Problem 1. Let P be a set of n points in $\mathbb{R}^d$. Define d_{max} as the maximum distance between two points in P, namely, $d_{max} = \max_{p,q\in P} dist(p,q)$, where $dist(p,q)$ is the Euclidean distance between p and q. Describe an $O(n)$-time algorithm to find a d-dimensional box such that:

1. the box has the same side length on each dimension, and the side length is $\Theta(d_{max})$;
2. the box covers all the points in P.

Solution. Let $p_i = (x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(d)}), \forall p_i \in P$.

Then, define $p_{k,max} = \max\{p_i : x_i^{(k)}\}, p_{k,min} = \min\{p_i : x_i^{(k)}\}$.

Then the box is defined to be the region $[p_{1,min}, p_{1,max}] \times [p_{2,min}, p_{2,max}] \times \ldots \times [p_{d,min}, p_{d,max}]$.

Justification:

1. the side length is $O(\max_k\{p_{k,max} - p_{k,min}\}) = O(d_{max})$. Since

$$\begin{aligned} d_{max} &= \sqrt{(p_{1,max} - p_{1,min})^2 + (p_{2,max} - p_{2,min})^2 + \ldots + (p_{d,max} - p_{d,min})^2} \\ &\geq \sqrt{(\max_k\{p_{k,max} - p_{k,min}\})^2} \\ &= \max_k\{p_{k,max} - p_{k,min}\}. \end{aligned}$$

   Consider the 1D case where the points are scattered, the box constructed at least have side length $|p_{1,max} - p_{1,min}| = \Omega(d_{max})$. Therefore, each side length of the box is $\Theta(d_{max})$.

2. the box can cover all points in P since all the coordinates should lie within the range, this should be trivial.

Problem 2. Let P be a set of n points in $\mathbb{R}^d$. Define d_{max} as the maximum distance between two points in P, namely, $d_{max} = \max_{p,q\in P} dist(p,q)$. Define d_{min} as the minimum distance between two distinct points in P, namely, $d_{min} = \min_{\text{distinct } p,q\in P} dist(p,q)$. Describe how to build a quadtree on P in time $O(n \cdot \log \frac{d_{max}}{d_{min}})$ time.

Solution. The initial cell of the quadtree is a hypercube bounding P, which has the side length $O(d_{max})$. The following process is then repeated recursively. Consider any unprocessed cell and its associated node u in the current tree. If this cell contains either zero or one point of P, then this is declared a leaf node of the quadtree, and the subdivision process terminates for this cell. Otherwise, the cell is subdivided into 2^d hypercubes whose side lengths are exactly half that of the original hypercube. For each of these 2^d cells we create a node of the tree, which is then made a child of u in the quadtree.

Considering the process of shrinking the region of a cell, note that the closest pair with distance d_{min} can never be contained in the same hypercube with side length $d_{min}/\sqrt{d}$. The height of the quadtree constructed by the set P can never exceed $O(\log \frac{d_{max}}{d_{min}})$.

There are at most n nodes on each layer of the tree, and thus the quadtree construction time is bounded within $O(n \cdot \log \frac{d_{max}}{d_{min}})$.

Problem 3. Let P be a set of points in $\mathbb{R}^d$. Suppose that we have constructed an s-well separated point decomposition for $P : \{\{A_1, B_1\}, \{A_2, B_2\}, \ldots, \{A_m, B_m\}\}$. Let $\{p, q\}$ be a closest pair of P (i.e., $dist(p, q) = d_{min}$, where d_{min} is as defined in Problem 2). Prove: if $s > 2$, then there exists an $i \in [1, m]$ such that A_i contains only p, and B_i contains only q.
Hint: Recall that there must exist an $i \in [1, m]$ such that $p \in A_i$ and $q \in B_i$.

Proof. There must exist an $u \in [m]$ such that $p \in A_u, q \in B_u$.

Suppose that there exists another $r \in A_u$ and $r \neq p$. Then, since $dis(A_u, B_u) \geq sr > 2r$, $|pr| \leq 2d_{min} < sd_{min} \leq |pq|$, yielding a contradiction. Therefore, A_i contains only p. By symmetry, it must be that B_i contains only q. □

Problem 4. Let P be a set of points in $\mathbb{R}^d$. Suppose that we have constructed an s-well separated point decomposition for $P : \{\{A_1, B_1\}, \{A_2, B_2\}, \ldots, \{A_m, B_m\}\}$. For each $i \in [1, m]$, let a_i be an arbitrary point in A_i, and b_i be an arbitrary point in B_i. Let us construct an undirected graph $G = (V, E)$ as follows:

1. $V = P$, namely, each vertex of V is a point in P.
2. For each $i \in [1, m]$, add to E an edge $\{a_i, b_i\}$.

Prove: If $s > 2$, then G must be connected (i.e., for any two points $p, q \in P$, G has a path from p to q).
Hint: Imagine listing all distinct pairs of the points in P in ascending order of distance. Apply induction on the sorted list.

Proof.

Base Case: Let (p, q) be the closest pair of points. Then $(\{p\}, \{q\})$ must belong to the decomposition, for $s > 2$. Hence, $(p, q) \in G$.

Induction Hypothesis: Fix a pair (p, q). Every pair (a,b) with $|a - b| < |p - q|$ is connected.

Inductive Step: By definition, $\exists i \in [m]$, such that $p \in A_i, q \in B_i$. Consider the representative $p' \in A_i, q' \in B_i$. Since $|pq| \leq |pp'| + |p'q'| + |q'q|$, and note that p, p' and q, q' are connected by shorter edges from the inductive hypothesis, and p', q' are connected by the construction of the graph. We know that p, q is connected. □