Orthogonal Range Search in High Dims?

(appl: L_∞ exact nearest neighbors...)

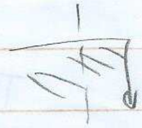
k-d tree: $O(n)$ space, $O(n^{1-1/d})$ time

range tree: $O(n \log^d n)$ space, $O(\log^d n)$ time

nontriv. only for $d \ll \frac{\log n}{\log\log n}$

better analysis:

consider dominance range



$$Q_d(n) \leq Q_d(\tfrac{n}{2}) + Q_{d-1}(n)$$

$$\Rightarrow Q_d(n) \leq O\left(\binom{\log n + d}{d}\right)$$

$$\lesssim O\left(\frac{\log n + d}{d}\right)^d$$

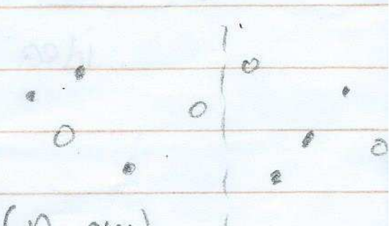
nontriv for $d < 0.29 \log n$

larger d?

Consider offline case of dominance, given n data pts (blue) & m query pts (red)

Old Method ①

divide by median blue pt

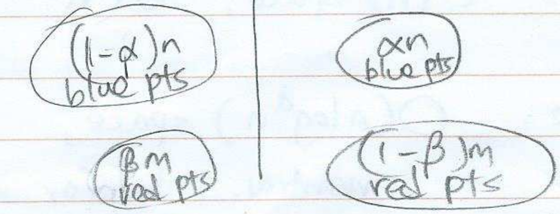


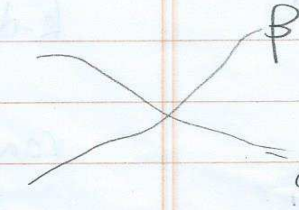
$$T_d(n,m) \leq \max_{\alpha \in [0,1]} T_d\left(\tfrac{n}{2}, (1-\alpha)m\right) + T_d\left(\tfrac{n}{2}, \alpha m\right) + T_{d-1}(n, m)$$

as above

New Method 1 (Impagliazzo et al. '14 / C.'15)

idea – lop-sided divide & conquer





choose dividing hyperplane s.t. $\alpha = \beta$

$$T_d(n,m) \le \max_{\alpha \in (0,1)} T_d((1-\alpha)n, \alpha m) + T_d(\alpha n, (1-\alpha)m) + T_{d-1}((1-\alpha)n, (1-\alpha)m)$$

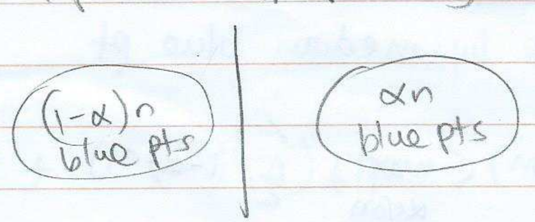
$$\Rightarrow T_d(n,n) \le \boxed{n^{2 - \frac{1}{\Theta(c \log^2 c)}}} \text{ for } d = c \log n.$$

nontriv for $c \ll \log n$ i.e. $d \ll \log^2 n$.

online queries?

Method 2 (C'17)

idea – lop-sided d & c again!



Pick $\alpha \sim \frac{1}{(c \log c)^4}$

\# subsets $\binom{c \log n}{\frac{\delta \log n}{\log c}} \sim n^{\delta}$

- pick a rand. axis to divide
- store range trees for all subsets of $\frac{\delta \log n}{\log c}$ dims.

$\Rightarrow O(n^{1+\delta})$ space, $n^{1 - \frac{1}{\Theta(c \log c)}}$ query time (rand.)

Method 3 (Abboud-Williams-Yu '16)

back to offline: $\boxed{n^{2-o(\log c)}}$ time for $d = c\log n$

idea - reduce to $\{0,1\}$ case by divide & conquer

- solve $\{0,1\}$ case by the polynomial method + matrix mult

Divide into $\frac{n}{s}$ groups of s pts.

Given groups A, B,

$$E(A,B) = \bigwedge_{a\in A,\, b\in B} [a \text{ not dom by } b]$$

$$= \bigwedge_{a\in A,\, b\in B} \underbrace{\bigvee_{i=1}^{d} a_i \bar{b}_i}_{\text{poly of large deg}}$$

$$\left(\bigvee_{i=1}^{d} x_i = 1 - \prod_{i=1}^{d} (1-x_i) \right)$$

rand. trick - replace $\bigvee_{i=1}^{d} x_i$ by $\underbrace{\bigvee_{j=1}^{t} \sum_{i\in R^{(j)}} x_i \bmod 2}_{\text{poly of } O(\log s) \text{ deg.}}$ ($R^{(j)}$: rand. subset)

err prob $\frac{1}{2^t}$

Set $t = \Theta(\log s)$

Conditional Lower Bd.

If $\exists$ alg'm for $\{0,1\}$ dominance on N pts with $O(N^{1.99})$ time for $d = \omega(\log N)$

then $\exists$ alg'm for SAT with kn clauses with $O(1.999^n)$ time for all const k

(disproving Strong Exp Time Hypothesis (SETH))

Pf by Ex:

$(x_1 \vee x_2 \vee x_3 \vee x_4) \wedge (x_1 \vee \overline{x_2} \vee \overline{x_3} \vee x_4) \wedge \ldots$ is satisfiable

iff some pt $(\overline{x_1 \vee x_2}, \overline{x_1 \vee \overline{x_2}}, \ldots)$ is dominated by $(x_3 \vee x_4, \overline{x_3} \vee x_4, \ldots)$.

k-SAT reduces to dominance for $N = 2^{n/2}$ pts in dim $d \leq kn$ $\leq O(k \log N)$

$(N^{1.99} = 2^{.995n})$ □