CS 388R: Randomized Algorithms Fall 2015

Lecture 20 — 16 Nov 2015

Prof. Eric Price *Scribe: Jessica Hoffmann, Ashish Bora*

1 Overview

In this lecture we will look at the **Graph Sparsification** Problem which is the following: Given a dense graph $G = (V, E_G, W_G)$, find a sparse graph $H = (V, E_H, W_H)$, which *approximately* preserves some properties of G. The vertex set will remain the same, but the edge set and their weights can be different. We will henceforth denote $|V|$ by n.

What are the different properties of G that we would like to preserve?

2 Cut-Sparsifier

In the first lecture we studied a randomized algorithm to compute the min-cut in a graph. Here we study a related problem of finding a **cut-sparsifier**, namely, a sparse graph H, that approximately **preserves all the cuts** in G. Formally,

For a given graph $G = (V, E, W)$, a cut $S \subseteq V$ has size:

$$C_G(S) = \sum_{(u,v)\in E} W(u,v) \cdot \mathbb{I}_{\{u\in S, v\notin S\}}$$

Definition 1 (Cut-sparsifier)**.** *H is a cut-sparsifier for G if:*

$$\forall S \subseteq V, C_H(S) = (1 \pm \epsilon) C_G(S)$$

Definition 2 (Expander)**.** *A d-regular unweighted graph is an expander if:*

$$\forall S, 2|S| < n \implies (1-\epsilon)d|S| \leq C_H(S) \leq d|S|$$

Note that the rightmost inequality holds for all d-regular graphs. So intuitively, what we are saying is that **every subset has a large neighborhood**, which implies that **every two vertices are connected by a short path** (length $\mathcal{O}(\log n)$). Also, note that for the definition to be meaningful, it is necessary to restrict $|S|$, since by choosing $S = V$, we have $C_G(S) = 0$ which does not satisfy our requirement.

Example of cut sparsifier: Suppose $G = K_n$ = complete graph on n vertices. For $|S| < \epsilon n$, we have:

$$C_G(S) = |S|(n - |S|) \in n|S| \cdot [1-\epsilon, 1]$$

Now, if H is chosen to be a degree d-expander and all edge weights $= n/d$, then:

$$C_H(S) \in d|S|\frac{n}{d} \cdot [1-\epsilon, 1] = n|S| \cdot [1-\epsilon, 1]$$

$$\implies \frac{C_G(S)}{C_H(S)} \in \frac{[1-\epsilon, 1]}{[1-\epsilon, 1]} \in [1 \pm \mathcal{O}(\epsilon)]$$

So, H approximates cuts in G for all subsets with small size. If $\epsilon > 1/2$, then H is a cut sparsifier for G.

3 Spectral Sparsifier

Here we generalize the notion of cut-sparsification [2]. First let us define what a spectral sparsifier is. For this purpose here is a cool analogy:

Think of the given graph as a network of resisitors. The weight of an edge can be thought of as the conductance (inverse of resistance) of that edge. Now if we apply some external voltage at the vertices, some current will flow along the edges. This will result in dissipation of power (Joule effect). So now, we can think of a function $P_G(\cdot)$, that takes applied node voltages as input and gives the total power dissipation as output. Obviously this function depends on the graph. We will denote the vector of voltages by $x \in \mathbb{R}^n$. Thus,

$$P_G(x)\text{: Voltage} \rightarrow \text{Power}.$$

Since power is non-negative, $P_G(x) : \mathbb{R}^n \rightarrow \mathbb{R}^+$.

By Ohms law, for every edge,

$V = IR = \frac{I}{C}$,

where I is the current along an edge, V is the voltage difference, and $C = \frac{1}{R}$ is the conductance of the edge.

Let P_e = power dissipated along edge e. $P_e = IV = V^2C$. Total power is the sum of all edge powers, and also, conductance is same as edge weights. So

$$P_G(x) = \sum_{e=(u,v)\in E_G} W(u,v)(x_u - x_v)^2$$

There is a more compact way of writing this expression. For this, consider only one edge e.

$$\begin{aligned}
P_e &= W(u,v)(x_u - x_v)^2 \\
&= W(u,v)(x_u^2 - 2x_u x_v + x_v^2) \\
&= x^\top \begin{bmatrix} & \vdots & & \vdots & \\ \dots & W(u,v) & \dots & -W(u,v) & \dots \\ & \vdots & & \vdots & \\ \dots & -W(u,v) & \dots & W(u,v) & \dots \\ & \vdots & & \vdots & \end{bmatrix} x \\
&:= x^\top A_{u,v} x.
\end{aligned}$$

In $A_{u,v}$ defined above, all the other coefficients are 0. Define

$$L_G = \sum_{(u,v)\in E_G} A_{u,v}.$$

So we have

$$L_G(u,v) = \begin{cases} -W(u,v) & u \neq v \\ \sum_t W(u,t) & u = v \end{cases}$$

L_G is called the graph Laplacian. So we have $P_G(x) = \sum_{e\in E_G} P_e = x^\top L_G x$

Definition 3 (Spectral Sparsifier)**.** *A spectral sparsifier is a graph that spectrally approximates the graph Laplacian. i.e. for all voltages x, we should have*

$$\begin{aligned} P_H(x) &= (1 \pm \epsilon) P_G(x) \\ \Leftrightarrow (1-\epsilon)x^\top L_G x \leq x^\top L_H x &\leq (1+\epsilon) x^\top L_G x \quad \forall\, x \in \mathbb{R}^n \\ \Leftrightarrow (1-\epsilon) L_G \preceq L_H &\preceq (1+\epsilon) L_G \end{aligned}$$

Notation: $\preceq$ is the generalized matrix inequality on symmetric matrices: two symmetric matrices A and B satisfy $A \preceq B$ iff $(B - A)$ is positive semidefinite.

Theorem 4. *Spectral Sparsifier $\implies$ Cut-sparsifier*

Proof. Let S be any subset of vertices. Let $x = \mathbb{I}_S$. Since $(x_u - x_v)^2 = 1$ iff (u,v) has an endpoint in S, and the other endpoint out of S (so (u,v) is in the cut defined by S), we have $P_G(x) = C_G(x)$ □

Thus, for cut sparsification, we just need

$$(1-\epsilon)x^\top L_G x \leq x^\top L_H x \leq (1+\epsilon) x^\top L_G x \quad \forall\, x \in \{0,1\}^n$$

4 Sampling Edges

Notice that the condition for spectral sparsification can be rewritten as

$$-\epsilon L_G \preceq L_H - L_G \preceq \epsilon L_G$$

This is very similar to what we wanted to achieve in Johnson-Lindenstrauss Lemma, i.e. we want a compact representation such that we do not deviate too much from our input. Thus, in this section, we consider **sampling the edges** to obtain an approximation to L_G.

Let p_e be the probability of sampling the edge e. We know that if an edge is included with weight c_e, then the Laplacian must have c_e entry in the pattern of $A_{u,v}$.

Define $Y_e = c_e(e_u - e_v)$. Where e_u is the vector of all zeros except for 1 at u-th position. Now let $Z_1, Z_2, \ldots Z_m$ be i.i.d. samples such that $Z_i = Y_e$ with probability p_e. So,

$$
\begin{aligned}
\mathbb{E}[Z_i Z_i^\top] &= \sum_e p_e Y_e Y_e^\top \\
&= \sum_e p_e c_e^2 (e_u - e_v)(e_u - e_v)^\top
\end{aligned}
$$

Now if we choose $c_e = \dfrac{\sqrt{W(u,v)}}{\sqrt{p_e}}$, then

$$
\begin{aligned}
\mathbb{E}[Z_i Z_i^\top] &= \sum_e W(u,v)(e_u - e_v)(e_u - e_v)^\top \\
&= \sum_{e=(u,v)} A_{u,v} \\
&= L_G
\end{aligned}
$$

So, we output $L_H = \sum_{i=1}^{m} Z_i Z_i^\top$, which converges to L_G if m is large and the H thus formed is at least m-sparse (we only took m edges).

Next, we try to find how large m should be. For this, we first consider a simpler goal of making $\|L_H - L_G\|$ small. This will imply the desired spectral bound if L_G is round, i.e. if all its eigenvalues are similar. Such graphs are also called isometric.

Notation: $\|A\| = \sup_{\|x\| \le 1} \|Ax\|_2$.

We will show $\mathbb{E}[\|L_H - L_G\|] \le \sqrt{\frac{n \log n}{m}} \|L_G\|$, which gives us an ϵ-approximation if $m = \mathcal{O}(\frac{n \log n}{\epsilon^2})$.

Theorem 5 (Non-commutative Bernstein inequality)**.** *Extension of Bernstein-type inequalities to matrices [1].*

- X_i *independent symmetric matrices,* $i = 1 \dots n$
- $\mathbb{E}[X_i] = 0$
- $\|X_i\| \le K$
- $\left\| \sum_{i=1}^{n} \mathbb{E}[X_i^2] \right\| \le \sigma^2$

Then: $\exists\ C < 0$, *such that*

$$
\mathbb{P}\left(\left\| \sum_{i=1}^{n} X_i \right\| \ge t \right) \le 2n \cdot e^{C \min\left(\frac{t^2}{\sigma^2}, \frac{t}{K}\right)}
$$

We omit the proof of this theorem.

Theorem 6 (R-V theorem)**.** *Suppose we have:*

- X_i *i.i.d. vectors in* $\mathbb{R}^n$, $i = 1 \dots m$
- $\|X_i\|_2 \leq Q$, $Q \geq 1$
- $\|\mathbb{E}[XX^\top]\| \leq 1$

Then:

$$\mathbb{E}\left[\left\|\frac{1}{m}\sum_{i=1}^{m} X_i X_i^\top - \mathbb{E}[XX^\top]\right\|\right] \lesssim Q\sqrt{\frac{\log n}{m}}$$

Proof. Let $Y = XX^\top - \mathbb{E}[XX^\top]$. We want to apply the non-commutative Bernstein theorem to Y.

$\|Y\| \overset{?}{\leq} K$:

$$\begin{aligned}
\|Y\| &= \sup_{\|u\|\leq 1} u^\top Y u \\
&\leq \frac{X^\top XX^\top X}{\|X\|_2^2} - \inf_{\|u\|_2 \leq 1} u^\top \mathbb{E}[XX^\top] u \\
&\leq \|X\|_2^2 - \inf_{\|u\|_2\leq 1} \mathbb{E}[u^\top XX^\top u] \\
&\leq \|X\|_2^2 \\
&\leq Q^2
\end{aligned}$$

($XX^\top$ is positive semi-definite (Gram matrix), so $0 \preceq XX^\top$, or equivalently, $\forall u, u^\top XX^\top u \geq 0$.)

$\left\|\sum_{i=1}^{m} \mathbb{E}[Y_i^2]\right\| \overset{?}{\leq} \sigma^2$:

$$\begin{aligned}
\left\|\sum_{i=1}^{m} \mathbb{E}[Y_i^2]\right\| &\leq m\|\mathbb{E}[Y_1^2]\| \\
&= m\left\|\mathbb{E}\left[(XX^\top)^2 - \mathbb{E}[XX^\top]^2\right]\right\| \\
&\leq m\left(\left\|\mathbb{E}[\|X\|_2^2 \cdot XX^\top]\right\| + \left\|\mathbb{E}[XX^\top]\right\|^2\right)
\end{aligned}$$

$$\begin{aligned}
\left\|\mathbb{E}[\|X\|_2^2 \cdot XX^\top]\right\| &= \sup_{\|u\|_2\leq 1} \mathbb{E}\left[u^\top \|X\|_2^2 \cdot XX^\top u\right] \\
&\leq Q^2 \sup_{\|u\|_2 \leq 1} \mathbb{E}\left[u^\top \cdot XX^\top u\right] \\
&= Q^2 \cdot \|\mathbb{E}[XX^\top]\| \\
&\leq Q^2 \cdot 1 \\
&\leq Q^2
\end{aligned}$$

$$\left\|\sum_{i=1}^{m} \mathbb{E}[Y_i^2]\right\| \leq m\left(\left\|\mathbb{E}[\|X\|_2^2 \cdot XX^\top]\right\| + \left\|\mathbb{E}[XX^\top]\right\|^2\right)$$
$$\leq m(Q^2 + 1^2)$$
$$\leq 2mQ^2$$

Since the Y_i are symmetric and independent, and $\mathbb{E}[Y_i] = 0$, we can apply the non-commutative Bernstein inequality:

$$\mathbb{P}\left(\left\|\sum_{i=1}^{m} \mathbb{E}[Y_i]\right\| \geq mt\right) \leq 2m \cdot e^{C \min\left(\frac{mt^2}{2Q^2}, \frac{mt}{Q^2}\right)}$$

Hence, for $t = \mathcal{O}\left(\frac{Q^2}{m}\log n + Q\sqrt{2\frac{\log n}{m}}\right)$:

$$\mathbb{P}\left(\left\|\sum_{i=1}^{m} \mathbb{E}[Y_i]\right\| \geq mt\right) \leq \frac{1}{n^K}$$

□

More on the subject can be found here [3].

References

[1] R. Ahlswede and A. Winter. Strong converse for identification via quantum channels. *Information Theory, IEEE Transactions on*, 48(3):569–579, Mar 2002.

[2] J. Batson, D. A. Spielman, N. Srivastava, and S.-H. Teng. Spectral sparsification of graphs: Theory and algorithms. *Commun. ACM*, 56(8):87–94, Aug. 2013.

[3] D. A. Spielman and N. Srivastava. Graph sparsification by effective resistances. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, STOC '08, pages 563–568, New York, NY, USA, 2008. ACM.