CS264: Homework #6

Due by midnight on Thursday, February 23, 2017

Instructions:

(1) Form a group of 1-3 students. You should turn in only one write-up for your entire group. See the course site for submission instructions.

(2) Please type your solutions if possible and feel free to use the LaTeX template provided on the course home page.

(3) All students should complete all of the exercises. Students taking the course for a letter grade should complete Problems 22–24.

(4) Write convincingly but not excessively. Exercise solutions rarely need to be more than 1-2 paragraphs. Problem solutions rarely need to be more than a half-page (per part), and can often be shorter.

(5) You may refer to your course notes, and to the textbooks and research papers listed on the course Web page *only*. You cannot refer to textbooks, handouts, or research papers that are not listed on the course home page. (Exception: feel free to use your undergraduate algorithms textbook.) Cite any sources that you use, and make sure that all your words are your own.

(6) If you discuss solution approaches with anyone outside of your team, you must list their names on the front page of your write-up.

(7) Exercises are worth 5 points each. Problem parts are labeled with point values.

(8) No late assignments will be accepted.

Lecture 11 Exercises

Exercise 34

Recall the heuristic for finding a planted k-clique that simply returns the k vertices with the highest degrees. Consider a semirandom adversary A, which given a graph G with a planted k-clique Q and non-clique edge probabilities $\frac{1}{2}$, returns a graph $A(G)$ with the same k-clique Q and possibly fewer non-clique edges.

Prove that there is a constant $c > 0$ and a semirandom adversary such that, with high probability over the choice of G, the 'top k degrees" algorithm fails to return a clique of $A(G)$ when $k \leq cn$.

Exercise 35

Throughout our discussion of the planted bisection problem, we've been assuming that:

(i) p is at least some constant $c_1 > 0$ (independent of n);

(ii) for a sufficiently large constant $c_2 > 0$, $p - q \geq c_2\sqrt{\frac{\log n}{n}}$.

Where in our arguments in lecture did we use assumption (i)?

Lecture 12 Exercises

Exercise 36

Let $\mathbf{B}$ denote the adjacency matrix of an n-vertex graph (with n even) and $\mathbf{J}$ the $n \times n$ all-ones matrix. Fix a bisection (S,T) of V, and for $i \in V$ define

$$y_i = (\text{\# of } i\text{'s neighbors on the other side of } (S,T)) - (\text{\# of } i\text{'s neighbors on the same side of } (S,T)).$$

Let $\mathbf{Y}$ denote the diagonal matrix with diagonal entries $y_1, \ldots, y_n$. Prove that

$$-\mathbf{B} + \mathbf{J} - \mathbf{Y}$$

is singular (i.e., has a zero eigenvalue).

Exercise 37

Recall the quadratic programming relaxation of the maximum independent set problem:[1]

$$\max \sum_{i,j \in V} z_i z_j$$

subject to:

$$\begin{aligned} z_i z_j &= 0 \qquad \text{for every } (i,j) \in E \\ \sum_{i \in V} z_i^2 &= 1 \end{aligned}$$

Prove that a feasible solution z to this quadratic program is optimal if and only if it has the following form: for some maximum independent set S of the graph,

$$z_i = \begin{cases} \frac{1}{\sqrt{|S|}} & \text{if } i \in S; \\ 0 & \text{otherwise.} \end{cases}$$

Exercise 38

Assume the result in lecture, that for $k \geq c\sqrt{n}$ (with c a sufficiently large constant), the optimal value of the SDP relaxation for the maximum independent set problem equals the actual size of the planted independent set (with high probability over the planted instance).

Prove that this result continues to hold in the semirandom model, with an arbitrary monotone adversary (who can add any edges that it wants, except for edges inside the planted independent set).

Problems

Problem 22

For both of the following parts, you can assume that the planted solution is the unique optimal solution (with high probability). You can also consider just the planted version of the problem, without a monotone adversary. When building on results, you may change the choice of constant(s) that the result holds for.

(a) (5 points) Assume the planted independent set recovery result from lecture: for $k \geq c\sqrt{n}$ (and sufficiently large c), with high probability over the input (say at least $1 - \frac{1}{n^3}$), the optimal value of the SDP relaxation is k. Building on this result, give a polynomial-time algorithm that recovers the planted independent set (and not just its value).

[1] Whenever we refer to a psd matrix, it goes without saying that the matrix is symmetric.

(b) (5 points) Assume the planted bisection recovery result from lecture: for $p \geq c_1$ and $p - q \geq c_2\sqrt{\frac{\log n}{n}}$ (and sufficiently large $c_1, c_2 > 0$), with high probability over the input (say at least $1 - \frac{1}{n^3}$), the optimal value of the SDP relaxation is the size of the planted bisection. Building on this result, give a polynomial-time algorithm that recovers the planted bisection (and not just its value).

Problem 23

(15 points) This problem considers a planted model for graph coloring, to complement the ones we saw in lecture for the minimum bisection and maximum clique problems. Fix an integer k (which you should view as a constant), a number $p \in (0, 1)$ (also constant), and an integer n (which you should think of as going to infinity). Consider generating a random k-colorable graph as follows:

1. Each vertex is independently given a label uniformly at random from $\{1, 2, \ldots, k\}$.
2. For each pair of vertices with endpoints with different labels, include the corresponding edge (independently) with probability p.

The ensuing graph is k-colorable with probability 1, with the color classes corresponding to the subsets of same-labeled vertices.

Design a polynomial-time algorithm that recovers the planted k-coloring in such a random graph with high probability (for large n).

Problem 24

This problem fills in the gap from our proof in lecture of weak SDP duality, namely that $\mathbf{A} \bullet \mathbf{B} \geq 0$ whenever $\mathbf{A}, \mathbf{B}$ are psd matrices. (Recall that $\mathbf{A} \bullet \mathbf{B}$ is defined as $\sum_{i,j} a_{ij} b_{ij}$—in effect, the inner product between $\mathbf{A}$ and $\mathbf{B}$ when viewed as vectors of length n^2.)

(a) (6 points) Prove that a symmetric $n \times n$ matrix $\mathbf{A}$ is a psd matrix if and only if there exist (correlated) random variables $Y_1, \ldots, Y_n$, on a common state space, such that

$$a_{ij} = \mathbf{E}[Y_i \cdot Y_j]$$

for every $i, j \in \{1, 2, \ldots, n\}$.

[Hint: recall that a symmetric matrix $\mathbf{A}$ is psd if and only if it can be written as $\mathbf{X}^T\mathbf{X}$ for some $n \times n$ matrix $\mathbf{X}$.]

(b) (5 points) Define the *Hadamard product* (also called *Schur product*) $\mathbf{A} \circ \mathbf{B}$ as the matrix $\mathbf{C}$ obtained via entrywise multiplication, i.e., with $C_{ij} = A_{ij} B_{ij}$ for every $i, j \in \{1, \ldots, n\}$. Prove that the Hadamard product of psd matrices is again psd.

[Hint: use (a), with the random variables that define $\mathbf{A}$ defined to be independent of those that define $\mathbf{B}$.]

(c) (4 points) Conclude that if $\mathbf{A}, \mathbf{B}$ are psd matrices, then $\mathbf{A} \bullet \mathbf{B} \geq 0$.

Problem 25 (Extra Credit)

(10 extra credit points) Consider the planted bisection problem with $p = \frac{1}{2}$ and $p - q = \frac{c}{\sqrt{n}}$. Prove that for all sufficiently small constants c, there is no algorithm (polynomial-time or otherwise) that recovers the planted bisection (S, T) with high probability (i.e., probability approaching 1 as $n \to \infty$).

[Thus, our recovery result in Lecture #12 achieves the minimum-possible gap between p and q, even for the standard planted model (without an adversary), and even information-theoretically.]