Crossing Patterns of Semi-algebraic Sets*

Noga Alon† János Pach‡ Rom Pinchasi§ Radoš Radoičić¶ Micha Sharir‖

Abstract

We prove that, for every family $\mathcal{F}$ of n semi-algebraic sets in $\mathbb{R}^d$ of constant description complexity, there exist a positive constant ε that depends on the maximum complexity of the elements of $\mathcal{F}$, and two subfamilies $\mathcal{F}_1, \mathcal{F}_2 \subseteq \mathcal{F}$ with at least εn elements each, such that either every element of $\mathcal{F}_1$ intersects all elements of $\mathcal{F}_2$ or no element of $\mathcal{F}_1$ intersects any element of $\mathcal{F}_2$. This implies the existence of another constant δ such that $\mathcal{F}$ has a subset $\mathcal{F}' \subseteq \mathcal{F}$ with n^δ elements, so that either every pair of elements of $\mathcal{F}'$ intersect each other or the elements of $\mathcal{F}'$ are pairwise disjoint. The same results hold when the intersection relation is replaced by any other semi-algebraic relation. We apply these results to settle several problems in discrete geometry and in Ramsey theory.

1 Introduction

Complete bipartite interaction in graph theory and in geometry. Let $V(G)$ and $E(G)$ denote the vertex set and the edge set of a graph G, respectively. Let H be a fixed graph on k vertices. Erdős, Hajnal and Pach [EHP00] proved that every graph G with n vertices, which does not contain an induced subgraph isomorphic to H, has two disjoint subsets of vertices $V_1, V_2 \subseteq V(G)$, such that $|V_1|, |V_2| \geq \frac{1}{2}n^{1/(k-1)}$, and either all edges between V_1 and V_2 belong to G, or no edge between V_1 and V_2 belongs to G.

Note that the weaker result, where the sizes of V_1, V_2 are roughly $\log n$, instead of $n^{1/(k-1)}$, holds for any n-vertex graph, and immediately follows from Ramsey's theorem [ES35]. A related result of Erdős and Hajnal [EH89] guarantees the existence of a complete or an empty induced subgraph with $e^{c\sqrt{\log n}}$ vertices, where $c = c(H) > 0$ is a constant. See [G97, APS01] for details concerning the well known conjecture that this bound can be further improved to n^c, for some constant c, and for some partial results in this direction.

*János Pach and Micha Sharir have been supported by NSF Grant CCR-00-98246, and by a grant from the U.S.-Israeli Binational Science Foundation. Work by János Pach has also been supported by PSC-CUNY Research Award 63352-0036, and by OTKA T-032458. Noga Alon has been supported by a grant from the U.S.-Israeli Binational Science Foundation, by a grant from the Israel Science Foundation and by the Hermann Minkowski Minerva Center for Geometry at Tel Aviv University. Work by Micha Sharir has also been supported by a grant from the Israel Science Fund, Israeli Academy of Sciences, for a Center of Excellence in Geometric Computing at Tel Aviv University, and by the Hermann Minkowski–MINERVA Center for Geometry at Tel Aviv University.

†School of Mathematical Sciences, Tel Aviv University, Tel Aviv, Israel; nogaa@post.tau.ac.il

‡City College, CUNY and Courant Institute of Mathematical Sciences, New York University, New York, NY, USA; pach@cims.nyu.edu

§Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, USA; room@math.mit.edu

¶Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, USA; rados@math.mit.edu

‖School of Computer Science, Tel Aviv University, Tel Aviv, Israel, and Courant Institute of Mathematical Sciences, New York University, New York, NY, USA; michas@post.tau.ac.il

The result of [EHP00] has many geometric applications, where G encodes some pattern of interaction between geometric entities, and where one only needs to find an appropriate forbidden graph H. For example, it is well known [EET76, PS01] that, as k tends to infinity, almost all graphs with k vertices cannot be obtained as the intersection graph of a family $\mathcal{F}$ of arcwise connected sets in the plane. Therefore, there exists a constant $\delta > 0$ such that every family $\mathcal{F}$ of arcwise connected sets in the plane has two subfamilies $\mathcal{F}_1, \mathcal{F}_2 \subseteq \mathcal{F}$ with at least n^δ elements each, such that either every element of $\mathcal{F}_1$ intersects all elements of $\mathcal{F}_2$ or no element of $\mathcal{F}_1$ intersects any element of $\mathcal{F}_2$.

In the special case when $\mathcal{F}$ consists of straight-line segments, Pach and Solymosi [PS01] improved the lower bound in the last statement from n^δ to εn. As we will show, this improvement also applies to the case of general arcs, provided they have constant description complexity (see below).

The goal of this paper is to show that in many geometric applications, that involve a family $\mathcal{F}$ of n geometric objects and a relation R on $\mathcal{F}$, one can find subfamilies $\mathcal{F}_1, \mathcal{F}_2$ *of linear size*, such that either $\mathcal{F}_1 \times \mathcal{F}_2$ is fully contained in R, or $\mathcal{F}_1 \times \mathcal{F}_2$ is disjoint from R. As a consequence, we show that one can find a single subfamily $\mathcal{F}' \subseteq \mathcal{F}$ of size n^δ, for some constant δ that depends on the problem characteristics, such that either every pair of distinct elements in $\mathcal{F}' \times \mathcal{F}'$ belongs to R, or every pair of distinct elements in $\mathcal{F}' \times \mathcal{F}'$ does not belong to R.

We present a few applications of these general results. They include subsets of line segments, arcs, disks, or more general regions in the plane (or in higher fixed dimension), such that either every pair of elements in the two subsets intersect each other, or every pair of elements are disjoint; subsets of lines in 3-space, such that all lines in one subset pass above all lines in the second subset; and a few additional applications.

Complete bipartite interaction in a general semi-algebraic setting. A *real semi-algebraic set* in $\mathbb{R}^d$ is the locus of all points that satisfy a given finite Boolean combination of polynomial equations and inequalities in the d coordinates. We say that the *description complexity* of such a set is at most κ if in some representation the number of equations and inequalities is at most κ, and each of them has degree at most κ. We refer to such a representation as a *quantifier-free* representation, and note that semi-algebraic sets can also be defined using quantifiers involving additional variables, but these quantifiers can always be eliminated and yield a more explicit, quantifier-free representation of the set. See [BCR98, BPR03] for details concerning semi-algebraic sets, including quantifier elimination in such sets.

In what follows, we are given a family $\mathcal{F}$ of semi-algebraic sets of constant description complexity, and a relation R on $\mathcal{F} \times \mathcal{F}$. We assume that R is also semi-algebraic, in the following sense. Since the sets of $\mathcal{F}$ have constant description complexity, there exists a constant q, such that each set $f \in \mathcal{F}$ can be represented by a point f^* in $\mathbb{R}^q$ (say, the point whose coordinates are the coefficients of the monomials in the polynomials that define f). Then we say that R is semi-algebraic if its corresponding representation

$$R^* = \{(f^*, g^*) \in \mathbb{R}^{2q} \mid f, g \in \mathcal{F},\ (f, g) \in R\}$$

is a semi-algebraic set.

The main general result of this paper is the following.

Theorem 1.1. *Let $\mathcal{F}$ be a family of n semi-algebraic sets in $\mathbb{R}^d$ of constant description complexity, and let $R \subseteq \mathcal{F} \times \mathcal{F}$ be a fixed semi-algebraic relation on $\mathcal{F}$. Then there exist a constant $\varepsilon > 0$, which depends only on the maximum description complexity of the sets in $\mathcal{F}$ and of R, and two subfamilies $\mathcal{F}_1, \mathcal{F}_2 \subseteq \mathcal{F}$ with at least εn elements each, such that either $\mathcal{F}_1 \times \mathcal{F}_2 \subseteq R$, or $(\mathcal{F}_1 \times \mathcal{F}_2) \cap R = \emptyset$.*

A typical application of Theorem 1.1 is with R being the intersection relation. It is easy to verify that this relation is indeed semi-algebraic, as will be detailed in Section 4. Thus we obtain two subfamilies $\mathcal{F}_1, \mathcal{F}_2 \subseteq \mathcal{F}$ with at least εn elements each, such that either every element of $\mathcal{F}_1$ intersects all the elements of $\mathcal{F}_2$, or no element of $\mathcal{F}_1$ intersects any element of $\mathcal{F}_2$.

We remark that Theorem 1.1 also holds if we have two sets $\mathcal{F}, \mathcal{G}$ of semi-algebraic sets of constant description complexity, and a semi-algebraic relation $R \subseteq \mathcal{F} \times \mathcal{G}$. In this case we obtain $\varepsilon > 0$, and subsets $\mathcal{F}_1 \subseteq \mathcal{F}$, $\mathcal{G}_1 \subseteq \mathcal{G}$, with $|\mathcal{F}_1| \geq \varepsilon|\mathcal{F}|$, $|\mathcal{G}_1| \geq \varepsilon|\mathcal{G}|$, such that either $\mathcal{F}_1 \times \mathcal{G}_1 \subseteq R$, or $(\mathcal{F}_1 \times \mathcal{G}_1) \cap R = \emptyset$. This remark carries over to essentially all the applications established in this paper.

A natural extension of Theorem 1.1 is to the case where R is symmetric, and we seek a *single* subset $\mathcal{F}' \subseteq \mathcal{F}$ such that either every pair of distinct elements in $\mathcal{F}'$ satisfies R, or no such pair satisfies R. It turns out that this extension is a corollary of Theorem 1.1, except that we can no longer guarantee that $\mathcal{F}'$ has linear size. Specifically, we show:

Theorem 1.2. *Let $\mathcal{F}$ and R be as in Theorem 1.1, so that R is symmetric. Then there exist a constant $\delta > 0$, which depends only on the maximum description complexity of the sets in $\mathcal{F}$ and of R, and a subfamily $\mathcal{F}' \subseteq \mathcal{F}$ with at least n^δ elements, such that either every pair of distinct elements of $\mathcal{F}'$ belongs to R, or no such pair belongs to R.*

Let us call an n-vertex graph t-*Ramsey* if it contains no clique and no independent set of size at least t. The known quantitative proofs of Ramsey Theorem, like the one given in [ES35], show that no n-vertex graph is $\frac{1}{2}\log_2 n$-Ramsey. As shown by Erdős [E47] in one of the first applications of the probabilistic method, this is tight, up to a constant factor, namely, there are n-vertex graphs which are $2\log_2 n$-Ramsey. Despite the simplicity of Erdős' proof, there is no constructive version of it, in the sense that there is no known deterministic algorithm that constructs a $C\log n$-Ramsey graph on n vertices, where C is any absolute constant, in time which is polynomial in n. The problem of finding such an explicit construction received a considerable amount of attention, but is still wide open. Theorem 1.2 above shows that such a construction cannot be given by defining the graph using a semi-algebraic relation on a family of semi-algebraic sets of constant description complexity in fixed dimension. In fact, any n-vertex graph constructed in such a way will necessarily have a clique or an independent set of size at least n^δ for some $\delta > 0$. This can be viewed as a partial explanation of the fact that explicit constructions of $O(\log n)$-Ramsey graphs have so far remained elusive.

In particular, the above implies that if the vertices of a graph are given by n vectors in $\mathbb{R}^d$, and the adjacency relation is determined by the signs of some fixed set of (symmetric) polynomials evaluated at the corresponding vectors, the resulting graph cannot be t-Ramsey for any $t = n^{o(1)}$. This (nearly) settles a conjecture of Babai [B76], and improves a previous result of the first author [A90] that showed that such graphs cannot be t-Ramsey for $t = e^{o(\sqrt{\log n})}$.

The problem of finding explicit constructions of graphs G_n on n vertices so that neither G_n nor its complement contain large complete bipartite graphs with vertex classes of equal size is even more challenging than that of finding explicit $t(n)$-Ramsey graphs for some slowly growing functions $t(n)$. In fact, there is no known explicit construction of a graph G on n vertices such that neither G nor its complement contain a complete bipartite graph with color classes of size $n^{1/2-\varepsilon}$ each, for any $\varepsilon > 0$. Constructions of this type may yield interesting applications in the process of extracting random bits from weak sources of randomness, and have thus been considered by various researchers, with no real success. See [PR04] for the best known polynomial time construction. Here, too, Theorem 1.1 can be viewed as a partial explanation of the fact that such explicit constructions have so far remained elusive.

All the specific geometric applications that are established in this paper, as well as many other similar results, follow easily from Theorem 1.1 or from its corollary Theorem 1.2. We present two proofs of Theorem 1.1. The first proof uses a standard linearization process (see [AM94]) to transform the elements of $\mathcal{F}$ into vectors in a higher-dimensional space, and the relation R to the set of all pairs of vectors with a nonnegative scalar product. One then applies the beautiful partition theorem of Yao and Yao [YY85] (see below for details), to derive the following "linearized" version of Theorem 1.1 in which $\langle u, v\rangle$ denotes the scalar product of u and v.

Theorem 1.3. *Let U and V be finite multisets of vectors in $\mathbb{R}^d$. Then there are subsets $U' \subset U$ and $V' \subset V$ such that $|U'| \geq \frac{1}{2^{d+1}}|U|$, $|V'| \geq \frac{1}{2^{d+1}}|V|$, and either $\langle u, v\rangle \geq 0$ for all $u \in U'$, $v \in V'$, or $\langle u, v\rangle < 0$ for all $u \in U'$, $v \in V'$.*

The second proof of Theorem 1.1 uses more advanced machinery from geometric range searching, notably the results of Agarwal and Matoušek [AM94] on range searching with semi-algebraic sets. The resulting proof is somewhat simpler, more general, and more direct (since it uses heavier machinery), but supplies, in some cases, weaker estimates of the constants ε and δ.

Although both proofs use fairly standard machinery from real algebraic geometry, they are somewhat involved because they aim to establish Theorem 1.1 in full generality. However, in most applications, the linearization process used in the first proof is easy to do "by hand", and the relation R is just a conjunction of (what become bilinear) inequalities. In such cases the proof becomes much simpler, and there is no need to explicitly involve the theory of semi-algebraic sets. We will present direct derivations of several instances of the theorem, including the intersection relations for line segments and disks in the plane, and for the above/below relation for lines in 3-space.

Applications.

Intersecting segments, disks, and regions. We first give an alternative and simpler proof of the result of Pach and Solymosi [PS01]. That is, we show that, if S is a family of segments in general position in the plane, then there exist two subfamilies $S_1, S_2 \subseteq S$ of linear size, such that either every segment in S_1 crosses all segments in S_2, or no segment in S_1 crosses any segment in S_2. As a consequence, any set S of n segments in general position in the plane has a subset S' of at least n^δ segments, so that either every pair of them intersect or no such pair intersect. The constants appearing in these bounds substantially improve those given in [PS01].

We then demonstrate the generality of our approach by first obtaining similar results for the intersection relation between disks in the plane, where the linearization can also be done "by hand". In fact, as has already been mentioned, the result continues to hold for the intersection relation of any family of simply shaped regions in the plane or in any fixed dimension, and we conclude this set of applications by formulating and proving it for arbitrary semi-algebraic sets (of constant description complexity).

Lines in 3-space. Using the fact that there exists no perfect weaving pattern of five lines in $\mathbb{R}^3$ [PPW93], Erdős, Hajnal and Pach [EHP00] proved that there exists a positive constant δ such that every family $\mathcal{L}$ of n straight lines in general position in 3-space has two subfamilies $\mathcal{L}_1, \mathcal{L}_2 \subseteq \mathcal{L}$ with at least n^δ elements each, such that every element of $\mathcal{L}_1$ passes above all elements of $\mathcal{L}_2$. They have raised the question whether one can replace the bound n^δ by εn. In Section 5, we answer their question in the affirmative. Specifically, we show in Theorem 5.1 that any family $\mathcal{L}$ of n straight lines in general position in 3-space has two subfamilies $\mathcal{L}_1, \mathcal{L}_2 \subseteq \mathcal{L}$ with at least $n/64$ elements each, such that every element of $\mathcal{L}_1$ passes above all elements of $\mathcal{L}_2$.

Erdős, Hajnal and Pach [EHP00] also raised the question whether there exists a positive constant δ such that every family $\mathcal{L}$ of n straight lines in general position in 3-space contains a *tournament* on $k \geq n^\delta$ lines, that is, a sequence $\ell_1, \ell_2, \ldots, \ell_k$ of $k \geq n^\delta$ lines, such that ℓ_i passes above ℓ_j for all $i < j$. We answer this question in the affirmative as well, with $\delta = 1/6$; it is an easy corollary of Theorem 5.1, or rather a specialized version of Theorem 1.2.

Miscellaneous results. Clearly, the technique in this paper can be applied to a wide variety of similar relations. Here are two representative applications:

(a) Let C be a set of n circles in 3-space. Then there exist two subsets C_1, C_2 of C of linear size, such that either every pair in $C_1 \times C_2$ forms a link, or no such pair forms a link. Moreover, C contains a subset C' of at least n^δ circles, for some constant δ, such that either every pair of distinct circles in C' forms a link, or no such pair forms a link.

(b) Two line segments in the plane are in *T-position* if the line containing one of the segments intersects the other segment. A *segment T-graph* is a graph whose vertices are a collection of pairwise disjoint line segments in the plane, where two vertices are adjacent iff the corresponding segments are in T-position. The study of segment T-graphs has been motivated by the investigation of certain problems on common transversals for families of disjoint segments in the plane. In [AKS90] it is shown that some graphs are not segment T-graphs. Our results here imply the following stronger statement, showing that typical graphs are not segment T-graphs: Any segment T-graph contains two linear-size subsets of vertices, so that either every vertex of the first set is adjacent to every vertex of the second, or no vertex of the first set is adjacent to any vertex of the second.

The paper is organized as follows. In Section 2 we present the proof of Theorem 1.3, and then describe the first proof of Theorem 1.1 and the derivation of its corollary, Theorem 1.2, in Section 3. The second proof is given later, in Section 6. We first present the applications to intersecting segments, disks, and regions (Section 4), and to lines in 3-space (Section 5). In many of these applications, the first proof can be applied with the linearization done explicitly "by hand". The final section, Section 7, contains a brief discussion of the other problems mentioned above, together with some concluding remarks.

2 Proof of Theorem 1.3

A major tool in our analysis is the following result of Yao and Yao [YY85], that has been an important stepping stone in the early development of the theory of geometric range searching, and whose proof uses the Borsuk-Ulam theorem (see, e.g., [M03]).

Theorem 2.1 (Yao and Yao [YY85]). *Given a continuous and everywhere positive density function on $\mathbb{R}^d$, one can partition $\mathbb{R}^d$ into 2^d regions, each with mass equal to $\frac{1}{2^d}$, such that every hyperplane in $\mathbb{R}^d$ must avoid at least one of the regions.*

Moreover, the partition of $\mathbb{R}^d$ yielded by the theorem is such that each region is a convex polyhedral cone, and all cones have a common apex (the *center* in [YY85]).

It is an immediate corollary of the discrete version of the Yao-Yao theorem that, given a finite set V of vectors in $\mathbb{R}^d$, one can partition $\mathbb{R}^d$ into 2^d convex polyhedral cones with a common apex c, such that the closure of each cone contains at least $\frac{|V|}{2^d}$ vectors of V. In addition, this partition has the property that any closed halfspace fully contains one of the cones, and any open halfspace contains one of the cones, possibly without its apex c.

Let us now turn to the proof of Theorem 1.3. Observe that we may assume that at most $\frac{|U|}{2^d}$ of the vectors in U are equal to 0, and that at most $\frac{|V|}{2^d}$ of the vectors in V are equal to c. Otherwise, Theorem 1.3 follows readily.

To each vector $0 \neq u \in U$ we assign the hyperplane $H_u = \{x \in \mathbb{R}^d : \langle u, x\rangle = 0\}$. It induces a partition of $\mathbb{R}^d$ into the two halfspaces

$$\begin{aligned} H_u^+ &= \{x \in \mathbb{R}^d : \langle u, x\rangle \geq 0\}, \\ H_u^- &= \{x \in \mathbb{R}^d : \langle u, x\rangle < 0\}. \end{aligned}$$

There are two possible cases:

Case 1. For at least half of the vectors u, the positive halfspace H_u^+ contains a cone of the partition. In this case at least $\frac{1}{2^d}$ of those halfspaces contain the same cone, so they all contain the endpoints of all the vectors in a subset of V of size $\frac{|V|}{2^d}$. Thus, we have found a subset $U' \subseteq U$ of size $\frac{1}{2^{d+1}}|U|$ and a subset $V' \subseteq V$ of size $\frac{|V|}{2^d}$, such that $\langle u, v\rangle \geq 0$ for every $u \in U', v \in V'$.

Case 2. For at least half of the vectors u, the negative halfspace H_u^- contains a cone minus the center c. In this case at least $\frac{1}{2^d}$ of those halfspaces contain the same cone (minus its apex). We denote the vectors whose endpoints lie in this cone (excluding the vectors equal to c) by V'. Clearly, $|V'| \geq \frac{1}{2^d}(1 - \frac{1}{2^d})|V| > \frac{1}{2^{d+1}}|V|$. Let U' denote the set of nonzero vectors $u \in U$ such that H_u^- contains all the endpoints of the vectors of V'. Then $|U'| \geq \frac{1}{2^d}(1 - \frac{1}{2^d})|U| > \frac{1}{2^{d+1}}|U|$, and any pair of vectors $u \in U'$ and $v \in V'$ satisfies $\langle u, v\rangle < 0$. $\Box$

Remark: If all the elements of U are distinct and all the elements of V are distinct, as will be the case in most of our applications, then the sizes of the sets U', V' yielded by the theorem are at least $\frac{1}{2^d}|U| - 1$ and $\frac{1}{2^d}|V| - 1$, respectively. In addition, if the elements of U and V are in general position, again, a situation that holds in most of our applications, then the sizes of U', V' slightly further improve to $\frac{1}{2^d}|U|$ and $\frac{1}{2^d}|V|$, respectively.

3 First Proof of Theorem 1.1

We recap the discussion in the introduction: Since each of the semi-algebraic sets in $\mathcal{F}$ has description complexity $\leq \kappa$, there exists a constant $q = q(\kappa)$, such that each $f \in \mathcal{F}$ can be parametrized as a point $f^* \in \mathbb{R}^q$. Let $\mathcal{F}^*$ denote the set of these points. In addition, the relation R can be mapped into a semi-algebraic set R^* in $\mathbb{R}^{2q}$. More precisely, for any pair of sets $f, g \in \mathcal{F}$, we can express the condition $(f, g) \in R$ as a Boolean combination of polynomial equations and inequalities in the coordinates of the points f^*, g^*, and this defines the representation R^*. For each $g \in \mathcal{F}$, let Σ_g denote the set $\{f^* \in \mathbb{R}^q \mid (f^*, g^*) \in R^*\}$.

The next step is to transform the problem further so that the polynomials appearing in the definition of any of the sets Σ_g become linear. This linearization process is fairly standard, and is described in detail by Agarwal and Matoušek [AM94]. It results in an embedding φ of $\mathbb{R}^q$ as an algebraic variety within some space $\mathbb{R}^Q$ of larger dimension Q, and a transformation, which we also denote by φ, of each set Σ_g into a polyhedral region in $\mathbb{R}^Q$. More specifically, the Boolean combination that defines Σ_g remains the same, and each of the equations and inequalities that appear there is mapped into a bilinear equation or inequality.

We first replace each equation of the form $P = 0$ in the definition of R^* by the two inequalities $P \geq 0$ and $P \leq 0$. Suppose that there are now k bilinear inequalities in the definition of R^*. We run a k-step process, where the j-th step starts with two subsets $\mathcal{F}'_{j-1}, \mathcal{F}''_{j-1}$ of $\mathcal{F}$, and extracts from them subsets $\mathcal{F}'_j \subseteq \mathcal{F}'_{j-1}$, $\mathcal{F}''_j \subseteq \mathcal{F}''_{j-1}$, such that $|\mathcal{F}'_j| \geq \frac{1}{2^{Q+1}}|\mathcal{F}'_{j-1}|$, $|\mathcal{F}''_j| \geq \frac{1}{2^{Q+1}}|\mathcal{F}''_{j-1}|$, and either every pair $(f, g) \in \mathcal{F}'_j \times \mathcal{F}''_j$ is such that (f^*, g^*) satisfy the j-th inequality in the definition of R^*, or no such pair satisfies this inequality. Starting the process with $\mathcal{F}'_0, \mathcal{F}''_0 := \mathcal{F}$, it is then clear that the final pair of subsets $\mathcal{F}'_k, \mathcal{F}''_k$ are such that $|\mathcal{F}'_k|, |\mathcal{F}''_k| \geq \frac{1}{2^{k(Q+1)}}|\mathcal{F}|$, and either every pair $(f, g) \in \mathcal{F}'_k \times \mathcal{F}''_k$ satisfies R, or none of these pairs satisfies R. This is because each of the inequalities that appear in the representation of R^* has a fixed sign for every pair f^*, g^*, with $(f, g) \in \mathcal{F}'_k \times \mathcal{F}''_k$. Since R only depends on these signs, the claim follows. This completes the proof of the theorem. $\square$

Remark: By the remark at the end of the preceding section, if we assume that the sets in $\mathcal{F}$ are in general position, we can improve the constant $\frac{1}{2^{k(Q+1)}}$ yielded by the proof to $\frac{1}{2^{kQ}}$.

3.1 Proof of Theorem 1.2

Define a family of perfect graphs $\mathcal{G}$ as follows: the trivial graph with one vertex belongs to the family, and if two graphs H_1, H_2 belong to the family, then so does their disjoint union, and their join (that is, the graph obtained from their disjoint union by adding all edges between vertices of H_1 and vertices of H_2). The family $\mathcal{G}$ is the family of all *complement reducible graphs*, or *cographs* for short; see, e.g., [CPS85]. Obviously, every induced subgraph of a cograph is also a cograph, and it is easy to prove by induction that every cograph is perfect, that is, the chromatic number of every induced subgraph of it is equal to the size of the largest clique in this subgraph. It follows that any cograph on m vertices contains either a clique or an independent set of size at least $\sqrt{m}$, since if it contains no clique of size $\sqrt{m}$ its chromatic number is at most $\sqrt{m}$ and hence it contains an independent set of size at least $\sqrt{m}$.

Suppose, now, that $\mathcal{F}$ and R are as in Theorem 1.1, so that R is symmetric. Let G be a graph whose n vertices are the members of $\mathcal{F}$, where two such vertices f, g are adjacent if and only if $(f, g) \in R$. Let $h(t)$ denote the largest number h such that any induced subgraph of G on t vertices contains an induced subgraph on h vertices which is a member of $\mathcal{G}$. Clearly $h(1) = 1$. In addition, we claim that there exists an $\varepsilon > 0$ that depends only on the the maximum description complexity of the elements of $\mathcal{F}$ and of R, so that for every t, $h(t) \geq 2h(\varepsilon t)$. Indeed, in any induced subgraph of G with t vertices we can find, by Theorem 1.1, two disjoint sets of vertices $\mathcal{F}_1, \mathcal{F}_2$, each of size at least εt, such that either G contains all edges connecting a member of $\mathcal{F}_1$ and a member of $\mathcal{F}_2$, or it contains none of these edges. (Note that the theorem does not ensure that the two sets $\mathcal{F}_1, \mathcal{F}_2$ are disjoint, but this can clearly be ensured by replacing, if needed, each set $\mathcal{F}_i$ by a subset of half its size, so that the two subsets are disjoint.) By definition, the induced subgraph of G on $\mathcal{F}_i$, for $i = 1, 2$, contains an induced subgraph H_i on at least $h(\varepsilon t)$ vertices, and the desired claim follows from the definition of the class $\mathcal{G}$ of cographs. Solving the recurrence, we conclude that $h(n) \geq n^\gamma$, where $\gamma = \log_{(1/\varepsilon)} 2 > 0$ depends only on the maximum description complexity of the members of $\mathcal{F}$ and of R, implying that our graph G contains an induced subgraph on at least n^γ vertices that belongs to $\mathcal{G}$. By the discussion in the beginning of the proof, this implies that G contains either a clique or an independent set of size at least $n^{\gamma/2}$, implying the assertion of the theorem. $\square$

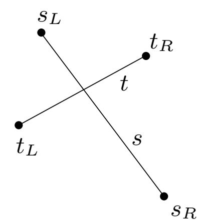

Figure 1: Intersection of the segments s and t.

4 Crossing Patterns of Segments, Disks, and Regions

In deriving the first two results, we construct the corresponding linearization explicitly, and rely directly on Theorem 1.3, thereby bypassing the general Theorem 1.1.

4.1 Crossing segments.

We first provide an alternative proof of the result of Pach and Solymosi [PS01], with considerably improved constants.

Theorem 4.1 ([PS01]). *Let S be a family of segments in general position in the plane. Then there exist two subfamilies $S_1, S_2 \subseteq S$, such that $|S_1|, |S_2| \geq \frac{1}{2^{13}}|S|$, and either every segment in S_1 crosses all segments in S_2, or no segment in S_1 crosses any segment in S_2.*

Proof: We may assume that no segment in S is vertical. We split S into two subsets S', S'' of equal size, such that the slope of every segment in S' is smaller than the slopes of all segments in S''.

Represent each segment $s \in S$ by the pair (s_L, s_R) of its left and right endpoints. Let $s \in S', t \in S''$. Then $s \cap t \neq \emptyset$ if and only if (see [dBvKOS00] and Figure 1)

$$\begin{array}{rcl} \text{Left-Turn}(s_L, s_R, t_L) & < & 0 \\ \text{Left-Turn}(s_L, s_R, t_R) & > & 0 \\ \text{Left-Turn}(t_L, t_R, s_L) & > & 0 \\ \text{Left-Turn}(t_L, t_R, s_R) & < & 0, \end{array} \tag{1}$$

where

$$\text{Left-Turn}(a, b, c) = \begin{vmatrix} 1 & x_a & y_a \\ 1 & x_b & y_b \\ 1 & x_c & y_c \end{vmatrix}.$$

We next rewrite each of the conditions in (1) as an inequality involving the scalar product of a vector that depends on s and a vector that depends on t. For example, the first inequality can be rewritten as $\langle u_1(s), v_1(t) \rangle > 0$, where

$$\begin{array}{rcl} u_1(s) & = & (x_{s_L} y_{s_R} - y_{s_L} x_{s_R},\, y_{s_R} - y_{s_L},\, x_{s_R} - x_{s_L}) \\ v_1(t) & = & (-1,\, x_{t_L},\, -y_{t_L}), \end{array}$$

and similarly for the other inequalities, where we rewrite the j-th inequality, for $j = 2, 3, 4$, as $\langle u_j(s), v_j(t) \rangle > 0$, with $u_j(s)$ and $v_j(t)$ appropriately defined.

To enforce all inequalities in (1), we apply Theorem 1.3 four times, where in each step we enforce one of the inequalities. In the first step we map S' to the set $U_1 = \{u_1(s) \mid s \in S'\}$, and map S'' to the set $V_1 = \{v_1(t) \mid t \in S''\}$. Applying Theorem 1.3 to these sets, and using the general position assumption, we conclude that there exist subsets $S_1' \subseteq S'$, $S_1'' \subseteq S''$, such that $|S_1'| \geq \frac{1}{8}|S'|$, $|S_1''| \geq \frac{1}{8}|S''|$, and either every pair of segments $(s,t) \in S_1' \times S_1''$ satisfies the first inequality in (1), or no such pair of segments satisfies it. In the latter case, no segment of S_1' intersects any segment of S_1'' and we are done. In the former case, we proceed to the next pruning step with S_1' and S_1'', and extract from them subsets S_2', S_2'' such that either all pairs in $S_2' \times S_2''$ satisfies the second inequality in (1), or no such pair satisfies it. Continuing this process for at most two more steps, we end up with subsets $S_1 \subseteq S'$, $S_2 \subseteq S''$, such that $|S_1| \geq \left(\frac{1}{8}\right)^4 \frac{|S|}{2} = \frac{1}{2^{13}}|S|$, $|S_2| \geq \frac{1}{2^{13}}|S|$, and either every pair of segments $(s,t) \in S_1 \times S_2$ intersect each other, or all such pairs are disjoint. $\square$

We remark that our constant $c = \frac{1}{2^{13}}$ is much larger than the one provided by the analysis of [PS01], and the new proof is conceptually simpler.

4.2 Crossing disks.

Our approach can be easily applied to prove Ramsey-type results of this kind for families of other geometric objects. For example, we have:

Theorem 4.2. *Let S be a family of disks in the plane. Then there exist two subfamilies $S_1, S_2 \subseteq S$, such that $|S_1|, |S_2| \geq \frac{1}{2^{10}}|S|$, and either every disk in S_1 intersects all the disks in S_2, or every disk in S_1 is disjoint from all the disks in S_2.*

Proof: Represent a disk d by the coordinates (x_d, y_d) of its center and by its radius r_d. Then a pair of disks $s, t \in S$ intersect each other if and only if $\langle u(s), v(t) \rangle \geq 0$, where

$$\begin{array}{rcl} u(s) & = & (\; -x_s^2, \; 2x_s, \; -1, \; -y_s^2, \; 2y_s, \; -1, \; r_s^2, \; 2r_s, \; 1 \;) \\ v(t) & = & (\; 1, \; x_t, \; x_t^2, \; 1, \; y_t, \; y_t^2, \; 1, \; r_t, \; r_t^2 \;). \end{array}$$

The assertion now follows from Theorem 1.3, applied in 9-space. $\square$

4.3 Crossing regions.

Finally, we consider the crossing pattern of general semi-algebraic sets, and show:

Theorem 4.3. *Let $\mathcal{F}$ be a family of semi-algebraic sets of constant description complexity in $\mathbb{R}^d$. Then there exists $\varepsilon > 0$ that depends only on the maximum description complexity of the sets in $\mathcal{F}$, and there exist two subfamilies $\mathcal{F}', \mathcal{F}'' \subseteq \mathcal{F}$ such that $|\mathcal{F}'|, |\mathcal{F}''| \geq \varepsilon|\mathcal{F}|$, and either every element of $\mathcal{F}'$ intersects all the elements of $\mathcal{F}''$, or no element of $\mathcal{F}'$ intersects any element of $\mathcal{F}''$.*

Proof: This is an immediate application of Theorem 1.1, with the relation R defined as $\{(f,g) \in \mathcal{F} \times \mathcal{F} \mid f \cap g \neq \emptyset\}$. We need to show that R is indeed semi-algebraic, in the sense defined in the introduction. Following the notation in that definition, let q be a constant dimension such that the elements of $\mathcal{F}$ can be represented as points in $\mathbb{R}^q$. Then we can represent R as

$$R^* = \{(f^*, g^*) \in \mathbb{R}^{2q} \mid f, g \in \mathcal{F} \text{ and } \exists x \in \mathbb{R}^d \mid x \in f \text{ and } x \in g\}.$$

This is clearly a semi-algebraic set in $\mathbb{R}^{2q}$. We can apply quantifier elimination (see, e.g., Theorem 2.74 of [BPR03]) to rewrite R^* as a quantifier-free semi-algebraic set. Then, for each $g \in \mathcal{F}$, the corresponding region

$$\Sigma_g = \{f^* \mid (f^*, g^*) \in R^*\}$$

is also given as a quantifier-free semi-algebraic set, and all these sets have constant description complexity. The theorem is now an immediate corollary of Theorem 1.1. □

By a similar reasoning, Theorem 1.2 implies the following:

Theorem 4.4. *Let $\mathcal{F}$ be a family of semi-algebraic sets of constant description complexity in $\mathbb{R}^d$. Then there exist $\delta > 0$ that depends on the maximum description complexity of the sets in $\mathcal{F}$, and a subfamily $\mathcal{F}' \subseteq \mathcal{F}$ of size at least n^δ, such that either every element of $\mathcal{F}'$ intersects all other elements of $\mathcal{F}'$, or no element of $\mathcal{F}'$ intersects any other element of $\mathcal{F}'$.*

Remarks.

(a) Clearly, Theorem 4.4 also applies to the two special cases studied above. For the case of segments, we obtain a subset S' of at least $n^{1/26}$ segments, so that either all of them are pairwise crossing, or all of them are pairwise disjoint. For the case of disks, the corresponding subset has at least $n^{1/20}$ disks. A related result by Aronov et al. [AEG+94] considers the set of all $\binom{n}{2}$ segments that connect n points in the plane in general position, and shows the existence of a subset of $\Omega(n^{1/2})$ segments, every pair of which intersect.

(b) Let $\mathcal{F}$ and $\mathcal{G}$ be two families of semi-algebraic sets of constant description complexity Δ with $|\mathcal{F}| = m, |\mathcal{G}| = n$. Define their *intersection graph* as a bipartite graph with vertex classes $\mathcal{F}$ and $\mathcal{G}$, where $f \in \mathcal{F}$ and $g \in \mathcal{G}$ are connected by an edge if and only if they have a point in common. As pointed out in the Introduction, Theorem 4.3 also holds in the following bipartite form: There is a constant $\varepsilon = \varepsilon(\Delta) > 0$ and subfamilies $\mathcal{F}' \subseteq \mathcal{F}, \mathcal{G}' \subseteq \mathcal{G}$ with $|\mathcal{F}'| \geq \varepsilon|\mathcal{F}|, |\mathcal{G}'| \geq \varepsilon|\mathcal{G}|$, such that either all edges between $\mathcal{F}'$ and $\mathcal{G}'$ belong to the intersection graph or none of them do.

Clearly, the total number of labeled bipartite graphs with m and n elements in their vertex classes is 2^{mn}. However, it follows from the last statement that only a negligible proportion of them can be obtained as intersection graphs of families of semi-algebraic sets of constant description complexity. Denoting by $f(m,n)$ the (base two) logarithm of the number of all such graphs, we easily obtain the recurrence:

$$f(m,n) \leq 1 + H(\varepsilon)(m+n) + f(\varepsilon m, (1-\varepsilon)n) + f((1-\varepsilon)m, \varepsilon n) + f((1-\varepsilon)m, (1-\varepsilon)n),$$

where $H(x) = -x\log_2 x - (1-x)\log_2(1-x)$ is the binary entropy function. This implies $f(m,n) = O((mn)^{1-\gamma})$, for a suitable $\gamma = \gamma(\Delta) > 0$. This bound can be further improved to $O((m+n)\log(m+n))$, by applying the Thom-Milnor-Warren theorem from real algebraic geometry to the polynomials that define the intersection relation [A90, BPR03]. We omit the details.

5 Lines in Space

In this section we show the following result.

Theorem 5.1. *Any family $\mathcal{L}$ of n straight lines in general position in 3-space has two subfamilies $\mathcal{L}_1, \mathcal{L}_2 \subseteq \mathcal{L}$ with at least $n/64$ elements each, such that every element of $\mathcal{L}_1$ passes above all elements of $\mathcal{L}_2$.*

We exploit a standard representation of lines, using Plücker coordinates (see [CEG+96]), which we briefly review here for the convenience of the reader. Let ℓ be an oriented line in $\mathbb{R}^3$, and let a, b be two points on ℓ such that ℓ is oriented from a to b. Let $[a_0, a_1, a_2, a_3]$, $[b_0, b_1, b_2, b_3]$ be the homogeneous coordinates of a and b, with a_0, b_0 being the homogenizing weights.[1] The Plücker coordinates of ℓ are the six real numbers

$$\pi(\ell) = [\pi_{01}, \pi_{02}, \pi_{12}, \pi_{03}, \pi_{13}, \pi_{23}],$$

where $\pi_{ij} = a_i b_j - a_j b_i$, for $0 \le i < j \le 3$. The most important property of this representation is that incidence between lines is a bilinear predicate. Specifically, define a second set of Plücker coordinates by

$$\varpi(\ell) = [\pi_{23}, -\pi_{13}, \pi_{03}, \pi_{12}, -\pi_{02}, \pi_{01}].$$

Then $\ell^{(1)}$ is incident to $\ell^{(2)}$ if and only if their Plücker coordinates satisfy the relationship

$$\ell^{(1)} \diamond \ell^{(2)} := \langle \pi(\ell^{(1)}),\, \varpi(\ell^{(2)}) \rangle = \pi_{01}^{(1)}\pi_{23}^{(2)} - \pi_{02}^{(1)}\pi_{13}^{(2)} + \pi_{12}^{(1)}\pi_{03}^{(2)} + \pi_{03}^{(1)}\pi_{12}^{(2)} - \pi_{13}^{(1)}\pi_{02}^{(2)} + \pi_{23}^{(1)}\pi_{01}^{(2)} = 0, \quad (2)$$

where $\pi^{(1)} = \pi(\ell^{(1)})$ and $\pi^{(2)} = \pi(\ell^{(2)})$.

The Plücker coordinates are homogeneous, and yield a mapping of lines in 3-space to points in the real projective 5-space. If we assume that the given lines are in general position, we can normalize the Plücker coordinates by setting the homogenizing weights a_0, b_0 to 1, thus obtaining points in $\mathbb{R}^5$. With some care, we can then use the $\diamond$-relation to express the relation that one line passes above another. Specifically, under this normalization, the sign of $\ell^{(1)} \diamond \ell^{(2)}$ is positive if and only if the orientation of $\ell^{(1)}$ relative to $\ell^{(2)}$, namely, the orientation of the simplex $abcd$ where $a, b \in \ell^{(1)}$, $c, d \in \ell^{(2)}$, $\ell^{(1)}$ is oriented from a to b and $\ell^{(2)}$ is oriented from c to d, is positive. Denote by $\bar{\ell}$ the projection of a nonvertical line ℓ onto the xy-plane. If we assume that neither $\ell^{(1)}$ nor $\ell^{(2)}$ is vertical, and if we orient them so that $\bar{\ell}^{(2)}$ lies clockwise to $\bar{\ell}^{(1)}$, then $\ell^{(1)} \diamond \ell^{(2)} > 0$ if and only if $\ell^{(2)}$ passes above $\ell^{(1)}$.

Now, we are ready to prove Theorem 5.1. Orient the lines of $\mathcal{L}$ so that their xy-projections are oriented from left to right. Let $\mathcal{L}^+$ (resp., $\mathcal{L}^-$) denote the subset of the $n/2$ lines of $\mathcal{L}$ whose xy-projections have the largest (resp., smallest) slopes. Set $U := \{\pi(\ell) \mid \ell \in \mathcal{L}^+\}$ and $V := \{\varpi(\ell) \mid \ell \in \mathcal{L}^-\}$. By Theorem 1.3, and the fact that our lines are all distinct and in general position, there exist subsets $\mathcal{L}_1 \subseteq \mathcal{L}^+$, $\mathcal{L}_2 \subseteq \mathcal{L}^-$, each of size at least $\frac{1}{2^5} \cdot \frac{n}{2} = \frac{n}{64}$, such that either $\ell^{(1)} \diamond \ell^{(2)} > 0$ for every pair $\ell^{(1)} \in \mathcal{L}_1$, $\ell^{(2)} \in \mathcal{L}_2$, or $\ell^{(1)} \diamond \ell^{(2)} < 0$ for every pair $\ell^{(1)} \in \mathcal{L}_1$, $\ell^{(2)} \in \mathcal{L}_2$. (We do not have equality since we have assumed that the lines are in general position). In other words, either every line of $\mathcal{L}_2$ passes above all the lines of $\mathcal{L}_1$, or every line of $\mathcal{L}_2$ passes below all the lines of $\mathcal{L}_1$. This completes the proof of Theorem 5.1. $\square$

Let $f(n)$ denote the largest integer so that any collection of n lines in general position in 3-space contains a *tournament* of $f(n)$ lines, as defined in the introduction. Then, by Theorem 5.1, we have $f(n) \ge 2f(n/64)$. Solving the recurrence, we get $f(n) \ge n^{1/6}$. This yields an affirmative answer to the question of Erdős et al. [EHP00]:

Corollary 5.2. *Every family $\mathcal{L}$ of n straight lines in general position in 3-space contains $k \ge n^{1/6}$ elements $\ell_1, \ell_2, \ldots, \ell_k$, such that ℓ_i passes above ℓ_j for all $i < j$.*

It is very likely that the exponent $1/6$ in the last statement can be replaced by a better constant c. J. Cooper and U. Wagner (personal communication) showed by an easy modification of a construction in [PT00] that c cannot exceed $\log_7 3 \approx 0.565$.

[1] This means that when $a_0 \ne 0$, the Cartesian coordinates of a are $(a_1/a_0, a_2/a_0, a_3/a_0)$, and similarly for b.

6 Second Proof of Theorem 1.1

As in the introduction, we represent the elements of $\mathcal{F}$ as points in $\mathbb{R}^q$, represent the relation R as a semi-algebraic set in $\mathbb{R}^{2q}$, and construct the regions Σ_f, for $f \in \mathcal{F}$. For the convenience of the proof, we slightly modify this notation, and consider the problem in the following setup. We have a set $\mathcal{F}$ of points in $\mathbb{R}^q$, and a family $\mathcal{G}$ of semi-algebraic sets of constant description complexity in $\mathbb{R}^q$. The goal is to show the existence of linear-size subsets $\mathcal{F}' \subseteq \mathcal{F}$, $\mathcal{G}' \subseteq \mathcal{G}$, such that either $f \in g$ for every pair $(f, g) \in \mathcal{F}' \times \mathcal{G}'$, or $f \notin g$ for every pair $(f, g) \in \mathcal{F}' \times \mathcal{G}'$. Put $m := |\mathcal{F}|$ and $n := |\mathcal{G}|$.

The arrangement $\mathcal{A}(\mathcal{G})$ of $\mathcal{G}$ is the decomposition of $\mathbb{R}^q$ into relatively open maximal connected sets (*cells*), such that each cell is contained in a fixed subset of elements of $\mathcal{G}$ and avoids all the other elements (see [SA95]). Since the elements of $\mathcal{G}$ have constant description complexity, the standard theory of real algebraic geometry (see [BPR03]) implies that the complexity of $\mathcal{A}(\mathcal{G})$, namely, the number of cells in this decomposition, is $O(n^q)$, with a constant of proportionality that depends on q and on the maximum complexity of the elements of $\mathcal{G}$.

We fix a constant parameter r, choose a random sample $\mathcal{G}_0$ of r elements of $\mathcal{G}$, and construct the arrangement $\mathcal{A}(\mathcal{G}_0)$. Next, we construct the *vertical decomposition* $\mathcal{A}^{\|}(\mathcal{G}_0)$ of $\mathcal{A}(\mathcal{G}_0)$ [CEGS89]. This is a recursively-defined decomposition of the cells of $\mathcal{A}(\mathcal{G}_0)$ into subcells of constant description complexity (which, in general, is much larger than the complexity of the elements of $\mathcal{G}$, but still a constant); see [SA95, AS00] for more details concerning vertical decompositions. As shown in [CEGS89], and enhanced by the recent improvement of [K01], the number of cells in $\mathcal{A}^{\|}(\mathcal{G}_0)$ is at most cr^2 for $q = 2$, at most $cr^3\beta(r)$ for $q = 3$, where $\beta(r)$ is an extremely slowly growing function of r related to the inverse Ackermann function, and at most $cr^{2q-4+\varepsilon}$, for any $\varepsilon > 0$, for $q \geq 4$, where in all cases c is a constant that depends on q and on the description complexity of the elements of $\mathcal{G}$ (and on ε for $q \geq 4$). We continue the proof assuming that $q \geq 4$. The other cases can be handled in a similar (and simpler) manner.

Let τ be a cell of $\mathcal{A}^{\|}(\mathcal{G}_0)$, and let $g \in \mathcal{G}$. We say that g *crosses* τ if $g \cap \tau \neq \emptyset$ but g does not fully contain τ. The standard theory of random sampling (see, e.g., [AE98, CS89, S03]) implies that, with high probability, each cell of $\mathcal{A}^{\|}(\mathcal{G}_0)$ is *crossed* by (i.e., intersects but not contained in) at most $\frac{c_1 n}{r} \log r$ elements of $\mathcal{G}$, where c_1 is a constant that depends on q and on the description complexity of the elements of $\mathcal{G}$ (but is independent of r). Let us then assume that $\mathcal{G}_0$ does indeed satisfy this property.

For each cell τ of $\mathcal{A}^{\|}(\mathcal{G}_0)$, let $\mathcal{G}_\tau$ be the subset of the elements of $\mathcal{G}$ that cross τ, and set $\mathcal{F}_\tau := \mathcal{F} \cap \tau$. There must exist a cell τ satisfying $|\mathcal{F}_\tau| \geq \frac{m}{cr^{2q-4+\varepsilon}}$. Then every element $g \in \mathcal{G} \setminus \mathcal{G}_\tau$ either fully contains τ or is disjoint from τ. Setting

$$\begin{aligned} \alpha &= \frac{1}{cr^{2q-4+\varepsilon}}, \\ \beta &= \frac{1}{2}\left(1 - \frac{c_1}{r} \log r\right) \approx \frac{1}{2}, \end{aligned}$$

we conclude that there exist a subset $\mathcal{F}' = \mathcal{F}_\tau$ of at least αm elements of $\mathcal{F}$, and a subset $\mathcal{G}'$ of at least βn elements of $\mathcal{G}$, such that either each element of $\mathcal{F}'$ is contained in every element of $\mathcal{G}'$, or no element of $\mathcal{F}'$ is contained in any element of $\mathcal{G}'$. $\square$

Discussion. (a) The second proof of Theorem 1.1 does not depend on the linearization of the elements of $\mathcal{G}$, and is therefore more general than the preceding one. Such a linearization is easy to obtain when each element of $\mathcal{G}$ is defined by a single polynomial equality or inequality, but when each element of $\mathcal{G}$ is defined by a Boolean combination of constraints, such a linearization may be difficult to obtain, without resorting to additional levels of decomposition. See, e.g., the case of crossing segments in the plane (Theorem 4.1), where the linearization-based technique had to be applied four levels in succession.

(b) It is also interesting to note that the size of $\mathcal{G}'$ can be guaranteed to be almost half the size of $\mathcal{G}$. It is not clear which of the two proofs yields a better lower bound on the size of $\mathcal{F}'$. The advantage of the first proof of Theorem 1.1 is that there are no additional hidden constants in the fractions $\frac{1}{2^{d+1}}$ (or $\frac{1}{2^d}$), whereas the constant c in the second proof is typically quite large. On the other hand, the dimension d in the first approach depends on the linearization of the elements of $\mathcal{G}$ and can be much larger than the dimension q of the ambient space in which $\mathcal{F}$ is naturally defined.

(c) From a historical perspective, the theorem of Yao and Yao is a precursor to the more general and advanced decomposition methods that have been later developed for range searching and related applications, and that we have used in the second proof. Problems that can be reduced to the setup where the Yao-Yao result can be applied benefit from this simpler and more elegant decomposition, but the new techniques allow us to extend the analysis to considerably more general situations.

7 Miscellaneous Applications and Conclusion

Theorem 1.1 easily implies the statement about segment T-graphs mentioned in the introduction: there exists an $\varepsilon > 0$ such that every collection S of n segments in the plane contains two subsets S_1, S_2, each of size at least εn, such that either every $s_1 \in S_1$ and $s_2 \in S_2$ are in T-position, or no $s_1 \in S_1$, $s_2 \in S_2$ are in T-position. Indeed, the condition of being in T-position can be expressed as the conjunction of just the first two inequalities in (1), and the proof is then just a simplified variant of the proof of Theorem 4.1, yielding the constant $\varepsilon = \frac{1}{2^7}$. By a similar reasoning, one can deduce from Theorem 1.1 that every collection of n circles in 3-space contains two subsets C_1, C_2 of linear size such that either every pair in $C_1 \times C_2$ forms a link, or no such pair forms a link. Many other variants of our general results can be similarly established.

Let $\mathcal{P}$ be a family of semi-algebraic sets in $\mathbb{R}^d$. Define its *crossing density*, $\delta(\mathcal{P})$, as the number of crossing pairs (p, p') in $\mathcal{P} \times \mathcal{P}$, divided by $|\mathcal{P}|^2$. Clearly, we have $0 \leq \delta(\mathcal{P}) \leq 1$. Similarly, define the *non-crossing density*, $\bar{\delta}(\mathcal{P})$. Then we can use the combinatorial machinery of Pach and Solymosi which is based on the regularity lemma of Szemerédi (see Theorem 3.3 in [PS01]), combined with Theorem 1.1, and obtain the following density Ramsey-type results for semi-algebraic sets.

Corollary 7.1. *Let $\mathcal{P}$ be a family of n semi-algebraic sets of constant description complexity in $\mathbb{R}^d$, such that $\delta(\mathcal{P}) \geq c > 0$. Then there exist a constant $\varepsilon > 0$, depending on c and on the maximum description complexity of the sets in $\mathcal{P}$, and two disjoint subfamilies $\mathcal{P}', \mathcal{P}'' \subseteq \mathcal{P}$, such that $|\mathcal{P}'|, |\mathcal{P}''| \geq \varepsilon n$, and every set in $\mathcal{P}'$ crosses all the sets in $\mathcal{P}''$.*

Corollary 7.2. *Let $\mathcal{P}$ be a family of n semi-algebraic sets of constant description complexity in $\mathbb{R}^d$, such that $\bar{\delta}(\mathcal{P}) \geq c > 0$. Then there exist a constant $\varepsilon > 0$, depending on c and on the maximum description complexity of sets in $\mathcal{P}$, and two disjoint subfamilies $\mathcal{P}', \mathcal{P}'' \subseteq \mathcal{P}$, such that $|\mathcal{P}'|, |\mathcal{P}''| \geq \varepsilon n$, and no set in $\mathcal{P}'$ crosses any set in $\mathcal{P}''$.*

As above, these density Ramsey-type results can be extended to cases where the intersection relation is replaced by any other semi-algebraic relation.

The lower estimate for the cardinalities of U', V' in Theorem 1.3 contains a $\frac{1}{2^{d+1}}$-factor. This exponentially small factor is indeed needed, though we do not know if the base of the exponent is tight. That is, there are examples of two sets $U, V \subset \mathbb{R}^d$ of n points each, such that the conclusion of the theorem does not hold for any two subsets $U' \subset U,\ V' \subset V$ of sizes bigger than $\frac{n}{c^d}$ for some $c > 1$. One simple example is obtained by taking $U = V = \{+1, -1\}^d$. If there are are U', V' such that $\langle u, v\rangle \geq 0$ for all $u \in U', v \in V'$ we can replace V' by $-V'$ and conclude that $\langle u, v\rangle \leq 0$ for all $u \in U',\ v \in V'$. As this holds in the second possible conclusion of the theorem as well, we can assume that this is always the case. This means that the Hamming distance between each member of U' and each member of V' is at least $d/2$, implying, by the known isoperimetric inequality for the Hamming cube (see [H66]), that $\min\{|U'|, |V'|\} \leq \sum_{i=0}^{d/4} \binom{d}{i} \leq 2^{H(1/4)d}$. This gives the required exponential dependence on d (and if we wish to have a fixed d and large n we can simply duplicate every point $n/2^d$ times). A somewhat better, similar, example can be obtained by using the usual isoperimetric inequality on the continuous unit sphere in $\mathbb{R}^d$. It is known (see, e.g., [Sch03]) that if we have two sets on the unit sphere in $\mathbb{R}^d$ and the distance between them is at least f, we can replace each set by a cap of the same measures, where the centers of the caps are antipodal points, keeping the distance at least f. It follows that if U', V' are two measurable sets on the unit sphere, and $\langle u, v\rangle \leq 0$ for all $u \in U',\ v \in V'$, then the relative measure of at least one of these sets is at most $(\frac{1+o(1)}{\sqrt{2}})^d$. By repeating the reasoning above and by letting U and V be two random sets on the unit sphere of size n each, where n tends to infinity, this implies that the assertion of Theorem 1.3 does not hold if we replace the $\frac{n}{2^{d+1}}$ estimate by more than $\frac{n}{((1-o(1))\sqrt{2})^d}$. We omit the details.

The Ramsey-type conjecture of Erdős and Hajnal, mentioned in the introduction, that any graph on n vertices which does not contain an induced copy of some fixed graph H, contains either a clique or an independent set of size n^c for some $c = c(H) > 0$, remains open. Theorem 1.2 shows that the assertion of this conjecture holds for a wide class of graphs defined by semi-algebraic relations of constant description complexity, and it may well be the case that the assertion holds for additional classes of graphs defined by geometric conditions. In particular, as mentioned in the introduction, it is known that intersection graphs of any family $\mathcal{F}$ of n arcwise connected sets in the plane contains two subfamilies $\mathcal{F}_1, \mathcal{F}_2$ of size at least n^δ each, so that either every element of $\mathcal{F}_1$ intersects every element of $\mathcal{F}_2$, or no element of $\mathcal{F}_1$ intersects any element of $\mathcal{F}_2$. It will be interesting to decide if a stronger conclusion holds: there is always one subfamily $\mathcal{F}' \subset \mathcal{F}$ of size at least n^δ such that either every two distinct elements of $\mathcal{F}'$ intersect, or no two distinct elements of $\mathcal{F}'$ intersect. This remains open.

References

[AS00] P. K. Agarwal and M. Sharir, Arrangements and their applications, in: *Handbook of Computational Geometry* (J. -R. Sack and J. Urrutia, eds.), Elsevier Science, North-Holland, Amsterdam, 2000, pp. 49–119.

[AE98] P. K. Agarwal and J. Erickson, Geometric range searching and its relatives, in: *Advances in Discrete and Computational Geometry* (B. Chazelle, J. E. Goodman and R. Pollack, eds.), AMS Press, Providence, RI, 1998, pp. 1–56.

[AM94] P.K. Agarwal and J. Matoušek, Range searching with semialgebraic sets, *Discrete Comput. Geom.* **11** (1994), 393–418.

[A90] N. Alon, Ramsey graphs cannot be defined by real polynomials, *J. Graph Theory* **14** (1990), 651–661.

[AKS90] N. Alon, M. Katchalski and E.R. Scheinerman, Not all graphs are segment T-graphs, *European J. Combinat.* **11** (1990), 7-13.

[APS01] N. Alon, J. Pach, and J. Solymosi, Ramsey-type theorems with forbidden subgraphs, *Combinatorica* **21** (2001), 155–170.

[AEG+94] B. Aronov, P. Erdős, W. Goddard, D.J. Kleitman, M. Klugerman, J. Pach and L.J. Schulman, Crossing families, *Combinatorica* **14** (1994), 127–134.

[B76] L. Babai, Open problem, in: Proc. of the Conference in Keszthely, Hungary, 1976, Vol. 2, North Holland (1978), 1189.

[BPR03] S. Basu, R. Pollack, and M.-F. Roy, *Algorithms in Real Algebraic Geometry,* Springer-Verlag, Berlin, 2003.

[dBvKOS00] M. de Berg, M. van Kreveld, M. Overmars, and O. Schwarzkopf, *Computational Geometry: Algorithms and Applications*, 2nd Edition, Springer Verlag, Heidelberg 2000.

[BCR98] J. Bochnak, M. Coste, and M.-F. Roy, *Real Algebraic Geometry,* (translated from the 1987 French original), Springer-Verlag, Berlin, 1998.

[CEGS89] B. Chazelle, H. Edelsbrunner, L. Guibas, and M. Sharir, A singly exponential stratification scheme for real semi–algebraic varieties and its applications, *Proc. 16th Int. Colloq. on Automata, Languages and Programming* (1989) pp. 179–193. Also in *Theoretical Computer Science* **84** (1991), 77–105.

[CEG+96] B. Chazelle, H. Edelsbrunner, L. J. Guibas, M. Sharir, and J. Stolfi, Lines in space: Combinatorics and algorithms, *Algorithmica* **5** (1996), 428–447.

[CPS85] D. G. Corneil, Y. Perl, and L. K. Stewart, A linear recognition algorithm for cographs, *SIAM J. Comput.* **14** (1985), 926–934.

[CS89] K. L. Clarkson and P. W. Shor, Applications of random sampling in computational geometry, II, *Discrete Comput. Geom.* **4** (1989), 387–421.

[EET76] G. Ehrlich, S. Even, and R. E. Tarjan, Intersection graphs of curves in the plane, *J. Combinat. Theory* **21** (1976), 8–20.

[E47] P. Erdős, Some remarks on the theory of graphs, *Bull. Amer. Math. Soc.* **53** (1947), 292–294.

[EH89] P. Erdős and A. Hajnal, Ramsey-type theorems, *Discrete Applied Math.* **25** (1989), 37–52.

[EHP00] P. Erdős, A. Hajnal, and J. Pach, A Ramsey-type theorem for bipartite graphs, *Geombinatorics* **10** (2000), 64–68.

[ES35] P. Erdős and G. Szekeres, A combinatorial problem in geometry, *Compositio Mathematica* **2** (1935), 463–470.

[G97] A. Gyárfás, Reflections on a problem of Erdős and Hajnal, in: *The Mathematics of Paul Erdős* (R. L. Graham and J. Nešetřil, eds.), Algorithms and Combinatorics 14, Volume II, Springer-Verlag, Heidelberg, 1997, 93–98.

[H66] L. H. Harper, Optimal numberings and isoperimetric problems on graphs, *J. Combinat. Theory* **1** (1966), 385–394.

[K01] V. Koltun, Almost tight upper bounds for vertical decomposition in four dimensions, *Proc. 42nd Annu. IEEE Sympos. Found. Comput. Sci.*, 2001, 56–65.

[M03] J. Matoušek, *Using the Borsuk-Ulam Theorem, Lectures on Topological Methods in Combinatorics and Geometry,* Springer, Heidelberg, 2003.

[PPW93] J. Pach, R. Pollack, and E. Welzl, Weaving patterns of lines and line segments in space, *Algorithmica* **9** (1993), 561–571.

[PR04] P. Pudlak and V. Rödl, Pseudorandom sets and explicit constructions of Ramsey graphs, *Quderni di Matematica*, Università di Napoli, to appear.

[PS01] J. Pach and J. Solymosi, Crossing patterns of segments, *J. Combinat. Theory, Ser. A* **96** (2001), 316–325.

[PT00] J. Pach and G. Tardos, Cutting glass, *Discrete Comput. Geom.* **24** (2000), 481–495.

[S03] M. Sharir, The Clarkson-Shor technique revisited and extended, *Combin. Probab. Comput.* **12** (2003), 191–201.

[Sch03] G. Schechtman, Concentration results and applications, in: *Handbook of the Geometry of Banach Spaces*, Vol. 2, North-Holland, Amsterdam, 2003, 1603–1634,

[SA95] M. Sharir and P.K. Agarwal, *Davenport-Schinzel Sequences and Their Geometric Applications*, Cambridge University Press, New York, 1995.

[YY85] A. C. Yao and F. F. Yao, A general approach to d-dimensional geometric queries, *Proc. 17th Annu. ACM Sympos. Theory Comput.* (1983), 163–168.