# Randomized Algorithms: Closest Pair of Points

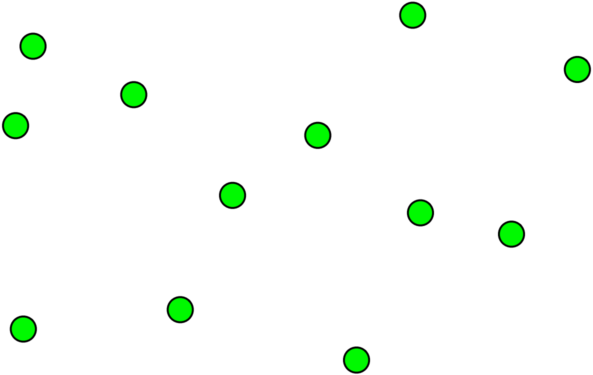
Slides by Carl Kingsford

May 2, 2014

Based on Khuller and Matias

# The problem

Problem. Given a set of points $S = \{p_1, \ldots, p_n\}$ in the plane find the pair of points $\{p_i, p_j\}$ that are closest together.

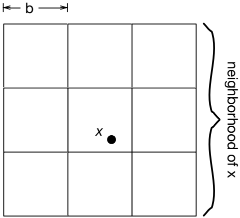


# Estimate

Let $\delta(S)$ be the smallest distance in $S$. Suppose you had a good estimate $b$ of $\delta(S)$ such that:

$$b/3 \leq \delta(S) \leq b$$

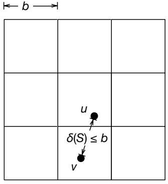
Then you could find the closest points in $O(n)$ time as follows. Create a grid of boxes of side-length $= b$:



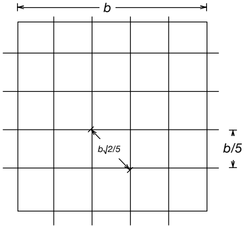

Compare each point to the points in its neighborhood.

# Why $O(n)$?

The closest pair of points lie in each other's neighborhood of the $b$-grid:




Each grid box contains $\leq 25$ points:




Largest distance inside of a smaller grid point $= \frac{\sqrt{2}}{5} b < \frac{b}{3} \leq \delta(S)$.

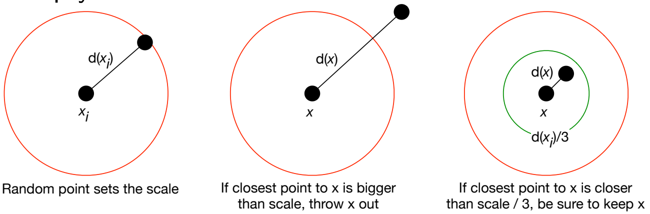
# Randomized approach to estimating $b$

While $S$ is not empty:

1. Choose random point $x_i \in S$.
2. Compute $d(x_i) :=$ smallest distance from $x_i$ to any other point currently in $S$
3. For all points $x \in S$: If

| | |
|---|---|
| far | $d(x) > d(x_i) \rightarrow$ *throw out* $x$ |
| close | $d(x) \leq d(x_i)/3 \rightarrow$ *keep* $x$ |
| medium | $d(x) > d(x_i)/3$ *but* $d(x) \leq d(x_i) \rightarrow$ *do what you want.* |

Return $b = d(x_i) = d(x^*)$ where $x^* :=$ the last $x_i$ chosen before $S$ became empty.

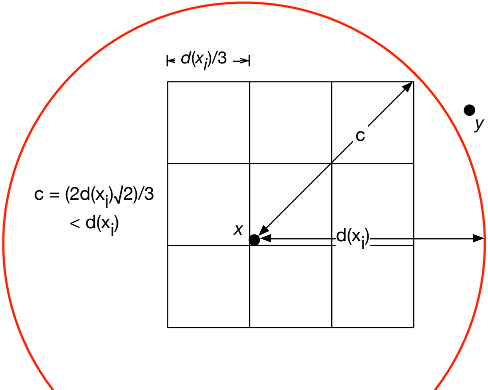


# Implementing Step 3

Build a $d(x_i)/3$-grid.

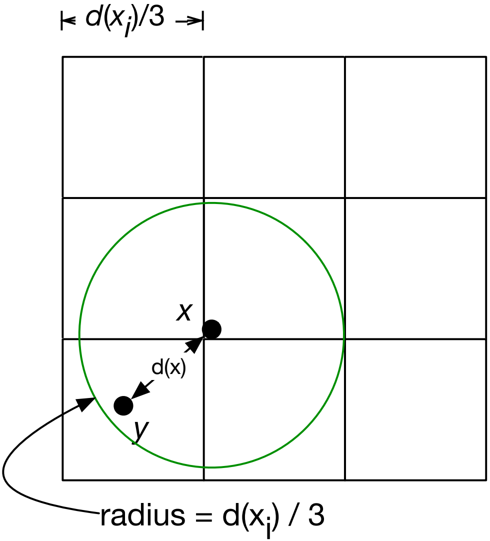
Step 3 Rule: A point $x$ should be thrown out if it's the only point in its neighborhood.

This will definitely throw out all points with $d(x) > d(x_i)$:



# Step 3, 2

Step 3 Rule will definitely keep all points with $d(x) \leq d(x_i)/3$.



# $b = d(x^*)$ is a good estimate for $\delta(S)$

We have $d(x^*) = b \leq \delta(S)$ by definition.

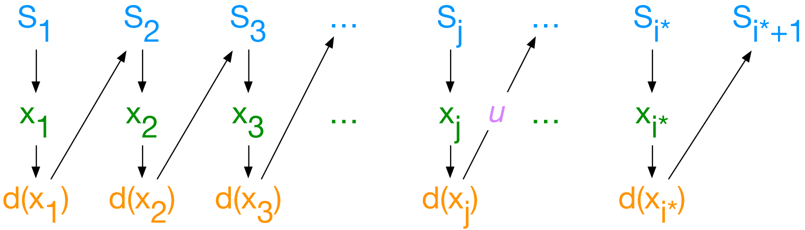
**Theorem.** $\delta(S) \geq b/3$

Proof. Let $(u, v)$ be the closest pair of points. Since $S$ eventually becomes empty, $u, v$ are deleted from $S$ at some point. Suppose $u$ was deleted first, and let $j$ be the stage at which $u$ was deleted. At that time:

- $d(u) \geq d(x_j)/3$ because otherwise $u$ would have been kept.
- $d(u) = \delta(S)$ because both $u, v$ were in $S$ at the start of stage $j$.
- $d(x_j) \geq d(x_i)$ because $i > j$ and at stage $j$ we removed all points with $d(x) \geq d(x_j)$ [rule 3.1] so there are no points left with $d(x) \geq d(x_j)$ from which $x_i$ could have been selected.

So: $d(u) = \delta(S) \geq d(x_j)/3 \geq d(x_i)/3 = b/3$ □

# Proof: picture



# Runtime Analysis

Let $S_i$ be the set of points at stage $i$.

Let $s_i$ be $|S_i|$.

**Theorem.** $\mathbb{E}[s_i] \leq \frac{n}{2^{i-1}}$

*Proof.* Assume true for $i$. Then:

$$\mathbb{E}[s_{i+1}] = \mathbb{E}\mathbb{E}[s_{i+1}] \leq \mathbb{E}[s_i/2] = \frac{1}{2}\mathbb{E}[s_i] \leq \frac{1}{2}\frac{n}{2^{i-1}} = \frac{n}{2^i}$$

Here $\mathbb{E}[s_i/2] \leq s_i/2$ because we chose $x_i$ randomly. If you consider all the points, about half would have larger $d(x)$ and half would be smaller.

□

# Runtime, 2

We then have that the total running time is

$$\mathbb{E}\left(\sum_{i=1}^{i^*} s_i\right) \leq \mathbb{E}\left(\sum_{i=1}^{n} s_i\right) = \sum_{i=1}^{n} \mathbb{E}[s_i] \leq \sum_{i=1}^{n} \frac{n}{2^{i-1}} \leq 2n$$

where $i^*$ is the number of stages, and the inequalities and equalities follow from linearity of expectation, the theorem above, and the sum of a geometric series.