UNEXPECTED APPLICATIONS OF POLYNOMIALS IN COMBINATORICS

LARRY GUTH

In the last six years, several combinatorics problems have been solved in an unexpected way using high degree polynomials. The most well-known of these problems is the distinct distance problem in the plane. In [Erdős46], Erdős asked what is the smallest number of distinct distances determined by n points in the plane. He noted that a square grid determines $\sim n(\log n)^{-1/2}$ distinct distances, and he conjectured that this is sharp up to constant factors. Recently, an estimate was proven which is sharp up to logarithmic factors.

Theorem 0.1. *(*[Guth-Katz11]*, building on* [Elekes-Sharir10]*) For any n point set in the plane, the number of distinct distances is $\geq cn(\log n)^{-1}$.*

The main new thing in the proof is the use of high-degree polynomials. This new technique first appeared in Dvir's paper [Dvir09], which solved the finite field Nikodym and Kakeya problems. Experts had considered these problems very difficult, but the proof was essentially one page long. The method has had several other applications. The joints problem was resolved in [Guth-Katz10]. The argument was simplified and generalized in [KSS10] and [Quilodrán10], leading to another one page proof. A higher-dimensional generalization of the Szemerédi-Trotter theorem was proven in [Solymosi-Tao12]. And several fundamental theorems in incidence geometry were reproved in the paper [KMS12].

The new trick in these proofs can be summarized as follows. We want to understand some finite set S in a vector space. We consider a minimal degree (non-zero) polynomial that vanishes on the set S. Then we use this polynomial to study the problem. This strategy is somewhat surprising because the statements of the problems often involve only points and lines. The joints problem and the finite field Nikodym problem can be solved in a page each using high degree polynomials but seem very difficult to solve without polynomials. Why polynomials play such a crucial role in these problems is somewhat mysterious.

The point of this essay is to explain how these new methods work and to reflect on them philosophically. The main theme is the connection between combinatorics and algebra (polynomials).

Here is an outline of the essay.

We begin by giving two detailed examples of the polynomial method: the finite field Nikodym problem and the joints problem. This is the subject of Section 1: Examples of the polynomial method.

Once we've seen a couple examples of this method, we're going to work on understanding "where it comes from". In Section 2, we discuss where the method comes from historically.

We discuss related arguments from other areas of mathematics. Polynomials are fundamental mathematical objects, and there are many different perspectives about them. Section 2 is called ‘Perspectives on polynomials’. We will see perspectives about polynomials coming from number theory, coding theory, and differential geometry. Each of these perspectives helps to understand why polynomials are useful in these combinatorial problems.

In Section 3, we describe the new results in incidence geometry proven with polynomials, and we put them in perspective in the field. We recall the Szemerédi-Trotter theorem - a central result in the field - and discuss why the problem is difficult. We discuss one of the important methods in the field - the cutting method of [CEGSW90]. The Szemerédi-Trotter theorem involves lines in the plane. More generally, it is interesting to try to study k-dimensional objects in n-dimensional space. There are new challenges in higher dimensions. In particular, there is a new difficulty in dealing with objects of codimension > 1, such as lines in $\mathbb{R}^3$ or 2-planes in $\mathbb{R}^4$. We take some time to explain why this type of problem is hard to understand using previous methods. The distinct distance problem appears at first sight (and second and third...) as a problem about circles in the plane, but Elekes found a way to rephrase it as a problem about curves in three dimensions. In particular, we will meet two theorems about lines in $\mathbb{R}^3$ which are closely connected to the distinct distance problem and which illustrate the difficulties of incidence geometry in codimension > 1.

In Section 4, we explain how polynomials can be used to study incidence geometry. Section 4 is called ‘Combinatorial structure and algebraic structure’. We will explain the main ideas in the proofs of the two theorems at the end of Section 3. More broadly, we will try to explain the mechanisms why a configuration with a lot of combinatorial structure is forced to have a special polynomial structure.

This essay is for a volume on the mathematics of Paul Erdős. Erdős’s ideas influenced the work we describe in many ways. He posed the distinct distance problem in [Erdős46]. This paper was one of the first papers in incidence geometry, perhaps the first, and the problem has shaped many ideas in the field. I am a big admirer of hard problems that are simple to state. The most exciting - in my opinion - is a simply stated problem that is hard *for a new reason.* I think Erdős’s distance problems are such problems. They helped create and guide a whole field of math. Mathematicians working in incidence geometry have made a great effort to clarify the nature of the difficulty of these problems, and then to find methods to deal with these difficulties. We describe here one chapter of this story.

1. Examples of the polynomial method

Because some of the arguments are so short, I think the best introduction to the polynomial method is to look at some proofs. We give detailed sketches of two proofs, and then we will step back and talk about them.

1.1. The main ingredients.

There are two basic facts about polynomials which are the main ingredients in the arguments. If $\mathbb{F}$ is a field, let $Poly_D(\mathbb{F}^n)$ be the space of polynomials over $\mathbb{F}$ with degree $\leq D$ and n variables. $Poly_D(\mathbb{F}^n)$ is a vector space over $\mathbb{F}$.

Proposition 1.1. *The vector space $Poly_D(\mathbb{F}^n)$ has dimension $\binom{D+n}{n} \geq D^n/n!$.*

Proof. A basis is given by the monomials $x_1^{D_1}...x_n^{D_n}$ with $D_1 + ... + D_n \leq D$. By the 'stars and stripes' argument, the number of monomials is $\binom{D+n}{n}$. □

As a corollary, we can estimate the degree of a polynomial that vanishes at prescribed points.

Corollary 1.2. *(Parameter counting) If $S \subset \mathbb{F}^n$ is a finite set, then there is a non-zero polynomial that vanishes on S with degree $\leq n|S|^{1/n}$.*

In rough terms, when we choose a polynomial in $Poly_D(\mathbb{F}^n)$, we have $\binom{D+n}{n}$ parameters at our disposal. As long as $\binom{D+n}{n} > S$, we have enough parameters to arrange a non-zero polynomial that vanishes at every point of S. Linear algebra makes this heuristic rigorous.

Proof. We let $Fcn(S, \mathbb{F})$ be the vector space of functions from S to $\mathbb{F}$. Restricting polynomials to the set S gives a (linear) map $Poly_D(\mathbb{F}^n) \to Fcn(S, \mathbb{F})$. There is a non-zero polynomial of degree $\leq D$ vanishing on S if and only if this linear map has a non-trivial kernel. As long as the dimension of the domain is bigger than the dimension of the range, the linear map does have a non-trivial kernel. The dimension of the domain is $\binom{D+n}{n}$, and the dimension of the range is $|S|$. By a brief computation, we can always choose $D \leq n|S|^{1/n}$ so that $\binom{D+n}{n} > |S|$. □

The second main fact is that a non-zero polynomial in one variable cannot have more zeroes than its degree. A little more generally, we have the following.

Lemma 1.3. *(Vanishing lemma) If L is a line in a vector space and P is a polynomial of degree $\leq D$, and if P vanishes at $D+1$ points of L, then P vanishes on L.*

With little more than these tools, we will solve two hard problems about how lines intersect in vector spaces.

1.2. **The Nikodym problem in finite fields.**

Let $\mathbb{F}_q$ be a finite field with q elements. A set $N \subset \mathbb{F}_q^n$ is called a Nikodym set if for each point $x \in \mathbb{F}_q^n$, there is a line L so that $L \setminus \{x\} \subset N$. The question is, "how big does a Nikodym set need to be?" The paper [Dvir09] proves that a Nikodym set needs to have at least $c_n q^n$ elements - it needs to contain a definite fraction of the points in $\mathbb{F}_q^n$.

The history. The problem above is a finite-field adaptation for a problem in Euclidean geometry. A set $N \subset [0,1]^n$ is called a Nikodym set if for each $x \in [0,1]^n$, there is a line L so that N contains $L \cap [0,1]^n \setminus \{x\}$. The main question is "how big does a Nikodym set need to be?" Nikodym proved in the 20's that there are Nikodym sets of measure 0. The Nikodym conjecture says that the (Hausdorff or Minkowski) dimension of a Nikodym set is always n. (This roughly means that the δ neighborhood of a Nikodym set must contain nearly δ^{-n} δ-boxes.)

The Nikodym conjecture is a major open question in harmonic analysis. From our brief description, it's not at all clear why the problem is considered important. The Nikodym problem turns out to have connections to fundamental problems in Fourier analysis and

PDE, including the restriction problem. The restriction problem was raised by Stein in the 1960's, and it has played a major Fourier analysis ever since then. The Nikodym conjecture is a close cousin of the more well-known Kakeya conjecture. The connection between these geometrical questions and problems in Fourier analysis and PDE is described in [Laba08] and [Tao01].

Mathematicians have put a lot of effort into the Nikodym and Kakeya problems but remain far from a complete solution. Because the problems seem difficult, analysts have begun working on a variety of cousins and model problems that may shed some light back on the original problems. In [Wolff99], Wolff proposed looking at the finite field analogues of these questions. Proving that the Minkowski dimension of a Nikodym set in $[0,1]^n$ is at least α is analogous to proving that a Nikodym set in $\mathbb{F}_q^n$ has $\gtrsim q^\alpha$ elements. In particular, Dvir's theorem is analogous to the Nikodym conjecture.

The proof of the finite field Nikodym conjecture. Let us assume that $N \subset \mathbb{F}_q^n$ is a Nikodym set with $< (10n)^{-n} q^n$ elements. Let P be a non-zero polynomial that vanishes on N with minimal degree.

1. By parameter counting, the degree of P is $\le n|N|^{1/n} < q-1$.

2. By the vanishing lemma, $P(x) = 0$ at every point $x \in \mathbb{F}_q^n$. To see this, consider the line L given by the definition of the Nikodym set. We know that $x \in L$ and that $|L \cap N| \ge q-1$. So P vanishes on $q-1$ points of L, and since $deg(P) < q-1$, P must vanish on all of L.

3. Once we know that P vanishes at every point (and that $deg(P) < q-1$), it's not hard to show that all the coefficients of P are zero. In other words, P is the zero polynomial and we have a contradiction.

The Kakeya problem The Nikodym problem is a close cousin of the more well-known Kakeya problem. A Kakeya set in $\mathbb{R}^n$ is a set containing a unit line segment in each direction. The Kakeya conjecture says that any Kakeya set in $\mathbb{R}^n$ must have dimension n. A Kakeya set in $\mathbb{F}_q^n$ is a set containing a line "in every direction". More precisely, a Kakeya set contains a translate of any line in $\mathbb{F}_q^n$. By a small modification of the argument above, [Dvir09] proves that any Kakeya set in $\mathbb{F}_q^n$ contains $\ge c(n) q^n$ points.

The influence. This proof shocked the harmonic analysis community. Analysts exchange stories about where they were when they heard about it. In [Erdős], Erdős told a story about how hard it is to judge the difficulty of a problem. This is the most dramatic example that I have personally encountered. The Nikodym and Kakeya and restriction problems are closely connected, notoriously difficult problems of analysis. I believe that the finite field version was considered roughly as difficult as the original version until it was proven in one page. (To be fair, I should also say that the finite field version was only open for about ten years, and it was much less studied than the original problem.)

After the shock, people tried to adapt the new method to the original Nikodym and Kakeya problems in Euclidean space. So far, not much has been proven this way. It remains to be seen whether these methods will lead to progress in harmonic analysis. But the polynomial method has had a lot of influence in combinatorics. In this section we give one more example: the joints problem.

1.3. **The joints problem.** Suppose that $\mathfrak{L}$ is a set of L lines in $\mathbb{R}^n$. (The case $n = 3$ is a good case to keep in mind.) A joint is a point that lies in n lines of $\mathfrak{L}$ with linearly independent tangent directions. In other words, if the lines of $\mathfrak{L}$ thru x do not all lie in a hyperplane, then x is a joint. The problem is, how many joints can we make with L lines? The joints theorem says that the number of joints is $\lesssim L^{\frac{n}{n-1}}$. This number is sharp up to constant factors. For example, consider S hyperplanes in general position. Any $n-1$ hyperplanes intersect in a line, giving $L = \binom{S}{n-1}$ lines. Any n hyperplanes intersect in a point, and each of these points is a joint for our set of lines. So the number of joints is $\binom{S}{n} \sim L^{\frac{n}{n-1}}$.

The history. The joints problem was posed by Chazelle, Edelsbrunner, Guibas, Pollack, Seidel, Sharir, and Snoeyink in [CEGPSSS92]. They thought of the problem as a model problem for some difficult (still open) problems connected with computer vision. The original problem was in three dimensions. The best known bound before the polynomial method was that the number of joints is $\lesssim L^{1.62}$, [Feldman-SharirS05]. The paper [Guth-Katz10] proved the joints conjecture in three dimensions using the polynomial method. The papers [KSS10] and [Quilodrán10] simplified the proof and generalized the result to any dimension.

The proof We will prove the following main lemma: In any arrangement of lines in $\mathbb{R}^n$ with J joints, one of the lines contains $\lesssim J^{1/n}$ joints. The theorem follows from this main lemma by elementary counting. Given L lines and J joints, we remove the lines one at a time, using the main lemma to find an unpopular line to take out. Each time we remove a line, at most $J^{1/n}$ joints disappear. Therefore, $J \lesssim LJ^{1/n}$, and rearranging gives the theorem.

To prove the main lemma, we let P be a non-zero polynomial of minimal degree that vanishes on all the joints.

1. By parameter counting, the degree of P is $\lesssim J^{1/n}$.

2. If a line l contains $> deg(P)$ joints, then P vanishes on the whole line. So it suffices to find a line $l \in L$ so that P is not identically zero on l.

3. If P vanishes on all of the lines of $\mathfrak{L}$ going thru a joint x, then ∇P vanishes at x. This is because $\nabla P(x)$ vanishes in the direction tangent to each line, and the span of the tangent directions is all of $\mathbb{R}^n$. So if P vanishes on all the lines in $\mathfrak{L}$, then each partial derivative $\partial_j P$ vanishes at each joint. We know that P is not constant, so one of these partial derivatives is non-zero, and it has degree $< deg(P)$. This contradicts the definition of P as having minimal degree.

The influence. Starting from these two proofs, this little trick with high degree polynomials has become a major tool in incidence geometry. It has helped resolve several old problems and led to new proofs and perspectives about fundamental theorems. We will discuss the resulting ideas in Sections 3-4.

1.4. **Why polynomials?** The proofs of the finite field Nikodym conjecture and the joints conjecture feel like the "right" proofs to me because they are so short and because the problems seemed very difficult before. But the proofs still seems a little mysterious to me. These are questions about points and lines, and yet it seems to be crucially important to use high degree polynomials to understand them. Is it really much harder to prove these results without using high degree polynomials? If so, why are polynomials so connected with these problems? I have been thinking about these questions and discussing them with people for several years. In this essay, I will share the observations that I know. I still wish I understood the questions better.

If we play around with questions about how lines intersect in $\mathbb{R}^3$, then we will come to an important example that involves a degree 2 algebraic surface. Let's try a few questions, beginning very naively. If $\mathfrak{L}$ is a set of lines, an intersection point is a point that lies in at least two lines.

Question 0. Given L lines in $\mathbb{R}^3$, how many intersection points can there be?

There can be at most $\binom{L}{2}$ intersection points, since any two lines intersect at most once. This upper bound is sharp. If all the lines lie in a plane, and if they lie in general position within the plane, then there will be $\binom{L}{2}$ distinct intersection points.

Perhaps a set of lines in space can have many intersection points only by clustering in a plane? We can probe this issue with the following question.

Question 1. Suppose that $\mathfrak{L}$ is a set of L lines in $\mathbb{R}^3$ with ≤ 10 lines in any plane. How many intersection points can there be?

Remarkably, there can still be $\sim L^2$ intersection points. Here we come to a crucial example involving a degree 2 algebraic surface. The surface is defined by the equation $z = xy$. This surface contains a lot of lines. For any number $b \in \mathbb{R}$, let H_b be the "horizontal" line $(x, b, bx), x \in \mathbb{R}$. For any number $a \in \mathbb{R}$, let V_a be the vertical line $(a, y, ay), y \in \mathbb{R}$. The horizontal lines and the vertical lines both lie in the surface $z = xy$. The horizontal line H_b and the vertical line V_a intersect at (a, b, ab). Let $\mathfrak{L}$ consist of $L/2$ horizontal lines and $L/2$ vertical lines. These lines intersect at $L^2/4$ distinct points, so $\mathfrak{L}$ has $\gtrsim L^2$ intersection points. The intersection of a plane with the surface $z = xy$ is a degree 2 algebraic curve, and so it contains at most two lines. Therefore, any plane contains ≤ 2 lines of $\mathfrak{L}$.

(This degree 2 surface is an example of a regulus. Reguli play an important role in the approach to the joints problem in [CEGPSSS92].)

Although Question 1 is about points and lines, the key examples do not just involve linear objects (lines, planes, etc.) - they also involve algebraic surfaces. This example gives one motivation why polynomials play a role in incidence problems about lines and points.

Let's follow our investigation a bit further. Lines may have many intersection points by clustering in a plane or in a degree 2 surface. Let's forbid both types of clustering.

Many Intersections Problem. *Suppose that $\mathfrak{L}$ is a set of L lines in $\mathbb{R}^3$ with ≤ 10 lines in any plane or degree 2 surface. How many intersection points can there be?*

This time, there will be far less than L^2 intersection points. The methods of [CEGPSSS92] show that the number of intersection points is $\lesssim L^{5/3}$. Using the polynomial method, the paper [Guth-Katz11] shows that the number of intersection points is $\lesssim L^{3/2}$. This estimate plays a role in the distinct distance estimate, and we will discuss it more later.

The many intersection problem is a significant open problem. The best current upper bound on the number of intersection points is $\sim L^{3/2}$. The examples I know all have $\lesssim L$ intersection points.

We can get a little perspective on this problem by naive parameter counting. The set of lines in $\mathbb{R}^3$ is a 4-dimensional manifold. If we choose L lines, we are choosing $4L$ parameters. In fact, there is no real loss in generality in assuming that each line is given by a graph $x = az + b, y = cz + d$. So we can specify L lines by $4L$ real parameters $a_1, ..., a_L, b_1, ..., b_L$, etc. The condition that line i intersects line j can be described by one algebraic equation in the parameters $a_i, b_i, c_i, d_i, a_j, b_j, c_j, d_j$. If we want our lines to have I intersection points, then we need to solve I equations in $4L$ variables. This naive parameter counting suggests that getting significantly more than $4L$ intersection points requires some kind of structure or coincidence. A bit more rigorously, I believe that if we replace the set of lines by a "generic" 4-parameter set of curves in $\mathbb{R}^3$, then no arrangement will have more than $4L$ intersection points.

Here is the philosophical question behind the many intersections problem. Morally, any arrangement with more than $4L$ intersection points exists only because of some special structure in the set of lines. Now what special structures could the set of lines have? There is some structure from linear algebra. There is also some structure from polynomials and algebraic geometry. Are there any other 'special structures' of the set of lines in $\mathbb{R}^3$?

In summary, some important examples in incidence geometry come from algebraic surfaces. It is interesting to ask whether *all the examples* come from algebraic surfaces. The polynomial method gives an approach to prove this type of statement in some cases. The main goal of Section 4 is to explain how this works.

2. Perspectives on polynomials

In this section, we explore how this polynomial trick is connected to other parts of math. We will consider three other areas. The areas are diophantine problems in number theory, error-correcting codes in computer science, and surface area estimates in differential geometry. These areas give different perspectives on what makes polynomials special and useful functions.

2.1. There are lots of polynomials – Thue's work on diophantine approximation.

Let's begin with a warmup problem. What is the smallest possible degree of a non-zero polynomial $P \in \mathbb{R}[x, y]$ that vanishes at the million points $(j, 2^j)$ where j is an integer in the range $[1, 10^6]$?

The first approach one might try is to write down polynomials that vanish at the prescribed points. For example, we might try $\prod_{j=1}^{10^6}(x - j)$ or $\prod_{j=1}^{10^6}(y - 2^j)$. Either of these options has degree 10^6. We might try to craft a more clever formula that improves the degree. I don't know how to write down any explicit formula with degree $\le 10^5$. But the optimal degree is less than 1500. This follows by parameter counting, as in Section 1.1. The dimension of $Poly_{1498}(\mathbb{R}^2)$ is $\binom{1500}{2} > 10^6$, and so there is a non-zero polynomial of degree ≤ 1498 vanishing at all million points. This type of situation appeared in Thue's work on diophantine equations and approximation. Here is Thue's central result.

Theorem 2.1. *(Thue 1909) Suppose that β is an irrational algebraic number of degree $d > 2$. If p/q is any rational number, then*

$$|\beta - p/q| \geq c(\beta) q^{-\frac{d}{2} - 1.01}.$$

As an immediate corollary, Thue proved that a huge class of diophantine equations in two variables have only finitely many integer solutions. For example, the following equations have only finitely many integer solutions.

(1) $x^3 - 2y^3 = 6$.
(2) $x^4 + 11xy^3 + 17y^4 = 29$.
(3) $x^5 + 2x^2y^3 + 9y^5 = 9$.

Thue's corollary can be stated as follows:

Corollary 2.2. *If $P(x,y) \in \mathbb{Z}[x,y]$ is a homogeneous polynomial of degree $d \geq 3$ which is irreducible, and if n is an integer, then the equation $P(x,y) = n$ has only finitely many integer solutions $(x,y) \in \mathbb{Z}^2$.*

Thue's result was dramatically more general than any previous theorem about diophantine equations. To get a sense of how Thue's diophantine approximation result implies the corollary, consider equation 1 above. Dividing through by y^3 we get $(x/y)^3 - 2 = 6|y|^{-3}$. This formula shows that x/y is a very good rational approximation of $2^{1/3}$. With a little manipulation, it follows that $|2^{1/3} - (x/y)| \leq 100|y|^{-3}$. In contrast, Thue's theorem on diophantine approximation says that $|2^{1/3} - (x/y)| \geq c|y|^{-2.51}$. Comparing these inequalities, we see that $|y|$ is uniformly bounded, and then it follows that there are only finitely many solutions.

We will give a very partial sketch of Thue's proof, and we will see how it connects with our warmup question about polynomials.

Before Thue, the main theorem about diophantine approximation was Liouville's theorem.

Theorem 2.3. *(Liouville 1840's) If β is an algebraic number of degree $d > 1$, and p/q is any rational number, then*

$$|\beta - p/q| \geq c(\beta) q^{-d}.$$

The idea of the proof is simple and we describe it here. By assumption, β is a root of a degree d polynomial $Q(x) \in \mathbb{Z}[x]$. Since d is the minimal degree of such a polynomial, it's not hard to check that $Q(p/q)$ is non-zero. But $Q(p/q)$ is a rational number with denominator q^d, and so $|Q(p/q)| \geq q^{-d}$. But $Q(\beta) = 0$, and since Q is minimal, it's not hard to check that $Q'(\beta) \neq 0$. So $|Q(p/q)|$ has the same order of magnitude as $|\beta - p/q|$, and we see that $|\beta - p/q| \geq c(\beta) q^{-d}$. In rough terms, the polynomial Q "protects" β from rational approximations because $Q(\beta) = 0$ but $Q(p/q)$ cannot be too small.

Liouville's theorem is not strong enough to prove finiteness for any diophantine equation. When $d = 2$, Liouville's theorem is optimal, but for any $d > 2$, Thue was able to improve the exponent $-d$. Any improvement of the exponent in Liouville's theorem implies the finiteness corollary. In other words, once we know that $|\beta - p/q| \geq c(\beta) q^{-d+\epsilon}$ for any $\epsilon > 0$, then Thue's finiteness result follows.

Thue had the idea to use other polynomials besides just Q in order to protect β. Looking for other polynomials of one variable doesn't turn up anything, but Thue had the remarkable idea to use polynomials of two variables. If $P(x,y) \in \mathbb{Z}[x,y]$ is a polynomial of two variables that vanishes (maybe to high order) at (β, β), then P can "protect" β from pairs of good rational approximations $(p_1/q_1, p_2/q_2)$. To prove that $|2^{1/3} - (x/y)| \ge c|y|^{-2.51}$, Thue requires an infinite sequence of auxiliary polynomials $P_j(x,y) \in \mathbb{Z}[x,y]$ which vanish at $(2^{1/3}, 2^{1/3})$ to different orders. Each of these polynomials protects $(2^{1/3}, 2^{1/3})$ from rational approximations $(p_1/q_1, p_2/q_2)$ in certain ranges, and working all together they provide enough protection to prove Thue's theorem.

Thue carefully by hand crafted this infinite sequence of polynomials $P_j(x,y)$. He was able to construct the desired polynomials by hand when β is a d^{th} root of a rational number. He became stuck trying to generalize his method to other algebraic numbers, because he didn't know how to construct the auxiliary polynomials. The problem of looking for these auxiliary polynomials is similar to our warmup problem. At a certain point, Thue gave up trying to craft the polynomials he needed. Instead, he proved that they must exist by counting parameters, essentially as we did above.

At the 1974 ICM, Schmidt gave a lecture [Schmidt74] on Thue's work and its influence in number theory. He wrote,

"The idea of asserting the existence of certain polynomials rather than explicitly constructing them is the essential new idea in Thue's work. As Siegel [1970] points out, a study of Thue's papers reveals that Thue first tried hard to construct the polynomials explicitly (and he actually could do so in case β^d is rational)."

This idea reminds me of the probabilistic method. Thue proved that his auxiliary polynomials exist using the pigeon-hole principle. No one knows how to give an explicit formula for these polynomials, but there are so many polynomials that some of them are guaranteed to work.

Thue's wonderful argument has many similarities to the proofs in Section 1. All the arguments have the following general outline. First, by counting parameters, we find a polynomial that vanishes at certain places. Second, we use basic facts about polynomials to understand what the polynomial does at other places. Polynomials work in these arguments because they have a combination of rigidity and flexibility. Polynomials obey rigid properties like the vanishing lemma, which make them useful in the second step. On the other hand, there are lots of polynomials, which make them rather flexible in the first step. It's somewhat remarkable that such a large space of functions obeys such rigid properties.

2.2. **The resilience of polynomials – polynomials in coding theory.**

The two main ingredients in the proofs of finite field Nikodym and joints are the parameter counting lemma and the vanishing lemma. This team of ingredients appeared together earlier in the theory of error-correcting codes. Dvir has a background in coding theory, and this circle of ideas may have influenced his proof of the finite field Nikodym conjecture.

Let $\mathbb{F}_q$ be a finite field with q elements, and let $Poly_D(\mathbb{F}_q)$ be the vector space of all polynomials over $\mathbb{F}_q$ of degree $\le D$. Because of the vanishing lemma, any two polynomials in $Poly_D(\mathbb{F}_q)$ can only agree at $\le D$ points. As long as D is much less than q, any two polynomials in $Poly_D(\mathbb{F}_q)$ look very different from each other. This makes them interesting tools for building error correcting codes.

Here is a typical situation in coding theory. Q is a polynomial over $\mathbb{F}_q$ of degree $\le q/1000$. We want to transmit or save Q, but the data gets corrupted, and instead we end up with a function $F : \mathbb{F}_q \to \mathbb{F}_q$. Suppose we know that $F(x) = Q(x)$ for at least $(51/100)q$ values of x. Is it possible to recover Q from F?

It follows immediately from the vanishing lemma that Q can be recovered from F in theory. Suppose that Q_1 and Q_2 are polynomials of degree $\le q/1000$ that agree with F for $\ge (51/100)q$ values of x. Then $Q_1 - Q_2$ vanishes for at least $(2/100)q$ values of x, and so $Q_1 - Q_2$ is zero by the vanishing lemma. Hence there is only one polynomial $Q \in Poly_{q/1000}(\mathbb{F}_q)$ consistent with the data F.

But there's a deeper question that remains: can we recover Q from F *in a practical way*? The argument above tells us that we can find Q by testing all the polynomials of degree $\le q/1000$ - but the length of this procedure is more than exponential in q. In the mid-80's, Berlekamp and Welch gave a polynomial-time algorithm to recover Q from F ([BW86]).

We consider the graph of F: the set $\{(x, y) \in \mathbb{F}_q^2 | F(x) = y\}$. This graph looks like a cloud of points. Inside the cloud of points a certain algebraic structure is hidden: most of the points lie on the graph of Q. How can we search out this algebraic structure hidden in the cloud of points?

The main idea of the algorithm is to find the lowest degree non-zero polynomial $P(x, y)$ that vanishes on the graph of F. On the one-hand, we can find an optimal P with an efficient algorithm. On the other hand, this optimal P uncovers the hidden algebraic structure in the cloud of points: looking at the zero set of P, the graph of Q jumps off the page.

We begin by explaining how to find this optimal P. This discussion is closely connected to the parameter counting argument in Section 1. Suppose we want to check whether there is a non-zero polynomial of degree $\le D$ that vanishes on a set $S \subset \mathbb{F}_q^2$. Let $Poly_D(\mathbb{F}_q^2)$ denote the space of all the polynomials with degree $\le D$. Let $Fcn(S, \mathbb{F}_q)$ be the vector space of all the functions from the set S to $\mathbb{F}_q$. This is a vector space of dimension $|S|$. Let $R : Poly_D(\mathbb{F}_q^2) \to Fcn(S, \mathbb{F}_q)$ be the restriction map which restricts each polynomial to the set S. The map R is a linear map between vector spaces, and it's not hard to write down an explicit matrix for it. The basic operations of linear algebra can be done in polynomial time. We can check whether R has a non-trivial kernel, and if it does we can find a non-zero element in the kernel. Doing this for each degree D, we find in polynomial time a non-zero polynomial P which vanishes on the graph of F and has minimal degree.

In the discussion so far, we treated the variables x and y on equal terms. Berlekamp and Welch actually treat them differently. This makes sense if we look back at the problem we're trying to attack. We're hoping to find the graph of Q, which is defined by $y - Q(x) = 0$. This defining equation has degree 1 in y and high degree in x. In order to adapt to the problem, it turns out to be a good idea to use polynomials $P(x, y)$ of degree 1 in y and high degree in x. From now on we just consider polynomials $P(x, y) = P_0(x) + yP_1(x)$. By the same linear algebra argument, we can find such a polynomial P which vanishes on the graph of F, and where $\max(degP_0, degP_1)$ is as small as possible.

We can also give an estimate for this degree. If we consider P_0, P_1 of degree $\le D$, then we get a vector space of polynomials of dimension $2D+2$. We want to find a polynomial that vanishes on the graph of F, which has q points. As long as $2D+2 > q$, such a polynomial is guaranteed to exist by parameter counting. Therefore, we know that the degree of P_0, P_1 is $\le q/2$.

Let's summarize. We found a polynomial $P(x,y) = P_0(x) + yP_1(y)$ which vanishes on the graph of F, where the degrees of P_0 and P_1 are as small as possible and definitely $\le q/2$. This polynomial will help us to unlock the information hidden in F.

The key point is that P vanishes on the graph of Q! This follows in a few simple steps.

1. We know $P = 0$ on the graph of F. In other words, $P(x, F(x)) = 0$ for all x.

2. But we know that F usually agrees with Q. So $P(x, Q(x)) = 0$ for at least $(51/100)q$ values of x.

3. But $P(x, Q(x)) = P_0(x) + Q(x)P_1(x)$ is a polynomial in x of degree $\le q/2 + q/1000 < (51/100)q$.

4. By the vanishing lemma, $P(x, Q(x))$ is the zero polynomial!

We have proven that $P(x, Q(x)) = P_0(x) + Q(x)P_1(x)$ is identically zero. Hence $Q(x)P_1(x) = -P_0(x)$. We know P_0 and P_1, and now we can recover Q by doing polynomial division. This is the Berlekamp-Welch algorithm.

There is a more visual way of explaining how to recover Q, which makes the graph of Q jump off the page. We let the set of errors be $E := \{x \in \mathbb{F}_q | F(x) \neq Q(x)\}$. Adding a few more lines to the argument above, one can prove that the zero set of our polynomial P is the union of the graph of Q and a vertical line $x = e$ at each error $e \in E$. Looking at the zero set of P, the set of errors is immediately visible, together with a large chunk of the graph of Q. From this large chunk of the graph of Q, we can quickly recover Q itself.

Computer scientists working on error-correcting codes found a new set of questions about polynomials, very different from questions that pure mathematicians have considered. Working on these questions gave new perspectives about polynomials. Writing about coding theory in [Sudan95], Sudan referred to the resilience of polynomials: we can significantly distort the polynomial Q, but the information in Q survives. There is a lot more work on polynomials and coding theory. Some of it is described in [Sudan95] and in [Trevisan04]. The parameter counting lemma and the vanishing lemma continue to be important players.

2.3. **Efficiency of polynomials - polynomials in geometry.**

The last step of the proof of the distinct distance problem was influenced by ideas about polynomials in differential geometry. The overarching idea is that polynomials are geometrically efficient.

We begin with an older result about the efficiency of complex polynomials. The zero sets of complex polynomials are minimal surfaces. Let's formulate a precise result. We identify $\mathbb{C}^n$ with $\mathbb{R}^{2n}$ and equip it with the standard Euclidean metric. Let P be a complex polynomial $\mathbb{C}^n \to \mathbb{C}$. Let $Z(P)$ denote the zero set of P. If the zero set of P does not contain any critical points of P, then $Z(P)$ is a submanifold of real dimension $2n-2$.

Theorem 2.4. *(*[Federer69]*) Suppose that $P : \mathbb{C}^n \to \mathbb{C}$ is a complex polynomial, and that $F : \mathbb{R}^{2n} \to \mathbb{R}^2$ is a smooth function, so that $P = F$ outside of the unit ball $B^{2n} \subset \mathbb{R}^{2n} = \mathbb{C}^n$. Also, suppose that $Z(P)$ and $Z(F)$ don't contain any critical points, which implies that they are both manifolds. Then*

$$Vol_{2n-2} Z(P) \cap B^{2n} \le Vol_{2n-2} Z(F) \cap B^{2n}.$$

This theorem says that complex algebraic surfaces do not waste any volume.

In this section, we will be interested in analogous results for real polynomials. Initially, it seems that there can be no such result. The Weierstrauss approximation theorem says that any continuous function on a compact subset of $\mathbb{R}^n$ can be C^0 approximated by real polynomials. This basically means that real polynomials have no special properties at all.

But if we slightly shift the question, there is an interesting theorem discovered only in the last ten years. Instead of focusing on one polynomial at a time, we focus on the space $Poly_D(\mathbb{R}^n)$, the space of all polynomials of degree $\le D$. Individual polynomials may be wasteful with volume, but we will see that the space $Poly_D(\mathbb{R}^n)$ is efficient with volume. This follows from two results, one old and one new.

Proposition 2.5. *If P is a non-zero polynomial in $Poly_D(\mathbb{R}^n)$, then*

$$Vol_{n-1} Z(P) \cap B^n \le C(n) D.$$

This is a classical result. Because P is a degree D polynomial, a line can intersect $Z(P)$ at most D times unless the whole line lies in $Z(P)$. The Crofton formula describes how the volume of a hypersurface can be reconstructed in terms of the number of intersections between the surface and all of the lines in space. When we plug our estimate on the intersection numbers into the Crofton formula, it follows that the volume of $Z(P) \cap B^n$ is $\le C(n) D$.

Now comes the new result. Gromov compared $Poly_D(\mathbb{R}^n)$ with other vector spaces of the same dimension and saw that $Poly_D(\mathbb{R}^n)$ has approximately the smallest zero sets.

Theorem 2.6. *(*[Gromov03]*, see also* [Guth09]*) If W is a vector space of continuous functions $B^n \to \mathbb{R}$, and if $\dim W = \dim Poly_D(\mathbb{R}^n)$, then there exists $F \in W$ so that*

$$Vol_{n-1} Z(F) \cap B^n \ge c(n) D.$$

The proof uses a result from topology, but in some ways it is similar to the proof of finite field Nikodym or joints. A leading role is played by the fact that $\dim Poly_D(\mathbb{R}^n) \sim D^n$.

The contribution from topology is the Stone-Tukey ham sandwich theorem. The original ham sandwich theorem says that given three finite volume sets in $\mathbb{R}^3$, there is a plane that bisects all three. This theorem was proven by Banach in the late 30's. Stone and Tukey generalized the result. For one thing, they generalized it to higher dimensions, but they did much more than that. They realized that the argument does not apply only to perfectly flat planes but also to many other families of surfaces. Stone and Tukey figured out the right way to formulate the theorem, making it much more general. The formulation is based on functions instead of hypersurfaces.

We say that a continuous function F bisects a finite volume set U if the subset of U where $F > 0$ has half the volume of U and the subset where $F < 0$ has half the volume of U.

Theorem 2.7. *(Stone-Tukey 1942) Suppose W is a vector space of continuous functions on a domain $\Omega \subset \mathbb{R}^n$, so that for every non-zero $F \in W$, $Z(F)$ has measure 0. Let $U_1, ..., U_N \subset \Omega$ be finite volume sets, where $N < \dim W$. Then there is a non-zero $F \in W$ which bisects each U_i.*

We can now sketch the proof of Gromov's theorem. If there is a non-zero function $F \in W$ so that $Z(F)$ has positive (n-dimensional!) measure, then it has infinite (n-1)-dimensional volume, and we are done. So we can assume that each $Z(F)$ has measure 0, and we can apply the Stone-Tukey ham sandwich theorem. Let U_i be $\sim D^n$ disjoint balls in B^n each of radius $\sim D^{-1}$. We choose a non-zero function $F \in W$ that bisects each ball. A classical result in geometry says that a surface bisecting a ball needs to have a certain minimal volume. In fact, the smallest bisecting surface is a disk through the center of the ball.

Bisection lemma. If a hypersurface bisects $B^n(r)$, then it has volume at least $c(n)r^{n-1}$.

Therefore, $Z(F) \cap U_i \geq c(n)D^{-(n-1)}$. And $Z(F) \cap B^n \gtrsim D^n D^{-(n-1)} = D$. This finishes the sketch of Gromov's estimate.

These ideas from geometry/topology give a new twist to the polynomial method. Using linear algebra, we can find a non-zero polynomial $P \in Poly_D(\mathbb{R}^n)$ that vanishes on a set of points $p_1, ..., p_N$ as long as $N < \dim Poly_D(\mathbb{R}^n)$. This fact plays a key role in the solutions of the finite field Nikodym problem and the joints problem. Now using the Stone-Tukey theorem from topology, we can find a non-zero polynomial $P \in Poly_D(\mathbb{R}^n)$ that bisects some sets $U_1, ..., U_N$ as long as $N < \dim Poly_D(\mathbb{R}^n)$. The proof of the distinct distance estimate uses this new twist. We will explain how to use it in Section 4.

3. Some methods and problems in incidence geometry

In this section, we describe the impact of the polynomial method in incidence geometry. We begin by recalling some important results and methods in the subject. Then we will come to the new applications of the polynomial method. We will try to motivate these results, and we will discuss why they are hard to prove with previous methods.

This section motivates the results, and in the next section, we will discuss the proofs of these results.

3.1. **Incidence theory in the plane.**

Suppose that $\mathfrak{L}$ is a set of lines in the plane. Let $S_r(\mathfrak{L})$ be the set of r-rich points: the set of points that lie in $\geq r$ lines of $\mathfrak{L}$. One of the basic questions in the field is, "for a given number of lines and a given number of r, how big can $S_r(\mathfrak{L})$ be ?" This question was answered in a fundamental theorem of Szemerédi and Trotter.

Theorem 3.1. [ST83] *If $\mathfrak{L}$ is a set of L lines in the plane, then $|S_r(\mathfrak{L})| \lesssim L^2 r^{-3} + L r^{-1}$.*

This theorem is a central result of incidence geometry.

The first estimates about this problem exploit the following basic fact:

Basic Fact. *Two lines intersect in at most one point.*

Using just this fact and doing some counting arguments, we get some basic estimates. We call these estimates 'basic' because they follow just from the basic fact above.

Basic estimate 1. $|S_r(\mathfrak{L})| \lesssim L^2 r^{-2}$.

At each point of $S_r(\mathfrak{L})$, there are $\binom{r}{2}$ pairs of lines intersecting. In total, there are only $\binom{L}{2}$ pairs of lines, and each pair only intersects once. Therefore, $|S_r(\mathfrak{L})| \le \binom{L}{2}\binom{r}{2}^{-1} \sim L^2 r^{-2}$. Another short counting argument gives the following further estimate.

Basic estimate 2. If $r \ge 2L^{1/2}$, then $|S_r(\mathfrak{L})| \lesssim L/r$.

These estimates are not as strong as the conclusion of the theorem. For example, if $r = L^{1/2}$, then the theorem says that $|S_r(\mathfrak{L})| \lesssim L^{1/2}$, but the basic estimates give only $\lesssim L$.

There is a crucial example in the story showing that a proof of the Szemerédi-Trotter theorem requires some quite different ideas. The example involves lines over finite fields. Let $\mathbb{F}_q$ denote the finite field with q elements. Let $\mathfrak{L}$ be the set of q^2 non-vertical lines $y = mx + b$, $m, b \in \mathbb{F}_q$. Each point of $\mathbb{F}_q^2$ lies in q different lines of $\mathfrak{L}$. So we have $|S_q(\mathfrak{L})| = q^2$. Since $q = L^{1/2}$, we have $|S_{L^{1/2}}(\mathfrak{L})| = L$. For L lines in $\mathbb{R}^2$, the Szemerédi-Trotter theorem gives the much better bound $|S_{L^{1/2}}(\mathfrak{L})| \lesssim L^{1/2}$. Now it is still true in $\mathbb{F}_q^2$ that two lines intersect in at most one point. Therefore, we cannot possibly prove the Szemerédi-Trotter theorem just by exploiting the fact that two lines intersect in at most one point.

The main philosophical issue in the proof is to figure out what other information about lines in $\mathbb{R}^2$ we can use. We need to use something that is true in $\mathbb{R}^2$ but false in $\mathbb{F}_q^2$. There are several approaches to the problem, and in some way they all use the topology of the plane.

3.2. The cutting method.

The cutting method was introduced by Clarkson, Edelsbrunner, Guibas, Sharir, and Welzl in [CEGSW90]. They used the method to give an elegant proof of the Szemerédi-Trotter theorem. They were also able to prove incidence geometry results in higher dimensions. We will discuss this more below. Cutting plays a crucial role in the later applications of the polynomial method.

We illustrate the cutting method by describing the main idea of the proof of the Szemerédi-Trotter theorem. The proof is a divide-and-conquer argument. We cut the plane into pieces using D red lines. Here $D << L$ is a parameter we can play with, and the D red lines don't have to be lines from $\mathfrak{L}$. The complement of the red lines consists of convex polygonal cells. The idea is that we use the basic estimates for the points and lines in each cell, and then sum up the pieces. This idea works well as long as the lines of $\mathfrak{L}$ and the points of $S_r(\mathfrak{L})$ are evenly distributed among the cells.

Let's be a little more precise about what we may hope for. The D red lines cut the plane into $\sim D^2$ cells. If the points were evenly distributed among the cells, we would have the following:

Equidistribution 1. Each cell contains $\lesssim |S_r(\mathfrak{L})| D^{-2}$ points of $S_r(\mathfrak{L})$.

Now a line may enter at most $D+1$ cells, because it can only cross each red line once. Since there are $\sim D^2$ cells, each line enters only a small fraction of the cells. If the lines were evenly distributed among the cells, we would have the following

Equidistribution 2. Each open cell intersects $\lesssim LD^{-1}$ lines of $\mathfrak{L}$.

If we are allowed to choose any D, and find D red lines that evenly distribute $S_r(\mathfrak{L})$ and $\mathfrak{L}$, then using the basic estimates in each cell and adding the results we get the conclusion of the Szemerédi-Trotter theorem. In fact, we don't need to evenly distribute both $S_r(\mathfrak{L})$ and $\mathfrak{L}$ - either one will suffice. We state this precisely as a proposition.

Proposition 3.2. *Let $\mathfrak{L}$ be a set of L lines in the plane and fix some r. Let $i=1$ or 2. Suppose that for any $1 \le D \le L$, we can find D lines cutting the plane into $\sim D^2$ cells so that Equidistribution(i) holds. Then $|S_r(\mathfrak{L})| \lesssim L^2 r^{-3} + Lr^{-1}$.*

The proof of this result is just a calculation. When I first did this calculation, I thought I had understood the main idea of the proof of Szemerédi-Trotter. Getting the points or lines to evenly distribute among the cells seemed like a minor point to me. My wrong intuition went like this: if I just put down the dividing lines without thinking too much, then the points wouldn't have a reason to concentrate in any particular cells, so they would probably end up pretty evenly distributed. With a little more experience, I think that this intuition was totally wrong.

Here's an alternate perspective. If I choose D red lines, then I have $2D$ real parameters at my disposal. I would like each of D^2 cells to contain $\sim |S_r(\mathfrak{L})|/D^2$ points of $S_r(\mathfrak{L})$. I am trying to satisfy $\sim D^2$ conditions. In essence, I have $2D$ variables, and I am hoping to solve D^2 equations. Without other information, this is a plan that sounds unlikely to work.

Here's an example of a set of points which is impossible to equidistribute. Take any set of points lying on a closed convex curve in the plane. Each red line intersects the curve in at most 2 points. Therefore, D red lines cut the curve into $\le 2D$ pieces. It follow that most of the $\sim D^2$ cells contain no points of the set.

This divide-and-conquer plan actually does work, driven by one further crucial idea. The crucial idea is to choose the D red lines independently at random from among the lines of $\mathfrak{L}$. If we do this, the lines of $\mathfrak{L}$ interact with the red lines in a good way, and we get something close to Equidistribution 2. We briefly give intuition why this may work. Suppose that we first randomly pick $D/10$ red lines from the lines of $\mathfrak{L}$ and look at the resulting cells. If one of these cells contains $\ge 100LD^{-1}$ lines of $\mathfrak{L}$, then it is very likely that one of them will be chosen among the next $D/10$ red lines, and the cell will get cut into smaller pieces. Cells intersecting more than $100LD^{-1}$ lines have a brief half-life, and this suggests that at the end of the process almost all cells will intersect $\lesssim LD^{-1}$ lines of $\mathfrak{L}$. This gives (a bit of) the flavor of the random line argument. We have left out some important details. The cutting method involves some further care, and the random cutting needs to be refined a little. But choosing a random subset of D lines from $\mathfrak{L}$ is a crucial first step.

3.3. **Problems in higher dimensions.** Generalizations of the Szemerédi-Trotter theorem are a central subject of incidence geometry. One natural direction is to work in higher dimensions. Instead of lines in the plane, we can consider k-planes in $\mathbb{R}^n$. Some of the proofs of the Szemerédi-Trotter theorem are very planar, and it is difficult to generalize them to $\mathbb{R}^n$ for $n \ge 3$. For example, [Székely97] gives a beautiful proof of the theorem using crossing numbers of graphs. This proof generalizes to a huge variety of problems in the plane, but it seems very difficult to generalize it to higher dimensions. The cutting method was invented partly in order to attack higher-dimensional problems.

Let's summarize how to adapt the method to higher dimensions. The general divide-and-conquer strategy still makes sense. To divide $\mathbb{R}^n$ into cells, we need D red hyperplanes instead of D red lines. They divide $\mathbb{R}^n$ into $\sim D^n$ cells. If we have some kind of equidistribution, we still get interesting estimates. Moreover, if we are studying a set of (n-1)-dimensional planes in $\mathbb{R}^n$, then we can randomly choose D hyperplanes from our set, and we get some type of equidistribution. The objects don't necessarily have to be planes - we can also study codimension 1 spheres, paraboloids or other shapes.

But if we are studying k-planes in $\mathbb{R}^n$ for $k < n-1$, then there is a major difficulty: k-planes do not divide $\mathbb{R}^n$ into cells. If we try to choose (n-1)-planes so that the k-planes are equidistributed among the cells, we cannot use the key random trick above. We are stuck with $\sim D$ parameters hoping to satisfy $\sim D^n$ conditions. Moreover, there are examples of arrangments of k-planes in $\mathbb{R}^n$ where no arrangement of hyperplanes gives equidistribution. These examples generalize the set of points on a convex curve described above.

In summary, there is a major obstacle in dealing with objects of codimension > 1. The joints problem is one of the simplest incidence problems in codimension > 1. That's one reason the joints problem is interesting and important. Following the joints theorem, it looks reasonable to use the polynomial method to attack other incidence problems in codimension > 1. We will see a number of results in this direction.

Before the polynomial method, I only know of one sharp estimate about incidences in codimension > 1. This is Toth's complex generalization of the Szemerédi-Trotter theorem [Toth03]. If $\mathfrak{L}$ is a set of L complex lines in $\mathbb{C}^2$, Toth proved that $|S_r(\mathfrak{L})| \lesssim L^2 r^{-3} + L r^{-1}$ - the same estimate as for real lines in $\mathbb{R}^2$. From the point of view of topology, $\mathbb{C}^2$ is homeomorphic to $\mathbb{R}^4$ and the complex lines are homeomorphic to $\mathbb{R}^2$, and so in a topological sense the codimension is 2. Toth's proof is adapted from the first proof of Szemerédi and Trotter, and it is technically difficult.

In his work on the complex problem, Toth raised the following question. Suppose that $\mathfrak{L}$ is a set of k-planes in $\mathbb{R}^{2k}$, and that any two k-planes of $\mathfrak{L}$ intersect in ≤ 1 point. (In other words, we forbid two k-planes to contain a common line.) Is it still true that the number of r-rich points is $\lesssim L^2 r^{-3} + L r^{-1}$. This is a bold higher-dimensional generalization of the Szemerédi-Trotter theorem (and it also includes the complex version of the Szemerédi-Trotter theorem). Recently, Solymosi and Tao proved Toth's conjecture up to a factor of L^ϵ using the polynomial method.

Theorem 3.3. ([Solymosi-Tao12]) *If $\mathfrak{L}$ is a set of L k-planes in $\mathbb{R}^{2k}$, and if any two planes of $\mathfrak{L}$ intersect in ≤ 1 point, then for any $\epsilon > 0$, the number of r-rich points of $\mathfrak{L}$ is $\leq C(\epsilon) L^{\epsilon}(L^2 r^{-3} + L r^{-1})$.*

3.4. **Distance problems in the plane.** There are many deep open problems in incidence geometry even for curves in the plane. One example is the unit distance problem (which Erdős's posed in [Erdős46] alongside the distinct distance problem). It asks, given n points in the plane, how many pairs of points can have distance 1? In all known examples, the number of unit distances is $\lesssim n^{1+\epsilon}$. (In a square grid with a well-chosen spacing, the number of unit distances is slightly superlinear, but $\lesssim n^{1+\epsilon}$ for any $\epsilon > 0$.) The paper [SST] the best currently known bound: the number of unit distances is $\lesssim n^{4/3}$. This bound is closely connected with the Szemerédi-Trotter theorem. The unit distance problem is analogous to the Szemerédi-Trotter problem with unit circles in place of lines.

The reason for the difficulty seems to be the following. If we replace unit circles by "unit parabolas" (parabolas of the form $y = x^2 + ax + b$), then the bound $n^{4/3}$ is tight. To improve the $n^{4/3}$ bound, we need to find and use a property which is true for unit circles and false for unit parabolas. There's no clear candidate for this property or how to use it.

The distinct distance problem can also be phrased as a problem about circles in the plane, and it is difficult for similar reasons.

Elekes found a completely different way of thinking about the distinct distance problem, connecting it to problems in higher codimension like the ones we discussed in the last section.

3.5. **Partial symmetries.** Suppose G is a group acting on a space X. If $P \subset X$ is a finite set, then we can look at the symmetries of P under the group action. We define

$$G(P) := \{g \in G \text{ such that } g(P) = P\}.$$

Elekes started a study of partial symmetries. A partial symmetry of P is a group element that maps a large chunk of P to another large chunk of P. More precisely we define the r-rich partial symmetries by

$$G_r(P) := \{g \in G \text{ such that } |g(P) \cap P| \geq r\}.$$

It's interesting to try to understand the size and structure of $G_r(P)$ in different situations. Elekes realized that this natural problem is closely connected to the distinct distance problem and to the incidence geometry of curves in 3-dimensional space. In these connections, the group G is the group of orientation-preserving rigid motions of the plane.

Conjecture 3.4. ([Elekes-Sharir10]) *If P is a finite set in the plane, and $r \geq 2$, then*

$$|G_r(P)| \lesssim |P|^3 r^{-2}.$$

(If P is a square grid, then this bound is tight up to a constant factor for all $2 \leq r \leq |P|/10$.)

Elekes and Sharir proved this conjecture for $r = 3$ using the polynomial method. Nets Katz and I proved the conjecture in [Guth-Katz11].

This conjecture is closely related to the distinct distance problem. Elekes realized that if a set P has few distinct distances, then it must have lots of partial symmetries. We sketch the reason. Let $Q(P)$ be the set of distance quadruples, defined as follows.

$$Q(P) := \{(p_1, q_1, p_2, q_2) | dist(p_1, q_1) = dist(p_2, q_2)\}.$$

If there are few distinct distances, then it stands to reason that there will be many pairs of points at the same distance. By a Cauchy-Schwarz argument, one gets $|d(P)||Q(P)| \gtrsim |P|^4$, where $d(P)$ is the number of distinct distances of the set P. So if there are few distinct distances, then $|Q(P)|$ will be large.

On the other hand, each quadruple in $Q(P)$ suggests a partial symmetry of P. For each quadruple of $Q(P)$, there is a unique rigid motion $g \in G$ so that $g(p_1) = p_2$ and $g(q_1) = q_2$. The rigid motion takes two points of P to two other points of P, so it belongs to $G_2(P)$. In this way, we get a map $E : Q(P) \to G_2(P)$. We want to use this map to count $Q(P)$. If the map E were injective, we would have $|Q(P)| \leq |G_2(P)|$, which in turn is $\lesssim |P|^3$. The map E is actually not injective. If $|g(P) \cap P| = r$, then the preimage $E^{-1}(g)$ has size $\sim r^2$, because there are $\binom{r}{2}$ pairs of points in $g(P) \cap P$, and each pair yields a distance quadruple. Based on this observation, it's straightforward to relate $Q(P)$ and $G_r(P)$:

$$|Q(P)| \sim \sum_{r=2}^{|P|} r|G_r(P)|.$$

Plugging in the Elekes-Sharir conjecture gives $|Q(P)| \lesssim \sum_{r=2}^{|P|} |P|^3 r^{-1} \sim |P|^3 \log |P|$, and so $|d(P)| \gtrsim |P| / \log |P|$. So the Elekes-Sharir conjecture implies the new bound for the distinct distance problem.

The next observation of Elekes is that understanding the size of $|G_r(P)|$ is an incidence geometry problem where the background is the group G instead of Euclidean space. Instead of lines in $\mathbb{R}^3$, we consider the following special curves in G. For any two points $p_1, p_2 \in \mathbb{R}^2$, define

$$S_{p_1,p_2} := \{g \in G \text{ such that } g(p_1) = p_2\}.$$

These curves are natural objects from the point of view of the group structure of G. The curves S_{p_1,p_1} are 1-dimesional subgroups of G, and the curves S_{p_1,p_2} are their cosets.

For a finite set $P \subset \mathbb{R}^2$, let $S(P)$ denote the $|P|^2$ curves $\{S_{p_1,p_2}\}_{p_1,p_2 \in P}$. Next, we observe that a group element g is in $G_r(P)$ if and only if g lies in $\geq r$ of the curves of $S(P)$. This follows directly from the definition. If g is in $G_r(P)$, then it means that $g : P_1 \to P_2$ bijectively, where P_1 and P_2 are subsets of P with size r. For each point $p_1 \in P_1$, we have $g \in S_{p_1, g(p_1)}$, so g lies in r curves of $S(P)$. The converse direction is similar. So we can redefine $G_r(P)$ in the following way:

$$G_r(P) = \{g | g \text{ lies in } \geq r \text{ curves of } S(P)\}.$$

Understanding the size of $G_r(P)$ is closely analogous to understanding the number of r-rich points of a set of lines in $\mathbb{R}^3$. In particular, both problems involve objects of codimension 2, and they involve the difficulties discussed in Section 3.3.

In the future, mathematicians may consider the incidence theory of subgroups and cosets inside of a Lie group G by working intrinsically inside of G. For the time being, we are much more comfortable in Euclidean space, and we choose coordinates on G so that we get a problem about curves in Euclidean space. In the particular case of our curves $S(P)$ in our group G, there is a good choice of coordinates where the curves become straight lines in $\mathbb{R}^3$. In these coordinates, the Elekes-Sharir conjecture reduces to the following two theorems about straight lines in $\mathbb{R}^3$. The theorems were proven in [Guth-Katz11].

Theorem A. *Suppose that $\mathfrak{L}$ is a set of L lines in $\mathbb{R}^3$ with $\le L^{1/2}$ lines in any plane or regulus. Prove that the number of intersection points of lines of $\mathfrak{L}$ is $\lesssim L^{3/2}$.*

In particular, this theorem gives the best known estimate on the many intersection problem that we discussed in Section 1.

Theorem B. *Suppose that $\mathfrak{L}$ is a set of L lines in $\mathbb{R}^3$ with $\le L^{1/2}$ lines in any plane. For $3 \le r \le L^{1/2}$, prove that the number of r-rich intersection points of $\mathfrak{L}$ is $\lesssim L^{3/2} r^{-2}$.*

I like to think of this theorem as a generalization of the Szemerédi-Trotter theorem to lines in $\mathbb{R}^3$. There are probably many generalizations of that theorem to higher dimensions. Toth's conjecture is one generalization, and Theorem B is another generalization with a different flavor.

3.6. **Conclusion.** Studying incidence geometry problems in codimenson > 1 presents particular challenges. The polynomial method is the most effective tool currently available for studying these problems. The simplest case is the case of lines in $\mathbb{R}^3$. For lines in $\mathbb{R}^3$, the joints theorem and theorems A and B give a good picture of what we now understand. The many intersections problem is a good example of what we still don't understand. In higher dimensions, the Solymosi-Tao result on the Toth conjecture is the main example of what we now know. This result is remarkable (partly) because it works with arbitrarily high dimensions and arbitrarily high codimensions.

Several problems can be transformed into incidence geometry problems in higher codimension. We have seen that the distinct distance problem in the plane and the partial symmetries of plane sets are both related to the incidence structure of lines in $\mathbb{R}^3$.

In the next section we will describe how to attack these problems using high-degree polynomials, extending the ideas from the proofs of finite field Nikodym and joints.

4. Combinatorial structure and algebraic structure

In this section, we will discuss the proofs of Theorems A and B. The proofs are based on the polynomial method, and the key point is the connection between combinatorial structure and algebraic structure.

We saw earlier that lines in $\mathbb{R}^3$ may have many intersection points by clustering into either a plane or a degree 2 surface. Theorem A is a (partial) converse to this observation:

more than $L^{\frac{3}{2}+\epsilon}$ intersection points may be formed *only if* the lines cluster into a plane or a degree 2 surface. Theorem A says that a certain combinatorial structure forces a certain algebraic structure. Our goal in this section is to explore how combinatorial structure can force algebraic structure.

We will see two different mechanisms how combinatorial structure can force algebraic structure. We begin by considering what we mean by algebraic structure.

4.1. **Algebraic structure for finite sets.** If $X \subset \mathbb{R}^n$, let $deg(X)$ be the smallest degree of a non-zero polynomial that vanishes on X. We have seen that for a finite set X, $deg(X) \lesssim |X|^{1/n}$. Of course particular finite sets can have much lower degree. For instance, any subset of a plane has degree 1. For generic sets, the $|X|^{1/n}$ bound is sharp. So a generic finite set has $deg(X) \sim |X|^{1/n}$. Any set with degree significantly smaller than $|X|^{1/n}$ has non-trivial algebraic structure.

There is a similar discussion for finite unions of lines. If X is a union of L lines in $\mathbb{R}^n$, then $deg(X) \lesssim L^{\frac{1}{n-1}}$. The proof is straightforward, so we sketch it here. Suppose that D is a degree so that $(D+1)L < \dim Poly_D(\mathbb{R}^n)$. Then we can choose a non-zero polynomial of degree $\leq D$ that vanishes at $D+1$ points on each of the L lines. By the vanishing lemma, this polynomial vanishes on each line. A short calculation shows that we can choose $D \lesssim L^{\frac{1}{n-1}}$. In summary, any union of L lines has degree $\lesssim L^{\frac{1}{n-1}}$. If a union of L lines has degree significantly smaller than this, then it has some non-trivial algebraic structure.

With this definition of algebraic structure, we can begin to explore how combinatorial structure forces algebraic structure.

Proposition 4.1. *(Degree reduction) Suppose that $\mathfrak{L}$ is a set of L lines in $\mathbb{R}^3$. Suppose that each line contains $\geq A$ (distinct) intersection points with other lines of $\mathfrak{L}$. Then the degree of the union of the lines is $\leq 10^5 L/A$.*

(If $A \leq L^{1/2}$, then the conclusion of the proposition is worthless, because every set of L lines has degree $\leq 4L^{1/2}$ anyway. But if A is much bigger than $L^{1/2}$, then the lines have non-trivial algebraic structure.)

Here is the idea of the proof. We saw above that for any L' lines of $\mathfrak{L}$, there is a non-zero polynomial which vanishes on those lines with degree $\leq 10(L')^{1/2}$. We let $\mathfrak{L}' \subset \mathfrak{L}$ be a subset of L' random lines of $\mathfrak{L}$, and we consider the polynomial P that vanishes on them. If A and L' are large enough, then this polynomial has to vanish on many other lines. Let l be another line of $\mathfrak{L}$. If l intersects the lines of $\mathfrak{L}'$ at $> deg(P)$ points, then P will vanish on l also. The expected number of intersection points between l and the lines of $\mathfrak{L}'$ is $A(L'/L)$. Whenever $A(L'/L) > 100(L')^{1/2}$, the expected number of intersection points is $> 10deg(P)$. In this situation, the polynomial P will vanish on the vast majority of the lines of $\mathfrak{L}$. Choosing L' optimally, we get a polynomial of degree $\leq 10^5 L/A$ that vanishes on most of the lines of $\mathfrak{L}$. (And with a little extra technique, we can get a polynomial that vanishes on all of the lines of $\mathfrak{L}$.)

I think this proposition is fundamental to the polynomial method. It shows that a set of lines with a lot of intersections must have an algebraic structure. This algebraic structure is an important clue to try to understand such sets of lines. Once we know that the set of lines has a non-trivial algebraic structure, it's natural to try to use algebra and algebraic geometry to understand the set better.

4.2. **Ruled surfaces.** The proof of Theorem A is based on the theory of ruled surfaces. An algebraic surface $Z(P) \subset \mathbb{R}^3$ is called ruled if each point of $Z(P)$ lies in a line in $Z(P)$. If is called doubly ruled if each point of $Z(P)$ lies in two different lines in $Z(P)$. There is a classification of doubly ruled surfaces, and in particular the following result is relevant for us.

Proposition 4.2. *A doubly ruled algebraic surface $Z(P) \subset \mathbb{R}^3$ is a union of planes and degree 2 surfaces.*

Theorem A is a discrete analogue of this proposition from algebraic geometry. To try to make the analogy as close as possible, we state a small variation of Theorem A.

Theorem A'. *Suppose that $\mathfrak{L}$ is a set of L lines in $\mathbb{R}^3$, and that each line contains $\geq 10^{10}L^{1/2}$ intersection points with other lines of $\mathfrak{L}$. Then the lines of $\mathfrak{L}$ are contained in a union of $10^{-5}L^{1/2}$ planes and degree 2 surfaces.*

In this analogy, the set of intersection points of the lines of $\mathfrak{L}$ is a 'discrete approximation of a surface'. Each of these points lies in two lines of $\mathfrak{L}$, and each line of $\mathfrak{L}$ contains many points of our 'discrete surface'. The hypothesis is that we have a kind of 'discrete doubly ruled surface', and the conclusion is that $\mathfrak{L}$ is contained in a union of planes and degree 2 surfaces.

The degree reduction argument is a first step to prove Theorem A'. It tells us that the lines of $\mathfrak{L}$ are contained in the zero set of a polynomial P of degree $\leq 10^{-5}L^{1/2}$. This is the right bound for the degree, but we still have to prove that the polynomial factors into polynomials of degree 1 and 2. We have to understand better the structure of the polynomial P. We will see that the combinatorial structure of the lines of $\mathfrak{L}$ is connected with the geometric structure of $Z(P)$. We explain the connection in the next subsection.

4.3. **Contagious structures.** Suppose that l is a line in $Z(P)$. If there are $> deg(P)$ critical points of P on l, then every point of l is critical. The property of being critical is 'contagious'.

Let's give another example of a contagious property. Suppose now that $l \subset Z(P)$ and that each point of l is non-critical. A regular point x in $Z(P)$ is called flat if the curvature of $Z(P)$ vanishes at x - equivalently if there is a plane thru x which is tangent to $Z(P)$ to second order. If the line l contains more than $3deg(P)$ flat points, then every point on the line is flat. So being flat is also a contagious property.

These properties are contagious because they are described by (other) polynomials. A point is critical if and only if $\partial_1 P, \partial_2 P$, and $\partial_3 P$ all vanish. These partial derivatives have degree $\leq deg(P) - 1$. It follows by the vanishing lemma that being critical is contagious. With a little more work, being flat is also described by polynomials. For any polynomial P, there exists a finite list of polynomials SP with degree $\leq 3deg(P)$, and a (regular) point

$x \in Z(P)$ is flat if and only if $SP(x) = 0$. It doesn't take that much work to construct SP, and then we see that being flat is contagious too.

To see how to use contagious properties, we will begin by discussing triple intersection points, because the method is a little easier. Suppose that $\mathfrak{L}$ is a set of L lines in $\mathbb{R}^3$ and each line contains $\geq 10^{10} L^{1/2}$ triple intersection points. By degree reduction, these lines lie in $Z(P)$ for a polynomial P of degree $\leq 10^{-5} L^{1/2}$. Since the number of triple points on each line is much more than the degree, any contagious property of the triple points will spread to all of the lines.

Triple intersection points indeed have interesting properties. If x lies in 3 lines in $Z(P)$ and the lines are not coplanar, then x is a critical point of P, as we saw in the proof of the joints theorem. On the other hand, if x is not a critical point and x lies in three lines of $Z(P)$, then x is a flat point. The three lines must lie in the tangent plane of $Z(P)$, and then the tangent plane hugs $Z(P)$ along three lines, which forces it to be tangent to $Z(P)$ to second order. Anyway, every triple intersection point is either critical or flat. Since these properties are contagious, every point in the union of the lines of $\mathfrak{L}$ must be either critical or flat.

Contagious properties don't just spread from points to lines. If there are many lines with a contagious property, then it can spread to a whole surface. This follows from the following version of Bezout's theorem.

Theorem 4.3. *If P and Q are polynomials in three variables, and if they have no common factor, then $Z(P) \cap Z(Q)$ contains at most $deg(P) \cdot deg(Q)$ lines. In particular, if P is irreducible and $Z(P) \cap Z(Q)$ contains $> deg(P) \cdot deg(Q)$ lines, then Q vanishes on $Z(P)$.*

Remember that our L lines lie in $Z(P)$ where $deg(P) \leq 10^{-5} L^{1/2}$. Each of the lines is either critical or flat. Suppose for a moment that they are all flat. (The critical case is similar.) For the sake of exposition, let's also assume that P is irreducible. The number of flat lines is $L \geq 10^{10} (deg(P))^2$. Each polynomial of SP vanishes on these lines. The degree of each polynomial of SP is $\leq 3deg(P)$. By the Bezout theorem, SP vanishes on $Z(P)$. This means that every point of $Z(P)$ is flat. Then it follows that $Z(P)$ is a plane.

In general, the polynomial P may be reducible and there may be several components, but a similar argument shows that they are all planes. We have sketched the proof of the following result, which essentially appears in [EKS11].

Theorem 4.4. *(*[EKS11]*) If $\mathfrak{L}$ is a set of L lines in $\mathbb{R}^3$, and each line contains $\geq 10^{10} L^{1/2}$ triple intersection points, then the union of the lines is contained in $\leq 10^{-5} L^{1/2}$ planes.*

This theorem is the case $r = 3$ of Theorem B.

It is harder to understand intersection points than triple interesection points. The problem is that if x lies in two lines in $Z(P)$, then it doesn't imply that x is either critical or flat. It's not clear right away if there is another contagious property that we can use instead.

To approach this question, let's step back and try to understand where contagious properties come from. We can build contagious properties by looking at polynomials in P and the derivatives of P. If RP is a polynomial of degree $\leq C$ in P and its derivatives, then

$RP(x)$ is a polynomial in x of degree $\leq Cdeg(P)$. We can use any such RP in place of SP in the argument above. Algebraic geometry helps understand what geometric properties of $Z(P)$ at a point x can be described by some polynomial equations in P and its derivatives. In short there are a lot of contagious properties.

We give one more example. A point $x \in Z(P)$ is called flecnodal if and only if there is a non-zero vector v so that P vanishes in the direction v to third order at x. It's not immediately obvious that being flecnodal is contagious, but it is. There is a polynomial FP, called the flecnode polynomial, of degree at most $11 DegP$, and a point $x \in Z(P)$ is flecnodal if and only if $FP(x) = 0$. This polynomial and this result were discoved by Salmon in the 1800's.

Stepping back from the details, we can describe the moral of the proof of Theorem A'. If x lies in two lines in $Z(P)$, it leads to some equations about P and the derivatives of P at x. These equations are all contagious, and so they end up holding at every point of $Z(P)$. So all the points of $Z(P)$ have a lot in common with the intersection points. After working out the details, it follows that *every point of $Z(P)$ lies in two lines in $Z(P)$*. The surface $Z(P)$ is doubly ruled. By the classification of doubly ruled surfaces, $Z(P)$ is a union of planes and degree 2 surfaces. We also know that the degree of P is $\leq 10^{-5} L^{1/2}$. Hence all the lines of $\mathfrak{L}$ lie in $\leq 10^{-5} L^{1/2}$ planes and degree 2 surfaces.

4.4. **Polynomial cell decompositions.** Theorem B involves a combination of all of the difficulties we have encountered in this essay so far. It is a problem about lines in $\mathbb{R}^3$, so the codimension is > 1. This suggests that the proof needs to use high degree polynomials. We saw in the last section how to prove the case $r = 3$ with the polynomial method, and I don't have any idea how to approach the problem without it. But for large r, Theorem B is false over finite fields like the Szemerédi-Trotter theorem. This suggests that the proof needs to use the topology of $\mathbb{R}^3$.

The proof of Theorem A does not generalize to Theorem B. It breaks down in the very first step: the degree reduction argument does not work.

The proof of Theorem B involves a combination of (almost) all of the methods that we've discussed in this essay. The key step is to build cell decompositions using polynomial surfaces, combining the cutting method and the polynomial method.

Instead of cutting space with D hyperplanes, we cut space with a degree D polynomial. A degree D polynomial surface has many good features in common with a union of D hyperplanes. The complement of D hyperplanes consists of $\sim D^n$ components. The complement of a degree D polynomial surface consists of $\lesssim D^n$ components, and there are $\sim D^n$ components in many examples. We will call these componenents cells. In each case, a line can only enter at most $D + 1$ cells.

The union of D hyperplanes is a special case of a degree D polynomial surface, but there are many more polynomial surfaces. Using polynomial surfaces gives us much more flexibility, and we have a better chance to prove equidistribution. Recall that we would like some equidistribution among $\sim D^n$ cells, which means we are trying to achieve $\sim D^n$ conditions. Choosing D hyperplanes gives us $\sim D$ degrees of freedom. But choosing a degree D polynomial surface gives us $\sim D^n$ degrees of freedom. Having so much more freedom, it looks more realistic to get equidistribution. Here is a precise result about building cell decompositions with polynomial surfaces.

Lemma 4.5. *(Polynomial cell decomposition lemma) If $\mathfrak{S}$ is any finite set in $\mathbb{R}^n$, and if $D \ge 1$ is any integer, then there is a non-zero polynomial $P \in Poly_D(\mathbb{R}^n)$ so that each component of the complement of $Z(P)$ contains $\le C(n)|\mathfrak{S}|D^{-n}$ points of $\mathfrak{S}$.*

We should make an important caveat right away. The lemma does not say that all the points of $\mathfrak{S}$ are in the complement of $Z(P)$. Some or even all the points of $\mathfrak{S}$ could lie in $Z(P)$.

The proof of the cell decomposition lemma is based on the Stone-Tukey ham sandwich theorem, which we discussed in the Section 2.3. The ham sandwich theorem allows us to cut a bunch of sets in half. By using it repeatedly, we can cut our set of points into halves, then quarters, then eighths... Here is a detailed sketch.

1. The ham sandwich theorem says that given N finite volume open sets, we can choose a polynomial of degree $\lesssim N^{1/n}$ that bisects all of them.

We are dealing with finite sets, which have volume zero. Suppose that we have N finite sets $S_1, ..., S_N$. We let U_j be the ϵ-neighborhood of S_j. We apply the theorem to U_j and take the limit as ϵ goes to zero. In this way we get the following more combinatorial result.

2. If $S_1, ..., S_N$ are finite sets, then there is a polynomial P of degree $\lesssim N^{1/n}$ so that $P > 0$ on at most half the points of S_j and $P < 0$ on at most half the points of S_j. (Remark: P might vanish on some or even all of the points of S_j.)

3. We have a set S that we want to divide into 2^J fairly even pieces. Pick a plane that bisects S. Then pick a surface that bisects each half, leaving us with four sets of cardinality at most $|\mathfrak{S}|/4$. Next pick a surface that bisects each of these four sets. Continuing in this way, we have cut $\mathfrak{S}$ into 2^J pieces of cardinality at most $|\mathfrak{S}|2^{-J}$ by a union of J algebraic hypersurfaces. The degrees of these hypersurfaces are bounded by step 2, and adding up we get a total degree $\lesssim 2^{J/n}$ as desired.

Next we discuss how to use the polynomial cell decomposition lemma. We consider an arrangement of lines $\mathfrak{L}$, and we let $\mathfrak{S}$ be the set of r-rich points. We build a polynomial cell decomposition. If all the points of $\mathfrak{S}$ lie in the cells, then we can proceed by a divide-and-conquer argument as in the cutting method. We know that each cell has the same number of points of $\mathfrak{S}$, and we know the number of lines that enter an average cell. In each cell, we can use a more elementary method to count r-rich points. Adding up the contributions from all of the cells, we see that the number of r-rich points is $\lesssim L^{3/2}r^{-2}$ - the conclusion of Theorem B.

This is not a complete proof of Theorem B. It may happen that most or all of the points of $\mathfrak{S}$ lie in $Z(P)$, and then the argument breaks down. Here is a slightly more optimistic way of looking at the situation.

The polynomial cell decomposition argument gives a second, completely different mechanism by which combinatorial structure forces algebraic structure. If $\mathfrak{L}$ is a set of L lines with significantly more than $L^{3/2}r^{-2}$ r-rich points, then the argument above shows that almost all of the r-rich points lie in $Z(P)$ for a polynomial P of surprisingly low degree. Since there are many r-rich points on each line, it follows that the lines lie in $Z(P)$ also, and the conclusion is that the degree of $\mathfrak{L}$ is far below $L^{1/2}$. The combinatorial structure of having many r-rich points forces algebraic structure.

Once the set of lines has algebraic structure, the rest of the proof of Theorem B is similar to the proof of Theorem A, using contagious properties.

The polynomial cell decomposition has had several other applications. The paper [Solymosi-Tao12] uses it to prove the higher-dimensional generalization of the Szemerédi-Trotter theorem. The paper [KMS12] uses it to give new proofs and perspectives on several fundamental theorems of incidence geometry.

4.5. **Final summary.** The proofs we have been studying get off the ground by proving that arrangements with a lot of combinatorial structure must have unexpectedly low degree. We have seen two mechanisms to find these unexpectedly low degree polynomials. One mechanism is the degree reduction lemma. This lemma is proven by combining the parameter counting argument and the vanishing lemma. It's based on the proof of the finite field Nikodym conjecture and recovery algorithms for error-correcting codes. The second mechanism is the polynomial cell decomposition method. This mechanism is based on the polynomial method, but also on the cutting method and surface area estimates from differential geometry.

Once we know that the arrangement we are studying lies in the zero set of a polynomial of unexpectedly low degree, then it's natural to try to use that polynomial to study the set. The contagious structures are one tool to do that.

References

[BW86] E. Berlekamp and L. Welch, Error correction of algebraic block codes. US Patent Number 4,633,470. 1986.

[CEGPSSS92] B. Chazelle, H. Edelsbrunner, L. Guibas,R. Pollack, R.Seidel, M. Sharir, and J. Snoeyink, Counting and cutting cycles of lines and rods in space, Computational Geometry: Theory and Applications, 1(6) 305-323 (1992).

[CEGSW90] K.L. Clarkson, H. Edelsbrunner, L. Guibas, M Sharir, and E. Welzl, Combinatorial Complexity bounds for arrangements of curves and spheres, Discrete Comput. Geom. (1990) 5, 99-160.

[Dvir09] Z. Dvir, On the size of Kakeya sets in finite fields, J. Amer. Math Soc. (2009) 22, 1093-1097.

[Erdős46] P. Erdős, On sets of distances of n points, Amer. Math. Monthly (1946) 53, 248-250.

[Erdős] P. Erdős, Some of my favorite problems and results, in *The Mathematics of Paul Erdős*, Springer, 1996.

[EKS11] Gy. Elekes, H. Kaplan, and M. Sharir, On lines, joints, and incidences in three dimensions, Journal of Combinatorial Theory, Series A (2011) 118, 962-977.

[Elekes-Sharir10] Gy. Elekes and M. Sharir, Incidences in three dimensions and distinct distances in the plane, Proceedings 26th ACM Symposium on Computational Geometry (2010) 413-422.

[Federer69] H. Federer, *Geometric measure theory.* Die Grundlehren der mathematischen Wissenschaften, Band 153 Springer-Verlag New York Inc., New York 1969.

[Feldman-SharirS05] S. Feldman and M. Sharir, *An improved bound for joints in arrangements of lines in space*, Discrete Comput. Geom. (2005) 33, 307-320.

[Gromov03] M. Gromov, Isoperimetry of waists and concentration of maps. Geom. Funct. Anal. 13 (2003), no. 1, 178-215.

[Guth-Katz10] L. Guth and N. Katz, Algebraic methods in discrete analogs of the Kakeya problem. Adv. Math. 225 (2010), no. 5, 2828-2839.

[Guth-Katz11] L. Guth and N. Katz, On the Erdős distinct distance problem in the plane, arXiv:1011.4105.

[Guth09] Minimax problems related to cup powers and Steenrod squares. Geom. Funct. Anal. 18 (2009), no. 6, 1917-1987.

[KMS12] H. Kaplan, J. Matoušek, and M. Sharir, Simple proofs of classical theorems in discrete geometry via the Guth–Katz polynomial partitioning technique. Discrete Comput. Geom. 48 (2012), no. 3, 499-517.

[KSS10] H. Kaplan, M. Sharir, and E. Shustin, On lines and joints, Discrete Comput Geom (2010) 44, 838-843.

[Laba08] I. Laba, From harmonic analysis to arithmetic combinatorics. Bull. Amer. Math. Soc. (N.S.) 45 (2008), no. 1, 77-115.

[Quilodrán10] R. Quilodrán, The joints problem in $\mathbf{R}^\mathbf{n}$, Siam J. Discrete Math, Vol. 23, 4, p. 2211-2213.

[Schmidt74] Schmidt, Wolfgang M. Applications of Thue's method in various branches of number theory. Proceedings of the International Congress of Mathematicians (Vancouver, B.C., 1974), Vol. 1, pp. 177-185. Canad. Math. Congress, Montreal, Que., 1975.

[Solymosi-Tao12] J. Solymosi and T. Tao, An incidence theorem in higher dimensions. Discrete Comput. Geom. 48 (2012), no. 2, 255-280

[SST] J. Spencer, E. Szemerdi, and W. Trotter, Unit distances in the Euclidean plane. Graph theory and combinatorics (Cambridge, 1983), 293-303, Academic Press, London, 1984.

[Sudan95] M. Sudan, Efficient checking of polynomials and proofs and the hardness of approximation problems, ACM Distinguished Thesees, Springer 1995.

[Székely97] L. Székely, Crossing numbers and hard Erdős problems in discrete geometry. Combin. Probab. Comput. 6 (1997), no. 3, 353-358.

[ST83] E. Szemerédi and W. T. Trotter Jr., Extremal Problems in Discrete Geometry, Combinatorica (1983) 3, 381-392.

[Tao01] T. Tao, From rotating needles to stability of waves: emerging connections between combinatorics, analysis, and PDE. Notices Amer. Math. Soc. 48 (2001), no. 3, 294-303.

[Toth03] C. Toth, The Szemerédi-Trotter theorem in the complex plane. aXiv:math/0305283, 2003.

[Trevisan04] L. Trevisan, Some applications of coding theory in computational complexity. Complexity of computations and proofs, 347-424, Quad. Mat., 13, Dept. Math., Seconda Univ. Napoli, Caserta, 2004.

[Wolff99] T. Wolff. Recent work connected with the Kakeya problem. Prospects in mathematics (Princeton, NJ, 1996). pages 129-162, 1999.