Math 270: Geometry of Polynomials	**Fall 2015**

Lecture 6: Interlacing polynomials, restricted invertibility

Lecturer: Nick Ryder — *Scribe: Ahmed El Alaoui*

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

In this lecture we introduce the concept of *interlacing* for polynomials. This concept provide a convenient way of reasoning about orderings of the roots of real-rooted polynomials. Specifically, it provides a way of tying the roots of the average polynomial (obtained by averaging the coefficients) to the roots of the individual polynomials. This can for example —in the spirit of the probabilistic method— be used to prove that among a collection of symmetric matrices, there exist one that has eigenvalues lying in a certain range, just by examining the roots of the average characteristic polynomial of the matrices. This technique will be used to provide an alternative proof of the Bourgain-Tzafriri *restricted invertibility* theorem. This material is mostly taken from the survey [MSS14] and the blog post [Sri].

6.1 Interlacing families of polynomials

A motivating question for introducing interlacing is : what convex sets are contained in the set of real-rooted polynomials? i.e. when are convex combinations of real-rooted polynomials real-rooted?

Definition 6.1 (Interlacing)**.** Let f be a monic real-rooted polynomial of degree n and g a monic real-rooted polynomial of degree n or $n-1$, with roots $\alpha_n \le \cdots \le \alpha_1$, $\beta_n \le \cdots \le \beta_1$ respectively (ignoring β_n if g is of degree $n-1$). We say that g interlaces f if

$$\beta_n \le \alpha_n \le \beta_{n-1} \le \cdots \le \beta_1 \le \alpha_1.$$

We adopt the notation $g \longrightarrow f$ to indicate that f has the largest root.

Definition 6.2 (Common interlacing)**.** Let f, g be monic, real-rooted polynomials of degree n with roots $\alpha_n \le \cdots \le \alpha_1$, $\beta_n \le \cdots \le \beta_1$ respectively. We say that f and g have a common interlacing if there exist a polynomial that interlaces both of them. Equivalently, f and g have a common interlacing if there exist a real sequence $a_n \le \cdots \le a_1$ such that

$$\forall 1 \le i \le n-1 \quad \alpha_i,\ \beta_i \in [a_{i+1}, a_i].$$

The next theorem (see e.g. [Ded92]) characterizes real rootedness of convex combinations of real-rooted polynomials.

Theorem 6.3. *Let f and g be two monic, real-rooted polynomials. The following two statements are equivalent:*

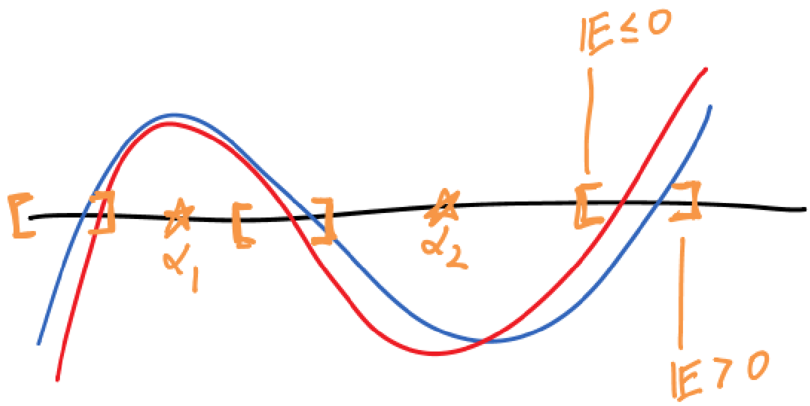

Figure 6.1: Two polynomials with common interlacing and an illustration of Theorem 6.5.

1. *f and g have a common interlacing.*
2. *The polynomial $h_t = tf + (1-t)g$ is real rooted for all $t \in [0, 1]$.*

The proof requires one simple ingredient:

Lemma 6.4. *Let $\epsilon > 0$ and f a real-rooted polynomial of degree n. Then the polynomial*

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

is real-rooted and has simple roots.

Proof. Let $T_\epsilon = I - \epsilon\partial$. If f has a multiple root z (with multiplicity ≥ 2) then $f(z) - \epsilon f'(z) = 0$ since $f(z) = f'(z) = 0$. On the other hand the rational function $f'/f = \sum_i \frac{m_i}{x - z_i}$ (where m_i is the multiplicity of the root z_i) has poles at z_i and is equal to $1/\epsilon$ after each z_i (see Figure 6.2). Hence $T_\epsilon f$ has n real roots, and the multiplicity of each multiple root z_i in $T_\epsilon f$ has dropped by one. Applying the operator T_ϵ n times to f ensures that the resulting n-degree polynomial has all real and simple roots. □

Proof of Theorem 6.3. **(**1. $\Rightarrow$ 2.**)** This will be proven in the next theorem. **(**2. $\Rightarrow$ 1.**)** Assume as a first step that f and g have no common roots and that their roots are simple. Under these assumptions the roots of h_t trace n different intervals I_i on the real line as t varies from 0 to 1, starting from the roots of g and ending at the roots of f. Each one of these intervals contains exactly one root of f and one root of g. Otherwise, (taking g as an example) there would exist a $t \neq 0$ and $z \in \mathbb{R}$ such that $h_t(z) = g(z) = 0$ which would imply that $f(z) = 0$ which in turn contradicts the no-common-roots assumption. Therefore one can choose subintervals $J_i \subset I_i$ with pairwise disjoint interiors containing one root of f and one root of g only, hence establishing interlacing. To prove the general case, notice that the no-common-roots assumption is not problematic since one can always factor the common roots out and put them back at the end. One could easily get an interlacing sequence for f and g from an interlacing sequence of the factored-out polynomials by adding one additional point to the sequence on one or the other side of each common root. As for root multiplicity, we consider a sequence of polynomials f_ϵ and g_ϵ as defined in Lemma 6.4. These have simple roots for any $\epsilon > 0$ and converge to f and g respectively uniformly on any bounded interval as $\epsilon \to 0$. We conclude by a limiting argument using the continuity of the roots as a function of the coefficients. □

This result is very useful when it comes to proving the existence of a common interlacing. One can prove the real-rootedness of the convex combinations instead of coming up with a clever construction of an interlacer. Next, we state the main theorem relating the roots of the averaged polynomial to the roots of the individual polynomials.

Theorem 6.5 (Lamma 4.1 in [MSS13]). *Suppose $f_1, \cdots, f_m$ are monic, real-rooted polynomials of degree n. Let $\lambda_k(f_j)$ denote the kth largest root of f_j and let μ be any probability measure on the set $\{1, \cdots, m\}$. If $f_1, \cdots, f_m$ have a common interlacing, then for all $k \in \{1, \cdots, n\}$,*

$$\min_i \lambda_k(f_i) \le \lambda_k\left(\mathbb{E}_{I\sim\mu}\left[f_I\right]\right) \le \max_i \lambda_k(f_i).$$

Proof. Fix $k \in \{1, \cdots, n\}$. Let $a_n \le \cdots \le a_1$ be the common interlacing sequence for the polynomials $f_1, \cdots, f_m$, i.e. . $a_{k+1} \le \lambda_k(f_i) \le a_k$. The polynomials (f_i) are all monic, so they all have the same sign at a_{k+1} and the same *opposite* sign at a_k. Hence their average changes sign in the interval $[a_{k+1}, a_k]$, so it has to vanish in the same interval by the intermediate value theorem. It is therefore real-rooted by a simple counting argument (thereby also proving the implication $1. \Rightarrow 2.$ in Theorem 6.3). Moreover, it is easy to see that the root is sandwiched between the smallest and the largest root of the f_i's on the interval $[a_{k+1}, a_k]$. □

The connection to the probabilistic method is made readily clear in the above theorem: if one can bound the kth root of the expected polynomial $\mathbb{E}_{I\sim\mu}\left[f_I\right]$, then there exist at least one polynomial f_i the kth root of which obeys the same bound.

6.2 Restricted Invertibility

The above technique will be used to prove that a version of Bourgain and Tzafriri's restricted invertibility theorem that says any matrix of high enough rank contains a well invertible sub-matrix.

Theorem 6.6 (Bourgain-Tzafriri [BT87], Vershynin [Ver01]). *Suppose $v_1, \cdots, v_m \in \mathbb{R}^n$ are vectors with $\sum_{i=1}^m v_i v_i^\intercal = I_n$. Then for every $k < n$ there exist a subset $S \subset \{1, \cdots, m\}$ of size k with*

$$\lambda_k\left(\sum_{i\in S} v_i v_i^\intercal\right) \ge \left(1 - \sqrt{\frac{k}{n}}\right)^2 \frac{n}{m}.$$

The above theorem says that if the vectors v_i are isotropic (i.e. $\mathbb{E}[\langle u, v_i\rangle^2] = 1$ for any unit norm vector u) then there exist a well conditioned subset of the vectors v_i of any size. Before diving into the proof machinery, let us pause for a second to discuss the sharpness of the statement in the above theorem. If the vectors $v_1, \cdots, v_m$ are the columns of a $n \times m$ Gaussian matrix G with centered entries of variance $1/m$ then $E[GG^\top] = I$, and by concentration of measure $\sum_{i=1}^m v_i v_i^\intercal = I(1+o(1))$ with very high probability. Moreover, the matrix $GG^\top$ is a Wishart matrix and its spectrum is very well understood and gives rise to the Marchenko-Pastur distribution in the limit $n/m \to a$ for a fixed. It is in particular known that the spectrum is supported on the interval $[(1-\sqrt{a})^2, (1+\sqrt{a})^2]$ in the limit, and very sharp (sub-Gaussian) concentration bounds hold for finite n and m on both boundaries of this interval. This in particular means that for any submatrix S of G of size $n \times k$, $k/n \sim a < 1$, the matrix $SS^\top = \sum_{i \in S} v_i v_i^\intercal$ has a smallest eigenvalue greater than $(1-\sqrt{a}+o(1))^2 \frac{n}{m}$ with high probability, which means that the above bound is un-improvable in general.

The proof of this result requires essentially two main ideas: a construction of an interlacing family of polynomials related to the matrices $v_i v_i^\intercal$ and a way to analyze and bound the roots of the average polynomial. The relevant polynomials to be used here are characteristic polynomials of symmetric matrices.

Recall that the characteristic polynomial of a square matrix A is the polynomial in x

$$\chi(A)(x) = \det(xI - A).$$

A basic property of the characteristic polynomial is that if A is symmetric, then $\chi(A)$ is real-rooted. Moreover, the characteristic behaves nicely under rank one updates of the matrix A, this is illustrated in Cauchy's interlacing theorem:

Theorem 6.7 (Cauchy's interlacing theorem). *Let A be a symmetric matrix and v be a vector then*

$$\chi(A) \longrightarrow \chi(A + vv^\intercal),$$

i.e. $\chi(A)$ interlaces $\chi(A + vv^\intercal)$.

Proof. We use here a useful rank one update formula for the determinant:

$$\begin{aligned} \chi(A + vv^\intercal)(x) &= \det\left(xI - A - vv^\intercal\right) \\ &= \det\left(xI - A\right)\det\left(I - (xI - A)^{-1} vv^\intercal\right) \\ &= \chi(A)(x)\left(1 - v^\intercal (xI - A)^{-1} v\right). \end{aligned}$$

Letting (λ_j, u_j) be the jth eigenvalue-eigenvector pair of A, we have

$$\chi(A + vv^\intercal)(x) = \chi(A)(x)\left(1 - \sum_{j=1}^n \frac{\langle v, u_j\rangle^2}{x - \lambda_j}\right). \tag{6.1}$$

Therefore, an eigenvalue of $A + vv^\intercal$ is either an eigenvalue of A (if v is orthogonal to the corresponding eigenvector) or such that $\phi(x) := \sum_{j=1}^n \frac{\langle v, u_j\rangle^2}{x - \lambda_j} = 1$. ϕ is a rational function that has poles on the λ_j's and takes the value 1 after each poles and before the next one. This establishes the interlacing property. □

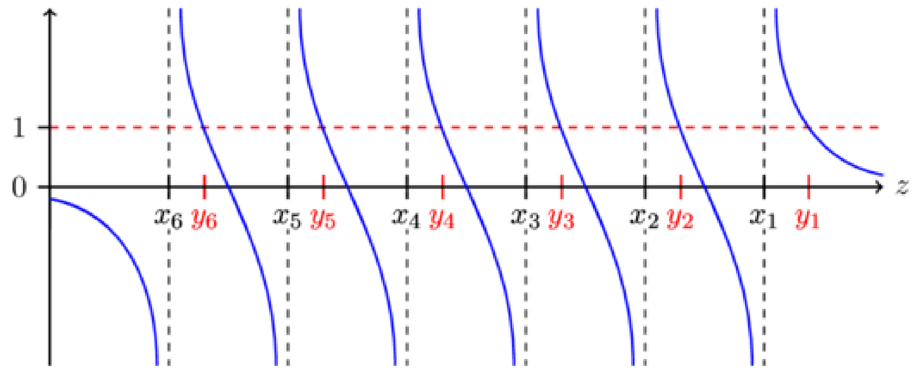

Figure 6.2: Illustration of the behavior of the function ϕ.

Now consider a fixed real symmetric matrix A and vectors $v_1, \cdots, v_m \in \mathbb{R}^n$. Theorem 6.7 tell us that

$$\chi(A + v_1 v_1^\intercal), \cdots, \chi(A + v_m v_m^\intercal)$$

have a common interlacing given by the polynomial $\chi(A)$. Moreover, if we choose the index i to be uniformly random on $\{1, \cdots, m\}$, then

$$\mathbb{E}\left[\chi(A + v_i v_i^\intercal)\right](x) = \chi(A)(x)\left(1 - \sum_{j=1}^{n} \frac{\mathbb{E}\left[\langle v_i, u_j\rangle^2\right]}{x - \lambda_j}\right).$$

When $\sum_{i=1}^m v_i v_i^\intercal = I$, we have $\mathbb{E}\left[\langle v_i, u_j\rangle^2\right] = 1/m$. Hence

$$\mathbb{E}\left[\chi(A + v_i v_i^\intercal)\right](x) = \chi(A)(x) - \frac{1}{m}\chi(A)'(x).$$

This means that performing a rank one update to the matrix A (when the v_i's are isotropic) has the effect in expectation of subtracting off a multiple of the derivative from the characteristic polynomial. The above expression can written equivalently as

$$\mathbb{E}\left[\chi(A + v_i v_i^\intercal)\right] = \left(I - \frac{1}{m}\partial\right)\chi(A). \tag{6.2}$$

Hence the expected characteristic polynomial of the (random) matrices $A + v_i v_i^\intercal$ is the image of $\chi(A)$ by the differential operator $I - \frac{1}{m}\partial$. These operators are going to play a crucial role in the proof of the restricted invertibility result. In particular, a central property that will allow us to apply the above results inductively on many random variables is that these operator preserve the interlacing property in the following sense:

Lemma 6.8. *Let $f_1, \cdots, f_m$ be monic real-rooted polynomials that have a common interlacing. Then the polynomials $(I - \alpha\partial)f_1, \cdots, (I - \alpha\partial)f_m$ are also real-rooted and have a common interlacing for any $\alpha \geq 0$.*

Proof. First, if f is monic and real-rooted then $f = \chi(A)$ where A is a diagonal matrix containing the roots of f on its diagonal. Letting $v = \sqrt{\alpha}\ (1, \cdots, 1)$, we have $(I - \alpha\partial)f = \chi(A + vv^\intercal)$ by equation (6.1). And by the same argument that lead us to prove Cauchy's interlacing theorem, we conclude that $(I - \alpha\partial)f$ is real rooted as well. Next, by Theorem 6.3 all convex combinations $f_1, \cdots, f_m$ are real-rooted, hence so are all convex combination of $(I - \alpha\partial)f_1, \cdots, (I - \alpha\partial)f_m$, and we conclude using Theorem 6.3 again. □

Now we are in position of proving the restricted invertibility theorem.

Proof of Theorem 6.6. To prove the theorem, we need to find a subset of the vectors $\{v_1, \cdots, v_m\}$ of size k with large smallest singular value. We will prove existence of these vectors one by one in an inductive way. To proceed, for a partial assignment $i_1, \cdots, i_l \in \{1, \cdots, m\}$ for $l \leq k$ and random vectors $X_j \in \mathbb{R}^n$ uniformly distributed on the set $v_1, \cdots, v_m$, we define the conditional expected characteristic polynomial

$$q_{i_1,\cdots,i_l} = \mathbb{E}\left[\chi\left(\sum_{j=1}^{l} v_{i_j}v_{i_j}^\intercal + \sum_{j=l+1}^{k} X_j X_j^\intercal\right)\right]$$

where the expectation is taken with respect to the vectors X_i. Note in particular that $q_{i_1,\cdots,i_k} = \chi\left(\sum_{j=1}^{k} v_{i_j}v_{i_j}^\intercal\right)$. We use the convention $q_\emptyset$ for $l = 0$. Observe that if we apply the rank one update formula (6.2) $k - l$ times we have

$$q_{i_1,\cdots,i_l} = \left(I - \frac{1}{m}\partial\right)^{k-l} \chi\left(\sum_{j=1}^{l} v_{i_j}v_{i_j}^\intercal\right).$$

Fix a partial assignment $i_1, \cdots, i_l$ for $l \leq k-1$ and let $A = \sum_{j=1}^{l} v_{i_j}v_{i_j}^\intercal$. By Cauchy's interlacing interlacing theorem (Theorem 6.7), the family of polynomials $\chi\left(A + v_{i_{l+1}}v_{i_{l+1}}^\intercal\right)$ for $i_{l+1} \in \{1, \cdots, m\}$ have a common interlacing, namely $\chi(A)$. Then, since applying the operator $I - \frac{1}{m}\partial$ preserves interlacing (Lemma 6.8), the polynomials

$$\left(I - \frac{1}{m}\partial\right)^{k-(l+1)} \chi(A + v_{i_{l+1}}v_{i_{l+1}}^\intercal) = q_{i_1,\cdots,i_{l+1}}$$

have a common interlacing. Therefore, by Theorem 6.5 there exist a particular $i_{l+1} \in \{1, \cdots, m\}$ such that

$$\lambda_k\left(q_{i_1,\cdots,i_{l+1}}\right) \geq \lambda_k(q_{i_1,\cdots,i_l}). \tag{6.3}$$

Now by induction on l, there exist a complete assignment $i_1, \cdots, i_k$ such that

$$\lambda_k(q_{i_1,\cdots,i_k}) \geq \lambda_k(q_\emptyset).$$

Remark: notice that since $\mathbb{E}_{i_{l+1}}[q_{i_1,\cdots,i_{l+1}}] = q_{i_1,\cdots,i_l}$, inequality (6.3) is obtained by essentially "swapping" the operators $\mathbb{E}_{i_{l+1}}$ and λ_k and appealing to the probabilistic method (i.e. if $\mathbb{E}[X] > c$

then there exist a realization of X that is greater than c). Usually, one can perform this swapping on simple functions by relying on convexity properties of the function of interest (e.g. Jensen's inequality), but the map $p \to \lambda_k(p)$ is highly non linear and non convex. The crucial property that allows this swapping operation is interlacing, which by Theorem 6.3 transfers "enough" convexity to the map λ_k for this to be possible.

Now all is left is to study the roots of the polynomial

$$q_\emptyset = \mathbb{E}\left[\chi\left(\sum_{j=1}^{k} X_j X_j^\intercal\right)\right] = \left(I - \frac{1}{m}\partial\right)^k x^n.$$

It turns out that the rescaled polynomial

$$p_k(x) = \mathbb{E}\left[\chi\left(m\sum_{j=1}^{k} X_j X_j^\intercal\right)\right] = (I - \partial)^k x^n$$

can be written as $p_k(x) = x^{n-k} L_k^{(n-k)}(x)$ where $L_k^{(n-k)}$ is the associated Laguerre polynomial of degree k. It is in particular known [Kra06] that the roots of $L_k^{(n-k)}$ lie in the interval

$$\left[n\left(1 - \sqrt{\frac{k}{n}}\right)^2, n\left(1 + \sqrt{\frac{k}{n}}\right)^2\right],$$

and this completes the proof.

□

References

[BT87] J Bourgain and L Tzafriri. Invertibility of large submatrices with applications to the geometry of banach spaces and harmonic analysis. *Israel journal of mathematics*, 57(2):137–224, 1987.

[Ded92] Jean Pierre Dedieu. Obreschkoff's theorem revisited: what convex sets are contained in the set of hyperbolic polynomials? *Journal of pure and applied algebra*, 81(3):269–278, 1992.

[Kra06] Ilia Krasikov. On extreme zeros of classical orthogonal polynomials. *Journal of computational and applied mathematics*, 193(1):168–182, 2006.

[MSS13] Adam W Marcus, Daniel Spielman, and Nikhil Srivastava. Interlacing families i: Bipartite ramanujan graphs of all degrees. In *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*, pages 529–537. IEEE, 2013.

[MSS14] Adam W Marcus, Daniel A Spielman, and Nikhil Srivastava. Ramanujan graphs and the solution of the kadison-singer problem. *arXiv preprint arXiv:1408.4421*, 2014.

[Sri] Nikhil Srivastava. Restricted invertibility by interlacing polynomials. `http://windowsontheory.org/2014/04/15/restricted-invertiblity-by-interlacing-polynomials`.

[Ver01] Roman Vershynin. John's decompositions: selecting a large part. *Israel Journal of Mathematics*, 122(1):253–277, 2001.