STAT260 Mean Field Asymptotics in Statistical Learning	Lecture 17 - 03/29/2021

Lecture 17: Random Matrices and Stieltjes Transforms

Lecturer: Song Mei *Scriber: Ryan Theisen* *Proof reader: Alexander Tsigler*

1 Motivation

Many quantities that we are interested in are of the form

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[f(\lambda_i(\boldsymbol{A}))] = \mathbb{E}\int_{\mathbb{R}} f(\lambda)\widehat{\mu}_{\boldsymbol{A}}(\mathrm{d}\lambda),$$

where $\widehat{\mu}_{\boldsymbol{A}}(\mathrm{d}\lambda) = n^{-1}\sum_{i=1}^{n}\delta_{\lambda_i(\boldsymbol{A})}$ is the empirical measure on the eigenvalues of $\boldsymbol{A}$. For example, in Lecture 16 we saw that the bias and variance of the linear model can be written as:

$$\begin{aligned}
\mathbb{B}(\lambda) &= \lambda^2\mathbb{E}\int_{\mathbb{R}}\frac{1}{(s+\lambda)^2}\widehat{\mu}_{\boldsymbol{A}}(\mathrm{d}s) \quad \text{(Bias)},\\
\mathbb{V}(\lambda) &= \sigma^2\mathbb{E}\int_{\mathbb{R}}\frac{s}{(s+\lambda)^2}\widehat{\mu}_{\boldsymbol{A}}(\mathrm{d}s) \quad \text{(Variance)},
\end{aligned}$$

where here $\boldsymbol{A} = \boldsymbol{X}^\top\boldsymbol{X}/d$, and $\boldsymbol{X}$ is the design matrix. To analyze these expressions asymptotically, we hope to understand $\widehat{\mu}_{\boldsymbol{A}}(\mathrm{d}s)$ for large n.

2 Stieltjes Transforms and Stieltjes Functions

We begin with the definition of the Stieltjes transform.

Definition 1 (Stieltjes Transform)**.** *The Stieltjes transform $\mathcal{S}$ is a map*

$$\begin{aligned}
\mathcal{S}: \mathcal{P}(\mathbb{R}) &\longrightarrow \mathcal{F}(\mathbb{C}),\\
\mu &\longmapsto m = \mathcal{S}(\mu),
\end{aligned} \tag{1}$$

where $\mu \in \mathcal{P}(\mathbb{R})$ is a probability measure and $m : \mathbb{C}\setminus supp(\mu) \to \mathbb{C}$ is given by

$$m(z) = \int \frac{1}{x-z}\mu(\mathrm{d}x). \tag{2}$$

The function m is called the Stieltjes function *of μ.*

Remark 2. *Note that for $\mu = \widehat{\mu}_{\boldsymbol{A}} = n^{-1}\sum_{i=1}^{n}\delta_{\lambda_i(\boldsymbol{A})}$ the empirical measure on the eigenvalues of $\boldsymbol{A}$, we have that the associated Stieltjes function is given by*

$$m(z) = \int\frac{1}{x-z}\widehat{\mu}_{\boldsymbol{A}}(\mathrm{d}x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_i(\boldsymbol{A})-z} = \frac{1}{n}\mathsf{tr}((\boldsymbol{A}-z\boldsymbol{I})^{-1}).$$

In particular, when $\boldsymbol{A} = \boldsymbol{X}^\top\boldsymbol{X}$ and $z = -\lambda$, then $m(z) = n^{-1}\mathsf{tr}((\boldsymbol{X}^\top\boldsymbol{X}+\lambda\boldsymbol{I})^{-1})$.

Let's see some properties of the Stieltjes function $m = \mathcal{S}(\mu)$, for μ fixed.

- For $z \in \mathbb{C}$, write $z = \lambda + i\eta$ where $\lambda, \eta \in \mathbb{R}$ are the real and imaginary parts of z; we'll use the notation $\mathrm{Re}(z) = \lambda$ and $\mathrm{Im}(z) = \eta$ to denote these two numbers. Then we have that

$$\begin{aligned} \mathrm{Im}(m(z)) &= \int \mathrm{Im}\Big(\frac{1}{x-z}\Big)\mu(\mathrm{d}x) = \int \mathrm{Im}\Big(\frac{\overline{(x-z)}}{(x-z)\overline{(x-z)}}\Big)\mu(\mathrm{d}x) \\ &= -\int \frac{\mathrm{Im}(x-z)}{(x-\lambda)^2+\eta^2}\mu(\mathrm{d}x) = \int \frac{\eta}{(x-\lambda)^2+\eta^2}\mu(\mathrm{d}x), \end{aligned} \tag{3}$$

and furthermore

$$\mathrm{Re}(m(z)) = \int \frac{x-\lambda}{(x-\lambda)^2+\eta^2}. \tag{4}$$

- Notice from eqn. 3 that $\eta > 0$ implies that $\mathrm{Im}(m(z)) > 0$, and hence we have that m maps $\mathbb{C}_+$ to $\mathbb{C}_+$, where $\mathbb{C}_+ = \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}$ is the set of complex numbers with positive imaginary part.

- Based on the expression in eqn. 3, we have the following useful interpretation for $\mathrm{Im}(m(z))$. Recall the Cauchy(λ, η) distribution has density

$$p(x) = \frac{1}{\pi}\frac{\eta}{(x-\lambda)^2+\eta^2}.$$

Then observe that

$$\frac{1}{\pi}\mathrm{Im}(m(z)) = \int p(x-z)\mu(\mathrm{d}x) = \mu \star p,$$

where $\mu \star p$ is the *convolution* of μ and p. In other words, if $X \sim \mu$ and $W \sim \mathrm{Cauchy}(\lambda, \eta)$, then $\pi^{-1}\mathrm{Im}(m(z))$ is the density of the random variable $X + W$. Intuitively, $\mathrm{Im}(m(z))$ is a "smoothed" version of the density of μ.

- For $z \in \mathbb{C} \setminus \mathrm{supp}(\mu)$, the following hold:

$$\begin{aligned} |m(z)| &\le \int_{\mathbb{R}} \frac{1}{|x-z|}\mu(\mathrm{d}x) \le \frac{1}{\mathrm{dist}(z, \mathrm{supp}(\mu))} \le \frac{1}{\mathrm{Im}(z)} = \frac{1}{\eta}, \\ m'(z) &= \int_{\mathbb{R}} \frac{1}{(x-z)^2}\mu(\mathrm{d}x), \\ |m'(z)| &\le \int_{\mathbb{R}} \frac{1}{|x-z|^2}\mu(\mathrm{d}x) \le \frac{1}{\eta^2}, \\ |m^{(k)}(z)| &\le (k-1)!\frac{1}{\eta^k}. \end{aligned}$$

- As we saw above, the Stieltjes transform m is infinitely differentiable, and moreover it is an *analytic function* on the set $\mathbb{C} \setminus \mathrm{supp}(\mu)$, meaning that the following hold:
  - The Taylor expansion

  $$\sum_{k=0}^{\infty} \frac{m^{(k)}(z_0)}{k!}(z-z_0)^n$$

  converges to $m(z)$ locally in a neighborhood around z_0, for any $z_0 \in \mathbb{C} \setminus \mathrm{supp}(\mu)$.
  - m can be determined by only a countable set of points. That is, if $(z_n)_{n\ge 1} \subseteq D \subseteq \mathbb{C}_+$, where D is an open set, and z_n converges to some $z_\star \in D$, then for any two Stieltjes functions $m, \widetilde{m}$ we have that if $m(z_n) = \widetilde{m}(z_n)$ for all n, then $m(z) = \widetilde{m}(z)$ on all of D.

- Suppose that μ has bounded support, say $\mathrm{supp}(\mu) \subseteq [-M, M]$. Then for any $z \in \mathbb{C}$ with $|z| > M$, by Taylor expanding, we have

$$\begin{aligned} m(z) &= \int_{\mathbb{R}} \frac{1}{x - z} \mu(\mathrm{d}x) = -z^{-1} \int_{\mathbb{R}} \Big(1 - \frac{x}{z}\Big)^{-1} \mu(\mathrm{d}x) \\ &= -z^{-1} \int_{\mathbb{R}} \sum_{k=0}^{\infty} \Big(\frac{x}{z}\Big)^{k} \mu(\mathrm{d}x) = -\sum_{k=0}^{\infty} z^{-(k+1)} \int_{\mathbb{R}} x^k \mu(\mathrm{d}x). \end{aligned}$$

  In particular, if we denote $\widetilde{m}(y) = m(1/y)$, then we can write

$$\widetilde{m}(y) = -\sum_{k=0}^{\infty} y^{k+1} \underbrace{\int_{\mathbb{R}} x^k \mu(\mathrm{d}x)}_{=\mathbb{E}_{X\sim\mu}[X^k]}.$$

  That is, the coefficients of the Taylor expansion of $\widetilde{m}$ (and m) are given by the moments of the probability measure μ. In particular, for $|z|$ large, we have

$$m(z) = -\frac{1}{z} + O\Big(\frac{1}{|z|^2}\Big).$$

We also have the following proposition, which allows us to *invert* the Stieltjes function to get the original measure μ back from $m(z)$.

Proposition 3 (Stieltjes inversion). *For two points $a < b$ where the CDF of μ is continuous, then*

$$\mu([a, b]) = \lim_{\eta} \int_a^b \frac{1}{\pi} Im(m(\lambda + i\eta))\mathrm{d}\lambda. \tag{5}$$

Intuitively, this result follows from our understanding of $\mathrm{Im}(m(z))$ as the convolution of μ with the $\mathrm{Cauchy}(0, \eta)$ density, since

$$\frac{1}{\pi}\mathrm{Im}(m(\lambda + i\eta)) = \mu \star \mathrm{Cauchy}(0, \eta) \xrightarrow{\eta\to 0} \mu.$$

What this inversion formula means is that there is a one-to-one correspondence between Stieltjes functions $m(z)$ and probability measures μ on $\mathbb{R}$. An important implication of this fact is that if two distributions $\mu, \widetilde{\mu}$ have the same Stieltjes functions $m = \widetilde{m}$, then μ and $\widetilde{\mu}$ are equal. This also holds for approximate statements; that is, if $m \approx \widetilde{m}$, then $\mu \approx \widetilde{\mu}$. In order to formalize this claim, we recall the definition of weak convergence of measures on $\mathbb{R}$.

Definition 4 (Weak convergence). *Let $(\mu_n)_{n\geq 1} \subseteq \mathcal{P}(\mathbb{R})$ be a sequence of probability measures. Then we say that $\mu_n \to \mu$ weakly if for any continuous function $f \in C(\mathbb{R})$, we have*

$$\int f(x)\mu_n(\mathrm{d}x) \xrightarrow{n\to\infty} \int f(x)\mu(\mathrm{d}x).$$

We can now state the following theorem, which formalizes our above claim. Note that this theorem can also be made quantitative.

Theorem 5. *Let $(\mu_n)_{n\geq 1} \subseteq \mathcal{P}(\mathbb{R})$ be a sequence of probability measures, with Stieltjes functions $(m_n) \subseteq \mathcal{F}(\mathbb{C}_+)$. If there exists $m : \mathbb{C}_+ \to \mathbb{C}_+$ such that:*

- *For any $z \in \mathbb{C}_+$, $m_n(z) \to m(z)$ and*
- *m is the Stieltjes function of some probability mesaure μ,*

then $\mu_n \to \mu$ weakly.

Remark 6. *Recall that our goal is to characterize*

$$\lim_{n\to\infty} F_n = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f(\lambda_i(\boldsymbol{A}))$$

for a random matrix $\boldsymbol{A}$. Then we can proceed with the following approach.

- *First, we figure out the limit*

$$m(z) = \lim_{n\to\infty} \mathbb{E}[m_n(z)] = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}[\mathsf{tr}((\boldsymbol{A} - z\boldsymbol{I})^{-1})],$$

  where $m_n(z)$ is the Stieltjes transform of the empirical distribution $\widehat{\mu}_{\boldsymbol{A}}(\mathrm{d}\lambda) = n^{-1}\sum_{i=1}^{n} \delta_{\lambda_i(\boldsymbol{A})}$. Also, show that $m_n(z)$ concentrates around its expectation, so that this limit can be stated almost surely.

- *Then, we express F_n as a function of the Stieltjes function m_n, i.e. a map $\mathcal{G} : \mathcal{F}(\mathbb{C}_+) \to \mathbb{R}$ such that*

$$F_n = \mathcal{G}(m_n) = \lim_{\eta\to 0} \int f(\lambda)\frac{1}{\pi} Im(m_n(\lambda + i\eta))\mathrm{d}\lambda.$$

  For example, the bias term we saw before in the linear model can be written as $B_n(\lambda) = \lambda^2 \cdot m'(-\lambda)$.

- *Finally, we want to show that the map $\mathcal{G}$ is continuous, so that $\lim_{n\to\infty} F_n = \mathcal{G}(m)$.*

In the next section, we see an example of this general recipe.

3 Limiting Stieltjes Transform for the GOE Matrix

In this section, we look at a 'toy' example of the procedure described in the previous section. Namely, we will compute the limiting Stieltjes transform for the GOE matrix, and use this to characterize the limit of the empirical distribution of its eigenvalues. First, let's recall the definition of the Gaussian Orthogonal Ensemble.

Definition 7 (Gaussian Orthogonal Ensemble)**.** *We write $\boldsymbol{W} \sim \mathrm{GOE}(n)$ to mean the random $n \times n$ matrix $\boldsymbol{W}$ for which*

$$\begin{aligned} &\boldsymbol{W}_{ij} \sim \mathcal{N}(0, 1/n) \quad 1 \le i < j \le n, \\ &\boldsymbol{W}_{ii} \sim \mathcal{N}(0, 2/n) \quad 1 \le i \le n, \\ &\boldsymbol{W}_{ji} = \boldsymbol{W}_{ij}. \end{aligned}$$

We also define the *semicircle law*.

Definition 8 (Semicircle Law)**.** *The semicircle law is the measure μ such that*

$$\mu(\mathrm{d}x) = p(x)\mathrm{d}x,$$

where $p(x) = (2\pi)^{-1}\sqrt{4 - x^2}\mathbf{1}\{x \in [-2, 2]\}$. See Figure 1 for an illustration of the density p.

The main result of this section is that for $\boldsymbol{W} \sim \mathrm{GOE}(n)$, the distribution $\mu_n = \widehat{\mu}_{\boldsymbol{W}}$ converges to the semicircle law μ. This is stated in the following theorem.

Theorem 9. *Let $\boldsymbol{W} \sim \mathrm{GOE}(n)$. Then almost surely as $n \to \infty$, $\mu_n = n^{-1}\sum_{i=1}^{n} \delta_{\lambda_i(\boldsymbol{W})}$ converges weakly to the semicircle law μ.*

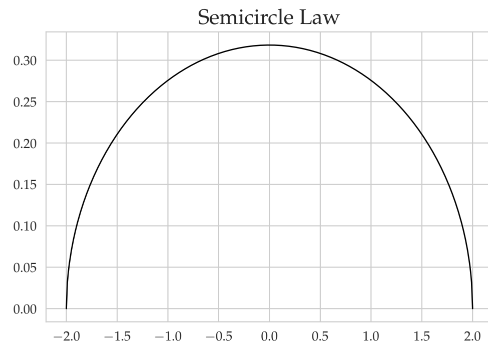

Figure 1: Illustration of the semicircle law $p(x) = (2\pi)^{-1}\sqrt{4 - x^2}\mathbf{1}\{x \in [-2, 2]\}$.

Remark 10. *The above theorem can be generalized to a Wigner random matrix, that is, a random symmetric matrix $\boldsymbol{W}$ with independent entries in the lower-left part, first and second moments matching* $\mathrm{GOE}(n)$*, and some other high order moment conditions.* $\mathrm{GOE}(n)$ *is a special case of a Wigner matrix, but the assumption that the entries are Gaussian can be relaxed.*

Proof sketch of Theorem 9. The main idea is to show the convergence of $m_n(z)$ — the Stieltjes function associated with $\mu_n = n^{-1}\sum_{i=1}^n \delta_{\lambda_i(\boldsymbol{W})}$ — to

$$m(z) = \frac{\sqrt{z^2 - 4} - z}{2},$$

which is the Stieltjes function associated with the semicircle law μ. Then, we apply Theorem 5 to obtain the desired result. The idea is to use the fact that $m(z)$ satisfies the self-consistent equation

$$m^2 + zm + 1 = 0. \tag{6}$$

To do this, we use a "leave-one-out" analysis. Heuristically, the approach is as follows.

- We want to study $m_n(z) = n^{-1}\mathsf{tr}((\boldsymbol{W} - z\boldsymbol{I})^{-1})$. Intuitively, we think of this as follows: for $\boldsymbol{W} \in \mathbb{R}^{n\times n}$, consider the smaller matrix $\boldsymbol{W}^{(n)} \in \mathbb{R}^{(n-1)\times(n-1)}$ where $\boldsymbol{W}^{(n)}_{ij} = \boldsymbol{W}_{ij}$, that is, we just delete the last row and column of $\boldsymbol{W}$. Then we observe that

  $$\boldsymbol{W}^{(n)} \sim \frac{\sqrt{n}}{\sqrt{n-1}}\mathrm{GOE}(n-1).$$

  Intuitively, the spectral properties of $\boldsymbol{W}$ and $\boldsymbol{W}^{(n)}$ should be approximately the same. Said other way, we would have that

  $$\frac{1}{n}\mathsf{tr}((\boldsymbol{W}^{(n)} - z\boldsymbol{I}_{n-1})^{-1}) \approx \frac{1}{n}\mathsf{tr}((\boldsymbol{W} - z\boldsymbol{I}_n)^{-1}), \tag{7}$$

  where the $\approx$ can be quantified. This intuition is the key to the leave-one-out approach.

- This approach leads to the following algebra. We want to compute $(\boldsymbol{W} - z\boldsymbol{I})^{-1}$. Recall the following expression for the inverse of a block matrix.

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

  The $\star$ terms can also be calculated, though for our purposes we are just interested in the first block term. We will use the following notation.

  $$\boldsymbol{\omega}_i := (\boldsymbol{W}_{si})_{s\neq i} \in \mathbb{R}^{n-1},$$

and

$$\boldsymbol{W}^{(i)} = (\boldsymbol{W}_{st})_{s,t\neq i} \in \mathbb{R}^{(n-1)\times(n-1)}.$$

Here $\boldsymbol{\omega}_i$ is the ith column of $\boldsymbol{W}$, omitting the ith row, and $\boldsymbol{W}^{(i)}$ (called the minor matrix of $\boldsymbol{W}$) is the matrix obtained by omitting the ith row and ith column of $\boldsymbol{W}$. To apply our block inversion formula, we take $\boldsymbol{A} = \boldsymbol{W}_{ii}$, $\boldsymbol{B}^\top = \boldsymbol{C} = \boldsymbol{\omega}_i$, and $\boldsymbol{D} = \boldsymbol{W}^{(i)} - z\boldsymbol{I}_{n-1}$. Using these definitions, we get that

$$(\boldsymbol{W} - z\boldsymbol{I})^{-1}_{ii} = \frac{1}{\boldsymbol{W}_{ii} - z - \boldsymbol{\omega}_i^\top(\boldsymbol{W}^{(i)} - z\boldsymbol{I}_{n-1})^{-1}\boldsymbol{\omega}_i}.$$

- Now recall that

$$m_n(z) = \frac{1}{n}\mathsf{tr}((\boldsymbol{W} - z\boldsymbol{I})^{-1}) = \frac{1}{n}\sum_{i=1}^n \frac{1}{\boldsymbol{W}_{ii} - z - \boldsymbol{\omega}_i^\top(\boldsymbol{W}^{(i)} - z\boldsymbol{I}_{n-1})^{-1}\boldsymbol{\omega}_i}. \tag{8}$$

Importantly, note that $\boldsymbol{\omega}_i$ is independent of $\boldsymbol{W}^{(i)}$, so we have that

$$\mathbb{E}[\boldsymbol{\omega}_i^\top(\boldsymbol{W}^{(i)} - z\boldsymbol{I}_{n-1})^{-1}\boldsymbol{\omega}_i \mid \boldsymbol{W}^{(i)}] = \frac{1}{n}\mathsf{tr}((\boldsymbol{W}^{(i)} - z\boldsymbol{I}_{n-1})^{-1}) \overset{(i)}{\approx} m_n(z), \tag{9}$$

where (i) comes from our heuristic in eqn. 7, which again can be made quantitative. Then, using concentration, we argue that

$$\boldsymbol{\omega}_i^\top(\boldsymbol{W}^{(i)} - z\boldsymbol{I}_{n-1})^{-1}\boldsymbol{\omega}_i \overset{(ii)}{\approx} \frac{1}{n}\mathsf{tr}((\boldsymbol{W}^{(i)} - z\boldsymbol{I}_{n-1})^{-1}).$$

Moreover, for fixed z, we have $z = O(1)$, and since $\boldsymbol{W}_{ii} \sim \mathcal{N}(0, 2/n)$, we have that $|\boldsymbol{W}_{ii}| = O(n^{-1/2})$. Thus, asymptotically, we ignore the $\boldsymbol{W}_{ii}$ term in eqn. 8. Combining these intuitions, we get that

$$m_n(z) \approx \frac{1}{n}\sum_{i=1}^n \frac{1}{-z - m_n(z)} + o_n(1) = \frac{1}{-z - m_n(z)} + o_n(1).$$

After rearranging to solve for m_n, we get

$$m_n(z)(z + m_n(z)) + 1 \approx 0 \iff m_n^2 + zm_n + 1 \approx 0, \tag{10}$$

which gives us back the desired self-consistent eqn. 6.

- What remains is to make the approximations (i) and (ii) rigorous, and show that after doing so the limit of the approximate self-consistent equation eqn. 10 is exact.

The approximation (ii) requires a concentration inequality (called the Hanson-Wright inequality) of the form

$$\mathbb{P}_{\boldsymbol{g}\sim\mathcal{N}(0,\boldsymbol{I}_n)}\Big(|\langle \boldsymbol{g}, \boldsymbol{A}\boldsymbol{g}\rangle/n - \mathsf{tr}(\boldsymbol{A})/n| \geq \frac{t}{n}\Big) \leq 2\cdot\exp\Big(-c\cdot\min\Big(\frac{t^2}{\|\boldsymbol{A}\|_F^2}, \frac{t}{\|\boldsymbol{A}\|_{op}}\Big)\Big), \tag{11}$$

where here we'll use $\boldsymbol{A} = (\boldsymbol{W}^{(i)} - z\boldsymbol{I})^{-1}$ and $\boldsymbol{g} = \sqrt{n}\boldsymbol{\omega}_i$. Indeed, one can verify that

$$\|(\boldsymbol{W}^{(i)} - z\boldsymbol{I})^{-1}\|_{op} \leq \frac{1}{\eta} \qquad \text{and}$$

$$\|(\boldsymbol{W}^{(i)} - z\boldsymbol{I})^{-1}\|_F \leq \frac{\sqrt{n}}{\eta}.$$

By plugging these into eqn. 11, we can verify that

$$\sup_{i\in[n]} |\langle \boldsymbol{g}^{(i)}, \boldsymbol{A}^{(i)}\boldsymbol{g}^{(i)}\rangle/n - \mathsf{tr}(\boldsymbol{A}^{(i)}/n| \approx O\Big(\sqrt{\frac{\log(n)}{n}}\Big).$$

Next, we need to deal with the approximation (i). To do this, we use the *eigenvalue interpolacing property*. To see what this means, let $\boldsymbol{W} \in \mathbb{R}^{n\times n}$ be symmetric, with $\boldsymbol{W}^{(n)}$ obtained as above by deleting the ith row and column of $\boldsymbol{W}$. Define the eigenvalues of these matrices as

$$\lambda_1(\boldsymbol{W}) \geq \lambda_2(\boldsymbol{W}) \geq \cdots \geq \lambda_n(\boldsymbol{W}),$$
$$\lambda_1(\boldsymbol{W}^{(i)}) \geq \lambda_2(\boldsymbol{W}^{(i)}) \geq \cdots \geq \lambda_n(\boldsymbol{W}^{(i)}).$$

Then the eigenvalue interpolacing property states that

$$\lambda_1(\boldsymbol{W}) \geq \lambda_1(\boldsymbol{W}^{(i)}) \geq \lambda_2(\boldsymbol{W}) \geq \cdots \geq \lambda_{n-1}(\boldsymbol{W}) \geq \lambda_{n-1}(\boldsymbol{W}^{(i)}) \geq \lambda_n(\boldsymbol{W}).$$

One can check that this allows us to make the approximation (i) quantitative.

Finally, we need to check that the approximate self-consistent equation eqn. 10 isn't changed when we take the limit. To see this, let

$$m_n(z) = \frac{1}{-z - m_n(z) - r_n(z)} \quad \text{where } r_n(z) \xrightarrow{a.s.} 0.$$

This gives the two solutions

$$m_n(z) = \frac{-(z + r_n(z)) \pm \sqrt{(z + r_n(z))^2 - 4}}{2}. \tag{12}$$

We'd like to argue that this approaches $m(z) = 2^{-1}(\sqrt{z^2-4} - z)$ as $n \to \infty$. To do this, we claim that for n large, only the positive solution from eqn. 12 will be in $\mathbb{C}_+$, which gives

$$m_n(z) = \frac{-(z + r_n(z)) + \sqrt{(z + r_n(z))^2 - 4}}{2} \xrightarrow{n\to\infty} m(z),$$

and hence the desired result.

□

For additional references, see [TV12, AGZ09, MP67].

References

[AGZ09] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni, *An introduction to random matrices*, Cambridge University Press, 2009.

[MP67] V. Marchenko and L. Pastur, *The eigenvalue distribution in some ensembles of random matrices*, Math. USSR (1967).

[TV12] Terence Tao and Van Vu, *Random matrices: The universality phenomenon for wigner ensembles*, Tech. report, 2012.