15-750
2/24/16

Hashing

2D-Closest Pair

Universe of Keys $= \mathcal{U}$

Note Often $|\mathcal{U}| \geq 10^{100} \geq 10^{60}$ universal large constant!

$K \subseteq \mathcal{U}$ actual keys

K is small, say, $|K| \approx 10^{15}$

Our solution so far been

Sort K Incurs a $O(\lg n)$ cost per op!

Hashing: Randomly map $\mathcal{U} \to T$ table

$|T| \approx |K|$

$O(1)$ time operation (expected)

Insert

Lookup

Delete

2D-Closest Pair using Hashing & Randomization.

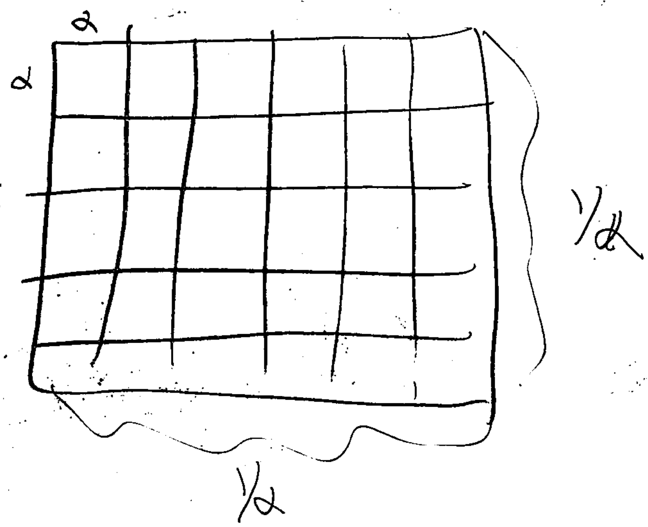
The Closest Pair Prob:

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

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

Placing Points into boxes using hashing

Idea: Partition UnitBox into boxes of side length α

$G_\alpha \equiv$ 

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

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

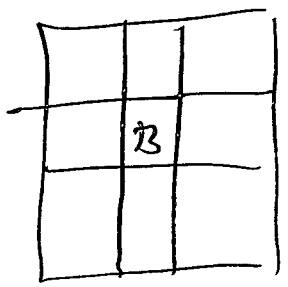
10^{40} boxes!

Let names of boxes be key space.
Make hash table $O(n)$ size, say, H. (dynamic sizing)

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

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

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

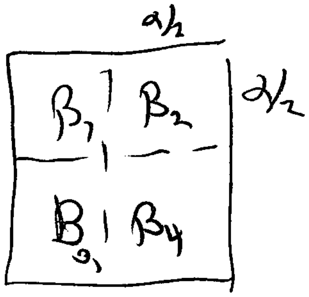


9 boxes.

Denoted $Ext(B)$

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

pf Split B into 4 boxes



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

Thus each B_i contains ≤ 1 point.

$\alpha > 0$

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

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

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 $\min \text{dis}(p_i, P_i \cap \text{Ext}(B)) = \beta$
    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 "$CP(P) > \alpha$"

Claim TEST is linear time and Correctly Tests $CP(P) \,?\, \alpha$.

Input $P \subseteq$ Unit Box

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

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

Correctness:

If $n \le 4$ 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.

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

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

note We restart TEST for each i st

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

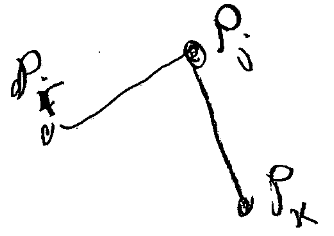
Goal: $\text{Prob}(\alpha_i < \alpha_{i-1})$?

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

say (P_j, P_k) then only restart for

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

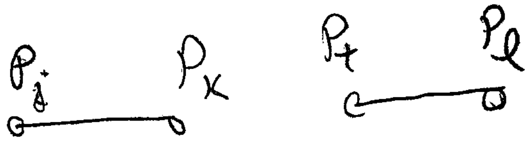
Case 2 $\neg \exists!$ CP in $\{P_1, \cdots, P_i\}$



restart for $i = j$

$\text{Prob} = 1/i$

<u>Case 3</u>



no restarts! : Prob = 0

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

Total expected work $\leq$

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