ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Master Thesis at the Department of Computer Science, ETH Zurich, externally supervised at the Department of Computer Science, FU Berlin

Spanning Trees with Low (Shallow) Stabbing Number

Johannes Obenaus
obenausj@student.ethz.ch

Advisors: Prof. Dr. Emo Welzl (ETH Zurich)
Prof. Dr. Wolfgang Mulzer (FU Berlin)
Dr. Michael Hoffmann (ETH Zurich)

Berlin, September 19, 2019

Acknowledgements

First of all, I'd like to express my gratitude to my advisors Prof. Dr. Emo Welzl, Prof. Dr. Wolfgang Mulzer, and Dr. Michael Hoffmann for their continued support, particularly I'd like to thank Wolfgang Mulzer for his excellent supervision and guidance, always having an open door for me.

I had the great opportunity to work in the environment of a fantastic team at "AG Theoretische Informatik" at FU Berlin. I'd like to thank all team members not only for great discussions on several topics but also for a relaxed and enjoyable atmosphere. I'm particularly grateful to Prof. László Kozma and Katharina Klost for many helpful suggestions.

In addition, I'm indebted to my good friend Zacharias Fisches not only for his moral support and very helpful advice but also for proofreading my thesis.

Finally, it is hard to phrase in words my deep gratitude for my wife, Vivian, for her unfailing support and encouragement in every situation.

Abstract

For a set $P \subseteq \mathbb{R}^2$ of n points and a geometric graph $G = (P, E)$ spanning this point set, the *stabbing number* of G is the maximum number of intersections any line determines with the edges of G. Of particular interest is the situation where G is a tree. Chazelle and Welzl [CW89] showed that any point set (in the plane) admits a spanning tree of stabbing number $O(\sqrt{n})$, which is tight. This thesis investigates to what extend this result carries over to the *k-shallow* stabbing number, where only lines having at most k points on one side are considered. Har-Peled and Sharir [HS09] proved that any planar n-point set has a spanning tree of k-shallow stabbing number $O(\sqrt{k} \log \frac{n}{k})$, leaving a gap to the (simple) lower bound $\Omega(\sqrt{k})$. For small k, namely $k \leq \log n$, we prove a lower bound of $\Omega(k)$, implying that the upper bound cannot be improved to $O(\sqrt{k})$ in general.

Moreover, the *tree stabbing number* of a point set $P \subseteq \mathbb{R}^2$ is the minimum stabbing number over all spanning trees (and similar for any other class of graphs). It is easy to see that the path stabbing number is a monotone parameter, i.e. for $P \subseteq P'$ one has $\textsc{path-stab}(P) \leq \textsc{path-stab}(P')$. Eppstein [Epp18, Open Problem 17.5] asked whether this holds for the tree stabbing number as well, which we answer negatively.

Furthermore, we prove that the triangulation stabbing number is not monotone either.

Lastly, we consider applications of spanning trees with low stabbing number in the context of crossing families, an active field of research, and give a detailed analysis of very recent results.

Contents

Chapter 1

Introduction

Spanning trees with low stabbing number play a crucial role for various topics in computational geometry, e.g. in range searching or ray shooting [Aga92]. Agarwal [KA17] describes the work that has been done in the 1980s and 1990s in the context of simplex range searching to be ”the beginning of a new chapter and revitalization of computational geometry as a whole”.

The notion of spanning trees with low stabbing number has been introduced by Welzl [Wel88] (see also [CW89, Wel92]) as a tool for simplex range searching. In simplex range searching one wants to preprocess a point set in such a way that the points in any query simplex can be reported / counted efficiently. In the planar case, Welzl [Wel88] showed – using spanning trees with low stabbing number – that triangle range searching queries can be answered in time $O(\sqrt{n}\log n)$, which was improved to optimal $O(\sqrt{n})$ by Matoušek [Mat93b]. A series of new tools have been delevoped in the context of range searching and especially the work of Matoušek [Mat91, Mat92a, Mat92b, Mat93a, Mat93b] stands out.

The focus of this thesis differs a bit from this classical approach and discusses the stabbing number in a ”shallow” setting on the one hand and its combinatorial properties on the other hand. Let us first introduce two central definitions and postpone the definition of basic terms, such as *points*, *hyperplanes*, *graphs*, etc. to the following section. Also note that we usually state theorems and definitions in the more general setting of d-dimensional space but often focus on the planar case.

Definition 1.1 (stabbing number) *Let $P \subseteq \mathbb{R}^d$ be a point set and $G = (P, E)$ be a geometric graph spanning P. The* stabbing number *of G is the maximum number of proper intersections any hyperplane determines with the edges of G.*

The starting point of this thesis was the following result of Har-Peled and Sharir about spanning trees with low *shallow* stabbing number, which is defined as above but only considering *k-shallow* hyperplanes.

Definition 1.2 (k-shallow hyperplanes) *For a point set $P \subseteq \mathbb{R}^d$, the* weight *w_h of a hyperplane h is the minimum number of points contained in either of the two halfspaces bounded by h (one of them open and the other closed, say the "upper" open). A hyperplane h is k-*shallow *for some positive integer k if $w_h \leq k$.*

Har-Peled and Sharir [HS09] proved that any point set in the plane has a spanning tree of k-shallow stabbing number $O(\sqrt{k}\log\frac{n}{k})$, generalizing the result of Chazelle and Welzl [CW89] that any point set (in the plane) has a spanning tree of stabbing number $O(\sqrt{n})$. To state it explicitly, the result of Chazelle and Welzl holds in arbitrary dimension guaranteeing a spanning tree of stabbing number $O(n^{1-1/d})$, whereas the result of Har-Peled and Sharir is a generalization only in the planar case.

In Chapter 2, we survey some of these results and show that the upper bound of Har-Peled and Sharir cannot be improved to $O(\sqrt{k})$ for certain k, specifically for $k \leq \log n$ we give a construction (which is based on personal communication with Peyman Afshani [Afs]) of a point set such that any k-shallow line intersects at least $\Omega(k)$ edges of any spanning tree.

Chapter 3 is concerned with the monotonicity of certain stabbing numbers. For a point set P, the *tree stabbing number* TREE-STAB(P) is the minimum stabbing number over all spanning trees. We prove that TREE-STAB(P) is not a monotone parameter, i.e. there exist point sets $P \subsetneq P'$ such that TREE-STAB(P) $>$ TREE-STAB(P'). This answers a question by Eppstein [Epp18, Open Problem 17.5]. We also show the same result for triangulations.

Knowing whether a parameter is monotone or not is an important information, since it gives a lot additional structure and thereby simplifies many (combinatorial) questions. For example, a very famous question [ES35] (known as the "happy ending problem") asks:

> For a given integer k, how large must we choose the number $f(k)$ as a function of k in order to guarantee that any point set of (at least) this size contains k points in convex position?

The (obvious) monotonicity of the parameter LARGEST-CONVEX-SUBSET(P) ensures that it is enough to find the smallest $f(k)$ such that the above statement holds (which, incidentally, is still open).

In Chapter 4, we consider applications of spanning trees with low stabbing number in the context of crossing families. We repeat a very recent breakthrough of Pach, Rubin and Tardos [PRT19] concerning the size of crossing families and examine crossing families from a variety of other perspectives.

The following section introduces basic terms, such as *points*, *order types*, *graphs*, *drawings*, etc. and can safely be skipped by the reader familiar with the subject (but be aware – if not mentioned otherwise – we consider planar and finite point sets in general position throughout this thesis).

1.1 Preliminaries and Definitions

Many problems in computational geometry consider geometric properties of embedded objects, whereas other related areas are concerned with abstract properties. For example, an abstract graph consists of a set of vertices and a set of pairs of vertices (called edges). In order to exploit the geometric properties of a graph, the first thing we need is a *drawing* using geometric objects (points, curves, line segments, lines, etc.). Intuitively, these terms are probably clear to any elementary school student, but the following introduction shall illustrate that one really needs to be careful with the subtleties of these terms.

In the context of Euclidean geometry a *point* is commonly referred to as a unique location in Euclidean space. However, there is some ambiguity when speaking of a *point set*, which on the one hand, would be natural to describe as a set of d-dimensional vectors of coordinates. On the other hand, in many geometric applications the exact locations are not important, instead all relevant information is captured in the relative position of the points. To make this distinction very clear, some authors use the terms *configuration* or *order type*.

The notion of *order types* has been introduced by Goodman and Pollack [GP83] (see also [GP93] and [Epp18]). The *order type* of a point set is the mapping that assigns every ordered triple its orientation (clockwise, counterclockwise or collinear, i.e. on a common line). A *configuration* is a class of *order-equivalent* point sets and two point sets are *order-equivalent* if they have the same order type. Strictly speaking, a point set would then be a *realization* of a configuration. For example, 4 points in the plane have only four different order types (two of them involving collinear triples and the other two being a triangle with one point inside and 4 points in convex position).

Throughout this thesis we are mostly interested in the relative position of points and not their exact location, but in order to be consistent with the literature, we use the term point set (it will be clear from the context).

If not mentioned otherwise, we consider finite and planar ($d = 2$) point sets (the size will be denoted by n). If no three points lie on a common line, the point set is in *general position*, which we will assume – if not explicitly stated otherwise – for all point sets throughout this thesis. Furthermore, a set of line segments is often defined on a point set P, which means that any line segment must have two distinct points of P as endpoints.

In $\mathbb{R}^d$, a *hyperplane* H is a $(d-1)$-dimensional affine subspace, which can be described by $(d+1)$ coefficients $a_1, \ldots, a_d, b$ (at least one of the a_i's being non-zero) as follows:

$$H = \{x \in \mathbb{R}^d : a_1 x_1 + \cdots + a_d x_d = b\}.$$

Replacing the equality by an inequality yields the definition of a *halfspace* (*open* if the inequality is strict and *closed* otherwise). In $\mathbb{R}^2$, hyperplanes are called *lines* and halfspaces are called *halfplanes*. For distinct points $p, q \in \mathbb{R}^2$, *lines* or *line segments* can also be characterized as the set of all points $r = p + \lambda \cdot (p - q)$ for $\lambda \in \mathbb{R}$ in the case of lines and $\lambda \in [0, 1]$ for line segments.

A *graph* $G = (V, E)$ consists of a set of *vertices* V and a set of *edges* $E \subseteq \binom{V}{2}$ each joining two vertices. A *curve* is a set $C = \{\gamma(t) : t \in [0, 1]\} \subseteq \mathbb{R}^2$ which is parameterized by a continuous function $\gamma : [0, 1] \to \mathbb{R}^2$ and *simple* if γ is injective on $[0, 1]$. A *drawing* $f : (V, E) \to \mathbb{R}^2$ maps every vertex v to a distinct point $f(v) \in \mathbb{R}^2$ and every edge $e = \{v, w\}$ to a simple curve $\gamma : [0, 1] \to \mathbb{R}^2$ such that

$$\gamma(0) = f(v),\ \gamma(1) = f(w),\ \text{ and } \gamma(t) \neq f(u) \text{ for } t \in (0, 1) \text{ and } u \in V.$$

Two edges *cross* or *intersect* if they share an interior point. When referring to crossing objects in the following chapters, we always mean that their open intersection is non-empty (sometimes emphasized by adding the adverb ”properly”).

A *geometric graph* is a drawing where edges are realized as straight-line segments. A graph $G = (V, E)$ *spans* a point set P if G is connected and $f(V) = P$.

A standard technique in computational geometry uses the *projective duality transform* (in the plane), which gives a one-to-one correspondence between points and (non-vertical) lines. A point $p = (p_x, p_y)$ is mapped to the line $p^\star$ parameterized by $f(x) = p_x x - p_y$ and a line $\ell : f(x) = mx + b$ to the point $\ell^\star = (m, -b)$ (more details can be found in [BCKO08] for example). This duality transform is order and incidence preserving [BCKO08, Observation 8.3], i.e. a point p lies above a line ℓ if and only if $\ell^\star$ lies above $p^\star$ and p lies on ℓ if and only if $\ell^\star$ lies on $p^\star$. Often it is much easier to handle a problem in the dual setting.

Chapter 2

Spanning Trees with Low Stabbing Number

1989	2003	2009	2009	2012
(Chazelle, Welzl)	**(Fekete, Lübbecke, Meijer)**	**(Har-Peled)**	**(Har-Peled, Sharir)**	**(Mulzer, Werner)**
Any planar n-point set has a spanning tree of stabbing number $O(\sqrt{n})$ (which is asymptotically tight).	Computing a spanning tree with minimum stabbing number is $\mathcal{NP}$-hard.	log-approximation algorithm for computing a spanning tree of stabbing number $O(\min(t \log n, n^{1-1/d}))$ with t being the true stabbing number.	Any point set has a spanning tree of k-shallow stabbing number $O(\sqrt{k} \log \frac{n}{k})$ (uses shallow partitions).	Crossing number of a shallow partition is at least $\Omega(\frac{\log n/k}{\log \log n/k})$.

Figure 2.1: Timeline of a few important results

In this chapter, we first repeat some general results about the stabbing number of spanning trees, the second part gives a detailed analysis of the shallow version of this result, and the last two sections discuss the possibility of generalization to higher dimensions as well as the complexity of finding a spanning tree with minimum stabbing number.

Note that also the term crossing number is used in the literature. To avoid confusion with the crossing number of an abstract graph, which is the minimal total number of crossing edges among all drawings, we decided to use stabbing number. Often it is easier (and more readable) to spell the names of parameters out instead of using symbols, but instead of writing half a sentence, we might use a notation like STABBING-NUMBER(G). A line with STABBING-NUMBER(G) intersections does not pass through any vertex of G, since otherwise one could perturb it to create a line with more crossings.

Furthermore, the stabbing number clearly depends on the drawing, but the tree stabbing number, for example, depends only on the order type (see Figure 2.2 for an example).

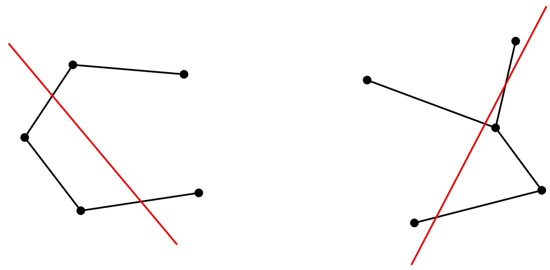


Figure 2.2: Two point sets with different order types and different tree stabbing numbers. For both point sets a spanning tree with minimum stabbing number and a line with the maximum number of intersections is drawn.

2.1 General Setting

The following result that any point set in $\mathbb{R}^d$ has a spanning tree of stabbing number $O(n^{1-1/d})$ was proved in 1989 by Chazelle and Welzl [CW89] in the more general setting of set systems with finite VC-dimension. Another proof for the planar case, using an iterative reweighting technique, was given by Welzl [Wel92], also proving a more precise (asymptotically equivalent) upper bound.

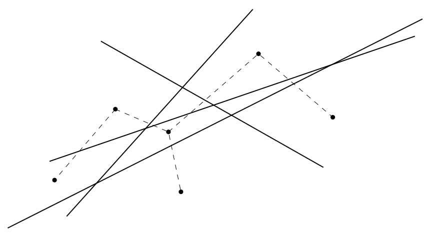
Theorem 2.1 (Chazelle, Welzl, 1989).

Let $P \subseteq \mathbb{R}^d$ be a set of n points. Then there exists a (straight-line) spanning tree T on P such that any hyperplane crosses at most $O(n^{1-1/d})$ edges of T.

We will not repeat the proof here, since the shallow version in the following section uses an entirely different approach.

The bound in Theorem 2.1 is asymptotically tight, meaning that for any $n \in \mathbb{N}$, one can construct a set of n points in $\mathbb{R}^d$ such that an arbitrary spanning tree has stabbing number at least $\Omega_d(n^{1-1/d})$ (the index d indicates that the constants depend on d). A construction proving this for the case $d = 2$ was given by Welzl [Wel92] and can be generalized to arbitrary (fixed) dimension d as follows.

Consider a *simple arrangement* of $dn^{1/d}$ hyperplanes. An *arrangement* is a subdivision of $\mathbb{R}^d$ induced by a finite set H of hyperplanes and is *simple* if H is in *general position*. A set of hyperplanes in $\mathbb{R}^d$ is in *general position* if the intersection of any $k \leq d$ hyperplanes is $(d-k)$-dimensional and empty if $k > d$. Place n points into different (d-dimensional)



cells of this arrangement. Indeed, there are enough such cells, as Lemma 2.2 shows. Each edge of any spanning tree now needs to intersect at least one hyperplane. In particular, $dn^{1/d}$ hyperplanes need to accomodate at least n intersections. Hence, at least one of them has $\frac{1}{d} \cdot n^{1-1/d} \in \Omega_d(n^{1-1/d})$ intersections.

Lemma 2.2 ([Mat02, Proposition 6.1.1]) *Let $\mathcal{A}(H)$ be a simple arrangement of n hyperplanes. Then the number of d-dimensional faces in $\mathcal{A}(H)$ is $\sum_{k=0}^{d} \binom{n}{k}$.*

In particular, Lemma 2.2 shows – for a simple arrangement of $dn^{1/d}$ hyperplanes – that the number of d-dimensional faces is at least

$$\sum_{k=0}^{d} \binom{dn^{1/d}}{k} > \binom{dn^{1/d}}{d} \geq \left(\frac{dn^{1/d}}{d}\right)^{d} = n\,.$$

Note that this estimate is rather crude and with a bit more work one can show that a simple arrangement of $(dn)^{1/d}$ hyperplanes also suffices, though yielding the same asymptotic bound $\Omega_d(n^{1-1/d})$.

2.2 Shallow Setting

The following result of Har-Peled and Sharir [HS09] is the shallow analogue of Theorem 2.1 (in the plane).

Theorem 2.3 (Har-Peled, Sharir, 2009).

Let $P \subseteq \mathbb{R}^2$ be a set of n points and $k \in \{1, \ldots, \lfloor \frac{n}{2} \rfloor\}$. Then there exists a (straight-line) spanning tree T on P such that any k-shallow line intersects at most $O(\sqrt{k}\log(\frac{n}{k}))$ edges of T.

However, the additional log-factor is a bit annoying because it creates a gap to the obvious lower bound of $\Omega(\sqrt{k})$. This lower bound can be obtained as above using an arrangement of $\sqrt{2k}$ lines (with the only difference that we place k points into distinct cells and $n-k$ into one cell that does not affect the shallowness of any line). Improving either of the bounds (upper or lower) will be the focus of this chapter. Unfortunately, we will only give a partial answer for certain values of k.

First, we repeat the main steps in the proof of Theorem 2.3 to understand where this log-factor comes from. The proof relies on partitioning the point set into smaller sets in an "onion-peeling" fashion and constructing spanning trees for the smaller sets. We need another definition:

Definition 2.4 ([MW12]) *Let $P \subseteq \mathbb{R}^d$ be a set of n points and $k \in \{1, \ldots, n\}$. A* k-partition *consists of* *i) a sequence $P_1, \ldots, P_{\lceil n/k \rceil}$ of pairwise disjoint subsets of P such that $\bigcup_i P_i = P$ and $|P_i| = k$, for $i \in \{1, \ldots, \lfloor n/k \rfloor\}$*

ii) a sequence $\Delta_1, \ldots, \Delta_{\lceil n/k \rceil}$ of d-dimensional simplices such that $P_i \subseteq \Delta_i$, for all i.

A hyperplane h crosses *a simplex Δ if $h \cap \Delta \neq \emptyset$ and the* (shallow) crossing number *of a k-partition is the maximum number of simplices that are intersected by any (k-shallow) hyperplane.*

In the following, we focus on the shallow setting. Matoušek proved an upper bound on the shallow crossing number of k-partitions:

Theorem 2.5 (Matoušek, 1992, [Mat92b, Theorem 3.1]).

Let $P \subseteq \mathbb{R}^d$ be a set of n points and $k \in \{1, \ldots, n\}$. Then there exists a k-partition of P with shallow crossing number $O(\log(n/k))$ (for $d = 2, 3$) and $O((n/k)^{1-1/\lfloor d/2 \rfloor})$ (for $d \geq 4$).

Note that Har-Peled and Sharir use Theorem 2.5 in a slightly different form specifically designed for the planar case (see Lemma 2.6). For a set P of n points in the plane and a positive integer k define $D_k = D(P, k)$ to be the intersection of all closed halfplanes containing at least $n - k$ points of P.

Lemma 2.6 *The set $P \setminus D_k$ can be covered by pairwise openly disjoint triangles $C_1, \ldots, C_u$, each containing at most $2k$ points of $P \setminus D_k$ such that any line intersects at most $O(\log(\frac{n}{k}))$ of these triangles. Moreover, $C_i \cap \partial D_k \neq \emptyset$ for each $i = 1, \ldots, u$.*

Lemma 2.6 is specifically constructed for planar point sets and in fact, it does not generalize to higher dimensions as we will see in Section 2.5. We omit the proof of Lemma 2.6, which can be found in [HS09], but we repeat the proof of Theorem 2.3, which relies on iteratively peeling out k-shallow slices of the point set and finding a spanning tree for these slices by applying Lemma 2.6 (see Figure 2.3). We need another lemma:

Lemma 2.7 *For a point set $P \subseteq \mathbb{R}^2$ and $k \in \{1, \ldots, n\}$, one can construct a spanning tree T for $P \setminus D_k$ such that any line intersects at most $O(\sqrt{k} \log(\frac{n}{k}))$ edges of T.*

Proof First, construct a shallow partition of $P \setminus D_k$ into covering triangles $C_1, \ldots, C_u$ as in Lemma 2.6 and for each of the subsets $P \cap C_i$ construct a spanning tree T_i with stabbing number $O(\sqrt{k})$ (Theorem 2.1).

Next, we need to connect all the T_i's to a single tree T. For $i \in \{1, \ldots, u\}$, connect an arbitrary vertex from T_i to an arbitrary point in $C_i \cap \partial D_k$ and let T' be a spanning tree of the connected graph consisting of ∂D_k, all T_i's,

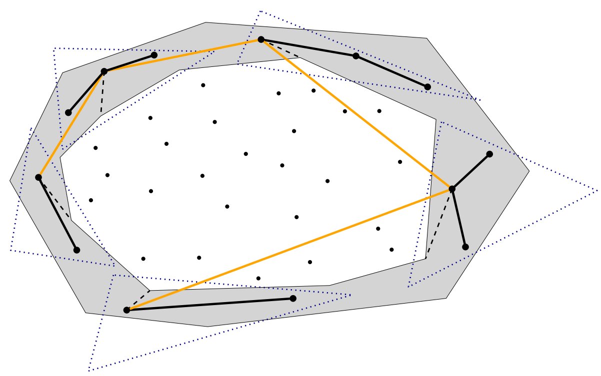


Figure 2.3: Illustration of the proof of Theorem 2.3. The solid, black segments form spanning trees of the subsets and together with the orange segments a spanning tree of the shaded layer.

and the extra edges. T' fulfills the desired stabbing number, since any line ℓ may intersect at most $O(\log(\frac{n}{k}))$ triangles, $O(\sqrt{k})$ edges within each triangle and at most two edges of ∂D_k. Now, we need to get rid of the extra Steiner vertices not belonging to $P \setminus \partial D_k$ (without increasing the stabbing number) by replacing every path $u, w_1, w_2, \ldots, v$ where $u, v \in P \setminus \partial D_k$ and $w_i \notin P$ with the straight-line segment $\overline{uv}$. It is easy to see that this does not increase the stabbing number, which we will also prove later in Lemma 3.1.

In total, these are at most $O(\sqrt{k}\log(\frac{n}{k}))$ crossings. □

Proof (of Theorem 2.3) Let $P \subseteq \mathbb{R}^2$ be a set of n points. Then we construct a spanning tree T as follows. First define a sequence $P_0, P_1, \ldots, P_{\log n}$ of subsets of P as follows:

$$P_0 = P$$
$$P_i = P_{i-1} \cap Q_i \quad \text{with} \quad Q_i = D(P_{i-1}, 2^i).$$

Since $|P_i| \leq |P_{i-1}| - 2^i$, we have at most $\log n$ such subsets. For each i construct a spanning tree T_i for the "onion-slice" $P_{i-1} \setminus Q_i$ such that any line intersects at most $O(\sqrt{2^i}\log(\frac{n_i}{2^i}))$ edges (n_i denotes the number of points in P_i) using Lemma 2.7.

Next, we only need to add at most $\log n$ edges to connect subsequent spanning trees T_i, T_{i+1} to get our desired spanning tree T. Indeed, let ℓ be a k-shallow line. Clearly, ℓ can only intersect subsets of index at most

$U := \lceil \log_2 k \rceil$. Hence, the number of edges intersected by ℓ is at most:

$$\begin{aligned}\sum_{i=1}^{U} O(\sqrt{2^i}\log(\frac{n_{i-1}}{2^i})) &= \sum_{i=1}^{U} O(\sqrt{2^i}\log(\frac{n-2^i}{2^i})) \\ &= O\left(\sum_{i=1}^{U}\sqrt{2^i}(\log n - i)\right) \\ &= O\left(\log n \cdot 2^{\log_2 k/2} - \log k \cdot 2^{\log_2 k/2}\right) \\ &= O(\sqrt{k}\cdot\log(\frac{n}{k}))\end{aligned}$$

□

A natural strategy to improve the upper bound in Theorem 2.3 would be to improve the bound in Theorem 2.5. In fact, Matoušek asked whether it is possible to improve the $O(\log(\frac{n}{k}))$ upper bound to $O(1)$, which would immediately imply the tight bound $O(\sqrt{k})$ for the shallow stabbing number of spanning trees. However, Mulzer and Werner [MW12] recently proved a lower bound of $\Omega(\frac{\log(n/k)}{\log\log(n/k)})$ for the shallow crossing number of k-partitions.

In fact, the bound in Theorem 2.3 cannot be improved to $O(\sqrt{k})$, as we shall see in Section 2.4.

2.3 A Lower Bound for Shallow Partitions

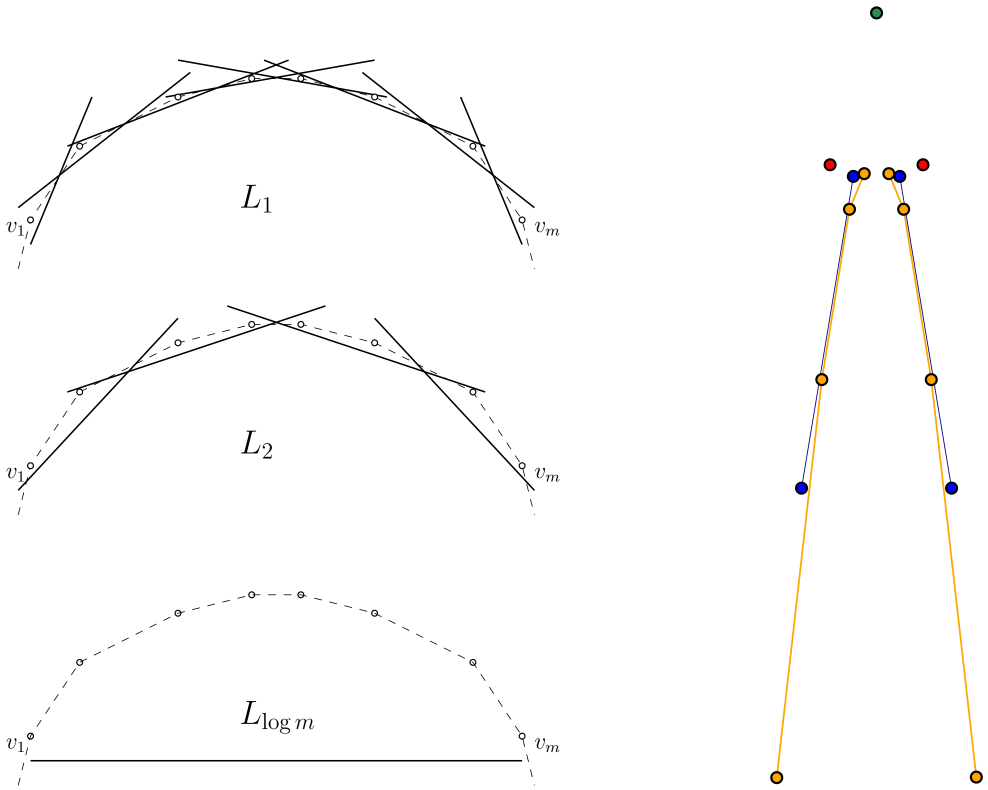
In this section we present the construction for the lower bound of shallow partitions (as done by Mulzer and Werner [MW12]). Again, we repeat the major arguments, since their construction gives insights about how a construction in the setting of shallow spanning trees might look like.

Theorem 2.8 (Mulzer, Werner, 2012).

For any $n \in \mathbb{N}$ and $k \in \{\log n, \ldots, \frac{n}{4}\}$ there exists a set $P \subseteq \mathbb{R}^2$ of n points such that any k-partition of P has crossing number at least $\Omega(\frac{\log(n/k)}{\log\log(n/k)})$.

It is convenient to describe their construction of P in the dual setting $L = P^\star$.

Let $m = 2^\beta$ be a power of two and consider an x-monotone convex chain C with m vertices $v_1, \ldots, v_m$. We define $\log m = \beta$ "layers" of lines such that every line in each layer has the same number of vertices above (lines in the first layer have exactly 1 point above, then $2, 4, 8, \ldots$). Precisely, for $j \in \{0, \ldots, \beta\}$, let L_j be the set of $\frac{m}{2^j}$ lines such that the first line in L_j lies below the vertices $v_1, \ldots, v_{2^j}$, the second line lies below $v_{2^j+1}, \ldots, v_{2\cdot 2^j}$, the third line lies below $v_{2\cdot 2^j+1}, \ldots, v_{3\cdot 2^j}, \ldots$, etc. Then, L consists of $\frac{k}{\beta+1}$ copies of $\cup_{j=0}^{\beta} L_j$ (assuming k is a multiple of $\beta+1$) perturbed in such a way that



(a) Dual setting (redrawn from [MW12]).

(b) Corresponding points in the primal.

Figure 2.4: Construction of the point set for the lower bound of shallow partitions. Points of equal color correspond to dual lines of the same layer.

all lines are distinct but keep their relative relationship. Figure 2.4 gives an illustration, also of the point set in the primal that corresponds to this set L of dual lines. The proof that this point set actually satisfies the required properties can be found in [MW12].

It would be interesting to reprove Theorem 2.8 in the primal setting gaining additional insights on the geometric structure of the point set. In fact, there is still an unsettled gap between the lower and upper bound of shallow partitions:

Open Problem 2.9.

Is the bound in Theorem 2.5 tight for $d = 2$? That is, does there exist for any $n \in \mathbb{N}$ an n-point set $P \subseteq \mathbb{R}^2$ such that any k-partition has crossing number $\Omega(\log(n/k))$?

2.4 The Upper Bound

In this section we give a construction, similar to the one for shallow partitions, proving that the bound $O(\sqrt{k})$ in Theorem 2.3 is not attainable for $k \leq \log n$, which is based on personal communication with Peyman Afshani [Afs]. For $k > \log n$, the situation turned out to be more difficult (we will give some explanation at the end of this section).

Theorem 2.10.

For any $n \in \mathbb{N}$ and positive integer $k \leq \lfloor \log_2 n \rfloor$, there exists a set $P \subseteq \mathbb{R}^2$ of n points such that for an arbitrary spanning tree T there is a k-shallow line intersecting at least $\Omega(k)$ edges.

Proof We construct a set $P \subseteq \mathbb{R}^2$ of $\Theta(n)$ points as follows. Start with a "lower" x-monotone convex chain containing n line segments and let $L = \{\ell_1, \ldots, \ell_n\}$ be the set of lines supporting those segments (these will be the k-shallow lines). Place k layers of points as follows (see also Figure 2.5):

- For $i \in \{0, \ldots, k-1\}$ the i'th layer contains $\frac{n}{2^i}$ points such that the j'th point in this layer is exactly below the lines $\ell_{2^i(j-1)+1}$ to $\ell_{2^i \cdot j}$ (these are 2^i lines).

In total, we placed $\sum_{i=0}^{k} n/2^i \in \Theta(n)$ points. We never placed two points of the same layer below the same line, hence at most k points are contained in the halfplane below each ℓ_i (which are therefore k-shallow). Also note that any layer contains at least one point, since $\frac{n}{2^i} \geq 1$ for $i \leq \log_2 n$.

There are n lines and we will show that any spanning tree must determine at least $\frac{nk}{4}$ intersections with these lines in total. First observe that any two points in our construction are separated by many lines:

Observation 2.11 *Let $p \in P$ be a point in the i'th layer. Then at least 2^{i-1} lines from L separate p from any other point $q \in P$.*

To see this, let p be the l'th point in the i'th layer and q the m'th point in the j'th layer ($i, j \in \{1, \ldots, k-1\}$, $k \in \{1, \ldots, \frac{n}{2^i}\}$ and $l \in \{1, \ldots, \frac{n}{2^j}\}$). The points p and q are distinct, so it is enough to consider the following cases:

Case 1: $i \neq j$. Without loss of generality $i > j$.

By construction, p is below 2^i lines and q is below 2^j lines. Since i and j are integers and $i > j$, we have $2^{i-1} \geq 2^j$ and hence, there are at least 2^{i-1} lines that p is below of, but q is not.

Case 2: $i = j$ and $l \neq m$. Without loss of generality $m < l$.

p is below the lines $\ell_{2^i(l-1)+1}$ to $\ell_{2^i \cdot l}$ and q is below $\ell_{2^i(m-1)+1}$ to $\ell_{2^i \cdot m}$. Thus, it suffices to show that these lines are all distinct, which follows

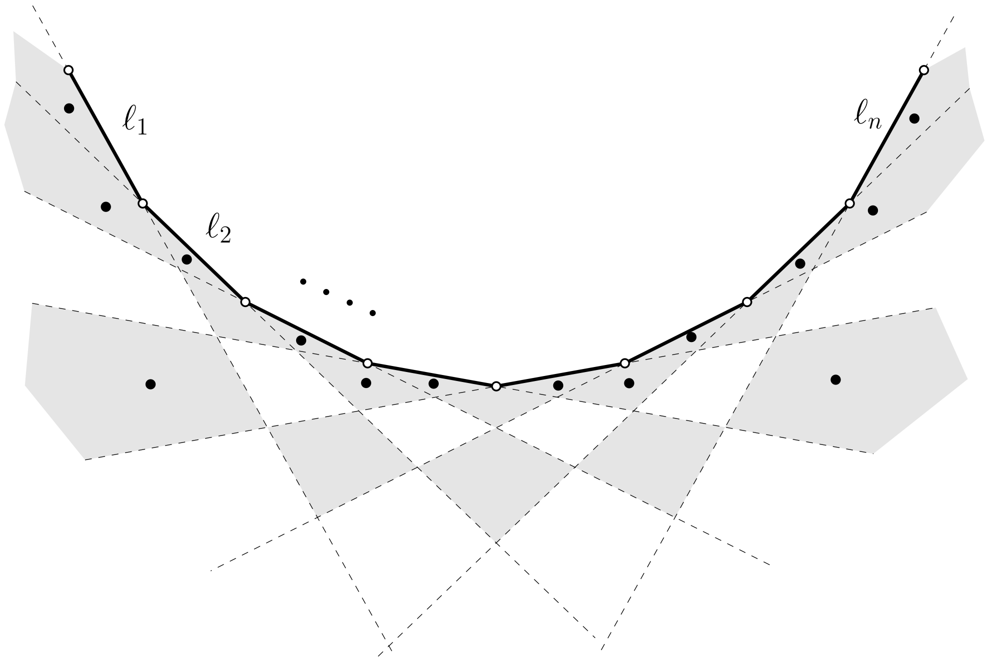


Figure 2.5: A convex chain with n segments and $\Theta(n)$ points arranged in $\log n$ layers (the first three layers are shaded).

from the fact that the largest index of the one line set is smaller than the smallest index of the other set:

$$2^i \cdot m \le 2^i(l-1) < 2^i(l-1)+1\,.$$

Also note that each point in the first layer ($i = 0$) is separated by at least one line.

Using Observation 2.11 and the fact that layer i contains $\frac{n}{2^i}$ points, it follows that the points of layer i contribute at least $\frac{1}{2} \cdot \frac{n}{2^i} \cdot 2^{i-1} = \frac{n}{4}$ intersections in any spanning tree, yielding a total of $\frac{nk}{4}$ intersections. The leading $1/2$-factor compensates for the fact that two distinct points might use the same edge (which is a bit loose but not affecting the asymptotic behaviour).

In particular, one of the n lines must have at least $\frac{k}{4} \in \Omega(k)$ intersections. $\square$

As mentioned earlier, this construction does not generalize to larger k easily. For example, for $k = \sqrt{n}$ the natural strategy to adapt above construction would be to place again k layers of points such that no two points of the same layer are below the same line, which guarantees k-shallowness of the lines.

So, it would be natural to place the points as follows:

- For $i \in \{1, \ldots, k\}$ the i'th layer contains $\lfloor \frac{n}{i^2} \rfloor$ points such that the j'th point is exactly below the lines $\ell_{i^2(j-1)+1}$ to $\ell_{i^2 j}$ (these are i^2 lines).

However, there are a few problems. First, the point set contains too many points, namely $\sum_{i=1}^{\sqrt{n}} n/i^2 \in \Theta(n \log n)$. Of course, we could remove some (e.g. place only $\frac{n}{i^2 \cdot \log n}$ points per layer instead of $\frac{n}{i^2}$), but this does not solve the second problem that we cannot guarantee enough intersections. Since points of subsequent layers are only separated by $2i-1$ lines we get a total number of intersections for an arbitrary spanning tree of

$$\frac{n}{\log n} \sum_{i=1}^{\sqrt{n}} \frac{2i-1}{i^2} \in \Theta(n)\,.$$

Having n lines, this means that we can only guarantee a line having $\Omega(1)$ intersections.

In fact, it makes sense that the construction does not generalize so easily. The lower bound of $\Omega(k)$ for $k \leq \log n$ in Theorem 2.10 is surprisingly strong compared to the upper bound in Theorem 2.3. Even for $k = \log^2(n)$ the analogous result, $\Omega(\log^2 n)$, would already exceed the upper of Theorem 2.3:

$$O(\sqrt{k} \log(\frac{n}{k})) \underset{k=\log^2 n}{=} O(\log n(\log n - \log \log n)).$$

In conclusion, it seems that a new idea is necessary to fully answer the following

Open Problem 2.12.

Is the bound in Theorem 2.3 tight? That is, does there exist, for any $n \in \mathbb{N}$, an n-point set $P \subseteq \mathbb{R}^2$ with k-shallow stabbing number $\Omega(\sqrt{k} \log(\frac{n}{k}))$?

2.5 Higher Dimensions

We saw that the shallow partition theorem (in the form of Lemma 2.6) played a crucial role for constructing spanning trees with low shallow stabbing number. We also saw that Theorem 2.5 applies in higher dimensions (with the same bound in $\mathbb{R}^3$), giving hope that the construction of Har-Peled and Sharir carries over to at least 3-space. However, as mentioned earlier, this is not the case. The following construction [HS09] shows that Lemma 2.6 fails in $\mathbb{R}^3$.

Consider an $m \times m$-grid in the xy-plane and let P be the set of $n = m^2$ points lifted to the unit paraboloid, i.e. every grid point (x, y) is lifted to $z = x^2 + y^2$.

Clearly, P is in convex position and hence every point is 1-shallow (a point is called k-shallow if there exists a halfspace containing this point and at most $k-1$ other points of P).

Lemma 2.13 *For any k-partition of P there is a hyperplane crossing $\Omega(\sqrt{n/k})$ sets in this partition.*

Note that the covering triangles of the k-partition are now 3-dimensional simplices, but it is sufficient to consider the convex hull of the subsets instead, since any hyperplane that intersects the convex hull of a point set will also intersect any simplex covering the point set.

Proof Let $\mathcal{P} = \{P_1, \ldots, P_u\}$ be a k-partition of P (remember that we have $u \in \Theta(\frac{n}{k})$) and consider the set H of $2m-2$ vertical hyperplanes, $m-1$ parallel to the xz-plane and $m-1$ parallel to the yz-plane such that each hyperplane separates the grid points in a distinct way into two non-empty subsets (see Figure 2.6). Let Π_i be the projection of the set P_i onto the xy-plane.

Since all hyperplanes $h \in H$ are vertical, h intersects the convex hull of P_i if and only if it intersects the convex hull of Π_i.

For each i, the convex hull $\mathrm{conv}(\Pi_i)$ has a certain extent in x-direction and in y-direction (the difference between largest and smallest coordinate in x-/y-direction) and at least one of these extents is at least $\sqrt{|P_i|}$ (in order to accommodate all points in P_i). Hence, the overall sum of all extents is at least

$$\Omega(\frac{n}{k} \cdot \sqrt{k}) = \Omega(\frac{n}{\sqrt{k}}).$$

Furthermore, the number of hyperplanes that intersect $\mathrm{conv}(\Pi_i)$ is equal to the sum of its extents in x- and y-direction (see Figure 2.6). Since there are $2\sqrt{n}-2$ hyperplanes, it follows that at least one of them intersects at least $\Omega(\frac{n}{\sqrt{k}} \cdot \frac{1}{\sqrt{n}}) = \Omega(\sqrt{n/k})$ sets. $\square$

But that is not the end of the story. Let's compare this lower bound to what we get from Matoušeks non-shallow partition theorem:

Theorem 2.14 (Matoušek, 1992, [Mat92a, Theorem 3.1]).

Let P be a set of n points in $\mathbb{R}^d$ $(d \geq 2)$ and $k \in \{1, \ldots, n\}$. Then there exists a k-partition such that any *hyperplane crosses at most $O((\frac{n}{k})^{1-1/d})$ sets of the partition.*

Of course, we can apply this theorem in our case to get the upper bound $O((\frac{n}{k})^{2/3})$ for the number of sets crossed by any hyperplane in 3-space. But

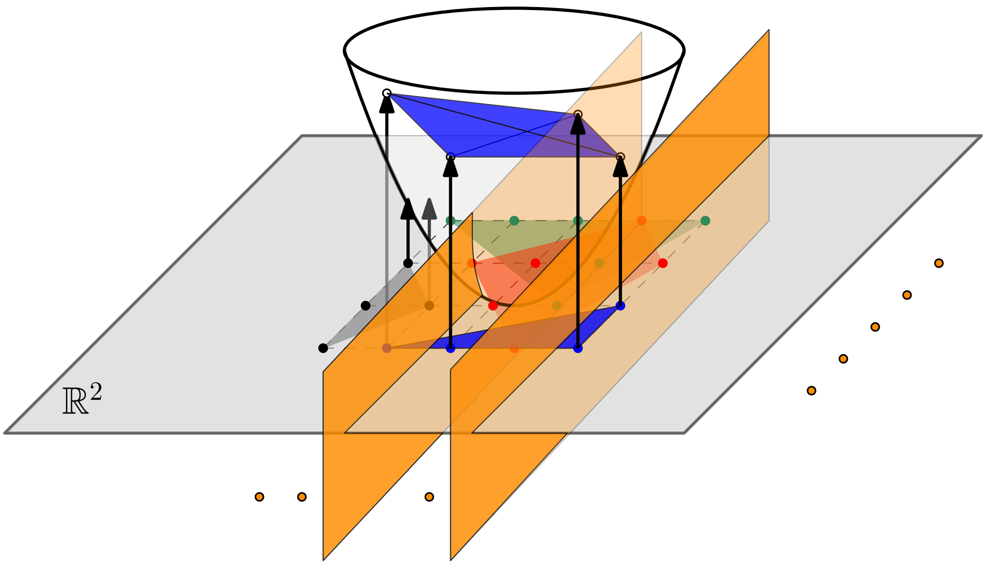


Figure 2.6: An $m \times m$ grid lifted to the unit paraboloid. For any partition (indicated by colors) there is a vertical hyperplane intersecting many convex hulls of subsets.

note that Theorem 2.14 completely throws away the information about the shallowness of the point set.

In fact, the gap between these bounds ($\Omega(\sqrt{n/k})$ due to the simple unit paraboloid example versus $O((n/k)^{2/3})$ derived from Theorem 2.14) is not settled.

Open Problem 2.15 ([HS09]).

Can any set P of n k-shallow points in $\mathbb{R}^3$ be partitioned into $\Theta(n/k)$ subsets (of size roughly k) such that any hyperplane intersects at most $O(\sqrt{n/k})$ subsets?

If this was the case, one can use Theorem 2.1 to construct a spanning tree of stabbing number $O(k^{1-1/d}) \underset{d=3}{=} O(k^{2/3})$ for each subset. And again, putting these together as before results in a spanning tree of stabbing number $O(\sqrt{n/k} \cdot k^{2/3}) = O(n^{1/2}k^{1/6})$.

2.6 $\mathcal{NP}$-hardness and Approximation Algorithms

Fekete, Lübbecke and Meijer [FLM03] proved that computing a spanning tree of minimum stabbing number is $\mathcal{NP}$-hard. They also showed $\mathcal{NP}$-hardness for the minimum stabbing number of other graph classes, such as triangulations and matchings. For all these hardness results they give a reduction from 3-SAT.

Knowing this, the natural next step is to construct approximation algorithms. Theorem 2.1 trivially implies a $\sqrt{n}$-approximation for the minimum stabbing number of spanning trees. Fekete, Lübbecke and Meijer [FLM03] also give a characterization of minimizing the stabbing number of spanning trees (and of matchings) as an integer program (IP) and conjecture that their corresponding linear programming relaxation, which can be solved in polynomial time, yields a constant factor approximation.

However, the currently best provable approximation algorithm is due to Har-Peled [Har09], who constructs a linear programming based algorithm computing a spanning tree (of a point set in $\mathbb{R}^d$) with stabbing number $O(\min(\log n \cdot \mathrm{OPT}, n^{1-1/d}))$ yielding a log-factor approximation.

Chapter 3

Monotonicity of some Stabbing Numbers

In Chapter 2, we studied the stabbing number of geometric graphs, which is the maximum number of crossings any line in the plane determines with the edges of this graph. As mentioned in the introduction, one can study in a similar fashion the *stabbing number of a point set* $P \subseteq \mathbb{R}^2$, which is the minimum stabbing number of any geometric graph G that spans P where G belongs to a certain class of graphs (e.g. paths, trees, matchings, triangulations, etc.). We denote the corresponding stabbing numbers by $\textsc{tree-stab}(P)$, $\textsc{path-stab}(P)$, $\textsc{tri-stab}(P)$. In this chapter, we consider the monotonicity of these parameters and show that $\textsc{tree-stab}(P)$ is not monotone, answering a question by Eppstein [Epp18, Open Problem 17.5].

3.1 Path Stabbing Number

The monotonicity of $\textsc{path-stab}$ is obtained from the simple observation that replacing a vertex of degree 2 in a geometric graph does not increase its stabbing number, precisely:

Lemma 3.1 *Let $G = (V, E)$ be a geometric graph. The following two operations do not increase the stabbing number of G:*

1. *Removing a vertex of degree 1.*
2. *Replacing a vertex v of degree 2 with the segment connecting the two neighbours w_1, w_2 of v.*

Proof Clearly, the first operation cannot increase the stabbing number, since it does not add any new segments.

For the second part, let G' be the geometric obtained from G by performing operation 2 and let ℓ be an arbitrary line. If ℓ has strictly less than STABBING-

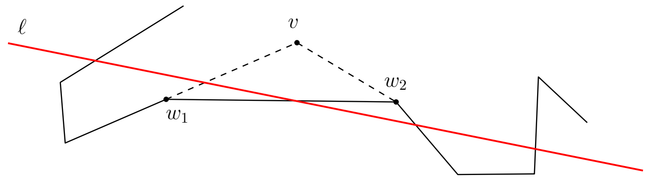


Figure 3.1: Any line that intersects the line segment $\overline{w_1w_2}$ must also intersect either $\overline{w_1v}$ or $\overline{w_2v}$.

Number(G) intersections in G, it has at most Stabbing-Number(G) intersections in G' since we added only one segment. Otherwise, if ℓ has Stabbing-Number(G) intersections in G, it does not pass through any vertex of G (as noted in Chapter 2). If ℓ intersects the newly inserted segment $\overline{w_1w_2}$ it must have also intersected either $\overline{w_1v}$ or $\overline{vw_2}$ (see Figure 3.1). In any case the number of intersection of ℓ did not increase. $\square$

Corollary 3.2 *For all point sets $P \subseteq P'$, it holds that* path-stab$(P) \leq$ path-stab(P').

Proof Let π' be a spanning path of P' with stabbing number path-stab(P'). Iteratively perform the operations of Lemma 3.1 on π' until the remaining vertices are precisely the points of P. This is possible, since all vertices in a spanning path are of degree 1 or 2 and neither operation destroys the property of being a spanning path nor increases the stabbing number. $\square$

Unfortunately, this argument does not apply for tree-stab. When removing a vertex x of degree larger than 2, locally reconnecting the components can result in more crossings, as we will see in the following section.

In fact, the next section provides a construction proving that tree-stab is not monotone.

3.2 Tree Stabbing Number

We construct two planar point sets P_1, P_2 consisting of 9 respectively 10 points disproving the monotonicity of tree-stab and afterwards generalize this construction to arbitrarily large point sets. The tree stabbing numbers of both sets have been computed by an exhaustive computer-aided brute-force search (the source code is available as github repository, see Appendix B).

Before going into details, let's describe the idea intuitively. We want to construct a point set such that removing a point increases its tree stabbing number. Clearly, the removed point must have at least 3 neighbours in any tree of minimum stabbing number, otherwise Lemma 3.1 applies.

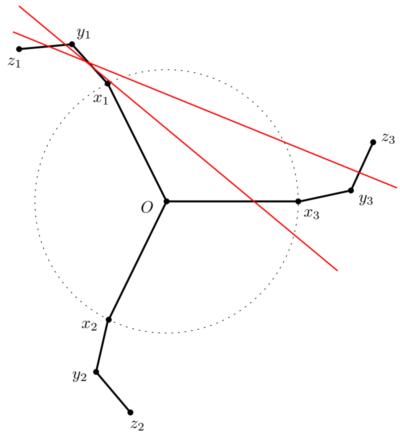
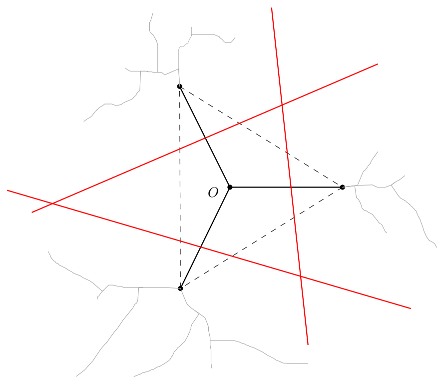


Figure 3.2: Illustration of a point set such that removing the point O increases the tree stabbing number.



If the point O (in the figure on the right) is the one we want to remove, we need to enforce that the red lines (those that intersect two dashed but only one solid segment) are the lines of maximum number of intersections with edges of the spanning tree. Then, locally repairing the removal of O would increase the number of intersections.

The construction is as follows (see also Figure 3.2). Start with a unit circle around the origin O and place 3 evenly distributed points x_1, x_2, x_3 on this circle (in counterclockwise order). Now add an "arm" consisting of 2 points y_i, z_i ($i = 1, 2, 3$) at each of the x_i (outside the circle) such that the points O, x_i, y_i, z_i form a convex chain for $i = 1, 2, 3$. These arms need to be flat enough, i.e. the line supporting the segment $\overline{x_i y_i}$ must intersect the interior of the segment $\overline{O x_{i+2}}$ (indices are taken modulo 3), but also curved enough, i.e. the line supporting the segment $\overline{y_i z_i}$ must have the remaining 8 points on the same side. In particular, there are lines intersecting the segments $\overline{x_i y_i}$, $\overline{y_i z_i}$ and also $\overline{O x_{i+2}}$ on the one hand and $\overline{y_{i+2} z_{i+2}}$ on the other hand (the red lines in Figure 3.2). In order to improve the readability, we might omit that indices are taken modulo 3 if there is no danger of confusion (as in the previous sentence).

Define the two point sets P_1, P_2 (which are both in general position) to be

$$P_1 = \{x_1, y_1, z_1, x_2, y_2, z_2, x_3, y_3, z_3\},$$
$$P_2 = P_1 \cup \{O\}.$$

Theorem 3.3.
For the point sets P_1 and P_2 as defined above we have TREE-STAB$(P_1) = 4$ *and* TREE-STAB$(P_2) \leq 3$.

Proof As noted above, this result was obtained by a computer-aided brute-force search. In order to compute the stabbing number of a given geometric graph spanning some point set, it is enough to consider a *representative set* H_P of lines (and not all infinitely many lines in the plane). For any line ℓ, that partitions the point set into two non-empty subsets, there is a line in the representative set inducing the same partitioning. For an n-point set in general position, the size of such a representative set is $\binom{n}{2}$ (see Appendix, Lemma A.1).

In our case we have

$$|H_{P_1}| = 36 \text{ and } |H_{P_2}| = 45.$$

The sets H_{P_1} and H_{P_2} were also obtained by computer assistance. Any pair of points induces four different representative lines, computing all these lines and removing duplicates yields the sets H_{P_1} and H_{P_2} (as in [GKM14] for example).

Now, it is enough to compute – for all $9^7 = 4782969$ possible spanning trees on P_1 – their intersections with the lines in H_{P_1}, yielding TREE-STAB$(P_1) = 4$.

On the other hand, for P_2 the spanning tree depicted in Figure 3.2 has stabbing number 3 (again by computing all intersections with lines in H_{P_2}) implying TREE-STAB$(P_2) \leq 3$. □

Hence, the tree stabbing number is not a monotone parameter. This idea also generalizes to larger point sets, which we shall discuss next.

Theorem 3.4.
For any integer $n \geq 9$, there exist (planar) point sets $P_1 \subsetneq P_2$ of size $|P_1| = n$ and $|P_2| = n + 1$ such that TREE-STAB$(P_1) >$ TREE-STAB(P_2).

The construction is very similar to the one above, but we need to be more careful this time. We simply replace one of the z_i (say z_1), which was a single point before, by a convex chain C consisting of k points $p_1, \dots, p_k$. Denote the convex chains x_1y_1C, $x_2y_2z_2$ and $x_3y_3z_3$ by C_1, C_2 and C_3.

Our goal will be to remove all but two points of $C \cup \{y_1\}$ to get back to our 9 points setting. Of course, it is crucial to keep the relative position of the points as it is in the 9-point set. Thus, place the points $p_1, \dots, p_k$ as follows:

(i) $O, x_1, y_1, p_1, \dots, p_k$ forms a convex chain.

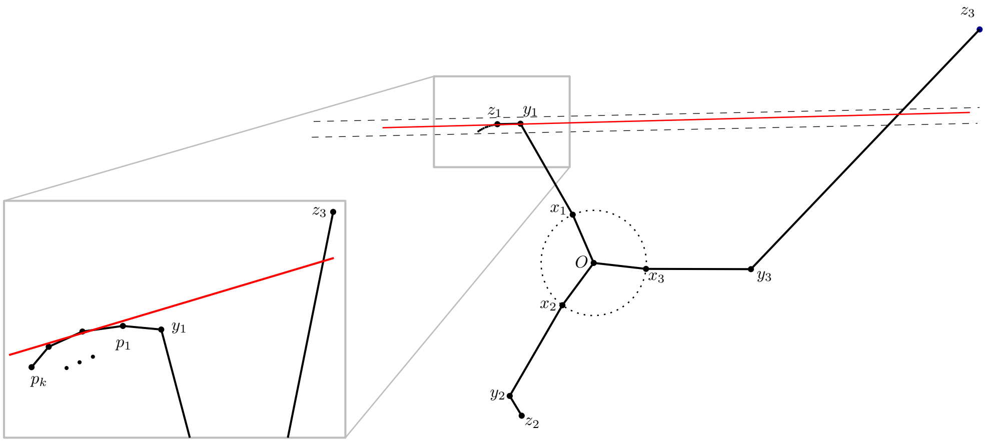


Figure 3.3: Modified point set, where z_1 is replaced by $p_1, \ldots, p_k$ and z_3 pushed out.

(ii) close enough to y_1, so that the order type of the resulting point set is the same no matter which $k-1$ of the points in $C \cup \{y_1\}$ we remove. In particular, no line through any two points not belonging to $y_1, p_1, \ldots, p_k$ may separate these points.

(iii) for any two segments formed by any triple of points in C_1 (consecutively along the convex chain) there is a line intersecting these two segments and also $\overline{y_3 z_3}$. To achieve this, C needs to be sufficiently flat and z_3 needs to be pushed further away.

See Figure 3.3 for an illustration of the modified point set. Also note that Theorem 3.3 has been verified to still hold after the modification of pushing z_3 further out.

Define P'_1 and P'_2 as above (replacing z_1 by $p_1, \ldots, p_k$):

$$P'_1 = \{x_1, y_1, p_1, \ldots, p_k, x_2, y_2, z_2, x_3, y_3, z_3\},$$
$$P'_2 = P'_1 \cup \{O\}.$$

We split the proof of Theorem 3.4 into two parts, first showing $\textsc{tree-stab}(P'_2) \leq 3$ and then $\textsc{tree-stab}(P'_1) \geq 4$, starting with the simpler.

Lemma 3.5 *For the point set P'_2 it holds that* $\textsc{tree-stab}(P'_2) \leq 3$.

Proof We show that the spanning tree depicted in Figure 3.3 has stabbing number 3. Let ℓ be a line in the plane. The following case distinction shows that ℓ intersects at most 3 segments.

Case 1: ℓ intersects none of the convex chains C_i $(i \in \{1, 2, 3\})$.

Then, ℓ has at most two intersections with the three inner segments.

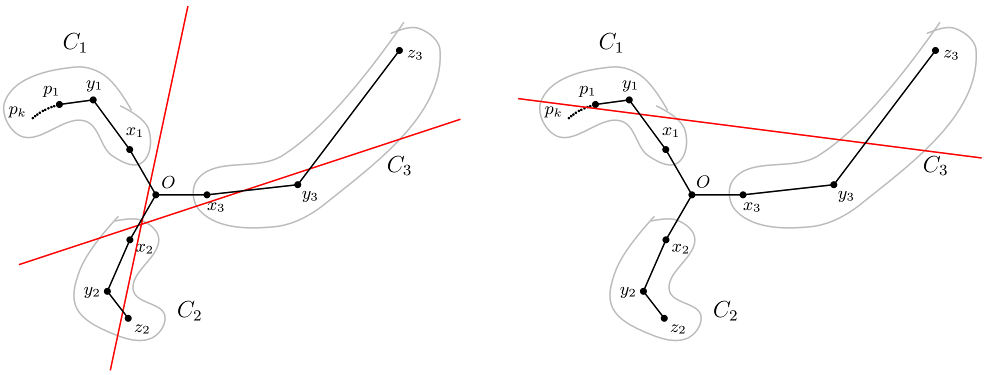


Figure 3.4: Illustration of Case 2 (left) and Case 3 (right) of the case distinction in the proof of Lemma 3.5.

Case 2: ℓ intersects exactly one of the C_i's $(i \in \{1,2,3\})$.

If ℓ intersects one segment of C_i, then it can have at most two more intersections with inner segments. If ℓ intersects two segments of C_i, then it can neither intersect $\overline{Ox_i}$ nor $\overline{Ox_{i+1}}$, since $x_{i+1}OC_i$ forms a convex chain (see Figure 3.4). In any case, ℓ has at most three intersections.

Case 3: ℓ intersects two of the C_i's $(i \in \{1,2,3\})$.

ℓ cannot intersect any inner segment and by construction of the convex chains ℓ cannot intersect two segments of different C_i's simultaneously (see again Figure 3.4). □

The second part (TREE-STAB$(P_1) \geq 4$) is a bit more cumbersome but uses only basic tools as well. We need the following lemma (see also Figure 3.5).

Lemma 3.6 *Let $G = (V, E)$ be a forest with c connected components and $|V| \geq 4$. Mark three of the vertices as special (call them v_1, v_2, v_3) and iteratively perform the following two operations on* non-special *vertices:*

- *remove a vertex of degree 1.*
- *replace a vertex v of degree 2 with the edge connecting the two neighbours of v.*

Repeat these operations (in any order) until no non-special vertex of degree ≤ 2 remains. Then the resulting graph is a forest and consists of the three special vertices and at most one non-special vertex.

Proof Let $G' = (V', E')$ be the graph that was obtained from G by repeatedly performing above operations and let n' denote its number of vertices (including the three special). Clearly G' is a forest, since both operations decrease the number of vertices and the number of edges by exactly 1 and cannot create cycles.

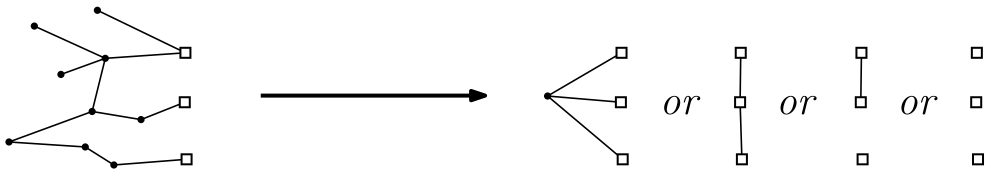


Figure 3.5: Illustration of Lemma 3.6. Special vertices are depicted as squares. Other vertices of degree 1 or 2 are successively removed.

Furthermore, all non-special vertices have degree at least three. Then – using the handshaking lemma and the fact that the forest G' has $n' - c'$ edges, where c' denotes the number of connected components in G' – we obtain:

$$2|E'| = \sum_{i=1}^{n'} \deg(v_i) \geq 3(n' - 3) + \sum_{i=1}^{3} \deg(v_i)\,. \tag{3.1}$$

Observe that $c' \leq 3$ holds, since all connected components not containing a special vertex are completely removed, which follows inductively from the fact that any tree has a leaf. Therefore, it suffices to consider the following three cases:

Case 1: $c' = 1$.

This implies $\sum_{i=1}^{3} \deg(v_i) \geq 3$ and hence (using Equation 3.1),

$$2(n' - 1) \geq 3n' - 6,$$

which only holds for $n' \leq 4$.

Case 2: $c' = 2$.

This implies $\sum_{i=1}^{3} \deg(v_i) \geq 2$ and hence (using Equation 3.1),

$$2(n' - 2) \geq 3n' - 7,$$

which only holds for $n' \leq 3$.

Case 3: $c' = 3$.

This implies $\sum_{i=1}^{3} \deg(v_i) \geq 0$ and hence (using Equation 3.1),

$$2(n' - 3) \geq 3n' - 9,$$

which only holds for $n' \leq 3$. $\square$

Keep in mind that the removal process of Lemma 3.6 does not increase the stabbing number (see Lemma 3.1). Now we are prepared to prove the second part of Theorem 3.4.

Lemma 3.7 *For P_1' as defined above it holds that* TREE-STAB$(P_1') \geq 4$.

Proof For the sake of contradiction assume that there is a spanning tree T of P_1' with stabbing number 3.

Our goal will be to carefully remove points from P_1 such that the stabbing number of T cannot increase until there are only 9 points left in exactly the same relative position as in Theorem 3.3. Clearly, this would be a contradiction.

Consider the set of edges (line segments) of T with at least one endpoint among the points in C_1. There are at most 3 edges having only one endpoint in C_1 (we call them *bridges*). If there would be more than 3 bridges, there would be a line that intersects at least 4 line segments, namely a line that separates C_1 from the rest. Because of the same reason, not all three bridges can go to the same other component (C_2 or C_3).

There are at most 3 points in C_1 that are incident to a bridge and if they are distinct, one of them needs to be x_1, otherwise the line separating x_1 from the rest of C_1 has 4 intersections. Pick three vertices v_1, v_2, v_3 in C_1 such that x_1 and any point incident to a bridge is among them and mark them as special.

Next, we apply Lemma 3.6 to the subforest induced by C_1:

Case 1: No non-special vertex in C_1 survives the removal process.

Then 9 points with the same order type as in Theorem 3.3 and a spanning tree with stabbing number 3 remain, which is a contradiction to Theorem 3.3.

Case 2: One non-special vertex v in C_1 survives the removal process.

Then v is incident to all special vertices v_1, v_2, v_3. Since these 4 points form a convex chain, there is a line ℓ that separates v from v_1, v_2, v_3 and at the same time z_3 from the rest of the point set (by construction). In particular, ℓ has only z_3 and v on one side and all other points on the other. z_3 cannot be adjacent to v, since v is not incident to a bridge and therefore contributes another intersection to ℓ. This is a contradiction to the assumption that T was a spanning tree of stabbing number 3 (see also Figure 3.6). □

Theorem 3.4 now follows from Lemma 3.5 and 3.7.

Relation to Euclidean Minimum Spanning Tree

It is interesting to observe that the two point sets in Theorem 3.3 have the property that the overall length of a Euclidean minimum spanning tree (EMST) is smaller for the larger point set. Intuitively, it makes sense to use

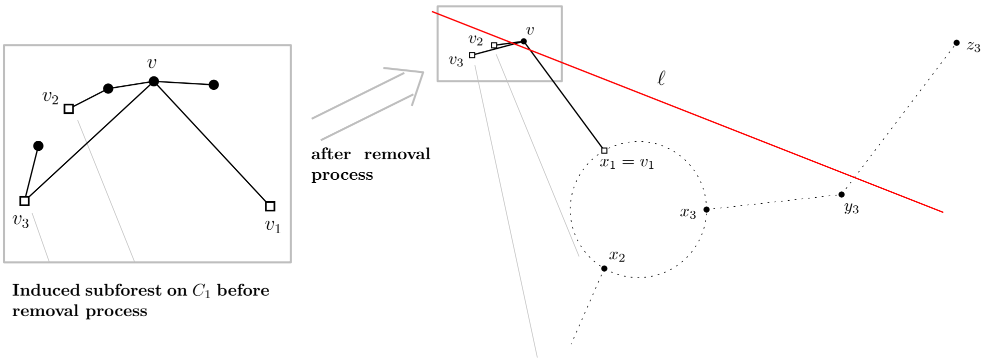
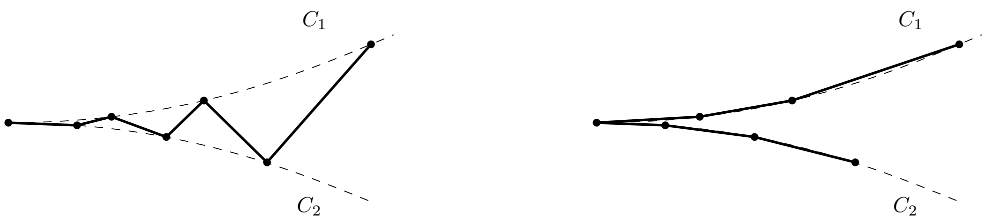


Figure 3.6: Illustration of Case 2. If a non-special vertex survives the removal process, the red line has too many intersections.



(a) EMST with large stabbing number.

(b) Spanning tree with small stabbing number.

Figure 3.7: A point set with points on two sufficiently flat convex chains C_1 and C_2 and distances on the chains large enough such that the spanning tree on the left forms an EMST.

the EMST as an approximation for a spanning tree with low stabbing number, since shorter segments are less likely to be intersected by some line. In fact, our spanning tree for the 10 point set is an EMST. Mitchell and Packer [MP07] analyzed this and other heuristics on random point sets.

However, the EMST yields an arbitrarily bad approximation in general, as Figure 3.7 shows.

The question, whether there is a relation between the non-monotonicity of the EMST and the non-monotonicity of the tree stabbing number remains open.

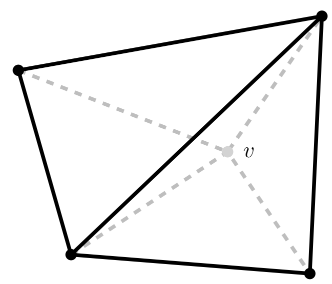
Open Problem 3.8.

Do there exist point sets $P \subsetneq P'$ with TREE-STAB*(P) $>$* TREE-STAB*(P'), but* EMST*(P) $\leq$* EMST*(P')?*

In the following section we discuss the triangulation stabbing number and show that it is not a monotone parameter either. Furthermore, the minimum weight triangulation is also not monotone [MR06], supporting the claim about a possible connection.

3.3 Triangulation Stabbing Number

In some sense the triangulation stabbing number is more difficult to handle. The vertex to be removed must have degree at least 5, otherwise we can easily re-triangulate locally without increasing the stabbing number by a similar argument as in Lemma 3.1 (see the illustration on the right).



On the other hand, triangulations enforce a lot more structure than trees for example. And this time, we are exploiting this structure to construct point sets disproving the monotonicity of TRI-STAB.

Theorem 3.9.
For any $n \in \mathbb{N}$ (large enough), there exist point sets $P \subsetneq P'$ in the plane such that TRI-STAB$(P) >$ TRI-STAB(P').

First, we need a lemma for the special situation of points in convex position:

Lemma 3.10 *Let $P \subseteq \mathbb{R}^2$ be a set of n points in convex position. Then*

$$\textsc{tri-stab}(P) \leq 2\lceil \log_2 n \rceil$$

holds.

Proof Enumerate the points of P along the convex hull by $p_0, \ldots, p_{n-1}$ in clockwise order and construct a triangulation "from outside to inside" by first connecting all points along the convex hull, then p_0 with p_2, p_2 with p_4, p_4 with p_6, …, then p_0 with p_4, p_4 with p_8, …, etc., precisely:

For $i \in \{0, \ldots, \lceil \log_2 n \rceil - 1\}$:
  For $j \in \{0, \ldots, \lfloor n/2^{i+1} \rfloor\}$:
    Connect the points with indices $2^{i+1} \cdot j$ and $2^{i+1} \cdot (j+1)$.

In each step ($i \in \{0, \ldots, k-1\}$) we took a subset of points and connected them with a convex chain (see Figure 3.8). Hence, any line ℓ can intersect at most two segments of each layer and since we placed at most $\lceil \log_2 n \rceil$ layers, ℓ intersects at most $2\lceil \log_2 n \rceil$ segments of the triangulation. □

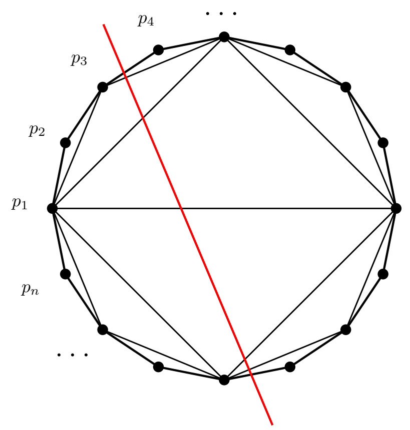


Figure 3.8: Any line can intersect at most two segments of each of the $\log n$ layers.

The proof of Theorem 3.9 exploits the additional structure of triangulations, using two convex chains facing each other.

Proof (of Theorem 3.9) Consider two symmetric convex chains $C_1 = \{p_1, \ldots, p_n\}$ and $C_2 = \{p'_1, \ldots, p'_n\}$ (sufficiently flat) each consisting of n points and facing each other as depicted in Figure 3.9 (a). These points constitute the point set P consisting of $2n$ points. P' consists of the same $2n$ points and two more (slightly perturbed) points added on the line segment connecting the two middle points of C_1 and C_2 (as in Figure 3.9 (b)).

Any triangulation of P must have $2n - 1$ segments connecting a point from C_1 with a point from C_2 (the green area in Figure 3.9 (a)). Hence,

$$\text{TRI-STAB}(P) \geq 2n - 1.$$

On the other hand, the triangulation of P' depicted in Figure 3.9 (b) has stabbing number $n + 4 \log n + 3$, which can be seen as follows. The two green areas contain n segments each and are constructed in such a way that any line ℓ may intersect at most n segments from both green areas. For this, the two points $p_{n/2-1}$ and $p_{n/2+1}$ need to be sufficiently far from $p_{n/2}$. Figure 3.10 illustrates the area that contains all lines which intersect line segments of "upper" $(p_1, \ldots, p_{n/2})$ and "lower" $(p_{n/2}, \ldots, p_n)$ half of the convex chain in the green region.

Since C_1 and C_2 are convex, ℓ may accumulate $4 \log n$ more intersections in the blue areas by Lemma 3.10. The white area contains only a constant number of segments, in total ℓ has at most $n + 4 \log n + 3$ intersections. Hence,

$$\text{TRI-STAB}(P') \leq n + 4 \log n + 3.$$

Taking n large enough the claim follows. $\square$

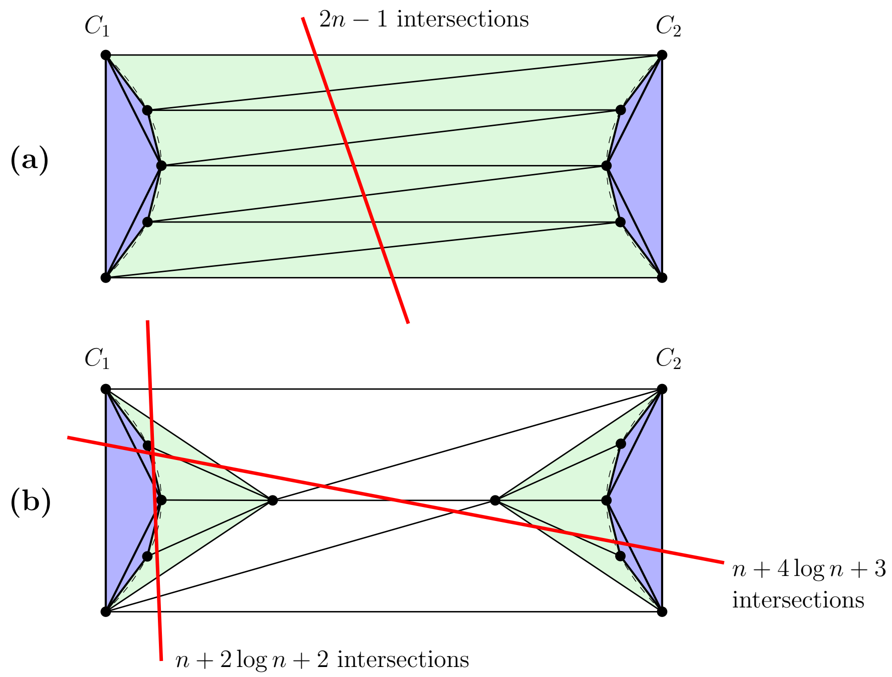


Figure 3.9: Adding points to a point set can increase its triangulation stabbing number.

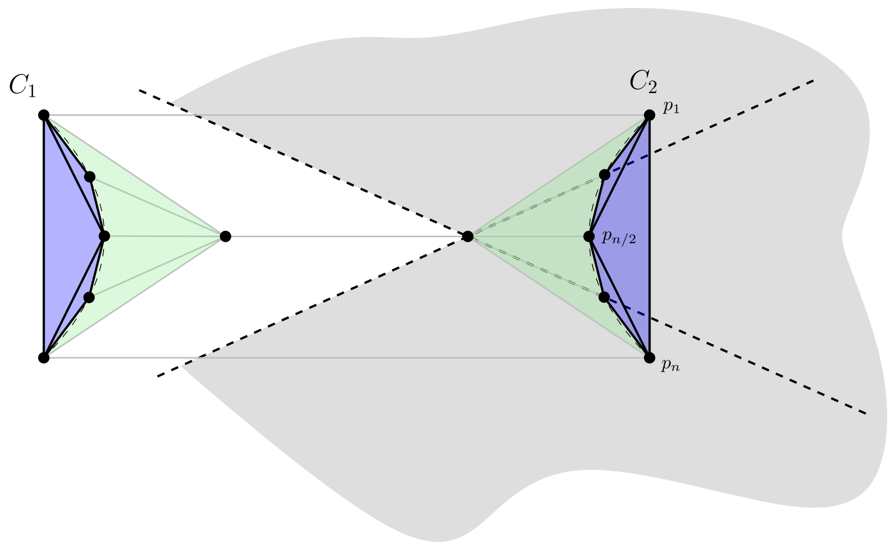


Figure 3.10: Any line that simultaneously intersects a segment connected to the "upper" half of the convex chain C_2 and a segment connected to the "lower" half of C_2 (both in the green region) must be fully contained in the shaded area and hence, cannot intersect any green segment of C_1 (and also the other way around).

Chapter 4

Crossing Families

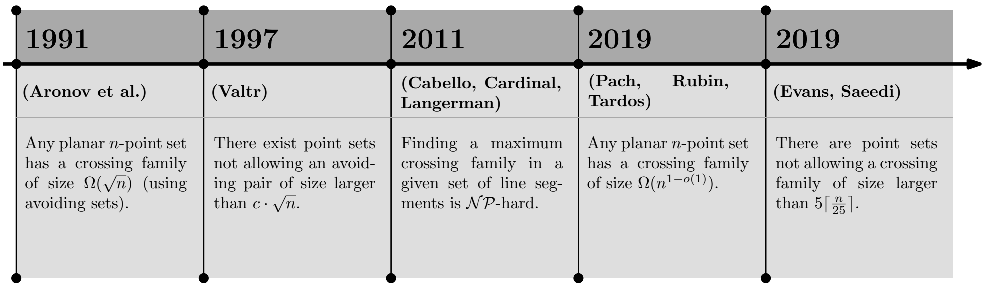


Figure 4.1: Timeline of a few important results.

In this chapter we discuss one of the many and broad applications of spanning trees with low stabbing number. The first section motivates the connection between spanning trees and crossing families, whereas the remaining sections analyze crossing families from various points of view.

A set L of line segments is called *crossing family* if any two segments in L intersect (properly) and *parallel family* if the lines through any two segments intersect outside both segments.

There is a very elegant proof using spanning trees with low stabbing to show a lower bound on the size of parallel families. Since the problems of finding large crossing families and parallel families are equivalent in the following sense, we repeat the proof for parallel families and then dive into more details of crossing families.

For a given set $L \subseteq \mathbb{R}^2$ of line segments, denote by MAX-CROSSING-FAMILY(L) a maximum crossing family determined by L, i.e. a maximum size subset of pairwise crossing segments from L, and similarly, for a given set $P \subseteq \mathbb{R}^2$ of points, MAX-CROSSING-FAMILY(P) denotes the size of a maximum crossing family determined by the set of all $\binom{n}{2}$ line segments on P. For $n \in \mathbb{N}$, define

$$c(n) = \min_{n-\text{point configuration } P} \textsc{max-crossing-family}(P)$$

$$p(n) = \min_{n-\text{point configuration } P} \textsc{max-parallel-family}(P).$$

Aronov et al. [AEG+91] showed that $c(n) = p(n)$ holds. In particular, there cannot exist point sets of n points that have a smaller maximum crossing family than the maximum parallel family of any other n-point set and vice versa.

Theorem 4.1 ([AEG+91, Wel92]).
Any set $P \subseteq \mathbb{R}^2$ of n points has a parallel family of size $\Omega(\sqrt{n})$.

Proof Let $p_1, \ldots, p_n$ be a spanning path of P of minimum stabbing number and add the edge $\{p_1, p_n\}$ to create a spanning cycle $\mathcal{C}$ with stabbing number c. The bound on the stabbing number of spanning trees (Theorem 2.1) implies also $c \in O(\sqrt{n})$, which can be seen as follows. Given a spanning tree T with stabbing number c', one can construct a spanning path of stabbing number at most $2c'$ by doubling all edges (i.e. creating a multigraph). This graph has a Eulerian tour and following this tour while replacing already visited vertices by a shortcut does not increase the stabbing number (Lemma 3.1).

Consider the graph G that has a vertex for each edge of $\mathcal{C}$ and two vertices are adjacent if and only if the corresponding line segments are *not* parallel. G has n vertices and an independent set in G corresponds to a parallel family on P.

The goal will be to bound the number of edges in G and then apply Turáns theorem.

By definition of the stabbing number, no line (extending a line segment of $\mathcal{C}$) can intersect more than c other line segments of $\mathcal{C}$ and therefore the degree d_v of any vertex $v \in V(G)$ is at most c. Hence, G has at most cn edges. By Turáns theorem [AS16], G has an independent set of size at least

$$\sum_{v \in V(G)} \frac{1}{d_v + 1} \geq \frac{n}{c+1} = \frac{n}{O(\sqrt{n})} = \Omega(\sqrt{n}) \,. \qquad \square$$

Next, we focus more on crossing families.

4.1 Bounds on the Size of Crossing Families

The original proof of Aronov et al. [AEG+91] showing that any point set has a crossing family of size $\Omega(\sqrt{n})$ used *avoiding sets*. In fact, this approach turned out to be very effective. Very recently, in 2019, Pach, Rubin and Tardos [PRT19] used a relaxation of the notion of avoiding sets to improve this lower bound to $\Omega(n^{1-o(1)})$ coming very close to the (trivial) upper bound $O(n)$.

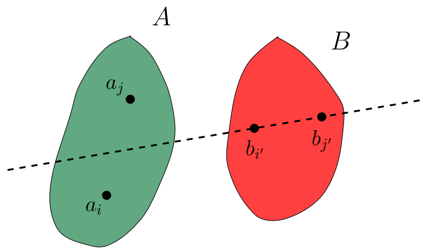
Let $A, B \subseteq \mathbb{R}^2$ be separated point sets (i.e. $\text{conv}(A) \cap \text{conv}(B) = \emptyset$). A *avoids* B if the line through any two points in A does not intersect the convex hull of B. Two sets A and B are called *avoiding* if A avoids B and B avoids A.

The benefit of avoiding sets gets apparent in the following lemma, which shows that an avoiding set determines a crossing family (and also a parallel family).

Lemma 4.2 *Let $A = \{a_1, \ldots, a_r\}$ and $B = \{b_1, \ldots, b_r\}$ be avoiding sets, each of size r. Then, $A \cup B$ determines a crossing family of size r (and a parallel family of size r).*

The *circular ordering* of some points $p_1, p_2, \ldots$ around a point q corresponds to the order the points p_i are encountered by a ray that is continuously rotated around q.

Proof Note that the circular orderings of the points $b_1, \ldots, b_r$ around any point a_i have to be the same. Indeed, if this was not the case and there exists points a_i and a_j such that two points $b_{i'}$ and $b_{j'}$ appear in different order, the line $\overline{b_{i'}b_{j'}}$ must have a_i on one side and a_j on the other. Hence, $\overline{b_{i'}b_{j'}}$ intersects the convex hull of A, which would be in contradiction to A and B being avoiding.



Rename the points in A and B in such a way that $a_1, \ldots, a_r$ is the (counterclockwise) circular ordering of the points in A around a point in B and $b_1, \ldots, b_r$ the ordering of the points in B around a point in A. Then the set $\{\overline{a_1b_1}, \ldots, \overline{a_rb_r}\}$ is a crossing family and $\{\overline{a_1b_r}, \ldots, \overline{a_rb_1}\}$ a parallel family (see Figure 4.2). Again, because otherwise we could find a line through a pair of points that intersects the convex hull of the other set. $\square$

Aronov et al. [AEG+91] proved that any set of n points has an avoiding pair of size $\Omega(\sqrt{n})$. It is interesting to note that there are point sets which do not contain an avoiding pair of size larger than $O(\sqrt{n})$, as proved by Valtr [Val97]. This implies that the approach of Aronov et al. cannot be further improved (avoiding sets are a stronger notion than crossing families). Nevertheless, a relaxation of this notion was the key for the following improved lower bound.

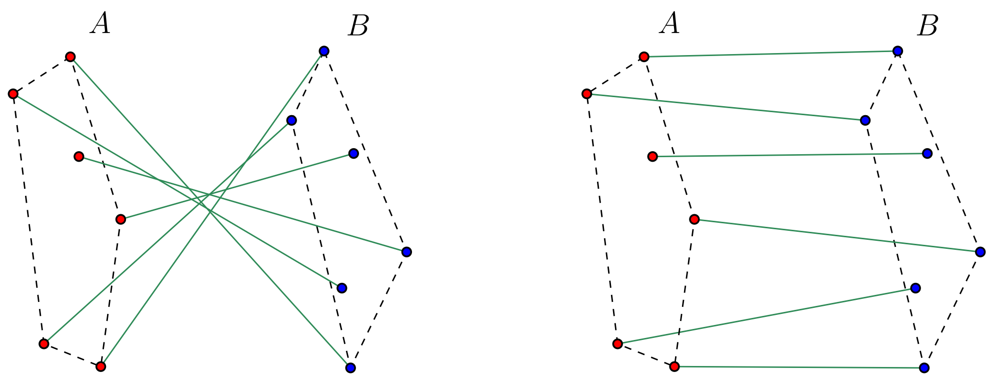


Figure 4.2: Two avoiding sets A, B determine a crossing family (left) and a parallel family (right).

Theorem 4.3 (Pach, Rubin, Tardos, 2019).

Any set $P \subseteq \mathbb{R}^2$ of n points determines a crossing family of size $\Omega(n^{1-o(1)})$.

We will not repeat the proof, which is a bit involved, but come back to this result in Section 4.5 when discussing a possible generalization to higher dimensions. Next, we briefly discuss some upper bounds on the size of crossing families.

The Upper Bound

The asymptotic lower and upper bounds differ only by a factor $2^{O(\sqrt{\log n})}$, guaranteeing a near linear size crossing family. However, one can still try to improve the constant of the upper bound. A crossing family can, of course, contain at most $n/2$ segments.

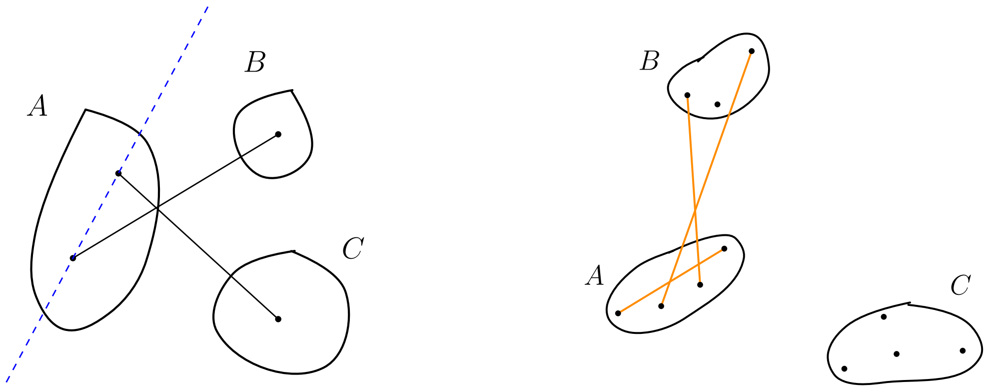
In order to improve this upper bound, one needs to construct a point set P such that a maximum crossing family consists of at most cn segments. In section 4.2 we will see that it is $\mathcal{NP}$-hard to find maximum crossing families. However, there are other ways to construct point sets not allowing large crossing families, as we argue next.

We illustrate the construction of Evans and Saeedi [ES19] of a point set not allowing a crossing family of size larger than $\lceil \frac{n}{3} \rceil$, which is not the tightest known constant, but easy to illustrate and the basis for tighter constructions.

A set A *separates* B from C if

(C1) the convex hulls of A and $B \cup C$ are disjoint

(C2) for any line ℓ through two points of A, the convex hulls of B and C are fully contained on different sides of ℓ.



(a) If AB- and AC-segments intersect, A cannot be separating.

(b) A crossing family emanating from A (A separates B from C).

Figure 4.3: Illustration of Lemma 4.4

For point sets P, Q, a segment connecting a point from P with one of Q is called *PQ-segment*.

Lemma 4.4 ([ES19, Lemma 1]) *Let A, B and C be disjoint point sets in the plane such that A separates B from C. Then any crossing family in $A \cup B \cup C$ emanates from either A, B or C, i.e. all segments have an endpoint in the same component.*

Proof First, no crossing family can contain an AB- and AC-segment, since the line through the two endpoints in A would have both other endpoints (one from B and one from C) on the same side, violating the separating property **(C2)** (see also Figure 4.3 (a)).

Second, the convex hulls of the union of any two of the sets A, B, C is disjoint from the convex hull of the third: $\operatorname{conv}(A) \cap \operatorname{conv}(B \cup C) = \emptyset$ holds by definition **(C1)** and the other two follow again from property **(C2)**. Hence, no crossing family can contain any XY- and ZZ-segments for $X, Y \in \{A, B, C\}$ and $X, Y \neq Z$.

This implies that any crossing family emanates from either A, B or C (see Figure 4.3 (b) for an example). $\square$

The current best upper bound is also due to Evans and Saeedi [ES19]:

Theorem 4.5 (Evans, Saeedi, 2019).

For any $n \in \mathbb{N}$ there exist n-point sets, whose maximum crossing family is of size at most $5\lceil \frac{n}{24} \rceil$.

We discussed the best known lower and upper bounds on the size of crossing families and next, we analyze the computational complexity of actually computing a maximum crossing family of a given point set.

4.2 Complexity of Finding Maximum Crossing Families

In this section we discuss the complexity of finding a maximum crossing family for a given point set (this problem can be phrased in terms of *segment intersection graphs*). For a finite set $S = \{s_1, \ldots, s_n\}$ of objects (e.g. line segments), the intersection graph G_S has a vertex for each s_i and two vertices in G_S are adjacent if and only if the two corresponding objects intersect. If S is a set of points, we define G_S to be the segment intersection graph of the set of all $\binom{n}{2}$ line segments on S.

Decision Problem (A) *Given a set $P \subseteq \mathbb{R}^2$ of n points and a positive integer k. Does P determine a crossing family of size at least k?*

The complexity of answering this question is not settled, but we will give some evidence why it could be $\mathcal{NP}$-hard. We can modify Decision Problem (A) slightly by changing n points to n line segments:

Decision Problem (B) *Given a set $L \subseteq \mathbb{R}^2$ of n line segments and a positive integer k. Does L determine a crossing family of size at least k?*

In (A), the set of given line segments is determined by all possible line segments on a point set, whereas in (B) an arbitrary set of line segments is given. On a first glance these problems look very similar. Clearly, problem (B) is the more general and an algorithm for (B) naturally works for (A) too (but not the other way around).

Hence, proving that (B) is $\mathcal{NP}$-hard does not help us immediately but let's have a look at this anyway (it might give some useful insights for (A) as well).

Clearly, problem (B) is equivalent to finding a maximum clique in the segment intersection graph G_L, a problem which is known to be $\mathcal{NP}$-hard:

Theorem 4.6 (Cabello, Cardinal, Langerman, 2011, [CCL11]).

Let $L \subseteq \mathbb{R}^2$ be a set of line segments. Finding a maximum clique in the segment intersection graph G_L is $\mathcal{NP}$-hard.

In fact, it is even $\mathcal{NP}$-hard for ray intersection graphs. We omit the details of their construction. They give a reduction from the maximum independent set problem in planar graphs. So their main result is showing that any planar graph can be transformed to a ray intersection graph, more precisely any planar graph has an even subdivision whose complement is a ray intersection graph.

Corollary 4.7 *Decision Problem (B) is $\mathcal{NP}$-hard.*

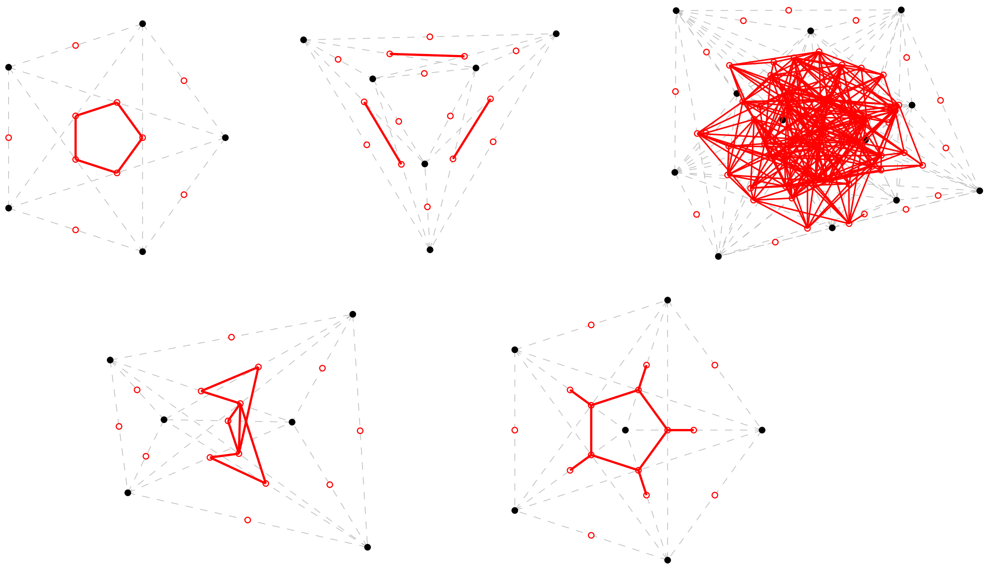


Figure 4.4: Examples of some point sets (solid, black points) and the corresponding segment intersection graphs (red). It is hard to imagine how to use the geometric structure of the point set to find the maximum cliques in the red graphs.

If we replace the set of line segments L in Theorem 4.6 by a set of points, which is a lot more restrictive, the arguments of Cabello, Cardinal and Langerman do not apply anymore:

Observation 4.8 *There exist planar graphs that do not have an even subdivision whose complement is a segment intersection graph where the segments are defined by a complete geometric graph.*

Proof Let $C = (V, E)$ be a cycle on at least four vertices and C' be a subdivision where every edge is subdivided an even number of times. Clearly, C' is still a cycle on at least four vertices and hence, the complement does not contain any vertex of degree zero. However, the segment intersection graph of any complete geometric graph contains at least three vertices of degree zero, namely those that correspond to segments of the convex hull. □

It is very unclear how a reduction (similar to the one given by Cabello et al.) for problem (A) might look like, but on the other hand it is also unclear how the additional structural information about the line segments (determined by a point set) can be used to find an efficient algorithm (see Figure 4.4 for a few examples).

Answering the following question would also settle the complexity of finding a maximum crossing family.

Open Problem 4.9.

Let $P \subseteq \mathbb{R}^2$ be a set of points. What is the complexity of finding a maximum clique in the segment intersection graph G_P of segments on P.

4.3 Greedy Algorithm for Finding a Crossing Family

The complexity of finding a maximum crossing family remains open, but still it makes sense to analyze possible (approximation) algorithms. The proof of Theorem 4.3 is fully algorithmic and hence provides a polynomial time algorithm to compute a $2^{O(\sqrt{\log n})}$-approximation. In this section we analyze a much simpler greedy algorithm.

Also note that a naive brute-force algorithm checking all subsets of the $\binom{n}{2}$ line segments has running time $2^{\binom{n}{2}}$.

Our algorithm greedily adds the line segment with most crossings (among all eligible segments) to the crossing family. Of course only line segments that intersect all previously chosen are eligible. The algorithm in pseudocode:

Algorithm 1 Greedy Crossing Family

Input: a set $P \subseteq \mathbb{R}^2$ of n points
Output: a set of pairwise crossing line segments (joining points of P)

```
1: crossing_family ← [ ]
2: eligible ← [ all (n choose 2) line segments ]
3: while eligible not empty do
4:     s_max ← segment in eligible with maximum number of intersections
5:     crossing_family ← crossing_family + s_max
6:     remove all segments from eligible that do not intersect s_max
7: end while
8: return crossing_family
```

Runtime & Correctness. Algorithm 1 can be implemented to run in $O(n^4)$. In addition to initializing the lists in step 1 and 2, which takes $O(n^2)$ time, we precompute a list with all intersections. Using a sweep line algorithm, this can be achieved in $O(n^2 \log n + k)$ time where k is the total number of intersections, i.e. $k \in O(n^4)$. This preprocessing is the expensive part. The while loop is executed at most $O(n)$ times, since in step i at least $n - i - 1$ line segments are discarded (namely those that share an endpoint with s_max). Each execution of the while loop iterates over the list of eligible segments

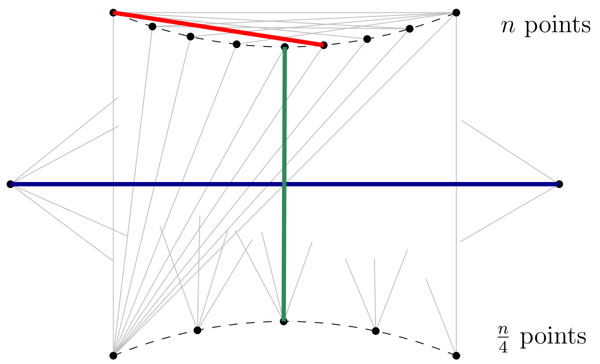


Figure 4.5: Example of a set of $n + \frac{n}{4} + 2$ points, where the greedy algorithm does not find the optimal crossing family. The black segment has $n^2/4$ intersections, the red $n^2/4 - n + 1$, and the green $1/8(n^2 + 5n + 4)$.

exactly once ($O(n^2)$). Note that we did not seek to optimize this runtime, which is likely to be improvable, e.g. by not precomputing all intersections but keeping some kind of priority queue where we can efficiently find s_max and compute its intersections with other segments.

Correctness follows from the fact that every segment in the eligible list intersects all previously chosen line segments.

The most interesting question is how large are the crossing families the greedy algorithm finds. Clearly they are not optimal in general as the example in Figure 4.5, consisting of two convex chains and two additional points, illustrates.

In Figure 4.5, the horizontal line segment has $\frac{n^2}{4}$ intersections, whereas the line segment depicted in the upper convex chain (it connects the first point with the $(n/2)$'th) has only

$$(\frac{n-2}{2})^2 = \frac{n^2}{4} - n + 1$$

intersections (neglecting rounding issues). Note that any other segment connecting two points within the same chain cannot have more intersections. Similarly, a line segment connecting two points from upper and lower convex chain has at most

$$2(\frac{n-1}{2} + 1)(\frac{n/4 - 1}{2} + 1) = \frac{1}{8}(n^2 + 5n + 4)$$

intersections.

For n large enough, this shows that the greedy algorithm will not find the largest crossing in the upper convex chain, which is of size $\frac{n}{2}$. In fact, it selects the horizontal line segment first and after that only $\frac{n}{4}$ more segments

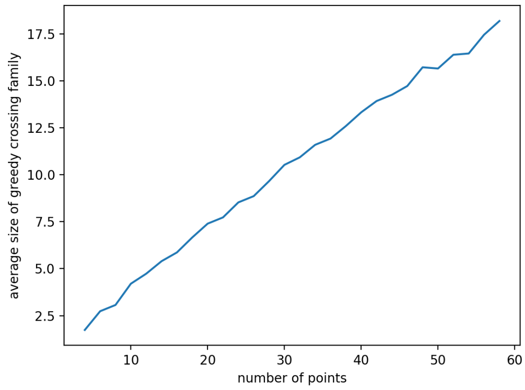


Figure 4.6: Average size of a crossing family obtained by the greedy algorithm for points drawn uniformly at random from a unit square.

can be chosen. However, in this case the greedy algorithm is only off by a constant factor, leaving hope to prove a good approximation ratio.

Open Problem 4.10.

What is the approximation ratio of the greedy crossing family?

Another fact giving hope that the greedy algorithm yields a good approximation ratio are the results of the following experiment where n points were drawn uniformly at random from a unit square and the average size of a greedy crossing family was computed over multiple rounds (see Figure 4.6). But we will see in the following section that random point sets might behave much more nicely.

4.4 Behaviour on Random Point Sets

So far we studied properties in worst-case scenarios giving guarantees like every set of n points has a spanning tree with stabbing number at most $O(\sqrt{n})$. However, often the point sets violating a property are extremely rare and one finds them only by a very careful (often degenerate or near-degenerate) construction. Therefore, it makes sense to study the situation where points are distributed randomly and analyze the behaviour of parameters in expectation, especially as a function of n (possibly for n large enough).

Of course, we first need to specify what we mean by random point sets. In the following we first fix a bounded region from which we draw points uniformly at random. A natural choice would be the unit disk. However, also other shapes, like a unit square or just some polygonal shape are often considered. Interestingly, the shape of the region has a major influence on the behaviour of parameters. For example, the number of extremal points (vertices on the convex hull) is $O(n^{1/3})$ in expectation for points drawn from a unit disk, but only $O(\log(n))$ for points drawn from a unit square or more generally $O(k \log(n))$ for points from a k-sided convex polygon (see [Har11]).

In this chapter we want to study the size of a crossing family on point sets drawn uniformly at random from a unit disk and a regular simple polygon with an even number of vertices. A *simple polygon* $S \subseteq \mathbb{R}^2$ is a compact region bounded by a simple, closed curve consisting of a finite number of line segments. A simple polygon is *regular* if all interior angles and all side lengths are equal. Any regular simple polygon is convex [Cox73].

Random Point Sets in the Unit Disk and Regular Polygons

It is known that a point set drawn uniformly at random from the unit disk has a linear size crossing family with high probability [AEG+91, Val94]. In this chapter we are extending the arguments of Valtr [Val94] to show the same result for point sets drawn from a regular simple polygon.

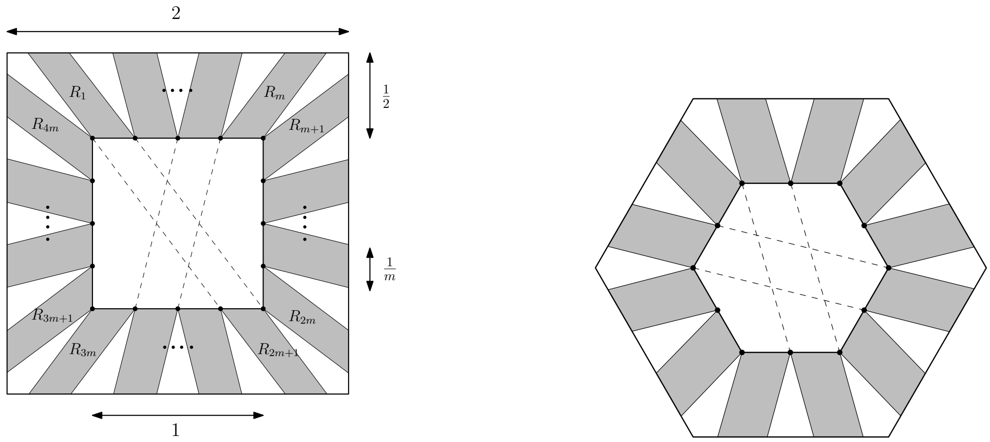
Theorem 4.11.

Any set $P \subseteq \mathbb{R}^2$ of n points drawn uniformly and independently from a regular, convex $2k$-gon ($k \in \mathbb{N}, k > 2$) contains a crossing family of size $\frac{n}{20} - O(\sqrt{n})$ with high probability.

The proof requires some work to write up but is very simple in essence. No matter which of the above shapes we are drawing our points from, we will split this shape into regions of opposing pairs as in Figure 4.7 and using an area argument compute the number of opposing pairs (as a random variable) that are occupied with points. These opposing pairs will contribute to a crossing family. And lastly we prove concentration around the expectation using a Chernoff bound.

Proof Let $\mathcal{R}$ be a regular, convex $2k$-gon. Also let the radius of the inscribed circle of $\mathcal{R}$ be 1 and call the center O. We draw a set P of n points uniformly at random. To show that P determines a crossing family of claimed size with high probability, we split $\mathcal{R}$ into regions as follows.

Let $m = \frac{n}{9k}$ be an integer and place an identical but smaller regular $2k$-gon $\mathcal{R}'$ with inscribed circle radius $1/2$ centered at O. Put $2km$ evenly distributed points $q_1, \ldots, q_{2km}$ on the boundary of $\mathcal{R}'$ (there is a point on each corner of $\mathcal{R}'$ and $m-1$ points in the interior of each edge) in clockwise order, starting



(a) $2k = 4$ and $m = 4$. Two opposite pairs are depicted by dashed lines (R_1 and R_{2m+1} for example).

(b) $2k = 6$ and $m = 2$.

Figure 4.7: Illustration of the regions forming opposite pairs for two different values of k.

from the upper left corner (these points have nothing to do with P). For $j \in \{0, \ldots, k-1\}$ and $i \in \{1, \ldots, m\}$ define R_{jm+i} and $R_{(j+k)m+i}$ to be the smaller parallelograms bounded by the boundary of $\mathcal{R}$ and $\mathcal{R}'$ and the two lines $\overline{q_{jm+i}\, q_{(j+k)m+i+1}}$ and $\overline{q_{jm+i+1}\, q_{(j+k)m+i}}$ (the gray regions in Figure 4.7).

For $i \in \{1, \ldots, km\}$, the regions R_i and R_{km+i} form opposite pairs (they share two common lines which are neither boundary lines of $\mathcal{R}$ nor $\mathcal{R}'$, the dashed lines in Figure 4.7). Clearly, any line segment joining two points of opposite pairs contributes to a crossing family (only one for each pair). Hence, we are interested in computing how many such opposing pairs are occupied by points of P.

Let X be the random variable counting the number of occupied pairs and for $i \in \{1, \ldots, km\}$ define the indicator random variables

$$X_i = \begin{cases} 1 & \text{if the } i\text{'th pair is occupied} \\ 0 & \text{otherwise} \end{cases}$$

Clearly, $X = \sum_{i=1}^{km} X_i$.

In order to compute the expectation of X, we need the probability that a fixed opposing pair is occupied. Since we drew the points uniformly and independently, we get

$$\Pr[X_i = 1] =$$
$$= \left(1 - (1 - \frac{\text{Area}(R_i)}{\text{Area}(2k\text{-gon})})^n\right)\left(1 - (1 - \frac{\text{Area}(R_{km+i})}{\text{Area}(2k\text{-gon})})^{n-1}\right). \quad (4.1)$$

Using the standard formulas [1] for computing areas of regular polygons and parallelograms we have

- $\text{Area}(2k\text{-gon}) = k \cdot \sin(\frac{\pi}{k})$
- $\text{Area}(R_i) = \frac{1}{2m} \cdot \cos(\frac{\pi}{2k}) \cdot \sin(\frac{\pi}{2k})$.

Hence,

$$\frac{\text{Area}(R_i)}{\text{Area}(2k\text{-gon})} = \frac{1}{2km} \cdot \cos(\frac{\pi}{2k}) \cdot \frac{\sin(\frac{\pi}{2k})}{\sin(\frac{\pi}{k})} \geq \frac{1}{2km} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8km}, \quad (4.2)$$

where the inequality follows from basic trigonometry (see Appendix, Lemma A.2).

Using linearity of expectation we get:

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{i=1}^{km} \mathbb{E}[X_i] = \sum_{i=1}^{km} \Pr[X_i = 1] \\
&\geq km\left(1 - (1 - \frac{1}{8km})^n\right) \cdot \left(1 - (1 - \frac{1}{8km})^{n-1}\right) \qquad [\text{plug 4.1, 4.2 in}] \\
&= \frac{n}{9}\left(1 - (1 - \frac{9}{8n})^n\right) \cdot \left(1 - (1 - \frac{9}{8n})^n - O(\frac{1}{n})\right) \qquad [m = \frac{n}{9k}] \\
&\geq \frac{n}{9}(1 - e^{-9/8}) \cdot (1 - e^{-9/8} - O(\frac{1}{n})) \\
&\geq \frac{n}{9}(1 - e^{-9/8})^2 - O(1) \\
&> \frac{n}{20} - O(1).
\end{aligned}$$

Lastly, using a Chernoff bound we show that the probability of X being less than $\mathbb{E}[X] - c \cdot \sqrt{n}$ gets arbitrarily small for large enough n and c. To simplify notation we set $\mu = \mathbb{E}[X]$.

$$\begin{aligned}
\Pr[X \leq \mu - c\sqrt{n}] &= \Pr[X \leq \mu(1 - c'\frac{1}{\sqrt{n}})] \\
&\leq e^{-\frac{1}{2}\mu \cdot \frac{c'^2}{n}} \\
&\leq e^{-\frac{c'^2}{40} + \frac{O(1)}{n}}
\end{aligned}$$

For large enough c' and n this term tends to zero. $\square$

We restricted our attention to regular polygons because in arbitrary convex polygons there is no natural way to split the region into opposing pairs. It is probably hard to adapt the simple probabilistic arguments from above to non-regular shapes. And in the light of Theorem 4.3, which already shows that any point set determines a near-linear size crossing family, this is not a very tempting problem.

4.5 Crossing Families in Higher Dimensions

Aronov et al. [AEG+91] used the notion of avoiding sets to show that any point set in the plane has a crossing family of size $\Omega(\sqrt{n})$. Pach, Rubin and Tardos [PRT19] improved this result to $\Omega(n^{1-o(1)})$ using a relaxation of avoiding sets and combinatorial results about posets.

In this section we discuss whether these methods generalize to higher dimensions. For a point set $P \subseteq R^d$ in d-dimensional space, a *crossing family* is a set of pairwise (properly) intersecting $(d-1)$-dimensional simplices (each joining d distinct points of P). Aronov et al. noted that their construction also applies in higher dimensions and guarantees a crossing family of polynomial size, however omitting a proof. We prove the following theorem (noting that this proof was recently also presented by Mirzaei and Suk [MS18]).

Theorem 4.12.
Let $P \subset \mathbb{R}^d$ be a set of n points in general position. Then P contains a crossing family of size $\Omega(n^{\frac{1-o(1)}{\prod_{i=3}^{d}(i^2-i+1)}})$.

The proof relies on finding avoiding pairs. In $\mathbb{R}^d$, two sets A and B form an avoiding pair if no hyperplane through any d points of the one set intersects the convex hull of the other set.

Lemma 4.13 (Aronov et al. [AEG+91]) *Any set $P \subseteq \mathbb{R}^d$ of n points has an avoiding pair of size $\Omega(n^{\frac{1}{d^2-d+1}})$.*

We need a few more observations before proving Theorem 4.12.

So far, we were only concerned with order types of planar point sets. In $\mathbb{R}^d$, the definition is analogous. For $P \subseteq \mathbb{R}^d$, the *order type* is a mapping that assigns any $(d+1)$-tuple $p_1, \ldots, p_{d+1}$ of points its orientation, i.e. 0 if all $d+1$ points lie on a d-dimensional hyperplane and $+1$ or -1 depending on which side of the hyperplane through $p_1, \ldots, p_d$ the point p_{d+1} lies.

One of the key steps will be to separate the avoiding pair A, B by a $(d-1)$-dimensional hyperplane $\mathcal{H}$ and consider rays from a point in A through the points in B creating intersections with $\mathcal{H}$ and inductively find a crossing family in this lower dimensional space. These intersection points have corresponding points in B and if we label them accordingly, we will show in Lemma 4.15 that the *labeled* order types are equal no matter from which point in A we shoot the rays. But let's start with the simple observation that point sets of equal order type have equal crossing families.

Lemma 4.14 *Let $P = \{p_1, \ldots, p_n\}$ and $P' = \{p'_1, \ldots, p'_n\} \subseteq \mathbb{R}^d$ be two point sets of equal (labeled) order type. Then $\mathcal{C}$ is a crossing family on P if and only if $\mathcal{C}'$, which is obtained by replacing each vertex p_i of a simplex in $\mathcal{C}$ with p'_i, is a crossing family on P'.*

Proof For the sake of contradiction, let $\Delta_1, \Delta_2 \in \mathcal{C}$ be two simplices that intersect on P but the corresponding Δ'_1, Δ'_2 do not intersect on P'. Remember that a simplex is the convex hull of d points.

Then, there is a hyperplane $\mathcal{H}$ separating Δ'_1 and Δ'_2, i.e. all extremal points of Δ'_1 and Δ'_2 are on different sides of $\mathcal{H}$, whereas no such separating hyperplane can exist for Δ_1 and Δ_2. Hence, P and P' have different (labeled) order types. □

Lemma 4.15 *Let $A, B \subseteq R^d$ be an avoiding pair and $\mathcal{H}$ a $(d-1)$-dimensional hyperplane separating A and B. For $a_i \in A$, let*

$$P_i = \{\mathcal{H} \cap \overline{a_i b} : b \in B\}.$$

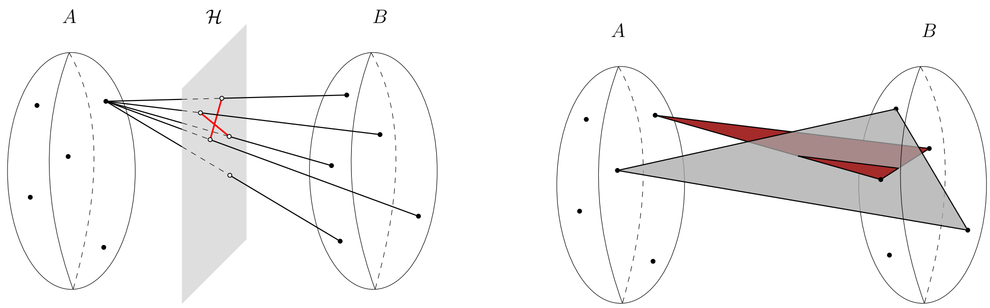
Then, for any $a_i, a_j \in A$, P_i and P_j have the same labeled order type.

Proof Consider an arbitrary d-tuple in P_i and assume that the corresponding d-tuple in P_j has different orientation (remember that P_i and P_j are $(d-1)$-dimensional). Let $b_1, \ldots, b_d$ be the corresponding points in B. Then, also the $(d+1)$-tuples $b_1, \ldots, b_d, a_i$ and $b_1, \ldots, b_d, a_j$ must have different orientations. In other words, the hyperplane through $b_1, \ldots, b_d$ has a_i and a_j on different sides, which is a contradiction to A and B being avoiding. □

Now we are ready to prove Theorem 4.12 (see also Figure 4.8).

Proof (of Theorem 4.12) We proceed by induction on d. The base case, $d = 2$, follows immediately from Theorem 4.3.

Let $P \subseteq \mathbb{R}^d$, $d > 2$, be a set of n points and assume the statement holds for any $d' < d$. Lemma 4.13 implies that there exist $A, B \subseteq P$ forming an avoiding pair of size $\Omega(n^{1/(d^2-d+1)})$. Let $\mathcal{H}$ be a $(d-1)$-dimensional hyperplane separating A and B and define $P_i = \{\mathcal{H} \cap \overline{a_i b} : b \in B\}$, as in Lemma 4.15.



(a) Separate A and B by a hyperplane $\mathcal{H}$ and find a crossing family (red) on $\mathcal{H}$.

(b) The $(d+1)$-dimensional simplices corresponding to the crossing family in (a) must intersect.

Figure 4.8: Illustration of Theorem 4.12

By the induction hypothesis, P_i has a crossing family $\mathcal{C} = \{\Delta_1, \ldots, \Delta_k\}$ of $(d-2)$-dimensional simplices of size

$$\Omega\left(|P_i|^{\frac{1-o(1)}{\prod_{i=3}^{d-1}(i^2-i+1)}}\right) = \Omega\left(n^{\frac{1-o(1)}{\prod_{i=3}^{d}(i^2-i+1)}}\right)$$

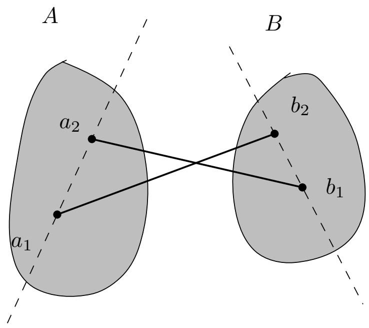
Note that by Lemma 4.14 and 4.15 we get the same crossing family for any P_j, $j \neq i$. Let B_i be the set of points in B that corresponds to Δ_i. Then $\mathcal{C}' = \{\Delta_i' = \mathrm{conv}(B_i \cap a_i) : i = 1, \ldots, k\}$ forms a crossing family on P, which can be seen as follows.

Consider two elements $\Delta_1', \Delta_2' \in \mathcal{C}'$ and let their vertices in A be a_1 and a_2. If we replace a_2 in Δ_2' by a_1 clearly both simplices intersect, namely on $\mathcal{H}$. If we shift the vertex back to a_2 and Δ_1' and Δ_2' would not be intersecting we would also have different order types induced by the rays from a_1 and a_2 on $\mathcal{H}$. Intuitively, if we think of this shifting as a continuous process, the moment Δ_1' and Δ_2' become non-intersecting, i.e. a vertex from one simplex changes from inside to outside of the other, this vertex must also change the side of the hyperplane through the $(d-2)$-dimensional simplex on $\mathcal{H}$. □

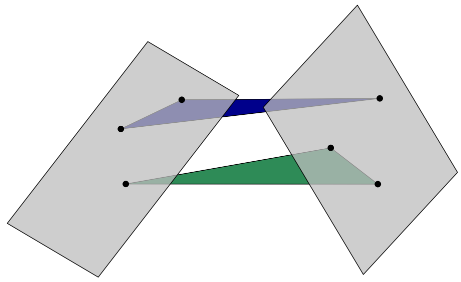
The interesting question now is, whether the arguments of Pach, Rubin and Tardos also apply in higher dimensions and yield a better bound. Many of their arguments seem to generalize to higher dimensions, though a very fundamental does not.

For sets $A, B \subset \mathbb{R}^2$ with $\mathrm{conv}(A) \cap \mathrm{conv}(B) = \emptyset$, they define a partial order $<_B$ on A as follows: $a_i <_B a_j$ if $\mathrm{conv}(B)$ is fully contained in the left halfplane of the directed line through a_i and a_j. Moreover, they prove the following lemma, which is straightforward in $\mathbb{R}^2$ (see Figure 4.9 (a)).

Lemma 4.16 ([PRT19]) *Let $A, B \subset \mathbb{R}^2$ such that $\mathrm{conv}(A) \cap \mathrm{conv}(B) = \emptyset$, $a_1 <_B a_2$, and $b_1 <_A b_2$. Then the line segments a_1b_1 and a_2b_2 intersect.*



(a) $a_2 <_B a_1$ and $b_1 <_A b_2$ implies that $a_2 b_1$ and $a_1 b_2$ intersect in $\mathbb{R}^2$.



(b) In $\mathbb{R}^3$, this does not hold anymore.

Figure 4.9: Illustration of Lemma 4.16 in 2- and 3-dimensional space.

First, it is not so clear how to define the partial order in higher dimensions. One possibility would be to define $a_1 <_B \{a_2, \ldots, a_k\}$ if $\operatorname{conv}(B)$ is fully contained above the hyperplane through $a_1, \ldots, a_k$. However, Lemma 4.16 does not guarantee us any intersections anymore (see Figure 4.9 (b)).

It seems that more work is required to apply their arguments in higher dimensions.

For large dimension d the bound in Theorem 4.12 is getting very small and is probably far from being tight. In $\mathbb{R}^2$, it has long been conjectured that any point set admits a crossing family of linear size. However, in higher dimensions this is clearly not the case anymore. Since any simplex of a crossing family contains d vertices, the trivial upper bound is $O(n/d)$. Hence, only for constant dimension we can even hope to find a linear-size crossing family. However, we believe that also for constant $d > 2$, one can construct a point set not allowing a linear size crossing family.

Conjecture 4.17.

In $\mathbb{R}^3$, there exist point sets not allowing a crossing family of linear size.

4.6 Other Objects

Lastly, we consider very briefly crossing families of certain other geometric objects, such as triangles, 2-paths, 3-paths etc. Alvarez-Rebollar, Cravioto-Lagos and Urrutia [ARCLU15] recently proved that any planar point set admits a linear size crossing family of triangles and 3-paths. A 3-path joins four distinct points of P by three line segments and two 3-paths intersect if they share at least one interior point (not belonging to P). Their proof uses very basic geometric tools and is depicted in Figure 4.10.

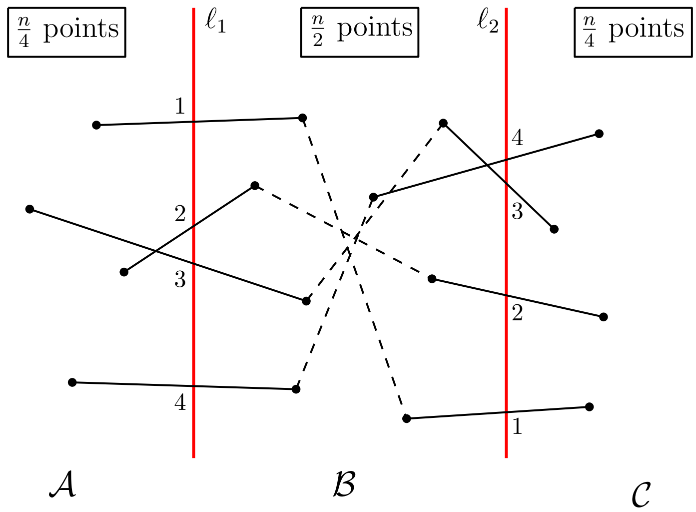


Figure 4.10: Redrawn from [ARCLU15]. Split the point set by two vertical lines ℓ_1 and ℓ_2 as depicted and match the points from $\mathcal{A}$ and $\mathcal{C}$ to points in $\mathcal{B}$ arbitrarily. Connect the endpoints in $\mathcal{B}$ according to the intersection along ℓ_1 and ℓ_2 (the dashed line segments). Then any two 3-paths intersect by the intermediate value theorem.

However, it is still open whether this also holds for 2-paths. In the light of Theorem 4.3, it probably makes more sense to ask whether there is a similarly simple geometric proof.

Another interesting observation is that there is not only always a linear size crossing family of 3-paths, but actually *all* points can be covered. Does the same hold for 2-paths?

Open Problem 4.18.

Can any set $P \subseteq \mathbb{R}^2$ of n points be partitioned into a crossing family of pairwise intersecting 2-paths (and potentially single segments if n is not divisible by 3) such that all n points are covered?

Chapter 5

Conclusion

In Chapter 2, we showed – for certain k – that the upper bound $O(\sqrt{k})$ is not achievable for the shallow stabbing number of spanning trees (Theorem 2.10). However, in general there is still a gap between lower and upper bound:

Question 5.1 *Is the bound in Theorem 2.3 tight? That is, does there exist, for any $n \in \mathbb{N}$, an n-point set $P \subseteq \mathbb{R}^2$ with k-shallow stabbing number $\Omega(\sqrt{k}\log(\frac{n}{k}))$?*

A closely related question is concerned with shallow partitions, which played a crucial role in the proof of Theorem 2.3.

Question 5.2 *Is the bound in Theorem 2.5 tight for $d = 2$? That is, does there exist for any $n \in \mathbb{N}$ an n-point set $P \subseteq \mathbb{R}^2$ such that any k-partition has crossing number $\Omega(\log(n/k))$?*

In Chapter 3, we proved that neither the tree stabbing number nor the triangulation stabbing number are monotone. Both constructions decreased the overall euclidean weight of the graphs by including more points.

Question 5.3 *Is there a relation between the non-monotonicity of stabbing numbers and the fact that the added points decrease the overall weight?*

Question 5.4 *Is there a characterization for which graph classes the stabbing number is monotone and for which not?*

In Chapter 4, we discussed a possible application of spanning trees with low stabbing number in the context of crossing families, a topic with high amount of attention during the past decades. We surveyed recent results, including the breakthrough concerning the size of crossing families by Pach, Rubin and Tardos [PRT19].

Question 5.5 *Can the methods of Pach, Rubin and Tardos be adapted for crossing families in higher dimensions?*

Another interesting question that was also raised in [PRT19] is concerned with the possibility of allowing topological curves as edges instead of straight-line segments.

Question 5.6 *For a point set $P \subseteq \mathbb{R}^2$, what size crossing family can we guarantee if the edges are topological curves that are allowed to intersect other curves only a bounded number of times?*

The notorious question about the complexity of finding a maximum crossing family also remains open.

Question 5.7 *Is it $\mathcal{NP}$-hard to compute a maximum crossing family for a given point set? Equivalently, is it $\mathcal{NP}$-hard to find a maximum clique in segment intersection graphs of complete geometric graphs?*

From an algorithmic point of view it would be interesting to evaluate the quality of the greedy crossing family.

Question 5.8 *What is the approximation ratio of the greedy crossing family?*

Finally, we considered crossing families of other geometric objects and asked the following questions.

Question 5.9 *Is there a simple geometric proof to show that any point set has a linear size family of pairwise intersecting 2-paths?*

Question 5.10 *Can any set $P \subseteq \mathbb{R}^2$ of n points be partitioned into a crossing family of pairwise intersecting 2-paths (and potentially single segments if n is not divisible by 3) such that all n points are covered?*

Bibliography

[1] https://en.wikipedia.org/wiki/Regular_polygon.

[AEG+91] Boris Aronov, Paul Erdős, Wayne Goddard, Daniel J. Kleitman, Michael Klugerman, János Pach, and Leonard J. Schulman. Crossing families. In *Proceedings of the Seventh Annual Symposium on Computational Geometry*, SCG '91, pages 351–356, New York, NY, USA, 1991. ACM.

[Afs] Peyman Afshani. personal communication via e-mail from Timothy Chan.

[Aga92] Pankaj K. Agarwal. Ray shooting and other applications of spanning trees with low stabbing number. *SIAM J. Comput.*, 21(3):540–570, June 1992.

[ARCLU15] Jose Luis Alvarez-Rebollar, Jorge Cravioto-Lagos, and Jorge Urrutia. Crossing families and self crossing hamiltonian cycles. July 2015.

[AS16] Noga Alon and Joel H. Spencer. *The Probabilistic Method*. Wiley Publishing, 4th edition, 2016.

[BCKO08] Mark de Berg, Otfried Cheong, Marc van Kreveld, and Mark Overmars. *Computational Geometry: Algorithms and Applications*. Springer-Verlag TELOS, Santa Clara, CA, USA, 3rd edition, 2008.

[CCL11] Sergio Cabello, Jean Cardinal, and Stefan Langerman. The clique problem in ray intersection graphs. *CoRR*, abs/1111.5986, 2011.

[Cox73] Harold S. M. Coxeter. *Regular Polytopes*. Dover Publication Inc., New York, 3rd edition, 1973.

[CW89] Bernard Chazelle and Emo Welzl. Quasi-optimal range searching in spaces of finite vc-dimension. *Discrete & Computational Geometry*, 4(5):467–489, Oct 1989.

[Epp18] David Eppstein. *Forbidden Configurations in Discrete Geometry*. Cambridge University Press, 2018.

[ES35] Pál Erdös and G. Szekeres. A combinatorial problem in geometry. *Compositio Mathematica*, 2:463–470, 1935.

[ES19] William Evans and Noushin Saeedi. On problems related to crossing families. *arXiv e-prints*, June 2019.

[FLM03] Sándor P. Fekete, Marco E. Lübbecke, and Henk Meijer. Minimizing the stabbing number of matchings, trees, and triangulations. *CoRR*, cs.CG/0310034, 2003.

[GKM14] Panos Giannopoulos, Maximilian Konzack, and Wolfgang Mulzer. Low-crossing spanning trees: an alternative proof and experiments. 2014.

[GP83] J. Goodman and R. Pollack. Multidimensional sorting. *SIAM Journal on Computing*, 12(3):484–507, 1983.

[GP93] Jacob E. Goodman and Richard Pollack. *Allowable Sequences and Order Types in Discrete and Computational Geometry*, pages 103–134. Springer-Verlag, Berlin, Heidelberg, 1993.

[Har09] Sariel Har-Peled. Approximating spanning trees with low crossing number. *CoRR*, abs/0907.1131, 2009.

[Har11] Sariel Har-Peled. On the expected complexity of random convex hulls. *CoRR*, abs/1111.5340, 2011.

[HS09] Sariel Har-Peled and Micha Sharir. Relative (p,epsilon)-approximations in geometry. *CoRR*, abs/0909.0717, 2009.

[KA17] Pankaj K. Agarwal. *Simplex Range Searching and Its Variants: A Review*, pages 1–30. May 2017.

[Mat91] Jiři Matoušek. Cutting hyperplane arrangements. *Discrete & Computational Geometry*, 6(3):385–406, Sep 1991.

[Mat92a] Jiří Matoušek. Efficient partition trees. *Discrete & Computational Geometry*, 8(3):315–334, Sep 1992.

[Mat92b] Jiří Matoušek. Reporting points in halfspaces. *Computational Geometry*, 2(3):169 – 186, 1992.

[Mat93a] Jiří Matoušek. Linear optimization queries. *Journal of Algorithms*, 14(3):432 – 448, 1993.

[Mat93b] Jiří Matoušek. Range searching with efficient hierarchical cuttings. *Discrete & Computational Geometry*, 10(2):157–182, Aug 1993.

[Mat02] Jiří Matoušek. *Lectures on Discrete Geometry*. Springer-Verlag, Berlin, Heidelberg, 2002.

[MP07] Joseph Mitchell and E Packer. Computing geometric structures of low stabbing number in the plane. Jan 2007.

[MR06] Wolfgang Mulzer and Günter Rote. Minimum-weight triangulation is np-hard. *CoRR*, abs/cs/0601002, 2006.

[MS18] Mozhgan Mirzaei and Andrew Suk. A positive fraction mutually avoiding sets theorem. *arXiv e-prints*, Feb 2018.

[MW12] Wolfgang Mulzer and Daniel Werner. A lower bound for shallow partitions. Jan 2012.

[PRT19] János Pach, Natan Rubin, and Gábor Tardos. Planar Point Sets Determine Many Pairwise Crossing Segments. *arXiv e-prints*, Apr 2019.

[Val94] Pavel Valtr. *Planar point sets with bounded ratios of distances*. dissertation, Freie Universität Berlin, 1994.

[Val97] Pavel Valtr. *On Mutually Avoiding Sets*, pages 324–328. Springer-Verlag, Berlin, Heidelberg, 1997.

[Wel88] Emo Welzl. Partition trees for triangle counting and other range searching problems. In *Symposium on Computational Geometry*, 1988.

[Wel92] Emo Welzl. *On spanning trees with low crossing numbers*, pages 233–249. Springer-Verlag, Berlin, Heidelberg, 1992.

Appendix A

Appendix

Lemma A.1 *For any set $P \subseteq \mathbb{R}^2$ of n points in general position, a set of representative lines contains exactly $\binom{n}{2}$ lines.*

Proof Let $\mathcal{P}$ be the set of *realizable partitions* of P. Two subsets $P_1, P_2 \subsetneq P$ form a *realizable partition* of P if $P_1 \cup P_2 = P$ and $\text{conv}(P_1) \cap \text{conv}(P_2) = \emptyset$, i.e. there is a line separating the convex hulls.

Define the function $f : \{(p,q) : p,q \in P\} \to \mathcal{P}$, which takes an ordered pair of points from P as input and returns a *realizable partition* (P_1, P_2) of P as follows. For an ordered pair (p,q) consider the directed line ℓ_{pq} through the two points (directed from p to q). Let

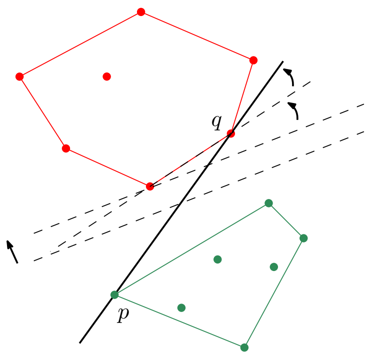
$$P_l = \{s \in P : s \text{ is left of } \overline{pq}\} \cup p$$
$$P_r = \{s \in P : s \text{ is right of } \overline{pq}\} \cup q$$

and define $f((p,q)) = (P_l, P_r)$.

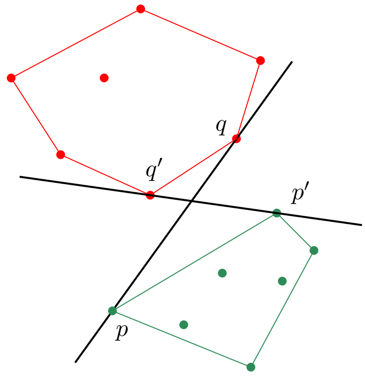
Since P is in general position, (P_l, P_r) forms a realizable partition (the line $\overline{pq}$ rotated infinitesimally in counterclockwise order separates P_l and P_r). Also note that the line $\overline{pq}$ must be tangent to $\text{conv}(P_1)$ and $\text{conv}(P_2)$ and both convex hulls are contained on different sides of $\overline{pq}$.

We will show that for any realizable partition there exist exactly two pairs (p,q) and (p',q') which are mapped to this partition. Since there are $2\binom{n}{2}$ ordered pairs on P, this proves the lemma.

Let $(P_1, P_2) \in \mathcal{P}$ be a realizable partition. Two pairs of points that are mapped to this partition can be found as follows. Consider a line ℓ separating P_1 and P_2 and shift ℓ towards P_1 until it hits a vertex. Next, rotate ℓ (once in clockwise and once in counterclockwise order) until it hits a vertex of the other set. Note that if we a hit a vertex of the same set, we just continue to rotate around this one (see Figure A.1 (a)). This way we get exactly



(a) Rotation process to find a tuple (p, q) that is mapped to the given partition.



(b) Result of both rotations (clockwise and counterclockwise).

Figure A.1: Illustration of Lemma A.1. The red and green sets form a realizable partition.

two distinct pairs of points (because of general position) that are mapped to the partition (P_1, P_2) if ordered accordingly (see Figure A.1 (b)).

On the other hand, there is no other line tangent to P_1 and P_2 simultaneously and having both sets entirely contained on different sides. □

Lemma A.2 *For any integer $k \geq 2$ the following identities hold:*

$$(i)\ \cos(\frac{\pi}{2k}) \geq \frac{1}{2} \qquad (ii)\ \frac{\sin(\frac{\pi}{2k})}{\sin(\frac{\pi}{k})} \geq \frac{1}{2}.$$

Proof The first identity follows from the monotonicity of $\cos(x)$ on $[0, \frac{\pi}{4}]$ and $\cos(\frac{\pi}{4}) \geq \frac{1}{2}$.

For the second identity, note

$$\begin{aligned} \sin(\frac{\pi}{k}) &= \sin(\frac{\pi}{2k} + \frac{\pi}{2k}) \\ &= \sin(\frac{\pi}{2k})\cos(\frac{\pi}{2k}) + \cos(\frac{\pi}{2k})\sin(\frac{\pi}{2k}) \\ &= 2\sin(\frac{\pi}{2k})\cos(\frac{\pi}{2k}). \end{aligned}$$

And hence,

$$\frac{\sin(\frac{\pi}{2k})}{\sin(\frac{\pi}{k})} = \frac{1}{2\cos(\frac{\pi}{2k})} \geq \frac{1}{2}.$$

□

Appendix B

Source Code

The source code that we used to prove Theorem 3.3 and to generate Figure 4.6 is available as github repository and can be downloaded as follows:

```
git clone https://github.com/jogo23/stabbing_number_thesis.git
```

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

Spanning Trees with Low (Shallow) Stabbing Number

Authored by (in block letters):
For papers written by groups the names of all authors are required.

Name(s):	**First name(s):**
Obenaus	Johannes

With my signature I confirm that
- I have committed none of the forms of plagiarism described in the 'Citation etiquette' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date	**Signature(s)**
Berlin, September 19, 2019	J. Obenaus

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.