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

Lecture 8: Kadison-Singer and Weaver's Conjecture

Lecturer: Aaron Schild *Scribe: Alex Rusciano*

1.1 Motivation: Ramanujan Graphs and Spectral Sparsifiers

We start by giving one formulation of the Kadison-Singer conjecture, and its graph theoretic significance. Ramanujan graphs provide motivation.

Definition 1.1. A Ramanujan Graph G of parameter d has eigenvalues $\lambda_i(A_G) < 2\sqrt{d-1}$ for $i < n$ and all vertices have degree d. Here A_G is its adjacency matrix, and eigenvalues are labeled in increasing order up to n, the number of vertices.

To interpret this in terms of spectral sparsifiers, it helps to introduce the graph Laplacian. The graph Laplacian is a positive semi-definite (PSD) version of the adjacency matrix that is cleaner to work with. In particular, it is defined by subtracting the adjacency matrix from the diagonal vertex degree matrix

$$L_G = D_G - A_G$$

One can check that the eigenvalues of the complete graph's Laplacian, L_{K_n} are 0 with multiplicity 1 and the others are n. On the other hand, those of Ramanujan L_G are larger than $d - 2\sqrt{d-1}$.

Consequently, it directly follows that

$$(1 - \frac{2\sqrt{d-1}}{d})L_{K_n} \preceq \frac{n}{d}L_G \preceq (1 + \frac{2\sqrt{d-1}}{d})L_{K_n}$$

Here $\preceq$ denotes the PSD ordering. As $d << n$ (potentially held fixed), in the spectral sense, the Ramanujan graphs are sparsifiers of complete graphs.

A result of Spielman and Srivastava shows that all graphs admit spectral sparsification. Something along the lines of

Theorem 1.2. *One can find a subgraph H of G with average degree $O(\frac{1}{\epsilon^2})$ such that*

$$(1-\epsilon)L_G \leq L_H \leq (1+\epsilon)L_G$$

The goal here is not to prove this result, but to bring to light one annoying feature of the result. Namely, even if G was an unweighted graph, H is a weighted graph with potentially very uniform weights. This contrasts to the Ramanujan case. A step in this direction will be when we prove in this and the next lecture:

Theorem 1.3. *Define the effective resistances of G to be $r_e := b_e^T L_G^+ b_e$. Here b_e is the vertex-incidence vector of the arbitrarily oriented version of edge e. Then so long as $\max_e r_e \leq \epsilon$, we may partition the edges of G into two halfs H_1, H_2 so that*

$$(\frac{1}{2} - \sqrt{\epsilon})L_G \leq L_{H_i} \leq (\frac{1}{2} + \sqrt{\epsilon})L_G$$

There are good reasons to care about spectral sparsifiers. Here we just mention that spectral sparsifiers preserve cut sizes up to a multiplicative factor. To see this, first observe that $L_G = \sum_e b_e b_e^T$. Now the cut represented by $0, 1$ label vector x is given by

$$\sum_e (x_i - x_j)^2 = \sum_e x^T b_e b_e^T x = \|x^T L_G x\|^2$$

1.2 A weaker guarantee by concentration inequalities

A matrix concentration inequality of Tropp forms the basis of this section. It is

Theorem 1.4.

$$\mathbb{P}[\|\sum \psi_i A_i - \mathbb{E}\sum \epsilon_i A_i\| > t] \leq 2n\exp(\frac{-t^2}{\|\sum A_i^2\|})$$

where ψ_i are i.i.d. Bernoulli-$\frac{1}{2}$.

We can readily prove the following version with high probability

Theorem 1.5. *Suppose $r_e \leq \epsilon$. Then in high probability, by simply including edges with probability $\frac{1}{2}$ one has*

$$(\frac{1}{2} - \sqrt{\epsilon \log(n)})G \leq H \leq (\frac{1}{2} + \sqrt{\epsilon \log(n)})G$$

Proof. One could immediately apply the concentration inequality to the $b_e b_e^T$ but then we would only get closeness in the operator norm, not the PSD ordering. (Roughly / intuitively, closeness of 1 eigenvalue vs closeness of all). So first we apply a normalizing procedure: Define $c_e = (L_G^+)^{1/2}$ so that $\sum_e c_e c_e^T = I$. Then apply Tropp's theorem with these outer products as the A_i, after calculating

$$\|\sum A_i^2\| \leq \max r_e \leq \epsilon$$

one gets

$$\mathbb{P}[\|\sum_e c_e c_e^T - \frac{1}{2}I\| > t] \le 2n\exp(\frac{-t^2}{\epsilon})$$

so taking $t = O(\sqrt{\epsilon\log(n)})$ achieves concentration. Let M' be resulting matrix. Then

$$(\frac{1}{2} - O(\sqrt{\epsilon\log(n)}))I \preceq M \preceq (\frac{1}{2} + O(\sqrt{\epsilon\log(n)}))I$$

and conjugating and recalling how c_e were defined,

$$M = L_G^{1/2} M' L_G^{1/2} = \sum_{e|\psi_e=1} b_e b_e^T$$

satisfies the desired bound

$$(\frac{1}{2} - O(\sqrt{\epsilon\log(n)}))L_G \le M \le (\frac{1}{2} + O(\sqrt{\epsilon\log(n)}))L_G$$

□

However, this result is only non-trivial in the realm $\log(n)\epsilon < 1$. An example shows that concentration based attempts to strengthen the result fail.

1.3 Existance Proof by Interlacing

To give the existance result, we prove the following theorem. It is known to imply Weaver's Conjecture, which in turn implies Kadison-Singer. Kadison-Singer itself is phrased more in terms of operator theory.

Theorem 1.6. *Suppose* $\mathbb{E}v_i v_i^T = I$*, and* $\|v_i\|^2 \le \epsilon$

Then there exists an instance where $\|\sum v_i v_i^T\| < (1+\sqrt{\epsilon})^2$

Before starting the proof, let's see how this implies our graph-theoretic result. Just take $v_i = (0, c_e)$ or $v_i = (c_e, 0)$ with probability $\frac{1}{2}$ each. Then $\|c_e\|^2 \le \epsilon$ is equivalent to $r_e \le \epsilon$. Moreover, $\sum v_i v_i^T$ is consists of 2 blocks, which correspond to the graph partition. The remaining details of the connection are readily filled in.

To prove the theorem, we pursue an interlacing strategy. Recall the first step is to show there's an interlacing family, and the second is to bound the roots. Next week's lecture focuses on the latter, today we established the former.

Proposition 1.7. *For any product distribution over the* $\{v_i\}_{i=1}^n$*, the partially averaged characteristic polynomials* $\mathbb{E}_{s_1,\ldots,s_k}[\det(xI - \sum v_i v_i^T)]$ *form an interlacing family. This is taken to mean*

$$\mathbb{E}[\det(xI - \sum v_i v_i^T)|v_1 = s_1, \ldots, v_k = s_k, v_{k+1} = s]$$

interlace over all choices of s, and for $0 \leq k \leq n-1$. ($k = 0$ corresponds to averaging out all but the first distribution, while $k = n-1$ corresponds to nothing averaged). The definition also requires all of these polynomials to be real rooted. Recall that an inductive argument shows that there will be an instance of the v_i who's largest eigenvalue is at least as large as the largest root of the fully averaged characteristic polynomial.

Proof. From past work, this is equivalent to showing that convex combinations of

$$\mathbb{E}[\det(xI - \sum v_i v_i^T)|v_1 = s_1, \ldots, v_k = s_k, v_{k+1} = s]$$

are real rooted, for the values s can take. Consequently, because v_i are independent and hence the expectations are over product distributions, it suffices to prove

$$\mathbb{E}_{\text{any product distribution}} \det(xI - \sum v_i v_i^T)$$

is real rooted. As it's a product distribution (alternately, the v_i are independent), this is given by

$$\prod (1 - \partial_i) \det(xI + \sum t_i \mathbb{E} v_i v_i^T)|_{t=0}$$

and real stability theorems are applicable to this polynomial. First $\det(xI + \sum t_i \mathbb{E} v_i v_i^T)$ is real stable, because in general things of the form

$$\det(\sum z_i A_i)$$

are real stable so long the $A_i \succeq 0$. Then show $1 - \partial_i$ preserves real stability, by using the classification theorem of Borcea and Branden. That is, just establish the real real stability of

$$(1 - \partial_{x_i}) \prod (x_j + y_j)^{k_j} = (1 - \frac{k_i}{x_i + y_i}) \prod (x_j + y_j)^{k_j}$$

But this is clear, as the imaginary parts would need to cancel, so not all imaginary parts can be positive. Finally, the restriction of a real stable polynomial to one variable (by setting the others to a real number) gives a real rooted polynomial. □

1.4 Next week

Next lecture, an upper bound of $(1 + \sqrt{\epsilon})^2$ on the completely averaged polynomial will be given to complete the proof. The interlacing family property inductively means that some instance of the v_i share the upper bound.

Dan Spielman, Nikhil Srivastava: Interlacing Families II: Mixed Characteristic Polynomials and the Kadison-Singer Problem, http://arxiv.org/abs/1306.3969