

Near-Optimal Dimension Reduction for Facility Location

Lingxiao Huang¹ Shaofeng H.-C. Jiang² Robert Krauthgamer³ **Di Yue**²

August 29, 2024

¹Nanjing University

²Peking University

³Weizmann Institute of Science

Dimension Reduction

Theorem (Johnson-Lindenstrauss lemma [\[Johnson-Lindenstrauss 84\]](#))

For all $n > 0$, $\varepsilon \in (0, 1)$, there exists a random linear map $\pi: \mathbb{R}^d \rightarrow \mathbb{R}^m$, for $m = O(\varepsilon^{-2} \log n)$, such that for every $X \subset \mathbb{R}^d$ with $|X| = n$, with high probability

$$\forall x, y \in X, \|\pi(x) - \pi(y)\| \in (1 \pm \varepsilon) \|x - y\|.$$

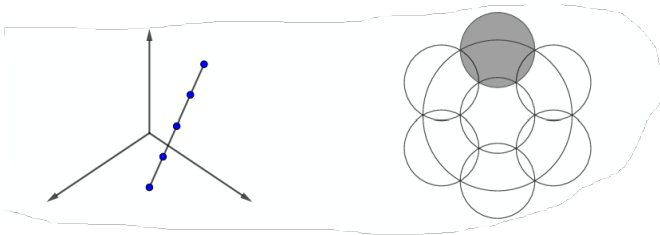
- JL mapping: $\pi: x \mapsto \frac{1}{\sqrt{m}} Gx$, where $G \in \mathbb{R}^{m \times d}$ with $g_{i,j} \sim i.i.d. N(0, 1)$.
- $m = O(\varepsilon^{-2} \log n)$ is tight [\[Larsen, Nelson, FOCS 2017\]](#).
- Curse of dimensionality: target dimension $m = \Theta(\log n)$ is always too high to afford.
 - TSP does not admit a PTAS in dimension $m = \Theta(\log n)$ [\[Trevisan, SIAM J. Comput. 00\]](#).
 - Many problems have 2^{2^d} dependence in the (low-dimensional) PTAS.

Doubling Dimension

[Gupta, Krauthgamer, Lee, FOCS 03]

Solution: Use *intrinsic dimension*.

- $\text{ddim}(X) :=$ the minimum $t \geq 0$, such that $\forall r > 0$, each ball in X of radius r can be covered by at most 2^t balls of radius $r/2$.



- Fundamental question: refine JL lemma such that m only depends on $\text{ddim}(X)$.
- Preserve objective value for specific computational problems instead of pairwise distances. (i.e. $\text{opt}(\pi(X)) \approx \text{opt}(X)$)

Related Work on JL

- Preserve objective value for specific computational problems instead of pairwise distances. (i.e. $\text{opt}(\pi(X)) \approx \text{opt}(X)$)

Problems	Approximation	Target Dimension	References
Nearest Neighbor	$1 + \varepsilon$	$O(\varepsilon^{-2} \text{ddim})$	IN07
k -Center Clustering	$1 + \varepsilon$	$O(\varepsilon^{-2}(\log k + \text{ddim}))$	JKS24
k -Median / k -Means	$1 + \varepsilon$	$O(\varepsilon^{-2} \log k)$	MMR19
Max-Cut	$1 + \varepsilon$	$O(1/\varepsilon^2)$	CJK23
MST	$1 + \varepsilon$	$O(\varepsilon^{-2} \text{ddim} \log \log n)$	NSIZ21
UFL	$O(1)$	$O(\varepsilon^{-2} \text{ddim})$	NSIZ21
UFL	$1 + \varepsilon$	$\tilde{O}(\varepsilon^{-2} \text{ddim})$	This work

Uniform Facility Location (UFL)

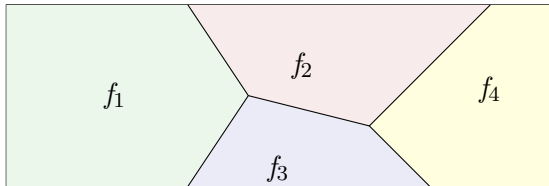
Input: Point set $X \subset \mathbb{R}^d$ with bounded doubling dimension ddim , opening cost $f > 0$.

Goal: Find a set of facilities $F \subset \mathbb{R}^d$, so as to minimize the objective

$$\text{cost}(X, F) := \underbrace{f \cdot |F|}_{\text{opening cost}} + \underbrace{\sum_{x \in X} \text{dist}(x, F)}_{\text{connection cost}},$$

where $\text{dist}(x, F) := \min_{y \in F} \text{dist}(x, y)$ and $\text{dist}(x, y) := \|x - y\|_2$.

- W.l.o.g., $f \equiv 1$ (by scaling the point set).
- Denote the optimal value by $\text{ufl}(X)$.



Problem (Dimension reduction for UFL)

Given $\varepsilon, \delta > 0$, find target dimension $m = f(\text{ddim}, \varepsilon, \delta)$, such that $\Pr[\text{ufl}(\pi(X)) \in (1 \pm \varepsilon) \text{ufl}(X)] \geq 1 - \delta$, where $\pi: \mathbb{R}^d \rightarrow \mathbb{R}^m$ is the JL mapping.

Results

Theorem (Dimension reduction for UFL)

$\Pr[\text{ufl}(\pi(X)) \in (1 \pm \varepsilon) \text{ufl}(X)] \geq 1 - \delta$, for *target dimension*
 $m := O(\varepsilon^{-2} \text{ddim} \cdot \log(\delta^{-1} \varepsilon^{-1} \text{ddim}))$.

Results

Theorem (Dimension reduction for UFL)

$\Pr[\text{ufl}(\pi(X)) \in (1 \pm \varepsilon) \text{ufl}(X)] \geq 1 - \delta$, for target dimension $m := O(\varepsilon^{-2} \text{ddim} \cdot \log(\delta^{-1} \varepsilon^{-1} \text{ddim}))$.

- Previous results: $O(1)$ -approximation with $m = O(\varepsilon^{-2} \text{ddim})$ [Narayanan, Silwal, Indyk, Zamir, ICML 21] or $(1 + \varepsilon)$ -approximation with $m = O(\varepsilon^{-2} \log n)$ [Makarychev, Makarychev, Razenshteyn, STOC 19].

Results

Theorem (Dimension reduction for UFL)

$\Pr[\text{ufl}(\pi(X)) \in (1 \pm \varepsilon) \text{ufl}(X)] \geq 1 - \delta$, for target dimension $m := O(\varepsilon^{-2} \text{ddim} \cdot \log(\delta^{-1} \varepsilon^{-1} \text{ddim}))$.

- Previous results: $O(1)$ -approximation with $m = O(\varepsilon^{-2} \text{ddim})$ [Narayanan, Silwal, Indyk, Zamir, ICML 21] or $(1 + \varepsilon)$ -approximation with $m = O(\varepsilon^{-2} \log n)$ [Makarychev, Makarychev, Razenshteyn, STOC 19].

Theorem (PTAS for UFL)

There is an algorithm that computes a $(1 + \varepsilon)$ -approximate solution for UFL, running in time $(2^{m'} d + 2^{2^{m'}}) \cdot \tilde{O}(n)$, for $m' = O(\text{ddim} \cdot \log(\text{ddim}/\varepsilon))$.

Results

Theorem (Dimension reduction for UFL)

$\Pr[\text{ufl}(\pi(X)) \in (1 \pm \varepsilon) \text{ufl}(X)] \geq 1 - \delta$, for target dimension $m := O(\varepsilon^{-2} \text{ddim} \cdot \log(\delta^{-1} \varepsilon^{-1} \text{ddim}))$.

- Previous results: $O(1)$ -approximation with $m = O(\varepsilon^{-2} \text{ddim})$ [Narayanan, Silwal, Indyk, Zamir, ICML 21] or $(1 + \varepsilon)$ -approximation with $m = O(\varepsilon^{-2} \log n)$ [Makarychev, Makarychev, Razenshteyn, STOC 19].

Theorem (PTAS for UFL)

There is an algorithm that computes a $(1 + \varepsilon)$ -approximate solution for UFL, running in time $(2^{m'} d + 2^{2^{m'}}) \cdot \tilde{O}(n)$, for $m' = O(\text{ddim} \cdot \log(\text{ddim}/\varepsilon))$.

- Facilities are allowed to be picked from the (d -dimensional) ambient space.
- Previous result: Running in time $2^{2^{O(\text{ddim}^2)}} d \cdot \tilde{O}(n)$, with facilities restricted to the same doubling metric [Cohen-Addad, Feldmann, Saulpic, JACM 21].

Proof Overview

- Upper bound $\text{ufl}(\pi(X))$: $\text{ufl}(\pi(X)) \leq \text{cost}(\pi(X), \pi(F^*)) \lesssim \text{cost}(X, F^*) = \text{ufl}(X)$.

Proposition ([Johnson, Lindenstrauss 84] [Makarychev, Makarychev, Razenshteyn, STOC 19])

$\forall x, y \in \mathbb{R}^d$ and $t > 0$,

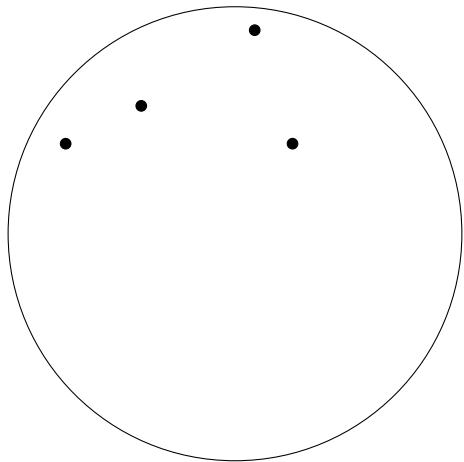
$$\mathbb{E} \left[\max \left\{ 0, \frac{\|\pi(x) - \pi(y)\|}{\|x - y\|} - (1 + t) \right\} \right] \leq \frac{1}{mt} e^{-t^2 m/2}.$$

Proof Overview

- Upper bound $\text{ufl}(\pi(X))$: $\text{ufl}(\pi(X)) \leq \text{cost}(\pi(X), \pi(F^*)) \lesssim \text{cost}(X, F^*) = \text{ufl}(X)$.
- Lower bound $\text{ufl}(\pi(X))$
 - (1) Partition X into “light” clusters $\Lambda = \{C_1, C_2, \dots, C_{|\Lambda|}\}$, where $\text{ufl}(C_i) = \Theta(\text{ddim}/\varepsilon)^{O(\text{ddim})}$.
 - (2) Sub-additivity: $\text{ufl}(X) \leq \sum_{C \in \Lambda} \text{ufl}(C)$.
 - (3) On each cluster $C \in \Lambda$, $\text{ufl}(C) \lesssim \text{ufl}(\pi(C))$ in expectation.
 - (4) $\sum_{C \in \Lambda} \text{ufl}(\pi(C)) \lesssim \text{ufl}(\pi(X))$.

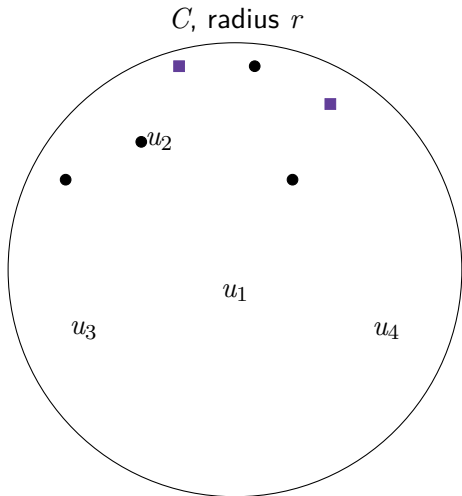
Step 1: Hierarchical Decomposition

C , radius r



Recursively partition a large cluster into small sub-clusters of half radius.

Step 1: Hierarchical Decomposition



Recursively partition a large cluster into small sub-clusters of half radius.

- Construct $(r/2)$ -net N on C .

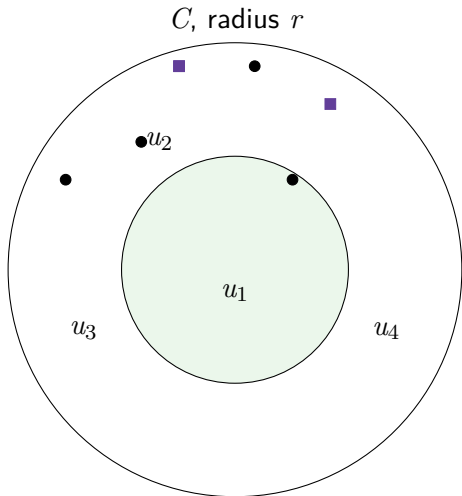
Definition (ρ -net)

- ρ -packing: $\forall x, y \in N, \text{dist}(x, y) \geq \rho$.
- ρ -covering: $\forall x \in C, \exists y \in N, \text{s.t. } \text{dist}(x, y) \leq \rho$.
- N is a ρ -net if it is both ρ -packing and ρ -covering for C .

Proposition (Packing property)

If S is ρ -packing then $|S| \leq (2 \text{diam}(S)/\rho)^{\text{ddim}(S)}$.

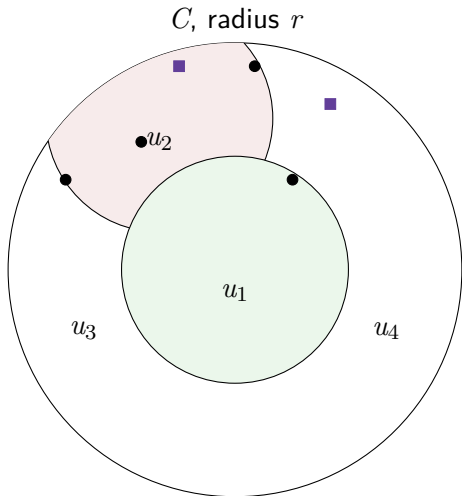
Step 1: Hierarchical Decomposition



Recursively partition a large cluster into small sub-clusters of half radius.

- Construct a level i child cluster $C_u \leftarrow C \cap B(u, r/2) \setminus \bigcup_{v \in N: \sigma(v) < \sigma(u)} B(v, r/2)$.

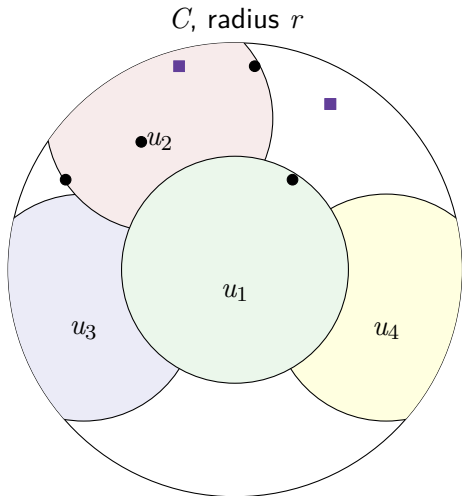
Step 1: Hierarchical Decomposition



Recursively partition a large cluster into small sub-clusters of half radius.

- Construct a level i child cluster $C_u \leftarrow C \cap B(u, r/2) \setminus \bigcup_{v \in N: \sigma(v) < \sigma(u)} B(v, r/2)$.

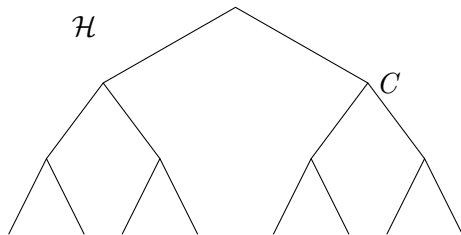
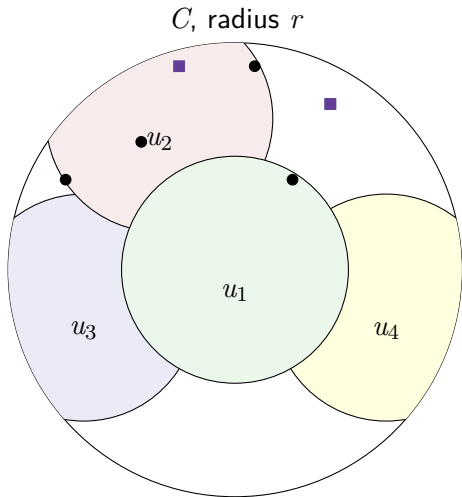
Step 1: Hierarchical Decomposition



Recursively partition a large cluster into small sub-clusters of half radius.

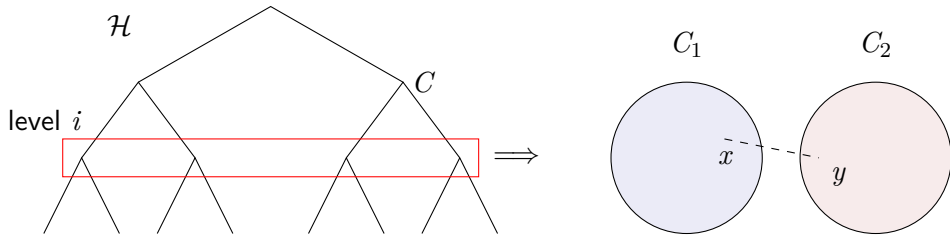
- Construct a level i child cluster $C_u \leftarrow C \cap B(u, r/2) \setminus \bigcup_{v \in N: \sigma(v) < \sigma(u)} B(v, r/2)$.

Step 1: Hierarchical Decomposition



- Node \leftrightarrow cluster.
 - Root: X ;
 - leaves: singletons;
 - level i : diameter $\Theta(2^i)$.
- Each node (cluster) has $2^{O(\text{ddim})}$ child nodes (clusters).

Cut



- Cut [Talwar, STOC 04] [Cohen-Addad, Feldmann, Saulpic, JACM 21]: a pair of points (x, y) is cut by cluster C if $x \in C$ and $y \notin C$.
- Badly cut: (x, y) is *badly cut* if (x, y) is cut by some cluster C with $\text{diam}(C) \geq \frac{\text{ddim}}{\varepsilon^2} \cdot \|x - y\|$.
- There is a (random) hierarchical decomposition, such that $\forall x, y \in X$, $\Pr[(x, y) \text{ is badly cut}] \leq O(\varepsilon^2)$.

Step 2: Eliminating Badly-Cut Pairs

Fix facility set $F \subseteq X$.

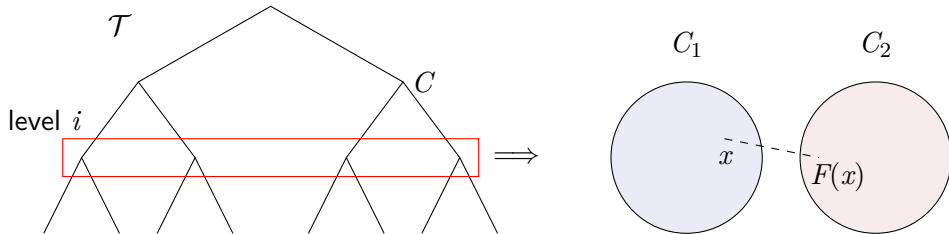
Goal: $\forall x \in X$, if $(x, F(x))$ is cut by cluster C , then x and $F(x)$ are separated ($\|x - F(x)\| \gtrsim \text{diam}(C)$).

Step 2: Eliminating Badly-Cut Pairs

Fix facility set $F \subseteq X$.

Goal: $\forall x \in X$, if $(x, F(x))$ is cut by cluster C , then x and $F(x)$ are separated ($\|x - F(x)\| \gtrsim \text{diam}(C)$).

- For each level i :
 - For each $x \in X$, if $x \in C$, $F(x) \notin C$ and $\text{diam}(C) \geq \frac{\text{ddim}}{\epsilon^2} \cdot \|x - F(x)\|$, then “move” x into the same level i cluster as $F(x)$.

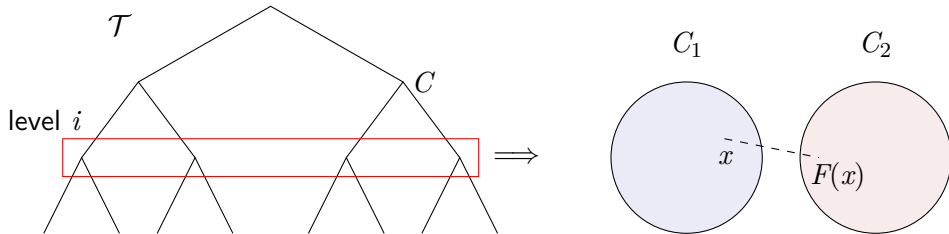


Step 2: Eliminating Badly-Cut Pairs

Fix facility set $F \subseteq X$.

Goal: $\forall x \in X$, if $(x, F(x))$ is cut by cluster C , then x and $F(x)$ are separated ($\|x - F(x)\| \gtrsim \text{diam}(C)$).

- For each level i :
 - For each $x \in X$, if $x \in C$, $F(x) \notin C$ and $\text{diam}(C) \geq \frac{\text{ddim}}{\epsilon^2} \cdot \|x - F(x)\|$, then “move” x into the same level i cluster as $F(x)$.

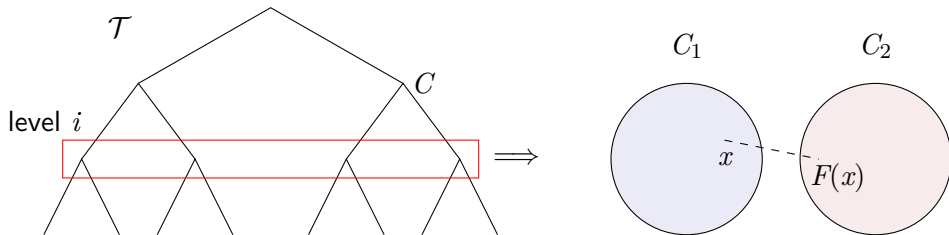


Step 2: Eliminating Badly-Cut Pairs

Fix facility set $F \subseteq X$.

Goal: $\forall x \in X$, if $(x, F(x))$ is cut by cluster C , then x and $F(x)$ are separated ($\|x - F(x)\| \gtrsim \text{diam}(C)$).

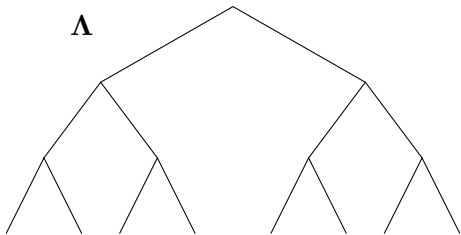
- For each level i :
 - For each $x \in X$, if $x \in C$, $F(x) \notin C$ and $\text{diam}(C) \geq \frac{d \dim}{\epsilon^2} \cdot \|x - F(x)\|$, then “move” x into the same level i cluster as $F(x)$.



- On the new decomposition \mathcal{T} , $(x, F(x))$ is not badly-cut.
 - Separation: If $(x, F(x))$ is cut by cluster C , then $\|x - F(x)\| \geq \frac{\epsilon^2}{d \dim} \text{diam}(C)$.

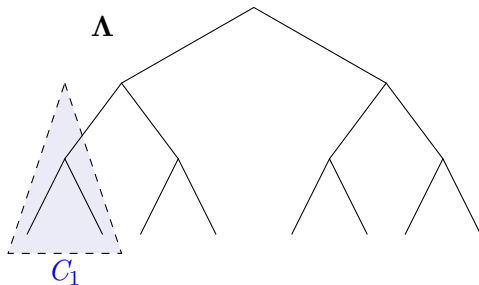
Step 3: Partition

Given threshold $\kappa = \Theta(\text{ddim}/\varepsilon)^{O(\text{ddim})}$, find the lowest level “heavy cluster” ($\text{ufl}(C) \geq \kappa$) in a bottom-up manner.



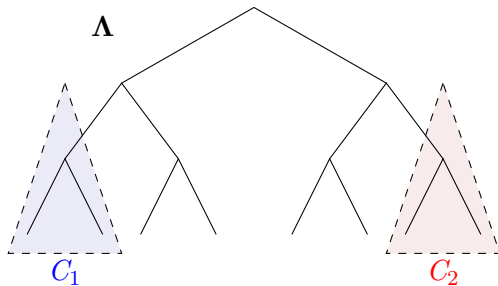
Step 3: Partition

Given threshold $\kappa = \Theta(\text{ddim}/\varepsilon)^{O(\text{ddim})}$, find the lowest level “heavy cluster” ($\text{ufl}(C) \geq \kappa$) in a bottom-up manner.



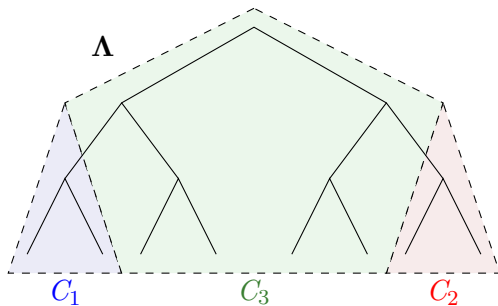
Step 3: Partition

Given threshold $\kappa = \Theta(\text{ddim}/\varepsilon)^{O(\text{ddim})}$, find the lowest level “heavy cluster” ($\text{ufl}(C) \geq \kappa$) in a bottom-up manner.



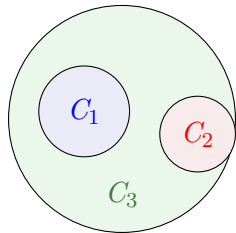
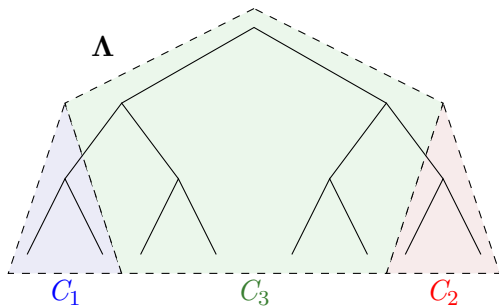
Step 3: Partition

Given threshold $\kappa = \Theta(\text{ddim}/\varepsilon)^{O(\text{ddim})}$, find the lowest level “heavy cluster” ($\text{ufl}(C) \geq \kappa$) in a bottom-up manner.



Step 3: Partition

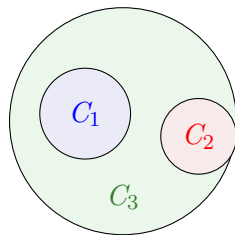
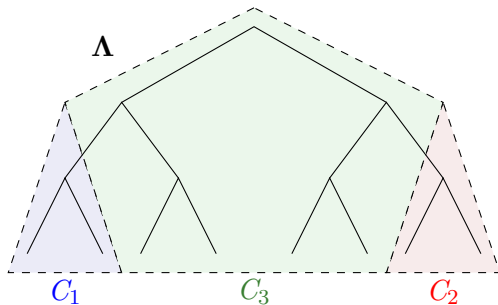
Given threshold $\kappa = \Theta(\text{ddim}/\varepsilon)^{O(\text{ddim})}$, find the lowest level “heavy cluster” ($\text{ufl}(C) \geq \kappa$) in a bottom-up manner.



$$\Lambda = \{C_1, C_2, C_3\}$$

Step 3: Partition

Given threshold $\kappa = \Theta(\text{ddim}/\varepsilon)^{O(\text{ddim})}$, find the lowest level “heavy cluster” ($\text{ufl}(C) \geq \kappa$) in a bottom-up manner.



$$\Lambda = \{C_1, C_2, C_3\}$$

- Bounded optimal value: $\forall C \in \Lambda, \kappa \leq \text{ufl}(C) \leq 2^{O(\text{ddim})}\kappa$.

JL on Each Cluster

- Bounded optimal value: $\forall C \in \Lambda, \kappa \leq \text{uf}(C) \leq 2^{O(\text{ddim})} \kappa$.
- Equivalent to τ -median on C , where $\tau = 2^{O(\text{ddim})} \kappa = \Theta(\text{ddim}/\varepsilon)^{O(\text{ddim})}$.

Lemma (Refined from [\[Makarychev, Makarychev, Razenshteyn, STOC 19\]](#))

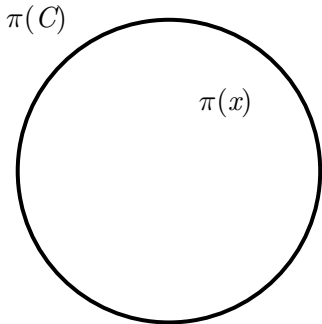
Assume $\text{uf}(C) \leq \tau$. Then for $m = \Omega(\varepsilon^{-2} \log(1/\varepsilon))$,

$$\Pr \left[\text{uf}(\pi(C)) \leq \frac{1}{1+\varepsilon} \text{uf}(C) \right] \leq \tau^3 \cdot e^{-\Omega(\varepsilon^2 m)}.$$

- Choose target dimension $m = O(\varepsilon^{-2} \log \tau) = O(\varepsilon^{-2} \text{ddim} \cdot \log(\text{ddim}/\varepsilon))$.
- On each cluster $C \in \Lambda$, $\text{uf}(C) \lesssim \text{uf}(\pi(C))$ in expectation.

Lower Bound $\text{ufl}(\pi(X))$

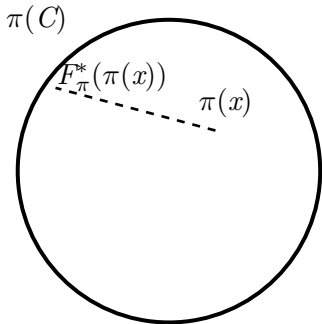
- Goal: $\sum_{C \in \Lambda} \text{ufl}(\pi(C)) \lesssim \text{ufl}(\pi(X))$.
- Idea: Open a “local” facility set $F'_{\pi(C)}$ for each $\pi(C)$ and show $\text{dist}(\pi(x), F'_{\pi(C)}) \lesssim \text{dist}(\pi(x), F_{\pi}^*), \forall x \in C$.



$$F'_{\pi(C)} :=$$

Lower Bound $\text{ufl}(\pi(X))$

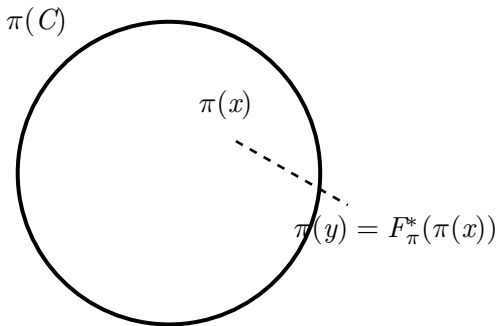
- Goal: $\sum_{C \in \Lambda} \text{ufl}(\pi(C)) \lesssim \text{ufl}(\pi(X))$.
- Idea: Open a “local” facility set $F'_{\pi(C)}$ for each $\pi(C)$ and show $\text{dist}(\pi(x), F'_{\pi(C)}) \lesssim \text{dist}(\pi(x), F_{\pi}^*), \forall x \in C$.



$$F'_{\pi(C)} := (F_{\pi}^* \cap \pi(C))$$

Lower Bound $\text{ufl}(\pi(X))$

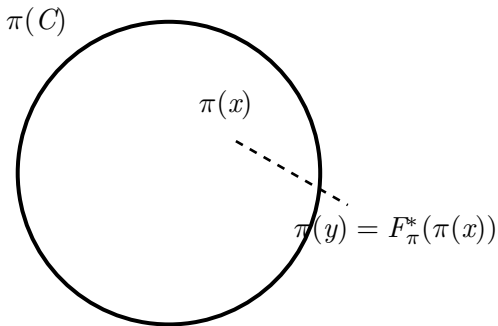
- Goal: $\sum_{C \in \Lambda} \text{ufl}(\pi(C)) \lesssim \text{ufl}(\pi(X))$.
- Idea: Open a “local” facility set $F'_{\pi(C)}$ for each $\pi(C)$ and show $\text{dist}(\pi(x), F'_{\pi(C)}) \lesssim \text{dist}(\pi(x), F_{\pi}^*), \forall x \in C$.



$$F'_{\pi(C)} := (F_{\pi}^* \cap \pi(C))$$

Lower Bound $\text{ufl}(\pi(X))$

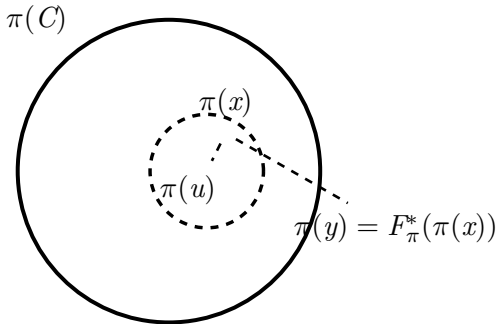
- Goal: $\sum_{C \in \Lambda} \text{ufl}(\pi(C)) \lesssim \text{ufl}(\pi(X))$.
- Idea: Open a “local” facility set $F'_{\pi(C)}$ for each $\pi(C)$ and show $\text{dist}(\pi(x), F'_{\pi(C)}) \lesssim \text{dist}(\pi(x), F_{\pi}^*), \forall x \in C$.



$$F'_{\pi(C)} := (F_{\pi}^* \cap \pi(C)) \cup \pi(\overbrace{N_C}^{\sigma \cdot \text{diam}(C)\text{-net on } C}).$$

Lower Bound $\text{ufl}(\pi(X))$

- Goal: $\sum_{C \in \Lambda} \text{ufl}(\pi(C)) \lesssim \text{ufl}(\pi(X))$.
- Idea: Open a “local” facility set $F'_{\pi(C)}$ for each $\pi(C)$ and show $\text{dist}(\pi(x), F'_{\pi(C)}) \lesssim \text{dist}(\pi(x), F_{\pi}^*), \forall x \in C$.



$$F'_{\pi(C)} := (F_{\pi}^* \cap \pi(C)) \cup \pi(\overbrace{N_C}^{\sigma \cdot \text{diam}(C)\text{-net on } C}).$$

Lower Bound $\text{ufl}(\pi(X))$

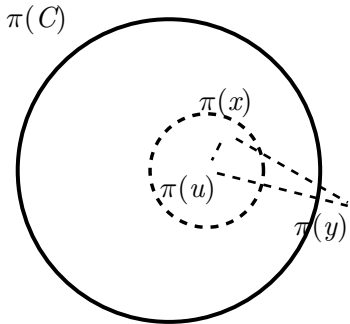
- Goal: $\sum_{C \in \Lambda} \text{ufl}(\pi(C)) \lesssim \text{ufl}(\pi(X))$.

- Idea: Open a “local” facility set $F'_{\pi(C)}$ for each $\pi(C)$ and show

$$\text{dist}(\pi(x), F'_{\pi(C)}) \lesssim \text{dist}(\pi(x), F_{\pi}^*), \forall x \in C.$$

$\sigma \cdot \text{diam}(C)$ -net on C

$$F'_{\pi(C)} := (F_{\pi}^* \cap \pi(C)) \cup \pi(\underbrace{N_C}_{\sigma \cdot \text{diam}(C)\text{-net on } C}).$$



- $\pi(x)$ and $\pi(u)$ are close.

- $\|x - u\| \leq \sigma \cdot \text{diam}(C) \implies \|\pi(x) - \pi(u)\| \leq O(\sigma) \cdot \text{diam}(C)$ w.h.p.

- $\pi(x)$ and $\pi(y)$ are separated.

- (x, y) is not badly cut $\implies \|x - y\| \geq \frac{\epsilon^2}{\text{ddim}} \cdot \text{diam}(C)$.

- $\|\pi(x) - \pi(y)\| \geq \frac{\Omega(\epsilon^2)}{\text{ddim}} \cdot \text{diam}(C)$ w.h.p.

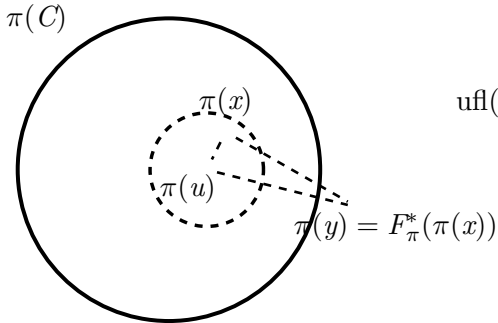
- Conclusion: $\|\pi(x) - \pi(u)\| \leq \|\pi(x) - \pi(y)\|$.

Lower Bound $\text{ufl}(\pi(X))$

- Goal: $\sum_{C \in \Lambda} \text{ufl}(\pi(C)) \lesssim \text{ufl}(\pi(X))$.
- Idea: Open a “local” facility set $F'_{\pi(C)}$ for each $\pi(C)$ and show

$$\text{dist}(\pi(x), F'_{\pi(C)}) \lesssim \text{dist}(\pi(x), F_{\pi}^*), \forall x \in C.$$

$$F'_{\pi(C)} := (F_{\pi}^* \cap \pi(C)) \cup \pi(\underbrace{\sigma \cdot \text{diam}(C)\text{-net on } C}_{N_C}).$$



$$\begin{aligned} \text{ufl}(\pi(C)) &\leq \sum_{x \in C} \text{dist}(\pi(x), F'_{\pi(C)}) + |F'_{\pi(C)}| \\ &\leq \underbrace{\sum_{x \in C} \text{dist}(\pi(x), F_{\pi}^*)}_{\text{connection cost on } \pi(C)} + \underbrace{|F_{\pi}^* \cap \pi(C)|}_{\text{opening cost on } \pi(C)} + |N_C|. \end{aligned}$$

$$\implies \sum_{C \in \Lambda} \text{ufl}(\pi(C)) \leq \text{ufl}(\pi(X)) + \text{err}.$$

PTAS for UFL

- Construct the partition Λ .
 - Hierarchical decomposition \mathcal{H} .
 - Eliminate badly-cut pairs $(x, F(x)) \implies \mathcal{T}$.
 - Construct Λ via \mathcal{T} .
- Construct near-optimal clustering \mathcal{C} .
 - For each $C \in \Lambda$, construct the near optimal clustering \mathcal{C}_C for $\pi(C)$.
 - $\mathcal{C} := \bigcup_{C \in \Lambda} \mathcal{C}_C$.
- Construct facility set F .
 - For each $X_i \in \mathcal{C}$, compute the (approximate) 1-median center f_i for X_i .
 - $F := \{f_1, f_2, \dots, f_s\}$.

Thank you!