CSE 200 – Fall 2021
Homework 3
Due Monday, November 15th, 11:59pm

Let G=(V,E) be an undirected graph on n nodes and m edges. To simplify notations, we will assume that $V = [n]$, where $[n] = \{1, \ldots, n\}$ and $[0] = \{\}$. A subset of nodes $S \subset V$ defines a cut $E(S) = \{(u, v): u \in S, v \notin S\}$ in the graph. Define $e(S) = |E(S)|$ to be the number of edges in the cut.

The MAX-CUT problem is to find the maximal cut in the graph. It is known to be NP-complete, so we will try to approximate the solution. Concretely, we will consider a 2-approximation: outputting a cut which has at least ½ of number of edges in a maximal cut.

Problem 1: Randomized 2-approximation of MAX-CUT

We give a randomized algorithm that (on expectation) gives a 2-approximation.

Algorithm:
Choose the cut S randomly as follows: for every node $v \in V$ include $v \in S$ independently with probability ½.

(10 points) Prove that if the graph has m edges, then the expected value of $E_S[e(S)]$ over this random choice is $m/2$. Argue why this is at least ½ of the MAX-CUT value in the graph.

Problem 2: Derandomized MAX-CUT algorithm using conditional expectations

We now derandomize this algorithm. To do so, we will use the method of **conditional expectations.** We will construct S in n steps, where at the i-th step we will decide whether to include node $i \in S$ or not.

In order to model this, for $t = 0, \ldots, n$ we denote by $S_t = S \cap [t]$ the set of nodes out of $[t]$ that happen to be in S. Given $S_t \subseteq [t]$, consider the randomized algorithm which fixes S_t and randomly chooses for each $i = t + 1, \ldots, n$ whether to include i in S or not independently with probability ½. Denote $e(t, S_t)$ the expected value of the cut obtained in this way.

(a) (3 points) Show that $e(0, \{\}) = m/2$.
(b) (4 points) Prove that for every $t = 0, \ldots, n$ and every $S_t \subseteq [t]$, the value $e(t, S_t)$ can be computed in deterministic polynomial time (hint: use linearity of expectation).
(c) (4 points) Prove that for every $t = 0, \ldots, n - 1$ and every $S_t \subseteq [t]$, there exists $S_{t+1} \subseteq [t + 1]$ such that $S_t \subseteq S_{t+1}$ and $e(t + 1, S_{t+1}) \geq e(t, S_t)$.
(d) (4 points) Complete the proof, and show that a 2-approximation of the MAX-CUT value can be found in deterministic polynomial time.

Problem 3: Pairwise independence

A random variable $X \in \{0,1\}^n$ is called pairwise-independent if for all $1 \le i < j \le n$ and all $a, b \in \{0,1\}$, $Pr[X_i = a \text{ and } X_j = b] = 1/4$. That is, the restriction of X to any 2 coordinates is uniformly distributed in $\{0,1\}^2$.

A function $H: \{0,1\}^r \to \{0,1\}^n$ is pairwise independent if the random variable $X = H(U)$ is pairwise independent, where $U \in \{0,1\}^r$ is uniformly chosen. In this case, we say that X has seed length r. We will show a construction of such a H with seed length $r = \log n + O(1)$. H is called a "pseudorandom generator".

Assume that $n = 2^k$. We will define a function $H: \{0,1\}^r \to \{0,1\}^n$, where we identify the coordinates of the output of H with $\{0,1\}^k$, which is the binary expansion of the coordinate. Let $U \in \{0,1\}^{k+1}$ be uniformly chosen. Write $U = u_1, \ldots, u_{k+1}$ where $u_i \in \{0,1\}$. Define $H(U) \in \{0,1\}^n$ as follows. For every $x \in \{0,1\}^k$, the x-coordinate of $H(U)$ is defined to be

$$H(U)_x = \left(\sum_{i=1}^{k} u_i x_i\right) + u_{k+1} \qquad (mod\ 2)$$

(a) (5 points) Prove that $H(U)$ is pairwise independent.

(b) (5 points) Prove that for general n (not necessarily a power of 2) this can be used to give a pairwise independent random variable $X \in \{0,1\}^n$ with seed length $r = \log n + O(1)$.

(c) (5 points) Prove that the construction is optimal: for any $H: \{0,1\}^r \to \{0,1\}^n$ which is pairwise independent, it must hold that $r \ge \log n$.
Hint: consider the $2^r \times n$ matrix $M_{u,i} = (-1)^{H(u)_i}$. Prove that its columns are pairwise orthogonal. Conclude that the columns must be linearly independent, and hence $2^r \ge n$.

Problem 4: Derandomized MAX-CUT algorithm using pairwise independence

Recall the randomized algorithm in Problem 1. For $x \in \{0,1\}^n$ define its associated set $S(x) = \{v_i : x_i = 1\}$. One way to interpret the randomized algorithm in question 1 is that if $X \in \{0,1\}^n$ is chosen uniformly, then the expected value of the cut is $E_X[e(S(X))] = m/2$.

(a) (5 points) Prove that if $X \in \{0,1\}^n$ is chosen from a pairwise independent distribution then also $E_X[e(S(X))] = m/2$.

(b) (5 points) Combine this with the construction from problem 3, to give an alternative deterministic algorithm which finds a 2-approximation of the MAX-CUT value. Briefly justify your algorithm's correctness.