CS 388R: Randomized Algorithms, Fall 2019 November 12th, 2019

Lecture 22: Spectral Sparsification

Prof. Eric Price *Scribe: Sebastian Oberhoff, Rachit Garg*

1 Overview

In the previous lecture we discussed:

- Bernstein Inequality and Matrix Bernstein Inequality.
- Rudelson-Vershynin Inequality.
- An introduction to graph sparsification.

In this lecture we look at spectral sparsification for complete graphs using the Rudelson-Vershynin Lemma and then analyze it for all undirected graphs.

2 Setup

Last time we saw the following Lemma due to Rudelson and Vershynin [RV05]:

Theorem 1 (RV Lemma)**.** *Let $x_1, \dots, x_m \in \mathbb{R}^n$ be independent and satisfying*

$$\|x_i\|_2 \le K \quad {\scriptstyle (K \ge 1)} \qquad \textit{and} \qquad \left\|\mathbb{E}\left[x_i x_i^T\right]\right\| \le 1\,.$$

Then

$$\mathbb{E}\left[\left\|\frac{1}{m}\sum_{i=1}^{m} x_i x_i^T - \mathbb{E}\left[x_i x_i^T\right]\right\|\right] \le \sigma K \sqrt{\frac{\log(n)}{m}}\,.$$

Note: For a matrix M we define

$$\begin{aligned} \|M\| &= \text{spectral norm} \\ &= \text{maximum eigenvalue (if positive semidefinite)} \\ &= \sup_{y \neq 0} \frac{y^T M y}{\|y\|_2^2} \end{aligned}$$

Given a graph $G = (V, E)$ define the diagonal matrix

$$D = \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{bmatrix}$$

where the d_i are the degrees, as well as the (weighted) adjacency matrix A.

This lets us further define the Laplacian

$$L_G = D - A = \sum_{e \in E} w_e u_e u_e^T$$

where w_e is the weight on edge e and u_e is of the form

$$u_e = \begin{bmatrix} -1 \\ \\ 1 \end{bmatrix} \begin{matrix} j \\ \\ i \end{matrix} \qquad \text{(e goes from vertex i to j).}$$

The goal is now: given a dense G, find a sparse H ($m = n\log(n)$) such that $L_H \approx L_G$. That is,

$$\forall x \in \mathbb{R}^n\colon \underbrace{x^T L_H x}_{P_H} = (1 \pm \epsilon) \underbrace{x^T L_G x}_{P_G} \iff (1-\epsilon) \preceq L_H \preceq (1+\epsilon) L_G$$

(P_H, P_G is the power on the graphs H, G respectively.)

Approach

Pick probabilities p_e. Set $z_i \coloneqq \sqrt{\frac{w_e}{p_e}} \cdot u_e$ with probability p_e. Then

$$\mathbb{E}\left[z_i z_i^T\right] = \sum_e p_e \frac{w_e}{p_e} u_e u_e^T = L_G \,.$$

So after obtaining $z_i, \ldots, z_m$ in this way we can construct

$$L_H \coloneqq \frac{1}{m} \sum_{i=1}^m z_i z_i^T$$

obeying $\mathbb{E}[L_H] = L_G$.

Question: for what m do we have $L_H \to L_G$?

3 Warmup: Complete Graph

For a complete graph we have

$$L_G = \begin{bmatrix} n-1 & -1 & -1 & \\ -1 & n-1 & -1 & \cdots \\ -1 & -1 & n-1 & \\ \vdots & & & \ddots \end{bmatrix} = nI - \mathbf{1}\mathbf{1}^T$$

$$P_G(x) = x^T L_G x^T = n\|x\|_2^2 - (\mathbf{1}^T x)^2 \qquad \forall x \text{ s.t. } \mathbf{1}^T x = 0$$

And what we want to show is

$$\frac{x^T(L_H - L_G)x}{\underbrace{x^T L_G x}_{n\|x\|_2^2}} \leq \epsilon \qquad \Longleftrightarrow \qquad \|L_G - L_H\| \leq n\epsilon\,.$$

By the RV Lemma we have

$$\mathbb{E}\big[\|L_H - L_G\|\big] \leq \sigma K\sqrt{\tfrac{\log(n)}{m}}$$

where

$$K = \max_i \|z_i\| = \max_e \sqrt{\tfrac{w_e}{p_e}\cdot 2} = \sqrt{n(n-1)} \approx n \qquad \text{and} \qquad \sigma = \|L_G\| = n\,.$$

Hence,

$$\mathbb{E}\big[\|L_H - L_G\|\big] \leq n\sqrt{n\tfrac{\log(n)}{m}}\,.$$

So $n = O\big(\frac{n\log(n)}{\epsilon^2}\big)$ suffices.

4 Spectral Sparsification for General Graphs

In the previous section we discussed spectral sparsification for complete graphs using the RV lemma [RV05]. The main observation here is that in the case for complete graphs the distribution was spherical, i.e. each edge had the same weight. But in the case of general graphs there could be some edges which are more important than the others. (Imagine the case of a barbell graph for an intuition. Here the connecting edge should be preserved in order for the two parts of the graph to remain connected). Thus we must appropriately handle the skewed distribution of a non-complete graph. We first introduce some notation and consider the simplifying assumption that the graph is an **unweighted undirected graph**.

Define $U \in \mathbb{R}^{|E|\times n}$ to be the matrix below, where $|E|$ is the number of edges and n is the number of vertices of the graph (we define u_e for edge $e = (a, b)$ to be $(e_a - e_b)$ where e_x is the elementary vector $\in \{0,1\}^n$):

$$U = \begin{bmatrix} — & u_1^T & — \\ — & u_2^T & — \\ — & \vdots & — \\ — & u_m^T & — \end{bmatrix}$$

$$L_G = U^T U,\ \ L_H = U^T S U$$

where $S \in \mathbb{R}^{|E|\times|E|}$ is a diagonal matrix such that $S_{e,e} = \frac{\#\text{times } e \text{ sampled}}{m\,p_e}$ and m is the number of samples.

4.1 Intuition from Electrical Engineering

Let v be the voltage at each vertex $\in \mathbb{R}^n$ and y be the currents across each edge $\in \mathbb{R}^{|E|}$.

Ohm's Law: states that $V = IR$ or $I = VC$, where C is the conductance. As we consider an unweighted graph, the conductance is 1 throughout. Hence we get that $I = V$ which says that the potential difference is equal to the current. This implies that $\forall i, j \colon y_{i,j} = v_i - v_j \implies y = VU$.

Kirchhoff's Current Law: Consider injecting current, $x \in \mathbb{R}^n$ into the graph. By Kirchhoff's current law, $x^T \mathbf{1} = 0$, due to charge conservation, sum of all incoming and outgoing currents is zero. Also observe that $x = U^T y = U^T U v = L_G v$. Since every matrix with real entries must have a pseudoinverse, we can write that $v = L_G^+ x$. The use of the voltage vector will be in describing the exact probabilities with which we should choose each edge and sparsify the graph.

Let r_e be the effective resitance of an edge, i.e. it is the potential difference induced when we inject one ampere of current between two junctions.

$$\begin{aligned} V = IR \implies r_e &= \text{voltage gap when we force } 1A \text{ current across an edge } e \\ &= u_e^T v \qquad \text{(when } x = u_e\text{)} \\ &= u_e^T L_G^+ u_e \end{aligned}$$

More generally, if we write the effective resistance R as a matrix where the diagonal entries are the effective resistances, then

$$R = U L_G^+ U^T$$

where $r_e = R_{e,e}$.

Claim 2. *[SS08] If $p_e \propto r_e (p_e = \frac{r_e}{n-1})$, then $m = O(\frac{n \log n}{\epsilon^2})$ suffices for L_H to be spectral sparsifier i.e. L_H approximates L_G with high probability.*

We now look at some properties of R

Lemma 3. *R is a projection matrix*

Lemma 4. *R has $n-1$ eigenvalues $= 1$, (rest $= 0$)*

Lemma 5. $\mathbb{E}\big[\|RSR - R\|\big] \leq \epsilon$

Lemma 6. $(1-\epsilon) L_G \preceq L_H \preceq (1+\epsilon) L_G$

Lemma 3

Proof.

$$\begin{aligned} R^2 &= U L_G^+ (U^T U) L_G^+ U^T \\ &= U (L_G^+ L_G) L_G^+ U^T \\ &= U L_G^+ U^T \\ &= R\,. \end{aligned}$$

□

Lemma 4

Proof. This can be seen easily due to R being a projection matrix. Let v be an eigenvector and λ be the corresponding eigenvalue i.e. $Rv = \lambda v$. $R^2 v = R\lambda v = \lambda^2 v$. But R^2 and R are the same matrices, hence $\lambda^2 = \lambda \implies \lambda \in \{0, 1\}$.

Theorem 7 (Foster's theorem)**.** *R has $n-1$ eigenvalues = 1 (rest 0).*

Using Foster's theorem we also get that $\sum_e r_e = \mathrm{Tr}(R) = n - 1$. ☐

Lemma 6

Proof. Assume that Lemma 5 is true for now. Using Markov's we have the fact that $\|RSR - R\| \leq c\epsilon$ with good probability for some constant c. Note that we can achieve high probability over here if we revisit the proof of [RV05]. Ignoring the constant c as we perform an asymptotic analysis at the end.

We have

$$\begin{aligned} \epsilon &\geq \|RSR - R\| \\ &= \|R(S-I)R\| \\ &= \sup_{y \neq 0} \frac{y^T R(S-I)Ry}{y^T y} \,. \end{aligned}$$

We want to bound

$$\begin{aligned} \sup_x \frac{x^T(L_H - L_G)x}{x^T L_G x} &= \sup_x \frac{x^T U^T (S-I) U x}{x^T U^T U x} \\ &= \sup_{y = Ux} \frac{y^T (S-I) y}{y^T y} \end{aligned}$$

(Note that $y = Ux = UL_G^+ U U^T x = RU^T x$. Thus, y lies in the range of R always.)

$$\begin{aligned} &= \sup_{y \in \mathrm{range}(R),\ y = Ry'} \frac{y'^T R^T (S-I) R y'}{y'^T R^T R y'} \\ &= \sup_{y \in \mathrm{range}(R),\ y = Ry'} \frac{y'^T R (S-I) R y'}{y'^T R R y'} \\ &= \sup_{y \in \mathrm{range}(R),\ y = Ry'} \frac{y'^T R (S-I) R y'}{y'^T y'} \qquad \big(\text{as } \|R\| = 1\big) \\ &\leq \epsilon \,. \end{aligned}$$

This gives us that

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

☐

Lemma 5

Proof.

$$\begin{aligned} RSR &= \sum_e S_{e,e} \cdot R_e R_e^T && \text{(where } R_e \in \mathbb{R}^n \text{ is the } e^{th} \text{ column of } R\text{)} \\ &= \sum_e \frac{\#\text{ times e sampled}}{m\, p_e} R_e R_e^T\,. \end{aligned}$$ □

Set $z_i = \frac{R_e}{\sqrt{p_e}}$ with probability p_e, then $S = \sum_{i=1}^m z_i z_i^T$, has $\mathbb{E}[S] = R$.

RV lemma uses

$$\|\mathbb{E}[z_i z_i^T]\| = \|R\| = 1$$

$$\|z_i\| = \frac{1}{\sqrt{p_e}}\|R_e\|_2 = \frac{\sqrt{r_e}}{\sqrt{p_e}} = \sqrt{n-1} \tag{1}$$

and from the statement of the lemma, we get that

$$\mathbb{E}\big[\|RSR - R\|\big] \leq \sqrt{\frac{(n-1)\log n}{m}} \leq \epsilon\,.$$

(1) uses the fact that $\|R_e\|_2 = \sqrt{r_e}$ which will see in the next class to complete the proof.

References

[RV05] Mark Rudelson, Roman Vershynin. Sampling from large matrices: an approach through geometric functional analysis. *arXiv: math/0503442 [math.FA]*, 2005.

[SS08] Daniel Spielman, Nikhil Srivastava. Graph Sparsification by Effective Resistances. *arXiv: 0803.0929 [cs.DS]*, 2008.