Geometry of polynomials, a course by Sergei Tabachnikov.

Adeel A. Khan <adeel2@umbc.edu>

6-15 July 2010

Contents

§ 1. When does a polynomial have multiple roots? 2

§ 2. Projective duality. 3

§ 3. When does a polynomial have multiple roots?, part 2. 5

§ 4. The resultant and discriminant.. 6

§ 5. Fundamental theorem of algebra. 10

§ 6. Hypersurface of polynomials with multiple roots in the complex case.. . 11

§ 7. Analytical interpretation of $S^n(\mathbb{RP}^2) = \mathbb{RP}^{2n}$. 13

§ 8. Chebyshev polynomials. 15

§ 9. Roots of trigonometric polynomials. 17

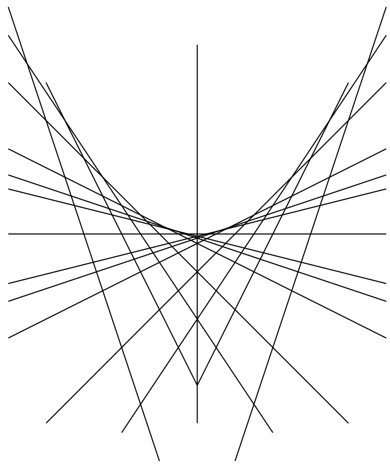
§ 10. The four-vertex theorem. 20

§ 11. Trigonometric and periodic polynomials, part 2. 22

The latest version of this document can be found at `<http://umbc.edu/~adeel2>`.

§ 1. When does a polynomial have multiple roots?

Quadratic polynomials as families of lines. Take the equation $x^2 + px + q = 0$. Usually we take the parameters to be p, q, and we have a two-parameter family of quadratic polynomials. We could also consider x to be the parameter and look at it as a one-parameter family of lines. That is, for each value of x, we have a line in the pq-plane. For example, $x = 0$ gives us $q = 0$, $x = -1$ gives us $q = p - 1$, $x = 1$ gives us $q = -p - 1$, and so on. We can graph all these lines and as below, a curve appears such that each of these lines are tangent to it. It appears to be a parabola.



For each line, write the corresponding value of x at the point of tangency with the curve. Then we have something of an "axis", or a parametrization of the curve.

Now for any choice (p_0, q_0) we can find the roots x_1, x_2 of $x^2 + p_0x + q_0 = 0$ as follows. Look at the lines passing through (p_0, q_0) that are tangent to the curve. x_1 and x_2 are the x-values of the points of tangency by the parametrization above. Obviously this isn't the most efficient algorithm to find the roots of a quadratic polynomial and we're actually more interested in something else here.

Envelope of a family of lines. The curve above is called the *envelope* of this family of lines. We can define it rigorously as follows.

Definition. Given a family of lines $\ell(x)$ let $P(x)$ be the intersection of $\ell(x)$ with an infinitesimally close line $\ell(x + \varepsilon)$ as $\varepsilon \to 0$. That is, $P(x) := \lim_{\varepsilon\to 0} \ell(x) \cap \ell(x + \varepsilon)$. Then the curve $P(x)$ is called *envelope* of $\ell(x)$.

It's not hard to see that the envelope is then just the intersection of f and f'. For example in this case we solve the system

$$f = x^2 + px + q = 0$$
$$f' = 2x + p = 0$$

to get $q = (1/4)p^2$. As we suspected, it is indeed a parabola.

Multiple roots. The point (p, q) lies on the envelope exactly when $x^2 + px + q = 0$ has a multiple root (when considered as a quadratic in x). This follows from the following elementary lemma.

Lemma (condition for multiple roots)**.** *For a polynomial $f(x)$, a is a multiple root iff $f(a) = f'(a) = 0$.*

Proof. Suppose a is a multiple root of f. Then $f(a) = 0$ and we can write $f(x) = (x-a)^2 g(x)$ for some $g(x)$. Then $f'(x) = 2(x-a)g(x) + (x-a)^2 g'(x)$ so $f'(a) = 0$.

If $f(a) = f'(a) = 0$ we can write $f(x) = (x-a)g(x)$. Then $f'(x) = (x-a)g'(x) + g(x)$ and since $f'(a) = 0$, $g(a) = 0$. Thus $g(x) = (x-a)h(x)$ for some $h(x)$. Thus $f(x) = (x-a)^2 h(x)$ so a is a multiple root of f. □

§ 2. Projective duality.

Point-line duality. Consider the equation $y + px + q = 0$ as a two-parameter family of lines with x, y as parameters. Any choice of (x, y) gives a (nonvertical) linear equation in p, q. In this case note that there is a symmetry in the equation in the sense that considering the equation as a two-parameter family with p, q as parameters would be exactly the same up to renaming of variables. So there is a duality here. A point (x_0, y_0) in xy-plane corresponds to the line $q = -x_0 p - y_0$ in pq-plane. Further, a point A on a line ℓ in the xy-plane is corresponds to a point L on a line a in the pq-plane.

Theorem (Pappus)**.** *If we have two sets A, B, C and A', B', C' of collinear points, then the intersection points $I := \overline{AB'} \cap \overline{BA'}$, $J := \overline{AC'} \cap \overline{CA'}$, and $K := \overline{BC'} \cap \overline{CB'}$ are also collinear.*

Remark. By duality we can construct an equivalent theorem. This is an exercise.

Example (dual of $y = x^a$)**.** Let γ be the curve $y = x^a$. At an arbitrary point $(t, t^a) \in \gamma$ we have the tangent line

$$y - t^a = -at^{a-1}(x - t).$$

We can rewrite this in the form $y + px + q = 0$.

$$y + (at^{a-1})x + -(a+1)t^a = 0.$$

Then we have $p(t) = at^{a-1}$ and $q(t) = -(a+1)t^a$. So the dual curve γ^* is given by $q = C \cdot p^{a/(a-1)}$ for some constant C (that can be calculated easily).

Projective duality. The projective plane is the space of lines through the origin in 3-space. Duality is built into the projective plane. For a vector space V consider its projectivization $P(V)$ and the projectivization of its dual, $P(V*)$. We can take a point in $P(V)$ and assign to it a line (hyperplane) in $P(V^*)$. So we consider the linear operators in $P(V^*)$ that vanish on V: $\{f : f(v) = 0\}$.

$$\begin{aligned} \mathbb{RP}^2 = P(V) &\to P(V^*) \\ P \in P(V) &\mapsto \{f : f(P) = 0\} \end{aligned}$$

An arbitrary hyperplane in $P(V)$ can be written $px + qy + rz = 0$. Above we considered the special case where $p = z = 1$. Now we can get rid of our clumsy assumption that lines are non-vertical.

A version of projective duality on the sphere. There is another way to view this duality on the sphere instead. We consider a unit sphere centered at the origin. Every line intersects it at two points, so we can define the projective plane as the sphere with diametrically opposite points identified. Now there is a version of projective duality on the sphere.

We want a correspondence between points and “lines” on the sphere, but we need to define what a line on the sphere is. Let a line on the sphere be a great circle. Then for every point A on the sphere we associate to it the natural great circle, for which A becomes a pole with the great circle being an equator.

Now we consider smooth curves on the sphere. What do they correspond to? Take a smooth γ on the sphere. Given a curve, we need to consider the one-parameter family of its tangent “lines” (great circles). To construct the dual curve we just translate every point on γ by $\pi/2$ in the perpendicular direction. ($\pi/2$ is the distance between a pole and its equator.) Then we have a new curve γ^*. Note that $(\gamma^*)^*$ is just the antipodal curve of γ. This makes sense because if we have two points A and B then the distance between them corresponds to the angle between their great circles.

Example (circle of radius r)**.** If γ is a circle of radius r it is easy to see that γ^* is also a circle, with radius $\pi/2 - r$.

Singularities in curves in the projective plane.

Example (curve with 3 cusps)**.** Let's come back to the projective plane. If γ is a curve with 3 cusps, its dual curve must have 3 inflections. Since γ has no inflections, γ^* has no cusps. Thus γ^* must be a smooth curve in the projective plane with 3 inflections and no cusps. Suppose a line intersects γ^* at one point. This line corresponds to a point in $P(V)$ that is tangent to γ at one point. If a line intersects γ^* at three points, then the corresponding point in $P(V)$ is tangent in three places.

The Möbius theorem arises naturally from this example.

Theorem (Möbius)**.** *A simple closed non-contractible curve in the projective plane must have at least three inflection points.*

§ 3. When does a polynomial have multiple roots?, part 2.

Three-parameter family of polynomials. Consider the three-parameter family of polynomials $x^4 + px^2 + qx + r = 0$. We don't need a cubic term because it can be easily killed. Thus we are considering a general quartic polynomial as a one-parameter family of planes (the parameter is x). As in the quadratic case let's find the envelope curve.

Let $\pi(x)$ denote the plane obtained by some x. We intersect two infinitesimally close planes $\pi(x)$ and $\pi(x+\varepsilon)$:

$$\ell(x) := \lim_{\varepsilon\to 0} \pi(x) \cap \pi(x+\varepsilon).$$

$\ell(x)$ is a family of lines that spans a surface Σ, the *envelope* of our planes, and that is the object of our study. Σ is made of lines and is therefore called *ruled.* Other examples of ruled surfaces are $x^2 + y^2 - z^2 = 1$ (hyperboloid) and $xy = z$ (hyperbolic paraboloid). In fact these are both doubly ruled.

Now $\ell(x)$ is given by the equations

$$x^4 + px^2 + qx + r = 0$$
$$4x^3 + 2px + q = 0.$$

Let $P(x) := \lim_{\varepsilon\to 0} \pi(x-\varepsilon)\cap\pi(x)\cap\pi(x+\varepsilon) = \ell(x-\varepsilon)\cap\ell(x)$ be the intersection of three infinitesimally close planes, or equivalently the intersection of two infinitesimally close lines. Typically two such lines are skew but these ones come from a family of equations and thus an intersection point is guaranteed. Now the equation of $P(x)$ is given by

$$\begin{aligned} x^4 + px^2 + qx + r &= 0 \\ 4x^3 + 2px + q &= 0 \\ 12x^2 + 2p &= 0 \end{aligned}$$

Now we draw the surface Σ in p, q, r-space. It is a 3-dimensional version of what we did yesterday. This surface is called a *swallowtail* and it consists of polynomials with multiple roots.

We can solve for p, q, r to get $p = -6x^2, q = 8x^3, r = -3x^4$. This curve forms the red *cuspidal edge* in the picture.

Suppose you have a curve $\gamma(x)$ in 3-space. Consider the surface $R(x,t)$ which is made up of tangent lines to this curve. R is given parametrically by

$$R(x,t) := \gamma(x) + t\gamma'(x),$$

where x is the parameter along the curve and t is the parameter along the tangent line. What about tangent plane to this surface at points on the tangent line at some point? We have partial derivatives $\overline{R}_x = \gamma' + t\gamma''$ and $\overline{R}_t = \gamma'$, so that $\text{Span}(R_x, R_t) = \text{Span}(\gamma', \gamma'')$.

We find that this surface is always of zero curvature. A sheet of paper in 3-space is an example of a surface of zero curvature. A classical result due to Euler says that a surface of zero curvature can be classified. When we take a sheet of paper in 3-space, it is made of straight lines and all these lines are tangent to a curve. This curve is not within the sheet itself but does exist. In particular, the swallow-tail Σ is closely similar to the sheet of paper. You need to extend the sheet of paper.

We use Euclid's algorithm on $f = x^4 + px^2 + qx + r$ and $f' = 4x^3 + 2px + q$ to get the equation of the surface Σ:

$$16p^4r - 4p^3q^2 - 128p^2r^2 + 144pq^2r + 256r^3 - 27q^4 = 0.$$

§ 4. The resultant and discriminant.

The resultant. Let $f(x) = \sum_{i=0}^{n} a_i x^{n-i}$ and $g(x) = \sum_{j=0}^{m} b_j x^{m-j}$. Question: when do they have a common factor?

Having a common factor is equivalent to having a common root. There is an easy approach to this using linear algebra. If f, g have a common factor w we can write $f = vw$, $g = uw$. Then $fu = gv$, and $\deg u \le m - 1$, $\deg v \le n - 1$. So let u and v have coefficients u_α and v_β, respectively: $u = u_0 x^{m-1} + u_1 x^{m-2} + \cdots + u_{m-1}$, $v = v_0 x^{n-1} + v_1 x^{n-2} + \cdots + v_{n-1}$. Then we get

$$\begin{aligned} a_0 u_0 &= b_0 v_0 \\ a_1 u_0 + a_0 u_1 &= b_1 v_0 + b_0 v_1 \\ a_2 u_0 + a_1 u_1 + a_0 u_2 &= b_2 v_0 + b_1 v_1 + b_0 v_2 \\ \vdots &= \vdots \end{aligned}$$

We get a system of linear equations. We have a nontrivial solution if the determinant of the corresponding matrix (the *Sylvester matrix*) is nonzero.

Definition. The Sylvester matrix is constructed as follows. The first row is

$$\begin{bmatrix} a_0 & a_1 & \cdots & a_n & 0 & 0 & \cdots & 0 \end{bmatrix},$$

and the next $(m-1)$ rows are obtained by shifting to the right. The $(m+1)$-th row is

$$\begin{bmatrix} b_n & b_{n-1} & \cdots & b_1 & b_0 & 0 & \cdots & 0 \end{bmatrix},$$

and the rest are obtained by shifting to the right.

For example, if $n = 3$ and $m = 2$ we get the 5×5 matrix

$$\begin{bmatrix} a_0 & a_1 & a_2 & a_3 & 0 \\ 0 & a_0 & a_1 & a_2 & a_3 \\ b_0 & b_1 & b_2 & 0 & 0 \\ 0 & b_0 & b_1 & b_2 & 0 \\ 0 & 0 & b_0 & b_1 & b_2 \end{bmatrix}.$$

The determinant of the Sylvester matrix is called the *resultant* $R(f, g)$ of f and g.

Example (Sylvester matrix for quartic)**.** Let's consider the Sylvester matrix for $f(x) =$

$x^4 + px^2 + qx + r$ and $f'(x) = 4x^3 + 2px + q$. We get

$$\begin{bmatrix} 1 & 0 & 0 & -4 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -4 & 0 & 0 \\ p & 0 & 1 & -2p & 0 & -4 & 0 \\ q & p & 0 & -q & -2p & 0 & -4 \\ r & q & p & 0 & -q & -2p & 0 \\ 0 & r & q & 0 & 0 & -q & -2p \\ 0 & 0 & r & 0 & 0 & 0 & -q \end{bmatrix}.$$

By taking the discriminant of this matrix we get precisely the formula above for the equation of Σ.

It shouldn't be a great surprise that the resultant of the two polynomials can be written in terms of their roots since we know the relationships between the roots and coefficients of a polynomial.

Theorem. *Suppose f has roots $x_1, \ldots, x_n$ and g has roots $y_1, \ldots, y_n$. Then*

$$\begin{aligned} R(f,g) &= a_0^m b_0^n \prod_{i,j} (x_i - y_j) \\ &= a_0^m \prod_i g(x_i) = b_0^n \prod_j f(y_j). \end{aligned}$$

Proof. We can write

$$f(x) = a_0 \prod_{i=1}^{n} (x - x_i) = \sum_{i=0}^{n} a_i x^{n-1},$$

and so we have

$$\begin{aligned} a_n &= (-1)^n a_0 (x_1 x_2 \cdots x_n) \\ a_{n-1} &= \pm a_0 \left(x_1 \cdots x_{n-1} + x_1 x_2 \cdots x_{n-2} x_n + \cdots \right) \\ \vdots &= \vdots \\ a_1 &= \pm a_0 \left(x_1 + \cdots + x_n \right) \end{aligned}$$

These sums are *symmetric sums* σ_j of x_i.

So in the Sylvester matrix we replace coefficients a_i and b_i by terms in x_i and y_i. It is clear that the determinant will then have the factor a_0 m times and the factor b_0 n times, so this explains the $a_0^m b_0^n$ factor in $R(f,g)$.

Lemma. *The determinant is a homogeneous polynomial of degree mn.*

What we see at position (i,j) on the left half of the matrix is the symmetric sum $\sigma_{i-j}(x)$

which has degree $i - j$. On the right half we have $\sigma_{i-j+m}(y)$ which has degree $i - j + m$. We have the numbers $1, 2, \ldots, m + n$ and we have the permutations $\alpha_1, \alpha_2, \ldots, \alpha_{m+n}$. For the degree of the determinant we have $\sum i = \sum \alpha_i = 0$ and what remains is mn.

Thus the determinant has degree mn and vanishes whenever there are common roots, so we conclude that it is divisible by the product $\prod_{i,j}(x_i - y_j)$. So $R(f, g)$ is equal to some scalar multiple of $\prod_{i,j}(x_i - y_j)$, and the claim follows. □

The discriminant. The discriminant is constructed similarly to the resultant but with only one polynomial instead of two. The question the discriminant answers is: when does a polynomial have multiple roots? We have discussed this question many times over the last few days and of course the answer is when $R(f, f') = 0$. But the definition you find in most algebra books for the discriminant is different.

Definition. The *discriminant* of f is defined to be

$$D(f) := a_0^{2n-2} \prod_{i<j} (x_i - x_j)^2.$$

For example, for the standard quadratic polynomial ax^2+bx+c we know the discriminant is $b^2 - 4ac$.

Theorem (relationship between determinant and $R(f, f')$)**.** *For a polynomial f we have*

$$R(f, f') = \pm a_0 D(f).$$

This is an unsurprising theorem given the fact that they both answer the same question. The proof is similar to the previous theorem and we will not dwell on it.

Example. Let's consider the following polynomial which is interesting in many situations.

$$f_n(x) = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots + \frac{1}{n!}x^n,$$

the truncated exponential series. Exercise: prove

$$D(n!f(x)) = (-1)^{\binom{n}{2}}(n!)^n.$$

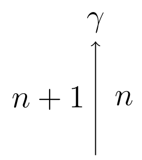
Remark. Whenever you deal with a relatively complicated object which you want to understand (like the polynomial $x^4 + px^2 + qx + r$) it's useful and important to include this object into a family, and then try to understand the hypersurface in the space of these objects which consists of singular objects. (Example regarding the latest big development in knot theory which was in 1990 by the Russian topologist Victor Vassiliev.)

§ 5. Fundamental theorem of algebra.

Theorem (Fundamental Theorem of Algebra)**.** *Every polynomial with complex coefficients with degree n has exactly n roots, not necessarily distinct.*

This is a very well-known theorem and because it is so familiar it is hard to appreciate that it is not such a trivial statement. We will look at a few topological proofs.

Proof 1. We have a continuous closed oriented curve γ and a point x not on the curve. We assign a *rotation number* (winding number) $R(\gamma, x) \in \mathbb{Z}$ which informally is the number of times the curve goes around the point x. This number changes continuously when you change the curve and point continuously as long as x is not on the curve. Let's consider a combinatorial approach to finding $R(\gamma, x)$. What happens when the point crosses the point γ? When a point crosses the curve towards the positive direction the rotation number increases by one.



Let $f(z) = z^n + a_1 z^{n-1} + \cdots + a_n$. Consider the family of concentric circles whose radius ranges from very small to very large. We assume that f has no roots. This means that the image of the circle for any radius r does not touch the origin. $\gamma_r := \{f(z) : |z| = r\}$ is the image of radius r. Now we look at $R(\gamma_r, 0)$. If a continuous function takes integral values it must be constant. So to arrive at a contradiction we need to compute the values at two values of r and see that they are different. Take $r = \varepsilon$ and $r = 1/\varepsilon$, i.e. ∞. For $r = \varepsilon$ it is clear that $R(\gamma_r, 0) = 0$. For a very large circle the value of f is dominated by z^n. So the answer should be the same as for z^n. But z^n just turns things n times around, so we claim the rotation number is $R(\gamma_\infty, 0) = n$. First of all we can write our polynomial as

$$f(z) = z^n \left(1 + \frac{a_1}{z} + \cdots + \frac{a_n}{z^n}\right).$$

We see

$$\left|\frac{a_1}{z} + \cdots + \frac{a_n}{z^n}\right| \leq \frac{|a_1|}{|z|} + \cdots + \frac{|a_n|}{|z^n|}$$

by the triangle inequality.

$$\frac{|a_1|}{r} < \frac{1}{2n}, \cdots$$

To finish the argument let's take $t \in [0, 1]$ and write

$$f(z) = z^n \left(1 + t\left(\frac{a_1}{z} + \cdots + \frac{a_n}{z^n}\right)\right).$$

We have a continuous family of curves and none of these curves passes through the origin. For $t = 0$ we know the answer is 1 and for $t = 1$ we see the answer must be 1 as well. And we are done. □

Another combinatorial proof due to Gauss was presented.

§ 6. Hypersurface of polynomials with multiple roots in the complex case.

Let's consider the swallow-tail curve again. This picture represents the hypersurface of polynomials with multiple roots inside the space of polynomials. This hypersurface can be considered for the space of complex polynomials as well.

Consider the space of monic polynomials in $\mathbb{C}^n$. Take a polynomial with n distinct roots. Then take any other point, another polynomial, also with n distinct roots. We take a path between them that avoids the hypersurface of polynomials with multiple roots. Then the claim is that it is possible to do this, and there is no point where the number of roots changes, unlike in the real case. Let's justify this picture. Consider the space $\mathbb{C}^n$ again with the hypersurface Δ inside. This contains tuples $(\alpha_1, \ldots, \alpha_n) \in \mathbb{C}^n$ which are roots of the polynomials in the first hypersurface. Then there is a projection F from this to the first, $F(\alpha_1, \ldots, \alpha_n) = \prod_{i=1}^n (z - \alpha_i)$.

I have a polynomial $f(z)$ with n distinct roots $\alpha_1, \ldots, \alpha_n$. Then I move this polynomial infinitesimally to get $f + \varepsilon g$. I'm restricted to monic polynomials so g is actually a polynomial of degree $n - 1$. Now what I want to find are roots of this polynomial which will be in the form $\alpha_1 - \varepsilon\beta_1, \ldots, \alpha_n - \varepsilon\beta_n$. I claim that for every g this problem has a solution. This is a nice algebraic exercise. What we are proving that the derivative dF of the the smooth map F is surjective. Such maps are called *submersions*. Write

$$f + \varepsilon g = \prod_{i=1}^{n} (z - \alpha_i + \varepsilon\beta_i).$$

Now calculus tells us what to do.

$$g(z) = \sum_{i=1}^{n-1} \beta_i \prod_{i \neq j} (z - \alpha_j).$$

We can substitute α_i to get

$$g(\alpha_i) = \beta_i \prod_{i \neq j} (\alpha_i - \alpha_j).$$

We must have

$$\beta_i = \frac{g(\alpha_i)}{\prod(\alpha_i - \alpha_j)}.$$

We can verify that this works.

$$g(z) \equiv \sum_{i=1}^{n} \frac{g(\alpha_i)\prod(z - \alpha_j)}{\prod(\alpha_i - \alpha_j)}.$$

Both polynomials have degree $n-1$ and they coincide at the n points α_i, so they must be identical.

Now it's possible that some root or collection of roots escapes to infinity. For example, consider a hyperbola that is vertically projected to the real line. But as we saw before if there is a bound on the absolute values of the coefficients of a polynomial then there is a bound on its roots. Here we have bounded coefficients and thus bounded roots, and thus this case is impossible. We can thus conclude that dF is surjective.

Symmetric powers. We have $S^n(\mathbb{C}) = \mathbb{C}^n$ where $S^n(\mathbb{C}) = \mathbb{C}^n/S_n$ is the nth symmetric power of $\mathbb{C}$. This is a topological statement, which means that this correspondence is a continuous correspondence. If you move the roots a little bit then the coefficients move a little bit, and vice versa.

The classical concept of a *configuration space* is a collection of n distinct points that are not allowed to collide. In our situations we allow collisions but do not care about order. So we consider n points $x_i \in \mathcal{M}$. The configuration space is $S^n(\mathcal{M})$.

Examples.

- $S^2(\mathbb{R})$ is the half-plane $\{(x, y) : x \geq y\}$.
- $S^2(S^1)$ is the Möbius band.

Projective space. Take a vector space V, consider the space V^{n+1} and remove 0. Then factor out by the relation $\sim$ where $v_1 \sim v_2$ iff $\exists t \neq 0 : v_1 = tv_2$. The result is called the projective space over V, $\mathbb{P}(V) = \mathbb{P}^n = (V^{n+1} \setminus 0)/\sim$. In other words, projective space is the space of one-dimensional subspaces (lines).

We can extend our original statement $S^n(\mathbb{C}) = \mathbb{C}^n$ to get the more interesting statement $S^n(\mathbb{CP}^1) = \mathbb{CP}^n$. $\mathbb{CP}^1$ is the Riemann sphere, S^2 from a topological point of view and $\mathbb{C} \cup \infty$ from an analysis point of view. A bijection between the sphere and $\mathbb{C} \cup \infty$ can be seen by stereographic projection.

Can we get an analogous result for $\mathbb{R}$? We have seen that for $n = 2$ we have $S^2(\mathbb{RP}^1) = S^2(S^1) = \text{Mb}$ so this doesn't work. We need to make the following modification:

Theorem. $S^n(\mathbb{RP}^2) = \mathbb{RP}^{2n}$.

This can be proved by topology.

§ 7. Analytical interpretation of $S^n(\mathbb{RP}^2) = \mathbb{RP}^{2n}$.

Harmonic functions and polynomials. A function $f(x_1, \dots, x_n)$ is harmonic if $\Delta f = 0$ where

$$\Delta f := f_{x_1x_1} + f_{x_2x_2} + \cdots + f_{x_nx_n}.$$

Recall that Δ denotes the Laplacian, i.e. $\Delta f = \operatorname{div}\operatorname{grad} f$. Defining grad requires a metric while div requires a volume, which is given when we have a metric. Thus the definition of Δ is valid whenever we have a Riemannian manifold with a metric. The sphere $S^{n-1} \subset \mathbb{R}^n$ for example is a Riemannian manifold so Δ is defined on the sphere. Denote the Laplacian on the sphere by $\tilde{\Delta}$.

Example. Take $S^1 \subset \mathbb{R}^2$. So we have two variables. It is clear that the Laplacian operator preserves homogeneous polynomials. In particular if we are interested in harmonic polynomials in $\mathbb{R}^2$ then this linear space decomposes into components corresponding to the degree of homogeneity. Let's look at the number of harmonic polynomials of degree d for small d.

$\deg d$	0	1	2	3	$\cdots$
dim	1	2	2	2	$\cdots$

For $d = 3$ we find that the basis of the space is $\{x^3 - 3xy^2, y^3 - 3yx^2\}$. Note that these are just $\operatorname{Re} z^3$ and $\operatorname{Im} z^3$. In fact for any degree d we have two basis elements which are precisely $\operatorname{Re} z^d$ and $\operatorname{Im} z^d$. This is not hard to check.

If we restrict harmonic polynomials of degree d on the unit sphere to pure harmonic functions then we get $\cos d\alpha$ and $\sin d\alpha$. We have $\tilde{\Delta} f(\alpha) = f''(\alpha)$, and $\cos''(d\alpha) = -d^2\cos(d\alpha)$, $\sin''(d\alpha) = -d^2\sin(d\alpha)$. These are eigenfunctions.

Something similar happens in every dimension.

Lemma (extension of Laplacian on sphere)**.** *We have the Laplacian Δ over $\mathbb{R}^n$ and $\tilde{\Delta}$ over S^{n-1}. Take a homogeneous polynomial f of degree k. Then, taking the restriction of f to S^{n-1}, we have $\tilde{\Delta} f = r^2\Delta f - [k^2 + k(n-2)]f$.*

In general if you have a function on the sphere you can extend it to the ambient space and choose any degree of homogeneity.

If f is harmonic, then $\tilde{\Delta}$ is not 0 but f is an eigenfunction for the spherical Laplacian operator with eigenvalue $-(k^2 + k(n-2))$, i.e. $\tilde{\Delta} = -(k^2 + k(n-2))f$.

Example. Now let's consider dimension $n = 3$ and see how many harmonic polynomials of various degrees there are. We have the space of polynomials of degree k $S^k\mathbb{R}^3$ and the map $\Delta : S^k\mathbb{R}^3 \to S^{k-2}\mathbb{R}^3$. We are interested in the dimension of the kernel. The dimension of $S^k\mathbb{R}^n$ is $\binom{k+2}{2}$. We need to know that the map is surjective. We prove this for all n by induction on k. Consider the monomial x^py^q where x is the first coordinate and y is the product of the other coordinates. In the "pure" case x^p is clearly in the image of the Laplacian because it is equal to

$$x^p = \Delta\left(\frac{x^{p+2}}{(p+2)(p+1)}\right).$$

For x^py^q consider

$$\Delta\left(\frac{x^{p+2}y^q}{(p+2)(p+1)}\right) = x^py^q + \Delta_y(u).$$

The term on the right has a smaller degree than q and thus is covered by our inductive assumption, so we can write $\Delta_y(u) = \Delta(v)$. Thus $x^py^q = \Delta u - \Delta v = \Delta(u - v)$.

Thus we conclude that the dimension of the space of harmonic polynomials of degree k is $\binom{k+2}{2} - \binom{k}{2} = 2k + 1$.

Now we claim that the projective space $\mathbb{RP}^{2n}$ is the projectivization of the space of harmonic polynomials of degree n.

Spherical functions and a way to construct them. Start with $\frac{1}{r}$ which is a harmonic function. Given this we can build many more harmonic functions. Choose a vector $v \in \mathbb{R}^3$ and take the directional derivative of $1/r$ with respect to v. This is a new function and we claim that this function is also harmonic. This is because the two operations, taking the Laplacian and taking the directional derivative, are commutative. Thus

$$\Delta\left(L_v\left(\frac{1}{r}\right)\right) = L_v\left(\Delta\left(\frac{1}{r}\right)\right) = L_v(0) = 0.$$

By restricting to the unit sphere we get $r = 1$ and we will end up with only the numerator, which is some polynomial. So we get all spherical functions, and Maxwell's theorem says this construction gives all of them.

Example. Take a polynomial P of degree r. Take the directional derivative in the direction of some x.

$$r_x = \frac{1}{2} \cdot 2x \cdot (x^2 + y^2 + z^2)^{1/2} = \frac{x}{r}$$

so that

$$L_x \left(\frac{P}{r^a} \right) = \frac{P_x}{r^a} - \frac{Par^{a-1}r_x}{r^{2a}} = \frac{P_x}{r^a} - \frac{aPx}{r^{a+2}} = \frac{P_x(x^2 + y^2 + z^2) - axP}{r^{a+2}}.$$

Now we look at the theorem.

Theorem (Maxwell)**.** *All spherical functions in* $\mathbb{R}^3$ *are obtained as* $L_{v_1} L_{v_2} \cdots L_{v_k} \left(\frac{1}{r}\right)$ *restricted to the sphere* S^2*, where* v_i *are uniquely defined by a function, up to order and nonzero factors.*

Note that the topological statement $S^n(\mathbb{RP}^2) = \mathbb{RP}^{2n}$ follows from this.

§ 8. Chebyshev polynomials.

Consider monic polynomials $x^n + a_1 x^{n-1} + \cdots + a_n$ of fixed degree. Consider some interval, for example, $[-2, 2]$. We are interested in the smallest possible deviation from 0 on this segment.

Definition. Given a polynomial we consider its maximum M and its minimum m on this segment. Then the *deviation* is the larger of the two, $\max\{|M|, |m|\}$.

Example. Let's consider degree 1. The polynomial is then $x + c$. It's clear that if we want the derivation to be as small as possible we should take the line through the origin, $f(x) = x$. The deviation is then $D = 2$.

Now take $n = 2$. Then $x^2 + px + q$ is a parabola pointing upwards. The best choice will be a curve symmetric about the y-axis with y-intercept -2. We get $f(x) = x^2 - 2$ and in this case the deviation is also $D = 2$.

For $n = 3$ we can experiment a little bit to see that the best curve is $x^3 - 3x$.

Consider the rectangular area bounded by $x = -2$, $x = 2$, $y = -d$, $y = d$. Suppose there is a polynomial such that in the bounded area it takes the minimum and maximum alternatingly $n + 1$ times. Then we can prove that this is the "best" polynomial, with minimal deviation. We can divide the box by vertical segments at the maximums and minimums. Suppose there is another polynomial with derivation less than or equal to the derivation of that of this polynomial. Then consider its graph. It must intersect our polynomial in each of the little boxes. We will have n intersection points. But these are harmonic polynomials and their difference has degree $n - 1$. Thus their difference must be identically zero, and we are done. But there is a small inaccuracy. What if the intersection point lies at the boundary of two little boxes? This can be fixed easily, we leave it as an exercise.

Note $2\cos 2\alpha = (2\cos\alpha)^2 - 2$ and $2\cos 3\alpha = (2\cos\alpha)^3 - 3(2\cos\alpha)$. This gives us the answer, given by the following lemma.

Lemma. *For every n, $2\cos n\alpha$ is a monic polynomial P_n of degree n in $2\cos\alpha$.*

Proof. We have

$$\cos(n+1)\alpha + cos(n-1)\alpha = 2\cos\alpha\cos n\alpha.$$

This implies that $2\cos(n+1)\alpha = 4\cos\alpha\cos n\alpha - 2\cos(n-1)\alpha$. Now we can proceed by induction on n. We have $P_{n+1}(x) = xP_n(x) - P_{n-1}(x)$, where $x = 2\cos\alpha$. □

Now we claim that these polynomials are the solutions to our problem. When α ranges between 0 and π, x covers each value in $[-2, 2]$ exactly once and $P_n(x)$ covers the interval n times. This is precisely the property we want.

What about other intervals? This is an exercise. Find the least deviation from zero for a monic polynomial of degree n on an interval $[a, b]$. You need to change variables to go to this new interval.

Chebyshev polynomials also give a solution to a completely different problem. Suppose you have two monic polynomials $P(x)$ and $Q(x)$. Is it possible that $P(Q(x)) = Q(P(x))$? There is a complete solution to this problem. We can take a polynomial $R(x)$ and let $P(x) = \underbrace{R\circ R\circ\cdots\circ R}_{n}$ and $Q(x) = \underbrace{R\circ R\circ\cdots\circ R}_{m}$. This is one infinite set of solutions. A more interesting set is the Chebyshev polynomials.

Theorem (closed form of P_n)**.**

$$P_n(x) = x^n - \frac{n}{n-1}\binom{n-1}{1}x^{n-2} + \frac{n}{n-2}\binom{n-2}{2}x^{n-4} - \cdots = \sum_{j=0}^{\lfloor n/2\rfloor}(-1)^j\frac{n}{n-j}\binom{n-j}{j}x^{n-2j}.$$

Consider the series

$$F(x,z) = 2 + zx + z^2(x^2-2) + \cdots + z^nP_n(x) + \cdots.$$

This is a generating function for the sequence of Chebyshev polynomials. The recurrence relation for P_n can be rewritten as an identity in F. We can write

$$F(x,z) = 2+zx+z^2(xP_1-P_0)+z^3(xP_2-P_1)+\cdots = 2+xz+xz(zP_1+z^2P_2+\cdots)-z^2(P_0+zP_1+\cdots),$$

or

$$F(x,z) = 2 + xz + xz(F(x,z) - 2) - z^2(F(x,z)) = 2 - xz + (xz - z^2)F(x,z)$$

and

$$F(x,z) = \frac{2 - xz}{1 - xz + z^2}.$$

Now when we expand this series the coefficient of z^n will be P_n.

$$F(x,z) = (2 - xz)\left(1 + (xz - z^2) + (xz - z^2)^2 + \cdots\right).$$

By the binomial formula

$$(xz - z^2)^k = x^k z^k - kx^{k-1}z^{k+1} + \cdots = \sum_{j=0}^{k} (-1)^j \binom{k}{j} x^{k-j} z^{k+j}.$$

We are interested in the coefficient of z^n. We therefore set $k + j = n$ and plug in $j = n - k$ to get the coefficient. When we do this we get the closed form of P_n.

Another approach is given by continued fractions.

$$\frac{P_n(x)}{P_{n-1}(x)} = x - \cfrac{1}{x - \cfrac{1}{x - \cfrac{1}{x - \ddots \cfrac{1}{x - \frac{2}{x}}}}}.$$

A third approach involves large determinants. Let

$$D_n(x) := \begin{bmatrix} x & 1 & & & \\ 1 & x & 1 & & \\ 0 & 1 & x & 1 & \\ & & & & \ddots \end{bmatrix}$$

Then we have

$$P_n(x) = D_n(x) - D_{n-2}(x).$$

§ 9. Roots of trigonometric polynomials.

Consider an arbitrary trigonometric polynomial of the form

$$f(\alpha) = c + a_1 \cos\alpha + b_1 \sin\alpha + a_2 \cos 2\alpha + b_2 \sin 2\alpha + \cdots + a_n \cos n\alpha + b_n \sin n\alpha.$$

The product of two trigonometric polynomials of degree m and n is a trigonometric polynomial of degree $m + n$ (exercise).

We consider two questions: (1) what is the upper bound on the number of roots of the polynomial (between 0 and 2π)?, and (2) what is the lower bound on the number of roots (Sturm-Hurwitz theorem)?

Upper bound on number of roots.

Theorem. *A trigonometric polynomial has at most $2n$ roots.*

Proof 1. Given a trigonometric polynomial of degree n, $f(\alpha)$, we construct a complex polynomial $F(z)$ of degree $2n$ such that the restriction to S^1 of $F(z)/z^n$ is $f(\alpha)$. Let $F(z) = u_0 + \cdots + u_{2n}z^{2n}$. Then we want to restrict

$$\frac{u_0}{z^n} + \frac{u_1}{z^{n-1}} + \cdots + u_{2n}z^n$$

to S^1. We make use of the formulas

$$\cos\phi = \frac{1}{2}\left(e^{i\phi} + e^{-i\phi}\right), \qquad \sin\phi = \frac{1}{2i}\left(e^{i\phi} - e^{-i\phi}\right).$$

On S^1 we have $r = 1$ and we can write $z = e^{i\phi}$, and these become

$$\cos\phi = \frac{z + z^{-1}}{2}, \qquad \sin\phi = \frac{z - z^{-1}}{2i}.$$

Take $\phi = k\alpha$. □

Proof 2. We construct a complex polynomial $G(z)$ of degree n such that $f(\alpha)$ is the restriction of $\operatorname{Re} G(z)$ to S^1. We write $G(z)$ as $(u + iv)(\cos k\alpha + i\sin k\alpha)$ and we solve

$$\operatorname{Re}(u + iv)(\cos k\alpha + i\sin k\alpha) = \cos k\alpha$$

to see $u = 1$, $v = 0$. Now we are talking about the intersection of two curves. The first is given by $\operatorname{Re} G(z) = 0$ which has degree n and the second is the unit circle $x^2 + y^2 = 1$ which is also an algebraic curve which has degree 2. There is a fundamental result in algebraic geometry, Bezout's thereom, that states that two algebraic curves of degree m and n intersect at no more than mn points. □

Periodic functions. Periodic functions can be written in the same form as trigonometric polynomials but with infinitely many terms. The *first harmonic* refers to the first nonzero term. The number of roots of a periodic function is not less than the number of roots of its first harmonic.

Theorem (Sturm-Hurwitz)**.** *If $f(\alpha) = a_k\cos k\alpha + b_k\sin k\alpha + \cdots$ with at least one of a_k, b_k nonzero, then f has at least $2k$ roots.*

Example. Suppose we start with $\cos 7\alpha$. Then by adding terms like $100\sin 9\alpha$, $-99\sin 99\alpha$, ..., we can add roots but we can not make the number of roots smaller.

Proof 1, by Rolle's theorem. Recall that Rolle's theorem tells us that between any two roots of a differentiable function there must lie a root of the derivative. We can reformulate this in terms of the number of sign changes $Z(f)$ of f on the unit circle.

Theorem (Rolle, reformulation)**.**

$$Z(f') \geq Z(f).$$

Now not every periodic function is the derivative of some other periodic function. It needs to have zero average.

Lemma. *Given a periodic function $f(\alpha)$ there exists a periodic function g such that $f = g'$ if and only if $\int_0^{2\pi} f d\alpha = 0$.*

The proof is left as an exercise.

Let D represent differentiation and D^{-1} integration. We see that D^{-1} is a well-defined operation on the space of functions with zero average. Thus we can reformulate Rolle's theorem again as $Z(f) \geq Z(D^{-1}f)$.

$$D^{-2}(\sin k\alpha) = -\frac{1}{k^2}\sin k\alpha.$$

If we repeat this operation sufficiently many times this will "eat" the "tail" of our polynomial. Even better we can consider

$$-k^2 \cdot D^{-2}(\sin k\alpha) = \sin k\alpha.$$

So

$$Z(f) \geq Z(D^{-1}f) \geq Z(D^{-2}f) \geq \cdots \geq Z(D^{-i}f) \geq \cdots$$

Then we can look at the graph for sufficiently large i. The number of roots is then $2k$ and it is clear that a small perturbation of the graph cannot change the number of roots. □

Proof 2, by heat equation. This proof is inspired by mathematical physics.

Consider $f(\alpha)$ as the distribution of heat over a circle. Then we look at how the heat propagates over time.

We have a partial differential equation $\frac{\delta F}{\delta t} = \frac{\delta^2 F}{\delta \alpha^2}$ where $F(\alpha, t)$ is such that $f(\alpha) = F(\alpha, 0)$. This is the heat equation. We want to say something about the number of zeros over time, and we want to see what happens when t is very large.

The first claim is that the number of zeros with time can only decrease. It's obvious physically. Essentially what we are saying is that propagation of heat is an average process. If you have a warm ocean and an iceberg, then the iceberg will eventually melt, i.e. two zeros will disappear. Mathematically this claim is the maximal principle and is not hard to prove.

Now what happens as t goes to infinity? This equation can be solved because you can solve it for every pure harmonic. The solution is

$$F(\alpha, t) = \sum_{j \geq k}^{N} e^{-j^2 t}(a_j \cos j\alpha + b_j \sin j\alpha).$$

For $t = 0$ we end up with $f(\alpha)$. $\square$

Proof 3, by integration. We define an inner product $\langle f, g\rangle := \int_0^{2\pi} fg d\alpha$. Given $f(\alpha) = a_k \cos k\alpha + b_k \sin k\alpha + \cdots$, suppose $Z(f) < 2k$ (i.e. $Z(f) \leq 2k - 2$). We have a collection of points $\alpha_1, \ldots, \alpha_{2k-2}$ on the circle, roots of our polynomial. We want to construct a trigonometric polynomial $g(\alpha)$ of degree $k - 1$ such that $g(\alpha_j) = 0$. Then α_j must be the full collection of its roots. Suppose such a g exists. Then we have a function f and a function g. Looking at their graphs, we see they must have the same intervals of constant sign. Their product therefore is almost always positive, excluding zeros. Thus the integral $\int_0^{2\pi} fg d\alpha > 0$. On the other hand they must be orthogonal with respect to the inner product, because the first term of f is k while the last term of g is $k - 1$. $\square$

Example. Suppose you have two points α_1, α_2 between 0 and 2π. Can we construct a linear harmonic that changes sign at these two points? Suppose it is $a \cos \alpha + b \sin \alpha + c$. Then we have the two conditions

$$a \cos \alpha_1 + b \sin \alpha_1 + c = 0$$
$$a \cos \alpha_2 + b \sin \alpha_2 + c = 0$$

The answer is

$$g(\alpha) = \sin \frac{\alpha - \alpha_1}{2} \sin \frac{\alpha - \alpha_2}{2} \cdots \sin \frac{\alpha - \alpha_{2k-2}}{2}.$$

Exercise: prove this is a trigonometric polynomial (remember the formula for the product $\sin \alpha \sin \beta$), and show that it satisfies our conditions.

§ 10. The four-vertex theorem.

Theorem (four-vertex)**.** *A plain oval has at least 4 vertices.*

Definition. A *vertex* is a local maximum or minimum of curvature.

Definition. We have a smooth curve γ and we define an arc-length parametrization on it. The magnitude of the acceleration vector $|\gamma''|$ is the *curvature* of γ.

Proof (S.Mukhopadhyay, 1909). The problem is how to characterize a curve. The arc-length parametrization is very good for theoretical purposes but an ellipse for example is a very simple curve for which you cannot find an arc-length parametrization using elementary functions.

Definition. Choose an origin O inside the oval. For every angle α we consider the tangent line to the curve that is perpendicular to this angle. The distance from O to the tangent line (*support line*) is a function $p(\alpha)$. This function is called the *support function.*

We can use support functions to parametrize the oval. For example, if the curve is a circle and the origin is at the center, then the support function is constant. All support lines are at a constant distance from the center. How does $p(\alpha)$ depend on the choice of the origin? What happens if we translate O to another point O'? Clearly the support function will change. In fact the new support function is given by $\overline{p}(\alpha) = p(\alpha) + a\cos\alpha + b\sin\alpha$ where $\overline{(a, b)} = \overline{OO'}$. This can be proved easily (exercise). It follows that the support functions of circles are first harmonics, $c + a\cos\alpha + b\sin\alpha$.

Definition. If you have a plain curve, at every point there is a circle, called the *osculating circle*, that approximates this point better than any other circle, by taking three infinitesimally close points to the curve. (Given three points we have a unique circle.)

Let $p(\alpha)$ be the support function for the curve and let $q(\alpha)$ be the support function for the osculating circle. At the point of tangency T we have $p(T) = q(T)$, $p'(T) = q'(T)$, and $p''(T) = q''(T)$.

Now what happens a vertex? A vertex is a point where the curvature is stationary. At a vertex the osculating circle is "more tangent" to the curve than at an ordinary point. Thus the circle shares at least 4 points with the curve, and at a point of tangency T we have also $p'''(T) = q'''(T)$.

Let's consider the following differential operator which kills linear harmonics: $f \mapsto f''' + f'$. We want to approximate the curve by a linear harmonic at 4 points. We claim this is possible iff $p'''(\alpha) + p'(\alpha)$. That is, the vertices are precisely the roots of this equation.

Now we can reformulate the problem in terms of analysis: Given a periodical and differential function $p(\alpha)$ on the circle, there are at least four different points x such that $p'''(\alpha) + p'(\alpha) = 0$. In fact this is even stronger than the original theorem because we don't impose any conditions on p. Why is this true? Let $f(\alpha) := p'''(\alpha) + p'(\alpha)$. If p is harmonic in the form $c + a_1\cos\alpha + b_1\sin\alpha + a_2\cos 2\alpha + b_2\sin 2\alpha + \cdots + a_n\cos n\alpha + b_n\sin n\alpha$, then the terms $c + a_1\cos\alpha + b_1\sin\alpha$ are killed. □

Similar results.

Theorem (Möbius)**.** *Every simple, non-intersecting, non-contractible curve has at least three inflection points.*

What is an inflection point? Given a curve at every point it can be approximated by a tangent line. This approximation is up to the first order. There may be some points on the curve where a tangent line approximation is "better" than at other points. These points are inflection points. In this sense the theorem is of the same nature.

Theorem. *Suppose you have an oval in the plane and you want to approximate it by conics instead of circles. A conic as you remember is determined by five points. At every point you can take five infinitesimally close points and that will determine your conic. This gives us an approximation of the fourth order. Again there may be points on the curve where the approximations are better. Such points are called* affine vertices *or* sextactic points*. There are at least six of them.*

This theorem was proved in the same paper by Mukhopadhyay.

Theorem (tennis ball)**.** *If you have a spherical curve which is closed, simple and doesn't intersect itself that bisects the area, then such a curve has at least four inflection points.*

An inflection point here is where a great circle approximates the curve better than usual.

Theorem (generalization of four-vertex theorem to other curves)**.** *We consider* immersed *or* embedded curves*, which allow self-intersections. We can define an equivalence relation on such curves. When a curve is equivalent to the plain oval, the 4-vertex theorem holds.*

§ 11. Trigonometric and periodic polynomials, part 2.

Consider the trigonometric polynomial $c + a_1 \cos\alpha + b_1 \sin\alpha + \cdots + a_n \cos n\alpha + b_n \sin n\alpha$. We can prove it has no more than $2n$ roots by taking the second derivative D^2. Since $a_n \cos n\alpha + b_n \sin n\alpha$ has only $2n$ roots, the result follows.

Analogy between periodic functions and curves on the sphere. Consider a simple spherical curve γ. There is another operation on the sphere: you can move every point a distance $\pi/2$ in the direction of the positive tangent. So we obtain a new curve γ' which we call the *derivative* of γ. For example, if γ is a circle of latitude, then γ' is an equator.

Now we can note some parallels with periodic functions.

1. As mentioned earlier, a periodic function is the derivative of some other if and only if its average is zero. Here the curve γ' bounds half the area of the sphere.

2. Given a periodic function we can find the inverse derivative but not uniquely: we have a family of solutions (for example, $D^{-1}(x) = \{(1/2)x^2 + t\}_{t\in\mathbb{R}}$). We have exactly the same thing on the sphere. The family of curves constructed by moving every point on γ in the orthogonal direction with constant speed all have the same derivative γ'.

Nonzero coefficients and the number of real roots. We know that a polynomial $f(x)$ of degree n has no more roots than its degree. It's a lesser known fact that what really matters as far as real roots are concerned is not the degree but the number of nonzero coefficients.

Definition. A *fewnomial* is polynomial that has only a "few" nonzero terms.

A typical fewnomial is something like $x^{100} - 1$ or $ax^n + bx^n$ with high degree but few terms.

Lemma (a fewnomial has few roots)**.** *If f has k nonzero coefficients then there are no more than $2k - 1$ real roots.*

Proof. We proceed by induction. If $k = 1$ our polynomial is just ax^n which has just one real root. Assume the claim is true for $1, \dots, k$, and we will prove it's true for $k + 1$. Write $f(x) = x^r g(x)$, where r is the biggest power of x that divides f. We have at most one root from x^r and the rest from $g(x)$. Now consider $g'(x)$ which has k terms. By the inductive hypothesis it has no more $2k - 1$ real roots. By Rolle's theorem between every two roots of g we have at most one root of g'. Thus it follows that g has no more than $2k$ roots, and f has no more than $2k + 1$, as desired. □

For example, $x(x^2 - 1) \cdots (x^2 - k^2)$ has $k + 1$ terms and $2k + 1$ roots. Thus we have a strong bound.

Bound on the number of positive real roots.

Theorem (Descartes' rule)**.** *The number of positive roots of a real polynomial is less than or equal to the number of sign changes in the sequence of coefficients.*

Proof. We need to analyse what happens when a polynomial acquires a new root. Suppose you have a polynomial $g(x)$ and we build a new polynomial $f(x) := (x - b)g(x)$ that has the additional root b. Now we consider the number of sign changes in the sequence of coefficients in g and f. Write $g(x) = a_0x^n + a_1x^{n-1} + \cdots + a_n$. Then the coefficients of f

are $a_0, a_1 - ba_0, a_2 - ba_2, \ldots, a_n - ba_{n-1}, -ba_n$. It is easy to see that the leftmost sign in each cluster reappears in the new sequence. In the very beginning a_0 reappears so we have the same sign. In the end we have $-ba_n$ so we have opposite sign. So every leftmost sign repeats while the last one is opposite. Thus a new root gives at least one new sign change.

To finish the proof take $f(x) = (x - b_1) \cdots (x - b_k)g(x)$ where b_i are all the positive roots of f and g has only negative roots. Then the number of sign changes is at least k and we are done. □

Sturm's method. Surprisingly we can calculate the number of real roots a real polynomial has on any interval. We fix an interval $[a, b]$. Given a polynomial $f(x)$ (we assume f has no multiple roots and does not vanish on a and b) we will construct a sequence of polynomials $p_0(x), \ldots, p_n(x)$ of decreasing degrees with the following properties:

(a) $p_0(x) = f(x)$ and $p_1(x) = f'(x)$.

(b) If $p_k(t) = 0$ for some $t \in \mathbb{R}$ then $p_{k-1}(t)$ and $p_{k+1}(t)$ are both nonzero and have opposite signs.

(c) $p_n(x)$ has no roots.

Such sequences are called *Sturm sequences*. For every $x \in [a, b]$ we consider the number $S(x)$ of sign changes in the sequence $p_0(x), \ldots, p_n(x)$.

Theorem (Sturm)**.** *The number of roots of $f(x)$ on $[a, b]$ is $S(a) - S(b)$.*

Proof. We need to analyze what happens to the number of sign changes as x goes from a to b. This number is usually the same, i.e. it is locally constant. It changes only when x is a root of one of p_i.

Suppose x is a root of $p_0 = f$, so that $f(x) = 0$. Then there are two pictures to consider. Either the sign changes from negative to positive or it changes from positive to negative. In the first case the derivative $f'(x) = p_1(x)$ is positive and in the second it is negative. If $p_1(x) > 0$, then we have the signs $-, +, \cdots$ initially and then $+, +, \cdots$. That is, $S(x + \varepsilon) - S(x - \varepsilon) = 1$. Similarly in the second case, we have $+, -, \cdots$ and then $-, -, \cdots$, so that $S(x + \varepsilon) - S(x - \varepsilon) = -1$.

If $p_k(x) = 0$ and $0 < k < n$ our assumption is $p_{k-1}(x)$ and $p_{k+1}(x)$ have opposite sign. Suppose the graph of $p_k(x)$ changes from positive to negative. Then we will have $\cdots, -, +, +, \cdots$ and after passing through the point x it becomes $\cdots, -, -, +, \cdots$. Thus the number of sign changes doesn't change. In the second case, where $p_k(x)$ changes from negative to positive, we have $\cdots, +, +, -, \cdots$ and then $\cdots, +, -, -, \cdots$ and we draw the same conclusion.

Now how do we construct p_i? Given p_{k-1} and p_k, to construct p_{k+1} we divide p_{k-1} by p_k and take the opposite of the remainder, so that $p_{k-1} = qp_k - p_{k+1}$. We have decreasing degrees as desired. Now p_n will be the greatest common divisor of f and f'. But if it has a root, then f and f' have a common root, so f has a multiple root. This contradicts our assumption. Also no two neighbors p_{k-1} and p_k may both vanish because then p_{k+1} vanishes as well, and continuing we see that p_n vanishes. But this is impossible. So we are done. □

Example. Take the polynomial $f(x) := x^5 - x + a$ for some a. We get $f'(x) = 5x^4 - 1$ and we can let $p_1(x) = x^4 - (1/5)$ since we are concerned only with sign changes. We compute $p_2(x) = x - (5/4)a$ and $p_3(x) = (1/5) - (5/4)^4 a^4$.

We can use Sturm's method to find that on the interval $(-\infty, \infty)$ there is 1 root when the last term is negative and 3 when the last term is positive.

Exercises.

1. Suppose we have a smooth function f and we know its graph $y = f(x)$ intersects the graph of a polynomial of degree $n - 1$ in at least $n + 1$ points. Prove that the nth derivative of this function has a root.

2. Prove that $1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots + \frac{1}{n!}x^n$ has one root if n is odd and none otherwise.

3. (Fourier-Budan Theorem.) f is a polynomial of degree n and $S(x)$ is the number of sign changes in the sequence $(f(x), f'(x), f''(x), \cdots)$. Let $[a, b]$ be an interval such that f does not vanish at a or b. The number of roots on $[a, b]$ is no more than $S(a) - S(b)$.

4. (Hyperbolic polynomials.)

   Theorem (Gauss-Lucas theorem)**.** *Let $f(z)$ be a complex polynomial without multiple roots. The roots of f' lie in the convex hull of the roots of f.*

   The *convex hull* of a finite collection of points in the plane is the smallest convex set containing them.

   Proof. Let $f(z) = (z - z_1)\cdots(z - z_n)$. We have

   $$\frac{f'(z)}{f(z)} = \frac{1}{z - z_1} + \cdots + \frac{1}{z - z_n}.$$

   If $f'(w) = 0$ then the above sum must vanish at w. The convex hull is the intersection of all half-planes containing the set. $1/z = \overline{z}/|z|^2$. □

A real polynomial f of degree n is called *hyperbolic* if it has n distinct real roots. If f is hyperbolic then so is f', by Rolle's theorem. Suppose f has roots $x_1 < x_2 < \cdots < x_n$ and f' has roots $z_1 < z_2 < \cdots < z_{n-1}$. What happens if we deform f by increasing its roots by a small amount? Then the derivative still has $n-1$ roots and it turns out that they will also move in the same direction.

Lemma. *Let $\overline{f}(x)$ be another polynomial with roots $\overline{x}_1 < \cdots < \overline{x}_n$ and let the roots of $\overline{f}'(x)$ be $\overline{z}_1 < \cdots < \overline{z}_n$. Then $x_i \leq \overline{x}_i$ for all i then $z_i \leq \overline{z}_i$ for all i.*

Proof. Assume $\overline{z}_k < z_k$ for some k. This implies $\overline{z}_k - \overline{x}_i < \overline{z}_k - \overline{x}_i$. Now $\overline{x}_k < \overline{z}_k < \overline{x}_{k+1}$ and $x_k < z_k < x_{k+1}$. This means that the signs of both $\overline{z}_k - \overline{x}_i$ and $\overline{z}_k - \overline{x}_i$ are the same. Then their reciprocals satisfy $\frac{1}{\overline{z}_k - \overline{x}_i} > \frac{1}{\overline{z}_k - \overline{x}_i}$. Now when we sum over i we arrive at a contradiction since we know

$$\sum_{i=1}^{n} \frac{1}{z_k - x_i} = 0, \qquad \sum_{i=1}^{n} \frac{1}{\overline{z}_k - \overline{x}_i} = 0$$

for each k. □

Now we have the following open problem. Suppose you have a hyperbolic polynomial of degree n. Consider the sequence $f, f', f'', \ldots$ which are all hyperbolic. Are there any forced relations between the set of combined roots of these polynomials? For example, suppose f has roots $a_0 < a_1 < a_2 < a_3$, f' has roots $b_0 < b_1 < b_2$, f'' has roots $c_1 < c_2$ and f''' has root d. Then one such relation is the following: if $a_1 < c_0$ and $a_2 < c_1$, then $b_1 < d$. The proof is an exercise.

We end with a remark related to the swallow-tail curve we studied in the beginning.

Braid group. A *braid group* is made of "streams" which go from top to bottom. The operation in this group is concatenation, the identity element is the trivial braid and the inversion is changing every over-crossing to an under-crossing. If you intersect a braid by a horizontal plane you get n points. If you move the plane from top to bottom you see a one-parameter family of polynomials. This is the fundamental group of the component to the swallow-tail.