

# COVARIANCE ESTIMATION FOR DISTRIBUTIONS WITH $2 + \varepsilon$ MOMENTS

NIKHIL SRIVASTAVA AND ROMAN VERSHYNIN

ABSTRACT. We study the minimal sample size  $N = N(n)$  that suffices to estimate the covariance matrix of an  $n$ -dimensional distribution by the sample covariance matrix in the operator norm, with an arbitrary fixed accuracy. We establish the optimal bound  $N = O(n)$  for every distribution whose  $k$ -dimensional marginals have uniformly bounded  $2 + \varepsilon$  moments outside the sphere of radius  $O(\sqrt{k})$ . In the specific case of log-concave distributions, this result provides an alternative approach to the Kannan-Lovasz-Simonovits problem, which was recently solved by Adamczak, Litvak, Pajor and Tomczak-Jaegermann [1]. Moreover, a lower estimate on the covariance matrix holds under a weaker assumption – uniformly bounded  $2 + \varepsilon$  moments of one-dimensional marginals. Our argument consists of randomizing the deterministic spectral sparsification technique of Batson, Spielman and Srivastava [4]. The new randomized method allows one to control the spectral edges of the sample covariance matrix via the Stieltjes transform evaluated at carefully chosen random points.

## 1. INTRODUCTION

**1.1. Covariance estimation problem.** Estimating covariance matrices of high dimensional distributions is a basic problem in statistics and its numerous applications. Consider a random vector  $X$  valued in  $\mathbb{R}^n$  and let us assume for simplicity that  $X$  is centered, i.e.  $\mathbb{E} X = 0$ ; this restriction will not be needed later. The covariance matrix of  $X$  is the  $n \times n$  positive semidefinite matrix

$$\Sigma = \mathbb{E} X X^T.$$

Our goal is to estimate  $\Sigma$  from a sample  $X_1, \dots, X_N$  taken from the same distribution as  $X$ . A classical unbiased estimator for  $\Sigma$  is the sample covariance matrix

$$\Sigma_N = \frac{1}{N} \sum_{i=1}^N X_i X_i^T.$$

A basic question is to *determine the minimal sample size  $N$  which guarantees that  $\Sigma$  is accurately estimated by  $\Sigma_N$* . More precisely, for a given accuracy  $\varepsilon > 0$  we are interested in the minimal  $N = N(n, \varepsilon)$  so that

$$\mathbb{E} \|\Sigma_N - \Sigma\| \leq \varepsilon \|\Sigma\|$$

where  $\|\cdot\|$  denotes the spectral (operator) norm. Replacing  $X$  by  $\Sigma^{-1/2} X$  and  $X_i$  by  $\Sigma^{-1/2} X_i$ , we reduce the problem to the distributions for which  $\Sigma = I$ , i.e. to *isotropic distributions*.

---

*Date:* June 14, 2011.

R. V. is partially supported by NSF grants FRG DMS 0918623 and DMS 1001829.

N. S. is supported by NSF grants CCF 0832797 and DMS 0835373.

**1.2. Sampling from isotropic distributions.** We consider independent isotropic random vectors  $X_i$  valued in  $\mathbb{R}^n$ , i.e. such that  $\mathbb{E} X_i X_i^T = I$ . Our goal is to determine the minimal sample size  $N = N(n, \varepsilon)$  such that

$$\mathbb{E} \|\Sigma_N - \Sigma\| \leq \varepsilon.$$

For obvious dimension reasons, one must have  $N \geq n$ . A remarkably general result of M. Rudelson ([12], see [16, Section 4.3]) yields that if  $\|X\|_2 = O(\sqrt{n})$  almost surely, then

$$(1.1) \quad N = O(n \log n)$$

where the  $O(\cdot)$  notation hides the dependence on  $\varepsilon$  here and thereafter. It is well known that the logarithmic oversampling factor cannot be removed from (1.1) in general, for example if the distribution is supported on  $O(n)$  points; see Section 1.8.

Nevertheless, it is also known that for sufficiently regular distributions the logarithmic oversampling factor is not needed in (1.1). This is a property of the standard normal distribution in  $\mathbb{R}^n$  and, more generally, of the distributions with *sub-gaussian* one-dimensional marginals. Namely,

$$N = O(n)$$

holds for every distribution that satisfies

$$(1.2) \quad \sup_{\|x\|_2 \leq 1} (\mathbb{E} |\langle X, x \rangle|^p)^{1/p} = O(\sqrt{p}) \quad \text{for } p \geq 1.$$

This result can be obtained by a standard covering argument, see [16, Section 4.3].

It is an open problem to describe the distributions for which the logarithmic oversampling is not needed, i.e. for which  $N = O(n)$ . The gap between sub-gaussian distributions where this bound holds and discrete distributions on  $O(n)$  points where it fails is quite large.

It is already a difficult problem to relax the sub-gaussian moment assumption (1.2) to anything weaker while keeping  $N = O(n)$ . A major step was made by R. Adamczak, A. Litvak, A. Pajor and N. Tomczak-Jaegermann [1], who showed that  $N = O(n)$  still holds (in fact, with high probability) under the *sub-exponential* moment assumptions:

$$(1.3) \quad \|X\|_2 = O(\sqrt{n}) \text{ a.s.}, \quad \sup_{\|x\|_2 \leq 1} (\mathbb{E} |\langle X, x \rangle|^p)^{1/p} = O(p) \quad \text{for } p \geq 1.$$

As an application, it was shown in [1] that  $N = O(n)$  holds for log-concave distributions, and in particular for the uniform distributions on isotropic convex bodies in  $\mathbb{R}^n$ . This answered a question posed by R. Kannan, L. Lovasz and M. Simonovits in [9].

The second author of the present paper speculated in [15] that  $N = O(n)$  should hold for a much wider class of distributions than sub-exponential, perhaps *for all distributions with  $2 + \varepsilon$  moments*. (The second moment – the variance – is assumed to be finite by the nature of the problem, as otherwise the covariance matrix is not defined.) The goal of the the current paper is to provide a result of this type.

**Theorem 1.1.** *Consider independent isotropic random vectors  $X_i$  valued in  $\mathbb{R}^n$ . Assume that  $X_i$  satisfy the strong regularity assumption: for some  $C, \eta > 0$ , one has*

$$(SR) \quad \mathbb{P}\{\|PX_i\|_2^2 > t\} \leq Ct^{-1-\eta} \quad \text{for } t > C \text{rank}(P)$$

*for every orthogonal projection  $P$  in  $\mathbb{R}^n$ . Then, for  $\varepsilon \in (0, 1)$  and for*

$$N \geq C_{\text{main}} \varepsilon^{-2-2/\eta} \cdot n$$

one has

$$(1.4) \quad \mathbb{E} \left\| \frac{1}{N} \sum_{i=1}^N X_i X_i^T - I \right\| \leq \varepsilon.$$

Here  $C_{\text{main}} = 512(48C)^{2+2/\eta}(6+6/\eta)^{1+4/\eta}$ , and as before  $\|\cdot\|$  denotes the spectral (operator) matrix norm and  $\|\cdot\|_2$  denotes the Euclidean norm in  $\mathbb{R}^n$ .

*Remark.* Since the distribution of  $PX_i$  is isotropic in the range of  $P$ , we have  $\mathbb{E} \|PX_i\|_2^2 = \text{rank}(P)$ . This explains why (SR) concerns only the tail values of  $t$  which are above  $\text{rank}(P)$ .

**1.3. Covariance estimation.** Returning to the covariance estimation problem, we deduce the following.

**Corollary 1.2** (Covariance estimation). *Consider a random vector  $X$  valued in  $\mathbb{R}^n$  with covariance matrix  $\Sigma$ . Assume that for some  $C, \eta > 0$ , the isotropic random vector  $Z = \Sigma^{-1/2}X$  satisfies*

$$(SR) \quad \mathbb{P}\{\|PZ\|_2^2 > t\} \leq Ct^{-1-\eta} \quad \text{for } t > C \text{rank}(P)$$

for every orthogonal projection  $P$  in  $\mathbb{R}^n$ . Then, for every  $\varepsilon \in (0, 1)$  and

$$N \geq C_{\text{main}} \varepsilon^{-2-2/\eta} \cdot n$$

the sample covariance matrix  $\Sigma_N$  obtained from  $N$  independent copies of  $X$  satisfies

$$\mathbb{E} \|\Sigma_N - \Sigma\| \leq \varepsilon \|\Sigma\|.$$

This result follows by applying Theorem 1.1 for the independent copies of the random vectors  $Z_i = \Sigma^{-1/2}X_i$  instead of  $X_i$ , and by multiplying the matrix  $\frac{1}{N} \sum_{i=1}^N X_i X_i^T - I$  in (1.4) by  $\Sigma^{1/2}$  on the left and on the right. Thus, for distributions satisfying (SR) we conclude that *the minimal sample size for the covariance estimation is  $N = O(n)$ .*

Let us illustrate these results with two important examples.

**1.4. Sampling from log-concave distributions and convex sets.** A notable class of examples where Corollary 1.2 applies is formed by the log-concave distributions, which includes the uniform distributions on convex bodies. Consider a random vector  $X$  with a log-concave distribution in  $\mathbb{R}^n$ , i.e. whose density has the form  $e^{-V(x)}$  where  $V(x)$  is a convex function on  $\mathbb{R}^n$ . A concentration inequality of G. Paouris [11] implies that regularity assumption (SR) holds for  $X$ . Indeed, consider an orthogonal projection  $P$  in  $\mathbb{R}^n$  and let  $k = \text{rank}(P)$ . The distribution of the isotropic random vector  $Z = \Sigma^{-1/2}X$  is log-concave in  $\mathbb{R}^n$ , and so is the distribution of  $PZ$  in the  $k$ -dimensional space  $\text{range}(P)$ . The theorem of G. Paouris then states that

$$\mathbb{P}\{\|PZ\|_2^2 > t\} \leq \exp(-ct) \quad \text{for } t > Ck$$

where  $C, c > 0$  are absolute constants. This is obviously stronger than assumption (SR), so Corollary 1.2 applies.

We conclude that *the minimal sample size for estimating the covariance matrix of a log-concave distribution is  $N = O(n)$ .* This matches the bound obtained by R. Adamczak et al. [1], though it should be noted that the guarantee of [1] holds with probability that converges to 1 exponentially fast as  $n \rightarrow \infty$ , whereas ours holds only in expectation. We have not tried

to obtain probability bounds of this type; note however that under our general assumption (SR), the probability can not converge to 1 faster than at a polynomial rate in  $n$ .

**1.5. Sampling from product distributions.** A distribution does not have to be log-concave in order to satisfy the regularity assumptions in Theorem 1.1 and Corollary 1.2. For example, all product distributions with finite  $4 + \varepsilon$  moments have the required regularity property. We can deduce this from the following thin shell estimate:

**Proposition 1.3** (Thin shell probability for product distributions). *Let  $p \geq 2$ , and consider a random vector  $X = (\xi_1, \dots, \xi_n)$ , where  $\xi_i$  are independent random variables with zero means, unit variances and with uniformly bounded  $(2p)$ -th moments. Then for every  $1 \leq k \leq n$  and for every orthogonal projection  $P$  in  $\mathbb{R}^n$  with  $\text{rank } P = k$ , one has*

$$(1.5) \quad \mathbb{E} \left| \|PX\|_2^2 - k \right|^p \lesssim k^{p/2}.$$

The factor implicit in (1.5) depends only on  $p$  and on the bound on the  $(2p)$ -th moments.

The proof of Proposition 1.3 is given in the Appendix.

Applying Chebychev's inequality together with (1.5) we obtain for  $t \geq k$  that

$$\mathbb{P}\{\|PX\|_2^2 > k + t\} \leq t^{-p} \cdot \mathbb{E} \left| \|PX\|_2^2 - k \right|^p \lesssim t^{-p} k^{p/2} \leq t^{-p/2}.$$

Thus for  $p > 2$  we get a sub-linear tail, as required in the regularity assumption (SR).

This shows that *Theorem 1.1 applies for product distributions in  $\mathbb{R}^n$  with uniformly bounded  $4 + \varepsilon$  moments, and it gives  $N = O(n)$  for their covariance estimation.* Note that this moment assumption is almost tight: without control of the fourth moments it is easy to construct a distribution for which one of the samples  $X_1, \dots, X_N$  is likely to have an unacceptably large coordinate much larger than  $\sqrt{N}$ , whenever  $N = O(n)$ .

**1.6. Extreme eigenvalues.** Theorem 1.1 states that, for sufficiently large  $N$ , all eigenvalues of the sample covariance matrix  $\Sigma_N = \frac{1}{N} \sum_{i=1}^N X_i X_i^T$  are concentrated near 1. It is easy to extend this to a result that holds for all  $N$ , as follows.

**Corollary 1.4.** *Let  $n, N$  be arbitrary positive integers, suppose  $X_i$  are independent isotropic random vectors in  $\mathbb{R}^n$  satisfying (SR), and let  $y = n/N$ . Then the sample covariance matrix  $\Sigma_N = \frac{1}{N} \sum_{i=1}^N X_i X_i^T$  satisfies:*

$$(1.6) \quad 1 - C_1 y^c \leq \mathbb{E} \lambda_{\min}(\Sigma_N) \leq \mathbb{E} \lambda_{\max}(\Sigma_N) \leq 1 + C_1(y + y^c).$$

Here  $c = \frac{\eta}{2\eta+2}$ ,  $C_1 = 512(16C)^{1+2/\eta}(6 + 6/\eta)^{1+4/\eta}$ , and  $\lambda_{\min}(\Sigma_N)$ ,  $\lambda_{\max}(\Sigma_N)$  denote the smallest and the largest eigenvalues of  $\Sigma_N$  respectively.

We deduce this result in Section 3. One can view (1.6) as a non-asymptotic form of the *Bai-Yin law* for the extreme eigenvalues of sample covariance matrices [3]. This law, associated with the work of S. Geman, Z. Bai, Y. Yin, P. Krishnaiah and J. Silverstein, applies for product distributions, specifically for random vectors  $X = (\xi_1, \dots, \xi_n)$  with i.i.d. components  $\xi_i$  with zero mean, unit variance and finite fourth moment. For such distributions one has asymptotically almost surely that

$$(1.7) \quad (1 - \sqrt{y})^2 - o(1) \leq \lambda_{\min}(\Sigma_N) \leq \lambda_{\max}(\Sigma_N) \leq (1 + \sqrt{y})^2 + o(1)$$

as  $n \rightarrow \infty$  and  $n/N \rightarrow y \in [0, 1]$ , see the rigorous statement in [3]. This limit law is sharp. On the other hand, the inequalities (1.6) hold in any fixed dimensions  $N, n$  and for general distributions (as in Theorem 1.1), *without any independence requirements* for the coordinates.

*Remark.* Comparing (1.6) with (1.7) one can ask about the optimal value of the exponent  $c$ , in particular whether  $c = 1/2$ . In a recent paper [2], R. Adamczak et al. obtained the optimal exponent  $c = 1/2$  for log-concave distributions, and more generally for sub-exponential distributions in the sense of (1.3). As (1.3) implies (SR) with  $\eta = (p-1)/2$  and  $C \leq (O(p))^p$ , Theorem 1.1 recovers a bound of  $c = 1/2 - 1/(p+1) = 1/2 - o(1)$  as  $p \rightarrow \infty$ .

*Remark* (Random matrices with independent rows). Corollary 1.4 can be interpreted as a result about the spectrum of random matrices with independent rows. Indeed, if  $A$  is the matrix with rows  $X_i$  then  $\Sigma_N = \frac{1}{N} \sum_{i=1}^N X_i X_i^T = \frac{1}{N} A^T A$ . So the singular values of the matrix  $\frac{1}{\sqrt{N}} A$  are the same as the eigenvalues of the matrix  $\Sigma_N$ , and they are controlled as in (1.6). In particular, under the regularity assumption (SR) on  $X_i$  we obtain that

$$(\mathbb{E} \|A\|^2)^{1/2} \leq C_2(\sqrt{N} + \sqrt{n})$$

where  $C_2 = \sqrt{2C_1}$  and  $C_1$  is as in Corollary 1.4.

Notice that while the rows of matrix  $A$  are independent, *the columns of  $A$  may be dependent*. The simpler case where all *entries* of  $A$  are independent is well understood by now. In the latter case, if the entries have zero mean and uniformly bounded fourth moments, the bound  $\mathbb{E} \|A\| \lesssim \sqrt{N} + \sqrt{n}$  follows, for example, from a general inequality of R. Latała [10].

**1.7. Smallest eigenvalue.** Our proof of Theorem 1.1 consists of two separate arguments for upper and lower bounds for the spectrum of the sample covariance matrix. It turns out that the full power of the strong regularity assumption (SR) is not needed for the lower bound. It suffices to assume *2 +  $\eta$  moments for one-dimensional marginals* rather than for marginals in all dimensions. This is only slightly stronger than the isotropy assumption, which fixes the *second moments* of one-dimensional marginals, and it broadens the class of distributions for which the result applies. We state this as a separate theorem.

**Theorem 1.5** (Smallest eigenvalue). *Consider independent isotropic random vectors  $X_i$  valued in  $\mathbb{R}^n$ . Assume that  $X_i$  satisfy the weak regularity assumption: for some  $C, \eta > 0$ ,*

$$(WR) \quad \sup_{\|x\|_2 \leq 1} \mathbb{E} |\langle X_i, x \rangle|^{2+\eta} \leq C.$$

*Then, for  $\varepsilon > 0$  and for*

$$(1.8) \quad N \geq C_{lower} \varepsilon^{-2-2/\eta} \cdot n,$$

*the minimum eigenvalue of the sample covariance matrix  $\Sigma_N = \frac{1}{N} \sum_{i=1}^N X_i X_i^T$  satisfies*

$$\mathbb{E} \lambda_{\min}(\Sigma_N) \geq 1 - \varepsilon.$$

*Here  $C_{lower} = 40(10C)^{2/\eta}$ .*

*Remark* (Moments vs. Tails). We have chosen to write (WR) in terms of moments rather than in terms of tail bounds as in (SR). By integration of the tails one can check that, for any given  $\eta > 0$ , (SR) with parameter  $C$  implies (WR) with parameter  $C' = C(2 + 2/\eta)$ .

In the remainder of the paper we will use (WR) for theorems regarding only the smallest eigenvalue and (SR) for theorems which involve the largest one.

*Remark* (Product distributions with  $2 + \eta$  moments). Many distributions of interest satisfy (WR). For example, let  $X = (\xi_1, \dots, \xi_n)$  have i.i.d. components  $\xi_i$  with zero mean, unit variance and finite  $(2 + \eta)$  moment. Then a standard application of symmetrization and Khintchine's inequality (or a direct application of Rosenthal's inequality [13], see [8]) shows that one-dimensional marginals of  $X$  also have bounded  $(2 + \eta)$  moments, i.e. (WR) holds.

In the context of the Bai-Yin law discussed in Section 1.6, this indicates that the *smallest* eigenvalue of a random matrix can be approximately controlled (as in (1.6)) even if the *fourth moment is infinite*. However, as we have already recalled, four moments are necessary to control the *largest* eigenvalue in the classical Bai-Yin law [5].

*Remark* (Covariance estimation). Theorem 1.5 can be used to obtain a *lower* estimate for the covariance matrix under the weak regularity assumption (WR).

**1.8. Optimality of the regularity assumptions.** Let us briefly mention two simple and known examples that illustrate the role of regularity assumptions (SR) and (WR) in the control of the largest and smallest eigenvalues respectively.

For the largest eigenvalue as in Theorem 1.1, it is not sufficient to put a regularity assumption of the type (SR) only on *one-dimensional* marginals, as it is done in Theorem 1.5 for the smallest eigenvalue. Even the following very strong (exponential) moment assumption is insufficient:

$$(1.9) \quad \sup_{\|x\|_2 \leq 1} \mathbb{P}\{|\langle X, x \rangle| > t\} \leq C \exp(-ct) \quad \text{for } t > 0.$$

Indeed, consider a random vector  $X = \xi Z$  where  $Z$  is a random vector uniformly distributed in the Euclidean sphere in  $\mathbb{R}^n$  centered at the origin and with radius  $\sqrt{n}$ , and where  $\xi$  is a standard normal random variable. Then  $X$  is isotropic, and all one-dimensional marginals of  $X$  have exponential tail decay (1.9). However, the multiplier  $\xi$  produces a dimension-free tail decay of the norm of  $Z$ , namely  $\mathbb{P}\{\|Z\|_2 > t\sqrt{n}\} = \mathbb{P}\{\xi > t\} \gtrsim \exp(-C't^2)$  for  $t > 0$ . It follows that a sample of  $N$  independent copies  $X_1, \dots, X_N$  of  $X$  satisfies  $\mathbb{E} \max_{i \leq N} \|X_i\|_2^2 \gtrsim N \log N$ , so the matrix  $\Sigma_N = \frac{1}{N} \sum_{i=1}^N X_i X_i^T$  satisfies

$$\mathbb{E} \|\Sigma_N - I\| \geq N^{-1} \mathbb{E} \max_{i \leq N} \|X_i\|_2^2 - 1 \gtrsim \log N,$$

which contradicts the conclusion of Theorem 1.1. This example is essentially due to G. Aubrun, see [1, Remark 4.9].

*Remark.* It is not clear whether Theorem 1.1 would hold if, in addition to  $(2 + \eta)$  moments on one-dimensional marginals, one puts a total boundedness assumption

$$\|X\| = O(\sqrt{n}) \quad \text{almost surely.}$$

A conjecture of this type is discussed in [15] where a version of Theorem 1.1 is proved under this assumption, with  $\eta = 2$  but with an additional  $(\log \log n)^{O(1)}$  oversampling factor.

Furthermore, we note that for the smallest eigenvalue as in Theorem 1.5, one can not drop the regularity assumption (WR), i.e. the assumption with  $\eta = 0$  is not sufficient. This is seen for  $X_i$  uniformly distributed in the set of  $2n$  points  $(\pm e_k)$  where  $(e_k)_{k=1}^n$  is an orthonormal

basis in  $\mathbb{R}^n$ . Indeed, in order that the smallest eigenvalue of the matrix  $\Sigma_N = \frac{1}{N} \sum_{i=1}^N X_i X_i^T$  be different from zero, one needs  $\Sigma_N$  to have full rank, for which all  $n$  basis vectors  $e_k$  need be present in the sample  $X_1, \dots, X_N$ . By the coupon collector's problem, for this to happen with constant probability one needs a sample of size  $N \gtrsim n \log n$ . For  $N = o(n \log n)$ , the smallest eigenvalue is zero with high probability, so the conclusion of Theorem 1.5 fails.

**1.9. The argument: randomizing the spectral sparsifier.** Our proof of Theorem 1.1 consists of randomizing the deterministic *spectral sparsification* method invented recently by J. Batson, D. Spielman and N. Srivastava [4] (see [14]). The randomization we propose makes this method appear naturally in the context of random matrix theory. The method is based on evaluating the Stieltjes transform of  $\Sigma_N$  while making rank one updates; however, in contrast to typical methods of random matrix theory (and to the deterministic process described in [4]), we shall evaluate the Stieltjes transform *at random real points*.

Let us illustrate the method by working out a crude upper bound  $O(1)$  for the largest eigenvalue of  $\Sigma_N$ . Equivalently, we want to show that a general Wishart matrix  $A_N := N\Sigma_N = \sum_{i=1}^N X_i X_i^T$  has all eigenvalues bounded by  $O(N)$ . We evaluate the Stieltjes transform

$$(1.10) \quad m_{A_N}(u) = \text{tr}(uI - A_N)^{-1} = \sum_{i=1}^n (u - \lambda_i(A_N))^{-1}, \quad u \in \mathbb{R},$$

where  $\lambda_i(A_N)$  denote the eigenvalues of  $A_N$ . This function has singularities at the points  $\lambda_i(A_N)$  and it vanishes at infinity. So the largest eigenvalue of  $A_N$  is the largest  $u$  where  $m_{A_N}(u) = \infty$ . However, such  $u$  is difficult to compute. So we soften this quantity by considering the largest number  $u_N := u_N(A_N)$  that satisfies

$$(1.11) \quad m_{A_N}(u_N) = \phi$$

where  $\phi$  is a fixed sensitivity parameter, for example  $\phi = 1$ .

Thus the *soft spectral edge*  $u_N$  provides an upper bound for the actual spectral edge,  $\lambda_{\max}(A_N) < u_N$ . So our goal is to show that

$$\mathbb{E} u_N = O(N).$$

This is the same problem as in [4] except the eigenvalues and hence the soft spectral edge  $u_N$  are now *random* points. The randomized problem is more difficult as we note below.

As opposed to the largest eigenvalue of  $A$ , the soft spectral edge  $u_N$  can be computed inductively using rank-one updates to the matrix. Specifically,  $u_k := u_k(A_k)$  will move to the right by a random amount at each step as we replace  $A_{k-1}$  by  $A_k = A_{k-1} + X_k X_k^T$ . Initially,  $A_0 = 0$  so  $u_0 = n$ . It suffices to prove that the  $u_k$  moves by  $O(1)$  on average at each step:

$$(1.12) \quad \mathbb{E}(u_k - u_{k-1}) = O(1).$$

Indeed, by summing up we would obtain the desired estimate  $\mathbb{E} u_N = n + O(1)N = O(N)$ .

The soft edge  $u_k$  can be recomputed at each step because it is determined by the Stieltjes transform  $m_{A_k}(u)$ , which in turn can be recomputed using Sherman-Morrison formula as is done in [4], which gives for every  $u \in \mathbb{R}$  that

$$(1.13) \quad m_{A_k}(u) = m_{A_{k-1}}(u) + \frac{X_k^T (uI - A)^{-2} X_k}{1 - X_k^T (uI - A)^{-1} X_k}.$$

This reduces proving (1.12) to a probabilistic problem which is essentially governed by the distribution of the random vector  $X_k$ .

The difficulty is that we are facing a non-linear inverse problem. Indeed, for an appropriate fixed  $u$  it is not difficult to show using (1.13) that  $m_{A_k}(u)$  remains bounded by  $\phi$  with nonzero probability whenever  $X_k$  is isotropic; this is done in [4]. However, here we require the identity  $m_{A_k}(u) = \phi$  to hold *deterministically*, because the largest  $u$  that satisfies it *defines* the soft spectral edge  $u_k$  as in (1.11). The task of computing the expectation of a random number  $u$  for which  $m_{A_k}(u) = \phi$  is a highly non-linear inverse problem [6, Section 4.1]. This is where some regularity of  $X_k$  with respect to the eigenstructure of  $A_{k-1}$  becomes essential. The technical part of our argument developed in most of the remaining sections is to realize and prove that a small amount of regularity encoded by (SR) or (WR) is already sufficient to control the solution to the inverse problem, and ultimately to control the spectral edges of  $A$ .

**1.10. Organization of the paper.** The rest of the paper is organized as follows. We start with the somewhat simpler Theorem 1.5 for the smallest eigenvalue in Section 2. A corresponding result for the largest eigenvalue, Theorem 3.1, is proved in Section 3. Corollary 1.4 is also deduced in Section 3. Combining Theorems 1.5 and 3.1 in Section 4, we obtain the main Theorem 1.1 on the spectral norm. In the Appendix, we prove Proposition 1.3 on the regularity of product distributions.

## 2. THE LOWER EDGE

We begin by proving Theorem 1.5 about the the lower edge of the spectrum, which is slightly simpler and requires fewer assumptions than the upper edge. As in [4], the tool that we use to do this is the *lower Stieltjes transform*

$$\underline{m}_A(\ell) = \text{tr}(A - \ell I)^{-1} = \sum_{i=1}^n (\lambda_i(A) - \ell)^{-1}, \quad \ell \in \mathbb{R}.$$

Note that  $\underline{m}_A(\ell) = -m_{-A}(-\ell)$  where  $m_A$  is the usual Stieltjes transform in (1.10).

For a sensitivity value  $\phi > 0$ , we define the *lower soft spectral edge*  $\ell_\phi(A)$  to be the smallest  $\ell$  for which

$$\underline{m}_A(\ell) = \phi.$$

Since  $\underline{m}_A(\ell)$  increases from 0 to  $\infty$  as  $\ell$  increases from  $-\infty$  to the lower spectral edge  $\lambda_{\min}(A)$ , the value  $\ell_\phi(A)$  is defined uniquely, and we always have the bound

$$\ell_\phi(A) < \lambda_{\min}(A).$$

For  $\phi \rightarrow \infty$  we have  $\ell_\phi(A) \rightarrow \lambda_{\min}(A)$ . However, we will work with small sensitivity  $\phi \in (0, 1)$ , which will make the soft spectral edge  $\ell_\phi(A)$  softer and easier to control.

The crucial property of  $\ell_\phi(A)$  is that it grows steadily under rank-one updates. Consider what happens when we add a random rank-one matrix  $XX^T$  to  $A \succ \ell I$ , where  $X$  is chosen from an isotropic distribution on  $\mathbb{R}^n$ . As  $\mathbb{E} \text{tr}(A + XX^T) = \text{tr}(A) + \text{tr} \mathbb{E} XX^T = \text{tr}(A) + n$ , we expect the eigenvalues of  $A + XX^T$  to have increased by 1 on average. It turns out that  $\ell_\phi(A)$  behaves almost as nicely as this if the distribution of  $X$  is sufficiently regular and the sensitivity  $\phi$  is sufficiently small. This is established in the following theorem.



**Theorem 2.1** (Random Lower Shift). *Suppose  $X$  is an isotropic random vector in  $\mathbb{R}^n$  satisfying the weak regularity assumption: for some  $C, \eta > 0$ ,*

$$(WR) \quad \sup_{\|x\| \leq 1} \mathbb{E} |\langle X, x \rangle|^{2+\eta} \leq C.$$

Let  $\varepsilon > 0$  and

$$\phi \leq c_{2.1} \varepsilon^{1+2/\eta},$$

where  $c_{2.1}^{-1} = 10(5C)^{2/\eta}$ . Then for every symmetric  $n \times n$  matrix  $A$  one has

$$\mathbb{E} \ell_\phi(A + XX^T) \geq \ell_\phi(A) + 1 - \varepsilon.$$

Iterating Theorem 2.1 easily yields a proof of Theorem 1.5 as follows.

*Proof of Theorem 1.5.* Let  $A_0 = 0$  and  $A_k = A_{k-1} + X_k X_k^T$  for  $k \leq N$ . Setting  $\phi = c_{2.1} \varepsilon^{1+2/\eta}$ , we find that

$$\ell_\phi(A_0) = \frac{-n}{\phi}.$$

Applying Theorem 2.1 inductively to  $A_0, A_1, \dots, A_N$ , we find that

$$\mathbb{E} [\ell_\phi(A_k) - \ell_\phi(A_{k-1}) | A_{k-1}] \geq 1 - \varepsilon \quad \text{for all } k \leq N,$$

where we take the conditional expectation with respect to the random vector  $X_k$  given the random vectors  $X_1, \dots, X_{k-1}$ , i.e. given  $A_{k-1}$ . Summing up these bounds yields

$$(2.1) \quad \mathbb{E} \ell_\phi(A_N) \geq \ell_\phi(A_0) + N(1 - \varepsilon).$$

Recalling that  $\lambda_{\min}(A_N) > \ell_\phi(A_N)$  and dividing both sides of (2.1) by  $N$ , we conclude that

$$\mathbb{E} \lambda_{\min} \left( \frac{1}{N} \sum_{i=1}^N X_i X_i^T \right) > \frac{\ell_\phi(A_0)}{N} + 1 - \varepsilon = 1 - \varepsilon - \frac{n}{\phi N}.$$

For  $N \geq n/\varepsilon\phi$ , the bound becomes  $1 - 2\varepsilon$ . Substituting the value of  $\phi$  and replacing  $\varepsilon$  by  $\varepsilon/2$  gives the promised result.  $\square$

The rest of this section is devoted to proving Theorem 2.1. Given a matrix  $A$ , a real number  $\ell < \lambda_{\min}(A)$ , and a vector  $x \in \mathbb{R}^n$ , we say that  $\delta \geq 0$  is a *feasible lower shift* if

$$A \succ (\ell + \delta)I \quad \text{and} \quad \underline{m}_{A+xx^T}(\ell + \delta) \leq \underline{m}_A(\ell).$$

The definition of the soft spectral edge  $\ell = \ell_\phi(A)$  along with monotonicity of the Stieltjes transform implies that

$$\ell_\phi(A + xx^T) \geq \ell_\phi(A) + \delta$$

for every feasible lower shift  $\delta$ . So we will be done if we can produce a feasible shift  $\delta$  such that  $\mathbb{E} \delta \geq 1 - \varepsilon$  where the expectation is over random  $X$ .

We begin by reducing the feasibility for a shift  $\delta$  to an inequality involving two quadratic forms. The following lemma appeared in [4], and we include it with proof for completeness.

**Lemma 2.2** (Feasible Lower Shift). *Consider numbers  $\ell \in \mathbb{R}$ ,  $\delta > 0$ , a matrix  $A \succ (\ell + \delta)I$  and a vector  $x$ . Then a sufficient condition for*

$$(2.2) \quad \underline{m}_{A+xx^T}(\ell + \delta) \leq \underline{m}_A(\ell)$$

is<sup>1</sup>

$$(2.3) \quad \frac{1}{\delta} \frac{x^T(A - \ell - \delta)^{-2}x}{\text{tr}(A - \ell - \delta)^{-2}} - x^T(A - \ell - \delta)^{-1}x =: \frac{1}{\delta}q_2(\delta, x) - q_1(\delta, x) \geq 1.$$

*Proof.* We begin by expanding  $\underline{m}_{A+xx^T}(\ell + \delta)$  using the Sherman-Morrisson formula:

$$\underline{m}_{A+xx^T}(\ell + \delta) = \text{tr}(A + xx^T - \ell - \delta)^{-1} = \text{tr}(A - \ell - \delta)^{-1} - \frac{x^T(A - \ell - \delta)^{-2}x}{1 + x^T(A - \ell - \delta)^{-1}x}.$$

Furthermore,

$$\text{tr}(A - \ell - \delta)^{-1} = \underline{m}_A(\ell) + \text{tr}[(A - \ell - \delta)^{-1} - (A - \ell)^{-1}].$$

The assumption  $A \succ (\ell + \delta)I$  implies that

$$(A - \ell - \delta)^{-1} - (A - \ell)^{-1} \preceq \delta(A - \ell - \delta)^{-2}.$$

Combining these estimates we see that (2.2) holds as long as

$$\delta \cdot \text{tr}(A - \ell - \delta)^{-2} - \frac{x^T(A - \ell - \delta)^{-2}x}{1 + x^T(A - \ell - \delta)^{-1}x} \leq 0$$

which we can rearrange into (2.3) observing that all quadratic forms involved are positive.  $\square$

The inequality (2.3) is quite nontrivial in the sense that  $\delta$  appears in many places, and it is not immediately clear from looking at it what the largest feasible  $\delta$  is given  $A, x$ , and  $\ell$ . In the following lemma, we present a tractable and explicit quantity defined solely in terms of  $q_1(0, x)$  and  $q_2(0, x)$  which always satisfies (2.3) and thus provides a lower bound on the best possible  $\delta$ .

**Lemma 2.3** (Explicit Feasible Shift). *Consider numbers  $\ell \in \mathbb{R}$ ,  $\phi > 0$ , a matrix  $A \succ \ell I$  satisfying  $\underline{m}_A(\ell) \leq \phi$ , and a vector  $x$ . Then for every  $t \in (0, 1)$ , the shift*

$$\delta := (1 - t)^3 q_2(0, x) \mathbf{1}_{\{q_1(0, x) \leq t\}} \mathbf{1}_{\{q_2(0, x) \leq t/\phi\}}$$

*satisfies  $A \succ (\ell + \delta)I$  and condition (2.3). Therefore  $\delta$  is a feasible lower shift, i.e.  $\underline{m}_{A+xx^T}(\ell + \delta) \leq \underline{m}_A(\ell)$ .*

The proof is based on regularity properties of the quadratic forms  $q_1$  and  $q_2$ , which we state in the following two lemmas.

**Lemma 2.4** (Regularity of Quadratic Forms). *Consider numbers  $\ell \in \mathbb{R}$ ,  $\phi > 0$ , a matrix  $A \succ \ell I$  satisfying  $\underline{m}_A(\ell) \leq \phi$ , and a vector  $x$ . Then for every positive number  $\delta < 1/\phi$ , one has  $A \succ (\ell + \delta)I$ , and moreover:*

- (i)  $q_1(0, x) \leq q_1(\delta, x) \leq (1 - \delta\phi)^{-1}q_1(0, x)$ ;
- (ii)  $(1 - \delta\phi)^2 q_2(0, x) \leq q_2(\delta, x) \leq (1 - \delta\phi)^{-2}q_2(0, x)$ .

---

<sup>1</sup>To ease the notation, we sometimes write  $A - u$  instead of  $A - uI$ .

*Proof.* The assumption  $A \succ \ell I$  states that all eigenvalues  $\lambda_i$  of  $A$  satisfy  $\lambda_i > \ell$ . Together with the assumption  $\underline{m}_A(\ell) = \sum_i (\lambda_i - \ell)^{-1} \leq \phi$  this implies that  $(\lambda_i - \ell)^{-1} \leq \phi$  for all  $i$ , hence  $\lambda_i - \ell \geq 1/\phi > \delta$  and  $A \succ (\ell + \delta)I$  as claimed.

(i) Let  $(\psi_i)_{i \leq n}$  denote the eigenvectors of  $A$ ; then

$$(2.4) \quad q_1(\delta, x) = \sum_{i=1}^n \frac{\langle x, \psi_i \rangle^2}{\lambda_i - \ell - \delta}.$$

Recalling that  $\lambda_i - \ell \geq 1/\phi$ , we have the comparison inequalities

$$(1 - \delta\phi)(\lambda_i - \ell) = \lambda_i - \ell - \phi\delta(\lambda_i - \ell) \leq \lambda_i - \ell - \delta \leq \lambda_i - \ell.$$

Using these for every term in (2.4) we complete the proof of (i).

(ii) Similar to (i), noting that the numerator and denominator of  $q_2$  are increasing in  $\delta$ .  $\square$

**Lemma 2.5** (Moments of Quadratic Forms). *Consider numbers  $\ell \in \mathbb{R}$ ,  $\phi > 0$  and a matrix  $A \succ \ell I$  satisfying  $\underline{m}_A(\ell) \leq \phi$ . If  $X$  is an isotropic random vector satisfying (WR), then for  $p = 1 + \eta/2$  the following moment bounds hold:*

- (i)  $\mathbb{E} q_1(0, X) = \underline{m}_A(\ell) \leq \phi$  and  $\mathbb{E} q_1(0, X)^p \leq C\phi^p$ ;
- (ii)  $\mathbb{E} q_2(0, X) = 1$  and  $\mathbb{E} q_2(0, X)^p \leq C$ .

*Proof.* (i) As in the proof of the previous lemma, let  $(\psi_i)_{i \leq n}$  denote the eigenvectors of  $A$ . By isotropy we have

$$\mathbb{E} q_1(0, X) = \sum_{i=1}^n \frac{\mathbb{E} \langle X, \psi_i \rangle^2}{\lambda_i - \ell} = \underline{m}_A(\ell) \leq \phi.$$

For the moment bound we use Minkowski's inequality to obtain

$$(\mathbb{E} q_1(0, X)^p)^{1/p} \leq \sum_{i=1}^n \frac{(\mathbb{E} \langle X, \psi_i \rangle^{2p})^{1/p}}{\lambda_i - \ell} \leq \sum_{i=1}^n \frac{C^{1/p}}{\lambda_i - \ell} = C^{1/p} \underline{m}_A(\ell) \leq C^{1/p} \phi.$$

(ii) Analogous to (i).  $\square$

We can now finish the proof of Lemma 2.3.

*Proof of Lemma 2.3.* First observe that by construction

$$(2.5) \quad \delta \leq q_2(0, x) \mathbf{1}_{\{q_2(0, x) \leq t/\phi\}} \leq t/\phi < 1/\phi,$$

so that we always have  $A \succ (\ell + \delta)I$  by Lemma 2.4.

If either of the indicators in the definition of the shift  $\delta$  is zero, then  $\delta = 0$ , which is trivially feasible and we are done. So assume both indicators are nonzero, that is  $q_1(0, x) \leq t$  and  $q_2(0, x) \leq t/\phi$ . By Lemma 2.2, it suffices to prove inequality (2.3), which is equivalent to

$$\frac{q_2(\delta, x)}{1 + q_1(\delta, x)} \geq \delta.$$

We can show this by replacing  $\delta$  with zero using Lemma 2.4:

$$\begin{aligned} \frac{q_2(\delta, x)}{1 + q_1(\delta, x)} &\geq \frac{q_2(0, x)(1 - \delta\phi)^2}{1 + q_1(0, x)(1 - \delta\phi)^{-1}} \\ &\geq \frac{q_2(0, x)(1 - t)^2}{1 + t(1 - t)^{-1}} \quad (\text{as } \delta\phi \leq t \text{ by (2.5) and } q_1(0, x) \leq t) \\ &= q_2(0, x)(1 - t)^3 = \delta. \end{aligned}$$

The proof is complete.  $\square$

We now complete the proof of Theorem 2.1 by using the regularity properties of  $X$  to show that the expectation of  $\delta$  as defined in Lemma 2.3 is large. Roughly speaking, this happens because (1)  $\delta$  is defined to be slightly less than  $q_2(0, X)$  whenever *both*  $q_1(0, X)$  and  $q_2(0, X)$  are not too large; (2) that event occurs with very high probability when  $\phi$  is sufficiently small (3) the expectation of  $q_2(0, X)$  equals 1.

*Proof of Theorem 2.1.* Let  $\ell = \ell_\phi(A)$ ; then  $\underline{m}_A(\ell) = \phi \leq c_{2.1}\varepsilon^{1+2/\eta}$  by assumption. Define a feasible shift  $\delta$  as in Lemma 2.3 for  $t = \varepsilon/5$ . Recall that it suffices to prove that  $\mathbb{E} \delta \geq 1 - \varepsilon$ .

According to Lemma 2.3,

$$\begin{aligned} \mathbb{E} \delta &= (1 - t)^3 \left[ \mathbb{E} q_2(0, X) - \mathbb{E} q_2(0, X) \mathbf{1}_{\{q_1(0, X) > t \vee q_2(0, X) > t/\phi\}} \right] \\ &\geq (1 - t)^3 \left[ 1 - (\mathbb{E} q_2(0, X)^p)^{\frac{1}{p}} \cdot (\mathbb{P}\{q_1(0, X) > t \vee q_2(0, X) > t/\phi\})^{\frac{1}{q}} \right] \end{aligned}$$

where we used Hölder's inequality with exponents  $p = 1 + \eta/2$  and  $q = \frac{p}{p-1} = 2/\eta + 1$ . By Lemma 2.5, we have  $\mathbb{E} q_2(0, X)^p \leq C$ . Next, the probability can be estimated by union bound, Markov's inequality and the moment bounds of Lemma 2.5, which gives

$$\begin{aligned} \mathbb{P}\{q_1(0, X) > t \vee q_2(0, X) > t/\phi\} &\leq \mathbb{P}\{q_1(0, X)^p > t^p\} + \mathbb{P}\{q_2(0, X)^p > (t/\phi)^p\} \\ &\leq \frac{C\phi^p}{t^p} + \frac{C}{(t/\phi)^p} = 2C(\phi/t)^p. \end{aligned}$$

We conclude that

$$\begin{aligned} \mathbb{E} \delta &\geq (1 - t)^3 \left[ 1 - C^{1/p} \cdot (2C(\phi/t)^p)^{1/q} \right] \\ &\geq (1 - t)^3 \left[ 1 - 2C(\phi/t)^{\frac{\eta}{2}} \right] \quad (\text{as } 1/p + 1/q = 1 \text{ and } p/q = \eta/2) \\ &= (1 - \varepsilon/5)^3 (1 - \varepsilon/5) \quad (\text{substituting } t \text{ and the bound for } \phi) \\ &\geq 1 - \varepsilon, \end{aligned}$$

as promised.  $\square$

### 3. THE UPPER EDGE

In this section we establish the following estimate for the expected largest eigenvalue, analogous to Theorem 1.5 for the smallest one.

**Theorem 3.1** (Largest eigenvalue). *Consider independent isotropic random vectors  $X_i$  valued in  $\mathbb{R}^n$ . Assume that  $X_i$  satisfy (SR) for some  $C, \eta > 0$ . Then, for  $\varepsilon \in (0, 1)$  and for*

$$N \geq C_{\text{upper}} \varepsilon^{-2-2/\eta} \cdot n$$

*the maximum eigenvalue of the sample covariance matrix  $\Sigma_N = \frac{1}{N} \sum_{i=1}^N X_i X_i^T$  satisfies*

$$(3.1) \quad \mathbb{E} \lambda_{\max}(\Sigma_N) \leq 1 + \varepsilon.$$

*Here  $C_{\text{upper}} := 512(16C)^{1+2/\eta}(6 + 6/\eta)^{1+4/\eta}$ .*

We shall control the largest eigenvalue of a symmetric matrix  $A$  using the (upper) *Stieltjes transform*

$$\bar{m}_A(u) = \text{tr}(uI - A)^{-1} = \sum_{i=1}^n (u - \lambda_i(A))^{-1}, \quad u \in \mathbb{R}.$$

Similarly to our argument for the lower edge, for a sensitivity value  $\phi > 0$  we define the *upper soft spectral edge*  $u_\phi(A)$  to be the largest  $u$  for which

$$\bar{m}_A(u) = \phi.$$

Since  $\bar{m}_A(u)$  decreases from  $\infty$  to 0 as  $u$  increases from the upper spectral edge  $\lambda_{\max}(A)$  to  $\infty$ , the value  $u_\phi(A)$  is defined uniquely, and

$$u_\phi(A) > \lambda_{\max}(A).$$

For  $\phi \rightarrow \infty$  we have  $u_\phi(A) \rightarrow \lambda_{\max}(A)$ , but as before we shall work with small sensitivity values  $\phi \in (0, 1)$ . Our goal is to show that  $u_\phi(A)$  increases by about 1 on average with every rank-one update.

**Theorem 3.2** (Random Upper Shift). *Suppose  $X$  is an isotropic random vector satisfying the strong regularity assumption (SR) for some  $C, \eta > 0$ . Assume  $\varepsilon \in (0, 1)$  and*

$$(3.2) \quad \phi \leq c_{3.2} \varepsilon^{1+2/\eta}$$

*where  $c_{3.2}^{-1} = 256(8C)^{1+2/\eta}(6 + 6/\eta)^{1+4/\eta}$ . Then for every symmetric matrix  $A$  one has*

$$(3.3) \quad \mathbb{E} u_\phi(A + XX^T) \leq u_\phi(A) + 1 + \varepsilon.$$

Iterating Theorem 3.2 yields a proof of Theorem 3.1.

*Proof of Theorem 3.1.* The argument is similar to the proof of Theorem 1.5 given in Section 2. We set  $\phi = \phi(\varepsilon) = c_{3.2} \varepsilon^{1+2/\eta}$ . Then we start with  $A_0 = 0$  where  $u_\phi(A_0) = n/\phi$  and inductively apply Theorem 3.2 for  $A_k = A_{k-1} + X_k X_k^T$  to obtain

$$\mathbb{E} \lambda_{\max}\left(\frac{1}{N} \sum_{i=1}^N X_i X_i^T\right) < \frac{u_\phi(A_0)}{N} + 1 + \varepsilon = 1 + \varepsilon + \frac{n}{\phi N}.$$

For  $N \geq n/\varepsilon\phi$ , the bound becomes  $1 + 2\varepsilon$ . Substituting the value of  $\phi$  and replacing  $\varepsilon$  by  $\varepsilon/2$  gives the promised result.  $\square$

The above proof works for  $\varepsilon, \phi(\varepsilon) < 1$  and thus for  $N = \Omega(n)$ , but it may be extended to smaller  $N$  as follows.

*Proof of Corollary 1.4.* In the proof of Theorem 3.1 we have shown that for every  $\varepsilon \in (0, 1)$  and every positive integer  $N$ , we have

$$E := \mathbb{E} \lambda_{\max}(\Sigma_N) < 1 + \varepsilon + \frac{n}{\phi(\varepsilon)N}$$

where  $\phi(\varepsilon) = c_{3.2} \varepsilon^{1+2/\eta}$ . Optimizing in  $\varepsilon$ , we apply this estimate with  $\varepsilon = (n/N)^{\frac{1}{2+2/\eta}}$  when  $n < N$  and with  $\varepsilon = 1/2$  when  $n \geq N$  to obtain:

$$\begin{aligned} E &< 1 + (1 + c_{3.2}^{-1}) \left(\frac{n}{N}\right)^{\frac{1}{2+2/\eta}} \quad \text{if } n < N, \\ E &< \frac{3}{2} + \frac{n}{\phi(1/2)N} \leq 1 + 2^{2+2/\eta} c_{3.2}^{-1} \left(\frac{n}{N}\right) \quad \text{if } n \geq N. \end{aligned}$$

Combining these, for every  $n$  and  $N$  we conclude that

$$E < 1 + (1 + c_{3.2}^{-1}) \left(\frac{n}{N}\right)^{\frac{1}{2+2/\eta}} + 2^{2+2/\eta} c_{3.2}^{-1} \left(\frac{n}{N}\right),$$

as required.

A similar bound for  $\mathbb{E} \lambda_{\min}(\Sigma_N)$  is immediate from Theorem 1.5 (see the remark after its proof).  $\square$

The rest of this section is devoted to proving Theorem 3.2. Given a matrix  $A$ , a real number  $u > \lambda_{\max}(A)$ , and a vector  $x \in \mathbb{R}^n$ , we say that  $\Delta \geq 0$  is a *feasible upper shift* if

$$(3.4) \quad A + xx^T \prec (u + \Delta)I \quad \text{and} \quad \overline{m}_{A+xx^T}(u + \Delta) \leq \overline{m}_A(u).$$

The definition of the soft spectral edge  $u = u_\phi(A)$  along with monotonicity of the Stieltjes transform implies that

$$(3.5) \quad u_\phi(A + xx^T) \leq u_\phi(A) + \Delta$$

for every feasible upper shift  $\Delta$ . So will be done if we can produce a feasible shift  $\Delta$  such that  $\mathbb{E} \Delta \leq 1 + \varepsilon$  where the expectation is over random  $X$ .

As in our argument for the lower edge, we begin by reducing the feasibility for a shift  $\delta$  to an inequality involving two quadratic forms.

**Lemma 3.3** (Feasible Upper Shift). *Consider numbers  $u \in \mathbb{R}$ ,  $\Delta > 0$ , a matrix  $A \prec uI$  and a vector  $x$ . Then a sufficient condition for  $\Delta \geq 0$  to be a feasible upper shift is*

$$(3.6) \quad \frac{x^T(u + \Delta - A)^{-2}x}{\overline{m}_A(u) - \overline{m}_A(u + \Delta)} + x^T(u + \Delta - A)^{-1}x =: Q_2(\Delta, x) + Q_1(\Delta, x) \leq 1.$$

*Proof.* Note that  $A \prec uI \prec (u + \Delta)I$  so that all quadratic forms are positive, and assume  $x \neq 0$  since otherwise the claim is trivial. As in the proof of Lemma 2.2, we use the Sherman-Morrisson formula to write:

$$\begin{aligned} \overline{m}_{A+xx^T}(u + \Delta) &= \text{tr}(u + \Delta - A - xx^T)^{-1} \\ &= \overline{m}_A(u + \Delta) + \frac{x^T(u + \Delta - A)^{-2}x}{1 - x^T(u + \Delta - A)^{-1}x} \\ &= \overline{m}_A(u) - (\overline{m}_A(u) - \overline{m}_A(u + \Delta)) + \frac{x^T(u + \Delta - A)^{-2}x}{1 - x^T(u + \Delta - A)^{-1}x}. \end{aligned}$$

Rearranging reveals that  $\overline{m}_{A+xx^T}(u + \Delta) \leq \overline{m}_A(u)$  exactly when (3.6) holds.

To establish the second condition

$$(3.7) \quad xx^T \prec u + \Delta - A,$$

we recall that

$$R \prec S \iff S^{-1/2}RS^{-1/2} \prec I$$

for all positive matrices  $R, S$  (this can be seen, for instance, using the Courant-Fischer theorem). Applying this fact to (3.7), we see that it suffices to have

$$(u + \Delta - A)^{-1/2}xx^T(u + \Delta - A)^{-1/2} \prec I$$

or equivalently

$$x^T(u + \Delta - A)^{-1}x < 1$$

which follows from (3.6) and  $Q_2(\Delta, x) > 0$ .  $\square$

We will reason about the two quantities  $Q_1$  and  $Q_2$  separately, producing two separate shifts  $\Delta_1$  and  $\Delta_2$  for them and eventually combining these into a single  $\Delta := \Delta_1 \vee \Delta_2$  as required by Lemma 3.3.

For some fixed parameter  $\tau \in (0, 1)$ , let us define  $\Delta_1 = \Delta_1(A, x, u)$  and  $\Delta_2 = \Delta_2(A, x, u)$  to be the smallest non-negative numbers such which satisfy

$$(3.8) \quad Q_1(\Delta_1, x) \leq \tau, \quad Q_2(\Delta_2, x) \leq 1 - \tau.$$

For  $u = u_\phi(A)$  and for a random vector  $x = X$ , Lemmas 3.4 and 3.6 will allow us to control the expected value of each of these shifts:

$$(3.9) \quad \mathbb{E} \Delta_1 \leq \varepsilon/2, \quad \mathbb{E} \Delta_2 \leq 1 + \varepsilon/2,$$

whenever the sensitivity parameter  $\phi = \phi(\tau, \varepsilon)$  is sufficiently small. From this we will obtain Theorem 3.2 quickly as follows.

*Proof of Theorem 3.2.* Let  $u_\phi(A) = u$ , so the condition  $A \prec uI$  of Lemma 3.3 holds. Consider the shifts  $\Delta_1 = \Delta_1(A, X, u)$  and  $\Delta_2 = \Delta_2(A, X, u)$  defined above. By (3.8), we have

$$Q_1(\Delta_1, X) + Q_2(\Delta_2, X) \leq 1.$$

Moreover, a quick inspection of the quadratic forms in Lemma 3.3 shows that  $Q_1(\Delta, X)$  and  $Q_2(\Delta, X)$  are decreasing in  $\Delta$ , hence

$$Q_1(\Delta_1 \vee \Delta_2, X) + Q_2(\Delta_1 \vee \Delta_2, X) \leq 1.$$

Then Lemma 3.3 guarantees that  $\Delta_1 \vee \Delta_2$  is a feasible upper shift, which implies by (3.5) that

$$u_\phi(A + XX^T) \leq u_\phi(A) + \Delta_1 \vee \Delta_2.$$

Furthermore, (3.9) yields a bound on the expected shift

$$\mathbb{E} \Delta_1 \vee \Delta_2 \leq \mathbb{E} \Delta_1 + \mathbb{E} \Delta_2 \leq 1 + \varepsilon,$$

which gives the conclusion (3.3) of Theorem 3.2.

It remains to note that Lemmas 3.4 and 3.6 only guarantee that the bounds (3.9) when the sensitivity  $\phi$  is sufficiently small, namely  $\phi \leq \phi_1(\tau, \varepsilon/2) \wedge \phi_2(\tau, \varepsilon/2)$ . With  $\tau = \varepsilon/16$ , we can simplify this inequality into the assumption of Theorem 3.2.  $\square$

The rest of this section is devoted to controlling the shifts  $\Delta_1$  and  $\Delta_2$ .

*Remark.* It is easy to check that the proofs of Lemmas 3.4 and 3.6 which follow, and consequently Theorem 3.2, only require

$$(3.10) \quad \mathbb{E} X_i X_i^T \prec cI,$$

for some constant  $c = c(\varepsilon)$ . Thus if we desire a bound of  $\lambda_{\max}(\frac{1}{N} \sum_{i=1}^N X_i X_i^T) < 1 + \varepsilon$  in Theorem 3.1, then  $\mathbb{E} X_i X_i^T = I$  can be replaced by the weaker condition (3.10).

### 3.1. Control of $\Delta_1$ .

**Lemma 3.4.** *Consider numbers  $u \in \mathbb{R}$ ,  $\phi > 0$  and a matrix  $A \prec uI$  satisfying  $\bar{m}_A(u) \leq \phi$ . Let  $X$  be a random vector satisfying (SR) for some  $C, \eta > 0$ , and let  $\varepsilon, \tau \in (0, 1)$ . If the sensitivity satisfies*

$$\phi \leq \phi_1(\tau, \varepsilon) := \frac{\tau^{1+1/\eta} \varepsilon^{1/\eta}}{(4C)^{1+1/\eta} (4 + 4/\eta)^{1+3/\eta}},$$

then the shift  $\Delta_1 = \Delta_1(A, X, u)$  satisfies

$$\mathbb{E} \Delta_1 \leq \varepsilon.$$

*Proof.* Let  $(\psi_i)_{i \leq n}$  and  $(\lambda_i)_{i \leq n}$  denote the eigenvectors and eigenvalues of  $A$ , and let  $\xi_i = \langle X, \psi_i \rangle^2$ . We know that  $\bar{m}_A(u) = \sum_{i=1}^n (u - \lambda_i)^{-1} \leq \phi$ , and  $\Delta_1$  is the smallest non-negative number satisfying

$$\sum_{i=1}^n \frac{\xi_i}{u - \lambda_i + \Delta_1} \leq \tau.$$

Rescaling everything by  $\phi$  and setting  $\mu_i := \phi(u - \lambda_i)$  so that

$$\sum_{i=1}^n \frac{1}{\mu_i} = \sum_{i=1}^n \frac{1}{\phi(u - \lambda_i)} \leq 1,$$

the problem becomes equivalent to bounding the least  $\mu := \phi \Delta_1$  for which

$$\sum_{i=1}^n \frac{1}{\mu_i + \mu} \leq \frac{\tau}{\phi}.$$

Applying the following somewhat more general probabilistic lemma to  $(\xi_i)_{i \leq n}$ , we conclude that

$$\mathbb{E} \Delta_1 \leq \frac{1}{\phi} \mathbb{E} \mu \leq \frac{1}{\phi} \frac{C(4 + 4/\eta)^{3+\eta} (4\phi)^{1+\eta}}{\tau^{1+\eta}}$$

whenever

$$\phi \leq \frac{\tau}{4C}.$$

Substituting  $\phi = \phi_1(\tau, \varepsilon)$  gives the promised bound. □

**Lemma 3.5.** *Suppose  $\{\xi_i\}_{i \leq n}$  are positive random variables with  $\mathbb{E} \xi_i = 1$  and:*

$$(3.11) \quad \mathbb{P} \left\{ \sum_{i \in S} \xi_i \geq t \right\} \leq \frac{C}{t^{1+\eta}} \quad \text{provided} \quad t > C|S| = C \sum_{i \in S} \mathbb{E} \xi_i.$$



for all subsets  $S \subset [n]$  and some constants  $C, \eta > 0$ . Consider positive numbers  $\mu_i$  such that

$$\sum_{i=1}^n \frac{1}{\mu_i} \leq 1.$$

Let  $\mu$  be the minimal positive number such that

$$\sum_{i=1}^n \frac{\xi_i}{\mu_i + \mu} \leq K,$$

for some  $K \geq 4C$ . Then

$$\mathbb{E} \mu \leq \frac{C(4 + 4/\eta)^{3+\eta}}{(K/4)^{1+\eta}}.$$

*Proof.* For simplicity of calculations, assume for the moment that the values of all  $\mu_i$  are dyadic, i.e.

$$\mu_i \in \{2^0, 2^1, 2^2, \dots\}.$$

For each dyadic number  $k$ , let

$$I_k := \{i : \mu_i = k\}, \quad n_k := |I_k|.$$

By assumptions, we have

$$1 \geq \sum_{i=1}^n \frac{1}{\mu_i} = \sum_{k \text{ dyadic}} \sum_{i \in I_k} \frac{1}{k} = \sum_{k \text{ dyadic}} \frac{n_k}{k},$$

and  $\mu$  is the smallest positive number such that

$$(3.12) \quad \sum_{i=1}^n \frac{\xi_i}{\mu_i + \mu} = \sum_{k \text{ dyadic}} \frac{1}{k + \mu} \sum_{i \in I_k} \xi_i \leq K.$$

We estimate  $\mu$  by replacing it with a bigger but easier quantity  $\mu'$ . Define  $\mu'$  to be the smallest positive number such that, for every dyadic  $k$ , one has

$$\frac{1}{k + \mu'} \sum_{i \in I_k} \xi_i \leq \varepsilon_k \quad \text{where} \quad \varepsilon_k := \frac{K}{2} \frac{n_k}{k} \vee \frac{K}{2\sigma} k^{-\frac{\eta}{2+2\eta}}$$

where

$$(3.13) \quad \sigma := \sum_{k \text{ dyadic}} k^{-\frac{\eta}{2+2\eta}} \leq \frac{2+2\eta}{\eta} \sum_{k \text{ dyadic}} \frac{1}{k} \leq 4 + 4/\eta.$$

Since

$$\sum_{k \text{ dyadic}} \frac{1}{k + \mu'} \sum_{i \in I_k} \xi_i \leq \sum_{k \text{ dyadic}} \varepsilon_k \leq \frac{K}{2} \sum_{k \text{ dyadic}} \frac{n_k}{k} + \frac{K}{2\sigma} \sum_{k \text{ dyadic}} k^{-\frac{\eta}{2+2\eta}} \leq K,$$

the definition of  $\mu$  given in (3.12) yields

$$\mu \leq \mu'.$$

It remains to bound  $\mathbb{E} \mu'$ .

By definition,

$$\mu' = \max_{k \text{ dyadic}} \left( \frac{1}{\varepsilon_k} \sum_{i \in I_k} \xi_i - k \right)_+.$$

Let  $\theta_k = \frac{1}{\varepsilon_k} \sum_{i \in I_k} \xi_i - k$ . For every  $t \geq 0$ , one has

$$\mathbb{P}\{\theta_k > t\} = \mathbb{P}\left\{\sum_{i \in I_k} \xi_i > (k+t)\varepsilon_k\right\}.$$

Since  $\varepsilon_k \geq \frac{Kn_k}{2k}$  by definition, we have

$$(k+t)\varepsilon_k \geq k\varepsilon_k \geq \frac{Kn_k}{2} = \frac{K}{2} \mathbb{E}\left(\sum_{i \in I_k} \xi_i\right) \geq C \mathbb{E}\left(\sum_{i \in I_k} \xi_i\right).$$

So by the regularity assumption (3.11)

$$\mathbb{P}\{\theta_k > t\} \leq \frac{C}{(k+t)^{1+\eta} \varepsilon_k^{1+\eta}}.$$

A union bound then gives

$$\begin{aligned} \mathbb{P}\{\mu' > t\} &\leq \sum_{k \text{ dyadic}} \frac{C}{(k+t)^{1+\eta} \varepsilon_k^{1+\eta}} \\ &\leq \frac{C}{(K/2\sigma)^{1+\eta}} \sum_{k \text{ dyadic}} \frac{k^{\eta/2}}{(k+t)^{1+\eta}} \quad (\text{by definition of } \varepsilon_k) \\ &\leq \frac{C}{(K/2\sigma)^{1+\eta}} \sum_{k \text{ dyadic}} \frac{1}{(k+t)^{1+\eta/2}}. \end{aligned}$$

This implies that

$$\begin{aligned} \mathbb{E} \mu' &= \int_0^\infty \mathbb{P}\{\mu' > t\} dt \leq \frac{C}{(K/2\sigma)^{1+\eta}} \sum_{k \text{ dyadic}} \int_0^\infty \frac{dt}{(k+t)^{1+\eta/2}} \\ &= \frac{C}{(K/2\sigma)^{1+\eta}} \sum_{k \text{ dyadic}} \frac{k^{-\eta/2}}{\eta/2} \\ &\leq \frac{C}{(K/2\sigma)^{1+\eta}} \frac{2}{\eta} \cdot \frac{4}{\eta} \quad (\text{by a calculation similar to (3.13)}) \\ &\leq \frac{C}{(K/2)^{1+\eta}} (4 + 4/\eta)^{3+\eta} \quad (\text{by (3.13)}). \end{aligned}$$

The promised bound for general (non-dyadic)  $\mu_i$  follows by rounding each  $\mu_i$  down to the nearest power of 2 and replacing  $K$  by  $K/2$ .  $\square$

*Remark* (Necessity of the strong regularity assumption (SR)). The preceding lemma is the only place in the proof where the full power of (SR) is used. To see that it is necessary, consider the following situation. Fix any  $S \subset [n]$  and let  $\frac{1}{\mu_i} = \mathbf{1}_{\{i \in S\}}/|S|$  so that  $\sum_i \frac{1}{\mu_i} = 1$ . Then the smallest  $\mu \geq 0$  for which  $\sum_i \frac{1}{\mu_i + \mu} \leq K$  is just

$$\mu = \left(\frac{1}{K} \sum_{i \in S} \xi_i - |S|\right)_+.$$

We now lowerbound the tail probability

$$\mathbb{P}\{\mu \geq t\} = \mathbb{P}\left\{\sum_{i \in S} \xi_i \geq K(|S| + t)\right\} \geq \mathbb{P}\left\{\sum_{i \in S} \xi_i \geq 2Kt\right\} \quad \text{for } t \geq |S|.$$

In order to have  $\mathbb{E} \mu = O(1)$ , this probability must be  $O(1/t)$  by Markov's inequality, which is essentially assumption (3.11) of the lemma. In the proof of Theorems 1.1 and 3.2, the sums of random variables  $\xi_i$  arise from projections of the random vector  $X$  onto varying eigenspaces of  $A$ ; the only succinct way to guarantee (3.11) for all such projections is essentially (SR).

### 3.2. Control of $\Delta_2$ .

**Lemma 3.6.** *Consider numbers  $u \in \mathbb{R}$ ,  $\phi > 0$  and a matrix  $A \prec uI$  satisfying  $\bar{m}_A(u) \leq \phi$ . Let  $X$  be a random vector satisfying (SR) for some  $C, \eta > 0$ , and let  $\varepsilon \in (0, 1)$ ,  $0 < \tau < \varepsilon/2$  be parameters. If the sensitivity satisfies*

$$\phi \leq \phi_2(\tau, \varepsilon) := \frac{\varepsilon^{2/\eta}(\varepsilon - 4\tau)}{128 \cdot (2C)^{2/\eta}(4 + 6/\eta)^{4/\eta}},$$

then the shift  $\Delta_2 = \Delta_2(A, X, u)$  satisfies

$$\mathbb{E} \Delta_2 \leq 1 + \varepsilon.$$

It will more convenient to work with the quadratic form

$$Q'_2(\Delta, x) := \frac{x^T(u + \Delta - A)^{-2}x}{\text{tr}(u + \Delta - A)^{-2}},$$

for which we have

$$(3.14) \quad \frac{1}{\Delta} Q'_2(\Delta, x) \geq Q_2(\Delta, x) \quad \text{for } \Delta > 0,$$

since the denominators satisfy:

$$\bar{m}_A(u) - \bar{m}_A(u + \Delta) = \text{tr}[(uI - A)^{-1} - (u + \Delta - A)^{-1}] \geq \Delta \text{tr}(u + \Delta - A)^{-2}.$$

*Remark.* The reason for working with  $Q_2$  rather than directly with  $Q'_2$  in Lemma 3.3 is that  $Q_2(\Delta, x)$  is decreasing in  $\Delta$ ; this monotonicity is required when arguing that the maximum of the two shifts  $\Delta = \Delta_1 \vee \Delta_2$  is feasible in the proof of Theorem 3.2.

We begin by recording some regularity properties of  $Q'_2(\Delta, X)$ .

**Lemma 3.7** (Regularity and Moments of  $Q'_2(\Delta, X)$ ). *Consider numbers  $u \in \mathbb{R}$ ,  $\phi > 0$  and a matrix  $A \prec uI$  satisfying  $\bar{m}_A(u) \leq \phi$ . Let  $X$  be a random vector satisfying (SR) for some  $C, \eta > 0$ . Then for every  $\Delta \geq 0$  one has:*

- (i)  $Q'_2(\Delta, X) \leq (1 + \phi\Delta)^2 Q'_2(0, X)$ ;
- (ii)  $\mathbb{E} Q'_2(\Delta, X) = 1$ ;
- (iii)  $\mathbb{E} Q'_2(\Delta, X)^p \leq C(3 + 3/\eta)$  for  $p = 1 + 2\eta/3$ .

*Proof.* (i) is analogous to Lemma 2.4. In a similar way, we show that all eigenvalues  $\lambda_i$  of  $A$  satisfy  $u - \lambda_i \geq 1/\phi$ , which implies the comparison inequality

$$u - \lambda_i \leq u + \Delta - \lambda_i \leq (1 + \phi\Delta)(u - \lambda_i).$$

Denoting  $(\psi_i)_{i \leq n}$  the eigenvectors of  $A$ , we express

$$(3.15) \quad Q'_2(\Delta, X) = \frac{\sum_{i=1}^n (u + \Delta - \lambda_i)^{-2} \langle X, \psi_i \rangle^2}{\sum_{i=1}^n (u + \Delta - \lambda_i)^{-2}}.$$

The comparison inequality yields (i).

(ii) We note that (3.15) can be rearranged as a convex combination of  $\langle X, \psi_i \rangle^2$ :

$$Q'_2(\Delta, X) = \sum_i \alpha_i \langle X, \psi_i \rangle^2 \quad \text{where} \quad \alpha_i \geq 0, \quad \sum_{i=1}^n \alpha_i = 1.$$

Then (ii) follows since  $\mathbb{E} \langle X, \psi_i \rangle^2 = 1$  by isotropy.

(iii) We apply Minkowski's inequality to obtain

$$(\mathbb{E} Q'_2(\Delta, X)^p)^{1/p} \leq \sum_{i=1}^n \alpha_i (\mathbb{E} \langle X, \psi_i \rangle^{2p})^{1/p}.$$

Now a simple integration of tails implies that each

$$\mathbb{E} \langle X, \psi_i \rangle^{2p} = \mathbb{E} \langle X, \psi_i \rangle^{2+4\eta/3} \leq C(3 + 3/\eta),$$

which concludes the proof.  $\square$

Next, we see how the regularity properties of  $Q'_2(\Delta, X)$  translate into the corresponding properties of  $\Delta_2$ :

**Lemma 3.8** (Regularity of  $\Delta_2$ ). *Consider numbers  $u \in \mathbb{R}$ ,  $\phi > 0$  and a matrix  $A \prec uI$  satisfying  $\bar{m}_A(u) \leq \phi$ . Let  $X$  be a random vector satisfying (SR) for some  $C, \eta > 0$ , and let  $0 < \tau < 1/2$ . Then the shift  $\Delta_2 = \Delta_2(A, X, u)$  satisfies:*

$$(i) \quad \mathbb{E} \Delta_2^{1+\eta/2} \leq 2^{1+\eta} C(4 + 6/\eta)^2;$$

$$(ii) \quad \mathbb{E} \Delta_2 \mathbf{1}_{\{Q'_2(0, X) \leq (t-2\tau)/8\phi\}} \leq 1 + t \text{ for every } t \in [0, 1].$$

*Proof.* (i) By definition of  $\Delta_2$  and using (3.14), we have for all  $t > 0$ :

$$\mathbb{P}\{\Delta_2 > t\} \leq \mathbb{P}\{Q_2(t, X) > 1 - \tau\} \leq \mathbb{P}\{Q'_2(t, X) > t(1 - \tau)\}.$$

This probability can be controlled using Lemma 3.7 (iii) and Markov's inequality, so we obtain

$$\mathbb{P}\{\Delta_2 > t\} \leq \frac{C(3 + 3/\eta)}{t^{1+2\eta/3}(1 - \tau)^{1+2\eta/3}} \leq \frac{C(3 + 3/\eta)}{(1/2)^{1+2\eta/3} t^{1+2\eta/3}}$$

as  $\tau < 1/2$ . Integration of tails yields

$$\mathbb{E} \Delta_2^{1+\eta/2} \leq 2^{1+2\eta/3} \cdot C(3 + 3/\eta)(4 + 6/\eta),$$

which implies the claim.

(ii) Let  $s_0$  denote the smaller solution of the quadratic equation

$$(1 + s\phi)^2 Q'_2(0, X) = s(1 - \tau),$$

whenever a solution exists. In this case  $s_0 > 0$  and Lemma 3.7 (i) yields that

$$Q'_2(s_0, X) \leq s_0(1 - \tau).$$

By (3.14), this yields  $Q_2(s_0, X) \leq s_0(1 - \tau)$ . By definition of  $\Delta_2$ , this in turn implies that

$$\Delta_2 \leq s_0.$$

An elementary calculation shows that if  $Q'_2(0, X) \leq (t - 2\tau)/8\phi$  then the solution  $s_0$  exists and satisfies

$$s_0 \leq (1 + t)Q'_2(0, X).$$

It follows that

$$\mathbb{E} s_0 \mathbf{1}_{\{Q'_2(0, X) \leq (t-2\tau)/8\phi\}} \leq (1 + t) \mathbb{E} Q'_2(0, X) = 1 + t,$$

where we used Lemma 3.7 (i) in the last step.  $\square$

We can now complete the proof of Lemma 3.6.

*Proof of Lemma 3.6.* We decompose

$$\mathbb{E} \Delta_2 = \mathbb{E} \Delta_2 \mathbf{1}_{\{Q'_2(0, X) \leq (t-2\tau)/8\phi\}} + \mathbb{E} \Delta_2 \mathbf{1}_{\{Q'_2(0, X) > (t-2\tau)/8\phi\}} =: E_1 + E_2.$$

By Lemma 3.8 (ii), we have  $E_1 \leq 1 + t$ . Next, we estimate  $E_2$  using Hölder's inequality:

$$E_2 \leq \left( \mathbb{E} \Delta_2^{1+\eta/2} \right)^{\frac{1}{1+\eta/2}} \left( \mathbb{P}\{Q'_2(0, X) > (t - 2\tau)/8\phi\} \right)^{\frac{\eta/2}{1+\eta/2}}.$$

The two terms here can be estimated using Lemma 3.8 (i) and Lemma 3.7 along with Markov's inequality:

$$\begin{aligned} E_2 &\leq \left( 2^{1+\eta} C(4 + 6/\eta)^2 \right)^{\frac{1}{1+\eta/2}} \left( \frac{C(3 + 3/\eta)}{((t - 2\tau)/8\phi)^{1+\eta/2}} \right)^{\frac{\eta/2}{1+\eta/2}} \\ &\leq 2^{1+\eta} C(4 + 6/\eta)^2 \cdot \left( \frac{8\phi}{t - 2\tau} \right)^{\eta/2}. \end{aligned}$$

Finally, we set  $t = \varepsilon/2$  and use the assumptions  $\phi \leq \phi_2(\tau, \varepsilon)$  and  $\tau < \varepsilon/2$  to conclude that  $E_2 \leq \varepsilon/2$ . Together with  $E_1 \leq 1 + t = 1 + \varepsilon/2$  this implies

$$\mathbb{E} \Delta_2 \leq 1 + \varepsilon$$

as claimed.  $\square$

*Remark.* Although for convenience of application Lemma 3.6 is stated under the strong regularity assumption (SR), the latter is not used in the proof. The argument above uses only the weak regularity assumption (WR).

#### 4. THE SPECTRAL NORM

In this section we prove Theorem 1.1 by showing that whenever  $X_1, \dots, X_N$  are independent and satisfy (SR), the spectral norm estimate

$$(4.1) \quad \mathbb{E} \|\Sigma_N - I\| \leq \varepsilon$$

follows from the spectral edge estimates

$$(4.2) \quad \mathbb{E} \lambda_{\min}(\Sigma_N) \geq 1 - \varepsilon/3; \quad \mathbb{E} \lambda_{\max}(\Sigma_N) \leq 1 + \varepsilon/3$$

obtained in Theorems 1.5 and 3.1. The basic idea is to show using independence that

$$\lambda_{\text{average}}(\Sigma_N) = \frac{1}{n} \text{tr}(\Sigma_N)$$

is concentrated near its expectation of 1. Combining this with

$$\mathbb{E} (\lambda_{\max}(\Sigma_N) - \lambda_{\min}(\Sigma_N)) \leq 2\varepsilon/3,$$

which follows immediately from (4.2), yields (4.1).

We rely on the following elementary proposition regarding sums of independent random variables.

**Proposition 4.1.** *Let  $Z_i$  be independent random variables with  $\mathbb{E} Z_i = 1$  and satisfying the following tail bounds for some  $C, \eta > 0$ :*

$$\mathbb{P}\{|Z_i| > t\} \leq Ct^{-1-\eta}, \quad t > 0.$$

If  $\varepsilon \in (0, 1)$  and

$$N \geq \frac{(2C)^{2/\eta}(1 + 1/\eta)^{2/\eta}}{(\varepsilon/2)^{2+2/\eta}},$$

then

$$\mathbb{E} \left| \frac{1}{N} \sum_{i=1}^N Z_i - 1 \right| \leq \varepsilon.$$

Postponing the proof of Proposition 4.1, we use this fact to control

$$\frac{1}{n} \operatorname{tr}(\Sigma_N) = \frac{1}{n} \sum_{i=1}^N \frac{\|X_i\|_2^2}{N}$$

and prove the main theorem as follows.

*Proof of Theorem 1.1.* Assume the random vectors  $X_i$  are isotropic and satisfy (SR) with parameters  $C, \eta$ . This implies that the random variables

$$Z_i = \frac{\|X_i\|_2^2}{n}$$

satisfy the requirements of Proposition 4.1 with parameters  $C^{1+\eta}, \eta$ . It follows that

$$(4.3) \quad \mathbb{E} \left| \frac{1}{n} \operatorname{tr}(\Sigma_N - I) \right| = \mathbb{E} \left| \frac{1}{N} \sum_{i=1}^N Z_i - 1 \right| \leq \varepsilon$$

whenever

$$(4.4) \quad N \geq \frac{(4C)^{2+2/\eta}(1 + 1/\eta)^{2/\eta}}{\varepsilon^{2+2/\eta}} =: \frac{C_{\operatorname{trace}}}{\varepsilon^{2+2/\eta}}.$$

Now consider the random variables

$$L = \lambda_{\min}(\Sigma_N - I), \quad U = \lambda_{\max}(\Sigma_N - I), \quad M = \frac{1}{n} \operatorname{tr}(\Sigma_N - I).$$

We have

$$L \leq M \leq U,$$

and we are interested in

$$(4.5) \quad \|\Sigma_N - I\| = U \vee -L \leq U - L + |M|.$$

When  $N \geq C_{\operatorname{upper}} n / \varepsilon^{2+2/\eta}$ , Theorem 3.1 gives  $\mathbb{E} U \leq \varepsilon$ . To show that  $\mathbb{E} L \geq \varepsilon$ , we recall that (SR) with parameters  $C, \eta$  implies (WR) with parameters  $C(2 + 2/\eta), \eta$  and invoke Theorem 1.5, noting that its requirement (1.8) is satisfied as

$$C_{\operatorname{upper}} = 512(16C)^{1+2/\eta}(6 + 6/\eta)^{1+4/\eta} > 40(10C(2 + 2/\eta))^{2/\eta} = C_{\operatorname{lower}}.$$

Now that we have both bounds  $\mathbb{E}U \leq \varepsilon$  and  $\mathbb{E}L \geq \varepsilon$ , we can combine them with (4.3) and (4.5), which yields

$$\mathbb{E} \|\Sigma_N - I\| \leq 2\varepsilon + \varepsilon,$$

whenever

$$(4.6) \quad N \geq C_{\text{upper}} \frac{n}{\varepsilon^{2+2/\eta}} \vee C_{\text{trace}} \frac{1}{\varepsilon^{2+2/\eta}}.$$

Replacing  $\varepsilon$  by  $\varepsilon/3$  and taking

$$N \geq C_{\text{main}} \frac{n}{\varepsilon^{2+2/\eta}}$$

where

$$C_{\text{main}} := 512 \cdot 3^{2+2/\eta} \cdot (16C)^{2+2/\eta} (6 + 6/\eta)^{1+4/\eta}$$

always satisfies (4.6). This completes the proof of the theorem.  $\square$

*Proof of Proposition 4.1.* Fix a parameter  $K > 0$ , and decompose

$$Z_i = Z_i \mathbf{1}_{\{|Z_i| \leq K\}} + Z_i \mathbf{1}_{\{|Z_i| > K\}} =: Z'_i + Z''_i.$$

Using  $\mathbb{E}Z'_i + \mathbb{E}Z''_i = \mathbb{E}Z_i = 1$  and by triangle inequality we obtain

$$\mathbb{E} \left| \frac{1}{N} \sum_{i=1}^N Z_i - 1 \right| \leq \mathbb{E} \left| \frac{1}{N} \sum_{i=1}^N Z'_i - \mathbb{E} \frac{1}{N} \sum_{i=1}^N Z'_i \right| + \mathbb{E} \left| \frac{1}{N} \sum_{i=1}^N Z''_i - \mathbb{E} \frac{1}{N} \sum_{i=1}^N Z''_i \right| =: E' + E''.$$

By Jensen's inequality, independence and the bound on  $Z'_i$ , we have

$$(E')^2 \leq \text{Var} \left( \frac{1}{N} \sum_{i=1}^N Z'_i \right) = \frac{1}{N^2} \sum_{i=1}^N \text{Var}(Z'_i) \leq \frac{K^2}{N}.$$

Moreover, by triangle and Jensen's inequalities,

$$E'' \leq 2 \mathbb{E} \left| \frac{1}{N} \sum_{i=1}^N Z''_i \right| \leq \frac{2}{N} \sum_{i=1}^N \mathbb{E} |Z''_i|.$$

The assumption on the tails of  $Z_i$  implies that  $\mathbb{P}\{|Z''_i| > t\} \leq C/(t \vee K)^{1+\eta}$  for  $t > 0$ , thus

$$\mathbb{E} |Z''_i| = \int_0^\infty \mathbb{P}\{|Z''_i| > t\} dt \leq \frac{C}{K^\eta} + \frac{C}{\eta K^\eta} = C \left(1 + \frac{1}{\eta}\right) K^{-\eta}.$$

Hence

$$E'' \leq 2C \left(1 + \frac{1}{\eta}\right) K^{-\eta},$$

and

$$E' + E'' \leq \frac{K}{\sqrt{N}} + 2C \left(1 + \frac{1}{\eta}\right) K^{-\eta}.$$

Choosing  $K = (\varepsilon/2)\sqrt{N}$  and using the assumption on  $N$ , one easily checks that

$$E' + E'' \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \leq \varepsilon$$

as desired.  $\square$

APPENDIX. PROOF OF PROPOSITION 1.3.

In this section we prove Proposition 1.3, which states that product distributions satisfy the regularity assumption in Theorem 1.1. Note that this result and its proof are not needed in the proof of Theorem 1.1.

Consider a random vector  $X$  and an orthogonal projection  $P$  in  $\mathbb{R}^n$  as in Proposition 1.3. Denoting by  $(P_{ij})$  the  $n \times n$  matrix of the operator  $P$ , we express

$$\|PX\|_2^2 = \langle X, PX \rangle = \sum_{i,j=1}^n \xi_i \xi_j P_{ij}.$$

The contribution of the diagonal of  $P$  to this sum is

$$D := \sum_{i=1}^n \xi_i^2 P_{ii}.$$

Denote by  $P_0$  the matrix  $P$  with diagonal removed; then

$$(4.7) \quad \|PX\|_2^2 - D = \langle X, P_0X \rangle.$$

We can estimate  $\langle X, P_0X \rangle$  using a standard decoupling argument. Let  $X'$  denote an independent copy of  $X$ , and let  $\mathbb{E}_X, \mathbb{E}_{X'}$  denote the expectations with respect to  $X$  and  $X'$  respectively. Since the matrix  $P_0$  has zero diagonal, we have<sup>2</sup>

$$(4.8) \quad \mathbb{E} |\langle X, P_0X \rangle|^p \lesssim \mathbb{E}_{X'} \mathbb{E}_X |\langle X, P_0X' \rangle|^p.$$

This inequality can be obtained from general decoupling results, see [7, Theorem 3.1.1]; a simple and well known proof of (4.8) is given in [17].

Next, an application of a standard symmetrization argument and Khintchine inequality (or a direct application of Rosenthal's inequality [13], see [8]) yields for every  $a \in \mathbb{R}^n$  that

$$\mathbb{E} |\langle X, a \rangle|^p = \mathbb{E} \left| \sum_{i=1}^n a_i \xi_i \right|^p \lesssim \|a\|_2^p.$$

Therefore, by conditioning on  $X'$  we obtain from (4.8) that

$$(4.9) \quad \mathbb{E} |\langle X, P_0X \rangle|^p \lesssim \mathbb{E}_{X'} \|P_0X'\|_2^p = \mathbb{E} \|P_0X\|_2^p.$$

Since  $P_0$  equals  $P$  without the diagonal, the triangle inequality yields

$$\|P_0X\|_2 \leq \|PX\|_2 + \left( \sum_{i=1}^n \xi_i^2 P_{ii}^2 \right)^{1/2}$$

Since  $0 < P_{ii} \leq \|P\| \leq 1$ , we can replace  $P_{ii}^2$  by  $P_{ii}$ , so

$$\|P_0X\|_2 \leq \|PX\|_2 + D^{1/2} \lesssim (\|PX\|_2^2 + D)^{1/2}.$$

Hölder's inequality then implies that

$$(4.10) \quad \mathbb{E} \|P_0X\|_2^p \lesssim (\mathbb{E} (\|PX\|_2^2 + D)^p)^{1/2}.$$

Putting (4.7), (4.9) and (4.10) together, we arrive at the inequality

$$\mathbb{E} \|\|PX\|_2^2 - D\|^p \lesssim (\mathbb{E} (\|PX\|_2^2 + D)^p)^{1/2}.$$

---

<sup>2</sup>Throughout this proof, we write  $a \lesssim b$  if  $a \leq Cb$  for some constant  $C$  which is independent of  $n$ .



Put in different words, the random variable  $Z := \|PX\|_2^2 - D$  satisfies the inequality

$$\|Z\|_{L_p}^2 \lesssim \|Z + 2D\|_{L_p} \leq \|Z\|_{L_p} + 2\|D\|_{L_p}.$$

Solving this quadratic inequality we obtain that

$$(4.11) \quad \|Z\|_{L_p} \lesssim 1 + \|D\|_{L_p}^{1/2}.$$

In order to bound  $\|D\|_{L_p}$  we consider

$$\|D - k\|_{L_p}^p = \mathbb{E} \left| \sum_{i=1}^n \xi_i^2 P_{ii} - k \right|^p = \mathbb{E} \left| \sum_{i=1}^n (\xi_i^2 - 1) P_{ii} \right|^p,$$

where we used that  $\sum_{i=1}^n P_{ii} = \text{tr}(P) = k$ . Recall that by the assumptions we have  $\mathbb{E}(\xi_i^2 - 1) = 0$  and  $\|\xi_i^2 - 1\|_{L_p} \leq \|\xi_i^2\|_{L_p} + 1 = \|\xi_i\|_{L_{2p}}^2 + 1 \lesssim 1$ . An application of Khintchine's inequality or Rosenthal's inequality (as before) and the bound  $P_{ii}^2 \leq P_{ii}$  yield that

$$(4.12) \quad \|D - k\|_{L_p}^p \lesssim \left( \sum_{i=1}^n P_{ii}^2 \right)^{p/2} \leq \left( \sum_{i=1}^n P_{ii} \right)^{p/2} = (\text{tr}(P))^{p/2} = k^{p/2}.$$

It follows that

$$\|D\|_{L_p} \leq \|D - k\|_{L_p} + k \lesssim k^{1/2} + k \lesssim k.$$

Putting this into (4.11), we see that

$$(4.13) \quad \|Z\|_{L_p} \lesssim k^{1/2}.$$

Finally, by definition of  $Z$  and using the triangle inequality and bounds (4.13), (4.12), we conclude that

$$\left\| \|PX\|_2^2 - k \right\|_{L_p} \leq \|Z\|_{L_p} + \|D - k\|_{L_p} \lesssim k^{1/2} + k^{1/2} \lesssim k^{1/2}.$$

Proposition 1.3 is proved. □

## REFERENCES

- [1] R. Adamczak, A. Litvak, A. Pajor, N. Tomczak-Jaegermann, *Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles*, J. Amer. Math. Soc. 23 (2010), 535–561.
- [2] R. Adamczak, A. Litvak, A. Pajor, N. Tomczak-Jaegermann, *Sharp bounds on the rate of convergence of the empirical covariance matrix*, submitted (2010).
- [3] Z. D. Bai, Y. Q. Yin, *Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix*, Ann. Probab. 21 (1993), 1275–1294.
- [4] J. Batson, D. Spielman, N. Srivastava, *Twice-Ramanujan Sparsifiers*, STOC 2009. SICOMP, to appear.
- [5] Z. D. Bai, J. Silverstein, Y. Q. Yin, *A note on the largest eigenvalue of a large-dimensional sample covariance matrix*, J. Multivariate Anal. 26 (1988), 166–168.
- [6] F. Benaych-Georges, R. Rao Nadakuditi, *The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices*, Advances in Mathematics 227 (2011), 494–521.
- [7] V. de la Pena, E. Gine, *Decoupling*. Probability and its Applications. Springer-Verlag, New York, 1999.
- [8] T. Figiel, P. Hitczenko, W. B. Johnson, G. Schechtman, *Extremal properties of Rademacher functions with applications to the Khintchine and Rosenthal inequalities*, Trans. Amer. Math. Soc. 349 (1997), 997–1027.
- [9] R. Kannan, L. Lovasz, M. Simonovits, *Random walks and an  $O^*(n^5)$  volume algorithm for convex bodies*, Random Structures Algorithms 11 (1997), 1–50.

- [10] R. Latała, *Some estimates of norms of random matrices*, Proc. Amer. Math. Soc. 133 (2005), 1273–1282.
- [11] G. Paouris, *Concentration of mass on convex bodies*, Geom. Funct. Anal. 16 (2006), 1021–1049.
- [12] M. Rudelson, *Random vectors in the isotropic position*, J. Funct. Anal. 164 (1999), 60–72.
- [13] H. Rosenthal, *On the subspaces of  $L^p$  ( $p > 2$ ) spanned by sequences of independent random variables*, Israel J. Math. 8 (1970), 273–303.
- [14] N. Srivastava, *Spectral Sparsification and Restricted Invertibility*, Ph.D. Thesis, March, 2010.
- [15] R. Vershynin, *How close is the sample covariance matrix to the actual covariance matrix?* Journal of Theoretical Probability, to appear (2010).
- [16] R. Vershynin, *Introduction to the non-asymptotic analysis of random matrices*. In: Compressed Sensing: Theory and Applications, eds. Yonina Eldar and Gitta Kutyniok. Cambridge University Press, to appear (2010). <http://www-personal.umich.edu/~romanv/papers/non-asymptotic-rmt-plain.pdf>
- [17] R. Vershynin, *A simple decoupling inequality in probability theory*, <http://www-personal.umich.edu/~romanv/papers/decoupling-simple.pdf> (2011).

SCHOOL OF MATHEMATICS, INSTITUTE FOR ADVANCED STUDY, 1 EINSTEIN DRIVE, PRINCETON, NJ 08540, U.S.A.

*E-mail address:* nikhils@math.ias.edu

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MICHIGAN, 530 CHURCH ST., ANN ARBOR, MI 48109, U.S.A.

*E-mail address:* romanv@umich.edu