Plane Algebraic Curves

Dhruv Goel

June 2024

Contents

Preface

These are lecture notes for a course taught at Ross/Ohio 2024 intended for peer mentors and counselors.

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 $\mathbb{A}_k^2$ (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 $C, D \subset \mathbb{A}_k^2$ 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 $C \subset \mathbb{A}_k^2$, 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.

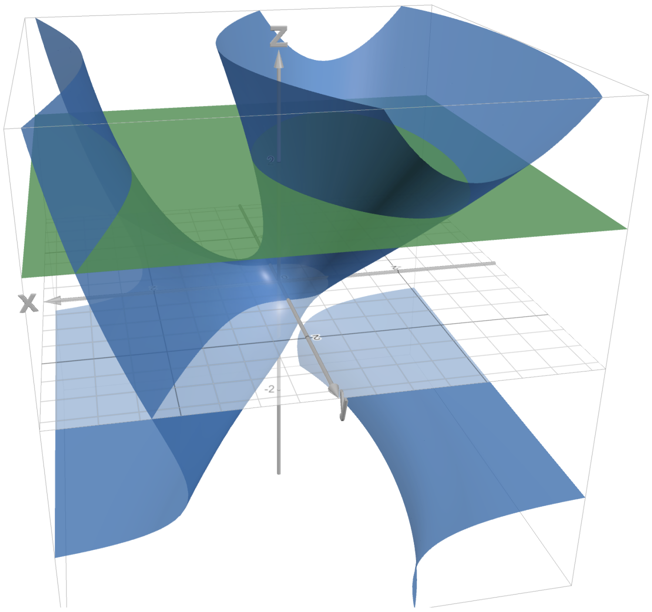


Figure 1.1: The elliptic curve over $k = \mathbb{C}(z)$ defined by $y^2 = x^3 - 3zx + (z^3 + 1)(z + 2)^{-1}$ in blue, along with its hyperplane section at $z = 2$, which is the elliptic curve $y^2 = x^3 - 6x + 9/4$. Picture made with Desmos 3D.

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^kg 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

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

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

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

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

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

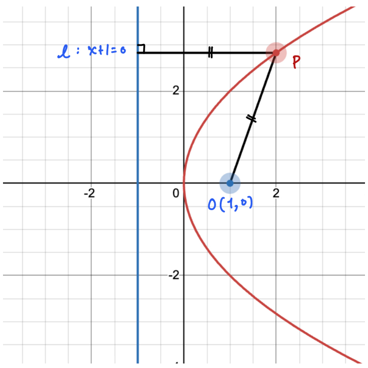


Figure 1.2: The synthetic construction of the parabola. Picture made with Desmos.

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.

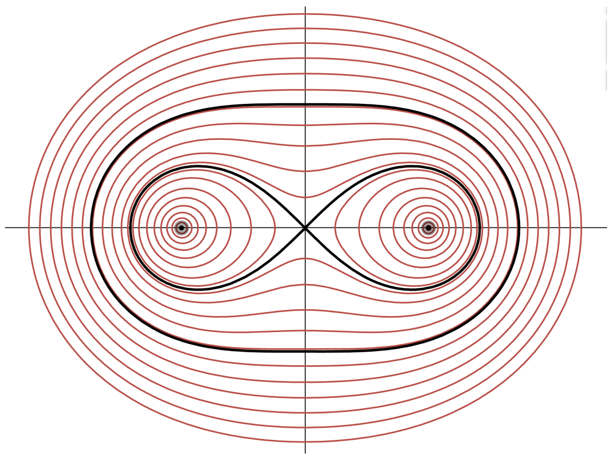


Figure 1.3: The Cassini ovals. Picture made with Desmos.

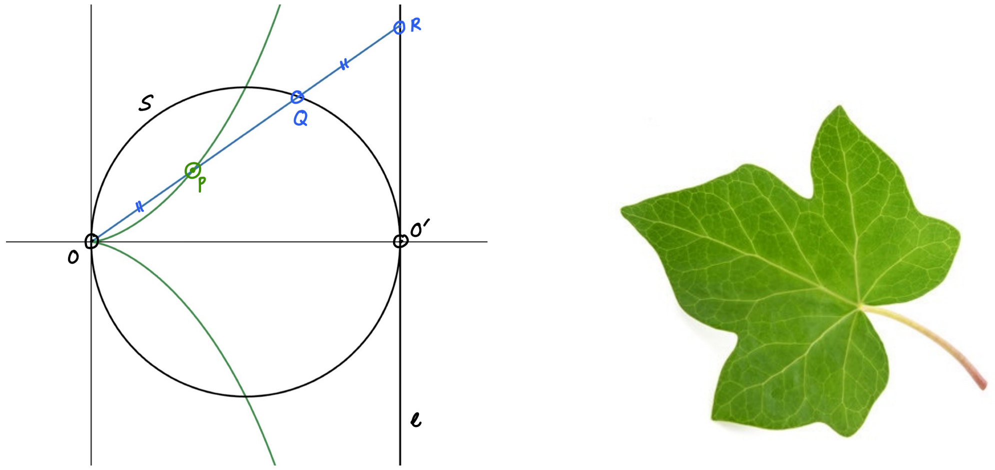
Example 1.2.13 (Cissoid of Diocles). This curve is named after the ancient Greek mathematician Diocles. To construct it, start with a circle $S \subset \mathbb{A}^2_{\mathbb{R}}$ and a point $O \in S$. Construct the diameter OO' to S through O as well as the tangent line ℓ to S through O'. Now for each point $Q \in S$, extend the line OQ to meet ℓ in R, and mark off the point P on OQ such that $\mathrm{dist}(OP) = \mathrm{dist}(QR)$. 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.

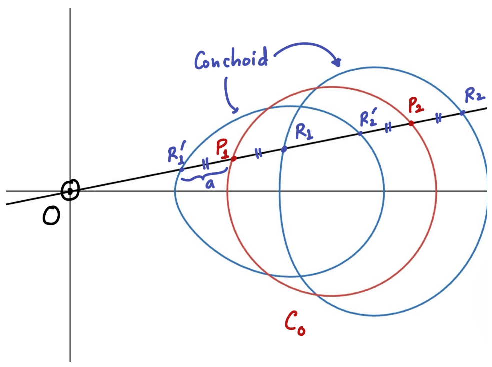
There are many other constructions of this curve: for instance, it is the curve obtained by inverting a parabola in a circle centered at its vertex, and also, if two congruent parabolae are set vertex-to-vertex, and one rolls on the other, then the vertex of the rolling parabola traces out the cissoid.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

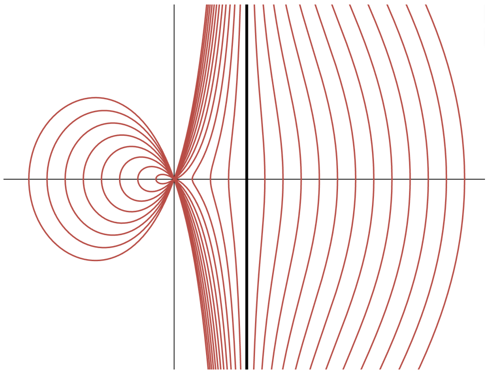
Example 1.2.14 (Conchoids). Our final example of a synthetic construction is that of conchoids. To construct a conchoid, you need a triple (O, C_0, a), where $O \in \mathbb{A}^2_{\mathbb{R}}$ is a point, $C_0 \subset \mathbb{A}^2_{\mathbb{R}}$ is the "base curve" and $a \in [0, \infty)$. Then the conchoid 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

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

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.

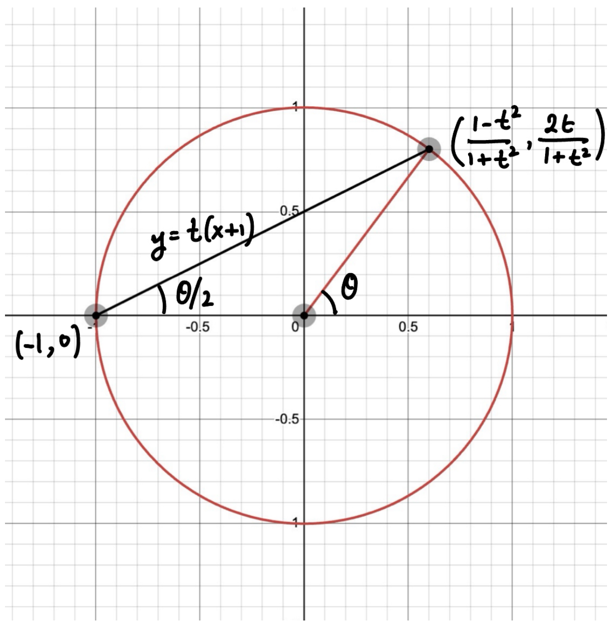


Figure 1.6: Parametrizing the circle $x^2+y^2=1$.

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 $\mathbb{A}^2_k \setminus C$ 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 2.3.1. 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 $f_1, \dots, f_n \in R$ such that

$$f = f_1 f_2 \cdots f_n$$

and if there is some other integer $m \geq 1$ and irreducible elements $g_1, \dots, g_m \in R$ 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 $\sigma : \{1, \dots, n\} \to \{1, \dots, n\}$ and units $c_1, \dots, c_n \in R^\times$ 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) ⇒ (b) We only need to show that every irreducible in a UFD is prime; I leave this to the reader.

(b) ⇒ (c) Clear.

(c) ⇒ (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 $p_1, \dots, p_n$, 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, $[t^{m+n}]fg = a_0 b_0 \neq 0$, 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 $f_1, \dots, f_n$ 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 $\tilde{f}\tilde{g}$ 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 $f(t) = a_0t^m + \cdots + a_m$ and $g(t) = b_0t^n + \cdots + b_n$, 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.

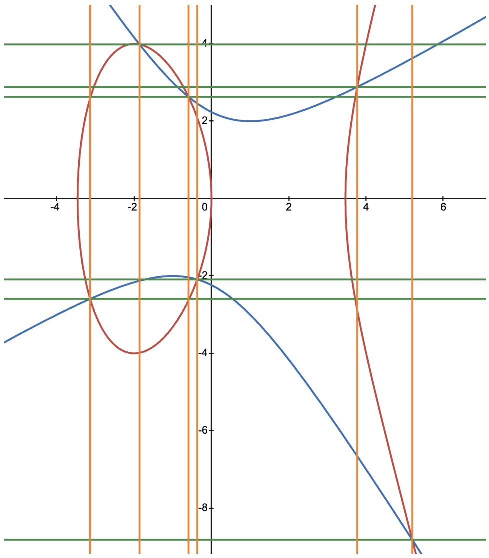


Figure 1.7: An illustration of Theorem 1.6.6 for the f and g in Example 1.6.4. The red curve is C_f, the blue curve is C_g, the green lines correspond to the roots of c_y, and the orange lines correspond to the roots of c_x. The intersection $C_f \cap C_g$ is contained in the finitely many points of 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 $\operatorname{rad}(f) \mid g$. In particular, $\operatorname{rad}(f)$ 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 $\operatorname{rad}(f) \mid f^n$, we are done. Finally, $\operatorname{rad}(f)$ is simply the nonzero polnyomial of least degree in $\mathbb{I}(C)$ (up to nonzero scalars). ∎

We say that $\operatorname{rad}(f)$ 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

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

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 $k[x, y]$.

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

Definition 1.7.6. 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 $\operatorname{rad}(f)$. This gives us a perfect translation between alegbra and geometry. For instance, we can use this to define the degree of curve.

Definition 1.7.7 (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 1.7.8 can be restated as

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

Corollary 1.7.8 (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 1.7.9 (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 1.7.8(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 1.7.8(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 1.7.10 (Finite Intersection Revisited). If $C, D \subset \mathbb{A}^2_k$ are two curves that don't share any common 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 1.7.9. 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 1.7.8(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 1.7.11. 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 1.7.12. 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 1.7.13.** 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 1.7.12, this can only happen if $f \mid g$ in $\mathbb{C}[x, y]$, so that $g/f \in \mathbb{C}[x, y] \cap \mathbb{R}(x, y) = \mathbb{R}[x, y]$. ∎

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 1.7.14. 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 2.3.2. 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 1.7.14 has two components. Again, however, $C_f(\mathbb{C})$ is a punctured torus (hence connected, even irreducible) and the two components

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

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 1.7.13 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 1.7.10). 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 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 $f(x,y) = \sum_{i,j\geq 0} a_{i,j}x^iy^j$ 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 $F(tp,tq) = t^dF(p,q)$. 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 $f = \sum_{i=0}^d a_ix^{d-i}y^i$. 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 $f(t,1) = \sum_{i=i_0}^d a_it^{d-i}$ of degree $d - i_0$ as

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

for some $\alpha_j \in k$, and then taking $a_{i_0}^{-1}\ell_1 = \ell_2 = \cdots = \ell_{i_0} = y$ and $\ell_{i_0+j} = x - \alpha_jy$ 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 Lemma **??**), 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 (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 $D : R \to R$ satisfyng 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 $R = k[x_1, x_2, \dots, x_n]$, 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 $R = k[x_1, \dots, x_n]$ be the polynomial ring over R in $n \geq 1$ variables $x_1, \dots, x_n$. Then

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

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

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 $j = 1, \dots, n$, showing uniqueness. Conversely, if $c_1, \dots, c_n$ 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 (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

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

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 $\phi : \mathbb{A}^2_k \to \mathbb{A}^2_k$ 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. If $C \subset \mathbb{A}^2_k$ is a reducible curve and we write $C = D \cup E$ for $D, E \subsetneq C$, then any point $P \in D \cap E$ is a singular point of C. In particular, a curve $C \subset \mathbb{A}^2_k$ is smooth iff its components are both individually smooth and pairwise disjoint.

Proof. Suppose f and g are minimal polynomials for D and E; then a minimal polynomial for C is $\mathrm{rad}(fg)$. Note that f and g are not constant multiples of each other since $D \neq E$ (why?). I'll do the case where f and g are distinct irreducibles, so that C has minimal polynomial fg, leaving the general case to the reader. By Theorem 1.8.8, it suffices to show that if $P \in D \cap E$, 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 1.7.9). For each $1 \leq i < j \leq n$, the intersection $C_i \cap C_j$ is finite by Theorem 1.7.10; 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 1.7.8(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 $C, D \subset \mathbb{A}_k^2$ 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

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

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 $f_1, f_2, g \in k[x,y] \setminus \{0\}$ and $P \in \mathbb{A}_k^2$.
(5) (Coordinate Ring Dependence) $i_P(f,g) = i_P(f, g + hf)$ for all $f,g,h \in k[x,y] \setminus \{0\}$.
(6) (Invariance under ACOCs) If $\phi : \mathbb{A}_k^2 \to \mathbb{A}_k^2$ 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 $\mathcal{O}_P$, 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

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

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 $(f,g)\mathcal{O}_P \subset \ker \mathrm{eval}_P$, so

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

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 $p \in \ker \mathrm{eval}_P$, 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 $x^i y^j$ 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

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

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 of elements (X, Y, Z) of k, not all zero, subject to the equivalence relation that $(X, Y, Z) \sim (cX, cY, cZ)$ for all $c \in k \smallsetminus \{0\} = k^*$, i.e.

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

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 funcitons 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 $\mathbb{A}_k^3$) 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 $P = [X : Y : Z] \in \mathbb{P}_k^2$ 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 $C \subset \mathbb{P}_k^2$ of the form

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

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 $C_F \cap \mathbb{A}_k^2$ 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 $\mathbb{A}_{\mathbb{R}}^2$ connect up to form one "continuous loop" in $\mathbb{P}_{\mathbb{R}}^2$.

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 $C_f \setminus P$? 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?

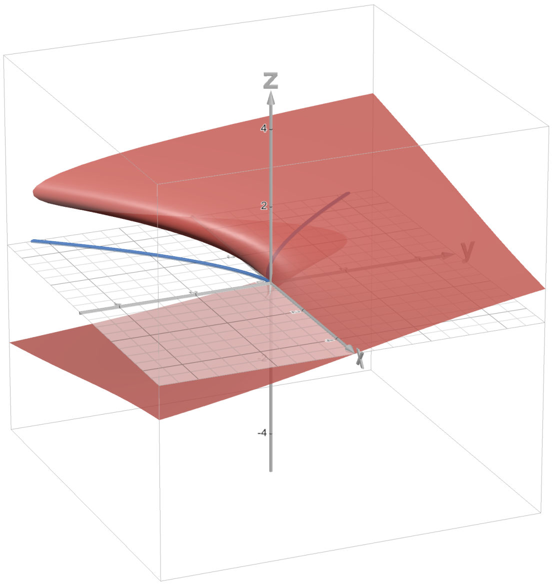


Figure 2.1: The surface $z^3 + xz - y = 0$ when orthogonally projected onto the xy-plane outlines a cuspidal curve. Picture made with Desmos 3D.

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?

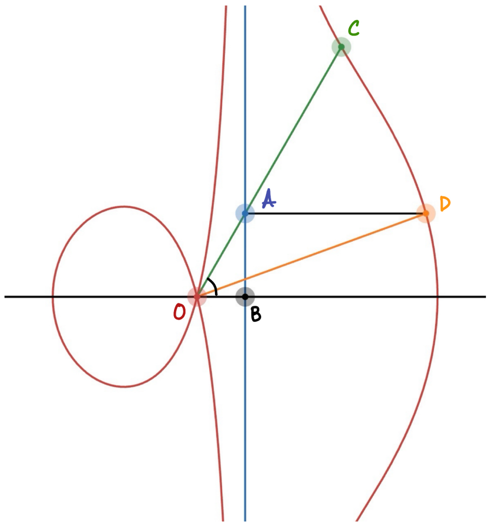


Figure 2.2: The Conchoid of Nichomedes and Angle Trisection. Picture made with Desmos and edited in Notability.

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?

This proof due to Kapferer has been adapted from [5]. With minor modifications, the same proof shows that any over a field k with $\operatorname{ch} k = 0$, every smooth projective curve of degree at least 3 cannot be parametrized by rational functions. In modern algebraic geometry, this result (in arbitrary characteristic) is often seen as a consequence of the Riemann-Hurwitz formula.[5]

[4] If you haven't seen this notion before, then define it.

[5] If you know what that is, do you see why this result is a consequence of it?

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[6] if $0 < e < 1$ and one piece if $e \geq 1$.
(b) The curve C_e is smooth[7] if and only if $e \neq 1$.
(c) For $e > 1$, the unique oval in C_e is convex[8] 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

$$t^{m+n-1}, t^{m+n-2}, \ldots, 1$$

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 $f(x) = a_0 t^m + \cdots + a_m$ and $g(x) = b_0 t^n + \cdots + b_n$. This matrix is called the Sylvester matrix of f and g.

[6] Here the word "piece" means "connected component".
[7] What does that mean?
[8] What does that mean?

(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

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

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 $\mathrm{disc}(f)$, is the resultant of f and its (formal) derivative f' with respect to t, up to scalar factors. More precisely, if $f(t) = a_0t^m + \cdots + a_m$ with $a_j \in K$ and $a_0 \neq 0$, then we define

$$\mathrm{disc}(f) := \frac{(-1)^{m(m-1)/2}}{a_0} \cdot \mathrm{Res}(f, f').$$

Let's do a few examples.

(a) Show that if $f(t) = at^2 + bt + c$, with $a \neq 0$, then $\mathrm{disc}(f) = b^2 - 4ac$.

(b) Show that if $f(t) = t^3 + pt + q$, then $\mathrm{disc}(f) = -4p^3 - 27q^2$. 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

$$\mathrm{disc}(f) = a_0^{2n-2} \prod_{1\le i<j\le n} (\alpha_i - \alpha_j)^2.$$

(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.

(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 2.3.2 (Eisenstein's Irreducibility Criterion)**.**

(a) Let R be a domain and let $f \in R[t]$. Let f have degree $n \geq 1$ and write $f = a_0 t^n + a_1 t^{n-1} + \cdots + a_n$ 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 is 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 2.3.3. 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. $\operatorname{disc}(p) \neq 0$.[9]

[9] 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 $\operatorname{Der}_{\mathbb{Q}}\left(\mathbb{Q}[\sqrt{-3}]\right)$? $\operatorname{Der}_{\mathbb{Q}}(\mathbb{Q}[\sqrt{7}, \cos(2\pi/5)])$? $\operatorname{Der}_{\mathbb{R}}(\mathbb{C})$? $\operatorname{Der}_{\mathbb{Q}}(\mathbb{Q}[\pi])$? $\operatorname{Der}_{\mathbb{Q}}(\mathbb{C})$? $\operatorname{Der}_{\mathbb{F}_p(t)}(\mathbb{F}_p(t)[s]/(s^p - t))$? What about $\operatorname{Der}_k K$, 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 $\phi : \mathbb{P}^2_k \to \mathbb{P}^2_k$ is a projective change of coordinates, 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 $\operatorname{ch} k \neq 2, 3$, 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.

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.

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] F. Lemmermeyer, “Parametrizing Algebraic Curves.” https://doi.org/10.48550/arXiv.1108.6219, 2011.