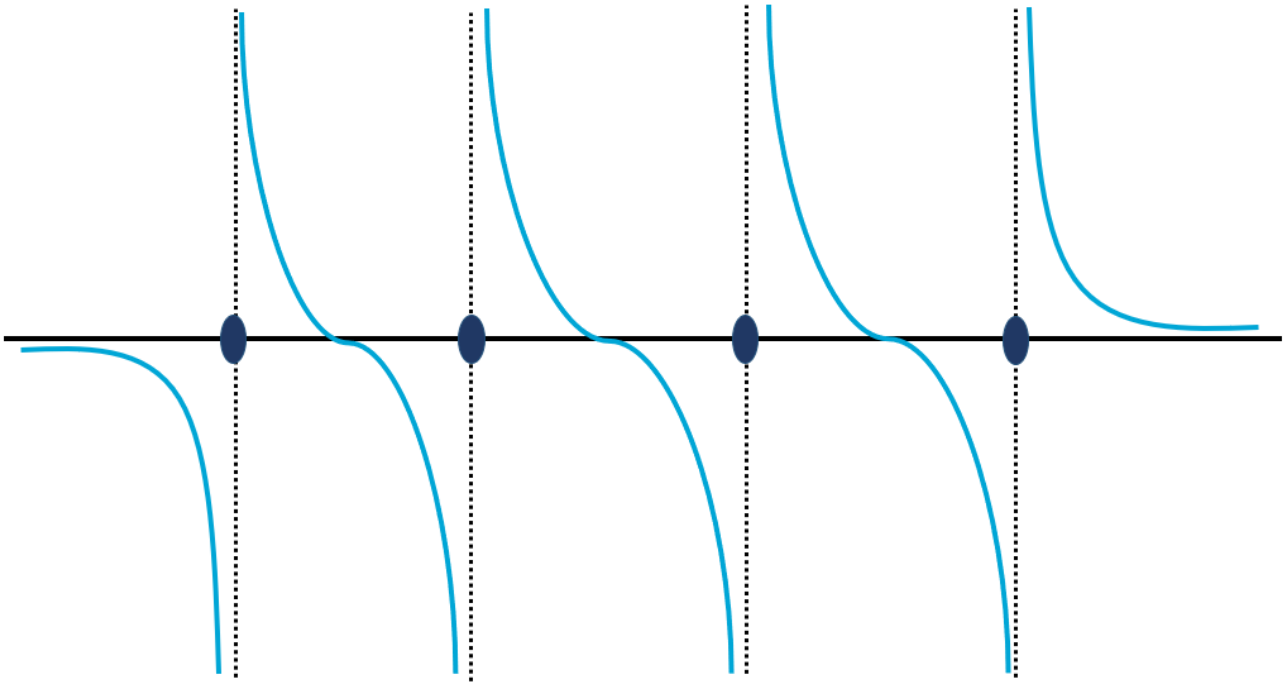
Controlling the behaviour of eigenvalues

The interlacing method

Lander Verlinde



TU Delft

Controlling the behaviour of eigenvalues

The interlacing method

by

Lander Verlinde

to obtain the degree of Bachelor of Science
at the Delft University of Technology,
to be defended publicly on Monday June 27, 2022 at 9:30 AM.

Student number: 5109353
Project duration: April 19, 2022 – June 27, 2022
Thesis committee:

Prof. A. Bishnoi,	TU Delft, supervisor
Prof. A. W. Marcus,	EPFL, supervisor
Prof. B. Janssens,	TU Delft

TU Delft

Layman Abstract

It is hard to find the roots of a polynomial with a higher degree than two. When calculating the eigenvalues of a matrix, one encounters unfortunately regularly polynomials of such a higher degree. However, we do want to solve problems for eigenvalues for all types of matrices. Hence it is desirable to have a method that avoids those difficult computations involving roots of polynomials.

This thesis suggests such a method for a specific type of matrix. When you add a so-called *rank-one matrix* to a *symmetric matrix*, the new eigenvalues are related to the original ones of the symmetric matrix: they are interlacing. So instead of redoing a lot of calculations you get an estimate for the eigenvalues for free.

Even though these types of matrices seem very specific, they are actually used in a lot of applications. Therefore we can use the *interlacing method* in a lot of different settings.

This thesis has two main goals. The first one is to guide the reader through three of such applications, all from different mathematical fields to show the versatility of the method.

Yet, as it is still an approximation on the actual eigenvalues, some problems may arise. We don't know where the eigenvalues are exactly, only how they more or less behave after a rank-one update. This creates some new challenges that have to be solved. Those challenges lead us to the second goal of this thesis: offer some solutions to problems one may encounter with the interlacing method, to create an intuition in which cases the method can be useful and in which not.

Abstract

This thesis uses the method of interlacing polynomials to study the behaviour of eigenvalues of a matrix after a rank-one update. Specifically, interlacing polynomials, common interlacing and interlacing families are exhaustively studied. These are excellent tools to find bounds on the eigenvalues of updated matrices by keeping track of how they move.

This enables us to prove results in different mathematical fields. We investigate three of them. The first one is from spectral graph theory: we prove the existence of a sharp κ-approximation for any graph.

The second result is from linear algebra. It states that if a matrix has a high stable rank, it must contain a large column submatrix with large least singular value.

Lastly, the proof of the Kadison-Singer Problem is discussed. Despite being a problem from analysis, it was solved with the interlacing method, which is originally a method from discrete mathematics.

This thesis shows how these three seemingly different problems are all connected by the same method, highlighting its advantages. The objective is to present a clear framework of the different facets of the interlacing method and provide an insight in the situations where one can expect the said method to be useful.

Contents

1

Introduction

This thesis provides a review for the *method of interlacing polynomials*. The method of interlacing polynomials is actually a collective term for multiple methods that all use either interlacing polynomials or interlacing families. In short, we say that polynomial $g(x) = \prod_{i=1}^{d+1}(x - \alpha_i)$ interlaces a polynomial $f(x) = \prod_{i=1}^{d}(x - \beta_i)$ if

$$\alpha_1 \geq \beta_1 \geq \alpha_2 \geq \beta_2 \geq \alpha_3 \geq \cdots \geq \alpha_d \geq \beta_d \geq \alpha_{d+1}.$$

An interlacing family is a tree of interlacing polynomials. It is surprising how such a simple concept on the roots of polynomials has some highly non-trivial applications. [7]

This thesis discusses three papers that each prove a result with the interlacing method. All proofs involve characteristic polynomials of matrices. It is in this facet that the reviewed papers are innovative. Traditionally, eigenvalues are controlled through moments $\mathrm{Tr}(A^k) = \sum_i \lambda_i(A)^k$ or Stieltjes-transforms $\mathrm{Tr}(A-z)^{-1} = \sum_i (\lambda_i(A)-z)^{-1}$. [19] Instead, we are going to consider them as roots of the characteristic polynomial $p_A(x)$. The main objective is to fully understand the strength of the interlacing method. Therefore, the main results of each paper are presented sequentially to show the historic evolution of how these techniques were shaped.

The first chapter shortly explains the most basic notions. First, interlacing polynomials are introduced, followed by common interlacing. Those definitions are then used to rigorously define an interlacing tree.

Since we are mainly going to focus on characteristic polynomials of matrices, it is important to sketch in which situations one can expect interlacing polynomials or common interlacing when dealing with these characteristic polynomials before even starting the proofs.

Next, the papers are reviewed. The first one, *Twice Ramanujan Sparsifiers* by Batson, Spielman and Srivastava proves a result from spectral graph theory. [1] More precisely, it proves the existence of a sharp κ-approximation for any graph G. This was an early paper where the notion of interlacing polynomials for eigenvalues was introduced. Hence the method is still in its early stages. The proof uses the idea of iteratively constructing a matrix and at each step keeping track of the zeros of its characteristic polynomial. This idea will return in a more complicated form later on and hence this proof is the perfect starting point to explore the method.

The second paper, by Marcus, Spielman and Srivastava, is called *Interlacing Families III: Sharper Restricted Invertibility Estimates* and proves a linear algebraic statement. [12] This section will focus on interlacing families and their advantages when constructing them carefully. The goal of this section is showing the link between interlacing families of characteristic polynomials on one side and Laguerre polynomials on the other. The latter show up in multiple fields of mathematics and have their origin in differential equations. The fact that these highly studied and powerful polynomials will appear in our interlacing families will support the idea that this method deserves attention and study.

Lastly, the greatest achievement of the interlacing method is discussed: the proof of the Kadison-Singer problem. This famous open problem was solved in the paper *Interlacing Families II: Mixed Characteristic Polynomials and The Kadison-Singer Problem* also by by Marcus, Spielman and Srivastava. [14] This time, the problem has an analytical nature. Notions from the first paper are combined with notions from the second one. This will enable us to fully understand the proof.

2

Core of the interlacing method

Firstly, we are going to review the most important definitions for the method of interlacing polynomials: *interlacing polynomials*, *common interlacing* and an *interlacing family*.

Definition 2.0.1. A polynomial $g(x) = \prod_{i=1}^{d+1}(x - \alpha_i)$ interlaces a polynomial $f(x) = \prod_{i=1}^{d}(x - \beta_i)$ if

$$\alpha_1 \geq \beta_1 \geq \alpha_2 \geq \beta_2 \geq \alpha_3 \geq \cdots \geq \alpha_d \geq \beta_d \geq \alpha_{d+1}.$$

That means that the roots of the two polynomials will always look as in Figure 2.1. The blue dots represent the roots of polynomial g, the orange ones of f. Since the inequalities hold, it is clear that all roots are real. Furthermore, they follow each other sequentially as in the picture, but it is possible that certain blue and orange roots coincide. There is also no condition on where the polynomials are positive or negative, only where their roots lie respectively to the roots of the other polynomial.

Figure 2.1: Sketch of how the roots should succeed each other.

Definition 2.0.2. Polynomials $f_1, \ldots, f_m$ have a common interlacing if there is a single polynomial g that interlaces f_i for each i.

The above are the two most basic concepts involving interlacing polynomials. A third key concept is an *interlacing family*. Its definition is a bit less straightforward.

Definition 2.0.3. An interlacing family is a finite rooted tree T whose vertices $v \in T$ are labelled by monic real-rooted polynomials $f_v(x)$ with the following two properties:

1. Every polynomial corresponding to a non-leaf vertex v is a convex combination of the polynomials corresponding to its children.
2. For all nodes $v_1, v_2 \in T$ with a common parent, all convex combinations of f_{v_1} and f_{v_2} are real-rooted.

Hence, a set of polynomials is an interlacing family if they are the labelling of such a tree.

But why is this called an interlacing family? In first instance, there doesn't seem to be a clear link with interlacing polynomials. To see where this name comes from, it is worth noting that the second condition implies that all convex combinations of the children of a node are real-rooted. That means that we can apply the following theorem:

Theorem 2.0.4. Let $f_1, \ldots, f_m$ be real-rooted polynomials of the same degree with positive leading coefficient. Then $f_1, \ldots, f_m$ have a common interlacing if and only if $\sum_{i=1}^{m} t_i f_i$ is real rooted for all convex combinations $t_i \geq 0$, $\sum_{i=1}^{m} t_i = 1$.

This theorem will be proven later on in section 5.2.

It implies that for all vertices with common parent in the tree, there exists a common interlacing.

2.1. Characteristic polynomials

All chapters consider the interlacing method in the setting of characteristic polynomials. The first indication that the said method is interesting for these polynomials is given by the following basic result: the eigenvalues of the real symmetric matrix A *almost* interlace those of $A + vv^T$ where v is a non-zero vector.

Theorem 2.1.1 (Eigenvalue Interlacing Theorem)**.** Let A be a real symmetric matrix and v any non-zero vector. If A has eigenvalues λ_i and $A + vv^T$ has eigenvalues μ_i, then

$$\lambda_n \leq \mu_n \leq \lambda_{n-1} \leq \mu_{n-1} \leq \cdots \leq \lambda_1 \leq \mu_1.$$

This sum $A + vv^T$ of a matrix A and the rank-one matrix vv^T formed with any non-zero vector v is called a rank-one update of the matrix A.

From the proof of the Eigenvalue Interlacing Theorem, an intuitive model can be derived fluently. But before starting with the proof, we need two lemmas. [6]

Lemma 2.1.2 (Sherman-Morrison Formula)**.** Suppose A is a non-singular $n \times n$ matrix and v is a vector. If $1 + v^T A^{-1} v$ is non-zero[1], then

$$\left(A + vv^T\right)^{-1} = A^{-1} - \frac{A^{-1} vv^T A^{-1}}{1 + v^T A^{-1} v}.$$

Lemma 2.1.3 (Matrix Determinant Lemma)**.** If A is non-singular and v is a vector, then

$$\det(A + vv^T) = \det(A)(1 + v^T A^{-1} v).$$

Proof of Eigenvalue Interlacing Theorem. Suppose $p_A(x)$ is the characteristic polynomial of the real symmetric matrix A, then the characteristic polynomial of the rank-one updated matrix is given by:

$$p_{A+vv^T}(x) = \det(xI - A - vv^T) = \det(xI - A)(1 - v^T (xI - A)^{-1} v).$$

Now let $A = \sum_j \lambda_j u_j u_j^T$ be the diagonalisation of A. λ_j are the eigenvalues of A and u_j the corresponding eigenvectors. Hence,

$$(xI - A)^{-1} = \sum_j \frac{1}{x - \lambda_j} u_j u_j^T.$$

If we plug in this identity in the above expression and replace the determinant by $p_A(x)$, we find

$$p_{A+vv^T}(x) = p_A(x) \left(1 - \sum_j \frac{\langle v, u_j \rangle^2}{x - \lambda_j}\right).$$

While it may not seem that way, the right-hand side is indeed a polynomial. The $x - \lambda_j$ in the denominator will cancel out against $p_A(x)$. From this expression, it is clear that the newly obtained polynomial has two types of zeros:

- $p_A(x) = 0$
- $1 - \sum_j \frac{\langle v, u_j \rangle^2}{x - \lambda_j} = 0$

The first case is when the added vector v is orthogonal to an eigenvector u_j and hence the zero is equal to λ_i. One can verify this by substituting $p_A(x) = \prod_{j=1}^{n} (x - \lambda_j)$ in the expression for $p_{A+vv^T}(x)$ and evaluating it at $x = \lambda_i$. So in this case the corresponding eigenvalue stays equal. Suppose now that the vector we have added is not orthogonal to any of the eigenvectors.

The second case for the roots considers all the points x such that $\sum_j \frac{\langle v, u_j \rangle^2}{x - \lambda_j} = 1$. This rational function is continuous everywhere except at the different eigenvalues λ_j. In that case one of the terms is divided by zero. Calculating the different limits is straightforward:

[1] The case where $1 + v^T A^{-1} v = 0$ is exactly the case where $A + vv^T$ is singular.

$$\lim_{x \uparrow \lambda_j} \left(\sum_j \frac{\langle v, u_j \rangle^2}{x - \lambda_j} \right) = -\infty$$

$$\lim_{x \downarrow \lambda_j} \left(\sum_j \frac{\langle v, u_j \rangle^2}{x - \lambda_j} \right) = +\infty$$

$$\lim_{x \to -\infty} \left(\sum_j \frac{\langle v, u_j \rangle^2}{x - \lambda_j} \right) = 0$$

$$\lim_{x \to +\infty} \left(\sum_j \frac{\langle v, u_j \rangle^2}{x - \lambda_j} \right) = 0$$

Using these limits we can make a sketch on what the function $\left(\sum_j \frac{\langle v, u_j \rangle^2}{x - \lambda_j} \right)$ looks like:

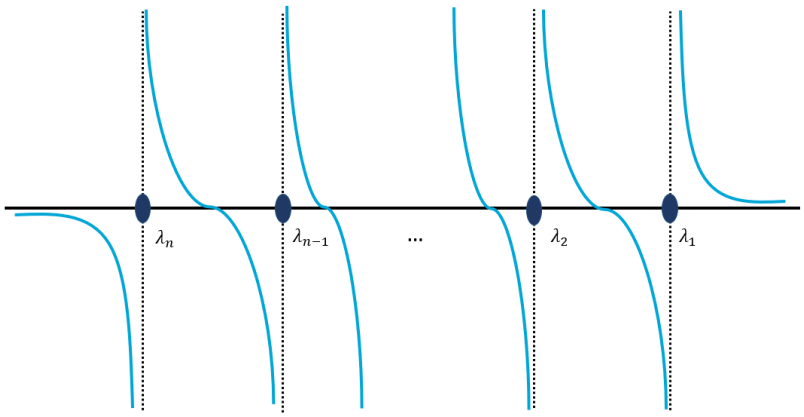


Figure 2.2: Sketch of $\left(\sum_j \frac{\langle v, u_j \rangle^2}{x - \lambda_j} \right)$.

Because of continuity there should be at least one value of x for which the function becomes 1 between each λ_j and λ_{j_1}. Furthermore there is also such an x larger than λ_1. But since the matrix only has n eigenvalues, we can conclude that there is indeed exactly one new eigenvalue between two successive λ_j and one eigenvalue larger than λ_1.This shows that if v is not orthogonal to any u_i the eigenvalues are even strictly interlacing: $\lambda_n < \mu_n < \cdots < \lambda_1 < \mu_1$.

When we do add a vector v orthogonal to u_i for some $1 \leq i \leq n$ we have to combine the two types of roots. The limits of $\sum_j \frac{\langle v, u_j \rangle^2}{x - \lambda_j}$ stay the same for all λ_j except for λ_i. Since

$$\frac{\langle v, u_i \rangle^2}{x - \lambda_i} = 0$$

for any value of x, even in its limit at λ_i, we can just ignore this term and the rational function becomes:

$$\sum_{j \neq i} \frac{\langle v, u_j \rangle^2}{x - \lambda_j}.$$

The exact same analysis as above applies to this function. So we find such a zero in all intervals $[\lambda_n, \lambda_{n-1}], \ldots, [\lambda_{i+2}, \lambda_{i+1}], [\lambda_{i+1}, \lambda_{i-1}], [\lambda_{i-1}, \lambda_{i-2}], \ldots, [\lambda_2, \lambda_1]$. By the same counting argument as before, we only need to find one more zero.

We have already established that there is a zero at λ_i itself, so this ensures that there is a root in $[\lambda_{i+1}, \lambda_i]$ and another one in $[\lambda_i, \lambda_{i-1}]$. Hence also in this case the Eigenvalue Interlacing Theorem holds. This argument can of course be extended when the added vector is orthogonal to multiple eigenvectors.

□

As previously said, this proof enables us to look at the rank-one update of a matrix in a very intuitive way. Consider the eigenvalues λ as balls lying on a slope as depicted in Figure 2.3. At each eigenvalue of A a barrier is put. Each ball is resting against such a barrier because of gravity. Hence these are indeed the locations of the matrix A.

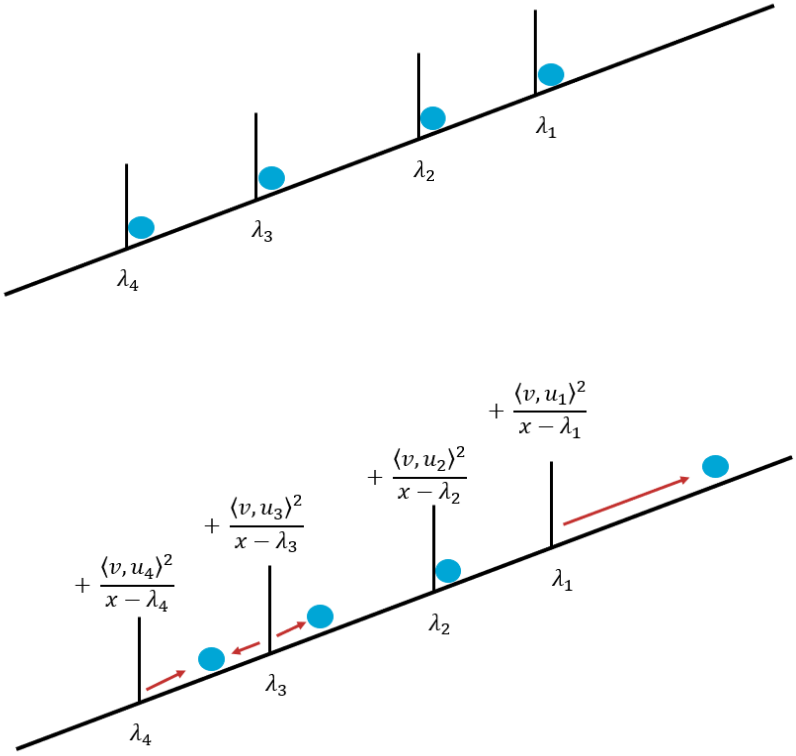


Figure 2.3: Physical model of eigenvalues

Adding a rank-one matrix corresponds to adding a charge to the barriers. These charges will push away the balls. The eigenvalues will start rolling until the forces pushing upwards, the forces pushing downwards and gravity cancel each other out and an equilibrium state is reached, or when it is pushed against the barrier above and cannot move any further. However, a key observation is that the largest eigenvalue cannot be stopped by a barrier and it is hence possible that it rolls quite far away until it reaches its equilibrium. This means that the largest eigenvalue of the rank-one update can end up much larger than all its other eigenvalues.

In any case, since the eigenvalues are blocked by the barriers one will always have:

$$p_A(x) = \prod_{i=1}^{d}(x-\lambda_i) \text{ and } p_{A+vv^T} = \prod_{i=1}^{d}(x-\mu_i)$$

$$\Rightarrow \mu_1 \geq \lambda_1 \geq \mu_2 \geq \lambda_2 \geq \cdots \geq \mu_d \geq \lambda_d.$$

This is why these eigenvalues are *almost* interlacing. To be really interlacing a λ_0 larger than μ_1 is still needed. This lack of an upper bound for μ_1 will result in some new challenges which will be discussed and solved in the next chapter.

Using the Eigenvalue Interlacing Theorem, one can easily construct a set of polynomials that have a common interlacing.

Lemma 2.1.4. Let A be a d-dimensional symmetric matrix and let $v_1, v_2, \dots, v_k \in \mathbb{R}^d$. The characteristic polynomials

$$f_j(x) = \det(xI - A - v_j v_j^T)$$

have a common interlacing.

Proof. Let $g(x) = \det(xI - A) = \prod_{i=1}^{d}(x - \lambda_i)$ and for arbitrary j: $f_j(x) = \prod_{i=1}^{d}(x - \mu_i)$. As already established:

$$\mu_1 \geq \lambda_1 \geq \mu_2 \geq \lambda_2 \geq \cdots \geq \mu_d \geq \lambda_d$$

Hence for large enough λ_0, $(x - \lambda_0)g(x)$ interlaces every $f_j(x)$ □

Now that we have built up the basic concepts for the method of interlacing polynomials we can look into their use in more elaborate proofs.

3

Twice-Ramanujan Sparsifiers

Maybe the most obvious advantage of interlacing polynomials is that even though we do not know what the exact values of the zeros of certain polynomials are, it gives us an approximate idea of their locations. This is a very strong tool in finding upper and lower bounds on these zeros. That was exactly the very first reason why this method was used by Batson, Spielman and Srivastava to prove is the existence of a *Twice-Ramanujan Sparsifier* for any graph. [1] [16]

3.1. Statement of the problem

A Twice-Ramanujan Sparsifier is - surprise, surprise - a sparsifier. A sparsifier of a graph is an approximation of a graph in a useful way. It is mostly used in computer science. The approximation has less edges and hence computations on this graphs will become faster. [2] [5] But of course the computations on this approximation still need to be accurate. Therefore different conditions can be imposed on the sparsifier, depending on the problem.

This section considers a κ-approximation as type of sparsifier.

Definition 3.1.1. Consider an edge weighted graph $G = (V, E, w)$ where $w : E \to \mathbb{R}$. $H = (V, F, \tilde{w})$ is a κ-approximation of G if for all $x \in \mathbb{R}^V$:

$$x^T L_G x \leq x^T L_H x \leq \kappa \cdot x^T L_G x$$

with L_G and L_H respectively the Laplacian matrices of G and H.

A Laplacian matrix of a weighted graph $G = (V, E, w)$ with $w_{(i,j)}$ is the weight of edge (i, j) is defined as follows:

$$L(i,j) = \begin{cases} \sum_z w_{(i,z)} & \text{if } i = j \\ -w_{(i,j)} & \text{if } i \text{ is adjacent to } j \\ 0 & \text{otherwise} \end{cases}$$

A natural question for such a definition is ”How small can we make κ such that for any graph G there exists a κ-approximation H?”

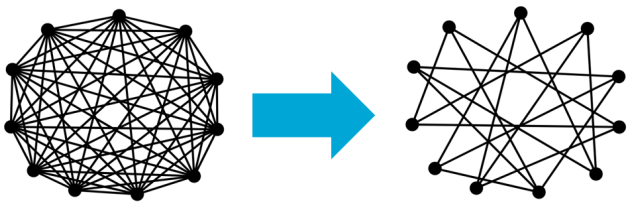


Figure 3.1: A sparsifier is an approximation of a graph while keeping certain properties of the graph.

In the case where G is a complete graph, an excellent sparsifier is given by the *Ramanujan Graphs*: d-regular graphs whose non-zero Laplacian eigenvalues are in the interval $[d - 2\sqrt{d-1}, d + 2\sqrt{d-1}]$. If we have such a Ramanujan Graph on n vertices, then we can multiply the weight of every edge by $n/(d - 2\sqrt{d-1})$. This gives a κ-approximation for the complete graph on n vertices for

$$\kappa = \frac{d + 2\sqrt{d-1}}{d - 2\sqrt{d-1}}.$$

However, when the paper Twice-Ramanujan Sparsifiers was written, the existence of a Ramanujan Graph for any degree d wasn't proven yet. Hence the existence of such a sparsifier wasn't even proven for an arbitrary complete graph G.

In this chapter we are going to prove the existence of a κ-approximation for *any* graph G with a κ as least as good as the one obtained from Ramanujan Graphs.

Theorem 3.1.2. For every $d > 1$, every undirected weighted graph $G = (V, E, w)$ on n vertices contains a weighted subgraph $H = (V, F, \tilde{w})$ with $\lceil d(n-1) \rceil$ edges that satisfies

$$x^T L_G x \leq x^T L_H x \leq \frac{d + 1 + 2\sqrt{d}}{d + 1 - 2\sqrt{d}} \cdot x^T L_G x.$$

One can verify that this κ is indeed slightly better than the one from the Ramanujan Graphs. However, this construction requires twice as many edges than the sparsifier obtained from the said Ramanujan Graphs. Ergo the name *Twice-Ramanujan Sparsifiers*.

The most important theorem to prove this existence is actually a statement that looks more like linear algebra than graph theory.

Theorem 3.1.3. Suppose $d > 1$ and $v_1, v_2, \ldots, v_m$ are vectors in $\mathbb{R}^n$ with

$$\sum_{i \leq m} v_i v_i^T = I.$$

Then there exist scalars $s_i \geq 0$ with $|\{i : s_i \neq 0\}| \leq dn$ do that

$$I \preceq \sum_{i \leq m} s_i v_i v_i^T \preceq \left(\frac{d + 1 + 2\sqrt{d}}{d + 1 - 2\sqrt{d}} \right) I.$$

Where $A \preceq B$ means that $B - A$ is positive semi-definite.

3.2. Idea of the proof

To prove Theorem 3.1.3, the matrix $A_m := \sum_{i \leq m} s_i v_i v_i^T$ will be constructed iteratively. Hence we will create a sequence $A_0 = 0, A_1, \ldots, A_Q$ (with $Q \leq dn$) by at each ith step adding $s_i v_i v_i^T$ with a nonzero s_i. The reason this approach seems promising is that each step is a rank-one update of the previous matrix. From section 2.1, we known that the original eigenvalues will almost interlace the new ones. Hence we have more or less an idea where the new eigenvalues will end up. So the main idea of the proof is to start with a certain interval. This will be the "goal" to aim the eigenvalues of A_1 in by choosing the right v_1 and s_1. Next we will move this interval and aim the eigenvalues of A_2 in this new goal. This is illustrated in Figure 3.2. These steps will be repeated until we end up with A_m such that eventually:

$$\frac{\lambda_{max}(A_Q)}{\lambda_{min}(A_Q)} \leq \frac{d + 1 + 2\sqrt{d}}{d + 1 - 2\sqrt{d}}$$

which implies Theorem 3.1.3.

Unfortunately, this aiming of the eigenvalues in the intervals is not always that easy. Problems arise for example when the eigenvalues end up too close to each other. This may seems counter intuitive, since if they are close to each other it seems easier to aim them in an interval. Why we don't want them too close together is illustrated in Figure 3.3. The real difficulty occurs at the next iteration because of the almost interlacing. From the figure it is clear that the smallest eigenvalues don't have any room to move to the right-hand side.

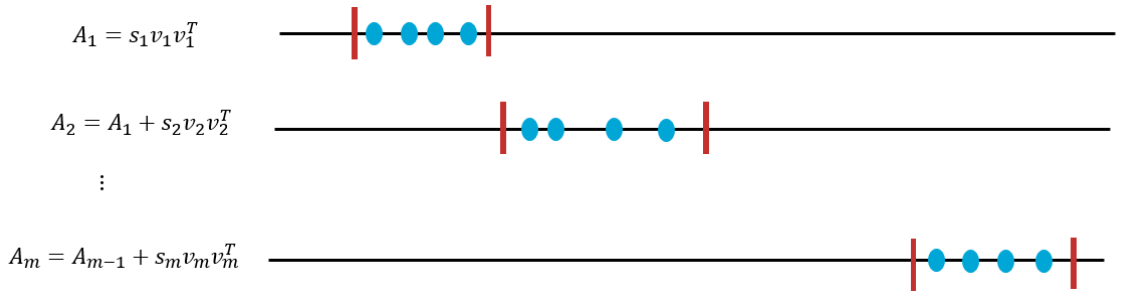


Figure 3.2: Sketch of iterations to construct A_m. At each step we choose the right v_i and s_i to aim the blue eigenvalues in the interval.

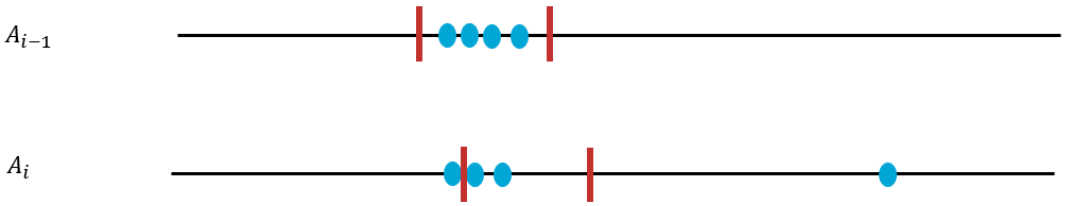


Figure 3.3: When the eigenvalues are too close, the next iteration encounters difficulties

To use the physical model from Figure 2.3, the first eigenvalues immediately hit the next wall. But since these walls are so close to each other, the greatest eigenvalue is repelled by the forces of all the walls behind him. This ball will shoot far away and hence the largest eigenvalue of the rank-one update will be way larger than the other ones. In this case the quotient of the largest eigenvalue and the smallest one cannot be bounded by small κ. This is an unfavorable case that we should avoid under all circumstances. So at each iteration we want the eigenvalues to end up in this interval *and* also be spread out. The difficulty lies of course in the fact that this seems to contradict each other.

3.3. The need for barrier functions

To measure this spreading of the eigenvalues we will introduce a new function: the barrier function. These functions are used a lot in optimisation and in physics and thanks to the physical model we have built in chapter 2 we can introduce them for eigenvalues as well. [22] [24] We will use an upper and a lower barrier function:

Definition 3.3.1. For $u, l \in \mathbb{R}$ and A a symmetric matrix with eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$, we define:

$$\Phi^u(A) := \mathrm{Tr}(uI - A)^{-1} = \sum_i \frac{1}{u - \lambda_i} \text{ (Upper potential)}$$

$$\Phi_l(A) := \mathrm{Tr}(A - lI)^{-1} = \sum_i \frac{1}{\lambda_i - l} \text{ (Lower potential)}$$

As long as $\lambda_{max} < u$ and $\lambda_{min} > l$, these potentials measure how far the eigenvalues of A are from the upper barrier u and the lower barrier l respectively. The really useful property of these barrier functions is that they tell something about the locations of all eigenvalues combined. For example, one can verify that if $\Phi^u(A) < 1$, then there is no eigenvalue of A within distance 1 from u, nor are there 2 λ_i's within distance 2, etc. More generally, it tells us that there are no k λ_i's within distance k from the upper bound.

This barrier function is exactly the tool we need to formulate the iterations in such a way that the eigenvalues don't come too close to the barrier and are spread out enough.

3.4. The proof

So let's make the proof and its iterations more rigorous. At each step, we want to satisfy the following conditions:

1. Initially, $A_0 = 0$ and we place a barrier at $u = u_0$ and $l = l_0$. The potentials are
$$\Phi^{u_0}(A_0) = \epsilon_U \text{ and } \Phi_{l_0}(A_0) = \epsilon_L$$
2. Each new iteration is obtained by a rank-one update of the previous one:
$$A_{q+1} = A_q + tvv^T \text{ for some } v \in \{v_i\} \text{ and } t \geq 0$$
3. At each step we move the lower barrier l by δ_L and the upper barrier u by δ_U. Furthermore, we don't want the potentials to increase at any step. For $q = 0, 1, \ldots, Q$:
$$\Phi^{u+\delta_U}(A_{q+1}) \leq \Phi^u(A_q) \leq \epsilon_U \text{ for } u = u_0 + q\delta_U$$
$$\Phi_{l+\delta_L}(A_{q+1}) \leq \Phi_l(A_q) \leq \epsilon_L \text{ for } l = l_0 + q\delta_L$$
4. Lastly, we want to make sure that no eigenvalue crosses a barrier. For every $q = 0, 1, \ldots, Q$:
$$\lambda_{max}(A_q) < u_0 + q\delta_U \text{ and } \lambda_{min}(A_q) > l_0 + q\delta_L$$

So the final step in the proof will be to pick positive values for $u_0, l_0, \delta_U, \delta_L, \epsilon_U$ and ϵ_L such that after $Q = dn$ steps we obtain A_n with

$$\frac{\lambda_{\max}(A_Q)}{\lambda_{\min}(A_Q)} \leq \frac{u_0 + dn\delta_U}{l_0 + dn\delta_L} = \frac{d + 1 + 2\sqrt{d}}{d + 1 - 2\sqrt{d}}$$

3.4.1. Moving the upper barrier

At each iteration we want to move the upper barrier by δ_U. When moving this barrier, the upper potential $\Phi^u(A)$ decreases, since all eigenvalues will be situated further away from the barrier. Hence we create some room to add $tv_iv_i^T$ for some $t > 0$ and counteract the decrease due to shifting of the barrier. To know whether we can always add such a vector and how large we can make t, we use the following theorem.

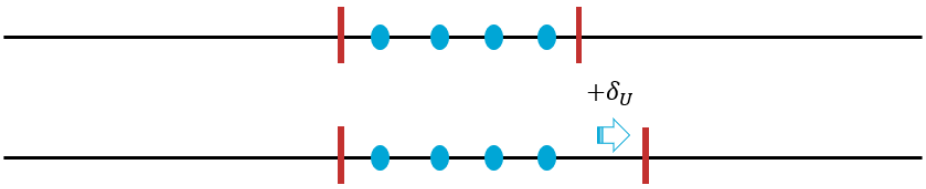


Figure 3.4: Shifting the upper barrier creates room to move the eigenvalues without increasing the upper potential $\Phi^u(A)$

Theorem 3.4.1 (Upper Barrier Shift)**.** Suppose $\lambda_{max}(A) < u$, δ_U a positive constant and v is any vector. If

$$\frac{1}{t} \geq \frac{v^T \left((u + \delta_U) I - A\right)^{-2} v}{\Phi^u(A) - \Phi^{u+\delta_U}(A)} + v^T \left((u + \delta_U) I - A\right)^{-1} v =: U_A(v)$$

then

$$\Phi^{u+\delta_U}\left(A + tvv^T\right) \leq \Phi^u(A) \quad \text{and } \lambda_{\max}\left(A + tvv^T\right) < u + \delta_U.$$

This theorem states that if we increase the upper barrier by δ_U and then add t that satisfies the above conditions, we do not increase the upper potential and all eigenvalues stay smaller than the barrier.

Proof. To ease notation, let $u' = u + \delta_U$.

$$\begin{aligned}
\Phi^{u+\delta_U}\left(A + tvv^T\right) &= \operatorname{Tr}\left(u'I - A - tvv^T\right)^{-1} \\
&= \operatorname{Tr}\left((u'I - A)^{-1} + \frac{t\,(u'I - A)^{-1} vv^T (u'I - A)^{-1}}{1 - tv^T (u'I - A)^{-1} v}\right) \\
&\quad \text{because of Theorem 2.1.2} \\
&= \operatorname{Tr}(u'I - A)^{-1} + \frac{t \operatorname{Tr}\left(v^T (u'I - A)^{-1} (u'I - A)^{-1} v\right)}{1 - tv^T (u'I - A)^{-1} v} \\
&\quad \text{since } \operatorname{Tr} \text{ is linear and } \operatorname{Tr}(XY) = \operatorname{Tr}(YX) \\
&= \Phi^{u+\delta_U}(A) + \frac{tv^T (u'I - A)^{-2} v}{1 - tv^T (u'I - A)^{-1} v} \\
&= \Phi^{u}(A) - \left(\Phi^{u}(A) - \Phi^{u+\delta_U}(A)\right) + \frac{v^T (u'I - A)^{-2} v}{1/t - v^T (u'I - A)^{-1} v} \\
&\leq \Phi^{u}(A) \\
&\quad \text{since } U_A(v) \leq \frac{1}{t}
\end{aligned}$$

To prove that $\lambda_{\max}\left(A + tvv^T\right) < u + \delta_U$ we note that $U_A(v) > v^T (u'I - A)^{-1} v$.
So for $U_A(v) \leq \frac{1}{t}$, $\frac{v^T (u'I - A)^{-2} v}{1/t - v^T (u'I - A)^{-1} v}$ is finite. Suppose that $\lambda_{\max}\left(A + tvv^T\right) \geq u + \delta_U$, then there would be some positive $t' \leq t$ for which $\lambda_{max}\left(A + t'vv^T\right) = u + \delta_U$. For this t', $\Phi^{u+\delta_U}\left(A + t'vv^T\right)$ would blow up, while we have just argued that it should be finite. □

3.4.2. Moving the lower barrier

Besides the upper barrier, we also want to shift the lower barrier at each iteration. When we add δ_L to this barrier, we move it closer to the eigenvalues and thus the lower potential $\Phi_l(A)$ increases. So in contrast to shifting the upper barrier, where we *could* move the eigenvalues, now we *have to* move them. The following theorem tells us how much of vv^T we have to add such that the lower potential decreases again.

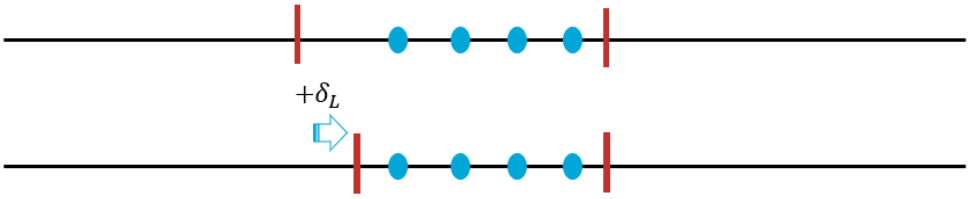


Figure 3.5: Shifting the lower barrier takes away room between the barrier and the eigenvalues. Hence we must move the eigenvalues to decrease the lower potential $\Phi_l(A)$.

Theorem 3.4.2 (Lower Barrier Shift)**.** Suppose $\lambda_{min}(A) > l, \Phi_l(A) \leq \frac{1}{\delta_L}$ for some real positive constant δ_L and v is any vector. If

$$0 < \frac{1}{t} \leq \frac{v^T (A - (l + \delta_L) I)^{-2} v}{\Phi_{l+\delta_L}(A) - \Phi_l(A)} - v^T (A - (l + \delta_L) I)^{-1} v =: L_A(v)$$

then

$$\Phi_{l+\delta_L}\left(A + tvv^T\right) \leq \Phi_l(A) \text{ and } \lambda_{\min}\left(A + tvv^T\right) > l + \delta_L.$$

The proof of this theorem is very similar to the one of Theorem 3.4.1:

Proof.

$$\begin{aligned}
\Phi_{l+\delta_L}\left(A+tvv^T\right) &= \operatorname{Tr}\left(A+tvv^T-l'I\right)^{-1} \\
&= \operatorname{Tr}\left((A-l'I)^{-1}+\frac{t\,(A-l'I)^{-1}\,vv^T\,(A-l'I)^{-1}}{1-tv^T\,(A-l'I)^{-1}\,v}\right) \\
&\quad \text{because of Theorem 2.1.2} \\
&= \operatorname{Tr}(A-l'I)^{-1}+\frac{t\operatorname{Tr}\left(v^T\,(A-l'I)^{-1}\,(A-l'I)^{-1}\,v\right)}{1-tv^T\,(A-l'I)^{-1}\,v} \\
&\quad \text{since Tr is linear and } \operatorname{Tr}(XY)=\operatorname{Tr}(YX) \\
&= \Phi_{l+\delta_L}(A)+\frac{tv^T\,(A-l'I)^{-2}\,v}{1-tv^T\,(A-l'I)^{-1}\,v} \\
&= \Phi_l(A)-\left(\Phi_{l+\delta_L}(A)-\Phi_l(A)\right)+\frac{v^T\,(A-l'I)^{-2}\,v}{1/t-v^T\,(A-l'I)^{-1}\,v} \\
&\leq \Phi_l(A) \\
&\quad \text{since } L_A(v)\geq\frac{1}{t}
\end{aligned}$$

Furthermore, note that $\lambda_{min} > l$ and $\Phi_l(A) \leq \frac{1}{\delta_L}$. This implies that $\lambda_{min}(A+) > l+\delta_L$, so for any $t>0$, $\lambda_{min}(A+tvv^T > l+\delta_L$. □

3.4.3. Shifting the two barriers at once

We have established what happens when we shift the upper barrier and what happens when we shift the lower barrier. For the upper barrier we have quantified how much of vv^T we *can* add and for the lower barrier how much we *must* add. Then big question becomes hence if we can always add more than we must. This turns out to be the case.

So suppose we are in the iteration and have constructed some matrix A_q and have a set of v_i which we haven't picked in the q iterations before. Then the following theorem tells us how to choose a vector v_{q+1} and a positive scalar t_{q+1} to add $t_{q+1}v_{q+1}v_{q+1}^T$ to A_q.

Theorem 3.4.3 (Both Barriers Shift)**.** If $\lambda_{max}(A) < u$, $\lambda_{min}(A) > l$, $\Phi^u(A) \leq \epsilon_U$, $\Phi_l(A) \leq \epsilon_L$ and ϵ_U, ϵ_L, δ_U and δ_L satisfy

$$0 \leq \frac{1}{\delta_U}+\epsilon_U \leq \frac{1}{\delta_L}-\epsilon_L$$

then there exists an i and a positive t for which:

$$L_A(v_i) \geq \frac{1}{t} \geq U_A(v_i)$$

$$\lambda_{max}\left(A+tv_iv_i^T\right) < u+\delta_U$$

$$\lambda_{min}\left(A+tv_iv_i^T\right) > l+\delta_L$$

Proof. We will prove that

$$\sum_i L_A(v_i) \geq \sum_i U_A(v_i).$$

If this is proven, then it follows that there exists an index i and a scalar t such that

$$U_A(v_i) \leq \frac{1}{t} \leq L_A(v_i).$$

Firstly, consider $\sum_i U_A(v_i)$:

$$\sum_i U_A(v_i) = \frac{((u+\delta_U)I - A)^{-2} \bullet \left(\sum_i v_i v_i^T\right)}{\Phi^u(A) - \Phi^{u+\delta_U}(A)} + ((u+\delta_U)I - A)^{-1} \bullet \left(\sum_i v_i v_i^T\right)$$

where $\bullet$ is the inner product.

$$= \frac{\mathrm{Tr}\left((u+\delta_U)I - A\right)^{-2}}{\Phi^u(A) - \Phi^{u+\delta_U}(A)} + \mathrm{Tr}\left((u+\delta_U)I - A\right)^{-1}$$

since $\sum_i v_i v_i^T = I$

$$= \frac{\mathrm{Tr}\left((u+\delta_U)I - A\right)^{-2}}{\Phi^u(A) - \Phi^{u+\delta_U}(A)} + \Phi^{u+\delta_U}(A)$$

$$\leq \frac{\mathrm{Tr}\left((u+\delta_U)I - A\right)^{-2}}{\Phi^u(A) - \Phi^{u+\delta_U}(A)} + \Phi^u(A)$$

since the upper barrier increases when moving the barrier to the left.

$$\leq \frac{\mathrm{Tr}\left((u+\delta_U)I - A\right)^{-2}}{\Phi^u(A) - \Phi^{u+\delta_U}(A)} + \epsilon_U$$

$$= \frac{-\frac{\partial}{\partial_{u+\delta_U}}\Phi^{u+\delta_U}(A)}{\Phi^u(A) - \Phi^{u+\delta_U}(A)} + \epsilon_U$$

$$\leq \frac{-\frac{\partial}{\partial_{u+\delta_U}}\Phi^{u+\delta_U}(A)}{\delta_U\left(-\frac{\partial}{\partial_{u+\delta_U}}\Phi^{u+\delta_U}(A)\right)} + \epsilon_U$$

since the barrier function is convex. [14]

$$= \frac{1}{\delta_U} + \epsilon_U$$

In a similar way we can prove

$$\frac{1}{\delta_L} - \epsilon_L \leq \sum_i L_A(v_i).$$

So putting both inequalities together, we find

$$\sum_i U_A(v_i) \leq \frac{1}{\delta_U} + \epsilon_U \leq \frac{1}{\delta_L} - \epsilon_L \leq \sum_i L_A(v_i).$$

This is exactly what we wanted, because indeed there is then always an i such that

$$U_A(v_i) \leq L_A(v_i)$$

and hence we can always squeeze $1/t$ in between for some $t > 0$.

□

This theorem tells exactly at which step we should choose which v_i and how large we have to take t. The last task to prove Theorem 3.1.3 is to choose the right ϵ_U, ϵ_L, δ_U, δ_L, u_0 and l_0 that satisfy the conditions from Theorem 3.4.3 as well as

$$\frac{\lambda_{\max}(A_Q)}{\lambda_{\min}(A_Q)} \leq \frac{u_0 + dn\delta_U}{l_0 + dn\delta_L} = \frac{d+1+2\sqrt{d}}{d+1-2\sqrt{d}}.$$

These magical values are given by

$$\delta_L = 1 \qquad \epsilon_L = \frac{1}{\sqrt{d}} \qquad l_0 = \frac{-n}{\epsilon_L}$$

$$\delta_U = \frac{\sqrt{d}+1}{\sqrt{d}-1} \quad \epsilon_U = \frac{\sqrt{d}-1}{\sqrt{d}(\sqrt{d}+1)} \quad u_0 = \frac{n}{\epsilon_U}$$

Where we can check the conditions:

$$\frac{1}{\delta_U} + \epsilon_U = \frac{\sqrt{d}-1}{\sqrt{d}+1} + \frac{\sqrt{d}-1}{\sqrt{d}(\sqrt{d}+1)} = 1 - \frac{1}{\sqrt{d}} = \frac{1}{\delta_L} - \epsilon_L$$

and

$$\frac{\lambda_{\max}(A_Q)}{\lambda_{\min}(A_Q)} \leq \frac{u_0 + dn\delta_U}{l_0 + dn\delta_L} = \frac{\frac{n}{\epsilon_U} + dn\delta_U}{\frac{-n}{\epsilon_L} + +dn\delta_L} = \frac{\frac{d+\sqrt{d}}{\sqrt{d}-1} + d\frac{\sqrt{d}+1}{\sqrt{d}-1}}{d-\sqrt{d}} = \frac{d+1+2\sqrt{d}}{d+1-2\sqrt{d}}$$

Performing dn steps of the iteration with these proposed constants will always result in the matrix from Theorem 3.1.3.

3.5. The actual sparsifier

Now that we finally have this result, we can actually prove the existence of the κ-approximations. Suppose we have a graph $G = (V, E, w)$. Without loss of generality we can assume that G is connected. If we arbitrarily orient its edges, we can write its Laplacian as $L_G = B^T W B$ where W is the diagonal matrix with $W(e,e) = w_e$ and B is the *signed edge-vertex incidence matrix*:

$$B(e,v) = \begin{cases} 1 & \text{if } v \text{ is } e\text{'s head} \\ -1 & \text{if } v \text{ is } e\text{'s tail} \\ 0 & \text{otherwise} \end{cases}$$

Another type of matrix that we need is the *Moore-Penrose Pseudoinverse* of a Laplacian L.

Since L is symmetric, we can diagonalise it:

$$L = \sum_{i=1}^{n-1} \lambda_i u_i u_i^T$$

.

The Moore-Penrose Pseudoinverse L^+ is then defined as:

$$L+ = \sum_{i=1}^{n-1} \frac{1}{\lambda_i} u_i u_i^T$$

.

This is called a pseudoinverse since

$$LL^+ = L^+L = \sum_{i=1}^{n-1} u_i u_i^T$$

,

which is the projection onto the span of the nonzero eigenvectors of L. Hence $LL^+ = L^+L$ is the identity on im(L).

So for the rest of the proof we are going to restrict our attention to im$(L_G) \cong \mathbb{R}^{n-1}$. Then we can apply Theorem 3.1.3 to the columns $\{v_i\}_{i \leq m}$ of

$$V = \left(L_G^+\right)^{\frac{1}{2}} B^T W^{\frac{1}{2}}.$$

These columns indeed satisfy

$$\sum_{i\leq m} v_i v_i^T = VV^T = \left(L_G^+\right)^{\frac{1}{2}} B^T W B \left(L_G^+\right)^{\frac{1}{2}}$$
$$= \left(L_G^+\right)^{\frac{1}{2}} L_G \left(L_G^+\right)^{\frac{1}{2}} = I$$

Applying the aforementioned theorem gives us scalars $s_i \geq 0$ that we can write in the diagonal matrix $S(i,i) = s_i$. Set $L_H = B^T W^{\frac{1}{2}} S W^{\frac{1}{2}} B$. This is the Laplacian of the subgraph H of G with edge weights $\{\tilde{w}_i = w_i s_i\}$. Hence H has at most $d(n-1)$ edges. Moreover,

$$I \preceq \sum_{i\leq m} s_i v_i v_i^T = VSV^T \preceq \kappa \cdot I \text{ for } \kappa = \frac{d+1+2\sqrt{d}}{d+1-2\sqrt{d}}.$$

By the Courant-Fischer Theorem, this is equivalent to:

$$1 \leq \frac{y^T VSV^T y}{y^T y} \leq \kappa \quad \forall y \in \text{im}((L_G)^{\frac{1}{2}}) = \text{im}(L_G)$$
$$\Leftrightarrow 1 \leq \frac{y^T (L_G^+)^{\frac{1}{2}} L_H (L_G^+)^{\frac{1}{2}} y}{y^T y} \leq \kappa \quad \forall y \in \text{im}((L_G)^{\frac{1}{2}})$$
$$\Leftrightarrow 1 \leq \frac{x^T L_G^{\frac{1}{2}} (L_G)^{\frac{1}{2}} L_H (L_G^+)^{\frac{1}{2}} L_G^{\frac{1}{2}} x}{x^T L_G^{\frac{1}{2}} L_G^{\frac{1}{2}} x} \leq \kappa \quad \forall x \perp (1,\dots,1)$$
$$\Leftrightarrow 1 \leq \frac{x^T L_H x}{x^T L_G x} \leq \kappa \quad \forall x \perp (1,\dots,1)$$

and therefore H is the κ-approximation we were looking for.

3.6. Conclusion on Twice-Ramanujan Sparsifiers

The greatest challenge for this proof is to keep the eigenvalues relatively close to each other while still being spread out. Since the characteristic polynomial of a matrix *almost* interlaces the one of any rank-one update of the matrix, there is no upper bound on the largest root of this updated matrix. Hence this largest root can shoot off in the far distance which complicates finding a small interval to fit all eigenvalues in.

The solution is to introduce barrier functions, which measure the cumulative distance from the roots to the barrier. The barrier function is originally a function used in physics, but with the physical representation of eigenvalues[1] it can nicely be implemented in this setting as well. This enables us to quantify on one hand how far the zeros can move per iteration and on the other hand how far they have to move. So we can reduce the problem to checking that at each step there is portion of a vector that we can add to end up in this sweet spot. We prove this to be the case and find an explicit expression on how much of each vector we can add.

So the insight that the polynomials are almost interlacing is not what actually let us prove the existence of the sparsifier. Without the iterations and the barrier functions we wouldn't have come very far. These two elements will also be very important in chapter 5. But in that chapter a more detailed notion of interlacing families is needed. This is introduced in the next chapter.

[1] Figure 2.3

4

Restricted Invertibility

The proof of the Twice Ramanujan sparsifiers gave already a nice insight in the case where we have two almost interlacing polynomials. Now we will investigate what happens when we have more polynomials. [12] In that case we can construct an interlacing family. This knowledge will be the base for the next step in chapter 5.

4.1. Restricted Invertibility Theorem

In short, the theorem states that if a matrix has a high *stable rank* then it contains a large column submatrix with large *least singular value*. So what does that mean?

Definition 4.1.1. The *stable rank* of a matrix B is the quotient of its Frobenius norm squared and its 2-norm squared:

$$\text{srank}(B) := \frac{\|B\|_F^2}{\|B\|_2^2}.$$

Where the Frobenius norm $\|B\|_F$ is the square root of the sum of the squares of the entries of B. The 2-norm $\|B\|_2$ is the square root of largest eigenvalue of $B^T B$.

Definition 4.1.2. The *least singular value* of a real matrix B is the smallest eigenvalue of the matrix $B^T B$. Equivalently, it can be defined by

$$\sigma_{min}(B) := \min_{x \neq 0} \frac{\|Bx\|}{\|x\|}.$$

From this second definition it is clear that the least singular value is a measure for how far the matrix is from being singular. When a matrix is singular and has dependent columns, there is always a vector x such that $\|Bx\| = 0$. The smaller the least singular value, the more the columns will come close to being dependent.

Rigorously, the theorem then becomes:

Theorem 4.1.3 (Restricted Invertibility Theorem)**.** Suppose B is an $d \times m$ matrix and $k \leq \text{srank}(B)$ is an integer. Then there exists a subset S of $[m]$ such that

$$\sigma_{min}(B_S)^2 \geq \left(1 - \sqrt{\frac{k}{\text{srank}(B)}}\right)^2 \frac{\|B\|_F^2}{m}.$$

While this theorem holds for any matrix B, we are only going to prove it in a particular case. This case will turn out to be a very nice example on the use of interlacing families without too many technical complications. It is called the isotropic case and happens when $BB^T = I$

4.2. The Isotropic Case

When we assume that $BB^T = I$, the theorem becomes a bit simpler. The Frobenius norm, which is also equal to the square root of the trace of BB^T, then becomes $\sqrt{d}$. The 2-norm of B is clearly 1. Hence srank$(B) = d$ and the bound on the least singular value that we want to prove then becomes:

$$\sigma_{min}(B_S)^2 \geq \left(1 - \sqrt{\frac{k}{d}}\right)\frac{d}{m}$$

The proof will require interlacing families. As already seen in Definition 2.0.3, an interlacing family is a tree of polynomials with specific properties. In order to discuss these polynomials and ease the arguments in this section, we are firstly going to label all the nodes in the tree. A polynomial in the tree at distance t from the root is labelled with with sequences $s_1, \ldots, s_t \in [m]^t$. The empty sequence at the root is denoted by $f_\emptyset$.

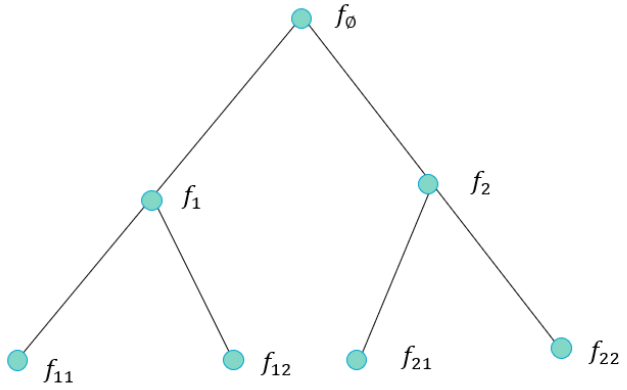


Figure 4.1: Example of a tree with labelling by sequences

The leaves are denoted by sequences of length k: $f_{s_1,\ldots,s_k}$. All the intermediate nodes are labelled by the average of the polynomials of its children.

$$f_{s_1,\ldots,s_t}(x) = \frac{1}{m}\sum_{j=1}^{m} f_{s_1,\ldots,s_t,j}(x)$$

So now there is a clear convention on how to label the polynomials, we can start looking at some properties. The following theorem relates the roots of a polynomial to the roots of its children in the tree.

Theorem 4.2.1. Let $f_1, \ldots, f_m$ be real-rooted degree d polynomials that have a common interlacing. Then for every index $1 \leq j \leq d$ and for every non-negative $\mu_1, \ldots, \mu_m$ such that $\sum_{i=1}^{m} \mu_i = 1$, there exists an a and a b such that

$$\lambda_j(f_a) \geq \lambda_j\left(\sum_i \mu_i f_i\right) \geq \lambda_j(f_b)$$

where $\lambda_j(f)$ denotes the jth root of the polynomial f.

Proof. Of course, when $\mu_i = 0$ for some i, $\mu_i f_i = 0$ and we can just ignore that particular f_i. Therefore we can assume without loss of generality that all μ_i are strictly positive.

The key to solving this problem is to rewrite it as a tree of depth 1. The polynomial given by the convex combination of the polynomials then becomes the root. Let's call its zeros β_i:

$$f_\emptyset(x) = \sum_i \mu_i f_i(x) = \prod_{i=1}^{d}(x - \beta_i)$$

So with this notation, we are looking for a and b for which $\lambda_j(f_a) \geq \beta_j \geq \lambda_j(f_b)$. Consider

$$g(x) = \prod_{i=1}^{d+1}(x - \alpha_i)$$

a polynomial that interlaces every f_i and that has roots α_i. Furthermore, we are firstly going to consider the case where $f_1, \ldots f_m$ do not have any zero in common.

All f_i have positive leading coefficient. That means that for every $i : f_i(x) > 0$ for $x > \beta_1$. Since the f_i have a common interlacing, but no zeros in common, all zeros have multiplicity one. This also implies that $\alpha_1 > \alpha_2 > \ldots > \alpha_{d+1}$. So we can see that $f_i(\alpha_k) \geq 0$ for k odd and $f_i(\alpha_k) \leq 0$ for k even. Moreover there is an i such that $f_i(\alpha_k) \neq 0$, because if not, that would also contradict the fact that there is no zero in common. So given that all μ_i are positive, $f_\emptyset(\alpha_k) > 0$ for k odd and $f_\emptyset(\alpha_k) < 0$ for k even.

Additionally, we know there must be an i such that $f_i(\beta_j) \neq 0$. Since $f_\emptyset(\beta_j) = 0$ and $f_\emptyset$ is the convex sum of the different f_i, there must be an a for which $f_a(\beta_j) < 0$ and a b for which $f_b(\beta_j) > 0$. We can suppose for now that j is odd (the case where j is even is exactly the same but a and b flip). We have already established that $f_a(\alpha_j) \geq 0$, so it must have a zero between β_j and α_j. Since g interlaces f_a, this is the jth largest zero of f_a. Hence $\lambda_j(f_a) \geq \beta_j$.

Similarly, since $f_b(\alpha_{j+1}) \leq 0$, it has to have its jth zero between α_{j+1} and β_j. Therefore $\lambda_j(f_a) \geq \beta_j \geq \lambda_j(f_b)$, which is what we wanted.

To prove the case where $f_1, \ldots, f_m$ have a common zero, we let f_0 be their greatest common divisor. Then we can apply the already proven part to the different $\hat{f}_i = \frac{f_i}{f_0}$, which do not have any zero in common. Then multiplying the $\hat{f}_i$ again with f_0 adds all the same zeros to $f_1, \ldots, f_m$ and $f_\emptyset$. Hence the theorem also holds in this case. □

Please note that this theorem directly implies one side of Theorem 2.0.4: all zeros of any convex combination are within real intervals. The other side is more complicated and will be proven in chapter 5, when all further needed concepts are introduced. For now, we will first improve the above theorem to a slightly stronger result.

Theorem 4.2.2. Let f be an interlacing family of degree d polynomials with root labeled by $f_\emptyset$ and leaves by $\{f_l(x)\}_{l \in L}$. Then for all indices $1 \leq j \leq n$, there exist leaves a and b such that

$$\lambda_j(f_a) \geq \lambda_j(f_\emptyset) \geq \lambda_j(f_b)$$

Proof. We can use induction on the height of the tree and Theorem 4.2.1. If the tree height is 1, the result immediately follows. If the tree has a greater height, the theorem states that there are children of the root $v_{\hat{a}}$ and $v_{\hat{b}}$ whose jth zero satisfies the inequality. Since $v_{\hat{a}}$ is not a leaf, it is the root of its own interlacing family and hence by the induction hypothesis should have a leaf v_a such that

$$\lambda_j(v_a) \geq \lambda_j(v_{\hat{a}}) \geq \lambda_j(f_\emptyset).$$

This argument can be repeated for $v_{\hat{b}}$. □

The theorems above provide us the tools to prove Theorem 4.1.3. First of all, $BB^T = I$ is equivalent to $\sum_i u_i u_i^T = I$ with u_i the columns of B. That means that for a set S of size k of these columns:

$$\sigma_{min}(B_S)^2 = \lambda_k(B_S^T B_S) = \lambda_k(B_S B_S^T) = \lambda_k\left(\sum_{i \in S} u_i u_i^T\right)$$

The main idea of the proof will be to pick the columns of B at random to form the column submatrix B_S. After having picked a column we will replace it amongst the choices, so it will be possible to add a column multiple times to B_S. This may not seem as the best thing to do: when a column is picked twice, $\sigma_{min}(B_S) = 0$ because it clearly has dependent columns. However, this case will not bother us too much as we are going to prove there is always a way to pick the columns in such a way that the inequality holds.

To do so, consider the expected characteristic polynomials of the outer products of the chosen vectors. These polynomials will form an interlacing family where the expected characteristic polynomials after having chosen k vectors from m possible choices form the leaves.

$$f_{s_1,\dots,s_k}(x) := \det\left[xI - \sum_{i=1}^{k} u_{s_i} u_{s_i}^T\right]$$

The root polynomial then becomes the weighted sum over all these leaves:

$$f_\emptyset(x) = \frac{1}{m^k} \sum_{s_1,\dots,s_k \in [m]^k} f_{s_1,\dots,s_k}(x)$$

So the different levels in the tree correspond to the number of picked vectors. At the root, we haven't picked anything yet, so we have the average polynomial of all different possible characteristic polynomials. At the first level, the jth node corresponds to the average polynomial of the characteristic polynomials given that we picked the jth vector as a first column.

This tree forms an interlacing polynomial by Lemma 2.1.4: each level of the tree has a common interlacing and every non-leaf node is a convex combination of its children. And in the isotropic case the tree actually becomes even nicer: the root polynomial has an explicit formula. [12]

$$f_\emptyset(x) = \left(1 - \frac{1}{m}\partial_x\right)^k x^d$$

From this formula one can see that as long as $d \geq k$, the polynomial $f_\emptyset$ is divisible by x^{d-k}. So the kth largest root of $f_\emptyset$ is equal to the smallest zero of

$$x^{-(d-k)} f_\emptyset(x) = x^{-(d-k)} \left(1 - \frac{1}{m}\partial_x\right)^k x^d.$$

This is the step that shows why this isotropic case is so interesting. The expression resembles the famous Laguerre polynomial of degree n and parameter α: [18]

$$L_n^{(\alpha)} = \frac{x^{-\alpha}}{n!}(\partial_x - 1)^n x^{n+\alpha}.$$

Indeed,

$$x^{-(d-k)} f_\emptyset(x) = (-1)^k \frac{k!}{m^k} L_k^{(d-k)}(mx).$$

This link between our root polynomial and the Laguerre polynomial is quite surprising and shows the great strength of this method. The root polynomial is an average over characteristic polynomials, while the Laguerre polynomial is the solution to a differential equation. This polynomial appears in different fields in mathematics: quantum mechanics, random matrix theory and now also interlacing families. This raises the idea that there is a larger mathematical structure behind this method of interlacing families. Moreover, since these polynomials are heavily studied, there are already known bounds on the roots, which is exactly what we need.

More specifically, we are going to use the bound on the smallest root of a Laguerre polynomial due to Krasikov. [10]

Theorem 4.2.3 (Krasikov's bound)**.** For $\alpha > -1$,

$$\lambda_k(L_k^{(\alpha)}(x)) \geq V^2 + 3V^{4/3}(U^2 - V^2)^{-1/3}$$

where $V = \sqrt{k+\alpha+1} - \sqrt{k}$ and $U = \sqrt{k+\alpha+1} + \sqrt{k}$.

Corollary 4.2.4.

$$\lambda_k(f_\emptyset(x)) > \frac{1}{m}(\sqrt{d}-\sqrt{k})^2$$

Proof. When applying Theorem 4.2.3 with $\alpha = d-k$ and hence $V = (\sqrt{d+1}-\sqrt{k})$, we obtain

$$\lambda_k(f_\emptyset(x)) = \lambda_k(L_k^{(d-k)}(mx)) = \frac{1}{m}\lambda_k(L_k^{(d-k)}(x)) \geq V^2 > \frac{1}{m}(\sqrt{d}-\sqrt{k})^2.$$

□

So now that we have a bound on the zeros of the root polynomial, we can use Theorem 4.2.2 to tell something about the zeros of one of the leaf polynomials: there exists a sequence $s_1, \ldots, s_k, \in [m]^k$ for which

$$\lambda_k(f_{s_1,\ldots,s_k}) \geq \lambda_k(f_\emptyset) > \frac{(\sqrt{d}-\sqrt{k})^2}{m}.$$

We know that $f_{s_1,\ldots,s_k}$ is the characteristic polynomial of $\sum_{i=1}^k u_{s_i}u_{s_i}^T$, so it is impossible to have picked the same vector twice. If that would be the case, the matrix would have at most rank $k-1$, which implies $\lambda_k = 0$.

This shows that there has to be a set $S \subset [m]$ of k different vectors such that

$$\sigma_{min}(B_S)^2 = \lambda_k\left(\sum_{i\in S} u_i u_i^T\right) > \frac{(\sqrt{d}-\sqrt{k})^2}{m} = \left(1-\sqrt{\frac{k}{d}}\right)^2 \frac{d}{m}$$

which is exactly the Restricted Invertibility Theorem.

4.3. Conclusion on the Restricted Invertibility Theorem

This chapter focuses on interlacing families as trees of polynomials. The first important result holds for any such tree. It relates the zeros of the roots with the zeros of the leafs: every jth zero of the root polynomial is bounded above by the jth zero of a leaf polynomial and bounded below by the jth zero of another leaf polynomial. This theorem will also play a crucial role in the next chapter.

Next, we concentrate on a specific choice for the polynomials. Just as in the previous chapter, we use the characteristic polynomial of a rank-one update of a matrix. The greatest difference is now that we have multiple choices in which vector is used to construct this update, since we don't need to aim the eigenvalues in an interval. Therefore we keep track of all possible matrices by putting them in a tree. One can walk from the root to the leaf by at each node picking one of the vectors, with replacement. The final leaf polynomial is then the characteristic polynomial of the matrix obtained by all these rank-one updates.

It turns out that the root polynomial in the Isotropic case of the Restricted Invertibility Theorem has a very nice explicit expression, which is similar to the Laguerre polynomial. This is a crucial point in the proof. The link between on one hand the characteristic polynomial of a matrix and on the other hand the Laguerre polynomial, which is a solution to a differential equation. This is an indicative example that interlacing polynomials have a strong structure and appear in different settings.

Since Laguerre polynomials are heavily studied, there is a known bound on its zeros. Using then the relation with the zeros of the leaf polynomials we proved the existence of a choice of vectors giving the column submatrix in the theorem.

5

The Kadison-Singer problem

The last theorem that we are going to prove with the method of interlacing polynomials is probably the most important one: the Kadison-Singer problem. [9] This problem posed by Kadison and Singer in 1959 remained unsolved for years until in 2014 Marcus, Spielman and Srivastava finally found a solution thanks to their interlacing method. [14] In this chapter, we combine the concepts from chapter 3 with those from chapter 4. Specifically, we will use a proof with iteration and barrier functions, but this time the polynomials will be multivariate instead of univariate. This will complicate things to keep track of locations of the zeros and bring us back to interlacing families.

Question 5.0.1 (Kadison-Singer Problem)**.** Does every pure state on the (abelian) von Neumann algebra $\mathbb{D}$ of bounded diagonal operators on l_2 have a unique extension to a pure state on $B(l_2)$, the von Neumann algebra of all bounded operators on l_2?

This question is not similar at all to the previous theorems that we have proven: it seems more analytical and algebraic. And that is also how, at first, this problem was approached. The Kadison-Singer problem became famous when one of the more known analytical mathematicians of that time, Jean Bourgain, didn't succeed in proving the positive result. [21] [20]

Using modern analytical methods was apparently not the right approach. This was the realisation of Weaver, so he went for a different approach. [25] He proposed the following conjecture, called KS_2.

Conjecture 5.0.2 (KS_2)**.** There exists universal constants $\eta \geq 2$ and $\theta > 0$ so that the following holds. Let $w_1, \ldots, w_n \in \mathbb{C}^d$ satisfy $\|w_i\| \leq 1$ for all i and suppose

$$\sum_{i=1}^{m} |\langle u, w_i \rangle|^2 = \eta$$

for every unit vector $u \in \mathbb{C}^d$. Then there exists a partition S_1, S_2 of $\{1, \ldots, m\}$ so that

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

for every unit vector $u \in \mathbb{C}^d$ and each $j \in \{1, 2\}$.

This conjecture is equivalent to the Kadison-Singer problem, yet it does not look very similar. And that is the great strength of Weaver's Conjecture. Because of this new formulation, Weaver hoped that more discrete mathematical methods could be used to finally prove the problem.

And this turns out to be the case. As said, the proof was given by Marcus, Spielman and Srivastava by the following theorem.

Theorem 5.0.3 (Marcus-Spielman-Srivastava Theorem)**.** If $\epsilon > 0$ and $v_1, \ldots, v_m$ are independent random vectors in $\mathbb{C}^d$ with finite support such that

$$\sum_{i=1}^{m} \mathbb{E} v_i v_i^* = I$$

and

$$\mathbb{E}\|v_i\|^2 \leq \epsilon \text{ for all } i,$$

then

$$\mathbb{P}\left[\left\|\sum_{i=1}^{m} v_i v_i^*\right\| \leq (1+\sqrt{\epsilon})^2\right] > 0.$$

This theorem is stated in a way that will enable us to prove Weaver's Conjecture. However, with the current phrasing, it is hard to use the method of interlacing polynomials: there are no polynomials in the statement and hence certainly no roots of polynomials.

The largest part of the proof will actually be to step-by-step rewrite it in similar statements that when tracing back our steps will eventually imply the theorem above. First of all we are going to replace the $v_i v_i^*$ by Hermitian rank-one matrices A_i.

Theorem 5.0.4. Let $A_1, \ldots, A_m$ be jointly independent random rank-one Hermitian positive semi-definite $d \times d$ matrices such that the sum $A := \sum_{i=1}^{m} A_i$ has mean

$$\mathbb{E} A = I$$

and such that

$$\mathbb{E}\,\mathrm{Tr}[A_i] \leq \epsilon$$

for some $\epsilon > 0$ and all $i = 1, \ldots m$.

Then

$$\mathbb{P}\left[\|A\| \leq (1+\sqrt{\epsilon})^2\right] > 0.$$

This makes it easier to include polynomials. Since for a Hermitian positive semi-definite matrix $\|A\|$ is the largest eigenvalue of A, we will prove $\lambda_{max}\left(p_A(x)\right) \leq (1+\sqrt{\epsilon})$ with positive probability. [19] However, we don't have a lot of information on this characteristic polynomial. Since A is the sum of all random matrices, it is hard to directly find a bound on its eigenvalues. Therefore we are going to use the following result to reformulate the theorem once more. But this time, the rephrasing is a bit more complicated and we will need two new formulas.

5.1. Mixed characteristic polynomial

First of all, we are going to use a new formula for the characteristic polynomial $p_A(x)$ of a matrix A. [14]

Proposition 5.1.1. Let A be the sum of deterministic rank-one $d \times d$ matrices $A_1, \ldots, A_m$. Then

$$p_A(x) = \left(\prod_{i=1}^{m} 1 - \partial_{z_i}\right) \det\left(xI + \sum_{i=1}^{m} z_i A_i\right)\Big|_{z_1 = \ldots = z_m = 0}$$

This polynomial is called the *mixed characteristic polynomial* and is denoted by $\mu\left[A_1, \ldots, A_m\right](x)$. The main reason why this formula is interesting is because now we include that A is the sum over all the different A_i. Another advantage is that if we take the expected characteristic polynomial we get the following lemma.

Lemma 5.1.2. Let A be the sum of jointly independent rank-one $d \times d$ matrices $A_1, \ldots, A_m$. Then we have

$$\mathbb{E}p_A(x) = \mu\left[\mathbb{E}A_1, \ldots, \mathbb{E}A_m\right](x).$$

Proof. Since x is not a random variable, we can keep it fixed. In that case $\det\left(xI + \sum_{i=1}^m z_i A_i\right)$ is a polynomial combination of the different $z_i A_i$. Furthermore, $\prod_{i=1}^m (1 - \partial_{z_i})$ is a linear combination of the differential operators $\frac{\partial^j}{\partial_{i_1} \ldots \partial_{i_j}}$ for $1 \leq i_1 < \ldots < i_j \leq d$. So we can rewrite the mixed characteristic polynomial as a linear combination of terms. And each of these terms is a multilinear combination of $A_{i_1}, \ldots, A_{i_j}$ for $1 \leq i_1 < \ldots < i_j \leq d$. So if we take the expectation at both sides, we firstly take the expectation over the different terms. Of course this will result in the sum over the expectations. Then using the joint independence of the different A_i, the result follows. □

This lemma seems very promising: while we have no info on what A looks like, we do know that $\mathbb{E}A = \sum_i \mathbb{E}A_i = I$. Using this condition, we can simplify the mixed characteristic polynomial.

Lemma 5.1.3. Let $A_1, \ldots, A_m$ be Hermitian positive semidefinite matrices satisfying $\sum_i A_i = I$. In that case

$$\mu\left[A_1, \ldots, A_m\right](x) = \left(\prod_{i=1}^m 1 - \partial_{y_i}\right) \det\left(\sum_{i=1}^m y_i A_i\right)\Big|_{y_1 = \ldots = y_m = x}$$

Proof. The expression for the mixed characteristic polynomial was given by

$$\begin{aligned}
\mu[A_1, \ldots, A_m](x) &= \left(\prod_{i=1}^m 1 - \partial_{z_i}\right) \det\left(xI + \sum_{i=1}^m z_i A_i\right)\Big|_{z_1 = \ldots = z_m = 0} \\
&= \left(\prod_{i=1}^m 1 - \partial_{z_i}\right) \det\left(\sum_{i=1}^m (z_i + x) A_i\right)\Big|_{z_1 = \ldots = z_m = 0} \\
&= \left(\prod_{i=1}^m 1 - \partial_{y_i}\right) \det\left(\sum_{i=1}^m y_i A_i\right)\Big|_{y_1 = \ldots = y_m = x}
\end{aligned}$$

Where the last expression is obtained by substituting $y_i = z_i + x$. □

5.2. Comparison with mean

It seems a good idea to rewrite the theorem to a statement including the expected characteristic polynomial $\mathbb{E}p_A(x)$. The following theorem will allow us to do so, but to prove it on its turn we will need some new concepts and results.

Theorem 5.2.1. Let A be a random matrix, which is the sum of $A = \sum_{i=1}^m A_i$ of m independent Hermitian rank-one $d \times d$ matrices A_i, each taking a finite number of values. Then,

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

and

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

The most important notion needed is the one of stable polynomials.

Definition 5.2.2. A polynomial $p : \mathbb{C}^m \to \mathbb{C}$ is stable if it has no zeros in the region $\{(z_1, \ldots, z_m) : \mathrm{Im}(z_1) > 0, \ldots, \mathrm{Im}(z_m) > 0\}$.

A polynomial is real stable if it is stable and has real coefficients.

Note that when a polynomial of one variable is real stable, the polynomial is real-rooted because of the symmetry of the roots. Only a minority of polynomials is real stable, but if you ever encounter them in the wild, it is a reason to be very happy. They have very nice properties that we are going to exploit. The polynomials that we are going to consider are real stable. [3]

Proposition 5.2.3. If $A_1, \ldots, A_m$ are positive semi-definite Hermitian matrices, then the polynomial

$$\det\left(\sum z_i A_i\right)$$

is real stable.

Now we are mostly interested in the mixed characteristic polynomial, which still has a product of all $1 - \partial_{z_i}$ in front of it. To investigate in which cases the mixed characteristic is stable, the next theorem will turn out to be useful. [15]

Theorem 5.2.4. If all the zeros of a degree d polynomial $q(z)$ lie in a closed circular region A, then for $\lambda \in \mathbb{C}$, all zeros of

$$q(z) - \lambda q'(z)$$

lie in the convex region obtained by translating A in the magnitude and direction of the vector $d\lambda$.

We are particularly interested in what happens when we take $\lambda = 1$.

Corollary 5.2.5. If $p : \mathbb{R}^m \to \mathbb{R}$ is real stable, then so is

$$(1 - \partial_{z_1})p(z_1, \ldots, z_m)$$

.

Proof. Let $x_2, \ldots, x_m$ be numbers with positive imaginary part. Then the univariate polynomial

$$q(z_1) = p(z_1, z_2, \ldots, z_m)|_{z_2 = x_2, \ldots, z_m = x_m}$$

is stable. If that would not be the case, it would imply that p has a root where all z_i have positive imaginary part. That contradicts the stability of p. Hence all zeros of q lie in the circular region consisting of numbers with negative imaginary part. Since translating this region by d will once again give a region of numbers with negative imaginary part, $(1-\partial_{z_i})q(z)$ is stable. Hence $(1-\partial_{z_i})p$ has no roots of which all variables have positive imaginary part and is stable as well. □

Another property of real stable polynomials that we will need is preservation when setting variables to real numbers. [23]

Proposition 5.2.6. If $p : \mathbb{R}^m \to \mathbb{R}$ is real stable and $a \in \mathbb{R}$, then $p|_{z_1 = a} : \mathbb{R}^{m-1} \to \mathbb{R}$ is real stable as well.

If we combine the above results, the following corollary follows immediately:

Corollary 5.2.7. If $A_1, \ldots, A_m$ are positive semi-definite matrices then $\mu\,[A_1, \ldots, A_m]\,(x)$ is real stable.

One of the reasons why this real stability is interesting, is because there is a clear link with interlacing polynomials.

It will actually enable us to finally prove Theorem 2.0.4. [17]

Proof of Theorem 2.0.4. As already mentioned, Theorem 4.2.1 proves that if $f_1, \ldots, f_m$ are real-rooted polynomials of the same degree with positive leading coefficients and with a common interlacing, then $\sum_i t_i f_i$ is real rooted for any convex combination of t_i. So now we still have to prove the converse.

Consider two polynomials f and g such that $h_t = tf + (1-t)g$ has real roots for any $0 \leq t \leq 1$. We will show that f and g have a common interlacing.

To start, suppose f and g have no common zero and the zeros have multiplicity one. In this case, when t goes from 0 to 1, the roots of h_t trace from the roots of g to the roots of f. The n different roots hence trace in n different intervals I_i and each I_i contains exactly one root from g and one from f. If such an interval would without loss of generality contain two zeros from g, there would be a $t \neq 0$ and an $x \in \mathbb{R}$ such that $h_t(x) = g(x) = 0$. This implies that $f(x) = 0$, contradicting the assumption of having no common zeros. Therefore the I_i should have pairwise disjoint interiors and f and g have a common interlacing. For the case where there are roots in common, we apply the same reasoning as in Theorem 4.2.1: let f_0 be their greatest common divisor.

Then we can apply the already proven part to $\hat{f} = \frac{f}{f_0}$ and $\hat{g} = \frac{f}{f_0}$, which do not have any zero in common. Then multiplying all the $\hat{f_i}$ again with f_0 adds all the same zeros to f and g, which we can also add to the common interlacing. Hence the theorem also holds in this case.

The last thing we have to deal with is root multiplicity. We need an intermediate result. Consider an $\epsilon > 0$ and f a real-rooted polynomial of degree n. Then the polynomial

$$f_\epsilon = (1 - \epsilon\partial)^n f$$

is real-rooted and has simple roots.

We have already established that f_ϵ is real rooted, so now we still have to count the roots. Consider $(1 - \epsilon\partial)f$. This polynomial is zero whenever $\frac{f'}{f} = 1/\epsilon$. Furthermore $\frac{f'}{f} = \sum_i \frac{m_i}{x - z_i}$ where m_i is the multiplicity of root z_i of function f. This expression allows us to do the exact same analysis as we did for Figure 2.2. Hence we find that $\frac{f'}{f} = 1/\epsilon$ exactly once between each root of f. So applying $(1 - \epsilon\partial)$ to f decreases the multiplicity of each root by 1. Applying it n times ensures that we have simple roots.

Therefore we can consider a sequence of polynomials f_ϵ and g_ϵ defined as above. These have simple, real roots for any $\epsilon > 0$ and converge respectively uniformly to f and g on any bounded interval as $\epsilon \to 0$. Using continuity on roots as a function of the coefficients, we can conclude that the theorem also holds if f or g has roots with multiplicity ≥ 2.

Lastly, we have to prove the case where we have more than two polynomials. Suppose that we have $f_1, \ldots, f_m$ such that for any non-negative convex combination of t_i, $\sum_i t_i f_i$ is real rooted. Then by choosing two of the t_i to be 0.5 implies that any pair of f_i have a common interlacing. This implies in its turn that all f_i have a common interlacing.

To see this, suppose there is no common interlacing. Then it is impossible to construct closed intervals such that there is exactly one root of each polynomial in each interval. So that means there are two polynomials f_i and f_j such that

$$\lambda_a(f_i) \leq \lambda_{a+1}(f_i) \leq \lambda_a(f_j)$$

for some a which obstruct the construction of the above mentioned intervals. But this contradicts the common interlacing of f_i and f_j.

□

Theorem 4.2.1 will also enable us to link the characteristic polynomial with its expectation: Theorem 5.2.1 follows by iterating the following result.

Lemma 5.2.8. Let $A_1, \ldots, A_m$ be jointly independent random rank-1 positive semi-definite Hermitian matrices, with each A_i taking finitely many values. For any $1 \leq j < m$ and any fixed choice $A_1, \ldots, A_{j-1}$ we have

$$\mathbb{P}\left[\lambda_{max}\left(\mu[A_1, \ldots, A_{j-1}, A_j, \mathbb{E}A_{j+1}, \ldots \mathbb{E}A_m]\right) \leq \lambda_{max}\left(\mu[A_1, \ldots, A_{j-1}, \mathbb{E}A_j, \mathbb{E}A_{j+1}, \ldots \mathbb{E}A_m]\right)\right] > 0$$

and

$$\mathbb{P}\left[\lambda_{max}\left(\mu[A_1, \ldots, A_{j-1}, A_j, \mathbb{E}A_{j+1}, \ldots \mathbb{E}A_m]\right) \geq \lambda_{max}\left(\mu[A_1, \ldots, A_{j-1}, \mathbb{E}A_j, \mathbb{E}A_{j+1}, \ldots \mathbb{E}A_m]\right)\right] > 0$$

Proof. $\mathbb{E}A_j$ is a convex combination over all values of A_j that occur with positive probability. Hence by the same properties we exploited for Lemma 5.1.2, we obtain that $\mu[A_1, \ldots, A_{j-1}, \mathbb{E}A_j, \mathbb{E}A_{j+1}, \ldots \mathbb{E}A_m]$ is a convex combination of the $\mu[A_1, \ldots, A_{j-1}, A_j, \mathbb{E}A_{j+1}, \ldots \mathbb{E}A_m]$ over the possible A_j. Moreover, by Corollary 5.2.7 all possible combinations are real stable. Therefore we can construct an interlacing family of level one with $\mu[A_1, \ldots, A_{j-1}, \mathbb{E}A_j, \mathbb{E}A_{j+1}, \ldots \mathbb{E}A_m]$ as root polynomial.

The children are then the $\mu[A_1, \ldots, A_{j-1}, A_j, \mathbb{E}A_{j+1}, \ldots \mathbb{E}A_m]$ for the different possibilities of A_j. Hence we can apply Theorem 4.2.1 to obtain the result we want. □

5.3. Proof of Weaver's Conjecture

Combining section 5.1 and section 5.2, Theorem 5.0.4 becomes:

Theorem 5.3.1. Let $A_1, \dots, A_m$ be jointly independent random rank-one Hermitian positive semi-definite $d \times d$ matrices such that the sum $A := \sum_{i=1}^{m}$ has mean

$$\mathbb{E}A = I$$

and such that

$$\mathbb{E}\,\mathrm{Tr}[A_i] \leq \epsilon$$

for some $\epsilon > 0$ and all $i = 1, \dots, m$.

Then

$$\lambda_{max}(\mathbb{E}p_A(x)) \leq \left(1 + \sqrt{\epsilon}\right)^2$$

which would indeed imply the Marcus-Spielman-Srivastava Theorem.

So now we are going to use our built up knowledge on the mixed characteristic polynomial to plug in the explicit expression of $\mathbb{E}p_a(x)$. The problem is that we will replace a univariate polynomial by a multivariate one. That means that an upper bound for the largest root is not well defined anymore, since we are dealing with vectors. Therefore we will use the notion of being *above the roots*.

Definition 5.3.2. Let $p(z_1, \dots, z_m)$ be a multivariate polynomial. We say that $z \in \mathbb{R}^m$ is above the roots of p if

$$p(z + t) > 0 \text{ for all } t = (t_1, \dots, t_m) \in \mathbb{R}^m, t_i \geq 0$$

So if we can prove that the vector $\left((1 + \sqrt{\epsilon})^2, \dots, (1 + \sqrt{\epsilon})^2\right)$. is above the roots of

$$Q(x, \dots, x) := \mathbb{E}p_A(x) = \mu[\mathbb{E}A_1, \dots, \mathbb{E}A_m](x) = \left(\prod_{i=1}^{m} 1 - \partial_{y_i}\right) \det\left(\sum_{i=1}^{m} y_i \mathbb{E}A_i\right)\Big|_{y_1 = \dots = y_m = x}$$

then we prove the theorem above.

To do this, we use the same idea as in chapter 3: we are going to iteratively apply an operator to a polynomial while keeping track of the roots of the zeros. More precisely, we are going to start with the polynomial

$$\det\left(\sum_{i=1}^{m} y_i A_i\right)$$

and the operator at hand is multiplying with $(1 - \partial_i)$ for some direction i that hadn't been used before.

In short, we are going to prove the following theorem:

Theorem 5.3.3. Let $A_1, \dots, A_m$ be symmetric positive semi-definite $d \times d$-matrices such that

$$\sum_{i=1}^{m} A_i = I$$

and

$$\mathrm{Tr}[A_i] \leq \epsilon$$

for some $\epsilon > 0$ and all $i = 1, \dots, m$.

Let $p(z_1, \dots, z_m)$ be the polynomial

$$p(z_1, \dots, z_m) := \det\left(\sum_{i=1}^{m} z_i A_i\right)$$

.

Then $\left((1 + \sqrt{\epsilon})^2, \dots, (1 + \sqrt{\epsilon})^2\right)$ lies above the roots of $\left(\prod_{i=1}^{m}(1 - \partial_i)\right) p$.

From the previous section, we know that the starting polynomial p is real stable and multiplying with $(1-\partial_i)$ keeps it that way. That means that at no iteration there are roots of which all variables have positive imaginary parts.

However, that still doesn't give a lot of insight on how the roots behave during the iterations. Therefore we are going to reintroduce the barrier functions from section 3.3. But in the mentioned section the upper and lower barrier were enough because the polynomials were univariate. For the multivariate polynomial now, we need redefine the barrier functions to keep track of the zeros in all directions.

Definition 5.3.4. Given a real stable polynomial p and a point $z = (z_1, \ldots, z_m)$ above the roots of p. The barrier function of p in direction i at z is

$$\Phi_p^i(z) = \frac{\partial_{z_i} p(z)}{p(z)} = \partial_{z_i} \log p(z)$$

.

There is an equivalent definition that resembles more Definition 3.3.1:

$$\Phi_p^i(z_1, \ldots, z_m) = \frac{q'_{z,i}(z_i)}{q_{z,i}(z_i)} = \sum_{j=1}^{r} \frac{1}{z_i - \lambda_j}$$

where the univariate restriction

$$q_{z,i}(t) = p(z_1, \ldots, z_{i-1}, t, z_{i+1}, \ldots, z_m)$$

has roots $\lambda_1, \ldots, \lambda_r$ which are real by Proposition 5.2.6.

Once again, the barrier functions are very useful. Two nice properties that they have are monotonicity and convexity: [14]

Proposition 5.3.5. Suppose p is real stable and z is above the roots of p. Then for all $i, j \leq m$ and $\delta \geq 0$:

- $\Phi_p^i(z + \delta e_j) \leq \Phi_p^i(z)$

(monotonicity)

- $\Phi_p^i(z + \delta e_j) \leq \Phi_p^i(z) + \delta \cdot \partial_{z_i} \Phi_p^i(z + \delta e_j)$

(convexity)

From these properties, we almost immediately find a result that tells us something about the roots when applying the operator.

Lemma 5.3.6. Suppose that p is real stable, that z is above its roots and that $\Phi_p^i(z) < 1$. Then z is above the roots of $p - \partial_{z_i} p$

Proof. Let t have non-negative entries. Using the monotonicity of Φ_p^i we obtain:

$$\begin{aligned} &\Phi_t^i(z + t) < 1 \\ \Rightarrow\ &\partial_{z_i} p(z + t) < p(z + t) \\ \Rightarrow\ &(p - \partial_{z_i} p)(x + t) > 0 \end{aligned}$$

□

While this lemma allows us to prove that a vector is above the roots of $p - \partial_i p$, it is not strong enough to iterate this argument. The problem is that we have no control over $\Phi^j_{p-\partial_i p}$ for any other direction j. So it is very well possible that $\Phi^j_{p-\partial_i p}(z)$ becomes larger than one and then we are stuck.

The solution to this problem lies in taking another bound on the barrier functions. [14]

Proposition 5.3.7. Suppose that $p(z_1, ..., z_m)$ is real stable, that x is above its roots and that $\delta > 0$ satisfies

$$\Phi_p^j(x) \leq 1 - \frac{1}{\delta}.$$

Then $x + \delta e_j$ lies above the roots of $p - \partial_j p$ and furthermore,

$$\Phi_{p-\partial_j p}^i(x + \delta e_j) \leq \Phi_p^i(x)$$

for all $1 \leq i \leq m$.

This proposition is the one we need, because now we know that at each step the barrier function will decrease in every direction. Hence we can require that the barrier function is less than $1 - 1/\delta$ for all directions and apply the result iteratively to easily find

Lemma 5.3.8. Suppose that $p(z_1, ..., z_m)$ is real stable, that $x = (x_1, ..., x_m)$ is above its roots and that $\delta > 0$ satisfies for all $1 \leq j \leq m$

$$\Phi_p^j(x) \leq 1 - \frac{1}{\delta}.$$

Then $x + (\delta, ..., \delta)$ lies above the roots of $\left(\prod_{i=1}^m (1 - \partial_i)\right) p$.

So to prove Theorem 5.3.3 using the above lemma, the only elements that we still need to find are a vector x that lies above the roots of

$$\det\left(\sum_{i=1}^m z_i A_i\right)$$

and a $\delta > 0$ such that

$$\Phi_p^j(x) \leq 1 - \frac{1}{\delta}$$

for every direction j.

Proof of Theorem 5.3.3. Let

$$p(z_1, ..., z_m) = \det\left(\sum_{i=1}^m z_i A_i\right)$$

It turns out that finding a vector that is above the roots of p is not that hard. Since all matrices A_i are positive semidefinite and for any $t > 0$

$$\det\left(t \sum_i A_i\right) = \det(tI) > 0$$

the all-t vector $(t, ..., t)$ lies above the roots p as long as t is strictly positive.

Finding an appropriate δ requires some more work, but to keep the essence in the proof, the technical computations with identities from linear algebra are left out. One can verify that the barrier function for the specific polynomial becomes:

$$\Phi_p^i(z_1, ..., z_m) = \mathrm{Tr}\left(\left(\sum_{i=1}^m z_i A_i\right)^{-1} A_i\right)$$

So if we plug in the all-t vector in this barrier function we obtain

$$\Phi_p^i(t, ..., t) = \frac{\mathrm{Tr}(A_i)}{t} \leq \frac{\epsilon}{t}$$

for any direction i.

Hence if we choose t and δ such that $\frac{\epsilon}{t}+\frac{1}{\delta}\leq 1$, Lemma 5.3.8 tells us that $(t+\delta,\dots,t+\delta)$ lies above the roots of $\left(\prod_{i=1}^{m}(1-\partial_i)\right)p$.

Choosing $t=\sqrt{\epsilon}+\epsilon$ and $\delta=1+\sqrt{\epsilon}$ gives

$$\frac{\epsilon}{t}+\frac{1}{\delta}=1$$

and

$$t+\delta=(1+\sqrt{\epsilon})^2$$

as required.

□

As previously discussed, from Theorem 5.3.3 we can work our way back to eventually proof the original formulation of the Marcus-Spielman-Srivastava Theorem (Theorem 5.0.3).

So the only thing left to prove is Weaver's Conjecture KS_2, starting from this theorem.

To do this, we need one more corollary:

Corollary 5.3.9. Let r be a positive integer and let $u_1,\dots,u_m\in\mathbb{C}^d$ be vectors such that

$$\sum_{i=1}^{m}u_iu_i^*=I$$

and $\|u_i\|^2\leq\delta$ for all i. Then there exists a partition $\{S_1,\dots,S_r\}$ of $[m]$ such that

$$\left\|\sum_{i\in S_j}u_iu_i^*\right\|\leq\left(\frac{1}{\sqrt{r}}+\sqrt{\delta}\right)^2 \text{ for } j=1,\dots,r.$$

Proof. For each $i\in[m]$ and $k\in[r]$, define $w_{i,k}$ as the direct sum of r vectors from $\mathbb{C}^d$, where every vector is the 0-vector in $\mathbb{C}^d$ except for the kth one, which is a copy of u_i. That is,

$$w_{i,1}=\begin{pmatrix}u_i\\0^d\\\vdots\\0^d\end{pmatrix},w_{i,2}=\begin{pmatrix}0^d\\u_i\\\vdots\\0^d\end{pmatrix},\dots,w_{i,r}=\begin{pmatrix}0^d\\\vdots\\0^d\\u_i\end{pmatrix}.$$

Next, we let $v_1,\dots,v_m$ be independent random vectors where v_i takes values $\{\sqrt{r}w_{i,k}\}_{k=1}^{r}$ each with probability $\frac{1}{r}$. We define these vectors in this way because then they satisfy

$$\mathbb{E}v_iv_i*=\begin{pmatrix}u_iu_i* & 0_{d\times d} & \dots & 0_{d\times d}\\0_{d\times d} & u_iu_i^* & \dots & 0_{d\times d}\\\vdots & & \ddots & \vdots\\0_{d\times d} & 0_{d\times d} & \dots & u_iu_i^*\end{pmatrix}.$$

So

$$\sum_{i=1}^{m}\mathbb{E}v_iv_i^*=I_{rd}.$$

Furthermore,

$$\|v_i\|^2=r\|u_i\|^2\leq r\delta$$

which means we can apply Theorem 5.0.3 with $\epsilon=r\delta$ and find an assignment of each v_i so that

$$\left(1+\sqrt{r\delta}\right)^2\geq\left\|\sum_{i=1}^{m}v_iv_i^*\right\|=\left\|\sum_{k=1}^{r}\sum_{i:v_i=w_{i,k}}\left(\sqrt{r}w_{i,k}\right)\left(\sqrt{r}w_{i,k}\right)^*\right\|.$$

So if we now set $S_k = \{i : v_i = w_{i,k}\}$, we obtain for every k:

$$\left\| \sum_{i \in S_k} u_i u_i^* \right\| = \left\| \sum_{i \in S_k} w_{i,k} w_{i,k}^* \right\| \leq \frac{1}{r} \left\| \sum_{k=1}^{r} \sum_{i : v_i = w_{i,k}} \left(\sqrt{r} w_{i,k}\right) \left(\sqrt{r} w_{i,k}\right)^* \right\| \leq \left(\frac{1}{\sqrt{r}} + \sqrt{d} \right)^2 .$$

□

This allows us to finally prove KS_2.

Proof of Weaver's Conjecture. Set $r = 2$ and $\delta = \frac{1}{18}$ in the corollary above. We will prove that this implies Weaver's Conjecture for $\eta = 18$ and $\theta = 2$.

Set $u_i = \frac{w_i}{\sqrt{\eta}}$. The condition $\sum_{i=1}^{m} |\langle u, w_i \rangle|^2 = \eta$ for every unit vector $u \in \mathbb{C}^d$, implies that

$$\sum_{i=1} w_i w_i^* = \eta I.$$

Substituting u_i in this expression gives $\sum_i u_i u_i^* = I$.

Moreover, the δ from Corollary 5.3.9 becomes in that case $\delta = 1/\eta$ and hence $\eta = 18$.

If we now multiply the result from the corollary back with η, we obtain the result for Weaver's Conjecture with $\eta - \theta = 16$, which in its turn shows that $\theta = 2$. And this finally concludes the proof. □

And so by combining the idea of iterative proof and barrier functions from chapter 3, the notion of interlacing families from chapter 4 and the mixed characteristic polynomial, we were finally able to prove the Kadison-Singer problem. The fact that this method could solve a famous open problem from analysis once more highlights its use and its value.

5.4. Conclusion on Kadison-Singer

To prove the Kadison-Singer Problem we have shown Weaver's conjecture to be true using the Marcus-Spielman-Srivastava Theorem. The largest part of the proof of is actually rewriting the statement step by step where each steps implies the previous one. Hence if we then prove the last formulation, we can trace our steps to prove the theorem.

In this rewriting quite some new concepts are introduced. First of all, instead of the usual characteristic polynomial, the mixed characteristic polynomial is considered. This because we want to include the expected characteristic polynomial, since we don't know what the actual characteristic polynomial of the random matrix looks like. This in contrast with the expected characteristic polynomial which is the mixed characteristic polynomial over the expected rank-one matrices.

Of course, it has to be investigated if we can just replace the characteristic polynomial by its expectation. One of the reasons why this proof works out so nicely is because this is indeed possible. The fact that the mixed characteristic polynomial is real stable ensures that we can take the expectation without too much trouble.

But by rewriting the univariate polynomial as a multivariate one, it is made impossible to say that its largest root is smaller than a constant. Hence the introduction of being above the roots. This leads to the final form of the theorem that we can finally prove.

In this chapter there are clearly quite a lot concepts that come together. The idea of a proof with iteration and barrier functions from chapter 3 is refined. The fact that the polynomials form an interlacing family, studied in chapter 4, is needed to take the expectation of the characteristic polynomial. All is then combined with the mixed characteristic polynomial and real stability.

This makes it the perfect apotheosis for my study of the interlacing method.

6

Discussion and conclusion

My thesis reviewed the method of interlacing polynomials. But unlike the name suggests, this is not just one method. It is a collective term for all sorts of methods using interlacing polynomials, common interlacing and interlacing families.

Since these three concepts allow for a great range of applications, the method can be used in a large amount of problems. The versatility was shown by considering three papers from three different mathematical fields: spectral graph theory, linear algebra and analysis. Of course, since the results of all papers could be shown using related methods, their cores could be boiled down to pretty similar mathematics.

In all problems we considered the roots of a matrix, constructed by rank-one matrices. The eigenvalues of a rank-one update of matrix are almost interlaced by the eigenvalues of the original matrix. That makes that we have more or less control on how these eigenvalues behave.

This control is used to iteratively construct matrices whose eigenvalues are bounded in a small interval (chapter 3) or bounded above by a small value (chapter 5). In both chapters we keep track on how these eigenvalues move at each step in every direction with barrier functions. Those functions were originally used in physics, but through the physical representation of eigenvalues after a rank-one update they also found their way to linear algebra.

Chapter 4 also considers eigenvalues of a matrix built up with rank-one matrices. However, we don't keep track of the zeros using barrier functions, but by constructing a tree of polynomials. This tree forms an interlacing family. The leaves of the tree are the characteristic polynomials of the possible matrices after having chosen k columns. The levels above become average polynomials over their children. This construction resulted in the root polynomial being a Laguerre polynomial. In each interlacing family there is a relation between the zeros of the root polynomial and those of the leaf polynomials. The beautiful correspondence between the root polynomial and the Laguerre polynomial provided us with a bound on the zeros of the root polynomial. Using the said relationship we are able to pick a certain leaf polynomial with exactly the bound on its zeros that we wanted.

There is a clear evolution through the paper in how the interlacing method is used. In *Twice Ramanujan Sparsifiers* the set of vector from which we construct rank-one updates is deterministic. There is no probability theory involved, we are only constructing an order to add the different updates in such a way that we control the behaviour of the eigenvalues. This absence of probability theory allows us to just keep working with the univariate characteristic polynomial.

In the chapter *Restricted Invertibility* the idea of picking vectors at random is introduced the first time. We prove that if you pick the vectors with replacement, there is a possible choice of vectors with the wanted properties. We use the most basic probabilistic method: we just keep track of all possibilities and then pick the one that we like.

In the last chapter, on the *Kadison-Singer Problem* the probabilistic method is more complicated. To prove there is a possible choice of vectors we use the expected characteristic polynomial. This resulted in the multivariate mixed characteristic polynomial which also complicated the use of similar methods to the those of the chapter on the sparsifiers.

Of course, this should not be the end of the evolution of research on the interlacing method. There are still problems worth investigating. Professor Marcus suggested a follow-up question to the Marcus-Spielman-Srivastava theorem.

Question 6.0.1. Theorem 5.0.3 proves that the largest eigenvalue of the matrix is bounded above by a constant. Can we find a sharp lower bound on the smallest eigenvalue as well?

The reason why this is an interesting question is that it is in the first place very similar to the mentioned theorem. However, it is not that straightforward.

A possible first step would be to note that the smallest root of a polynomial $p(x)$ is (-1) times the largest root of $p(-x)$. So one could try to prove that a certain vector is above the roots of $\mu[A_1, \dots, A_m](-x)$. But then the operator in each iteration becomes $(1 + \partial_i)$. Hence a new investigation is needed on how the zeros behave at each step. This would be very interesting because besides a lower bound on the eigenvalues it would also give insight in the different bounds one can construct using barrier functions for different operators.

A second possible continuation of the research is to dive in the proof of the existence of bipartite Ramanujan graphs for any number of vertices and any degree. As stated in chapter 3, these graphs are notorious for being excellent sparsifiers for the complete graph.

Question 6.0.2. Does there exist a Ramanujan graph for any degree d and any number of vertices n?

This turns out to be the case. Their existence was proven by Marcus, Spielman and Srivastava by again considering the mixed characteristic polynomial. So even though it isn't an open problem, the papers on this topic are a nice extension for people interested in the interlacing method. [13] [11]

A more computational continuation could be to implement the proofs discussed in code. In the papers polynomial run time algorithms are suggested. But of course it is interesting to actually write the code and research whether the run time can be further reduced.

Question 6.0.3. How fast can we make algorithms implementing the problems from the three papers?

Lastly, method of interlacing polynomials is a tool to keep track of the zeros of a polynomial. The most famous problem involving roots of a function is of course the Riemann hypothesis. [4] There are people looking in the similarities between the method and this hypothesis. [8] The probability of actually proving the hypothesis with a related method is extremely low, but beating those odds would without a doubt be a interesting continuation in the evolution of the interlacing method.

Acknowledgement

First and foremost I would like to thank my two supervisors: Anurag Bishnoi and Adam Marcus. It was only because of their time, input and guidance that I was able to get a full insight in the interlacing method. I am extremely grateful that they both accepted my request to do a project on this theme.

I would like to extend my sincere thanks to Bas Janssens. Even though he is not one of the official supervisors he attended the intermediate presentations and his feedback was highly valuable. Moreover, I am really thankful for him being part of my graduation committee.

I'd also like to express my appreciation for all the support received from my mentor Jan van Neerven. He has guided me in all my important decisions regarding my studies so far, of which the exchange to the EPFL was one.

Lastly, I would like to thank Olof Lissmats for proofreading my thesis, his incredible knowledge on Latex and just being an amazing friend.

Lander Verlinde
Delft, June 2022

Bibliography

[1] J.D. Batson, D.A. Spielman, and N. Srivastava. *Twice-Ramanujan Sparsifiers*. 2008. URL: https://doi.org/10.48550/arXiv.0808.0163.

[2] A. Benczur and D. R. Karger. *Randomized Approximation Schemes for Cuts and Flows in Capacitated Graphs*. 2002. URL: https://doi.org/10.48550/arxiv.cs/0207078.

[3] J. Borcea and P. Brändén. *Applications of stable polynomials to mixed determinants: Johnson's conjectures, unimodality, and symmetrized Fischer products*. 2006. URL: https://doi.org/10.48550/arxiv.math/0607755.

[4] B. Conrey. *Riemann's hypothesis*. 2019. URL: https://aimath.org/~kaur/publications/90.pdf.

[5] T. Gale, E. Elsen, and S. Hooker. *The State of Sparsity in Deep Neural Networks*. 2019. URL: https://doi.org/10.48550/arxiv.1902.09574.

[6] G.H. Golub and C.F. Van Loan. "Matrix Analysis". In: *Matrix Computations*. 3rd ed. The Johns Hopkins University Press, 1996, pp. 48–52.

[7] W.H. Haemers. "Interlacing eigenvalues and graphs". In: *Linear Algebra and its Applications* 226–228 (1995), pp. 593–616. URL: https://doi.org/10.1016/0024-3795(95)00199-2.

[8] A. Higashitani, M. Kummer, and M. Michałek. "Interlacing Ehrhart polynomials of reflexive polytopes". In: *Selecta Mathematica* 23.4 (2017), pp. 2977–2998. URL: https://doi.org/10.1007/s00029-017-0350-6.

[9] R.V. Kadison and I.M. Singer. "Extensions of Pure States". In: *American Journal of Mathematics* 81.2 (1959), pp. 383–400. URL: http://www.jstor.org/stable/2372748.

[10] I. Krasikov. *On extreme zeros of classical orthogonal polynomials*. 2003. URL: https://doi.org/10.48550/arxiv.math/0306286.

[11] A. W. Marcus, D. A. Spielman, and N. Srivastava. *Interlacing Families IV: Bipartite Ramanujan Graphs of All Sizes*. 2015. URL: https://doi.org/10.48550/arxiv.1505.08010.

[12] A. W. Marcus, D. A. Spielman, and N. Srivastava. *Interlacing Families III: Sharper Restricted Invertibility Estimates*. 2017. URL: https://doi.org/10.48550/arxiv.1712.07766.

[13] A.W Marcus, D. A. Spielman, and N. Srivastava. *Interlacing Families I: Bipartite Ramanujan Graphs of All Degrees*. 2013. URL: https://doi.org/10.48550/arxiv.1304.4132.

[14] A.W. Marcus, D. A. Spielman, and N. Srivastava. *Interlacing Families II: Mixed Characteristic Polynomials and the Kadison-Singer Problem*. 2013. URL: https://doi.org/10.48550/arxiv.1306.3969.

[15] M. Marden. "Combinations of a polynomial and its derivatives". In: *Geometry of polynomials*. American Mathematical Society, 1949, pp. 81–89. URL: https://doi.org/http://dx.doi.org/10.1090/surv/003.

[16] N. Srivastava. "Graph Sparsification II: Rank one updates, Interlacing, and Barriers". [Slides]. Aug. 26, 2014. URL: https://simons.berkeley.edu/sites/default/files/docs/1772/slidessrivastava2.pdf.

[17] N. Srivastava. *Restricted Invertiblity by Interlacing Polynomials*. Apr. 15, 2014. URL: https://windowsontheory.org/2014/04/15/restricted-invertiblity-by-interlacing-polynomials/.

[18] G. Szegö. "Laguerre and hermite polynomials". In: *Orthogonal polynomials*. Vol. 23. American Mathematical Society, 1939, p. 101. URL: https://doi.org/http://dx.doi.org/10.1090/coll/023.

[19] T. Tao. *Real stable polynomials and the Kadison-Singer problem*. Nov. 4, 2013. URL: https://terrytao.wordpress.com/2013/11/04/real-stable-polynomials-and-the-kadison-singer-problem/.

[20] T. Tao. “Jean Bourgain, problem solver”. In: *Proceedings of the National Academy of Sciences* 116.28 (2019), pp. 13717–13718. URL: https://doi.org/10.1073/pnas.190196511.

[21] L. Tzafriri and J. Bourgain. “On a problem of Kadison and Singer.” In: *Journal für die angewandte Mathematik* 420 (1991), pp. 1–44. URL: https://doi.org/10.1515/crll.1991.420.1.

[22] V.S. Vassiliadis and C.A. Floudas. “The modified barrier function approach for large-scale optimization”. In: *Computers & Chemical Engineering* 28.8 (1997), pp. 855–874. URL: https://doi.org/10.1016/S0098-1354(96)00313-4.

[23] D. G. Wagner. *Multivariate stable polynomials: theory and applications*. 2009. URL: https://doi.org/10.48550/arxiv.0911.3569.

[24] G.W. Watson et al. “Lattice Energy and Free Energy Minimization Techniques”. In: *Computer Modeling in Inorganic Crystallography*. Academic Press, 1997, pp. 55–81. URL: https://doi.org/10.1016/B978-012164135-1/50004-6.

[25] N. Weaver. “The Kadison–Singer problem in discrepancy theory”. In: *Discrete Mathematics* 278.1 (2004), pp. 227–239. URL: https://doi.org/10.1016/S0012-365X(03)00253-X.