The Annals of Probability
2013, Vol. 41, No. 5, 3081–3111
DOI: 10.1214/12-AOP760

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

BY NIKHIL SRIVASTAVA[1] AND ROMAN VERSHYNIN[2]

Institute for Advanced Study and University of Michigan

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 et al. [*J. Amer. Math. Soc.* **23** (2010) 535–561]. 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 spectral sparsifier, a deterministic tool developed recently by Batson, Spielman and Srivastava [*SIAM J. Comput.* **41** (2012) 1704–1721]. 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, that is, $\mathbb{E}X = 0$; this restriction will not be needed later. The covariance matrix of X is the $n \times n$ positive semidefinite matrix

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

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

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

Received June 2011; revised March 2012.
[1]Supported by NSF Grants CCF 0832797 and DMS-08-35373.
[2]Supported in part by NSF Grants FRG DMS-09-18623 and DMS-10-01829.
MSC2010 subject classifications. Primary 62H12; secondary 60B20.
Key words and phrases. Covariance matrices, high-dimensional distributions, Stieltjes transform, log-concave distributions, random matrices.

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

$$\mathbb{E}\|\Sigma_N - \Sigma\| \le \varepsilon \|\Sigma\|,$$

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 $\Sigma = I$, that is, to *isotropic distributions*.

1.2. *Sampling from isotropic distributions.* We consider independent isotropic random vectors X_i valued in $\mathbb{R}^n$, that is, 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

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

For obvious-dimensional reasons, one must have $N \ge n$. Rudelson's remarkably general result ([13], see [17], Section 4.3) yields that if $\|X\|_2 = O(\sqrt{n})$ almost surely, then

$$N = O(n \log n), \tag{1.1}$$

where the $O(\cdot)$ notation hides the dependence on ε 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

$$\sup_{\|x\|_2 \le 1} \left(\mathbb{E}\left|\langle X, x\rangle\right|^p\right)^{1/p} = O\left(\sqrt{p}\right) \qquad \text{for } p \ge 1. \tag{1.2}$$

This result can be obtained by a standard covering argument; see [17], Section 4.3.

It is an open problem to describe the distributions for which the logarithmic oversampling is not needed, that is, 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

Adamczak et al. [1], who showed that $N = O(n)$ still holds (in fact, with high probability) under the *sub-exponential* moment assumptions

$$\begin{aligned} &\|X\|_2 = O\big(\sqrt{n}\big) \qquad \text{a.s.},\\ &\sup_{\|x\|_2 \le 1} \big(\mathbb{E}\big|\langle X, x\rangle\big|^p\big)^{1/p} = O(p) \qquad \text{for } p \ge 1. \end{aligned} \tag{1.3}$$

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 Kannan, Lovasz and Simonovits in [9].

The second author of the present paper speculated in [16] 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*

$$\mathbb{P}\big\{\|PX_i\|_2^2 > t\big\} \le Ct^{-1-\eta} \qquad \text{for } t > C\operatorname{rank}(P) \tag{SR}$$

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

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

one has

$$\mathbb{E}\left\|\frac{1}{N}\sum_{i=1}^N X_i X_i^T - I\right\| \le \varepsilon. \tag{1.4}$$

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 $\mathbb{E}\|PX_i\|_2^2 = \operatorname{rank}(P)$. This explains why (SR) concerns only the tail values of t which are above $\operatorname{rank}(P)$.

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* Σ. *Assume that for some* $C, \eta > 0$, *the isotropic random vector* $Z = \Sigma^{-1/2}X$ *satisfies*

$$\mathbb{P}\big\{\|PZ\|_2^2 > t\big\} \le Ct^{-1-\eta} \qquad \text{for } t > C\operatorname{rank}(P) \tag{SR}$$

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 Σ_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 $\mathbb{R}^n$, that is, whose density has the form $e^{-V(x)}$ where $\log V(x)$ is a convex function on $\mathbb{R}^n$. Paouris's concentration inequality [11] implies that regularity assumption (SR) holds for X. Indeed, consider an orthogonal projection P in $\mathbb{R}^n$, and let $k = \operatorname{rank}(P)$. 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 $\operatorname{range}(P)$. Paouris's theorem then states that

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

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 Adamczak et al. [1], though it should be noted that the guarantee of [1] holds with probability that converges to 1 exponentially fast as $n \to \infty$, 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 cannot 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 $X = (\xi_1, \ldots, \xi_n)$, where ξ_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* rank $P = k$, *one has*

$$\mathbb{E}\big|\|PX\|_2^2 - k\big|^p \lesssim k^{p/2}. \tag{1.5}$$

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

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

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—according to [3], if the components ξ_i are i.i.d. and have infinite fourth moment, then $\limsup \|\Sigma_N\| \to \infty$ as $n \to \infty$ and $n/N \to y > 0$. (This is because in this situation at least one of the Nn i.i.d. coordinates of $X_1, \ldots, X_N$ will likely to be large.)

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*

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

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 Σ_N, respectively.

We deduce this result in Section 3. One can view (1.6) as a nonasymptotic form of the *Bai–Yin law* for the extreme eigenvalues of sample covariance matrices [4]. This law, associated with the works of Geman, Bai, Yin, Krishnaiah and Silverstein applies for product distributions, specifically for random vectors $X = (\xi_1, \ldots, \xi_n)$ with i.i.d. components ξ_i with zero mean, unit variance and finite fourth moment. For such distributions one has asymptotically almost surely that

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

as $n \to \infty$ and $n/N \to y \in [0, 1)$; see the rigorous statement in [4]. This limit law is sharp. On the other hand, 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], 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 $C \le (O(p))^p$, Theorem 1.1 recovers a bound of $c = 1/2 - 1/(p+1) = 1/2 - o(1)$ as $p \to \infty$.

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 Σ_N, and they are controlled as in (1.6). In particular, under the regularity assumption (SR) on X_i we obtain that

$$\bigl(\mathbb{E}\|A\|^2\bigr)^{1/2} \le C_2\bigl(\sqrt{N} + \sqrt{n}\bigr),$$

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 Latala's general inequality [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 following* weak regularity assumption: *for some $C, \eta > 0$,*

$$\sup_{\|x\|_2 \le 1} \mathbb{E}\bigl|\langle X_i, x\rangle\bigr|^{2+\eta} \le C. \tag{WR}$$

Then, for $\varepsilon > 0$ and for

$$N \geq C_{\text{lower}} \varepsilon^{-2-2/\eta} \cdot n, \tag{1.8}$$

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

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 $X = (\xi_1, \ldots, \xi_n)$ have i.i.d. components ξ_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 [12], see [8]) shows that one-dimensional marginals of X also have bounded $(2 + \eta)$ moments; that is, (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 already recalled, four moments are necessary to control the *largest* eigenvalue in the classical Bai–Yin law [3].

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:

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

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 ξ 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 ξ produces a dimension-free tail decay of the norm of Z, namely $\mathbb{P}\{\|X\|_2 > t\sqrt{n}\} = \mathbb{P}\{\xi > t\} \gtrsim \exp(-C't^2)$ for $t > 0$. It follows that a sample of N independent copies $X_1, \dots, X_N$ of X safisfies $\mathbb{E}\max_{i\le N} \|X_i\|_2^2 \gtrsim N \log N$, so the matrix $\Sigma_N = \frac{1}{N}\sum_{i=1}^N X_i X_i^T$ satisfies

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

which contradicts the conclusion of Theorem 1.1. This example is essentially due to 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

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

A conjecture of this type is discussed in [16] where a version of the theorem 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 cannot drop the regularity assumption (WR); that is, 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 Σ_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 *spectral sparsifier* invented by Batson, Spielman and Srivastava [5]; see [14]. The randomization makes the spectral sparsifier appear naturally in the context of random matrix theory. The method is based on evaluating the Stieltjes transform of Σ_N while making rank one updates. However, in contrast to typical methods of random matrix theory (and to the spectral sparsifier itself), 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 Σ_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

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

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 that satisfies

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

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

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 [5], 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; u_N 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:

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

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 [5], which gives for every $u \in \mathbb{R}$ that

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

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 nonlinear inverse problem. Indeed, for a fixed u it is not difficult to compute the expectation of $m_{A_k}(u)$ from (1.13), and in particular to bound the expectation by ϕ; this is done in [5]. However, we require the identity $m_{A_k}(u) = \phi$ to hold *deterministically*, because the largest u that satisfies it defines the soft spectral edge of A_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 nonlinear 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. A technical part of our argument developed in most of the remaining sections is to realize and prove that a small amount or 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 [5], the tool that we use to do this is the *lower Stieltjes transform*

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

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 ℓ for which

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

Since $\underline{m}_A(\ell)$ increases from 0 to ∞ as ℓ 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 $\phi \to \infty$ we have $\ell_\phi(A) \to \lambda_{\min}(A)$. 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 $\mathbb{E}\operatorname{tr}(A + XX^T) = \operatorname{tr}(A) + \operatorname{tr}\mathbb{E}XX^T = \operatorname{tr}(A) + n$, 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 ϕ 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 $C, \eta > 0$,*

$$\sup_{\|x\| \le 1} \mathbb{E}\big|\langle X, x\rangle\big|^{2+\eta} \le C. \tag{WR}$$

Let $\varepsilon > 0$ and

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

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

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

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 $A_0, A_1, \ldots, A_N$, we find that

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

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

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

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 ϕ and replacing ε by $\varepsilon/2$ gives the promised result. □

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

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

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

We begin by reducing the feasibility for a shift δ to an inequality involving two quadratic forms. The following lemma appeared in [5], and we include it with a proof for completeness.

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

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

is[3]

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

PROOF. We begin by expanding $\underline{m}_{A+xx^T}(\ell+\delta)$ using the Sherman–Morisson formula,

$$\begin{aligned}\underline{m}_{A+xx^T}(\ell+\delta) &= \operatorname{tr}\bigl(A+xx^T-\ell-\delta\bigr)^{-1}\\ &= \operatorname{tr}(A-\ell-\delta)^{-1} - \frac{x^T(A-\ell-\delta)^{-2}x}{1+x^T(A-\ell-\delta)^{-1}x}.\end{aligned}$$

Furthermore,

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

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

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

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

Inequality (2.3) is quite nontrivial in the sense that δ appears in many places, and it is not immediately clear from looking at it what the largest feasible δ is given A, x and ℓ. 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 δ.

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

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

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

[3]To ease the notation, we sometimes write $A-u$ instead of $A-uI$.

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

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

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

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

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

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

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

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 δ. □

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

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

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

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

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

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

(ii) Analogous to (i). □

We can now finish the proof of Lemma 2.3.

PROOF OF LEMMA 2.3. First observe that by construction,

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

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

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

We can show this by replacing δ with zero using Lemma 2.4:

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

The proof is complete. □

We now complete the proof of Theorem 2.1 by using the regularity properties of X to show that the expectation of δ, as defined in Lemma 2.3, is large. Roughly speaking, this happens because (1) δ 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 ϕ is sufficiently small; (3) the expectation of $q_2(0,X)$ equals 1.

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

According to Lemma 2.3,

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

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 $\mathbb{E}q_2(0,X)^p \le C$. Next, the probability can be estimated by a union bound, Markov's inequality and the moment bounds of Lemma 2.5, which gives

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

We conclude that

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

as promised. □

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

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

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*

$$\overline{m}_A(u) = \operatorname{tr}(uI-A)^{-1} = \sum_{i=1}^n \big(u-\lambda_i(A)\big)^{-1}, \qquad u\in\mathbb{R}.$$

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

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

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

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

For $\phi\to\infty$ we have $u_\phi(A)\to\lambda_{\max}(A)$, 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*

$$\phi \le c_{3.2}\varepsilon^{1+2/\eta}, \tag{3.2}$$

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

$$\mathbb{E}u_\phi\big(A + XX^T\big) \le u_\phi(A) + 1 + \varepsilon. \tag{3.3}$$

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 we 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 $N \ge n/\varepsilon\phi$, the bound becomes $1 + 2\varepsilon$. Substituting the value of ϕ and replacing ε by $\varepsilon/2$ gives the promised result. □

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 ε, we apply this estimate with $\varepsilon = (n/N)^{1/(2+2/\eta)}$ when $n < N$ and with $\varepsilon = 1/2$ when $n \ge N$ to obtain

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

$$E < \frac{3}{2} + \frac{n}{\phi(1/2)N} \le 1 + 2^{2+2/\eta}c_{3.2}^{-1}\left(\frac{n}{N}\right) \qquad \text{if } n \ge N.$$

Combining these, for every n and N we conclude that

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

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

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

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

$$u_\phi\big(A + xx^T\big) \leq u_\phi(A) + \Delta \tag{3.5}$$

for every feasible upper shift Δ. So will be done if we can produce a feasible shift Δ 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 δ to an inequality involving two quadratic forms.

LEMMA 3.3 (Feasible upper shift). *Consider the 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*

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

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–Morisson formula to write

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

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

$$xx^T \prec u + \Delta - A, \tag{3.7}$$

we recall that

$$R \prec S \quad \Longleftrightarrow \quad S^{-1/2} R S^{-1/2} \prec I$$

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

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

or equivalently

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

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

We will reason about the two quantities Q_1 and Q_2 separately, producing two separate shifts Δ_1 and Δ_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 nonnegative numbers such which satisfy

$$Q_1(\Delta_1, x) \le \tau, \qquad Q_2(\Delta_2, x) \le 1 - \tau. \tag{3.8}$$

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, so

$$\mathbb{E}\Delta_1 \le \varepsilon/2, \qquad \mathbb{E}\Delta_2 \le 1 + \varepsilon/2, \tag{3.9}$$

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

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

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 Δ, and hence

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

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

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

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

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

which gives 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) hold when the sensitivity ϕ is sufficiently small, namely $\phi \le \phi_1(\tau, \varepsilon/2) \wedge \phi_2(\tau, \varepsilon/2)$. With $\tau = \varepsilon/16$, we can simplify this inequality into the assumption of Theorem 3.2. □

The rest of this section is devoted to controlling the shifts Δ_1 and Δ_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

$$\mathbb{E}X_i X_i^T \prec cI \tag{3.10}$$

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* Δ_1.

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

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

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

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

PROOF. Let $(\psi_i)_{i\le n}$ and $(\lambda_i)_{i\le n}$ denote the eigenvectors and eigenvalues of A, and let $\xi_i = \langle X, \psi_i\rangle^2$. We know that $\overline{m}_A(u) = \sum_{i=1}^n (u-\lambda_i)^{-1} \le \phi$, and Δ_1 is the smallest nonnegative number satisfying

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

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

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

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

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

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

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

whenever

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

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

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

$$\mathbb{P}\left\{\sum_{i\in S}\xi_i \ge t\right\} \le \frac{C}{t^{1+\eta}} \quad \text{provided} \quad t > C|S| = C\sum_{i\in S}\mathbb{E}\xi_i \tag{3.11}$$

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

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

Let μ be the minimal positive number such that

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

for some $K \ge 4C$. Then

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

PROOF. For simplicity of calculations, assume for the moment that the values of all μ_i are dyadic, that is,

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

For each dyadic number k, let

$$I_k := \{i : u_i = k\}, \qquad n_k := |I_k|.$$

By assumption, we have

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

and μ is the smallest positive number such that

$$\sum_{i=1}^{n}\frac{\xi_i}{\mu_i+\mu} = \sum_{k \text{ dyadic}}\frac{1}{k+\mu}\sum_{i\in I_k}\xi_i \le K. \tag{3.12}$$

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

$$\frac{1}{k+\mu'}\sum_{i\in I_k}\xi_i \le \varepsilon_k \qquad \text{where } \varepsilon_k := \frac{K}{2}\frac{n_k}{k} \vee \frac{K}{2\sigma}k^{-\eta/(2+2\eta)},$$

where

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

Since

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

the definition of μ given in (3.12) yields

$$\mu \le \mu'.$$

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 $t \ge 0$, one has

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

Since $\varepsilon_k \ge \frac{K n_k}{2k}$ by definition, we have

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

So by regularity assumption (3.11),

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

A union bound then gives

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

This implies that

$$\begin{aligned} \mathbb{E}\mu' = \int_0^\infty \mathbb{P}\{\mu' > t\}\, dt &\le \frac{C}{(K/2\sigma)^{1+\eta}} \sum_{k \text{ dyadic}} \int_0^\infty \frac{dt}{(k+t)^{1+\eta/2}} \\ &= \frac{C}{(K/2\sigma)^{1+\eta}} \sum_{k \text{ dyadic}} \frac{k^{-\eta/2}}{\eta/2} \end{aligned}$$

$$\leq \frac{C}{(K/2\sigma)^{1+\eta}} \frac{2}{\eta} \cdot \frac{4}{\eta} \qquad \text{[by a calculation similar to (3.13)]}$$

$$\leq \frac{C}{(K/2)^{1+\eta}} (4 + 4/\eta)^{3+\eta} \qquad \text{[by (3.13)].}$$

The promised bound for general (nondyadic) μ_i follows by rounding each μ_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

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

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 ξ_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* Δ_2.

LEMMA 3.6. *Consider numbers* $u \in \mathbb{R}$, $\phi > 0$ *and a matrix* $A \prec uI$ *satisfying* $\overline{m}_A(u) \leq \phi$. *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 be more convenient to work with the quadratic form

$$Q_2'(\Delta, x) := \frac{x^T (u + \Delta - A)^{-2} x}{\operatorname{tr}(u + \Delta - A)^{-2}},$$

for which we have

$$\frac{1}{\Delta}Q_2'(\Delta, x) \geq Q_2(\Delta, x) \qquad \text{for } \Delta > 0, \tag{3.14}$$

since the denominators satisfy

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

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 Δ; 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 of $Q_2'(\Delta, X)$]. *Consider numbers $u \in \mathbb{R}$, $\phi > 0$ and a matrix $A \prec uI$ satisfying $\overline{m}_A(u) \leq \phi$. 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 λ_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

$$Q_2'(\Delta, X) = \frac{\sum_{i=1}^n (u + \Delta - \lambda_i)^{-2} \langle X, \psi_i\rangle^2}{\sum_{i=1}^n (u + \Delta - \lambda_i)^{-2}}. \tag{3.15}$$

The comparison inequality yields (i).

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

$$Q_2'(\Delta, X) = \sum_i \alpha_i \langle X, \psi_i\rangle^2 \qquad \text{where } \alpha_i \geq 0, \sum_{i=n} \alpha_i = 1.$$

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

(iii) We apply Minkowski's inequality to obtain

$$\big(\mathbb{E}Q_2'(\Delta, X)^p\big)^{1/p} \leq \sum_{i=1}^n \alpha_i \big(\mathbb{E}\langle X, \psi_i\rangle^{2p}\big)^{1/p}.$$

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. □

Next, we see how the regularity properties of $Q_2'(\Delta, X)$ translate into the corresponding properties of Δ_2:

LEMMA 3.8 (Regularity of Δ_2). *Consider numbers $u \in \mathbb{R}$, $\phi > 0$ and a matrix $A \prec uI$ satisfying $\overline{m}_A(u) \le \phi$. 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) $\mathbb{E}\Delta_2^{1+\eta/2} \le 2^{1+\eta} C(4 + 6/\eta)^2$;
(ii) $\mathbb{E}\Delta_2 \mathbf{1}_{\{Q_2'(0,X) \le (t-2\tau)/8\phi\}} \le 1 + t$ *for every* $t \in [0, 1]$.

PROOF. (i) By definition of Δ_2 and using (3.14), we have for all $t > 0$,

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

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

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

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

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

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

$$Q_2'(s_0, X) \le s_0(1 - \tau).$$

By (3.14), this yields $Q_2(s_0, X) \le s_0(1 - \tau)$. By definition of Δ_2, this in turn implies that

$$\Delta_2 \le s_0.$$

An elementary calculation shows that if $Q_2'(0, X) \le (t - 2\tau)/8\phi$, then the solution s_0 exists and satisfies

$$s_0 \le (1 + t) Q_2'(0, X).$$

It follows that

$$\mathbb{E} s_0 \mathbf{1}_{\{Q_2'(0,X) \le (t-2\tau)/8\phi\}} \le (1 + t)\mathbb{E} Q_2'(0, X) = 1 + t,$$

where we used Lemma 3.7(i) in the last step. □

We can now complete the proof of Lemma 3.6.

PROOF OF LEMMA 3.6. We decompose

$$\mathbb{E}\Delta_2 = \mathbb{E}\Delta_2 \mathbf{1}_{\{Q_2'(0,X)\le (t-2\tau)/8\phi\}} + \mathbb{E}\Delta_2 \mathbf{1}_{\{Q_2'(0,X)>(t-2\tau)/8\phi\}} =: E_1 + E_2.$$

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

$$E_2 \le \bigl(\mathbb{E}\Delta_2^{1+\eta/2}\bigr)^{1/(1+\eta/2)} \bigl(\mathbb{P}\{Q_2'(0,X) > (t-2\tau)/8\phi\}\bigr)^{(\eta/2)/(1+\eta/2)}.$$

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

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

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

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

as claimed. □

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

$$\mathbb{E}\|\Sigma_N - I\| \le \varepsilon \tag{4.1}$$

follows from the spectral edge estimates

$$\mathbb{E}\lambda_{\min}(\Sigma_N) \ge 1-\varepsilon/3; \qquad \mathbb{E}\lambda_{\max}(\Sigma_N) \le 1+\varepsilon/3 \tag{4.2}$$

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

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

is concentrated near its expectation of 1. Combining this with

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

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 $C, \eta > 0$:*

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

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

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

then

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

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, η. 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

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

whenever

$$N \ge \frac{(4C)^{2+2/\eta}(1+1/\eta)^{2/\eta}}{\varepsilon^{2+2/\eta}} =: \frac{C_{\text{trace}}}{\varepsilon^{2+2/\eta}}. \tag{4.4}$$

Now consider the random variables

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

We have

$$L \le M \le U,$$

and we are interested in

$$\|\Sigma_N - I\| = U \vee -L \le U - L + |M|. \tag{4.5}$$

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

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

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

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

Replacing ε 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. □

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

$$\begin{aligned}\mathbb{E}\left|\frac{1}{N}\sum_{i=1}^N Z_i - 1\right| &\leq \mathbb{E}\left|\frac{1}{N}\sum_{i=1}^N Z_i' - \mathbb{E}\frac{1}{N}\sum_{i=1}^N Z_i'\right| + \mathbb{E}\left|\frac{1}{N}\sum_{i=1}^N Z_i'' - \mathbb{E}\frac{1}{N}\sum_{i=1}^N Z_i''\right| \\ &=: E' + E''.\end{aligned}$$

By Jensen's inequality, independence and the bound on Z_i', we have

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

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

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

and

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

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

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

as desired. □

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

$$\|PX\|_2^2 - D = \langle X, P_0 X\rangle. \tag{A.1}$$

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[4]

$$\mathbb{E}\big|\langle X, P_0X\rangle\big|^p \lesssim \mathbb{E}_{X'}\mathbb{E}_X\big|\langle X, P_0X'\rangle\big|^p. \tag{A.2}$$

This inequality can be obtained from general decoupling results; see [7], Theorem 3.1.1; a simple and well-known proof of (A.2) is given in [15].

[4]Throughout this proof, we write $a \lesssim b$ if $a \le Cb$ for some constant C which is independent of n.

Next, an application of a standard symmetrization argument and Khintchine inequality (or a direct application of Rosenthal's inequality [12], 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 (A.2) that

$$\mathbb{E}|\langle X, P_0 X\rangle|^p \lesssim \mathbb{E}_{X'}\|P_0 X'\|_2^p = \mathbb{E}\|P_0 X\|_2^p. \tag{A.3}$$

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

$$\|P_0 X\|_2 \le \|PX\|_2 + \left(\sum_{i=1}^n \xi_i^2 P_{ii}^2\right)^{1/2}.$$

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

$$\|P_0 X\|_2 \le \|PX\|_2 + D^{1/2} \lesssim \left(\|PX\|_2^2 + D\right)^{1/2}.$$

Hölder's inequality then implies that

$$\mathbb{E}\|P_0 X\|_2^p \lesssim \left(\mathbb{E}\left|\|PX\|_2^2 + D\right|^p\right)^{1/2}. \tag{A.4}$$

Putting (A.1), (A.3) and (A.4) together, we arrive at the inequality

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

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

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

Solving this quadratic inequality we obtain that

$$\|Z\|_{L_p} \lesssim 1 + \|D\|_{L_p}^{1/2}. \tag{A.5}$$

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

$$\|D - k\|_{L_p}^p \lesssim \left(\sum_{i=1}^n P_{ii}^2\right)^{p/2} \le \left(\sum_{i=1}^n P_{ii}\right)^{p/2} = \left(\operatorname{tr}(P)\right)^{p/2} = k^{p/2}. \tag{A.6}$$

It follows that

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

Putting this into (A.5), we see that

$$\|Z\|_{L_p} \lesssim k^{1/2}. \tag{A.7}$$

Finally, by definition of Z and using the triangle inequality and bounds (A.7), (A.6), we conclude that

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

Proposition 1.3 is proved.

Acknowledgments. The authors are grateful to the referees whose comments improved the presentation of the paper.

REFERENCES

[1] ADAMCZAK, R., LITVAK, A. E., PAJOR, A. and TOMCZAK-JAEGERMANN, N. (2010). Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles. *J. Amer. Math. Soc.* **23** 535–561. MR2601042

[2] ADAMCZAK, R., LITVAK, A. E., PAJOR, A. and TOMCZAK-JAEGERMANN, N. (2011). Sharp bounds on the rate of convergence of the empirical covariance matrix. *C. R. Math. Acad. Sci. Paris* **349** 195–200. MR2769907

[3] BAI, Z. D., SILVERSTEIN, J. W. and YIN, Y. Q. (1988). A note on the largest eigenvalue of a large-dimensional sample covariance matrix. *J. Multivariate Anal.* **26** 166–168. MR0963829

[4] BAI, Z. D. and YIN, Y. Q. (1993). Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. *Ann. Probab.* **21** 1275–1294. MR1235416

[5] BATSON, J. D., SPIELMAN, D. A. and SRIVASTAVA, N. (2012). Twice-Ramanujan sparsifiers. *SIAM J. Comput.* **41** 1704–1721.

[6] BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.* **227** 494–521. MR2782201

[7] DE LA PEÑA, V. H. and GINÉ, E. (1999). *Decoupling*: *From Dependence to Independence*: *Randomly Stopped Processes U-Statistics and Processes Martingales and Beyond*. Springer, New York. MR1666908

[8] FIGIEL, T., HITCZENKO, P., JOHNSON, W. B., SCHECHTMAN, G. and ZINN, J. (1997). Extremal properties of Rademacher functions with applications to the Khintchine and Rosenthal inequalities. *Trans. Amer. Math. Soc.* **349** 997–1027. MR1390980

[9] KANNAN, R., LOVÁSZ, L. and SIMONOVITS, M. (1997). Random walks and an $O^*(n^5)$ volume algorithm for convex bodies. *Random Structures Algorithms* **11** 1–50. MR1608200

[10] LATALA, R. (2005). Some estimates of norms of random matrices. *Proc. Amer. Math. Soc.* **133** 1273–1282 (electronic). MR2111932

[11] PAOURIS, G. (2006). Concentration of mass on convex bodies. *Geom. Funct. Anal.* **16** 1021–1049. MR2276533

[12] ROSENTHAL, H. P. (1970). On the subspaces of L^p $(p > 2)$ spanned by sequences of independent random variables. *Israel J. Math.* **8** 273–303. MR0271721

[13] RUDELSON, M. (1999). Random vectors in the isotropic position. *J. Funct. Anal.* **164** 60–72. MR1694526

[14] SRIVASTAVA, N. (2010). Spectral sparsification and restricted invertibility. Ph.D. thesis, Yale Univ. MR2941475

[15] VERSHYNIN, R. (2011). A simple decoupling inequality in probability theory. Available at http://www-personal.umich.edu/~romanv/papers/decoupling-simple.pdf.

[16] VERSHYNIN, R. (2012). How close is the sample covariance matrix to the actual covariance matrix? *J. Theoret. Probab.* **25** 655–686.

[17] VERSHYNIN, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing*: *Theory and Applications* (Y. Eldar and G. Kutyniok, eds.) 210–268. Cambridge Univ. Press, Cambridge.

SCHOOL OF MATHEMATICS
INSTITUTE FOR ADVANCED STUDY
1 EINSTEIN DRIVE
PRINCETON, NEW JERSEY 08540
USA
E-MAIL: nikhils@math.ias.edu

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MICHIGAN
530 CHURCH ST.
ANN ARBOR, MICHIGAN 48109
USA
E-MAIL: romanv@umich.edu