$$

Then, conditionally on $P_sX_j : s\ne r$, $j=1,\dots,n$, the random vector $(\alpha^{(1)}(u),\alpha^{(2)}(u))$ has the same distribution as mean zero Gaussian random vector in $\mathbb{R}^2$ with covariance

$$\frac{\mu_r}{n}\Bigg(n^{-1}\sum_{j=1}^n\zeta_j^{(k_1)}(u)\zeta_j^{(k_2)}(u)\Bigg),\quad k_1,k_2=1,2.$$

Note that

$$\begin{aligned}n^{-1}\sum_{j=1}^n\big|\zeta_j(u)\big|^2 &= n^{-1}\sum_{j=1}^n\big|\big\langle\big(\tilde R_\Sigma(\eta)E\big)^{s-2}\tilde R_\Sigma(\eta)X_j,u\big\rangle\big|^2\\ &= \big\langle\hat\Sigma\big(\tilde R_\Sigma(\eta)E\big)^{s-2}\tilde R_\Sigma(\eta)u,\big(\tilde R_\Sigma(\eta)E\big)^{s-2}\tilde R_\Sigma(\eta)u\big\rangle\le\|\hat\Sigma\|_\infty\big\|\tilde R_\Sigma(\eta)\big\|_\infty^{2(s-1)}\|E\|_\infty^{2(s-2)}\|u\|^2\\ &\le\big(\|\Sigma\|_\infty\big\|\tilde R_\Sigma(\eta)\big\|_\infty^{2(s-1)}\|E\|_\infty^{2(s-2)}+\big\|\tilde R_\Sigma(\eta)\big\|_\infty^{2(s-1)}\|E\|_\infty^{2s-3}\big)\|u\|^2.\end{aligned}$$

Under the assumption $\delta_n<\frac{\bar g_r}{2}$, the following inclusion holds:

$$\big\{\|E\|_\infty\le\delta_n\big\}\subset\Bigg\{n^{-1}\sum_{j=1}^n\big|\zeta_j(u)\big|^2\le 2\|\Sigma\|_\infty\bigg(\frac{2}{\bar g_r}\bigg)^{2(s-1)}\delta_n^{2(s-2)}\|u\|^2\Bigg\}=:G.$$

Therefore, we have

$$\begin{aligned}&\mathbb{E}\big|\big\langle\big(\tilde R_\Sigma(\eta)E\big)^{s-2}\tilde R_\Sigma(\eta)E\theta_l,u\big\rangle\big|^2 I\big(\|E\|_\infty\le\delta_n\big)\\ &\quad\le\mathbb{E}\big|\big\langle\big(\tilde R_\Sigma(\eta)E\big)^{s-2}\tilde R_\Sigma(\eta)E\theta_l,u\big\rangle\big|^2 I_G\end{aligned}$$

$$= \mathbb{E}\mathbb{E}\big(\big|\big\langle\big(\tilde{R}_\Sigma(\eta)E\big)^{s-2}\tilde{R}_\Sigma(\eta)E\theta_l, u\big\rangle\big|^2 I_G \mid P_s X_j, s \neq r, j = 1, \dots, n\big)$$

$$= \frac{\mu_r}{n}\mathbb{E}\mathbb{E}\left(n^{-1}\sum_{j=1}^n \big|\zeta_j(u)\big|^2 I_G \;\middle|\; P_s X_j, s \neq r, j = 1, \dots, n\right)$$

$$= \frac{\mu_r}{n}\mathbb{E}n^{-1}\sum_{j=1}^n \big|\zeta_j(u)\big|^2 I_G \le 2\|\Sigma\|_\infty \frac{\mu_r}{n}\left(\frac{2}{\bar{g}_r}\right)^{2(s-1)} \delta_n^{2(s-2)}\|u\|^2. \tag{4.12}$$

By (4.11) and (4.12),

$$\begin{aligned}
&\big|\mathbb{E}\big\langle E\big(R_\Sigma(\eta)E\big)^{k-s}P_r v, \theta_l\big\rangle\big\langle\big(\tilde{R}_\Sigma(\eta)E\big)^{s-2}\tilde{R}_\Sigma(\eta)E\theta_l, u\big\rangle I\big(\|E\|_\infty \le \delta_n\big)\big| \\
&\quad \le \big(\mathbb{E}\big|\big\langle E\big(R_\Sigma(\eta)E\big)^{k-s}P_r v, \theta_l\big\rangle\big|^2 I\big(\|E\|_\infty \le \delta_n\big)\big)^{1/2}\big(\mathbb{E}\big|\big\langle\big(\tilde{R}_\Sigma(\eta)E\big)^{s-2}\tilde{R}_\Sigma(\eta)E\theta_l, u\big\rangle\big|^2 I\big(\|E\|_\infty \le \delta_n\big)\big)^{1/2} \\
&\quad \le \sqrt{2}\frac{\|\Sigma\|_\infty}{\sqrt{n}}\left(\frac{2\delta_n}{\bar{g}_r}\right)^{k-1}\|u\|\|v\|
\end{aligned} \tag{4.13}$$

and it follows from (4.10) and (4.13) that

$$\begin{aligned}
&\big|\mathbb{E}\big\langle\big(\tilde{R}_\Sigma(\eta)E\big)^{s-1}P_r E\big(R_\Sigma(\eta)E\big)^{k-s}P_r v, u\big\rangle I\big(\|E\|_\infty \le \delta_n\big)\big| \\
&\quad \le \sum_{l\in\Delta_r}\big|\mathbb{E}\big\langle E\big(R_\Sigma(\eta)E\big)^{k-s}P_r v, \theta_l\big\rangle\big\langle\big(\tilde{R}_\Sigma(\eta)E\big)^{s-2}\tilde{R}_\Sigma(\eta)E\theta_l, u\big\rangle I\big(\|E\|_\infty \le \delta_n\big)\big| \\
&\quad \le \sqrt{2}m_r\frac{\|\Sigma\|_\infty}{\sqrt{n}}\left(\frac{2\delta_n}{\bar{g}_r}\right)^{k-1}\|u\|\|v\|.
\end{aligned} \tag{4.14}$$

Similarly, we also have

$$\big|\mathbb{E}\big\langle\big(\tilde{R}_\Sigma(\eta)E\big)^k P_r v, u\big\rangle\big| \le \sqrt{2}m_r\frac{\|\Sigma\|_\infty}{\sqrt{n}}\frac{2}{\bar{g}_r}\left(\frac{2\delta_n}{\bar{g}_r}\right)^{k-1}\|u\|\|v\|. \tag{4.15}$$

Now use (4.8), (4.14) and (4.15) to get (under assumption that $\delta_n \le (1-\gamma/2)\frac{\bar{g}_r}{2}$)

$$\begin{aligned}
&\big|\mathbb{E}\big\langle P_r^\perp S_r(E)P_r v, u\big\rangle I\big(\|E\|_\infty \le \delta_n\big)\big| \\
&\quad \le \sum_{k\ge 2}\frac{1}{2\pi}\oint_{\gamma_r}\Bigg[\frac{1}{|\mu_r-\eta|^2}\sum_{s=2}^k\big|\mathbb{E}\big\langle\big(\tilde{R}_\Sigma(\eta)E\big)^{s-1}P_r E\big(R_\Sigma(\eta)E\big)^{k-s}P_r v, u\big\rangle I\big(\|E\|_\infty \le \delta_n\big)\big| \\
&\qquad + \frac{1}{|\mu_r-\eta|}\big|\mathbb{E}\big\langle\big(\tilde{R}_\Sigma(\eta)E\big)^k P_r v, u\big\rangle I\big(\|E\|_\infty \le \delta_n\big)\big|\Bigg]d\eta \\
&\quad \le \sum_{k\ge 2}\frac{1}{2\pi}2\pi\frac{\bar{g}_r}{2}\left(\frac{2}{\bar{g}_r}\right)^2\sqrt{2}m_r\frac{\|\Sigma\|_\infty}{\sqrt{n}}k\left(\frac{2\delta_n}{\bar{g}_r}\right)^{k-1}\|u\|\|v\| \\
&\quad = \sqrt{2}m_r\frac{\|\Sigma\|_\infty}{\sqrt{n}}\frac{2}{\bar{g}_r}\sum_{k\ge 2}k\left(\frac{2\delta_n}{\bar{g}_r}\right)^{k-1}\|u\|\|v\| \\
&\quad = \sqrt{2}m_r\frac{\|\Sigma\|_\infty}{\sqrt{n}}\frac{2}{\bar{g}_r}\left(\left(1-\frac{2\delta_n}{\bar{g}_r}\right)^{-2}-1\right)\|u\|\|v\| \le \frac{32\sqrt{2}}{\gamma^2}m_r\frac{\|\Sigma\|_\infty}{\sqrt{n}}\frac{\delta_n}{\bar{g}_r^2}\|u\|\|v\|.
\end{aligned}$$

Therefore,

$$\big\|\mathbb{E}P_r^{\perp}S_r(E)P_rI\big(\|E\|_\infty\le\delta_n\big)\big\|_\infty\le\frac{32\sqrt{2}}{\gamma^2}m_r\frac{\|\Sigma\|_\infty}{\sqrt{n}}\frac{\delta_n}{\bar{g}_r^2}. \tag{4.16}$$

Obviously, the same bound holds for $\|\mathbb{E}P_rS_r(E)P_r^{\perp}I(\|E\|_\infty\le\delta_n)\|_\infty$. Moreover, similarly, it can be proved that

$$\big\|\mathbb{E}P_r^{\perp}S_r(E)P_r^{\perp}I\big(\|E\|_\infty\le\delta_n\big)\big\|_\infty\le c_\gamma m_r\frac{\|\Sigma\|_\infty}{\sqrt{n}}\frac{\delta_n}{\bar{g}_r^2} \tag{4.17}$$

with some constant $c_\gamma>0$.

To complete the proof, note that

$$\begin{aligned}&\big\|\mathbb{E}\big(P_r^{\perp}S_r(E)P_r+P_rS_r(E)P_r^{\perp}+P_r^{\perp}S_r(E)P_r^{\perp}\big)I\big(\|E\|_\infty>\delta_n\big)\big\|_\infty\\&\quad\le\mathbb{E}\big\|P_r^{\perp}S_r(E)P_r+P_rS_r(E)P_r^{\perp}+P_r^{\perp}S_r(E)P_r^{\perp}\big\|_\infty I\big(\|E\|_\infty>\delta_n\big)\\&\quad\le\mathbb{E}\big\|S_r(E)\big\|_\infty I\big(\|E\|_\infty>\delta_n\big).\end{aligned}$$

Next, using bound (2.10), Theorem 1 and bound (4.3), we get

$$\begin{aligned}\mathbb{E}\big\|S_r(E)\big\|_\infty I\big(\|E\|_\infty>\delta_n\big)&\le14\frac{\mathbb{E}\|E\|_\infty^2I(\|E\|_\infty>\delta_n)}{\bar{g}_r^2}\le14\frac{\mathbb{E}^{1/2}\|E\|_\infty^4\mathbb{P}^{1/2}\{\|E\|_\infty>\delta_n\}}{\bar{g}_r^2}\\&\lesssim\frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\frac{\mathbf{r}(\Sigma)}{n}\left(\frac{1}{\mathbf{r}(\Sigma)}\right)^{1/2}=\frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\frac{\sqrt{\mathbf{r}(\Sigma)}}{n}=\frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\sqrt{\frac{\mathbf{r}(\Sigma)}{n}}\frac{1}{\sqrt{n}}.\end{aligned}$$

Consequently, we get that

$$\big\|\mathbb{E}\big(P_r^{\perp}S_r(E)P_r+P_rS_r(E)P_r^{\perp}+P_r^{\perp}S_r(E)P_r^{\perp}\big)I\big(\|E\|_\infty>\delta_n\big)\big\|_\infty\lesssim\frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\sqrt{\frac{\mathbf{r}(\Sigma)}{n}}\frac{1}{\sqrt{n}}. \tag{4.18}$$

Bound (4.2) now follows from representation (4.6), bounds (4.16), (4.17) and (4.18). □

5. Asymptotics of bilinear forms of empirical spectral projectors

In this section, we study the asymptotic behavior of the bilinear forms

$$\big\langle(\hat{P}_r-\mathbb{E}\hat{P}_r)u,v\big\rangle,\quad u,v\in\mathbb{H}$$

in the case when the sample size n and the effective rank $\mathbf{r}(\Sigma)$ are both large. To describe this precisely, one has to deal with a sequence of problems in which the data is sampled from Gaussian distributions in $\mathbb{H}$ with mean zero and covariance $\Sigma=\Sigma^{(n)}$. This leads to the following asymptotic framework. 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 $\Sigma^{(n)}$, $\mu_r^{(n)}, r\ge1$ be distinct non-zero eigenvalues of $\Sigma^{(n)}$ arranged in decreasing order and $P_r^{(n)}, r\ge1$ 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 $\{\sigma_j(\hat{\Sigma}_n), j\in\Delta_r^{(n)}\}$.

The next assumption means that, for large enough n, there exists a unique eigenvalue $\mu^{(n)}$ of $\Sigma^{(n)}$ isolated inside a fixed interval from the rest of the spectrum of $\Sigma^{(n)}$.

Assumption 2. *There exists an interval $(\alpha,\beta)\subset\mathbb{R}_+$ and a number $\delta>0$ such that, for all large enough n, the set $\sigma(\Sigma^{(n)})\cap(\alpha,\beta)$ consists of a single eigenvalue $\mu^{(n)}=\mu_{r_n}^{(n)}$ of $\Sigma^{(n)}$ and*

$$\sigma\big(\Sigma^{(n)}\big)\setminus\big\{\mu^{(n)}\big\}\subset\mathbb{R}_+\setminus(\alpha-\delta,\beta+\delta).$$

Denote by $P^{(n)}$ the spectral projector corresponding to the eigenvalue $\mu^{(n)}$ and define the following sequence of operators:

$$C^{(n)} := \sum_{\mu_s^{(n)} \neq \mu^{(n)}} \frac{1}{\mu^{(n)} - \mu_s^{(n)}} P_s^{(n)}.$$

Consider the spectral measures associated with the covariance operators $\Sigma^{(n)}$:

$$\Lambda_{u,v}^{(n)}(A) := \sum_{r=1}^{\infty} \langle P_r^{(n)} u, v \rangle I_A\big(\mu_r^{(n)}\big), \quad u, v \in \mathbb{H}, A \in \mathcal{B}(\mathbb{R}_+),$$

where $\mathcal{B}(\mathbb{R}_+)$ denotes the Borel σ-algebra in $\mathbb{R}_+$.

Assumption 3. *For all $u, v \in \mathbb{H}$, the sequence of measures $\Lambda_{u,v}^{(n)}$ converges weakly to a measure $\Lambda_{u,v}$ in $\mathbb{R}_+$. Also assume that there exists $u \in \mathbb{H}$ such that $\Lambda_{u,u}([\alpha, \beta]) > 0$.*

Denote

$$\Gamma_1(u, v) := \int_{\alpha}^{\beta} \lambda \Lambda_{u,v}(d\lambda), \qquad \Gamma_2(u, v) := \int_{\mathbb{R}_+ \setminus [\alpha,\beta]} \frac{\lambda}{(\mu - \lambda)^2} \Lambda_{u,v}(d\lambda).$$

It will be shown in the proof of Theorem 5 that the limit covariance function Γ of normalized bilinear forms $n^{1/2}\langle(\hat{P}^{(n)} - \mathbb{E}\hat{P}^{(n)})u, v\rangle$ of empirical spectral projectors $\hat{P}^{(n)} := \hat{P}_{r_n}^{(n)}$ can be expressed in terms of functions Γ_1 and Γ_2 (see formula (5.3)). A step in this direction is the following lemma that provides the limits of bilinear forms $\langle P^{(n)}u, v\rangle$, $\langle P^{(n)}\Sigma^{(n)}P^{(n)}u, v\rangle$ and $\langle C^{(n)}\Sigma^{(n)}C^{(n)}u, v\rangle$ under Assumptions 2 and 3.

Lemma 6. *Under Assumptions 2 and 3, the following statements hold.*

(i) *There exists $\mu \in [\alpha, \beta]$ such that*

$$\mu^{(n)} \to \mu \quad \textit{as } n \to \infty.$$

(ii) *For all $u, v \in \mathbb{H}$,*

$$\big\langle P^{(n)}u, v\big\rangle \to \Lambda_{u,v}\big([\alpha, \beta]\big) \quad \textit{as } n \to \infty.$$

(iii) *For all $u, v \in \mathbb{H}$,*

$$\big\langle P^{(n)}\Sigma^{(n)}P^{(n)}u, v\big\rangle \to \Gamma_1(u, v) \quad \textit{as } n \to \infty.$$

(iv) *For all $u, v \in \mathbb{H}$,*

$$\big\langle C^{(n)}\Sigma^{(n)}C^{(n)}u, v\big\rangle \to \Gamma_2(u, v) \quad \textit{as } n \to \infty.$$

Proof. We start with proving (ii). In view of Assumption 2, for all $\delta' < \delta$,

$$\Lambda_{u,v}^{(n)}\big((\alpha - \delta', \beta + \delta')\big) = \Lambda_{u,v}^{(n)}\big(\big\{\mu^{(n)}\big\}\big) = \big\langle P^{(n)}u, v\big\rangle.$$

We can choose δ' such that $\alpha - \delta'$ and $\beta + \delta'$ are not atoms of $\Lambda_{u,v}$. Therefore, by Assumption 3,

$$\big\langle P^{(n)}u, v\big\rangle = \Lambda_{u,v}^{(n)}\big((\alpha - \delta', \beta + \delta')\big) \to \Lambda_{u,v}\big((\alpha - \delta', \beta + \delta')\big) \quad \text{as } n \to \infty$$

for all such δ'. Note that the limit does not depend on δ'. It is enough now to let $\delta' \to 0$ to get (ii).

To prove (iii), note that, for the same δ' as in the previous step,

$$\langle P^{(n)}\Sigma^{(n)}P^{(n)}u, v\rangle = \int_{\alpha-\delta'}^{\beta+\delta'} \lambda\Lambda_{u,v}^{(n)}(d\lambda) \to \int_{\alpha-\delta'}^{\beta+\delta'} \lambda\Lambda_{u,v}(d\lambda),$$

and, again, it is enough to let $\delta' \to 0$.

To prove (i), take $v = u \in \mathbb{H}$ such that $\Lambda_{u,u}([\alpha, \beta]) > 0$. By (iii), we have

$$\mu^{(n)}\langle P^{(n)}u, u\rangle = \langle P^{(n)}\Sigma^{(n)}P^{(n)}u, u\rangle \to \int_{\alpha}^{\beta} \lambda\Lambda_{u,v}(d\lambda)$$

and, by (ii),

$$\langle P^{(n)}u, u\rangle \to \Lambda_{u,u}\bigl([\alpha, \beta]\bigr) > 0.$$

This implies that

$$\mu^{(n)} \to \mu := \frac{\int_\alpha^\beta \lambda\Lambda_{u,u}(d\lambda)}{\Lambda_{u,u}([\alpha, \beta])}$$

that clearly belongs to $[\alpha, \beta]$ (and does not depend on the choice of u).

Finally, we prove (iv). To this end, note that, for all $\delta' < \delta$,

$$\langle C^{(n)}\Sigma^{(n)}C^{(n)}u, v\rangle = \int_{\mathbb{R}_+\setminus(\alpha-\delta',\beta+\delta')} \frac{\lambda}{(\mu^{(n)} - \lambda)^2}\Lambda_{u,v}^{(n)}(d\lambda).$$

Due to bilinearity, it will be enough to consider the case when $v = u$. Let $\delta' < \delta$ and suppose that $\alpha - \delta'$, $\beta + \delta'$ are not atoms of $\Lambda_{u,u}$. Since $\mu^{(n)} \to \mu$ and Assumption 2 holds,

$$\frac{\lambda}{(\mu^{(n)} - \lambda)^2} \to \frac{\lambda}{(\mu - \lambda)^2} \quad \text{as } n \to \infty$$

uniformly in $\mathbb{R}_+ \setminus (\alpha - \delta', \beta + \delta')$. Due to the weak convergence of $\Lambda_{u,u}^{(n)}$ to $\Lambda_{u,u}$, it is easy to show that

$$\begin{aligned}\langle C^{(n)}\Sigma^{(n)}C^{(n)}u, u\rangle &= \int_{\mathbb{R}_+\setminus(\alpha-\delta',\beta+\delta')} \frac{\lambda}{(\mu^{(n)} - \lambda)^2}\Lambda_{u,u}^{(n)}(d\lambda) \\ &\to \int_{\mathbb{R}_+\setminus(\alpha-\delta',\beta+\delta')} \frac{\lambda}{(\mu - \lambda)^2}\Lambda_{u,u}(d\lambda),\end{aligned}$$

and it remains to let $\delta' \to 0$. □

It turns out that the following Assumption 4, which is somewhat easier to understand, implies Assumption 3 and even its stronger version (stated in Proposition 1 below).

Assumption 4. *Suppose the sequence of covariance operators $\Sigma^{(n)}$ with $\sup_{n\geq 1}\|\Sigma^{(n)}\|_\infty < +\infty$ converges strongly to a bounded symmetric nonnegatively definite operator $\Sigma : \mathbb{H} \mapsto \mathbb{H}$ (that is, $\Sigma^{(n)}u \to \Sigma u$ as $n \to \infty$ for all $u \in \mathbb{H}$). Let $\mathcal{E}(\cdot)$ be the resolution of the identity associated with Σ.[3] Suppose also that there exists $u \in \mathbb{H}$ such that $\langle\mathcal{E}([\alpha, \beta])u, u\rangle > 0$.*

[3]This means that $\mathcal{E}(\cdot)$ is a projector valued measure on Borel subsets of $\mathbb{R}_+$, such that $\mathcal{E}(\Delta)\mathcal{E}(\Delta') = \mathcal{E}(\Delta \cap \Delta')$, $\mathcal{E}(\mathbb{R}_+) = I$ and $\Sigma = \int_{\mathbb{R}_+} \lambda\mathcal{E}(d\lambda)$.

The next proposition shows that the limit spectral measure $\Lambda_{u,v}$ can be defined in terms of the limit resolution of the identity $\mathcal{E}$:

$$\Lambda_{u,v}(\Delta) = \langle \mathcal{E}(\Delta)u, v\rangle, \quad \Delta \in \mathcal{B}(\mathbb{R}_+), u, v \in \mathbb{H}.$$

Proposition 1. *Assumption 4 implies Assumption 3. Moreover, it implies that, for all $u, v \in \mathbb{H}$ and for all sequences $u_n \to u$, $v_n \to v$ as $n \to \infty$, the sequence of measures $\Lambda^{(n)}_{u_n,v_n}$ converges weakly to $\Lambda_{u,v}$.*

Proof. Indeed, let $\mathcal{E}^{(n)}(\cdot)$ be the resolution of the identity associated with $\Sigma^{(n)}$. Then $\Lambda^{(n)}_{u,v}(\cdot) = \langle \mathcal{E}^{(n)}(\cdot)u, v\rangle$. It is well known (see, e.g., [28], Ch. IX, Section 134) that the uniform boundedness of $\|\Sigma^{(n)}\|_\infty$ and strong convergence of operators $\Sigma^{(n)}$ to Σ implies strong convergence of $\mathcal{E}^{(n)}([0,\lambda])$ to $\mathcal{E}([0,\lambda])$ for all λ that do not belong to the point spectrum of Σ, which easily implies the weak convergence of measures $\Lambda^{(n)}_{u_n,v_n}$ to $\Lambda_{u,v}$. □

We will also need the following simple proposition (its proof is elementary).

Proposition 2. *Suppose Assumptions 2 and 4 hold. Suppose also that $\mu^{(n)}$ is an eigenvalue of multiplicity 1. Then, the corresponding spectral projector $P^{(n)} = \theta^{(n)} \otimes \theta^{(n)}$, where $\theta^{(n)}$ is the eigenvector corresponding to $\mu^{(n)}$ and, for some $\theta \in \mathbb{H}$, $\theta^{(n)} \to \theta$ as $n \to \infty$.*

As a typical example where Assumption 4 holds, consider the case of $\Sigma^{(n)} = P_{L_n} \Sigma P_{L_n}$ for a sequence of subspaces $L_n \subset \mathbb{H}$ with $\dim(L_n) \to \infty$ and $\bigcup_{n\geq 1} L_n$ being dense in $\mathbb{H}$ (see also the discussion of general spiked covariance models in Section 1).

We will now state the main result of this section.

Theorem 5. *Suppose that*

$$\sup_{n\geq 1} \|\Sigma^{(n)}\|_\infty < \infty \tag{5.1}$$

and

$$\mathbf{r}(\Sigma^{(n)}) = o(n) \quad \text{as } n \to \infty. \tag{5.2}$$

Also, suppose that Assumptions 2 and 3 hold. Let $\hat{P}^{(n)} := \hat{P}^{(n)}_{r_n}$. Then, the finite dimensional distributions of stochastic processes

$$n^{1/2}\langle (\hat{P}^{(n)} - \mathbb{E}\hat{P}^{(n)})u, v\rangle, \quad u, v \in \mathbb{H}$$

converge weakly as $n \to \infty$ to the finite dimensional distributions of the centered Gaussian process $Y(u,v)$, $u, v \in \mathbb{H}$ with covariance function Γ defined as follows:

$$\Gamma(u, v; u', v') := \Gamma_1(v, v')\Gamma_2(u, u') + \Gamma_1(v, u')\Gamma_2(u, v') + \Gamma_1(u, u')\Gamma_2(v, v') + \Gamma_1(u, v')\Gamma_2(v, u'). \tag{5.3}$$

If, in addition, Assumption 4 holds, then, for all $\varphi_n, \psi_n : \mathbb{H} \mapsto \mathbb{H}$ such that $\varphi_n(u) \to u$, $\psi_n(u) \to u$ as $n \to \infty$ for all $u \in \mathbb{H}$, the finite dimensional distributions of stochastic processes

$$n^{1/2}\langle (\hat{P}^{(n)} - \mathbb{E}\hat{P}^{(n)})\varphi_n(u), \psi_n(v)\rangle, \quad u, v \in \mathbb{H}$$

converge weakly as $n \to \infty$ to the same limit.

Proof. We prove only the first claim. The modifications needed to establish the second claim are rather obvious. The proof is based on the following representation of $\hat{P}^{(n)} - P^{(n)}$:

$$\hat{P}^{(n)} - \mathbb{E}\hat{P}^{(n)} = L^{(n)}(E^{(n)}) + R^{(n)}, \tag{5.4}$$

where

$$L^{(n)}\big(E^{(n)}\big) = P^{(n)}E^{(n)}C^{(n)} + C^{(n)}E^{(n)}P^{(n)}, \qquad E^{(n)} := \hat{\Sigma}_n - \Sigma^{(n)}$$

and where the remainder $R^{(n)}$ will be controlled using Theorem 3.

In addition to this, to show the asymptotic normality of $\langle L^{(n)}(E^{(n)})u, v\rangle$, we rely on Lemma 6 and also on Lemma 7 below (both lemmas are based on Assumptions 2 and 3). Observe that

$$n^{1/2}\big\langle L^{(n)}\big(E^{(n)}\big)u, v\big\rangle = n^{-1/2}\sum_{j=1}^{n}\big(\xi_j^{(n)}(u,v) + \xi_j^{(n)}(v,u)\big), \tag{5.5}$$

where $\xi_j^{(n)}(u,v) := \langle X_j^{(n)}, P^{(n)}v\rangle\langle X_j^{(n)}, C^{(n)}u\rangle$ are independent copies of random variable $\xi^{(n)}(u,v) := \langle X^{(n)}, P^{(n)}v\rangle\langle X^{(n)}C^{(n)}u\rangle$. Recall also that Gaussian random variables $\langle X^{(n)}, P^{(n)}v\rangle$, $\langle X^{(n)}, C^{(n)}u\rangle$ are uncorrelated and, hence, independent. Therefore, $\xi^{(n)}(u,v)$ is mean zero and, by Lemma 6, for all $u, v, u', v' \in \mathbb{H}$,

$$\mathbb{E}\xi^{(n)}(u,v)\xi^{(n)}\big(u',v'\big) = \big\langle P^{(n)}\Sigma^{(n)}P^{n}v, v'\big\rangle\big\langle C^{(n)}\Sigma^{(n)}C^{n}u, u'\big\rangle \quad \to \quad \bar{\Gamma}\big(u,v;u',v'\big) := \Gamma_1\big(v,v'\big)\Gamma_2\big(u,u'\big),$$

which implies

$$\mathbb{E}\big(\xi^{(n)}(u,v) + \xi^{(n)}(v,u)\big)\big(\xi^{(n)}\big(u',v'\big) + \xi^{(n)}\big(v',u'\big)\big) \to \Gamma\big(u,v;u'v'\big).$$

Lemma 7. *Under Assumptions 2 and 3, the sequence of finite dimensional distributions of*

$$n^{1/2}\big\langle L^{(n)}\big(E^{(n)}\big)u, v\big\rangle, \quad u, v \in \mathbb{H}$$

converges weakly as $n \to \infty$ to the finite dimensional distributions of the centered Gaussian process $Y(u,v), u, v \in \mathbb{H}$ with covariance function Γ.

Proof. In view of (5.5), it is enough to show the convergence of finite dimensional distributions of the process $n^{-1/2}\sum_{j=1}^{n}\xi_j^{(n)}(u,v), u, v \in \mathbb{H}$ to the finite dimensional distributions of the centered Gaussian process $\bar{Y}(u,v), u, v \in \mathbb{H}$ with covariance function $\bar{\Gamma}$. To this end, one has to check the Lindeberg condition, which reduces to

$$\frac{\mathbb{E}|\xi^{(n)}(u,v)|^2 I(|\xi^{(n)}(u,v)| \geq \tau\sqrt{n}\mathbb{E}^{1/2}|\xi^{(n)}(u,v)|^2)}{\mathbb{E}|\xi^{(n)}(u,v)|^2} \to 0 \quad \text{as } n \to \infty$$

for all $\tau > 0$. Note that

$$\frac{\mathbb{E}|\xi^{(n)}(u,v)|^2 I(|\xi^{(n)}(u,v)| \geq \tau\sqrt{n}\mathbb{E}^{1/2}|\xi^{(n)}(u,v)|^2)}{\mathbb{E}|\xi^{(n)}(u,v)|^2} \leq \frac{\mathbb{E}|\xi^{(n)}(u,v)|^4}{\tau^2 n(\mathbb{E}|\xi^{(n)}(u,v)|^2)^2}.$$

Since

$$\mathbb{E}\big|\xi^{(n)}(u,v)\big|^2 = \big\langle P^{(n)}\Sigma^{(n)}P^{(n)}v, v\big\rangle\big\langle C^{(n)}\Sigma^{(n)}C^{(n)}u, u\big\rangle$$

and

$$\mathbb{E}\big|\xi^{(n)}(u,v)\big|^4 = \mathbb{E}\big\langle X^{(n)}, P^{(n)}v\big\rangle^4\mathbb{E}\big\langle X^{(n)}, C^{(n)}u\big\rangle^4 = 9\big\langle P^{(n)}\Sigma^{(n)}P^{(n)}v, v\big\rangle^2\big\langle C^{(n)}\Sigma^{(n)}C^{(n)}u, u\big\rangle^2$$

(where we used the fact that, for a centered normal random variable g, $\mathbb{E}g^4 = 3(\mathbb{E}g^2)^2$), we get

$$\limsup_{n\to\infty}\frac{\mathbb{E}|\xi^{(n)}(u,v)|^4}{\tau^2 n(\mathbb{E}|\xi^{(n)}(u,v)|^2)^2} = \frac{9\langle P^{(n)}\Sigma^{(n)}P^{(n)}v, v\rangle^2\langle C^{(n)}\Sigma^{(n)}C^{(n)}u, u\rangle^2}{\langle P^{(n)}\Sigma^{(n)}P^{(n)}v, v\rangle^2\langle C^{(n)}\Sigma^{(n)}C^{(n)}u, u\rangle^2}\lim_{n\to\infty}\frac{1}{\tau^2 n} = 0,$$

and the result follows. □

To complete the proof of Theorem 5, it is enough to use representation (5.4) and bound (3.2) of Theorem 3. Since $\mathbf{r}(\Sigma^{(n)}) = o(n)$, it follows from bound (3.2) that

$$\langle R^{(n)}u, v\rangle = o_{\mathbb{P}}(n^{-1/2}),$$

and the result follows from Lemma 7. $\square$

Remark 4. *Under the assumption*

$$\mathbf{r}(\Sigma^{(n)}) = o(n^{1/2}) \quad \text{as } n \to \infty, \tag{5.6}$$

the finite dimensional distributions of stochastic processes

$$n^{1/2}\langle(\hat{P}^{(n)} - P^{(n)})u, v\rangle, \quad u, v \in \mathbb{H}$$

converge weakly as $n \to \infty$ to the finite dimensional distributions of Y. Indeed, by Theorem 4 and bound (4.1),

$$\|\mathbb{E}\hat{P}^{(n)} - P^{(n)}\|_\infty = O\left(\frac{\mathbf{r}(\Sigma^{(n)})}{n}\right) = o(n^{-1/2}),$$

and the claim follows from Theorem 5.

6. Asymptotics and concentration bounds for linear forms of empirical eigenvectors corresponding to a simple eigenvalue

We will discuss special versions of some of the results of the previous sections in the case of spectral projectors corresponding to an isolated simple eigenvalue. In this case, it becomes natural to state the results in terms of eigenvectors rather than spectral projectors.

Suppose μ_r is a simple eigenvalue of Σ, that is, μ_r is of multiplicity $m_r = 1$ so that the spectral projector P_r is of rank 1: $P_r = \theta_r \otimes \theta_r$, where θ_r is a unit eigenvector corresponding to μ_r. This implies that the projector $\hat{P}_r$ is also of rank 1 with a high probability (provided that $\mathbb{E}\|\hat{\Sigma} - \Sigma\|_\infty \le (1-\gamma)\frac{\bar{g}_r}{2}$ which will be assumed in what follows). Let $\hat{P}_r = \hat{\theta}_r \otimes \hat{\theta}_r$ and suppose that the sign of $\hat{\theta}_r$ is chosen in such a way that $\langle\hat{\theta}_r, \theta_r\rangle \ge 0$. Since the eigenvectors $\hat{\theta}_r, \theta_r$ are defined only up to their signs, there is no loss of generality in such an assumption.

Under the assumptions of Theorem 4,

$$\mathbb{E}\hat{P}_r = P_r + P_r W_r P_r + T_r,$$

where $W_r = \mathbb{E}\hat{P}_r - P - r$ is the bias of $\hat{P}_r$ and the remainder T_r satisfies bound (4.2):

$$\|T_r\|_\infty \le D_\gamma \frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\sqrt{\frac{\mathbf{r}(\Sigma)}{n}}\frac{1}{\sqrt{n}}. \tag{6.1}$$

Note that

$$\langle P_r W_r P_r u, v\rangle = \langle P_r W_r \theta_r, v\rangle\langle\theta_r, u\rangle = \langle W_r\theta_r, \theta_r\rangle\langle\theta_r, u\rangle\langle\theta_r, v\rangle.$$

Therefore,

$$P_r W_r P_r = \langle W_r\theta_r, \theta_r\rangle(\theta_r \otimes \theta_r) = b_r P_r$$

and

$$\mathbb{E}\hat{P}_r = (1 + b_r)P_r + T_r, \tag{6.2}$$

where $b_r := \langle W_r\theta_r, \theta_r\rangle$ is a real number characterizing the bias of $\hat{P}_r$. Note that

$$b_r = \langle \mathbb{E}\hat{P}_r\theta_r, \theta_r\rangle - \langle P_r\theta_r, \theta_r\rangle = \mathbb{E}\langle\hat{\theta}_r, \theta_r\rangle^2 - 1,$$

implying that $b_r \in [-1, 0]$ (with $b_r = -1$ being equivalent to $\hat{\theta}_r \perp \theta_r$ a.s. and $b_r = 0$ being equivalent to $\hat{\theta}_r = \theta_r$ a.s.). In what follows, we will often assume that b_r is bounded away from -1 which would ensure that the bias is not too large. In fact, it follows from bound (4.1) that, under the assumption that $\mathbf{r}(\Sigma) \lesssim n$,

$$|b_r| \lesssim \frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\frac{\mathbf{r}(\Sigma)}{n}, \tag{6.3}$$

so, b_r is small provided that $\frac{\|\Sigma\|_\infty}{\bar{g}_r}$ remains bounded and $\mathbf{r}(\Sigma) = o(n)$.

Theorem 6. *Suppose that condition (3.1) holds for some $\gamma \in (0, 1)$ and also that*

$$1 + b_r \geq \frac{\gamma}{2}. \tag{6.4}$$

Then, there exists a constant $C_\gamma > 0$ such that for all $t \geq 1$ with probability at least $1 - e^{-t}$

$$\left|\langle\hat{\theta}_r - \sqrt{1 + b_r}\theta_r, u\rangle\right| \leq C_\gamma \frac{\|\Sigma\|_\infty}{\bar{g}_r}\left(\sqrt{\frac{t}{n}} \vee \frac{t}{n}\right)\|u\|.$$

Proof. We need the following lemma that provides a representation of the linear functional $\langle\hat{\theta}_r - \theta_r, u\rangle$ in terms of bilinear form of operator $\hat{P}_r - P_r$.

Lemma 8. *For all $u \in \mathbb{H}$,*

$$\langle\hat{\theta}_r - \theta_r, u\rangle = \frac{\langle(\hat{P}_r - P_r)\theta_r, u\rangle - (\sqrt{1 + \langle(\hat{P}_r - P_r)\theta_r, \theta_r\rangle} - 1)\langle\theta_r, u\rangle}{\sqrt{1 + \langle(\hat{P}_r - P_r)\theta_r, \theta_r\rangle}}. \tag{6.5}$$

Proof. The following representation is obvious

$$(\hat{P}_r - P_r)\theta_r = \hat{\theta}_r - \theta_r + \langle\hat{\theta}_r - \theta_r, \theta_r\rangle\theta_r + \langle\hat{\theta}_r - \theta_r, \theta_r\rangle(\hat{\theta}_r - \theta_r)$$

and it implies that

$$\langle\hat{\theta}_r - \theta_r, u\rangle = \frac{\langle(\hat{P}_r - P_r)\theta_r, u\rangle - \langle\hat{\theta}_r - \theta_r, \theta_r\rangle\langle\theta_r, u\rangle}{1 + \langle\hat{\theta}_r - \theta_r, \theta_r\rangle}. \tag{6.6}$$

For $u = \theta_r$, it yields

$$\langle\hat{\theta}_r - \theta_r, \theta_r\rangle^2 + 2\langle\hat{\theta}_r - \theta_r, \theta_r\rangle = \big\langle(\hat{P}_r - P_r)\theta_r, \theta_r\big\rangle$$

and, since $\langle\hat{\theta}_r, \theta_r\rangle \geq 0$, we easily get that

$$\langle\hat{\theta}_r, \theta_r\rangle = \sqrt{1 + \big\langle(\hat{P}_r - P_r)\theta_r, \theta_r\big\rangle}.$$

Substituting this into (6.6) gives the result. □

Denote

$$\rho_r(u) := \big\langle\big(\hat{P}_r - (1 + b_r)P_r\big)\theta_r, u\big\rangle.$$

We can rewrite (6.5) as follows:

$$\begin{aligned}
\langle \hat{\theta}_r - \theta_r, u\rangle &= \frac{b_r\langle \theta_r, u\rangle + \rho_r(u) - (\sqrt{1+b_r+\rho_r(\theta_r)} - 1)\langle \theta_r, u\rangle}{\sqrt{1+b_r+\rho_r(\theta_r)}} \\
&= \left(\frac{1+b_r}{\sqrt{1+b_r+\rho_r(\theta_r)}} - 1\right)\langle \theta_r, u\rangle + \frac{\rho_r(u)}{\sqrt{1+b_r+\rho_r(\theta_r)}} \\
&= (\sqrt{1+b_r} - 1)\langle \theta_r, u\rangle + \left(\frac{1+b_r}{\sqrt{1+b_r+\rho_r(\theta_r)}} - \sqrt{1+b_r}\right)\langle \theta_r, u\rangle + \frac{\rho_r(u)}{\sqrt{1+b_r+\rho_r(\theta_r)}} \\
&= (\sqrt{1+b_r} - 1)\langle \theta_r, u\rangle + \frac{\rho_r(u)}{\sqrt{1+b_r+\rho_r(\theta_r)}} \\
&\quad - \frac{\sqrt{1+b_r}}{\sqrt{1+b_r+\rho_r(\theta_r)}(\sqrt{1+b_r+\rho_r(\theta_r)} + \sqrt{1+b_r})}\rho_r(\theta_r)\langle \theta_r, u\rangle,
\end{aligned}$$

which implies

$$\begin{aligned}
\langle \hat{\theta}_r - \sqrt{1+b_r}\theta_r, u\rangle &= \frac{\rho_r(u)}{\sqrt{1+b_r+\rho_r(\theta_r)}} \\
&\quad - \frac{\sqrt{1+b_r}}{\sqrt{1+b_r+\rho_r(\theta_r)}(\sqrt{1+b_r+\rho_r(\theta_r)} + \sqrt{1+b_r})}\rho_r(\theta_r)\langle \theta_r, u\rangle. \qquad (6.7)
\end{aligned}$$

The next bound on $\rho_r(u)$ follows from Corollary 1 and from Theorem 4, and it holds with some constant $D_\gamma \geq 1$ and with probability at least $1 - e^{-t}$:

$$\rho_r(u) \leq D\frac{\|\Sigma\|_\infty}{\bar{g}_r}\sqrt{\frac{t}{n}}\|u\| + D_\gamma\frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{t}{n}} \vee \frac{t}{n}\right)\sqrt{\frac{t}{n}}\|u\|.$$

Recall that, under condition (3.1),

$${C_1}^{-1}\|\Sigma\|_\infty\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} < \frac{\bar{g}_r}{2}$$

with some constant $C_1 \geq 1$ and assume that also

$${C_1}^{-1}\|\Sigma\|_\infty\left(\sqrt{\frac{t}{n}} \vee \frac{t}{n}\right) < \frac{\bar{g}_r}{2}. \qquad (6.8)$$

Then we have

$$\left|\rho_r(u)\right| \leq (D + C_1 D_\gamma)\frac{\|\Sigma\|_\infty}{\bar{g}_r}\sqrt{\frac{t}{n}}\|u\|. \qquad (6.9)$$

Assuming that

$$(D + C_1 D_\gamma)\|\Sigma\|_\infty\left(\sqrt{\frac{t}{n}} \vee \frac{t}{n}\right) \leq \frac{\gamma}{2}\frac{\bar{g}_r}{2}, \qquad (6.10)$$

which, of course, implies condition (6.8), we get that, with probability at least $1 - e^{-t}$, $\rho_r(\theta_r) \leq \frac{\gamma}{4}$. As a consequence, under condition (6.4), with the same probability,

$$1 + b_r + \rho_r(\theta_r) \geq \frac{\gamma}{4},$$

and it follows from (6.7) and (6.9) that with probability at least $1-2e^{-t}$

$$\left|\langle\hat{\theta}_r-\sqrt{1+b_r}\theta_r,u\rangle\right|\lesssim_\gamma \frac{\|\Sigma\|_\infty}{\bar{g}_r}\sqrt{\frac{t}{n}}\|u\|.$$

It remains to consider the case when condition (6.10) does not hold. In this case,

$$\frac{4(D+C_1D_\gamma)}{\gamma}\frac{\|\Sigma\|_\infty}{\bar{g}_r}\left(\sqrt{\frac{t}{n}}\vee\frac{t}{n}\right)\geq 1,$$

and, under the assumption $\|u\|\leq 1$, we simply have

$$\left|\langle\hat{\theta}_r-\sqrt{1+b_r}\theta_r,u\rangle\right|\leq\|\hat{\theta}_r\|+\sqrt{1+b_r}\|\theta_r\|\leq 2\leq\frac{8(D+C_1D_\gamma)}{\gamma}\frac{\|\Sigma\|_\infty}{\bar{g}_r}\left(\sqrt{\frac{t}{n}}\vee\frac{t}{n}\right),$$

implying the bound of the theorem. To complete the proof, it is enough to adjust the constants properly (to write the probability bound as $1-e^{-t}$). □

Remark 5. *In view of Remark 2 of Section 3, the bound on $\rho_r(\theta_r)$ that appeared in the above proof could be improved as follows: with probability at least $1-e^{-t}$,*

$$\left|\rho_r(\theta_r)\right|\leq D_\gamma\frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}}\vee\sqrt{\frac{t}{n}}\vee\frac{t}{n}\right)\sqrt{\frac{t}{n}}.$$

This implies that with the same probability

$$\left|\langle\hat{\theta}_r-\sqrt{1+b_r}\theta_r,\theta_r\rangle\right|\leq C_\gamma\frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}}\vee\sqrt{\frac{t}{n}}\vee\frac{t}{n}\right)\left(\sqrt{\frac{t}{n}}\vee\frac{t}{n}\right). \tag{6.11}$$

Based on Theorem 6, it is easy to develop a simple $\sqrt{n}$-consistent estimator of the bias parameter b_r and suggest an approach to bias reduction in the problem of estimation of linear functionals of eigenvectors of Σ. Suppose, for simplicity, that the sample size is an even number $2n$ and divide the sample $(X_1,\dots,X_{2n})$ into two subsamples of size n each (the first n observations and the rest). Let $\hat{\Sigma}_n$ be the sample covariance based on the first subsample and $\hat{\Sigma}_n'$ be the sample covariance based on the second subsample. For a simple eigenvalue μ_r with an eigenvector θ_r, denote by $\hat{\theta}_r$ the corresponding eigenvector of $\hat{\Sigma}_n$ and by $\hat{\theta}_r'$ the corresponding eigenvector of $\hat{\Sigma}_n'$. Assume that their signs are chosen in such a way that $\langle\hat{\theta}_r,\hat{\theta}_r'\rangle\geq 0$. Define

$$\hat{b}_r:=\langle\hat{\theta}_r,\hat{\theta}_r'\rangle-1$$

and

$$\tilde{\theta}_r:=\frac{\hat{\theta}_r}{\sqrt{1+\hat{b}_r}}.$$

Proposition 3. *Under the assumptions and notations of Theorem 6, for some constant $C_\gamma>0$ with probability at least $1-e^{-t}$,*

$$|\hat{b}_r-b_r|\leq C_\gamma\frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}}\vee\sqrt{\frac{t}{n}}\vee\frac{t}{n}\right)\left(\sqrt{\frac{t}{n}}\vee\frac{t}{n}\right), \tag{6.12}$$

and, for all $u\in\mathbb{H}$,

$$\left|\langle\tilde{\theta}_r-\theta_r,u\rangle\right|\leq C_\gamma\frac{\|\Sigma\|_\infty}{\bar{g}_r}\left(\sqrt{\frac{t}{n}}\vee\frac{t}{n}\right)\|u\|. \tag{6.13}$$

Proof. It follows from the definition of $\hat{b}_r$ that

$$
\begin{aligned}
|\hat{b}_r - b_r| &= \big|\langle\hat{\theta}_r, \hat{\theta}_r'\rangle - (1+b_r)\big| \\
&= \big|\sqrt{1+b_r}\langle\hat{\theta}_r - \sqrt{1+b_r}\theta_r, \theta_r\rangle + \sqrt{1+b_r}\langle\hat{\theta}_r' - \sqrt{1+b_r}\theta_r, \theta_r\rangle \\
&\quad + \langle\hat{\theta}_r - \sqrt{1+b_r}\theta_r, \hat{\theta}_r' - \sqrt{1+b_r}\theta_r\rangle\big| \\
&\le \big|\sqrt{1+b_r}\langle\hat{\theta}_r - \sqrt{1+b_r}\theta_r, \theta_r\rangle\big| + \big|\sqrt{1+b_r}\langle\hat{\theta}_r' - \sqrt{1+b_r}\theta_r, \theta_r\rangle\big| \\
&\quad + \big|\langle\hat{\theta}_r - \sqrt{1+b_r}\theta_r, \hat{\theta}_r' - \sqrt{1+b_r}\theta_r\rangle\big|.
\end{aligned} \tag{6.14}
$$

By bound (6.11), with probability at least $1 - e^{-t}$

$$
\big|\langle\hat{\theta}_r - \sqrt{1+b_r}\theta_r, \theta_r\rangle\big| \le C_\gamma \frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{t}{n}} \vee \frac{t}{n}\right)\left(\sqrt{\frac{t}{n}} \vee \frac{t}{n}\right)
$$

and with the same probability

$$
\big|\langle\hat{\theta}_r' - \sqrt{1+b_r}\theta_r, \theta_r\rangle\big| \le C_\gamma \frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{t}{n}} \vee \frac{t}{n}\right)\left(\sqrt{\frac{t}{n}} \vee \frac{t}{n}\right).
$$

By Theorem 6, conditionally on the second sample, with probability at least $1 - e^{-t}$

$$
\big|\langle\hat{\theta}_r - \sqrt{1+b_r}\theta_r, \hat{\theta}_r' - \sqrt{1+b_r}\theta_r\rangle\big| \le C_\gamma \frac{\|\Sigma\|_\infty}{\bar{g}_r}\left(\sqrt{\frac{t}{n}} \vee \frac{t}{n}\right)\big\|\hat{\theta}_r' - \sqrt{1+b_r}\theta_r\big\|. \tag{6.15}
$$

To bound the norm $\|\hat{\theta}_r' - \sqrt{1+b_r}\theta_r\|$ in the right-hand side, note that

$$
\begin{aligned}
\big\|\hat{\theta}_r' - \sqrt{1+b_r}\theta_r\big\| &\le \big\|\hat{\theta}_r' - \theta_r\big\| + |\sqrt{1+b_r} - 1| \\
&\le \sqrt{2}\big\|\hat{P}_r' - P_r\big\|_\infty + \frac{|b_r|}{1+\sqrt{1+b_r}} \le \sqrt{2}\big\|\hat{P}_r' - P_r\big\|_\infty + |b_r|,
\end{aligned}
$$

where $\hat{P}_r' := \hat{\theta}_r' \otimes \hat{\theta}_r'$ and we used the bound

$$
\big\|\hat{\theta}_r' - \theta_r\big\|^2 = 2 - 2\langle\hat{\theta}_r, \theta_r\rangle \le 2 - 2\langle\hat{\theta}_r, \theta_r\rangle^2 = 2 - 2\langle\hat{P}_r', P_r\rangle = \big\|\hat{P}_r' - P_r\big\|_2^2 \le 2\big\|\hat{P}_r' - P_r\big\|_\infty^2.
$$

Using bounds (2.7), (6.3) and Theorem 2, it is easy to show that with probability at least $1 - e^{-t}$

$$
\big\|\hat{\theta}_r' - \sqrt{1+b_r}\theta_r\big\| \lesssim_\gamma \frac{\|\Sigma\|_\infty}{\bar{g}_r}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{t}{n}} \vee \frac{t}{n}\right).
$$

Together with (6.15) this implies that, for some $C_\gamma > 0$, with probability at least $1 - 2e^{-t}$

$$
\big|\langle\hat{\theta}_r - \sqrt{1+b_r}\theta_r, \hat{\theta}_r' - \sqrt{1+b_r}\theta_r\rangle\big| \le C_\gamma \frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{t}{n}} \vee \frac{t}{n}\right)\left(\sqrt{\frac{t}{n}} \vee \frac{t}{n}\right).
$$

It remains to deduce from (6.14) that (6.12) holds with probability at least $1 - 4e^{-t}$. To write the probability bound as $1 - e^{-t}$, it is enough to adjust the constants.

Under the assumptions of Theorem 6, the proof of bound (6.13) is straightforward. □

We turn now to asymptotic normality of empirical spectral projectors. It is easy to see that (6.2), bound (6.1) on $\|T_r\|_\infty$ and Theorem 5 yield the following corollary.

Corollary 2. *Suppose that*

$$\sup_{n\geq 1}\left\|\Sigma^{(n)}\right\|_{\infty}<+\infty \tag{6.16}$$

and

$$\mathbf{r}\big(\Sigma^{(n)}\big)=o(n)\quad \text{as } n\to\infty. \tag{6.17}$$

Suppose also that Assumptions 2 and 3 of Section 5 hold. Finally, suppose that $\mu^{(n)}=\mu^{(n)}_{r_n}$ is an eigenvalue of $\Sigma^{(n)}$ of multiplicity 1. Denote $b^{(n)}=b^{(n)}_{r_n}$. Then, the finite dimensional distributions of stochastic processes

$$n^{1/2}\big\langle\big(\hat{P}^{(n)}-\big(1+b^{(n)}\big)P^{(n)}\big)u,v\big\rangle,\quad u,v\in\mathbb{H}$$

converge weakly as $n\to\infty$ to the finite dimensional distributions of centered Gaussian process $Y(u,v)$, $u,v\in\mathbb{H}$ with covariance Γ.

If, in addition, Assumption 4 holds, then, for all $\varphi_n,\psi_n:\mathbb{H}\mapsto\mathbb{H}$ such that $\varphi_n(u)\to u$, $\psi_n(u)\to u$ as $n\to\infty$ for all $u\in\mathbb{H}$, the finite dimensional distributions of stochastic processes

$$n^{1/2}\big\langle\big(\hat{P}^{(n)}-\big(1+b^{(n)}\big)P^{(n)}\big)\varphi_n(u),\psi_n(v)\big\rangle,\quad u,v\in\mathbb{H}$$

converge weakly as $n\to\infty$ to the same limit.

Note that under the assumptions of Corollary 2, $P^{(n)}=\theta^{(n)}\otimes\theta^{(n)}$ and, with probability tending to 1, $\hat{P}^{(n)}=\hat{\theta}^{(n)}\otimes\hat{\theta}^{(n)}$ for eigenvectors $\theta^{(n)}$ of $\Sigma^{(n)}$ and $\hat{\theta}^{(n)}$ of $\hat{\Sigma}_n$. We will be able to rephrase the corollary in terms of linear forms of eigenvectors rather than bilinear forms of spectral projectors.

Theorem 7. *Suppose that*

$$\sup_{n\geq 1}\left\|\Sigma^{(n)}\right\|_{\infty}<+\infty \tag{6.18}$$

and

$$\mathbf{r}\big(\Sigma^{(n)}\big)=o(n)\quad \text{as } n\to\infty. \tag{6.19}$$

Suppose also that Assumptions 2 and 4 hold and recall that, under these assumptions, $\theta^{(n)}\to\theta\in\mathbb{H}$ as $n\to\infty$. Finally, assume that the sign of $\hat{\theta}^{(n)}$ is chosen to satisfy the condition $\langle\hat{\theta}^{(n)},\theta^{(n)}\rangle\geq 0$. Then, the finite dimensional distributions of stochastic processes

$$n^{1/2}\big\langle\hat{\theta}^{(n)}-\sqrt{1+b^{(n)}}\theta^{(n)},u\big\rangle,\quad u\in\mathbb{H}$$

converge weakly as $n\to\infty$ to the finite dimensional distributions of centered Gaussian process $Y(\theta,u)$, $u\in\mathbb{H}$.

Proof. Denote

$$\rho^{(n)}(u):=\rho^{(n)}_{r_n}(u):=\big\langle\big(\hat{P}^{(n)}-\big(1+b^{(n)}\big)P^{(n)}\big)\theta^{(n)},u\big\rangle,\quad u\in\mathbb{H}.$$

It follows from Corollary 2 and the fact that $Y(\theta,\theta)=0$ (see also Theorem 5 and the definition of the process Y and its covariance) that the finite dimensional distributions of stochastic processes

$$n^{1/2}\big(\rho^{(n)}(u),\rho^{(n)}\big(\theta^{(n)}\big)\big),\quad u\in\mathbb{H} \tag{6.20}$$

converge weakly to the Gaussian process $(Y(\theta,u),0)$, $u\in\mathbb{H}$. In particular, this implies that

$$\rho^{(n)}\big(\theta^{(n)}\big)=O_{\mathbb{P}}\big(n^{-1/2}\big)=o_{\mathbb{P}}(1).$$

Under the conditions of the corollary, we also have that

$$b^{(n)} = O\left(\frac{\mathbf{r}(\Sigma^{(n)})}{n}\right) = o(1).$$

It follows from (6.7) that

$$\begin{aligned} n^{1/2}\langle\hat\theta^{(n)} - \sqrt{1+b^{(n)}}\theta^{(n)}, u\rangle &= \frac{n^{1/2}\rho^{(n)}(u)}{\sqrt{1+b^{(n)}+\rho^{(n)}(\theta^{(n)})}} \\ &\quad - \frac{\sqrt{1+b^{(n)}}}{\sqrt{1+b^{(n)}+\rho^{(n)}(\theta^{(n)})}(\sqrt{1+b^{(n)}+\rho^{(n)}(\theta^{(n)})}+\sqrt{1+b^{(n)}})} \\ &\quad \times n^{1/2}\rho^{(n)}\big(\theta^{(n)}\big)\langle\theta^{(n)}, u\rangle. \end{aligned}$$

This representation, the convergence of finite dimensional distribution of the process (6.20) and the fact that $\rho^{(n)}(\theta^{(n)}) = o_{\mathbb{P}}(1), b^{(n)} = o(1)$, imply the result. □

It turns out that the asymptotic normality also holds for the estimator with bias correction $\tilde\theta^{(n)} := \frac{\hat\theta^{(n)}}{\hat b^{(n)}}$, where $\hat b^{(n)} := \langle\hat\theta^{(n)}, \hat\theta'^{(n)}\rangle - 1$, $\hat\theta^{(n)}$, $\hat\theta'^{(n)}$ being empirical eigenvectors based on the first and on the second subsamples (of size n each) of a sample of size $2n$. As before, it is assumed that $\langle\hat\theta^{(n)}, \hat\theta'^{(n)}\rangle \geq 0$. We state the result without proof.

Theorem 8. *Under assumptions of Theorem 7, the finite dimensional distributions of stochastic processes*

$$\sqrt{n}\langle\tilde\theta^{(n)} - \theta^{(n)}, u\rangle, \quad u \in \mathbb{H}$$

converge weakly to the finite dimensional distributions of stochastic process $Y(\theta, u), u \in \mathbb{H}$.

Suppose $\mathbb{H} = \mathbb{R}^p$ and let $e_1, \dots, e_p$ be an orthonormal basis of the space $\mathbb{R}^p$. For $u \in \mathbb{R}^p$, let

$$\|u\|_{\ell_\infty} := \max_{1\leq j\leq p}\big|\langle u, e_j\rangle\big| = \max_{1\leq j\leq p}\big|u^{(j)}\big|.$$

We present now a non-asymptotic bound on $\|\tilde\theta_r - \theta_r\|_{\ell_\infty}$ that immediately follows from Proposition 3.

Corollary 3. *Suppose the assumptions of Theorem 6 hold. Then, with probability at least $1 - e^{-t}$,*

$$\|\tilde\theta_r - \theta_r\|_{\ell_\infty} \leq C_\gamma \frac{\|\Sigma\|_\infty}{\bar g_r}\left(\sqrt{\frac{t+\log p}{n}} \vee \frac{t+\log p}{n}\right).$$

Example 3 (Eigenvector support recovery). *Our goal is to recover the support of eigenvector θ_r denoted by*

$$J_r := \operatorname{supp}(\theta_r) := \big\{j : \theta_r^{(j)} \neq 0\big\}.$$

It follows from Corollary 3 that a simple hard-thresholding procedure can achieve support recovery. Define $\tilde J_r = \{j : |\tilde\theta_r^{(j)}| > \beta_n\}$, where

$$\beta_n := C_\gamma \frac{\|\Sigma\|_\infty}{\bar g_r}\left(\sqrt{\frac{t+\log p}{n}} \vee \frac{t+\log p}{n}\right).$$

If $\rho := \min_{j\in J_r}|\theta_r^{(j)}| > 2\beta_n$, then we can immediately deduce from Corollary 3 that $\mathbb{P}(\tilde J_r = J_r) \geq 1 - e^{-t}$. It is well known that the theoretical threshold to perform support recovery in the Gaussian sequence space model is $\beta_n^* \asymp \sigma\sqrt{\frac{t+\log p}{n}}$ where σ is the noise variance. The above threshold β_n in eigenvector support recovery is similar with the noise variance σ replaced by $\frac{\|\Sigma\|_\infty}{\bar g_r}$.

Example 4 (Sparse PCA oracle inequality). *We propose a new estimator of θ_r that satisfies a sparsity oracle inequality with sharp minimax l_2-norm rate (see [31] for more details about minimax rates in sparse PCA). This estimator is computationally feasible and also adaptive in the sense that no prior knowledge about the sparsity of θ_r is required. Consider the estimator $\tilde{\theta}_r \in \mathbb{R}^p$ obtained by keeping all the components of $\tilde{\theta}_r$ with their indices in $\tilde{J}_r$ and setting all the remaining components equal to 0. We denote by $\|\theta_r\|_{l_0}$ the number of non-zero components of θ_r. Combining the above support recovery property with Corollary 3, we immediately get the following result.*

Theorem 9. *Let the conditions of Theorem 6 be satisfied. Assume in addition that $\rho = \min_{j\in J_r} |\theta_r^{(j)}| \geq 2\beta_n$. Then, with probability at least $1 - e^{-t}$*

$$\|\tilde{\theta}_r - \theta_r\|_{l_2}^2 \leq C_\gamma^2 \frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2} \|\theta_r\|_{l_0} \left(\frac{t+\log p}{n} \vee \left(\frac{t+\log p}{n}\right)^2\right). \tag{6.21}$$

Remark 6. *The main focus of this paper was on concentration and asymptotic properties of bilinear forms of spectral projectors, in other words, on their asymptotic properties in the weak topology in the space of bounded operators. In this case, it was possible to show that the bilinear forms (centered with their expectations) are asymptotically normal with the standard rate $\sqrt{n}$. Such results do not hold in the case of strong operator topology. For instance, in the case when the true spectral projector $P_r = \theta_r \otimes \theta_r$ is of rank 1, it is easy to deduce from the bounds of this section that*

$$\mathbb{E}\|\hat{P}_r u - \mathbb{E}\hat{P}_r u\|^2 \sim -b_r \langle\theta_r, u\rangle^2 \asymp \frac{\|\Sigma\|_\infty^2}{\bar{g}_r^2} \frac{\mathbf{r}(\Sigma)}{n} \langle\theta_r, u\rangle^2$$

(under the assumptions that $\mathbf{r}(\Sigma) \to \infty$ and $\mathbf{r}(\Sigma) = o(n)$). This means that, if $\langle\theta_r, u\rangle \neq 0$, then $\mathbb{E}\|\hat{P}_r u - \mathbb{E}\hat{P}_r u\|^2$ and $\mathbb{E}\|\hat{P}_r u - P_r u\|^2$ are essentially of the same order as $\mathbb{E}\|\hat{P}_r - P_r\|_2^2$ (see[17] for the computation of the last quantity).

References

[1] A. A. Amini and M. J. Wainwright. High-dimensional analysis of semidefinite relaxations for sparse principal components. *Ann. Statist.* **37** (5B) (2009) 2877–2921. MR2541450

[2] T. W. Anderson. Asymptotic theory for principal component analysis. *Ann. Math. Stat.* **34** (1963) 122–148. MR0145620

[3] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*, 3rd edition. *Wiley Series in Probability and Statistics*. Wiley-Interscience, Hoboken, NJ, 2003. MR1990662

[4] J. Baik, G. Ben Arous and S. Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33** (5) (2005) 1643–1697. MR2165575

[5] A. Birnbaum, I. M. Johnstone, B. Nadler and D. Paul. Minimax bounds for sparse PCA with noisy high-dimensional data. *Ann. Statist.* **41** (3) (2013) 1055–1084. MR3113803

[6] G. Blanchard, O. Bousquet and L. Zwald. Statistical properties of kernel principal component analysis. *Mach. Learn.* **66** (2–3) (2007) 259–294.

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

[8] T. T. Cai, Z. Ma and Y. Wu. Sparse PCA: Optimal rates and adaptive estimation. *Ann. Statist.* **41** (6) (2013) 3074–3110. MR3161458

[9] A. d'Aspremont, L. El Ghaoui, M. I. Jordan and G. R. G. Lanckriet. A direct formulation for sparse PCA using semidefinite programming. *SIAM Rev.* **49** (3) (2007) 434–448 (electronic). MR2353806

[10] J. Dauxois, A. Pousse and Y. Romain. Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *J. Multivariate Anal.* **12** (1) (1982) 136–154. MR0650934

[11] I. M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** (2) (2001) 295–327. MR1863961

[12] I. M. Johnstone and A. Y. Lu. On consistency and sparsity for principal components analysis in high dimensions. *J. Amer. Statist. Assoc.* **104** (486) (2009) 682–693. MR2751448

[13] I. M. Johnstone and Z. Ma. Fast approach to the Tracy–Widom law at the edge of GOE and GUE. *Ann. Appl. Probab.* **22** (5) (2012) 1962–1988. MR3025686

[14] T. Kato. *Perturbation Theory for Linear Operators*. Springer, New York, 1980.

[15] A. Kneip and K. J. Utikal. Inference for density families using functional principal component analysis. *J. Amer. Statist. Assoc.* **96** (454) (2001) 519–532. MR1946423

[16] V. Koltchinskii. Asymptotics of spectral projections of some random matrices approximating integral operators. In *High Dimensional Probability (Oberwolfach, 1996)* 191–227. *Progr. Probab.* **43**. Birkhäuser, Basel, 1998. MR1652327

[17] V. Koltchinskii and K. Lounici. Normal approximation and concentration of spectral projectors of sample covariance, 2015. Available at arXiv:1504.07333.

[18] V. Koltchinskii and K. Lounici. Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli*. To appear. Available at arXiv:1405.2468. MR2555200

[19] K. Lounici. Sparse principal component analysis with missing observations. In *High Dimensional Probability VI* 327–356. *Prog. Proba., Institute of Mathematical Statistics (IMS) Collections* **66**, 2013.

[20] K. Lounici. High-dimensional covariance matrix estimation with missing observations. *Bernoulli* **20** (3) (2014) 1029–1058. MR3217437

[21] Z. Ma. Accuracy of the Tracy–Widom limits for the extreme eigenvalues in white Wishart matrices. *Bernoulli* **18** (1) (2012) 322–359. MR2888709

[22] Z. Ma. Sparse principal component analysis and iterative thresholding. *Ann. Statist.* **41** (2) (2013) 772–801. MR3099121

[23] V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb.* **72** (114) (1967) 507–536. MR0208649

[24] A. Mas and F. Ruymgaart. High dimensional principal projections. *Complex Anal. Oper. Theory* **9** (1) (2015) 35–63. MR3300524

[25] D. Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica* **17** (4) (2007) 1617–1642. MR2399865

[26] D. Paul and I. M. Johnstone. Augmented sparse principal component analysis for high dimensional data, 2007. Available at arXiv:1202.1242.

[27] J. O. Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer, New York, 1997. MR2168993

[28] F. Riesz and B. Sz.-Nagy. *Functional Analysis*. Dover, New York, 1990. MR1068530

[29] B. Schölkopf, A. Smola and K. R. Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.* **10** (5) (1998) 1299–1319.

[30] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing* 210–268. Cambridge University Press, Cambridge, 2012. MR2963170

[31] V. Vu and J. Lei. Minimax rates of estimation for sparse PCA in high dimensions. *J. Mach. Learn. Res.* **22** (2012) 1278–1286.