CMSC 754: Spring 2026 Dave Mount

Homework 7

Due by the start of class on Tuesday, April 21. (Submissions will be through Gradescope.) Late homeworks are not accepted (unless an extension has been prearranged) so please turn in whatever you have completed by the due date. Unless otherwise specified, you may assume that all inputs are given in *general position.*

Problem 1. As mentioned in class, a WSPD is an efficient (approximate) representation for the complete graph on a set of points P. Another important structure is the complete *bipartite graph* on a pair of point sets. In this problem, we will explore this topic.

You are given two sets of points in $\mathbb{R}^d$, called R (for red) and B (for blue). Throughout, let $n = |R| + |B|$. A *bichromatic pair* is any pair of points (p, q), where $p \in R$ and $q \in B$. Given a parameter $s > 0$, define a *bichromatic* s*-WSPD* is a collection of pairs of subsets $\{(R_1, B_1), (R_2, B_2), \ldots\}$ such that

(i) $R_i \subseteq R$ and $B_i \subseteq B$,

(ii) R_i and B_i are s-well separated, and

(iii) for every bichromatic pair (p, q) there is exactly one pair (R_i, B_i) such that $p \in R_i$ and $q \in B_i$.

Given this definition, answer the following questions:

(a) Explain how to modify the standard WSPD algorithm given in class to produce a bichromatic s-WSPD for the sets R and B. (You can explain what changes to make to the algorithm given in class.)

(b) Show that the asymptotic running time and total size of your bichromatic WSPD construction are the same as for the standard WSPD construction.

(c) Explain how to enhance your WSPD construction so that, with the same time and space requirements, every well-separated pair (R_i, B_i) stores a pair of *representative points* $\text{rep}(R_i) \in R_i$ and $\text{rep}(B_i) \in B_i$, and additionally, each WSPD stores the *sizes* $|R_i|$ and $|B_i|$.

Problem 2. Given two sets R and B in $\mathbb{R}^d$, define the *average bichromatic distance* to be

$$\overline{D}(R, B) \;=\; \frac{1}{|R| \cdot |B|} \sum_{p \in R} \sum_{q \in B} \|p - q\|.$$

Present an efficient algorithm that, given R, B, and $0 < \varepsilon < 1$, computes an ε-approximation to the average bichromatic distance. That is, your algorithm should return value $\widehat{D}$ such that

$$\frac{\overline{D}(R, B)}{1 + \epsilon} \;\le\; \widehat{D} \;\le\; (1 + \varepsilon) \cdot \overline{D}(R, B).$$

Your algorithm should run in $O(n \log n + n/\varepsilon^d)$ time, where $n = |R| + |B|$. (**Hint:** Use the answer to Problem 1.)

Problem 3. A set P of n points in $\mathbb{R}^d$ in general position determines a set of $\binom{n}{2}$ distinct distances. Define $\Delta(P)$ to be this set of distances $\{\|p_i - p_j\| : 1 \leq i < j \leq n\}$. Given an integer k, where $1 \leq k \leq \binom{n}{2}$, we are interested in computing the kth smallest distance from this set. Normally, this would take $O(n^2)$ time, so let's consider a fast approximation algorithm.

Let $\delta(P, k)$ denote the exact kth smallest distance in $\Delta(P)$. Given $\varepsilon > 0$, a distance value x is an ε-approximation to $\delta(P, k)$ if

$$\frac{\delta(P, k)}{1 + \varepsilon} \;\leq\; x \;\leq\; (1 + \varepsilon)\delta(P, k).$$

Present an efficient algorithm to compute such a value x. Aim for a running time of $O(n \log n + n/\varepsilon^d)$. Justify your algorithm's correctness and derive its running time.

(**Hint:** The following utility may be of use. Suppose that you are given a set $X = \{x_1, \ldots, x_n\}$ of real numbers, where each number x_i is associated with a positive integer *multiplicity* m_i. Let $M = \sum_{i=1}^{n} m_i$. Let $\widehat{X}$ denote the multiset, where the element x_i occurs m_i times. Given any k, where $1 \leq k \leq M$, define the *weighted kth smallest element* to be the kth smallest of $\widehat{X}$. You may assume that you have access to an algorithm that computes the weighted kth smallest element of such a multiset $\widehat{X}$ in $O(n)$ time, irrespective of how large M is.)

Problem 4. (Optional–Ungraded) A desirable property of spanners is that every vertex should have constant degree.

(a) Show that there exists a set of points in the plane such that the WSPD-based spanner construction given in class results in at least one vertex of degree $\Omega(n)$. That is, the degree is at least cn for some $c > 0$, which might depend on the dimension and separation factor, but not on n. (**Hint:** The algorithm for constructing the quadtree did not specify how representatives are chosen in the quadtree. To obtain the worst case, you will need to select representatives carefully.)

(b) Show that it is possible to modify the WSPD-based spanner construction so that the number of edges is the same, but each vertex has constant degree (where the constant depends on the stretch factor). For this, you may make the simplifying assumption that the quadtree is not compressed and that every internal node has at least two nonempty children. (**Hint:** Show that in such a tree it is possible to distribute representatives so that each point occurs as the representative for at most a constant number nodes of the tree.)

Note: Challenge problems are not graded as part of the homework. The grades are recorded separately. After final grades have been computed, I may "bump-up" a grade that is slightly below a cutoff threshold based on these extra points. (But there is no formal rule for this.)

Challenge Problem: In this problem, we'll consider a quick-and-dirty method for computing a crude lower bound for the weight of the Euclidean minimum spanning tree in $\mathbb{R}^d$.

You are given a set of n points $P = \{p_1, \ldots, p_n\} \subset \mathbb{R}^d$. Let $\mathrm{emst}(P)$ denote the weight of P's Euclidean minimum spanning tree. Place the points of P in a unit square grid, that is,

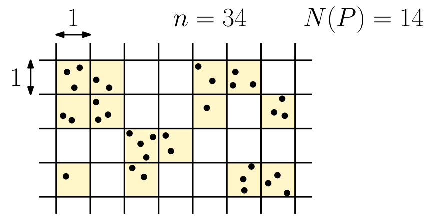

Figure 1: Lower bound for Euclidean minimum spanning tree.

a grid of hypercubes, each of side length 1 (see Fig. 1). Let $N(P)$ denote the number of grid squares that contain at least one point of P.

Prove that there exist positive constants a_d and c_d (depending on the dimension d, but not on n), such that if $N(P) \geq a_d$ then

$$\text{emst}(P) \;\geq\; \frac{N(P)}{c_d}.$$

It is not necessary to derive the best values of a_d and c_d, but as a hint, both quantities vary exponentially with d.

Hint: Show that there exists a subset of size $O(N(P))$ of grid squares where each contains a point of P and these grid squares are not adjacent to each other (that is, each pair is separated by a distance of at least one).