CS 860 Spectral graph theory. Spring 2019, Waterloo.

Lecture 15: Weaver's conjecture

We study the proof of Weaver's conjecture using the method of interlacing polynomials.

A multivariate barrier argument is used to bound the max root of the expected characteristic polynomial.

Overview

Today we see the proof of the probabilistic statement formulated by Marcus, Spielman, Srivastava.

Theorem Let $v_1, v_2, \ldots, v_m \in \mathbb{R}^n$ be independent random vectors with finite support such that

$$E\left[\sum_{i=1}^m v_i v_i^T\right] = I_n \quad \text{and} \quad E\left[\|v_i\|^2\right] \leq \varepsilon \quad \text{for } 1 \leq i \leq m.$$

Then $\Pr\left[\left\| \sum_{i=1}^m v_i v_i^T \right\| \leq (1+\sqrt{\varepsilon})^2\right] > 0$.

We saw in L13 that it implies Weaver's conjecture, which in turn implies a positive resolution of the Kadison-Singer problem.

The proof of the theorem has two main steps.

① First, establish that with positive probability,

$$\left\|\sum_{i=1}^m v_i v_i^T\right\| = \operatorname{maxroot}\left(\det\left(\lambda I - \sum_{i=1}^m v_i v_i^T\right)\right) \leq \operatorname{maxroot}\left(\mathop{E}_{v_1,\ldots,v_m}\left[\det\left(\lambda I - \sum_{i=1}^m v_i v_i^T\right)\right]\right).$$

The inequality is exactly the theorem of the new probabilistic method that we proved in L13, so the first part is done.

② The second part is to prove that $\operatorname{maxroot}\left(\mathop{E}_{v_1,\ldots,v_m}\left[\det\left(\lambda I - \sum_{i=1}^m v_i v_i^T\right)\right]\right) \leq (1+\sqrt{\varepsilon})^2$, given the assumptions that $E[\|v_i\|^2] \leq \varepsilon$ and $E\left[\sum_{i=1}^m v_i v_i^T\right] = I_n$.

In L14 when we construct bipartite Ramanujan graphs, the expected characteristic polynomial turns out to be exactly the matching polynomial and there were results bounding its max root.

For Weaver's conjecture, bounding the maxroot of the expected polynomial is a major technical challenge.

Recall the multilinear formula $\displaystyle E\left[\det\left(\lambda I - \sum_{i=1}^m v_i v_i^T\right)\right] = \prod_{i=1}^m \left(1 - \frac{\partial}{\partial x_i}\right) \det\left(\lambda I + \sum_{i=1}^m x_i E[v_i v_i^T]\right) \Big|_{x_1 = \cdots = x_n = 0}$.

This formula is crucial in the first step, by showing that the LHS is real-rooted to establish common interlacing for the new probabilistic method to hold.

It turns out that the formula is also crucial in the second step.

The idea is to first show an upper bound of the "maxroot" of the multivariate polynomial, and then maintain a good upper bound after each $(1-\frac{\partial}{\partial x_i})$ differential operator is applied.

To establish the upper bound, they use a barrier argument similar to the one in constructing linear-sized spectral sparsifier in L12, but generalize it to the multivariate setting.

Multivariate maxroot argument

To bound $\mathrm{maxroot}\left(E\left[\det(\lambda I - \sum_{i=1}^m v_i v_i^T)\right]\right)$, we bound $\mathrm{maxroot}\left(\prod_{i=1}^m (1-\frac{\partial}{\partial x_i}) \det(\lambda I + \sum_{i=1}^m x_i E[v_i v_i^T])\Big|_{x_1=\cdots=x_n=0}\right)$

First, we use the assumption that $E[\sum_{i=1}^m v_i v_i^T] = I$ to rewrite the polynomial in the RHS as

$$\prod_{i=1}^m (1-\frac{\partial}{\partial x_i}) \det\left(\sum_{i=1}^m (\lambda+x_i) E[v_i v_i^T]\right)\Big|_{x_1=\cdots=x_n=0} = \prod_{i=1}^m (1-\frac{\partial}{\partial x_i}) \det\left(\sum_{i=1}^m x_i E[v_i v_i^T]\right)\Big|_{x_1=x_2=\cdots=x_n=\lambda}.$$

Call the matrix $E[v_i v_i^T] = A_i$. Note that $A_i \succcurlyeq 0$.

Call the polynomial $\prod_{i=1}^k (1-\frac{\partial}{\partial x_i}) \det\left(\sum_{i=1}^m x_i A_i\right) = p_k(x_1, x_2, \ldots, x_m)$ for $0 \le k \le n$, so that

$$p_0(x_1,\ldots,x_n) = \det\left(\sum_{i=1}^m x_i A_i\right) \quad \text{and} \quad p_m(x_1,\ldots,x_m) = \prod_{i=1}^m (1-\frac{\partial}{\partial x_i}) \det\left(\sum_{i=1}^m x_i A_i\right).$$

Our task is to find a small x^* and show that $p_m(x,x,\ldots,x) > 0 \quad \forall\, x > x^*$.

This would imply that $\mathrm{maxroot}\left(\prod_{i=1}^m (1-\frac{\partial}{\partial x_i}) \det\left(\sum_{i=1}^m x_i E[v_i v_i^T]\right)\Big|_{x_1=x_2=\cdots=x_n=\lambda}\right) \le x^*$, and thus by the multilinear formula $\mathrm{maxroot}\left(E\left[\det(\lambda I - \sum_{i=1}^m v_i v_i^T)\right]\right) \le x^*$.

Univariate approach

Initially, since $\sum_{i=1}^m A_i = I_n$, we have $p_0(x,x,\ldots,x) = \det(xI) > 0$ for all $x > 0$.

One natural strategy is to set $x_i = x_{i-1} + \gamma$ for some small γ and prove inductively that

$$p_i(x,\ldots,x) > 0 \quad \forall\, x > x_i.$$

For this to work to give us the final bound $(1+\sqrt{\varepsilon})^2$, we would need γ to be as small as $\frac{1}{m}(1+\sqrt{\varepsilon})^2$, but since each v_i could have $\|v_i\|^2 = \varepsilon$, it is not possible for the induction hypothesis to go through.

Multivariate approach

By doing explicit calculations, it is possible to show that if $p_0(x,x,\ldots,x) > 0$ for all $x > 0$, then $p_1(x+\delta, x, x, \ldots, x) > 0$ for all $x > 0$ for a large enough δ.

This suggests that the correct induction hypothesis should be multivariate, and applying the $1-\frac{\partial}{\partial x_i}$

operation should only change the i-th variable.

Definition Given a multivariate real-stable polynomial $p(x_1,\dots,x_m)$. We say a point $y \in \mathbb{R}^m$ is "above the roots" of p if $p(\vec{y}+\vec{t}) > 0$ for all $\vec{t} \geq 0$.

Our goal is to prove that $(1+\sqrt{\varepsilon})^2 \cdot \vec{1}$ is above the roots of $p_m(x_1,\dots,x_m)$.

To do this, we start with a point $(t,t,\dots,t)$ for some $t>0$ to be chosen later, and we know that $(t,t,\dots,t)$ is above the roots of p_0.

The induction hypothesis is that $(\underbrace{t+\delta, t+\delta, \dots, t+\delta}_{k \text{ coordinates}}, t, \dots, t)$ is above the roots of p_k.

This will imply that $(t+\delta, t+\delta, \dots, t+\delta)$ is above the roots of p_m.

Multivariate barrier functions

To execute the above proof plan, we use a similar approach as in L12, to use barrier functions to establish a "comfortable" upper bound (i.e. a stronger hypothesis) for the induction to go through.

Univariate barrier function

In L12, we set an upper bound u and keep track of a potential function $\Phi_u(A) = \mathrm{Tr}(uI-A)^{-1}$ and maintain the invariant that $\Phi_u(A) \leq 1$, to guarantee that u is well above the roots.

Recall that $\mathrm{Tr}(uI-A)^{-1} = \sum_{i=1}^{n} \frac{1}{u-\lambda_i}$ where $\lambda_1,\dots,\lambda_n$ are the eigenvalues of A.

We can rewrite it in terms of the roots of the characteristic polynomials.

Note that $\det(xI-A) = \prod_{i=1}^{n}(x-\lambda_i)$ and $\partial_x \log\det(xI-A) = \frac{\partial_x \det(xI-A)}{\det(xI-A)} = \sum_{i=1}^{n} \frac{1}{x-u_i}$.

So, we could understand the potential function $\Phi_x(A)$ as $\partial_x \log\det(xI-A)$.

Multivariate barrier functions

The above barrier function is generalized to the multivariate setting.

Given a polynomial $p \in \mathbb{R}[x_1,x_2,\dots,x_m]$ and a point $y \in \mathbb{R}^m$ above the roots of p, the barrier function of p in direction i at y is $\Phi_p^i(y) = \frac{\partial x_i\, p(y)}{p(y)} = \partial x_i \log p(y)$.

Equivalently, we can define Φ_p^i by $\Phi_p^i(y) = \frac{q_{y,i}'(y_i)}{q_{y,i}(y_i)} = \sum_{j=1}^{d} \frac{1}{y_i - \lambda_j}$, where $q_{y,i}(t)$ is the univariate restriction $q_{y,i}(t) = p(y_1,\dots,y_{i-1},t,y_{i+1},\dots,y_m)$ where $\lambda_1,\dots,\lambda_d$ are roots of this univariate polynomial, which is real-rooted as substituting real numbers preserve real stability.

Induction hypothesis

Initially, we find a point x_0 above the roots of p_0 such that $\Phi^i_{p_0}(x_0) \le \phi$ for all $1 \le i \le m$.

Specifically, $x_0 = (t, t, \dots, t)$ for some $t > 0$.

Then, as stated before, let $x_k = (\underbrace{t+\delta, \dots, t+\delta}_{k \text{ coordinates}}, t, \dots, t)$ and $p_k = \prod_{i=1}^{k} (1 - \frac{\partial}{\partial x_i}) \det(\sum_{i=1}^{m} x_i A_i)$,

and the induction hypothesis is to maintain that $\Phi^i_{p_k}(x_k) \le \phi$ for all $1 \le i \le m$.

Informally, we prove inductively that x_k is well above the roots of p_k.

Properties of barrier functions

The following properties of the barrier functions are important in the proof.

Theorem Suppose $p \in \mathbb{R}[x_1, \dots, x_m]$ is real-stable and y is above the roots of p.

Then, for all $i, j \in [m]$ and $\delta \ge 0$, the following properties hold:

(monotonicity) $\Phi^i_p(y + \delta e_j) \le \Phi^i_p(y)$ where e_j is the j-th standard vector.

(convexity) $\Phi^i_p(y + \delta e_j) \le \Phi^i_p(y) + \delta\, \partial_{x_j} \Phi^i_p(y + \delta e_j)$.

The second property can be written as $\frac{\Phi^i_p(y+\delta e_j) - \Phi^i_p(y)}{\delta} \le \partial_{x_j} \Phi^i_p(y + \delta e_j)$, saying that the slope is non-decreasing, equivalent to the function being convex.

Note that the theorem holds for any real-stable polynomial.

We will first use this theorem to bound the max-root, and then present some proof ideas after that.

Inductive proof

As a warm up, we first see that when a point y is well above the roots (i.e. $\Phi^i_p(y) < 1 \ \forall i$), then y is still above the roots after the operation $1 - \partial_{x_i}$.

Claim Suppose that p is real stable and y is above the roots of p, with the additional property that $\Phi^i_p(y) < 1$ for $1 \le i \le m$. Then y is still above the roots of $(1 - \partial_{x_i}) p$ for $1 \le i \le m$.

proof Let z be above y, i.e. $z \ge y$. By monotonicity, $\Phi^i_p(z) \le \Phi^i_p(y) < 1$ for $1 \le i \le m$.

This implies that $1 > \Phi^i_p(z) = \frac{\partial_{x_i} p(z)}{p(z)}$, and thus $0 < p(z) - \partial_{x_i} p(z) = (1 - \partial_{x_i}) p(z)$, and so there is still no roots above y in $(1-\partial x_i)p$. $\square$

The claim shows that y is still above the roots after one differentiation operation, but we could not repeat this argument because the condition $\Phi^j_{(1-\partial x_i)p}(y) < 1$ may no longer hold.

To maintain the invariant, we will increase the upper bound in the corresponding coordinate to decrease the potential function to maintain the invariant.

<u>Lemma</u> Suppose that p is real stable and y is above the roots of p.

Suppose further that $\Phi^i_p(y) \le 1 - \frac{1}{\delta}$ for $1 \le i \le m$ for some $\delta > 0$.

Then, $\Phi^i_{(1-\partial x_j)p}(y+\delta e_j) \le \Phi^i_p(y)$ for $1 \le i, j \le m$.

In particular, $y+\delta e_j$ is still above the roots of $(1-\partial x_i)p$.

<u>Proof</u> Recall that $\Phi^j_p = \partial x_j \log p = \frac{\partial x_j p}{p}$, and this implies that $(1-\partial x_j)p = (1-\Phi^j_p)p$.

Taking log on both sides, $\log(1-\partial x_j)p = \log(1-\Phi^j_p) + \log p$

Applying ∂x_i on both sides, $\Phi^i_{(1-\partial x_j)p} = \Phi^i_p - \frac{\partial x_i \Phi^j_p}{1-\Phi^j_p}$

Therefore, $\Phi^i_{(1-\partial x_j)p}(y+\delta e_j) = \Phi^i_p(y+\delta e_j) - \frac{\partial x_i \Phi^j_p(y+\delta e_j)}{1-\Phi^j_p(y+\delta e_j)}$.

To prove $\Phi^i_{(1-\partial x_j)p}(y+\delta e_j) \le \Phi^i_p(y)$, it is equivalent to $\Phi^i_p(y) - \Phi^i_p(y+\delta e_j) \ge -\frac{\partial x_i \Phi^j_p(y+\delta e_j)}{1-\Phi^j_p(y+\delta e_j)}$.

By convexity, we have that $\Phi^i_p(y) - \Phi^i_p(y+\delta e_j) \ge -\delta \cdot \partial x_j \Phi^i_p(y+\delta e_j)$.

So, the above inequality holds if we could prove that $\delta \cdot \partial x_j \Phi^i_p(y+\delta e_j) \le \frac{\partial x_i \Phi^j_p(y+\delta e_j)}{1-\Phi^j_p(y+\delta e_j)}$.

Note that $\partial x_j \Phi^i_p = \partial x_j \partial x_i \log p = \partial x_i \partial x_j \log p = \partial x_i \Phi^j_p$, and so the numerators cancel out, and the above is equivalent to $\delta \ge 1/(1-\Phi^j_p(y+\delta e_j))$, as $\partial x_j \Phi^i_p(y+\delta e_j) \le 0$ as the function is monotonically decreasing.

Our assumption implies that $\delta \ge 1/(1-\Phi^j_p(y)) \ge 1/(1-\Phi^j_p(y+\delta e_j))$ as desired, where the second inequality is again by monotonicity.

Thus, we always have $\Phi^i_{(1-\partial x_j)p}(y+\delta e_j) \le \Phi^i_p(y)$ for $1 \le i, j \le m$. $\square$

<u>Choosing the parameters</u>

If we choose the initial $x_0 = (t, \dots, t)$ such that $\Phi^i_{p_0}(x_0) \le 1 - \frac{1}{\delta}$ for $1 \le i \le m$ for some $\delta > 0$.

Then, by induction, $x_m = (t+\delta, t+\delta, \ldots, t+\delta)$ is above the roots of p_m.

This would imply that $\operatorname{maxroot}\left(\mathbb{E}\det(\lambda I - \sum_{i=1}^{m} v_i v_i^T)\right) \leq t+\delta$.

It remains to optimize t and δ to prove the best upper bound.

Finally, we need to compute $\partial x_i\, p$ explicitly, for which we use the following formula.

Lemma (Jacobi's formula) $\partial_t \det(A+tB) = \det(A+tB)\operatorname{Tr}\left((A+tB)^{-1}B\right)$.

proof First, we consider $\partial_t \det(A+tB)|_{t=0}$.

Note that $\partial_t \det(A+tB) = \det(A)\,\partial_t \det(I + tA^{-1}B) = \det(A)\,\partial_t \prod_i (1+t\lambda_i)$,

where λ_i are the eigenvalues of $A^{-1}B$.

Since the coefficient of the linear term in $\prod_i (1+t\lambda_i)$ is $\sum_i \lambda_i = \operatorname{Tr}(A^{-1}B)$, we have

$\partial_t \det(A+tB)|_{t=0} = \det(A)\,\partial_t \prod_i (1+t\lambda_i)|_{t=0} = \det(A)\operatorname{Tr}(A^{-1}B)$, as only

the coefficient of the linear term remains after substituting $t=0$.

To compute $\partial_t \det(A+tB)$, we can compute $\partial_x \det(A+tB+xB)|_{x=0}$, which is equal

to $\det(A+tB)\operatorname{Tr}\left((A+tB)^{-1}B\right)$ by the above calculation. □

Initial value

Recall that $p_0(x_1,\ldots,x_m) = \det\left(\sum_{i=1}^m x_i A_i\right)$ where $A_i = \mathbb{E}[v_i v_i^T] \succcurlyeq 0$ and $\sum_{i=1}^m A_i = \sum_{i=1}^m \mathbb{E}[v_i v_i^T] = I_n$.

We need to compute

$$\Phi_{p_0}^{j}(x) = \partial x_j \log\det\left(\sum_{i=1}^m x_i A_i\right) = \frac{\partial x_j \det\left(\sum_{i=1}^m x_i A_i\right)}{\det\left(\sum_{i=1}^m x_i A_i\right)}$$

$$= \frac{\det\left(\sum_{i=1}^m x_i A_i\right)\operatorname{Tr}\left(\left(\sum_{i=1}^m x_i A_i\right)^{-1} A_j\right)}{\det\left(\sum_{i=1}^m x_i A_i\right)} \quad \text{by Jacobi}$$

$$= \operatorname{Tr}\left(\left(\sum_{i=1}^m x_i A_i\right)^{-1} A_j\right).$$

Put in $y = x_0 = (t,\ldots,t)$, it follows that $\Phi_{p_0}^{j}(x) = \operatorname{Tr}\left((tI)^{-1}A_j\right)$ using $\sum_{i=1}^m A_i = I$

$= \frac{1}{t}\operatorname{Tr}(A_j) \leq \frac{\varepsilon}{t}$ where we finally use the assumption $\operatorname{Tr}(A_i) \leq \varepsilon$.

If we set t so that $\Phi_{p_0}^{j}(y) \leq \frac{\varepsilon}{t} \leq 1 - \frac{1}{\delta}$, then we will get the final bound $t+\delta$.

So, we should set t so that $\frac{\varepsilon}{t} = 1 - \frac{1}{\delta}$, so the final bound is $t + \frac{1}{1-\varepsilon/t}$.

This is minimized when $t = \sqrt{\varepsilon} + \varepsilon$, then the final bound is $(1+\sqrt{\varepsilon})^2$.

This completes the proof of the main theorem.

Monotonicity and convexity

We need to establish the following two properties of the barrier functions:

Monotonicity: $\Phi_p^i(y+\delta e_j) \le \Phi_p^i(y)$ where e_j is the j-th standard vector.

Convexity: $\Phi_p^i(y+\delta e_j) \le \Phi_p^i(y) + \delta\, \partial_{x_j} \Phi_p^i(y+\delta e_j)$.

<u>When $i=j$</u>

First, we see that when $i=j$, it is easy to show that the two properties hold.

Recall that $\Phi_p^i(y) = \sum_{j=1}^{d} \frac{1}{y_i - \lambda_j}$ where $\lambda_1, \dots, \lambda_d$ are the roots of the univariate polynomial restricted to y_i by fixing the values on the remaining variables.

Since y_i is above the root, $y_i > \lambda_j$ for all $1 \le j \le d$.

So, if we increase y_i, the function value will decrease (i.e. $\Phi_p^i(y+\delta) < \Phi_p^i(y)$), and thus monotonicity.

To prove convexity, we need to show that $\Phi_p^{i\,''}(y) \ge 0$ for all y above the roots.

A direct calculation shows that $\Phi_p^{i\,''}(y) = \sum_{j=1}^{d} \frac{2}{(y_i - \lambda_j)^3} > 0$ as $y_i > \lambda_j$ $\forall j$ when y is above the roots.

<u>When $i \neq j$</u>

The proof is nontrivial when $i \neq j$, as we have a bivariate polynomial rather than a univariate polynomial.

The proof of [MSS] uses a deep result by that any bivariate real-stable polynomial $p(x_1, x_2)$ can be written as $\pm \det(x_1 A + x_2 B + C)$ for some $A, B \succcurlyeq 0$ and some symmetric C.

Then they do some explicit computations from this to prove monotonicity and convexity.

We present the proof (sketch) by Tao, which is more elementary and self-contained.

By freezing other variables and relabeling, we assume $i=1$ and $j=2$ and consider the resulting real stable polynomial $p(x_1, x_2)$.

Let $x = (x_1, x_2)$ be a point above the roots of p.

The monotonicity property and convexity follows from the more general claim that

$$0 \le (-1)^k \frac{\partial}{\partial x_2^k} \Phi_p^1(x) = (-1)^k \frac{\partial}{\partial x_2^k} \frac{\partial}{\partial x_1} \log p(x) = \frac{\partial}{\partial x_1}\left((-1)^k \frac{\partial}{\partial x_2^k} \log p(x) \right).$$

Therefore, we just need to show that $(-1)^k \frac{\partial}{\partial x_2^k} \log p(x)$ is non-decreasing when we increase x_1.

For fixed $x_1 \in \mathbb{R}$, the polynomial $p_{x_1}: x_2 \to p(x_1, x_2)$ is real-stable, and thus real-rooted, and we denote the roots by $y_1(x_1), \dots, y_d(x_1)$.

Then, $(-1)^k \frac{\partial}{\partial x_2^k} \log p(x) = -(k-1)! \sum_{j=1}^{d} \frac{1}{(x_2 - y_j(x_1))^k}$.

the roots by $y_1(x_1), \ldots, y_d(x_1)$.

Then, $$(-1)^k \frac{\partial}{\partial x_2^k} \log p(x) = -(k-1)! \sum_{i=1}^{d} \frac{1}{(x_2 - y_i(x_1))^k}.$$

To show that the LHS is non-decreasing when we increase x_1, it suffices to show that each term $1/(x_2 - y_i(x_1))^k$ is non-increasing when we increase x_1.

As x_2 is above the roots, it suffices to show that $y_j(x_1)$ is non-increasing when we increase x_1.

Suppose, by contradiction, that this is not the case.

Then, for a "generic" x_1, using some arguments in complex analysis, one can argue that there exists a point $x^0 \in \mathbb{R}$ such that the function $y_j(x_1)$ has positive derivative in an open interval containing x^0 in the real line.

Furthermore, by more complex analysis arguments, there is an open neighborhood containing x^0 in the complex plane such that the function $y_j(x_1)$ is complex analytic, meaning that it is differentiable along any direction on the complex plane.

Now, we would like to argue that the roots $y_j(x^0 + \varepsilon i)$ for some small $\varepsilon > 0$ will have a positive imaginery part, and this would imply that there exist a root $(x^0 + \varepsilon i, y_j(x^0 + \varepsilon i))$ of the bivariate polynomial p where both coordinates have positive imaginery parts, contradicting real stability of p.

To see this, as the function $x_1 \to y_j(x_1)$ is complex analytic around x^0, it holds that

$$\lim_{\varepsilon \to 0} \frac{y_j(x^0 + \varepsilon i) - y_j(x^0)}{\varepsilon i} = c$$ where c is a positive real number by our assumption (for contradiction).

It follows that there exists a small enough ε such that $\frac{y_j(x^0 + \varepsilon i) - y_j(x^0)}{\varepsilon i} = \tilde{c} + di$ where $\tilde{c} > 0$.

For this ε, the root $y_1(x^0 + \varepsilon i)$ has positive imaginery part.

This completes the proof sketch.

We skipped many arguments using complex analysis; see [Tao] for details.

References

[MSS] Interlacing families II: Mixed characteristic polynomials and the Kadison-Singer problem.

[Tao] Real stable polynomials and the Kadison-Singer problem, blogpost.