CS 5968/6968: Mid-term examination

Date: March 8, 2016, **Duration:** 1hr, 20 min

NAME: [solutions]

UID :

Rules: You are allowed to reference any course material that you bring with you, but using a laptop is not allowed (ask instructor if you want to use it only to look up course notes). Please write down the solutions in the space provided below the questions. Attaching a rough sheet with your name/UID is OK, but shouldn't be necessary.

Problem 1 (15 points)

Please answer each of the following questions in a couple of lines each.

(a) (3 points) Let $X_1, X_2, \ldots, X_n$ be 0/1 random variables that take value 1 with probability p. If $p < 1/n$, show that the probability that none of the variables is 1 is non-zero.

By a union bound, $\Pr[(X_1 = 1) \vee (X_2 = 1) \vee \cdots \vee (X_n = 1)] \leq \sum_i \Pr[X_i = 1] < 1$. Thus the probability that none of the X_i is 1 is > 0.

(b) (3 points) In the above question, suppose the X_i are all *independent*. Then show that for *any* $p < 1$, the probability that none of the variables is 1 is non-zero.

Because of independence, we have

$$\Pr[(X_1 = 0) \wedge (X_2 = 0) \wedge \cdots \wedge (X_n = 0)] = \prod_{i=1}^{n} \Pr[(X_i = 0)] = (1-p)^n > 0.$$

(c) (3 points) Consider a complete binary tree with 2^k nodes (figure on the board for $k = 3$). All nodes have degree $1, 2$, or 3. Show that for any subset S of the vertices with $|S| \leq n/2$,

$$\frac{E(S, \overline{S})}{|S|} \geq \frac{2}{n}.$$

The graph is connected, so for any subset of the vertices S, $E(S, \overline{S}) \geq 1$. Thus for $|S| \leq n/2$, we have

$$\frac{E(S, \overline{S}}{|S|} \geq \frac{1}{(n/2)} = \frac{2}{n}.$$

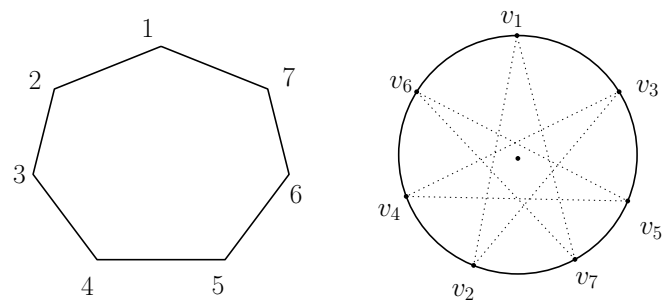

Figure 1: Cycle on 7 vertices, and its SDP solution.

(d) (6 points) Consider the max-cut problem when the graph is a cycle of length 7, and consider the solution to the semidefinite program shown in Figure 1. (The vectors are all unit length, in two dimensions.) Consider any hyperplane through the center (in this case, it is simply a line through the center). How many edges does this cut? What is the maximum cut in the graph?

Every line through the origin will cut exactly 6 edges. This is because it will split the graph into two sets of vertices, one of size 4 and the other of size 3. The smaller side will not have any edges, while the larger has precisely one. Since there are 7 edges total, the cut size is 6.

The max cut cannot have size 7 (odd cycle), thus it has size ≤ 6. The above shows it's equal to 6.

Problem 2 (10 points)

Let G be a graph on n vertices, with every vertex having degree d. Let us show, via the probabilistic method, that there exists an independent set of size $n/9d$. Suppose we sample every vertex independently with probability $1/3d$, and consider the induced subgraph G' on the sampled vertices. (I.e., the vertex set is V' – the set of sampled vertices, and the edge set is E' – all the edges in G both of whose end points are in V'.)

(a) (2 points) What is the expected size of V'?

Let X_i be an indicator for i being picked. Then $\mathbb{E}[X_i] = 1/3d$. Thus

$$\mathbb{E}[|V'|] = \mathbb{E}[\sum_i X_i] = \sum_i \mathbb{E}[X_i] = n/3d.$$

(b) (2 points) What is the expected size of E'?

Every edge ij exists in the sample iff both the end points i, j are picked. This happens with probability $\frac{1}{9d^2}$. Let X_{ij} be the indicator for this event. Then, as before,

$$\mathbb{E}[|E'|] = \mathbb{E}[\sum_{ij \in E} X_{ij}] = \sum_{ij \in E} \frac{1}{9d^2} = \frac{nd}{18d^2} = \frac{n}{18d}.$$

We used the fact that the total number of edges is $nd/2$ in a d-regular graph.

(c) (4 points) Show that there exists a subgraph that satisfies

$$|V'| - 2|E'| \geq \frac{n}{9d}. \qquad \text{(Bonus: point out the mild subtlety.)}$$

Putting together the two expectations above, we have

$$\mathbb{E}[|V'| - 2|E'|] = \frac{n}{3d} - \frac{n}{9d} = \frac{2n}{9d} > n/9d.$$

Thus there exists a subgraph (V', E') such that $|V'| - 2|E'| > n/9d$.
(The subtle point is that we cannot use parts (a) and (b) to 'directly' conclude that there exists a graph in which $|V'| \geq n/3d$ **and** $|E'| \leq n/18d$. Taking the difference is a trick to do that.)

(d) (2 points) Conclude that there exists an independent set of size $n/9d$.

Consider the graph (V', E'), and simply remove all the vertices that have an edge incident to them. The number of vertices removed is at most $2|E'|$. Thus there are at least $|V'| - 2|E'|$ vertices remaining, and they form an independent set.
Part (c) then gives the size bound.

Problem 3 (15 points)

Consider a dumb-bell graph D_n, which is a graph on n vertices, that are divided into two subsets L and R of $n/2$ vertices each. Every two vertices in L are connected by an edge, as are every two vertices in R. Additionally, there is one edge between L and R, that goes between two vertices which we call ℓ and r.

Consider a particle that starts at some vertex in the graph, and does a simple random walk for $T + n$ steps, where T is large ($> n^6$, say, so we assume that the walk has mixed completely). Let us denote by X_i the random variable that is 1 if <u>at time step $T + i$</u>, the particle is at some vertex in L, and 0 if it is at a vertex in R.

Let us consider $X = X_1 + X_2 + \cdots + X_n$.

(a) (3 points) What is $\mathbb{E}[X]$?

We assumed that the walk has mixed completely by time T. Thus by symmetry, we have that for each i, the probability of the particle being at one of the L vertices is the same as it being in one of the R vertices. Thus $\mathbb{E}[X_i] = 1/2$ for all i. Thus

$$\mathbb{E}[X] = \sum_i \mathbb{E}[X_i] = n/2.$$

(b) (6 points) How does the distribution of X look like? In particular is it *concentrated* around $\mathbb{E}[X]$? Argue intuitively.

The key observation is that if $X_1 = 1$, then it is extremely likely that $X_2 = 1$ (i.e., if the particle was at an L vertex at time $T + 1$, even if it is at the vertex ℓ, it's likelihood of taking the edge to r is $< 2/n$. In fact, if $X_1 = 1$, the expected number of steps needed to move to R is roughly n^2. Thus $X_2, X_3, \ldots, X_n$ will all be 1, with good probability. Thus X will be n.

Similarly, if $X_1 = 0$, then with good probability, X_i will be 0 for all $i \leq n$. Thus X is essentially distributed as n with probability 1/2 and 0 w.p. 1/2. It is *not* concentrated around $n/2$.

(c) (6 points) Let us make the above more formal. Suppose $Y_i = X_i - \mathbb{E}[X_i]$. Then give a lower bound for $\mathbb{E}[(Y_1 + Y_2 + \ldots Y_n)^2]$ (note that this quantity is the variance of X).
(*HINT:* find a way to lower bound Y_iY_j for $i \neq j$.)

We have

$$\mathbb{E}[(\sum_i Y_i)^2] = \sum_i \mathbb{E}[Y_i^2] + 2\sum_{i<j} \mathbb{E}[Y_iY_j].$$

Since $\mathbb{E}[X_i] = 1/2$, we have $Y_i = X_i - 1/2$. Thus, each Y_i is either $1/2$ or $-1/2$, and each occurs w.p. 1/2. Thus the first term is $\Theta(n)$. The second term is more interesting. Fix some $i < j$ and consider $\mathbb{E}[Y_iY_j]$.

$$\mathbb{E}[Y_iY_j] = (1/2)\mathbb{E}[Y_iY_j \mid Y_i = 1/2] + (1/2)\mathbb{E}[Y_iY_j \mid Y_i = -1/2].$$

We will show that each of the $\mathbb{E}[Y_iY_j \mid Y_i = \cdots]$ terms is $\geq 1/4 - O(1/n)$. Let us denote $n' = n/2$, for convenience, and let us condition on $Y_i = 1/2$, i.e., $X_i = 1$. (The other case is symmetric.) Let pos_r denote the position of the particle at time $T + r$. The key definitions are following: (for $r \geq i$)

$$a_{r+1} = \Pr[pos_{r+1} \in L \setminus \{\ell\} \wedge (X_i = X_{i+1} = \cdots = X_r = 1)], \text{ and}$$
$$b_{r+1} = \Pr[pos_{r+1} = \ell \wedge (X_i = X_{i+1} = \cdots = X_r = 1)].$$

It is easy to see that we have the recurrences:

$$a_{r+1} = a_r \cdot \frac{n'-2}{n'-1} + b_r \cdot \frac{n'-1}{n'} \;;\quad b_{r+1} = a_r \cdot \frac{1}{n'-1}. \tag{1}$$

Thus, we can inductively show that for any $j > i$ (a) $b_j \leq \frac{1}{n'-1}$, and (b) $a_j + b_j \geq a_i + b_i - \frac{j-i}{n'(n'-1)}$.

Part (a) is easy, from (1). To see part (b), note that $a_{r+1} + b_{r+1} = a_r + b_r - \frac{b_r}{n'} \geq a_r + b_r - \frac{1}{n'(n'-1)}$.

If we condition on $Y_i = 1/2$, or equivalently $X_i = 1$, we get $a_r + b_r \geq 1 - \frac{r}{n'}n' - 1$, which for $r \leq n$, says that $a_r + b_r \geq 1 - O(1/n)$. Thus for $j - i \leq n$ (as in our setting), we have $\Pr[X_j = 1 \mid X_i = 1] = 1 - O(1/n)$. This gives $\mathbb{E}[Y_i Y_j | Y_i = 1/2] \geq 1/4 - O(1/n)$, as claimed.

Problem 4 (10 points)

Let us develop an algorithm that approximately computes the minimum distance from a convex set K to a point p. Denote this minimum distance by $d(p, K)$. Suppose we have an efficient separation oracle $\mathcal{A}$ for K, and suppose we are given an R, with the guarantee that $R \leq d(p, K) \leq 10R$.

The goal is to compute $d(p, K)$ to within a $(1 + \epsilon)$ factor, for any constant $\epsilon > 0$.

(a) (4 points) Show a seemingly unrelated observation: if two convex sets $K, K' \subseteq \mathbb{R}^n$ have efficient separation oracles, then $K \cap K'$ has an efficient separation oracle.

Given a point x, we can simply run the two separation oracles. If both of them say $x \in K$ (and $x \in K'$), we can return that $x \in K \cap K'$. Else we return the separating plane output by either of the oracles (that is a valid separator because $K' \cap K$ is a subset of both K and K'.

(b) (6 points) Give the outline of an efficient algorithm to compute $d(p, K)$ using this idea (you do not need to go into the full details of the proof).

The observation is that $\text{Ball}(p, r) \cap K$ is non-empty iff $d(p, K) \geq r$. For any given r, we can use the ellipsoid algorithm to check non-emptiness (because there is an efficient separation oracle, by the above).

How do we make a small number of checks? We just use binary search: we know that $R \leq d(p, K) \leq 10R$ to start with. At some point, if we know $A \leq d(p, K) \leq B$, we can run the check above with $r = (A + B)/2$, and if $d(p, K) \geq (A + B)/2$, we recurse with the bounds $((A + B)/2, B)$, and if not, with $(A, (A + B)/2)$. We always reduce the length of the *candidate* interval by $1/2$, thus in $\log(9/\epsilon)$ iterations, we will obtain an interval of length $< \epsilon R$.

This means we have a good approximation.