Recent Advances in Approximation Algorithms **Spring 2015**

Lecture 15 & 16: Proof of Kadison-Singer Conjecture and the Extensions

Lecturer: Shayan Oveis Gharan *May 18th and 20th*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

In this lecture we finally prove the Kadison-Singer conjecture. First, we recall the main statement of the theorem.

Theorem 15.1 (Marcus, Spielman, Srivastava [MSS13b]). *Given a set of vectors $v_1, \ldots, v_m \in \mathbb{R}^d$ in isotropic position, if $\max_{1 \leq i \leq m} \|v_i\|^2 \leq \epsilon$, then there is a 2 partitioning S_1, S_2 of $[m]$ such that for each $j \in \{1, 2\}$,*

$$1/2 - O(\sqrt{\epsilon}) \leq \left\| \sum_{i \in S_j} v_i v_i^\intercal \right\| \leq 1/2 + O(\sqrt{\epsilon})$$

Recall that we are interested in Kadison Singer's theorem to prove the existence of (spectrally) thin trees. The following is a direct corollary of the above theorem.

Corollary 15.2. *Any graph G has an $O(\max_{e \in G} \text{Reff}(e))$-spectrally thin tree.*

Proof sketch. e sketch the main idea of the proof. Let $\epsilon = \max_{e \in E} \text{Reff}(e)$. First we need to construct an isotropic set of vector. For any edge $e \in E$ we let $v_e = L_G^{\dagger/2} b_e$. It is easy to see that $\sum_{e \in E} v_e v_e^\intercal = I$ and $\|v_e\|^2 = \text{Reff}(e)$. So, by the above theorem, the edges of G can be partitioned into two sets E_1, E_2 such that for $j \in \{1, 2\}$,

$$(1/2 - O(\sqrt{\epsilon}))L_G \preceq L_{E_j} \preceq (1/2 + O(\sqrt{\epsilon}))L_G.$$

The left inequality implies that the effective resistance of each edge of $e \in E_1$ with respect to L_{E_1} is about $2\text{Reff}(e)$. So, we can recursively divide E_1, E_2 into two subgraphs until the effective resistance are close to 1. after $\log(1/\epsilon) - o(\log(1/\epsilon))$ divisions we get to a $O(\epsilon)$ thin (connected) subgraph of G. □

The general strategy to prove the above theorem is very similar to what we did in the previous lecture. We define random vectors based on $v_1, \ldots, v_m$ and we show that in the corresponding random matrix there is a point of small norm. Perhaps the simplest

We construct random vectors $r_1, \ldots, r_m \in \mathbb{R}^{2d}$ where for each $1 \leq i \leq m$, $r_i = \sqrt{2} \begin{pmatrix} v_i \\ 0^d \end{pmatrix}$ with probability $1/2$ and $r_i = \sqrt{2} \begin{pmatrix} 0^d \\ v_i \end{pmatrix}$ with probability $1/2$. Observe that this implies that for each i, $\mathbb{E} \|r_i\|^2 = 2 \|v_i\|^2$. In addition,

$$\sum_{i=1}^m \mathbb{E} r_i r_i^\intercal = \begin{pmatrix} 2 \sum_{i=1}^m v_i v_i^\intercal & 0 \\ 0 & 2 \sum_{i=1}^m v_i v_i^\intercal \end{pmatrix} = I_{2d}$$

Therefore, we can reformulate Theorem 15.1 in terms of the random vectors $r_1, \ldots, r_m$ as follows. The following is indeed the main theorem of [MSS13b].

Theorem 15.3 (Marcus, Spielman, Srivastava [MSS13b]). *If $\epsilon > 0$ and $r_1, \ldots, r_m$ are independent random vectors in $\mathbb{R}^d$ with finite support such that*

$$\sum_{i=1}^m \mathbb{E} r_i r_i^\intercal = I,$$

and for all i,

$$\mathbb{E}\,\|r_i\|^2 \leq \epsilon,$$

then

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

It is easy to see that the above theorem implies Theorem 15.1. Letting $r_1, \ldots, r_m$ be as defined above, we get that there exists a point in the support, i.e., a partitioning S_1, S_2 of $[m]$ where S_1 are the coordinates i where $r_i = \sqrt{2}\begin{pmatrix} v_i \\ 0^d \end{pmatrix}$ and S_2 are the rest of the coordinates, such that

$$\begin{pmatrix} 2\sum_{i\in S_1} v_i v_i^{\intercal} & 0 \\ 0 & 2\sum_{i\in S_2} v_i v_i^{\intercal} \end{pmatrix} \preceq (1+\sqrt{2\epsilon})^2 I_{2d}.$$

This implies that for $j \in \{1,2\}$,

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

Since $v_1, \ldots, v_m$ are in isotropic position, the minimum eigenvalues of each of the above matrices is at least $1 - (1+\sqrt{2\epsilon})^2/2$. This proves Theorem 15.1.

Let us conclude this section by a few remarks about the above theorem. The first remark is about the importance of the isotropic assumption in the above theorems.

Remark 15.4. *Observe that the isotropic assumption is not necessary for Theorem 15.3. Even if $r_1, \ldots, r_m$ are sub-isotropic, i.e., if $\sum_{i=1}^m \mathbb{E} r_i r_i^{\intercal} \preceq I$, then the conclusion holds. To see that, observe that given sub-isotropic vectors $r_1, \ldots, r_m$ we can add several other (deterministic) vectors of small norm such that the whole set is in the isotropic position. Then, by the above theorem there is a point of small norm in the support of the probability distribution. But removing the artificially added vectors can only decrease the norm of the point in the support. On the other hand, the isotropic assumption is crucially used in the proof of Theorem 15.1 (using Theorem 15.3).*

The second remark is about the limitation of the method of interlacing polynomials.

Remark 15.5. *As we alluded to in the previous lecture, the main (existential) step of the proof is that for any set of (univariate) real rooted polynomials $p_1, \ldots, p_m$ that have a common interlacer, there is $1 \leq i \leq m$ such that the maximum root of p_i is at most the maximum root of $p_1 + \cdots + p_m$. Similar arguments show that for any $k \geq 1$ there exists $1 \leq k_i \leq m$ such that the k-th largest root of p_{k_i} is at most the k-th largest root of $p_1 + \cdots + p_m$.*

The main disadvantage of this argument is that it gives a bound only on a single root. For example, it is not possible to say that there is $1 \leq i \leq m$ such that the maximum root of p_i is at most the maximum root of the sum and the smallest root of p_i is at most the smallest root of the sum. That is the reason that Marcus, Spielman and Srivastava in their main theorem 15.3 only upper bound one root, i.e., the maximum root, of a point in the support of the distribution of independent random vectors.

Our last comment is about the style of the proof of the above theorem. Recall that in the last lecture we proved a seemingly similar statement where the covariance matrices $\mathbb{E} r_i r_i^{\intercal}$ are multiplies of the identity so they commute. This implies that all of the expected characteristic polynomials are simple linear transformation of the type $(1-D)$ of univariate real rooted polynomials. In the above statement we need to study sums of independent rank 1 matrices which are from non identically distribution random vectors. Because of this we need to work with multivariate *real stable* polynomials. The proof correspondingly need to use multivariate differential operators and multivariate barrier functions.

15.1 Extensions and Applications of Theorem 15.3

Before getting to the proof of the above theorem we talk about the extensions and applications of the above theorem. The first extension of Theorem 15.3 is by Brändén [Bra14] who used the techniques in [MSS13b] to prove upper bound on the maximum root of hyperbolic polynomials (as opposed to the characteristic polynomials) defined on rank 1 vectors[1] on the closure of the hyperbolicity cone. We will not elaborate on this and we refer interested readers to [Bra14] for details.

As we mentioned earlier, our main motivation in studying [MSS13b] is for the application in existence of (spectrally) thin tree. It turns out that Corollary 15.2 is not strong enough for our applications in bounding the integrality gap of the Held-Karp relaxation. We need to prove a stronger version of the above theorem. Apart from this specific application it is quite interesting to understand the importance of the independence assumption in the statement of the above theorem. In other words, under what families of distributions on the random vectors $v_1, \dots, v_m$ can we expect to have a point of small norm in the support of the distribution. The following theorem partially answers this question, roughly speaking the above theorem holds for any (homogeneous) strong Rayleigh distribution of vectors.

Theorem 15.6 (Anari, Oveis Gharan [AO14])**.** *Let μ be a homogeneous strongly Rayleigh probability distributions on $[m]$ such that the marginal probability*[2] *of each element is at most ϵ_1, and let $v_1, \dots, v_m \in \mathbb{R}^d$ be vectors such that*

$$\sum_{i=1}^{m} v_i v_i^{\intercal} = I,$$

and for all i, $\|v_i\|^2 \le \epsilon_2$. Then,

$$\mathbb{P}_{S\sim\mu}\left[\left\|\sum_{i\in S} v_i v_i^{\intercal}\right\| \le O((\epsilon_1 + \epsilon_2)\right] > 0.$$

Similar to Remark 15.4 even if $v_1, \dots, v_m$ are sub-isotropic in the above theorem, i.e., if $\sum_{i=1}^{m} v_i v_i^{\intercal} \preceq I$, still the conclusion holds because we can fill up the space with vectors of small square norm and zero marginal probability. The main advantage of the above theorem compared to Theorem 15.3 is that we can conclude there is a set of small norm *in the support of* μ. As alluded to in Remark 15.5, method of interlacing polynomials only gives a bound on one eigenvalue, this is the reason that restricting the family of feasible solutions to those in the support of μ is nontrivial.

It is an interesting question to find out other families of probability distributions that satisfy the conclusion of the above theorem. In this course we will not prove the above theorem and we refer interested readers to [AO14].

Next, we explain an application of the above theorem in the thin basis problem and then we use that to prove a weak sufficient condition for the existence of spectrally thin trees in a graph G.

Corollary 15.7 (Thin Basis Problem)**.** *Given a set of vectors $v_1, \dots, v_m \in \mathbb{R}^d$ in sub-isotropic position,*

$$\sum_{i=1}^{m} v_i v_i^{\intercal} \preceq I,$$

if $\max_{1\le i\le m} \|v_i\|^2 \le \epsilon$, and the set $\{v_1, \dots, v_m\}$ contains k disjoint bases, then there exists a basis $T \subseteq [m]$ such that

$$\sum_{i\in T} v_i v_i^{\intercal} \preceq O(\epsilon + 1/k)I.$$

[1] Suppose h is a hyperbolic polynomial with respect to $e \in \mathbb{R}^d$. We say x has rank 1 with respect to h if the univariate polynomial $h(x - te)$ has only one nonzero root.

[2] Recall that the marginal probability of i is $\mathbb{P}_{S\sim\mu}[i \in S]$.

Proof. For any basis T of $v_1, \ldots, v_m$ let

$$\mu(T) = \det(\sum_{i \in T} v_i v_i^\intercal),$$

be the square of the volume of the vectors in T. First, observe that μ is a strong Rayleigh distribution. This is because

$$\det(\sum_{i=1}^m z_i v_i v_i^\intercal) = \sum_T \mu(T) z^T,$$

where as usual $z^T = \prod_{i \in T} z_i$. In addition μ is a homogeneous distribution because any basis T has exactly d vectors. To use Theorem 15.6 we only need to make sure that the marginal probability of each vector is small. It is easy to see that if $\sum_{i=1}^m v_i v_i^\intercal = I$, then for any i, $\mathbb{P}[i \in T] = \|v_i\|^2 \leq \epsilon$. However, when we have the sub-isotropic assumption this is no longer the case. In fact by the Matrix tree theorem that we discussed in Lecture 3 for any i,

$$\mathbb{P}[i \in T] = v_i^\intercal \left(\sum_{j=1}^m v_j v_j^\intercal \right)^{-1} v_i,$$

which may be very close to 1.

So, we need to perturb the distribution $\mu(.)$ to decrease the marginal probabilities, this is very we crucially use the fact that $\{v_1, \ldots, v_m\}$ contains k disjoint bases. Let $T_1, \ldots, T_k$ be these bases.

The idea is to use the maximum entropy convex program to assign weights $w(i)$ to each vector v_i with and then to choose each basis T with probability $\mu'(T) = \mu(T) \prod_{i \in T} w(i)$. Such a distribution is an external field operator on μ, so μ' is strong Rayleigh. The only nontrivial step is to choose the weights such that the marginal probability of each vector is at most $O(1/k)$. To do that, we use the maximum entropy convex program.

$$\begin{aligned} \min \quad & \sum_T \mu(T) w(T) \log w(T) \\ \text{subject to} \quad & \sum_{T: i \in T} \mu(T) w(T) = 1/k \quad \forall i \in \bigcup_{i=1}^k T_i, \\ & p(T) \geq 0. \end{aligned} \tag{15.1}$$

Since $T_1, \ldots, T_k$ are k disjoint bases the point $\frac{T_1}{k} + \cdots + \frac{T_k}{k}$ is inside the convex hull of the indicator vectors of the all bases of $\{v_1, \ldots, v_m\}$. So, the above convex program is feasible. To make sure that the Sleator conditions hold we need to make sure that the point $\frac{T_1}{k} + \cdots + \frac{T_k}{k}$ is in the interior of this convex hull; that is achievable by slightly moving this point away from the faces of the polytope.

Using KKT conditions similar to Lecture 7 we can show that the optimum $w(.)$ is a product distribution, that is we can assign nonnegative weights to the vectors such that for any base T, $w(T) = \prod_{i \in T} w(i)$. □

An implication of the above corollary is that if a graph G has edges with effective resistance very close to 1 but in any cut of G there are several edges with small effective resistance, then G has a spectrally thin tree.

Corollary 15.8. *Given a graph $G = (V, E)$, let $F = \{e : \text{Reff}(e) \leq \epsilon\}$. If (V, F) is a k-edge connected subgraph of G, then G has a spanning tree $T \subseteq F$ such that*

$$L_T \preceq O(\epsilon + 1/k) L_G.$$

Proof. Let $v_e = L_G^{\dagger/2} b_e$ for each $e \in F$. Then, by the definition of F, for each edge $e \in F$,

$$\|v_e\|^2 = b_e L_G^\dagger b_e = \text{Reff}(e) \leq \epsilon.$$

In addition,

$$\sum_{e \in F} v_e v_e^\intercal = L_G^{\dagger/2} \left(\sum_{e \in F} b_e b_e^\intercal \right) L_G^{\dagger/2} = L_G^{\dagger/2} L_F L_G^{\dagger/2} \preceq L_G^{\dagger/2} L_G L_G^{\dagger/2} = I_{n-1},$$

where the matrix inequality follows by the fact that for any symmetric matrix C, if $A \preceq B$ then $CAC \preceq CBC$. Now, by Corollary 15.7 □

As we will see in the future lectures, the above corollary is a major tool in the recent improved bounds on the integrality gap of the Held-Karp LP relaxation for Asymmetric TSP.

15.2 Expected Characteristic Polynomial

The general strategy to prove Theorem 15.3 is similar to what we did in the last lecture. First, we show that

$$\mathbb{E} \det(tI - r_1 r_1^\intercal - \cdots - r_m r_m^\intercal)$$

is a real rooted polynomial, then we show that there is a point in the support of the independent distribution whose largest root is at most the largest root of the above polynomial. Finally, we use an extension of the barrier argument to upper bound the the maximum root of the above expected characteristic polynomial is small.

In this section we write the expected characteristic polynomial in terms of (multivariate) differential operators. Let us open up the above polynomial. Using Theorem 3.2 of Lecture 3 we can write

$$\begin{aligned} \mathbb{E} \det(tI - r_1 r_1^\intercal - \cdots - r_m r_m^\intercal) &= \sum_{k=0}^{d} t^{d-k} (-1)^k \mathbb{E} \left[\sum_{S \in \binom{[m]}{k}} \det_k \left(\sum_{i \in S} r_i r_i^\intercal \right) \right] \\ &= \sum_{k=0}^{d} t^{d-k} (-1)^k \sum_{S \in \binom{[m]}{k}} \mathbb{E} \det_k \left(\sum_{i \in S} r_i r_i^\intercal \right) \qquad (15.2) \end{aligned}$$

Note that the above identities crucially use that each $r_i r_i^\intercal$ is a rank 1 matrix. Observe that similar to the previous lecture each term $\mathbb{E} \det_k(\sum_{i \in S} r_i r_i^\intercal)$ is a linear function of $\mathbb{E} r_i r_i^\intercal$. Therefore, similar to the previous lecture, the above equation does not depend on the actual vectors in the support of $r_1, \ldots, r_m$. However, the previous analysis was significantly easier, we had $\mathbb{E} \det_k(\sum_{i \in S} r_i r_i^\intercal) = d!/(d-k)!$ for any set S of size k because $\mathbb{E} r_i r_i^\intercal = I$ for all i. This made the description of the mixed characteristic polynomial simpler.

Suppose for each $1 \leq i \leq m$,

$$r_i = \begin{cases} v_{i,1} & \text{with prob } p_{i,1} \\ v_{i,2} & \text{with prob } p_{i,2} \\ \vdots & \\ v_{i,\ell_i} & \text{with prob } p_{i,\ell_i}. \end{cases}$$

If we use a univariate mixed characteristic polynomial,

$$f(D) \det(tI - \mathbb{E} r_1 r_1^\intercal - \cdots - \mathbb{E} r_m r_m^\intercal).$$

for some differential operator $f(D)$, we get k-th order determinants involving multiple vectors in the support of a single random vector r_i, e.g., $\det_3(v_{1,1} v_{1,1}^\intercal + v_{1,2} v_{1,2}^\intercal + v_{2,1} v_{2,1}^\intercal)$. But there is no such term in (15.2). So, we use multivariate polynomials to eliminate any such term. Here is the idea: Suppose we scale every vector in the support of r_i with an indeterminant z_i. Then, any bad term will have indeterminants with exponents more than 1, for example the above $\det_3(.)$ will scale by $z_1^2 z_2$. Now, we can just eliminate any non-multilinear monomial, simply by differentiating and zeroing out the indeterminants, e.g., if we differentiate by $\partial_1 \partial_2$ and the we let $z_1 = z_2 = \cdots = 0$ then the monomial $z_1^2 z_2$ will map to zero.

We use the following mixed characteristic polynomial

$$\prod_{i=1}^{m}(1-\partial_i)\det(tI - z_1 \mathbb{E} r_1 r_1^{\intercal} - \cdots - z_m \mathbb{E} r_m r_m^{\intercal})\Big|_{z_1=\cdots=z_m=0},$$

where as usual $\partial_i = \partial/\partial z_i$. Now, for any set $S \subseteq [m]$ of size k (suppose $S = \{1, 2, \ldots, k\}$ after renaming the vectors) we have

$$\begin{aligned}\left(\prod_{i=1}^{k}\partial_i\right)\det\left(tI - \sum_{i=1}^{m} z_i \mathbb{E} r_i r_i^{\intercal}\right)\Big|_{z_1=\cdots=z_m=0} &= t^{d-k}\sum_{j_1\in[\ell_1],\ldots,j_k\in[\ell_k]}\left(\prod_{i=1}^{k} p_{i,j_j}\right)\det_k\left(\sum_{i=1}^{k} v_{i,j_i} v_{i,j_i}^{\intercal}\right)\\ &= t^{d-k}\mathbb{E}\det_k\left(\sum_{i=1}^{k} r_i r_i^{\intercal}\right),\end{aligned}$$

where the second identity uses the independence of $r_1, \ldots, r_k$. The RHS is exactly what we had in (15.2). So, we get the following lemma.

Lemma 15.9. *If $r_1, \ldots, r_m$ are independent random vectors then*

$$\mathbb{E}\det(tI - r_1 r_1^{\intercal} - \cdots - r_m r_m^{\intercal}) = \prod_{i=1}^{m}(1-\partial_i)\det(tI + z_1 \mathbb{E} r_1 r_1^{\intercal} + \cdots + z_m \mathbb{E} r_m r_m^{\intercal})\Big|_{z_1=\cdots=z_m=0}.$$

We call the polynomial in the RHS the *mixed characteristic polynomial* and we denote it by $\mathcal{I}[\mathbb{E} r_1 r_1^{\intercal}, \ldots, \mathbb{E} r_m r_m^{\intercal}](t)$.

Remark 15.10 (Computability of Mixed Polynomial)**.** *It turns out that the coefficients of the mixed characteristic polynomial can be #P-hard to compute exactly. Because of that the proof that we elaborate in this lecture is not algorithmic. The #P-hardness follows by carefully choosing the random vectors $r_1, \ldots, r_m$ such that the coefficients of the expected characteristic polynomial are the number matchings of sizes $1, \ldots, m$ in a bipartite graph of $2m$ vertices. Here, we do not give more details and we refer interested readers to [MSS13a].*

We can extend the above lemma and write the expected characteristic polynomial when the vectors $r_1, \ldots, r_m$ are coming from a strong Rayleigh distribution.

Lemma 15.11. *For $v_1, \ldots, v_m \in \mathbb{R}^d$ and a homogeneous strong Rayleigh distribution $\mu : 2^{[m]} \to \mathbb{R}_+$,*

$$t^{d_\mu - d}\mathbb{E}_{T\sim\mu}\left[\det\left(t^2 I - \sum_{i\in T} 2 v_i v_i^{\intercal}\right)\right] = \prod_{i=1}^{m}(1-\partial_i^2)\left(g_\mu(t+z_1, \ldots, t+z_m)\cdot\det\left(tI + \sum_{i=1}^{m} z_i v_i v_i^{\intercal}\right)\right)\Big|_{z_1=\cdots=z_m=0},$$

where g_μ is the generating polynomial of μ (see Lecture 11 for the definition), and d_μ is the degree of g_μ, or equivalently the size of the samples of μ.

We will use $\mu[v_1, \ldots, v_m](t)$ to denote the polynomial in the RHS of the above. Let us open up the LHS of the above equation.

$$\mathbb{E}_{T\sim\mu}\left[\det(tI - \sum_{i\in T} v_i v_i^{\intercal})\right] = \sum_{k=0}^{d_\mu} t^{d_\mu - k}(-1)^k \sum_{S\in\binom{[m]}{k}} \mathbb{P}_{T\sim\mu}\left[S\subseteq T\right]\det_k\left(\sum_{i\in S} v_i v_i^{\intercal}\right).$$

Note that because of the (negative) dependence between the underlying elements of μ, we can not simply write a term $\mathbb{P}[T\sim\mu]\, S\subseteq T \det_k(\sum_{i\in S} v_i v_i^\intercal)$ as expectation of an independent distribution. The idea is to note that $\mathbb{P}_{T\sim\mu}[S\subseteq T]$ is simply the coefficient of z^S in the polynomial $g_\mu(t\mathbf{1}+z)$, and, $\det_k(\sum_{i\in S} v_i v_i^\intercal)$ is the coefficient of z^S in $\det(tI+\sum_{i=1}^m z_i v_i v_i^\intercal)$. Since $g_\mu(t\mathbf{1}+z)$ and $\det(tI+\sum_{i=1}^m z_i v_i v_i^\intercal)$ are multilinear polynomials, if we apply take the partial derivative $\prod_{i\in S}\partial_i^2$ and then we zero out all variables, the product of the above two terms survive,

$$\prod_{i\in S}\partial_i^2 (g_\mu(t+z_1,\dots,t+z_m)\cdot \det(tI+\sum_{i=1}^m z_i v_i v_i^\intercal)) = t^{d+d_\mu-2k}\mathbb{P}_{T\sim\mu}[S\subseteq T]\cdot \det_k\left(\sum_{i\in S} v_i v_i^\intercal\right).$$

The above lemma simply follows by taking the sum over all sets of size at most d_μ.

Let us give an explicit example of $\mu[v_1,\dots,v_m](t)$.

Example 15.12. *Let G be the ℓ dimensional hypercube, μ be the uniform distribution on all spanning trees of G and $v_{e_i} = L_G^{\dagger/2} b_{e_i}$ for any edge e_i. Let Y be the transfer current matrix that we studied in the first few lecture, i.e., $Y_{e_i,e_j} = b_{e_i} L_G^\dagger b_{e_j}$. Then, for any set S of k edges,*

$$\det_k\left(\sum_{e_i\in S} v_{e_i} v_{e_i}^\intercal\right) = \det(Y_S) = \mathbb{P}_{T\sim\mu}[S\subseteq T].$$

Therefore,

$$\mathbb{E}_{T\sim\mu}\left[\det(tI - \sum_{e_i\in T} v_{e_i} v_{e_i}^\intercal)\right] = \sum_{k=0}^{n-1} t^{n-1-k}(-1)^k \sum_{S\in\binom{[m]}{k}} \det(Y_S)^2.$$

For example, for $k=1$, the inner sum in the RHS is equal to $\sum_e \mathrm{Reff}(e)^2$. It turns out that although we know a lot about the effective resistance of edges and their sums, we know very little about sum of square of effective resistance (or in general $\sum_{\binom{S\in[m]}{k}} \det(Y_S)^2$). In part (c) of Problem 4 of Assignment 2 we proved upper bounds on sum of square of effective resistance of edges in a k connected graph. That is one of the few examples that we know how to analyze these sums, but there is a lot to be done in this direction.

15.3 Interlacing

First, observe that the mixed characteristic polynomial $\mathcal{I}[A_1,\dots,A_m](t)$ is real rooted for any set of matrices $A_1,\dots,A_m$. This simply follows by the closure properties of real stable polynomials that we discussed in Lecture 10. First,

$$\det(tI + z_1 A + \dots + z_m A_m)$$

is real stable by Lemma 10.10. Second,

$$\prod_{i=1}^m (1-\partial_i)\det(tI + z_1\mathbb{E}A_1 + \dots + z_m A_m)$$

is real stable by the closure property of the operators $1-\partial_i$. Finally, $\mathcal{I}[A_1,\dots,A_m](t)$ is real stable just by the closure of specialization operators, i.e., $z_1=\dots=z_m=0$ preserves real stable. Since $\mathcal{I}[A_1,\dots,A_m](t)$ is univariate, it must be real rooted.

Lemma 15.13. *For any PSD matrices $A_1,\dots,A_m$, $\mathcal{I}[A_1,\dots,A_m](t)$ is real rooted.*

Similarly, we can argue that $\mu[v_1, \ldots, v_m](t)$ is real rooted. To start recall that the product of any two real stable polynomials is real stable, since μ is strong Rayleigh, g_μ is real stable, so

$$g_\mu(t+z_1, \ldots, t+z_m) \cdot \det(tI + z_1 v_1 v_1^\intercal + \cdots + z_m v_m v_m^\intercal)$$

is real stable. Note that if $p(z_1, \ldots, z_m)$ is a real stable polynomial then so is $p(t+z_1, \ldots, t+z_m)$ for a new variable t. Therefore, $\mu[v_1, \ldots, v_m](t)$ is real stable by the closure property of $1 - \partial_i^2$ operators. Take a look at Lecture 11 to see general form of differential operators that preserve real stability.

The following is the main statement of this section

Lemma 15.14. *With positive probability,*

$$\lambda_{\max}(\det(tI - r_1 r_1^\intercal - \cdots - r_m r_m^\intercal)) \leq \lambda_{\max}(\mathbb{E} \det(tI - r_1 r_1^\intercal - \cdots - r_m r_m^\intercal)).$$

Similar to the previous lecture we iteratively round $r_1, \ldots, r_m$. So, we show the first step and the rest simply follows by induction.

Lemma 15.15. *With positive probability,*

$$\lambda_{\max}(\mathbb{E}_{r_2,\ldots,r_m} \det(tI - r_1 r_1^\intercal - \cdots - r_m r_m^\intercal)) \leq \lambda_{\max}(\mathbb{E}_{r_1,\ldots,r_m} \det(tI - r_1 r_1^\intercal - \cdots - r_m r_m^\intercal)).$$

Similar to the previous lecture all we need to do is to show that the polynomials $\mathbb{E}_{r_2,\ldots,r_m} \det(tI - r_1 r_1^\intercal - r_m r_m^\intercal)$ have a common interlacer and then the above lemma follows by Claim 14.13 of Lecture 14.

To show that these polynomials have a common interlacer we use Theorem 10.25 (Lecture 10) by Dedieu.

Theorem 15.16. *Two real rooted polynomials p, q have a common interlacer if and only if for any $a, b \geq 0$, $ap + bq$ is real rooted.*

Note the difference with Obrechkoff, Dedieu theorem that we used in the last lecture, if $ap + bq$ is real rooted for any $a, b \in \mathbb{R}$ then p, q are interlacing. In other words, one can read the above theorem as that real rooted polynomials that interlace a polynomial p form a convex set.

Let $v_{1.1}, \ldots, v_{1,\ell_1}$ be the vectors in the support of r_1. By the above theorem all we need to show is that for any nonnegative numbers $a_1, \ldots, a_{\ell_1}$,

$$\sum_{i=1}^{\ell_1} a_{1,i} \mathbb{E} \det(tI - v_{1,i} v_{1,i}^\intercal - r_2 r_2^\intercal - \cdots - r_m r_m^\intercal),$$

is real rooted. But then we can simply define a new random vector r_1' where $r_1' = v_{1,i}$ with probability $a_{1,i}$. Then by Lemma 15.13, $\mathcal{I}[\mathbb{E} r_1' {r_1'}^\intercal, \ldots, \mathbb{E} r_m r_m^\intercal](t)$ is real rooted so the above polynomial is real rooted.

15.4 Multivariate Barrier Argument

In this section we upper bound the largest root of the mixed characteristic polynomial.

Lemma 15.17. *Let $A_1, \ldots, A_m$ be PSD matrices such that $\sum_{i=1}^m A_i = I$ and $\mathrm{Tr}(A_i) \leq \epsilon$ for all i. Then the largest root of $\mathcal{I}[A_1, \ldots, A_m](t)$ is at most $(1+\sqrt{\epsilon})^2$.*

First observe that

$$\mathcal{I}[A_1, \ldots, A_m](t) = \prod_{i=1}^m (1 - \partial_i) \det\left(\sum_{i=1}^m z_i A_i\right)\Bigg|_{z_1 = \cdots = z_m = t}.$$

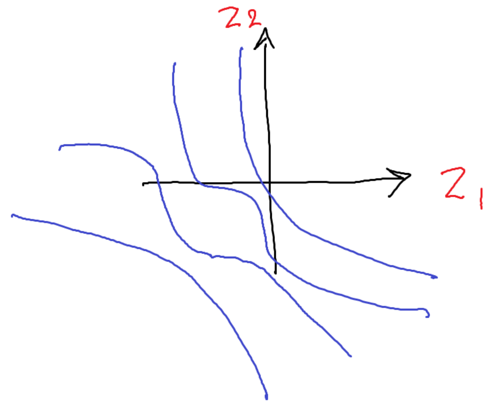

Figure 15.1: An example of the roots of a degree 4 homogeneous real stable polynomial in $\mathbb{R}^2$.

This is because $tI + \sum_{i=1}^m z_i A_i = \sum_{i=1}^m (z_i + t) A_i$, so instead of letting $z_i = 0$ we can let $z_i + t = t$.

The polynomial

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

is real stable, so for vector $e \in \mathbb{R}^m_{>0}$ if we substitute $z = x - te$ we get a real rooted polynomial. To prove the above lemma we need to find the maximum root for e being the all 1 vector and $x = 0$.

The general strategy is similar to what we had in the previous lecture. First of all, the polynomial

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

is a real stable homogeneous polynomial. So, the largest root along the all 1 vectors is zero. The simple way to see this is to note that $A_1 + \cdots + A_m = I$ so if $z_1 = \cdots = z_m = t > 0$, then $\det(z_1 A_1 + \cdots + z_m A_m) = t^d$. But, indeed any homogeneous real stable polynomial has the positive orthant in the hyperbolicity cone of the all 1 vector. So, any homogeneous real stable polynomial $p(z)$ has no root for any $z \in \mathbb{R}_{>0}$. See Figure 15.1 for the structure of the roots of a homogeneous real stable polynomial in $\mathbb{R}^2$.

When we apply the $1 - \partial_i$ operators, we loose the homogeneity and the largest root can get bigger and bigger. In the last lecture we proved that for any real rooted polynomial p, the largest root of $(1 - D)p$ is at most $1/(1-\alpha)$ bigger than the largest root of p. Such a bound is not good in our case because $1/(1-\alpha) \geq 1$. Since we apply m differential operators such a bound can only give an upper bound of m on the largest root.

Here, we will use the fact that indeed we are dealing with multivariate (real stable) polynomials. So, the upper barrier should not be a single number, because along every direction along the positive orthant we see a different set of roots. Instead, the upper barrier of a real stable polynomial $p(.)$ will be a point in $z \in \mathbb{R}^m$ such that for any point $z' \geq z$, $p(z') > 0$, where $z' \geq z$ if for all $1 \leq i \leq m$, $z'_i \geq z_i$. For example, in Figure 15.1, the all zero vector is an upper barrier. The goal is to show that for a carefully chosen starting upper barrier $z = t\mathbf{1}$ of $\det(z_1 A_1 + \cdots + z_m A_m)$ whenever we apply a $1 - \partial_i$ operator all we need to do is to shift the upper barrier along the direction $\mathbf{1}_i$ by $1/(1-\alpha)$, for an appropriate choice of α. If we prove that then by the time that we get to $\prod_{i=1}^m (1 - \partial_i)$ we have moved the upper barrier by $1/(1-\alpha)$ along the all ones direction. This means that the largest root of the expected characteristic polynomial is smaller than $1/(1-\alpha) + t$. At the end of the day we will have $t = \sqrt{\epsilon} + \epsilon$ and $\alpha = \epsilon/t \approx \sqrt{\epsilon}$.

We will use barrier functions analogous to what we had in the last lecture. For any $1 \le i \le m$ we define

$$\Phi_i^p(z) = \frac{\partial_i p(z)}{p(z)}.$$

Fix $z_2, \ldots, z_m$, then $p(z)$ is a univariate real rooted polynomial of z_1. So,

$$q(z_1) = p(z_1, z_2, \ldots, z_m) = \prod_{i=1}^{d} (z_1 - \lambda_i).$$

Then, similar to the previous lectures,

$$\Phi_i^p(z) = \sum_{i=1}^{d} \frac{1}{z_1 - \lambda_i}.$$

We will go over the details, we just elaborate new ideas of the proof. The main lemma of this part of the proof is the following.

Lemma 15.18. *Suppose $p(z_1, \ldots, z_m)$ is a real stable polynomial and z is an upper barrier of p. If*

$$\delta \ge \frac{1}{1 - \Phi_j^p(z)},$$

then

$$\Phi_i^{(1-\partial_j)p}(z + \delta e_j) \le \Phi_i^p(z).$$

We emphasize that once we apply $1 - \partial_j$ just by shifting the upper barrier along $\mathbf{1}_j$ a little bit we can make sure that none of the barrier functions increase. We will not talk about the proof of the above lemma because of the similarity with the last lecture, instead we prove an essential step of the proof, namely monotonicity and convexity of barrier functions.

Recall that the univariate barrier arguments that we discussed in the last two lectures we proved a univariate version of the above lemma where we used two essential properties of the barrier functions, monotone decreasing, and convexity. The proof of the above lemma also exploits these two properties, but we need to show it between every pair of variables.

Lemma 15.19. *Suppose p is real stable and z is an upper barrier of p. Then for all $1 \le i, j \le m$ and $\delta \ge 0$,*

$$\begin{aligned} \Phi_i^p(z + \delta e_j) &\le \Phi_i^p(z), \\ \Phi_i^p(z + \delta e_j) - \delta \partial_j \Phi_i^p(z + \delta e_j) &\le \Phi_i^p(z). \end{aligned}$$

Note that if $i = j$ then the above lemma is almost trivial, because $1/t - \lambda_i$ is a monotone decreasing and convex function as we also argued in the last lecture. In particular, although p is multivariate, in this case we can just think of p as a univariate real rooted polynomial and the same proof follows. So, suppose $i \ne j$. The above lemma is the main place that we use the real stability of the polynomial p throughout the multivariate barrier argument. Roughly speaking, the above lemma follows from nice properties of the structure of the roots of real stable polynomials Take a look at Figure 15.2 for an example. Perhaps it is easiest to observe the above properties just by looking at the largest root of p. It is not hard to see that the diagram of the largest roots is convex (this is the same as the convexity of the hyperbolicity cone that we discussed in Lecture 10). Because of that when we move away from the origin along one coordinate, the largest root decreases.

There are several proof of Lemma 15.19. Perhaps the most elementary one is by Tao [Tao14]. The proof that we discuss here uses the characterization of real stable bivariate polynomials that we discussed in Lecture 10.

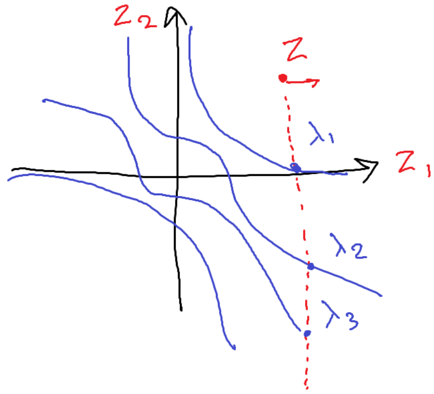

Figure 15.2: In this figure the barrier function Φ_2^p at point z is simply $\sum_{i=1}^m \frac{1}{z_2-\lambda_i}$. Observe that when we increase z_1, λ_i decreases for all i, so $\Phi_i^p(z+\delta e_j)$ is a decreasing function of δ.

Lemma 15.20. *If $p(z_1, z_2)$ is a bivariate real stable polynomial of degree exactly d, then there exist $d \times d$ PSD matrices A, B and a symmetric matrix C such that*

$$p(z_1, z_2) = \pm \det(z_1 A + z_2 B + C).$$

Now, we are ready to prove Lemma 15.19. *Proof of Lemma 15.19.* Let z be the upper barrier, fix all variables except z_i, z_j and let

$$\pm \det(z_i B_i + z_j B_j + C) = p(z_1, \ldots, z_m).$$

First, we argue that $z_i B_i + z_j B_j + C \succ 0$. First, note that since p is a degree d polynomial $z_i B_i + z_j B_j$ must have full rank. So, $z_i B_i + z_j B_j \succ 0$. Now, if $z_i B_i + z_j B_j + C \not\succ 0$, then we can slightly increase z_i, z_j and make the smallest eigenvalue 0 which makes the determinant 0. But, because z is an upper barrier that is not possible, so $z_i B_i + z_j B_j + C \succ 0$.

Let $M = z_i B_i + z_j B_j + C$ and note that M an invertible mapping of z_i, z_j. Now, let us write down $\Phi_i^p(z)$.

$$\begin{aligned} \Phi_i^p(z) &= \frac{\partial_i p(z)}{p(z)} \\ &= \frac{\partial_i \det(M)}{\det(M)} \\ &= \frac{\det(M) \operatorname{Tr}(M \bullet \partial_i M)}{\det(M)} = \operatorname{Tr}(M \bullet B_i). \end{aligned}$$

In the third equality we used part (b) of Problem 4 of Assignment 3. Also, observe that the above shows that the barrier function is nonnegative because $\operatorname{Tr}(AB) \geq 0$ for any two PSD matrices (see Problem 2 of Assignment 2).

Now, to prove the monotonicity it is enough to differentiate the above function

$$\begin{aligned} \partial_j \Phi_i^p(z) &= \operatorname{Tr}(M \bullet B_i) \\ &= \operatorname{Tr}(-M^{-1} \partial_j M M^{-1} B_i) \\ &= -\operatorname{Tr}(M^{-1} B_j M^{-1} B_i). \end{aligned}$$

The second equality uses Lemma 15.21 below. But the RHS is always nonpositive. This is because $M^{-1} B_j M^{-1} \succeq 0$ and $B_i \succeq 0$, so $\operatorname{Tr}(M^{-1} B_j M^{-1} B_i) \geq 0$. So, $\Phi_i^p(z)$ is a nonincreasing function of j.

Finally, to prove the convexity we show that the derivative of the above is nonnegative.

$$\begin{aligned}\partial_j \operatorname{Tr}(-M^{-1}B_jM^{-1}B_i) &= -\operatorname{Tr}(\partial_j(M^{-1}B_j)M^{-1}B_i) - \operatorname{Tr}(M^{-1}B_j\partial_j(M^{-1}B_i)) \\ &= \operatorname{Tr}(M^{-1}B_jM^{-1}B_jM^{-1}B_i) + \operatorname{Tr}(M^{-1}B_jM^{-1}B_jM^{-1}B_i).\end{aligned}$$

Similar to above, $M^{-1}(B_j(M^{-1})B_j)M^{-1} \succeq 0$ and $B_i \succeq 0$. So, both of the above traces are nonnegative. □

Lemma 15.21. *For an invertible matrix A which is a differentiable function of t,*

$$\frac{\partial A^{-1}}{\partial t} = -A^{-1}(\partial_t A)A^{-1}.$$

Proof. Differentiating both sides of the identity $A^{-1}A = I$ with respect to t, we get

$$A^{-1}\frac{\partial A}{\partial t} + \frac{\partial A^{-1}}{\partial t}A = 0.$$

Rearranging the terms and multiplying with A^{-1} gives the lemma's conclusion. □

References

[AO14] N. Anari and S. Oveis Gharan. "The Kadison-Singer Problem for Strongly Rayleigh Measures and Applications to Asymmetric TSP". 2014. URL: http://arxiv.org/abs/1412.1143 (cit. on p. 15-3).

[Bra14] P. Brändén. "Hyperbolic polynomials and the Marcus-Spielman-Srivastava theorem". 2014. URL: http://arxiv.org/abs/1412.0245 (cit. on p. 15-3).

[MSS13a] A. Marcus, D. A. Spielman, and N. Srivastava. "Interlacing Families I: Bipartite Ramanujan Graphs of All Degrees". In: *FOCS*. 2013, pp. 529–537 (cit. on p. 15-6).

[MSS13b] A. Marcus, D. A. Spielman, and N. Srivastava. "Interlacing Families II: Mixed Characteristic Polynomials and the Kadison-Singer Problem". 2013. URL: http://arxiv.org/abs/1306.3969 (cit. on pp. 15-1, 15-3).

[Tao14] T. Tao. *Real stable polynomials and the Kadison-Singer problem.* 2014. URL: https://terrytao.wordpress.com/tag/kadison-singer-problem/ (cit. on p. 15-10).