These notes have not received the scrutiny of publication. They could be missing important references, etc.

Random Matrices IV

1 A lower bound on the largest eigenvalue

In the last lecture, we proved $\mathbb{E}[\lambda_{\max}(J)] \le 2 + o(1)$ for a GOE $J \in \mathbb{R}^{n\times n}$ through the Sudakov–Fernique inequality. In this lecture, we establish a matching lower bound by studying the distribution of the spectrum $\mathrm{spec}(J) = \{\lambda_1, \cdots, \lambda_n\}$.

1.1 Idea: the power method

A standard way to compute the largest eigenvalue $\lambda_{\max}$ is through the power iteration

$$\begin{cases} g_0 = g \\ g_{t+1} = \dfrac{Jg_t}{\|Jg_t\|} \end{cases}$$

which can be shown to converge to the leading eigenvector under mild conditions. This is related to the fact that direction of the vector $J^k g$ is dominated by the eigenvalue of the largest magnitude. To see this, consider a random vector $g \sim \mathcal{N}(\mathbf{0}, I)$. From the eigendecomposition $J = \sum_{i=1}^n \lambda_i f_i f_i^\top$ of J, we can write $g = \sum_{i=1}^n h_i f_i$ where $h_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$. Then we have

$$J^k g = \sum_{i=1}^n \lambda_i^k h_i f_i \,.$$

Hence, by studying $\langle g, J^k g\rangle$ we gain information about the largest eigenvelue $\lambda_{\max}$. Note that

$$\begin{aligned} \mathbb{E}_J \mathbb{E}_g[\langle g, J^k g\rangle] &= \mathbb{E}_J \mathbb{E}_g[\langle gg^\top, J^k\rangle] \\ &= \mathbb{E}_J \langle I, J^k\rangle \\ &= \mathbb{E}[\mathrm{Tr}(J^k)] \end{aligned}$$

which equals the expected value of $\sum_{i=1}^n \lambda_i^k$.

In fact, it is a fundamental result in random matrix theory that the empirical distribution of $\mathrm{spec}(J)$, called the *empirical spectral distribution* (ESD), converges (in some probability sense) to the *Wigner semicircle law* defined by

$$\mu_{\mathrm{SC}}(x) := \frac{1}{2\pi}\sqrt{4 - x^2}\,, \quad -2 \le x \le 2\,.$$

One way of proving this is through the method of moments. In connection with the point above, we prove that their kth moments match for every fixed k:

$$\frac{1}{n}\mathbb{E}[\mathrm{Tr}(J^k)] = \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^n \lambda_i^k\right] \to \mathbb{E}_{\mu_{\mathrm{SC}}}[\lambda^k] = \int \lambda^k d\mu_{\mathrm{SC}}(\lambda)\,. \tag{1}$$

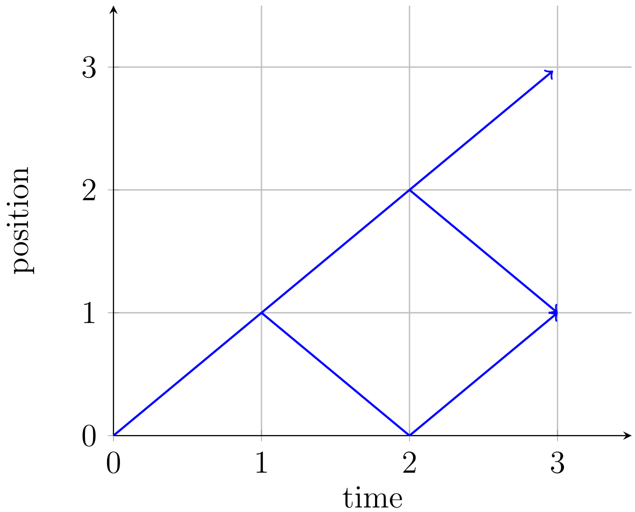


Figure 1: A diagram illustrating the computation process in (4). At time 3, two paths arrive at 1 and one path at 3, represented by $2\phi_1 + \phi_3$.

1.2 The moments of the semicircle law

We start with the RHS of (1). We make use of the *Chebyshev polynomials* ϕ_k, which is defined in terms of polynomials u_k by

$$\phi_k(x) = u_k(x/2)\,, \quad u_k(\cos\theta) = \frac{\sin((k+1)\theta)}{\sin\theta}\,.$$

Here, u_k is well defined due to the trigonometric identity

$$\sin((k+2)\theta) = 2\cos\theta\sin((k+1)\theta) - \sin(k\theta)\,.$$

This identity gives rise to a useful recurrence relation

$$\begin{cases} \phi_0(x) = 1 \\ \phi_1(x) = x \\ \phi_{k+1}(x) = x\phi_k(x) - \phi_{k-1}(x) & \text{if } k \geq 1. \end{cases}$$

The second identity can equivalently be written as

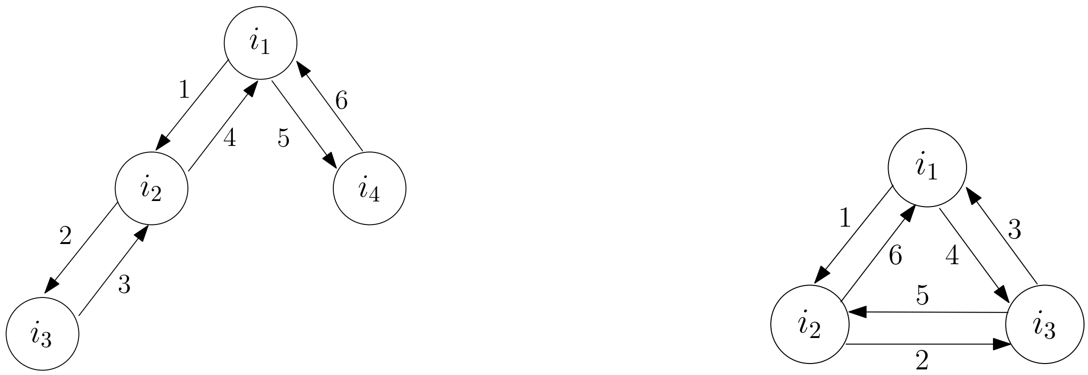
$$x\phi_k(x) = \phi_{k-1}(x) + \phi_{k+1}(x)\,. \tag{2}$$

The Chebyshev polynomials are in fact *orthogonal* under the semicircle law, in the sense that

$$\begin{aligned} \mathbb{E}_{\mu_{\mathrm{SC}}}[\phi_k(X)\phi_\ell(X)] &= \frac{1}{2\pi}\int_{-2}^{2} \phi_k(x)\phi_\ell(x)\sqrt{4-x^2}dx \\ &= \frac{2}{\pi}\int_0^\pi u_k(\cos\theta)u_\ell(\cos\theta)\sin^2\theta d\theta \\ &= \frac{2}{\pi}\int_0^\pi \sin((k+1)\theta)\sin((\ell+1)\theta)d\theta \\ &= \mathbf{1}_{k=\ell}\,. \end{aligned}$$

Since $\phi_0 \equiv 1$, this in particular implies that

$$\mathbb{E}_{\mu_{\mathrm{SC}}}[\phi_k(X)] = \mathbf{1}_{k=0}\,. \tag{3}$$



(a) A 6-walk passing through 4 vertices. The walk traverses a tree.

(b) A 6-walk passing through 3 vertices.

Figure 2: Example of two closed 6-walks using each edge twice.

Returning to our calculation of the kth moments, the trick is to repeatedly reduce the exponent using (2):

$$\begin{aligned}\mathbb{E}_{\mu_{\mathrm{SC}}}[X^k] &= \mathbb{E}_{\mu_{\mathrm{SC}}}[X^{k-1}\phi_1(X)] \\ &= \mathbb{E}_{\mu_{\mathrm{SC}}}[X^{k-2}(\phi_0(X)+\phi_2(X))] \\ &= \mathbb{E}_{\mu_{\mathrm{SC}}}[X^{k-3}(2\phi_1(X)+\phi_3(X))] \\ &= \mathbb{E}_{\mu_{\mathrm{SC}}}[X^{k-4}(2\phi_0(X)+2\phi_2(X)+\phi_4(X))] \\ &\ \ \vdots\end{aligned} \tag{4}$$

and after the kth step, due to (3), the coefficient of ϕ_0 becomes the exact value of the expectation. This corresponds to the number of k-step walks on a 1-dimensional integer lattice $\mathbb{Z}_{\geq 0}$ starting and ending at zero, which is called the *Dyck paths* in combinatorics. Figure 1 illustrates the process of computing the kth moment we just described. This gives

$$\mathbb{E}_{\mu_{\mathrm{SC}}}[X^k] = \begin{cases} 0 & \text{if } k \text{ is odd,} \\ C_m & \text{if } k = 2m \end{cases}$$

where C_m is the mth *Catalan number* defined by

$$C_m = \frac{1}{m+1}\binom{2m}{m} \tag{5}$$

which precisely counts the number of Dyck paths of length $2m$.

1.3 The moments of the empirical spectral distribution

Now we derive a formula for $\mathbb{E}[\mathrm{Tr}(J^k)]$ for a *fixed* k. By the definition of trace, this is written as

$$\mathrm{Tr}(J^k) = \sum_{i_1,\cdots,i_k=1}^{n} J_{i_1,i_2}J_{i_2,i_3}\cdots J_{i_k,i_1}\,.$$

Each of the sum is the product of the entries of J along the closed walk $i_1 \to i_2 \to \cdots \to i_k \to i_1$. Note that some of the edges in the walk might be loops, i.e., an edge $i \to i$. Since J is a GOE, all upper diagonal entries are independent mean zero Gaussian random variables. Since mean zero Gaussian has zero moments at odd orders, this yields

$$\mathbb{E}[\mathrm{Tr}(J^k)] = \mathbb{E}\left[\sum_{\substack{P\text{:closed } k\text{-walk using}\\ \text{each edge even times}}} \prod_{(i,j)\in P} J_{ij}\right] . \tag{6}$$

This already gives $\mathbb{E}[\mathrm{Tr}(J^k)] = 0$ if k is odd. Now let $k = 2m$. Since any edge must be used at least twice, there are at most m edges in the walk and thus the walk passes through at most $m + 1$ vertices (see Figure 2a for an example $k = 6$). A crucial observation here that further simplifies (6) is that the contribution of walks passing through at most m vertices is negligible. As a toy example, we can compute the contribution of the closed 6-walks depicted in Figure 2b as

$$\begin{aligned}\mathbb{E}\left[\sum_{i_1,i_2,i_3} J^2_{i_1,i_2} J^2_{i_2,i_3} J^2_{i_3,i_1}\right] &= \sum_{i_1,i_2,i_3} \frac{1}{n^3}\\ &= \frac{n(n-1)(n-2)}{n^3}\\ &= O(1)\end{aligned}$$

which is negligible compared to that of Figure 2a:

$$\begin{aligned}\mathbb{E}\left[\sum_{i_1,i_2,i_3,i_4} J^2_{i_1,i_2} J^2_{i_2,i_3} J^2_{i_1,i_4}\right] &= \sum_{i_1,i_2,i_3,i_4} \frac{1}{n^3}\\ &= \frac{n(n-1)(n-2)(n-3)}{n^3}\\ &= (1+o(1))n\,.\end{aligned}$$

To formally see this in general, let

$$\mathcal{P}_{k,\ell} := \{P : \text{closed } k\text{-walks through } \ell \text{ vertices using each edge even times}\}\,.$$

Then we have

$$\begin{aligned}|\mathcal{P}_{k,\ell}| &= \binom{n}{\ell} \cdot \#\{\text{closed } k\text{-walks through the fixed } \ell \text{ vertices using each edge even times}\}\\ &= \binom{n}{\ell} \cdot f(k,\ell)\\ &= O_k(n^\ell)\end{aligned} \tag{7}$$

where $f(k, \ell)$ is literally a function of k and ℓ, which is constant in n. Thus, (6) becomes

$$\begin{aligned}\mathbb{E}[\mathrm{Tr}(J^k)] &= \mathbb{E}\left[\sum_{\ell=1}^{m+1} \sum_{P\in\mathcal{P}_{k,\ell}} \prod_{(i,j)\in P} J_{ij}\right]\\ &= \sum_{\ell=1}^{m} \sum_{P\in\mathcal{P}_{k,\ell}} \mathbb{E}\left[\prod_{(i,j)\in P} J_{ij}\right] + \sum_{P\in\mathcal{P}_{k,m+1}} \mathbb{E}\left[\prod_{(i,j)\in P} J_{ij}\right] .\end{aligned} \tag{8}$$

We claim that the first term in the RHS is $o_k(n)$. Since P uses each edge even times, the product inside the expectation can be written in the form

$$\prod_{(i,j)\in P} J_{ij} = \prod_{i<j} J_{ij}^{2d_{ij}}\,, \qquad \sum_{i<j} d_{ij} = m$$

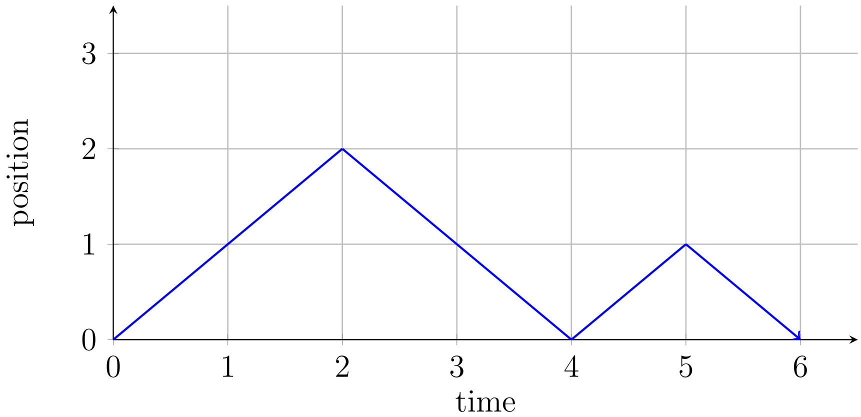


Figure 3: The Dyck path corresponding to the tree traversal in Figure 2a.

where now the terms in the product are independent. Since each term is mean zero Gaussian with variance either $1/n$ or $2/n$, we have that $\mathbb{E}[J_{ij}^{2d_{ij}}] = O_k(n^{-d_{ij}})$. Therefore,

$$\mathbb{E}\left[\prod_{(i,j)\in P} J_{ij}\right] = \prod_{i<j} O_k(n^{-d_{ij}}) = O_k(n^{-m})$$

which then by (7) gives

$$\begin{aligned}\sum_{\ell=1}^{m} \sum_{P\in\mathcal{P}_{k,\ell}} \mathbb{E}\left[\prod_{(i,j)\in P} J_{ij}\right] &= \sum_{\ell=1}^{m} |\mathcal{P}_{k,\ell}| \cdot O_k(n^{-m}) \\ &= O_k(1)\end{aligned}$$

proving the claim. Hence, the only thing left to analyze is

$$\sum_{P\in\mathcal{P}_{k,m+1}} \mathbb{E}\left[\prod_{(i,j)\in P} J_{ij}\right].$$

Since now we use exactly $m+1$ vertices and each of the m edges exactly twice, we are in fact traversing a *tree*. Thus, the product inside the expectation is simply the product of the squares of m independent $\mathcal{N}(0, 1/n)$ random variables, implying that

$$\sum_{P\in\mathcal{P}_{k,m+1}} \mathbb{E}\left[\prod_{(i,j)\in P} J_{ij}\right] = |\mathcal{P}_{k,m+1}| \cdot n^{-m}.$$

Now it remains to calculate $|\mathcal{P}_{k,m+1}|$, which can be written as

$$|\mathcal{P}_{k,m+1}| = n(n-1)\cdots(n-m) \cdot \#\{\text{trees with } m+1 \text{ vertices with fixed visiting order}\}.$$

Here, the visiting order means the order of vertices the walk passes through for the first time. For example, in Figure 2a, the visiting order is $i_1 \to i_2 \to i_3 \to i_4$. A tree with prescribed vertices visiting order is called an *ordered tree* or a *plane tree*. Now if we consider the y-axis of Figure 1 as the *depth* of the tree, one can see that there is a one-to-one correspondence between the ordered trees with $m+1$ vertices and the Dyck paths of length $k = 2m$. For instance, Figure 3 depicts the Dyck path corresponding to Figure 2a. Hence, we arrive at

$$|\mathcal{P}_{k,m+1}| = n(n-1)\cdots(n-m)C_m$$

where C_m is the Catalan number as in (5). As a result, (8) gives

$$\frac{1}{n}\mathbb{E}[\mathrm{Tr}(J^k)] = (1 + o_k(1))C_m$$

proving (1).

1.4 From the method of moments to the lower bound

Combined with the fact that the moments concentrate for all fixed k (HW), it follows that the ESD converges in probability to the semicircle law, in the sense that

$$\frac{1}{n}\sum_{i=1}^{n}\varphi(\lambda_i) \xrightarrow{p} \int \varphi(\lambda)d\mu_{\mathrm{SC}}(\lambda)$$

for any continuous φ on $[-2, 2]$. In particular, this implies that with high probability there is a nontrivial mass at $[2-\epsilon, 2]$ in the ESD for any $\epsilon > 0$, proving $\mathbb{E}[\lambda_{\max}(J)] \geq 2 - o(1)$.

Note that this argument does not prove the upper bound. We could try

$$\lambda_i \leq \left(\frac{1}{n}\sum_{i=1}^{n}\lambda_i^k\right)^{1/k} \cdot n^{1/k}$$

but then k would need to grow faster than $\log n$ for $n^{1/k}$ to vanish, which is not achieved from our argument.