Recent Advances in Approximation Algorithms **Spring 2015**

Lecture 13: A Construction of Linear size Thin Forests

Lecturer: Shayan Oveis Gharan *May 11th*

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

In the next few lectures we will talk about the recent polyloglog(n) bound on the integrality gap of the Held-Karp relaxation for Asymmetric TSP [AO14]. Along this way we first go over the recent breakthrough of Marcus, Spielman and Srivastava [MSS13] who proved the long standing Kadison, Singer conjecture [KS59].

As it will be clear, the proof of [MSS13] although not being long is very involved and has several novel ideas. So, we start with a significantly simpler setup and we prove relatively weaker statements. As the ideas become more clear we will go over the actual ideas in [MSS13]. The materials of this lecture are based on the work of Batson, Spielman and Srivastava [BSS14] and lecture notes of Nick Harvey [Har13].

13.1 Combinatorial and Spectral Thin Trees

In lecture 6 we defined a thin spanning tree with respect to feasible solutions of the Held-Karp LP relaxation of ATSP. It is not hard to see that by rescaling the edges we can study thin spanning trees with respect to any unweighted (undirected) graph G.

Definition 13.1 (Thin Tree)**.** *Given an unweighted (undirected) graph $G = (V, E)$, we say a spanning tree $T \subseteq E$ is α-thin with respect to G if for any set $S \subset V$,*

$$|T(S, \overline{S})| \leq \alpha \cdot |E(S, \overline{S})|.$$

Note that because of the rescaling we are interested in finding a tree with $\alpha \ll 1$. It is conjectured that if G is sufficiently connected then we can always find a tree with $\alpha < .99$. Recall that a graph G is k-edge connected (or k-connected for short) if for any set $S \subset V$, $|E(S, \overline{S})| \geq k$.

Conjecture 13.2 (Weak Thin Tree Conjecture)**.** *There exists a number k_0 such that for any $k \geq k_0$, any k-connected graph (of arbitrary size) has a 0.99 thin spanning tree.*

The above conjecture is still open. Note that the main difficulty in proving the above conjecture is for values of k which are significantly small compared to the size of G. In the strong thin tree conjecture we are looking to find a spanning tree T that is $O(1/k)$ thin with respect to G. This question can be significantly hard in some well-known families of graphs. For example, try to construct an $O(1/k)$-thin spanning tree in a k-dimensional hypercube. See Problem 3 of Assignment 3 for the solution of this problem.

We recall that a proof of the weak thin tree conjecture gives an $O(\log(n)^{1-\epsilon})$ bound on the integrality gap of the Held-Karp LP relaxation for some constant $\epsilon > 0$ independent of n. Any proof of the strong thin tree conjecture implies a constant bound on the integrality gap of the Held-Karp LP relaxation. In [AO14] it is shown that any k-connected graph has polyloglog$(n)/k$-thin tree. All of these statements essentially follow from the Hoffman circulation argument that we discussed in lecture 7.

One of the major difficulties in understanding thin trees is that proving the thinness of a tree is not an easy problem. In particular, we are not aware of any polynomial sized certificate to prove the thinness of a given spanning tree. Note that the thinness of T is

$$\min_{S \subset V} \frac{|E(S, \overline{S})|}{|T(S, \overline{S})|},$$

which is an instance of the generalized sparsest cut problem. The best approximation algorithm for this problem is an algorithm of Arora, Lee, Naor [ALN08] that has an approximation factor of $O(\sqrt{\log(n)}\log\log(n))$.

Instead, the idea is to study a generalization of thin trees which is known as spectrally thin trees.

Definition 13.3 (Spectrally Thin Tree). *A spanning tree T is α-spectrally thin with resect to G if*

$$L_T \preceq \alpha \cdot L_G.$$

Recall that $L_T = \sum_{e\in T} b_e b_e^\intercal$ is the Laplacian of T.

First observe that if T is α-spectrally thin, then it is also α-thin. This is because for any set $S \subset V$,

$$|T(S,\overline{S})| = \mathbf{1}_S^\intercal L_T \mathbf{1}_S \leq \alpha \cdot \mathbf{1}_S^\intercal L_G \mathbf{1}_S = \alpha \cdot |E(S,\overline{S})|.$$

In other words, T is α thin if for every test vector x, (that is not necessarily integral), $x^\intercal L_T x \leq \alpha \cdot x^\intercal L_G x$. Take a look at Problem 5 of assignment 1 to see the converse of this statement does not necessarily hold.

Second, observe that spectral thinness is very easy to calculate (computationally or analytically), it is just the maximum eigenvalue of the symmetric matrix

$$L_G^{\dagger/2} L_T L_G^{\dagger/2}.$$

Recall that $L_G^\dagger$ is the pseudo-inverse of L_G and $L_G^{\dagger/2}$ is the square root of the PSD matrix $L_G^\dagger$ (see lectures 1-2 for the definition of pseudo-inverse).

This definition has its own drawbacks. A k-connected graph (for arbitrary large k) does not necessarily have a spectrally thin tree. We will talk more about this in the future lectures. Right now, we are seeking for a very simpler question. Suppose instead of looking for a tree we are just interested in finding an $\Omega(n)$ set of edges of G which is $O(1/k)$-thin. The goal of this lecture is to use spectral techniques to answer this question. In particular, we will find a set F which is a forest such that $|F| \geq \Omega(n)$ and that

$$L_F \preceq L_G/k.$$

It is a very interesting open problem to find a linear sized $O(1/k)$-thin set of edges in a k-edge connected graph using randomized rounding (without appealing to spectral techniques).

13.2 Searching for Thin Linear size Forests

In this lecture we design a simple algorithm that finds a linear size thin forest in a k-edge connected graph.

Theorem 13.4. *For any (unweighted) k-edge connected graph $G = (V, E)$, there is a polynomial time algorithm that finds a forest $F \subseteq E$ such that $|F| \geq \Omega(n)$ and*

$$L_F \preceq O(L_G/k).$$

We start by normalizing b_e vectors and putting them in an isotropic position. We have already studied this in the first lecture, we just need to work with vectors $y_e = L_G^{\dagger/2} b_e$. The main motivation for this normalization is to classify small normed vectors (after normalization) as easy to handle edges. Note that although all edges have the same norm before the normalization that is not necessarily true after the normalization. In fact, after the normalization for any edge e, $\|y_e\|^2 = \text{Reff}(e)$. It is easy to see that if F has an edge of norm $\|y_e\|^2 \approx 1$, then $L_F \not\preceq L_G/2$. We leave this as an exercise. Roughly speaking, we should never pick big edges with respect to isotropic position in a spectrally thin subgraph.

So, we can reformulate the theorem as finding a linear size forest $F \subseteq E$ such that

$$\|\sum_{e\in F} y_e y_e^\intercal\| \le O(1/k).$$

Note that if the above equation holds, then we have

$$L_G^{\dagger/2} L_F L_G^{\dagger/2} \preceq O(1/k)I,$$

or equivalently, $L_F \preceq O(L_G/k)$.

In the rest of this lecture we prove the following stronger theorem.

Theorem 13.5. *Given a set of vectors $v_1, \ldots, v_m \in \mathbb{R}^n$ such that*

$$\sum_{i=1}^{m} v_i v_i^\intercal \preceq I.$$

If the set contains k disjoint bases then there is a set F of linearly independent vectors such that $|F| \ge \Omega(n)$ and

$$\|\sum_{i\in F} v_i v_i^\intercal\| \le O(1/k).$$

It is easy to see that Theorem 13.4 follows directly from the above theorem. So, in the rest of this lecture we prove the above theorem.

In addition, there are examples where for linearly independent set of vectors F with $|F| > n - n/k$, the above norm is 1. So, in a sense the statement of the above theorem is tight.

Lemma 13.6. *There is a set of vectors $v_1, \ldots, v_m \in \mathbb{R}^n$ that form k disjoint bases such that*

$$\sum_{i=1}^{m} v_i v_i^\intercal \preceq I$$

and for any set F of linearly independent vectors where $|F| > n - n/k$,

$$\|\sum_{i\in F} v_i v_i^\intercal\| \ge \Omega(1).$$

We leave the proof of the above lemma as an exercise.

13.3 Main Proof

First of all, it is instructive to see that a random sample does not necessarily work, i.e., if we choose each vector with probability $1/k$ the sample may have norm 1 with probability very close to 1.

The basic idea is to start with $F = \{\}$ and iteratively add a "good" vector to F while maintaining that $A = \sum_{i\in F} v_i v_i^\intercal$ has a small norm. So, we use an inductive argument. Perhaps, the most interesting idea in this work is to choose the right induction hypothesis. First, assume we use $\|A\|$ as the potential function (or the induction hypothesis). The problem with this is that

$$\|A + vv^\intercal\| = \|A\| + \|v\|$$

in the worst case. Note that we have a degree of freedom in choosing the right v. But in the worst case it could be that all of the eigenvalues of A are equal to $\|A\|$ so any v will shift the norm of A by $\|v\|$.

So, perhaps a better idea is to assume a bound on all of the eigenvalues of A, say we use $\mathrm{Tr}(A)$ as the potential function. In this case it is actually true that $\mathrm{Tr}(A)$ only shifts very little $\langle v, v\rangle$ when we add a vector v. Unfortunately, assuming that $\mathrm{Tr}(A)$ is small, say $O(n)$, we can not conclude that the maximum eigenvalue is small.

The idea of Batson, Spielman and Srivastava [BSS14] is to use the following $\alpha_{\max}(.)$ function, say $\lambda_1 \geq \cdots \geq \lambda_n$ are the eigenvalues of A,

$$\alpha_{\max}(A) = \max\left\{t : \frac{1}{t-\lambda_1} + \cdots + \frac{1}{t-\lambda_n} = \alpha\right\}.$$

Note that

$$\sum_{i=1}^{n} \frac{1}{t-\lambda_i} = \mathrm{Tr}((tI-A)^{-1}).$$

There are many interpretations of the above function and we will see generalizations in the future lectures. Another equivalent definition is the following

$$\mathrm{Tr}((tI-A)^{-1}) = \frac{\frac{\mathrm{d}}{\mathrm{d}t}\chi[A](t)}{\chi[A](t)} = \frac{\mathrm{d}}{\mathrm{d}t}\log\chi[A](t).$$

We will prove this in the next lecture. Perhaps the main source of inspiration is by studying the barrier functions used in the interior point method. This is out of the scope of this course.

To understand $\alpha_{\max}$ it is instructive to plugin different values of α, if $\alpha \approx \infty$, then $\alpha_{\max}(A)$ is the largest eigenvalue of A. On the other hand if $\alpha \approx 0$, then $\alpha_{\max}(A)$ is just n/t; in the latter case it is very easy to see that $\alpha_{\max}(A)$ changes very slowly when we apply a rank 1 update. The idea is by choosing the right α we want to have the best of the both worlds, on one hand we want to have a good approximation of the largest eigenvalue of A and on the other hand we want to be able to say that $\alpha_{\max}(A)$ changes slowly when we apply a rank 1 update.

To show that $\alpha_{\max}(A)$ is a good potential function we need to prove two properties.

i) Any upper bound on $\alpha_{\max}(A)$ gives and upper bound on $\|A\|$.

ii) $\alpha_{\max}(A)$ increases slowly when we apply a rank 1 update operation.

These two properties are proved in the following two claims.

Claim 13.7. *For any symmetric matrix A, $\alpha_{\max}(A) \geq \|A\|$.*

Proof. Let $\lambda_1 \geq \cdots \geq \lambda_n$ be the eigenvalues of A. For $t = \lambda_1$,

$$\frac{1}{t-\lambda_1} + \cdots + \frac{1}{t-\lambda_n}$$

is infinity and for $t > \lambda_1$ the above sum is a decreasing continuous function of t. Therefore, $\alpha_{\max}(A) > \lambda_1$. □

Next, we want to prove (ii). Ideally, we want to prove that for a vector v_i,

$$\alpha_{\max}(A + v_i v_i^{\intercal}) \leq \alpha_{\max}(A) + O(1/kn).$$

Note that if the above equation holds we can add $\Omega(n)$ vectors to F while making sure that $\|A\| \leq \alpha_{\max}(A) \leq O(1/k)$ is small. But, of course the above equation can not hold for any arbitrary vector v_i because v_i can have the same direction as the largest eigenvector of A. Instead we use an averaging argument, we say if we have $\Theta(kn)$ vectors such that the sum of their quadratic form is at most identity, one of the satisfies the above equation.

Lemma 13.8. *Given a symmetric matrix $A \in \mathbb{R}^{n\times n}$, and a set of vectors $v_1, \ldots, v_m$ such that*

$$\sum_{i=1}^{m} v_i v_i^\intercal \preceq I,$$

For any bounded $\alpha > 0$, there exists $1 \leq i \leq m$ such that

$$\alpha_{\max}(A + v_i v_i^\intercal) \leq \alpha_{\max}(A) + \frac{1}{m/2 - \alpha}.$$

Suppose $\alpha_{\max}(A) = t$. It is enough to show that there is a vector v_i such that for any $\delta \geq \frac{1}{m/2-\alpha}$,

$$\mathrm{Tr}(((t+\delta)I - A - v_i v_i^\intercal)^{-1}) \leq \mathrm{Tr}((tI - A)^{-1}). \tag{13.1}$$

To analyze the left hand side we use the Sherman-Morrison Formula.

Theorem 13.9. *For any invertible matrix $A \in \mathbb{R}^{n\times n}$ and any vector $v \in \mathbb{R}^n$, $(A - vv^\intercal)^{-1}$ exists if and only if $1 - v^\intercal A^{-1} v \neq 0$, and in this case*

$$(A - vv^\intercal)^{-1} = A^{-1} + \frac{A^{-1} v v^\intercal A^{-1}}{1 - v^\intercal A^{-1} v}$$

Let us use $B = (\delta + t)I - A$ for the brevity of notation. Note that $B \succ 0$ by definition. We can rewrite (13.1) as follows:

$$\mathrm{Tr}\left(\frac{B^{-1} v v^\intercal B^{-1}}{1 - v_i^\intercal B^{-1} v}\right) \leq \mathrm{Tr}((B - \delta I)^{-1}) - \mathrm{Tr}(B^{-1}) \tag{13.2}$$

First, we show the RHS is at most $\delta \cdot \mathrm{Tr}((\delta I + B)^{-2})$. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of B. Then,

$$\begin{aligned} \mathrm{Tr}((B - \delta I)^{-1}) - \mathrm{Tr}(B^{-1}) &= \sum_{i=1}^{n} \frac{1}{t - \lambda_i} - \frac{1}{t + \delta - \lambda_i} \\ &\geq \sum_{i=1}^{n} \frac{\delta}{(t + \delta - \lambda_i)^2} = \delta \cdot \mathrm{Tr}(B^{-2}). \end{aligned}$$

Rewriting (13.2) and using $\mathrm{Tr}(XY) = \mathrm{Tr}(YX)$, all we need to show is that there is a vector v_i such that

$$\frac{v_i^\intercal B^{-2} v_i}{1 - v_i^\intercal B^{-1} v_i} \leq \delta \cdot \mathrm{Tr}(B^{-2}). \tag{13.3}$$

Now, let's see how does the above equation look like for a random v_i. First, observe that for any invertible

C,

$$\begin{aligned}
\mathbb{E}\left[v_i^\intercal C v_i\right] &= \sum_{i=1}^{m} \frac{1}{m} \operatorname{Tr}(v_i v_i^T C) \\
&= \operatorname{Tr}\left(\sum_{i=1}^{m} \frac{1}{m} v_i v_i^\intercal C\right) \\
&= \operatorname{Tr}\left(C^{1/2} \sum_{i=1}^{m} \frac{1}{m} v_i v_i^\intercal C^{1/2}\right) \\
&\leq \frac{1}{m} \operatorname{Tr}(C).
\end{aligned}$$

The inequality uses that

$$C^{1/2} \sum_{i=1}^{m} v_i v_i^\intercal C^{1/2} \preceq C^{1/2} I C^{1/2} = C.$$

See Lemma 13.10 below for the proof of the PSD inequality. By Markov's inequality exists $1 \leq i \leq m$ such that

$$\begin{aligned}
v_i B^{-2} v_i &\leq \frac{2\operatorname{Tr}(B^{-2})}{m}, \\
v_i B^{-1} v_i &\leq \frac{2\operatorname{Tr}(B^{-1})}{m} \leq \frac{2\operatorname{Tr}((B-\delta I)^{-1})}{m} = \frac{2\alpha}{m}.
\end{aligned}$$

So,

$$\frac{v_i^\intercal B^{-2} v_i}{1 - v_i^\intercal B^{-1} v_i} \leq \frac{2\operatorname{Tr}(B^{-2})/m}{1 - 2\alpha/m} \leq \delta \cdot \operatorname{Tr}(B^{-2}).$$

So, it is enough to have $\delta \geq \frac{1}{m/2-\alpha}$. This completes the proof of Lemma 13.8.

We just prove Lemma 13.10 below.

Lemma 13.10. *For any three symmetric matrices $A, B, C \in \mathbb{R}^{n\times n}$, if $A \preceq B$, then*

$$CAC \preceq CBC.$$

Proof. For any test vector x, and $y = Cx$,

$$x^\intercal CACx = y^\intercal Ay \leq y^\intercal By = x^\intercal CBCx,$$

where the inequality follows by $A \preceq B$. □

To prove Theorem 13.5 it is enough to repeatedly apply the above lemma.

Proof of Theorem 13.5. Let $\alpha = m/4$. We start with $A_0 = 0$. Then,

$$\alpha = \operatorname{Tr}((tI - A)^{-1}) = \frac{n}{t},$$

so let $\alpha_{\max}(A_0) = 4n/m$. We apply Lemma 13.8 $n/2$ times. Let F_j be the set of vectors that we have chosen in the first j iterations and let $A_j = \sum_{i\in F_j} v_i v_i^T$.

Now, let $S \subseteq \{v_1, \ldots, v_m\}$ be all of the vectors which are linearly independent of the vectors in F_j. It is easy to see that

$$|S| \geq k(n - |F_j|) = k(n - j) \geq kn/2.$$

This is because the given set of vectors $v_1, \dots, v_m$ include k disjoint bases and each of these bases have at most $|F_j|$ vectors that are linearly dependent to the vectors in F_j.

Applying Lemma 13.8 to S there exists a vector v_i such that

$$\alpha_{\max}(A_j + v_i v_i^\intercal) \le \alpha_{\max}(A_j) + \frac{4}{|S|}.$$

We let $F_{j+1} = F_j \cup \{v_i\}$, $A_{j+1} = A_j + v_i v_i^\intercal$ and we recurse.

After $n/2$ iterations we get a set $F_{n/2}$ of linearly independent vectors such that

$$\alpha_{\max}(A_{n/2}) \le \frac{4n}{m} + \sum_{j=0}^{n/2-1} \frac{4}{k(n-j)} \le \frac{16}{k}.$$

In the first inequality we also used $m \ge nk$. □

Batson, Spielman and Srivastava [BSS14] used $\alpha_{\max}$ function together with the following $\alpha_{\min}$ and argued that any graph has a linear size weighted spectral sparsifier. For a matrix A,

$$\alpha_{\min}(A) = \min\left\{t : \frac{1}{\lambda_1 - t} + \dots + \frac{1}{\lambda_n - t} = \alpha\right\}$$

They show that there is always a vector (with a suitable weight) such that a rank 1 update of A with that vector increases $\alpha_{\max}, \alpha_{\min}$ by almost similar amounts. Therefore, after $O(n/\epsilon^2)$ steps we get a set F of $O(n/\epsilon^2)$ vectors where the corresponding $\alpha_{\max} \le (1+\epsilon)\alpha_{\min}$. This implies that the quadratic form associated with F very well approximates the identity matrix, and in the graph language, they correspond to a very well spectral sparsifier.

References

[ALN08] S. Arora, J. R. Lee, and A. Naor. "Euclidean distortion and the Sparsest Cut". In: *J. Amer. Math. Soc.* 21.1 (2008), pp. 1–21 (cit. on p. 13-2).

[AO14] N. Anari and S. Oveis Gharan. "Effective-Resistance-Reducing Flows and Asymmetric TSP". 2014. URL: http://arxiv.org/abs/1411.4613 (cit. on p. 13-1).

[BSS14] J. D. Batson, D. A. Spielman, and N. Srivastava. "Twice-Ramanujan Sparsifiers". In: *SIAM Review* 56.2 (2014), pp. 315–334 (cit. on pp. 13-1, 13-4, 13-7).

[Har13] N. Harvey. *Lecture 16 of Sparse Approximations Course*. 2013. URL: http://www.cs.ubc.ca/~nickhar/W13/Scribe/Lecture16Scribe.pdf (cit. on p. 13-1).

[KS59] R. V. Kadison and I. M. Singer. "Extensions of Pure States". In: *American Journal of Mathematics* 81.2 (1959), pp. 383–400 (cit. on p. 13-1).

[MSS13] A. Marcus, D. A. Spielman, and N. Srivastava. "Interlacing Families II: Mixed Characteristic Polynomials and the Kadison-Singer Problem". 2013. URL: http://arxiv.org/abs/1306.3969 (cit. on p. 13-1).