Eigenvalue bounds on sums of random matrices

Adam W. Marcus

Princeton University
Institute for Advanced Study
adam.marcus@princeton.edu

Research supported by the National Science Foundation.

Outline

Conventions

In this talk, I will (hopefully) stick to the following conventions:

1. u, v, w: vector
2. A, B, C: matrix
3. $\hat{u}, \hat{v}, \hat{w}$: random vector
4. $\hat{A}, \hat{B}, \hat{C}$: random matrix

Conventions

In this talk, I will (hopefully) stick to the following conventions:

1. u, v, w: vector
2. A, B, C: matrix
3. $\hat{u}, \hat{v}, \hat{w}$: random vector
4. $\hat{A}, \hat{B}, \hat{C}$: random matrix

Warning: I will say things that involve probability.

This is *NOT* a talk about probability.

Conventions

In this talk, I will (hopefully) stick to the following conventions:

1. u, v, w: vector
2. A, B, C: matrix
3. $\hat{u}, \hat{v}, \hat{w}$: random vector
4. $\hat{A}, \hat{B}, \hat{C}$: random matrix

Warning: I will say things that involve probability.

This is *NOT* a talk about probability.

This is a talk about the combinatorics/geometry of vector spaces.

"Random matrix" will always mean there are finitely many possibilities, each with some nonzero probability.

Problem

For $\hat{A}_1, \dots \hat{A}_n$ independent, random, self adjoint matrices, let

$$\theta(\hat{A}_1, \dots, \hat{A}_n) = \min_{A_i \in supp(\hat{A}_i)} \lambda_{max} \left(\sum_i \hat{A}_i \right)$$

where $\lambda_{max}(X)$ is the largest eigenvalue of matrix X:

$$\lambda_{max}(X) = \max_v \frac{v^* X v}{\|v\|^2} = \max_{v : \|v\| = 1} v^* X v.$$

Problem

For $\hat{A}_1, \dots \hat{A}_n$ independent, random, self adjoint matrices, let

$$\theta(\hat{A}_1, \dots, \hat{A}_n) = \min_{A_i \in supp(\hat{A}_i)} \lambda_{max} \left(\sum_i \hat{A}_i \right)$$

where $\lambda_{max}(X)$ is the largest eigenvalue of matrix X:

$$\lambda_{max}(X) = \max_v \frac{v^* X v}{\|v\|^2} = \max_{v:\|v\|=1} v^* X v.$$

Such a quantity appears in numerous contexts (as we will see).

Problem

For $\hat{A}_1, \dots \hat{A}_n$ independent, random, self adjoint matrices, let

$$\theta(\hat{A}_1, \dots, \hat{A}_n) = \min_{A_i \in supp(\hat{A}_i)} \lambda_{max} \left(\sum_i \hat{A}_i \right)$$

where $\lambda_{max}(X)$ is the largest eigenvalue of matrix X:

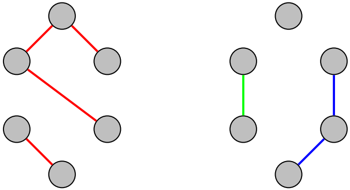
$$\lambda_{max}(X) = \max_v \frac{v^* X v}{\|v\|^2} = \max_{v : \|v\| = 1} v^* X v.$$

Such a quantity appears in numerous contexts (as we will see).

Much of my recent work has concerned finding upper bounds for θ.

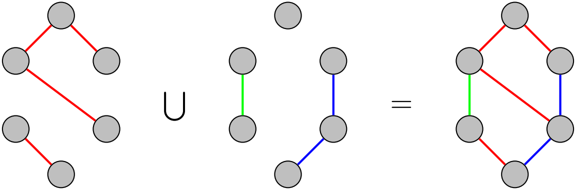
Example: Graphs

Let G_1 and G_2 be graphs with adjacency matrices A_1 and A_2.



Example: Graphs

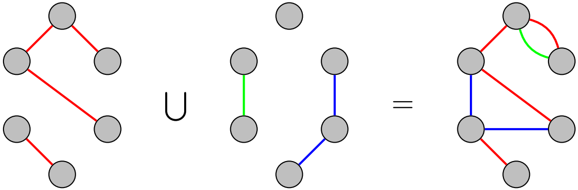
Let G_1 and G_2 be graphs with adjacency matrices A_1 and A_2.



$$A_{new} = A_1 + A_2$$

Example: Graphs

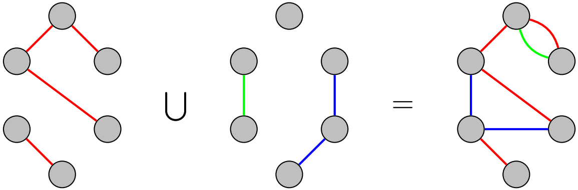
Let G_1 and G_2 be graphs with adjacency matrices A_1 and A_2.



$$A_{new} = A_1 + P^T A_2 P$$

Example: Graphs

Let G_1 and G_2 be graphs with adjacency matrices A_1 and A_2.



$$A_{new} = A_1 + P^T A_2 P$$

Can treat $\hat{A}_2 = P^T A_2 P$ as a random matrix with support size $|V|!$.

If A_1, A_2 are regular bipartite graphs, then

$$\theta(\Pi_{\perp\vec{1}}(A_1), \Pi_{\perp\vec{1}}(\hat{A}_2))$$

gives the best spectral gap.

Example: Spectral discrepancy

For positive semidefinite matrices $A_1, \ldots, A_n$ with $\sum_i A_i = I$, let

$$\hat{A}_i \in \left\{ \begin{array}{c|c} A_i & 0 \\ \hline 0 & 0 \end{array}, \begin{array}{c|c} 0 & 0 \\ \hline 0 & A_i \end{array} \right\}$$

have independent, uniform distributions.

Example: Spectral discrepancy

For positive semidefinite matrices $A_1, \ldots, A_n$ with $\sum_i A_i = I$, let

$$\hat{A}_i \in \left\{ \begin{array}{c|c} A_i & 0 \\ \hline 0 & 0 \end{array}, \begin{array}{c|c} 0 & 0 \\ \hline 0 & A_i \end{array} \right\}$$

have independent, uniform distributions.
Then $\theta(\hat{A}_1, \ldots \hat{A}_n)$ gives the "fairest partition":

$$\min_{S \subset [n]} \left\{ \left\| \sum_{i \in S} A_i \right\|, \left\| \sum_{i \notin S} A_i \right\| \right\}.$$

Example: Spectral discrepancy

For positive semidefinite matrices $A_1, \dots, A_n$ with $\sum_i A_i = I$, let

$$\hat{A}_i \in \left\{ \begin{array}{c|c} A_i & 0 \\ \hline 0 & 0 \end{array}, \begin{array}{c|c} 0 & 0 \\ \hline 0 & A_i \end{array} \right\}$$

have independent, uniform distributions.
Then $\theta(\hat{A}_1, \dots \hat{A}_n)$ gives the "fairest partition":

$$\min_{S \subset [n]} \left\{ \left\| \sum_{i \in S} A_i \right\|, \left\| \sum_{i \notin S} A_i \right\| \right\}.$$

For $\sum A_i = A \neq I$, one can set $B_i = A^{-1/2} A_i$ (so $\sum B_i = I$).

Probabilistic method

One way to try to bound θ is using the probabilistic method.

If we can show that

$$\mathbb{P}\left[\lambda_{max}\left(\sum_i \hat{A}_i\right) < t\right] > 0$$

then certainly

$$\theta(\hat{A}_1, \dots, \hat{A}_n) < t.$$

Probabilistic method

One way to try to bound θ is using the probabilistic method.

If we can show that

$$\mathbb{P}\left[\lambda_{max}\left(\sum_i \hat{A}_i\right) < t\right] > 0$$

then certainly

$$\theta(\hat{A}_1, \ldots, \hat{A}_n) < t.$$

There are *numerous* techniques for bounding such quantities.

Example: Matrix Chernoff

Theorem (Matrix Chernoff)

Let Let $\hat{A}_1, \dots, \hat{A}_n \in \mathbb{R}^{m \times m}$ *be positive semidefinite with*

$$\|\hat{A}_k\| \leq R \text{ a.s.} \qquad \text{and} \qquad \lambda_{max}\left(\sum_k \mathbb{E}\left\{\hat{A}_k\right\}\right) = \mu.$$

Then

$$\mathbb{P}\left[\lambda_{max}\left(\sum_k \hat{A}_k\right) \geq (1+\delta)\mu\right] \leq m\left[\frac{e^\delta}{(1+\delta)^{1+\delta}}\right]^{\mu/R}.$$

Matrix Bernstein

Theorem (Matrix Bernstein)

Let $\hat{A}_1, \ldots, \hat{A}_n \in \mathbb{R}^{m \times m}$ *be positive semidefinite with*

$$\lambda_{max}(\hat{A}_k) \leq R \text{ a.s.}$$

Then for $\sigma^2 = \| \sum_k \mathbb{E}\left\{ \hat{A}_k^2 \right\} \|$.

$$\mathbb{P}\left[\lambda_{max}\left(\sum_k \hat{A}_k \right) \geq t \right] \leq m e^{-3t^2/8\sigma^2}$$

for all $t \geq \sigma^2/R$.

Matrix Hoeffding

Theorem (Matrix Hoeffding)

Let $\hat{A}_1, \dots, \hat{A}_n \in \mathbb{R}^{m \times m}$ *be self adjoint with*

$$\mathbb{E}\left\{\hat{A}_k\right\} = 0 \qquad \textit{and} \qquad \mathbb{E}\left\{\hat{A}_k^2\right\} \preceq B_k^2 \ a.s.$$

Then for $\sigma^2 = \| \sum_k B_k^2 \|$.

$$\mathbb{P}\left[\lambda_{max}\left(\sum_k \hat{A}_k\right) \geq t\right] \leq m e^{-t^2/8\sigma^2}$$

for all $t > 0$.

Master tail Bound

Theorem (Tropp ('10))

Let $\hat{A}_1, \ldots, \hat{A}_n \in \mathbb{R}^{m \times m}$ be self adjoint and let

$$M_k(\theta) = \mathbb{E}\left\{ e^{\theta \hat{A}_k} \right\}$$

be their moment generating functions. Then

$$\mathbb{P}\left[\lambda_{max}\left(\sum_k \hat{A}_k \right) \geq t \right] \leq e^{-\theta t} \mathrm{Tr}\left[e^{\sum_k \log M_k(\theta)} \right]$$

for all $t \in \mathbb{R}$ and all $\theta > 0$.

Implies all previous bounds.

Known tools

Theorem (Matrix Chernoff/Bernstein/Hoeffding/etc)

If $\hat{A}_1, \dots, \hat{A}_n \in \mathbb{R}^{m \times m}$ are independent random self adjoint matrices then

$$\mathbb{P}\left[\lambda_{max}\left(\sum \hat{A}_i\right) > t\right] \leq m \cdot e^{-f(t, \hat{A}_1, \dots, \hat{A}_n)}$$

Similar inequalities by Rudelson ('99), Ahlswede–Winter ('02), Tropp ('10).

Known tools

Theorem (Matrix Chernoff/Bernstein/Hoeffding/etc)

If $\hat{A}_1, \ldots, \hat{A}_n \in \mathbb{R}^{m \times m}$ are independent random self adjoint matrices then

$$\mathbb{P}\left[\lambda_{max}\left(\sum \hat{A}_i\right) > t\right] \leq m \cdot e^{-f(t, \hat{A}_1, \ldots, \hat{A}_n)}$$

Similar inequalities by Rudelson ('99), Ahlswede–Winter ('02), Tropp ('10).

All such inequalities have two things in common:

1. They are all concentration bounds
2. The bounds depend on the dimension

The bad seed

Define $\hat{A}_1, \dots, \hat{A}_n \in \mathbb{R}^{n \times n}$ to be one of the n elementary diagonal matrices (with uniform probability).

The bad seed

Define $\hat{A}_1, \ldots, \hat{A}_n \in \mathbb{R}^{n \times n}$ to be one of the n elementary diagonal matrices (with uniform probability).

Then this is a balls and bins problem:

$$\theta(\hat{A}_1, \ldots, \hat{A}_n) = 1$$

but

$$\mathbb{P}\left[\lambda_{max}\left(\hat{A}_i\right) \geq \Omega\left(\frac{\log n}{\log\log n}\right)\right] \geq 1 - 1/n^{1/3}.$$

The bad seed

Define $\hat{A}_1, \ldots, \hat{A}_n \in \mathbb{R}^{n \times n}$ to be one of the n elementary diagonal matrices (with uniform probability).

Then this is a balls and bins problem:

$$\theta(\hat{A}_1, \ldots, \hat{A}_n) = 1$$

but

$$\mathbb{P}\left[\lambda_{max}\left(\hat{A}_i\right) \geq \Omega\left(\frac{\log n}{\log \log n}\right)\right] \geq 1 - 1/n^{1/3}.$$

Master tail bound gives:

$$\theta(\hat{A}_1, \ldots, \hat{A}_n) \leq O\left(\frac{\log n}{\log \log n}\right).$$

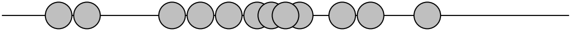
Want to do better

Similar examples show that any sufficiently generic bound that asserts $\lambda_{max} \geq t$ with "high probability" will need to depend on the dimension.

Want to do better

Similar examples show that any sufficiently generic bound that asserts $\lambda_{max} \geq t$ with "high probability" will need to depend on the dimension.

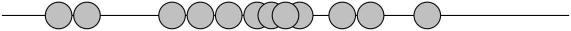
Fortunately, for our purposes, "high probability" is suboptimal.



Want to do better

Similar examples show that any sufficiently generic bound that asserts $\lambda_{max} \geq t$ with "high probability" will need to depend on the dimension.

Fortunately, for our purposes, "high probability" is suboptimal.



Need to find a way to capture "low probability" events.

"Low probability" means exponentially small (but still positive).

Outline

Key Idea

The key idea is to switch from random matrices to random polynomials. For any self adjoint matrix A,

$$\lambda_{\mathsf{max}}(A) = \mathrm{maxroot}\left\{\det\left[xI - A\right]\right\}.$$

Key Idea

The key idea is to switch from random matrices to random polynomials. For any self adjoint matrix A,

$$\lambda_{\max}(A) = \operatorname{maxroot}\left\{\det\left[xI - A\right]\right\}.$$

Hence for random self adjoint matrix $\hat{A}$,

$$\mathbb{P}\left[\lambda_{\max}(\hat{A}) < t\right] = \mathbb{P}\left[\operatorname{maxroot}\left\{\det\left[xI - \hat{A}\right] < t\right\}\right]$$

This suggests studying random characteristic polynomials.

Key Idea

The key idea is to switch from random matrices to random polynomials.
For any self adjoint matrix A,

$$\lambda_{\mathsf{max}}(A) = \mathrm{maxroot}\,\{\det[xI - A]\}\,.$$

Hence for random self adjoint matrix $\hat{A}$,

$$\mathbb{P}\left[\lambda_{\mathsf{max}}(\hat{A}) < t\right] = \mathbb{P}\left[\mathrm{maxroot}\left\{\det\left[xI - \hat{A}\right] < t\right\}\right]$$

This suggests studying random characteristic polynomials.

Need to have a way to compare the roots of a collection of polynomials with the roots of the average (which in general is not possible).

Main Lemma

Lemma (Separation Lemma)

Let $p_1, \dots, p_k$ be polynomials and $[s, t]$ an interval such that

- *Each $p_i(s)$ has the same sign (or is 0)*
- *Each $p_i(t)$ has the same sign (or is 0)*
- *each p_i has exactly one real root in $[s, t]$.*

Then $\sum_i p_i$ has exactly one real root in $[s, t]$ and it lies between the roots of some p_a and p_b.

Main Lemma

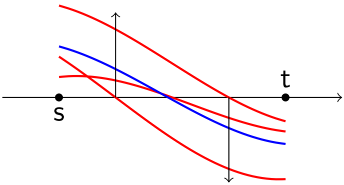
Lemma (Separation Lemma)

Let $p_1, \ldots, p_k$ be polynomials and $[s, t]$ an interval such that

- *Each $p_i(s)$ has the same sign (or is 0)*
- *Each $p_i(t)$ has the same sign (or is 0)*
- *each p_i has exactly one real root in $[s, t]$.*

Then $\sum_i p_i$ has exactly one real root in $[s, t]$ and it lies between the roots of some p_a and p_b.

Proof by picture:



Finding separation

Polynomial theory gives us a nice characterization of interlacing:

Lemma (Chudnovsky–Seymour, among others)

Let $\{p_i\}$ be degree d monic polynomials. The following are equivalent:

- *Every polynomial in the convex hull of $\{p_i\}$ has d real roots.*
- *The polynomials have all d of their roots separated.*

Finding separation

Polynomial theory gives us a nice characterization of interlacing:

Lemma (Chudnovsky–Seymour, among others)

Let $\{p_i\}$ be degree d monic polynomials. The following are equivalent:

- *Every polynomial in the convex hull of $\{p_i\}$ has d real roots.*
- *The polynomials have all d of their roots separated.*

We will say that p forms an *interlacing star* with $\{q_i\}$ if

1. The $\{q_i\}$ are degree d monic polynomials.
2. All convex combinations of the q_i are real rooted.
3. p is a convex combination of the $\{q_i\}$

Finding separation

Polynomial theory gives us a nice characterization of interlacing:

Lemma (Chudnovsky–Seymour, among others)

Let $\{p_i\}$ be degree d monic polynomials. The following are equivalent:

- *Every polynomial in the convex hull of $\{p_i\}$ has d real roots.*
- *The polynomials have all d of their roots separated.*

We will say that p forms an *interlacing star* with $\{q_i\}$ if

1. The $\{q_i\}$ are degree d monic polynomials.
2. All convex combinations of the q_i are real rooted.
3. p is a convex combination of the $\{q_i\}$

Corollary

If p forms an interlacing star with $\{q_i\}$, then there exist i, j such that

$$\mathrm{k^{th}root}\left(q_i\right) \leq \mathrm{k^{th}root}\left(p\right) \leq \mathrm{k^{th}root}\left(q_j\right).$$

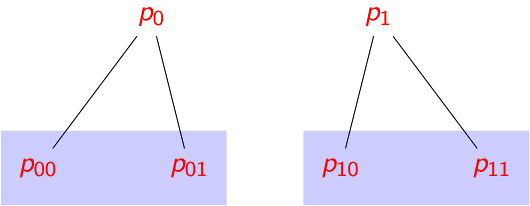
Interlacing families

To make this idea more versatile, we can iterate.

p_{00} p_{01} p_{10} p_{11}

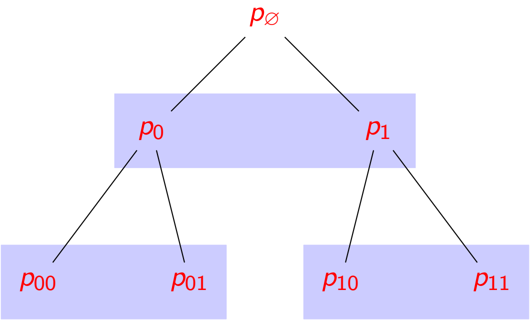
Interlacing families

To make this idea more versatile, we can iterate.



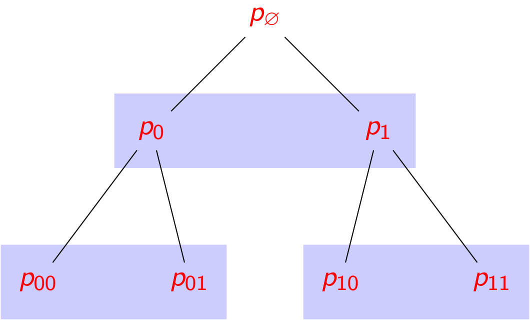
Interlacing families

To make this idea more versatile, we can iterate.



Interlacing families

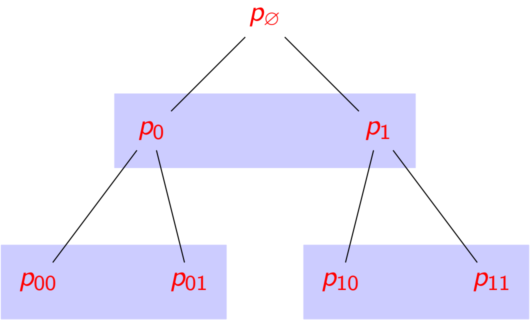
To make this idea more versatile, we can iterate.



We will call a rooted, connected tree where each node forms an interlacing star with its children an *interlacing family*.

Interlacing families

To make this idea more versatile, we can iterate.



We will call a rooted, connected tree where each node forms an interlacing star with its children an *interlacing family*.

The punchline

Corollary

Every interlacing family contains leaf nodes p_{leaf_1} *and* p_{leaf_2} *such that*

$$\mathrm{k^{th}root}\left(p_{leaf_1}\right) \leq \mathrm{k^{th}root}\left(p_{root}\right) \leq \mathrm{k^{th}root}\left(p_{leaf_2}\right).$$

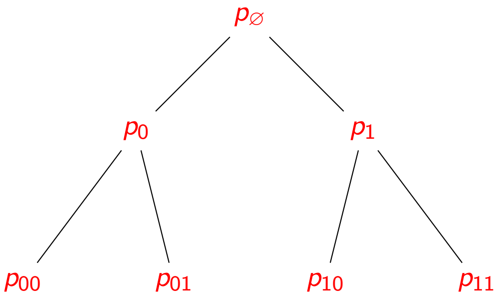
The punchline

Corollary

Every interlacing family contains leaf nodes p_{leaf_1} *and* p_{leaf_2} *such that*

$$\mathrm{k}^{\mathrm{th}}\mathrm{root}\left(p_{leaf_1}\right) \leq \mathrm{k}^{\mathrm{th}}\mathrm{root}\left(p_{root}\right) \leq \mathrm{k}^{\mathrm{th}}\mathrm{root}\left(p_{leaf_2}\right).$$

To find p_{leaf_i}:



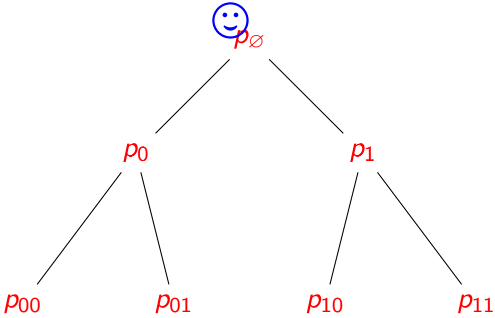
The punchline

Corollary

Every interlacing family contains leaf nodes p_{leaf_1} *and* p_{leaf_2} *such that*

$$\mathrm{k^{th}root}\left(p_{leaf_1}\right) \leq \mathrm{k^{th}root}\left(p_{root}\right) \leq \mathrm{k^{th}root}\left(p_{leaf_2}\right).$$

To find p_{leaf_i}:



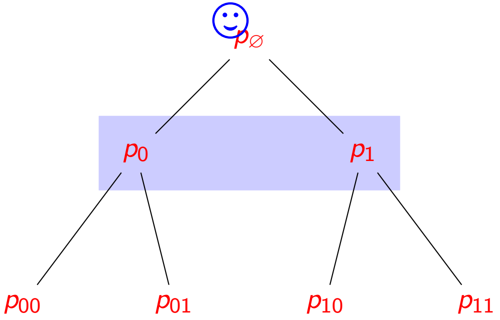
The punchline

Corollary

Every interlacing family contains leaf nodes p_{leaf_1} *and* p_{leaf_2} *such that*

$$\mathrm{k}^{\mathrm{th}}\mathrm{root}\left(p_{leaf_1}\right) \leq \mathrm{k}^{\mathrm{th}}\mathrm{root}\left(p_{root}\right) \leq \mathrm{k}^{\mathrm{th}}\mathrm{root}\left(p_{leaf_2}\right).$$

To find p_{leaf_i}:



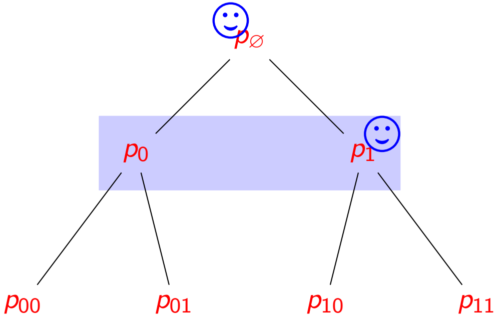
The punchline

Corollary

Every interlacing family contains leaf nodes p_{leaf_1} *and* p_{leaf_2} *such that*

$$\mathrm{k^{th}root}\left(p_{leaf_1}\right) \leq \mathrm{k^{th}root}\left(p_{root}\right) \leq \mathrm{k^{th}root}\left(p_{leaf_2}\right).$$

To find p_{leaf_i}:



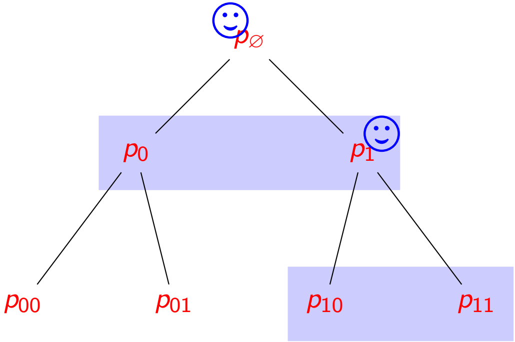
The punchline

Corollary

Every interlacing family contains leaf nodes p_{leaf_1} *and* p_{leaf_2} *such that*

$$\mathrm{k^{th}root}\left(p_{leaf_1}\right) \leq \mathrm{k^{th}root}\left(p_{root}\right) \leq \mathrm{k^{th}root}\left(p_{leaf_2}\right).$$

To find p_{leaf_i}:



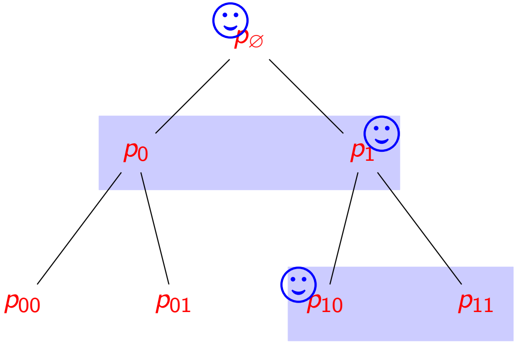
The punchline

Corollary

Every interlacing family contains leaf nodes p_{leaf_1} *and* p_{leaf_2} *such that*

$$\mathrm{k}^{\mathrm{th}}\mathrm{root}\left(p_{leaf_1}\right) \leq \mathrm{k}^{\mathrm{th}}\mathrm{root}\left(p_{root}\right) \leq \mathrm{k}^{\mathrm{th}}\mathrm{root}\left(p_{leaf_2}\right).$$

To find p_{leaf_i}:



Rank 1

In the rank 1 case, a bound on any root can then be obtained:

Theorem (MMS, ('13))

Let $\hat{A}_1, \hat{A}_2, \ldots \hat{A}_n \in \mathbb{R}^{m \times m}$ be random, independent rank 1 positive semidefinite matrices. Then the polynomials

$$\left\{ \det \left[xI - \sum_i A_i \right] \right\}_{A_i \in supp(\hat{A}_i)}$$

form an interlacing family. In particular

$$p_\emptyset(x) = \mathbb{E} \left\{ \det \left[xI - \sum_i \hat{A}_i \right] \right\}$$

has only real roots, and $\theta(\hat{A}_1, \ldots, \hat{A}_n) \leq \mathrm{maxroot}\, \{ p_\emptyset \}$.

Rank-1-ification

For higher rank matrices, a bound on θ can be obtained by "rank-1-ifying" them.

Theorem (Cohen ('16))

Let $\hat{A}_1, \hat{A}_2, \dots \hat{A}_n \in \mathbb{R}^{m \times m}$ be random, independent (any rank) positive semidefinite matrices, and let $\hat{B}_1, \hat{B}_2, \dots \hat{B}_n \in \mathbb{R}^{m \times m}$ be random rank 1 positive semidefinite matrices such that $\mathbb{E}\left\{ \hat{A}_i \right\} = \mathbb{E}\left\{ \hat{B}_i \right\}$ for all i. Then

$$\theta(\hat{A}_1, \dots, \hat{A}_n) \leq \operatorname{maxroot}\left\{ \mathbb{E}\left\{ \det\left[xI - \sum_i \hat{B}_i \right] \right\} \right\}.$$

Doesn't work for other roots.

Outline

Poisson Paradigm

"When X is the sum of many rare indicator "mostly independent" random variables and $\lambda = \mathbb{E}\{X\}$, we would like to say that X is close to a Poisson distribution with mean λ. We call this rough statement the Poisson Paradigm." (Alon, Spencer)

Poisson Paradigm

"When X is the sum of many rare indicator "mostly independent" random variables and $\lambda = \mathbb{E}\{X\}$, we would like to say that X is close to a Poisson distribution with mean λ. We call this rough statement the Poisson Paradigm." (Alon, Spencer)

Can we do something similar for matrices?

Poisson Paradigm

"When X is the sum of many rare indicator "mostly independent" random variables and $\lambda = \mathbb{E}\{X\}$, we would like to say that X is close to a Poisson distribution with mean λ. We call this rough statement the Poisson Paradigm." (Alon, Spencer)

Can we do something similar for matrices?

- $X = \sum_i X_i$ is sum of many nonnegative random variables
- each X_i has small expectation
- the X_i are "mostly independent"
- X behaves like a Poisson distribution with mean $\mathbb{E}\{X\}$.

Note the change of "rare, indicator" to "nonnegative, small expectation".

Noncommutative probability

Translation to noncommutative probability:

Classical	Noncommutative
distribution	eigenvalue distribution
random variable	linear operator
expectation	normalized trace
nonnegative	positive semidefinite

Noncommutative probability

Translation to noncommutative probability:

Classical	Noncommutative
distribution	eigenvalue distribution
random variable	linear operator
expectation	normalized trace
nonnegative	positive semidefinite

- $X = \sum_i X_i$ is sum of many ~~random variables~~ PSD matrices
- each ~~event~~ matrix X_i has ~~low probability~~ small trace
- the X_i are "mostly independent"
- X behaves like a Poisson distribution with mean $\mathbb{E}\{X\}$.

Noncommutative probability

Translation to noncommutative probability:

Classical	Noncommutative
distribution	eigenvalue distribution
random variable	linear operator
expectation	normalized trace
nonnegative	positive semidefinite

- $X = \sum_i X_i$ is sum of many ~~random variables~~ PSD matrices
- each ~~event~~ matrix X_i has ~~low probability~~ small trace
- the X_i are "mostly independent"
- X behaves like a Poisson distribution with mean $\mathbb{E}\{X\}$.

What does dependence mean?

Free probability

For two matrices A, B, the eigenvalues of $f(A, B)$ depend on

1. the eigenvalues of A
2. the eigenvalues of B
3. the dot product of the corresponding eigenvectors

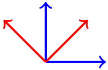
Free probability

For two matrices A, B, the eigenvalues of $f(A, B)$ depend on

1. the eigenvalues of A
2. the eigenvalues of B
3. the dot product of the corresponding eigenvectors

Quick and dirty explanation of "free independence":

- classical independence: f depends only on the marginal distributions
- free independence: f depends only on the eigenvalues (all dot products are equal)



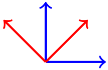
Free probability

For two matrices A, B, the eigenvalues of $f(A, B)$ depend on

1. the eigenvalues of A
2. the eigenvalues of B
3. the dot product of the corresponding eigenvectors

Quick and dirty explanation of "free independence":

- classical independence: f depends only on the marginal distributions
- free independence: f depends only on the eigenvalues (all dot products are equal)



Properties:

1. the identity is freely independent from everything (in any dimension)
2. Randomly rotated matrices are "asymptotically free"
3. In some sense, "as far away from commuting as possible".

Back to translation

Now we have

- $X = \sum_i X_i$ is sum of many ~~random variables~~ PSD matrices
- each ~~event~~ matrix X_i has ~~low probability~~ small trace
- the X_i are "mostly ~~independent~~ freely independent"
- X behaves like a Poisson distribution with mean $\mathbb{E}\{X\}$.

Back to translation

Now we have

- $X = \sum_i X_i$ is sum of many ~~random variables~~ PSD matrices
- each ~~event~~ matrix X_i has ~~low probability~~ small trace
- the X_i are "mostly ~~independent~~ freely independent"
- X behaves like a Poisson distribution with mean $\mathbb{E}\{X\}$.

Lastly we need to understand what Poisson would mean in this scenario.

Back to translation

Now we have

- $X = \sum_i X_i$ is sum of many ~~random variables~~ PSD matrices
- each ~~event~~ matrix X_i has ~~low probability~~ small trace
- the X_i are "mostly ~~independent~~ freely independent"
- X behaves like a Poisson distribution with mean $\mathbb{E}\{X\}$.

Lastly we need to understand what Poisson would mean in this scenario.

If original variables were *truly* independent, then Poisson distribution appears as "law of small numbers".

$$\binom{n}{k} p^k (1-p)^{n-k} \xrightarrow[np \to \lambda]{n \to \infty} e^{-\lambda} \frac{\lambda^k}{k!}$$

Free convolution

So need to know the eigenvalue distribution of $A + B$ when A and B are freely independent.

Free convolution

So need to know the eigenvalue distribution of $A+B$ when A and B are freely independent.

Theorem (Voiculescu ('91))

Let A_n and B_n be a sequence of $n \times n$ matrices with eigenvalue distributions converging to μ_A and μ_B (both compactly supported). Let Q be a random unitary matrix distributed via the Haar measure.

- *For all k, the sequence of k^{th} moments*

$$\mathbb{E}\left\{(A_n + QB_nQ^*)^k\right\}$$

  converges (weakly) to some m_k.
- *There exists a unique distribution which has $\mathbb{E}\{X^k\} = m_k$.*

New distribution is called the *free convolution* and written $\mu_A \boxplus \mu_B$.

Free Poisson distribution

Let $\mu_1, \ldots, \mu_n$ have Bernoulli(p) eigenvalue distributions.
Then the n-times free convolution

$$\mu_1 \boxplus \cdots \boxplus \mu_n \xrightarrow[np \to \lambda]{n \to \infty} \mu_{MP}$$

converges in distribution to the "Free Poisson distribution"

$$\mu_{MP}(t) = \frac{1}{2\pi t}\sqrt{4\lambda - (t - (1+\lambda))^2}dt.$$

Free Poisson distribution

Let $\mu_1, \dots, \mu_n$ have Bernoulli(p) eigenvalue distributions.
Then the n-times free convolution

$$\mu_1 \boxplus \cdots \boxplus \mu_n \xrightarrow[np \to \lambda]{n \to \infty} \mu_{MP}$$

converges in distribution to the "Free Poisson distribution"

$$\mu_{MP}(t) = \frac{1}{2\pi t}\sqrt{4\lambda - (t - (1 + \lambda))^2}dt.$$

In particular, μ_{MP} is supported on the interval

$$[(1 - \sqrt{\lambda})^2, (1 + \sqrt{\lambda})^2].$$

More on μ_{MP}

In fact μ_{MP} was discovered long before free probability existed in the field of Random Matrix Theory.

Theorem (Marcenko–Pastur)

Consider the random matrix

$$Y_{m,n} = \frac{1}{n} X X^T$$

where X is an $m \times n$ random matrix with i.i.d. $N(0,1)$ entries (often times called a Wishart matrix).
If $m, n \to \infty$ in such a way that $m/n \to \lambda \in \mathbb{R}$, then the empirical eigenvalue distribution of $Y_{m,n}$ distribution converges in distribution to μ_{MP}.

More on μ_{MP}

In fact μ_{MP} was discovered long before free probability existed in the field of Random Matrix Theory.

Theorem (Marcenko–Pastur)

Consider the random matrix

$$Y_{m,n} = \frac{1}{n} X X^T$$

where X is an $m \times n$ random matrix with i.i.d. $N(0,1)$ entries (often times called a Wishart matrix).
If $m, n \to \infty$ in such a way that $m/n \to \lambda \in \mathbb{R}$, then the empirical eigenvalue distribution of $Y_{m,n}$ distribution converges in distribution to μ_{MP}.

In particular, the support depends on the *ratio* m/n.

Free Poisson Paradigm

Recall our inspiration:

"When X is the sum of many rare indicator "mostly independent" random variables and $\lambda = \mathbb{E}\{X\}$, we would like to say that X is close to a Poisson distribution with mean λ. We call this rough statement the Poisson Paradigm." (Alon, Spencer)

Free Poisson Paradigm

Recall our inspiration:

"When X is the sum of many rare indicator "mostly independent" random variables and $\lambda = \mathbb{E}\{X\}$, we would like to say that X is close to a Poisson distribution with mean λ. We call this rough statement the Poisson Paradigm." (Alon, Spencer)

This becomes:

"When X is the sum of n "mostly freely independent" positive semidefinite $m \times m$ random matrices, we would like to say that X is close to a free Poisson distribution with parameter n/m. We call this rough statement the free Poisson Paradigm."

Mostly freely independent

Theorem

If $\mathbb{E}\left\{\hat{A}_k\right\} = \alpha_k I$ *for all* k*, then the free Poisson paradigm holds.*

Mostly freely independent

Theorem

If $\mathbb{E}\left\{\hat{A}_k\right\} = \alpha_k I$ *for all* k*, then the free Poisson paradigm holds.*

Recall the bad seed:

Define $\hat{A}_1, \ldots, \hat{A}_n \in \mathbb{R}^{n \times n}$ to be one of the n elementary diagonal matrices (with uniform probability).

Master tail bound gives:

$$\theta(\hat{A}_1, \ldots, \hat{A}_n) \leq O\left(\frac{\log n}{\log\log n}\right).$$

Mostly freely independent

Theorem

If $\mathbb{E}\left\{\hat{A}_k\right\} = \alpha_k I$ *for all* k, *then the free Poisson paradigm holds.*

Recall the bad seed:

Define $\hat{A}_1, \ldots, \hat{A}_n \in \mathbb{R}^{n \times n}$ to be one of the n elementary diagonal matrices (with uniform probability).

Master tail bound gives:

$$\theta(\hat{A}_1, \ldots, \hat{A}_n) \leq O\left(\frac{\log n}{\log \log n}\right).$$

Poisson paradigm gives:

$$\theta(\hat{A}_1, \ldots, \hat{A}_n) \leq 1 + \frac{2}{\sqrt{n}} + \frac{1}{n}$$

Outline

"Bounds"

So now we have (some sort of) bounds, but...

1. How can someone actually compute them?
2. Are they any good?

“Bounds”

So now we have (some sort of) bounds, but...

1. How can someone actually compute them?
2. Are they any good?

Such questions prompted the development of “finite free probability.”

Can be used for arbitrary self adjoint matrices (unlike method of interlacing polynomials).

Finite free additive convolution

Let
$$p(x) = \prod_{i=1}^{m}(x - a_i) \qquad \text{and} \qquad q(x) = \prod_{i=1}^{m}(x - b_i).$$

The *finite free additive convolution* of p and q is defined to be
$$[p \boxplus_m q](x) = \mathbb{E}_\sigma \left\{ \prod_{i=1}^{m}(x - a_i - b_{\sigma(i)}) \right\}.$$

Finite free additive convolution

Let

$$p(x) = \prod_{i=1}^{m} (x - a_i) \qquad \text{and} \qquad q(x) = \prod_{i=1}^{m} (x - b_i).$$

The *finite free additive convolution* of p and q is defined to be

$$[p \boxplus_m q](x) = \mathbb{E}_\sigma \left\{ \prod_{i=1}^{m} (x - a_i - b_{\sigma(i)}) \right\}.$$

Can be obtained without factoring:

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

Relation to random matrices

Let A, B be $m \times m$ self adjoint matrices and let

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

Then

$$[p \boxplus_m q](x) = \int \det [xI - A - QBQ^*] \, dQ.$$

Relation to random matrices

Let A, B be $m \times m$ self adjoint matrices and let

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

Then

$$[p \boxplus_m q](x) = \int \det[xI - A - QBQ^*]\, dQ.$$

Can take dQ to be

- Haar measure over orthogonal matrices
- Haar measure over unitary matrices
- uniformly distributed signed permutation matrices

Relation to random matrices

Let A, B be $m \times m$ self adjoint matrices and let

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

Then

$$[p \boxplus_m q](x) = \int \det[xI - A - QBQ^*] \, dQ.$$

Can take dQ to be

- Haar measure over orthogonal matrices $(\beta = 1)$
- Haar measure over unitary matrices $(\beta = 2)$
- uniformly distributed signed permutation matrices $(\beta = 0)$

Independent with respect to β.

Relation to free probability

Theorem (M ('16))

Let A and B be $m \times m$ self adjoint matrices and let $\mathcal{A}$ and $\mathcal{B}$ be freely independent random variables with the same eigenvalue distributions as A and B (respectively). Set

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

Then the root distribution of

$$[p^k \boxplus_{km} q^k]$$

converges (in distribution, as $k \to \infty$) to the eigenvalue distribution of

$$\mathcal{A} + \mathcal{B}.$$

Relation to free probability

Theorem (MSS ('15))

Let A and B be $m \times m$ self adjoint matrices and let $\mathcal{A}$ and $\mathcal{B}$ be freely independent random variables with the same eigenvalue distributions as A and B (respectively). Set

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

Then

$$\mathrm{maxroot}\left\{[p \boxplus_m q]\right\} \leq \lambda_{sup}(\mathcal{A} + \mathcal{B})$$

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

Relation to free probability

Theorem (MSS ('15))

Let A and B be $m \times m$ self adjoint matrices and let $\mathcal{A}$ and $\mathcal{B}$ be freely independent random variables with the same eigenvalue distributions as A and B (respectively). Set

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

Then

$$\mathrm{maxroot}\left\{[p \boxplus_m q]\right\} \leq \lambda_{sup}(\mathcal{A} + \mathcal{B})$$

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

Conjecture: $\mathrm{maxroot}\left\{[p^k \boxplus_{km} q^k]\right\}$ is increasing in k.

Relation to free probability

Theorem (MSS ('15))

Let A and B be $m \times m$ self adjoint matrices and let $\mathcal{A}$ and $\mathcal{B}$ be freely independent random variables with the same eigenvalue distributions as A and B (respectively). Set

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

Then

$$\mathrm{maxroot}\left\{[p \boxplus_m q]\right\} \leq \lambda_{sup}(\mathcal{A} + \mathcal{B})$$

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

Relation to free probability

Theorem (MSS ('15))

Let A and B be $m \times m$ self adjoint matrices and let $\mathcal{A}$ and $\mathcal{B}$ be freely independent random variables with the same eigenvalue distributions as A and B (respectively). Set

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

Then

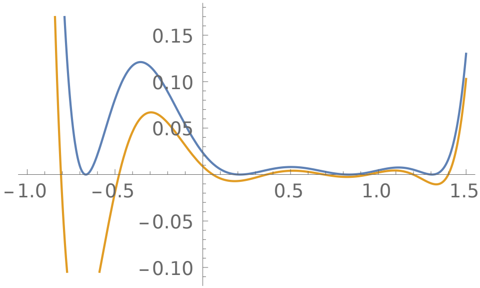
$$\operatorname{maxroot}\{[p \boxplus_m q]\} \leq \lambda_{sup}(\mathcal{A} + \mathcal{B})$$

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

Conjecture: $\operatorname{maxroot}\left\{[p^k \boxplus_{km} q^k]\right\}$ is increasing in k.

Sequence

The sequence $[p^k \boxplus_{km} q^k]$ can give higher correlations as well.



Blue: $[p \boxplus_4 q]^2$
Yellow: $[p^2 \boxplus_4 q^2]$

Conjectures:

- the roots of $[p^2 \boxplus_m q^2]$ majorize the roots of $[p \boxplus_m q]^2$
- $[p^2 \boxplus_m q^2] \leq [p \boxplus_m q]^2$ for all $x \in \mathbb{R}$

Algebra

Let $L = \sum_i c_i \partial^i$ be a linear differential operator. Then

$$L\{p \boxplus_m q\} = L\{p\} \boxplus_m q = p \boxplus_m L\{q\}.$$

Algebra

Let $L = \sum_i c_i \partial^i$ be a linear differential operator. Then

$$L\left\{p \boxplus_m q\right\} = L\{p\} \boxplus_m q = p \boxplus_m L\{q\}.$$

So if P and Q are linear differential operators such that

$$p(x) = P\{x^m\} \qquad \text{and} \qquad q(x) = Q\{x^m\}$$

then

$$\begin{aligned}[p \boxplus_m q] &= [P\{x^m\} \boxplus_m Q\{x^m\}] \\ &= P\{Q\{[x^m \boxplus_m x^m]\}\} \\ &= PQ\{x^m\}.\end{aligned}$$

Algebra

Let $L = \sum_i c_i \partial^i$ be a linear differential operator. Then

$$L\{p \boxplus_m q\} = L\{p\} \boxplus_m q = p \boxplus_m L\{q\}.$$

So if P and Q are linear differential operators such that

$$p(x) = P\{x^m\} \qquad \text{and} \qquad q(x) = Q\{x^m\}$$

then

$$\begin{aligned}[p \boxplus_m q] &= [P\{x^m\} \boxplus_m Q\{x^m\}] \\ &= P\{Q\{[x^m \boxplus_m x^m]\}\} \\ &= PQ\{x^m\}.\end{aligned}$$

Isomorphic to $\mathbb{R}[x]/\langle x^{m+1}\rangle$ under multiplication.

Example: Hermite polynomials

The *Hermite polynomials* are defined as

$$H_m(x) = \sum_{k=0}^{\lfloor m/2 \rfloor} \frac{m!}{k!(m-2k)!} \left(\frac{-1}{2} \right)^k x^{m-2k} = e^{-\partial^2/2} \{ x^m \}$$

Example: Hermite polynomials

The *Hermite polynomials* are defined as

$$H_m(x) = \sum_{k=0}^{\lfloor m/2 \rfloor} \frac{m!}{k!(m-2k)!} \left(\frac{-1}{2} \right)^k x^{m-2k} = e^{-\partial^2/2}\{x^m\}$$

Hence we have

$$\begin{aligned} [H_m \boxplus_m H_m] &= \left[e^{-\partial^2/2}\{x^m\} \boxplus_m e^{-\partial^2/2}\{x^m\} \right] \\ &= e^{-\partial^2}\{[x^m \boxplus_m x^m\} \\ &= e^{-\partial^2}\{x^m\} \\ &= H_m(\sqrt{2}x) \end{aligned}$$

First glimpse of free probability

Theorem (Central Limit Theorem)

Let $A_1, \ldots, A_n, \ldots$ *be* $m \times m$ *real, symmetric matrices such that*

1. $\phi[A] = 0$
2. $\phi\left[A^2\right] = 1$
3. $p_k(x) = \det[xI - A_k]$

First glimpse of free probability

Theorem (Central Limit Theorem)

Let $A_1, \dots, A_n, \dots$ *be* $m \times m$ *real, symmetric matrices such that*

1. $\phi[A] = 0$
2. $\phi\left[A^2\right] = 1$
3. $p_k(x) = \det[xI - A_k]$

Then the roots of

$$\lim_{n\to\infty} [p_1(\sqrt{n}x) \boxplus_m p_2(\sqrt{n}x) \boxplus_m \cdots \boxplus_m p_n(\sqrt{n}x)]$$

converge to the roots of

$$H_m(x\sqrt{m-1}).$$

First glimpse of free probability

Theorem (Central Limit Theorem)

Let $A_1, \ldots, A_n, \ldots$ *be* $m \times m$ *real, symmetric matrices such that*

1. $\phi[A] = 0$
2. $\phi\left[A^2\right] = 1$
3. $p_k(x) = \det[xI - A_k]$

Then the roots of

$$\lim_{n \to \infty} [p_1(\sqrt{n}x) \boxplus_m p_2(\sqrt{n}x) \boxplus_m \cdots \boxplus_m p_n(\sqrt{n}x)]$$

converge to the roots of

$$H_m(x\sqrt{m-1}).$$

Note: as $m \to \infty$, root distribution approaches the semicircle law.

Proof

For each k, we have $p_k(\sqrt{n}x)$ has the same roots as $T_k\{x^m\}$ where

$$T_k = 1 - \frac{1}{2n(m-1)}\partial^2 + O(n^{-3/2}).$$

Proof

For each k, we have $p_k(\sqrt{n}x)$ has the same roots as $T_k\{x^m\}$ where

$$T_k = 1 - \frac{1}{2n(m-1)}\partial^2 + O(n^{-3/2}).$$

Then

$$[p_1(\sqrt{n}x) \boxplus_m p_2(\sqrt{n}x) \boxplus_m \cdots \boxplus_m p_n(\sqrt{n}x)] = \left[\prod_{k=1}^{n} T_k\right]\{x^m\}$$

where

$$\prod_{k=1}^{n} T_k = \left(1 - \frac{1}{2n(m-1)}\partial^2 + O(n^{-3/2})\right)^n \to e^{-\partial^2/2(m-1)}$$

Proof

For each k, we have $p_k(\sqrt{n}x)$ has the same roots as $T_k\{x^m\}$ where

$$T_k = 1 - \frac{1}{2n(m-1)}\partial^2 + O(n^{-3/2}).$$

Then

$$[p_1(\sqrt{n}x) \boxplus_m p_2(\sqrt{n}x) \boxplus_m \cdots \boxplus_m p_n(\sqrt{n}x)] = \left[\prod_{k=1}^{n} T_k\right]\{x^m\}$$

where

$$\prod_{k=1}^{n} T_k = \left(1 - \frac{1}{2n(m-1)}\partial^2 + O(n^{-3/2})\right)^n \to e^{-\partial^2/2(m-1)}$$

So

$$\prod_{k=1}^{n} T_k\{x^m\} \to e^{-\partial^2/2(m-1)}\{x^m\} = H_m(x\sqrt{m-1}).$$

Second glimpse of free probability

A similar computation can be used in the free Poisson Paradigm.

Theorem (Poisson Limit Theorem)

Let $p(x) = x^{m-1}(x-1)$.
Then

$$[\underbrace{p \boxplus_m p \boxplus_m \cdots \boxplus_m p}_{\lambda m \text{ times}}] = m!(-m)^m L_m^{((\lambda-1)m)}(mx) \qquad (*)$$

where $L_n^{(\alpha)}(x)$ *is an associated Laguerre polynomial.*

Second glimpse of free probability

A similar computation can be used in the free Poisson Paradigm.

Theorem (Poisson Limit Theorem)

Let $p(x) = x^{m-1}(x-1)$.
Then

$$\underbrace{[p \boxplus_m p \boxplus_m \cdots \boxplus_m p]}_{\lambda m \text{ times}} = m!(-m)^m L_m^{((\lambda-1)m)}(mx) \qquad (*)$$

where $L_n^{(\alpha)}(x)$ *is an associated Laguerre polynomial.*

Note: as $m \to \infty$, root distribution approaches the Marcenko–Pastur law.

Second glimpse of free probability

A similar computation can be used in the free Poisson Paradigm.

Theorem (Poisson Limit Theorem)

Let $p(x) = x^{m-1}(x-1)$.
Then

$$\underbrace{[p \boxplus_m p \boxplus_m \cdots \boxplus_m p]}_{\lambda m \text{ times}} = m!(-m)^m L_m^{((\lambda-1)m)}(mx) \qquad (*)$$

where $L_n^{(\alpha)}(x)$ *is an associated Laguerre polynomial.*

Note: as $m \to \infty$, root distribution approaches the Marcenko–Pastur law.

[Ismail, Li ('92)]

$$\mathrm{maxroot}\,\{(*)\} \leq (1+\sqrt{\lambda})^2 - \frac{\left(\lambda^{1/4}+\lambda^{-1/4}\right)^2}{m} + O(m^{-2})$$

Max roots

Of course we cannot hope that all polynomials have extensive literature giving bounds on their roots.

Unfortunately, the $\mathrm{maxroot}\{\}$ operation is unstable with respect to our convolutions (and really any operation).

Max roots

Of course we cannot hope that all polynomials have extensive literature giving bounds on their roots.

Unfortunately, the $\mathrm{maxroot}\left\{\right\}$ operation is unstable with respect to our convolutions (and really any operation).

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

$$\mathrm{maxroot}\left\{p\right\} = \mathrm{maxroot}\left\{q\right\} = 1.$$

Max roots

Of course we cannot hope that all polynomials have extensive literature giving bounds on their roots.

Unfortunately, the $\mathrm{maxroot}\{\}$ operation is unstable with respect to our convolutions (and really any operation).

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

Max roots

Of course we cannot hope that all polynomials have extensive literature giving bounds on their roots.

Unfortunately, the $\mathrm{maxroot}\,\{\}$ operation is unstable with respect to our convolutions (and really any operation).

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 then gives an upper bound of 2.

Max roots

Solution: use smoother version of the $\mathrm{maxroot}\{\}$ function.

Max roots

Solution: use smoother version of the $\mathrm{maxroot}\{\}$ function.

Definition

For a real rooted polynomial p, we define

$$\alpha\max(p) = \mathrm{maxroot}\left\{p - \alpha p'\right\}.$$

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

Max roots

Solution: use smoother version of the $\mathrm{maxroot}\{\}$ function.

Definition

For a real rooted polynomial p, we define

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

So $\alpha = 0$ is the usual $\mathrm{maxroot}\{\}$ 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.

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

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

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

Proof uses implicit differentiation and Newton inequalities.

Lemma

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

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

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 $x \geq \mathrm{maxroot}\{p\}$.

Corollary

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

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

Main inequality

Theorem

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

$$\alpha\mathrm{max}\,(p \boxplus_m q) \leq \alpha\mathrm{max}\,(p) + \alpha\mathrm{max}\,(q) - m\alpha$$

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

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

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

	$\mathrm{maxroot}\left\{\cdot\right\}$	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

Main inequality

Theorem

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

$$\alpha\mathrm{max}\left(p \boxplus_m q\right) \leq \alpha\mathrm{max}\left(p\right) + \alpha\mathrm{max}\left(q\right) - m\alpha$$

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

	$\mathrm{maxroot}\left\{\cdot\right\}$	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

Multiple convolutions: keep as a function of α, then optimize at the end.

Main inequality

Theorem

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

$$\lim_{k\to\infty} \alpha\mathrm{max}\left(p^k \boxplus_{km} q^k\right) = \alpha\mathrm{max}\left(p^k\right) + \alpha\mathrm{max}\left(q^k\right) - km\alpha$$

for all p and q.

Main inequality

Theorem

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

$$\lim_{k \to \infty} \alpha\mathrm{max}\left(p^k \boxplus_{km} q^k\right) = \alpha\mathrm{max}\left(p^k\right) + \alpha\mathrm{max}\left(q^k\right) - km\alpha$$

for all p and q.

Converges to the R-transform identity for free convolution:

$$R_{A \boxplus B}(x) = R_A(x) + R_B(x).$$

Main inequality

Theorem

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

$$\lim_{k\to\infty} \alpha\mathrm{max}\left(p^k \boxplus_{km} q^k\right) = \alpha\mathrm{max}\left(p^k\right) + \alpha\mathrm{max}\left(q^k\right) - km\alpha$$

for all p and q.

Converges to the R-transform identity for free convolution:

$$R_{A\boxplus B}(x) = R_A(x) + R_B(x).$$

Implies Poisson paradigm is asymptotically sharp (for the given information).

Outline

Restricted Invertibility

Theorem (Bourgain, Tzafriri)

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

$$\sum_{i=1}^{n} v_i v_i^T = I$$

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

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

Many applications in computer science, functional analysis, convex geometry.

Spectral Graph Theory

Let G_1 be a d_1-regular graph and G_2 be a d_2-regular graph with adjacency matrices A_1 and A_2. Then the (random) matrices

$$B_1 = \Pi_{\perp \vec{1}}(A_1) \qquad \text{and} \qquad B_2 = \Pi_{\perp \vec{1}}(\hat{A}_2)$$

obey the Poisson paradigm. In particular, there exists a rotation P such that

$$\lambda_2(A_1 + P^T A_2 P) \leq \mathrm{maxroot}\left\{\left[\det\left[xI - B_1\right] \boxplus_n \det\left[xI - B_2\right]\right]\right\}.$$

Spectral Graph Theory

Let G_1 be a d_1-regular graph and G_2 be a d_2-regular graph with adjacency matrices A_1 and A_2. Then the (random) matrices

$$B_1 = \Pi_{\perp \vec{1}}(A_1) \qquad \text{and} \qquad B_2 = \Pi_{\perp \vec{1}}(\hat{A}_2)$$

obey the Poisson paradigm. In particular, there exists a rotation P such that

$$\lambda_2(A_1 + P^T A_2 P) \leq \mathrm{maxroot}\left\{[\det[xI - B_1] \boxplus_n \det[xI - B_2]]\right\}.$$

Leads to the construction of d-regular Ramanujan graphs as union of d randomly permuted perfect matchings.

Recap

We have a new way to capture low probability events.

Useful when "Gaussian random matrix" is the conjectured worst case scenario.

Recap

We have a new way to capture low probability events.

Useful when "Gaussian random matrix" is the conjectured worst case scenario.

Method of interlacing polynomials is used to show eigenvalues meet *some* bound with nonzero probability.

Using finite free probability, we can explicitly calculate these bounds.

All such bounds will be asymptotically tight (example showing tightness comes from free probability).

Much to learn

The "furthest from freely independent" situation we know:

Theorem (MSS ('13) + Cohen ('16))

If $\hat{A}_1, \hat{A}_2, \dots \hat{A}_n \in \mathbb{R}^{m \times m}$ are independent random positive semidefinite matrices with

$$\sum_i \mathbb{E}\left\{ \hat{A}_i \right\} = I_m \qquad (*) \qquad \text{and} \qquad \mathbb{E}\left\{ \mathrm{Tr}\left[\hat{A}_i \right] \right\} \leq \epsilon$$

then

$$\theta(\hat{A}_1, \dots, \hat{A}_n) \leq (1 + \sqrt{\epsilon})^2.$$

Much to learn

The "furthest from freely independent" situation we know:

Theorem (MSS ('13) + Cohen ('16))

If $\hat{A}_1, \hat{A}_2, \ldots \hat{A}_n \in \mathbb{R}^{m \times m}$ are independent random positive semidefinite matrices with

$$\sum_i \mathbb{E}\left\{ \hat{A}_i \right\} = I_m \qquad (*) \qquad \textit{and} \qquad \mathbb{E}\left\{ \mathrm{Tr}\left[\hat{A}_i \right] \right\} \leq \epsilon$$

then

$$\theta(\hat{A}_1, \ldots, \hat{A}_n) \leq (1 + \sqrt{\epsilon})^2.$$

Not captured (I believe) by additive convolution — suggests need for a multivariate extension.

Lyapunov theorem

Example application of a stronger theory:

Theorem (Akemann–Weaver ('14) + Cohen ('16))

Let $A_1, \dots, A_n \in \mathbb{C}^{m \times m}$ be positive semidefinite matrices with

$$\sum_i A_i \leq I \qquad \text{and} \qquad \mathrm{Tr}\left[A_i\right] \leq \epsilon$$

for all i. Then for all values of $t_1, \dots, t_n \in [0, 1]$, there exists a set of indices $S \subset [n]$ such that

$$\left\| \sum_{i \in S} A_i - \sum_{i=1}^{n} t_i A_i \right\| = O(\epsilon^{1/8}).$$

Lyapunov theorem

Example application of a stronger theory:

Theorem (Akemann–Weaver ('14) + Cohen ('16))

Let $A_1, \dots, A_n \in \mathbb{C}^{m \times m}$ *be positive semidefinite matrices with*

$$\sum_i A_i \leq I \qquad \textit{and} \qquad \mathrm{Tr}\left[A_i\right] \leq \epsilon$$

for all i. *Then for all values of* $t_1, \dots, t_n \in [0, 1]$, *there exists a set of indices* $S \subset [n]$ *such that*

$$\left\| \sum_{i \in S} A_i - \sum_{i=1}^{n} t_i A_i \right\| = O(\epsilon^{1/8}).$$

Applications to semidefinite programming?

Open problems

Can we find a way to bound

$$\min_{A_i \in supp(\hat{A}_i)} \left\| \sum_i A_i \right\|$$

for general self adjoint matrices?

Open problems

Can we find a way to bound

$$\min_{A_i \in supp(\hat{A}_i)} \left\| \sum_i A_i \right\|$$

for general self adjoint matrices?

Can we get bounds in terms of norms other than the expected trace?

Bounds in terms of Frobenius norm would be particularly interesting.

Open problems

Can we find a way to bound

$$\min_{A_i \in supp(\hat{A}_i)} \left\| \sum_i A_i \right\|$$

for general self adjoint matrices?

Can we get bounds in terms of norms other than the expected trace?

Bounds in terms of Frobenius norm would be particularly interesting.

Multivariate extensions?

Open applications

Other ideas from free probability can be "finitized".

1. Multiplicative convolution
2. Asymmetric additive convolution (singular values)
3. Additive/multiplicative brownian motion
4. Entropy
5. Fisher information
6. Combinatorial theory

Open applications

Other ideas from free probability can be "finitized".

1. Multiplicative convolution
2. Asymmetric additive convolution (singular values)
3. Additive/multiplicative brownian motion
4. Entropy
5. Fisher information
6. Combinatorial theory

Can these be used in similar ways?

(There are noticable similarities to ideas in the "discrete log gas" literature.)

Thanks

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

And thank you for your attention!