Interlacing Families: A New Technique for Controlling Eigenvalues

Adam W. Marcus

Princeton University
adam.marcus@princeton.edu

Acknowledgements:

Joint work with:

Dan Spielman
Yale University

Nikhil Srivastava
University of California, Berkeley

My involvement partially supported by:

Crisply

National Science Foundation
Mathematical Sciences Postdoctoral Research Fellowship

Some notation

First some conventions:

1. $\alpha, \beta, \ldots$ will be real numbers
2. $u, v, \ldots$ will be vectors in $\mathbb{R}^d$
3. $U, V, \ldots$ will be $d \times d$ symmetric, real matrices

1. $\widehat{u}, \widehat{v}, \ldots$ will be random vectors in $\mathbb{R}^d$
2. $\widehat{U}, \widehat{V}, \ldots$ will be random matrices

Some notation

First some conventions:

1. $\alpha, \beta, \ldots$ will be real numbers
2. $u, v, \ldots$ will be vectors in $\mathbb{R}^d$
3. $U, V, \ldots$ will be $d \times d$ symmetric, real matrices

1. $\widehat{u}, \widehat{v}, \ldots$ will be random vectors in $\mathbb{R}^d$
2. $\widehat{U}, \widehat{V}, \ldots$ will be random matrices

And please interrupt if you have any questions!

Outline

Motivation

I want to look at self-adjoint linear operators.

Motivation

I want to look at self-adjoint linear operators.

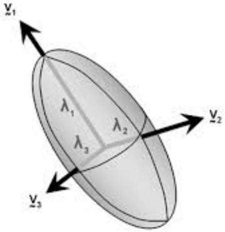
Algebraically, think: real, square, symmetric, matrices.

Motivation

I want to look at self-adjoint linear operators.

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

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

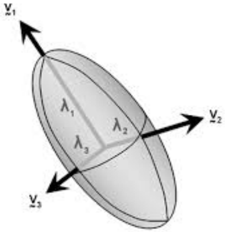


Motivation

I want to look at 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*

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

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

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

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 frame*.

Known tools

Well-known techniques exist for bounding the eigenvalues of random frames. For example,

Theorem (Matrix Chernoff)

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

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

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

Known tools

Well-known techniques exist for bounding the eigenvalues of random frames. For example,

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).

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).

The goal

I want to find a bound on the eigenvalues that is *independent of dimension*.

The goal

I want to find a bound on the eigenvalues that is *independent of dimension*.

Furthemore, I want to keep the "probabilistic" nature:

Theorem

If $\widehat{\theta}$ is a random variable with finite support, then

$$\mathbb{P}\left[\widehat{\theta} \geq \mathbb{E}\widehat{\theta}\right] > 0 \quad \textit{and} \quad \mathbb{P}\left[\widehat{\theta} \leq \mathbb{E}\widehat{\theta}\right] > 0$$

In other words, I want to study one object (here $\mathbb{E}\widehat{\theta}$) and then be able to assert the existence of something at least as good (in both directions).

In fairy-tale land

So given a random frame $\widehat{V} = \sum_i \widehat{v}_i \widehat{v}_i^T$, I would like to say:

$$\mathbb{P}\left[\lambda_{max}(\widehat{V}) \geq \lambda_{max}(\mathbb{E}\widehat{V})\right] > 0$$

and

$$\mathbb{P}\left[\lambda_{max}(\widehat{V}) \leq \lambda_{max}(\mathbb{E}\widehat{V})\right] > 0$$

In fairy-tale land

So given a random frame $\widehat{V} = \sum_i \widehat{v}_i \widehat{v}_i^T$, I would like to say:

$$\mathbb{P}\left[\lambda_{max}(\widehat{V}) \geq \lambda_{max}(\mathbb{E}\widehat{V})\right] > 0$$

and

$$\mathbb{P}\left[\lambda_{max}(\widehat{V}) \leq \lambda_{max}(\mathbb{E}\widehat{V})\right] > 0$$

But this isn't true (pick just $\widehat{v}$ as $(0, 1)$ or $(1, 0)$ uniformly).

In fairy-tale land

So given a random frame $\widehat{V} = \sum_i \widehat{v}_i \widehat{v}_i^T$, I would like to say:

$$\mathbb{P}\left[\lambda_{max}(\widehat{V}) \geq \lambda_{max}(\mathbb{E}\widehat{V})\right] > 0$$

and

$$\mathbb{P}\left[\lambda_{max}(\widehat{V}) \leq \lambda_{max}(\mathbb{E}\widehat{V})\right] > 0$$

But this isn't true (pick just $\widehat{v}$ as $(0, 1)$ or $(1, 0)$ uniformly).

So instead, we make an observation:

Observation

If A is a $d \times d$ real, symmetric matrix with eigenvalues $\lambda_1, \ldots, \lambda_d$, then

$$\chi_A(x) := \det\left[xI - A\right] = \prod_{i=1}^{d}(x - \lambda_i)$$

Called the *characteristic polynomial* of A.

REAL fairy-tale land

So *now*, maybe we can do what we want in terms of polynomials!

REAL fairy-tale land

So *now*, maybe we can do what we want in terms of polynomials!

That is, given a random frame $\widehat{V} = \sum_i \widehat{v}_i \widehat{v}_i^T$, maybe we can say:

$$\mathbb{P}\left[\text{maxroot}\left(\chi_{\widehat{V}}\right) \geq \text{maxroot}\left(\mathbb{E}\left[\chi_{\widehat{V}}\right]\right)\right] > 0$$

and

$$\mathbb{P}\left[\text{maxroot}\left(\chi_{\widehat{V}}\right) \leq \text{maxroot}\left(\mathbb{E}\left[\chi_{\widehat{V}}\right]\right)\right] > 0$$

REAL fairy-tale land

So *now*, maybe we can do what we want in terms of polynomials!

That is, given a random frame $\widehat{V} = \sum_i \widehat{v}_i \widehat{v}_i^T$, maybe we can say:

$$\mathbb{P}\left[\text{maxroot}\left(\chi_{\widehat{V}}\right) \geq \text{maxroot}\left(\mathbb{E}\left[\chi_{\widehat{V}}\right]\right)\right] > 0$$

and

$$\mathbb{P}\left[\text{maxroot}\left(\chi_{\widehat{V}}\right) \leq \text{maxroot}\left(\mathbb{E}\left[\chi_{\widehat{V}}\right]\right)\right] > 0$$

Certainly this is nonsense, but let's play along with a toy problem:

Let A be a matrix and $\widehat{w}$ a random vector (taking values u or v uniformly).

What can we say about the eigenvalues of $A + \widehat{w}\widehat{w}^T$?

Still playing along

We would (naively) start by looking at the expected polynomial

$$p(x) = \frac{1}{2}\chi_{A+uu^T}(x) + \frac{1}{2}\chi_{A+vv^T}(x)$$

Why is this naive?

Still playing along

We would (naively) start by looking at the expected polynomial

$$p(x) = \frac{1}{2}\chi_{A+uu^T}(x) + \frac{1}{2}\chi_{A+vv^T}(x)$$

Why is this naive?

Adding polynomials is a function of the *coefficients* and we are interested in the *roots*.
In general, it is easy to get the coefficients from the roots but hard to get the roots from the coefficients.

Still playing along

We would (naively) start by looking at the expected polynomial

$$p(x) = \frac{1}{2}\chi_{A+uu^T}(x) + \frac{1}{2}\chi_{A+vv^T}(x)$$

Why is this naive?

Adding polynomials is a function of the *coefficients* and we are interested in the *roots*.
In general, it is easy to get the coefficients from the roots but hard to get the roots from the coefficients.

Example: $p(x) = (x-2)^2 - 1$ (has double root at 1) and $q(x) = (x+2)^2 - 1$ (has double root at -1).

$$p(x) + q(x) = x^2 + 6$$

does not have any real roots (roots are $\pm\sqrt{-6}$).

Unless...

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.

Unless...

Lemma (Separation Lemma)

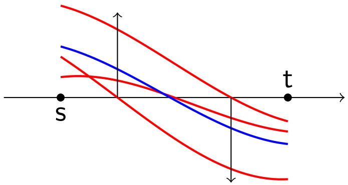
Let $p_1, \ldots, p_k$ be polynomials and $[s, t]$ an interval such that

- *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:



$\square$

A ray of hope

So if we have the right structure, using characteristic polynomials could actually work!

A ray of hope

So if we have the right structure, using characteristic polynomials could actually work!

Pros:

- All eigenvalues are tracked in a compact form
- Maybe take advantage of polynomial techniques that "don't make sense" to matrices

A ray of hope

So if we have the right structure, using characteristic polynomials could actually work!

Pros:

- All eigenvalues are tracked in a compact form
- Maybe take advantage of polynomial techniques that "don't make sense" to matrices

Cons:

- You lose rotation (how can we add without knowing rotation?)
- Have to worry about matrix operations that "don't make sense" to polynomials

A ray of hope

So if we have the right structure, using characteristic polynomials could actually work!

Pros:

- All eigenvalues are tracked in a compact form
- Maybe take advantage of polynomial techniques that "don't make sense" to matrices

Cons:

- You lose rotation (how can we add without knowing rotation?)
- Have to worry about matrix operations that "don't make sense" to polynomials

What do I mean by "polynomial techniques"?

Polynomial Techniques

Univariate polynomials inherit techniques from

- Convex Analysis
- Complex Analysis
- Combinatorics

Polynomial Techniques

Univariate polynomials inherit techniques from

- Convex Analysis
- Complex Analysis
- Combinatorics

Multivariate polynomials inherit techniques from

- (Real) Algebraic Geometry
- Matroid theory
- Control Theory

Polynomial Techniques

Univariate polynomials inherit techniques from

- Convex Analysis
- Complex Analysis
- Combinatorics

Multivariate polynomials inherit techniques from

- (Real) Algebraic Geometry
- Matroid theory
- Control Theory

Both inherit from recent work in polynomial geometry:

- Hyperbolic polynomials
- Stable polynomials

What you need to know

We are interested in the eigenvalues of random frames:

$$\widehat{V} = \sum_i \widehat{v}_i \widehat{v}_i^T$$

What you need to know

We are interested in the eigenvalues of random frames:

$$\widehat{V} = \sum_i \widehat{v}_i \widehat{v}_i^T$$

All known techniques for this require concentration of measure and (as a result) weaken as the dimension grows.

What you need to know

We are interested in the eigenvalues of random frames:

$$\widehat{V} = \sum_i \widehat{v}_i \widehat{v}_i^T$$

All known techniques for this require concentration of measure and (as a result) weaken as the dimension grows.

We will look for new techniques by doing something seemingly absurd: study their (random) characteristic polynomials.

What you need to know

We are interested in the eigenvalues of random frames:

$$\widehat{V} = \sum_i \widehat{v}_i \widehat{v}_i^T$$

All known techniques for this require concentration of measure and (as a result) weaken as the dimension grows.

We will look for new techniques by doing something seemingly absurd: study their (random) characteristic polynomials.

In the case that we have root separation, we actually have a chance for this to work.

What you need to know

We are interested in the eigenvalues of random frames:

$$\widehat{V} = \sum_i \widehat{v}_i \widehat{v}_i^T$$

All known techniques for this require concentration of measure and (as a result) weaken as the dimension grows.

We will look for new techniques by doing something seemingly absurd: study their (random) characteristic polynomials.

In the case that we have root separation, we actually have a chance for this to work.

In exchange for requiring extra structure, we are hoping to get some new "polynomial techniques" that we can use.

Outline

Return on investment

To find separating intervals, we can use results in polynomial theory.

Return on investment

To find separating intervals, we can use results in polynomial theory.

Let p be a real rooted polynomial of degree d and q a real rooted polynomial of degree $d-1$

$$p(x) = \prod_{i=1}^{d}(x-\alpha_i) \quad \text{and} \quad q(x) = \prod_{i=1}^{d-1}(x-\beta_i)$$

with $\alpha_1 \leq \cdots \leq \alpha_d$ and $\beta_1 \leq \cdots \leq \beta_{d-1}$.

Return on investment

To find separating intervals, we can use results in polynomial theory.

Let p be a real rooted polynomial of degree d and q a real rooted polynomial of degree $d-1$

$$p(x) = \prod_{i=1}^{d}(x-\alpha_i) \quad \text{and} \quad q(x) = \prod_{i=1}^{d-1}(x-\beta_i)$$

with $\alpha_1 \leq \cdots \leq \alpha_d$ and $\beta_1 \leq \cdots \leq \beta_{d-1}$.

We say q *interlaces* p if $\alpha_1 \leq \beta_1 \leq \alpha_2 \cdots \leq \alpha_{d-1} \leq \beta_{d-1} \leq \alpha_d$.

Think: The roots of q separate the roots of p.

Return on investment

To find separating intervals, we can use results in polynomial theory.

Let p be a real rooted polynomial of degree d and q a real rooted polynomial of degree $d-1$

$$p(x) = \prod_{i=1}^{d}(x - \alpha_i) \quad \text{and} \quad q(x) = \prod_{i=1}^{d-1}(x - \beta_i)$$

with $\alpha_1 \leq \cdots \leq \alpha_d$ and $\beta_1 \leq \cdots \leq \beta_{d-1}$.

We say q *interlaces* p if $\alpha_1 \leq \beta_1 \leq \alpha_2 \cdots \leq \alpha_{d-1} \leq \beta_{d-1} \leq \alpha_d$.

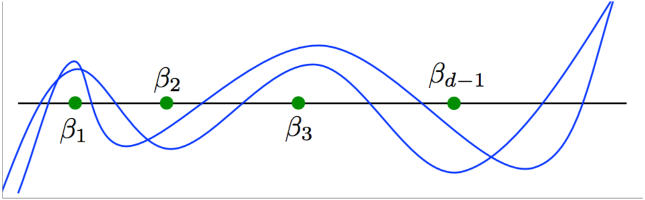
Think: The roots of q separate the roots of p.

Example: $p'(x)$ interlaces $p(x)$.

Common Interlacer

We say that degree d real rooted polynomials $p_1, \ldots, p_k$ have a *common interlacer* if there exists a q such that q interlaces *every* p_i simultaneously.

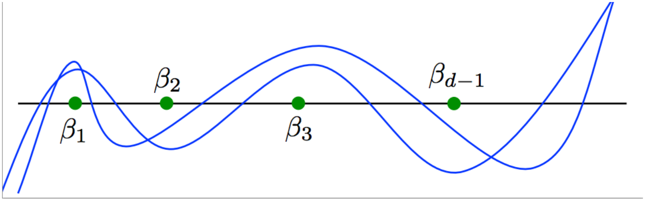
Think: the roots of q split up $\mathbb{R}$ into d intervals, each of which contains exactly one root of each p_i.



Common Interlacer

We say that degree d real rooted polynomials $p_1, \ldots, p_k$ have a *common interlacer* if there exists a q such that q interlaces *every* p_i simultaneously.

Think: the roots of q split up $\mathbb{R}$ into d intervals, each of which contains exactly one root of each p_i.



Note: if the p_i have a common interlacer (say q), then the intervals defined by the β_i can serve as separators for the lemma!

Back to the toy problem

Recall our goal was to understand the roots of

$$
\begin{aligned}
p(x) &= \frac{1}{2}\chi_{A+uu^T}(x) + \frac{1}{2}\chi_{A+vv^T}(x) \\
&= \frac{1}{2}q_0(x) + \frac{1}{2}q_1(x)
\end{aligned}
$$

Back to the toy problem

Recall our goal was to understand the roots of

$$\begin{aligned} p(x) &= \frac{1}{2}\chi_{A+uu^T}(x) + \frac{1}{2}\chi_{A+vv^T}(x) \\ &= \frac{1}{2}q_0(x) + \frac{1}{2}q_1(x) \end{aligned}$$

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

1. p and $\{q_i\}$ have the same degree and are all real rooted
2. The leading coefficients of the $\{q_i\}$ have the same sign
3. The collection of polynomials $\{q_i\}$ has a common interlacer
4. p is a convex combination of the $\{q_i\}$

Back to the toy problem

Recall our goal was to understand the roots of

$$p(x) = \frac{1}{2}\chi_{A+uu^T}(x) + \frac{1}{2}\chi_{A+vv^T}(x)$$
$$= \frac{1}{2}q_0(x) + \frac{1}{2}q_1(x)$$

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

1. p and $\{q_i\}$ have the same degree and are all real rooted
2. The leading coefficients of the $\{q_i\}$ have the same sign
3. The collection of polynomials $\{q_i\}$ has a common interlacer
4. 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

$$\mathrm{k^{th}root}\,(q_i) \le \mathrm{k^{th}root}\,(p) \le \mathrm{k^{th}root}\,(q_j)$$

More help from polynomials

Polynomial theory gives us a nice characterization of interlacing:

Lemma (Chudnovsky–Seymour, among others)

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

- *Every polynomial in the convex hull of $\{p_i\}$ has d real roots.*
- *The collection $\{p_i\}$ has a common interlacer.*

More help from polynomials

Polynomial theory gives us a nice characterization of interlacing:

Lemma (Chudnovsky–Seymour, among others)

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

- *Every polynomial in the convex hull of $\{p_i\}$ has d real roots.*
- *The collection $\{p_i\}$ has a common interlacer.*

Recall (again) our equation

$$p(x) = \frac{1}{2}\chi_{A+uu^T}(x) + \frac{1}{2}\chi_{A+vv^T}(x)$$

More help from polynomials

Polynomial theory gives us a nice characterization of interlacing:

Lemma (Chudnovsky–Seymour, among others)

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

- *Every polynomial in the convex hull of $\{p_i\}$ has d real roots.*
- *The collection $\{p_i\}$ has a common interlacer.*

Recall (again) our equation

$$p(x) = \frac{1}{2}\chi_{A+uu^T}(x) + \frac{1}{2}\chi_{A+vv^T}(x)$$

If we could show that

$$p(x) = \lambda\chi_{A+vv^T}(x) + (1-\lambda)\chi_{A+uu^T}(x)$$

was real rooted for all $\lambda \in [0,1]$, then we would get the interlacing for free.

Back to reality

But remember we are interested in random frames — that is, sums of *multiple* random vectors.

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

Back to reality

But remember we are interested in random frames — that is, sums of *multiple* random vectors.

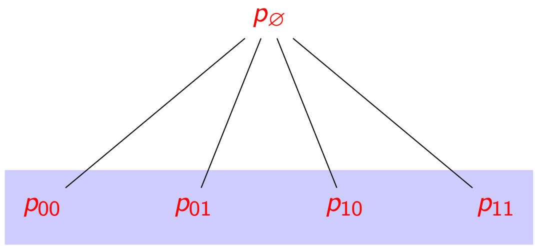
If all of the resulting characteristic polynomials had a common interlacer,

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

Back to reality

But remember we are interested in random frames — that is, sums of *multiple* random vectors.

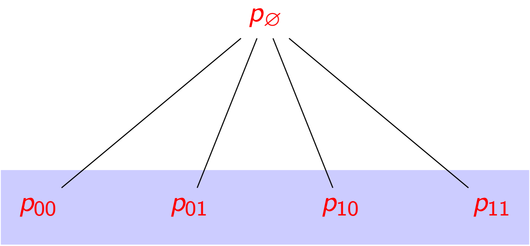
If all of the resulting characteristic polynomials had a common interlacer, we could study some convex combination and be able to use the lemma.



Back to reality

But remember we are interested in random frames — that is, sums of *multiple* random vectors.

If all of the resulting characteristic polynomials had a common interlacer, we could study some convex combination and be able to use the lemma.



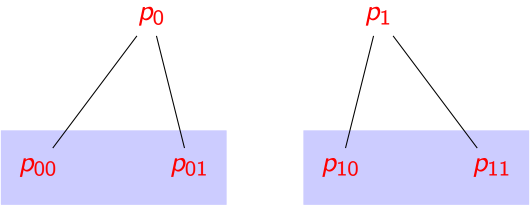
But in general they don't have a common interlacer...

Instead...

p_{00} $\quad$ p_{01} $\quad$ p_{10} $\quad$ p_{11}

Instead...

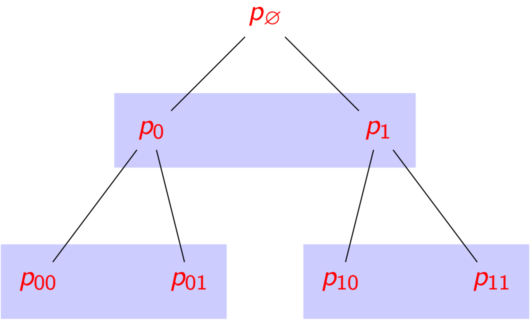
We can try to group them into smaller stars.



Instead...

We can try to group them into smaller stars.

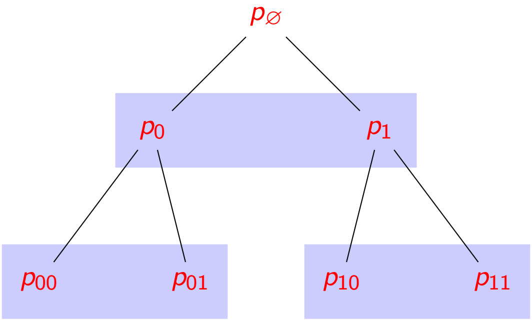
And then try to iterate.



Instead...

We can try to group them into smaller stars.

And then try to 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*

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

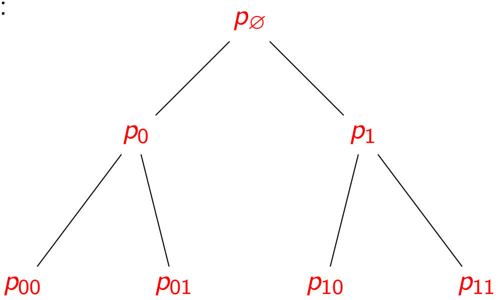
The punchline

Corollary

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

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

To find p_{leaf_i}:



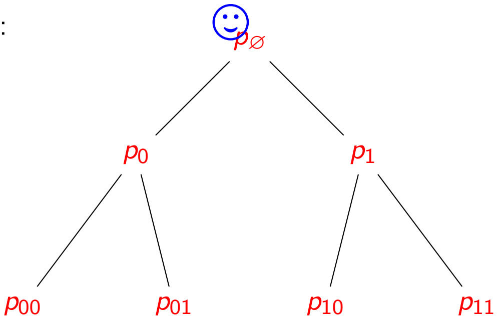
The punchline

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}:



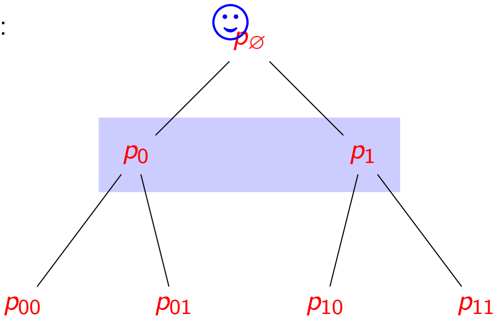
The punchline

Corollary

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

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

To find p_{leaf_i}:



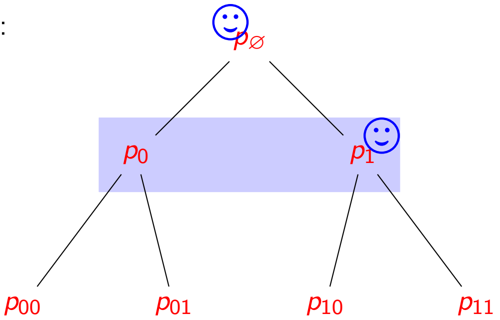
The punchline

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}:



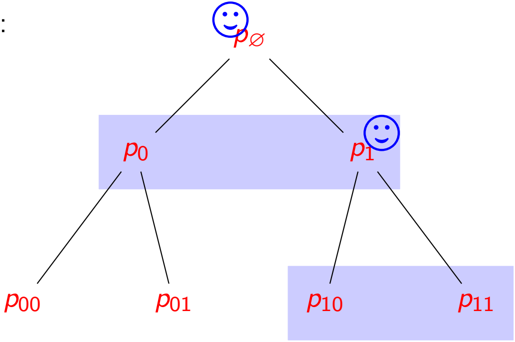
The punchline

Corollary

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

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

To find p_{leaf_i}:



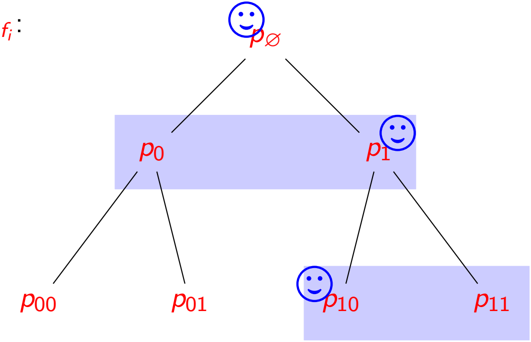
The punchline

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}:



Outline

Building an interlacing family

Consider the random frame

$$\widehat{V} = \sum_{i=1}^{m} \widehat{v}_i \widehat{v}_i^T$$

where the $\widehat{v}_i$ have support size at most n.

Building an interlacing family

Consider the random frame

$$\widehat{V} = \sum_{i=1}^{m} \widehat{v}_i \widehat{v}_i^T$$

where the $\widehat{v}_i$ have support size at most n.

We will define a *choice vector* $\sigma \in [n]^m$ where σ_i is the index of a vector in the support of $\widehat{v}_i$. Then the characteristic polynomial of a fixed frame V in the support $\widehat{V}$ can be denoted

$$p_\sigma(x) = \chi_V(x)$$

Building an interlacing family

Consider the random frame

$$\widehat{V} = \sum_{i=1}^{m} \widehat{v}_i \widehat{v}_i^T$$

where the $\widehat{v}_i$ have support size at most n.

We will define a *choice vector* $\sigma \in [n]^m$ where σ_i is the index of a vector in the support of $\widehat{v}_i$. Then the characteristic polynomial of a fixed frame V in the support $\widehat{V}$ can be denoted

$$p_\sigma(x) = \chi_V(x)$$

We then define *partial choice* vectors $\sigma' \in [n]^k$ for $k < m$; the corresponding polynomial will be the conditional expectation.

$$p_{\sigma'} = \mathbb{E}_{\widehat{v}_{k+1},\dots,\widehat{v}_d}\left[\chi(\widehat{V})(x) \mid \widehat{v}_i = v_i^{\sigma'_i} \text{ for } 1 \le i \le k\right]$$

Building an interlacing family

Consider the random frame

$$\widehat{V} = \sum_{i=1}^{m} \widehat{v}_i \widehat{v}_i^T$$

where the $\widehat{v}_i$ have support size at most n.

We will define a *choice vector* $\sigma \in [n]^m$ where σ_i is the index of a vector in the support of $\widehat{v}_i$. Then the characteristic polynomial of a fixed frame V in the support $\widehat{V}$ can be denoted

$$p_\sigma(x) = \chi_V(x)$$

We then define *partial choice* vectors $\sigma' \in [n]^k$ for $k < m$; the corresponding polynomial will be the conditional expectation.

$$p_{\sigma'} = \mathbb{E}_{\widehat{v}_{k+1},\ldots,\widehat{v}_d}\left[\chi(\widehat{V})(x) \mid \widehat{v}_i = v_i^{\sigma'_i} \text{ for } 1 \leq i \leq k\right]$$

This forms an n-ary tree with fixed assignments at the leaves and $p_\varnothing = \mathbb{E}\left[\chi_{\widehat{V}}(x)\right]$ at the root.

Mixed characteristic polynomials

If we choose the vectors independently, the polynomials take a special form:

Mixed characteristic polynomials

If we choose the vectors independently, the polynomials take a special form:

Theorem

Let $\widehat{v}_1, \ldots \widehat{v}_m$ be independent random vectors such that $\mathbb{E}\left[\widehat{v}_i \widehat{v}_i^T\right] = A_i$. Then

$$\mathbb{E}\left[\chi_{\widehat{V}}(x)\right] = \prod_{i=1}^{m} \left(1 - \frac{\partial}{\partial z_i}\right) \det\left[xI + \sum_{i=1}^{m} z_i A_i\right] \Bigg|_{z_1 = \cdots = z_m = 0}$$

In particular, the expected polynomial does not depend on the vectors or the probabilities — only the expected outer product.

Mixed characteristic polynomials

If we choose the vectors independently, the polynomials take a special form:

Theorem

Let $\widehat{v}_1, \dots \widehat{v}_m$ be independent random vectors such that $\mathbb{E}\left[\widehat{v}_i \widehat{v}_i^T\right] = A_i$. Then

$$\mathbb{E}\left[\chi_{\widehat{V}}(x)\right] = \prod_{i=1}^{m} \left(1 - \frac{\partial}{\partial z_i}\right) \det\left[xI + \sum_{i=1}^{m} z_i A_i\right] \Bigg|_{z_1 = \cdots = z_m = 0}$$

In particular, the expected polynomial does not depend on the vectors or the probabilities — only the expected outer product.

We call this a *mixed characteristic polynomial* and denote it $\mu[A_1, \dots, A_m]$.

A world of mixed characteristic polynomials

Every polynomial we defined previously is a mixed characteristic polynomial.

A world of mixed characteristic polynomials

Every polynomial we defined previously is a mixed characteristic polynomial.

1. Normal characteristic polynomials (for an assignment $\sigma = v_1, \ldots, v_m$ with $\sum_i v_i v_i^T = V$)

$$p_\sigma(x) = \chi_V(x) = \mu[v_1 v_1^T, \ldots, v_m v_m^T](x)$$

A world of mixed characteristic polynomials

Every polynomial we defined previously is a mixed characteristic polynomial.

1. Normal characteristic polynomials (for an assignment $\sigma = v_1, \ldots, v_m$ with $\sum_i v_i v_i^T = V$)

$$p_\sigma(x) = \chi_V(x) = \mu[v_1 v_1^T, \ldots, v_m v_m^T](x)$$

2. The expected characteristic polynomial (with $\mathbb{E}\left[\widehat{v}_i \widehat{v}_i^T\right] = A_i$)

$$\mathbb{E}\left[\chi_{\widehat{V}}(x)\right] = \mu[A_1, \ldots, A_m](x)$$

A world of mixed characteristic polynomials

Every polynomial we defined previously is a mixed characteristic polynomial.

1. Normal characteristic polynomials (for an assignment $\sigma = v_1, \ldots, v_m$ with $\sum_i v_i v_i^T = V$)

$$p_\sigma(x) = \chi_V(x) = \mu[v_1 v_1^T, \ldots, v_m v_m^T](x)$$

2. The expected characteristic polynomial (with $\mathbb{E}\left[\widehat{v}_i \widehat{v}_i^T\right] = A_i$)

$$\mathbb{E}\left[\chi_{\widehat{V}}(x)\right] = \mu[A_1, \ldots, A_m](x)$$

3. The partial assignment polynomials

$$\begin{aligned} p_{\sigma'} &= \mathbb{E}_{\widehat{v}_{k+1}, \ldots, \widehat{v}_d}\left[\chi_{\widehat{V}}(x) \mid \widehat{v}_i = v_i^{\sigma'_i} \text{ for } 1 \leq i \leq k\right] \\ &= \mu[v_1 v_1^T, \ldots, v_k v_k^T, A_{k+1}, \ldots, A_m] \end{aligned}$$

Real stable polynomials

The advantage of having a multivariate formula is that we can utilize the theory of *real stable polynomials*, a multivariate extension of real rooted polynomials. Let
$\mathbb{H} = \{x \in \mathbb{C} \mid \Im(z_i) > 0\}$.

Real stable polynomials

The advantage of having a multivariate formula is that we can utilize the theory of *real stable polynomials*, a multivariate extension of real rooted polynomials. Let
$\mathbb{H} = \{x \in \mathbb{C} \mid \Im(z_i) > 0\}$.

An n-variate polynomial p is called *stable* if it is never 0 in $\mathbb{H}^n$. (i.e. if $p(z_1, \dots, z_n) = 0$, then some z_i has nonnegative imaginary part). If, in addition, all coefficients of p are real, it is called *real stable*.

Real stable polynomials

The advantage of having a multivariate formula is that we can utilize the theory of *real stable polynomials*, a multivariate extension of real rooted polynomials. Let $\mathbb{H} = \{x \in \mathbb{C} \mid \Im(z_i) > 0\}$.

An n-variate polynomial p is called *stable* if it is never 0 in $\mathbb{H}^n$. (i.e. if $p(z_1, \ldots, z_n) = 0$, then some z_i has nonnegative imaginary part). If, in addition, all coefficients of p are real, it is called *real stable*.

Two important properties:

- Univariate polynomials are real rooted if and only if they are real stable.
- Real stable polynomials are closed under substitution of reals $(z_1, z_2, \ldots, z_n) \to (a, z_2, \ldots, z_n)$ for $a \in \mathbb{R}$.

Similar to *hyperbolic polynomials*.

Real stable techniques

There are numerous techniques for showing real stability. In particular,

Lemma

Let $A_1, \dots, A_m$ be Hermitian positive semidefinite matrices and $x_1 \dots x_m$ variables. Then

$$p(x_1, \dots, x_m) = \det \left[\sum_{i=1}^{m} x_i A_i \right]$$

is a real stable polynomial.

Real stable techniques

There are numerous techniques for showing real stability. In particular,

Lemma

Let $A_1, \ldots, A_m$ *be Hermitian positive semidefinite matrices and* $x_1 \ldots x_m$ *variables. Then*

$$p(x_1, \ldots, x_m) = \det \left[\sum_{i=1}^{m} x_i A_i \right]$$

is a real stable polynomial.

Lemma

If $p(x_1, \ldots, x_m)$ *is a real stable polynomial, then*

$$\left(1 - \frac{\partial}{\partial x_i} \right) p(x_1, \ldots, x_m) = p(\vec{x}) - \frac{\partial p(\vec{x})}{\partial x_i}$$

is a real stable polynomial.

Cutting to the chase

Theorem

Mixed characteristic polynomials are real rooted.

Cutting to the chase

Theorem

Mixed characteristic polynomials are real rooted.

Proof.

Follows directly from the formula:

$$\mu[A_1, \ldots, A_m](x) = \prod_{i=1}^{m} \left(1 - \frac{\partial}{\partial z_i}\right) \det \left[xI + \sum_{i=1}^{m} z_i A_i \right] \Bigg|_{z_1 = \cdots = z_m = 0}$$

$\square$

Cutting to the chase

Theorem

Mixed characteristic polynomials are real rooted.

Proof.

Follows directly from the formula:

$$\mu[A_1, \ldots, A_m](x) = \prod_{i=1}^{m} \left(1 - \frac{\partial}{\partial z_i}\right) \det \left[xI + \sum_{i=1}^{m} z_i A_i \right] \Big|_{z_1 = \cdots = z_m = 0}$$

□

This provides an easy way to generate interlacing families.

Corollary

Any tree of polynomials resulting from choosing independent random vectors forms an interlacing family.

Full Circle

So what have we accomplished?

Full Circle

So what have we accomplished?

We now have a “probabilistic” way to deal with roots of polynomials (under certain conditions).

Full Circle

So what have we accomplished?

We now have a "probabilistic" way to deal with roots of polynomials (under certain conditions).

In the case that we are choosing vectors independently and wanting to track the eigenvalues, those conditions are satisfied.

Full Circle

So what have we accomplished?

We now have a "probabilistic" way to deal with roots of polynomials (under certain conditions).

In the case that we are choosing vectors independently and wanting to track the eigenvalues, those conditions are satisfied.

Hence we have a "probabilistic" way to deal with eigenvalues. That is, for any given k, let R be the k^{th} root of the *expected characteristic polynomial* (under whatever product distribution you want). Then there exists

1. an assignment of the random vectors that has $\lambda_k \geq R$
2. an assignment of the random vectors that has $\lambda_k \leq R$

Outline

Motivation and the Fundamental Lemma

Exploiting separation: Interlacing Families

Interlacing Families associated with Mixed Characteristic Polynomials

Applications of Mixed Characteristic Polynomials
- Ramanujan Families
- Kadison–Singer
- Traveling Salesman

Summary

Who cares?

Matrices appear in a *lot* of places.

Because of this, a new tool for understanding eigenvalues can lead to new understanding (in a lot of places).

Who cares?

Matrices appear in a *lot* of places.

Because of this, a new tool for understanding eigenvalues can lead to new understanding (in a lot of places).

To use *our* tool, however, we must have added structure (an interlacing family).

So how useful is this new tool?

Who cares?

Matrices appear in a *lot* of places.

Because of this, a new tool for understanding eigenvalues can lead to new understanding (in a lot of places).

To use *our* tool, however, we must have added structure (an interlacing family).

So how useful is this new tool?

As it turns out*, just the subset of interlacing families that comes from mixed characteristic polynomials can be used to address a number of open problems.

Outline

Expander graphs

Expander Graphs are sparse, regular, well-connected graphs that "approximate" random graphs.

Expander graphs

Expander Graphs are sparse, regular, well-connected graphs that "approximate" random graphs.

- Sets of vertices have many external (equivalently, few internal) neighbors
- No "small" cuts
- Random walks mix quickly
- Can get from a to b using few edges

Expander graphs

Expander Graphs are sparse, regular, well-connected graphs that "approximate" random graphs.

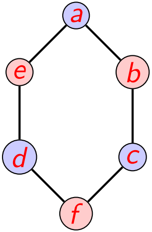
- Sets of vertices have many external (equivalently, few internal) neighbors
- No "small" cuts
- Random walks mix quickly
- Can get from a to b using few edges

Extremely important in theoretical computer science:

- Error-correcting codes
- Pseudorandom generators
- Computational complexity
  - PCP theorem (Dinur 2007)
  - SL=L (Reingold 2005)

Adjacency Matrix

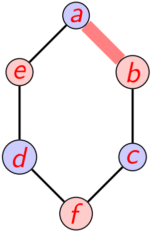
Given G with n vertices, the adjacency matrix A is defined as



$$\begin{array}{c|cccccc} & a & c & d & b & e & f \\ \hline a & 0 & 0 & 0 & 1 & 1 & 0 \\ c & 0 & 0 & 0 & 1 & 0 & 1 \\ d & 0 & 0 & 0 & 0 & 1 & 1 \\ b & 1 & 1 & 0 & 0 & 0 & 0 \\ e & 1 & 0 & 1 & 0 & 0 & 0 \\ f & 0 & 1 & 1 & 0 & 0 & 0 \end{array}$$

Adjacency Matrix

Given G with n vertices, the adjacency matrix A is defined as

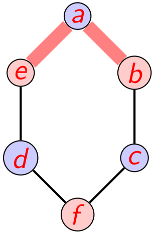


$$\begin{array}{c|cccccc} & a & c & d & b & e & f \\ \hline a & 0 & 0 & 0 & 1 & 1 & 0 \\ c & 0 & 0 & 0 & 1 & 0 & 1 \\ d & 0 & 0 & 0 & 0 & 1 & 1 \\ b & 1 & 1 & 0 & 0 & 0 & 0 \\ e & 1 & 0 & 1 & 0 & 0 & 0 \\ f & 0 & 1 & 1 & 0 & 0 & 0 \end{array}$$

1. $A_{i,j} = 1$ if and only if $\{v_i, v_j\} \in E$

Adjacency Matrix

Given G with n vertices, the adjacency matrix A is defined as

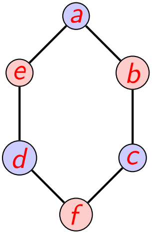


$$\begin{array}{c|cccccc} & a & c & d & b & e & f \\ \hline a & 0 & 0 & 0 & 1 & 1 & 0 \\ c & 0 & 0 & 0 & 1 & 0 & 1 \\ d & 0 & 0 & 0 & 0 & 1 & 1 \\ b & 1 & 1 & 0 & 0 & 0 & 0 \\ e & 1 & 0 & 1 & 0 & 0 & 0 \\ f & 0 & 1 & 1 & 0 & 0 & 0 \end{array}$$

1. $A_{i,j} = 1$ if and only if $\{v_i, v_j\} \in E$
2. If the graph is d-regular, each row sums to d

Adjacency Matrix

Given G with n vertices, the adjacency matrix A is defined as



$$\begin{array}{c|cccccc} & a & c & d & b & e & f \\ \hline a & 0 & 0 & 0 & 1 & 1 & 0 \\ c & 0 & 0 & 0 & 1 & 0 & 1 \\ d & 0 & 0 & 0 & 0 & 1 & 1 \\ b & 1 & 1 & 0 & 0 & 0 & 0 \\ e & 1 & 0 & 1 & 0 & 0 & 0 \\ f & 0 & 1 & 1 & 0 & 0 & 0 \end{array}$$

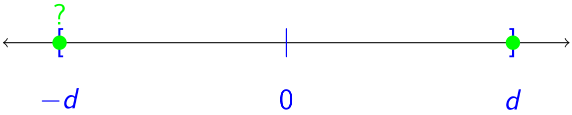
1. $A_{i,j} = 1$ if and only if $\{v_i, v_j\} \in E$
2. If the graph is d-regular, each row sums to d

Since A is symmetric, it has n real eigenvalues.

Eigenvalues

A d-regular graph has either 1 or 2 *trivial* eigenvalues

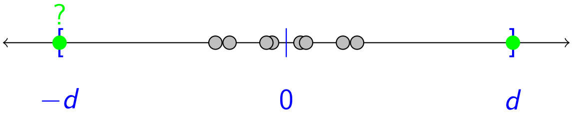
1. d is always the largest eigenvalue
2. G is bipartite if and only if $-d$ is an eigenvalue



Eigenvalues

A d-regular graph has either 1 or 2 *trivial* eigenvalues

1. d is always the largest eigenvalue
2. G is bipartite if and only if $-d$ is an eigenvalue

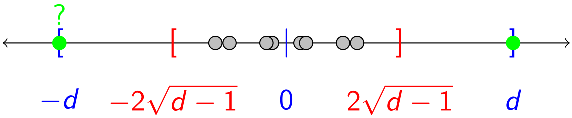


G is a good expander (spectrally) if all nontrivial eigenvalues are small (in absolute value).

Eigenvalues

A d-regular graph has either 1 or 2 *trivial* eigenvalues

1. d is always the largest eigenvalue
2. G is bipartite if and only if $-d$ is an eigenvalue



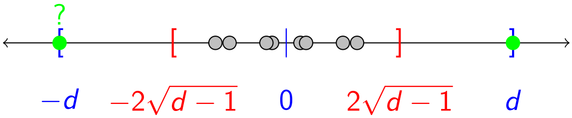
G is a good expander (spectrally) if all nontrivial eigenvalues are small (in absolute value).

A d-regular graph with all nontrivial eigenvalues inside $[-2\sqrt{d-1}, 2\sqrt{d-1}]$ is called a *Ramanujan graph* and an infinite collection (all d-regular) a *Ramanujan family*.

Eigenvalues

A d-regular graph has either 1 or 2 *trivial* eigenvalues

1. d is always the largest eigenvalue
2. G is bipartite if and only if $-d$ is an eigenvalue



G is a good expander (spectrally) if all nontrivial eigenvalues are small (in absolute value).

A d-regular graph with all nontrivial eigenvalues inside $[-2\sqrt{d-1}, 2\sqrt{d-1}]$ is called a *Ramanujan graph* and an infinite collection (all d-regular) a *Ramanujan family*.

Theorem (Alon, Boppana (1996))

No smaller interval can contain all nontrivial eigenvalues of an infinite collection of d-regular graphs.

Previous results

Theorem (Margulis, Lubotzky–Phillips–Sarnak (1988))

Ramanujan families exist for $d = p + 1$ where p is a prime number.

Previous results

Theorem (Margulis, Lubotzky–Phillips–Sarnak (1988))

Ramanujan families exist for $d = p + 1$ where p is a prime number.

Extended by Morganstern to $d = p^k + 1$, unknown for all other d.

All known constructions are algebraic — they are Cayley graphs of highly structured groups.

Previous results

Theorem (Margulis, Lubotzky–Phillips–Sarnak (1988))

Ramanujan families exist for $d = p + 1$ where p is a prime number.

Extended by Morganstern to $d = p^k + 1$, unknown for all other d.

All known constructions are algebraic — they are Cayley graphs of highly structured groups.

On the other hand, almost everything is almost Ramanujan:

Theorem (Friedman (2008))

A randomly chosen d-regular graph has its non-trivial eigenvalues in the interval

$$[-2\sqrt{d-1} - \epsilon, 2\sqrt{d-1} + \epsilon]$$

with high probability.

Previous results

Theorem (Margulis, Lubotzky–Phillips–Sarnak (1988))

Ramanujan families exist for $d = p + 1$ where p is a prime number.

Extended by Morganstern to $d = p^k + 1$, unknown for all other d.

All known constructions are algebraic — they are Cayley graphs of highly structured groups.

On the other hand, almost everything is almost Ramanujan:

Theorem (Friedman (2008))

A randomly chosen d-regular graph has its non-trivial eigenvalues in the interval

$$[-2\sqrt{d-1} - \epsilon, 2\sqrt{d-1} + \epsilon]$$

with high probability.

Obvious question: are Ramanujan families really that special?

Lifts

Using a technique they called *lifting*, Bilu and Linial (2006) suggested a method for finding Ramanujan families.

Lifts

Using a technique they called *lifting*, Bilu and Linial (2006) suggested a method for finding Ramanujan families.

For each edge $e \in G$, assign either $+1$ or -1. The vector of assignments $s \in \{\pm 1\}^{|E|}$ is called a *signing*.

Lifts

Using a technique they called *lifting*, Bilu and Linial (2006) suggested a method for finding Ramanujan families.

For each edge $e \in G$, assign either $+1$ or -1. The vector of assignments $s \in \{\pm 1\}^{|E|}$ is called a *signing*.

Multiplying each value in A by the corresponding sign from s gives the *signed adjacency matrix* A_s.

$$
\begin{array}{c|cccccc}
 & a & c & d & b & e & f \\
\hline
a & 0 & 0 & 0 & \boxed{-1} & 1 & 0 \\
c & 0 & 0 & 0 & 1 & 0 & 1 \\
d & 0 & 0 & 0 & 0 & 1 & 1 \\
b & \boxed{-1} & 1 & 0 & 0 & 0 & 0 \\
e & 1 & 0 & 1 & 0 & 0 & 0 \\
f & 0 & 1 & 1 & 0 & 0 & 0
\end{array}
$$

Main Eigenvalue lemma

To each signing s, they associate a graph G_s they call a *2-lift*.

Theorem (Bilu–Linial (2006))

Let G be a d-regular Ramanujan graph with n vertices and let s be a signing of G. If all eigenvalues of A_s lie in the interval

$$[-2\sqrt{d-1}, 2\sqrt{d-1}]$$

then the 2-lift G_s is a d-regular Ramanujan graph with $2n$ vertices.

Main Eigenvalue lemma

To each signing s, they associate a graph G_s they call a *2-lift*.

Theorem (Bilu–Linial (2006))

Let G be a d-regular Ramanujan graph with n vertices and let s be a signing of G. If all eigenvalues of A_s lie in the interval

$$[-2\sqrt{d-1}, 2\sqrt{d-1}]$$

then the 2-lift G_s is a d-regular Ramanujan graph with $2n$ vertices.

Conjecture (Bilu–Linial (2006))

Every *d-regular graph contains a signing s for which the eigenvalues of A_s lie inside the interval*

$$[-2\sqrt{d-1}, 2\sqrt{d-1}]$$

Main Eigenvalue lemma

To each signing s, they associate a graph G_s they call a *2-lift*.

Theorem (Bilu–Linial (2006))

Let G be a d-regular Ramanujan graph with n vertices and let s be a signing of G. If all eigenvalues of A_s lie in the interval

$$[-2\sqrt{d-1}, 2\sqrt{d-1}]$$

then the 2-lift G_s is a d-regular Ramanujan graph with $2n$ vertices.

Conjecture (Bilu–Linial (2006))

Every *d-regular graph contains a signing s for which the eigenvalues of A_s lie inside the interval*

$$[-2\sqrt{d-1}, 2\sqrt{d-1}]$$

We prove the conjecture for every *bipartite* graph G.

Bipartite Adjacency Matrices

What is so special about being bipartite?

Bipartite Adjacency Matrices

What is so special about being bipartite?

In this case, the signed adjacency matrix can be written in block form

$$\left(\begin{array}{c|c} 0 & B \\ \hline B^T & 0 \end{array}\right)$$

causing eigenvalues/vectors to come in pairs

$$v_i = [u_i \mid u_i] \quad \text{and} \quad v_{n-i} = [u_i \mid -u_i]$$

for $1 \leq i \leq n/2$ and so the eigenvalues satisfy $\lambda_i = -\lambda_{n-i}$.

Bipartite Adjacency Matrices

What is so special about being bipartite?

In this case, the signed adjacency matrix can be written in block form

$$\left(\begin{array}{c|c} 0 & B \\ \hline B^T & 0 \end{array}\right)$$

causing eigenvalues/vectors to come in pairs

$$v_i = [u_i \mid u_i] \quad \text{and} \quad v_{n-i} = [u_i \mid -u_i]$$

for $1 \leq i \leq n/2$ and so the eigenvalues satisfy $\lambda_i = -\lambda_{n-i}$.

Corollary

A bipartite signed adjacency matrix A_s has all of its eigenvalues in the interval

$$[-2\sqrt{d-1}, 2\sqrt{d-1}]$$

if and only if all of its eigenvalues are at most $2\sqrt{d-1}$.

Main idea

For each signing, we consider the characteristic polynomial of the signed adjacency matrix.

Main idea

For each signing, we consider the characteristic polynomial of the signed adjacency matrix.

These correspond to picking either

- $(\delta_i + \delta_j)$, or
- $(\delta_i - \delta_j)$

independently for each edge (v_i, v_j).

Main idea

For each signing, we consider the characteristic polynomial of the signed adjacency matrix.

These correspond to picking either

- $(\delta_i + \delta_j)$, or
- $(\delta_i - \delta_j)$

independently for each edge (v_i, v_j).

Corollary

The tree corresponding to these polynomials forms an interlacing family.

Main idea

For each signing, we consider the characteristic polynomial of the signed adjacency matrix.

These correspond to picking either

- $(\delta_i + \delta_j)$, or
- $(\delta_i - \delta_j)$

independently for each edge (v_i, v_j).

Corollary

The tree corresponding to these polynomials forms an interlacing family.

Hence it suffices to bound the largest root of the expected characteristic polynomial.

The expected characteristic polynomial

Theorem (Godsil–Gutman (1981))

For any graph G,

$$\mu_G(x) := \mathbb{E}_{s \in \{\pm\}^m} \chi_{A_s}(x) = \sum_i x^{n-2i}(-1)^i m_i$$

where m_i is the number of matchings in G of size i.

The expected characteristic polynomial

Theorem (Godsil–Gutman (1981))

For any graph G,

$$\mu_G(x) := \mathbb{E}_{s \in \{\pm\}^m} \chi_{A_s}(x) = \sum_i x^{n-2i} (-1)^i m_i$$

where m_i is the number of matchings in G of size i.

Heilmann and Lieb had introduced this polynomial in their study of monomers and dimers, and proved the following bound:

Theorem (Heilmann–Lieb (1972))

Let G be a graph with maximum degree Δ. Then

$$\operatorname{maxroot}(\mu_G) \leq 2\sqrt{\Delta - 1}$$

Piecing things together

Theorem

There exist bipartite Ramanujan families of degree d for any d.

Piecing things together

Theorem

There exist bipartite Ramanujan families of degree d for any d.

Proof.

Set $G_0 = K_{d,d}$ (which is bipartite, d-regular, and Ramanujan for any d). Given G_i, form G_{i+1} as follows:

Piecing things together

Theorem

There exist bipartite Ramanujan families of degree d for any d.

Proof.

Set $G_0 = K_{d,d}$ (which is bipartite, d-regular, and Ramanujan for any d). Given G_i, form G_{i+1} as follows:

For each possible signing $s \in \{\pm 1\}^{|E_i|}$, form the polynomials p_s. By our theorem, this is an interlacing family.

Piecing things together

Theorem

There exist bipartite Ramanujan families of degree d for any d.

Proof.

Set $G_0 = K_{d,d}$ (which is bipartite, d-regular, and Ramanujan for any d). Given G_i, form G_{i+1} as follows:

For each possible signing $s \in \{\pm 1\}^{|E_i|}$, form the polynomials p_s. By our theorem, this is an interlacing family.

Combining this with Godsil–Gutman and Heilmann–Lieb ensures some p_{s^*} such that $\mathrm{maxroot}(p_{s^*}) \leq \mathrm{maxroot}(p_\varnothing) \leq 2\sqrt{d-1}$

Piecing things together

Theorem

There exist bipartite Ramanujan families of degree d for any d.

Proof.

Set $G_0 = K_{d,d}$ (which is bipartite, d-regular, and Ramanujan for any d). Given G_i, form G_{i+1} as follows:

For each possible signing $s \in \{\pm 1\}^{|E_i|}$, form the polynomials p_s. By our theorem, this is an interlacing family.

Combining this with Godsil–Gutman and Heilmann–Lieb ensures some p_{s^*} such that $\mathrm{maxroot}(p_{s^*}) \leq \mathrm{maxroot}(p_\varnothing) \leq 2\sqrt{d-1}$

Set G_{i+1} to be the 2-lift associated with s^* — this is bipartite, d-regular, and (by Bilu and Linial) Ramanujan — and proceed by induction.

□

Outline

The original problem

In 1959, Kadison and Singer asked the following question:

The original problem

In 1959, Kadison and Singer asked the following question:

Question (Kadison–Singer)

Let $\mathcal{A}$ be a discrete maximal abelian subalgebra of $\mathcal{B}(\mathcal{H})$, the algebra of bounded linear operators on a (separable, complex) Hilbert space. Let $\rho : \mathcal{A} \to \mathbb{C}$ be a pure state on that subalgebra. Is the (pure) extension $\rho' : \mathcal{B}(\mathcal{H}) \to \mathbb{C}$ of ρ to all of $\mathcal{B}(\mathcal{H})$ unique?

The original problem

In 1959, Kadison and Singer asked the following question:

Question (Kadison–Singer)

Let $\mathcal{A}$ be a discrete maximal abelian subalgebra of $\mathcal{B}(\mathcal{H})$, the algebra of bounded linear operators on a (separable, complex) Hilbert space. Let $\rho : \mathcal{A} \to \mathbb{C}$ be a pure state on that subalgebra. Is the (pure) extension $\rho' : \mathcal{B}(\mathcal{H}) \to \mathbb{C}$ of ρ to all of $\mathcal{B}(\mathcal{H})$ unique?

Related to the mathematical formalization of quantum physics via C^* algebras — can a quantum state be uniquely determined without altering it?

The original problem

In 1959, Kadison and Singer asked the following question:

Question (Kadison–Singer)

Let $\mathcal{A}$ be a discrete maximal abelian subalgebra of $\mathcal{B}(\mathcal{H})$, the algebra of bounded linear operators on a (separable, complex) Hilbert space. Let $\rho : \mathcal{A} \to \mathbb{C}$ be a pure state on that subalgebra. Is the (pure) extension $\rho' : \mathcal{B}(\mathcal{H}) \to \mathbb{C}$ of ρ to all of $\mathcal{B}(\mathcal{H})$ unique?

Related to the mathematical formalization of quantum physics via C^* algebras — can a quantum state be uniquely determined without altering it?

Kadison and Singer showed it is NOT true for continuous subalgebras!

The original problem

In 1959, Kadison and Singer asked the following question:

Question (Kadison–Singer)

Let $\mathcal{A}$ be a discrete maximal abelian subalgebra of $\mathcal{B}(\mathcal{H})$, the algebra of bounded linear operators on a (separable, complex) Hilbert space. Let $\rho : \mathcal{A} \to \mathbb{C}$ be a pure state on that subalgebra. Is the (pure) extension $\rho' : \mathcal{B}(\mathcal{H}) \to \mathbb{C}$ of ρ to all of $\mathcal{B}(\mathcal{H})$ unique?

Related to the mathematical formalization of quantum physics via C^* algebras — can a quantum state be uniquely determined without altering it?

Kadison and Singer showed it is NOT true for continuous subalgebras!

And then people worked on it.

A World of Equivalences

Kadison–Singer
(1959)

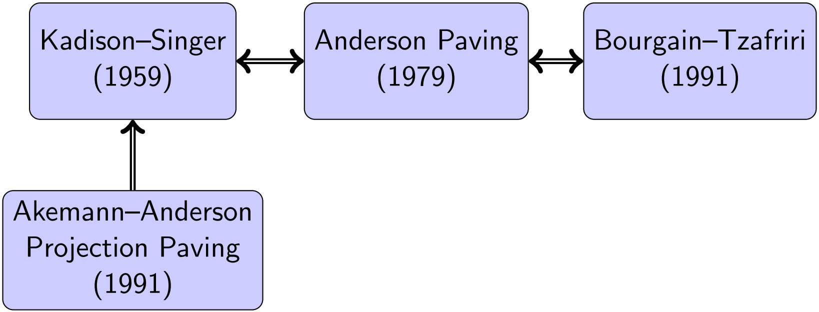
A World of Equivalences

Kadison–Singer (1959) $\Longleftrightarrow$ Anderson Paving (1979)

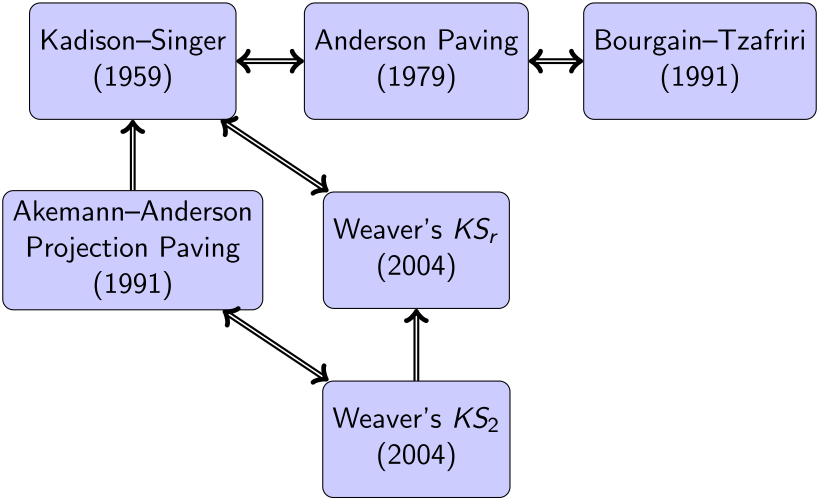
A World of Equivalences

Kadison–Singer (1959) $\Longleftrightarrow$ Anderson Paving (1979) $\Longleftrightarrow$ Bourgain–Tzafriri (1991)

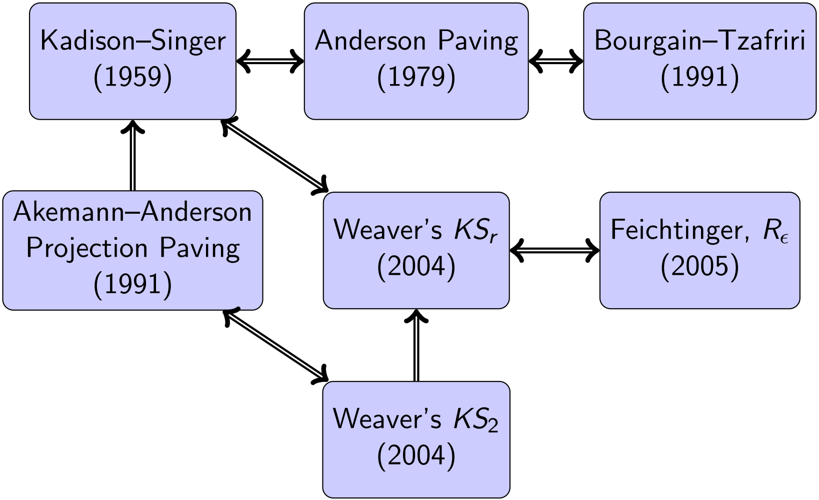
A World of Equivalences



A World of Equivalences



A World of Equivalences



Weaver's Conjecture

Conjecture (KS_2)

There exist **universal constants** $\epsilon, \theta > 0$ *such that the following holds: for all* $w_1, \ldots, w_m \in \mathbb{C}^d$ *satisfying*

$$\sum_i w_i w_i^* = I \quad \text{and} \quad \|w_i\|^2 \leq \epsilon$$

for all i*, there exists a subset of the vectors* S *such that*

$$\theta I \prec \sum_{i \in S} w_i w_i^* \prec (1 - \theta) I$$

Weaver's Conjecture

Conjecture (KS_2)

There exist **universal constants** $\epsilon, \theta > 0$ *such that the following holds: for all* $w_1, \ldots, w_m \in \mathbb{C}^d$ *satisfying*

$$\sum_i w_i w_i^* = I \quad \textit{and} \quad \|w_i\|^2 \leq \epsilon$$

for all i*, there exists a subset of the vectors* S *such that*

$$\theta I \prec \sum_{i \in S} w_i w_i^* \prec (1 - \theta) I$$

Can a frame be "split" into two pieces such that both pieces have similar eigenvalues?

The obvious next question:

Bourgain and Tzafriri (1991) showed that a uniformly random choice works with high probability if

$$\|w_i\|^2 \leq \frac{C}{\log d}$$

Uses matrix concentration inequalities similar to Rudelson (1999) and Ahlswede–Winter (2002).

The obvious next question:

Bourgain and Tzafriri (1991) showed that a uniformly random choice works with high probability if

$$\|w_i\|^2 \leq \frac{C}{\log d}$$

Uses matrix concentration inequalities similar to Rudelson (1999) and Ahlswede–Winter (2002).

Recall: all of these inequalities have two things in common:

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

The obvious next question:

Bourgain and Tzafriri (1991) showed that a uniformly random choice works with high probability if

$$\|w_i\|^2 \leq \frac{C}{\log d}$$

Uses matrix concentration inequalities similar to Rudelson (1999) and Ahlswede–Winter (2002).

Recall: all of these inequalities have two things in common:

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

Also recall: this will *always* be true — tight concentration (in this respect) depends on the dimension (consider multiple copies of the basis vectors).

The obvious next question:

Bourgain and Tzafriri (1991) showed that a uniformly random choice works with high probability if

$$\|w_i\|^2 \leq \frac{C}{\log d}$$

Uses matrix concentration inequalities similar to Rudelson (1999) and Ahlswede–Winter (2002).

Recall: all of these inequalities have two things in common:

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

Also recall: this will *always* be true — tight concentration (in this respect) depends on the dimension (consider multiple copies of the basis vectors).

Weaver's conjecture: you can trade the $\log d$ factor in exchange for nonzero (instead of high) probability.

Probabilistic framework

First, we need to put this in a probabilistic framework:
Rather than saying $w_i \in S$ or $w_i \notin S$, we can say $\widehat{v}_i$ is a random vector choosing between

$$\left\{ \begin{pmatrix} w_i \\ 0^d \end{pmatrix}, \begin{pmatrix} 0^d \\ w_i \end{pmatrix} \right\}$$

each with probability $1/2$.

Probabilistic framework

First, we need to put this in a probabilistic framework:
Rather than saying $w_i \in S$ or $w_i \notin S$, we can say $\widehat{v_i}$ is a random vector choosing between

$$\left\{ \begin{pmatrix} w_i \\ 0^d \end{pmatrix}, \begin{pmatrix} 0^d \\ w_i \end{pmatrix} \right\}$$

each with probability $1/2$.

So for a given S, the matrix

$$M_S = \begin{pmatrix} \sum_{i \in S} w_i w_i^* & 0_{d \times d} \\ 0_{d \times d} & \sum_{i \notin S} w_i w_i^* \end{pmatrix}$$

Probabilistic framework

First, we need to put this in a probabilistic framework:
Rather than saying $w_i \in S$ or $w_i \notin S$, we can say $\widehat{v}_i$ is a random vector choosing between

$$\left\{ \begin{pmatrix} w_i \\ 0^d \end{pmatrix}, \begin{pmatrix} 0^d \\ w_i \end{pmatrix} \right\}$$

each with probability $1/2$.

So for a given S, the matrix

$$M_S = \begin{pmatrix} \sum_{i \in S} w_i w_i^* & 0_{d \times d} \\ 0_{d \times d} & \sum_{i \notin S} w_i w_i^* \end{pmatrix}$$

Bounding the largest eigenvalue of M_S bounds the largest *and* smallest eigenvalues of the subset!

Mixed characteristic polynomials

If we choose the vectors independently, the polynomials form an interlacing family.

Mixed characteristic polynomials

If we choose the vectors independently, the polynomials form an interlacing family.

So suffices to bound the largest root of

$$\mu[A_1, \ldots, A_m](x) = \prod_{i=1}^{m} \left(1 - \frac{\partial}{\partial z_i}\right) \det \left[xI + \sum_{i=1}^{m} z_i A_i\right] \Big|_{z_1 = \cdots = z_m = 0}$$

over all A_i with $\sum_i A_i = I$ and $\mathrm{Tr}\left[A_i\right] \leq \epsilon$.

Mixed characteristic polynomials

If we choose the vectors independently, the polynomials form an interlacing family.

So suffices to bound the largest root of

$$\mu[A_1, \ldots, A_m](x) = \prod_{i=1}^{m} \left(1 - \frac{\partial}{\partial z_i}\right) \det \left[xI + \sum_{i=1}^{m} z_i A_i \right] \Bigg|_{z_1 = \cdots = z_m = 0}$$

over all A_i with $\sum_i A_i = I$ and $\mathrm{Tr}\,[A_i] \leq \epsilon$.

The bound uses the multivariate structure: start at $\det\left[xI + \sum_{i=1}^{m} z_i A_i\right]$, and apply $\left(1 - \frac{\partial}{\partial z_i}\right)$ one at a time.

What happens to the roots?

“Roots” of multivariate polynomials

Rather than having roots that are points, multivariate polynomials have *zero surfaces*.

"Roots" of multivariate polynomials

Rather than having roots that are points, multivariate polynomials have *zero surfaces*.

Allows for techniques from *real algebraic geometry and convex optimization.*

Potential function

We use a *multivariate potential function* to help understand the behavior as the operators are applied.

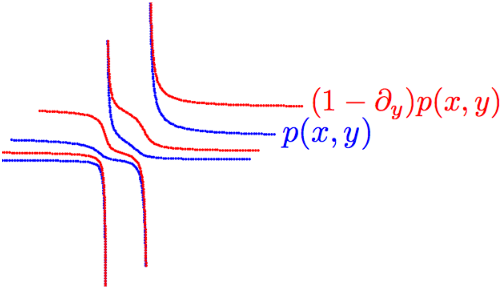
$$\Phi_p^i(z_1, \ldots, z_m) = \frac{\partial}{\partial z_i} \log p(z_1, \ldots, z_m)$$

Potential function

We use a *multivariate potential function* to help understand the behavior as the operators are applied.

$$\Phi_p^i(z_1, \ldots, z_m) = \frac{\partial}{\partial z_i} \log p(z_1, \ldots, z_m)$$

Applying the operator $(1 - \partial_{z_j})$ causes the roots to shift closer.

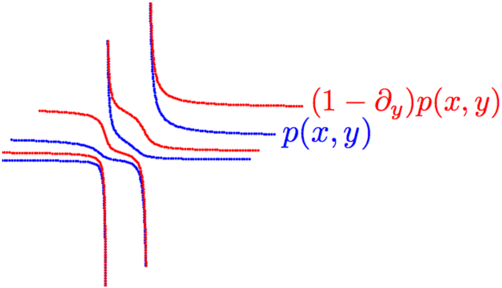


Potential function

We use a *multivariate potential function* to help understand the behavior as the operators are applied.

$$\Phi_p^i(z_1, \ldots, z_m) = \frac{\partial}{\partial z_i} \log p(z_1, \ldots, z_m)$$

Applying the operator $(1 - \partial_{z_j})$ causes the roots to shift closer.



Φ_p^j tells us how much we need to move in the direction of the shift to get back the original cushion (in all other directions).

Back to Weaver's problem

This leads to the following theorem:

Theorem

Let $w_1, \ldots w_m \in \mathbb{C}^d$ be vectors such that

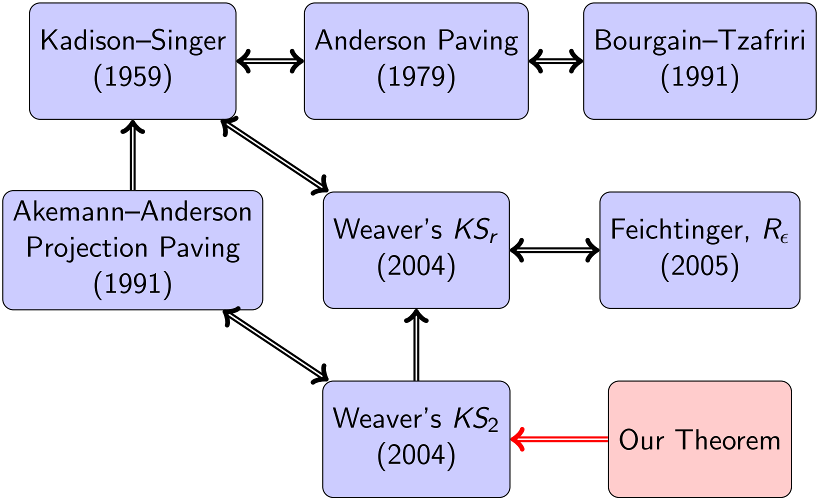
$$\sum_{i=1}^{m} w_i w_i^* = I \quad \textit{and} \quad \|w_i\|^2 \leq \epsilon$$

for all i. Then there exists a partition of $[m]$ into sets $S_1, \ldots, S_r$ such that

$$\left\| \sum_{i \in S_j} w_i w_i^* \right\| \leq \frac{1}{r}(1 + \sqrt{r\epsilon})^2.$$

for all j.

Proof 1


Kadison–Singer (1959)
Anderson Paving (1979)
Bourgain–Tzafriri (1991)
Akemann–Anderson Projection Paving (1991)
Weaver's KS_r (2004)
Feichtinger, R_ϵ (2005)
Weaver's KS_2 (2004)
Our Theorem

Outline

Extension

Akemann and Weaver extend the previous theorem to arbitrary subsums.

Theorem (Akemann, Weaver (2014))

Let $w_1, \ldots, w_m \in \mathbb{C}^d$ *such that*

$$\sum_i w_i w_i^* \leq I \quad \textit{and} \quad \|w_i\|^2 \leq \epsilon$$

for all i. *Then for any collection of real numbers* $0 \leq t_1, \ldots, t_m \leq 1$, *there exists an* $S \subseteq [m]$ *such that*

$$\left\| \sum_{i \in S} w_i w_i^* - \sum_i t_i w_i w_i^* \right\| \leq O(\epsilon^{1/16})$$

Extension

Akemann and Weaver extend the previous theorem to arbitrary subsums.

Theorem (Akemann, Weaver (2014))

Let $w_1, \ldots, w_m \in \mathbb{C}^d$ *such that*

$$\sum_i w_i w_i^* \leq I \quad \textit{and} \quad \|w_i\|^2 \leq \epsilon$$

for all i. *Then for any collection of real numbers* $0 \leq t_1, \ldots, t_m \leq 1$, *there exists an* $S \subseteq [m]$ *such that*

$$\left\| \sum_{i \in S} w_i w_i^* - \sum_i t_i w_i w_i^* \right\| \leq O(\epsilon^{1/16})$$

A *Lyapunov-type theorem*: says fractional sums can be well-approximated by discrete sums.

Linear Programming

Given vectors $c, a_1, \ldots, a_m \in \mathbb{R}^d$ and $b_1, \ldots, b_m \in \mathbb{R}$

$$
\begin{array}{lll}
(P^{\mathbb{Z}}) & \text{minimize} & \langle c, x \rangle \\
& \text{subject to} & \langle a_i, x \rangle \leq b_i \text{ for all } i \\
& & x \in \mathbb{Z}^d
\end{array}
$$

Linear Programming

Given vectors $c, a_1, \ldots, a_m \in \mathbb{R}^d$ and $b_1, \ldots, b_m \in \mathbb{R}$

$$
\begin{array}{lll}
(P^{\mathbb{Z}}) & \text{minimize} & \langle c, x \rangle \\
 & \text{subject to} & \langle a_i, x \rangle \leq b_i \text{ for all } i \\
 & & x \in \mathbb{Z}^d
\end{array}
$$

Optimization over integers is typically hard to solve! So instead...

Linear Programming

Given vectors $c, a_1, \ldots, a_m \in \mathbb{R}^d$ and $b_1, \ldots, b_m \in \mathbb{R}$

$$(P^{\mathbb{Z}}) \qquad \begin{array}{ll} \text{minimize} & \langle c, x \rangle \\ \text{subject to} & \langle a_i, x \rangle \le b_i \text{ for all } i \\ & x \in \mathbb{Z}^d \end{array}$$

Optimization over integers is typically hard to solve! So instead...

$$(P^{\mathbb{R}}) \qquad \begin{array}{ll} \text{minimize} & \langle c, x \rangle \\ \text{subject to} & \langle a_i, x \rangle \le b_i \text{ for all } i \\ & x \in \mathbb{R}^d \end{array}$$

$(P^{\mathbb{R}})$ is called the *relaxation* of $(P^{\mathbb{Z}})$.

Linear Programming

Given vectors $c, a_1, \ldots, a_m \in \mathbb{R}^d$ and $b_1, \ldots, b_m \in \mathbb{R}$

$$\begin{array}{lll} (P^{\mathbb{Z}}) & \text{minimize} & \langle c, x \rangle \\ & \text{subject to} & \langle a_i, x \rangle \leq b_i \text{ for all } i \\ & & x \in \mathbb{Z}^d \end{array}$$

Optimization over integers is typically hard to solve! So instead...

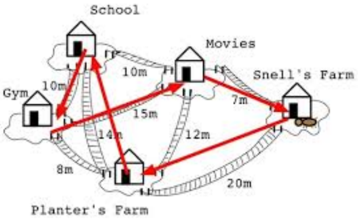
$$\begin{array}{lll} (P^{\mathbb{R}}) & \text{minimize} & \langle c, x \rangle \\ & \text{subject to} & \langle a_i, x \rangle \leq b_i \text{ for all } i \\ & & x \in \mathbb{R}^d \end{array}$$

$(P^{\mathbb{R}})$ is called the *relaxation* of $(P^{\mathbb{Z}})$.

If *all* fractional solutions are close to some discrete solution, these are close.

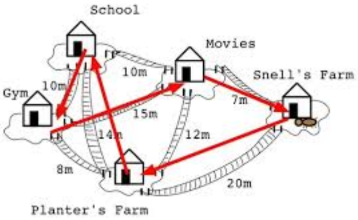
Traveling Salesman

What is the shortest "tour"?



Traveling Salesman

What is the shortest "tour"?

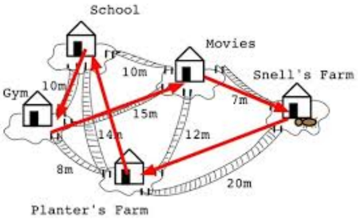


In general, can have *asymmetric* costs (one way roads, NYC tolls)

Asymmetric case notoriously harder to understand than symmetric case.

Traveling Salesman

What is the shortest "tour"?



In general, can have *asymmetric* costs (one way roads, NYC tolls)

Asymmetric case notoriously harder to understand than symmetric case.

Known to be NP-hard and representable by an integer linear program $P^{\mathbb{Z}}$.

Approximation Algorithm

Theorem (Anari–Oveis Gharan, 2014)

For an asymmetric traveling salesman problem on an n vertex graph, we have

$$\frac{\mathrm{opt}(P^{\mathbb{Z}})}{\mathrm{opt}(P^{\mathbb{R}})} \leq O(poly \log\log(n))$$

Previously best known bound: $O(\log n)$.

Approximation Algorithm

Theorem (Anari–Oveis Gharan, 2014)

For an asymmetric traveling salesman problem on an n vertex graph, we have

$$\frac{\mathrm{opt}(P^{\mathbb{Z}})}{\mathrm{opt}(P^{\mathbb{R}})} \leq O(\mathit{poly}\log\log(n))$$

Previously best known bound: $O(\log n)$.

Proof involves an extension of mixed characteristic polynomials to *homogeneous strong Rayleigh measures*.

Then uses a *spanning-tree measure* to show existence of *spectrally thin trees* (which can be used to build a good tour).

Approximation Algorithm

Theorem (Anari–Oveis Gharan, 2014)

For an asymmetric traveling salesman problem on an n vertex graph, we have

$$\frac{\mathrm{opt}(P^{\mathbb{Z}})}{\mathrm{opt}(P^{\mathbb{R}})} \leq O(\mathit{poly} \log \log(n))$$

Previously best known bound: $O(\log n)$.

Proof involves an extension of mixed characteristic polynomials to *homogeneous strong Rayleigh measures*.

Then uses a *spanning-tree measure* to show existence of *spectrally thin trees* (which can be used to build a good tour).

Interesting part: proof is completely existential.

State of the world

Just to be clear:

Assume you are given an asymmetric distance graph on n vertices.

State of the world

Just to be clear:
Assume you are given an asymmetric distance graph on n vertices.

You can provably approximate the best *total cost* of a tour within a factor of $O(\mathit{poly} \log \log n)$, but never find a tour that achieves it.

State of the world

Just to be clear:
Assume you are given an asymmetric distance graph on n vertices.

You can provably approximate the best *total cost* of a tour within a factor of $O(poly \log \log n)$, but never find a tour that achieves it.

In particular, given a tour T, you can know that T is suboptimal without being able to do better.

State of the world

Just to be clear:
Assume you are given an asymmetric distance graph on n vertices.

You can provably approximate the best *total cost* of a tour within a factor of $O(poly \log \log n)$, but never find a tour that achieves it.

In particular, given a tour T, you can know that T is suboptimal without being able to do better.

Could have interesting repercussions in complexity theory...

Outline

Takeaways

So what should you take from this talk?

Takeaways

So what should you take from this talk?

We have developed a new technique that uses polynomials to understand eigenvalues of random matrices.

This involves considering the (seemingly unrelated) expectation of characteristic polynomials.

Takeaways

So what should you take from this talk?

We have developed a new technique that uses polynomials to understand eigenvalues of random matrices.

This involves considering the (seemingly unrelated) expectation of characteristic polynomials.

By using polynomials, we are able to leverage results from areas such as real algebraic geometry and convex optimization.

As a result, we can show bounds that occur with *low* probability — something that was previously impossible.

Moving Forward

Other results are in progress:

1. Connections to Noncommutative (Free) Probability

2. General techniques for proving inequalities on roots

Moving Forward

Other results are in progress:

1. Connections to Noncommutative (Free) Probability
2. General techniques for proving inequalities on roots

And other directions are ripe for exploration:

1. Connections to Convex Optimization
2. Weakening the dependence on real rootedness
3. Finding more interlacing families!!

Thanks

Thank you for inviting me to speak today.

Thanks

Thank you for inviting me to speak today.

And thank you for your attention!