CS 599B: Math for TCS, Spring 2022

Exercise Set 11

Due: 10:00PM, Monday May 2, 2022 on Gradescope.

Instructions. You're encouraged to discuss these problems with other students as you solve them, but the written solutions you hand in must be in your own words. Don't worry about polishing the presentation of your solutions; these are primarily intended to keep you thinking about and engaged with the material.

Problem 1 (Correlation clustering)**.** In the correlation clustering, you are given a graph $G = (V, E)$ and two edge weight functions $w_{in}, w_{out} : E \to \mathbb{R}_{\geq 0}$. For a partition (clustering) $S = \{S_1, \ldots, S_k\}$ of the vertices, let $I(S)$ be the set of edges that connect two vertices within the same cluster, and let $O(S)$ be the set of edges connecting two vertices in different clusters. The goal is to find a clustering S that approximately maximizes

$$val(S) = \sum_{e \in I(S)} w_{in}(e) + \sum_{e \in O(S)} w_{out}(e).$$

(a) Show the following quadratic integer program exactly captures the correlation clustering problem.

$$\begin{aligned} \max \sum_{(u,v) \in E} & w_{in}(u,v)\langle x_u, x_v\rangle + w_{out}(u,v)(1 - \langle x_u, x_v\rangle) \\ \text{s.t. } & x_v \in \{e_1, \ldots, e_n\} && \forall v \in V \end{aligned}$$

where $e_i \in \mathbb{R}^n$ is the vector with a 1 in the ith index and 0s everywhere else.

(b) Explain why

$$\begin{aligned} \max \sum_{(u,v) \in E} & w_{in}(u,v)\langle x_u, x_v\rangle + w_{out}(u,v)(1 - \langle x_u, x_v\rangle) \\ \text{s.t. } & \langle x_v, x_v\rangle = 1 && \forall v \in V \\ & \langle x_u, x_v\rangle \geq 0 && \forall u \neq v \in V. \end{aligned}$$

is an SDP relaxation of the program from part (a).

(c) Let $(x_u)_{u \in V}$ maximize the SDP in part (b). Partition V into 4 clusters by splitting up the unit sphere using two independent random hyperplanes. Show that the expected value of this clustering is at least $\frac{3}{4} \cdot \max_S val(S)$.

You can use without proof the facts that $(1 - \theta/\pi)^2 \geq \frac{3}{4}\cos\theta$ and $1 - (1 - \theta/\pi)^2 \geq \frac{3}{4}(1 - \cos\theta)$ for all $\theta \in [0, \pi/2]$.

Problem 2 (Sherali-Adams for Clique)**.** You may optionally turn in a carefully written solution to this problem by **10PM on Wednesday, May 4th** to contribute to your homework grade.

Let D be the distribution over n-vertex graphs such that each edge is included independently with probability $1/2$. Recall (Exercise 4.3a) that with high probability, G does not contain a clique of size greater than $2 \log n$. Given a graph $G = ([n], E)$, consider the integer program:

$$\begin{aligned} \max \sum_{i \in [n]} z_i & \\ \text{s.t. } z_i + z_j \leq 2 & \qquad \forall (i,j) \in E \\ z_i + z_j \leq 1 & \qquad \forall (i,j) \notin E \\ z_i \in \{0,1\} & \qquad \forall i \in [n]. \end{aligned}$$

Show that with high probability over $G \sim D$, the level-$O(\log n)$ Sherali-Adams relaxation of this IP can certify that G does not contain a size-$100 \log n$ clique. That is, the level-$O(\log n)$ SA relaxation has value at most $100 \log n$ with high probability.