Anisotropic local laws for random matrices

Antti Knowles

ETH Zürich

with Jun Yin

A few examples of random matrices

Consider a Hermitian $M \times M$ random matrix Q normalized so that $\|Q\| \asymp 1$.

(a) Wigner matrix. The entries $(Q_{ij} : 1 \leqslant i \leqslant j \leqslant M)$ are independent and satsify

$$\mathbb{E}Q_{ij} = 0\,, \qquad \mathbb{E}|Q_{ij}|^2 = M^{-1}\,.$$

(Hamiltonian of a disordered mean-field quantum system.)

(b) Band matrix. Like a Wigner matrix, except that

$$\mathbb{E}|Q_{ij}|^2 \;=\; W^{-1}\,\mathbf{1}(|i-j| \leqslant W)\,,$$

where $1 \ll W \ll M$ is the band width.

(Hamiltonian with spatial structure.)

(c) Sample covariance matrix. $Q = XX^*$, where $X \in \mathbb{C}^{M \times N}$ has independent entries satisfying

$$\mathbb{E}X_{ij} = 0\,, \qquad \mathbb{E}|X_{ij}|^2 = N^{-1}\,.$$

(Sample covariance matrix of uncorrelated data.)

(d) Random graph. Graph on M vertices,

$$Q_{ij} \;:=\; \alpha\,\mathbf{1}(i \sim j)$$

is the (rescaled) adjacency matrix. Example: Erdős-Rényi graph $G(M,p)$ with $M^{-1} \ll p \ll 1$ and $\alpha = (pM)^{-1/2}$.

The resolvent

Goal: distribution of eigenvalues

$$\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_M$$

and eigenvectors

$$\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_M \in \mathbb{S}^{M-1}$$

of Q.

Right tool: the resolvent

$$R(z) := (Q - zI)^{-1}, \qquad z = E + \mathrm{i}\eta \in \mathbb{C}_+ .$$

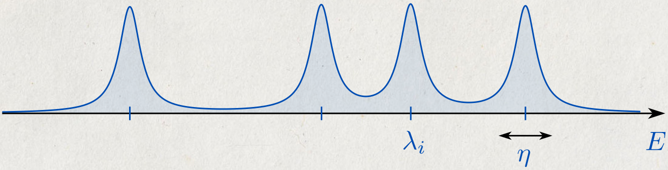
Contains the complete information about the eigenvalues and eigenvectors:

$$R(z) = \sum_{i=1}^{M} \frac{\mathbf{u}_i \mathbf{u}_i^*}{\lambda_i - z} .$$

Global and local laws

From the spectral decomposition of Q we get

$$\operatorname{Im} \frac{1}{M} \operatorname{Tr} R(z) \;=\; \pi \, \frac{1}{M} \sum_{i=1}^{M} \frac{\eta/\pi}{(\lambda_i - E)^2 + \eta^2}$$



Observation: η is the spectral resolution.

- Global law: Control of $R(z)$ for $\eta \asymp 1$.
- Local law: Control of $R(z)$ for $\eta \gg M^{-1}$.

A local law is required to understand the distribution of individual eigenvalues and eigenvectors.

In fact, for all of these applications one has to control $R(z)$ as a matrix.

Isotropy

For the models (a)–(d) one can show that

$$R(z) \approx m(z)I \qquad (\eta \gg M^{-1})$$

with high probability, where m is the Stieltjes transform of the asymptotic eigenvalue density. (Erdős-Schlein-Yau-Yin (2009-2010), Erdős-K-Yau-Yin (2011-2013), K-Yin (2012), Pillai-Yin (2012).)

$R(z)$ is asymptotically isotropic.

From this one can deduce (under some additional assumptions) that $\mathbf{u}_i \sim \mathrm{Unif}(\mathbb{S}^{M-1})$ for all i.

More complicated models, typically with correlated entries, are anisotropic.

Main example: sample covariance matrix

Correlated M-dimensional data $\mathbf{a} = (a_1, \dots, a_M)^* \in \mathbb{R}^M$ with population covariance matrix

$$\Sigma_{ij} \ := \ \mathbb{E}\big[(a_i - \mathbb{E}a_i)(a_j - \mathbb{E}a_j)\big] \,.$$

Take N independent copies $A = [\mathbf{a}_1 \cdots \mathbf{a}_N] \in \mathbb{R}^{M \times N}$ of $\mathbf{a}$, and define the sample covariance matrix

$$\mathcal{Q}_{ij} \ := \ \frac{1}{N-1} \sum_{\mu=1}^{N} (A_{i\mu} - [A]_i)(A_{j\mu} - [A]_j) \,, \qquad [A]_i := \frac{1}{N} \sum_{\mu=1}^{N} A_{i\mu} \,.$$

Without loss of generality, $\mathbb{E}\mathbf{a} = 0$.

For simplicity, consider

$$Q_{ij} \ := \ \frac{1}{N} \sum_{\mu=1}^{N} A_{i\mu} A_{j\mu}$$

instead of $\mathcal{Q}_{ij}$. All of the following results hold for Q and $\mathcal{Q}$.

Model for population $\mathbf{a}$

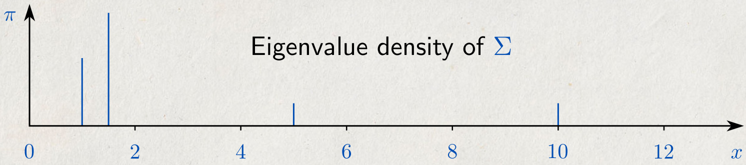
- $\mathbf{a} = T\mathbf{b}$, $T \in \mathbb{R}^{M \times \widehat{M}}$ is deterministic and $\mathbf{b} \in \mathbb{R}^{\widehat{M}}$ has independent entries.
- Entries of $\mathbf{b}$ have enough uniformly bounded moments.
- $M \asymp \widehat{M} \asymp N$.
- $\Sigma := \mathbb{E}\mathbf{a}\mathbf{a}^* \leqslant C$.

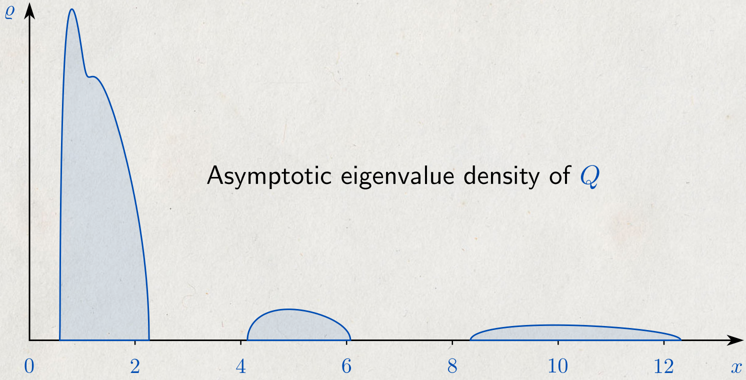
Example. Signal + noise model

$$\mathbf{a} \;=\; \sum_{l=1}^{r} y_l \mathbf{u}_l + \mathbf{z}\,,$$

where $y_1, \dots, y_r$ (signal) and $z_1, \dots, z_M$ (noise) are independent random variables, and $\mathbf{u}_l$ are deterministic vectors.

The asymptotic eigenvalue density





Denote by π the empirical spectral measure of Σ: $\pi := \frac{1}{M}\sum_{i=1}^{M} \delta_{\sigma_i}$ where $\{\sigma_i\}$ are the eigenvalues of Σ.

Define

$$f(x) \;:=\; -\frac{1}{x} + \frac{M}{N}\int \frac{\pi(\mathrm{d}s)}{x + s^{-1}}\,.$$

Then for each $z \in \mathbb{C}_+$ the equation $z = f(m)$ has a unique solution $m \equiv m(z) \in \mathbb{C}_+$

The function $m(z)$ is the Stieltjes transform of a probability measure ϱ.

Global law:

Theorem [Marchenko-Pastur (1967), Silverstein (1995)]. For $\eta \asymp 1$ we have

$$\frac{1}{M}\operatorname{Tr} R(z) \;=\; m(z) + o_P(1)\,.$$

The anisotropic local law

Theorem [K–Yin (2014)]. Suppose that π satisfies a regularity condition (see later). Then for $\eta \gg M^{-1}$ we have

$$\langle \mathbf{v} , R(z)\mathbf{w} \rangle \;=\; \langle \mathbf{v} , P(z)\mathbf{w} \rangle + O_{\mathrm{HP}}(\Psi(z)|\mathbf{v}||\mathbf{w}|)\,,$$

where

$$P(z) \;:=\; -\big(z(I + m(z)\Sigma)\big)^{-1}\,, \qquad \Psi(z) \;:=\; \sqrt{\frac{\operatorname{Im} m(z)}{M\eta}} + \frac{1}{M\eta}\,.$$

The rate of convergence given by Ψ is optimal.

Previously, an anisotropic global law (for $\eta \asymp 1$) for a related model was derived in Hachem-Loubaton-Najim-Vallet (2013).

Trivial consequence: complete delocalization of eigenvectors

For $\eta \geqslant \alpha M^{-1}$ we have $\|P(z)\| \leqslant C$ and $\Psi(z) \leqslant C$.

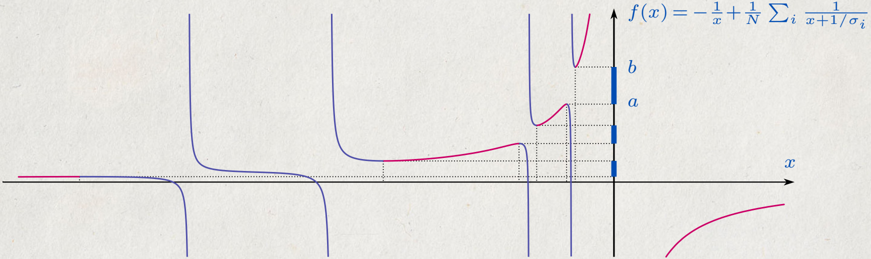
With $z := \lambda_i + \mathrm{i}\alpha M^{-1}$ we get for any $\mathbf{v} \in \mathbb{S}^{M-1}$

$$
\begin{aligned}
C \;&\gtrsim_{\mathrm{HP}}\; \mathrm{Im}\langle \mathbf{v}\,, R(z)\mathbf{v}\rangle \\
&=\; \sum_{j=1}^{M} \frac{\eta}{(\lambda_i - \lambda_j)^2 + \eta^2} |\langle \mathbf{v}\,, \mathbf{u}_j\rangle|^2 \\
&\geqslant\; \eta^{-1} |\langle \mathbf{v}\,, \mathbf{u}_i\rangle|^2 \\
&=\; \frac{M}{\alpha}\, |\langle \mathbf{v}\,, \mathbf{u}_i\rangle|^2\,.
\end{aligned}
$$

Complete delocalization of all eigenvectors with respect to an arbitrary basis.

The regularity condition

Fact: the edges of ϱ in $(0,\infty)$ are given by $E := \{f(x) : f'(x) = 0\}$.

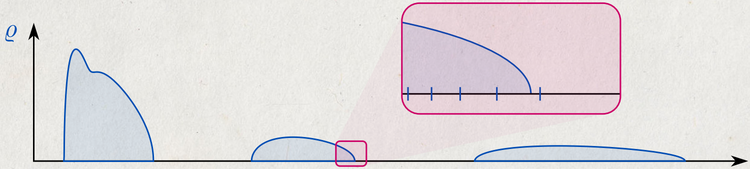


- The edge $a = f(x)$ is regular if $\exists \delta > 0$: $\min_i |x + 1/\sigma_i| \geqslant \delta$ and $\min_{b \in E \setminus \{a\}} |a - b| \geqslant \delta$. (El Karoui (2007), Hachem-Hardy-Najim (2014))
- The bulk component $[a, b]$ is regular if $\forall \varepsilon > 0\, \exists \delta > 0$: $\mathrm{d}\varrho(E)/\mathrm{d}E \geqslant \delta$ for $E \in [a+\varepsilon, b-\varepsilon]$.

The anisotropic local law holds in the vicinity of every regular edge, in every regular bulk component, and outside of the spectrum.

Application: edge universality

Theorem. [K-Yin (2014)]. The asymptotic joint eigenvalue distribution of any finite family of eigenvalues near the regular edges depends only on π. (Independent of distribution of X, left and right singular vectors of T, and dimensions of T.)



Combine with Hachem-Hardy-Najim (2014) for Gaussian case: Tracy-Widom-Airy-statistics, distribution of condition number, etc.

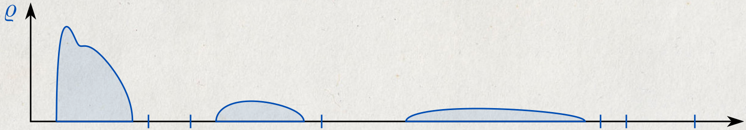
Previously: Tracy-Widom-Airy-statistics near top edge if $\mathbf{a}$ are uncorrelated. (El Karoui (2007), Onatski (2008), Bao-Pan-Zhou (2014), Lee-Schnelli (2014)).

Application: outliers

Suppose Σ has a finite number of spikes that violate the regularity condition. Treat them separately:

$$\Sigma = \Sigma_0 \left(I + VDV^* \right),$$

where Σ_0 satisfies the regularity condition and D is $r \times r$ with fixed $r \in \mathbb{N}$.



Deduce information about R using identities such as

$$V^* R V \; = \; \frac{1}{z} \left(D^{-1} - \frac{\sqrt{I+D}}{D} \frac{1}{D^{-1} + V^* R_0 V} \frac{\sqrt{I+D}}{D} \right).$$

Use that R_0 satisfies anisotropic local law. Allows a very precise analysis of eigenvalues and eigenvectors, also near BBP transition.

Prelude to the proof

For simplicity, let $Q = \Sigma^{1/2} X X^* \Sigma^{1/2}$, where $(X_{ik}) \in \mathbb{R}$ are independent and satisfy $\mathbb{E} X_{ik} = 0$ and $\mathbb{E} X_{ik}^2 = N^{-1}$.

Linearization. Replace $R(z)$ and $P(z)$ with the block matrices

$$G(z) := \begin{pmatrix} -\Sigma^{-1} & X \\ X^* & -zI \end{pmatrix}^{-1}, \quad \Pi(z) := \begin{pmatrix} -\Sigma(I + m(z)\Sigma)^{-1} & 0 \\ 0 & m(z) I \end{pmatrix}.$$

We estimate $G(z) - \Pi(z)$. Corollary: estimate of $R(z) - P(z)$ (from Schur's complement formula).

Three main steps to prove the anisotropic local law

$$\langle \mathbf{v}, G(z) \mathbf{w} \rangle = \langle \mathbf{v}, \Pi(z) \mathbf{w} \rangle + O_{\mathrm{HP}}(\Psi(z) |\mathbf{v}| |\mathbf{w}|).$$

- Step (A). X Gaussian; Σ diagonal; $\mathbf{v}, \mathbf{w}$ standard basis vectors. ("easy")
- Step (B). X Gaussian; Σ general; $\mathbf{v}, \mathbf{w}$ general. ("easy")
- Step (C). X general; Σ general; $\mathbf{v}, \mathbf{w}$ general. (hard)

Step (A)

Extension of previous works (Erdős-K-Yau-Yin (2013)). Main new observation: the equation $z = f(m)$ arises as a double application of Schur's complement formula. Formally, this is very easy to see.

Use notations $i \leqslant M$ and $\mu > M$. Let $G^{(\mu)}$ denote the matrix obtained from G by setting $X_{i\mu} = 0$ for all $i = 1, \dots, M$. Similarly, $G^{(i)}$ is obtained from G by setting $X_{i\mu} = 0$ for all $\mu = M+1, \dots, M+N$.

Using the approximation $G_{\mu\mu} \approx m$ we find

$$\frac{1}{m} \approx \frac{1}{G_{\mu\mu}} = -z - \left(X^* G^{(\mu)} X\right)_{\mu\mu} \approx -z - \frac{1}{N} \sum_i G_{ii}^{(\mu)} \approx -z - \frac{1}{N} \sum_i G_{ii} \,.$$

Moreover,

$$\frac{1}{G_{ii}} = -\frac{1}{\sigma_i} - (X G^{(i)} X^*)_{ii} \approx -\frac{1}{\sigma_i} - \frac{1}{N} \sum_\mu G_{\mu\mu} \approx -\frac{1}{\sigma_i} - m \,.$$

This yields $z = f(m)$.

The actual proof requires a control of errors in $\approx$. Stability of the resulting self-consistent equation follows from regularity assumptions.

Conclusion

- We prove the anisotropic local law $R(z) = P(z) + (\mathrm{error})$, where

$$R(z) := (Q - zI)^{-1}, \qquad P(z) := -\big(z(I + m(z)\Sigma)\big)^{-1}.$$

- Applications: complete delocalization of eigenvectors, edge universality, spikes and outliers.

- Similar results hold for deformed Wigner matrices $Q = W + A$, where $W = W^*$ is a Wigner matrix and $A = A^*$ is deterministic.

Coming up in Jun Yin's talk:

- Application: distribution of eigenvectors.

- Core of the proof: self-consistent comparison and Step (C).