137 NOTES, PART 1: POLYNOMIALS IN ONE VARIABLE AND THEIR ZEROS

ZIV RAN

Contents

0. prelude

What is the subject of of Plane Curves all about? The answer is surprisingly involved and will take a while to develop, but to fix ideas let's start with a provisional definition of our objects of study.

Definition 0.1. *A real plane curve is a set of the form*

$$C = \{(x, y) \in \mathbb{R}^2 : f(x, y) = 0\}$$

where $f(x, y)$ is a polynomial. C is said to be the 'zero-set' of f.

Having said that, the obvious question is: what sorts of questions do we study about these curves ? An obvious answer might be: 'determine C explicitly' or perhaps 'try to draw nice sketches of them, like in Calculus texts'. Unfortunately, neither of these turns out to be possible, except for some very special polynomials. So the answer as to what the 'right' questions are to study about curves is not so obvious. To get a better idea what it might be, one may try to travel back in time a bit. In an earlier era, high-school students used to spend much time studying what was called *Analytic Geometry*. A highlight of this study was the so-called *Classification Theorem of Real conics*, one version of which would run as follows.

Theorem 0.2. *Let $f(x,y)$ be a real quadratic polynomial with zero-set*

$$C = \{(x,y) : f(x,y) = 0\}$$

(called a conic). Then there is a substitution of the form

$$x = a_{11}x^* + a_{12}y^* + b_1$$
$$y = a_{21}x^* + a_{22}y^* + b_2$$

with $a_{11}a_{22} - a_{12}a_{21} \neq 0$, and a constant $c \neq 0$, such that if $g(x^*, y^*) = cf(x,y)$ denotes the resulting polynomial, then g is one, and only one (depending on f) of the following (with zero-set in parentheses)

- *(nondegenerate type)*
  - *(hyperbola)* $(x^*)^2 - (y^*)^2 - 1$
  - *(parabola)* $x^* - (y^*)^2$
  - *(circle)* $(x^*)^2 + (y^*)^2 - 1$
- *(degenerate type)*
  - *(line-pair)* $(x^*)^2 - (y^*)^2$
  - *(double line)* $(x^*)^2$
  - *(point)* $(x^*)^2 + (y^*)^2$
- *(empty)* $(x^*)^2 + (y^*)^2 + 1$ □

The *proof* of this theorem as given in old texts is perhaps not particularly interesting and will not be discussed at this time (some parts of it are subsumed in results we shall develop later). But the *statement* nonetheless is interesting and may serve as a prototype for what we call a *Classification Theorem* in Algebraic Geometry, and encapsulates some of the important themes for our course and for the subject as a whole.

- Under study is a class of *objects* with both an algebraic and a geometric aspect (viz. polynomials of degree 2 and their zero-sets). This is essentially, but not exactly, a subclass of the class of plane curves as defined previously; the algebraic data of the polynomial are included as well as the curve.
- We set up a notion of *equivalence* between objects, which tells us when two of them are are to be considered 'essentially the same' (viz. when one polynomial can be transformed to the other by a substitution as above).
- Finally, we state a *classification* of our class of objects, which means that we provide a particular list of objects (preferably finite – in our case containing just 7 objects), and state that any object in our class is equivalent to one (and only one) object on the list.

- As a broad theme, we like to talk- and draw- Geometry; but when it comes to precise statements and calculations, Algebra seems to be the way to go. What is the precise relation between the Algebra and Geometry ? e.g. does a (nondegenerate) conic have a uniquely determined equation ?

This type of results and questions are a prototype and goal of much of what one is trying to do in Algebraic Geometry and in this course. The classes of objects considered usually are rather more complex than conics, so considerable work is required before one can get anywhere near a classification; certainly, there will be more than enough work to keep us busy this quarter.

As a warmup for our discussion of polynomials in 2 variables, we start with the 1-variable case.

1. Inhomogeneous polynomials and the affine line

We can talk about polynomials with coefficients that are rational, real, or complex. To save some writing, let's introduce an ad-hoc definition.

Definition 1.1. *A concrete field* $\mathbb{F}$ *is either the rational numbers* $\mathbb{Q}$*, the real numbers* $\mathbb{R}$*, or the complex numbers* $\mathbb{C}$*.*

What's important about a concrete field $\mathbb{F}$ is that the usual arithmetic operations (addition, subtraction, multiplication and division) make sense for its elements (as long as we don't divide by zero, of course!). Though we don't need this here, there is an abstract algebraic notion of 'field' of which these are all examples; another example is the 'binary field' $\mathbb{F}_2 = \{0, 1\}$, indeed polynomials with $\mathbb{F}_2$ coefficients have many important applications, e.g. in coding theory. The natural numbers $\mathbb{N}$ or the integers $\mathbb{Z}$ are not a field because, e.g. we cannot always divide two integers and get another. $\mathbb{Z}$ is an example of a number system called a *ring*; later, polynomials with ring coefficients will play an important role in what we do, but that is a rather more involved story than the case of field coefficients, to which we'll stick for now.

Another important fact about concrete fields (and fields in general) $\mathbb{F}$ is that the basic theory of linear algebra, including vector spaces, bases, dimension, matrices, linear transformations ... makes sense taking scalars in $\mathbb{F}$ (to specify that a given set V is a vector space with scalars in $\mathbb{F}$ we call V a vector space *over* $\mathbb{F}$). This is because all this linear algebra only uses the formal properties of the arithmetical operations.

Here right away is an important example of a vector space. Denote by $\mathbb{F}[x]$ the set of polynomials with coefficients in $\mathbb{F}$, i.e. formal expressions of the form

$$f(x) = a_n x^n + ... + a_1 x + a_0, a_0, ..., a_n \in \mathbb{F}, n \in \mathbb{N}$$

If $a_n \neq 0$, we call it the *leading coefficient* of f and call n the *degree* of f. Then $\mathbb{F}[x]$ is a vector space over $\mathbb{F}$. It has a basis of the form $1, x, ...$

This basis is infinite, and that might be problematic. So we are led to look at finite-dimensional subspaces of $\mathbb{F}[x]$. Let $\mathbb{F}[x]_n$ denote the set of polynomials of degree n or less, i.e. all expressions as in the above display, for fixed $n \in \mathbb{N}$ (whether $a_n \neq 0$ or not). $\mathbb{F}[x]_n$ forms a vector space with ('standard') basis $1, x, x^2, ..., x^n$; therefore it has dimension $n+1$. Besides the standard basis, $\mathbb{F}[x]_n$ has many others, for example $1, x-a, (x-a)^2, ..., (x-a)^n$, for any $a \in \mathbb{F}$.

Exercise 1.1. *Prove that* $1, x-a, (x-a)^2, ..., (x-a)^n$*, for any* $a \in \mathbb{F}$*, is a basis of* $\mathbb{F}[x]_n$*.*

Our objective is to study the algebra/geometry interplay. The 'geometry' of a polynomial is, essentially, its set of zeros, so the geometric counterpart of $\mathbb{F}[x]$ is $\mathbb{F}$. More precisely, it is conceptually important to take a copy of $\mathbb{F}$ and denote it by $\mathbb{A}^1_{\mathbb{F}}$. We call this the *affine line* over $\mathbb{F}$, and think of it as a purely geometric object (unlike $\mathbb{F}$ which as its arithmetic operations). The zeros of a polynomial will consist of a finite set of points (the zeros), each with a well-defined multiplicity. To formalize this, consider a formal symbol $[p]$ for each $p \in \mathbb{A}^1_{\mathbb{F}}$ and let $Z(\mathbb{A}^1_{\mathbb{F}})$ be the set of formal sums of the form

$$z = \sum_{i=1}^{m} n_i [p_i]$$

with $n_i \in \mathbb{Z}$ (positive or negative). Such a sum z is called a *cycle*; its *degree* is by definition $\sum n_i$. If all $n_i \geq 0$, z is said to be a *positive* or *effective* cycle and the set of all of these is denoted by $Z(\mathbb{A}^1_{\mathbb{F}})_+$. The we can define an important mapping called the *cycle map*

$$Z = Z_{\mathbb{F}} : \mathbb{F}[x] \setminus \{0\} \to Z(\mathbb{A}^1_{\mathbb{F}})_+$$

by the rule that for any nonzero polynomial $f(x)$,

$$Z_{\mathbb{F}}(f) = \sum_{i=1}^{m} n_i [p_i]$$

where $p_1, ..., p_m$ are the zeros of f in $\mathbb{F}$ and for each i, n_i is the multiplicity of p_i.

One nice property of $Z_{\mathbb{F}}$ is its compatibility with multiplication of polynomials:

$$Z_{\mathbb{F}}(fg) = Z(f) + Z(g). \tag{1}$$

Example 1.2.
- $Z_{\mathbb{Q}}(x^2 - 4) = [2] + [-2]$;
- $Z_{\mathbb{Q}}(x^3 - x^2) = 2[0] + [1]$,
- $Z_{\mathbb{Q}}(x^2 + 4) = 0; Z_{\mathbb{C}}(x^2 + 4) = [2i] + [-2i]$ *(where* $i = \sqrt{-1}$*.)*

This last example shows $Z_{\mathbb{F}}$ depends on the field $\mathbb{F}$! It also shows that, in general, we cannot recover f from its zero-cycle $Z_{\mathbb{F}}(f)$.
We cannot, that is, *unless* we insist on taking $\mathbb{F} = \mathbb{C}$; that is due to the following important classical result

Theorem 1.3. *(Fundamental Theorem of Algebra) Every nonconstant in* $\mathbb{C}[x]$ *has at least one root in* $\mathbb{C}$*.*

Exercise 1.2. Project: *Look up an elementary proof of this theorem somewhere (internet, the book 'What is Mathematics by Courant/Robbins or...) and present it.*

Another, stronger way to formulate this is:

Theorem 1.4. *Every polynomial* $f \in \mathbb{C}[x]$ *can be expressed in the form*

$$f(x) = c\prod_{i=1}^{m}(x - b_i)^{n_i}, c, b_i \in \mathbb{C} \tag{2}$$

Note that if f is nonzero of degree n, then $a \neq 0$ and $\sum n_i = n$. This implies that $Z(f)$ is a cycle of degree n (same as f), and that f *can be recovered from* $Z(f)$ (up to constant multiple, of course). In fact, we can set up an inverse mapping to Z,

$$W : Z(\mathbb{A}^1_{\mathbb{C}})_+ \to \mathbb{C}[x]$$

defined by

$$W(\sum n_i[b_i]) = \prod(x - b_i)^{n_i}.$$

Then it is immediate (in fact, for any $\mathbb{F}$) that

$$Z \circ W(\sum n_i[b_i]) = \sum n_i[b_i]$$

and that

$$W \circ Z(c\prod_{i=1}^{m}(x - b_i)^{n_i}) = \prod_{i=1}^{m}(x - b_i)^{n_i}$$

In particular, when $F = \mathbb{C}$, any polynomial f factors as in (2), and therefore W and Z together yield a 1-1 correspondence between, on the one hand, nonzero polynomials, up to scalar multiple (or what is

the same *monic* polynomials), and on the other hand, cycles. Thus, we have essentially a complete mirror correspondence between the algebra of polynomials and the geometry of cycles– provided we work over $\mathbb{C}$. This correspondence does not work over $\mathbb{Q}$ or $\mathbb{R}$, which provides a major motivation for choosing $\mathbb{C}$ for scalars.

Exercise 1.3. *Show that the $Z-W$ correspondence can be extended to a correspondence between rational functions and (not necessarily positive) cycles by defining, for f, g polynomials*

$$Z(f/g) = Z(f) - Z(g);$$

$$W(\sum_{i=1}^{m} n_i[b_i] - \sum_{i=1}^{m'} n_i'[b_i']) = \frac{\prod_{i=1}^{m}(x-b_i)^{n_i}}{\prod_{i=1}^{m'}(x-b_i')^{n_i'}}, n_i, n_i' \geq 0.$$

Show that Z, W are well-defined, mutually inverse maps between $Z(\mathbb{A}^1_{\mathbb{C}})$ and the set of nonzero complex rational functions up to constant factor. Moreover, for any rational functions F, G, we have

$$Z(FG) = Z(F) + Z(G).$$

2. Affine substitutions

There is much more structure to $\mathbb{F}[x]$ than just vector space: e.g. polynomials can be multiplied, not only added, and this gives $\mathbb{F}[x]$ a structure known in abstract algebra as *ring* - more on this later. We now discuss another important structural element, motivated by our brief discussion of the classification of conics, which has to do with substitutions. Starting on the 'geometric side', let us consider two copies of $\mathbb{A}^1_{\mathbb{F}}$ with respective coordinates x, x', and denote them by $\mathbb{A}^1_x, \mathbb{A}^1_{x'}$. For any pair of scalars $a, b \in \mathbb{F}$, we can define a mapping

$$T = T_{(a,b)} : \mathbb{A}^1_x \to \mathbb{A}^1_{x'}$$

$$x' = T(x) = ax + b.$$

Of course, the same formula also defines a mapping from $\mathbb{A}^1_x$ to itself. Whenever $a \neq 0$, T admits an inverse T^{-1} given by $x = x'/a - b/a$, i.e.

$$(T_{a,b})^{-1} = T_{1/a,-b/a}.$$

Such a mapping $T_{a,b}$ (with $a \neq 0$) is known as an *affine transformation* (of $\mathbb{A}^1$, known as the *affine line*). We have just seen that the inverse of an affine transformation is affine; it's also true that the composition of two affine transformations is affine.

Exercise 2.1. *Prove this.*

On the algebra side, T corresponds to a mapping going *the other way*

$$T^* : \mathbb{F}[x'] \to \mathbb{F}[x],$$
$$T^*(f(x')) = f(ax+b) (= f(T(x))).$$

Again, the same formula also defines a mapping from $\mathbb{F}[x]$ to itself. The assignment $T \mapsto T^*$ is what's called a *contravariant* operation, which means that for a pair of affine transformations T_1, T_2, we have

$$(T_2 \circ T_1)^* = T_1^* \circ T_2^*$$

(recall that $T_2 \circ T_1$ means apply T_1 first, then T_2).

Exercise 2.2. *Prove this.*

Another simple property is that, for any two polynomials f, g, we have

$$T^*(fg) = T^*(f)T^*(g).$$

Now two polynomials f, g are said to be *affine equivalent* if there is an affine transformation T and a nonzero constant $c \in \mathbb{F}$ such that

$$f = cT^*(g).$$

Exercise 2.3. *(i) Prove that for any $\mathbb{F}$, any polynomial in $\mathbb{F}[x]$ of degree 1 is affine equivalent to $g(x) = x$.*

(ii) Prove that if $\mathbb{F} = \mathbb{C}$, any $f \in \mathbb{F}[x]$ of degree 2 is affine equivalent either to $x^2 - x$ or to x^2 but not to both.

(iii) Prove that if $\mathbb{F} = \mathbb{R}$, any $f \in \mathbb{F}[x]$ of degree 2 is affine equivalent to precisely one of $x^2 - x$, $x^2 + 1$ or x^2 (hint: complete the square...)

How does affine equivalence relate to zeros ? Note that if $f = cT^*(g)$ and $[p] \in \mathbb{A}^1_{\mathbb{F}}$ is a zero of f, then $f(p) = cg(T(p)) = 0$, therefore $[T(p)]$ is a zero of g. For example, if $g = x - u$ then

$$f := T^*(g) = ax + b - u = a(x - (\frac{1}{a}u - b/a))$$

so

$$p := \frac{1}{a}u - b/a = T^{-1}(u)$$

is the unique zero of f and indeed $T(p) = u$ is the unique zero of g.

Now for a general polynomial g and its zero u, we can write

$$g = (x-u)^n g_1,$$

where n is the multiplicity of u and $g_1(u) \neq 0$. Then for $f = cT^*(g)$ we can write

$$f = (x-p)^n f_1$$

where $f_1 = cT^*(g_1)$, so $f_1(p) \neq 0$. Therefore, the multiplicity of p as zero of f equals that of u as zero of g. It follows that the cycles of f and g are nicely related: if

$$Z_{\mathbb{F}}(f) = \sum n_i[p_i]$$

then

$$Z_{\mathbb{F}}(g) = \sum n_i[T(p_i)].$$

To encode this kind of relation on cycles we may may define an 'action' of an affine tranformation T on cycles by

$$T(\sum n_i[p_i]) = \sum n_i[T(p_i)].$$

Then two cycles $\sum n_i[p_i], \sum m_j[q_j]$ are said to be affine equivalent if there is an affine transformation T such that

$$\sum m_j[q_j] = T(\sum n_i[p_i]).$$

So we get the nice statement that that *two affine equivalent polynomials have affine equivalent cycles.*

Exercise 2.4. *Prove that if $\mathbb{F} = \mathbb{C}$, then two polynomials f, g are affine equivalent if and only if their cycles are.* (Hint: 'only if' has been shown above; for 'if', suppose f, g are polynomials such that $T(Z_{\mathbb{C}}(f)) = Z_{\mathbb{C}}(g)$. Then f and $T^*(g)$ have the same cycle, therefore are proportional by the $Z - W$ correspondence in §1)

3. Compactification: the projective line

Everyone with visual experience is familiar with the situation of a point on a line 'moving out to infinity'. As mathematicians, we would like to shrug off the metaphysical or mythical qualities of 'infinity' and make it part of our framework. As a hint for how to do so, consider in the affine plane $\mathbb{F}^2$ the fixed vertical line H given by $x = 1$, which may be identified in an obvious way with $\mathbb{A}^1_{\mathbb{F}}$. For any point $P_c = (1, c) \in H$, we have a unique line through the origin containing it, that is the line L_a with equation $y = cx$ (or, if $c \neq 0$, $x = y/c$). This line is always non-vertical, and conversely any non-vertical line through the origin has this form. So we can identify $\mathbb{A}^1_{\mathbb{F}}$ with the set of non-vertical lines through the origin in the plane. As P_a moves 'out toward infinity' a goes to ∞ and our line L_a simply approaches the vertical line $x = 0$. So if we want to incorporate infinity into our affine line $\mathbb{A}^1$ simply by identifying $\mathbb{A}^1$ with the set of all non-vertical lines through the origin and letting infinity be the vertical line. This leads naturally to the *projective line* $\mathbb{P}^1 = \mathbb{P}^1_{\mathbb{F}}$ which is the set of all lines (vertical or not) through the origin.

A bit more formally, consider the 2-dimensional vector space $\mathbb{F}^2$. Any 1-dimensional subspace of $\mathbb{F}^2$ has a basis consisting of 1 nonzero vector (a, b), and two vectors generate the same subspace iff they are proportional. We denote by $[a, b]$ the equivalence class of a nonzero vector (a, b) with respect to proportionality; thus $[a, b] = [a', b']$ iff there is a nonzero constant $c \in \mathbb{F}$ such that $a' = ca, b' = cb$. We call (a, b) a *representative* (or a set of homogeneous coordinates) of $[a, b]$. We denote by $\mathbb{P}^1 = \mathbb{P}^1_{\mathbb{F}}$ the collection of all these equivalence classes. Note that if $a \neq 0$ then $[a, b] = [1, b/a]$; if $a = 0$ then $b \neq 0$, hence $[a, b] = [0, 1]$. Therefore the subset $U_0 \subset \mathbb{P}^1$ consisting of all $[a, b]$ with $a \neq 0$ can be identified with the line H we saw above (hence with $\mathbb{A}^1$), and the only other point in $\mathbb{P}^1$ is $[0, 1]$, which we naturally refer to as the 'point at infinity' or ∞. Thus

$$\mathbb{P}^1 = U_0 \cup \{\infty\}.$$

If we similarly let

$$U_1 = \{[a, b] = [a/b, 1] : b \neq 0\} = \{[c, 1] : c \in \mathbb{F}\},$$

then we can naturally identify U_1 with the horizontal line $y = 1$ in the plane, hence again with $\mathbb{A}^1$, and we have

$$\mathbb{P}^1 = U_1 \cup \{[1, 0]\}.$$

Thus, $\infty = [0, 1]$ and the point $[1, 0]$ are entirely analogous, and we have demystified infinity. More on this later.

Example 3.1. 1. $\mathbb{F} = \mathbb{R}$. For $P = [a, b] \in \mathbb{P}^1$, $a^2 + b^2 > 0$ so scaling (a, b) by $\sqrt{a^2 + b^2}$, then by ± 1 we get a representative of P of the form (a, b) with $a^2 + b^2 = 1, a \geq 0$, i.e. on the right half of the unit circle, that is

$$S_+ = \{(a, b) : a^2 + b^2 = 1\} = \{(\sqrt{1 - b^2}, b) : b \in [-1, 1]\}.$$

Two points on this semicircle are always non-proportional except $[0, 1] = [0, -1]$. To account for this single exception, write

$$a = \cos(\theta), b = \sin(\theta)$$

and map (a, b) to

$$(\cos(2\theta), \sin(2\theta) = (a^2 - b^2, 2ab) = (1 - 2b^2, 2\sqrt{1 - b^2}b) = (u, v).$$

This mapping wraps the semicircle S_+ over an entire unit circle

$$S = \{u^2 + v^2 = 1\}$$

and establishes a 1-1 correspondence between $\mathbb{P}^1_{\mathbb{R}}$ and the unit circle S. An explicit inverse mapping is given by

$$S \to \mathbb{P}^1_{\mathbb{R}}$$

$$(u, v) \mapsto [\frac{1 + u}{2}, \frac{v}{\sqrt{2 + 2u}}], u \neq -1$$

$$(-1, 0) \mapsto [0, 1]$$

This explains why Physicists like to talk about $\mathbb{P}^1_{\mathbb{R}}$ as 'curling up' the real line (Mathematicians usually prefer 'compactification').

2. A somewhat more involved computation establishes a 1-1 correspondence between $\mathbb{P}^1_{\mathbb{C}}$ and the unit sphere in $\mathbb{R}^3$, bases on stereographic projection from the north pole $(0, 0, 1)$.

It is convenient to represent a variable point of $\mathbb{P}^1$ as $[X_0, X_1]$ and call X_0, X_1 *homogeneous coordinates* on $\mathbb{P}^1$. Of course, X_0, X_1 are only defined up to proportionality, they are not well-defined functions– but their ratios are. More precisely $x_{10} = X_1/X_0$ is a well-defined function on U_0 (which becomes just the usual x coordinate, under our identification of U_0 with $\mathbb{A}^1$). Likewise, $x_{01} = X_0/X_1$ is a well-defined function on U_1 (which on $U_1 \cap U_0$ becomes $1/x$). This also gives us a hint as to how to extend the 'geometry of polynomials' from the affine to the projective setting: viz. use homogeneous polynomials.

Now let's digress briefly to discuss homogeneous polynomials in some generality, as they will play an important role in our study. A polynomial $F \in \mathbb{F}[X_0, ..., X_n]$ is said to be *homogeneous of degree* d if it

is a linear combination of monomials $X_0^{m_0} \cdots X_n^{m_n}$ whose *weight*, (i.e. degree) which by definition is $m_0 + ... + m_n$, equals d.

Proposition 3.2. *A polynomial F is homogeneous of degree d iff*

$$F(tX_0, ..., tX_n) = t^d F(X_0, ..., X_n).$$

Proof. 'Only if' is clear. For 'if', we may assume F is nonzero and write

$$F = F_{m_1} + ... + F_{m_k}$$

with each F_{m_i} nonzero homogeneous of degree m_i and $m_1 < ... < m_k$ (this is the so-called 'homogeneous decomposition' of F). Then

$$\begin{aligned} t^d F(X) = F(tX) &= F_{m_1}(tX) + ... + F_{m_k}(tX) \\ &= t^{m_1} F_{m_1}(X) + ... + t^{m_k} F_{m_k}(X) \end{aligned}$$

Then the polynomial in t

$$t^d F(X) - t^{m_1} F_{m_1}(X) - ... - t^{m_k} F_{m_k}(X)$$

vanishes identically, therefore is coefficients must vanish. This forces $k = 1, m_1 = d$, so F is homogenous of degree d □

Another way to characterize homogeneous polynomials is via the *Euler identity*

$$\sum_{i=1}^{n} X_i \partial F / \partial X_i = dF \tag{3}$$

Proposition 3.3. *A polynomial F is homogeneous of degree d iff the Euler identity (3) holds.*

Proof. 'Only if' can be proven by inspection on monomials. Another proof starts with the identity

$$F(tX_0, ..., tX_n) = t^d F(X_0, ..., X_n).$$

Now differentiate both sides with respect to t, using the chain rule for the LHS. This yields

$$\sum X_i \frac{\partial F}{\partial X_i}(tX_0, ..., tX_n) = dt^{d-1} F(X_0, ..., X_n).$$

Now setting $t = 1$ yields Euler's identity.

For 'if', suppose F satisfies Euler's identity and decompose it into homogeneous parts as above:

$$F = F_{m_1} + ... + F_{m_k}, m_1 < ... < m_k.$$

From the Euler identity we get

$$\sum_i X_i \partial F/\partial X_i = dF = dF_{m_1} + ... + dF_{m_k}$$

$$= \sum_{i,j} X_i \partial F_{m_j}/\partial X_i = \sum_j m_j F_{m_j}$$

Thus

$$\sum_j (d - m_j) F_{m_j} = 0$$

which is only possible if F is homogeneous of degree d. □

Next we need the algebraic processes of homogenization and dehomogenization of polynomials. Thus, let $f \in \mathbb{F}[x_1, ..., x_n]$ be a polynomial of degree $\leq m$. The *mth homogenization* of f is by definition the homogeneous polynomial

$$F(X_0, ..., X_n) = \hom_m(f) := X_0^m f(X_1/X_0, ..., X_n/X_0)$$

(if f has degree m exactly, we call $\hom_m(f)$ simply the homogenization of f and denote it by $\hom(f)$). In the other direction, let us denote the set of homogeneous polynomials of degree m in $X_0, ..., X_n$ with coefficients in $\mathbb{F}$ by $F[X_0, ..., X_n]_m$. Given such a polynomial F, its dehomogenization is by definition

$$\deh(F) = f(x) = F[1, x_1, ..., x_n]$$

This will, in general, be a polynomial of degree $\leq m$.

Example 3.4. *(i) If* $f(x) = x - b$ *then*

$$\hom(f) = X_1 - bX_0, \hom_2(f) = X_1X_0 - bX_0^2.$$

If $f(x) = \prod_1^d (x - b_i)$ *then for all* $m \geq d$,

$$\hom_m(f) = X_0^{m-d} \prod_i^d (X_1 - b_i X_0).$$

(ii) If $F(X) = X_0^2 + X_1^2$ *then* $\deh(F) = 1 + x_1^2$. *If* $F(X) = X_0^2 X_1$ *then* $\deh(F) = x_1$.

Here are some basic properties of these operations.

Lemma 3.5. *(i) If* $d = \deg(f)$, *then* $\hom_m(f) = X_0^{m-d} h(f)$ *and* $\deh(\hom_m(f)) = f, \forall m \geq d$.

(ii) if F *is homogeneous of degree* m *and* $X_0^k, k \geq 0$ *is the highest power dividing* F, *then* $\deh(F)$ *is of degree* $m - k$ *and* $\hom_m(\deh(F)) = F/X_0^k$.

Now note that if F is a homogeneous polynomial of degree m and $P \in \mathbb{P}^1$ is a point with homogenous coordinates $[u, v]$, then any representative of P has the form $(u', v') = (au, av), a \in \mathbb{F}^* := \mathbb{F} \setminus \{0\}$ and $F(u', v') = a^m F(u, v)$. Therefore, F does not define a function on $\mathbb{P}^1$ but $F(u, v) = 0$ iff $F(u', v') = 0$, so this condition depends only on P and not on the particular representative. So it makes sense to say that P is a zero of F if $F(u, v) = 0$ for any, or every, representative. For example, if $F = uX_1 - vX_0$, then $P = [u, v]$ is a zero of F. Now the following is easy to prove via homogenizing/dehomogenizing

Lemma 3.6. *If $P = [u, v]$ is a zero of F, then $uX_1 - vX_0$ divides F.*

Proof. Either $u \neq 0$ or $v \neq 0$. We assume the former case as the latter is similar. If $u \neq 0$, then $0 = F(u, v) = u^d F(1, v/u)$ so v/u is a zero of the dehomogenization

$$f(x) = F(1, x).$$

Then $(x - v/u)|f$ so we can write

$$f(x) = (x - v/u)g(x).$$

Homogenizing, we get

$$\mathrm{hog}(f) = (X_1 - (v/u)X_0)\mathrm{hog}(g) = (uX_1 - vX_0)(\mathrm{hog}(g)/u).$$

Since F is in any case a multiple of $\mathrm{hog}(f)$, it follows that $uX_1 - vX_0$ divides F. □

Corollary 3.7. *If F is not the zero polynomial, the number of zeros of F in $\mathbb{P}^1_{\mathbb{F}}$ is finite.*

The highest power of $uX_1 - vX_0$ dividing F is called the *multiplicity* of P as zero of F and denoted $m_P(F)$. As in the case of $\mathbb{A}^1$, we may define the set of cycles $Z(\mathbb{P}^1_{\mathbb{F}})_+$ as the set of formal linear combinations $\sum_{i-1}^{r} m_i P_i$ where $m_1, ..., m_r$ are natural numbers and $P_1, ..., P_r \in \mathbb{P}^1_{\mathbb{F}}$, and there is a map called the cycle map

$$Z : \mathbb{F}[X_0, X_1]_m \to Z(\mathbb{P}^1_{\mathbb{F}})_+$$

defined by $Z(F) = \sum_P m_P(F)[P]$, the sum being over all zeros P of F. Let $Z_m(\mathbb{P}^1_{\mathbb{F}})_+ \subset Z(\mathbb{P}^1_{\mathbb{F}})_+$ denote the set of cycles of degree m. Then going in the other direction we can try to define a map

$$W : Z_m(\mathbb{P}^1_{\mathbb{F}})_+ \to \mathbb{F}[X_0, X_1]_m$$

by

$$W(\sum_{i=1}^{r} m_i[u_i, v_i]) = \prod_{i=1}^{r} (u_i X_1 - v_i X_0)^{m_i}.$$

This is well-defined only up to proportionality, but that is sufficient. The following homogeneous version of the Fundamental Theorem of Algebra holds:

Theorem 3.8. *If $\mathbb{F} = \mathbb{C}$, every homogeneous polynomial $F \in \mathbb{F}[X_0, X_1]$ has the form*

$$F(X_0, X_1) = \prod_{i=1}^{r} (a_i X_1 - b_i X_0)^{m_i}.$$

Moreover, $\sum m_i$ is the degree of F and the points $[a_1, b_1], ..., [a_r, b_r] \in \mathbb{P}^1_{\mathbb{C}}$ may be taken distinct.

Note that the Theorem implies, for $\mathbb{F} = \mathbb{C}$, that $Z(F)$ and F have the same degree, so we have shown

Corollary 3.9. *Via $Z - W$, there is a 1-1 correspondence between the set of nonzero elements in $\mathbb{C}[X_0, X_1]_m$, up to proportionality, and $Z_m(\mathbb{P}^1)_+$.*

Exercise 3.1. *Let $F \in \mathbb{F}[X_0, X_1]$ be nonzero and $f = \deh(F)$, and identify $U_0 \subset \mathbb{P}^1_{\mathbb{F}}$ with $\mathbb{A}^1_{\mathbb{F}}$, thus identifying $Z(\mathbb{A}^1_{\mathbb{F}})$ with a subset of $Z(\mathbb{P}^1_{\mathbb{F}})$ known as the set of 'finite cycles'. Then show that*

$$Z(F) = Z(f) + m_\infty[\infty]$$

4. Projective substitutions

Projective transformations, or substitutions, are no less important than their affine analogues– and the good news is, they are in some ways simpler. To see why, let's go back to our transformation $T = T_{(a,b)}$ mapping x to $ax+b$. From our projective 'perspective', x is represented by the point $[1, x] \in \mathbb{P}^1$, and T maps this to

$$\begin{bmatrix} 1 \\ ax+b \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ b & a \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix}$$

Thus, via our projective notation, an affine transformation becomes a linear one, given by multiplication by the 2×2 matrix on the right. Such a matrix is called affine, and it is always nonsingular provided $a \neq 0$. But now notice that if $A = [a_{i,j}]$ is any nonsingular 2×2 matrix, then we can always define a transformation

$$T = T_A : \mathbb{P}^1 \to \mathbb{P}^1$$

by the rule that $T([X_0, X_1]) = [X_0', X_1']$ where

$$\begin{bmatrix} X_0' \\ X_1' \end{bmatrix} = \begin{bmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{bmatrix} \begin{bmatrix} X_0 \\ X_1 \end{bmatrix}.$$

These are known as *projective transformations*. Note that if we formally set $x = X_1/X_0$ then T maps x to

$$x' = \frac{a_{10} + a_{11}x}{a_{00} + a_{01}x}.$$

Under this guise, these transformation are known as Möbius or fractional-linear transformations. These have been studied a great deal and they have many interesting properties. Before discussing some of these, here is an example.

Example 4.1. Let

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Then $T_A[X_0, X_1] = [X_1, X_0]$; or in fractional-linear guise, T_A takes x to $1/x$. Not surprisingly, this is known as *inversion*. It interchanges $0 = [1, 0]$ and ∞ and fixes $[1, 1]$.

Let's continue with some elementary general facts about projective transformations:

Lemma 4.2. *(i)* $T_{AB} = T_A \circ T_B$

(ii) T_{cI_2} *is the identity transformation (where* I_2 *is the identity matrix).*

(iii) Conversely, if T_A is the identity transformation, then A is a scalar matrix, i.e. $A = cI_2$ for some $c \neq 0$.

Let us prove (iii). If $A = [a_{ij}]$ is such that T_A is the identity, then

$$A \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

This means $a_{10} = 0$. Similarly, $a_{01} = 0$. A being nonsingular, it follows that a_{00}, a_{11} are both nonzero. Now use

$$A \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

i.e.

$$\begin{bmatrix} a_{00} \\ a_{11} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Thus the vectors (a_{00}, a_{11}) and $(1, 1)$ are proportional, so A is a scalar matrix.

Actually, this proof also establishes the following important fact

Corollary 4.3. *Any projective transformation that fixes the three points*

$$'0' = [1, 0], '1' = [1, 1], '\infty' = [1, 0]$$

is the identity.

Another important consequence is

Corollary 4.4. *(i) The composition of any number of projective transformations is projective, as is the inverse of any projective transformation.*

(ii) Given nonsingular matrices A, B, the projective transformations T_A and T_B are equal iff $A = cB$ for some scalar c.

Proof. (i) For inverses,

$$T_A \circ T_{A^{-1}} = T_{AA^{-1}} = Id$$

Therefore $T_{A^{-1}} = (T_A)^{-1}$.

(ii) If $T_A = T_B$ then

$$T_{AB^{-1}} = T_A T_{B^{-1}} = T_A (T_B)^{-1} = Id$$

Hence $AB^{-1} = cI_2$ is scalar so $A = cB$.

□

Here is another remarkable property of projective transformations:

Proposition 4.5. *Given any 3 distinct points $P, Q, R \in \mathbb{P}^1$, there is a unique projective transformation T mapping $0, 1, \infty$ to P, Q, R in order.*

Proof. First existence. Let's write $P = [p_0, p_1]$ etc. Then the vectors $(p_0, p_1), (q_0, q_1)$ are not proportional, so a first try might be the matrix

$$A_1 = \begin{bmatrix} p_0 & r_0 \\ p_1 & r_1 \end{bmatrix}.$$

Then T_{A_1} indeed takes 0 to P and ∞ to R. What about Q? Well, notice that any diagonal matrix $D = \mathrm{Diag}(d_1, d_2)$ with $d_1 d_2 \neq 0$ fixes 0 and ∞, therefore $T_{A_1 D}$ also takes 0 to P and ∞ to R. Can we find D such that $T_{A_1 D}$ also takes 1 to Q? If so, $A = A_1 D$ solves our problem. As system of linear equations, this condition reads

$$A_1 D \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} q_0 \\ q_1 \end{bmatrix}$$

(proportionality rather than equality would suffice). As A_1 is nonsingular, this is equivalent to

$$D \begin{bmatrix} 1 \\ 1 \end{bmatrix} = A_1^{-1} \begin{bmatrix} q_0 \\ q_1 \end{bmatrix}$$

in other words,

$$\begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = A_1^{-1} \begin{bmatrix} q_0 \\ q_1 \end{bmatrix}.$$

Clearly d_1, d_2 cannot both be zero. If $d_1 = 0$ then $[d_1, d_2] = \infty$ so we conclude $T_{A_1}(\infty) = Q$, whereas by construction $T_{A_1}(\infty) = R$, so $Q = R$ against our assumption P, Q, R distinct. Therefore $d_1 \neq 0$ and similarly, $d_2 \neq 0$. Therefore $D = \mathrm{Diag}(d_1, d_2)$ is our solution. This shows existence.

Uniqueness: suppose A, B are such that T_A, T_B both take $0, 1, \infty$ to the same P, Q, R. Then $T_A \circ (T_B)^{-1} = T_{AB^{-1}}$ takes $0, 1, \infty$ to themselves, hence it is the identity, so that A, B differ by a scalar and $T_A = T_B$. □

A special case of this is that no projective transformation other than the identity can fix more than 3 points.

Corollary 4.6. *Given 2 triples of distinct points P, Q, R and P', Q', R' on $\mathbb{P}^1$, there is a unique projective transformation taking P, Q, R to P', Q', R' in order.*

Proof. If T takes P, Q, R to P', Q', R', let T_1 (resp. T_2) take $0, 1, \infty$ to P, Q, R (resp. P', Q', R'). Then $T \circ T_1$ takes $0, 1, \infty$ to P', Q', R', so by uniqueness $T \circ T_1 = T_2$. Therefore $T = T_2 \circ T_1^{-1}$ is uniquely determined. □

Exercise 4.1. *(i) Show that any projective transformation fixing ∞ is affine.*

(ii) Show that any projective transformation fixing 0 and ∞ is diagonal (comes from a diagonal matrix)

(iii) Show that any projective transformation interchanging 0 and ∞ is of the form $[X_0, X_1] \mapsto [X_1, cX_0], c \neq 0$.

Now given *four* distinct points $P, Q, R, S \in \mathbb{P}^1_{\mathbb{F}}$, i.e. a distinct quadruple, let T be the unique projective transformation taking P, Q, R to $0, 1, \infty$. Then $T(S) \in \mathbb{P}^1_{\mathbb{F}}$ is a point distinct from $0, 1, \infty$, so it can be represented in the form $[1, \lambda], \lambda \in \mathbb{F}\backslash\{0, 1\}$. λ is called the *cross ratio* of P, Q, R, S and denoted $[P, Q, R, S]$; similarly if $x_1, ..., x_4 \in \mathbb{F} = \mathbb{A}^1_{\mathbb{F}}$ are distinct, their cross-ratio, denoted $[x_1, x_2, x_3, x_4]$, is by definition the cross-ratio of the corresponding points in $\mathbb{P}^1_F$, that is, $[1, x_1], ..., [1, x_4]$. As in the proof of Cor. 4.6, we can show:

Corollary 4.7. *Two distinct quadruples $(x_1, ..., x_4), , (y_1, ..., y_4)$ on $\mathbb{P}^1$ are projectively equivalent iff $[x_1, ..., x_4] = [y_1, ..., y_4]$*

Exercise 4.2. *Compute the cross-ratio $[1, 2, 3, 4]$.*

Exercise 4.3. *Given $\lambda = [x_1, x_2, x_3, x_4]$, compute $[x_2, x_1, x_3, x_4]$, $[x_4, x_2, x_3, x_1]$ and $[x_1, x_4, x_3, x_2]$.*

Now let's look over on the algebraic side at the relation of projective transformations (literally, substitutions) and homogeneous polynomials. For any homogeneous polynomial $F \in \mathbb{F}[X_0, X_1]_m$ and 2x2 matrix $A = [a_{ij}]$, define another such polynomial, denoted $T^*_A(F) \in \mathbb{F}[X_0, X_1]_m$, by

$$T^*_A F(X_0, X_1) = F(A(X_0, X_1)) = F(a_{00}X_0 + a_{01}X_1, a_{10}X_0 + a_{11}X_1)$$

Note that by definition, if $P \in \mathbb{P}^1$ and $Q := T_A(P)$ is a zero of F, then P is a zero of $T^*_A(F)$ and vice versa, in other words, Q is a zero of F iff $T^{-1}_A(Q)$ is a zero of $T^*_A(F)$.

Example 4.8. Let $L_P = vX_0 - uX_1$ be the linear polynomial (unique up to constant multiple) with unique zero $P = [u, v]$. By direct calculation,

$$T^*_A(L_P) = (va_{00} - ua_{01})X_0 - (-va_{01} + ua_{11})X_1 = L_Q$$

where

$$Q = [ua_{11} - va_{01}, -ua_{01} + va_{10}] = T_{adj(A)}(P)$$

where

$$adj(A) = \begin{bmatrix} a_{11} & -a_{01} \\ -a_{10} & a_{00} \end{bmatrix}$$

is the adjoint matrix, which coincides with A^{-1} up to the scalar $\det(A)$, therefore $T_{adj(A)} = T_A^{-1}$.

Now for any nonzero homogeneous polynomial F and point P, we can write $F = L_P^m F_1$ with F_1 nonzero at P and $m \geq 0$ equal to the multiplicity of F at P. Then we get a nice relation between the cycles associated to F and $T_A^*(F)$:

$$Z(T_A^*(F)) = T_{A^{-1}}(Z(F)) \tag{4}$$

where we define, for any matrix B and cycle $Z = \sum m_i[P_i]$,

$$T_B(Z) = \sum m_i[T_B(P_i)].$$

Equivalently,

$$T_A(Z(T_A^*(F))) = Z(F).$$

Two polynomials $F, G \in \mathbb{F}[X_0, X_1]_m$ are said to be *projectively equivalent* over $\mathbb{F}$ if $G = cT_A^*(F)$ for some 2x2 matrix A and nonzero constant $c \in \mathbb{F}$. This relation is easily seen to be an equivalence relation, and the equivalence classes are also called projective orbits. Similarly, two cycles Z, Z' are said to be projectively equivalent if there is a projective transformation T_A such that $T_A(Z) = Z'$.

Corollary 4.9. *If F, G are projectively equivalent polynomials, then $Z(F), Z(G)$ are projectively equivalent cycles.*

Exercise 4.4. *If $\mathbb{F} = \mathbb{C}$, then two polynomials F, G are projectively equivalent iff $Z(F), Z(G)$ are projectively equivalent cycles.*

Example 4.10. Some low degree examples:

- $m = 1$ Here there is clearly just 1 orbit, regardless of $\mathbb{F}$.
- $m = 2$ Here the nature of the field $\mathbb{F}$ begins to play a role. First, the usual 'completing the square' technique shows that any F is equivalent to $G = X_1^2 + aX_0^2, a \in \mathbb{F}$. If $\mathbb{F} = \mathbb{R}$ and $a \neq 0$, G is equivalent to $X_1^2 \pm X_0^2$ and these are inequivalent, therefore there are precisely 3 orbits. If $\mathbb{F} = \mathbb{C}$, $X_1^2 \pm X_0^2$ are equivalent (substitute $X_0 \mapsto \sqrt{-1}X_0$), so there are just 2 orbits. If $\mathbb{F} = \mathbb{Q}$, there are infinitely many nonsquares and infinitely many orbits.
- $m = 3$ Here we will just consider the case $\mathbb{F} = \mathbb{C}$. Any cubic F splits into linear factors $F = L_1L_2L_3$ with each L_i homogeneous linear with unique zero P_i.

  Case 1: P_i distinct. Then there is a projective transformation T_A taking $0, 1, \infty$ to P_1, P_2, P_3, hence L_1, L_2, L_3 to $X_1, X_1 - X_0, X_0$, respectively (up to constant factor). Then $T_A^*(F) = X_0X_1(X_1 - X_0)$.

Case 2: $P_1 = P_2 \neq P_3$. Then we can transform $0, \infty$ to P_1, P_3 and similarly conclude F is equivalent to $X_1^2 X_0$.

Case 3: P_i all equal. Then clearly F is equivalent to X_1^3.

Thus in total there are 3 orbits.

Over $\mathbb{R}$, a slightly more complicated analysis. based on the fact that a real cubic has precisely 1 or 3 real roots, yields 4 orbits: in addition to 3 orbits analogous to the above, there is the orbit of $X_1(X_1^2 + X_0^2)$.

Exercise 4.5. *Fill in the details in the above argument classifying orbits in* $\mathbb{R}[X_0, X_1]_2$ *and* $\mathbb{C}[X_0, X_1]_2$.

Exercise 4.6. *Classify projective equivalence classes in* $\mathbb{R}[X_0, X_1]_3$.

As soon as $m \geq 4$, the set of projective equivalence classes in $\mathbb{F}[X_0, X_1]_m$ is no longer finite and indeed gets rather complicated. In the simplest case $m = 4, \mathbb{F} = \mathbb{C}$, this set can be identified with the set of equivalence classes of 4 distinct point in $\mathbb{P}^1_{\mathbb{C}}$, and hence with $\mathbb{P}^1_{\mathbb{C}} \setminus \{0, 1, \infty\}$.

Remark. Underlying some of the topics we discussed above is the fact that the set of all projective transformations of $\mathbb{P}^1_{\mathbb{F}}$ forms an abstract-algebraic structure known as a *group* (studied in courses like 171-2):in our context this essentially means that compositions and inverses of projective transformations are themselves projective transformations. The group of all projective transformations of $\mathbb{P}^1_{\mathbb{F}}$ is denoted $\mathrm{PGL}_2(\mathbb{F})$ or just PGL_2. This group is closely related to the group of all nonsingular 2x2 matrices, which is denoted $\mathrm{GL}_2(\mathrm{F})$ or GL_2. Indeed $\mathrm{PGL}_2(\mathbb{F})$ is the quotient or factor group of $\mathrm{GL}_2(\mathbb{F})$ by the (normal) subgroup of scalar matrices, i.e. $\mathbb{F}^* I_2$ (which reflects the fact that two matrices define the same projective transformation iff they are scalar multiples of each other.