WHAT'S SPECIAL ABOUT POLYNOMIALS? (A GEOMETRIC PERSPECTIVE)

This section is for context and background. We discuss some results about polynomials from the point of view of geometry/topology. I think there are some interesting philosophical ideas here. We build up to an application of the general ham sandwich theorem to prove a geometric estimate about polynomials. This geometric argument is a precursor of the applications of the ham sandwich theorems in this section. We will not give complete proofs here. We sketch the main ideas when we can.

From the point of view of differential geometry and topology, polynomials (over $\mathbb{C}$ or $\mathbb{R}$) are strikingly efficient. I learned this point of view from V. I. Arnold's essay on the "Topological economy principle in algebraic geometry", in the Arnoldfest.

We begin with examples about complex polynomials. In fact, all these examples are true more generally of holomorphic functions. Polynomials in one variable are efficient in terms of the number of zeroes. We make this precise in the following proposition, which is closely related to material in a first course on complex analysis or in differential topology.

Theorem 0.1. *Suppose that $P : \mathbb{C} \to \mathbb{C}$ is a complex polynomial in one variable (or just a holomorphic function). We identify $\mathbb{C}$ with $\mathbb{R}^2$, and suppose that $F : \mathbb{R}^2 \to \mathbb{R}^2$ is any smooth function which agree with P outside of the unit disk $\mathbb{D}$. Finally, assume that 0 is a regular value for both P and F. Then the number of zeroes of P in $\mathbb{D}$ is less than or equal to the number of zeroes of F in $\mathbb{D}$.*

We sketch the main idea of the proof. A point $x \in \mathbb{R}^2$ is a regular point of F if the derivative $dF_x : \mathbb{R}^2 \to \mathbb{R}^2$ is an isomorphism. When we say that zero is a regular value, it means that each point x with $F(x) = 0$ is a regular point. Let $x_1, ..., x_N$ be the points in the unit disk where $F(x) = 0$. Each such x can be given a multiplicity of $+1$ if $det dF_x > 0$ and -1 if $det dF_x < 0$. We denote the multiplicity by $m(x_i)$. Let us assume for the sketch that F and P don't vanish on the unit circle S^1. Then $F : S^1 \to \mathbb{R}^2 \setminus \{0\}$, and F has a well-defined winding number around 0, denoted $W(F)$. In differential topology, one proves that the winding number $W(F)$ is equal to the sum of the multiplicities: $W(F) = \sum_i m(x_i)$. There is a similar formula for the polynomial P. Since P and F agree on S^1, $W(F) = W(P)$. But P is a holomorphic function, and so dP_x is a complex linear map which must be orientation preserving. The multiplicity of P at each of its zeroes is 1, and so the number of zeroes of P in $\mathbb{D}$ is exactly $W(P) = W(F)$. Therefore, the number of zeroes of F in $\mathbb{D}$ is at least the number of zeroes of P in $\mathbb{D}$.

The result says that P has no unnecessary zeroes. Also, there is nothing special about 0. If $w \in C$ denotes any regular value of P and F, then there are at least as many points in $\mathbb{D}$ where $F(z) = w$ as points where $P(z) = w$. There is also nothing special about the unit disk, which can be replaced by other open sets. I don't know the history of this result. It may have been known in the 19th century.

This result holds for all holomorphic functions, and in fact just for functions whose derivatives are orientation preserving.

Complex polynomials in several variables are efficient in terms of the surface area of their zero sets.

Theorem 0.2. *Suppose that $P : \mathbb{C}^n \to \mathbb{C}$ is a complex polynomial (or just a holomorphic function). We identify $\mathbb{C}$ with $\mathbb{R}^2$ and $\mathbb{C}^n$ with $\mathbb{R}^{2n}$, and suppose that $F : \mathbb{R}^{2n} \to \mathbb{R}^2$ is any smooth function which agree with P outside of the unit ball $\mathbb{B}^{2n}$. Finally, assume that 0 is a regular value for both P and F. Let $Z(P)$ denote the zero set of P, and let $Z(F)$ denote the zero set of F. Since 0 is a regular value, these are both smooth manifolds of dimension $2n-2$.*

Then the volume of $Z(P) \cap B$ is smaller than the volume of $Z(F) \cap B$:

$$Vol_{2n-2}[Z(P) \cap B] \leq Vol_{2n-2}[Z(F) \cap B].$$

Here we are using the standard Euclidean metric on $\mathbb{R}^{2n}$. If we take $n = 1$, then this theorem reduces to the first theorem, because $Z(P)$ is a finite set of points and its volume is just the number of points. A related result is that $Z(P)$ is a minimal surface. If we take $Z(P) \cap \partial B$, we get a closed $(2n-3)$-dimensional surface, and $Z(P) \cap B$ is the smallest surface with that boundary.

This result plays an important role in the theory of minimal surfaces and in differential geometry. (I am not sure of its history either. I have seen it attributed to DeRham or to Federer. I believe it dates from the 1950's.) The proof uses differential forms. It has had a significant influence in geometry - many other arguments modelled on it have appeared since then. This type of argument was dubbed a calibration argument by Harvey and Lawson who generalized it to many other settings. A good place to read about this material is their paper "Calibrated geometries" in Acta Math. 148 (1982), 47-157.

We can give some idea of the argument without mentioning differential forms as follows. Let L denote any complex line in $\mathbb{C}^n$. The intersection $L \cap Z(P)$ is just the points of L where P vanishes, and $L \cap Z(F)$ is just the points of L where F vanishes. Let us therefore consider F as a function from L to $\mathbb{C}$. It won't necessarily happen that zero is a regular value for this function, but for almost every complex line L, zero is a regular value for both F and P. Then we can apply the one-dimensional result, and we get the following.

Lemma 0.3. *For almost every complex line $L \in C^n$,*

$|L \cap Z(P) \cap B| \leq |L \cap Z(F) \cap B|$.

The intersections of a surface X with various lines and the volume of X are connected. The branch of math that studies this connection is called integral geometry. Carefully assembling the information in the last lemma, it's possible to prove that $Vol[Z(P) \cap B] \leq Vol[Z(F) \cap B]$. We won't sketch the proof here, but we give a tiny introduction to integral geometry below.

Complex polynomials are also efficient in terms of the topological complexity of their zero sets. In particular, there is a striking theorem about polynomials in two variables.

Theorem 0.4. *(Kronheimer-Mrowka) Suppose that $P : \mathbb{C}^2 \to \mathbb{C}$ is a complex polynomial in two variables. We identify $\mathbb{C}^2$ with $\mathbb{R}^4$, and suppose that $F : \mathbb{R}^4 \to \mathbb{R}^2$ is any smooth function which agree with P outside of the unit ball $\mathbb{B}^4$. Assume that 0 is a regular value for both P and F. Let $Z(P)$ denote the zero set of P, and let $Z(F)$ denote the zero set of F. Let's also assume that $Z(P)$ and $Z(F)$ are connected. Then the genus of $Z(P)$ is at most the genus of $Z(F)$.*

If $Z(P)$ or $Z(F)$ is disconnected, some form of this theorem still holds, but it takes more care to state it. The theorem was proven in the paper "Gauge theory for embedded surfaces. I." in Topology 32 (1993), no. 4, 773-826. The proof of the theorem uses gauge theory, and we can't even sketch it here. It has applications in low-dimensional topology, for example in knot theory. I believe this theorem is also true more generally for holomorphic functions (because an arbitrary holomorphic function can be well-approximated by a polynomial in any compact set). I'm curious whether some version of this topological efficiency holds for complex polynomials $P : \mathbb{C}^n \to \mathbb{C}$ – as far as I know, this is an open problem.

So far, we have seen examples of the efficiency of complex polynomials, and more generally of holomorphic functions. The reader may well say that the key property involved is being holomorphic, not being polynomial. What about real polynomials? Are any of these theorems true for polynomials over $\mathbb{R}$?

All three theorems are completely false for polynomials over $\mathbb{R}$. For example, a real polynomial may have $P(-1) = -1$, $P(1) = 1$, and may have 113 zeroes in $(-1, 1)$. A competitor function F may have only 1 zero in $(-1, 1)$ - the other 112 zeroes are unnecessary. Modifying this example a bit, it's easy to check that the second theorem is false, and the it's not hard to see that the third theorem is false too. In fact, any smooth function can be well approximated by a real polynomial, which suggests that real polynomials cannot have any special properties at all.

But if we switch our point of view from individual polynomials to the whole space of polynomials, then some version of the first two theorems survives for polynomials over $\mathbb{R}$. Let $V_n(d)$ denote the vector space of all polynomials of degree $\leq d$ in n variables. This vector space of functions is "efficient" in a certain sense.

Theorem 0.5. *Pick a degree d and consider the space of polynomials of degree $\leq d$ in one variable: $V_1(d)$. This space has dimension $d+1$. Let W be any other vector space of real-valued functions with dimension $d+1$. Every polynomial in $V_1(d) \setminus \{0\}$ has at most d zeroes. Then some function $F \in W \setminus \{0\}$ has at least d zeroes.*

This is a basic dimension counting argument, of the kind we have used many times.

Proof. Pick any d points $x_1, ..., x_d \in \mathbb{R}$. Let E be the evaluation map $E : W \to \mathbb{R}^d$ given by $E(F) = (F(x_1), ..., F(x_d))$. The map E is linear, and the dimension of the domain is greater than the dimension of the range. Therefore E has a non-trivial kernel. Let F be a non-zero element in $\ker E$. □

The set of real polynomials in n variables is also efficient in a similar way.

Theorem 0.6. *(Gromov) For any $d \geq 1$,*

$$\sup_{0 \neq P \in V_n(d)} Vol_{n-1} Z(P) \cap B^n \sim d.$$

If W is any vector space of continuous functions defined on the unit n-ball B^n, with $\dim W \geq \dim V_n(d)$, then

$$\sup_{0 \neq F \in W} Vol_{n-1} Z(F) \cap B^n \gtrsim d.$$

This theorem says that the vector space $V_n(d)$ is fairly efficient in terms of the volumes of zero sets. For a space of functions W from the unit ball B^n to $\mathbb{R}$, define $MaxVol(W)$ to be $\sup_{0 \neq F \in W} Vol_{n-1} Z(F) \cap B^n$. The theorem says that if $dimW = dimV_n(d)$, then $MaxVolV_n(d) \leq C_n MaxVolW$. (It's an open problem whether $MaxVolV_n(d) \leq MaxVolW$.)

The first half of the result comes from integral geometry, and it was known in the early 20th century. The second half is much more recent. It was proven by Gromov in the paper, "Isoperimetry of waists and concentration of maps" in Geom. Funct. Anal. 13 (2003), no. 1, 178-215.

We describe the proof of each half.

1. Let P be a non-zero polynomial of degree $\leq d$. For a line $l \subset \mathbb{R}^n$, either $|l \cap Z(P)| \leq d$ or else $l \subset Z(P)$. If $X^{n-1} \subset \mathbb{R}^n$ is a hypersurface, then the volume of X is connected to the number of intersections $|l \cap X|$ with different lines. The connection is made by the Crofton formula, which we now describe.

Let $AG(1,n)$ be the set of affine lines in $\mathbb{R}^n$. The group of rigid motions of $\mathbb{R}^n$, G_{rigid}, acts transitively on $AG(1,n)$. In fact, $AG(1,n)$ is the quotient of the group of rigid motions by the stabilizer of one line. Using the Haar measure on G_{rigid}, we get a G_{rigid}-invariant measure on $AG(1,n)$, $d\mu$. This measure is unique up to scaling.

Theorem 0.7. *(Crofton) There exists a constant α_n so that the following equation holds for every (smooth) hypersurface $X \subset \mathbb{R}^n$:*

$$Vol_{n-1}(X) = \alpha_n \int_{AG(1,n)} |l \cap X| d\mu(l).$$

We give the idea of the proof. We abbreviate the RHS by $Crof(X)$. We want to prove that the two sides are equal, and we note some qualities that the two sides have in common.

1. Disjoint unions. If X is the disjoint union of X_1 and X_2, we have $Vol_{n-1}X = Vol_{n-1}X_1 + Vol_{n-1}X_2$ and $Crof(X) = Crof(X_1) \cap Crof(X_2)$.

2. Rigid motion invariance. If g is a rigid of $\mathbb{R}^n$, then $Vol_{n-1}(gX) = Vol_{n-1}(X)$ and $Crof(gX) = Crof(X)$.

We choose α_n so that $Crof([0,1]^{n-1}) = 1 = Vol_{n-1}([0,1]^{n-1})$. By the two properties above, we easily see that $Crof([0,s]^{n-1}) = s^{n-1}$ for any s. (We start with positive integers and with $s = 1/N$, and then rational s, and then take a limit to get all s.) Next if X is a finite union of (n-1)-cubes with various side lengths, we see that $Vol_{n-1}X = Crof(X)$.

Finally, given an arbitrary hypersurface X, we approximate X by X_{cub} - a finite union of (n-1)-cubes. We just have to check that $Vol_{n-1}(X_{cub})$ approximates $Vol_{n-1}(X)$ and that $Crof(X_{cub})$ approximates $Crof(X)$.

Now using the Crofton formula, we can bound the volume of $Z(P) \cap B^n$. Note that if l is any line which intersects B^n, then $|S^{n-1} \cap l| = 2$. The set of lines l with $l \subset Z(P)$ has measure 0. So we see that for $d\mu$ almost every $l \subset \mathbb{R}^n$,

$$|Z(P) \cap B^n \cap l| \leq (d/2)|S^{n-1} \cap l|.$$

Using the Crofton formula, we see that $Vol_{n-1}Z(P) \cap B \leq (d/2)VolS^{n-1}$. This inequality is sharp for every even d by taking $Z(P)$ to be a union of $d/2$ spheres with radii very close to 1. (The sharp argument and example were explained to me by Jake Solomon.)

The second half of the theorem follows from the general ham sandwich theorem, which we recall.

Theorem 0.8. *(Stone-Tukey) If W is a vector space of continuous functions from B^n to $\mathbb{R}$, and $U_1, ..., U_N \subset B^n$ are finite volume open sets, with $N < dimW$, and if each function $F \in W \setminus \{0\}$ has $meas(Z(F)) = 0$, then there is a non-zero $F \in W$ which bisects each set U_i.*

In our case, $dimW \geq dimV_n(d) \sim d^n$. (If any non-zero $F \in W$ has $Z(F)$ with positive Lebesgue measure, then $Vol_{n-1}Z(F)$ is infinite.) We can apply the theorem. We let $U_1, ..., U_N$ be $\sim d^n$ disjoint balls in B^n, each with radius $\sim d^{-1}$.

A hypersurface which bisects the unit n-ball must have (n-1)-volume at least $c_n > 0$. This fact follows, for example, from the isoperimetric ienquality. By scaling, $Z(F) \cap U_i$ must have (n-1)-volume at least $c_n d^{-(n-1)}$. Therefore, $Vol_{n-1} Z(F) \gtrsim d^n d^{-(n-1)} = d$.

To end this chapter, let's mention a couple themes that appear in both the geometry today and the combinatorics we've been studying. We always exploit the fundamental fact that a non-zero degree d polynomial in one variable vanishes at most d times. Next we come to polynomials in several variables. This is a very large space, and it has the simple but remarkable property that if we restrict a degree d polynomial in several variables to a line, then we get a degree d polynomial in one variable. So we get a lot of information about what the polynomial is doing on each line. In both settings, we want to assemble that information to give global information about what the polynomial is doing globally. In the geometric setting, integral geometry gives an important tool for assembling the information, leading to some of the geometric estimates above.

We've also seen the general ham sandwich theorem in both settings. The way it's applied is a little different, but the geometric theorem on efficiency of real polynomials is still a kind of precursor for the approach in this chapter.