Random Matrices I

Saturday, April 12, 2025 12:21 AM

Focus: spectrum / eigenvalue distribution

- Motivation:

1950s E. Wigner nuclear physics: spectrum of heavy atoms $\approx$ eigenvalues of random matrix

General tool to understand random operators / large disordered Hamiltonians

Statistics/CS: data analysis, hypothesis testing, social network

Number theory: zeros of Riemann zeta function = eigenvalue distribution?

- Particular models: iid matrix; Wigner matrix $X_{ij} = X_{ji}$ (symmetric), $X_{ij} = \overline{X_{ji}}$ (Hermitian); Covariance matrix $X^t X$, X $m \times n$ iid matrix

- In lec 3, we have seen: $\mathbb{P}[\|X\|_2 > C(\sqrt{n} + t)] < e^{-t^2}$, for iid matrix (using ε-net; also for Wigner matrices)

$\left(\|X\|_2 = \sup_{u \in \mathbb{R}^n} \frac{\|Xu\|_2}{\|u\|_2} = \max_{1 \le i \le n} |\lambda_i| \right)$

For Wigner: all $\lambda_i \in \mathbb{R}$

$\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n$ (order $\sqrt{n}$)

Question:
- Joint distribution, precisely? $\mathbb{P}((\lambda_1, \dots, \lambda_n) = (x_1, \dots, x_n)) = ?$ (In some ensemble, yes!)
- What does it look like, globally? $\frac{1}{n} \sum_i f(\frac{\lambda_i}{\sqrt{n}}) \to \int f \, d\mu$, for $f \in C_c$? $\left[\frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i/\sqrt{n}} \to \mu \text{ in vague topology} \right]$ (First order understanding)
- what does it look like, locally? Edge: $(\lambda_1 - c\sqrt{n})\, n^{\alpha} \to ?$ (statistics / computer science / combinatorics)
  Bulk: # eigenvalues in $[a\sqrt{n} - \frac{1}{\sqrt{n}}, a\sqrt{n} + \frac{1}{\sqrt{n}}] \to ?$ (physics / number theory)

Many of these limiting objects are first found in certain matrix models, then universal.

This week: global law.

Matrix: Wigner symmetric / Hermitian: $X_{ij} = X_{ji}$ / $X_{ij} = \overline{X_{ji}}$; X_{ij} for all $i \le j$ independent; $\mathbb{E} X_{ij} = 0$, $\operatorname{Var} X_{ij} = 1$ ($\mathbb{E}|X_{ij}|^2 = 1$) for $i < j$; $\mathbb{E} X_{ii}$, $\operatorname{Var} X_{ii} < C$

(For simplicity, think of Rademacher: $\mathbb{P}(X_{ij} = 1) = \mathbb{P}(X_{ij} = -1) = \frac{1}{2}$)

Moment Method: $\sum_{i=1}^{n} \lambda_i^k = \operatorname{tr} X^k$; want $n^{-\frac{k}{2}-1} \operatorname{tr} X^k = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\lambda_i}{\sqrt{n}}\right)^k \to \int y^k \, d\mu(y)$ for some μ

> $(X_{ij})_{i,j=1}^{\infty}$, $n \times n$ upper-left corner $X^{(n)}$
> almost sure convergence holds if $\sum_n \mathbb{P}[|n^{-\frac{3}{2}} \operatorname{tr} X^{(n)}| > \varepsilon] < \infty$ for any $\varepsilon > 0$ (Borel-Cantelli)

$\operatorname{tr} X = \sum_{i=1}^{n} X_{ii}$, mean zero, var $\sim n$, $n^{-\frac{3}{2}} \operatorname{tr} X \to 0$ as $n \to \infty$ (in probability, or even almost surely)

$\operatorname{tr} X^2 = \sum_{i,j=1}^{n} X_{ij} X_{ji} = \sum_{i,j=1}^{n} |X_{ij}|^2$ ($X_{ij} \overline{X_{ji}}$ in Hermitian), $\mathbb{E} \operatorname{tr} X^2 = n^2$, $\operatorname{Var}(\operatorname{tr} X^2) \sim n^2$, $n^{-2} \operatorname{tr} X^2 \to 1$ as $n \to \infty$

$\operatorname{tr} X^3 = \sum_{i,j,k=1}^{n} X_{ij} X_{jk} X_{ki}$, $\mathbb{E} \operatorname{tr} X^3 = \sum_{i=1}^{n} \mathbb{E} X_{ii}^3 = O(n)$, $\operatorname{Var}(\operatorname{tr} X^3) = O(n^3)$, $\left(= \sum_{i,j,k} \sum_{i',j',k'} \mathbb{E} X_{ij} X_{jk} X_{ki} X_{i'j'} X_{j'k'} X_{k'i'} \right)$ $\ne 0$ when $(i,j,k) = (i',j',k')$ or $i = j$, $i' = j'$, $k = k'$ $\Rightarrow n^{-\frac{5}{2}} \operatorname{tr} X^3 \to 0$ (Can also use concentration ineq)

$\operatorname{tr} X^4 = \sum_{i_1 i_2 i_3 i_4 = 1}^{n} X_{i_1 i_2} X_{i_2 i_3} X_{i_3 i_4} X_{i_4 i_1}$, $\mathbb{E} X_{i_1 i_2} X_{i_2 i_3} X_{i_3 i_4} X_{i_4 i_1} = 0$, if $\exists e \in \{1, \dots, n\}$ appear exactly once in $(i_1 i_2)(i_2 i_3)(i_3 i_4)(i_4 i_1)$

$\Rightarrow$ must $i_1 = i_3$, or $i_2 = i_4$, $\Rightarrow \mathbb{E} \operatorname{tr} X^4 = \sum_{i,j,k} \mathbb{E} X_{ij} X_{ji} X_{ik} X_{ki} + \sum_{i,j,k} \mathbb{E} X_{ij} X_{jk} X_{kj} X_{ji} - \sum_{i,j} \mathbb{E} X_{ij} X_{ji} X_{ij} X_{ji}$

$= 2n^3 - n^2$

$\operatorname{Var}(\operatorname{tr} X^4) = O(n^4)$; compute directly: $\mathbb{E}(\operatorname{tr} X^4)^2 = \sum_{i_1 i_2 i_3 i_4, j_1 j_2 j_3 j_4} \mathbb{E} X_{i_1 i_2} X_{i_2 i_3} X_{i_3 i_4} X_{i_4 i_1} X_{j_1 j_2} X_{j_2 j_3} X_{j_3 j_4} X_{j_4 j_1} = (\mathbb{E} \operatorname{tr} X^4)^2 + O(n^4)$

Main term: $i_1 = i_3$ or $i_2 = i_4$; $j_1 = j_3$ or $j_2 = j_4$

$\Rightarrow \frac{\operatorname{tr} X^4}{n^3} \to 2$ (in prob / a.s.)

For general k:

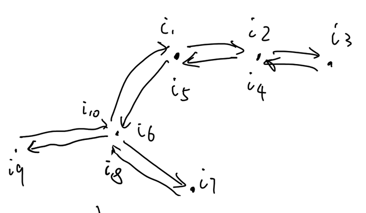
$\operatorname{tr} X^k = \sum_{i_1 \cdots i_k} X_{i_1 i_2} X_{i_2 i_3} \cdots X_{i_{k-1} i_k} X_{i_k i_1}$

$\mathbb{E} X_{i_1 i_2} \cdots X_{i_k i_1} = 0$ if $\exists e \in \{1, \dots, n\}^2$, appear exactly once in $(i_1 i_2) \cdots (i_k i_1)$

$\Rightarrow$ at most $\frac{k}{2}$ different edges among $(i_1 i_2) \cdots (i_k i_1)$;

k odd: $|\mathbb{E} \operatorname{tr} X^k| \lesssim n \cdot n^{\frac{k-1}{2}}$ (# rooted connected graphs with $\le \frac{k-1}{2}$ edges)

$n^{-\frac{k}{2}-1} |\mathbb{E} \operatorname{tr} X^k| \lesssim n^{-\frac{1}{2}} \to 0$

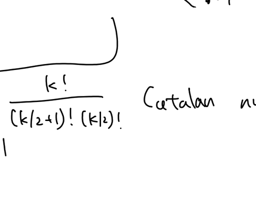


k even: $\mathbb{E} \operatorname{tr} X^k \approx$ # rooted trees with $\frac{k}{2}$ edges

Dyck word: $(())(()())$

$C_{k/2}$: # of Dyck words of length k $= \frac{k!}{(k/2+1)!\,(k/2)!}$ Catalan number

proof: up-right path from $(1,1)$ to $(\frac{k}{2}, \frac{k}{2})$ below $y = x+1$



$\#\text{path} = \binom{k}{\frac{k}{2}} - \binom{k}{\frac{k}{2}-1}$

$\binom{k}{\frac{k}{2}}$: # of path from $(1,1)$ to $(\frac{k}{2}, \frac{k}{2})$

$\binom{k}{\frac{k}{2}-1}$: # of path from $(1,1)$ to $(\frac{k}{2}, \frac{k}{2})$ that touches $y = x+1$ $=$ # of path from $(1,1)$ to $(\frac{k}{2}-1, \frac{k}{2}+1)$

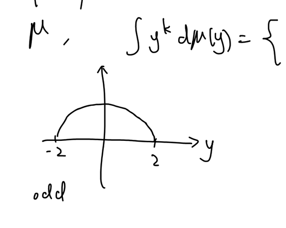
$\Rightarrow \mathbb{E} \operatorname{tr} X^k \approx C_{k/2}\, n(n-1) \cdots (n-k/2) \approx C_{k/2}\, n^{k/2+1}$

$n^{-\frac{k}{2}-1} \cdot \mathbb{E} \operatorname{tr} X^k \to C_{k/2}$ as $n \to \infty$

(upgrade to in prob / a.s. convergence: $\operatorname{Var} \operatorname{tr} X^k = o(n^k)$)

- For what measure μ, $\int y^k \, d\mu(y) = \begin{cases} 0 & k \text{ odd} \\ C_{k/2} & k \text{ even} \end{cases}$?

$\mu_{sc} = \frac{1}{2\pi} \sqrt{4 - y^2}\, dy$



Claim: $\int_{-2}^{2} y^k \, d\mu_{sc}(y) = \begin{cases} 0 & \text{odd} \\ C_{k/2} & \text{even} \end{cases}$

proof: odd: by symmetry

even: $y = 2\cos\theta$

$\int_{-2}^{2} \frac{1}{2\pi} y^k \sqrt{4-y^2}\, dy = \int_0^\pi \frac{2^{k+1}}{\pi} \cos^k\theta \sin^2\theta \, d\theta = \int_0^\pi \frac{2^{k+1}}{\pi} (\cos^k\theta - \cos^{k+2}\theta)\, d\theta$

$\int_0^\pi \cos^k\theta \, d\theta = \pi \cdot \frac{1}{2} \cdot \frac{3}{4} \cdots \frac{k-1}{k}$, by induction

(Wigner semi-circle law) Wigner symmetric/Hermitian (independent on/above diagonal; mean $=0$, var $=1$ above diagonal; bounded mean, var on diagonal)
$\lambda_1^{(n)}, \dots, \lambda_n^{(n)}$ be eigenvalues of upper-left $n \times n$ corner of infinite symmetric/Hermitian Wigner matrix

For $f \in C_c$, $\frac{1}{n} \sum_{i=1}^{n} f(\frac{\lambda_i^{(n)}}{\sqrt{n}}) \to \int f(y)\, d\mu_{sc}(y)$ in expectation / in probability / almost surely

($\frac{1}{n} \sum \delta_{\lambda_i^{(n)}/\sqrt{n}} \to \mu_{sc}$ in vague topology)

Also, since μ_{sc} is a continuous distribution

$\frac{1}{n} \#\{\lambda_i^{(n)} : \lambda_i < a\sqrt{n}\} \to \mu_{sc}((-2, a))$ in expectation / in probability / almost surely.

Note: does not provide any upper bound for operator norm $= \max\{|\lambda_1|, |\lambda_n|\}$!!! (Lower bound: $\mathbb{P}(\lambda_1 < (2-\varepsilon)\sqrt{n}) \to 0$, by semi-circle law)

(Bai-Yin) $\frac{1}{\sqrt{n}} \|X^{(n)}\|_2 \to 2$ almost surely, if X_{ij} for $i \le j$ are iid, with $\mathbb{E} X_{ij}^4 < \infty$

(can alter diagonal X_{ii} entries, to be iid with $\mathbb{E} X_{ii}^2 < \infty$)

Upper bound proof idea: upper-left $n \times n$ corner

$\|X^{(n)}\|_2^k = \max\{|\lambda_1|^k, |\lambda_n|^k\} \le \operatorname{tr}(X^{(n)})^k \le (C_{k/2} + o_k(1))\, n^{k/2+1} \le (2^k + o_k(1))\, n^{\frac{k}{2}+1}$ (may lose a factor of n)

want: $\|X^{(n)}\|_2 < (2+\varepsilon)\sqrt{n}$, take k s.t. $n^{\frac{1}{k}} < 1 + \frac{\varepsilon}{2}$; $k = C \log n$.

need be careful: error term $o_k(1)$ can be subtle; can be large without 4-th moment bound.

Stieltjes transform

Given a measure μ on $\mathbb{R}$, define its Stieltjes transform S_μ on $\mathbb{C} \setminus \operatorname{supp}(\mu)$ as

$S_\mu(z) = \int \frac{d\mu(x)}{x - z}$

Property: $\overline{S_\mu(z)} = S_\mu(\bar{z})$; $\left|\frac{d^j}{dz^j} S(z)\right| \le j! \frac{\mu(\mathbb{R})}{\operatorname{Im}(z)^{j+1}}$

Formally (or for $|z|$ large), $S_\mu(z) = -\sum_{k=0}^{\infty} \frac{\int x^k \, d\mu(x)}{z^{k+1}}$ generate moments at $z = \infty$

example.
For $\mu_{sc} = \frac{1}{2\pi} \sqrt{4 - x^2}\, dx$ supported on $[-2, 2]$

$S_{\mu_{sc}}(z) = \frac{1}{2\pi} \int_{-2}^{2} \frac{\sqrt{4-x^2}}{x - z}\, dx = \frac{-z + \sqrt{z^2 - 4}}{2}$, for $z \in \mathbb{C} \setminus [-2, 2]$

example.
For $n \times n$ matrix X (symmetric/Hermitian) with eigenvalues $\lambda_1 \ge \lambda_2 \cdots \ge \lambda_n$, $\mu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i/\sqrt{n}}$

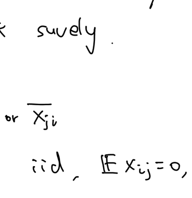
$S_{\mu_n}(z) = \frac{1}{n} \operatorname{tr}(\frac{1}{\sqrt{n}} X - z I_n)^{-1} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\lambda_i/\sqrt{n} - z}$

Why might Stieltjes transform be better than moments? Capture μ very well.

- For $z = a + ib$ $(b > 0)$

$\operatorname{Im} \frac{1}{x - z} = \frac{1}{2i} \left(\frac{1}{x - a - ib} - \frac{1}{x - a + ib} \right) = \frac{b}{(x-a)^2 + b^2} > 0$

$\Rightarrow \operatorname{Im}(S_\mu(a + ib)) = \int \frac{b \, d\mu(x)}{(x-a)^2 + b^2} = \pi (\mu * P_b)(a)$ (convolution: average at scale b at location a.) $\qquad P_b(x) = \frac{1}{\pi} \frac{b}{x^2 + b^2} = \frac{1}{b} P_1(\frac{x}{b})$



$\Rightarrow \operatorname{Im}(S_\mu(\cdot + ib))\, dx \to \pi\mu$ as $b \to 0^+$ (in vague topology)

Using these it can be shown that vague topology conv. is equiv. to Stieltjes transform conv.

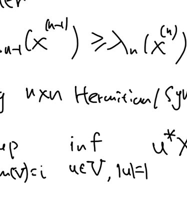
(Stieltjes continuity theorem) Let μ_n be a sequence of random prob. measures on $\mathbb{R}$ (i.e. $\mu_n(\mathbb{R}) = 1$)
Let μ be a deterministic prob. measure on $\mathbb{R}$

$\mu_n \to \mu$ in vague topology (i.e. $\int f \, d\mu_n \to \int f \, d\mu$ for $f \in C_c$) in expectation / in probability / almost surely

iff $S_{\mu_n}(z) \to S_\mu(z)$ for any $z \in \mathbb{C}_+$, in expectation / in probability / almost surely.

Next: use these to give an alternative proof of semi-circle law.

For $S_n(z) = S_{\mu_n}(z) = \frac{1}{n} \operatorname{tr}(\frac{1}{\sqrt{n}} X^{(n)} - z I_n)^{-1}$



$X_{ij} = X_{ji}$ or $\overline{X_{ji}}$

X_{ij}, $i<j$ iid, $\mathbb{E} X_{ij} = 0$, $\operatorname{Var} X_{ij} = 1$

X_{ii}, iid, $\mathbb{E} X_{ii}$, $\operatorname{Var} X_{ii} < C$

①. $S_n(z)$ concentrates around $\mathbb{E} S_n(z)$

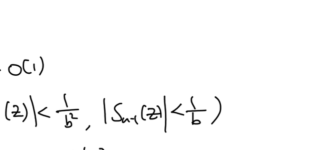
② $\mathbb{E} S_n(z) \approx -\frac{1}{z + \mathbb{E} S_n(z)}$

(Assuming these, can solve $\mathbb{E} S_n(z) \approx -\frac{z \pm \sqrt{z^2 - 4}}{2}$; since $S_n(z) \sim \frac{1}{z}$ as $|z| \to \infty$, $S_n(z) = \frac{-z + \sqrt{z^2 - 4}}{2}$)

(Does not need to know the limiting law μ_{sc} a priori !!)

For ① (Cauchy interlacing law) $X^{(n)}$ $n \times n$ Hermitian/symmetric matrix
$X^{(n-1)}$ upper-left $n-1 \times n-1$ corner

$\lambda_1(X^{(n)}) \ge \lambda_1(X^{(n-1)}) \ge \lambda_2(X^{(n)}) \ge \lambda_2(X^{(n-1)}) \ge \cdots \ge \lambda_{n-1}(X^{(n-1)}) \ge \lambda_n(X^{(n)})$



proof Use Cauchy-Fisher min-max theorem: for $X^{(n)}$ being $n \times n$ Hermitian/Symmetric

$\lambda_i(X^{(n)}) = \sup_{\dim(V) = i} \inf_{u \in V, |u| = 1} u^* X^{(n)} u$

$= \inf_{\dim(V) = n+1-i} \sup_{u \in V, |u| = 1} u^* X^{(n)} u$

Then by interlacing, for $z = a + ib \in \mathbb{C}_+$

$\left| \sum_{i=1}^{n} \frac{\lambda_i(X^{(n)})/\sqrt{n} - a + ib}{(\lambda_i(X^{(n)})/\sqrt{n} - a)^2 + b^2} - \sum_{i=1}^{n-1} \frac{\lambda_i(X^{(n-1)})/\sqrt{n} - a + ib}{(\lambda_i(X^{(n-1)})/\sqrt{n} - a)^2 + b^2} \right| = O(1)$

first sum $\to n S_n(a + ib)$; second sum $\to \sqrt{n(n-1)}\, S_{n-1}(\frac{\sqrt{n}}{\sqrt{n-1}}(a + ib)) = n S_{n-1}(a + ib) + O(1)$

(since $|\partial_z S_{n-1}(z)| < \frac{1}{b^2}$, $|S_{n-1}(z)| < \frac{1}{b}$)

$\Rightarrow S_n(z) - S_{n-1}(z) = O(\frac{1}{n})$

In words: changing the n-th row and n-th column alters $S_n(z)$ by $O(\frac{1}{n})$

By symmetry, changing any row/column alters $S_n(z)$ by $O(\frac{1}{n})$

$\Rightarrow \mathbb{P}[|S_n(z) - \mathbb{E} S_n(z)| > \beta] \le \exp(-\frac{\beta^2}{n \cdot \frac{C}{n^2}}) = \exp(-n\beta^2/C)$ (McDiarmid's ineq)

$\Rightarrow |S_n(z) - \mathbb{E} S_n(z)| = O(\frac{1}{\sqrt{n}})$

For ②: $\mathbb{E} S_n(z) = \frac{1}{n} \mathbb{E} \operatorname{tr}(X^{(n)}/\sqrt{n} - z I_n)^{-1} = \mathbb{E}\left[(X^{(n)}/\sqrt{n} - z I_n)^{-1}\right]_{nn}$ ($n \cdot n$ entry of the resolvent matrix)

$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & \\ & \end{bmatrix}$

$X^{(n)} = \begin{pmatrix} X^{(n-1)} & u \\ u^* & X_{nn} \end{pmatrix}$ $\xrightarrow{\text{linear algebra}}$ $= -\mathbb{E} \frac{1}{z - \frac{X_{nn}}{\sqrt{n}} + \frac{1}{n} u^* (X^{(n-1)}/\sqrt{n} - z I_{n-1})^{-1} u}$

Note: $\mathbb{P}[|u^* R u - \mathbb{E} u^* R u| > C\beta] \le \exp(-\beta^2/n)$ (by concentration ineq applied to $(u^* R u)^{\frac{1}{2}} = \|R^{\frac{1}{2}} u\|$; Exercise)

R: given deterministic, positive semi-def $(n-1) \times (n-1)$ matrix, with operator norm $\|R\|_2 = 1$.

$\Rightarrow \frac{1}{n} u^* (X^{(n-1)}/\sqrt{n} - z I_{n-1})^{-1} u$ concentrates around its expectation $= \mathbb{E} \operatorname{tr}(X^{(n-1)}/\sqrt{n} - z I_{n-1})^{-1} = \mathbb{E} \sqrt{n(n-1)}\, S_{n-1}(\sqrt{\frac{n}{n-1}} z) = \mathbb{E} S_{n-1}(z) + O(\frac{1}{n})$

$\Rightarrow \mathbb{E} S_n(z) = -\frac{1}{z + \mathbb{E} S_n(z)} + o(1)$

Sample covariance matrix & Marchenko-Pastur law

X: $p \times n$ matrix, iid, $\mathbb{E} X_{ij} = 0$, $\operatorname{Var} X_{ij}$ ($\mathbb{E}|X_{ij}|^2$) $= 1$ Application: data analysis, test independence

$\frac{1}{n} X X^T$ (or $\frac{1}{n} X X^*$) $p \times p$ matrix, has eigenvalues $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_p$ (of order 1)

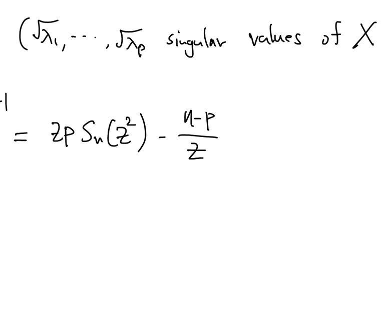
$n \to \infty$, $\frac{p}{n} \to \lambda \in (0, \infty)$

(Marchenko-Pastur) $\frac{1}{p} \sum_{i=1}^{p} \delta_{\lambda_i} \to \mu_{mp}^{\lambda}$, in vague topology in expectation | in probability | almost surely

Where (1) $0 < \lambda < 1$: $\mu_{mp}^{\lambda} = \frac{\sqrt{(\lambda_+ - x)(x - \lambda_-)}}{2\pi\lambda x} dx$, $\lambda_\pm = (1 \pm \sqrt{\lambda})^2$

(2) $\lambda = 1$: $\mu_{mp}^{\lambda} = \frac{1}{2\pi} \sqrt{\frac{4-x}{x}}\, dx$

(3) $\lambda > 1$: $\mu_{mp}^{\lambda} = \frac{1}{\lambda} \frac{\sqrt{(\lambda_+ - x)(x - \lambda_-)}}{2\pi\lambda x} dx + (1 - \frac{1}{\lambda}) \delta_0$, $\lambda_\pm = (\sqrt{\lambda} \pm 1)^2$



How to prove?

- Moments also work: $\int x^k \, d\mu_{mp}^{\lambda}(x) = \frac{1}{k} \sum_{i=0}^{k} \binom{k}{i} \binom{k}{i+1} \lambda^i$

- derive using Stieltjes transform $S_n(z) = \frac{1}{p} \sum_{i=1}^{p} \frac{1}{\lambda_i - z} = \frac{1}{p} \operatorname{tr}(X X^T/n - z)^{-1}$

Assume $n \ge p$

① Linearize: $Y = \begin{bmatrix} 0 & X^T/\sqrt{n} \\ X/\sqrt{n} & 0 \end{bmatrix}$ has eigenvalues $\pm\sqrt{\lambda_1}, \pm\sqrt{\lambda_2}, \dots, \pm\sqrt{\lambda_p}$, and $n-p$ zeros. ($\sqrt{\lambda_1}, \dots, \sqrt{\lambda_p}$ singular values of X)

$\Rightarrow (Y - z)^{-1} = \begin{bmatrix} -z I_n & X^T/\sqrt{n} \\ X/\sqrt{n} & -z I_p \end{bmatrix}^{-1} \underset{\text{(linear algebra)}}{=} \begin{bmatrix} (\frac{X^T X}{nz} - z I_n)^{-1} & \\ & \end{bmatrix}$; $\sum_{i=1}^{n} (Y - z)^{-1}_{ii} = \operatorname{tr}\left(\frac{X^T X}{nz} - z I_n\right)^{-1} = z p S_n(z^2) - \frac{n-p}{z}$

② Similar to semi-circle case:

$(Y - z)^{-1} = \begin{bmatrix} -z & 0 & u^T/\sqrt{n} \\ 0 & -z I_{n-1} & X^{(1)T}/\sqrt{n} \\ u/\sqrt{n} & X^{(1)}/\sqrt{n} & -z I_p \end{bmatrix}^{-1}$ $\qquad X = [u \mid X^{(1)}]$

via linear algebra: $(Y - z)^{-1}_{11} = \left(-z - \frac{1}{n} u^T \left(\frac{X^{(1)} X^{(1)T}}{nz} - z\right)^{-1} u\right)^{-1}$

concentrates around its mean (same as in sc)

$\frac{1}{n} \mathbb{E} u^T \left(\frac{X^{(1)} X^{(1)T}}{nz} - z\right)^{-1} u = \frac{1}{n} \mathbb{E} \operatorname{tr}\left(\frac{X^{(1)} X^{(1)T}}{nz} - z\right)^{-1} = \frac{1}{n} \mathbb{E} \operatorname{tr}\left(\frac{X X^T}{nz} - z\right)^{-1} + O(\frac{1}{n}) = \mathbb{E} \frac{pz}{n} S_n(z^2) + O(\frac{1}{n})$

Interlacing of eigenvalues in n & $n-1$ (★ this is a different interlacing: eigenvalues of A and $A + vv^T$ interlace, for symmetric A)

(as in semi-circle, interlacing also gives concentration of $S_n(z)$ around $\mathbb{E} S_n(z)$)

$\Rightarrow (Y - z)^{-1}_{11} \approx \frac{1}{-z - \frac{pz}{n} \mathbb{E} S_n(z^2)}$

also for $(Y - z)^{-1}_{ii}$, for any $1 \le i \le n$

$\Rightarrow \frac{pz}{n} \mathbb{E} S_n(z^2) - \frac{n-p}{zn} = \frac{1}{-z - \frac{pz}{n} \mathbb{E} S_n(z^2)} + o(1)$

let $w = z^2$, and send $n \to \infty$

$\Rightarrow \lambda w S(w) + \lambda - 1 = \frac{1}{-1 - \lambda S(w)}$

$\Rightarrow S(w) = \frac{-(w + \lambda - 1) \pm \sqrt{(w - (1-\sqrt{\lambda})^2)(w - (1+\sqrt{\lambda})^2)}}{2\lambda w}$

take $+$ to match $S(w) \sim -\frac{1}{w}$ for $|w| \to \infty$

check: $\frac{1}{\pi} \operatorname{Im} S(w) \to \mu_{mp}^{\lambda}(x)$ as $w \to x$ from $\mathbb{C}^+$