Coresets

Sándor Kisfaludi-Bak

Computaional Geometry
Summer semester 2020

max planck institut
informatik

Overview

- Minimum volume bounding box

Overview

- Minimum volume bounding box

- Coreset for directional witdh, usage

Overview

- Minimum volume bounding box

- Coreset for directional witdh, usage

- Computing a corset for directional width

Overview

- Minimum volume bounding box
- Coreset for directional witdh, usage
- Computing a corset for directional width
- r-division for planar graphs

The John ellipsoid

Ellipsoid: picture of unit ball under invertible linear transformation

Theorem. Any compact convex object $K \subset \mathbb{R}^d$ has a unique maximum volume ellipsoid $\mathcal{E} \subseteq K$.

The John ellipsoid

Ellipsoid: picture of unit ball under invertible linear transformation

Theorem. Any compact convex object $K \subset \mathbb{R}^d$ has a unique maximum volume ellipsoid $\mathcal{E} \subseteq K$.

Theorem (John 1948). For any compact convex $K \subset \mathbb{R}^d$ with $\mathcal{E}$ centered at the origin, $\mathcal{E} \subseteq K \subseteq d\mathcal{E}$.

Minimum volume bounding box

Min. volume bounding box of P: smallest volume box (of arbitrary rotation) containing P

Theorem. A bounding box B of P can be computed in $O(d^2 n)$ time s.t.

$$(i)\ \mathrm{Vol}(B_{opt}(P)) \leq \mathrm{Vol}(B) \leq 2^d d! \mathrm{Vol}(B_{opt}(P))$$

and (ii) there is a shift $x \in \mathbb{R}^d$ and $c > 0$ that depends only on d, s.t. $x + cB \subset \mathrm{conv}(P)$.

Minimum volume bounding box

Min. volume bounding box of P: smallest volume box (of arbitrary rotation) containing P

Theorem. A bounding box B of P can be computed in $O(d^2 n)$ time s.t.

$$(i)\ \mathrm{Vol}(B_{opt}(P)) \leq \mathrm{Vol}(B) \leq 2^d d! \mathrm{Vol}(B_{opt}(P))$$

and (ii) there is a shift $x \in \mathbb{R}^d$ and $c > 0$ that depends only on d, s.t. $x + cB \subset \mathrm{conv}(P)$.

Let $s \in P$ arbitrary and let $s' \in P$ most distant form s.
If t, t' realize the diameter of P, then

$$\mathrm{diam}(P) = |tt'| \leq |ts| + |st'| \leq 2|ss'|$$

Minimum volume bounding box

Min. volume bounding box of P: smallest volume box (of arbitrary rotation) containing P

Theorem. A bounding box B of P can be computed in $O(d^2 n)$ time s.t.

$$(i)\ \mathrm{Vol}(B_{opt}(P)) \leq \mathrm{Vol}(B) \leq 2^d d! \mathrm{Vol}(B_{opt}(P))$$

and (ii) there is a shift $x \in \mathbb{R}^d$ and $c > 0$ that depends only on d, s.t. $x + cB \subset \mathrm{conv}(P)$.

Let $s \in P$ arbitrary and let $s' \in P$ most distant form s.
If t, t' realize the diameter of P, then

$$\mathrm{diam}(P) = |tt'| \leq |ts| + |st'| \leq 2|ss'|$$

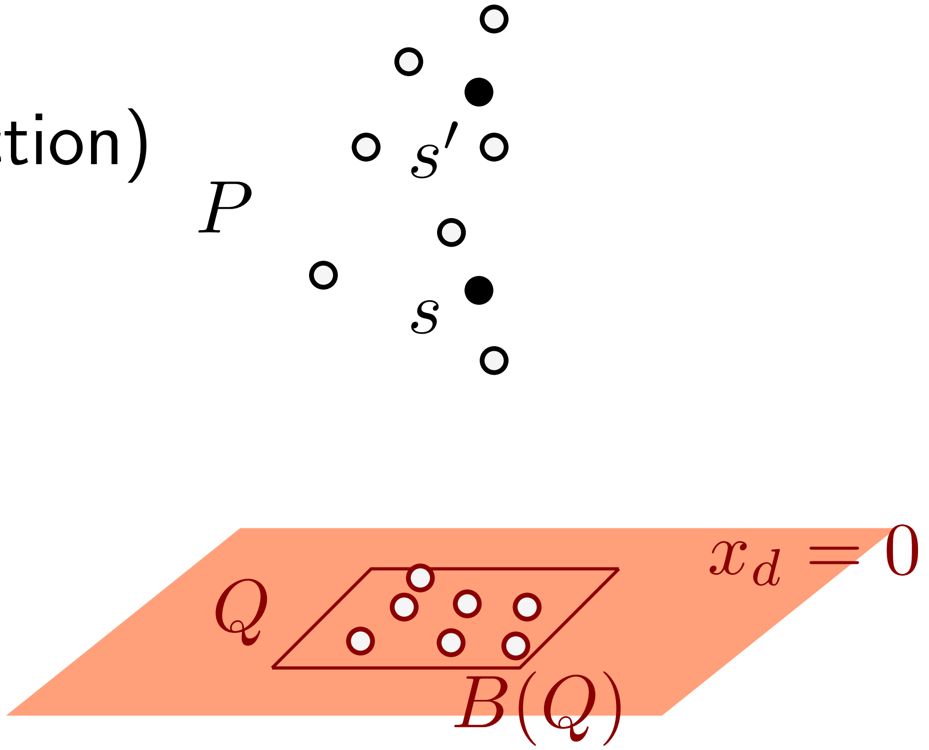
Wlog. ss' parallel to x_d axis.
$\pi :=$ perpendicular projection to $x_d = 0$.

Induction setup

Use induction on dimension.

$d = 1$ trivial.

$Q := \pi(P)$
$B(Q) :=$ bounding box of Q (induction)



Induction setup

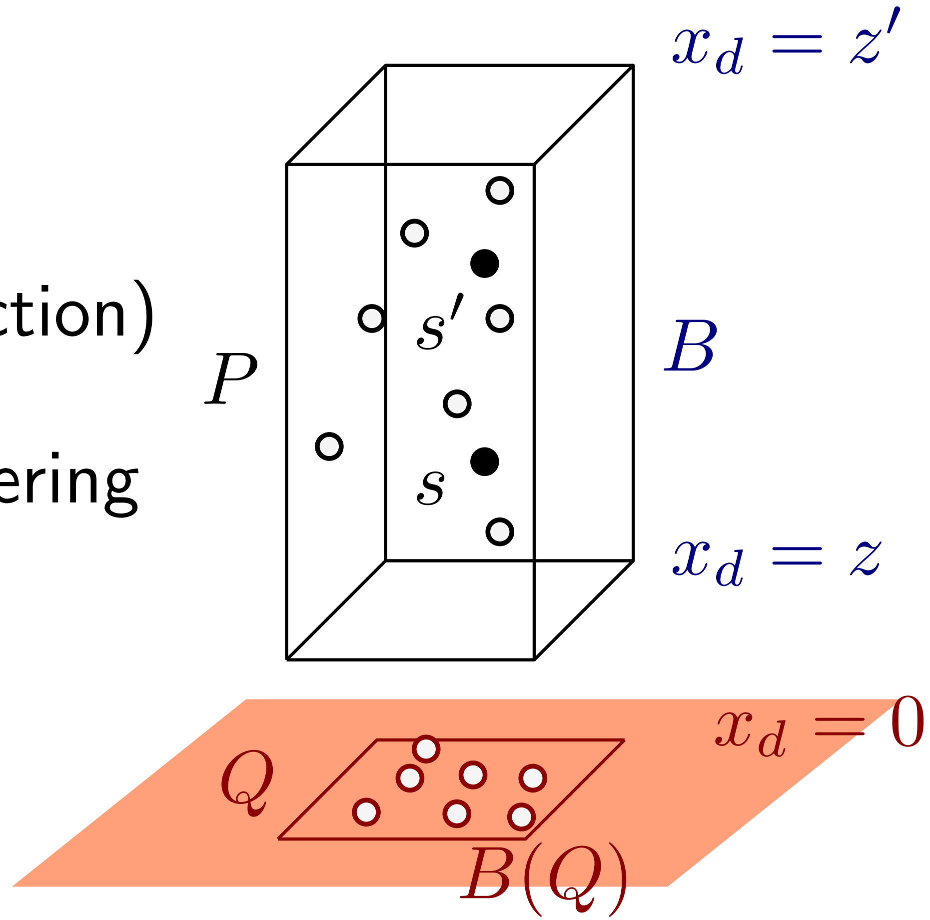
Use induction on dimension.

$d = 1$ trivial.

$Q := \pi(P)$
$B(Q) :=$ bounding box of Q (induction)

$[z, z'] :=$ shortest iv on x_d axis covering projection of P

$$B := B(Q) \times [z, z']$$



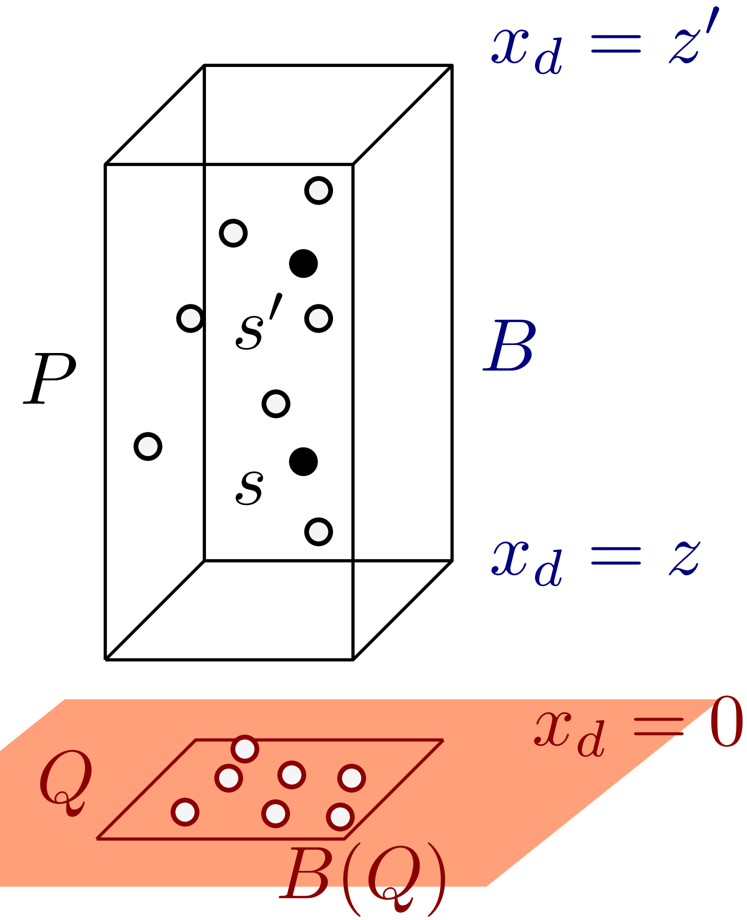
Induction setup

Use induction on dimension.

$d = 1$ trivial.

$Q := \pi(P)$
$B(Q) :=$ bounding box of Q (induction)

$[z, z'] :=$ shortest iv on x_d axis covering projection of P



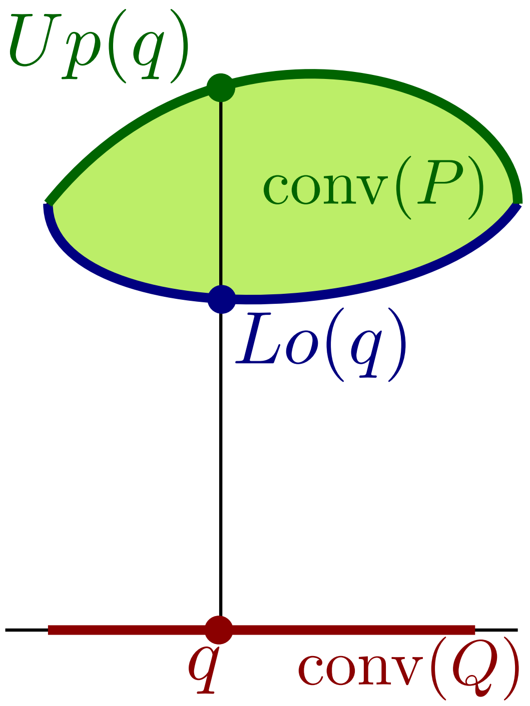
$$B := B(Q) \times [z, z']$$

Need: $\mathrm{Vol}_d(\mathrm{conv}(P)) \geq \mathrm{Vol}_d(B)/(2^d d!)$

Shifting down, pyramid

Upper hull $\mathrm{conv}^{\uparrow}(P)$ as function:
$Up : \mathrm{conv}(Q) \to \mathbb{R}^d$ is concave

Lower hull $\mathrm{conv}^{\downarrow}(P)$ as function:
$Lo : \mathrm{conv}(Q) \to \mathbb{R}^d$ is convex



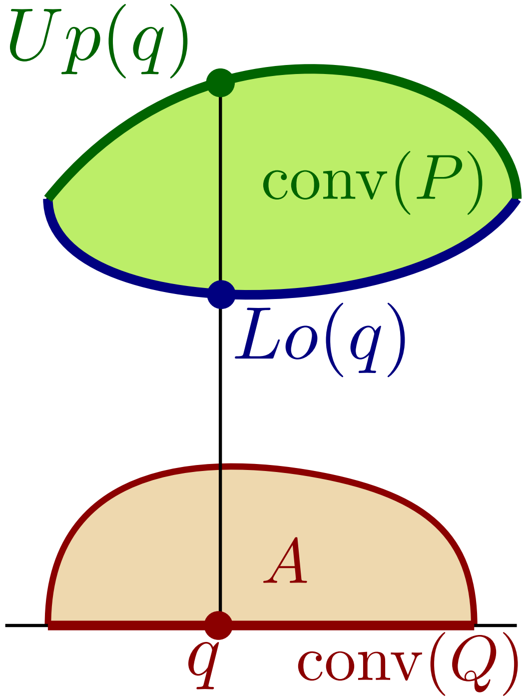
Shifting down, pyramid

Upper hull $\mathrm{conv}^{\uparrow}(P)$ as function:
$Up : \mathrm{conv}(Q) \to \mathbb{R}^d$ is concave

Lower hull $\mathrm{conv}^{\downarrow}(P)$ as function:
$Lo : \mathrm{conv}(Q) \to \mathbb{R}^d$ is convex

$Up - Lo$ is concave

$$\Rightarrow A := \bigcup_{q \in \mathrm{conv}(Q)} [0, Up(q) - Lo(q)] \text{ is convex}$$



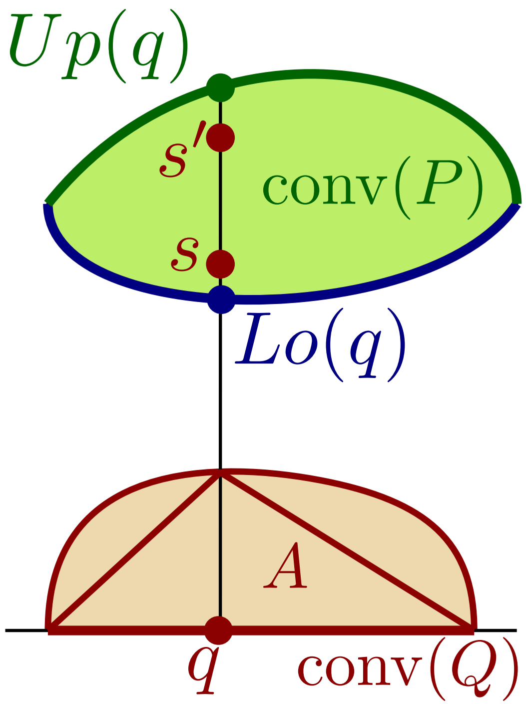
Shifting down, pyramid

Upper hull $\operatorname{conv}^{\uparrow}(P)$ as function:
$Up : \operatorname{conv}(Q) \to \mathbb{R}^d$ is concave

Lower hull $\operatorname{conv}^{\downarrow}(P)$ as function:
$Lo : \operatorname{conv}(Q) \to \mathbb{R}^d$ is convex

$Up - Lo$ is concave

$$\Rightarrow A := \bigcup_{q \in \operatorname{conv}(Q)} [0, Up(q) - Lo(q)] \text{ is convex}$$



At $\pi(s)$, height of A is at least $|ss'|$.
A contains *pyramid* with base $\operatorname{conv}(Q)$ and pole length $\geq |ss'|$.

Bounding box approximation quality

$$
\begin{aligned}
\mathrm{Vol}_d(B) &\geq \mathrm{Vol}_d(B_{opt}) \\
&\geq \mathrm{Vol}_d(\mathrm{conv}(P)) = \mathrm{Vol}_d(A) \\
&\geq \mathrm{Vol}(pyramid) \\
&\geq \frac{\mathrm{Vol}_{d-1}(\mathrm{conv}(Q))|ss'|}{d} \\
&\geq \frac{\mathrm{Vol}_{d-1}\Big(B(Q)/(2^{d-1}(d-1)!)\Big)2|ss'|}{2d} \\
&\geq \frac{\mathrm{Vol}_{d-1}(B(Q))|zz'|}{2^d d!} \\
&= \frac{\mathrm{Vol}_d(B)}{2^d d!}
\end{aligned}
$$

Bounding box approximation quality

$$\begin{aligned}
\mathrm{Vol}_d(B) &\geq \mathrm{Vol}_d(B_{opt}) \\
&\geq \mathrm{Vol}_d(\mathrm{conv}(P)) = \mathrm{Vol}_d(A) \\
&\geq \mathrm{Vol}(pyramid) \\
&\geq \frac{\mathrm{Vol}_{d-1}(\mathrm{conv}(Q))|ss'|}{d} \\
&\geq \frac{\mathrm{Vol}_{d-1}\Big(B(Q)/(2^{d-1}(d-1)!)\Big)2|ss'|}{2d} \\
&\geq \frac{\mathrm{Vol}_{d-1}(B(Q))|zz'|}{2^d d!} \\
&= \frac{\mathrm{Vol}_d(B)}{2^d d!}
\end{aligned}$$

Running time: $T(n, d) = O(nd) + T(n, d-1)$

$\Rightarrow$ Runs in $O(nd^2)$.

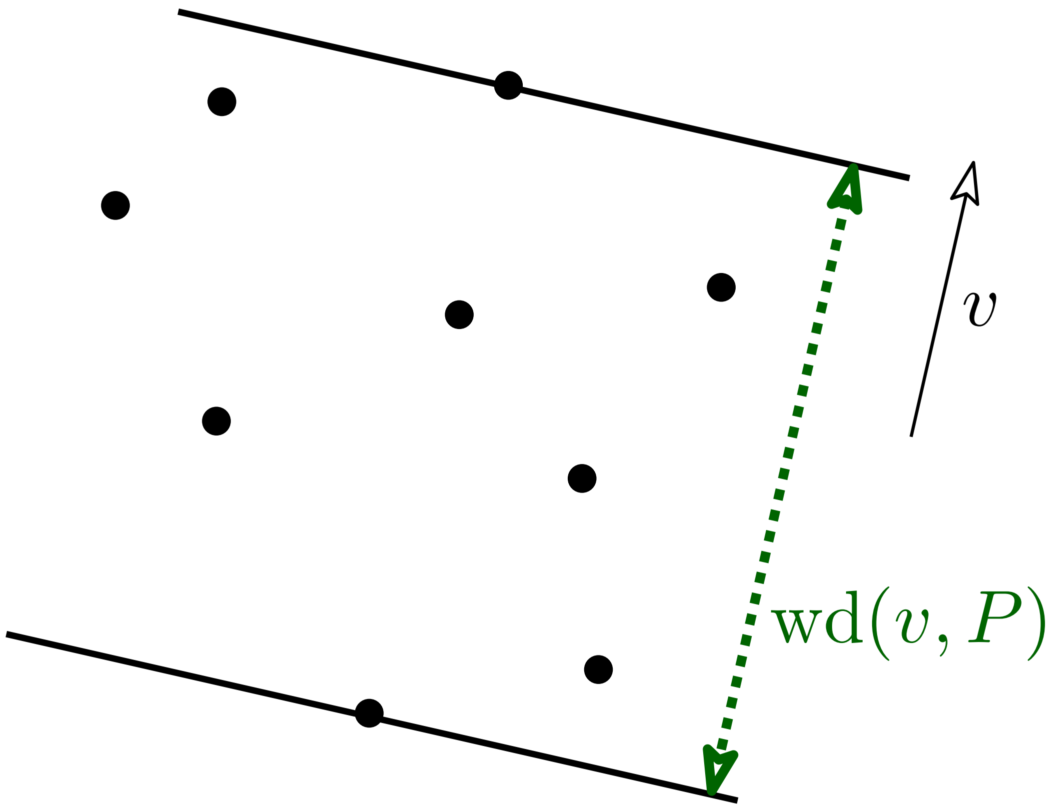
(i)

Coreset for directional width, usage

Directional width

Definition. The directional width of $P \subset \mathbb{R}^d$ w.r.t. $v \in \mathbb{R}^d \setminus \{0\}$ is

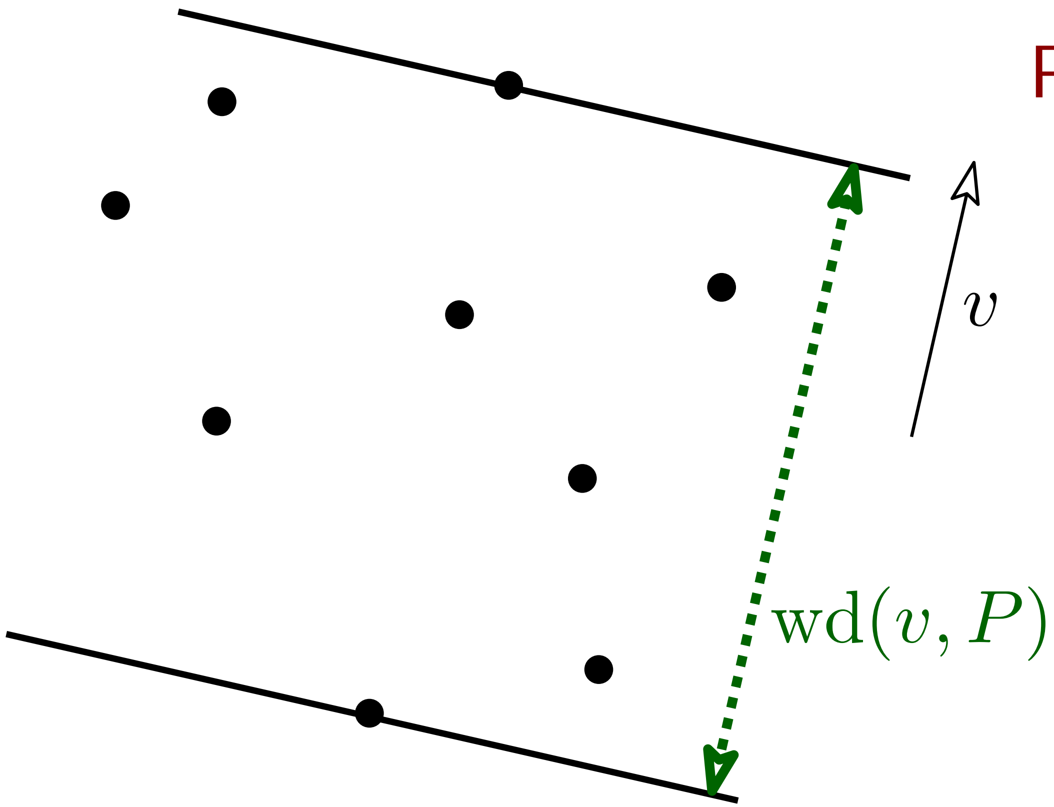
$$\mathrm{wd}(v, P) := \max_{p \in P} \langle v, p \rangle - \min_{p \in P} \langle v, p \rangle$$



Directional width

Definition. The directional width of $P \subset \mathbb{R}^d$ w.r.t. $v \in \mathbb{R}^d \setminus \{0\}$ is

$$\mathrm{wd}(v, P) := \max_{p \in P} \langle v, p \rangle - \min_{p \in P} \langle v, p \rangle$$



Properties:

- translation invariant
- scales linearly
- $\mathrm{wd}(v, P) = \mathrm{wd}(v, \mathrm{conv}(P))$
- monotone: if $Q \subset P$, then $\mathrm{wd}(v, Q) \leq \mathrm{wd}(v, P)$

Coreset for directional width

Definition. The set $S \subset P$ is an ε-coreset for directional width if for every unit vector v

$$\mathrm{wd}(v, S) \geq (1 - \varepsilon)\mathrm{wd}(v, P)$$

Coreset for directional width

Definition. The set $S \subset P$ is an ε-coreset for directional width if for every unit vector v

$$\mathrm{wd}(v, S) \geq (1 - \varepsilon)\mathrm{wd}(v, P)$$

- Captures the geometry of P, and (hopefully) is much smaller than P.

Coreset for directional width

Definition. The set $S \subset P$ is an ε-coreset for directional width if for every unit vector v

$$\mathrm{wd}(v, S) \geq (1-\varepsilon)\mathrm{wd}(v, P)$$

- Captures the geometry of P, and (hopefully) is much smaller than P.
- If S is ε-coreset of S' and S' is ε'-coreset of P, then S is $(\varepsilon + \varepsilon')$-coreset of P

Coreset for directional width

Definition. The set $S \subset P$ is an ε-coreset for directional width if for every unit vector v

$$\mathrm{wd}(v, S) \geq (1-\varepsilon)\mathrm{wd}(v, P)$$

- Captures the geometry of P, and (hopefully) is much smaller than P.
- If S is ε-coreset of S' and S' is ε'-coreset of P, then S is $(\varepsilon + \varepsilon')$-coreset of P
- If S is ε-coreset of P and S' is ε-coreset of P', then $S \cup S'$ is ε-coreset of $P \cup P'$

Usage 1: min volume bounding box

Lemma. Let $\varepsilon > 0$, $P \subset \mathbb{R}^d$, and let S be a δ-coreset of P for directional width ($\delta = \varepsilon/(8d)$). Then

$$\mathrm{Vol}((1+3\delta)\mathcal{B}(S)) \leq (1+\varepsilon)\mathrm{Vol}(\mathcal{B}(P)).$$

and $(1+3\delta)\mathcal{B}(S)$ contains P.

Proof.

Usage 1: min volume bounding box

Lemma. Let $\varepsilon > 0$, $P \subset \mathbb{R}^d$, and let S be a δ-coreset of P for directional width ($\delta = \varepsilon/(8d)$). Then

$$\mathrm{Vol}((1+3\delta)\mathcal{B}(S)) \leq (1+\varepsilon)\mathrm{Vol}(\mathcal{B}(P)).$$

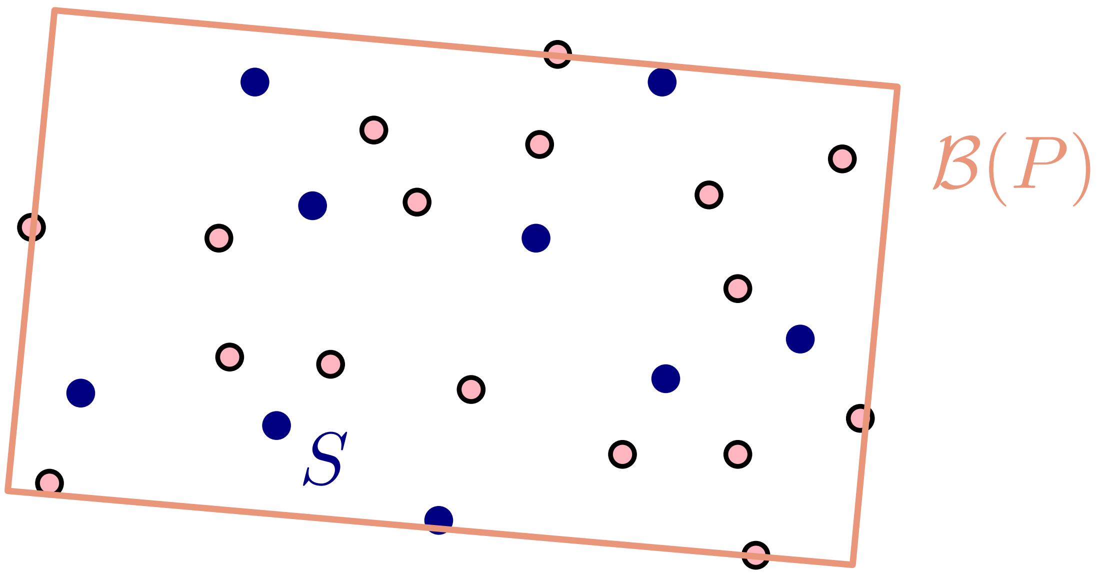
and $(1+3\delta)\mathcal{B}(S)$ contains P.

Proof.

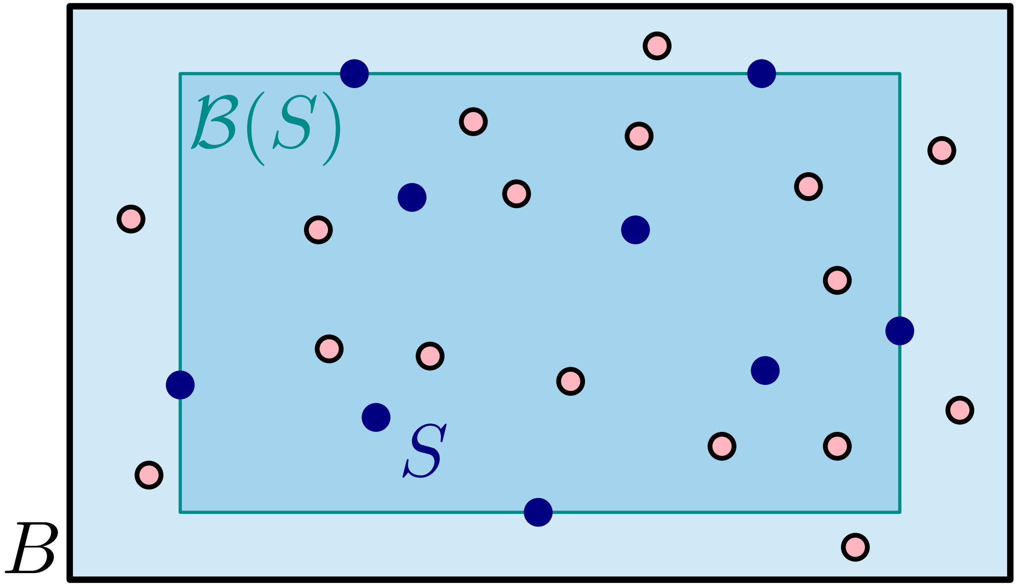
Volume claim: $(1+3\delta)^d < (1+\varepsilon)$

Need: $B := (1+3\delta)\mathcal{B}(S)$ contains P

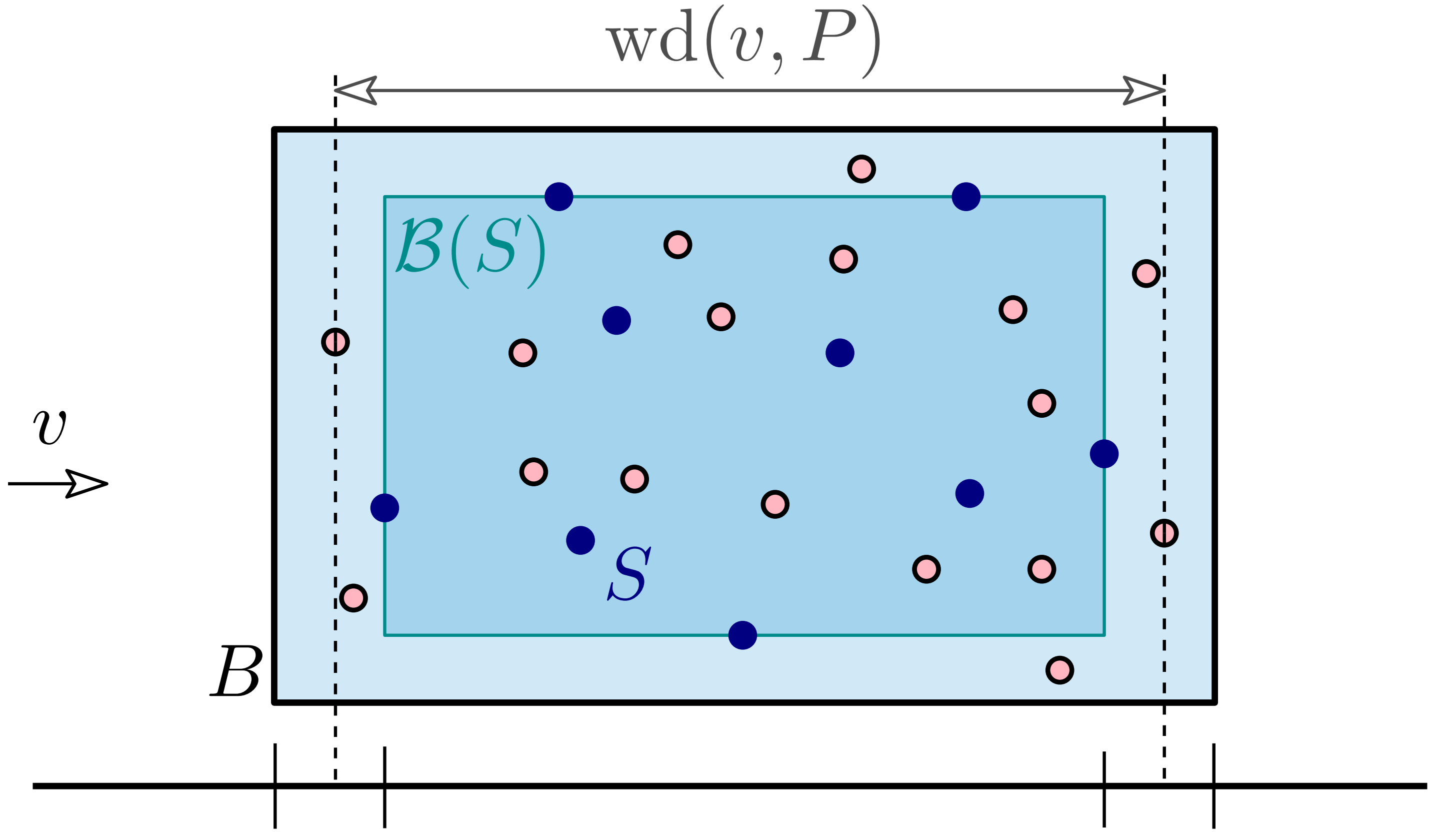
Projecting to a line



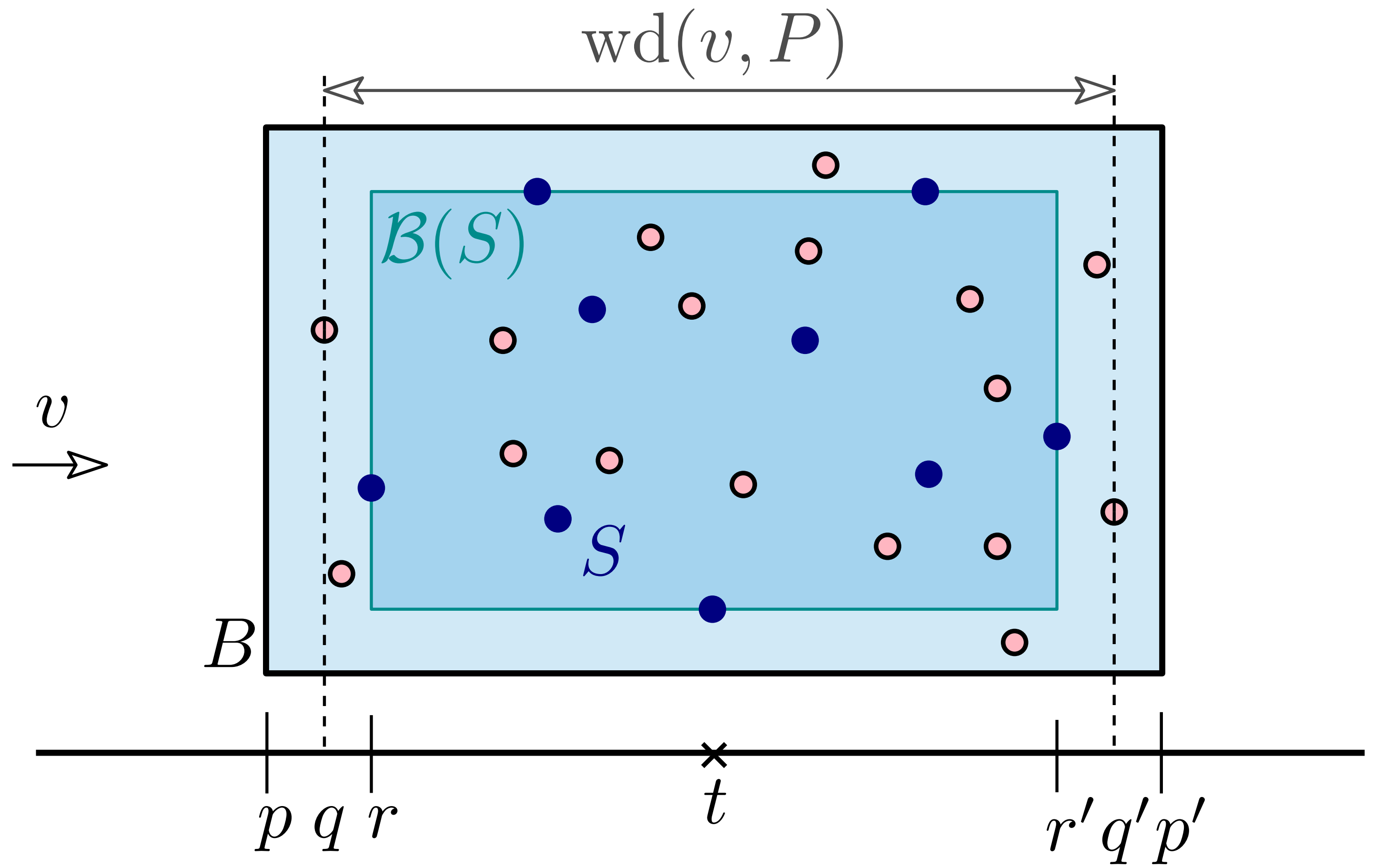
Projecting to a line



Projecting to a line



Projecting to a line



$(1-\delta)|qq'| \leq |rr'| = 2|tr|$ as S is δ-coreset.

$$|tq| \leq |tr| + \delta|qq'| \leq \left(1 + \frac{2\delta}{1-\delta}\right)|tr| \leq (1+3\delta)|tr| = |tp|$$

Usage 2: minimum enclosing ball

Lemma. If S is an $\varepsilon/4$-coreset of P for directional width, then $(1+\varepsilon)\mathcal{B}(S)$ contains P.

Proof.

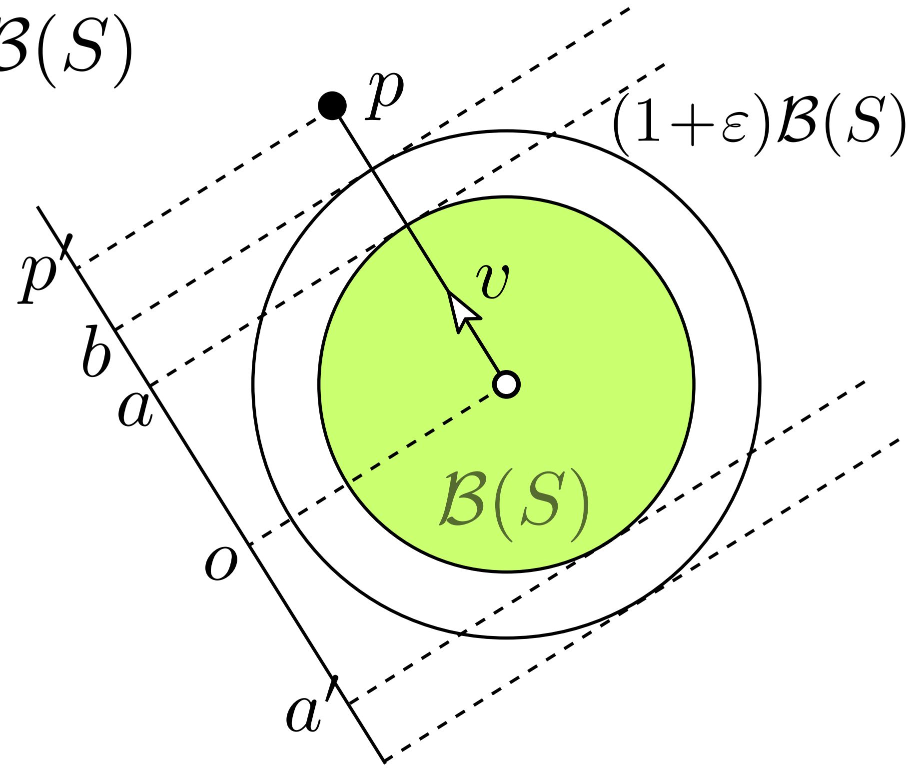
Usage 2: minimum enclosing ball

Lemma. If S is an $\varepsilon/4$-coreset of P for directional width, then $(1+\varepsilon)\mathcal{B}(S)$ contains P.

Proof.

Suppose $p \in P$ is outside $(1+\varepsilon)\mathcal{B}(S)$

$\Rightarrow |op'| > |ob| = (1+\varepsilon)|oa|$



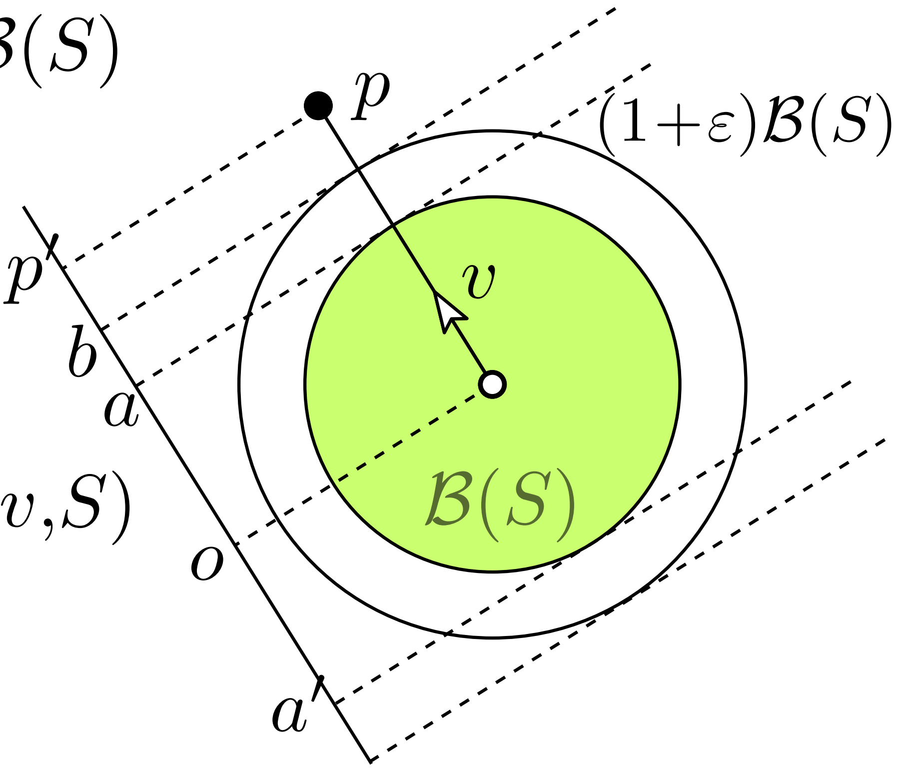
Usage 2: minimum enclosing ball

Lemma. If S is an $\varepsilon/4$-coreset of P for directional width, then $(1+\varepsilon)\mathcal{B}(S)$ contains P.

Proof.

Suppose $p \in P$ is outside $(1+\varepsilon)\mathcal{B}(S)$

$$\Rightarrow |op'| > |ob| = (1+\varepsilon)|oa|$$



$$\mathrm{wd}(v,P) - \mathrm{wd}(v,S) \geq \frac{\varepsilon}{2}|aa'| \geq \frac{\varepsilon}{2}\mathrm{wd}(v,S)$$

$$\mathrm{wd}(v,S) \leq \frac{1}{1+\varepsilon/2}\mathrm{wd}(v,P) < (1-\frac{\varepsilon}{4})\mathrm{wd}(v,P)$$

□

Computing a coreset for directional width (Agarwal, Har-Peled, Varadarajan 2004)

Coreset computation: partition into pillars

Theorem. Given $\varepsilon > 0$ and $P \subset \mathbb{R}^d$, we can compute an ε-coreset $S \subseteq P$ of size at most $|S| = O(1/\varepsilon^{d-1})$ in $O(n)$ time (where d is a fixed constant).

Coreset computation: partition into pillars

Theorem. Given $\varepsilon > 0$ and $P \subset \mathbb{R}^d$, we can compute an ε-coreset $S \subseteq P$ of size at most $|S| = O(1/\varepsilon^{d-1})$ in $O(n)$ time (where d is a fixed constant).

Proof.
B is bounding box s.t. $c_d B \subset \operatorname{conv}(P) \subset B$

(takes $O(d^2 n)$)

Let $M = \lceil \frac{2}{\varepsilon c_d} \rceil$, divide B into $M \times \cdots \times M$ grid
Pillar of cell $(i_1, \ldots, i_d)$ is $(i_1, \ldots, i_{d-1})$

Coreset computation: partition into pillars

Theorem. Given $\varepsilon > 0$ and $P \subset \mathbb{R}^d$, we can compute an ε-coreset $S \subseteq P$ of size at most $|S| = O(1/\varepsilon^{d-1})$ in $O(n)$ time (where d is a fixed constant).

Proof.
B is bounding box s.t. $c_d B \subset \mathrm{conv}(P) \subset B$

(takes $O(d^2 n)$)

Let $M = \lceil \frac{2}{\varepsilon c_d} \rceil$, divide B into $M \times \cdots \times M$ grid
Pillar of cell $(i_1, \ldots, i_d)$ is $(i_1, \ldots, i_{d-1})$

$$\underbrace{\text{For each of the } M^{d-1} \text{ pillars, find max and min } x_d\text{-coordiante}}_{S}$$

Coreset computation: partition into pillars

Theorem. Given $\varepsilon > 0$ and $P \subset \mathbb{R}^d$, we can compute an ε-coreset $S \subseteq P$ of size at most $|S| = O(1/\varepsilon^{d-1})$ in $O(n)$ time (where d is a fixed constant).

Proof.
B is bounding box s.t. $c_d B \subset \mathrm{conv}(P) \subset B$

(takes $O(d^2 n)$)

Let $M = \lceil \frac{2}{\varepsilon c_d} \rceil$, divide B into $M \times \cdots \times M$ grid
Pillar of cell $(i_1, \ldots, i_d)$ is $(i_1, \ldots, i_{d-1})$

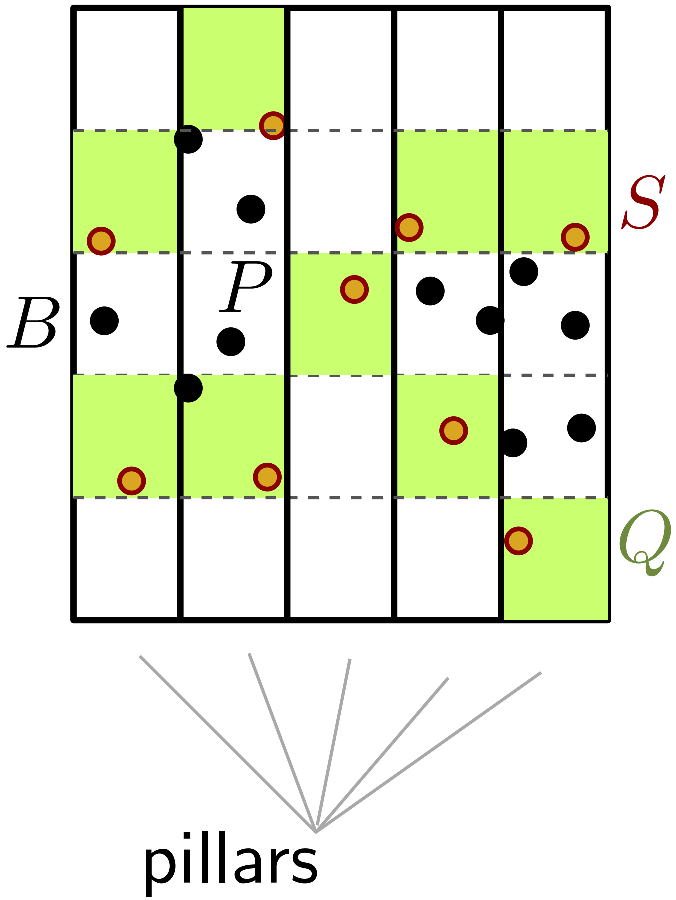
$$\underbrace{\text{For each of the } M^{d-1} \text{ pillars, find max and min } x_d\text{-coordiante}}_{S}$$

$|S| = 2M^{d-1} = O(1/\varepsilon^{d-1})$, need: S is coreset

Coreset from pillars

Let Q = union of cells containg a point of S.

$P \subset \mathrm{conv}(Q) \Rightarrow \mathrm{wd}(v, \mathrm{conv}(Q)) \geq \mathrm{wd}(v, P)$ for all v

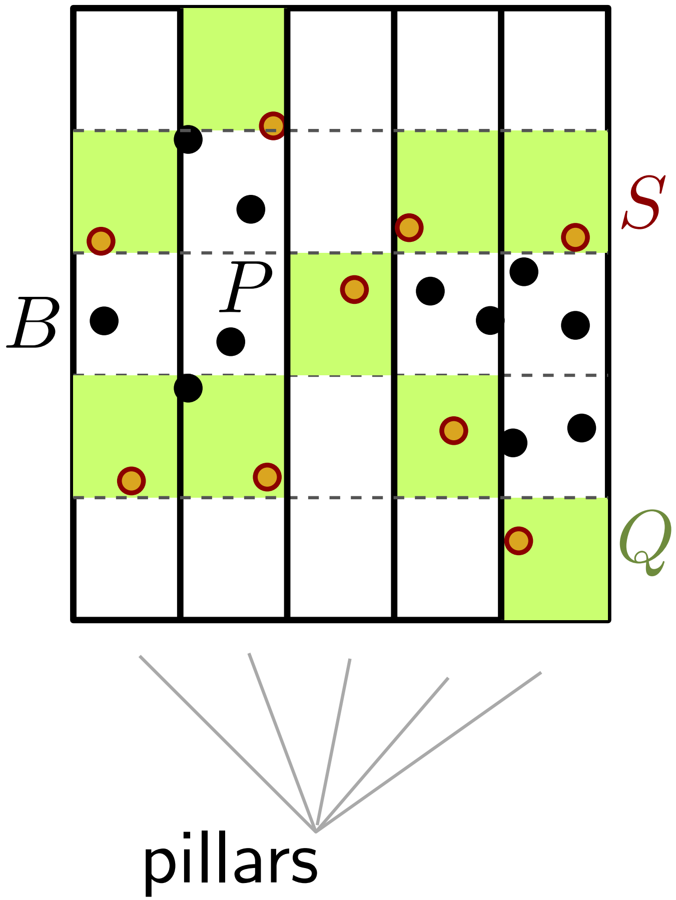


Coreset from pillars

Let Q = union of cells containg a point of S.

$P \subset \mathrm{conv}(Q) \Rightarrow \mathrm{wd}(v, \mathrm{conv}(Q)) \geq \mathrm{wd}(v, P)$ for all v

$\mathrm{wd}(v, S) + 2\mathrm{wd}(v, B/M) \geq \mathrm{wd}(v, Q) \geq \mathrm{wd}(v, P)$



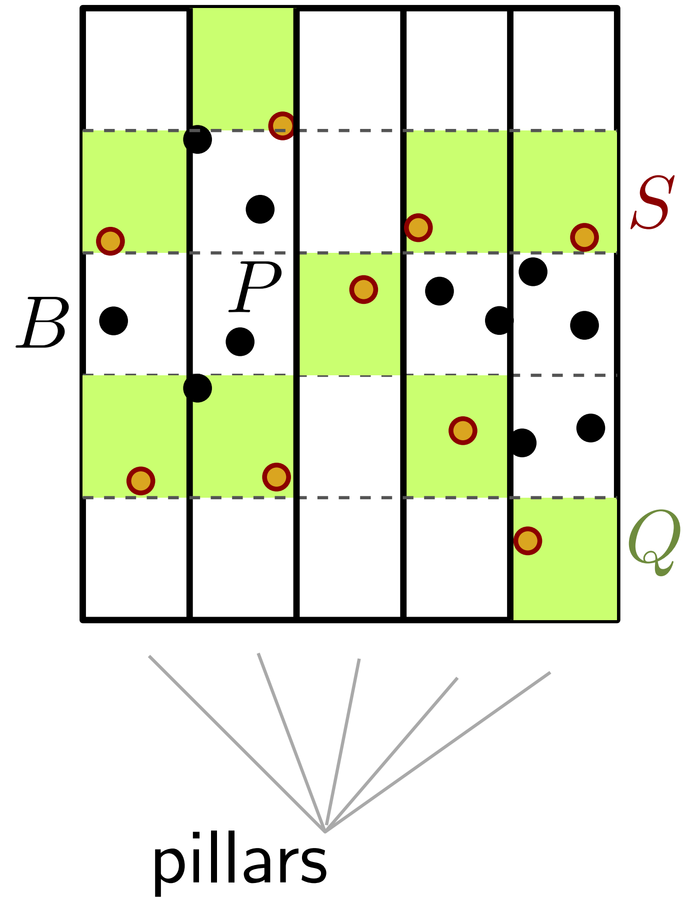
Coreset from pillars

Let Q = union of cells containg a point of S.

$P \subset \mathrm{conv}(Q) \Rightarrow \mathrm{wd}(v, \mathrm{conv}(Q)) \geq \mathrm{wd}(v, P)$ for all v

$\mathrm{wd}(v, S) + 2\mathrm{wd}(v, B/M) \geq \mathrm{wd}(v, Q) \geq \mathrm{wd}(v, P)$

Let $s \in \mathbb{R}^d, c_d > 0$ such that
$s + c_d B \subset \mathrm{conv}(P)$



$$\mathrm{wd}(B/M) = \frac{\mathrm{wd}(B)}{M} = \frac{\mathrm{wd}(c_d B)}{c_d M}$$

$$\leq \frac{\mathrm{wd}(P)}{c_d M} \leq \frac{\mathrm{wd}(P)}{2/\varepsilon} = \frac{\varepsilon}{2}\mathrm{wd}(v, P)$$

Better coreset, other uses

Theorem. Given $\varepsilon > 0$ and $P \subset \mathbb{R}^d$, we can compute an $\varepsilon/2$-coreset $S \subseteq P$ of size at most $|S| = O(1/\varepsilon^{(d-1)/2})$ in $O(n + 1/\varepsilon^{3(d-1)/2})$ time (where d is a fixed constant).

Proof ideas:

- two stages: first the previous algo. for $\varepsilon/2$ gives S, then this (slower) algorithm for $\varepsilon/2$ on S
- make $\mathrm{conv}(P)$ fat via affine transformation into unit hypercube
- find small enclosing ball B (radius $\sqrt{d}$)
- $X :=$ a $c\sqrt{\varepsilon}$-packing in ∂B
- Let $S =$ nearest point to each $x \in X$

Better coreset, other uses

Theorem. Given $\varepsilon > 0$ and $P \subset \mathbb{R}^d$, we can compute an $\varepsilon/2$-coreset $S \subseteq P$ of size at most $|S| = O(1/\varepsilon^{(d-1)/2})$ in $O(n + 1/\varepsilon^{3(d-1)/2})$ time (where d is a fixed constant).

Proof ideas:

- two stages: first the previous algo. for $\varepsilon/2$ gives S, then this (slower) algorithm for $\varepsilon/2$ on S
- make $\mathrm{conv}(P)$ fat via affine transformation into unit hypercube
- find small enclosing ball B (radius $\sqrt{d}$)
- $X :=$ a $c\sqrt{\varepsilon}$-packing in ∂B
- Let $S =$ nearest point to each $x \in X$

Coresetes have been used for

- width, diameter
- clustering (k-means)

and in streaming, dynamic problems, online algorithms, machine learnig, etc.

Planar graphs: r-division (Frederickson 1986)

Balanced separator for a subset

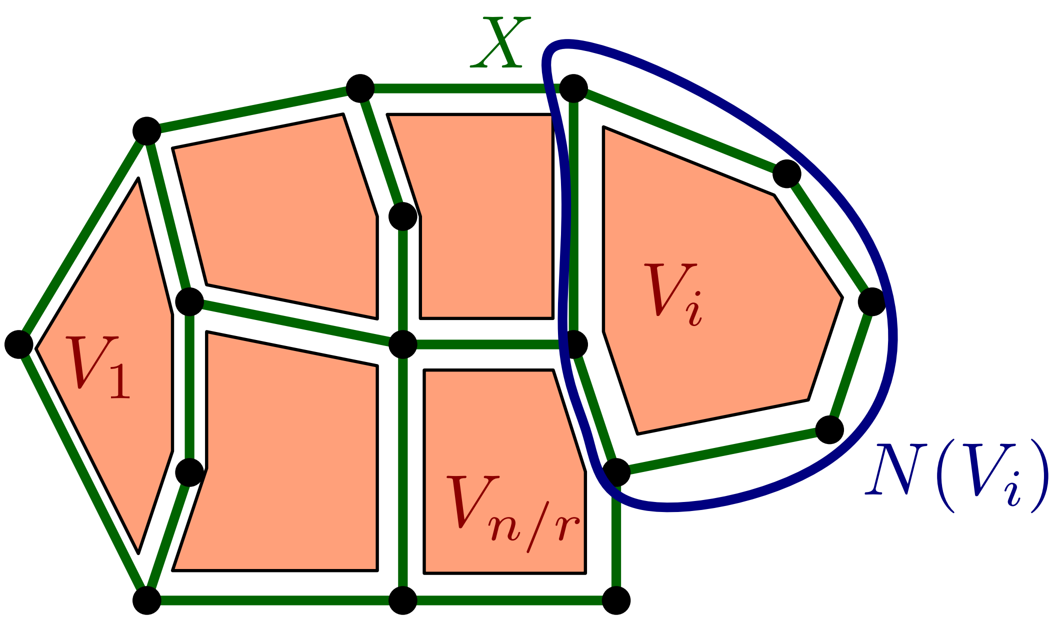
Observation. Given planar graph $G = (V, E)$ and a vertex set $W \subset V$, G has separator of size $O(\sqrt{n})$ s.t. each side has $\leq 36/37|W|$ vertices from W.

Proof.
Start proof with smallest square that encloses $\geq |W|/37$ disks from W.

r-divisions

Theorem There are constants c_1, c_2, c_3 s.t. for any $r \in \mathbb{Z}_+$ and planar graph G, there is a boundary set X of size $\leq c_1 n/\sqrt{r}$ and a partition of $V \setminus X$ into n/r sets $V_1, \ldots, V_{n/r}$ satisfying

- $|V_i| \leq c_2 r$
- $N(V_i) \cap V_j = \emptyset$ if $i \neq j$
- $|N(V_i) \cap X| \leq c_3 \sqrt{r}$



Computing an r-division

Proof sketch. Use planar separator theorem.

Recursively divide until sides have size $\leq c_2 r$
$X :=$ union of separators throughout.
group size ✓ group number ✓ disjointness ✓

Computing an r-division

Proof sketch. Use planar separator theorem.

Recursively divide until sides have size $\leq c_2 r$
$X :=$ union of separators throughout.
group size ✓ group number ✓ disjointness ✓

$$|X| \leq c\sqrt{n} + 2\sqrt{\frac{2}{3}n} + \cdots + 2^i\sqrt{\left(\frac{2}{3}\right)^i n}$$

$$= c\sqrt{n} \sum_{i=0}^{\log_{3/2}(n/r)} (2\sqrt{2/3})^i$$

Computing an r-division

Proof sketch. Use planar separator theorem.

Recursively divide until sides have size $\leq c_2 r$
$X :=$ union of separators throughout.

group size ✓ group number ✓ disjointness ✓

$$|X| \leq c\sqrt{n} + 2\sqrt{\frac{2}{3}n} + \cdots + 2^i\sqrt{\left(\frac{2}{3}\right)^i n}$$

$$= c\sqrt{n} \sum_{i=0}^{\log_{3/2}(n/r)} (2\sqrt{2/3})^i$$

Computing an r-division

Proof sketch. Use planar separator theorem.

Recursively divide until sides have size $\leq c_2 r$
$X :=$ union of separators throughout.
group size ✓ group number ✓ disjointness ✓

Separator divides into size αn and $(1-\alpha)n$ for some $\alpha \in [\frac{1}{3}, \frac{2}{3}]$.

$X(n) \leq c\sqrt{n} + X(\alpha n) + X((1-\alpha)n)$ if $n > r$, else $X(n) = 0$

With substitution and indction, $|X| = X(n) = c_1 n/\sqrt{r}$

Computing an r-division

Proof sketch. Use planar separator theorem.

Recursively divide until sides have size $\leq c_2 r$
$X :=$ union of separators throughout.
group size ✓ group number ✓ disjointness ✓

Separator divides into size αn and $(1-\alpha)n$ for some $\alpha \in [\frac{1}{3}, \frac{2}{3}]$.

$X(n) \leq c\sqrt{n} + X(\alpha n) + X((1-\alpha)n)$ if $n > r$, else $X(n) = 0$

With substitution and indction, $|X| = X(n) = c_1 n/\sqrt{r}$

Still need small boundaries for each group!

Idea: while $N(V_i) \cap X > c_3\sqrt{r}$, separate $N(V_i)$ with balance wrt. $N(V_i) \cap X$.
Gives $O(n/\sqrt{r})$ new boundary vertices and $O(n/r)$ new groups