L16: Back analysis II

$P \subseteq \mathbb{R}^2 \qquad |P| = n$

$$cl(P) = \min_{\substack{p,q \in P \\ p \neq q}} \|pq\|$$

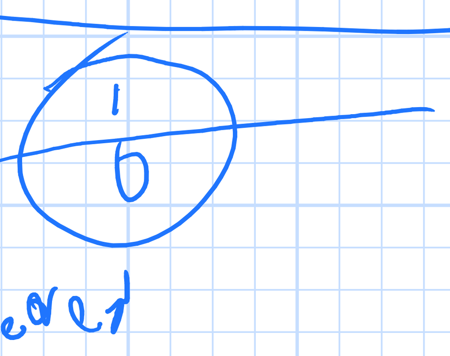
UNIQNESS / distinctness

$X \subseteq \mathbb{R}$ n numbers

Q: Are they all distinct?

Theorem UNIQNESS takes $\Omega(n \log n)$ time in the comparison model.

$n!$



$\log_2 n! \approx$ # of comp needed

$\log_2 n! = \Omega(n \log n)$

$$n! = 1 \cdot 2 \cdot 3 \cdots n \geq \underbrace{\frac{n}{2} \cdot \frac{n}{2} \cdot \frac{n}{2} \cdots \frac{n}{2}}_{\frac{n}{2} \text{ times}} = \left(\frac{n}{2}\right)^{n/2}$$

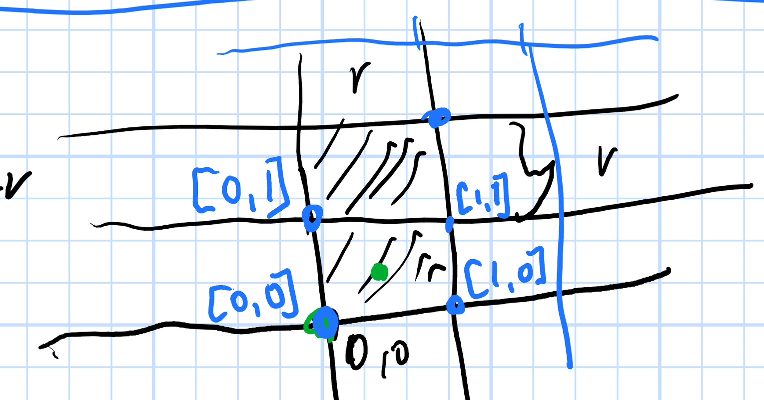
$$\log n! \geq \log_2 \left(\frac{n}{2}\right)^{n/2} = \frac{n}{2} \log_2 \frac{n}{2}$$

Extensions operations

- Floor function $\lfloor x \rfloor$ = integer part of x.
- hashing $O(1)$ per operation.

Grid

r – parameter



$$id(x,y) \longrightarrow \left(\left\lfloor \frac{x}{r} \right\rfloor, \left\lfloor \frac{y}{r} \right\rfloor\right)$$

$0.999\ldots = 1.000$

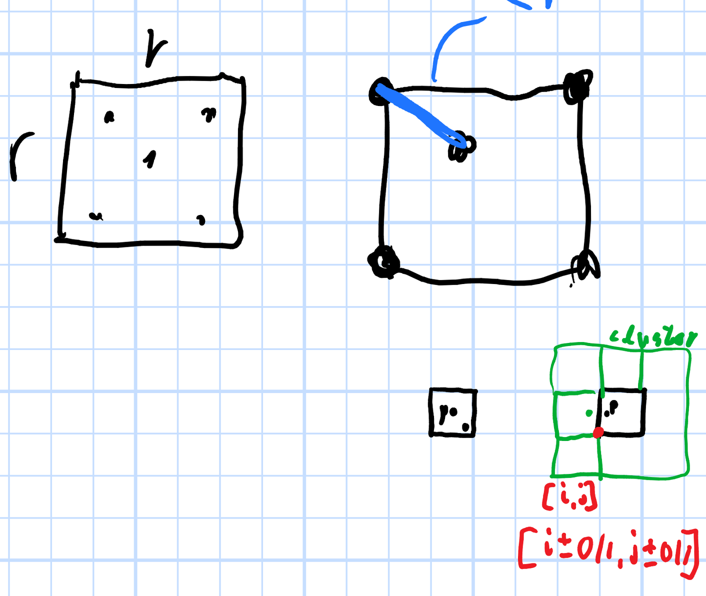
$P \to G_r$

$\forall (x,y) \in P$ store $Map[id(p)] \leftarrow p$

Storing point set in a grid in linear time.

Decision version: Given P and r decide if $cp(P) \leq r$ or $> r$?

- store points in the grid G_r.
- For all $\square \in G_r$ that contains points of P:
  - if $|\square \cap P| > 4$ then output "$cp(P) < r$"
- For all $\square \in G_r$ containing points of P
  - For all $p \in \square \cap P$ compute it's "r-closest pair distance"
  - Extract all points in the cluster of p, compute p distance to all these points



If min distance computed smaller than r then "$cp < r$"

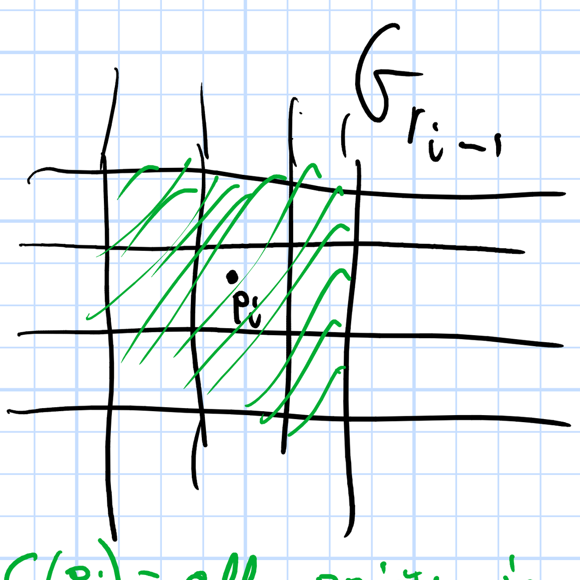
else "$cp > r$".

$p_1, p_2, \ldots, p_n$ is a random permutation

$r_i = cp(p_1, p_2, \ldots, p_i)$

In the ith iteration we insert p_i.

$P_{i-1} = \{p_1, \ldots, p_{i-1}\}$



Store P_{i-1} in the grid $G_{r_{i-1}}$

If $r_i = r_{i-1}$ then continue to next iteration

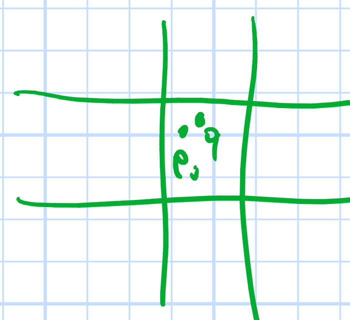
If $r_i < r_{i-1}$ then rebuild and recompute $G_{r_i}(p_1, \ldots, p_i)$

loop $i = 1 \ldots, n$

return r_n

$C(p_i)$ = all points in P_{i-1} in the cluster of p_i

if $\forall q \in C(p_i)$ $\|qp_i\| > r_{i-1}$ then $r_i = r_{i-1}$

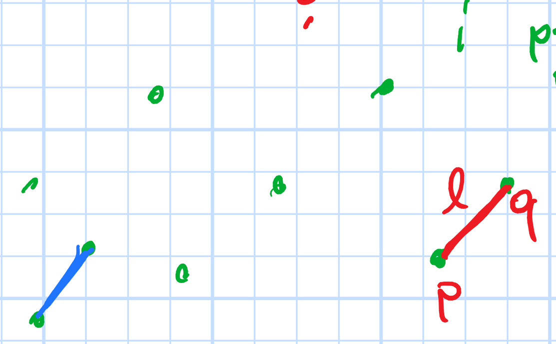


The new value of r_i is computed. Rebuild the grid $G_{r_i}(P_i)$

$$P[r_i < r_{i-1}] = P\left[cp(p_1, \ldots, p_i) < cp(p_1, \ldots, p_{i-1})\right]$$

$$\leq \frac{2}{i}$$

i points of P_i



R_i

$$E\left[\begin{array}{c}\text{Running time in} \\ \text{ith iteration}\end{array}\right]$$

$$= P\left[\text{no rebuild grid in ith iter}\right] O(1)$$

$$+ P\left[\text{Rebuild grid in ith iter} \Leftrightarrow r_i < r_{i-1}\right] O(i)$$

$$\leq 1 \cdot O(1) + \frac{2}{i} O(i) = O(1)$$

Total expected running time

$$\sum R_i = \sum_{i=1}^{n} O(1) = O(n).$$

$$\sum_{i=1}^{n} \frac{2}{i} = O(\log n)$$

$O(nX)$

$X \sim Geom\left(\frac{1}{2}\right)$