

# On the convergence of the extremal eigenvalues of empirical covariance matrices with dependence

Djalil Chafai<sup>1</sup>  · Konstantin Tikhomirov<sup>2</sup>

Received: 12 November 2015 / Revised: 15 April 2017 / Published online: 26 April 2017  
© Springer-Verlag Berlin Heidelberg 2017

**Abstract** Consider a sample of a centered random vector with unit covariance matrix. We show that under certain regularity assumptions, and up to a natural scaling, the smallest and the largest eigenvalues of the empirical covariance matrix converge, when the dimension and the sample size both tend to infinity, to the left and right edges of the Marchenko–Pastur distribution. The assumptions are related to tails of norms of orthogonal projections. They cover isotropic log-concave random vectors as well as random vectors with i.i.d. coordinates with almost optimal moment conditions. The method is a refinement of the rank one update approach used by Srivastava and Vershynin to produce non-asymptotic quantitative estimates. In other words we provide a new proof of the Bai and Yin theorem using basic tools from probability theory and linear algebra, together with a new extension of this theorem to random matrices with dependent entries.

**Keywords** Convex body · Random matrix · Covariance matrix · Singular value · Operator norm · Sherman–Morrison formula · Thin-shell inequality · Log-concave distribution · Dependence

**Mathematics Subject Classification** 52A23 · 15B52 · 60B20 · 62J10

---

Dedicated to Alain Pajor & Nicole Tomczak-Jaegermann.

---

✉ Djalil Chafai  
djalil@chafai.net

<sup>1</sup> CEREMADE, PSL, IUF, Université Paris-Dauphine, Paris, France

<sup>2</sup> Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Canada

## Contents

- 1 Introduction . . . . . 848
  - 1.1 Outline of the argument and novelty . . . . . 854
  - 1.2 Discussion and extensions . . . . . 855
  - 1.3 Structure . . . . . 856
  - 1.4 Notations . . . . . 856
- 2 Proof of Propositions 1.3–1.5 . . . . . 856
- 3 Proof of Theorem 1.6 . . . . . 866
  - 3.1 Feasible lower shift . . . . . 868
  - 3.2 Randomization and control of expectations . . . . . 870
  - 3.3 Proof of Theorem 1.6, completed . . . . . 873
- 4 Proof of Theorem 1.8 . . . . . 875
  - 4.1 Feasible upper shift . . . . . 876
  - 4.2 Randomization and control of expectations . . . . . 880
  - 4.3 Proof of Theorem 1.8, completed . . . . . 884
- References . . . . . 888

## 1 Introduction

Let  $\mathbb{N} = \{1, 2, \dots\}$  and let  $(X_n)_{n \in \mathbb{N}}$  be a sequence of random vectors where for each  $n \in \mathbb{N}$  the random vector  $X_n$  takes values in  $\mathbb{R}^n$ , is centered, with unit covariance (isotropy):

$$\mathbb{E}(X_n) = 0 \quad \text{and} \quad \mathbb{E}(X_n \otimes X_n) = I_n \tag{1.1}$$

where  $I_n$  is the  $n \times n$  identity matrix. Let  $(m_n)_{n \in \mathbb{N}}$  be a sequence in  $\mathbb{N}$  such that

$$0 < \underline{\rho} := \liminf_{n \rightarrow \infty} \frac{n}{m_n} \leq \limsup_{n \rightarrow \infty} \frac{n}{m_n} =: \bar{\rho} < \infty. \tag{1.2}$$

For every  $n \in \mathbb{N}$ , let  $X_n^{(1)}, \dots, X_n^{(m_n)}$  be i.i.d. copies of  $X_n$ . Their empirical covariance matrix is the  $n \times n$  symmetric positive semidefinite random matrix

$$\widehat{\Sigma}_n := \frac{1}{m_n} \sum_{k=1}^{m_n} X_n^{(k)} \otimes X_n^{(k)}. \tag{1.3}$$

If  $\mathbb{X}_n$  denotes the  $m_n \times n$  rectangular random matrix with i.i.d. rows  $X_n^{(1)}, \dots, X_n^{(m_n)}$  then

$$\widehat{\Sigma}_n = \frac{1}{m_n} \mathbb{X}_n^\top \mathbb{X}_n.$$

Note that  $\mathbb{E}\widehat{\Sigma}_n = \mathbb{E}(X_n \otimes X_n) = I_n$ . For convenience we define the random matrix

$$A_n := m_n \widehat{\Sigma}_n = \mathbb{X}_n^\top \mathbb{X}_n = \sum_{k=1}^{m_n} X_n^{(k)} \otimes X_n^{(k)}. \tag{1.4}$$

The eigenvalues of  $A_n$  are squares of the singular values of  $\mathbb{X}_n$ , and in particular

$$\lambda_{\max}(A_n) = s_{\max}(\mathbb{X}_n)^2 = \max_{\|x\|=1} \|\mathbb{X}_n x\|^2 = \|\mathbb{X}_n\|^2.$$

When  $m_n \geq n$  then the smallest eigenvalue of  $A_n$  satisfies

$$\lambda_{\min}(A_n) = s_{\min}(\mathbb{X}_n)^2 = \min_{\|x\|=1} \|\mathbb{X}_n x\|^2 = \|\mathbb{X}_n^{-1}\|^{-2},$$

where by  $\mathbb{X}_n^{-1}$  we denote the inverse linear operator for  $\mathbb{X}_n$  defined on the image of  $\mathbb{X}_n$ . Of course, the last formula holds only when  $\mathbb{X}_n$  is invertible (impossible if  $m_n < n$ ). Above and in the sequel we denote by  $\|x\| = (x_1^2 + \dots + x_n^2)^{1/2}$  the Euclidean norm of  $x \in \mathbb{R}^n$ .

Let us denote by  $(X_{n,k})_{1 \leq k \leq n}$  the components of the random vector  $X_n$ . If the random variables  $(X_{n,k})_{n \geq 1, 1 \leq k \leq n}$  are i.i.d. standard Gaussians, then the law of the random matrix  $\widehat{\Sigma}_n$  is known as the real Wishart law, and constitutes a sort of a matrix version of the  $\chi^2(n)$  law. The law of the eigenvalues of  $\widehat{\Sigma}_n$  is then called the Laguerre Orthogonal Ensemble, a Boltzmann–Gibbs measure with density on  $\{\lambda \in [0, \infty)^m : \lambda_1 \geq \dots \geq \lambda_n\}$  proportional to

$$\lambda \mapsto \exp \left( -\frac{m_n}{2} \sum_{k=1}^n \lambda_k + \frac{m_n - n + 1}{2} \sum_{k=1}^n \log(\lambda_k) + \sum_{i < j} \log |\lambda_i - \lambda_j| \right),$$

see [38, (7.4.2) p. 192]. This exactly solvable Gaussian model allows to deduce sharp asymptotics for the empirical spectral distribution as well as for the extremal eigenvalues of  $\widehat{\Sigma}_n$ . The famous Marchenko–Pastur theorem [35, 38] states that if

$$\frac{n}{m_n} \xrightarrow{n \rightarrow \infty} \rho \quad \text{with } \rho \in (0, \infty)$$

then almost surely as  $n \rightarrow \infty$ , the empirical spectral distribution of  $\widehat{\Sigma}_n$  tends weakly with respect to continuous and bounded test functions to a non-random distribution, namely

$$a.s. \quad \frac{1}{m_n} \sum_{k=1}^n \delta_{\lambda_k(\widehat{\Sigma}_n)} \xrightarrow[n \rightarrow \infty]{C_b} \mu_\rho \tag{1.5}$$

where  $\mu_\rho$  is the Marchenko–Pastur distribution on  $[a^-, a^+]$  with  $a^\pm = (1 \pm \sqrt{\rho})^2$  given by

$$\mu_\rho(dx) = \frac{\rho - 1}{\rho} \mathbf{1}_{\rho > 1} \delta_0 + \frac{\sqrt{(a^+ - x)(x - a^-)}}{\rho 2\pi x} \mathbf{1}_{[a^-, a^+]}(x) dx.$$

It is a mixture between a Dirac mass at point 0 and an absolutely continuous law. The atom at 0 disappears when  $\rho \leq 1$  and is a reminiscence of the rank of  $\widehat{\Sigma}_n$ . The asymptotic phenomenon (1.5) holds beyond the Gaussian case. In particular it was shown by

Pajor and Pastur [36] that it holds if for every  $n \in \mathbb{N}$  the distribution of the isotropic random vector  $X_n$  is log-concave. Recall that a probability measure  $\mu$  on  $\mathbb{R}^n$  with density  $\varphi$  is log-concave when  $\varphi = e^{-V}$  with  $V$  convex, see [18, 20]. Log-concavity allows some kind of geometric dependence but imposes sub-exponential tails. The asymptotic phenomenon (1.5) also holds if  $(X_{n,k})_{n \geq 1, 1 \leq k \leq n}$  are i.i.d. with finite second moment [9, 38]. An extension to various other models can be found in Bai and Zhou [13], Pastur and Shcherbina [38], Adamczak [1], and Adamczak and Chafaï [3].

The weak convergence (1.5) does not provide much information at the edge on the behavior of the extremal atoms, and what one can actually extract from (1.5) is that

$$a.s. \limsup_{n \rightarrow \infty} \lambda_{\min}(\widehat{\Sigma}_n) \leq (1 - \sqrt{\rho})^2 \leq (1 + \sqrt{\rho})^2 \leq \liminf_{n \rightarrow \infty} \lambda_{\max}(\widehat{\Sigma}_n) \tag{1.6}$$

where the first inequality is considered only in the case where  $m_n \geq n$ . If  $(X_{n,k})_{n \geq 1, 1 \leq k \leq n}$  are i.i.d. with finite fourth moment then it was shown by Bai and Yin [11, 12, 53] using combinatorics that the convergence in (1.6) holds:

$$a.s. (1 - \sqrt{\rho})^2 = \lim_{n \rightarrow \infty} \lambda_{\min}(\widehat{\Sigma}_n) \leq \lim_{n \rightarrow \infty} \lambda_{\max}(\widehat{\Sigma}_n) = (1 + \sqrt{\rho})^2, \tag{1.7}$$

where the first equality is considered only in the case where  $m_n \geq n$ . The convergence of the largest eigenvalue in the right hand side of (1.7) does not take place if  $(X_{n,k})_{n \geq 1, 1 \leq k \leq n}$  are i.i.d. with infinite fourth moment, see [9]. It was understood recently that the convergence of the smallest eigenvalue in the left hand side of (1.7) holds actually as soon as  $(X_{n,k})_{n \geq 1, 1 \leq k \leq n}$  are i.i.d. with finite second moment, see Tikhomirov [49].

An analytic proof of (1.7) based on the resolvent is also available, and we refer to Bordenave [17] for the i.i.d. case, and to Bai and Silverstein [8], Pillai and Yin [41], and Richard and Guionnet [42] for more sophisticated models still not including the case in which the law of  $X_n$  is log-concave for every  $n \in \mathbb{N}$ . Note that the analytic approach was also used for various models of random matrices by Haagerup and Thorbjørnsen [28], Schultz [45], and by Capitaine and Donati-Martin [21].

The study of quantitative high dimensional non-asymptotic properties of the smallest and of the largest eigenvalues of empirical covariance matrices was the subject of an intense line of research in the recent years; in connection with log-concavity, see for instance Adamczak et al. [5, 6], Rudelson and Vershynin [44], Koltchinskii and Mendelson [32], and references therein.

Non-asymptotic estimates for (1.7) were obtained by Srivastava and Vershynin [48] using a rank one update strategy which takes advantage of the decomposition (1.4). This approach is an elementary interplay between probability and linear algebra, which is remarkably neither analytic nor combinatorial. The outcome is that with high probability

$$\left(1 - c^- \sqrt{\frac{n}{m_n}}\right)^2 \leq \lambda_{\min}(\widehat{\Sigma}_n) \leq \lambda_{\max}(\widehat{\Sigma}_n) \leq \left(1 + c^+ \sqrt{\frac{n}{m_n}}\right)^2, \tag{1.8}$$

where the first inequality is considered only in the case where  $m_n \geq n$ . Here  $c^\pm > 0$  are constants, and thus one cannot deduce (1.7) from (1.8). The approach of Srivastava

and Vershynin is a randomization of the spectral sparsification method developed by Batson et al. [14, 15]; the idea of using rank one updates can also be found in the works of Benaych-Georges and Nadakuditi [16]. This approach requires the control of tails of norms of projections of  $X_n$ , a condition which is satisfied by log-concave distributions thanks to the thin-shell phenomenon. This condition is also satisfied if  $(X_{n,k})_{n \geq 1, 1 \leq k \leq n}$  are i.i.d. with a finite moment of order  $2 + \varepsilon$  for the lower bound on the smallest eigenvalue in (1.7) and i.i.d. with a finite moment of order  $4 + \varepsilon$  for the upper bound on the largest eigenvalue in (1.7). This method was also used recently by Yaskov [51, 52].

Our main results below lie between the original Bai–Yin theorem and the more recent work of Srivastava and Vershynin. Our contribution is to show that the non-asymptotic approach of Srivastava and Vershynin is indeed suitable to prove and extend the sharp Bai–Yin theorem, which is an asymptotic result, under fairly general assumptions on tails of norms of projections of  $X_n$ , which allow heavy tailed i.i.d. as well as log-concavity! When the coordinates of  $X_n$  are i.i.d. our approach reaches the (almost) optimal moment condition: finite second moment for the smallest eigenvalue and finite fourth moment for the largest.

Our results are based on the following tail conditions on the norm of projections.

**Definition 1.1** (*Weak Tail Projection property (WTP)*) Let  $X_n, n \in \mathbb{N}$ , be as in (1.1). We say that the *Weak Tail Projection* (WTP) property holds when the following is true:

- (a) The family  $(\langle X_n, y \rangle^2)_{n \in \mathbb{N}, y \in S^{n-1}}$  is uniformly integrable, in other words

$$\lim_{M \rightarrow \infty} \sup_{\substack{n \in \mathbb{N} \\ y \in S^{n-1}}} \mathbb{E} \left( \langle X_n, y \rangle^2 \mathbf{1}_{\{\langle X_n, y \rangle^2 \geq M\}} \right) = 0, \tag{WTP-a}$$

where  $S^{n-1} := \{y \in \mathbb{R}^n : \|y\| = 1\}$  denotes the unit sphere of  $\mathbb{R}^n$ ;

- (b) There exist two functions  $f : \mathbb{N} \rightarrow [0, 1]$  and  $g : \mathbb{N} \rightarrow \mathbb{R}_+$  such that  $f(r) \rightarrow 0$  and  $g(r) \rightarrow 0$  as  $r \rightarrow \infty$  and for every  $n \in \mathbb{N}$  and any orthogonal projection  $P : \mathbb{R}^n \rightarrow \mathbb{R}^n$  with  $P \neq 0$ ,

$$\mathbb{P} \left( \|PX_n\|^2 - \text{rank}P \geq f(\text{rank}P)\text{rank}P \right) \leq g(\text{rank}P). \tag{WTP-b}$$

This can be less formally written as  $\mathbb{P}(\|PX_n\|^2 - r \geq o(r)) \leq o(1)$  where  $r := \text{rank}P$  and where the “ $o$ ” are with respect to  $r$  and are uniform in  $n$ .

**Definition 1.2** (*Strong Tail Projection property (STP)*) Let  $X_n, n \in \mathbb{N}$ , be as in (1.1). We say the *Strong Tail Projection* (STP) property holds when there exist  $f : \mathbb{N} \rightarrow [0, 1]$  and  $g : \mathbb{N} \rightarrow \mathbb{R}_+$  such that  $f(r) \rightarrow 0$  and  $g(r) \rightarrow 0$  as  $r \rightarrow \infty$ , and for every  $n \in \mathbb{N}$ , for any orthogonal projection  $P : \mathbb{R}^n \rightarrow \mathbb{R}^n$  with  $P \neq 0$ , for any real  $t \geq f(\text{rank}P)\text{rank}P$  we have

$$\mathbb{P} \left( \|PX_n\|^2 - \text{rank}P \geq t \right) \leq \frac{g(\text{rank}P)\text{rank}P}{t^2}. \tag{STP}$$

This can be less formally written as  $\mathbb{P}(\|PX_n\|^2 - r \geq t) \leq o(r)t^{-2}$  where  $t \geq o(r)$  and  $r := \text{rank}P$  and where the “ $o$ ” are with respect to  $r$  and are uniform in  $n$ .

Note that  $\mathbb{E}(\|PX_n\|^2) = \text{rank}P$  since  $X_n$  is isotropic. The properties (WTP) and (STP) were inspired by the “strong regularity assumption” used by Srivastava and Vershynin [48]. They are specially designed to obtain a sharp Bai–Yin type asymptotic result in the i.i.d. case with (almost) optimal moment assumptions, as well as in the log-concave case.

It is easy to see that (STP) implies that for any  $\varepsilon > 0$ , the  $4 - \varepsilon$  moments of 1-dimensional marginals of  $X_n$ ’s are uniformly bounded; in particular, in this case the vectors  $(X_n)_{n \in \mathbb{N}}$  satisfy condition (WTP-a). Next, condition (WTP-b) is clearly weaker than (STP); thus, (STP) implies (WTP), hence the qualifiers “Strong” and “Weak” for these properties.

Proposition 1.3 below, proved in Sect. 2, implies that if  $(X_n)_{n \in \mathbb{N}}$  is as in (1.1) and if  $X_n$  is log-concave for every  $n \in \mathbb{N}$  then properties (WTP) and (STP) are satisfied. It is a consequence of the thin-shell and sub-exponential tails phenomena for these distributions.

**Proposition 1.3** (Log-concave random vectors) *Let  $X_n, n \in \mathbb{N}$ , be as in (1.1), and suppose that the centered isotropic random vector  $X_n$  is log-concave for any  $n \in \mathbb{N}$ . Then a stronger form of (STP) holds with*

$$\{\|PX_n\|^2 - \text{rank}P \geq t\} \text{ replaced by } \{|\|PX_n\|^2 - \text{rank}P| \geq t\}.$$

The next proposition, proved in Sect. 2, implies that if  $(X_{n,k})_{n \geq 1, 1 \leq k \leq n}$  are i.i.d. then property (WTP) holds, and if moreover these i.i.d. random variables have finite fourth moment then property (STP) holds.

**Proposition 1.4** (Random vectors with i.i.d. coordinates) *Let  $X_n, n \in \mathbb{N}$ , be as in (1.1), and suppose that the coordinates  $(X_{n,k})_{n \geq 1, 1 \leq k \leq n}$  are i.i.d. distributed as a real random variable  $\xi$  with zero mean and unit variance. Then, denoting  $r := \text{rank}P$ ,*

- a stronger version of (WTP) holds, with, in (WTP-b),

$$\{\|PX_n\|^2 - r \geq f(r)r\} \text{ replaced by } \{|\|PX_n\|^2 - r| \geq f(r)r\};$$

- if moreover  $\mathbb{E}(\xi^4) < \infty$  then a stronger version of (STP) holds, with

$$\{\|PX_n\|^2 - r \geq t\} \text{ replaced by } \{|\|PX_n\|^2 - r| \geq t\}.$$

The next proposition, proved in Sect. 2, was suggested by an anonymous referee.

**Proposition 1.5** (Marchenko–Pastur law holds under (WTP)) *Let  $X_n, n \in \mathbb{N}$ , be as in (1.1). If  $\lim_{n \rightarrow \infty} \frac{n}{m_n} = \rho$  with  $\rho \in (0, 1)$ , in other words  $\underline{\rho} = \bar{\rho} \in (0, 1)$ , and if (WTP-a) and (WTP-b) are satisfied, then the Marchenko–Pastur law (1.5) holds.*

Our main results are stated in the following couple of theorems and corollaries.

**Theorem 1.6** (Smallest eigenvalue) *Let  $X_n, m_n,$  and  $A_n, n \in \mathbb{N},$  be as in (1.1), (1.2), and (1.4) respectively. If  $\bar{\rho} < 1$  (in particular  $m_n > n$  for  $n \gg 1$ ), and if (WTP-a) and (WTP-b) are satisfied then*

$$\liminf_{n \rightarrow \infty} \frac{\mathbb{E}(\lambda_{\min}(A_n))}{(\sqrt{m_n} - \sqrt{n})^2} \geq 1.$$

*If additionally  $\lim_{n \rightarrow \infty} \frac{n}{m_n} = \rho$  with  $\rho \in (0, 1)$ , in other words  $\underline{\rho} = \bar{\rho} \in (0, 1)$ , then*

$$\lambda_{\min}(\widehat{\Sigma}_n) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} (1 - \sqrt{\rho})^2.$$

Theorem 1.6 is proved in Sect. 3. The second part, which is the Bai–Yin edge convergence (1.7) of the smallest eigenvalue in probability, is obtained by combining the first part of the theorem with the Marchenko–Pastur bound (1.6) for the smallest eigenvalue obtained from the Marchenko–Pastur law (1.5) provided by Proposition 1.5.

Combining Theorem 1.6 with Propositions 1.3 and 1.4, we obtain the following corollary.

**Corollary 1.7** (Smallest eigenvalue convergence) *Let  $X_n, m_n, \widehat{\Sigma}_n,$  and  $A_n, n \in \mathbb{N},$  be as in (1.1), (1.2), (1.3), and (1.4) respectively. If  $\bar{\rho} < 1$  (in particular  $m_n > n$  for  $n \gg 1$ ) and if the centered isotropic random vector  $X_n$  is log-concave for every  $n \in \mathbb{N}$  or if  $(X_{n,k})_{n \geq 1, 1 \leq k \leq n}$  are i.i.d. then*

$$\liminf_{n \rightarrow \infty} \frac{\mathbb{E}(\lambda_{\min}(A_n))}{(\sqrt{m_n} - \sqrt{n})^2} \geq 1.$$

*If additionally  $\lim_{n \rightarrow \infty} \frac{n}{m_n} = \rho$  with  $\rho \in (0, 1)$ , in other words  $\underline{\rho} = \bar{\rho} \in (0, 1)$ , then*

$$\lambda_{\min}(\widehat{\Sigma}_n) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} (1 - \sqrt{\rho})^2.$$

**Theorem 1.8** (Largest eigenvalue) *Let  $X_n, m_n,$  and  $A_n, n \in \mathbb{N},$  be as in (1.1), (1.2), and (1.4) respectively. If (STP) holds then*

$$\limsup_{n \rightarrow \infty} \frac{\mathbb{E}(\lambda_{\max}(A))}{(\sqrt{m_n} + \sqrt{n})^2} \leq 1.$$

*If additionally  $\lim_{n \rightarrow \infty} \frac{n}{m_n} = \rho$  with  $\rho \in (0, \infty)$  in other words  $\underline{\rho} = \bar{\rho} \in (0, \infty)$ , then*

$$\lambda_{\max}(\widehat{\Sigma}_n) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} (1 + \sqrt{\rho})^2.$$

Theorem 1.8 is proved in Sect. 4. Again, the second part uses the Marchenko–Pastur bound (1.6) obtained from the Marchenko–Pastur law (1.5) provided by Proposition 1.5.

Combining Theorem 1.8 with Propositions 1.3 and 1.4, we obtain the following statement:

**Corollary 1.9** (Largest eigenvalue convergence) *Let  $X_n, m_n, \widehat{\Sigma}_n$  and  $A_n, n \in \mathbb{N}$ , be as in (1.1), (1.2), (1.3), and (1.4) respectively. If the centered isotropic random vector  $X_n$  is log-concave for every  $n \in \mathbb{N}$  or if  $(X_{n,k})_{n \geq 1, 1 \leq k \leq n}$  are i.i.d. with finite 4th moment then*

$$\limsup_{n \rightarrow \infty} \frac{\mathbb{E}(\lambda_{\max}(A_n))}{(\sqrt{m_n} + \sqrt{n})^2} \leq 1.$$

If additionally  $\lim_{n \rightarrow \infty} \frac{n}{m_n} = \rho$  with  $\rho \in (0, \infty)$  in other words  $\underline{\rho} = \bar{\rho} \in (0, \infty)$ , then

$$\lambda_{\max}(\widehat{\Sigma}_n) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} (1 + \sqrt{\rho})^2.$$

### 1.1 Outline of the argument and novelty

Assume we are given a random matrix

$$A = \sum_{k=1}^m X^{(k)} \otimes X^{(k)},$$

where  $X^{(k)}$  are i.i.d. isotropic random vectors. As we already mentioned above, the key ingredient in estimating the extremal eigenvalues of  $A$  is the following rank one update formula known as the Sherman–Morrison formula:

$$(M + xx^\top)^{-1} = M^{-1} - \frac{M^{-1}xx^\top M^{-1}}{1 + x^\top M^{-1}x},$$

which is valid for any non-singular  $n \times n$  matrix  $M$  and a vector  $x$  with  $1 + x^\top M^{-1}x \neq 0$ . Using the above identity and assuming that  $M$  is symmetric, the restriction on  $\mathbb{R}$  of the Cauchy–Stieltjes transform of the spectral distribution of  $M + xx^\top$ , which is defined as an appropriately scaled trace of  $(u - M - xx^\top)^{-1}$ ,  $u \in \mathbb{R}$ , can be in a simple way expressed in terms of the Cauchy–Stieltjes transform of  $M$ . To be more precise, setting

$$A^{(0)} := 0 \quad \text{and} \quad A^{(k)} := A^{(k-1)} + X^{(k)} \otimes X^{(k)}, \quad k = 1, \dots, m,$$

we get, for any  $k = 1, \dots, m$  and any  $u \in \mathbb{R}$ ,

$$\text{tr}(u - A^{(k)})^{-1} = \text{tr}(u - A^{(k-1)})^{-1} + \frac{X^{(k)\top}(u - A^{(k-1)})^{-2}X^{(k)}}{1 - X^{(k)\top}(u - A^{(k-1)})^{-1}X^{(k)}}. \tag{1.9}$$

A crucial observation, made already in [15] and further developed in [48] is that the Cauchy–Stieltjes transform on the real line can be efficiently used to control the



extreme eigenvalues of a positive semi-definite matrix. In our setting, starting with some specially chosen *non-random* number  $u_0 \neq 0$ , we inductively define a sequence of *random* numbers  $u_k$  ( $k \leq m$ ) in such a way that all  $u_k$ 's stay on the same side of the spectra of  $A^{(k)}$ 's ( $k \leq m$ ), at the same time not departing too far from the spectrum. Then the expectation of the corresponding extreme eigenvalue of  $A = A^{(m)}$  can be estimated by  $\mathbb{E}u_m$ . The increments  $u_k - u_{k-1}$  are defined with help of (1.9) as certain functions of  $A^{(k-1)}$ ,  $u_{k-1}$  and  $X^{(k)}$ , and their expectations are controlled using the information about the distribution of  $X^{(k)}$  as well as certain induction hypotheses. At this level, the approach used in the present paper is similar to [48].

On the other hand, as our result is asymptotically sharp and covers the i.i.d. case with almost optimal moment conditions, the technical part of our argument differs significantly from [48]. In particular, we introduce the “regularity shifts”, which are designed in such a way that  $u_k$ 's stay “sufficiently far” from the spectrum of  $A^{(k)}$ 's, which guarantees validity of certain concentration inequalities, whereas at the same time not departing “too far” so that one still gets a satisfactory estimate of the expectation of the spectral edges. The shifts (which we denote by  $\delta_R^k$  and  $\Delta_R^k$ ) are defined in Sects. 3.3 and 4.3 (see, in particular, Lemmas 3.5 and 4.7).

Let us emphasize once more that the proofs we obtain are simpler and shorter than the original combinatorial approach of Bai–Yin and the analytic approach based on the resolvent (more precisely the Cauchy–Stieltjes transform on the complex plane outside the real axis), while the class of distributions covered in our paper is larger. In particular, Theorem 1.6 of the present paper essentially recovers a recent result [49]. It must be noted, however, that in our paper we replace convergence almost surely with the weaker convergence in probability.

## 1.2 Discussion and extensions

In this note we restrict our analysis to random vectors with real coordinates because we think that this is simply more adapted to geometric dependence such as log-concavity. It is likely that the method remains valid for random vectors with complex entries. The Bai–Yin theorem is also available for random symmetric matrices (which are the sum of rank two updates which are no longer positive semidefinite) but it is unclear to us if one can adapt the method to this situation. One can ask in another direction if the method remains usable for non-white population covariance matrices, and for the so-called information plus noise covariance matrices, two models studied at the edge by Bai and Silverstein [8, 10] among others. One can ask finally if the method allows to extract at least the scaling of the Tracy–Widom fluctuation at the edge. For the Tracy–Widom fluctuation at the edge of empirical covariance matrices we refer to Johansson [29], Johnstone [30], Borodin and Forrester [19], Soshnikov [47], Péché [39], Feldheim and Sodin [23], Pillai and Yin [41], and references therein. It was shown by Lee and Yin [33] that for centered Wigner matrices the finiteness of the fourth moment is more than enough for the Tracy–Widom fluctuation of the largest eigenvalue. One can ask the same for the largest eigenvalue of the empirical covariance matrix of random vectors with i.i.d. entries, and one can additionally ask if a finite second moment is enough for the Tracy–Widom fluctuation of the smallest eigenvalue.

### 1.3 Structure

The rest of the article is structured as follows. Section 2 provides the proof of Propositions 1.3 and 1.4. In Sects. 3 and 4 we prove Theorems 1.6 and 1.8, respectively.

### 1.4 Notations

We set  $\|v\| := \sqrt{v_1^2 + \dots + v_n^2}$  and  $\|v\|_\infty := \max(|v_1|, \dots, |v_n|)$  for any vector  $v \in \mathbb{R}^n$ . We denote  $\|f\|_\infty := \sup_{x \in S} |f(x)|$  for any function  $f : S \rightarrow \mathbb{R}$ . We often use the notation  $|S| := \text{card}(S)$  for a set  $S$ . Further, we denote by

$$\lambda_{\max}(M) := \lambda_1(M) \geq \dots \geq \lambda_n(M) =: \lambda_{\min}(M)$$

the eigenvalues of a symmetric  $n \times n$  matrix  $M \in \mathcal{M}_n(\mathbb{R})$ . We denote by  $I_n$  the  $n \times n$  identity matrix. For any real number  $u$  we sometimes abridge  $uI_n$  into  $u$  in matrix expressions such as in  $M - uI_n = M - u$ .

## 2 Proof of Propositions 1.3–1.5

*Proof of Proposition 1.3* Assume  $X$  is a centered isotropic log-concave vector in  $\mathbb{R}^n$  and  $P : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is a non-zero orthogonal projection. The random vector  $PX$  is log-concave with mean zero and covariance matrix  $PP^T = P^2 = P$ . Restricted to the image of  $P$ , the vector  $Y = PX$  is log-concave with covariance matrix  $I_r$  where  $r = \text{rank} P$ . The so-called thin-shell phenomenon [7] states that “most” of the distribution of  $Y$  is supported in a thin-shell around the Euclidean sphere of radius  $\sqrt{r}$ . Quantitative estimates have been obtained notably by Klartag [31], Fleury [25], Guédon and Milman [27], see also Guéon [26]. On the other hand, it is also known that the tail of the norm of  $Y$  is sub-exponential, see Paouris [37], and also Admaczak et al. [4]. The following inequality, taken from [27, Theorem 1.1], captures both phenomena: there exist absolute constants  $c, C \in (0, \infty)$  such that for any real  $u \geq 0$ ,

$$\mathbb{P}(|\|Y\| - \sqrt{r}| \geq u\sqrt{r}) \leq C \exp\left(-c\sqrt{r} \min(u, u^3)\right).$$

This is more than enough for our needs. Namely, let  $\beta \in (0, 1/20)$ , and let  $u = u(r) \in (0, \infty)$  and  $\alpha = \alpha(r) \in (0, 1)$  be such that  $\alpha \geq (1+r)^{-\beta}$  and  $u \geq \max((1+r)^{-\beta}, 2\alpha/(1-\alpha^2))$ . Note that  $2\alpha/(1-\alpha^2) \in (0, 1)$  when  $\alpha \in (0, \sqrt{2}-1)$ , and that  $2\alpha/(1-\alpha^2) \rightarrow 0$  as  $\alpha \rightarrow 0$ . Now, using the inequality  $\exp(-2t) \leq t^{-4}$  for  $t > 0$ , we get, if  $\alpha u \in (0, 1]$ ,

$$\begin{aligned} \mathbb{P}(\|Y\|^2 - r \geq u^2r) &= \mathbb{P}(\|Y\| \geq \sqrt{r + u^2r}) \\ &\leq \mathbb{P}(\|Y\| \geq \sqrt{r} + \alpha u\sqrt{r}) \\ &\leq C \exp\left(-c\sqrt{r}(\alpha u)^3\right) \\ &\leq \frac{2^4 C}{c^4(u^2r)^2 \alpha^{12} u^8} = \frac{o(r)}{(u^2r)^2}, \end{aligned}$$

while if  $\alpha u \in [1, \infty)$  we get

$$\mathbb{P}(\|Y\|^2 - r \geq u^2 r) \leq \dots \leq C \exp(-c\sqrt{r}\alpha u) \leq \frac{2^4 C}{c^4(u^2 r)^2 \alpha^4} = \frac{o(r)}{(u^2 r)^2}.$$

Similarly, for an arbitrary  $\beta \in (0, 1/20)$ , let  $u = u(r) \in (0, \infty)$  and  $\alpha = \alpha(r) \in (0, 1)$  be such that  $\alpha \geq (1 + r)^{-\beta}$  and  $u \geq \max((1 + r)^{-\beta}, 2\alpha/(1 + \alpha^2))$ . If  $u > 1$  then necessarily

$$\mathbb{P}(\|Y\|^2 - r \leq -u^2 r) = 0.$$

Otherwise,  $\alpha u \in (0, 1]$  and using again the inequality  $\exp(-2t) \leq t^{-4}$  for  $t > 0$ , we get

$$\begin{aligned} \mathbb{P}(\|Y\|^2 - r \leq -u^2 r) &\leq \mathbb{P}(\|Y\| \leq \sqrt{r - u^2 r}) \\ &\leq \mathbb{P}(\|Y\| \leq \sqrt{r} - \alpha u \sqrt{r}) \\ &\leq C \exp(-c\sqrt{r}(\alpha u)^3) \\ &\leq \frac{2^4 C}{c^4(u^2 r)^2 \alpha^{12} u^8} = \frac{o(r)}{(u^2 r)^2}. \end{aligned}$$

Thus, we obtain  $\mathbb{P}(|\|Y\|^2 - r| \geq t) \leq o(r)t^{-2}$  for  $t \geq o(r)$  as expected. □

*Proof of Proposition 1.4*

- *Proof of the first part (uniform integrability (WTP-a))* Recall that we are given a random variable  $\xi$  with zero mean and unit variance and that for every  $n \in \mathbb{N}$  the coordinates of  $X_n$  are independent copies of  $\xi$ . We want to show that

$$\lim_{M \rightarrow \infty} \sup_{\substack{n \in \mathbb{N} \\ y \in S^{n-1}}} \mathbb{E} \left( \langle X_n, y \rangle^2 \mathbf{1}_{\{\langle X_n, y \rangle^2 \geq M\}} \right) = 0.$$

For every  $x \in \mathbb{R}^n$ , we define  $f_n(x) := \langle X_n, x \rangle$ . Clearly,  $\mathbb{E}(f_n^2(x)) = \|x\|^2$  since  $X_n$  is isotropic. Let us start with some comments to understand the problem. The random variables  $(f_n^2(y))_{n \in \mathbb{N}, y \in S^{n-1}}$  belong to the unit sphere of  $L^1$ . If  $\xi$  has finite fourth moment  $B := \mathbb{E}(\xi^4) < \infty$  then by expanding and using isotropy we get  $\mathbb{E}(f_n^4(y)) \leq \max(B, 3)$  which shows that the family  $(f_n^2(y))_{n \in \mathbb{N}, y \in S^{n-1}}$  is bounded in  $L^2$  and thus uniformly integrable. How to proceed when  $\xi$  does not have a finite fourth moment? If  $y$  belongs to the canonical basis of  $\mathbb{R}^n$  then  $f_n(y)$  is distributed as  $\xi$  and has thus the same integrability. On the other hand, if  $y$  is far from being sparse, say  $y = (n^{-1/2}, \dots, n^{-1/2})$ , then  $f_n(y)$  is distributed as  $n^{-1/2}(\xi_1 + \dots + \xi_n)$  where the  $\xi_i$ 's are independent copies of  $\xi$ , which is close in distribution to the standard Gaussian law  $\mathcal{N}(0, 1)$  by the Central Limit Theorem (CLT).

We will use the following uniform quantitative CLT. Even though it probably exists somewhere in the literature, we provide a short proof for convenience. It can probably

also be proved using the classical Fourier analytic smoothing inequality [24, equation (3.13) XVI.3 p. 538] which is the basis of many uniform CLT estimates.

**Lemma 2.1** (Uniform quantitative CLT) *Let  $\xi$  be a random variable with zero mean and unit variance and let  $\Phi$  be the cumulative distribution of the standard real Gaussian law  $\mathcal{N}(0, 1)$  of zero mean and unit variance. For any  $\varepsilon > 0$  there exists  $\delta > 0$  depending only on  $\varepsilon$  and the law of  $\xi$  with the following property: if  $n \in \mathbb{N}$  and  $y \in S^{n-1}$  is such that  $\|y\|_\infty \leq \delta$ , then the cumulative distribution function  $F_n$  of  $\sum_{i=1}^n y_i \xi_i$  satisfies*

$$\|F_n - \Phi\|_\infty \leq \varepsilon.$$

*Proof of Lemma 2.1* To prove the lemma, let us assume the contrary. Then there exists  $\varepsilon > 0$  and a sequence  $(y_m)_{m \geq 1}$  in  $\ell^2(\mathbb{N})$  such that  $\|y_m\| = 1$  and  $\|y_m\|_\infty \leq 1/m$  for every  $m \in \mathbb{N}$ , and such that  $\|F_m - \Phi\|_\infty > \varepsilon$  where  $F_m$  is the cumulative distribution function of  $S_m := \sum_{i=1}^\infty y_{m,i} \xi_i$ . Let  $\varphi_m$  be the characteristic function of  $S_m$ . We have  $\varphi_m(t) = \prod_{i=1}^\infty \varphi(y_{m,i}t)$  where  $\varphi$  is the characteristic function of  $\xi$ . Fix any  $t \in \mathbb{R}$ . By assumption on  $\xi$ , we get  $\varphi(t) = 1 - \frac{t^2}{2} + o(t^2)$ . Hence, using the identities  $\|y_m\|^2 := \sum_{i=1}^\infty y_{m,i}^2 = 1$  and  $\|y_m\|_\infty := \max_{i \in \mathbb{N}} |y_{m,i}| \leq 1/m$  together with the formula (valid for  $m \rightarrow \infty$ ):

$$\begin{aligned} \varphi_m(t) &= \prod_{i=1}^\infty \left( 1 - \frac{y_{m,i}^2 t^2}{2} + o\left(\left(y_{m,i}^2 t^2\right)\right) \right) \\ &= \exp\left(-\frac{t^2}{2} \sum_{i=1}^\infty y_{m,i}^2 + \sum_{i=1}^\infty o\left(y_{m,i}^2 t^2\right)\right), \end{aligned}$$

we get  $\lim_{m \rightarrow \infty} \varphi_m(t) = e^{-t^2/2}$ . By the Lévy theorem for characteristic functions, it follows that  $F_m \rightarrow \Phi$  pointwise as  $m \rightarrow \infty$ , which yields  $S_m \rightarrow \mathcal{N}(0, 1)$  weakly as  $m \rightarrow \infty$ , contradicting to our initial assumption. □

Let us continue with the proof of the uniform integrability. Since  $\xi^2 \in L^1$  we get, by dominated convergence,

$$h(M) := \mathbb{E}\left(\xi^2 \mathbf{1}_{\{\xi^2 \geq M\}}\right) \xrightarrow{M \rightarrow \infty} 0.$$

Let  $\varepsilon > 0$ , and let  $\delta > 0$  be defined from  $\varepsilon$  and the law of  $\xi$  as in the above lemma (we can, of course, assume that  $\delta \rightarrow 0$  with  $\varepsilon \rightarrow 0$ ). Let  $M > 0$  and  $n \in \mathbb{N}$  and  $y \in S^{n-1}$ .

Let us write  $y = w + z$  where  $w_i := y_i \mathbf{1}_{|y_i| \leq \delta^2}$  and  $z_i := y_i \mathbf{1}_{|y_i| > \delta^2}$  for any  $i = 1, \dots, n$  (here, one can recall decomposition of the unit sphere into sets of “close to sparse” and “far from sparse” vectors, employed, in particular, in [34, 43]). Then it is easily seen that

- $\text{supp}(z) \cap \text{supp}(w) = \emptyset$ ;
- $\|w\| \leq 1$  and  $\|w\|_\infty \leq \delta^2$ ;
- $\|z\| \leq 1$  and  $|\text{supp}(z)| \leq 1/\delta^4$ ,

where  $\text{supp}(x) := \{i : x_i \neq 0\}$ . Now we have  $f_n^2(y) \leq 2(f_n^2(w) + f_n^2(z))$  and

$$\begin{aligned} \mathbb{E} \left( f_n^2(y) \mathbf{1}_{\{f_n^2(y) \geq M\}} \right) &\leq 2\mathbb{E} \left( \left( f_n^2(w) + f_n^2(z) \right) \left( \mathbf{1}_{\{f_n^2(w) \geq \frac{M}{4}\}} + \mathbf{1}_{\{f_n^2(z) \geq \frac{M}{4}\}} \right) \right) \\ &\leq 2\|z\|^2 \mathbb{P} \left( f_n^2(w) \geq \frac{1}{4}M \right) + 2\|w\|^2 \mathbb{P} \left( f_n^2(z) \geq \frac{1}{4}M \right) \\ &\quad + 2\mathbb{E} \left( f_n^2(z) \mathbf{1}_{\{f_n^2(z) \geq \frac{M}{4}\}} \right) + 2\mathbb{E} \left( f_n^2(w) \mathbf{1}_{\{f_n^2(w) \geq \frac{M}{4}\}} \right) \\ &=: (*) + (**) + (***) + (****). \end{aligned}$$

Now, by Markov’s inequality

$$(*) + (**) \leq \frac{16}{M}.$$

Second, using that  $f_n^2(x) \leq \|P_x X_n\|^2$ , valid for any  $x \in \mathbb{R}^n$  with  $\|x\| \leq 1$  and orthogonal projection  $P_x$  onto its support, we obtain

$$\begin{aligned} (***) &\leq 2\mathbb{E} \left( \left( \sum_{i:z_i \neq 0} \xi_i^2 \right) \sum_{i:z_i \neq 0} \mathbf{1}_{\{\xi_i^2 \geq \frac{\delta^8 M}{4}\}} \right) \\ &= 2\mathbb{E} \left( \sum_{i,j \in \text{supp}(z)} \xi_i^2 \mathbf{1}_{\{\xi_j^2 \geq \frac{\delta^8 M}{4}\}} \right) \\ &\leq 2 \left( \frac{1}{\delta^8} \frac{4}{\delta^8 M} + \frac{h \left( \frac{\delta^8 M}{4} \right)}{\delta^4} \right) = \frac{8}{\delta^{16} M} + \frac{2h \left( \frac{\delta^8 M}{4} \right)}{\delta^4}, \end{aligned}$$

where in the third line we used Markov’s inequality. Third, we write

$$(****) = 2\mathbb{E} \left( f_n^2(w) \right) - 2\mathbb{E} \left( f_n^2(w) \mathbf{1}_{\{f_n^2(w) < \frac{M}{4}\}} \right).$$

Now if  $\|w\| \leq \delta$  then  $(***) \leq 2\delta^2$ . Suppose in contrast that  $\|w\| > \delta$ , and denote  $w_* := w/\|w\|$ . Then  $\|w_*\|_\infty \leq \delta$ , and therefore, by the CLT of Lemma 2.1, the distribution of  $f_n(w_*)$  is  $\varepsilon$ -close to  $\mathcal{N}(0, 1)$ , and in particular, there exist  $M_*(\varepsilon) > 0$  and  $\rho(\varepsilon) > 0$  depending *only* on  $\varepsilon$  such that  $\rho(\varepsilon) \rightarrow 0$  as  $\varepsilon \rightarrow 0$  and

$$\mathbb{E} \left( f_n^2(w_*) \mathbf{1}_{\{f_n^2(w_*) < M_*(\varepsilon)\}} \right) \geq 1 - \rho(\varepsilon).$$

Thus, having in mind that  $f_n^2(w_*) = f_n^2(w)/\|w\|^2$ , we get

$$(***) \leq 2\|w\|^2 \rho(\varepsilon) \leq 2\rho(\varepsilon)$$

provided that  $M \geq 4M_*(\varepsilon)$ . So, in all cases, as long as  $M$  is sufficiently large,

$$(***) \leq 2\delta^2 + 2\rho(\varepsilon).$$

Finally, take any  $M > 0$  such that  $M \geq \max(16/\varepsilon, 8/(\delta^{16}\varepsilon), 4M_*(\varepsilon))$  and such that  $2h(\delta^8 M/4)/\delta^{14} \leq \varepsilon$ ; then the desired result follows from

$$(*) + (**) + (***) + (****) \leq 4\varepsilon + 2\delta^2 + 2\rho(\varepsilon).$$

*Remark 2.2* (Alternative proof of (WTP-a)) An anonymous referee has suggested the following alternative proof of (WTP-a), which is shorter, but relies on a non-elementary inequality. For any  $n \geq 1$  and any  $y = (y_1, \dots, y_n) \in S^{n-1}$ , Petrov’s inequality [40, §2.6.26 p. 83]—a consequence of Fuk and Nagaev’s inequality [40, (2.33) p. 61]—yields:

$$\mathbb{E}(w(\langle X_n, y \rangle)) \leq \sum_{k=1}^n \mathbb{E}(w(2y_k X_{n,k})) + 2e^2 \int_0^\infty \frac{dw(x)}{(1 + \frac{1}{2}x^2)^2}$$

for any even function  $w : \mathbb{R} \rightarrow \mathbb{R}_+$  which is non-decreasing on  $\mathbb{R}_+$  with  $w(0) = 0$ . Now take  $w(x) := x^2 \mathbf{1}_{\{x^2 > M\}}$  where  $M > 0$  is fixed. Then, on the one hand, for any  $1 \leq k \leq n$ , as  $y_k \in [0, 1]$ , we have, denoting  $X_{n,k}$  the  $k$ -th component of  $X_n$ ,

$$\mathbb{E}(w(2y_k X_{n,k})) = 4y_k^2 \mathbb{E}\left(X_{n,k}^2 \mathbf{1}_{\{(2y_k X_{n,k})^2 > M\}}\right) \leq 4y_k^2 \mathbb{E}\left(X_{n,k}^2 \mathbf{1}_{\{(2X_{n,k})^2 > M\}}\right)$$

while on the other hand,

$$\int_0^\infty \frac{dw(x)}{(1 + \frac{1}{2}x^2)^2} = \frac{M}{(1 + \frac{1}{2}M)^2} + \int_{\sqrt{M}}^\infty \frac{dx^2}{(1 + \frac{1}{2}x^2)^2} \leq \frac{4}{M} + \frac{4}{M + 2} \leq \frac{8}{M},$$

and we obtain the following inequality:

$$\mathbb{E}\left(\langle X_n, y \rangle^2 \mathbf{1}_{\{\langle X_n, y \rangle^2 > M\}}\right) \leq 4 \max_{1 \leq k \leq n} \mathbb{E}\left(X_{n,k}^2 \mathbf{1}_{\{X_{n,k}^2 > \frac{M}{4}\}}\right) + \frac{8}{M}.$$

This implies that the family  $\{\langle X_n, y \rangle^2 : n \geq 1, y \in S^{n-1}\}$  is uniformly integrable.

- *Proof of the first part (improved (WTP-b))* As before, we assume that  $\xi$  is a random variable with zero mean and unit variance, and denote by  $(\xi_i)$  a sequence of i.i.d. copies of  $\xi$ . Let us recall a kind of the weak law of large numbers for weighted sums, taken from [49, Lemma 5], which can be seen as a consequence of Lévy’s continuity theorem for characteristic functions: if  $(\eta_i)_{i \in I}$  is a sequence (finite or infinite) of i.i.d. real random variables with zero mean then for every  $\varepsilon > 0$  there exists  $\delta > 0$  which may depend on the law of the random variables with the

following property: for every deterministic sequence  $(t_i)_{i \in I}$  in  $[0, \infty)$  such that  $\|t\|_1 := \sum_{i \in I} t_i = 1$  and  $\|t\|_\infty := \max_{i \in I} t_i \leq \delta$ , we have

$$\mathbb{P} \left( \left| \sum_{i \in I} t_i \eta_i \right| > \varepsilon \right) < \varepsilon.$$

Setting  $\eta_i := \xi_i^2 - 1$ , it follows that there exists  $h : (0, 1] \rightarrow (0, 1]$  such that, given any  $\varepsilon > 0$  and a sequence  $(t_i)_{i \in I}$  in  $[0, \infty)$  with  $\|t\|_1 = 1$  and  $\|t\|_\infty \leq h(\varepsilon)$ , we have

$$\mathbb{P} \left( \left| \sum_{i \in I} t_i \xi_i^2 - 1 \right| > \varepsilon \right) \leq \varepsilon. \tag{2.1}$$

Without loss of generality, one can take  $h$  strictly monotone (i.e. invertible).

We now proceed similarly to [48, Proposition 1.3]. Fix  $n \in \mathbb{N}$  and let  $P$  be a non-zero orthogonal projection of  $\mathbb{R}^n$  of rank  $r$ . Let  $X = (\xi_k)_{1 \leq k \leq n}$  be a random vector of  $\mathbb{R}^n$  with  $\xi_1, \dots, \xi_n$  i.i.d. copies of  $\xi$ . We have

$$\|PX\|^2 = \langle X, PX \rangle.$$

Let us also denote the matrix of  $P$  in the canonical basis as  $P$ . We have  $P^2 = P = P^\top$  and  $\text{tr}(P) = \text{rank}(P) = r$ . Let  $P_0$  be the matrix obtained from  $P$  by zeroing the diagonal. We have  $P - P_0 = \text{diag}(P)$ . A standard decoupling inequality proved in [50] (see also the book [22]) states that for any convex function  $F$ ,

$$\mathbb{E}(F(\langle X, P_0 X \rangle)) \leq \mathbb{E}(F(4\langle X, P_0 X' \rangle)), \tag{2.2}$$

where  $X'$  is an independent copy of  $X$ . In particular the choice  $F(u) = u^2$  gives

$$\mathbb{E}(\langle X, P_0 X \rangle^2) \leq 16 \mathbb{E}(\langle X, P_0 X' \rangle^2).$$

Now recall that if  $Z$  is a centered random vector of  $\mathbb{R}^n$  with covariance matrix  $\Gamma$  and if  $B$  is a  $n \times n$  matrix then  $\mathbb{E}(\langle Z, BZ \rangle) = \text{tr}(\Gamma B^\top)$ . Seeing  $X$  and  $X'$  as column vectors,

$$\begin{aligned} \mathbb{E}(\langle X, P_0 X' \rangle^2) &= \mathbb{E} \left( X'^\top P_0 X X^\top P_0 X' \right) \\ &= \mathbb{E} \left( X'^\top P_0 \mathbb{E}(X X^\top) P_0 X' \right) \\ &= \mathbb{E} \left( X'^\top P_0^2 X' \right) \\ &= \text{tr} \left( P_0^2 \right) \\ &\leq 2\text{tr} \left( (P - P_0)^2 \right) + 2\text{tr}(P^2) \\ &\leq 2\text{tr}(P^2) + 2\text{tr}(P^2) \\ &= 4r. \end{aligned}$$

Therefore

$$\mathbb{E}(\langle X, P_0 X \rangle^2) \leq 64r$$

and by Markov’s inequality we get

$$\mathbb{P} \left( |\langle X, P_0 X \rangle| > r^{3/4} \right) \leq \frac{64}{\sqrt{r}}. \tag{2.3}$$

Next note that

$$\langle X, (P - P_0) X \rangle = \sum_{i=1}^n P_{i,i} \xi_i^2$$

with  $0 \leq P_{i,i} \leq 1$  and  $\sum_{i=1}^n P_{i,i} = r$ . Hence taking  $t_i = P_{i,i}/r$  and  $\varepsilon := h^{-1}(1/r)$  with  $\varepsilon := 1$  if  $1/r$  is outside of the range of  $h$ , we get, using (2.1),

$$\mathbb{P}(|\langle X, (P - P_0) X \rangle - r| > \varepsilon r) \leq \varepsilon.$$

Finally, by combining with (2.3) we obtain

$$\mathbb{P}(|\langle X, P X \rangle - r| > (\varepsilon + r^{-1/4}) r) \leq \varepsilon + \frac{64}{\sqrt{r}}$$

and this implies the desired result, namely

$$\mathbb{P}(|\langle X, P X \rangle - r| > o(r)) = o(1).$$

- *Proof of the second part (improved (STP))* Let  $\xi$  be a random variable with zero mean, unit variance and a finite fourth moment. Further, let  $P = (P_{ij})$ ,  $P_0$  and  $r$  have the same meaning as before, and  $X = (\xi_k)_{1 \leq k \leq n}$ , where  $\xi_k$ ’s are i.i.d. copies of  $\xi$ . For any  $u > 0$  we have

$$\begin{aligned} \mathbb{P} \left( \left| \|P X\|^2 - r \right| > u \right) &\leq \mathbb{P}(|\langle X, P_0 X \rangle| > u/2) + \mathbb{P}(|\langle X, (P - P_0) X \rangle - r| > u/2) \\ &=: (*) + (**). \end{aligned}$$

Let us estimate the last two quantities (\*) and (\*\*) separately. In both cases we will compute selected moments and use Markov’s inequality. Set  $B := \mathbb{E}(\xi_k^4)$ . Note that  $B \geq 1$  since  $\mathbb{E}\xi^2 = 1$ . Since  $\xi_k$ ’s are independent, we get, for any unit vector  $y = (y_k)_{1 \leq k \leq n}$ ,

$$\mathbb{E}(\langle X, y \rangle^4) = \sum_i \mathbb{E}(\xi_i^4) y_i^2 y_i^2 + 3 \sum_{i \neq j} \mathbb{E}(\xi_i^2) \mathbb{E}(\xi_j^2) y_i^2 y_j^2 \leq \max(B, 3). \tag{2.4}$$

The decoupling inequality (2.2) with  $F(u) = u^4$  gives

$$\mathbb{E}(\langle X, P_0 X \rangle^4) \leq 256 \mathbb{E}(\langle X, P_0 X' \rangle^4),$$



where  $X'$  is an independent copy of  $X$ . Next, in view of (2.4), we get

$$\mathbb{E}(\langle X, P_0 X' \rangle^4) \leq \mathbb{E}(\|P_0 X'\|^4) \max_{\|y\|=1} \mathbb{E}(\langle X, y \rangle^4) \leq \max(B, 3)\mathbb{E}(\|P_0 X'\|^4).$$

Since  $X'$  has independent coordinates of zero mean and unit variance, we get

$$\begin{aligned} \mathbb{E}(\|P_0 X'\|^4) &\leq 8\mathbb{E}(\|P X'\|^4) + 8\mathbb{E}(\|(P - P_0)X'\|^4) \\ &= 8\mathbb{E}\left(\sum_{i,j} P_{ij} \xi_i \xi_j\right)^2 + 8\mathbb{E}\left(\sum_{i=1}^n \xi_i^2 P_{ii}\right)^2 \\ &\leq 8 \max(B, 2) \sum_{i,j} P_{ij}^2 + 8 \sum_{i \neq j} P_{ii} P_{jj} + 8B \sum_{i,j} P_{ii}^2 P_{jj}^2 \\ &\leq 8 \max(B, 2)r + 8r^2 + 8Br^2 \\ &\leq 32Br^2. \end{aligned}$$

Hence,  $\mathbb{E}(\langle X, P_0 X \rangle^4) \leq (256Br)^2$ , and applying Markov's inequality, we get the following bound for (\*):

$$(*) \leq \frac{(1024Br)^2}{u^4}.$$

Let us turn to estimating (\*\*). We will use symmetrization, truncation, and concentration. Let  $\tilde{X} = (\tilde{\xi}_k)_{1 \leq k \leq n}$  be an independent copy of  $X$ . Note that

$$\mathbb{E}(\langle \tilde{X}, (P - P_0)\tilde{X} \rangle - r)^2 = \mathbb{E}\left(\left(\sum_{k=1}^n P_{kk}(\tilde{\xi}_k^2 - 1)\right)^2\right) \leq Br,$$

so, using the independence of  $X$  and  $\tilde{X}$  and applying Markov's inequality, we get

$$\begin{aligned} (**) &\leq 2\mathbb{P}\left(\left|\langle X, (P - P_0)X \rangle - \langle \tilde{X}, (P - P_0)\tilde{X} \rangle\right| > u/2 - \sqrt{2Br}\right) \\ &= 2\mathbb{P}\left(\left|\sum_{k=1}^n P_{kk}(\xi_k^2 - \tilde{\xi}_k^2)\right| > u/2 - \sqrt{2Br}\right). \end{aligned}$$

Clearly, the variables  $\xi_k^2 - \tilde{\xi}_k^2$  ( $1 \leq k \leq n$ ) are symmetrically distributed, with the variance bounded from above by  $2B$ . Let  $h : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  be a function defined as

$$h(t) := \mathbb{E}\left(\left(\xi_k^2 - \tilde{\xi}_k^2\right)^2 \chi_t\right),$$

where  $\chi_t$  denotes the indicator of the event  $\{|\xi_k^2 - \tilde{\xi}_k^2| \geq t\}$ . Clearly,  $h(t) \rightarrow 0$  when  $t$  tends to infinity. Note that, by Hoeffding’s inequality, we have

$$\begin{aligned} \mathbb{P}\left(\left|\sum_{k=1}^n P_{kk} \left(\xi_k^2 - \tilde{\xi}_k^2\right) \left(1 - \chi_{r^{1/4}}\right)\right| > u/4\right) &\leq 2 \exp\left(-\frac{2u^2}{16\sqrt{r}} \left(4 \sum_k P_{kk}^2\right)^{-1}\right) \\ &\leq 2 \exp\left(-u^2/(32r\sqrt{r})\right). \end{aligned}$$

On the other hand, since the random variable  $(\xi_k^2 - \tilde{\xi}_k^2)\chi_{r^{1/4}}$  has zero mean, we have

$$\mathbb{E}\left(\left(\sum_{k=1}^n P_{kk} \left(\xi_k^2 - \tilde{\xi}_k^2\right) \chi_{r^{1/4}}\right)^2\right) \leq rh(r^{1/4}).$$

Applying Markov’s inequality, we get for all  $u > 4\sqrt{2Br}$ ,

$$\mathbb{P}\left(\left|\sum_{k=1}^n P_{kk} \left(\xi_k^2 - \tilde{\xi}_k^2\right) \chi_{r^{1/4}}\right| > u/4 - \sqrt{2Br}\right) \leq \frac{rh(r^{1/4})}{(u/4 - \sqrt{2Br})^2}.$$

Combining the estimates, we obtain

$$\begin{aligned} (**) &\leq 2\mathbb{P}\left(\left|\sum_{k=1}^n P_{kk} \left(\xi_k^2 - \tilde{\xi}_k^2\right) \left(1 - \chi_{r^{1/4}}\right)\right| > u/4\right) \\ &\quad + 2\mathbb{P}\left(\left|\sum_{k=1}^n P_{kk} \left(\xi_k^2 - \tilde{\xi}_k^2\right) \chi_{r^{1/4}}\right| > u/4 - \sqrt{2Br}\right) \\ &\leq 4 \exp\left(-u^2/(32r\sqrt{r})\right) + \frac{128rh(r^{1/4})}{u^2} \quad \text{for all } u > 8\sqrt{2Br}. \end{aligned}$$

Finally, grouping together the bounds for (\*) and (\*\*), we get for all  $u > 8\sqrt{2Br}$ ,

$$\mathbb{P}\left(\left|\|PX\|^2 - r\right| > u\right) \leq \frac{(1024Br)^2}{u^4} + 4 \exp\left(-u^2/(32r\sqrt{r})\right) + \frac{128rh(r^{1/4})}{u^2}.$$

Set  $\alpha \in (0, 1/6)$ . For any  $u > 8\sqrt{2Br}^{1-\alpha} = o(r)$ , using the inequality  $e^{-2t} \leq 1/t^4$  for  $t > 0$ , the right hand side of the last equation above is bounded above by

$$\frac{r}{u^2} \frac{(1024B)^2 r}{u^2} + \frac{r}{u^2} 4 \frac{64^4 r^5}{u^6} + \frac{r}{u^2} 128h(r^{1/4}) = \frac{o(r)}{u^2}.$$

This proves the desired result. We note that the proof can be shortened and simplified under the stronger assumption that  $\mathbb{E}(|\xi_1|^p) < \infty$  for some  $p > 4$ , see also [48, Proposition 1.3] for thin-shell estimates in the same spirit.  $\square$

*Proof of Proposition 1.5* This proof was suggested by an anonymous referee. Let  $X_n, n \in \mathbb{N}$ , be as in (1.1) and assume that  $\lim_{n \rightarrow \infty} \frac{n}{m_n} = \rho$  with  $\rho \in (0, 1)$ , in other words that  $\underline{\rho} = \bar{\rho} \in (0, 1)$ . It is known—see [2, Theorem 2.4]—that a sufficient condition for the Marchenko–Pastur theorem (1.5) to hold is

$$\lim_{n \rightarrow \infty} \sup_A \frac{1}{n} \mathbb{E} \left| X_n^\top A X_n - \text{tr}(A) \right| = 0$$

where the supremum runs over all  $n \times n$  complex matrices such that  $\|A\| \leq 1$  (let us note that a similar condition was exploited in [36]). Recall that  $\|A\| = \max_{\|x\|=1} \|Ax\| = s_{\max}(A)$ . Note that the mapping  $A \mapsto v^\top A v - \text{tr}(A)$  is linear for any  $v \in \mathbb{R}^n$ . Actually, as  $\|\Re A\|, \|\Im A\| \leq \|A\|$ , one can consider only real matrices in the above condition. Moreover, as  $v^\top A v = 0 = \text{tr}(A)$  when  $A$  is antisymmetric, it is enough to verify the condition for real symmetric matrices. Furthermore, by splitting the spectrum it can be shown that one can consider only real symmetric positive semi-definite matrices. Now, if  $A \neq 0$  is a real symmetric positive semi-definite matrix, then  $A/\|A\|$  can be represented as a convex linear combination of some orthogonal projectors. Indeed, write

$$\begin{aligned} \frac{A}{\|A\|} &= \sum_{k=1}^n \lambda_k v_k v_k^\top \\ &= \lambda_n \sum_{k=1}^n v_k v_k^\top + (\lambda_{n-1} - \lambda_n) \sum_{k=1}^{n-1} v_k v_k^\top + \dots + (\lambda_1 - \lambda_2) v_1 v_1^\top \\ &=: \lambda_n \Pi^{(n)} + (\lambda_{n-1} - \lambda_n) \Pi^{(n-1)} + \dots + (1 - \lambda_2) \Pi^{(1)}, \end{aligned}$$

where  $(v_k)_{1 \leq k \leq n}$  is an orthonormal basis of eigenvectors of  $A/\|A\|$  associated to the eigenvalues  $\lambda_1, \dots, \lambda_n$  ordered as  $0 \leq \lambda_n \leq \lambda_{n-1} \leq \dots \leq \lambda_1 = 1$ . Now, due to the convexity of the  $L^1$ -norm, we get that a sufficient condition for the Marchenko–Pastur theorem to hold is

$$\lim_{n \rightarrow \infty} \sup_{\Pi} \frac{1}{n} \mathbb{E} \left| X_n^\top \Pi X_n - \text{tr}(\Pi) \right| = 0,$$

where the supremum runs over all orthogonal projectors in  $\mathbb{R}^n$ . In turn, the latter is equivalent to

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \left| X_n^\top \Pi_n X_n - \text{tr}(\Pi_n) \right| = 0 \tag{2.5}$$

for any sequence  $(\Pi_n)_{n \geq 1}$  where  $\Pi_n$  is an orthogonal projector in  $\mathbb{R}^n$  for every  $n \geq 1$ . Since  $X_n$  is isotropic for every  $n \geq 1$ , it turns out that the latter is in turn equivalent to

$$\frac{1}{n} (X_n^\top \Pi_n X_n - \text{tr}(\Pi_n)) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0 \tag{2.6}$$

for any sequence  $(\Pi_n)_{n \geq 1}$  where  $\Pi_n$  is an orthogonal projector in  $\mathbb{R}^n$  for every  $n \geq 1$ . To see that (2.5) and (2.6) are equivalent, consider variables  $Z_n := \frac{1}{n}(X_n^\top \Pi_n X_n - \text{tr}(\Pi_n)) + 1 \geq 0$  which satisfy  $\mathbb{E}(Z_n) = 1$  for every  $n \geq 1$ . Note that, assuming condition (2.6), we have  $Z_n \xrightarrow{\mathbb{P}} 1$ , and then use the fact that if a sequence  $(Z_n)_{n \geq 1}$  satisfies  $Z_n \xrightarrow{\mathbb{P}} Z$ ,  $\mathbb{E}(|Z_n|) \rightarrow \mathbb{E}(|Z|)$ , and  $\mathbb{E}(|Z|) < \infty$  then  $\mathbb{E}(|Z_n - Z|) \rightarrow 0$ , see for instance [46, Problem 18, p. 211].

Let us show now that (WTP) implies (2.6). We may assume without loss of generality that  $\text{rank}(\Pi_n) \rightarrow \infty$  as  $\mathbb{E}(X_n^\top \Pi_n X_n) = \text{rank}(\Pi_n) = \text{tr}(\Pi_n)$ . By (WTP-b), there exists a sequence  $(\varepsilon_n)_{n \geq 1}$  such that  $\varepsilon_n \rightarrow 0$  and, denoting  $P_n := \frac{1}{\text{rank}(\Pi_n)} X_n^\top \Pi_n X_n$ , we have

$$\lim_{n \rightarrow \infty} \mathbb{P}(P_n > 1 + \varepsilon_n) = 0.$$

We only need to show that  $P_n \xrightarrow{\mathbb{P}} 1$ . For every fixed real number  $M > 1$ , we have

$$1 = \mathbb{E}(P_n) = \mathbb{E}(P_n \mathbf{1}_{\{P_n \leq 1 + \varepsilon_n\}}) + \mathbb{E}(P_n \mathbf{1}_{\{1 + \varepsilon_n < P_n \leq M\}}) + \mathbb{E}(P_n \mathbf{1}_{\{P_n > M\}}).$$

First note that  $\mathbb{E}(P_n \mathbf{1}_{\{1 + \varepsilon_n < P_n \leq M\}}) \leq M \mathbb{P}(P_n > 1 + \varepsilon_n) \xrightarrow{n \rightarrow \infty} 0$ . Now by (WTP-a), the family  $\mathcal{A} := \{X_n, y\}^2 : n \geq 1, y \in S^{n-1}\}$  is uniformly integrable. Hence, the family

$$\mathcal{B} := \bigcup_{n \geq 1} \left\{ \frac{X_n^\top \Pi X_n}{\text{rank}(\Pi)} : \Pi \neq 0 \text{ is an orthogonal projector in } \mathbb{R}^n \right\}$$

is also uniformly integrable, because, denoting by  $(v_k)_{1 \leq k \leq n}$  the orthonormal basis of eigenvectors of  $\Pi$ ,

$$\frac{X_n^\top \Pi X_n}{\text{rank}(\Pi)} = \frac{1}{\text{rank}(\Pi)} \sum_{k=1}^{\text{rank}(\Pi)} \langle X_n, v_k \rangle^2$$

is a convex linear combination of some elements of  $\mathcal{A}$ .

Next, the uniform integrability of  $\mathcal{B}$  implies that  $\lim_{M \rightarrow \infty} \sup_{n \geq 1} \mathbb{E}(P_n \mathbf{1}_{\{P_n > M\}}) = 0$  and then

$$\liminf_{n \rightarrow \infty} \mathbb{E}(P_n \mathbf{1}_{\{P_n \leq 1 + \varepsilon_n\}}) \geq 1.$$

The latter is possible only if  $\mathbb{E}(P_n \mathbf{1}_{\{P_n \leq 1 + \varepsilon_n\}}) \rightarrow 1$ . This yields  $P_n \xrightarrow{\mathbb{P}} 1$ . □

### 3 Proof of Theorem 1.6

As in [14, 48], for every  $n \times n$  symmetric matrix  $S$  and  $t \in \mathbb{R} \setminus \{\lambda_1(S), \dots, \lambda_n(S)\}$ , we set

$$m_S(t) := \text{tr}((S - tI_n)^{-1}) = \sum_{k=1}^n \frac{1}{\lambda_k(S) - t}$$

The function  $t \mapsto m_S(t)$  is positive and strictly increasing on  $(-\infty, \lambda_n(S))$ . We note that  $n^{-1}m_S$  is the restriction on  $\mathbb{R}$  of the Cauchy–Stieltjes transform of the empirical spectral distribution of  $S$ .

Let us give an overview of the proof. Assume that  $n \leq m$  and  $B := \sum_{k=1}^m X^{(k)} \otimes X^{(k)}$ , and let

$$B^{(0)} := 0 \quad \text{and} \quad B^{(k)} := B^{(k-1)} + X^{(k)} \otimes X^{(k)}, \quad k = 1, \dots, m,$$

where  $X^{(k)}$  ( $k \leq m$ ) are independent isotropic random vectors in  $\mathbb{R}^n$ . We want to find a good lower bound for  $\lambda_{\min}(B)$ . As was mentioned in the introduction, we define inductively a sequence of numbers  $u_0, u_1, \dots, u_m$  such that  $u_k < \lambda_{\min}(B^{(k)})$  ( $k \leq m$ ) with probability one, implying  $\mathbb{E}\lambda_{\min}(B) \geq \mathbb{E}u_m$ . More precisely, we set  $u_0 := n - \sqrt{mn} < 0$ , and for every  $k \leq m$ , define  $u_k := u_{k-1} + \delta^k - \delta_R^k$ , where both  $\delta^k$  and  $\delta_R^k$  are random non-negative numbers. The quantities  $\delta^k$  and  $\delta_R^k$  are designed so that **1**)  $u_{k-1} + \delta^k$  (whence,  $u_k$ ) stays on the left of the spectrum of  $B^{(k)}$  with probability one, and **2**) the sum of expectations  $\sum_{k=1}^m (\delta^k - \delta_R^k)$  is close to  $m - \sqrt{mn}$ , implying that  $\mathbb{E}u_m$  is close to the (scaled) left edge of the Marchenko–Pastur distribution  $(\sqrt{m} - \sqrt{n})^2$ . The numbers  $\delta^k$ , in a non-random “rank one update” setting, are defined in Lemma 3.2 below. The definition is quite technical, and even the inequality  $u_{k-1} + \delta^k < \lambda_{\min}(B^{(k)})$  is not immediately clear. To take care of this condition, we utilize the concept of a *feasible lower shift* which was previously applied in [48]. With help of an auxiliary Lemma 3.1, based on the Sherman–Morrison formula, we verify that our  $\delta^k$  is a feasible lower shift, meaning that condition **1** is satisfied.

Up to this point, the argument is quite similar to [48]. Still, let us note that the formula for  $\delta^k$  which we use in our paper is different from the one in [48] (the reader might compare Lemma 3.2 with [48, Lemma 2.3]). Indeed, unlike in [48], here we are interested in getting a sharp limiting result; hence, our estimates for the expectation (condition **2**) must be more precise than in [48]. However, just “tuning” the definition of  $\delta^k$  from [48] turns out to be insufficient to reach our goal; that is the reason for introducing *regularity shifts*  $\delta_R^k$  which are the main technical novelty of the proof. We set

$$\delta_R^k := \min \left\{ \ell \in \{0, 1, \dots\} : \underline{m}_{B^{(k)}}(u_{k-1} + \delta^k - \ell) - \underline{m}_{B^{(k)}}(u_{k-1} + \delta^k - \ell - 1) \leq \frac{1}{\varepsilon n} \right\},$$

where  $\varepsilon > 0$  is a parameter which can be made arbitrarily small. The impact of the regularity shifts might seem counter-intuitive at first glance since they move  $u_k$ ’s in the direction away from the spectrum increasing the gap between  $u_k$  and  $\lambda_{\min}(B^{(k)})$ . For every  $k < m$ , the purpose of  $\delta_R^k$  is to guarantee *at the next step* some concentration inequalities for the quadratic form

$$(X^{(k+1)})^\top (B^{(k)} - u_k)^{-1} X^{(k+1)}$$

(this form will appear below as  $q_1$ ). The nature of this concentration property is revealed in Lemma 3.3, which in turn enables us to obtain a satisfactory estimate for the expectation of  $\delta^k$  in Lemma 3.4. In fact, the condition (3.4) in Lemma 3.3 is exactly what is provided by  $\delta_R^k$ 's. Note that the regularity shifts stay useful only if their cumulative effect is of lower order of magnitude compared to  $(\sqrt{m} - \sqrt{n})^2$ . In Lemma 3.5, we show that

$$\sum_{k=1}^m \delta_R^k \leq \frac{\varepsilon n^2}{\sqrt{mn} - n}$$

deterministically; thus, taking  $\varepsilon$  arbitrarily small, we can estimate the left edge of the spectrum of  $B$  with the required precision.

We find it useful to separate deterministic (linear algebraic) arguments from probabilistic estimates. In Sect. 3.1 below, we work with a non-random matrix  $A$  and a fixed vector  $x$ , thus performing a non-random rank one update to  $A + xx^\top$ . The shift  $\delta^k$  is defined deterministically as a function of the matrix and the vector. In the next Sect. 3.2, we replace the fixed vector  $x$  with a random isotropic vector  $X$ , but the matrix  $A$  stays non-random. Finally, in Sect. 3.3, we make the setting ‘‘fully random’’ (the way it was described above, but with matrices  $A^{(k)}$  in place of  $B^{(k)}$ 's).

### 3.1 Feasible lower shift

Let  $A$  be an  $n \times n$  positive semi-definite non-random matrix with eigenvalues  $\lambda_{\max} := \lambda_1 \geq \dots \geq \lambda_n =: \lambda_{\min} \geq 0$  and a corresponding orthonormal basis of eigenvectors  $(x_i)_{i=1}^n$ , and let  $u < \lambda_{\min}$ . Further, assume that  $x$  is a (non-random) vector in  $\mathbb{R}^n$ . We are interested in those numbers  $\delta \geq 0$  that (deterministically) satisfy

$$\lambda_{\min} > u + \delta \quad \text{and} \quad \underline{m}_{A+xx^\top}(u + \delta) \leq \underline{m}_A(u). \tag{3.1}$$

Following [48], any value of  $\delta$  satisfying (3.1), will be called a *feasible lower shift* with respect to  $A$ ,  $x$  and  $u$ . The following statement is taken from [48]; we provide its proof for reader’s convenience.

**Lemma 3.1** (Feasible lower shift—[48, Lemma 2.2]) *Let  $\delta \geq 0$  be such that  $u + \delta < \lambda_{\min}$ . Let us define*

$$q_1(\delta) := x^\top (A - u - \delta)^{-1} x = \sum_{i=1}^n \frac{\langle x, x_i \rangle^2}{\lambda_i - u - \delta}$$

$$q_2(\delta) := \frac{x^\top (A - u - \delta)^{-2} x}{\text{tr}((A - u - \delta)^{-2})} = \left( \sum_{i=1}^n (\lambda_i - u - \delta)^{-2} \right)^{-1} \sum_{i=1}^n \frac{\langle x, x_i \rangle^2}{(\lambda_i - u - \delta)^2}.$$

Then a sufficient condition for (3.1) to be satisfied is  $q_2(\delta) \geq \delta(1 + q_1(\delta))$ .

*Proof* If  $S$  is a symmetric matrix and if  $x$  is a vector, and if both  $S$  and  $S + xx^\top$  are invertible, then  $1 + x^\top S^{-1}x \neq 0$  since  $x^\top S^{-1}x = -1$  gives  $(S + xx^\top)S^{-1}x = 0$ . Moreover the inverse  $(S + xx^\top)^{-1}$  of the rank one update  $S + xx^\top$  of  $S$  can be expressed as

$$(S + xx^\top)^{-1} = S^{-1} - \frac{S^{-1}xx^\top S^{-1}}{1 + x^\top S^{-1}x}. \tag{3.2}$$

This is known as the Sherman–Morrison formula. This allows to write

$$\begin{aligned} \underline{m}_{A+xx^\top}(u + \delta) &= \text{tr}((A - u - \delta + xx^\top)^{-1}) \\ &= \text{tr}((A - u - \delta)^{-1}) - \frac{x^\top(A - u - \delta)^{-2}x}{1 + x^\top(A - u - \delta)^{-1}x} \\ &= \underline{m}_A(u) + \text{tr}((A - u - \delta)^{-1} - (A - u)^{-1}) \\ &\quad - \frac{x^\top(A - u - \delta)^{-2}x}{1 + x^\top(A - u - \delta)^{-1}x}. \end{aligned}$$

Now since  $A - (u + \delta)I_n$  is positive definite, it follows that

$$\delta(A - u - \delta)^{-2} - ((A - u - \delta)^{-1} - (A - u)^{-1})$$

is positive definite, and therefore

$$\underline{m}_{A+xx^\top}(u + \delta) - \underline{m}_A(u) \leq \delta \text{tr}((A - u - \delta)^{-2}) - \frac{x^\top(A - u - \delta)^{-2}x}{1 + x^\top(A - u - \delta)^{-1}x}.$$

Finally it can be checked that the right hand side is  $\leq 0$  if  $\delta(1 + q_1(\delta)) - q_2(\delta) \leq 0$ .  $\square$

**Lemma 3.2** (Construction of the feasible shift) *Let  $A, x, u$  and  $q_1, q_2$  be as above,  $\varepsilon \in (0, 1)$  and assume that*

$$\lambda_{\min} - u \geq 2/\varepsilon^2.$$

*Then the quantity*

$$\delta := \frac{(1 - \varepsilon)\mathbf{1}_{\{q_1(1/\varepsilon) \leq (1+\varepsilon)\underline{m}_A(u)+\varepsilon\}}}{(1 + \varepsilon)(1 + \underline{m}_A(u))} \left( \sum_{i=1}^n (\lambda_i - u)^{-2} \right)^{-1} \sum_{i=1}^n \frac{\langle x, x_i \rangle^2 \mathbf{1}_{\{(x, x_i)^2 \leq 1/\varepsilon\}}}{(\lambda_i - u)^2}$$

*satisfies  $\delta \leq 1/\varepsilon$  and is a feasible lower shift w.r.t.  $A, x$  and  $u$ , i.e.  $\lambda_{\min} > u + \delta$  and  $q_2(\delta) \geq \delta(1 + q_1(\delta))$ .*

*Proof* First, note that the condition  $\lambda_{\min} - u \geq 2/\varepsilon^2$  immediately implies that

$$\varepsilon(\lambda_i - u)^2 - \frac{2(\lambda_i - u)}{\varepsilon} + \frac{1}{\varepsilon^2} \geq 0, \quad i = 1, \dots, n,$$

which is in turn equivalent to the relation

$$\frac{1}{(1 - \varepsilon)(\lambda_i - u)^2} \geq \frac{1}{(\lambda_i - u - 1/\varepsilon)^2}, \quad i = 1, \dots, n. \tag{3.3}$$

Now, let us return to  $\delta$ . The inequality  $\delta \leq 1/\varepsilon$  follows directly from its definition. Next, since  $q_1(\cdot)$  is an increasing function on  $[0, \lambda_{\min} - u)$ , we have

$$\frac{(1 + q_1(\delta))\mathbf{1}_{\{q_1(1/\varepsilon) \leq (1+\varepsilon)\underline{m}_A(u)+\varepsilon\}}}{(1 + \varepsilon)(1 + \underline{m}_A(u))} \leq \frac{(1 + q_1(1/\varepsilon))\mathbf{1}_{\{q_1(1/\varepsilon) \leq (1+\varepsilon)\underline{m}_A(u)+\varepsilon\}}}{(1 + \varepsilon)(1 + \underline{m}_A(u))} \leq 1.$$

Thus, we obtain

$$\delta(1 + q_1(\delta)) \leq (1 - \varepsilon)q_2(0).$$

Finally, note that by the definition of  $q_2$  and in view of relation (3.3), we have

$$\begin{aligned} q_2(\delta) &= \left( \sum_{i=1}^n (\lambda_i - u - \delta)^{-2} \right)^{-1} \sum_{i=1}^n \frac{\langle x, x_i \rangle^2}{(\lambda_i - u - \delta)^2} \\ &\geq \left( \sum_{i=1}^n (\lambda_i - u - 1/\varepsilon)^{-2} \right)^{-1} \sum_{i=1}^n \frac{\langle x, x_i \rangle^2}{(\lambda_i - u - \delta)^2} \\ &\geq \left( (1 - \varepsilon)^{-1} \sum_{i=1}^n (\lambda_i - u)^{-2} \right)^{-1} \sum_{i=1}^n \frac{\langle x, x_i \rangle^2}{(\lambda_i - u - \delta)^2} \\ &\geq \left( (1 - \varepsilon)^{-1} \sum_{i=1}^n (\lambda_i - u)^{-2} \right)^{-1} \sum_{i=1}^n \frac{\langle x, x_i \rangle^2}{(\lambda_i - u)^2} \\ &= (1 - \varepsilon)q_2(0), \end{aligned}$$

and the statement follows. □

### 3.2 Randomization and control of expectations

Let, as before,  $A$  be an  $n \times n$  non-random positive semidefinite matrix with eigenvalues  $\lambda_1 \geq \dots \geq \lambda_n \geq 0$ , and let  $u < \lambda_n$ . We define a (random) quantity  $\delta$  as in Lemma 3.2, replacing the fixed vector  $x$  with a random isotropic vector  $X$ .

**Lemma 3.3** *Let  $\varepsilon \in (0, 12^{-3})$ ,  $n \geq 12/\varepsilon^4$ , and let  $X$  be a random isotropic vector in  $\mathbb{R}^n$  such that*

$$\mathbb{P}\left(\|PX\|^2 - \text{rank}P \geq \varepsilon \text{rank}P/6\right) \leq \varepsilon^4$$



for any non-zero orthogonal projection  $P$  of rank at least  $\varepsilon^{11}n/72$ . Further, let the non-random matrix  $A$ , the numbers  $(\lambda_i)_{i \leq n}$  and vectors  $(x_i)_{i \leq n}$  be as above, and  $u \in \mathbb{R}$  be such that  $\lambda_{\min} - u \geq 6\varepsilon^{-2} + \varepsilon^{-1}$ . Assume additionally that

$$\sum_{i=1}^n \frac{1}{(\lambda_i - u)(\lambda_i - u + 1)} \leq \frac{1}{\varepsilon n}. \tag{3.4}$$

Then, with  $q_1$  defined as in Lemma 3.1 (with  $X$  replacing the non-random vector  $x$ ), we have

$$\mathbb{P} \left\{ q_1(1/\varepsilon) + 1 \geq (1 + \varepsilon)(1 + \underline{m}_A(u)) \right\} \leq 4\varepsilon^2.$$

*Proof* First, note that the lower bound on  $\lambda_{\min} - u$  implies that  $q_1(1/\varepsilon) \leq (1 + \varepsilon/6)q_1(0)$  (deterministically). Let us split the index set  $\{1, \dots, n\}$  into several subsets in the following way: First, let  $I := \{i \leq n : \lambda_i - u \leq \varepsilon^4 n/12\}$ . Next, we set

$$J := \{i \leq n : \lambda_i - u \geq 6n/\varepsilon^3\},$$

so that

$$\{1, \dots, n\} \setminus (I \cup J) = \left\{ i \leq n : \varepsilon^4 n/12 < \lambda_i - u < 6n/\varepsilon^3 \right\}.$$

Note that, by the choice of  $\varepsilon$ , we have  $\exp(1/(12\varepsilon)) \geq 72/\varepsilon^7$ . Hence, the interval  $(\ln(\varepsilon^4 n/12), \ln(6n/\varepsilon^3))$  can be partitioned into  $\lfloor \varepsilon^{-2} \rfloor$  subintervals  $\mathcal{S}_k$  ( $k \leq \varepsilon^{-2}$ ) of length at most  $\varepsilon/6$  each. Then we let

$$I_k := \{i \leq n : \ln(\lambda_i - u) \in \mathcal{S}_k\}, \quad k \leq \varepsilon^{-2}.$$

Obviously,

$$q_1(0) = \sum_{i \in I} \frac{\langle X, x_i \rangle^2}{\lambda_i - u} + \sum_{i \in J} \frac{\langle X, x_i \rangle^2}{\lambda_i - u} + \sum_{k \leq \varepsilon^{-2}} \sum_{i \in I_k} \frac{\langle X, x_i \rangle^2}{\lambda_i - u} = (*) + (**) + (***) .$$

Let us estimate the three quantities separately.

First, in view of the condition (3.4), the lower bound on  $n$  and the definition of  $I$ , we have

$$\mathbb{E} (*) = \sum_{i \in I} \frac{1}{\lambda_i - u} \leq \sum_{i \in I} \frac{\varepsilon^4 n/12 + 1}{(\lambda_i - u)(\lambda_i - u + 1)} \leq \frac{\varepsilon^3}{6}.$$

Hence, by Markov’s inequality,

$$\mathbb{P} \{ (*) \geq \varepsilon/6 \} \leq \varepsilon^2.$$

Similarly,

$$\mathbb{E}(**) = \sum_{i \in J} \frac{1}{\lambda_i - u} \leq \frac{\varepsilon^3}{6},$$

whence

$$\mathbb{P}\{(**) \geq \varepsilon/6\} \leq \varepsilon^2.$$

Now, we consider the quantity (\*\*\*) . First, assume that for some  $k \leq \varepsilon^{-2}$  we have  $|I_k| \leq \varepsilon^{11}n/72$ . Since for every  $i \in I_k$  we have  $\lambda_i - u \geq \varepsilon^4n/12$ , we obtain

$$\mathbb{E} \sum_{i \in I_k} \frac{\langle X, x_i \rangle^2}{\lambda_i - u} \leq \frac{\varepsilon^7}{6},$$

whence, by Markov’s inequality,

$$\mathbb{P} \left\{ \sum_{i \in I_k} \frac{\langle X, x_i \rangle^2}{\lambda_i - u} \geq \varepsilon^3/6 \right\} \leq \varepsilon^4.$$

Now, if for some  $k \leq \varepsilon^{-2}$  we have  $|I_k| > \varepsilon^{11}n/72$ , then, by the condition on projections and the definition of  $I_k$ , denoting by  $P_k$  the orthogonal projection onto the span of  $(x_i)_{i \in I_k}$ , we obtain

$$\mathbb{P} \left\{ \sum_{i \in I_k} \frac{\langle X, x_i \rangle^2}{\lambda_i - u} \geq \sum_{i \in I_k} \frac{\exp(\varepsilon/6)(1 + \varepsilon/6)}{\lambda_i - u} \right\} \leq \mathbb{P} \left\{ \|P_k X\|^2 \geq (1 + \varepsilon/6)\text{rank}P_k \right\} \leq \varepsilon^4.$$

Combining all the above estimates together, we get

$$\begin{aligned} &\mathbb{P} \left\{ q_1(1/\varepsilon) \geq (1 + \varepsilon/6) (\varepsilon/2 + \exp(\varepsilon/6)(1 + \varepsilon/6)\underline{m}_A(u)) \right\} \\ &\leq \mathbb{P} \left\{ (*) + (**) + (***) \geq \varepsilon/6 + \varepsilon/6 + \varepsilon/6 + \exp(\varepsilon/6)(1 + \varepsilon/6)\underline{m}_A(u) \right\} \\ &\leq 4\varepsilon^2. \end{aligned}$$

It remains to note that

$$(1 + \varepsilon/6) (\varepsilon/2 + \exp(\varepsilon/6)(1 + \varepsilon/6)\underline{m}_A(u)) + 1 \leq (1 + \varepsilon)(1 + \underline{m}_A(u)).$$

□

**Lemma 3.4** (Control of  $\mathbb{E}\delta$ ) *Let  $\varepsilon \in (0, 12^{-3})$ ,  $n \geq 12/\varepsilon^4$ , and let  $A$ ,  $(\lambda_i)_{i \leq n}$ ,  $(x_i)_{i \leq n}$ ,  $X$  and  $\delta$  be the same as in Lemma 3.3 (and satisfy the same conditions).*

Assume additionally that  $\mathbb{E}(\langle X, x_i \rangle^2 \mathbf{1}_{\{\langle X, x_i \rangle^2 \leq 1/\varepsilon\}}) \geq r$  ( $i \leq n$ ) for some  $r \in [0, 1]$ . Then

$$\mathbb{E}\delta \geq \frac{(1 - \varepsilon)r}{(1 + \varepsilon)(1 + \underline{m}_A(u))} - 4\varepsilon.$$

*Proof* Since

$$\begin{aligned} \delta &= \frac{(1 - \varepsilon)(1 - \mathbf{1}_{\{q_1(1/\varepsilon) > (1 + \varepsilon)\underline{m}_A(u) + \varepsilon\}})}{(1 + \varepsilon)(1 + \underline{m}_A(u))} \\ &\quad \times \left( \sum_{i=1}^n (\lambda_i - u)^{-2} \right)^{-1} \sum_{i=1}^n \frac{\langle X, x_i \rangle^2 \mathbf{1}_{\{\langle X, x_i \rangle^2 \leq 1/\varepsilon\}}}{(\lambda_i - u)^2}, \end{aligned}$$

and in view of the bound  $\delta \leq 1/\varepsilon$ , we get

$$\begin{aligned} \mathbb{E}\delta &\geq \frac{(1 - \varepsilon)}{(1 + \varepsilon)(1 + \underline{m}_A(u))} \left( \sum_{i=1}^n (\lambda_i - u)^{-2} \right)^{-1} \sum_{i=1}^n \frac{\mathbb{E}(\langle X, x_i \rangle^2 \mathbf{1}_{\{\langle X, x_i \rangle^2 \leq 1/\varepsilon\}})}{(\lambda_i - u)^2} \\ &\quad - \varepsilon^{-1} \mathbb{P} \{q_1(1/\varepsilon) > (1 + \varepsilon)\underline{m}_A(u) + \varepsilon\}. \end{aligned}$$

Finally, applying Lemma 3.3 to the last expression, we get the result. □

### 3.3 Proof of Theorem 1.6, completed

Let  $(X_n)_{n \in \mathbb{N}}$  be as in the statement of the theorem. Without loss of generality, we can assume that both functions  $f, g : \mathbb{N} \rightarrow \mathbb{R}_+$  in the Weak Tail Projection property (WTP-b) are non-increasing. Additionally let us define

$$h(M) := \sup_{\substack{n \in \mathbb{N} \\ y \in \mathcal{S}^{n-1}}} \mathbb{E} \left( \langle X_n, y \rangle^2 \mathbf{1}_{\{\langle X_n, y \rangle^2 \geq M\}} \right), \quad M \geq 0$$

Note that (WTP-a) gives  $\lim_{M \rightarrow \infty} h(M) = 0$ .

Take any  $0 < \varepsilon < \min(12^{-3}, (\bar{\rho}^{-1/2} - 1)^2/4)$ , and define  $n_\varepsilon$  as the smallest integer greater than  $12/\varepsilon^4$  such that

- (a)  $g(\varepsilon^{11}n_\varepsilon/72) \leq \varepsilon^4$  and  $f(\varepsilon^{11}n_\varepsilon/72) \leq \varepsilon/6$  and
- (b) for all  $n \geq n_\varepsilon$  we have  $m_n/n \geq 3\bar{\rho}^{-1}/4 + 1/4$ .

Note that condition (b) implies that  $(\sqrt{m_n/n} - 1)^{-1} \leq 2/(\bar{\rho}^{-1/2} - 1)$ , whence  $(\sqrt{m_n/n} - 1)^2 \geq \varepsilon$  for all  $n \geq n_\varepsilon$ . From now on, we fix an  $n \geq n_\varepsilon$ , let  $m := m_n$  and let  $X^{(1)}, \dots, X^{(m)}$  be i.i.d. copies of  $X_n$ . We define

$$A^{(0)} := 0 \quad \text{and} \quad A^{(k)} := A^{(k-1)} + X^{(k)} \otimes X^{(k)}, \quad 1 \leq k \leq m,$$

so that  $A_n = A^{(m)}$ . Set

$$u_0 := n - \sqrt{mn}$$

and let  $u_1, \dots, u_m$  be a collection of *random* numbers defined inductively as follows:

$$u_k := u_{k-1} + \delta^k - \delta_R^k,$$

where  $\delta^k$  is defined as  $\delta$  in Lemma 3.2 (with  $A^{(k-1)}$ ,  $u_{k-1}$  and  $X^{(k)}$  replacing  $A$ ,  $u$  and  $x$ , respectively) and *the regularity shift*  $\delta_R^k$  is defined by

$$\delta_R^k := \min \left\{ \ell \in \{0, 1, \dots\} : \underline{m}_{A^{(k)}}(u_{k-1} + \delta^k - \ell) - \underline{m}_{A^{(k)}}(u_{k-1} + \delta^k - \ell - 1) \leq \frac{1}{\varepsilon n} \right\}.$$

Note that Lemmas 3.1 and 3.2 imply that we have, (deterministically) for all  $k \geq 0$ ,

$$\underline{m}_{A^{(k)}}(u_k) \leq \underline{m}_{A^{(0)}}(u_0) = \frac{n}{\sqrt{mn} - n}.$$

The ultimate purpose of the shift  $\delta_R^k$  is to guarantee relation (3.4) which was an important condition in proving the concentration Lemma 3.3. Of course, since  $\delta_R^k$  moves  $u_k$  away from the spectrum, one must make sure that the cumulative impact of the shifts  $\delta_R^k$  is small enough and does not destroy the desired asymptotic estimate.

**Lemma 3.5** *With  $\delta_R^k$  ( $1 \leq k \leq m$ ) defined above, the following holds deterministically:*

$$\sum_{k=1}^m \delta_R^k \leq \frac{\varepsilon n^2}{\sqrt{mn} - n}.$$

*Proof* Take any admissible  $k \geq 1$ . The definition of  $\delta_R^k$  immediately implies that for all  $0 \leq \ell < \delta_R^k$  we have

$$\underline{m}_{A^{(k)}}(u_{k-1} + \delta^k - \ell) - \underline{m}_{A^{(k)}}(u_{k-1} + \delta^k - \ell - 1) > \frac{1}{\varepsilon n}.$$

Hence,

$$\begin{aligned} \underline{m}_{A^{(k)}}(u_k) &= \underline{m}_{A^{(k)}}\left(u_{k-1} + \delta^k - \delta_R^k\right) \\ &\leq \underline{m}_{A^{(k)}}(u_{k-1} + \delta^k) - \frac{\delta_R^k}{\varepsilon n} \\ &\leq \underline{m}_{A^{(k-1)}}(u_{k-1}) - \frac{\delta_R^k}{\varepsilon n}. \end{aligned}$$

Thus,  $\underline{m}_{A^{(k)}}(u_k) \leq \underline{m}_{A^{(k-1)}}(u_{k-1}) - \frac{\delta_R^k}{\varepsilon n}$  for all  $k \geq 1$ , which, together with the relations  $0 < \underline{m}_{A^{(m)}}(u_m)$  and  $\underline{m}_{A^{(0)}}(u_0) = \frac{n}{\sqrt{mn} - n}$  implies the result.  $\square$

Now, fix for a moment any  $k \in \{1, \dots, m\}$  and let  $\mathcal{F}_{k-1}$  be the  $\sigma$ -algebra generated by the vectors  $X^{(1)}, \dots, X^{(k-1)}$ , with the convention  $\mathcal{F}_0 = \{\emptyset, \Omega\}$ . We will first estimate the conditional expectation  $\mathbb{E}(\delta^k \mid \mathcal{F}_{k-1})$ .

Note that, by the definition of  $u_{k-1}$ , we have (deterministically)

$$\underline{m}_{A^{(k-1)}}(u_{k-1}) - \underline{m}_{A^{(k-1)}}(u_{k-1} - 1) \leq \frac{1}{\varepsilon n}$$

(the above relation holds for  $k > 1$  in view of the definition of  $\delta_R^{k-1}$  and for  $k = 1$  — because of the definition of  $n_\varepsilon$ ). Together with the lower bound on  $n$  (which implies the conditions on orthogonal projections assumed in Lemma 3.4) and the condition

$$\underline{m}_{A^{(k-1)}}(u_{k-1}) \leq \underline{m}_{A^{(0)}}(u_0) = \frac{n}{\sqrt{mn} - n},$$

we get from Lemma 3.4 that

$$\begin{aligned} \mathbb{E}(\delta^k \mid \mathcal{F}_{k-1}) &\geq \frac{(1 - \varepsilon)(1 - h(1/\varepsilon))}{(1 + \varepsilon)(1 + \underline{m}_{A^{(0)}}(u_0))} - 4\varepsilon \\ &= \frac{(1 - \varepsilon)(1 - h(1/\varepsilon))}{1 + \varepsilon} \left(1 - \sqrt{\frac{n}{m}}\right) - 4\varepsilon. \end{aligned}$$

Hence, by the definition of  $u_k$ 's and Lemma 3.5, we obtain

$$\mathbb{E}u_m \geq u_0 + \frac{(1 - \varepsilon)(1 - h(1/\varepsilon))}{1 + \varepsilon} (m - \sqrt{mn}) - 4\varepsilon m - \frac{\varepsilon n^2}{\sqrt{mn} - n}.$$

Since  $u_m < \lambda_{\min}(A_n)$  (deterministically), we get from the above relation

$$\begin{aligned} \mathbb{E}\lambda_{\min}(A_n) &\geq (\sqrt{m} - \sqrt{n})^2 - (3\varepsilon + h(1/\varepsilon)) (m - \sqrt{mn}) - 4\varepsilon m - \frac{\varepsilon n^2}{\sqrt{mn} - n} \\ &\geq (\sqrt{m} - \sqrt{n})^2 - 7\varepsilon m - h(1/\varepsilon)m - \frac{2\varepsilon n}{\rho^{-1/2} - 1}. \end{aligned}$$

Since the above estimate holds for arbitrarily small  $\varepsilon$ , and having in mind that  $h(1/\varepsilon) \rightarrow 0$  with  $\varepsilon \rightarrow 0$ , we get the desired result.

### 4 Proof of Theorem 1.8

As in [14, 48], for any  $n \times n$  symmetric matrix  $S$  and every  $t \in \mathbb{R} \setminus \{\lambda_1(S), \dots, \lambda_n(S)\}$ , we set

$$\bar{m}_S(t) := \text{tr}((tI_n - S)^{-1}) = \sum_{k=1}^n \frac{1}{t - \lambda_k(S)}.$$

The function  $t \mapsto \bar{m}_S(t) = -\underline{m}_S(t)$  is positive and strictly decreasing on  $(\lambda_1(S), +\infty)$ .

The scheme of the proof of Theorem 1.8 in many respects resembles the proof of Theorem 1.6 from the previous section. Thus, we prefer to omit a detailed discussion here. Let us make just a few remarks. The concepts of *feasible shifts* and *regularity shifts* play analogous role. This time, we are talking about feasible *upper* shifts, again following [48]. Since the case of the largest eigenvalue is technically more complicated, our feasible upper shift  $\Delta^k$  is split into two components  $\Delta_1^k$  and  $\Delta_2^k$ , each designed to guarantee certain conditions on quadratic forms, and their expectations are estimated separately. As with the smallest eigenvalue, the purpose of regularity shifts  $\Delta_R^k$ , in cooperation with the Strong Tail Projection property, is to enable us to apply some concentration inequalities, which are used to obtain sharp estimates on the expectations of the shifts  $\Delta^k$ .

### 4.1 Feasible upper shift

Let  $A$  be an  $n \times n$  positive semi-definite non-random matrix with eigenvalues  $\lambda_{\max} := \lambda_1 \geq \dots \geq \lambda_n =: \lambda_{\min} \geq 0$  and a corresponding orthonormal basis of eigenvectors  $(x_i)_{i=1}^n$ , and let  $u > \lambda_1$  be such that  $\bar{m}_A(u) < 1$ . Further, assume that  $x$  is a (non-random) vector in  $\mathbb{R}^n$ . In this section, we consider those numbers  $\Delta \geq 0$  that (deterministically) satisfy

$$\lambda_{\max}(A + xx^\top) < u + \Delta \quad \text{and} \quad \bar{m}_{A+xx^\top}(u + \Delta) \leq \bar{m}_A(u). \tag{4.1}$$

Following [48], any value of  $\Delta$  satisfying (4.1), will be called a *feasible upper shift* with respect to  $A$ ,  $x$  and  $u$ . The following statement is taken from [48]; we provide its proof for completeness.

**Lemma 4.1** (Feasible upper shift—[48, Lemma 3.3]) *Let  $u > \lambda_1$  and  $\bar{m}_A(u) < 1$ . For any  $\Delta > 0$ , define*

$$Q_1(\Delta) := x^\top (u + \Delta - A)^{-1} x = \sum_{i=1}^n \frac{\langle x, x_i \rangle^2}{u + \Delta - \lambda_i}$$

$$Q_2(\Delta) := \frac{x^\top (u + \Delta - A)^{-2} x}{\bar{m}_A(u) - \bar{m}_A(u + \Delta)} = (\bar{m}_A(u) - \bar{m}_A(u + \Delta))^{-1} \sum_{i=1}^n \frac{\langle x, x_i \rangle^2}{(u + \Delta - \lambda_i)^2}.$$

Then a sufficient condition for (4.1) to be satisfied is

$$Q_1(\Delta) < 1 \quad \text{and} \quad Q_2(\Delta) \leq 1 - Q_1(\Delta).$$

*Proof* We can assume without loss of generality that  $x \neq 0$ , so that  $Q_2(\Delta) > 0$ . The Sherman–Morrison formula (3.2) gives

$$\begin{aligned} \bar{m}_{A+xx^\top}(u + \Delta) &= \text{tr}((u + \Delta)I_n - A - xx^\top)^{-1} \\ &= \text{tr}((u + \Delta)I_n - A)^{-1} + \frac{\|((u + \Delta)I_n - A)^{-1}x\|^2}{1 - x^\top((u + \Delta)I_n - A)^{-1}x} \\ &= \bar{m}_A(u + \Delta) + \left(1 - \sum_{i=1}^n \frac{\langle x, x_i \rangle^2}{u + \Delta - \lambda_i}\right)^{-1} \sum_{i=1}^n \frac{\langle x, x_i \rangle^2}{(u + \Delta - \lambda_i)^2}. \end{aligned}$$

Hence, we have  $\bar{m}_{A+xx^\top}(u + \Delta) \leq \bar{m}_A(u)$  if and only if

$$\left(1 - \sum_{i=1}^n \frac{\langle x, x_i \rangle^2}{u + \Delta - \lambda_i}\right)^{-1} \sum_{i=1}^n \frac{\langle x, x_i \rangle^2}{(u + \Delta - \lambda_i)^2} \leq \bar{m}_A(u) - \bar{m}_A(u + \Delta).$$

But the latter inequality clearly holds if  $Q_1(\Delta) < 1$  and  $Q_2(\Delta) \leq 1 - Q_1(\Delta)$ .

Next, the rank one matrix

$$(u + \Delta - A)^{-1/2}x((u + \Delta - A)^{-1/2}x)^\top = (u + \Delta - A)^{-1/2}xx^\top(u + \Delta - A)^{-1/2}$$

has eigenvalues 0 and  $\|(u + \Delta - A)^{-1/2}x\|^2$ . But the condition  $Q_1(\Delta) < 1$  implies  $\|(u + \Delta - A)^{-1/2}x\|^2 = x^\top(u + \Delta - A)^{-1}x < 1$ . Therefore the matrix

$$I_n - (u + \Delta - A)^{-1/2}xx^\top(u + \Delta - A)^{-1/2}$$

is positive definite, implying that

$$(u + \Delta - A)^{1/2}I_n(u + \Delta - A)^{1/2} - xx^\top = (u + \Delta)I_n - (A + xx^\top)$$

is positive definite (recall that for any two positive definite matrices  $S$  and  $T$ , the matrix  $S^{1/2}TS^{1/2}$  is positive definite). Hence,  $u + \Delta > \lambda_{\max}(A + xx^\top)$ .  $\square$

Following [48], we will treat the quantities  $Q_1$  and  $Q_2$  separately: for a specially chosen number  $\tau \in (0, 1)$ , we will take  $\Delta_1$  such that  $Q_1(\Delta_1) \leq \tau$ , and  $\Delta_2$  such that  $Q_2(\Delta_2) \leq 1 - \tau$ . Then, in view of monotonicity of  $Q_1$  and  $Q_2$ , the sum  $\Delta_1 + \Delta_2$  will be a feasible upper shift. In our case,  $\tau$  shall be close to  $\bar{m}_A(u)$ .

Let us introduce “level sets”  $I_j$  as follows:

$$I_j := \left\{i \leq n : 4^{j-1} \leq u - \lambda_i < 4^j\right\}, \quad j \in \mathbb{N}. \tag{4.2}$$

Note that the condition  $\bar{m}_A(u) < 1$  immediately implies that  $|I_j| < 4^j$  for all  $j \geq 1$ . Moreover, if we additionally assume that  $\max_{\ell \in \mathbb{N}} \frac{|I_\ell|}{16^\ell}$  is sufficiently small then the following estimate of the ratio  $\frac{|I_j|}{4^j}$  is valid:

**Lemma 4.2** *Let the sets  $I_j$  be as above and assume that for some  $\varepsilon \in (0, 1]$  we have*

$$\max_{\ell \in \mathbb{N}} \frac{|I_\ell|}{16^\ell} \leq \frac{1}{\varepsilon n}.$$

Then

$$\frac{|I_j|}{4^j} \leq \sqrt{\frac{|I_j|}{\varepsilon n}}, \quad j \in \mathbb{N}.$$

*Proof* Fix any natural  $j$  such that  $I_j \neq \emptyset$ . Then, obviously,

$$\frac{|I_j|}{4^j} \leq \frac{4^j}{\varepsilon n}.$$

Fix for a moment any  $\alpha > 0$ . If  $|I_j| \geq \alpha 4^j$  then

$$\frac{|I_j|}{4^j} \leq \frac{\alpha^{-1}|I_j|}{\varepsilon n}.$$

Otherwise,  $\frac{|I_j|}{4^j} \leq \alpha$ . It remains to choose  $\alpha := \sqrt{\frac{|I_j|}{\varepsilon n}}$ . □

For every  $j \geq 1$ , denote

$$h_j := \sum_{i \in I_j} \langle x, x_i \rangle^2 - |I_j|. \tag{4.3}$$

**Lemma 4.3** (Definition of  $\Delta_1$ ) *Let  $A, x$  and  $u$  be as above, with  $u > \lambda_{\max}$  and  $\bar{m}_A(u) < 1$ , and let  $\varepsilon \in (0, 1/4]$  be a real parameter. Define a number  $\Delta_1$  as follows:*

$$\Delta_1 := \Delta'_1 + \sum_{j=1}^{\lfloor \log_4(n/\varepsilon^2) \rfloor} \Delta_{1,j},$$

where

$$\Delta'_1 := \varepsilon^{-1/2} \|x\|^2 \mathbf{1}_{\{\varepsilon \|x\|^2 \geq n\}}$$

and, for each natural  $j \leq \log_4(n/\varepsilon^2)$ , we let

$$\Delta_{1,j} := \varepsilon^{-1} h_j \mathbf{1}_{\{h_j > \varepsilon^2 2^j \sqrt{|I_j|}\}}.$$

Let  $Q_1$  be as in Lemma 4.1. Then  $\Delta_1$  satisfies

$$Q_1(\Delta_1) \leq \bar{m}_A(u + \Delta_1) + 6\sqrt{\varepsilon} + 8\varepsilon\sqrt{n} \sqrt{\max_{\ell \in \mathbb{N}} \frac{|I_\ell|}{16^\ell}}.$$



*Proof* Denote by  $J$  the set of all  $j \leq \log_4(n/\varepsilon^2)$  such that  $h_j > \varepsilon^2 2^j \sqrt{|I_j|}$  and let  $J'$  be its complement inside  $\{1, \dots, \lfloor \log_4(n/\varepsilon^2) \rfloor\}$ . Then

$$\begin{aligned} Q_1(\Delta_1) &= \bar{m}_A(u + \Delta_1) + \sum_{j=1}^{\infty} \sum_{i \in I_j} \frac{\langle x, x_i \rangle^2 - 1}{u + \Delta_1 - \lambda_i} \\ &\leq \bar{m}_A(u + \Delta_1) + \sum_{j > \log_4(n/\varepsilon^2)} \sum_{i \in I_j} \frac{\langle x, x_i \rangle^2}{u + \Delta_1 - \lambda_i} \\ &\quad + 4 \sum_{j \in J'} \frac{h_j}{\Delta_1 + 4^j} + 4 \sum_{j \in J} \frac{h_j}{\Delta_1 + 4^j} \\ &=: \bar{m}_A(u + \Delta_1) + (*) + 4(**) + 4(***) . \end{aligned}$$

We shall estimate the last three quantities separately. First, note that

$$(*) \leq \frac{4\|x\|^2}{\Delta_1 + n/\varepsilon^2} \leq \frac{4\sqrt{\varepsilon}\|x\|^2}{\|x\|^2 \mathbf{1}_{\{\varepsilon\|x\|^2 \geq n\}} + n/\varepsilon^{3/2}} \leq 4\sqrt{\varepsilon}.$$

Next,

$$(**) \leq \varepsilon^2 \sum_{j \in J'} \sqrt{\frac{|I_j|}{4^j}} \leq \varepsilon^2 \sqrt{\max_{\ell \in \mathbb{N}} \frac{|I_\ell|}{16^\ell}} \sum_{j \leq \log_4(n/\varepsilon^2)} 2^j \leq 2\varepsilon \sqrt{n} \sqrt{\max_{\ell \in \mathbb{N}} \frac{|I_\ell|}{16^\ell}}.$$

Finally, we have

$$(***) \leq \frac{\sum_{j \in J} h_j}{\sum_{j \in J} \Delta_{1,j}} \leq \varepsilon.$$

Summing up the estimates, we get the result. □

Denote

$$F_2(\Delta) := \left( \sum_{i=1}^n \frac{1}{(u + \Delta - \lambda_i)(u - \lambda_i)} \right)^{-1} \sum_{i=1}^n \frac{\langle x, x_i \rangle^2}{(u + \Delta - \lambda_i)^2}, \quad \Delta \geq 0$$

(clearly,  $F_2(\Delta) = \Delta Q_2(\Delta)$  for all  $\Delta > 0$ ).

**Lemma 4.4** (Definition of  $\Delta_2$ ) *Let  $A, u$  and  $x$  be as above, with  $u > \lambda_1$ , and let  $\alpha > 0$  satisfy  $\bar{m}_A(u) + \alpha < 1$ . Define  $\Delta_2$  according to the following procedure: if*

$$(1 + \alpha)F_2(0) \leq \alpha(u - \lambda_1)(1 - \bar{m}_A(u) - \alpha)$$

then set

$$\Delta_2 := \frac{(1 + \alpha)F_2(0)}{1 - \bar{m}_A(u) - \alpha},$$

otherwise, take the smallest non-negative integer  $j$  such that

$$Q_2\left(2^j(u - \lambda_1)\right) \leq 1 - \bar{m}_A(u) - \alpha$$

and let

$$\Delta_2 := 2^j(u - \lambda_1).$$

Then  $\Delta_2 = 0$  whenever  $x = 0$ , and, for  $x \neq 0$ , we have  $\Delta_2 \neq 0$  and

$$Q_2(\Delta_2) \leq 1 - \bar{m}_A(u) - \alpha.$$

*Proof* The case  $x = 0$  is trivial, so further we assume that  $x \neq 0$  implying  $\Delta_2 \neq 0$ . First, suppose that

$$(1 + \alpha)F_2(0) \leq \alpha(u - \lambda_1)(1 - \bar{m}_A(u) - \alpha).$$

Then, in view of the definition of  $\Delta_2$ , we have  $\Delta_2 \leq \alpha(u - \lambda_1)$ , and

$$\begin{aligned} F_2(\Delta_2) &\leq \left( \sum_{i=1}^n \frac{1}{(u + \Delta_2 - \lambda_i)(u - \lambda_i)} \right)^{-1} \sum_{i=1}^n \frac{\langle x, x_i \rangle^2}{(u - \lambda_i)^2} \\ &\leq (1 + \alpha)F_2(0). \end{aligned}$$

Hence,

$$Q_2(\Delta_2) = F_2(\Delta_2)/\Delta_2 \leq 1 - \bar{m}_A(u) - \alpha.$$

If

$$(1 + \alpha)F_2(0) > \alpha(u - \lambda_1)(1 - \bar{m}_A(u) - \alpha)$$

then the statement follows directly from the definition of  $\Delta_2$ . □

### 4.2 Randomization and control of expectations

In this subsection, we “randomize” Lemmas 4.3 and 4.4. Let, as before,  $A$  be an  $n \times n$  (non-random) positive semidefinite matrix with eigenvalues  $\lambda_1 \geq \dots \geq \lambda_n \geq 0$ , let  $u > \lambda_1$  satisfy  $\bar{m}_A(u) < 1$ . We define (random) quantities  $\Delta_1$  and  $\Delta_2$  as in Lemmas 4.3 and 4.4, replacing the fixed vector  $x$  with a random isotropic vector  $X$ . In the following two lemmas, we will estimate the expectations of  $\Delta_1$  and  $\Delta_2$ .

**Lemma 4.5** (Control of  $\mathbb{E}\Delta_1$ ) *Let  $\varepsilon \in (0, 1/4]$  and let  $A, \lambda_i, u$  be as above and  $I_j$  be defined according to (4.2). Assume additionally that*

$$\max_{\ell \in \mathbb{N}} \frac{|I_\ell|}{16^\ell} \leq \frac{1}{\varepsilon n}.$$

Further, let  $X \in \mathbb{R}^n$  be an isotropic random vector and  $K \geq 1$ . Assume that for any non-zero orthogonal projection  $P : \mathbb{R}^n \rightarrow \mathbb{R}^n$  we have

$$\mathbb{P} \left\{ \|P(X)\|^2 - \text{rank}P \geq t \right\} \leq \frac{g(\text{rank}P)\text{rank}P}{t^2}, \quad t \geq f(\text{rank}P)\text{rank}P,$$

where functions  $f : \mathbb{N} \rightarrow [0, 1]$  and  $g : \mathbb{N} \rightarrow [0, K]$  satisfy

$$g(k) \leq \varepsilon^{11/2} \quad \text{and} \quad f(k) \leq \varepsilon^2 \quad \forall k \geq \varepsilon^{17}n.$$

Define a random variable  $\Delta_1$  as in Lemma 4.3 replacing the non-random vector  $x$  with  $X$ . Then

$$\mathbb{E}\Delta_1 \leq 32K\sqrt{\varepsilon}.$$

*Proof* Let us estimate separately the expectations of  $\Delta'_1$  and  $\Delta_{1,j}$ , ( $j \leq \log_4(n/\varepsilon^2)$ ), where the quantities are defined as in Lemma 4.3. First,

$$\begin{aligned} \mathbb{E}\Delta'_1 &= \varepsilon^{-3/2}n\mathbb{P}\{\|X\|^2 \geq n/\varepsilon\} + \int_{\varepsilon^{-3/2}n}^\infty \mathbb{P}\{\|X\|^2 \geq \sqrt{\varepsilon}\tau\} d\tau \\ &\leq \frac{\varepsilon^4}{(\varepsilon^{-1} - 1)^2} + \int_{\varepsilon^{-3/2}n - \varepsilon^{-1/2}n}^\infty \frac{\varepsilon^{9/2}n}{\tau^2} d\tau \\ &\leq \frac{\varepsilon^4}{(\varepsilon^{-1} - 1)^2} + \frac{\varepsilon^5}{\varepsilon^{-1} - 1} \\ &\leq 4\varepsilon^6. \end{aligned}$$

Next, fix any natural  $j \leq \log_4(n/\varepsilon^2)$  with  $I_j \neq \emptyset$ . We let  $h_j$  to be defined as in (4.3), with the random vector  $X$  replacing the non-random  $x$ . We have

$$\mathbb{E}\Delta_{1,j} = \varepsilon 2^j \sqrt{|I_j|} \mathbb{P} \left\{ h_j > \varepsilon^2 2^j \sqrt{|I_j|} \right\} + \int_{\varepsilon 2^j \sqrt{|I_j|}}^\infty \mathbb{P}\{h_j \geq \varepsilon\tau\} d\tau.$$

Note that  $h_j = \|P_j(X)\|^2 - \text{rank}P_j$  and  $|I_j| = \text{rank}P_j$ , where  $P_j$  is the orthogonal projection onto the span of  $\{x_i, i \in I_j\}$ . Let us consider two cases.

- $|I_j| \geq \varepsilon^{17}n$ . Since  $\overline{m}_A(u) < 1$ , we have  $|I_j| \leq 4^j$ , and  $\varepsilon^2|I_j| \leq \varepsilon^2 2^j \sqrt{|I_j|}$ . Hence, from the above formula for  $\mathbb{E}\Delta_{1,j}$  and the conditions on projections of  $X$  we obtain

$$\begin{aligned} \mathbb{E}\Delta_{1,j} &\leq \frac{\varepsilon^{5/2} \sqrt{|I_j|}}{2^j} + \frac{\varepsilon^{5/2} \sqrt{|I_j|}}{2^j} \\ &\leq 2^{j+1} \varepsilon^{5/2} \sqrt{\max_{\ell \in \mathbb{N}} \frac{|I_\ell|}{16^\ell}} \\ &\leq 2^{j+1} \varepsilon^2 n^{-1/2}. \end{aligned}$$

2.  $|I_j| < \varepsilon^{17}n$ . Then, in view of Lemma 4.2,  $|I_j| \leq \varepsilon^8 4^j$ , implying

$$f(\text{rank}P_j) \leq 1 \leq |I_j| = \sqrt{|I_j|}\sqrt{|I_j|} \leq \varepsilon^4 2^j \sqrt{|I_j|} < \varepsilon^2 2^j \sqrt{|I_j|}.$$

Applying again the formula for  $\mathbb{E}\Delta_{1,j}$  together with the projection conditions and Lemma 4.2, we obtain

$$\mathbb{E}\Delta_{1,j} \leq \frac{2K\sqrt{|I_j|}}{\varepsilon^3 2^j} \leq \min\left(\frac{2^{j+1}K}{\varepsilon^{7/2}\sqrt{n}}, 2\varepsilon K\right).$$

Thus, in any case  $\mathbb{E}\Delta_{1,j} \leq 2^{j+1}\varepsilon^2 n^{-1/2} + \min\left(\frac{2^{j+1}K}{\varepsilon^{7/2}\sqrt{n}}, 2\varepsilon K\right)$ . Summing over all  $j \leq \log_4(n/\varepsilon^2)$ , we get

$$\begin{aligned} \sum_{j \leq \log_4(n/\varepsilon^2)} \mathbb{E}\Delta_{1,j} &\leq 4\varepsilon + \sum_{j \leq \log_4(\varepsilon^9 n)} \frac{2^{j+1}K}{\varepsilon^{7/2}\sqrt{n}} + 2K\varepsilon(-11 \log_4 \varepsilon + 1) \\ &\leq 4\varepsilon + 4K\varepsilon + 22K\sqrt{\varepsilon} + 2K\varepsilon, \end{aligned}$$

and the statement follows. □

The next lemma does not require assumptions on projections of  $X$  other than of rank one.

**Lemma 4.6** (Control of  $\mathbb{E}\Delta_2$ ) *Let  $A$  and  $\lambda_i$  ( $i \leq n$ ) be as above and let  $X$  be an isotropic random vector in  $\mathbb{R}^n$  such that for some  $\kappa, K > 0$  we have  $\max_{\|y\|=1} \mathbb{E}|(X, y)|^{2+\kappa} \leq K$ , and let  $u > \lambda_1$  and  $\alpha \in (0, 1/2]$  satisfy  $\bar{m}_A(u) + \alpha < 1$ . Then for*

$$w_{4.6} := K2^{1+\kappa}/(\alpha^{1+\kappa/2}(1 - 2^{-\kappa/2}))$$

we have

$$\mathbb{E}\Delta_2 \leq \frac{1 + \alpha}{1 - \bar{m}_A(u) - \alpha} \left( 1 + \frac{w_{4.6}}{(1 - \bar{m}_A(u) - \alpha)^{\kappa/2}(u - \lambda_1)^{\kappa/2}} \right),$$

where  $\Delta_2$  is defined as in Lemma 4.4, with  $X$  replacing  $x$ .

*Proof* An inspection of the definition of  $\Delta_2$  in Lemma 4.4 immediately shows that

$$\Delta_2 \leq \frac{(1 + \alpha)F_2(0)}{1 - \bar{m}_A(u) - \alpha} + \sum_{j=0}^{\infty} 2^j (u - \lambda_1)\chi_j,$$

where  $\chi_0$  is the indicator function of the event

$$\{(1 + \alpha)F_2(0) > \alpha(u - \lambda_1)(1 - \bar{m}_A(u) - \alpha)\}$$

and for each  $j > 0$ ,  $\chi_j$  is the indicator of the event

$$\left\{ Q_2 \left( 2^{j-1}(u - \lambda_1) \right) > 1 - \bar{m}_A(u) - \alpha \right\}.$$

Note that for each fixed number  $r \geq 0$ ,  $F_2(r)$  is a random variable representable in the form

$$F_2(r) = \sum_{i=1}^n \beta_i \langle X, x_i \rangle^2,$$

where the numbers  $\beta_i = \beta_i(r)$  are positive and  $\sum_{i=1}^n \beta_i \leq 1$ . By the Minkowski inequality, for any  $p \geq 1$  we have

$$\begin{aligned} (\mathbb{E}F_2(r)^p)^{1/p} &\leq \sum_{i=1}^n \left( \mathbb{E}(\beta_i \langle X, x_i \rangle^2)^p \right)^{1/p} \\ &= \sum_{i=1}^n \beta_i \left( \mathbb{E}|\langle X, x_i \rangle|^{2p} \right)^{1/p} \leq \max_{\|y\|=1} \left( \mathbb{E}|\langle X, y \rangle|^{2p} \right)^{1/p}. \end{aligned}$$

In particular,  $\mathbb{E}F_2(r)^{1+\kappa/2} \leq \max_{\|y\|=1} \mathbb{E}|\langle X, y \rangle|^{2+\kappa} \leq K$ . Hence, from the definition of the indicator functions  $\chi_j$  and applying Markov's inequality, we get

$$\mathbb{E}\chi_0 \leq K \left( \alpha(1 + \alpha)^{-1}(u - \lambda_1)(1 - \bar{m}_A(u) - \alpha) \right)^{-1-\kappa/2},$$

and for every  $j \geq 1$ :

$$\mathbb{E}\chi_j \leq K \left( 2^{j-1}(u - \lambda_1)(1 - \bar{m}_A(u) - \alpha) \right)^{-1-\kappa/2}.$$

Finally, we obtain

$$\begin{aligned} \mathbb{E}\Delta_2 &\leq \frac{1 + \alpha}{1 - \bar{m}_A(u) - \alpha} \\ &\quad + K(u - \lambda_1) \left( \alpha(1 + \alpha)^{-1}(u - \lambda_1)(1 - \bar{m}_A(u) - \alpha) \right)^{-1-\kappa/2} \\ &\quad + \sum_{j=1}^{\infty} 2^j K(u - \lambda_1) \left( 2^{j-1}(u - \lambda_1)(1 - \bar{m}_A(u) - \alpha) \right)^{-1-\kappa/2} \\ &= \frac{1 + \alpha}{1 - \bar{m}_A(u) - \alpha} + \frac{K}{(1 - \bar{m}_A(u) - \alpha)^{1+\kappa/2}(u - \lambda_1)^{\kappa/2}} \\ &\quad \times \left( \frac{(1 + \alpha)^{1+\kappa/2}}{\alpha^{1+\kappa/2}} + \sum_{j=1}^{\infty} 2^{1+\kappa/2-\kappa j/2} \right) \end{aligned}$$

$$\leq \frac{(1 + \alpha)}{1 - \bar{m}_A(u) - \alpha} \times \left( 1 + \frac{K(1 + \alpha)^{\kappa/2}}{\alpha^{1+\kappa/2}(1 - \bar{m}_A(u) - \alpha)^{\kappa/2}(u - \lambda_1)^{\kappa/2}} \sum_{j=0}^{\infty} 2^{1+\kappa/2-\kappa j/2} \right),$$

and the result follows. □

### 4.3 Proof of Theorem 1.8, completed

Let  $X_n, m_n$  and  $A_n, n \in \mathbb{N}$ , be as in (1.1), (1.2), and (1.4), respectively. Without loss of generality, we can assume that both functions  $f$  and  $g$  in (STP) are non-increasing. Denote

$$\gamma := \inf_{n \geq 1} \frac{m_n}{n} > 0$$

and fix any  $\alpha \in (0, \sqrt{\gamma}/(1 + \sqrt{\gamma}))$ . Further, let  $\varepsilon := \alpha^2/225$  and let  $n_\varepsilon \in \mathbb{N}$  be such that

$$g(k) \leq \varepsilon^{11/2} \quad \text{and} \quad f(k) \leq \varepsilon^2 \quad \forall k \geq \varepsilon^{17} n_\varepsilon$$

and, additionally,

$$\frac{2(48 + 192g(1))\alpha^{-3/2}}{(\sqrt{\gamma}/(1 + \sqrt{\gamma}) - \alpha)^{3/2} (\sqrt{\varepsilon n_\varepsilon}/4)^{1/2}} \leq \varepsilon$$

(the last condition will be needed later to simplify the estimate coming from Lemma 4.6). From now on, we fix  $n \geq n_\varepsilon$ . For convenience, we let  $m := m_n$  and  $X^{(1)}, \dots, X^{(m)}$  be i.i.d. copies of  $X_n$ . We define

$$A^{(0)} := 0 \quad \text{and} \quad A^{(k)} = A^{(k-1)} + X^{(k)} \otimes X^{(k)}, \quad 1 \leq k \leq m,$$

so that  $A_n = A^{(m)}$ . Set

$$u_0 := n + \sqrt{mn}$$

and let  $u_1, \dots, u_m$  be a collection of random numbers defined inductively as follows: For every  $k = 1, 2, \dots, m$ , assuming that  $u_{k-1}$  has already been defined, we let  $I_j^k$  ( $j \in \mathbb{N}$ ) be random subsets of  $\{1, 2, \dots, n\}$  analogous to those given in (4.2), with  $A^{(k-1)}$  and  $u_{k-1}$  replacing  $A$  and  $u$ , respectively, i.e.

$$I_j^k := \left\{ i \leq n : 4^{j-1} \leq u_{k-1} - \lambda_i(A^{(k-1)}) < 4^j \right\}, \quad j \in \mathbb{N}.$$

Then we set

$$u_k := u_{k-1} + \Delta_1^k + \Delta_2^k + \Delta_R^k,$$

where  $\Delta_1^k$  is taken from Lemma 4.3 (with  $A^{(k-1)}$ ,  $u_{k-1}$ ,  $X^{(k)}$  and  $I_j^k$  replacing  $A$ ,  $u$ ,  $x$  and  $I_j$ , respectively);  $\Delta_2^k$  is defined as in Lemma 4.4 (again, with  $A^{(k-1)}$ ,  $u_{k-1}$  and  $X^{(k)}$  taking place of  $A$ ,  $u$  and  $x$ ), and  $\Delta_R^k$  is the *regularity shift* defined by

$$\Delta_R^k := \min \left\{ \ell \in \mathbb{N} \cup \{0\} : \bar{m}_{A^{(k)}} \left( u_{k-1} + \Delta_1^k + \Delta_2^k + \ell \right) - \bar{m}_{A^{(k)}} \left( u_{k-1} + \Delta_1^k + \Delta_2^k + \ell + 1 \right) \leq \frac{1}{2\epsilon n} \right\}$$

(the factor “2” in front of  $\epsilon n$  in the above definition does not carry any special meaning, and introduced for a purely technical reason). The quantities  $\Delta_R^k$  play a role analogous to numbers  $\delta_R^k$  from the section dealing with the smallest eigenvalue.

Before going further, we must make sure that the above definition is consistent, i.e.  $\bar{m}_{A^{(k)}}(u_k) + \alpha < 1$  and  $u_k > \lambda_{\max}(A^{(k)})$  *deterministically* for all  $k \in \{0, 1, \dots, m\}$  (otherwise, we would not be able to apply Lemmas 4.1, 4.3, 4.4). We shall verify this by induction. For  $k = 0$ , the inequalities hold by the definition of  $u_0$  and  $\alpha$ . Now, let  $k \geq 1$ , and assume that  $u_{k-1}$  has the property that  $\bar{m}_{A^{(k-1)}}(u_{k-1}) + \alpha < 1$  and  $u_{k-1} > \lambda_{\max}(A^{(k-1)})$  *deterministically* (everywhere on the probability space). Let  $Q_1$  and  $Q_2$  be defined as in Lemma 4.1, with  $A^{(k-1)}$ ,  $u_{k-1}$  and  $X^{(k)}$  replacing  $A$ ,  $u$  and  $x$ . Note that if  $k = 1$  then we have

$$\sum_{j=1}^{\infty} \frac{|I_j^1|}{16^j} \leq \frac{16n}{u_0^2} < \frac{1}{\epsilon n};$$

and if  $k > 1$  then the summand  $\Delta_R^{k-1}$  in the definition of  $u_{k-1}$  makes sure that

$$\sum_{j=1}^{\infty} \frac{|I_j^k|}{16^j} \leq \sum_{j=1}^{\infty} \frac{2|I_j^k|}{4^j(4^j + 1)} < 2(\bar{m}_{A^{(k-1)}}(u_{k-1}) - \bar{m}_{A^{(k-1)}}(u_{k-1} + 1)) \leq \frac{1}{\epsilon n}.$$

Thus, by Lemma 4.3,

$$Q_1 \left( \Delta_1^k \right) \leq \bar{m}_{A^{(k-1)}}(u_{k-1}) + 14\sqrt{\epsilon}.$$

Further, by Lemma 4.4, we have

$$Q_2 \left( \Delta_2^k \right) \leq 1 - \bar{m}_{A^{(k-1)}}(u_{k-1}) - \alpha.$$

The definition of  $\epsilon$  and monotonicity of  $Q_1$ ,  $Q_2$  then imply that  $Q_1(u_k - u_{k-1}) < 1$  and  $Q_1(u_k - u_{k-1}) + Q_2(u_k - u_{k-1}) \leq 1$ . Thus, by Lemma 4.1 we have  $\bar{m}_{A^{(k)}}(u_k) \leq \bar{m}_{A^{(k-1)}}(u_{k-1}) < 1 - \alpha$  and  $u_k > \lambda_{\max}(A^{(k)})$  *deterministically*.

The following lemma gives an estimate of the cumulative impact of the regularity shifts:

**Lemma 4.7** *With  $\Delta_R^k$  ( $1 \leq k \leq m$ ) defined above, the following holds deterministically:*

$$\sum_{k=1}^m \Delta_R^k \leq 2\varepsilon n.$$

*Proof* Take any admissible  $k \geq 1$ . The definition of  $\Delta_R^k$  immediately implies that for all  $0 \leq \ell < \Delta_R^k$  we have

$$\bar{m}_{A^{(k)}}(u_{k-1} + \Delta_1^k + \Delta_2^k + \ell) - \bar{m}_{A^{(k)}}(u_{k-1} + \Delta_1^k + \Delta_2^k + \ell + 1) > \frac{1}{2\varepsilon n}.$$

Hence,

$$\begin{aligned} \bar{m}_{A^{(k)}}(u_k) &= \bar{m}_{A^{(k)}}(u_{k-1} + \Delta_1^k + \Delta_2^k + \Delta_R^k) \\ &\leq \bar{m}_{A^{(k)}}(u_{k-1} + \Delta_1^k + \Delta_2^k) - \frac{\Delta_R^k}{2\varepsilon n} \\ &\leq \bar{m}_{A^{(k-1)}}(u_{k-1}) - \frac{\Delta_R^k}{2\varepsilon n}. \end{aligned}$$

Thus,  $\bar{m}_{A^{(k)}}(u_k) \leq \bar{m}_{A^{(k-1)}}(u_{k-1}) - \frac{\Delta_R^k}{2\varepsilon n}$  for all  $k \geq 1$ , which, together with the relations  $0 < \bar{m}_{A^{(k)}}(u_k) < 1$  ( $0 \leq k \leq m$ ), implies the result.  $\square$

Now, fix for a moment any  $k \in \{1, \dots, m\}$  and let  $\mathcal{F}_{k-1}$  be the  $\sigma$ -algebra generated by the vectors  $X^{(1)}, \dots, X^{(k-1)}$ . We will first estimate the conditional expectation

$$\mathbb{E}(\Delta_1^k + \Delta_2^k | \mathcal{F}_{k-1}) = \mathbb{E}(\Delta_1^k | \mathcal{F}_{k-1}) + \mathbb{E}(\Delta_2^k | \mathcal{F}_{k-1}).$$

Applying Lemma 4.5, we obtain

$$\left\| \mathbb{E}(\Delta_1^k | \mathcal{F}_{k-1}) \right\|_\infty \leq 32g(1)\sqrt{\varepsilon}.$$

Next, note that the assumptions on 1-dimensional projections imply that for any unit vector  $y$  we have

$$\begin{aligned} \mathbb{E}|\langle y, X_n \rangle|^3 &= \int_0^\infty \mathbb{P}\left\{ \langle y, X_n \rangle^2 \geq \tau^{2/3} \right\} d\tau \\ &\leq \sqrt{8} + \int_{\sqrt{8}}^\infty \mathbb{P}\left\{ \langle y, X_n \rangle^2 - 1 \geq \tau^{2/3} - 1 \right\} d\tau \\ &\leq \sqrt{8} + 4g(1) \int_{\sqrt{8}}^\infty \tau^{-4/3} d\tau \\ &\leq \sqrt{8} + 12g(1). \end{aligned}$$



Next, the upper bound on  $\frac{|I_j^k|}{16^j}$  implies that  $I_j^k = \emptyset$  whenever  $j \leq \log_{16}(\varepsilon n)$ , whence

$$u_{k-1} - \lambda_1(A^{(k-1)}) \geq \frac{\sqrt{\varepsilon n}}{4}.$$

Then, by Lemma 4.6 (applied with  $\kappa := 1$ ), and in view of the inequality  $\overline{m}_{A^{(k-1)}}(u_{k-1}) \leq \overline{m}_{A^{(0)}}(u_0)$ , we get

$$\left\| \mathbb{E}(\Delta_2^k | \mathcal{F}_{k-1}) \right\|_\infty \leq \frac{1 + \alpha}{1 - \overline{m}_{A^{(0)}}(u_0) - \alpha} \left( 1 + \frac{w_{4.6}}{(1 - \overline{m}_{A^{(0)}}(u_0) - \alpha)^{1/2} (\sqrt{\varepsilon n}/4)^{1/2}} \right),$$

where

$$w_{4.6} = 4 \sup_{\|y\|=1} \mathbb{E}| \langle y, X_n \rangle |^3 / \left( \alpha^{3/2} \left( 1 - 2^{-1/2} \right) \right) \leq (48 + 192g(1))\alpha^{-3/2}.$$

It is not difficult to check that for  $\beta := (1 + 2\sqrt{\gamma})\alpha / (\sqrt{\gamma} - (1 + \sqrt{\gamma})\alpha)$  we have

$$\frac{1 + \alpha}{1 - \overline{m}_{A^{(0)}}(u_0) - \alpha} \leq \frac{1 + \beta}{1 - \overline{m}_{A^{(0)}}(u_0)},$$

and by the choice of  $n_\varepsilon$ , we have

$$\frac{w_{4.6}(1 + \alpha)}{(1 - \overline{m}_{A^{(0)}}(u_0) - \alpha)^{3/2} (\sqrt{\varepsilon n}/4)^{1/2}} \leq \varepsilon.$$

Hence, we obtain

$$\left\| \mathbb{E} \left( \Delta_2^k | \mathcal{F}_{k-1} \right) \right\|_\infty \leq (1 + \beta) \left( 1 + \sqrt{\frac{n}{m}} \right) + \varepsilon.$$

Summing up over  $k \in \{1, \dots, m\}$  and applying Lemma 4.7, we get

$$\begin{aligned} \mathbb{E}\lambda_1(A_n) &\leq \mathbb{E}u_m \\ &\leq n + \sqrt{mn} + 32g(1)\sqrt{\varepsilon}m + (1 + \beta)(m + \sqrt{mn}) + \varepsilon m + 2\varepsilon n \\ &= (\sqrt{n} + \sqrt{m})^2 + 32g(1)\sqrt{\varepsilon}m + \varepsilon(m + 2n) + \beta(m + \sqrt{mn}). \end{aligned}$$

Since  $\alpha$  can be chosen arbitrarily small and both  $\beta$  and  $\varepsilon$  tend to zero with  $\alpha$ , we get the result.

**Acknowledgements** D.C. would like to warmly thank Radosław Adamczak, Charles Bordenave, and Alain Pajor for discussions on this topic in Warsaw, Toulouse, and Paris. K.T. would like to thank Nicole Tomczak-Jaegermann for her support, Pierre Youssef for introducing him to the result of Batson–Spielman–Srivastava, and especially thank Alain Pajor for the invitation to visit Université Paris-Est Marne-la-Vallée in November–December, 2014. A significant part of the present work was done during that period. Both authors are grateful to the anonymous referees for valuable suggestions.

## References

1. Adamczak, R.: On the Marchenko–Pastur and circular laws for some classes of random matrices with dependent entries. *Electron. J. Probab.* **16**(37), 1068–1095 (2011)
2. Adamczak, R.: Some remarks on the Dozier–Silverstein theorem for random matrices with dependent entries. *Random Matrices Theory Appl.* **2**(2), 1250017 (2013)
3. Adamczak, R., Chafaï, D.: Circular law for random matrices with unconditional log-concave distribution. *Commun. Contemp. Math.* **17**(4), 1550020 (2015)
4. Adamczak, R., Latała, R., Litvak, A.E., Oleszkiewicz, K., Pajor, A., Tomczak-Jaegermann, N.: A short proof of Paouris’ inequality. *Can. Math. Bull.* **57**(1), 3–8 (2014)
5. Adamczak, R., Litvak, A.E., Pajor, A., Tomczak-Jaegermann, N.: Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles. *J. Am. Math. Soc.* **23**(2), 535–561 (2010)
6. Adamczak, R., Litvak, A.E., Pajor, A., Tomczak-Jaegermann, N.: Sharp bounds on the rate of convergence of the empirical covariance matrix. *C. R. Math. Acad. Sci. Paris* **349**(3–4), 195–200 (2011)
7. Anttila, M., Ball, K., Perissinaki, I.: The central limit problem for convex bodies. *Trans. Am. Math. Soc.* **355**(12), 4723–4735 (2003). (electronic)
8. Bai, Z.-D., Silverstein, J.W.: No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.* **26**(1), 316–345 (1998)
9. Bai, Z.-D., Silverstein, J.W.: *Spectral Analysis of Large Dimensional Random Matrices*. Springer Series in Statistics, 2nd edn. Springer, New York (2010)
10. Bai, Z.-D., Silverstein, J.W.: No eigenvalues outside the support of the limiting spectral distribution of information-plus-noise type matrices. *Random Matrices Theory Appl.* **1**(1), 1150004, 44 (2012)
11. Bai, Z.-D., Yin, Y.-Q.: Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix. *Ann. Probab.* **16**(4), 1729–1741 (1988)
12. Bai, Z.-D., Yin, Y.-Q.: Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. *Ann. Probab.* **21**(3), 1275–1294 (1993)
13. Bai, Z.-D., Zhou, W.: Large sample covariance matrices without independence structures in columns. *Stat. Sin.* **18**(2), 425–442 (2008)
14. Batson, J., Spielman, D.A., Srivastava, N.: Twice-Ramanujan sparsifiers. *SIAM J. Comput.* **41**(6), 1704–1721 (2012)
15. Batson, J.D., Spielman, D.A., Srivastava, N.: Twice-Ramanujan sparsifiers. In: *STOC’09—Proceedings of the 2009 ACM International Symposium on Theory of Computing*, pp. 255–262. ACM, New York (2009)
16. Benaych-Georges, F., Nadakuditi, R.R.: The singular values and vectors of low rank perturbations of large rectangular random matrices. *J. Multivar. Anal.* **111**, 120–135 (2012)
17. Bordenave, C.: Notes on random matrices (2014) (Available on the author webpage)
18. Borell, C.: Convex measures on locally convex spaces. *Ark. Mat.* **12**, 239–252 (1974)
19. Borodin, A., Forrester, P.J.: Increasing subsequences and the hard-to-soft edge transition in matrix ensembles. *J. Phys. A* **36**(12), 2963–2981 (2003). *Random matrix theory*
20. Brazitikos, S., Giannopoulos, A., Valettas, P., Vritsiou, B.-H.: *Geometry of Isotropic Convex Bodies*, Mathematical Surveys and Monographs, vol. 196. American Mathematical Society, Providence (2014)
21. Capitaine, M., Donati-Martin, C.: Strong asymptotic freeness for Wigner and Wishart matrices. *Indiana Univ. Math. J.* **56**(2), 767–803 (2007)
22. de la Peña, V.H., Giné, E.: *Decoupling. From Dependence to Independence, Randomly Stopped Processes. U-Statistics and Processes Martingales and Beyond*. Probability and its Applications (New York). Springer, New York (1999)
23. Feldheim, O.N., Sodin, S.: A universality result for the smallest eigenvalues of certain sample covariance matrices. *Geom. Funct. Anal.* **20**(1), 88–123 (2010)
24. Feller, W.: *An Introduction to Probability Theory and its Applications*, vol. II, 2nd edn. Wiley, New York, London, Sydney (1971)
25. Fleury, B.: Concentration in a thin Euclidean shell for log-concave measures. *J. Funct. Anal.* **259**(4), 832–841 (2010)
26. Guédon, O.: Concentration phenomena in high dimensional geometry. In: *Journées MAS 2012, ESAIM Proceedings*, vol. 44, pp. 47–60. EDP Sciences, Les Ulis (2014)
27. Guédon, O., Milman, E.: Interpolating thin-shell and sharp large-deviation estimates for isotropic log-concave measures. *Geom. Funct. Anal.* **21**(5), 1043–1068 (2011)

28. Haagerup, U., Thorbjørnsen, S.: A new application of random matrices:  $\text{Ext}(C_{\text{red}}^*(F_2))$  is not a group. *Ann. of Math. (2)* **162**(2), 711–775 (2005)
29. Johansson, K.: Shape fluctuations and random matrices. *Commun. Math. Phys.* **209**(2), 437–476 (2000)
30. Johnstone, I.M.: On the distribution of the largest eigenvalue in principal components analysis. *Ann. Stat.* **29**(2), 295–327 (2001)
31. Klartag, B.: Power-law estimates for the central limit theorem for convex sets. *J. Funct. Anal.* **245**(1), 284–310 (2007)
32. Koltchinskii, V., Mendelson, S.: Bounding the smallest singular value of a random matrix without concentration, preprint [arXiv:1312.3580](https://arxiv.org/abs/1312.3580) to appear in *International Mathematics Research Notices* (2015)
33. Lee, J.O., Yin, J.: A necessary and sufficient condition for edge universality of Wigner matrices. *Duke Math. J.* **163**(1), 117–173 (2014)
34. Litvak, A.E., Pajor, A., Rudelson, M., Tomczak-Jaegermann, N.: Smallest singular value of random matrices and geometry of random polytopes. *Adv. Math.* **195**(2), 491–523 (2005)
35. Marchenko, V.O., Pastur, L.A.: Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)* **72**(114), 507–536 (1967)
36. Pajor, A., Pastur, L.A.: On the limiting empirical measure of eigenvalues of the sum of rank one matrices with log-concave distribution. *Stud. Math.* **195**(1), 11–29 (2009)
37. Paouris, G.: Concentration of mass on convex bodies. *Geom. Funct. Anal.* **16**(5), 1021–1049 (2006)
38. Pastur, L.A., Shcherbina, M.: *Eigenvalue Distribution of Large Random Matrices*, *Mathematical Surveys and Monographs* (2011). ISBN:978-0-8218-5285-9, <http://www.ams.org/mathscinet-getitem?mr=2808038>
39. Péché, S.: Universality results for the largest eigenvalues of some sample covariance matrix ensembles. *Probab. Theory Relat. Fields* **143**(3–4), 481–516 (2009)
40. Petrov, V.V.: *Limit Theorems of Probability Theory*, *Oxford Studies in Probability. Sequences of Independent Random Variables*, Oxford Science Publications, vol. 4. The Clarendon Press, Oxford University Press, New York (1995)
41. Pillai, N.S., Yin, J.: Universality of covariance matrices. *Ann. Appl. Probab.* **24**(3), 935–1001 (2014)
42. Richard, K., Guionnet, A.: Central limit theorem and convergence of the support for Wishart matrices with correlated entries, preprint [arXiv:1401.7367](https://arxiv.org/abs/1401.7367) (2014)
43. Rudelson, M., Vershynin, R.: The Littlewood–Offord problem and invertibility of random matrices. *Adv. Math.* **218**(2), 600–633 (2008)
44. Rudelson, M., Vershynin, R.: Non-asymptotic theory of random matrices: extreme singular values. In: *Proceedings of the International Congress of Mathematicians*, vol. III., pp. 1576–1602. Hindustan Book Agency, New Delhi (2010)
45. Schultz, H.: Non-commutative polynomials of independent Gaussian random matrices. The real and symplectic cases. *Probab. Theory Relat. Fields* **131**(2), 261–309 (2005)
46. Shiryaev, A.N.: *Probability. 1. Graduate Texts in Mathematics*, 3rd ed., vol. 95. Springer, New York (2016). Translated from the fourth (2007) Russian edition by R. P. Boas and D. M. Chibisov
47. Soshnikov, A.: A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *J. Stat. Phys.* **108**(5–6), 1033–1056 (2002). Dedicated to David Ruelle and Yasha Sinai on the occasion of their 65th birthdays
48. Srivastava, N., Vershynin, R.: Covariance estimation for distributions with  $2+\varepsilon$  moments. *Ann. Probab.* **41**(5), 3081–3111 (2013)
49. Tikhomirov, K.: The limit of the smallest singular value of random matrices with i.i.d. entries. *Adv. Math.* **284**, 1–20 (2015)
50. Vershynin, R.: A simple decoupling inequality in probability theory, unpublished, available on the Internet (2011)
51. Yaskov, P.: Lower bounds on the smallest eigenvalue of a sample covariance matrix. *Electron. Commun. Probab.* **19**(83), 1–10 (2014)
52. Yaskov, P.: Sharp lower bounds on the least singular value of a random matrix without the fourth moment condition. *Electron. Commun. Probab.* **20**(44), 1–9 (2015)
53. Yin, Y.-Q., Bai, Z.-D., Krishnaiah, P.R.: On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix. *Probab. Theory Relat. Fields* **78**(4), 509–521 (1988)