Nearest Neighbor Search

Input: $S \subseteq \mathbb{R}^d$ $|S| = n$

Output: Search data structure DS

st

$\forall x \in \mathbb{R}^d$ $DS(x) \in S$ closest point

Cost:

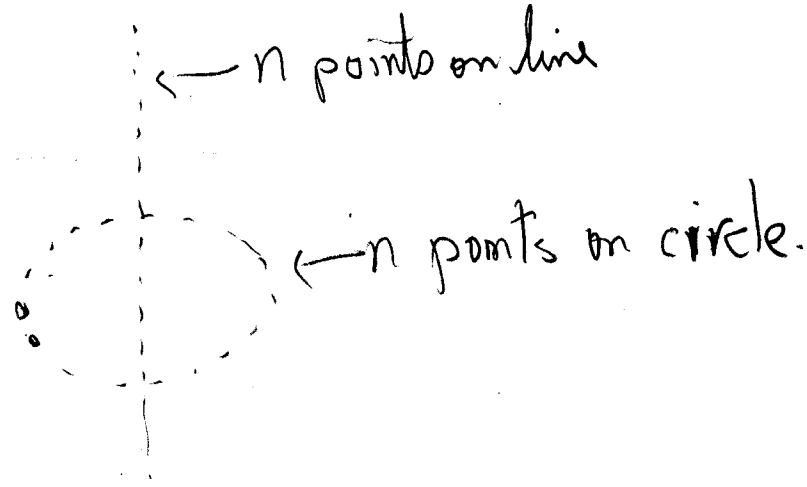
1) Time to compute DS
2) Size of DS
3) Query time

Equivalent to Finding the Voronoi cell containing x.

Prob: # of Voronoi points on d maybe of size $O(n^{\lceil d/2 \rceil})$

# of Delaunay Simplices

Preparata's Ex.



Claim: nei pair on circle + nei pair on line form a Del-Tet

$\therefore$ n^2 Tets $d=3$ $n^{\lceil d/2 \rceil} = n^2$

Naive DS may have size $\Omega(n^2)$ in 3D.

Work around: Approximate NN

Def $P \in S$ is a $(1+\varepsilon)$-approx NN of q if

$$\frac{dist(P,q)}{dist(P',q)} \leq (1+\varepsilon) \quad \forall P' \neq q \in S$$

Goal for d fixed

$O(n \log n)$ preprocessing

$O(n)$ space DS

$O(\log n)$ query time $(1+\varepsilon)$-approx

Assume $P \subseteq \mathbb{R}^d$ $P \subseteq$ Unit box

Build Quadtree (P)

Init B = UnitBox & $p \in P$

Root = (B, p) ; $rep(B) = p$

While a box $|B' \cap P| \geq 2$ . . . .

Split B' into B_1, B_2, B_3, B_4

for each i s.t $|B_i \cap P| \geq 1$ pick $p_i \in B_i \cap P$

add box (B_i, p_i) ; $rep(B_i) = p_i$

ANN-Search(q, Q) q = point
Q = Quad tree DS

Init Current = rep(B)
r_{cur} = d(q, current)
$Queue_1$ = { children(B) } $i=1$

While $Queue_i \neq \emptyset$
Search($Queue_i$); $i = i+1$ (**)

Return current

Search($Queue_i$)

While $Queue_i \neq \emptyset$ do B = Dequene($Queue_i$)

1) If $d(q, rep(B)) < r_{cur}$ then
current = rep(B); $r_{cur} = d(q, rep(B))$

2) If $d(q, rep(B)) - dia(B) < (1 - \varepsilon/2) r_{cur}$ (*)
then add children(B) to $Queue_{i+1}$

Claim If $d(B) < (\frac{\varepsilon}{2})\delta$ then B is Discarded.

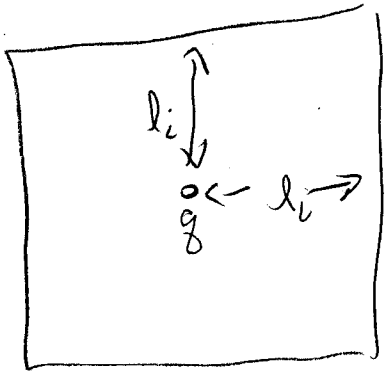
$$\begin{aligned} d(q, \text{rep}(B)) - d(B) &\geq d(q, \text{rep}) - (\tfrac{\varepsilon}{2})\delta \\ &\geq r_{cur} - (\tfrac{\varepsilon}{2}) r_{cur} \\ &= (1 - \tfrac{\varepsilon}{2}) r_{cur} \end{aligned}$$

Claim After Search(Q_i) $l_i = \delta + d(B_i) = \delta + \sqrt{d}\, 2^{-i} \geq r_{cur}$.

Consider box B s.t. $nn(q) \in B$

$$r_{cur} \leq d(q, \text{rep}(B)) \leq \delta + d(B)$$

Claim Live box at time i contained in $2l_i \times 2l_i$ box $\bar{B}$ centered at q_i



Correctness

To show It is OK to discard a box

ie Let B be box contains $NN(q) \in P$.
If B is discarded then $rep(B)$ is a $(1+\varepsilon)$-ANN.

Suppose B is discarded

$$(1-\varepsilon/2)\, r_{cur} \leq d(q, rep(B)) - dia(B)$$

$$\leq_{\Delta} d(q, nn(q)) = \delta$$

$$\frac{\delta}{(1-\varepsilon/2)} \geq r_{cur}$$

but $\frac{1}{(1-\varepsilon/2)} \leq 1+\varepsilon$ for $0 < \varepsilon \leq 1$

$$r_{cur} \leq (1+\varepsilon)\,\delta$$

$$\text{Vol}(\bar{B}) = (2l_i)^d$$

$$\text{Vol}(B_i) = (1/2^i)^d \qquad n_i \leq \left(\frac{2l_i}{\bar{2}^i}\right)^d$$

$$= \left[2\left(\frac{\delta + \sqrt{d}\,\bar{2}^i}{\bar{2}^i}\right)\right]^d = O\left(\left(1 + \frac{\delta}{\bar{2}^i}\right)^d\right) = O(1 + 2^i\delta)^d$$

$$= O\left(1 + (2^i\delta)^d\right)$$

Total # Box

$$\sum_{i=0}^{h} n_i = O\left(\sum_{0}^{-\lg(\frac{\varepsilon}{2}\delta)} \left[1 + (2^i\delta)^d\right]\right)$$

$$= O\left(\log\left(\frac{1}{\varepsilon\delta}\right) + \left(\frac{\delta}{\frac{\varepsilon}{2}\delta}\right)^d\right)$$

$$= O\left(\log\left(\frac{1}{\delta}\right) + \frac{1}{\varepsilon^d}\right)$$

Lemma $P \subseteq \mathbb{R}^d$ is n points $\wedge$ QT $\mathcal{T}$ of P

$dia(P) \approx 1$ then $(1+\varepsilon)$-ANN queries in

$O\left(1/\varepsilon^d + \log(\Phi)\right)$ Φ = spread of P.