Approximation Schemes for Euclidean k-Medians and Related Problems
S. Arora, P. Raghavan, S. Rao STOC '98

A Nearly Linear-Time App. Scheme for the Euclidean k-median Problem
S. Kolliopoulos, S. Rao ESA '99
k-medians Problem

Given

• $S = \{x_i\}$ be $n$ points in metric space $\mathbb{R}^d$
• Positive integer $k$

Goal

• Find $M = \{m_i\} \in \mathbb{R}^d$ which minimizes

$$
\sum_{i=1}^{n} \min_{1 \leq j \leq k} d(x_i, m_j)
$$
k-medians Problem

Paper 1

Paper 2
k-medians Problem

Paper 1

Paper 2
k-medians Problem

Paper 1

Paper 2
Facility Location Problem

Given
• $S=\{x_i\}$ be $n$ points in metric space $\mathbb{R}^d$
• Positive cost function $c()$

Goal
• Find $M=\{m_i\} \in S$ which minimizes

$$\sum_{m_j \in M} c(m_j) + \sum_{i=1}^{n} \min_{1 \leq j \leq k} d(x_i, m_j)$$
Facility Location Problem
Facility Location Problem
<table>
<thead>
<tr>
<th>Approximation Ratio</th>
<th>Author(s)</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>O(log n)</td>
<td>Hochbaum</td>
<td>'82</td>
</tr>
<tr>
<td>3.16</td>
<td>Shmoys et al.</td>
<td>'97</td>
</tr>
<tr>
<td>2.41</td>
<td>Guha and Khuller</td>
<td>'98</td>
</tr>
<tr>
<td>1.74</td>
<td>Chudak</td>
<td>'98</td>
</tr>
</tbody>
</table>

but no \((1+\varepsilon)\) approx before this paper
Previous Results: k-medi ans

$(1+\varepsilon)$ approx using $(1+1/\varepsilon)(1+\ln n)k$ medians
Lin and Vitter '92

$2(1+\varepsilon)$ approx using $(1+1/\varepsilon)k$ medians
Lin and Vitter '92
Results: Paper 1

In 2D, for the *k-median problem*, given any constant positive $\varepsilon$ with probability $1-o(1)$ achieve solution at most $(1+\varepsilon)OPT$ in time $O(n^{O(1/\varepsilon)}nk \log n)$

- Proof of existence
- Dynamic programming: bound table size
Results: Paper 1

In 2D, for the facility problem, given any constant positive $\varepsilon$ with probability $1-o(1)$ achieve solution at most $(1+\varepsilon)\text{OPT}$ in time $O(n^{1+O(1/\varepsilon)}\log n)$.
Results: Paper 2

In 2D, for the *k-median problem*, given any constant positive $\varepsilon$ with probability $1-o(1)$ achieve solution at most $(1+\varepsilon)OPT$ in time $O(2^{O(1+\log(1/\varepsilon)/\varepsilon)}nk\log n)$

- Proof of existence
- Dynamic programming: bound table size
Approximation Schemes for Euclidean $k$-Medians and Related Problems
S. Arora, P. Raghavan, S. Rao  
STOC '98
Integer Point Coordinates

- Given n points, minmax solution D is 2OPT for facility assignment
- Optimal facility cost in \([D/2, Dn]\)
- Arora-TSP '98: Mapping each point to \(jD/n^2\) increases cost by maximum \(O(D/n)\) (<\(\varepsilon\)?)

- Length of BBox(S)
  - \(L=O(n^4)\)
Terminology

- Dissection vs quadtree
- \( L = O(n^4) \)
- Number of nodes \( O(L^2) \)
- Minimum size of box = 1
- Depth of tree \( \log(L) \)

- Dissection with \((a,b)\) shift
- Quadtree derived from \((a,b)\)-shifted dissection
A Result from Arora-TSP-'98

$S=\{l_i\}$ is a collection of line segments

t$(S,l)$ number of lines in $S$ crossing $l$

- $|l_i| \geq 4$
- $\sum_{l_i \in S} |l_i| = cost(S)$

$$\sum_{i} t(S, l_i) = \theta(cost(S))$$
More Notations

- **BBox** at level 0
  - its 4 children at level 1, ...
- **Level** of an edge
  - level of the corresponding box
- **Edge** is in level \( i \) \( \Rightarrow \) in \((i+1), (i+2), \ldots, \log L\)
- **maximal level**

\[
\Pr_a[l \text{ is at level } i] = \frac{2^i}{L}
\]
Charging Scheme

R-charging

For a maximal edge which is crossed \( g \) times by \( S \), charge

\[
g/R \times \text{length of edge}
\]
Charging Lemma

• Expected total cost for top i-level edges is $O(i\text{cost}(S)/R)$

• Maximal level of grid line l be j
  – Length $L/2^j$
  – Charge $t(S,l)/R \cdot L/2^j$
  – Probability $2^j/L$

by linearity of expectation,

$$\sum_{j \leq i} 2^j/L \cdot t(S,l)/R \cdot L/2^j \leq i/R \cdot t(S,l)$$
Charging Lemma

by linearity of expectation,

\[ \sum_{j \leq i} 2^j / L \ t(S, l) / R \ L / 2^j \leq i / R \ t(S, l) \]

• \( i \leq \log L \)
• Choose \( R = \log L / \varepsilon \)

• Cost \( \leq \varepsilon \ t(S, l) \)
• Total \( \leq O(\varepsilon \ \text{cost}(S)) \)
m-portal

- m-regular portals for a shifted dissection
- Total number of points $\Rightarrow 4m$
m-portal

- m-regular portals for a shifted dissection
- Total number of points ⇒ 4m
m-light Solution
m-light Solution
Existence

• \( m > 1 \) and \((a, b) \in U[0, L]\)

• \( \text{OPT} \Rightarrow \) deflect to form \( m \)-portal
  – for \( l \) sized square, cost of deflection \( O(l/m) \)
  – \( m \)-charging scheme

• w.p. \( \frac{1}{2} \) or in expectation or w.h.p.
  \( m \)-light solution exists for \((a, b)\) shifted
dissection with cost at most

\[
\sum_{l} O(i/R \ t(S, l))
\]

\( \Rightarrow \ O(\log L/m \ \text{OPT}) \)
Existence

\[ O \left( \frac{i}{R} \ t(S,l) \right) \]
\[ \Rightarrow O \left( \frac{\log L}{m} \ \text{OPT} \right) \]

Solution is \( O( (1+\log L/m)\text{OPT} ) \)

- \( L = O(n^4) \)
- \( \text{Choose } m \) to get \( O( (1+\epsilon)\text{OPT} ) \)
Dynamic Programming: Sketch

• Finds solution within \((1+1/4m)\) of m-light OPT

\begin{align*}
\Rightarrow & \quad O((1+\varepsilon/4\log n)(1+\varepsilon)OPT) \\
& = O((1+\varepsilon)OPT)
\end{align*}

Dynamic Programming
1. Nearest facility within \((1+1/4m)\)
2. Nearest facilities similar for neighbors
Dynamic Programming: Sketch

Given f and sub-boxes (children boxes $S_i$-s) solve for current level

- Table build for all choices $f(\leq k)$ and S
- Table size $O(n^c)$
  $\Rightarrow$ worst case time for algorithm
Summary

In 2D, for the *k-median problem*, given any constant positive ε with probability 1-o(1) achieve solution at most (1+ε)OPT in time $O(n^{O(1/\varepsilon)}nk \log n)$

- Proof of existence
- Dynamic programming: bound table size
A Nearly Linear-Time App. Scheme for the Euclidean k-median Problem
S. Kolliopoulos, S. Rao  ESA '99
k-medians

– Given
  • $S=\{x_i\}$ be $n$ points in metric space $\mathbb{R}^d$
  • Positive integer $k$

– Goal
  • Find $M=\{m_i\} \in S$ in metric space $\mathbb{R}^d$ which minimizes

$$\sum_{i=1}^{n} \min_{1 \leq j \leq k} d(x_i, m_j)$$
Paper 2

\(O(n^{O(1/\varepsilon)}nk \log n)\)

reduced to

\(O(O(1/\varepsilon)nk \log n)\)

• Reduce the number of portals from

\(O(\log n/\varepsilon) \Rightarrow O(1+\log(1/\varepsilon)/\varepsilon)\)

• Different construction (no \((a,b)\) shifting)
Adaptive Dissection

Sub-rectangle
Adaptive Dissection

Sub-rectangle
Adaptive Dissection

Sub-rectangle

$$B_s = B \cap B'$$
Adaptive Dissection

Cut-rectangle (randomization)
Adaptive Dissection

Cut-rectangle (randomization)
Adaptive Dissection

Cut-rectangle (randomization)
Lemma (just one of many)

- Given two parallel cut-lines (due to cut-rectangle) are $L$ apart, the line segments has side length less than $3L$. 
Structure Theorem

Error due to assignment using m-portal respecting paths is bounded by

\[ O(\max\{1, \log(m)\} / m \ OPT) \]

Choose

\[ \log(m) / m \leq c \varepsilon \]
Extensions

• d-dimension $\Rightarrow m^{d-1}$-portal
• Facility location
  – Same structure
• Capacitated k-median
  – Tweak dynamic programming
• Few medians (small k)
  – Guess position of facilities
  – Number of choices $\binom{m^2}{k}$
k-medians Problem

\[ \sum_{i=1}^{n} \min_{1 \leq j \leq k} d(x_i, m_j) \]
k-centers Problem

\[ \sum_{i=1}^{n} \min_{1 \leq j \leq k} d^2(x_i, m_j) \]