

Lecture 22

*Lecturer: David P. Williamson**Scribe: Shijin Rajakrishnan*

1 Recap

In the previous lectures, we explored the problem of finding spectral sparsifiers and saw both deterministic and randomized algorithms that given a graph G with n vertices finds a spectral sparsifier with $O\left(\frac{n \log n}{\epsilon^2}\right)$ edges. In this lecture, we improve upon this result to find linear-sized spectral sparsifiers, or more concretely, spectral sparsifiers with $O\left(\frac{n}{\epsilon^2}\right)$ edges.

This is an interesting result since we noted in the past that a spectral sparsifier is a generalization of the cut sparsifier of a graph, and the best results known previous to this one for cut sparsifiers were that a cut sparsifier with $n \log^{O(1)} n / \epsilon^2$ edges can be found in nearly linear time [1].

Recall that given a graph $G = (V, E)$, a weighted graph $H = (V, E')$ with weights $w(i, j)$ is an ϵ -spectral sparsifier of G if

$$(1 - \epsilon)L_G \preceq L_H \preceq (1 + \epsilon)L_G.$$

In the last lecture, we showed that

$$(1 - \epsilon)L_G \preceq \sum_{(i,j) \in E} w(i, j)(\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T \preceq (1 + \epsilon)L_G$$

if and only if

$$(1 - \epsilon)I \preceq \sum_{(i,j) \in E} w(i, j)\mathbf{x}_{ij}\mathbf{x}_{ij}^T \preceq (1 + \epsilon)I,$$

where the vectors $\mathbf{x}_{ij} = L_G^{\frac{1}{2}}(\mathbf{e}_i - \mathbf{e}_j)$, and that $\sum_{(i,j) \in E} \mathbf{x}_{ij}\mathbf{x}_{ij}^T = I^*$ (recall that I^* is our continuing fudge of an identity matrix, which is actually $L_G L_G^\dagger$; for any vector \mathbf{v} such that $\mathbf{v}^T \mathbf{e} = 0$, $I^* \mathbf{v} = \mathbf{v}$).

Our goal is, given vectors $\mathbf{v}_1, \dots, \mathbf{v}_m$ such that $\sum_{i=1}^m \mathbf{v}_i \mathbf{v}_i^T = I$, to show that there exists some weights $w_i \geq 0$ and a subset $S \subseteq [m]$ s.t. $|S| \leq \lceil \frac{n}{\epsilon^2} \rceil$ and

$$(1 - \epsilon)^2 I \preceq \sum_{i \in S} w_i \mathbf{v}_i \mathbf{v}_i^T \preceq (1 + \epsilon)^2 I.$$

⁰This lecture is derived from Spielman 2015, Lecture 18, <http://www.cs.yale.edu/homes/spielman/561/lect18-15.pdf>.

Definition 1 The vectors \mathbf{v}_i are said to be in isotropic position if

$$\sum_{i=1}^m \mathbf{v}_i \mathbf{v}_i^T = I.$$

Note that for vectors \mathbf{v}_i in isotropic position, for any A ,

$$\sum_{i=1}^m \mathbf{v}_i^T A \mathbf{v}_i = \sum_{i=1}^m (\mathbf{v}_i \mathbf{v}_i^T) \cdot A = \left(\sum_{i=1}^m \mathbf{v}_i \mathbf{v}_i^T \right) \cdot A = I \cdot A = \text{tr}(A).$$

In this lecture, we prove a weaker version of the theorem first and then mention how to extend it to the general case. The version we prove is as follows.

Theorem 1 Given vectors $\mathbf{v}_1, \dots, \mathbf{v}_m$ that are in isotropic position, we can find a subset $S \subseteq [m]$, and weights $w_i \geq 0$ such that $|S| \leq 6n$ and

$$\frac{1}{\sqrt{13}} I \preceq \frac{1}{\sqrt{13n}} \sum_{i \in S} w_i \mathbf{v}_i \mathbf{v}_i^T \preceq \sqrt{13} I.$$

2 The Algorithm

The basic idea behind the algorithm is that we greedily pick vectors and weights such that we control how the maximum and minimum eigenvalues change.

Algorithm 1: Linear sized spectral sparsifier

```

 $l \leftarrow -n; \quad u \leftarrow n$ 
 $\Delta l \leftarrow 1/3; \quad \Delta u \leftarrow 2$ 
 $w_i \leftarrow 0 \quad \forall i$ 
 $A \leftarrow 0$ 
for  $k \leftarrow 1$  to  $6n$  do
    Pick  $\mathbf{v}_i, c$  s.t.  $\lambda_{\min}(A + c\mathbf{v}_i \mathbf{v}_i^T) \geq l + \Delta l$  and  $\lambda_{\max}(A + c\mathbf{v}_i \mathbf{v}_i^T) \leq u + \Delta u$ 
     $w_i \leftarrow w_i + c;$ 
     $A \leftarrow A + c\mathbf{v}_i \mathbf{v}_i^T$ 
     $l \leftarrow l + \Delta l$ 
     $u \leftarrow u + \Delta u$ 
end
return  $A$ 

```

Note that initially $\lambda_{\min}(A) = \lambda_{\max}(A) = 0$, which satisfies $l = -n, u = n$ as bounds. At the end of the algorithm, $\lambda_{\min}(A) \geq -n + (6n)\frac{1}{3} = n$, and $\lambda_{\max}(A) \leq n + (6n)(2) = 13n$. So we divide the matrix by $\sqrt{13n}$ to get the bound as given in the theorem.

3 Barrier Functions

The genius of the result is that rather than working with λ_{\min} and λ_{\max} directly, we use barrier functions, $L(l, A)$ and $U(u, A)$, which we now define.¹ We let

$$L(l, A) = \sum_{i=1}^m \frac{1}{\lambda_i - l} = \text{tr}((A - lI)^{-1}),$$

and

$$U(u, A) = \sum_{i=1}^m \frac{1}{u - \lambda_i} = \text{tr}((uI - A)^{-1}),$$

for $l \leq \lambda_1 \leq \dots \leq \lambda_n \leq u$ where the λ_i are the eigenvalues of A .

Initially, $l_0 \leq \lambda_{\min}(A) \leq \lambda_{\max}(A) \leq u_0$. This implies that $U(u, A)$ and $L(l, A)$ are both positive and bounded. During the course of the algorithm, we pick vectors and increase l and u so that the barrier functions don't increase in value, and so we can ensure that the eigenvalues lie in the range $[l, u]$.

Initially $U(u, 0) = \sum_{i=1}^n \frac{1}{u} = 1$ and $L(-n, 0) = \sum_{i=1}^n \frac{1}{-n} = 1$.

What we want in each iteration is to find a vector \mathbf{v} and a weight c such that

$$U(u + \Delta u, A + c\mathbf{v}\mathbf{v}^T) \leq U(u, A) \leq 1$$

and

$$L(l + \Delta l, A + c\mathbf{v}\mathbf{v}^T) \leq L(l, A) \leq 1,$$

so that the barrier functions do not increase as we update A, u, l through the run of the algorithm.

4 The Analysis

We now turn towards proving that such a selection of \mathbf{v} and c can be made in each iteration of the loop. We will show that there exists matrices U_A and L_A such that the following two lemmas hold.

Lemma 2 *The barrier functions do not increase in an iteration; that is,*

$$U(u + \Delta u, A + c\mathbf{v}\mathbf{v}^T) \leq U(u, A)$$

and

$$L(l + \Delta l, A + c\mathbf{v}\mathbf{v}^T) \leq L(l, A)$$

holds for c, \mathbf{v} , if

$$\mathbf{v}^T U_A \mathbf{v} \leq \frac{1}{c} \leq \mathbf{v}^T L_A \mathbf{v}.$$

¹In the original paper due to Batson, Spielman, and Srivastava, as well as other notes on their result, the notation $\Phi^u(A)$ is used for $U(u, A)$ and $\Phi_l(A)$ is used for $L(l, A)$.

Lemma 3

$$\sum_{i=1}^m \mathbf{v}_i^T U_A \mathbf{v}_i \leq \frac{1}{\Delta u} + U(u, A) \leq \frac{1}{\Delta l} - \frac{1}{1/L(l, A) - \Delta l} \leq \sum_{i=1}^m \mathbf{v}_i^T L_A \mathbf{v}_i.$$

From the choice of our parameters, note that we get

$$\frac{1}{\Delta u} + U(u, A) \leq \frac{1}{2} + 1 = \frac{3}{2},$$

and

$$\frac{1}{\Delta l} - \frac{1}{1/L(l, A) - \Delta l} \geq 3 - \frac{1}{1 - 1/3} = \frac{3}{2}.$$

From Lemma 3, we see that this implies

$$\sum_{i=1}^m v_i^T U_A v_i \leq \frac{3}{2} \leq \sum_{i=1}^m v_i^T L_A v_i,$$

and therefore, there exists some c, \mathbf{v}_i such that

$$\mathbf{v}_i^T U_A \mathbf{v}_i \leq \frac{1}{c} \leq \mathbf{v}_i^T L_A \mathbf{v}_i.$$

Then Lemma 2 ensures that there exists a vector \mathbf{v}_i and weight c so that we can add $c\mathbf{v}_i\mathbf{v}_i^T$ to A without increasing the barrier functions.

The only remaining part is to prove the two lemmata.

To prove Lemma 2, we first analyze what the addition of the vector $c\mathbf{v}\mathbf{v}^T$ does to the matrix A , by using the following formula.

Theorem 4 (Sherman-Morrison formula) *For a nonsingular symmetric matrix X and a vector \mathbf{v}*

$$(X - \mathbf{v}\mathbf{v}^T)^{-1} = X^{-1} + \frac{X^{-1}\mathbf{v}\mathbf{v}^T X^{-1}}{1 - \mathbf{v}^T X^{-1} \mathbf{v}}.$$

The formula expresses a rank-1 update to the inverse of a matrix.

Proof of Lemma 2:

On adding the matrix $c\mathbf{v}\mathbf{v}^T$ to A , the barrier function changes to

$$\begin{aligned} U(u, A + c\mathbf{v}\mathbf{v}^T) &= \text{tr}((uI - (A + c\mathbf{v}\mathbf{v}^T))^{-1}) \\ &= \text{tr}((uI - A - c\mathbf{v}\mathbf{v}^T)^{-1}) \\ &= \text{tr}((uI - A)^{-1}) + \frac{c \cdot \text{tr}((uI - A)^{-1} \mathbf{v}\mathbf{v}^T (uI - A)^{-1})}{1 - c\mathbf{v}^T (uI - A)^{-1} \mathbf{v}} \\ &= U(u, A) + \frac{c \cdot \text{tr}((uI - A)^{-1} \mathbf{v}\mathbf{v}^T (uI - A)^{-1})}{1 - c\mathbf{v}^T (uI - A)^{-1} \mathbf{v}} \\ &= U(u, A) + c \frac{\text{tr}(\mathbf{v}^T (uI - A)^{-2} \mathbf{v})}{1 - c\mathbf{v}^T (uI - A)^{-1} \mathbf{v}} \\ &= U(u, A) + c \frac{\mathbf{v}^T (uI - A)^{-2} \mathbf{v}}{1 - c\mathbf{v}^T (uI - A)^{-1} \mathbf{v}} \end{aligned}$$

The third inequality follows from the Sherman-Morrison formula, the fourth by the definition of $U(u, A)$, and the fifth by the cyclic property of the trace.

So as we add $c\mathbf{v}\mathbf{v}^T$, the barrier function increases. To counteract this, we increase the value of u so to keep the barrier function constant. Let $\hat{u} \equiv u + \Delta u$. Then we want to see under what values of c and v the barrier function doesn't increase. So then

$$\begin{aligned} U(\hat{u}, A + c\mathbf{v}\mathbf{v}^T) &= U(\hat{u}, A) + c \frac{\mathbf{v}^T(\hat{u}I - A)^{-2}\mathbf{v}}{1 - c\mathbf{v}^T(\hat{u}I - A)^{-1}\mathbf{v}} \\ &= U(u, A) - (U(u, A) - U(\hat{u}, A)) + c \frac{\mathbf{v}^T(\hat{u}I - A)^{-2}\mathbf{v}}{1 - c\mathbf{v}^T(\hat{u}I - A)^{-1}\mathbf{v}} \end{aligned}$$

We want $U(\hat{u}, A + c\mathbf{v}\mathbf{v}^T) \leq U(u, A)$, which will be true if

$$U(u, A) - U(\hat{u}, A) \geq c \frac{\mathbf{v}^T(\hat{u}I - A)^{-2}\mathbf{v}}{1 - c\mathbf{v}^T(\hat{u}I - A)^{-1}\mathbf{v}},$$

which holds if

$$\frac{1}{c} \geq \frac{\mathbf{v}^T(\hat{u}I - A)^{-2}\mathbf{v}}{U(u, A) - U(\hat{u}, A)} + \mathbf{v}^T(\hat{u}I - A)^{-1}\mathbf{v} = \mathbf{v}^T U_A \mathbf{v},$$

where we define

$$U_A \equiv \frac{(\hat{u}I - A)^{-2}}{U(u, A) - U(\hat{u}, A)} + (\hat{u}I - A)^{-1}.$$

This proves one half of the Lemma. The other half can be proved similarly, but with $\hat{l} = l + \Delta l$, and

$$L_A = \frac{(A - \hat{l}I)^{-2}}{L(\hat{l}, A) - L(l, A)} - (A - \hat{l}I)^{-1}.$$

□

We now turn to the proof of the second Lemma.

Proof of Lemma 3, upper bound: Notice that if $u > \lambda_{\max}(A)$, $\delta > 0$, then the barrier function is decreasing in u , ie, $U(u + \delta, A) < U(u, A)$.

Now, since the vectors \mathbf{v}_i are in isotropic position,

$$\begin{aligned} \sum_{i=1}^m \mathbf{v}_i^T U_A \mathbf{v}_i &= \text{tr}(U_A) \\ &= \frac{\text{tr}((\hat{u}I - A)^{-2})}{U(u, A) - U(\hat{u}, A)} + \text{tr}((\hat{u}I - A)^{-1}) \\ &\leq \frac{\text{tr}((\hat{u}I - A)^{-2})}{U(u, A) - U(\hat{u}, A)} + U(u, A). \end{aligned}$$

To bound the first term, we note that

$$\frac{d}{du}U(u, A) = \frac{d}{du} \sum_{i=1}^n \frac{1}{u - \lambda_i} = - \sum_{i=1}^n \frac{1}{(u - \lambda_i)^2} = -\text{tr}((uI - A)^{-2}),$$

and

$$\frac{d^2}{du^2}U(u, A) = 2 \sum_{i=1}^n \frac{1}{(u - \lambda_i)^3} > 0,$$

and thus $U(u, A)$ is decreasing and convex in u .

In particular, convexity implies that

$$U(u, A) - U(\hat{u}, A) \geq (-\Delta u) \frac{d}{du}U(\hat{u}, A) = \Delta u \cdot \text{tr}((\hat{u}I - A)^{-2}),$$

which yields

$$\frac{\text{tr}((uI - A)^{-2})}{U(u, A) - U(\hat{u}, A)} \leq \frac{1}{\Delta u},$$

leading to

$$\frac{\text{tr}((\hat{u}I - A)^{-2})}{U(u, A) - U(\hat{u}, A)} \leq \frac{1}{\Delta u}.$$

□

We can prove the lower bound for Lemma 3 in a similar way.

Proof of Lemma 3, lower bound:

In this case, we have

$$\begin{aligned} \sum_{i=1}^m \mathbf{v}_i^T L_A \mathbf{v}_i &= \text{tr}(L_A) \\ &= \frac{\text{tr}((A - \hat{l}I)^{-2})}{L(\hat{l}, A) - L(l, A)} - \text{tr}((A - \hat{l}I)^{-1}) \\ &= \frac{\text{tr}((A - \hat{l}I)^{-2})}{L(\hat{l}, A) - L(l, A)} - L(\hat{l}, A). \end{aligned}$$

Now,

$$\frac{d}{dl}L(l, A) = \frac{d}{dl} \sum_{i=1}^n \frac{1}{\lambda_i - l} = \sum_{i=1}^n \frac{1}{(\lambda_i - l)^2} = \text{tr}((A - lI)^{-2}),$$

and

$$\frac{d^2}{dl^2}L(l, A) = 2 \sum_{i=1}^n \frac{1}{(\lambda_i - l)^3} > 0,$$

and thus $L(l, A)$ is increasing and convex in l . Then by convexity,

$$L(l, A) - L(\hat{l}, A) \geq -\Delta l \frac{d}{dl}L(\hat{l}, A) \geq -\Delta l \text{tr}((A - \hat{l}I)^{-2}).$$

Rearranging terms, we get

$$\frac{\text{tr}((A - \hat{l}I)^{-2})}{L(\hat{l}, A) - L(l, A)} \geq \frac{1}{\Delta l}.$$

To bound the second term, we claim that

$$\frac{L(\hat{l}, A) - L(l, A)}{\Delta l} \leq L(l, A)L(\hat{l}, A).$$

If the claim is true, then by rearranging terms we get

$$L(\hat{l}, A) \leq \frac{1}{1/L(l, A) - \Delta l},$$

as desired. To prove the claim we observe that

$$\begin{aligned} \frac{L(\hat{l}, A) - L(l, A)}{\Delta l} &= \frac{1}{\Delta l} \sum_{i=1}^n \left[\frac{1}{\lambda_i - \hat{l}} - \frac{1}{\lambda_i - l} \right] \\ &= \frac{1}{\Delta l} \sum_{i=1}^n \left[\frac{\lambda_i - l - (\lambda_i - \hat{l})}{(\lambda_i - l)(\lambda_i - \hat{l})} \right] \\ &= \frac{1}{\Delta l} \sum_{i=1}^n \left[\frac{\Delta l}{(\lambda_i - l)(\lambda_i - \hat{l})} \right] \\ &= \sum_{i=1}^n \left[\frac{1}{(\lambda_i - l)(\lambda_i - \hat{l})} \right] \\ &\leq \left[\sum_{i=1}^n \frac{1}{\lambda_i - l} \right] \left[\sum_{i=1}^n \frac{1}{\lambda_i - \hat{l}} \right] = L(l, A)L(\hat{l}, A), \end{aligned}$$

as long as $l \leq \lambda_{\min}$, as we have been guaranteeing. \square

To get a stronger (more general) result, we change the parameters $\Delta u, \Delta l$, and then prove

$$\text{tr}(L_A) \geq \frac{1}{\Delta l} - L(l, A)$$

instead of

$$\text{tr}(L_A) \geq \frac{1}{\Delta l} - \frac{1}{1/L(l, A) - \Delta l}.$$

The proof of this is messier and involves more algebra, but is not really bad.

Having seen the correctness of the algorithm, we turn towards the question of its runtime. The general usage of a spectral sparsifier is to reduce the dependence of the runtime of algorithms that depend on the number of edges by creating a sparse approximation. The algorithm described runs far too slowly to realize gains using a sparsifier for cut algorithms (for instance). The faster algorithm due to Lee and Sun 2015, constructs a sparsifier with $O(\frac{qn}{\epsilon^2})$ edges in $\tilde{O}(qmn^{\frac{1}{2}})$ time.

References

- [1] Wai Shing Fung, Ramesh Hariharan, Nicholas JA Harvey, and Debmalya Panigrahi, *A General Framework for Graph Sparsification*. Proceedings of the Forty-Third Annual ACM Symposium on the Theory of Computing, pp 71-80, 2011.