Applications of WSPD

Closest Pair

Theorem 22 Let S be a set of n points in $\mathbb{R}^d$

A "closest pair" can be computed in $O(n \log n)$ time.

(d is a constant)

Proof: Construct WSPD w.r.t. $s > 2$

Let (A_i, B_i) contain the closest pair (p, q)

$p \in A_i$, $q \in B_i$ w.l.o.g.

Assume there is some $p' \neq p$ with $p' \in A_i$

$$\Rightarrow \quad |pq| \leq |pp'| \underset{\substack{\uparrow \\ \text{Lemma 17}}}{\leq} \frac{2}{s} |pq| < pq \qquad \lightning$$

$\Rightarrow \{p\} = A_i$ and analogously $q = \{B_i\}$

Consider all pair $(A_i, B_i) \rightsquigarrow O(n \log n)$ $\square$

$|A_i| = |B_i| = 1$

k-nearest neighbors / all nearest neighbors

Can we generalize the given approach?

Same proof shows: b nearest neighbor of a

$a \in A_i$ $b \in B_i$ (A_i, B_i) in WSPD for $s > 2$

$\Rightarrow A_i = \{a\}$ (a need not be the nearest neighbor of b)

First simple idea:

Check all pair (A_i, B_i) in WSPD for $(\{a\}, B_i)$

For all such $\{a\}$ check all $b \in B_i$ for the nearest neighbor

$O(|B_i|)$ for single $\{a\}$ could be quadratic!

Improvement and generalization:

$S = \{p_1, p_2, \dots, p_n\}$ n points in $\mathbb{R}^d$, d constant

Definition 23: $p \in S$ $k \leq n-1$

$kNN(p, S) :=$ Set of k nearest neighbors of p in S

$= \{p_{i_1}, p_{i_2}, \dots, p_{i_k}\} \subseteq S \setminus \{p\}$ satisfying

$|pp_{i_1}| \leq |pp_{i_2}| \leq \dots \leq |pp_{i_k}|$ and for each

$p_l \in S \setminus \{p, p_{i_1}, \dots, p_{i_k}\}$ we have

$|pp_{i_k}| = |pp_l|$ or $|pp_{i_k}| < |pp_l|$

Theorem 24 All sets $kNN(p,S)$ $p \in S$ can be computed in opt. $O(n \log n + nk)$ time.

Corollary 25: All nearest neighbors can be computed in optimal $O(n \log n)$ time.

Lower bounds Reduction of ε-closeness ✓
Check the nearest neighbors ✓

Now general position ✓ All $|pq|$ distance are different ✓

Lemma 26 Given n numbers x_i $i=1,\dots,n$, $k \in \{1,\dots,n\}$
The k-th largest (smallest) x_i can be computed in time $O(n)$. Black Box

($\rightsquigarrow$ order statistics / median alg. (Randomized / Deterministic)
$k=1$ min, $k=n$ max, $k=\lfloor \frac{n+1}{2} \rfloor$ $k=\lceil \frac{n+1}{2} \rceil$ median)

Lemma 27 (A,B) in WSPD for $s>2$
$a \in A, b \in B$ $b \in kNN(a,S)$ $\Rightarrow$ $|A| \leq 4$

Proof: $|A| > 2$ let $A' := A \setminus \{a\}$

A,B well-sep w.r.t. $s>0$ $\Rightarrow$ $|aa'| \overset{\text{Lemma 17}}{<} \frac{2}{s}|ab| < ab$

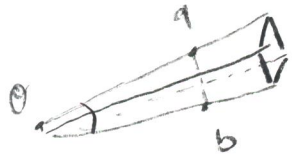
$\Rightarrow b \notin kNN(a,S)$ ↯ □

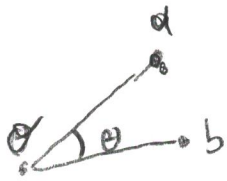
Definition 28:

$K \subseteq \mathbb{R}^d$ is a Θ-cone at $o \in \mathbb{R}^d$ $\quad \Theta \in [0, 2\pi]$

$:\Longleftrightarrow$ $o \in K$ and $\forall a, b \in K$ angle $(a, o, b) \leq \Theta$

Example: $\mathbb{R}^2$





(↑ spanned by some vectors)

(Black box Narasimhan/Smid Geometric spanner networks)

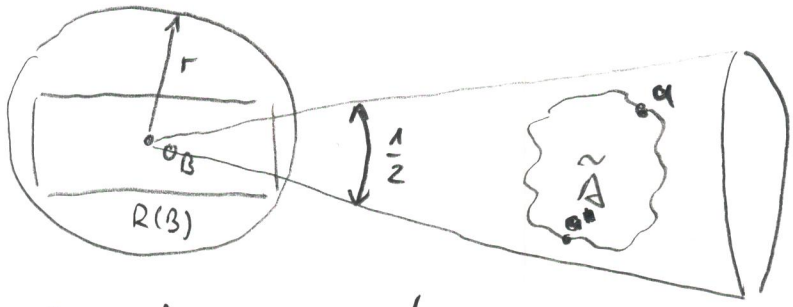
Lemma 29: $\forall d \; \forall \Theta \; \forall o$:

There is a decomposition of $\mathbb{R}^d$ into a constant (depend. on d, Θ) number of Θ cones at o. (need not be disjoint)

(So-called Θ-frame at o)

Can be constructed in constant time.

Lemma 30: B part of well-sep. pair (B, A_i)



Smallest sphere enclosing $R(B)$

$\frac{1}{2}$-cone at $o_B =: K$

$\tilde{A} \subseteq K$ collection of points from sets A_i well-sep. from B for $s > 2$

Then $a \in \tilde{A} \Rightarrow$ all $a' \in \tilde{A}$ closer to o_B than a are closer to a than any point of B is to a

<u>Proof:</u> w.l.o.g (scaling) radius 1.

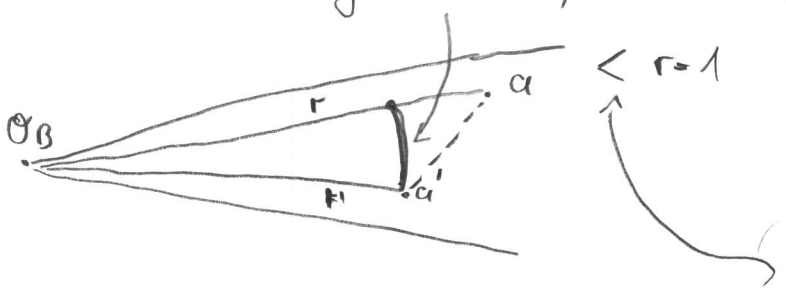
(Clear if a, a' are out of the same set $\bar{A}$ with $(B, \bar{A})$ well-sep.)

Let $a, a' \in \tilde{A}$ such that

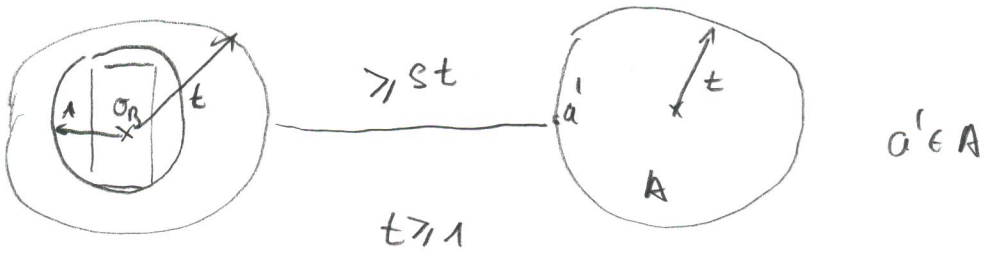
$$r' := |a' o_B| \leq |a\, o_B| =: r$$

and let $b \in B$. We show that $|aa'| < |ab|$ holds

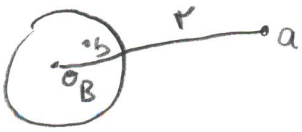
Clearly $|aa'| \leq r - r' + r' \operatorname{angle}(a, o_B a') < r - r' + \frac{1}{2} r' = r - \frac{1}{2} r'$



well-sep. condition $\Rightarrow$ $r' = |a' o_B| \geq st \geq s > 2$



On the other hand: $|ab| \geq r - 1$



$\square$

Proof of Theorem 24:

Idea compute $b \in kNN(a)$! $b \in B$

For each set $B \subseteq S$ that is generated as a node set $B := S_v$ for a node v during the construction of the Split tree T, we recursively compute a set $N(B) \subseteq S$ as follows:

($N(B)$ with $|N(B)| \in O(k)$ and $N(B)$ contains only points $a \in A$ where A is well-sep. from B w.r.t. s)

Invariant

Algorithm:

$N(\text{root}) := \emptyset$

For node set $(S_v =) B$ and $B \neq$ root (identify B with v)

$N(B)_{init} := N(\text{parent}(B)) \cup I(B)$ where

$I(B) := \{ a \mid \exists A \text{ with } a \in A, |A| \leq k \ (A,B) \text{ in WSPD} \}$

// Compute $N(B)$ from $N(B)_{init}$ //

Place $\frac{1}{2}$-frame at center O_B of smallest sphere around B

For each cone K of this frame

take the k elements closest to O_B of $N(B)_{init} \cap K$

$N(B) :=$ Union of these point sets, over all cones K in $\frac{1}{2}$-frame.

Observation:

Fulfills invariant: Number of cones is constant,
k points each
invariant parent(B), definition $\mathcal{I}(B)$

Now assume: $b \in kNN(a)$

$\Rightarrow$ ex. A, B $a \in A$ $|A| \leq k$, $b \in B$ (A,B) in WSPD

Lem. 27

$\Rightarrow a \in \mathcal{I}(B) \subseteq N(B)_{init}$

Suppose a does not survive in $N(B)$

$\Rightarrow$ a lies in some cone K that has $\geq k$ elements a' of $N(B)_{init}$ that are closer to o_B than a is

$\Rightarrow$ those a' are closer to a than b is $\Rightarrow$ ↯ $b \notin kNN(a)$

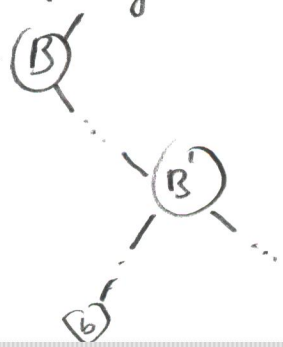
Lem. 30

$\tilde{A} := N(B)_{init} \cap K$

$\Rightarrow a \in N(B)$.

Show at $N(\{b\})$: Split Tree ends in leafs ✓

Recursion of computing $N(\,)$ to children B' of B on path to b



a is included in $N(B')$ init as member of $N(\text{parent}(B'))$

a is included in $N(B')$ by the same argument ($b \in B'$, $b \in kNN(a)$)

$\Rightarrow$ $\underline{\underline{a \in N(\{b\})}}$!

Construction of all sets $kNN(a,S)$ as follows:

- traverse all sets $N(\{b\})$ just computed
- for each a, collect all b where $a \in N(\{b\})$ $\rightsquigarrow$ List $L(a)$
- compute the k closest elements of a in $L(a)$ $\rightsquigarrow$ $kNN(a)$ [$b \in kNN(a) \Rightarrow a \in N(\{b\})$
  $\Rightarrow b \in L(a)$
  $\Rightarrow b$ is a kNN in $L(a)$]

<u>Correctness</u>: Just proved!

<u>Running time</u>: We have $\sum_{B \text{ node of } T} I(B) \in O(nk)$

($O(n)$ pairs, at most k from each)

Invariant $\Rightarrow$ $|N(\text{parent}(B))| \in O(k)$ (for each of n nodes)

All $I(B)$ computed in $O(nk)$ time $\Rightarrow$ $N(B)$ init in $O(nk)$

Subdividing $N(B)$ init into cases in overall $O(nk)$ time. (Size $O(nk)$)