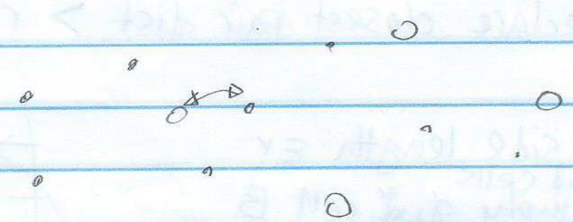
Proximity / Nearest Neighbors

Problem 1 Given n red pts P & n blue pts Q in $\mathbb{R}^d$, find bichromatic closest pair

Problem 2 find nearest blue neighbor for every red pt



Known exact alg'ms:

$d=2$ $O(n \log n)$

$d=3$ $\tilde{O}(n^{4/3})$

$d>3$ $\tilde{O}(n^{2-\frac{2}{\lceil d/2\rceil+1}})$

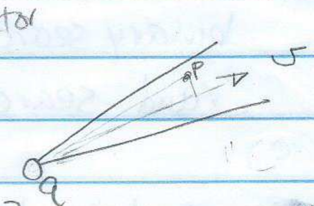
Method 0 - rounding dirs

for each red pt, q,

for each dir $v \in V_\delta$, (unit vector)

reduces to range searching:

find blue pt in cone $\{p : \angle(p-q, v) < \delta\}$ with smallest $p \cdot v$



Analysis:

$$\|p-q\| \geq (p-q)\cdot v \geq \|p-q\| \cos\delta$$
$$\sim (1-O(\delta^2))\|p-q\|$$

set $\delta = \sqrt{\varepsilon}$

$$\Rightarrow O\left(\frac{1}{\delta^{d-1}}\right) = O\left(\frac{1}{\varepsilon^{(d-1)/2}}\right) \text{ dirs}$$

not obvious how to do cone range search ...

Method 1 - grid

approx decision problem: given fixed r,

return some pair of dist $\leq (1+\varepsilon) r$

or declare closest pair dist $> r$.

Method 1.1

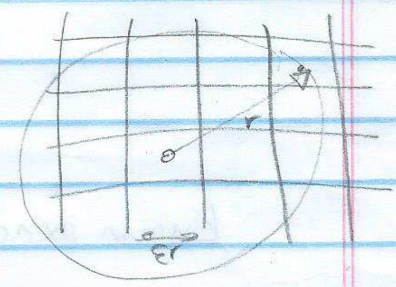
form grid of side length εr

hash pts to grid cells

for each nonempty grid cell B

check all grid cells B'

with $d(B, B') \leq r$.



(exact for monochromatic)

runtime $\boxed{O\left(\frac{1}{\varepsilon^d} n\right)}$

how to reduce to decision problem?

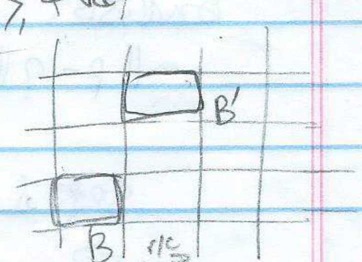
binary search - extra $\log U$ factor

rand. search ...

better ε-dependence?

Method 1.2

$(c \geq 4\sqrt{d})$



form grid of side length $\frac{r}{c}$

if some grid cell contains

both red & blue, done

for each nonempty red grid cell B

for each nonempty blue grid cell B' with $d(B, B') \leq r$

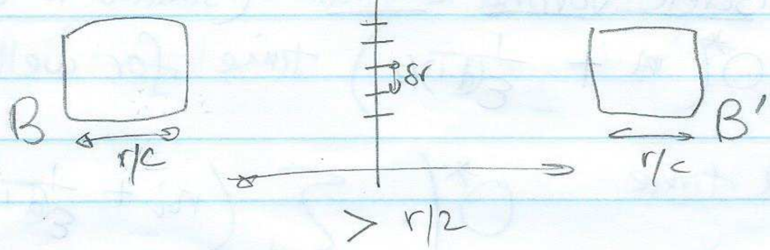
if $d(B, B') \leq r/2$, done,

if $d(B, B') > r/2$

B & B' are well-separated

solve subproblem in B & B'

How to solve well-separated case?



build grid G_δ of side length δr over middle hyperplane H

for each grid pt $g \in G_\delta$,

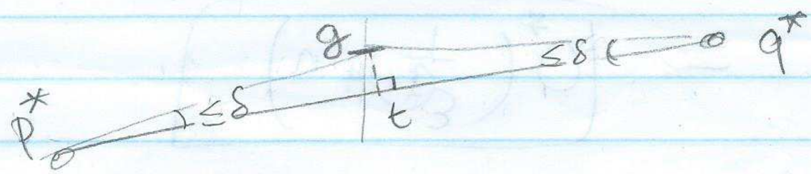
$O\left(\frac{1}{\delta^{d-1}}\right)$ time

p_g = nearest neighbor of g in B

q_g = " " " " B'

return closest among p_g, q_g

Analysis: say $p^* q^*$ is closest pair



Let g be closest grid pt to $\overline{p^* q^*} \cap H$

$$
\begin{aligned}
\| p_g - q_g \| &\le \| p_g - g \| + \| g - q_g \| \\
&\le \| p^* - g \| + \| g - q^* \| \\
&\le \frac{\| p^* - t \|}{\cos \delta} + \frac{\| t - q^* \|}{\cos \delta} \\
&= \frac{\| p^* - q^* \|}{\cos \delta}
\end{aligned}
$$

approx factor $1+\varepsilon$ by setting $\delta \sim \sqrt{\varepsilon}$

total time $\boxed{O\left(\frac{1}{\varepsilon^{(d-1)/2}} n\right)}$

Improvement:

by discrete Voronoi diagram (similar to ε-kernels!)

near $O^*\left(n + \frac{1}{\varepsilon^{(d-1)/2}}\right)$ time for well-separated case

total time $$O^*\left(\sum_i \left(n_i + \frac{1}{\varepsilon^{(d-1)/2}}\right)\right)$$ with $\sum_i n_i = O(n)$

$$= O^*\left(\frac{1}{\varepsilon^{(d-1)/2}}\, n\right)$$ no better!

Further improvement: (C.'17 / Arya-da Fonseca-Mount '17)

$$O^*\left(\sum_i \min\left\{n_i + \frac{1}{\varepsilon^{(d-1)/2}},\ n_i^2\right\}\right)$$

if $n_i > \frac{1}{\varepsilon^{(d-1)/4}}$

if $n_i < \frac{1}{\varepsilon^{d-1}}$

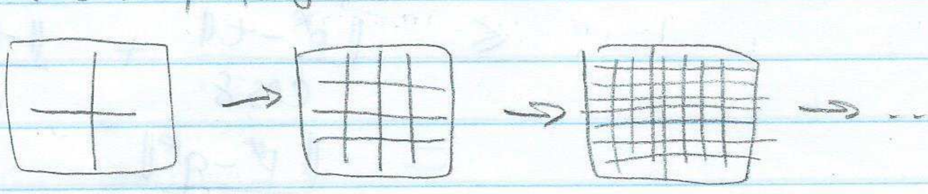
$$\leq O^*\left(\sum_i \frac{1}{\varepsilon^{(d-1)/4}}\, n_i\right)$$

$$= \boxed{O^*\left(\frac{1}{\varepsilon^{(d-1)/4}}\, n\right)}$$

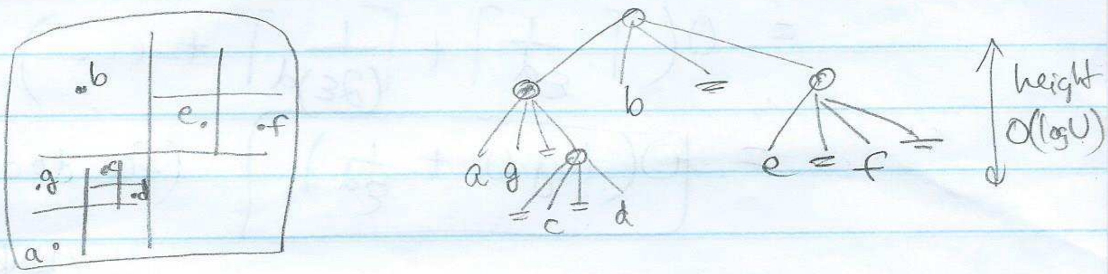
What if r is not fixed? e.g. nearest neighbor of each red pt

Method 2 - Quadtrees

idea - hierarchy of grids



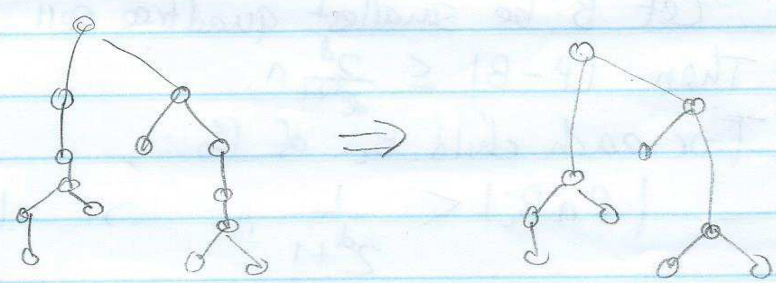
e.g.



<u>Def</u> A <u>quadtree cell</u> B is a grid cell of side length 2^l for some $l \in \{0, 1, \ldots, \log U\}$.

space $O(n \log U)$

can be reduced to $\boxed{O(n)}$
by compressed quadtree (short-cutting deg-1 nodes)



decis-query(B, q, r): // find pt of dist $\le (1+\varepsilon) r$ from q or declare nearest neighbor dist $> r$)

   if ball(q, r) does not intersect B return
   if B has side length $\le \varepsilon r$ return any pt in B
   for each child B_i of B
      decis-query(B_i, q, r).

$\Rightarrow$ query time $= O($ # quadtree cells of side length $> \varepsilon r$ intersecting ball$(q, r))$

$$= O\left(\sum_{l:\, 2^l > \varepsilon r} \left\lceil \frac{r^d}{(2^l)^d} \right\rceil \right)$$

$$= O\left(\left\lceil \frac{1}{\varepsilon^d}\right\rceil + \left\lceil \frac{1}{(2\varepsilon)^d}\right\rceil + \cdots\right)$$

$$= \boxed{O\left(\log U + \frac{1}{\varepsilon^d}\right)} \quad \text{(for decis.)}$$

(Rmk. for nearest neighbor, do binary search or "priority" search)

can we reduce $\log U$ to $\log n$?

<u>Method 2.2.</u> <u>Balanced Quadtrees</u>

<u>Lemma</u> $\exists$ quadtree cell B s.t. $|P \cap B|, |P - B| \le \frac{2^d}{2^d+1} n$.

Pf: Let B be smallest quadtree cell with $|P \cap B| \ge \frac{1}{2^d+1} n$

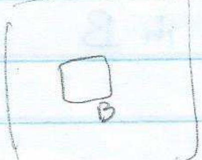
Then $|P - B| \le \frac{2^d}{2^d+1} n$.

For each child B_i of B,

$$|P \cap B_i| < \frac{1}{2^d+1} n \implies |P \cap B| \le \frac{2^d}{2^d+1} n. \qquad \square$$

recurse in $P \cap B$ & $P - B$

$\implies$ binary tree of height $\log_{\frac{2^d+1}{2^d}} n = O(\log n)$.



query time $O\big(($# quadtree cells B of side length $> \varepsilon r$ intersecting $\text{ball}(q,r)) * \log n\big)$

← a cell may occur in multiple nodes

$$= \boxed{O\left(\frac{1}{\varepsilon^d} \log n\right)}$$

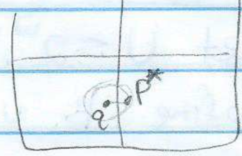
Rmk: Known alternatives:

BBD trees (Arya, Mount et al. '95),

BAR trees, fair-split trees, ring-cover trees, skip quadtrees, ...

(aiming for const-factor approx)

Method 2.3 ... with shifting



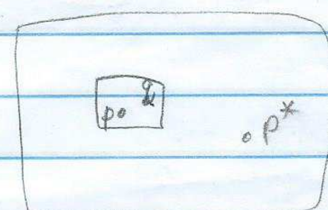
naive-query (B, q) :
  find child B_i containing q
  naive-query (B_i, q)

when does this work?

Let r^* = nearest neighbor dist from q.

Def ball(q, r^*) is k-good if it lies in a quadtree cell of side length $< 4kr^*$

Analysis; assuming ball(q, r^*) is k-good
  let p be output pt
  $\Rightarrow \|p-q\| < 4k\sqrt{d}\, r^*$
  $\Rightarrow$ O(1) approx if k const



$< 4cr^*$

Query time: $\boxed{O(\log U)}$

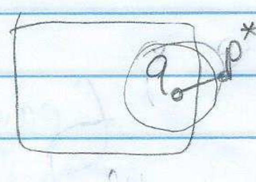
can be reduced to $\boxed{O(\log n)}$ by binary search in "Z-order"

[Bern '93] can make $v_1 = \cdots = v_d$.

Shifting Lemma (Rand. Version)
  Shift all pts by rand. vector $v = (v_1, \ldots, v_d) \in [U]^d$.
Then ball(q, r^*) is k-good with const prob if $k > cd$.

Pf: Say $2^l \le 4kr^* < 2^{l+1}$.

$\Pr[\text{bad}] = \Pr$(ball(q, r^*) crosses boundary of quadtree cell of side length 2^l]



$$\le d \cdot \frac{2r^*}{2^l} \le \frac{d}{c} \qquad \square$$

/7

Shifting Lemma (Det. Version) [C, '98]

Let $U = 2^w$. Say k is odd with $k > d$.

Define k vectors $v^{(i)} = \left(\frac{i 2^w}{k}, \ldots, \frac{i 2^w}{k} \right)$ $i = 0, \ldots, k-1$

Then $\forall$ ball, $\exists i$ s.t. it is k-good after shifting by $v^{(i)}$.
in fact, for all but d indices.

Pf: Fix q, r^*. Fix i.

I bad after shifting by $v^{(i)}$ $\Rightarrow$ $\left(\left(q_j + \frac{i 2^w}{k} \right) \bmod 2^l \right) \notin [r^*, 2^l - r^*)$

for some $j \in \{1, \ldots, d\}$

$\Rightarrow \left(\frac{k}{2^l} q_j + i 2^{w-l} \right) \bmod k$

$\notin \Big[\underbrace{\frac{k}{2^l} r^*}_{< \frac{1}{2}}, \; k - \frac{k}{2^l} r^* \Big)$

for some j

$\Rightarrow 2^{w-l} i + \underbrace{\left[\!\left[\frac{k q_j}{2^l} \right]\!\right]}_{\text{(round)}} \equiv 0 \bmod k$ for some j

($ax + b \equiv 0 \bmod k$ has uniq. sol'n if a, k rel. prime) $\Rightarrow$ at most d choices of i. $\square$

Rmk: const approx $\longrightarrow$ $1+\varepsilon$ approx

query time $O\left(\frac{1}{\varepsilon^d} \log n \right)$

(can be improved to $\tilde{O}\left(\frac{1}{\varepsilon^{d/4}} \log n \right)$

with $\tilde{O}\left(\frac{1}{\varepsilon^{d/4}} n \right)$ space)

[time/space trade-offs)