Recent Advances in Approximation Algorithms **Spring 2015**

Lecture 14: Spectrally Thin Forests using Interlacing Polynomials

Lecturer: Shayan Oveis Gharan *May 13th*

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

In the last lecture we showed that every k connected graph G has a linear sized $O(1/k)$-spectrally thin forest. We also briefly argued that if G has edges with effective resistance close to 1, then those edges cannot be in any spectrally thin subgraph of G. Suppose we are given a graph where the effective resistance of every edge is at most $1/k$. Can we say that this graph has a $O(1/k)$-spectrally thin tree? It turns out that the answer to this question is yes and it follows from the Kadison-Singer conjecture. That is indeed the motivation that we are talking about the proof of the Kadison-Singer conjecture by Marcus, Spielman and Srivastava [MSS13] in this class. In particular, we will prove the following theorem (note that this theorem is incomparable with the statement of the Kadison-Singer conjecture).

Theorem 14.1 ([MSS13; HO14]). *Given a set of vectors $v_1, \dots, v_m \in \mathbb{R}^d$ such that $\sum_{i=1}^m v_i v_i^\intercal = I$ and that for each $1 \le i \le m$, $\|v_i\|^2 \le \epsilon$. There exists a basis T, i.e., d linearly independent vectors $\{v_i\}_{i \in T}$ such that*

$$\left\| \sum_{i \in T} v_i v_i^\intercal \right\| \le O(\epsilon).$$

Unfortunately, the same technique that we used in the last lecture cannot be used to prove the above theorem without a logarithmic loss in the dimension. First, it is not hard to see that the same analysis can give us a basis of norm $O(\epsilon \cdot \log(d))$. This is because for any number ℓ, if $|F| = \ell$, in the next iteration, the upper barrier, t, will increase by at most $O(\frac{\epsilon}{d-\ell})$. So, by the time that we have a basis ($|F| = d$), the norm of the vectors in F is at most $O(\epsilon \cdot (1 + 1/2 + \cdots + 1/d)) = O(\epsilon \cdot \log d)$. It is instructive to take a look at [HO14] for an argument based on pipage rounding with an $O(\log d / \log \log d)$ loss.

Now, let us see that there is an underlying barrier to get a bound better than $O(\epsilon \cdot \log d)$ in the greedy algorithm. Roughly speaking, the impossibility is due to the memoryless property of the greedy algorithm. In particular, observe that in each iteration of the algorithm we choose one vector v_i from a given set of vectors $\{v_1, \dots, v_m\}$ where $\sum_{i=1}^m v_i v_i^\intercal = I$ such that v_i is linearly independent of the vectors that are already in F and that it only increases the upper barrier, t, by a small amount. The same analysis would work if the set of vectors that we feed into the algorithm in ℓ-th iteration is different from those that we feed in the $\ell + 1$-th iteration. It turns out that an adversary can feed in a bad sequence of vectors for which there is no basis of norm less than $O(\epsilon \cdot \log d)$. Let us elaborate an example by Harvey [Har14].

Example 14.2 ([Har14]). *Suppose we are at the $d - \ell + 1$ iteration of the greedy algorithm and $|F| = d - \ell$. As the vectors in F are linearly independent, $\ker(F) = \ell$. Let $u_1, \dots, u_\ell$ be an orthonormal basis of $\ker(F)$ and let u_0 be the largest eigenvector of the matrix $\sum_{i \in F} v_i v_i^\intercal$. We choose the vectors for the $d - \ell + 1$ iteration as follows: There are $(\ell + 1)/\epsilon$ vectors of the form*

$$\pm \sqrt{\frac{\epsilon}{\ell + 1}} u_0 \pm \cdots \pm \sqrt{\frac{\epsilon}{\ell + 1}} u_\ell,$$

where the $+$ or $-$ sign is chosen such that the sum of the quadratic forms of these vectors is $u_0 u_0^\intercal + \cdots + u_\ell u_\ell^\intercal$. The rest of the vectors are chosen such that each of them has square norm ϵ and the whole set add up to the identity. To choose a vector that is linearly independent of those that are already in F, the algorithm has to choose one of the above vectors. But, any such vector increases the spectral norm by $\epsilon/(\ell + 1)$. So, by the time where F is a basis the spectral norm is $O(\epsilon \cdot \log d)$.

In summary, to prove Theorem 14.1 we need to exploit a different technique. In this lecture we reprove (a weaker version) the thin forest result using a different technique. We will use the barrier argument to bound the norm, but instead of using the greedy algorithm we use properties of real rooted polynomials to find a set F of small norm. The materials of this lecture are based on the survey by Marcus, Spielman and Srivastava [MSS14]. We prove the following theorem.

Theorem 14.3. *Given a set of vectors $v_1, \ldots, v_m \in \mathbb{R}^d$, such that*

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

There exists a multiset F of vectors of size $|F| \geq \Omega(d)$ such that

$$\left\| \sum_{i \in F} v_i v_i^\intercal \right\| \leq O(d/m).$$

Note that unlike the previous lecture we do not guarantee that vectors in F are linearly independent, so in a sense we want to prove a weaker statement. But as we will see the technique has the potential to be extended to prove Theorem 14.1.

We would like to use the probabilistic method. Let $r_1, \ldots, r_n$ be uniformly random vectors where for all $1 \leq i \leq n$ we have $r_i = \sqrt{m} v_j$ with probability $1/m$. Note that $\mathbb{E} r_i r_i^\intercal = I$. To prove the above theorem, all we need to say is that with a positive probability

$$\left\| \sum_{i=1}^{n} r_i r_i^\intercal \right\| \leq O(d+n).$$

Theorem 14.4. *For any given independent random vectors $r_1, \ldots, r_n$ such that $\mathbb{E} r_i r_i^\intercal = I$ for all $1 \leq i \leq n$, with a positive probability,*

$$\left\| \sum_{i=1}^{n} r_i r_i^\intercal \right\| \leq 2d + 1 + 2n.$$

Letting $n = d/2$ in the above theorem proves Theorem 14.3. Note that the dependency on d is necessary in the above theorem. In particular, even if $n = 1$ it may be that every vector in the support of the random vector r_1 as norm at least d.

First Idea: The first idea to prove the above theorem is to show that the expected norm is small, $\mathbb{E} \| \sum_{i=1}^{n} r_i r_i^\intercal \| \leq O(d+n)$. But, unfortunately, it may be that only with an exponentially small probability this norm is small. Let us give an example.

Example 14.5. *Suppose for all $1 \leq i \leq n$, $r_i = 0$ with probability $\frac{c-1}{c}$ and for all $1 \leq j \leq d$, $r_i = \sqrt{c \cdot d} \mathbf{1}_j$ with probability $\frac{1}{c \cdot d}$ for some $c \gg 1$. It follows that $\mathbb{E} r_i r_i^\intercal = I$ for all i. Observe that any sample of this distribution has norm $c \cdot d$ except the one where $r_i = 0$ for all $1 \leq i \leq n$. So, for $c \geq 10$ and $n < d/2$ there is only one solution to Theorem 14.4, or only one good point in the support of our independent distribution and that point has an exponentially small probability $(1 - 1/c)^n$.*

The above example shows the delicacy of the proof of Theorem 14.4. We can also use the above example to point out the importance of the word *independence* in the statement of Theorem 14.4. Suppose that in the above example the point where $r_i = 0$ for all $1 \leq i \leq n$ was not in the support of our distribution. Then, obviously any point in the support of the distribution would have norm $2d$. Note that such a distribution is indeed not very far from an independent distribution. As we will see in the future lectures instead of a product distribution, i.e., an independence assumption, we may as well work with a strongly Rayleigh distribution and obtain similar quality statements.

Second Idea: The second idea that comes to mind is to show that

$$\left\| \mathbb{E} \sum_{i=1}^{n} r_i r_i^{\intercal} \right\| \leq O(n),$$

and then use the independence assumption to show that there exists a point in the support of this distribution of small spectral norm. First note that $\mathbb{E} \sum_{i=1}^{n} r_i r_i^{\intercal} = nI$ so the above obviously holds. But, if we only use this assumption (as opposed to $\mathbb{E} r_i r_i^{\intercal} = I$ for all i) it may be that $r_2 = r_3 = \cdots = r_n = 0$ with probability 1 and $r_1 = \sqrt{nd}\mathbf{1}_j$ with probability $1/d$. In that case the best bound that we can get is $O(d \cdot n)$.

Third Idea: The idea of Marcus, Spielman and Srivastava is that instead of averaging the matrices of the samples of our independent distribution we should average out their characteristic polynomial. This can help us to use algebraic techniques and in particular the properties of real stable polynomials to prove our goal. In other words, it is to only consider the eigenvalues when we are averaging out different samples and just drop the eigenvectors. They defined the expected characteristic polynomial as follows:

$$\mathbb{E}\chi[r_1 r_1^{\intercal} + \cdots + r_n r_n^{\intercal}](t).$$

In other words, for any point t, the above polynomial is simply the average of the characteristic polynomial $\det(tI - r_1 r_1^{\intercal} - \cdots - r_n r_n^{\intercal})$ over all points of the independent distribution.

We emphasize that the expected characteristic polynomial is not equal to the characteristic polynomial of the expected matrix. Note that the characteristic polynomial of the expected matrix is simply $(t-n)^d$ but as we see in the first lemma the expected characteristic polynomial is $(1-D)^n t^d$.

Lemma 14.6. $\mathbb{E}\chi\left[r_1 r_1^{\intercal} + \cdots + r_n r_n^{\intercal}\right](t) = (1-D)^n t^d$

The above lemma gives a lot of information about the expected characteristic polynomial. Perhaps most importantly, the expected characteristic polynomial is a real rooted polynomial. This simply follows from the closure properties of the real stable polynomials under the operators $1 - D$ (take a look at Lecture 10 for more details).

A natural idea that comes to mind is that analogous to idea 2, first show that the largest root of the expected characteristic polynomial is small and the show that there is a point in the support of the distribution such that the largest root of its characteristic polynomial is smaller than the largest root of the expected polynomial. Note that in Idea 2 the proof of the first step was easy but the second step was not necessarily true. It turns out that in this case we can use the fact that we have an independent distribution and interlacing properties to prove the second step. Finally, we will use a variant of the barrier argument that we discussed in the last lecture to prove the first step.

Lemma 14.7. *With a positive (possibly exponentially small) probability*

$$\lambda_{\max}(\chi[r_1 r_1^{\intercal} + \cdots + r_n r_n^{\intercal}](t)) \leq \lambda_{\max}(\mathbb{E}\chi[r_1 r_1^{\intercal} + \cdots + r_n r_n^{\intercal}](t)).$$

In fact the proof of the above lemma is algorithmic.

It remains to bound the largest root of the expected characteristic polynomial or equivalently, the largest root of $(1-D)^n t^d$. The polynomial $(1-D)^n t^d$ is called the mixed characteristic polynomial in [MSS14] and we adopt similar terminology.

Perhaps, the easiest way to bound the maximum root of the mixed characteristic polynomial is to note that this is a constant multiple of a well known polynomial known as *associated Laguerre polynomial* and the roots of these polynomials are very well studied [Kra06]. Here, we use a variant of the barrier argument to analyze the roots, the main reason is that we will use a similar idea when we extend this argument to multivariate polynomials in order to prove Theorem 14.1.

Lemma 14.8. *The largest of the polynomial* $(1-D)^n t^d$ *is at most* $2d + 2n$*.*

Roughly speaking, we use the same barrier argument and we show that any time that we apply a $1-D$ operator to any real rooted polynomial this increases the upper barrier no more than $1/(1-\alpha)$ where α is the cumulative distance of the roots to the upper barrier as it is also used in the last lecture.

14.1 The Mixed Characteristic Polynomial

In this part we prove Lemma 14.6.

The main intuition in the proof of Lemma 14.6 is that for all $1 \le i \le n$, the expected characteristic polynomial is a linear function of the actual realization of r_i, therefore, as long as we have an independent distribution of these random vectors the LHS is only a function of $\mathbb{E} r_1 r_1^\intercal, \ldots, \mathbb{E} r_n r_n^\intercal$ and not the actual realization of the vectors.

We give two proofs of this lemma, the first one is simpler and in a sense less illuminative, it uses the determinant rank 1 update formula. Fix a (symmetric) matrix A, for a random vector r satisfying $\mathbb{E} rr^\intercal = I$ we can write

$$\begin{aligned} \mathbb{E}\chi[tI - A - rr^\intercal] &= \mathbb{E}\det(tI - A)(1 - r^\intercal (tI-A)^{-1} r) \\ &= \det(tI-A)(1 - \mathrm{Tr}(\mathbb{E}[rr^\intercal](tI-A)^{-1})) \\ &= \det(tI-A)(1-\mathrm{Tr}((tI-A)^{-1})). \end{aligned} \tag{14.1}$$

Say $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A, we can write the RHS as

$$\begin{aligned} \prod_{i=1}^n (t-\lambda_i)\left(1 - \sum_{i=1}^n \frac{1}{t-\lambda_i}\right) &= \chi[A](t) - \sum_{i=1}^n \prod_{j\neq i}(t-\lambda_i) \\ &= (1-D)\chi[A](t). \end{aligned}$$

Now, the lemma follows by a simple induction on n and noting that $(1-D)$ and $\mathbb{E}$ operators are commuting. In particular,

$$\begin{aligned} \mathbb{E}\det(tI - r_1 r_1^\intercal - \cdots - r_n r_n^\intercal) &= \mathbb{E}[\mathbb{E}[\det(tI - r_1 r_1^\intercal - \cdots - r_n r_n^\intercal)|r_1, \ldots, r_{n-1}]] \\ &= \mathbb{E}(1-D)\det(tI - r_1 r_1^\intercal - \cdots - r_{n-1} r_{n-1}^\intercal) \\ &= (1-D)\mathbb{E}\det(tI - r_1 r_1^\intercal - \cdots - r_{n-1} r_{n-1}^\intercal). \end{aligned}$$

Now, we can induct. This completes the first proof.

The second proof is a more direct approach, we write down both sides of the identity of Lemma 14.6 explicitly and we show that the coefficient of t^i is the same in both sides of the equality. Let us start with the RHS it is easy to see that

$$\begin{aligned}(1-D)^n t^d &= \sum_{k=0}^{n} \binom{n}{k} (-D)^k x^d \\ &= \sum_{k=0}^{n} t^{d-k} (-1)^k \frac{d!}{(d-k)!} \binom{n}{k}. \end{aligned} \tag{14.2}$$

Now, we want to show that the coefficient of t^{d-k} in the LHS is the same. First, we recall the reformulation of the characteristic polynomial that we discussed in Lecture 3.

Fact 14.9. *For rank 1 symmetric matrices $W_1, \ldots, W_n \in \mathbb{R}^{d\times d}$,*

$$\det(tI + \sum_{i=1}^{n} z_i W_i) = \sum_{k=0}^{n} t^{d-k} (-1)^k \sum_{S \in \binom{n}{k}} z^S \det_k \left(\sum_{i\in S} W_i \right),$$

where $z^S = \prod_{i\in S} z_i$.

It follows that

$$\mathbb{E}\chi[r_1 r_1^\intercal + \cdots + r_n r_n^\intercal](t) = \sum_{k=0}^{n} t^{d-k} \cdot \mathbb{E} \left[(-1)^k \sum_{S\in\binom{n}{k}} \det_k \left(\sum_{i\in S} r_i r_i^\intercal \right) \right]$$

Note that as we defined earlier, when we take the expected characterisitic polynomial it means that we take the average of the coefficients of t^i for all $0 \le i \le d$.

Comparing the above with (14.2) and using the linearity of expectation, all we need to show is that for any $S \in \binom{n}{k}$,

$$\mathbb{E} \det_k \left(\sum_{i\in S} r_i r_i^\intercal \right) = \frac{d!}{(d-k)!}.$$

The above equality essentially follows from the fact that $\det_k$ is a linear function of $r_i r_i^\intercal$. To see this geometrically, recall that $\det_k(\sum_{i\in S} r_i r_i^\intercal)$ is the square of the volume of the parallelepiped of $\{r_i\}_{i\in S}$. The volume square changes quadratically with r_i and linearly with $r_i r_i^\intercal$. Therefore, by linearity of expectation, the LHS of the above only depends on $\mathbb{E} r_i r_i^\intercal$ for $i \in S$.

We need to sum up the expected determinant of all $k \times k$ minors of $\sum_{i\in S} r_i r_i^\intercal$. There are $\binom{d}{k}$ possibility it is easy to see that the expected determinant of each one of them is exactly $k!$. We leave the proof of this as an exercise.

14.2 Interlacing

Next, we prove Lemma 14.7. Before proving this we prove the Cauchy's interlacing theorem. It says that for any (symmetric) matrix A and any vector v, the roots of $\det(tI - A)$ interlaces the roots of $\det(tI - A - vv^\intercal)$.

Lemma 14.10 (Cauchy's interlacing theorem)**.** *For any (symmetric) matrix A and any vector v,* $\det(tI - A)$ *interlaces* $\det(tI - A - vv^\intercal)$.

Proof. Recall that for two monic (univariate) polynomials $p(t), q(t)$ if the Wronskian $W[p,q] = p'q - pq' \leq 0$ over $\mathbb{R}$ (and the roots of p, q have multiplicity 1) then p interlaces q. See Lecture 10 for applications and generalizations of the Wronskian. Here, we show that

$$W[\det(tI - A - vv^\intercal), \det(tI - A)] \geq 0,$$

and we leave it as an exercise to verify the case where eigenvalues of A may have multiplicity more than 1.

Equivalently, we show that

$$\frac{\mathrm{d}}{\mathrm{d}t} \frac{\det(tI - A - vv^\intercal)}{\det(tI - A)} \geq 0 \tag{14.3}$$

over the entire $\mathbb{R}$. This is because for any p, q the sign of $p'q - pq'$ is the same as the sign of $(p/q)' = (p'q - pq')/q^2$.

Now, by (14.1), we can write

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} \frac{\det(tI - A - vv^\intercal)}{\det(tI - A)} &= \frac{\mathrm{d}}{\mathrm{d}t} \frac{\det(tI - A)(1 - v^\intercal (tI - A)^{-1} v)}{\det(tI - A)} \\ &= \frac{\mathrm{d}}{\mathrm{d}t} (1 - v^\intercal (tI - A)^{-1} v) \end{aligned}$$

Let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of A with corresponding (orthonormal) eigenvectors $u_1, \dots, u_n$. We can rewrite the above as follows

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} (1 - v^\intercal (tI - A)^{-1} v) &= \frac{\mathrm{d}}{\mathrm{d}t} - \sum_{i=1}^{n} \frac{\langle v, u_i \rangle^2}{t - \lambda_i} \\ &= \sum_{i=1}^{n} \frac{\langle v, u_i \rangle^2}{(t - \lambda_i)^2} \geq 0, \end{aligned}$$

as desired. □

Having this in hand, we can round $r_1, \dots, r_n$ inductively.

Lemma 14.11. *Fix a (symmetric) matrix A and let $r_1, \dots, r_n$ are independent random vectors such that $\mathbb{E} r_i r_i^\intercal = I$ for all i. There is a point in the support of r_1 such that*

$$\lambda_{\max}(\mathbb{E}_{r_2, \dots, r_n} \det(tI - A - r_1 r_1^\intercal - \dots - r_n r_n^\intercal)) \leq \lambda_{\max}(\mathbb{E} \det(tI - A - r_1 r_1^\intercal - \dots - r_n r_n^\intercal)), \tag{14.4}$$

where the expectation in the LHS is only over random vectors $r_2, \dots, r_n$.

It is easy to see that Lemma 14.7 simply follows from the above lemma.

Observe that by Lemma 14.6 the LHS of (14.4) is $(1 - D)^{n-1} \det(tI - A - r_1 r_1^\intercal)$ and the RHS is $(1 - D)^n \det(tI - A)$. We prove the above lemma using the following two claims: First we show that the (real rooted) polynomials $(1 - D)^{n-1} \det(tI - A - r_1 r_1^\intercal)$ have a common interlacer. Then we show that for any set of monic real rooted polynomials $p_1, \dots, p_m$ 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 $p_1 + \dots + p_m$. The lemma follows from the fact that $\mathbb{E}(1 - D)^{n-1} \det(tI - A - r_1 r_1^\intercal) = (1 - D)^n \det(tI - A)$.

Claim 14.12. *For any point in the support of r_1, $(1 - D)^{n-1} \det(tI - A)$ interlaces $(1 - D)^{n-1} \det(tI - A - r_1 r_1^\intercal)$.*

Proof. By Cauchy's interlacing theorem, for any point in the support of r_1, $\det(tI - A)$ interlaces $\det(tI - A - r_1 r_1^\intercal)$. By Obrechkoff, Dedieu theorem that we alluded to in Lecture 10, two real rooted polynomials p, q (of the same degree) are interlacing if and only if $ap + bq$ is real rooted for all $a, b \in \mathbb{R}$. This implies that for all $a, b \in \mathbb{R}$,

$$a \det(tI - A) + b \det(tI - A - r_1 r_1^\intercal)$$

is real rooted. By the closure property of real rooted polynomials,

$$(1 - D)(a \det(tI - A) + b \det(tI - A - r_1 r_1^\intercal)) = a(1 - D) \det(tI - A) + b(1 - D) \det(tI - A - r_1 r_1^\intercal)$$

is real rooted, so for all point in the support of r_1, $(1 - D) \det(tI - A)$ interlaces $(1 - D) \det(tI - A - r_1 r_1^\intercal)$. Similarly, we can show that $(1 - D)^{n-1} \det(tI - A)$ interlaces $(1 - D)^{n-1} \det(tI - A - r_1 r_1^\intercal)$. □

The above claim implies that $(1 - D)^{n-1} \det(tI - A)$ is a common interlacer of the polynomials $(1 - D)^{n-1} \det(tI - A - r_1 r_1^\intercal)$.

Claim 14.13. *Let $p_1, \dots, p_m$ be degree d real rooted univariate (monic) polynomials that have a common interlacer. Then, there is $1 \leq i \leq m$ such that the maximum root of p_i is at most the maximum root of $p_1 + \dots + p_m$*

Proof. Perhaps after renaming assuming that p_1 has the smallest largest root and let λ_1 be the largest root of p_1. Since $p_1, \dots, p_m$ have a common interlacer, each p_i has exactly one root which is at least λ_1. In addition since these are monic, i.e., $p_i(\infty) = \infty$, $p_i(\lambda_1) \leq 0$ for all $1 \leq i \leq m$. Therefore,

$$p_1(\lambda_1) + \dots + p_m(\lambda_1) \leq 0.$$

But, the sum of these polynomials is infinity at $t = \infty$. Therefore, the sum has a root that is at least λ_1. □

This completes the proof of Lemma 14.7.

14.3 Barrier Argument

In this section we prove Lemma 14.8. First, we define our barrier function: For a (univariate) polynomial $p(t)$ we define

$$\Phi^p(t) = \frac{\frac{\mathrm{d}}{\mathrm{d}t} p(t)}{p(t)}.$$

This definition essentially follows from the barrier function that we talked about in the last lecture. In particular, if $p(t) = \det(tI - A)$ for a matrix A with eigenvalues $\lambda_1, \dots, \lambda_n$, then

$$\Phi^p(t) = \frac{\sum_{i=1}^n \prod_{j \neq i} (t - \lambda_j)}{\prod_{i=1}^n (t - \lambda_i)} = \sum_{i=1}^n \frac{1}{t - \lambda_i}.$$

We use the same $\alpha_{\max}$ function that we defined in the last lecture. Let α be the upper barrier. For a polynomial p and a number α we define

$$\alpha_{\max}(p) := \max\{t : \Phi^p(t) = \alpha\}.$$

The following is analogous to the barrier function argument that we had in the last lecture.

Lemma 14.14. *For any univariate real rooted polynomial p and $\alpha > 0$,*

$$\alpha_{\max}((1-D)p) \le \alpha_{\max}(p) + \frac{1}{1-\alpha}.$$

Proof. First, we show that

$$\Phi^{(1-D)p} = \Phi^p - \frac{D\Phi^p}{1-\Phi^p}.$$

This is mainly an algebraic argument.

$$\begin{aligned}
\Phi^{(1-D)p} &= \frac{(p-p')'}{p-p'} \\
&= \frac{(p(1-\Phi^p))'}{p(1-\Phi^p)} \\
&= \frac{p'(1-\Phi^p) + p(-(\Phi^p)')}{p(1-\Phi^p)} \\
&= \Phi^p - \frac{D\Phi^p}{1-\Phi^p}.
\end{aligned}$$

Let $t = \alpha_{\max}(p)$. Therefore, it is enough to show that for $\delta \ge 1/(1-\alpha)$,

$$-\frac{D\Phi^p(t+\delta)}{1-\Phi^p(t+\delta)} \le \Phi^p(t) - \Phi^p(t+\delta). \tag{14.5}$$

Similar to the previous lecture $\Phi^p(t)$ is a convex function of t. Therefore,

$$\Phi^p(t) \ge \Phi^p(t+\delta) - \delta(D\Phi^p(t+\delta)),$$

or

$$\Phi^p(t) - \Phi^p(t+\delta) \ge -\delta(D\Phi^p(t+\delta)).$$

On the other hand, $\Phi^p(t)$ is a monotone decreasing function of t, so the RHS of above is nonnegative. So, by (14.5) it is enough to show

$$\frac{1}{1-\Phi^p(t+\delta)} \le \delta.$$

Again using $\Phi^p(t)$ is monotonically decreasing, $\Phi^p(t+\delta) \le \Phi^p(t) \le \alpha$ as required. □

To prove Lemma 14.8 we let $\alpha = 1/2$. Since t^d has d roots which are all 1, $\alpha_{\max}(t^d) = 2d+1$. By the above lemma each $1-D$ operator increases t by at most $\frac{1}{1-\alpha} = 2$. So,

$$\alpha_{\max}((1-D)^n t^d) \le 2d + 1 + 2n.$$

We note that this is not necessarily the best optimization of the parameters. Lemma 14.8 follows from the fact that $\alpha_{\max}$ upper bounds the maximum root.

References

[Har14] N. Harvey. Personal Communication. 2014 (cit. on p. 14-1).

[HO14] N. J. A. Harvey and N. Olver. "Pipage Rounding, Pessimistic Estimators and Matrix Concentration". In: *SODA*. 2014, pp. 926–945 (cit. on p. 14-1).

[Kra06] I. Krasikov. "On extreme zeros of classical orthogonal polynomials". In: *Journal of computational and applied mathematics* 193.1 (2006), pp. 168–182 (cit. on p. 14-4).

[MSS13] 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 p. 14-1).

[MSS14] A. W. Marcus, D. A. Spielman, and N. Srivastava. "Ramanujan Graphs and the Solution of the Kadison-Singer Problem". 2014. URL: http://arxiv.org/abs/1408.4421 (cit. on pp. 14-2, 14-4).