CHAPTER 1

Introduction

This book is about some applications of polynomials to problems in combinatorics. What I think is interesting about these arguments is that the statements of the problems do not involve polynomials, but polynomials provide a crucial structure under the surface. The starting point of the book is Dvir's solution of the finite field Kakeya problem [**D**]. This is a problem on the border between combinatorics and harmonic analysis. People in the field had believed that it was a very hard problem, but the proof is only a few pages long, and it only requires an undergraduate background to understand.

Here is the statement of the finite field Kakeya problem. Let $\mathbb{F}_q$ denote the finite field with q elements. A set $K \subset \mathbb{F}_q^n$ is called a Kakeya set if it contains a line in every direction. (In other words, K is a Kakeya set if it contains a translate of every 1-dimensional subspace of $\mathbb{F}_q^n$.) The question is, what is the smallest possible cardinality of a Kakeya set $K \subset \mathbb{F}_q^n$?

THEOREM 1.1. ([**D**]) If $K \subset \mathbb{F}_q^n$ is a Kakeya set, then

$$|K| \geq (10n)^{-n} q^n.$$

For a fixed dimension n, this estimate says that the size of a Kakeya set is at least a constant factor times the size of the whole vector space $\mathbb{F}_q^n$.

The proof begins by considering a lowest degree (non-zero) polynomial that vanishes on the set K. Using this polynomial, the proof is short and clean. But without mentioning this polynomial, proving Theorem 1.1 seems to be very difficult, and people worked hard to prove much weaker estimates. Over the course of the book, we will explore different variations of this trick with polynomials, and we will discuss applications of this method to different problems.

A large piece of the book is about the distinct distance problem in the plane, a combinatorics problem raised by Erdős in the 1940s [**Er1**]. The problem asks, given a set of N points in the plane, what is the minimum possible number of distinct distances between the points. For example, if the points are evenly spaced along a line, then there are $N - 1$ distinct distances. Erdős checked that arranging the points in a square grid is slightly more efficient, giving on the order of $N(\log N)^{-1/2}$ distinct distances. He conjectured that grids are nearly optimal. We will prove the following lower bound, which nearly matches the grid example:

THEOREM 1.2. (Guth and Katz, [**GK2**], building on [**ElSh**]) If P is a set of N points in $\mathbb{R}^2$, then the number of distinct distances between the points of P is at least $cN(\log N)^{-1}$.

The main applications in the book are to problems in combinatorics. But it is also striking to me that this trick with polynomials has connections to several

other areas of mathematics. We will see related arguments connected with error-correcting codes in computer science, inequalities about surface area in differential geometry, diophantine equations in number theory, and geometric estimates related to Fourier analysis. Each of these arguments has some significant ingredients in common with the proof of the finite field Kakeya conjecture. Also, each of these fields offers its own perspective about why polynomials are special functions and what makes them useful in these applications.

I tried to make the book self-contained, and I hope that it will be accessible to readers with a first-year graduate background or a strong undergraduate background.

In the rest of the introduction, we give an overview of the book. Some readers might want to begin by reading the overview. Other readers might want to begin by reading the proof of finite field Kakeya in Chapter 2.

1.1. Incidence geometry

When the distinct distance problem was first raised, in [**Er1**] in the 1940's, it didn't fit into any well-developed field of mathematics. There were a small number of isolated problems of this flavor that different people had raised in different situations. Some of these problems - including the distinct distance problem - turned out to be surprisingly hard. Over the next few decades, people began to study this circle of problems in a systematic way, and they developed a field of combinatorics called incidence geometry. Broadly speaking, incidence geometry is the study of combinatorial problems about basic geometric objects, like lines, circles, angles, or distances. To give a little sense of this area, let us describe one of the important theorems and some open problems.

One fundamental question in the field has to do with the possible intersection patterns of lines in the plane. If $\mathfrak{L}$ is a set of lines, a point x is called an r-rich point of $\mathfrak{L}$ if x lies in at least r lines of $\mathfrak{L}$. The set of r-rich points of $\mathfrak{L}$ is denoted $P_r(\mathfrak{L})$. Given a certain number of lines, how many r-rich points can they form? In the early 1980s, Szemerédi and Trotter solved this problem up to a constant factor.

THEOREM 1.3. ([**SzTr**], 1983) There are constants $0 < c < C$ so that the following holds. If $2 \leq r \leq L^{1/2}$, then

$$cL^2 r^{-3} \leq \max_{|\mathfrak{L}|=L} |P_r(\mathfrak{L})| \leq CL^2 r^{-3}.$$

If $L^{1/2} \leq r \leq L$, then

$$cLr^{-1} \leq \max_{|\mathfrak{L}|=L} |P_r(\mathfrak{L})| \leq CLr^{-1}.$$

The lower bound comes from a fairly simple example involving a grid of points. The difficult part is the upper bound. A remarkable thing about this proof is that it is based on topology. The topological approach was developed further by other mathematicians in the field, in papers such as [**CEGSW**] and [**Sz**], leading to a range of tools that apply to many problems. Developing topological methods to prove combinatorial estimates of this kind is one of the main achievements of incidence geometry.

Incidence geometry also has many simply stated open problems. For instance, in Theorem 1.3, if we replace lines by circles, we get a difficult open problem. Replacing lines by unit circles gives a different difficult open problem. Replacing

lines by ellipses or parabolas gives two more difficult open problems. These problems have been studied intensively for decades. The topological methods discussed above give interesting bounds for these problems, but the best current bounds don't match any known examples, and most people believe the bounds are not sharp. The distinct distance problem was also studied by these topological methods. People proved interesting bounds, but the method runs into similar issues as it does in problems about circles.

In the last decade, polynomial methods have developed into a second major approach to incidence geometry. Here is an example of an incidence geometry problem that seemed out of reach with topological methods but which has a short proof using polynomials. The joints problem is a problem about lines in $\mathbb{R}^3$, which was raised in the early 90s by [**CEGPSSS**]. If $\mathfrak{L}$ is a set of lines in $\mathbb{R}^3$, then a point x is a joint of $\mathfrak{L}$ if x lies in three non-coplanar lines of $\mathfrak{L}$. It is not hard to find examples with L lines and on the order of $L^{3/2}$ joints, and [**CEGPSSS**] conjectured that the number of joints is always at most $CL^{3/2}$. Before the polynomial method, the best known bound was $L^{1.62\ldots}$ ([**FS**]) and the argument was fairly complex.

THEOREM 1.4. ([**GK1**], proof simplified by [**KSS**], [**Q**]) A set of L lines in $\mathbb{R}^3$ forms at most $CL^{3/2}$ joints.

We will prove this result in Chapter 2, right after the proof of the finite field Kakeya conjecture.

In [**ElSh**], Elekes and Sharir proposed a new approach to the distinct distance problem, which connects it to the incidence geometry of lines in $\mathbb{R}^3$. This approach led to new questions about lines in $\mathbb{R}^3$, which I think are natural questions in their own right. These questions were resolved by Nets Katz and the author in [**GK2**]. The proofs use polynomial methods, and they also bring into play the topological methods described above and more tools from algebraic geometry. Explaining these results and their applications is one of the main goals of the book.

Suppose that $\mathfrak{L}$ is a set of L lines in $\mathbb{R}^3$. Let us first consider the 2-rich points of $\mathfrak{L}$. Since any two lines intersect in at most one point, the number of 2-rich points is at most $\binom{L}{2}$, and this can actually happen if all the lines lie in a plane. But the scenario that all lines lie in a plane is a special situation. If we rule out this situation, can we get a better bound? For instance, in the approach to the distinct distance problem from [**ElSh**], one is led to a set $\mathfrak{L}$ of L lines in $\mathbb{R}^3$ with at most $L^{1/2}$ lines in any plane. For such a set, can we prove a significantly stronger bound for $|P_2(\mathfrak{L})|$?

Suprisingly the answer is no. The counterexample comes from a degree 2 algebraic surface, such as the surface defined by $z = xy$. This surface is doubly ruled – every point in the surface lies in two lines in the surface. Choosing L lines contained in this degree 2 surface, we get a set $\mathfrak{L}$ with $L^2/4$ 2-rich points, while any plane contains at most 2 lines of $\mathfrak{L}$. This doubly ruled surface has been known in algebraic geometry for a long time, and it plays an important role in the first paper on the joints problem [**CEGPSSS**]. This example, involving a polynomial surface, helps to explain why polynomials play an important role in studying the intersection patterns of lines in space.

What if we assume that $\mathfrak{L}$ contains at most $L^{1/2}$ lines in any plane or any degree 2 algebraic surface? With these stronger hypotheses, can we prove a significantly stronger bound on $|P_2(\mathfrak{L})|$? This time, the answer is yes. The methods from [**CEGPSSS**] give a significant improvement, and [**GK2**] gives a sharp estimate.

THEOREM 1.5. If $\mathfrak{L}$ is a set of L lines in $\mathbb{R}^3$, and at most $L^{1/2}$ lines of $\mathfrak{L}$ lie in any plane or any degree 2 algebraic surface, then

$$|P_2(\mathfrak{L})| \leq CL^{3/2}.$$

The proof of Theorem 1.5 uses polynomial methods, and it also draws on the theory of ruled surfaces from algebraic geometry. (An algebraic surface is called ruled if it contains a line through every point.)

What about r-rich points for $r > 2$? If all the lines of $\mathfrak{L}$ lie in a plane, then the Szemerédi-Trotter theorem gives a sharp upper bound for $|P_r(\mathfrak{L})|$. We focus on the range $3 \leq r \leq L^{1/2}$, which is more challenging and interesting. In this range, Theorem 1.3 gives

$$|P_r(\mathfrak{L})| \leq CL^2 r^{-3}.$$

It's not hard to extend this bound to any set of L lines in $\mathbb{R}^3$. But suppose that $\mathfrak{L}$ contains at most $L^{1/2}$ lines in any plane. Can we prove a significantly better upper bound? The answer is yes, and the following sharp upper bound was proven in [**GK2**].

THEOREM 1.6. ([**GK2**]) If $\mathfrak{L}$ is a set of L lines in $\mathbb{R}^3$, and at most $L^{1/2}$ lines of $\mathfrak{L}$ lie in any plane, and if $3 \leq r \leq L^{1/2}$ then

$$|P_r(\mathfrak{L})| \leq CL^{3/2} r^{-2}.$$

The proof of Theorem 1.6 combines polynomial methods with topological methods that come from the proof of Theorem 1.3.

Theorems 1.5 and 1.6 give a lot of understanding of the incidence geometry of lines in $\mathbb{R}^3$. The distinct distance estimate, Theorem 1.2, follows by combining them with the framework from [**ElSh**].

1.2. Connections with other areas

The proofs of these combinatorial results have some similarities to proofs in other fields, and we will discuss a number of these connections.

One connection involves error-correcting codes in computer science. Dvir's background is in computer science. His interests include error-correcting codes, and perspectives from coding theory helped lead to the proof of finite field Kakeya. Here is a typical problem from error-correcting codes. Suppose that $\mathbb{F}_q$ is the finite field with q elements and that $Q : \mathbb{F}_q \to \mathbb{F}_q$ is a polynomial whose degree is not too high. Suppose that we have access to a corrupted version of Q. More precisely, suppose that $F : \mathbb{F}_q \to \mathbb{F}_q$ is a function which is known to agree with Q for a certain fraction of points $x \in \mathbb{F}_q$. By looking at F, we would like to recover the original polynomial Q, and we would like to do so efficiently. Berlekamp and Welch [**BW**] discovered an interesting trick for recovering the original polynomial, and this trick has common ingredients with the proof of finite field Kakeya.

It turns out that even if F contains quite a lot of corruption, it is still possible to efficiently recover the polynomial Q. In the field of error-correcting codes, polynomials are known for their resiliency - a polynomial code can tolerate a high level of error, and the original information can still be recovered. Polynomials are important in error-correcting codes because they are an especially resilient class of functions in this sense.

In differential geometry, polynomials are known for their efficiency. There are many examples of this efficiency. To mention one classical example, the zero set

of a complex polynomial is an area-minimizing surface – a surface with the least possible area given its boundary. Most of these results about efficiency involve very different ideas from the ones in this book, but there is one recent example involving closely related ideas. This result is a theorem of Gromov [**Gr**] proving surface area estimates for families of functions. It takes a little work to set up the statement of this theorem, so we postpone it to the chapter on polynomial methods in differential geometry. Roughly speaking, the theorem says that polynomials are a particularly efficient class of functions in terms of the surface areas of their zero sets. The proof from [**Gr**] has a parallel structure to the proof of finite field Kakeya. It also involves a different tool coming from topology, the polynomial ham sandwich theorem. This tool coming from the geometry literature plays a role in the proof of the distinct distance estimate.

To summarize the last few paragraphs, polynomials are efficient from the point of view of differential geometry, and polynomials are resilient from the point of view of error-correcting codes. These two facts are related to each other, and the proofs in both fields share some common ingredients with the proof of finite field Kakeya.

A third connection involves diophantine equations in number theory. In the early 20th century, Thue proved that a broad class of diophantine equations in two variables have only finitely many integer solutions. His theorem was important in part because it covers a much broader class of equations than any previous work in the subject. Here is the statement of the theorem.

THEOREM 1.7. Suppose that $P(x, y) \in \mathbb{Z}[x, y]$ is a homogeneous polynomial of degree $d \geq 3$ which is irreducible over $\mathbb{Z}$. (For instance $P(x, y) = y^d - 2x^d$ for $d \geq 3$.) Then, for any integer A, the diophantine equation $P(x, y) = A$ has only finitely many integer solutions $(x, y) \in \mathbb{Z}^2$.

The proof of Theorem 1.7 also involves some similar ideas to the proof of finite field Kakeya. The statement of Theorem 1.7 involves a polynomial $P(x, y)$, but the proof also involves a lot of other polynomials, called auxiliary polynomials. The auxiliary polynomials in the proof play a similar role to the polynomial in the proof of finite field Kakeya. In the chapter on diophantine equations, we will prove Theorem 1.7 and discuss the parallels with the other proofs in the book.

Finally, we mention the original Kakeya problem. The finite field Kakeya problem was invented as a cousin for the original Kakeya problem, which involves the behavior of lines in $\mathbb{R}^n$. Recall that a finite field Kakeya set $K \subset \mathbb{F}_q^n$ is a set which contains a line in every direction. Similarly, a Kakeya set $K \subset \mathbb{R}^n$ is a set which contains a unit line segment in every direction. There are several variations of the Kakeya problem, but they all have to do with how big a Kakeya set needs to be. For instance, one version asks about the minimum possible Hausdorff dimension of a Kakeya set. All known Kakeya sets $K \subset \mathbb{R}^n$ have Hausdorff dimension n. The Kakekya problem is about the possible intersection patterns of lines in $\mathbb{R}^n$, but unlike in incidence geometry, we consider infinitely many lines instead of finitely many lines. The Kakeya problem can also be rephrased in terms of the intersection patterns of finitely many thin tubes in $\mathbb{R}^n$. This description in terms of thin tubes is the most useful for working on the problem and also the most useful in applications, so we emphasize it in the book. Here is a version of the Kakeya problem involving the intersection patterns of long thin tubes.

QUESTION 1.8. Fix a dimension n and let $\delta > 0$ be a small parameter. Suppose that $\mathcal{T}$ is a set of cylindrical tubes in $\mathbb{R}^n$, each of radius $\delta > 0$ and length 1. For a tube T, let $v(T)$ be a unit vector in the direction of T. Suppose that for any two tubes $T_1, T_2 \in \mathcal{T}$, the angle between $v(T_1)$ and $v(T_2)$ is at least δ, and suppose that for any unit vector w, there is some $T \in \mathcal{T}$ so that the angle between $v(T)$ and w is at most 10δ. What is the minimum possible volume of the union of the tubes of $\mathcal{T}$?

If the tubes of $\mathcal{T}$ are disjoint, then it is easy to check that the volume of the union is on the order of 1. In the 1920s, Besicovitch constructed an ingenious example where the volume of the union goes to zero with δ. A slightly improved version of this construction [**Sch**] gives logarithmic decay:

$$|\cup_{T\in\mathcal{T}} T| \leq C_n \frac{1}{|\log \delta|}.$$

This construction is still the best one known. The Kakeya conjecture asserts that, for any $\varepsilon > 0$, the volume of the union of the tubes in $\mathcal{T}$ is at least $c(\varepsilon)\delta^{\varepsilon}$. The best known lower bounds for the volume are much weaker: for instance, if $n = 3$, we know that the volume of the union is at least $c\delta^{1/2}$.

In the 1970's, mathematicians discovered that the Kakeya problem is closely connected to a circle of problems in Fourier analysis. This connection encouraged a lot of interest in the problem, and it has been studied intensively ever since.

It is not clear how much the polynomial method can contribute to the original Kakeya problem. The proof of finite field Kakeya seems like it might be an important clue, but there are major difficulties in trying to adapt the proof from lines in $\mathbb{F}_q^n$ to thin tubes in $\mathbb{R}^n$. On the other hand, the polynomial method has had some successes proving harmonic analysis estimates related to the Kakeya problem. We will discuss all these issues in the chapter on harmonic analysis.

1.3. Outline of the book

The first part of the book is about introducing the polynomial methods we will study. In Chapter 2, we prove the finite field Kakeya theorem and the joints theorem, and we outline the ingredients of the method. In Chapter 3, we discuss why these problems were difficult to solve without polynomials and what features of polynomials make them useful. The proofs in Chapter 2 are partly based on ideas from error-correcting codes. In Chapter 4, we study the Berlekamp-Welch algorithm and other work in error-correcting codes, and we see how it relates to the proofs in Chapter 2. In Chapter 5, we discuss some earlier work in combinatorics with a similar flavor. In Chapter 6, we prove the Bezout theorem, a fundamental theorem of algebraic geometry. We will use this result in the later chapters, and we also give a proof with a somewhat similar flavor to the other proofs in the book.

The second part of the book gives background in incidence geometry. In Chapter 7, we prove the Szemerédi-Trotter theorem, and introduce some of the topological methods in the area. We discuss the distinct distance problem as well as some hard open problems in the field. In Chapter 8, we discuss incidence geometry in higher dimensions, especially dimension three. In Chapter 9, we discuss the partial symmetry approach to the distinct distance problem.

The third part of the book is about applications of the polynomial method in incidence geometry. In this part of the book, we prove Theorems 1.5 and 1.6. These

proofs involve several different tools, and we introduce one tool in each chapter. Chapter 10 introduces polynomial partitioning. This is an important tool, and it turns out to be enough to prove a slightly weaker form of the distinct distance estimate. Chapter 11 explores the connection between combinatorial structure and algebraic structure. Chapter 12 combines these tools and finishes the proof of Theorem 1.6. Chapter 13 introduces tools from ruled surface theory in algebraic geometry and proves Theorem 1.5.

The fourth part of the book discusses connections with a few other areas. Chapter 14 discusses connections with differential geometry. Chapter 15 discusses the Kakeya problem and Fourier analysis. Chapter 16 discusses Thue's work on diophantine equations.

1.4. Other connections between polynomials and combinatorics

There are a lot of interesting connections between polynomials and combinatorics. I wanted to mention a few interesting directions that have a similar flavor to the topics in this book.

The book *Linear algebra methods in combinatorics* [**BF**], by Babai and Frankl, develops a circle of ideas involving polynomials, linear algebra, and combinatorics. The recent book by Matousek, [**Ma2**], discusses many of the same ideas. We will touch on this work briefly in Chapter 5.

Alon proved a variant of the Hilbert Nullstellensatz from algebraic geometry, called the combinatorial nullstellensatz. This is a theorem about polynomials, related to classical theorems of Chevalley and Warning. He and others applied this theorem to problems in combinatorics, including some additive number theory and some graph theory. See the survey [**Al**] and the references therein for an introduction to this area.

Recently, Green and Tao [**GT**] proved some old conjectures in incidence geometry with an argument that uses a combination of topology and polynomials. We will say more about these results in Section 7.5.

1.5. Notation

I would like to introduce one convenient piece of notation here. We write $A \lesssim B$ to mean that there is some constant C so that $A \leq CB$. We write $A \sim B$ to mean that $A \lesssim B$ and $B \lesssim A$.

We will introduce other notation as it comes up, but we record here for reference a few basic pieces of notation that will come up a lot in the book. We let $\mathbb{F}$ denote a field, and we let $\mathbb{F}_q$ denote the finite field with q elements. We let $\mathrm{Poly}_D(\mathbb{F}^n)$ be the space of polynomials in n variables, with coefficients in $\mathbb{F}$, and with total degree at most D. If P is a polynomial, then we write $Z(P)$ for the zero set of P. If $\mathfrak{L}$ is a set of lines, then we write $P_r(\mathfrak{L})$ for the set of r-rich points of $\mathfrak{L}$ - the set of points that lie in at least r lines of $\mathfrak{L}$.