# Concentration of noncommutative polynomials of random matrices

Mark Meckes

Case Western Reserve University

College Station, July 23, 2009

# Random matrices

$X = (X_{ij})$ is a random $n \times n$ symmetric (or Hermitian) matrix.

Typical examples of interest:

- Wigner matrices: $X_{ij}$ are independent for $1 \leq i \leq j \leq n$.
- Orthogonal/unitary ensembles: $X$ has density

  $$e^{-\operatorname{Tr} U(A)} = e^{-\sum_{i=1}^{n} U(\lambda_i(A))}$$

  with respect to Lebesgue measure $dA$ on $M_n^{sa}$ for some $U : \mathbb{R} \to \mathbb{R}$.
- The intersection of these examples are the Gaussian orthogonal/unitary ensembles: $U(x) \propto x^2$, with independent Gaussian entries above the diagonal.

# Wigner's theorem

## Theorem

*If $X$ is one of the types of random matrix described on the last slide, with*

$$\mathbb{E}X_{ij} = 0, \qquad \operatorname{Var} X_{ij} = 1$$

*(and a little more), and $f : \mathbb{R} \to \mathbb{R}$ is bounded and continuous or a polynomial, then*

$$\frac{1}{n} \operatorname{Tr} f\left(\frac{1}{\sqrt{n}} X\right) \xrightarrow{n\to\infty} \frac{1}{2\pi} \int_{-2}^{2} f(x)\sqrt{4 - x^2}\, dx$$

*almost surely.*

# Wigner's theorem

## Theorem

*If $X$ is one of the types of random matrix described on the last slide, with*

$$\mathbb{E}X_{ij} = 0, \qquad \operatorname{Var} X_{ij} = 1$$

*(and a little more), and $f : \mathbb{R} \to \mathbb{R}$ is bounded and continuous or a polynomial, then*

$$\frac{1}{n} \operatorname{Tr} f\left(\frac{1}{\sqrt{n}} X\right) \xrightarrow{n\to\infty} \frac{1}{2\pi} \int_{-2}^{2} f(x)\sqrt{4 - x^2}\, dx$$

*almost surely.*

Wigner's theorem is a SLLN for functionals $\frac{1}{n}\operatorname{Tr} f$, and implies a SLLN for the spectral measure of $\frac{1}{\sqrt{n}}X$.

Central limit theorems also have a rich history.

# Multiple random matrices – free probability theory

Let $X^{(1)}, \ldots, X^{(m)}$ be independent random matrices (with some more conditions) and let $P$ be a noncommutative polynomial in $m$ variables.

Voiculescu's free probability theory implies that

$$\lim_{n\to\infty} \frac{1}{n} \operatorname{Tr} P\left(\frac{1}{\sqrt{n}}X^{(1)}, \ldots, \frac{1}{\sqrt{n}}X^{(m)}\right)$$

exists and gives a method to compute it.

# Multiple random matrices – free probability theory

Let $X^{(1)}, \ldots, X^{(m)}$ be independent random matrices (with some more conditions) and let $P$ be a noncommutative polynomial in $m$ variables.

Voiculescu's free probability theory implies that

$$\lim_{n\to\infty} \frac{1}{n} \operatorname{Tr} P\left(\frac{1}{\sqrt{n}} X^{(1)}, \ldots, \frac{1}{\sqrt{n}} X^{(m)}\right)$$

exists and gives a method to compute it.

It is also of interest to consider

$$\frac{1}{n} \operatorname{Tr} P\left(\frac{1}{\sqrt{n}} X^{(1)}, \ldots, \frac{1}{\sqrt{n}} X^{(m)}\right),$$

when $X^{(1)}, \ldots, X^{(m)}$ are non-Hermitian and $P$ is a $*$-polynomial.

# Concentration hypothesis

A random vector $Y$ in a normed space $V$ satisfies the convex concentration property if for any 1-Lipschitz function $F : V \to \mathbb{R}$,

$$\mathbb{P}\big[\,|F(Y) - \mathbb{M}F(Y)| \geq t\big] \leq Ce^{-ct^2}$$

for every $t > 0$, where $\mathbb{M}$ denotes a median.

# Concentration hypothesis

A random vector $Y$ in a normed space $V$ satisfies the convex concentration property if for any 1-Lipschitz function $F : V \to \mathbb{R}$,

$$\mathbb{P}\big[\,|F(Y) - \mathbb{M}F(Y)| \geq t\big] \leq Ce^{-ct^2}$$

for every $t > 0$, where $\mathbb{M}$ denotes a median.

Typical examples of interest ($V = \mathbb{R}^N$):

- $Y \sim N(0, I_N)$ (Gaussian isoperimetric inequality).
- $Y$ has a density $e^{-U(y)}$ for $U : \mathbb{R}^N \to \mathbb{R}$ such that $D^2 U \geq c$ (log-Sobolev inequalities).
- $Y_1, \ldots, Y_N$ are independent and $|Y_i| \leq c$ (Talagrand's convex distance inequality).

Standard observation: with some modification of the constants, this is equivalent to the same kind of concentration about a mean.

# Back to random matrices

We equip $M_n^{sa}$ with the Hilbert-Schmidt norm:

$$\|A\|_2 = \sqrt{\sum_{i=1}^{n} \lambda_i(A)^2} = \sqrt{\sum_{i,j=1}^{n} \left|a_{ij}\right|^2}.$$

# Back to random matrices

We equip $M_n^{sa}$ with the Hilbert-Schmidt norm:

$$\|A\|_2 = \sqrt{\sum_{i=1}^{n} \lambda_i(A)^2} = \sqrt{\sum_{i,j=1}^{n} |a_{ij}|^2}.$$

With respect to this norm, $X$ satisfies the CCP in the following cases:

- $X_{ij}$ are independent for $1 \leq i \leq j \leq n$ and $|X_{ij}| \leq c$.
- $X$ is drawn from an orthogonal or unitary ensemble with density $e^{-\operatorname{Tr} U(A)}$ with $U''(x) \geq c$ (in particular, the GUE/GOE).

# Guionnet and Zeitouni's theorem

## Theorem (Guionnet-Zeitouni)

*Suppose* $X$ *satisfies the CCP on* $M_n^{sa}$*. If* $f : \mathbb{R} \to \mathbb{R}$ *is convex and* 1*-Lipschitz, then*

$$\mathbb{P}\left[\left|\frac{1}{n}\operatorname{Tr} f\left(\frac{1}{\sqrt{n}}X\right) - \mathbb{M}\frac{1}{n}\operatorname{Tr} f\left(\frac{1}{\sqrt{n}}X\right)\right| \geq t\right] \leq Ce^{-cn^2t^2}$$

*for every* $t > 0$*.*

# Guionnet and Zeitouni's theorem

## Theorem (Guionnet-Zeitouni)

*Suppose $X$ satisfies the CCP on $M_n^{sa}$. If $f : \mathbb{R} \to \mathbb{R}$ is convex and 1-Lipschitz, then*

$$\mathbb{P}\left[\left|\frac{1}{n}\operatorname{Tr} f\left(\frac{1}{\sqrt{n}}X\right) - \mathbb{M}\frac{1}{n}\operatorname{Tr} f\left(\frac{1}{\sqrt{n}}X\right)\right| \geq t\right] \leq Ce^{-cn^2t^2}$$

*for every $t > 0$.*

Guionnnet and Zeitouni used this to show that the spectral measure of $\frac{1}{\sqrt{n}}X$ is concentrated about its mean.

To apply CCP to $\frac{1}{n}\operatorname{Tr} f$, Guionnet and Zeitouni just needed some basic facts from matrix analysis.

# Matrix analytic facts

### Lemma (Davis, "Klein's lemma")

*If $F : \mathbb{R}^n \to \mathbb{R}$ is convex, then*

$$A \mapsto F\big(\lambda_1(A), \ldots, \lambda_n(A)\big)$$

*is a convex function of the Hermitian matrix A. In particular, if $f : \mathbb{R} \to \mathbb{R}$ is convex, then $\operatorname{Tr} f : M_n^{sa} \to \mathbb{R}$ is convex.*

# Matrix analytic facts

## Lemma (Davis, “Klein’s lemma”)

*If* $F : \mathbb{R}^n \to \mathbb{R}$ *is convex, then*

$$A \mapsto F\big(\lambda_1(A), \ldots, \lambda_n(A)\big)$$

*is a convex function of the Hermitian matrix* $A$*. In particular, if* $f : \mathbb{R} \to \mathbb{R}$ *is convex, then* $\operatorname{Tr} f : M_n^{sa} \to \mathbb{R}$ *is convex.*

## Lemma

*The map*

$$A \mapsto \big(\lambda_1(A), \ldots \lambda_n(A)\big)$$

*is a* $1$*-Lipschitz map* $M_n^{sa} \to \mathbb{R}^n$*. In particular, if* $f : \mathbb{R} \to \mathbb{R}$ *is* $1$*-Lipschitz, then* $\operatorname{Tr} f$ *is* $\sqrt{n}$*-Lipschitz.*

# Concentration for polynomials

What can we say about $\frac{1}{n}\operatorname{Tr} P\left(\frac{1}{\sqrt{n}}X\right)$ when $P : \mathbb{R} \to \mathbb{R}$ is a polynomial?

The major obstacle here is that polynomials are not Lipschitz.

# Concentration for polynomials

What can we say about $\frac{1}{n}\operatorname{Tr} P\left(\frac{1}{\sqrt{n}}X\right)$ when $P : \mathbb{R} \to \mathbb{R}$ is a polynomial?

The major obstacle here is that polynomials are not Lipschitz.

From the point of view of free probability theory, it's interesting to consider even

$$\frac{1}{n}\operatorname{Tr} P\left(\frac{1}{\sqrt{n}}X^{(1)}, \ldots, \frac{1}{\sqrt{n}}X^{(m)}\right),$$

where $X^{(1)}, \ldots, X^{(m)}$ are independent non-Hermitian random matrices and $P$ is a noncommutative $*$-polynomial.

# Main theorem

## Theorem (M.-Szarek)

*Let* $X^{(1)}, \ldots, X^{(m)}$ *be independent Hermitian or non-Hermitian random matrices which satisfy the CCP and* $\mathbb{E}X^{(i)} = 0$*, and let* $P$ *be a noncommutative *-polynomial in* $m$ *variables with degree* $d$ *and coefficients bounded by 1 in modulus. Define*

$$Y_P = \frac{1}{n} \operatorname{Tr} P\left(\frac{1}{\sqrt{n}}X^{(1)}, \ldots, \frac{1}{\sqrt{n}}X^{(m)}\right).$$

*Then*

$$\mathbb{P}\big[\,|Y_P - \mathbb{E}Y_P| \geq t\big] \leq C_{m,d} \exp\big[-c_{m,d} \min\big\{n^2 t^2, n^{1+2/d} t^{2/d}\big\}\big]$$

*for* $t > 0$.

We will begin by proving the following special case.

### Proposition

*Let $X$ be Hermitian and satisfy the CCP, and let $d \geq 1$. Then*

$$\mathbb{P}\left[\left|\operatorname{Tr} X^d - \mathbb{M} \operatorname{Tr} X^d\right| \geq t\right] \leq C \exp\left[-c_d \min\left\{n^{-d} t^2, t^{2/d}\right\}\right]$$

*for $t > 0$.*

We will begin by proving the following special case.

## Proposition

*Let $X$ be Hermitian and satisfy the CCP, and let $d \geq 1$. Then*

$$\mathbb{P}\Big[\big|\operatorname{Tr} X^d - \mathbb{M}\operatorname{Tr} X^d\big| \geq t\Big] \leq C\exp\Big[-c_d \min\big\{n^{-d}t^2, t^{2/d}\big\}\Big]$$

*for $t > 0$.*

Let $F : M_n^{sa} \to \mathbb{R}$, $F(A) = \operatorname{Tr} A^d = \sum_{i=1}^n \lambda_i(A)^d$. Then

$$\|\nabla F(A)\|_2 = d\sqrt{\sum_{i=1}^n \lambda_i(A)^{2(d-1)}} = d\,\|A\|_{2(d-1)}^{d-1},$$

and $F$ is convex if $d$ is even.

The idea now is to control the Lipschitz behavior of $F$ by controlling $\|X\|_{2(d-1)}$.

Assume $d$ is even for now.

It is trivial to bound the Hilbert-Schmidt norm, and a standard $\varepsilon$-net argument bounds the operator norm:

$$\mathbb{E}\left\|X\right\|_2 \leq \sqrt{\sum_{i,j=1}^{n} \mathbb{E}\left|X_{ij}\right|^2} \leq cn, \qquad \mathbb{E}\left\|X\right\|_\infty \leq c\sqrt{n}.$$

Interpolating between these, we obtain

$$\mathbb{E}\left\|X\right\|_{2(d-1)} \leq cn^{d/2(d-1)}.$$

It is trivial to bound the Hilbert-Schmidt norm, and a standard $\varepsilon$-net argument bounds the operator norm:

$$\mathbb{E}\|X\|_2 \le \sqrt{\sum_{i,j=1}^{n} \mathbb{E}\left|X_{ij}\right|^2} \le cn, \qquad \mathbb{E}\|X\|_\infty \le c\sqrt{n}.$$

Interpolating between these, we obtain

$$\mathbb{E}\|X\|_{2(d-1)} \le cn^{d/2(d-1)}.$$

Since $2(d-1) \ge 2$, $A \mapsto \|A\|_{2(d-1)}$ is a convex 1-Lipschitz function on $M_n^{sa}$, CCP gives the needed control of $\|X\|_{2(d-1)}$:

$$\mathbb{P}\left[\,\|X\|_{2(d-1)} \ge n^{d/2(d-1)}a\right] \le C\exp\left[-cn^{d/(d-1)}a^2\right]$$

for $a \ge c$.

Now let $F_a : M_n^{sa} \to \mathbb{R}$ be a family of functions for $a \geq c$ such that:

- $F_a(A) = \operatorname{Tr} A^d$ if $\|A\|_{2(d-1)} \leq n^{d/2(d-1)} a$.
- $F_a$ is convex.
- $F_a$ is $(dn^{d/2} a^{d-1})$-Lipschitz.
- $F_a(A) \leq \operatorname{Tr} A^d$ for every $A$.
- For every $A$, $F_a(A)$ increases monotonically to $\operatorname{Tr} A^d$ as $a \to \infty$.

Now let $F_a : M_n^{sa} \to \mathbb{R}$ be a family of functions for $a \geq c$ such that:

- $F_a(A) = \operatorname{Tr} A^d$ if $\|A\|_{2(d-1)} \leq n^{d/2(d-1)} a$.
- $F_a$ is convex.
- $F_a$ is $(dn^{d/2} a^{d-1})$-Lipschitz.
- $F_a(A) \leq \operatorname{Tr} A^d$ for every $A$.
- For every $A$, $F_a(A)$ increases monotonically to $\operatorname{Tr} A^d$ as $a \to \infty$.

By the CCP applied to $F_c$,

$$\begin{aligned}\mathbb{P}\left[\left|\operatorname{Tr} X^d - \mathbb{M}F_c(X)\right| \geq t\right] \\ \leq \mathbb{P}\left[\left|F_c(X) - \mathbb{M}F_c(X)\right| \geq t\right] + \mathbb{P}\left[\|X\|_{2(d-1)} \geq n^{d/2(d-1)} c\right]\end{aligned}$$

Now let $F_a : M_n^{sa} \to \mathbb{R}$ be a family of functions for $a \geq c$ such that:

- $F_a(A) = \operatorname{Tr} A^d$ if $\|A\|_{2(d-1)} \leq n^{d/2(d-1)} a$.
- $F_a$ is convex.
- $F_a$ is $(dn^{d/2} a^{d-1})$-Lipschitz.
- $F_a(A) \leq \operatorname{Tr} A^d$ for every $A$.
- For every $A$, $F_a(A)$ increases monotonically to $\operatorname{Tr} A^d$ as $a \to \infty$.

By the CCP applied to $F_c$,

$$\begin{aligned}\mathbb{P}\left[\left|\operatorname{Tr} X^d - \mathbb{M} F_c(X)\right| \geq t\right] \\ &\leq \mathbb{P}\left[\left|F_c(X) - \mathbb{M} F_c(X)\right| \geq t\right] + \mathbb{P}\left[\|X\|_{2(d-1)} \geq n^{d/2(d-1)} c\right] \\ &\leq C \exp\left[-c_d \frac{t^2}{n^d}\right] + C \exp\left[-cn^{d/(d-1)}\right],\end{aligned}$$

and so $\left|\mathbb{M} \operatorname{Tr} X^d - \mathbb{M} F_a(X)\right| \leq c_d n^{d/2}$ for $a \geq c$.

Now by the CCP applied to $F_a$,

$$\begin{aligned}\mathbb{P}\big[\big|\operatorname{Tr} X^d - \mathbb{M}\operatorname{Tr} X^d\big| \geq t\big] \\ &\leq \mathbb{P}\big[\big|F_a(X) - \mathbb{M}\operatorname{Tr} X^d\big| \geq t\big] + \mathbb{P}\big[\,\|X\|_{2(d-1)} \geq n^{d/2(d-1)} a\big]\end{aligned}$$

Now by the CCP applied to $F_a$,

$$\begin{aligned}\mathbb{P}\big[\big|\operatorname{Tr} X^d - \mathbb{M}\operatorname{Tr} X^d\big| \geq t\big] &\leq \mathbb{P}\big[\big|F_a(X) - \mathbb{M}\operatorname{Tr} X^d\big| \geq t\big] + \mathbb{P}\big[\,\|X\|_{2(d-1)} \geq n^{d/2(d-1)}a\big] \\ &\leq C\exp\left[-c_d\frac{(t - c_d n^{d/2})^2}{n^d a^{2(d-1)}}\right] + C\exp\big[-cn^{d/(d-1)}a^2\big].\end{aligned}$$

Now by the CCP applied to $F_a$,

$$\begin{aligned}\mathbb{P}\big[\big|\operatorname{Tr} X^d - \mathbb{M}\operatorname{Tr} X^d\big| \geq t\big] &\leq \mathbb{P}\big[\big|F_a(X) - \mathbb{M}\operatorname{Tr} X^d\big| \geq t\big] + \mathbb{P}\big[\,\|X\|_{2(d-1)} \geq n^{d/2(d-1)} a\big] \\ &\leq C \exp\left[-c_d \frac{(t - c_d n^{d/2})^2}{n^d a^{2(d-1)}}\right] + C \exp\big[-c n^{d/(d-1)} a^2\big].\end{aligned}$$

For each fixed $t > c_d n^{d/2}$, we optimize over $a \geq c$ to get

$$\mathbb{P}\big[\big|\operatorname{Tr} X^d - \mathbb{M}\operatorname{Tr} X^d\big| \geq t\big] \leq C \exp\big[-c_d \min\big\{n^{-d} t^2, t^{2/d}\big\}\big].$$

Picking constants appropriately makes the inequality hold vacuously for $t \leq c_d n^{d/2}$.

Changing the normalization gives the special case of a monomial of even degree $d$ in one Hermitian random matrix.

Changing the normalization gives the special case of a monomial of even degree $d$ in one Hermitian random matrix.

If $d$ is odd, note that

$$x^d = \max\left\{x^d, 0\right\} - \max\left\{(-x)^d, 0\right\}$$

is the difference of two convex functions. Therefore we can apply the same argument to each of these.

Changing the normalization gives the special case of a monomial of even degree *d* in one Hermitian random matrix.

If *d* is odd, note that

$$x^d = \max\left\{x^d, 0\right\} - \max\left\{(-x)^d, 0\right\}$$

is the difference of two convex functions. Therefore we can apply the same argument to each of these.

For the general theorem, by the triangle inequality we can first reduce to monomials.

To get from powers of a single Hermitian random matrix to $*$-monomials in *m* non-Hermitian random matrices, we need an ad hoc polarization trick.

The basic idea: define the $dn \times dn$ block matrix

$$B = \begin{bmatrix} 0 & A_1 & 0 & & \\ & 0 & A_2 & 0 & \\ & & \ddots & \ddots & 0 \\ 0 & & & 0 & A_{d-1} \\ A_d & 0 & & & 0 \end{bmatrix}$$

where each $A_i$ is equal to $X^{(k)}$ or $\left(X^{(k)}\right)^*$ for some $k$.

The basic idea: define the $dn \times dn$ block matrix

$$B = \begin{bmatrix} 0 & A_1 & 0 & & & \\ & 0 & A_2 & 0 & & \\ & & \ddots & \ddots & 0 & \\ 0 & & & 0 & A_{d-1} & \\ A_d & 0 & & & 0 & \end{bmatrix}$$

where each $A_i$ is equal to $X^{(k)}$ or $\left(X^{(k)}\right)^*$ for some $k$.

Then

$$B^d = \begin{bmatrix} A_1 A_2 \cdots A_{d-1} A_d & & & 0 \\ & A_2 A_3 \cdots A_d A_1 & & \\ & & \ddots & \\ 0 & & & A_d A_1 \cdots A_{d-1} \end{bmatrix},$$

so $\operatorname{Tr} B^d = d \operatorname{Tr}\left(A_1 A_2 \cdots A_d\right)$.

$B$ satisfies the CCP with a constant that now depends on $d$, but $B$ is not Hermitian so the Proposition doesn't apply.

$B$ satisfies the CCP with a constant that now depends on $d$, but $B$ is not Hermitian so the Proposition doesn't apply.

Define the $dn \times dn$ Hermitian random matrix

$$B_+ = B + B^* = \begin{bmatrix} 0 & A_1 & & & & A_d^* \\ A_1^* & 0 & A_2 & & & \\ & A_2^* & 0 & & & \\ & & & \ddots & & \\ & & & & 0 & A_{d-1} \\ A_d & & & & A_{d-1}^* & 0 \end{bmatrix}.$$

The Proposition applies to show $\operatorname{Tr} B_+^d$ is concentrated, and if $d$ is odd then

$$\operatorname{Tr} B_+^d = 2d \operatorname{Re} \operatorname{Tr} \left( A_1 A_2 \cdots A_d \right).$$

The imaginary part can be handled by multiplying (say) $A_d$ by $i$.

If $d$ is even, then we also define

$$B_- = \begin{bmatrix} 0 & A_1 & & & & -A_d^* \\ A_1^* & 0 & A_2 & & & \\ & A_2^* & 0 & & & \\ & & & \ddots & & \\ & & & & 0 & A_{d-1} \\ -A_d & & & & A_{d-1}^* & 0 \end{bmatrix}.$$

Then the Proposition again applies to show $\operatorname{Tr} B_-^d$ is concentrated, and

$$\operatorname{Tr} B_+^d - \operatorname{Tr} B_-^d = 4d \operatorname{Re} \operatorname{Tr} \left( A_1 A_2 \cdots A_d \right).$$

□

# A stronger result for bounded entries

For low-degree polynomials of random matrices with independent bounded entries, a completely different argument yields the following.

## Theorem (M.)

*If* $X^{(1)}, \ldots, X^{(m)}$ *have independent real entries such that* $\mathbb{E}X^{(k)} = 0$ *and* $\left|X_{ij}^{(k)}\right| \le c$, *then*

$$\mathbb{P}\big[\,|Y_P - \mathbb{E}Y_P| \ge t\big] \le C_m \exp\big[-c_m n^2 t^2\big]$$

*if* $d = 2$, *and*

$$\mathbb{P}\big[\,|Y_P - \mathbb{E}Y_P| \ge t\big] \le C'_m \exp\big[-c'_m n t^2\big]$$

*if* $d = 3$.

# The end

Thank you.