Geometric clustering in the normed plane

Pedro Martín

University of Extremadura, Badajoz

Cáceres, March 2016
$M^2 = (\mathbb{R}^2, \| \cdot \|)$ is a 2-dimensional normed (or Minkowski) plane.
Geometric clustering

$\mathbb{M}^2 = (\mathbb{R}^2, \| \cdot \|)$ is a 2-dimensional *normed* (or *Minkowski*) plane. Let $S$ be a set of $n$ points in the normed plane and $k$ a fixed number.
Geometric clustering

$\mathbb{M}^2 = (\mathbb{R}^2, \| \cdot \|)$ is a 2-dimensional normed (or Minkowski) plane. Let $S$ be a set of $n$ points in the normed plane and $k$ a fixed number.

How can $S$ be separated (by an algorithm) in $k$ clusters verifying some conditions?
$\mathbb{M}^2 = (\mathbb{R}^2, \| \cdot \|)$ is a 2-dimensional normed (or Minkowski) plane. Let $S$ be a set of $n$ points in the normed plane and $k$ a fixed number.

How can $S$ be separated (by an algorithm) in $k$ clusters verifying some conditions?
Geometric clustering

$k = 1$, minimizing the radius of a enclosing disc:
- Elzinga-Hearn and Shamos-Hoey (Euclidean plane).
- Alonso-Martini-Spirova and Jahn (general normed plane).

$k = 2$, minimizing the maximum Euclidean diameter of the clusters:
- Avis, $O(n^2 \log n)$.
- Asano-Bhattacharya-Keil-Yao, $O(n \log n)$.

$k = 2$, minimizing the sum of the two Euclidean diameters:
- Monma-Suri, $O(n^2)$.

$k = 2$, $\mu$ a measure, $\mu_1 > 0$ and $\mu_2 > 0$, splitting $S$ into two clusters $A$ and $B$ such that $\mu(A) \leq \mu_1$ and $\mu(B) \leq \mu_2$:
- Hershberger and Suri,
  - $\mu =$Euclidean diameter, $O(n \log n)$.
  - $\mu =$area, perimeter, or diagonal of the smallest rectangle with sides parallel to the coordinates axes ($O(n \log n)$ time).
  - $\mu =$radius of the smallest enclosing sphere with the norms $L_1$ ($O(n \log n)$ time) or the Euclidean norm ($O(n^2 \log n)$ time).
Geometric clustering

\( k = 2 \), the **2-center problem**: cover \( S \) by (the union of) two congruent closed disks whose radius is as small as possible.

- Eppstein and Sharir (1997), near linear time cost (Euclidean case).

\( k = 3 \), minimizing the maximum Euclidean diameter

- Hagauer-Rote, \( O(n^2 \log^2 n) \)

Any \( k \), minimizing any monotone function \( \mathcal{F} \) (\( \mathcal{F} : \mathbb{R}^k \to \mathbb{R} \)) of the Euclidean diameters or the Euclidean radii of the clusters.

Examples of \( \mathcal{F} \):

- The sum of the diameters (or the radii)
- The maximum of the diameters (or the radii)
- The sum of the squares of the diameters (or the radii).

- Capoyleas-Rote-Woeginger, polynomial time.
Hagauer-Rote and Capoyleas-Rote-Woeginger obtain their results from this theorem

**Theorem (Capoyleas-Rote-Woeginger)**

*Let $A$ and $B$ be two sets of points in the Euclidean plane. Then, there are two linearly separable sets $A'$ and $B'$ such that $\text{diam}(A') \leq \text{diam}(A)$, $\text{diam}(B') \leq \text{diam}(B)$, and $A' \cup B' = A \cup B$.***

*Figure: Non linearly separable (left) and linearly separable sets (right)*
Linear separation of clusters

This first statement is used in the proof of the Theorem: *In every triangle with an obtuse angle, the side lying opposite to the obtuse angle is the (Euclidean) longest side in the triangle.*
This first statement is used in the proof of the Theorem: *In every triangle with an obtuse angle, the side lying opposite to the obtuse angle is the (Euclidean) longest side in the triangle.*

Figure: The side opposite to the obtuse angle is not the longest side in the triangle $\triangle abc$. 
Linear separation of clusters

This second statement is used in the proof of Theorem:

1. \( \text{diam}(A) \geq \text{diam}(B) \)

2. \( \{a_i, a'_i, a_m\} \subset A, \{b_j, b'_j\} \subset B \)

Clockwise order: \( a'_i, b'_j, a_m, b_j, a_i \)

3. \( \langle b_j, b'_j \rangle \) separates \( \{a_i, a'_i\} \) from \( a_m \).

\[
\begin{cases}
\{ \|a_i - b_j\|, \|a'_i - b'_j\| \} \\
(\mathbb{E}^2) \leq \text{diam}(A).
\end{cases}
\]
Linear separation of clusters

But this point configuration is possible in a general normed plane:

Figure: $\|a_i - b_j\|$ and $\|a'_i - b'_j\|$ are longer than the diameter of $A$. 
Linear separation of clusters

Objective: to prove the Theorem for any normed plane.
Linear separation of clusters

Step 1: \( \{u_1, u_2, \ldots, u_{2k}\} = \partial(\text{conv}(A)) \cap \partial(\text{conv}(B)). \)
Linear separation of clusters

We can assume that $\text{diam}(A) \geq \text{diam}(B)$

We say that...

- $(A_i, B_j)$ is a \textit{bad pair} if $\text{diam}(A_i \cup B_j) > \text{diam}(A)$.
  Then, $A_i$ and $B_j$ are \textit{bad partners}.

- $a_i \in A_i$ and $b_j \in B_j$ are \textit{bad points} if $\|a_i - b_j\| > \text{diam}(A)$.
  Then, $a_i$ and $b_j$ are \textit{bad partners},
  and the segment $a_i b_j$ is a \textit{bad segment}.
Lemma
Let \((A_i, B_j)\) and \((A_i', B_j')\) two disjoint bad pairs. Let us choose \(a_i \in A_i, b_j \in B_j, a_i' \in A_i', b_j' \in B_j'\) such that \(a_i b_j\) and \(a_i' b_j'\) are bad segments. Then, either these bad segments intersect, or any point \(a \in A_m\) belonging to the halfplane defined by \(\langle b_j b_j' \rangle\) where \(a_i\) and \(a_i'\) are not contained, is not bad.
Linear separation of clusters

Lemma
Let \((A_i, B_j)\) and \((A'_i, B'_j)\) two disjoint bad pairs. Let us choose \(a_i \in A_i, b_j \in B_j, a'_i \in A'_i, b'_j \in B'_j\) such that \(a_i b_j\) and \(a'_i b'_j\) are bad segments. Then, either these bad segments intersect, or any point \(a \in A_m\) belonging to the halfplane defined by \(< b_j b'_j >\) where \(a_i\) and \(a'_i\) are not contained, is not bad.

Sketch of the proof. Possible clockwise order (up to symmetries):

Case 1: \(a_i, b'_j, a'_i, b_j\)  
Case 2: \(a_i, a'_i, b'_j, b_j\)
Linear separation of clusters

Case 1: clockwise order

\[ a_i, b_j', a_i', b_j \]

We get a contradiction:

\[
\text{diam}(A) + \text{diam}(B) \geq \| a_i - a_i' \| + \| b_j - b_j' \| \geq \\
\| a_i - b_j \| + \| a_i' - b_j' \| > 2 \text{ diam}(A).
\]
Case 2: clockwise order $a_i, a_{i'}, b_{j'}, b_j$:

Figure: $(a_i, b_j), (a_{i'}, b_{j'})$ are bad partners $\implies \nexists$ any bad partner for $a_m$
Linear separation of clusters

Case 2: clockwise order $a_i, a_{i'}, b_{j'}, b_j$:

Figure: $(a_i, b_j)$ and $(a_{i'}, b_{j'})$ bad partners $\implies$ $\nexists$ any bad partner for $a_m$
Linear separation of clusters

Step 2: Maximal cyclic subsequences of polygons.
**Linear separation of clusters**

**Step 2: Maximal cyclic subsequences of polygons.**

- Consider maximal cyclic subsequences of adjacent bad polygons $A_i$.
  - No ”good” polygon $A_k$ belongs to one of this maximal cyclic subsequences of bad $A_i$-polygons.
  - Some intervening ”good” polygon $B_j$ can belong to this maximal cyclic subsequences of $A_i$-polygons.
- Similarly with adjacent bad polygons $B_j$.
- These maximal cyclic sequences are noted by $\bar{A}_1, \bar{A}_2, \ldots, \bar{A}_p$ and $\bar{B}_1, \bar{B}_2, \ldots, \bar{B}_q$. 
Linear separation of clusters

Example with 3 maximal cyclic subsequences of $A_i$-polygons and 3 maximal subsequences of $B_j$-polygons:

$\bar{A}_1 = \{A_1\}$
$\bar{B}_1 = \{B_1\}$
$\bar{A}_2 = \{A_2\}$
$\bar{B}_2 = \{B_3\}$
$\bar{A}_3 = \{A_4\}$
$\bar{B}_3 = \{B_5\}$
Linear separation of clusters

Example with 3 maximal cyclic subsequences of $A_i$-polygons, 3 maximal subsequences of $B_j$-polygons, and "good" intervening polygons:

\[
\tilde{A}_1 = \{A_1\} \\
\tilde{B}_1 = \{B_1\} \\
\tilde{A}_2 = \{A_2, B_2, A_3\} \\
\tilde{B}_2 = \{B_3\} \\
\tilde{A}_3 = \{A_4\} \\
\tilde{B}_3 = \{B_4, A_5, B_5\}
\]
Linear separation of clusters

Properties

- Let \((A_i, B_j)\) and \((A_i', B_j')\) be two disjoint bad pairs. Then

\[A_i, A_i' \in \bar{A}_k \implies B_j, B_j' \in \bar{B}_t\]

- The number of maximal cyclic sequences of adjacent bad \(A_i\)-polygons and \(B_j\)-polygons is the same.

- If \((\bar{A}_i, \bar{B}_j)\) and \((\bar{A}_i', \bar{B}_j')\) are disjoint bad pairs of maximal subsequences, then there exist two (one from every pair) bad-crossing segments.

- There is an odd number of subsequences from each cluster, and they must be completely interlacing.
Linear separation of clusters

Step 3: Separate the sets.
Linear separation of clusters

- Let $A_i$ be the last polygon of a maximal cyclic subsequence (in clockwise order).
- Let $B_j$ be the last bad partner of $A_i$.
- Let $B_{j'}$ be the first bad polygon after $A_i$.
- Let $A_{i'}$ be the first bad partner of $B_{j'}$.
- Choose the line $L$ going through the point just before $B_j$ and the point just after $B_{j'}$.
- Define $B'$ to be the points in $A \cup B$ lying on the same side of $L$ as $B_j$ and $B_{j'}$, and $A'$ as the remaining points.
Linear separation of clusters

$A_i$ and $B_j$
Linear separation of clusters

Proposition

\[ \text{diam}(A') \leq \text{diam}(A), \quad \text{diam}(B') \leq \text{diam}(B). \]

Theorem

Let \( A \) and \( B \) be two sets of points in a general normed plane. Then, there are two linearly separable sets \( A' \) and \( B' \) such that \( \text{diam}(A') \leq \text{diam}(A), \text{diam}(B') \leq \text{diam}(B), \text{ and } A' \cup B' = A \cup B. \)

Corollary

In the construction in the Theorem,

\[ \text{perimeter}(A) + \text{perimeter}(B) \geq \text{perimeter}(A') + \text{perimeter}(B') \]

holds. If \( \text{conv}(A) \cap \text{conv}(B) \neq \emptyset \), then the inequality is strict.
Some consequences

The 2-clustering problem for diameter respect to the minimum: Dividing $S$ in two sets minimizing the maximum diameter of the sets.

**Theorem**

*Given a set $S$ of $n$ points in a normed plane, the 2-clustering problem for diameter respect to the minimum can be computed in $O(n^2 \log^2 n)$ time.*

- Sort the distances $d_i$ between the points of $S$ into increasing order.
- By a binary search, locate the minimum $d_i$ that admits a *stabbing line* for the set of segments meeting point of $S$ at distance greater than $d_i$. 

Some consequences

The $k$-clustering problem for diameter respect to a function $\mathcal{F}$ (for example, $\mathcal{F}$ can be the $maximum$, the $sum$, or the $sum$ of $squares$):

Dividing $S$ in $k$ sets minimizing a function $\mathcal{F}$ of the diameters of the sets.

**Theorem**

Consider the optimal $k$-clustering problem for the diameter respect to a monotone increasing function $\mathcal{F}$ of such as diameters. For every set $S$ of $n$ points in a general normed plane,

- There is an optimal $k$-clustering such that each pair of clusters is linearly separable.
- The problem is solvable by an algorithm in polynomial time.
Thank you very much!