Homework set 3

Due: Monday, June 16 by 1pm. Solve all five problems.

1. A multiset is a set with repeated elements allowed. So for multisets, $\{a, b, c\}$ is not the same as $\{a, b, a, c, b\}$. We consider an algorithm to decide whether two integer multisets are identical in the sense that each integer occurs the same number of times in both sets. Suggest a way to represent this as a problem involving the verification of a polynomial identity, and thereby obtain an efficient randomized algorithm.

   The problem can also be solved by sorting the two sets. Compare the sorting-based solution with the method using polynomial identity testing in terms of their efficiency. Discuss the relative merits of the two algorithms, considering issues such as the size of the integers being operated upon, and ways the algorithm can fail.

2. Given a complete Euclidean graph G we find an approximate TSP cycle through G as follows. We build a cycle C by adding vertices to C one at a time. We start by choosing any vertex y and letting $C = \{y\}$. y forms a single vertex cycle. Repeat the following step until all vertices of G have been added to C: Choose a vertex z not in the current cycle C and where the distance from z to any vertex in C is minimum. Find the vertex u in C whose distance to z is smallest. Let us assume that the vertex following u in C is v. Then replace the edge (u, v) in C by $(u, z), (z, v)$. Prove that this method results in a TSP cycle C whose cost is at most $2\mathrm{OPT}(G)$.

3. Design a polynomial-time algorithm for the k-center problem when the input consists of 1-dimensional points. Write the pseudo code and provide running-time analysis.

4. Consider a stream of elements that come one-at-a time. Design an algorithm that picks k elements from the already-seen set of elements such that the picked k elements are a *uniform random* sample of the elements seen so far.

5. Consider a set of points $X = \{x_1, ..., x_n\}$ such that $x_i \in \mathbb{R}^d$. Also assume a distance function $D()$ that gives the distance between two points and that is also a metric. We call the representative of X the point $x^* \in \mathbb{R}^d$ such that:

   $$x^* = \text{argmin} \sum_{x_i \in X} D(x^*, x_i).$$

   Show that there exists a point $x \in X$ such that

   $$\sum_{x_i \in X} D(x, x_i) <= 2 \sum_{x_i \in X} D(x_i, x^*).$$