

Randomized Algo: Closest Pair

def: Given $P_1 = (x_1, y_1) \dots P_n = (x_n, y_n) \in \mathbb{Z}^2$. The closest pair is to find

$$\arg \min_{i \neq j} \frac{\text{dist}((x_i, y_i), (x_j, y_j))}{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}$$

Conventions: All numbers $O(\lg n)$ bits.

$$|x_i|, |y_i| \leq N = \text{poly}(n)$$

Thm (prev) Closest pair in $O(n \lg n)$ expected time.

Thm in $O(n)$ expected time.

idea: data structure \Rightarrow efficient way to store data.

Warmup Can we verify minimum dist $\geq \Delta$, $\Delta \in \mathbb{Z}$?

idea: process points in order.

def: $\Delta_k = \min_{i,j \leq k} \text{dist}(P_i, P_j)$

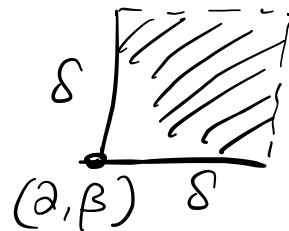
Q: Compare Δ_n v.s. Δ ? Δ_{k-1} v.s. Δ_k ?

idea: coarse discretization of space

def: $\delta > 0$, $\delta \in \mathbb{R}$, a δ -subsquare is a subset of \mathbb{Z}^2 given

$$S_{\alpha, \beta} = \int (x, y) \in \mathbb{Z}^2 : \begin{array}{l} \alpha \leq x \leq \alpha + \delta \\ \beta \leq y \leq \beta + \delta \end{array}$$

$\uparrow \quad \uparrow$
 \mathbb{R}



claim: $\Delta_k \geq \Delta \Rightarrow$ each $\Delta/2$ subsquare have ≤ 1 points from P_i to P_k .

Pf: $\frac{\Delta}{2} \times \frac{\Delta}{2} \Rightarrow \text{distance} \leq \Delta/\sqrt{2} < \Delta$ (contrapositive)

Idea: round coordinates:

def: $(a,b) \in \mathbb{N}^2$, the $(a,b) \frac{\Delta}{2}$ subsquare of \mathbb{N}^2

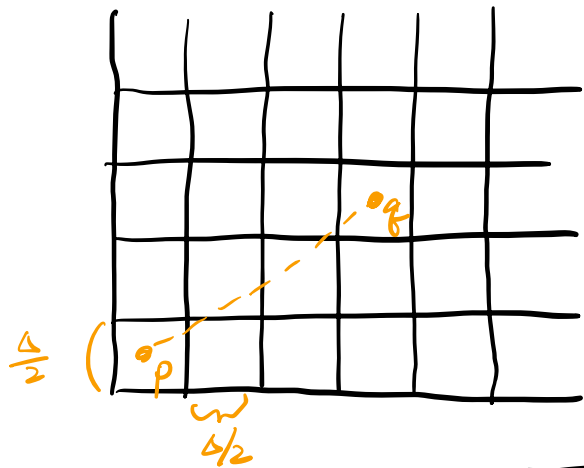
is $S_{a \cdot \Delta/2, b \cdot \Delta/2}$ Partition

The $\frac{\Delta}{2}$ -grid is $\{S_{a,b}, 0 \leq a,b \leq \frac{2N}{\Delta}\} = T_{a,b}$

prop $d(p,q) \leq \Delta$, $p \in S_{a,b}$, $q \in S_{c,d}$

$$\Rightarrow |a-c|, |b-d| \leq 2$$

Pf: By contrapositive:



$$\text{If } |a-c| > 2 \Rightarrow \text{dist}(p,q) = \sqrt{(x-z)^2 + (y-w)^2} \geq |x-z| > \Delta$$

def A dictionary over U is a data structure for a set $S \subseteq U$
of keys x along with value y .

Supports $\left\{ \begin{array}{l} \text{insert}(x, y) \\ \text{lookup}(z) \end{array} \right.$

Algo:

- Init dictionary A on $U = \left\{ (a, b), 0 \leq a, b \leq \frac{2N}{\Delta} \right\}$
- for $1 \leq i \leq n$
 - $(a, b) = \left(\left\lfloor \frac{x_i}{\Delta/2} \right\rfloor, \left\lfloor \frac{y_i}{\Delta/2} \right\rfloor \right)$ *round down*
 - Compute $\Delta_i' = \min d(p_i, p), p \in A[a, b], |a-c|, |b-d| \leq 2$.
 - if $\Delta_i' < \Delta$, return p_i, p, Δ_i'
 - Insert p_i into $A[a, b]$
- return $\Delta_n \geq \Delta$

Prop Suppose $\Delta_i \geq \Delta$, then

(a) algo reach insertion

(b) $A[a,b]$ empty when reach insertion

Cor Suppose $\Delta_i \geq \Delta$ then ≤ 25 lookups and return ≤ 25 points.

Thm $S \subseteq U$ $|S| = n$, $|U| \leq N = \text{poly}(n)$

One can in deterministic $O(n)$ time, construct hash family $\mathcal{H}: U \rightarrow T$

- $|T| \leq O(n)$

- Choose $h \in \mathcal{H}$ takes $\begin{cases} O(1) \text{ space.} \\ O(1) \text{ eval} \end{cases}$

- insertions, lookup in $O(1)$ expected.

Cor: If $\Delta_n \geq \Delta$ then $\begin{cases} - \text{ algo is correct} \\ - \text{ run in } O(n) \text{ expected.} \end{cases}$

Cor $\exists k, \Delta_{k-1} \geq \Delta, \Delta_k < \Delta$

Then algo is correct, returns $\Delta' = \Delta_k$

Runs in $O(k)$ expected time.

Use this with random order $P_1 \dots P_n$

```
(p, q) = (P_1, P_2)
Δ = d(p, q)
while
  verify  $\Delta_n \geq \Delta$ 
  if  $\Delta_n \geq \Delta$ 
    return (p, q)
  else
    (p, q), Δ ← (P', Q'), Δ'
```

} \Rightarrow correct

prop Algo update some $\Delta \rightarrow \Delta'$ once

def $I_k = \begin{cases} 1 & \text{verification algo updates } \Delta \rightarrow \Delta_k \\ 0 & \text{o.w.} \end{cases}$

prop # dict operations is $O(n + \sum_{k=1}^n k \cdot I_k)$

pf: $O(n)$ ops in final pass

$O(k)$ ops to update $\Delta \rightarrow \Delta' = \Delta_k$ happen iff $I_k = 1$

Key: What is $\Pr[I_k = 1]$?

prop $\Pr[I_k = 1] \leq \frac{2}{k}$

pf: $I_k = 1 \iff \text{update } \Delta \rightarrow \Delta_k \iff \Delta_1, \dots, \Delta_{k-1} \triangleright \Delta$
 $\iff \min_{i,j < k} d(p_i, p_j) > \min_{i < k} d(p_i, p_k)$

\Rightarrow in a random ordering, in the first k points, $\exists 2$ points that are close (will update $A \rightarrow \Delta_k$). The probability that one of these two stays in the end is $\frac{2}{k}$.

$\Rightarrow \leq \frac{2}{k}$

Cor algo in $O(n)$ expected

$$\mathbb{E}_0 \mathbb{E}_A \left[D + \sum_k D_k \circ I_k \right]$$

\uparrow $O(n)$ \uparrow $O(k)$

$$= \mathbb{E}_0 \left\{ \underbrace{\mathbb{E}_A[D]}_{O(n)} + \sum_k \underbrace{\mathbb{E}_A[D_k \circ I_k]}_{\substack{\downarrow \\ \mathbb{E}_A[D_k]}} \right\} \quad \text{not related to } A$$

$\leftarrow O(k)$ ops, k random variable independent from A .

$$= I_k O(k)$$

$$\begin{aligned} &= \mathbb{E}_0 \left\{ O(n) + \sum I_k O(k) \right\} = O(n) + \sum O(k) \Pr[I_k=1] \\ &\leq O(n) + \sum_{k=1} O(k) \cdot \frac{k}{2} \\ &\leq O(n) \end{aligned}$$