Problem Set II

Solutions to many homework problems, including problems on this set, are available on the Internet or can be obtained by an LLM, either for exactly the same problem formulation or for some minor perturbation. It is *not acceptable* to copy such solutions. It is hard to make strict rules on what information from the Internet you may use and hence whenever in doubt contact Michael Kapralov. You are, however, allowed to discuss problems in groups of up to three students.

1 PTAS for MaxCut. *(34 pts)* In this problem, you will design and analyze a *polynomial-time approximation scheme*[1] for the MaxCut problem restricted to graphs that are "dense" and "almost bipartite". Hereafter, let $G = (V, E)$ be an undirected graph on n vertices. We denote by $\textsc{opt}(G) = \max_{U \subseteq V} |E(U, V \setminus U)|$ the maximum size of a cut, and let $Q^* \subseteq V$ be a maximum cut of G, i.e. $|E(Q^*, V \setminus Q^*)| = \textsc{opt}(G)$. For any cut $U \subseteq V$, we use $\mathbb{1}_Q$ to denote the vector in $\{\pm 1\}^V$ with $(\mathbb{1}_U)_u = 1$ if $u \in U$ and $(\mathbb{1}_U)_u = -1$ if $u \notin U$.

1a *(10 pts)* For any assignment $\sigma \in \{\pm 1\}^V$, we define for each $u \in V$ the *bias* of u with respect to σ as $\beta(u; \sigma) = -\sigma_u \sum_{v \in N(u)} \sigma_v$. Furthermore, for a tuple of t vertices $S = (s_1, \ldots, s_t) \in V^t$, we define the *$S$-projected bias* of u with respect to σ as $\hat{\beta}_S(u; \sigma) = -\sigma_u \sum_{i \in [t] \colon s_i \in N(u)} \sigma_{s_i}$.

Consider sampling S by picking t vertices uniformly at random from V with replacement. Show that for any assignment $\sigma \in \{\pm 1\}^V$ and any vertex $u \in V$ such that $\beta(u; \sigma) > 0$, we have

$$\Pr_S \left[\hat{\beta}_S(u; \sigma) \leq 0 \right] \leq \exp\left(-\frac{t \cdot (\beta(u; \sigma))^2}{2n^2} \right) .$$

Hint: for m independent random variables $Y_1, \ldots, Y_m$ over $[-1, 1]$, and any $\theta > 0$, we have $\Pr[\sum_{i=1}^m (Y_i - \mathbb{E}[Y_i]) \geq \theta] \leq \exp(-\theta^2/(2m))$ (a.k.a. Hoeffding's inequality for independent bounded variables).

1b *(10 pts)* Show that for any cut $Q \subseteq V$, one has

$$\textsc{opt}(G) - |E(Q, V \setminus Q)| \leq \sum_{u \in V \colon (\mathbb{1}_{Q^*})_u \neq (\mathbb{1}_Q)_u} \beta(u; \mathbb{1}_{Q^*}) + 2\left(|E| - \textsc{opt}(G)\right) .$$

[1] The term "polynomial-time approximation scheme" (PTAS, for short) generically refers to a polynomial-time algorithm which computes a $(1-\epsilon)$-approximate solution (or $1+\epsilon$, for minimization problems) in polynomial time for any constant $\epsilon > 0$.

1c *(14 pts)* Fix a constant $c > 0$. Describe and prove the correctness of a randomized algorithm that, given an accuracy parameter $\epsilon > 0$ and an undirected n-vertex graph $G = (V, E)$ with $|E| \geq cn^2$ (i.e. G is *dense*) and $\textsc{opt}(G) \geq (1-\epsilon)|E|$ (i.e. G is *almost bipartite*), runs in time $2^{O(\epsilon^{-2}\log(1/\epsilon))}\mathrm{poly}(n)$ and outputs a cut $Q \subseteq V$ such that

$$\Pr\left[|E(Q, V \setminus Q)| \geq (1 - O(\epsilon))\textsc{opt}(G)\right] \geq \frac{2}{3}\,.$$

Hint: the analysis might require arguing that for every $u \in V$ one has $\beta(u; \mathbb{1}_{Q^*}) \geq 0$ (that is, very $u \in V$ has at least as many neighbors on the opposite side of Q^* as compared to the side where u lies).

2 Knapsack. *(30 pts)* We want to come up with an approximation algorithm for the Knapsack problem. In this problem we are given n items with prices $p_1, \ldots p_n$, weights $w_1, \ldots, w_n$ and a capacity bound W. We want to pack a subset of items $S \subseteq [n]$ that does not exceed the capacity bound and achieves a maximum profit. More formally, we would like to find $\max_{S \subseteq [n]} \sum_{i \in S} p_i$ subject to $\sum_{i \in S} w_i \leq W$. Design and analyze a $(1-\epsilon)$-approximation algorithm for the Knapsack problem that runs in time $n^{O(1/\epsilon)}$ for any positive constant $\epsilon > 0$.

Hint: Guess the $\frac{1}{\epsilon}$ most profitable items in the optimal solution and than use a "bang-for-the-buck" greedy strategy.

3 **Bin Packing.** *(36 pts)* In this problem we want to study the bin packing problem. Here, we are given items $U = [n]$ of sizes $\{a_1, \dots a_n\}$ and a capacity bound $W \in \mathbb{N}$. We want to place items in bins of capacity B such that each item is contained in a bin. Moreover, we want to use the minimal number of bins to do so. A solution to this problem corresponds to a partition $B_1 \cup \dots \cup B_m = U$ of the items such each B_i respects the capacity constraint, $\sum_{j \in b_i} a_j \leq B$ for each $i \in [m]$. You can assume that each item fits in a bin, $a_i \leq W$ for all $i \in U$. This problem is NP-complete so we are interested in an efficient approximation algorithm. An important relaxation for the problem is the *configuration linear program.* A configuration C is a subset of items that respects the capacity bound W, i.e. $\sum_{i \in C} a_i \leq W$. In other words, C is a configuration of a bin we can use for the partition. Let $\mathcal{C} = \{C \subseteq U \mid \sum_{i \in C} a_i \leq W\}$ be the set of all possible configurations. The configuration LP will have one variable x_C for each configuration $C \in \mathcal{C}$ indicating whether we use the configuration C in the partition.

$$\text{minimize} \quad \sum_{C \in \mathcal{C}} x_C \tag{1}$$

$$\text{subject to} \quad \sum_{C \in \mathcal{C}:\, i \in C} x_C \geq 1 \qquad \forall i \in U \tag{2}$$

$$x_C \geq 0 \qquad \forall C \in \mathcal{C}. \tag{3}$$

Constraint (2) ensures each item a is placed in at least one (possibly fractional) bin. This relaxation is well studied and can be rounded to a solution of cost at most OPT + log(OPT). The standard methods for solving LPs run in exponential time in this case since the LP has exponentially many variables. Therefore, your task is to implement the Hedge strategy to obtain an algorithm that solves the configuration LP approximately in polynomial time. Here, the solution x can be described as a list of non-zero coordinates.

3a *(12 pts)* Show that the oracle for the reduced problem in your Hedge strategy can be solved up to a $(1 - \epsilon)$ multiplicative error in time $n^{O(1/\epsilon)}$.

Hint: Make use of the algorithm designed in Problem 2.

3b *(12 pts)* Prove that the solution you compute is a $(1 + \epsilon)$-approximation to the configuration LP. In other words, you find a solution x such that,

$$\sum_{C \in \mathcal{C}} x_C \leq (1 + \epsilon) \sum_{C \in \mathcal{C}} x_C^\star,$$

where $x^\star$ is an optimal solution to the configuration LP.

3c *(12 pts)* Bound the parameter ρ, i.e. bound the absolute value of the cost vectors in your hedge strategy, in order to achieve the overall runtime bound of $n^{O(1/\epsilon)}$. For example, one can achieve a bound of $\rho \leq n$.