Chapter 17

More Interlacing Families

We see some further developments in the method of interlacing family of polynomials in this chapter.

17.1 Interlacing Family for Strongly Rayleigh Measure

Motivated by the thin tree problem, Anari and Oveis Gharan [AO14] developed an interlacing family for strongly Rayleigh distributions and applied it to the asymmetric traveling salesman problem. We will first study the interlacing family and then discuss the application.

Recall that the interlacing family and the probabilistic formulation by Marcus, Spielman, and Srivastava in Theorem 13.28 and Theorem 15.2 crucially depend on the random variables being independent. Anari and Oveis Gharan proved a beautiful generalization of the probabilistic formulation to strongly Rayleigh measures. In the following, a strongly Rayleigh measure μ is homogeneous if every non-zero monomial in the generating polynomial g_μ is of the same degree.

Theorem 17.1 (Probabilistic Method for Strongly Rayleigh Measure)**.** *Let $\mu : \{0,1\}^m \to \mathbb{R}$ be a homogeneous strongly Rayleigh measure such that the marginal probability $\Pr_{S\sim\mu}[i \in S]$ of each element $1 \leq i \leq m$ is at most ϵ_1. Let $v_1, \ldots, v_m \in \mathbb{R}^n$ be vectors in isotropic condition $\sum_{i=1}^m v_i v_i^T = I_n$ and $\|v_i\|_2^2 \leq \epsilon_2$ for $1 \leq i \leq m$. Then*

$$\Pr_{S\sim\mu}\left[\left\|\sum_{i\in S} v_i v_i^T\right\| \leq 4(\epsilon_1+\epsilon_2) + 2(\epsilon_1+\epsilon_2)^2\right] > 0.$$

Product distributions are strongly Rayleigh distributions, so Theorem 17.1 should be more general than Theorem 15.2, but the leading constant is just slightly larger that it cannot be used to prove Weaver's Conjecture 15.1. It can still be used to prove a multi-partitioning version of Weaver's conjecture for any $r \geq 5$, using a similar reduction as in Section 15.2.

Problem 17.2 (Multi-Partitioning Weaver's Problem)**.** *Let $v_1, \ldots, v_m \in \mathbb{R}^n$ be vectors in isotropic condition $\sum_{i=1}^m v_i v_i^T = I_n$ and $\|v_i\|_2^2 \leq \epsilon$ for $1 \leq i \leq m$. Then, for any r, there is an r partitioning of $[m]$ into $S_1, \ldots, S_r$ such that for any $1 \leq j \leq r$,*

$$\left\|\sum_{i\in S} v_i v_i^T\right\| \leq 4\Big(\frac{1}{r}+\epsilon\Big) + 2\Big(\frac{1}{r}+\epsilon\Big)^2.$$

The proof of Theorem 17.1 is based on the same two key steps as in Chapter 14 and Chapter 15:

1. Prove that the family of polynomials $\left\{\det\left(xI - \sum_{i\in S} v_i v_i^T\right)\right\}_{S\in\text{supp}(\mu)}$ forms an interlacing family, and apply the probabilistic method in Theorem 12.12 to show that there exists $S \in \text{supp}(\mu)$ with
$$\lambda_{\max}\left(\det\left(xI - \sum_{i\in S} v_i v_i^T\right)\right) \le \lambda_{\max}\left(\mathbb{E}_{S\sim\mu}\left[\det\left(xI - \sum_{i\in S} v_i v_i^T\right)\right]\right).$$

2. Bound the maximum root of $\mathbb{E}_{S\sim\mu}\left[\det\left(xI - \sum_{i\in S} v_i v_i^T\right)\right] = \sum_S \mu(S)\cdot\det\left(xI - \sum_{i\in S} v_i v_i^T\right)$.

Expected Characteristic Polynomial

Recall that in the solution to the Weaver's conjecture in Chapter 15, both steps depend crucially on the multilinear formula in Theorem 13.20. Anari and Oveis Gharan proved a generalization incorporating the probability measure μ.

Theorem 17.3 (Expected Characteristic Polynomial of Strongly Rayleigh Distribution)**.** *Let g_μ be the homogeneous generating polynomial of a measure $\mu : \{0,1\}^m \to \mathbb{R}$ with degree d. Let $v_1, v_2, \ldots, v_m \in \mathbb{R}^n$. For any $c \in \mathbb{R}$,*
$$c^{d-n}\sum_{R\subseteq[m]} \mu(R)\det\left(c\lambda I - \sum_{i\in R} 2v_i v_i^T\right) = \prod_{i=1}^m (1-\partial_{x_i}^2) g_\mu(c\vec{1}+x)\cdot\det\left(\lambda I + \sum_{i=1}^m x_i v_i v_i^T\right)\Big|_{\vec{x}=0}.$$

Proof. Start with the LHS. Let $A_i = v_i v_i^T$, which is rank one so that $\det(\lambda I + \sum_i x_i A_i)$ is multilinear in x_i. Then
$$\begin{aligned}\sum_R \mu(R)\det\left(\lambda I + \sum_{i\in R} x_i A_i\right) &= \sum_R \mu(R)\sum_{S\subseteq R} x^S\left(\prod_{i\in S}\partial_{x_i}\det\left(\lambda I + \sum_{j\in R} x_j A_j\right)\big|_{\vec{x}=0}\right) \qquad (17.1)\\ &= \sum_R \mu(R)\sum_{S\subseteq R} x^S\left(\prod_{i\in S}\partial_{x_i}\det\left(\lambda I + \sum_{j=1}^m x_j A_j\right)\big|_{\vec{x}=0}\right)\\ &= \sum_S x^S\Big(\sum_{R:R\supseteq S}\mu(R)\Big)\left(\prod_{i\in S}\partial_{x_i}\det\left(\lambda I + \sum_{j=1}^m x_j A_j\right)\big|_{\vec{x}=0}\right)\end{aligned}$$
where the first equality is by the multilinear expression of $\det(\lambda I + \sum_{i\in R} x_i A_i)$ as described in Subsection 13.4.

The idea is to come up with one polynomial g with coefficient $\sum_{R:R\supseteq S}\mu(R)$ on the monomial x^S, and another polynomial f with coefficient $\prod_{i\in S}\partial_{x_i}\det(\lambda I + \sum_{j=1}^m x_j A_j)|_{\vec{x}=0}$ on the monomial x^S. Clearly $f(x) := \det(\lambda I + \sum_{i=1}^m x_i A_i)$.

Consider $g_\mu(c\vec{1}+x) = \sum_R \mu(R)\prod_{i\in R}(c+x_i)$. Each R with $R\supseteq S$ contributes $\mu(R)\cdot c^{|R|-|S|}$ to x^S. Therefore, since μ is homogeneous, the coefficient of x^S in g_μ is $\sum_{R:R\supseteq S}\mu(R)\cdot c^{|R|-|S|} = c^{d-|S|}\sum_{R:R\supseteq S}\mu(R)$. Let $g(x) := g_\mu(c\vec{1}+x)$.

Since both f and g are multilinear in x_i, the coefficient of $\prod_{i\in S} x_i^2$ in $f\cdot g$ is the product of the coefficients of x^S in f and in g. We can read the coefficient of $\prod_{i\in S} x_i^2$ in $f\cdot g$ using the formula $2^{-|S|}\prod_{i\in S}\partial_{x_i}^2 f\cdot g|_{\vec{x}=0}$. So,
$$2^{-|S|}\prod_{i\in S}\partial_{x_i}^2\, f\cdot g\big|_{\vec{x}=0} = c^{d-|S|}\sum_{R:R\supseteq S}\mu(R)\prod_{i\in S}\partial_{x_i}\det\left(\lambda I + \sum_{j=1}^m x_j A_j\right)\big|_{\vec{x}=0}.$$

Therefore,

$$\prod_{i=1}^{m}(1-\partial_{x_i}^2)f\cdot g\Big|_{\vec{x}=0} = \sum_S(-1)^{|S|}\prod_{i\in S}\partial_{x_i}^2 f\cdot g\Big|_{\vec{x}=0}$$
$$= c^d\sum_S(-1)^{|S|}\cdot\Big(\frac{2}{c}\Big)^{|S|}\Big(\sum_{R:R\supseteq S}\mu(R)\Big)\Big(\prod_{i\in S}\partial_{x_i}\det\Big(\lambda I+\sum_{j=1}^{m}x_jA_j\Big)\big|_{\vec{x}=0}\Big).$$

Note that the LHS of this equality is equal to the RHS of the statement, and the RHS of this equality is equal to the LHS of the statement by plugging in $x_i = {-2}/{c}$ in Equation 17.1 for $1 \le i \le m$. □

Using Theorem 17.3 and Corollary 16.8, we have the following formula for the expected characteristic polynomial for the volume measure.

Problem 17.4. *Let $v_1, v_2, \ldots, v_m \in \mathbb{R}^n$ with $\sum_{i=1}^m v_iv_i^T = I_n$ and let μ be its volume measure as in Definition 16.5. Then, for any $c \in \mathbb{R}$,*

$$c^{d-n}\sum_{S\subseteq[m]}\mu(S)\det\Big(c\lambda I-\sum_{i\in S}2v_iv_i^T\Big)=\prod_{i=1}^{m}\Big(1-\partial_{x_i}^2\Big)\det\Big(cI+\sum_{i=1}^{m}x_iv_iv_i^T\Big)\cdot\det\Big(\lambda I+\sum_{i=1}^{m}x_iv_iv_i^T\Big)\Big|_{\vec{x}=0}.$$

Interlacing Family

Once the formula in Theorem 17.3 is established, the proof that the family of polynomials $\big\{\det\big(xI-\sum_{i\in S}v_iv_i^T\big)\big\}_{S\in\mathrm{supp}(\mu)}$ forms an interlacing family is similar to that of Theorem 13.28. First, using that g_μ is real stable as μ is strongly Rayleigh, we can prove that the expected characteristic polynomial is real-rooted.

Exercise 17.5 (Expected Characteristic Polynomial is Real-Rooted)**.** *The expected characteristic polynomial $\mathbb{E}_{S\sim\mu}\big[\det\big(xI-\sum_{i\in S}v_iv_i^T\big)\big]$ is real-rooted for any strongly Rayleigh distribution μ.*

Then, using a tree with depth m where each internal node has at most 2 children, and associating each non-leaf node with the conditional expected polynomial, we can use a similar argument as in Theorem 13.28 to establish an interlacing family for strongly Rayleigh measure.

Problem 17.6 (Interlacing Family of Strongly Rayleigh Measure)**.** *Let $\mu : \{0,1\}^m \to \mathbb{R}$ be a strongly Rayleigh measure with homogeneous generating polynomial g_μ. The set of all polynomials in $\big\{\det\big(xI-\sum_{i\in S}v_iv_i^T\big)\big\}_{S\in\mathrm{supp}(\mu)}$ form an interlacing family, where the root polynomial can be chosen to be $\mathbb{E}_{S\sim\mu}\big[\det\big(xI-\sum_{i\in S}v_iv_i^T\big)\big]$.*

Multivariate Barrier Argument

The second step is to upper bound the maximum root of the expected characteristic polynomial. The proof structure is similar to that in the induction hypothesis in Definition 15.6. The same multivariate barrier functions $\Phi_p^i(y) = \partial_{x_i}p(y)/p(y)$ as in Definition 15.5 are used. Starting with a point $(t,\ldots,t)$ which is above the roots of the multivariate polynomial $g_\mu(\lambda\vec{1}+x)\cdot\det\big(\lambda I+\sum_{i=1}^m x_iv_iv_i^T\big)$, Anari and Oveis Gharan proved that $(t+\delta,\ldots,t+\delta,t,\ldots,t)$ with the first k coordinates being $t+\delta$ for a small δ is still above the roots after applying the differential operator $(1-\partial_{x_i}^2)$ for $1 \le i \le k$.

Also, the monotoncity and the convexity of the barrier functions in Proposition 15.7 are important in the analysis.

The main difference is that the differential operator $(1 - \partial_{x_i}^2)$ is different. So, not only they need to keep track of $\Phi_p^i(y) = \partial_{x_i} p(y)/p(y)$, but also the second derivative $\Psi_p^i(z) = \partial_{x_i}^2 p(y)/p(y)$ as well. They prove a new lemma that $\partial_{x_i} \Psi_p^j(y)/\partial_{x_i} \Phi_p^j(y) \leq 2\Phi_p^j(y)$, also using the result that a bivariate real-stable polynomial can be written as $\det(x_1 A + x_2 B + C)$ for $A, B \succcurlyeq 0$ and C Hermitian. As in Chapter 15, the assumptions $\Pr(i \in S) \leq \epsilon_1$ and $\|v_j\|_2^2 \leq \epsilon_2$ are (only) used in the computation of the initial value of the barrier functions. Because of the differential operator $(1-\partial_{x_i}^2) = (1+\partial_{x_i})(1-\partial_{x_i})$, they could prove that the shift δ is much smaller as $(1-\partial_{x_i})$ shifts the root up while $(1+\partial_{x_i})$ shifts the roots down, hence getting a final bound that is much smaller than that in Theorem 15.2.

We refer the reader to [AO14] for details. It would be very nice if one could strengthen their result to prove Weaver's Conjecture 15.1.

Thin Tree and Asymmetric Traveling Salesman Problem

The main motivation of their work is the thin tree problem and its application to the asymmetric traveling salesman problem (ATSP).

Definition 17.7 (Thin Tree). *Given an undirected graph $G = (V, E)$ and $0 < \alpha < 1$, we say a spanning tree T is α-thin if $|\delta_T(S)| \leq \alpha \cdot |\delta_G(S)|$ for all $S \subseteq V$. In words, a spanning tree is α-thin if it uses at most α-fraction of edges in every cut.*

There is a strong conjecture about the existence of a thin tree.

Conjecture 17.8 (Goddyn's Conjecture). *Every k-edge-connected graph has a $O(\frac{1}{k})$-thin tree.*

If the conjecture is true and a $O(\frac{1}{k})$-thin tree can be found in polynomial time, then it would imply a constant factor approximation algorithm for ATSP [AGM+17].

It can be proved that a random spanning tree is a $O\big(\frac{\log n}{\log\log n} \cdot \frac{1}{k}\big)$-thin tree [AGM+17]. The argument is similar to that in cut sparsification, using Chernoff bound and careful union bound. The reason that we can apply Chernoff bound is that the edges in a random spanning tree are negatively associated as shown in Chapter 16.

As in graph sparsification, one can define a stronger spectral notion of a thin tree.

Definition 17.9 (Spectrally Thin Tree). *Given an undirected graph $G = (V, E)$ and $0 < \alpha < 1$, we say a spanning tree T is α-spectrally-thin if $L(T) \preccurlyeq \alpha \cdot L(G)$, where $L(T)$ and $L(G)$ are the Laplacian matrices of T and G respectively.*

Exercise 17.10 (Spectral Thin Tree is Combinatorially Thin Tree). *Prove that an α-spectrally thin tree is also an α-thin tree.*

One advantage of this stronger notion is that it is easier to work with. For example, given a tree, it is easy to check whether it is α-spectrally thin, while it is not known how to check whether it is (combinatorially) α-thin. Moreover, the solution to the Weaver's conjecture in Corollary 15.3 provides a non-trivial sufficient condition for the existence of a spectrally thin tree.

Proposition 17.11 (Sufficient Condition for Spectrally Thin Tree [HO14]). *If the maximum effective resistance of an edge in a graph G is α, then G has a $O(\alpha)$-spectrally thin tree.*

The solution to Weaver's conjecture in Corollary 15.3 implies that if the maximum effective resistance of an edge in G is α, then the edge set of G can be partitioned into two subgraphs H_1 and H_2 such that for $i \in \{1, 2\}$,

$$\frac{1}{2}\big(1-\sqrt{2\alpha}\big)L_G \preccurlyeq L_{H_i} \preccurlyeq \frac{1}{2}\big(1+\sqrt{2\alpha}\big)L_G.$$

The idea in Proposition 17.11 is to recursively apply this partitioning in each subgraph (with slightly weaker bound on the maximum effective resistance of an edge) until we cannot apply again, by that time there will be $O(\frac{1}{\alpha})$ edge-disjoint subgraphs of G, each is connected and $O(\alpha)$-spectrally thin.

This gives hope that the techniques developed in the method of interlacing family of polynomials can be used to prove Goddyn's conjecture. Proposition 17.11 gives us a spectrally thin tree, which is combinatorially thin, but it requires the assumption that hte maximum effective resistance of an edge is small, which is not necessarily satisfied in a k-edge-connected graph. The breakthrough by Anari and Oveis Gharan [AO15], in a high level, is to reduce the combinatorial problem to the spectral problem, and use Theorem 17.1 to prove the following result.

Theorem 17.12 (Anari, Oveis Gharan [AO15])**.** *Every k-edge-connected graph has a $O\big(\log\log n \cdot \frac{1}{k}\big)$-thin tree.*

The reduction, however, is very complicated and technically challenging. Also, there is now a constant factor approximation algorithm for ATSP. So we just highlight some underlying mathematics of the thin tree result. First, check that the probabilistic formulation for strongly Rayleigh measure can be used to prove Proposition 17.11 without using recursion.

Exercise 17.13 (Direct Proof of Sufficient Condition)**.** *Show that Theorem 17.1 can be used to prove Proposition 17.11. You may use the fact that the probability that an edge is in a uniform spanning tree is equal to the effective resistance of an edge, i.e.* $\Pr_T[e \in T] = \mathrm{Reff}(e)$*.*

The main advantage of Theorem 17.1 is that the output is guaranteed to be a spanning tree, so that we get connectivity for free (without worrying about the minimum eigenvalue), which is important for the thin tree problem. The following is the fundamental building block of their approach.

Theorem 17.14 (Spectral Thin Tree from Subgraph)**.** *Given a graph $G = (V, E)$ and a subset of edges $F \subseteq E$ such that (V, F) is k-edge-connected, if $\mathrm{Reff}(e) \leq \epsilon$ for all $e \in F$, then G has a $O\big(\frac{1}{k} + \epsilon\big)$-spectrally thin tree in F.*

Proof Idea: Since F is k-edge-connected, there are at least $k/2$ edge-disjoint spanning trees by Tutte or Nash-Williams' theorem. This implies that there is a point in the spanning tree polytope with maximum edge value $O(1/k)$. By expressing this point using "maximum entropy distribution", it can be proved that there is a weighting of the edges, so that the weighted random spanning tree distribution μ (which is still homogeneous strongly Rayleigh by Exercise 16.11) has maximum marginal probability of an edge $O(1/k)$, i.e. $\epsilon_1 = O(1/k)$ in Theorem 17.1. The assumption about effective resistance implies that $\epsilon_2 \leq \epsilon$, and so Theorem 17.1 can be applied. □

With Theorem 17.14, their strategy is to add "short-cut" edges in the graph, so that they don't change the cut structures much, while creating many edges with small effective resistance. They do it in $O(\log\log n)$ iterations so that the edges with small effective resistance form a k-edge-connected subgraph. The most difficult step is to prove the existence of good short-cut edges, which they proved by using an involved analysis of a semidefinite program. They managed to prove Theorem 17.12 after 80 pages of work after Theorem 17.14.

17.2 Interlacing Family for Matrix Discrepancy

The result by Marcus, Spielman, and Srivastava on Weaver's conjecture can be interpreted as an improved discrepancy bound over the matrix Chernoff bound in Theorem 9.13. Taking this perspective, Kyng, Luh, and Song [KLS20] considered the following more refined matrix concentration result.

Theorem 17.15 (Matrix Concentration with Variance [Tro12]). *Let $\xi_i \in \{\pm 1\}$ be independent random signs, and let $A_1, \ldots, A_m \in \mathbb{R}^{n\times n}$ be symmetric matrices. Let $\sigma^2 = \|\sum_{i=1}^m \mathrm{var}[\xi_i] A_i^2\|$. Then,*

$$\Pr\left(\left\|\sum_{i=1}^m \mathbb{E}\left[\xi_i\right] A_i - \sum_{i=1}^m \xi_i A_i\right\| \geq t\cdot\sigma\right) \leq 2ne^{-\frac{t^2}{2}}.$$

This theorem implies that with high probability the discrepancy is at most $O(\sqrt{\log n})\cdot\sigma$. The main result of Kyng, Luh, and Song is to prove that there exists a signing with a stronger discrepancy bound.

Theorem 17.16 (Matrix Discrepancy of Rank One Matrices [KLS20]). *Consider any independent scalar random variables $\xi_1, \ldots, \xi_m$ with finite support. Let $u_1, \ldots, u_m \in \mathbb{R}^n$ and*

$$\sigma^2 = \left\|\sum_{i=1}^m \mathrm{var}[\xi_i](u_i u_i^T)^2\right\|.$$

Then there exists a choice of outcomes $\epsilon_1, \ldots, \epsilon_m$ in the support of $\xi_1, \ldots, \xi_m$ such that

$$\left\|\sum_{i=1}^m \mathbb{E}\left[\xi_i\right] u_i u_i^T - \sum_{i=1}^m \epsilon_i u_i u_i^T\right\| \leq 4\sigma.$$

Note that if $\|u_i\|_2^2 \leq \epsilon$ and $\sum_{i=1}^m u_i u_i^T = I_n$, then $\sigma^2 \leq \epsilon$, and the conclusion is that there is a signing $\epsilon_1, \ldots, \epsilon_m \in \{\pm 1\}$ with $\left\|\sum_{i=1}^m \epsilon_i u_i u_i^T\right\| \leq O(\sqrt{\epsilon})$. Check that this matches the result of Marcus, Spielman and Srivastava in Corollary 15.3 applied to the same setting (i.e. with $\{\pm 1\}$ instead of $\{0,1\}$). Theorem 17.16 is more flexible that allow arbitrary biased ± 1 random variables, instead of only zero mean random variables. Also, Theorem 17.16 is more refined in that it proves stronger bounds when there are only a few vectors with $\|u_i\|_2^2 = \epsilon$ while other vectors are much shorter.

Two-Sided Spectral Rounding

One interesting application of Theorem 17.16 is the two-sided spectral rounding problem from Chapter 11.

Problem 17.17 (Two-Sided Spectral Rounding). *Let $v_1, ..., v_m \in \mathbb{R}^n$ and $x \in [0,1]^m$. Suppose $\sum_{i=1}^m x_i v_i v_i^T = I_n$ and $\|v_i\|_2^2 \leq \epsilon$ for all $i \in [m]$. Prove that there exists a subset $S \subseteq [m]$ satisfying*

$$\big(1 - O(\sqrt{\epsilon})\big)\cdot I_n \preccurlyeq \sum_{i\in S} v_i v_i^T \preccurlyeq \big(1 + O(\sqrt{\epsilon})\big)\cdot I_n.$$

This result can be slightly extended to incorporate one non-negative linear constraint, which has some applications in network design.

Theorem 17.18 (Two-Sided Spectral Rounding with Costs [LZ20]). *Let $v_1, ..., v_m \in \mathbb{R}^n$, $x \in [0,1]^m$ and $c \in \mathbb{R}^m_{\geq 0}$. Suppose $\sum_{i=1}^m x_i v_i v_i^T = I_n$, $\|v_i\|_2^2 \leq \epsilon$ for all $i \in [m]$ and $c_\infty \leq \epsilon^2 \langle c, x\rangle$. Then there exists $z \in \{0,1\}^m$ such that*

$$\big(1-O(\sqrt{\epsilon})\big)\cdot I_n \preccurlyeq \sum_{i=1}^m z_i v_i v_i^T \preccurlyeq \big(1+O(\sqrt{\epsilon})\big)\cdot I_n \quad \text{and} \quad \big(1-O(\sqrt{\epsilon})\big)\cdot\langle c, x\rangle \leq \langle c, z\rangle \leq \big(1+O(\sqrt{\epsilon})\big)\cdot\langle c, x\rangle.$$

Proof Ideas

Theorem 17.16 needs to bound the maximum eigenvalue and the minimum eigenvalue of the difference. Their main idea is to consider the polynomial

$$\det\Big(x^2 I - \Big(\sum_{i=1}^m \xi_i u_i u_i^T\Big)^2\Big) = \det\Big(xI - \sum_{i=1}^m \xi_i u_i u_i^T\Big)\cdot \det\Big(xI + \sum_{i=1}^m \xi_i u_i u_i^T\Big)$$

Note that the largest root of this polynomial is

$$\lambda_{\max}\Bigg(\det\Big(x^2 I - \Big(\sum_{i=1}^m \xi_i u_i u_i^T\Big)^2\Big)\Bigg) = \Big\|\sum_{i=1}^m \xi_i u_i u_i^T\Big\|.$$

They found a nice formula for the expected characteristic polynomial.

Proposition 17.19 (Expected Characteristic Polynomial for Matrix Discrepancy). *Let $u_1, \ldots, u_m \in \mathbb{R}^n$. Consider independent random variables ξ_i with means μ_i and variances γ_i^2. Let $Q \in \mathbb{R}^{m\times m}$ be a symmetric matrix. Then*

$$\begin{aligned}&\mathbb{E}_\xi\Bigg[\det\Big(x^2 I - \Big(Q + \sum_{i=1}^m (\xi_i - \mu_i) u_i u_i^T\Big)^2\Big)\Bigg] \\ =\ &\prod_{i=1}^m \Big(1 - \frac{1}{2}\partial_{z_i}^2\Big) \det\Big(xI - Q + \sum_{i=1}^m z_i\gamma_i u_i u_i^T\Big)\cdot\det\Big(xI + Q + \sum_{i=1}^m z_i \gamma_i u_i u_i^T\Big)\Big|_{z_1=\ldots=z_m=0}.\end{aligned}$$

This formula is obtained inductively by the following lemma, as in the inductive proof of the multilinear formula in Subsection 13.4.

Problem 17.20 (Expected Characteristic Polynomial after One Step). *For positive semidefinite matrices $M, N \in \mathbb{R}^{m\times m}$, $v \in \mathbb{R}^m$ and ξ a random variable with zero mean and variance γ^2,*

$$\mathbb{E}_\xi\Big[\det\big(M - \xi vv^T\big)\cdot\det\big(N + \xi vv^T\big)\Big] = \Big(1 - \frac{1}{2}\frac{d^2}{dt^2}\Big)\det\big(M + t\gamma vv^T\big)\det\big(N + t\gamma vv^T\big)\Big|_{t=0}.$$

Proposition 17.19 implies that the expected characteristic polynomial is real-rooted. Then, using a similar argument as in Theorem 13.28 and Problem 17.6, one can prove that the set of all possible characteristic polynomials form an interlacing family. Therefore, by the probabilistic method in Theorem 12.12, there exists a choice of outcomes $\epsilon_1, \ldots, \epsilon_m$ in the finite support of $\xi_1, \ldots, \xi_m$ such that

$$\Big\|\sum_{i=1}^m \epsilon_i u_i u_i^T - \sum_{i=1}^m \mathbb{E}[\xi_i]\, u_i u_i^T\Big\| \leq \lambda_{\max}\Bigg(\mathbb{E}_{\xi_1,\ldots,\xi_m}\Bigg[\det\Big(x^2 I - \Big(\sum_{i=1}^m (\xi_i - \mathbb{E}[\xi_i]) u_i u_i^T\Big)^2\Big)\Bigg]\Bigg).$$

Then the second step is to bound the maximum root of the expected polynomial. The differential operator $1 - \partial_{z_i}^2$ in Proposition 17.19 is the same as in the formula for strongly Rayleigh measure in Theorem 17.3. It turned out that the same setup (including the induction hypothesis and the multivariate barrier functions) and many calculations in [AO14] can be reused. The base case is different, and once again the assumptions are (only) used in computing the initial values of the barrier functions.

17.3 Interlacing Family for Higher Rank Matrices

The interlacing families that we have seen so far all involve sum of rank one matrices. It was remarked that the same approach would fail spectacularly for sum of higher rank matrices, e.g. the multilinear formula in Theorem 13.20 does not hold and the expected characterisitic polynomials may not even be real-rooted anymore. Cohen [Coh16] found a very clever solution to bypass expected characteristic polynomials and proved the following generalization of Theorem 15.2.

Theorem 17.21 (Cohen [Coh16]). *Let $A_1, \ldots, A_m \in \mathbb{R}^{n \times n}$ be independent random positive semidefinite matrices with finite support such that*

$$\mathbb{E}\Big[\sum_{i=1}^m A_i\Big] = I_n \quad \text{and} \quad \mathbb{E}\big[\operatorname{Tr}(A_i)\big] \le \epsilon \text{ for } 1 \le i \le m.$$

Then

$$\Pr\left[\lambda_{\max}\Big(\sum_{i=1}^m A_i\Big) \le \big(1+\sqrt{\epsilon}\big)^2\right] > 0.$$

The insight of Cohen is to focus on the RHS of the multilinear formula, the mixed characteristic polynomial in Definition 13.21. The following is a generalization with a "non-mixed" matrix M, where the mixed characteristic polynomial in Definition 13.21 is a special case with $M = 0$.

Definition 17.22 (Generalized Mixed Characteristic Polynomial). *The generalized mixed characteristic polynomial of $n \times n$ matrices $M, B_1, \ldots, B_m$ (not necessarily rank-one) is defined as*

$$\mu[M; B_1, \ldots, B_m](\lambda) = \prod_{i=1}^m (1 - \partial_{x_i}) \det\Big(\lambda I - M + \sum_{i=1}^m x_i \cdot B_i\Big)\Big|_{x_1 = x_2 = \cdots = x_m = 0}$$

The multivariate barrier argument in Chapter 15 proved that

$$\lambda_{\max}\Big(\mu\big[\mathbb{E}[A_1], \ldots, \mathbb{E}[A_m]\big](\lambda)\Big) = \lambda_{\max}\bigg(\prod_{i=1}^m (1 - \partial_{x_i}) \det\Big(\lambda I + \sum_{i=1}^m x_i \cdot \mathbb{E}[A_i]\Big)\Big|_{\vec{x}=0}\bigg) \le (1+\sqrt{\epsilon})^2.$$

Cohen's proof has two steps. The first step is to prove that the set of all possible mixed characteristic polynomials form an interlacing family and so the probabilistic method in Theorem 12.12 applies.

Problem 17.23 (Interlacing Family for Mixed Characteristic Polynomials). *Let $A_1, \ldots, A_m \in \mathbb{R}^{n \times n}$ be independent random positive semidefinite matrices, where each A_j has k possibilities $M_{j,1}, \ldots, M_{j,k}$. Prove that the set of all k^m possible mixed characteristic polynomials $\mu[M_{1,i_1}, \ldots, M_{m,i_m}](\lambda)$ where each $1 \le i_j \le k$ for $1 \le j \le m$ form an interlacing family, and the root polynomial can be chosen to be $\mu\big[\mathbb{E}[A_1], \ldots, \mathbb{E}[A_m]\big]$. Conclude that there exists a choice $M_j \in \operatorname{supp}(A_j)$ for $1 \le j \le m$ such that*

$$\lambda_{\max}\Big(\mu\big[M_1, \ldots, M_m\big](\lambda)\Big) \le \lambda_{\max}\Big(\mu\big[\mathbb{E}[A_1], \ldots, \mathbb{E}[A_m]\big](\lambda)\Big).$$

The second step is to prove that the maximum root of the characteristic polynomial can only be smaller than the maximum root of the mixed characteristic polynomial.

Proposition 17.24 (Maximum Root of Mixed Characteristic Polynomials)**.** *For any positive semidefinite matrices $M_1, \ldots, M_m$,*

$$\lambda_{\max}\Big(\det\Big(\lambda I - \sum_{i=1}^{m} M_i\Big)\Big) \leq \lambda_{\max}\Big(\mu\big[M_1, \ldots, M_m\big](\lambda)\Big)$$

The proof of Proposition 17.24 is by applying the following lemma repeatedly, where we move the mixed part to the non-mixed part one at a time. The proof of the following lemma nicely uses the convexity of real-stable polynomials for points above the roots.

Lemma 17.25 (Inductive Step for Proposition 17.24)**.** *Let $M, M_1, \ldots, M_m \in \mathbb{R}^{n\times n}$ be positive semidefinite matrices. Then*

$$\lambda_{\max}\Big(\mu\big[M + M_m; M_1, \ldots, M_{m-1}\big](\lambda)\Big) \leq \lambda_{\max}\Big(\mu\big[M; M_1, \ldots, M_m\big](\lambda)\Big)$$

Proof Sketch: Consider the bivariate polynomial

$$p(\lambda, x) := \prod_{i=1}^{m-1}(1 - \partial_{x_i}) \det\Big(\lambda I - M + xM_m + \sum_{j=1}^{m-1} x_i M_i\Big)\Big|_{x_1=\ldots=x_{m-1}=0}.$$

Note that

$$p(\lambda, -1) = \mu\big[M + M_m; M_1, \ldots, M_{m-1}\big](\lambda) \quad \text{and} \quad (1 - \partial_x)p(\lambda, x)|_{x=0} = \mu\big[M; M_1, \ldots, M_m\big](\lambda).$$

Let λ^* be the maximum root of $p(\lambda, -1)$. Note that both $p(\lambda, -1)$ and $(1 - \partial_x)p(\lambda, x)|_{x=0}$ are real-rooted with positive leading coefficients. So, to prove the statement of the lemma, it suffices to prove that $(1 - \partial_x)p(\lambda^*, x)|_{x=0} \leq 0$.

Using the result that any bivariate real-stable polynomial $p(x_1, x_2)$ can be written as $\pm\det(x_1 A + x_2 B + C)$ for some $A, B \succcurlyeq 0$ and some symmetric C (or the complex analysis argument in Tao's blogpost), it can be shown that the roots of the univariate polynomial $p_x(\lambda)$ can only decrease when we increase x. This implies that the point $(\lambda^*, -1)$ is above the roots of $p(\lambda, x)$. Since $(\lambda^*, -1)$ is above the roots of $p(\lambda, x)$, the point $(\lambda^*, 0)$ is also above the roots of $p(\lambda, x)$. Again, by the result of bivariate real-stable polynomial, it follows that $p(\lambda^*, x)$ is convex along the interval $x \in [-1, 0]$, which implies that

$$p(\lambda^*, 0) - p(\lambda^*, -1) \leq \partial_x p(\lambda^*, 0) \quad \Longrightarrow \quad (1 - \partial_x)p(\lambda^*, x)|_{x=0} \leq p(\lambda^*, -1) = 0.$$

□

More Results

The interlacing family for permutations in [MSS18] and the interlacing family for the paving problem in [RL20] are also very interesting.

17.4 References

[AGM+17] Arash Asadpour, Michel X. Goemans, Aleksander Madry, Shayan Oveis Gharan, and Amin Saberi. An $O(\log n/\log \log n)$-approximation algorithm for the asymmetric traveling salesman problem. *Oper. Res.*, 65(4):1043–1061, 2017. 164

[AO14] Nima Anari and Shayan Oveis Gharan. The Kadison-Singer problem for strongly Rayleigh measures and applications to asymmetric TSP. *arXiv preprint arXiv:1412.1143*, 2014. 161, 164, 168

[AO15] Nima Anari and Shayan Oveis Gharan. Effective-resistance-reducing flows, spectrally thin trees, and asymmetric TSP. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 20–39. IEEE, 2015. 165

[Coh16] Michael Cohen. Talk of "Improved spectral sparsification and kadison-singer for sums of higher-rank matrices" at Banff. 2016. https://www.birs.ca/events/2016/5-day-workshops/16w5111/videos/watch/201608011534-Cohen.html. 168

[HO14] Nicholas J. A. Harvey and Neil Olver. Pipage rounding, pessimistic estimators and matrix concentration. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 926–945. SIAM, 2014. 164

[KLS20] Rasmus Kyng, Kyle Luh, and Zhao Song. Four deviations suffice for rank 1 matrices. *Advances in Mathematics*, 375:107366, 2020. 166

[LZ20] Lap Chi Lau and Hong Zhou. A spectral approach to network design. In *Proccedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020, Chicago, IL, USA, June 22-26, 2020*, pages 826–839. ACM, 2020. 109, 110, 114, 167

[MSS18] Adam W. Marcus, Daniel A. Spielman, and Nikhil Srivastava. Interlacing families IV: bipartite Ramanujan graphs of all sizes. *SIAM J. Comput.*, 47(6):2488–2509, 2018. 142, 169

[RL20] Mohan Ravichandran and Jonathan Leake. Mixed determinants and the Kadison–Singer problem. *Mathematische Annalen*, 377(1):511–541, 2020. 169

[Tro12] Joel A Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12(4):389–434, 2012. 93, 166