- **Coresets** — a powerful tool for efficiently approximating various extent measures of a point set P.

- **Extent measures**: An extent measure of P either computes certain statistics of P itself or of a geometric shape (e.g., sphere, box, cylinder, etc.) enclosing P.
  - → Diameter or kth largest distance between pairs of points in P.
  - → compute smallest radius of a sphere, min volume of a box, smallest width of a slab that contains P.

- Smallest volume bounding box/ball containing P in $\mathbb{R}^3$.
  - → Expensive in general: $O(n^3)$ time.
  - → Can we find $(1+\epsilon)$-approximation in $O(n f(\varepsilon))$ time?

$$\begin{tikzcd}
P \arrow[r, "\text{coresets}"'] \arrow[d, dashed, "\text{improved algorithm}"'] & Q \ \leftarrow \text{size } 1/\varepsilon^{O(1)} \arrow[d, "\text{inefficient algorithm}"] \\
\mu(P) & \mu(Q) \geq (1-\varepsilon)\mu(P) \arrow[l]
\end{tikzcd}$$

§ Kernels for Point Sets:

- μ is a measure function (e.g., the width of a point set)

$$\mu : \mathbb{R}^d \longrightarrow \mathbb{R}^+ \cup \{0\}$$

and μ is monotone $[P_1 \subseteq P_2 \Rightarrow \mu(P_1) \leq \mu(P_2)]$.

Given a parameter $\varepsilon > 0$, $Q \subseteq P$ is an ε-coreset of P (w.r.t. μ) if $(1-\varepsilon)\mu(P) \leq \mu(Q)$.

• Problem with ϵ-nets & ϵ-samples:

Approach via VC-dimension:

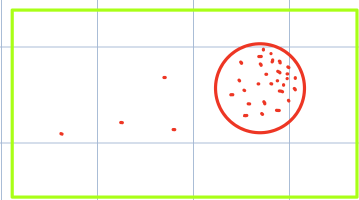
Define $\mathcal{R}$ to be set ranges that are complement of balls.

One can show VC-dim $(X, \mathcal{R}) = O(1)$.

So one can obtain ε-net R of size $O(\frac{1}{\varepsilon} \log \frac{1}{\varepsilon})$.

If an enclosing ball g for R contains $\geq \varepsilon |P|$ points of P outside it.

Then $\bar{g}$ contains $\geq \varepsilon |P|$ points of P inside it.



$\Rightarrow$ $\bar{g}$ must contain a point of R (an ϵ-net).

$\Rightarrow$ g does not enclose all points of R.

So any ball that encloses R covers all but an ε-fraction of points of P. ✓ ☺

However, remaining points can be quite far, thus this will not "approximate" the minimum enclosing ball. ✗ ☹

→ coresets capture the whole structure of the input, random samples capture the structure for "most" of the points.

Coresets trade off geometric error (i.e., an ε-fraction of the radius of the enclosing ball) while random samples trade off statistical error (hold for all except for an ε-fraction of the input).

- Agarwal, Har-Peled, Varadarajan introduced the notion of ε-kernel and showed it is $f(\varepsilon)$-coreset for numerous minimization problems.

- ε-kernel:

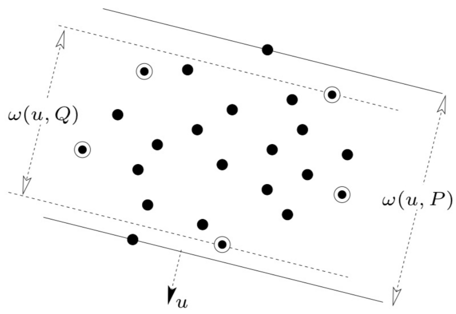


Figure 1. Directional width and ε-kernel.

ε-kernel. Let $\mathbb{S}^{d-1}$ denote the unit sphere centered at the origin in $\mathbb{R}^d$. For any set P of points in $\mathbb{R}^d$ and any direction $u \in \mathbb{S}^{d-1}$, we define the *directional width* of P in direction u, denoted by $\omega(u, P)$, to be

$$\omega(u, P) = \max_{p \in P} \langle u, p \rangle - \min_{p \in P} \langle u, p \rangle ,$$

where $\langle \cdot, \cdot \rangle$ is the standard inner product. Let $\varepsilon > 0$ be a parameter. A subset $Q \subseteq P$ is called an ε-*kernel* of P if for each $u \in \mathbb{S}^{d-1}$,

$$(1 - \varepsilon)\omega(u, P) \leq \omega(u, Q).$$

- A measure function μ is faithful if $\exists$ constant c s.t. $\forall\, P \subseteq \mathbb{R}^d$ and any constant $\varepsilon > 0$, an ε-kernel of P is a $c\varepsilon$-coreset for P w.r.t. μ.

$\rightarrow$ E.g. diameter, width, radius of the smallest enclosing ball, volume of the smallest enclosing box.

$\rightarrow$ A common property of these μ: $\mu(P) = \mu(\mathrm{conv}(P))$.

- α-fat point set P: For $\alpha \leq 1$, if $\exists$ point $p \in \mathbb{R}^d$ and a hypercube $\mathcal{C}$ centered at the origin s.t.

$$p + \alpha\mathcal{C} \subseteq \mathrm{conv}(P) \subseteq p + \mathcal{C}.$$

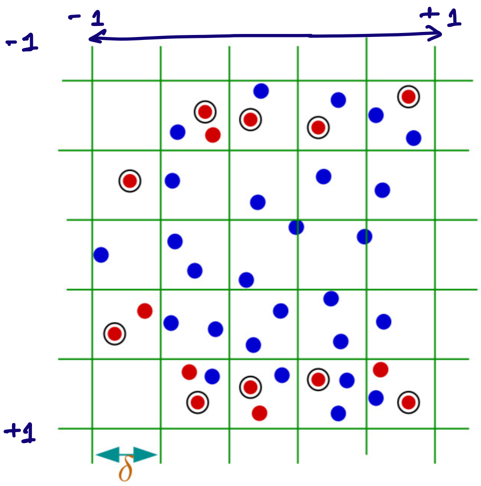
• Algorithms for computing Kernels :

can be thought of as a good appx of conv hull

Lemma [AHV] : Let $P \subseteq \mathbb{R}^d$, $|P| = n$ s.t. vol (conv(P)) > 0, and let $\mathcal{C} = [-1, 1]^d$. One can compute in $O(n)$ time an affine transform J s.t. $J(P)$ is an α-fat point set (α depends on d) and s.t. $Q \subseteq P$ is an ε-kernel of P iff $J(Q)$ is an ε-kernel of $J(P)$.

(via approximation to Löwner-John Ellipsoid).

$\Rightarrow P \subseteq [-1, +1]^d$ that is α-fat.



(i)

① Grid-based Algorithms :

Take largest $\delta \leq \left(\frac{\varepsilon}{\sqrt{d}}\right)\alpha$, $1/\delta \in \mathbb{Z}$.

d-dim grid of size δ.

For each column, choose one point (arbitrary) from highest nonempty cell of the column and one point from lowest nonempty cell.

Need to maintain 2 points in each of $1/\delta$ cols in $[-1, 1]^2$.
in general, 2 points in each of $(1/\delta)^{d-1}$ cols in $[-1, 1]^d$.

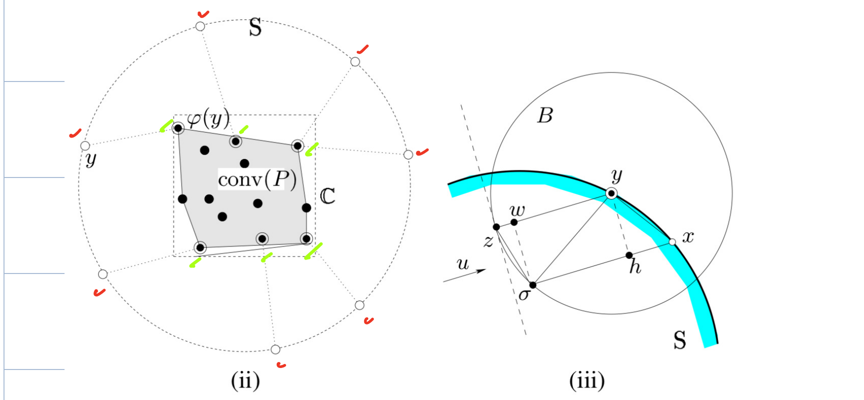
$|Q| = O\left(\frac{1}{(\alpha\varepsilon)^{d-1}}\right)$ & can be computed in $O\left(n + \frac{1}{(\alpha\varepsilon)^{d-1}}\right)$ time.

---

② NN-based Algorithms :

S = sphere of radius of $(\sqrt{d} + 1)$ centered at origin.

$\delta = \sqrt{\varepsilon\alpha} \leq 1/2$



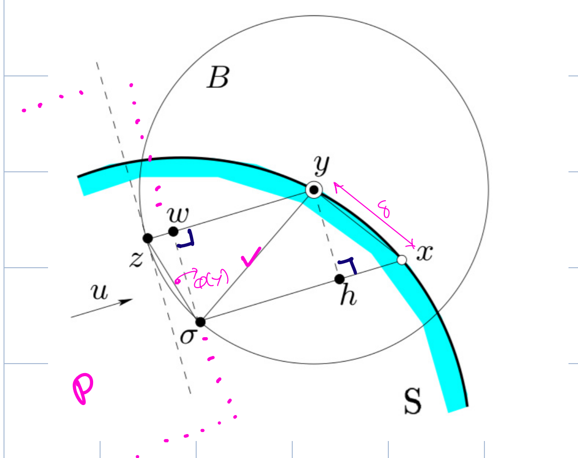
Construct a set $\mathcal{J}$ of $O(1/\delta^{d-1}) = O(1/\varepsilon^{(d-1)/2})$ points on S, s.t. $\forall x \in S, \exists y \in \mathcal{J}$ s.t. $\|x-y\| \le \delta$.

Process P into data structure that can answer ε-appx nearest neighbor queries. [Arya et al.]

For each point $y \in \mathcal{J}$, compute its (ϵ-appx) NN in P $\to \varphi(y)$

Return $Q = \{\varphi(y) \mid y \in \mathcal{J}\}$.

- Intuition on why Q is an ϵ-kernel of P.



For simplicity assume $\varphi(y) =$ exact NN.

Consider a direction u.

Let $\sigma \in P$ be the point that max $\langle u, p\rangle$ over all $p \in P$.

Ray emanating from σ hits S at x.

Let $y \in \mathcal{J}$ s.t. $\|x-y\| \le \delta$.

If $\varphi(y) = \sigma$, then $\sigma \in Q$ and

$$\max_{p\in P} \langle u,p\rangle - \max_{q \in Q} \langle u,q\rangle = 0$$

Otherwise, $\varphi(y) \ne \sigma$. Let B be d-dim ball of radius $\|y-\sigma\|$ centered at y.

Let $z \in \partial B$ that is hit by ray from y in dir $-u$.
w, h are projections (see fig.)

Since, $\phi(y)$ is closer to σ, $\phi(y)$ lies inside B.

$\langle u, \phi(y)\rangle \geq \langle u, z\rangle$

further, $\langle u, \sigma\rangle - \langle u, \phi(y)\rangle \leq \alpha\varepsilon$. [HW: by geometry]

Hence, $\max_{p\in P}\langle u, p\rangle - \max_{q\in Q}\langle u, q\rangle \leq \langle u, \sigma\rangle - \langle u, \phi(y)\rangle \leq \alpha\varepsilon$

Similarly, $\min_{p\in P}\langle u, p\rangle - \min_{q\in Q}\langle u, q\rangle \geq -\alpha\varepsilon$.

Hence, $w(u, Q) \geq w(u, P) - 2\alpha\varepsilon$.

Since $\alpha C \subset conv(P)$, $w(u, P) \geq 2\alpha$

Hence, $w(u, Q) \geq (1-\varepsilon)\, w(u, P)$ for any direction u.

→ Chan showed that $\phi(y)$ for all $y \in J$ can be computed in a total time $O(n + 1/\varepsilon^{d-1})$ time [using appx Voronoi diagrams].

• Theorem [Chan]: Given a set of n points in $\mathbb{R}^d$ and parameter $\varepsilon > 0$, one can compute an ε-kernel of P of size $O(1/\varepsilon^{(d-1)/2})$ in time $O(n + 1/\varepsilon^{d-\frac{3}{2}})$.

→ So if a faithful measure μ can be computed in $O(n^\alpha)$-time.

Then, by above thm compute an (ε/c)-kernel Q and then use $O(n^\alpha)$-algo. $\Rightarrow O\left(n + \frac{1}{\varepsilon^{d-3/2}} + \frac{1}{\varepsilon^{\alpha(d-1)/2}}\right)$ time algo.

• Exact diameter computation takes $O(n^2)$ time.
Above approach computes it in near-linear time.

⊙ Coresets for Clustering:

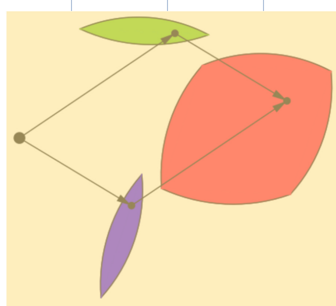
Clustering: Given a set P of n points in $\mathbb{R}^d$. $k \in \mathbb{Z}_+$.
Partition P into k subsets (clusters) $P_1, P_2, \ldots, P_k$ s.t. certain objective is minimized.

- centered clustering: obj: $\max_{1 \le i \le k} \mu(P_i)$ → k-center, → k-line-center

summed clustering: obj: $\sum_{i=1}^{k} \mu(P_i)$ → k-median, → k-means

• Generalized k-center: cluster: (f, S). f: q-dim subspace for $q \le d$; $S \subseteq P$

$$\mu(f, S) = \max_{p \in S} d(p, f).$$



The red figure is the Minkowski sum of blue and green figures.

$A \oplus B = \{a \oplus b : a \in A, b \in B\}$

Define $B(f, r)$ to be $f \oplus B(0, r)$ — Ball of radius r centered at O

→ ball of radius r if f = point $(q=0)$
→ cylinder of radius r if f = line $(q=1)$
→ slab of width $2r$ if f = hyperplane. $(q=2)$

• Define $\mathcal{C} = \{(f_1, P_1), \ldots, (f_k, P_k)\}$ a k-clustering (of dim q) if each f_i is a q-dim subspace and $P = \bigcup_{i=1}^{k} P_i$.

$\mu(\mathcal{C}) = \max_{1 \le i \le k} \mu(f_i, P_i)$, $r_{opt}(P, k, q) = \min_{\mathcal{C}} \mu(\mathcal{C})$.

Let $\mathcal{C}_{opt}(P, k, q)$ be the opt clustering

- $q = 0$ k-center : covering P by k balls
- $ = 1$ k-line-center : " k cylinders
- $ = d-1$ k-hyperplane-center : " k slabs

} of min radius

Greedy 2-appx for k-center :

first iteration

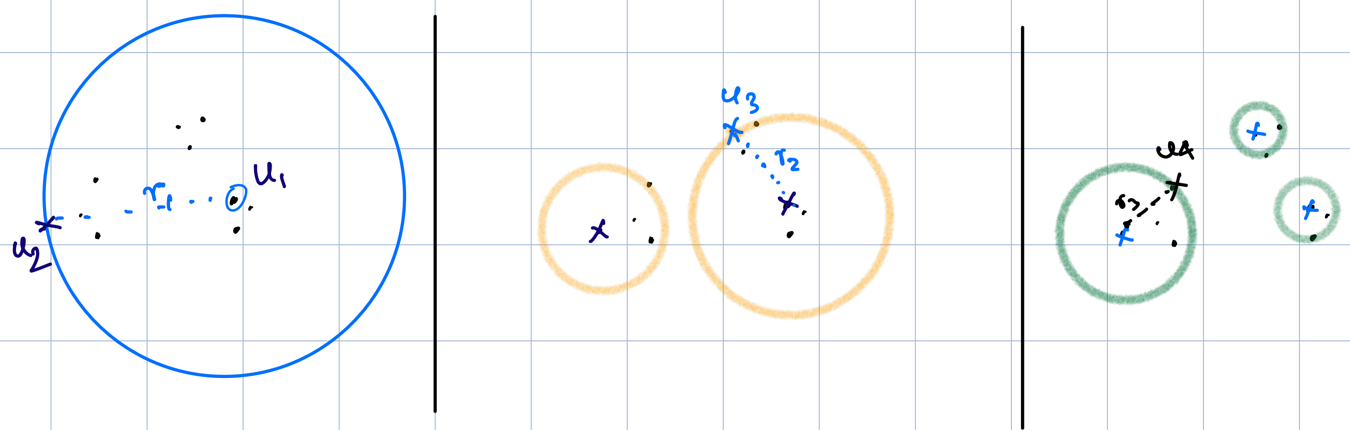
- Pick an arbitrary point u_1 & make it a center : $U_1 = \{u_1\}$
- for next $(k-1)$ iterations :
  - pick the point furthest away from current set of centers
  - add this point to the set of centers.

[Let U_i be the set of centers at the end of iteration i.

$r_{i-1} = \max_{p \in P} d(p, U_{i-1})$ & u_i be such selected point p

$U_i = U_{i-1} \cup \{u_i\}$]

We return U_k.



Proof : ① $r_1 \geq r_2 \geq \cdots \geq r_k$.

② The distance between any pair of centers $\geq r_k$.

For contradiction, assume $r_k > 2 \cdot r^*$ $\longrightarrow$ OPT radius

Consider OPT that covers P with k balls of radius r^*.

By Δ-inequality, any two points in a ball are at dist $\leq 2r^*$.

Thus none of these balls can cover two points of U_{k+1}.

Hence, all points in U_{k+1} can be covered by k balls

as $|U_{k+1}| = k+1$. $\longrightarrow$ Contradiction.

- Additive & Multiplicative coresets:

$Q \subseteq P$ is an ε-coreset of P (for measure μ) if for every k-clustering $\mathcal{C} = \{(f_1, Q_1), \ldots, (f_k, Q_k)\}$ of Q with $r_i = \mu(f_i; Q_i)$,

$P \subseteq \bigcup_{i=1}^{k} \mathcal{B}(f_i, r_i + \varepsilon\mu(\mathcal{C}))$ → additive coresets

or $P \subseteq \bigcup_{i=1}^{k} \mathcal{B}(f_i, (1+\varepsilon) r_i)$ → multiplicative coresets.

small expansion of balls for Q cover P.

- Additive coreset:

$r^* = r_{opt}(P, k, 0)$, $\mathcal{B} = \{B_1, \ldots, B_k\}$. family of k balls of radius r^* that cover P.

→ Draw d-dim grid with side length $\varepsilon r^*/(2d)$.

→ $O(k/\varepsilon^d)$ of these gridcells intersect the balls in $\mathcal{B}$.

→ for each such cell $\square$ with $P \cap \square \neq \emptyset$, select a point $p \in P \cap \square$ arbitrarily.

→ One can show that these $O(k/\varepsilon^d)$ points form an additive coreset.

- To estimate r^*, one can run greedy & get $\tilde{r} \in [r^*, 2r^*]$ & draw grids with side length $\varepsilon \tilde{r}/(4d)$.

More complicated constructions give better coresets. ↓

⊙ Thm: Let P be a set of n points in $\mathbb{R}^d$, $0 < \varepsilon < 1/2$. $\exists$ a mult. ε-coresets of size $O(k!/\varepsilon^{dk})$ of P for k-center.

- See Feldman's survey for other constructions/applications:
  - uniform sampling [Braverman et al., FOCS'22] + importance sampling
  - Grids
  - Greedy construction [Feldman-Rus, NeurIPS'16].