15-456
3/01/13

2D-Closest Pair using Hashing & Randomization.

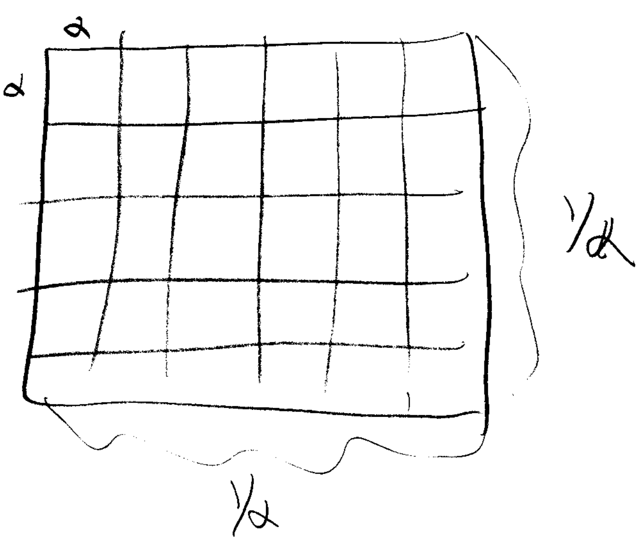
The Closest Pair Prob:

Input: $P \subseteq \mathbb{R}^2$ $P \subseteq$ Unit Box $|P| = n$

Output: $CP(P) = \min_{p \neq q \in P} \|p - q\|$

Placing Points into boxes using hashing

Idea: Partition Unit Box into boxes of side length α

$G_\alpha \equiv$ 

$(1/\alpha)^2$ boxes.

<u>Note</u> If α is small $\approx 10^{-20}$ then

10^{40} boxes!

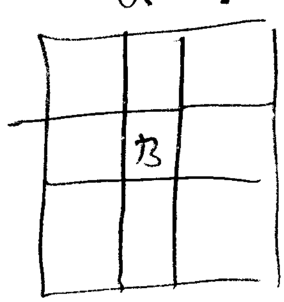
Let names of boxes be key space.

Make hash table $O(n)$, say H

Hash points $\in P$ using name of containing box as key.

<u>Thm</u> Hash points "into" its box in $O(1)$ time

<u>Def</u> B box of G_α then <u>extended neig</u> of B are:

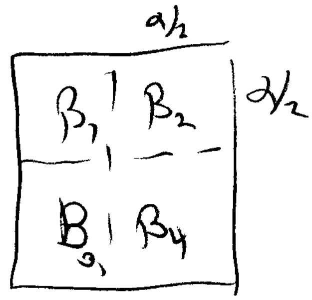


9 boxes.

Denoted $Ext(B)$

<u>Packing Lemma</u>: Box B with sidelength α
$\alpha < CP(P)$ $P \subseteq B$ Then
$|P| \leq 4$

pf Split B into 4 boxes



$\operatorname{dia}(B_i) = \frac{\alpha}{\sqrt{2}}$

Thus each B_i contains ≤ 1 point.

$\alpha > 0$

<u>Def</u> $\text{Test}(\alpha, P) = \begin{cases} \beta < \alpha & \text{if } \exists p \neq q \in P \ \ \|p-q\| = \beta < \alpha \\ \alpha & \text{if } CP(P) = \alpha \\ \text{False} & \text{o.w.} \end{cases}$

Proc $TEST(\alpha, P)$ $\quad \alpha > 0$

1) Make hash table H_α for grid G_α
2) "Insert" P_1 into G_α
3) For $i = 2$ to n $\qquad$ Let $P_i = \{P_1, \ldots, P_{i-1}\}$
   a) Insert P_i into its box B.
   b) compute min $\text{dis}(P_i, P_i \cap Ext(B)) = \beta$
   c) If $\beta < \alpha$ return "$CP(P) \leq \beta < \alpha$" (restart)
   d) If $\beta = \alpha$ set Flag = true
4) If Flag then return "$CP(P) = \alpha$"
   Else return "$CP(P) > \alpha$"

<u>Claim</u> TEST is linear time.
Correctly Tests $CP(P) \overset{?}{} \alpha$.

Input $P \subseteq$ Unit Box

Proc: $\overline{CP}(P)$

1) If $n \leq 4$ check all pairs.
2) Randomly permute $P = \{P_1, \ldots, P_n\}$; $\alpha = 1$
3) If TEST(α, P) returns $\beta < \alpha$
   then set $\alpha \Leftarrow \beta$; repeat 3). (restart)
4) Return α.

Correctness:

If $n \leq 4$ done

By Lemma if $n > 4$ then $\alpha < 1$.

Thm $\overline{CP}(P)$ is expected linear time.

Use Backwards Analysis.

Def α_i be random variable.

$$\alpha_i = CP(P_1, \dots, P_i) \quad i \geq 2$$

note $\alpha_{i+1} \leq \alpha_i$

note We restart TEST for each i st

$$\alpha_i < \alpha_{i-1}.$$

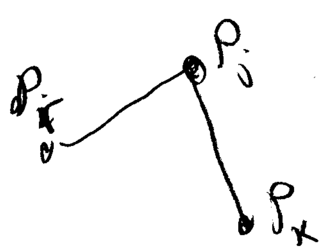
Goal: $Prob(\alpha_i < \alpha_{i-1})$?

Case 1 $\exists!$ CP in $\{P_1, \dots, P_i\}$

say (P_j, P_k) then only restart for

$i = j$ or k $\quad Prob(\alpha_i < \alpha_{i-1}) = 2/i$

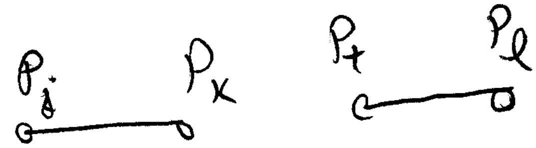
Case 2 $\neg\exists!$ CP in $\{P_1, \dots, P_i\}$



restart for $i = j$

$Prob = 1/i$

Case 3



no restarts! $\therefore$ Prob $= 0$

Each restart is $O(i)$ new work.

Total expected work $\leq$

$$O\left(\sum \left(\frac{2}{i}\right) i\right) = O(n)$$

k-enclosing disk

$D_{opt}(P,k) \equiv$ min radius disk contain k points.

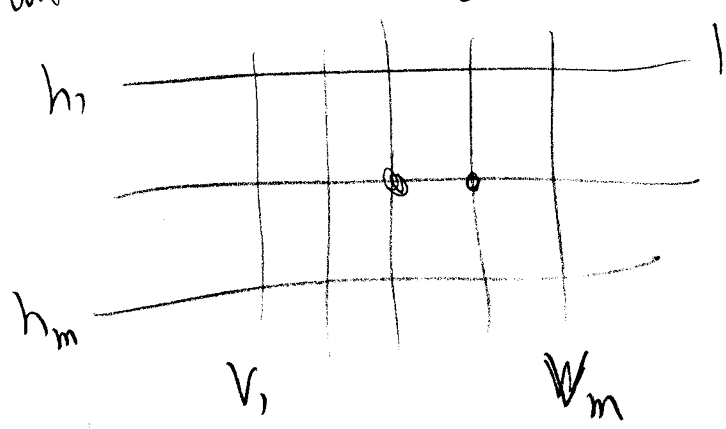
$r_{opt}(P,k)$ = radius of D_{opt}

2-approx for k-enclosing disk.

Make horizonal lines $h_1 \dots h_m$ $m = O(n/k)$
Using a selection algorithm on y values of P.
s.t. between each line we have at most $k/4$ points.

Make vertical lines $v_1, \dots, v_m$ as above.

Thus non-uniform grid



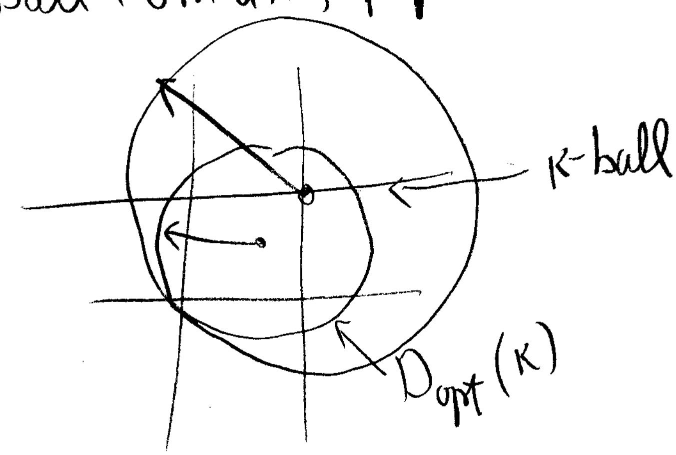
$O((\frac{n}{k})^2)$ intersections.

$O(n \log(n/k))$ time

<u>Claim</u> no ball disjoint from the intersections contains more than $k/2$ points

thus a ball containing an k points contains an intersection

<u>Alg</u> for each intersection point compute ball containing k points. Pick min radius ball.



<u>Thm</u> $O(n(n/k)^2)$ time 2-approx