Plane Algebraic Curves

Dhruv Goel

June 2024

Contents

Preface

These are lecture notes for a course on classical algebraic geometry that I taught at Ross/Ohio 2024 intended for peer mentors and counselors. The course covers the fundamental theory of plane algebraic curves, up to a proof of Bézout's Theorem and an introduction to the theory of elliptic curves. The course only assumes familiarity with the material on the Ross first-year sets.

Chapter 1

Lecture Notes

1.1 06/10/24 - Introduction

Example 1.1.1 (Student Examples). Get Desmos to plot the subsets of the plane (over $k = \mathbb{R}$) defined by the vanishing of the following polynomials

(a) $3x + 4y - 7$ (line)
(b) $x^2 + y^2 - 1$ (circle),
(c) $y - x^2$ (parabola),
(d) $y^2 + x^3$ (semicubical parabola/cuspidal cubic),
(e) $y^2 - x^3 - x$ (one-component elliptic curve),
(f) $y^2 - x^3 + x$ (two-component elliptic curve),
(g) $(x^2 + y^2)(x + y - 1)$ (line and point not on it),
(h) $xy - 1$ (hyperbola), and
(i) $x^2 + y^2 + 1$ (empty set).

These are all examples of algebraic curves. Now get Desmos to plot

(a) $y - \sin(1/x)$, and
(b) $y - |x|$.

These are not plane algebraic curves (why?). See also Exercise 2.1.8.

We will fix a field k throughout (see Remark 1.1.17).

Definition 1.1.2. The affine plane over k, denoted $\mathbb{A}^2_k$, is the set of ordered pairs of elements of k, so that

$$\mathbb{A}^2_k := \{(p, q) : p, q \in k\}.$$

If you want, see Remark 1.1.18 for an explanation of why we use $\mathbb{A}^2_k$ to denote the set others sometimes denote by k^2.

Given a function $F : \mathbb{A}^2_k \to k$, we can look at its vanishing locus, denoted variously by

$$F^{-1}(0) = C_F = \mathbb{V}(F) = \mathrm{Z}(F) = \{(p, q) : F(p, q) = 0\}.$$

We will usually stick to the notation C_F.

Remark 1.1.3. More generally, we can look at the level sets $F^{-1}(a)$ for all $a \in k$. Why does this perspective not add anything new?

Any polynomial $f(x, y) \in k[x, y]$ gives rise to a function $F_f : \mathbb{A}^2_k \to k$ by evaluation.

Remark 1.1.4. Why is it important to keep the notions of a polynomial and polynomial function separate? See Exercise 2.2.6.

Definition 1.1.5. An affine plane algebraic curve is the vanishing locus of a polynomial function in the affine plane given by a nonconstant polynomial, i.e. a subset $C \subset \mathbb{A}^2_k$ of the form $C = C_{F_f}$ for some nonconstant polynomial $f(x, y) \in k[x, y]$.

For simplicity, we'll use the notation $C_f := C_{F_f}$. We will sometimes write $C_f(k)$ to denote C_f if we want to emphasize the underlying field. Finally, we will often abbreviate "affine plane algebraic curves" to simply "curves," since we will not have occasion to deal with other kinds of curves, at least initially.

Remark 1.1.6. Our definition is currently a little weird. For instance, with our current definition, for certain fields k, a curve can be

- empty (think $x^2 + y^2 + 1 = 0$ over $\mathbb{R}$),
- a finite collection of points (think $x^2 + y^2 = 0$ over $\mathbb{R}$ and Proposition 1.1.7, or think of what happens when $k = \mathbb{F}_q$ is a finite field),
- and all of $\mathbb{A}^2_k$ (again think of $k = \mathbb{F}_q$ being a finite field).

Neither of these sets seem to be "1-dimensional," which is the elusive notion we are trying to capture. We could either choose to restrict ourselves to working over infinite fields or algebraically closed fields (even in positive characteristic–see Exercise 2.2.8), but this misses a lot of important number theory (see Examples 1.1.11 and 1.1.15). Alternatively, we can accept that our definition is broader than initially intended, and try to study its consequences.

Proposition 1.1.7. Let k be a field. If $C, D \subset \mathbb{A}^2_k$ are curves, then so is $C \cup D$.

Proof. If $C = C_f$ and $D = C_g$ for $f, g \in k[x, y]$, then $C \cup D = C_{fg}$. ∎

Remark 1.1.8. Here we are using that $k[x, y]$ is a ring (how?), and that k is a field (or at least that it is a domain–what happens if k is not even a domain?). We will say more about this when we talk about irreducibility and reducedness of curves.

1.1.1 Motivating Questions

Given a field k and a curve $C \subset \mathbb{A}^2_k$, we can ask several questions about it.

Question 1.1.9. Is $C = \emptyset$?

This is not at all as trivial as it seems. Many number-theoretic questions can be phrased in this language, if we take k to be $\mathbb{Q}$ or a finite field $\mathbb{F}_q$, for instance.

Example 1.1.10. Take $k = \mathbb{Q}$, fix a prime p, and look at the curve C defined by

$$f(x, y) := x^2 + y^2 - p \in \mathbb{Q}[x, y].$$

Then $C = \emptyset$ iff p satisfies a certain congruence condition (which?). See Exercise 2.1.1.

Example 1.1.11. Take $k = \mathbb{F}_p$ to be a finite field of prime order and $a \in k$ to be any element, and look at the curve C defined by

$$f(x, y) = x^2 - a \in \mathbb{F}_p[x, y].$$

Then $C = \emptyset$ iff a is quadratic nonresidue modulo p, i.e. $\left(\frac{a}{p}\right) = -1$.

Remark 1.1.12. For any field k, if $f(x, y) \in k[x, y]$ is a polynomial of x only, then the curve C_f defined by f is a finite (possibly empty) union of "vertical lines". Can you make this precise?

Example 1.1.13. Take $k = \mathbb{Q}$ and $n \geq 1$ to be a positive integer. Let

$$f_n(x, y) := x^n + y^n - 1 \in \mathbb{Q}[x, y],$$

and $C_n := C_{f_n}$ be the curve defined by f_n. Then Fermat's Last Theorem says that

$$C_n(\mathbb{Q}) = \emptyset \Leftrightarrow n > 2.$$

Question 1.1.14. If C is nonempty, what can we say about the locus C? Is it finite or infinite? What can we say about its topology[a]?

[a] What's that?

Example 1.1.15. For instance, if k is finite, what is the cardinality of $C(k)$? Suppose $k = \mathbb{F}_q$ is a finite field, and that C is an elliptic curve[1], e.g. the curve defined by

$$f(x, y) = y^2 - x^3 - x \in \mathbb{F}_q[x, y]$$

when q is not a power of 2. The Hasse Theorem says that, in the above case,

$$(\sqrt{q} - 1)^2 \leq \#C(\mathbb{F}_q) \leq (\sqrt{q} + 1)^2.$$

In particular, we have $\#C(F_q) \sim q$ for all large q. (What does that even mean? Aren't we starting with a fixed q to begin with?) We will not prove this theorem in this course.

Example 1.1.16. If $k = \mathbb{R}$ or $k = \mathbb{C}$, how many pieces (i.e. connected components) does $C(k)$ have? How are they related to each other? See Exercise 2.1.2 for the case when $k = \mathbb{R}$. Another theorem, which will not prove in this course, asserts that if $k = \mathbb{C}$, then any irreducible curve[2] is connected.

1.1.2 Some Unimportant Remarks

Remark 1.1.17. Why did we require k to be a field? What would happen if k were just a ring–does the notion of an affine plane curve over a ring make sense? [Hint: some things make sense, whereas other things like Proposition 1.1.7 break down. See Remark 1.1.8.] Can you see how far you can go till things break down and what you can salvage by adapting definitions?

Remark 1.1.18. As sets, $\mathbb{A}^2_k$ and $k^2 = k \times k$ are identical[3], but $\mathbb{A}^2_k$ does not come equipped with additional structure that k^2 is often (implicitly) interpreted to have: k^2 is often seen (by students who have seen some linear algebra) as a vector space with an additive structure and a distinguished origin, but for us $\mathbb{A}^2_k$ is just a set[4] and, as will become clear when we discuss affine changes of coordinates, there is no distinguished point in $\mathbb{A}^2_k$–all points "look the same". In slightly more grown-up terminology, the affine plane over k is a principal homogenous space or torsor for the (underlying additive group) of the vector space k^2. If you do not understand what this remark means, you can safely ignore it.

Remark 1.1.19. Regarding the different choices of the field k: it's often easiest to plot curves over $k = \mathbb{R}$, but plots can also be made over other fields such as $k = \mathbb{C}$ (using some ingenuity and imagination–how?) or $k = \mathbb{F}_q$ (this may be a silly, uninstructive plot, but not always!). We will see throughout the course that it is, in fact, easier to work with curves over $k = \mathbb{C}$ than over $k = \mathbb{R}$ (why do you think this might be?). However, curves over other fields are equally important:

(a) Fields such as $k = \mathbb{Q}, \mathbb{F}_p$ (or finite extensions and completions of these–such as $k = \mathbb{Q}_p$) show up a lot in solving number-theoretic questions. See Examples 1.1.10, 1.1.11 and 1.1.13.

[1] We will define this notion formally later.

[2] Now, what's that?

[3] Only according to our definition! There are other accepted definitions of $\mathbb{A}^2_k$, such as $\mathbb{A}^2_k = \operatorname{Spec} k[x, y]$, for which this is no longer the case. You don't have to wrorry too much about this right now.

[4] Later on in your studies, it can, and will, be given the structure of a topological space, and in fact a locally ringed space (even affine scheme).

(b) Another case of interest is when $k = K(t)$ for some other field k. When $K = \mathbb{F}_q$ is a finite field, working with curves over $k = \mathbb{F}_q(t)$ is known as a the "function field analog" of the theory of curves. Many important questions which are unsolved in the "usual case" have been solved in the function field case (such as the Riemann Hypothesis), and this provides (one strand of) evidence for the Riemann Hypothesis.

(c) In (b), when we take $K = \mathbb{C}$, so that we are looking at curves over $k = \mathbb{C}(t)$, we are *really* looking at one-parameter families of curves that fit together into an algebraic surface. For instance, elliptic curves over $\mathbb{C}(t)$ often give rise to elliptic K3 surfaces. This perspective is very helpful in the study of higher-dimensional algebraic varieties as well.

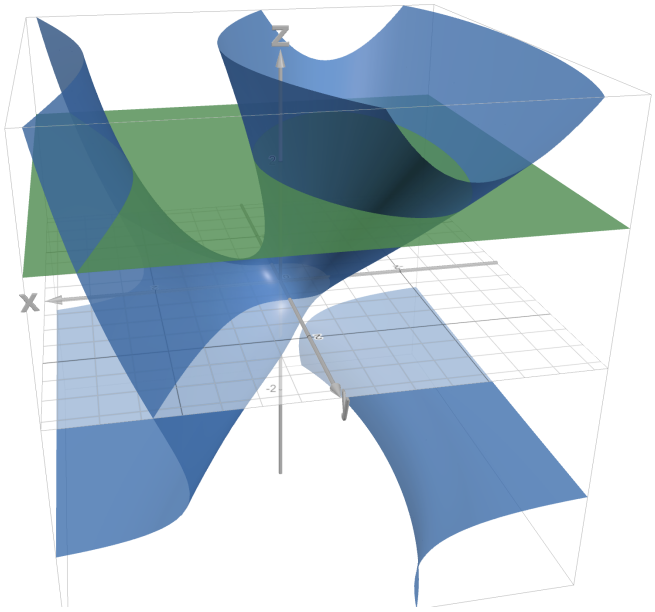


Figure 1.1: The elliptic curve over $k = \mathbb{C}(z)$ defined by $y^2 = x^3 - 3zx + (z^3 + 1)(z + 2)^{-1}$ in blue, along with its hyperplane section at $z = 2$, which is the elliptic curve $y^2 = x^3 - 6x + 9/4$. Picture made with Desmos 3D.

Therefore, it is helpful to have the flexibility to work over arbitrary fields from the beginning.

1.2 06/12/24 - Degree I, More Examples

Today, I want to start discussing an important notion, namely that of the *degree* of an algebraic curve, and give more examples of curves.

1.2.1 Degree I

Clearly, the "degree" of a line should be one, whatever the word "degree" means. Similarly, the degree of the parabola defined by $y - x^2$ should be two.

So we can start defining the degree of a polynomial $f \in k[x, y]$ as follows: the degree of a monomial $cx^i y^j$ where $0 \neq c \in k$ and $i, j \geq 0$ is $i+j$, and the degree of f is the maximal degree of the (finitely many) monomials appearing in it. Here's one definition we can now propose:

Definition 1.2.1 (Degree–Attempt I)**.** For a field k and curve $C \subset \mathbb{A}^2_k$, pick a nonconstant $f \in k[x, y]$ such that $C = C_f$ (this exists because C is a curve!), and define the degree of C by

$$\deg C := \deg f.$$

Is this a definition? Well, not really. For this to be a definition, we have to check that if for $f, g \in k[x, y]$ we have $C_f = C_g$, then $\deg f = \deg g$. Unfortunately, this is not quite the case with our definitions. Consider the following examples:

(a) When $k = \mathbb{R}$, we can take $f(x, y) = x^3 - y^3$ and $C = C_f$. Then C_f is also C_ℓ where $\ell(x, y) := x - y$, but $\deg f = 3$ while $\deg \ell = 1$.

(b) What happens to the empty set? E.g. when $k = \mathbb{R}$, then for any $n \geq 1$ we have $C_{f_n} = \emptyset$, where $f_n := x^{2n} + y^{2n} + 1 \in k[x, y]$. Therefore, the empty set should have degree every positive even integer.

(c) Maybe (a) and (b) illustrate that there is something wrong with the field $k = \mathbb{R}$. But, in fact, this notion is problematic over other fields too: for any field $f \in k[x, y]$, we have thanks to the proof of Proposition 1.1.7 that

$$C_{f^2} = C_f \cup C_f = C_f.$$

If f is nonconstant, then $\deg f^2 = 2 \deg f > \deg f$, and this is a problem.

What should we do? One salvage (proposed by students) could be:

Definition 1.2.2 (Degree–Attempt II)**.** For a field k and curve $C \subset \mathbb{A}^2_k$, look at the set

$$\{\deg f : \text{nonconstant } f \in k[x, y] \text{ such that } C = C_f\}.$$

This set is a nonempty subset of the positive integers by definition, and so we may use the Well-Ordering Principle to define the degree of C, written $\deg C$, to be the least element of this set.

This is at least a definition. However, again we have some weird properties. For instance, by this definition, in example (a) above, the curve defined by $f(x, y) = x^3 - y^3$ will have degree 1, whereas the empty set of example (b) will have degree 2 (why?). Let's use this as a provisional definition for now–we will revisit it in a few lectures.

Let's now do some more examples of curves.

1.2.2 Polar Curves

I'll assume some familiarity with polar coordinates.

Definition 1.2.3. Given any function $G : [0, \infty) \times \mathbb{R} \to \mathbb{R}$, the polar curve $P_G \subset \mathbb{A}^2_{\mathbb{R}}$ implicitly defined by the vanishing of G is the subset

$$P_G := \{(r\cos\theta, r\sin\theta) : (r, \theta) \in [0, \infty) \times \mathbb{R} \text{ such that } G(r, \theta) = 0\} \subset \mathbb{A}^2_{\mathbb{R}}.$$

Example 1.2.4. The Archimedean spiral is the polar curve defined by $G(r, \theta) = r - \theta$. (Get Desmos to draw a picture!)

Remark 1.2.5. Note that there is some redundancy here: for any $(r, \theta) \in [0, \infty) \times \mathbb{R}$, the polar coordinates (r, θ) and $(r, \theta + 2\pi)$ define the same point in $\mathbb{A}^2_{\mathbb{R}}$, and for all $\theta \in \mathbb{R}$, the polar coordinates $(0, \theta)$ define only the origin $(0, 0) \in \mathbb{A}^2_{\mathbb{R}}$. Could we perhaps come up with a better domain of definition for G?

A natural question to ask is: which of these curves is an algebraic curve? Here's one thing you can do: any nonconstant polynomial $g(r, c, s) \in \mathbb{R}[r, c, s]$ in the variables r, c, and s[5] defines a function G_g of r and θ by

$$G_g(r, \theta) = g(r, \cos\theta, \sin\theta).$$

The vanishing set of G_g will be denoted by $P_g := P_{G_g}$; this is the curve implicitly defined by the "polar polynomial" g.

Example 1.2.6. What curve do you get by taking $g(r, c, s) = (r^2 - 1)^3 - r^5c^2s^3$?

Example 1.2.7. What's the equation of a line $\ell \subset \mathbb{A}^2_{\mathbb{R}}$ defined by say $ax+by+c = 0$ for $a, b, c \in \mathbb{R}$ with not both a and b zero, in polar coordinates?

But how do we know that such a subset is always an algebraic curve in our definition (using x and y coordinates)? Here's the result we need:

Proposition 1.2.8. Given any nonconstant $g(r, c, s) \in \mathbb{R}[r, c, s]$, there is a nonconstant $f(x, y) \in \mathbb{R}[x, y]$ such that

$$P_g \subset C_f.$$

Proof. We give an algorithm to produce an f. Firstly, find $k \geq 0$ such that $r^k g$ is a polynomial in the variables r, rc and rs. Next, rearrange to separate odd powers of r, i.e. find polynomials $p(t, u, v), q(t, u, v) \in \mathbb{R}[t, u, v]$ such that

$$r^k g = r \cdot p(r^2, rc, rs) - q(r^2, rc, rs).$$

Finally, take

$$f(x, y) := (x^2 + y^2) \cdot p(x^2 + y^2, x, y)^2 - q(x^2 + y^2, x, y)^2.$$

■

We leave it to the reader to verify details of the proof (why is f nonconstant?), as well as the fact that this procedure works; it is, of course, essentially the only natural thing to do.

[5] Even any element in the quotient ring $\mathbb{R}[r, c, s]/(c^2 + s^2 - 1)$.

Example 1.2.9. Consider $g(r, c, s) = r^2 - s$. Take $k = 1$ and $p = t$ and $q = v$ to get

$$f(x, y) = (x^2 + y^2)^3 - y^2.$$

Use Desmos to plot the curves P_g and C_f.

Here are two issues with this approach:

(a) From Example 1.2.9, it is clear that the "squaring" at the last step introduces extraneous components. Can these components be avoided? We will eventually develop more tools to answer such questions, but for right now you are invited to explore this in Exercise 2.1.3.

(b) Is the f produced in Proposition 1.2.9 here unique? It is not because we can always multiply f with anything else: for any $h \in \mathbb{R}[x, y]$, we have $C_f \subset C_{fh}$. Here's a better question: is this f unique (up to scalars) if we require it to be of smallest degree? You are invited to explore this in Exercise 2.1.10.

1.2.3 Synthetic Constructions

Sometimes, we can give "synthetic constructions" for curves. Instead of telling you what that means, I'll just go over a few examples. For now, we'll stick to $k = \mathbb{R}$.

Example 1.2.10. Given a line $\ell \subset \mathbb{A}^2_{\mathbb{R}}$ (the "directrix") and a point $O \in \mathbb{A}^2_{\mathbb{R}}$ not on it (the "focus"), we can look at the locus

$$C := \{P \in \mathbb{A}^2_{\mathbb{R}} : \text{dist}(P, \ell) = \text{dist}(P, O)\}$$

of points at an equal distinct from ℓ and O. This is, of course, one classical definition of the parabola. Taking the line ℓ to be $x + a = 0$ and the point O to be $(a, 0)$ for some $0 \neq a \in \mathbb{R}$ (see Figure 1.2) gives us the algebraic equation

$$f(x, y) = y^2 - 4ax.$$

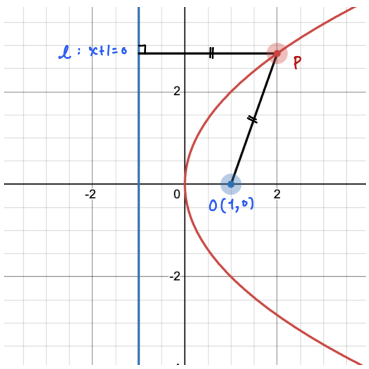


Figure 1.2: The synthetic construction of the parabola. Picture made with Desmos.

Other conic sections–ellipses and hyperbolae–also admit such synthetic descriptions. One way to connect these synthetic definitions to the definitions as sections of a cone is to use Dandelin spheres; see this fantastic video by 3Blue1Brown for more on this. Finally, note that an ellipse limits to a circle as the foci coincide, and a pair of lines as well as a "double" line can be obtained as a "limit" of these conic sections as well–for instance, as $a \to 0$, the above parabola limits to the "double" line $y^2 = 0$. This suggests that we should also count pairs of lines and double lines as conic sections, at least if we the set of conic sections to be closed under limits of coefficients. This motivates the following definition over arbitrary fields:

Definition 1.2.11. For a field k, a conic section, or conic, is a curve $C \subset \mathbb{A}^2_k$ defined by the vanishing of a quadratic polynomial of the form

$$f(x,y) = ax^2 + hxy + by^2 + ex + fy + c \in k[x,y]$$

for some $a, b, c, e, f, h \in k$, not all zero.

Note how this definition encapsulates all the above notions: of ellipses, hyperbolae, parabolae, pairs of lines, and double lines. In Exercise 2.1.6, you'll show that at least when $k = \mathbb{C}$, these are *all* the conics, up to affine changes to coordinates (to be defined soon). When $\operatorname{ch} k \neq 2$, it is often traditional to replace h, e, f in the above with $2h, 2e, 2f$–this is because it allows us to think of this vanishing locus as the set of (x, y) such that

$$\begin{bmatrix} x & y & 1 \end{bmatrix} \begin{bmatrix} a & h & e \\ h & b & f \\ e & f & c \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = [0]$$

and then to use tools of linear algebra to help us study conics. More on this later.

Example 1.2.12 (Cassini Ovals and Lemniscate)**.** For any two points $A, B \in \mathbb{A}^2_{\mathbb{R}}$ and constant $b \geq 0$, we can consider the locus

$$C_b := \{P \in \mathbb{A}^2_{\mathbb{R}} : \operatorname{dist}(P, A) \cdot \operatorname{dist}(P, B) = b^2\}.$$

For varying values of b, these give a family of curves, whose members are called Cassini ovals. These are named after the 17^{th} century astronomer Giovanni Domencio Cassini, who used these in his study of planetary motion. Taking A and B to be at $(\pm a, 0)$ for $0 \neq a \in \mathbb{R}$ yields the equation

$$f_{a,b}(x,y) := \left((x-a)^2 + y^2\right)\left((x+a)^2 + y^2\right) - b^4 \in \mathbb{R}[x,y].$$

The shape of these ovals depends only on the eccentricity $e := b/a$. When $e = 0$, the curve is two points; when $0 < e < 1$, the curve consists of two oval pieces (i.e. connected components); when $e = 1$, the curve is the Lemniscate of Bernoulli–the ∞ symbol–which has a node at the origin; when $e > 1$, the curve is connected. For $1 < e < \sqrt{2}$, the curve is not convex, but for $e \geq \sqrt{2}$ it is. The limiting case of $e \to \infty$ is the circle. You are invited to prove these results in Exercise 2.2.2. See Figure 1.3 in which I have drawn these ovals for some values of e between 0 and 2, and marked the special cases $e = 0, 1, \sqrt{2}$ in black.

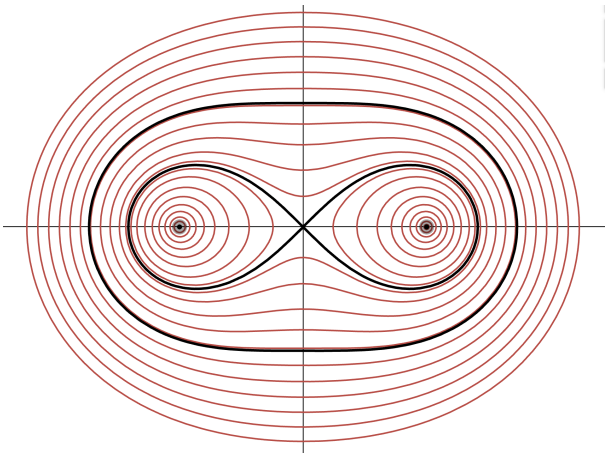


Figure 1.3: The Cassini ovals. Picture made with Desmos.

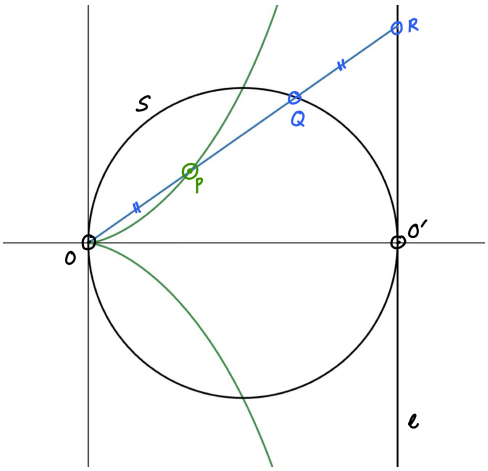
Example 1.2.13 (Cissoid of Diocles)**.** This curve is named after the ancient Greek mathematician Diocles. To construct it, start with a circle $S \subset \mathbb{A}^2_{\mathbb{R}}$ and a point $O \in S$. Construct the diameter OO' to S through O as well as the tangent line ℓ to S through O'. Now for each point $Q \in S$, extend the line OQ to meet ℓ in R, and mark off the point P on OQ such that $\text{dist}(OP) = \text{dist}(QR)$. As Q varies on S, the path that P traces out is called the cissoid; see Figure 1.4a. Taking $O = (0,0)$ and S to have center $(a,0)$ and radius a for $a \in (0,\infty)$ yields the polar equation

$$r = 2a(\sec\theta - \cos\theta),$$

which is easily seen (check!) to correspond to the Cartesian description as the vanishing locus of

$$f_a(x,y) = (x^2+y^2)x - 2ay^2 \in \mathbb{R}[x,y].$$

For all nonzero values of a, this polynomial f_a defines a plane cuspidal cubic. The name of this curve is derived from the Greek κισσοειδής, which means "ivy-shaped", presumably because of the similarity to the shape of ivy leaf edges (see Figure 1.4b).



(a) Cissoid of Diocles. Made with Desmos.

(b) An ivy leaf. Picture from the internet.

Figure 1.4: Comparison of the cissoid and the edgy of an ivy leaf.

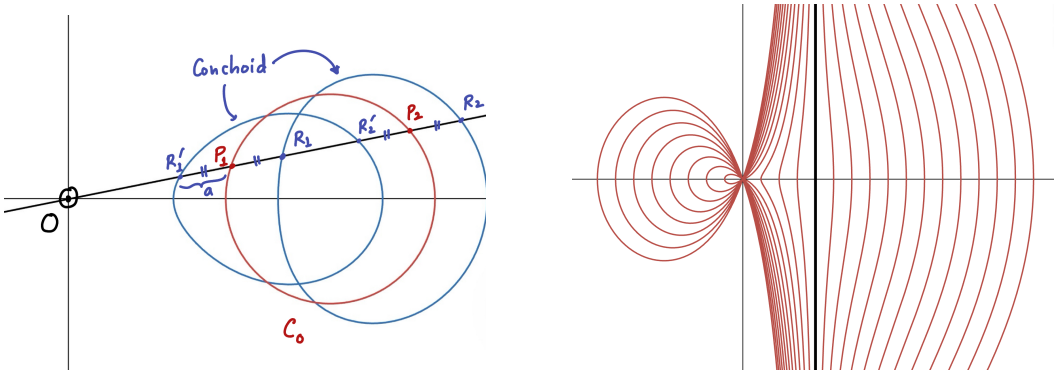
There are many other constructions of this curve: for instance, it is the curve obtained by inverting a parabola in a circle centered at its vertex, and also, if two congruent parabolae are set vertex-to-vertex, and one rolls on the other, then the vertex of the rolling parabola traces out the cissoid. It is a fun exercise, left to the reader, to try to prove these assertions.

It was a classical observation that the cissoid can be used to construct two mean proportionals to a given length $a > 0$, i.e. to construct the length $\sqrt[3]{a}$, given the length a. You are invited to explore this in Exercise 2.1.5.

Example 1.2.14 (Conchoids)**.** Our final example of a synthetic construction is that of conchoids. To construct a conchoid, you need a triple (O, C_0, a), where $O \in \mathbb{A}^2_{\mathbb{R}}$ is a point, $C_0 \subset \mathbb{A}^2_{\mathbb{R}}$ is the "base curve" and $a \in [0,\infty)$. Then the conchoid with these parameters is constructed as follows: for each point $P \in C_0$, draw the line segment OP joining O and P, and let R, R' be points on the line OP on either side of P (with say R in the direction of the ray OP from P) satisfying

$$\text{dist}(PR) = \text{dist}(PR') = a.$$

As P varies on C_0, the points R and R' trace out a curve, and this is the curve we call the conchoid. (Sometimes the locus traced by either R or R' is also called the conchoid.) See Figure 1.5a.



(a) Conchoid with base curve a circle.

(b) Various conchoids of Nichomedes.

Figure 1.5: Conchoids of various forms. Pictures made with Desmos.

If we set $O = (0, 0)$ and suppose that C_0 is given by the polar equation $r = f(\theta)$ for some function f, the the conchoid has polar equation

$$r = f(\theta) \pm a.$$

For instance, taking C_0 to be the line $x = t$ yields the curve called the conchoid of Nichomedes, and it is easy to see (check!) that it has the Cartesian description as the vanishing locus of

$$f(x, y) = (x - t)^2(x^2 + y^2) - a^2x^2 \in \mathbb{R}[x, y].$$

See Figure 1.5b for a plot of conchoids for various values of the parameters. The name comes from the Greek word κόγχη meaning "conch" or "shell"–I'll let you be the judge of whether this curve resembles the shape of a conch.

The conchoid of Nichomedes constructed with appropriate parameters can be used to trisect a given angle. You are invited to prove this in Exercise 2.1.5.

Many more examples of such synthetic constructions can be found in Brieskorn and Knörrer's *Plane Algebraic Curves*, [1, Chapter I].

1.3 06/14/24 - Parametric Curves

Today we'll discuss parametrization of curves, and what you can do with them.

Example 1.3.1. Given a field k and $u, v, w, z \in k$ with not both u, w zero, you can look at the subset given parametrically by

$$C := \{(ut + v, wt + z) : t \in k\} \subset \mathbb{A}^2_k.$$

This is the line C_ℓ defined by the polynomial

$$\ell(x, y) := wx - uy - wv + uz \in k[x, y].$$

Conversely, any line ℓ can be similarly parametrized (this uses that ℓ is not constant!).

Example 1.3.2. For any field k, the parametrization (t, t^2) traces the parabola $y - x^2 = 0$.

Example 1.3.3. Take $k = \mathbb{R}$ and the subset

$$C := \{(t^2, t^2 + 1) : t \in \mathbb{R}\} \subset \mathbb{A}^2_{\mathbb{R}}.$$

This is the ray defined by $y - x - 1 = 0$ and $x \geq 0$. This example shows that a "quadratic" parametrization can give rise to a linear curve, and the image of a parametrization of this sort need not be an entire algebraic curve, even if it is part of one.

One might argue that the above phenomenon occurs only because t^2 cannot be negative in $\mathbb{R}$, i.e. that $\mathbb{R}$ is not algebraically closed. However, as the following example shows, the same thing can happen also over any field.

Example 1.3.4. For any field k, the subset

$$C := \left\{ \left(\frac{t+1}{t+3}, \frac{t-2}{t+5} \right) : t \in k \smallsetminus \{-3, -5\} \right\} \subset \mathbb{A}^2_k$$

traces out the hyperbola defined by

$$f(x, y) = 2xy + 5x - 4y - 3 \in k[x, y],$$

except for the point $(1, 1)$, i.e.

$$C = C_f \smallsetminus \{(1, 1)\}.$$

As we shall see, this is the typical situation–that over an algebraically closed fied k, a rational parametrization of an algebraic curve C can miss at most one point–more on that next time.

Here's one example of a thing we can *do* with parametrizations.

Theorem 1.3.5 (Primitive Pythagorean Triples)**.** If $X, Y, Z \in \mathbb{Z}$ are pairwise coprime positive integers such that $X^2 + Y^2 = Z^2$, then there are coprime integers m, n of different parity such that $m > n > 0$ and either (X, Y, Z) or (Y, X, Z) is $(m^2 - n^2, 2mn, m^2 + n^2)$.

Of course, this result can be used to produce or characterize *all* Pythagorean triples, not just primitive ones (how?).

Proof. Over any field k (of characteristic other than 2 for simplicity), we can parametrize the circle C defined by $x^2+y^2-1 \in k[x,y]$ by projection from the point $(-1,0)$. In other words, for each $t \in k$, we may look at the line through $(-1,0)$ with slope t, which is given by the vanishing of $y - t(x+1)$, and consider its intersection with the circle C. We can now solve the system of equations

$$\begin{aligned} x^2 + y^2 - 1 &= 0 \\ y - t(x+1) &= 0 \end{aligned}$$

by substituting the expression for y from the second line in the first to get

$$0 = x^2 + t^2(x+1)^2 - 1 = (x+1)\left((1+t^2)x - (1-t^2)\right).$$

One of the roots of this quadratic equation is the expected $x = -1$, and, as long as $1 + t^2 \neq 0$, the other root is

$$x = \frac{1-t^2}{1+t^2},$$

which yields the point

$$\left(\frac{1-t^2}{1+t^2}, \frac{2t}{1+t^2}\right) \in C.$$

This recipe tells us that, in fact, this is a parametrization of all of C–except the point $(-1,0)$ itself, i.e.

$$\left\{\left(\frac{1-t^2}{1+t^2}, \frac{2t}{1+t^2}\right) : t \in k, 1+t^2 \neq 0\right\} = C \smallsetminus \{(-1,0)\}.$$

Make sure you understand this! Of course, this is the familiar "half-angle" parametrization of the circle, i.e. we have the trigonometric identities

$$\cos\theta = \frac{1-\tan^2\theta/2}{1+\tan^2\theta/2} \text{ and } \sin\theta = \frac{2\tan\theta/2}{1+\tan^2\theta/2}.$$

See Figure 1.6.

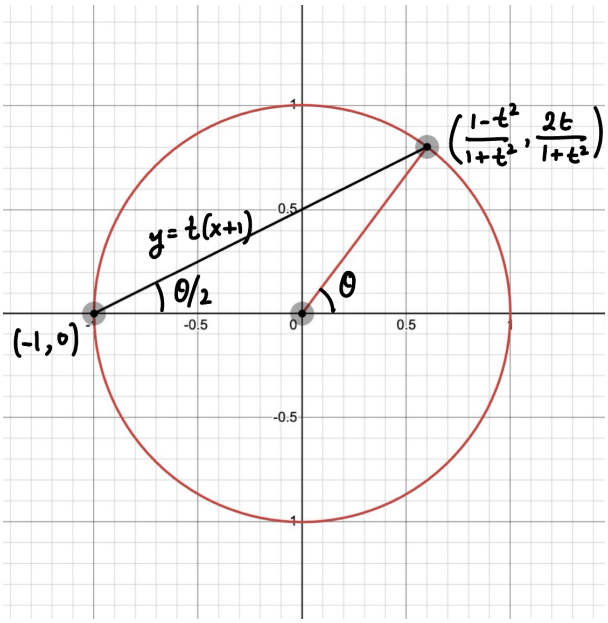


Figure 1.6: Parametrizing the circle $x^2 + y^2 = 1$.

Now, let's specialize to the case $k = \mathbb{Q}$. If X, Y, Z are as in the statement, then the point

$$(x, y) := \left(\frac{X}{Z}, \frac{Y}{Z}\right) \in C(\mathbb{Q}) \smallsetminus \{(-1, 0)\},$$

so there is a $t \in \mathbb{Q}$ such that

$$\left(\frac{X}{Z}, \frac{Y}{Z}\right) = \left(\frac{1-t^2}{1+t^2}, \frac{2t}{1+t^2}\right).$$

Then $0 < t < 1$ because $X, Y > 0$. Write $t = m/n$ for some positive coprime integers m, n with $m > n > 0$ to get

$$\left(\frac{X}{Z}, \frac{Y}{Z}\right) = \left(\frac{1-t^2}{1+t^2}, \frac{2t}{1+t^2}\right) = \left(\frac{m^2-n^2}{m^2+n^2}, \frac{2mn}{m^2+n^2}\right).$$

If m and n are of opposite parity, then the expression on the right is in lowest terms (check!) and hence we conclude that

$$(X, Y, Z) = (m^2 - n^2, 2mn, m^2 + n^2)$$

as needed. If m and n are both odd, then

$$\gcd(m^2 - n^2, m^2 + n^2) = \gcd(2mn, m^2 + n^2) = 2,$$

from which we conclude that

$$\begin{aligned} 2X &= m^2 - n^2, \\ 2Y &= 2mn, \\ 2Z &= m^2 + n^2. \end{aligned}$$

In this case, we can take

$$m' := \frac{m+n}{2} \text{ and } n' := \frac{m-n}{2},$$

which are again coprime, of different parity (check!), such that $m' > n' > 0$ and

$$(Y, X, Z) = \left((m')^2 - (n')^2, 2m'n', (m')^2 + (n')^2\right).$$

■

Let's now do some parametrizations of higher degree curves.

Example 1.3.6 (Cuspidal Cubic). For any field k, consider the set

$$C := \{(t^2, t^3) : t \in k\} \subset \mathbb{A}^2_k.$$

If we let

$$f(x, y) := y^2 - x^3 \in k[x, y],$$

then it is clear that

$$C \subset C_f.$$

To go the other direction, suppose we have a point $(p, q) \in C_f$. If $p = 0$, then $q = 0$ as well, and then $(p, q) = (t^2, t^3)$ for $t = 0$. Else, if $p \neq 0$, then it is easy to see (check!) that $(p, q) = (t^2, t^3)$ for $t := q/p$. This tells us that

$$C = C_f.$$

Again, what we are doing geometrically is that we are parametrizing points of the cuspidal cubic by the slope of the line joining the point to the cusp.

Example 1.3.7 (Nodal Cubic). For any field k, consider the curve C_f defined by the vanishing of

$$f(x,y) = y^2 - x^3 - x^2 \in k[x,y].$$

This is a nodal cubic with a node at $(0,0)$. For any $t \in k$, consider the line of slope t through the node, which has the equation $y - tx = 0$. We may now solve the system of equations

$$\begin{aligned} y^2 - x^3 - x^2 &= 0 \\ y - tx &= 0 \end{aligned}$$

as before by subsituting the second line into the first to get

$$0 = t^2x^2 - x^3 - x^2 = x^2(-x + t^2 - 1).$$

This is a cubic equation with a "double root" at $x = 0$; this captures the fact that the point $(0,0)$ is a node (how?). The third root is then the unique point of intersection of this line with the curve C_f other than the origin, and has x-coordinate $x = t^2 - 1$ and hence coordinates

$$(x,y) = (t^2 - 1, t^3 - t^2).$$

This is easily seen to be (check!) a parametrization of C_f, i.e.

$$C_f = \{(t^2 - 1, t^3 - t^2) : t \in k\}.$$

The above examples lead us to ask the following natural questions:

Question 1.3.8. Does every curve $C \subset \mathbb{A}^2_k$ admit a rational parametrization? In other words, given any curve $C \subset \mathbb{A}^2_k$, are there rational functions $u(t), v(t) \in k(t)$ such that

$$C = \{(u(t), v(t)) : t \in k \smallsetminus S\},$$

where $S \subset k$ is the finite set of poles of $u(t)$ and $v(t)$?

Question 1.3.9. Is every subset of $\mathbb{A}^2_k$ given parametrically by rational functions an algebraic curve? In other words, given any $u(t), v(t) \in k(t)$ and S as before, can we always find an $f(x,y) \in k[x,y]$ such that

$$\{(u(t), v(t)) : t \in k \smallsetminus S\} = C_f?$$

The answer to Question 1.3.8 is "yes" if C is a line (Example 1.3.1), "almost yes" if C is a conic, and "no, in general" if C has higher degree. Here's what the "almost yes" means: it means that if C is a conic and $C(k) \neq \emptyset$, then given any point $P \in C(k)$, there is a parametrization of $C(k) \smallsetminus P$ (by projection from the point P to any line not containing P, as in the proof of Theorem 1.3.5), and in some cases we may have a complete parametrization of $C(k)$ as well[6], as in Example 1.3.2. For curves of higher degree, the situation is drastically different: *most* curves of higher degree (in some sense of the word) do not admit rational parametrizations. However, proving this is beyond our tools at the moment. The simplest example of a curve that does *not* admit a rational parametrization is probably given by taking

$$f(x,y) := y^2 - x^3 + x \in k[x,y]$$

[6]This happens precisely when $\overline{C} \smallsetminus C$ contains a k-rational point, where $\overline{C} \subset \mathbb{P}^2_k$ is the projective closure of C. If you don't know what this means, you can ignore it now.

when $\operatorname{ch} k \neq 2$. In Exercise 2.2.1, you will be guided through a proof of this result, at least when $\operatorname{ch} k = 0$.

The answer to Question 1.3.9 is also "no", at least the way it is currently stated, as Examples 1.3.3 and 1.3.4 illustrate. However, the claim actually admits a very nice salvage; as it turns out, we can always find an f such that $C \subset C_f$, and at least when k is algebraically closed (a notion to be discussed soon), either C is all of C_f or all of C_f except perhaps one point. We will not prove this general statement here, although see Remark 1.3.11.

Given u and v, finding such an f as in Question 1.3.9 amounts to "eliminating" t from the system of equations

$$\begin{aligned} u(t) - x &= 0 \\ v(t) - y &= 0. \end{aligned}$$

This is the beginning of a vast subject called elimination theory; we won't get into the general theory here, and only discuss specific examples. Let's start with one.

Example 1.3.10 (Student Example)**.** For any field k, consider the curve given parametrically as

$$C = \{(t^3 - 2t^2 + 7, t^2 + 1) : t \in k\} \subset \mathbb{A}^2_k.$$

To produce such an f, perform Euclid's algorithm on the polynomials

$$\begin{aligned} A &= t^3 - 2t^2 + 7 - x \\ B &= t^2 + 1 - y \end{aligned}$$

in the polynomial ring $K[t]$ where $K = k(x, y)$ is the field of rational functions in two variables x and y. The algorithm runs to give us

$$\begin{aligned} A &= Bq_1 + r_1, \\ B &= r_1 q_2 + r_2, \text{ and} \\ r_1 &= r_2 q_3, \end{aligned}$$

where

$$\begin{aligned} q_1 &= t - 2, & r_1 &= (y-1)t - (x + 2y - 9), \\ q_2 &= \frac{1}{y-1}t + \frac{x + 2y - 9}{(y-1)^2}, & r_2 &= \frac{(x + 2y - 9)^2 - (y-1)^3}{(y-1)^2}, \end{aligned}$$

and $q_3 = r_1 r_2^{-1}$. We claim that taking

$$f(x, y) = (x + 2y - 9)^2 - (y - 1)^3 \in k[x, y]$$

suffices in the sense that at least $C \subset C_f$. To see this, use backward substitution in Euclid's algorithm to obtain the polynomial identity

$$f = P \cdot A + Q \cdot B \in k[x, y, t]$$

where

$$\begin{aligned} P &= -(y-1)t - (x + 2y - 9), t \text{ and} \\ Q &= (y-1)t^2 + (x - 7)t + y^2 - 2x - 6y + 19. \end{aligned}$$

This identity tells us that if for some $x, y, t \in k$ we have $(x, y) = (t^3 - 2t^2 + 7, t^2 + 1)$, then $A = B = 0$ and hence $f(x, y) = 0$, proving that $C \subset C_f$. Note that

$$f(x,y) = \det \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ -2 & 1 & 0 & 1 & 0 \\ 0 & -2 & 1-y & 0 & 1 \\ 7-x & 0 & 0 & 1-y & 0 \\ 0 & 7-x & 0 & 0 & 1-y \end{bmatrix}.$$

(Where on earth did this matrix come from?) In this case, we have in fact that $C = C_f$ when k is algebraically closed; you are invited to solve the mystery of this matrix and show this last result in Exercise 2.2.4. Get Desmos to plot the curve C of Example 1.3.10 over $k = \mathbb{R}$. Geometrically, we are taking the intersection of the surfaces in (x, y, t) space defined by the vanishing of A and B and projecting the resulting curve to the (x, y)-plane–can you get Desmos 3D to illustrate this?

Here's a slightly more advanced explanation that I do not expect you to fully understand right now; I include it for the sake of completeness and for when you revisit this topic later.

Remark 1.3.11. Suppose we are given a parametrization of the form

$$C = \{(u(t), v(t)) : t \in k \smallsetminus S\}$$

for some rational functions $u(t), v(t) \in k(t)$ and finite set S of all poles of $u(t)$ and $v(t)$; for the sake of nontriviality, we'll assume that $S \subsetneq k$. Write

$$u(t) = \frac{p(t)}{q(t)} \text{ and } v(t) = \frac{r(t)}{s(t)}$$

for some $p, q, r, s \in k[t]$ with $qs \neq 0$ and $(p, q) = (r, s) = (1)$. Consider the elements

$$A := p - xq \text{ and } B := r - ys$$

of $k[x, y, t] \subset K[t]$ where $K = k(x, y)$. Now consider the ideal $(A, B) \subset K[t]$. Since $K[t]$ is a Euclidean domain and hence a PID, either $(A, B) = (q)$ for some $q \in K[t]$ of positive degree, or $(A, B) = (1)$. In fact, the former case cannot happen, although we don't quite yet have the tools to prove this.[7] It follows that the Euclidean algorithm can be used as above to produce $P, Q \in k[x, y, t]$ and nonzero[8] $f \in k[x, y]$ such that

$$f = P \cdot A + Q \cdot B \in k[x, y, t]. \tag{1.1}$$

The polynomial f then cannot be constant: if it were a nonzero constant c, then we could take any value of $t \in k \smallsetminus S$ and substitute $x = u(t), y = v(t)$ in (1.1) to produce the contradiction $c = 0$. It follows as before that

$$C \subset C_f.$$

[7] Here's a proof: if A and B had a common factor $q \in K[t]$ of positive degree, then there would be an $\alpha \in \overline{K} = \overline{k(x,y)}$ such that $p(\alpha) - xq(\alpha) = r(\alpha) - ys(\alpha) = 0$. Now, we claim that $q(\alpha) \neq 0$. Indeed, if $q(\alpha) = 0$, then $p(\alpha) = 0$ as well, but already there are $m, n \in k[t]$ such that $mp + nq = 1$, so plugging in $t = \alpha$ would give $0 = 1$, which is false. Similarly, $s(\alpha) \neq 0$. Therefore, in $K(\alpha)$, we have

$$x = \frac{p(\alpha)}{q(\alpha)} \text{ and } y = \frac{r(\alpha)}{s(\alpha)}.$$

Therefore, $k(\alpha) \supset k(x, y)$ is a finite algebraic extension, but that cannot happen because the transcendence degree of $k(x, y)$ over k is 2. Alternatively, more "elementary" proofs can be given using the theory of Gröbner bases.

[8] This uses that $(A, B) = (1)$ in $K[t]$.

In fact, if f is chosen to be of minimal degree such that an equation like (1.1) holds (e.g. such as when f is coprime to P and Q–which we always do by cancelling common factors), then this f is none other than the resultant of A and B with respect to t, i.e. $f = \mathrm{Res}_t(A, B)$.

Finally, it is not always true that $C_f \subset C$, although if k is algebraically closed then C is either all of C_f or C_f minus at most one point; we certainly don't have the tools to prove this (at least at this level of generality) either.[9]

[9]Here's a proof: the rational parametrization amounts to a morphism

$$\varphi : \mathbb{A}^1_k \smallsetminus S \to C_f$$

which extends by smoothness of $\mathbb{P}^1_k$ to a morphism

$$\varphi : \mathbb{P}^1_k \to \overline{C}_f \subset \mathbb{P}^2_k,$$

where $\overline{C}_f$ is the projective closure of $\mathbb{P}^2_k$. Since, by assumption, φ is not constant, it follows from the general theory of curves that this morphism is surjective on k-points. Note that any point in S must map to $\overline{C}_f \smallsetminus C_f$ by the hypothesis that S is the set of poles of $u(t)$ and $v(t)$. If we let ∞ denote the unique k-point of $\mathbb{P}^1_k \smallsetminus \mathbb{A}^1_k$, then we have two cases: either $\varphi(\infty) \in \overline{C}_f \smallsetminus C_f$, in which case it follows that $\varphi : \mathbb{A}^1_k \smallsetminus S \to C_f$ is surjective on k-points, or $\varphi(\infty) \in C_f$, in which case $\varphi : \mathbb{A}^1_k \smallsetminus S \to C_f$ is surjective onto $C_f(k) \smallsetminus \{\varphi(\infty)\}$.

1.4 06/17/24 - Changes of Coordinates, Nonempty Curves

1.4.1 Affine Changes of Coordinates

Definition 1.4.1. An affine change of coordinates is a transformation

$$\phi : \mathbb{A}^2_k(x', y') \to \mathbb{A}^2_k(x, y)$$

of the form

$$(x, y) = \phi(x', y') = (ax' + by' + p, cx' + dy' + q),$$

for some $a, b, c, d, p, q \in k$, where $ad - bc \neq 0$.

Here $\mathbb{A}^2_k(x', y')$ is just the plane $\mathbb{A}^2_k$, which we think of as having coordinates x', y' (and similarly for $\mathbb{A}^2_k(x, y)$). The $ad - bc \neq 0$ condition guarantees that ϕ is invertible (why?). Affine changes of coordinates comprise of a linear map following by a translation; in particular, the image $\phi(0, 0) = (p, q)$ of the "origin" $(0, 0) \in \mathbb{A}^2_k$ can be any point, i.e. all points look the same (see also Remark 1.1.18).

Note that such a transformation induces a map on the polynomial rings in the opposite direction, i.e. we have a ring homomorphism (even a k-algebra homomorphism)

$$\phi^* : k[x, y] \to k[x', y'], \quad x \mapsto ax' + by' + c, y \mapsto cx' + dy' + q$$

which records the same information. For instance, ϕ is an isomorphism iff ϕ^* is. The reason for this switching of direction, also called "contravariance," is that you should think of $k[x, y]$ as the ring of polynomial functions $f : \mathbb{A}^2_k \to k$, so a coordinate transformation $\phi : \mathbb{A}^2_k(x', y') \to \mathbb{A}^2_k(x, y)$, or more properly ϕ^*, takes a function $f : \mathbb{A}^2_k(x, y) \to k$ to the function

$$\phi^* f = f \circ \phi : \mathbb{A}^2_k(x', y') \to k$$

obtained via precomposition. (This is the ultimate root of all contravariance in algebraic geometry.) Of course, thinking of polynomials as functions is not *quite* right, as you are invited to explore in Exercise 2.2.6; however, this suffices to get good intuition.

Here are a few things you can do with these: check that given any point $(p, q) \in \mathbb{A}^2_k$ and line ℓ through (p, q), there is an affine change of coordinates $\phi : \mathbb{A}^2_k(x', y') \to \mathbb{A}^2_k(x, y)$ such that $\phi(0, 0) = (p, q)$ and $\phi^{-1}\ell = C_x$, i.e. such that in the coordinate system (x', y'), the point (p, q) moves to the origin and the line ℓ moves to the y-axis C_x. We shall often define things in this course in good coordinate systems–it is then *your* job to check that these definitions are invariant under affine changes of coordinates. You are invited to play with the transformation of conics under affine changes of coordinates in Exercise 2.1.6.

1.4.2 Algebraically Closed Fields

As we have seen many times previously, it may very well happen over an arbitrary (even infinite) field k that the vanishing locus $C_f \subset \mathbb{A}^2_k$ of a polynomial function corresponding to a nonconstant polynomial $f \in k[x, y]$ is just empty. One example of this situation is when

$$f(x, y) = x^n + a_1 x^{n-1} + \cdots + a_n \in k[x, y],$$

i.e. that f is a polynomial of x alone. In this case, the corresponding locus C_f is nonempty iff this equation has a root in k, in which case C_f is the union of some vertical lines (see Remark 1.1.12). This suggests that the problem lies already in finding solutions to polynomial in one variable.

Definition 1.4.2. A field k is said to be algebraically closed if for every nonconstant polynomial $f(x) \in k[x]$, there is a root of f in k, i.e. there is an $\alpha \in k$ such that $f(\alpha) = 0$.

Example 1.4.3. The fields $\mathbb{Q}, \mathbb{R}$ and $\mathbb{F}_q$ for any q are not algebraically closed (why?).

Here are two facts which I will take for granted–these are important theorems in their own right, but this course is perhaps not the right place for them.

Theorem 1.4.4 (Fundamental Theorem of Algebra)**.** The field $\mathbb{C}$ is algebraically closed.

Theorem 1.4.5. Given any field k, there is an algebraically closed field k' containing k.

Theorem 1.4.5 says that every field k can be embedded into some algebraically closed one, although in many different ways in general.[10] This theorem says that we lose little when passing to algebraically closed fields, even when working in positive characteristic. The "smallest"[11] algebraically closed field containing k is often called the algebraic closure of k, and is often denoted $\overline{k}$; then the condition of being algebraically closed reads $k = \overline{k}$. This is notation I will occasionally slip and use, although we don't really need to dwell on the notion of algebraic closures at the moment.

One last thing to think about: can an algebraically closed field be finite? You are invited to explore this in Exercise 2.2.8. The following lemma might help.

Lemma 1.4.6. Let k be an algebraically closed field. If $f(x) \in k[x]$ is a polynomial such that $f(\alpha) = 0$ for all $\alpha \in k$, then f is the zero polynomial.

Proof. The polynomial $f + 1$ has no roots in k and is hence a constant polynomial. ∎

In fact, the condition of being algebraically closed is sufficient but not necessary; this result is, of course, the one-dimensional analog of Exercise 2.2.6. This result now allows us to prove nonemptiness results for curves.

Theorem 1.4.7. If $C \subset \mathbb{A}^2_k$ is a curve over an algebraically closed field k, then $C(k) \neq \emptyset$.

Proof. Suppose $C = C_f$ for some nonconstant $f(x, y) \in k[x, y]$. Write

$$f(x, y) = a_n(x)y^n + a_{n-1}(x)y^{n-1} + \cdots + a_0(x)$$

for some integer $n \geq 0$ and polynomials $a_0(x), \ldots, a_n(x) \in k[x]$ with $a_n(x) \neq 0$. If $n = 0$, then f is a polynomial of x alone; since f is nonconstant and k is algebraically closed, we may pick a root $\alpha \in k$ of this polynomial and any $\beta \in k$ whatsoever to give us the point $(\alpha, \beta) \in C$. If $n \geq 1$, then Lemma 1.4.6 gives us an $\alpha \in k$ such that $a_n(\alpha) \neq 0$; then the polynomial $f(\alpha, y) \in k[y]$ is nonconstant, so again, since k is alegbraically closed, there is a root $\beta \in k$ of $f(\alpha, y)$, giving us again $(\alpha, \beta) \in C$. ∎

[10] This is a subtlety which we will not have the need to discuss right now, and a true discussion of which belongs to algebra courses anyway.

[11] What would that mean?

This statement–every algebraic curve $C \subset \mathbb{A}^2_k$ is nonempty–is a characterization of algebraically closed fields, although not an awfully useful one. In fact, as you can check, the proof gives us more: the proof above shows that if C is not already the union of finitely many vertical lines, then for all but finitely many values of a (namely the roots of $a_n(x)$, if any), the curve C will intersect the vertical line $x = a$. In particular, if k is infinite (see Exercise 2.2.8), then this argument shows that $C(k)$ must be infinite as well. (So we are leaving behind the nonsense of a curve being finitely many points as well.) In Exercise 2.2.7, you are invited to discuss whether the complement $\mathbb{A}^2_k \smallsetminus C$ of C in $\mathbb{A}^2_k$ is infinite as well. The picture is therefore somewhat easier to understand over algebraically closed fields than over general fields–this is the reason that we shall essentially restrict ourselves to working with algebraically closed fields from now on.

Example 1.4.8. Considering the hyperbola defined by the vanishing of $f(x, y) = xy - 1$ and taking the line $x = 0$ shows that it is not necessarily true than an algebraic curve C intersects *every* vertical line. Somehow, the point of intersection of $f(x, y) = xy - 1$ with $x = a$ "moves to infinity" as $a \to 0$; this is a situation we will rectify in projective space, where every curve will intersect every other. More on that soon!

1.5 06/19/24 - Irreducibility I and Unique Factorization I

Last time, we showed that if $C \subset \mathbb{A}^2_k$ is an algebraic curve over an algebraically closed field k, then C is nonempty (and, in fact, infinite). Let's record this fact here, since I left some of it to you as an exercise.

Lemma 1.5.1. If k is an algebraically closed field, then any curve $C \subset \mathbb{A}^2_k$ is infinite.

Henceforth, we will always assume that our base field k is algebraically closed; this will simplify life for us tremendously. If time permits, we will return to non algebraically closed fields towards the end of the course.

1.5.1 Irreducibility I

Today I want to spend some more time relating the algebra of $k[x, y]$ to the geometry of curves in $\mathbb{A}^2_k$. Consider the following parallel definitions:

Definition 1.5.2. Let R be a ring.

(a) An element $f \in R$ is said to be irreducible if it is not zero, not a unit, and if $f = gh$ for some $g, h \in R$, then either g or h is a unit.
(b) An element $f \in R$ is said to be a prime if it is not zero, not a unit, and if $f|gh$ for some $g, h \in R$, then either $f|g$ or $f|h$.

Definition 1.5.3. A curve $C \subset \mathbb{A}^2_k$ is said to be irreducible if whenever $C = D \cup E$ for curves $D, E \subset \mathbb{A}^2_k$, then either $D = C$ or $E = C$.

Remark 1.5.4. The condition in Definition 1.5.2(b) says that a nonzero $f \in R$ is prime iff the principal ideal $(f) \subset R$ generated by f is a prime ideal. If R is an integral domain, then every prime is irreducible, but the converse need not hold in general–see Exercise 2.3.2. The converse does, however, hold if R is a UFD; see Proposition 1.5.8.

What is the relationship between the irreducibility of a polynomial and that of the curve defined by it? In light of Proposition 1.1.7, one could reasonably make

Conjecture 1.5.5. Give a nonconstant polynomial $f \in k[x, y]$, the algebraic curve C_f defined by f is irreducibe iff f is.

However, a moment's reflection shows that this cannot be correct as stated. For instance, if $f(x, y) = x^2$, then f is not irreducible, but the algebraic curve C_f is a line, which is irreducible thanks to Exercise 2.1.7 (how?). One correct salvage of this statement would be

Theorem 1.5.6. If an $f \in k[x, y]$ is irreducible, then C_f is irreducible, and conversely if $C \subset \mathbb{A}^2_k$ is an irreducible curve, then there is an irreducible $f \in k[x, y]$ such that $C = C_f$.

Our next order of business is to develop tools to prove Theorem 1.5.6.

1.5.2 Unique Factorization I

The first fact we would need is that $k[x, y]$ is UFD. Let's recall the definition of such a ring.

Definition 1.5.7. A ring is said to be a unique factorization domain, abbreviated UFD, if R is a domain[a], if every nonzero nonunit in it is a product of finitely many irreducible elements, and the decomposition into irreducible factors is unique up to order and multiplication by units. In other words, a domain R is a UFD if given any nonzero nonunit $f \in R$, there is an integer $n \geq 1$ and irreducible elements $f_1, \ldots, f_n \in R$ such that

$$f = f_1 f_2 \cdots f_n$$

and if there is some other integer $m \geq 1$ and irreducible elements $g_1, \ldots, g_m \in R$ such that

$$f = f_1 f_2 \cdots f_n = g_1 g_2 \cdots g_m,$$

then we must have $n = m$, a bijection $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ and units $c_1, \ldots, c_n \in R^\times$ such that for all i with $1 \leq i \leq n$ we have $c_i g_i = f_{\sigma(i)}$.

[a] This means the same thing as "integral domain".

A field is vacuously a UFD–there *are* no nonzero nonunits. Here's one way to identify UFD's.

Proposition 1.5.8. Let R be a domain. Then the following are equivalent:

(a) R is a UFD.
(b) Every nonzero nonunit in R is a product of finitely many irreducible elements and each irreducible element is prime.
(c) Every nonzero nonunit in R is a product of finitely many prime elements.

Proof.

(a) $\Rightarrow$ (b) We only need to show that every irreducible in a UFD is prime; I leave this to the reader.

(b) $\Rightarrow$ (c) Clear.

(c) $\Rightarrow$ (a) Since primes are irreducible, all that remains to be shown is uniqueness of factorization. For this, we first show that if (c) holds, then every irreducible element is prime: indeed, if $f \in R$ is irreducible and we write $f = p_1 \cdots p_n$ for some integer $n \geq 1$ and primes $p_1, \ldots, p_n$, then irreducibility of f tells us (how?) that $n = 1$ and $f = p_1$ is prime. We show uniqueness of the irreducible decomposition of a nonzero nonunit $f \in R$ by inducting on the minimal number $n \geq 1$ of irreducible factors in such a decomposition. For the base case $n = 1$, our $f = f_1$ itself is irreducible, so if $f = g_1 \cdots g_m$ for some $m \geq 1$ and irreducibles $g_j \in R$, then irreducibility of f tells us (how?) that $m = 1$ and $f = g_1$. Inductively, if we have for some $m \geq n \geq 2$ that

$$f = f_1 \cdots f_n = g_1 \cdots g_m,$$

then primality of g_1 tells us that $g_1 \mid f_j$ for some j with $1 \leq j \leq n$, so let $c_1 \in R$ be such that $c_1 g_1 = f_j$. Now f_j is irreducible and g_1 is not a unit, so c_1 must be a unit. Therefore, cancelling f_1 from both sides, we are left with

$$f_1 \cdots f_{j-1} f_{j+1} \cdots f_n = (c_1^{-1} g_2) g_3 \cdots g_m,$$

so we are done by induction (how?).

■

The one technique we have seen at Ross so far of showing that a domain is a UFD is to work with Euclidean functions. Let's define those now.

Definition 1.5.9.

(a) Let R be a domain. A Euclidean function on R is a map $d : R \setminus \{0\} \to \mathbb{Z}_{\geq 0}$ such that for all $A, B \in R$ with $B \neq 0$, there are $q, r \in R$ such that

$$A = Bq + r$$

and either $r = 0$ or $d(r) < d(B)$.

(b) A domain R is said to be a Euclidean domain if it admits a Euclidean function.

Here are a few key examples.

Example 1.5.10.

(a) For $R = K$ a field, the function $d \equiv 1$ is Euclidean.
(b) For $R = \mathbb{Z}$, the function $d(n) = |n|$ is Euclidean.
(c) For $R = \mathbb{Z}[i]$ or $R = \mathbb{Z}[\omega]$, the norm function $d(\alpha) = \mathrm{N}(\alpha)$ is Euclidean.
(d) For $R = K[t]$, the polynomial ring over the field K, the function $d(f) = \deg f$ is Euclidean.
(e) For $R = K[\![t]\!]$, the $d(f) = \mathrm{ord}_t f$ taking a power series to the highest power of t dividing it is Euclidean.

The key reason we like Euclidean domains is

Theorem 1.5.11. Every Euclidean domain is a UFD.

Proof Sketch. The key idea is that Euclidean functions allow us to perform the Euclidean algoritheorem to produce the greatest common divisor of any two elements, although I do want to warn you that the proof at this level of generality needs some work. See [2] for a direct proof, or any algebra textbook. ■

The result that we really need, however, is that the ring $R = k[x, y]$ is a UFD. This cannot be done using Theorem 1.5.11–indeed, the ring $k[x, y]$ is not a Euclidean domain.[12] How do we proceed then?

We will prove

Theorem 1.5.12. If R is a UFD, then so is the polynomial ring $R[t]$.

Remark 1.5.13. In fact, one can check that if R is any ring such that $R[t]$ is a UFD, then so is R. (Prove this!) This makes the statement in Theorem 1.5.12 an "if-and-only-if" statement.

The way we will we use Theorem 1.5.12 is via

Corollary 1.5.14. If R is a UFD, then so is the polynomial ring $R[t_1, \ldots, t_n]$ for each $n \geq 1$. In particular, for any field k, the ring $k[x, y]$ is a UFD.

[12] This is because Euclidean domains are principal ideal domains, while $k[x, y]$ is not one. If you don't know what this means, you can ignore this comment. If you do know what this means, there are also examples of principal ideal domains which are not Euclidean, but such rings are harder to come by. The simplest examples I know of are $R = \mathcal{O}_{\mathbb{Q}[\sqrt{-19}]}$ and $R = \mathbb{R}[x, y]/(x^2 + y^2 + 1)$, but proving these claims needs some work.

To prove Theorem 1.5.12, we need some preparation. In what follows, we will fix a UFD R and let $K = \operatorname{Frac} R$ be its fraction field, so that $K = \{p/q : p, q \in R, q \neq 0\}$. Also, for any $f \in R[t]$ and $n \geq 0$, we will denote the coefficient of t^n by $[t^n]f$. The first order of business is to show that $R[t]$ is a domain.

Lemma 1.5.15.

(a) If R is a domain, then so is $R[t]$.
(b) If $p \in R$ is prime, then p is also prime in $R[t]$.

Proof.

(a) Write $0 \neq f, g \in R[t]$ as $f = \sum_{i=0}^{n} a_{n-i}t^i$ and $g = \sum_{j=0}^{m} b_{m-j}t^j$ for some $m, n \geq 0$, with $a_i, b_j \in R$ and $a_0 \neq 0$ and $b_0 \neq 0$. Since R is a domain, $[t^{m+n}]fg = a_0b_0 \neq 0$, so $fg \neq 0$.
(b) We can either reduce to (a) by noticing that $R[t]/(p) \cong (R/p)[t]$ (how?), or argue directly as before: if $f \in R[t]$ is such that $p \nmid f$ and we write $f = \sum_{i=0}^{n} a_{n-i}t^i$ for some $n \geq 0$ and $a_i \in R$ with $a_0 \neq 0$, then there is some i with $0 \leq i \leq n$ and $p \nmid a_i$; let i_0 be the smallest such i. Similarly, if $p \nmid g$, then write $g = \sum_{j=0}^{m} b_{m-j}t^j$ as in (a) and pick the smallest j_0 with $0 \leq j_0 \leq m$ such that $p \nmid b_{j_0}$. Then, $p \nmid [t^{(m-i_0)+(n-j_0)}]fg$ (check!) so that $p \nmid fg$.

■

Definition 1.5.16. A polynomial $f \in R[t]$ is said to be primitive if the following equivalent conditions hold:

(a) If $\alpha \in R$ is such that $\alpha \mid f$, then α is a unit.
(b) There is no prime $p \in R$ such that $p \mid f$, i.e. $p \mid [t^i]f$ for all $i \geq 0$.
(c) The greatest common divisor of all coefficients of f is (1).

Note that 0 is not primitive. Any $f \in K[t]$ can be written as $f = \operatorname{cont}(f) \cdot \tilde{f}$ for some $\operatorname{cont}(f) \in K$ and primitive $\tilde{f} \in R[t]$. If $f \neq 0$, then $\operatorname{cont}(f)$ and $\tilde{f}$ are uniquely determined up to units in R; then $\operatorname{cont}(f)$ is called the content of f, and $\tilde{f}$ is called the primitive part of f, defined uniquely only up to units in R.[13] Here are some basic properties that we will need:

Lemma 1.5.17. If $0 \neq f \in K[t]$, then

(a) $\deg \tilde{f} = \deg f$,
(b) $\operatorname{cont}(f) = f$ iff f is constant,
(c) $f \in R[t]$ iff $\operatorname{cont}(f) \in R$,
(d) if (c) holds, then f is primitive iff $\operatorname{cont}(f)$ is a unit in R, and
(e) $\tilde{\tilde{f}} = \tilde{f}$.

Proof. Left to the reader. ■

The key result that allows us to relate $R[t]$ and $K[t]$ is

[13] One way to make this precise is to say that the fractional ideal $(\operatorname{cont} f)$ of R and the (integral) ideal $(\tilde{f})$ of $R[t]$ are uniquely determined. We will not need these notions. When we assert an equality involving $\operatorname{cont}(f)$ or $\tilde{f}$, that equality will always be assumed to hold up to units.

Lemma 1.5.18 (Gauss's Lemma)**.**

(a) If $f, g \in R[t]$ are primitive, then so is fg. In general, if we have nonzero $f, g \in K[t]$, then $\operatorname{cont}(fg) = \operatorname{cont}(f)\operatorname{cont}(g)$ and $\widetilde{fg} = \tilde{f}\tilde{g}$ (up to units). The same holds for any number $f_1, \ldots, f_n$ of elements with $n \geq 1$.
(b) If $f, g \in R[t]$ are nonzero such that $f \mid g$ in $K[t]$ and f is primitive, then $f \mid g$ in $R[t]$.
(c) If $f \in R[t]$ is primitive and prime in $K[t]$, then f is prime in $R[t]$.

Proof.

(a) The general case follows by induction, so we do the case $n = 2$. If $f, g \in R[t]$ are primitive and if a prime $p \in R$ were to divide fg, then it would divide either f or g by Lemma 1.5.15(b). In general, given nonzero $f, g \in K[t]$, we have $fg = \operatorname{cont}(f)\operatorname{cont}(g) \cdot \tilde{f}\tilde{g}$, and $\widetilde{\tilde{f}\tilde{g}}$ is primitive by the first part, so by the uniqueness of this deomposition we must have $\widetilde{fg} = \tilde{f} \cdot \tilde{g}$, and hence that $\operatorname{cont}(fg) = \operatorname{cont}(f) \cdot \operatorname{cont}(g)$.
(b) If $g = fq$ for some nonzero $q \in K[t]$, then $\operatorname{cont}(g) = \operatorname{cont}(f) \cdot \operatorname{cont}(q)$. Since $f, g \in R[t]$, Lemma 1.5.17(c) tells us that $\operatorname{cont}(f), \operatorname{cont}(g) \in R$, and since f is primitive, Lemma 1.5.17(d) tells us that $\operatorname{cont}(f)$ is a unit, so that $\operatorname{cont}(q) = \operatorname{cont}(g)\operatorname{cont}(f)^{-1} \in R$, and hence by Lemma 1.5.17(c) again we conclude that $q \in R[t]$.
(c) Suppose $f \in R[t]$ is primitive and prime in $K[t]$ (and hence nonzero), and suppose $f | gh$ for some $g, h \in R[t]$. Then $f \mid gh$ also in $K[t]$, and so by primality either $f | g$ or $f | h$ in $K[t]$, and hence also in $R[t]$ by (b), showing that f is prime in $R[t]$.

■

In Lemma 1.5.18(b), we certainly need f to be primitive; a simple counterexample otherwise is given by taking $R = \mathbb{Z}$ and $f(t) = 2t$ and $g(t) = t$. We are now ready to prove Theorem 1.5.12.

Proof of Theorem 1.5.12. Suppose R is a UFD and $K = \operatorname{Frac} R$. By Proposition 1.5.8(c), it suffices to show that every nonzero nonunit $f \in R[t]$ is a product of finitely many primes. Since $f = \operatorname{cont}(f) \cdot \tilde{f}$, it suffices to show that each of $\operatorname{cont}(f)$ and $\tilde{f}$ is a product of finitely many primes in $R[t]$.[14]

Since $0 \neq \operatorname{cont}(f) \in R$ and R is a UFD, either $\operatorname{cont}(f)$ is a unit in R (and hence in $R[t]$), or it is a product of one or more primes in R. Since primes in R are primes in $R[t]$ by Lemma 1.5.15(b), it follows that $\operatorname{cont}(f)$ is a product of finitely many primes in $R[t]$.

Now consider the primitive part $0 \neq \tilde{f} \in R[t]$. Since $K[t]$ is a UFD, it follows that either $\tilde{f}$ is a unit in $K[t]$ or it is the product of one or more primes in $K[t]$. In the former case, $\tilde{f}$ is constant[15] and so since it is primitive, it must be a unit in R (by Lemma 1.5.17(b) and (d)). In the latter case, $\tilde{f}$ is the product of one or more primes in $K[t]$, say $\tilde{f} = f_1 \cdots f_n$ for some $n \geq 1$, where for $1 \leq j \leq n$, each $f_j \in K[t]$ is prime. Then using Lemma 1.5.17(e) and Lemma 1.5.18(a), we find that

$$\tilde{f} = \tilde{\tilde{f}} = \tilde{f}_1 \cdots \tilde{f}_n.$$

For each j, the element $\tilde{f}_j \in R[t]$ is primitive and prime in $K[t]$ (since it is a nonzero constant, i.e. unit, times the prime f_j in $K[t]$), and so by Lemma 1.5.18(c) is a prime in $R[t]$. Therefore, we have exhibited $\tilde{f}$ as a product of one or more primes in $R[t]$, finishing the proof. ■

[14] Note that finitely many also includes zero many–i.e. it is okay for $\operatorname{cont}(f)$ or $\tilde{f}$ to be a unit in $R[t]$, but if both are units in $R[t]$, then so is $f = \operatorname{cont}(f) \cdot \tilde{f}$.

[15] This is because the only units in $K[t]$ are constants, i.e. elements of $K^\times = K \smallsetminus \{0\}$. If you haven't seen this before, prove it!

1.6 06/21/24 - Nullstellensatz, Irreduciblity II, and Unique Factorization II

Last time, we proved that if R is a UFD, then so is $R[t]$. The same circle of ideas allows us to compare irreducibles in $R[t]$ and $K[t]$. Let's prove two results in this direction, and then return to the theory of curves to see their applications.

As before, in what follows we will take R to be a UFD and $K = \operatorname{Frac} R$ to be its fraction field.

Lemma 1.6.1.

(a) If $f \in R[t]$ is irreducible and of positive degree, then f is irreducible in $K[t]$.
(b) If $f \in R[t]$ is primitive and irreducible in $K[t]$, then f is irreducible in $R[t]$.

Proof.

(a) In this case, f is a nonzero nonunit in $K[t]$. If $f = gh$ for $g, h \in K[t]$, then Lemma 1.5.18(a) tells us that $\tilde{f} = \tilde{g}\tilde{h}$, and then $f = (\operatorname{cont}(f) \cdot \tilde{g}) \cdot \tilde{h}$. Since f is irreducible in $R[t]$, either $\operatorname{cont}(f) \cdot \tilde{g}$ is a unit in R, in which case $\tilde{g}$ is a (nonzero) constant and hence $g \in K[t]^\times$ by Lemma 1.5.17(a), or simlarly $\tilde{h}$ is a unit in R, in which case $h \in K[t]^\times$.
(b) This is Lemma 1.5.18(c), given that the terms "prime" and "irreducible" are interchangable in $R[t]$ and $K[t]$ thanks to Proposition 1.5.8 and Theorem 1.5.12.

■

In any UFD S, we say that two elements $f, g \in S$ are relatively prime if there is no prime $p \in S$ such that $p \mid f$ and $p \mid g$.

Lemma 1.6.2. If $f, g \in R[t]$ are relatively prime in $R[t]$, then

(a) they are relatively prime in $K[t]$, and
(b) there are $a, b \in R[t]$ and $0 \neq c \in R$ such that $af + bg = c$.

Proof.

(a) If there is a prime $q \in K[t]$ such that $q \mid f$ and $q \mid g$ in $K[t]$, then by rescaling we can assume without loss of generality that $q \in R[t]$ is primitive (how?), and then Lemma 1.5.18(b) tells us that $q \mid f$ and $q \mid g$ in $R[t]$, and Lemma 1.5.18(c) tells us that q is prime in $R[t]$. This can't happen if $f, g \in R[t]$ are relatively prime in $R[t]$.
(b) This is clear from the Euclidean algorithm and backward substitution if R is a field (make sure you understand this!). In the general case, the first observation and part (a) combine to tell us that there are $a_1, b_1 \in K[t]$ and $0 \neq c_1 \in K$ such that $a_1 f + b_1 g = c_1$. Now we can simply "clear denominators": find a $0 \neq d \in R$ such that $a := a_1 \cdot d$ and $b := b_1 \cdot d$ are in $R[t]$, and $c := c_1 d \in R$.

■

Example 1.6.3. Take $R = \mathbb{Z}$ and $f(t) = t^3 + 1$ and $g(t) = t^2 - 7$. Then we can take $a = -7t + 1$ and $b = 7t^2 - t + 49$ with $c = -342$ via the identity

$$(-7t + 1)(t^3 + 1) + (7t^2 - t + 49)(t^2 - 7) = -342 = -2 \cdot 3^2 \cdot 19.$$

Note that the same polynomial identity holds over any ring R, but something special happens over $R = \mathbb{Z}/2, \mathbb{Z}/3$ and $\mathbb{Z}/19$: the polynomials f and g end up being not relatively prime. In

fact, f and g are not relatively prime in $\mathbb{Z}/p$ iff $p \in \{2, 3, 19\}$. This fascinating observation has to do with resultants again–see Remark 1.6.5.

Example 1.6.4. Consider the polynomials $f(x, y) = x^3 - 12x - y^2$ and $g(x, y) = x^2 - xy - y^2 + 5$ in $k[x, y]$ for some field k (e.g. $k = \mathbb{C}$). Applying the above procedure to $R = k[y]$ with variable $t = x$ yields

$$\begin{aligned} a_y &= (-2y^2 + 17)x + y(3y^2 - y - 22), \\ b_y &= (2y^2 - 17)x^2 + y(-y^2 + y + 5)x + (y^4 + y^3 - 46y^2 + 289), \text{ and} \\ c_y &= -y^6 - 4y^5 + 52y^4 + 27y^3 - 519y^2 + 1445. \end{aligned}$$

On the other hand, applying the above procedure to $R = k[x]$ with variable $t = y$ yields

$$\begin{aligned} a_x &= (-x)y + (x^3 - 2x^2 - 12x - 5), \\ b_x &= xy + (-x^3 + x^2 + 12x + 5), \text{ and} \\ c_x &= x^6 - 3x^5 - 23x^4 + 26x^3 + 154x^2 + 120x + 25. \end{aligned}$$

Remark 1.6.5 (Resultants)**.** If we fix integers $m, n \geq 1$, and take $R = \mathbb{Z}[a_0, \ldots, a_m, b_0, \ldots, b_n]$ with $f(t) = a_0 t^m + \cdots + a_m$ and $g(t) = b_0 t^n + \cdots + b_n$, then Lemma 1.6.2 gives us $a, b \in R[t]$ and $0 \neq c \in R$ such that $af + bg = c$.[16] The c of least such degree is (up to a negative sign perhaps) none other than the resultant $\mathrm{Res}_t(f, g)$ of f and g with respect to t, essentially because it is the "universal" polynomial which in the coefficients which tests the coprimality of f and g. This is not a hard result, but we won't need it directly, so I won't give a proof; you are invited to prove it (perhaps using the definition from Exercise 2.2.4) if you'd like. Lemma 1.6.2 then gives us the important consequence that the resultant of two polynomials can be written as a polynomial-linear combination of them with coefficients in the ring generated by *their* coefficients.

1.6.1 Finite Intersection of Curves, Nullstellensatz, and Irreducibility II

Let's now return to the theory of curves. One important consequence of Lemma 1.6.2, evident already from Example 1.6.4 is

Theorem 1.6.6 (Finite Intersection)**.** If k is any field and $f, g \in k[x, y]$ are nonconstant relatively prime polynomials, then the intersection $C_f \cap C_g$ is finite.

Proof. Applying Lemma 1.6.2 to $R = k[y]$ with variable $t = x$ yields $a, b \in k[x, y]$ and $0 \neq c \in k[y]$ such that $af + bg = c$. Therefore, if $(p, q) \in C_f \cap C_g$, then $c(q) = 0$, so q is one of the finitely many roots of c, and hence can only take on finitely many values. Reversing the roles of x and y, we conclude that p can only take on finitely many values as well, and hence $C_f \cap C_g$ is finite. ∎

This result generalizes the one from Exercise 2.1.7 (how?). Geometrically, what is happening is this: the roots of the polynomial c are (or at least include) the projections of the points in $C_f \cap C_g$ to the y-axis, and similarly for the corresponding polynomial in x. This yields a finite grid of horizontal and vertical lines, the finitely many intersection points of which contain $C_f \cap C_g$. See Figure 1.7 for an illustration of this phenomenon for the polynomials f and g of Example 1.6.4. We have now arrived at one of the most important results in this theory.

[16] Technically, you have to check that $f(t)$ and $g(t)$ are relatively prime in $R[t]$, but this follows because they are the "universal" polynomials–if they were not, then every pair of polynomials over any ring would have a common factor, which is absurd.

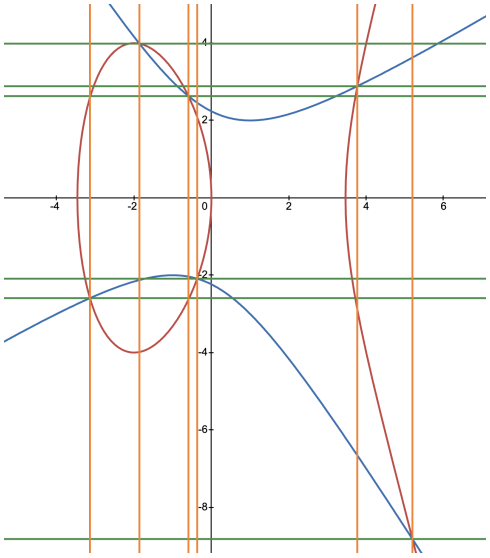


Figure 1.7: An illustration of Theorem 1.6.6 for the f and g in Example 1.6.4. The red curve is C_f, the blue curve is C_g, the green lines correspond to the roots of c_y, and the orange lines correspond to the roots of c_x. The intersection $C_f \cap C_g$ is contained in the finitely many points of the green-orange grid. Picture made with Desmos.

Theorem 1.6.7 (Hilbert's Nullstellensatz for Curves)**.** If k is an algebraically closed field, and $f, g \in k[x, y]$ are nonconstant polynomials, then $C_g \subset C_f$ iff there is some integer $n \geq 1$ such that $g \mid f^n$.

Proof. One direction is clear (which?). For the other direction, it suffices to show that if $q \in k[x, y]$ is a prime factor of g, then $q \mid f$. If there were a prime factor q for which this were not the case, then q and f would be relatively prime in $k[x, y]$, and so by Theorem 1.6.6, the intersection $C_q \cap C_f$ would be finite. But now, $C_q \subset C_g \subset C_f$ implies that $C_q \cap C_f = C_q$, which is infinite by the fact that q is nonconstant and Lemma 1.5.1.[17] ■

Note that the Nullstellensatz–German for "the theorem on the location of zeroes"–uses crucially that k is algebraically closed. We will henceforth return to our convention that k is an algebraically closed field. One important corollary we can extract is

Corollary 1.6.8. If $f, g \in k[x, y]$ are nonconstant polynomials with f irreducible, then $C_g \subset C_f$ implies $C_g = C_f$.

Proof. By Theorem 1.6.7, there is some $n \geq 1$ such that $g \mid f^n$. Then primality of f (using Corollary 1.5.14 and Proposition 1.5.8) tells us that $f \mid g$, so the easy direction of Theorem 1.6.7 implies that $C_f \subset C_g$ as needed. ■

We are now ready to prove Theorem 1.5.6, which we restate here.

[17] This is the only step where we use that k is algebraically closed.

Theorem 1.5.6. If an $f \in k[x,y]$ is irreducible, then C_f is irreducible, and conversely if $C \subset \mathbb{A}^2_k$ is an irreducible curve, then there is an irreducible $f \in k[x,y]$ such that $C = C_f$.

Proof. If f is irreducible and $C_f = C_g \cup C_h$ for nonconstant $g, h \in k[x,y]$, then Corollary 1.6.8 gives us that $C_f = C_g = C_h$, showing irreducibility of C_f. Conversely, if $C = C_{f_0} \subset \mathbb{A}^2_k$ is an irreducible curve for some $f_0 \in k[x,y]$, then we claim that there is an irreducible $f \in k[x,y]$ and an integer $n \geq 1$ such that $f_0 = f^n$. If this were not the case, we would be able to write $f_0 = gh$ for nonconstant relatively prime g, h, from which it would follow that $C = C_g \cup C_h$. Then irreducibility of C would tell us that either $C = C_g$ or $C = C_h$; suppose, without loss of generality, that $C = C_g$. Then Theorem 1.6.7 applied to the containment $C \subset C_g$ would imply that there is some $n \geq 1$ such that $f_0 \mid g^n$, which is a contradiction to the factorization $f_0 = gh$ in the UFD $k[x,y]$, since g and h are relatively prime. ∎

1.6.2 Unique Factorization II

Here's the picture that we are building to: there is a parallel between the unique factorization in $k[x,y]$ and of curves in $\mathbb{A}^2_k$, namely each curve $C \subset \mathbb{A}^2_k$ can be decomposed as a finite union of irreducible curves

$$C = C_1 \cup C_2 \cup \cdots \cup C_n,$$

and these are determined uniquely upto ordering the factors. For this, the first question we can ask is:

Question 1.6.9. To what extent does a curve $C \subset \mathbb{A}^2_k$ determine a defining polynomial $f \in k[x,y]$, i.e. a polynomial f such that $C = C_f$?

The answer here is: almost, the only problem being multiplicity. Specifically, consider

Definition 1.6.10. Let R be a UFD.

(a) If a nonzero $f \in R$ is decomposed as

$$f = c f_1^{m_1} \cdots f_n^{m_n}$$

where $c \in R^\times$ is a unit, $n \geq 1$ an integer, $f_1, \ldots, f_n \in R$ irreducibles and $m_1, \ldots, m_n \geq 1$, then we define the radical of f by

$$\mathrm{rad}(f) := f_1 \cdots f_n.$$

Note that this is well-defined up to units in R.

(b) We say that a nonzero $f \in R$ is reduced if $f = \mathrm{rad}(f)$ (up to units).

Taking $R = k[x,y]$ in this definition and given any nonconstant $f \in k[x,y]$, the radical $\mathrm{rad}(f)$ is again nonconstant, and we have that

$$C_f = C_{\mathrm{rad}(f)}.$$

Therefore, a curve C cannot distinguish a polynomial from its radical. The Nullstellensatz tells us, however, that the radical can however be recovered from the curve.

Definition 1.6.11. Given a curve $C \subset \mathbb{A}^2_k$, the subset

$$\mathbb{I}(C) := \{g \in k[x,y] \text{ nonconstant} : C \subset C_g\} \cup \{0\} \subset k[x,y]$$

is called the (vanishing) ideal of C.

We will define the term "ideal" properly next time. The key claim here is then

Theorem 1.6.12. If k is algebraically closed, and $f \in k[x,y]$ is a nonconstant polynomial, then a polynomial $g \in k[x,y]$ is in $\mathbb{I}(C_f)$ iff $\text{rad}(f) \mid g$. In particular, $\text{rad}(f)$ is uniquely determined (up to nonzero scalars) by the curve C.

Proof. If g is nonconstant, then $C_f \subset C_g$ implies by Theorem 1.6.7 that for some $n \geq 1$, we have $f^n \mid g$. Since $\text{rad}(f) \mid f^n$, we are done. Finally, $\text{rad}(f)$ is simply the nonzero polnyomial of least degree in $\mathbb{I}(C)$ (up to nonzero scalars). ■

We say that $\text{rad}(f)$ is a generator $\mathbb{I}(C)$, and call it the minimal polynomial of C.

Corollary 1.6.13 (Hilbert's Nullstellensatz for Curves, Version II)**.** Over an algebraically closed field k, there is a bijective correspondence

$$\{\text{curves } C \subset \mathbb{A}^2_k\} \dashleftarrow\dashrightarrow \{\text{nonconstant reduced } f \in k[x,y]\}/(\text{nonzero scalars})$$

given by sending an f to C_f and a curve C to its minimal polynomial.

Under this correspondence,

(a) the curve C is irreducible iff its minimal polynomial is, and
(b) the union of curves corresponds to taking the product of the minimal polynomials (and then the radical).

Also,

(c) Two nonconstant reduced polynomials define the same curve iff they are nonzero scalar multiples of each other.

This result is one of the earliest manifestations of the systematization of the parallels between algebra and geometry, which is the heart and soul of algebraic geometry. We will discuss more consequences of this bijective correspondence next time.

1.7 06/24/24 - Ideals, Irreducible Components, Degree II

Today, I want to review some algebra to express our observations from last time in a cleaner way.

1.7.1 Crash Course on Ideals

Definition 1.7.1. Let R be a ring. An ideal of R is an additive subgroup $I \subset R$ such that for all $f \in I$ and $g \in R$, we have $fg \in I$.

The terminology historically comes from thinking of ideals as "ideal numbers". In the 19th century, people came to realize that in some natural rings in number theory, such as $\mathbb{Z}[\sqrt{-5}]$, unique factorization into prime numbers failed. Kummer and Dedekind salvaged this by saying that in these number rings, or in what are now known more generally as Dedekind domains, we do get a unique factorization of numbers into *prime ideal numbers*, i.e. these objects behave the way prime numbers "ideally" would.

If $I \subset R$ is an ideal, we can define an equivalence relation on R called congruence modulo I, by saying $f \sim g$ iff $f - g \in I$. The set of equivalence classes R/I then admits a structure of a ring such that then natural surjection $R \to R/I$ is a ring homomorphism (and this determines the ring structure on R/I completely). This ring R/I is called the quotient of the ring R by the ideal I.

Example 1.7.2.

(a) In any ring R, the set $I = \{0\} \subset R$ is an ideal called the zero ideal. Similarly, $I = R$, i.e. all of R, is also an ideal. We say an ideal $I \subset R$ is a proper ideal if I is a proper subset of R, i.e $I \subsetneq R$.

(b) Given a ring R and an element $f \in R$, we define the principal ideal generated by f to be the ideal $(f) := \{g \in R : f \mid g\}$. An ideal $I \subset R$ is said to be a principal ideal if $I = (f)$ for some $f \in R$; in general, this f is not unique. (E.g. $(2) = (-2)$ in $\mathbb{Z}$.) Note that (0) is the zero ideal, whereas $(1) = R$; more generally, $(u) = R$ iff $u \in R$ is a unit.

(c) More generally, given any subset $S \subset R$, the ideal generated by S is the ideal

$$(S) = \left\{ \sum_{i=1}^{n} a_i s_i : a_i \in R, s_i \in S \right\} \subset R.$$

This is the smallest (with respect to inclusion) ideal containing S, or equivalently the intersection of all ideals containing S.

(d) Any additive subgroup $S \subset \mathbb{Z}$ is of the form (n) for some unique $n \in \mathbb{Z}_{\geq 0}$. In particular, these are all the ideals in $\mathbb{Z}$. (Proof: if $S \cap \mathbb{Z}_{>0} = \emptyset$, then $S = (0)$; else, there is a least $n \in S \cap \mathbb{Z}_{>0}$ by the well-ordering principle, and then $S = (n)$.) A ring R is said to be a principal ideal ring if every ideal of R is principal; a domain R that is a principal ideal ring is called a principal ideal domain, abbreviated PID.

In general, principal ideals don't determine generators (e.g. in $R = \mathbb{Z}/6$, we have $(2) = (4)$); however, in domains[18], principal ideals determine generators up to units.

[18] Fascinatingly, this is not quite a characterization of domains. Other rings, such as local rings, also satisfy this property. I do not know of a complete characterization of rings with this property.

Lemma 1.7.3. If R is a domain and $f, g \in R$, then $(f) = (g)$ iff there is a unit $u \in R^\times$ such that $f = ug$. In other words, a principal ideal in R is determined by, and determines, its generator up to units.

Proof. One direction is clear (which, and why?). For the other direction, by assumption, there are $u, v \in R$ such that $f = ug$ and $g = vf$. Then $f(uv - 1) = 0$, so since R is a domain, one of f and $uv - 1$ is zero. If $f = 0$, then $g = vf = 0$, and $0 = 1 \cdot 0$. Otherwise, $uv = 1$ implies $u \in R^\times$. ∎

Proposition/Definition 1.7.4. For a ring R and a proper ideal $P \subset R$, the following are equivalent:

(a) If $f, g \in R$, then $fg \in P$ implies either $f \in P$ or $g \in P$.
(b) The quotient ring R/P is a domain.

A proper ideal $P \subset R$ satisfying these equivalent conditions is called a prime ideal.

Example 1.7.5.

(a) A ring R is a domain iff $(0) \subset R$ is a prime ideal.[19]
(b) An ideal $I \subset \mathbb{Z}$ is prime iff either $I = 0$ or $I = (p)$ for some prime integer p.
(c) In general, if R is any ring, then and $0 \neq f \in R$, then f is a prime element iff (f) is a prime ideal.

In Exercise 2.3.3, you are invited to find all prime ideals of the ring $k[x, y]$ when k is algebraically closed. Finally, we will need one more fact about ideals.

Proposition 1.7.6. Let R be a ring and $I \subset R$ be a proper ideal (i.e. $I \neq R$). Then there is a prime ideal $Q \subset R$ containing I.

Proof. Let $\mathcal{C}$ be the partially ordered set of all proper ideals of R containing I ordered by inclusion; this is nonempty because $I \in \mathcal{C}$. If (I_α) is an ascending chain of ideals in $\mathcal{C}$, then $\bigcup_\alpha I_\alpha \subset R$ is also a proper ideal of R (check!); this proves that every chain in $\mathcal{C}$ has an upper bound, and hence $\mathcal{C}$ has a maximal element Q (this element need not be unique). We claim that Q is prime. Indeed, if it were not, then there would $p, q \in R$ such that $pq \in Q$ but neither $p \in Q$ nor $q \in Q$. Then we claim that $Q + pR$ is a strictly larger ideal in $\mathcal{C}$; that it contains I is clear, that it is strictly larger follows from $p \notin Q$, and that it is proper follows from the following argument. If $Q + pR = R$, then we can write $1 = s + pt$ for some $s \in Q$ and $t \in R$. Multiplying by q yields $q = qs + pqt$, but $qs \in Q$ (because $s \in Q$) and $pqt \in Q$ (because $pq \in Q$) and hence $q \in Q$, which is a contradiction. ∎

In fact, this maximal element Q of $\mathcal{C}$ as in the above proof if actually a maximal ideal of R, i.e. an ideal not contained in any other proper ideals of R (almost by definition!), and it is a general fact, which we showed in this proof, that any maximal ideal is prime.

1.7.2 Irreducible Components and Degree II

Let's return to the theory of curves; recall that we are over an algebraically closed field k. The idea here is that if $C \subset \mathbb{A}^2_k$ is a curve, then the vanishing ideal $\mathbb{I}(C)$ of C defined in Definition

[19] Here, and always, we use the convention that domains are nonzero.

1.6.11 is an ideal of the ring $k[x, y]$, and, in fact, by Theorem 1.6.12, a principal ideal.

Definition 1.7.7. Given a curve $C \subset \mathbb{A}^2_k$, a minimal polynomial of C is a generator of the principal ideal $\mathbb{I}(C) \subset k[x, y]$.

Note that any minimal polynomial must necessarily be reduced (why?). By Lemma 1.7.3, any two minimal polynomials of C differ by multiplication by units in $k[x, y]$, i.e. nonzero scalars–this is why we sometimes speak of "the minimal polyomial". If $C = C_f$ for a nonconstant $f \in k[x, y]$ then a minimal polynomial of C can be taken to $\mathrm{rad}(f)$. This gives us a perfect translation between alegbra and geometry. For instance, we can use this to define the degree of curve.

Definition 1.7.8 (Degree)**.** Given a curve $C \subset \mathbb{A}^2_k$, the degree of C is defined to be the degree of any minimal polynomial for C.

You may verify that if $k = \overline{k}$, then this definition agrees with Definition 1.2.2. Similarly, Corollary 1.6.13 can be restated as

Corollary 1.7.9 (Hilbert's Nullstellensatz for Curves, Version III)**.** Over an algebraically closed field k, there is a bijective correspondence

$$\{\text{curves } C \subset \mathbb{A}^2_k\} \dashleftarrow\dashrightarrow \{\text{pr. ideals of } k[x, y] \text{ gen. by nonconst. reduced } f \in k[x, y]\}$$

given by sending a curve C to $\mathbb{I}(C)$ and an ideal I to C_f for any generator f of I. Under this correspondence, the curve C is irreducible iff $\mathbb{I}(C)$ is a prime ideal.

Finally, from unique factorization in $k[x, y]$, we also obtain a decomposition for curves.

Theorem 1.7.10 (Unique Factorization/Irreducible Decomposition for Curves)**.** If $k = \overline{k}$, then given any curve $C \subset \mathbb{A}^2_k$, there is an integer $n \geq 1$ and irreducible curves $C_1, \ldots, C_n \subset \mathbb{A}^2_k$ such that $C_i \neq C_j$ for $i \neq j$ and

$$C = C_1 \cup C_2 \cup \cdots \cup C_n.$$

Further, if $m \geq 1$ is any other integer and $D_1, \ldots, D_m \subset \mathbb{A}^2_k$ irreducible curves such that $D_i \neq D_j$ for $i \neq j$ and

$$C = D_1 \cup D_2 \cup \cdots \cup D_m,$$

then $m = n$ and for all i, we have $C_i = D_{\sigma(i)}$ for some bijection $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$.

Proof. If f is a minimal polynomial of C, and we write $f = f_1 \cdots f_n$ for some $n \geq 1$ and distinct irreducible $f_1, \ldots, f_n \in k[x, y]$, then taking $C_i = C_{f_i}$ for $1 \leq i \leq n$ gives us the indicated decomposition, where we are using both that f is reduced and Corollary 1.6.8 to conclude that $C_i \neq C_j$ for $i \neq j$ (how?). If we have a decomposition $C = D_1 \cup \cdots \cup D_m$, and for each j with $1 \leq j \leq m$, we take a minimal polynomial $g_j \in k[x, y]$ for D_j, then each g_j is irreducible by Corollary 1.6.13(a), and for $i \neq j$, the polynomials g_i and g_j are not scalar multiples of each other by the hypothesis that $D_i \neq D_j$. Then the reduced polynomials f and $g := g_1 \cdots g_m$ define the same curve C, and hence by Corollary 1.6.13(c) are related by nonzero scalars; then we are done by unique factorization in $k[x, y]$, which is Corollary 1.5.14 (how?). ∎

The curves $C_1, \ldots, C_n \subset C$ occuring in such a decomposition are called the irreducible components of C, and they correspond to the irreducible factors of any minimal polynomial of C. Finally, we can upgrade Theorem 1.6.6 slightly to get

Theorem 1.7.11 (Finite Intersection Revisited)**.** If $C, D \subset \mathbb{A}^2_k$ are two curves that don't share any common irreducible components, then the intersection $C \cap D$ is finite.

Proof. Decompose $C = C_1 \cup \cdots \cup C_n$ and $D = D_1 \cup \cdots \cup D_m$ into irreducible components as in Theorem 1.7.10. For each pair (i, j) with $1 \leq i \leq n$ and $1 \leq j \leq m$, if we take minimal polynomials f_i and g_j of C_i and D_j respectively, then f_i and g_j are irreducible (by Corollary 1.6.13(a)) and $C_i \neq D_j$ implies that f_i and g_j are not scalar multiples of each other and hence relatively prime. It follows from Theorem 1.6.6 that each $C_i \cap D_j$ is finite, and hence so is

$$C \cap D = \bigcup_{1 \leq i \leq n} \bigcup_{1 \leq j \leq m} C_i \cap D_j.$$

■

1.7.3 A Few Examples of Irreducible Curves

That's enough general theory. Let's work out a few specific examples.

Example 1.7.12. For any field k, the linearpolynomial $\ell = x + y + 1 \in k[x, y]$ is irreducible: indeed, applying Lemma 1.6.1 to $R = k[x]$ with $t = y$, it suffices to show that ℓ is irreducible in $K[y]$ where $K = k(x)$, but that is true simply because $\ell \in K[y]$ is a linear polynomial.[20] Therefore, the line $C_\ell \subset \mathbb{A}^2_k$ is irreducible. The same argument shows that any line in $\mathbb{A}^2_k$ is irreducible, or many generally, that if $f(x, y) \in k[x, y]$ is any polynomial that is linear in either x or y, then $f(x, y)$ is irreducible. For instance, the polynomial $f(x, y) = y - x^2 \in k[x, y]$ is irreducible, so that the parabola $C = \{(t, t^2) : t \in k\} \subset \mathbb{A}^2_k$ is as well.

Example 1.7.13. For any field k, the polynomial $f(x, y) := xy - 1 \in k[x, y]$ is irreducible thanks to Lemma 1.6.1 applied to $R = k[x]$ with $t = y$–note that although $f(x, y)$ is not monic in y, it is still primitive. Over $k = \mathbb{R}$, the polynomial $f(x, y) = xy - 1 \in \mathbb{R}[x, y]$ defines the rectangular hyperbola C with two components. Why does this not contradict irreducibility? Well, firstly: the connection between (topological) irreducibility of curves and polynomials only works over algebraically closed fields such as $k = \mathbb{C}$: over $k = \mathbb{C}$, the "hyperbola" defined by f is a topologically a sphere punctured at two points, which is connected. Secondly, the rectangular hyperbola $C \subset \mathbb{A}^2_{\mathbb{R}}$ is still algebraically irreducible:

Lemma 1.7.14. If $g(x, y) \in \mathbb{R}[x, y]$ is a polynomial that vanishes on one branch of the hyperbola C, (or, in fact, any infinite subset of C) then $f \mid g$ in $\mathbb{R}[x, y]$, so that g must also vanish on the second branch.

Proof. Either g is zero and we are done, or g is nonconstant, in which case we may consider $C_g(\mathbb{C}) \subset \mathbb{A}^2_{\mathbb{C}}$. By hypothesis, $C_g(\mathbb{C})$ and $C_f(\mathbb{C})$ intersect in infinitely many points, so it follows from Theorem 1.6.6 that $f, g \in \mathbb{C}[x, y]$ are not relatively prime. Since $f \in \mathbb{C}[x, y]$ is irreducible by Example 1.7.13, this can only happen if $f \mid g$ in $\mathbb{C}[x, y]$, so that $g/f \in \mathbb{C}[x, y] \cap \mathbb{R}(x, y) = \mathbb{R}[x, y]$. ■

[20] This uses that we understand irreducibility in the polynomial ring $K[y]$ in one variable y over a field K really well.

In other words, just one branch of the hyperbola C is not an algebraic curve by itself. This proposition illustrates that sometimes we can prove results over non algebraically closed fields by using Theorem 1.4.5, and also that curves are incredibly rigid: any polynomial vanishing on any collection of infinite points of one curve must vanish on all of it. This is a manifestation of the coarseness of the Zariski topology.

Example 1.7.15. For any field k, the polynomial $f(x,y) := y^2 - x^3 + x \in k[x,y]$ is irreducible as well. There are a few ways to prove this. One way is sketched in Exercise 2.3.1. Another way to invoke Lemma 1.6.1 again to reduce the problem to showing that $y^2 - x^3 + x \in K[y]$ is irreducible where $K = k(x)$. If it were not irreducible, then it would split into two linear factors; we can assume without loss of generality that these factors of the form $y \pm p(x)$ for some $p(x) \in K$ (why?). Then $p^2 = x^3 - x \in K$, and there are many ways to see why this can't happen. One possible approach is to note that although $x^3 - x$ is not squarefree in general (when $\operatorname{ch} K = 2$), the power of x dividing $x^3 - x$ is still exactly one, and in particular odd. Therefore, if we use that $k[x]$ is a UFD to write $p = r/s$ for some coprime $r, s \in k[x]$ with $s \neq 0$, then $r^2 = x(x^2-1)s^2$ leads to a contradiction to unique factorization.

Again, over $k = \mathbb{R}$, the curve C_f of Example 1.7.15 has two components. Again, however, $C_f(\mathbb{C})$ is a punctured torus (hence connected, even irreducible) and the two components visible in $C_f(\mathbb{R})$ are vestiges of slicing this torus and the fact that $\mathbb{R}$ is not algebraically closed. Finally, an argument identical to that in the proof of Lemma 1.7.14 shows that neither of the pieces of $C_f(\mathbb{R})$ are algebraic curves by themselves.

1.7.4 A Sneak Peek at Curve Intersections

Given two curves $C, D \subset \mathbb{A}^2_k$, in how many points do C and D intersect? Well, they could share a component and have infinitely many points in common, but at least when they don't share a component this intersection is finite (this was Theorem 1.7.11). A little experimenting seems to suggest that if C and D are curves of degree m and n respectively, then C and D usually intersect in mn points, but this is not always true. For instance:

(a) When $k = \mathbb{R}$, the parabola C_f defined by $f(x,y) = y - x^2$ and the line C_ℓ defined by $\ell(x,y) = y - x + 1$ do not intersect at all, since $x^2 - x + 1 \in \mathbb{R}[x]$ has no real roots. However, this problem doesn't really appear over algebraically closed fields such as $k = \mathbb{C}$.

(b) Even over fields such as $k = \mathbb{C}$, we have to account for tangency. For instance, if we take $f(x,y) = y - x^2$ again and $\ell(x,y) := y - 2x + 1$, then the polynomial $x^2 - 2x + 1 = (x-1)^2 \in \mathbb{C}[x]$ still has only one root over $\mathbb{C}$. This is because this line C_ℓ is tangent to the parabola, and should really count as having "intersection multiplicity" two.

(c) Finally, even if we account for intersection multiplicites, we can have asymptotes or parallel lines. For instance, the lines defined by $\ell_1(x,y) := y - x$ and $\ell_2(x,y) = y - x + 1$ never intersect in $\mathbb{A}^2_k$ for any field k because they are "parallel". To rectify this situation, we need to account for intersections "at infinity".

As it turns out, these are the only four problems. Our eventual goal is to show the theorem of Bézout (Theorem 1.14.1) which says that if k is an algebraically closed field, then any two projective plane curves $C, D \subset \mathbb{P}^2_k$ of degrees $m, n \geq 1$ respectively that do not share a common component intersect in exactly mn points, when counted with multiplicity. Over the next few lectures, we'll develop tools to prove this theorem, starting with smoothness and intersection multiplicity.

1.8 06/26/24 - Smoothness, Multiplicity, Tangent Lines

Today, we will talk about smoothness of algebraic curves. What should smoothness mean–i.e. what should it mean to say that a curve $C \subset \mathbb{A}^2_k$ is smooth at a point $P \in C$? One definition is that at each point, we have a well-defined tangent direction, i.e. that the curve is well-approximated by a linear polynomial. Certainly, whatever this notion is, it should be invariant under affine changes of coordinates, so we may focus on the case when $P = (0,0)$, and then considering a few examples naturally leads us to the following definition.

Definition 1.8.1.

(a) A polynomial $f(x,y) \in k[x,y]$ is said to be homogeneous of degree $d \geq 0$ if in the ring $k[x,y,t]$, we have the polynomial identity

$$f(tx,ty) = t^d f(x,y).$$

This is equivalent to saying that in an expression of the form $f(x,y) = \sum_{i,j\geq 0} a_{i,j}x^i y^j$ with $a_{i,j} \in k$, we have $a_{i,j} = 0$ unless $i+j = d$. For each $d \geq 0$, the set of all polynomials in $k[x,y]$ of degree d will be denoted by $k[x,y]_d$.

(b) Any $f(x,y) \in k[x,y]$ can be written uniquely as

$$f = f_0 + f_1 + \cdots + f_d,$$

where $d = \deg f \geq 0$, and for each i with $0 \leq i \leq d$, the polynomial $f_i \in k[x,y]$ is homogeneous of degree i. If $0 \neq f$, then there is a unique smallest index i_0 such that $f_{i_0} \neq 0$; in this case, we define the multiplicity of f at the origin $O = (0,0)$, written $m_O(f)$, and the initial part of f, written $\mathrm{in}(f)$, to be, respectively,

$$m_O(f) = i_0 \text{ and } \mathrm{in}(f) := f_{i_0}.$$

Example 1.8.2. If $f(x,y) = y^2 - x^3$, then $m_O(f) = 2$ with $\mathrm{in}(f) = y^2$.

We say that a function $F : \mathbb{A}^2_k \to k$ is homogeneous of degree $d \geq 0$ if for all $(p,q) \in \mathbb{A}^2_k$ and $t \in k$, we have $F(tp,tq) = t^d F(p,q)$. If a polynomial $f \in k[x,y]$ is homogeneous of degree $d \geq 0$, then so is the associated function F_f, and the converse holds if k is infinite. Note that the zero polynomial $0 \in k[x,y]$ is homogenous of degree d for every $d \geq 0$, and for each $d \geq 0$, the subset $k[x,y]_d \subset k[x,y]$ is a vector subspace of dimension $d+1$ with basis $x^d, x^{d-1}y, \cdots, xy^{d-1}, y^d$, with $k[x,y] = \bigoplus_{d\geq 0} k[x,y]_d$. Finally, if $f \in k[x,y]_d$ and $g \in k[x,y]_e$, then $fg \in k[x,y]_{d+e}$. This structure on $k[x,y]$ is called the structure of a graded k-algebra.

Lemma 1.8.3. If $k = \overline{k}$, then for any $d \geq 0$ and $f \in k[x,y]_d$, there are homogeneous linear polynomials $\ell_1, \ldots, \ell_d \in k[x,y]_1$ such that $f = \ell_1\ell_2\cdots\ell_d$. If f is nonzero, then these factors are uniquely determined up to reordering and nonzero scalars.

Proof. Write $f = \sum_{i=0}^d a_i x^{d-i}y^i$. If $f \neq 0$, let i_0 be the least index such that $a_{i_0} \neq 0$. Since $k = \overline{k}$, we can factor the polynomial $f(t,1) = \sum_{i=i_0}^d a_i t^{d-i}$ of degree $d - i_0$ as

$$f(t,1) = \sum_{i=i_0}^{d} a_i t^{d-i} = a_{i_0} \prod_{j=1}^{d-i_0} (t - \alpha_j)$$

for some $\alpha_j \in k$, and then taking $a_{i_0}^{-1}\ell_1 = \ell_2 = \cdots = \ell_{i_0} = y$ and $\ell_{i_0+j} = x - \alpha_j y$ for $j = 1, \ldots, d-i_0$ suffices. Uniqueness is clear because $k[x,y]$ is a UFD, and each ℓ_j is prime. ∎

Definition 1.8.4.

(a) Given a curve $C \subset \mathbb{A}^2_k$, we define the multiplicity of C at the origin $O = (0,0)$ to be

$$m_O(C) := m_O(f_C),$$

where $f_C \in k[x,y]$ is any minimal polynomial for C. If $\mathrm{in}(f_C) = \ell_1 \cdots \ell_m$ is the factorization of $\mathrm{in}(f_C)$ into linear factors as in Lemma 1.8.3, where $m := m_O(C)$, then we define the tangent lines to C at O to be the lines $L_j := C_{\ell_j}$ for $j = 1, \ldots, m$. (These need not all be distinct, and are independent of the choice of f_C.) Finally, the tangent cone to C at O is define to be

$$\mathrm{TC}_O\, C := C_{\mathrm{in}(f)} = L_1 \cup L_2 \cup \cdots \cup L_m.$$

(b) Given a curve $C \subset \mathbb{A}^2_k$ and an arbitrary point $P \in \mathbb{A}^2_k$, we define the multiplicity of C at P to be

$$m_P(C) := m_O(\phi^{-1}C),$$

where $\phi : \mathbb{A}^2_k \to \mathbb{A}^2_k$ is any affine change of coordinates such that $\phi(O) = P$. We define the tangent lines to C at P to be the lines $\phi(L_j)$ for $j = 1, \ldots, m$ where $m = m_P(C)$, and similarly the tangent cone to C at P to be

$$\mathrm{TC}_P C = \phi(\mathrm{TC}_O(\phi^{-1}C)).$$

(c) Given a curve $C \subset \mathbb{A}^2_k$ and point $P \in \mathbb{A}^2_k$, we have $m_P(C) \geq 1$ iff $P \in C$, in which case we say that P is a smooth point of C iff $m_p(C) = 1$. The curve C is said to be smooth if every $P \in C$ is a smooth point. A point $P \in C$ that is not a smooth point is called a singular point or multiple point of C.[a]

[a] Outside of mathematics, the terms "singular" and "multiple" are usually antonyms; in this case, they are not, because "singular" here means "exceptional" or "extraordinary" (see Theorem 1.9.7), while "multiple" means "of higher (i.e. > 1) multiplicity".

Note that a smooth point on a curve has a unique tangent line, which we will denote by $\mathrm{T}_P C$. The coordinate-invariance of smoothness and multiplicity is baked into the definition–if we can show that it is well-defined. To do this, we need that if $\phi : \mathbb{A}^2_k \to \mathbb{A}^2_k$ is an affine change of coordinates such that $\phi(O) = O$, then for any polynomial $f \in k[x,y]$ we have $m_O(f) = m_O(\phi^*(f))$. By considering the homogeneous parts separately, this reduces to showing

Lemma 1.8.5. If $\phi : \mathbb{A}^2_k \to \mathbb{A}^2_k$ is an affine change of coordinates such that $\phi(O) = O$, and if $0 \neq f \in k[x,y]$ is homogeneous of degree $n \geq 0$, then so is $\phi^*(f)$.

Proof. Note that ϕ is of the form $\phi(x', y') = (ax' + by', cx' + dy')$ for some $a, b, c, d \in k$ with $ad - bc \neq 0$. The claim is clear when $n = 0$, since then f is a nonzero constant and $\phi^*(f) = f$. When $n = 1$, we have $f = \lambda x + \mu y$ for some $\lambda, \mu \in k$, not both zero, and then

$$\phi^*(f) = \lambda(ax' + by') + \mu(cx' + dy') = (a\lambda + c\mu)x' + (b\lambda + d\mu)y'.$$

Now, since one of λ and μ is not zero, and since $ad - bc \neq 0$, it follows easily that at least one of $a\lambda + c\mu$ and $b\lambda + d\mu$ is nonzero (this is basic linear algebra, but can also be shown directly–how?). Therefore, we are done in this case. If $n \geq 2$, then by Lemma 1.8.3, we can write $f = \ell_1 \cdots \ell_n$ for some ℓ_j homogeneous of degree 1; then we are done by the case $n = 1$ and the observation $\phi^*(f) = \phi^*(\ell_1)\phi^*(\ell_2) \cdots \phi^*(\ell_n)$. This finishes the proof when $k = \overline{k}$ (which is the only case we care about), but in general, we can use Theorem 1.4.5 to reduce to this case. ∎

Example 1.8.6. The parabola C defined by $f(x,y) = y - x^2 \in k[x,y]$ has is smooth at the point $(1,1) \in \mathbb{A}^2_k$ with tangent line L defined by the vanishing of $y - 2x + 1 = 0$.

Example 1.8.7. A curve C is said to have a simple node at P iff $m_P(C) = 2$ and C has two distinct tangent lines at P. For instance, the curve C defined by $f(x,y) = y^2 - x^2(x+1) \in k[x,y]$ over a field k with $\operatorname{ch} k \neq 2$ has a simple node at the origin, with tangent lines L_1, L_2 defined by the vanishing of $y \pm x$, and tangent cone $T_O(C) = L_1 \cup L_2$. (What happens when $\operatorname{ch} k = 2$?)

Of course, this definition is not very convenient when we want to locate all singular points of a given curve C. For this, we need a more convenient criterion. This is provided by

Theorem 1.8.8 (Affine Jacobi Criterion)**.** Suppose we are given a curve $C \subset \mathbb{A}^2_k$ and a point $P = (p,q) \in \mathbb{A}^2_k$. Let $f \in k[x,y]$ be a minimal polynomial for C. Then

(a) $P \in C$ iff $f|_P := f(p,q) = 0$, and in this case
(b) P is a singular point of C iff

$$\left.\frac{\partial f}{\partial x}\right|_P = \left.\frac{\partial f}{\partial y}\right|_P = 0.$$

(c) If $P \in C$ is a smooth point, then the tangent line $\mathrm{T}_P C$ is defined by the vanishing of

$$\left.\frac{\partial f}{\partial x}\right|_P (x-p) + \left.\frac{\partial f}{\partial y}\right|_P (y-q) \in k[x,y].$$

Wait, what? What are these partial derivative symbols? Why can we do this over any field k? We'll discuss this more next time, but for now let's work out an example to see how conveniently Theorem 1.8.8 allows us to locate singular points of a curve C.

Example 1.8.9. If $f(x,y) = y - x^2$, then $\partial f/\partial y \equiv 1$ tells us that f is smooth everywhere. At the point $P = (t,t^2)$, the tangent line to C is given by the vanishing of

$$-2t(x-t) + 1(y-t) = y - 2tx + t^2 \in k[x,y].$$

Note that when $\operatorname{ch} k = 2$, this tangent line is always horizontal–which is incredibly weird. In general, weird stuff happens to curves of degree p in characteristic p–watch out for this over the next few weeks!

Example 1.8.10. If $f(x,y) = y^2 - x^3$, then the system of equations we need to solve for the singular points of C is

$$\begin{aligned} y^2 - x^3 &= 0, \\ -3x^2 &= 0, \\ 2y &= 0, \end{aligned}$$

which in any characteristic has the unique solution $(x,y) = (0,0)$ (check!). Therefore, the unique singular point of C is the origion O, where C has the unique tangent line $y = 0$, i.e. the x-axis.

1.9 06/28/24 - Derivations, Intersection Multiplicity

Today, we'll prove the Jacobi Criterion (Theorem 1.8.8), and start talking about intersection multiplicity for two curves.

1.9.1 Derivations and the Jacobi Criterion

We want to first discuss an algebraic way to differentiate things, for which we introduce derivations.

Definition 1.9.1. Let k be a field, and R be a ring containing k. A k-derivation on R is a k-linear map $D : R \to R$ satisyfing the Leibniz rule, i.e. a map $D : R \to R$ such that

(a) for all $a, b \in k$ and $f, g \in R$, we have $D(af + bg) = a \cdot D(f) + b \cdot D(g)$, and
(b) for all $f, g \in R$, we have $D(fg) = D(f) \cdot g + f \cdot D(g)$.

The set of all k-derivations of R is denoted by $\mathrm{Der}_k(R)$.

Remark 1.9.2. The definition works also if k is any ring–then R can be any k-algebra, i.e. a ring with a homomorphism $\rho : k \to R$. Note that $\mathrm{Der}_k(R)$ is an R-module and a k-Lie algebra.[21]

Note that if $D \in \mathrm{Der}_k(R)$, then $D(c) = 0$ for all $c \in k$. This follows from

$$D(1) = D(1^2) = D(1) \cdot 1 + 1 \cdot D(1) = 2\,D(1),$$

so that $D(1) = 0$ and $D(c) = c \cdot D(1) = 0$. Therefore, a k-derivation on R captures the notion of differentiating elements of R, where elements of k function as "constants".

Example 1.9.3. If $R = k[x]$, then the operation $\sum_{i\geq 0} a_i x^i \mapsto \sum_{i\geq 1} i a_i x^{i-1}$ is a k-derivation on R, denoted ∂_x or $\partial/\partial x$. Note that if $c \in R$ is any element, then the operation $f \mapsto c \cdot \partial_x f$ is also a derivation of R. More generally, if $R = k[x_1, x_2, \ldots, x_n]$, then the operations ∂_{x_j} are all derivations on R, and hence so are $\sum_{j=1}^n c_j \partial_{x_j}$. In fact, these are all the k-derivations of R.

Theorem 1.9.4. Let k be a field, and let $R = k[x_1, \ldots, x_n]$ be the polynomial ring over R in $n \geq 1$ variables $x_1, \ldots, x_n$. Then

$$\mathrm{Der}_k(R) = \bigoplus_{j=1}^{n} R \cdot \partial_{x_j}.$$

In other words, given any $c_1, c_2, \ldots, c_n \in R$, there is a unique k-derivation $D : R \to R$ such that $D(x_j) = c_j$ for each $j = 1, \ldots, n$.

Proof. It follows from the Leibniz rule that if $D : R \to R$ is any derivation and $f \in R$, then

$$D(f) = \sum_{j=1}^{n} \partial_{x_j}(f) \cdot D(x_j).$$

Therefore, a k-derivation D of R is determined by $D(x_j)$ for $j = 1, \ldots, n$, showing uniqueness. Conversely, if $c_1, \ldots, c_n$ are given, taking $D = \sum_{j=1}^n c_j \partial_{x_j}$ works, showing existence. ∎

[21] As usual, if you don't know what this means, you can ignore it. If you do, what is the Lie algebra structure on $\mathrm{Der}_k(R)$?

It is now possible to derive algebraically the multivariable chain rule for polynomials. Let's do a special case–the only case we will need–to illustrate the process.

Lemma 1.9.5. Let $\phi : \mathbb{A}^2_k(x', y') \to \mathbb{A}^2_k(x, y)$ be an affine change of coordinates of the form $(x, y) = \phi(x', y') = (ax' + by' + p, cx' + dy' + q)$, where $a, b, c, d, p, q \in k$ satisfy $ad - bc \neq 0$. If $\phi^* : k[x, y] \to k[x', y']$ denotes the associated ring homomorphism, then for any $f \in k[x, y]$, we have

$$\partial_{x'}(\phi^* f) = a \cdot \phi^*(\partial_x f) + c \cdot \phi^*(\partial_y f) \text{ and}$$
$$\partial_{y'}(\phi^* f) = b \cdot \phi^*(\partial_x f) + d \cdot \phi^*(\partial_y f).$$

In particular, given any $Q \in \mathbb{A}^2_k$ and $f \in k[x, y]$, we have

$$\partial_x f|_Q = \partial_y f|_Q = 0 \Leftrightarrow \partial_{x'}(\phi^* f)|_{\phi^{-1}(Q)} = \partial_{y'}(\phi^* f)|_{\phi^{-1}(Q)} = 0.$$

The more traditional way to express the change of variables formula from Lemma 1.9.5 is to write

$$\frac{\partial f}{\partial x'} = \frac{\partial f}{\partial x} \cdot \frac{\partial x}{\partial x'} + \frac{\partial f}{\partial y} \cdot \frac{\partial y}{\partial x'} \text{ and}$$
$$\frac{\partial f}{\partial y'} = \frac{\partial f}{\partial x} \cdot \frac{\partial x}{\partial y'} + \frac{\partial f}{\partial y} \cdot \frac{\partial y}{\partial y'},$$

written which way, this formula is valid for other types of changes of coordinates as well.

Proof. We'll show the first identity; the proof of the second is similar. Since ϕ^* is a ring isomorphism, in light of Theorem 1.9.4, it suffices to show that the map

$$D : k[x, y] \to k[x, y] \text{ defined by } D(f) = (\phi^*)^{-1}\partial_{x'}(\phi^* f)$$

is a k-derivation, and that $D(x) = a$ and $D(y) = c$. This last part is easy: indeed,

$$D(x) = (\phi^*)^{-1}\partial_{x'}(\phi^* x) = (\phi^*)^{-1}\partial_{x'}(ax' + by' + p) = (\phi^*)^{-1}a = a,$$

and similarly $D(y) = c$. To check that this D is a derivation, note that condition (a) in Definition 1.9.1 is clear because ϕ^*, $\partial_{x'}$ and $(\phi^*)^{-1}$ are all k-linear, and condition (b) follows from the check that for all $f, g \in k[x, y]$ we have

$$\begin{aligned} D(fg) &= (\phi^*)^{-1}\partial_{x'}(\phi^*(fg)) \\ &= (\phi^*)^{-1}\partial_{x'}(\phi^* f \cdot \phi^* g) \\ &= (\phi^*)^{-1}\left[\partial_{x'}(\phi^* f) \cdot \phi^* g + \phi^* f \cdot \partial_{x'}(\phi^* g)\right] \\ &= \left((\phi^*)^{-1}\partial_{x'}(\phi^* f)\right) \cdot (\phi^*)^{-1}\phi^* g + (\phi^*)^{-1}\phi^* f \cdot \left((\phi^*)^{-1}\partial_{x'}(\phi^* g)\right) \\ &= D(f) \cdot g + f \cdot D(g). \end{aligned}$$

The second statement follows from the first by the same linear algebra as before, since again $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ has nonzero determinant, i.e. is an invertible matrix. ∎

We are now ready to prove the Jacobi criterion, which we restate here for convenience.

Theorem 1.8.8 (Affine Jacobi Criterion)**.** Suppose we are given a curve $C \subset \mathbb{A}_k^2$ and a point $P = (p, q) \in \mathbb{A}_k^2$. Let $f \in k[x, y]$ be a minimal polynomial for C. Then

(a) $P \in C$ iff $f|_P := f(p, q) = 0$, and in this case
(b) P is a singular point of C iff

$$\left.\frac{\partial f}{\partial x}\right|_P = \left.\frac{\partial f}{\partial y}\right|_P = 0.$$

(c) If $P \in C$ is a smooth point, then the tangent line $\mathrm{T}_P C$ is defined by the vanishing of

$$\left.\frac{\partial f}{\partial x}\right|_P (x - p) + \left.\frac{\partial f}{\partial y}\right|_P (y - q) \in k[x, y].$$

Proof. The statement in (a) is clear. First , let's prove (b) and (c) for $P = O = (0, 0)$. If we write $f = f_1 + f_2 + \cdots + f_d$, where $d = \deg f$ and each f_j is homogeneous of degree j (note $P \in C$ is equivalent to $f_0 = 0$), then

$$f_1 = \lambda x + \mu y$$

for some $\lambda, \mu \in k$. Then

$$\partial_x f = \lambda + \partial_x f_2 + \cdots + \partial_x f_d,$$

and for each $j \geq 2$, we have $\partial_x f_j|_P = 0$, whence $\partial_x f|_P = \lambda$. Similarly, $\partial_y f|_P = \mu$. Therefore,

$$m_P(C) \geq 2 \Leftrightarrow f_1 = 0 \Leftrightarrow \lambda = \mu = 0 \Leftrightarrow \partial_x(f)|_P = \partial_y(f)|_P = 0.$$

Since

$$f_1 = \partial_x f|_P \cdot (x - 0) + \partial_y f|_P \cdot (y - 0),$$

the result of (c) is also clear. In general, let $\phi : \mathbb{A}_k^2 \to \mathbb{A}_k^2$ be an affine change of coordinates such that $\phi(O) = P$. It is easy to see then that $\phi^* f$ is a minimal polynomial for $\phi^{-1}C$, and so we have

$$\begin{aligned} m_P(C) \geq 2 &\Leftrightarrow m_O(\phi^{-1}C) \geq 2 \\ &\Leftrightarrow \partial_{x'}(\phi^* f)|_O = \partial_{y'}(\phi^* f)|_O = 0 \\ &\Leftrightarrow \partial_x f|_P = \partial_y f|_P = 0 \end{aligned}$$

as needed, where in the last step we have used Lemma 1.9.5. The proof of (c) is similar, but can be simplified even more by noting that it suffices to consider a change of coordinates of the simple form $(x, y) = \phi(x', y') = (x' + p, y' + q)$; the details are left to the reader. ∎

From this criterion, we can derive many important results. Here are a couple.

Theorem 1.9.6. A plane curve is singular at the points of intersection of its components. In particular, an affine curve is smooth iff its components are both individually smooth and pairwise disjoint.

Proof. Let f, g be two distinct irreducibles, and suppose $C = C_f \cup C_g = C_{fg}$; the general case is similar. By Theorem 1.8.8, it suffices to show that if $P \in C_f \cap C_g$, then $\partial_x(fg)|_P = \partial_y(fg)|_P = 0$, but this is clear because, for instance, we have

$$\partial_x(fg)|_P = \partial_x(f)|_P \cdot g|_P + f|_P \cdot \partial_x(g)|_P = 0$$

because $f|_P = g|_P = 0$. ∎

Recall now our base assumption that k is algebraically closed.

Theorem 1.9.7. If $C \subset \mathbb{A}^2_k$ is any curve, then C has only finitely many singular points.

Proof. Let $C = C_1 \cup C_2 \cup \cdots \cup C_n$ be the irreducible decomposition of C (Theorem 1.7.10). For each $1 \leq i < j \leq n$, the intersection $C_i \cap C_j$ is finite by Theorem 1.7.11; therefore, it suffices to show the result for an irreducible C. Let $f \in k[x, y]$ be a minimal polynomial for C; then f is irreducible by Corollary 1.6.13(a). By Theorem 1.8.8, it suffices to show that the system of polynomial equations

$$f = \partial_x f = \partial_y f = 0$$

has only finitely many solutions in $\mathbb{A}^2_k$.[22] First suppose that $\partial_x f \neq 0$ (i.e. as a polynomial in $k[x, y]$). Since $\deg \partial_x f < \deg f$, it follows that either $\partial_x f$ is a nonzero constant (in which case C is smooth, and we are done), or that f and $\partial_x f$ are relatively prime (since f is prime and $\partial_x f$ cannot be a nonzero polynomial multiple of f for degree reasons), in which case we are done by Theorem 1.6.6. Similarly, if $\partial_y f \neq 0$, we are done.

This finishes the proof when $\operatorname{ch} k = 0$, because if $\operatorname{ch} k = 0$ and $f \in k[x, y]$ is any nonconstant polynomial, then one of $\partial_x f$ and $\partial_y f$ is nonzero. Unfortunately, when $\operatorname{ch} k = p > 0$, there are nonconstant $f \in k[x, y]$ such that $\partial_x f = \partial_y f = 0$, such as $f = x^p + y^p$. We will show that this cannot happen if f is irreducible: we will show that even if $\operatorname{ch} k = p > 0$, as long as $f \in k[x, y]$ is irreducible, then one of $\partial_x f$ and $\partial_y f$ is nonzero. Indeed, suppose not. Then to say that $\partial_x f = 0$ means that if we write $f = \sum_{i,j} a_{i,j} x^i y^j$, then $a_{i,j} = 0$ unless $p \mid i$. Similarly, $\partial_y f = 0$ implies that $a_{i,j} = 0$ unless $p \mid j$. Therefore, we conclude that

$$f = \sum_{i,j \geq 0} a_{pi,pj} x^{pi} x^{pj}.$$

Since k is algebraically closed, for each $i, j \geq 0$, we can find a p^{th} root $\alpha_{i,j} \in k$ of $a_{pi,pj}$, i.e. an element such that $\alpha_{i,j}^p = a_{pi,pj}$. Then, since we are in characteristic p,

$$f = \sum_{i,j \geq 0} \alpha_{i,j}^p x^{pi} y^{pj} = \left(\sum_{i,j \geq 0} \alpha_{i,j} x^i y^j \right)^p = g^p,$$

where $g := \sum_{i,j \geq 0} \alpha_{i,j} x^i y^j$, contradicting irreducibility of f. This completes the proof when $k = \overline{k}$; in general, we can reduce to this case by Theorem 1.4.5 as before. ∎

Example 1.9.8. For any field k, consider the circle C defined by $f(x, y) := x^2 + y^2 - 1 \in k[x, y]$. This has partial derivatives

$$\partial_x f = 2x \text{ and } \partial_y f = 2y.$$

When $\operatorname{ch} k \neq 2$, it follows that this system $f = \partial_x f = \partial_y f = 0$ has no solutions, so that C is smooth. When $\operatorname{ch} k = 2$, it seems that $\partial_x f = \partial_y f = 0$, so that any point on C should be singular–why does this not contradict Theorem 1.9.7? Well, if we are to follow the proof of Theorem 1.9.7, we will observe that when $\operatorname{ch} k = 2$, in fact, we have that

$$f(x, y) = (x + y + 1)^2 \in k[x, y],$$

so that f is not reduced. In this case, the curve C is just a line with minimal polynomial $g(x, y) = x + y + 1$, which is also smooth. This example shows that when applying the Jacobi Criterion (Theorem 1.8.8), it is crucial to use a minimal polynomial for your curve. Another way to think about this is: a "curve" defined by a nonreduced polynomial is singular everywhere. This can be made precise using the language of schemes; we won't discuss this in this course.

[22] Here, I'm being a little sloppy about the distinction between polynomials and polynomial functions–given that we're in week 3, I'll presume you know what I mean and how to make this rigorous.

1.9.2 Intersection Multiplicity

Given curves $C, D \subset \mathbb{A}^2_k$ and a point $P \in C \cap D$, we want to make precise what we mean by the intersection multiplicity of C and D at P. Again, whatever this notion means, it should be invariant under affine (or even other kinds of) changes of coordinates, and as we observed in the previous sections, it is helpful to have this notion already for polynomials and not just curves–after all, we want to capture nonreduced behavior.

The goal, therefore, is to find a function

$$i : (k[x,y] \smallsetminus \{0\}) \times (k[x,y] \smallsetminus \{0\}) \times \mathbb{A}^2_k \to \mathbb{Z}_{\geq 0} \cup \{\infty\}, \quad (f,g,P) \mapsto i_P(f,g)$$

that satisfies some reasonable properties. What properties should we have? Here I list a few.

(1) (Symmetry) $i_P(f,g) = i_P(g,f)$ for all f, g, P.
(2) (Finiteness for Proper Intersection) $i_P(f,g) = \infty$ iff f and g have a common component through P, i.e. there is a $q \in k[x,y]$ such that $q \mid f$ and $q \mid g$ and $q|_P = 0$.
(3) (Non-Intersection) $i_P(f,g) = 0$ iff $P \notin C_f \cap C_g$, i.e. either $f|_P \neq 0$ or $g|_P \neq 0$.
(4) (Additivity) $i_P(f_1 f_2, g) = i_P(f_1, g) + i_P(f_2, g)$ for all $f_1, f_2, g \in k[x,y] \smallsetminus \{0\}$ and $P \in \mathbb{A}^2_k$.
(5) (Coordinate Ring Dependence) $i_P(f,g) = i_P(f, g + hf)$ for all $f, g, h \in k[x,y] \smallsetminus \{0\}$.
(6) (Invariance under ACOCs) If $\phi : \mathbb{A}^2_k \to \mathbb{A}^2_k$ is an ACOC, then $i_P(f,g) = i_{\phi^{-1}(P)}(\phi^*(f), \phi^*(g))$.
(7) (Normalization) For $P = O = (0,0)$, we have $i_O(x,y) = 1$.

The amazing result is, then, that these properties characterize intersection multiplicity uniquely.

Theorem 1.9.9. There is a unique function i satisfying (1)-(7) above.

We'll sketch a proof next time; today, let's work out a few examples this time. Firstly, by (3) and (4), scaling f or g by nonzero scalars does not change the intersection multiplicity.

Example 1.9.10. If $f = y^2 - x^2(x+1)$ and $g = x$ and $P = (0,0)$, then

$$i_P(y^2 - x^2(x+1), x) = i_P(y^2, x) = 2i_P(x,y) = 2.$$

If $g = y - tx$ for $t \in k$, then

$$\begin{aligned} i_P(y^2 - x^2(x+1), y - tx) &= i_P(y - tx, y^2 - x^2(x+1) - (y+tx)(y-tx)) \\ &= i_P(y - tx, x^2(-x + t^2 - 1)) \\ &= 2i_P(y - tx, x) + i_P(y - tx, -x + t^2 - 1) \\ &= 2 + \begin{cases} 1, & \text{if } t^2 - 1 = 0, \\ 0, & \text{else.} \end{cases} \end{aligned}$$

This confirms our intuition that each line through P intersects the curve C_f at least twice, with even higher multiplicity (at most three) iff it is tangent to C_f at P.

Example 1.9.11. If C is a smooth curve with tangent line $L = \mathrm{T}_P C$ at $P \in C$ such that $C \neq L$, and f and ℓ are minimal polynomials for C and L, then $i_P(f, \ell) \geq 2$. Indeed, we can choose a suitable coordinate system so that $P = (0,0)$ and $\ell = y$; then $f_0 = 0$ and $f_1 = y$, whence

$$i_P(f, \ell) = i_P(y + (f - y), y) = i_P(f - y, y) \geq 2,$$

where in the last step we have used that $f - y$ is nonzero and homogeneous of degree at least 2. (How does this result follow?)

Example 1.9.12. Let $p(x) \in k[x]$ be a nonconstant polynomial of x alone, and let $f := y - p(x)$ and $g = y$. Then P is a point of intersection of the curves C_f (i.e. the graph of p) and C_g (i.e. the x-axis) iff $P = (\alpha, 0)$ for some root α of p. To compute the intersection multiplicity at this point, we factor $p(x) = (x - \alpha)^m q(x)$ for some integer $m \geq 1$ and $q(x) \in k[x]$ with $q(\alpha) \neq 0$, and then note that

$$i_P(f, g) = i_P((x - \alpha)^m q(x), y) = m \cdot i_P(x - \alpha, y) + i_P(q(x), y) = m \cdot 1 + 0 = m.$$

Therefore, the intersection multiplicity of f and g at P is exactly the multiplicity $m_\alpha(p)$ of α as a root of $p(x)$. In particular, we have

$$\sum_{P \in C_f \cap C_g} i_P(f, g) = \sum_{\alpha : p(\alpha) = 0} m_p(\alpha) = \deg p = (\deg f)(\deg g).$$

This is one simple manifestation of Bézout's Theorem, which we will soon get to. When $p(x) = 0$, every point on the x-axis is a point of infinite multiplicity, while if $p(x) = c$ is a nonzero constant, then there are no points of intersection, althought $(\deg f)(\deg g) = 1$; this is because the lines C_f and C_g are parallel (i.e. meet "at infinity"). We will soon develop tools to make this more precise.

1.10 07/01/24 - Intersection Multiplicity, the Projective Plane

Today, we'll finish the proof of Theorem 1.9.9, and start talking about the projective plane and projective curves.

1.10.1 Intersection Multiplicity

Let's proceed to the proof of Theorem 1.9.9. We need to show two things: existence and uniqueness of i. We'll start with uniqueness.

Proof of Uniqueness in Theorem 1.9.9. We will give an algorithm that takes as input (f, g, P) and returns $i_P(f, g)$ in finitely many steps, using only the axioms (1) - (7).

(a) By (6), we can reduce to the case $P = (0, 0)$.
(b) By (2) and (3), we are done if either f and g have a common component through P, or if $P \notin C_f \cap C_g$, so assume that we are not in either of these cases (we then say that C_f and C_g intersect properly at P).
(c) Consider the polynomials $f(x, 0), g(x, 0) \in k[x]$, and suppose they have degrees $d, e \geq 0$ respectively, where we use the convention that $\deg 0 = 0$. By (1), we may assume by switching f and g if needed that $0 \leq d \leq e$. Now we split into two cases:

Case 1. If $d > 0$, then we may perform the Euclidean algorithm to produce an integer $n \geq 1$ and polynomials $q_1, q_2, \ldots, q_{n+1}, r_1, \ldots, r_n, r_{n+1} \in k[x]$ such that for $i = 0, 1, \ldots, n$, we have

$$r_{i-1} = r_i \cdot q_{i+1} + r_{i+1},$$

and $\deg r_{i+1} < \deg r_i$, where $r_{-1} := g(x, 0)$, $r_0 := f(x, 0)$, $r_1 \cdots r_n \neq 0$, and $r_{n+1} = 0$; then $r_n = \gcd(f(x, 0), g(x, 0))$. Define polynomials $h_1, \ldots, h_n, h_{n+1} \in k[x, y]$ by

$$h_i = h_{i-2} - q_i \cdot h_{i-1}$$

for $i = 1, \ldots, n + 1$, where we set $h_{-1} := g$ and $h_0 = f$. We find inductively using (5) that

$$i_P(f, g) = i_P(h_1, f) = i_P(h_2, h_1) = \cdots = i_P(h_n, h_{n-1}) = i_P(h_{n+1}, h_n),$$

and $h_i(x, 0) = r_i(x)$ for each $i = 1, \ldots, n + 1$, and hence $h_{n+1}(x, 0) = 0$. We replace (f, g) by (h_{n+1}, h_n) and land on

Case 2. If $d = 0$, then $y \mid f$, and so we can write $f = y^N p$ for some $N \geq 1$ and $p \in k[x, y]$ such that $y \nmid p$. Then by (4) we have

$$i_P(f, g) = N \cdot i_P(y, g) + i_P(p, g).$$

By (5), we have

$$i_P(y, g) = i_P(y, g(x, 0)) = i_P(y, y - g(x, 0)) = m_0(g(x, 0)),$$

where in the last step we have used the computation in Example 1.9.12 (this uses (7)). By our assumption that $g|_P = 0$, we have $m_0(g(x, 0)) \geq 1$, and hence $i_P(y, g) \geq 1$, whence $i_P(p, g) < i_P(f, g)$. Either $i_P(p, g) = 0$, in which case we are done; else, return to the beginning of Step (c) with (f, g) replaced by (g, p).

■

It is clear that if such an i exists, then the above algorithm terminates in finitely many steps, and determines the function i uniquely. Let's work out an example in detail to see this in practice.

Example 1.10.1. Let's take

$$f(x,y) = y^2 - x^3 + x,$$
$$g(x,y) = (x^2 + y^2 - 3x)^2 - 4x^2(2-x),$$

and $P = (0,0)$. For simplicity, we work out the case when $\operatorname{ch} k \neq 2$, and leave this (easier) case to the reader. Note that C_f and C_g do not share a component because f is irreducible and $C_f \not\subseteq C_g$: and plugging in $y^2 = x^3 - x$ into g recovers nonzero polynomial

$$x^2(x-1)(x^3 + 3x^2 - 4x - 8),$$

which has finitely many roots. Let's now apply Step (c).

(1) We have

$$f(x,0) = -x(x-1)(x+1) \text{ and } g(x,0) = x^2(x-1)^2,$$

so that $d = 3$ and $e = 4$. Applying the Euclidean algorithm gives us $n = 1$ with

$$q_1(x) = -x + 2, \qquad r_1(x) = 2x(x-1),$$
$$q_2(x) = -\frac{1}{2}(x+1), \quad r_2(x) = 0.$$

Then

$$h_1 = y^4 + (2x^2 - 5x - 2)y^2 + 2x(x-1) \text{ and}$$
$$h_2 = \frac{1}{2}y^2\left((1+x)y^2 + x(2x^2 - 3x - 7)\right).$$

Setting $(f_1, g_1) := (h_2, h_1)$, we are now in Case 2.

(2) Here $N = 2$ and

$$p_1 = \frac{1}{2}\left((1+x)y^2 + x(2x^2 - 3x - 7)\right).$$

Then

$$i_P(f_1, g_1) = 2 \cdot m_0(g_1(x,0)) + i_P(p_1, g_1) = 2 + i_P(p_1, g_1).$$

Setting $(f_2, g_2) := (g_1, p_1)$ (switching for degree reasons), we are again in Case 1.

(3) We have

$$f_2(x,0) = 2x(x-1) \text{ and } g_2(x,0) = \frac{1}{2}x(2x^2 - 3x - 7),$$

so that $d = 2$ and $e = 3$. Again, we get $n = 1$ with

$$q_1(x) = \frac{1}{2}x - \frac{1}{4}, \qquad r_1(x) = -4x,$$
$$q_2(x) = -\frac{1}{2}x + \frac{1}{2}, \quad r_2(x) = 0.$$

Then

$$h_1 = \left(-\frac{1}{2}x + \frac{1}{4}\right)y^4 + x\left(-x^2 + 3x + \frac{1}{4}\right)y^2 - 4x$$
$$h_2 = -\frac{1}{8}y^2\left((2x^2 - 3x - 7)y^2 + (4x^4 - 16x^3 - 5x^2 + 41x + 16)\right).$$

Setting $(f_3, g_3) := (h_2, h_1)$, we are now in Case 2.

(4) Here again $N = 2$ and

$$p_3 = -\frac{1}{8}\left((2x^2 - 3x - 7)y^2 + (4x^4 - 16x^3 - 5x^2 + 41x + 16)\right).$$

Then

$$i_P(f_3, g_3) = 2 \cdot m_0(g_3(x, 0)) + i_P(p_3, g_3) = 2 + i_P(p_3, g_3).$$

At this point, we have $i_P(p_3, g_3) = 0$, and the algorithm terminates.

We conclude that $i_P(f, g) = 4$. Get Desmos to draw some pictures to make sure you believe this!

To show existence, we first define the local ring of $\mathbb{A}^2_k$ at a point $P \in \mathbb{A}^2_k$.

Definition 1.10.2. Given a $P \in \mathbb{A}^2_k$, the local ring of $\mathbb{A}^2_k$ at P, denoted $\mathcal{O}_P$ or $\mathcal{O}_{\mathbb{A}^2_k, P}$, is the ring

$$\mathcal{O}_P := \{r \in k(x, y) : \text{there are } s, t \in k[x, y] \text{ s.t. } r = s/t \text{ and } t|_P \neq 0.\} \subset k(x, y).$$

Since $k[x, y]$ is a UFD and $k(x, y) = \operatorname{Frac} k[x, y]$, this ring can equivalently be defined as the set of $r \in k(x, y)$, which, when written in lowest terms as $r = s/t$ with $s, t \in k[x, y]$ and $t \neq 0$ satisfy $t|_P \neq 0$. We are now ready to sketch the proof of existence.

Proof Sketch of Existence in Theorem 1.9.9. Define

$$i_P(f, g) := \dim_k \mathcal{O}_P/(f, g)\mathcal{O}_P.$$

Properties (1), (5), and (6) are reasonably clear. To show (7), note that for $P = O = (0, 0)$, there is an evaluation map

$$\operatorname{eval}_P : \mathcal{O}_P \to k;$$

this is clearly surjective, and it is easy to see that its kernel is generated by x and y, whence we get an isomorphism

$$\mathcal{O}_P/(x, y)\mathcal{O}_P \xrightarrow{\sim} k$$

and so $i_P(x, y) = 1$. To show (3), note that if $f|_P \neq 0$, then $f \in \mathcal{O}_P^\times$, and so $(f, g)\mathcal{O}_P = \mathcal{O}_P$, and similarly if $g|_P \neq 0$. Conversely, if $f|_P = g|_P = 0$, then $(f, g)\mathcal{O}_P \subset \ker \operatorname{eval}_P$, so

$$\mathcal{O}_P/(f, g)\mathcal{O}_P \twoheadrightarrow \mathcal{O}_P/\ker \operatorname{eval}_P \cong k \text{ implies that } i_P(f, g) \geq 1.$$

To show (2), we may assume $P = O = (0, 0)$. First suppose that we have such a q; then $(f, g)\mathcal{O}_P \subset (q)\mathcal{O}_P$, and we get $\mathcal{O}_P/(f, g)\mathcal{O}_P \twoheadrightarrow \mathcal{O}_P/(q)\mathcal{O}_P$, so it suffices to show that $\mathcal{O}_P/(q)\mathcal{O}_P$ is not finite dimensional over k. To do this, we may assume by a linear change of coordinates that $y \nmid q$; we show that the classes of $1, y, y^2, \ldots$ in $\mathcal{O}_P/(q)\mathcal{O}_P$ are linearly independent. If they were not, then there would be a nonzero $p \in k[y]$ of least degree such that $p \in (q)\mathcal{O}_P$, which is to say that $p = qs/t$ for some nonzero $s, t \in k[x, y]$ with $t|_P \neq 0$. Then $p|_P = 0$ implies $y \mid p$, so if $y \nmid q$, then $y \mid s$, and we may cancel a y from both sides, contradicting our choice of p. Conversely, suppose that f and g have no common components through P. Since irreducible factors of f and g not through P are units in $\mathcal{O}_P$, we may assume by dividing by these factors that f and g are relatively prime in $k[x, y]$. Then, as in Example 1.6.4, Lemma 1.6.2 tells us that there are nonzero $p \in k[x]$ and $q \in k[y]$ such that $p, q \in (f, g)k[x, y] \subset (f, g)\mathcal{O}_P$. Now if we write $p = x^m p_0$ for some $m \geq 0$ and $p_0 \in k[x]$ with $p_0(0) \neq 0$, then $m \geq 1$ because $p \in \ker \operatorname{eval}_P$, and $p_0 \in \mathcal{O}_P^\times$, so that $x^m \in (f, g)\mathcal{O}_P$. Similarly, from q we get an integer $n \geq 1$ such that $y^n \in (f, g)\mathcal{O}_P$. Then it follows that any rational function of the form $1/t$ with $t|_P \neq 0$ can be

expanded in $\mathcal{O}_P/(f,g)\mathcal{O}_P$ as $\sum_{i\geq 0}(1-t)^i$, where all but finitely many terms are zero because of $[x^n]=[y^m]=0$. It is then easy to see that the classes of the monomials x^iy^j with $0\leq i\leq m-1$ and $0\leq j\leq n-1$ span $\mathcal{O}_P/(f,g)\mathcal{O}_P$ as a k-vector space. Finally, to show (4), the result boils down to showing that there is a short exact sequence of the form

$$0\to \mathcal{O}_P/(f_1,g)\mathcal{O}_P \xrightarrow{\cdot f_2} \mathcal{O}_P/(f_1f_2,g)\mathcal{O}_P \to \mathcal{O}_P/(f_1,g)\mathcal{O}_P\to 0,$$

and the rank-nullity theorem. For full details, see [3, §3.3, Theorem 3] or [4, Chapter 2]. ■

1.10.2 The Projective Plane

As we have observed before, to count intersection points of curves properly, we have the need for a systematic way to study intersection points "at infinity". One way to do this is to note that every collection of parallel lines has a unique representative through the origin, and so points at infinity should correspond to lines through the origin–which are determined by their slope. Therefore, one approach would be to parametrize points at infinity via a parameter $t\in k$, where t corresponds to the point at infinity along the line $y-tx=0$. However, this misses exactly one line: namely the vertical line $x=0$, for which the value of t "would be" ∞.

A more symmetrical approach is to note that lines through the origin can be written as $\lambda x-\mu y=0$, where $\lambda,\mu\in k$ are not both zero, and the pair (λ,μ) determines the same line as $(c\lambda,c\mu)$ for every $c\in k\smallsetminus\{0\}$, so when $\mu\neq 0$, this corresponds to the above with $t=\lambda/\mu$, but when $\mu=0$, this adds the line $x=0$. In this case, we denote the "coordinates" of the line by $[\lambda:\mu]$ to emphasize that only the ratio between the coordinates matters. This gives us a way to think of the "projective plane" $\mathbb{P}^2_k$ as the disjoint union of points $(p,q)\in\mathbb{A}^2_k$ and the directions $[\lambda:\mu]$, but in fact there is a more symmetric way to do it. This leads us to

Definition 1.10.3. The projective plane over k, denoted $\mathbb{P}^2_k$, is the set of equivalence classes of ordered triples (X,Y,Z) of elements of k, not all zero, subject to the equivalence relation that $(X,Y,Z)\sim(\lambda X,\lambda Y,\lambda Z)$ for all $\lambda\in k\smallsetminus\{0\}=k^*$, i.e.

$$\mathbb{P}^2_k=\frac{\{(X,Y,Z)\in k^3\smallsetminus\{(0,0,0)\}\}}{(X,Y,Z)\sim(\lambda X,\lambda Y,\lambda Z)\,\forall\,\lambda\in k^*}.$$

The class of a triple (X,Y,Z) in $\mathbb{P}^2_k$ is usually denoted by $[X:Y:Z]$, and X,Y,Z are called the homogeneous coordinates on $\mathbb{P}^2_k$.

Note that the homogeneous coordinates are not well-defined functions on $\mathbb{P}^2_k$–only their ratios are, and those too only away from the loci where the denominator vanishes. Note also that $[0:0:0]$ is not a well-defined point in $\mathbb{P}^2_k$. Homogeneous coordinates were introduced by Möbius in his 1827 treatise *Der Barycentrische Calcül*. This way of thinking about $\mathbb{P}^2_k$ is in a sense the same as that from before: if $Z\neq 0$, then the point $[X:Y:Z]$ has a unique representative of the form $[x:y:1]$ where $x:=X/Z$ and $y:=Y/Z$, and these are the points that compose the $\mathbb{A}^2_k\subset\mathbb{P}^2_k$. When $Z=0$, however, we get points of the form $[X:Y:0]$, and these are exactly the points at ∞. One way to think about them is to think of them as the points that are limits of affine the form $[X/\varepsilon:Y/\varepsilon:1]$ as $\varepsilon\to 0$. The advantage of this formulation is that it makes some additional symmetry–namely that between X,Y, and Z, obvious–which we will leverage to great effect.

Note that in the case of the projective plane, the distinction between polynomials and polynomial functions becomes even more crucial: an arbitrary polynomial $F\in k[X,Y,Z]$ does not even define a well-defined function $F:\mathbb{P}^2_k\to k$ because picking a different representatives

(X, Y, Z) of a point $P = [X : Y : Z]$ will in general (i.e. for nonconstant F) yield different values under the polynomial function (on $\mathbb{A}^3_k$) arising from F. However, if F is homogenous of degree $d \geq 0$, then we see that for any $c \in k^\times$ we have

$$F(cX, cY, cZ) = c^d F(X, Y, Z),$$

whence the locus of points $P = [X : Y : Z] \in \mathbb{P}^2_k$ where $F|_P = 0$ still makes sense. This leads us to

Definition 1.10.4. A projective plane algebraic curve is the vanishing locus of a nonconstant homogeneous polynomial F in the projective plane, i.e. a subset $C \subset \mathbb{P}^2_k$ of the form

$$C = C_F := \{P \in \mathbb{P}^2_k : F|_P = 0\}$$

for a nonconstant homogeneous polynomial $F(X, Y, Z) \in k[X, Y, Z]$.

Next time, we'll define the homogenization of a polynomial and the projective closure of algebraic curves in more detail. Today, I want to end with one example.

Example 1.10.5. Consider the hyperbola C_f defined by $f(x, y) = xy - 1 \in k[x, y]$. Then the homogenization of f is $F = f^h = XY - Z^2 \in k[X, Y, Z]$, and the projective closure of C is the curve

$$\overline{C_f} = C_F = \{P = [X : Y : Z] \in \mathbb{P}^2 : XY - Z^2 = 0\}.$$

The intersection $C_F \cap \mathbb{A}^2_k$ is exactly C_f; on the other hand, the new points at infinity correspond to solutions to $XY - Z^2 = Z = 0$, which are the two points $[1 : 0 : 0]$ and $[0 : 1 : 0]$. These are the two points corresponding to the two asymptotes of C_f, namely the lines $x = 0$ and $y = 0$. In particular, over $k = \mathbb{R}$, the two branches which are disjoint in $\mathbb{A}^2_{\mathbb{R}}$ connect up to form one "continuous loop" in $\mathbb{P}^2_{\mathbb{R}}$.

1.11 07/03/24 - Projective Duality, (De)Homogenization, Projective Nullstellensatz

Last time, we introduced the projective plane and projective curves. Let's start by looking at an extended example first.

1.11.1 Projective Lines and Projective Duality

Definition 1.11.1. A projective line is a projective curve of the form $L = C_F \subset \mathbb{P}^2_k$ for a nonconstant linear homogeneous polynomial $F \in k[X, Y, Z]_1$.

Once we have the notion of degrees for projective curves below (§1.11.3), then we'll see that a projective line is a projective curve of degree 1.

Example 1.11.2. If $F = Z$, then $L = C_F$ is called the line at infinity and denoted L_∞.

Any linear homogeneous F is specified as $F = AX + BY + CZ$, where $A, B, C \in k$ are not all zero. Note that multiplying F by a nonzero scalar $\lambda \in k^\times$ does not affect C_F. Analogously to the affine case, we will see (in Theorem 1.11.17) that $L = C_F$ recovers F up to nonzero scalars, and hence we get a bijection between the set of lines $L \subset \mathbb{P}^2_k$ and the set of ordered triples (A, B, C) of elements of k, not all zero, subject to the equivalence $(A, B, C) \sim (\lambda A, \lambda B, \lambda C)$ for all $\lambda \in k^*$–but that's just another projective plane! We denote this projective plane by $\mathbb{P}^{2*}_k := \mathbb{P}^2_k(A, B, C)$, so we have a bijection

$$\{\text{lines } L \subset \mathbb{P}^2_k\} \leftrightarrow \mathbb{P}^{2*}_k.$$

Note that points in $\mathbb{P}^{2*}_k$ correspond to lines in $\mathbb{P}^2_k$, but the symmetry of the equation

$$AX + BY + CZ = 0$$

tells us that *lines* in $\mathbb{P}^{2*}_k$ correspond to *points* in $\mathbb{P}^2_k$–and indeed, if a point $P \in \mathbb{P}^2_k$ corresponds to the line $P^* \in \mathbb{P}^{2*}_k$, and the line $L \subset \mathbb{P}^2_k$ corresponds to the point $L^* \in \mathbb{P}^{2*}_k$, then we have

$$P \in L \Leftrightarrow P^* \ni L^*.$$

This funny phenomenon of interchanging the set of lines in one projective plane with the set of points in another is called the phenomenon of projective duality. Duality is a powerful tool that allows us to start with statements about points, lines, and incidences, and produce corresponding "dual" statements–effectively doubling the number of statements we can prove about the projective plane with very little effort. This is because this duality carries with it a lot of structure.

Consider, for instance, the following asymmetry: in $\mathbb{A}^2_k$, given any two points, there is a unique line passing through them, but given any two lines, they either interesect in a unique point or not at all (i.e. if they are parallel). In the projective plane, duality asserts that this asymmetry cannot happen.

Proposition 1.11.3. Given any two distinct points $P_1, P_2 \in \mathbb{P}^2_k$, there is a unique line $L \subset \mathbb{P}^2_k$ through them, and given two distinct lines $L_1, L_2 \subset \mathbb{P}^2_k$, they intersect in a unique point.

Proof. The second assertion follows from the first applied to $\mathbb{P}_k^{2*}$ (i.e. by duality), and so it suffices to show the first one. Suppose we write $P_1 = [X_1 : Y_1 : Z_1]$ and $P_2 = [X_2 : Y_2 : Z_2]$; then we are trying to solve simultaneously the system of equations

$$AX_1 + BY_1 + CZ_1 = 0$$
$$AX_1 + BY_2 + CZ_2 = 0$$

for A, B, C, not all zero, up to scaling. Multiplying the first equation by Y_2 and the second by Y_1 and subtracting yields

$$A(X_1Y_2 - X_2Y_1) + C(Z_1Y_2 - Z_2Y_1) = 0.$$

Similarly, we obtain two other equations of this sort. It follows easily (check!) that there is a solution to the above system of equations, up to scalars, given by

$$[A : B : C] = [Y_1Z_2 - Y_2Z_1 : Z_1X_2 - Z_2X_1 : X_1Y_2 - X_2Y_1],$$

where at least one of the expressions $Y_1Z_2 - Y_2Z_1, Z_1X_2 - Z_2X_1$, and $X_1Y_2 - X_2Y_1$ is nonzero because $P_1 \neq P_2$ (why?). ∎

Similarly, the question of collinearity of three points in $\mathbb{P}_k^2$ is answered by

Proposition 1.11.4. Given points $P_1, P_2, P_3 \in \mathbb{P}_k^2$, write $P_i = [X_i : Y_i : Z_i]$ for $i = 1, 2, 3$. The points P_1, P_2 and P_3 are collinear iff

$$\det \begin{bmatrix} X_1 & Y_1 & Z_1 \\ X_2 & Y_2 & Z_2 \\ X_3 & Y_3 & Z_3 \end{bmatrix} = 0.$$

Proof. The points P_1, P_2, P_3 are collinear iff there are $A, B, C \in k$, not all zero, such that $AX_i + BY_i + CZ_i = 0$ for $i = 1, 2, 3$. This can be rephrased by asking for A, B, C, not all zero, such that

$$\begin{bmatrix} X_1 & Y_1 & Z_1 \\ X_2 & Y_2 & Z_2 \\ X_3 & Y_3 & Z_3 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

and then the result follows from simple linear algebra: if the determinant of this matrix were nonzero, it would be invertible (by Cramer's rule, say), and so we would conclude from such an equation that $A = B = C = 0$, and conversely, if the determinant *is* zero, then there is a nonzero vector in the kernel of the linear map determined by it. ∎

Note that projective duality tells us that concurrent triples of lines $L_1, L_2, L_3 \subset \mathbb{P}_k^2$ correspond exactly to collinear triples of points in $\mathbb{P}_k^{2*}$, and we get a corresponding criterion for concurrency of lines, which I will leave to you to formulate.

Of course, this statement automatically implies a corresponding statement in the affine plane (Corollary 1.11.5)) as well, but somehow I have always found the projective ase easier to understand conceptually.

Corollary 1.11.5. Given points $p_1, p_2, p_3 \in \mathbb{A}^2_k$ with coordinates $p_i = (x_i, y_i)$, we have that p_1, p_2, p_3 are collinear iff

$$\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} = (x_2y_3 - x_3y_2) + (x_3y_1 - x_1y_3) + (x_1y_2 - x_2y_1) = 0.$$

Proof. The points p_i are collinear in $\mathbb{A}^2_k$ iff the points $P_i := [x_i : y_i : 1]$ are collinear in $\mathbb{P}^2_k$. ■

Remark 1.11.6. Note that similarly to how projective lines in $\mathbb{P}^2_k$ are parametrized by another $\mathbb{P}^2_k$, it is clear that curves $C \subset \mathbb{P}^2_k$ of a fixed degree (interpreted appropriately, i.e. with multiplicity) are also parametrized by a projective space of higher dimension. For instance, a conic section $C \subset \mathbb{P}^2_k$ is specified by a homogeneous quadratic polynomial

$$F = AX^2 + BXY + CY^2 + DXZ + EXZ + FZ^2,$$

which amounts to giving a 6-tuple (A, B, C, D, E, F) of elements of k, not all zero, up to simultaneous scaling: in other words, the set of all conics $C \subset \mathbb{P}^2_k$ is a $\mathbb{P}^5_k$. More generally, the set of all degree $d \geq 1$ curves $C \subset \mathbb{P}^2_k$ is a projective space $\mathbb{P}^{d(d+3)/2}_k$, and even more generally, the set of all degree $d \geq 1$ hypersurfaces $Z \subset \mathbb{P}^n_k$ for $n \geq 1$ is given by a projective space $\mathbb{P}^{\binom{d+n}{n}-1}_k$. (Think about what this could mean–I haven't defined projective spaces of higher dimensions for you yet!) This idea of parameter spaces in our own category is unique to algebraic geometry–for instance, the set of submanifolds of a smooth manifold does not have the structure of a finite-dimensional manifold in any way. This notion of parameter spaces is one of the most powerful tools in modern algebraic geometry: the geometry of a parameter space often dictates the behavior of the objects it parametrizes. I will not dwell on this further, but I would encourage you to think about this as and when this idea shows up in your further studies.

1.11.2 (De)Homogenization, Projective Closure and Affine Part

Let's now start talking about the relationship between affine and projective curves. For this, we first need some algebraic definitions.

Definition 1.11.7.

(a) Given a polynomial $f \in k[x, y]$ of degree $d \geq 0$, we define its homogenization, written f^{h}, to be

$$f^{\mathrm{h}}(X, Y, Z) := Z^d f\left(\frac{X}{Z}, \frac{Y}{Z}\right) \in k[X, Y, Z]_d.$$

In other words, if $f \neq 0$ and we write $f = f_0 + f_1 + \cdots + f_d$ with each f_i homogeneous of degree i, and $f_d \neq 0$, then we have

$$f^{\mathrm{h}}(X, Y, Z) = Z^d f_0(X, Y) + Z^{d-1} f_1(X, Y) + \cdots + f_d(X, Y).$$

(b) Given a homogeneous polynomial $F \in k[X, Y, Z]$, we define the inhomogeneous part or dehomogenization of F with respect to Z, denoted F^{i}, to be

$$F^{\mathrm{i}}(x, y) := F(x, y, 1) \in k[x, y].$$

We will use simple properties of these operations such as $(fg)^{\mathrm{h}} = f^{\mathrm{h}} g^{\mathrm{h}}$ for nonzero $f, g \in k[x, y]$ without further comment. Note that although we have for any $f \in k[x, y]$ that

$(f^{\mathrm{h}})^{\mathrm{i}} = f$, the operations $f \mapsto f^{\mathrm{h}}$ and $F \mapsto F^{\mathrm{i}}$ are not inverse bijections in general. For any nonzero $f \in k[x,y]$ of degree d, the homogenization f^{h} is homogeneous of degree $d = \deg f$, and $Z \nmid f^{\mathrm{h}}$ because $f_d \neq 0$; therefore, if $Z \mid F$, then we cannot possibly have $F = (F^{\mathrm{i}})^{\mathrm{h}}$. However, this is the only problem: we have for any nonzero $F \in k[X,Y,Z]$ that is homogeneous of degree $d \geq 0$ that if $F = Z^m F_0$ for some $m \geq 0$ and $F_0 \in k[X,Y,Z]_{d-m}$ not divisible by d, then $(F^{\mathrm{i}})^{\mathrm{h}} = F_0$, whence $F = Z^m (F^{\mathrm{i}})^{\mathrm{h}}$. In particular, if $Z \nmid F$, then $F = (F^{\mathrm{i}})^{\mathrm{h}}$. Phrased slightly differently, we have

Lemma 1.11.8. For any $d \geq 0$, the operations $f \mapsto f^{\mathrm{h}}$ and $F \mapsto F^{\mathrm{i}}$ give inverse bijections between the set of all nonzero polynomials $f \in k[x,y]$ of degree d and the set of all nonzero homgeneous polynomials $F \in k[X,Y,Z]$ of degree d such that $Z \nmid F$.

The parallel definitions in geometry are as follows.

Definition 1.11.9.

(a) Given an affine curve $C \subset \mathbb{A}^2_k$, we define its projective closure, denoted $\overline{C}$, to be $\overline{C} := C_{f^{\mathrm{h}}}$, where $f \in k[x,y]$ is any polynomial such that $C = C_f$. Given any affine curve $C \subset \mathbb{A}^2_k$, we define the set of points at infinity along C to be $\overline{C} \cap L_\infty$.

(b) Given a projective curve $C \subset \mathbb{P}^2_k$, we define its affine part in the chart defined by $Z \neq 0$ to be $C^\circ := C \cap \mathbb{A}^2_k = \{P \in C : P = [X : Y : Z] \text{ and } Z \neq 0\} = C_{F^{\mathrm{i}}}$, where $F \in k[X,Y,Z]$ is any homogeneous polynomial such that $C = C_F$.

The first thing to note here is that if $f, g \in k[x,y]$ are polynomials such that $C_f = C_g$, then $C_{f^{\mathrm{h}}} = C_{g^{\mathrm{h}}}$, making the projective closure well-defined; similarly, if $F, G \in k[X,Y,Z]$ are homogeneous polynomials such that $C_F = C_G$, then $C_{F^{\mathrm{i}}} = C_{G^{\mathrm{i}}}$ (which is somewhat easier to see from the alternative description). Next, we note that if $C \subset \mathbb{A}^2_k$ has degree $d \geq 1$, then $\overline{C}$ is obtained by attaching at most d new points to C (i.e. there are at most d points at infinity along C); namely, if we write $f = f_0 + \cdots + f_d$, then points of $\overline{C} \smallsetminus C$ correspond to roots of the homogeneous polynomial $f_d(X,Y)$, of which there are at most d distinct values. This observation has the amusing consequence that an algebraic curve of degree d in $\mathbb{A}^2_k$ can have at most d distinct asymptotes.[23] Finally, we have as before that if $C \subset \mathbb{A}^2_k$ is an affine curve, then $(\overline{C})^\circ = C$, but the operations $C \mapsto \overline{C}$ and $C \mapsto C^\circ$ are not inverse bijections: if we consider the line at infinity L_∞, then $L_\infty^\circ = \emptyset$, whence $\overline{L_\infty^\circ} = \emptyset$ as well. Again, this is the only problem, and if $C \subset \mathbb{P}^2_k$ is any projective curve other than L_∞, then C° is a nonempty affine curve. In fact, we have

Lemma 1.11.10. If $C \subset \mathbb{P}^2_k$ is a projective curve, then either $L_\infty \not\subset C$, in which case we have $C = \overline{C^\circ}$, or we have $C = \overline{C^\circ} \cup L_\infty$.

Proof. Left to the reader. ∎

Remark 1.11.11. The terminology "projective closure" comes from topology: there is a topology on $\mathbb{P}^2_k$ called the Zariski topology, in which $\overline{C}$ is just the ordinary topological closure of $C \subset \mathbb{A}^2_k \subset \mathbb{P}^2_k$. Understanding the Zariski topology is absolutely fundamental to appreciating more advanced algebraic geometry, but we don't need to worry too much about it right now.

The goal of this translation is that it allows us to port over the work that we did in the affine case to the projective case without a lot of additional effort. This is what we do now. Let's do a couple of examples.

[23] What are those?

Proposition 1.11.12. If $C, D \subset \mathbb{P}^2_k$ are projective curves, then so is $C \cup D$.

Proof. If $C = C_F$ and $D = C_G$, then $C \cup D = C_{F \cdot G}$. ∎

Proposition 1.11.13. If k is an algebraically closed field and $C \subset \mathbb{P}^2_k$ is a projective curve, then $C = C(k)$ is infinite.

Proof. Either $C = L_\infty$, in which case we are done because k is infinite (how?), or C° is an affine curve, so we are done by Lemma 1.5.1. ∎

Let's now move on to a few more things that follow easily.

1.11.3 Homogeneous Unique Factorization, Nullstellensatz, etc.

Lemma 1.11.14. If $F, G \in k[X, Y, Z]$ are such that F is homogeneous and $G \mid F$, then G is homogeneous.

Proof. Write $F = GH$, and suppose that the degrees of F, G, H are $d, m, n \geq 0$ with $m + n = d$. If $m = 0$ or $n = 0$, then the result is clear; hence assume that $m, n \geq 1$, so $d \geq 2$. Expand $G = G_0 + G_1 + \cdots + G_m$ and $H = H_0 + \cdots + H_n$ with each G_i (resp. each H_j) homogeneous of degree i (resp. j), and $G_m \neq 0$ (resp. $H_n \neq 0$). Let i be the least non-negative integer such that $G_i \neq 0$, so that $0 \leq i \leq m$; similarly, let j be the least non-negative integer such that $H_j \neq 0$. Then the degree $i + j$ component of $F = GH$ is $G_i H_j$, which is nonzero; since we assumed that F is homogeneous of degree d, this implies that $i + j = d$, whence $i = m$ and $j = n$, showing that both G and H are homogeneous. ∎

From this, we immediately obtain a homogeneous analog of unique factorization in $k[X, Y, Z]$, namely

Theorem 1.11.15 (Homogeneous Unique Factorization). Every nonconstant homogeneous $F \in k[X, Y, Z]$ can be factored as

$$F = F_1 \cdots F_n,$$

a product of finitely many homogeneous irreducible elements $F_1, \ldots, F_n \in k[X, Y, Z]$, and this factorization is unique up to the order of the elements and multiplication by units.

I will leave to you to make the last statement precise (say along the lines of Definition 1.5.7.)

Proof. Immediate consequence of unique factorization in $k[X, Y, Z]$ (Corollary 1.5.14) and the Lemma 1.11.14 above. ∎

Now we can mimic the affine theory as follows. Firstly, the analog of Theorem 1.6.6 is

Theorem 1.11.16 (Projective Finite Intersection). Let $F, G \in k[X, Y, Z]$ be nonconstant relatively prime homogeneous polynomials. Then $C_F \cap C_G$ is finite.

Proof. Note that Z cannot divide both F and G; without loss of generality, suppose that $Z \nmid G$. Since

$$C_F \cap C_G \subset (C_F^\circ \cap C_G^\circ) \cup (L_\infty \cap C_G),$$

it suffices to show that both $C_F^\circ \cap C_G^\circ$ and $L_\infty \cap C_G$ are finite. The latter is easy: if $\deg G = d \geq 0$, and we write

$$G = Z^d G_0(X,Y) + Z^{d-1} G_1(X,Y) + \cdots + G_d(X,Y),$$

where each $G_j(X,Y) \in k[X,Y]_j$ is homogeneous of degree d, then $Z \nmid G$ implies that $G_d \neq 0$, whence $L_\infty \cap C_G$ corresponds to the finitely many roots of the homogeneous polynomial $G_d(X,Y)$, of which there are at most d.[24] To show the former, note that $C_F^\circ \cap C_G^\circ = C_{F^\mathrm{i}} \cap C_{G^\mathrm{i}}$, so in light of Theorem 1.6.6, it suffices to show that if $F, G \in k[X,Y,Z]$ are nonconstant relatively prime homogeneous polynomials, then the dehomogenizations $F^\mathrm{i}, G^\mathrm{i} \in k[x,y]$ are also relatively prime (although no longer necessarily nonconstant). To show this statement, it suffices note that if $q \in k[x,y]$ is such that $q | F^\mathrm{i}$, then $F^\mathrm{i} = pq$ for some $p \in k[x,y]$, whence $q^\mathrm{h} \mid p^\mathrm{h} q^\mathrm{h} = (F^\mathrm{i})^\mathrm{h} \mid F$; then, if a nonconstant $q \in k[x,y]$ were to divide both F^i and G^i, then the nonconstant[25] $q^\mathrm{h} \in k[X,Y,Z]$ would divide F and G, contradicting their relative primality. ∎

This theorem was the key to the Nullstellensatz, and all of its corollaries, which we collect in one theorem here.

Theorem 1.11.17 (Projective Nullstellensatz)**.** Suppose that k is an algebraically closed field.

(a) If $F, G \in k[X,Y,Z]$ are nonconstant homogeneous polynomials, then $C_G \subset C_F$ iff there is some integer $n \geq 1$ such that $G \mid F^n$.
(b) If $F, G \in k[X,Y,Z]$ are nonconstant homogeneous polynomials with F irreducible, then $C_G \subset C_F$ implies $C_G = C_F$.
(c) If $F \in k[X,Y,Z]$ is a nonconstant homogeneous polynomial, then C_F is irreducible.[a] Conversely, if $C \subset \mathbb{P}_k^2$ is an irreducible projective curve, then there is an irreducible homogeneous $F \in k[X,Y,Z]$ such that $C = C_F$.

[a] You were invited to define the notion of irreducibility for projective curves in Exercise 2.4.2.

Proof.

(a) Identical to the proof of Theorem 1.6.7: if Q is a prime factor of G, then Q is homogeneous by Lemma 1.11.14, and then if Q and F were relatively prime, then $C_Q \cap C_F = C_Q$ would be finite by Theorem 1.11.16 but infinite by Proposition 1.11.13.
(b) Identical to the proof of Corollary 1.6.8 using (a) instead of Theorem 1.6.7.
(c) Identical to the proof of Theorem 1.5.6, and left to the reader.

∎

Similarly to the affine case, given a projective curve $C \subset \mathbb{P}_k^2$, we can try to define a vanishing ideal $\mathbb{I}(C) \subset k[X,Y,Z]$ of C consisting of homogeneous polynomials vanishing on C, but the problem is that the sum of two homogeneous polynomials of different degrees is not homogeneous. The correct definition is

[24] In other words, we have $[X_0 : Y_0 : Z_0] \in L_\infty \cap C_G$ iff $Z_0 = 0$ and $G_d(X_0, Y_0) = 0$, but the latter condition constrains the ratio $[X_0 : Y_0]$ to be one of the homogeneous roots of $G_d(X_0, Y_0)$, i.e. if we factor G_d using Lemma 1.8.3 (and Theorem 1.4.5 if needed) as $G_d = \prod_{i=1}^d (\lambda_i X + \mu_i Y)$, then $[X_0 : Y_0]$ can only be one of the d possible choices for $[-\mu_i : \lambda_i]$.

[25] This uses $\deg q^\mathrm{h} = \deg q$.

Definition 1.11.18. Given a projective curve $C \subset \mathbb{P}^2_k$, we define the vanishing ideal of C to be

$$\mathbb{I}(C) := \{F \in k[X,Y,Z] : \text{if } F = F_0 + \cdots + F_d \text{ with } F_j \in k[X,Y,Z]_j \text{ then } C \subset C_{F_j} \text{ for all } j.\}$$

This is, in fact, an ideal of $k[X,Y,Z]$–and, indeed, a special kind of ideal called a homogeneous ideal.[26] Then the analog of Theorem 1.6.12 still holds: $\mathbb{I}(C)$ is a principal ideal generated by $\text{rad}(F)$ for any homogeneous $F \in k[X,Y,Z]$ such that $C = C_F$. A generator of $\mathbb{I}(C)$ is again called a minimal polynomial of C; any two of these differ by a nonzero scalar, and we define the degree of C to be the degree of any minimal polynomial for C. The analog of Corollary 1.6.13 still holds: over $k = \overline{k}$, there is a bijective correspondence between projective curves $C \subset \mathbb{P}^2_k$ and principal ideals of $k[X,Y,Z]$ generated by nonconstant reduced homogeneous $F \in k[X,Y,Z]$, and the curve C is irreducible iff $\mathbb{I}(C)$ is a prime ideal. Finally, we also have an analog of Theorem 1.7.10; let's write this down in some detail.

Theorem 1.11.19 (Projective Unique Decomposition)**.** If $k = \overline{k}$, then given any curve $C \subset \mathbb{P}^2_k$, there is an integer $n \geq 1$ and irreducible curves $C_1, \ldots, C_n \subset \mathbb{P}^2_k$ such that $C_i \neq C_j$ for $i \neq j$, such that

$$C = C_1 \cup C_2 \cup \cdots \cup C_n.$$

The integer n is uniquely determined, as are the C_j up to reordering.

Proof. Identical to the proof of Theorem 1.7.10. ∎

The curves $C_1, \ldots, C_n \subset C$ occuring in such a decomposition are called the irreducible comopnents of C. Finally, the analog of Theorem 1.7.11 is

Theorem 1.11.20 (Projective Finite Intersection Revisited)**.** If $C, D \subset \mathbb{P}^2_k$ are two curves that don't share any common irreducible components, then the intersection $C \cap D$ is finite.

Proof. Identical to the proof of Theorem 1.7.11. ∎

The three things from the affine case that we haven't transferred yet are (a) parametric curves, (b) changes of coordinates, and (c) (intersection) multiplicity. This we will do in the next two lectures.

1.11.4 Addendum: Irreducible Projective Curves

I did not have time to cover this in lecture, but I do want to explain the relationship between minimal polynomials and irreducibility of affine curves and their projective counterparts. This is the content of

[26] Can you come up with a good definition of this notion?

Theorem 1.11.21.

(a) If $f \in k[x,y]$ is irreducible (resp. reduced), then so is $f^{\mathrm{h}} \in k[X,Y,Z]$. Conversely, if a homogeneous $F \in k[X,Y,Z]$ is irreducible (resp. reduced), then so is $F^{\mathrm{i}} \in k[x,y]$, unless $F = \lambda Z^m$ for some $\lambda \in k^\times$ and $m \geq 0$ (resp. $m = 0, 1$), in which case, and only in which case, $F^{\mathrm{i}} = \lambda$ is a nonzero constant.

(b) If an affine curve $C \subset \mathbb{A}^2_k$ has minimal polynomial f, then its projective closure $\overline{C}$ has minimal polynomial f^{h}; in particular, $\deg C = \deg \overline{C}$. If $C \subset \mathbb{P}^2_k$ has minimal polynomial F, then its affine part C°, if nonempty, has minimal polynomial F^{i} and either

  (i) $L_\infty \subset C$ and $\deg C^\circ = \deg C - 1$ (where $\deg C^\circ = 0$ says just that $C^\circ = \emptyset$), or

  (ii) $L_\infty \not\subset C$ and $\deg C^\circ = \deg C$.

(c) If $C \subset \mathbb{A}^2_k$ is an irreducible affine curve, then its projective closure $\overline{C}$ is an irreducible projective curve. If $C \subset \mathbb{P}^2_k$ is an irreducible projective curve, then either $C^\circ = \emptyset$ (which happens iff $C = L_\infty$), or C° is an irreducible affine curve.

Proof.

(a) Let's treat irreducibility; the proof for reducedness is similar and left to the reader. If given an $f \in k[x,y]$, there is a $G \in k[X,Y,Z]$ such that $G \mid f^{\mathrm{h}}$ and $0 < \deg G < \deg f^{\mathrm{h}} = \deg f$, then G is homogeneous by Lemma 1.11.14 and $Z \nmid G$ because $Z \nmid f^{\mathrm{h}}$, from which we get that $G^{\mathrm{i}} \mid (f^{\mathrm{h}})^{\mathrm{i}} = f$ and $0 < \deg G^{\mathrm{i}} = \deg G < \deg f$; therefore, if f is irreducible, then so is f^{h}. Conversely, given a homogeneous $F \in k[X,Y,Z]$ that is not of the form λZ^m, we must have $\deg F^{\mathrm{i}} \geq 1$; if $g \in k[x,y]$ is such that $g \mid F^{\mathrm{i}}$ and $0 < \deg g < \deg F^{\mathrm{i}} \leq \deg F$, then $g^{\mathrm{h}} \mid (F^{\mathrm{i}})^{\mathrm{h}} \mid F$ with $0 < \deg g^{\mathrm{h}} = \deg g < \deg F$; therefore, if F is irreducible, then so is F^{i}.

(b) If an affine curve C has minimal polynomial f, then f is reduced, and so by (a) so is f^{h}; since f^{h} is a reduced homogeneous polynomial vanishing on $\overline{C}$, it follows that f^{h} is a minimal polynomial for $\overline{C}$. I will leave the rest to the reader.

(c) If C is an irreducible affine curve, then any minimal polynomial f for C is irreducible; then f^{h} is irreducible by (a) and a minimal polynomial for $\overline{C}$ by (b), and so it follows that $\overline{C}$ is an irreducible projective curve. The converse is again left to the reader.

■

Finally, the symmetry in X, Y, Z tells us that irreducibility of a given homogeneous $F \in k[X,Y,Z]$ is testable by dehomogenization with respect to any of the variables.

1.12 07/05/24 - Projective Changes of Coordinates, Multiplicity and Smoothness, Classification of Projective Conics

1.12.1 Projective Changes of Coordinates

We defined affine changes of coordinates by setting x and y to be linear polynomials in x' and y' subject to a nondegeneracy condition. We want to mimic the situation in the projective case: we want to set X, Y, Z to be three homogeneous linear polynomials $L, M, N \in k[X', Y', Z']$, but now we need to ensure nondegeneracy as well. If L, M, N were concurrent in $\mathbb{P}^2$, then this point of concurrency would be mapped to $[0:0:0]$, which doesn't make any sense; therefore, we need to at least ask that L, M, N be nonconcurrent. It turns out that in the projective case, this condition is also sufficient. The discussion in §1.11.1 gives us a direct condition to check to ensure nonconcurrency, and leads us to

Definition 1.12.1. A projective change of coordinates is a transformation

$$\Phi : \mathbb{P}^2_k(X', Y', Z') \to \mathbb{P}^2_k(X, Y, Z)$$

of the form

$$[X : Y : Z] = \Phi[X' : Y' : Z'] = [AX' + BY' + CZ' : DX' + EY' + FZ' : GX' + HY' + IZ']$$

for some $A, B, C, D, E, F, G, H, I \in k$ such that

$$\det \begin{bmatrix} A & B & C \\ D & E & F \\ G & H & I \end{bmatrix} \neq 0.$$

Again, the nondegeneracy condition on the determinant ensures that the transformation is both well-defined and, in fact, invertible: this is because the transformation is given before homogenization (i.e. quotienting by the equivalence relation of scaling) by the map

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} A & B & C \\ D & E & F \\ G & H & I \end{bmatrix} \begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix},$$

so if this transformation matrix has nonzero determinant, then by Cramer's rule it is an invertible matrix, and we can recover $[X' : Y' : Z']$ from $[X : Y : Z]$ using

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \begin{bmatrix} A & B & C \\ D & E & F \\ G & H & I \end{bmatrix}^{-1} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}.$$

Since, of course, two such transformation matrices define the same transformation if they differ by scalar multiples, the group of all projective changes of coordinates is the group $\mathrm{PGL}_3 k \subset \mathbb{P}^8_k$ of all 3×3 matrices in k with nonzero determinant up to simultaneously scaling by a nonzero scalar, i.e. $\mathrm{GL}_3 k$ subject to the equivalence relation $M \sim \lambda M$ for all $M \in \mathrm{GL}_3 k$ and $\lambda \in k^\times$. This scaling invariance implies that, unlike the affine case, a projective change of coordinates Φ does not quite give us a pullback map on the homogeneous polynomial ring $\Phi^* : k[X, Y, Z] \to k[X', Y', Z']$, but we can always choose such a pullback map which does what we want[27]; such a pullback map would then necessarily be an isomorphism, and any two such maps would be related by a nonscalar scalar.

[27] What do we want?

The key fact to note here is that projective changes of coordinates respect incidence. This is captured by

Lemma 1.12.2. Let $\Phi : \mathbb{P}^2_k(X', Y', Z') \to \mathbb{P}^2_k(X, Y, Z)$ be a projective change of coordinates. Then three points $P_1, P_2, P_3 \in \mathbb{P}^2_k(X', Y', Z')$ are collinear iff $\Phi(P_1), \Phi(P_2)$ and $\Phi(P_3)$ are.

Proof. Write $P_i = [X'_i : Y'_i : Z'_i]$ for $i = 1, 2, 3$. Using Proposition 1.11.4 and the fact that determinants are multiplicative and invariant under taking transposes, we conclude that

$$\begin{aligned} P_1, P_2, P_3 \text{ are collinear } &\Leftrightarrow \det \begin{bmatrix} X'_1 & X'_2 & X'_3 \\ Y'_1 & Y'_2 & Y'_3 \\ Z'_1 & Z'_2 & Z_3 \end{bmatrix} = 0 \\ &\Leftrightarrow \det \left(\begin{bmatrix} A & B & C \\ D & E & F \\ G & H & I \end{bmatrix} \begin{bmatrix} X'_1 & X'_2 & X'_3 \\ Y'_1 & Y'_2 & Y'_3 \\ Z'_1 & Z'_2 & Z_3 \end{bmatrix} \right) = 0 \\ &\Leftrightarrow \Phi(P_1), \Phi(P_2), \Phi(P_3) \text{ are collinear.} \end{aligned}$$

■

Lemma 1.12.2 and Proposition 1.11.3 tell us that projective changes of coordinates preserve all incidence geometry of $\mathbb{P}^2_k$: they take lines to lines, and incidences of points on lines to incidence of points on lines, concurrency of lines to concurrency of lines, etc.

Example 1.12.3. An affine change of coordinates of the form $(x, y) = (ax' + by' + p, cx' + dy' + q)$ is the affine "shadow" of a projective change of coordinates given by the matrix

$$\begin{bmatrix} a & b & p \\ c & d & q \\ 0 & 0 & 1 \end{bmatrix},$$

where the affine and projective nondegeneracy conditions are identical because the determinant of this matrix is $ad - bc$. Note that this "projectivization" of any affine change of coordinates fixes the line at infinity $L_\infty \subset \mathbb{P}^2_k$ as a set (although perhaps not pointwise!), and conversely, any projective change of coordinates that fixes the line at infinity must arise from an affine change of coordinates. Projective changes of coordinates are, however, more powerful, and treat all points (resp. lines) "equally," including points (resp. the line) at infinity.

From the construction itself, it is pretty clear that given any three tuple $L, M, N \in k[X, Y, Z]_1$ of homogeneous linear polynomials which vanish on three nonconcurrent lines, there is a change of coordinates taking the lines given by the vanishing of X, Y, Z to those given by L, M, N; in Exercise 2.4.8 you are invited to make this precise, and to explore whether such a transformation is unique. This incredible flexibility of projective transformations often makes explicit computations with projective curves really easy. Here's some terminology and a proposition we will have repeated ocassion to use.

Definition 1.12.4. A subset $S \subset \mathbb{P}^2_k$ is said to be in general position if no three points in S are collinear. We also say that the points $P_j \in S$ are in general position.

Proposition 1.12.5. Given any two ordered 4-tuples

$$P = (P_1, P_2, P_3, P_4) \text{ and } Q = (Q_1, Q_2, Q_3, Q_4)$$

of points in $\mathbb{P}^2_k$, both in general position, there is a unique projective change of coordinates $\Phi : \mathbb{P}^2_k \to \mathbb{P}^2_k$ taking one to the other, i.e. such that $\Phi(P_i) = Q_i$ for $i = 1, 2, 3, 4$.

Proof. It suffices to show the result when

$$P_1 = E_1 := [1:0:0], P_2 = E_2 := [0:1:0], P_3 = E_3 := [0:0:1] \text{ and } P_4 = E_4 := [1:1:1],$$

because then we can first uniquely take an arbitrary 4-tuple P to this standard 4-tuple E (because projective changes of curves are invertible), and then further take this standard 4-tuple to an arbitrary collection Q.[28] If we write $Q_i = [X_i : Y_i : Z_i]$, then any Φ taking $E_i \mapsto Q_i$ for $i = 1, 2, 3$ must be given by a matrix of the form

$$\begin{bmatrix} X_1 & X_2 & X_3 \\ Y_1 & Y_2 & Y_3 \\ Z_1 & Z_2 & Z_3 \end{bmatrix} \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \mu & 0 \\ 0 & 0 & \nu \end{bmatrix}$$

for some $\lambda, \mu, \nu \in k^\times$; that this matrix has nonzero determinant uses that Q_1, Q_2, Q_3 are non-collinear. Then the condition $E_4 \mapsto Q_4$ uniquely determines the triple (λ, μ, ν), up to simultaneous scaling, by the requirement that

$$\begin{bmatrix} X_1 & X_2 & X_3 \\ Y_1 & Y_2 & Y_3 \\ Z_1 & Z_2 & Z_3 \end{bmatrix} \begin{bmatrix} \lambda \\ \mu \\ \nu \end{bmatrix} = t \cdot \begin{bmatrix} X_4 \\ Y_4 \\ Z_4 \end{bmatrix}$$

for some $t \in k^\times$, since the matrix on the left is invertible; the fact that the resulting λ, μ, ν from this matrix equation are nonzero then is equivalent to saying that Q_4 does not lie in the lines $\overline{Q_2Q_3}, \overline{Q_1Q_3}$ and $\overline{Q_1Q_2}$ respectively.[29] ∎

1.12.2 Multiplicity, Smoothness, and Intersection Multiplicity

We would like to define the notions of multiplicity, smoothness, tangent lines, and intersection multiplicity in a way that is both invariant under projective changes of coordinates and compatible with dehomogenization. One way to do this is to define these local notions by first changing coordinates so that the point in consideration is $P = [0:0:1]$, and then use dehomogenization, and then rehomogenize–so for instance, the tangent line to a projective curve at a point would be the projective closure of its affine tangent line in some chart. This approach works, but has the disadvantage that checking invariance under projective changes of coordinates is a much more daunting task than in the affine case. A slightly more elegant approach is given by thinking about local rings.

Recall from Definition 1.10.2 that given a point $P \in \mathbb{A}^2_k$, we define its local ring $\mathcal{O}_{\mathbb{A}^2_k, P} \subset k(x, y)$ to consist of all rational functions on $\mathbb{A}^2_k$ which can be evaluated at P, in which case evaluation at P gives us a ring homomorphism

$$\text{eval}_P : \mathcal{O}_{\mathbb{A}^2_k, P} \to k$$

with kernel

$$I_{\mathbb{A}^2_k, P} := \ker \text{eval}_P$$

[28] Make this statement precise, particularly if it doesn't obviously make sense!
[29] Check this! This uses Cramer's rule.

consisting of all rational functions that vanish at P.[30] If $P = (0,0)$ is the origin, then $I_{\mathbb{A}^2_k,P}$ is an ideal of $\mathcal{O}_{\mathbb{A}^2_k,P}$ generated by x and y. It follows from this that $I^2_{\mathbb{A}^2_k,P}$ is generated by x^2, xy, and y^2, or more generally that for any $n \geq 1$, the ideal $I^n_{\mathbb{A}^2_k,P}$ is generated over $\mathcal{O}_{\mathbb{A}^2_k,P}$ by $x^n, x^{n-1}y, \ldots, xy^{n-1}, y^n$. In particular, we have that

$$\bigcap_{n\geq 0} I^n_{\mathbb{A}^2_k,P} = 0,$$

and hence given any nonzero polynomial $f \in k[x,y] \subset \mathcal{O}_{\mathbb{A}^2_k,P}$, there is a unique largest integer $m \geq 0$ such that $f \in k[x,y] \cap I^m_{\mathbb{A}^2_k,P}$. A moment's reflection shows that this m is nothing but the multiplicity $m_P(f)$ of f at P. This could have been used as an alternative definition of multiplicity, and this notion would then be somewhat visibly invariant under changes of coordinates, since the local ring $\mathcal{O}_{\mathbb{A}^2_k,P}$ and its maximal ideal $I_{\mathbb{A}^2_k,P}$ visibly behave well under changes of coordinates. Further, the above discussion gives us some added flexibility: the same definition applies to any nonzero element of $\mathcal{O}_{\mathbb{A}^2_k,P}$, and so we are now allowed to talk about intersection multiplicities of rational functions that one can evaluate at P. This is a crucial generalization needed to check the invariance of multiplicity under projective changes of coordinates irrespective of the chosen definition. Similarly, the notion of local intersection multiplicity is local: we observed in the proof of existence in Theorem 1.9.9 that $i_P(f,g)$ is just $\dim_k \mathcal{O}_{\mathbb{A}^2_k,P}/(f,g)\mathcal{O}_{\mathbb{A}^2_k,P}$, again dependent only on the local ring.

The above discussion tells us that if we can define projective analogs of the rational function field of $\mathbb{A}^2_k$ and of these local rings in a way that is compatible with taking affine charts, then we would be in good shape to define multiplicity in this case. And indeed, this is possible.

Definition 1.12.6. The rational function field of $\mathbb{P}^2_k$ is the subfield

$$k(\mathbb{P}^2_k) := \left\{ \frac{F}{G} \in k(X,Y,Z) : F, G \text{ are homogeneous of the same degree} \right\} \subset k(X,Y,Z).$$

Given a point $P \in \mathbb{P}^2_k$, we define the local ring of $\mathbb{P}^2_k$ at P to be the ring

$$\mathcal{O}_{\mathbb{P}^2_k,P} := \{r \in k(\mathbb{P}^2_k) : r = F/G \text{ for some homogeneous } F, G \in k[X,Y,Z] \text{ s.t. } G|_P \neq 0.\}$$

Evaluation at P gives us a surjective map

$$\text{eval}_P : \mathcal{O}_{\mathbb{P}^2_k,P} \to k$$

whose kernel we will denote by $I_{\mathbb{P}^2_k,P}$. Finally, given any integer $n \geq 1$ and homogeneous $F_1, \ldots, F_n \in k[X,Y,Z]$, we define the ideal $(F_1, \ldots, F_n)\mathcal{O}_{\mathbb{P}^2_k,P}$ to consist of all linear combinations of the form

$$\sum_{i=1}^{n} \frac{H_i}{G_i} \cdot F_i,$$

where the $H_i, G_i \in k[X,Y,Z]$ are homogeneous such that $\deg H_i = \deg F_i + \deg G_i$ and after cancellation of common factors we have $G_i|_P \neq 0$.

Given this, we are now ready to handle defining multiplicity of a homogeneous polynomial at a point $P \in \mathbb{P}^2_k$ and the intersection multiplicity of two polynomials, etc.

[30] Some textbooks denote this kernel by $\mathfrak{m}_{\mathbb{A}^2_k,P}$ or $\mathfrak{m}_P$ to emphasize that it is a maximal ideal of $\mathcal{O}_{\mathbb{A}^2_k,P}$, but we will not need this idea and I will stick to $I_{\mathbb{A}^2_k,P}$ or I_P.

Definition 1.12.7. Let $P \in \mathbb{P}^2_k$ be a point.

(a) Given a nonzero homogeneous polynomial $F \in k[X,Y,Z]$, we define the multiplicity of F at P to be the largest integer $m \geq 0$ such that

$$(F)\mathcal{O}_{\mathbb{P}^2_k,P} \subset I^m_{\mathbb{P}^2_k,P}.$$

(b) Given a curve $C \subset \mathbb{P}^2_k$ and point $P \in \mathbb{P}^2_k$, we define the multiplicity of C at P to be

$$m_P(C) := m_P(F)$$

where F is any minimal polynomial for C.

(c) Given two nonzero homogeneous polynomials $F, G \in k[X,Y,Z]$, we define the local intersection multiplicity of F and G at P to be

$$i_P(F,G) := \dim_k \mathcal{O}_{\mathbb{P}^2_k,P}/(F,G)\mathcal{O}_{\mathbb{P}^2_k,P}.$$

(d) Given two curves $C, D \subset \mathbb{P}^2_k$ and point $P \in \mathbb{P}^2_k$, we define the intersection multiplicity of C and D at P to be

$$i_P(C,D) := i_P(F,G)$$

where F, G are any minimal polynomials for C and D.

These definitions have the advantage of being visibly invariant under projective changes of coordinates, but we observe also that they are compatible with definitions from the affine case: setting $x := X/Z$ and $y := Y/Z$ gives us an isomorphism

$$k(\mathbb{P}^2_k) \xrightarrow{\sim} k(x,y)$$

with the property that if $P = [x_0 : y_0 : 1] \in \mathbb{A}^2_k \subset \mathbb{P}^2_k$, then this map takes

$$\mathcal{O}_{\mathbb{P}^2_k,[x_0:y_0:1]} \xrightarrow{\sim} \mathcal{O}_{\mathbb{A}^2_k,(x_0,y_0)} \text{ and } I_{\mathbb{P}^2_k,[x_0:y_0:1]} \xrightarrow{\sim} I_{\mathbb{A}^2_k,(x_0,y_0)}.$$

From this isomorphism and our above discussion on multiplicity, it follows immediately that if $P \in \mathbb{A}^2_k \subset \mathbb{P}^2_k$ and $F \in k[X,Y,Z]$ is a nonzero homogeneous polynomial, then

$$m_P(F) = m_P(F^{\mathrm{i}}),$$

and similarly that if $F, G \in k[X,Y,Z]$ are nonzero homogeneous polynomials, then

$$i_P(F,G) = i_P(F^{\mathrm{i}}, G^{\mathrm{i}}).$$

It follows from this that the function i satisfies axioms similar to (1)-(7) and is also completely characterized by them. Henceforth, we will use notions of (intersection) multiplicity for projective curves without further comment.

Remark 1.12.8. One can reasonably ask: which subring of $k(x,y)$ does $\mathcal{O}_{\mathbb{P}^2_k,P}$ map to when $P \in L_\infty$? The answer is pretty fun to work out and straightforward: if $P = [1:0:0]$, then $\mathcal{O}_{\mathbb{P}^2_k,P} \subset k(x,y)$ corresponds to the ring

$$k\left[\frac{y}{x}, \frac{1}{x}\right]_{(y/x,1/x)} \subset k(x,y)$$

which is the localization of the polynomial ring $k[y/x, 1/x]$ at the maximal ideal $(y/x, 1/x)$. If you do not know what this remark means, you can safely ignore it.

1.12.3 Projective Jacobi Criterion

One useful result that we would like to have in our toolkit is a projective analog of Theorem 1.8.8. For this, we will need

Lemma 1.12.9 (Euler). Suppose $F \in k[X,Y,Z]$ is a homogeneous polynomial of degree $d \geq 0$. If $\partial_X F$ (resp. $\partial_Y F$, $\partial_Z F$) denotes the formal partial derivative of F with respect to X (resp. Y, Z), then

$$X \cdot \partial_X F + Y \cdot \partial_Y F + Z \cdot \partial_Z F = d \cdot F.$$

Proof. Both sides of the equation are k-linear in F, so it suffices to show the result for a monomial of the form $F = X^a Y^b Z^c$, where $a + b + c = d$; but then the statement is clear. ∎

Of course, there is nothing special about the polynomial ring in three variables, and a similar result holds in any number of variables. Lemma 1.12.9 tells us also that if $\operatorname{ch} k \nmid d$ (in particular always in characteristic zero), then the conditions $\partial_X F|_P = \partial_Y F|_P = \partial_Z F|_P = 0$ also imply $F|_P = 0$. We are now ready to prove

Theorem 1.12.10 (Projective Jacobi Criterion). Suppose we are given a curve $C \subset \mathbb{P}^2_k$ and a point $P \in \mathbb{P}^2_k$. Let $F \in k[X,Y,Z]$ be a minimal polynomial for C. Then

(a) $P \in C$ iff $F|_P = 0$, and in this case
(b) P is a singular point of C iff

$$\partial_X F|_P = \partial_Y F|_P = \partial_Z F|_P = 0.$$

(c) If $P \in C$ is a smooth point, then the tangent line $\mathrm{T}_P C$ is defined by the vanishing of

$$\partial_X F|_P \cdot X + \partial_Y F|_P \cdot Y + \partial_Z F|_P \cdot Z = 0,$$

where in these evaluations we use the same representative (X_0, Y_0, Z_0) for the point $P = [X_0 : Y_0 : Z_0]$.

Proof. The statement in (a) is clear. As in the proof of Theorem 1.8.8, all parts are invariant under projective coordinate changes,[31] so it suffices to do the case $P = [0:0:1]$, and so we may work in the affine chart $\mathbb{A}^2_k$. For (b), we note that P is a singular point for C iff it is a singular point for C°, which by Theorem 1.11.21 has minimal polynomial F^{i}. Theorem 1.8.8 tells us that this happens iff

$$\partial_x F^{\mathrm{i}}|_P = \partial_y F^{\mathrm{i}}|_P = 0.$$

But now we observe that

$$\begin{aligned} \partial_x F^{\mathrm{i}} &= (\partial_X F)^{\mathrm{i}}, \\ \partial_y F^{\mathrm{i}} &= (\partial_Y F)^{\mathrm{i}}, \text{ and} \\ \partial_Z F|_P &= d \cdot F|_P, \end{aligned}$$

where in the last equality we are using Lemma 1.12.9. It follows that if (a) holds, then

$$\partial_x F^{\mathrm{i}}|_P = \partial_y F^{\mathrm{i}}|_P = 0 \Leftrightarrow \partial_X F|_P = \partial_Y F|_P = \partial_Z F|_P = 0,$$

[31] Check! This is the reason for the symmetric shape of the statement, although we will break the symmetry by invoking the affine Jacobi criterion below.

proving (b). Theorem 1.8.8 also tells us that the affine tangent line to C° at P is

$$\partial_x F^{\mathrm{i}}|_{(0,0)} \cdot x + \partial_y F^{\mathrm{i}}|_{(0,0} \cdot y = 0$$

which has projective closure

$$\partial_X F|_P \cdot X + \partial_Y F|_P \cdot Y + \partial_Z F|_P \cdot Z = 0$$

as needed. ■

One immediate consequence of this criterion is an analog of Theorem 1.9.7; this is

Theorem 1.12.11. If $C \subset \mathbb{P}^2_k$ is any curve, then C has only finitely many singular points.

Proof. Identical to the proof of Theorem 1.9.7, using Theorem 1.12.10 instead of Theorem 1.11.21. We can also reduce to the affine case. I leave the details to the reader. ■

That's more than enough abstract theory for now. Let's return to some concrete examples now.

1.12.4 Bézout's Theorem for a Line, Classification of Projective Conics up to Changes of Coordinates

Let's first prove Bézout's theorem for a line.

Theorem 1.12.12. If k is an algebraically closed field, $C \subset \mathbb{P}^2_k$ is a curve of degree $d \geq 1$, and $L \subset \mathbb{P}^2_k$ is a line such that $L \not\subset C$, then

$$\sum_{P \in C \cap L} i_P(C, L) = d.$$

Of course, $\deg L = 1$, so that $d = (\deg C)(\deg L)$.

Proof. By a projective change of coordinates, we can assume $L = L_\infty$. By Theorem 1.11.21, if F is a minimal polynomial for C, then C° has minimal polynomial $f := F^{\mathrm{i}}$, and $L_\infty \not\subset C$ implies that $Z \nmid F$ and so $\deg f = \deg F = \deg C = d$. If we write $f = f_0 + \cdots + f_d$, where each $f_j \in k[x, y]$ is homogeneous of degree j, then

$$F = f^{\mathrm{h}} = Z^d f_0(X, Y) + \cdots + f_d(X, Y).$$

Then points $P \in C \cap L$ are exactly points of the form $[X_0 : Y_0 : 0]$, where $f_d(X_0, Y_0) = 0$, and there are exactly d such points counted with multiplicity, by Lemma 1.8.3, where the two notions of multiplicity coincide by the computation in Example 1.9.12; I leave the details of this verification to the reader, since we will do the more general case soon. ■

We can now use this to classify all projective conics–at least when the base field k is algebraically closed.

Theorem 1.12.13. If k is algebraically closed and $Q \in k[X,Y,Z]_2$ is a nonzero homogeneous polynomial of degree 2, then there is a projective change of coordinates $\Phi : \mathbb{P}^2_k \to \mathbb{P}^2_k$ and a lift to the homogeneous polynomial ring $\Phi^* : k[X,Y,Z] \to k[X,Y,Z]$ such that $\Phi^* Q$ is either X^2, XY or $YZ - X^2$.

It is also clear that three cases are disjoint, since the corresponding projective curves are not isomorphic. One immediate consequence of this algebraic result is

Corollary 1.12.14 (Classification of Projective Conics)**.** If k is an algebraically closed field and $C \subset \mathbb{P}^2_k$ a conic (i.e. curve of degree 2), then there is a projective change of coordinates $\Phi : \mathbb{P}^2_k \to \mathbb{P}^2_k$ taking C to one, and only one, of the following forms:

(a) $C = C_{XY}$, which is a union of two lines that is singular at $[0:0:1]$, and
(b) $C = C_{YZ-X^2}$, which is a smooth conic.

In particular, it follows that any irreducible conic is smooth; compare this with the proof of this result from Exercise 2.3.4(b).

Proof of Theorem 1.12.13. Either $Q = \ell^2$ for some $L \in k[X,Y,Z]_1$, in which case we can take ℓ to X via some Φ^*; or $Q = \ell_1 \ell_2$ for some distinct irreducibles $\ell_1, \ell_2 \in k[X,Y,Z]_1$, in which case we can take $\ell_1 \mapsto X$ and $\ell_2 \mapsto Y$ by a simple application of Proposition 1.12.5; or Q is irreducible. Consider the curve C defined by Q; then C is also irreducible. By Proposition 1.11.13, C is infinite, but by Theorem 1.12.11, C has only finitely many singular points; in particular, all but finitely many points on C are smooth.

Let $P_1, P_2 \in C$ be any two distinct smooth points, and let $L_i = \mathrm{T}_{P_i} C$ for $i = 1, 2$ be the tangent lines at those points. We claim that $P_1 \notin L_2$ (and so, by symmetry, we have $P_2 \notin L_1$), and in particular $L_1 \neq L_2$. Indeed, if $P_1 \in L_2$, then we get that

$$\sum_{P \in C \cap L_2} i_P(C, L_2) \geq i_{P_1}(C, L_2) + i_{P_2}(C, L_2) \geq 1 + 2 = 3,$$

where $i_{P_2}(C, L_2) \geq 2$ because L_2 is tangent to C at P_2 (check!). This, combined with Theorem 1.12.12 tells us that $L_2 \subset C$, which by 1.11.17(b) implies that $L_2 = C$, contradicting the fact that $\deg C = 2$.[32] Since $L_1 \neq L_2$, we conclude from Proposition 1.11.3 that L_1 and L_2 intersect in a unique point, say P_3. Since $P_1 \notin L_2$, it follows that $P_1 \neq P_3$; similarly, $P_2 \neq P_3$. In fact, it follows that P_1, P_2, P_3 are not collinear: if they were collinear, then Proposition 1.11.3 would tell us that the line containing them would have to be both L_1 and L_2, contradicting $L_1 \neq L_2$.

It then follows from 1.12.5 that there is a projective change of coordinates Φ taking $P_1 \mapsto [0:0:1]$, $P_2 \mapsto [0:1:0]$ and $P_3 \mapsto [1:0:0]$. In this coordinate system, L_1 is the line $Y = 0$, and L_2 is the line $Z = 0$. For this Φ and any choice of Φ^*, if we write

$$\Phi^* Q = aX^2 + bXY + cY^2 + dXZ + eYZ + fZ^2,$$

then $f = 0$ (because $P_1 \in C$), $c = 0$ (because $P_2 \in C$), $d = 0$ (because $L_1 = C_Y$) and $b = 0$ (because $L_2 = C_Z$). In particular, we will have

$$\Phi^* Q = aX^2 + eYZ.$$

Now neither a nor e is zero, because otherwise Q would be reducible. Then we may scale Φ^* by $-a^{-1}$ and further change coordinates so Y is replaced by $-ae^{-1}Y$ to bring Q into the form $YZ - X^2$. ∎

[32] The fact that $P_1 \notin L_2$ uses crucially that C is a conic–for instance, a tangent to a cubic or higher degree curve meets the curve in at least one other point in general.

Remark 1.12.15. A careful analysis of the proof shows that we did not really use, in the last case, that k is algebraically closed, but only that C has at least two points. The above proof can be upgraded, with some care, to also obtain a classification over other fields: namely that if k is *any* field and $Q \in k[X, Y, Z]_2$ is a homogeneous irreducible element of degree 2 such that C_Q has at least two points, then, after a suitable change of coordinates, $Q = YZ - X^2$. This is, in fact, the best we can do in general: if $k = \mathbb{R}$, then the possibilities for Q include, in addition to $X^2, XY, YZ - X^2$, also the "conics" $X^2 + Y^2$ (which defines one point) and $X^2 + Y^2 + Z^2$ (which defines the empty set). The classification of projective conics over an arbitrary field is closely related to the classification of binary quadratic forms in 3 variables over that field. See, for instance, [5, §1.6] for another perspective on this result via this approach.

Next time, we will start by discussing very cool applications of these results–including Pascal's Theorem!

1.13 07/08/24 - Parametric Projective Curves, Pascal's Theorem, and More on Conics

Today, I want to prove Bézout's theorem for conics, and derive some delicious applications. For this, I will need to talk about parametric projective curves.

1.13.1 Parametric Projective Curves and Bézout's Theorem for a Conic

In the affine case, we defined a parametric curve to be an image of $\mathbb{A}^1_k$ under two rational functions. In the projective case, we can always clear denominators and work with $\mathbb{P}^1_k$ instead. This leads us to

Definition 1.13.1. A parametric projective algebraic curve is the image of a map $\Psi : \mathbb{P}^1_k \to \mathbb{P}^2_k$ of the form

$$\Psi[U : V] = [F_1(U, V) : F_2(U, V) : F_3(U, V)],$$

where $F_i(U, V) \in k[U, V]$ for $i = 1, 2, 3$ are homogeneous polynomials of the same degree, not all zero.

This definition corresponds to the affine one by considering $\mathbb{A}^1_k \subset \mathbb{P}^2_k$ as the set where $V \neq 0$ with coordinate $t = U/V$, in which case the affine part of this parametric projective curve is given by

$$t \mapsto \left(\frac{F_1(t, 1)}{F_3(t, 1)}, \frac{F_2(t, 1)}{F_3(t, 1)}\right),$$

which is a parametric affine curve. One can show, either using techniques similar to those in §1.3 or by reducing to the affine case, that a parametric projective algebraic curve is, in fact, a projective algebraic curve (at least when not all F_i are proportional, in which case the image is a single point). I will not do this here, but I encourage you to carry this out yourselves.

Remark 1.13.2. Note that I did not ask for the $F_j(U, V)$ to not have a common root on $\mathbb{P}^1_k$, because if they did, then I would very easily be able to just cancel this common factor from each $F_j(U, V)$. This is a manifestation of the completeness of projective curves–projective curves have no holes, and rational maps out a smooth projective curve always eextends to a regular morphism out of it. As usual, if this doesn't make sense, please ignore it.

Example 1.13.3. The smooth conic defined by $YZ - X^2 \in k[X, Y, Z]_2$ can be parametrized via the map Ψ given as

$$\Psi[U : V] = [UV : U^2 : V^2].$$

This is the projective version of affine parametrization $t \mapsto (t, t^2)$ of the parabola defined by $y - x^2 = 0$. By Corollary 1.12.14, this gives us a parametrization of every smooth conic curve. In particular, every smooth conic curve admits a parametrization.

From this parametrization, we can now prove Bézout's theorem for a conic.

Theorem 1.13.4. If k is an algebraically closed field, $C \subset \mathbb{P}^2_k$ is a conic (i.e. a curve with $\deg C = 2$), and $D \subset \mathbb{P}^2_k$ a curve of degree $d \geq 1$ such that C and D do not share a component, then

$$\sum_{P \in C \cap L} i_P(C, D) = 2d.$$

Of course, $\deg C = 2$ and so $2d = (\deg C)(\deg D)$.

Proof. By Corollary 1.12.14, we can choose coordinates such that C is either the union of the two lines C_X and C_Y, or that $C = C_{YZ-X^2}$; make a change of coordinates so that we are working with this coordinate system. In the first case, neither of the two lines C_X or C_Y can be contained in D, and we are done by additivity of intersection multiplicity and Theorem 1.12.12 (make sure you believe this!). In the second case, C is irreducible and Example 1.13.3 tells us that we can parametrize C as the image of the map

$$[U : V] \mapsto [UV : U^2 : V^2].$$

If F is a minimal polynomial for D, then F is a homogeneous polynomial of degree d, and the intersection points of C and D correspond exactly to $[U_0 : V_0] \in \mathbb{P}^1_k$ such that

$$F(U_0V_0, U_0^2, V_0^2) = 0.$$

Now $F(UV, U^2, V^2) \in k[U, V]_{2d}$ is a homogeneous polynomial of degree $2d$. If it is identically zero, then we conclude that $C \subset D$, contrary to assumption that C and D do not share any components; therefore, this polynomial is not identically zero, and so has exactly $2d$ roots counted with multiplicity, again by Lemma 1.8.3. I will again leave it to the reader to check, perhaps using techniques similar to those from Example 1.9.12, that the intersection multiplicity of C and D at a point $[U_0V_0 : U_0^2 : V_0^2]$ agrees with the multiplicity of $[U_0 : V_0]$ as a root of $F(U_0V_0, U_0^2, V_0^2) = 0$.[33] ∎

We are now ready for some delicious applications!

1.13.2 Pascal's Theorem, Pappus's Theorem, Brocard's Theorem, etc.

Theorem 1.13.5 (Pascal). Let k be an algebraically closed field, $C \subset \mathbb{P}^2_k$ a smooth conic and $P_1, \ldots, P_6$ distinct points on C. For $i = 1, \ldots, 6$, let L_i be the line joining P_i and P_{i+1} (where $P_7 := P_1$), and for $j = 1, 2, 3$, let $Q_j := L_j \cap L_{j+3}$. Then the points $Q_1, Q_2, Q_3 \in \mathbb{P}^2_k$ are collinear, i.e. there is a line $L_0 \subset \mathbb{P}^2_k$ such that $Q_j \in L_0$ for $j = 1, 2, 3$.

Let's first make a few observations about the statement:

(a) The lines L_i are all distinct: if $L_i = L_{i'}$ for some $i \neq i'$, then this line intersects C in at least 3 distinct points, and is hence contained in C (by either Theorem 1.12.12 or 1.13.4); this would mean that C is reducible and hence (by Corollary 1.12.14 if needed) not smooth. In particular, by Proposition 1.11.3, the points Q_1, Q_2, Q_3 are uniquely determined.

(b) Each P_i lies on exactly two lines $L_{i'}$, namely L_{i-1} and L_i, and, in particular, these lines have indices that differ by 1 (modulo 6); conversely, each L_i contains exactly two points P_i and P_{i+1} of C, because again if it contained a third point of C, it would be contained in C entirely.

(c) We have $P_i \neq Q_j$ for all i, j. Indeed, let us take the indices i, j modulo 6; then $P_i = Q_j$ cannot happen because this implies that $P_i \in L_j \cap L_{j+3}$, violating the observation (b).

(d) Finally, we have $Q_1, Q_2, Q_3 \notin C$. Indeed, if some $Q_j \in C$, then $Q_j \in L_j \cap C = \{P_j, P_{j+1}\}$ implies that $Q_j = P_i$ for some i, j, violating (c).

(e) In fact, although we will not need this for the proof, all the 9 points P_i, Q_j are distinct:

Let's now proceed to the proof, which is rather simple given the tools we have.

[33] I'm being lazy partly because, in the proof of Pascal's Theorem (Theorem 1.13.5) below, we will only need the result that C and D intersect in at most $2d$ points unless they share a component, and this we have already proven. Also, we shall do a full proof of the general Bézout Theorem very soon.

Proof. For $i = 1, \dots, 6$, let $\ell_i \in k[X, Y, Z]_1$ be a homogeneous linear polynomial vanishing on L_i. Consider the family D_Λ of cubic curves parametrized by $\Lambda = [\lambda : \mu] \in \mathbb{P}^1_k$, where D_Λ is defined by the vanishing of the polynomial

$$\lambda \ell_1 \ell_3 \ell_5 + \mu \ell_2 \ell_4 \ell_6.$$

Note that each curve D_Λ in this family passes through all the P_i's and Q_j's, and we have that $D_{[1:0]} = L_1 \cup L_3 \cup L_5$ and $D_{[0:1]} = L_2 \cup L_4 \cup L_6$.[34] Now pick a point $R \in C \smallsetminus \{P_1, \dots, P_6\}$, which exists because C is infinite (Proposition 1.11.13). From observation (b) above, we conclude that $R \notin D_{[1:0]} \cup D_{[0:1]}$, from which it follows that there is a unique $\Lambda_0 \in \mathbb{P}^1_k$ such that $R \in D_{\Lambda_0}$.[35] Let $D = D_{\Lambda_0}$. Since D is a cubic curve and C and D intersect in at least 7 points, it follows from Theorem 1.13.4 that C and D share a component. Since C is irreducible (Corollary 1.12.14) and $\deg D = 3$, this can only happen if $C \subset D$ and $D = C \cup L_0$ for some line $L_0 \subset \mathbb{P}^2_k$. But now, D contains Q_1, Q_2, Q_3 (because each D_Λ does), while C does not contain Q_1, Q_2, Q_3 (this was observation (c) above), and hence $Q_1, Q_2, Q_3 \in L_0$. ∎

See Figure 1.8 for a visual demonstration of the proof technique.

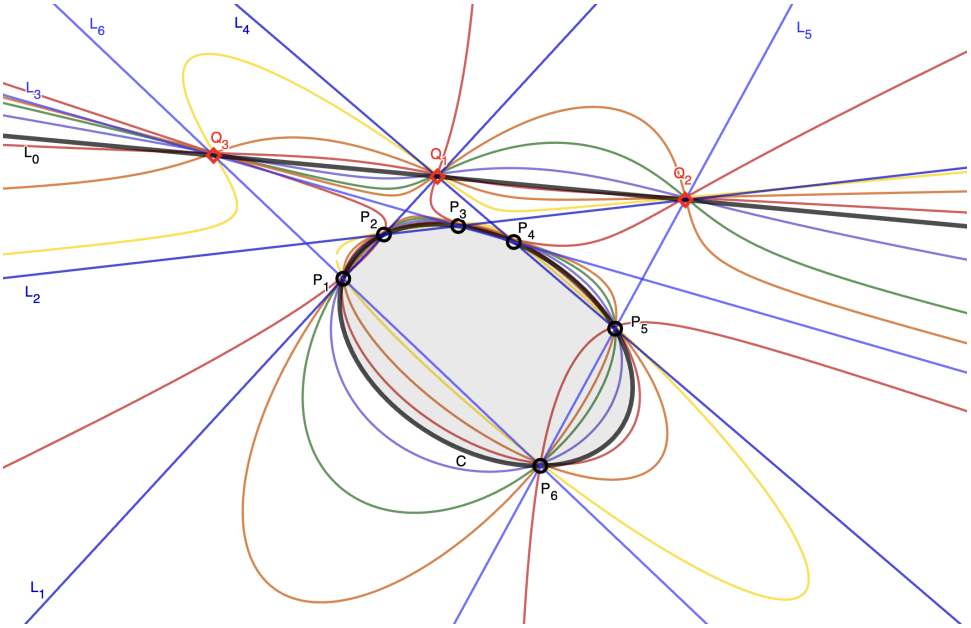


Figure 1.8: Pascal's Theorem. The conic (here ellipse) C and the line L_0 are in thick black style. The various colorful curves represent various members of the one parameter family D_Λ, one member of which is also $C \cup L_0$. Picture made with Geogebra.

Remark 1.13.6. Note that the actual statement of Theorem 1.13.5 does not use an ordering whatsoever on the points $P_1, \dots, P_6$–indeed, for general fields, it does not even make sense to order points of a conic. In particular, if we start with a collection of 6 distinct unordered points on a conic C, then they can be connected into a hexagon in 60 different ways, and resulting in 60 different instances of Pascal's Theorem and 60 different "Pascal" lines; this configuration of 60 lines associated to 6 points on a hexagon is often called the Hexagrammum Mysticum. Finally, although we have proven the theorem over algebraically closed fields, it follows also immediately

[34] If you were not convinced of this already, then this observation tells us that every curve D_Λ has degree 3: indeed, if it did not, then some D_Λ would be either a line or a union of two lines, neither of which can contain all the P_i's, since no three of them are collinear (why?).

[35] Indeed, if we pick a representative (X_0, Y_0, Z_0) for $R = [X_0 : Y_0 : Z_0]$, then neither of $\ell_1 \ell_3 \ell_5|_{(X_0,Y_0,Z_0)}$ and $\ell_2 \ell_4 \ell_6|_{(X_0,Y_0,Z_0)}$ is zero, and this unique Λ_0 is $\Lambda_0 = [-\ell_2 \ell_4 \ell_6|_{(X_0,Y_0,Z_0)} : \ell_1 \ell_3 \ell_5|_{(X_0,Y_0,Z_0)}] \in \mathbb{P}^1_k$.

over all fields (e.g. over $k = \mathbb{R}$), thanks to Theorem 1.4.5 and the observation that Proposition 1.11.3 does not use that the base field is algebraically closed, which implies, for instance, that if three points $Q_1, Q_2, Q_3 \in \mathbb{P}^2_{\mathbb{R}} \subset \mathbb{P}^2_{\mathbb{C}}$ are collinear in $\mathbb{P}^2_{\mathbb{C}}$, then they are collinear in $\mathbb{P}^2_{\mathbb{R}}$, i.e. the line joining them is real. Over the real numbers, other proofs can also be given; after all, Pascal did not actually have Bézout's Theorem. One approach involves using a variant of the classification of projetive conics over $\mathbb{R}$ (see Remark 1.12.15) to conclude that any smooth conic can be taken by a projective change of coordinates over $\mathbb{R}$ to a circle $X^2 + Y^2 - Z^2 = 0$, and then to use other techniques from Euclidean geometry (e.g. Menelaus's Theorem).

In the proof of Pascal's Theorem, we did not really use that C was a smooth conic other than to rule out certain degenerate cases. Therefore, the same proof technique also yields

Theorem 1.13.7 (Pappus)**.** Let k be any field. Let $L_1, L_2 \subset \mathbb{P}^2_k$ two distinct lines, and $P_1, Q_1, R_1 \in L_1 \smallsetminus L_2$ and $P_2, Q_2, R_2 \in L_2 \smallsetminus L_1$ be distinct points. If

$$\begin{aligned} S_1 &= \overline{Q_1R_2} \cap \overline{Q_2R_1}, \\ S_2 &= \overline{P_1R_2} \cap \overline{{}_2R_1}, \text{ and} \\ S_3 &= \overline{P_1Q_2} \cap \overline{P_2Q_1}. \end{aligned}$$

. Then $S_1, S_2, S_3 \subset \mathbb{P}^2_k$ are collinear.

Proof. By Theorem 1.4.5 and Proposition 1.11.3, we may replace k by an algebraically closed field and still have the same result (check!), and then the same proof technique as in Theorem 1.13.5 works. I leave the verification of the nondegeneracy conditions to the diligent reader. ■

Finally, Pascal's Theorem can also be applied with "multiplicities". The key result needed to do this is

Lemma 1.13.8. Let $C \subset \mathbb{P}^2_k$ be a curve and $P \in C$ be a smooth point. Let F be a minimal polynomial for C, and let $G, H \in k[X, Y, Z]$ be homogeneous polynomials such that $G, H, G + H \neq 0$. Then

$$i_P(F, G + H) \geq \max\{i_P(F, G), i_P(F, H)\}$$

with equality if $i_P(F, G) \neq i_P(F, H)$.

Proof Sketch. This is a local property invariant under changes of coordinates, and so we may work in the affine chart $Z \neq 0$ and assume that $P = (0, 0)$ and that the tangent line to C at P is the x-axis C_y. Let $f = F^{\mathrm{i}}$. The claim is that for any $0 \neq g \in \mathcal{O}_P$, there is a unique integer $n \geq 0$ such that for some unit $u \in \mathcal{O}_P^\times$ we have $g - ux^n \in (f)\mathcal{O}_P$. Uniqueness is clear, because then $i_P(f, g) = i_P(f, ux^n) = n$. For existence, scale f and write it as $f = y + x^n p + y^2 q$ for some $p \in k[x]$ such that $p(0) \neq 0$ and then $y - p(1 + yq)^{-1}x^n \in (f)\mathcal{O}_P$, proving the claim for $g = y$. The statement for $g = x$ is clear, and so is the fact that if such an n exists for g and h, then it does also for $g \cdot h$. Showing the result for the sum $g + h$ when $0 \neq g + h$ is slightly more involved, but in any case if $g \equiv ux^n \pmod{f\mathcal{O}_P}$ and $h \equiv vx^m \pmod{f\mathcal{O}_P}$ for some $n, m \geq 0$ and $u, v \in \mathcal{O}_P^\times$, then $f + g \equiv (ux^{n-m} + v)x^m \pmod{f\mathcal{O}_P}$ showing that the result holds for $f + g$ (as well as the claim in the lemma), unless we have $n = m$ and $u + v = 0$; this case needs some more effort, but is not too difficult. See, for instance, the discussion in the proof of [3, §3.3, Theorem 3(8)]. ■

Remark 1.13.9. The grown-up way to prove (and understand) Lemma 1.13.8 is to say that if $C \subset \mathbb{P}^2_k$ is a curve and $P \in C$ is a smooth point, then the local ring $\mathcal{O}_{C,P}$ of C at P is a discrete valuation ring with uniformizer given by the class of any line not tangent to C at P. I haven't defined what those terms are yet, so do not worry too much about this at the moment.

Given Lemma 1.13.8, however, it is very straightforward to extend the proof of Pascal's Theorem to cases where the points "degenerate". Here's one example of how to do this; you are invited to explore other examples of this sort in Exercise 2.5.3.

Theorem 1.13.10 (Brocard)**.** Let k be an algebraically closed field and $C \subset \mathbb{P}^2_k$ be a smooth conic and $P_1, P_2, P_3, P_4 \in C$ be distinct points. For $i = 1, \ldots, 4$, let $T_i := \mathrm{T}_{P_i} C$, and for $1 \leq i, j \leq 4$, let L_{ij} be the line joining P_i and P_j. Let

$$\begin{aligned} S_1 &= L_{12} \cap L_{34}, \\ S_2 &= L_{23} \cap L_{41}, \\ Q_{13} &= T_1 \cap T_3, \\ Q_{24} &= T_2 \cap T_4. \end{aligned}$$

The points S_1, S_2, Q_{13} and Q_{24} in $\mathbb{P}^2_k$ are collinear.

I will leave to the reader the verification of many implicit claims in the statement of the theorem, e.g. the definition of S_1 uses Proposition 1.11.3 and that $L_{12} \neq L_{34}$. The line joining S_1, S_2, Q_{13} and Q_{24} is called the polar of the last intersection point $S_3 := L_{13} \cap L_{41}$ with respect to the conic C. Again, the ordering of the points P_1, P_2, P_3, P_4 does not matter, and we end up with 3 different such configurations.

Proof. It suffices to show that S_1, S_2 and Q_{13} are collinear, because then S_2, S_1 and Q_{24} are collinear by an application of the proven claim to P_2, P_3, P_4, P_1 in that order. To show the first claim, apply Pascal's Theorem (Theorem 1.13.5) to the "hexagon" $P_1 P_1 P_2 P_3 P_3 P_4$. To say more, take

$$\begin{aligned} L_1 &= T_1, \\ L_2 &= L_{12}, \\ L_3 &= L_{23}, \\ L_4 &= T_3, \\ L_5 &= L_{34}, \text{ and} \\ L_6 &= L_{41} \end{aligned}$$

in the setup of Theorem 1.8, so that $Q_1 = Q_{13}$, $Q_2 = S_1$ and $Q_3 = S_2$. Take linear polynomials as ℓ_i as before, and again consider the 1-parameter family D_Λ. Again, take a new point R and a unique Λ_0 such that $R \in D$. Since $L_1 \cup L_3 \cup L_5$ and $L_2 \cup L_4 \cup L_6$ each meet C in multiplicity at least 2 at both P_1 and P_3, it follows from Lemma 1.13.8 then every memebr of the family D_Λ meets C both passes through the points $P_1, P_2, P_3, P_4, S_1, S_2$ and Q_{13}, and meets C to multiplicity at least two at both P_1 and P_3. It follows as before from Bézout's Theorem for a conic (Theorem 1.13.4), but this time applied with multiplicities, that $C \subset D$, and the rest of the proof is identical to that of Theorem 1.13.5. ∎

See Figure 1.9 for an illustration of Theorems 1.13.7 and 1.13.10.

Remark 1.13.11. Over $k = \mathbb{R}$ or $k = \mathbb{C}$, the proof of these "degenerate" cases can also be given by continuity. Similarly, once you have Pascal's Theorem, you can also derive from it Pappus's

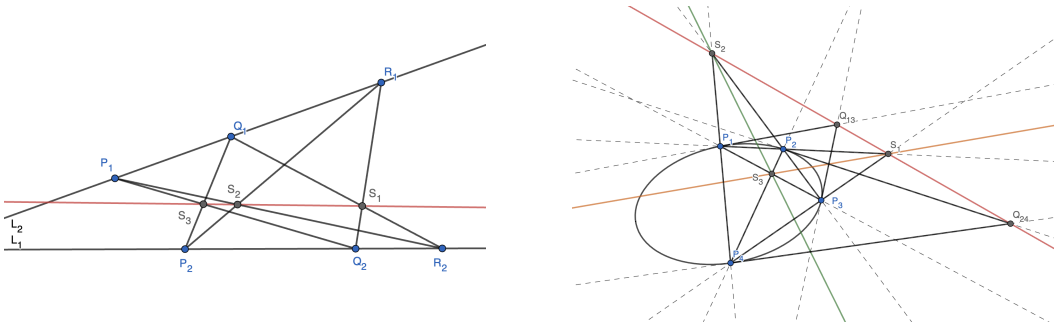


Figure 1.9: Theorems of Pappus and Brocard. Pictures made with Geogebra.

Theorem by continuity (by letting a hyperbola degenerate to a pair of lines). Such proofs are also available over other fields, but only with significantly more sophisticated tools.

1.13.3 More on Conics

Finally, let's talk about how Bézout's Theorem can be used to solve interpolation problems, i.e. problems involving finding curves of certain degrees passing through given points in $\mathbb{P}^2_k$.

Theorem 1.13.12. Let $S \subset \mathbb{P}^2_k$ be a set with 5 elements.

(a) There is a conic $C \subset \mathbb{P}^2_k$ passing through S, i.e. such that $S \subset C$.
(b) If no four distinct points in S are collinear, then such a conic as in (a) is unique.
(c) If no three distinct points in S are collinear (i.e. S is in general position), then the unique conic as in (b) is smooth.

Note that (b) and (c) are the best possible refinements of (a): if four points in S were collinear, then (at least if k is infinite), there would be infinitely many (reducible) conics containing S, and similarly if three points in S are collinear, then there is no hope of a conic containing S being irreducible or equivalently smooth (thanks again to Theorem 1.13.4).

Proof.

(a) Let $S = \{P_1, \ldots, P_5\}$, and pick representatives (X_i, Y_i, Z_i) for $P_i = [X_i : Y_i : Z_i]$ for $i = 1, 2, \ldots, 5$. The vector space of homogeneous quadratic polynomials in
(b) If there are two distinct conics $C, D \subset \mathbb{P}^2_k$ through S, then by Bézout's Theorem (Theorem 1.13.4), C and D must have a common component. Then neither C nor D can be irreducible, and, in fact, we must have $C = L_1 \cup L_2$ and $D = L_2 \cup L_3$ for some distinct lines $L_1, L_2, L_3 \subset \mathbb{P}^2_k$ (check!). In this case, $S \subset C \cap D = L_2 \cup (L_1 \cap L_3)$. Since $L_1 \cap L_3$ is one point, at least four points of S must lie on L_2.
(c) If the unique conic C as in (b) is singular, then it is reducible and hence a union of two lines. By the Pigeonhole Principle, at least three elements of S must lie on a line.

■

Remark 1.13.13. You are invited to explore similar interpolation problems in Exercise 2.5.1. In the above result, there is some subtlety involving whether of not we're working over algebraically closed fields; I'll let you work through the details of that. Remark 1.12.15 may be of some help.

We will have just a little more to say about conics in the next few lectures–when we talk about one-parameter families (i.e. pencils) of conics. Next time, we will finally go over two proofs of Bézout's Theorem.

1.14 07/10/24 - Proof(s) of Bézout's Theorem

We are now finally ready to prove Bézout's Theorem, which we state here.

Theorem 1.14.1 (Bézout). If k is an algebraically closed field, and $C, D \subset \mathbb{P}^2_k$ algebraic curves that do not share a common component, then

$$\sum_{P \in C \cap D} i_P(C, D) = (\deg C)(\deg D).$$

We showed in Theorem 1.11.20 that if C and D do not share a component, then C and D intersect in finitely many points. We will give two proofs of Theorem 1.14.1 below. The proof strategy in both case is going to be to choose a suitable coordinate system in which C and D do not intersect at infinity–that it all what we will need the projective plane for. Having done that, the rest of the proof becomes a computation in the affine plane.

1.14.1 Proof 1: Dimension Count

Proof 1 of Theorem 1.14.1. Pick a line L not meeting $C \cap D$ (this is possible by Theorem 1.11.20 and the correct salvage to Exercise 2.6.7), and choose a system of coordinates such that (i.e. assume by a projective change of coordinates that) $L = L_\infty$. Then neither C nor D contains L as a component–indeed, if, say, $L \subset C$, then it would follow from Theorem 1.12.12 that $L \cap D$ is nonempty, and then $L \cap C \cap D$ is nonempty, contrary to assumption. In particular, if F (resp. G) is a minimal polynomial for C (resp. D), and we let $f := F^{\mathrm{i}}$ (resp. $g := G^{\mathrm{i}}$) and $\deg C = n \geq 1$ (resp. $\deg D = m \geq 1$), then we have by Theorem 1.11.21 that

$$\deg f = \deg F = \deg C = m \text{ and } \deg g = \deg G = \deg D = n.$$

If we write $f = f_0 + \cdots + f_m$ and $g = g_0 + \cdots + g_n$, where each f_i and g_i is homogeneous of degree i in x and y, then $f_m g_n \neq 0$, and it follows from the assumption that $L \cap C \cap D = \emptyset$ that $f_m, g_n \in k[x, y]$ are relatively prime (for instance, thanks to Lemma 1.8.3). Finally, the fact that C and D do not share a common component implies that f and g are relatively prime. We now divide the rest of the proof into two lemmas, whose proofs we postpone for a moment.

Lemma 1.14.2. If k is an algebraically closed field and $f, g \in k[x, y]$ are relatively prime, then the following map is an isomorphism:

$$k[x, y]/(f, g) \xrightarrow{\sim} \prod_{P \in C_f \cap C_g} \mathcal{O}_P/(f, g)\mathcal{O}_P.$$

Lemma 1.14.3. If k is a field and $f, g \in k[x, y]$ have degree $m, n \geq 1$ such that f and g are relatively prime and the leading terms f_m and g_n are relatively prime, then

$$\dim_k k[x, y]/(f, g) = mn.$$

By our definition of intersection multiplicity (as in the existence part of the proof of Theorem 1.9.9), the two lemmas above combined prove Theorem 1.14.1. ∎

The first lemma is a local-to-global principle (often called Max Noether's $af + bg$ theorem), and is a sort of Chinese Remainder Theorem for curves, if you will. The second result is the global dimension computation that proves the result. Let's now prove the lemmas.

Lemma 1.14.2. If k is an algebraically closed field and $f, g \in k[x, y]$ are relatively prime, then the following map is an isomorphism:

$$k[x, y]/(f, g) \xrightarrow{\sim} \prod_{P \in C_f \cap C_g} \mathcal{O}_P/(f, g)\mathcal{O}_P.$$

Proof. To show surjectivity, note that we showed in the proof of existence in Theorem 1.9.9 that if $f, g \in k[x, y]$ are relatively prime and if $P = (p, q) \in C_f \cap C_g$, then there is an $N \geq 1$ such that $(x - p)^N, (y - q)^N \in (f, g)\mathcal{O}_P$. Since, by Theorem 1.6.6, the intersection $C_f \cap C_g$ is finite, there is an $N \geq 1$ that works for all $P \in C_f \cap C_g$. In other words, there is an $N \geq 1$ such that if we enumerate $C_f \cap C_g = \{P_i\}$ with $P_i = (p_i, q_i)$, then $(x - p_i)^N, (y - q_i)^N \in (f, g)\mathcal{O}_{P_i}$ for all i. Now, to show injectivity, it suffices to show that for each i, there is a polynomial $f_i \in k[x, y]$ such that f_i maps to 0 in $\mathcal{O}_{P_j}/(f, g)\mathcal{O}_{P_j}$ for all $j \neq i$, but to a unit in $\mathcal{O}_{P_i}/(f, g)\mathcal{O}_{P_i}$; for this, simply take

$$f_i := \prod_{j : p_j \neq p_i} (x - p_j)^N \prod_{j : q_j \neq q_i} (y - q_j)^N,$$

which maps to zero in each $\mathcal{O}_{P_j}/(f, g)\mathcal{O}_{P_j}$ for $j \neq i$ because of our choice of N, while it is a unit already in $\mathcal{O}_{P_i}$ and hence also in $\mathcal{O}_{P_i}/(f, g)\mathcal{O}_{P_i}$.[36]

To show injectivity, we have to show that if $h \in k[x, y]$ is such that $h \in (f, g)\mathcal{O}_P$ for all $P \in C_f \cap C_g$, then $h \in (f, g)k[x, y]$. For that, given an h, consider the ideal

$$I := \{q \in k[x, y] : qh \in (f, g)\} \subset k[x, y].$$

Then $I \supset (f, g)k[x, y]$, and we want to show that $1 \in I$, i.e. that $I = k[x, y]$.[37] If I is not a proper ideal, then by Proposition 1.7.6, there is a prime ideal $Q \subset k[x, y]$ containing I.[38] Since Q cannot be 0 or of the form (r) for some irreducible $r \in k[x, y]$ (because $f, g \in Q$ are nonzero and relatively prime), by Exercise 2.3.3, we must have $Q = (x - p, y - q)$ for some $p, q \in k$ (this uses that k is algebraically closed). Now $f, g \in Q = (x - p, y - q)$ implies that if $P = (p, q)$, then $P \in C_f \cap C_g$. Since, by hypothesis, we have $h \in (f, g)\mathcal{O}_P$, we conclude that there are $a, b, c \in k[x, y]$ such that $ch = af + bg$ with $c|_P \neq 0$. But this implies that $c \in I \smallsetminus Q$, which is a contradiction, finishing the proof. ∎

Lemma 1.14.3. If k is a field and $f, g \in k[x, y]$ have degree $m, n \geq 1$ such that f and g are relatively prime and the leading terms f_m and g_n are relatively prime, then

$$\dim_k k[x, y]/(f, g) = mn.$$

Proof. For each integer $d \geq 0$, let $k[x, y]_{\leq d}$ denote the k-vector subspace of $k[x, y]$ consisting of polynomials of degree at most d, which has dimension $\binom{d+2}{2}$ over k. The proof idea is to approximate $\dim_k k[x, y]/(f, g)$ by the images of the projections of $k[x, y]_d$ for $d \gg 1$. To do this, for any $d \geq m + n$, consider the sequence of k-vector spaces and k-linear maps given by

$$0 \to k[x, y]_{\leq d-m-n} \xrightarrow{\alpha} k[x, y]_{\leq d-m} \times k[x, y]_{\leq d-n} \xrightarrow{\beta} k[x, y]_{\leq d} \xrightarrow{\pi_d} k[x, y]/(f, g), \tag{1.2}$$

[36] The surjectivity result does not actually need k to be algebraically closed.

[37] The ideal I is often called the ideal quotient of (f, g) by (h) and is denoted $(f, g) : (h)$.

[38] In our case, we did not quite need a fact this general, since we already have $f, g \in I$ and so we may conclude from this that there are polynomials in x only and y only in I, but Proposition 1.7.6 (which is a good fact to know in general) simplifies things tremendously.

where

$$\begin{aligned}\alpha : c &\mapsto (cg, -cf),\\ \beta : (a,b) &\mapsto af + bg,\end{aligned}$$

and π_d is the restriction of the natural projection map $\pi : k[x,y] \to k[x,y]/(f,g)$ to the subspace $k[x,y]_{\leq d}] \subset k[x,y]$. In the sequence (1.2), the compositions of each pair of successive maps are all zero, i.e. $\beta \circ \alpha = 0$ and $\pi_d \circ \beta = 0$. The key claim is that, under our hypotheses, this sequence (1.2) is exact, i.e. α is injective, and we have $\operatorname{im}\alpha = \ker\beta$ and $\operatorname{im}\beta = \ker\pi_d$. Assuming this, we conclude from repeated applications of the Rank-Nullity Theorem that

$$\begin{aligned}\dim_k \operatorname{im}\pi_d &= \binom{d+2}{2} - \dim_k \ker\pi_d\\ &= \binom{d+2}{2} - \dim_k \operatorname{im}\beta\\ &= \binom{d+2}{2} - \binom{d-m+2}{2} - \binom{d-n+2}{2} + \dim_k \ker\beta\\ &= \binom{d+2}{2} - \binom{d-m+2}{2} - \binom{d-n+2}{2} + \dim_k \operatorname{im}\alpha\\ &= \binom{d+2}{2} - \binom{d-m+2}{2} - \binom{d-n+2}{2} + \binom{d-m-n+2}{2}\\ &= mn,\end{aligned}$$

where the last step is a trivial simplification. In particular, for all $d \geq m+n$, the dimension of $\operatorname{im}\pi_d$ is independent of d. Since the $\operatorname{im}\pi_d \subset k[x,y]/(f,g)$ for $d \geq 0$ form an increasing sequence of subspaces with union $\operatorname{im}\pi = k[x,y]/(f,g)$, it follows from this constancy of dimensions that

$$\operatorname{im}\pi_{m+n} = \operatorname{im}\pi_{m+n+1} = \operatorname{im}\pi_{m+n+2} = \cdots = \operatorname{im}\pi = k[x,y]/(f,g),$$

and hence

$$\dim k[x,y]/(f,g) = \dim \operatorname{im}\pi_{m+n} = mn.$$

It remains to show that under our hypothesis, the sequence (1.2) is exact, which we do now.

(a) The map α is visibly injective, since $k[x,y]$ is a domain and $f, g \neq 0$.

(b) Clearly, $\operatorname{im}\alpha \subset \ker\beta$. Conversely, if $(f,g) \in \ker\beta$, then $af + bg = 0$. Since f and g are relatively prime, it follows from this that $g \mid a$ and $f \mid b$, and in fact that there is a $c \in k[x,y]$ such that $a = cg$ and $b = -cf$. If $\deg a \leq d-m$ and $\deg b \leq d-n$, then we must also have $\deg c \leq d-m-n$. This proves that $\ker\beta \subset \operatorname{im}\alpha$.

(c) Again, clearly $\operatorname{im}\beta \subset \ker\pi_d$. Conversely, if $h \in \ker\pi_d$, then $h \in (f,g)$. Write $h = af + bg$ for some $a, b \in k[x,y]$ and suppose that this representation is chosen so that $\deg a$ is minimal (here we take $\deg 0 = 0$). We will show that $\deg a \leq d-m$ and $\deg b \leq d-n$, from which it follows that $h \in \operatorname{im}\beta$, finishing the proof. Suppose to the contrary that $p := \deg a > d-m$ or that $q := \deg b > d-n$, so that either af or bg contains a term of degree greater than d. Since $\deg h \leq d$ and $h = af + bg$, it follows that the leading terms of af and bg must cancel, i.e. $p + m = q + n$ and if we write $a = a_0 + \cdots + a_p$ and $b = b_0 + \cdots + b_q$, where each a_i, b_i is homogeneous of degree i with $a_p b_q \neq 0$, then

$$a_p f_m + b_q g_n = 0.$$

Now, since the terms f_m and g_n are relatively prime, it follows as before that there is some nonzero $c \in k[x,y]$ of degree $p - n = q - m$ such that $a_p = g c_n$ and $b_q = -c f_m$. Then

$$h = (a - cg)f + (b + cf)g$$

is another representation of h with $\deg(a - cg) < \deg a$, contrary to our choice of a.

■

1.14.2 Proof 2: Resultants

Sketch of Proof 2 of Theorem 1.14.1. Consider the finite set S consisting of all lines that join two or more points of $C \cap D$ and all tangent lines to C and D at all the points of intersection $C \cap D$. Pick a point $P_0 \in \mathbb{P}^2_k$ that is not on $C \cup D$ and not on any line in S. Pick a coordinate system so that $P_0 = [1 : 0 : 0]$. It follows from this choice that each "horizontal" line $Z_0Y - Y_0Z = 0$ meets at most one point of $C \cap D$, i.e. all the points of intersection have distinct y-coordinates. The idea of the proof is to project the intersection points $C \cap D$ onto the y-axis, and use this to count then number intersection points (with multiplicity).

For this, let $\deg C = m$ (resp. $\deg D = n$), and let F (resp. G) be a minimal polynomial for C (resp. D). Write

$$F = F_0X^m + \cdots + F_m \text{ and } G = G_0X^n + \cdots + G_n,$$

where each F_i (resp. G_i) is a polynomial only of Y and Z and homogeneous of degree i. The assumption that $P_0 \notin C \cup D$ implies that $F_0G_0 \neq 0$. Since F, G are relatively prime in $k[X, Y, Z]$, by Lemma 1.6.2(b) there are $A, B \in k[X, Y, Z]$ and $0 \neq R \in k[Y, Z]$ such that $AF + BG = R$. In fact, we can choose R to be the resultant

$$R = \mathrm{Res}_X(F, G) \in k[Y, Z]_{mn}$$

with A and B homogeneous as well.[39] Then a point $[Y_0 : Z_0]$ is a root of R iff the polynomials $F(X, Y_0, Z_0)$ and $G(X, Y_0, Z_0)$ have common root X_0 over k (Exercise 2.2.4(d)), which happens iff the horizontal line $Z_0Y - Y_0Z = 0$ intersects the curve. In other words, the roots of R correspond exactly to the projection of the intersection of F and G to the y-axis, since we chose our coordinate system so that no two points of intersection lie on the same horizontal line.

Since R has exactly mn roots counted with multiplicity, to complete the proof, it suffices to show that for each root $[Y_0 : Z_0]$ of R, the intersection multiplicity of C and D at the unique point of intersection on the line $Z_0Y - Y_0Z = 0$ is exactly the multiplicity of $[Y_0 : Z_0]$ as a root of R. There are many ways to do this. One way to show this is to prove that this definition satisfies (with respect to any choice of P_0) satisfies the axioms (1)-(7), and use the uniqueness result from Theorem 1.9.9; this is, for instance, the approach followed in [6, Theorem 3.18]. Another way to do this is to note that the problem is local at P, so by an affine translation (so preserving P_0), we may assume that $P = (0, 0)$ is the point of intersection on line $y = 0$. Since resultants are stable under dehomogenization, we conclude that if f and g are the dehomogenizations of F and G, then we have to show that $i_P(f, g)$ is the multiplicity $m_0(r)$ of $r = \mathrm{Res}_x(f, g)$ at 0, which is the highest power of y dividing r. Let this highest power be N. The claim then follows from the observation in the local ring $\mathcal{O}_P$, we have $(f, g)\mathcal{O}_P = (x+yq, y^N)\mathcal{O}_P$ for some $q \in k[x, y]$. The result follows from this from because then

$$i_P(f, g) = \dim_k \mathcal{O}_P/(f, g)\mathcal{O}_P = i_P(x + yq, y^N) = N \cdot i_P(x + yq, y) = N \cdot i_p(x, y) = N.$$

To show that $(f, g)\mathcal{O}_P = (x + yq, y^N)\mathcal{O}_P$, note first that $r \in (f, g)k[x, y]$ can be written as $y^N r_0$ for some $r_0 \in k[y]$ with $r_0(0) \neq 0$, whence $y^N \in (f, g)\mathcal{O}_P$. Also, we can write $f = xf_1 + yf_2$ and $g = xg_1 + yg_2$ for some polynomials $f_1, g_1 \in k[x]$ and $f_2, g_2 \in k[x, y]$. Then the assumption that P is the only intersection point of C and D on $y = 0$ implies that f_1 and g_1 are coprime, whence from Bézout's *Lemma* it follows that there are $a, b \in k[x]$ such that $af_1+bg_1 = 1$. It follows then that $af+bg = x+yq$ for $q = af_2+bg_2$, and hence $x+yq \in (f, g)\mathcal{O}_P$. This shows $(x + yq, y^N)\mathcal{O}_P \subset (f, g)\mathcal{O}_P$. The other inclusion is similar, but needs more work of reconstructing the polynomials f and g from the resultant and powers of x. ∎

[39] We haven't quite shown this, but it is not very hard to do with the tools that we have developed. A fuller discussion of the theory of resultants would include this result. The resultant R is homogeneous of degree mn precisely because $F_0G_0 \neq 0$.

1.15 07/12/24 - More Applications, Pencils of Curves, Introduction to Elliptic Curves

Today, we'll do more applications of Bézout's Theorem and start talking about elliptic curves, which will be our main focus for the last few lectures. Before we do that though, it is helpful to have a few handy corollaries and ideas. Here are two immediate applications of Bézout's Theorem.

Theorem 1.15.1. Let k be an algebraically closed field.

(a) If $C, D \subset \mathbb{P}^2_k$ are any two projective curves, then $C \cap D \neq \emptyset$.
(b) Any smooth projective curve is irreducible.

Proof. The statement (a) is an immediate corollary of Bézout's Theorem (Theorem 1.14.1). For (b), if a projective curve has multiple components, then some two of these components must intersect somewhere by (a), and then by Theorem 1.9.6, this point of intersection is a singular point of the curve. ■

Note that (a) is sharp in the sense that it is possible for two curves of any degrees $m, n \geq 1$ to intersect in a single point with multiplicity mn. We shall have occasion to use (b) repeatedly below.

1.15.1 Pencils of Curves and the Quartic Equation

Let's now talk about linear one parameter families of curves, starting with a couple of examples.

Example 1.15.2. The family $\mathcal{C} = \{C_\lambda\}_{\lambda \in k}$ of curves, where C_λ is the horizontal line defined by $y - \lambda = 0$ is a one-parameter family of curves of degree 1. When $\lambda \to \infty$, curve C_λ seems to disappear; one way to rectify this is to write this family projectively as given by the vanishing locus of $\mu Y - \lambda Z = 0$ for $\Lambda = [\lambda : \mu] \in \mathbb{P}^1_k$, so when λ is "infinity", i.e. $\Lambda = [1 : 0]$, then the corresponding line is simply $Z = 0$, the line at infinity–we could have predicted that. Note that in this case, each member of the family has degree exactly 1.

Example 1.15.3. Now consider the family $\mathcal{C} = \{C_\Lambda\}_{\Lambda \in \mathbb{P}^1_k}$ of curves, where C_Λ for $\Lambda = [\lambda : \mu]$ is the vanishing locus of $\lambda YZ - \mu X^2 = 0$. This is a one-parameter family of conics (specifically parabolae), and the member C_Λ is singular iff $\Lambda = [1 : 0]$ or $\Lambda = [0 : 1]$; in the former case, it is the union of the x-axis and L_∞, and in the latter case, it is the ("doubled") y-axis. Note that $\deg C_\Lambda = 2$ for all Λ except $[0 : 1]$, where $\deg C_{[0:1]} = 1$.

These examples motivate the following definition.

Definition 1.15.4.

(a) A pencil $\mathcal{C}$ of projective plane curves of degree d is a one-parameter linear family $\mathcal{C} = \{C_\Lambda\}_{\Lambda \in \mathbb{P}^1_k}$ of projective curves, all but finitely many members of which have degree d.
(b) Given a pencil $\mathcal{C}$ of curves, we define the base locus of $\mathcal{C}$ to be

$$\mathrm{BL}(\mathcal{C}) := \bigcap_{C \in \mathcal{C}} C$$

the intersection of all the curves in the pencil.

Concretely, a pencil $\mathcal{C}$ of degree d is given by specifying two linearly independent $F, G \in k[X, Y, Z]_d$ and then defining

$$C_\Lambda := C_{\lambda F + \mu G}$$

for $\Lambda = [\lambda : \mu] \in \mathbb{P}^1_k$. In this case, we have

$$\mathrm{BL}(\mathcal{C}) = C_F \cap C_G.$$

Of course, the choices for F and G are not unique: any two F', G' that form a basis for the span $k\langle F, G\rangle$ of F and G can be chosen as our F and G spanning the pencil, at the cost of changing the parameter Λ representing each curve C_Λ (by a projective change of coordinates in $\mathbb{P}^1_k$.) Saying that all but finitely many members of $\mathcal{C}$ have degree d is equivalent to saying that there is no homogeneous polynomial $H \in k[X, Y, Z]$ such that $H^2 \mid F, G$ (check!); this is a condition we will assume from henceforth as well.

Remark 1.15.5. With our description of the parameter space $\mathbb{P}^{d(d+3)/2}_k$ for all curves of degree $d \geq 1$, a pencil corresponds exactly to a line $\mathbb{P}^1_k \cong L \subset \mathbb{P}^{d(d+3)/2}_k$. Similarly, a two-parameter family (given by a plane $\mathbb{P}^2_k \cong \Lambda \subset \mathbb{P}^{d(d+3)/2}$) is called a net and a three-parameter family is called a web (which are some rather pictorial names); in general, a k-dimensional linear family of curves of degree d is also called a k-dimensional linear system of degree d curves. Note also that we cannot, in general, expect all curves in our pencil to have degree exactly d, as Example 1.15.3 illustrates that we cannot ask all members of our pencil to have the same degree; this can be done (e.g. if we consider the "double" y-axis to have degree 2), but needs the language of schemes. As we shall see below, the notion of base locus also behaves most nicely when we are in the world of schemes, so we can keep track of tangency of the members of our pencil as well.

Example 1.15.6. A pencil of lines is just the family of all lines in $\mathbb{P}^2_k$ passing through some fixed point $P \in \mathbb{P}^2_k$; in particular, there is only one kind of pencil of lines up to projective changes of coordinates, and the family of all pencils of lines in $\mathbb{P}^2_k$ is exactly $\mathbb{P}^2_k$.

Example 1.15.7. Over an algebraically closed field of characteristic other than 2, there are exactly 8 types of pencils of conics up to projective changes of coordinates. Ifa pencil $\mathcal{C}$ contains at least one smooth member, then the base locus of $\mathcal{C}$ consists of at most 4 distinct points, and the intersection multiplicities of at the base locus add up to 4; in other words, family containing one smooth member are indexed by partitions of 4, of which there are five. Conversely, if two pencils, each containing one smooth member, give rise to the same partition, then either one can be taken to the other by a projective change of coordinates. If all members of $\mathcal{C}$ are singular, then the base locus can be either a point, the union of a point and a line not passing through it, or a line. If it is a point P_0, then the pencil consists only of pairs of lines intersecting at that point, and no line is common to all such pairs. If it is the union $\{P_0\} \cup L$ for some point P_0 and line L such that $P_0 \notin L$, then the pencil consists of all reducible conics of the form $C_\Lambda = L \cup L_\Lambda$, where L_Λ is the pencil of all lines through P_0 (see Example 1.15.6). Finally, if the base locus is a line L, then there is a point $P_0 \in L$ such that the pencil again consists of all reducible conics of the form $C_\Lambda = L \cup L_\Lambda$, where L_Λ is the pencil of all lines through P_0. In these three degenerate case, the base locus completely determines the pencil up to projective changes of coordinates.[40] See Figure 1.10 for a picture illustrating these eight types, as well as their names. You are invited to prove these results in Exercise 2.6.2.

Example 1.15.8. We met examples of pencils of cubic curves in the proof of Pascal's Theorem (Theorem 1.13.5); see Figure 1.8 for an illustration.

[40] This happens also in the first case (i.e. when $\mathcal{C}$ has at least one smooth member), if we think of the base locus scheme-theoretically, i.e. as remembering what the multiplicities at each point of intersection are.

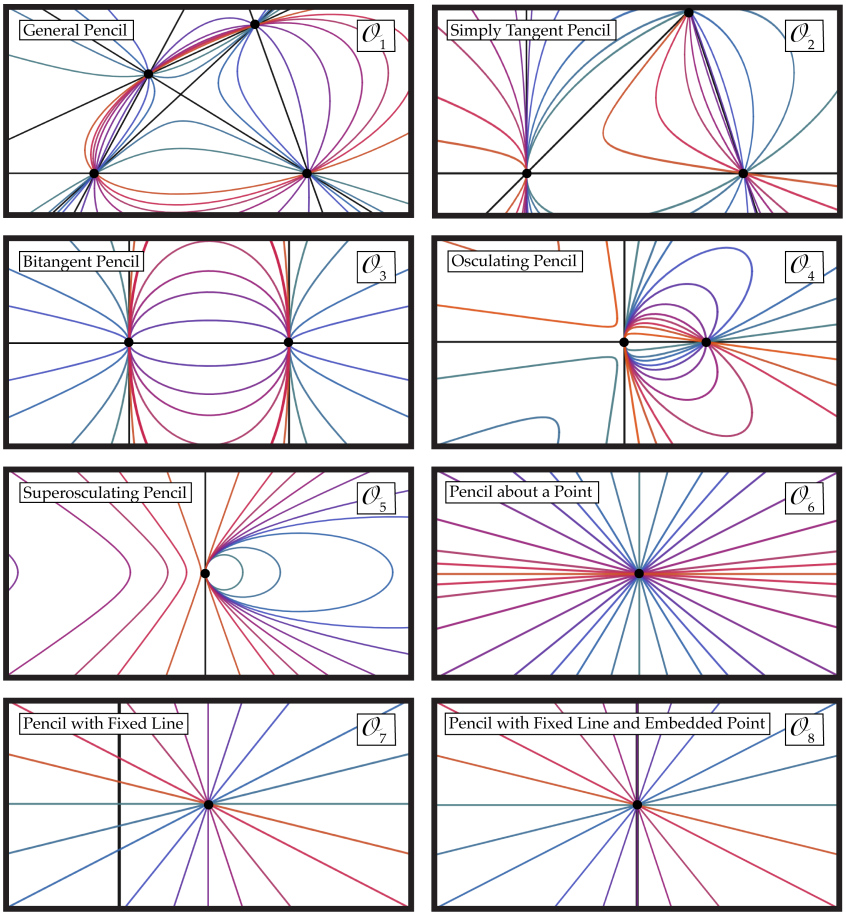


Figure 1.10: The eight types of pencils of conics up to projective changes of coordinates. Picture(s) made with Desmos.

Unfortunately, there are only finitely many types of pencils of degree d curves in $\mathbb{P}^2_k$, up to projective changes of coordinates, iff $d \in \{1, 2\}$. In general, for $d \geq 3$, classification of all pencils of curves of degree d, even in $\mathbb{P}^2_k$, is a very difficult problem. We will discuss the case of $d = 3$ in detail when we talk about the classification of elliptic curves in $\mathbb{P}^2_k$.

Here's one cool thing we can say about pencils of conics.

Theorem 1.15.9. Let k be an algebraically closed field of characteristic other than 2, and let $\mathcal{C}$ be a pencil of conics in $\mathbb{P}^2_k$. Then either every member of $\mathcal{C}$ is reducible, or at most 3 are.

Proof. Note that if $\operatorname{ch} k \neq 2$, then a quadratic homogeneous polynomial $Q \in k[X, Y, Z]_2$ can be written as

$$Q = \begin{bmatrix} X & Y & Z \end{bmatrix} \begin{bmatrix} A & H & E \\ H & B & F \\ E & F & C \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix},$$

where the matrix in the middle determines, and iis uniquely determined, by Q.[41] If we denote

[41] This is the reason that the classification of projective conics is intimately related to the theory of binary quadratic forms. See Remark 1.12.15.

this matrix by M_Q, then we see that

$$\begin{bmatrix} \partial_X Q \\ \partial_Y Q \\ \partial_Z Q \end{bmatrix} = 2 \cdot M_Q \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}.$$

In particular, it follows from the Projective Jacobi Criterion (Theorem 1.12.10) that the conic C_Q defined by Q (when $Q \neq 0$) is singular iff M_Q has a nonzero kernel (i.e. a nonzero eigenvector with eigenvalue 0), which happens iff $\det M_Q = 0$, as we have talked about several times. Now given two such linearly independent Q_1, Q_2 and corresponding matrices $M_i := M_{Q_i}$ for $i = 1, 2$, the pencil $\mathcal{C}$ containing $C_i = C_{Q_i}$ for $i = 1, 2$ is given by taking C_Λ to be the curve defined by the vanishing of $\lambda Q_1 + \mu Q_2 = 0$. The matrix representative of this quadric is given exactly by

$$\lambda M_{Q_1} + \mu M_{Q_2}.$$

By the first observation, the reducible conics of the pencil $\mathcal{C}$ correspond exactly to the roots Λ of the equation

$$\det(\lambda M_{Q_1} + \mu M_{Q_2}) = 0.$$

Since this is homogeneous cubic equation in λ and μ, it is either identically zero (in which case every member of $\mathcal{C}$ is reducible), or it has at most three roots, in which case at most three members of $\mathcal{C}$ are reducible, and the rest smooth. ■

Note that a pencil can have any number of singular members between 1 and 3 (inclusive)– the precise number corresponds to the multiplicities of the roots of the cubic polynomial $\det(\lambda M_{Q_1} + \mu M_{Q_2})$, and can also be read off from the geometry of the base locus (how?).

Example 1.15.10. Let k be an algebraically closed field and let $C, D \subset \mathbb{P}^2_k$ be two conics that intersect in exactly 4 distinct points $P_1, \ldots, P_4$. In this case, these four points must be in general position (Definition 1.12.4); indeed, if some three of them were to lie on a line L, then every conic through them would have to contain L (by Bézout's Theorem for lines or conics), and hence any two distinct conics passing through them would intersect in all points along L, of which there are infinitely many (Proposition 1.11.13).

In this case, the pencil of conics containing C and D is said to be a general pencil; see the case $\mathcal{O}_1$ in Figure 1.10 for an illustration of this type of pencil. The claim is that such a pencil consists of all conics passing through these four points (and, in particular, always contains smooth members). This can be proven using Max Noether's Fundamental Theorem (Theorem 1.16.1) which we will use to prove Chasles's Theorem (Theorem 1.15.14) next time, or using a dimension argument on the number of linear constraints imposed on conics by four points in general position, but an alternative, direct, proof runs as follows. Let E be any other conic passing through these four points, and pick a fifth point P_5 on E distinct from $P_1, \ldots, P_4$. Since no four of $P_1, \ldots, P_5$ are collinear, it follows that E is the unique conic passing through $P_1, \ldots, P_5$ (Theorem 1.13.12(b)). In particular, if we can find a conic E' in the pencil spanned by C and D that contains P_5, then we would have shown that $E = E'$ and hence that E is in the pencil spanned by C and D.

For this, we claim first that $P_5 \notin C \cup D$; indeed, if $P_5 \in C$, then by Bézout's Theorem for conics (Theorem 1.13.4), we know that E and C share a component. Since E and C are distinct conics, this can only happen in $E = L_1 \cup L_2$ and $C = L_2 \cup L_3$ for some distinct lines $L_1, L_2, L_3 \subset \mathbb{P}^2_k$ with $P_5 \in L_2$. Since L_2 contains exactly two of the four points P_i, say P_1 and P_2, and both E and C pass through P_3 and P_4 as well, it follows that both L_1 and L_3 are lines joining P_1 and P_4, whence $L_1 = L_3$, which is a contradiction. Therefore, as in the proof of Theorem 1.13.5, if we take F and G to be homogeneous equations defining C and D respectively, and

pick a representative (X_0, Y_0, Z_0) for $P_5 = [X_0 : Y_0 : Z_0]$, then $F(X_0, Y_0, Z_0) \cdot G(X_0, Y_0, Z_0) \neq 0$, and the curve $E' = C_\Lambda = C_{\lambda F + \mu G}$ in the pencil spanned by C and D, where

$$\Lambda = [\lambda : \mu] = [-G(X_0, Y_0, Z_0) : F(X_0, Y_0, Z_0)]$$

contains P_5, and we are done. (That Λ is well-defined uses that $P \notin C \cup D$, or at least that one of $P \notin C$ and $P \notin D$ holds.)

Therefore, we have shown that a general pencil of conics is exactly the set of all conics that pass through four points $P_1, \ldots, P_4$ in $\mathbb{P}^2_k$ in general position. Since any such tuple of points can be taken to any other by a projective change of coordinates (this was Proposition 1.12.5), it follows that any two general pencils are related by a projective change of coordinates. This is an $1/8^{\text{th}}$ of the solution to Exercise 2.6.2.

Finally, note that if $\mathcal{C}$ is a general pencil of conics through the points $P_1, \ldots, P_4$, then we can see explicitly what the exactly three reducible conics in $\mathcal{C}$, as suggested by Theorem 1.15.9 are: namely, they are the three pairs of lines that are opposite edges of the complete quadrilateral with vertices $P_1, \ldots, P_4$; i.e. if for $1 \leq i < j \leq 4$, we let L_{ij} be the line joining P_i and P_j, then the three reducible conics are exactly $L_{12} \cup L_{34}, L_{13} \cup L_{24}$ and $L_{14} \cup L_{23}$.

This observation gives us a way to find the intersection points of two conics that intersect in 4 points as follows. Given equations Q_1 and Q_2 of conics intersecting in 4 distinct points, we find the roots of the cubic polynomial $\det(\lambda M_{Q_1} + \mu M_{Q_2})$ (say via Cardano's method), and use this to find the singular members of the pencil spanned by Q_1 and Q_2. Then we decomposition the equation of these singular members into equations of the corresponding lines (by solving quadratic equations). Finally, the four intersection points of the original conics will be contained in the 6 pairwise intersection points of these lines, and lines are easy enough to intersect.

Example 1.15.11. Here's an example of how to use pencils of conics to solve the quartic equation, at least when the characteristic of the base field is other than 2. Suppose we are trying to solve the equation

$$x^4 + ax^3 + bx^2 + cx + d = 0$$

over a field k with $\operatorname{ch} k \neq 2$. It is easy to see (check!) that solving this equation is equation amounts to finding the intersection points of the two parabolae given by the vanishing of the homogeneous polynomials

$$\begin{aligned} Q_1 &= Y^2 + aXY + bYZ + cXZ + dZ^2, \text{ and} \\ Q_2 &= YZ - X^2, \end{aligned}$$

since they do not intersect on the line at infinity. Then the corresponding matrices M_{Q_1} and M_{Q_2} are easily seen to be

$$M_{Q_1} := \begin{bmatrix} 0 & a/2 & c/2 \\ a/2 & 1 & b/2 \\ c/2 & b/2 & d \end{bmatrix} \text{ and } M_{Q_2} := \begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 1/2 \\ 0 & 1/2 & 0 \end{bmatrix},$$

whence

$$\det(\lambda M_{Q_1} + \mu M_{Q_2}) = -\frac{1}{4}\left[\left(a(ad - bc) + c^2\right)\lambda^3 + (ac - b^2 + 4d)\lambda^2\mu + 2b\lambda\mu^2 - \mu^3\right].$$

Then, we may solve this cubic, and use this as suggested in Example 1.15.10 to find the intersection points of C_{Q_1} and C_{Q_2}, and hence the roots of the quartic equation. You are invited to work out one (carefully chosen) example in detail in Exercise 2.6.3. The whole procedure above can be simplified slightly by first depressing the quartic (i.e. replacing X by $X - (1/4)a$) and then applying the above procedure. For a (slightly) more detailed explanation of the procedure and its connection to Galois theory, as well as references, see [5, §1.14].

1.15.2 An Introduction to Elliptic Curves

We now want to focus on the next simplest case of curves after the conics, namely the cubic curves. We already classified all singular plane cubics up to projective changes of coordinates (at least over algebraically closed fields of characteristic other than 3) in Exercise 2.4.4, so we may now focus on the case of smooth cubics–it turns out that such curves admit a very rich theory, which makes them very powerful objects in modern algebraic geometry.

Definition 1.15.12. An elliptic curve (over a field k) is a pair (E, O), where $E \subset \mathbb{P}^2_k$ is a smooth cubic curve, and $O \in E$.

The reader will not lose much by imagining k to be algebraically closed (otherwise our definition of smoothness is not quite the right one), and soon we will be assuming $\operatorname{ch} k \neq 2, 3$ as well for convenience, but it is helpful to have the right level of generality and to be able to talk about points of elliptic curves over finite fields, for instance.

Now consider the binary operation $+ : E \times E \to E$ defined as follows: given a pair $(A, B) \in E \times E$, let the line[42] $L_{A,B}$ joining A and B intersect the curve E in the third point D.[43] Then we define $A + B := +(A, B)$ to be the third point of intersection of E and the line $L_{O,D}$ joining O and D. See Figure 1.11. The key claim, from which the power of elliptic curves comes, is

Theorem 1.15.13. Let (E, O) be an elliptic curve. Then the binary operation $+ : E \times E \to E$ defined above makes E into an abelian group with identity O.

Proof. Commutativity of $+$ is clear, as is the fact that $A + O = A$ for all $A \in E$: indeed, if the line $L_{A,O}$ meets the curve again in A', then the line $L_{O,A'}$ meets the curve again in A. To find inverses, consider once and for all the point $O' \in E$ which is the third point of intersection of the tangent line $L_{O,O} = \mathrm{T}_O E$ with E; then it is easy to see that given any $A \in E$, the third intersection point A'' of $L_{AO'}$ with E has the property that $A + A'' = O$. Finally, we have to show associativity.

For this, consider points $A, B, C \in E$. Let D denote the third intersection of $L_{A,B}$ with E, let F denote the third intersection of $L_{A+B,C}$ with E, and let G denote the third intersection of $L_{B,C}$ with E. (See Figure 1.11.) To show associativity, it suffices to show that the line $L_{A,B+C}$ passes through F (check!). Temporarily denote the third intersection point of $L_{A,B+C}$ with E by F'; then we have to show that $F = F'$.

Consider the cubic curves $\Gamma := L_{A,B} \cup L_{C,F} \cup L_{O,G}$ and $\Sigma := L_{B,C} \cup L_{A,B+C} \cup L_{O,D}$, and note that

$$E \cap \Gamma = \{O, A, B, C, D, G, A + B, B + C, F\} \text{ and}$$
$$E \cap \Sigma = \{O, A, B, C, D, G, A + B, B + C, F'\}.$$

In particular, Σ is a cubic curve that passes through 8 of the 9 intersection points of the cubic curves E and Γ. Therefore, the proof is finished by the following theorem (Theorem 1.15.14). ∎

[42] When $A = B$, we take $L_{A,B}$ to be the tangent line to E at A, which we can do uniquely since E smooth.

[43] Here we are using Bézout's Theorem (Theorem 1.14.1 or at least Theorem 1.12.12). We do not disallow the possibility that $D = A, B, O$. For instance, $D = A$ if $A \neq B$ but the line $L_{A,B}$ is tangent to E at A, or if $A = B$ and $L_{A,B}$ meets E with multiplicity three at A (i.e. A is an inflection point of E). I will leave such considerations to the reader, but see also Remark 1.15.16.

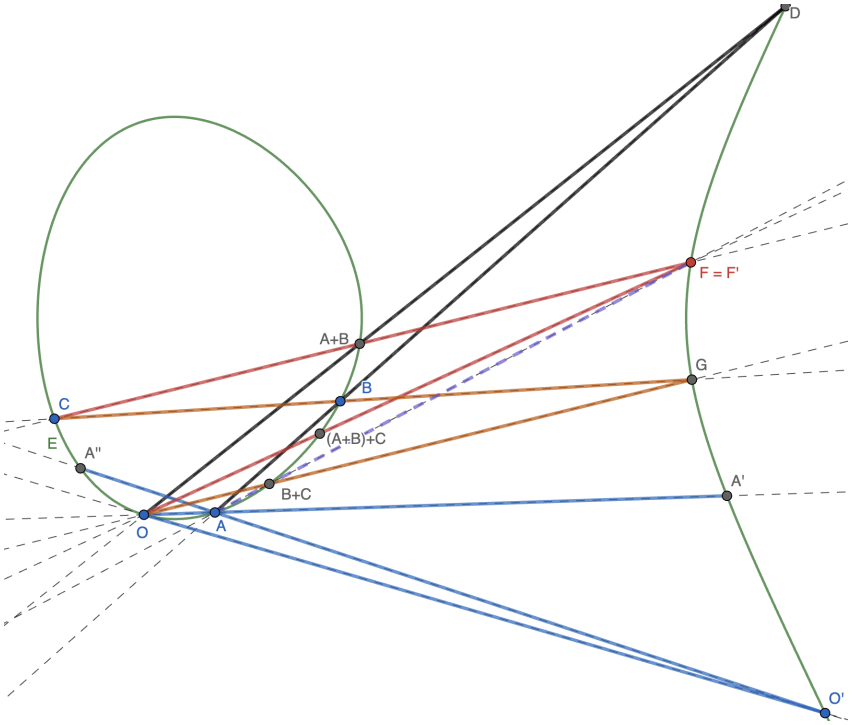


Figure 1.11: The addition law on an elliptic curve. Picture made with GeoGebra.

Theorem 1.15.14 (Chasles). Let $D, E \subset \mathbb{P}^2_k$ be two cubic curves that intersect in 9 points, and suppose one of D or E is irreducible. If $X \subset \mathbb{P}^2_k$ is another cubic curve that passes through 8 of 9 of these points, then X also passes through the 9th one.

There are many ways to prove Theorem 1.15.14. One approach is to use a dimension count: each point of intersection imposes one linear condition on the space $\mathbb{P}^9_k$ of homogeneous cubic equations, and so imposing 8 such general conditions brings us down to a $\mathbb{P}^1_k \subset \mathbb{P}^9_k$, i.e. a pencil of cubic curves. If C and D are two members of this pencil, then any other cubic curve passing through these 8 points belongs to the pencil spanned by C and D and hence also passes through the 9th point. For this argument to work, the points need to be in sufficiently general position–it turns out to be sufficient to assume that no 7 of the points $P_1, \ldots, P_9$ lie on a conic. For an argument along these lines, see either [5, Prop. 2.6], or this blog post [7] by Terry Tao. This $8 \Rightarrow 9$ phenomenon can be fruitfully generalized in the direction of the number of linear conditions imposed by points in projective space, resulting in the so-called the Cayley-Bacharach Theorem. Sometimes Theorem 1.15.14 itself is called the Cayley-Bacharach theorem, but this is a misnomer–see this paper [8] by Eisenbud, Green, and Harris[44] for an explanation of the Cayley-Bacharach Theorem and its relation to Chasles's Theorem. Next time, we will give a proof of Theorem 1.15.14 using a local-to-global principle called Max Noether's Fundamental Theorem (Theorem 1.16.1).

Remark 1.15.15. Chasles's Theorem (Theorem 1.15.14) immediately implies those of Pascal (Theorem 1.13.5) and Pappus (Theorem 1.13.7); for instance, to deduce Pascal's Theorem in the notation used in that section, we can take the two cubic curves to be $D := L_1 \cup L_3 \cup L_5$ and $E := L_2 \cup L_4 \cup L_6$ (so the intersection points are $P_1, \ldots, P_6, Q_1, Q_2, Q_3$), and then take X to be the union of the conic C and the line L joining Q_1 and Q_2.

[44] Harris was my advisor!

Remark 1.15.16. The proof of Theorem 1.15.13 and the statement of Theorem 1.15.14 certainly work as written when all the 9 involved points are distinct, but that is not quite sufficient to prove Theorem 1.15.13. We also need to take into account intersection multiplicites and tangencies. There are a few ways to get around this. Over fields such as $k = \mathbb{R}$ or $k = \mathbb{C}$, we may use continuity arguments, as in indicated for instance in [5, §I.2]. Over general fields, we can use a similar argument, but using the rigidity of complete varieties instead, as explained in [9, Chapter 3]. Alternatively, one can write down explicit formulae for the group law and verify all the claims directly via (very) tedious computation. Finally, we can treat the whole theory as above somewhat more carefully using the notion of intersection multiplicities already introduced, and note that Theorem 1.15.14 also works when we counts point with intersection multiplicity.[45] This last one is, generally speaking, the approach we will take, as we shall see in the proofs next time.

Remark 1.15.17. Suppose that an elliptic curve E defined over a field k is smooth over its algebraic closure $\overline{k}$. The above addition law tells us then that the set of k-rational points $E(k)$ of E form a subgroup of $E(\overline{k})$–indeed, this follows from the group law because the third intersection point of a L joining two k-points with a cubic curve defined over k is also defined over k, because a cubic equation with coefficients in k and two roots in k must also have its last root in k.

In particular, for instance, it makes sense to talk about, say, the subgroup real points of a complex elliptic curve which is defined over the real numbers and has $O \in E(\mathbb{R})$. Such a "real elliptic curve" is then a topological–even Lie–group. It seems also from Figure 1.11 above that the sums $A+B, B+C$ and $A+B+C$ lie on the same component of the two-component elliptic curve as A, B, C, as long as this component contains O, i.e. that the component containing O of a real two-component elliptic curve is a subgroup of the whole curve under the addition law, although it is not an algebraic curve itself (Example 1.7.15). You are invited to explore this in Exercise 2.6.8.

Next time, we will prove Theorem 1.15.14 and start working with explicit examples of elliptic curves.

[45] For instance, if instead of 9 distinct points $P_1, \ldots, P_9$ we have only 8 distinct points $P_1, \ldots, P_8$ of intersection but tangency at P_8, then the statement says also that if X passes through $P_1, \ldots, P_8$, then it is also tangent to both D and E at P_8.

1.16 07/15/24 - Max Noether's Theorem, Proof of Chasles's Theorem, Weierstrass Normal Form

The first order of business today is to prove Chasles's Theorem, for which we will need

Theorem 1.16.1 (Max Noether). Let $F, G, H \in k[X, Y, Z]$ be relatively prime homogeneous polynomials of degrees $m, n, d \geq 1$ such that F and G are relatively prime. Then H can be written as

$$H = AF + BG$$

for some homogeneous $A, B \in k[X, Y, Z]$ of degrees $d - m, d - n$ iff for each point $P \in \mathbb{P}^2_k$, we have

$$(H)\mathcal{O}_{\mathbb{P}^2_k, P} \subset (F, G)\mathcal{O}_{\mathbb{P}^2_k, P}.$$

This theorem, often called Max Noether's $AF + BG$ Theorem, or Max Noether's Fundamental Theorem, is again an upgraded version of the local-to-global principal Lemma 1.14.2, and says that H is globally a polynomial-linear combination of F, G iff it is locally a polynomial-linear combination of F and G at each point P.

Proof. One direction is clear. For the other, assume that $(H)\mathcal{O}_{\mathbb{P}^2_k, P} \subset (F, G)\mathcal{O}_{\mathbb{P}^2_k, P}$ for all $P \in \mathbb{P}^2_k$, and suppose by a projective change of coordinates that all points of $C_F \cap C_G$ are in the finite plane, i.e. not on L_∞. If $f, g, h \in k[x, y]$ are the dehomogenizations of F, G, H respectively, then it follows that $h \in (f, g)\mathcal{O}_P$ for all $P \in C_f \cap C_g$, so from Lemma 1.14.2, it follows that $h \in (f, g)k[x, y]$, i.e. $h = af + bg$ for some $a, b \in k[x, y]$. Homogenization then yields

$$Z^r H = AF + BG$$

for some $r \geq 0$ and $A, B \in k[X, Y, Z]$ homogeneous of degrees $d+r-m$ and $d+r-n$ respectively. The result then follows by induction from the following lemma. ∎

Lemma 1.16.2. Let $F, G \in k[X, Y, Z]$ be relatively prime homogeneous polynomials of degrees $m, n \geq 1$ such that $C_F \cap C_G \cap L_\infty = \emptyset$. If $H \in k[X, Y, Z]$ is a homogeneous polynomial of degree $d \geq 1$ such that

$$ZH = AF + BG$$

for some homogeneous $A, B \in k[X, Y, Z]$ of degrees $d + 1 - m, d + 1 - n$ respectively, then there are $A', B' \in k[X, Y, Z]$, homogeneous of degrees $d - m, d - n$ respectively such that

$$H = A'F + B'G.$$

In other words, if C_F and C_G do not intersect on the line at infinity, then multiplication by Z is injective on the quotient ring $k[X, Y, Z]/(F, G)$.

Proof. For $P \in k[X, Y, Z]$, let P° denote the specialization $P^\circ := P(X, Y, 0) \in k[X, Y, Z]$; then $Z \mid P$ iff $P^\circ = 0$. Specializing the equation $ZH = AF + BG$ yields

$$A^\circ F^\circ + B^\circ G^\circ = 0.$$

Since $C_F \cap C_G \cap L_\infty = \emptyset$, the polynomials $F^\circ, G^\circ \in k[X, Y]$ are relatively prime, and hence there is a $C \in k[X, Y]$ such that $A^\circ = CG^\circ$ and $B^\circ = -CF^\circ$. In this case, the polynomial

$A - CG$ has the property that $(A - CG)^\circ = A^\circ - CG^\circ = 0$, whence there is an $A' \in k[X, Y, Z]$ such that $A - CG = A'Z$. Similarly, there is a $B' \in k[X, Y, Z]$ such that $B + CF = B'Z$. These A' and B' work. ■

We are now ready to prove Chasles's Theorem. For simplicity, I will do the case when the nine points of intersection are distinct, leaving the general case (with multiplicities) to the dedicated reader. This is not too unfair, since we have developed all the necessary tools for this extension already. The advantage of working with distinct points is that it makes Theorem 1.16.1 very easy to apply.

Lemma 1.16.3. Let $D, E \subset \mathbb{P}^2_k$ be projective curves of degrees $m, n \geq 1$ which intersect in exactly mn distinct points, and let $Y \subset \mathbb{P}^2_k$ be a curve that passes through all mn of these points. If $F, G, H \in k[X, Y, Z]$ are minimal polynomials for D, E, Y respectively, then there are homogeneous polynomials $A, B \in k[X, Y, Z]$ of degrees $\deg(H) - \deg(F), \deg(H) - \deg(G)$ respectively such that $H = AF + BG$.

Proof. By Theorem 1.16.1, it suffices to show that $(H)\mathcal{O}_{\mathbb{P}^2_k, P} \subset (F, G)\mathcal{O}_{\mathbb{P}^2_k, P}$ for all $P \in \mathbb{P}^2_k$. When $P \notin D \cap E$, this is clear, since the right hand side is all of $\mathcal{O}_{\mathbb{P}^2_k, P}$. Now suppose that $P \in D \cap E$. Our hypothesis coupled with Bézout's Theorem implies that $i_P(D, E) = 1$, and we have to show that this combined with $P \in Y$ implies the result. This is clearly a local computation, so we can pass to the affine case; let f, g, h denote the respective dehomogenizations. Then $\mathrm{eval}_P : \mathcal{O}_{\mathbb{A}^2_k, P} \to k$ is surjective with kernel containing (f, g) such that the quotient $\mathcal{O}_{\mathbb{A}^2_k, P}/(f, g)$ has dimension one; this gives us an isomorphism $\mathrm{eval}_P : \mathcal{O}_{\mathbb{A}^2_k, P}/(f, g) \to k$. In particular, $Y \ni P$ iff h lies in the kernel of this evaluation map iff $h \in (f, g)\mathcal{O}_{\mathbb{A}^2_k, P}$. ■

The only difference in the general case is that one needs to check the "Noether condition" $(H)\mathcal{O}_{\mathbb{P}^2_k, P} \subset (F, G)\mathcal{O}_{\mathbb{P}^2_k, P}$ by hand for each $P \in D \cap E$, so to speak. See [3, §5.5]. We are now ready to prove

Theorem 1.15.14 (Chasles). Let $D, E \subset \mathbb{P}^2_k$ be two cubic curves that intersect in 9 points, and suppose one of D or E is irreducible. If $X \subset \mathbb{P}^2_k$ is another cubic curve that passes through 8 of 9 of these points, then X also passes through the 9th one.

Proof. Suppose that D is irreducible, and write $D \cap E = \{P_1, \ldots, P_9\}$, with $P_i \in X$ for $i = 1, \ldots, 8$. Let the ninth point of intersection of X with D be Q, and suppose for the sake of contradiction that $Q \neq P_9$. Pick a general line L through P_9; it suffices to take one not passing through Q and meeting D in two distinct other points R, S. Then $E \not\ni R, S$. Applying Lemma 1.16.3 to $Y := X \cup L$, we conclude that if F, G, H are the homogeneous cubic polynomials defining D, E, X respectively, then there are homogeneous linear polynomials $A, B \in k[X, Y, Z]$ such that

$$LH = AF + BG.$$

(Here we are using L also to denote the linear polynomial defining the line L; we will also do this for A and B.) Now $R, S \notin E$ implies that the line G contains R and S, and hence must be identical with L. It follows that $L \mid AF$, but since F is assumed to be irreducible, this can only happen if $L = A$ (upto scaling). Cancelling the factor of L tells us that $H = \alpha F + \beta G$ for some scalars $\alpha, \beta \in k$, i.e. that X is in the pencil spanned by D and E. In particular, $X \ni P_9$, which is a contradiction. This shows that our assumption $Q \neq P_9$ is false, proving $X \ni P_9$ as needed. ■

With this we have now completed the proof of the associativity of the elliptic curve addition law–at least as long as all the points involved are distinct; see Remark 1.15.16. Let us now move on to some explicit examples illustrating how to work with elliptic curves.

1.16.1 Weierstrass Normal Form and Legendre Form, Two and Three Torsion

Recall our convention that k is an algebraically closed field of characteristic other than 2 or 3. In these circumstances, we given a smooth cubic $E \subset \mathbb{P}^2_k$, we can make a convenient choice of basepoint $O \in E$ and coordinates that makes the study of the elliptic curve (E, O) particularly convenient.

Firstly, the choice of basepoint O doesn't really matter all that much (see Exercise 2.6.9), but a convenient choice of O can make the addition law particularly easy. Namely, by Exercise 2.5.5, E has exactly 9 inflection points, and we pick O to be one of these flexes. The upshot of this is that in the addition law on E, we have $O' = O$ by definition (see the proof of Theorem 1.15.13), and hence the $-A, O$ and A are collinear for each $A \in E$; in fact, it is easy to see in this case (check!) that three points $A, B, C \in E$ (counted with multiplicity) are collinear iff $A + B + C = 0$ in the group law.

As a first consequence, not that this means that given a fixed $P \in E$, the point P is an inflection point on E iff the "points" P, P, P are collinear iff $3P = 0$ iff $P \in E[3]$ is a 3-torsion point. In particular, Exercise 2.5.5 gives us that $E[3]$ is an abelian group with 9 elements, each of order 3, and hence that $E[3] \cong \mathbb{Z}/3 \times \mathbb{Z}/3$. This is the first observation in a very large story, another part of which we shall see below and which you will be asked flesh out in detail in Exercise 2.6.10.

Given an elliptic curve (E, O) with $O \in E$ an inflection point, we can now bring E into what is called the (reduced) Weierstrass normal form. Here's how this goes: pick a coordinate system in which $O = [0 : 1 : 0]$ with the tangent line $\mathrm{T}_O E$ being the line at infinity $Z = 0$. Let F be the minimal polynomial of E, and write F as

$$F = A_0 Y^3 + A_1 Y^2 + A_2 Y + A_3$$

for $A_i \in k[X, Z]_i$ homogeneous of degree i for $i = 0, \ldots, 3$. The condition $O \in E$ implies $A_0 = 0$, the condition $\mathrm{T}_O E = \mathbb{V}(Z)$ implies that $A_1 = Z$ (possibly after scaling, which we do), and the condition that $O \in E$ is an inflection point says that $Z \mid A_2$. Therefore, the polynomial F looks like

$$Y^2 Z + (\lambda X + \mu Z) Y Z + A_3.$$

Since $\operatorname{ch} k \neq 2$, we can replace Y by $Y - (\lambda X + \mu Z)/2$ to eliminate the middle term, so that the equation looks like

$$Y^2 Z = \alpha_0 X^3 + \alpha_1 X^2 Z + \alpha_2 X Z^2 + \alpha_3 Z^3$$

for some $\alpha_i \in k$ for $i = 0, \ldots, 3$. Since E is irreducible, we must have $\alpha_0 \neq 0$; replacing Z by $\alpha_0 Z$, we may assume that $\alpha_0 = 1$ to get an equation of the form

$$Y^2 Z = X^3 + \beta_1 X^2 Z + \beta_2 X Z^2 + \beta_3 Z^3.$$

Finally, using $\operatorname{ch} k \neq 3$, we may replace X by $X - \frac{1}{3}\beta_1 Z$ to depress this last cubic to obtain the reduced Weierstrass normal form

$$Y^2 Z = X^3 + p X Z^2 + q Z^3$$

for some $p, q \in k$, or in affine coordinates

$$y^2 = x^3 + px + q.$$

By (a salvage of) Exercise 2.3.10 combined with Exercise 2.2.5(b), this curve is smooth iff

$$4p^3 + 27q^2 \neq 0.$$

One thing this form enables us to see immediately is the two-torsion $E[2]$ on E. Firstly, the only point on E at infinity (i.e. on $Z = 0$) is the point O. Next, given a(n) (affine) point $P = (x, y)$ on E, when E is in Weierstrass form, since P, O and $-P$ are collinear, we see that $-P = (x, -y)$. In particular, $2P = 0$ iff $P = -P$ iff $P = O$ or $P = (x, y)$ with $y = 0$. In other words, the two-torsion points other than O correspond directly to the roots of $x^3 + px + q$; if these roots are $e_1, e_2, e_3 \in k$ (using here that $k = \overline{k}$), then

$$E[2] = \{O, (e_1, 0), (e_2, 0), (e_3, 0)\}.$$

Note that the discriminant condition $4p^3 + 27q^2 \neq 0$ (or equivalently the nonsingularity of E) implies the roots e_1, e_2, e_3 are pairwise distinct, whence $E[2]$ is an abelian group of size 4; since every nontrivial element of $E[2]$ has order 2, we see immediately that

$$E[2] \cong \mathbb{Z}/2 \times \mathbb{Z}/2.$$

The two examples here suggest the following generalization: is it always true that for any $n \geq 1$ we have

$$E[n] \cong \mathbb{Z}/n \times \mathbb{Z}/n,$$

as we have shown for $n = 1, 2, 3$? In fact, this is always true in characteristic zero, or more generally if $\operatorname{ch} k \nmid 2n$; for a proof, see Exercise 2.6.10. The best way I know of understanding this result, however, involves seeing connections to a different branch of math, namely complex analysis; I'll cover this in the story time during the next lecture–see §1.17.2.

The above version of the Weierstrass normal form is convenient, but it doesn't make it clear how the isomorphism class of E depends on (p, q). For starters, replacing Z by uZ tells us that the curves given by (p, q) and (u^2p, u^3q) are the same for any $u \in k^\times$. It turns out, but is more difficult to prove, that two elliptic curves are in short Weierstrass form are isomorphic iff there is such a transformation between them. We'll pursue a slightly different line of study, via a slightly different variant of the Weierstrass form.

Namely, recall as above that we by a chance of coordinates assume that the curve is given as

$$Y^2Z = X^3 + \beta_1 X^2 Z + \beta_2 X Z^2 + \beta_3 Z^3.$$

This time, we'll factor the right hand side as

$$(X - e_1 Z)(X - e_2 Z)(X - e_3 Z)$$

for some distinct $e_i \in k$. Next, replacing X by $X - e_1 Z$, we will assume that $e_1 = 0$; then $e_2 e_3 \neq 0$. Finally, by replacing Z by $e_2^{-1} Z$ and Y by $e_2^{1/2} Y$ (again using $k = \overline{k}$), we arrive at the Legendre form

$$Y^2Z = X(X - Z)(X - \lambda Z)$$

for some $\lambda \in k \smallsetminus \{0, 1\}$. Written in affine coordinates, this is

$$y^2 = x(x - 1)(x - \lambda).$$

Let us denote this curve by E_λ. One can then ask: when are E_λ and E_μ for $\lambda, \mu \in k \smallsetminus \{0, 1\}$ related by a projective change of coordinates? Giving a complete answer to this question will allow us to give a classification of elliptic curves. This is what we will pursue next time.

1.17 07/17/24 - Classification of Elliptic Curves, Story Time

Recall our standing assumption that the base field k is algebraically closed of characteristic other than 2 or 3. We showed in lecture last time that every elliptic curve (E, O) with $O \in E$ an inflectionary point can be put into Legendre form, i.e. there is some $\lambda \in k \smallsetminus \{0, 1\}$ and a projective change of coordinates taking E to the curve E_λ which is the projective closure of the affine curve defined by

$$y^2 = x(x-1)(x-\lambda)$$

with basepoint $O = [0 : 1 : 0]$. For a given E, how many such λ work? In other words, when are the curves E_λ and E_μ for $\lambda, \mu \in k \smallsetminus \{0, 1\}$ isomorphic (i.e. related by a change of coordinates)? Answering this question will enable us to "classify" all elliptic curves in the sense that we will be able to tell exactly when two such cubics are related by a change of coordinates, somewhat similarly to Theorem 1.12.13. For this, we need to introduce a key invariant of elliptic curves–the j-invariant.

1.17.1 The j-Invariant

I have used the word "isomorphism" quite a few times already; let me explain what I mean by that at the moment.

Definition 1.17.1.

(a) Two curves $D, E \subset \mathbb{P}^2_k$ are said to be projectively isomorphic if there is a projective change of coordinates $\Phi : \mathbb{P}^2_k \to \mathbb{P}^2_k$ such that $\Phi(D) = E$.

(b) Let (E, O) and (E', O') be two plane elliptic curves. We say that (E, O) and (E', O') are weakly isomorphic if the underlying cubic curves $E, E' \subset \mathbb{P}^2_k$ are projectively isomorphic; we say that (E, O) and (E', O') are strongly isomorphic if the projective change of coordinates $\Phi : \mathbb{P}^2_k \to \mathbb{P}^2_k$ taking $\Phi(E) = E'$ can be chosen to satisfy $\Phi(O) = O'$. Such a Φ is called a strong isomorphism $(E, O) \to (E', O')$.

Note that projective isomorphism (i.e. being projectively isomorphic) is an equivalence condition (often denoted by $\cong$) on the set of all projective plane curves; projectively isomorphic curves have the same number and degrees of irreducible components, singular points, etc. We will often denote projective isomorphism via the notation $\cong$.Effectively, they are "the same" curve, just viewed under different coordinates. Note also that strongly isomorphic elliptic curves are weakly isomorphic, and a strong isomorphism Φ is automatically a group isomorphism thanks to the geometric nature of the group law on E. Finally, if $E \subset \mathbb{P}^2_k$ is a smooth cubic, and $O \in E$ an inflection point while $O' \in E$ *not* an inflection point, then the elliptic curves (E, O) and (E, O') are *not* strongly isomorphic, since the property of being an inflection point of a curve is preserved under projective changes of coordinates.

Remark 1.17.2. The terminology here is my own and not standard. Further, in slightly more advanced treatments of algebraic geometry, there is yet another notion of isomorphism: the notion of an "abstract" isomorphism of algebraic curves, given by polynomial or rational functions. For instance, the parametrization in Example 1.13.3 combined with the projection explained in Lecture 1.3 tells us that a line $L \subset \mathbb{P}^2_k$ and a smooth conic $C \subset \mathbb{P}^2_k$ are abstractly isomorphic, although they cannot be projectively isomorphic because they have different degrees. However, it turns out that for elliptic curves the notions of abstractly isomorphic and projectively isomorphic agree, although showing this needs more work. (See the grown-up text [9] if you are curious.) Here we will restrict ourselves to the study of projective isomorphisms.

Finally, one last definition that we will need is

Definition 1.17.3. The j-function $j : k \setminus \{0,1\} \to k$ is the rational function defined by

$$j(\lambda) := 2^8 \frac{(\lambda^2 - \lambda + 1)^3}{\lambda^2(\lambda - 1)^2}.$$

The origin of this mysterious function will be explained in Remark 1.17.7 below; before that, we arrive at the (somewhat surprising) main result of this section.

Theorem 1.17.4. Let $\lambda, \mu \in k \setminus \{0,1\}$, and let $O := [0:1:0]$ be the usual basepoint. The following are equivalent:

(a) The elliptic curves (E_λ, O) and (E_μ, O) are strongly isomorphic.
(b) The curves $E_\lambda, E_\mu \subset \mathbb{P}^2_k$ are projectively isomorphic.
(c) We have $\mu \in M_\lambda := \{\lambda, 1/\lambda, 1-\lambda, 1/(1-\lambda), \lambda/(\lambda-1), (\lambda-1)/\lambda\}$.
(d) We have $j(\lambda) = j(\mu)$.

Proof. For this, we need the following two observations:

(1) Let $E \subset \mathbb{P}^2_k$ be a smooth cubic curve, and $O, O' \in E$ be two inflection points. Then the elliptic curves (E, O) and (E, O') are strongly isomorphic. For a proof, consider the line $L_{O,O'}$, and let the third point of its intersection with E be O''; then $O'' \in E$ is also an inflection point by Exercise 2.5.5(b), and $O'' \neq O, O'$ if $O \neq O'$. Put (E, O'') in Weierstrass normal form; then since O and O' are collinear with O'', we have on this elliptic curve that $O + O' = 0$. It follows that the projective change of coordinates $Y \mapsto -Y$ is the required strong isomorphism between (E, O) and (E, O').
(2) If $E \subset \mathbb{P}^2_k$ is a smooth cubic curve and $O \in E$ an inflection point, then except for the flex tangent there are three other tangents to E that pass through O, and the points of contact of these three lines with E are collinear. This is immediate by putting (E, O) in Weierstrass or Legendre form.

We are now ready to proceed to the main proof.

(a) $\Leftrightarrow$ (b) Strongly isomorphic elliptic curves are weakly isomorphic by definition. follows from (1). Indeed, if $\Phi : \mathbb{P}^2_k \to \mathbb{P}^2_k$ is a change of coordinates such that $\Phi(E_\lambda) = E_\mu$, then $\Phi(O) \in E_\mu$ is an inflection point. By (1), there is a strong isomorphism Ψ taking $(E_\mu, \Phi(O))$ to (E_μ, O). Then $\Psi \circ \Phi$ is a strong isomorphism taking (E_λ, O) and (E_μ, O).

(a) $\Leftrightarrow$ (c) For one direction, let Φ be the strong isomorphism taking (E_λ, O) to (E_μ, O). Then Φ takes $\mathrm{T}_O E_\lambda$ to $\mathrm{T}_O E_\mu$, i.e. preserves the line $Z = 0$ as a set (although not necessarily pointwise). By (2), Φ must take the set the points $\{(0,0), (1,0), (\lambda, 0)\}$ to $\{(0,0), (1,0), (\mu, 0)\}$. In particular, Φ must fix the line $Y = 0$ as well, and hence the point $[1:0:0]$. This combined with the fact that $\Phi(O) = O$ implies that Φ must be of the form $[X:Y:Z] \mapsto [sX + tZ : Y : Z]$ for some $s, t \in k$ with $s \neq 0$ (check!). In particular, the automorphism $x \mapsto sx + t$ takes the set $\{0, 1, \lambda\}$ to $\{0, 1, \mu\}$. In particular, $t \in \{0, 1, \mu\}$, and for each choice of t, we are left with two possibilities for s; these correspond to the six choices for μ above. Conversely, the same argument shows that when $\mu \in M_\lambda$, a transformation of this sort gives us the required strong isomorphism.

(c) $\Leftrightarrow$ (d) This follows from the identity

$$j(\lambda) - j(\mu) = 2^8 \frac{(\lambda - \mu)(\lambda\mu - 1)(\lambda + \mu - 1)(\lambda\mu - \mu + 1)(\lambda\mu - \lambda + 1)(\lambda\mu - \lambda - \mu)}{\lambda^2(\lambda - 1)^2\mu^2(\mu - 1)^2}.$$

■

This key theorem allows us to define a crucial invariant for smooth cubic curves: the j-invariant.

Definition 1.17.5. Let $E \subset \mathbb{P}^2_k$ be a smooth cubic curve. Define the j-invariant of E as follows: pick a projective change of coordinates $\Phi : \mathbb{P}^2_k \to \mathbb{P}^2_k$ such that $\Phi(E)$ is in Legendre form, say $\Phi(E) = E_\lambda$, and define

$$j(E) := j(\lambda).$$

That this is well-defined follows from Theorem 1.17.4. From this we finally arrive at the required classification theorem for all cubic curves.

Corollary 1.17.6. Up to projective changes of coordinates, a cubic curve in $\mathbb{P}^2_k$ is of exactly one of the following seven types.

(a) The union of three concurrent lines.
(b) The union of three nonconcurrent lines.
(c) The union of a smooth conic and a line tangent to it.
(d) The union of a smooth conic and a line transverse to it (i.e. meeting it in two distinct points).
(e) A nodal cubic curve.
(f) A cuspidal cubic curve.
(g) A smooth cubic curve, i.e. after choosing a basepoint, an elliptic curve.

Further:

(a) The types (a) - (d) correspond to the reducible cubics, and the types (e) - (g) to the irreducible cubics. Of these, all curves of types (a) - (f) are singular.
(b) Any two curves of the same type from (a) - (f) are projectively isomorphic.
(c) Two smooth cubic curves are projectively isomorphic iff they have the same j-invariant. Further, given any specified $\alpha \in k$, there is a smooth cubic $E \subset \mathbb{P}^2_k$ with j-invariant α.

In particular, the j-invariant is a complete isomorphism invariant of smooth cubic curves, and can take any value in k.

Proof. The case of the reducible cubics is easy and left to the reader and the case of the irreducible but singular cubics was handled in Exercise 2.4.4, so we'll take the classification as well as statements (a) and (b) as proven. For (c), it is firstly clear that the j-invariant of smooth cubic is a projective isomorphism invariant; this is the content of Theorem 1.17.4. Conversely, suppose that $E, E' \subset \mathbb{P}^2_k$ are two smooth cubic curves with $j(E) = j(E')$. By the discussion in §1.16.1, there are $\lambda, \mu \in k \smallsetminus \{0, 1\}$ and projective isomorphisms $E \cong E_\lambda$ and $E' \cong E_\mu$. It follows then from the first part that

$$j(\lambda) = j(E_\lambda) = j(E) = j(E') = j(E_\mu) = j(\mu),$$

so from Theorem 1.17.4 we conclude that $E_\lambda \cong E_\mu$. It then follows that

$$E \cong E_\lambda \cong E_\mu \cong E'$$

as needed. Finally, given an $\alpha \in k$, solve the equation

$$2^8(\lambda^2 - \lambda + 1)^8 - \alpha\lambda^2(\lambda - 1)^2 = 0$$

to get a $\lambda \in k \smallsetminus \{0, 1\}$ (using $k = \overline{k}$ and $\operatorname{ch} k \neq 2$); then $E_\lambda \subset \mathbb{P}^2_k$ is a smooth cubic with j-invariant α. ∎

Remark 1.17.7. In a more advanced perspective on the theory of elliptic curves, it is seen that elliptic curves are $2:1$ covers of $\mathbb{P}^1_k$ branched over four points, and the location of the 4 points in $\mathbb{P}^1_k$ (up to projective changes) determines the isomorphism type of corresponding elliptic curve. In the above set-up (i.e. when E is in say Legendre form $y^2 = x(x-1)(x-\lambda)$), this map to $\mathbb{P}^1_k$ is given by taking the x-coordinate; for most values of x, there are two values of y, i.e. two points in E, mapping to it–except for the values $x = 0, 1, \infty, \lambda$. Now given an ordered quadruple of points (a, b, c, d) of $\mathbb{P}^1_k$, we can associate to them a quantity–the cross ratio–which is invariant under coordinate changes; however, permuting the 4 points gives rise to up to six different numbers, each of which is an equal candidate for the title of the cross ratio of an unordered quadruple of points. To systematize this, we can note that any four points on $\mathbb{P}^1_k$ can be brought via a projective change of coordinates into a tuple of the form $(0, 1, \infty, \lambda)$ for some $\lambda \in k \smallsetminus \{0, 1\} = \mathbb{P}^1_k \smallsetminus \{0, 1, \infty\}$–and indeed, this λ then *is* the cross-ratio. The set M_λ is the set of values $\mu \in \mathbb{P}^1_k \smallsetminus \{0, 1, \infty\}$ such that the quadruple $\{0, 1, \infty, \mu\}$ has the same cross-ratio as that of $\{0, 1, \infty, \lambda\}$ when taken in some order, and as the proof of Theorem 1.17.4 shows, the j-function captures precisely this set, and provides a true invariant (under coordinate changes) of unordered quadruples of points on $\mathbb{P}^1_k$. In more grown-up terminology, there is an S_3 action on $\mathbb{P}^1_k$, the orbit of a fixed $\lambda \in \mathbb{P}^1_k \smallsetminus \{0, 1, \infty\}$ under this action is exactly M_λ, and the j function $j : \mathbb{P}^1_k \to \mathbb{P}^1_k$ is a rational function of degree 6 that exhibits $\mathbb{P}^1_k$ as the quotient $\mathbb{P}^1_k/S_3$. (The factor of 2^8 is there for further normalization purposes; I will not explain here what that means.)

Remark 1.17.8. The most interesting values of λ are the ones for which the set M_λ has fewer than 6 elements; these correspond to elliptic curves E_λ with additional symmetries. A little computation shows that there are (in $\operatorname{ch} k \neq 2, 3$) exactly 5 such values corresponding to

$$M_{-1} = M_2 = M_{1/2} = \{-1, 1/2, 2\} \text{ and } M_\rho = M_{\overline{\rho}} = \{\rho, \overline{\rho}\},$$

where ρ is a primitive sixth root of unity, i.e. $\rho^2 - \rho + 1 = 0$, and $\overline{\rho} = 1 - \rho$. In the former case, we are looking at the elliptic curve $y^2 = x^3 - x$ which has $j = 1728$ and has a $\mathbb{Z}/4$-symmetry $(x, y) \mapsto (-x, \mathrm{i}y)$; in the latter case, we are looking at the curve $y^2 = x^3 - 1$ which has $j = 0$ and the $\mathbb{Z}/6$-symmetry $(x, y) \mapsto (\rho^2 x, \rho^3 y)$. This is (part of) the reason for the specialness of the values $j = 0$ and $j = 1728$ in the theory of elliptic curves. (Surprisingly, for $\operatorname{ch} k = 2, 3$ curves with $j = 0 = 1728$, we have automorphism groups $\mathrm{SL}_2 \mathbb{F}_3$ and $\mathbb{Z}/3 \rtimes \mathbb{Z}/4$ of sizes 24 and 12 respectively; see [9, Theorem 10.1].)

There is, of course, much more to say about elliptic curves, but that is all we have time for, because I want to spend some time narrating some stories.

1.17.2 Story Time

The theory of plane algebraic curves is a classical yet very rich subject which is both the starting point of several deep stories (the theory of abstract curves, algebraic geometry in general, and elliptic curves to name a few) and the source of still many unsolved problems and not-as-well-understood phenomenen (e.g. the moduli of curves). In this section, I want to end the course by mentioning some of the directions the study of curves can take from here.

- For starters, there is much, much more to say about elliptic curves. For a good introduction you can start reading now, see [10]; for a more advanced perspective, the classical textbook is [9]. Here are three facts I want to mention:
  (a) Over the field $k = \mathbb{C}$, an elliptic curve $E \subset \mathbb{P}^2_{\mathbb{C}}$ is "the same" as a complex torus, i.e. $\mathbb{C}/\Lambda$ for some lattice $\Lambda \subset \mathbb{C}$, relating the theory of plane cubics to doubly periodic meromorphic functions in the plane (which is–via the theory of elliptic integrals–ultimately the source of the nomenclature "elliptic", since it is otherwise not so clear what these curves have to do with ellipses). The addition law on E is then

induced from the usual addition law on $\mathbb{C}$ (or more precisely on the quotient group $\mathbb{C}/\Lambda$); this perspective makes abundantly clear why for each $n \geq 1$ we can expect $E[n] \cong \mathbb{Z}/n \times \mathbb{Z}/n$–these correspond exactly to the n^2 points in the fundamental parallelogram of Λ corresponding to $(1/n)\Lambda$. Further, if for each τ in the upper half plane $\mathbb{H}$, we let E_τ be the elliptic curve corresponding to the lattice $\mathbb{Z} \oplus \mathbb{Z}\tau$, then the function $\tau \mapsto j(E_\tau)$ is a holomorphic function $\mathbb{H} \to \mathbb{C}$ which very beautiful properties (it is invariant under the action of the modular group $\mathrm{PSL}_2 \mathbb{Z}$, and intimately related to the theory of modular forms etc.).

(b) Elliptic curves over finite fields $k = \mathbb{F}_q$ are both theoretically important and the key to a lot of modern day cryptography. For instance, the Hasse bound says that if $E \subset \mathbb{P}^2$ is an elliptic curve defined over $\mathbb{F}_q$, then the number $\#E(\mathbb{F}_q)$ of $\mathbb{F}_q$-points on E is bound by

$$|\#E(\mathbb{F}_q) - (q+1)| \leq 2\sqrt{q}.$$

Such counts of points an elliptic curve for varying $q = p^n$ are the start of another beautiful story–that of the Weil conjectures.

(c) Finally, over number fields such as $k = \mathbb{Q}$ the theory is still fascinating and at times mysterious. The famous Mordell-Weil Theorem asserts that if k is a number field and E an elliptic curve defined over k, then the group of k-rational points $E(k)$ is a finitely generated abelian group. In particular, it is isomorphic to $\mathbb{Z}^r \oplus T$ for some unique integer $r \geq 0$ and finite abelian group T (the torsion subgroup)–this r is called the algebraic rank of $E(k)$. A theorem of Barry Mazur (a professor of mine!) asserts that when $k = \mathbb{Q}$, the torsion subgroup T can be only one of 15 types–it can be either $\mathbb{Z}/n$ for $n = 1, \ldots, 10$ or $n = 12$, or it can be $\mathbb{Z}/2n \times \mathbb{Z}/2$ for $n = 1, 2, 3, 4$. Much work has been done to extend this result to general number fields k. Finally, the rank r is another fascinating quantity. The largest known rank over $k = \mathbb{Q}$ as of this writing (September 2024) is 29, and this elliptic curve was found by Elkies and Klagsbrun (Elkies was another professor of mine); see [11] for the equation. The rank is part of some very important unsolved conjectures as well: associated to an elliptic curve E over $k = \mathbb{Q}$, we also have a an L-function holomorphic in a neighborhood of $s = 1$; the order of vanishing of this function at $s = 1$ is called the analytic rank of E. The famous Birch and Swinnerton-Dyer conjecture asserts that the analytic rank of an elliptic curve agrees with its algebraic rank–if you show this, you get a million dollars (among other things)!

- Another theme that is still an area of active research that touched upon in Exercise 2.1.2 is real algebraic geometry. This field studies the topology and geometry of algebraic curves (and, more generally, varieties) over the field $k = \mathbb{R}$ of real numbers, which is harder than the case $k = \mathbb{C}$ because $\mathbb{R}$ is not algebraically closed. It is theorem due to Harnack from the 19th century that a real projective algebraic curve of degree $d \geq 1$ has at most $\binom{d-1}{2}+1$ connected components; for a proof sketch, see [12, Lect. 19]. Harnack also showed with this proof that this bound is actually achieved–curves with this maximal number of connected components are called M-curves. The classification of all possible isotopy types (roughly, the nesting type) of M-curves still remains quite mysterious for $d \geq 8$. In his list of 23 mathematical problems presented by Hilbert before the Paris conference of the International Congress of Mathematicians in 1900, the study of real algebraic geometry was in the sixteenth place, and this problem has occupied researchers fruitfully for almost a century with lots still to be explored.
- In our focus on smooth curves, one fascinating area of study we completely missed out on was singularity theory, which studies the singularities of curves (or more generally varieties). For instance, if you look at the nodal curve C defined by $y^2 = x^3 - x^2$ over $k = \mathbb{C}$ and take a small cross-section near the singularity $(0, 0)$ in $\mathbb{A}^2_{\mathbb{C}} \cong \mathbb{C}^2$ by a 3-sphere $S^3 \subset \mathbb{C}^2$, the intersection $C \cap S^3$ is a link in S^3, namely the Hopf link–two circles that don't intersect

but cannot be "pulled apart" because they link exactly once. Similarly, if C were to be the cuspidal cubic $y^2 = x^3$, then the intersection $C \cap S^3$ would be the trefoil knot. Therefore, the resulting link $C \cap S^3$ in S^3 is somehow capturing the nature of the singularity of the corresponding curve C–the more complicated the singularity, the more complicated the corresponding link (this observation was first made by K. Brauner). Knots and links arising in this way are called knots of singularities, and were studied extensively by Milnor, the standard resource on the subject still being [13]. Here's one thing to think about: the figure-8 knot 4_1 does *not* arise as a complex knot singularity. Can you think of how you would prove something like this?

- One more connection I want to mention, already somewhat manifest in our discussion of elliptic curves over $k = \mathbb{C}$ above, is the relationship between algebraic geometry and complex analysis. Classically, these subjects were not considered separate at all, with the main focus being the study of complex algebraic curves. The idea is that if $X \subset \mathbb{P}^2_{\mathbb{C}}$ is a smooth curve, then X is a compact complex manifold of dimension one–a Riemann surface. By a topological classification theorem of orientable closed surfaces, each such X is a g-holed torus for some $g \geq 0$ (think of a the surface of a donut with g holes; the case $g = 0$ is the sphere, and $g = 1$ the standard torus). This integer g–called the genus of the curve X–is directly computable from X and contains a lot of information about it. If $X \subset \mathbb{P}^2_{\mathbb{C}}$ is a smooth curve of degree d, then the genus of X can be shown to be $\binom{d-1}{2}$. One piece of information contained in g is about a natural geometry on X. It turns out that any compact Riemann surface carries in a natural way a metric, which for $g = 0, 1, \geq 2$ corresponds to round, flat, and hyperbolic metrics. The round case corresponds to $g = 0$ being $X \cong \mathbb{P}^1_{\mathbb{C}}$, which is topologically a sphere (the Reimann sphere). The flat case $g = 1$ corresponds to the case of plane cubics–$d = 3$–which, as mentioned above, are naturally isomorphic to Riemann surfaces of the form $\mathbb{C}/\Lambda$ for lattices $\Lambda \subset \mathbb{C}$; the flat metric on the elliptic curve then comes from the Λ-translation-invariant flat metric on $\mathbb{C}$. By far the most interesting and mysterious case is $g \geq 2$, when we have a natural hyperbolic metric on X, i.e. a metric of constant negative curvature. Much work has been done to study the moduli theory of such curves, although a complete understanding is far beyond our means at the moment. Finally, a famous theorem (a generalization of which is due to Chow) asserts conversely that any complex submanifold $X \subset \mathbb{P}^2_{\mathbb{C}}$ is a smooth algebraic curve, so in dimension 1 the theory of compact complex manifolds and smooth algebraic varieties (i.e. curves) are identical (although these theories, importantly, diverge in higher dimensions).
- There's a way to bring a lot of the above discussions together–and, indeed, syntheses of this sort are the biggest triumphs of 20^{th} century algebraic geometry. If $X \subset \mathbb{P}^2_{\mathbb{Q}}$ is a curve defined over $\mathbb{Q}$, then we can rescale the defining equation of X to have only integer coefficients in a minimal manner (i.e. such that the minimal polynomial is primitive as a trivariate polynomial over $\mathbb{Z}$). It then makes sense to talk about not only $X(\mathbb{Q})$, but in fact $X(k)$ for any field k. Under the additional assumption that $X(\mathbb{C})$ is smooth, it turns out that there is a beautiful relationship between the topology of the curves $X(\mathbb{Q}), X(\mathbb{C})$ and $X(\mathbb{F}_q)$ over different q–this is again brought out in detail by the Weil conjectures, which is a(nother beautiful) story for some other time. Here's a different punchline I want to leave you with. The Mordell-Weil theorem mentioned above asserts that if $X(\mathbb{C})$ has genus 1, then $X(\mathbb{Q})$ is a finitely generated abelian group; this result was shown by Mordell already. Based on this result (and additional considerations), Mordell conjectured in 1922 that if $X(\mathbb{C})$ has genus $g \geq 2$, then, in fact, $X(\mathbb{Q})$ is finite. This fascinating conjecture remained open for a while until it was proven by Faltings in 1983. Isn't this result simply amazing? Somehow, the "rational part" $X(\mathbb{Q})$ of the complex curve $X(\mathbb{C})$ "sees" the topology of the complex curve and decides accordingly whether it wants to be very infinite ($g = 0$), "somewhat infinite" or finitely generated ($g = 1$), or finite ($g \geq 2$).

This is a good ending point for this course. I hope you enjoyed and that it has made you excited to go and learn more algebraic geometry in the future!

Chapter 2

Exercise Sheets

2.1 Exercise Sheet 1

2.1.1 Numerical and Exploration

Exercise 2.1.1. For an ordered pair (a, b) of rational numbers, consider the polynomial

$$f_{a,b}(x, y) := ax^2 + by^2 - 1 \in \mathbb{Q}[x, y].$$

Let $C(a, b) = C_{f_{a,b}} \subset \mathbb{A}^2_{\mathbb{Q}}$ be the rational affine plane algebraic curve defined by $f_{a,b}$.

(a) Show that $C(2/5, 1/5) = \emptyset$.
(b) Characterize all primes p such that $C(1/p, 1/p) = \emptyset$.
(c) Characterize all pairs (a, b) such that $C(a, b) = \emptyset$.

Exercise 2.1.2.

(a) Play around with graphs of real affine plane algebraic curves (RAPACs) on, say, Desmos or WolframAlpha. What is the coolest thing you can get a graph to do (cross itself thrice, look like a heart, etc.)?
(b) How many pieces (i.e. connected components) can a RAPAC of degree $d = 2$ have? How about $d = 3$? What about $d \in \{4, 5, 6, 7\}$?
(c) What can you say in general? Can you come up with upper or lower bounds for the number of pieces?
(d) Does the number of pieces depend on the nesting relations[1] between them? Does it depend on (or dictate) their shapes (e.g. convexity)?[2]

Exercise 2.1.3.

(a) Let $P \subset \mathbb{A}^2_{\mathbb{R}}$ be the polar curve implicitly defined by the equation

$$r^3 + r\cos\theta - \sin 4\theta = 0.$$

Find a nonconstant polynomial $f(x, y) \in \mathbb{R}[x, y]$ such that the curve $C_f \subset \mathbb{A}^2_{\mathbb{R}}$ defined by f contains P, i.e. satisfies $P \subset C_f$.[3]
(b) What is the degree of your f? What is the smallest possible degree of such an f?
(c) By your choice of f, we have the containment $P \subset C_f$. Is P all of C_f? If so, can you explain why (perhaps by retracing steps)? If not, how would you describe the extraneous components of $C_f \smallsetminus P$? Could you have predicted them? Can you pick an f that provably minimizes the number of extraneous components?
(d) Repeat the same analysis as in (a) through (c) for other such implicitly defined polar curves of your own devising.
(e) Can you perform the same analysis as above for the Archimedean spiral, which is the polar curve implicitly defined by the equation $r = \theta$?

Draw pictures, or get a computer to draw them for you, but beware–is your software doing exactly what you think it is?

[1] What does that mean? What are those?

[2] Here's a harder result to whet your appetite: if $d = 4$ and there is a nested pair of closed ovals, then the inner oval must be convex and there cannot be more components, although there may be up to 4 non-convex components in general. You may not be able to prove this now, but you should be able to solve this problem by the end of the course.

[3] I like to use the symbol $\subset$ to mean "is contained in or equal to". Others prefer the symbol $\subseteq$ to denote the same thing. I will use the symbol $\subsetneq$ when I want to exclude the possibility of equality.

Exercise 2.1.4. Consider the surface defined by the equation $z^3 + xz - y = 0$, pictured in Figure 2.1. The orthogonal projection of this surface to the xy-plane outlines a cuspidal curve.

(a) Find the equation describing this cuspidal curve, and prove the assertion made above.
(b) How does all of this relate to the Cardano formula for the solution to the cubic equation?

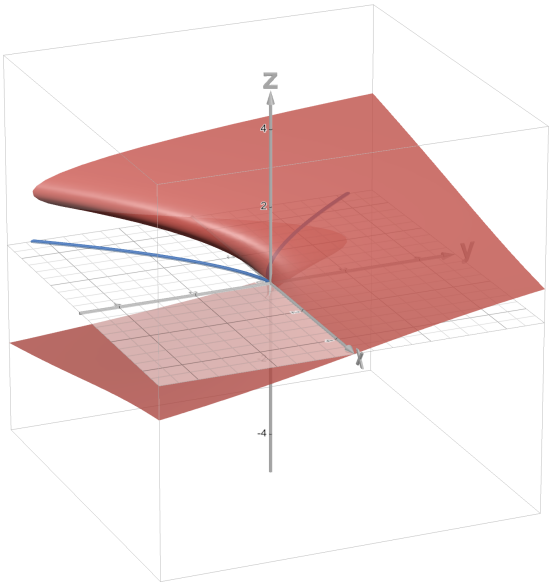


Figure 2.1: The surface $z^3 + xz - y = 0$ when orthogonally projected onto the xy-plane outlines a cuspidal curve. Picture made with Desmos 3D.

Exercise 2.1.5. Can you find a way to use the conchoid of Nichomedes (Example 1.2.14) to trisect a given angle? You may suppose that you know how to construct a conchoid with any given parameters. (Hint: see Figure 2.2.) Once you've done that, use the cissoid of Diocles to give a compass and ruler (and cissoid) construction of $\sqrt[3]{2}$, or of $\sqrt[3]{a}$ for any given $a > 0$. How far can you take this–what else can you do with the cissoid and conchoids of different parameters? Why do these constructions not contradict results from Galois theory you may have seen?

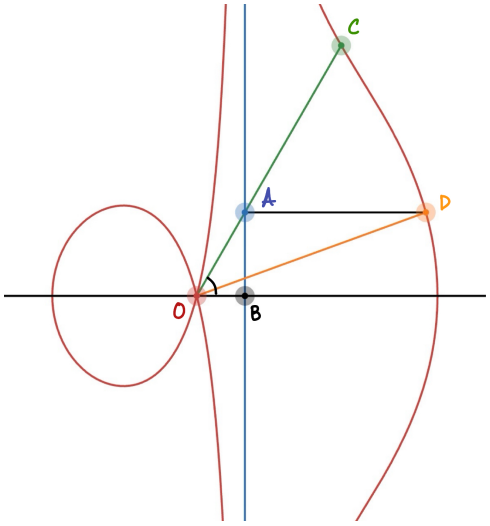


Figure 2.2: The Conchoid of Nichomedes and Angle Trisection. Picture made with Desmos and edited in Notability.

Exercise 2.1.6. Show that over $k = \mathbb{C}$, every affine conic section, i.e. plane curve defined by a quadratic polynomial of the form

$$f(x, y) = ax^2 + 2hxy + by^2 + 2ex + 2fy + c \in \mathbb{C}[x, y]$$

for some $a, b, c, e, f, h \in \mathbb{C}$, not all zero, can be brought by an affine change of coordinates into one and only one of the following forms:

(a) an ellipse/circle/hyperbola defined by $x^2 + y^2 = 1$,
(b) a parabola defined by $y = x^2$, or
(c) a pair of lines defined by $xy = 0$, or
(d) a double line defined by $x^2 = 0$.

Note that the equivalence of the circle $x^2 + y^2 = 1$ and hyperbola $x^2 - y^2 = 1$ in $\mathbb{A}^2_{\mathbb{C}}$ uses that $\mathbb{C}$ contains a square root of -1 (how?). Can you come up with a similar classification over $k = \mathbb{R}$? What about other fields like $k = \mathbb{F}_q$?

2.1.2 PODASIPs

Prove or disprove and salvage if possible the following statements.

Exercise 2.1.7. Let k be a field, $C \subset \mathbb{A}^2_k$ be an algebraic curve, and $\ell \subset \mathbb{A}^2_k$ be a line. Then the intersection $C \cap \ell \subset \mathbb{A}^2_k$ of C and ℓ is finite.

Exercise 2.1.8. Given any field k and function $f : k \to k$, we define its graph to be the subset

$$\Gamma_f := \mathbb{V}(y - f(x)) = \{(x, f(x)) : x \in k\} \subset \mathbb{A}^2_k.$$

(a) When $k = \mathbb{R}$ and $f(x) = \sin x$, the graph $\Gamma_f \subset \mathbb{A}^2_{\mathbb{R}}$ is an algebraic curve.
(b) When $k = \mathbb{R}$ and $f(x) = \mathrm{e}^x$, the graph $\Gamma_f \subset \mathbb{A}^2_{\mathbb{R}}$ is an algebraic curve.
(c) In the setting of (b), every line $\ell \subset \mathbb{A}^2_{\mathbb{R}}$ meets Γ_f in at most two points.
(d) When $k = \mathbb{C}$ and $f(x) = \mathrm{e}^x$, the graph $\Gamma_f \subset \mathbb{A}^2_{\mathbb{C}}$ is an algebraic curve.

[Possible Hints: For (a), see Exercise 2.1.7. For (b), the exponential function grows *very fast*, so that your solution to (a) may not work for (b) thanks to (c). You may either use this growth to your advantage, or you may first solve (d) and use a little bit of complex analysis.]

Exercise 2.1.9 (Apparently Transcendental Curves)**.**

(a) The curve $C_1 \subset \mathbb{A}^2_{\mathbb{R}}$ given parametrically as

$$C_1 = \{(\mathrm{e}^{2t} + \mathrm{e}^t + 1, \mathrm{e}^{3t} - 2) : t \in \mathbb{R}\}$$

is an algebraic curve.
(b) The curve $C_2 \subset \mathbb{A}^2_{\mathbb{R}}$ defined by the vanishing of the function f defined by

$$f(x, y) = x^2 + y^2 + \sin^2(x + y)$$

is an algebraic curve.

These examples are a little silly, but they illustrate important points (what?). Can we improve our definition of a plane algebraic curve to avoid such silliness?

Exercise 2.1.10. Given any $g(r, c, s) \in \mathbb{R}[r, c, s]$, there is a unique polynomial $f(x, y) \in \mathbb{R}[x, y]$ such that the polar algebraic curve P_g implicitly defined by g (see §1.2.2) is contained in the algebraic curve C_f defined by f, i.e. satisfies $P_g \subset C_f$.

2.2 Exercise Sheet 2

2.2.1 Numerical and Exploration

Exercise 2.2.1. Show that if k is any field of characteristic zero (e.g. $k = \mathbb{R}$ or $k = \mathbb{C}$), then the affine curve $C = C_f \subset \mathbb{A}^2_k$ defined by the vanishing of the polynomial

$$f(x, y) = y^2 - x^3 + x \in k[x, y]$$

cannot be parametrized by rational functions, using the following proof outline.

(a) Suppose to the contrary that it can, and use this to produce polynomials $f, g, h \in k[t]$ that satisfy all of the following properties simultaneously:

(i) $h \neq 0$ and not all of f, g, h are constant,

(ii) the polynomials f, g, h are coprime as a triple, i.e. that $(f, g, h) = (1)$ in $k[t]$, and

(iii) $g^2h - f^3 + fh^2 = 0$.

(b) Verify the following matrix identities over the ring $k[t]$ (or equivalently field $K = k(t)$):

$$\begin{bmatrix} f & g & h \\ f' & g' & h' \end{bmatrix} \cdot \begin{bmatrix} -3f^2 + h^2 \\ 2gh \\ g^2 + 2fh \end{bmatrix} = \begin{bmatrix} f & g & h \\ f' & g' & h' \end{bmatrix} \cdot \begin{bmatrix} gh' - hg' \\ hf' - fh' \\ fg' - gf' \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Here f' denotes the formal derivative[4] of f with respect to t, and similarly for g' and h'.

(c) Show that the 2×3 matrix

$$\begin{bmatrix} f & g & h \\ f' & g' & h' \end{bmatrix}$$

has full rank, i.e. that at least one of $gh' - hg', hf' - fh', fg' - gf' \in k[t]$ is nonzero. (Hint: Exercise 2.2.11(a).)

(d) Use (b), (c), and basic linear algebra over the field $K = k(t)$ to conclude that there are relatively prime polynomials $p(t), q(t) \in k[t]$ with $q(t) \neq 0$ satisfying

$$q(t) \cdot \begin{bmatrix} -3f^2 + h^2 \\ 2gh \\ g^2 + 2fh \end{bmatrix} = p(t) \cdot \begin{bmatrix} gh' - hg' \\ hf' - fh' \\ fg' - gf' \end{bmatrix}. \tag{2.1}$$

(e) Show that the polynomials $-3f^2 + h^2, 2gh, g^2 + 2fh \in k[t]$ are coprime as a triple, i.e. in $k[t]$, we have that

$$(-3f^2 + h^2, 2gh, g^2 + 2fh) = (1).$$

Conclude that $p(t)$ is a nonzero constant.

(f) Use the equation (a)(iii) and the matrix equation (2.1) to derive a contradiction. (Hint: do some case-work on the possible relationships between the degrees of f, g and h.)

(g) Why do the polynomials $-3f^2 + h^2, 2gh$ and $g^2 + 2fh$ show up in this proof? What goes wrong in the above proof if you try to repeat it for $f(x, y) = y^2 - x^3 - x^2 \in k[x, y]$ instead? (We showed in Example 1.3.7 that this curve admits a rational parametrization.)

(h) Where in the proof did you use $\operatorname{ch} k = 0$? Investigate what happens in positive characteristic. Is the result still true? If not, can you come up with a parametrization? If yes, then does the same proof work? If the result is true but the proof doesn't work, can you come up with a different proof?

[4] If you haven't seen this notion before, then define it.

This proof due to Kapferer has been adapted from [14]; with minor modifications, the same proof shows that any over a field k with $\operatorname{ch} k = 0$, any smooth projective curve of degree at least 3 cannot be parametrized by rational functions. For a different proof of this specific case using Fermat's method of infinite descent, see [5, §I.2.2]. In modern algebraic geometry, the more general result (in arbitrary characteristic) is often seen as a consequence of the Riemann-Hurwitz formula.

Exercise 2.2.2. Let $C_e \subset \mathbb{A}^2_{\mathbb{R}}$ denote the Cassini curve of eccentricity $e \in (0, \infty)$ (see Example 1.2.12). For concreteness, you may take $C_e := C_{f_e}$, where

$$f_e(x, y) := \left((x-1)^2 + y^2\right)\left((x+1)^2 + y^2\right) - e^4 \in \mathbb{R}[x, y].$$

Show that:

(a) The curve C_e consists of two pieces[5] if $0 < e < 1$ and one piece if $e \geq 1$.
(b) The curve C_e is smooth[6] if and only if $e \neq 1$.
(c) For $e > 1$, the unique oval in C_e is convex[7] iff $e \geq \sqrt{2}$.

Exercise 2.2.3 (More Parametric Curves)**.** Using the proof strategy from Example 1.3.10 and Remark 1.3.11 or otherwise, come up with Cartesian equations defining the parametric curves given by the following parametrizations.

(a) $(t^4 + 2t - 3, t^3 + 2t^2 - 5)$
(b) $\left(\dfrac{t(t^2+1)}{t^4+1}, \dfrac{t(t^2-1)}{t^4+1}\right)$

Now come up with a few examples of your own devising, and repeat the same. Can you write a program that does these (somewhat tedious) calculations for you?

Exercise 2.2.4 (Resultants)**.** For those who know a little linear algebra, this exercise provides a different perspective on the resultant of two polynomials than is presented in the Ross set on this topic (which you should now solve if you haven't done so previously!).

For a field K and for each integer $N \geq 0$, let $K[t]_N \subset K[t]$ denote the subspace of polynomials of degree strictly less than N, so that $\dim_K K[t]_N = N$. Given polynomials $f, g \in K[t]$ of degree $m, n \geq 0$ respectively, we can investigate whether or not f and g have a common factor in $K[t]$ as follows.

(a) Consider the linear map $\phi : K[t]_n \times K[t]_m \to K[t]_{m+n}$ given by $\phi(u, v) := uf + vg$. Show that f and g have a common factor in $K[t]$ of positive degree iff the map ϕ is not injective. (Hint: use that $K[t]$ is a UFD.)
(b) Show that if we choose the ordered basis

$$(t^{n-1}, 0), (t^{n-2}, 0), \ldots, (1, 0), (0, t^{m-1}), (0, t^{m-2}), \ldots, (0, 1)$$

of the domain and

$$t^{m+n-1}, t^{m+n-2}, \ldots, 1$$

[5] Here the word "piece" means "connected component".
[6] What does that mean?
[7] What does that mean?

of the range, then the matrix representative of ϕ with respect to these bases is

$$\mathrm{Syl}(f,g) := \begin{bmatrix} a_0 & 0 & \cdots & 0 & b_0 & 0 & \cdots & 0 \\ a_1 & a_0 & \cdots & 0 & b_1 & b_0 & \cdots & 0 \\ a_2 & a_1 & \ddots & 0 & b_2 & b_1 & \ddots & 0 \\ \vdots & \vdots & \ddots & a_0 & \vdots & \vdots & \ddots & b_0 \\ a_m & a_{m-1} & \cdots & \vdots & b_n & b_{n-1} & \cdots & \vdots \\ 0 & a_m & \ddots & \vdots & 0 & b_n & \ddots & \vdots \\ \vdots & \vdots & \ddots & a_{m-1} & \vdots & \vdots & \ddots & b_{n-1} \\ 0 & 0 & \cdots & a_m & 0 & 0 & \cdots & b_n \end{bmatrix},$$

where $f(x) = a_0t^m + \cdots + a_m$ and $g(x) = b_0t^n + \cdots + b_n$. This matrix is called the Sylvester matrix of f and g.

(c) The determinant of the Sylvester matrix of f and g is called the called the resultant of f and g with respect to t, often written $\mathrm{Res}_t(f,g)$ or simply $\mathrm{Res}(f,g)$, so that

$$\mathrm{Res}(f,g) := \det \mathrm{Syl}(f,g) \in \mathbb{Z}[a_0,\ldots,a_m,b_0,\ldots,b_n] \subset K.$$

Show, using some basic linear algebra, that f and g share a common factor in $K[t]$ iff

$$\mathrm{Res}(f,g) = 0 \in K.$$

(Hint: the domain and range of ϕ have the same dimension over K.)

(d) Conclude that if K is algebraically closed and $a_0b_0 \neq 0$, then f and g have a common root $t = t_0 \in K$ iff

$$\mathrm{Res}(f,g) = 0.$$

(What happens if $a_0b_0 = 0$?) Use this to show that, even if K is not algebraically closed, and $\alpha_1,\ldots,\alpha_m$ and $\beta_1,\ldots,\beta_n$ are roots of f and g, respectively, in some extension field $K' \supset K$ of K, then

$$\mathrm{Res}(f,g) = a_0^n b_0^m \prod_{i=1}^{m}\prod_{j=1}^{n}(\alpha_i - \beta_j) = a_0^n \prod_{i=1}^{m} g(\alpha_i) = (-1)^{mn} b_0^m \prod_{j=1}^{n} f(\beta_j).$$

(e) Let's do one example computation: show that if $m = n = 2$ and

$$f(t) = a_1t^2 + b_1t + c_1 \text{ and}$$
$$g(t) = a_2t^2 + b_2t + c_2,$$

then

$$\mathrm{Res}(f,g) = (a_1c_2 - a_2c_1)^2 - (a_1b_2 - a_2b_1)(b_1c_2 - b_2c_1).$$

In particular, these quadratic equations have a common root (in K, or if necessary, a quadratic extension of K) iff this polynomial of degree 4 in the coefficients vanishes.

(f) (Finishing Example 1.3.10.) Show that if $u(t), v(t) \in k[t]$ are any nonconstant polynomials which define the parametric curve

$$C = \{(u(t), v(t)) : t \in k\} \subset \mathbb{A}_k^2$$

and if

$$f(x,y) := \mathrm{Res}_t(u(t) - x, v(t) - y) \in k[x,y],$$

then $C \subset C_f$ with equality if k is algebraically closed.

Exercise 2.2.5 (Discriminants). Given a field K and a polynomial $f(t) \in K[t]$, the discriminant of f, written $\operatorname{disc}(f)$, is the resultant of f and its (formal) derivative f' with respect to t, up to scalar factors. More precisely, if $f(t) = a_0 t^m + \cdots + a_m$ with $a_j \in K$ and $a_0 \neq 0$, then we define

$$\operatorname{disc}(f) := \frac{(-1)^{m(m-1)/2}}{a_0} \cdot \operatorname{Res}(f, f').$$

Let's do a few examples.

(a) Show that if $f(t) = at^2 + bt + c$, with $a \neq 0$, then $\operatorname{disc}(f) = b^2 - 4ac$.

(b) Show that if $f(t) = t^3 + pt + q$, then $\operatorname{disc}(f) = -4p^3 - 27q^2$. How does this relate to Exercise 2.1.4?

(c) Show that if over an extension field $K' \supset K$, the polynomial f splits into linear factors as

$$f(t) = a_0 \prod_{i=1}^{m} (t - \alpha_i) \in K'[t]$$

for some $\alpha_i \in K'$, then

$$\operatorname{disc}(f) = a_0^{2n-2} \prod_{1 \leq i < j \leq n} (\alpha_i - \alpha_j)^2.$$

(d) Show that the polynomial $f(t)$ has a repeated root over an algebraic closure of K iff $\operatorname{disc}(f) = 0$. In other words, if there is an α some extension field $K' \supset K$ and a polynomial $q(t) \in K'[t]$ such that

$$f(t) = (t - \alpha)^2 q(t),$$

then $\operatorname{disc}(f) = 0$, and conversely, if $\operatorname{disc}(f) = 0$, then we can find such α, K and q.

2.2.2 PODASIPs

Prove or disprove and salvage if possible the following statements.

Exercise 2.2.6. For a field k, let $\operatorname{Fun}(\mathbb{A}^2_k, k)$ be the set of all functions $F : \mathbb{A}^2_k \to k$. Claim: for any field k, the map

$$k[x, y] \to \operatorname{Fun}(\mathbb{A}^2_k, k), \quad f \mapsto F_f$$

which sends a polynomial to the corresponding polynomial function is injective. In other words, if two polynomials $f, g \in k[x, y]$ agree at all points $(p, q) \in \mathbb{A}^2_k$, then $f = g$.

Exercise 2.2.7. If k is any infinite field and $C \subset \mathbb{A}^2_k$ an algebraic curve, then the complement

$$\mathbb{A}^2_k \smallsetminus C$$

of C in $\mathbb{A}^2_k$ is infinite.

Exercise 2.2.8. A field is algebraically closed if and only if it is infinite.

Exercise 2.2.9. For any field k, if $f, g \in k[t]$ are polynomials such that

$$f(t)^2 + g(t)^2 = 1$$

as polynomials, then $f(t)$ and $g(t)$ are constant. In other words, the "unit circle" $C \subset \mathbb{A}^2_k$ does not admit a polynomial parametrization.

Exercise 2.2.10 (Separability). For any field K and polynomial $f(t) \in K[t]$, we say that f is separable if an algebraic closure of K separates the roots of f, i.e. that $\operatorname{disc}(f) \neq 0 \in K$. (See Exercise 2.2.5.) Claim: for any field K and $f(t) \in K[t]$, the polynomial f is separable if and only if it is irreducible as an element of the ring $K[t]$.

Exercise 2.2.11 (Wronskians).

(a) For any field k and polynomials $f, g \in k[t]$ in one variable t over k, we have $fg' = gf'$ iff there are $\alpha, \beta \in k$, not both zero, such that $\alpha f + \beta g = 0$. Here, as before, f' (resp. g') denotes the formal derivative of f (resp. g) with respect to t.

(b) More generally, for any field k, integer $n \geq 1$, and polynomials $f_1, \ldots, f_n \in k[t]$ in one variable t over k, the determinant

$$W(f_1, \ldots, f_n) = \det \begin{bmatrix} f_1 & f_2 & \cdots & f_n \\ f_1' & f_2' & \cdots & f_n' \\ \vdots & \vdots & \ddots & \vdots \\ f_1^{(n-1)} & f_2^{(n-1)} & \cdots & f_n^{(n-1)} \end{bmatrix} \in k[t]$$

vanishes (i.e. we have $W(f_1, \ldots, f_n) = 0$ as a polynomial) iff the $f_1, \ldots, f_n \in k$ are linearly dependent, i.e. there are $\alpha_1, \ldots, \alpha_n \in k$, not all zero, such that

$$\alpha_1 f_1 + \alpha_2 f_2 + \cdots + \alpha_n f_n = 0.$$

Here, for any $f \in k[t]$ and $j \geq 0$, the symbol $f^{(j)}$ denotes the j^{th} formal derivative of f with respect to t, so that $f^{(0)} = f$ and we have $f^{(1)} = f'$, $f^{(2)} = f''$, etc.

2.3 Exercise Sheet 3

2.3.1 Standard Exercises

Exercise 2.3.1 (Eisenstein's Irreducibility Criterion)**.**

(a) Let R be a domain, let $f \in R[t]$ have degree $n \geq 1$, and write $f = a_0t^n + a_1t^{n-1} + \cdots + a_n$ for $a_0, \ldots, a_n \in R$ with $a_0 \neq 0$. Show that if there is a prime ideal $P \subset R$ such that
   (i) $a_0 \notin P$,
   (ii) for each j with $1 \leq j \leq n$ we have $a_j \in P$, and
   (iii) $a_n \notin P^2$,
   then f cannot be written as a product of two nonconstant polynomials in $R[t]$. In particular, if f is primitive, the existence of such a P implies that f irreducible.

(b) Show that for each integer $r \geq 1$ and integer prime $p > 0$, the prime-power cyclotomic polynomial

$$\Phi_{p^r}(t) := \frac{t^{p^r} - 1}{t^{p^{r-1}} - 1} = \sum_{j=0}^{p-1} t^{p^{r-1}j} \in \mathbb{Z}[t]$$

is irreducible. (Hint: an $f(t) \in R[t]$ is irreducible iff for some $a \in R$, the shift $f(t+a)$ is.)

(c) Show that the polynomial $f(x, y) = x^2 + y^2 - 1 \in \mathbb{Q}[x, y]$ is irreducible.

(d) Show that if k is any field, then the polynomial $f(x, y) = y^2 - x^3 + x \in k[x, y]$ is irreducible.

(e) Given a field k, an integer $n \geq 1$, and a polynomial $p(x) \in k[x]$ of x alone, can you come up with a criterion for the irreducibility of the polynomial $f(x, y) := y^n - p(x) \in k[x, y]$?

Exercise 2.3.2.

(a) Show that if R is any integral domain, then any prime element of R is irreducible.

(b) Show that if R is a UFD, then any irreducible element of R is prime as well, so that the terms "prime" and "irreducible" mean the same thing in UFDs.

(c) Show that the ring $R := \mathbb{C}[x, y, z]/(z^2 - xy)$ is an integral domain and that the class of z in R is an irreducible element that is not prime. Conclude that R is not a UFD.

Exercise 2.3.3. Show that if k is an algebraically closed field and $\mathfrak{p} \subset k[x, y]$ is a prime ideal, then one and exactly one of the following holds:

(a) $\mathfrak{p} = (0)$;

(b) there is an irreducible $f \in k[x, y]$ such that $\mathfrak{p} = (f)$;

(c) there are $p, q \in k$ such that $\mathfrak{p} = (x - p, y - q)$.

Compare with your knowledge of the prime ideals of $\mathbb{Z}[x]$. Can you prove an analogous result for prime ideals of $R[t]$ for any PID R? Also, what happens if k is not algebraically closed?

Exercise 2.3.4. Let k be an algebraically closed field, and $C \subset \mathbb{A}^2_k$ be a curve of degree $n \geq 2$.

(a) Show that if $P \in C$ is such that $m_P(C) = n$, then C is a union of n lines through P.

(b) Conclude that if C is irreducible, then for any point $P \in C$, the multiplicity of C at P satisfies

$$1 \leq m_P(C) \leq n - 1.$$

In particular, any irreducible conic $C \subset \mathbb{A}^2_k$ is smooth.

(c) Show that if C is irreducible and if some $P \in C$ has multiplicity $m_P(C) = n - 1$, then C admits a rational parametrization.

Finally,

(d) For each $n \geq 2$ and integer j with $1 \leq j \leq n - 1$, find an irreducible curve $C \subset \mathbb{A}^2_k$ and a point $P \in C$ such that $m_P(C) = j$.

Exercise 2.3.5. (Taken from [3, Problems 3.22-23].) Let k be an algebraically closed field, $C = C_f \subset \mathbb{A}^2_k$ be a curve, and $P \in C$.

(a) Suppose that $m_P(C) \geq 2$ and that C has a unique tangent line C_ℓ at P. Show that $i_P(f, \ell) \geq m_p(C)+1$. The curve C is said to have an ordinary hypercusp of order $n := m_p(C)$ at P if equality holds; a hypercusp of order $n = 2$ is called simply a cusp.
(b) Suppose we pick coordinates so that $P = (0, 0)$ and $\ell = y$. Show that if $\operatorname{ch} k \neq 2, 3$, then P is a cusp iff $\partial^3 f/\partial x^3|_P \neq 0$. Use this to give examples. What happens if $\operatorname{ch} k \in \{2, 3\}$?
(c) Show that if P is a cusp of C, then there is only one component of C through P.
(d) Generalize (b) and (c) to the case of hypercusps.

2.3.2 Numerical and Exploration

Exercise 2.3.6. (Adapted from [3, Problem 3.2].) Suppose $k = \mathbb{C}$. Find the multiple points, and the tangent lines at the multiple points, for each of the following curves:

(a) $y^3 - y^2 + x^3 - x^2 + 3xy^2 + 3x^2y + 2xy$,
(b) $x^3 + y^3 - 3x^2 - 3y^2 + 3xy + 1$,
(c) $(x^2 + y^2 - 3x)^2 - 4x^2(2 - x)$, and
(d) $(x^2 + y^2 - 1)^m + x^n y^n$ for $m, n \geq 1$.

Be sure to draw (or get a computer to draw) tons of pictures! Which of you answers change in positive characteristic, and what are the answers there?

Exercise 2.3.7. Let $k = \mathbb{C}$ and $P = (0, 0)$. Consider the affine plane curves C_i containing P defined by the polynomials f_i for $1 \leq i \leq 7$ below:

(i) $f_1 = y - x^2$,
(ii) $f_2 = y^2 - x^3 + x$,
(iii) $f_3 = y^2 - x^3$,
(iv) $f_4 = y^2 - x^3 - x^2$,
(v) $f_5 = (x^2 + y^2)^3 + 3x^2y - y^3$,
(vi) $f_6 = (x^2 + y^2)^3 - 4x^2y^2$, and
(vii) $f_7 = (x^2 + y^2 - 3x)^2 - 4x^2(2 - x)$.

For each pair of integers i, j with $1 \leq i < j \leq 7$, compute the local intersection multiplicity $i_P(f_i, f_j)$ of C_i and C_j at P. What patterns do you observe? Make some conjectures.

Exercise 2.3.8. Over a field $k = \overline{k}$, how many singular points can a curve $C \subset \mathbb{A}^2_k$ of degree $n \geq 1$ have? Come up with an upper bound and a conjecture for when it is achieved.

2.3.3 PODASIPs

Prove or disprove and salvage if possible the following statements.

Exercise 2.3.9. A cubic curve $C \subset \mathbb{A}^2_k$ over a field k can have at most one singular point.

Exercise 2.3.10. Given a field k, an integer $n \geq 1$, and a polynomial $p(x) \in k[x]$, the curve $C_f \subset \mathbb{A}^2_k$ defined by the vanishing of the polynomial

$$f(x, y) := y^n - p(x) \in k[x, y]$$

is smooth iff the polynomial $p(x)$ is separable, i.e. $\operatorname{disc}(p) \neq 0$.[8]

[8] See Exercise 2.2.10. When $\operatorname{ch} k \neq 2$, smooth curves of the form C_f with $n = 2$ are called hyperelliptic curves.

2.4 Exercise Sheet 4

2.4.1 Numerical and Exploration

Exercise 2.4.1. What can you say about $\mathrm{Der}_{\mathbb{Q}}\left(\mathbb{Q}[\sqrt{-3}]\right)$? $\mathrm{Der}_{\mathbb{Q}}(\mathbb{Q}[\sqrt{7}, \cos(2\pi/5)])$? $\mathrm{Der}_{\mathbb{R}}(\mathbb{C})$? $\mathrm{Der}_{\mathbb{Q}}(\mathbb{Q}[\pi])$? $\mathrm{Der}_{\mathbb{Q}}(\mathbb{C})$? $\mathrm{Der}_{\mathbb{F}_p(t)}(\mathbb{F}_p(t)[s]/(s^p - t))$? What about $\mathrm{Der}_k K$, where k is any field and $K = \operatorname{Frac} k[x, y]/(f)$ for $f = y, y - x^2, y^2 - x^3, y^2 - x^3 + x$? Make (and prove) conjectures.

Exercise 2.4.2. (Adapted from [3, Exercise 5.2].) Define what it means for a projective plane curve to be irreducible. For each of the following polynomials F, identify whether the projective curve $C_F \subset \mathbb{P}^2_k$ is irreducible, find all the multiple points, their multiplicities, and tangent lines at the multiple poins.

(a) $XY^4 + YZ^4 + ZX^4$.
(b) $X^2Y^3 + Y^2Z^3 + Z^2X^3$.
(c) $Y^2Z - X(X - Z)(X - \lambda Z)$ for $\lambda \in k$.
(d) $X^n + Y^n + Z^n$ for $n \geq 1$.

What is the relationship between the irreducibility of F and that of C_F? Do your answers depend on the characteristic of the base field?

Exercise 2.4.3. (Adapted from [3, Exercise 5.3].) Find all points of intersection of the following pairs of curves, and the intersection numbers at these points.

(a) $X^2 + Y^2 - Z^2$ and Z.
(b) $(X^2 + Y^2)Z + X^3 + Y^3$ and $X^3 + Y^3 - 2XYZ$.
(c) $Y^5 - X(Y^2 - XZ)^2$ and $Y^4 + Y^3Z - X^2Z^2$.
(d) $(X^2 + Y^2)^2 + 3X^2YZ - Y^3Z$ and $(X^2 + Y^2)^3 - 4X^2Y^2Z^2$.

Do your answers depend on the base field?

Exercise 2.4.4 (Singular Plane Cubics)**.** Let $F \in k[X, Y, Z]$ be an irreducible homogeneous cubic polynomial, and suppose that $C = C_F$ has a cusp at a point $P \in C$ (see Exercise 2.3.5).

(a) Show that there is a projective change of coordinates such that $P = [0 : 0 : 1]$ and $\mathrm{T}_P C$ is defined by $Y = 0$. Show that in these coordinates,

$$F = Y^2Z - AX^3 - BX^2Y - CXY^2 - DY^3$$

for some $A, B, C, D \in k$ with $A \neq 0$, up to scaling F by a nonzero scalar.

(b) Find a projective change of coordinates to make $C = D = 0$. In other words, find a projective change of coordinates $\phi : \mathbb{P}^2_k(X_1, Y_1, Z_1) \to \mathbb{P}^2_k(X, Y, Z)$ such that we have $\phi^* F = Y_1 Z_1^2 - AX_1^3 - BX_1^2 Y$.

(c) Now suppose that k is algebraically closed (or even that $k^\times = (k^\times)^3$, i.e. that every nonzero element is a cube) and also that $\operatorname{ch} k \neq 3$. Find a projective change of coordinates to make $A = 1$ and $B = 0$. Conclude that when k satisfies the above hypotheses (e.g. $k = \mathbb{C}$ or $k = \mathbb{F}_5$), there is a unique cuspidal plane cubic up to projective changes of coordinates, and this has no other singularities. What happens when these hypotheses on k are not satisfied?

(d) Similarly, show that under suitable hypotheses on k, there is a unique nodal plane cubic up to projective changes of coordinates, and this has no other singularities. Explore what happens when these hypothesis on k do not apply.

(e) Give at least two proofs of the following fact: under suitable hypothesis on the base field k, any irreducible projective plane cubic is either nonsingular, or has at most one singular point of multiplicity at most 2, which must be either a node or a cusp. (Hint: For one, use (c) and (d). For the other, use the correct salvage of Exercise 2.4.9 below.)

(f) What can you say about irreducible singular plane quartic curves? Can you come up with a similar clasification? What about singular plane quintic curves? Can you explore and make some general conjectures?

Exercise 2.4.5 (Hessian). (Adapted from [4, Exercise 3.29].) Let $F \in k[X,Y,Z]$ be a homogeneous polynomial. We define the Hessian polynomial of F to be

$$\mathrm{Hess}(F) := \det \begin{bmatrix} \partial^2 F/\partial X^2 & \partial^2 F/\partial X \partial Y & \partial^2 F/\partial X \partial Z \\ \partial^2 F/\partial X \partial Y & \partial^2 F/\partial Y^2 & \partial^2 F/\partial Y \partial Z \\ \partial^2 F/\partial X \partial Z & \partial^2 F/\partial Y \partial Z & \partial^2 F/\partial Z^2 \end{bmatrix}.$$

(a) Show that if $\Phi : \mathbb{P}^2_k(X',Y',Z') \to \mathbb{P}^2_k(X,Y,Z)$ is a projective change of coordinates and we pick a lift $\Phi^* : k[X,Y,Z] \to k[X',Y',Z']$ representing it, then, then we have that $\mathrm{Hess}(\Phi^* F) = C \cdot \Phi^*(\mathrm{Hess}(F))$ for some nonzero constant C. What is C in terms of F and Φ^*?

(b) Compute the Hessian for

$$F_\lambda := Y^2 Z - X(X-Z)(X-\lambda Z),$$

where $\lambda \in k$, and describe the intersection $C_{F_\lambda} \cap C_{\mathrm{Hess}(F_\lambda)}$? (If the general case is too hard, can you do this for some special values of λ?)

(c) Show that if $\mathrm{ch}\, k \neq 2, 3$, if F is irreducible of $\deg F \geq 2$ and if $P \in C_F$ is a smooth point of C_F, then $P \in C_F \cap C_{\mathrm{Hess}(F)}$ iff $i_P(C_F, \mathrm{T}_P C_F) \geq 3$. Such a point is called an inflection point of C_F.

(d) How many inflection points can a smooth curve of degree 2 have? What about 3? 4? 5? Find patterns and make some conjectures.

See also [3, Exercises 5.23-24].

2.4.2 PODASIPs

Prove or disprove and salvage if possible the following statements.

Exercise 2.4.6. If k is any field and $f \in k[t]$ a nonconstant polynomial, then $\partial_t f \neq 0$.

Exercise 2.4.7. If k is any infinite field and $C \subset \mathbb{P}^2_k$ a projective plane curve, then C is infinite.

Exercise 2.4.8. Given any two ordered sets of nonconcurrent lines (L_1, L_2, L_3) and (L_1', L_2', L_3') in $\mathbb{P}^2_k$, there is a unique projective change of coordinates $\phi : \mathbb{P}^2_k \to \mathbb{P}^2_k$ such that $\phi(L_i) = L_i'$ for $i = 1, 2, 3$.

Exercise 2.4.9 (Bézout's Theorem for a Line). If k is any field and $C \subset \mathbb{P}^2_k$ a projective curve of degree $n \geq 1$ with minimal polynomial $F \in k[X,Y,Z]_n$, then for any line $C_L \subset \mathbb{P}^2_k$ where $L \in k[X,Y,Z]_1$, we have

$$\sum_{P \in C_F \cap L} i_P(F, L) = n.$$

Exercise 2.4.10. If $F \in k[X,Y,Z]$ is a nonconstant homogenous polynomial, the the projective curve C_F defined by F is irreducible iff F is.

2.5 Exercise Sheet 5

2.5.1 Standard Exercises/Numerical and Exploration

Exercise 2.5.1. Given a nonempty finite set $S \subset \mathbb{P}^2_k$ of points in $\mathbb{P}^2_k$, let $d(S)$ be the smallest degree of a curve $C \subset \mathbb{P}^2_k$ through S, i.e. such that $C \supset S$. Let's investigate the relationship between S, its size $n := \#S$, and the integer $d(S)$.

(a) Show that if $n \in \{1, 2\}$, then $d(S) = 1$.
(b) Show that if $n \in \{3, 4\}$, then $d(S) \in \{1, 2\}$. When does each case hold?
(c) Show that if $n = 5$, then $d(S) \in \{1, 2\}$, or equivalently that given any five distinct points $P_1, \ldots, P_5 \in \mathbb{P}^2_k$, there is at least one (possibly reducible) conic $C \subset \mathbb{P}^2_k$ passing through each P_i.
(d) Show that, in general, we have

$$1 \leq d(S) \leq \left\lceil \frac{\sqrt{9 + 8n} - 3}{2} \right\rceil,$$

where $\lceil \cdot \rceil$ denotes the ceiling function. (Hint: When does a system of N linear equations in M variables always have a solution that is not identically zero?)
(e) (Cramer's Theorem) Show that the bound in (d) is sharp in general: for each $n \geq 1$, come up with a collection S of n points such that $d(S)$ equals the upper bound from (d). Can you characterize the sets S for which this equality holds? What possible intermediate values of $d(S)$ are possible?

Exercise 2.5.2. Let k be a field.

(a) Suppose $\operatorname{ch} k \neq 2$, and consider the collection of 9 points $S := \{(i, j) \in \mathbb{A}^2_k : 0 \leq i, j \leq 2\}$. How many distinct cubic curves $C \subset \mathbb{A}^2_k$ pass through S? (Hint: by Exercise 2.5.1(d), there is at least one such C. Does your answer change if the question is about projective cubics instead? Does the choice of base field matter? Can you come up with an analog if $\operatorname{ch} k = 2$?)
(b) Can you formulate an analog of (a) for a configuration of n^2 points

$$S := \{(i, j) \in \mathbb{A}^2_k : 0 \leq i, j \leq n - 1\},$$

where $n \geq 2$ is any integer (say when $\operatorname{ch} k = 0$ for convenience)?

Exercise 2.5.3 (More on Pascal)**.** (Adapted from [3, Exercise 5.31].) If in Pascal's Theorem, we let some adjacent vertices coincide (the side being tangent), then we get many new theorems.

(a) State and sketch what happens if $P_1 = P_2$, $P_3 = P_4$ and $P_5 = P_6$.
(b) Let $P_1 = P_2$ and the other four points be distinct. Deduce a rule for constructing a tangent to a given conic at a given point, using only a straight-edge.

Exercise 2.5.4. Let $C \subset \mathbb{P}^2_k$ be a curve of degree d over an algebraically closed field k.

(a) Make sense of the following statement: a "general" line $L \subset \mathbb{P}^2_k$ intersects C in exactly d distinct points.
(b) Given a "general" point $P \in \mathbb{P}^2_k$, how many lies through P are tangent to C?

(Hint: How is this exercise is related to Exercises 2.5.8, 2.5.9, and 2.5.10? For (b), you may suppose for convenience that $\operatorname{ch} k = 0$. What happens in positive characteristic?)

Exercise 2.5.5. Let k be an algebraically closed field, and let $C \subset \mathbb{P}^2_k$ be a smooth cubic curve.

(a) Show that C has exactly 9 inflection points. The set of inflection points on C is usually denoted by $C[3]$. (Hint: Exercise 2.4.5. You may assume $\operatorname{ch} k \neq 2, 3$ for convenience.)

(b) Show that $C[3]$ is not contained in a line, but any line passing through any two points in $C[3]$ passes through a third point in $C[3]$. Why does this not violate the Sylvester-Gallai Theorem?

(c) Suppose that $\operatorname{ch} k \neq 3$. Show that by a projective change of coordinates, we can bring $C[3]$ to be the nine points

$$[0:1:\xi], [\xi:0:1], [1:\xi:0],$$

where ξ runs over the three roots of $t^3+1=0$ in k.[9]

(d) Keeping the hypothesis that $\operatorname{ch} k \neq 3$, show that every cubic curve passing through the 9 points from (c) has the equation

$$F_\Lambda = \lambda(X^3+Y^3+Z^3) + 3\mu XYZ \in k[X,Y,Z]$$

for some $\Lambda := [\lambda:\mu] \in \mathbb{P}^1_k$. This curve is singular iff Λ is either $[0:1]$ or $[1:\xi]$ where $\xi^3+1=0$. In each of these cases the curve $C_\Lambda := C_{F_\Lambda}$ degenerates into a product of three lines. If C_Λ is irreducible, then the flexes of C_Λ are exactly the 9 points above.

(e) Conclude, using either (b) or both (c) and (d), that if $k=\mathbb{C}$, then

$$\#\left(C[3] \cap C(\mathbb{R})\right) \leq 3,$$

i.e. at most three of the flexes of a complex smooth cubic curve can be real. Come up with a curve C for which this bound is achieved. Can this intersection have fewer than 3 points? Can it have exactly 2?

Exercise 2.5.6. If $f, g \in k[x,y]$ are nonconstant polynomials and $P \in \mathbb{A}^2_k$, then

$$i_P(f,g) \geq m_P(f) \cdot m_P(g).$$

When does equality hold? (This is a very hard exercise, and you may not be able to do it with the tools we have developed so far; nonetheless, it is very valuable to work out special cases. Try doing the case when f or g is linear. Next, try the case when $m_P(f)=1$ or $m_P(g)=1$. Finally, see how far you can extend your techniques to the next (or general) case; once you've done that, see [3, §3.3, Theorem 3] or [15, Theorem 7.4].)

2.5.2 PODASIPs

Prove or disprove and salvage if possible the following statements.

Exercise 2.5.7 (Braikenridge-Maclaurin Theorem/Converse to Pascal's Theorem)**.** If the intersection points of opposite sides of a hexagon lie on a straight line, then the vertices of the hexagon lie on a conic.

Exercise 2.5.8. (Adapted from [3, Exercise 5.26].) If $C \subset \mathbb{P}^2_k$ is a curve of degree $n \geq 1$, and $P \in \mathbb{P}^2_k$ a point of multiplicity $m := m_P(C) \geq 0$, then for all but finitely many lines L through P, the line L intersects C in $n-m$ distinct points other than P.

Exercise 2.5.9. Given a curve $C \subset \mathbb{P}^2_k$ and a point $P \in \mathbb{P}^2_k$, there is at least one tangent line L to C that does not pass through P.

Exercise 2.5.10 (Dual Curve)**.** Let $C \subset \mathbb{P}^2_k$ be a curve. Let

$$C^* := \{L \in \mathbb{P}^{2*}_k : L \text{ is tangent to } C \text{ at some point } P \in C\} \subset \mathbb{P}^{2*}_k.$$

Then $C^* \subset \mathbb{P}^{2*}_k$ is a curve, and $C^{**} = C$. (Hint: Can you work out a few examples in low degrees? What is the relationship between the degrees of C and C^*?)

[9] That these roots are distinct uses $\operatorname{ch} k \neq 3$.

2.6 Exercise Sheet 6

2.6.1 Numerical and Exploration

Exercise 2.6.1 (Brianchon's Theorem). Let $C \subset \mathbb{P}^2_k$ be a smooth conic, and $(L_1, \ldots, L_6)$ an ordered six-tuple of pairwise distinct lines tangent to it. For $i = 1, \ldots, 6$, let $P_i := L_i \cap L_{i+1}$, where $L_7 := L_1$, and for $1 \leq i < j \leq 6$, let M_{ij} denote the line joining P_i and P_j.

(a) Show that the lines M_{14}, M_{25} and M_{36} are concurrent. See Figure 2.3.
(b) How many such distinct configurations can you produce from an unordered set of 6 distinct lines $L_1, \ldots, L_6$?
(c) Explore what happens when some of the lines $L_1, \ldots, L_6$ "collide"–what theorems can you obtain then?

(Hint: Theorem 1.13.5 and Exercise 2.5.10.)

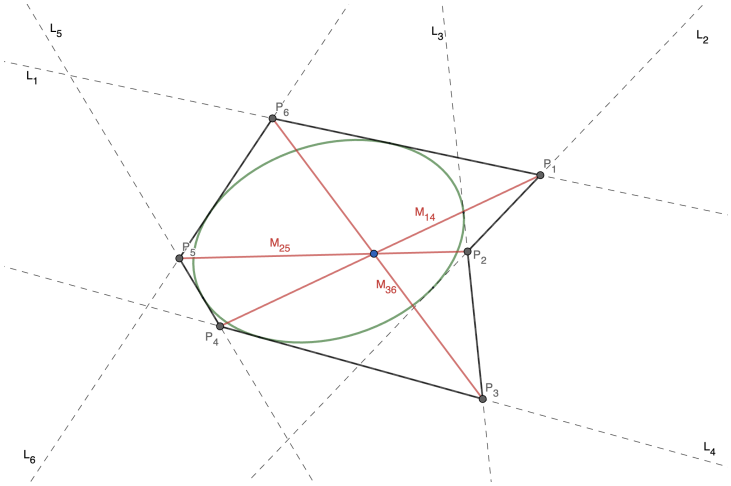


Figure 2.3: Brianchon's Theorem. Picture made with Geogebra.

Exercise 2.6.2. Suppose that k is an algebraically closed field of characteristic other than 2. Show that there are, up to projective changes of coordinates, exactly 8 types of pencils of conics in $\mathbb{P}^2_k$, as described in Example 1.15.7. Explore what happens when k is not algebraically closed or has characteristic 2.

Exercise 2.6.3. Solve, by hand, the quartic equation

$$x^4 - 4x^3 - 22x^2 + 116x - 119 = 0$$

over an arbitrary field k. In other words, given a arbitrary field k, determine how many roots this equation has in k and what are their multiplicities are. (Hint: Example 1.15.11.)

Exercise 2.6.4. Suppose that k is a field of characteristic other than 2 or 3.

(a) For each $\alpha \in k$, let $F_\alpha := X^3 + Y^3 + \alpha Z^3 \in k[X, Y, Z]$, and let $E_\alpha := C_{F_\alpha}$ be the corresponding cubic curve. Show that when $\alpha \neq 0$, the curve E_α is smooth, and so becomes an elliptic curve when equipped with the base point $O = [1 : -1 : 0]$.
(b) Find a projective change of coordinates that brings E_α into Weierstrass normal form, and use this to find $j(E_\alpha)$ as a function of α.
(c) Next, suppose that $k = \mathbb{Q}$. Determine $E_\alpha(\mathbb{Q})$, i.e. the $\mathbb{Q}$-rational points of E_α for $\alpha \in \{\pm 1, \pm 2\}$. Show that if α is an integer other than $\pm 1, \pm 2$, then $E_\alpha(\mathbb{Q})$ is infinite. Conclude that for each integer α other than $\pm 1, \pm 2$, there are infinitely many coprime triples (X, Y, Z) of integers such that $X^3 + Y^3 + \alpha Z^3 = 0$.

(d) Using a computer, determine $\#E_1(\mathbb{F}_p)$, i.e. the number of points on E_1 over the finite field $k = \mathbb{F}_p$ with p elements, for all primes $p \in [5, 1000]$. What patterns do you observe? Make conjectures, and prove them. (Hint: Consider the cases $p \equiv 1, 2 \pmod 3$ separately.)

Exercise 2.6.5. (Adapted from [10, Exercise 1.18].) Consider the elliptic curve E defined in Weierstrass normal form by

$$y^2 = x^3 + 17$$

over $k = \mathbb{Q}$. Note that E contains the rational points

$$Q_1 = (-2, 3), Q_2 = (-1, 4), Q_3 = (2, 5), Q_4 = (4, 9), \text{ and } Q_5 = (8, 23).$$

(a) Show that Q_2, Q_4 and Q_5 can be expressed as $mQ_1 + nQ_2$ for appropriate choices of $m, n \in \mathbb{Z}$.
(b) Compute the points $Q_6 = -Q_1 + 2Q_3$ and $Q_7 = 3Q_1 - Q_3$.
(c) Notice that the points $Q_1, \ldots, Q_7$ and there inverses all have integer coordinates. There is exactly one more rational point Q_8 on this curve that has integer coordinates and $y > 0$. Find it.

If you are up for a real challenge, here are a few more things to think about in this example:

(d) Show the claim made in (c) about the set of all integral points on E.
(e) Show that $E(\mathbb{Q}) \cong \mathbb{Z}^2$, i.e. there are no nontrivial rational torsion points on E and $E(\mathbb{Q})$ has rank 2. Can some two of the above points $Q_1, \ldots, Q_8$ be taken to be two generators for $E(\mathbb{Q})$, and if so, which ones?

Exercise 2.6.6. (Adapted from [10, Exercise 2.13].) Let k be a field of characteristic other than 2, let $t \in k$, and consider the projective closure $E_t \subset \mathbb{P}^2_k$ of the locus defined by

$$y^2 = x^3 - (2t - 1)x^2 + t^2 x.$$

(a) Prove that E_t is nonsingular iff $t \notin \{0, 1/4\}$, in which case (E_t, O) is an elliptic curve over k with $O = [0 : 1 : 0]$. What is $j(E_t)$?
(b) Show that, in the situation in (a), the point $(t, t) \in E(k)$ has order 4.
(c) Show that if $E \subset \mathbb{P}^2_k$ is any elliptic curve over a field k of characteristic other than 2 or 3 such that there is a point $P \in E(k)$ of order 4, then there is a projective change of coordinates $\Phi : \mathbb{P}^2_k \to \mathbb{P}^2_k$ such that $\Phi(E) = E_t$ and $\Phi(P) = [t : t : 1]$ for some $t \notin \{0, 1/4\}$.
(d) For a given pair (E, P) as in (c), how many values of t work?

2.6.2 PODASIPs

Prove or disprove and salvage if possible the following statements.

Exercise 2.6.7. If k is a field, and $S \subset \mathbb{P}^2_k$ a finite subset, then there is a line $L \subset \mathbb{P}^2_k$ such that $S \cap L = \emptyset$, i.e. in projective space, a line can be chosen that avoids any finite set of points. Can we produce two such lines L_1, L_2? Can we produce n such lines for any $n \geq 1$? Can we produce infinitely many?

Exercise 2.6.8. Every connected component of a real elliptic curve is a subgroup of it under the elliptic curve addition law. A real elliptic curve is isomorphic as a group (in fact, as a Lie group[10]) to the circle group $S^1 := \{z \in \mathbb{C} : |z| = 1\}$.

Exercise 2.6.9. Let $E \subset \mathbb{P}^2_k$ be a smooth cubic curve, and let $O, O' \in E$ be two points. There is a projective change of coordinates $\Phi : \mathbb{P}^2_k \to \mathbb{P}^2_k$ such that $\Phi(E) = E$ and $\Phi(O) = \Phi(O')$; in

[10] What's that?

particular, as abelian groups, $(E, O) \cong (E, O')$. (Hint: For a very strong salvage, consider the map $\alpha : E \to E$ defined as follows. Let $L_{O,O'}$ intersect E in the third point T, and consider the map $\alpha : E \to E$ which sends a $P \in E$ to the third intersection point of the line $L_{P,T}$ with E.)

Finally, here are a couple more really challenging exercises to keep you occupied all (the rest of) summer.

Exercise 2.6.10 (Division Polynomials)**.** Let $R := \mathbb{Z}[p, q]$ be the polynomial ring in two variables p, q. Take the polynomial $f := x^3 + px + q \in R[x]$, and let $f' = 3x^2 + p$ and $f'' = 6x$ be the first and second formal derivatives of f with respect to x.

(a) Define the sequence $(f_n)_{n\geq 0}$ of polynomials in $R[x]$ recursively by $f_0 = 0, f_1 = f_2 = 1$,

$$\begin{aligned} f_3 &:= 2f \cdot f'' - (f')^2, \\ f_4 &:= -16f^2 + 4f \cdot f' \cdot f'' - 2(f')^3, \\ f_{2n+1} &:= f_{n+2} \cdot f_n^3 - 16f^2 \cdot f_{n-1} \cdot f_{n+1}^3 \quad \text{for } n \geq 2 \text{ odd}, \\ f_{2n+1} &:= 16f^2 \cdot f_{n+2} \cdot f_n^3 - f_{n-1} \cdot f_{n+1}^3 \quad \text{for } n \geq 2 \text{ even, and} \\ f_{2n} &:= f_n(f_{n+2} \cdot f_{n-1}^2 - f_{n-2} \cdot f_{n+1}^2) \quad \text{for } n \geq 3. \end{aligned}$$

For $n \geq 1$, we have

$$f_n = \begin{cases} nx^{(n^2-1)/2} + \cdots, & \text{for } n \text{ odd, and} \\ (n/2)x^{(n^2-4)/2} + \cdots, & \text{for } n \text{ even,} \end{cases}$$

where $\cdots$ denotes terms of lower degree.

(b) The equation $y^2 = f$ defines an elliptic curve E in Weierstrass normal form (over $k = \mathbb{Q}(p, q)$ or over any field k of characteristic other than 2 when given specific $p, q \in k$ such that $4p^3 + 27q^2 \neq 0 \in k$). In this case,

$$\gcd(f_n, f \cdot f_{n+1} \cdot f_{n-1}) = (1)$$

when n is odd and

$$\gcd(f \cdot f_n, f_{n+1} \cdot f_{n-1}) = (1)$$

when $n \geq 2$ is even.

(c) If $P = (x, y) \in E$, then the coordinates of $nP \in E$ are given as

$$nP = \left(x - \frac{4 \cdot f \cdot f_{n+1} \cdot f_{n-1}}{f_n^2}, y \cdot \frac{f_{2n}}{f_n^4} \right)$$

when n is odd and

$$nP = \left(x - \frac{f_{n+1} \cdot f_{n-1}}{4f \cdot f_n^2}, y \cdot \frac{f_{2n}}{16f^2 \cdot f_n^4} \right)$$

when n is even.

(d) Now fix an $n \geq 1$, and suppose that k is an algebraically closed field with $\operatorname{ch} k \nmid 2n$.

(1) For $P = (x, y) \in E$, we have $nP = O$ iff the x-coordinate $x(P)$ of P satisfies $f_n(x) = 0$ when n is odd or satisfies $f(x) \cdot f_n(x) = 0$ when n is even.

(2) When n is odd, the polynomial f_n is separable, and when n is even, the polynomial $f \cdot f_n$ is separable (Exercise 2.2.10).

(3) There are exactly n^2 points of order dividing n in E, and, in fact, we have

$$E[n] \cong \mathbb{Z}/n \times \mathbb{Z}/n.$$

(Hint: If G is an abelian group of order n^2 for some $n \geq 1$ such that for each divisor $d \mid n$ we have $\#G[d] = d^2$, where $G[d] \subset G$ is the subgroup of all points of order dividing d, then $G \cong \mathbb{Z}/n \times \mathbb{Z}/n$.)

(e) Now suppose that $p, q \in \mathbb{R}$. How many real roots can $f_3(x) \in \mathbb{R}[x]$ have? Use this to give another solution to Exercise 2.5.5(e).

Exercise 2.6.11 (Elliptic Divisibility Sequences). (Adapted from [9, Exercises 3.34-3.36].) Let k be a field. A (nondegenerate) elliptic divisibility sequence (EDS) over k is a sequence $a = (a_n)_{n\geq 1}$ defined by four initial parameters a_1, a_2, a_3, a_4 with $a_1 a_2 a_3 \neq 0$ subject to the recursive relations

$$a_{2n+1} = \frac{1}{a_1^3}\left(a_{n+2}a_n^3 - a_{n-1}a_{n+1}^3\right), \text{ and}$$
$$a_{2n} = \frac{1}{a_1^2 a_2} a_n (a_{n+2}a_{n-1}^2 - a_{n-2}a_{n+1}^2)$$

for all $n \geq 2$.

(a) The sequence a defined by $a_n = n$ is an EDS. The sequence a defined by $a_n = F_n$, where F_n is the n^{th} Fibonacci number, is an EDS. More generally, given $a_1, a_2, x, y \in k$, the sequence a defined by the linear recursive relation

$$a_n = x a_{n-1} + y a_{n-2}$$

for $n \geq 2$ is an EDS.

(b) If $(a_n)_{n\geq 1}$ is an EDS, then for each $m \geq 1$ such that $a_m \neq 0$, so is the sequence $(a_{mn}/a_m)_{n\geq 1}$. An EDS such that $a_1 = 1$ is said to be normalized; given any sequence a we define its normalization $\tilde{a}$ to be given by $\tilde{a}_n = a_n/a_1$ for $n \geq 1$. Given a normalized EDS $(a_n)_{n\geq 1}$, we define its discriminant to be

$$\Delta := a_4 a_2^{15} - a_3^3 a_2^{12} + 3a_4^2 a_2^1 0 - 20 a_4 a_3^3 a_2^7 + 3a_4^3 a_2^5 + 16 a_3^6 a_2^4 + 8 a_4^2 a_3^2 a_2^2 + a_4^4.$$

We say that a EDS is singular if the discriminant of its normalization is zero; else it is said to be nonsingular. Which of the sequences from (a) are nonsingular?

(c) Let $E : y^2 = x^3 + px + q$ be an elliptic curve over k, and let $P = (x_0, y_0) \in E(k)$. The sequence $a = (a_n)_{n\geq 1}$ defined by

$$a_n = \begin{cases} f_n(x_0) & n \text{ odd, and} \\ 2y_0 \cdot f_n(x_0), & n \text{ even,} \end{cases}$$

is an EDS, where the polynomials f_n are as in Exercise 2.6.10. What is the discriminant of (the normalization of) this sequence a_n? Is this sequence singular?

(d) The sequence $a = (a_n)_{n\geq 1}$ is an EDS iff for each $m > n > r > 0$, we have

$$a_{m+n}a_{m-n}a_r^2 = a_{m+r}a_{m-r}a_n^2 - a_{n+r}a_{n-r}a_m^2.$$

(e) Now suppose that $k = \operatorname{Frac} R$ for some integral domain R, and let $a = (a_n)$ be an EDS over k such that $a_1, a_2, a_3, a_4 \in R$ and such that $a_1 \mid a_i$ for $i = 2, 3, 4$ and $a_2 \mid a_4$. Then a is a divisibility sequence in the sense that each $a_n \in R$ and if $m, n \geq 1$ are integers, then

$$m \mid n \Rightarrow a_n \mid a_m.$$

If, further, R is a PID and $\gcd(a_3, a_4) = 1$, then for all $m, n \geq 1$ we have

$$a_{\gcd(m,n)} = \gcd(a_m, a_n),$$

up to units. In particular, these properties hold for the Fibonacci sequence F_n.

(f) Finally suppose that $k = \mathbb{R}$. Suppose that a is a nonsingular, non-periodic EDS. Then there is a real number $h > 0$ such that

$$\lim_{n\to\infty} \frac{\log |a_n|}{n^2} = h.$$

Bibliography

[1] E. Brieskorn and H. Knörrer, *Plane Algebraic Curves*, vol. 38. Springer Basel AG, 1986.

[2] K. Conrad, "Remarks about Euclidean Domains." `https://kconrad.math.uconn.edu/blurbs/ringtheory/euclideanrk.pdf`.

[3] W. Fulton, *Algebraic Curves: An Introduction to Algebraic Geometry*. third ed., 2008. `https://dept.math.lsa.umich.edu/~wfulton/CurveBook.pdf`.

[4] A. Gathmann, "Plane Algebraic Curves: Class Notes RPTU Kaiserslautern." `https://agag-gathmann.math.rptu.de/en/curves.php`, 2023.

[5] M. Reid, *Undergraduate Algebraic Geometry*, vol. 12 of *London Mathematical Society Student Texts*. Cambridge University Press, 2001. Available also at https://homepages.warwick.ac.uk/staff/Miles.Reid/MA4A5/UAG.pdf.

[6] F. Kirwan, *Complex Algebraic Curves*. No. 23 in London Mathematical Society Student Texts, Cambridge University Press, 1992.

[7] T. Tao, "Pappus's Theorem and Elliptic Curves." `https://terrytao.wordpress.com/2011/07/15/pappuss-theorem-and-elliptic-curves/`, 2011.

[8] D. Eisenbud, M. L. Green, and J. Harris, "Cayley-Bacharach Theorems and Conjectures," *Bulletin of the American Mathematical Society*, vol. 33, pp. 295–324, 1996.

[9] J. H. Silverman, *The Arithmetic of Elliptic Curves*, vol. 106 of *Graduate Texts in Mathematics*. Springer, 2nd ed.

[10] J. H. Silverman and J. T. Tate, *Rational Points on Elliptic Curves*, vol. 106 of *Undergraduate Texts in Mathematics*. Springer, 2nd ed.

[11] Elkies and Klagsbrun, "Rank $\geq$ 29." `https://web.math.pmf.unizg.hr/~duje/tors/rk29.html`.

[12] J. Harris, *Algebraic Geometry: A First Course*, vol. 133 of *Graduate Texts in Mathematics*. Springer.

[13] J. Milnor, *Singular Points of Complex Hypersurfaces*. Princeton University Press and the University of Tokyo Press, 1968.

[14] F. Lemmermeyer, "Parametrizing Algebraic Curves." `https://doi.org/10.48550/arXiv.1108.6219`, 2011.

[15] E. Kunz, *Introduction to Plane Algebraic Curves*. Birkhäuser, 2005.