Threshold phenomena for quantum marginals

Stanislaw Szarek

Case Western Reserve/Paris 6

Cambridge, October 15, 2013

Collaborators: G. Aubrun, D. Ye

Comm. Pure Appl. Math. (2014), arXiv:1106.2264v3
Phys. Rev. A. 85(R) (2012), arXiv:1112.4582v2

http://www.cwru.edu/artsci/math/szarek/

Abstract

Consider a quantum system consisting of N identical particles and assume that it is in a random pure state (i.e., uniformly distributed over the sphere of the corresponding Hilbert space) and two subsystems $\mathcal{A}$ and $\mathcal{B}$ consisting of k particles each. Are $\mathcal{A}$ and $\mathcal{B}$ likely to share entanglement? Is the $\mathcal{AB}$ -marginal typically PPT?

For many natural properties there is a sharp "phase transition." E.g., there is a threshold $K \sim N/5$ such that - if $k > K$, then $\mathcal{A}$ and $\mathcal{B}$ typically share entanglement - if $k < K$, then $\mathcal{A}$ and $\mathcal{B}$ typically do not share entanglement.

The first statement was (essentially) shown in the talk by G. Aubrun. Here we present a general scheme for handling such questions and sketch the analysis specific to entanglement.

Talk summary

• setup, notation; random quantum states and ensembles of random matrices

• emergence of entanglement (the main result)

• a sketch of the proof using the tools of geometric functional analysis and random matrix theory

Sets of states

The set of (mixed) states on $\mathcal{H}$ is denoted by $\mathcal{D} = \mathcal{D}(\mathcal{H})$

Separable (unentangled) states on $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$: $\mathcal{S} = \mathcal{S}(\mathcal{H})$

When we talk about $\mathcal{S}(\mathbb{C}^n)$, we implicitly assume $n = d^2$ and $\mathbb{C}^n \sim \mathbb{C}^d \otimes \mathbb{C}^d$ (a fixed bipartition).

Sets of states

The set of (mixed) states on $\mathcal{H}$ is denoted by $\mathcal{D} = \mathcal{D}(\mathcal{H})$

Separable (unentangled) states on $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$: $\mathcal{S} = \mathcal{S}(\mathcal{H})$

When we talk about $\mathcal{S}(\mathbb{C}^n)$, we implicitly assume $n = d^2$ and $\mathbb{C}^n \sim \mathbb{C}^d \otimes \mathbb{C}^d$ (a fixed bipartition).

The sets $\mathcal{D}(\mathbb{C}^n)$ and $\mathcal{S}(\mathbb{C}^n)$ are convex bodies in the hyperplane $H_1 := \{\mathrm{tr}(\cdot) = 1\} \subset \mathcal{M}_n^{sa}$ with $\frac{1}{n}\mathbb{I}$ in the interior. $\frac{1}{n}\mathbb{I}$ is the only point invariant under the isometries of either set.

Sets of states

The set of (mixed) states on $\mathcal{H}$ is denoted by $\mathcal{D} = \mathcal{D}(\mathcal{H})$

Separable (unentangled) states on $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$: $\mathcal{S} = \mathcal{S}(\mathcal{H})$

When we talk about $\mathcal{S}(\mathbb{C}^n)$, we implicitly assume $n = d^2$ and $\mathbb{C}^n \sim \mathbb{C}^d \otimes \mathbb{C}^d$ (a fixed bipartition).

The sets $\mathcal{D}(\mathbb{C}^n)$ and $\mathcal{S}(\mathbb{C}^n)$ are convex bodies in the hyperplane $H_1 := \{\mathrm{tr}(\cdot) = 1\} \subset \mathcal{M}_n^{sa}$ with $\frac{1}{n}\mathbb{I}$ in the interior. $\frac{1}{n}\mathbb{I}$ is the only point invariant under the isometries of either set.

Our approach will (in principle) work for any property in place of $d \times d$ separability, provided it has some minimal permanence properties.

Setup, notation; partial trace

The state of the entire system is described by

$$|\psi\rangle \in \mathcal{H} = \mathcal{A} \otimes \mathcal{B} \otimes \mathcal{E} = (\mathbb{C}^D)^{\otimes k} \otimes (\mathbb{C}^D)^{\otimes k} \otimes (\mathbb{C}^D)^{\otimes N-2k}$$

$$= \mathbb{C}^d \otimes \mathbb{C}^d \otimes \mathbb{C}^s$$

with $d = D^k$, $s = D^{N-2k}$

If $N = 5k$, then $s = D^{N-2k} = D^{3k} = d^3$

Setup, notation; partial trace

The state of the entire system is described by

$$
\begin{aligned}
|\psi\rangle \in \mathcal{H} = \mathcal{A} \otimes \mathcal{B} \otimes \mathcal{E} &= (\mathbb{C}^D)^{\otimes k} \otimes (\mathbb{C}^D)^{\otimes k} \otimes (\mathbb{C}^D)^{\otimes N-2k} \\
&= \mathbb{C}^d \otimes \mathbb{C}^d \otimes \mathbb{C}^s \\
&\text{with } d = D^k,\ s = D^{N-2k} \\
&\text{If } N = 5k, \text{ then } s = D^{N-2k} = D^{3k} = d^3
\end{aligned}
$$

If the entire system is in the pure state $|\psi\rangle \in \mathcal{H}$ or $|\psi\rangle\langle\psi|$ in the density matrix formalism, then the $\mathcal{AB}$-marginal is given by the partial trace $\mathrm{tr}_{\mathcal{E}} \ |\psi\rangle\langle\psi| = \mathrm{tr}_{\mathbb{C}^s} \ |\psi\rangle\langle\psi|$.

Setup, notation; partial trace

The state of the entire system is described by

$$|\psi\rangle \in \mathcal{H} = \mathcal{A} \otimes \mathcal{B} \otimes \mathcal{E} = (\mathbb{C}^D)^{\otimes k} \otimes (\mathbb{C}^D)^{\otimes k} \otimes (\mathbb{C}^D)^{\otimes N-2k}$$
$$= \mathbb{C}^d \otimes \mathbb{C}^d \otimes \mathbb{C}^s$$

with $d = D^k$, $s = D^{N-2k}$

If $N = 5k$, then $s = D^{N-2k} = D^{3k} = d^3$

If the entire system is in the pure state $|\psi\rangle \in \mathcal{H}$ or $|\psi\rangle\langle\psi|$ in the density matrix formalism, then the $\mathcal{AB}$-marginal is given by the partial trace $\mathrm{tr}_{\mathcal{E}} \, |\psi\rangle\langle\psi| = \mathrm{tr}_{\mathbb{C}^s} \, |\psi\rangle\langle\psi|$.

If $|\psi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d \otimes \mathbb{C}^s \sim \mathbb{C}^n \otimes \mathbb{C}^s$ is identified with a matrix $A \in \mathcal{M}_{n\times s}$, then $\rho = AA^\dagger$.

Random marginals and “standard” ensembles of random matrices

Random marginals and "standard" ensembles of random matrices

- If $|\psi\rangle$ is random on $\mathbb{C}^n \otimes \mathbb{C}^s$, then $\mathrm{tr}_{\mathbb{C}^s} |\psi\rangle\langle\psi|$ is a random state on $\mathbb{C}^n$

Random marginals and "standard" ensembles of random matrices

• If $|\psi\rangle$ is random on $\mathbb{C}^n \otimes \mathbb{C}^s$, then $\mathrm{tr}_{\mathbb{C}^s}\ |\psi\rangle\langle\psi|$ is a random state on $\mathbb{C}^n$

• If $|\psi\rangle$ is distributed uniformly on the sphere of $\mathbb{C}^n \otimes \mathbb{C}^s$, then A is uniform on the Hilbert-Schmidt/Frobenius sphere in $\mathcal{M}_{n\times s}$.

Random marginals and "standard" ensembles of random matrices

• If $|\psi\rangle$ is random on $\mathbb{C}^n \otimes \mathbb{C}^s$, then $\mathrm{tr}_{\mathbb{C}^s} |\psi\rangle\langle\psi|$ is a random state on $\mathbb{C}^n$

• If $|\psi\rangle$ is distributed uniformly on the sphere of $\mathbb{C}^n \otimes \mathbb{C}^s$, then A is uniform on the Hilbert-Schmidt/Frobenius sphere in $\mathcal{M}_{n\times s}$.

Either way, $\rho_{n,s} := \mathrm{tr}_{\mathbb{C}^s} |\psi\rangle\langle\psi| \sim AA^\dagger$ is a random state on $\mathbb{C}^n$ and s is a parameter.

Random marginals and "standard" ensembles of random matrices

- If $|\psi\rangle$ is random on $\mathbb{C}^n \otimes \mathbb{C}^s$, then $\mathrm{tr}_{\mathbb{C}^s} |\psi\rangle\langle\psi|$ is a random state on $\mathbb{C}^n$
- If $|\psi\rangle$ is distributed uniformly on the sphere of $\mathbb{C}^n \otimes \mathbb{C}^s$, then A is uniform on the Hilbert-Schmidt/Frobenius sphere in $\mathcal{M}_{n\times s}$.

Either way, $\rho_{n,s} := \mathrm{tr}_{\mathbb{C}^s} |\psi\rangle\langle\psi| \sim AA^\dagger$ is a random state on $\mathbb{C}^n$ and s is a parameter.

Another way to represent such random states on $\mathbb{C}^n$ is to consider

$$\rho_{n,s} = \frac{BB^\dagger}{\mathrm{tr}BB^\dagger},$$

where B is the standard $n\times s$ complex Gaussian matrix

Random marginals and "standard" ensembles of random matrices

• If $|\psi\rangle$ is random on $\mathbb{C}^n \otimes \mathbb{C}^s$, then $\mathrm{tr}_{\mathbb{C}^s}\, |\psi\rangle\langle\psi|$ is a random state on $\mathbb{C}^n$

• If $|\psi\rangle$ is distributed uniformly on the sphere of $\mathbb{C}^n \otimes \mathbb{C}^s$, then A is uniform on the Hilbert-Schmidt/Frobenius sphere in $\mathcal{M}_{n\times s}$.

Either way, $\rho_{n,s} := \mathrm{tr}_{\mathbb{C}^s}\, |\psi\rangle\langle\psi| \sim AA^\dagger$ is a random state on $\mathbb{C}^n$ and s is a parameter.

Another way to represent such random states on $\mathbb{C}^n$ is to consider

$$\rho_{n,s} = \frac{BB^\dagger}{\mathrm{tr}BB^\dagger},$$

where B is the standard $n\times s$ complex Gaussian matrix; so $\rho_{n,s}$ is a normalized complex Wishart matrix.

Entanglement emergence for $\rho_{n,s}$, $\mathbb{C}^n \sim \mathbb{C}^d \otimes \mathbb{C}^d$

Theorem There is a sharp "entanglement emergence" threshold $s_{ent} = s_{ent}(d)$ for $\rho_{n,s}$ verifying $cd^3 \leq s_{ent} \leq Cd^3 \log^2 d$

Entanglement emergence for $\rho_{n,s}$, $\mathbb{C}^n \sim \mathbb{C}^d \otimes \mathbb{C}^d$

Theorem There is a sharp "entanglement emergence" threshold $s_{ent} = s_{ent}(d)$ for $\rho_{n,s}$ verifying $cd^3 \leq s_{ent} \leq Cd^3 \log^2 d$, where $C, c > 0$ are universal constants.

Entanglement emergence for $\rho_{n,s}$, $\mathbb{C}^n \sim \mathbb{C}^d \otimes \mathbb{C}^d$

Theorem There is a sharp "entanglement emergence" threshold $s_{ent} = s_{ent}(d)$ for $\rho_{n,s}$ verifying $cd^3 \leq s_{ent} \leq Cd^3 \log^2 d$, where $C, c > 0$ are universal constants.

More precisely, if $\varepsilon > 0$, then

If $s \leq (1-\varepsilon)s_{ent}(d)$, then $\mathbb{P}(\rho_{n,s} \text{ is separable}) \leq 2\exp(-c(\varepsilon)s_{ent})$

If $s \geq (1+\varepsilon)s_{ent}(d)$, then $\mathbb{P}(\rho_{n,s} \text{ is entangled}) \leq 2\exp(-c(\varepsilon)s)$

Entanglement emergence for $\rho_{n,s}$, $\mathbb{C}^n \sim \mathbb{C}^d \otimes \mathbb{C}^d$

Theorem There is a sharp "entanglement emergence" threshold $s_{ent} = s_{ent}(d)$ for $\rho_{n,s}$ verifying $cd^3 \leq s_{ent} \leq Cd^3 \log^2 d$, where $C, c > 0$ are universal constants.

More precisely, if $\varepsilon > 0$, then

If $s \leq (1-\varepsilon)s_{ent}(d)$, then $\mathbb{P}(\rho_{n,s} \text{ is separable}) \leq 2\exp(-c(\varepsilon)s_{ent})$

If $s \geq (1+\varepsilon)s_{ent}(d)$, then $\mathbb{P}(\rho_{n,s} \text{ is entangled}) \leq 2\exp(-c(\varepsilon)s)$

Corollary The statement in the abstract

Entanglement emergence for $\rho_{n,s}$, $\mathbb{C}^n \sim \mathbb{C}^d \otimes \mathbb{C}^d$

Theorem There is a sharp "entanglement emergence" threshold $s_{ent} = s_{ent}(d)$ for $\rho_{n,s}$ verifying $cd^3 \leq s_{ent} \leq Cd^3 \log^2 d$, where $C, c > 0$ are universal constants.

More precisely, if $\varepsilon > 0$, then

If $s \leq (1-\varepsilon)s_{ent}(d)$, then $\mathbb{P}(\rho_{n,s} \text{ is separable}) \leq 2\exp(-c(\varepsilon)s_{ent})$

If $s \geq (1+\varepsilon)s_{ent}(d)$, then $\mathbb{P}(\rho_{n,s} \text{ is entangled}) \leq 2\exp(-c(\varepsilon)s)$

Corollary The statement in the abstract

Earlier results in various directions/cases were in Kendon–Życzkowski–Munro 2002, Hayden–Leung–Winter 2006, Aubrun-S. 2006, Ye 2010

Why the sharp threshold?

Why the sharp threshold?

Replacing k by $k-1$ results in replacing

$d = D^k$ with $d_1 = D^{k-1} = \frac{d}{D} \leq \frac{d}{2}$ and

$s = D^{N-2k}$ with $s_1 = D^{N-2(k-1)} = D^2 s \geq 4s$

Why the sharp threshold?

Replacing k by $k-1$ results in replacing

$$d = D^k \text{ with } d_1 = D^{k-1} = \frac{d}{D} \leq \frac{d}{2} \qquad \text{and}$$

$$s = D^{N-2k} \text{ with } s_1 = D^{N-2(k-1)} = D^2 s \geq 4s$$

and it takes only an increase in s by a factor $\frac{1+\varepsilon}{1-\varepsilon}$ to switch from "generic entanglement" to "generic separability."

Why the sharp threshold?

Replacing k by $k-1$ results in replacing

$$d = D^k \text{ with } d_1 = D^{k-1} = \frac{d}{D} \leq \frac{d}{2} \qquad \text{and}$$

$$s = D^{N-2k} \text{ with } s_1 = D^{N-2(k-1)} = D^2 s \geq 4s$$

and it takes only an increase in s by a factor $\frac{1+\varepsilon}{1-\varepsilon}$ to switch from "generic entanglement" to "generic separability."

Remark $\pi_{d,s} := \mathbb{P}(\mu_{d^2,s} \text{is separable})$ is a decreasing function of d.

Why the sharp threshold?

Replacing k by $k-1$ results in replacing

$$d = D^k \text{ with } d_1 = D^{k-1} = \frac{d}{D} \leq \frac{d}{2} \qquad \text{and}$$

$$s = D^{N-2k} \text{ with } s_1 = D^{N-2(k-1)} = D^2 s \geq 4s$$

and it takes only an increase in s by a factor $\frac{1+\varepsilon}{1-\varepsilon}$ to switch from "generic entanglement" to "generic separability."

Remark $\pi_{d,s} := \mathbb{P}(\mu_{d^2,s} \text{is separable})$ is a decreasing function of d.

Problem Is $\pi_{d,s}$ an increasing function of s?

Rephrasing the entanglement property

$\frac{\mathbb{I}}{n} \in \mathcal{S} \subset \mathcal{D} \subset H_1 := \{\mathrm{tr}(\cdot) = 1\} \subset \mathcal{M}_n^{sa}$

Rephrasing the entanglement property

$$\frac{\mathbb{I}}{n} \in \mathcal{S} \subset \mathcal{D} \subset H_1 := \{\mathrm{tr}(\cdot) = 1\} \subset \mathcal{M}_n^{sa}$$

Translate everything to the origin:

$$0 \in \mathcal{S}_0 := \mathcal{S} - \frac{\mathbb{I}}{n} \subset \mathcal{D}_0 \ldots \subset H_0 := \{\mathrm{tr}(\cdot) = 0\} \subset \mathcal{M}_n^{sa}$$

Rephrasing the entanglement property

$$\frac{\mathbb{I}}{n} \in \mathcal{S} \subset \mathcal{D} \subset H_1 := \{\mathrm{tr}(\cdot) = 1\} \subset \mathcal{M}_n^{sa}$$

Translate everything to the origin:

$$0 \in \mathcal{S}_0 := \mathcal{S} - \frac{\mathbb{I}}{n} \subset \mathcal{D}_0 \ldots \subset H_0 := \{\mathrm{tr}(\cdot) = 0\} \subset \mathcal{M}_n^{sa}$$

Then

$$\rho \ \text{ is unentangled } \iff \rho - \frac{\mathbb{I}}{n} \in \mathcal{S}_0 \iff \left\| \rho - \frac{\mathbb{I}}{n} \right\|_{\mathcal{S}_0} \leq 1$$

where $\|x\|_K := \min\{t \geq 0 \ : \ x \in tK\}$ is the gauge of a convex set K

Rephrasing the entanglement property

$$\frac{\mathbb{I}}{n} \in \mathcal{S} \subset \mathcal{D} \subset H_1 := \{\mathrm{tr}(\cdot) = 1\} \subset \mathcal{M}_n^{sa}$$

Translate everything to the origin:

$$0 \in \mathcal{S}_0 := \mathcal{S} - \frac{\mathbb{I}}{n} \subset \mathcal{D}_0 \ldots \subset H_0 := \{\mathrm{tr}(\cdot) = 0\} \subset \mathcal{M}_n^{sa}$$

Then

$$\rho \quad \text{is unentangled} \quad \iff \rho - \frac{\mathbb{I}}{n} \in \mathcal{S}_0 \iff \left\| \rho - \frac{\mathbb{I}}{n} \right\|_{\mathcal{S}_0} \leq 1$$

where $\|x\|_K := \min\{t \geq 0 \ : \ x \in tK\}$ is the gauge of a convex set K ($0 \in$ the interior of K)

Rephrasing the entanglement property

$$\frac{\mathbb{I}}{n} \in \mathcal{S} \subset \mathcal{D} \subset H_1 := \{\mathrm{tr}(\cdot) = 1\} \subset \mathcal{M}_n^{sa}$$

Translate everything to the origin:

$$0 \in \mathcal{S}_0 := \mathcal{S} - \frac{\mathbb{I}}{n} \subset \mathcal{D}_0 \ldots \subset H_0 := \{\mathrm{tr}(\cdot) = 0\} \subset \mathcal{M}_n^{sa}$$

Then

$$\rho \;\text{ is unentangled } \iff \rho - \frac{\mathbb{I}}{n} \in \mathcal{S}_0 \iff \left\| \rho - \frac{\mathbb{I}}{n} \right\|_{\mathcal{S}_0} \leq 1$$

where $\|x\|_K := \min\{t \geq 0 \;:\; x \in tK\}$ is the gauge of a convex set K ($0 \in$ the interior of K) and of course

$$\rho \text{ is entangled } \iff \left\| \rho - \frac{\mathbb{I}}{n} \right\|_{\mathcal{S}_0} > 1$$

Rephrasing the entanglement property

$$\frac{\mathbb{I}}{n} \in \mathcal{S} \subset \mathcal{D} \subset H_1 := \{\mathrm{tr}(\cdot) = 1\} \subset \mathcal{M}_n^{sa}$$

Translate everything to the origin:

$$0 \in \mathcal{S}_0 := \mathcal{S} - \frac{\mathbb{I}}{n} \subset \mathcal{D}_0 \ldots \subset H_0 := \{\mathrm{tr}(\cdot) = 0\} \subset \mathcal{M}_n^{sa}$$

Then

$$\rho \;\; \text{is unentangled} \iff \rho - \frac{\mathbb{I}}{n} \in \mathcal{S}_0 \iff \left\| \rho - \frac{\mathbb{I}}{n} \right\|_{\mathcal{S}_0} \leq 1$$

where $\|x\|_K := \min\{t \geq 0 \;:\; x \in tK\}$ is the gauge of a convex set K ($0 \in$ the interior of K) and of course

$$\rho \text{ is entangled} \iff \left\| \rho - \frac{\mathbb{I}}{n} \right\|_{\mathcal{S}_0} > 1$$

Need to show that, for appropriate values of n, s and for $\rho = \rho_{n,s}$, the above occur with probability close to 1 or 0.

The strategy : concentration

Step 1: Show that, for the appropriate values of n, s and $\varepsilon > 0$,

$\mathbb{E}\|\rho_{n,s} - \frac{\mathbb{I}}{n}\|_{\mathcal{S}_0} \leq 1 - \varepsilon$ or $\mathbb{E}\|\rho_{n,s} - \frac{\mathbb{I}}{n}\|_{\mathcal{S}_0} \geq 1 + \varepsilon$, as needed.

Step 2: Show that $A \to f(A) := \|AA^\dagger - \frac{\mathbb{I}}{n}\|_{\mathcal{S}_0}$ is smooth enough and so it concentrates around its median (and its mean).

Recall that A varies over the Frobenius sphere in $\mathcal{M}_{n\times s}$; the relevant metric will also given by the Frobenius norm $\|\cdot\|_2$.

Step 2 is nontrivial, but relatively routine.

Step 1 is harder and requires a few new tricks.

Step 2 : smoothness and concentration

$A \to f(A) = \|AA^\dagger - \frac{\mathbb{I}}{n}\|_{\mathcal{S}_0}$ is a composition of two operations

Step 2 : smoothness and concentration

$A \to f(A) = \|AA^\dagger - \frac{\mathbb{I}}{n}\|_{\mathcal{S}_0}$ is a composition of two operations

- $A \to AA^\dagger - \frac{\mathbb{I}}{n}, \quad \|\cdot\|_2 \to \|\cdot\|_2$

Step 2 : smoothness and concentration

$A \to f(A) = \|AA^\dagger - \frac{\mathbb{I}}{n}\|_{\mathcal{S}_0}$ is a composition of two operations

- $A \to AA^\dagger - \frac{\mathbb{I}}{n}$, $\quad \|\cdot\|_2 \to \|\cdot\|_2$
- $X \to \|X\|_{\mathcal{S}_0}$, $\quad$ with $\|\cdot\|_2$ in the domain

Step 2 : smoothness and concentration

$A \to f(A) = \|AA^\dagger - \frac{\mathbb{I}}{n}\|_{\mathcal{S}_0}$ is a composition of two operations

- $A \to AA^\dagger - \frac{\mathbb{I}}{n}$, $\quad \|\cdot\|_2 \to \|\cdot\|_2$
- $X \to \|X\|_{\mathcal{S}_0}$, $\quad$ with $\|\cdot\|_2$ in the domain

The latter is a Lipschitz function with constant r^{-1}, where r is the (Euclidean/Frobenius) inradius of $\mathcal{S}_0$, which is known to be the same as the inradius of $\mathcal{D}$ (Gurvits-Barnum 2002), which is $1/\sqrt{n(n-1)} \sim 1/n$.

Step 2 : smoothness and concentration

$A \to f(A) = \|AA^\dagger - \frac{\mathbb{I}}{n}\|_{\mathcal{S}_0}$ is a composition of two operations

- $A \to AA^\dagger - \frac{\mathbb{I}}{n}, \quad \|\cdot\|_2 \to \|\cdot\|_2$
- $X \to \|X\|_{\mathcal{S}_0}, \quad$ with $\|\cdot\|_2$ in the domain

The latter is a Lipschitz function with constant r^{-1}, where r is the (Euclidean/Frobenius) inradius of $\mathcal{S}_0$, which is known to be the same as the inradius of $\mathcal{D}$ (Gurvits-Barnum 2002), which is $1/\sqrt{n(n-1)} \sim 1/n$.

So the Lipschitz constant is $\sim n$, in fact $< n$.

Smoothness of $A \to AA^\dagger - \frac{\mathbb{I}}{n}$, $\|\cdot\|_2 \to \|\cdot\|_2$

Smoothness of $A \to AA^\dagger - \frac{\mathbb{I}}{n}$, $\quad \|\cdot\|_2 \to \|\cdot\|_2$

This operation is not Lipschitz, but its restriction to the Hilbert-Schmidt sphere is 2-Lipschitz, so f is $2n$-Lipschitz.

Smoothness of $A \to AA^\dagger - \frac{\mathbb{I}}{n}$, $\quad \|\cdot\|_2 \to \|\cdot\|_2$

This operation is not Lipschitz, but its restriction to the Hilbert-Schmidt sphere is 2-Lipschitz, so f is $2n$-Lipschitz.

We can now appeal to **Levy's lemma**:

Let $f : S^{m-1} \to \mathbb{R}$ be an L-Lipschitz function and let $\varepsilon > 0$. Then

$$\mathbb{P}(\{|f - M_f| > \varepsilon\}) \leq C \exp(-\tfrac{m}{2L^2}\,\varepsilon^2),$$

where M_f is the median (or mean) of f, $\mathbb{P}$ is the normalized uniform measure on the sphere and $C > 0$ is a universal constant.

Smoothness of $A \to AA^\dagger - \frac{\mathbb{I}}{n}$, $\quad \| \cdot \|_2 \to \| \cdot \|_2$

This operation is not Lipschitz, but its restriction to the Hilbert-Schmidt sphere is 2-Lipschitz, so f is $2n$-Lipschitz.

We can now appeal to **Levy's lemma**:

Let $f : S^{m-1} \to \mathbb{R}$ be an L-Lipschitz function and let $\varepsilon > 0$. Then

$$\mathbb{P}(\{|f - M_f| > \varepsilon\}) \le C \exp(-\tfrac{m}{2L^2}\, \varepsilon^2),$$

where M_f is the median (or mean) of f, $\mathbb{P}$ is the normalized uniform measure on the sphere and $C > 0$ is a universal constant.

Verification of the exponent:

$$m = 2ns,\ L \sim 2n, \text{ so } \tfrac{m}{2L^2} \sim \tfrac{2ns}{8n^2} = \tfrac{s}{4n}.$$

Smoothness of $A \to AA^\dagger - \frac{\mathbb{I}}{n}$, $\quad \|\cdot\|_2 \to \|\cdot\|_2$

This operation is not Lipschitz, but its restriction to the Hilbert-Schmidt sphere is 2-Lipschitz, so f is $2n$-Lipschitz.

We can now appeal to **Levy's lemma**:

Let $f : S^{m-1} \to \mathbb{R}$ be an L-Lipschitz function and let $\varepsilon > 0$. Then

$$\mathbb{P}(\{|f - M_f| > \varepsilon\}) \leq C \exp(-\frac{m}{2L^2}\, \varepsilon^2),$$

where M_f is the median (or mean) of f, $\mathbb{P}$ is the normalized uniform measure on the sphere and $C > 0$ is a universal constant.

Verification of the exponent:

$$m = 2ns,\ L \sim 2n, \text{ so } \frac{m}{2L^2} \sim \frac{2ns}{8n^2} = \frac{s}{4n}.$$

Not quite the $\Omega(s)$ that we wanted, but since ultimately we are interested in the range $s = \Omega(d^3) = \Omega(n^{3/2})$, this yields some concentration for $\varepsilon = o(1)$ and so is "marginally sufficient" (even if not optimal).

A better argument

Let $\mathcal{T} := \{A : \|A\|_\infty = O(\frac{1}{\sqrt{n}})\}$. Then

- $\mathbb{P}(\mathcal{T}) \geq 1 - e^{-cs}$
- $A \to AA^\dagger$ is "locally Lipschitz on $\mathcal{T}$" with constant $O(\frac{1}{\sqrt{n}})$.

The bottom line is that the local Lipschitz constant of f on $\mathcal{T}$ is $O(\frac{1}{\sqrt{n}} \times n) = O(\sqrt{n})$.

This follows from $AA^\dagger$ being a rescaled Wishart distribution, whose eigenvalue distribution approximates (for large n, s) the rescaled Marchenko-Pastur distribution. In particular, the singular values of A are typically in the interval $\frac{1}{\sqrt{n}} \times [1 - \sqrt{\frac{n}{s}}, 1 + \sqrt{\frac{n}{s}}]$, hence $\|A\|_\infty = O(\frac{1}{\sqrt{n}})$. The probability estimate stated above is a consequence of the corresponding large deviation bound.

Local Levy's lemma

Let $\mathcal{T} \subset S^{m-1}$ be a subset of measure larger than $3/4$. Let $f : S^{m-1} \to \mathbb{R}$ be a function such that the restriction of f to $\mathcal{T}$ is L-Lipschitz. Then, for every $\varepsilon > 0$,

$$\mathbb{P}(\{|f - M_f| > \varepsilon\}) \leq \mathbb{P}(S^{m-1} \setminus \mathcal{T}) + C \exp(-\frac{m}{2L^2} \varepsilon^2),$$

where M_f is the median of f and $C > 0$ a universal constant.

Recalculation of the exponent:

$$m = 2ns,\ L = O(\sqrt{n}), \text{ so } \tfrac{m}{2L^2} = \tfrac{ns}{L^2} = \Omega(s).$$

We had $\exp(-c(\varepsilon)s)$ in the Theorem, so this is about right.

Step 1, estimating $\mathbb{E}\|\rho_{n,s} - \frac{\mathbb{I}}{n}\|_{\mathcal{S}_0}$: *the difficulty*

The gauge $\|\cdot\|_{\mathcal{S}_0}$ is hard to work with directly (NP-hard to calculate; Gurvits 2003).

Yesterday's argument was based on working with the "dual picture" and the somewhat easier dual quantity, $\|\cdot\|_{\mathcal{S}_0^\circ}$, where for $K \subset \mathbb{R}^m$

$$K^\circ := \{x \in \mathbb{R}^m : \langle x, y\rangle \leq 1 \text{ for all } y \in K\} ,$$

and subsequently on estimating $w(\mathcal{S}_0) := \int_{S_{H_0}} \|u\|_{\mathcal{S}_0^\circ}\, du.$

Step 1, estimating $\mathbb{E}\|\rho_{n,s} - \frac{\mathbb{I}}{n}\|_{\mathcal{S}_0}$: *the argument*

The argument is based on two facts/substeps, both non-trivial:

Step 1, estimating $\mathbb{E}\|\rho_{n,s} - \frac{\mathbb{I}}{n}\|_{\mathcal{S}_0}$: *the argument*

The argument is based on two facts/substeps, both non-trivial:

Substep (i) [Random matrices] : If $n, \frac{s}{n} \to \infty$, then

$$\mathbb{E}\|\rho_{n,s} - \frac{\mathbb{I}}{n}\|_{\mathcal{S}_0} \sim \frac{1}{\sqrt{s}}\, w(\mathcal{S}_0^\circ) = \frac{1}{\sqrt{s}} \int_{S_{H_0}} \|u\|_{\mathcal{S}_0}\, du.$$

There are two-sided estimates with universal constants if $n \leq s$.

Note: S_{H_0} is the Hilbert-Schmidt sphere in the space of trace 0 matrices. The equality is the definition of the mean width $w(\cdot)$.

Step 1, estimating $\mathbb{E}\|\rho_{n,s} - \frac{\mathbb{I}}{n}\|_{\mathcal{S}_0}$: *the argument*

The argument is based on two facts/substeps, both non-trivial:

Substep (i) [Random matrices] : If $n, \frac{s}{n} \to \infty$, then

$$\mathbb{E}\|\rho_{n,s} - \frac{\mathbb{I}}{n}\|_{\mathcal{S}_0} \sim \frac{1}{\sqrt{s}}\, w(\mathcal{S}_0^\circ) = \frac{1}{\sqrt{s}} \int_{S_{H_0}} \|u\|_{\mathcal{S}_0}\, du.$$

There are two-sided estimates with universal constants if $n \leq s$.

Note: S_{H_0} is the Hilbert-Schmidt sphere in the space of trace 0 matrices. The equality is the definition of the mean width $w(\cdot)$.

Substep (ii) [Geometry] : $cd^{3/2} \leq w(\mathcal{S}_0^\circ) \leq Cd^{3/2} \log d$

Step 1, estimating $\mathbb{E}\|\rho_{n,s} - \frac{\mathbb{I}}{n}\|_{\mathcal{S}_0}$: *the argument*

The argument is based on two facts/substeps, both non-trivial:

Substep (i) [Random matrices] : If $n, \frac{s}{n} \to \infty$, then

$$\mathbb{E}\|\rho_{n,s} - \tfrac{\mathbb{I}}{n}\|_{\mathcal{S}_0} \sim \frac{1}{\sqrt{s}}\, w(\mathcal{S}_0^\circ) = \frac{1}{\sqrt{s}} \int_{S_{H_0}} \|u\|_{\mathcal{S}_0}\, du.$$

There are two-sided estimates with universal constants if $n \leq s$.

Note: S_{H_0} is the Hilbert-Schmidt sphere in the space of trace 0 matrices. The equality is the definition of the mean width $w(\cdot)$.

Substep (ii) [Geometry] : $cd^{3/2} \leq w(\mathcal{S}_0^\circ) \leq Cd^{3/2} \log d$

Once these are shown, we set $s_{ent} := w(\mathcal{S}_0^\circ)^2$ and everything falls nicely into place.

Step 1, estimating $\mathbb{E}\|\rho_{n,s} - \frac{\mathbb{I}}{n}\|_{\mathcal{S}_0}$: *the argument*

The argument is based on two facts/substeps, both non-trivial:

Substep (i) [Random matrices] : If $n, \frac{s}{n} \to \infty$, then

$$\mathbb{E}\|\rho_{n,s} - \tfrac{\mathbb{I}}{n}\|_{\mathcal{S}_0} \sim \tfrac{1}{\sqrt{s}}\, w(\mathcal{S}_0^\circ) = \tfrac{1}{\sqrt{s}} \int_{S_{H_0}} \|u\|_{\mathcal{S}_0}\, du.$$

There are two-sided estimates with universal constants if $n \leq s$.

Note: S_{H_0} is the Hilbert-Schmidt sphere in the space of trace 0 matrices. The equality is the definition of the mean width $w(\cdot)$.

Substep (ii) [Geometry] : $cd^{3/2} \leq w(\mathcal{S}_0^\circ) \leq Cd^{3/2} \log d$

Once these are shown, we set $s_{ent} := w(\mathcal{S}_0^\circ)^2$ and everything falls nicely into place. For example, if $s \geq (1+\varepsilon)s_{ent}$, then

$$\mathbb{E}\|\rho_{n,s} - \tfrac{\mathbb{I}}{n}\|_{\mathcal{S}_0} \sim \tfrac{1}{\sqrt{s}}\, w(\mathcal{S}_0^\circ) = \sqrt{\tfrac{s_{ent}}{s}} \leq \tfrac{1}{\sqrt{1+\varepsilon}} \sim 1 - \tfrac{\varepsilon}{2}$$

Step 1, estimating $\mathbb{E}\|\rho_{n,s} - \frac{\mathbb{I}}{n}\|_{\mathcal{S}_0}$: *the argument*

The argument is based on two facts/substeps, both non-trivial:

Substep (i) [Random matrices] : If $n, \frac{s}{n} \to \infty$, then

$$\mathbb{E}\|\rho_{n,s} - \tfrac{\mathbb{I}}{n}\|_{\mathcal{S}_0} \sim \tfrac{1}{\sqrt{s}}\, w(\mathcal{S}_0^\circ) = \tfrac{1}{\sqrt{s}} \int_{S_{H_0}} \|u\|_{\mathcal{S}_0}\, du.$$

There are two-sided estimates with universal constants if $n \leq s$.
Note: S_{H_0} is the Hilbert-Schmidt sphere in the space of trace 0 matrices. The equality is the definition of the mean width $w(\cdot)$.

Substep (ii) [Geometry] : $cd^{3/2} \leq w(\mathcal{S}_0^\circ) \leq Cd^{3/2} \log d$

Once these are shown, we set $s_{ent} := w(\mathcal{S}_0^\circ)^2$ and everything falls nicely into place. For example, if $s \geq (1+\varepsilon)s_{ent}$, then

$$\mathbb{E}\|\rho_{n,s} - \tfrac{\mathbb{I}}{n}\|_{\mathcal{S}_0} \sim \tfrac{1}{\sqrt{s}}\, w(\mathcal{S}_0^\circ) = \sqrt{\tfrac{s_{ent}}{s}} \leq \tfrac{1}{\sqrt{1+\varepsilon}} \sim 1 - \tfrac{\varepsilon}{2}$$

Thus the mean (and the median) of $\mathbb{E}\|\rho_{n,s} - \frac{\mathbb{I}}{n}\|_{\mathcal{S}_0}$ are $\frac{\varepsilon}{4}$-separated from 1 and so, by concentration, $\mathbb{P}(\|\rho_{n,s} - \frac{\mathbb{I}}{n}\|_{\mathcal{S}_0} \leq 1) \approx 1$.

Substep (ii): $cd^{3/2} \leq w(\mathcal{S}_0^\circ) \leq Cd^{3/2} \log d$

Substep (ii): $cd^{3/2} \leq w(\mathcal{S}_0^\circ) \leq Cd^{3/2} \log d$

Recall: $w(\mathcal{S}_0^\circ)$ is the average of $\| \cdot \|_{\mathcal{S}_0}$, which is hard to work with directly.

Substep (ii): $cd^{3/2} \le w(\mathcal{S}_0^\circ) \le Cd^{3/2}\log d$

Recall: $w(\mathcal{S}_0^\circ)$ is the average of $\|\cdot\|_{\mathcal{S}_0}$, which is hard to work with directly. However, for the dual quantity $w(\mathcal{S}_0) = w(\mathcal{S})$ we have

Fact (Aubrun-S.) The mean width $w(\mathcal{S}) = \Theta(n^{-3/4}) = \Theta(d^{-3/2})$.

Substep (ii): $cd^{3/2} \leq w(\mathcal{S}_0^\circ) \leq Cd^{3/2} \log d$

Recall: $w(\mathcal{S}_0^\circ)$ is the average of $\|\cdot\|_{\mathcal{S}_0}$, which is hard to work with directly. However, for the dual quantity $w(\mathcal{S}_0) = w(\mathcal{S})$ we have

Fact (Aubrun-S.) The mean width $w(\mathcal{S}) = \Theta(n^{-3/4}) = \Theta(d^{-3/2})$.

One always has $w(K)w(K^\circ) \geq 1$, which yields $w(\mathcal{S}_0^\circ) \geq cd^{3/2}$.
If K is centrally symmetric and is in the so-called ℓ-position, the reverse inequality almost holds:

Substep (ii): $cd^{3/2} \leq w(\mathcal{S}_0^\circ) \leq Cd^{3/2} \log d$

Recall: $w(\mathcal{S}_0^\circ)$ is the average of $\| \cdot \|_{\mathcal{S}_0}$, which is hard to work with directly. However, for the dual quantity $w(\mathcal{S}_0) = w(\mathcal{S})$ we have

Fact (Aubrun-S.) The mean width $w(\mathcal{S}) = \Theta(n^{-3/4}) = \Theta(d^{-3/2})$.

One always has $w(K)w(K^\circ) \geq 1$, which yields $w(\mathcal{S}_0^\circ) \geq cd^{3/2}$.
If K is centrally symmetric and is in the so-called ℓ-position, the reverse inequality almost holds:

Fact (The *MM^*-estimate*; Pisier, Figiel–Tomczak-Jaegermann)
$w(K)w(K^\circ) = O(\log \dim K)$, so $w(K^\circ) = O\big(w(K)^{-1} \log \dim K\big)$.

Substep (ii): $cd^{3/2} \leq w(\mathcal{S}_0^\circ) \leq Cd^{3/2} \log d$

Recall: $w(\mathcal{S}_0^\circ)$ is the average of $\|\cdot\|_{\mathcal{S}_0}$, which is hard to work with directly. However, for the dual quantity $w(\mathcal{S}_0) = w(\mathcal{S})$ we have

Fact (Aubrun-S.) The mean width $w(\mathcal{S}) = \Theta(n^{-3/4}) = \Theta(d^{-3/2})$.

One always has $w(K)w(K^\circ) \geq 1$, which yields $w(\mathcal{S}_0^\circ) \geq cd^{3/2}$.
If K is centrally symmetric and is in the so-called ℓ-position, the reverse inequality almost holds:

Fact (The *MM**-estimate; Pisier, Figiel–Tomczak-Jaegermann)
$w(K)w(K^\circ) = O(\log \dim K)$, so $w(K^\circ) = O\big(w(K)^{-1} \log \dim K\big)$.

Note: K being in the ℓ-position means that K is isotropic in some precise technical sense; this can always be achieved by rescaling.

Substep (ii): $cd^{3/2} \leq w(\mathcal{S}_0^\circ) \leq Cd^{3/2} \log d$

Recall: $w(\mathcal{S}_0^\circ)$ is the average of $\| \cdot \|_{\mathcal{S}_0}$, which is hard to work with directly. However, for the dual quantity $w(\mathcal{S}_0) = w(\mathcal{S})$ we have

Fact (Aubrun-S.) The mean width $w(\mathcal{S}) = \Theta(n^{-3/4}) = \Theta(d^{-3/2})$.

One always has $w(K)w(K^\circ) \geq 1$, which yields $w(\mathcal{S}_0^\circ) \geq cd^{3/2}$. If K is centrally symmetric and is in the so-called ℓ-position, the reverse inequality almost holds:

Fact (The *MM^*-estimate*; Pisier, Figiel–Tomczak-Jaegermann) $w(K)w(K^\circ) = O(\log \dim K)$, so $w(K^\circ) = O\big(w(K)^{-1} \log \dim K\big)$.

Note: K being in the ℓ-position means that K is isotropic in some precise technical sense; this can always be achieved by rescaling.

If we had the MM^*-estimate for $\mathcal{S}_0$, it would follow that $w(\mathcal{S}_0^\circ) = O\big(d^{3/2} \log d\big)$, hence $s_{ent} = w(\mathcal{S}_0^\circ)^2 = O(d^3 \log^2 d)$, exactly as needed.

The workaround

One guarantee for $K \subset \mathbb{R}^m$ to be in the ℓ-position is when the isometry group of K acts irreducibly on $\mathbb{R}^m$.

The workaround

One guarantee for $K \subset \mathbb{R}^m$ to be in the ℓ-position is when the isometry group of K acts irreducibly on $\mathbb{R}^m$.

This is not the case if $K = \mathcal{S}_0$, but still there are sufficiently many isometries to imply the MM^*-estimate via simple representation theory.

The workaround

One guarantee for $K \subset \mathbb{R}^m$ to be in the ℓ-position is when the isometry group of K acts irreducibly on $\mathbb{R}^m$.

This is not the case if $K = \mathcal{S}_0$, but still there are sufficiently many isometries to imply the MM^*-estimate via simple representation theory.

Specifically, the action of the isometry group (local unitaries) splits into three irreducible factors, one of dimension $(d^2 - 1)^2$ and two of dimension $d^2 - 1$. It then follows from general theory that to bring $\mathcal{S}_0$ to the ℓ-position we only need to apply some dilations in the two smaller factors, and since their dimensions are relatively small, this does not affect in a major way the mean width of $\mathcal{S}_0$ or its polar.

The workaround

One guarantee for $K \subset \mathbb{R}^m$ to be in the ℓ-position is when the isometry group of K acts irreducibly on $\mathbb{R}^m$.

This is not the case if $K = \mathcal{S}_0$, but still there are sufficiently many isometries to imply the MM^*-estimate via simple representation theory.

Specifically, the action of the isometry group (local unitaries) splits into three irreducible factors, one of dimension $(d^2 - 1)^2$ and two of dimension $d^2 - 1$. It then follows from general theory that to bring $\mathcal{S}_0$ to the ℓ-position we only need to apply some dilations in the two smaller factors, and since their dimensions are relatively small, this does not affect in a major way the mean width of $\mathcal{S}_0$ or its polar.

Oops. . . need central symmetry for the MM^*-estimate and $\mathcal{S}_0$ is not symmetric. . .

The workaround

One guarantee for $K \subset \mathbb{R}^m$ to be in the ℓ-position is when the isometry group of K acts irreducibly on $\mathbb{R}^m$.

This is not the case if $K = \mathcal{S}_0$, but still there are sufficiently many isometries to imply the MM^*-estimate via simple representation theory.

Specifically, the action of the isometry group (local unitaries) splits into three irreducible factors, one of dimension $(d^2-1)^2$ and two of dimension d^2-1. It then follows from general theory that to bring $\mathcal{S}_0$ to the ℓ-position we only need to apply some dilations in the two smaller factors, and since their dimensions are relatively small, this does not affect in a major way the mean width of $\mathcal{S}_0$ or its polar.

Oops. . . need central symmetry for the MM^*-estimate and $\mathcal{S}_0$ is not symmetric. . . there is another workaround based on Santaló, inverse Santaló, Urysohn, Rogers-Shephard inequalities . . .

Substep (i): $\mathbb{E}\|\rho_{n,s} - \frac{\mathbb{I}}{n}\|_{\mathcal{S}_0} \sim \frac{1}{\sqrt{s}} \int_{S_{H_0}} \|u\|_{\mathcal{S}_0}\, du.$

(true for any gauge $\|\cdot\|_K$)

Substep (i): $\mathbb{E}\|\rho_{n,s} - \frac{\mathbb{I}}{n}\|_{\mathcal{S}_0} \sim \frac{1}{\sqrt{s}} \int_{S_{H_0}} \|u\|_{\mathcal{S}_0}\, du.$

(true for any gauge $\|\cdot\|_K$)

The trick : linearization

We approximate $\rho_{n,s} - \frac{\mathbb{I}}{n}$ by $\frac{1}{n\sqrt{s}}\, GUE_0$, where GUE_0 is the standard H_0-valued random Gaussian matrix, and use the relationship between the spherical and the Gaussian mean

$$\mathbb{E}\|\cdot\|_K \sim \sqrt{m} \int_{S^{m-1}} \|u\|_K\, du$$

($\mathbb{E}$ with respect to the standard Gaussian measure on $\mathbb{R}^m$)

Substep (i): $\mathbb{E}\|\rho_{n,s} - \frac{\mathbb{I}}{n}\|_{\mathcal{S}_0} \sim \frac{1}{\sqrt{s}} \int_{S_{H_0}} \|u\|_{\mathcal{S}_0}\, du.$

(true for any gauge $\|\cdot\|_K$)

The trick : linearization

We approximate $\rho_{n,s} - \frac{\mathbb{I}}{n}$ by $\frac{1}{n\sqrt{s}}\, GUE_0$, where GUE_0 is the standard H_0-valued random Gaussian matrix, and use the relationship between the spherical and the Gaussian mean

$$\mathbb{E}\|\cdot\|_K \sim \sqrt{m} \int_{S^{m-1}} \|u\|_K\, du$$

($\mathbb{E}$ with respect to the standard Gaussian measure on $\mathbb{R}^m$)

The bottom line is then

$$\mathbb{E}\|\rho_{n,s} - \frac{\mathbb{I}}{n}\|_{\mathcal{S}_0} \sim \frac{1}{n\sqrt{s}} \|GUE_0\|_{\mathcal{S}_0} \sim \frac{\sqrt{n^2-1}}{n\sqrt{s}} \int_{S_{H_0}} \|u\|_{\mathcal{S}_0}\, du$$

as needed.

Approximating $\rho_{n,s} - \frac{\mathbb{I}}{n}$ by $\frac{1}{n\sqrt{s}}\, GUE_0$, sanity check

Approximating $\rho_{n,s} - \frac{\mathbb{I}}{n}$ by $\frac{1}{n\sqrt{s}}\, GUE_0$, sanity check

Asymptotically, the spectrum of $\rho_{n,s} = AA^\dagger$ lives on the interval $\frac{1}{n} \times [(1-\sqrt{\frac{n}{s}})^2, (1+\sqrt{\frac{n}{s}})^2] \approx [\frac{1}{n} - \frac{2}{\sqrt{ns}}, \frac{1}{n} + \frac{2}{\sqrt{ns}}]$, from Marchenko-Pastur, while – by Wigner – the spectrum of $\frac{1}{\sqrt{n}}\, GUE_0$ lives on $[-2, 2]$, so at least the scaling is right:

Approximating $\rho_{n,s} - \frac{\mathbb{I}}{n}$ by $\frac{1}{n\sqrt{s}}\, GUE_0$, sanity check

Asymptotically, the spectrum of $\rho_{n,s} = AA^\dagger$ lives on the interval $\frac{1}{n} \times [(1-\sqrt{\frac{n}{s}})^2, (1+\sqrt{\frac{n}{s}})^2] \approx [\frac{1}{n} - \frac{2}{\sqrt{ns}}, \frac{1}{n} + \frac{2}{\sqrt{ns}}]$,
from Marchenko-Pastur, while – by Wigner – the spectrum of $\frac{1}{\sqrt{n}}\, GUE_0$ lives on $[-2, 2]$, so at least the scaling is right: both ensembles have spectrum on $[-\frac{2}{\sqrt{ns}}, \frac{2}{\sqrt{ns}}]$,

Approximating $\rho_{n,s} - \frac{\mathbb{I}}{n}$ by $\frac{1}{n\sqrt{s}}\, GUE_0$, for real

Fact (Wigner, Geman?, Bai-Yin) Both $\sqrt{ns}\left(\rho_{n,s} - \frac{\mathbb{I}}{n}\right)$ and $\frac{1}{\sqrt{n}}\, GUE_0$ converge to the standard semicircular law as $n, \frac{s}{n} \to \infty$.

Approximating $\rho_{n,s} - \frac{\mathbb{I}}{n}$ by $\frac{1}{n\sqrt{s}} GUE_0$, for real

Fact (Wigner, Geman?, Bai-Yin) Both $\sqrt{ns}(\rho_{n,s} - \frac{\mathbb{I}}{n})$ and $\frac{1}{\sqrt{n}} GUE_0$ converge to the standard semicircular law as $n, \frac{s}{n} \to \infty$.

Usually cited: weak convergence of $n^{-1} \sum_{i=1}^{n} \delta_{\lambda_i(X)}$, but also extreme eigenvalues converge to the endpoints of the support of the limit spectrum.

Approximating $\rho_{n,s} - \frac{\mathbb{I}}{n}$ by $\frac{1}{n\sqrt{s}} GUE_0$, for real

Fact (Wigner, Geman?, Bai-Yin) Both $\sqrt{ns}(\rho_{n,s} - \frac{\mathbb{I}}{n})$ and $\frac{1}{\sqrt{n}} GUE_0$ converge to the standard semicircular law as $n, \frac{s}{n} \to \infty$.

Usually cited: weak convergence of $n^{-1} \sum_{i=1}^{n} \delta_{\lambda_i(X)}$, but also extreme eigenvalues converge to the endpoints of the support of the limit spectrum.

Can be subsumed as convergence in the ∞-Wasserstein distance:

$$d_\infty(\mu, \nu) := \inf \|X - Y\|_{L^\infty},$$

with infimum over all couples (X, Y) of random variables with laws μ and ν, defined on a common probability space.

Approximating $\rho_{n,s} - \frac{\mathbb{I}}{n}$ by $\frac{1}{n\sqrt{s}}\, GUE_0$, for real

Fact (Wigner, Geman?, Bai-Yin) Both $\sqrt{ns}(\rho_{n,s} - \frac{\mathbb{I}}{n})$ and $\frac{1}{\sqrt{n}}\, GUE_0$ converge to the standard semicircular law as $n, \frac{s}{n} \to \infty$.

Usually cited: weak convergence of $n^{-1} \sum_{i=1}^{n} \delta_{\lambda_i(X)}$, but also extreme eigenvalues converge to the endpoints of the support of the limit spectrum.

Can be subsumed as convergence in the ∞-Wasserstein distance:

$$d_\infty(\mu, \nu) := \inf \|X - Y\|_{L^\infty},$$

with infimum over all couples (X, Y) of random variables with laws μ and ν, defined on a common probability space.

The needed equivalence $\mathbb{E}\|\rho_{n,s} - \frac{\mathbb{I}}{n}\|_{\mathcal{S}_0} \sim \mathbb{E}\|\frac{1}{n\sqrt{s}}\, GUE_0\|_{\mathcal{S}_0}$ will now follow from the general theory. However, since the gauge $\|\cdot\|_{\mathcal{S}_0}$ is rather poorly continuous, some finesse is needed.

∞-Wasserstein distance and gauges

∞-Wasserstein distance and gauges

Let $\mathbb{R}^{n,0} := \{(x_j) \in \mathbb{R}^n : \sum_j x_j = 0\}$; $\nu_x := \frac{1}{n} \sum_j \delta_{x_j}$ for $x \in \mathbb{R}^n$.

∞-Wasserstein distance and gauges

Let $\mathbb{R}^{n,0} := \{(x_j) \in \mathbb{R}^n : \sum_j x_j = 0\}$; $\nu_x := \frac{1}{n} \sum_j \delta_{x_j}$ for $x \in \mathbb{R}^n$.

Proposition Let μ be a 0-centered non-point mass measure.
Then $\forall \varepsilon > 0 \; \exists \eta > 0$ such that if
• $x, y \in \mathbb{R}^{n,0}$ verify $d_\infty(\nu_x, \mu) \leq \eta$ and $d_\infty(\nu_y, \mu) \leq \eta$, and
• ϕ is a permutationally invariant convex function on $\mathbb{R}^{n,0}$,
then $\phi(x) \leq \phi((1+\varepsilon)y)$.

∞-Wasserstein distance and gauges

Let $\mathbb{R}^{n,0} := \{(x_j) \in \mathbb{R}^n : \sum_j x_j = 0\}$; $\nu_x := \frac{1}{n}\sum_j \delta_{x_j}$ for $x \in \mathbb{R}^n$.

Proposition Let μ be a 0-centered non-point mass measure.
Then $\forall \varepsilon > 0 \; \exists \eta > 0$ such that if
• $x, y \in \mathbb{R}^{n,0}$ verify $d_\infty(\nu_x, \mu) \leq \eta$ and $d_\infty(\nu_y, \mu) \leq \eta$,and
• ϕ is a permutationally invariant convex function on $\mathbb{R}^{n,0}$,
then $\phi(x) \leq \phi((1+\varepsilon)y)$.

The link to our context is as follows. If K is a convex body in $H_0 \subset \mathcal{M}_n^{sa}$ (containing 0 in the interior), we set

$$\phi_K(x) = \int_{\mathcal{U}(n)} \|U\mathrm{Diag}(x)U^\dagger\|_K \, dU$$

∞-Wasserstein distance and gauges

Let $\mathbb{R}^{n,0} := \{(x_j) \in \mathbb{R}^n : \sum_j x_j = 0\}$; $\nu_x := \frac{1}{n}\sum_j \delta_{x_j}$ for $x \in \mathbb{R}^n$.

Proposition Let μ be a 0-centered non-point mass measure.
Then $\forall \varepsilon > 0\ \exists \eta > 0$ such that if
- $x, y \in \mathbb{R}^{n,0}$ verify $d_\infty(\nu_x, \mu) \leq \eta$ and $d_\infty(\nu_y, \mu) \leq \eta$,and
- ϕ is a permutationally invariant convex function on $\mathbb{R}^{n,0}$,

then $\phi(x) \leq \phi((1+\varepsilon)y)$.

The link to our context is as follows. If K is a convex body in $H_0 \subset \mathcal{M}_n^{sa}$ (containing 0 in the interior), we set

$$\phi_K(x) = \int_{\mathcal{U}(n)} \|U\mathrm{Diag}(x)U^\dagger\|_K dU$$

We now apply the Proposition with $\phi = \phi_{\mathcal{S}_0}$, μ–the semicircular distribution, x, y – the spectra of the two ensembles, and take the expected values. In addition to the convergence of the ensembles to a common limit spectral distribution, the calculation uses their invariance under conjugation by unitaries.

The concept behind the Proposition: *majorization*

For $x, y \in \mathbb{R}^{n,0}$, we write $x \prec y$ if, for every $k \in \{1, \ldots, n\}$,

$$\sum_{i=1}^{k} x_i^{\downarrow} \leq \sum_{i=1}^{k} y_i^{\downarrow},$$

where $x^{\downarrow}$ the non-increasing rearrangement of x.

The concept behind the Proposition: *majorization*

For $x, y \in \mathbb{R}^{n,0}$, we write $x \prec y$ if, for every $k \in \{1, \ldots, n\}$,

$$\sum_{i=1}^{k} x_i^{\downarrow} \leq \sum_{i=1}^{k} y_i^{\downarrow},$$

where $x^{\downarrow}$ the non-increasing rearrangement of x. This is equivalent to each of the following

(a) whenever ϕ is a permutationally invariant convex function on $\mathbb{R}^{n,0}$, then $\phi(x) \leq \phi(y)$

(b) for every $t \in \mathbb{R}$, we have $\sum_{i=1}^{n} |x_i - t| \leq \sum_{i=1}^{n} |y_i - t|$

The concept behind the Proposition: *majorization*

For $x, y \in \mathbb{R}^{n,0}$, we write $x \prec y$ if, for every $k \in \{1, \dots, n\}$,

$$\sum_{i=1}^{k} x_i^{\downarrow} \leq \sum_{i=1}^{k} y_i^{\downarrow},$$

where $x^{\downarrow}$ the non-increasing rearrangement of x. This is equivalent to each of the following

(a) whenever ϕ is a permutationally invariant convex function on $\mathbb{R}^{n,0}$, then $\phi(x) \leq \phi(y)$

(b) for every $t \in \mathbb{R}$, we have $\sum_{i=1}^{n} |x_i - t| \leq \sum_{i=1}^{n} |y_i - t|$

Note that the assertion of the Proposition is in the spirit of (a), while $\frac{1}{n}\sum_{i=1}^{n} |x_i - t|$ from (b) can be rewritten as $\int \Phi \, d\nu_x$, where $\Phi(u) := |u - t|$, and so can be related to convergence of measures.

THANK YOU