Problem 1. Let P be a set of points in $\mathbb{R}^d$. Consider the following algorithm for finding an approximation of $\mathrm{diam}(P)$ (i.e. the diameter of P).

procedure DIAMETERAPPROXIMATION(P)
  pick an arbitrary point $p \in P$
  identify a point $q \in P$ maximizing $dist(p, q)$ (the Euclidean distance between p and q)
  return $dist(p, q)$
end procedure

Prove: The algorithm returns a $\frac{1}{2}$-approximate, answer, i.e. $1/2 \cdot \mathrm{diam}(P) \leq dist(p, q) \leq \mathrm{diam}(P)$.

Proof. Suppose the otherwise that $dist(p, q) < 1/2 \cdot diam(P)$.

Place a circle centered at p with radius $dist(p, q)$. Since $2dist(p, q) < \mathrm{diam}(P)$, there must be points lying outside of the circle. This contradicts with the assumption since q is the point maximizing $dist(p, q)$ and any point t outside of the circle has distance $dist(p, t) > dist(p, q)$. □

Problem 2. Let P be a set of points in $\mathbb{R}^d$. Describe how to find in $O(n \log n + n/\varepsilon^d)$ time a value Δ satisfying $\mathrm{diam}(P) \leq \Delta \leq (1 + \varepsilon) \cdot \mathrm{diam}(P)$ for any $0 < \varepsilon < 1$.

Solution. The original approximation of finding Δ satisfying $(1 - \varepsilon)\mathrm{diam}(P) \leq \Delta \leq \mathrm{diam}(P)$ can be done in $O(n \log n + n/\varepsilon^d)$ time by s-WSPD with approximation factor $s = 4(1 - h)/h$.

In this problem, consider scaling up the grid with the transformation $T(\mathbf{p}) = (1+\varepsilon)\mathbf{p}$. Now, apply the above algorithm with approximation factor $4/\varepsilon$, which is done in $O(n \log n + n/\varepsilon^d)$ time.

Denote the transformed coordinate distance $dist(p', q')$ and transformed diameter $\mathrm{diam}(P')$. Δ returned should satisfy

$$\frac{1}{1+\varepsilon}\mathrm{diam}(P') \leq \Delta \leq \mathrm{diam}(P') \Leftrightarrow \frac{1}{1+\varepsilon}(1+\varepsilon)\mathrm{diam}(P) \leq \Delta \leq (1+\varepsilon)\mathrm{diam}(P),$$

which is the required.

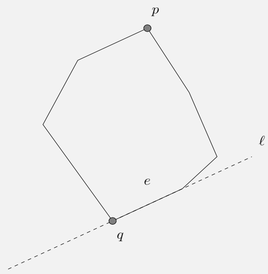
Problem 3. Let P be a set of 2D points. Let $p, q \in P$ be two points such that $dist(p, q) = \mathrm{diam}(P)$. Prove: p and q must be vertices of the convex hull of P.
Hint: Use the fact that if a point p is not a vertex of the convex hull, then every line ℓ passing p has the property that there are points of P falling on both sides of ℓ.

Proof. Suppose the otherwise that p is not a vertex of the convex hull of P. Let p, q define the diameter of a unique circle C. This circle must enclose all points $p \in P$. Let ℓ be a tangent line of C at p. If p is not a vertex of the convex hull, there must be a point lying on the opposite side of the circle with respect to ℓ, and thus, out of the circle, which is a contradiction.

By symmetry, it immediately follows that q must be a vertex of the convex hull. □

Problem 4. Let G be a convex polygon. Define p and q as the two vertices of G maximizing $dist(p, q)$. Prove: G has an edge e such that

- one of p and q is a vertex of e;
- if p is a vertex of e, then q is a vertex of G having the maximum distance to the line ℓ passing e.



Hint: First find two parallel lines passing p and q, respectively, and enclosing G in between. Then, rotate these lines.
Remark: Problem 3 and 4 suggest that the (precise) diameter of a 2D point set P can be found in $O(n \log n)$ time.

Proof. Let $\overline{pq}$ be the segment connecting points p and q. Let ℓ_1, ℓ_2 be lines crossing p, q respectively, and $\overline{pq} \perp \ell_1, \overline{pq} \perp \ell_2$.

Now, ℓ_1, ℓ_2 are lines parallel to each other and encloses G in between. Rotate these lines (in the same random direction specified, either clockwise or counterclockwise) around the pivots p, q. At any moment of the rotation before hitting any line segments, q is a vertex of G having the maximum distance to the line ℓ_1 and p is a vertex of G having the maximum to the line ℓ_2.

Proof: Observe that any point being the first to be hit by the lines is on the convex hull (gift wrapping) and the two lines bounds G at any time of the rotation before stopping. WLOG, suppose ℓ_2 hit some segment e. Since $\ell_1 \parallel \ell_2$, any points within the slab between ℓ_1 and ℓ_2 has distance less than or equal to the distance of ℓ_1 and ℓ_2, and p retains to be the point having the maximum distance to ℓ_2.

The segment e being hit satisfies the requirement. $\square$

Problem 5. Prove the correctness of the algorithm discusses in lecture for computing a t-spanner graph for a set of points in $\mathbb{R}^d$.
Hint: you can use an inductive argument.

Proof. Let $W = \{\{A_1, B_1\}, \{A_2, B_2\}, \cdots, \{A_m, B_m\}\}$ be the WSPD of the given set of points. Generate the graph as follows: for every pair $\{A_i, B_i\}, i \in [m]$, let $u = rep(A_i), v = rep(B_i)$ be the representative points (an arbitrary point in the set) of A_i and B_i, respectively, connecting them with an undirected edge bearing the weight $dist(u, v)$. Then, by Exercise 6 Problem 4, the graph is connected and all points in the given set occurs in the graph generated.

This graph is constructed in $O(n \log n + s^d \cdot n)$ time, where s is the separation factor.

Now we prove by induction on the number of edges on the generated spanner graph $G = (V, E)$, that $|xy| \leq \delta_G(x, y) \leq t \cdot |xy|$ for every pair (x, y), where $x, y \in V$ and $\delta_G(x, y)$ is the shortest path between x and y.

The first inequality is trivially true, since no path in the graph can be shorter than the distance between these two points (by triangle inequality).

- **Base Case**: Observe that if x and y is joined by an edge in G, then clearly $\delta_G = |xy| \leq t \cdot |xy|$ for all $t > 1$.
- **Inductive Step**: Suppose that there is no edge connecting x and y. Then, there must be a path connecting these two points, and a well-separated pair $\{A_i, B_i\}, i \in [m]$ with the pairs of nodes $\{x, y\}$. Let $p_x = rep(A_i), p_y = rep(B_i)$, then

$$|xy| \leq \delta_G(xp_x) + |p_x p_y| + \delta_G(p_y y).$$

Note that there can only be $p_x = x$ or $p_y = y$, $xp_x, p_y y$ must be shorter than the overall path. Therefore, we could apply the inductive hypothesis, which yields $\delta_G(x, p_x) \leq t \cdot |xp_x|, \delta_G(p_y, y) \leq t \cdot |p_y y|$, yielding

$$\delta_G(x, y) \leq t \cdot (|xp_x| + |p_y y|) + |p_x p_y|.$$

Now, since A_i, B_i are s-well separated, we know that these point sets can each be enclosed within a ball with radius at most r such that the two balls are separated by distance of at least sr. Thus, $\max(|xp_x|, |p_y y|) \leq 2r, |xy| \geq sr$. By triangle inequality,

$$|p_x p_y| \leq |p_x x| + |xy| + |yp_y| \leq 2r + |xy| + 2r.$$

Combining with the inequality of $\delta_G(x, y)$, we have

$$\delta_G(x, y) \leq t \cdot (|xp_x| + |p_y y|) + 4r + |xy| \leq t \cdot 4r + 4r + |xy|.$$

Combining $r \leq |xy|/s$, we have

$$\delta_G(x, y) \leq (4t/s) \cdot |xy| + (4/s) \cdot |xy| + |xy| = \left(1 + \frac{4(t+1)}{s}\right) |xy|.$$

Select $t \geq \left(1 + \frac{4(t+1)}{s}\right)$ where $s = 4\left(\frac{t+1}{t-1}\right)$ suffices and $\delta_G(x, y) \leq t \cdot |xy|$, completing the induction.

The number of edges in the spanner graph is $O(s^d \cdot n) = O((1/(t-1)^d) \cdot n)$ and can be computed in $O(n \log n + (1/(t-1)^d) \cdot n)$ time. □

Problem 6. Prove the correctness of the algorithm discussed in the lecture for computing a $(1 + \varepsilon)$-approximate EMST for a set of points in $\mathbb{R}^d$.

Proof. First, obtain the $(1 + \varepsilon)$-spanner graph by the algorithm describe above in $O(n \log n + n/\varepsilon^d)$ time with $O(n/\varepsilon^d)$ edges.

Let MST$= (P, E)$ be the true Euclidean minimum spanning tree for the given set of points P. We claim that the spanner graph has an $(1 + \varepsilon)$-approximate EMST.

Consider substituting the every edge (u, v) on MST with the weight $\delta_G(u, v)$ as defined in the spanner graph. Now

$$\sum_{(u,v)\in \text{MST}} dist(u, v) \leq \sum_{(u,v)\in \text{MST}} \delta_G(u, v) \leq \sum_{(u,v)\in \text{MST}} (1+\varepsilon) \cdot dist(u, v) = (1+\varepsilon) \sum_{(u,v)\in \text{MST}} dist(u, v),$$

which is justified by Problem 5.

The $(1 + \varepsilon)$-spanner of the original complete graph thus has an MST with weight no larger than $(1 + \varepsilon) \cdot w(\text{MST})$ due to minimality. □