The Polynomial Method in Combninatorics

Steve Zhang

March 20, 2023

Abstract

In the past decade, the polynomial method in combinatorics has rapidly risen in popularity. It has shown up many times as an unexpected way to solve combinatorial problems by analyzing their algebraic structure through polynomials. The method is rooted in computer science and is comparable to finding the line of best fit on a given data set. In this paper, we explain the method and touch on a few relevant examples of its application.

1 The Method

In general, a high-level outline of the approach consists of:

1. Expressing the problem in terms of points in a vector space,
2. Finding a polynomial of low degree which vanishes on all of the points,
3. Analyzing the algebraic properties of the polynomial and forming a conclusion.

We begin by demonstrating how the polynomial is determined. Let $\mathbb{F}$ be a finite field. Suppose $p_1, \ldots, p_k$ are points in $\mathbb{F}^n$. Then, we propose the question:

Question 1.1. *Does there exist a polynomial P with degree at most d such that P vanishes on all the points $p_1, \ldots, p_k$?*

Answering this question is really just a matter of linear algebra. We define $V(d)$ to be the vector space of all polynomials in n variables with degree less than or equal to d. For example, if $n = 2$ and $d = 2$, then

$$V(d) = \{ax_1^2 + bx_2^2 + cx_1x_2 + dx_1 + ex_2 + f \mid a, b, c, d, e, f \in \mathbb{F}\}.$$

Now, let E be the linear map sending $V(d)$ to $\mathbb{F}^n$ such that

$$E(P) = (P(p_1), \ldots, P(p_k)).$$

Then, there exists a nonzero vanishing polynomial with degree less than or equal to d if and only if E has a non-trivial kernel, i.e. there exists a polynomial P

such that $E(P) = (0, \ldots, 0)$. Since $\mathbb{F}$ is finite, it is possible to iteratively check every degree d until we find a vanishing polynomial. While we cannot say much about the vanishing polynomial yet, we will see later how it can help reveal the algebraic structure of a problem.

2 Recovering Polynomials from Corrupted Data

In this section, we analyze a problem dealing with single-variable polynomials. We first provide a few elementary facts regarding polynomials.

Lemma 2.1. *A nonzero single-variable polynomial with degree d can have at most d zeroes.*

From this lemma, we can easily derive the following result.

Corollary 2.2. *If P and Q are nonzero single-variable polynomials with degree at most d, then either*

1. *$P(x) = Q(x)$ for at most d values of x, or*
2. *$P = Q$.*

Proof. Consider the polynomial $G = P - Q$. Then, $G(x) = 0$ if and only if $P(x) = Q(x)$. If G is nonzero, it must have degree at most d so it has at most d roots. Thus, when $P \neq Q$, $P(x) = Q(x)$ also has at most d roots. If $G = 0$, then we simply have $P = Q$. ■

We now introduce the problem, followed by a proof provided in [1].

Theorem 2.3. *Let $\mathbb{F}$ be a finite field with q elements such that q is large. Let $P(x)$ be a polynomial with degree $d \leq q^{\frac{1}{3}}$. Let $F(x)$ be a polynomial that agrees with $P(x)$ at least 51% of the time (i.e. $F(x) = P(x)$ for at least $\frac{51q}{100}$ distinct values of x). Then, given the function F, it is possible to determine P.*

Proof. Let $q > 10^4$. Suppose $Q(x)$ is some polynomial agreeing with $F(x)$ on at least $\frac{51q}{100}$ distinct values of x. Then, by the pigeonhole principle, $Q(x)$ must agree with $P(x)$ on at least $\frac{2q}{100}$ distinct values of x. Since

$$\frac{2q}{100} > q^{\frac{1}{3}}$$

for $q > 10^4$, we must have that $Q = P$ by Corollary 2.2. Hence, it suffices to find some polynomial $Q(x)$ agreeing with $F(x)$ on at least $\frac{51q}{100}$ distinct values of x. Since $\mathbb{F}$ is a finite field, we can simply check all polynomials of degree less than or equal to $q^{\frac{1}{3}}$ to determine P. ■

Alternatively, we can set

$$G := \{x, F(x) \mid x \in \mathbb{F}\}$$

to be the graph of F in $\mathbb{F}^2$. Let $Q(x, y)$ be the polynomial of minimal degree vanishing on G. Then, factor Q until it is irreducible. The polynomial $y - P(x)$ must divide $Q(x, y)$.

3 Finite Field Kakeya Conjecture

We first prove a lemma that is essential to the proof of the Finite Field Kakeya Conjecture. This bounds the number of zeroes of a degree d polynomial in a finite field.

Lemma 3.1 (Schwartz-Zippel)**.** *Let $p(x_1, \ldots, x_n)$ be a nonzero polynomial over a field $\mathbb{F}$ with degree d. Given some finite set $S \subseteq \mathbb{F}$, for randomly selected $a_1, \ldots, a_n \in S$,*

$$\Pr[p(a_1, \ldots, a_n) = 0] \leq \frac{d}{|S|}.$$

Proof. We will use induction on n, the number of variables in p. For $n = 1$, p has at most d roots, so the statement is trivial.

Suppose the statement holds for all $n \leq k - 1$ variables. Then, consider a polynomial $p(x_1, \ldots, x_k)$. We can rewrite p as a polynomial in x_1 by setting

$$p(x_1, \ldots, x_k) = \sum_{i=0}^{k} x_1^i p_i(x_2, \ldots, x_k).$$

Let m be the maximum value such that p_m is not the zero polynomial. Since $x_1^m p_m$ has degree at most d, we know that p_m has degree at most $d-m$. Suppose we randomly choose $a_1, a_2, \ldots, a_k \in S$. By our induction hypothesis,

$$\Pr[p_m(a_2, \ldots, a_k) = 0] \leq \frac{d-m}{|S|}.$$

Now, define the polynomial

$$q(x_1) = \sum_{i=0}^{k} x_i^i p_i(r_2, \ldots, r_k).$$

We will denote the events $p_m(a_2, \ldots, a_k) = 0$ by A and $q(a_1) = 0$ by B. Note that if A doesn't occur, then $q(x_1)$ has degree m, so by our induction hypothesis

$$\Pr[B \mid \neg A] \leq \frac{m}{|S|}.$$

Combining everything, we have that

$$\begin{aligned}
\Pr[B] &= \Pr[B \wedge A] + \Pr[B \wedge \neg A] \\
&= \Pr[B \wedge A] + \Pr[B \mid \neg A] \cdot \Pr[\neg A] \\
&\leq \Pr[A] + \Pr[B \mid \neg A] \\
&\leq \frac{d-m}{|S|} + \frac{m}{|S|} = \frac{d}{|S|}.
\end{aligned}$$

■

A Kakeya Set K in $\mathbb{R}^n$ is a compact set containing a unit line segment in every direction, and the general Kakeya Conjecture states that any such set must have dimension n. However, we look at a simpler version of the problem, involving Kakeya sets in finite fields.

Definition 3.2 (Kakeya Set)**.** Let $\mathbb{F}$ be a finite field with q elements. A Kakeya Set is a set $K \subset \mathbb{F}^n$ containing a line in every direction. In particular, K is a Kakeya set if for every slope $x \in \mathbb{F}^n$, there exists some point $y \in \mathbb{F}^n$ such that

$$\{y + ax \mid a \in \mathbb{F}\} \subseteq K.$$

Theorem 3.3 (Finite Field Kakeya Conjecture)**.** *Let K be a Kakeya set in $\mathbb{F}^n$ where $\mathbb{F}$ is some finite field with q elements. Then,*

$$|K| \geq C_n \cdot q^{n-1}$$

for some constant C_n that only depends on n.

We outline the proof from [2].

Proof. Assume for the sake of contradiction that

$$|K| < \binom{q+n-3}{n-1}.$$

Consider the set of all monomials $x_1^{d_1} x_2^{d_2} \dots x_n^{d_n}$ with degree exactly $q-2$. The size of this set is the number of solutions to

$$d_1 + d_2 + \ldots + d_n = q - 2,$$

which is exactly equal to

$$\binom{q+n-3}{n-1}.$$

Suppose we have some nonzero homogeneous polynomial $P(x_1, \ldots, x_n)$ of degree $q-2$. There are $\binom{q+n-3}{n-1}$ terms in P and less than $\binom{q+n-3}{n-1}$ points in K. Hence, there exists some arrangement of coefficients for P such that it vanishes on all points in K, as this reduces to solving a system of $|K|$ equations in more than $|K|$ variables.

We seek to show that P vanishes on all points in $\mathbb{F}^n$. Clearly, we have that $P(0, \ldots, 0) = 0$. Consider some nonzero vector $x \in \mathbb{F}^n$. By the definition of a Kakeya set, there exists some point $y \in \mathbb{F}^n$ such that

$$\{y + ax \mid a \in \mathbb{F}\} \subseteq K.$$

By the homogeneity of P, we know that if $P(z) = 0$ for some $z \in \mathbb{F}^n$, then $P(cz) = 0$ for any $c \in \mathbb{F}$. Hence, since P vanishes on all points in K, we can determine that

$$P(ay + x) = P\left(a \cdot \left(y + a^{-1}x\right)\right) = 0.$$

Since P is a degree $q-2$ polynomial, $P(ay+x)$ must be a degree $q-2$ polynomial in a. However, it has at least $q-1$ roots, corresponding to the nonzero elements of $\mathbb{F}$, so it must be the zero polynomial. Hence, plugging in $a = 0$, we find that $P(x) = 0$.

We have shown that $P(x) = 0$ for all $x \in \mathbb{F}^n$. However, by the Schwartz-Zippel lemma,

$$\Pr[P(x) = 0] \leq \frac{d}{|S|} = \frac{q-2}{q} < 1.$$

This is a contradiction, so we must have

$$|K| \geq \binom{q+n-3}{n-1} \approx \frac{q^{n-1}}{(n-1)!}.$$

■

We can come up with an even better bound using the simple observation that a product of Kakeya sets is also a Kakeya set.

Theorem 3.4. *Let K be a Kakeya set in $\mathbb{F}^n$ where $\mathbb{F}$ is some finite field with q elements. Then, for any $\epsilon > 0$,*

$$|K| \geq C_{n,\epsilon} \cdot q^{n-\epsilon}$$

for some constant $C_{n,\epsilon}$ that only depends on n and ε.

Proof. Note that for any integer r, the Cartesian product $K^r \subset \mathbb{F}^{n \cdot r}$ is also a Kakeya set. Then, applying Theorem 3.3, we find that

$$|K|^r \geq C_{n \cdot r} \cdot q^{n \cdot r - 1},$$

so

$$|K| \geq C_{n,\frac{1}{r}} \cdot q^{n - \frac{1}{r}}$$

where $\left(C_{n,\frac{1}{r}}\right)^r = C_{n \cdot r}$. ■

References

[1] Guth, L. (2012, January 4). The polynomial method in combinatorics. *Larry Guth's webpage*. `https://math.mit.edu/~lguth/polynomialmethbeamer.pdf`

[2] Dvir, Z. (2009). On the size of Kakeya sets in finite fields. *Journal of the American Mathematical Society*, 22(4), 1093-1097. `https://www.cs.princeton.edu/~zdvir/papers/Dvir09.pdf`

[3] Valiant, G. (2019). CS265/CME309: Randomized Algorithms and Probabilistic Analysis Lecture# 1: Computational Models, and the Schwartz-Zippel Randomized Polynomial Identity Test. `https://people.csail.mit.edu/madhu/ST03/scribe/lect06.pdf`