CS 860 Spectral graph theory. Spring 2019, Waterloo.

Lecture 16: Strongly Rayleigh measure

We see an interesting variation of the method of interlacing polynomials in tackling the thin tree conjecture.

We study the underlying mathematics behind this development, the concept of strongly Rayleigh probability measures.

Thin tree

Given an undirected graph $G=(V,E)$, for $0<\alpha<1$, we say a spanning tree T is α-thin if for all $S \subseteq V$, we have $|\delta_T(S)| \leq \alpha \cdot |\delta_G(S)|$.

In words, a spanning tree is α-thin if it uses at most α fraction of edges in every cut.

There is a very strong conjecture about the existence of a thin tree.

Goddyn's conjecture Every k-edge-connected graph has a $O(\frac{1}{k})$-thin spanning tree.

If the conjecture is true and a $O(\frac{1}{k})$-thin tree can be found in polynomial time, then it would imply a constant factor approximation algorithm for the asymmetric traveling salesman problem.

It can be proved that a random spanning tree is a $O(\frac{\log n}{\log\log n} \cdot \frac{1}{k})$-thin tree.

The argument is similar to that in cut sparsification, using Chernoff bound and careful union bound.

The reason that we can apply Chernoff bound is that the edges in a random spanning tree are negatively correlated, a result that we will study soon.

Spectrally thin tree

As in graph sparsification, we consider a spectral generalization of thinness of a tree.

We say a tree T is α-spectrally-thin if $L_T \preccurlyeq \alpha L_G$.

Note that it is a stronger notion than (combinatorial) thinness, as $L_T \preccurlyeq \alpha L_G$ implies that $|\delta_T(S)| = x_S^T L_T x_S \leq \alpha\, x_S^T L_G x_S = \alpha |\delta_G(S)|$, just as in spectral sparsification.

One advantage of this stronger notion is that it is easier to work with. For example, given a tree, it is easy to check whether it is α-spectrally-thin, while it is not known how to check whether it is (combinatorially) α-thin.

Moreover, the result by Marcus-Spielman-Srivastava implies a non-trivial sufficient condition for the existence of a spectrally thin tree.

Corollary [MSS] If the maximum effective resistance of an edge in G is α, then G has a $O(\alpha)$-spectrally-thin tree.

Recall that the MSS theorem in L15 implies that if the maximum effective resistance of an edge in G is α, then the edge set of G can be partitioned into two subgraphs H_1 and H_2 such that $\frac{1}{2}(1-\sqrt{2\alpha})^2 L_G \preccurlyeq L_{H_i} \preccurlyeq \frac{1}{2}(1+\sqrt{2\alpha})^2 L_G$ for $i \in \{1,2\}$.

We can then recursively apply MSS theorem again in each subgraph (with slightly weaker bounds on maximum effective resistance of an edge) until we cannot apply again, by that time there will be $O(\frac{1}{\alpha})$ edge-disjoint subgraphs of G, each is connected and $O(\alpha)$-spectrally-thin.

See Harvey-Olver or the blogpost by Srivastava for details.

This gives hope that the techniques developed by MSS can be used to prove Goddyn's conjecture.

The corollary gives us a spectrally thin tree, which is combinatorially thin, but it requires a stronger assumption that the maximum effective resistance of an edge is small, which is not necessarily satisfied in an k-edge-connected graph.

The recent breakthrough by Anari and Oveis Gharan, in a very high level, can be seen as a way to reduce the combinatorial problem to the spectral problem, and to use an interesting variant of the MSS theorem to obtain the following result.

Theorem [Anari, Oveis Gharan] Every k-edge-connected graph has a $O(\log\log n \cdot \frac{1}{k})$-thin tree.

The reduction, however, is very complicated and technically challenging.

Also, there is now a constant factor approximation algorithm for ATSP, without using thin trees.

So, we just focus on the underlying mathematics of this development, and discuss the beautiful variant of the MSS theorem that Anari and Oveis Gharan proved.

Probability distribution and real-rooted polynomial

As a warm up to strongly Rayleigh measure, we discuss a simpler one-dimensional analog.

Let X be a random variable over $[d] = \{0, 1, \dots, d\}$ with $\Pr(X=i) = \mu_i$.

Let $p(x) = \sum_{i=0}^{d} \mu_i x^i$ be its generating polynomial.

We would like to understand the relations between the properties of X and the properties of its generating polynomial $p(x)$.

For example, what if $p(x)$ is real-rooted?

<u>Proposition</u> $p(x)$ is real-rooted if and only if μ can be written as the sum of independent Bernoulli random variables.

The proof can be found in Oveis Gharan course notes (L10 - L11).

By chernoff bound, we expect that the distribution $\mu = (\mu_0, \dots, \mu_d)$ should look like a Bell curve, e.g. log-concave and in particular unimodal (i.e. $\exists k$ s.t. $\dots \leq \mu_{k-2} \leq \mu_{k-1} \leq \mu_k \geq \mu_{k+1} \geq \dots$).

<u>Theorem</u> (Newton inequality) For any real-rooted polynomial $p(x) = \sum_{i=0}^{d} \mu_i x^i$, the sequence of coefficients $\{\mu_0, \mu_1, \dots, \mu_d\}$ is ultra-log-concave, i.e. $\frac{\mu_{k-1}}{\binom{d}{k-1}} \cdot \frac{\mu_{k+1}}{\binom{d}{k+1}} \leq \left(\frac{\mu_k}{\binom{d}{k}}\right)^2$ for all $0 < k < d$.

<u>proof</u> The proof is by closure properties of real-rooted polynomials in L13.

First, $p_1(x) = \frac{d^{i-1}}{dx^{i-1}} p(x)$ is real-rooted, and this shaves off the coefficients $\mu_0, \dots, \mu_{i-2}$.

Then, $p_2(x) = t^{d-i+1} p_1(\frac{1}{t})$ is real-rooted, and this reverse the coefficients of $p_1(x)$.

Finally, $p_3(x) = \frac{d^{d-i-1}}{dt^{d-i-1}} p_2(x)$ is real-rooted, and this shaves off the coefficients $\mu_{i+2}, \dots, \mu_d$.

Now, $p_3(x)$ is a quadratic real-rooted polynomial $p_3(x) = \frac{d!}{2}\left(\frac{\mu_{i-1}}{\binom{d}{i-1}} x^2 + \frac{2\mu_i}{\binom{d}{i}} x + \frac{\mu_{i+1}}{\binom{d}{i+1}}\right)$.

The theorem follows from that a quadratic polynomial is real-rooted iff its discriminant is non-negative. $\square$

It follows from Newton inequality that the sequence of coefficients is log-concave (i.e. $a_{k-1} \cdot a_{k+1} \leq a_k^2$ for all $0 < k < d$), and in particular unimodal.

Another consequence is that the density function of a sum of independent Bernoulli random variable is an ultra log-concave sequence of numbers.

<u>Strongly Rayleigh measures</u>

Let X be a random variable over $\{0,1\}^m$ (i.e. a random subset of m elements), with probability distribution $\mu: \{0,1\}^m \to \mathbb{R}$ such that $\mu(S) \geq 0$ for every subset $S \subseteq [m]$ and $\sum_{S \subseteq [m]} \mu(S) = 1$.

Definition (strongly Rayleigh measure) Given a probability distribution $\mu : \{0,1\}^m \to \mathbb{R}$, we define its generating polynomial $g_\mu(x_1,\dots,x_m) := \sum_{S \subseteq [m]} \mu(S) \prod_{i \in S} x_i$, i.e. the coefficient of the monomial $\prod_{i\in S} x_i$ is $\mu(S)$.

We say μ is strongly Rayleigh if its generating polynomial g_μ is a real-stable polynomial.

Determinantal measure

This is an important class of strongly Rayleigh measure.

Definition Let X be a random variable over $\{0,1\}^m$ with probability distribution $\mu : \{0,1\}^m \to \mathbb{R}$.

We say μ is determinantal if there exists a matrix $A \in \mathbb{R}^{m\times m}$ such that

$$\Pr(S \subseteq X) = \sum_{T : T \supseteq S} \mu(T) = \det(A_{S,S}) \quad \text{for every subset } S \subseteq [m],$$

where $A_{S,S}$ is the $|S|\times|S|$-submatrix of A restricting to the rows and columns corresponding to S.

Theorem If μ is determinantal with a matrix $0 \preceq A \preceq I$, then μ is strongly Rayleigh.

Proof Let $h(x) = \det(I - A + A \cdot \operatorname{diag}(x))$ where $x = (x_1,\dots,x_m)$ is an m-dimensional vector.

We claim that $h(x)$ is the generating polynomial of μ which is $g_\mu(x) = \sum_{S\subseteq[m]} \mu(S) \prod_{i\in S} x_i$, and also that $h(x)$ is a real stable polynomial. These two claims will imply the theorem.

First, we check that $h(x)$ is a real-stable polynomial.

We prove the claim when $0 \prec A \preceq I$, and the claim for $0 \preceq A \preceq I$ will follow by a continuity argument that we have seen in L13.

Note that $h(x) = \det(I - A + A\cdot\operatorname{diag}(x)) = \det(A)\det(A^{-1} - I + \operatorname{diag}(x))$, where we used $A \succ 0$ so that A^{-1} exists and also $\det(A) > 0$.

Since $0 \preceq A \preceq I$, it follows that $B := A^{-1} - I \succeq 0$, and so $\det(A^{-1} - I + \operatorname{diag}(x))$ can be written as $\det(B + \sum_{i=1}^m x_i \operatorname{diag}(e_i))$ where $B \succeq 0$ and $\operatorname{diag}(e_i) \succeq 0$ for $1 \le i \le m$.

By the result in L13, $\det(B + \sum_{i=1}^m x_i \operatorname{diag}(e_i))$ is a real stable polynomial. (Actually, we only proved in L13 the case when $B \succ 0$, but the case $B \succeq 0$ follows from the same continuity argument.)

Next, we check that $h(x) = g_\mu(x)$.

To do this, we will prove that $h(x_S) = g_\mu(x_S) = \sum_{R\subseteq S} \mu(R)$ for every subset $S \subseteq [m]$ where x_S is the characteristic vector of the subset S, and it will follow that $h(x) = g_\mu(x)$.

Now, $h(x_S) = \det(I - A + A\cdot \operatorname{diag}(x_S)) = \det\left(I - A + A\begin{pmatrix} \overbrace{1 \ \ddots \ 1}^{|S|} & \\ & 0 \end{pmatrix}\right) = \det\begin{pmatrix} I_{|S|} & -A_{S,\bar{S}} \\ \end{pmatrix} = \det(I_{m-|S|} - A_{\bar{S},\bar{S}})$.

characteristic vector of the subset S, and it will follow that $h(x) = g_\mu(x)$.

Now, $h(x_S) = \det(I - A + A \cdot \mathrm{diag}(x_S)) = \det\left(I - A + A \begin{pmatrix} \overbrace{1_{\ddots}{}_{1}}^{|S|} & \\ & 0_{\ddots}{}_{0} \end{pmatrix} \right) = \det \begin{pmatrix} I_{|S|} & -A_{S,\bar{S}} \\ 0 & I_{m-|S|} - A_{\bar{S},\bar{S}} \end{pmatrix} = \det(I_{m-|S|} - A_{\bar{S},\bar{S}})$.

We use the formula $\det(\lambda I_n - A) = \sum_{k=0}^{n} \lambda^{n-k} (-1)^k \sum_{S \in \binom{n}{k}} \det(A_{S,S})$ where $\det(A_{\phi,\phi}) = 0$

So, $h(x_S) = \det(I_{m-|S|} - A_{\bar{S},\bar{S}}) = 1 + \sum_{k=1}^{m-|S|} (-1)^k \sum_{R \subseteq \bar{S} : |R| = k} \det(A_{R,R})$

$= 1 + \sum_{k=1}^{m-|S|} (-1)^k \sum_{R \subseteq \bar{S} : |R| = k} \Pr(X \text{ contains } R)$ by the definition of determinantal measure where X is the random outcome

$= 1 - \Pr(X \text{ contains some element in } \bar{S})$ by inclusion-exclusion principle

$= \Pr(X \text{ contains no element in } \bar{S}) = \Pr(X \subseteq S)$

$= \sum_{R \subseteq S} \mu(R)$.

This proves the second claim and thereby completes the proof of the theorem. □

Spanning tree measure

One interesting example of determinantal measure is the uniform spanning tree measure.

Let $G = (V, E)$ be an undirected graph. Let the edge set E be the ground set with $|E| = m$.

Let N be the number of spanning trees in G.

Let $\mu : \{0,1\}^E \to \mathbb{R}$ be the probability distribution with $\mu(T) = \frac{1}{N}$ if T is a spanning tree of G and zero otherwise.

Burton and Pemantle proved that this measure is determinantal.

Theorem: Let Y be the $m \times m$ matrix where each row and column corresponds to an edge, and $Y_{e,f} = \langle L_G^{+/2} b_e, L_G^{+/2} b_f \rangle$ where L_G^{+} is the pseudo-inverse of the Laplacian matrix and $b_e = x_i - x_j$ for an edge ij.

Then, for a subset of edges $F \subseteq E$, $\Pr_{T \sim \mu}(F \subseteq T) = \det(Y_{F,F})$.

Proof sketch: The base case is $\Pr_{T \sim \mu}(e \in T) = \det(Y_{e,e}) = \langle L_G^{+/2} b_e, L_G^{+/2} b_e \rangle = b_e L_G^{+} b_e = R_{eff}(e)$, and this is known to hold by a classical result of Kirchhoff.

Let $\tilde{L}_G = L_G + 11^T/n$ where n is the number of vertices in G, so that $\tilde{L}_G$ is of full rank.

The proof of the matrix tree theorem can be used to show that the number of spanning trees is equal to $\frac{1}{n} \det(\tilde{L}_G)$.

Then, we can write $\Pr_{T \sim \mu}[T \cap F = \phi] = \dfrac{\det(\tilde{L}_G - \sum_{e \in F} b_e b_e^T)}{\det(\tilde{L}_G)} = \det(I - \sum_{e \in F} b_e b_e^T)$.

On one hand, $\Pr_{T\sim\mu}[T\cap F=\phi] = 1-\Pr[T\cap F\neq\phi] = 1-\Pr(T \text{ contains some element in } F)$, and this can be computed using the inclusion-exclusion principle as in the previous proof.

On the other hand, we can expand $\det(I-\sum_{e\in F} b_e b_e^T)$ using the formula as in the previous proof, and using the Cauchy-Binet formula we can see that $\det(Y_{F',F'})$ for $F'\subseteq F$ satisfy the same inclusion-exclusion formula.

By induction, each of the term $\Pr(F'\subseteq T)$ for some $F'\subseteq E$ can be written as $\det(Y_{F',F'})$, and it follows that $\Pr(F\subseteq T)=\det(Y_{F,F})$.

Roughly speaking, the statement holds because the formula for characteristic polynomial and the inclusion-exclusion formula are the same, and each term equals because the base cases hold.

See the notes of Oveis-Shayan (L03) for details. □

Finally, note that $Y=(B^T L_G^{+/2})(L_G^{+/2}B) = B^T L_G^{+} B$ where B is the $n\times m$ matrix where the e-th column is b_e.

This implies that $Y^2 = B^T L_G^+ B B^T L_G^+ B = B^T L_G^+ L_G L_G^+ B = B^T L_G^+ B = Y$ as $BB^T = L_G$.

So, Y is a projection matrix and thus each eigenvalue is either zero or one.

Therefore, $0 \preccurlyeq Y \preccurlyeq I$ and so the previous theorem can be used to prove that the uniform random spanning tree measure is a determinantal measure.

A remark is that the proof can be extended to the edge weighted case where the probability of a spanning tree is proportional to $\prod_{e\in T} w_e$, to show that the edge weighted distribution is also determinantal.

Properties of strongly Rayleigh measure

Some useful properties of strongly Rayleigh measures follow from closure properties of real stable polynomials.

Conditioning

For a variable x_i, the conditional probability distributions $\mu|_{x_i=0}$ and $\mu|_{x_i=1}$ of a strongly Rayleigh distribution μ are also strongly Rayleigh.

To see this, let $g_\mu(x_1,\ldots,x_m)$ be the generating polynomial of μ.

The generating polynomial of $\mu|_{x_i=0}$ is simply $g(x_1,\ldots,x_{i-1},0,x_{i+1},\ldots,x_m)/\sum_{S:x_i\notin S}\mu(S)$, and thus is real stable because of substitution of x_i by a real number, and hence $\mu|_{x_i=0}$ is strongly Rayleigh.

Note that the generating polynomial of $\mu|_{x_i=1}$ is $x_i\frac{\partial}{\partial x_i} g(x_1,\ldots,x_m)/\sum_{S:x_i\in S}\mu(S)$, and this is real stable because differentiation is a stability preserving operation (either by Borcea-Branden or by direct proof by reducing to the univariate case by fixing other variables), and hence $\mu_{|x_i=1}$ is strongly Rayleigh.

So, fixing the values of a subset of variables, the conditional probability measure remains strongly Rayleigh.

Projection

For any set $S \subseteq [m]$, the projection of μ onto S, denoted by $\mu_{|S}$ is the measure where for any set $R \subseteq S$,

$$\mu_{|S}(R) = \sum_{T \subseteq [m] : T \cap S = R} \mu(T).$$

The generating polynomial $g_{\mu_{|S}}$ is obtained from g_μ by simply substituting $x_j = 1$ for all variables $j \notin S$, and hence $\mu_{|S}$ is strongly Rayleigh.

Negative correlation

This is probably the most important property of strongly Rayleigh measure, as for instance it allows us to apply Chernoff bounds on the variables to prove concentration results.

The simplest form of negative dependency is $\Pr(x_i = 1 \mid x_j = 1) \leq \Pr(x_i = 1)$ for any two variables $i \neq j$.

Note that the probability $\Pr(x_i = 1)$ can be read from the generating polynomial $g_\mu(x_1, \ldots, x_m)$ of μ, as $\Pr(x_i = 1) = \frac{\partial}{\partial x_i} g(x_1, \ldots, x_m)|_{x_1 = x_2 = \cdots = x_m = 1} = \sum_{S : S \ni i} \mu(S)$, the sum of coefficients of μ containing i.

Therefore, we can rewrite the negative correlation inequality as $\Pr(x_i = 1 \text{ and } x_j = 1) \leq \Pr(x_i = 1)\Pr(x_j = 1)$, and then express it using generating polynomial as $\left(\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} g(\vec{1})\right) \cdot g(\vec{1}) \leq \frac{\partial}{\partial x_i} g(\vec{1}) \cdot \frac{\partial}{\partial x_j} g(\vec{1}) \quad \forall i \neq j$.

Strongly Rayleigh measures satisfy this inequality for any $y \in \mathbb{R}^m$ (i.e. not just for $y = \vec{1}$).

Theorem Let g be a multiaffine real stable polynomial.

Then $\left(\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} g(y)\right) \cdot g(y) \leq \frac{\partial}{\partial x_i} g(y) \cdot \frac{\partial}{\partial x_j} g(y) \quad \forall i \neq j$ and $\forall y \in \mathbb{R}^m$.

proof For any $y \in \mathbb{R}^m$, consider the bivariate restriction $f(s,t) = g(y_1, \ldots, y_{i-1}, y_i + s, y_{i+1}, \ldots, y_{j-1}, y_j + t, y_{j+1}, \ldots, y_m)$

Then f is a bivariate real stable polynomial.

Since g is multiaffine, note that $f(s,t) = g(y) + \left(\frac{\partial}{\partial x_i} g(y)\right)s + \left(\frac{\partial}{\partial x_j} g(y)\right)t + \left(\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} g(y)\right)st$.

The univariate polynomial $h(s) = f(s,i)$ is stable (but not necessarily real).

Let $a+bi$ be a root of $h(s)$.

Then $\mathrm{Re}(h(a+bi)) = g(y) + \left(\frac{\partial}{\partial x_i} g(y)\right) \cdot a - \left(\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} g(y)\right) \cdot b = 0$, and

$\mathrm{Im}(h(a+bi)) = \left(\frac{\partial}{\partial x_i} g(y)\right) \cdot b + \frac{\partial}{\partial x_j} g(y) + \left(\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} g(y)\right) \cdot a = 0$.

Solving the two equations by eliminating a, we get

$$\left(\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} g(y)\right)\cdot g(y) - \left(\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} g(y)\right)^2\cdot b = \left(\frac{\partial}{\partial x_i} g(y)\right)^2\cdot b + \left(\frac{\partial}{\partial x_i} g(y)\right)\left(\frac{\partial}{\partial x_j} g(y)\right).$$

By stability of h, we have $b \le 0$.

This implies that $\left(\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} g(y)\right)\cdot g(y) - \left(\frac{\partial}{\partial x_i} g(y)\right)\left(\frac{\partial}{\partial x_j} g(y)\right) \le 0$. □

Negative association

A stronger form of negative dependency is called negative association:

Definition We say the binary random variables $\{X_1, \dots, X_m\}$ are negatively associated if for any two nondecreasing functions $f, g \in \{0,1\}^m \to \mathbb{R}$ that depend on disjoint set of variables, it holds that $E[f(X_1,\dots,X_m)\, g(X_1,\dots,X_m)] \le E[f(X_1,\dots,X_m)]\cdot E[g(X_1,\dots,X_m)]$, where a function f is nondecreasing if $f(x) \ge f(y)$ for $x \ge y$.

Note that negative correlation is a special case of negative association.

Feder and Mihail used negative correlation as the base case in an induction to prove that the random variables of a strongly Rayleigh measure are negative associated.

Borcea, Brändén and Liggett developed the theory of strongly Rayleigh measure and use it to answer many questions about negatively dependent random variables.

Probabilistic method for strongly Rayleigh measure

Recall the probabilistic statement that Marcus, Spielman, Srivastava proved.

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

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

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

The proof crucially uses the assumption that the vectors $v_1, \dots, v_m$ are independent random variables.

Anari and Oveis Gharan proved a beautiful variation for strongly Rayleigh measure.

In the following, a strongly Rayleigh measure μ is homogenous if every non-zero monomial in the generating polynomial g_μ is of the same degree.

Theorem Let $\mu: \{0,1\}^m \to \mathbb{R}$ be a homogenous strongly Rayleigh measure with the maximum marginal probability of a variable/coordinate is ε_1 (i.e. $\Pr_{S\sim\mu}(i \in S) \le \varepsilon_1$ for $1 \le i \le m$).

Given $v_1, \dots, v_m \in \mathbb{R}^n$ with $\sum_{i=1}^m v_i v_i^T = I_n$ and $\|v_i\|_2^2 \le \varepsilon_2$ for $1 \le i \le m$, it holds that

$$\Pr_{S\sim\mu}\left[\left\| \sum_{i\in S} v_i v_i^T \right\| \le 4(\varepsilon_1+\varepsilon_2) + 2(\varepsilon_1+\varepsilon_2)^2 \right] > 0.$$

Let's see how this can be used to prove directly the corollary about spectrally thin tree, without using recursion.

As before, we set $v_i = L^{+/2} b_{e_i}$ so that $\sum_{i=1}^m v_i v_i = I$ and each $\|v_i\|^2 = b_{e_i}^T L_G^+ b_{e_i} = \mathrm{Reff}(e_i) \le \alpha$ by assumption.

We set μ to be the uniform random spanning tree measure, so that it is homogenous strongly Rayleigh,

and furthermore $\Pr_{T\sim\mu}(e \in T) = \mathrm{Reff}(e) \le \alpha$.

So, we can apply the new theorem with $\varepsilon_1 = \varepsilon_2 = \alpha$, and directly get a spanning tree T^* with

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

A key advantage of this theorem is that the output is guaranteed to be a spanning tree, so that we get connectivity for free (without worrying about the minimum eigenvalue), and this is very important in the thin tree problem.

The following is the fundamental building block of the new thin tree result.

<u>Theorem</u> Given a graph $G=(V,E)$ and a subset of edges $F \subseteq E$ such that (V,F) is k-edge-connected, if $\mathrm{Reff}(e) \le \varepsilon$ for all $e \in F$, then G has a $O(\frac{1}{k}+\varepsilon)$-spectrally thin tree in F.

<u>Proof idea</u> Since F is k-edge-connected, there are at least $k/2$ edge disjoint spanning trees in F.

This implies that there is a point in the spanning tree polytope with maximum edge value $O(\frac{1}{k})$.

By expressing this point using "maximum entropy distribution", it will turn out that there is a weighting of the edges, so that the weighted random spanning tree distribution μ (which is homogenous strongly Rayleigh) has maximum marginal probability of an edge $O(\frac{1}{k})$, i.e. $\varepsilon_1 = O(\frac{1}{k})$.

The assumption about effective resistance implies that $\varepsilon_2 \le \varepsilon$, and so the theorem applies. □

With this theorem, their strategy is to add "short-cut" edges in the graph, so that it doesn't change the cut structures much, while creating many edges with small effective resistance.

They do it in $O(\log\log n)$ iterations so that the edges with small effective resistance form a k-edge-connected subgraph.

The most difficult step is to prove the existence of good short-cut edges, which they prove by using a delicate analysis of a semidefinite program.

With 80 pages of work after the above theorem, they manage to prove the existence of a $O(\log\log n \cdot \frac{1}{k})$-thin tree.

Proof ideas

We discuss the proof of the probabilistic statement for strongly Rayleigh measure here.

The proof is also based on the method of interlacing family, with the same two key steps:

① Proving the family of polynomials $\det(\lambda I - \sum_{i\in S} v_i v_i^T)_{S \in supp(\mu)}$ form an interlacing family.

This would imply that $\exists S^* \in supp(\mu)$ such that $\text{maxroot}\left(\det(\lambda I - \sum_{i\in S^*} v_i v_i^T)\right) \le \text{maxroot}\left(\mathbb{E}_{S\sim\mu}\left(\det(\lambda I - \sum_{i\in S} v_i v_i^T)\right)\right)$.

② Upper bound the maximum root of $\mathbb{E}_{S\sim\mu}\left(\det(\lambda I - \sum_{i\in S} v_i v_i^T)\right) = \sum_S \mu(S) \det(\lambda I - \sum_{i\in S} v_i v_i^T)$.

Recall that in [MSS], both steps depend crucially on the multilinear formula.

Anari and Oveis Gharan proved a generalization incorporating the probability measure μ.

Theorem Let $v_1,\dots,v_m \in \mathbb{R}^n$ and the degree of the homogenous generating polynomial g_μ of the measure $\mu: \{0,1\}^m \to \mathbb{R}$ be d.

Then $$\lambda^{d-n} \sum_{S\subseteq[m]} \mu(S) \det(\lambda^2 I - \sum_{i\in S} v_i v_i^T) = \prod_{i=1}^m \left(1 - \frac{\partial^2}{\partial x_i}\right) g_\mu(\lambda\vec{1} + x) \det(\lambda I + \sum_{i=1}^m x_i v_i v_i^T) \Big|_{x_1=x_2=\dots=x_m=0}.$$

Proof The main idea is to write two multilinear polynomials, one with $\mu(S)$ as the coefficient of $\prod_{i\in S} x_i$, and another with $\det(\lambda^2 I - \sum_{i\in S} v_i v_i^T)$ as the coefficient of $\prod_{i\in S} x_i$.

If this can be done, then we can multiply the two polynomials, and take out $\mu(S)\det(\lambda^2 I - \sum_{i\in S} v_i v_i^T)$ as the coefficient of $\prod_{i\in S} x_i^2$ by differentiating $\prod_{i\in S}\frac{\partial^2}{\partial x_i}$ and substituting zero on all variables.

Let's start with the LHS. Let $A_i = v_i v_i^T$, which is rank one so that $\det(\lambda I + \sum_i x_i A_i)$ is multilinear in x_i.

$$\sum_R \mu(R)\det(\lambda I + \sum_{i\in R} x_i A_i) = \sum_R \mu(R) \sum_{S\subseteq R}\left(\prod_{i\in S} x_i\right)\left(\prod_{i\in S}\frac{\partial}{\partial x_i} \det(\lambda I + \sum_{j\in R} x_j A_j)\Big|_{x_1=\dots=x_m=0}\right) \quad \text{(multilinear coefficients)}$$

$$= \sum_R \mu(R) \sum_{S\subseteq R}\left(\prod_{i\in S} x_i\right)\left(\prod_{i\in S}\frac{\partial}{\partial x_i} \det(\lambda I + \sum_{j=1}^m x_j A_j)\Big|_{x_1=\dots=x_m=0}\right) \quad \text{(expanding and including all variables)}$$

$$(*) \quad = \sum_S \left(\sum_{R: R\supseteq S} \mu(R)\right)\cdot\left(\prod_{i\in S}\frac{\partial}{\partial x_i} \det(\lambda I + \sum_{j=1}^m x_j A_j)\Big|_{x_1=\dots=x_m=0}\right)\cdot\left(\prod_{i\in S} x_i\right)$$

Now, let's come up with one polynomial g with coefficient $\sum_{R:R\supseteq S}\mu(R)$ on the monomial $\prod_{i\in S} x_i$, and another polynomial f with coefficient $\prod_{i\in S}\frac{\partial}{\partial x_i}\det(\lambda I + \sum_{j=1}^m x_j A_j)\Big|_{x_1=\dots=x_m=0}$ on the monomial $\prod_{i\in S} x_i$.

Consider $g_\mu(\lambda\vec{1} + x) = \sum_R \mu(R) \prod_{i\in R}(\lambda + x_i)$.

Each R with $R \supseteq S$ contributes $\mu(R)\lambda^{|R|-|S|}$ to $\prod_{i\in S} x_i$.

Therefore, since μ is homogenous, the coefficient of $\prod_{i\in S} x_i$ in g_μ is $\sum_{R:R\supseteq S}\mu(R)\cdot\lambda^{|R|-|S|} = \lambda^{d-|S|}\sum_{R:R\supseteq S}\mu(R)$.

Consider $f(x) = \det(\lambda I + \sum_{i=1}^{m} x_i A_i)$.

As A_i is rank-one, $f(x)$ is multilinear in x_i.

Therefore, the coefficient of $\prod_{i\in S} x_i$ is $\prod_{i\in S} \frac{\partial}{\partial x_i} \det(\lambda I + \sum_{j=1}^{m} x_j A_j)\Big|_{x_1=\cdots=x_m=0}$.

So, since both f and g are multilinear in x_i, the coefficient of $\prod_{i\in S} x_i^2$ in $f\cdot g$ is the product

of the coefficients of $\prod_{i\in S} x_i$ in f and in g.

We can read the coefficient of $\prod_{i\in S} x_i^2$ in $f\cdot g$ as $2^{-|S|} \prod_{i\in S} \frac{\partial}{\partial x_i^2} f\cdot g\Big|_{x_1=\cdots=x_m=0}$.

Therefore, $2^{-|S|} \prod_{i\in S} \frac{\partial}{\partial x_i^2} f\cdot g\Big|_{x_1=\cdots=x_m=0} = \Big(\lambda^{d-|S|} \sum_{R: R\supseteq S} \mu(R)\Big)\cdot\Big(\prod_{i\in S} \frac{\partial}{\partial x_i} \det(\lambda I + \sum_{j=1}^{m} x_j A_j)\Big|_{x_1=\cdots=x_m=0}\Big)$.

$$\text{Hence, } \prod_{i=1}^{m}\Big(1-\frac{\partial}{\partial x_i^2}\Big) f\cdot g\Big|_{x_1=\cdots=x_m=0} = \sum_{S\subseteq[m]} (-1)^{|S|} \prod_{i\in S} \frac{\partial}{\partial x_i^2} f\cdot g\Big|_{x_1=\cdots=x_m=0}$$

$$= \sum_{S\subseteq[m]} (-1)^{|S|}\cdot 2^{|S|}\cdot \lambda^{d-|S|}\cdot\Big(\sum_{R:R\supseteq S} \mu(R)\Big)\cdot\Big(\prod_{i\in S}\frac{\partial}{\partial x_i} \det(\lambda I + \sum_{j=1}^{m} x_j A_j)\Big|_{x_1=\cdots=x_m=0}\Big).$$

The LHS of this identity is exactly the RHS of the theorem.

The RHS of this identity is equal to λ^d times the LHS of (*) when we plug in $x_i = \frac{-2}{\lambda}$ for $1\le i\le m$.

Therefore, $\prod_{i=1}^{m}\Big(1-\frac{\partial}{\partial x_i^2}\Big) f\cdot g\Big|_{x_1=\cdots=x_m=0} = \lambda^d \sum_R \mu(R)\cdot\det(\lambda I - \frac{2}{\lambda}\sum_{i\in R} A_i) = \lambda^{d-n} \sum_R \mu(R)\cdot\det(\lambda^2 I - 2\sum_{i\in R} v_i v_i^T)$. □

Once this formula is established, the remaining proof has a similar structure to that in MSS.

First, the formula shows that the expected characteristic polynomial is real-rooted, because g is real-stable

as μ is strongly Rayleigh, and so $f\cdot g$ is real stable as both f, g are, and then the operation

$1-\frac{\partial^2}{\partial x_i}$ is real-stability preserving as it is equal to $(1-\frac{\partial}{\partial x_i})(1+\frac{\partial}{\partial x_i})$ as both are real-stability

preserving as shown in L13.

Then, it follows that the family of polynomials $\det(\lambda I - \sum_{i\in R} v_i v_i^T)$ form an interlacing family.

The more difficult step is to upper bound of the maxroot of the expected characteristic polynomial

using the multivariate barrier argument.

The potential functions are the same $\frac{\partial}{\partial x_i}\log p(x_1,\ldots,x_m)$, where $p = f\cdot g$.

They need to compute the second derivative of the potential functions because of $1-\frac{\partial^2}{\partial x_i}$.

A key difference is that now the target upper bound is $O(\varepsilon_1+\varepsilon_2)$, much smaller than $(1+\sqrt{\varepsilon})^2$ as in MSS,

but it turns out that the $1-\frac{\partial^2}{\partial x_i}$ operation allows for much smaller shift.

References

- The Kadison-Singer problem for strongly Rayleigh measures and applications to ATSP, by Anari, Oveis Gharan.
- Effective-resistance-reducing flows, spectrally thin trees, and asymmetric TSP, by Anari, Oveis Gharan.
- Negative dependency and the geometry of polynomials, by Borcea, Brändén, Liggett.
- Course notes of "Recent developments in approximation algorithms", by Oveis Gharan.