Abstract

Given two sets of points $A$ and $B$ in a normed plane, we prove that 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 $A' \cup B' = A \cup B$. As a result, some Euclidean clustering algorithms are adapted to normed planes, for instance, those that minimize the maximum, the sum, or the sum of squares of the diameters (or the radii) of $k$ clusters. Some specific solutions are presented for $k = 2$ and $k = 3$ that minimize the diameter of the clusters. The 2-clustering problem when two different bounds are imposed to the diameters is also studied.

Keywords: geometric clustering, normed plane.
1 Introduction and preliminaries

Given a set $S$ of $n$ points in the plane, a cluster is any non-empty subset of $S$, and a $k$-clustering is a set of $k$ disjoint clusters such that any point of $S$ belongs to some cluster. For a fixed distance function on the plane, in general, a clustering problem asks for a $k$-clustering of $S$ that verify some conditions, for example minimizing the maximum diameter of the clusters.

From now on, we denote by $\mathbb{E}^2$ the Euclidean plane, and by $\mathbb{M}^2$ a normed plane. We call $B(x;r)$ to the ball with center $x \in \mathbb{M}^2$ and radius $r > 0$, and $S(x;r)$ to the sphere of $B(x;r)$. We use the usual abbreviations $\text{diam}(A)$ and $\text{conv}(A)$ for the diameter and the convex hull of a set $A$, $\overline{ab}$ for the line segment meeting two points $a, b \in \mathbb{M}^2$, and $\langle a, b \rangle$ for its affine hull. $p(A)$ denotes the perimeter of $\text{conv}(A)$.

We say that two sets of points in $\mathbb{M}^2$ are linearly separable if there exists a line $L$ such that every set is situated in a different closed half-plane defined by $L$. We present the following theorem that extends an Euclidean result ([4]) to any normed plane.

**Theorem 1.1** Let $A$ and $B$ be two finite sets in $\mathbb{M}^2$. Then, there are two linearly separable sets $A'$ and $B'$ such that $\text{diam}(A') \leq \text{diam}(A)$, $\text{diam}(B') \leq \text{diam}(B)$, $A' \cup B' = A \cup B$. Besides, $p(A) + p(B) \geq p(A') + p(B')$.

**Sketch of the proof.** The structure of our proof is similar to that presented for $\mathbb{E}^2$, but some important modifications are necessary due to the general framework of $\mathbb{M}^2$. Let us assume that $\text{diam}(A) \geq \text{diam}(B)$. The first step is to consider the sequences of polygons (clockwise) $\{A_1, A_2, \ldots, A_k\}$ and $\{B_1, B_2, \ldots, B_k\}$ such that $\bigcup_{i=1}^k A_i = \text{conv}(A) \setminus \text{conv}(B)$ and $\bigcup_{j=1}^k B_j = \text{conv}(B) \setminus \text{conv}(A)$. We say that $(A_i, B_j)$ is a bad pair if $\text{diam}(A_i \cup B_j) > \text{diam}(A)$. In such a case we say that both $A_i$ and $B_j$ are bad polygons. The second step is to make a group with each subsequences of adjacent (clockwise) bad polygons $A_i$ from $A$ (intervening subsets $B_j$ of $B$ inside a group from $A$ must not be bad), and the same is made with $B$. In the final step, we separate the sets: for a fixed a group, let $A_j$ be the last bad set (clockwise) of such a group, and let $B_{j'}$ be the last bad partner of $A_j$; let $B_{j'}$ be the first bad set after $A_i$, and $A_{j'}$ be the first bad partner of $B_j$; the separating line $L$ goes through the point of $\text{conv}(A) \cap \text{conv}(B)$ before $B_j$ and the point $\text{conv}(A) \cap \text{conv}(B)$ after $B_{j'}$. We 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.
2 Some algorithms for clustering problems

We assume that in our computation model an oracle answers the required questions about the unit ball of $\mathbb{M}^2$ (see Section 3.3 of [6]).

**Theorem 2.1** Given a set of $n$ points in $\mathbb{M}^2$, the 2-clustering problem of minimizing the maximum diameter can be solved in $O(n^2 \log^2 n)$ time.

**Proof.** The following works correctly (Theorem 1.1): sort the distances $d_i$ between the points of $S$ into increasing order ($O(n^2 \log n)$ time); locate the minimum $d_i$ that admits a stabbing line by a binary search. Use the graph $(S, E_{d_i})$, where $E_{d_i}$ is the set of edges meeting two points of $S$ at distance more than $d_i$, and the algorithm by Edelsbrunner et al. ([5]) in order to find the stabbing line for $E_{d_i}$ as a subroutine ($O(m \log m)$, where $m$ is the number of edges). □

The above approach was presented by Avis [1] for the Euclidean case ($O(n^2 \log^2 n)$ time) and was improved by Asano et al. ([2], $O(n \log n)$ time) using a maximum spanning tree of $S$. There is not an efficient method for building a maximum spanning tree for any normed plane.

**Theorem 2.2** Given a set $S$ of $n$ points in $\mathbb{M}^2$, and $d_1 \geq d_2 > 0$, the 2-clustering problem of dividing $S$ into two sets $S_1$ and $S_2$ such that $\text{diam}(S_1) \leq d_1$ and $\text{diam}(S_2) \leq d_2$ can be computed in $O(n^2 \log^2 n)$ time.

**Proof.** Let $E_{d_1}$ be the set of edges meeting two points of $S$ at a distance of more than $d_1$. Sort the distances between the points of $S$ into increasing order and build the graph $(S, E_{d_1})$ in $O(n^2 \log n)$ time. Test if $E_{d_1}$ has a stabbing line (in $O(n \log n)$ time with the algorithm presented in [5]). If the stabbing line does not exist, there is no solution (Theorem 1.1). Check if a stabbing line separates a set with a diameter less than or equal to $d_2$. □

The Euclidean approach to this problem by Hershberger and Suri ([8], $O(n \log n)$ time) does not work in $\mathbb{M}^2$ because they use the fact (not true in $\mathbb{M}^2$) that if $\|a - b\| \geq d_1$, then $B(a, d_2) \cap B(b, d_1)$ can always be split into two subsets whose diameters are at most $d_1$ and $d_2$, respectively.

**Theorem 2.3** Let $S$ be a set of $n$ points in $\mathbb{M}^2$. Consider the k-clustering problem of minimizing a monotone increasing function $F : \mathbb{R}^k \rightarrow \mathbb{R}$ to the diameters or to the radii of $k$ subsets of $S$. Then there is an optimal $k$-clustering such that each pair of clusters is linearly separable. A solution can be obtained by the algorithm presented by Capoyleas, Rote, and Woeginger ([4]), and it takes polynomial time for the case of the diameters.
Given a set of \(S\) and \(S\), the part of \(S\) defined by the line that contains \(p\) points orthogonal to or has only one connected component, \(C_i\) and \(C_j\) are separable. If \(S(c_i, r_i) \cap S(c_j, r_j)\) has two different components, let us consider a line meeting two \(c_i\) and \(c_j\) assigning to \(C_i\) and \(C_j\) respectively, such that \(C_i \subset B(c_i, r_i)\) and \(C_j \subset B(c_j, r_j)\). If \(S(c_i, r_i) \cap S(c_j, r_j)\) is the empty set or has only one connected component, \(C_i\) and \(C_j\) are separable. If \(S(c_i, r_i) \cap S(c_j, r_j)\) has two different components, let us consider a line meeting two points \(p^1\) and \(p^2\), one point from each component. Let \(H_{ij}\) be the half-plane defined by the line that contains \(p^1 - (c_j - c_i)\) and \(p^2 - (c_j - c_i)\), and \(H_{ji}\) be the other half-plane. Let \(S_1(c_i, r_i)\) be the part of \(S(c_i, r_i)\) on \(H_{ij}\), and \(S_2(c_i, r_i)\) be the part of \(S(c_i, r_i)\) on \(H_{ji}\). Let \(S_1(c_j, r_j)\) be the part of \(S(c_j, r_j)\) on \(H_{ji}\), and \(S_2(c_j, r_j)\) be the part of \(S(c_j, r_j)\) on \(H_{ij}\). Then, \(S_2(c_i, r_i) \subseteq \text{conv}(S_1(c_j, r_j))\) and \(S_2(c_j, r_j) \subseteq \text{conv}(S_1(c_i, r_i))\) (see Banasiak [3]).

We can reassign the points according to their position relative to the separating lines:

\[C'_i := S \cap B(c_i, r_i) \cap \left( \bigcap_{j=1, j \neq i}^k H_{ij} \right) \quad i = 1, \ldots, k.\]

The new clusters are separable and the value of \(\mathcal{F}\) does not increase.

\[\square\]

**Theorem 2.4** Given a set of \(n\) points in \(\mathbb{M}^2\) and \(d > 0\), we can determine in \(O(n^3 \log^2 n)\) time with the approach by Hagauer and Rote ([7]) whether there is a partition of \(S\) into sets \(A, B, C\) with diameters at most \(d\), and construct in \(O(n^3 \log^3 n)\) time a 3-partition of \(S\) such that the largest of the three diameters is as small as possible.

**Proof.** (Scheme) We fix a normal basis \(\{x, y\}\) in \(\mathbb{M}^2\) such that \(x\) is Birkhoff orthogonal to \(y\) (namely, such that \(\|x\| \leq \|x + \lambda y\|\) for every \(\lambda \in \mathbb{R}\)). It is assumed that two given points of \(S\) have different \(x\) and \(y\) coordinates (the points are rotated if it is necessary). Given \(d > 0\), the algorithm searches all the possible linearly separable subsets \(A, B, C\), such that the maximum diameter is less than or equal to \(d\). The point \(a \in S\) with the minimum \(x\)-coordinate is placed in \(A\), and each point \(a' \in S\) such as \(\|a - a'\| \leq d\) is tested as the possible point of \(A\) with the maximum \(x\)-coordinate. Any \(u \in S \cap aa'^\perp\) is assigned to \(A\). The plane is divided into the following three zones by the lines \(\langle a, a' \rangle\) and \(a' + \beta y\) (\(\beta \in \mathbb{R}\)). EAST contains the points of \(S\) on the ”right” of the line \(a' + \beta y\). The points of \(S\) on the left of \(a' + \beta y\) are contained either...
in NORTH (if they are "above" \(\overline{ad}\)) or in SOUTH (if they are "below" \(\overline{ad}\)). Solutions are tested in three different cases: Case 1, NORTH \(\subseteq A\); Case 2, SOUTH \(\subseteq A\); and Case 3, NORTH and SOUTH are not completely contained in A. We note \(A_{cand}\) to the set of points that could be placed in A for every candidate \(a'\):

\[
A_{cand} = S \cap B(a, d) \cap B(a', d).
\]

Theorem 2.4 is proved in \(\mathbb{E}^2\) using some lemmas (from Lemma 3 to Lemma 6 in [7]) and Theorem 1.1 (for \(\mathbb{E}^2\)). We prove results similar to the rest of the lemmas in [7] for any normed plane (using Birkhoff orthogonality). We present below Lemma 2.5 as an example. Regarding the complexity of the algorithm, we can justify that the data structure introduced by Hershberger and Suri ([8]) is usable in the same way as in [4]. Finally, a binary search on the \(\binom{n}{2}\) distances occurring in \(S\) solves the optimization problem. □

**Lemma 2.5** Let us assume the following conditions in \(\mathbb{M}^2\): A, B, C are separable; \(\max\{\text{diam}(A), \text{diam}(B), \text{diam}(C)\} \leq d\); \(B \cap \text{NORTH} \neq \emptyset\) and \(C \cap \text{SOUTH} \neq \emptyset\). If there exists a pair of points \(u = (u_x, u_y), v = (v_x, v_y) \in \text{EAST}\) such that \(\|u - v\| > d\) and \(u_y > v_y\), then \(u \in B\) and \(v \in C\).

**Proof.** Since \(\|u - v\| > d\), the points \(u\) and \(v\) cannot be situated in the same subset of the partition \(A, B, C\). We can choose \(u' = (u'_x, u'_y) \in B \cap \text{NORTH}\) and \(v' = (v'_x, v'_y) \in C \cap \text{SOUTH}\).

Let us consider the shaded zone, that is the part of the East hidden behind the segment \(\overline{u'v'}\) or behind the \(\overline{ad}\) from the point of view \(u'\) (see Figure 1).

Let see what happens if \(u'_y > u_y > v'_y\) (the analysis is similar in the rest of the possible relative positions of \(u\), \(u'\) y \(v'\)). There are three cases (regarding the position of \(v\)).

Case 1: \(v\) belongs to the shaded zone. Then \(v\) must belong to \(C\), because in another case either the pair \(\overline{vu'}\) and \(\overline{wu}\) or the pair \(\overline{vu'}\) and \(\overline{a'd}\) cross.

Case 2: \(v\) does not belong to the shaded zone and \(u_x < v_x\) (for instance, \(v = v_1\) in Figure 1). Then, we consider the two intersection points of the line \(u + \lambda y\) with the line \(v + \lambda x\) and with the line \(v + \lambda (u' - v)\), that we denote by \(\overline{u} \) and \(\overline{u}\), respectively. Since \(x\) is Birkhoff orthogonal to \(y\), \(u + \lambda y\) supports \(S(v, \|\overline{u} - v\|)\) on \(\overline{u}\), and \(\|v - u'\| \geq \|v - \overline{u}\| \geq \|v - u\| \geq \|v - \overline{u}\|\). As a result of \(\|v - u'\| \geq \|v - u\| > d, v \in C\).

Case 3: \(v\) does not belong to the shaded zone and \(u_x > v_x\) (for instance, \(v = v_2\) in Figure 1). Then, we consider the two intersection points of the line \(v + \lambda y\) with the line \(u + \lambda x\) and with the line \(u + \lambda (u - v')\), that we note by \(\overline{v}\) and \(\overline{v}\), respectively. Since \(x\) is Birkhoff orthogonal to \(y\), the line \(v + \lambda y\) is the support line of \(S(u, \|u - \overline{v}\|)\) on \(\overline{v}\), and \(\|u - v'\| \geq \|u - \overline{v}\| \geq \|u - v\| \geq \|u - \overline{v}\|\).
As a result of $\|u - v'\| \geq \|u - v\| > d$, $u \in B$.

References


