A Short Introduction to Operator Limits of Random Matrices

Diane Holcomb and Bálint Virág

Abstract. These are notes to a four-lecture minicourse given at the 2017 PCMI Summer Session on Random Matrices. We give a quick introduction to the theory of large random matrices by taking limits that preserve their operator structure, rather than just their eigenvalues. The operator structure takes the role of exact formulas, and allows for results in the context of general β-ensembles. Along the way, we cover a non-computational proof of the Wiegner semicircle law, a quick proofs for the Füredi-Komlós result on the top eigenvalue, as well as the BBP phase transition.

1. The Gaussian Ensembles

1.1. The Gaussian Orthogonal and Unitary Ensembles. One of the earliest appearances of random matrices in mathematics was due to Eugene Wigner in the 1950's. Let G be an $n \times n$ matrix with independent standard normal entries. Then

$$M_n = \frac{G + G^t}{\sqrt{2}}.$$

This distribution on symmetric matrices is called the Gaussian Orthogonal Ensemble, because it is invariant under orthogonal conjugation. For any orthogonal matrix OM_nO^{-1} has the same distribution as M_n. To check this, note that OG has the same distribution as G be the rotation invariance of the Gaussian column vectors, and the same is true for OGO^{-1} by the rotation invariance of the row vectors. To finish note that orthogonal conjugation commutes with symmetrization.

If we instead start with a matrix with independent standard complex Gaussian entries, we get the Gaussian Unitary ensemble. To see how the eigenvalues behave, we recall the following classical theorem.

Theorem 1.1.1. *Suppose M_n has GOE or GUE distribution then M_n has eigenvalue density*

$$f(\lambda_1, \dots, \lambda_n) = \frac{1}{Z_n} \prod_{k=1}^{n} e^{-\frac{\beta}{4}\lambda_k^2} \prod_{i<j} |\lambda_i - \lambda_j|^\beta \tag{1.1.2}$$

with $\beta = 1$ for the GOE and $\beta = 2$ for the GUE.

2010 *Mathematics Subject Classification.* Primary 60B20; Secondary 15B52.
Key words and phrases. Park City Mathematics Institute.

©0000 (copyright holder)

For convenience we will take $\Lambda = \Lambda_n = \{\lambda_i\}_{i=1}^n$ to denote the set of eigenvalues of the GOE or GUE. This notation will be used later to denote the eigenvalues or points in whatever random matrix model is being discussed at the time.

From this we can see that this is a model for n particles that would like to be Gaussian, but the Vandermonde term pushes them apart. Note that Tr $M_n^2/n^2 \to 1$ in probability (the sum of squares of Gaussians), so the empirical quadratic mean of the eigenvalues is asymptotically $\sqrt{n}$, rather than order 1. The interaction term has a very strong effect.

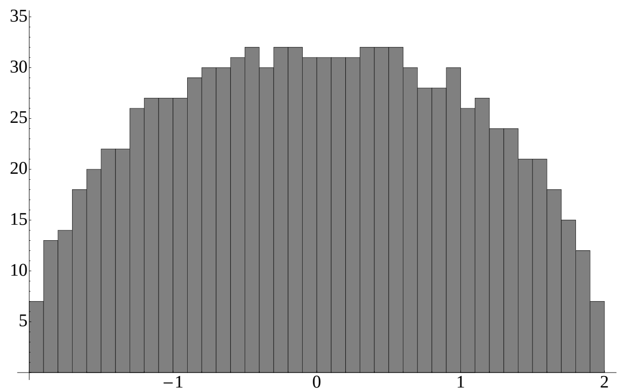

FIGURE 1.1.3. Rescaled eigenvalues of a 1000×1000 GOE matrix

1.2. Tridiagonalization and spectral measure. The spectral measure of a matrix at a coordinate (which we will take to be the first one) is a measure supported on the eigenvalues that reflects the local structure of the matrix there.

Definition 1.2.1. For a symmetric matrix A, its spectral measure (at the first coordinate) σ_A is the measure for which

$$\int x^k d\sigma_A = A_{11}^k.$$

From this definition, its unclear whether the spectral measure exists or is unique. Nevertheless, that is the case for finite matrices.

Exercise 1.2.2. Check that if $\lambda_1, ..., \lambda_n$ are the eigenvalues of A then

$$\sigma_A = \sum_i \delta_{\lambda_i} \varphi_i(1)^2$$

where φ_i is the ith normalized eigenvector of A.

Now measures with finite support are determined by their moments, so in fact the definition above works.

The spectral measure is a complete invariant for a certain set of symmetries. For this, first recall something more familiar.

We say two symmetric matrices are equivalent if they have the same eigenvalues with multiplicity. This equivalence is well understood: two matrices are equivalent if and only if they are conjugates by an orthogonal matrix. In group theory language, the equivalence classes are the orbits of the conjugation action of the orthogonal group. There is a canonical representative in each class, a diagonal matrix with non-increasing diagonals, and the set of eigenvalues is a complete invariant.

Can we have a similar characterization for matrices with the same spectral measure? The answer is yes, for a generic class of matrices.

Definition 1.2.3. A vector v is cyclic for an $n \times n$ matrix A if $v, Av, ..., A^{n-1}v$ is a basis for the vector space $\mathbb{R}^n$.

Theorem 1.2.4. *Let A and B be two matrices for which the first coordinate vector is cyclic. Then $\sigma_A = \sigma_B$ if an only if $O^{-1}AO = B$ where O is orthogonal matrix fixing the first coordinate vector.*

Let's find a nice set of class representatives.

Definition 1.2.5. A Jacobi matrix is a real symmetric tridiagonal matrix with positive off-diagonals.

Theorem 1.2.6. *For all A there exists a unique Jacobi matrix J such that $\sigma_J = \sigma_A$*

Proof of Existence. We can conjugate a symmetric matrix to a Jacobi matrix by hand. Write our matrix in a block form,

$$A = \left[\begin{array}{c|c} a & b^t \\ \hline b & C \end{array}\right]$$

Now let O be an $(n-1) \times (n-1)$ orthogonal matrix, and let

$$Q = \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & O \end{array}\right]$$

Then Q is orthogonal and

$$QAQ^t = \left[\begin{array}{c|c} a & (Ob)^t \\ \hline Ob & OCO^t \end{array}\right]$$

Now we can choose the orthogonal matrix O so that Ou is in the direction of the first coordinate vector, namely $Ou = |u|\mathbf{e}_1$.

An explicit option for O is the following Householder reflection:

$$Ov = v - 2\frac{\langle v, w\rangle}{\langle w, w\rangle}w \quad \text{ where } \quad w = b - |b|\mathbf{e}_1$$

Check that $OO^t = I$, $Ob = |b|\mathbf{e}_1$.

Therefore

$$QAQ^t = \left[\begin{array}{c|cccc} a & |b| & 0 & \dots & 0 \\ \hline |b| & & & & \\ 0 & & & & \\ \vdots & & & OCO^t & \\ 0 & & & & \end{array}\right].$$

We now repeat the previous step, but this time choosing the first two rows and columns to be 0 except having 1's in the diagonal entries, and than again until the matrix becomes tridiagonal. □

There is a lot of choices that can be made for the orthogonal matrices during the tridiagonalization. However, these choices do not affect the final result. J is unique, as shown in the following exercise.

Exercise 1.2.7. Show that two Jacobi matrices with the same spectral measure are equal. (Hint: express the moments $J_{1,1}^k$ of the spectral measure as sums over products of matrix entries.)

The procedure presented above may have a familiar feeling. It turns out that Gram-Schmidt is lurking in the background

Exercise 1.2.8. Suppose that the first coordinate vector e_1 is cyclic. Apply Gram-Schmidt to the vectors $(e_1, Ae_1, ..., A^{n-1})$ to get a new orthonormal basis. Show that A written in this basis will be a Jacobi matrix.

1.2.1. Tridiagonalization and the GOE. When we apply the tridiagonalization procedure to the GOE we get a very clean result because of the invariance of the distribution under independent orthogonal transformation.

Proposition 1.2.9 (Trotter [12])**.** *Let A be GOE_n. There exists a random orthogonal matrix fixing the first coordinate vector e so that*

$$OAO^t = \begin{bmatrix} a_1 & b_1 & 0 & \dots & 0 \\ b_1 & a_2 & \ddots & & \\ 0 & \ddots & \ddots & \ddots & \\ \vdots & & \ddots & a_{n-1} & b_{n-1} \\ 0 & & & b_{n-1} & a_n \end{bmatrix}$$

with $a_i \sim \mathcal{N}(0,2)$ and $b_i \sim \chi_{n-i}$ and independent. In particular, OAO^t has the same spectral measure as A.

Recall that for a vector v of independent $\mathcal{N}(0,1)$ random variables of length k, then $\chi_k \overset{d}{=} \|v\|$. The density of a χ random for $k > 0$ is given by

$$f_{\chi_k}(x) = \frac{1}{2^{\frac{k}{2}-1}\Gamma(k/2)} x^{k-1} e^{-x^2/2},$$

where $\Gamma(x)$ is the Gamma function.

Proof. The argument above can be applied to the random matrix. After the first step, OCO^t will be independent of a, b and have a GOE distribution. This is because GOE is invariant by conjugation with a fixed O, and O is only a function of b. So conditionally on a,b, both C and OCO^t have GOE distribution. □

Exercise 1.2.10. Let X be an $n \times m$ matrix with $X_{i,j} \sim \mathcal{N}(0, 1)$ (not symmetric nor Hermitian). The distribution of this matrix is invariant under left and right multiplication by independent unitary matrices. Show that such a matrix X may be lower bidiagonalized such that the distribution of the singular values is the same for both matrices. Note that the singular values of a matrix are unchanged by multiplication by a unitary matrix.

(1) Start by coming up with a matrix that right multiplied with A gives you a matrix where the first row is 0 except the 11 entry.
(2) What can you say about the distribution of the rest of the matrix after this transformation to the first row?
(3) Next apply a left multiplication. Continue using right and left multiplication to finish the bidiagonalization.

1.3. β-ensembles. Let's consider the spectral measure as a map $J \mapsto \sigma_J$ from Jacobi matrices of dimension n to probability measures on at most n points. We have seen that this map is one-to-one. First we see that in fact spectral measures in the image are supported on exactly n points.

Exercise 1.3.1. Show that a Jacobi matrix cannot have an eigenvector whose first coordinate is zero. Conclude that all eigenspaces are 1-dimensional.

Second, we check that for the set of such probability measure measures, the map $J \mapsto \sigma_J$ is onto.

Exercise 1.3.2. For every probability measure on n points there exists a symmetric matrix with that spectral measure. This implies that there exists a Jacobi matrix with this spectral measure.

Since $J \mapsto \sigma_J$ is a bijection, we could compute probability distributions by the change-of-variable formula, as long as we know the Jacobian determinant.

Let

$$A_n = \frac{1}{\sqrt{\beta}} \begin{bmatrix} a_1 & b_1 & 0 & \dots & 0 \\ b_1 & a_2 & \ddots & & \\ 0 & \ddots & \ddots & \ddots & \\ \vdots & & \ddots & a_{n-1} & b_{n-1} \\ 0 & & & b_{n-1} & a_n \end{bmatrix}$$

That is a tridiagonal matrix with $a_1, a_2, ..., a_n \sim N(0,2)$ on the diagonal and $b_1, ..., b_{n-1}$ with $b_k \sim \chi_{\beta(n-k)}$ and everything independent. Recall that if $z_1, z_2, ...$ are independent standard normal random variables, then $z_1^2 + \cdots + z_k^2 \sim \chi_k^2$.

If $\beta = 1$ then A_n is similar to a GOE matrix (the joint density of the eigenvalues is the same). If $\beta = 2$ then A_n is similar to a GUE matrix.

Theorem 1.3.3 ([5]). *If $\beta > 0$ then the joint density of the eigenvalue of A_n is given by*

$$f(\lambda_1, ..., \lambda_n) = \frac{1}{Z_{n,\beta}} e^{-\frac{\beta}{4}\sum_{i=1}^n \lambda_i^2} \prod_{1 \leqslant i < j \leqslant n} |\lambda_i - \lambda_j|^\beta.$$

In the matrix A_n the off-diagonal entries are $b_i > 0$, and there are $2n-1$ variable in the matrix.

It is hard to compute the Jacobian of the map $(\overline{\lambda}, \overline{q}) \to (\overline{a}, \overline{b})$ directly, because the map is complicated. To work around this, we use moments, which have a simple connection to both representations:

$$m_k = \int x^k d\sigma = \sum \lambda_i^k q_i^2$$

We look at maps from both sets to $(m_1, ..., m_{2n-1})$. These are simple transformations. One can write down the appropriate matrices and then can find their determinants. These computations can be found in [9], and yield the following.

Theorem 1.3.4 (Dumitriu, Edelman, Krishnapur, Rider, Virág, [5, 9]). *Let V be a potential (think convex) and $\overline{a}, \overline{b}$ are chosen from then density proportional to*

$$\exp\left(-\mathit{Tr} V(J)\right) \prod_{k=1}^{n-1} b_{n-k}^{k\beta - 1}$$

then the eigenvalues have distribution

$$f(\lambda_1, .., \lambda_n) = \frac{1}{Z} \exp\left(-\sum_i V(\lambda_i)\right) \prod_{i<j} |\lambda_i - \lambda_j|^\beta$$

and the q_i are independent of the λ with $(q_1, ..., q_n) = (\varphi_1(1)^2, ..., \varphi_n(1)^2)$ have Dirichlet$(\frac{\beta}{2}, ..., \frac{\beta}{2})$ distribution.

Exercise 1.3.5. Show that when $V(x) = x^4$, the sequence $\{(a_i, b_i), i \geqslant 1\}$ with the distribution from the theorem forms a time-inhomogeneous Markov Chain.

A result like this holds for general polynomial V, though one needs to take bigger blocks of (a_i, b_i). This is exploited in [9] to get universality for the top eigenvalue.

2. The Wigner semicircle law

2.1. Graph convergence. The proof of the Wigner semicircle law given here will rest on a graph convergence argument. We begin by introducing the notions of convergence needed for the proof. Examples will make the convergence easier to understand.

We will be considering rooted graphs (G, ρ) where the root ρ is just the a marked vertex. The spectral measure of (G, ρ) is the spectral measure of the adjacency matrix A of G at the coordinate corresponding to ρ (which we will often just take to be the first entry).

Note that the kth moment of the spectral measure is just the number of paths of length k starting and ending at the root. In particular, moments up to $2k$ of the spectral measure are determined by the k-neighborhood of ρ in G.

Our definition of the spectral measure works even for infinite graphs, but it is again not a priori clear that it exists or it is unique.

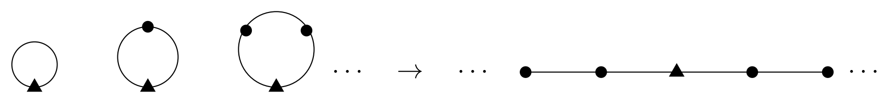

FIGURE 2.1.1. Rooted convergence: n-cycles to $\mathbb{Z}$

Definition 2.1.2. Rooted convergence. A sequence of rooted graphs (G_n, ρ) converges to a limit (G, ρ) if for any radius r, the ball of that radius in G_n about ρ equals that in G for all large enough n.

Examples 2.1.3. We give two examples

(1) The n cycle with any vertex chosen as the root converges to $\mathbb{Z}$.
(2) The k by k grid with vertices at the intersection. If we choose a vertex at the center of the grid as the root, we get convergence to $\mathbb{Z}^2$.

For bounded degree graphs G_n, if (G_n, ρ_n) converges to (G, ρ) is the sense of rooted convergence, then by definitions, the moments of the spectral measures σ_n converge.

This implies two things. First, since the spectral measures of bounded (by b) degree graphs are supported on $[-b, b]$, σ_n have subsequential weak limits on $[-b, b]$. But such measures are determined by their moments, so the limit of σ_n exists, and is the spectral measure of (G, o). Since any bounded degree rooted infinite graph is a rooted limit of balls around the root, we get

Proposition 2.1.4. *Bounded degree rooted infinite graphs have unique spectral measure.*

Exercise 2.1.5. Consider paths of length n rooted at the left end point. This sequence converges to $\mathbb{Z}_+$ in the limit. What is the limit of the spectrum? It is the Wigner semicircle law, since the moments are Dyck paths. But one can prove this directly, since the paths on length n are easy to diagonalize. This is an example where the spectral measure has a different limit than the eigenvalue distribution.

Exercise 2.1.6. Suppose you have the random d-regular graph on n vertices in the configuration model (for a given degree sequence we choose a uniform at random matching on the half edges attached to each vertex). In the limit this converges to the d-regular infinite tree in the limit.

Definition 2.1.7. Benjamini-Schramm convergence: We define this convergence for unrooted graphs. Choose a vertex uniformly at random to be the root. If this converges in distribution with respect to rooted convergence to a random rooted graph, we say the graphs converge in the Benjamini-Schramm sense.

Benjamini-Schramm convergence is equivalent to convergence of local statistics. This is the following statement. For every finite rooted graph (K, ρ) and every r, the proportion of vertices in G_n whose ball radius r is rooted-isomorphic to (K, ρ) converges to the probability that the ball of radius r in the limit is rooted-isomorphic to (K, ρ).

Benjamini-Schramm limits of finite graphs are unimodular. For the special case of regular graphs, this means that if we pick a uniform random neighbor ν of ρ in G, then (G, ρ, ν) has the same distribution as (G, ν, ρ).

Exercise 2.1.8. Show that if G is a fixed connected finite regular graph with a random vertex ρ, then (G, ρ) is unimodular if and only if ρ has uniform distribution.

In the general case the distributions have to first be biased by the degree of the root.

The most intriguing open problem in this area is whether all infinite unimodular random graphs are Benjamini-Schramm limits. Those that are are called sofic.

Proposition 2.1.9. *Suppose we have a finite graph G and we choose a vertex uniformly at random. This defines a random graph and its associated random spectral measure σ. Then $E\sigma = \mu$ is the eigenvalue distribution.*

Proof. Recall that for the spectral measure of a matrix (and so a graph) we have

$$\sigma_{(G,\rho)} = \sum_{i=1}^{n} \delta_{\lambda_i} \varphi_i^2(\rho)$$

Since φ_i is of length one, we have

$$\sum_{\rho \in V(G)} \varphi_i(\rho)^2 = 1$$

hence

$$E\sigma_{G,\rho} = \frac{1}{n} \sum_{\rho \in V(G)} \sigma_{G,\rho} = \mu_G.$$

□

Example 2.1.10. The following are examples of Banjamini-Schramm convergence:

(1) A cycle graph converges to the graph of $\mathbb{Z}$.
(2) A path of length n converges to the graph of $\mathbb{Z}$.
(3) Large box of $\mathbb{Z}^d$ converges to the full $\mathbb{Z}^d$ lattice.

Notice that for the last two examples the probability of being in a neighborhood of the edge goes to 0 and so the limiting graph doesn't see the edge effects.

Exercise 2.1.11. We say a sequence of d-regular graphs G_n with n vertices is of essentially large girth if for every k the number of k-cycles in G_n is $o(n)$. Show that G_n is essentially large girth if and only if it Benjamini-Schramm converges to the d-regular tree.

Exercise 2.1.12. Show that for $d \geqslant 3$ the d-regular tree is not the Benjamini-Schramm limit of finite trees. (Hint: consider the expected degree).

How does this help? We have that $\sigma_n \to \sigma_\infty$ in distribution. By bounded convergence, the eigenvalue distributions will converge as well:

$$\mu_n = E\sigma_n \to E\sigma_\infty.$$

One can consider a more general setting of weighted graphs. This corresponds to general symmetric matrices A. In this case we require that the neighborhoods stabilize and the weights also converge. Everything above goes through.

Example 2.1.13 (The spectral measure of $\mathbb{Z}$)**.** We use Benjamini-Schramm convergence of the cycle graph to $\mathbb{Z}$. We begin by computing the spectral measure of the n-cycle G_n. We get that $A = T + T^t$ where

$$T = \begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ 1 & & & & 0 \end{bmatrix}$$

The eigenvalues of T are the nth roots of unity η_i. We can think of this as $A = T + T{-1}$, so the eigenvalues of A are $\eta_i + \eta_i^{-1} = 2\Re\eta_i$. Geometrically, these are projections of the $2\eta_i$, that is points uniformly spaced on the circle of radius 2, to the real line.

In the limit measure converges to the projection of the uniform measure on that circle, also called the arcsine distribution

$$\sigma_{\mathbb{Z}} = \frac{1}{2\pi\sqrt{4-x^2}}\mathbf{1}_{x\in[-2,2]}\,dx.$$

Exercise 2.1.14. Let B_n be the unweighted finite binary tree with n levels. Suppose a vertex is chosen uniformly at random from the set of vertices. Give the distribution of the limiting graph.

The proof of the Wigner semicircle law is now just an exercise. We will give the solution as well.

Exercise 2.1.15. Let A be a rescaled $n \times n$ Dumitriu-Edelman tridiagonal matrix

$$A = \frac{1}{\sqrt{\beta n}} \begin{bmatrix} b_1 & a_1 & & & \\ a_1 & b_2 & a_2 & & \\ & a_2 & \ddots & \ddots & \\ & & \ddots & b_{n-1} & a_{n-1} \\ & & & a_{n-1} & b_n \end{bmatrix}, \quad a_i \sim \chi_{\beta(n-i)}, \quad b_i \sim \mathcal{N}(0,1)$$

all independent, and suppose that A is the adjacency matrix of a weighted graph.

(1) Draw the graph with adjacency matrix A. (There can be loops)
(2) Suppose a root for your graph is chosen uniformly at random, what is the limiting distribution of your graph?
(3) What is the limiting spectral measure of the graph rooted at the vertex corresponding to the first row and column?
(4) What is the limiting spectral measure of the unweighted graph?

2.2. Wigner semicircle law. Note that a Jacobi matrix can be thought of the adjacency matrix of a weighted path with loops.

Exercise 2.2.1. Check that $\chi_n - \sqrt{n} \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0, 1/2)$.

Proof 1. [15]

Take the previous graph, divide all the labels by $\sqrt{n}$ and then take a Benjamini-Schramm limit.

What is the limit? $\mathbb{Z}$, but then we need labels. The labels in a randomly rooted neighborhood will now be the square root of a single uniform random variable in $[0,1]$. Call this U. Then the edge weights are $\sqrt{U}$. Recall that $\mu_n \to E\sigma$. In the case where U was fixed we would just get a scaled arcsine measure.

Choose a point uniformly on the circle of radius $\sqrt{U}$ and project it down to the real line. But this point is in fact a uniform random point in the disk. This gives us the semicircle law.

$$\mu_{sc} = \frac{1}{2\pi}\sqrt{4-x^2}1_{x\in[-2,2]}\,dx.$$

□

Proof 2. We take the rooted limit where we choose the root to be the matrix $J/\sqrt{n}$ corresponding to the first coordinate vector.

In the limit this graph converges to $\mathbb{Z}_+$. Therefore $\sigma_n \to \sigma_{\mathbb{Z}_+} = \rho_{sc}$. This convergence is weak convergence in probability.

So going back to the full matrix model of GOE, we see that the spectral measure at an arbitrary root converges weakly in probability to μ_{sc}. But then this must hold also if we average the spectral measures over the choice of root (but not the randomness in the matrix).

Thus we get $\mu_n \to \mu_{sc}$ in probability. □

Dilemma: The limit of the spectral measure should have nothing to do with the limit of the eigenvalue distribution in the general case. This tells you that the Jacobi matrices that we get in the case of the GOE are very special.

3. The top eigenvalue and the Baik-Ben Arous-Pechet transition

3.1. The top eigenvalue. The eigenvalue distribution of the GOE converges after scaling by $\sqrt{n}$ to the Wigner semicircle law. From this, it follows that the top eigenvalue, $\lambda_1(n)$ satisfies for every $\varepsilon > 0$

$$P(\lambda_1(n)/\sqrt{n} > 2 - \varepsilon) \to 1$$

the 2 here is the top of the support of the semicircle law. However, the upper bound does not follow and needs more work. This is the content of the following theorem.

Theorem 3.1.1 (Füredi-Komlós).

$$\frac{\lambda_1(n)}{\sqrt{n}} \to 2 \textit{ in probability.}$$

This holds for more general Wigner matrices; we have a simple proof for the GOE case.

Lemma 3.1.2. *If* J *is a Jacobi matrix (a's diagonal, b's off-diagonal) then*

$$\lambda_1(J) \leqslant \max_i (a_i + b_i + b_{i-1})$$

Here we take the convention $b_0 = b_n = 0$.

Proof. Observe that J may be written as

$$J = -AA^T + \operatorname{diag}(a_i + b_i + b_{i-1})$$

where

$$A = \begin{bmatrix} 0 & \sqrt{b_1} & & & \\ & -\sqrt{b_1} & \sqrt{b_2} & & \\ & & -\sqrt{b_2} & \sqrt{b_3} & \\ & & & \ddots & \ddots \end{bmatrix}$$

and AA^t is nonnegative definite. So for the top eigenvalues we have

$$\lambda_1(J) \leqslant -\lambda_1(AA^T) + \lambda_1(\operatorname{diag}(a_i + b_i + b_{i-1})) \leqslant \max_i (a_i + b_i + b_{i-1}).$$

We used subadditivity of λ_1, which follows from the Rayleigh quotient representation. □

If we apply this to our setting we get that

$$\lambda_1(\mathrm{GOE}) \leqslant \max_i (N_i, \chi_{n-i} + \chi_{n-i+1}) \leqslant 2\sqrt{n} + c\sqrt{\log n} \tag{3.1.3}$$

the right inequality is an exercise (using the Gaussian tails in χ) and holds with probability tending to 1 if c is large enough. This completes the proof of Theorem 3.1.1.

This shows that the top eigenvalue cannot go further than an extra $\log n$ outside of the spectrum. Indeed we have that

$$\lambda_1(\mathrm{GOE}) = 2\sqrt{n} + TW_1 n^{-1/6} + o(n^{-1/6})$$

so the bound above is not optimal.

3.2. Baik-Ben Arous-Pechet transition. The approach we take here is a version of a section in Bloemendal's PhD thesis [2].

Historically one of the areas that random matrices have been used is to study correlations. To see whether correlations are significant, one compares to a case in which random data comes with no correlations at all.

Wishart in the 20s considered matrices $X_{n\times m}$ with independent normal entries and studied the eigenvalues of XX^t. The rank-1 perturbations below model the case one there is one significant trend in the data, but the rest is just noise. We consider the case $n = m$. A classical result is the following.

Theorem 3.2.1 (BBP transition)**.**

$$\frac{1}{n}\lambda_1\Big(X\ \mathrm{diag}(1+a^2, 1, 1, ..., 1)X^t\Big) \to \varphi(a)^2$$

where

$$\varphi(a) = \begin{cases} 2 & a \leqslant 1 \\ a + \frac{1}{a} & a \geqslant 1 \end{cases}$$

Heuristically, correlation in the populations appears in the asymptotics in the top eigenvalue of the data only if it is sufficiently large, $a > 1$. Otherwise, it gets washed out by the fake correlations coming from noise. . We will prove the GOE analogue of this theorem, and leave the Wishart case as an exercise.

One can also study the fluctuations of the eigenvalues. In the case $a < 1$ we get Tracy-Widom. When $a > 1$ we get Gaussian fluctuations. Very close to the point $a = 1$ we get a deformed Tracy-Widom, see [1], [3].

The GOE analogue answers the following question. Suppose that we add a common nontrivial mean to the entries of a GOE matrix. When does this influence the top eigenvalue on the semicircle scaling?

Theorem 3.2.2 (Top eigenvalue of GOE with nontrivial mean)**.**

$$\frac{1}{\sqrt{n}}\lambda_1(\mathrm{GOE}_n + \frac{a}{\sqrt{n}}11^t) \to \varphi(a)$$

where 1 *is the all-1 vector, and* 11^t *is the all-1 matrix.*

It may be surprising how little change in the mean in fact changes the top eigenvalue!

We will not use the following theorem, but will include it only to show where the function φ comes from. It will also motivate the proof for the GOE case.

For an infinite graph, we can define λ_1 by Rayleigh quotients using the adjacency matrix A

$$\lambda_1(G) = \sup_v \frac{\langle v, Av\rangle}{\|v\|_2^2}$$

Clearly, λ_1 is at most the maximal degree in G. This can be used to prove the $a \leqslant 1$ case of the following exercise.

Exercise 3.2.3 (BBP for $\mathbb{Z}^+$)**.**

$$\lambda_1(\mathbb{Z}^+ + \text{ loop of weight } a \text{ on } 0) = \varphi(a)$$

Hint. To prove the lower bound, use specific test functions. When $a > 1$, note that there is an eigenvector $(1, a^{-1}, a^{-2}, ...)$ with eigenvalue $a + \frac{1}{a}$. When $a \leqslant 1$ use the indicator of a large interval. The upper bound for $a > 1$ is more difficult; use rooted convergence and interlacing. □

We will need the following result.

Exercise 3.2.4. Let A be a symmetric matrix, let v be a vector of ℓ^2-norm at least 1, and let $x \in \mathbb{R}$ so that $\|Av - xv\| \leqslant \varepsilon$. Then there is an eigenvalue λ of A with $|\lambda - x| \leqslant \varepsilon$. Hint: consider the inverse of $A - Ix$.

Proof of GOE case. The first observation is that because the GOE is an invariant ensemble, we can replace 11^t by vv^t for any vector v having the same length as the vector 1. We can replace the perturbation with $\sqrt{n}ae_1e_1^t$. Such perturbations commute with tridiagonalization.

Therefore we can consider Jacobi matrices of the form

$$J(a) = \frac{1}{\sqrt{n}} \begin{bmatrix} a\sqrt{n} + N_1 & \chi_{n-1} & & \\ \chi_{n-1} & N_2 & \chi_{n-2} & \\ & \ddots & \ddots & \ddots \end{bmatrix}$$

Case 1: $a \leqslant 1$. Since the perturbation is positive, we only need an upper bound. We use the maximum bound from before. For $i = 1$, the first entry, there was space of size $\sqrt{n}$ below $2\sqrt{n}$. For $i = 1$ the max bound still holds.

Case 2: $a > 1$

Now fix k and let $v = (1, 1/a, 1/a^2, ..., 1/a^k, 0, ..., 0)$. We get that the error from the noise will be of order $1/\sqrt{n}$ so that

$$\left\| \frac{J(a)}{v} - v(a + \frac{1}{a}) \right\| \leqslant ca^{-k}$$

with probability tending to 1.

By Exercise 3.2.4, $J(a)$ has an eigenvalue λ^* that is ca^{-k}-close to $a + 1/a$.

We now need to check that this eigenvalue will actually be the maximum.

Lemma 3.2.5. *Consider adding a positive rank 1 perturbation to a symmetric matrix. Then the eigenvalues of the two matrices will interlace and the shift under perturbation will be to the right.*

By interlacing,

$$\lambda_2(J(a)) \leqslant \lambda_1(J) = 2 + o(1) < a + 1/a - ca^k$$

if we chose k large enough. Thus the eigenvalue λ^* we identified must be λ_1. $\square$

4. The Stochastic Airy Operator

4.1. Global and local scaling. In the Wigner semicircle law the rescaled eigenvalues $\{\lambda_i/\sqrt{n}\}_{i=1}^n$ accumulate on a compact interval and so in the limit become indistinguishable from each other. If we are instead interested in the local interactions between eigenvalues, we need to be able to see the behavior of individual points in the limit.

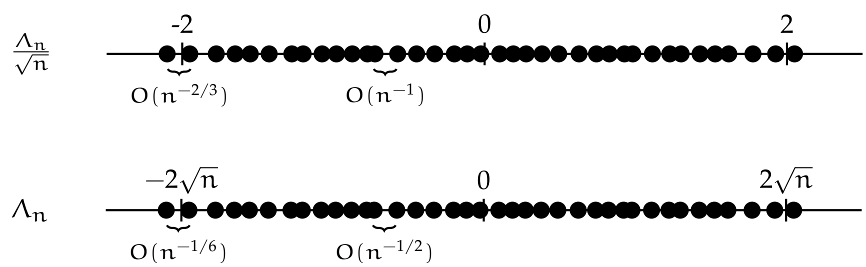

FIGURE 4.1.1. Spectrum of a GOE on the semicircle and original scales

Figure 4.1 shows the spectrum of a GOE on two different scales along with the order of magnitude of the spacing. Notice that at this point if we took $n \to \infty$ for either $\Lambda_n/\sqrt{n}$ or Λ_n all of the points would become indistinguishable even at the edge. To see where the spacing comes from consider the Wigner semicircle law. When n is large we get that for $a < b \in [-2\sqrt{n}, 2\sqrt{n}]$

$$\#\Lambda_n \cap [a,b] \approx n \int_{a/\sqrt{n}}^{b/\sqrt{n}} d\sigma_{sc}(x) = n \int_{a/\sqrt{n}}^{b/\sqrt{n}} \frac{1}{2\pi}\sqrt{4-x^2}\,dx$$

So we expect that for $a \in (-2, 2)$ the process $\sqrt{4-a^2}(\Lambda_n - a\sqrt{n})$ should have average spacing 2π.

Exercise 4.1.2. Check that the typical spacing at the edge $2\sqrt{n}$ of Λ_n is of order $n^{-1/6}$.

See Figure 4.1 for scales. The limiting spectral measure (in this case the semicircle) can be taken as a guide for the correct scale at which to see local interaction but is not guaranteed to give the right answer.

We will further discuss what happens at the edge of the spectrum shortly, but we will state the convergence statement in the bulk now.

Theorem 4.1.3 ([13],[14]). *Let Λ_n have β-Hermite distribution and $a \in (-2, 2)$ then*

$$\sqrt{4-a^2}\sqrt{n}(\Lambda_n - a\sqrt{n}) \Rightarrow \mathrm{Sine}_\beta$$

where Sine_β *is a point process that may be characterized as the eigenvalues of a certain random Dirac operator.*

Remark 4.1.4. These theorems were originally proved for $\beta = 1$ and 2 and stated using the integrable structure of the GOE and GUE. The GUE eigenvalue form a determinantal point process, and the GOE eigenvalues form a Pfaffian point process with kernels constructed from Hermite polynomials. The limiting processes may be identified to looking at the limit of the kernel in the appropriate scale. A version for unitary matrices is proved in [8].

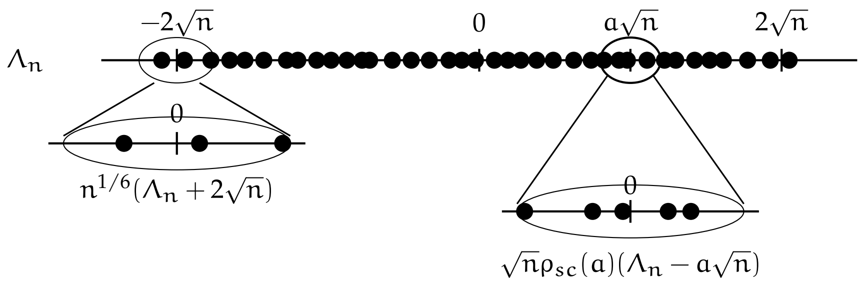

FIGURE 4.1.5. The scale of local interactions

4.2. The heuristic convergence argument at the edge. The goal here is to understand the limiting top eigenvalue of the Hermite β ensembles in terms of a random operator. To do this we look at the geometric structure of the tridiagonal matrix.

Simulations show that the eigenvectors corresponding to the top eigenvalues of the matrix tend to be supported in the first $o(n)$ coordinates. This suggests that the top corner of the matrix determines the behavior of the top eigenvalue.

Now consider the matrix

$$A = \begin{bmatrix} 0 & 1 & & \\ 1 & 0 & 1 & \\ & 1 & 0 & \ddots \\ & & \ddots & \ddots \end{bmatrix}$$

Let $m \leqslant n$ and let Let $f : \mathbb{R}^+ \to \mathbb{R}$ and $v_f = (f(0), f(1/m), f(2/m), ..., f(n/m))^t$. Then $B = m^2(A - 2I)$ acts as a discrete second derivative on f, in the sense that $Bv_f \approx v_{f''}$ as $m, n \to \infty$.

Returning to the β-Hermite case, by Exercise 2.2.1, for $k \ll n$ we have

$$\chi_{n-k} \approx \sqrt{\beta(n-k)} + \mathcal{N}(0, 1/2) \approx \sqrt{\beta}(\sqrt{n} - \frac{k}{2\sqrt{n}}) + \mathcal{N}(0, 1/2)$$

Now we consider the matrix

$$m^\gamma(2\sqrt{n}I - J) \approx \tag{4.2.1}$$

$$m^{\gamma}\sqrt{n}\begin{bmatrix} 2 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 0 & \ddots \\ & & \ddots & \ddots \end{bmatrix} + \frac{m^{\gamma}}{2\sqrt{n}}\begin{bmatrix} 0 & 1 & & & \\ 1 & 0 & 2 & & \\ & 2 & 0 & 3 & \\ & & 3 & 0 & \ddots \\ & & & \ddots & \ddots \end{bmatrix}$$

$$+\frac{m^{\gamma}}{\sqrt{\beta}}\begin{bmatrix} N_1 & \tilde{N}_1 & & \\ \tilde{N}_1 & N_2 & \tilde{N}_2 & \\ & \tilde{N}_2 & N_3 & \ddots \\ & & \ddots & \ddots \end{bmatrix}$$

and assume that we have $m = n^{\alpha}$ for some α. What choice of α should we make? For the first term we want

$$m^{\gamma}\sqrt{n}\begin{bmatrix} 2 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 0 & \ddots \\ & & \ddots & \ddots \end{bmatrix}$$

to behave like a second derivative. This means that $m^{\gamma}\sqrt{n} = m^2$ which gives $2\alpha = \alpha\gamma + 1/2$. We do a similar analysis on the second term. We want this term to behave like multiplication by t. For this we want $\frac{m^{\gamma}}{\sqrt{n}} = \frac{1}{m}$ which gives $\alpha\gamma - 1/2 = -\alpha$. Solving this system we get $\alpha = 1/3$ and $\gamma = 1/2$. For the noise term, multiplication by it should yield a distribution (in the Schwarz sense), which means that its integral over intervals should be of order 1. In other words, the average of m noise terms times m^{γ} should be of order 1. This gives $\gamma = 1/2$, consistent with the previous computations.

This means that we need to look at the section of the matrix that is $m = n^{1/3}$ and we rescale by $n^{1/6}$. That is we look at the matrix

$$n^{1/6}(2\sqrt{n} - J_n)$$

acting on functions with mesh size $n^{-1/3}$.

Exercise 4.2.2. Show that in this scaling, the second matrix in the expansion above has the same limit as the diagonal matrix with $0, 2, 4, 5, 6, ...$ on the diagonal (scaled the same way).

Conclusion. This matrix acting on functions with this mesh size behaves like a differential operator. That is

$$H_n = n^{1/6}(2\sqrt{n} - J_n) \approx -\partial_x^2 + x + \frac{2}{\sqrt{\beta}}b'_x = \mathrm{SAO}_{\beta}$$

here b'_x is white noise. This operator will be called the Stochastic Airy operator (SAO_{β}). We also set the boundary condition to be Dirichlet. This conclusion can be made precise. The heuristics are due to Edelman and Sutton [6], and the proof to Rider, Ramirez and Virag [11].

There are two problems at this point that must be overcome in order to make this convergence rigorous. The first is that we need to actually be able to make sense of that limiting operator. The second is that the matrix even embedded an operator on step functions acts on a different space that the SAO_β so we need to make sense of what the convergence statement should actually be. The following are the main ideas behind the operator convergence.

Remarks on operator convergence

(1) Embed $\mathbb{R}^n$ into $L^2(\mathbb{R})$ via
$$e_i \mapsto \sqrt{m}\mathbf{1}_{[\frac{i-1}{m},\frac{i}{m})}.$$
This gives an embedding of the matrix J acting on a subspace of $L^2(\mathbb{R}^+)$.

(2) It is not clear what functions the Stochastic Airy Operator acts on at this point. Certainly nice functions multiplied by the derivative of Brownian motion will not be functions, but distributions. The only way we get nice functions as results if this is cancelled out by the second derivative. Nevertheless, the domain of SAO_β can be defined.

In any case, these operators act on two completely different sets of functions. The matrix acts on piecewise constant functions, while SAO_β acts on some exotic functions.

(3) The nice thing is that if there are no zero eigenvalues, both H_n^{-1} and J^{-1} can be defined in their own domains, and the resulting operators have compact extensions to the entire L^2.

The sense of convergence we have is
$$\|H_n^{-1} - A_\beta^{-1}\|_{2\to 2} \to 0.$$
This is called norm resolvent convergence, and it implies convergence of eigenvalues and eigenvectors if the limit has discrete simple spectrum.

(4) The simplest way to deal with the limiting operator and the issues of white noise is to think of it as a bilinear form. This is the approach we follow in the next section. The kth eigenvalue can be identified using the Courant-Fisher characterization.

Exercise 4.2.3. We will consider cases where a matrix $A_{n\times n}$ can be embedded as an operator acting on the space of step function with mesh size $1/m_n$. In particular we can encode these step functions in to vectors $v_f = [f(\frac{1}{m_n}), f(\frac{2}{m_n}), \dots, f(\frac{n}{m_n})]^t$.

Let A be the matrix
$$A = \begin{bmatrix} -1 & 1 & & \\ & -1 & 1 & \\ & & \ddots & \ddots \\ & & & -1 \end{bmatrix}$$

For which k_n we get $k_n A v_f \to v_{f'}$?

Exercise 4.2.4. Let A be the diagonal matrix with diagonal entries $(1, 4..., n^2)$. Find a k_n such that $k_n A v_f$ converges to something nontrivial in the limit. What is k_n and what does the limit converge to?

Exercise 4.2.5. Let J be a Jacobi matrix (tridiagonal with positive off-diagonal entries) and v be an eigenvector with eigenvalue λ. The number of times that v changes sign is equal to the number of eigenvalues above λ. More generally the equation $Jv = \lambda v$ determines a recurrence for the entries of v. If we run this recurrence for an arbitrary λ (not necessarily an eigenvalue) and count the number of times that v changes sign this still gives the number of eigenvalues greater than λ.

(1) Based on this gives a description of the number of eigenvalues in the interval $[a, b]$.
(2) Suppose that $v^t = (v_1, ..., v_n)$ solves the recurrence defined by $Jv = \lambda v$. What is the recurrence for $r_k = v_{k+1}/v_k$? What are the boundary conditions for r that would make v an eigenvector?

4.3. The bilinear form SAO$_\beta$. Recall the Airy operator

$$A = -\partial_x^2 + x$$

acting on $f \in L^2(\mathbb{R}^+)$ with boundary condition $f(0) = 0$. The equation $Af = 0$ has two solutions $\mathrm{Ai}(x)$ and $\mathrm{Bi}(x)$, called Airy functions. Note that the solution of $(A - \lambda)f = 0$ is just a shift of these functions by λ.

Since only Ai^2 is integrable, the eigenfunctions of A are the shifts of Ai with the eigenvalues the amount of the shift. We know that the kth zero of the Ai function is at $z_k = -\left(\frac{3}{2}\pi k\right)^{2/3} + o(1)$, therefore to satisfy the boundary conditions the shift must place a 0 at 0, so the kth eigenvalue is given by

$$\lambda_k = -z_k = \left(\frac{3}{2}\pi k\right)^{2/3} + o(1) \tag{4.3.1}$$

The asymptotics are classical.

For the Airy operator A and a.e. differentiable, continuous functions f with $f(0) = 0$ we can define

$$\|f\|_*^2 := \langle Af, f\rangle = \int_0^\infty f^2(x)x + f'(x)^2\, dx. \tag{4.3.2}$$

Let L^* be the space of functions with $\|f\|_* < \infty$.

Exercise 4.3.3. Show that there is $c > 0$ so that

$$\|f\|_2 \leqslant c\|f\|_*$$

for every $f \in L^*$. In particular, $L^* \subset L^2$.

Recall the Rayleigh quotient characterization of the eigenvalues λ_1 of A.

$$\lambda_1 = \inf_{f \in L^*, \|f\|_2 = 1} \langle Af, f\rangle.$$

More generally, the Courant-Fisher characterization is

$$\lambda_k = \inf_{B \subset L^*, \dim B = k} \; \sup_{f \in B, \|f\|_2 = 1} \langle Af, f \rangle,$$

where the infimum is over subspaces B.

For two operators we say $A \leqslant B$ if any $f \in L^*$

$$\langle f, Af \rangle \leqslant \langle f, Bf \rangle.$$

Exercise 4.3.4. If $A \leqslant B$, then $\lambda_k(A) \leqslant \lambda_k(B)$.

Our next goal is to define the bilinear form associated with the Stochastic Airy operator on functions in L^*. Clearly, the only missing part is to define

$$\int_0^\infty f^2(x) b'(x)\, dx.$$

At this point you could say that this is defined in terms of stochastic integration, but the standard L^2 theory is not strong enough – we need it to be defined in the almost sure sense for all functions in L^*. We could define it in the following way:

$$\langle f, b'f \rangle \text{``} = \text{''} \int_0^\infty f^2(x) b'(x) dx = -\int_0^\infty 2f'(x) f(x) b(x) dx$$

This is now a perfectly fine integral, but it may not converge. The main idea will be to write b as its average together with an extra term.

$$b(x) = \int_x^{x+1} b(s) ds + \tilde{b}(x) = \bar{b}(x) + \tilde{b}(x)$$

In this decomposition we get that $\bar{b}$ is differentiable and $\tilde{b}$ is small. The averaging term decouples quickly (at time intervals of length 1), so this term is analogous to a sequence of i.i.d. random variables. We define the inner product in terms of this decomposition as follows.

$$\langle f, b'f \rangle := \langle f, \bar{b}'f \rangle - 2\langle f', \tilde{b}f \rangle$$

It follows from Lemma 4.3.7 below that the integrals on the right hand side are well defined.

Exercise 4.3.5. There exists a random constant C so that we have the following inequality of functions:

$$|\bar{b}'|, |\tilde{b}| \leqslant C\sqrt{\log(2+x)} \tag{4.3.6}$$

Now we return to the Stochastic Airy operator, the following lemma with give us that the operator is bounded from below.

Lemma 4.3.7. *For every $\varepsilon > 0$ there exists random C so that in the positive definite order,*

$$\pm b' \leqslant \varepsilon A + CI,$$

and therefore

$$-CI + (1-\varepsilon)A \leqslant SAO_\beta \leqslant (1+\varepsilon)A + CI.$$

The upper bound here implies that the bilinear form is defined for all functions $f \in L^*$.

Proof. For $f \in L_*$ by our definition,

$$\langle f, b' f\rangle = \langle f, \bar{b}' f\rangle - 2\langle f', \tilde{b} f\rangle$$

using integration by parts in the second term. Now using bounds of the form $-2yz \leqslant y^2/\varepsilon + z^2\varepsilon$ we get the upper bound

$$\langle f, (\bar{b}' + \tilde{b}^2/\varepsilon) f\rangle + \varepsilon\|f'\|^2.$$

By Exercise 4.3.5 there exists a random constant C so that

$$\bar{b}' + \tilde{b}^2/\varepsilon \leqslant \varepsilon x + C.$$

We get the desired bound for $+b'$, and the same arguments works for $-b'$. □

Corollary 4.3.8. *The eigenvalues of* SAO_β *satisfy*

$$\frac{\lambda_k^\beta}{k^{2/3}} \to \left(\frac{2\pi}{3}\right)^{2/3} \qquad a.s.$$

Proof. It suffices to show that a.s. for every rational $\varepsilon > 0$ there exists $C_\varepsilon > 0$ so that

$$(1-\varepsilon)\lambda_k - C_\varepsilon \leqslant \lambda_k^\beta \leqslant (1+\varepsilon)\lambda_k + C_\varepsilon$$

where the λ_k are the Airy eigenvalues (4.3.1). But this follows from the operator inequality of Lemma 4.3.7 and Exercise 4.3.4. □

If you look at the empirical distribution of the eigenvalues as $k \to \infty$ then the "density" behaves like $\sqrt{\lambda}$. More precisely, the number of eigenvalues less than λ is of order $\lambda^{3/2}$. This is the Airy-β version of the Wigner semicircle law. Only the edge of the semicircle appears here.

4.4. Convergence to the Stochastic Airy Operator. The goal of this section is to give a rigorous convergence argument for the extreme eigenvalues to those of the limiting operator. To avoid technicalities in the exposition, we will use a simplified model, which has the features of the tridiagonal beta ensembles. Consider the $n \times n$ matrix

$$H_n = n^{2/3}\begin{bmatrix} 2 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & \ddots \\ & & \ddots & \ddots \end{bmatrix} + n^{-1/3}\mathrm{diag}(1,2,3,\ldots) + \mathrm{diag}(N_{n,1}, N_{n,2}, \ldots) \tag{4.4.1}$$

Here for each n the $N_{n,i}$ are independent centered normal random variables with variance $\frac{4}{\beta}n^{-1/3}$. This is a simplified version of (4.2.1).

We couple the randomness by setting

$$N_{n,i} = b(in^{-1/3}) - b((i-1)n^{-1/3})$$

for a fixed Brownian motion b which, here for notational simplicity, has variance $4/\beta$. From now on we fix b and our arguments will be deterministic, so we drop the a.s. notation.

We now embed the domains $\mathbb{R}^n$ of H_n into $L^2(\mathbb{R}^+)$ by the map

$$e_i \mapsto n^{1/6}\mathbf{1}_{[\frac{i-1}{n^{1/3}},\frac{i}{n^{1/3}})},$$

and denote the $\tilde{\mathbb{R}}^n$ the isometric image of $\mathbb{R}^n$ in this embedding. Let $-\Delta_n$, x_n and b_n be the images of the three matrix terms on the right of (4.4.1) under this map, respectively. For $f \in \tilde{\mathbb{R}}^n$, let

$$\|f\|_{*n}^2 = \langle f, (-\Delta_n + x_n)f\rangle$$

and recall the L_* norm $\|f\|_*$ from (4.3.2).

We will need some standard analysis Lemmas.

Exercise 4.4.2. Let $f \in L_*$ of compact support. Let f_n be its orthogonal projection to $\tilde{\mathbb{R}}^n$. Then $f_n \to f$ in L^2, and $\langle f_n, H_n f_n\rangle \to \langle f, Hf\rangle$.

Let $\lambda_{n,k}, \lambda_k$ denote the kth lowest eigenvalue of H_n and the Stochastic Airy Operator $H = SAO_\beta = -\partial_x^2 + x + b'$, respectively.

Proposition 4.4.3. $\limsup \lambda_{n,1} \leqslant \lambda_1$.

Proof. For $\varepsilon > 0$ let f be of compact support and norm 1 so that $\langle f, Hf\rangle \leqslant \lambda_1 + \varepsilon$. Let f_n be the projection of f to $\tilde{\mathbb{R}}^n$. Then by Exercise 4.4.2 we have

$$\lambda_{n,1} \leqslant \frac{\langle f_n, H_n f_n\rangle}{\|f_n\|^2} \to \langle f, Hf\rangle \leqslant \lambda_1 + \varepsilon$$

since ε is arbitrary, the claim follows. □

For the upper bound, we need a tightness argument for eigenvectors and eigenvalues.

Exercise 4.4.4. Show that for every $\varepsilon > 0$ there is a random constant C so that

$$\pm b_n \leqslant \varepsilon(-\Delta_n + x_n) + CI$$

in the positive definite order for all n. Hint: use a version of the argument in Lemma 4.3.7.

Note that this exercise implies

$$H_n \geqslant (1-\varepsilon)(-\Delta_n + x_n) - CI$$

and since $-\Delta_n + x_n$ is positive definite, it follows that $\lambda_{n,1} \geqslant -C$, which is a Füredi-Komlós type bound, but now of the right order! (Compare to 3.1.3).

Finally another piece of necessary analysis given as an exercise.

Exercise 4.4.5. Let $f_n \in \tilde{\mathbb{R}}^n$ with $\|f_n\|_{*n} \leqslant c$ for all n. Then f_n has a subsequential limit f in L^2 so that along that subsequence

$$\liminf \langle f_n, H_n f_n \rangle \geqslant \langle f, Hf \rangle.$$

A solution to this analysis problem can be found in [11].

Proposition 4.4.6. $\liminf \lambda_{n,1} \geqslant \lambda_1$.

Proof. By Exercise 4.4.4, in the positive definite order,

$$H_n \leqslant (1+\varepsilon)(-\Delta_n + x_n) + CI$$

but since $\Delta_n + x$ is nonnegative definite, $\lambda_{1,n} \leqslant C$.

Now let $(f_n, \lambda_{n,1})$ be the eigenvector, lowest eigenvalue pair for H_n, so that $\|f_n\| = 1$. Then by Exercise 4.4.4

$$(1-\varepsilon)\|f_n\|_{*n} \leqslant \langle f_n, H_n f_n \rangle + C = \lambda_{n,1} + C \leqslant 2C.$$

Now consider a subsequence along which $\lambda_{n,1}$ converges to its lim inf. By Exercise 4.4.5 we can find a further subsequence of f_n so that $f_n \to f$ in L^2, and

$$\liminf \lambda_{n,1} = \liminf \langle f_n, H_n f_n \rangle \geqslant \langle f, Hf \rangle \geqslant \lambda_1,$$

as required. □

Exercise 4.4.7. Modify the proofs above using the Courant-Fisher characterization to show that for every k, we have $\lambda_{n,k} \to \lambda_k$.

4.5. Tails of the Tracy Widom$_\beta$ distribution.

Definition 4.5.1. We define the Tracy-Widom-β distribution

$$TW_\beta = -\lambda_1(A_\beta)$$

In the case $\beta = 1, 2$ this is consistent with the classical definition.

As for the tails, they are asymmetric. Our methods can be used to show that as $a \to \infty$ the right tail satisfies

$$P(TW_\beta > a) = \exp(-\frac{2+o(1)}{3}\beta a^{3/2})$$

and we will prove that the left tail satisfies the following.

Theorem 4.5.2 ([11])**.**

$$P(TW_\beta < -a) = \exp(-\frac{\beta + o(1)}{24} a^3) \qquad \textit{as} \quad a \to \infty.$$

Proof of the upper bound. Suppose we have $\lambda_1 > a$, this implies that for all $f \in L^*$ we get the bound

$$\langle f, A_\beta f \rangle \geqslant a\|f\|_2^2.$$

Therefore we are interested in the probability

$$P\left(\|f'\|_2^2 + \|\sqrt{x}f\|_2^2 + \frac{2}{\sqrt{\beta}}\int f^2 b' dx \geqslant a\|f\|_2^2\right)$$

The first two terms are deterministic, and for f fixed the third term is a Paley-Wiener integral. In particular, it has centered normal distribution with variance

$$\frac{4}{\beta}\Big(\int f^4 dx\Big) = \frac{4}{\beta}\|f\|_4^4.$$

This leads us to computing

$$P\left(\|f'\|_2^2 + \|\sqrt{x}f\|_2^2 + N\|f\|_4^2 \geqslant a\|f\|_2^2\right)$$

where N is a normal random variable with variance $4/\beta$. Using the standard tail bound for a normal random variable we get

(4.5.3)

$$P\left(\|f'\|_2^2 + \|\sqrt{x}f\|_2^2 + N\|f\|_4^2 \geqslant a\|f\|_2^2\right) \leqslant 2\exp\left(-\frac{\beta(a\|f\|_2^2 - \|f'\|_2^2 - \|f\sqrt{x}\|_2^2)^2}{8\|f\|_4^4}\right)$$

We want to optimize over possible choices of f. It turns out the optimal f will have small derivative, so we will drop the derivative term and then optimize the remaining terms. That is we wish to maximize.

$$\frac{(a\|f\|_2^2 - \|f\sqrt{x}\|_2^2)^2}{\|f\|_4^4}$$

With some work we can show that the optimal function will be approximately $f(x) \approx \sqrt{(a-x)^+}$. This needs to be modified a bit in order to keep the derivative small, so we cut this function off and replace is with a linear piece

$$f(x) = \sqrt{(a-x)^+} \wedge (a-x)^+ \wedge x\sqrt{a}$$

We can check that

$$a\|f\|_2^2 \sim \frac{a^3}{2} \qquad \|f\| \sim O(a) \qquad \|\sqrt{x}f\|^2 \sim \frac{a^3}{6} \qquad \|f\|_4^4 \sim \frac{a^3}{3}.$$

Using these values in equation (4.5.3) give us the correct upper bound. □

Proof of the lower bound. We begin by introducing the Riccati transform: Suppose we have an operator

$$L = -\partial_{xx} + V(x)$$

then the eigenvalue equation is

$$\lambda f = (-\partial_{xx} + V(x))f$$

We can pick a λ and attempt to solve this equation. The left boundary condition is given, so one can check if the solution satisfies $f \in L^*$, in which case we get an eigenfunction. Most of the time this won't be true, but we can still gain information by studying these solutions. To study this problem we first make the transformation

$$p = \frac{f'}{f}, \qquad \text{which gives} \qquad p' = V(x) - \lambda - p^2, \qquad p(0) = \infty.$$

The following is standard part of the theory for Schrdinger operators of the form SAO_β, although some technical work is needed because the potential is irregular.

Proposition 4.5.4. *Choose λ, we will have $\lambda \leqslant \lambda_1$ if an only if the solution to the Ricatti equation does not blow up.*

The slope field looks as follows. When $V(x) = x$ and $\lambda = 0$ there is a right facing parabola $p^2 = x$ where the upper branch is attracting and the lower branch is repelling. The drift will be negative outside the parabola and positive inside. Shifting the initial condition to the left is equivalent to shifting the λ to the right, so this picture may be used to consider the problem for all λ.

Now replace b'. The solution of the Ricatti equation is now a diffusion. In this case there is some positive chance of the diffusion moving against the drift, including crossing the parabola. If we use $P_{-\lambda,y}$ to denote the probability measure associated with starting our diffusion with initial condition $p(-\lambda) = y$, then we get

$$P(\lambda_1 > a) = P_{-a,+\infty}\,(\text{p does not blow up})$$

Because diffusion solution paths do not cross, we can bound this below by starting our particle at 1.

$$P_{-a,+\infty}\,(\text{p does not blow up}) \geqslant P_{-a,1}\,(\text{p does not blow up})$$

We now bound this below by requiring that our diffusion stays in $p(x) \in [0, 2]$ on the interval $x \in [-a, 0)$ and then choosing convergence to the upper edge of the parabola after 0. This gives

$$\begin{aligned}&P_{-a,1}\,(\text{p does not blow up})\\ &\qquad\geqslant P_{-a,1}\,(\text{p stays in } [0,2] \text{ for } x < 0) \cdot P_{0,0}\,(\text{p does not blow up})\,.\end{aligned}$$

The second probability is a constant not depending on a.We focus on the first event.

A Girsanov change of measure can be used to determine the probability. This change of measure moves us to working on the space where p is replaces by a standard Brownian motion (started at 1). The Radon-Nikodym derivative of this change of measure may be computed explicitly. We compute

$$\begin{aligned}&P_{-a,1}\,(\text{p stays in } [0,2] \text{ for } x < 0) = E_{-a,1}\left[\mathbf{1}(p_x \in [0,2], x \in (-a,0))\right]\\ &= E_{-a,1}\left[\exp\left(\frac{\beta}{4}\int_{-a}^{0}(x-b^2)db - \frac{\beta}{8}\int_{-a}^{0}(x-b^2)^2dx\right)\mathbf{1}(b_x \in [0,2], x \in (-a,0))\right].\end{aligned}$$

Notice that

$$\frac{\beta}{4}\int_{-a}^{0}(x-b^2)db \sim O(a), \qquad \text{and} \qquad \frac{\beta}{8}\int_{-a}^{0}(x-b^2)^2dx \approx -\frac{\beta}{24}a^3,$$

which gives us the desired lower bound. □

References

[1] J. Baik, G. Ben Arous, and S. Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probab.* 33(5), 1643-1697. 2005. 12

[2] A. Bloemendal. Finite Rank Perturbations of Random Matrices and Their Continuum Limits *PhD Thesis, University of Toronto*, 2011. 12

[3] A .Bloemendal, and B. Virág. Limits of spiked random matrices I. *Probab. Theory and Related Fields* 156.3-4: 795-825. 2013. 12

[4] P. Deift, T. Trogdon, and G. Menon. On the condition number of the critically-scaled Laguerre unitary ensemble. *Discrete and Cont. Dyn. Sys.*, 36(8):4287–4347, 2016.

[5] I. Dumitriu and A. Edelman. Matrix models for beta ensembles. *J. Math. Phys.*, 43(11):5830–5847, 2002. 6

[6] A. Edelman, B. Sutton From random matrices to stochastic operators. *Journal of Statistical Physics*, 127(6), 1121–1165, 2007. 17

[7] S. N. Ethier and T. G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986. Characterization and convergence.

[8] R. Killip, and M. Stoiciu. Eigenvalue statistics for CMV matrices: from Poisson to clock via random matrix ensembles. *Duke Math. Journal* 146.3: 361-399. 2009. 15

[9] M. Krishnapur, B. Rider, and B. Virág. Universality of the stochastic Airy operator. *Comm. Pure Appl. Math.*, 69:145–199, 2016. 6

[10] J. A. Ramírez and B. Rider. Diffusion at the random matrix hard edge. *Comm. Math. Phys.*, 288(3):887–906, 2009.

[11] J. A. Ramírez, B. Rider, and B. Virág. Beta ensembles, stochastic Airy spectrum, and a diffusion. *J. Amer. Math. Soc.*, 24(4):919–944, 2011. 17, 22

[12] H. F. Trotter. Eigenvalue distributions of large Hermitian matrices; Wigner's semicircle law and a theorem of Kac, Murdock, and Szegő. *Adv. in Math.*, 54(1):67–82, 1984. 4

[13] B. Valkó and B. Virág. Continuum limits of random matrices and the Brownian carousel. *Invent. Math.*, 177:463–508, 2009. 14

[14] B. Valkó and B. Virág. The Sine$_\beta$ operator. *Invent. Math.* 209:275-327, 2017. 14

[15] B. Virág. . Operator limits of random matrices. *Proceedings of the ICM.* 2014.

10

KTH, Lindstedtsvägen 25, 100 44 Stockholm, Sweden
E-mail address: holcomb@kth.se

University of Toronto, 40 St George St., Toronto, ON, M5S 2E4, Canada
E-mail address: balint@math.toronto.edu