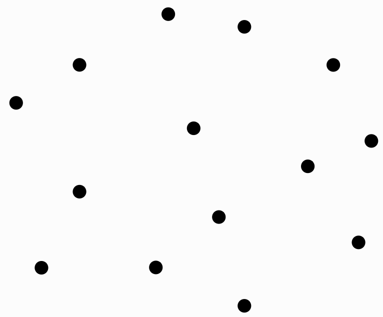
Das Closest-Pair-Problem



Literaturhinweis: Th. Ottmann, *Das Divide-and-Conquer-Prinzip*, in *Prinzipien des Algorithmenentwurfs*, Spektrum Akademischer Verlag 1997.

P endliche Menge von Punkten $p = (p_x, p_y) \in \mathbb{R}^2$

$\texttt{dist}(p, q)$: euklidischer Abstand von $p, q \in \mathbb{R}^2$

Gesucht

$$\delta(P) = \min \{ \texttt{dist}(p, q) \,;\, p, q \in P,\, p \neq q \}$$

allgemeiner für endliche $P, Q \subseteq \mathbb{R}^2$

$$\delta(P, Q) = \min \{ \texttt{dist}(p, q) \,;\, p \in P,\, q \in Q,\, p \neq q \}$$

also $\delta(P) = \delta(P, P)$

Algorithmisches Problem

- berechne $\delta(P)$ möglichst effizient
  (evtl. auch: bestimme ein Paar $p, q \in P$ mit $\texttt{dist}(p, q) = \delta(P)$)

Naive Lösung der Aufgabe

- berechne alle paarweisen Distanzen

$$\{ \texttt{dist}(p, q) \,;\, p, q \in P,\, p \neq q \}$$

- berechne das Minimum dieser Menge

Komplexität

- für $\sharp P = n$: $\binom{n}{2}$ Distanzen berechnen: $\mathcal{O}(n^2)$
- Minimum in $\mathcal{O}(n^2)$ Zahlen suchen: $\mathcal{O}(n^2)$ Vergleiche
- insgesamt ein $\mathcal{O}(n^2)$-Verfahren

Divide-and-Conquer Ansatz

- Zerlege P in zwei (ungefähr) gleich grosse Teile P_ℓ, P_r
- berechne (rekursiv) $\delta(P_\ell), \delta(P_r)$
- berechne $\delta(P_\ell, P_r)$
- berechne $\delta(P) = \min(\delta(P_\ell), \delta(P_r), \delta(P_\ell, P_r))$

Dieser Ansatz führt — bei naivem Vorgehen zur Berechnung von $\delta(P_\ell, P_r)$ — auf eine Rekursion für den Aufwand $t(n)$ an Operationen (bei $\sharp P = n$)

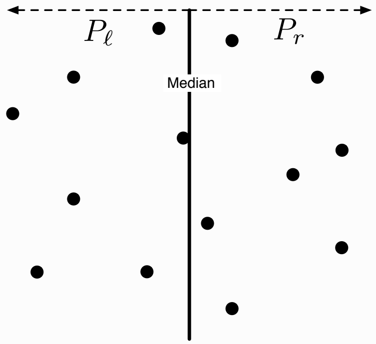
$$t(n) = 2 \cdot t(n/2) + 2 \cdot \left(\frac{n}{2}\right)^2 + 1$$
$$t(2) = 1$$

mit einer Lösung $t(n) \in \mathcal{O}(n^2)$

Geometrisches Divide-and-Conquer

- Zerlege P mit Hilfe des Medians m der x-Koordinaten der $p \in P$ in zwei (ungefähr) gleich grosse Teile

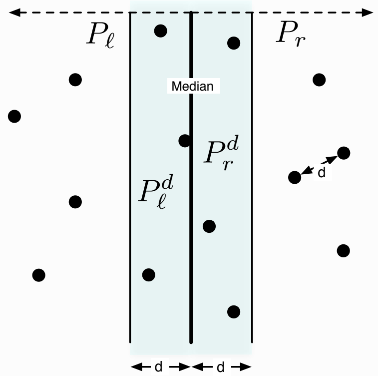
$$P_\ell = \{\, p \in P \,;\, p_x \le m \,\} \qquad P_r = \{\, p \in P \,;\, m < p_x \,\}$$



- Bemerkung: Median-Berechnung für n Elemente geht in $\mathcal{O}(n)$
- berechne (rekursiv) $\delta(P_\ell), \delta(P_r)$ und $d = \min(\delta(P_\ell), \delta(P_r))$

- bestimme

$$P_\ell^d = \{\, p \in P \,;\, m - d \le p_x \le m \,\} \qquad P_r^d = \{\, p \in P \,;\, m < p_x \le m + d \,\}$$

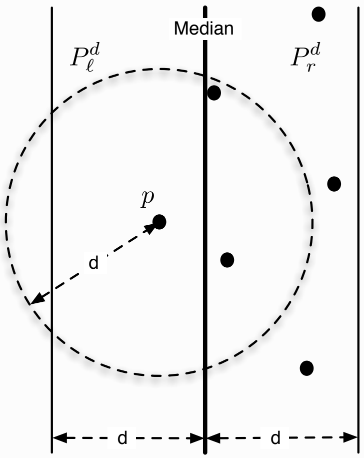


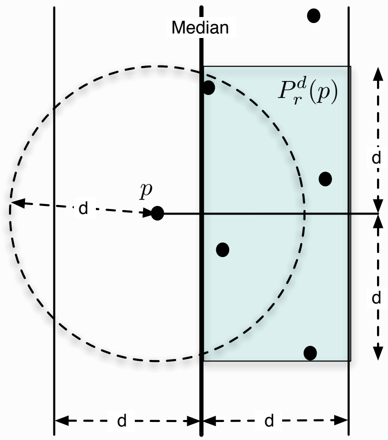
- berechne $\delta(P_\ell^d, P_r^d)$ und somit

$$\delta(P) = \min(d, \delta(P_\ell, P_r)) = \min(d, \delta(P_\ell^d, P_r^d))$$

Vorsicht: auch P_ℓ^d und P_r^d können noch $\mathcal{O}(n)$ Elemente enthalten!

ABER: zu jedem $p \in P_\ell^d$ kann es nur sehr wenige $q \in P_r^d$ geben, die "nahe" ($\approx d$) von p liegen:





$$\begin{aligned} \mathtt{dist}(p,q) < d \quad &\Rightarrow \quad p_y - d < q_y < p_y + d \\ &\Rightarrow \quad q \in \underbrace{P_r^d \cap \{\, q \,;\, p_y - d < q_y < p_y + d) \,\}}_{P_r^d(p)} \end{aligned}$$

$$\delta(P_\ell, P_r) = \delta(P_\ell^d, P_r^d)$$
$$= \min_{p \in P_\ell^d} \delta(p, P_r^d) = \min_{p \in P_\ell^d} \delta(p, P_r^d(p))$$

Aus der Geometrie folgt (grosszügig abgeschätzt!)

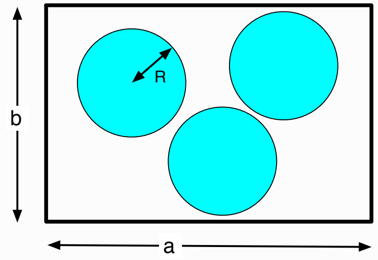
$$\sharp P_r^d(p) \leq 7$$

d.h. zu jedem $p \in P_\ell^d$ gibt es höchstens sieben $q \in P_r^d$ mit $\texttt{dist}(p,q) < d$,

d.h. man muss bei geschickter (!!) Organisation der Daten nur $\mathcal{O}(n)$ Operationen durchführen, um $\delta(P_\ell, P_r)$ zu berechnen!

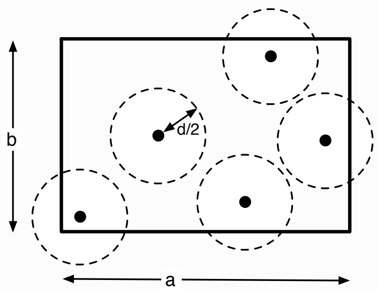
Eine geometrische Überlegung

- Wieviele disjunkte, offene Kreisscheiben mit Radius R kann man in einem $(a \times b)$-Rechteck unterbringen?



$$\text{Anzahl} \quad k(a,b,R) \leq \frac{a \cdot b}{\pi \cdot R^2}$$

- Wieviele verschiedene Punkte, die paarweise einen Abstand $\geq d$ haben, kann man in einem $(a \times b)$-Rechteck unterbringen?



$$\text{Anzahl} \leq k(a+d, b+d, d/2) \leq \frac{(a+d)(b+d)}{\pi \cdot (d/2)^2}$$

P_ℓ^d nach y-Koordinaten sortiert

$$\begin{array}{ccccccc} P_\ell^d & : & p^{(1)} & p^{(2)} & p^{(3)} & \dots & p^{(s)} \\ & & p_y^{(1)} \leq & p_y^{(2)} \leq & p_y^{(3)} \leq & \dots & < p_y^{(s)} \end{array}$$

P_r^d nach y-Koordinaten sortiert

$$\begin{array}{ccccccc} P_r^d & : & q^{(1)} & q^{(2)} & q^{(3)} & \dots & q^{(t)} \\ & & q_y^{(1)} \leq & q_y^{(2)} \leq & q_y^{(3)} \leq & \dots & < q_y^{(t)} \end{array}$$

Kandidaten für $\texttt{dist}(p,q)$ in Blöcken der Länge ≤ 7

$$\begin{array}{cccccccccc} P_\ell^d & : & p^{(1)} & p^{(2)} & p^{(3)} & \dots & \boxed{p} & \dots & p^{(s)} \\ P_r^d & : & q^{(1)} & q^{(2)} & q^{(3)} & \dots & \underbrace{\boxed{q^{(i)}\ q^{(i+1)}\ \dots\ q^{(i+j)}}}_{P_\ell^d(p)} & \dots & q^{(s)} \end{array}$$

cp-Algorithmus

- Zerlege P mit Hilfe des Medians m der x-Koordinaten der $p \in P$ in zwei (ungefähr) gleich grosse Teile

$$P_\ell = \{\, p \in P\,; p_x \leq m \,\} \qquad P_r = \{\, p \in P\,;\, m < p_x \,\}$$

- berechne (rekursiv) $\delta(P_\ell), \delta(P_r)$ und $d = \min(\delta(P_\ell), \delta(P_r))$
- bestimme

$$P_\ell^d = \{\, p \in P\,;\, m - d \leq p_x \leq m \,\} \quad P_r^d = \{\, p \in P\,;\, m < p_x \leq m + d \,\}$$

- sortiere die Mengen P_ℓ^d und P_r^d nach den y-Koordinaten ihrer Elemente
- berechne $\delta(P_\ell^d, P_r^d)$ mit $\mathcal{O}(n)$ Operationen
- $\delta(P) = \min(d, \delta(P_\ell^d, P_r^d))$

Komplexität $t(n)$ des cp-Algorithmus

- Medianberechnung und Zerlegung $P = P_\ell \uplus P_r \,:\, \mathcal{O}(n)$
- Herausziehen von P_ℓ^d und $P_r^d \,:\, \mathcal{O}(n)$
- Sortieren von P_ℓ^d und $P_r^d \,:\, \mathcal{O}(n \log n)$
- Berechnen von $\delta(P_\ell^d, P_r^d) \,:\, \mathcal{O}(n)$
- divide-and-conquer Rekursionsgleichung

$$\begin{aligned} t(n) &= 2\,t(n/2) + \mathcal{O}(n \log n) \\ t(2) &= 1 \end{aligned}$$

- Verhalten der Lösung der Rekursionsgleichung

$$t(n) \in \mathcal{O}(n \log^2 n)$$

- Man kann das Sortieren auf jeder Rekursionsstufe durch einmaliges Sortieren zu Beginn und geschickte Verwaltung dieser Information ersetzen, dadurch sogar Gesamtkomplexität $t(n) \in \mathcal{O}(n \log n)$ erreichen