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. It is a quick introduction to the theory of large random matrices through limits that preserve their operator structure, rather than just their eigenvalues. This 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, quick proofs for the Füredi-Komlós result on the top eigenvalue, the BBP phase transition, as well as local convergence of the soft-edge process and tail asymptotics for the TW_β distribution.

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. Consider the matrix

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

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 by 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.

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

Theorem 1.1.1 (see e.g. [2]). *Suppose M_n has GOE or GUE distribution then M_n has eigenvalue density*

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

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.

For convenience set $\Lambda = \Lambda_n = \{\lambda_i\}_{i=1}^n$ 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 the density in Theorem 1.1.1 we can see that this is a model for n particles that would like to be Gaussian, but the Vandermonde term $\prod_{i<j} |\lambda_i - \lambda_j|^\beta$ pushes them apart. Note that $\mathrm{Tr}\ M_n^2/n^2 \to 1$ 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.

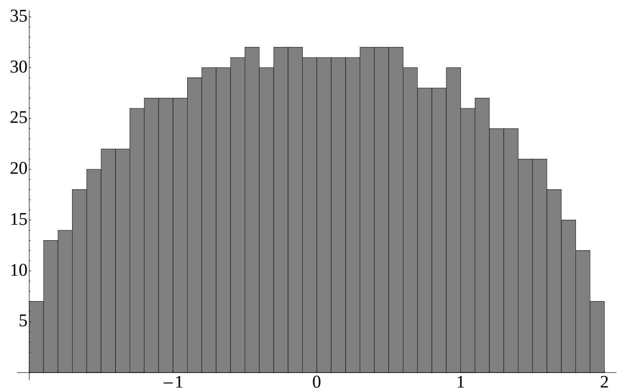

FIGURE 1.1.3. Rescaled eigenvalues of a 1000×1000 GOE matrix

We begin by introducing a tridiagonal matrix model that has the same joint density as the Gaussian ensembles. This model and Jacobi matrices more generally share many characteristics with differential operator theory including Sturm-Liouville theory. In this section we derive the tridiagonal model for the Gaussian unitary ensemble and then give tridiagonal models for a wider class of models called β-ensembles. In Section 2 we begin by introducing two different notions of graph convergence which are then used to prove the Wigner semicircle law. In Section 3 we study the behavior of the largest eigenvalue of the GOE under rank one perturbations. The behavior depends on the strength of the perturbation and is called the Baik-Ben Arous-Péché transition. In Section 4 we introduce the notion of local convergence and give a proof of local convergence at the soft edge. We also study the tail behavior of the smallest eigenvalue of the Stochastic Airy Operator which is the $\beta > 0$ generalization of the Tracy-Widom 1,2, and 4 laws. Section 5 gives a partial overview of results that are proved using operator methods including other local limit theorems.

1.2. Tridiagonalization and spectral measure. The spectral measure of a matrix at a vector (which we will take to be the first coordinate vector) is a measure supported on the eigenvalues that reflects the local structure of the matrix there. This is more easily seen in the case where the matrix is the adjacency matrix of a (possibly weighted) graph. In this case the spectral measure at coordinate vector j gives information about the graph in the neighborhood of vertex j.

Definition 1.2.1. The spectral measure σ_A of a symmetric matrix A at the first coordinate vector is the measure for which

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

From this definition it is not clear that the spectral measure exists. The advantage is that this definition can be extended to the operator setting in many cases. For bounded operators (e.g. matrices) uniqueness is clear since probability measures with compact support are determined by their moments (see e.g. Kallenberg [21]). For finite matrices, the following alternative explicit definition proves existence.

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.

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

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.

Is there 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 Hermitian 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 representatives for the class of matrices for which the first coordinate vector is cyclic.

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

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

Proof of Existence. It is possible to 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 choose the orthogonal matrix O so that Ob is in the direction of the first coordinate vector, namely $Ob = |b|\mathbf{e}_1$.

An explicit option for O is the following Householder reflection:

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

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

Therefore

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

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 are 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 $J^k_{1,1}$ 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 $\mathbf{e}_1$ is cyclic. Apply Gram-Schmidt to the vectors $(\mathbf{e}_1, A\mathbf{e}_1, ..., A^{n-1}\mathbf{e}_1)$ 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. The tridiagonalization procedure can be applied to the GOE matrix, as pioneered by Trotter [33]. The invariance of the distribution under independent orthogonal transformations yields a tractable Jacobi matrix.

Proposition 1.2.9 (Trotter [33])**.** *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.

Let v be a vector of independent $\mathcal{N}(0,1)$ random variables of length k, then the length of the vector has χ distribution with parameter k, $\chi_k \stackrel{d}{=} \|v\|$. The density of a χ random variable 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 tridiagonalization algorithm 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. The independence propagates throughout the algorithm meaning each rotation defined produces the relevant tridiagonal terms and an independent submatrix. □

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 orthogonal 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 orthogonal 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.

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

Second, for the set of such probability measures, the map $J \mapsto \sigma_J$ is onto. This is left as an exercise.

Exercise 1.2.12. For every probability measure σ supported on exactly n points there exists an $n \times n$ symmetric matrix with spectral measure σ. The existence part of Theorem 1.2.6 then implies that there exists a Jacobi matrix with spectral measure σ.

1.3. β-ensembles. Let

$$A_n = \frac{1}{\sqrt{\beta}} \begin{bmatrix} a_1 & b_1 & 0 & \dots & 0 \\ b_1 & a_2 & \ddots & & \\ 0 & \ddots & \ddots & \ddots & \\ \vdots & & \ddots & a_{n-1} & b_{n-1} \\ 0 & & & b_{n-1} & a_n \end{bmatrix}, \tag{1.3.1}$$

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. We will frequently use the notation $a_i = \mathcal{N}_i$ in order to refer more directly to the distribution of the random variable. 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 1.3.2 (Dumitriu, Edelman, [12])**.** *If $\beta > 0$ then the joint density of the eigenvalue of A_n is given by*

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

The spectral measure of a Jacobi matrix may be written as $\sigma_J = \sum_{j=1}^n \delta_{\lambda_j} q_j^2$, where $\lambda_1, ..., \lambda_n$ are the eigenvalues of the matrix and the $q_1, ..., q_n$ are the associated the spectral weights with $\sum q_i^2 = 1$. Recall that the map $J \mapsto \sigma_J$ is a bijection, so one possible strategy is to use this map to compute the distribution of the eigenvalues and spectral weights from the Jacobi matrix entries by the change-of-variable formula. This is possible since the Jacobian determinant can be computed

Dumitriu and Edelman used direct computation of the Jacobian of the map $(\overline{\lambda}, \overline{q}) \to (\overline{a}, \overline{b})$ to prove Theorem 1.3.2. This computation may be simplified by working through the moments, which have a simple connection to both representations:

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

One can look at maps from both sets to $(m_1, ..., m_{2n-1})$. Notice that $2n-1$ moments are required as there are $2n-1$ variables in both the spectral and tridiagonal basis. These are simple transformations and one can write down the appropriate matrices and then find their determinants. These computations can be found in [25], and yield the following.

Theorem 1.3.3 (Dumitriu, Edelman, Krishnapur, Rider, Virág, [12, 25])**.** *Let V a real-valued function, and $\overline{a}, \overline{b}$ are chosen from then density proportional to*

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

assuming such a denisty exists (i.e. the total integral is finite). Then the eigenvalues have distribution

$$f(\lambda_1, .., \lambda_n) = \frac{1}{Z_{n,\beta}} \exp\left(-\sum_i V(\lambda_i)\right) \prod_{i<j} |\lambda_i - \lambda_j|^\beta \tag{1.3.4}$$

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 [25] to get universality for the top eigenvalue.

2. Graph Convergence and the Wigner semicircle law

The approach of operator limits is a case of the "objective method" pioneered by Aldous. In limit theories, it is best to understand the limit of the a high-level structured object, such as a matrix or a graph, rather than just its statistics (such as eigenvalues, or graph-related quantities such as triangle density).

In this section we will use Jacobi matrices together with graph convergence in order to give proofs of the Wigner semicircle law. The graphs themselves are operators through their adjacency matrix. While it is not helpful here to formalize this correspondence, we still think of graph convergence as an example of operator limits.

We begin by defining the spectral measure of a graph and give an introduction to different notions of graph convergence. We finish by proving the Wigner semicircle law in two different ways using different notions of graph convergence.

2.1. Graph convergence. The proof of the Wigner semicircle law given here will rest on a graph convergence argument. First we introduce 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 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.

The definition of the spectral measure (1.2.1) works even for infinite graphs, but it is again not a priori clear that it exists or it is unique.

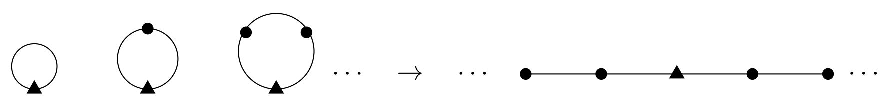

FIGURE 2.1.1. Rooted convergence: n-cycles to $\mathbb{Z}$

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 radius r in G_n about ρ equals that in G for all large enough n.

Examples 2.1.3. 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. With 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) in 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 graphs with degrees bounded by b 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, ρ).

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 defined by* (1.2.1).

Exercise 2.1.5. Consider a straight line with n vertices spaced evenly 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 path of length n is easy to diagonalize. This is an example where the spectral measure has a different limit than the eigenvalue distribution.

Exercise 2.1.6. Consider the random d-regular graph (G_n, ρ) on n vertices in the configuration model (for a given degree sequence, choose a uniform random matching on the half edges attached to each vertex). Show that (G_n, ρ) converges to the d-regular infinite tree in probability.

In the case where there is no designated root we will need a different notion of convergence.

Definition 2.1.7 (Benjamini-Schramm, [4])**.** For a sequence of graphs $\{G_n\}$ choose a vertex uniformly at random to be the root. The graphs converge in the **Benjamini-Schramm** sense if this random sequence of rooted graphs converges in distribution with respect to rooted convergence to a random rooted graph.

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 randomly rooted regular graphs (G, o) unimodularity means that if for a uniformly chosen random neighbor ν of ρ in G, the triple (G, ρ, ν) has the same distribution as (G, ν, ρ). For the general case, in order to define unimodularity the distributions have to first be biased by the degree of the root, see e.g. [27].

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.

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. For more on this see [1].

Proposition 2.1.9. *Let G be a fixed finite graph and choose a root ρ uniformly at random from the vertex set $V(G)$. This defines a random rooted 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

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

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. 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 is Benjamini-Schramm convergence related to the eigenvalue distribution? First, an exercise abound random probability measures, useful here and in the sequel.

Exercise 2.1.13. Let ν_n be a sequence of random probability measures, and assume that $\nu_n \to \nu$ in distribution with respect to the weak topology.

(1) Show that $E\nu_n \to E\nu$ weakly.

(2) Assume that ν is deterministic. Let $\mathcal{F}_n$ be an arbitrary sequence of σ-fields. Define the random probability measure $E(\nu_n|F_n)$ by $E(\nu_n|F_n)(A) = E(\nu_n(A)|F_n)$ for measurable sets A. Show that $E(\nu_n|F_n) \to \nu$ in probability with respect to the weak topology.

Assume that $G_n \to (G, \rho)$ in the Benjamini-Schramm sense. By the continuity theorem (applied to the rooted convergence topology), the distribution of the random spectral measures σ_n converges to the distribution of $\sigma_{G,\rho}$.

By Exercise 2.1.13 (1) this implies that the eigenvalue distributions converge as well:

Theorem 2.1.14. *Let G_n be a sequence of graphs with eigenvalue distributions μ_n. If $G_n \to (G, \rho)$ in the Benjamini-Schram sense, then $\mu_n = E\sigma_n \to E\sigma_{G,\rho}$ weakly.*

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 2.1.15 (The spectral measure of $\mathbb{Z}$)**.** We use Benjamini-Schramm convergence of the cycle graph to $\mathbb{Z}$. First we compute 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. Since $A = T + T^{-1}$ 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, the 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 2.1.16. 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.

2.2. Wigner's semicircle law.

Theorem 2.2.1 (Wigner's semicircle Law)**.** *Let Λ_n have β-Hermite distribution and let $\mu_n = \frac{1}{n}\sum_{j=1}^n \delta_{\lambda_j/\sqrt{n}}$ be the empirical distribution of $\Lambda_n/\sqrt{n}$. Then*

$$\mu_n \Rightarrow \mu_{sc}, \qquad \textit{where} \qquad \frac{d\mu_{sc}}{dx} = \frac{1}{2\pi}\sqrt{4-x^2}1_{x\in[-2,2]}\, dx.$$

The following exercise provides the necessary tools to give several different proofs of Wigner's semicircle law. You can attempt the exercise first or read on in order to see more details of the proof.

Exercise 2.2.2. Let A be a rescaled $n \times n$ Dumitriu-Edelman tridiagonal matrix

$$A = \frac{1}{\sqrt{\beta n}}\begin{bmatrix} a_1 & b_1 & & & \\ b_1 & a_2 & b_2 & & \\ & b_2 & \ddots & \ddots & \\ & & \ddots & a_{n-1} & b_{n-1} \\ & & & b_{n-1} & a_n \end{bmatrix}, \quad b_i \sim \chi_{\beta(n-i)}, \quad a_i \sim \mathcal{N}(0,2)$$

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?

Note that a Jacobi matrix can be thought of as the adjacency matrix of a weighted path with loops. For all of the proofs of the Wigner semi-circle law we will use the graph with the adjacency matrix given by the rescaled Dumitriu-Edelman model given in Exercise 2.2.2, see Figure 2.2.3.

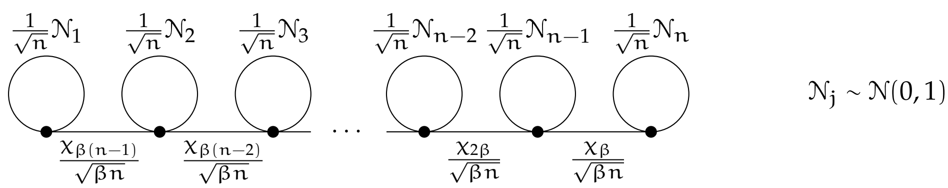

FIGURE 2.2.3. Unrooted rescaled Dumitriu-Edelman graph

Exercise 2.2.4. Check that $\chi_n - \sqrt{n} \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0, 1/2)$.

Proof 1. [37]

Take the graph associated to the rescaled Dumitriu-Edelman tridiagonal matrix shown in Figure 2.2.3, and then take a Benjamini-Schramm limit.

The convergence is in probability to a random rooted graph. Note that there are two levels of randomness, one coming from the fact that we take a limit of random weighted graphs (not just weighted graphs) and the second from the fact that even deterministic graphs have Benjamini-Schramm given by a random rooted graph. The first kind of randomness is lost in the limit, hence the convergence in probability.

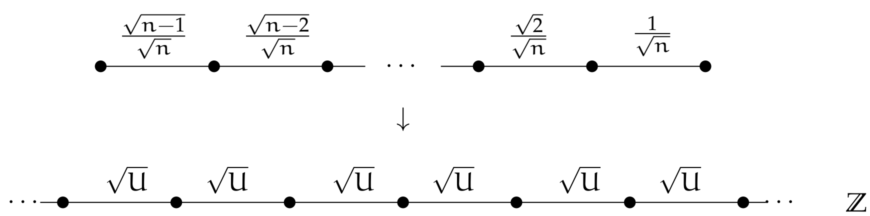

FIGURE 2.2.5. Terms from the graph visible in Benjamini-Schramm convergence.

Notice that an application of Exercise 2.2.4 will give us that it is enough to consider the graph with no loops and deterministic edge labels $\sqrt{\frac{n-k}{n}}$. See Figure 2.2.5.

What is the limit? The structure clearly converges to $\mathbb{Z}$. The edge weights in a randomly rooted neighborhood converge in distribution to $\sqrt{U}$ for a single random variable U uniform in $[0,1]$. Let σ denote the spectral measure $\mathbb{Z}$ with these edge weights. Theorem 2.1.14 and continuity theorem implies that $\mu_n \to E\sigma$ in probability.

Now σ is a rescaling of the spectral measure of $\mathbb{Z}$ by $\sqrt{U}$. By Exercise 2.1.15 σ is a scaled arcsine measure that corresponds to the projection of a circle of radius $\sqrt{U}$.

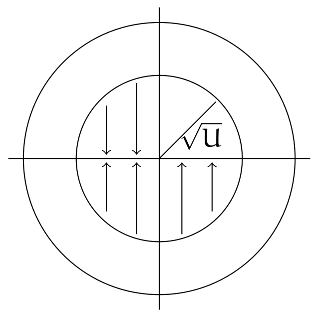

FIGURE 2.2.6. The role of U in the limiting distribution and projection

The distribution of a point chosen with radius $\sqrt{U}$ and uniform angle is in fact a uniform random point in the disk. Thus $E\sigma$ is the projection of the uniform measure on the disk to the real line. See figure 2.2.6. This is the semicircle law.

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

□

Proof 2. Going back to the full matrix model of GOE, let A_n be $1/\sqrt{n}$ times the tridiagonalization of GOE from a uniformly chosen random vertex v. This vertex will correspond to the first coordinate in the tridiagonalization. Let $\mathcal{F}_n$ be a sigma-field generated by the randomness in the nth GOE, but not the choice of the vertex. Then with $\sigma_n = \sigma_{A_n,v}$ the empirical eigenvalue distribution of GOE satisfies

$$\mu_n = E(\sigma_n|\mathcal{F}_n).$$

It suffices to show that σ_n converges to the semicircle law in probability, and conclude by Exercise 2.1.13 (2).

A_n has $1/\sqrt{n}$ times the distribution (1.3.1). Consider the rooted limit of (A_n, v). In Figure 2.2.3 v is the leftmost point on the graph.

This random weighted graph converges in probability with respect to the rooted convergence topology to $\mathbb{Z}_+$. Therefore by the continuity theorem and Theorem 2.1.14 $\sigma_n \to \sigma_{\mathbb{Z}_+} = \rho_{sc}$ in probability, as required. □

The limit of the spectral measure at the first vertex should have nothing to do with the limit of the eigenvalue distribution in the general case. The Jacobi matrices that we get in the case of the GOE are special.

3. The top eigenvalue and the Baik-Ben Arous-Péché 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 matching upper bound does not follow and needs more work. This is the content of the following theorem.

Theorem 3.1.1 (Füredi-Komlós [14]).

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

This holds for more general entry distributions in the original matrix model; 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

$$J = -AA^t + \mathrm{diag}(a_i + b_i + b_{i-1}),$$

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

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

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

Applying 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 $\sqrt{\log n}$ outside of the spectrum. Indeed we will see that

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

for some distribution TW_1, so the bound above is not optimal.

3.2. Baik-Ben Arous-Péché transition. The approach taken here is a version of a section in Bloemendal's PhD thesis [7].

Historically random matrices have been used to study correlations in data sets. To see whether correlations are significant, one compares to a case in which data is sampled randomly without correlations.

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 where 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)**.** *Let* X_n *be an* $n \times n$ *matrix with* $n <$ *and independent* $\mathcal{N}(0,1)$ *entries, then*

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

where

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

Heuristically, correlation in the population appears in the asymptotics in the top eigenvalue of the sample 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 distributional limit of the top eigenvalue. When $a < 1$ the distribution is unchanged from the unperturbed case, the limit being Tracy-Widom. When $a > 1$ the top eigenvalue separates and has limiting Gaussian fluctuations. Close to the point $a = 1$ a deformed Tracy-Widom distribution appears, see [3], [5].

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)**.**

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

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 next exercise in the proof of 3.2.2, but include it to show where the function φ comes from. It will also motivate the proof for the GOE case.

Exercise 3.2.3 (BBP for $\mathbb{Z}^+$)**.** For an infinite graph, we can define λ_1 by Rayleigh quotients using the adjacency matrix A

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

(1) Show that λ_1 is at most the maximal degree in G.
(2) Prove that for $a \leqslant 1$

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

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 Theorem 3.2.2. 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 $\sqrt{n}a\mathbf{e}_1\mathbf{e}_1^t$. 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 $v = (1, 1/a, 1/a^2, ..., 1/a^k, 0, ..., 0)$. Thus the error from the noise will be of order $1/\sqrt{n}$ so that

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

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.

Exercise 3.2.5. Let A, P be asymmetric matrices, with $P \geqslant 0$ of rank 1. Then the eigenvalues of A and $A + P$ interlace and the shift under perturbation is to the right.

Hint: use the Courant-Fisher characterization.

By interlacing,

$$\lambda_2(J(a)) \leqslant \lambda_1(J(0)) = 2 + o(1) < a + 1/a - ca^k$$

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

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. For the local interactions between eigenvalues, the behavior of individual points has to prevail in the limit.

To make a guess at the correct spacing required to see individual points in the limit we begin by pretending that they are quantiles of the Wigner semicircle law. When n is large we get that for $a < b \in [-2, 2]$

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

So we expect that for $a \in (-2, 2)$ the process $\sqrt{n(4 - a^2)}(\Lambda_n - a\sqrt{n})$ should have average spacing $\frac{1}{2\pi}$.

Exercise 4.1.1. Check that the typical spacing at the edge $2\sqrt{n}$ of Λ_n is of order $n^{-1/6}$.

The correct scales needed to obtain a local limit are give in Figure 4.1.2. These notes will focus on the convergence result for the edge of the spectrum. The statement for the bulk and more on the operator viewpoint will be discussed in Section 5.

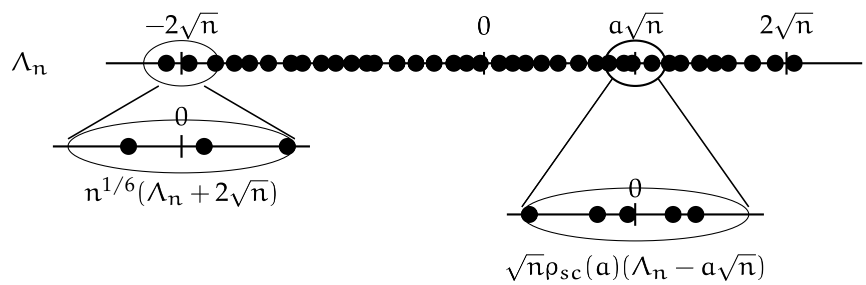

FIGURE 4.1.2. The scale of local interactions

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. Jacobi matrices are frequently associated with differential equations and sometimes studied under the name of discrete Schrödinger operators. To see the connection with Schrödinger operators consider the following example:

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

The semi-infinite version of this is frequently called the **discrete Laplacian**. To understand this name let m be large for $f : \mathbb{R}^+ \to \mathbb{R}$ define the discretization $\nu_f = (f(0), f(1/m), f(2/m), ...)^t$. Then $B = m^2(A - 2I)$ acts as a discrete second derivative on f, in the sense that $B\nu_f \approx \nu_{f''}$ as $m \to \infty$. For this to hold in the first entry we need to further assume that f satisfies a Dirichlet boundary condition $f(0) = 0$. This convergence argument may be easily extended to matrices of the form $(A - 2I) + D$ where D is a semi-infinite diagonal matrix with entries $D_k = V(\frac{k}{m})$ for some function $V : \mathbb{R} \to \mathbb{R}$. In this case the matrices converge to the Schrödinger operator $\Delta + V$.

Now apply this type of convergence argument to the tridiagonal model for the β-Hermite ensemble. To start we first need to determine which portions of the matrix contribute to the behavior of the largest eigenvalue. Recall the Dumitriu-Edelman matrix model A_n for the β-Hermite ensemble defined in equation (1.3.1). Take $u = c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + \cdots c_n\mathbf{e}_n$ where $\mathbf{e}_k$ is the kth coordinate vector, and observe that if we assume the c_k vary smoothly we have

$$A_n u = \frac{1}{\sqrt{\beta}} \sum_{k=1}^{n} (c_{k-1}a_k + c_{k+1}a_k + c_k b_k)\mathbf{e}_k \approx \sum_{k=1}^{n} 2c_k\sqrt{n-k}\,\mathbf{e}_k.$$

We are interested in which eigenvectors u give us $A_n u = (2\sqrt{n} + o(1))u$. This calculation suggests thar these eigenvectors should be concentrated on the first $k = o(n)$ coordinates. This suggests that the top corner of the matrix determines the behavior of the top eigenvalue.

Returning to the β-Hermite case, by Exercise 2.2.4, for $k \ll n$ we have

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

We can use this expansion to break the matrix $m^\gamma(2\sqrt{n}I - A_n)$ into terms:

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

$$+\frac{m^\gamma}{\sqrt{\beta}}\begin{bmatrix} N_1 & \tilde{N}_1 & & \\ \tilde{N}_1 & N_2 & \tilde{N}_2 & \\ & \tilde{N}_2 & N_3 & \ddots \\ & & \ddots & \ddots \end{bmatrix} \tag{4.2.1}$$

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 $2\alpha = \alpha\gamma + 1/2$. A similar analysis can be done on the second term. This term should 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

$$H_n = n^{1/6}(2\sqrt{n}I - A_n)$$

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 $0, 2, 4, 6, 8....$ on the diagonal (scaled the same way).

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

$$H_n = n^{1/6}(2\sqrt{n} - A_n) \approx -\partial_x^2 + x + \frac{2}{\sqrt{\beta}}b'_x = \mathrm{SAO}_\beta, \tag{4.2.3}$$

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 [13], and the proof to Ramírez, Rider, and Virág [29].

There are two problems at this point that must be overcome in order to make this convergence rigorous. The first is that we need 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 be.

Remarks on operator convergence

(1) Embed $\mathbb{R}^n$ into $L^2(\mathbb{R})$ via

$$\mathbf{e}_i \mapsto \sqrt{m}\mathbf{1}_{[\frac{i-1}{m},\frac{i}{m})}.$$

This gives an embedding of the matrix A_n 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.

We will not do this in these notes, but the sense of convergence that can be shown is

$$\|H_n^{-1} - SAO_\beta^{-1}\|_{2\to 2} \to 0.$$

This is called norm resolvent convergence, and it implies convergence of eigenvalues and eigenvectors if the limit has discrete simple spectrum. See e.g. Chapter 7 [31].

(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 4.2.4. 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 $v_f = [f(\frac{1}{m_n}), f(\frac{2}{m_n}), ..., f(\frac{n}{m_n})]^t$.

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 4.2.5. 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. What is k_n and what does the limit converge to?

Exercise 4.2.6. 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 give 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_β. 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. 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

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

where the infimum is over subspaces B.

For two operators $A \leqslant B$ if for all $f \in L^*$

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

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 average 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 will 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,

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

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 $-b'$. □

The above Lemma implies that the eigenvalues of Stochastic Airy should behave asymptotically the same as those of the Airy operator with the same boundary condition. From the discussion at the start of Section 4.3 we will get the following asymptotic result.

Corollary 4.3.8. *The eigenvalues of SAO_β satisfy*

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

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. □

One way to view the above Corollary is through the empirical distribution of the eigenvalues as $k \to \infty$. In this view 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

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

$$+\,\mathrm{diag}(N_{n,1}, N_{n,2}, \ldots).$$

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 by $\tilde{\mathbb{R}}^n$ 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 $\langle f_n, H_n f_n\rangle \to \langle f, Hf\rangle$ where $H = \mathrm{SAO}_\beta$.

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

Exercise 4.4.5. Show that if $\tilde{f}_n \to f$ uniformly on compact subsets with f_n differentiable and $f' \in L^2$, then

$$\liminf_{n\to\infty} \|f_n'\| \geqslant \|f'\|.$$

Exercise 4.4.6. Recall that b_n is defined to be the image of b under the embedding defined above. Show that for $\tilde{b}_n$ and $\bar{b}_n$ defined as before we have that

$$\tilde{b}_n \to \tilde{b} \qquad \text{and} \qquad \bar{b}_n' \to \bar{b}'$$

converge uniformly on compact subsets.

Proposition 4.4.7. *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*

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

Proof. Let

$$\tilde{f}_n(x) = \int_0^x \Delta_n f_n(s)ds.$$

Exercise 4.4.8. Show that $\tilde{f}_n - f_n \to 0$ uniformly on compact subsets.

Note that by Cauchy-Schwarz

$$\left|\tilde{f}_n(t+s) - \tilde{f}_n(t)\right| = \left|\int_t^{t+s} \Delta f_n(x)dx\right| \leqslant \sqrt{s}\|\Delta f_n\|, \qquad \text{with} \qquad f_n(0) = 0.$$

Therefore the $\tilde{f}_n$ form an equicontinuous family of functions and an application of the Arzela-Ascoli theorem gives us that there exists a continuous f and subsequence such that $\tilde{f}_n \to f$ uniformly on compacts. Moreover we have that $\tilde{f}_n'$ are in L^2 and so there exists $g \in L^2$ and a further subsequence along which $\tilde{f}_n' \to g$ weakly in L^2. This follows from the fact that the balls are weak*-compact. By testing against indicators of intervals we can show that we must have $f' = g$.

From the previous exercise we get the same convergence statements for the f_n and Δf_n.

Recalling the definition of the bilinear form we need to prove several different convergence statements. First observe that

$$\liminf_{n\to\infty} \langle f_n, x_n f_n\rangle \geqslant \langle f, xf\rangle.$$

This follows directly from the positivity of the integrand and Fatou's lemma. That second term $\liminf_{n\to\infty} \|\Delta f_n\|_n \geqslant \|f'\|_2$ follows from Exercise 4.4.5. For the final two terms involving $\tilde{b}$ and $\bar{b}'$ we will need to make use of the L^* bounds to cut off the integral at some large number K.

We will first consider the term $\int f_n^2 \bar{b}_n' dx$. For K large enough we have that

$$\left|\int_0^\infty f_n^2 \bar{b}_n' dx - \int_0^K f_n^2 \bar{b}_n' dx\right| \leqslant \int_K^\infty f_n^2 (C + \sqrt{x}) dx \leqslant \int_K^\infty f_n^2 x dx \leqslant \frac{C + \sqrt{K}}{K} \|f_n\|_*.$$

This error may be made arbitrarily small, therefore it will be enough to show the necessary inequality on compact subsets of $\mathbb{R}_+$. This we do by observing that $f_n \to f$ and $\bar{b}_n' \to \bar{b}'$ uniformly on compacts. The dominated convergence theorem implies convergence of the integrals. The following exercise completes the proof. □

Exercise 4.4.9. Prove that $\langle f_n, \tilde{b}_n \Delta f_n \rangle \to \langle f, \tilde{b} f' \rangle$. Use the same method of cutting off the integral at large K and use convergence on compact subsets.
Hint: $2\int_K^\infty |fg| ds \leqslant \varepsilon \|f\|_2^2 + \frac{1}{\varepsilon}\|g\|_2^2$.

Proposition 4.4.10. $\liminf \lambda_{n,1} \geqslant \lambda_1$.

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 4.4.7 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 4.4.11. 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

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

In the case $\beta = 1, 2,$ and 4 this is consistent with the classical definition. In these cases the soft edge or Airy process may be characterized as a determinantal or Pfaffian process. Tracy and Widom express the law of the lowest eigenvalue in terms of a Painlevé transcendent [32].

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

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

see [29]. Here we show that the left tail satisfies the following.

Theorem 4.5.2 ([29]).

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

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

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

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

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

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 replace the function at the ends of its support by linear pieces:

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

We can check that

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

Using these values in equation (4.5.3) gives 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 $f \in L^2$, 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 Schrödinger 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.

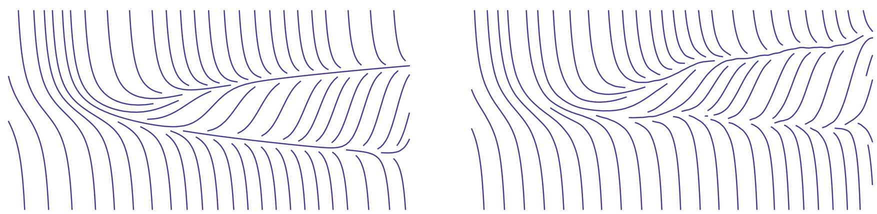

FIGURE 4.5.5. Drift trajectories for p and the random ODE $p - B$

Now replace $V(x) = x$ by $V(x) = x + \frac{2}{\sqrt{\beta}} b'$. The solution of the Ricatti equation is now an Itô diffusion given by

$$dp(x) = (x + \lambda - p(x)^2)dx + \frac{2}{\sqrt{\beta}} db_x, \qquad p(0) = \infty. \tag{4.5.6}$$

In this case there is some positive chance of the diffusion moving against the drift, including crossing the parabola. Drift trajectories for this slope field and an example of the random slope field for the ODE satisfied by $p - B$ are given in Figure 4.5.5. 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

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

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

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

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

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

The second probability is a positive 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 replaced by a standard Brownian motion (started at 1). The Radon-Nikodym derivative of this change of measure may be computed explicitly. We compute

$$E_{-a,1}\left[\mathbf{1}(p_x \in [0,2], x \in (-a,0))\right]$$
$$= E_{-a,1}\left[\exp\left(\frac{\beta}{4}\int_{-a}^{0}(x-b^2)db - \frac{\beta}{8}\int_{-a}^{0}(x-b^2)^2dx\right)\mathbf{1}(b_x \in [0,2], x \in (-a,0))\right].$$

Notice that when b stays in $[0,2]$, the density term can be controlled

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

while the probability of staying in $[0,2]$ is only exponentially small in a. This gives us the desired lower bound. □

5. Related Results

This section will give a brief partial survey of other work that makes use of the tridiagonal matrix models and operator convergence techniques that were introduced in these notes. We will discuss two other local limits that appear in the bulk and the hard-edge of a random matrix model. We will also briefly review results that can be obtained about the limiting processes, connections to sum laws and large deviations, connections to Painelevé, and an alternate viewpoint for operator convergence.

5.1. The Bulk Limit In Section 4 we proved a limit result about the local behavior of the β-Hermite ensemble at the edge of the spectrum. A similar result can be obtained for the local behavior of the spectrum near a point $a\sqrt{n}$ where $|a| < 2$.

The limiting process is the spectrum of the self-adjoint random differential operator $\mathtt{Sine}_\beta$ given by

$$f \mapsto 2R_t^{-1}\begin{pmatrix} 0 & -\frac{d}{dt} \\ \frac{d}{dt} & 0 \end{pmatrix} f, \qquad f:[0,1)\to\mathbb{R}^2, \tag{5.1.1}$$

where R_t is the positive definite matrix representation of hyperbolic Brownian motion with variance $4/\beta$ in logarithmic time. This operator is associated with a canonical system, see de Branges, [11]. It provides a link between the Montgomery-Dyson conjecture about the Sine_2 process and the non-trivial zeros of the Riemann zeta function, the Hilbert-Pólya conjecture and de Brange's attempt to prove the Riemann hypothesis, see [34].

To be more specific, we have

$$R_t = \frac{1}{2y}X_{s(t)}^{t}X_{s(t)}, \qquad s(t) = -\log(1-t). \tag{5.1.2}$$

where X satisfies the SDE

$$dX = \begin{pmatrix} 0 & dB_1 \\ 0 & dB_2 \end{pmatrix} X, \qquad X_0 = I,$$

and B_1, B_2 are two indepenent copies of Brownian motion with variance $4/\beta$. The boundary conditions are $f(0)\|(1,0)$ and when $\beta > 2$ also $f(1_-)\|X_\infty^{-1}(0,1)^t$. The ratio of entries in $X_t^{-1}(0,1)^t$ performs a hyperbolic Brownian motion in the Poincare half plane representation, see [35].

Theorem 5.1.3 ([34],[35])**.** *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 the point process of eigenvalues of the* $\mathtt{Sine}_\beta$ *operator.*

Remark 5.1.4. Local limit theorems including the bulk limit given in Theorem 5.1.3 were originally proved for $\beta = 1$ and 2 and stated using the integrable structure of the GOE and GUE. The GUE eigenvalues 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 circular ensembles with $\beta > 0$ is proved in [24].

The original description of the bulk limit process for the β-Hermite ensemble was through a process called the Browning Carousel first introduced by Valkó and Virág in [34]. The limiting process introduced there could also be described in terms of a system of coupled stochastic differential equations which gave the counting function of the process. In particular let α_λ satisfy

(5.1.5)
$$d\alpha_\lambda = \lambda \frac{\beta}{4} e^{-\frac{\beta}{4}t} dt + \mathrm{Re}\left[(e^{-i\alpha_\lambda} - 1)dZ\right],$$

where $Z_t = X_t + iY_t$ with X and Y standard Brownian motions and $\alpha_\lambda(0) = 0$. The α_λ are coupled through the noise term. Define $N_\beta(\lambda) = \frac{1}{2\pi}\lim_{t\to\infty}\alpha_\lambda(t)$, then $N_\beta(\lambda)$ is the counting function for Sine_β. This characterization is the one used to prove all of the results about Sine_β presented in Section 5.4.

The circular unitary β-ensemble is a distribution on $\mathbb{C}^n$ with joint density proportional to

$$\prod_{i<j} |\lambda_i - \lambda_j|^\beta$$

with respect to length measure on the unit circle. The local convergence to the Sine_β process was first shown by Killip and Stoiciu [24]. Using the Killip-Nenciu [22] representation, Valkó-Virág [35] show that the opeator $\mathtt{Circ}_{n,\beta}$ given by (5.1.1) with hyperbolic Brownian motion replaced by a certain hyperbolic random walk, has eigenvalues that are liftings of these λ_i to the universal cover $\mathbb{R}$. So the convergence of random matrices reduces to convergence of random walks!

In fact, the inverses of the finite-n and limiting operators can be coupled so that they are close in Hilbert-Schmidt norm.

Theorem 5.1.6 ([36]). *There exists a coupling so that for all large* n *we have*

$$\|\mathtt{Circ}_{\beta,n}^{-1} - \mathtt{Sine}_{\beta}^{-1}\|_{HS} \leqslant \frac{\log^6 n}{n}.$$

Also, if $\cdots < \lambda_{n,-1} < \lambda_{n,0} < 0 < \lambda_{n,1} < \cdots$ *are the eigenvalues of* $\mathtt{Circ}_{\beta,n}$ *and* λ_k *are the points of* Sine_β *ordered in the same way then for all large* n *we have*

$$\sum_{k\in\mathbb{Z}} \left(\frac{1}{\lambda_k} - \frac{1}{\lambda_{n,k}}\right)^2 \leqslant \frac{\log^6 n}{n}.$$

Moreover as $n \to \infty$ *we have a.s.*

$$\max_{|k|\leqslant \frac{n^{1/4}}{\log^2 n}} |\lambda_k - \lambda_{n,k}| \to 0.$$

It is the strongest coupling known so far, and it's open whether one can do better than the exponent 1/4.

5.2. The Hard–edge Limit There is another exciting local behavior that has a general $\beta > 0$ limit process description. This process does not appear as a limit of the β-Hermite ensemble, but does for the related β-Laguerre ensemble. Consider a rectangular matrix $n \times p$ matrix X_n with $p \geqslant n$ and $x_{i,j} \sim \mathcal{N}(0,1)$ all independent. The matrix

$$M_n = X_n X_n^t$$

is a symmetric matrix which may be thought of as a sample covariance matrix for a population with independent normally distributed traits. As in the case of the Gaussian ensembles we could have started with complex entries and looked instead at XX^* to form a Hermitian matrix. The eigenvalues of this matrix have distribution

$$f_{L,\beta}(\lambda_1, ..., \lambda_n) = \frac{1}{Z_{\beta,n,p}} \prod_{i=1}^{n} \lambda_i^{\frac{\beta}{2}(p-n+1)-1} e^{-\frac{\beta}{2}\lambda_i} \prod_{j<k} |\lambda_j - \lambda_k|^\beta, \tag{5.2.1}$$

with $\beta = 1$. This generalizes to the β-Laguerre ensemble which is a set of points with density $f_{L,\beta}$ for any $\beta > 0$. The matrix model M_n is part of a wider class of random matrix models called Wishart matrices. This class of models was originally introduced by Wishart in the 1920's.

As in the case of the Gaussian ensembles there is a limiting spectral measure when the eigenvalues are in the correct scale.

Theorem 5.2.2 (Marchenko-Pastur law). *Let* $\lambda_1, ..., \lambda_n$ *have β-Laguerre distribution,*

$$\nu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i/n},$$

and suppose that $\frac{n}{p} \to \gamma \in (0,1]$. *Then as* $n \to \infty$

(5.2.3)

$$\nu_n \Rightarrow \sigma_{mp}, \qquad \textit{where} \qquad \frac{d\sigma_{mp}}{dx} = \rho_{mp}(x) = \frac{\sqrt{(\gamma_+ - x)(x-\gamma_-)}}{2\pi\gamma x}\mathbf{1}_{[\gamma_-,\gamma_+]},$$

and $\gamma_\pm = (1 \pm \sqrt{\gamma})^2$.

Notice that this density can display different behavior at the lower end point depending on the value of γ. For any $\gamma < 1$ we get that the lower edge has the same $\sqrt{x}$ type behavior that we see at the edge of the semi-circle distribution. In this case the local limit is again the Airy_β process discussed in Section 4. We get something different if $\gamma = 1$. This gives us $\gamma_- = 0$ and the density simplifies to

$$\rho_{mp}(x) = \frac{1}{2\pi}\sqrt{\frac{4-x}{x}}\mathbf{1}_{[0,4]}.$$

In this case the lower edge has an asymptote at 0. In the case where $p - n \to \infty$ this is conjectured to still produce soft-edge behavior and there are limited results in this direction. In the case where $p - n \to a$ as $n \to \infty$ we obtain a different edge process at the lower edge called a hard-edge process. The name derives from the fact that the process occurs when the spectrum of a random matrix is forced against some hard constraint. Recalling the full matrix models for $\beta = 1, 2$ we observe that the matrices are positive definite. This gives a hard lower constraint of 0 for the eigenvalues. If p is close to n than this hard constraint on the lower edge will be felt and so result in different local behavior.

We begin by defining the positive random differential operator

$$\mathfrak{G}_{\beta,a} = -e^{(a+1)x + \frac{2}{\sqrt{\beta}}b(x)}\frac{d}{dx}\left[e^{-ax - \frac{2}{\sqrt{\beta}}b(x)}\frac{d}{dx}\right], \tag{5.2.4}$$

where $b(x)$ is a standard Brownian motion.

Theorem 5.2.5 (Ramírez, Rider, [28])**.** *Let* $0 < \lambda_1 < \lambda_2 < \cdots$ *have* β*−Laguerre distribution with* $p - n = a$ *and let* $\Lambda_1(a) < \Lambda_2(a) < \cdots$ *be the eigenvalues of the Stochastic Bessel Operator* $\mathfrak{G}_{\beta,a}$ *on the positive half-line with Dirichlet boundary conditions, then*

$$\{n\lambda_1, n\lambda_2, ..., n\lambda_k\} \Rightarrow \{\Lambda_1(a), \Lambda_2(a), ..., \Lambda_k(a)\}$$

(jointly in law) for any fixed $k < \infty$ *as* $n \to \infty$.

Remark 5.2.6. This result was originally conjectured with a different formulation by Edelman and Sutton [13] using intuition similar to the method of proof used for the soft edge. The actual result is proved instead by working with the inverses and a natural embedding of matrices as integral operators with piece-wise constant kernels.

5.3. Universality of local processes Recall the definition of β-ensembles introduced in (1.3.4) with the general potential function $V(x)$. The three local processes that we have discussed capture the local behavior for a wide range of these models. In particular for β-ensembles where the limiting spectral density has a single measure of support and is non-vanishing in the interval as long as $V(x)$ grows fast enough it can be proved that these are the correct limit processes. This was showed first for the bulk process by Bourgade, Erdős, and Yau [8]. For the soft edge this was showed by two groups with slightly different conditions on V and

β. Bourgade, Erdős, and Yau use analytical techniques involving Stieltjes transforms [9], while Krishnapur, Rider, and Virág give a proof that makes use of the operator convergence structure studied in these notes [25]. Finally a universality result for the hard edge was shown again using operator methods related to those introduced in these notes by Rider and Waters [30]

5.4. Properties of the limit processes For the Stochastic Airy Operator (4.2.3) we saw that it was useful to have a family of stochastic differential equations that characterizes the point process. The SDEs for SAO_β came from considering the Ricatti equation. We can build a similar family of diffusions for the Stochastic Bessel Operator introduced in (5.2.4). There is also a description for the counting function of the bulk process in terms of SDEs which was given in (5.1.5). These characterizations are be used to prove the results introduced in this section.

We will begin by discussing result for the Sine_β process. The first two results are asymptotic results for the number of points in a large interval $[0, \lambda]$. Let $N_\beta(\lambda)$ denote the number of points of Sine_β in $[0, \lambda]$. By looking at the integrated expression of α_λ we can check that $EN_\beta(\lambda) = \frac{\lambda}{2\pi}$ we consider fluctuation around the mean.

Theorem 5.4.1 (Kritchevski, Valkó, Virág [26])**.** *As $\lambda \to \infty$ we have*

$$\frac{1}{\sqrt{\log \lambda}}\left(N_\beta(\lambda) - \frac{\lambda}{2\pi}\right) \Rightarrow \mathcal{N}(0, \frac{2}{\beta\pi^2})$$

This result describes the distribution of the fluctuations on the scale of $\sqrt{\log \lambda}$ there are other regimes. In particular for fluctuation on the order of $c\lambda$ we have the following.

Theorem 5.4.2 (Holcomb, Valkó [19])**.** *The rescaled counting function $N_\beta(\lambda)/\lambda$ satisfies a large deviation principle with scale λ^2 and a good rate function $\beta I_{\mathrm{Sine}_\beta}(\rho)$ which can be written in terms of elliptic integrals.*

Roughly speaking, this means for large λ

$$P(N_\beta(\lambda) \sim \rho\lambda) \sim e^{-\lambda^2 I_{\mathrm{Sine}_\beta}(\rho)}.$$

Remark 5.4.3. Results similar to Theorems 5.4.1 and 5.4.2 may be shown for the hard edge process. The key observation is that there is an SDE description for the counting function that may be treated using mostly the same techniques as those used for the α_λ diffusion that characterizes the Sine_β process [17].

The next result give the asymptotic probability of having a large number of points in a small interval.

Theorem 5.4.4 (Holcomb, Valkó [20])**.** *Fix $\lambda_0 > 0$, then there exists c depending only on β and λ_0 such that for any $n \geqslant 1$ and $0 < \lambda \leqslant \lambda_0$ we have*

$$P(N_\beta(\lambda) \geqslant n) \leqslant e^{-\frac{\beta}{2}n^2\log(\frac{n}{\lambda}) + cn\log(n+1)\log(\frac{n}{\lambda}) + cn^2}. \tag{5.4.5}$$

Moreover, there exists an $n_0 \geqslant 1$ *so that for any* $n \geqslant n_0$, $0 < \lambda \leqslant \lambda_0$ *we also have*

$$P(N_\beta(\lambda) = n) \geqslant e^{-\frac{\beta}{2}n^2\log(\frac{n}{\lambda}) - cn\log(n+1)\log(\frac{n}{\lambda}) - cn^2}. \tag{5.4.6}$$

The previous three results focused on the number of points in a single interval. In this situation we have the advantage that the α_λ diffusion satisfies a simplified SDE

$$d\alpha_\lambda = \lambda\frac{\beta}{4}e^{-\frac{\beta}{4}t}dt + 2\sin\left(\frac{\alpha_\lambda}{2}\right)dB_t^{(\lambda)},$$

where the Brownian motion that appears $B^{(\lambda)}$ depends on the choice of parameter. The next two results require information multiple values of λ and so this simplification cannot be used. The first is a result on the maximum deviation of the counting function. This is closely related to questions on the maximum of $\operatorname{Im}\log\Phi_n(x)$ where $\Phi_n(x)$ is the characteristic polynomial of the $n \times n$ tridiagonal model.

Theorem 5.4.7 (Holcomb, Paquette [18])**.**

$$\frac{\max_{0\leqslant\lambda\leqslant x}[N_\beta(\lambda) + N_\beta(-\lambda) - \frac{\lambda}{\pi}]}{\log x} \xrightarrow[x\to\infty]{P} \frac{2}{\sqrt{\beta}\pi}.$$

The next result is a type of rigidity for the Sine_β point process.

Definition 5.4.8. A point process X on a complete separable metric space E is rigid if and only if for all bounded Borel subsets B of E, the number of points $X(B)$ in B is measurable with respect to the σ-algebra $\Sigma_{E\backslash B}$. Here $\Sigma_{E\backslash B}$ is the σ-algebra generated by all of the random variables $X(A)$ with $A \subset E\backslash B$.

A way of thinking about this is that if we have complete information about a point process X outside of a set B, then this determines the number of points in B. Notice that in a finite point process with n points this notion of rigidity follows immediately since we must have $X(B) + X(E\backslash B) = n$.

Theorem 5.4.9 (Chhaibi, Najnudel [10])**.** *The* Sine_β *point process is rigid in the sense of Definition 5.4.8.*

5.5. Spiked matrix models and more on the BBP transition Recall that is Section 3 we studied the impact of a rank one perturbation on the top eigenvalue of the GOE. In that case we studied the case where the perturbation was strong enough to be seen at the scale of the empirical spectral density. That is that the location of the top eigenvalue when scaled down by $\sqrt{n}$ depends on the strength of the perturbation. These types of results may be refined further to consider the impact of such a perturbation at the level of the local interactions. As in Section 3 we may look at two types of perturbations.

(1) For additive ensembles we study perturbations of the form $\mathrm{GOE}_n + \frac{a}{\sqrt{n}}11^t$. Here GOE_n is the $n \times n$ full matrix model, and 1 is the all-1 vector, with 11^t is the all-1 matrix.

(2) For multiplicative type ensembles we take $X_{n\times m}$ be an $n \times m(n)$ matrix with $n < m(n)$ and independent $\mathcal{N}(0,1)$ entries, then we study $X\,\mathrm{diag}(1+a^2, 1, 1, ..., 1)X^t$.

Here we will focus on the additive case. In this case it can be shown that if $\mathbf{T}$ is the tridiagonal matrix obtained by tridiagonalizing a GOE, then the corresponding tridiagonal model will have the form $\mathbf{T} + (a\beta n + aY)\mathbf{e}_1\mathbf{e}_1^t$ for some random variable Y with $EY = 0$ and EY^2 bounded (here $\mathbf{e}_1\mathbf{e}_1^t$ is the $n \times n$ matrix with a 1 in the top left corner and 0 everywhere else). In the analogue to the soft edge limit if we have $n^{1/3}(1-a) \to w \in (-\infty, \infty]$ then the top eigenvalues will converge to the eigenvalues of the Stochastic Airy operator, but with a modified boundary condition.

Exercise 5.5.1. Let

$$T = m_n^2 \begin{bmatrix} 1+\frac{w}{m_n} & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & \ddots & \\ & & \ddots & \ddots & -1 \\ & & & -1 & \ddots \end{bmatrix}.$$

Show that $T v_f \to -\partial_x^2$ for $v_f = [f(0), f(\frac{1}{m_n}), f(\frac{2}{m_n}), ..., f(\frac{n}{m_n})]$ with an appropriate boundary condition for f. Determine what the boundary condition should be.

We denote by $\mathcal{H}_{\beta,w}$ the Stochastic Airy Operator defined in equation (4.2.3), with boundary condition $f'(0) = wf(0)$, a Neumann or Robin condition with the $w = \infty$ case corresponding to the Dirichlet condition of the original soft edge process $f(0) = 0$.

Theorem 5.5.2 (Bloemendal, Virág, [5]). *Let* $a_n \in \mathbb{R}$, *and* $G_n \sim \mathrm{GOE}_n + \frac{a_n}{\sqrt{n}} 11^t$, *and suppose that* $n^{1/3}(1-a_n) \to w \in (-\infty, \infty]$ *and* $n \to \infty$. *Let* $\lambda_1 > \lambda_2 > \cdots > \lambda_n$ *be the ordered eigenvalues of* G_n. *Then jointly for* $k = 1, 2, ...$ *in the sense of finite dimensional distributions we have*

$$n^{1/6}(2\sqrt{n} - \lambda_k) \Rightarrow \Lambda_k \qquad \text{as } n \to \infty,$$

where $\Lambda_1 < \Lambda_2 < \cdots$ *are the eigenvalues of* $\mathcal{H}_{1,w}$.

As in the case of the eigenvalue problem for the stochastic Airy operator that was originally studied (with boundary condition $f(0) = 0$) we may study the Ricatti process introduced in (4.5.6). The new boundary condition for $\mathcal{H}_{\beta,w}$ leads us to the same diffusion with a different boundary condition

$$dp_\lambda = (x + \lambda - p^2)dx + \frac{d}{\sqrt{\beta}} db_x, \qquad p(0) = w.$$

Relationships between the law of the perturbed eigenvalues and the original Tracy-Widom distributions may be studied using the space-time generator for this SDE.

The generator gives a boudnary value problem representation for the Tracy-Widom$_\beta$ distribution. This can be solved explicitly for $\beta = 2, 4$. This gives fast derivations of the famous Painlevé representation for the Tracy-Widom distributions without the use of determinantal formulas. See [5,6] for further results and details. It is not known how to deduce Painlevé formulas for the Sine$_\beta$ process directly, even for $\beta = 2$.

5.6. Sum rules via large deviations Sum rules are a family of relationships used and studied in the field of orthogonal polynomials that give a relationship between a functional on a subset of probability measures and the recurrence (or Jacobi) coefficients of the orthogonal polynomials. It was recently recognized by Gamboa, Nagel, and Rouault that these relationships can be obtained using large deviation theory for random matrices [15]. This is a beautiful example of the power of large deviation theory as well as a demonstration of the relationship between the Jacobi data and spectral data of an operator. Here we will only state the result for the semicircle distribution, but the methods have been used for a wider range of models including the Marchenko-Pastur law and matrix valued measures. The proof of the theorem for the Hermite/semicircle case is originally due to Killip and Simon [23] using different methods. The real advance here is the recognition that Large Deviations may be used to prove sum rules. These techniques may then be used on a wider range of models.

Before introducing the theorem statement we introduce the Kullback-Leibler divergence or relative entropy between two probability measures μ and ν

$$\mathcal{K}(\mu|\nu) = \begin{cases} \int_{\mathbb{R}} \log \frac{d\mu}{d\nu} d\mu & \text{if } \mu \ll \nu \\ \infty & \text{otherwise.} \end{cases} \tag{5.6.1}$$

Here $\mu \ll \nu$ means that μ is absolutely continuous with respect to ν.

Now returning to the semicircle distribution we note that the Jacobi coefficients for the semicircle measure are given by

$$a_k^{sc} = 0, \qquad b_k^{sc} = 1, \qquad \text{for } k \geqslant 1. \tag{5.6.2}$$

The corresponding orthogonal polynomials are the Chebyshev polynomials of the second kind. Now suppose that μ is a probability measure on $\mathbb{R}$ with Jacobi coefficients $\{a_k, b_k\}_{k\geqslant 1}$ and define

$$\mathcal{I}_H(\mu) = \sum_{k\geqslant 1} \frac{a_k^2}{2} + b_k - 1 - \log b_k. \tag{5.6.3}$$

Now suppose that $\text{supp}(\mu) = I \cup \{\lambda_i^-\}_{i=1}^{N-} \cup \{\lambda_i^+\}_{i=1}^{N+}$ with $I \subset [-2, 2]$, $\lambda_1^- < \lambda_2^- < \cdots < -2$ and $\lambda_1^+ > \lambda_2^+ > \cdots > 2$. and define

$$\mathcal{F}_H^+(x) = \begin{cases} \int_2^x \sqrt{t^2 - 4}\,dt & \text{if } x \geqslant 2 \\ \infty \text{ otherwise,} \end{cases} \tag{5.6.4}$$

and $\mathcal{F}_H^-(x) = \mathcal{F}_H^+(-x)$.

Theorem 5.6.5 (Killip and Simon [23], Gamboa, Nagel, and Rouault [15]). *Let* J *be a Jacobi matrix with diagonal entries* $a_1, a_2, \ldots \in \mathbb{R}$ *and off-diagonal entries* $b_1, b_2, \ldots > 0$ *satisfying* $\sup_k b_k + \sup_k |a_k| < \infty$ *and let* μ *be the associated spectral measure. Then* $\mathcal{I}_H(\mu)$ *is infinite if* $\operatorname{supp}(\mu) \neq I \cup \{\lambda_i^-\}_{i=1}^{N-} \cup \{\lambda_i^+\}_{i=1}^{N+}$ *as given above. If* μ *has the desired support structure than*

$$\mathcal{I}_H(\mu) = \mathcal{K}(\mu_{sc}|\mu) + \sum_{i=1}^{N^+} \mathcal{F}_H^+(\lambda_i^+) + \sum_{i=1}^{N^-} \mathcal{F}_H^-(\lambda_i^-)$$

where both sides may be infinite simultaneously.

This is the same $\mathcal{I}_H$ given in (5.6.3). This gives us that for measures that are "close enough" to semicircular (where $\mathcal{I}_H(\mu)$ is finite) we get that

$$\sum_{k \geqslant 1} \frac{a_k^2}{2} + b_k - 1 - \log b_k = \mathcal{K}(\mu_{sc}|\mu) + \sum_{i=1}^{N^+} \mathcal{F}_H^+(\lambda_i^+) + \sum_{i=1}^{N^-} \mathcal{F}_H^-(\lambda_i^-).$$

The important observation is that $\mathcal{I}_H(\mu)$ is a large deviation rate function for the appropriate large deviation problem. Because the spectral data and the Jacobi coefficient data can both be used to describe the asymptotic likelihood of the same event the rate functions must coincide. See more details and more sum rules in (e.g. for Marchenko-Pastur) in [15].

5.7. The Stochastic Airy semigroup The idea here will be to show convergence of the moment generating function tridiagonal matrix model to the operator $e^{-\frac{T}{2}\mathrm{SAO}_\beta}$. This work by Gorin and Shkolnikov makes use of the moment method to prove this alternate version of convergence [16]. We begin by considering the tridiagonal matrix model with the coefficients reversed

(5.7.1)

$$M_N = \frac{1}{\sqrt{\beta}} \begin{bmatrix} a_1 & b_1 & & & \\ b_1 & a_2 & b_2 & & \\ & b_2 & a_3 & \ddots & \\ & & \ddots & \ddots & b_{N-1} \\ & & & b_{N-1} & a_n \end{bmatrix}, \qquad b_k \sim \chi_{\beta k}, \quad a_k \sim \mathcal{N}(0,2).$$

We take $A \subset \mathbb{R}_{\geqslant 0}$, $T > 0$, and define $[M_{N,A}]_{i,j} = [M_N]_{i,j} \mathbf{1}(\frac{N-i+1/2}{N^{1/3}}, \frac{N-j+1/2}{N^{1/3}} \in A)$. Then we study the moments

$$\mathcal{M}(T,A,N) = \frac{1}{2}\left(\left(\frac{M_{N,A}}{2\sqrt{N}}\right)^{\lfloor TN^{2/3} \rfloor} + \left(\frac{M_{N,A}}{2\sqrt{N}}\right)^{\lfloor TN^{2/3} \rfloor - 1} \right).$$

Theorem 5.7.2 (Gorin and Shkolnikov, [16]). *There exist an almost surely symmetric non-negative trace class operator* $\mathcal{U}_A(T)$ *on* $L^2(\mathbb{R}_{\geqslant 0})$ *with* $\mathcal{U}_{\mathbb{R}_{\geqslant 0}}(T) = e^{-\frac{T}{2}\mathrm{SAO}_\beta}$ *almost surely such that*

$$\lim_{N \to \infty} \mathcal{M}(T,A,N) = \mathcal{U}_A(T), \qquad T \geqslant 0$$

in the following senses:

(1) *Weak convergence: For any locally integrable* f, g *with subexponential growth at infinity and* $\pi_N f$ *denote the appropriate projection of* f *onto step functions, and* $T \geqslant 0$ *we have*

$$\lim_{N\to\infty} (\pi_N f)^t \mathcal{M}(T, A, N)(\pi_N g) = \int_{\mathbb{R}_{\geqslant 0}} (\mathcal{U}_A(T)f)(x)g(x)dx$$

in distribution and in the sense of moments.

(2) *Convergence of traces: For any* $T \geqslant 0$ *we have*

$$\lim_{N\to\infty} \mathrm{Trace}(\mathcal{M}(T, A, N)) = \mathrm{Trace}(\mathcal{U}_A(T))$$

in distribution and in the sense of moments.

This is an alternate notion of operator convergence.

References

[1] David Aldous and Russell Lyons, *Processes on unimodular random networks*, Electron. J. Probab. **12** (2007), no. 54, 1454–1508. MR2354165 ←9

[2] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni, *An introduction to random matrices*, Cambridge Studies in Advanced Mathematics, vol. 118, Cambridge University Press, Cambridge, 2010. MR2760897 ←1

[3] Jinho Baik, Gérard Ben Arous, and Sandrine Péché, *Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices*, Ann. Probab. **33** (2005), no. 5, 1643–1697. MR2165575 ←15

[4] Itai Benjamini and Oded Schramm, *Recurrence of distributional limits of finite planar graphs*, Electron. J. Probab. **6** (2001), no. 23, 13. MR1873300 ←9

[5] Alex Bloemendal and Bálint Virág, *Limits of spiked random matrices I*, Probab. Theory Related Fields **156** (2013), no. 3-4, 795–825. MR3078286 ←15, 35, 36

[6] Alex Bloemendal and Bálint Virág, *Limits of spiked random matrices II*, Ann. Probab. **44** (2016), no. 4, 2726–2769. MR3531679 ←36

[7] Alexander Bloemendal, *Finite Rank Perturbations of Random Matrices and Their Continuum Limits*, ProQuest LLC, Ann Arbor, MI, 2011. Thesis (Ph.D.)–University of Toronto (Canada). MR3004406 ←15

[8] Paul Bourgade, László Erdős, and Horng-Tzer Yau, *Bulk universality of general* β*-ensembles with non-convex potential*, J. Math. Phys. **53** (2012), no. 9, 095221, 19. MR2905803 ←32

[9] Paul Bourgade, László Erdős, and Horng-Tzer Yau, *Edge universality of beta ensembles*, Comm. Math. Phys. **332** (2014), no. 1, 261–353. MR3253704 ←33

[10] R. Chhaibi and J. Najnudel, *Rigidity of the* Sine_β *process*, ArXiv e-prints (April 2018), available at 1804.01216. ←34

[11] Louis de Branges, *Hilbert spaces of entire functions*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1968. ←29

[12] Ioana Dumitriu and Alan Edelman, *Matrix models for beta ensembles*, J. Math. Phys. **43** (2002), no. 11, 5830–5847. MR1936554 ←6, 7

[13] Alan Edelman and Brian D. Sutton, *From random matrices to stochastic operators*, J. Stat. Phys. **127** (2007), no. 6, 1121–1165. MR2331033 ←19, 32

[14] Z. Füredi and J. Komlós, *The eigenvalues of random symmetric matrices*, Combinatorica **1** (1981), no. 3, 233–241. MR637828 ←14

[15] Fabrice Gamboa, Jan Nagel, and Alain Rouault, *Sum rules via large deviations*, J. Funct. Anal. **270** (2016), no. 2, 509–559. MR3425894 ←36, 37

[16] Vadim Gorin and Mykhaylo Shkolnikov, *Stochastic Airy semigroup through tridiagonal matrices*, Ann. Probab. **46** (2018), no. 4, 2287–2344. MR3813993 ←37

[17] Diane Holcomb, *The random matrix hard edge: rare events and a transition*, Electron. J. Probab. **23** (2018), 1–20. ←33

[18] Diane Holcomb and Elliot Paquette, *The maximum deviation of the Sine$_\beta$ counting process*, Electron. Commun. Probab. **23** (2018), 1–13. ←34
[19] Diane Holcomb and Benedek Valkó, *Large deviations for the* Sine$_\beta$ *and* Sch$_\tau$ *processes*, Probab. Theory Related Fields **163** (2015), no. 1-2, 339–378. MR3405620 ←33
[20] Diane Holcomb and Benedek Valkó, *Overcrowding asymptotics for the* sine$_{fi}$ *process*, Ann. Inst. Henri Poincaré Probab. Stat. **53** (2017), no. 3, 1181–1195. MR3689965 ←33
[21] Olav Kallenberg, *Random measures, theory and applications*, Probability Theory and Stochastic Modelling, vol. 77, Springer, Cham, 2017. MR3642325 ←3
[22] Rowan and Nenciu Killip Irina, *Matrix models for circular ensembles*, International Mathematics Research Notices **2004** (2004), no. 50, 2665–2701. ←30
[23] Rowan Killip and Barry Simon, *Sum rules for Jacobi matrices and their applications to spectral theory*, Ann. of Math. (2) **158** (2003), no. 1, 253–321. MR1999923 ←36, 37
[24] Rowan Killip and Mihai Stoiciu, *Eigenvalue statistics for CMV matrices: from Poisson to clock via random matrix ensembles*, Duke Math. J. **146** (2009), no. 3, 361–399. MR2484278 ←30
[25] Manjunath Krishnapur, Brian Rider, and Bálint Virág, *Universality of the stochastic Airy operator*, Comm. Pure Appl. Math. **69** (2016), no. 1, 145–199. MR3433632 ←7, 33
[26] Eugene Kritchevski, Benedek Valkó, and Bálint Virág, *The scaling limit of the critical one-dimensional random Schrödinger operator*, Comm. Math. Phys. **314** (2012), no. 3, 775–806. MR2964774 ←33
[27] Russell Lyons and Yuval Peres, *Probability on trees and networks*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 42, Cambridge University Press, New York, 2016. MR3616205 ←9
[28] José A. Ramí rez and Brian Rider, *Diffusion at the random matrix hard edge*, Comm. Math. Phys. **288** (2009), no. 3, 887–906. MR2504858 ←32
[29] José A. Ramí rez, Brian Rider, and Bálint Virág, *Beta ensembles, stochastic Airy spectrum, and a diffusion*, J. Amer. Math. Soc. **24** (2011), no. 4, 919–944. MR2813333 ←19, 26
[30] B. Rider and P. Waters, *Universality of the Stochastic Bessel Operator*, ArXiv e-prints (October 2016), available at 1610.01637. ←33
[31] Barry Simon, *Operator theory*, A Comprehensive Course in Analysis, Part 4, American Mathematical Society, Providence, RI, 2015. MR3364494 ←20
[32] Craig A. Tracy and Harold Widom, *Level-spacing distributions and the Airy kernel*, Comm. Math. Phys. **159** (1994), no. 1, 151–174. MR1257246 ←26
[33] Hale F. Trotter, *Eigenvalue distributions of large Hermitian matrices; Wigner's semicircle law and a theorem of Kac, Murdock, and Szegő*, Adv. in Math. **54** (1984), no. 1, 67–82. MR761763 ←5
[34] Benedek Valkó and Bálint Virág, *Continuum limits of random matrices and the Brownian carousel*, Invent. Math. **177** (2009), no. 3, 463–508. MR2534097 ←29, 30
[35] Benedek Valkó and Bálint Virág, *The* Sine$_\beta$ *operator*, Invent. Math. **209** (2017), no. 1, 275–327. MR3660310 ←30
[36] Benedek Valkó and Bálint Virág, *Operator limit of the circular* β *ensemble*, ArXiv e-prints (2017), available at 1710.06988. ←31
[37] Bálint Virág, *Operator limits of random matrices*, Proceedings of the International Congress of Mathematicians—Seoul 2014. Vol. IV, 2014, pp. 247–271. MR3727611 ←12

KTH, Lindstedtsvägen 25, 100 44 Stockholm, Sweden
Email address: holcomb@kth.se

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