The Method of Interlacing Polynomials

H. T. Chau
Easter 2021

Part III Essay
under the supervision of
Prof W. T. Gowers

Contents

0 Introduction

In this essay, we discuss the method of interlacing families introduced by Marcus, Spielman, and Srivastava in [MSS15a] and [MSS15b], as well as some relevant background. Interlacing of polynomials was studied previously in analysis, such as in [Fel80] and [Ded92], but it was only much later that Marcus, Spielman, and Srivastava found a new variant of the probabilistic method that makes use of the interlacing property of polynomials (Section 1), and applied it to prove the existence of certain **Ramanujan graphs** in spectral graph theory (Section 2) and the **Anderson paving conjecture** that is known to imply a positive answer to the Kadison–Singer problem about C^*-algebras (Section 3).

The interlacing property is required in their probabilistic method because of the following: In the usual first moment method, we may show a bound $\mathbb{E}(X) \leq M$ for some random variable X, and use $\mathbb{P}(X \leq \mathbb{E}(X)) > 0$ to conclude that some instance with $X \leq M$ exists. However, for some combinatorial problems, we need to consider roots of polynomials[1], and for a random polynomial $f(x)$ of degree d with positive leading coefficient, a bound on the roots of $\mathbb{E}(f(x))$ does not tell us much about the roots of $f(x)$ at first sight, because the map λ_1 sending a polynomial f to its largest real root $\lambda_1(f)$ is non-linear[2].

Fortunately, as we shall see in Lemma 1.5, if the possible outcomes of f have a common interlacing, then with non-zero probability we have $\lambda_1(f) \leq \lambda_1(\mathbb{E}(f))$. Furthermore, by an inductive argument as in Lemma 1.11, the interlacing condition can be slightly relaxed. We just require f to take values from the leaves of a tree, where each internal node is the expectation of its children, and all siblings have a common interlacing. Such a tree is called an **interlacing family** (Definition 1.9).

It turns out a particular class of interlacing families (Lemma 1.23) is useful in both the Ramanujan graph and the Anderson paving conjecture applications. They arise from **mixed characteristic polynomials**, which is of the form $\mathbb{E}(\chi(\sum_i A_i))$ for some random rank one matrices A_i, and $\chi(X)$ denotes the characteristic polynomial of X.

After these lemmas, we shall be ready to apply the method of interlacing families. For each of the two problems we consider, we need to identify an interlacing family f_σ, and then prove a bound $\lambda_1(\mathbb{E}_\sigma(f_\sigma)) \leq M$, and this would immediately give the existence of some σ such that $\lambda_1(f_\sigma) \leq M$, which translates to a combinatorial property that we are looking for.

For Ramanujan graphs, the interlacing family comes from the characteristic polynomials of the **signed adjacency matrices** (Lemma 2.9), and in this case the expectation is a well-known polynomial called the **matching polynomial** of the graph, which can be defined recursively, and an inductive argument in [HL72] gives the desired bound on the roots of that polynomial (Theorem 2.12).

For the Anderson paving conjecture, the interlacing family comes from translating a problem about partition of vectors (Aim 3.14) into one about random vectors (Aim 3.15), but to bound the roots of the expectation we need a **barrier argument**. This requires convexity results (Lemma 3.20) about the log-derivative $\Phi_f^i = \frac{\partial_{z_i} f}{f}$ of a multivariate polynomial f, in order to estimate the effect on the zero set of the polynomial f when one apply the operator $(1 - \partial_{z_i})$ to it (Lemma 3.21).

[1] Rota said in an interview that "... all sorts of problems of combinatorics can be viewed as problems of location of the zeros of certain polynomials and in giving these zeros a combinatorial interpretation." The interview is avaibable at `https://fas.org/sgp/othergov/doe/lanl/pubs/00326965.pdf`

[2] Tao13, after Proposition 1.

1 Roots of Polynomials

1.1 Interlacing Polynomials

Definition 1.1 (Roots)**.** A polynomial in one variable is **real-rooted** if all of its roots are real. If $f(x) \in \mathbb{R}[x]$ is a real-rooted polynomial of degree $d \geq 1$, then we write $\lambda_i(f)$ for its i-th largest root ($1 \leq i \leq d$), i.e.

$$\lambda_d(f) \leq \lambda_{d-1}(f) \leq \cdots \leq \lambda_2(f) \leq \lambda_1(f)$$

are the d roots of f.

For convenience, we shall write $\lambda_0(f) = +\infty$ and $\lambda_{d+1}(f) = -\infty$. If the leading coefficient of f is positive, then for $0 \leq j \leq d$ and $x \in (\lambda_{j+1}(f), \lambda_j(f))$, $\operatorname{sgn} f(x) = (-1)^j$, where

$$\operatorname{sgn} a = \begin{cases} +1, & \text{if } a > 0, \\ 0, & \text{if } a = 0, \\ -1, & \text{if } a < 0. \end{cases}$$

Definition 1.2 (Interlacing)**.** [3] Let $f(x), g(x)$ be real-rooted polynomials in $\mathbb{R}[x]$ of degree $d, d-1$ respectively. We say $g(x)$ **interlaces** $f(x)$ if their roots alternate, i.e.

$$\lambda_d(f) \leq \lambda_{d-1}(g) \leq \lambda_{d-1}(f) \leq \cdots \leq \lambda_2(g) \leq \lambda_2(f) \leq \lambda_1(g) \leq \lambda_1(f).$$

If $f_1(x), \cdots, f_n(x) \in \mathbb{R}[x]$ are real-rooted and have degree d, and there is some $g(x)$ of degree $d-1$ that interlaces all $f_i(x)$, then we say $f_1(x), \cdots, f_n(x)$ have a **common interlacing**.

So having a common interlacing is equivalent to having

$$\max_i \lambda_{j+1}(f_i) \leq \min_i \lambda_j(f_i)$$

for $1 \leq j \leq d-1$ (so that we can put $\lambda_j(g)$ between these two numbers).

Example 1.3. If f is a real-rooted polynomial of degree d, then by Rolle's theorem, its derivative f' is a real-rooted polynomial that interlaces f.

In the rest of this subsection, we shall show the useful equivalence between having a common interlacing and having all convex combinations real-rooted.

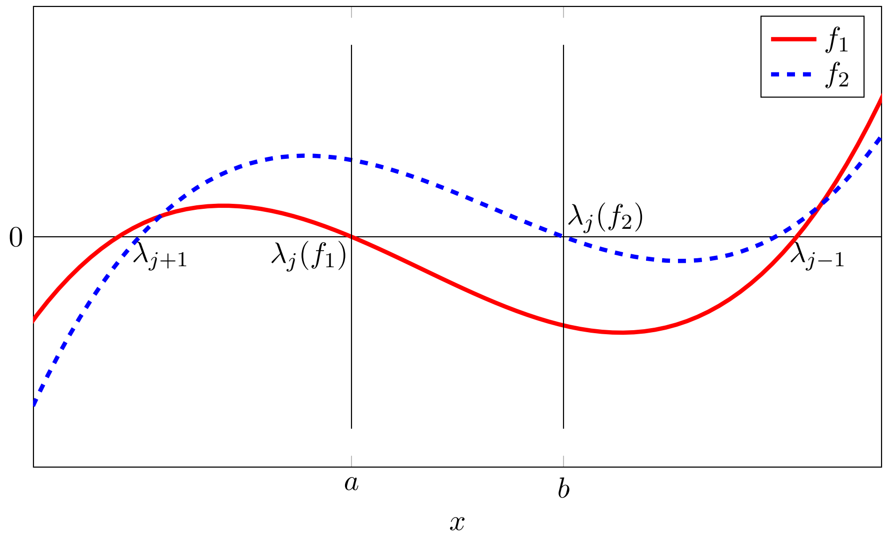
Lemma 1.4 (Root property from common interlacing)**.** [4] If $f_1(x), f_2(x) \in \mathbb{R}[x]$ have degree d, have positive leading coefficients, and have a common interlacing, then any convex combination $h(x) = (1-\mu)f_1(x) + \mu f_2(x)$ ($0 \leq \mu \leq 1$) is real-rooted, and we have

$$\min\{\lambda_j(f_1), \lambda_j(f_2)\} \leq \lambda_j(h) \leq \max\{\lambda_j(f_1), \lambda_j(f_2)\}$$

for $1 \leq j \leq d$.

Proof. [5] It suffices to consider strict convex combinations where $0 < \mu < 1$. First consider the generic case where all $2d$ roots of f_1 and f_2 are distinct.

[3] MSS15a, Definition 4.1.
[4] Fel80, Theorem 2'.
[5] Ded92, pp. 2.3–2.4.



Fix $j \in [d]$. Without loss of generality, let $a = \lambda_j(f_1) < \lambda_j(f_2) = b$. f_1, f_2 have a common interlacing, so $\lambda_j(f_1) < b < \lambda_{j-1}(f_1)$, so $\operatorname{sgn} f_1(b) = (-1)^{j-1}$. Similarly $\operatorname{sgn} f_2(a) = (-1)^j$. But then

$$\begin{aligned} h(a) &= \mu f_2(a) \text{ has sign } (-1)^j, \\ h(b) &= (1-\mu) f_1(b) \text{ has sign } (-1)^{j-1}, \end{aligned}$$

so h has a real root between $\lambda_j(f_1)$ and $\lambda_j(f_2)$. The same holds for $1 \le j \le d$, so all roots of h are real and its j-th root lies between $\lambda_j(f_1)$ and $\lambda_j(f_2)$.

As a corollary of Rouché's theorem, the (complex) roots of a polynomial of degree d vary continuously when the coefficients vary, as long as the leading coefficient never vanishes[6]. This allows us to perturb the polynomials to reduce to the generic case. Consider $f_{i,\varepsilon}$ with same leading coefficients as f_i but with $\lambda_j(f_{i,\varepsilon}) = \lambda_j(f_i) - (2j+i)\varepsilon$. Only for finitely many ε can some two of the $2d$ roots of $f_{1,\varepsilon}, f_{2,\varepsilon}$ be equal.

Since $\max\{\lambda_j(f_1), \lambda_j(f_2)\} \le \min\{\lambda_{j-1}(f_1), \lambda_{j-1}(f_2)\}$ by common interlacing, for $\varepsilon > 0$ we have

$$\begin{aligned} \max\{\lambda_j(f_{1,\varepsilon}), \lambda_j(f_{2,\varepsilon})\} &\le \max\{\lambda_j(f_1), \lambda_j(f_2)\} - (2j+1)\varepsilon \\ &< \min\{\lambda_{j-1}(f_1), \lambda_{j-1}(f_2)\} - (2(j-1)+2)\varepsilon \\ &\le \min\{\lambda_{j-1}(f_{1,\varepsilon}), \lambda_{j-1}(f_{2,\varepsilon})\}. \end{aligned}$$

This says for all sufficiently small $\varepsilon > 0$, $f_{1,\varepsilon}, f_{2,\varepsilon}$ have a common interlacing. They are also in the generic case above, so $h_\varepsilon = \mu f_{1,\varepsilon} + (1-\mu) f_{2,\varepsilon}$ is real-rooted with

$$\min\{\lambda_j(f_{1,\varepsilon}), \lambda_j(f_{2,\varepsilon})\} \le \lambda_j(h_\varepsilon) \le \max\{\lambda_j(f_{1,\varepsilon}), \lambda_j(f_{2,\varepsilon})\}.$$

Taking $\varepsilon \to 0$ gives the desired result. □

In Lemma 1.4, if g interlaces both f_1 and f_2, then it interlaces h as well. This gives a natural extension to n polynomials.

Lemma 1.5 (Root property from common interlacing)**.** [7] If $f_1(x), \cdots, f_n(x) \in \mathbb{R}[x]$ have degree d, have positive leading coefficients, and have a common interlacing, then any convex combination $h(x) = \sum_i \mu_i f_i(x)$ ($\mu_i \ge 0, \sum_i \mu_i = 1$) is real-rooted, and moreover

$$\min_i \lambda_j(f_i) \le \lambda_j(h) \le \max_i \lambda_j(f_i)$$

for $1 \le j \le d$.

[6] Ale, Corollary 3.3.
[7] CS07, Theorem 3.6.

Proof. Induct on n. The $n = 2$ case is Lemma 1.4.

Let g interlace all f_i. WLOG μ_1, μ_2 are not simultaneously zero. Then $f_{1,2} = \frac{\mu_1}{\mu_1+\mu_2} f_1 + \frac{\mu_2}{\mu_1+\mu_2} f_2$ is a convex combination of f_1 and f_2, so by Lemma 1.4, $f_{1,2}$ is real-rooted and g interlaces $f_{1,2}$, and $\min_{i\le 2} \lambda_j(f_i) \le \lambda_j(f_{1,2}) \le \max_{i\le 2} \lambda_j(f_i)$.

Now

$$h = (\mu_1 + \mu_2) f_{1,2} + \sum_{i\ge 3} \mu_i f_i$$

is a convex combination of $f_{1,2}, f_3, \cdots, f_n$, so the result follows from the induction hypothesis. □

A converse to Lemma 1.4 holds (Lemma 1.7), but in proving that, it is more convenient to consider the ratio

$$g_\mu(x) = \frac{\mu f_1(x) + (1-\mu) f_2(x)}{\mu f_2(x)} = \frac{1-\mu}{\mu} + \frac{f_1(x)}{f_2(x)}$$

instead of the convex combination $h_\mu = \mu f_1 + (1-\mu) f_2$ itself. g_μ and h_μ have the same zeros except when h_μ and f_2 have common factors that cancel in the fraction.

Lemma 1.6. [8] If $f_1(x), f_2(x) \in \mathbb{R}[x]$ have positive leading coefficients, same degree d, and are coprime, and all convex combinations $h_\mu = \mu f_1 + (1-\mu) f_2$ $(0 \le \mu \le 1)$ are real-rooted, then for $0 < \mu < 1$, all roots of $g_\mu = \frac{1-\mu}{\mu} + \frac{f_1}{f_2}$ are simple, and hence all roots of h_μ are simple.

Proof. The roots of g_μ are precisely the roots of h_μ, which are real and vary continuously in μ. Suppose x_0 is a root of g_{μ_0} of order $l \ge 1$, then for μ close to μ_0, g_μ must have l real roots (counting multiplicity) near x_0.

Near x_0, we have $g_{\mu_0}(x) = C(x-x_0)^l + \mathcal{O}((x-x_0)^{l+1})$ for some constant $C \ne 0$. For μ close to μ_0, the equation $g_\mu(x) = 0$ can be rewritten as

$$\frac{1-\mu_0}{\mu_0} - \frac{1-\mu}{\mu} = C(x-x_0)^l + \mathcal{O}((x-x_0)^{l+1}).$$

If we pick μ such that the left hand side has different sign as C, then there is no real root near x_0 for even l, and at most 1 real root for odd l. But we already know g_μ has l roots near x_0, so $l = 1$ as required.

For the last claim, since $f_1(x)$ and $f_2(x)$ are coprime, $h_\mu(x)$ and $f_2(x)$ are also coprime for $0 < \mu < 1$, so h_μ has the same roots as g_μ. □

Lemma 1.7 (Convex combination criterion)**.** [9] Suppose $f_1(x), f_2(x) \in \mathbb{R}[x]$ have positive leading coefficients and same degree d. Then they have a common interlacing if and only if all convex combinations $\sum \mu_i f_i$ are real-rooted.

Proof. [10] "$\Longrightarrow$" is Lemma 1.4.

For "$\Longleftarrow$", first we focus on the generic case where f_1, f_2 are coprime and all roots are simple, so that there are $2d$ distinct roots $\lambda_j(f_i)$ $(1 \le i \le 2, 1 \le j \le d)$. Suppose f_1 and f_2 do not have a common interlacing, then there is a largest $j \in [d-1]$ such that

$$\max(\lambda_{j+1}(f_1), \lambda_{j+1}(f_2)) > \min(\lambda_j(f_1), \lambda_j(f_2)).$$

[8] Ded92, Lemma 2.2.

[9] Ded92, Theorem 2.1.

[10] Ded92, Proofs 2.5–2.7.

By maximality of j and since the roots are distinct,

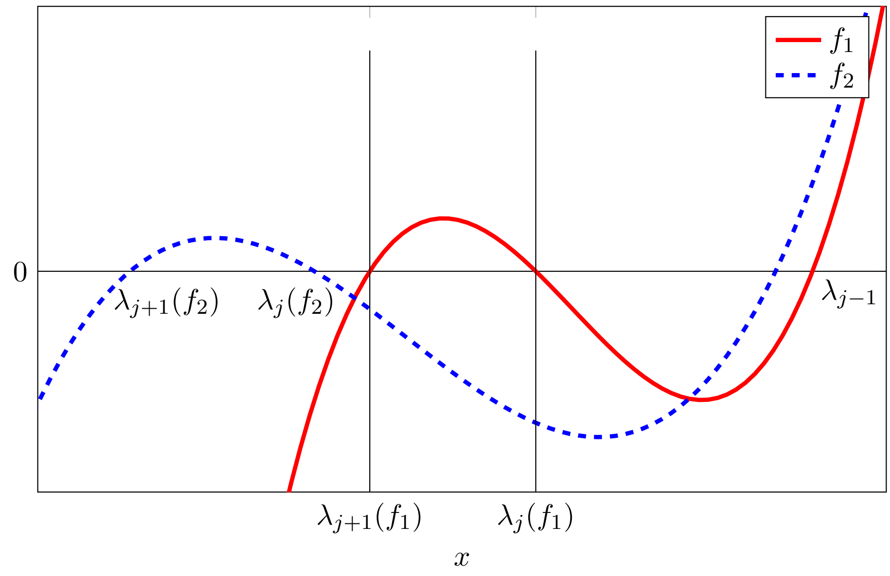
$$\max(\lambda_j(f_1), \lambda_j(f_2)) < \min(\lambda_{j-1}(f_1), \lambda_{j-1}(f_2)).$$

($j-1$ might be zero in which case the right hand side is $+\infty$.)

Without loss of generaltiy, $\lambda_j(f_1) > \lambda_j(f_2)$, then the inequalities above force

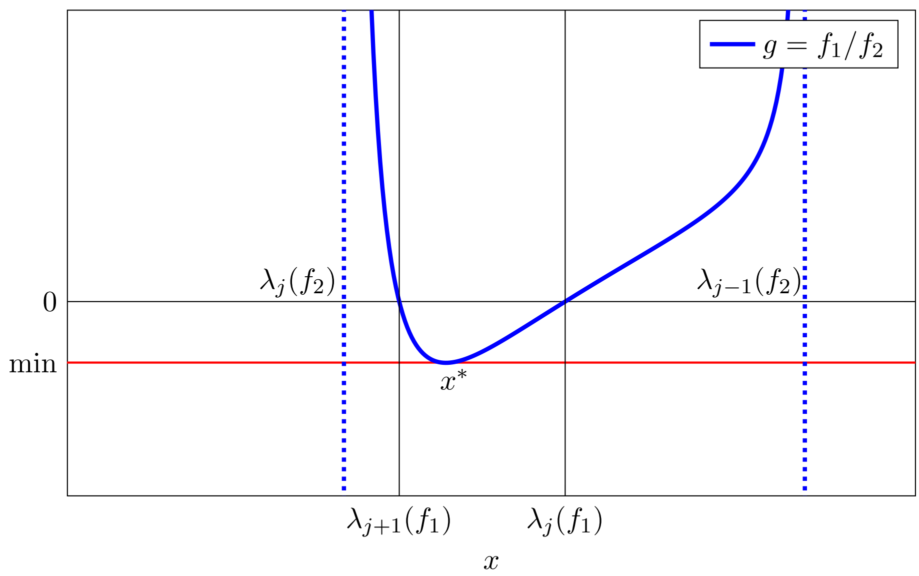
$$\lambda_{j+1}(f_2) < \lambda_j(f_2) < \lambda_{j+1}(f_1) < \lambda_j(f_1) < \min(\lambda_{j-1}(f_1), \lambda_{j-1}(f_2))$$

as shown in the diagram.



Consider the interval $I = (\lambda_{j+1}(f_1), \lambda_j(f_1))$. For any $x \in I$, $\operatorname{sgn} f_1(x) = (-1)^j$, but $I \subseteq (\lambda_j(f_2), \lambda_{j-1}(f_2))$, so $\operatorname{sgn} f_2(x) = (-1)^{j-1}$, so $\frac{f_1(x)}{f_2(x)} < 0$.

Now $\frac{f_1(x)}{f_2(x)}$ is 0 at the end points of I, and $\frac{f_1(x)}{f_2(x)} < 0$ in I, so it attains a minimum at some $x^* \in I$.



Pick $\mu \in (0, 1)$ such that

$$\frac{1-\mu}{\mu} + \frac{f_1(x^*)}{f_2(x^*)} = 0.$$

Then x^* is a zero and a minimum to g_μ defined by

$$g_\mu(x) = \frac{1-\mu}{\mu} + \frac{f_1(x)}{f_2(x)},$$

so it has multiplicity at least 2, contradicting Lemma 1.6 that the roots of g_μ are simple.

Now we have done the generic case. The next case is when $f_1(x)$ and $f_2(x)$ are coprime but each might have repeated roots. In this case, consider $f_{1,\varepsilon} = (1-\varepsilon)f_1 + \varepsilon f_2$ and $f_{2,\varepsilon} = \varepsilon f_1 + (1-\varepsilon)f_2$ for small $\varepsilon > 0$. By Lemma 1.6, both have simple roots which are real by assumption. Moreover, their roots are close to those of f_1 and f_2 respectively, so they remain coprime, so they have a common interlacing by the generic case, i.e. for all $j \in [d-1]$,

$$\max\{\lambda_{j+1}(f_{1,\varepsilon}), \lambda_{j+1}(f_{2,\varepsilon})\} \leq \min\{\lambda_j(f_{1,\varepsilon}), \lambda_j(f_{2,\varepsilon})\}.$$

By continuity of roots, we can take $\varepsilon \to 0$ to obtain

$$\max\{\lambda_{j+1}(f_1), \lambda_{j+1}(f_2)\} \leq \min\{\lambda_j(f_1), \lambda_j(f_2)\},$$

so f_1 and f_2 have a common interlacing.

Finally, we deal with common factors of $f_1(x)$ and $f_2(x)$. Induct on the degree of their gcd. If $f_i(x) = (x-\alpha)g_i(x)$, then the assumption "$h_\mu$ is real-rooted" says $(x-\alpha)(\mu g_1(x) + (1-\mu)g_2(x))$ is real-rooted, so all convex combinations of g_1 and g_2 are real-rooted and they have a common interlacing by induction hypothesis.

Now we can add back the root α and still have a common interlacing: EITHER

$$\max\{\lambda_{j+1}(g_1), \lambda_{j+1}(g_2)\} \leq \alpha \leq \min\{\lambda_j(g_1), \lambda_j(g_2)\},$$

for some j (allow $j = 0$ or $j = d-1$), in which case we can label the extra α as $\lambda_{j+1}(f_1)$ and $\lambda_{j+1}(f_2)$, OR α is strictly between $\lambda_j(g_1)$ and $\lambda_j(g_2)$ for some $j \in [d-1]$ (WLOG $\lambda_j(g_1) < \lambda_j(g_2)$), in which case we can label one α as $\lambda_j(f_1)$ to pair up with $\lambda_j(f_2) = \lambda_j(g_2)$, and the other α as $\lambda_{j+1}(f_2)$ to pair up with $\lambda_{j+1}(f_1) = \lambda_j(g_1)$. □

Lemma 1.8 (Convex combination criterion)**.** [11] Suppose $f_1(x), f_2(x), \cdots, f_n(x) \in \mathbb{R}[x]$ have positive leading coefficients and same degree d. Then they have a common interlacing if and only if all convex combinations $\sum \mu_i f_i$ are real-rooted.

Proof. "$\Longleftarrow$": If $f_1, \cdots, f_n$ have no common interlacing, then for some $j \in [d-1]$, $\max_i \lambda_j(f_i) > \min_i \lambda_{j+1}(f_i)$. Without loss of generality, assume $\lambda_j(f_1) > \lambda_{j+1}(f_2)$. This says f_1 and f_2 does not have common interlacing, so some convex combination $\mu f_1 + (1-\mu)f_2$ is not real-rooted by Lemma 1.7, which in particular is a convex combination of $f_1, \cdots, f_n$.

"$\Longrightarrow$": this is Lemma 1.5. □

1.2 Interlacing Families

The root property in Lemma 1.5 holds for a larger class of families of polynomials, because we may apply the inequality in several steps like

$$\min_a \left(\min_b \lambda_1(f_{ab})\right) \leq \min_a \lambda_1\left(\sum_j \nu_j f_{aj}\right) \leq \lambda_1\left(\sum_{ij} \mu_i \nu_j f_{ij}\right).$$

It suffices to have common interlacing at each step. More precisely:

[11] CS07, Theorem 3.6.

Definition 1.9 (Interlacing family)**.** [12] Let Σ be a finite set (the "alphabet") and $T \subseteq \Sigma^* = \{\sigma : \sigma \text{ finite sequence in } \Sigma\}$. We say T is a **finite tree** if T is finite, non-empty, and whenever $\sigma = (\sigma_0, \sigma_1, \cdots, \sigma_{n-1}) \in T$, we also have all its prefixes $\sigma \restriction i = (\sigma_0, \sigma_1, \cdots, \sigma_{i-1}) \in T$. Note that this implies $\varnothing \in T$.

For $\sigma \in T$, if for some $a \in \Sigma$, $\sigma a = (\sigma_0, \sigma_1, \cdots, \sigma_{n-1}, a)$ is also in T, then we say σa is a **child** of σ. σ is a **leaf** if it has no children. The children of the same σ are **siblings**.

An **interlacing family** is a family of polynomials $f_\sigma(x) \in \mathbb{R}[x]$ indexed by $\sigma \in T$, T a finite tree, satisfying:

(1) All f_σ have the same degree d.

(2) All f_σ have positive leading coefficient.

(3) For all $\sigma \in T$, f_σ is real-rooted.

(4) If $\sigma \in T$ is not a leaf, then f_σ is a convex combination of $f_{\sigma a}$ for children $\sigma a \in T$ of σ, and the children $f_{\sigma a}$ have a common interlacing. (One might note that (1) and (3) are implied by (2), (4), Lemma 1.5, and our definition of common interlacing.)

Sometimes we do not distinguish between σ and f_σ and say e.g. f_σ is a sibling of f_τ.

Example 1.10 (Interlacing family with no common interlacing)**.** Let $T = \{\varnothing, 1, 2, 11, 12, 21, 22\}$, and

$$\begin{aligned} f_{11}(x) &= x^2 - 1 = (x-1)(x+1), \\ f_{12}(x) &= x^2 - 49 = (x-7)(x+7), \\ f_1(x) = \frac{1}{2} f_{11}(x) + \frac{1}{2} f_{12}(x) &= x^2 - 25 = (x-5)(x+5). \end{aligned}$$

Then, let $f_{2i}(x) = f_{1i}(x-3)$ and $f_2(x) = f_1(x-3)$.

Since f_1 has roots $-5, 5$ and f_2 has roots $-2, 8$, they have a common interlacing. Similarly we can verify this gives an interlacing family (choose any convex combination $f_\varnothing$ of f_1, f_2). However, f_{11} and f_{21} have no common interlacing.

Nevertheless, the roots still satisfy an inequality similar to Lemma 1.5.

Lemma 1.11 (Root property of interlacing families)**.** [13] If $\{f_\sigma : \sigma \in T\}$ is an interlacing family of degree d, then for $1 \le j \le d$,

$$\min_{\sigma \text{ leaf}} \lambda_j(f_\sigma) \le \lambda_j(f_\varnothing) \le \max_{\sigma \text{ leaf}} \lambda_j(f_\sigma).$$

Proof. By Lemma 1.5, for each non-leaf σ, there is a child σa such that $\lambda_j(f_{\sigma a}) \le \lambda_j(f_\sigma)$. Therefore we can start at $f_\varnothing$ and iterate the above until we arrive at some leaf α such that $\lambda_j(f_\alpha) \le \lambda_j(f_\varnothing)$. Similarly there is a leaf β such that $\lambda_j(f_\varnothing) \le \lambda_j(f_\beta)$. $\square$

This provides a new probabilistic method in the following way: Let ξ be a random leaf of a finite tree T (with non-zero probability at each leaf). Suppose we have some polynomial f_σ for each leaf σ. Then for any $\sigma \in T$ of length n, we can define f_σ to be the conditional expectation of f_ξ given that $\xi \restriction n = \sigma$, i.e. the first n entries of ξ form σ.

[12] MSS15a, Definition 4.3.
[13] MSS15a, Theorem 4.4.

This gives $f_\sigma = \mathbb{E}_a(f_{\sigma a})$, where we are taking the conditional expectation on the $(n+1)$-th entry of ξ given that the first n entries form σ, so the convex combination condition in the definition of interlacing families is automatically satisfied. Moreover, $f_\varnothing = \mathbb{E}(f_\xi)$.

So if we can show that $\{f_\sigma : \sigma \in T\}$ defined this way is an interlacing family, then we know $\lambda_1(f_\xi) \le \lambda_1(f_\varnothing) = \lambda_1(\mathbb{E}(f_\xi))$ with non-zero probability. A particularly important example of such an interlacing family is described in Lemma 1.23. We shall use this key idea in Sections 2 and 3 to prove results on graphs and matrix paving.

Without the interlacing condition, it is difficult to say anything about the roots of f_ξ just from knowing the roots of $\mathbb{E}(f_\xi)$.

1.3 Real Rooted and Stable Polynomials

In order to prove that a family of polynomials is an interlacing family, we need to show that each polynomial is real-rooted, and all siblings have a common interlacing, which again is asserting all convex combinations are real-rooted by Lemma 1.8. Therefore, we need a systematic way of proving real-rootedness.

Definition 1.12 (Stability and real stability)**.** [14] A polynomial $f(z_1, \cdots, z_n) \in \mathbb{C}[z_1, \cdots, z_n]$ is **stable** if it has no zeros with positive imaginary parts, i.e. for all $(z_1, \cdots, z_n) \in \mathbb{C}^n$, if $\Im(z_i) > 0$ for all i, then $f(z_1, \cdots, z_n) \neq 0$. A polynomial is **real stable** if it is stable and has real coefficients.

Note that if a polynomial in one variable is real stable, then it is real-rooted (otherwise imaginary roots come in conjugate pairs). Therefore, stability of polynomials generalises real-rootedness to complex and multivariate polynomials.

Multivariate stability can be defined in terms of monovariate stability:

Lemma 1.13. [15] $f(z_1, \cdots, z_n)$ is stable if and only if $g(t) = f(\boldsymbol{a} + \boldsymbol{b}t) \in \mathbb{C}[t]$ is stable for all $\boldsymbol{a} \in \mathbb{R}^n, \boldsymbol{b} \in \mathbb{R}^n_{>0}$.

Proof. "$\Longrightarrow$": If $f(\boldsymbol{a} + \boldsymbol{b}t_0) = 0$ for some t_0 with $\Im t_0 > 0$, then $\Im(a_i + b_i t_0) = b_i \Im(t_0) > 0$, so f is not stable.

"$\Longleftarrow$": Suppose $f(c_1, \cdots, c_n) = 0$ with $\Im c_i > 0$ for all i. Then define $a_i = \Re c_i \in \mathbb{R}$ and $b_i = \Im c_i \in \mathbb{R}^+$. This says i is a root to $g(t) = f(\boldsymbol{a} + \boldsymbol{b}t) \in \mathbb{C}[t]$, so g is not stable. ☐

We now show that some determinants are stable (Example 1.14), and some transformations preserve (real) stability (Lemmas 1.15 and 1.16). Together, they generate a large class of stable polynomials.

Example 1.14 (Some determinants are real stable)**.** [16] If $A_1, \cdots, A_m$ are positive semi-definite self-adjoint matrices, then $f(z_1, \cdots, z_n) = \det\left(\sum_i z_i A_i\right) \in \mathbb{C}[z_1, \cdots, z_n]$ is real stable or identically zero.

Proof. Again, by continuity of roots with respect to coefficients, it suffices to consider the generic case where all A_i are positive definite.

[14] BB10, Definition 1.1.
[15] BB10, Lemma 2.1.
[16] BB10, Proposition 1.12.

"Stable": We shall apply Lemma 1.13. Fix $\boldsymbol{a} \in \mathbb{R}^n, \boldsymbol{b} \in \mathbb{R}^n_+$. We have

$$\det\left(\sum_i (a_i + b_i t) A_i\right) = \det\left(t\left(\sum_i b_i A_i\right) + \sum_i a_i A_i\right).$$

Write

$$P = \sum_i a_i A_i, \text{ and } Q = \sum_i b_i A_i$$

both of which are self-adjoint, and moreover Q is positive definite, so Q is invertible and has a self-adjoint square root, so we can write the polynomial as

$$\det(tQ + P) = \det Q^{1/2} \det(tI + Q^{-1/2} P Q^{-1/2}) \det Q^{1/2},$$

which is a constant times the characteristic polynomial of a self-adjoint matrix $-Q^{-1/2}PQ^{-1/2}$, whence real-rooted.

"Real": Note that when all z_i are real, $\sum_i z_i A_i$ is self-adjoint, so the polynomial has real value. This proves f has real coefficients. ☐

Lemma 1.15 (Evaluating stable polynomials)**.** [17] Let $n \geq 2$. If $f(z_1, \cdots, z_{n-1}, z_n) \in \mathbb{C}[z_1, \cdots, z_n]$ is stable, and c is a constant with $\Im c \geq 0$, then $g(z_1, \cdots, z_{n-1}) = f(z_1, \cdots, z_{n-1}, c) \in \mathbb{C}[z_1, \cdots, z_{n-1}]$ is either stable or identically zero. It might be identically zero only when c is real.

Proof. For the generic case where $\Im c > 0$, if $\Im z_i > 0$ for $1 \leq z_i \leq n-1$, then $(z_1, \cdots, z_{n-1}, c)$ is not a zero of f by stability, so $(z_1, \cdots, z_{n-1})$ is not a zero of g.

Now suppose $\Im c = 0$. Suppose g is not identically zero and has a zero $(a_1, \cdots, a_{n-1})$ with $\Im a_i > 0$. Consider $g_\varepsilon(z_1) = f(z_1, a_2, \cdots, a_{n-1}, c + \varepsilon i)$ for small $\varepsilon > 0$. This has a root a_1 with $\Im a_1 > 0$ when $\varepsilon = 0$. By continuity of roots, g_ε also has a root a_1' with $\Im a_1' > 0$ for sufficiently small $\varepsilon > 0$, contradicting the generic case. ☐

Remark. Therefore if f is real stable, and c is real, then g is also *real* stable or identically zero.

Lemma 1.16 (Lieb–Sokal lemma)**.** [18][19] If $f(z_1, \cdots, z_n) \in \mathbb{C}[z_1, \cdots, z_n]$ is stable, then $(1 - \partial_{z_1}) f(z_1, \cdots, z_n)$ is stable.

Proof. [20] Fix $a_2, \cdots, a_n \in \mathbb{C}$ with $\Im a_i > 0$, and let $g(z_1) = f(z_1, a_2, \cdots, a_n)$. Since f is stable, g is also stable by Lemma 1.15. Let the degree of g be $d \geq 0$, and the roots be $b_1, \cdots, b_d$, so $g(z_1) = C \prod_j (z_1 - b_j)$ for some non-zero constant $C \in \mathbb{C}$. Then $\Im b_j \leq 0$.

Now

$$(1 - \partial_{z_1}) g(z_1) = g(z_1)\left(1 - \sum_j \frac{1}{z_1 - b_j}\right).$$

When $\Im z_1 > 0$, $\Im(z_1 - b_j) > 0$, so $\Im\left(\sum_j \frac{1}{z_1 - b_j}\right) < 0$. Also $g(z_1) \neq 0$, so the right hand side is non-zero. This means $(1 - \partial_{z_1}) g(z_1)$ on the left hand side never vanishes when $\Im z_1 > 0$. ☐

[17] Tao13, Lemma 12.
[18] Special case of LS81, Lemma 2.3.
[19] MSS15b, Corollary 3.8.
[20] Tao13, Lemma 12.

There is a characterisation of all stability-preserving differential operators in the Weyl algebra $\mathbb{C}[z_1, \cdots, z_n, \partial_{z_1}, \cdots, \partial_{z_n}]$, of which Lemma 1.16 is a special case, but we do not require this theorem.

Theorem (Borcea–Brändén characterisation)**.** [21] Consider the operator

$$T = \sum_{\alpha_i, \beta_i \in \mathbb{Z}_{\geq 0}} c_{\alpha_1 \cdots \alpha_n \beta_1 \cdots \beta_n} z_1^{\alpha_1} \cdots z_n^{\alpha_n} \partial_{z_1}^{\beta_1} \cdots \partial_{z_n}^{\beta_n},$$

where only finitely many coefficients $c_{\alpha_1 \cdots \alpha_n \beta_1 \cdots \beta_n} \in \mathbb{C}$ are non-zero. This operator preserves stability of polynomials in $\mathbb{C}[z_1, \cdots, z_n]$ if and only if the corresponding polynomial

$$\sum_{\alpha_i, \beta_i \in \mathbb{Z}_{\geq 0}} c_{\alpha_1 \cdots \alpha_n \beta_1 \cdots \beta_n} z_1^{\alpha_1} \cdots z_n^{\alpha_n} (-w_1)^{\beta_1} \cdots (-w_n)^{\beta_n}$$

is stable in $\mathbb{C}[z_1, \cdots, z_n, w_1, \cdots, w_n]$.

Using this, $1 - \partial_{z_1}$ preserves stability if and only if $1 + w_1$ is stable, and indeed we see $1 + w_1$ is real-rooted.

1.4 Mixed Characteristic Polynomial

In this subsection, we shall put Section 1.3 to use and prove that a special family of polynomials is an interlacing family. This family arises from the characteristic polynomials of random matrices of the form $\sum_i \boldsymbol{v}_i \boldsymbol{v}_i^*$, where $\boldsymbol{v}_i$ are random vectors in $\mathbb{C}^n$. In particular, the matrix is a sum of rank 1 matrices.

Lemma 1.17 (Rank 1 updates are affine)**.** [22] For $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{C}^n$, $\det(I + \boldsymbol{u}\boldsymbol{v}^*) = 1 + \boldsymbol{v}^*\boldsymbol{u} = 1 + \operatorname{tr}(\boldsymbol{u}\boldsymbol{v}^*)$.

Proof. By passing a multiplicative constant from $\boldsymbol{v}$ to $\boldsymbol{u}$, we may assume $\|\boldsymbol{v}\| = 1$. Extend to an orthonormal basis $\mathcal{B} = \{\boldsymbol{v}, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$, so that $I = \boldsymbol{v}\boldsymbol{v}^* + \sum_{i=2}^n \boldsymbol{v}_i \boldsymbol{v}_i^*$. Now $I + \boldsymbol{u}\boldsymbol{v}^* = (\boldsymbol{u} + \boldsymbol{v})\boldsymbol{v}^* + \sum_i \boldsymbol{v}_i \boldsymbol{v}_i^*$. With respect to the basis $\mathcal{B}$, $I + \boldsymbol{u}\boldsymbol{v}^*$ has matrix

$$\begin{pmatrix} \boldsymbol{v}^*\boldsymbol{u} + 1 & 0 & 0 & \cdots & 0 \\ \boldsymbol{v}_2^*\boldsymbol{u} & 1 & 0 & \cdots & 0 \\ \boldsymbol{v}_3^*\boldsymbol{u} & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{v}_n^*\boldsymbol{u} & 0 & 0 & \cdots & 1 \end{pmatrix}.$$

So the determinant is $1 + \boldsymbol{v}^*\boldsymbol{u}$ and the final equality follows from $\operatorname{tr}(AB) = \operatorname{tr}(BA)$. □

Corollary 1.18 (Rank 1 updates are affine)**.** [23] If B is a fixed n by n matrix, and A is a (variable) rank 1 matrix, then $\det(B + A)$ is an affine function in the components in A, i.e. there is a linear function $f : \mathcal{M}_{n,n}(\mathbb{C}) \to \mathbb{C}$ (depending on B) such that

$$\det(B + A) = \det B + f(A)$$

holds for all rank one matrices A.

[21] BB10, Theorem 1.2.
[22] MSS15b, Lemma 3.10.
[23] Tao13, Lemma 8.

Proof. For the generic case where B is invertible, we have

$$\det(B+A) = \det(B)\det(I+B^{-1}A) = \det(B)(1+\operatorname{tr}(B^{-1}A))$$

by Lemma 1.17 (since $B^{-1}A$ has rank one). We can further rewrite this as $\det B + \operatorname{tr}(\operatorname{adj}(B)A)$, where $\operatorname{adj} B = \det(B)B^{-1}$ is the adjugate matrix, whose entries are polynomials in the entries of B.

Now $\det(B+A) = \det B + \operatorname{tr}(\operatorname{adj}(B)A)$ and both sides are polynomial in the entries of B, so they must be equal even for non-invertible B. ☐

We are considering the sum of many rank 1 updates, so the effect is "multiaffine".

Definition 1.19 (Multiaffine polynomials)**.** [24] A polynomial $f(z_1, \cdots, z_n) \in \mathbb{C}[z_1, \cdots, z_n]$ is **multiaffine** if it is affine in each z_i when we fix all other variables. In other words, if we write

$$f(z_1, \cdots, z_n) = \sum_{a_1, \cdots, a_n \geq 0} c_{a_1, a_2, \cdots, a_n} z_1^{a_1} \cdots z_n^{a_n},$$

then all coefficients $c_{a_1, \cdots, a_n}$ with some $a_i \geq 2$ are 0.

The **multiaffine part** of $f(z_1, \cdots, z_n)$, denoted by $\mathrm{MAP}(f)(z_1, \cdots, z_n)$, is

$$\sum_{0 \leq a_i \leq 1\ \forall i} c_{a_1, \cdots, a_n} z_1^{a_1} \cdots z_n^{a_n}$$

using the notation above. MAP removes all terms that has degree at least 2 in some variable.

Lemma 1.20 (Expression of MAP)**.** [25]

$$\mathrm{MAP}(f)(t_1, \cdots, t_n) = \left(\prod_i (1 + t_i \partial_{z_i}) \right) f(z_1, \cdots, z_n) \bigg|_{z_1 = \cdots = z_n = 0}.$$

Proof. For $1 \leq i_1 < i_2 < \cdots < i_d \leq n$,

$$\partial_{z_{i_1}} \partial_{z_{i_2}} \cdots \partial_{z_{i_d}} f(z_1, \cdots, z_n) \Big|_{z_1 = z_2 = \cdots = z_n = 0}$$

is precisely the coefficient of the $t_{i_1} t_{i_2} \cdots t_{i_d}$ term in the polynomial $f(t_1, \cdots, t_n)$. Since both sides are multiaffine in the t-variables, this says all their corresponding coefficients are equal. ☐

Lemma 1.21 (Mixed characteristic polynomial)**.** [26]

(1) If $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n$ are independent random vectors in $\mathbb{C}^d$, each taking finitely many possible values, then the mean of the characteristic polynomial

$$\mathbb{E}\chi\left(\sum_{i=1}^n \boldsymbol{v}_i \boldsymbol{v}_i^*\right)(x) = \mathbb{E}\det\left(xI - \sum_i \boldsymbol{v}_i \boldsymbol{v}_i^*\right)$$

only depends on the expectations $A_i = \mathbb{E}(\boldsymbol{v}_i \boldsymbol{v}_i^*)$. Note that A_i must be positive semi-definite.

[24] Von18, p. 1.
[25] Tao13, after Corollary 9.
[26] MSS15b, Theorem 4.1.

(2) This expression is called the **mixed characteristic polynomial** of $A_1, \cdots, A_n$, denoted by $\mu[A_1, \cdots, A_n](x)$, and equal to

$$\left(\prod_i (1-\partial_{z_i})\right) \det\left(xI + \sum_i z_i A_i\right)\Bigg|_{z_1=\cdots=z_n=0}.$$

Note that this defines $\mu[A_1, \cdots, A_n]$ for all matrices $A_1, \cdots, A_n$, not just those expressible as $\mathbb{E}(\boldsymbol{v}\boldsymbol{v}^*)$.

Proof. [27]

(1) By Corollary 1.18, $\det(xI - \sum_i \boldsymbol{v}_i\boldsymbol{v}_i^*)$ is affine in the entries of $\boldsymbol{v}_i\boldsymbol{v}_i^*$ when we fix other $\boldsymbol{v}_j$, so when we expand $\det(xI - \sum_i \boldsymbol{v}_i\boldsymbol{v}_i^*)$ as a polynomial in the entries of $\boldsymbol{v}_i\boldsymbol{v}_i^*$, it is multiaffine, i.e. each term is a product of entries from $\boldsymbol{v}_i\boldsymbol{v}_i^*$ for distinct i. By independence, the expectation of that product is the product of the respective expectations.

(2) Write U_i for the random rank one matrix $\boldsymbol{v}_i\boldsymbol{v}_i^*$. For general matrices $B_1 \cdots, B_n$, consider the polynomial $P_B(x)(t_1, \cdots, t_n) = \det(xI + \sum_i t_i B_i)$. We can think of this as a polynomial in $t_i B_{ijk}$ where B_{ijk} are entries of B_i, so $\mathrm{MAP}(P_B(x))(t_1, \cdots, t_n)$, which is multiaffine in t_i, must also be multiaffine in the entries of B_i.

By iterating Corollary 1.18, $P_U(x)(t_1, \cdots, t_n) = \det(xI + \sum_i t_i U_i)$ is already multiaffine in $t_1, \cdots, t_n$, so we have

$$P_U(x)(t_1, \cdots, t_n) = \mathrm{MAP}(P_U(x))(t_1, \cdots, t_n).$$

Taking expectations, by multiaffineness in matrix entries and independence, we have

$$\mathbb{E}P_U(x)(t_1, \cdots, t_n) = \mathrm{MAP}(P_{\mathbb{E}U}(x))(t_1, \cdots, t_n) = \mathrm{MAP}(P_A(x))(t_1, \cdots, t_n). \qquad (\heartsuit)$$

The characteristic polynomial is

$$\chi\left(\sum_{i=1}^n \boldsymbol{v}_i\boldsymbol{v}_i^*\right)(x) = P_U(x)(-1, -1, \cdots, -1),$$

so taking $t_i = -1$ for all i in $(\heartsuit)$ gives

$$\mathbb{E}\chi\left(\sum_{i=1}^n \boldsymbol{v}_i\boldsymbol{v}_i^*\right)(x) = \mathrm{MAP}(P_A(x))(-1, \cdots, -1).$$

Hence

$$\mu[A_1, \cdots, A_n](x) = \left(\prod_i (1-\partial_{z_i})\right) \det\left(xI + \sum_i z_i A_i\right)\Bigg|_{z_1=\cdots=z_n=0}$$

by Lemma 1.20 (with $t_i = -1$). ☐

Remark. We cannot go from $P_U(x)(t_1, \cdots, t_n)$ to $P_{\mathbb{E}U}(x)(t_1, \cdots, t_n)$. The reason is that $A_i = \mathbb{E}U_i$ might not have rank one, so $P_A(x)(t_1, \cdots, t_n) = \det(xI + \sum_i t_i A_i)$ might not be affine in A_i. It is therefore necessary to go through the multiaffine part of $\det(xI + \sum_i t_i A_i)$ as above.

Lemma 1.22. [28] If $A_1, \cdots, A_n$ are positive semi-definite self-adjoint d by d matrices, then their mixed characteristic polynomial $\mu[A_1, \cdots, A_n](x)$ is a real polynomial and is real-rooted.

[27] Tao13, Corollary 4.
[28] MSS15b, Corollary 4.4.

Proof. Example 1.14 says $\det(xI + \sum_i z_i A_i)$ is stable. By iterating Lemma 1.16, we know $(\prod_i (1 - \partial_{z_i}))\det(xI + \sum_i z_i A_i)$ is stable. Finally by Lemma 1.15, substituting $z_1 = z_2 = \cdots = z_n = 0$ preserves stability, so $\mu[A_1, \cdots, A_n](x)$ is stable.

$\mu[A_1, \cdots, A_n](x)$ is also real ($xI + \sum_i z_i A_i$ is self-adjoint for real x and z_i) and monovariate, so it is real-rooted. □

Finally, we describe how we can get an interlacing family from the mixed characteristic polynomials.

Lemma 1.23 (Mixed characteristic polynomials give rise to an interlacing family)**.** [29] Let $\boldsymbol{v}_i$ ($i = 1, 2, \cdots, n$) be independent random vectors, with $\boldsymbol{v}_i$ taking value from the set of constant vectors $\{\boldsymbol{w}_{i1}, \cdots, \boldsymbol{w}_{is_i}\} \subseteq \mathbb{C}$. Let ξ be the random sequence such that $\boldsymbol{v}_i = \boldsymbol{w}_{i\xi_i}$.

Let $T = \{\sigma = (\sigma_1, \cdots, \sigma_k) : 0 \leq k \leq n \text{ and } 1 \leq \sigma_i \leq s_i\}$ and for $\sigma = (\sigma_1, \cdots, \sigma_k) \in T$, define

$$\begin{aligned} f_\sigma(x) &= \mathbb{E}_{\boldsymbol{v}} \chi \left(\sum_{i=1}^{k} \boldsymbol{w}_{i\sigma_i} \boldsymbol{w}_{i\sigma_i}^* + \sum_{i=k+1}^{n} \boldsymbol{v}_i \boldsymbol{v}_i^* \right)(x) \\ &= \mu[\boldsymbol{w}_{1\sigma_1} \boldsymbol{w}_{1\sigma_1}^*, \cdots, \boldsymbol{w}_{k\sigma_k} \boldsymbol{w}_{k\sigma_k}^*, \mathbb{E}(\boldsymbol{v}_{k+1} \boldsymbol{v}_{k+1}^*), \cdots, \mathbb{E}(\boldsymbol{v}_n \boldsymbol{v}_n^*)](x). \end{aligned}$$

Then $\{f_\sigma : \sigma \in T\}$ is an interlacing family. In particular, there is some leaf σ such that $\lambda_1(f_\sigma) \leq \lambda_1(f_\varnothing)$. Equivalently, $\lambda_1(f_\xi) \leq \lambda_1(\mathbb{E}(f_\xi))$ with non-zero probability.

Proof. f_σ is monic and has degree d for all $\sigma \in T$. Moreover, they are real-rooted by Lemma 1.22.

It remains to show that for any non-leaf $\sigma = (\sigma_1, \cdots, \sigma_k)$ with $0 \leq k < n$, the polynomials $f_{\sigma a}$ ($1 \leq a \leq s_{k+1}$) have a common interlacing, and f_σ is a convex combination of $f_{\sigma a}$. The latter is clear from $f_\sigma = \mathbb{E}_{\xi_{k+1}}(f_{\sigma \xi_{k+1}})$. To show a common interlacing, we shall apply Lemma 1.8, so we need to consider an arbitrary convex combination.

A convex combination $\sum_a \mu_a f_{\sigma a}$ is just the expectation taken over a different distribution: if $\boldsymbol{v}'$ takes value $\boldsymbol{w}_{k+1,a}$ with probability μ_a, independent of other $\boldsymbol{v}_i$, then

$$\begin{aligned} \sum_a \mu_a f_{\sigma a}(x) &= \sum_a \mu_a \mathbb{E}_{\boldsymbol{v}} \chi \left(\sum_{i=1}^{k} \boldsymbol{w}_{i\sigma_i} \boldsymbol{w}_{i\sigma_i}^* + \boldsymbol{w}_{k+1,a} \boldsymbol{w}_{k+1,a}^* + \sum_{i=k+2}^{n} \boldsymbol{v}_i \boldsymbol{v}_i^* \right)(x) \\ &= \mathbb{E}_{\boldsymbol{v}'} \mathbb{E}_{\boldsymbol{v}} \chi \left(\sum_{i=1}^{k} \boldsymbol{w}_{i\sigma_i} \boldsymbol{w}_{i\sigma_i}^* + \boldsymbol{v}' \boldsymbol{v}'^* + \sum_{i=k+2}^{n} \boldsymbol{v}_i \boldsymbol{v}_i^* \right)(x) \\ &= \mu[\boldsymbol{w}_{1\sigma_1} \boldsymbol{w}_{1\sigma_1}^*, \cdots, \boldsymbol{w}_{k\sigma_k} \boldsymbol{w}_{k\sigma_k}^*, \mathbb{E}_{\boldsymbol{v}'}(\boldsymbol{v}' \boldsymbol{v}'^*), \mathbb{E}(\boldsymbol{v}_{k+2} \boldsymbol{v}_{k+2}^*), \cdots, \mathbb{E}(\boldsymbol{v}_n \boldsymbol{v}_n^*)](x), \end{aligned}$$

which is again real-rooted as a mixed characteristic polynomial.

Therefore they form an interlacing family, and the last claim follows from Lemma 1.11. □

Both interlacing families we consider in Sections 2 and 3 are of the form described by Lemma 1.23.

[29] MSS15b, Theorem 4.5.

2 Ramanujan Graphs of All Degrees

All graphs in this section are simple and undirected.

2.1 Ramanujan Graphs

Definition 2.1 (Spectrum of a graph)**.** For a graph $G = ([n], E)$, its **adjacency matrix** A is the symmetric n by n matrix defined by $A_{ij} = \begin{cases} 1, & \text{if } ij \in E, \\ 0, & \text{if } ij \notin E. \end{cases}$ The **eigenvalues** of the graph are the eigenvalues of A, denoted by

$$\lambda_n(G) \leq \lambda_{n-1}(G) \leq \cdots \leq \lambda_1(G).$$

They do not depend on the labelling of the vertices.

If G is d-regular, then we have $\lambda_1(G) = d$, and $\lambda_n(G) \geq -d$, so we say $\lambda_1(G) = d$ is a **trivial eigenvalue**. If in addition, G is bipartite, then $\lambda_n(G) = -d$ and $\lambda_{n+1-i}(G) = -\lambda_i(G)$ for all i. In this case $\lambda_n(G) = -d$ is another trivial eigenvalue.

Assuming connectivity, the other eigenvalues ($\lambda_2(G)$ to $\lambda_{n-1}(G)$, and $\lambda_n(G)$ if G non-bipartite) have absolute values less than d. They are the **non-trivial eigenvalues**.

Definition 2.2 (Ramanujan graphs)**.** [30] A (connected) d-regular graph G is **Ramanujan** if all of its non-trivial eigenvalues $\lambda_i(G)$ satisfies $|\lambda_i(G)| \leq 2\sqrt{d-1}$.

The aim of this section is to present the proof of the following theorem as one of the earliest applications of the interlacing family method.

Aim 2.3 (Marcus–Spielman–Srivastava 2015)**.** [31] For every $d \geq 3$, there is an infinite family of d-regular bipartite Ramanujan graphs.

To build this family of Ramanujan graphs, we start with some trivial Ramanujan graphs (Example 2.4), and build larger and larger graphs, each being twice as large as the previous one, and such that the eigenvalues of the new graph are the old ones together with some new ones (Lemma 2.8). In order to guarantee the new graph is still Ramanujan, we need to ensure that the new eigenvalues have absolute value at most $2\sqrt{d-1}$.

This is where Lemma 1.23 comes in, providing a bound on λ_1 of some polynomial. However, we cannot simultaneously control both the new λ_1 and the new λ_n (in the sense that there is some σ such that $\lambda_1(f_\sigma) \leq \lambda_1(\mathbb{E}(f))$, and some σ' such that $\lambda_n(f_{\sigma'}) \geq \lambda_n(\mathbb{E}(f))$, but we cannot guarantee $\sigma = \sigma'$). Therefore, the proof using Lemma 1.23 only works for bipartite graphs: they have $\lambda_{n-i+1} = -\lambda_i$, so once we know the new λ_1 is at most $2\sqrt{d-1}$, then we immediately know the new λ_n is also at least $-2\sqrt{d-1}$.

One reason why we care about Ramanujan graphs is that they are spectral expanders, which has nice quasi-randomness properties. For example, a version of the expander mixing lemma most relevant to bipartite Ramanujan graphs is stated below.

Theorem (Bipartite expander mixing lemma)**.** [32] If G is a d-regular bipartite graph with parts U, V, and $|U| = |V| = n$, and $\lambda = \max\{|\lambda_2|, |\lambda_{2n-1}|\}$, then for any $X \subseteq U$ and $Y \subseteq V$, the

[30] LPS88, Definition 1.1.
[31] MSS15a, Theorem 5.5.
[32] DSV12, Lemma 8.

number $e(X,Y)$ of edges between X and Y is close to what we expect for a random graph, in the sense that

$$\left| e(X,Y) - \frac{d}{n} |X| \, |Y| \right| \leq \lambda \sqrt{|X| \, |Y|}.$$

Moreover, Ramaunujan graphs are the best possible spectral expanders in the following sense:

Theorem (Alon–Boppana bound)**.** [33] If G is a d-regular graph of order n, then

$$\lambda_2(G) \geq 2\sqrt{d-1} - \mathcal{O}((\log n)^{-1})$$

for large n. In particular, for any $\varepsilon > 0$, there is no infinite family (G_i) of d-regular graphs such that $\lambda_2(G_i) < 2\sqrt{d-1} - \varepsilon$ for all i.

2.2 Identifying an Interlacing Family

First, we make precise how we would build the bipartite Ramanujan graphs.

Example 2.4 (Trivial Ramanujan graphs)**.** [34] For any $d \geq 1$, the complete bipartite graph $K_{d,d}$ is Ramanujan.

Proof. Its adjancency matrix is

$$\begin{pmatrix} 1 & \cdots & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 1 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & \cdots & 1 \end{pmatrix},$$

which has rank 2, so except $\lambda_1(K_{d,d}) = d$ and $\lambda_{2d}(K_{d,d}) = -d$, all other eigenvalues are 0. □

The operation that we use to double the order of a graph is the following:

Definition 2.5 (2-lifts)**.** [35] A **2-lift** of a graph $G = (V, E)$ is a graph $G' = (V', E')$ together with a homomorphism $\pi : G' \to G$ such that:

(1) For each $v \in V$, there are exactly two $v' \in V'$ with $\pi(v') = v$.

(2) For each edge $uv \in E$ and $u' \in V'$ with $\pi(u') = u$, there is a unique $v' \in V'$ such that $u'v' \in E'$ and $\pi(v') = v$.

Remark. If we think of the graph as a topological cell-complex, then a 2-lift is just a 2-sheeted covering space. Similarly for n-lifts and universal covers (which must be trees because universal covers are simply connected).

Definition 2.6 (Signed adjacency matrix of 2-lifts)**.** [36] Let $G = ([n], E)$. Let $G' = (V', E')$ and $\pi : G' \to G$ be a 2-lift. For $i \in [n]$, let $\pi^{-1}(i) = \{a_i, b_i\}$. By (2) in Definition 2.5, for any $ij \in E$, we have exactly one of the following:

[33] Alo86, p.95.
[34] MSS15a, Lemma 5.4.
[35] BL06, Section 2.
[36] BL06, Section 2.

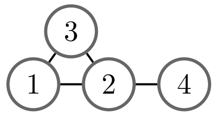
(a) $a_ia_j,\ b_ib_j \in E'$ and $a_ib_j,\ b_ia_j \notin E'$, or

(b) $a_ib_j,\ b_ia_j \in E'$ and $a_ia_j,\ b_ib_j \notin E'$.

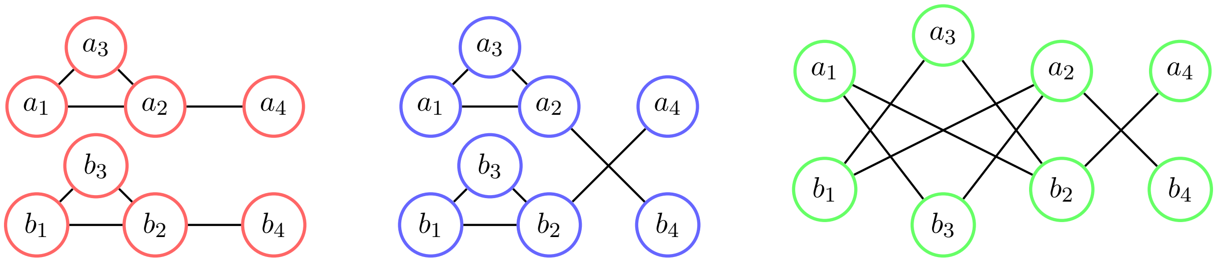
Define the **signing** $s : E \to \{\pm 1\}$ with respect to the partition $V' = \{a_1, \cdots, a_n\} \cup \{b_1, \cdots, b_n\}$ by $s(ij) = \begin{cases} +1, & \text{if (a) holds,} \\ -1, & \text{if (b) holds,} \end{cases}$ and we say the n by n matrix A_s with $(A_s)_{ij} = \begin{cases} s(ij), & \text{if } ij \in E, \\ 0, & \text{if } ij \notin E. \end{cases}$ is the **signed adjacency matrix**.

The signed adjacency matrix of a 2-lift is well-defined up to conjugation: if we swap the labels of a pair a_i, b_i, then all entries in the i-th row or i-th column of A_s change signs (the intersection $(A_s)_{ii}$ is always 0), which is the effect of changing basis from $\{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n\}$ to $\{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_{i-1}, -\boldsymbol{e}_i, \boldsymbol{e}_{i+1}, \cdots, \boldsymbol{e}_n\}$.

Example 2.7 (2-lifts)**.** Start with the graph G on the right.



There are different 2-lifts. Some of them are shown below, which we call G_1, G_2, G_3 from left to right. G_1 is just two disjoint copies of G, and its signing is given by $s_1(ij) = +1$ for all ij. G_2, despite having $s_2(24) = -1 \neq s_1(24)$, is isomorphic to G_1. G_3 has $s_3(ij) = -1$ for all ij, and is bipartite. In general, the 2-lift with $s(ij) = -1$ for all ij is always bipartite, and it is called the **canonical double cover**[37].



The 2-lift is a useful construction in spectral graph theory, because its eigenvalues can be described easily.

Lemma 2.8 (Eigenvalues of 2-lifts)**.** [38] Let $G = ([n], E)$, $G' = (V', E')$ with $\pi : G' \to G$ a 2-lift, and A_s be its signed adjacency matrix. Then the $2n$ eigenvalues of G' are precisely the n eigenvalues of G together with the n eigenvalues of A_s (counting multiplicity).

Proof. We may label the vertices of G' so that $V' = [2n]$ and $\pi(i) = \pi(n+i) = i \in G$, and let s be the signing with respect to the partition $[2n] = \{1, 2, \cdots, n\} \cup \{n+1, \cdots, 2n\}$. For $i, j \in [n]$, let $(A_s^+)_{ij} = \begin{cases} 1, & \text{if } s(ij) = +1, \\ 0, & \text{otherwise,} \end{cases}$ and $(A_s^-)_{ij} = \begin{cases} 1, & \text{if } s(ij) = -1, \\ 0, & \text{otherwise,} \end{cases}$, so that $A_s = A_s^+ - A_s^-$, and the unsigned adjacency matrix of G is $A = A_s^+ + A_s^-$.

By definition of the signing, G' has adjacency matrix

$$A' = \begin{pmatrix} A_s^+ & A_s^- \\ A_s^- & A_s^+ \end{pmatrix}.$$

If $A\boldsymbol{v} = \lambda \boldsymbol{v}$, then

$$\begin{pmatrix} A_s^+ & A_s^- \\ A_s^- & A_s^+ \end{pmatrix} \begin{pmatrix} \boldsymbol{v} \\ \boldsymbol{v} \end{pmatrix} = \lambda \begin{pmatrix} \boldsymbol{v} \\ \boldsymbol{v} \end{pmatrix},$$

[37]The name "bipartite double cover" is discouraged by Pis18.

[38]BL06, Lemma 3.1.

giving n eigenvalues of A'.

Similarly, if $A_s\boldsymbol{v} = \lambda\boldsymbol{v}$, then

$$\begin{pmatrix} A_s^+ & A_s^- \\ A_s^- & A_s^+ \end{pmatrix} \begin{pmatrix} \boldsymbol{v} \\ -\boldsymbol{v} \end{pmatrix} = \lambda \begin{pmatrix} \boldsymbol{v} \\ -\boldsymbol{v} \end{pmatrix},$$

giving another n eigenvalues of A' because any two vectors of the form $(\boldsymbol{v}_1, \boldsymbol{v}_1)$ and $(\boldsymbol{v}_2, -\boldsymbol{v}_2)$ must be orthogonal. Hence the $2n$ eigenvectors of A' are precisely those of A and A_s. □

Therefore, it suffices to show that for any d-regular graph G, there is a signing s such that the eigenvalues of A_s have absolute value at most $2\sqrt{d-1}$. This is the **Bilu–Linial conjecture**.[39] Our bipartite case is easier because the eigenvalues pair up, and it suffices to show that there is some s such that $\lambda_1(A_s) \le 2\sqrt{d-1}$.

Lemma 2.9. [40] For any graph $G = ([n], E)$, the characteristic polynomials $f_s(x) = \chi(A_s)(x)$ (s is a signing) are leaves of an interlacing family with the polynomial $f_\varnothing = \mathbb{E}(f_s)$ at the root of the tree, where $\mathbb{E}$ is taken with s uniformly at random. Therefore, there is some s such that $\lambda_1(f_s) \le \lambda_1(\mathbb{E}(f_s))$.

Proof. Think of a signing $s : E \to \{\pm 1\}$ as a string $s = (s_1, s_2, \cdots, s_m)$ of length $m = |E|$ with $s_i = \pm 1$. We would like to apply Lemma 1.23, but the contribution of s_i (corresponding to the i-th edge $a_ib_i \in E,\ a_i < b_i$) to A_s is either

$$\begin{pmatrix} 0 & +1 \\ +1 & 0 \end{pmatrix} \text{ or } \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$$

in some submatrix (the intersection of the a_i-th and b_i-th rows and columns), which is not a rank 1 update.

However, if we add 1 to the diagonal of that submatrix, then it has rank 1. Formally, let $\boldsymbol{w}_{i,+1} = \boldsymbol{e}_{a_i} + \boldsymbol{e}_{b_i}$ and $\boldsymbol{w}_{i,-1} = \boldsymbol{e}_{a_i} - \boldsymbol{e}_{b_i}$, then the contribution of s_i to A_s is

$$\boldsymbol{w}_{i,s_i}\boldsymbol{w}_{i,s_i}^* - \boldsymbol{e}_{a_i}\boldsymbol{e}_{a_i}^* - \boldsymbol{e}_{b_i}\boldsymbol{e}_{b_i}^*,$$

so summing over $1 \le i \le m$, we obtain

$$A_s = \sum_{i=1}^{m} \boldsymbol{w}_{i,s_i}\boldsymbol{w}_{i,s_i}^* - \sum_{j=1}^{n} d_j \boldsymbol{e}_j\boldsymbol{e}_j^*,$$

where d_j is the degree of vertex j. Let Δ be the maximum degree, then

$$A_s + \Delta I = \sum_{i=1}^{m} \boldsymbol{w}_{i,s_i}\boldsymbol{w}_{i,s_i}^* + \sum_{j=1}^{n} (\Delta - d_j)\boldsymbol{e}_j\boldsymbol{e}_j^*,$$

so by Lemma 1.23 (with random vectors $\boldsymbol{v}_i$ taking values $\boldsymbol{w}_{i,\pm 1}$ uniformly and independently at random, and for $1 \le j \le n$, $\Delta - d_j$ many auxilliary vectors taking values $\boldsymbol{e}_j$ surely), $\{\chi(A_s + \Delta I) : s \in \{\pm 1\}^m\}$ is the set of leaves of an interlacing family.

Since

$$\chi(A_s)(x) = \chi(A_s + \Delta I)(x + \Delta),$$

$\{\chi(A_s) : s \in \{\pm 1\}^m\}$ is also the set of leaves of an interlacing family, so $\lambda_1(f_s) \le \lambda_1(\mathbb{E}(f_s))$ for some s. □

Now, it suffices to prove $\lambda_1(\mathbb{E}(f_s)) \le 2\sqrt{d-1}$.

[39] BL06, Conjecture 3.1.

[40] MSS15a, Theorem 5.2.

2.3 Bounding the Roots

We can identify $\mathbb{E}(f_s)$ with a generating function that counts matchings, and [HL72] showed that its roots have absolute value at most $2\sqrt{d-1}$.

Lemma 2.10 (Matching polynomial)**.** [41] Let G be a graph on vertex set $[n]$ with m edges. Let $s \in \{\pm 1\}^m$ be uniformly at random, and A_s the signed adjacency matrix corresponding to s. Then $\mathbb{E}\chi(A_s)(x)$ is equal to the **matching polynomial**

$$\mu_G(x) = \sum_{i=0}^{\lfloor n/2 \rfloor} (-1)^i x^{n-2i} m_i$$

where m_i is the number of matchings with i edges in G, and it is real-rooted.

Proof. We need to compute $\mathbb{E}\chi(A_s)(x) = \mathbb{E}\det(xI - A_s)$, which is just $\mathbb{E}\det(xI + A_s)$ because A_s and $-A_s$ have the same distribution. Write $B_s = xI + A_s$.

Expand $\det B_s$ as the sum of the terms $\operatorname{sgn}(\sigma)\prod_{i=1}^{n}(B_s)_{i\sigma(i)}$, where σ traverses all permutations $[n] \to [n]$ and sgn denotes the sign of the permutation.

If for some fixed σ, the term does not vanish, then the only factors that can appear are either of the form $(B_s)_{ii}$, which is x, or $(B_s)_{ij}$ with $ij \in E$. If $(B_s)_{ij}$ appears, but $(B_s)_{ji}$ does not appear, then by independence and $\mathbb{E}(B_s)_{ij} = 0$, the expectation of that term also vanish. If both $(B_s)_{ij}$ and $(B_s)_{ji}$ appear, then they cancel each other (both are ± 1 with same sign).

Therefore, the terms with non-zero expectations are those with σ being a product of disjoint transpositions (and fixing all other $i \in [n]$). If σ is a product of k disjoint transpositions (corresponding to a matching with k edges), then $\operatorname{sgn}(\sigma) = (-1)^k$, and it fixes $n-2k$ indices, so it contribute $(-1)^k x^{n-2k}$ to the expectation. This shows $\mathbb{E}\chi(A_s)(x) = \mu_G(x)$.

The last claim follows from Lemma 2.9 because there is an interlacing family $\left\{f_a : a \in \{\pm 1\}^{\leq m}\right\}$ such that $f_\varnothing = \mathbb{E}\chi(A_s)$ and $f_a = \chi(A_a)$ for leaves $a \in \{\pm 1\}^m$. □

[HL72] investigated the roots of the matching polynomial for general weighted graphs. They did this in the context of statistical physics, where vertices are particles that can either exist in isolation (as "monomer") or bonded with another particle (as part of a "dimer"), and dimers can have different energies. They asked whether a phase transition will occur in the system when the fugacity of monomers varies (corresponding to the variable x in $\mu_G(x)$). The bound $|\lambda_i(\mu_G)| \leq 2\sqrt{d-1}$ (Theorem 2.12) was an intermediate step in their work.

Now we shall present their proof but we restrict to unweighted graphs.

Lemma 2.11 (Recurrence for the matching polynomial)**.** [42] Let i be a vertex of G, then

$$\mu_G(x) = x \cdot \mu_{G-i}(x) - \sum_{\substack{j \in G-i, \\ ij \in E}} \mu_{G-i-j}(x),$$

where $G - v$ is the the graph G with vertex v removed.

Proof. Write $m_k(G)$ for the number of matchings with k edges in G. Given a matching M consisting of k edges in G, exactly one of the following holds:

[41] MSS15a, Theorem 3.6.
[42] HL72, Equation (4.1).

(a) the vertex i is not matched, in which case M is a matching in $G - i$, or

(b) the vertex i is matched to some $j \in G - i$, with $ij \in E$, in which case $M \setminus \{ij\}$ is a matching of size $k - 1$ in $G - i - j$, so

$$m_k(G) = m_k(G - i) + \sum_{\substack{j \in G - i, \\ ij \in E}} m_{k-1}(G - i - j).$$

Multiply both sides by $(-1)^k x^{n-2k}$ and sum over k, and we obtain the recurrence. □

Knowing that μ_G has a recursive definition, we can bound its roots by an inductive argument.

Theorem 2.12 (Heilmann–Lieb 1972). [43] Let G be a graph of order n with max degree $\Delta \geq 2$. Then the largest root $\lambda_1(\mu_G)$ of the matching polynomial satisfies $\lambda_1(\mu_G) < 2\sqrt{\Delta - 1}$.

Proof. For a spanning subgraph H of G, we say $(H, H - i)$ is a good pair if $i \in H$ and there is an edge ij in G such that $j \notin H$. We shall prove by induction the following claim:

Claim. [44] *If $(H, H - i)$ is a good pair of spanning subgraphs of G, then whenever $x \geq 2\sqrt{\Delta - 1}$, we have $\mu_{H-i}(x) > 0$, $\mu_H(x) > 0$, and*

$$\frac{\mu_H(x)}{\mu_{H-i}(x)} > \sqrt{\Delta - 1}.$$

Proof of Claim. Base case: If H has only 1 vertex i, then $\mu_H(x) = x$ and $\mu_{H-i}(x) = 1$, so the result holds.

Inductive step: by Lemma 2.11,

$$\mu_H(x) = x\mu_{H-i}(x) - \sum_{\substack{j \in H - i, \\ ij \in E}} \mu_{H-i-j}(x).$$

Since $(H, H - i)$ is a good pair, i has an edge not in H, so the sum over j has at most $\Delta - 1$ terms. Moreover, for each term, since $j \in H - i$ and $ij \in E$, $(H - i, H - i - j)$ is a good pair (j has an edge ij not in $H - i$), so by the induction hypothesis, $\mu_{H-i}(x) > 0, \mu_{H-i-j}(x) > 0$, and $\mu_{H-i-j}(x) < \frac{\mu_{H-i}(x)}{\sqrt{\Delta-1}}$ whenever $x \geq 2\sqrt{\Delta - 1}$.

Therefore, when $x \geq 2\sqrt{\Delta - 1}$,

$$\begin{aligned}
\mu_H(x) &= x\mu_{H-i}(x) - \sum_{\substack{j \in H - i, \\ ij \in E}} \mu_{H-i-j}(x) \\
&> 2\sqrt{\Delta - 1} \cdot \mu_{H-i}(x) - (\Delta - 1) \cdot \frac{\mu_{H-i}(x)}{\sqrt{\Delta - 1}} \\
&= \sqrt{\Delta - 1} \cdot \mu_{H-i}(x).
\end{aligned}$$

In particular, $\mu_H(x) > 0$ when $x \geq 2\sqrt{\Delta - 1}$. □

[43] HL72, Theorem 4.3, unweighted version.

[44] HL72, Lemma 4.4.

$(G, G-i)$ is not a good pair for any i, so if we perform the inductive step as above, we do not know that the sum has at most $\Delta - 1$ terms. Nonetheless, it has at most Δ terms, so for $x \geq 2\sqrt{\Delta - 1}$ and any $i \in G$:

$$\begin{aligned}\mu_G(x) &= x\mu_{G-i}(x) - \sum_{\substack{j \in G-i, \\ ij \in E}} \mu_{G-i-j}(x) \\ &> 2\sqrt{\Delta - 1} \cdot \mu_{G-i}(x) - \Delta \cdot \frac{\mu_{G-i}(x)}{\sqrt{\Delta - 1}} \\ &\geq 0.\end{aligned}$$

(The last inequality holds because $2(\Delta - 1) \geq \Delta$ for $\Delta \geq 2$.)

Hence $\lambda_1(\mu_G) < 2\sqrt{\Delta - 1}$. ☐

To conclude this section:

Proof of Aim 2.3. Let $G_0 = K_{d,d}$, which is a d-regular bipartite Ramanujan graph by Example 2.4. By Lemma 2.9, Lemma 2.10, and Theorem 2.12 in that order, there is a signing $s : E(G_0) \to \{\pm 1\}$ such that

$$\lambda_1(\chi(A_s)) \leq \lambda_1(\mathbb{E}\chi(A_s)) = \lambda_1(\mu_G) \leq 2\sqrt{d-1}.$$

Since 2-lifts of bipartite graphs are bipartite, the 2-lift G_1 corresponding to the signing s is bipartite. By Lemma 2.8, its eigenvalues are the eigenvalues of G_0 and those of A_s, so the eigenvalues of A_s satisfies $\lambda_{2d-i+1}(A_s) = -\lambda_i(A_s)$ and we have $\lambda_{2d}(A_s) \geq -2\sqrt{d-1}$.

This means G_1 is a d-regular bipartite Ramanujan graph of order $4d$. We can repeat this procedure to build an infinite family $G_0, G_1, \cdots$ such that G_n is a d-regular bipartite Ramanujan graph of order $2^{n+1}d$. ☐

3 Kadison–Singer Problem

Definition 3.1 (Some C^*-algebras)**.** [45] Write $\boldsymbol{\ell_2}$ for the ℓ_2-space $\{(a_1, a_2, \cdots) : a_i \in \mathbb{C}, \sum_{i=1}^{\infty} |a_i|^2 < \infty\}$, which is a complex Hilbert space with inner product

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^* \boldsymbol{y} = \sum_{i=1}^{\infty} x_i^* y_i.$$

$\boldsymbol{B(\ell_2)}$ is the algebra of bounded operators $\ell_2 \to \ell_2$. Inside $B(\ell_2)$, there is an abelian subalgebra $\boldsymbol{D(\ell_2)}$ consisting of the diagonal operators, i.e. those $T : \ell_2 \to \ell_2$ with $T(a_1, a_2, \cdots) = (d_1 a_1, d_2 a_2, \cdots)$ for some $(d_1, d_2, \cdots)$ with $\sup_i |d_i| < \infty$).

Apart from the algebra operations $(+, -, \cdot)$ in $B(\ell_2)$, there are the **operator norm** $\|\cdot\| : B(\ell_2) \to \mathbb{R}_{\geq 0}$, defined by

$$\|T\| = \sup_{\substack{\|\boldsymbol{x}\|=1, \\ \boldsymbol{x} \in \ell_2}} \|T\boldsymbol{x}\|,$$

and the map $* : B(\ell_2) \to B(\ell_2)$ sending an operator $T \in B(\ell_2)$ to its **adjoint** T^*, defined by

$$\langle \boldsymbol{x}, T\boldsymbol{y} \rangle = \langle T^*\boldsymbol{x}, \boldsymbol{y} \rangle \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \ell_2.$$

[45] Tao13, before Theorem 29.

Together they make $B(\ell_2)$ into a ***C**-algebra** (whose precise definition we shall omit). Similarly $D(\ell_2)$ is a C^*-algebra.

For self-adjoint $A, B \in B(\ell_2)$, we write $A \succeq B$ if $A - B$ is positive semi-definite.

Definition 3.2 (States)**.** A **state** on $B(\ell_2)$ is a bounded linear map $f : B(\ell_2) \to \mathbb{C}$ that satisfies $f(I) = 1$ for the identity operator I and $f(A) \geq 0$ for all positive semi-definite $A \in B(\ell_2)$ (in other words, if $A \succeq B$, then $f(A) \geq f(B)$). A state is **pure** if it cannot be written as a convex combination of other states. Similarly for states on $D(\ell_2)$. A state $f : B(\ell_2) \to \mathbb{C}$ **extends** a state $g : D(\ell_2) \to \mathbb{C}$ if they agree on $D(\ell_2)$, i.e. $f \upharpoonright D(\ell_2) = g$.

The concept of a state is closedly related to quantum mechanics, where a system has a C^*-algebra (the "algebra of observables"), and a physical quantity that can be measured (an "observable") corresponds to a self-adjoint operator T in that algebra. There might be uncertianties when we measure an observable quantity of a state s, but the expected outcome of measuring T in s is $s(T)$.[46]

In the 1950's, Kadison and Singer were concerned about the functional analytic foundation of Dirac's work in quantum mechanics. Dirac assumed that each pure state on a maximal abelian C^*-subalgebra $X \subseteq B(\ell_2)$ has a unique state extension to $B(\ell_2)$, but Kadison and Singer gave some counterexample $X \neq D(\ell_2)$ such that the extension is not unique. Whether one has uniqueness for the case $X = D(\ell_2)$ remained open for decades. This is called the **Kadison–Singer problem**.[47]

Theorem 3.3 (Kadison–Singer problem)**.** Every pure state $f : D(\ell_2) \to \mathbb{C}$ extends uniquely to a state $f' : B(\ell_2) \to \mathbb{C}$.

Existence is clear: given $T \in B(\ell_2)$, we can take its diagonal part, and then apply f, and we can check this defines a state $B(\ell_2) \to \mathbb{C}$. Henceforth we focus on the uniqueness.

It is known that the Anderson paving conjecture below is equivalent to Theorem 3.3.[48] In Section 3.1, we shall present the proof that Anderson paving conjecture implies Kadison–Singer. In Sections 3.2 and 3.3, we shall prove the paving conjecture using interlacing families.

Aim 3.4 (Anderson paving conjecture)**.** [49] For every $\varepsilon > 0$, there is an $r \in \mathbb{Z}^+$ such that:

For every $n \times n$ self-adjoint complex matrix T whose diagonal entries are all 0, we can partition the indices in $[n]$ into r sets $S_1, S_2, \cdots, S_r$ such that $\|P_{S_i} T P_{S_i}\| \leq \varepsilon \|T\|$ for all i, where P_{S_i} is the orthogonal projection to the indices in S_i.

(In other words, $P_{S_i} T P_{S_i}$ is the submatrix of T that is the intersection of rows and columns whose indices are in S_i. We are "paving" the diagonal of T by smaller submatrices.)

3.1 From Paving Conjecture to Kadison–Singer Problem

In this subsection, we assume Aim 3.4 and prove Theorem 3.3. Most results in this subsection are in fact bidirectional, but we shall just include the proof that allows us to go from the paving conjecture to the Kadison–Singer problem.

The first step is a compactness argument that allows us to go from finite dimension (Aim 3.4) to infinite dimension, since Theorem 3.3 is about infinite dimensional operators.

[46] Bog+90, pp. 233–234.
[47] Cas+06, p.2.
[48] Har13, Theorems 5.11 and 6.1.
[49] MSS15b, Conjecture 1.3.

Lemma 3.5 (Compactness argument)**.** [50] (We have assumed Aim 3.4.) For all $\varepsilon > 0$, there is an $r \in \mathbb{Z}^+$ such that for every self-adjoint $T \in B(\ell_2)$ with zero diagonal, we can partition $\mathbb{Z}^+$ into r sets $S_1, \cdots, S_r$ such that $\|P_{S_i} T P_{S_i}\| \le \varepsilon \|T\|$ for all i.

Proof. Fix $\varepsilon > 0$. Consider the top-left n by n submatrix of T, denoted by T_n. By assumption, there is $r \in \mathbb{Z}^+$ such that for every n, there is a partition $S_1^n, \cdots, S_r^n$ of $[n]$, such that $\left\|P_{S_i^n} T_n P_{S_i^n}\right\| \le \varepsilon \|T_n\|$.

Think of the partition as a function $c^n : [n] \to [r]$ with $c^n(m) = i$ if $m \in S_i^n$. There is an infinite subset $A_1 \subseteq \mathbb{Z}^+$ such that $c^n(1)$ is constant for all $n \in A_1$. Similarly there is an infinite subset $A_2 \subseteq A_1$ such that $c^n(2)$ is constant for all $n \in A_2$, and so on. We therefore have a decreasing family $A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots$ with all A_k infinite and for all $n \in A_k$, c^n agrees on $[k]$, so we can define $c : \mathbb{Z}^+ \to [r]$ by $c(k) = c^n(k)$ for any $n \in A_k$, and this gives a partition $S_i = \{k : c(k) = i\}$ of $\mathbb{Z}^+$.

Consider the set $S_{i,m}$ of first m elements in S_i. If the largest of them is k, then $S_{i,m} \subseteq S_i^n$ for any $n \in A_k$, so

$$\left\|P_{S_{i,m}} T P_{S_{i,m}}\right\| \le \left\|P_{S_i^n} T_n P_{S_i^n}\right\| \le \varepsilon \|T_n\| \le \varepsilon \|T\|.$$

But we also know that when $m \to \infty$, $\left\|P_{S_{i,m}} T P_{S_{i,m}}\right\| \to \|P_{S_i} T P_{S_i}\|$ from below, so $\|P_{S_i} T P_{S_i}\| \le \varepsilon \|T\|$. □

Before moving on, we shall say very roughly the reason why the Kadison–Singer problem can be rewritten into something like Aim 3.4. We have already seen that any state $f : D(\ell_2) \to \mathbb{C}$ has an extension $f' : B(\ell_2) \to \mathbb{C}$ by first taking the diagonal part of $T \in B(\ell_2)$ and then apply f. So if we split T into a sum of the diagonal part D and the other part T', then to show uniqueness, we need to prove $f'(T) = f(D)$, which by linearity is equivalent to $f'(T') = 0$. So we can focus on operators T' with zero diagonal.

Now the partition in Lemma 3.5 comes in. For the given T', we have a partition of $\mathbb{Z}^+$ into finitely many parts, so one of the S_i is "large", in the sense that every extension $f' : B(\ell_2) \to \mathbb{C}$ of $f : D(\ell_2) \to \mathbb{C}$ only cares about the submatrix $P_{S_i} T' P_{S_i}$ but not the other entries of T', so that $f'(P_{S_i} T' P_{S_i}) = f'(T')$ (Lemma 3.11). But we can make $\|P_{S_i} T' P_{S_i}\|$ arbitrarily small, so by continuity of f', $f'(P_{S_i} T' P_{S_i})$ is arbitrarily small, so $f'(T') = 0$ (Lemma 3.12).

To fill in the gaps above, we need to understand what the pure states on $D(\ell_2)$ are, and to make sure that the "large" S_i has the desired property (that $P_{S_i} T' P_{S_i}$ determines the value of $f'(T')$). The pure states on $D(\ell_2)$ can be described using ultrafilters, and "large" means being in the ultrafilter.

Definition 3.6 (Ultrafilters)**.** [51] An **ultrafilter** on $\mathbb{Z}^+$ is a family $\mathcal{U}$ of subsets of $\mathbb{Z}^+$ such that:

(1) $\varnothing \notin \mathcal{U}$, $\mathbb{Z}^+ \in \mathcal{U}$;

(2) If $A \in \mathcal{U}$ and $A \subseteq B$, then $B \in \mathcal{U}$;

(3) If $A, B \in \mathcal{U}$, then $A \cap B \in \mathcal{U}$;

(4) For every $A \subseteq \mathbb{Z}^+$, either $A \in \mathcal{U}$ or $\mathbb{Z}^+ \setminus A \in \mathcal{U}$.

Equivalently, (4) can be replaced by "If $A \cup B \in \mathcal{U}$, then either $A \in \mathcal{U}$ or $B \in \mathcal{U}$", or its natural generalisation to n sets $A_1, \cdots, A_n$.

[50] Har13, Theorem 6.1 and Claim 6.2.
[51] Har13, Definitions 3.1 and 3.3.

We shall show that there is a one-to-one correspondence between pure states $f : D(\ell_2) \to \mathbb{C}$ and ultrafilters $\mathcal{U}$ over $\mathbb{Z}^+$. One direction is the following. The other is in Lemma 3.10.

Lemma 3.7 (Pure states must be $f_{\mathcal{U}}$)**.** [52] If $f : D(\ell_2) \to \mathbb{C}$ is a pure state, then $\{A \subseteq \mathbb{Z}^+ : f(P_A) = 1\}$ is an ultrafilter, where P_A is the diagonal projection as before.

Moreover, the map $f \mapsto \{A \subseteq \mathbb{Z}^+ : f(P_A) = 1\}$ is injective.

Proof. Suppose f is a pure state, and let $\mathcal{U} = \{A \subseteq \mathbb{Z}^+ : f(P_A) = 1\}$. We check the definition of an ultrafilter:

(1) $f(P_\varnothing) = f(0) = 0$, and $f(P_{\mathbb{Z}^+}) = f(I) = 1$.

(2) If $f(P_A) = 1$ and $B \supseteq A$, then $P_A \preceq P_B \preceq I$, so $1 = f(P_A) \le f(P_B) \le f(I) = 1$, so $f(P_B) = 1$.

(3) If $f(P_A) = f(P_B) = 1$, then from $P_A + P_B = I + P_{A\cap B}$ and linearity, we have $f(P_{A\cap B}) = f(P_A) + f(P_B) - f(I) = 1$.

(4) Since $P_A + P_{\mathbb{Z}^+\setminus A} = I$, by positivity we have $0 \le f(P_A) \le 1$, $0 \le f(P_{\mathbb{Z}^+\setminus A}) \le 1$, and by linearity they sum to 1. We need to show that one of them is 1. Suppose not, then we have $f(P_A) = \alpha \in (0, 1)$. Let $g(X) = \frac{1}{\alpha} f(P_A X)$ and $h(X) = \frac{1}{1-\alpha} f(P_{\mathbb{Z}^+\setminus A} X)$. Both g and h are states, and $f(X) = f(P_A X) + f(P_{\mathbb{Z}^+\setminus A} X) = \alpha g(X) + (1-\alpha)h(X)$, but f is pure, so $f = g = h$. However,

$$h(P_A) = \frac{1}{1-\alpha} f(P_{\mathbb{Z}^+\setminus A} P_A) = \frac{1}{1-\alpha} f(0) = 0 \neq 1 = g(P_A),$$

contradiction.

The map is injective: If f, g are two pure states $D(\ell_2) \to \mathbb{C}$ both mapped to $\mathcal{U}$, then we have seen above that f, g must map diagonal projections P_A to 1 if $A \in \mathcal{U}$, and 0 if $A \notin \mathcal{U}$, so they agree on all diagonal projections. However, $D(\ell_2)$ is the closed linear span of the diagonal projections, so f and g agree on $D(\ell_2)$. □

In the finite dimensional setting, we would have that $\mathcal{U}$ is an ultrafilter over a finite set $[n]$, which implies $\mathcal{U}$ is **principal**, i.e. there is some $k \in [n]$ such that $A \in \mathcal{U} \iff k \in A$, and so the pure states just send the n by n matrix X to X_{kk}. However, in the infinite dimensional setting, we need to consider the non-principal ultrafilters, giving rise to other pure states.

To see the other direction of the correspondence (i.e. given an ultrafilter $\mathcal{U}$ we can define a pure state $f_{\mathcal{U}}$), we need the help of $\mathcal{U}$-limits, which allows us to say something like "$f_{\mathcal{U}}$ sends $T \in D(\ell_2)$ to x if the majority of the diagonal entries T_{ii} are close to x".

Definition 3.8 ($\mathcal{U}$-limits)**.** [53] Given a sequence (x_n) in $\mathbb{C}$, we say $\lim_{\mathcal{U}} x_n = x$ if for all $\varepsilon > 0$, the set $\{n : |x_n - x| < \varepsilon\} \in \mathcal{U}$.

The usual limit has the **cofinite filter** $\mathcal{F} = \{A : \mathbb{Z}^+ \setminus A \text{ is finite}\}$ in place of $\mathcal{U}$, but $\mathcal{F}$ is not ultra. If we have an ultrafilter $\mathcal{U} \supset \mathcal{F}$, then more sets are in $\mathcal{U}$ than in $\mathcal{F}$, so it becomes easier to have $\{n : |x_n - x| < \varepsilon\} \in \mathcal{U}$, so more sequences have $\mathcal{U}$-limits, so this generalises the usual limit.

[52] Har13, Theorem 4.3.

[53] Har13, Definition 3.15.

Lemma 3.9 (Existence and uniqueness of $\mathcal{U}$-limits)**.** [54] For any sequence (x_n) in $\mathbb{C}$ and ultrafilter $\mathcal{U}$, there is at most one x such that $\lim_{\mathcal{U}} x_n = x$. If moreover (x_n) is bounded, then for any ultrafilter $\mathcal{U}$ on $\mathbb{Z}^+$, $\lim_{\mathcal{U}} x_n$ exists.

Proof. If $x \neq y$, then for $0 < \varepsilon < \frac{|x-y|}{2}$, the sets $\{n : |x_n - x| < \varepsilon\}$ and $\{n : |x_n - y| < \varepsilon\}$ are disjoint, so they cannot both be in $\mathcal{U}$, so (x_n) cannot have two distinct $\mathcal{U}$-limits.

Now suppose (x_n) is bounded and has no $\mathcal{U}$-limit. Then every $y \in \mathbb{C}$ is not a $\mathcal{U}$-limit, so there is an $\varepsilon_y > 0$ such that the set $A_y = \{n : |x_n - y| < \varepsilon_y\} \notin \mathcal{U}$. Since (x_n) lies in a compact region C and the open balls $B_y = \{x : |x - y| < \varepsilon_y\}$ for $y \in C$ form an open cover of C, there is a finite subcover, say

$$C \subseteq B_{y_1} \cup B_{y_2} \cup \cdots \cup B_{y_k}.$$

For every $n \in \mathbb{Z}^+$, $x_n \in C$, so $x_n \in B_{y_i}$ for some i, which gives $n \in A_{y_i}$. So $A_{y_1} \cup A_{y_2} \cup \cdots \cup A_{y_k} = \mathbb{Z}^+$, but this says one of the $A_{y_i} \in \mathcal{U}$, contradiction. ☐

Similar to the usual limit, $\lim_{\mathcal{U}}$ has properties such as $\lim_{\mathcal{U}} x_n + \lim_{\mathcal{U}} y_n = \lim_{\mathcal{U}}(x_n + y_n)$[55] and $\lim_{\mathcal{U}} x_n y_n = \lim_{\mathcal{U}} x_n \lim_{\mathcal{U}} y_n$[56]. The usual proof can be translated to the proof for $\mathcal{U}$-limits by changing "Take $N = \max(N_1, N_2) \in \mathbb{Z}^+$" to "Take $A = A_1 \cap A_2 \in \mathcal{U}$". Using this, we can define a state $f : D(\ell_2) \to \mathbb{C}$ from an ultrafilter $\mathcal{U}$.

Lemma 3.10 (Characterisation of pure states on $D(\ell_2)$)**.** [57] For every ultrafilter $\mathcal{U}$ on $\mathbb{Z}^+$, the map $f_{\mathcal{U}} : D(\ell_2) \to \mathbb{C}$ defined by $f_{\mathcal{U}}(X) = \lim_{\mathcal{U}} X_{nn}$ (where (X_{nn}) is the sequence of diagonal entries of X) satisfies $\{A : f_{\mathcal{U}}(P_A) = 1\} = \mathcal{U}$, and $f_{\mathcal{U}}$ is a pure state.

Therefore the map $f \mapsto \{A \subseteq \mathbb{Z}^+ : f(P_A) = 1\}$ as in Lemma 3.7 is also surjective, and is a one-to-one correspondence between ultrafilters on $\mathbb{Z}^+$ and pure states on $D(\ell_2)$.

Proof. If $X \in D(\ell_2)$, then (X_{nn}) is a bounded sequence, so by Lemma 3.9, $\lim_{\mathcal{U}} X_{nn}$ exists and is unique. For $A \subseteq \mathbb{Z}^+$, $(P_A)_{nn}$ is the indicator $1_{n\in A}$, so $f_{\mathcal{U}}(P_A) = \lim_{\mathcal{U}} 1_{n\in A} \in \{0, 1\}$, and

$$\lim_{\mathcal{U}} 1_{n\in A} = 1 \iff \{n : 1_{n\in A} = 1\} \in \mathcal{U} \iff A \in \mathcal{U}.$$

so $\{A : f_{\mathcal{U}}(P_A) = 1\} = \mathcal{U}$.

We can check $f_{\mathcal{U}}$ is linear, continuous, and satisfies $f_{\mathcal{U}}(I) = 1$ and $f_{\mathcal{U}}(X) \geq 0$ for $X \succeq 0$, so it is a state. Suppose g, h are states with $f_{\mathcal{U}} = \alpha g + (1 - \alpha)h$ for some $\alpha \in (0, 1)$. We have seen $g(P_A) \in [0, 1]$ for all $A \subseteq \mathbb{Z}^+$ and similarly for h, but $f_{\mathcal{U}}(P_A) \in \{0, 1\}$ is already at an endpoint of the interval, so $g(P_A) = h(P_A) = f_{\mathcal{U}}(P_A)$. Again $D(\ell_2)$ is the closed linear span of the diagonal projections, so we conclude $g(X) = h(X) = f_{\mathcal{U}}(X)$ for all $X \in D(\ell_2)$. So $f_{\mathcal{U}}$ is pure. ☐

The pure state $f_{\mathcal{U}}$ only care about the "majority" of the entries of $X \in D(\ell_2)$. More precisely, if $A \in \mathcal{U}$, then $f_{\mathcal{U}}(P_{\mathbb{Z}^+\setminus A}) = 0$, so for all $B \subseteq \mathbb{Z}^+$, $f_{\mathcal{U}}(P_{\mathbb{Z}^+\setminus A} P_B) = f_{\mathcal{U}}(P_{(\mathbb{Z}^+\setminus A)\cap B}) = 0$ since $0 \preceq P_{(\mathbb{Z}^+\setminus A)\cap B} \preceq P_{\mathbb{Z}^+\setminus A}$. Therefore by linearity, $f_{\mathcal{U}}(P_{\mathbb{Z}^+\setminus A} X) = 0$ for all $X \in D(\ell_2)$, and

$$f_{\mathcal{U}}(X) = f_{\mathcal{U}}(P_A X) + f_{\mathcal{U}}(P_{\mathbb{Z}^+\setminus A} X) = f_{\mathcal{U}}(P_A X).$$

This property also holds for extensions of $f_{\mathcal{U}}$.

[54] Har13, Claims 3.16 and 3.18.
[55] Har13, Claim 3.20.
[56] Har13, Claim 3.21.
[57] Har13, Theorem 4.2.

Lemma 3.11. [58] If $f : D(\ell_2) \to \mathbb{C}$ is a pure state and $f' : B(\ell_2) \to \mathbb{C}$ extends f, then for any $A \in \mathcal{U}$, $f'(P_A X) = f'(X)$ for all $X \in B(\ell_2)$. Similarly $f'(XP_A) = f'(X)$ for all $X \in B(\ell_2)$.

Proof. Since f' is linear and $f'(X) \geq 0$ for all $X \succeq 0$, the map $(X, Y) \mapsto f'(X^*Y)$ is a positive semi-definite sesquilinear form (inner product except $f'(X^*X)$ can be 0 for some $X \neq 0$), so by Cauchy–Schwarz inequality,

$$\left| f'(P^*_{\mathbb{Z}^+ \setminus A} Y) \right|^2 \leq f'(P^*_{\mathbb{Z}^+ \setminus A} P_{\mathbb{Z}^+ \setminus A}) f'(Y^*Y) = f(P_{\mathbb{Z}^+ \setminus A}) f'(Y^*Y) = 0,$$

so $f'(P_{\mathbb{Z}^+ \setminus A} Y) = 0$ and $f'(P_A Y) = f'(Y) - f'(P_{\mathbb{Z}^+ \setminus A} Y) = f'(Y)$. □

Lemma 3.12. [59] If $f : D(\ell_2) \to \mathbb{C}$ is a pure state, and $f' : B(\ell_2) \to \mathbb{C}$ is a state extending f, then for any $T \in B(\ell_2)$ self-adjoint with zero diagonal, $f'(T) = 0$.

Proof. By Lemma 3.10, $f = f_{\mathcal{U}}$ for some ultrafilter $\mathcal{U}$. Fix $\varepsilon > 0$ and $T \in B(\ell_2)$. By Lemma 3.5 with ε' sufficiently small, there is a partition $S_1, \cdots, S_r$ of $\mathbb{Z}^+$ such that $\|P_{S_i} T P_{S_i}\| \leq \varepsilon' \|T\| \leq \varepsilon$ for all $i \in [r]$. Since $S_1 \cup S_2 \cup \cdots \cup S_r = \mathbb{Z}^+ \in \mathcal{U}$, some $S_i \in \mathcal{U}$. By Lemma 3.11, $f'(T) = f'(P_{S_i} T) = f'(P_{S_i} T P_{S_i})$.

From $\|P_{S_i} T P_{S_i}\| \leq \varepsilon$, we know

$$-\varepsilon P_{S_i} \preceq P_{S_i} T P_{S_i} \preceq \varepsilon P_{S_i},$$

so

$$-\varepsilon f'(P_{S_i}) \leq f'(P_{S_i} T P_{S_i}) \leq \varepsilon f'(P_{S_i}),$$

i.e. $|f'(P_{S_i} T P_{S_i})| \leq \varepsilon$, so $|f'(T)| \leq \varepsilon$. This holds for all $\varepsilon > 0$, so $f'(T) = 0$. □

Lemma 3.12 is enough to show that the extension is unique to all $T \in B(\ell_2)$.

Proof of Theorem 3.3. Given $T \in B(\ell^2)$, we can decompose it as $T = T_1 + iT_2$ where $T_1 = \frac{1}{2}(T + T^*)$ and $T_2 = \frac{-i}{2}(T - T^*)$ are both self-adjoint. Since $f'(T) = f'(T_1) + if'(T_2)$, f' is uniquely determined by its values at the self-adjoint operators. Henceforth assume T is self-adjoint.

Write $T = D + T_0$ where $D \in D(\ell_2)$ is the diagonal part of T, and T_0 is self-adjoint with zero diagonal. Since f' extends f, we have $f'(D) = f(D)$, and $f'(T_0) = 0$ by Lemma 3.12. Hence $f'(T) = f'(D) + f'(T_0) = f(D)$, and the extension f' is uniquely determined by f. □

3.2 Identifying an Interlacing Family

In this section we shall see how we can reduce Aim 3.4 to a statement in terms of random vectors that would allow us to apply Lemma 1.23. We start with some linear algebra tricks.

Lemma 3.13. [60] The following are equivalent:

(1) [61] (Anderson paving conjecture, diagonal 0 self-adjoint, Aim 3.4)

For every $\varepsilon > 0$, there is an $r \in \mathbb{Z}^+$ such that:

For every n by n self-adjoint complex matrix T whose diagonal entries are all 0, we can partition $[n]$ into r sets $S_1, S_2, \cdots, S_r$ such that $\|P_{S_i} T P_{S_i}\| \leq \varepsilon \|T\|$ for all i.

[58] Har13, Corollary C.19.
[59] Har13, Lemma 5.9.
[60] Har13, Theorem 6.3.
[61] MSS15b, Theorem 6.1.

(2) [62] (Anderson paving conjecture, diagonal $\frac{1}{2}$ orthogonal projection)

For every $\varepsilon > 0$, there is an $r \in \mathbb{Z}^+$ such that:

For every n by n orthogonal projection matrix Q (i.e. $Q^* = Q$ and $Q^2 = Q$) whose diagonal entries are all $\frac{1}{2}$, we can partition $[n]$ into r sets $S_1, S_2, \cdots, S_r$ such that $\|P_{S_i} Q P_{S_i}\| \le \frac{1+\varepsilon}{2}\|Q\| = \frac{1+\varepsilon}{2}$ for all i.

Proof. (1) $\implies$ (2): Given ε, let r be given by (1). If Q is an orthogonal projection with diagonal $\frac{1}{2}$, then $T = 2Q - I$ is zero diagonal self-adjoint. Since the eigenvalues of Q are 0 or 1 (and not all 0), the eigenvalues of T are ± 1, so $\|Q\| = 1$ and $\|T\| = 1$. By (1), there is a partition $S_1, \cdots, S_r$ of $[n]$ such that $\|P_{S_i} T P_{S_i}\| \le \varepsilon$, so

$$\|P_{S_i} Q P_{S_i}\| = \left\| P_{S_i}\left(\frac{I+T}{2}\right) P_{S_i}\right\| \le \left\|\frac{P_{S_i}}{2}\right\| + \left\|\frac{P_{S_i} T P_{S_i}}{2}\right\| \le \frac{1+\varepsilon}{2}.$$

(2) $\implies$ (1): Given ε, let r be given by (2). Let T be a diagonal 0 self-adjoint $n \times n$ matrix. By rescaling, we may assume $\|T\| = 1$, so that $I - T^2$ is positive semi-definite self-adjoint and has a square root. Let Q be the $2n$ by $2n$ matrix

$$\frac{1}{2} I_{2n} + \frac{1}{2}\begin{pmatrix} T & \sqrt{I_n - T^2} \\ \sqrt{I_n - T^2} & -T \end{pmatrix}.$$

Then all diagonal entries of Q are $\frac{1}{2}$, and Q is self-adjoint. Also,

$$Q^2 = \frac{1}{4} I_{2n} + \frac{1}{2}\begin{pmatrix} T & \sqrt{I_n - T^2} \\ \sqrt{I_n - T^2} & -T \end{pmatrix} + \frac{1}{4}\begin{pmatrix} I_n & 0 \\ 0 & I_n \end{pmatrix} = Q.$$

So by (2), there is a partition $S_1, \cdots, S_r$ of $[2n]$ such that $\|P_{S_i} Q P_{S_i}\| \le \frac{1+\varepsilon}{2}$. Restricting to the top-left n by n block, we have a partition $S'_1, \cdots, S'_r$ of $[n]$ given by $S'_i = S_i \cap [n]$ such that

$$\left\| P_{S'_i} \frac{I_n + T}{2} P_{S'_i} \right\| \le \frac{1+\varepsilon}{2},$$

so the largest eigenvalue $\lambda_1\left(P_{S'_i} T P_{S'_i}\right) \le \varepsilon$.

Similarly, restricting to the bottom-right n by n block, we have another partition $S''_1, \cdots, S''_r$ of $[n]$ given by $S''_i = \{a - n : n+1 \le a \le 2n, a \in S_i\}$ such that

$$\left\| P_{S'_i} \frac{I_n - T}{2} P_{S'_i} \right\| \le \frac{1+\varepsilon}{2}.$$

So the smallest eigenvalue $\lambda_n\left(P_{S''_i} T P_{S''_i}\right) \ge -\varepsilon$.

Now we can take the coarsest common refinement $R_{ab} = S'_a \cap S''_b$ ($a, b \in [r]$), which is a partition of $[n]$ into r^2 sets such that $\|P_{R_{ab}} T P_{R_{ab}}\| \le \varepsilon$ for all a, b. ☐

Consider an n by n orthogonal projection Q with diagonal $\frac{1}{2}$. We have $Q_{ij} = \boldsymbol{e}_i^* Q \boldsymbol{e}_j = \boldsymbol{e}_i^* Q^* Q \boldsymbol{e}_j = \boldsymbol{u}_i^* \boldsymbol{u}_j$ if we write $\boldsymbol{u}_i = Q\boldsymbol{e}_i$, so in particular, $\|\boldsymbol{u}_i\|^2 = \boldsymbol{u}_i^* \boldsymbol{u}_i = Q_{ii} = \frac{1}{2}$. Moreover,

$$\|P_{S_k} Q P_{S_k}\| = \left\| (\boldsymbol{u}_i^* \boldsymbol{u}_j)_{i,j \in S_k} \right\| \le \left\| \sum_{i \in S_k} \boldsymbol{u}_i \boldsymbol{u}_i^* \right\|$$

[62] MSS15b, Theorem 6.2.

(if $P_{S_k} Q P_{S_k} \boldsymbol{v} = \lambda \boldsymbol{v}$ then $\sum_{i \in S_k} \boldsymbol{u}_i \boldsymbol{u}_i^*$ also acts on $\sum_{j \in S_k} v_j \boldsymbol{u}_j$ as multiplication by λ). Also, $\sum_i \boldsymbol{u}_i \boldsymbol{u}_i^* = \sum_i Q \boldsymbol{e}_i \boldsymbol{e}_i^* Q^* = QIQ^* = Q$.

Since all $\boldsymbol{u}_i$ live in the image of Q (which is an $\frac{n}{2}$-dimensional subspace because $\operatorname{tr} Q = \frac{n}{2}$, and Q acts as the identity on this subspace), we can instead think of $\boldsymbol{u}_i$ as vectors from $\mathbb{C}^{n/2}$ so that $\sum_{i=1}^n \boldsymbol{u}_i \boldsymbol{u}_i^* = I_{n/2}$ and still have $Q_{ij} = \boldsymbol{u}_i^* \boldsymbol{u}_j$ for $i, j \in [n]$. Summarising, we have reduced Aim 3.4 to the following:

Aim 3.14. [63] For every $\varepsilon > 0$, there is an $r \in \mathbb{Z}^+$ such that if $\sum_{i=1}^n \boldsymbol{u}_i \boldsymbol{u}_i^* = I_{n/2}$ and $\|\boldsymbol{u}_i\|^2 = \frac{1}{2}$, then there is a partition $S_1, \cdots, S_r$ of $[n]$ such that for all $k \in [r]$,

$$\left\| \sum_{i \in S_k} \boldsymbol{u}_i \boldsymbol{u}_i^* \right\| \leq \frac{1+\varepsilon}{2}.$$

Since $\sum_{i \in S_k} \boldsymbol{u}_i \boldsymbol{u}_i^*$ is positive semi-definite, the norm is just the largest eigenvalue $\lambda_1(\sum_{i \in S_k} \boldsymbol{u}_i \boldsymbol{u}_i^*)$, which we might be able to control using Lemma 1.11 once we identify an interlacing family.

Indeed, a random partition can be thought of as a random assignment

$$\boldsymbol{v}_i \leftarrow \left\{ \begin{pmatrix} \boldsymbol{u}_i \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{u}_i \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \end{pmatrix}, \cdots, \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{u}_i \end{pmatrix} \right\}.$$

This fits into the framework of Lemma 1.23, which gives an interlacing family. Say $\boldsymbol{v}_i$ takes each of the r possible values with equal probability, then

$$\begin{aligned} \sum_{i=1}^n \mathbb{E}(\boldsymbol{v}_i \boldsymbol{v}_i^*) = \sum_{i=1}^n \frac{1}{r} \begin{pmatrix} \boldsymbol{u}_i \boldsymbol{u}_i^* & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{u}_i \boldsymbol{u}_i^* & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{u}_i \boldsymbol{u}_i^* \end{pmatrix} &= \frac{1}{r} \begin{pmatrix} \sum_i \boldsymbol{u}_i \boldsymbol{u}_i^* & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \sum_i \boldsymbol{u}_i \boldsymbol{u}_i^* & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \sum_i \boldsymbol{u}_i \boldsymbol{u}_i^* \end{pmatrix} \\ &= \frac{1}{r} I_{rn/2}, \end{aligned}$$

and a realisation of $\sum_i \boldsymbol{v}_i \boldsymbol{v}_i^*$ is of the form

$$\begin{pmatrix} \sum_{i \in S_1} \boldsymbol{u}_i \boldsymbol{u}_i^* & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \sum_{i \in S_2} \boldsymbol{u}_i \boldsymbol{u}_i^* & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \sum_{i \in S_r} \boldsymbol{u}_i \boldsymbol{u}_i^* \end{pmatrix},$$

where $S_1, \cdots, S_r$ form some partition of $[n]$, so $\|\sum_i \boldsymbol{v}_i \boldsymbol{v}_i^*\| = \max_{j \in [r]} \left\| \sum_{i \in S_j} \boldsymbol{u}_i \boldsymbol{u}_i^* \right\|$ and it suffices to show that $\|\sum_i \boldsymbol{v}_i \boldsymbol{v}_i^*\| \leq \frac{1+\varepsilon}{2}$ with non-zero probability.

Therefore, we have reduced Aim 3.14 to the following.

Aim 3.15. [64] For every $\varepsilon > 0$, there is an $r \in \mathbb{Z}^+$ such that for all $n \in 2\mathbb{Z}^+$, if independent random vectors $\boldsymbol{v}_i \in \mathbb{C}^{rn/2}$ (taking finitely many possible values) satisfies $\sum_{i=1}^n \mathbb{E}(\boldsymbol{v}_i \boldsymbol{v}_i^*) = \frac{1}{r} I_{rn/2}$ and $\|\boldsymbol{v}_i\|^2 = \frac{1}{2}$, then with non-zero probability we have

$$\left\| \sum_{i=1}^n \boldsymbol{v}_i \boldsymbol{v}_i^* \right\| \leq \frac{1+\varepsilon}{2}.$$

[63] MSS15b, Corollary 1.5.
[64] MSS15b, Theorem 1.4.

We may recognise, by Lemma 1.23, that some suitable family of the mixed characteristic polynomials arising from these random vectors is an interlacing family. Aim 3.15 (in a more general form) will be shown in the next subsection by establishing an upper bound on the largest root of $\mathbb{E}\chi(\sum \boldsymbol{v}_i\boldsymbol{v}_i^*)$.

3.3 Bounding the Roots

By Lemma 1.23, any bound on the roots of the mixed characteristic polynomial $\mathbb{E}\chi(\sum \boldsymbol{v}_i\boldsymbol{v}_i^*)$ gives a bound in Aim 3.15.

We can rescale (write $A_i = r\mathbb{E}(\boldsymbol{v}_i\boldsymbol{v}_i^*)$) in Aim 3.15 so that we have $\sum A_i = I$ and $\operatorname{tr} A_i = r\mathbb{E}(\boldsymbol{v}_i^*\boldsymbol{v}_i) = \frac{r}{2}$, and A_i are positive semi-definite self-adjoint. Their mixed characteristic polynomial is then

$$\begin{aligned}
\mu[A_1, \cdots, A_n](x) &= \left(\prod_{i=1}^n (1 - \partial_{z_i})\right) \det\left(xI + \sum_{i=1}^n z_i A_i\right)\Bigg|_{z_1=\cdots=z_n=0} \\
&= \left(\prod_{i=1}^n (1 - \partial_{z_i})\right) \det\left(\sum_{i=1}^n (x + z_i) A_i\right)\Bigg|_{z_1=\cdots=z_n=0} \\
&= \left(\prod_{i=1}^n (1 - \partial_{z_i})\right) \det\left(\sum_{i=1}^n z_i A_i\right)\Bigg|_{z_1=\cdots=z_n=x}.
\end{aligned}$$

After the rescaling, we need to prove a bound $\lambda_1(\mu[A_1, \cdots, A_n]) < \frac{1+\varepsilon}{2}r$. Consider the polynomial

$$p(z_1, \cdots, z_n) = \left(\prod_{i=1}^n (1 - \partial_{z_i})\right) \det\left(\sum_{i=1}^n z_i A_i\right),$$

then $\mu[A_1, \cdots, A_n](x) = p(x, x, \cdots, x)$, so to show $\lambda_1(\mu[A_1, \cdots, A_n]) < \frac{1+\varepsilon}{2}r$, it suffices to show the stronger statement that $p(z_1, \cdots, z_n)$ is non-zero when $z_i \geq \frac{1+\varepsilon}{2}r$ for all i.

Definition 3.16 (Above)**.** [65] Let $p(z_1, \cdots, z_n) \in \mathbb{R}[z_1, \cdots, z_n]$. We say $\boldsymbol{a} \in \mathbb{R}^n$ is **above** all zeros of p if $p(z_1, \cdots, z_n) > 0$ whenever $z_i \geq a_i$ for all i.

We would like to find some $M = \frac{r}{2} + o(r) < \frac{1+\varepsilon}{2}r$ (for large r) such that $(M, M, \cdots, M)$ is above all zeros of p. We already know that $(\varepsilon, \varepsilon, \cdots, \varepsilon)$ is above all zeros of $\det(\sum_{i=1}^n z_i A_i)$ for any $\varepsilon > 0$. (If all $z_i \geq \varepsilon$, then $\sum_i A_i = I \succ 0$ and each $A_i \succeq 0$, so $\sum_i z_i A_i$ is positive definite, so the determinant is non-zero.)

But $p(z_1, \cdots, z_n) = (\prod_i (1 - \partial_{z_i})) \det(\sum_i z_i A_i)$, so we need to understand how $(1 - \partial_{z_i})$ affects the position of the zeros. First we look at a failed attempt. We might try to use to following

Lemma 3.17. [66] If $f(z) \in \mathbb{R}[z]$ has degree d and is real-rooted, and a is above all roots of f, then $a + d$ is above all roots of $(1 - \partial_z)f$. In other words, λ_1 goes up by at most d when we apply $1 - \partial_z$.

Proof. For $x > a > \lambda_1(f)$, we have $f(x) \neq 0$, so $(1 - \partial_z)f(x) = 0 \iff f(x) - f'(x) = 0 \iff \frac{f'(x)}{f(x)} = 1$. So if x is a root to $(1 - \partial_z)f$, then

$$\sum_i \frac{1}{x - \lambda_i(f)} = \frac{f'(x)}{f(x)} = 1,$$

[65] MSS15b, Definition 5.3.

[66] Special case of Mar66, Corollary 18.2a.

so

$$\frac{d}{x-\lambda_1(f)} \geq \sum_i \frac{1}{x-\lambda_i(f)} = 1,$$

so $x \leq \lambda_1(f) + d < a + d$. □

However, this is not enough. Since the A_i's are $\frac{rn}{2}$ by $\frac{rn}{2}$ matrices, $\det(\sum_i z_i A_i)$ has degree $\frac{rn}{2}$, so if we simply apply Lemma 3.17 in each coordinate, from $(\varepsilon, \varepsilon, \cdots, \varepsilon)$ is above all zeros of $\det(\sum_{i=1}^n z_i A_i)$ we can only conclude that $(\varepsilon + \frac{rn}{2}, \varepsilon + \frac{rn}{2}, \cdots, \varepsilon + \frac{rn}{2})$ is above all zeros of $p(z_1, \cdots, z_n)$. This is bad because we want a bound $\frac{r}{2} + o(r)$ but have got an extra factor of n.

Some reasons that this fails are:

(1) We did not take into account the condition $\operatorname{tr} A_i \leq \frac{r}{2}$.

(2) The bound in Lemma 3.17 (the "$+d$") is not tight in most cases. As we can see in the proof, equality holds if and only if $\lambda_i(f) = \lambda_1(f)$ for all i, i.e. all roots are equal.

If we know that the roots of f are not close to each other, or that the initial bound a is already very far from the largest root, then we should somehow be able to obtain an increment that is less than $+d$. This leads to more careful consideration of the following quantity, which has appeared in the proof above.

Definition 3.18 (Barrier function)**.** [67][68] Let $f(z_1, \cdots, z_n) \in \mathbb{R}[z_1, \cdots, z_n]$ and assume we are in a region where f is positive. The **barrier function** or **log-derivative** in the i-th coordinate is defined as

$$\Phi_f^i = \partial_{z_i} \log f = \frac{\partial_{z_i} f}{f}$$

In the one-variable case with all roots being real, $\Phi_f^1(a) = \sum_i \frac{1}{a-\lambda_i(f)}$ measures how far $a > \lambda_1(f)$ is away from the roots $\lambda_i(f)$, $1 \leq i \leq d$. When $a \to +\infty$, $\Phi_f^1(f) \to 0$, but when $a \to \lambda_1(f)^+$, there is a barrier $\Phi_f^1(f) \to +\infty$.

Note that $\det(\sum_i z_i A_i)$ and $p(z_1, \cdots, z_n)$ are both real stable by Example 1.14 and Lemma 1.16. For real stable polynomials f, there are a few convexity results regarding Φ_f^i that would help us control the zeros of $(1 - \partial_{z_i})f$ better than the failed attempt above.

Lemma 3.19 (Convexity lemma, 1 variable)**.** [69] If $f(z_1, \cdots, z_n)$ is a real stable polynomial, and $\boldsymbol{a} \in \mathbb{R}^n$ is above all zeros of f, then $\Phi_f^1(\boldsymbol{a}) > 0, \partial_{z_1} \Phi_f^1(\boldsymbol{a}) < 0, \partial_{z_1}^2 \Phi_f^1(\boldsymbol{a}) > 0$, i.e. Φ_f^1 is a positive decreasing convex function in z_1 when we are above all zeros of f.

Proof. Let $g(z_1) = f(z_1, a_2, \cdots, a_n)$. Since $\boldsymbol{a}$ is above all zeros of f, a_1 is above all zeros of g, and g is real-rooted by Lemma 1.15. Let $g(z_1) = C \prod_{i=1}^d (z_1 - \lambda_i(g))$, where $\lambda_i(g) < a_1$ are the real roots of g, then

$$\Phi_f^1(z_1, a_2, \cdots, a_n) = \frac{g'(z_1)}{g(z_1)} = \sum_{i=1}^d \frac{1}{z_1 - \lambda_i(g)}.$$

Hence the first derivative is $\sum_{i=1}^d \frac{-1}{(z_1-\lambda_i(g))^2}$ and the second derivative is $\sum_{i=1}^d \frac{2}{(z_1-\lambda_i(g))^3}$. Now substitute $z_1 = a_1$ to obtain the result. □

[67] MSS15b, Definition 5.4.
[68] Tao13, Section 3, before Lemma 16.
[69] Tao13, Lemma 16.

Lemma 3.20 (Convexity lemma, 2 variables)**.** [70] If $f(z_1, \cdots, z_n)$ is a real stable polynomial, and $\boldsymbol{a} \in \mathbb{R}^n$ is above all zeros of f, then $\partial_{z_2} \Phi_f^1(\boldsymbol{a}) \leq 0, \partial_{z_2}^2 \Phi_f^1(\boldsymbol{a}) \geq 0$, i.e. Φ_f^1 is a positive decreasing convex (not necessarily strictly) function in z_2 when we are above all zeros of f.

Combining with Lemma 3.19, this says for all i and j (not necessarily distinct), we have

$$\partial_{z_j} \Phi_f^i(\boldsymbol{a}) \leq 0, \quad \partial_{z_j}^2 \Phi_f^i(\boldsymbol{a}) \geq 0$$

when $\boldsymbol{a}$ is above all zeros of f.

Proof. Let $g_{z_1}(z_2) = f(z_1, z_2, a_3, \cdots, a_n)$. Let d be its degree in z_2. We may also assume $g_{z_1}(z_2)$ is irreducible (otherwise the log-derivative is just the sum of log-derivatives of the irreducible factors).

As in the proof of Lemma 3.19, g_b is real-rooted for any $b \in \mathbb{R}$. Write

$$\partial_{z_2} \log f(z_1, z_2, a_3, \cdots, a_n) = \frac{g'_{z_1}(z_2)}{g_{z_1}(z_2)} = \sum_{i=1}^{d} \frac{1}{z_2 - \lambda_i(g_{z_1})}, \tag{$\clubsuit$}$$

and

$$\partial_{z_2}^2 \log f(z_1, z_2, a_3, \cdots, a_n) = \sum_{i=1}^{d} \frac{-1}{(z_2 - \lambda_i(g_{z_1}))^2}. \tag{$\diamondsuit$}$$

Since $\partial_{z_2}\partial_{z_1} \log f(\boldsymbol{a}) = \partial_{z_1}\partial_{z_2} \log f(\boldsymbol{a})$, and similarly $\partial_{z_2}^2\partial_{z_1} \log f(\boldsymbol{a}) = \partial_{z_1}\partial_{z_2}^2 \log f(\boldsymbol{a})$, the inequalities we are required to show are that the left hand side of ($\clubsuit$) is non-increasing in z_1 and that of ($\diamondsuit$) is non-decreasing in z_1.

Since $\boldsymbol{a}$ is above all zeros of f, we have $a_2 > \lambda_i(g_{a_1})$, so it suffices to show that $\lambda_i(g_{z_1})$ is non-increasing in z_1. We would like to take ∂_{z_1}, but there are some technicalities before we know $\lambda_i(g_{z_1})$ is differentiable for most z_1.

By continuity, it suffices to prove the inequalities $\partial_{z_2} \Phi_f^1(\boldsymbol{a}) \leq 0$ and $\partial_{z_2}^2 \Phi_f^1(\boldsymbol{a}) \geq 0$ for generic $\boldsymbol{a}$, i.e. on a dense subset of $\{\boldsymbol{a} : \boldsymbol{a} \text{ above all roots of } f\}$.

Claim. *$g_b(z_2) \in \mathbb{R}[z_2]$ has d distinct real roots except for finitely many $b \in \mathbb{R}$.*

Proof of Claim. The coefficient of the highest term z_2^d in $g_{z_1}(z_2)$ is a polynomial in z_1 which only has finitely many roots, so g_b has d roots except for finitely many b.

The discriminant $\Delta(g_{z_1})$ is a polynomial in z_1. If $\Delta(g_{z_1})$ is identically zero, then the irreducible polynomial $g_{z_1}(z_2)$ over the field $\mathbb{R}(z_1)$ has repeated roots in the algebraic closure $\overline{\mathbb{R}(z_1)}$, which is impossible since $\mathbb{R}(z_1)$ has characteristic 0. So again $\Delta(g_{z_1})$ is a non-zero polynomial and has finitely many roots. All other choices of b give distinct roots to g_b. ☐

So for generic $z_1 \in \mathbb{R}$, g_{z_1} has roots $\lambda_d(g_{z_1}) < \lambda_{d-1}(g_{z_1}) < \cdots < \lambda_1(g_{z_1})$. Denote by A the set of such z_1, then A is dense open in $\mathbb{R}$.

Claim. *The map $z_1 \mapsto \lambda_i(g_{z_1})$ can be extended holomorphically to a complex neighbourhood of b for every $b \in A$.*

[70] Tao13, Lemma 17.

Proof of Claim. Consider the roots-to-coefficients map $\alpha : \mathbb{C}^n \to \mathbb{C}^n$ defined by

$$(\lambda_1, \lambda_2, \cdots, \lambda_n) \mapsto \left(\sum_i \lambda_i, \sum_{i<j} \lambda_i \lambda_j, \cdots, \prod_i \lambda_i \right).$$

One can compute that the derivative has determinant $\prod_{i<j}(\lambda_i - \lambda_j)$, so when this determinant is non-zero, α has a local holomorphic inverse α^{-1} ("coefficients-to-roots"), by the inverse function theorem.

Since the coefficient of z_2^d in $g_{z_1}(z_2)$ does not vanish when z_1 is at $b \in A$, the coefficients (divided by the highest coefficient) are also holomorphic in z_1 near b. Composing with α^{-1} and projecting to the i-th coordinate gives the desired holomorphic map. □

Having done these two claims, we know $\lambda_j(g_{z_1})$ is differentiable for most z_1. It remains to show that $\partial_1 \lambda_j(g_{z_1}) \leq 0$. Suppose not, then $\partial_{z_1} \lambda_j(g_{z_1})|_{z_1=b} = h > 0$ for some $b \in A$, so for sufficiently small $\varepsilon > 0$, $\lambda_j(g_{b+\varepsilon i}) = \lambda_j(g_b) + h\varepsilon i + o(\varepsilon)$ has positive real part. But this means $(b + \varepsilon i, \lambda_j(g_{b+\varepsilon i}), a_3, \cdots, a_n)$ is a zero of f, contradicting stability (and Lemma 1.15). □

It is possible to deduce these convexity results just from the following representation theorem of real stable polynomials in 2 variables[71], but we shall not use it:

Theorem (Helton–Vinnikov 2007)**.** [72][73] If $g(z_1, z_2)$ is a real stable polynomial, then there exist positive semi-definite self-adjoint matrices A_1 and A_2 and a self-adjoint matrix A_0 such that

$$g(z_1, z_2) = \pm \det(A_0 + z_1 A_1 + z_2 A_2).$$

Next, we shall see how we may apply the convexity result Lemma 3.20 to prove a bound with a smaller increment than in Lemma 3.17.

Lemma 3.21. [74][75] If $f(z_1, \cdots, z_n)$ is real stable polynomial and $\boldsymbol{a}$ lies above all zeros of f, and for some i and $\delta > 0$,

$$\Phi^i_f(\boldsymbol{a}) + \frac{1}{\delta} \leq 1,$$

then $\boldsymbol{a} + \delta \boldsymbol{e}_i$ lies above all zeros of $(1 - \partial_{z_i})f$, and for all j,

$$\Phi^j_{(1-\partial_{z_i})f}(\boldsymbol{a} + \delta \boldsymbol{e}_i) \leq \Phi^j_f(\boldsymbol{a}).$$

Proof. We start by showing that $\boldsymbol{a}$ (whence $\boldsymbol{a} + \delta \boldsymbol{e}_i$) is above all zeros of $(1 - \partial_{z_i})f$. Suppose $\boldsymbol{b}$ with $b_j \geq a_j$ for all j. Then $f(\boldsymbol{b}) \neq 0$, and by monotonicity in each coordinate (Lemma 3.20), $\Phi^i_f(\boldsymbol{b}) \leq \Phi^i_f(\boldsymbol{a}) < 1$, so $(1 - \partial_{z_i})f(\boldsymbol{b}) = \left(1 - \Phi^i_f(\boldsymbol{b})\right) f(\boldsymbol{b}) \neq 0$ as required.

Next we need to prove

$$\Phi^j_{(1-\partial_{z_i})f}(\boldsymbol{a} + \delta \boldsymbol{e}_i) \leq \Phi^j_f(\boldsymbol{a}).$$

On the left hand side,

$$\begin{aligned} \Phi^j_{(1-\partial_{z_i})f} &= \partial_{z_j} \log((1 - \Phi^i_f) \cdot f) \\ &= \partial_{z_j} \log(1 - \Phi^i_f) + \partial_{z_j} \log f \\ &= \frac{-\partial_{z_j} \Phi^i_f}{1 - \Phi^i_f} + \Phi^j_f, \end{aligned}$$

[71] MSS15b, Lemma 5.7.
[72] The version most relevant to us is BB10, Corollary 6.7.
[73] The original theorem is HV07, Theorem 2.2.
[74] MSS15b, Lemma 5.10.
[75] Tao13, Lemma 20.

so it suffices to prove

$$\frac{-\partial_{z_j}\Phi_f^i(\boldsymbol{a}+\delta\boldsymbol{e}_i)}{1-\Phi_f^i(\boldsymbol{a}+\delta\boldsymbol{e}_i)} \le \Phi_f^j(\boldsymbol{a}) - \Phi_f^j(\boldsymbol{a}+\delta\boldsymbol{e}_i)$$

By convexity in Lemma 3.20, on the right hand side we have

$$\Phi_f^j(\boldsymbol{a}) \ge \Phi_f^j(\boldsymbol{a}+\delta\boldsymbol{e}_i) - \delta\partial_{z_i}\Phi_f^j(\boldsymbol{a}+\delta\boldsymbol{e}_i),$$

so it suffices to prove

$$\frac{-\partial_{z_j}\Phi_f^i(\boldsymbol{a}+\delta\boldsymbol{e}_i)}{1-\Phi_f^i(\boldsymbol{a}+\delta\boldsymbol{e}_i)} \le -\delta\partial_{z_i}\Phi_f^j(\boldsymbol{a}+\delta\boldsymbol{e}_i).$$

Here, $\partial_{z_j}\Phi_f^i(\boldsymbol{a}+\delta\boldsymbol{e}_i) = \partial_{z_i}\Phi_f^j(\boldsymbol{a}+\delta\boldsymbol{e}_i)$, and it is non-positive by Lemma 3.20, so it suffices to prove

$$\frac{1}{1-\Phi_f^i(\boldsymbol{a}+\delta\boldsymbol{e}_i)} \le \delta,$$

which is true since $\Phi_f^i(\boldsymbol{a}+\delta\boldsymbol{e}_i) \le \Phi_f^i(\boldsymbol{a}) \le 1-\frac{1}{\delta}$ by assumption. ☐

Remark. The inequality

$$\Phi^j_{(1-\partial_{z_i})f}(\boldsymbol{a}+\delta\boldsymbol{e}_i) \le \Phi_f^j(\boldsymbol{a})$$

ensures that the condition $\Phi_f^i(\boldsymbol{a})+\frac{1}{\delta} \le 1$ is preserved under the transformation

$$\boldsymbol{z} \mapsto \boldsymbol{z}+\delta\boldsymbol{e}_i, \quad f \mapsto (1-\partial_{z_i})f,$$

so that we can iterate Lemma 3.21.

Finally, we can prove the bound that we have anticipated since the beginning of Section 3.3.

Lemma 3.22. [76] If $A_1, \cdots, A_n$ are positive semi-definite self-adjoint m by m matrices, and $\sum A_i = I, \operatorname{tr} A_i \le \varepsilon$, then $((1+\sqrt{\varepsilon})^2, (1+\sqrt{\varepsilon})^2, \cdots, (1+\sqrt{\varepsilon})^2)$ lies above all zeros of $p(z_1, z_2, \cdots, z_n) = (\prod_i(1-\partial_{z_i}))\det(\sum_i z_iA_i)$.

Remark. Comparing with what we want at the beginning of Section 3.3, ε corresponds to $\frac{r}{2}$, which can be large.

Proof. Let $f(z_1,\cdots,z_n) = \det(\sum_i z_iA_i)$. For any $t>0$, $(t,t,t,\cdots,t)$ is above all zeros of f (if $z_i \ge t$ for all i, then $\sum z_iA_i \succeq tI \succ 0$), so if we have some $\delta > 0$ such that

$$\Phi_f^i(t,t,\cdots,t)+\frac{1}{\delta} \le 1$$

for all i, then we can iterate Lemma 3.21 on each coordinate to show that $(t+\delta, t+\delta, \cdots, t+\delta)$ is above all zeros of $p = (\prod_{i=1}^n(1-\partial_{z_i}))\,f$.

To compute $\Phi_f^i(t,t,\cdots,t)$ (WLOG $i=1$), note that $f(t-h,t,\cdots,t) = \det(-hA_1+t(\sum_i A_i)) = \det(tI-hA_1) = h^m\det\left(\frac{t}{h}I-A_1\right)$, so $f(t-h,t,\cdots,t)=0$ if and only if $\frac{t}{h}$ is some eigenvalue $\lambda_k(A_1)$, so the corresponding values of h are $h_k = \frac{t}{\lambda_k(A_1)}$. So we have

$$\Phi_f^1(t,t,\cdots,t) = \frac{\partial_{z_1}f(t,t,\cdots,t)}{f(t,t,\cdots,t)} = \sum_k\frac{1}{h_k} = \sum_k\frac{\lambda_k(A_1)}{t} = \frac{\operatorname{tr}(A_1)}{t} \le \frac{\varepsilon}{t}. \qquad (\star)$$

Now we have shown that for any $t,\delta>0$ such that $\frac{\varepsilon}{t}+\frac{1}{\delta} \le 1$, $(t+\delta,\cdots,t+\delta)$ is above all zeros of p. It remains to minimise $t+\delta$ to obtain a good upper bound. By Cauchy–Schwarz, $t+\delta \ge (t+\delta)(\frac{\varepsilon}{t}+\frac{1}{\delta}) \ge (\sqrt{\varepsilon}+1)^2$, and equality can be attained. ☐

[76] MSS15b, Theorem 5.1.

Remark. ($\star$) is why we need $\operatorname{tr}(A_i) \le \varepsilon$. Our failed attempt only used the fact that the roots of $g(z_1) = f(z_1, t, t, \cdots, t)$ satisfies $\lambda_k(g) \le 0$, i.e. $h_k \ge t$ in the proof above. The bound this gives is

$$\Phi_f^1(t, t, \cdots, t) = \sum_{k=1}^{m} \frac{1}{h_k} \le \frac{m}{t},$$

but the bound we want must not depend on the dimension m.

Summarising:

Theorem 3.23 (Marcus–Spielman–Srivastava 2015)**.** [77] For every $\varepsilon > 0$, if independent random vectors $\boldsymbol{v}_i \in \mathbb{C}^m$ (taking finitely many possible values) satisfies $\sum_{i=1}^n \mathbb{E}(\boldsymbol{v}_i \boldsymbol{v}_i^*) = I_m$ and $\mathbb{E}(\|\boldsymbol{v}_i\|^2) \le \varepsilon$, then with non-zero probability we have

$$\left\| \sum_{i=1}^{n} \boldsymbol{v}_i \boldsymbol{v}_i^* \right\| \le (1 + \sqrt{\varepsilon})^2.$$

Proof. By Lemma 1.23, there is non-zero probability that

$$\lambda_1 \left(\chi \left(\sum_{i=1}^{n} \boldsymbol{v}_i \boldsymbol{v}_i^* \right) \right) \le \lambda_1 \left(\mathbb{E}\chi \left(\sum_{i=1}^{n} \boldsymbol{v}_i \boldsymbol{v}_i^* \right) \right).$$

As we have mentioned at the beginning of Section 3.3, a bound like Lemma 3.22 corresponds to a bound of $\lambda_1 \left(\mathbb{E}\chi \left(\sum_{i=1}^n \boldsymbol{v}_i \boldsymbol{v}_i^* \right) \right)$, so the right hand side is at most $(1 + \sqrt{\varepsilon})^2$. □

Remark. To prove Aim 3.15 and therefore Theorem 3.3, it suffices that the right hand side is $\varepsilon + o(\varepsilon)$ for sufficiently *large* ε, and we do not care about small ε.

Proof of Aim 3.15. Fix $\varepsilon' > 0$. After rescaling we have $\sum_{i=1}^n \mathbb{E}(\boldsymbol{v}_i \boldsymbol{v}_i^*) = I_{rn/2}$ and $\mathbb{E}(\|\boldsymbol{v}_i\|^2) = \frac{r}{2}$, so if we take $\varepsilon = \frac{r}{2}$ in Theorem 3.23, we have with non-zero probability,

$$\left\| \sum_{i=1}^{n} \boldsymbol{v}_i \boldsymbol{v}_i^* \right\| \le \left(1 + \sqrt{\frac{r}{2}} \right)^2 = \frac{1}{2} r + o(r) < \frac{1 + \varepsilon'}{2} r$$

for sufficiently large r, as required. □

4 Beyond This Essay

We have seen how the simply idea of Lemma 1.11 gives us a powerful new method. This essay only covers some applications of the particular case of mixed characteristic polynomials (Lemma 1.23), but interlacing families do not have to be in this form. For example, [MSS15c] used an interlacing family that comes from the expected characteristic polynomial of a sum of the form $\sum_i P_i A_i P_i^T$ (A_i being fixed symmetric matrices and P_i random permutation matrices) to show existence of bipartite Ramanujan graphs (allowing repeated edges) of any degree d and $2n$ vertices for any n. It would be interesting to find other combinatorial problems where interlacing families arise.

From a computational perspective, [MSS15a] noted that their proof (Section 2) of the existence of Ramanujan graphs does not give a polynomial time algorithm to build such graphs, because the first step would be to compute the matching polynomial μ_G, but its lowest term is the number

[77] MSS15b, Theorem 1.4.

of perfect matchings on G, which is a #P-complete problem[78], so there is no known polynomial time algorithm (as any such algorithm will imply $\mathsf{P} = \mathsf{NP}$). Nonetheless, based on [MSS15c], [Coh16] gave a polynomial time algorithm that builds a larger class of bipartite Ramanujan graphs than described in Section 2.

One might also investigate the quantitative version of Aim 3.4 or its equivalent formulations such as the diagonal $\frac{1}{2}$ version in Lemma 3.13, i.e. to get more precise bounds on how large r needs to be for fixed ε. For example, if we work out the details in Section 3, then we would have the bound $\|P_{S_i} Q P_{S_i}\| \leq \left(\sqrt{\frac{1}{2}} + \sqrt{\frac{1}{r}}\right)^2$ in the context of Lemma 3.13(2).[79]

In comparison, [RL20], which also uses the interlacing families method but with a generalisation of the characteristic polynomial, gives a bound of $\left(\sqrt{\frac{1}{2}} + \sqrt{\frac{1}{r} - \frac{1}{2(r-1)}}\right)^2$ for $r \geq 3$.[80]

5 References

[Ale] A. Alexanderian. *On continuous dependence of roots of polynomials on coefficients*. URL: `https://aalexan3.math.ncsu.edu/articles/polyroots.pdf`.

[Alo86] N. Alon. "Eigenvalues and expanders". In: *Combinatorica* 6.2 (1986), pp. 83–96. DOI: `10.1007/BF02579166`.

[BB10] J. Borcea and P. Brändén. "Multivariate Pólya–Schur classification problems in the Weyl algebra". In: *Proc. Lond. Math. Soc.* 101.1 (2010), pp. 73–104. DOI: `10.1112/plms/pdp049`.

[BL06] Y. Bilu and N. Linial. "Lifts, discrepancy and nearly optimal spectral gap". In: *Combinatorica* 26.5 (2006), pp. 495–519. DOI: `10.1007/s00493-006-0029-7`.

[Bog+90] N. N. Bogolubov et al. "Algebra of observables and state space". In: *General Principles of Quantum Field Theory*. Kluwer Academic Publishers, 1990. Chap. 6. DOI: `10.1007/978-94-009-0491-0_6`.

[Cas+06] P. G. Casazza et al. *The Kadison–Singer Problem in mathematics and engineering: A detailed account*. 2006. arXiv: `math/0510024`.

[Coh16] M. B. Cohen. "Ramanujan graphs in polynomial time". In: *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*. 2016, pp. 276–281. DOI: `10.1109/FOCS.2016.37`.

[CS07] M. Chudnovsky and P. Seymour. "The roots of the independence polynomial of a clawfree graph". In: *J Combin. Theory Ser. B* 97.3 (2007), pp. 350–357. DOI: `10.1016/j.jctb.2006.06.001`.

[Ded92] J. P. Dedieu. "Obreschkoff's theorem revisited: What convex sets are contained in the set of hyperbolic polynomials?" In: *J. Pure Appl. Algebra* 81.3 (1992), pp. 269–278. DOI: `10.1016/0022-4049(92)90060-S`.

[DSV12] S. De Winter, J. Schillewaert, and J. Verstraete. "Large incidence-free sets in geometries". In: *Electron. J. Combin.* 19.4 (2012). DOI: `10.37236/2831`.

[Fel80] H. J. Fell. "On the zeros of convex combinations of polynomials". In: *Pacific J. Math.* 89.1 (1980), pp. 43–50. DOI: `10.2140/pjm.1980.89.43`.

[Har13] N. J. A. Harvey. *An introduction to the Kadison–Singer Problem and the Paving Conjecture*. 2013. URL: `https://www.cs.ubc.ca/~nickhar/papers/KS/KS.pdf`.

[78] Val79, Theorem 1.
[79] Special case of Tao13, Corollary 24.
[80] Special case of RL20, Theorem 1.4.

[HL72] O. J. Heilmann and E. H. Lieb. “Theory of monomer-dimer systems”. In: *Comm. Math. Phys.* 25.3 (1972), pp. 190–232. DOI: 10.1007/BF01877590.

[HV07] J. W. Helton and V. Vinnikov. “Linear matrix inequality representation of sets”. In: *Comm. Pure Appl. Math.* 60.5 (2007), pp. 654–674. DOI: 10.1002/cpa.20155.

[LPS88] A. Lubotzky, R. Phillips, and P. Sarnak. “Ramanujan graphs”. In: *Combinatorica* 8.3 (1988), pp. 261–277. DOI: 10.1007/BF02126799.

[LS81] E. H. Lieb and A. D. Sokal. “A general Lee–Yang theorem for one-component and multicomponent ferromagnets”. In: *Comm. Math. Phys.* 80.2 (1981), pp. 153–179. DOI: 10.1007/BF01213009.

[Mar66] M. Marden. *Geometry of Polynomials*. 2nd ed. American Mathematical Society, 1966. URL: https://archive.org/details/geometryofpolyno0000mard.

[MSS15a] A. W. Marcus, D. A. Spielman, and N. Srivastava. “Interlacing families I: Bipartite Ramanujan graphs of all degrees”. In: *Ann. of Math.* 182.1 (2015), pp. 307–325. DOI: 10.4007/annals.2015.182.1.7.

[MSS15b] A. W. Marcus, D. A. Spielman, and N. Srivastava. “Interlacing families II: Mixed characteristic polynomials and the Kadison–Singer problem”. In: *Ann. of Math.* 182.1 (2015), pp. 327–350. DOI: 10.4007/annals.2015.182.1.8.

[MSS15c] A. W. Marcus, D. A. Spielman, and N. Srivastava. “Interlacing families IV: Bipartite Ramanujan graphs of all sizes”. In: *2015 IEEE 56th Annual Symposium on Foundations of Computer Science (FOCS)*. 2015, pp. 1358–1377. DOI: 10.1109/FOCS.2015.87.

[Pis18] T. Pisanski. “Not every bipartite double cover is canonical”. In: *Bull. Inst. Combin. Appl.* 82 (2018), pp. 51–55. URL: https://www.researchgate.net/publication/322696331_Not_every_bipartite_double_cover_is_canonical.

[RL20] M. Ravichandran and J. Leake. “Mixed determinants and the Kadison–Singer problem”. In: *Math. Ann.* 377 (2020), pp. 511–541. DOI: 10.1007/s00208-020-01986-7.

[Tao13] T. Tao. *Real stable polynomials and the Kadison-Singer problem*. 2013. URL: https://terrytao.wordpress.com/2013/11/04/real-stable-polynomials-and-the-kadison-singer-problem/.

[Val79] L. G. Valiant. “The complexity of computing the permanent”. In: *Theoret. Comput. Sci.* 8.2 (1979), pp. 189–201. DOI: 10.1016/0304-3975(79)90044-6.

[Von18] J. Vondrák. “Lecture 14. Characterization of linear stability-preserving transformations”. In: *Math 233A: Non-constructive methods in combinatorics*. 2018. URL: https://theory.stanford.edu/~jvondrak/MATH233A-2018/Math233-lec14.pdf.