CS 798 - Algorithmic Spectral Graph Theory, Fall 2015, Waterloo

Lecture 19: Interlacing polynomials

We go through some background on interlacing and real-stable polynomials for the next few lectures, which will use these to develop a new probabilistic method to solve several important problems.

Spectral sparsification and Kadison-Singer problem

It was observed that the linear-size spectral sparsification result is similar to the following conjecture by Weaver, which is known to be equivalent to the Kadison-Singer problem, whose positive resolution would have implications in several areas of mathematics.

Weaver's conjecture There exist positive constants α and ε so that for every m and n and every set of vectors $v_1, \ldots, v_m \in \mathbb{R}^n$ such that $\|v_i\| \leq \alpha$ for all i and $\sum_i v_i v_i^T = I$, there exists a partition of $\{1, \ldots, m\}$ into two sets S_1 and S_2 so that $\|\sum_{i \in S_j} v_i v_i^T\| < 1 - \varepsilon$ for $j \in \{1,2\}$.

Note that since $\sum_{i \in S_1} v_i v_i^T + \sum_{i \in S_2} v_i v_i^T = I$, the conclusion $\|\sum_{i \in S_1} v_i v_i^T\| < 1-\varepsilon$ is equivalent to

$$\varepsilon I \preccurlyeq \sum_{i \in S_1} v_i v_i^T \preccurlyeq (1-\varepsilon) I$$

, and so the vectors in S_1 is a spectral approximation of I.

In the BSS theorem in L18, the task was to find scalars with few non-zeros so that

$$(1-\varepsilon) I \preccurlyeq \sum_i w_i v_i v_i^T \preccurlyeq (1+\varepsilon) I, \text{ or equivalently } \frac{(1-\varepsilon)}{2} I \preccurlyeq \sum_i w_i' v_i v_i^T \preccurlyeq \frac{(1+\varepsilon)}{2} I.$$

If all the w_i' are either zero or one, then it would have given a positive solution to Weaver's.

This is not always possible, however, since if there is a long vector v_i (say $\|v_i\| > 1-\varepsilon$), then setting w_i to be zero or one would violate the minimum eigenvalue or maximum eigenvalue bound.

This is why there is an additional condition $\|v_i\| \leq \alpha$ in Weaver's conjecture, and with this we want to set the scalars to be zero or one (but not arbitrary real values), so that we get the stronger conclusion that the vectors can be partitioned into two groups.

An analogy is like the first graph sparsification result by Karger, where there is an additional condition that the min-cut size is large, and in return we can apply uniform sampling so that all non-zero weights equal.

The Weaver's conjecture in the spectral sparsification setting is asking if the maximum effective resistance of an edge is at most α, then there is a partitioning of the edges into two groups so that the subgraph formed by each group is a (somewhat) good spectral approximation of the original graph.

Some examples of graphs with maximum effective resistance small are expander graphs or edge-transitive graphs (since effective resistance will be the same for each edge, like hypercubes, Cayley graphs).

One can apply matrix Chernoff bounds to this problem, and get a solution with norm $= O(\sqrt{\log n})$ with high probability, but this is not good enough for our purpose.

The approach by BSS heavily depends on a careful choice of the scalar and also seems not applicable.

New probabilistic method

Marcus, Spielman and Srivastava developed a completely new approach to work with the maximum eigenvalue.

Perhaps surprisingly, they look at the characteristic polynomials, and use the algebraic and analytical properties of these polynomials to reason about their maximum root.

Recall that the characteristic polynomial of a matrix A is defined as $\det(\lambda I - A)$. If A is the adjacency matrix of the graph, then it is a degree n polynomial where the roots of the polynomial are the eigenvalues of A. To bound the maximum eigenvalue, it is equivalent to bounding the max root.

In standard probabilistic method, we compute the expectation of a random variable $E[X]$, and then conclude that there is a choice with value at most $E[X]$ or at least $E[X]$.

Let A be a sum of independent random Hermitian matrices, i.e. $A = \sum A_i$ where A_i are random.

MSS approach is to compare $\text{max-root}(\det(\lambda I - A))$ and $\text{max-root}\left(\underset{A_i}{E}\det(\lambda I - A)\right)$.

(Notice that $\text{max-root}\left(E\det(\lambda I - A)\right)$ is not the same as $\text{max-root}\left(\det(\lambda I - E[A])\right)$.)

In general, as we will discuss soon, it is not true that there exists some choices of A_i such that $\text{max-root}(\det(\lambda I - A)) \le \text{max-root}\left(E\det(\lambda I - A)\right)$.

The proofs using this approach consist of two main steps:

- Show that $\text{max-root}(\det(\lambda I - A)) \le \text{max-root}\left(E\det(\lambda I - A)\right)$ with positive probability.

  In particular, this is true if A_i are random Hermitian rank one matrices, and so this method is applicable to the setting in Weaver's conjecture.

  It turns out this step involves some nice mathematics about interlacing polynomials and real-stable polynomials, and there is a rich literature about these polynomials.

- Bound the maximum root of these expected characteristic polynomial.

In the case of constructing Ramanujan graphs, there are existing results in the literature.

In proving the Weaver's conjecture, the barrier argument in L18 is generalized to do this.

Today, we will go through the background on interlacing polynomials and real-stable polynomials.

Next week, we will show interesting applications of this method in constructing Ramanujan graphs and proving Weaver's conjecture.

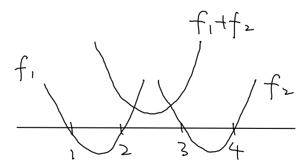
Interlacing polynomials

Recall that we would like to study under what conditions $\text{max-root}(\det(\lambda I - A)) \le \text{max-root}(\mathbb{E}_A \det(\lambda I - A))$.

Let's consider the more general question when $\min_i \text{max-root}(f_i) \le \text{max-root}(\sum_i f_i)$ when f_i are polynomials.

In general, it is usually not true.

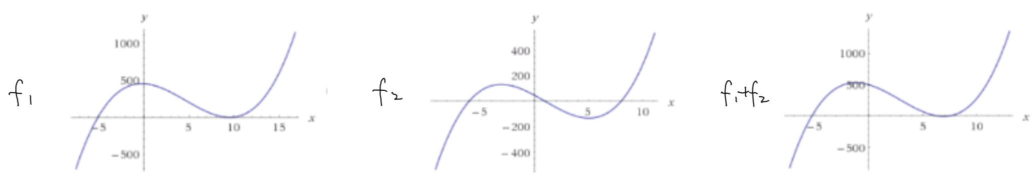
Consider $f_1 = (x-1)(x-2)$ and $f_2 = (x-3)(x-4)$.



The polynomial f_1+f_2 is not even real-rooted.

Even if f_1+f_2 is real-rooted, the relation may not hold.

For example, consider $f_1 = (x+5)(x-9)(x-10)$ and $f_2 = (x+6)(x-1)(x-8)$, f_1+f_2 has roots $\approx -5.3, 6.4, 7.4$.



pictures from Wolfram alpha

There are, however, some additional properties in the polynomials in the Weaver's setting.

Definition (Interlacing) Let f be a degree n polynomial with real roots $\alpha_1 \ge \alpha_2 \ge \dots \ge \alpha_n$ and let g be a degree n polynomial with real roots $\beta_1 \ge \beta_2 \ge \dots \ge \beta_n$ (or g a degree $n-1$ polynomial).

We say that g interlaces f if $\alpha_1 \ge \beta_1 \ge \alpha_2 \ge \beta_2 \ge \dots \ge \beta_{n-1} \ge \alpha_n \ge \beta_n$.

Characteristic polynomials after rank-one update

In Weaver's setting, we consider rank-one update $A + vv^T$ of a matrix A, and it is known as the Cauchy's interlacing theorem that $\det(\lambda I - A)$ interlaces $\det(\lambda I - A - vv^T)$.

To see this, we need the matrix determinantal formula that will be useful many times.

Lemma (matrix determinantal formula) For non-singular M, $\det(M + vv^T) = \det(M) \cdot (1 + v^T M^{-1} v)$

proof $\det(M + vv^T) = \det(M(I + M^{-1} vv^T)) = \det(M) \det(I + M^{-1} vv^T)$ since $\det(AB) = \det(A)\det(B)$.

Recall that determinant of a matrix is equal to the product of its eigenvalues.

For the matrix $I + M^{-1}vv^T$, the eigenvalues are the eigenvalues of $M^{-1}vv^T$ plus one.

Since $M^{-1}vv^T$ is a rank one matrix, its only non-zero eigenvalue is $Tr(M^{-1}vv^T) = v^T M^{-1} v$.

Therefore, the spectrum of $I + M^{-1}vv^T$ is $(1+v^T M^{-1} v, 1, 1, \ldots, 1)$, and thus $\det(I + M^{-1}vv^T) = (1 + v^T M^{-1} v)$. $\square$

Back to Cauchy's interlacing theorem, $\det(\lambda I - A - vv^T) = \det(\lambda I - A)\left(1 - v^T(\lambda I - A)^{-1} v\right)$

(where λ_j are eigenvalues of A and u_j eigenvectors) $\rightarrow$
$$= \det(\lambda I - A)\left(1 - v^T\left(\sum_{j=1}^n (\lambda - \lambda_j) u_j u_j^T\right)^{-1} v\right)$$
$$= \det(\lambda I - A)\left(1 - v^T\left(\sum_{j=1}^n \frac{u_j u_j^T}{\lambda - \lambda_j}\right) v\right)$$
$$= \det(\lambda I - A)\left(1 - \sum_{j=1}^n \frac{\langle u_j, v\rangle^2}{\lambda - \lambda_j}\right).$$

Without loss, assume the eigenvalues of A are distinct, $\lambda_1 > \lambda_2 > \lambda_3 > \ldots > \lambda_n$.

Between two eigenvalues λ_j and λ_{j+1}, $\det(\lambda I - A)$ is either positive or negative as there is no root in this range, say $\det(\lambda I - A)$ is entirely positive in $(\lambda_j, \lambda_{j+1})$.

On the other hand, the function $1 - \sum_{j=1}^n \frac{\langle u_j, v\rangle^2}{\lambda - \lambda_j}$ tends to ∞ when λ approaches λ_j from below, and tends to $-\infty$ when λ approaches λ_{j+1} from above, and so there is a root of $\det(\lambda I - A - vv^T)$ in $(\lambda_j, \lambda_{j+1})$.

Applying this argument for each interval, we get $\beta_1 > \lambda_1 > \beta_2 > \lambda_2 > \ldots > \beta_n > \lambda_n$ where β_i are the eigenvalues of $A + vv^T$ and so $\det(\lambda I - A)$ interlaces $\det(\lambda I - A - vv^T)$.

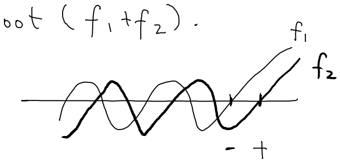
The above argument works when $\langle v, u_j\rangle \neq 0$ for all j. This can be made precise either by a pertubation argument, or special case handling of the shared roots.

Common interlacing and comparing max roots

Suppose two real-rooted polynomials with positive leading coefficients are interlacing.

Then it is easy to see that $\min_{i=\{1,2\}} \text{max-root}(f_i) \leq \text{max-root}(f_1 + f_2)$.

Say $\text{max-root}(f_1) \leq \text{max-root}(f_2)$.



Since both f_1 and f_2 have positive leading coefficients,

both are positive in the range $(\text{maxroot}(f_2), \infty)$, and $(f_1 + f_2)(\text{maxroot}(f_2)) > 0$.

Since f_1 interlaces f_2, the second largest root of f_2 is smaller than the largest root of f_1, and thus $f_2(\text{maxroot}(f_1))$ must be negative, and hence $(f_1 + f_2)(\text{maxroot}(f_1)) < 0$.

By the intermediate value theorem, there must be a root of $f_1 + f_2$ in the range $(\text{maxroot}(f_1), \text{maxroot}(f_2))$.

and this proves the claim.

This can be generalized to a set of polynomials in a natural way.

Definition (common interlacing) We say a set of polynomials $f_1, \dots, f_m$ have a common interlacing if there is a polynomial g which interlaces each f_i.

Equivalently, $f_1, \dots, f_m$ have a common interlacing if there are disjoint intervals $I_1 \geq I_2 \geq \dots \geq I_n$ so that the k-th largest root of each f_i is contained in I_k, i.e. $\beta_5\ I_5\ \beta_4\ I_4\ \beta_3\ I_3\ \beta_2\ I_2\ \beta_1\ I_1$

The following lemma follows by applying the intermediate value theorem on each I_k.

Lemma Suppose $f_1, \dots, f_m$ have a common interlacing where each has a positive leading coefficient. Let $\lambda_k(f_j)$ be the k-th largest root of f_j, and $\mu_1, \dots, \mu_m$ be non-negative numbers with $\sum_{i=1}^m \mu_i = 1$. Then
$$\min_j \lambda_k(f_j) \leq \lambda_k\Big(\sum_{j=1}^m \mu_j f_j\Big) \leq \max_j \lambda_k(f_j).$$

So, if we could show that a set of polynomials have a common interlacing, then we can apply the new probabilistic method to show that one polynomial has small max-root by showing that the expected polynomial has small max-root.

We will show how to use this approach for the Weaver's conjecture next week.

Today we focus on some general techniques to prove that a set of polynomials has common interlacing.

Common interlacing and real-rootedness

From the Lemma in the previous section, if $f_1, \dots, f_m$ have a common interlacing, then any convex combination of $f_1, \dots, f_m$ is also real-rooted (we assumed f_i is real-rooted).

It turns out that the converse is also true.

We need the following simple fact for the proof.

Fact $f_1, \dots, f_m$ have a common interlacing if and only if f_i, f_j have a common interlacing $\forall i \neq j$.

proof ($\Rightarrow$) is trivial.

($\Leftarrow$) if every pair has a common interlacing, then in any interval $[c, \infty]$, no polynomial can have two more roots than any polynomial, and so there is a common interlacing. □

Also, we use the following theorem from complex analysis without proof.

Theorem The roots of a polynomial are continuous functions of its coefficients.

Lemma Given $f_1, \dots, f_m$, if all convex combinations $\sum_{i=1}^{m} \mu_i f_i$ are real-rooted, then they have a common interlacing.

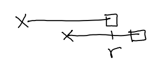
proof By the fact, it is enough to prove this for two polynomials f_1 and f_2.

We assume that f_1 and f_2 have no common roots; if they do, we divide them out by the common roots, prove that the remaining roots have common interlacing, and it is easy to see that putting back the common roots will preserve common interlacing.

Let $f_t = (1-t) f_1 + t f_2$ where $t \in [0,1]$.

By the theorem and our assumption on real-rootedness of f_t, the roots of f_t define n intervals in the real line, starting from a root in f_1 and ending at a root of f_2.

If two intervals overlap at a point r which is a root of f_1,



then $0 = f_t(r) = (1-t) f_1(r) + t f_2(r) = t f_2(r)$, and $f_2(r)$ must be zero if $t \in (0,1)$, contradicting that f_1 and f_2 have no common roots.

So, the intervals must be disjoint except at endpoints, and so f_1, f_2 have a common interlacing. $\square$

By the lemma, to prove a set a polynomials have a common interlacing (to apply the probabilistic method), it is equivalent to proving that all convex combinations are real-rooted.

What are some examples of real-rooted polynomials?

We used one class of examples all the time: the characteristic polynomial of real symmetric matrices

Other than showing that a polynomial is the characteristic polynomial of a real symmetric matrix, it is usually non-trivial to argue that a polynomial is real-rooted without explicitly computing roots.

Real stable polynomials

A general method to prove that a univariate polynomial is real-rooted is to consider a generalization to multivariate polynomial, and use some real stability preserving operations to get back the given univariate polynomial.

Definition (real stable polynomials) A multivariate polynomial $f \in \mathbb{R}[x_1, \dots, x_m]$ is real stable if there are no roots $(y_1, y_2, \dots, y_m)$ with $\mathrm{Im}(y_j) > 0$ for all $1 \le j \le m$.

Note that it is a generalization of real-rootedness for univariate polynomial with real coefficients.

Fact A univariate polynomial $f \in \mathbb{R}[x]$ is real stable if and only if it is real-rooted.

proof Let $f(x) = \sum_{k=0}^{d} c_k x^k$. The proof follows from the observation that complex roots come in pairs.

Suppose $f(a+ib) = \sum_{k=0}^{d} c_k (a+ib)^k = 0$.

Then $0 = \sum_{k=0}^{d} \overline{c_k (a+ib)^k} = \sum_{k=0}^{d} c_k \overline{(a+ib)^k} = \sum_{k=0}^{d} c_k \overline{(a+ib)}^k = \sum_{k=0}^{d} c_k (a-ib)^k = f(a-ib)$.

One of these must have positive imaginary part, contradicting to real stability of f. □

Examples

What are some examples of real-stable polynomials?

All examples today are from determinants.

Lemma If $A_1, \dots, A_m$ are positive semidefinite matrices,

then $f(z_0, z_1, \dots, z_m) := \det(z_0 I + \sum_{i=1}^{m} z_i A_i)$ is real stable.

proof We show that if $\operatorname{Im}(z_i) > 0$ for all $0 \le i \le m$, then the matrix $z_0 I + \sum_{i=1}^{m} z_i A_i$ is of full rank, and hence $\det(z_0 I + \sum_{i=1}^{m} z_i A_i) \neq 0$, implying real stability.

Let $\vec{v} = \vec{c} + i\vec{d}$ where $\vec{c}$ is the real part and $\vec{d}$ is the imaginary part of $\vec{v}$.

Let $Z = z_0 I + \sum_{i=1}^{m} z_i A_i$. Write it as $\operatorname{Re}(Z) + i \operatorname{Im}(Z)$.

When $\operatorname{Im}(z_j) > 0$ for all $0 \le j \le m$, this implies that $\operatorname{Im}(Z) \succ 0$, as $A_i \succcurlyeq 0$ and $I \succ 0$.

We will show that $Z\vec{v} = 0$ only if $\vec{c} = \vec{d} = 0$, and hence Z is of full rank.

To show this, we show that $(\vec{c} - i\vec{d})^T (\operatorname{Re}(Z) + i\operatorname{Im}(Z))(\vec{c} + i\vec{d}) = 0$ only if $\vec{c} = \vec{d} = 0$.

Note that $\operatorname{Im}\left[(\vec{c} - i\vec{d})^T (\operatorname{Re}(Z) + i\operatorname{Im}(Z))(\vec{c} + i\vec{d})\right] = \vec{c}^T \operatorname{Im}(Z)\vec{c} + \vec{d}^T \operatorname{Im}(Z)\vec{d} = 0$ only if $\vec{c} = \vec{d} = 0$, as $\operatorname{Im}(Z) \succ 0$ when all $\operatorname{Im}(z_j) > 0$. This completes the proof. □

Observe that the polynomial in this Lemma is similar to the characteristic polynomial of a sum of matrices A_i, those that appear in Weaver's setting, except it is a multivariate polynomial.

Next week, we will start from this multivariate polynomial to show interlacing properties of those characteristic polynomials in Weaver's setting.

The only missing piece are the following two real-stability preserving operations.

Real stability preserving operations

There are several real-stability preserving operations, and there are some deep results about these.

We just present two operations that we need for the Weaver's conjecture.

Specialization

This will be useful in reducing the number of variables of the polynomial.

Lemma Let $f(z_1, z_2, \ldots, z_m)$ be a non-zero real stable polynomial.

Then, for $t \in \mathbb{R}$, $f(t, z_2, \ldots, z_m)$ is a real stable polynomial.

proof As all coefficients of f are real and t is real, all coefficients of $f(t, z_1, \ldots, z_m)$ are real.

Suppose by contradiction that $f(t, z_2, \ldots, z_m)$ is not real stable.

This means that there exist $y_2, \ldots, y_m$ such that $\mathrm{Im}(y_j) > 0$ for $2 \le j \le m$,

and $f(t, y_2, \ldots, y_m) = 0$.

Consider the polynomial $f(t + i\delta, z_2, \ldots, z_m)$ for some small enough $\delta > 0$ to be chosen.

By the theorem that the roots of polynomials are continuous functions of the coefficients,

there exists a root $y_2', \ldots, y_m'$ such that $|y_j - y_j'| < \varepsilon$ and $f(t + i\delta, y_2', \ldots, y_m') = 0$.

For ε small enough, we still have $\mathrm{Im}(y_j') > 0$ for all $2 \le j \le m$.

Since it is continuous, for any ε, there exists δ small enough such that this happens,

but then it contradicts real stability of f, since all coordinates of this root

$(t + i\delta, y_2', \ldots, y_m')$ have positive imaginary part. $\square$

Differentiation

Lemma For any real t, the polynomial $\left(1 + t\frac{\partial}{\partial z_1}\right) f(z_1, z_2, \ldots, z_m)$ is real stable if f is.

proof We substitute $z_2 = y_2, \ldots, z_m = y_m$ with $\mathrm{Im}(y_j) > 0$ into f.

Since f is stable, the resulting univariate polynomial $g(z)$ is stable (note that

it may not be real since y_j are complex).

If we could prove that $g(z) + t g'(z)$ is also stable (assuming g is), then we prove

that $\left(1 + t\frac{\partial}{\partial z_1}\right) f(z_1, z_2, \ldots, z_m)$ is stable, because if $\left(1 + t\frac{\partial}{\partial z_1}\right) f(z_1, z_2, \ldots, z_m)$ is

not stable, then there exist $y_2, \ldots, y_m$ with $\mathrm{Im}(y_j) > 0$ such that $g(z) + t g'(z)$ not stable.

Since $g(z)$ is stable, it can be written as $c \prod_{j=1}^{n} (z - w_j)$ with $\mathrm{Im}(w_j) \le 0$ for all j.

Then $g(z) + t g'(z) = g(z) \left(1 + \sum_{j=1}^{n} \frac{t}{z - w_j} \right)$.

For z with $\mathrm{Im}(z) > 0$, $g(z) > 0$ as g is stable, and furthermore since $\mathrm{Im}(z) > 0$, we have $\mathrm{Im}\left(\frac{t}{z - w_j} \right) < 0$ for all j, and thus $1 + \sum_{j=1}^{n} \frac{t}{z - w_j} \neq 0$, proving that $g(z) + t g'(z)$ is stable $\square$

Quick review

We record the important results for uses in next lectures.

- $\det\left(zI + \sum_i z_i A_i \right)$ is real stable if $A_i \succeq 0$ $\forall i$.
- Specialization and differentiation are operations that preserve stability.
- real-stable univariate polynomial is real-rooted.
- Common interlacing if and only if all convex combinations are real-rooted.
- Common interlacing allows the new probabilistic method to apply.

References

- Ramanujan graphs and the solution of the Kadison-Singer problem, by Marcus, Spielman, Srivastava.
- Real stable polynomials and the Kadison-Singer problem, blogpost by Terence Tao.