CS 860 Spectral graph theory, Spring 2019, Waterloo.

Lecture 12: Linear-sized spectral sparsification

We study a polynomial time algorithm to construct a linear-sized spectral sparsifier using barrier functions.

Results

The main theorem that we study today is by Batson, Spielman and Srivastava.

Theorem [BSS] For any graph G, there is a $(1\pm\varepsilon)$-spectral sparsifier H with $O(\frac{n}{\varepsilon^2})$ edges.

This is an amazing result, especially when it was not know that linear-sized $(1\pm\varepsilon)$-cut sparsifiers exist.

As we have seen last time, the spectral sparsification problem can be reduced to the following linear algebraic statement.

Theorem [BSS] Suppose $\sum_{i=1}^{m} v_i v_i^T = I_n$ where $v_i \in \mathbb{R}^n$. There exist $w_i \in \mathbb{R}$ with at most dn non-zeros such that $(1-\frac{1}{\sqrt{d}})^2 I_n \preccurlyeq \sum_{i=1}^{m} w_i v_i v_i^T \preccurlyeq (1+\frac{1}{\sqrt{d}})^2 I_n$.

Random Sampling

Last time we used random sampling to construct a spectral sparsifier. here we explain that it will not work to give linear-sized sparsifiers.

Consider the simple example where $\sum_{i=1}^{n} e_i e_i^T = I_n$ where e_i is the i-th standard unit vector.

Since all vectors are of the same length, the sampling probabilities are the same, i.e. $p_i = \frac{1}{n}$ for $1 \le i \le n$.

Now, following L11, if we sample $k = cn$ iterations and set $w_i := w_i + \frac{1}{c}$ for some constant $c \geq 1$.

The resulting solution is equivalent to throwing cn balls into n bins.

It is well-known that if cn balls are thrown, the maximum load is $\Omega(\log n / \log\log n)$ while there is a constant fraction of empty bins.

So, the resulting solution is very unbalanced and not a good approximation to the identity at all.

Potential functions

The approach by Batson, Spielman and Srivastava is deterministic.

The algorithm adds one edge at a time, and uses a potential function to guide the choice of the next edge.

The beauty of the algebraic formulation is that we only have two parameters to keep track of, the maximum eigenvalue and the minimum eigenvalue of our solution.

We add one edge by one edge. Let A be the current matrix so far.

One natural attempt is to bound the maximum eigenvalue $\lambda_{max}(A)$, and make sure that it won't increase too quickly, e.g. to prove inductively $\lambda_{max}(A_{i+1}) \le \lambda_{max}(A_i) + \varepsilon$.

This way of measuring progress does not work well for this problem, as the matrix is n-dimensional, just focusing on the maximum direction can't distinguish the case where every direction is large or where one direction is large and all other (orthogonal) directions are small.

Ideally, we hope to say something like after n iterations, every direction is increased by one unit (think of the balls and bins example).

To prove it inductively, we want a potential function to let us argue that the maximum direction is increased by $\frac{1}{n}$ unit per edge on average, but the maximum eigenvalue won't do this for us.

So, we would like a more global quantity that will take into consideration of all directions.

One possible parameter of this kind is $Tr(A)/n$, the average eigenvalue. For this, we can easily argue that it increases slowly, but the problem is that we cannot conclude that the maximum eigenvalue is small by using that the average is small.

The key idea in [BSS] is to use the potential function $\Phi(A) = Tr(uI-A)^{-1} = \sum_{i=1}^{n} \frac{1}{u-\lambda_i}$, where u is intended to be an upper bound of the maximum eigenvalue and $\lambda_1, \dots, \lambda_n$ are the eigenvalues of A.

They maintain the invariant that $\Phi(A)$ is upper bounded, by increasing u slowly over time.

This ensures that the maximum eigenvalue will never get close to u, as otherwise the potential function will blow up and would not be bounded.

Suppose we guarantee that $\Phi(A) = \sum_{i=1}^{n} \frac{1}{u-\lambda_i} \le 1$. This will ensure that at most one eigenvalue $\ge u-1$, at most two eigenvalues $\ge u-2$, etc. This gives us a more global measure of the solution.

This is very similar to the use of barrier functions in interior point method in convex optimization.

<u>**Outline**</u>

To give a more concrete idea, we present the algorithm in a simple (but non-optimal) setting of parameters

Initially, $A=0$, $u_0 = n$ and $l_0 = -n$.

Invariants to maintain are $\Phi_u(A) = Tr(uI-A)^{-1} = \sum_{i=1}^{n} \frac{1}{u-\lambda_i} \leq 1$ and $\Phi_\ell(A) = Tr(A-\ell I)^{-1} = \sum_{i=1}^{n} \frac{1}{\lambda_i - \ell} \leq 1$,

where the second function is the lower barrier function to lower bound the minimum eigenvalue.

In each iteration, we choose a vector v_i and set $A \leftarrow A + t v_i v_i^T$ for some scalar t,

so that $\Phi_{u+\varepsilon_u}(A + t v_i v_i^T) \leq 1$ and $\Phi_{\ell+\varepsilon_\ell}(A + t v_i v_i^T) \leq 1$.

Think of both ε_u and ε_ℓ are constants, say $\varepsilon_u = 2$ and $\varepsilon_\ell = \frac{1}{3}$.

If all the iterations are possible, after $6n$ iterations, the maximum eigenvalue is at most $13n$,

while the minimum eigenvalue is at least n, and thus we have a spectral sparsifier with

$6n$ edges that approximate all quadratic forms up to a constant factor.

Analysis

The mathematics turns out to work very nicely.

The first thing to understand is how the potential functions change upon adding a vector.

We would like to know for what δ_u and t such that $\Phi_{u+\delta_u}(A + t vv^T) \leq \Phi_u(A)$.

By definition, $\Phi_{u+\delta_u}(A + t vv^T) = Tr(u'I - A - tvv^T)^{-1}$ where $u' = u + \delta_u$.

There is a well-known formula for updating the inverse after a rank one update.

Sherman-Morrison-Woodbury formula $\quad (M - txx^T)^{-1} = M^{-1} + \frac{tM^{-1}xx^TM^{-1}}{1 - tx^TM^{-1}x}$ for non-singular M.

Upper barrier

$$
\begin{aligned}
\text{So, } \Phi_{u+\delta_u}(A + t vv^T) &= Tr(u'I - A - tvv^T)^{-1} \\
&= Tr\left((u'I-A)^{-1} + \frac{t(u'I-A)^{-1}vv^T(u'I-A)^{-1}}{1 - t\, v^T(u'I-A)^{-1}v} \right) \\
&= Tr(u'I-A)^{-1} + t\, Tr\left(\frac{v^T(u'I-A)^{-2}v}{1 - t\, v^T(uI-A)^{-1}v} \right) \qquad \text{as } Tr(XY) = Tr(YX) \\
&= \Phi_{u'}(A) + \frac{t\, v^T(u'I-A)^{-2}v}{1 - t\, v^T(u'I-A)^{-1}v}.
\end{aligned}
$$

$$
\begin{aligned}
\text{Therefore, } \Phi_{u'}(A + tvv^T) \leq \Phi_u(A) &\iff \Phi_{u'}(A) - \Phi_u(A) + \frac{v^T(u'I-A)^{-2}v}{\frac{1}{t} - v^T(u'I-A)^{-1}v} \leq 0 \\
&\iff \frac{1}{t} \geq \frac{v^T(u'I-A)^{-2}v}{\Phi_u(A) - \Phi_{u'}(A)} + v^T(u'I-A)^{-1}v \qquad (*)
\end{aligned}
$$

So, for any v, we can increase t a bit without violating the upper barrier, which ensures

that the maximum eigenvalue after adding tvv^T is at most $u + \delta_u$ as long as $\Phi_u(A)$ is bounded.

Lower barrier.

This is mostly similar, with some subtle difference.

$$\Phi_{\ell+\delta_\ell}(A+tvv^T) = \mathrm{Tr}\,(A+tvv^T-\ell' I)^{-1} \qquad (\text{where } \ell' = \ell+\delta_\ell)$$

$$= \mathrm{Tr}\left((A-\ell' I)^{-1} - \frac{t(A-\ell' I)^{-1}vv^T(A-\ell' I)^{-1}}{1+tv^T(A-\ell' I)^{-1}v}\right)$$

$$= \Phi_{\ell'}(A) - \frac{tv^T(A-\ell' I)^{-2}v}{1+tv^T(A-\ell' I)^{-1}v}.$$

So, $\Phi_{\ell+\delta_\ell}(A+tvv^T) \le \Phi_\ell(A) \iff \Phi_{\ell'}(A) - \Phi_\ell(A) - \dfrac{v^T(A-\ell' I)^{-2}v}{\frac{1}{t}+v^T(A-\ell' I)^{-1}v} \le 0$

$$\iff \frac{1}{t} \le \frac{v^T(A-\ell' I)^{-2}v}{\Phi_{\ell'}(A)-\Phi_\ell(A)} - v^T(A-\ell' I)^{-1}v. \qquad (**)$$

The RHS could be negative, and so for some v, it may not be possible to find a $t \ge 0$ so that the lower barrier function doesn't increase.

Both barriers

We would like to find a v_i so that there exists t to satisfy both (*) and (**) simultaneously. The important idea here is to show that the sum of RHS of (**) is at least the sum of RHS of (*), and this would imply there exists a specific v_i with the RHS of (**) at least the RHS of (*), and so we can choose t for that v_i so that both potential functions don't increase.

Let's first consider the sum of RHS of (*), which is

$$\sum_i \left(\frac{v_i^T(u'I-A)^{-2}v_i}{\Phi_u(A)-\Phi_{u'}(A)} + v_i^T(u'I-A)^{-1}v_i \right)$$

$$= \sum_i \left(\frac{\mathrm{Tr}((u'I-A)^{-2}v_iv_i^T)}{\Phi_u(A)-\Phi_{u'}(A)} + \mathrm{Tr}((u'I-A)^{-1}v_iv_i^T) \right)$$

$$= \frac{\mathrm{Tr}((u'I-A)^{-2}\sum v_iv_i^T)}{\Phi_u(A)-\Phi_{u'}(A)} + \mathrm{Tr}((u'I-A)^{-1}\sum_i v_iv_i^T) \qquad \text{as trace is linear}$$

$$= \frac{\mathrm{Tr}(u'I-A)^{-2}}{\Phi_u(A)-\Phi_{u'}(A)} + \mathrm{Tr}(u'I-A)^{-1} \qquad v_iv_i^T = I \text{ by the assumption of the theorem}$$

$$= \frac{\sum_j \frac{1}{(u+\delta_u-\lambda_j)^2}}{\sum_j \frac{1}{u-\lambda_j} - \sum_j \frac{1}{u+\delta_u-\lambda_j}} + \Phi_{u'}(A)$$

$$= \frac{\sum_j (u+\delta_u-\lambda_j)^{-2}}{\sum_j \delta_u (u-\lambda_j)^{-1}(u+\delta_u-\lambda_j)^{-1}} + \Phi_{u'}(A)$$

$$< \frac{1}{\delta_u} + \Phi_u(A) \qquad \text{as } \Phi_u(A) > \Phi_{u'}(A)$$

The sum of the RHS of (**) is similar but more involved:

$$\sum_i \left(\frac{v_i^T (A-l'I)^{-2} v_i}{\Phi_{l'}(A) - \Phi_l(A)} - v_i^T (A-l'I)^{-1} v_i \right)$$

$$= \frac{\mathrm{Tr}(A-l'I)^{-2}}{\Phi_{l'}(A) - \Phi_l(A)} - \mathrm{Tr}(A-l'I)^{-1} \quad \text{using trace is linear and } \sum_i v_i v_i^T = I$$

$$= \frac{\sum_j (\lambda_j - l - \delta_l)^{-2}}{\sum_j (\lambda_j - l - \delta_l)^{-1} - \sum_j (\lambda_j - l)^{-1}} - \sum_j (\lambda_j - l - \delta_l)^{-1}$$

$$\geq \frac{1}{\delta_l} - \Phi_l(A) \quad \text{hiding the calculations in Claim 3.6 of [BSS].}$$

Therefore, if we choose δ_u, δ_l and set the initial potentials $\Phi_u(A), \Phi_l(A)$ in such a way that $\frac{1}{\delta_l} - \Phi_l(A) \geq \frac{1}{\delta_u} + \Phi_u(A)$, then we can choose some v_i and t such that $\Phi_{l'}(A+tv_iv_i^T) \leq \Phi_l(A)$ and $\Phi_{u'}(A+tv_iv_i^T) \leq \Phi_u(A)$, and thus the condition will continue to satisfy, and we can repeat as many times as we want.

<u>Setting the parameters</u>

Set $l_0 = -\sqrt{d}\, n$ and $u_0 = \left(\frac{d+\sqrt{d}}{\sqrt{d}-1}\right) n$ so that initially $\Phi_{l_0}(A) = \frac{1}{\sqrt{d}}$ and $\Phi_{u_0}(A) = \frac{\sqrt{d}-1}{d+\sqrt{d}}$.

Set $\delta_l = 1$ and $\delta_u = \frac{\sqrt{d}+1}{\sqrt{d}-1}$ so that the condition $\frac{1}{\delta_l} - \Phi_{l_0}(A) \geq \frac{1}{\delta_u} + \Phi_{u_0}(A)$ holds.

Therefore, after dn steps, $\frac{\lambda_{max}(A)}{\lambda_{min}(A)} \leq \frac{\Phi_{u_0}(A) + dn\,\delta_u}{\Phi_{l_0}(A) + dn\,\delta_l} = \frac{d+2\sqrt{d}+1}{d-2\sqrt{d}+1}$.

Reference Twice Ramanujan sparsifiers, by Batson, Spielman, Srivastava.