HAL
open science

Identification of points using disks

Valentin Gledel, Aline Parreau

To cite this version:

Valentin Gledel, Aline Parreau. Identification of points using disks. Discrete Mathematics, 2019, 342 (1), pp.256-269. 10.1016/j.disc.2018.10.002 . hal-01531101

HAL Id: hal-01531101

https://hal.science/hal-01531101v1

Submitted on 1 Jun 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Identification of points using disks

Valentin Gledel* Aline Parreau†

June 1, 2017

Abstract

We consider the problem of identifying n points in the plane using disks, i.e., minimizing the number of disks so that each point is contained in a disk and no two points are in exactly the same set of disks. This problem can be seen as an instance of the *test covering problem* with geometric constraints on the tests. We give tight lower and upper bounds on the number of disks needed to identify any set of n points of the plane. In particular, we prove that if there are no three colinear points nor four cocyclic points, then $2\lceil n/6 \rceil + 1$ disks are enough, improving the known bound of $\lceil (n+1)/2 \rceil$ when we only require that no three points are colinear. We also consider complexity issues when the radius of the disks is fixed, proving that this problem is NP-complete. In contrast, we give a linear-time algorithm computing the exact number of disks if the points are colinear.

1 Introduction

Let $\mathcal{P}$ be a set of n points of the plane $\mathbb{R}^2$. What is the minimum number of disks so that each point is contained in a disk and no two points are in exactly the same set of disks? In other words, we want to find a minimum set of disks such that every point is in a disk and the disks that contain a given point uniquely determine it. Such a set of disks (not necessarily minimum) is said to *identify* $\mathcal{P}$. See Figure 1 for an example of an identifying set of disks.

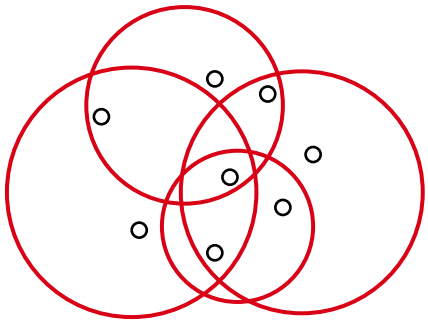

Figure 1: A set of four disks optimally identifying eight points.

The motivation of this problem comes from the localization of indviduals and more generally from the following setting of identification problems: Given a set of individuals with binary attributes that each individual can have or not, the goal is to choose a minimum number of attributes in such a way that each individual has a unique set of attributes. This problem is known in the literature as the *test covering problem* [16] or *identifying codes problem in hypergraph* [15] since one can represent the data by a hypergraph where individuals are vertices and attributes are hyperedges. It has many application in particular in medical diagnostics and pattern recognition [13, 16, 20].

*Univ Lyon, Université Claude Bernard Lyon 1, LIRIS - CNRS UMR 5205, F69622 (France). E-mail: valentin.gledel@univ-lyon1.fr

†Univ Lyon, Université Claude Bernard Lyon 1, CNRS, LIRIS - CNRS UMR 5205, F69622 (France). E-mail: aline.parreau@univ-lyon1.fr

In a context of localization, the attributes are defined by the metric of the space where individuals live. As an example, in the context of *identifying codes* [12], individuals are vertices of a graph. Then the attributes are defined by the closed neighbourhoods, meaning "to be closed to". Choosing some attributes is equivalent to setting detectors on some vertices that are able to detect errors in their neighbourhood. Then the set of detectors is able to detect any intrusion in the graph. Indeed, if there is an intrusion on a vertex, then the set of detectors that have detected something uniquely determines where the intrusion is. Locating-dominating sets [22, 23] and open locating dominating sets [21] are defined in a similar way. These concepts are studied by various authors since the 1970s and 1980s, and have been applied to various regions such as fault-detection in networks [12, 25] or graph isomorphism testing [2].

In this paper, we consider that individuals (which are the points in our problem) are living in $\mathbb{R}^2$. A detector can be placed anywhere, with any radius of detection and thus is represented by a disk. It can be formulated as a test covering instance: a set of individuals share an attribute if they can be isolated from the other individuals by a disk. It is also related to identifying codes in graphs: if the detectors must be located on points and have a fixed radius, a natural graph structure emerges. Then our problem is equivalent to the problem of identifying codes in unit disk graphs (in the general case) or in unit interval graphs (if points are colinear).

Another motivation comes from the notion of *geometric separator* in computational geometry [7]. Let $C_1, \ldots, C_k$ be k finite disjoint sets of $\mathbb{R}^2$. A finite set S of curves in the plane is a *separator* for the sets C_1,...,C_k if every connected component in $\mathbb{R}^2 - S$ contains points from only one set C_i. Finding separators is a classical problem of computationnal geometry, in particular when considering image analysis. The most studied case is $k = 2$ and separation with lines or circles [1]. Our problem – if we forget the condition that each point must be in a disk – can be considered as a separating problem where each set C_i contains only one point and S is a union of circles. This problem has been mentionned by Gerbener and Tóth [10] who have considered more generally separation with convex sets. They in particular proved that $\lceil n/2 \rceil$ circles are enough to separate n points even if they are in a general configuration (no three colinear points). Separators of single points have also been studied for lines. Bolland and Urrutia [19] gave an algorithm of time complexity $O(n \log n)$ to find a family of $n/2$ lines that separates any set of n points in a general configuration. Călinescu, Dumitrescu and Wan [5] proved that in the particular case where the lines are parallel to the axis, the problem is NP-complete and gave a constant approximation polynomial algorithm for this case. A natural extension in higher dimension, called *multi-modal sensor allocation problem*, has been defined in [14], making links with identification problems. Note that the separating problem with lines is a subproblem of ours. Indeed, if the points are given, one can consider a line as a very large circle.

In Section 2, we give formal definitions and background that will need along the paper. In Section 3, we study particular configurations of points: colinear or forming a grid. For colinear points, we give the exact number of disks needed if any radius can be used. If the points are on a grid, we give exact values for height 2 and bounds for larger heights. In Section 4, we give tight lower and upper bounds: we prove that at least $\Theta(\sqrt{n})$ and at most $\lceil (n+1)/2 \rceil$ disks are necessary (n is the number of points). If moreover there are no three colinear points nor four cocyclic points, then we prove, using Delaunay triangulation, that we need at most $2\lceil n/6 \rceil + 1$ disks. Finally, in Section 5, we discuss the complexity of the problem when the radius is fixed , we prove that it is NP-complete in the general case but that there is a linear algorithm to solve it when the points are colinear.

2 Definition and background

2.1 Formal definition

Let $\mathcal{P}$ be a set of points of $\mathbb{R}^2$. A disk of radius $r \in \mathbb{R}$ and center $c \in \mathbb{R}^2$ is the set of points of $\mathbb{R}^2$ at distance at most r of c. A point $P \in \mathcal{P}$ is *covered by* a disk if it belongs to it. Two points P and Q of $\mathcal{P}$ are *separated* by a disk D if exactly one of them is covered by D. A set of disks $\mathcal{D}$ is *identifying* $\mathcal{P}$ if it is covering all the points of $\mathcal{P}$ and separating all the pairs of points of $\mathcal{P}$. We denote by $\gamma_D^{\text{ID}}(\mathcal{P})$ the minimum number of disks needed to identify $\mathcal{P}$. Let $r \in \mathbb{R}$, we denote by $\gamma_{D,r}^{\text{ID}}(\mathcal{P})$ the minimum number of disks of radius r needed to identify $\mathcal{P}$. When r is large enough compare to the distances between the points of $\mathcal{P}$, any disk of radius r is separating the same pairs of points as some half-plane. Hence, identification with half-planes is a particular case of identification with disks of fixed size. We will denote by $\gamma_{D,\infty}^{\text{ID}}(\mathcal{P})$ the corresponding number.

Remark. In our definition, we ask that every point of $\mathcal{P}$ must be covered by at least one disk. This choice could be discussed. Indeed, it is not the case for similar notions like *separating families* or *test covers*. We choose this definition to be consistent with our first motivation: in a context of localization, our detection system must be able to detect if there is an intrusion or not, which is possible only if all the points are covered. This is the reason why in *identifying codes* there is the condition of domination (see Section 2.2 for formal definition of identifying codes). However, if a set of disks is separating all the pairs of points of $\mathcal{P}$, at most one point is not covered (otherwise all the points that are not covered will not be separated). Therefore, we need at most one more disk to obtain an identifying set. Hence the difference between the two values is at most one and our results can be easily adapted for only-separating sets.

For any radius r and any points P and Q of $\mathbb{R}^2$, there is always a disk of radius r that separates them. Hence, $\gamma_D^{\text{ID}}(\mathcal{P})$ and $\gamma_{D,r}^{\text{ID}}(\mathcal{P})$ are well-defined and always smaller than $\binom{|\mathcal{P}|}{2}$. Moreover, if the radius is not fixed or small enough, one can take for each point a disk containing only this point and then forms an identifying set of disks. Thus, we have $\gamma_D^{\text{ID}}(\mathcal{P}) \leq |\mathcal{P}|$. About the lower bound, consider a set $\mathcal{D}$ of k disks identifying $\mathcal{P}$. Since each point is contained in a unique non-empty subset of $\mathcal{D}$, there are at most $2^k - 1$ points in $\mathcal{P}$, leading to the following lower bound on $\gamma_D^{\text{ID}}(\mathcal{P})$:

Lemma 1. *Let $\mathcal{P}$ be a set of n points of $\mathbb{R}^2$, then $\gamma_D^{ID}(\mathcal{P}) \geq \lceil \log(n+1) \rceil$.*

These trivial lower and upper bounds are not tight and will be improved in Section 4.

Finally, since a set of disks identifying $\mathcal{P}$ is identifying any subset of $\mathcal{P}$, we have the following lemma:

Lemma 2. *Let $\mathcal{P}$ and $\mathcal{P}'$ be two sets of points of $\mathbb{R}^2$ with $\mathcal{P}' \subseteq \mathcal{P}$, then $\gamma_D^{ID}(\mathcal{P}) \geq \gamma_D^{ID}(\mathcal{P}')$.*

2.2 Related work

Among the related notions given in the introduction, we give formal definitions for three of them that we will need in the rest of the paper.

Separating families of disks. If $\mathcal{D}$ is only separating any pair of vertices of $\mathcal{P}$, $\mathcal{D}$ is a *separating family of disks*, studied by Gerbner and Tóth [10] in the more general context of convex sets. They in particular consider the parameter $s(n, \mathcal{D})$ and $s'(n, \mathcal{D})$) which stand for the maximum number of disks that are needed to separate any n-point set and any n-point set in general position (no three of its points are on a line). They prove that $s(n, \mathcal{D}) = s'(n, \mathcal{D}) = \lceil n/2 \rceil$. Since at most one more disk is necessary to obtain an identifying set of disks from a separating set of disks, it means that $\gamma_D^{\text{ID}}(\mathcal{P})$ is at most $\lceil n/2 \rceil + 1$. We will improve this bound in Section 4 to $2\lceil n/6 \rceil + 1$ if moreover no four points are cocyclic.

Identifying codes in unit interval and unit disk graphs. Let $G = (V, E)$ be a graph. A vertex c *dominates* a vertex x if c is in the closed neighbourhood of x (i.e: x and its neighbours). It *separates* two vertices x and y if it is dominating exactly one of them. An *identifying code* of G is a subset of vertices C such that each vertex is dominated by some vertex of C and each pair of vertices of G is separated by some vertex of C. We denote by $\gamma^{\text{ID}}(G)$ the minimum number of vertices in an identifying code of G. Note that $\gamma^{\text{ID}}(G)$ is not always well-defined since G might have two vertices with exactly the same neighbourhood and thus no vertex can separate them. Such vertices are called *twin vertices*. If a graph does not have any twins, then it has an identifying code (take for example all the vertices in C).

Identifying codes are closely related to identifying sets of disks when considering graphs of geometric intersections. Given a set of geometric objects, one can define its intersection graph as follows. Vertices are the objects and there is an edge between two objects if they intersect. A class of graphs of particular interest for us is the class of *unit disk graphs* that are the intersection graphs of disks of radius 1. Let G be a unit disk graph and denote by $\mathcal{P}$ the set of centers of the disks forming G. Then an identifying code of G is equivalent to an identifying set of $\mathcal{P}$ using disks that have radius 2 and are centered on points of $\mathcal{P}$. Indeed, a disk of radius 2 centered on a point P of $\mathcal{P}$ contains all the points that are centers of disks of the closed neighbourhood of the disk corresponding to P in G. Identifying codes in unit disk graphs have been studied by Müller and Sereni [17] who prove, in particular, that the minimization problem in NP-complete. If the points of $\mathcal{P}$ are colinear, then G is a *unit interval graph*. The complexity of identifying codes in unit interval graphs is surprisingly still open [9] (but has been proved to be NP-complete for interval graphs).

Junnila and Laihonen [11] studied identifying codes in the grid $\mathbb{Z}^2$ using Euclidean balls. The underlying graph has the set $\mathbb{Z}^2$ as vertices and the closed neighbourhood are given by the Euclidean balls of a fixed radius r. This graph can also be seen as an (infinite) unit disk graph. They give lower and upper bounds on the density of minimum identifying codes in this graph in function of r.

Identifying codes in hypergraphs. The notion of identifying codes can be extended to hypergraphs. Let $\mathcal{H} = (V, \mathcal{E})$ be a hypergraph. An *identifying code* of $\mathcal{H}$ is a set $C \subseteq \mathcal{E}$ of hyperedges such that:

- each vertex of $\mathcal{H}$ is in at least one element of C;
- for each pair of vertices of $\mathcal{H}$, there is an element of C containing exactly one element of the pair.

An identifying code in a graph G is equivalent to an identifying code in the hypergraph of the closed neighbourhoods of G. As said in the introduction, this notion is known under different names and has actually been introduced before identifying codes in graphs, see [15, 16]. Our problem can be reduced to identifying codes in hypergraph. Indeed, let $\mathcal{P}$ be a set of n points of $\mathbb{R}^2$. Let $\mathcal{H}(\mathcal{P})$ be the hypergraph with vertex set $\mathcal{P}$ and a set of points $E \subseteq \mathcal{P}$ is a hyperdege if there exists a disk D such that $D \cap \mathcal{P} = E$. Then finding an identifying set of disks identifying $\mathcal{P}$ is equivalent to finding an identifying code in $\mathcal{H}(\mathcal{P})$. Note that an hyperedge of $\mathcal{H}(\mathcal{P})$ of size k corresponds to a nonempty cell in the iterated Voronoï diagram of size k of $\mathcal{P}$ and can be computed in $O(n)$ time [18]. The whole hypergraph $\mathcal{H}(\mathcal{P})$ can be obtained by computing all iterated Voronoï diagrams of $\mathcal{P}$. This can be done in time $O(n^3)$ and the number of hyperedges of $\mathcal{H}(\mathcal{P})$ is of order $O(n^3)$ [8].

3 Particular configurations

3.1 Colinear points

When points are located on a single line, the problem is completly solved with the following theorem.

Theorem 3. *Let $\mathcal{P}$ be a set of n colinear points, then $\gamma_D^{ID}(\mathcal{P}) = \lceil \frac{n+1}{2} \rceil$.*

Proof. Let $\mathcal{P}$ be a set of n colinear points located on a line L. We denote by $x_1, ..., x_n$ the points, respecting their order on L.

Let $\mathcal{D}$ be a set of disks identifying $\mathcal{P}$. For any $i \in \{1, ..., n-1\}$, x_i and x_{i+1} are separated by $\mathcal{D}$. It means that there is a disk $D \in \mathcal{D}$, such that its perimeter intersects L between x_i and x_{i+1}. Moreover, x_1 and x_n are covered by $\mathcal{D}$, thus there is a disk whose perimeter intersects L before x_1 and after x_n. In total, there are at least $n+1$ intersections between L and some disks' perimeters. Since a circle intersects a line into at most two points, we necessarily have $|\mathcal{D}| \geq \lceil \frac{n+1}{2} \rceil$.

To prove the equality, note that for any subset of consecutives points $x_i, x_{i+1}, ..., x_j$ of $\mathcal{P}$, there exists a disk $D_{i,j}$ such that $D_{i,j} \cap \mathcal{P} = \{x_i, x_{i+1}, ..., x_j\}$. Then the set of disks

$$\mathcal{D} = \left\{ D_{i,i+\lceil n/2 \rceil} \mid i = 1, .., \left\lceil \frac{n+1}{2} \right\rceil \right\}$$

has size $\lceil \frac{n+1}{2} \rceil$ and is identifying $\mathcal{P}$. See Figure 2 for an illustration with nine points.

□

In the solution given in the proof of Theorem 3, some disks might have a big radius. Actually, if the radius of the disks is bounded by a constant r, n disks are sometimes needed and $\gamma_{D,r}^{\text{ID}}(\mathcal{P})$ can take any value between $\lceil \frac{n+1}{2} \rceil$ and n. In Section 5, we give an algorithm that computes $\gamma_{D,r}^{\text{ID}}(\mathcal{P})$ in linear time.

3.2 Points located on a grid

We now consider points located on a regular grid. Given two integers m and n, we denote by $\mathcal{P}_{m,n}$ the set of points (x, y) of $\mathbb{Z}^2$ such that $1 \leq y \leq m$ and $1 \leq x \leq n$.

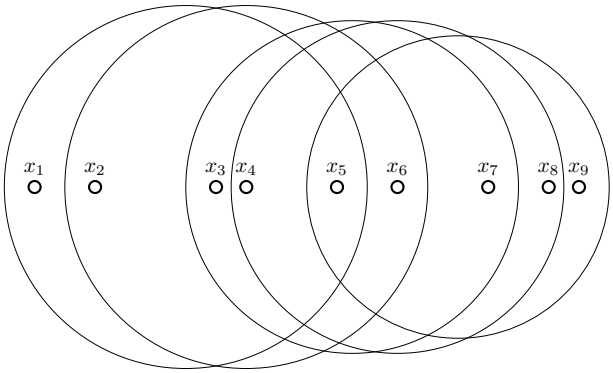

Figure 2: Identification of colinear points

3.2.1 Grids of height 2

When the grid contains only two lines, one can identify the points using the same number of disks than on a single line, except in few cases:

Theorem 4. *Let $n \geq 2$ be an integer. We have:*

$$\gamma_D^{ID}(\mathcal{P}_{2,n}) = \begin{cases} \lceil \frac{n+1}{2} \rceil + 1 & \text{if } n \in \{2,3,4,5,7\}, \\ \lceil \frac{n+1}{2} \rceil & \text{otherwise.} \end{cases}$$

Proof. We can first see that for all n, $\mathcal{P}_{2,n} \leq \lceil \frac{n+1}{2} \rceil + 1$. Indeed, to identify $\mathcal{P}_{2,n}$ one can use the method proposed in Theorem 3 and add an half-plane (which can be seen as a very large disk) to separate the lines as in Figure 3.

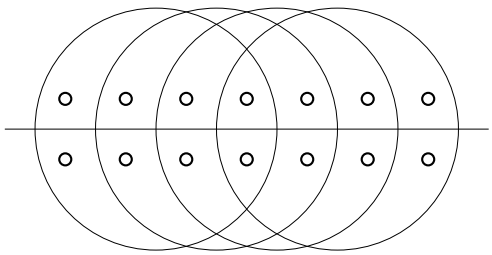

Figure 3: Identification of $\mathcal{P}_{2,n}$ with $\lceil \frac{n+1}{2} \rceil + 1$ disks

For grids $\mathcal{P}_{2,n}$ with $n \leq 5$, this solution is optimal by Lemma 1. In Section 4 Proposition 7, we show that at least five disks are needed to identify a set of 14 points, so there is no better solution for $\mathcal{P}_{2,7}$. For all the other cases, we show that we only need $\lceil \frac{n+1}{2} \rceil$ disks. We only have to study the case where n is odd or equal to 6. Indeed, by Lemma 2, solution for $P_{2,2q+1}$ is also a solution for $P_{2,2q}$ by removing the points of the last column.

We first give a characterization for a set $X \subseteq \mathcal{P}_{2,n}$ to be the intersection of $\mathcal{P}_{2,n}$ and a disk. Let X be such a set, X is the union of two sets of consecutive points of the first line $(a,1)$, ... , $(b,1)$ and of the second line $(c,2)$, ... , $(d,2)$, with $a,b,c,d \in \mathbb{N}$, $a \leq b$ and $c \leq d$. We must have either $[a,b] \subseteq [c,d]$ or $[c,d] \subseteq [a,b]$ and the difference between each extremities must differ of at most 1: $|(c-a)-(b-d)| \leq 1$. This condition is sufficient since for every a,b,c,d verifying this relation, there exist a disk $\mathcal{D}_{[a,b]}^{[c,d]}$ that contains exactly these consecutive points.

An explicit solution for the grids $P_{2,6}$, and $P_{2,9}$ are the following disks :

- $P_{2,6}$ can be identified by the set of disks : $\mathcal{D}_{[1,5]}^{[3,4]}$, $\mathcal{D}_{[3,6]}^{[4,5]}$, $\mathcal{D}_{[2,3]}^{[1,5]}$ and $\mathcal{D}_{[4,4]}^{[2,6]}$.

- $P_{2,9}$ can be identified by the set of disks : $\mathcal{D}^{[3,4]}_{[1,6]}$, $\mathcal{D}^{[4,6]}_{[2,9]}$, $\mathcal{D}^{[6,7]}_{[4,8]}$, $\mathcal{D}^{[1,8]}_{[3,4]}$ and $\mathcal{D}^{[2,9]}_{[6,7]}$.

We now give a solution for grids $\mathcal{P}_{2,4p+1}$, with $p \geq 3$. This solution use three different steps. Figure 4 gives an illustration of these three steps.

The first step is to use the disks $\mathcal{D}_1 = \mathcal{D}^{[p+2,2p]}_{[1,3p+1]}$, $\mathcal{D}_2 = \mathcal{D}^{[1,3p+1]}_{[p+2,2p]}$, $\mathcal{D}_3 = \mathcal{D}^{[2p+2,3p]}_{[p+1,4p+1]}$ and $\mathcal{D}_4 = \mathcal{D}^{[p+1,4p+1]}_{[2p+2,3p]}$. After adding these disks, the points of each line are separated from the other line. Indeed, the points of the first line in the intervals $[1, p+1]$ and $[2p+1, 3p+1]$ are in the disk $\mathcal{D}_1$ and are not in the disk $\mathcal{D}_2$ which separate them from all the points of the second line. Similarly the points of the first line and in the intervals $[p+2, 2p]$ and $[3p+2, 4p+1]$ are in the disk $\mathcal{D}_3$ and are not in the disk $\mathcal{D}_4$, which separate them from all the points of the second line.

In the second step, we add the disks $\mathcal{D}^{[p,p+2]}_{[p,p+2]}$ and $\mathcal{D}^{[3p,3p+2]}_{[3p,3p+2]}$. These two disks separate the points on the columns $p+1$, $2p+1$ and $3p+1$, which weren't until now.

After this, all the points are covered by at least one disk and the points that are no separated from each other are the same line and on the intervals $[1, p-1]$, $[3p+3, 4p+1]$, $[p+3, 2p]$ and $[2p+2, 3p-1]$ (the last two intervals occurs if $p \geq 4$).

In the third step, we can now finish identifying the points by adding the following concentric disks : $\mathcal{D}^{[2,4p]}_{[2,4p]}$, $\mathcal{D}^{[3,4p-1]}_{[3,4p-1]}$, ... , $\mathcal{D}^{[p-1,3p+3]}_{[p-1,3p+3]}$, $\mathcal{D}^{[p+4,3p+2]}_{[p+4,3p-2]}$, $\mathcal{D}^{[p+5,3p-3]}_{[p+5,3p-3]}$, ... , $\mathcal{D}^{[2p,2p+2]}_{[2p,2p+2]}$.

We use four disks in the first part, then two disks and finally $(p-2)+(p-3)$. So in total we use $2p+1$ disks, which is equal to $\frac{(4p+1)+1}{2}$ disks.

For the grids $\mathcal{P}_{2,4p-1}$, we can remove the points of the columns 1 and $4p+1$ and the disk $\mathcal{D}^{[2,4p]}_{[2,4p]}$.

So we can indeed identify the grids $\mathcal{P}_{2,n}$, with $n \geq 10$ with $\frac{n+1}{2}$ disks.

□

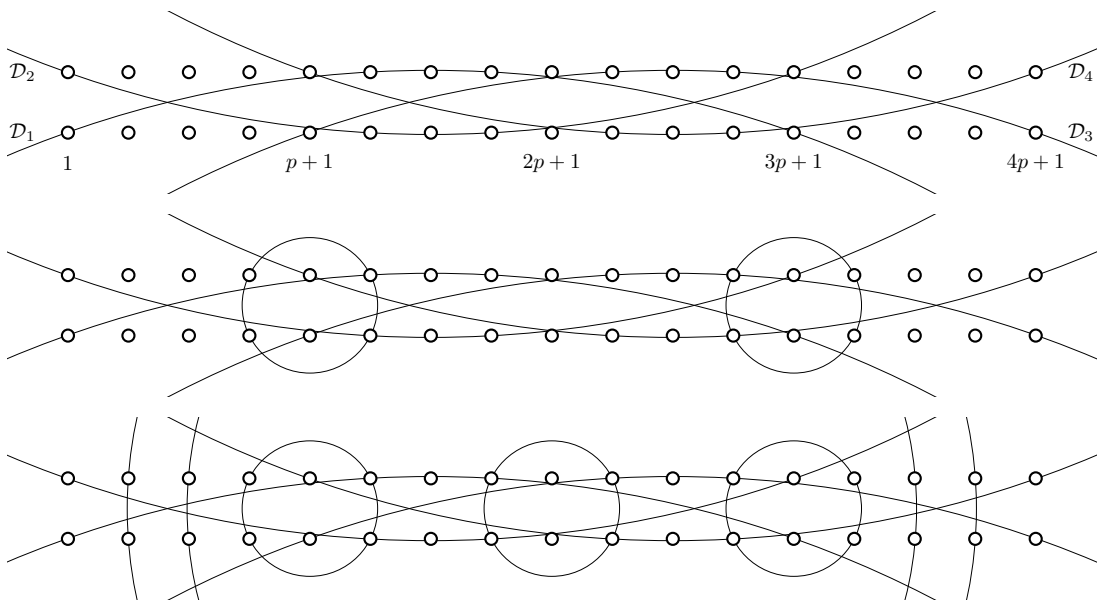

Figure 4: The three steps of the proof the grid $\mathcal{P}_{2,4p+1}$

3.2.2 General case

We now consider the general case of grids $m \times n$, $n \geq m \geq 3$. We first solve the case of identification with half-planes - which can be considered as disks with infinite radius.

Theorem 5. *Let $m, n \geq 3$ be two integers. Then, $\gamma^{ID}_{D,\infty}(\mathcal{P}_{m,n}) = m+n-2$.*

Proof. We denote by $x_1, ..., x_{2(m+n-2)}$ the points on the convex hull of $\mathcal{P}_{m,n}$, respecting their order.

Let $\mathcal{L}$ be a set of half-planes identifying $\mathcal{P}_{m,n}$. For any $i \in \{1, ..., 2(m+n-2)\}$, x_i and x_{i+1} are separated by $\mathcal{L}$ (with $x_{2(m+n-2)+1}$ associated with x_1). It means that the there is a half-plane $L \in \mathcal{L}$

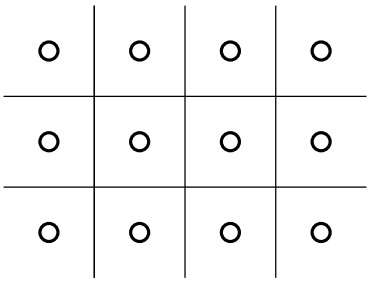

Figure 5: Identification of the edges of a grid using half-planes

whose boundary line intersects the convex hull of $\mathcal{P}_{m,n}$ between x_i and x_{i+1}. In total, there are at least $2(m+n-2)$ intersections between the convex hull of $\mathcal{P}_{m,n}$ and boundary lines of some half-planes of $\mathcal{L}$. Since a line intersects a convex polygon into at most two points, we necessarily have $|\mathcal{L}| \geq m+n-2$.

Consider any $m-1$ vertical lines between adjacent points and $n-1$ horizontal lines between adjacent points (see Figure 5 for an example). Then every pair of points is separated by a line. To obtain a solution, one just need to choose half-planes with these lines as boundary and in such a way that every point is covered. □

This theorem gives a bound for the general case: $\gamma_D^{\text{ID}}(\mathcal{P}_{m,n}) \leq n+m-2$. This bound is not tight, especially when n is large enough compared to m. Next theorem gives a better (but still not tight) bound in this case:

Theorem 6. *Let n and m be two integers such that $m \geq 3$ and $n \geq \frac{m^2}{2}-3$. Then $\gamma_D^{ID}(\mathcal{P}_{n,m}) \leq \lceil \frac{n}{2} \rceil + m - 1$.*

Proof. The idea is to use a method similar to the one described in Figure 3. We use half-planes to separates the lines and disks to separates the columns. When n is large enough the disks act on each line in the same way they act when there is only one line.

Since $m \geq 3$ we use half-planes to separate the lines and include all the points into a disk: the bottom half-plane includes all the points above itself and the top half-plane includes all the points below itself.

We now use disks of radius $\sqrt{\left(\frac{1}{2}\lceil \frac{n}{2} \rceil\right)^2 + \frac{m^2}{4}}$ and centered on $(\frac{1}{2}(\lceil \frac{n}{2} \rceil + 1) + k, m/2)$ with k an integer between 0 and $\lceil \frac{n}{2} \rceil - 1$. Since $n \geq \frac{m^2}{2} - 3$, those disks contains $\lceil \frac{n}{2} \rceil$ points on each line and they separates all the columns. An example of such disks can be seen in Figure 6. Since all the points are inside a half-plane, there is no need for all the columns to be inside a disk. That is why we only need $\lceil \frac{n}{2} \rceil$ disks instead of $\lceil \frac{n+1}{2} \rceil$ as in the case of one line.

There is $\lceil \frac{n}{2} \rceil$ disks to separate the columns and $m-1$ half-planes to separate the lines, this gives us $\lceil \frac{n}{2} \rceil + m - 1$ in total.

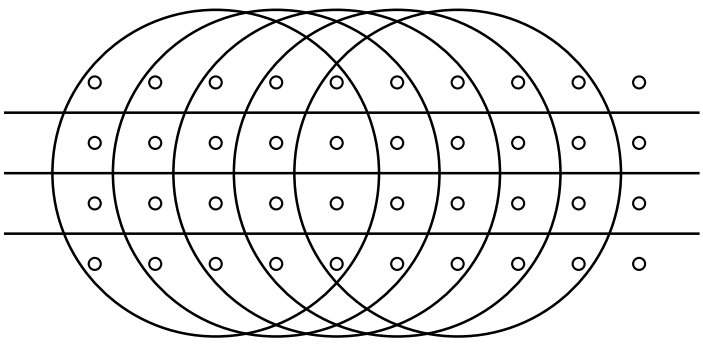

Figure 6: Example of identification of points of a grid with n sufficiently greater than m

□

4 Extremal cases

In this section, we give tight lower and upper bounds on $\gamma_D^{\text{ID}}(\mathcal{P})$ using the number of points of $\mathcal{P}$ when the points are in general configuration.

4.1 Lower bound

The logarithmic lower bound given in Lemma 1 is the natural lower bound for identifying codes in hypergraphs. It is tight if any hyperedge is allowed. But if there is some structure on the hyperedges, this is not always true. In particular, if the hyperedge set has bounded dual VC-dimension d^*, then the lower bound is at least of order n^{1/d^*} [3]. This is the case for our problem since the hypergraph induced by disks have bounded dual VC-dimension equals to 3, leading to a lower bound of order $n^{1/3}$. However, this bound is still not tight. Indeed, we provide in this section a lower bound of order $n^{1/2}$. This bound comes from the fact that an arrangment of k disks can create at most $k^2 - k + 1$ inner faces. This classical result can be proved by induction with the argument that each time one add a circle to a set of circles it cross each circle at most twice (see [24] for more details and references). Since if some disks are identifying a set of points, there is at most one point in each faces of the intersections of the disks, we have the following bound:

Proposition 7. *Let $\mathcal{P}$ be a set of n points of $\mathbb{R}^2$. Then, $\gamma_D^{ID}(\mathcal{P}) \geq \left\lceil \frac{1+\sqrt{1+4(n-1)}}{2} \right\rceil$. This bound is tight.*

To obtain a set $\mathcal{P}$ of n points reaching the bound, one can use an arrangment of k disks making $k^2 - k + 1$ inner faces and set one point in each face. Such an arrangment can be obtained by taking disks of radius $1 + \epsilon$, centered on vertices of a k-regular polygon that is inscribed in a circle of radius 1. See Figure 7 for a construction with k=5.

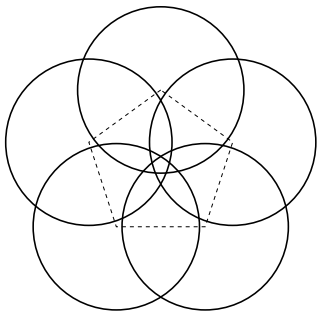

Figure 7: Five disks creating $21 = 5^2 - 5 + 1$ inner faces.

Since an identifying code of a unit disk graph can be seen as an identifying set of special disks, the lower bound of Property 7 is still true for identifying codes in unit disk graphs, improving the bound given in [4]. Moreover, this bound is also tight for this case. Indeed, the construction of Property 7 can be adapted for identifying codes of unit disk graphs since all the disks have the same radius and their center can be points.

4.2 Upper bound

We now consider the worse configurations of points. Otherwise said, what is the minimum number of disks that is enough to identify any set of n points? This question has already been solved by Gerbner and Tóth when one just wants to separate points [10]. They prove that $\lceil n/2 \rceil$ disks are always enough and that this is the best value one can obtain since there are point sets needing this number of disks. Since one more disk is enough to obtain an identifying set, it gives us the bound $\lceil n/2 \rceil + 1$. Actually, we can slightly improve it by noticing that in the proof of Gerbner and Tóth [10], all the points are covered if there is an odd number of points. For the sake of completeness we give the proof, it follows the one of [10].

Proposition 8. *Let $\mathcal{P}$ be a set of n points of $\mathbb{R}^2$. Then, $\gamma_D^{ID}(\mathcal{P}) \leq \lceil \frac{n+1}{2} \rceil$. This bound is tight.*

Proof. Since $\mathcal{P}$ is finite there exists a direction we can choose as abscissa, such that no pair of points have the same abscissa. Then there is a line L of constant abscissa which separates the points in two parts of the same size up to one. Let D be the half-plane defined by L and containing the biggest part of points as one of the disks. We choose a direction perpendicular to the abscissa as the ordinate.

At first, all the points are part of a set P' and the set of identifying disks contains only D. Then we repeat the following operation $\lfloor \frac{n}{2} \rfloor$ times. Consider the convex hull of P', exactly two of its edges cross L. Let (x, y) be the edge which intersects L on the largest ordinate. Since it is an edge of the convex hull, there is no point of P' with a larger ordinate than x and y. Therefore there is a disk $D_{x,y}$ that contains only x and y among the points of $\mathcal{P}'$. Add this disk to the set of identifying disks and remove x and y from P'. Iterate the process.

This algorithm gives a set of disks that identifies P of size $1 + \lfloor \frac{n}{2} \rfloor = \lceil \frac{n+1}{2} \rceil$. Indeed, at each step, x and y are separated from all the other points of P' by $D_{x,y}$ and since all the other points have been considered previously they are also separated from them. Moreover, since (x, y) is an edge that crosses L, x and y are separated from each other by D. At the end, if there is an odd number of points, there is a point that is only in D, since this is the only point that is only in D, it is identified.

Figure 8 illustrates some steps of the algorithm. This bound is tight when all the points are colinear (see Theorem 3). □

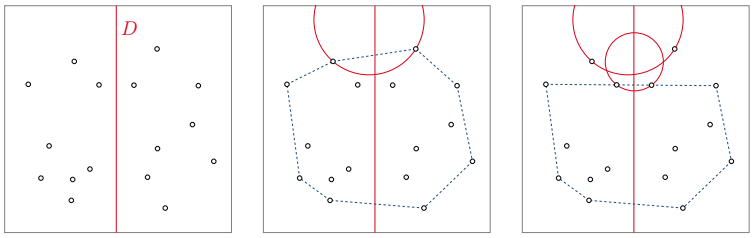

Figure 8: First steps of the algorithm for the general upper bound

All the values between the lower bound of Property 7 and the upper bound of Property 8 are reached:

Theorem 9. *Let $n \in \mathbb{N}$ and $k \in \mathbb{N}$ be such that $\lceil \frac{1+\sqrt{1+4(n-1)}}{2} \rceil \leq k \leq \lceil \frac{n+1}{2} \rceil$.*
There exists an n-point set $\mathcal{P}$ of $\mathbb{R}^2$ such that $\gamma_D^{ID}(\mathcal{P}) = k$.

Proof. Consider the optimal arrangement of k disks based on a regular polygon given on Figure 7. There is a line L cutting this construction into $2k - 1$ regions. Indeed, let L' be a line going through an intersection of disks and the center of the polygon, for symmetry reason, this line goes through k regions and $k - 1$ intersections. So by shifting the line infinitesimally and parrallely, it is still going through the previous regions but for each intersection we have a new region. Therefore, there is indeed $2k - 1$ regions crossed by this new line L.We set $2k - 1$ points on L in the different regions of the arrangement of disks. Since $k \geq \lceil \frac{1+\sqrt{1+4(n-1)}}{2} \rceil$ we can set the $n - (2k - 1)$ remaining points in the other regions of the arrangement. At the end, the set of k disks of the arrangment identifies the n points that we have put in its different regions and, since there are $2k - 1$ colinear points, no smaller set of disks can identify these points.

□

4.3 Improved upper bound for general configurations

The upper bound of Proposition 8 is tight for colinear points. A natural question is whether the bound is still tight if there are no three colinear points among $\mathcal{P}$. Actually, the bound is also tight if the points are cocyclic. But, if there are no three colinear points nor four cocyclic points in $\mathcal{P}$, the upper bound is not tight anymore. In this section, we said that a set of points of $\mathbb{R}^2$ is in *general configuration* if there are no three points of $\mathcal{P}$ on a line nor four points of $\mathcal{P}$ on a circle.

Theorem 10. *Let $\mathcal{P} \subseteq \mathbb{R}^2$ be a set of n points in general configuration. Then $\gamma_D^{ID}(\mathcal{P}) \leq 2\lceil n/6 \rceil + 1$.*

The idea of the proof of this theorem is to give an algorithm that constructs an identifying set of disks of size $2\lceil n/6 \rceil + 1$. The algorithm is based on the same principle that we used in the not restricted case:

1. Divide $\mathcal{P}$ in three equal parts using lines;
2. Choose a disk that contains exactly one point in each part, remove these points and repeat the operation.

The crucial part is to find the disk of the second step. For that, we use Delaunay triangulations - that is a triangulation of the points in such a way that the circumcircle of each triangle only contains its vertices. Since there are no three colinear nor four cocyclic points, such a triangulation always exists (and is unique). To find the disk of Step 2, we then need to find a Delaunay triangle that has a vertex in each part, as illustrated in Figure 9. To insure the existence of such a triangle, we need to be more precise at Step 1.

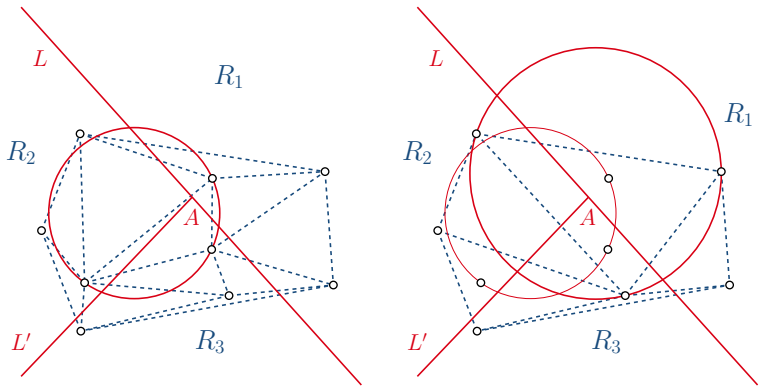

Figure 9: First steps of the method used in the main idea of the proof

Before going into details, we need two preliminary results.

Theorem 11 (Ceder [6])**.** *For n points of $\mathbb{R}^2$ with no three colinear points, there is a way to divide the plan in six regions containing each between $\lceil \frac{n}{6} \rceil - 1$ and $\lceil \frac{n}{6} \rceil$ points using three concurrent lines.*

Lemma 12. *Let $\mathcal{P}$ be a set of points of $\mathbb{R}^2$, L a line and L' a half-line with origin A on L. If each of the three regions R_1, R_2 and R_3 made by L and L' contains one point of $\mathcal{P}$ and if A is in the convex hull of $\mathcal{P}$, then every triangulation of $\mathcal{P}$ contains a triangle that has a vertex in each region.*

Proof. Let T be a triangulation of $\mathcal{P}$. Since the intersection A of L and L' is inside the convex hull, there is at least a segment of T between any pair of regions.

Consider the segments between the regions separated by L', namely R_2 and R_3. Let $[x, y]$ be the segment which cut L' the closest from A. Since A is in the convex hull of $\mathcal{P}$, there is at least one point z of $\mathcal{P}$ such that (x, y, z) is a triangle of T in the direction of A from this segment. If z is in R_2 or R_3, then the segment $[x, z]$ or $[y, z]$ would intersects L' closer to A than $[x, y]$, which contradicts the hypothesis that $[x, y]$ is the closest segment to A. Therefore z is in R_1 and (x, y, z) form a triangle with one vertex in each region.

□

Proof of Theorem 10. Using Theorem 11, there exist three concurrent lines L_1, L_2 and L_3 that divides the plane into six regions of the same size up to one. Let A be their common intersection. Let D_1, D_2 and D_3 be three half-planes defining by L_1, L_2 and L_3 such that every point is in at least one half-plane. Let a, b, c, d, e and f be the six regions of the plane created by these lines, as illustrated in Figure 10.

First consider the regions a, c and e. Each of these regions contains between $\lceil \frac{n}{6} \rceil - 1$ and $\lceil \frac{n}{6} \rceil$ points. For construction reasons, if there is a point in each region then A is inside the triangle formed by these points, so Lemma 12 always applies.

Consider the following process. Add D_1, D_2 and D_3 to the future set of identifying disks. Set in P' all the points of a, c and e. Then repeat the following operation $\lceil \frac{n}{6} \rceil - 1$ times. Consider the Delaunay triangulation of P', at least one triangle (x, y, z) has its vertices in the three different regions. Since it is a triangle of a Delaunay triangulation, its circumscribed circle C contains no other remaining points. Add $D_{x,y,z}$, the disk of perimeter C, to the set of identifying disks, remove x, y and z from P' and iterate the process.

We do the same iterated operation for the regions b, d and f.

At the end of each step of the process all the considered points are separated from all the other points. Indeed, at each step x, y and z are separated from the points of the other regions by the half-planes and, since each considered triangle comes from a Delaunay triangulation, their circumscribed circle contains no other points of P'.

If a point has not been considered at the end of the processes, it is alone in its regions and therefore is isolated from all the other points. Moreover, by the selection of the half-planes, it is inside at least one half-plane so it is covered.

Therefore, this algorithm constructs an identifying set of disks of size $3 + 2(\lceil \frac{n}{6} \rceil - 1) = 2\lceil \frac{n}{6} \rceil + 1$. $\square$

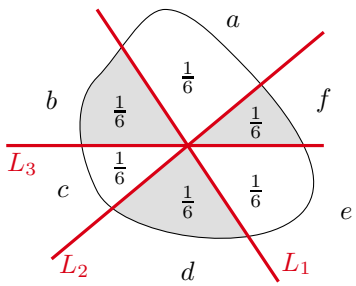

Figure 10: Dividing the points into six equal parts

The previous bound is tight, up to a constant 2, when points are located on an half-parabola, the curve constituted of one side of a parabola symmetry axis:

Proposition 13. *Let $\mathcal{P} \subseteq \mathbb{R}^2$ be a set of n points placed on an half-parabola. Then, $\gamma_D^{ID}(\mathcal{P}) \geq \frac{n}{3}$.*

Proof. Let $\mathcal{P}$ be a set of n points located on a half-parabola H. We denote by $x_1, ..., x_n$ the points, respecting their order on H, with x_1 the closest to the extrema of H.

Let $\mathcal{D}$ be a set of disks identifying $\mathcal{P}$. For any $i \in \{1, ..., n-1\}$, x_i and x_{i+1} are separated by $\mathcal{D}$. It means that there is a disk $D \in \mathcal{D}$ whose perimeter intersects H between x_i and x_{i+1}. Moreover, x_n is covered by $\mathcal{D}$, thus there is a disk whose perimeter intersects H after x_n. In total, there are at least n intersections between H and some perimeters of disks of $\mathcal{D}$.

We now prove that a circle C can intersect H into at most three points. Let $(x, y) \in C \cap H$. Without loss of generality, (x, y) satisfies the following set of equations, with x_0, y_0, r that are constant.

$$\begin{cases} y = x^2 \\ (x - x_0)^2 + (y - y_0)^2 = r^2 \\ x \geq 0 \end{cases}$$

In particular, x is a solution of:

$$(X - x_0)^2 + (X^2 - y_0)^2 = r^2 \tag{1}$$

There is no term in X^3 in the previous equation. Thus, if x_1, x_2, x_3 and x_4 are solutions of (1, we have $x_1 + x_2 + x_3 + x_4 = 0$. Since $x \geq 0$, there are at most three possible values for x.

Since there are at least n intersections between H and an identifying set of disks $\mathcal{D}$ and since a circle intersects a half-parabola at most three times, we necessarly have $\mathcal{D} \geq \frac{n}{3}$.

$\square$

5 Complexity when the radius is fixed

In this section, we consider the complexity of the following decision problems (with $r \in \mathbb{R}$):

Identification-Disk(r)
Instance: A finite set $\mathcal{P} \subseteq R^2$, an integer k.
Question: Is it true that $\gamma_{D,r}^{\text{ID}}(\mathcal{P}) \leq k$?

Colinear Identification-Disk(r)
Instance: A finite set $\mathcal{P} \subseteq R^2$ of colinear points, an integer k.
Question: Is it true that $\gamma^{\text{ID}}_{D,r}(\mathcal{P}) \leq k$?

Theorem 14. Identification-Disk*(r) is $\mathcal{NP}$-complete.*

Proof. We prove the result for $r = 1/2$ which is not restrictive. We reduce this problem from the problem of partitioning a grid graph into path on three vertices. A *grid graph* is a graph with vertex set included in $\mathbb{Z}^2$ and two vertices are adjacent if they are at Euclidean distance 1.

P_3-Partition-Grid
Instance: A grid graph G.
Question: Is there a partition of the vertices of G in such a way that each part induces a path on three vertices?

Bevern *et al.* [26] proved that this problem is $\mathcal{N}P$-complete.

Let G be an instance of P_3-Partition-Grid and $n = |V(G)|$. The instance of the identification problem is $\mathcal{P} = V(G)$ and $k = 2n/3$. We have to prove that G has a P_3-partition if and only if $V(G)$ is identified by $2n/3$ disks of radius $1/2$.

Assume first that there is a partition of G into path on three vertices. For each part (x, y, z) of the P_3 partition, add to the identifying set the disk $D_{x,y}$ of radius $1/2$ that contains x and y and the disk $D_{y,z}$ that contains y and z. Since (x, y) and (y, z) are edges of G, these disks exist and contains exactly two points. Furthermore, x is the only point that is contained only in $D_{x,y}$, z is the only point that is contained only in $D_{y,z}$ and y is the only point that is contained exactly in $D_{x,y}$ and $D_{y,z}$. Hence, we obtain an identifying set of disks of size $2n/3$.

Assume now that there is an identifying set of disks $\mathcal{D}$ that identify $V(G)$ with $2n/3$ disks. Since the points are at distance at least 1, every disk contains at most two points. Without loss of generality, we can suppose that if a disk contains only one point, then this point is not included in any other disk. Indeed, assume there are two points x, y and two disks D_1 and D_2 such that D_1 contains only x and D_2 contains both x and y, then we can replace D_2 by D_2' that contains only y and the situation is similar, $V(G)$ is still identified and the number of disks is the same.

Let a be the number of disks containing only one point. Let V' be the $n-a$ points not covered by these a disks. Let G' be the graph with vertices V' and edges (x, y) if x, y are contained together in a disk of $\mathcal{D}$. Note that G' is a subgraph of G. The graph G' has $n - a$ vertices and its connected components have at least three vertices. Indeed, if a component as only two vertices then these vertices are not identified. So there are $k \leq (n - a)/3$ connected components. We name these connected components $\{G_1, ..., G_k\}$. The number of edges of G' is $\sum_{i=1}^{k} E(G_i) \geq \sum_{i=1}^{k}(V(G_i) - 1) \geq (n - a) - k \geq 2(n - a)/3$.

Since a disk is either containing one point (and there are a such disks) or corresponds to an edge of G', there are at least $a + 2(n - a)/3 = 2n/3 + 2a/3$ disks in $\mathcal{D}$. Therefore we necessarily have $a = 0$ and there are exactly $n/3$ connected components in G', each of them being of size 3. This is a P_3-partition of G. □

However if all of the points are colinear then this problem can be solved in linear time:

Theorem 15. Colinear Identification-Disk*(r) can be solved in linear time.*

Note that if the disks are required to be centered on the points, this problem is equivalent to the problem of identifying codes in unit interval graphs, whose complexity is surprisingly still open.

To prove Theorem 15, we introduce few definitions and preliminary results. Let $\mathcal{P}$ be a set of n colinear points on a line L and $\mathcal{D}$ is a set of disks of radius r identifying $\mathcal{P}$. Let x_1,..x_n be the points of $\mathcal{P}$. We confuse x_i with its abscissa on L and we assume that $x_1 < ... < x_n$.

Note that, since the points are colinear and the centers of the points can be chose anywhere on the plane, a set of points of $\mathcal{P}$ can be the intersection of $\mathcal{P}$ and a disk of radius r if and only if there are consecutive points of $\mathcal{P}$ at distance at most $2r$. In the following, we will refer often directly to the set of points contained in a disk D instead of D itself.

The set $\mathcal{D}$ is *optimal* if $|\mathcal{D}| = \gamma_D^{\text{ID}}(\mathcal{P})$, it is *perfect* if n is odd and if $|\mathcal{D}| = \frac{n+1}{2}$ (in particular $\mathcal{D}$ is optimal).

The disks of $\mathcal{D}$ partitions the points of $\mathcal{P}$ on connected components. More formally, we define an equivalence relation $x \sim_{\mathcal{D}} y$ meaning that x and y are connected by a path of disks. For x and y two points of $\mathcal{P}$, we have $x \sim_{\mathcal{D}} y$ if and only if there is a disk D in $\mathcal{D}$ that contains both x and y, or there is a point z in $\mathcal{P}$ such that $x \sim_{\mathcal{D}} z$ and $y \sim_{\mathcal{D}} z$.

The equivalence classes $(\mathcal{P}_i)$ of $\sim_{\mathcal{D}}$ are made of consecutive points. Let $\mathcal{D}_i = \{D \in \mathcal{D} | D \cap \mathcal{P}_i \neq \emptyset\}$ be the disks containing points the points of $(\mathcal{P}_i)$.

A set of disks $\mathcal{D}$ is *piece-wise perfect* if it is optimal and if each $\mathcal{D}_i$ perfectly identifies $\mathcal{P}_i$.

Lemma 16. *For any set of colinear points $\mathcal{P}$, there is a set $\mathcal{D}$ of disks of radius r that identifies $\mathcal{P}$ and is piece-wise perfect.*

Proof. Let $\mathcal{D}$ be a set of disks that identifies optimally $\mathcal{P}$ and such that $\sum_{D \in \mathcal{D}} |D \cap \mathcal{P}|$ is minimal. We will prove that this set is piece-wise perfect.

Assume the contrary. It means that there is a set $\mathcal{P}_i$ which is not perfectly identified by $\mathcal{D}_i$. Following the proof of Theorem 3, this means that there are two disks D_1 and D_2 whose perimeters intersect L between the same pair of adjacent points of $\mathcal{P}_i$ or both before the first point of $\mathcal{P}_i$ or both after the last point of $\mathcal{P}_i$. Let $x_a, ..., x_b$ be the points covered by D_1 and $x_c, ..., x_d$ the points covered by D_2.

Case 1 : $a = c$ (the case $b = d$ is similar). Suppose, without loss of generality, that $d \leq b$. Let D_1' be a disk that contains the points from $x_a + 1$ to x_b, such a disk exist because its intersection with $\mathcal{P}$ is included in $D_1 \cap \mathcal{P}$. Then $\mathcal{D}' = \mathcal{D} \setminus \{D_1\} \cup \{D_1'\}$ identifies $\mathcal{P}$. Indeed, the only point of $\mathcal{P}$ for whom the situation is different for $\mathcal{D}$ and $\mathcal{D}'$ is x_a. For $\mathcal{D}'$ it is the only point of $\mathcal{P}$ that is inside D_2 and not inside D_1'. So we have a new set of disks that identifies $\mathcal{P}$ and such that the sum of the number of points contained in each disk is smaller. This is a contradiction to the minimal property of $\mathcal{D}$.

Case 2 : $c = b + 1$ (the case $a = d + 1$ is similar). Since $\mathcal{P}_i$ is an equivalence class for the relation $\sim_{\mathcal{D}}$, we must have $x_b \sim_{\mathcal{D}} x_{b+1}$ and there must be a disk D_3 such that D_3 contains both x_b and x_{b+1}. Let x_e the first point of D_3 and x_f its last point.

Subcase 2.1 : $a < e < b < f < d$.

Let D_1' be a disk that contains the point from x_a to x_{b-1}, such a disk can exist because it is smaller than D_1, and let $\mathcal{D}' = \mathcal{D} \setminus \{D_1\} \cup \{D_1'\}$. $\mathcal{D}'$ identifies $\mathcal{P}$ and contradict the minimal property of $\mathcal{D}$.

Subcase 2.2 : $e < a$ (the case $f > d$ is similar).

Let D_1' be a disk that contains the point from x_e to x_b, such a disk can exist because it is smaller than D_3 and D_3' be a disk that contains the points from x_a to x_f, such a disk can exist because it is smaller than D_3. The set of disks $\mathcal{D}' = \mathcal{D} \setminus \{D_1, D_3\} \cup \{D_1', D_3'\}$ identifies $\mathcal{P}$. Indeed, the only points of $\mathcal{P}$ for whom the situation is different for $\mathcal{D}$ and $\mathcal{D}'$ are those between x_e and x_{a-1}. They are still separated from each other by the disks that separates them in $\mathcal{D}$ and they are separated from the other points because they are the only points that are in D_1' and not in D_3'. The sum of the number of points contained in the disk is the same for $\mathcal{D}'$ and $\mathcal{D}$.

Finally, the sum of the number of points contained in the disks remains the same, but now the disk that contains both x_b and x_{b+1} does not contain the disks that intersect the region between x_b and x_{b+1}. So we are now in Subcase 2.1 and we can apply the method used in this case, concluding the proof. □

A set of disks $\mathcal{D}$ identifying a set $\mathcal{P}$ of n colinear points is in *normal form* if n is odd and if :

- $n = 1$ and $\mathcal{D}$ is composed of a unique disk containing the point of $\mathcal{P}$.

  or

- $n = 2p+1$, $p \geq 1$, and $\mathcal{D} = \{D_i\}_{i \in [0,p]}$ with D_0 containing the points x_1 and x_2, D_p containing the points x_{2p} and x_{2p+1} and, for $i \in [1, p-1]$, D_i containing the points x_{2i}, x_{2i+1} and x_{2i+2}.

In particular, $\mathcal{D}$ is perfect.

Lemma 17. *For any set of colinear points $\mathcal{P}$, if there is a set of disks that perfectly identifies $\mathcal{P}$, then there exists a set $\mathcal{D}$ that identifies $\mathcal{P}$ and is in normal form.*

Proof. If there is only one point in $\mathcal{P}$, then the only way to identify perfectly $\mathcal{P}$ is to have a set $\mathcal{D}$ that contains exactly one disk which contains the point of $\mathcal{P}$, it always exists and it is already in normal form.

Suppose that $\mathcal{P}$ is of size $2p+1$ with $p \geq 1$. We show that if there is no set of disks identifying $\mathcal{P}$ in normal form then there is no set perfectly identifying $\mathcal{P}$.

Assume that there is no set of disks identifying $\mathcal{P}$ in normal form but that a set $\mathcal{D}$ perfectly identifies $\mathcal{P}$. Necessarily, $\sim_{\mathcal{D}}$ has only one equivalnce class and all the adjacent points of $\mathcal{P}$ are distant at most $2r$.

Since there is no possible set in normal form, there exists $i \in [1, p-1]$ such that the distance between the points x_{2i} and x_{2i+2} is greater than $2r$. Let $\mathcal{P}_1$ be the set $\{x_1, ..., x_{2i}\}$ and $\mathcal{P}_2$ be the set $\{x_{2i+2}, ..., x_{2p+1}\}$. Let $\mathcal{D}_1$ (respectively $\mathcal{D}_2$) be the subset of disks of $\mathcal{D}$ that contains at least one point of $\mathcal{P}_1$ (resp. $\mathcal{P}_2$). The intersection between $\mathcal{D}_1$ and $\mathcal{D}_2$ is empty since the distance between x_{2i} and x_{2i+2} is at least $2r$. By Theorem 3, since $\mathcal{D}_1$ identifies $\mathcal{P}_1$ and $\mathcal{D}_2$ identifies $\mathcal{P}_2$, $|\mathcal{D}_1| \geq \lceil \frac{2i+1}{2} \rceil$ and $|\mathcal{D}_2| \geq \lceil \frac{2(p-i)+1}{2} \rceil$. So $|\mathcal{D}| \geq |\mathcal{D}_1| + |\mathcal{D}_2| \geq \lceil \frac{2i+1}{2} \rceil + \lceil \frac{2(p-i)+1}{2} \rceil = p+2$. Hence $\mathcal{D}$ does not identify $\mathcal{P}$ perfectly, a contradiction. □

Proof of Theorem 15. By Lemma 16, there exist a set identifying $\mathcal{P}$ that is piece-wise perfect. By Lemma 17, every perfect part of that set can be identified by a set of disks in normal form. So there is a piece-wise perfect set of disks $\mathcal{D}$ such that every $\mathcal{D}_i$ is in normal form.

We now give an algorithm that finds an optimal solution to identify $\mathcal{P}$ with connected sets of disks that are in normal form :

Require: the abscissas $x_1, ..., x_n$ of a set of colinear points $\mathcal{P}$
Ensure: $\mathcal{D}$ is a minimal identifying set that can be partitioned into subset in normal form.

```
i ← 0
D ← ∅
x_{n+1} ← ∞, x_{n+2} ← ∞, x_{n+3} ← ∞
while i ≤ n do
  if x_{i+1} − x_i > 2r or x_{i+2} − x_{i+1} > 2r then
    Add to D a disk that contains only x_i
    i ← i + 1
  else
    Add to D a disk that contains only x_i and x_{i+1}
    i ← i + 1
    while x_{i+2} − x_i ≤ 2r and x_{i+3} − x_{i+2} ≤ 2r do
      Add to D a disk that contains only x_i, x_{i+1} and x_{i+2}
      i ← i + 2
    end while
    Add to D a disk that contains only x_i and x_{i+1}
    x_i ← x_{i+2}
  end if
end while
```

This algorithm takes the biggest connected sets of disks in normal form starting with the first point of $\mathcal{P}$. We prove that this is optimal. Assumme it is not the case. Let $\mathcal{P}$ be a set of points such that the set $\mathcal{D}$ given by the algorithm is not an optimal solution. We choose $\mathcal{P}$ with a minimum number of points. Let $\mathcal{D}^{opt}$ be an optimal set in normal form. Its first connected component $\mathcal{D}_1^{opt}$ is smaller than the first connected component of $\mathcal{D}$, $\mathcal{D}_0$. Indeed, it cannot be bigger since the algorithm take the biggest connected component and it cannot be the same since, by minimality of $\mathcal{P}$, the algorithm is optimal on the rest of the points. So $\mathcal{D}^{opt}$ identifies $\mathcal{P}_1^{opt}$, the points of $\mathcal{P}$ that are not in the disks of $\mathcal{D}_1^{opt}$ with less disks than $\mathcal{D}$ uses to identify $\mathcal{P}_1$, the points of $\mathcal{P}$ that are not in $\mathcal{D}'$. Since $\mathcal{P}_1 \subset \mathcal{P}_1^{opt}$, $\mathcal{D}_{opt}$ also identifies $\mathcal{P}_1$, and thus with less disks than $\mathcal{D}$. This contradicts the minimality of $\mathcal{P}$.

This algorithm is linear since we consider each point at most once.

So there is a linear algorithm to find the maximum number of disks needed to identify a set of points if each connected part must be in normal form. By Lemma 16 and Lemma 17, this algorithm also gives a solution to Colinear Identification-Disk(r). □

6 Conclusion

We conclude with some open problems. About complexity issues, we do not know if computing a minimum identifying set of disks when the radius is not fixed is $\mathcal{NP}$-complete, but the contrary would be surprising. The complexity of identification with lines seems to be also open. An intersecting question is what is the number of disks needed if the points are randomly chosen in a 1×1 square. It would also be interesting to consider identifications with other sets or in higher dimensions using balls instead of disks.

References

[1] Esther M Arkin, Ferran Hurtado, Joseph SB Mitchell, Carlos Seara, and Steven Skiena. Some separability problems in the plane. In *EuroCG*, pages 51–54, 2000.

[2] László Babai. On the complexity of canonical labeling of strongly regular graphs. *SIAM Journal on Computing*, 9(1):212–216, 1980.

[3] Laurent Beaudou, Florent Foucaud, Peter Dankelmann, Michael A Henning, Arnaud Mary, and Aline Parreau. Bounding the order of a graph using its diameter and metric dimension: a study through tree decompositions and vc dimension. *arXiv preprint arXiv:1610.01475*, 2016.

[4] Nicolas Bousquet, Aurélie Lagoutte, Zhentao Li, Aline Parreau, and Stéphan Thomassé. Identifying codes in hereditary classes of graphs and vc-dimension. *SIAM Journal on Discrete Mathematics*, 29(4):2047–2064, 2015.

[5] Gruia Călinescu, Adrian Dumitrescu, Howard Karloff, and Peng-Jun Wan. Separating points by axis-parallel lines. *International Journal of Computational Geometry & Applications*, 15(06):575–590, 2005.

[6] Jack G Ceder. Generalized sixpartite problems. *Bol. Soc. Mat. Mexicana (2)*, 9:28–32, 1964.

[7] Olivier Devillers, Ferran Hurtado, Merce Mora, and Carlos Seara. Separating several point sets in the plane. In *In Proc. 13th Canadian Conference on Computational Geometry*. Citeseer, 2001.

[8] Herbert Edelsbrunner, Joseph O'Rourke, and Raimund Seidel. Constructing arrangements of lines and hyperplanes with applications. *SIAM Journal on Computing*, 15(2):341–363, 1986.

[9] Florent Foucaud, George B Mertzios, Reza Naserasr, Aline Parreau, and Petru Valicov. Identification, location-domination and metric dimension on interval and permutation graphs. ii. algorithms and complexity. *Algorithmica*, pages 1–31, 2016.

[10] Dániel Gerbner and Géza Tóth. Separating families of convex sets. *Computational Geometry*, 46(9):1056–1058, 2013.

[11] Ville Junnila and Tero Laihonen. Identification in $\mathbb{Z}^2$ using euclidean balls. *Discrete Applied Mathematics*, 159(5):335–343, 2011.

[12] Mark G Karpovsky, Krishnendu Chakrabarty, and Lev B Levitin. On a new class of codes for identifying vertices in graphs. *IEEE Transactions on Information Theory*, 44(2):599–611, 1998.

[13] Alexander Kogan. On the essential test sets of discrete matrices. *Discrete Applied Mathematics*, 60(1-3):249–255, 1995.

[14] Farinaz Koushanfar, Sasha Slijepcevic, Miodrag Potkonjak, and Alberto Sangiovanni-Vincentelli. Error-tolerant multimodal sensor fusion. In *IEEE CAS Workshop on Wireless Communication and Networking*, pages 5–6, 2002.

[15] Julien Moncel. *Codes identifiants dans les graphes*. PhD thesis, Université Joseph-Fourier-Grenoble I, 2005.

[16] Bernard ME Moret and Henry D Shapiro. On minimizing a set of tests. *SIAM Journal on Scientific and Statistical Computing*, 6(4):983–1003, 1985.

[17] Tobias Müller and J-S Sereni. Identifying and locating–dominating codes in (random) geometric networks. *Combinatorics, Probability and Computing*, 18(06):925–952, 2009.

[18] Joseph O'rourke, S Rao Kosaraju, and Nimrod Megiddo. Computing circular separability. *Discrete & Computational Geometry*, 1(1):105–113, 1986.

[19] Boland R. P. and J. Urrutia. Separating families of points on the plane. *Information Processing Letters*, 53:177–183, 1995.

[20] RW Payne and DA Preece. Identification keys and diagnostic tables: A review. *Journal of the Royal Statistical Society. Series A (General)*, pages 253–292, 1980.

[21] Suk J Seo and Peter J Slater. Open neighborhood locating–dominating in trees. *Discrete Applied Mathematics*, 159(6):484–489, 2011.

[22] Peter J. Slater. Dominating and reference sets in a graph. *Journal of Mathematical and Physical Sciences*, 22(4):445–455, 1988.

[23] Peter J. Slater and Douglas F. Rall. On location-domination numbers for certain classes of graphs. *Congressus Numerantium*, 45:97–106, 1984.

[24] Neil James Alexander Sloane. The On-Line Encyclopedia of Integer Sequences, Sequence A014206. `http://oeis.org/A014206`.

[25] Rachanee Ungrangsi, Ari Trachtenberg, and David Starobinski. An implementation of indoor location detection systems based on identifying codes. In *Intelligence in Communication Systems*, pages 175–189. Springer, 2004.

[26] René Van Bevern, Robert Bredereck, Laurent Bulteau, Jiehua Chen, Vincent Froese, Rolf Niedermeier, and Gerhard J Woeginger. Star partitions of perfect graphs. In *International Colloquium on Automata, Languages, and Programming*, pages 174–185. Springer, 2014.