Polynomials and (finite) free probability

Adam W. Marcus

Princeton University
adam.marcus@princeton.edu

Polynomials and (finite) free probability

(with color commentary by Zoidberg)

Adam W. Marcus

Princeton University
adam.marcus@princeton.edu

Acknowledgements:

Research aided tremendously by:

Beyond Kadison-Singer: paving and consequences (AIM)

Hot Topics: Kadison-Singer, Interlacing Polynomials, and Beyond (MSRI)

Outline

Motivation

Recently, I have been interested in self-adjoint linear operators.

Motivation

Recently, I have been interested in self-adjoint linear operators.

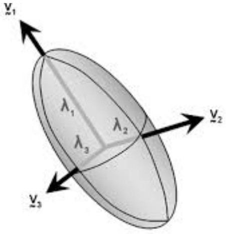
Algebraically, think: real, square, symmetric, matrices.

Motivation

Recently, I have been interested in self-adjoint linear operators.

Algebraically, think: real, square, symmetric, matrices.

Geometrically, think: image of the unit ball is an ellipse.

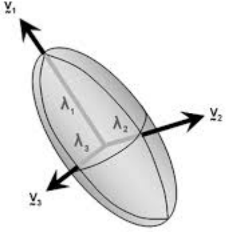


Motivation

Recently, I have been interested in self-adjoint linear operators.

Algebraically, think: real, square, symmetric, matrices.

Geometrically, think: image of the unit ball is an ellipse.



The λ are called *eigenvalues* and the v their associated *eigenvectors*.

Eigenvalues

Theorem (Spectral Decomposition)

Any $d \times d$ real symmetric matrix A can be decomposed as

$$A = \sum_{i=1}^{d} \lambda_i v_i v_i^T$$

where the v_i are orthonormal and each pair (λ_i, v_i) is an eigenpair.

Eigenvalues

Theorem (Spectral Decomposition)

Any $d \times d$ real symmetric matrix A can be decomposed as

$$A = \sum_{i=1}^{d} \lambda_i v_i v_i^T$$

where the v_i are orthonormal and each pair (λ_i, v_i) is an eigenpair.

In particular, if λ_{max} is the largest eigenvalue (in absolute value), then

$$\max_{x : \|x\| = 1} \|Ax\| = \lambda_{max}$$

and if λ_{min} is the smallest (in absolute value)

$$\min_{x : \|x\| = 1} \|Ax\| = \lambda_{min}$$

Frames

The number of non-zero eigenvalues of A is called the *rank*.

Frames

The number of non-zero eigenvalues of A is called the *rank*.

The spectral decomposition is a *rank-1 decomposition*. General rank-1 decompositions $V = \sum_i v_i v_i^T$ are called *frames*.

Frames

The number of non-zero eigenvalues of A is called the *rank*.

The spectral decomposition is a *rank-1 decomposition*. General rank-1 decompositions $V = \sum_i v_i v_i^T$ are called *frames*.

When the $\widehat{v}_i$ are random vectors, then $\widehat{V} = \sum_i \widehat{v}_i \widehat{v}_i^T$ is a random matrix.

Frames

The number of non-zero eigenvalues of A is called the *rank*.

The spectral decomposition is a *rank-1 decomposition*. General rank-1 decompositions $V = \sum_i v_i v_i^T$ are called *frames*.

When the $\widehat{v}_i$ are random vectors, then $\widehat{V} = \sum_i \widehat{v}_i \widehat{v}_i^T$ is a random matrix.

Example

If $\widehat{u}^T \in \{[1,0],[1,1]\}$ and $\widehat{v}^T \in \{[0,1],[1,1]\}$ with independent uniform distributions, then

$$\widehat{u}\widehat{u}^T + \widehat{v}\widehat{v}^T \in \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}, \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} \right\}$$

each with probability $1/4$.

Known tools

Well-known techniques exist for bounding eigenvalues of random frames.

Theorem (Matrix Chernoff)

Let $\widehat{v}_1, \dots, \widehat{v}_n$ *be independent random vectors with* $\|\widehat{v}_i\| \leq 1$ *and* $\sum_i \widehat{v}_i \widehat{v}_i^T = \widehat{V}$. *Then*

$$\mathbb{P}\left[\lambda_{max}(\widehat{V}) \leq \theta\right] \geq 1 - d \cdot e^{-nD(\theta \| \lambda_{max}(\mathbb{E}\widehat{V}))}$$

Similar inequalities by Rudelson (1999), Ahlswede–Winter (2002).

Known tools

Well-known techniques exist for bounding eigenvalues of random frames.

Theorem (Matrix Chernoff)

Let $\widehat{v}_1, \ldots, \widehat{v}_n$ be independent random vectors with $\|\widehat{v}_i\| \leq 1$ and $\sum_i \widehat{v}_i \widehat{v}_i^T = \widehat{V}$. Then

$$\mathbb{P}\left[\lambda_{max}(\widehat{V}) \leq \theta\right] \geq 1 - d \cdot e^{-nD(\theta \| \lambda_{max}(\mathbb{E}\widehat{V}))}$$

Similar inequalities by Rudelson (1999), Ahlswede–Winter (2002).

All such inequalities have two things in common:

1. They give results with *high probability*
2. The bounds depend on the dimension

This will *always* be true — tight concentration (in this respect) depends on the dimension (consider n/d copies of basis vectors).

New method

MSS developed a new technique for getting eigenvalue bounds called the *method of interlacing polynomials*.

New method

MSS developed a new technique for getting eigenvalue bounds called the *method of interlacing polynomials*.

Theorem (MSS; 13)

Let $\widehat{V} = \sum_i \widehat{v}_i \widehat{v}_i^T$ *be a random frame where all* $\widehat{v}_i$ *have finite support and are mutually independent. Now let*

$$p(x) = \mathbb{E}\left\{ \det \left[xI - \widehat{V} \right] \right\}$$

be its expected characteristic polynomial.

New method

MSS developed a new technique for getting eigenvalue bounds called the *method of interlacing polynomials*.

Theorem (MSS; 13)

Let $\widehat{V} = \sum_i \widehat{v}_i \widehat{v}_i^T$ *be a random frame where all* $\widehat{v}_i$ *have finite support and are mutually independent. Now let*

$$p(x) = \mathbb{E}\left\{\det\left[xI - \widehat{V}\right]\right\}$$

be its expected characteristic polynomial. Then

1. p *has all real roots* $r_1 \leq \cdots \leq r_m$*,*
2. *For all* $0 \leq k \leq m$*, we have*

$$\mathbb{P}\left[\lambda_k(\widehat{V}) \leq r_k\right] > 0 \qquad \textit{and} \qquad \mathbb{P}\left[\lambda_k(\widehat{V}) \geq r_k\right] > 0$$

And it works

By trading “high probability” for “nonzero probability”, the method is able to prove bounds *independent* of dimension.

And it works

By trading "high probability" for "nonzero probability", the method is able to prove bounds *independent* of dimension.

Applications have included:

1. Ramanujan graphs:
    - Of all degrees, using 2-lifts (MSS; 13)
    - Of all degrees, using k-lifts (Hall, Puder, Sawin; 14)
    - Of all degrees and sizes, using matchings (MSS; 15)
2. Functional Analysis:
    - Kadison–Singer (and equivalents) (MSS; 13)
    - Lyapunov theorems (Akemann, Weaver; 14)
3. Approximation algorithms:
    - Asymmetric Traveling Salesman (Anari, Oveis-Gharan; 15)

And it works

By trading "high probability" for "nonzero probability", the method is able to prove bounds *independent* of dimension.

Applications have included:

1. Ramanujan graphs:
   - Of all degrees, using 2-lifts (MSS; 13)
   - Of all degrees, using k-lifts (Hall, Puder, Sawin; 14)
   - Of all degrees and sizes, using matchings (MSS; 15)
2. Functional Analysis:
   - Kadison–Singer (and equivalents) (MSS; 13)
   - Lyapunov theorems (Akemann, Weaver; 14)
3. Approximation algorithms:
   - Asymmetric Traveling Salesman (Anari, Oveis-Gharan; 15)

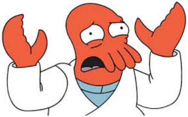
Numerous applications of Kadison–Singer and paving bounds as well.

Big question

Inquiring minds want to know:

Big question

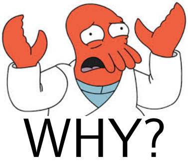
Inquiring minds want to know:



WHY?

Big question

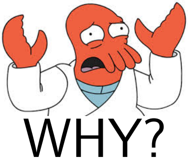
Inquiring minds want to know:



Why should any of this work? Why are all of these polynomials real rooted?

Big question

Inquiring minds want to know:

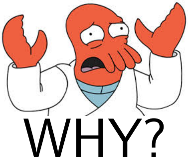


Why should any of this work? Why are all of these polynomials real rooted?

Why would the expected characteristic polynomial (of all things) provide decent bounds on anything worth bounding?

Big question

Inquiring minds want to know:



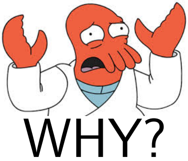
Why should any of this work? Why are all of these polynomials real rooted?

Why would the expected characteristic polynomial (of all things) provide decent bounds on anything worth bounding?

Finally, is there some way to know when the "method of interlacing polynomials" could work?

Big question

Inquiring minds want to know:



Why should any of this work? Why are all of these polynomials real rooted?

Why would the expected characteristic polynomial (of all things) provide decent bounds on anything worth bounding?

Finally, is there some way to know when the "method of interlacing polynomials" could work?

This talk: introduce a new theory that answers these questions (and more).

Outline

1. Introduction
2. Polynomial Convolutions
   - The issue with the characteristic map
   - The issue with maximum roots
3. Free probability
4. The Intersection
   - General ideas
   - Connecting polynomials and free probability
5. Application: Restricted Invertibility

Expected characteristic polynomials

"Prior to the work of [MSS], I think it is safe to say that the conventional wisdom in random matrix theory was that the representation

$$\|A\|_{op} = \operatorname{maxroot}\left(\det\left[xI - A\right]\right)$$

was not particularly useful, due to the highly non-linear nature of both the characteristic polynomial map $A \mapsto \det\left[xI - A\right]$ and the maximum root map $p \mapsto \operatorname{maxroot}(p)$."

Expected characteristic polynomials

"Prior to the work of [MSS], I think it is safe to say that the conventional wisdom in random matrix theory was that the representation

$$\|A\|_{op} = \mathrm{maxroot}\left(\det\left[xI - A\right]\right)$$

was not particularly useful, due to the highly non-linear nature of both the characteristic polynomial map $A \mapsto \det\left[xI - A\right]$ and the maximum root map $p \mapsto \mathrm{maxroot}(p)$."

"For instance, a fact as basic as the triangle inequality

$$\|A + B\|_{op} \leq \|A\|_{op} + \|B\|_{op}$$

is extremely difficult to establish [using characteristic polynomials]."

Expected characteristic polynomials

"Prior to the work of [MSS], I think it is safe to say that the conventional wisdom in random matrix theory was that the representation

$$\|A\|_{op} = \text{maxroot}\left(\det\left[xI - A\right]\right)$$

was not particularly useful, due to the highly non-linear nature of both the characteristic polynomial map $A \mapsto \det\left[xI - A\right]$ and the maximum root map $p \mapsto \text{maxroot}(p)$."

"For instance, a fact as basic as the triangle inequality

$$\|A + B\|_{op} \leq \|A\|_{op} + \|B\|_{op}$$

is extremely difficult to establish [using characteristic polynomials]."

Both are legitimate problems, but for different reasons.

The characteristic map

The problem with $A \mapsto \det[xI - A]$ is that it loses information (the rotation of A).

The characteristic map

The problem with $A \mapsto \det[xI - A]$ is that it loses information (the rotation of A).

So instead consider a rotation invariant operation:

Definition

For $m \times m$ symmetric matrices A and B with characteristic polynomials

$$p(x) = \det[xI - A] \qquad \text{and} \qquad q(x) = \det[xI - B],$$

the *symmetric additive convolution* of p and q is defined as

$$[p \boxplus_m q](x) = \mathbb{E}_Q \left\{ \det \left[xI - A - QBQ^T \right] \right\}$$

where the expectation is taken over orthonormal matrices Q distributed uniformly (via the Haar measure).

Some properties

For degree m polynomials p, q, we have

$$[p \boxplus_m q](x + y) = \sum_{i=0}^{m} p^{(i)}(x) q^{(m-i)}(y).$$

So symmetric and linear!

Some properties

For degree m polynomials p, q, we have

$$[p \boxplus_m q](x + y) = \sum_{i=0}^{m} p^{(i)}(x) q^{(m-i)}(y).$$

So symmetric and linear!

For any linear differential operator $R = \sum_i \alpha_i \partial^i$, we have

$$R\{[p \boxplus_m q]\} = [R\{p\} \boxplus_m q] = [p \boxplus_m R\{q\}]$$

So the algebra $(\mathbb{C}_{\leq m}[x], \boxplus_m)$ is isomorphic to $(\mathbb{C}[\partial] \mod [\partial^{m+1}], \times)$.

Some properties

For degree m polynomials p, q, we have

$$[p \boxplus_m q](x + y) = \sum_{i=0}^{m} p^{(i)}(x) q^{(m-i)}(y).$$

So symmetric and linear!

For any linear differential operator $R = \sum_i \alpha_i \partial^i$, we have

$$R\{[p \boxplus_m q]\} = [R\{p\} \boxplus_m q] = [p \boxplus_m R\{q\}]$$

So the algebra $(\mathbb{C}_{\leq m}[x], \boxplus_m)$ is isomorphic to $(\mathbb{C}[\partial] \mod [\partial^{m+1}], \times)$.

Lemma (Borcea, Brändén)

If p and q have all real roots, then $[p \boxplus_m q]$ has all real roots.

So (when real rooted), we get an easy triangle inequality.

Max roots

The second issue is the maximum root — this time the problem lies in stability.

Max roots

The second issue is the maximum root — this time the problem lies in stability.

Let $p(x) = x^{m-1}(x-1)$ and $q(x) = x(x-1)^{m-1}$. So

$$\text{maxroot}\,(p) = \text{maxroot}\,(q) = 1.$$

Max roots

The second issue is the maximum root — this time the problem lies in stability.

Let $p(x) = x^{m-1}(x-1)$ and $q(x) = x(x-1)^{m-1}$. So

$$\mathrm{maxroot}\,(p) = \mathrm{maxroot}\,(q) = 1.$$

But then

1. $\mathrm{maxroot}\,([p \boxplus_m p]) = 1 + \sqrt{1/m}$
2. $\mathrm{maxroot}\,([p \boxplus_m q]) = 1 + \sqrt{1 - 1/m}$

The triangle inequality says it can be at most 2. 

Max roots

Solution: use smoother version of the maxroot () function.

Max roots

Solution: use smoother version of the maxroot () function.

Definition

For a real rooted polynomial p, we define

$$\alpha\max(p) = \operatorname{maxroot}(p - \alpha p').$$

So $\alpha = 0$ is the usual maxroot () function (and grows with α).

Max roots

Solution: use smoother version of the maxroot () function.

Definition

For a real rooted polynomial p, we define

$$\alpha\mathrm{max}\,(p) = \mathsf{maxroot}\,(p - \alpha p')\,.$$

So $\alpha = 0$ is the usual maxroot () function (and grows with α).

Can we understand the $\alpha\mathrm{max}\,()$ function?

Brief aside

If you recall the barrier function of Batson, Spielman, Srivastava.

$$\Phi_p(x) = \partial \log p(x) = \frac{p'(x)}{p(x)}$$

defined for x above the largest root of (real rooted) p.

Brief aside

If you recall the barrier function of Batson, Spielman, Srivastava.

$$\Phi_p(x) = \partial \log p(x) = \frac{p'(x)}{p(x)}$$

defined for x above the largest root of (real rooted) p.

$$\begin{aligned}\alpha\mathrm{max}\,(p) = x &\iff \mathrm{maxroot}\,\left(p - \alpha p'\right) = x \\ &\iff p(x) - \alpha p'(x) = 0 \\ &\iff \frac{p'(x)}{p(x)} = \frac{1}{\alpha} \\ &\iff \Phi_p(x) = \frac{1}{\alpha}\end{aligned}$$

Brief aside

If you recall the barrier function of Batson, Spielman, Srivastava.

$$\Phi_p(x) = \partial \log p(x) = \frac{p'(x)}{p(x)}$$

defined for x above the largest root of (real rooted) p.

$$\begin{aligned} \alpha\mathrm{max}\,(p) = x &\iff \text{maxroot}\left(p - \alpha p'\right) = x \\ &\iff p(x) - \alpha p'(x) = 0 \\ &\iff \frac{p'(x)}{p(x)} = \frac{1}{\alpha} \\ &\iff \Phi_p(x) = \frac{1}{\alpha} \end{aligned}$$

That is, we are implicitly studying the barrier function.

Some max root results

If p is a degree m, real rooted polynomial, μ_p the average of its roots:

Lemma

$$1 \leq \frac{\partial}{\partial \alpha} \alpha\max(p) \leq 1 + \frac{m-2}{m+2}$$

Proof uses implicit differentiation and Newton inequalities.

Some max root results

If p is a degree m, real rooted polynomial, μ_p the average of its roots:

Lemma

$$1 \leq \frac{\partial}{\partial \alpha} \alpha\max(p) \leq 1 + \frac{m-2}{m+2}$$

Proof uses implicit differentiation and Newton inequalities.

Lemma

$$\alpha\max(p') \leq \alpha\max(p) - \alpha$$

Proof uses concavity of p/p' for $x \geq \mathrm{maxroot}(p)$.

Some max root results

If p is a degree m, real rooted polynomial, μ_p the average of its roots:

Lemma

$$1 \leq \frac{\partial}{\partial\alpha}\alpha\mathrm{max}\,(p) \leq 1 + \frac{m-2}{m+2}$$

Proof uses implicit differentiation and Newton inequalities.

Lemma

$$\alpha\mathrm{max}\,(p') \leq \alpha\mathrm{max}\,(p) - \alpha$$

Proof uses concavity of p/p' for $x \geq \mathrm{maxroot}\,(p)$.

Corollary

$$\mu_p \leq \alpha\mathrm{max}\,(p) - m\alpha \leq \mathrm{maxroot}\,(p)$$

Iterate the previous lemma $(m-1)$ times.

Main inequality

Theorem

Let p and q be degree m real rooted polynomials. Then

$$\alpha\max(p \boxplus_m q) \leq \alpha\max(p) + \alpha\max(q) - m\alpha$$

with equality if and only if p or q has a single distinct root.

Proof uses previous lemmas, induction on m, and "pinching".

Main inequality

Theorem

Let p and q be degree m real rooted polynomials. Then

$$\alpha\max(p \boxplus_m q) \leq \alpha\max(p) + \alpha\max(q) - m\alpha$$

with equality if and only if p or q has a single distinct root.

Proof uses previous lemmas, induction on m, and “pinching”.

Applying this to $p(x) = x^{m-1}(x-1)$ and $q(x) = x(x-1)^{m-1}$ gives

	$\mathrm{maxroot}(\cdot)$	best α in Theorem
$[p \boxplus_m p]$	$1 + 1/\sqrt{m}$	$\approx 1 + 2/\sqrt{m}$
$[p \boxplus_m q]$	$1 + \sqrt{1 - 1/m}$	2

Quick Review

We want to be able to work with expected characteristic polynomials, and had three concerns:

1. the real rootedness
2. the behavior of the map $A \mapsto \det[xI - A]$
3. the behavior of the map $p \mapsto \mathrm{maxroot}(p)$.

Quick Review

We want to be able to work with expected characteristic polynomials, and had three concerns:

1. the real rootedness
2. the behavior of the map $A \mapsto \det[xI - A]$
3. the behavior of the map $p \mapsto \text{maxroot}(p)$.

We addressed the first two by introducing a new convolution which is rotation independent and preserves real rootedness.

Quick Review

We want to be able to work with expected characteristic polynomials, and had three concerns:

1. the real rootedness
2. the behavior of the map $A \mapsto \det[xI - A]$
3. the behavior of the map $p \mapsto \text{maxroot}(p)$.

We addressed the first two by introducing a new convolution which is rotation independent and preserves real rootedness.

We addressed the third by using a smooth version of the maximum root function.

Quick Review

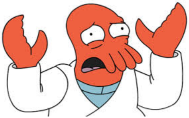
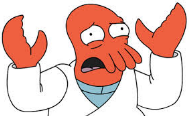
We want to be able to work with expected characteristic polynomials, and had three concerns:

1. the real rootedness
2. the behavior of the map $A \mapsto \det[xI - A]$
3. the behavior of the map $p \mapsto \operatorname{maxroot}(p)$.

We addressed the first two by introducing a new convolution which is rotation independent and preserves real rootedness.

We addressed the third by using a smooth version of the maximum root function.

On the other hand, we have more explaining to do:

 $-3 + 2 =$ 

Outline

Definition

A *von Neumann algebra* M on a Hilbert space H is a unital subalgebra of the space $B(H)$ of bounded operators so that

1. $T \in M \to T^* \in M$
2. $T_i \in M, \langle T_i u, v \rangle \to \langle Tu, v \rangle$ for all u, v implies $T \in M$ (closed on weak operator topology).

Definition

A *von Neumann algebra* M on a Hilbert space H is a unital subalgebra of the space $B(H)$ of bounded operators so that

1. $T \in M \to T^* \in M$
2. $T_i \in M, \langle T_i u, v \rangle \to \langle Tu, v \rangle$ for all u, v implies $T \in M$ (closed on weak operator topology).

We will designate a linear functional $\tau : M \to \mathbb{C}$ that is

1. continuous in the weak operator topology
2. unital: $\tau(\mathbb{1}) = 1$
3. positive: $\tau(T^* T) \geq 0$
4. tracial: $\tau(ST) = \tau(TS)$ for all $S, T \in M$.

to be the special *trace function* (we assume at least one exists).

Definition

A *von Neumann algebra* M on a Hilbert space H is a unital subalgebra of the space $B(H)$ of bounded operators so that

1. $T \in M \to T^* \in M$
2. $T_i \in M, \langle T_i u, v\rangle \to \langle Tu, v\rangle$ for all u, v implies $T \in M$ (closed on weak operator topology).

We will designate a linear functional $\tau : M \to \mathbb{C}$ that is

1. continuous in the weak operator topology
2. unital: $\tau(\mathbb{1}) = 1$
3. positive: $\tau(T^*T) \geq 0$
4. tracial: $\tau(ST) = \tau(TS)$ for all $S, T \in M$.

to be the special *trace function* (we assume at least one exists).

Example

1. $M = L^\infty(X, \mu)$, with $\tau(T) = \int T d\mu \ (= \mathbb{E}_\mu\{T\})$
2. $M = M_{n\times n}$ with $\tau(T) = \frac{1}{n}\mathrm{Tr}\,[T]$

Random variables

Each operator $T \in (M, \tau)$ defines a probability distribution μ_T on $\mathbb{C}$ by

$$\mu_T(U) = \tau(\delta_U(T))$$

for each Borel set $U \subseteq \mathbb{C}$ (δ_U is a WOT limit of polynomials, so $\delta_U(T) \in M$).

Random variables

Each operator $T \in (M, \tau)$ defines a probability distribution μ_T on $\mathbb{C}$ by

$$\mu_T(U) = \tau(\delta_U(T))$$

for each Borel set $U \subseteq \mathbb{C}$ (δ_U is a WOT limit of polynomials, so $\delta_U(T) \in M$).

When T is self adjoint, μ_T is a measure on the real line.

Random variables

Each operator $T \in (M, \tau)$ defines a probability distribution μ_T on $\mathbb{C}$ by

$$\mu_T(U) = \tau(\delta_U(T))$$

for each Borel set $U \subseteq \mathbb{C}$ (δ_U is a WOT limit of polynomials, so $\delta_U(T) \in M$).

When T is self adjoint, μ_T is a measure on the real line.

We will think of T is (some sort of) noncommutative random variable.

This generalizes the idea of a (classical) random variable.

Examples

Classic random variables:

- $M = L^\infty(X, \mu)$
- $\tau(T) = \int T d\mu \; (= \mathbb{E}\{T\})$
- $\mu_T = T_* \mu$ (the usual notion of a distribution on T)

Examples

Classic random variables:

- $M = L^{\infty}(X, \mu)$
- $\tau(T) = \int T d\mu \ (= \mathbb{E}\{T\})$
- $\mu_T = T_* \mu$ (the usual notion of a distribution on T)

Nonrandom matrices:

- $M = M_{n \times n}$
- $\tau(T) = \frac{1}{n} \mathrm{Tr}\,[T]$
- $\mu_T = \frac{1}{n} \sum_i \delta_{\lambda_i(T)}$

Examples

Classic random variables:

- $M = L^\infty(X, \mu)$
- $\tau(T) = \int T d\mu \;(= \mathbb{E}\{T\})$
- $\mu_T = T_* \mu$ (the usual notion of a distribution on T)

Nonrandom matrices:

- $M = M_{n \times n}$
- $\tau(T) = \frac{1}{n} \mathrm{Tr}\,[T]$
- $\mu_T = \frac{1}{n} \sum_i \delta_{\lambda_i(T)}$

Random matrices:

- $M = L^\infty \otimes M_{n \times n}$
- $\tau(T) = \frac{1}{n}\, \mathbb{E}\{\mathrm{Tr}\,[T]\}$
- $\mu_T = \frac{1}{n}\, \mathbb{E}\left\{\sum_i \delta_{\lambda_i(T)}\right\}$

Independence

Independence is a special joint distribution that allows one to reduce mixed traces to simpler ones.

Independence

Independence is a special joint distribution that allows one to reduce mixed traces to simpler ones.

Definition

T and S are *independent* if

1. $TS = ST$
2. $\tau(p(T)q(S)) = \tau(p(T))\tau(q(S))$ for all polynomials p, q

Independence

Independence is a special joint distribution that allows one to reduce mixed traces to simpler ones.

Definition

T and S are *independent* if

1. $TS = ST$
2. $\tau(p(T)q(S)) = \tau(p(T))\tau(q(S))$ for all polynomials p, q

Example

For $\tau = \mathbb{E}\{\}$ and X, Y independent classical random variables,

1. $\tau(X^2Y^2) = \tau(X^2)\tau(Y^2)$
2. $\tau(XYXY) = \tau(X^2)\tau(Y^2)$

Independence

Independence is a special joint distribution that allows one to reduce mixed traces to simpler ones.

Definition

T and S are *independent* if

1. $TS = ST$
2. $\tau(p(T)q(S)) = \tau(p(T))\tau(q(S))$ for all polynomials p, q

Example

For $\tau = \mathbb{E}\{\}$ and X, Y independent classical random variables,

1. $\tau(X^2Y^2) = \tau(X^2)\tau(Y^2)$
2. $\tau(XYXY) = \tau(X^2)\tau(Y^2)$

What's the point of being noncommutative!?!

Free Independence

Definition

T and S are called *freely independent* if

$$\tau(p_1(T)q_1(S)p_2(T)q_2(S)\dots p_m(T)q_m(S)) = 0$$

whenever $\tau(p_j(T)) = \tau(q_j(S)) = 0$ for all j.

Free Independence

Definition

T and S are called *freely independent* if

$$\tau(p_1(T)q_1(S)p_2(T)q_2(S)\dots p_m(T)q_m(S)) = 0$$

whenever $\tau(p_j(T)) = \tau(q_j(S)) = 0$ for all j.

Example

For S, T freely independent,

1. $\tau(T^2S^2) = \tau(T^2)\tau(S^2)$
2. $\tau(TSTS) = \tau(T^2)\tau(S)^2 + \tau(S^2)\tau(T)^2 - \tau(S)^2\tau(T)^2$

Free Independence

Definition

T and S are called *freely independent* if

$$\tau(p_1(T)q_1(S)p_2(T)q_2(S)\dots p_m(T)q_m(S)) = 0$$

whenever $\tau(p_j(T)) = \tau(q_j(S)) = 0$ for all j.

Example

For S, T freely independent,

1. $\tau(T^2S^2) = \tau(T^2)\tau(S^2)$
2. $\tau(TSTS) = \tau(T^2)\tau(S)^2 + \tau(S^2)\tau(T)^2 - \tau(S)^2\tau(T)^2$

Proof:

Let $S_0 = S - \tau(S)\mathbb{1}$ and $T_0 = T - \tau(T)\mathbb{1}$, so $\tau(S_0) = \tau(T_0) = 0$.

Free Independence

Definition

T and S are called *freely independent* if

$$\tau(p_1(T)q_1(S)p_2(T)q_2(S)\dots p_m(T)q_m(S)) = 0$$

whenever $\tau(p_j(T)) = \tau(q_j(S)) = 0$ for all j.

Example

For S, T freely independent,

1. $\tau(T^2S^2) = \tau(T^2)\tau(S^2)$
2. $\tau(TSTS) = \tau(T^2)\tau(S)^2 + \tau(S^2)\tau(T)^2 - \tau(S)^2\tau(T)^2$

Proof:

Let $S_0 = S - \tau(S)\mathbb{1}$ and $T_0 = T - \tau(T)\mathbb{1}$, so $\tau(S_0) = \tau(T_0) = 0$.

By free independence, $\tau(T_0S_0T_0S_0) = 0$, now substitute and use linearity.

Convolutions

Given r.v. $A \sim \mu_A$ and $B \sim \mu_B$, what is distribution of $A + B$?

Ill defined question (regardless of commutativity)!

Requires knowing the joint distribution!

Convolutions

Given r.v. $A \sim \mu_A$ and $B \sim \mu_B$, what is distribution of $A + B$?

Ill defined question (regardless of commutativity)!

Requires knowing the joint distribution!

However, we know two "special" joint distributions:

Definition

Let μ and ρ be probability distributions with $X \sim \mu$ and $Y \sim \rho$. The

1. *additive convolution* $\mu \oplus \rho$ is the distribution of $X + Y$ in the case that X, Y are independent.
2. *free additive convolution* $\mu \boxplus \rho$ is the distribution of $X + Y$ in the case that X, Y are freely independent.

Now how can we compute such things?

Computation

To compute the (classical) additive convolution, one uses the *moment generating function*

$$M_\mu(t) = \mathbb{E}_{X\sim\mu}\left\{ e^{tX} \right\}$$

Computation

To compute the (classical) additive convolution, one uses the *moment generating function*

$$M_\mu(t) = \mathbb{E}_{X \sim \mu}\left\{ e^{tX} \right\}$$

to form the *cumulant generating function*

$$K_\mu(t) = \log M_\mu(t)$$

Computation

To compute the (classical) additive convolution, one uses the *moment generating function*

$$M_\mu(t) = \mathbb{E}_{X\sim\mu}\left\{ e^{tX} \right\}$$

to form the *cumulant generating function*

$$K_\mu(t) = \log M_\mu(t)$$

and then adds them

$$K_{\mu\oplus\rho}(t) = K_\mu(t) + K_\rho(t)$$

Computation

To compute the (classical) additive convolution, one uses the *moment generating function*

$$M_\mu(t) = \mathbb{E}_{X\sim\mu}\left\{e^{tX}\right\}$$

to form the *cumulant generating function*

$$K_\mu(t) = \log M_\mu(t)$$

and then adds them

$$K_{\mu\oplus\rho}(t) = K_\mu(t) + K_\rho(t)$$

and then reverses

$$M_{\mu\oplus\rho}(t) = e^{K_{\mu\oplus\rho}(t)}.$$

Only computable up to moments!

Free Computation

To compute the free additive convolution, one uses the *Cauchy transform*

$$\mathcal{G}_{\mu_A}(t) = \int \frac{\mu_A(x)}{t-x} dx = \tau\left((t\mathbb{1} - A)^{-1}\right)$$

Free Computation

To compute the free additive convolution, one uses the *Cauchy transform*

$$\mathcal{G}_{\mu_A}(t) = \int \frac{\mu_A(x)}{t - x} dx = \tau\left((t\mathbb{1} - A)^{-1}\right)$$

to form the *R-transform*

$$\mathcal{R}_{\mu_A}(t) = \mathcal{G}_{\mu_A}^{-1}(t) - \frac{1}{t}$$

Free Computation

To compute the free additive convolution, one uses the *Cauchy transform*

$$\mathcal{G}_{\mu_A}(t) = \int \frac{\mu_A(x)}{t - x} dx = \tau\big((t\mathbb{1} - A)^{-1}\big)$$

to form the *R-transform*

$$\mathcal{R}_{\mu_A}(t) = \mathcal{G}_{\mu_A}^{-1}(t) - \frac{1}{t}$$

and then adds them

$$\mathcal{R}_{\mu_A \boxplus \mu_B}(t) = \mathcal{R}_{\mu_A}(t) + \mathcal{R}_{\mu_B}(t)$$

Free Computation

To compute the free additive convolution, one uses the *Cauchy transform*

$$\mathcal{G}_{\mu_A}(t) = \int \frac{\mu_A(x)}{t-x} dx = \tau\left((t\mathbb{1} - A)^{-1}\right)$$

to form the *R-transform*

$$\mathcal{R}_{\mu_A}(t) = \mathcal{G}_{\mu_A}^{-1}(t) - \frac{1}{t}$$

and then adds them

$$\mathcal{R}_{\mu_A \boxplus \mu_B}(t) = \mathcal{R}_{\mu_A}(t) + \mathcal{R}_{\mu_B}(t)$$

and reverses.

Free Computation

To compute the free additive convolution, one uses the *Cauchy transform*

$$\mathcal{G}_{\mu_A}(t) = \int \frac{\mu_A(x)}{t - x} dx = \tau\left((t\mathbb{1} - A)^{-1}\right)$$

to form the *R-transform*

$$\mathcal{R}_{\mu_A}(t) = \mathcal{G}_{\mu_A}^{-1}(t) - \frac{1}{t}$$

and then adds them

$$\mathcal{R}_{\mu_A \boxplus \mu_B}(t) = \mathcal{R}_{\mu_A}(t) + \mathcal{R}_{\mu_B}(t)$$

and reverses.

Note $\frac{1}{t} = \mathcal{G}_{\mu_0}^{-1}(t)$.

Free probability

Voiculescu developed an entire theory (constructed all of the spaces, showed everything converges, etc) which he called *free probability*.

Free probability

Voiculescu developed an entire theory (constructed all of the spaces, showed everything converges, etc) which he called *free probability*.

Furthermore, he showed a link between classical and free independence.

Theorem

Let $\{A_n\}$ and $\{B_n\}$ be sequences of $n \times n$ random matrices where each entry in each matrix is drawn independently from a standard normal distribution. Then there exist operators $\mathcal{A}$ and $\mathcal{B}$ such that

$$\mu_{A_n} \to \mu_{\mathcal{A}} \quad \text{and} \quad \mu_{B_n} \to \mu_{\mathcal{B}} \quad \text{and} \quad \mu_{A_n+B_n} \to \mu_{\mathcal{A}} \boxplus \mu_{\mathcal{B}}$$

in the weak operator topology.

Free probability

Voiculescu developed an entire theory (constructed all of the spaces, showed everything converges, etc) which he called *free probability*.

Furthermore, he showed a link between classical and free independence.

Theorem

Let $\{A_n\}$ and $\{B_n\}$ be sequences of $n \times n$ random matrices where each entry in each matrix is drawn independently from a standard normal distribution. Then there exist operators $\mathcal{A}$ and $\mathcal{B}$ such that

$$\mu_{A_n} \to \mu_{\mathcal{A}} \quad \text{and} \quad \mu_{B_n} \to \mu_{\mathcal{B}} \quad \text{and} \quad \mu_{A_n+B_n} \to \mu_{\mathcal{A}} \boxplus \mu_{\mathcal{B}}$$

in the weak operator topology.

The sequences $\{A_n\}$ and $\{B_n\}$ are called *asymptotically free*.

Free probability

Voiculescu developed an entire theory (constructed all of the spaces, showed everything converges, etc) which he called *free probability*.

Furthermore, he showed a link between classical and free independence.

Theorem

Let $\{A_n\}$ and $\{B_n\}$ be sequences of $n \times n$ random matrices where each entry in each matrix is drawn independently from a standard normal distribution. Then there exist operators $\mathcal{A}$ and $\mathcal{B}$ such that

$$\mu_{A_n} \to \mu_{\mathcal{A}} \quad \text{and} \quad \mu_{B_n} \to \mu_{\mathcal{B}} \quad \text{and} \quad \mu_{A_n+B_n} \to \mu_{\mathcal{A}} \boxplus \mu_{\mathcal{B}}$$

in the weak operator topology.

The sequences $\{A_n\}$ and $\{B_n\}$ are called *asymptotically free*.

Many examples of random matrices now known to be asymptotically free.

Quick Review

In free probability, one thinks of probability distributions μ_A and μ_B living on the spectrum of self adjoint operators A and B.

Then one wants to try to understand μ_{A+B} (for example).

Quick Review

In free probability, one thinks of probability distributions μ_A and μ_B living on the spectrum of self adjoint operators A and B.

Then one wants to try to understand μ_{A+B} (for example).

Free independence is a “special” joint distribution that allows one to compute the moments μ_{A+B} from the moments of μ_A and μ_B (which we know).

In particular, functions of freely independent random variables are rotation independent!!

Quick Review

In free probability, one thinks of probability distributions μ_A and μ_B living on the spectrum of self adjoint operators A and B.

Then one wants to try to understand μ_{A+B} (for example).

Free independence is a "special" joint distribution that allows one to compute the moments μ_{A+B} from the moments of μ_A and μ_B (which we know).

In particular, functions of freely independent random variables are rotation independent!!

This captures "Dysonian" behavior — independence on entries (often) translates to freeness in the spectrum.

Hence it can then be applied to random matrices, but only asymptotically.

Outline

Legendre transform

Definition

Let f be a function that is convex on an interval $X \subseteq \mathbb{R}$. The *Legendre transform* is

$$f^*(s) = \sup_{x \in X} \{xs - f(x)\}$$

Legendre transform

Definition

Let f be a function that is convex on an interval $X \subseteq \mathbb{R}$. The *Legendre transform* is

$$f^*(s) = \sup_{x \in X} \{xs - f(x)\}$$

Lemma

$f'(x)$ and $f^{*\prime}(s)$ are inverses: that is, $f'(f^{*\prime}(x)) = x$

Legendre transform

Definition

Let f be a function that is convex on an interval $X \subseteq \mathbb{R}$. The *Legendre transform* is

$$f^*(s) = \sup_{x \in X} \{xs - f(x)\}$$

Lemma

$f'(x)$ *and* $f^{*\prime}(s)$ *are inverses: that is,* $f'(f^{*\prime}(x)) = x$

Recall that the R-transform is achieved by inverting the Cauchy transform.

This allows us to achieve it via a sup.

L^p norm

Definition

The L_p norm of a function f on a measure space (X, μ) is

$$\|f\|_{L^p(X)} = \left(\int_X |f|^p \, \mathrm{d}\mu \right)^{1/p}$$

L^p norm

Definition

The L_p norm of a function f on a measure space (X, μ) is

$$\|f\|_{L^p(X)} = \left(\int_X |f|^p \, \mathrm{d}\mu \right)^{1/p}$$

Lemma

If f is nonnegative and continuous on X and μ is absolutely continuous with respect to Lebesguese measure, then

$$\lim_{p \to \infty} \|f\|_{L^p(X)} = \sup\{f(x) : x \in X\}$$

L^p norm

Definition

The L_p norm of a function f on a measure space (X, μ) is

$$\|f\|_{L^p(X)} = \left(\int_X |f|^p \, \mathrm{d}\mu \right)^{1/p}$$

Lemma

If f is nonnegative and continuous on X and μ is absolutely continuous with respect to Lebesguese measure, then

$$\lim_{p \to \infty} \|f\|_{L^p(X)} = \sup\{f(x) : x \in X\}$$

This will be our method of convergence.

Fuglede–Kadison determinants

For $n \times n$ positive definite matrix A, recall

$$\det[A] = \exp \mathrm{Tr}[\log A].$$

Fuglede–Kadison determinants

For $n \times n$ positive definite matrix A, recall

$$\det[A] = \exp \mathrm{Tr}[\log A].$$

This idea can be extended to von Neumann algebras:

Definition

Given a von Neumann algebra M and trace function τ, the *Fuglede–Kadison determinant* is defined by

$$\Delta(T) = \exp \tau(\log |T|) = \exp \int \log t \, \mathrm{d}\mu_{|T|}$$

where $|T| = (T^* T)^{1/2}$.

Fuglede–Kadison determinants

For $n \times n$ positive definite matrix A, recall

$$\det[A] = \exp \mathrm{Tr}[\log A].$$

This idea can be extended to von Neumann algebras:

Definition

Given a von Neumann algebra M and trace function τ, the *Fuglede–Kadison determinant* is defined by

$$\Delta(T) = \exp \tau(\log |T|) = \exp \int \log t \, \mathrm{d}\mu_{|T|}$$

where $|T| = (T^* T)^{1/2}$.

Example

For T positive semidefinite in $M_{n \times n}$, $\Delta(T) = (\det[T])^{1/n}$

U Transform

Let S be a multiset of complex numbers.

Claim: there exists a unique multiset T with $|S| = |T|$ such that

$$\prod_{s_i \in S} (x - s_i) = \frac{1}{|T|} \sum_{t_i \in T} (x - t_i)^m.$$

Called the *U transform*.

U Transform

Let S be a multiset of complex numbers.

Claim: there exists a unique multiset T with $|S| = |T|$ such that

$$\prod_{s_i \in S} (x - s_i) = \frac{1}{|T|} \sum_{t_i \in T} (x - t_i)^m.$$

Called the *U transform*.

Proof: Coefficient of x^k is a constraint on $\sum_i t_i^k$ (so $|T|$ constraints).

U Transform

Let S be a multiset of complex numbers.

Claim: there exists a unique multiset T with $|S| = |T|$ such that

$$\prod_{s_i \in S} (x - s_i) = \frac{1}{|T|} \sum_{t_i \in T} (x - t_i)^m.$$

Called the *U transform*.

Proof: Coefficient of x^k is a constraint on $\sum_i t_i^k$ (so $|T|$ constraints).

Newton identities: power sums $\iff$ elementary symmetric polynomials

Unique solution by fundamental theorm of algebra.

Finite transforms

Let A be an $m \times m$ real, symmetric matrix with maximum eigenvalue ρ_A.

Finite transforms

Let A be an $m \times m$ real, symmetric matrix with maximum eigenvalue ρ_A.

Definition

The *m-finite K-transform of* μ_A

$$\mathcal{K}^m_{\mu_A}(s) = -\frac{\partial}{\partial s} \ln \left\| e^{-xs} \Delta(xI - A) \right\|_{L^m(X)}$$
$$= -\frac{1}{m}\frac{\partial}{\partial s} \ln \int_X e^{-mxs} \Delta(xI - A)^m \, \mathrm{d}x$$

where $X = (\rho_A, \infty)$.

The *m-finite R-transform* is

$$\mathcal{R}^m_{\mu_A}(s) = \mathcal{K}^m_{\mu_A}(s) - \mathcal{K}^m_{\mu_0}(s)$$

where μ_0 is the constant 0 distribution.

The connection

Theorem

For all noncommutative random variables A with compact support, we have

$$\lim_{m \to \infty} \mathcal{R}^{m}_{\mu_A}(s) = \mathcal{R}_{\mu_A}(s)$$

Proof uses Legendre transform and convergence of L_p norm. Works for other measures?

The connection

Theorem

For all noncommutative random variables A with compact support, we have

$$\lim_{m\to\infty} \mathcal{R}^m_{\mu_A}(s) = \mathcal{R}_{\mu_A}(s)$$

Proof uses Legendre transform and convergence of L_p norm. Works for other measures?

Theorem

Let A and B be $m \times m$ real symmetric matrices. Then the following are equivalent:

1. $\mathcal{R}^m_{\mu_A}(s) + \mathcal{R}^m_{\mu_B}(s) \equiv \mathcal{R}^m_{\mu_C}(s) \mod [s^m]$
2. $\det[xI - A] \boxplus_m \det[xI - B] = \det[xI - C]$

Proof uses U transform.

Proof sketch

U transform turns polynomial convolutions into classical probability:

Lemma

If Y and Z are independent random variables, then

$$\mathbb{E}\{(x - Y)^m\} \boxplus_m \mathbb{E}\{(x - Z)^m\} = \mathbb{E}\{(x - Y - Z)^m\}.$$

Proof sketch

U transform turns polynomial convolutions into classical probability:

Lemma

If Y and Z are independent random variables, then

$$\mathbb{E}\{(x - Y)^m\} \boxplus_m \mathbb{E}\{(x - Z)^m\} = \mathbb{E}\{(x - Y - Z)^m\}.$$

So $\mathcal{R}^m_{\mu_A}(s)$ must become (linear function of) classical CGF.

Lemma

If A is an $m \times m$ matrix and Y is uniformly distributed over the U transform of $\lambda(A)$, then

$$\mathcal{R}^m_{\mu_A}(s) \equiv \left(\frac{1}{m}\frac{\partial}{\partial s}\log \mathbb{E}\left\{e^{mYs}\right\}\right) \mod [s^m]$$

The connection, ctd.

Theorem

Let A, B, C be $m \times m$ real, symmetric matrices such that

$$\mathcal{R}^m_{\mu_A}(s) + \mathcal{R}^m_{\mu_B}(s) \equiv \mathcal{R}^m_{\mu_C}(s) \mod [s^m].$$

Then for all w,

$$\mathcal{R}_{\mu_C}(w) \leq \mathcal{R}_{\mu_A \boxplus \mu_B}(w)$$

with equality if and only if A or B is a multiple of the identity.

The connection, ctd.

Theorem

Let A, B, C be $m \times m$ real, symmetric matrices such that

$$\mathcal{R}^m_{\mu_A}(s) + \mathcal{R}^m_{\mu_B}(s) \equiv \mathcal{R}^m_{\mu_C}(s) \mod [s^m].$$

Then for all w,

$$\mathcal{R}_{\mu_C}(w) \leq \mathcal{R}_{\mu_A \boxplus \mu_B}(w)$$

with equality if and only if A or B is a multiple of the identity.

Follows from "smoothed" triangle inequality:

$$\mathcal{R}_{\mu_A}\left(\frac{1}{m\alpha}\right) = \alpha \max(p) - m\alpha.$$

when $p(x) = \det[xI - A]$.

The connection, ctd.

Theorem

Let A, B, C be $m \times m$ real, symmetric matrices such that

$$\mathcal{R}^m_{\mu_A}(s) + \mathcal{R}^m_{\mu_B}(s) \equiv \mathcal{R}^m_{\mu_C}(s) \mod [s^m].$$

Then for all w,

$$\mathcal{R}_{\mu_C}(w) \leq \mathcal{R}_{\mu_A \boxplus \mu_B}(w)$$

with equality if and only if A or B is a multiple of the identity.

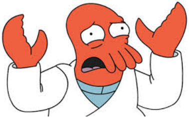
Follows from "smoothed" triangle inequality:

$$\mathcal{R}_{\mu_A}\left(\frac{1}{m\alpha}\right) = \alpha\max(p) - m\alpha.$$

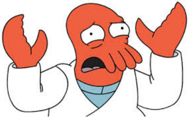
when $p(x) = \det[xI - A]$.

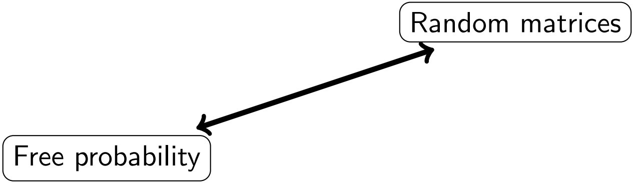
Implies support of finite convolution lies *inside* support of free convolution.

Answers

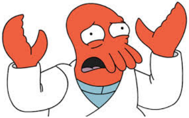


Answers





Answers

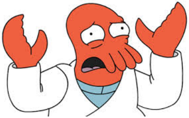


Random matrices

concentration of measure

Free probability

Answers



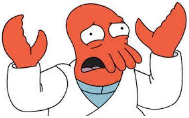
concentration of measure

Random matrices

Free probability

Expected characteristic polynomials

Answers



Random matrices

concentration of measure

Free probability

moments

Expected characteristic polynomials

Answers

Random matrices

concentration of measure

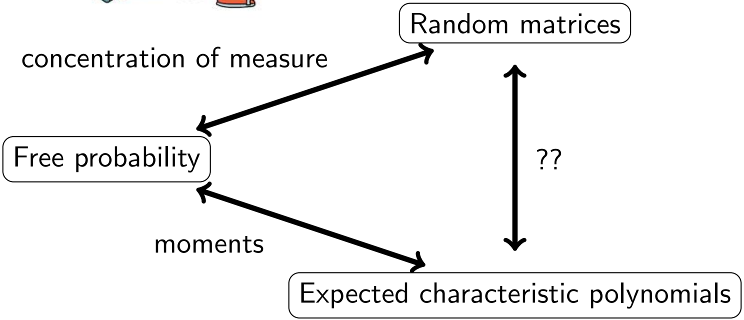
Free probability

moments

Expected characteristic polynomials

Expected characteristic polynomials are a finite approximation of an asymptotic approximation of random matrices.

Answers



Expected characteristic polynomials are a finite approximation of an asymptotic approximation of random matrices.

Remarks

So “method of interlacing polynomials” is transferring an asymptotic bound to a finite one.

Remarks

So “method of interlacing polynomials” is transferring an asymptotic bound to a finite one.

We should expect the finite case to do *better than* the asymptotics (at least when it comes to maximum eigenvalues).

Remarks

So "method of interlacing polynomials" is transferring an asymptotic bound to a finite one.

We should expect the finite case to do *better than* the asymptotics (at least when it comes to maximum eigenvalues).

Similar results for multiplicative convolution.

Remarks

So "method of interlacing polynomials" is transferring an asymptotic bound to a finite one.

We should expect the finite case to do *better than* the asymptotics (at least when it comes to maximum eigenvalues).

Similar results for multiplicative convolution.

Other (known) finite analogues:

1. Limit theorems (Central, Poisson)
2. Dyson Brownian motion
3. Entropy, Fisher information, Cramer–Rao (for one r.v.)

Open directions:

1. Bivariate polynomials (second order freeness?)
2. Entropy (and friends) for joint distributions

Conjecture

Relation to β-ensembles? Let A, B be $m \times m$ matrices with

$$a_1 = \mathrm{tr}\,[A] \qquad \text{and} \qquad a_2 = \mathrm{tr}\left[A^2\right]$$

and same for B.

Conjecture

Relation to β-ensembles? Let A, B be $m \times m$ matrices with

$$a_1 = \mathrm{tr}\,[A] \qquad \text{and} \qquad a_2 = \mathrm{tr}\left[A^2\right]$$

and same for B.

If A and B are freely independent, one gets

$$\mathrm{tr}\left[(AB)^2\right] = a_2 b_1^2 + b_2 a_1^2 - a_1^2 b_1^2 \quad (*)$$

Conjecture

Relation to β-ensembles? Let A, B be $m \times m$ matrices with

$$a_1 = \mathrm{tr}\,[A] \qquad \text{and} \qquad a_2 = \mathrm{tr}\left[A^2\right]$$

and same for B.

If A and B are freely independent, one gets

$$\mathrm{tr}\left[(AB)^2\right] = a_2 b_1^2 + b_2 a_1^2 - a_1^2 b_1^2 \quad (*)$$

If A and B are finite freely independent, one gets

$$\mathrm{tr}\left[(AB)^2\right] = (*) + \frac{1}{m-1}(a_2 - a_1^2)(b_2 - b_1^2) \quad (**)$$

Conjecture

Relation to β-ensembles? Let A, B be $m \times m$ matrices with

$$a_1 = \mathrm{tr}\,[A] \qquad \text{and} \qquad a_2 = \mathrm{tr}\left[A^2\right]$$

and same for B.

If A and B are freely independent, one gets

$$\mathrm{tr}\left[(AB)^2\right] = a_2 b_1^2 + b_2 a_1^2 - a_1^2 b_1^2 \quad (*)$$

If A and B are finite freely independent, one gets

$$\mathrm{tr}\left[(AB)^2\right] = (*) + \frac{1}{m-1}(a_2 - a_1^2)(b_2 - b_1^2) \quad (**)$$

And for β-ensembles, one gets (courtesy of Alan Edelman):

$$\mathbb{E}_Q\left\{\mathrm{tr}\left[(AQ^T BQ)^2\right]\right\} = (**) - \frac{2m}{(m-1)(m\beta+2)}(a_2 - a_1^2)(b_2 - b_1^2)$$

Giving back

Also potential applications:

1. Connes embedding conjecture?

   Asks how well vN algebras can be approximated by finite matrices.

   Likely requires one of the "open directions."

Giving back

Also potential applications:

1. Connes embedding conjecture?

   Asks how well vN algebras can be approximated by finite matrices.

   Likely requires one of the "open directions."

2. Random matrix universality?

   Universality can often be achieved by studying the asymptotic distribution of roots of certain polynomials.

   Which polynomials? Here is a recipe:

   Random matrix
   $\longrightarrow$ *free probability*
   $\longrightarrow$ *free convolutions*
   $\longrightarrow$ *finite free convolutions*
   $\longrightarrow$ *polynomial*

Outline

An application

Example: Restricted invertibility (special case)

Theorem

If $v_1, \dots, v_n \in \mathbb{C}^m$ *are vectors with*

$$\|v_i\|^2 = \frac{m}{n} \qquad \text{and} \qquad \sum_{i=1}^{n} v_i v_i^* = I,$$

then for all $k < n$*, there exists a set* $S \subset [n]$ *with* $|S| = k$ *such that*

$$\lambda_k \left(\sum_{i \in S} v_i v_i^* \right) \geq \left(1 - \sqrt{\frac{k}{m}} \right)^2 \left(\frac{m}{n} \right).$$

First proved by Bourgain and Tzafriri (in more generality, worse constants).

Translation

Translate to random matrices:

Given a random $m \times m$ rotation matrix R, and a random set S of size k, what do you expect the eigenvalue distribution of

$$R[S, \cdot]R[S, \cdot]^*$$

to look like?

Translation

Translate to random matrices:

Given a random $m \times m$ rotation matrix R, and a random set S of size k, what do you expect the eigenvalue distribution of

$$R[S, \cdot]R[S, \cdot]^*$$

to look like?

Similar ensembles are studied in random matrix theory, where they are called *Wishart matrices*.

Translation

Translate to random matrices:

Given a random $m \times m$ rotation matrix R, and a random set S of size k, what do you expect the eigenvalue distribution of

$$R[S, \cdot]R[S, \cdot]^*$$

to look like?

Similar ensembles are studied in random matrix theory, where they are called *Wishart matrices*.

Let's see what random matrix theory has to say.

Wishart matrices

Let X be an $M \times N$ random matrix whose entries are i.i.d. with mean 0 and variance σ^2. Set

$$Y_N = \frac{1}{N} X X^*$$

Wishart matrices

Let X be an $M \times N$ random matrix whose entries are i.i.d. with mean 0 and variance σ^2. Set

$$Y_N = \frac{1}{N} X X^*$$

If $M, N \to \infty$ in such a way that $M/N \to \lambda \in (0, \infty)$, then the asymptotic eigenvalue distribution of the resulting sequence of matrices has density function

$$\mathrm{d}\nu(x) = \frac{1}{2\pi\sigma^2} \frac{\sqrt{(\lambda_+ - x)(x - \lambda_-)}}{\lambda x} \mathbf{1}_{[\lambda_-, \lambda_+]} \, \mathrm{d}x$$

where $\lambda_\pm = \sigma^2(1 \pm \sqrt{\lambda})^2$.

Wishart matrices

Let X be an $M \times N$ random matrix whose entries are i.i.d. with mean 0 and variance σ^2. Set

$$Y_N = \frac{1}{N} X X^*$$

If $M, N \to \infty$ in such a way that $M/N \to \lambda \in (0, \infty)$, then the asymptotic eigenvalue distribution of the resulting sequence of matrices has density function

$$\mathrm{d}\nu(x) = \frac{1}{2\pi\sigma^2} \frac{\sqrt{(\lambda_+ - x)(x - \lambda_-)}}{\lambda x} \mathbf{1}_{[\lambda_-, \lambda_+]} \, \mathrm{d}x$$

where $\lambda_\pm = \sigma^2 (1 \pm \sqrt{\lambda})^2$.

Called the *Marchenko–Pastur distribution*.

Lower bound

In particular, we have the lower edge of the spectrum $\sigma^2(1-\sqrt{\lambda})^2$.

Lower bound

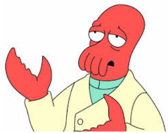
In particular, we have the lower edge of the spectrum $\sigma^2(1-\sqrt{\lambda})^2$.

Already suggests the optimal bound to try for.

Lower bound

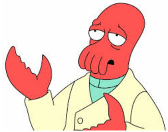
In particular, we have the lower edge of the spectrum $\sigma^2(1 - \sqrt{\lambda})^2$.

Already suggests the optimal bound to try for. But how?



Lower bound

In particular, we have the lower edge of the spectrum $\sigma^2(1-\sqrt{\lambda})^2$.

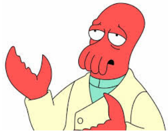
Already suggests the optimal bound to try for. But how? 

If this is going to work, it is because the random matrix acts (asymptotically) like a free distribution.

Lower bound

In particular, we have the lower edge of the spectrum $\sigma^2(1 - \sqrt{\lambda})^2$.

Already suggests the optimal bound to try for. But how?



If this is going to work, it is because the random matrix acts (asymptotically) like a free distribution.

If it acts like a free distribution, it should act like our polynomial convolutions.

To polynomials!

Translate to finite free probability: if

$$p(x) = \det\left[xI - vv^*\right] = x^m - \frac{m}{n}x^{m-1}$$

then

$$[\underbrace{p \boxplus_m p \boxplus_m \cdots \boxplus_m p}_{k \text{ times}}] = m!(-n)^{-m} L_m^{k-m}(nx)$$

where $L_m^{(\alpha)}(x)$ is the (very well studied) Laguerre polynomial.

To polynomials!

Translate to finite free probability: if

$$p(x) = \det\left[xI - vv^*\right] = x^m - \frac{m}{n}x^{m-1}$$

then

$$[\underbrace{p \boxplus_m p \boxplus_m \cdots \boxplus_m p}_{k \text{ times}}] = m!(-n)^{-m} L_m^{k-m}(nx)$$

where $L_m^{(\alpha)}(x)$ is the (very well studied) Laguerre polynomial.

In particular, the smallest nonzero root is (asymptotically)

$$\left(1 - \sqrt{\frac{k}{m}}\right)^2 \left(\frac{m}{n}\right).$$

To polynomials!

Translate to finite free probability: if

$$p(x) = \det\left[xI - vv^*\right] = x^m - \frac{m}{n}x^{m-1}$$

then

$$[\underbrace{p \boxplus_m p \boxplus_m \cdots \boxplus_m p}_{k \text{ times}}] = m!(-n)^{-m} L_m^{k-m}(nx)$$

where $L_m^{(\alpha)}(x)$ is the (very well studied) Laguerre polynomial.

In particular, the smallest nonzero root is (asymptotically)

$$\left(1 - \sqrt{\frac{k}{m}}\right)^2 \left(\frac{m}{n}\right).$$

Same bound can be calculated using $\alpha\max()$ (and picking optimal α).

Interlacing family

Still need to build an interlacing family.

Interlacing family

Still need to build an interlacing family.

Suffices to find distribution on $v_1, \ldots, v_n$ so that

1. Each choice of a vector is independent
2. expected polynomial after k vectors has same roots as $L_m^{k-m}(nx)$

This would imply the probability that one of the polynomials has a larger kth root.

Interlacing family

Still need to build an interlacing family.

Suffices to find distribution on $v_1, \ldots, v_n$ so that

1. Each choice of a vector is independent
2. expected polynomial after k vectors has same roots as $L_m^{k-m}(nx)$

This would imply the probability that one of the polynomials has a larger kth root.

Hence we want to find a (generic) discrete sum that equals the (generic) integral (for some subset of "generic").

Not possible in general, but *is* possible if we restrict to degree m matrices (since integral becomes a fixed degree polynomial).

Interlacing family

Still need to build an interlacing family.

Suffices to find distribution on $v_1, \ldots, v_n$ so that

1. Each choice of a vector is independent
2. expected polynomial after k vectors has same roots as $L_m^{k-m}(nx)$

This would imply the probability that one of the polynomials has a larger kth root.

Hence we want to find a (generic) discrete sum that equals the (generic) integral (for some subset of "generic").

Not possible in general, but *is* possible if we restrict to degree m matrices (since integral becomes a fixed degree polynomial).

Formulas of this type are known as *quadrature rules*.

Quadrature

For special case, choosing uniformly suffices:

Lemma

If A *is an* $m \times m$ *matrix and* $\{v_i\}_{i=1}^{n} \subseteq \mathbb{C}^m$ *are vectors with*

$$\|v_i\|^2 = \frac{m}{n} \qquad \text{and} \qquad \sum_i v_i v_i^* = I$$

then

$$\frac{1}{n} \sum_i \det\left[A + v_i v_i^*\right] = \mathbb{E}_Q\left\{\det\left[A + Q v_1 v_1^* Q^T\right]\right\}$$

Quadrature

For special case, choosing uniformly suffices:

Lemma

If A is an $m \times m$ matrix and $\{v_i\}_{i=1}^n \subseteq \mathbb{C}^m$ are vectors with

$$\|v_i\|^2 = \frac{m}{n} \qquad \textit{and} \qquad \sum_i v_i v_i^* = I$$

then

$$\frac{1}{n} \sum_i \det\left[A + v_i v_i^*\right] = \mathbb{E}_Q\left\{\det\left[A + Q v_1 v_1^* Q^T\right]\right\}$$

For full Bourgain-Tzafriri result, need to be more clever.

Quadrature in general

Quadrature rules exist for more general sums as well.

The larger the domain of possible integrals, the more nodes required:

Quadrature in general

Quadrature rules exist for more general sums as well.

The larger the domain of possible integrals, the more nodes required:

Theorem

For all $m \times m$ matrices A and B,

$$\mathbb{E}_P\left\{\det\left[A + \hat{P}B\hat{P}^T\right]\right\} = \mathbb{E}_Q\left\{\det\left[A + QBQ^T\right]\right\}$$

where

- *Q is an orthogonal matrix, distributed uniformly (via Haar measure)*
- *$\hat{P}$ is a signed permutation matrix, distributed uniformly ($2^n n!$ total)*

More connections

Recall the recipe for understanding random matrix distributions:

Random matrix
$\longrightarrow$ *free probability*
$\longrightarrow$ *free convolutions*
$\longrightarrow$ *finite free convolutions*
$\longrightarrow$ *polynomial*

More connections

Recall the recipe for understanding random matrix distributions:

Random matrix
→ *free probability*
→ *free convolutions*
→ *finite free convolutions*
→ *polynomial*

The free probability distribution is the *free Poisson distribution*.

The polynomials one studies to learn about Marchenko–Pastur distributions is precisely the collection of Laguerre polynomials we found.

Ramanujan Graphs

Application: existence of Ramanujan graphs of any size and degree.

Ramanujan Graphs

Application: existence of Ramanujan graphs of any size and degree.

Build from random matchings:

$$(x-1)^{m/2}(x+1)^{m/2}$$

Free probability distribution is *Kesten–McKay law*.

Ramanujan Graphs

Application: existence of Ramanujan graphs of any size and degree.

Build from random matchings:

$$(x-1)^{m/2}(x+1)^{m/2}$$

Free probability distribution is *Kesten–McKay law*.

Details are *far* more complicated:

1. generalization of characteristic polynomials to *determinant-like* polynomials.
2. special quadrature formula for Laplacian matrices
3. new convolution for asymmetric matrices

Thanks

Thank you to the organizers for providing me the opportunity to speak to you today.

And thank you for your attention!