CME 305: Discrete Mathematics and Algorithms

The following lecture notes are based heavily on the paper by Spielman and Srivastava, sometimes following it literally.

1 Introduction

The goal of sparsification is to approximate a given graph G by a sparse graph H on the same set of vertices. If H is close to G in some appropriate metric, then H can be used as a proxy for G in computations without introducing too much error. At the same time, since H has very few edges, computation with and storage of H should be cheaper.

Given a graph $G = (V, E)$, we wish to solve some problems that involve cuts of G. For example: Minimum Cut, Sparsest Cut, Maximim Cut, etc. The running time of algorithms for these problems typically depends on the number of edges in the graph, which might be as high as $\binom{n}{2}$. To bring down the running time of these algorithms, it would be nice if we could build another graph G' that approximates G across all cuts, but has fewer edges.

We will describe the method of Spielman and Svristava (2009) to 'sparsify' our graph, which uses effective resistances to drive a sampling procedure.

2 Random Sampling

To see how naive random sampling performs, we will sample each edge with the same probability p, and give weight $1/p$ to each edge in the sparse graph H. With these weights, each edge $e \in E$ will have expected contribution exactly 1 to any cut, and thus the expected weight of any cut in H will match G. It remains to see how many samples we need to have cut equivalence between G and H with high probability.

Consider a particular cut $S \subseteq V$. If it has c edges crossing it in G, the expected weight of edges crossing it in the new graph H is also c. Denote the total weight of edges between S and $V - S$ by $f_G(S) = c$, we have the following concentration result due to Chernoff:

$$P(|f_H(S) - c| \geq \epsilon c) \leq 2e^{-\epsilon^2 pc/2}$$

In particular, picking $p = \Omega(\frac{t \log n}{\epsilon^2 c})$ will make the RHS of the above less than $1/n^t$. To make sure this result holds for *all* cuts, we take c to be the global minimum cut, and apply union bound using Karger's cut-counting theorem. Karger's cut-counting says that the If G has a min-cut of size c, then the number of cuts of value αc is at most $n^{2\alpha}$.

This sampling procedure gives us a sparsified graph H with $O(\frac{n^2 \log(n)}{c\epsilon^2})$ edges. This has a major drawback: the dependence on c can be really disadvantageous when $c = 1$, which happens for example in the dumbbell graph. We alleviate this problem via importance sampling and effective resistances.

3 Sampling with effective resistances

The main idea is to include each edge of G in the sparsifier H with probability proportional to its effective resistance. The effective resistance of an edge is known to be equal to the probability that the edge appears in a random spanning tree of G, and is proportional to the commute time between the endpoints of the edge. We show how sampling according to these resistances yields a good sparsifier.

To define effective resistance, identify $G = (V, E)$ with an electrical network on n nodes in which each edge e corresponds to a link of conductance 1. Then the effective resistance R_e across an edge e is the potential difference induced across it when a unit current is injected at one end of e and extracted at the other end of e. The algorithm can now be stated as follows.

H = **Sparsify**(G, q)
for q times **do**
  Sample an edge e with probability $p_e = \frac{R_e}{n-1}$
  Add e to H with weight $\frac{1}{p_e q}$ (summing weights if needed)
end for

3.1 Laplacian

Recall that the Laplacian of a weighted graph is given by $L_G = D - A$ where A is the weighted adjacency matrix and D is the diagonal matrix of weighted degrees of each node. When we drop the subscript and write L, we have that L_G is intended. Notice that the quadratic form associated with L is just

$$x^T L x = \sum_{(u,v)\in E} (x(u) - x(v))^2 w_{uv}$$

Let L_G be the Laplacian of G and let L_H be the Laplacian of H. The main theorem is that if q is sufficiently large, then the quadratic forms of L_H and L_G are close.

Theorem 1 *Suppose G and $H = Sparsify(G, q)$ have laplacians L_G and L_H respectively, and $1\sqrt{n} < \epsilon \leq 1$. If $q = O(n \log(n)/\epsilon^2)$, then probability at least $1/2$, we have for all $x \in \mathcal{R}^n$:*

$$(1 - \epsilon) x^T L_H x \leq x^T L_G x \leq (1 + \epsilon) x^T L_H x$$

3.2 Effective Resistances via Linear Algebra

Let G be an undirected graph and B its corresponding $m \times n$ incidence matrix. First, arbitrarily orient each edge of E, then define the (e, v) entry of B as 1 if v is the head of e, and -1 if v is the tail. All other entries are zero. Then it is easy to see that $L_G = B^T B$. We denote the rows of B as b_e. Note that $b_e^T = \chi_v - \chi_u$.

It is immediate that L is positive semidefinite since $x^T L x = x^T B^T B x = ||Bx||_2^2 \geq 0$, any $x \in \mathcal{R}^n$. We also have $\ker(L) = \ker(B) = \text{span}(1)$ since

$$x^x L x = 0 \Leftrightarrow ||Bx||_2^2 = 0 \Leftrightarrow \sum_{(u,v)\in E} (x(u) - x(v))^2 w_{uv} = 0$$

$$\Leftrightarrow x(u) - x(v) = 0$$

$$\Leftrightarrow \ x \text{ is constant, since } G \text{ is connected}$$

Now we introduce the pseudoinverse of L, denoted L^+. Since L is symmetric we can diagonalize it as $L = \sum_{i=1}^{n-1} \lambda_i u_i u_i^T$, where $\lambda_1, \ldots, \lambda_{n-1}$ are the nonzero eigenvalues of L, and u_i the eigenvectors. The *Moore-Penrose* Pseudoinverse of L is defined

$$L^+ = \sum_{i=1}^{n-1} \frac{1}{\lambda_i} u_i u_i^T$$

Notice that

$$LL^+ = L^+L = \sum_{i=1}^{n-1} u_i u_i^T$$

Which is simply the projection onto the span of the nonzero eigenvectors of L. Thus LL^+ is the identity on $\mathrm{im}(L) = \mathrm{im}(L^+) = \mathrm{span}(1)^\perp$.

Now we define some electrical flow notation. For a vector $i_{\mathrm{ext}}(u)$ of currents injected at the vertices, let $i(e)$ be the currents induced in the edges (in the direction of orientation) and $v(u)$ the potentials induced at the vertices. By Kirchoffs current law, the sum of the currents entering a vertex is equal to the amount injected at the vertex:

$$B^T i = i_{\mathrm{ext}}$$

By Ohm's law, the current flow in an edge is equal to the potential difference across its ends times its conductance:

$$i = Bv$$

Combining these two facts

$$i_{\mathrm{ext}}(u) = B^T B v = Lv$$

Now since $i_{\mathrm{ext}} \perp 1$, and by the projection property of the pseudoinverse, we have

$$v = L^+ i_{\mathrm{ext}}$$

Recall that the effective resistance between two vertices u and v is defined as the potential difference induced between them when a unit current is injected at one and extracted at the other. We will derive an algebraic expression for the effective resistance in terms of L^+. To inject and extract a unit current across the endpoints of an edge $e = (u, v)$, we set $i_{\mathrm{ext}} = b_e^T = (\chi_v - \chi_u)$, which is clearly orthogonal to 1. The potentials induced by i_{ext} at the vertices are given by $v = L^+ b_e^T$; to measure the potential difference across $e = (u, v)$, we simply multiply by b_e on the left:

$$v(v) - v(u) = (\chi_v - \chi_u) = (\chi_v - \chi_u)^T L^+ b_e^T$$

It follows that the effective resistance across $e = (u, v)$ is given by $b_e L^+ b_e^T$.

3.3 The Projection Matrix Π

Define the matrix $\Pi = BL^+B^T$. It is clear that $\mathrm{im}(\Pi) \subseteq \mathrm{im}(B)$. We also have that $\mathrm{im}(B) \subseteq \mathrm{im}(\Pi)$ since for any $y \in \mathrm{im}(B)$ there exists an $x \perp 1$ such that $Bx = y$, and so $\Pi y = BL^+B^TBx = y$ by the projection property of LL^+. Thus $\mathrm{im}(\Pi) = \mathrm{im}(B)$.

It is easy to see that P is the orthogonal projection onto $\mathrm{im}(B)$, since

$$\Pi^2 = BL^+B^TBL^+B^T = BL^+LL^+B^T = \Pi$$

Furthermore Π is symmetric, and has its e'th diagonal entry equal to R_e. Since Π is a projection with an $n-1$ dimensional range, it has exactly $n-1$ eigenvalues equal to 1, and all others equal to 0. Therefore the trace of Π, is $\mathrm{tr}(\Pi) = n-1$. This implies the sum of all effective resistances across edges of a graph is $n-1$, and explains the normalization factor in the sampling algorithm Sparsify(G, q) introduced earlier.

4 The Main Result

First we get handle on the output of the sparsification algorithm using linear algebra. Define the $m \times m$ diagonal matrix S, which has on the e'th diagonal set to the number of times edge e was sampled, scaled by $1/qp_e$. More concretely, define the random variable d_e as the number of times edge e was sampled. Then

$$S(e,e) = \frac{d_e}{qp_e}$$

We then write the laplacian of H as $L_H = B^T S B$. Note that S and L_H are random matrices with $E[S] = I_m$ and $E[L_H] = L_G$.

Now, we will apply a concentration bound of Rudelson and Vershynin to bound the number of samples q needed to ensure $||\Pi S \Pi - \Pi\Pi||_2$ small. To do so, we interpret the sampling procedure as sampling columns of Π (of which there are m) from the distribution which draws column e as $\Pi(.,e)/\sqrt{qp_e}$ with probability p_e. With this definition for samples of y_i, and the definition of S, it is clear

$$\Pi S \Pi = \frac{1}{q}\sum_{i=1}^{q} y y_i^T$$

Furthermore $E[yy^T] = \Pi\Pi^T = \Pi\Pi$. We can now apply the bound of Rudelson & Vershynin. This concentration inequality is given in the following lemma.

Lemma 1 *Let $\boldsymbol{p}$ be a probability distribution over $\Omega \subseteq \mathcal{R}^d$ such that $\sup_{y\in\Omega} ||y||_2 \leq M$ and $||E_p yy^T||_2 \leq 1$. Let $y_1, \ldots, y_q$ be independent samples drawn from $\boldsymbol{p}$, then*

$$E||\frac{1}{q}\sum_{i=1}^{q} y_i y_i^T - Eyy^T||_2 \leq \min(CM\sqrt{\frac{\log q}{q}}, 1)$$

Where C is an absolute constant.

Applying the above lemma to $\Pi S \Pi$ and $\Pi\Pi$, we have by setting $q = O(\frac{n \log n}{\epsilon^2})$

$$E||\Pi S \Pi - \Pi\Pi||_2 = E||\frac{1}{q}\sum_{i=1}^{q} y y_i^T - Eyy^T||_2 \leq \epsilon/2$$

Assuming $\epsilon \leq 1/\sqrt{n}$ which will happen for n large enough. Finally, by Markov's inequality we have

$$||\Pi S \Pi - \Pi\Pi||_2 \leq \epsilon$$

with probability at least half.

It remains to show how $||\Pi S \Pi - \Pi\Pi||_2$ being small implies that L_H and L_G are close in relative error. To see this:

$$||\Pi S \Pi - \Pi\Pi||_2 \leq \epsilon \Leftrightarrow \sup \frac{|y^T \Pi (S-I)\Pi y|}{y^T y} \leq \epsilon,\ \forall y \in \mathcal{R}^m$$

$$\Rightarrow \sup \frac{|y^T \Pi (S-I)\Pi y|}{y^T y} \leq \epsilon,\ \forall y \in \text{im}(B)$$

$$\Rightarrow \sup \frac{|y^T (S-I) y|}{y^T y} \leq \epsilon,\ \forall y \in \text{im}(B)$$

$$\Rightarrow \sup \frac{|(Bx)^T (S-I)(Bx)|}{(Bx)^T (Bx)} \leq \epsilon,\ \forall y = Bx, x \perp 1$$

$$\Rightarrow \sup \frac{|x^T L_H x - x^T L_G x|}{x^T L_G x} \leq \epsilon, \, \forall x \perp 1$$
$$\Rightarrow (1-\epsilon) x^T L_G x \leq x^T L_H x \leq (1+\epsilon) x^T L_G x, \, \forall x \perp 1$$

Thus we have the required result for all $x \notin \ker(B)$. For $x \in \ker(B)$, $L_H x = L_G x = 0$, thus the result holds trivially. So we have for all $x \in \mathcal{R}^n$

$$(1-\epsilon) x^T L_G x \leq x^T L_H x \leq (1+\epsilon) x^T L_G x$$

As required.