

Inhomogeneous circular laws for random matrices with non-identically distributed entries

Nick Cook

with Walid Hachem (Telecom ParisTech), Jamal Najim (Paris-Est) and
David Renfrew (SUNY Binghamton/IST Austria)



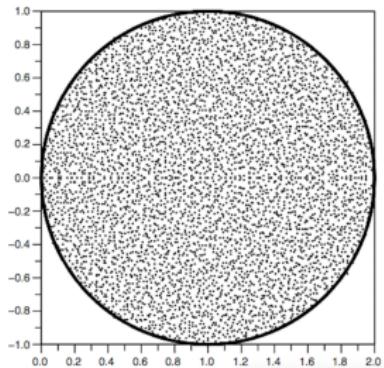
STANFORD
UNIVERSITY
Department of Mathematics

UC Davis, 2017/04/19

The Circular Law

Empirical spectral distribution (ESD) for $M \in \mathcal{M}_n(\mathbb{C})$ with eigenvalues $\lambda_1, \dots, \lambda_n \in \mathbb{C}$:

$$\mu_M = \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i}.$$



iid matrix model:

- Atom variable: $\xi \in \mathbb{C}$ with $\mathbb{E} \xi = 0$, $\mathbb{E} |\xi|^2 = 1$.
- For $n \geq 1$ let X_n be $n \times n$ with iid entries $\xi_{ij}^{(n)} \stackrel{d}{=} \xi$.

Theorem (Mehta '67, Girko '84, Edelman '97, Bai '97, Götze–Tikhomirov '05, '07, Pan–Zhou '07, Tao–Vu '07, **Tao–Vu '08**)

Almost surely, $\mu_{\frac{1}{\sqrt{n}}X_n}$ converges weakly to $\frac{1}{\pi}1_{B(0,1)}dxdy$ as $n \rightarrow \infty$.

Applications: stability for mean field models of dynamical systems

- Model a neural network with a nonlinear system of ODEs:

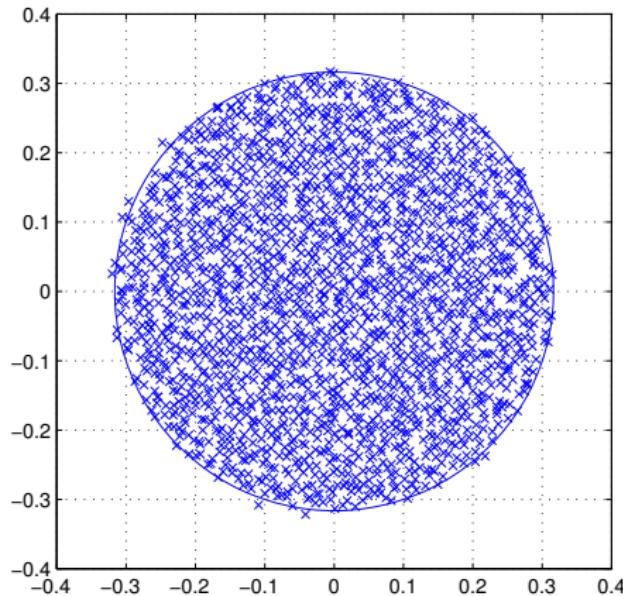
$$\dot{x}_i(t) = -x_i(t) + \sum_{j=1}^n Y_{ij} \tanh(x_j(t))$$

where x_i is the membrane potential for the i th neuron and $\tanh(x_i)$ is its firing rate.

- Sompolinsky, Crisanti, Sommers '88: modeled the *synaptic matrix* Y with a random matrix. (Similar to approach by [May '72] in ecology.)
- More recent work [Rajan, Abbott '06], [Tao '09], [Aljadeff, Renfrew, Stern '14] has tried to incorporate other structural features of neural networks into the distribution of Y .
- Desirable to consider Y with a general, possibly sparse variance profile, i.e.

$$Y_n = \frac{1}{\sqrt{n}} A_n \circ X_n = \left(\frac{1}{\sqrt{n}} \sigma_{ij} \xi_{ij} \right).$$

Simulated ESDs for $Y_n = \left(\frac{1}{\sqrt{n}} \sigma_{ij} \xi_{ij} \right)$.



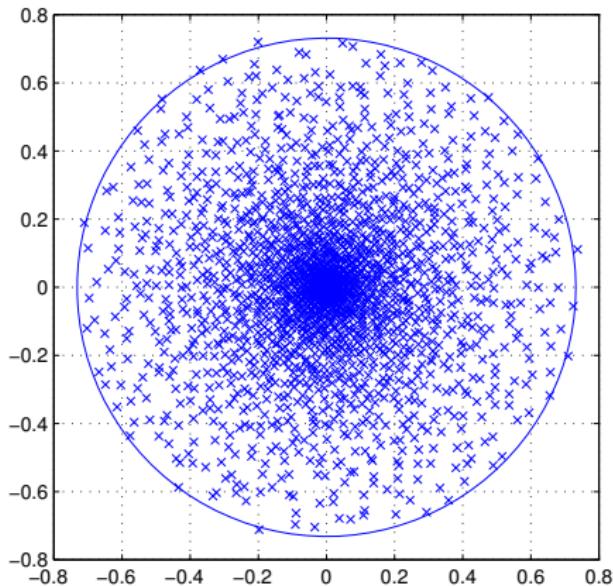
$$n = 2000$$

$$\xi \in \left\{ \pm \frac{1}{\sqrt{2}}, \pm i \frac{1}{\sqrt{2}} \right\} \text{ uniform.}$$

$$\sigma_{ij} = \sigma\left(\frac{i}{n}, \frac{j}{n}\right), \text{ with}$$

$$\sigma(x, y) = \mathbb{1}(|x - y| \leq 0.05)$$

Simulated ESDs for $Y_n = \left(\frac{1}{\sqrt{n}} \sigma_{ij} \xi_{ij} \right)$



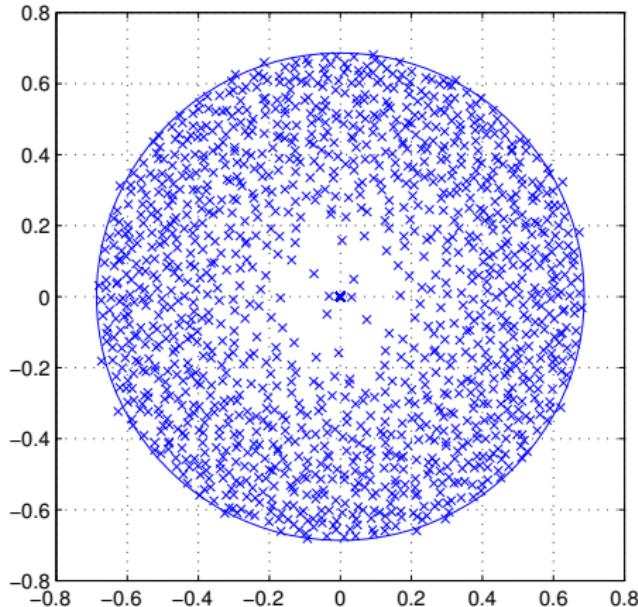
$n = 2000$

$\xi \in \{\pm \frac{1}{\sqrt{2}}, \pm i \frac{1}{\sqrt{2}}\}$ uniform.

$\sigma_{ij} = \sigma\left(\frac{i}{n}, \frac{j}{n}\right)$, with

$\sigma(x, y) = (x + y)^2 \mathbb{1}(|x - y| \leq 0.1)$

Simulated ESDs for $Y_n = \left(\frac{1}{\sqrt{n}} \sigma_{ij} \xi_{ij} \right)$



$n = 2001$

$\xi \in \{\pm \frac{1}{\sqrt{2}}, \pm i \frac{1}{\sqrt{2}}\}$ uniform.

$$A_n = (\sigma_{ij}) = \begin{pmatrix} 0 & \mathbf{1}_{n/3} & \mathbf{1}_{n/3} \\ \mathbf{1}_{n/3} & 0 & 0 \\ \mathbf{1}_{n/3} & 0 & 0 \end{pmatrix}.$$

The model and assumptions

$$Y_n = \frac{1}{\sqrt{n}} A_n \circ X_n = \left(\frac{1}{\sqrt{n}} \sigma_{ij}^{(n)} \xi_{ij}^{(n)} \right).$$

Inputs:

- Atom variable $\xi \in \mathbb{C}$ with $\mathbb{E} \xi = 0$, $\mathbb{E} |\xi|^2 = 1$, $\mathbb{E} |\xi|^{4+\varepsilon} < \infty$.
- $X_n = (\xi_{ij}^{(n)})$ sequence of iid matrices with atom variable ξ .
- Standard deviation profiles $A_n = (\sigma_{ij}^{(n)})$ with **uniformly bounded entries**, i.e. $\sigma_{ij}^{(n)} \in [0, 1]$.

Additionally assume A_n is “robustly irreducible” for all n .

Assumptions allow for vanishing variances ($\sigma_{ij}^{(n)} = 0$ for (say) 99% of entries).

Results

$$Y_n = \frac{1}{\sqrt{n}} A_n \circ X_n = \left(\frac{1}{\sqrt{n}} \sigma_{ij}^{(n)} \xi_{ij}^{(n)} \right), \text{ with}$$

- $\mathbb{E} |\xi|^{4+\varepsilon} < \infty$,
- $\sigma_{ij}^{(n)} \in [0, 1]$,
- A_n “robustly irreducible”.

Theorem (C., Hachem, Najim, Renfrew '16)

(Abridged) In the above setup, for each n , A_n **determines** a deterministic, compactly supported, rotationally invariant probability measure μ_n over \mathbb{C} such that $\mu_{Y_n} \sim \mu_n$ in probability, i.e.

$$\int_{\mathbb{C}} f \, d\mu_{Y_n} - \int_{\mathbb{C}} f \, d\mu_n \rightarrow 0 \quad \text{in probability } \forall f \in C_b(\mathbb{C}).$$

Refer to measures μ_n as deterministic equivalents for μ_{Y_n} .

Recent work of Alt, Erdős and Krüger gives a local version under stronger hypotheses (ξ having bounded density and all moments, $\sigma_{ij}^{(n)} \geq \sigma_{\min} > 0$).

Results

$$Y_n = \frac{1}{\sqrt{n}} A_n \circ X_n = \left(\frac{1}{\sqrt{n}} \sigma_{ij}^{(n)} \xi_{ij}^{(n)} \right), \text{ with}$$

- $\mathbb{E} |\xi|^{4+\varepsilon} < \infty$,
- $\sigma_{ij}^{(n)}$ uniformly bounded,
- ~~A_n “robustly irreducible”.~~ $V_n = (\frac{1}{n} \sigma_{ij}^2)$ doubly stochastic.

Theorem (Circular law for doubly stochastic variance profile)

With Y_n as above, $\mu_{Y_n} \rightarrow \frac{1}{\pi} 1_{B(0,1)} dx dy$ in probability.

Analogue of a result of Anderson and Zeitouni '06 for Hermitian random matrices.

How does A_n determine μ_n ? Through the *Master Equations*

For $s > 0$ consider the following system in unknowns $\mathbf{q}, \tilde{\mathbf{q}} \in \mathbb{R}^n$:

$$ME(s) : \begin{cases} q_i = \frac{(V_n^T \mathbf{q})_i}{s^2 + (V_n \tilde{\mathbf{q}})_i (V_n^T \mathbf{q})_i}, \\ \tilde{q}_i = \frac{(V_n \mathbf{q})_i}{s^2 + (V_n \tilde{\mathbf{q}})_i (V_n^T \mathbf{q})_i}, \end{cases} \quad \sum_{i=1}^n q_i = \sum_{i=1}^n \tilde{q}_i.$$

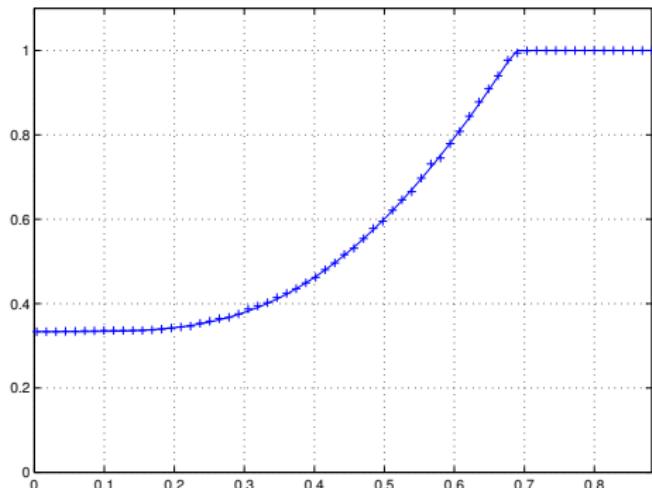
Can show that if V_n is irreducible, $s \in (0, \sqrt{\rho(V_n)})$,
 $\Rightarrow \exists$ unique non-trivial solution $(\mathbf{q}(s), \tilde{\mathbf{q}}(s)) \in \mathbb{R}_{\geq 0}^{2n}$.

μ_n is the radially symmetric probability measure on \mathbb{C} with

$$\mu_n(B(0, s)) = 1 - \frac{1}{n} \mathbf{q}(s)^T V_n \tilde{\mathbf{q}}(s) \quad \forall s \in (0, \infty)$$

where we set $\mathbf{q}(s) = \tilde{\mathbf{q}}(s) = \vec{0}$ for $s \geq \sqrt{\rho(V_n)}$.

Plot of $F_n(s) = 1 - \frac{1}{n} \mathbf{q}(s)^\top V_n \tilde{\mathbf{q}}(s)$ (curve) vs empirical realization (+)



$n = 2001$

$\xi \in \{\pm \frac{1}{\sqrt{2}}, \pm i \frac{1}{\sqrt{2}}\}$ uniform.

$$A_n = (\sigma_{ij}) = \begin{pmatrix} 0 & \mathbf{1}_{n/3} & \mathbf{1}_{n/3} \\ \mathbf{1}_{n/3} & 0 & 0 \\ \mathbf{1}_{n/3} & 0 & 0 \end{pmatrix}.$$

Useful transforms for proving convergence of measures

- For sums of independent scalar r.v.'s use the **Fourier transform** (characteristic function).
- For μ supported on \mathbb{R} (such as ESDs of Hermitian matrices) we have the **Stieltjes transform** $s_\mu : \mathbb{C}_+ \rightarrow \mathbb{C}_+$,

$$s_\mu(w) = \int_{\mathbb{R}} \frac{d\mu(x)}{x - w}, \quad \mu = \lim_{t \downarrow 0} \frac{1}{\pi} \operatorname{Im} s_\mu(\cdot + it) = \lim_{t \downarrow 0} \mu * \text{Cauchy}(t).$$

- For μ supported on \mathbb{C} (such as ESDs of non-Hermitian matrices) we have the **log transform** (or Coulomb potential)

$$U_\mu(z) = - \int_{\mathbb{C}} \log |\lambda - z| d\mu(\lambda), \quad \mu = -\frac{1}{2\pi} \Delta U_\mu.$$

Girko's Hermitization approach

log transform for probability measure μ on \mathbb{C} :

$$U_\mu(z) = - \int_{\mathbb{C}} \log |\lambda - z| \mu(d\lambda), \quad \mu = -\frac{1}{2\pi} \Delta U_\mu.$$

For the ESD of $Y_n = \frac{1}{\sqrt{n}} A_n \circ X_n$,

$$\begin{aligned} U_{\mu_{Y_n}}(z) &= -\frac{1}{n} \sum_{i=1}^n \log |\lambda_i(Y_n) - z| = -\frac{1}{n} \log |\det(Y_n - z)| \\ &= -\frac{1}{2n} \log |\det(\mathbf{Y}_n^z)| = - \int_{\mathbb{R}} \log |x| d\mu_{\mathbf{Y}_n^z}(x) \end{aligned}$$

where $\mathbf{Y}_n^z := \begin{pmatrix} 0 & Y_n - z \\ Y_n^* - \bar{z} & 0 \end{pmatrix}$ is a $2n \times 2n$ **Hermitian** matrix with eigenvalues $\{\pm s_i(Y_n - z)\}_{i=1}^n \subset \mathbb{R}$.

Girko's Hermitization approach

$$\mathbf{Y}_n^z = \begin{pmatrix} 0 & Y_n - z \\ Y_n^* - \bar{z} & 0 \end{pmatrix}, \quad U_{\mu_{Y_n}}(z) = - \int_{\mathbb{R}} \log |x| d\mu_{\mathbf{Y}_n^z}(x).$$

Two steps. For a.e. $z \in \mathbb{C}$:

- ① Find deterministic equivalents $\nu_n^z \sim \mu_{\mathbf{Y}_n^z}$.
- ② Prove these measures uniformly (in n) integrate $\log |\cdot|$.

Now we can use the Stieltjes transform

$$s_n^z(w) := \int_{\mathbb{R}} \frac{d\mu_{\mathbf{Y}_n^z}(x)}{x - w} = \frac{1}{2n} \sum_{i=1}^{2n} \frac{1}{\lambda_i(\mathbf{Y}_n^z) - w} = \frac{1}{2n} \operatorname{Tr} \mathbf{R}_n^z(w)$$

where $\mathbf{R}_n^z(w) := (\mathbf{Y}_n^z - w)^{-1}$ is the **resolvent** of \mathbf{Y}_n^z .

Deterministic equivalents for resolvents $\mathbf{R}_n^z(w)$: Complex Gaussian case

$$I_{2n} = \mathbf{R}_n^z(w)^{-1} \mathbf{R}_n^z(w) = \begin{pmatrix} -w & Y_n - z \\ Y_n^* - \bar{z} & -w \end{pmatrix} \begin{pmatrix} S & T \\ \tilde{T} & \tilde{S} \end{pmatrix} = \begin{pmatrix} -wS - z\tilde{T} + Y\tilde{T} & * \\ * & * \end{pmatrix}$$

$$1 = -w \mathbb{E} S_{ii} - z \mathbb{E} \tilde{T}_{ii} + \sum_{j=1}^n \mathbb{E} Y_{ij} \tilde{T}_{ji}$$

$$(\text{Gaussian IBP}) = -w \mathbb{E} S_{ii} - z \mathbb{E} \tilde{T}_{ii} + \sum_{j=1}^n \frac{1}{n} \sigma_{ij}^2 \mathbb{E} \partial_{Y_{ij}} \tilde{T}_{ji}$$

$$(\text{Resolvent derivative formula}) = -w \mathbb{E} S_{ii} - z \mathbb{E} \tilde{T}_{ii} - \frac{1}{n} \sum_{j=1}^n \sigma_{ij}^2 \mathbb{E} S_{ii} \tilde{S}_{jj}$$

$$(\text{Cauchy-Schwarz \& Poincaré}) \approx -w \mathbb{E} S_{ii} - z \mathbb{E} \tilde{T}_{ii} - \frac{1}{n} (\mathbb{E} S_{ii}) \sum_{j=1}^n \sigma_{ij}^2 (\mathbb{E} \tilde{S}_{jj}).$$

$$\text{Deterministic equivalents for resolvents } \mathbf{R}_n^z(w) = (\mathbf{Y}_n^z - w)^{-1} = \begin{pmatrix} S & T \\ \tilde{T} & \tilde{S} \end{pmatrix}$$

Similarly obtain equations for

$$\mathbb{E} S_{ii}, \mathbb{E} \tilde{S}_{ii}, \mathbb{E} T_{ii}, \mathbb{E} \tilde{T}_{ii}, \quad 1 \leq i \leq n$$

from diagonal entries of the other three blocks. Eventually reduce to a perturbed cubic system of $2n$ equations:

$$\mathbb{E} S_{ii} = \frac{(V_n^\top \mathbf{s})_i + w}{|z|^2 - [(V_n^\top \mathbf{s})_i + w][(V_n \tilde{\mathbf{s}})_i + w]} + \mathcal{E}_n,$$

$$\mathbb{E} \tilde{S}_{ii} = \frac{(V_n \tilde{\mathbf{s}})_i + w}{|z|^2 - [(V_n^\top \mathbf{s})_i + w][(V_n \tilde{\mathbf{s}})_i + w]} + \mathcal{E}'_n,$$

where $\mathbf{s} := (\mathbb{E} S_{ii})_{i=1}^n$, $\tilde{\mathbf{s}} := (\mathbb{E} \tilde{S}_{ii})_{i=1}^n$, and $\mathcal{E}_n, \mathcal{E}'_n$ are small errors.

Deterministic equivalents for resolvents $\mathbf{R}_n^z(w) = (\mathbf{Y}_n^z - w)^{-1} = \begin{pmatrix} S & T \\ \tilde{T} & \tilde{S} \end{pmatrix}$

Stability analysis $\Rightarrow \mathbf{s}, \tilde{\mathbf{s}}$ are close to solutions $\mathbf{p}(|z|, w), \tilde{\mathbf{p}}(|z|, w)$ of the unperturbed system (the *Schwinger–Dyson loop equations*):

$$LE(|z|, w) : \begin{cases} p_i &= \frac{(V_n^\top \mathbf{p})_i + w}{|z|^2 - [(V_n^\top \mathbf{p})_i + w][(V_n \tilde{\mathbf{p}})_i + w]}, \\ \tilde{p}_i &= \frac{(V_n \tilde{\mathbf{p}})_i + w}{|z|^2 - [(V_n^\top \mathbf{p})_i + w][(V_n \tilde{\mathbf{p}})_i + w]}. \end{cases}$$

In particular:

$$\begin{aligned} \mathbb{E} s_n^z(w) &= \mathbb{E} \frac{1}{2n} \operatorname{Tr} \mathbf{R}_n^z(w) = \mathbb{E} \frac{1}{2n} \left(\sum_{i=1}^n S_{ii} + \sum_{i=1}^n \tilde{S}_{ii} \right) \\ &\sim \frac{1}{2n} \left(\sum_{i=1}^n p_i + \sum_{i=1}^n \tilde{p}_i \right) = \frac{1}{n} \sum_{i=1}^n p_i. \end{aligned}$$

Deterministic equivalents for resolvents $\mathbf{R}_n^z(w) = (\mathbf{Y}_n^z - w)^{-1} = \begin{pmatrix} S & T \\ \tilde{T} & \tilde{S} \end{pmatrix}$

$$\mathbb{E} s_n^z(w) \sim \frac{1}{n} \sum_{i=1}^n p_i.$$

- Extend to the non-Gaussian case by a Lindeberg swapping argument, and remove the expectation with concentration of measure. Get

$$s_n^z(w) \sim \frac{1}{n} \sum_{i=1}^n p_i \quad a.s.$$

in the general case (and we can quantify the error).

- RHS is the Stieltjes transform of a probability measure ν_n^z on \mathbb{R} , so

$$\mu_{\mathbf{Y}_n^z} \sim \nu_n^z \quad a.s.$$

Integrability of log

We obtained deterministic equivalents ν_n^z for the “Hermitized” ESDs $\mu_{Y_n^z}$.

Remains to show for a.e. $z \in \mathbb{C}$, $\mu_{Y_n^z}$ and ν_n^z uniformly integrate $\log |x|$, i.e.

$$\forall \varepsilon > 0 \quad \exists T > 0 : \quad \left| \int_{|\log|x|| \geq T} \log |x| d\mu_{Y_n^z}(x) \right| \leq \varepsilon$$

with probability $\geq 1 - \varepsilon$, and similarly for ν_n^z .

Singularity of \log at ∞ is easy to handle.

Two-step approach to singularity at 0:

- ① (Wegner estimate) Use bounds on the Stieltjes transform to show $\mu_{Y_n^z}([-t, t]) = O(t)$ for $t \geq n^{-c}$.
- ② (Invertibility) Prove $|\lambda_{\min}|(Y_n^z) = s_{\min}(Y_n - z) \geq n^{-C}$ w.h.p.

Integrability of \log : Wegner estimates

Stieltjes transform controls the density of eigenvalues in short intervals:

$$\frac{1}{t} \mu_{Y_n^z}([-t, t]) \lesssim \mu_{Y_n^z} * \text{Cauchy}(t) = \text{Im } s_n^z(it).$$

From the loop equations:

$$\text{Im } s_n^z(it) \approx \frac{1}{n} \sum_{i=1}^n \text{Im } p_i(|z|, it).$$

Key Proposition

Let $z \neq 0, t > 0$, and let $p(|z|, it)$, $\tilde{p}(|z|, it)$ be solutions to $LE(|z|, it)$. If A_n is **robustly irreducible**, then

$$\frac{1}{n} \sum_{i=1}^n \text{Im } p_i(|z|, it) \leq K$$

for some constant $K < \infty$ independent of n, t .

Integrability of $\log \rightarrow$ Invertibility of structured random matrices

Want to show $s_{\min}(Y_n - z) \geq n^{-C}$ with high probability (w.h.p.).

Recall $s_{\min}(M) = 0$ iff M is singular, and otherwise $s_{\min}(M) = 1/\|M^{-1}\|$.

Say a random matrix M is *well-invertible w.h.p.* if

$$\mathbb{P} \left\{ \|M^{-1}\| \geq n^\alpha \right\} = O(n^{-\beta})$$

for some constants $\alpha, \beta > 0$.

Question: For what choices of $n \times n$ matrices $A = (a_{ij}), B = (b_{ij})$ is

$$Y = \frac{1}{\sqrt{n}} A \circ X + B = \left(\frac{1}{\sqrt{n}} a_{ij} \xi_{ij} + b_{ij} \right)$$

well-invertible w.h.p.?

(For convergence of ESDs we are interested in the shifts $B = -zI_n$.)

$$\text{Invertibility of } \frac{1}{\sqrt{n}} A \circ X + B = \left(\frac{1}{\sqrt{n}} a_{ij} \xi_{ij} + b_{ij} \right)$$

Condition number $\|X\| \|X^{-1}\|$ for iid matrices is well-studied
(von Neumann et al. '40s, Edelman '88, Sankar–Spielman–Teng '06,
Tao–Vu '05,'07, Rudelson '05, Rudelson–Vershynin '07).

Norm: $\|X\| = O(\sqrt{n})$ w.h.p. (folklore / Bai–Yin)

Norm of the inverse:

$$\mathbb{P} \left\{ \|X^{-1}\| \geq n^{\alpha(\beta)} \right\} \lesssim n^{-\beta} \quad \text{Tao–Vu '05,'07}$$

$$\mathbb{P} \left\{ \|X^{-1}\| \geq \sqrt{n}/\varepsilon \right\} \lesssim \varepsilon + e^{-cn} \quad \text{Rudelson–Vershynin '07}$$

(ξ sub-Gaussian).

So $\|X\| \|X^{-1}\| = n^{O(1)}$ w.h.p.

Invertibility of $\frac{1}{\sqrt{n}}A \circ X + B = (\frac{1}{\sqrt{n}}a_{ij}\xi_{ij} + b_{ij})$

Structured matrices: $\frac{1}{\sqrt{n}}A \circ X + B$ is well-invertible w.h.p. when

$$a_{ij} \geq \sigma_0 > 0, \quad \|B\| \leq n^{O(1)} \quad \text{Bordenave-Chafaï '11}$$

$$A \text{ is broadly connected}, \quad \|B\| = O(1) \quad \text{Rudelson-Zeitouni '12 for } \xi \sim N_{\mathbb{R}}(0, 1), \text{ C. '16 general case.}$$

The above give bounds that are uniform in the shift B , i.e.

$$\sup_{B: \|B\| \leq n^{O(1)}} \mathbb{P} \left\{ \left\| \left(\frac{1}{\sqrt{n}}A \circ X + B \right)^{-1} \right\| \geq n^\alpha \right\} = O(n^{-\beta}).$$

Can we further relax hypotheses on A if B is well-invertible (such as $B = -zI_n$ for fixed $z \neq 0$)?

Well-invertibility under diagonal perturbation

Theorem (C. '16)

Let $Y = \frac{1}{\sqrt{n}} A \circ X + Z$, where

- ① $X = (\xi_{ij})$ iid matrix with $\mathbb{E} \xi_{ij} = 0$, $\mathbb{E} |\xi_{ij}|^2 = 1$, $\mathbb{E} |\xi_{ij}|^{4+\varepsilon} = \gamma < \infty$;
- ② $A \in [0, 1]^{n \times n}$ arbitrary;
- ③ $Z = \text{diag}(z_i)_{i=1}^n$, $|z_i| \in [r, R] \subset (0, +\infty)$ for all $i \in [n]$.

There exist $\alpha, \beta > 0$, $C < \infty$ depending only on $\varepsilon, \gamma, r, R$ such that

$$\mathbb{P} (\|Y^{-1}\| \geq n^\alpha) \leq Cn^{-\beta}.$$

Note: Proof gives $\alpha = \text{twr}[O_\varepsilon(1) \exp((\gamma/r)^{O(1)})]$. The tower exponential is due to use of Szemerédi's regularity lemma.

Conjecture

Above holds if Z is any matrix with $s_i(Z) \in [r, R]$ for all $i \in [n]$.