Pseudospectra of structured random matrices

Oberwolfach, 2019/12/13

Nicholas Cook, Stanford University

Partly based on joint work with Alice Guionnet and Jonathan Husson

Outline

1. Geometric approach to RMT: successes and limitations
2. Spectral anti-concentration for structured Hermitian random matrices
3. Pseudospectra of i-non-id matrices
4. Pseudospectra and convergence to Brown measure for quadratic polynomials in Ginibre matrices (linearization $\rightsquigarrow$ pseudospectra for patterned block random matrices).

Geometric approach to RMT

Family of techniques originating from the local theory of Banach spaces (Grothendieck, Dvoretzky, Lindenstrauss, Milman, Schechtman ...).

Modern reference: Vershynin's text *High dimensional probability*.

Can often get quantitative bounds at finite N that are within a constant factor of the asymptotic truth, with arguments that are more flexible.

E.g. can show $\|X\|_{\mathrm{op}} = O(\sqrt{N})$ w.h.p. for X an iid matrix with sub-Gaussian entries with a simple net argument and concentration. Compare $\sim 2\sqrt{N}$ by the trace method.

* (For X GinOE can even get the right constant $\mathbb{E}\,\|X\|_{\mathrm{op}} \leq 2\sqrt{N}$ using Slepian's inequality!)

Net arguments and anti-concentration have been key for controlling the invertibility / condition number / pseudospectrum of random matrices.

Spectral anti-concentration

Consider H an $N \times N$ Hermitian random matrix with $\{H_{ij}\}_{i \le j}$ independent, centered, sub-Gaussian with variances $\sigma_{ij}^2 \in [0,1]$.

Denote by $\Sigma = (\sigma_{ij})_{i,j=1}^N$ the standard deviation profile.

How many eigenvalues can lie in an interval $\mathcal{I} \subset \mathbb{R}$? Under what conditions on Σ can we show

$$\mu_{\frac{1}{\sqrt{N}}H}(\mathcal{I}) \lesssim |\mathcal{I}| \qquad \forall \mathcal{I} \subset \mathbb{R},\ |\mathcal{I}| \ge N^{-1+\varepsilon}$$

with high probability (w.h.p.)?

Local semicircle law (Erdős–Schlein–Yau '08)

Suppose $\sigma_{ij} \equiv 1$. With high probability, for any interval $\mathcal{I} \subset \mathbb{R}$ with $|\mathcal{I}| \ge N^{-1+\varepsilon}$,

$$|\mu_{\frac{1}{\sqrt{N}}H}(\mathcal{I}) - \mu_{sc}(\mathcal{I})| = o(|\mathcal{I}|).$$

Extended to non-constant variance by Ajanki, Erdős & Krüger through careful analysis of associated vector Dyson equations.

Spectral anti-concentration

With $\operatorname{spt}(\Sigma) = \{(i,j) \in [N]^2 : \sigma_{ij} \geq \sigma_0\}$ (some fixed cutoff $\sigma_0 > 0$) say Σ is

- δ-**broadly connected** if $\forall I, J \subset [N]$ with $|I| + |J| \geq N$, $|\operatorname{spt}(\Sigma) \cap (I \times J)| \geq \delta |I||J|$ (Rudelson–Zeitouni '13);
- δ-**robustly irreducible** if $\forall J \subset [N]$, $|\operatorname{spt}(\Sigma) \cap (J \times J^c)| \geq \delta |J||J^c|$.

Robust irreducibility permits $\mu_{\frac{1}{\sqrt{N}}H}$ to have an atom at zero.

Theorem (C. '17, unpublished)

1. *Fix $\delta > 0$ and suppose Σ is δ-broadly connected. Then w.h.p., for any $\mathcal{I} \subset \mathbb{R}$ with $|\mathcal{I}| \geq C \frac{\log N}{N}$ we have $\mu_{\frac{1}{\sqrt{N}}H}(\mathcal{I}) \lesssim_\delta |\mathcal{I}|$.*
2. *Fix $\delta, \kappa > 0$ and suppose Σ is δ-robustly irreducible. Then w.h.p., for any $\mathcal{I} \subset \mathbb{R} \setminus (-\kappa, \kappa)$ with $|\mathcal{I}| \geq C \frac{\log N}{N}$ we have $\mu_{\frac{1}{\sqrt{N}}H}(\mathcal{I}) \lesssim_{\delta,\kappa} |\mathcal{I}|$.*

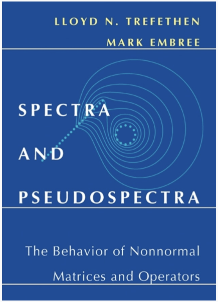
Related result of C.–Hachem–Najim–Renfrew '16 for deterministic equivalents.

Can reach intervals of length $N^{-1}\sqrt{\log N}$ using Bourgain–Tzafriri's restricted invertibility theorem as in independent work of Nguyen for case $\sigma_{ij} \equiv 1$.

Same strategy can be applied to e.g. $H_1 H_2 + H_2 H_1$ (local law by Anderson '15). Cf. Banna–Mai '18 on Hölder-regularity for distribution of NC-polynomials.

Pseudospectrum

(Already came up in talks of Capitaine, Fyodorov, Zeitouni and Vogel.)



For $A \in M_N(\mathbb{C})$, $\lambda \in \Lambda(A)$ is a qualitative statement.

More useful: For $\varepsilon > 0$ the ε-pseudospectrum is the set

$$\begin{aligned}\Lambda_\varepsilon(A) &= \Lambda(A) \cup \{z \in \mathbb{C} : \|(A-z)^{-1}\|_{\mathrm{op}} \geq 1/\varepsilon\} \\ &= \{z \in \mathbb{C} : \exists E \text{ with } \|E\|_{\mathrm{op}} \leq \varepsilon \text{ and } z \in \Lambda(A+E)\}.\end{aligned}$$

For A normal ($A^*A = AA^*$), $\Lambda_\varepsilon(A) = \Lambda(A) + \varepsilon\mathbb{D}$.
(We always have $\Lambda_\varepsilon(A) \supseteq \Lambda(A) + \varepsilon\mathbb{D}$.)

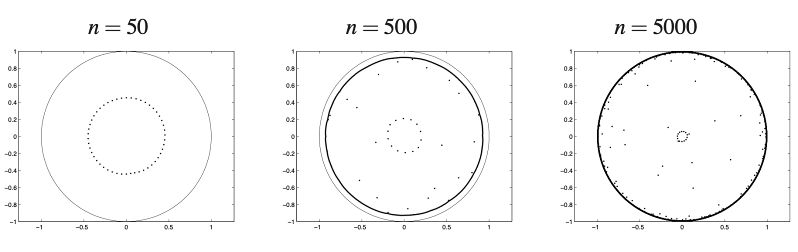
In particular, the spectrum of normal operators is stable: the spectrum is in a sense a 1-Lipschitz function of the matrix.

This can be extremely untrue for non-normal matrices!

The standard example: Left shift operator on $\mathbb{C}^N$

$$T_N = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ & & \cdots & & \\ 0 & & \cdots & & 1 \\ 0 & & \cdots & & 0 \end{pmatrix} \xrightarrow{*} \text{ Haar unitary element } u \in (\mathcal{A}, \tau).$$

ESDs $\equiv \delta_0$, $\Lambda_\varepsilon(T_N) \to \mathbb{D}$ for $\varepsilon = e^{-o(N)}$, Brown measure $=$ Unif$(\partial\mathbb{D})$.



Eigenvalues of $T_N + N^{-10} X_N$, with X_N GinOE. (Figure by Phil Wood.)

Pseudospectra of random matrices

Pseudospectrum of a **random** non-normal matrix is not so large.

For iid matrix X with sub-Gaussian entries,

$$\mathbb{P}\left\{z \in \Lambda_\varepsilon\left(\frac{1}{\sqrt{N}}X\right)\right\} = \mathbb{P}\left\{\left\|\left(\frac{1}{\sqrt{N}}X - z\right)^{-1}\right\|_{\mathrm{op}} \geq 1/\varepsilon\right\} \lesssim N\varepsilon + e^{-cN}$$

for any fixed $z \in \mathbb{C}$ ($\approx$ Rudelson–Vershynin '07).

In particular $\mathbb{E}\,\mathrm{Leb}(\Lambda_\varepsilon(\frac{1}{\sqrt{X}})) \lesssim N\varepsilon + e^{-cN}$.

Improves to $N^2\varepsilon^2$ for complex entries with independent real and imaginary parts [Luh '17] or real matrices with $\mathrm{dist}(z, \mathbb{R}) \gtrsim 1$ [Ge '17].

Compare deterministic bound $\mathrm{Leb}(\Lambda_\varepsilon(A)) \leq \pi N\varepsilon^2$ for normal matrices.

Pseudospectrum related to eigenvalue condition numbers (talk of Fyodorov):

$$\mathrm{Leb}(\Lambda_\varepsilon(M) \cap \Omega) \sim \pi\varepsilon^2 \sum_{j:\lambda_j \in \Omega} \kappa_j(M)^2 \qquad \text{as } \varepsilon \to 0.$$

Pseudospectra of structured random matrices

Applications to complex dynamical systems motivate understanding spectra and pseudospectra of sparse random matrices with non-iid entries (recall talk of David Renfrew).

Theorem (C. '16)

Let X have independent, centered entries of arbitrary variances $\sigma_{ij}^2 \in [0,1]$, $4+\varepsilon$ moments. For any $z \neq 0$,

$$\left\| \left(\frac{1}{\sqrt{N}} X - z \right)^{-1} \right\|_{\mathrm{op}} \leq N^{C(|z|,\varepsilon)} \quad \text{with probability } 1 - O(N^{-c(\varepsilon)}).$$

* $C(|z|,\varepsilon) = \mathrm{twr}(\exp(1/|z|^{O(1)}))$ ... Please improve!
* Conjecture: same holds with z replaced by any M with $s_{\min}(M) \gtrsim 1$.
* Assuming entries of bounded density, can improve probability bound to $1 - O(N^{-K})$ for arbitrary $K > 0$. Main difficulty is to allow $\sigma_{ij} = 0$.

This is a key ingredient for proof of the inhomogeneous circular law [C.–Hachem–Najim–Renfrew '16].

(Easier argument suffices for local law of [Alt–Erdős–Krüger '16] since they assume $\sigma_{ij} \gtrsim 1$ and bounded density. Cf. survey of Bordenave & Chafai '11.)

Pseudospectrum for quadratic polynomials in Ginibre matrices

Now let X denote a (complex) $N \times N$ Ginibre matrix having iid entries $X_{ij} \sim N_{\mathbb{C}}(0, 1/N)$.

Theorem (C.–Guionnet–Husson '19)

Let $m \geq 1$ and let p be a quadratic polynomial in non-commutative variables $x_1, \ldots, x_m$. Let $N \geq 2$ and $X_1, \ldots, X_m$ be iid $N \times N$ Ginibre matrices. Set $P = p(X_1, \ldots, X_m)$. For any $z \in \mathbb{C}$ and any $\varepsilon > 0$,

$$\mathbb{P}\{z \in \Lambda_\varepsilon(P)\} = \mathbb{P}\left\{ \|(P - z)^{-1}\|_{\mathrm{op}} \geq \frac{1}{\varepsilon} \right\} \leq N^C \varepsilon^c + e^{-cN}$$

for constants $C, c > 0$ depending only on p.

Motivation: convergence of ESDs

- Proofs of limits for the ESDs $\mu_X := \frac{1}{N}\sum_{j=1}^N \delta_{\lambda_j(X)}$ of non-normal random matrices $X = X_N$ hinge upon control of the pseudospectrum.

  In particular, the problem of the pseudospectrum is the reason non-Hermitian RMT has lagged behind the theory for Wigner matrices.

∗ A key idea: Hermitization

- Hermitian polynomials – some highlights:
  - Haagerup–Thorbjørnsen '05: No outliers (recent alternative proof by Collins–Guionnet–Parraud). Extensions by many authors.
  - Anderson '15: local law for the anti-commutator $H_1H_2 + H_2H_1$ of independent Wigner matrices.
  - Erdős–Krüger–Nemish '18: local law for polynomials satisfying a technical condition (includes homogeneous quadratic polynomials and symmetrized monomials in iid matrices $X_1X_2\cdots X_mX_m^*\cdots X_2X_1$).
- Products of independent iid matrices: limiting ESDs (Götze–Tikhomirov and O'Rourke–Soshnikov '10). No outliers and local law (Nemish '16, '17).

∗ A key idea: Linearization

Hermitization

- One can encode the ESD of a non-normal $M \in M_N(\mathbb{C})$ in a family of ESDs of Hermitian matrices $|M - z| = \sqrt{(M-z)^*(M-z)}$ as follows:

$$\mu_M = \frac{1}{N}\sum_{j=1}^{N} \delta_{\lambda_j(M)} = \frac{1}{2\pi}\Delta_z \int_0^{\infty} \log(s)\mu_{|M-z|}(ds).$$

- So it seems we can recover limit of μ_{X_N} if we know the limits of ESDs of the family of Hermitian matrices $\{|X_N - z|\}_{z\in\mathbb{C}}$.
- But not quite! Possible escape of mass to zero: Pseudospectrum
- Bai '97 controlled the pseudospectrum of iid matrices (under some technical assumptions) and obtained the Circular Law.

  Assumptions relaxed in works of Götze–Tikhomirov '07, Pan–Zhou '07, Tao–Vu '07, '08.

Brown measure

- Free probability gives tools to calculate limiting ESDs for polynomials in independent random matrices, at least if they're normal (e.g. $XY + YX$ for X, Y iid Wigner).
- For a normal element a of a non-commutative probability space $(\mathcal{A}, \tau)$, the spectral theorem provides us with a spectral measure μ_a determined by the $*$-moments $\tau(a^k (a^*)^l)$.
- For general (non-normal) elements a, can define the *Brown measure*:

$$\nu_a := \frac{1}{2\pi} \Delta_z \int_0^\infty \log(s) \mu_{|a-z|}(ds)$$

  which is determined by the $*$-moments ($|a - z|$ is self-adjoint).
- If A_N converge in $*$-moments to a, it doesn't follow that μ_{A_N} converge weakly to ν_a (Brown measure isn't continuous in this topology).
- Question: If A_N are non-normal *random* matrices, do the ESDs converge to the Brown measure? (Answer is yes for single iid matrix X_N.)

Convergence to the Brown measure for polynomials

Theorem (C.–Guionnet–Husson '19)

Let $m \geq 1$ and let p be a quadratic polynomial in non-commutative variables $x_1, \ldots, x_m$. For each N let $X_1^{(N)}, \ldots, X_m^{(N)}$ be iid $N \times N$ Ginibre matrices. Set $P^{(N)} = p(X_1^{(N)}, \ldots, X_m^{(N)})$. Almost surely,

$$\mu_{P^{(N)}} \to \nu_p \quad \text{weakly},$$

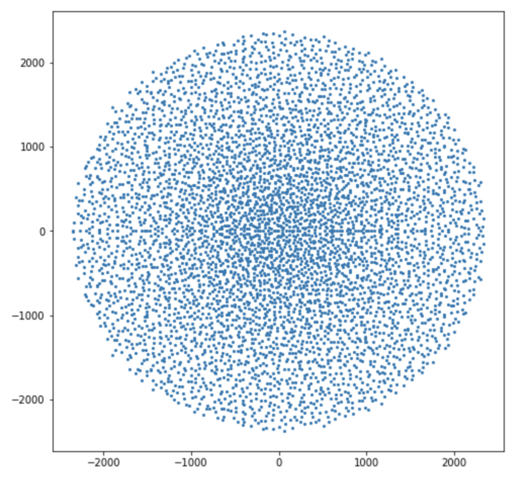
where ν_p is the Brown measure for $p(c_1, \ldots, c_m)$ with $c_1, \ldots, c_m$ free circular elements of a non-commutative probability space.

Partially answers a question raised in talk of Mireille Capitaine.

ν_p can be recovered from solution of an associated (matrix-valued) Schwinger–Dyson equation.

Hard to solve by hand! Numerics: ν_p has a "volcano" shape.

Simulation: $XY + YX$



$N = 5000$, entries Uniform $\in [-1, 1]$.

Pseudospectrum of $XY + YX$, Step 1: Linearization

- We'll illustrate ideas for the anti-commutator $P = XY + YX$ of independent Ginibre matrices.
- To control the pseudospectrum of P we need to bound $\|(P-z)^{-1}\|_{\mathrm{op}}$. Entries of P are highly correlated with complicated distribution, so previous approaches (Tao–Vu, Rudelson–Vershynin) don't apply.
- From the Schur complement formula, $(P-z)^{-1}$ is the top left block of $\boldsymbol{L}^{-1}$, where $\boldsymbol{L}$ is the $3N \times 3N$ linearized matrix

$$\boldsymbol{L} = \begin{pmatrix} -z & X & Y \\ Y & -I & 0 \\ X & 0 & -I \end{pmatrix}.$$

- So we've reduced to bounding $\|\boldsymbol{L}^{-1}\|_{\mathrm{op}}$, where we can view $\boldsymbol{L}$ as an $N \times N$ matrix with independent entries $L_{ij} \in M_3(\mathbb{C})$.

Pseudospectrum of $XY + YX$, Step 2: dimension reduction

- $\boldsymbol{L}$ is poorly-invertible (ill-conditioned) if one of its columns is close to the span of the remaining columns.
- **Reduction to bounded dimension:** Let $\hat{L}_j$ denote the projection of the jth column $\boldsymbol{L}_j = (L_{ij})_{i=1}^N \in M_3(\mathbb{C})^N$ to the span of the remaining $3N - 3$ columns. Can reduce our task to showing

$$\mathbb{P}\{\|(\hat{L}_1)^{-1}\|_{\mathrm{op}} \geq 1/\varepsilon\} \leq N^C \varepsilon^c + e^{-cN}.$$

- **Reduction to scalar anti-concentration:** We want to show $\hat{L}_1$ is well invertible. Giving up some powers of N, it's enough to show

$$\mathbb{P}\{|\det(\hat{L}_1)| \leq \varepsilon\} \leq N^{C'} \varepsilon^c + e^{-cN}.$$

  After conditioning on columns $\{\boldsymbol{L}_j\}_{j=2}^N$, $\det(\hat{L}_1)$ is a bounded-degree polynomial in the $2N$ independent Gaussian entries of $\boldsymbol{L}_1$.

Pseudospectrum of $XY + YX$, Step 3: anticoncentration

- Off-the-shelf anti-concentration (Carbery–Wright inequality): If f is a degree-d polynomial in iid Gaussian variables $g = (g_1, \dots, g_n)$, then

$$\sup_{t\in\mathbb{R}} \mathbb{P}\left\{ |f(g) - t| \le \varepsilon \sqrt{\operatorname{Var} f(g)} \right\} \lesssim_d \varepsilon^{1/d}.$$

So it's enough to show

$$\operatorname{Var}\big(\det(\hat{L}_1) \mid \{\boldsymbol{L}_j\}_{j=2}^N \big) \ge N^{-O(1)} \tag{1}$$

with high probability.

- Express $\hat{L}_1 = \boldsymbol{U}^* \boldsymbol{L}_1 = \sum_{i=1}^N U_i^* L_{i1}$ where $\boldsymbol{U} = (\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3)$ orthonormal in the orthocomplement of $\operatorname{Span}(\{\boldsymbol{L}_j\}_{j=2}^N)$.

  Expanding in the Gaussian variables X_{i1}, Y_{i1} and inspecting coefficients of highest degree (degree 3 in this case), one sees we get (1) unless $\boldsymbol{U}$ has a lot of geometric structure in its rows.

- Set of orthonormal bases with such structure has low metric entropy, so we can rule out such $\boldsymbol{U}$ using a net argument.

Further directions

- Higher degree polynomials, including deterministic matrices (as in Capitaine's talk)?
- General entry distributions?
- $*$-polynomials? (Includes polynomials in GUE matrices $\frac{1}{\sqrt{2}}(X + X^*)$.)
- Rational functions?

Thank you!