This problem sheet will be updated through the course of the semester. The answers must be typed in LaTeX, and emailed to the instructor in a single pdf file. **Collaboration is limited to groups of 2. Clearly write the name of your collaborator, if any, in every answer. Even if you have a collaborator, you should write your answer on your own, and be able to fully explain it if asked to.**

Problem 1
The weighted set cover problem is defined as follows: given a collection of subsets of a universe of elements, where each set has a positive weight, find a minimum weight collection of sets that covers all the elements.

(a) Consider the following iterative algorithm for weighted set cover: in each step, select an uncovered element uniformly at random and pick a set covering it according to some probability distribution depending on the weights of these sets. What probability distribution will you choose so that this algorithm has an expected approximation factor of f, the maximum number of sets that an element belongs to? Give an analysis to prove your claim.

(b) Now, consider a local search algorithm for weighted set cover. In this algorithm, a solution is not only a collection of sets that covers all elements, but also a mapping from each element to one of the sets covering it. Define the *cost-share* of an element e in a solution as w_S/n_S, where w_S is the of the set S that e is mapped to, and n_S is the number of elements that map to S in this solution. A local move is defined as adding a new set to the solution and mapping a set of elements to it so that the cost-share for every element decreases by at least a factor of 2. Show that if a solution does not have a local move, then it has an approximation factor of $O(\log n)$, where n is the number of elements.

Problem 2
In the metric facility location problem, there exists a set of facilities F and a set of clients C on a distance metric d. Every facility i has an opening cost o_i, and the connection cost of a client j is the distance d_{ij} to its closest open facility i. The goal is to open a subset of facilities that minimizes the sum of their opening costs and the connection costs of the clients given this set of open facilities.

Consider a local search algorithm for this problem defined by three local moves: *open* a new facility, *close* an existing facility, or *swap* an existing facility for a new one. The algorithm terminates when there is no local move that can improve the solution. In such a solution, define the *cost-share* of a client j as $s_j = o_{\sigma(j)}/n_{\sigma(j)} + d_{\sigma(j)j}$, where $\sigma(j)$ is the closest open facility to j and $n_{\sigma(j)}$ is the number of clients for which $\sigma(j)$ is the closest open facility. Show that for a large enough constant A, the following holds for any facility i:

$$\sum_{j \in C} \max(s_j/A - d_{ij}, 0) \leq o_i.$$

Problem 3
Recall that a submodular set function is one that has the *decreasing marginal gain property*, i.e., f is submodular if and only if for all sets $A \subseteq B$ and all elements x,

$$f(A \cup \{x\}) - f(A) \geq f(B \cup \{x\}) - f(B).$$

Answer the following questions on the problem of maximizing a submodular function subject to the restriction that the selected set must contain no more than k elements, for some input parameter k.

(a) Recall the iterative greedy submodular maximization algorithm that adds the element with the maximum marginal gain in each step. We saw in class that this algorithm has an approximation factor of $1 - 1/e$ for monotone submodular functions. Give an example of a *non-monotone* submodular function for which this algorithm does *not* achieve a constant approximation. [Hint: directed graph cuts.]

(b) Now, consider a variation of the greedy algorithm, where instead of adding the element with the maximum marginal gain in each step, the algorithm adds an element chosen uniformly at random from a set S of elements defined as follows: if there are at least k elements that have positive marginal gain, then S contains the k elements that achieve the maximum marginal gain, else S contains all the elements that achieve a positive marginal gain. Show that this algorithm continues to have an approximation factor of $(1 - 1/e)$ for monotone, submodular functions.

(c) Now, show that the algorithm in (b) also achieves a constant approximation factor for non-monotone submodular functions.

Problem 4
Recall that in the metric traveling salesman problem (metric-TSP), the goal is to find a minimum length tour that visits all points in a metric space. Consider a variation of TSP where the goal is to find the minimum length *path* that visits all points in a metric space and has two designated points s and t as its endpoints. Give an algorithm for this problem with an approximation factor of $5/3$. [Hint: same intuition as Christofides' algorithm.]

Problem 5
The non-metric facility location problem is identical to metric facility location, except that the distances between client-facility pairs no longer satisfy the triangle inequality. Show that the non-metric facility location problem is asymptotically identical to the set cover problem with respect to its approximability. In other words, show that an α-approximation algorithm for set cover gives an $O(\alpha)$-approximation algorithm for non-metric facility location, and vice-versa.

Problem 6
Recall that in the multiway cut problem, we are given an undirected graph with non-negative edge costs and a set of vertices $s_1, s_2, \ldots, s_k$, and the goal is to find a subset of edges of minimum cost whose removal disconnects all pairs s_i, s_j, $i \neq j$.

(a) Write an LP relaxation for this problem that embeds each vertex into a simplex of k dimensions where the vertices $s_1, s_2, \ldots, s_k$ are fixed at the k corners of the simplex and the objective for each edge is the $||\cdot||_1$-distance between its ends scaled by the cost of the edge. (A simplex is defined as the set of points in the non-negative orthant that are at an $||\cdot||_1$ distance of 1 from the origin.)

(b) Consider the following rounding algorithm. Pick a radius r uniformly at random from $(0, 1)$. Next, the k dimensions are ordered in a uniform random permutation. Now, for each dimension in this order, all vertices not rounded yet that are contained in an $||\cdot||_1$-ball of radius r centered at the corner corresponding to this dimension are rounded to the vertex at this corner. All vertices not rounded in the first $k-1$ rounds are rounded to the last corner. Show that this algorithm has an approximation factor of $3/2$ for the mutiway cut problem.