15-859N Spectral Graph Theory and The Laplacian Paradigm, Spring 2020

Homework 5 Version 1[1] **Due: Monday April 27th in class**

Gary Miller Zico Kolter Timothy Chu

Instructions. Collaboration is permitted in groups of size at most three. You must write the names of your collaborators on your solutions and must write your solutions alone.

Question	Points	Score
1	25	
2	25	
3	25	
Total:	75	

(25) 1. Sampling Multi-graphs

In class, we said that given two positively weighted graphs G and H on the same set of vertices the $G \preceq H$ if $\forall x \; x^T L_G x \leq x^T L_H x$. If G and H are weighted multi-graphs then we shall say that $G \preceq H$ if $\forall x \in \mathbb{R}^n$ we have that

$$\sum_{e(i,j) \in E(G)} W_e(i,j)(x_i - x_j)^2 \leq \sum_{e(i,j) \in E(H)} W_e(i,j)(x_i - x_j)^2$$

where $e(i,j)$ is the multiedge with end points i and j.

Suppose we now take each edge in G and partition its weight amongst some multiple copies of the edge, giving a multigraph H.

1. How is the spectral equivalence effected?
2. If we plan on using the procedure SAMPLE from lecture where edge e is sampled with frequency p'_e what should be the sampling frequency for the edges of H be to preserve spectral equivalence?
3. How does the two sparsified graphs for G and H compare when using AW?

(25) 2. **Leverage Scores and Resistors**

Recall that much of this class has focused on the theorems regrading effective resistance of graphs. The goal of this problem is to determine if we can generalized these theorems

[1] Corrected typo in problem 2-6.

to arbitrary matrices. Let B be the edge by vertex matrix of a connected graph G with a diagonal conductance matrix C then the Laplacian of G is B^TCB. If b_i is the ith row of B corresponding to the ith edge e_i of G then the effective resistance from one end of e_i to the other is $b_i(B^TCB)^\dagger b_i^T$. If we define $\bar{B} = C^{1/2}B$ then the resistance is $b_i(\bar{B}^T\bar{B})^\dagger b_i^T$. This motivates to following definition. Let $A^{m\times n}$ be a matrix of rank n.

Definition 0.1. *The* **Leverage Score** $\sigma(A,a)$ *where a is a column vector of size n is* $a^T(A^TA)^\dagger a$ *where $\dagger$ is the pseudoinverse. If a_i is the ith row of A the* $\sigma_i(A) = \sigma(A, a_i^T)$.

The goal of this problem is to determine what if any of the properties of effective resistance carry over the leverage scores.

1. Show that the leverage score of a nonzero row vector with itself is one.
2. Show that the column space $Col(A)$ of A and the left null space $Null_L(A)$ of A form an orthogonal bases of R^m. We will think of these vectors in $f \in \mathbb{R}^m$ as the flows. In the case when $A = \sqrt{C}B$ what kind of flows are the $Col(A)$ and the $Null_L(A)$?
3. We next prove a generalization of Foster's Theorem. Show that the sum of the row leverage scores of A is $rank(A)$. In this problem assume that A is not of full rank.
4. We next prove a generalization of Thomson's Principle. Suppose that x is a solution to the system $A^TAx = b$ were b is in the column space of A^T. Show that the flow $f = Ax$ is the unique minimum energy flow such that $A^Tf = b$. We define the energy of f to be f^Tf.
5. We next prove a generalization of Rayleigh's Monotonicity Law. If we increase a row of A by scaling it by $1+c$ for $c > 0$ or add a new row then no leverage score except the changed one will increase.
6. We next prove a generalization of Spielman-Srivastava Graph Sparsification Theorem. We say that $A^TA \approx_\epsilon B^TB$ if $(1-\epsilon)A^TA \preccurlyeq B^TB \preccurlyeq (1+\epsilon)A^TA$. Prove that that there exist a matrix QB where B is a subset of m' rows of A and Q is a nonnegative diagonal matrix where $m' = cn\log n$ for some constant c and $A^TA \approx_\epsilon B^TQ^TQB$.
7. Prove a variant of the fact that conductors add when placed in parallel. In particular prove a relationship between $\sigma(a,a)$, $\sigma(A,a)$, and $\sigma(\bar{A},a)$ where $\bar{A}$ is the matrix A with row a appended.

   Hint: Consider the Sherman-Morrison formula.

   Can you find a more general formula?

Research questions:

1. We also showed that the effective resistance forms a metric over the **vertices**. Thus in our generalization we should be looking for a metric on the columns of A. Suppose we define the score between two columns as $D_{ij} = \chi_{ij}^T(A^TA)^{-1}\chi_{ij}$ where χ_{ij} is the column vector with a 1 and -1 at i and j respectively. Does our score say anything interesting about the relationship of two columns? Is D_{ij} a metric on the columns of A.

2. Is there theory of random walks for leverage scores, either on the columns or rows of A?
3. If there in such a theory as randoms walks does commute time make sense and is related to leverage score?

(25) 3. **Recurrent Random Spanning Trees**

In class in order to prove Markov Chain Tree Theorem we need to consider a random walk over rooted (convergent) trees in a strongly connected direct graph G. In particular, we defined a random walk on trees given a random walk on the vertices of G.

Show that the walk on rooted trees defined by the **Last Visited** tree has a single recurrent class.