On the concentration of eigenvalues of random symmetric matrices

Noga Alon* Michael Krivelevich† Van H. Vu ‡

September 6, 2001

Abstract

It is shown that for every $1 \leq s \leq n$, the probability that the s-th largest eigenvalue of a random symmetric n-by-n matrix with independent random entries of absolute value at most 1 deviates from its median by more than t is at most $4e^{-t^2/32s^2}$. The main ingredient in the proof is Talagrand's Inequality for concentration of measure in product spaces.

1 Introduction

In this short paper we consider the eigenvalues of random symmetric matrices whose diagonal and upper diagonal entries are independent real random variables. Our goal is to study the concentration of the largest eigenvalues. For a symmetric real n-by-n matrix A, let $\lambda_1(A) \geq \lambda_2(A) \geq \ldots \geq \lambda_n(A)$ be its eigenvalues.

There are numerous papers dealing with eigenvalues of random symmetric matrices. The most celebrated result in this field is probably the so called *Semicircle Law* due to Wigner ([10], [11]) describing the limiting behavior of the bulk of the spectrum of random symmetric matrices under certain regularity assumptions.

The semi-circle law. *For $1 \leq i \leq j \leq n$ let a_{ij} be real valued independent random variables satisfying:*

1. *The laws of distributions of $\{a_{ij}\}$ are symmetric;*
2. $E[a_{ij}^2] = \frac{1}{4},\ 1 \leq i < j \leq n, \quad E[a_{ii}^2] \leq C,\ 1 \leq i \leq n;$
3. $E[(a_{ij})^{2m}] \leq (Cm)^m$, *for all $m \geq 1$,*

*Department of Mathematics, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel. E-mail: noga@math.tau.ac.il. Research supported in part by a USA Israeli BSF grant, by a grant from the Israel Science Foundation and by the Hermann Minkowski Minerva Center for Geometry at Tel Aviv University.

†Department of Mathematics, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel. E-mail: krivelev@math.tau.ac.il, Research supported in part by a USA-Israel BSF Grant and by a Bergmann Memorial Grant.

‡Microsoft Research, 1 Microsoft Way, Redmond, WA 98052, USA. E-mail: vanhavu@microsoft.com.

where $C > 0$ is an absolute constant. For $i < j$ set $a_{ji} = a_{ij}$. Let A_n denote the random matrix $(a_{ij})_1^n$. Finally, denote by $W_n(x)$ the number of eigenvalues of A_n not larger than x, divided by n. Then

$$\lim_{n\to\infty} W_n(x\sqrt{n}) = W(x) \ ,$$

in distribution, where $W(x) = 0$ if $x \leq -1, W(x) = 1$ if $x \geq 1$ and $W(x) = \frac{2}{\pi}\int_{-1}^{x}(1-x^2)^{1/2}dx$ if $-1 \leq x \leq 1$.

Many extensions and ramifications of the Semicircle Law have been proven since then. It is important to observe that the Semicircle Law provides a very limited information about the asymptotic behavior of any particular (say, the first) eigenvalue. There are, however, quite a few results describing the asymptotic distribution of the first few eigenvalues of random symmetric matrices. For example, Tracy and Widom [8], [9] found, for any fixed $k \geq 1$, the limiting distribution of the first k eigenvalues of the so called Gaussian Orthogonal Ensemble (GOE), corresponding to the case when the off-diagonal entries of the random symmetric matrix A are independent normally distributed random variables with parameters 0 and 1/2. Very recently, Soshnikov [6] generalized their result for a general Wigner Ensemble, i.e., for a random symmetric matrix meeting the conditions of the Semicircle Law. Füredi and Komlós [3] proved that if all off-diagonal entries $a_{ij}, \ , i < j$ of A have the same first moment $\mu > 0$ and the same second moment σ^2, while the expectation of all diagonal entries a_{ii} is $E[a_{ii}] = \nu$, then, assuming that all entries of A are uniformly bounded by an absolute constant $K > 0$, the first eigenvalue of A has asymptotically a normal distribution with expectation $(n-1)\mu + \nu + \delta^2/\mu$ and variance $2\delta^2$.

As we have mentioned already, our main goal here is to obtain concentration results for the eigenvalues of random symmetric matrices. Thus, instead of trying to calculate the limiting distribution of a particular eigenvalue we will rather be interested in bounding its tails. Of course, knowledge of the limiting distribution of a random variable (an eigenvalue, in our context) provides certain information about the decay of its tails. Sometimes, however, concentration results can be derived by applying powerful general tools dealing with concentration of measure to the particular setting of eigenvalues of random symmetric matrices. A detailed discussion of the later approach can be found in a recent survey of Davidson and Szarek [2].

Here we consider the following quite general model of random symmetric matrices. For $1 \leq i \leq j \leq n$, let a_{ij} be independent, real random variables with absolute value at most 1. Define $a_{ji} = a_{ij}$ for all admissible i, j, and let A be the n by n matrix $(a_{ij})_{n\times n}$. Our main result is as follows.

Theorem 1 *For every positive integer $1 \leq s \leq n$, the probability that $\lambda_s(A)$ deviates from its median by more than t is at most $4e^{-t^2/32s^2}$. The same estimate holds for the probability that $\lambda_{n-s+1}(A)$ deviates from its median by more than t.*

We wish to stress that our setting, though being incomparable with some other previously studied ensembles, like the Gaussian Orthogonal Ensemble, is very general and can potentially be applied to many particular cases. The proof is based on the so called Talagrand Inequality ([7], c.f. also [1], Chapter 7) and thus certainly fits the above mentioned framework of deriving concentration results for eigenvalues from the general measure concentration considerations.

The rest of the paper is organized as follows. In the next section we prove our main result, Theorem 1. Section 3 is devoted to a discussion of related results and open problems.

The main result of the paper for the first eigenvalue (i.e., the assertion of Theorem 1 for the special case $s = 1$) was first presented in [5], where it was used to design approximation algorithms for coloring and independent set problems, running in expected polynomial time over the space of random graphs $G(n, p)$.

2 The proof

Talagrand's Inequality is the following powerful large deviation result for product spaces.

Theorem 2 ([7]) *Let $\Omega_1, \Omega_2, \ldots, \Omega_m$ be probability spaces, and let Ω denote their product space. Let $\mathcal{A}$ and $\mathcal{B}$ be two subsets of Ω and suppose that for each $B = (B_1, \ldots, B_m) \in \mathcal{B}$ there is a real vector $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_m)$ such that for every $A = (A_1, \ldots, A_m) \in \mathcal{A}$ the inequality*

$$\sum_{i:A_i \neq B_i} \alpha_i \geq t(\sum_{i=1}^{m} \alpha_i^2)^{1/2}$$

holds. Then

$$Pr[\mathcal{A}]Pr[\mathcal{B}] \leq e^{-t^2/4}.$$

Talagrand's Inequality has already found a large number of applications in diverse areas. In particular, it has been used by Guionnet and Zeitouni [4] to derive concentration inequalities for the spectral measure of random matrices. In their paper, Guionnet and Zeitouni mention (without the detailed proof) the possibility of using similar tools to obtain concentration results for the spectral radius of random matrices.

We now apply Talagrand's Inequality to prove Theorem 1. Put $m = \binom{n+1}{2}$ and consider the product space Ω of the entries a_{ij}, $1 \leq i \leq j \leq n$. Fix a positive integer s, and let M, t be real numbers, where $t > 0$. Let $\mathcal{A}$ be the set of all matrices A in our space for which $\lambda_s(A) \leq M$ and let $\mathcal{B}$ denote the set of all matrices B for which $\lambda_s(B) \geq M + t$. By slightly abusing the notation we identify each member of $A = (a_{ij}) \in \mathcal{A} \cup \mathcal{B}$ with the vector of Ω consisting of its entries (a_{ij}) for $1 \leq i \leq j \leq n$.

Fix a vector $B = (b_{ij}) \in \mathcal{B}$. Let $v^{(1)}, v^{(2)}, \ldots, v^{(s)}$ be the eigenvectors of the s largest eigenvalues of (the matrix) B, where the l_2-norm of each $v^{(p)}$ is 1 and the vectors are orthogonal. Suppose $v^{(p)} = (v_1^{(p)}, v_2^{(p)}, \ldots, v_n^{(p)})$ and define a vector $\alpha = (\alpha_{ij})_{1 \leq i \leq j \leq n}$ of length m as follows.

$$\alpha_{ii} = \sum_{p=1}^{s} \left(v_i^{(p)}\right)^2 \quad \text{for} \quad 1 \leq i \leq n$$

and

$$\alpha_{ij} = 2\sqrt{\sum_{p=1}^{s} \left(v_i^{(p)}\right)^2} \sqrt{\sum_{p=1}^{s} \left(v_j^{(p)}\right)^2} \quad \text{for} \quad 1 \leq i < j \leq n.$$

Claim 1:

$$\sum_{1 \leq i \leq j \leq n} \alpha_{ij}^2 \leq 2s^2.$$

Proof: By definition,

$$\sum_{1\le i\le j\le n} \alpha_{ij}^2 = \sum_{i=1}^{n}\left[\sum_{p=1}^{s}(v_i^{(p)})^2\right]^2 + 4\sum_{1\le i<j\le n}\left[\sum_{p=1}^{s}(v_i^{(p)})^2\right]\left[\sum_{p=1}^{s}(v_j^{(p)})^2\right]$$

$$\le 2\left(\sum_{i=1}^{n}\sum_{p=1}^{s}(v_i^{(p)})^2\right)^2 = 2\left(\sum_{p=1}^{s}\sum_{i=1}^{n}(v_i^{(p)})^2\right)^2 = 2s^2,$$

where here we used the fact that each $v^{(p)}$ is a unit vector. $\square$

Claim 2: For every $A \in \mathcal{A}$,

$$\sum_{1\le i\le j\le n; a_{ij}\neq b_{ij}} \alpha_{ij} \ge t/2.$$

Proof: Fix $A \in \mathcal{A}$. Let $u = \sum_{p=1}^{s} c_p v^{(p)}$ be a unit vector in the span of the vectors $v^{(p)}$ which is orthogonal to the eigenvectors of the largest $s-1$ eigenvalues of A. Then $\sum_{p=1}^{s} c_p^2 = 1$ and $u^t A u \le \lambda_s(A) \le M$, whereas $u^t B u \ge \lambda_s(B) \ge M+t$. Recall that all entries of both A and B are bounded in their absolute values by 1, implying $|b_{ij} - a_{ij}| \le 2$ for all $1 \le i,j \le n$. It follows that if X is the set of all (ordered) pairs ij with $1 \le i,j \le n$ for which $a_{ij} \neq b_{ij}$ then

$$t \le u^t(B-A)u = \sum_{ij\in X}(b_{ij}-a_{ij})\sum_{p=1}^{s} c_p v_i^{(p)} \sum_{p=1}^{s} c_p v_j^{(p)}$$

$$\le 2\sum_{ij\in X}|\sum_{p=1}^{s} c_p v_i^{(p)}||\sum_{p=1}^{s} c_p v_j^{(p)}|$$

$$\le 2\sum_{ij\in X}\left(\sqrt{\sum_{p=1}^{s} c_p^2}\sqrt{\sum_{p=1}^{s}(v_i^{(p)})^2}\right)\left(\sqrt{\sum_{p=1}^{s} c_p^2}\sqrt{\sum_{p=1}^{s}(v_j^{(p)})^2}\right) \qquad \text{(by Cauchy-Schwartz)}$$

$$= 2\sum_{1\le i\le j\le n,\ a_{ij}\neq b_{ij}} \alpha_{ij},$$

as needed. $\square$

By the above two claims, and by Theorem 2, for every M and every $t > 0$

$$Pr[\lambda_s(A) \le M]Pr[\lambda_s(B) \ge M+t] \le e^{-\frac{t^2}{32s^2}}. \tag{1}$$

If M is the median of $\lambda_s(A)$ then, by definition, $Pr[\lambda_s(A) \le M] \ge 1/2$, implying that

$$Pr[\lambda_s(A) \ge M+t)] \le 2e^{-\frac{t^2}{32s^2}}.$$

Similarly, by applying (1) with $M+t$ being the median of $\lambda_s(A)$ we conclude that the probability that $\lambda_s(A)$ is smaller than its median minus t is bounded by the same quantity. This completes the proof of Theorem 1 for $\lambda_s(A)$. The proof for $\lambda_{n-s+1}(A)$ is analogous. $\square$

3 Concluding remarks

- In many cases concentration results are presented by giving bounds for the deviation of a random variable from its expectation, rather than its median as in our Theorem 1. Our result however easily enables to show that the expectation and the median of eigenvalues are very close. Indeed, recall that for any non-negative valued random variable X,

$$E[X] = \int_0^\infty Pr[X \geq t]dt \ .$$

  Denote by m_s the median of the s-th eigenvalue of A. By Theorem 1:

$$|E[\lambda_s(A)] - m_s| \leq E[|\lambda_s(A) - m_s|] = \int_0^\infty Pr[|\lambda_s - m_s| \geq t]dt \leq \int_0^\infty 4e^{-t^2/32s^2}dt = 8\sqrt{2\pi s} \ .$$

  Thus, the expectation of $\lambda_s(A)$ and its median are only $O(\sqrt{s})$ apart. Therefore, for all $t >> \sqrt{s}$ we get

$$Pr[|\lambda_s(A) - E[\lambda_s(A)]| \geq t] \leq e^{-(1-o(1))t^2/32s^2} \ .$$

- Our estimate from Theorem 1 is sharp, up to an absolute factor in the exponent, for the deviation of λ_1. Consider the following random symmetric matrix $A = (a_{ij})_1^n$. For each $1 \leq i < j$, a_{ij} takes value 1 with probability 1/2 and value 0 with probability 1/2; all diagonal entries a_{ii} are 0; set also $a_{ji} = a_{ij}$ for $1 \leq i < j \leq n$. (In fact, the obtained random matrix is the adjacency matrix of the binomial random graph $G(n, 1/2)$.) By a result of Füredi and Komlós [3] the expected value of $\lambda_1 = \lambda_1(A)$ is $n/2 + o(1)$. By the previous remark, the median of λ_1 and its expectation differ by at most a constant. On the other hand, λ_1 is at least the average number of ones in a row (the average degree of the graph $G(n, 1/2)$), and as this average degree is $2/n$ times a binomial random variable with parameters $\binom{n}{2}$ and 1/2, it follows that the probability that λ_1 exceeds its median by t is at least $\Omega(e^{-O(t^2)})$ for all admissible values of $t > 1$, say.

- Note that for the adjacency matrix of a random graph the entries of our matrix are in the range $[0, 1]$. In this case the estimate in Theorem 1 can be improved to $4e^{-t^2/8s^2}$, as each of the quantities $|b_{ij} - a_{ij}|$ in the proof of Claim 2 can be bounded by 1 (instead of bounding it by 2, as done in the present proof.)

- In certain cases, our concentration result can be combined with additional considerations to provide bounds for the expectations of eigenvalues of random symmetric matrices. Here is one example.

Proposition 3 *Let a_{ij}, $1 \leq i \leq j \leq n$ be independent random variable bounded by 1 in absolute values. Assume that for all $i < i$, the a_{ij} have a common expectation 0 and a common variance σ^2. Then*

$$E[\lambda_1(A)] \geq 2\sigma n^{1/2} - O(\sigma \log^{1/2} n) \ .$$

Consequently, with probability tending to 1,

$$\lambda_1(A) \geq 2\sigma n^{1/2} - O(\sigma \log^{1/2} n) \ .$$

Proof: Since $\lambda_1(cA) = c\lambda_1(A)$ for every scalar c, we may and will assume that $\sigma = 1/2$. Furthermore, set $\mu = n^{1/2}$, $k = \lceil \mu \log^{1/2} n \rceil$ and $x = a \log^{1/2} n$, where a is a positive constant chosen so that the following two inequalities hold:

$$\mu^k / k^{5/2} \geq 2(\mu - x/2)^k \tag{2}$$

$$\sum_{t=\frac{a}{2}\log^{1/2} n}^{\infty} e^{2t \log^{1/2} n - t^2/40} = o(1), \tag{3}$$

Without loss of generality, we assume that k is an even integer and let X be the trace of A^k. It is trivial that $E[X] \leq nE[\lambda_1^k]$. On the other hand, a simple counting argument (see [3]) shows that

$$E[X] \geq \frac{1}{(k/2)+1} \binom{k}{k/2} \sigma^k n(n-1)\dots(n-(k/2))$$
$$\geq \frac{2^k}{k^{3/2}} (\frac{1}{2})^k n(\mu^2 - \mu \log^{1/2} n)^{k/2} \geq n\mu^k / k^{5/2}.$$

It follows that

$$E[\lambda_1^k] \geq \mu^k / k^{5/2}\,. \tag{4}$$

Assume, for contradiction, that $E(\lambda_1) \leq \mu - x$. It follows from this assumption that

$$E[\lambda_1^k] \leq (\mu - x/2)^k + \sum_{t=x/2}^{\infty} (\mu - x + (t+1))^k Pr[\lambda_1 \geq \mu - x + t]\,. \tag{5}$$

By Theorem 1, $Pr(\lambda_1 \geq \mu - x + t) \leq e^{-t^2/40}$ for all $t \geq x/2$. Thus (2),(4) and (5) imply

$$\sum_{t=x/2}^{\infty} (\mu - x + (t+1))^k e^{-t^2/40} \geq \mu^k / k^{5/2} - (\mu - x/2)^k \geq (\mu - x/2)^k. \tag{6}$$

Since $(\mu - x + (t+1))^k / (\mu - x/2)^k \leq e^{(1+o(1))tk/\mu} = e^{(1+o(1))t \log^{1/2} n}$, (3) and (6) imply a contradiction, and this completes the proof. $\square$

This improves the error term in the bound $\lambda_1(A) \geq 2\sigma n^{1/2} - O(n^{1/3} \log n)$, stated by Füredi and Komlós in [3].

- The concentration provided by Theorem 1 for $\lambda_s(A)$ for larger values of s is weaker than that provided for $s = 1$. It seems this is only a feature of the proof, as it seems plausible to suspect that in fact each λ_s is as concentrated around its median as is λ_1, and in certain situations (like symmetric matrices with independenet, identically distributed entries) $\lambda_2(A)$ might be even more concentrated around its median than λ_1.

- Theorem 1 is obtained under very general assumptions on the distribution of the entries of a symmetric matrix A. Still, it will be very desirable to generalize its assertion even further, in particular, dropping or weakening the restrictive assumption about the uniform boundness of the entries of A. This task, however, may require the application of other tools, as the Talagrand inequality appears to be suited for the case of bounded random variables.
- Finally, it would be interesting to find further applications of our concentration results in algorithmic problems on graphs. The ability to compute the eigenvalues of a graph in polynomial time combined with an understanding of the potentially rich structural information encoded by the eigenvalues is likely to provide a basis for new algorithmic results exploiting the eigenvalues of graphs and their concentration.

Acknowledgment. The authors are grateful to Zeev Rudnick for his helpful comments.

References

[1] N. Alon and J. H. Spencer, *The Probabilistic Method*, Second Edition, Wiley, New York, 2000.

[2] K. R. Davidson and S. J. Szarek, *Local operator theory, random matrices and Banach spaces*, in: Handbook on the Geometry of Banach Spaces, W. B. Johnson, J. Lindenstrauss, eds., Elsevier Science, to appear. Available at: http://www.cwru.edu/artsci/math/szarek/TeX/DavSzaHB.pdf.

[3] Z. Füredi and J. Komlós, *The eigenvalues of random symmetric matrices*, Combinatorica 1 (3), 233–241 (1981).

[4] A. Guionnet and O. Zeitouni, *Concentration of the spectral measure for large matrices*, Electr. Commun. in Probab. Vol 5 (2000), 119–136.

[5] M. Krivelevich and V. H. Vu, *Approximating the independence number and the chromatics number in expected polynomial time*, Proceedings of the 7^{th} Int. Colloq. on Automata, Languages and Programming (ICALP'2000), 13-25.

[6] A. Soshnikov, *Universality of edge of the spectrum in Wigner random matrices*, Commun. Math. Phys. 207 (1999), 697–733.

[7] M. Talagrand, *Concentration of Measures and Isoperimetric Inequalities in product spaces*, Publications Mathematiques de l'I.H.E.S., 81, 73-205 (1996).

[8] C. Tracy and H. Widom, *Level-spacing distribution and Airy kernel*, Commun. Math. Phys. 159 (1994), 151–174.

[9] C. Tracy and H. Widom, *On orthogonal and symplectic matrix ensembles*, Commun. Math. Phys. 177 (1996), 727–754.

[10] E. P. Wigner, *Characteristic vectors of bordered matrices with infinite dimensions*, Ann. Math. 62 (1955), 548–564.

[11] E. P. Wigner, *On the distribution of the roots of certain symmetric matrices,* Ann. Math. 67, 325–328 (1958).