Transversals of Geometric Objects and Anagram-Free Colouring

Saman Bazargani

Thesis submitted to the University of Ottawa
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in
Computer Science

School of Electrical Engineering and Computer Science
Faculty of Engineering
University of Ottawa

Examining Committee

The following served on the Examining Committee for this thesis.

External Member: Anthony Bonato
Professor, Department of Mathematics
Toronto Metropolitan University

Carleton Member: Anil Maheshwari
Professor, School of Computer Science
Carleton Univeristy

Internal Member(s): Jean-Lou De Carufel
Associate Professor, School of Electrical Engineering & Computer Science
University of Ottawa

Robert Laganière
Professor, School of Electrical Engineering & Computer Science
University of Ottawa

Supervisor(s): Prosenjit Bose
Professor, School of Computer Science
Carleton University

Vida Dujmović
Professor, School of Electrical Engineering & Computer Science
University of Ottawa

Pat Morin
Professor, School of Computer Science
Carleton University

Declaration of Authorship

This integrated thesis is made up of three articles. Two of the articles have already been peer-reviewed and published, as listed below.

- Saman Bazargani, Ahmad Biniaz and Prosenjit Bose.
  Piercing Pairwise Intersecting Convex Shapes in the Plane.
  LATIN 2022: Theoretical Informatics - 15th Latin American Symposium.
  Pages: 679–695.
  [doi:10.1007/978-3-031-20624-5_41].

- Saman Bazargani, Paz Carmi, Vida Dujmović, and Pat Morin.
  $2 \times n$ Grids Have Unbounded Anagram-free Chromatic Number.
  Electronic Journal of Combinatorics, 29(3), 2022.
  [arxiv:2105.01916][doi:10.37236/10411].

The third article is yet to be published.

- Saman Bazargani and Pat Morin.
  Geodesic Anagram-free Colouring of Chordal Graphs.
  Pre-print.

As is typical in theoretical computer science, the authors are listed in alphabetical order by family name. Though it is difficult to pinpoint the exact source of any idea that appears in a multi-author paper, I made major contributions leading to the results in each of these three articles.

Abstract

This PhD thesis is comprised of 3 results in computational geometry and graph theory.

In the first paper, I demonstrate that the piercing number of a set S of pairwise intersecting convex shapes in the plane is bounded by $O(\alpha(S))$, where $\alpha(S)$ is the fatness of the set S, improving upon the previous upper-bound of $O(\alpha(S)^2)$.

In the second article, I show that anagram-free vertex colouring of a $2 \times n$ square grid requires a number of colours that increases with n. This answers an open question in Wilson's thesis and shows that even graphs of pathwidth 2 do not have anagram-free colouring with a bounded number of colours.

The third article is a study on the geodesic anagram-free chromatic number of chordal and interval graphs. *Geodesic anagram-free chromatic number* is defined as the minimum number of colours required to colour a graph such that all shortest paths between any pair of vertices are coloured anagram-free. In particular, I prove that the geodesic anagram-free chromatic number of a chordal graph G is $32p'w$, where p' is the pathwidth of the subtree intersection representation graph (tree) of G, and w is the clique number of G. Additionally, I prove that the geodesic anagram-free chromatic number of an interval graph is bounded by $32p$, where p is the pathwidth of the interval graph.

Dedication

I extend my heartfelt dedication to my thesis work to my beloved family.

I want to express my deepest gratitude to my beloved parents, Simin and Majid, whose unwavering support has been the foundation of my journey.

Mom, your strong and resilient spirit has been my anchor, teaching me to face challenges with determination and to treat others with kindness. Your words of encouragement have boosted my confidence, guiding me to give my best in every endeavor. During tough times, you've instilled in me the importance of discipline and the bravery to persevere through adversity.

Dad, your wisdom and constant encouragement have been my guiding stars. Your consistent presence has given me the assurance that I can tackle any challenge. Your unwavering support has been a driving force in my personal growth.

To my dear sister Sara, your unwavering support has been a continuous source of strength and comfort.

Furthermore, I dedicate this thesis to my friends and colleagues who have been by my side throughout this entire doctoral program. I am sincerely thankful for your support.

With warm regards, Saman

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Structure of the Thesis

This thesis is article-based. It is comprised of three articles, one in computational geometry and two in graph theory. Chapter 1 has three sections, each giving a brief motivation behind the problems studied in these three articles and presenting the obtained results. Chapters 2, 3 and 4 are the full copies of these three articles. The thesis ends with directions for some future work.

1.1 Article 1: Piercing Pairwise Intersecting Convex Shapes in the Plane

In this subsection, I outline the background and terminology required to understand the results in the first article. Then, I talk about the problems related to this article. The article itself can be found in Chapter 2.

Given a set C of convex shapes in the plane, what is the smallest integer k for which it is possible to partition C into k sets $C_1, \ldots, C_k$ such that the shapes in each C_i have a non-empty common intersection for every i, $1 \leq i \leq k$? We call k the *transversal number or piercing number* of the set C. A set of points P *pierces* a set C, if any element of C has a non-empty intersection with P.

In 1923, Helly [24] proved that if the intersection of any $d + 1$ shapes of a set of convex shapes in d-dimensions is non-empty, then the intersection of all of them is non-empty.

This result is known as the *Helly's Theorem*. For example, if H is a set of convex shapes in $\mathbb{R}^2$, such that every three of them have a common intersection, then, by Helly's theorem, all shapes in H have a common intersection. Thus, it is equivalent to saying that the piercing number of H is 1.

Helly's theorem in $\mathbb{R}^2$ requires that every triple of convex shapes intersect. It is easy to observe that "every triple" cannot be replaced by "every pair". Consider, for example, the easily verifiable fact that a set of n pairwise intersecting line segments in general position cannot be pierced with less than $\lceil \frac{n}{2} \rceil$ points. This gives rise to the following natural question: can a set of pairwise intersecting shapes be pierced with one, or a constant number of points, if we restrict possible shapes in some way. Bárány et al. [5] proved that the Helly number[1] of the Cartesian product of intervals is 2. This implies that the piercing number of a set of pairwise intersecting axis-parallel rectangles is 1, since every axis-parallel rectangle is the Cartesian product of 2 intervals. To pierce a set of pairwise intersecting homothetic[2] triangles, three points are sufficient, as shown by Chakerian et al. (1967) [13]. In the case of pairwise intersecting disks, four points are sufficient to pierce a set of pairwise intersecting disks. The proof of the existence of four piercing points was independently demonstrated by Danzer (1956, 1986) [14] and Stacho (1981) [43, 44].

As observed above, pairwise intersecting skinny objects (like line-segments) may require an unbounded number of piercing points, while pairwise intersecting fat objects (like disks) may be pierced by a constant number of points. This naturally leads to the following question: what role does the geometry of a shape, its roundness or fatness, play in the number of points needed to pierce the set?

This problem was considered by Agarwal et al. [2], Nielsen [37], Overmars et al. [38] and Katz [28]. The main approach used to pierce a set of pairwise intersecting fat convex shapes is to construct a grid of points whose resolution is quadratic in the fatness parameter α [2, 28, 37, 38]. In the first article, we were able to reduce the number of points required to pierce a set of pairwise intersecting fat convex shapes from quadratic to linear with respect to the fatness parameter. We achieved this by placing points near the perimeter of a convex shape that has a non-empty intersection with every other shape in the set. In essence, we show that it is possible to pierce the set by focusing on the perimeter of a shape as opposed to filling an area with points.

The details of this approach are given in Chapter 2. There, we look at various families

[1] For a family of sets F, the Helly number $h(F)$ of F is the minimal positive integer h such that if a finite subfamily $K \subset F$ satisfies $\cap K' \neq \emptyset$ for all $K' \subset K$ of cardinality $\leq h$, then $\cap K \neq \emptyset$

[2] An object B is *homothetic* to another object A if B can be obtained by scaling and translating the object A.

of pairwise intersecting convex shapes in two-dimensional Euclidean space. Define a shape to be α-fat when the ratio of the radius of the smallest disk that encloses the shape over the radius of the largest disk that is enclosed in the shape is at most α. Let C be a set of shapes. Define $\alpha(C)$ to be the minimum value where each shape in C is $\alpha(C)$-fat.

The following results are proved in Chapter 2.

- Any set C of pairwise intersecting arbitrary convex shapes with fatness $\alpha(C)$ can be pierced with at most $39.788\alpha(C) + 4 \in O(\alpha(C))$ points.
- Any set C of pairwise intersecting rectangles of arbitrary orientation with fatness $\alpha(C)$, can be pierced by $(5\sqrt{2} + 2)\alpha(C) + 25 + 5\sqrt{2} \in O(\alpha(C))$ points.
- Any set of pairwise intersecting convex homothets[3] can be pierced by 15 points.
- A set of pairwise intersecting homothets of regular hexagons can be pierced by 2 points.

1.2 Article 2: $2 \times n$ Grids have Unbounded Anagram-Free Chromatic Number

In this subsection, I provide the background and terminology required to understand the results in my second article and then I present the result. The article itself can be found in Chapter 3. My second and third contributions are related to a special kind of graph colouring.

A *graph*, $G = (V, E)$, consists of a set of *vertices* V and a set E of pairs of distinct vertices called *edges*. If $\{v, w\}$ is an edge then the vertices v and w are neighbours. Graph theory provides a versatile and powerful framework for modeling and analyzing complex relationships and structures in various fields. Graphs have been extensively studied. See, for example, the graph theory book authored by Diestel [15]. A *(proper) vertex colouring* of a graph is a function that assigns one colour to each vertex, such that neighbouring vertices receive distinct colours. Graph colouring is a topic of central importance in graph theory, both historically and in applications. The objective in graph colouring is to minimise the number of colours used. A famous example is the 4-colour theorem [4]. It states that the regions of any map can be coloured with four colours, such that neighbouring regions

[3] A set C of shapes is called *homothets*, if every pair of shapes in C is homothetic.

receive distinct colours. Graph colouring is one of the most well-known and well-studied topics in graph theory. See, for example, survey [40] and the 511 references there.

Numerous types of graph colourings have been studied in the literature. The focus in my second article is on so-called anagram-free colouring. A sequence of characters with an even length is called an *anagram* if, for every character c in the sequence, the number of occurrences of c in the first half of the sequence is equal to the number of occurrences of c in the second half of the sequence. A sequence s is *anagram-free* if every even-length subsequence of s is not an anagram. In 1961, Erdős [18] asked if arbitrarily long anagram-free sequences over an alphabet of size 4 exist. In 1968, Evdokimov [19, 19] proved the existence of arbitrarily long anagram-free sequence over an alphabet of size 25 and in 1971, Evdokimov [41] showed that an alphabet of size 5 is sufficient. Erdős' question was not fully resolved until 1992 when Keränen [30] answered it in the affirmative.

A notion of anagram-free sequences can be extended to graphs as follows. For a graph G, a k-vertex colouring $\Xi : V(G) \longrightarrow \{1, \ldots, k\}$ is *anagram-free* if the sequence formed by the colours assigned to the vertices along any even-length path in G is not an anagram. The minimum integer k for which G has an anagram-free colouring is called the *anagram-free chromatic number* of G.

To present the results obtained in both the second and the third article of this thesis, the notions of chordal graphs, interval graphs, treewidth and pathwidth need to be introduced next. Let X be a set of sets. The *intersection graph* G of X is the graph G with vertex set $V(G) = X$ and $vw \in E(G)$ if and only if $v \cap w \neq \emptyset$. When X is a set of real intervals the intersection graph of X is called an *interval graph*. Let $\mathcal{T}$ be a tree and let X be a set of subtrees of $\mathcal{T}$. Then, the intersection graph of X is called a *chordal graph*. Given any graph G, the *treewidth* of G is the minimum k such that G is a subgraph of a chordal graph with the maximum clique size $k + 1$. Given any graph G, the *pathwidth* of G is the minimum k such that G is a subgraph of interval graph with the maximum clique size $k+1$. Both treewidth and pathwidth are extensively studied notions of central importance in Robertson–Seymour's graph structure theory.

A corollary of Keränen's work [30] is that the anagram-free chromatic number of any path is 4; that is, it is bounded by a constant. In contrast, Wilson and Wood [47] showed that there exists trees, even binary trees, with arbitrary large anagram-free chromatic number. On the positive side, Wilson and Wood [47] showed that trees of pathwidth p have anagram-free chromatic number $4p + 1$. This gives rise to the following natural question: what are graph classes in which anagram-free chromatic number can be bounded by their pathwidth?

In 2018 Carmi et al. [11] proved that there are planar graphs[4] of pathwidth 3 with arbitrarily large anagram-free chromatic number. Since paths have pathwidth 1, the result of Wilson and Wood [47] and the result of Carmi et al. [11] leave as unresolved the question if graphs of pathwidth 2 have bounded anagram-free chromatic number? We answer this question in the negative by showing that a $2 \times n$ grid graph (that graph has pathwidth 2) has an unbounded anagram-free chromatic number.

The full article can be found in Chapter 3.

1.3 Article 3: Geodesic Anagram-Free Chromatic Number

In this subsection, I provide the background and terminology required to understand the results in my third article and then I present the result. This article is a follow up of the second article. The article itself can be found in Chapter 4.

For a graph G, let $\mathrm{pw}(G)$ denote the pathwidth of G. As mentioned in the previous section, Wilson and Wood [49] proved that the anagram-free chromatic number of a tree T is upper-bounded by $4\,\mathrm{pw}(T) + 1$. Their proof is inductive and relies on the existence of a path P such that $\mathrm{pw}(T - V(P)) < \mathrm{pw}(T)$. This observation allowed them to colour each tree in $T - V(P)$ anagram-freely using $4\,\mathrm{pw}(T) - 3$ colours, while colouring the path P with 4 distinct colours that differ from those used in the remainder of the tree. Observe that the path connecting any two vertices $u, v \in V(T)$ either avoids any vertex of P, hence, the path is anagram-free by induction or, it can enter and exit P at most once. That together with the fact that the colours assigned to P are distinct from those used in the rest of the tree gives angram-free colouring of the whole tree T with the claimed number of colours.

It is well known and easy to verify that every tree T is an intersection graph of subtrees of the same tree. Thus every tree is a chordal graph and in fact a graph with the maximum clique size 2. This points to one possible direction for generalizing the above result. Namely, one may wonder if the result can be extended to all chordal graphs with bounded maximum clique size, that is, if it is possible to prove that such graphs have anagram-free chromatic number bounded by their pathwidth. However, that is not possible by the previously mentioned result by Carmi et al. [11], as well as the result in the second article of this thesis. In particular, consider again $2 \times n$ grid graph G. G has pathwith 2 and thus it is well known that it is possible to add edges to G and obtain a graph G' that is an

[4] A graph is *planar* if it can be drawn (in two-dimensional space) without edge crossings.

interval (and thus a chordal) graph of pathwidth 2. Thus G' is a chordal graph of bounded pathwidth with maximum clique size 3. If G' had a bounded anagram-free chromatic number so would G, contradicting our result in the second article.

Notice next that, not only is every tree a chordal graph but every path in a tree is a shortest path (that is, a geodesic path). That observation leads to another possible direction for generalization of Wilson and Wood's [49] result on trees. Namely, one may wonder if the result can be extended in the following direction: is it true that for every chordal graph G with a bounded maximum clique number, where G is an intersection graph of a bounded pathwidth tree, can be coloured in a way that ensures the colouring of every geodesic path in G is anagram-free?

Specifically, a k-vertex colouring $\Xi : V(G) \longrightarrow \{1, \ldots, k\}$ of a graph G is *geodesic anagram-free* if the sequence of colours corresponding to any shortest path between any two vertices in $V(G)$ is anagram-free. The minimum integer k for which G has a geodesic anagram-free colouring is called the *geodesic anagram-free chromatic number* of a G.

In the third article, we show that the geodesic anagram-free chromatic number of an interval graph is bounded by $32p$ where p is the pathwidth of the interval graph. Then, based on Wilson and Wood's approach, we generalize this result to chordal graphs. In particular, we prove that the geodesic anagram free chromatic number of a chordal graph G is $32p'w$ where p' is the minimum pathwidth of a tree used in any subtree intersection representation of G and w is the clique number of G. This result generalizes the theorem by Wilson and Wood [49].

The full article can be found in Chapter 4.

Chapter 2

Piercing a Set of Pairwise Intersecting Fat Objects in The Plane

PIERCING PAIRWISE INTERSECTING CONVEX SHAPES IN THE PLANE[1]

Saman Bazargani[2] Ahmad Biniaz [3] Prosenjit Bose[4]

[1]This research was partly funded by NSERC.

ABSTRACT. Let C be any family of pairwise intersecting convex shapes in a two dimensional Euclidean space. Let $\tau(C)$ denote the piercing number of C, that is, the minimum number of points required such that every shape in C contains at least one of these points. Define a shape to be α-fat when the ratio of the radius of the smallest disk that encloses the shape over the radius of the largest disk that is enclosed in the shape is at most α. Define $\alpha(C)$ to be the minimum value where each shape in C is $\alpha(C)$-fat. We prove that $\tau(C) \approx 39.788\alpha(C) = O(\alpha(C))$ for any set C consisting of pairwise intersecting convex α-fat shapes. This improves the previous best known upper-bound of $O(\alpha(C)^2)$. This result has a number of implications on other results concerning fat shapes, such as designing data structures with less complexity for 3-D vertical ray shooting and computing depth orders. Additionally, our results reduce the time complexity of the query time of these data structures. We also get better bounds for some restricted families of shapes. We show that $(5\sqrt{2}+2)\alpha(C)+25+5\sqrt{2} \approx 9\alpha(C)+32 = O(\alpha(C))$ piercing points are sufficient to pierce a set of arbitrarily oriented α-fat rectangles. We also prove that $\tau(C) = 2$ when C is a set of pairwise intersecting homothets of regular hexagons. We show that the piercing number of a set of pairwise intersecting homothets of an arbitrary convex shape is at most 15. This improves the previous best upper-bound of 16. We also give an algorithm to calculate the exact location of the piercing points.

2.1 Introduction

Let H be a set of convex shapes in $d-$dimensions such that every subset of $d+1$ shapes in H has a non-empty intersection. In 1923, Helly [24] proved that the intersection of all shapes in H is non-empty. This result is known as the *Helly's theorem*. For example, if H is a set of convex shapes in $\mathbb{R}^2$ such that every three of them have a common intersection, then by Helly's theorem all shapes in H have a common intersection. In the other words, all shapes in H can be pierced with one point.

Consider the following fundamental geometric problem: What is the minimum number of points that is sufficient to pierce a given set of pairwise intersecting shapes in the plane? In the case of homothetic triangles, three points are sufficient, as was shown by Chakerian et al. (1967) [13]. In the case of disks, four points are sufficient. The proof of the existence of four piercing points was independently shown by Danzer (1956, 1986) [14] and Stacho (1981) [43, 44]. To pierce a set of pairwise intersecting line segments, $\Omega(n)$ points are sometimes required. This huge gap between the number of points required, from a constant to linear, to pierce different sets of pairwise intersecting shapes gives rise to many interesting problems. Notice that the linear lower-bound to pierce a set of pairwise intersecting line segments comes from the fact that line segments are essentially "thin". How round or fat an object is plays a vital role in the number of points needed to pierce the set. The main problem that we study in this paper is the following: How many points are sufficient to pierce a set of pairwise intersecting shapes in terms of their fatness parameter?

In the literature, the main approach used by researchers to pierce a set C of pairwise intersecting α-fat shapes is by constructing a grid whose resolution is quadratic in the fatness parameter [2,28,37,38]. In this article, we are able to reduce the number of points to linear with respect to the fatness parameter by placing points near the perimeter of a shape that has a non-empty intersection with every other shape in the set. In essence, we show that it is possible to pierce the set by focusing on the perimeter of an object as opposed to filling an area with points. The details of our approach are given in Section 2.2.

2.1.1 Preliminary

Informally the fatness of a shape is a parameter that tries to capture how close a shape is to a disk. There are many different definitions and variations of the fatness of a shape [1, 17, 29, 35, 38, 39, 46]. Most of them share some similarities. In this paper we use the following measure of fatness. The fatness of a shape c is the ratio of the radius of the smallest disk that encloses c over the radius of the largest disk that is enclosed in c. This measure of fatness will be denoted by α. We say that a shape c is α-fat if its fatness is at most α. A set C of shapes is referred to as $\alpha(C)$-fat if $\alpha(C)$ is the smallest value such that $\forall c_i \in C$, the fatness of c_i is greater than or equal to $\alpha(C)$. We note that a set of disks is 1-fat, since a disk is perfectly fat according to our fatness definition.

The *piercing number* of a family of sets $\mathcal{F}$ is the smallest integer k for which it is possible to partition $\mathcal{F}$ into subfamilies $\mathcal{F}_1, \ldots, \mathcal{F}_k$ such that the sets in each $\mathcal{F}_i$ have a non-empty intersection for every $1 \leq i \leq k$ [21]. We say that a set of points P pierces a set of shapes C if every shape in C contains at least one point of P.

A shape B is a *homothet* of a shape A if B can be obtained by scaling and translating the shape A. Two geometric figures are *homothetic* if one is a homothet of the other. A set C of shapes is called *homothets*, if every pair of shapes in C is homothetic. A set of shapes C is *unit* if all the shapes in C have the same area.

2.1.2 Our Contributions

In this paper we prove the following results in 2-dimensions:

- Any set C of pairwise intersecting arbitrary convex shapes with fatness $\alpha(C)$ can be pierced with $\approx 39.788\alpha(C) \in O(\alpha(C))$ points.
- Any set C of pairwise intersecting rectangles of arbitrary orientation with fatness $\alpha(C)$, can be pierced by $(5\sqrt{2}+2)\alpha(C) + 25 + 5\sqrt{2} \in O(\alpha(C))$ points.
- Any set of pairwise intersecting convex homothets can be pierced by 15 points.
- A set of pairwise intersecting homothets of regular hexagons can be pierced by 2 points.

Table 2.1: Known results for piercing sets of pairwise intersecting convex sets.

Family of convex shapes	Known results	Our results
Homothetic Triangles	3 Points [13]	_
Homothetic Rectangles	1 Point [folklore]	_
Homothetic regular Hexagons	Not known	2 Points, Theorem 2.5
Disks	4 Points [12, 14, 43, 44]	_
Centrally symmetric	7 Points [22]	_
Unit Shapes	3 Points [26]	_
Convex Homothets	16 [32]	15 Points, Theorem 2.3
α-fat Rectangles	$O(\alpha^2)$ [2, 28, 37, 38]	$\simeq 9\alpha + 32$, Theorem 2.2
α-fat Convex shapes	$O(\alpha^2)$ [2, 28, 37, 38]	$\simeq 39.788\alpha$, Theorem 2.1

2.1.3 Previous Results

Overmars et al. (1994) [38] proved that for a set of disjoint convex α-fat objects and a restricted range query (with diameter $h \times p$ where h is a constant and p is the radius of the minimal enclosing hyper-sphere among the objects in the set) in d-dimensions,

$O((\alpha d^d h)^d)$ points are enough to pierce all the shapes. They use a grid of points inside and around the range query to pierce such a set. Agarwal et al. (1995) [2], Katz (1996) [28] and Nielsen (2000) [37] among other results proved that $O(\alpha^2)$ points can pierce a set of pairwise intersecting α-fat shapes in 2-dimension. The definitions of fatness that they use are similar. The unifying theme among these proofs is to cover the area around and inside the smallest shape with a grid of $\Theta(\alpha^2)$ piercing points. To find the piercing points Nielsen (2000) [37] uses Fredman's sampling technique [28]. Agarwal et al. (1995) [2], Katz (1996) [28] use a similar gridding technique.

Let F be a family of pairwise intersecting and centrally symmetric convex homothets. Grünbaum (1959) [22] showed that $\tau(F) \leq 7$. He transforms all the shapes from Euclidean space into Minkowski space. The reason behind this transformation is that any centrally symmetric shape in Euclidean space can be treated as a disk in Minkowski space. This transformation maintains the pairwise intersecting property of the set. The fact that 4 points pierces a set of pairwise intersecting disks applies [12, 14, 43, 44]. Grünbaum (1959) [22] also showed that $\tau(F) = 3$ when F is a family of pairwise intersecting and centrally symmetric convex unit-shapes. He conjectured that $\tau(F) = 3$ for any family of pairwise intersecting convex unit-shapes. This conjecture was proved by Karasev (2000) [26]. Karasev (2001) [27] subsequently showed an upper-bound of $d + 1$ on the number of points sufficient to pierce a family of d-wise intersecting homothets of a simplex in $\mathbb{R}^d$. He also gave an upper-bound for a family F of d-wise intersecting spheres which is the following: $\tau(F) \leq 3(d + 1)$ when $d \geq 5$ and $\tau(F) \leq 4(d + 1)$ when $d \leq 4$.

In case of pairwise intersecting disks, Danzer (1956, 1986) [14] and Stacho (1981) [43,44] were the first to give a proof of the existence of 4 piercing points. However, both of their proofs are essentially non-constructive. Har-Peled et al. [23] were the first to present a deterministic and constructive algorithm. They find 5 piercing points, in $O(n)$ expected time, that pierces a set of n pairwise intersecting disks. Biniaz, Bose and Wang [7] gave a linear algorithm that finds 5 piercing points given a set of pairwise intersecting disks that does not use a LP-*type* framework unlike Har-Peled's algorithm. Carmi, Katz and Morin [12] gave a linear time algorithm to compute 4 piercing points which also uses LP-*type* machinery.

2.2 General Convex Shapes

2.2.1 Piercing a Set of fat Shapes

In this section we prove the following theorem which is our main result.

Theorem 2.1. *Any set C of pairwise intersecting arbitrary convex shapes on a plane with fatness $\alpha(C)$ can be pierced with $39.788\alpha(C) \in O(\alpha(C))$ points.*

Proof of Theorem 2.1. Let $S = \{S_0, S_1, \ldots, S_{n-1}\}$ be a set of pairwise intersecting convex shapes with fatness at most α. For any $S_i \in S$, we define α_i as the fatness of S_i. Let o be the smallest disk that has a non empty intersection with every shape in the set S. Let δ be the radius of o. Define sq_1 to be an axis-parallel square that is concentric with o. Let the side length of sq_1 be $2c\delta$ for a constant c. And let sq_2 be an axis-parallel square concentric with o with side length $2c_1\delta$ for a constant c_1 ($c_1 > c$). We specify the exact values of c and c_1 at the end of the proof.

If all the shapes in S have a common intersection we can pierce the whole set with one point and as a result o will have radius zero. Otherwise o is tangent to at least three shapes, say S_1, S_2, S_3. Let L_1, L_2, L_3 be the three tangent lines to o where S_1, S_2, S_3 intersect o. Note that L_1, L_2, L_3 form a triangle. Suppose they do not form a triangle, and two of them are parallel, denoted as L_1 and L_2. Consequently, S_1 and S_2 are located on opposite sides of L_1 and L_2, which contradicts the assumption that S_1 and S_2 have a non-empty intersection. (See Figure 2.2).

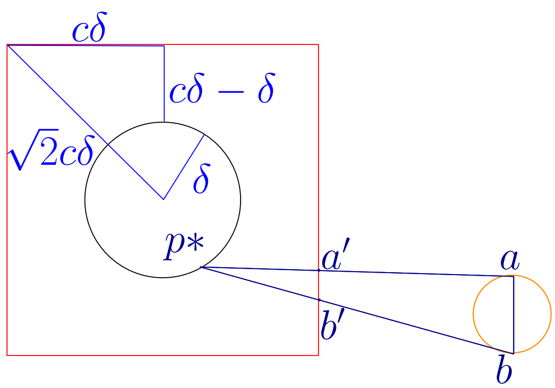

Figure 2.1: Outer case

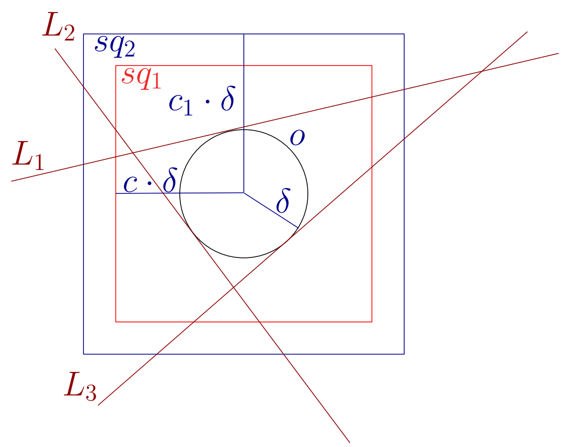

Figure 2.2: Initial setup and information required for the proof

We partition the set S into two groups, S^{gp1} and S^{gp2}. A shape $S_i \in S$ will be in S^{gp1} if the center of the largest enclosed disk in S_i or at least one of the largest enclosed disks in S_i (in case S_i has multiple largest enclosed disks) is located completely outside of sq_1. Otherwise S_i will be in S^{gp2}.

Piercing S^{gp1} By the definition of o, every shape in S^{gp1} intersects o. Every shape S_i in S^{gp1} is convex, intersects o and has at least one of the largest disk(s) enclosed in S_i centred outside of sq_1. We now show how to place a set of points on the boundary of sq_1 to pierce all the shapes in S^{gp1}. Let S_i be an arbitrary shape in S^{gp1}. Let p^* be an arbitrary point in the intersection of S_i and the boundary of o. Let o' be the largest disk enclosed in S_i and centered outside of sq_1(in the case of multiple disks satisfying these conditions, pick an arbitrary one). Without loss of generality, assume that S_i intersects the right vertical side of sq_1. Let ab be the diameter of o' parallel to the y-axis. Since S_i is convex, there exists a triangle p^*ab that is contained in S_i. Observe that the triangle p^*ab intersects a contiguous portion of the boundary of sq_1. Let the boundary of the triangle p^*ab cross the boundary of sq_1 at points a' and b'. Now the minimum possible length of the segment $a'b'$ gives us the required resolution of points to put on the boundary of sq_1 to pierce S^{gp1}. Recall that α_i is the fatness of S_i. The smallest disk that encloses S_i has a radius greater than or equal to $\frac{|p^*b|}{2}$ since the segment p^*b is in S_i and any disk with diameter less than $|p^*b|$ cannot have a segment of length $|p^*b|$ in it. Thus, $\alpha_i \geq \frac{\frac{|p^*b|}{2}}{\frac{|ab|}{2}} = \frac{|p^*b|}{|ab|}$. Moreover, since the two triangles p^*ab and $p^*a'b'$ are similar we get following equation: $\frac{|a'b'|}{|p^*b'|} = \frac{|ab|}{|p^*b|} \implies |a'b'| = \frac{|ab|\cdot|p^*b'|}{|p^*b|} \implies |a'b'| \geq \frac{|p^*b'|}{\alpha_i}$. Furthermore, note that $(c-1)\delta \leq |p^*b'| \leq 2\sqrt{2}c\delta$ and $\alpha(S) \geq \alpha_i$. Therefore, $|a'b'|$ is at least $\frac{(c-1)\delta}{\alpha(S)}$.

Exceptional Case The only exceptional case in this scenario is when a' is not located on the same side of sq_1 as b'. Considering the fact that the disk o' is centered outside of sq_1, the convexity of S_i implies that S_i contains a corner of sq_1 . To pierce such shapes we put points on the 4 corners of sq_1.

Proof of exceptional case (2.2.1) of S^{gp1} in Theorem 2.1. Without loss of generality, assume that the center of o' is to the right of the line that passes through the right side of sq_1. Let $side_1$ be the side of sq_1 that b' is located on. Without loss of generality, let $side_1$ be the bottom side of sq_1. Let $side_2$ be the side of sq_1 that a' is on. $side_2$ is either the right or the top side of sq_1.

Let D be the intersection of $side_1$ and $side_2$. If s_1 and s_2 were parallel, let D be the right end point of b'. Take the triangle p^*ab. By definition D is in the halfspace that is tangent to p^*b and contains p^*ab. D is in the halfspace that is tangent to p^*a and contains p^*ab. D is in the halfspace that is tangent to ab and contains p^*ab. Thus, D is in the convex hull of p^*, a, b. That implies that S_i contains D. □

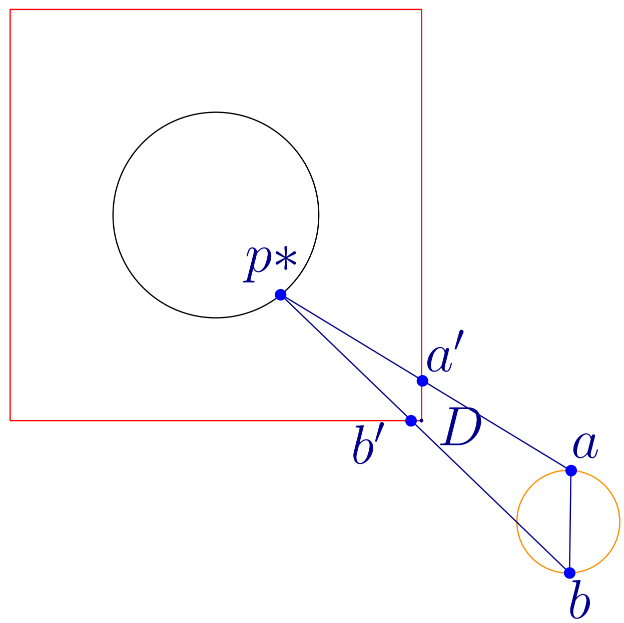

Figure 2.3: Exceptional case for a shape in S_{gp1}

The perimeter of sq_1 is $8c\delta$, therefore the number of points placed on the perimeter of sq_1 to pierce all the shapes in S^{gp1} is $4 + \frac{8c\delta}{\frac{(c-1)\delta}{\alpha(S)}} = 4 + \frac{8c}{c-1}\alpha(S)$.

Piercing S^{gp2} Let S_i be an arbitrary shape in S^{gp2}. Let $L_i^{'}$ be a line through the center of circle o and parallel to L_i for $i \in \{1, 2, 3\}$. We call a point p *proper* with respect to $L_i, i \in [1, 3]$ if it is located inside sq_2, and p is located on the same side of L_i and $L_i^{'}$ but p is closer to $L_i^{'}$. For an illustration see Figure 2.4.

Lemma 2.1. *Any point inside* sq_2 *is proper with respect to some* $L_i, i \in [1, 3]$.

Proof. Let $H_i, i \in [1, 3]$ be the halfspace that is tangent to $L_i^{'}$ and does not contain L_i. Since H_1, H_2, H_3 intersect at a point and the union of the angle that they cover is 2π (otherwise it contradicts with the fact L_1, L_2, L_3 form a triangle). Using the result of Bose et al. [8] we have that the $\cup H_i$, for $i \in [1, 3]$ covers the entire plane. Thus, they cover any point in sq_1 as well. □

Without loss of generality, assume that the center of at least one of the largest disks enclosed in S_i is a proper point respected to L_1. Such a disk exists, since, at least one of the largest disks enclosed in S_i is centred in sq_1.

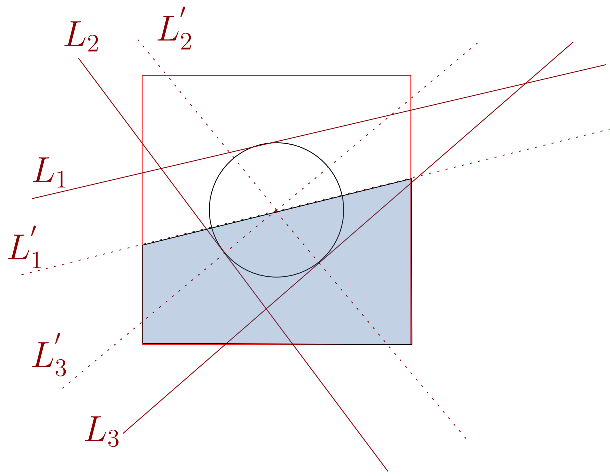

Figure 2.4: Any point in the blue region are proper points with respect to L_1

We analyze two cases, namely when $S_i \cap L_i \cap sq_2 \neq \emptyset$ and when $S_i \cap L_i \cap sq_2 = \emptyset$

Case 1. S_i has an intersection with L_1 inside sq_2.

Let p^* be an arbitrary point in the intersection of L_1 and S_i interior to sq_1. Let o' be a largest disk enclosed in S_i centred inside sq_1. Let ab be the diameter of o' parallel to L_1. Since the center of o' is proper point with respect to L_1, the triangle p^*ab intersects $L_1^{'}$ at two points a' and b'. Recall that α_i is the fatness of S_i, the smallest disk that encloses S_i has a radius greater than or equal to $\frac{|p^*b|}{2}$, since the segment p^*b is in S_i and any disk with diameter less than $|p^*b|$ cannot have a segment of length $|p^*b|$. Thus, $\alpha_i \geq \frac{\frac{|p^*b|}{2}}{\frac{|ab|}{2}} = \frac{|p^*b|}{|ab|}$. Moreover, since the triangles p^*ab and $p^*a'b'$ are similar we get following equation: $\frac{|a'b'|}{|p^*b'|} = \frac{|ab|}{|p^*b|} \implies |a'b'| = \frac{|ab|.|p^*b'|}{|p^*b|} \implies |a'b'| \geq \frac{|p^*b'|}{\alpha_i}$. By definition and the relation between L_1 and $L_1^{'}$, $\delta \leq |p^*b'| \leq 2\sqrt{2}c\delta$ and $\alpha(S) \geq \alpha_i$. Therefore, $|a'b'| \geq \frac{\delta}{\alpha(S)}$.

The length of $L_1^{'} \cap sq_2$ is at most $2\sqrt{2}c_1\delta$, which is the diameter of $sq2$. Hence, we place $\frac{2\sqrt{2}c_1\delta}{\frac{\delta}{\alpha(S)}} = 2\sqrt{2}c_1\alpha(S)$ points on $L_1^{'}$. So, in total $6\sqrt{2}c_1\alpha(S)$ points on $L_1^{'}, L_2^{'}, L_3^{'}$ are sufficient to pierce shapes in this case.

Case 2. S_i only intersects L_1 outside of sq_2.

As a result S_i intersects with the boundary of sq_2. Let p^* be a point in the intersection of S_i and the boundary of sq_2. Without loss of generality, assume that p^* is located on the right side of sq_2. Let o' be a largest disk enclosed in S_i centred inside sq_1. Let ab be the diameter of o' parallel to the y-axis. Recall that α_i is the fatness of S_i. The smallest disk that encloses S_i has a radius greater than or equal to $\frac{|p^*b|}{2}$. By an identical argument as in Case 2.2.1, we have $|a'b'| = \frac{|ab|.|p^*b'|}{|p^*b|} \implies |a'b'| \geq \frac{|p^*b'|}{\alpha_i}$. By definition of sq_1, sq_2 and $P^*a'b'$; $(c_1-c)\delta \leq |p^*b'| \leq 2\sqrt{2}c_1\delta$ and $\alpha(S) \geq \alpha_i$. Therefore, the minimum length of $|a'b'|$ is $\frac{(c_1-c)\delta}{\alpha(S)}$. The perimeter of sq_1 is $8c\delta$, therefore the number of points required to put on the perimeter of sq_1 is $\frac{8c\delta}{\frac{(c_1-c)\delta}{\alpha(S)}} = \frac{8c}{c_1-c}\alpha(S)$.

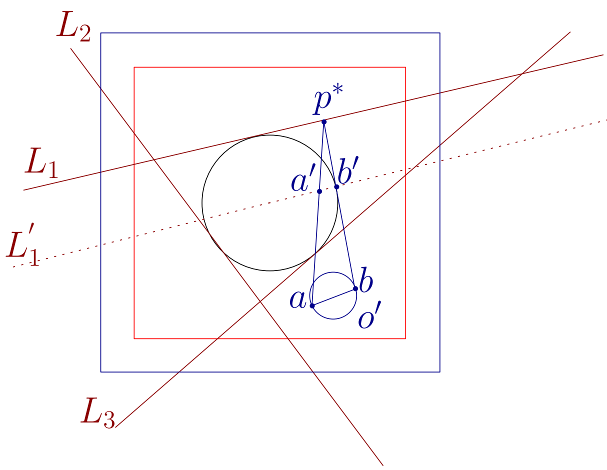

Figure 2.5: A convex shape with the largest enclosed disk centred in sq_1 and intersecting L_1 inside sq_2

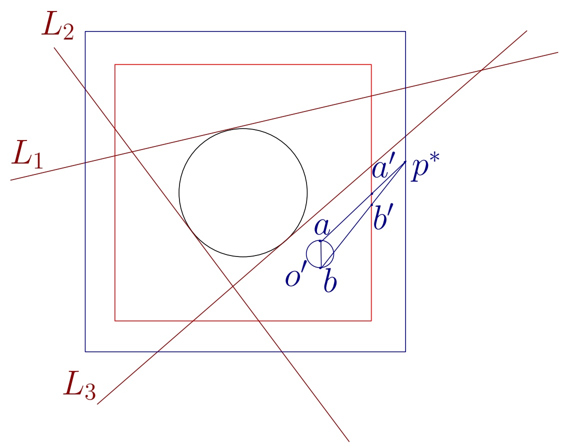

Figure 2.6: A convex shape with the largest enclosed disk centred in sq_1 and intersecting L_1 only outside of sq_2

Let $m = max(\frac{c}{c_1-c}, \frac{c}{c-1})$. The number of points sufficient to pierce the set S using the placements described above is $4 + (8m + 6\sqrt{2}c_1)\alpha(S)$. Our objective is to minimize $8m + 6\sqrt{2}c_1$. Here we break down this optimization problem into two simpler optimization problems:

1. Minimize $8c/(c_1 - c) + 6\sqrt{2}c_1$ subject to $1 < c < c_1 < 2c - 1$.
2. Minimize $8c/(c - 1) + 6\sqrt{2}c_1$ subject to $1 < c < 2c - 1 \leq c_1$.

The minimum value for both of these optimization problems is 39.788, which occurs when $2c - 1 = c_1$. Specifically, when $c = 1 + \frac{\sqrt[4]{2}}{\sqrt{3}}$ and $c_1 = 1 + \frac{2\sqrt[4]{2}}{\sqrt{3}}$. □

2.2.2 Implications

Our result has a number of implications on other research problems concerning sets of fat objects, such as computing depth orders, 3-D vertical ray shooting, 2-D point enclosure, range searching, and arc shooting for convex fat objects. The following are some corollaries where the asymptotic complexity is improved from $O(\alpha^2)$ to $O(\alpha)$ using our results:

- **Piercing a set of pairwise intersecting c-oriented convex polygon**
  A c-oriented convex polygon is a convex polygon whose edges are parallel to a fixed collection of c distinct directions.
  Nielsen [37][Lemma 3] showed that the best known upper-bound for the piercing number $\tau(\beta)$ when β is a set of pairwise intersecting c-oriented α-fat polygons is $O(\alpha)^2\phi(\beta)$ where $\phi(\beta)$ is the packing number of the set β. The packing number of a set B is the maximum number of disjoint objects in B.

- **Computing depth order for fat objects**
  An object is considered simply-shaped if it can be described by a constant number of polynomial inequalities, whose degrees are bounded by a small constant.
  A set of objects β in 3-D have the α-ratio property if the xy-projections of any pair of objects in β have the property that the ratio of the side length of the smallest axis-parallel square that contains K over the side length of the largest axis-parallel square that is contained in L is at most α.
  Agarwal et al. [2][Theorem 4.1]: The 2-dimensional linear-extension problem for a collection $\beta = \{\beta_1, \beta_2, \ldots, \beta_{n-1}\}$ of n convex, simply-shaped, α-fat objects with the α-ratio property is the following: Let $C = \{c_1, \ldots c_n\}$ be a set of n objects in the plane, and let $<$ be a partial binary anti-symmetric relation over C such that, for any pair of objects $c_i, c_j \in C$, if $c_i \cap c_j \neq \emptyset$ then either $c_i < c_j$ or $c_j < c_i$, and if $c_i \cap c_j = \emptyset$ then c_i and c_j are incomparable. Computing such a partial order is called *2-dimensional linear-extension*. Agarwal et al. [2] showed how to solve this problem in time $O(\alpha^2 n \lambda_s^{1/2}(n) \log^4 n)$, where s is the maximum number of intersections between the boundaries of any pair of objects in β. Within the same time bound, they show how to compute a depth order (if it exists) for a collection of n non-intersecting convex objects in 3-space, whose xy-projections have the above properties. The constant of proportionality is bounded by $c\alpha^2$, for some absolute constant c. $\lambda_s(n)$ is the maximum length of (n, s) Davenport-Schinzel sequences.

  To compute depth order for fat objects Agarwal et al. [2] used a tool such that given a collection β of pairwise intersecting fat objects, the tool partitions the collection into a number of subsets $L = \{\beta_1, \beta_2, \ldots, \beta_l\}$ such that each subset $\beta_i \in L$ has the property $\tau(\beta_i) = 1$. The number of groups is upper-bounded by $\tau(\beta)$.

- **3-D vertical ray shooting and 2-D point enclosure, range searching, and arc shooting amidst convex fat objects**

  Katz et al. [28] claim the following two theorems.

  Katz et al. [28] Theorem 2.4 Let C be a set of n non-intersecting convex simply-shaped flat objects in 3-space, and assume that the xy-projections of the objects in C are all α-fat. It is possible to construct a data structure of size $O(\lambda_s(n) \log^3 n)$ in $O(\lambda_s(n) \log^4 n)$ time, where s is the maximum number of intersections between the boundaries of the xy-projections of any pair of objects in C, so that, for a given query point p, the object of C lying immediately below p (if such an object exists) can be found in $O((\alpha^2 \log \alpha) \log^4 n)$ time.

  Katz et al. [28] Theorem 3.1 Let C be a set of n convex simply-shaped α-fat objects in the plane. It is possible to construct a data structure of size $O(\lambda_s(n) \log^3 n)$ in $O(\lambda_s(n) \log^4 n)$ time, where s is the maximum number of intersections between the boundaries of any pair of objects in C, so that, for a given query point p, the k objects of C containing p can be reported in $O((\alpha^2 \log \alpha) \log^3 n + k \log^2 n)$ time.

  Let κ be the set that consists of xy-projections of the objects in C. In both theorems stated above, Katz et al. [28] needed a tool to partition κ into smaller sets $\kappa_0, \kappa_1, \ldots, \kappa_l$, such that every $\kappa_i \in L$ has the *common point property*: "A set has a common point property if all objects in that set have a point in common".

1. **Piercing a set of pairwise intersecting c-oriented convex polygon**

   Corollary 2.1. *The piercing number $\tau(\beta)$ when β is a set of pairwise intersecting c-oriented α-fat polygons is $O(\alpha)$.*

2. **Computing depth order for fat objects**

   Corollary 2.2. *The time complexity of 2-dimensional linear-extension problem can is of $O(\alpha n \lambda_s^{1/2}(n) \log^4 n)$.*

3. **3-D vertical ray shooting and 2-D point enclosure, range searching, and arc shooting amidst convex fat objects**

In order compute this, that they used Fredman's sampling algorithm. This sampling process finds $O(\alpha^2)$ piercing points of a set of pairwise intersecting convex α-fat objects in $O(\alpha^2 \log \alpha)$ time. However, our result can replace Fredman's sampling algorithm and improve the query time complexity with respect to alpha for both cases.

Corollary 2.3. *For a given query point p, the object of C lying immediately below p (if such an object exists) can be found in $O(\alpha \log^4 n)$ time.*

Corollary 2.4. *For a given query point p, the k objects of C containing p can be reported in $O(\alpha \log^3 n + k \log^2 n)$ time.*

2.2.3 Piercing Fat Rectangles

In this section we demonstrate how to pierce a set C of pairwise intersecting rectangles of arbitrary orientation with fatness $\alpha(C)$. This theorem is a generalization of pairwise intersecting line segments in terms of fatness.

Theorem 2.2. *Any set C of pairwise intersecting rectangles of arbitrary orientation of fatness $\alpha(C)$ can be pierced with $(5\sqrt{2}+2)\alpha(C) + 25 + 5\sqrt{2} \approx 9\alpha(C) + 32 = O(\alpha(C))$ points.*

Lemma 2.2. *The maximum area of a square that does not have any lattice point in it is less than 2. A lattice point is a point with integer coordinates.*

Proof of Lemma 2.2. Let sq be an arbitrary oriented square of side length $\sqrt{2}$. There always exists a circle o of diameter $\sqrt{2}$ in sq. Since there exists an axis-parallel square in o of side length 1, we observe that, any axis-parallel square of side length 1 has at least one lattice point in it. Therefore, any square of side length $\sqrt{2}$ of arbitrary orientation has at least one lattice point in it. □

Proof of Theorem 2.2. Let $R = \{r_0, r_1, ..., r_{n-1}\}$ be a set of pairwise intersecting rectangles. Denote a rectangle r of width w and height h as (w, h). We assume that the longer side of an arbitrary oriented rectangle is the height of the rectangle ($h \geq w$). Without loss of generality, Let $r_1 = (w_1, h_1)$ be the rectangle with the minimum width among all rectangles in the set R. Without loss of generality, assume that r_1 is axis parallel.

Structure of the grid points Let $r_i = (w_i, h_i)$ be an arbitrary rectangle in R, let p^* be one of the intersection points of the boundary of r_1 with r_i. By definition of r_1 we have the following two inequalities: $w_i \geq w_1$ and $h_i \geq w_1$. For every $r_i \in R$ there exists a square s_i located inside r_i of side length w_1 such that, s_i contains the point p^*. Suppose that such a square does not exists. It implies that any square of side length w_1 that contains p^* intersects with the boundary of r_i. Thus, either $w_i < w_1$ or $h_i < w_1$, both of which contradicts with the minimality of w_1.

Let G be a grid of points whose resolution is 1. Let G' be a grid of points whose resolution is $\frac{w_1}{\sqrt{2}}$. By Lemma 2.2 we see that any square of side length at least w_1 must contain at least one point of G' (To see the argument simply scale down the squares and G' by factor of $\frac{w_1}{\sqrt{2}}$). By definition, every s_i intersects r_1, and, the distance from any point in any $s_i, i \in [0, n-1]$ to the boundary of r_1 is at most $\sqrt{2}w_1$(in the worst case the point can be on the opposite side of the diameter of a square). Therefore, we cover an axis-parallel rectangle centred with r_1, whose distance to r_1 is at most $\sqrt{2}w_1$, with a grid of points. See Figure 2.7 for illustration. That rectangle has width $w_1 + 2\sqrt{2}w_1$ and height $h_1 + 2\sqrt{2}w_1$. Therefore, The number of the grid points $(\frac{w_1+2\sqrt{2}w_1}{\frac{w_1}{\sqrt{2}}} + 1) \times (\frac{h_1+2\sqrt{2}w_1}{\frac{w_1}{\sqrt{2}}} + 1) = (\sqrt{2}+5) \times (\sqrt{2}\frac{h_1}{w_1} + 5) \leq (5\sqrt{2}+2)\alpha(C) + 25 + 5\sqrt{2}$. Therefore the sufficient number of points on the grid to pierce the set C is $(5\sqrt{2}+2)\alpha(C) + 25 + 5\sqrt{2}$ that is approximately $9\alpha(C) + 32$ $\square$

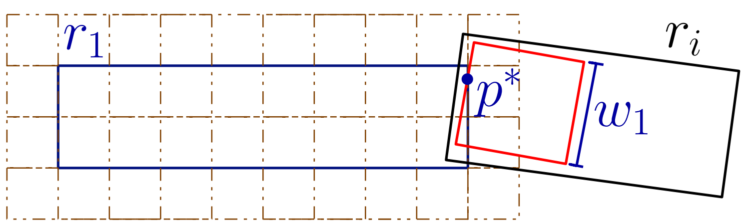

Figure 2.7: How an arbitrary rectangle in R gets pierced by a point on the grid.

2.3 Refined Results for Specific Shapes

In this subsection we study the number of points sufficient to pierce more specific sets of shapes. First we study sets of pairwise intersecting homothets and design an algorithm that computes the exact location of the points that pierce the set. Next, we show that 2 points are sometimes necessary and always sufficient to pierce a set of pairwise intersecting homothets of a regular hexagon.

2.3.1 Homothets of a convex shape

In this section, we show how one can pierce any set of pairwise intersecting homothetic shapes with a constant number of points. More precisely, we give an upper-bound of 15 piercing points. Kim et al. [32] proved that 16 points are sufficient to pierce any pairwise intersecting homothetic convex shapes. Since our approach is similar, we restate their proof with some small modifications. Let S be a set of pairwise intersecting homothetic shapes. We prove that 15 points are enough by eliminating one of the 16 points. Finally, given a set S of n k-gons, we give an algorithm of complexity $O(n + \log^2 k)$ to find the exact location of 16 piercing points and $O(n \log k + \log^2 k)$ to find 15 piercing points.

Theorem 2.3. *Any set of pairwise intersecting convex homothets can be pierced by 16 points.*

Proof of Theorem 2.3. Observe that the shapes in the set $R = \{R_0, R_1, ..., R_{n-1}\}$ will be pairwise intersecting, since S consists of pairwise intersecting shapes. Therefore, one point is enough to pierce the set R.

Since the set S is a set of convex homothets, sets R and $r = \{r_0, r_1, ..., r_{n-1}\}$ are also homothets (Proof is mentioned below). Without loss of generality, assume that the sets R and r consist of axis-parallel rectangles.

Without loss of generality, let $R^* = (W^*,\ L^*)$ be the rectangle with the minimum width in R and, $r^* = (w^*,\ l^*)$ be the corresponding rectangle in r. Schwarzkopf et al. [42] showed that $2w^* \geq W^*$ and $2l^* \geq L^*$. Now shrink the rest of the convex shapes in S while maintaining the following three conditions: First, the shapes should intersect S_1. Second, after shrinking, the resulting shapes should be enclosed by the corresponding original shapes. Third, after shrinking, the shapes should be the same size as S_1. Resulting shapes after shrinking are enclosed by original shapes, thus, piercing resulting shapes pierces the original shapes as well. Notice that S_1 and as a result R^* and r^* does not change. For any $R_i \in R$, define R_i^- to be the rectangle corresponding to R_i after shrinking S_i. Similarly, for any $r_i \in r$, define r_i^- to be the rectangle corresponding to r_i after shrinking S_i. And let $R^- = \{R_0^-, R_1^-, ..., R_{n-1}^-\}$.

For any rectangle r_i^-, the horizontal distance between the center points of r_1^- and r_i^- is at most W^*(which is less than $2w^*$) and the vertical distance is at most L^*(which is less than $2l^*$). Since, rectangles in R^- are of the same size as R^* and all of them intersect with R_1^-.

Without loss of generality, assume that center of r_1^- is on the origin. The center of any rectangle $r_i^-, i \in [0, n-1]$ can only be in the following range. $([-2w^*, +2w^*], [-2l^*, +2l^*])$, Figure 2.8. Therefore, the area that we need to cover is $(2w^* - (-2w^*)) \times (2l^* - (-2l^*)) = 16w^*l^*$.

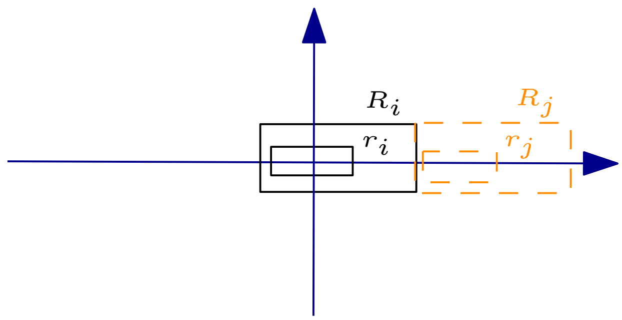

Figure 2.8: Relation between centers of two arbitrary $r_i, r_j \in C$

We cover that area by following 16 points. Each of these points will cover a rectangular area of size $w^* \times l^*$. Consider a point at the origin. Any axis-parallel rectangle with side lengths (w^*, l^*) contains the origin if their center point (p_x, p_y) is in following range: $-0.5w^* \leq p_x \leq 0.5w^*$ and $-0.5l^* \leq p_y \leq 0.5l^*$. Therefore, one piercing point can cover a rectangular area of size $(0.5w^* - (-0.5w^*)) \times (0.5l^* - (-0.5l^*)) = l^*w^*$. The total area that we need to cover is $16w^*l^*$, that can be partitioned into 16 rectangles of size $w^* \times l^*$. See the Figure 2.9 for illustration. Thus, $\frac{16w^*.l^*}{w^*.l^*} = 16$ piercing points are always sufficient.

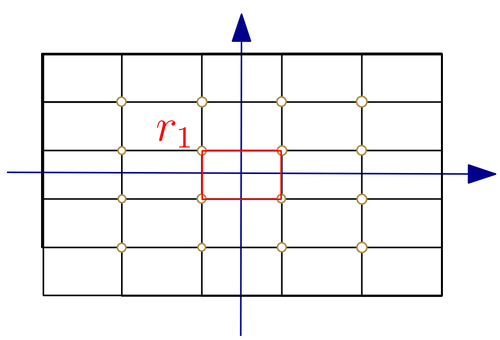

Figure 2.9: 16 points that pierce the set C

$\square$

Let $S = \{S_0, S_1, \ldots, S_{n-1}\}$ be a set of pairwise intersecting homothetic convex shapes in the plane. We transform every shape $S_i \in S$ into a pair of homothetic orthogonal rectangles (r_i, R_i), with each pair satisfying the following three conditions: First, r_i is enclosed in S_i, and, R_i encloses S_i. Second, the side length of r_i is at least half of the side length of R_i. Third, the vertices of r_i are located on the boundary of S_i.

For a shape S_i, define C_i to be a *cross-shaped* polygon with edges parallel to the edges of r_i, with $r_i \subseteq S_i \subseteq C_i \subseteq R_i$. Let V_i be the vertices of C_i which include the vertices of r_i.

Figure 2.10: r_i and C_i are enclosed in S_i and R_i

The existence of such a pair of enclosed and enclosing rectangles for any convex shape was shown by Schwarzkopf et al. [42]. They designed an algorithm to compute such a pair for a convex polygon in time $O(\log^2 k)$ when the k vertices of the polygon are given in a sorted array.

Let S^* be the smallest shape homothetic to the shapes in S that intersects every shape in S. Minimality of S^* implies that there exists at least 3 shapes in S, say S_1, S_2, S_3, that are tangent to S^* at points x_1, x_2, x_3. Let L_1, L_2, L_3 be the tangent lines to S^* at x_1, x_2, x_3. These three lines form a triangle. For simplicity we assume that S^* is an element of S.

Theorem 2.4. *Any set of pairwise intersecting convex homothets can be pierced by 15 points.*

Before proving this theorem, we prove a few helper lemmas. The 16 piercing points derived from the proof of Theorem 2.3 form a grid (see Figure 2.11). We label the points from 1 to 16 starting at the top left point. We show that a corner point can be removed from this set of piercing points.

According to the Theorem 2.3 every shape in S contains at least one of these 16 piercing points. If there is no shape in the set S that contains only one corner piercing points (points $1, 4, 13, 16$) we can simply remove one of the corner point and reduce the piercing number to 15. Otherwise, without loss of generality, let S_4(resp. $S_5, S_6, S_7) \in S$ be a shape that only contains the point 13 (resp. $1, 4, 16$).

Let $H_{i,j}^{k,+}$ (resp. $H_{i,j}^{k,-}$) be a halfspace defined by the line that goes through the piercing points i, j and contains the piercing point k (resp. does not contain the point k).

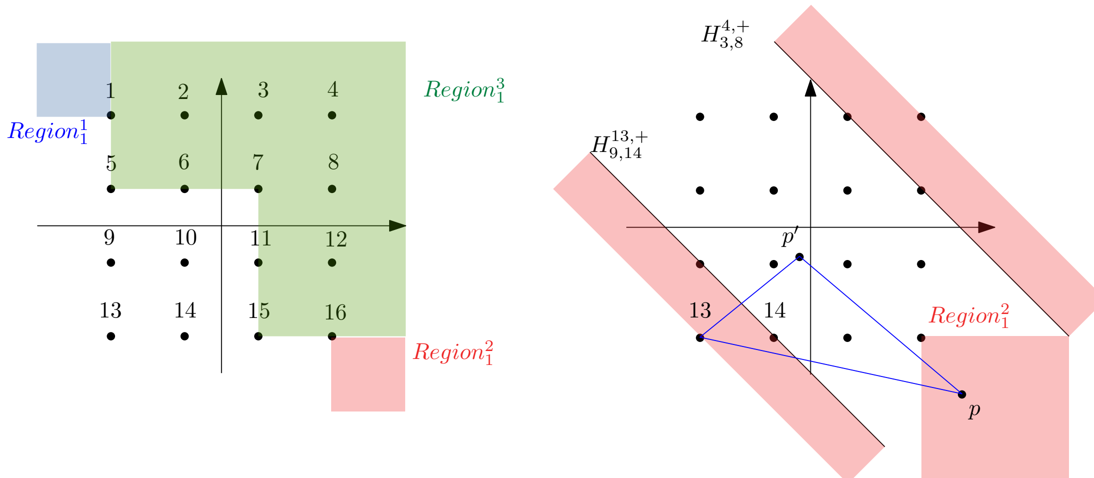

Figure 2.11: Dividing the space into different regions and The area that the intersection of S_4 and S_6 cannot be.

Let $Region_1$ (resp. $Region_2$, $Region_3$, $Region_4$) be the area defined as $H_{5,7}^{4,+} \cup H_{7,15}^{4,+}$ (resp. $H_{7,15}^{4,+} \cup H_{10,12}^{13,+}$, $H_{2,10}^{13,+} \cup H_{10,12}^{13,+}$, $H_{5,7}^{4,+} \cup H_{2,10}^{13,+}$). Let $Region_1^1$ be $H_{1,5}^{4,-} \cap H_{1,2}^{13,-}$. Let $Region_1^2$ be $H_{15,16}^{4,-} \cap H_{12,16}^{13,-}$. Let $Region_1^3$ be $Region_1 \cap (H_{1,5}^{4,+} \cap H_{15,16}^{4,+})$ and $Region_2^3$ be $Region_2 \cap (H_{1,2}^{13,+} \cap H_{12,16}^{13,+})$.

Lemma 2.3. *$S_4 \cap S_6 \nsubseteq Region_1^1 \cup Region_1^2$.*

Proof. Let $p \in S_4 \cap S_6$. Suppose, for the sake of a contradiction, $p \in Region_1^2$. Let $p' \in S_4 \cap S^*$ and let $p'' \in S_6 \cap S^*$.

- Suppose that p is not located in $Region_1^2 \cap H_{3,8}^{4,+}$.
  In this case the triangle formed by 4, p'' and p will contain point 8, which is a contradiction to the definition of S_6.

- Suppose that p is not located in $Region_1^2 \cap H_{9,14}^{13,+}$.
  Similarly, in this case the triangle formed by 13, p' and p will contain point 14, which is a contradiction to the definition of S_4.

Since $H_{3,8}^{4,+}$ and $H_{9,14}^{13,+}$ have an empty intersection, it implies that the point p cannot be in $Region_1^2$. The same argument holds for $Region_1^1$. Thus, S_4 and S_6 cannot intersect in $Region_1^1$ either.

□

Lemma 2.4. *$S_4 \cap Region_1^3$ has an empty intersection.*

Proof. Let p be a point on the boundary of $Region_1^3 \cap S_4$. Let S_4' be a shape homothetic to S_4 with the following conditions:

1. S_4' has the same size as S^*.
2. S_4' has p on its boundary.
3. S_4' is contained in S_4.

Let C_4' be the cross shaped polygon corresponding to S_4'. Let (r_4', R_4') be the pair of enclosed and enclosing rectangle corresponding to S_4'. Let v be the top right vertex of r_4'. By the definition of S_4', v cannot be inside of the rectangle defined by piercing points $9, 10, 13, 14$. Otherwise it contradicts S_4' having an intersection with the boundary of $Region_1^3$.

- If v is below point 13 then the triangle defined by point $13, p$ and v contains the piercing point 14.
- If v is to the left of point 13 then the triangle defined by point $13, p$ and v contains the piercing point 9.
- If v is to the right and above the piercing point 13, then since the resolution of the piercing points and resolution of the vertices that define r_4' (size of r_4') are equal it implies that r_4' as well as S_4 contains another piercing point beside point 13.

□

A similar argument holds for $S_6 \cap Region_2^3$. Lemmas 2.3 and 2.4 imply that $S_4 \cap S_6$ is located in the rectangle formed by piercing points $6, 7, 10, 11$.

Lemma 2.5. *$S_4 \cap Region_1$ has an empty intersection.*

Proof. Let p be a point from the intersection of the boundary of $Region_1$ and S_4. Notice that $Region_1 = Region_1^3 \cup (H_{11,15}^{4,+} \cap H_{15,16}^{4,-}) \cup (H_{1,5}^{4,-} \cap H_{5,6}^{4,+})$. Suppose that S_4 intersects $Region_1$. We analyze the following three cases:

1. $p \in Region_1^3$: According to the Lemma 2.3, p cannot be in $Region_1^3$

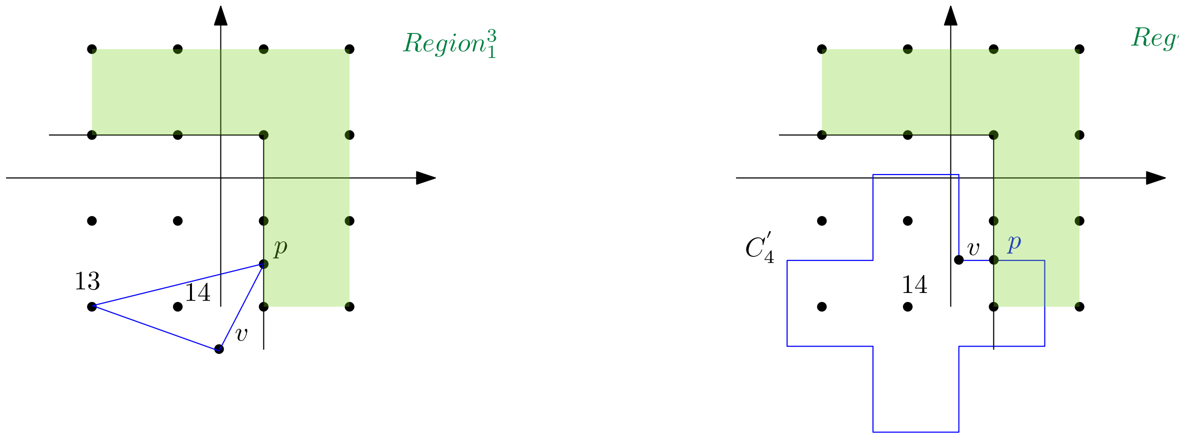

Figure 2.12: S_4 does not have an intersection with $Region_1^3$

2. $p \in H^{4,+}_{11,15} \cap H^{4,-}_{15,16}$: Let p' be a point from the intersection of S_4 and S_6. $S_4 \cap S_6$ is located in the rectangle formed by vertices $6, 7, 10, 11$. This means that p' is to the right of point 14. And it implies that the triangle formed by points 13, p and p' contains point 14 which is a contradiction to the definition of S_4.

3. $p \in H^{4,-}_{1,5} \cap H^{4,+}_{5,6}$: Let p' be a point from the intersection of S_4 and S_6. $S_4 \cap S_6$ is located in the rectangle formed by vertices $6, 7, 10, 11$. This means that p' is above point 9. And it implies that the triangle formed by points 13, p and p' contains point 9 which is a contradiction to the definition of S_4.

Thus, S_4 has an empty intersection with $Region_1$. □

For a region Reg, let $\overline{Reg}$ be the complement of the region Reg. More precisely, $\overline{Reg} = \{x | x \notin Reg, \forall x \in \mathbb{R}^2\}$.

Notice that S_1, S_2, S_3 are tangent to S^*. Also, S_4 intersects with S_1, S_2, S_3 and S^*. These two facts imply that S_4 intersects with L_1, L_2 and L_3. Moreover, Lemma 2.5 implies that the intersection of S_4 with each L_1, L_2, L_3 should be located in $\overline{Region_1}$. By symmetry $\overline{S_5 \text{ (resp. } S_6, S_7)}$ intersects with L_1, L_2, L_3. This intersection is located in $\overline{Region_2}$ (resp. $\overline{Region_3}, \overline{Region_4}$).

Lemma 2.6. *Each $\overline{Region_i}, i \in [1, 4]$ has a non-empty intersection with at least one of L_1, L_2, L_3.*

Proof. Notice that the bottom-left vertex of of r^* is in the $\overline{Region_1}$ and the triangle defined by the intersection of L_1, L_2, L_3 encloses r^*. Suppose that $\overline{Region_1}$ has empty intersection with all L_1, L_2, L_3. This implies that the triangle defined by the intersection of L_1, L_2, L_3 does not enclose r^*, which is a contradiction. Similarly this argument can be applied for $\overline{Region_2}$, $\overline{Region_3}$ and $\overline{Region_4}$. $\square$

Lemma 2.7. *At least two of the regions in* $\{\overline{Region_i}, i \in [1,4]\}$ *do not have an intersection with all three of* L_1, L_2, L_3.

Proof. Observe that each $L_i, i \in [1,3]$ can intersect with at most three regions of $\{\overline{Region_i} | i \in [1,4]\}$. Thus, we have at most 9 pairs of $(L_i, Region_j)$ when L_i intersects with $Region_j$ for $i \in [1,3], j \in [1,4]$. According to the Lemma 2.6 and the Pigeonhole theorem at least two of the regions in $\{\overline{Region_i} | i \in [1,4]\}$ does not intersect with all three of L_1, L_2, L_3. $\square$

Proof of Theorem 2.4. Lemma 2.7 implies that the regions corresponding to at least two of the shapes S_4, S_5, S_6, S_7 do not intersect with all three of L_1, L_2, L_3. This contradicts the fact that the shapes in the set S are pairwise intersecting. Thus, at least one piercing point can be removed from our piercing point set. $\square$

Algorithm to find the exact location of the piercing points First, we find the smallest shape, S_1, of the set in $O(n)$. Next we apply the Schwarzkopf et al. [42] algorithm on S_1 to compute the vertices of $r_1 = (w_1, h_1)$ in $O(\log^2 k)$. This allows us to compute the 16 points outlined in the proof of the Theorem 2.3, in a constant time. Thus the complexity of finding 16 piercing points is $O(n + \log^2 k)$. Next, we can determine in $O(n \log k)$ time which of the 4 corner points can be removed. Thus, we can find 15 piercing points in $O(n \log k + \log^2 k)$ time.

2.3.2 Hexagons

In this section we determine the piercing number of a set of pairwise intersecting homothets of a regular hexagon. We show that two points are always sufficient and sometimes necessary to pierce such a set. For a hexagon s with an edge parallel to the x-axis, we refer to its edges by *Bottom*, *BottomRight, TopRight, Top, TopLeft, BottomLeft* edges. We denote the *Bottom* edge of s by s^B. Respectively we refer to *BottomRight, TopRight, Top, TopLeft, BottomLeft* edges of s by $s^{BR}, s^{TR}, s^T, s^{TL}, s^{BL}$.

Theorem 2.5. *Any set of pairwise intersecting homothets of a regular hexagon can be pierced by two points.*

Proof of Theorem 2.5. Let $C = \{C_0, C_1, \dots C_{n-1}\}$ be a set of pairwise intersecting homothets of a regular hexagon. Without loss of generality, assume that the bottom edge of every hexagon in C is parallel with the x-axis. Let $TL = \{C_i^{TL} | \forall C_i \in C\}$, and $TR = \{C_i^{TR} | \forall C_i \in C\}$, and $BE = \{C_i^{B} | \forall C_i \in C\}$. Each element of these sets is a line segment. Segments of each set are associated with the same side of hexagons in C. $\forall C_i^{TL} \in TL$ let TL_i^+ be the halfspace defined by C_i^{TL} that does not contain the corresponding hexagon to C_i^{TL}. Let $TL^+ = \{TL_0^+, TL_1^+ \dots, TL_{n-1}^+\}$. $\forall C_i^{TR} \in TR$ let TR_i^+ be the halfspace defined by C_i^{TR} that does not contain the corresponding hexagon to C_i^{TR}. Let $TR^+ = \{TR_0^+, TR_1^+ \dots, TR_{n-1}^+\}$. Let tl^* be the halfspace in TL^+ that contains all other halfspaces in TL^+. Let tr^* be the halfspace in TR^+ that contains all other halfspaces in TR^+.

Let tl be the corresponding segment in TL to tl^*. Let tr be the corresponding segment in TR to tr^*. Let be be the top most segment in BE.

Define L_1 the line that is parallel to and goes through tl. Similarly, define R_1 to be the line that is parallel to and goes through tr, and B_1 to be the line that is parallel to and goes through be.

Assume that L_1, R_1, B_1 do not intersect at a point p. You will see that in such a case the two piercing points will be on top of each other at p, thus, p pierces the whole set. Observe that, the intersection points of L_1, R_1, B_1 form a equilateral triangle $T_h = ABD$. Let point B be the intersection point of lines L_1 and R_1. Let point D be the intersection point of lines L_1 and B_1 and let point A be the intersection point of lines B_1 and R_1. This triangle can have one of the two following possible shapes.

Case 1. In the first case, the point B is located above the segment AD. Therefore, the left top side of any hexagon in C is to the left of L_1. Similarly, the right top side of any hexagon in C is to the right of R_1 and any bottom side of any hexagon in the set is below B_1.

Case 2. In the second case, the point B is located below the segment AD. Therefore, the bottom right side of any hexagon in C is to the right of L_1 since, each pair of hexagons have to intersect. The top side of any hexagon in C is above B_1 and, the bottom left side of any hexagon is to the left of R_1, otherwise it contradicts the fact that each pair of hexagons intersects. Observe that this case is symmetric to the first case. Therefore, giving the proof for the first case is sufficient.

Proof for the case 1 Let M_L, M_R and M_B be the mid points corresponding to sides AB, BD and AD of the triangle $T_h = ABD$. We prove that every hexagon in C contains either M_B or B.

Take an arbitrary hexagon s from C. If s contains M_B then we are done. Suppose that s does not contain the point M_B. Without loss of generality, assume that the point M_B is to right of the s^{BR}. By the definition of R_1 the top right edge of any hexagon in C is to the right of R_1. Similarly, the bottom edge of any hexagon in C is to the bottom of B_1. This implies that the right bottom side of s should cross the lines B_1 and R_1. Let i_1 be the intersection point of s^{BR} and B_1, and i_2 be the intersection point of s^{BR} and R_1.

The point i_1 is to the left of M_B and i_2 is to the left of M_R. Observe that the triangle i_1i_2C is similar to M_BM_RC and $|Ci_2| > |CM_B|$. Therefore, the segment i_1i_2 is larger than M_RM_B. Moreover, the side length of s is greater than or equal to the length of the segment i_1i_2, and it is greater than M_BM_R ($|s^B| = |s^{BR}| \geq |i_1i_2| > |M_RM_B| = |M_RB| = |M_BA|$). Considering the facts that the side length of s is greater than $|M_RM_B|$, and s^{TR} is parallel to R_1 and to the right of R_1. Thus, by convexity of s, s^{TR} crosses L_1, and similarly s^B crosses B_1. Thus s contains the segment AB and in particular the point B. See Figure 2.13 for illustration.

$\square$

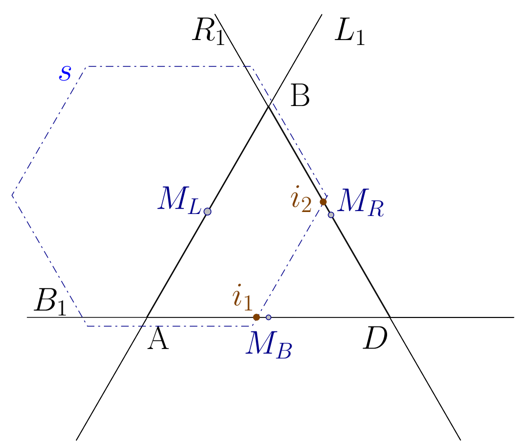

Figure 2.13: A hexagon $s \in C$ has to contain the point B if it does not contains M_B.

2.4 Conclusion

In this paper we showed that pairwise intersecting convex shapes of fatness α with arbitrary orientation can be pierced by a linear number of points with respect to the fatness parameter of the shapes in the set. The main idea to achieve our results is to avoid covering an area with a grid of high resolution but rather focusing on the perimeter of a specific shape. By using this idea we reduce the number of points sufficient to pierce any pairwise intersecting convex α-fat shapes from $O(\alpha^2)$ to $O(\alpha)$. Our theorem is an improvement over Fredman's sampling algorithm to find piercing points.

Moreover, for a set of pairwise intersecting homothets we showed that the piercing number is at most 15 points. The piercing number of a set of pairwise intersecting set of homothets of regular hexagons is 2 which is tight. We leave as an open problem to improve our upper bounds.

Chapter 3

$2 \times n$ Grids have Unbounded Anagram-Free Chromatic Number

$2 \times n$ GRIDS HAVE UNBOUNDED ANAGRAM-FREE CHROMATIC NUMBER[1]

Saman Bazargani[2] Paz Carmi[3] Vida Dujmović[4] Pat Morin[5]

[1]This research was partly funded by NSERC.

Abstract. We show that anagram-free vertex colouring a $2 \times n$ square grid requires a number of colours that increases with n. This answers an open question in Wilson's thesis and shows that even graphs of pathwidth 2 do not have anagram-free colourings with a bounded number of colours.

3.1 Introduction

Two words s and t are *anagrams* of each other if s is a permutation of t. A single word $w := w_1, \dots, w_{2r}$ is *anagramish* if its first half $w_1, \dots, w_r$ and its second half $w_{r+1}, \dots, w_{2r}$ are anagrams of each other. A word is *anagram-free* if it does not contain an anagramish factor. Anagramish words are also known as *abelian squares* and anagram-free words are also known as *abelian square free* words [30, 31], *strongly nonrepetitive* words [41], or *strongly asymmetric* sequences [19, 19].

In 1961, [18] asked if there exist arbitrarily long anagram-free words over an alphabet of size 4.[6] In 1968 [19, 19] showed the existence of arbitrarily long anagram-free words over an alphabet of size 25 and in 1971 [41] showed that an alphabet of size 5 is sufficient. Erdős's question was not fully resolved until 1992, when [30] answered it in the affirmative.

A path $v_1, \dots, v_{2r}$ in a graph G is *anagramish* under a vertex c-colouring $\phi : V(G) \to \{1, \dots, c\}$ if $\phi(v_1), \dots, \phi(v_{2r})$ is an anagramish word. The colouring ϕ is an *anagram-free colouring* of G if no path in G is anagramish under ϕ. The minimum integer c for which G has an anagram-free vertex c-colouring is called the *anagram-free chromatic number* of G and is denoted by $\dot{\chi}_\pi(G)$.

For a non-empty graph family $\mathcal{G}$, $\dot{\chi}_\pi(\mathcal{G}) := \max\{\dot{\chi}_\pi(G) : G \in \mathcal{G}\}$ or $\dot{\chi}_\pi(\mathcal{G}) := \infty$ if the maximum is undefined. The results on anagram-free words discussed in the preceding paragraph can be interpreted in terms of $\dot{\chi}_\pi(\mathcal{P})$ where $\mathcal{P}$ is the family of all paths. Indeed, Keränen's result shows that $\dot{\chi}_\pi(\mathcal{P}) \leq 4$. Slightly more complicated than paths are trees. [49] showed that $\dot{\chi}_\pi(\mathcal{T}) = \infty$ for the family $\mathcal{T}$ of trees and [25] showed that $\dot{\chi}_\pi(\mathcal{T}_2) = \infty$ even for the family $\mathcal{T}_2$ of binary trees.

One positive result in this context is that of [49], who showed that every tree T of pathwidth p has $\dot{\chi}_\pi(T) \leq 4p + 1$. Since trees are graphs of treewidth 1 it is natural to ask if this result can be extended to show that every graph G of treewidth t and pathwidth p

[6] This was an incredibly prescient question since it is not at all obvious that there exist arbitrarily long anagram-free words over *any* finite alphabet. The only justification for choosing the constant 4 is that a short case analysis rules out the possibility of length-8 anagram-free words over an alphabet of size 3.

has $\dot{\chi}_\pi(G) \le f(t,p)$ for some function $f : \mathbb{N}^2 \to \mathbb{N}$. [11] showed that such a generalization is not possible for any $t \ge 3$ by giving examples of n-vertex graphs of pathwidth 3 (and treewidth 3) with $\dot{\chi}_\pi(G) \in \Omega(\log n)$. The obvious remaining gap left by these two works is graphs of treewidth 2. Our main result is to show that G_n, the $2 \times n$ square grid has $\dot{\chi}_\pi(G_n) \in \omega_n(1)$. Since G_n has pathwidth 2, we have:

Theorem 3.1. *For every $c \in \mathbb{N}$, there exists a graph of pathwidth 2 that has no anagram-free vertex c-colouring.*

[47, Section 7.1] conjectured that $\dot{\chi}_\pi(G_n) \in \omega_n(1)$, so this work confirms this conjecture. Prior to the current work, it was not even known if the family of $n \times n$ square grids had anagram-free colourings using a bounded number of colours.

In a larger context, this lower bound gives more evidence that, except for a few special cases (paths [19,30,41], trees of bounded pathwidth [49], and highly subdivided graphs [48]), the qualitative behaviour of anagram-free chromatic number is not much different than that of treedepth/centered colouring [36]. Very roughly: For most graph classes, every graph in the class has an anagram-free colouring using a bounded number of colours precisely when every graph in the class has a colouring using a bounded number of colours in which every path contains a colour that appears only once in the path.

The remainder of this paper is organized as follows: Section 3.2 gives some definitions and states a key lemma which shows, under a certain periodicity condition, that every sufficiently long word contains a factor that is ϵ-close to being anagramish. In Section 3.3 we use this key lemma to prove Theorem 3.1. In Section 3.4 we prove the key lemma. Section 3.5 concludes with some final remarks about the (non-)constructiveness of our proof technique.

3.2 Periodicity in words

An *alphabet* Σ is a finite non-empty set and each element of Σ is called a *letter*. A *word* over Σ is a (possibly empty) sequence $w := w_1, \dots, w_n$ with $w_i \in \Sigma$ for each $i \in \{1, \dots, n\}$. The *length* $|w|$ of w is the length, n, of the sequence. A word of length at least 1 is *non-empty*. The unique word of length 0 is denoted by ε. For a word $w := w_1, \dots, w_n$, we let $\Sigma_w := \{w_i : i \in \{1, \dots, n\}\}$. For each $1 \le i \le j \le n+1$, $w_i, w_{i+1}, \dots, w_{j-1}$ is called a *factor* of w.[7] The set of all factors of w is denoted by $F(w)$.

[7]Note that this definition includes the empty factor ε obtained when $i = j$.

A *language* L over Σ is a set of words over Σ. A language L is *infinite* if $|L| = \infty$. We let $\Sigma_L := \bigcup_{w \in L} \Sigma_w$ and $F(L) := \bigcup_{w \in L} F(w)$. The language L is *factor-closed* if $F(w) \subseteq L$ for each $w \in L$. For any alphabet Σ and any $k \in \mathbb{N}$, Σ^k (the k-fold cartesian product of Σ with itself) is the language consisting of all length-k words over Σ. The *Kleene closure* $\Sigma^* := \bigcup_{k=0}^{\infty} \Sigma^k$ is the language consisting of all words over Σ.

Let $w := w_1, \ldots, w_n$ be a word over an alphabet Σ and, for each $a \in \Sigma$, define $|w|_a := |\{i \in \{1, \ldots, n\} : w_i = a\}|$. The *Parikh vector* of w is the integer-valued $|\Sigma|$-vector $\mathrm{p}(w) := (|w|_a : a \in \Sigma)$ that is indexed by elements of Σ. For example, the parikh vector of the word $w := abac$ over the alphabet $\Sigma = \{a, b, c, d\}$ is $(2, 1, 1, 0)$. Observe that a word $w_1, \ldots, w_{2r}$ is anagramish if and only if $\mathrm{p}(w_1, \ldots, w_r) = \mathrm{p}(w_{r+1}, \ldots, w_{2r})$ or, equivalently, $\mathrm{p}(w_1, \ldots, w_r) - \mathrm{p}(w_{r+1}, \ldots, w_{2r}) = \mathbf{0}$. For each $a \in \Sigma$, let $\delta_a(w) := |w_1, \ldots, w_r|_a - |w_{r+1}, \ldots, w_{2r}|_a$ and let $\tau_a(w) := |\delta_a|$. Then $\tau(w) := \sum_{a \in \Sigma} \tau_a(w)$ is a useful measure of how far a word is from being anagramish and $\tau(w) = 0$ if and only if w is anagramish.

A word $w := w_1, \ldots, w_n$ is *ℓ-periodic* over an alphabet $\Sigma \supseteq \Sigma_w$ if each length-ℓ factor of w contains every letter in Σ. A language L is *ℓ-periodic* if each of its words is ℓ-periodic over Σ_L. A language L is *periodic* if it is ℓ-periodic for some finite ℓ. We make use of the following key lemma, which states that every sufficiently long ℓ-periodic word contains arbitrarily long factors that are ϵ-close to being anagramish.

Lemma 3.1. *For each $r_0, \ell \in \mathbb{N}$ and each $\epsilon > 0$, there exists a positive integer n such that every ℓ-periodic word $w_1, \ldots, w_n$ contains a factor $w := w_{i+1}, \ldots, w_{i+2r}$ of length $2r \geq 2r_0$ such that $\tau(w) \leq \epsilon r$.*

The proof of Lemma 3.1 is deferred to Section 3.4. We now give some intuition as to how it is used. The process of checking if a word is anagramish can be viewed as finding common letters in the first and second halves and crossing them both out. If this results in a complete cancellation of all letters, then the word is an anagram. Lemma 3.1 tells us that we can always find a long factor w where, after exhaustive cancellation, only an ϵ-fraction of the original letters remain.

Stated differently, if up to ϵr letters in each half of w were each allowed to cancel two of the same letters in the other half of w, then it would be possible to complete the cancellation process. To achieve this type of one-versus-two cancellation in our setting, we decompose our coloured pathwidth-2 graph into pieces of constant size. The vertices in each piece can be covered with one path or partitioned into two paths (see Figure 3.4). In this way an occurrence H_z of a particular coloured piece in one half can be matched with two like-coloured pieces H_x and H_y in the other half. We construct a single path P that contains all vertices in H_z and only half the vertices in each of H_x and H_y. In this way, the colours of vertices $P \cap H_z$ can cancel the colours of the vertices in $P \cap (H_x \cup H_y)$.

Since Lemma 3.1 requires that the word w be ℓ-periodic, the following lemma will be helpful in obtaining words that can be used with Lemma 3.1. (Although Lemma 3.2 follows immediately from considerably stronger results in combinatorics on words—for example [33, Proposition 1.5.12]—we provide a proof here for the sake of completeness.)

Lemma 3.2. *Let L be an infinite factor-closed language over an alphabet Σ. Then there exists an infinite periodic factor-closed language $M \subseteq L$.*

Proof. Let $\Xi \subseteq \Sigma$ be any minimal subset of Σ with the property that $M := \Xi^* \cap L$ is infinite. Observe that, since L is factor closed, M is also factor-closed. Therefore M is infinite and factor-closed. All that remains is to show that M is ℓ-periodic for some integer ℓ.

Since Ξ is minimal, $\Pi^* \cap L$ is finite for any strict subset $\Pi \subsetneq \Xi$. Since L is factor-closed, $\Pi^* \cap L$ is factor-closed. Since $\Pi^* \cap L$ is factor-closed and finite, $|w|$ is finite for each $w \in \Pi^* \cap L$. Therefore, the value

$$\ell := 1 + \max\{|w| : w \in \Pi^* \cap L,\ \Pi \subsetneq \Xi\} \tag{3.1}$$

is finite. Now, consider any factor f of any word in M that does not contain every letter of Ξ, i.e., $f \in \Pi^*$ for some $\Pi \subsetneq \Xi$. Since $\Pi^* \cap L$ is factor closed, $f \in \Pi^* \cap L$ so, by Equation (3.1), $|f| < \ell$. Contrapositively, any factor of length at least ℓ of any word in M contains every element of Ξ, so M is ℓ-periodic. □

3.3 Proof of Theorem 3.1

For each $n \in \mathbb{N}$, let G_n be the $2 \times n$ square grid with top row $a_0, \ldots, a_{n-1}$ and bottom row $b_0, \ldots, b_{n-1}$ (see Figure 3.1). Our proof strategy is a proof by contradiction. Ultimately, we will assume (for the sake of contradiction) that there exists an integer c such that $\dot{\chi}_\pi(G_n) \leq c$ for each $n \in \mathbb{N}$ and use this to derive a contradiction. In the following, we will define several languages L_1, L_2, and L_3 that would be infinite if the preceding assumption were true. The final contradiction will occur if L_3 contains a sufficiently long string.

As a first step, observe that for any $n \in \mathbb{N}$ there is a bijection between vertex c-colourings of G_n and words of length $2n$ over the alphabet $\Sigma_0 := \{1, \ldots, c\}$ where a word

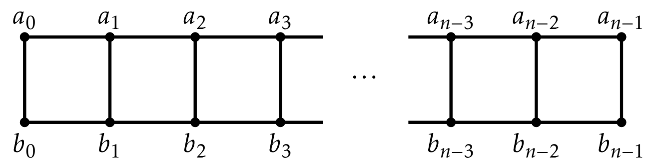

Figure 3.1: The graph G_n. These graphs are known as *ladder graphs*.

$w := w_0, \ldots, w_{2n-1}$ corresponds to the colouring $\phi_w : V(G_n) \to \{1, \ldots, c\}$ defined by $\phi_w(a_i) := w_{2i}$ and $\phi_w(b_i) := w_{2i+1}$, for each $i \in \{0, \ldots, n-1\}$.

For reasons that will become apparent, it is useful to break G_n into blocks consisting of 4 columns each. Let $\Sigma_1 := \Sigma_0^8$ and define the language L_1 to consist of exactly those words $w_1, \ldots, w_r \in \Sigma_1^*$ for which $\phi_{w_1 \cdots w_r}$ is an anagram-free colouring of G_{4r}.

By the definition of anagram-free colouring, every factor of a word in L_1 is also anagram-free, hence, L_1 is factor-closed. Define the language $L_2 \subseteq L_1$ as follows: If L_1 is finite, then $L_2 := L_1$. If L_1 is infinite then, by Lemma 3.2, there exists an infinite periodic factor-closed language contained in L_1 and we let L_2 be any such language. Note that in either case L_2 is periodic, since every finite language is periodic.

Let x be any letter in Σ_{L_2}. Note that the word xx is not in L_2 since the corresponding colouring ϕ_{xx} of G_8 has, for example, the anagramish path $a_0, \ldots, a_7$. Since L_2 is factor-closed, no word in L_2 contains xx as a factor, i.e., $xx \notin F(L_2)$. Let $\Sigma_3 \subset F(L_2)$ contain only those factors of L_2 that do not contain x. Thus, any word in L_2 is of the form $w_0 x w_1 x w_2 \cdots x w_r x w_{r+1}$ where each of $w_1, \ldots, w_r$ is in Σ_3 and each of w_0, w_{r+1} is in $\Sigma_3 \cup \{\varepsilon\}$.

Since L_2 is periodic, each word in Σ_3 has bounded length, so Σ_3 is finite, i.e., Σ_3 is an alphabet. Let L_3 be the language consisting of all words $w_1, \ldots, w_r \in \Sigma_3^*$ such that $x w_1 x w_2 \cdots x w_r \in L_2$. If L_2 is infinite, then so is L_3. Since L_2 is factor closed and periodic, L_3 is also factor-closed and periodic.

See Figure 3.2 for what follows. Each word $w := w_1, \ldots, w_r \in L_3$ defines a word $w' := x w_1 x w_2 \cdots x w_r \in L_2$. In graph colouring terms, w defines an integer $n(w) := 4 \sum_{i=1}^{r} (1 + |w_i|)$ and an anagram-free colouring $\phi_w := \phi_{w'}$ of $G_{n(w)}$. The word w' partitions the columns of $G_{n(w)}$ into *blocks*, each of which corresponds to a letter a of w' of length $4|a|$. There are two types of blocks:

- There are *boring* blocks $Q_0, \ldots, Q_{r-1}$, each containing 4 columns of $G_{n(w)}$, and each corresponding to an occurrence of x in w'.
- There are *colourful blocks* $H_1, \ldots, H_r$, where H_i is the subgraph of $G_{n(w)}$ corresponding to w_i and contains $4|w_i|$ columns of $G_{n(w)}$.

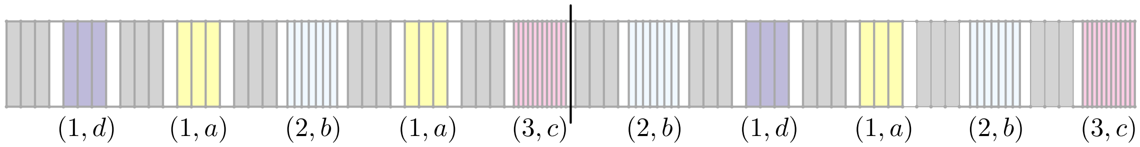

Figure 3.2: A word $w := dabacbdabc$, its corresponding word $w' := xdxaxbxaxcxbxdxaxbxc$, and its corresponding coloured graph $G_{n(w)}$.

The following lemma shows that any sufficiently long string w in L_3 must define a colouring ϕ_w of $G_{n(w)}$ that is not anagram-free.

Lemma 3.3. *There exists $\epsilon > 0$ and $r_0 \in \mathbb{N}$ such that, for any $w \in L_3$ that has even length $2r := |w| \geq 2r_0$ and that has $\tau(w) \leq \epsilon r$, the graph $G_{n(w)}$ contains a path that is anagramish under ϕ_w.*

Proof. Let $w_1, \ldots, w_{2r} := w$ so that the colouring ϕ_w is defined by the word $xw_1xw_2 \cdots xw_{2r}$. For each $a \in \Sigma_3 \setminus \{x\}$, define $X_a^+ := \{i \in \{1, \ldots, r-1\} : w_i = a\}$ and $Y_a^+ := \{i \in \{r+1, \ldots, 2r-1\} : w_i = a\}$. (The sets X_a^+ and Y_a^+ index the occurrences of a in the first and second half of w, respectively.) We will now define two sets $X_a \subseteq X_a^+$ and $Y_a \subseteq Y_a^+$ with the following properties:

1. If $\delta_a(w) = 0$, then $X_a = Y_a = \emptyset$.
2. If $\delta_a(w) > 0$, then $|X_a| = 2\delta_a(w) = 2|Y_a|$.
3. If $\delta_a(w) < 0$, then $|X_a| = \delta_a(w) = \frac{1}{2}|Y_a|$.

Let $X := \bigcup_{a \in \Sigma_3} X_a$, let $Y := \bigcup_{a \in \Sigma_3} Y_a$, and observe that the three preceding properties imply that $|X| = |Y| = 3\tau(w)/2$. The sets X_a and Y_a will be chosen so that $X \cup Y$ satisfies the following *global independence constraint*: There is no pair $i, j \in X \cup Y$ such that $i - j = 1$. We now explain why, with appropriately chosen ϵ and r_0, it is always possible to find sets X_a and Y_a (for each $a \in \Sigma_3 \setminus \{x\}$) that satisfy these three properties.

We choose the elements of $X \cup Y$ using the following greedy strategy. Suppose X_a or Y_a does not yet contain enough elements, for some $a \in \Sigma_3$. Without loss of generality, assume it is X_a. Choose an arbitrary element from $X_a^+ \setminus \bigcup_{i \in X \cup Y}\{i-1, i, i+1\}$ and add it to X_a.

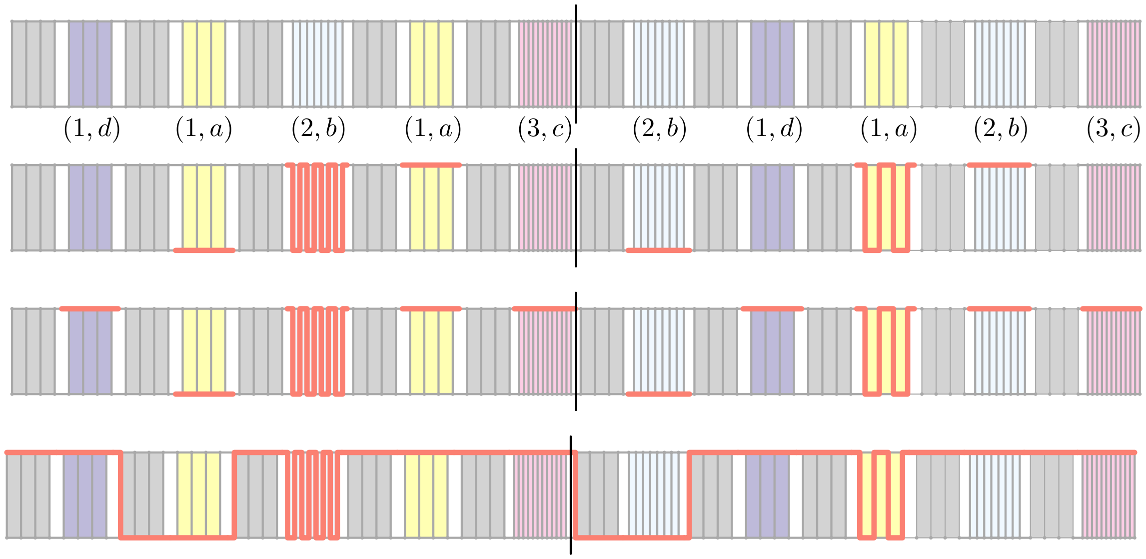

Figure 3.3: Constructing the anagramish path P: (1) pairs of top and bottom paths are matched with zig-zag paths; (2) all remaining colourful blocks receive top paths; (3) all boring blocks receive top, updown, or downup paths.

To see that this is always possible observe that at the beginning of each step,

$$\begin{aligned}\left| X_a^+ \setminus \bigcup_{i \in X \cup Y} \{i-1, i, i+1\} \right| &\geq |X_a^+| - 3|X \cup Y| \\ &\geq |X_a^+| - 3\,(3\tau(w) - 1) \\ &\geq |X_a^+| - 9\epsilon r - 3\ ,\end{aligned}$$

so there will always be an element to choose provided that $|X_a^+| > 9\epsilon r + 3$. Since $w \in L$, w is ℓ-periodic, so $|X_a^+| \geq \lfloor (r-1)/\ell \rfloor$. Therefore, this procedure will succeed in finding sets X and Y with the global independence property provided that $9\epsilon r - 3 < \lfloor (r-1)/\ell \rfloor$. In particular, $\epsilon < 1/(9\ell)$ satisfies this requirement.

To simplify notation, let $G := G_{n(w)}$. For each $i \in \{1, \dots, 2r\}$, let H_i be the subgraph of G induced by the vertices corresponding to w_i; the subgraphs $H_1, \dots, H_{2r}$ are referred to above as colourful blocks. We now construct the anagramish path P in a piecewise fashion, as illustrated in Figure 3.3.

- For each $a \in \Sigma_3$ such that $\delta_a(w) > 0$, group the elements of X_a into pairs. For each

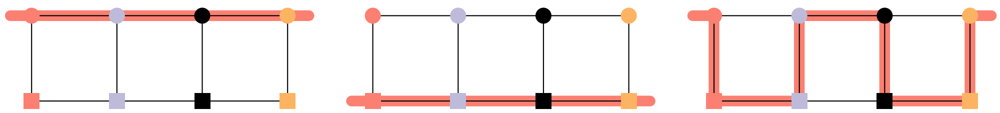

Figure 3.4: Subpaths of P through colourful blocks: A pair of top and bottom paths contribute the same amount as a single zig-zag path.

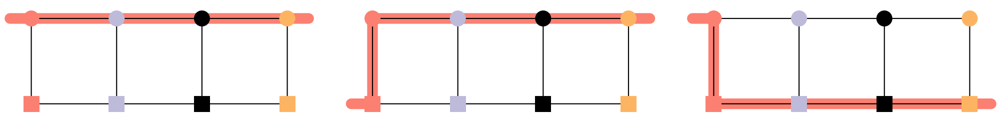

Figure 3.5: The paths taken by P through boring blocks: A top path, a downup path, and an updown path.

pair (i, j), P contains the path through the top row of H_i and the path through the bottom row of H_j. For each element $i \in Y_a$ the path P contains the zig-zag path with both endpoints in the top row of H_i and that contains every vertex of H_i.

- For each $a \in \Sigma$ such that $\delta_a(w) < 0$ we proceed symmetrically to the previous case, but reversing the roles of X_a and Y_a. Specifically, we group the elements of Y_a into pairs. For each pair (i, j), P contains the path through the top row of H_i and the path through the bottom row of H_j. For each element $i \in X_a$, P contains the zig-zag path with both endpoints in the top row of H_i and that contains every vertex of H_i.

- For each $i \in \{1, \dots, 2r\} \setminus (X \cup Y)$, P contains the top row of H_i.

The rules above define the intersection, P_i, of P with each colourful block H_i of G. If P_i is the path through the bottom row of H_i then we call H_i a *bottom block*. If P_i is the path through the top row of H_i then we call H_i a *top block*. If P_i is the zig-zag path that contains every vertex of H_i then we call H_i a *zig-zag block*. Note that $\sum_{a \in \Sigma_3} \delta_a(w) = 0$ and this implies that the number of bottom blocks among $H_1, \dots, H_{r-1}$ is the same as the number of bottom blocks among $H_{r+1}, \dots, H_{2r}$. Indeed, this number is exactly $\frac{1}{2}\tau(w)$.

We now define how P behaves for the boring blocks, denoted by $Q_0, \dots, Q_{2r-1}$. The first boring block Q_0 comes immediately before H_1. Each boring block Q_j, for $j \in \{1, \dots, 2r-1\}$ comes immediately after H_j and immediately before H_{j+1}. In almost every case, P uses the path through the top row of Q_j. The only exceptions are when H_j or H_{j+1} are bottom blocks. Note that, because of the global independence constraint, these two cases are mutually exclusive. See Figure 3.5.

- When H_j is a bottom block, P uses a path that begins at the bottom row of Q_j but moves immediately to the top row of Q_j and uses the entire path along the top row. We call this a *downup* path.
- When H_{j+1} is a bottom block, P uses a path that begins at the top row of Q_j and moves immediately to to the bottom row of Q_j and uses the entire path along the bottom row. We call this a *updown* path.

This completely defines the path $P := v_1, \dots, v_{2m}$. All that remains is to argue that the word $\rho := \phi_w(v_1), \dots, \phi_w(v_{2m})$ is anagramish.

Observe that the number of downup paths and the number of updown paths that appear in $Q_0, \dots, Q_{r-1}$ is exactly the same as the number of bottom blocks among $H_1, \dots, H_{r-1}$ which is exactly $\frac{1}{2}\tau(w)$. Similarly, the number of updown paths and downup paths that appear in $Q_{r+1}, \dots, Q_{2r-1}$ is exactly $\frac{1}{2}\tau(w)$. Now every path that is neither downup nor updown uses the top row. This implies that the sequence of colours contributed to ρ by the intersection of P with $Q_0, \dots, Q_{r-1}$ is a permutation of the sequence of colours contributed to ρ by the intersection of P with $Q_{r+1}, \dots, Q_{2r-1}$. These two sequences cancel perfectly.

Next, by construction, each pair of top and bottom blocks in $H_1, \dots, H_{r-1}$ contributes exactly the same amount as a single matching zig-zag block in $H_{r+1}, \dots, H_{2r-1}$. Specifically, if $x, y \in X_a$, $z \in Y_a$, H_x is a top block, H_y is a bottom block and H_z is a zig-zag block, then the contributions of P_x and P_y to ρ cancels out the contribution of P_z. After doing this cancellation exhaustively, all that remains are top blocks, which also cancel each other perfectly. This completes the proof. □

To be explicit, we finish the proof of Theorem 3.1 by pointing out the contradiction between Lemma 3.1 and Lemma 3.3:

Proof of Theorem 3.1. Assume for the sake of contradiction that there exists some $c \in \mathbb{N}$ such that $\dot{\chi}_\pi(G_n) \leq c$ for each $n \in \mathbb{N}$. Then the language L_1 is infinite and, by Lemma 3.2 so are L_2 and L_3. Therefore L_3 is an infinite periodic language. Therefore, Lemma 3.1 implies that for every $\epsilon > 0$ and $r_0 \in \mathbb{N}$ there exists a word $w \in L_3$ with $2r := |w| \geq 2r_0$, and $\tau(w) \leq \epsilon r$. In particular, this is true for the specific values of r_0 and ϵ that appear in Lemma 3.3. However, Lemma 3.3 implies that ϕ_w is not an anagram-free colouring of $G_{n(w)}$, which contradicts the fact that $w \in L_3$. □

3.4 Proof of Lemma 3.1

All that remains is to prove Lemma 3.1, which we do now.

Proof of Lemma 3.1. First observe that, for any ℓ-periodic word w, $|\Sigma_w| \le \ell$. Define an even-length word t to be *a-unbalanced* if $\tau_a(t) > \epsilon|t|/\ell$ and *a-balanced* otherwise. If t is a-balanced for each $a \in \Sigma_t$ then t is *balanced*. Observe that, if t is balanced then $\tau(t) \le |\Sigma_t|\epsilon|t|/\ell \le \epsilon|t|$. A word is *everywhere unbalanced* if it contains no balanced factor of length $r \ge r_0$. Our goal therefore is to show that there is an upper bound $n := n(\ell, \epsilon, r_0)$ on the length of any ℓ-periodic everywhere unbalanced word.

Let h be a positive integer whose value will be discussed later and let $n := r_0 2^h$. Let w be an ℓ-periodic everywhere unbalanced word of length n over the alphabet Σ. Assume, without loss of generality, that r_0 is a multiple of ℓ.

Consider the complete binary tree T of height h whose leaves, in order, are length-r_0 words whose concatenation is w and for which each internal node is the factor obtained by concatenating the node's left and right child. Note that for each $v \in V(T)$ and each $a \in \Sigma$, the fact that w is ℓ-periodic and r_0 is multiple of ℓ implies that $|v|_a \ge |v|/\ell$.

For each $a \in \Sigma$, let $S_a := \{v \in V(T) : v \text{ is } a\text{-unbalanced}\}$. Since w is everywhere unbalanced, $\bigcup_{a\in\Sigma} S_a = V(T)$. Therefore,

$$(h+1)n = \sum_{v\in V(T)} |v| \le \sum_{a\in\Sigma}\sum_{v\in S_a} |v| \ .$$

Therefore, there exists some $\alpha \in \Sigma$ such that $\sum_{v\in S_\alpha} |v| \ge (h+1)n/|\Sigma| \ge (h+1)n/\ell$. At this point we are primarily concerned with appearances of α, so let $X := S_\alpha$, and, for each node $v \in V(T)$, let $W(v) := |v|_\alpha$.

For each non-leaf node v of T, let $R(v)$ denote a child of v such that $W(R(v)) \le \frac{1}{2}\cdot W(v)$. It is helpful to think of T as being ordered so that each right child y with sibling x has $W(y) \le W(x)$. For a non-leaf node $v \in X$ the fact that v is α-unbalanced implies that

$$W(R(v)) \le \tfrac{1}{2}\cdot W(v) - \tfrac{\epsilon}{2\ell}\cdot|v| \le (\tfrac{1}{2} - \tfrac{\epsilon}{2\ell})W(v) \ .$$

From this point on we use the following shorthands. For any $S \subseteq V(T)$, $L(S) := \sum_{v\in S}|v|$, $W(S) := \sum_{v\in S} W(v)$, and $R(S) = \{R(v) : v \in S\}$. Summarizing, we have a complete binary tree T of height h and $X \subseteq V(T)$ with the following properties:

- For each $v \in V(T)$, $W(v) \ge |v|/\ell$.

- $L(X) \geq (h+1)n/\ell$.
- For each non-leaf node $v \in X$, $W(R(v)) \leq (\frac{1}{2} - \frac{\epsilon}{2\ell})W(v)$.

For each $i \in \{0,\ldots,h\}$, let $X_i \subseteq X$ denote the set of nodes $v \in X$ for which the path from the root of T to v contains exactly i nodes in X, excluding v. See Figure 3.6. Observe that, since each node in X_i has an ancestor in X_{i-1},

$$n \geq L(X_0) \geq L(X_1) \geq \cdots \geq L(X_h) \ .$$

We will show that there exists an integer $t := t(\epsilon, \ell, r_0)$ such that, for each $i \in \{0,\ldots,h-t\}$,

$$L(X_{i+t}) \leq (1-(1/2)^{t+1})L(X_i) \ . \tag{3.2}$$

In this way,

$$\begin{aligned}
(h+1)n/\ell \leq L(X) &= \sum_{i=0}^{h} L(X_i) \\
&\leq \sum_{i=0}^{h} L(X_{t\lfloor i/t \rfloor}) && \text{(since } i \geq t\lfloor i/t \rfloor\text{, so } L(X_i) \leq L(X_{i\lfloor i/t \rfloor})\text{)} \\
&= t \cdot \sum_{i=0}^{h/t} L(X_{it}) && \text{(for } h \text{ a multiple of } t\text{)} \\
&\leq t \cdot \sum_{i=0}^{\infty} (1-(1/2)^{t+1})^i L(X_0) && \text{(by Equation (3.2))} \\
&\leq tn \cdot \sum_{i=0}^{\infty} (1-(1/2)^{t+1})^i && \text{(since } |X_0| \leq n\text{)} \\
&= tn2^{t+1}
\end{aligned}$$

which is a contradiction for sufficiently large h; in particular, for $h > \ell t 2^{t+1} - 1$.

It remains to establish Equation (3.2), which we do now. Fix some $i \in \{0,\ldots,h-1\}$, define $A_0 := X_i$ and, for each $j \geq 1$, define A_j to be the subset of X_{i+j} that are descendants of some node in $R(A_{j-1})$. See Figure 3.6. To upper bound $L(X_{i+t})$ observe that X_{i+t} can be split into two sets A'_0 and B defined as follows: The nodes A'_0 do not have an ancestor

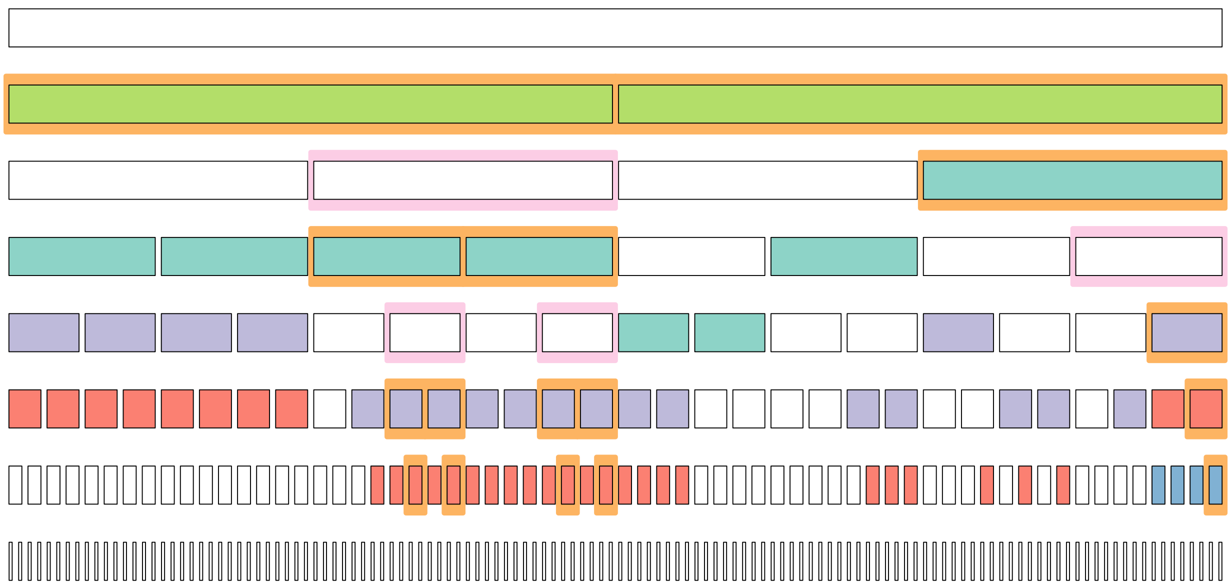

Figure 3.6: The partitioning of X into $X_0, \ldots, X_h$. Shaded nodes are in X and all nodes in X_i are shaded with the same colour. Starting with $A_0 = X_0$, the elements of $A_0, \ldots, A_h$ are highlighted (in orange). The elements of $R(A_0), \ldots, R(A_h)$ are also highlighted (in pink).

in $R(A_0)$ and therefore $L(A'_0) \leq (1/2)L(A_0)$. The nodes in B do have an ancestor in $R(A_0)$ and therefore have an ancestor in A_1. Iterating this argument, we obtain

$$
\begin{aligned}
L(X_{i+t}) &\leq \sum_{j=0}^{t-1} (1/2)L(A_j) + L(A_t) \\
&\leq \sum_{j=0}^{t-1} (1/2^{j+1})L(A_0) + L(A_t) \\
&= (1 - (1/2)^t)L(A_0) + L(A_t) \\
&= (1 - (1/2)^t)L(X_i) + L(A_t) \ .
\end{aligned}
$$

So all that remains to establish 3.2 is to prove that $L(A_t) \leq (1/2)^{t+1}L(X_i)$. To do this, observe that, for each $j \in \{1, \ldots, t\}$,

$$
W(A_j) \leq W(R(A_{j-1})) \leq (\tfrac{1}{2} - \tfrac{\epsilon}{2\ell}) \cdot W(A_{j-1}) \tag{3.3}
$$

which implies

$$
W(A_t) \leq (\tfrac{1}{2} - \tfrac{\epsilon}{2\ell})^t W(A_0) \leq (\tfrac{1}{2} - \tfrac{\epsilon}{2\ell})^t \cdot L(A_0) = (\tfrac{1}{2} - \tfrac{\epsilon}{2\ell})^t \cdot L(X_i)
$$

Since w is ℓ-periodic,

$$
L(A_t) \leq \ell \cdot W(A_t) \leq \ell \cdot (\tfrac{1}{2} - \tfrac{\epsilon}{2\ell})^t \cdot L(X_i) \leq L(X_i)/2^{t+1}
$$

for $t := \lceil \log(2\ell)/\log\bigl(1/(1 - \tfrac{\epsilon}{\ell})\bigr) \rceil$. $\square$

3.5 Reflections

Although an explicit upper bound on $n := n(\epsilon, \ell, r_0)$ could be extracted from the proof of Lemma 3.1 it would likely be far from tight. We suspect that there is a Fourier analytic proof that would give better quantitative bounds. We have not pursued this, because we have no idea how to explicitly upper bound ℓ, for reasons discussed in the next paragraph.

Lemma 3.2 and its proof give absolutely no clues to help find a concrete bound on ℓ or to find a minimal set Ξ. Indeed, for some choices of L, doing so can be a difficult problem. Consider the example where $|\Sigma| = 5$ and L is the language of all anagram free words in Σ^*. It is easy to see that this language L is factor-closed and the result of [41], published in 1970, shows that this L is infinite. The question of whether $|\Xi| = 4$ or $|\Xi| = 5$ is then the question of determining whether there exist arbitrarily long anagram-free words on an alphabet of size 4. This was the open problem posed by [18] in 1961 and again by [9] in 1971 and not resolved until 1992 when [30, 31] showed that the answer, in this case, is that $|\Xi| = 4$. However, if this were not the case, then determining ℓ would be the question of determining the length of the longest anagram-free word over an alphabet of size 4.

Our proof uses Lemma 3.2 twice and each uses a language L_3 that is considerably more complicated than the language of anagram-free words. It seems unlikely that we will obtain concrete upper bounds on ℓ as a function of c except, possibly, through the use of computer search. The resulting value ℓ is used in the application of Lemma 3.1 and also within the proof of Lemma 3.3.

Acknowledgement

We would like to thank an anonymous referee for pointing out existing results and terminology from the field of combinatorics on words that has helped streamline the presentation in Sections 3.2 and 3.3.

Chapter 4

Geodesic Anagram-Free Colouring of Chordal Graphs

GEODESIC ANAGRAM-FREE COLOURING OF CHORDAL GRAPHS.[1]

Saman Bazargani[2] Pat Morin[3]

[1]This research was partly funded by NSERC.

Abstract. A sequence of even length characters s is called an *anagram* if for every character c in s the number of occurrences of c in the first half of s is equal to the number of occurrences of c in the second half of s. Define a sequence s to be an *anagram*-free such that any subsequence of s is not an anagram. A colouring of a path p in a graph is an *anagram colouring* if the sequence corresponding to the colours of the vertices in p is an anagram. A graph colouring is an *anagram-free colouring* if the graph does not contain any anagram path. The anagram-free chromatic number of a graph is the minimum number of colours required to colour a graph anagram-freely. It is known that this number is unbounded for graphs of pathwidth ≥ 2 [6, 11]. Moreover, Wilson and Wood [49] showed that the anagram-free chromatic number of a tree T is upper-bounded by $4p + 1$ where p is the pathwidth of T. Define the *geodesic anagram-free chromatic number* of a graph to be the minimum number of colours required to colour a graph such that all shortest paths among any pair of vertices are coloured anagram-freely. In this chapter we give an upper bound on the geodesic anagram-free chromatic number of specific graph classes. A graph is *chordal* if every cycle of length at least 4 has a chord. We prove that the geodesic anagram free chromatic number of a chordal graph G is $32p'w$ where p' is the minimum pathwidth of a tree used in any subtree intersection representation of G and w is the clique number of G. Our result generalizes the theorem by Wilson and Wood [49]. Moreover, we show that the geodesic anagram-free chromatic number of an interval graph is bounded by $32p$ where p is the pathwidth of the interval graph.

4.1 Introduction

A sequence $w := w_1 w_2 \dots w_n$ is *nonrepetitive* if no two adjacent blocks (sub-sequences) are the same. In 1906 Thue [45] managed to construct an arbitrary long nonrepetitive sequence using only 3 symbols. Thue's work influenced developing numerous works in Combinatorics, specifically on Combinatorics of Words. A *vertex colouring(colouring)* of a graph G is a mapping $c : V(G) \to C$ that assigns each vertex v of G to a colour in some set C of colours. Call this a $|C|$*-colouring* of G. In 2002, Alon, Grytczuk, Hałuszczak, and Riordan [34] introduced a new graph colouring inspired by Thue's work. Let G be a graph. Define a vertex colouring $c : V(G) \longrightarrow C$ nonrepetitive if the sequence of colours corresponding to any path in G is nonrepetitive, where C is the set of colours.

In the literature, nonrepetitive colouring has been studied extensively in terms of path-nonrepetitive colouring, walk-nonrepetitive colouring, stroll-nonrepetitive. In 2002 Alon et al. [3] showed that the nonrepetitive chromatic number of a graph G is upper-bounded by

$c\Delta(G)^2$, where c is a constant and $\Delta(G)$ is the maximum degree of G. In 2016, Dujmović et al. [16] improved this bound to $\Delta(G)^2$ using entropy compression.

In the field of combinatorics on word, non-repetitive strings are also called square-free. A variant on this idea is abelian square-free words, a.k.a anagram-free words, which we now define.

A sequence $s := s_1, s_2, \ldots, s_{2r}$ is an anagram if the sequence $s_1, s_2, \ldots, s_r$ is a permutation of $s_{r+1}, s_{r+2}, \ldots, s_{2r}$. Notice that odd length sequences are not an anagram. A sequence s is an anagram-free sequence if no sub-sequence of s is an anagram.f Let G be a graph. Define a vertex colouring $c : V(G) \longrightarrow C$ anagram-free if the sequence of colours corresponding to any path in G is anagram-free, where C is the set of colours.

In 2018, Carmi, Dujmović, and Morin gave examples of graphs of pathwidth 3 that have unbounded anagram-free chromatic number. In 2021 Bazargani, Dujmović and Morin showed that graphs of pathwidth 2 also have unbounded anagram-free chromatic number, while Wilson and Wood [49] showed that the anagram-free chromatic number of a tree is bounded by its pathwidth in 2016. In particular, they proved that any tree can be coloured anagram-freely with $4p$ colours where p is the pathwidth of the tree.

As stated above, the anagram-free chromatic number is bounded by pathwidth in the case of trees. Also, the anagram-free chromatic number is a constant in the case of a path. Observe that the shortest path between any two vertices in either of these two graph classes is unique. Motivated by the works of Bazargani et al. [6] Carmi et al. [11] and Wilson et al. [49] and the above observation, a natural restriction can be added to the setting of anagram-free colouring. Hence, in this chapter we study the following colouring problem: Define the *geodesic anagram-free chromatic number* of a graph to be the minimum number of colours required such that the sequence of colours corresponding to any shortest path connecting any pair of vertices is anagram-free.

Theorem 4.1. *Any chordal graph G that can be represented as an intersection graph of subtrees of a tree T $(V_t : t \in V(T))$ has anagram-free chromatic number of $32p'w \in O(p'w)$, where p' is $\mathrm{pw}(T)$ and w is the clique number of G.*

The rest of the chapter is organized as following: In Section 4.2 we give the definitions of basic required background and the terminology to understand this chapter. In Section 4.3 we present a greedy algorithm that gives a geodesic anagram free colouring of an interval graph. Next, we show why this colouring is geodesic anagram free and we give an upper bound on the geodesic anagram-free chromatic number of interval graphs . In Section 4.4 we prove that chordal graphs have bounded geodesic anagram-free chromatic number. Next, we give an algorithm that shows how to colour a chordal graph geodesic anagram-freely.

4.2 Definitions

Intersection Graph Let X be a set of sets. The *intersection graph* G of X is the graph G with vertex set $V(G) = X$ and $vw \in E(G)$ if and only if $v \cap w \neq \emptyset$. When X is a set of real closed intervals the intersection graph of X is called an *interval graph*. Let T be a tree and let X be a set of subtrees of T. Then, the intersection graph of X is called a *chordal graph*. This definition of a chordal graph is equivalent to the previous definition. For more details, please refer to [20].

For an interval (vertex) i define $i[0]$ (resp. $i[1]$) to be the left endpoint (resp. right endpoint) of the interval i. For two intervals a and b, we say that a is to the right (resp. left) of b if $a[0] > b[0]$ and $a[1] > b[1]$ (resp. $a[0] < b[0]$ and $a[1] < b[1]$). An interval is called the *leftmost* (resp. *rightmost*) interval if its starting point (resp. end point) has the minimum (resp. maximum) value among all intervals.

Let $G = (V, E)$ be a graph. The length of a path $v_0 \dots v_r$ is r (the number of edges) where $v_i \in V$. Define $dist(u, w)$ to be the length of a shortest path between two vertices $u, w \in V$ (minimum number of edges that connect u to v.) We use the convention that if the graph is not connected and u and w are in two different disconnected components of the graph then $dist(u, w) = \infty$. A path is called *geodesic* if that path is a shortest path connecting two vertices.

Let $S \subseteq V$ be any subset of vertices of G. Then the *induced subgraph* $G[S]$ is the graph whose vertex set is S and whose edge set consists of all of the edges in E that have both endpoints in S.

Two vertices x, y are *neighbours* if xy is an edge of G. The set of neighbours of a vertex v in G is denoted by $N(v)$.

Tree Decomposition A tree decomposition of $G = (V, E)$ consists of a tree T and a subset $V_t \subseteq V$ associated with each node $t \in T$. We will call the subsets V_t bags of the tree decomposition. T and $\{V_t \colon t \in T\}$ must satisfy:

- (Node coverage) Every node of G belongs to at least one bag V_t.
- (Edge coverage) For every edge e of G, there is some bag V_t containing both ends of e.
- (Coherence) $T[\{t \in V(T) \colon v \in V_t\}]$ is connected for every $v \in V(G)$.

Treewidth The *width* of a tree decomposition T is $max_t|V_t|-1$, when V_t are the bags of T. The *treewidth* $tw(G)$ of a graph G is the minimum k such that G has a tree-decomposition of width k.

Pathwidth A *path decomposition* is a tree decomposition T for which T is a path. The *pathwidth* $\mathrm{pw}(G)$ of a graph G is the smallest k such that G has a path decomposition T of width k. Note that the treewidth of a chordal graph is equal to the size of its largest clique minus one. Similarly the pathwidth of an interval graph is equal to the size of its largest clique minus one.

Language An alphabet Σ is a finite and non-empty set. Each elements of Σ is called a *letter*. A *word* is a finite sequence of letters chosen from Σ. A *factor* of a word w is a sub-sequence of w. A *Language* L over Σ is a set of words over Σ. Let $s := s_1 s_2 \dots s_{n-1}$ be a sequence of characters. We denote the character at the index i by $s[i]$. Also, we denote the sub-sequence of s from the character located at the index i to j by $s[i : j] = s_i, s_{i+1} \dots s_j$. Define $|s|$ the length of the sequence s.

4.3 Geodesic anagram-free chromatic number of interval graphs

In this section, we prove that an interval graph has a geodesic anagram-free chromatic number, that is linear in its pathwidth. For an interval graph $G = (V, E)$, let P^* be the shortest path that connects the leftmost interval to the rightmost interval. On an abstract level any shortest path P between any two arbitrary intervals must follow the shadow of P^*, that means for any interval $I \in P$ there exists an interval $I^* \in P^*$ such that $(I, I^*) \in E$. Hence, the aim is to establish a colouring scheme such that the sequence of colours correspond to P be a mapping of sequence of colour correspond to P^*. We do this by presenting a greedy colouring scheme that gives a geodesic anagram free colouring for an interval graph.

Greedy colouring : Let $G = (V, E)$ be an interval graph. Recall that an interval graph is an undirected graph formed from a set of intervals on a real line. We assume without loss of generality that in the interval representation all endpoints of intervals are distinct integers in the range $\{1, \dots, 2|V|\}$. For a vertex v of G define $\tau(v)$ as follows:

1. If there exists an interval v' such that $(v', v) \in E$ and v' is to the right of v then $\tau(v) = \arg\max_{v' \in N(v)} v'[1]$.
2. If v is not connected to any intervals to its right then $\tau(v) = \arg\min_{v' \in V} v'[0]$ where $v'[0] > v[1]$.
3. Otherwise, $\tau(v) =$ None (indicating that v is the rightmost interval in the graph).

Layering the graph Let G be a connected interval graph with $\mathrm{pw}(G) = p$. Let the vertex v be the leftmost interval of G. Define P_1 as the vertex sequence of G defined by vertices $\{v, \tau(v), \tau(\tau(v)), \ldots, v'\}$ where v' is the rightmost interval of G. By the definition of τ; $G[P_1]$ is a path. Let G_1 be $G - \{P_1\}$. Let $P_2, \ldots, P_{p'}$ and $G_2, \ldots, G_{p'}$ be vertex sequences and induced subgraphs achieved by repeating the above process (taking a vertex sequence P_i that starts at the leftmost vertex of G_i, then uses τ repeatedly to get to the rightmost vertex of G_i. Next P_i is removed from G_i to get G_{i+1}) until no vertex remains in the graph, where $p' \leq p$. We call $P_1, P_2, \ldots, P_{p'}$ as first layer, second layer, $\ldots$, $p' - th$ layer. We get at most p layers since no two layers have a vertex in common and removing a layer from G reduces the pathwidth of G_i at least by one. Denote $P_i[j]$ as the j-th element of P_i for $i \in \{1, \ldots, p'\}$.

Observation 4.1. *For any integer i, at most two intervals of P_k contain i.*

Colouring the first layer In 1992, Keränen [30] proved the existence of an infinitely long anagram-free word using only four letters. This implies that a path has an anagram-free colouring using four colours.

4.3.1 Double anagram free sequence

Here we introduce a slightly different anagram-free colouring that is based on Keränen's [30] work.

Let L be an anagram free language over the alphabet $\Sigma = \{a, b, c, d\}$. Let $g : \Sigma^* \to \Sigma'^*$ be an endomorphism where $\Sigma'^* = \{a, a', b, b', c, c', d, d'\}$. We define g as following:

$$g(a) = aa', g(b) = bb', g(c) = cc', g(d) = dd'$$

Define a language L' as all factors of $g(w)$ for every word w in L. We call L' *double anagram free*.

Lemma 4.1. *A double anagram free language L' is anagram free.*

Proof. Let s be a word from L'. Let s_{even} (resp. s_{odd}) be a sequence of characters from s which their indices are even (resp. odd). Notice that the ordering of characters in s_{even} (resp. s_{odd}) will remain the same as s. Without loss of generality, assume that $s_{even} \in L$ thus, s_{even} is anagram-free. By the definition of L', $s_{odd} \in \Sigma'^*$ is also anagram-free since the characters of s_{odd} follow the same sequence as their corresponding word in L. Here we analyze two cases, when $|s|$ is even or odd.

- $|s|$ is even: Since s_{even} is anagram-free, s will be anagram free, as the rest of characters used in s are in $\Sigma'^* - \Sigma^*$. In other word there is a character $c \in s_{even}$ that appears more in one half of the sequence.
- $|s|$ is odd: An odd sequence cannot be an anagram.

□

Lemma 4.2. *Let s be a double anagram-free word. If we insert a character $c \in \Sigma'^*$ into any location in s, the resulting word s remains anagram-free.*

Proof. If the length of s is even, adding a character c makes the length of s odd, and by the definition of an anagram sequence, an odd sequence is not an anagram. Let s_{even} (resp. s_{odd}) be a sequence of characters from s which their indices are even (resp. odd). Notice that the ordering of characters in s_{even} (resp. s_{odd}) will remain the same as s. Without loss of generality, assume that $s_{even} \in L$. As stated above both s_{even} and s_{odd} are anagram free. Since both s_{even} and s_{odd} are constructed based on distinct alphabet and both are anagram-free. It implies that each of them has at least one unique character that appears more in one half of the sequence s than the other half of the sequence s. So, any double anagram-free sequence has two such characters. Thus, after inserting one character to s the new sequence remains anagram-free. □

Let the mapping (colouring) $\gamma : V(P_1) \to [1, 8]$ be a double anagram free colouring of P_1.

Lemma 4.3. *Any vertex $v \in V - P_1$ can only be connected to at most three consecutive vertices of P_1.*

Proof. Let $P_1 = v'_0, v'_1 \ldots, v'_{p-1}(p = |P_1|)$ and let v'_i be the leftmost vertex of P_1 that v is connected to. Supposed that v is connected to at least 4 consecutive vertices of P_1. This implies that $\tau(v)[1] > v'_{i+3}[0] > v'_{i+1}[1] > \tau(v'_i)[1]$. That is a contradiction with the choice of v'_{i+1}. $\square$

We will show that for each vertex $v \in V(G)$, there exists an integer j such that v falls into one of the following 5 categories. Notice that by our choice of intervals in P_1, any interval $v \in V(G)$ intersects with at least an interval of P_1 and by Lemma 4.3 v intersects at most 3 intervals of P_1. The interval corresponding to v is $P_1[j]$ if:

- Category 1: The only vertex of P_1 adjacent to v is $P_1[j]$. See Figure 4.1.

Figure 4.1: Category 1

- Category 2: The only two vertices of P_1 adjacent to v are $P_1[j]$ and $P_1[j-1]$ and $v[1] > P_1[j-1][1]$. See Figure 4.2.

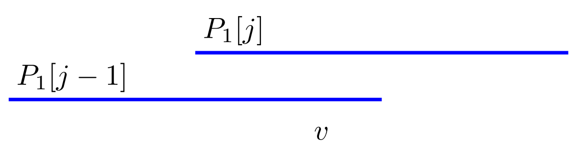

Figure 4.2: Category 2

- Category 3: The only two vertices of P_1 adjacent to v are $P_1[j]$ and $P_1[j+1]$ and $v[1] < P_1[j][1]$. See Figure 4.3.

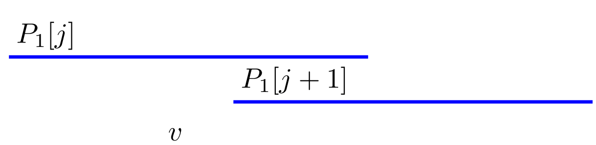

Figure 4.3: Category 3

- Category 4: The only three vertices of P_1 adjacent to v are $P_1[j-1]$ and $P_1[j]$ and $P_1[j+1]$. See Figure 4.4.

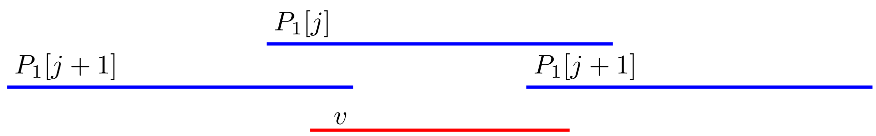

Figure 4.4: Category 4

- Category 5: v is $P_1[j]$.

For a vertex v denote $\Xi(v)$ to be the function that gives the corresponding vertex to v with respect to the five categories defined above.

We say that an interval is *contained* in another interval if it falls into category 1 or category 3. We say that an interval is a *bridge* if it falls into categories 2, 4 or 5, in which case we say that the bridge connects two intervals. In category 2 a bridge connects $P_1[j-1]$ to $P_1[j]$. In category 4 a bridge connects $P_1[j]$ to $P_1[j+1]$.

Box Representation of intervals The *box representation* of a layer of an interval graph is a partitioning of the intervals of that layer into a sequence of ordered sets such that all intervals in a set have the same relation with a unique interval of the first layer. These relations are defined as follows:

For a layer $K \in \{1, \ldots, p'\}$ let $\{B_{K,0}, B_{K,1}, \ldots, B_{K,n_0-1}\}$ s.t. $n_0 \leq |P_1|$ be the sequence for the boxes of the layer K.

1. K $= 1$: $B_{1,j} = \{P_1[j]\}$ for $0 \leq j \leq |P_1| - 1$.
2. K ≥ 2 : $B_{K,j} = \{i \in P_k : \Xi(i) = P_1[j]\}$ for $0 \leq j \leq |P_1| - 1$.

Notice that the box $B_{K,j}$ is the corresponding box for some intervals of the K-th layer that correspond $P_1[j]$. Moreover, observe that all of the intervals in $B_{K,j}$ are connected to $P_1[j]$. We order the intervals in each box so that the intervals inside each box maintain the same ordering as their original ordering in P_k.

Lemma 4.4. *A box B can have at most 2 bridges.*

Proof. Without loss of generality assume that $B = B_{k,i+1}$ for two integers i and k. Suppose that $B_{k,i+1}$ has more than two bridges. Recall that for an integer j any bridge that is in $B_k[j]$ is adjacent to $P_1[j-1]$ and $P_1[j]$ (categories 2 and 4), thus every bridge in $B_{k,i+1}$ is adjacent to $P_1[i]]$ and $P_1[i+1]$. Let I_1, I_2, I_3 be the three bridges that connect $P_1[i]$ to $P_1[i+1]$. By the definition of a bridge all three intervals I_1, I_2, I_3 should contain $P_1[i][1]$. This contradicts with Observation 4.1. □

Observation 4.2. *At most two bridges from a layer P_i can have the same corresponding vertex. This case happens when the first bridge is from category 2 (the bridge that connects two consecutive vertices of P_1, say $P_1[j-1], P_1[j]$), and the second bridge is from category 4 (the bridge that connects three consecutive vertices of P_1, say $P_1[j-1], P_1[j], P_1[j+1]$).*

In the following section we define a colouring scheme for vertices of G. As we demonstrate later the colour that the intervals inside a box take can be seen as a mapping of colours to the colours of intervals in the first layer. Define the colouring $\Pi : V \to C$ such that $C = [1, \mathrm{pw}(G)] \times [1, 8] \times [0, 3]$. Let $P_k[j]$ be the m-th interval of the box $B_{k,i}$.

1. If $P_k[j]$ falls into categories 1 or 3 with respect to the first layer: $\Pi(P_k[j]) := (k, \gamma(P_1[i]), (m \mod 2) + 2)$.
2. If $P_k[j]$ falls into categories 2 or 4 with respect to the first layer: $\Pi(P_k[j]) := (k, \gamma(P_1[i+1]), m \mod 2)$.
3. If $P_k[j]$ falls into category 5 with respect to the first layer: $\Pi(P_k[j]) := (1, \gamma(P_1[i]), 0)$.

4.4 Geodesic Anagram Free Colouring

In this section we prove that Π is a geodesic anagram free colouring of G.

Let $P^* = v_s, P^*[1], \ldots, v_t = P^*[0], P^*[1], \ldots, P^*[n_0]$ be the shortest path between two arbitrary vertices (intervals) v_s, v_t. Let s^* be the index of the rightmost vertex (interval) of the first layer connected to v_s. Let t^* be the index of the leftmost vertex (interval) of the first layer connected to v_t. Let seq_1 be the sequence of vertices (intervals) from the first layer that connect $P_1[s^*]$ to $P_1[t^*]$ ($seq_1 := P_1[s^* : t^*]$). Let set_1 be the set of vertices (intervals) of seq_1.

Let $seq_2 = (\Xi(P^*[0]), \Xi(P^*[1]), \Xi(P^*[2]), \ldots, \Xi(P^*[n_0]))$. Let set_2 be the set of vertices (intervals) of seq_2.

Lemma 4.5. $set_1 \subseteq set_2$.

Proof. Suppose that there is an interval $P_1[k] \in set_1 - set_2$ for an integer k. By the definition of a path and an interval graph the union of intervals of P^* form a continuous interval. This union includes an interval which starts at $v_s[0]$ and ends at $v_t[1]$. Notice that $P_1[k][0] > v_s[0]$ otherwise, $P_1[k]$ will be the corresponding vertex to v_s. Similarly, the end point of $P_1[k]$ can not be greater than $v_t[1]$ otherwise, $P_1[k]$ will be the corresponding vertex to v_t.

Let v_{i^*} be the rightmost interval in P^* that contains $P_1[k][0]$ thus, $v_{i^*}[0] < P_1[k][0]$ and let v_{j^*} be the leftmost interval in P^* that contains $P_1[k][1]$. Such two vertices v_{i^*}, v_{j^*} exist since the union of vertices in P^* form a continuous interval and $P_1[k][0] > v_s[0]$ and $P_1[k][1] < v_t[1]$. Since $P_1[k]$ is not the corresponding vertex to v_{i^*} it implies that $v_{i^*}[1] < P_1[k-1][1]$, moreover, since $\tau(P_1[k-1]) \neq v_{j^*}$ it implies that $v_{j^*}[0] > P_1[k-1][1]$. Thus, $v_{j^*}[0] > v_{i^*}[1]$ that means v_{j^*} is not connected to v_{i^*}. Moreover, $P_1[k-1][1]$ is contained in neither v_{i^*} nor v_{j^*}. Since the union of the intervals in P^* make a continuous interval it implies that there is a sequence of intervals (a path) in P^* that connect v_{i^*} to v_{j^*}. Let v^* be a vertex (interval) from that sequence that contains $P_1[k-1][1]$. By the definition of $v_{i^*}, P_1[k]$ and v^*, $v^*[0] > v_{i^*}[0]$ and $P_1[k][1] > v^*[1]$. This implies that $\Xi(v^*) = P_1[k]$ which is a contradiction. See Figure 4.5 for an illustration.

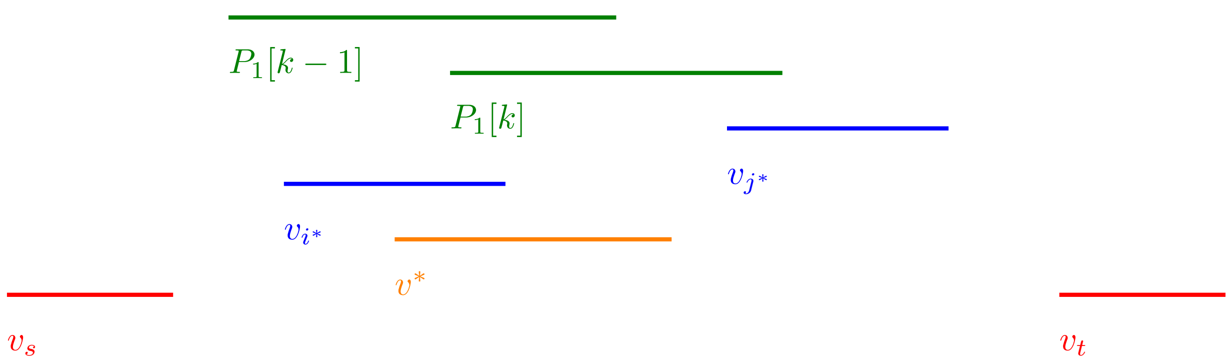

Figure 4.5: $set_1 \subseteq set_2$

□

Lemma 4.6. $|seq_2| - |seq_1| \leq 2$.

Proof. According to our greedy colouring scheme, the path connecting two vertices of P_1 only using the vertices of P_1 is a shortest path between these two vertices. Therefore,

$P_1[s^* : t^*]$ is a shortest path from $P_1[s^*]$ to $P_1[t^*]$. Recall that $|P^*| = |seq_2|$. Suppose that $|seq_2| - |seq_1| \geq 3$. It implies that $|P^*| \geq |seq_1| + 3$. This contradicts with minimality of P^* since the path $v_s.P_1[s^* : t^*].v_t$, that is of length $|seq_1| + 2$, is shorter than P^*. □

Let $\Pi(seq_1) = \Pi(seq_1[0]), \Pi(seq_1[1]), \ldots, \Pi(seq_1[|seq_1| - 1])$ be the sequence of colours of the vertices in seq_1.

Lemma 4.7. *At least $|seq_1| - 2$ intervals in P^* are bridges (the first or maybe second interval of a box).*

Proof. Let I_1, I_2 be two consecutive intervals in P^*, where I_2 is to the right of I_1. If $\Xi(I_1) \neq \Xi(I_2)$ it implies that $\Xi(I_1)[1] > I_2[0]$ and $\Xi(I_1)[1] < I_2[1]$. This implies that I_2 is a bridge. Since all of the intervals of seq_1 are distinct and $|seq_2| - 2 \leq |seq_1|$ and $set_1 \subseteq set_2$. Thus, there are at least $|seq_1| - 2$ consecutive pairs of intervals in P^* that their corresponding vertices are different. This implies P^* has at least $|seq_1| - 2$ bridges.

□

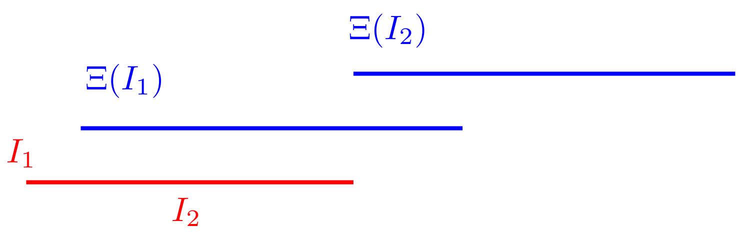

Figure 4.6: P^* has at least $|seq_2| - 2$ bridges

Lemma 4.8. *P^* either has a non-bridge interval or there is at most one $i \in [0, |P^*| - 1]$ such that $\Xi(P^*_i) = \Xi(P^*_{i+1})$.*

Proof. If P^* has a non-bridge interval we are done. According to Observation 4.2 the number of bridges in P^* that correspond to the same interval can be at most 2. Suppose that there are distinct i and j such that $\Xi(P^*_i) = \Xi(P^*_{i+1}) = P_1[s']$ and $\Xi(P^*_j) = \Xi(P^*_{j+1}) = P_1[t']$. Since $set_1 \subseteq set_2$ every interval in $P^*[i+1 : j-1]$ has a unique corresponding vertex (interval), this implies that $|P^*[i+1 : j-1]| = |P_1[s' : t'-1]|$. Moreover, $P_1[t'-1]$ is connected to both $P^*[j]$ and $P^*[j+1]$ since both of them are bridges thus, $P^*[j]$ has to be of category 1 and $P^*[j+1]$ has to be of category 4. This contradicts the minimality of P^* since the path $P^*[i].P_1[s' : t'-1].P^*[j+1]$ is shorter than $P^*[i].P^*[i+1 : j-1].P^*[j].P^*[j+1]$. □

Lemma 4.9. Π *is a geodesic anagram free colouring.*

Proof. According to Lemma 4.8 P^* either has a non-bridge interval or there is at most one $i \in \{0, \ldots, |P^*| - 1\}$ such that $\Xi(P_i^*) = \Xi(P_{i+1}^*)$.

1. P^* has a non-bridge vertex: Let I^* be a non-bridge vertex in P^*. It implies that the third index of I^*'s colour is 2 or 3. This implies that I^* has a distinct colour compared to the colour of any bridge. Since $|seq_2| - |seq_1| \leq 2$ and by Lemmea 4.7 P^* has $|seq_1| - 2$ bridges, either this colour is unique in one half, which means that this sequence is an anagram-free sequence. Otherwise, suppose that if there is a vertex that has the same colour in the other half. Cancel out these two vertices. By Lemma4.2, tge resulting sequence is anagram-free.

   The only exceptional case is when P^* is constituted of only two non-bridge intervals. In this case these two intervals have to be adjacent. If they belong to different layers the values of their first colour indices are different. If they belong to the same layer, it implies that the values of their third colour indices are different.

2. There is at most one $i \in \{0, \ldots, |P^*| - 1\}$ such that $\Xi(P_i^*) = \Xi(P_{i+1}^*)$: According to the colouring scheme the colouring of the first layer is a double-anagram free colouring. Moreover, the value of the second colour index of each vertex is the same as the value of the second colour index of their corresponding interval (vertex). Thus, despite these vertices having a different colour from their corresponding vertices they follow the same pattern as the first layer. seq_1 is a double anagram free sequence this implies if we remove that one extra character, say c, from seq_2 $(seq_2 - c)$. The new sequence will be double anagram-free sequence. According to Lemma 4.2 if we add the character c back the whole sequence remains anagram-free.

□

Theorem 4.2. *The geodesic anagram-free chromatic number of an interval graph G is less than or equal to $32p$ where p is the pathwidth of G.*

Proof. As we showed in this section Π gives an anagram-free colouring for any interval graph G. The number of used colours is $\mathrm{pw}(G) \times 8 \times 4 = 32p$. □

Theorem 4.3. *Any chordal graph G that can be represented as an intersection graph of subtrees of a tree T $(V_t : t \in V(T))$ has anagram-free chromatic number of $32p'w \in O(p'w)$, where p' is $\mathrm{pw}(T)$ and w is the clique number of G.*

Colouring Algorithm The colouring algorithm consists of the following steps:

1. If G has more than one component then process each separately.
2. Take a maximal path P in T. Since G is chordal it implies that the induced graph $G^* = G[\cup_{t\in V(P)} V_t]$ is an interval graph. By Theorem 4.2 G^* can be coloured geodesic anagram-freely with at most $32\,\mathrm{pw}(G^*) \leq 32w$ colours.
3. Remove G^* from G to get a graph $G' = G - (\cup_{t\in V(P)} V_t)$. G' is a chordal graph.
4. Remove P from T. $T' = T - V(P)$ is a forest. Each component of T' has pathwidth at most $\mathrm{pw}(T) - 1$.
5. $(V_t \backslash V(G^*) : t \in V(T)\backslash V(P))$ is a forest decomposition of G' and for each component C of G' some tree in this forest gives a tree-intersection representation of C.
6. Inductively colour G' using $32\,\mathrm{pw}(T')w \leq 32(p'-1)w$ colours that are distinct from those used in Step 2.

To show that $32p'w$ colours are sufficient, we give a proof by induction on $\mathrm{pw}(T)$.

Base Case In the base case $\mathrm{pw}(T) = 1$. In this case T is a path which implies $G^* = G$ and $\mathrm{pw}(G^*) = w$. We obtain a colouring using $32\,\mathrm{pw}(G^*) = 32w$ as required.

The total number of colours used by this colouring is at most $32w + 32(p'-1)w = 32p'w$. All that remains is to show that this is indeed a geodesic anagram-free colouring.

Lemma 4.10. *The colouring of G given by the colouring algorithm is a geodesic anagram free colouring.*

Proof. We give a proof by the induction on $\mathrm{pw}(T)$. In the base case $\mathrm{pw}(T) = 1$. This case follows from Theorem 4.2. Assume that $\mathrm{pw}(T) \geq 2$, and consider some geodesic path P^* in G. If P^* does not contain any vertices of G^*, then P^* is contained in G' and is anagram-free by the induction hypothesis. Otherwise, P^* contains some vertices of G^*. Consider the subsequence P' of P^* consisting of only the vertices in G^*. We claim that P' is a subpath of P^*. Since G^* is coloured anagram-freely and the colour of vertices of G^* are distinct from all other colours, therefore the colouring of P^* is anagram-free. □

Proof of the claim. Suppose that P' is not a subpath of P^*. This means that P' consists at least of two subpaths of P^*. This implies that P^* leaves G^* at a vertex $v_i \in P^*$ and enters back to G^* at a vertex $v_{i+k} \in P^*$ for some $k \geq 2$. Let P" be a shortest path in G^* connecting v_{i+k} and v_i. Here we show that $v_i v_{i+k}$ is an edge of G^*. Suppose otherwise. This means $P^*[i : i+k-1].P$" is either a cycle of length 3 which affirms the claim since G is chordal, or $P^*[i : i+k-1].P$" is a cycle of length greater than 3 with no chord, which contradicts with G being a chordal graph. This contradicts with P^* being a geodesic path since we can get a shorter path by avoiding $P^*[i+1 : i+k-1]$. $\square$

The colouring algorithm plus Lemma 4.10 proves Theorem 4.1.

Chapter 5

Conclusion

We close this thesis with a summary of our results and directions for the future work. In Chapter 2, we investigate the piercing number of families of pairwise intersecting convex shapes in a two-dimensional Euclidean space. Recall that the piercing number of a set C is denoted by $\tau(C)$ and the fatness of C is denoted by $\alpha(C)$. We establish that $\tau(C)$ is approximately equal to $39.788\alpha(C)$, improving upon the previous upper-bound of $O(\alpha(C)^2)$. Our findings have implications in various domains, such as designing efficient data structures for 3-D vertical ray shooting and computing depth orders. Furthermore, we provide improved bounds for specific families of shapes, including arbitrarily oriented α-fat rectangles and homothets of regular hexagons. Specifically, we show that $(5\sqrt{2}+2)\alpha(C)+25+5\sqrt{2}$ and $9\alpha(C)+32$ piercing points are sufficient for these families, respectively. Additionally, we establish that the piercing number for pairwise intersecting homothets of an arbitrary convex shape is at most 15, surpassing the previous bound of 16, and present an algorithm to determine the exact locations of the piercing points.

As for the future work one shall look into the following problems:

- We believe that the number of points required to pierce a set of pairwise intersecting homothetic shapes is less than 15. Specifically, we conjecture that the number of points required is 8.

- We believe that the piercing number of a set of arbitrary pairwise intersecting congruent shapes can be generalized in terms of fatness. In 2022, Caoduro [10] showed that a set of pairwise intersecting congruent squares can be pierced by a constant number. They also proved that a set of pairwise intersecting congruent rectangles can be pierced by $12 \log \alpha$ points, where α represents the fatness of any rectangle in the set. We believe that their result is applicable to any arbitrary shapes.

- Many researchers in the literature studied the relation between the ratio of the piercing number over packing number of a set of shapes. We believe that this ratio has a relation to fatness. One shall investigate finding this relation.

In Chapter 3, we demonstrate that the anagram-free vertex colouring of a $2 \times n$ square grid necessitates an increasing number of colours as n grows. This finding resolves an unresolved inquiry in Wilson's thesis and establishes that graphs with a pathwidth of 2 cannot be coloured in an anagram-free manner using a finite number of colours.

In Chapter 4, we introduce and investigate the geodesic anagram-free colouring of graphs, focusing specifically on interval and chordal graphs. We demonstrate that the geodesic anagram-free chromatic number of an interval graph is bounded by $32p$, where p represents the pathwidth of the graph. Furthermore, we prove that the geodesic anagram-free chromatic number of a chordal graph G is $32p'w$, where p' is the minimum pathwidth of a tree used in any sub-tree intersection representation of G, and w represents the clique number of G.

As for the future work one shall look into the following problems:

- For which other graph classes is the geodesic anagram-free chromatic number bounded?
- We conjecture that the anagram-free chromatic number of a chordal graph G is in $O(w \log p')$, indicating that our current bound of $O(p'w)$ is not optimal.

References

[1] Pankaj K. Agarwal, Matthew J. Katz, and Micha Sharir. Computing depth orders and related problems. In *Proceedings of the 4th Scandinavian Workshop on Algorithm Theory*, SWAT '94, page 1–12, Berlin, Heidelberg, 1994. Springer-Verlag.

[2] Pankaj K. Agarwal, Matthew J. Katz, and Micha Sharir. Computing depth orders for fat objects and related problems. *Computational Geometry: Theory and Applications*, 5(4):187–206, 1995.

[3] Noga Alon, Jaroslaw Grytczuk, Mariusz Haluszczak, and Oliver Riordan. Nonrepetitive colorings of graphs. *Random Struct. Algorithms*, 21(3-4):336–346, 2002.

[4] Kenneth Appel and Wolfgang Haken. Every planar map is four colorable. part i: Discharging. *Illinois Journal of Mathematics*, 21:429–490, 1977.

[5] Imre Bárány and Gil Kalai. Helly-type problems. *Bulletin of the American Mathematical Society*, 59(4):471–502, 2022.

[6] Saman Bazargani, Paz Carmi, Vida Dujmović, and Pat Morin. $2{\times}n$ grids have unbounded anagram-free chromatic number. *CoRR*, abs/2105.01916, 2021.

[7] Ahmad Biniaz, Prosenjit Bose, and Yunkai Wang. Simple linear time algorithms for piercing pairwise intersecting disks. In Meng He and Don Sheehy, editors, *Proceedings of the 33rd Canadian Conference on Computational Geometry, CCCG 2021, August 10-12, 2021, Dalhousie University, Halifax, Nova Scotia, Canada*, pages 228–236, 2021.

[8] Prosenjit Bose, Leonidas J. Guibas, Anna Lubiw, Mark H. Overmars, Diane L. Souvaine, and Jorge Urrutia. The floodlight problem. *Int. J. Comput. Geom. Appl.*, 7(1/2):153–163, 1997.

[9] T. C. Brown. Is there a sequence on four symbols in which no two adjacent segments are permutations of one another? *The American Mathematical Monthly*, 78(8):886–888, 1971.

[10] Marco Caoduro and András Sebő. Packing, hitting and coloring squares. *arXiv preprint arXiv:2206.02185*, 2022.

[11] Paz Carmi, Vida Dujmović, and Pat Morin. Anagram-free chromatic number is not pathwidth-bounded. In *Graph-Theoretic Concepts in Computer Science: 44th International Workshop, WG 2018, Cottbus, Germany, June 27–29, 2018, Proceedings 44*, pages 91–99. Springer, 2018.

[12] Paz Carmi, Matthew J. Katz, and Pat Morin. Stabbing pairwise intersecting disks by four points. *CoRR*, abs/1812.06907, 2018.

[13] GD Chakerian and SK Stein. Some intersection properties of convex bodies. *Proceedings of the American Mathematical Society*, 18(1):109–112, 1967.

[14] Ludwig Danzer. To solve the gallean problem äbout circular disks in the euclidean plane. *Studia Sci. Math. Hungary*, 21(1-2):111–134, 1986.

[15] Reinhard Diestel. *Graph theory*, volume 173 of *Graduate Texts in Mathematics*. Springer, Berlin, fifth edition, 2018.

[16] Vida Dujmovic, Gwenaël Joret, Jakub Kozik, and David R. Wood. Nonrepetitive colouring via entropy compression. *Comb.*, 36(6):661–686, 2016.

[17] Alon Efrat, Günter Rote, and Micha Sharir. On the union of fat wedges and separating a collection of segments by a line. In A. Lubiw and J. Urrutia, editors, *Proceedings of the Fifth Canadian Conference on Computational Geometry, Waterloo*, pages 115–120, 1993.

[18] Paul Erdős. Some unsolved problems. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 6:221–254, 1961.

[19] Alexander Andreevich Evdokimov. Strongly asymmetric sequences generated by a finite number of symbols. In *Doklady Akademii Nauk*, volume 179, pages 1268–1271. Russian Academy of Sciences, 1968.

[20] Fănică Gavril. The intersection graphs of subtrees in trees are exactly the chordal graphs. *Journal of Combinatorial Theory, Series B*, 16(1):47–56, 1974.

[21] Jacob E. Goodman and Joseph O'Rourke, editors. *Handbook of Discrete and Computational Geometry.* CRC Press, Inc., USA, 1997.

[22] Branko Grünbaum. On intersections of similar sets. *Portugal. Math.*, 18:155–164, 1959.

[23] Sariel Har-Peled, Haim Kaplan, Wolfgang Mulzer, Liam Roditty, Paul Seiferth, Micha Sharir, and Max Willert. Stabbing pairwise intersecting disks by five points. In *29th International Symposium on Algorithms and Computation (ISAAC 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

[24] Ed Helly. Über mengen konvexer körper mit gemeinschaftlichen punkte. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 32:175–176, 1923.

[25] Nina Kamčev, Tomasz Łuczak, and Benny Sudakov. Anagram-free colourings of graphs. *Comb. Probab. Comput.*, 27(4):623–642, 2018.

[26] Roman N Karasev. Transversals for families of translates of a two-dimensional convex compact set. *Discrete & Computational Geometry*, 24:345–354, 2000.

[27] Roman N Karasev. Piercing families of convex sets with the d-intersection property in $\mathbb{R}^d$. *Discrete & Computational Geometry*, 39(4):766–777, 2008.

[28] Matthew J. Katz. 3-d vertical ray shooting and 2-d point enclosure, range searching, and arc shooting amidst convex fat objects. *Computational Geometry*, 8(6):299–316, 1997.

[29] Matthew J. Katz, Mark H. Overmars, and Micha Sharir. Efficient hidden surface removal for objects with small union size. *Computational Geometry*, 2(4):223–234, 1992.

[30] Veikko Keränen. Abelian squares are avoidable on 4 letters. In *International Colloquium on Automata, Languages, and Programming*, pages 41–52. Springer, 1992.

[31] Veikko Keränen. A powerful abelian square-free substitution over 4 letters. *Theor. Comput. Sci.*, 410(38-40):3893–3900, 2009.

[32] Seog-Jin Kim, Kittikorn Nakprasit, Michael J. Pelsmajer, and Jozef Skokan. Transversal numbers of translates of a convex body. *Discrete Math.*, 306(18):2166–2173, sep 2006.

[33] Monsieur Lothaire. *Algebraic combinatorics on words*, volume 90. Cambridge university press, 2002.

[34] Dániel Marx and Marcus Schaefer. The complexity of nonrepetitive coloring. *Discrete Applied Mathematics*, 157(1):13–18, 2009.

[35] Jirí Matousek, János Pach, Micha Sharir, Shmuel Sifrony, and Emo Welzl. Fat triangles determine linearly many holes. *SIAM J. Comput.*, 23:154–169, 1994.

[36] Jaroslav Nešetřil and Patrice Ossona de Mendez. Tree-depth, subgraph coloring and homomorphism bounds. *Eur. J. Comb.*, 27(6):1022–1041, 2006.

[37] Frank Nielsen. On point covers of c-oriented polygons. *Theo. Comp. Sci.*, 265(1–2):17–29, 2001.

[38] Mark H. Overmars and Frank A. van der Stappen. Range searching and point location among fat objects. *Journal of Algorithms*, 21(3):629–656, 1996.

[39] J. Pach, Ido Safruti, and Micha Sharir. The union of congruent cubes in three dimensions. In *Proceedings of the Seventeenth Annual Symposium on Computational Geometry*, SCG '01, page 19–28, New York, NY, USA, 2001. Association for Computing Machinery.

[40] Panos M. Pardalos, Thelma Mavridou, and Jue Xue. *The Graph Coloring Problem: A Bibliographic Survey*, pages 1077–1141. Springer US, Boston, MA, 1998.

[41] Peter AB Pleasants. Non-repetitive sequences. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 68, pages 267–274. Cambridge University Press, 1970.

[42] Otfried Schwarzkopf, Ulrich Fuchs, Günter Rote, and Emo Welzl. Approximation of convex figures by pairs of rectangles. *Computational Geometry*, 10(2):77–87, 1998.

[43] Lajos Stachó. Über ein Problem für Kreisscheibenfamilien. *Acta Scientiarum Mathematicarum (Szeged)*, 26:273–282, 1965.

[44] Lajos Stachó. A solution of Gallai's problem on pinning down circles. *Mat. Lapok*, 32(1-3):19–47, 1981/84.

[45] Axel Thue. Uber unendliche zeichenreihen. *Norske Vid Selsk. Skr. I Mat-Nat Kl.(Christiana)*, 7:1–22, 1906.

[46] Marc J. van Kreveld. On fat partitioning, fat covering and the union size of polygons. *Comput. Geom.*, 9(4):197–210, 1998.

[47] Tim E. Wilson. *Anagram-free Graph Colouring and Colour Schemes*. PhD thesis, Monash University, February 2019.

[48] Tim E. Wilson and David R. Wood. Anagram-free colorings of graph subdivisions. *SIAM J. Discret. Math.*, 32(3):2346–2360, 2018.

[49] Tim E. Wilson and David R. Wood. Anagram-free graph colouring. *Electron. J. Comb.*, 25(2):P2.20, 2018.