Interlacing Families and Kadison–Singer

Adam W. Marcus

Crisply, LLC
Yale University
adam.marcus@yale.edu

December 6, 2013

Joint work with:

Dan Spielman
Yale University

Nikhil Srivastava
Microsoft Research, India

My involvement partially supported by:

National Science Foundation
Mathematical Sciences Postdoctoral Research Fellowship

Notations and Simplifications

Throughout this talk, the following will (hopefully) hold:
Notations:

- d is the dimension of the vector space
- m is the number of vectors
- v is a (deterministic) vector
- $\widehat{v}$ is a random variable that takes deterministic vectors as its possible values
- V is a matrix

Notations and Simplifications

Throughout this talk, the following will (hopefully) hold:
Notations:

- d is the dimension of the vector space
- m is the number of vectors
- v is a (deterministic) vector
- $\widehat{v}$ is a random variable that takes deterministic vectors as its possible values
- V is a matrix

Simplifications:

- All vectors will be real vectors (in $\mathbb{R}^d$) [though all proofs will hold for complex vectors by replacing transposes with adjoints]
- All random vectors will choose from 2 possible values [though proofs can be extended to any type of random vector]

Goals

In this talk I plan to

1. Give a brief history of Kadison–Singer and its relatives
2. Convey how we went about attacking the problem
3. Introduce a technique for showing the existence of combinatorial objects we call "the method of interlacing polynomials"
4. Introduce a class of polynomials we call "mixed characteristic polynomials".
5. Use these to prove two known equivalents of Kadison–Singer
6. Discuss some related open questions

not necessarily (but mostly) in this order .

Goals

In this talk I plan to

1. Give a brief history of Kadison–Singer and its relatives
2. Convey how we went about attacking the problem
3. Introduce a technique for showing the existence of combinatorial objects we call "the method of interlacing polynomials"
4. Introduce a class of polynomials we call "mixed characteristic polynomials".
5. Use these to prove two known equivalents of Kadison–Singer
6. Discuss some related open questions

not necessarily (but mostly) in this order .

And *please* interrupt if there are any questions.

Outline

The beginning of quantum mechanics

Dirac wrote in his 1930 book *Principles of Quantum Mechanics*, "how to introduce a representation [of a quantum state] in practice."

The beginning of quantum mechanics

Dirac wrote in his 1930 book *Principles of Quantum Mechanics*, "how to introduce a representation [of a quantum state] in practice."

- Measure a *complete commuting set* of observables (as much as you can simultaneously without introducing uncertainty).
- Express the observed probability distribution as an ensemble over pure states (extremal points of the algebra)
- Generalize this to the entire system

The beginning of quantum mechanics

Dirac wrote in his 1930 book *Principles of Quantum Mechanics*, "how to introduce a representation [of a quantum state] in practice."

- Measure a *complete commuting set* of observables (as much as you can simultaneously without introducing uncertainty).
- Express the observed probability distribution as an ensemble over pure states (extremal points of the algebra)
- Generalize this to the entire system

This predates the formalization of quantum mechanics using C^* algebras.

Kadison–Singer

Kadison and Singer read this in 1959, and (knowing the C^* algebra formalism) asked if the "generalization" step was unique.

Kadison–Singer

Kadison and Singer read this in 1959, and (knowing the C^* algebra formalism) asked if the "generalization" step was unique.

Precisely:

Question

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?

Note: Pure states are the rank 1 operators *plus* a bunch more guaranteed by Axiom of Choice (these are the ones Dirac ignored).

Kadison–Singer

Kadison and Singer read this in 1959, and (knowing the C^* algebra formalism) asked if the "generalization" step was unique.

Precisely:

Question

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?

Note: Pure states are the rank 1 operators *plus* a bunch more guaranteed by Axiom of Choice (these are the ones Dirac ignored).

Showed it is *not* true in the continuous case (their counterexample was $\mathcal{H} = L^2([0, 1])$).

A World of Equivalences

Kadison–Singer
(1959)

A World of Equivalences

Conjecture (Anderson Paving Conjecture (1979))

For every $\epsilon > 0$, there is a universal positive integer $k = k(\epsilon)$ so that for every zero-diagonal finite matrix A with n rows (and columns), there exists a partition $\{S_1, \ldots, S_k\}$ of $[n]$ so that

$$\|A[S_i, S_i]\| \leq \epsilon \|A\|$$

for all $i \in [k]$.

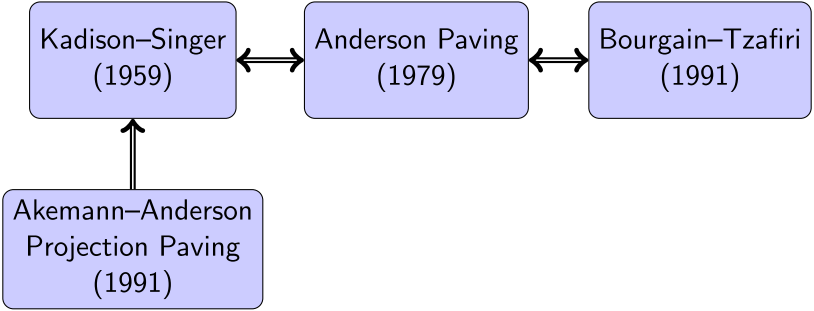
A World of Equivalences

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

A World of Equivalences

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

A World of Equivalences



A World of Equivalences

Conjecture (Weaver's KS_r (2004))

There exist universal constants $\eta \geq 2$ and $\theta > 0$ such that the following holds: if $w_1, \ldots, w_m \in \mathbb{C}^d$ satisfy $\|w_i\| \leq 1$ for all i and

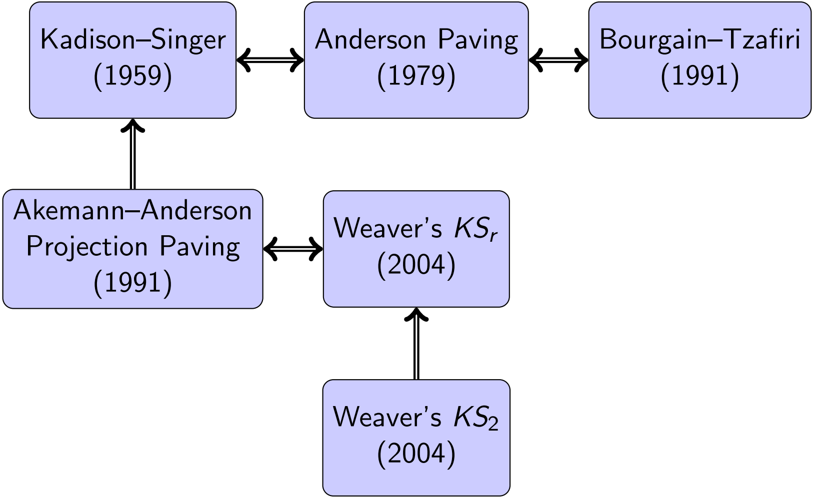
$$\sum_i |\langle u, w_i \rangle|^2 = \eta$$

for all unit vectors $u \in \mathbb{C}^d$. Then there exists a partition of the vectors into parts $\{S_1, \ldots, S_r\}$ so that

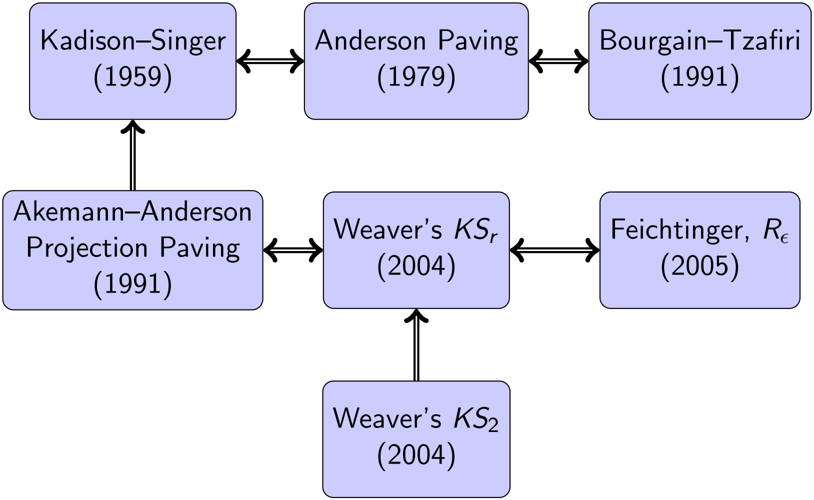
$$\sum_{i \in S_j} |\langle u, w_i \rangle|^2 \leq \eta - \theta$$

for all unit vectors $u \in \mathbb{C}^d$ and each $j \in [r]$.

A World of Equivalences



A World of Equivalences



Partitioning Identity

A collection of vectors $v_1, \ldots, v_m$ is called a *Parseval frame* if

$$\sum_i v_i v_i^T = I$$

Partitioning Identity

A collection of vectors $v_1, \ldots, v_m$ is called a *Parseval frame* if

$$\sum_i v_i v_i^T = I$$

Parseval frames act like orthonormal bases, but allow you to split the work over more vectors.

Examples:

1. Orthonormal bases
2. Regular simplex centered at the origin (properly scaled)
3. The union of other Parseval frames (properly scaled)
4. Any collection you want transformed by $\left(\sum_i v_i v_i^T\right)^{-1/2}$

Partitioning Identity

A collection of vectors $v_1, \ldots, v_m$ is called a *Parseval frame* if

$$\sum_i v_i v_i^T = I$$

Parseval frames act like orthonormal bases, but allow you to split the work over more vectors.
Examples:

1. Orthonormal bases
2. Regular simplex centered at the origin (properly scaled)
3. The union of other Parseval frames (properly scaled)
4. Any collection you want transformed by $\left(\sum_i v_i v_i^T\right)^{-1/2}$

Can a Parseval frame be partitioned into subsets which are "almost Parseval"?

Partitioning Identity

A collection of vectors $v_1, \ldots, v_m$ is called a *Parseval frame* if

$$\sum_i v_i v_i^T = I$$

Parseval frames act like orthonormal bases, but allow you to split the work over more vectors.
Examples:

1. Orthonormal bases
2. Regular simplex centered at the origin (properly scaled)
3. The union of other Parseval frames (properly scaled)
4. Any collection you want transformed by $\left(\sum_i v_i v_i^T\right)^{-1/2}$

Can a Parseval frame be partitioned into subsets which are "almost Parseval"?

In general, no (there may be huge vectors). So what if the vectors are all bounded in size?

What if I choose randomly?

Tropp (2011) showed that a uniformly random choice of vectors works with high probability if

$$\|v_i\| \leq \frac{C(\epsilon)}{\log d}$$

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

What if I choose randomly?

Tropp (2011) showed that a uniformly random choice of vectors works with high probability if

$$\|v_i\| \leq \frac{C(\epsilon)}{\log d}$$

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

This is known to be best possible (it is tight when each v_i is a scaled elementary basis vector).

What if I choose randomly?

Tropp (2011) showed that a uniformly random choice of vectors works with high probability if

$$\|v_i\| \le \frac{C(\epsilon)}{\log d}$$

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

This is known to be best possible (it is tight when each v_i is a scaled elementary basis vector).

The goal would be to trade the $\log d$ factor in exchange for nonzero (instead of high) probability.

In case you missed something

What you need to know for this talk:

1. Kadison and Singer asked a fundamental question about the mathematical foundations of quantum physics

In case you missed something

What you need to know for this talk:

1. Kadison and Singer asked a fundamental question about the mathematical foundations of quantum physics
2. Numerous other problems have since been shown to be equivalent

In case you missed something

What you need to know for this talk:

1. Kadison and Singer asked a fundamental question about the mathematical foundations of quantum physics
2. Numerous other problems have since been shown to be equivalent
3. The fundamental question is whether a Parseval frame can be partitioned into two "almost Parseval" frames

In case you missed something

What you need to know for this talk:

1. Kadison and Singer asked a fundamental question about the mathematical foundations of quantum physics
2. Numerous other problems have since been shown to be equivalent
3. The fundamental question is whether a Parseval frame can be partitioned into two “almost Parseval” frames
4. The fundamental question is true (with high probability) when the vectors have norm $O(1/\log d)$

In case you missed something

What you need to know for this talk:

1. Kadison and Singer asked a fundamental question about the mathematical foundations of quantum physics
2. Numerous other problems have since been shown to be equivalent
3. The fundamental question is whether a Parseval frame can be partitioned into two "almost Parseval" frames
4. The fundamental question is true (with high probability) when the vectors have norm $O(1/\log d)$
5. We want to know what happens when the vectors bounded in norm by a *constant*

Outline

Probabilistic Approach

We want to understand the “random” version of the problem.

Probabilistic Approach

We want to understand the “random” version of the problem.

Let $\widehat{v}_1, \ldots, \widehat{v}_m$ be random vectors that are Parseval in expectation:

$$\mathbb{E}\left[\widehat{v}_i \widehat{v}_i^T\right] = I$$

Will there always be a realization of these vectors that is close to Parseval?

Probabilistic Approach

We want to understand the "random" version of the problem.

Let $\widehat{v}_1, \ldots, \widehat{v}_m$ be random vectors that are Parseval in expectation:

$$\mathbb{E}\left[\widehat{v}_i \widehat{v}_i^T\right] = I$$

Will there always be a realization of these vectors that is close to Parseval?

Again, no. Consider when some vector takes values $2e_1$ and $2e_2$ each with probability $1/2$.

Probabilistic Approach

We want to understand the "random" version of the problem.

Let $\widehat{v}_1, \ldots, \widehat{v}_m$ be random vectors that are Parseval in expectation:

$$\mathbb{E}\left[\widehat{v}_i \widehat{v}_i^T\right] = I$$

Will there always be a realization of these vectors that is close to Parseval?

Again, no. Consider when some vector takes values $2e_1$ and $2e_2$ each with probability $1/2$.

But what if (as before) the vectors were also bounded in norm?

Main Theorem

Our main technical theorem says the following:

Theorem

Let $\epsilon > 0$ and $\widehat{v}_1, \ldots \widehat{v}_m$ be independent random vectors such that

$$\sum_{i=1}^{m} \mathbb{E}\left[\widehat{v}_i \widehat{v}_i^T\right] = I$$

and

$$\mathbb{E}\left[\|\widehat{v}_i\|^2\right] \leq \epsilon$$

for all i. Then there exists an assignment $\widehat{v}_i = v_i$ such that

$$\left\| \sum_{i=1}^{m} v_i v_i^T \right\| \leq (1 + \sqrt{\epsilon})^2.$$

In this talk, I will assume $\epsilon < 1/4$ and prove this for $(1 + 3\sqrt{\epsilon})$ (a slight weakening).

Translation and experimentation:

There are two benefits to dealing with vectors (rather than algebras):

1. Translation into the world of polynomials
   Given vectors $v_1, \ldots, v_m$, set $V = \sum_i v_i {v_i}^T$. Then the maximum eigenvalue of V is the largest root of $\chi_V(x)$ (the characteristic polynomial). So we can turn this into a question about the behavior of a special class of polynomials.
2. This formulation is ripe for experimentation.
   We can optimize over collections of vectors that satisfy the given constraints to see what the worst case scenarios are. We can see what the average scenarios are. We can see what the worst case average scenarios are.

Experimental Observations

So that is what we did, and these are the observations we made.

Experimental Observations

So that is what we did, and these are the observations we made.

Observation 1: The expected characteristic polynomial $\mathbb{E}\left[\chi_{\widehat{V}}(x)\right]$ always seemed to be real-rooted (that is, all roots r had $\Im(r) = 0$).

Experimental Observations

So that is what we did, and these are the observations we made.

Observation 1: The expected characteristic polynomial $\mathbb{E}\left[\chi_{\widehat{V}}(x)\right]$ always seemed to be real-rooted (that is, all roots r had $\Im(r) = 0$).

Observation 2: When we looked at all possible values of $\widehat{V} = \sum_i \widehat{v}_i \widehat{v}_i^T$ (satisfying the hypotheses of the theorem), there always seemed to be one whose characteristic polynomial $\chi_V(x)$ had a smaller largest root than the expected characteristic polynomial $\mathbb{E}\left[\chi_{\widehat{V}}(x)\right]$.

Experimental Observations, cont.

Observation 3: The expected characteristic polynomial $\mathbb{E}\left[\chi_{\widehat{V}}(x)\right]$ seemed to have maximal largest root when $\mathbb{E}\left[\|\widehat{v}_i\|^2\right] = \epsilon$ and $\mathbb{E}\left[\widehat{v}_i \widehat{v}_i^T\right] = \frac{\epsilon}{d} I$ for all i.

Experimental Observations, cont.

Observation 3: The expected characteristic polynomial $\mathbb{E}\left[\chi_{\widehat{V}}(x)\right]$ seemed to have maximal largest root when $\mathbb{E}\left[\|\widehat{v}_i\|^2\right] = \epsilon$ and $\mathbb{E}\left[\widehat{v}_i\widehat{v}_i^T\right] = \frac{\epsilon}{d}I$ for all i.

Observation 4: In the case where $\mathbb{E}\left[\|\widehat{v}_i\|^2\right] = \epsilon$ and $\mathbb{E}\left[\|\widehat{v}_i\widehat{v}_i^T\|^2\right] = \frac{\epsilon}{d}I$ and for all i, the expected characteristic polynomial is an associated Laguerre polynomial (a classical orthogonal polynomial whose roots satisfy the bounds we were hoping for).

Suggests an approach

If we could

1. bound the largest root of $\mathbb{E}\left[\chi_{\widehat{V}}(x)\right]$ over the set of random vectors that satisfy the hypotheses of the theorem, and then
2. show there always exists an assignment $v_1, \ldots, v_m$ such that the largest root of $\chi_V(x)$ is smaller than the largest root of $\mathbb{E}\left[\chi_{\widehat{V}}(x)\right]$

then this would prove our theorem (and this is our approach).

Suggests an approach

If we could

1. bound the largest root of $\mathbb{E}\left[\chi_{\widehat{V}}(x)\right]$ over the set of random vectors that satisfy the hypotheses of the theorem, and then
2. show there always exists an assignment $v_1, \ldots, v_m$ such that the largest root of $\chi_V(x)$ is smaller than the largest root of $\mathbb{E}\left[\chi_{\widehat{V}}(x)\right]$

then this would prove our theorem (and this is our approach).

We start with part (2).

Suggests an approach

If we could

1. bound the largest root of $\mathbb{E}\left[\chi_{\widehat{V}}(x)\right]$ over the set of random vectors that satisfy the hypotheses of the theorem, and then
2. show there always exists an assignment $v_1, \ldots, v_m$ such that the largest root of $\chi_V(x)$ is smaller than the largest root of $\mathbb{E}\left[\chi_{\widehat{V}}(x)\right]$

then this would prove our theorem (and this is our approach).

We start with part (2). Main Idea: define a process by which we pick the assignments one by one and try to understand how the (now conditional) expected characteristic polynomial changes.

Outline

Formalization

We establish notation to help keep track of assignments that have been made.

Formalization

We establish notation to help keep track of assignments that have been made.

We define a *choice vector* $\sigma \in \{0,1\}^m$ where σ_i corresponds to which realization vector $\widehat{v_i}$ takes.
Then we can reference the characteristic polynomial of a fixed assignment as

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

Formalization

We establish notation to help keep track of assignments that have been made.

We define a *choice vector* $\sigma \in \{0,1\}^m$ where σ_i corresponds to which realization vector $\widehat{v}_i$ takes.
Then we can reference the characteristic polynomial of a fixed assignment as

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

We also define a *partial choice* vector $\sigma' \in \{0,1\}^k$ $(k < m)$. The corresponding polynomial will be the conditional expectation.

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

Formalization

We establish notation to help keep track of assignments that have been made.

We define a *choice vector* $\sigma \in \{0,1\}^m$ where σ_i corresponds to which realization vector $\widehat{v}_i$ takes.
Then we can reference the characteristic polynomial of a fixed assignment as

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

We also define a *partial choice* vector $\sigma' \in \{0,1\}^k$ $(k < m)$. The corresponding polynomial will be the conditional expectation.

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

Note that $p_\varnothing = \mathbb{E}\left[\chi_{\widehat{V}}(x)\right]$, the expected characteristic polynomial we are interested in.

Sums of polynomials

We have the relation

$$p_{\sigma'}(x) = p_{\sigma',0}(x)\mathbb{P}\left[\widehat{v}_{k+1} = v_{k+1}^{0}\right] + p_{\sigma',1}(x)\mathbb{P}\left[\widehat{v}_{k+1} = v_{k+1}^{1}\right]$$

but now we have reached our first major issue.

Sums of polynomials

We have the relation

$$p_{\sigma'}(x) = p_{\sigma',0}(x)\mathbb{P}\left[\widehat{v}_{k+1} = v_{k+1}^0\right] + p_{\sigma',1}(x)\mathbb{P}\left[\widehat{v}_{k+1} = v_{k+1}^1\right]$$

but now we have reached our first major issue.

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.

Sums of polynomials

We have the relation

$$p_{\sigma'}(x) = p_{\sigma',0}(x)\mathbb{P}\left[\widehat{v}_{k+1} = v^0_{k+1}\right] + p_{\sigma',1}(x)\mathbb{P}\left[\widehat{v}_{k+1} = v^1_{k+1}\right]$$

but now we have reached our first major issue.

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.

Approach: forget this issue and see what we can prove.

A Lemma

Lemma

Let f *and* g *be monic polynomials. Assume there exists a point* $c \in \mathbb{R}$ *such that* f *and* g *each has* exactly *one real root larger than* c *(call these the "extreme roots"). Then the largest real root of* $f + g$ *lies between these extreme roots.*

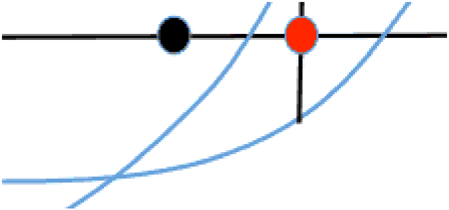
A Lemma

Lemma

Let f *and* g *be monic polynomials. Assume there exists a point* $c \in \mathbb{R}$ *such that* f *and* g *each has* exactly *one real root larger than* c *(call these the "extreme roots"). Then the largest real root of* $f + g$ *lies between these extreme roots.*

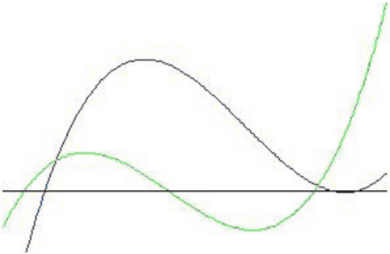
Proof.

By picture

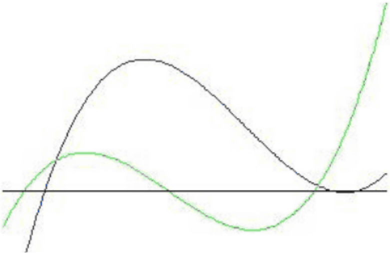


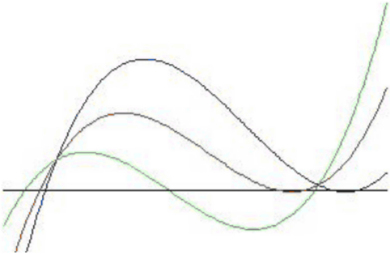
□

Without c to "anchor"



Without c to "anchor"





So what?

So now we can say something about the real roots, but not necessarily the complex roots.

But recall Observation 1: the expected characteristic polynomial seemed to have all real roots. If this was *always* true, we would be in good shape.

So what?

So now we can say something about the real roots, but not necessarily the complex roots.

But recall Observation 1: the expected characteristic polynomial seemed to have all real roots. If this was *always* true, we would be in good shape.

While our original polynomials (characteristic polynomials of Hermitian matrices) are real-rooted, in general the sums of real-rooted polynomials can be arbitrary.

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 $\pm\sqrt{-6}$).

Equation Revisited

Back to our equation

$$p_{\sigma'}(x) = p_{\sigma',0}(x)\mathbb{P}\left[\widehat{v}_{k+1} = v_{k+1}^{0}\right] + p_{\sigma',1}(x)\mathbb{P}\left[\widehat{v}_{k+1} = v_{k+1}^{1}\right]$$

Equation Revisited

Back to our equation

$$p_{\sigma'}(x) = p_{\sigma',0}(x)\mathbb{P}\left[\widehat{v}_{k+1} = v_{k+1}^0\right] + p_{\sigma',1}(x)\mathbb{P}\left[\widehat{v}_{k+1} = v_{k+1}^1\right]$$

The lemma tells us that if

1. $p_{\sigma'}(x)$ is real-rooted
2. $p_{\sigma',0}(x)$ is real-rooted
3. $p_{\sigma',1}(x)$ is real-rooted
4. There exists a c "anchoring" the largest roots of $p_{\sigma',0}(x)$ and $p_{\sigma',1}(x)$

Then we know the largest root of $p_{\sigma'}$ lies between the largest root of $p_{\sigma',0}$ and the largest root of $p_{\sigma',1}$.

Equation Revisited

Back to our equation

$$p_{\sigma'}(x) = p_{\sigma',0}(x)\mathbb{P}\left[\widehat{v}_{k+1} = v_{k+1}^0\right] + p_{\sigma',1}(x)\mathbb{P}\left[\widehat{v}_{k+1} = v_{k+1}^1\right]$$

The lemma tells us that if

1. $p_{\sigma'}(x)$ is real-rooted
2. $p_{\sigma',0}(x)$ is real-rooted
3. $p_{\sigma',1}(x)$ is real-rooted
4. There exists a c "anchoring" the largest roots of $p_{\sigma',0}(x)$ and $p_{\sigma',1}(x)$

Then we know the largest root of $p_{\sigma'}$ lies between the largest root of $p_{\sigma',0}$ and the largest root of $p_{\sigma',1}$.

Let's worry about c for the moment (keeping real-rootedness on the back burner).

Interlacing polynomials

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

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

with $\alpha_1 \leq \cdots \leq \alpha_n$ and $\beta_1 \leq \cdots \leq \beta_{n-1}$

Interlacing polynomials

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

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

with $\alpha_1 \leq \cdots \leq \alpha_n$ and $\beta_1 \leq \cdots \leq \beta_{n-1}$

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

Think: The roots of q separate the roots of p

Interlacing polynomials

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

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

with $\alpha_1 \leq \cdots \leq \alpha_n$ and $\beta_1 \leq \cdots \leq \beta_{n-1}$

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

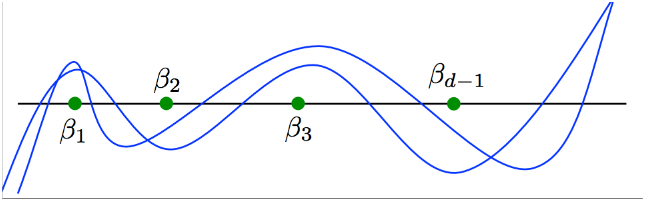
Think: The roots of q separate the roots of p

Example 1: $p'(x)$ interlaces $p(x)$
Example 2: If p has no multiple roots (and largest root R), then let $q = p/(x - R)$. Then $q(x + \epsilon)$ interlaces $p(x)$

Common Interlacers

We say that two degree n polynomials p and r have a *common interlacer* if there exists a q such that q interlaces *both* p and r simultaneously.

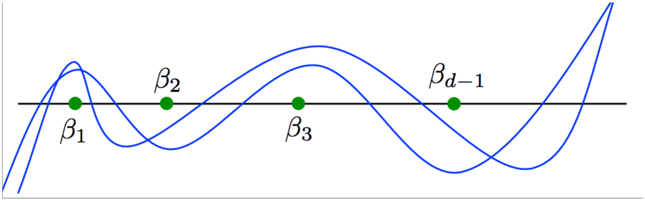
Think: the roots of q split up $\mathbb{R}$ into n intervals, each of which contains exactly one root of p and one root of r



Common Interlacers

We say that two degree n polynomials p and r have a *common interlacer* if there exists a q such that q interlaces *both* p and r simultaneously.

Think: the roots of q split up $\mathbb{R}$ into n intervals, each of which contains exactly one root of p and one root of r



Note, if p and r have a common interlacer (say q), then $c = \beta_{d-1}$ can serve as the anchor from the lemma!

Interlacing families

We say $\{p\}_{\sigma \in \Sigma}$ is an *interlacing family* if for all partial assignments σ' we have that

1. Each polynomial $p_{\sigma'}$ is real-rooted, and
2. The polynomials $p_{\sigma',0}$ and $p_{\sigma',1}$ have a common interlacer

Interlacing families

We say $\{p\}_{\sigma \in \Sigma}$ is an *interlacing family* if for all partial assignments σ' we have that

1. Each polynomial $p_{\sigma'}$ is real-rooted, and
2. The polynomials $p_{\sigma',0}$ and $p_{\sigma',1}$ have a common interlacer

Corollary

If $\{p\}_{\sigma}$ forms an interlacing family, then there exists an assignment σ_0 such that the largest root of p_{σ_0} is at most the largest root of $p_{\varnothing}$ (the expected polynomial).

Proof.

Start at the expected polynomial and walk backwards. $\square$

Interlacing for free

Fortunately, the interlacing follows directly from a well-known lemma:

Lemma (folklore, Fisk)

Let f, g be polynomials of the same degree such that every $\lambda f + (1 - \lambda) g$ is real-rooted for all $\lambda \in [0, 1]$. Then f and g have a common interlacer.

Interlacing for free

Fortunately, the interlacing follows directly from a well-known lemma:

Lemma (folklore, Fisk)

Let f, g be polynomials of the same degree such that every $\lambda f + (1-\lambda) g$ is real-rooted for all $\lambda \in [0,1]$. Then f and g have a common interlacer.

Recall (again) our equation

$$p_{\sigma'}(x) = p_{\sigma',0}(x)\mathbb{P}\left[\widehat{v}_{k+1} = v^0_{k+1}\right] + p_{\sigma',1}(x)\mathbb{P}\left[\widehat{v}_{k+1} = v^1_{k+1}\right]$$

Interlacing for free

Fortunately, the interlacing follows directly from a well-known lemma:

Lemma (folklore, Fisk)

Let f, g *be polynomials of the same degree such that every* $\lambda f + (1-\lambda) g$ *is real-rooted for all* $\lambda \in [0,1]$. *Then* f *and* g *have a common interlacer.*

Recall (again) our equation

$$p_{\sigma'}(x) = p_{\sigma',0}(x)\mathbb{P}\left[\widehat{v}_{k+1} = v^0_{k+1}\right] + p_{\sigma',1}(x)\mathbb{P}\left[\widehat{v}_{k+1} = v^1_{k+1}\right]$$

If we are able to show that $p_{\sigma'}$ is real-rooted (independent of the probabilities on the vectors) then we get interlacing for free!

Interlacing for free

Fortunately, the interlacing follows directly from a well-known lemma:

Lemma (folklore, Fisk)

Let f, g be polynomials of the same degree such that every $\lambda f + (1-\lambda)g$ is real-rooted for all $\lambda \in [0,1]$. Then f and g have a common interlacer.

Recall (again) our equation

$$p_{\sigma'}(x) = p_{\sigma',0}(x)\mathbb{P}\left[\widehat{v}_{k+1} = v^0_{k+1}\right] + p_{\sigma',1}(x)\mathbb{P}\left[\widehat{v}_{k+1} = v^1_{k+1}\right]$$

If we are able to show that $p_{\sigma'}$ is real-rooted (independent of the probabilities on the vectors) then we get interlacing for free!

Time to pull real-rootedness from the back burner.

Where to start?

The issue with real-rooted polynomials is that it is hard to see how to get from one to another.

Parking garage phenomenon

The issue with real-rooted polynomials is that it is hard to see how to get from one to another.

Unless you consider them to be a projection of higher dimensional objects.

Real stable polynomials

There have been many recent advances in understanding real-rootedness using theory of *real stable polynomials*, a multivariate extension of real-rooted polynomials.

Real stable polynomials

There have been many recent advances in understanding real-rootedness using theory of *real stable polynomials*, a multivariate extension of real-rooted polynomials.

A polynomial p is *real stable* if all coefficients are real and $p(z_1, \ldots, z_n) \neq 0$ whenever $\Im(z_i) > 0$ for all i (if $p(z_1, \ldots, z_n) = 0$ then some z_i has $\Im(z_i) \leq 0$).

Real stable polynomials

There have been many recent advances in understanding real-rootedness using theory of *real stable polynomials*, a multivariate extension of real-rooted polynomials.

A polynomial p is *real stable* if all coefficients are real and $p(z_1, \ldots, z_n) \neq 0$ whenever $\Im(z_i) > 0$ for all i (if $p(z_1, \ldots, z_n) = 0$ then some z_i has $\Im(z_i) \leq 0$).

Some 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 $a \in \mathbb{R}$.

Real stable polynomials

There have been many recent advances in understanding real-rootedness using theory of *real stable polynomials*, a multivariate extension of real-rooted polynomials.

A polynomial p is *real stable* if all coefficients are real and $p(z_1, \ldots, z_n) \neq 0$ whenever $\Im(z_i) > 0$ for all i (if $p(z_1, \ldots, z_n) = 0$ then some z_i has $\Im(z_i) \leq 0$).

Some 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*.

Borcea and Brändén

Borcea and Brändén developed 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 real stable.

Lemma

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

$$p(x_1, \ldots, x_m) - \frac{\partial p(x_1, \ldots, x_m)}{\partial x_i}$$

is real stable.

Our polynomials

Fortunately, our polynomials have a nice general form.

Theorem

Let $\widehat{v}_1, \ldots \widehat{v}_m$ be 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.

Our polynomials

Fortunately, our polynomials have a nice general form.

Theorem

Let $\widehat{v}_1, \ldots \widehat{v}_m$ be 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 $\mu[A_1, \ldots, A_m]$.

A world of mixed characteristic polynomials

Every polynomial we have seen so far is a mixed characteristic polynomial.

A world of mixed characteristic polynomials

Every polynomial we have seen so far 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
$$\begin{aligned} p_{\sigma'} &= \mathbb{E}_{v_{k+1},\ldots,v_d}\left[\chi_{\widehat{V}}(x) \mid \widehat{v}_i = v_i^{\sigma'_i} \text{ for } 1 \le i \le k\right] \\ &= \mu[v_1 v_1^T, \ldots, v_k v_k^T, A_{k+1}, \ldots, A_m] \end{aligned}$$

Putting it all together

Theorem

Mixed characteristic polynomials are real-rooted.

Proof.

By the first lemma of Borcea and Brändén,

$$p(z_1, \ldots, z_m) = \det\left[xI + \sum_{i=1}^{m} z_i A_i\right]$$

is real stable and so by the second lemma of Borcea and Brändén,

$$\prod_{i=1}^{m}\left(1 - \frac{\partial}{\partial z_i}\right)\det\left[xI + \sum_{i=1}^{m} z_i A_i\right]$$

is real stable.

Putting it all together, cont.

Since real stability is preserved under substitution by reals, (setting $z_1 = \cdots = z_m = 0$), we have

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

is univariate and real stable (and therefore real-rooted). $\square$

Putting it all together, cont.

Since real stability is preserved under substitution by reals, (setting $z_1 = \cdots = z_m = 0$), we have

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

is univariate and real stable (and therefore real-rooted). $\square$

Corollary

Our polynomials form an interlacing family.

Intermission

A quick review:

Intermission

A quick review:

1. We defined *interlacing families* and showed that any such family has a polynomial whose largest root is smaller than the largest root of the expected polynomial

Intermission

A quick review:

1. We defined *interlacing families* and showed that any such family has a polynomial whose largest root is smaller than the largest root of the expected polynomial
2. We argued that (for our polynomials) the interlacing condition was implied by the real-rootedness of the partial assignment polynomials (and the recurrence equation)

Intermission

A quick review:

1. We defined *interlacing families* and showed that any such family has a polynomial whose largest root is smaller than the largest root of the expected polynomial
2. We argued that (for our polynomials) the interlacing condition was implied by the real-rootedness of the partial assignment polynomials (and the recurrence equation)
3. We defined *mixed characteristic polynomials* and showed that our partial assignment polynomials belonged to this class.

Intermission

A quick review:

1. We defined *interlacing families* and showed that any such family has a polynomial whose largest root is smaller than the largest root of the expected polynomial
2. We argued that (for our polynomials) the interlacing condition was implied by the real-rootedness of the partial assignment polynomials (and the recurrence equation)
3. We defined *mixed characteristic polynomials* and showed that our partial assignment polynomials belonged to this class.
4. We showed that mixed characteristic polynomials were real-rooted by using Borcea and Brändén's theory of real stable polynomials.

Intermission

A quick review:

1. We defined *interlacing families* and showed that any such family has a polynomial whose largest root is smaller than the largest root of the expected polynomial
2. We argued that (for our polynomials) the interlacing condition was implied by the real-rootedness of the partial assignment polynomials (and the recurrence equation)
3. We defined *mixed characteristic polynomials* and showed that our partial assignment polynomials belonged to this class.
4. We showed that mixed characteristic polynomials were real-rooted by using Borcea and Brändén's theory of real stable polynomials.

So we are left with bounding the largest root of the expected characteristic polynomial.

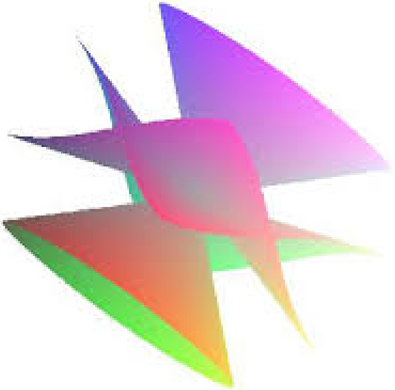
Outline

"Roots" of multivariate polynomials

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

"Roots" of multivariate polynomials

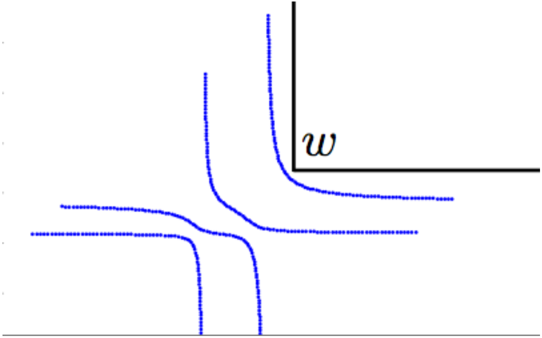
Rather than having roots that are points, multivariate polynomials have *zero surfaces*.



Above the roots

Let $p(x_1, \ldots, x_n)$ be a multivariate real stable polynomial.

We say a point $\vec{w} = (w_1, \ldots, w_n)$ is *above the roots* of p if $p(w_1 + t_1, w_2 + t_2, \ldots, w_n + t_n)$ is nonzero whenever $t_1, \ldots, t_n > 0$.



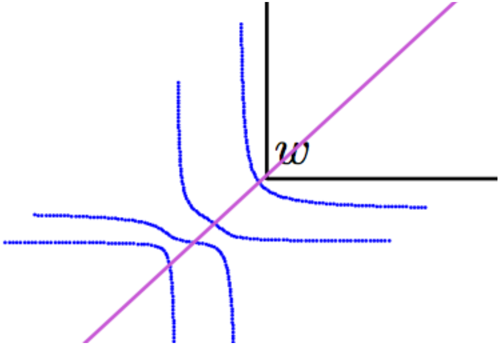
Diagonalization

The *diagonalization* of $p(x_1, \ldots, x_n)$ is the (univariate) polynomial $p(x, x, \ldots, x)$.

Diagonalization

The *diagonalization* of $p(x_1, \ldots, x_n)$ is the (univariate) polynomial $p(x, x, \ldots, x)$.

If $t\mathbb{1}$ is above the roots of p, then t is an upper bound on largest root of its diagonalization.



Shift

Fortunately, we can transform our target polynomial into a diagonalization:

Lemma

In the case that $\sum_i A_i = I$*, we have*

$$\mu[A_1, \ldots, A_m](x) = \prod_{i=1}^{m} \left(1 - \frac{\partial}{\partial z_i}\right) \det \left[xI + \sum_{i=1}^{m} z_i A_i \right] \Bigg|_{z_1 = \cdots = z_m = 0}$$
$$= \prod_{i=1}^{m} \left(1 - \frac{\partial}{\partial y_i}\right) \det \left[\sum_{i=1}^{m} y_i A_i \right] \Bigg|_{y_1 = \cdots = y_m = x}$$

Proof.

Substitute $y_i = z_i + x$. □

The new framework

Recall we are interested in bounding the roots of $\mathbb{E}\left[\chi_{\widehat{V}}(x)\right]$ in the case that

$$\mathbb{E}\left[\sum_{i=1}^{m} \widehat{v}_i \widehat{v}_i^T\right] = I \quad \text{and} \quad \mathbb{E}\left[\|\widehat{v}_i\|^2\right] \leq \epsilon.$$

The new framework

Recall we are interested in bounding the roots of $\mathbb{E}\left[\chi_{\widehat{V}}(x)\right]$ in the case that

$$\mathbb{E}\left[\sum_{i=1}^{m} \widehat{v}_i \widehat{v}_i^T\right] = I \quad \text{and} \quad \mathbb{E}\left[\|\widehat{v}_i\|^2\right] \leq \epsilon.$$

Given what we know about mixed characteristic polynomials, this is equivalent to showing (for some t) that $t\mathbb{1}$ is above the roots of

$$\prod_{i=1}^{m}\left(1 - \frac{\partial}{\partial z_i}\right) \det\left[\sum_{i=1}^{m} z_i A_i\right]$$

whenever $\sum_i A_i = I$ and $Tr[A_i] \leq \epsilon$ for all i.

The new framework

Recall we are interested in bounding the roots of $\mathbb{E}\left[\chi_{\widehat{V}}(x)\right]$ in the case that

$$\mathbb{E}\left[\sum_{i=1}^{m} \widehat{v}_i \widehat{v}_i^T\right] = I \quad \text{and} \quad \mathbb{E}\left[\|\widehat{v}_i\|^2\right] \leq \epsilon.$$

Given what we know about mixed characteristic polynomials, this is equivalent to showing (for some t) that $t\mathbb{1}$ is above the roots of

$$\prod_{i=1}^{m}\left(1 - \frac{\partial}{\partial z_i}\right) \det\left[\sum_{i=1}^{m} z_i A_i\right]$$

whenever $\sum_i A_i = I$ and $Tr[A_i] \leq \epsilon$ for all i.

We will apply the operators one by one and see what happens to the roots.

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.

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

- Blows up whenever a variable x_i gets close to a zero surface of p
- Monotone nonincreasing at any $\vec{w}$ that is above the roots of p
- Convex at any $\vec{w}$ that is above the roots of p
- Measures the amount of "cushion" we have in a given direction

Potential function

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

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

- Blows up whenever a variable x_i gets close to a zero surface of p
- Monotone nonincreasing at any $\vec{w}$ that is above the roots of p
- Convex at any $\vec{w}$ that is above the roots of p
- Measures the amount of "cushion" we have in a given direction

Generalization of potential function from Batson, Spielman, Srivastava (2008)

$$\Phi_p(x) = \frac{\partial}{\partial x} \log p(x) = \frac{p'(x)}{p(x)} = \sum_i \frac{1}{x - r_i}$$

Testing the water lemma

Lemma

If p is real stable, $\vec{w}$ is above the roots of p, and

$$\Phi_p^i(\vec{w}) < 1$$

then $\vec{w}$ is above the roots of $(1 - \partial_{z_i})p$

Testing the water lemma

Lemma

If p is real stable, $\vec{w}$ is above the roots of p, and

$$\Phi_p^i(\vec{w}) < 1$$

then $\vec{w}$ is above the roots of $(1 - \partial_{z_i})p$

Proof.

Since Φ_p^i is nonincreasing, $\Phi_p^i(\vec{w} + \vec{t}) \leq \Phi_p^i(\vec{w}) < 1$ for all $\vec{t} \geq 0$.

Testing the water lemma

Lemma

If p *is real stable,* $\vec{w}$ *is above the roots of* p*, and*

$$\Phi_p^i(\vec{w}) < 1$$

then $\vec{w}$ *is above the roots of* $(1 - \partial_{z_i})p$

Proof.

Since Φ_p^i is nonincreasing, $\Phi_p^i(\vec{w} + \vec{t}) \leq \Phi_p^i(\vec{w}) < 1$ for all $\vec{t} \geq 0$.
Therefore

$$\frac{\partial}{\partial z_i} p(\vec{w} + \vec{t}) < p(\vec{w} + \vec{t})$$

(this is just the definition Φ_p^i).

Testing the water lemma

Lemma

If p *is real stable,* $\vec{w}$ *is above the roots of* p*, and*

$$\Phi_p^i(\vec{w}) < 1$$

then $\vec{w}$ *is above the roots of* $(1 - \partial_{z_i})p$

Proof.

Since Φ_p^i is nonincreasing, $\Phi_p^i(\vec{w} + \vec{t}) \leq \Phi_p^i(\vec{w}) < 1$ for all $\vec{t} \geq 0$.
Therefore

$$\frac{\partial}{\partial z_i} p(\vec{w} + \vec{t}) < p(\vec{w} + \vec{t})$$

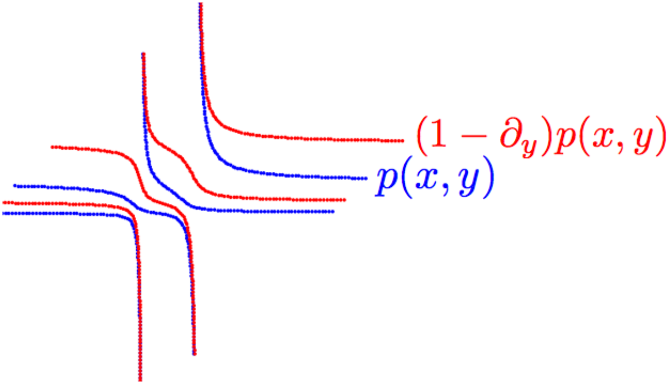
(this is just the definition Φ_p^i). Rearranging, gives

$$\left(1 - \frac{\partial}{\partial z_i}\right) p(\vec{w} + \vec{t}) > 0$$

□

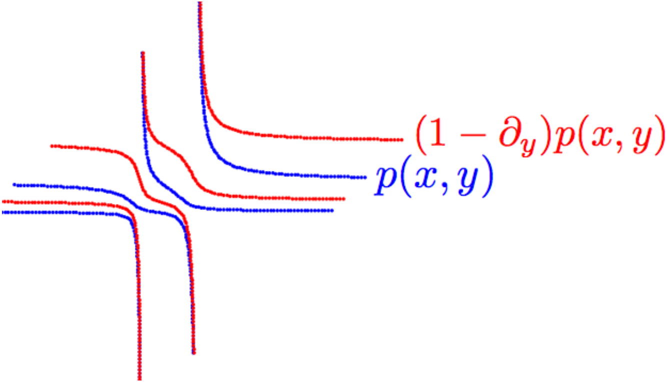
Interpreting the lemma

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



Interpreting the lemma

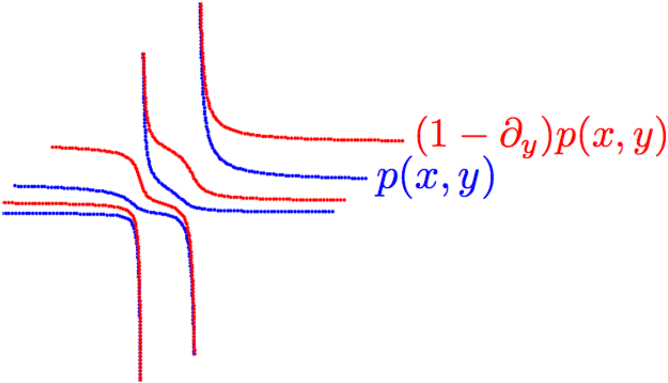
Applying the operator $(1 - \partial_{z_j})$ causes the roots to get closer.



If $\Phi_p^j < 1$ then we are still above the roots after the shift.

Interpreting the lemma

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



If $\Phi_p^j < 1$ then we are still above the roots after the shift.

But we have messed with the potential functions in the other directions (we decreased the cushion)!

Jumping in lemma

Lemma

If p is real stable, $\vec{w}$ is above the roots of p, and

$$\Phi_p^i(\vec{w}) < 1 - \frac{1}{\delta}$$

then $\vec{w}$ is above the roots of $(1 - \partial_{z_i})p$ and

$$\Phi_{p-p_j}^i(\vec{w} + \delta e_j) \leq \Phi_p^i(\vec{w})$$

for all i (where $p_j = \partial p / \partial z_j$).

Jumping in lemma

Lemma

If p is real stable, $\vec{w}$ is above the roots of p, and

$$\Phi_p^i(\vec{w}) < 1 - \frac{1}{\delta}$$

then $\vec{w}$ is above the roots of $(1 - \partial_{z_i})p$ and

$$\Phi_{p-p_j}^i(\vec{w} + \delta e_j) \leq \Phi_p^i(\vec{w})$$

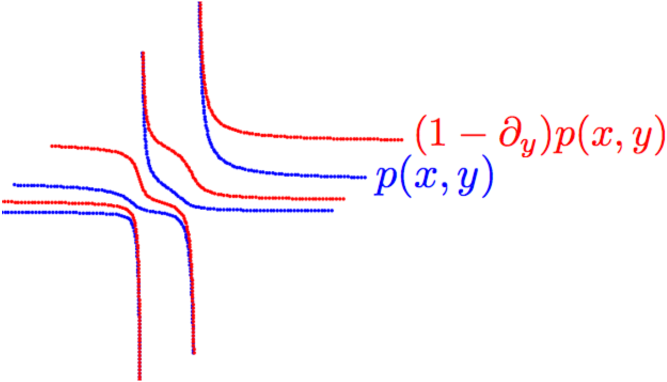
for all i (where $p_j = \partial p / \partial z_j$).

Proof.

Uses convexity mentioned above. □

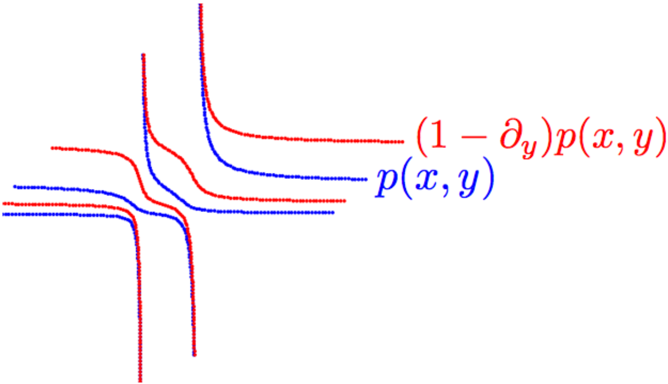
Interpreting the lemma

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



Interpreting the lemma

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



If $\Phi_p^j < 1 - 1/\delta$ then we are still above the roots after the shift *and* if we then move δ in the direction of the shift, we can get back the original cushion we had (in all other directions).

Proof of bound

Theorem

$(1+3\sqrt{\epsilon})\mathbb{1}$ *is above the roots of*

$$\prod_{i=1}^{m}\left(1-\frac{\partial}{\partial z_i}\right)\det\left[\sum_{i=1}^{m} z_i A_i\right]$$

for all $\epsilon < 1/4$.

Proof of bound

Theorem

$(1+3\sqrt{\epsilon})\mathbb{1}$ *is above the roots of*

$$\prod_{i=1}^{m}\left(1-\frac{\partial}{\partial z_i}\right)\det\left[\sum_{i=1}^{m} z_i A_i\right]$$

for all $\epsilon < 1/4$.

Proof.

Start with

$$Q_0(z_1,\ldots,z_m) := \det\left[\sum_{i=1}^{m} z_i A_i\right]$$

so that $t\mathbb{1}$ is above the roots of Q_0 for any $t > 0$.

Proof of bound

Theorem

$(1 + 3\sqrt{\epsilon})\mathbb{1}$ *is above the roots of*

$$\prod_{i=1}^{m} \left(1 - \frac{\partial}{\partial z_i}\right) \det \left[\sum_{i=1}^{m} z_i A_i\right]$$

for all $\epsilon < 1/4$.

Proof.

Start with

$$Q_0(z_1, \ldots, z_m) := \det \left[\sum_{i=1}^{m} z_i A_i\right]$$

so that $t\mathbb{1}$ is above the roots of Q_0 for any $t > 0$.

We will use $t = \sqrt{\epsilon}$ (we'll need the extra cushion).

Set $\vec{w}_0 = t\mathbb{1}$, so that

$$\Phi^{i}_{Q_0}(\vec{w}_0) = Tr\left[\left(\sum_{j=1}^{m} tA_j\right)^{-1} A_i\right] = \frac{Tr[A_i]}{t} \leq \frac{\epsilon}{\sqrt{\epsilon}} = \sqrt{\epsilon}.$$

This satisfies the cushion lemma for any

$$\delta > \frac{1}{1-\sqrt{\epsilon}}$$

so pick $\delta = 1 + 2\sqrt{\epsilon}$ (here we use $\epsilon < 1/4$).

Apply the operator $1 - \partial_{z_1}$ and then move δ in the direction of $\vec{e}_1$. By the lemma, $\vec{w}_1 = \vec{w}_0 + \delta\vec{e}_1$ is above the roots of

$$Q_1 = \left(1 - \frac{\partial}{\partial z_1}\right) Q_0$$

and we still satisfy the cushion lemma with $\delta = 1 + 2\sqrt{\epsilon}$.

Apply the operator $1 - \partial_{z_1}$ and then move δ in the direction of $\vec{e}_1$. By the lemma, $\vec{w}_1 = \vec{w}_0 + \delta\vec{e}_1$ is above the roots of

$$Q_1 = \left(1 - \frac{\partial}{\partial z_1}\right) Q_0$$

and we still satisfy the cushion lemma with $\delta = 1 + 2\sqrt{\epsilon}$.

Do this for $i = 2, \ldots, m$ (using the lemma each time). This shows that

$$\vec{w}_m = \vec{w}_0 + \delta \sum_i \vec{e}_i = (\delta + t)\mathbb{1} = (1 + 3\sqrt{\epsilon})\mathbb{1}$$

is above the roots of

$$Q_m = \prod_{i=1}^{m} \left(1 - \frac{\partial}{\partial z_i}\right) Q_0 = \prod_{i=1}^{m} \left(1 - \frac{\partial}{\partial z_i}\right) \det\left[\sum_i z_i A_i\right]$$

as required. $\square$

Second intermission

A quick review:

Second intermission

A quick review:

1. We defined *interlacing families* and showed that any such family has a polynomial p_σ such that the largest root of p_σ is smaller than the largest root of the expected polynomial ($p_\emptyset$)

Second intermission

A quick review:

1. We defined *interlacing families* and showed that any such family has a polynomial p_σ such that the largest root of p_σ is smaller than the largest root of the expected polynomial ($p_\emptyset$)
2. We showed that our polynomials formed an interlacing family by showing they were mixed characteristic polynomials

Second intermission

A quick review:

1. We defined *interlacing families* and showed that any such family has a polynomial p_σ such that the largest root of p_σ is smaller than the largest root of the expected polynomial ($p_\emptyset$)
2. We showed that our polynomials formed an interlacing family by showing they were mixed characteristic polynomials
3. We defined a multivariate barrier function to help us understand the evolution of the zero surfaces of multivariate polynomials

Second intermission

A quick review:

1. We defined *interlacing families* and showed that any such family has a polynomial p_σ such that the largest root of p_σ is smaller than the largest root of the expected polynomial ($p_\emptyset$)
2. We showed that our polynomials formed an interlacing family by showing they were mixed characteristic polynomials
3. We defined a multivariate barrier function to help us understand the evolution of the zero surfaces of multivariate polynomials
4. We used this to show that (for our polynomials) the largest root of $p_\emptyset$ was at most $1 + 3\sqrt{\epsilon}$.

Outline

Our theorem

We have proved our main technical theorem:

Theorem

Let $0 < \epsilon < 1/4$ *and* $\widehat{v}_1, \ldots \widehat{v}_m$ *be independent random vectors such that*

$$\sum_{i=1}^{m} \mathbb{E}\left[\widehat{v}_i \widehat{v}_i^T\right] = I$$

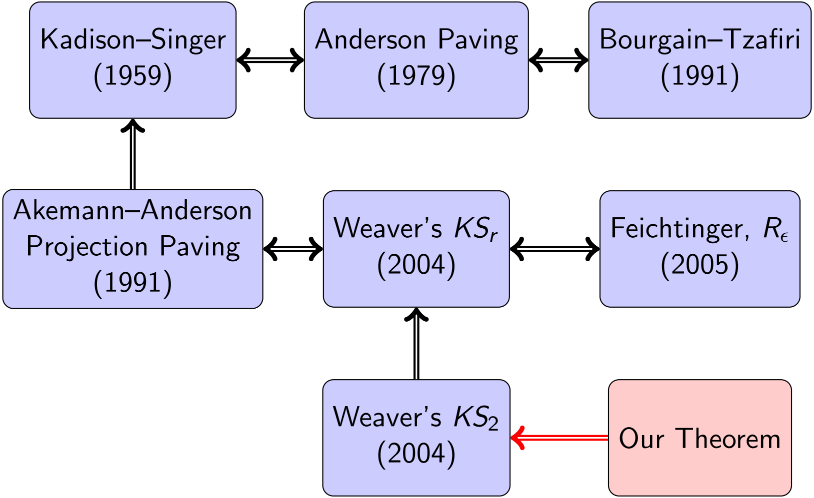
and

$$\mathbb{E}\left[\|\widehat{v}_i\|^2\right] \leq \epsilon$$

for all i*. Then there exists an assignment* $\widehat{v}_i = v_i$ *such that*

$$\left\| \sum_{i=1}^{m} v_i v_i^T \right\| \leq 1 + 3\sqrt{\epsilon}.$$

Claim:



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

Recall what KS_2 says:

Conjecture (KS_2)

There exist universal constants $\eta \geq 2$ and $\theta > 0$ such that the following holds: if $w_1, \ldots, w_m \in \mathbb{C}^d$ satisfy $\|w_i\| \leq 1$ for all i and

$$\sum_i |\langle u, w_i \rangle|^2 = \eta$$

for all unit vectors $u \in \mathbb{C}^d$. Then there exists a partition of the vectors into two parts S_0, S_1 so that

$$\sum_{i \in S_j} |\langle u, w_i \rangle|^2 \leq \eta - \theta$$

for all unit vectors $u \in \mathbb{C}^d$ and each $j \in \{0, 1\}$.

Recall what KS_2 says:

Conjecture (KS_2)

There exist universal constants $\eta \geq 2$ and $\theta > 0$ such that the following holds: if $w_1, \ldots, w_m \in \mathbb{C}^d$ satisfy $\|w_i\| \leq 1$ for all i and

$$\sum_i |\langle u, w_i \rangle|^2 = \eta$$

for all unit vectors $u \in \mathbb{C}^d$. Then there exists a partition of the vectors into two parts S_0, S_1 so that

$$\sum_{i \in S_j} |\langle u, w_i \rangle|^2 \leq \eta - \theta$$

for all unit vectors $u \in \mathbb{C}^d$ and each $j \in \{0, 1\}$.

Again, we prove the real case (though the complex case is identical).

Proof.

Given the w_i, let $\widehat{v}_i$ be the random vector (in $\mathbb{R}^{2d}$!) taking values in

$$\left\{ \sqrt{\frac{2}{\eta}} \begin{pmatrix} w_i \\ 0^d \end{pmatrix}, \sqrt{\frac{2}{\eta}} \begin{pmatrix} 0^d \\ w_i \end{pmatrix} \right\}$$

each with probability $1/2$ and set $\epsilon = 2/\eta$.

Proof.

Given the w_i, let $\widehat{v}_i$ be the random vector (in $\mathbb{R}^{2d}$!) taking values in

$$\left\{ \sqrt{\frac{2}{\eta}} \begin{pmatrix} w_i \\ 0^d \end{pmatrix}, \sqrt{\frac{2}{\eta}} \begin{pmatrix} 0^d \\ w_i \end{pmatrix} \right\}$$

each with probability $1/2$ and set $\epsilon = 2/\eta$.

Then (this is just a rescaling to fit our main theorem)

$$\sum_{i=1}^{m} \mathbb{E}\left[\widehat{v}_i \widehat{v}_i^T\right] = I \quad \text{and} \quad \mathbb{E}\left[\|\widehat{v}_i\|^2\right] \leq \epsilon$$

for all i, so let σ be the assignment guaranteed by our main theorem.

For the given σ, let

$$M_0 = \frac{2}{\eta} \sum_{i:\sigma(i)=0} w_i w_i^T \quad \text{and} \quad M_1 = \frac{2}{\eta} \sum_{i:\sigma(i)=1} w_i w_i^T$$

For the given σ, let

$$M_0 = \frac{2}{\eta} \sum_{i:\sigma(i)=0} w_i w_i^T \quad \text{and} \quad M_1 = \frac{2}{\eta} \sum_{i:\sigma(i)=1} w_i w_i^T$$

Then by the theorem, the matrix

$$\begin{pmatrix} M_0 & 0_{d\times d} \\ 0_{d\times d} & M_1 \end{pmatrix} = \begin{pmatrix} M_0 & 0_{d\times d} \\ 0_{d\times d} & I - M_0 \end{pmatrix} = \begin{pmatrix} I - M_1 & 0_{d\times d} \\ 0_{d\times d} & M_1 \end{pmatrix}$$

and so M_σ has largest eigenvalue at most $1 + 3\sqrt{2/\eta}$.

For the given σ, let

$$M_0 = \frac{2}{\eta} \sum_{i:\sigma(i)=0} w_i w_i^T \quad \text{and} \quad M_1 = \frac{2}{\eta} \sum_{i:\sigma(i)=1} w_i w_i^T$$

Then by the theorem, the matrix

$$\begin{pmatrix} M_0 & 0_{d\times d} \\ 0_{d\times d} & M_1 \end{pmatrix} = \begin{pmatrix} M_0 & 0_{d\times d} \\ 0_{d\times d} & I - M_0 \end{pmatrix} = \begin{pmatrix} I - M_1 & 0_{d\times d} \\ 0_{d\times d} & M_1 \end{pmatrix}$$

and so M_σ has largest eigenvalue at most $1 + 3\sqrt{2/\eta}$.

Set

$$S_0 = \{w_i \mid \sigma_i = 0\} \quad \text{and} \quad S_1 = \{w_i \mid \sigma_i = 1\}$$

Then for all $j \in \{0,1\}$ and $u \in \mathbb{C}^d$

$$\frac{2}{\eta} \sum_{i \in S_j} |\langle u, w_i \rangle|^2 \leq 1 + 3\sqrt{\frac{2}{\eta}}$$

Then for all $j \in \{0, 1\}$ and $u \in \mathbb{C}^d$

$$\frac{2}{\eta} \sum_{i \in S_j} |\langle u, w_i \rangle|^2 \leq 1 + 3\sqrt{\frac{2}{\eta}}$$

Setting $\eta = 32$ gives

$$\sum_{i \in S_j} |\langle u, w_i \rangle|^2 \leq 28$$

proving the theorem for $\eta = 32$ and $\theta = 4$. $\square$

Then for all $j \in \{0,1\}$ and $u \in \mathbb{C}^d$

$$\frac{2}{\eta} \sum_{i \in S_j} |\langle u, w_i \rangle|^2 \leq 1 + 3\sqrt{\frac{2}{\eta}}$$

Setting $\eta = 32$ gives

$$\sum_{i \in S_j} |\langle u, w_i \rangle|^2 \leq 28$$

proving the theorem for $\eta = 32$ and $\theta = 4$. $\square$

Using the (stronger) original theorem, we can get $\eta = 18$ and $\theta = 2$.

Then for all $j \in \{0, 1\}$ and $u \in \mathbb{C}^d$

$$\frac{2}{\eta} \sum_{i \in S_j} |\langle u, w_i \rangle|^2 \leq 1 + 3\sqrt{\frac{2}{\eta}}$$

Setting $\eta = 32$ gives

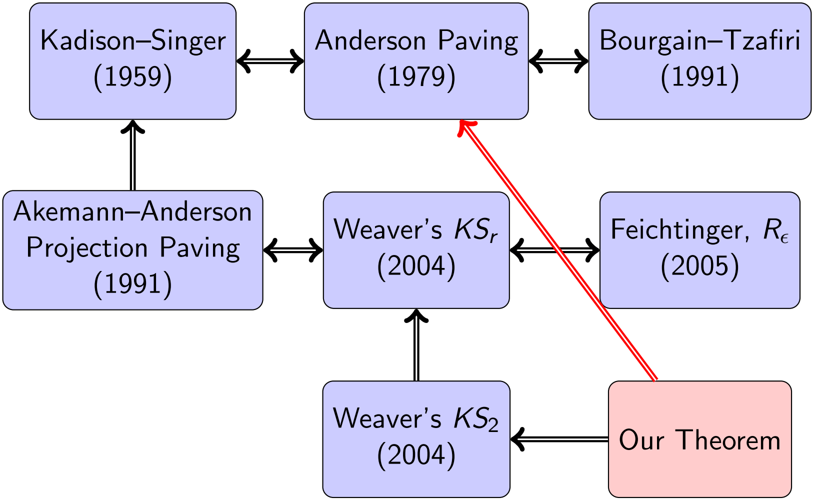
$$\sum_{i \in S_j} |\langle u, w_i \rangle|^2 \leq 28$$

proving the theorem for $\eta = 32$ and $\theta = 4$. □

Using the (stronger) original theorem, we can get $\eta = 18$ and $\theta = 2$.

Casazza showed $\eta = 2$ is not possible (optimal answer lies somewhere in between).

Or if you prefer paving



Paving

Conjecture (Casazza)

For all $\epsilon > 0$ *and even integers* $N > 0$*, there exists* $r = r(N, \epsilon)$ *such that for any* $d > 0$ *and any vectors* $v_1, \ldots, v_m \in \mathbb{C}^d$ *satisfying*

$$\sum_{j=1}^{m} v_j {v_j}^* = I_d \quad \textit{and} \quad \|v_j\|^2 \leq \frac{1}{N}$$

for all j*, there exists a partition* $\{A_1, \ldots, A_r\}$ *of* $[m]$ *such that for all* $i \in [r]$

$$\sum_{j \in A_i} v_j {v_j}^* \leq \frac{1 + \epsilon}{N}.$$

Paving

Conjecture (Casazza)

For all $\epsilon > 0$ *and even integers* $N > 0$, *there exists* $r = r(N, \epsilon)$ *such that for any* $d > 0$ *and any vectors* $v_1, \ldots, v_m \in \mathbb{C}^d$ *satisfying*

$$\sum_{j=1}^{m} v_j {v_j}^* = I_d \quad \text{and} \quad \|v_j\|^2 \leq \frac{1}{N}$$

for all j, *there exists a partition* $\{A_1, \ldots, A_r\}$ *of* $[m]$ *such that for all* $i \in [r]$

$$\sum_{j \in A_i} v_j {v_j}^* \leq \frac{1+\epsilon}{N}.$$

Casazza, et al. (2007) and Harvey (2013) showed equivalence to Anderson's paving conjecture (including evolution of constants).

Paving

Conjecture (Casazza)

For all $\epsilon > 0$ *and even integers* $N > 0$, *there exists* $r = r(N, \epsilon)$ *such that for any* $d > 0$ *and any vectors* $v_1, \ldots, v_m \in \mathbb{C}^d$ *satisfying*

$$\sum_{j=1}^{m} v_j {v_j}^* = I_d \quad \text{and} \quad \|v_j\|^2 \leq \frac{1}{N}$$

for all j, *there exists a partition* $\{A_1, \ldots, A_r\}$ *of* $[m]$ *such that for all* $i \in [r]$

$$\sum_{j \in A_i} v_j {v_j}^* \leq \frac{1+\epsilon}{N}.$$

Casazza, et al. (2007) and Harvey (2013) showed equivalence to Anderson's paving conjecture (including evolution of constants).

Argument similar to proof of KS_2 shows this holds for $r \geq 6N/\epsilon^2$.

Outline

Polynomial problems

Find more interlacing families and use them to solve problems!

Both results here used characteristic polynomials, but any real rooted polynomials will do (for example, Gurvits has interesting results using matching polynomials).

Polynomial problems

Find more interlacing families and use them to solve problems!

Both results here used characteristic polynomials, but any real rooted polynomials will do (for example, Gurvits has interesting results using matching polynomials).

Find quantitative extensions of Borcea and Brändén theory.

Polynomial problems

Find more interlacing families and use them to solve problems!

Both results here used characteristic polynomials, but any real rooted polynomials will do (for example, Gurvits has interesting results using matching polynomials).

Find quantitative extensions of Borcea and Brändén theory. What happens to roots of polynomials under different transformations?

Polynomial problems

Find more interlacing families and use them to solve problems!

Both results here used characteristic polynomials, but any real rooted polynomials will do (for example, Gurvits has interesting results using matching polynomials).

Find quantitative extensions of Borcea and Brändén theory. What happens to roots of polynomials under different transformations?

Find a computationally efficient version of the method of interlacing polynomials (or show it is not possible).

Polynomial problems

Find more interlacing families and use them to solve problems!

Both results here used characteristic polynomials, but any real rooted polynomials will do (for example, Gurvits has interesting results using matching polynomials).

Find quantitative extensions of Borcea and Brändén theory. What happens to roots of polynomials under different transformations?

Find a computationally efficient version of the method of interlacing polynomials (or show it is not possible).

Can any of the real-rootedness conditions be relaxed?

In general

Would direct application of the main theorem give better paving results (higher up the chain of implications)?

In general

Would direct application of the main theorem give better paving results (higher up the chain of implications)?

What is the worst (in terms of largest root) $\mu[A_1, \ldots, A_m](x)$?
We conjecture it is when all A_i are the same (and multiples of the identity).
This would improve all of the constants in this talk.

In general

Would direct application of the main theorem give better paving results (higher up the chain of implications)?

What is the worst (in terms of largest root) $\mu[A_1, \ldots, A_m](x)$? We conjecture it is when all A_i are the same (and multiples of the identity).
This would improve all of the constants in this talk.

"The general feeling in the community is that the original question (and therefore all equivalent forms) have a negative solution" (Casazza–Kutinyiok, 2013).

What are the implications of a *positive* solution?

Thanks

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

Thanks

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

And thank you for your attention!