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 $\mathbb{A}^2_k$ (again think of $k = \mathbb{F}_q$ being a finite field).

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

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

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

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

1.1.1 Motivating Questions

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

Question 1.1.9. Is $C = \emptyset$?

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

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

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

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

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

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

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

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

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

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

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

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

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

[a] What's that?

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

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

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

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

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

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

1.1.2 Some Unimportant Remarks

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

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

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

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

[1] We will define this notion formally later.

[2] Now, what's that?

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

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

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

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

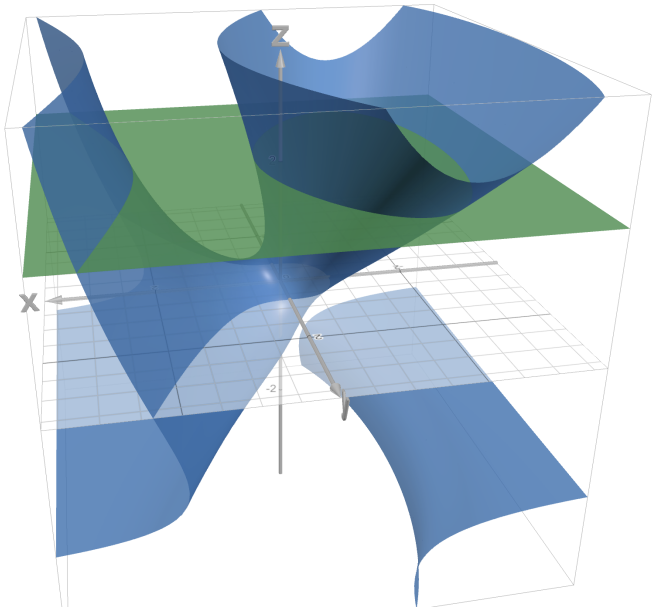


Figure 1.1: The elliptic curve over $k = \mathbb{C}(z)$ defined by $y^2 = x^3 - 3zx + (z^3 + 1)(z + 2)^{-1}$ in blue, along with its hyperplane section at $z = 2$, which is the elliptic curve $y^2 = x^3 - 6x + 9/4$. Picture made with Desmos 3D.

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

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

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

1.2.1 Degree I

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Let's now do some more examples of curves.

1.2.2 Polar Curves

I'll assume some familiarity with polar coordinates.

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

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

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

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

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

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

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

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

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

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

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

$$P_g \subset C_f.$$

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

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

Finally, take

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

■

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

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

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

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

Use Desmos to plot the curves P_g and C_f.

Here are two issues with this approach:

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

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

1.2.3 Synthetic Constructions

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

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

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

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

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

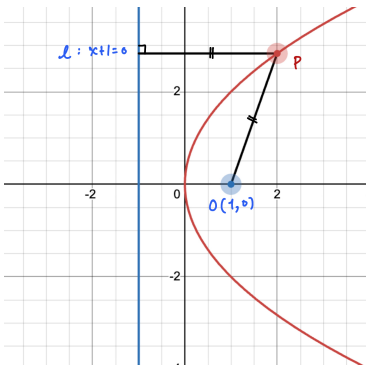


Figure 1.2: The synthetic construction of the parabola. Picture made with Desmos.

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

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

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

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

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

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

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

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

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

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

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

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

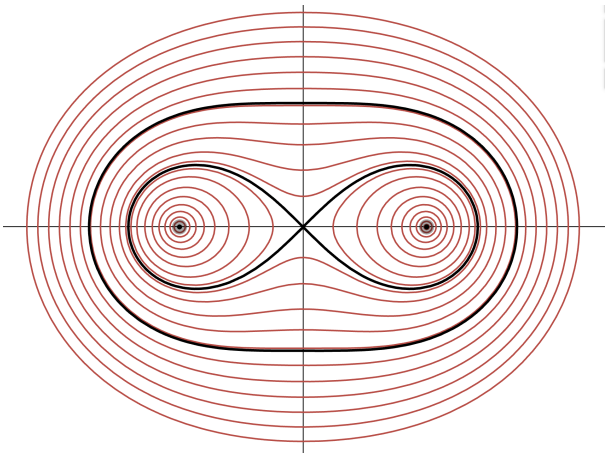


Figure 1.3: The Cassini ovals. Picture made with Desmos.

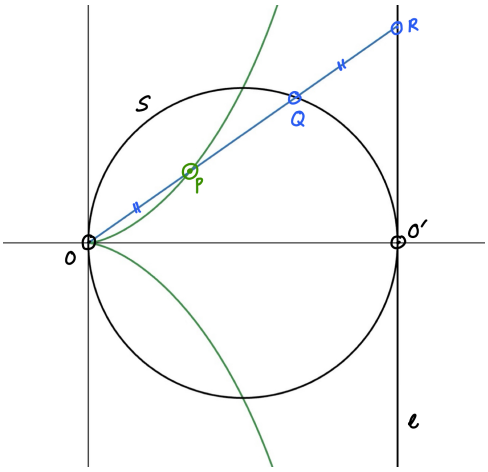
Example 1.2.13 (Cissoid of Diocles)**.** This curve is named after the ancient Greek mathematician Diocles. To construct it, start with a circle $S \subset \mathbb{A}^2_{\mathbb{R}}$ and a point $O \in S$. Construct the diameter OO' to S through O as well as the tangent line ℓ to S through O'. Now for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

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

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

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

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



(a) Cissoid of Diocles. Made with Desmos.

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

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

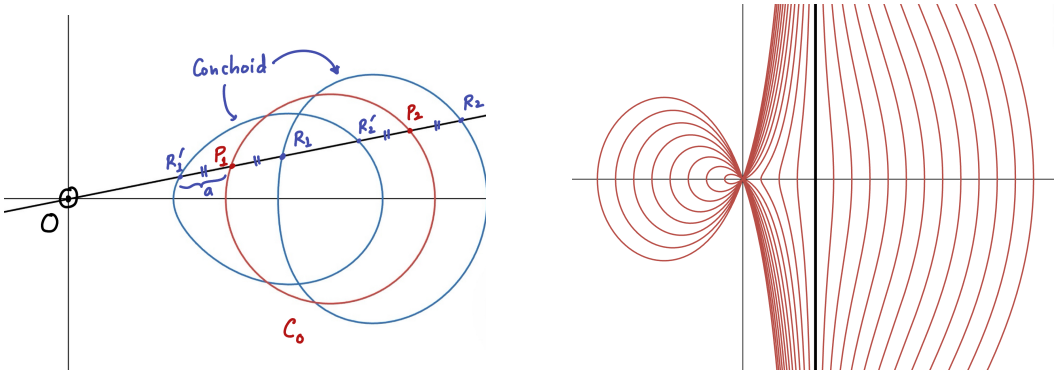
There are many other constructions of this curve: for instance, it is the curve obtained by inverting a parabola in a circle centered at its vertex, and also, if two congruent parabolae are set vertex-to-vertex, and one rolls on the other, then the vertex of the rolling parabola traces out the cissoid.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

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

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

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

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



(a) Conchoid with base curve a circle.

(b) Various conchoids of Nichomedes.

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

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

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

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

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

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

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

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

1.3 06/14/24 - Parametric Curves

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

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

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

This is the line C_ℓ defined by the polynomial

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

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

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

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

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

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

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

Example 1.3.4. For any field k, the subset

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

traces out the hyperbola defined by

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

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

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

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

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

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

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

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

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

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

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

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

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

which yields the point

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

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

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

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

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

See Figure 1.6.

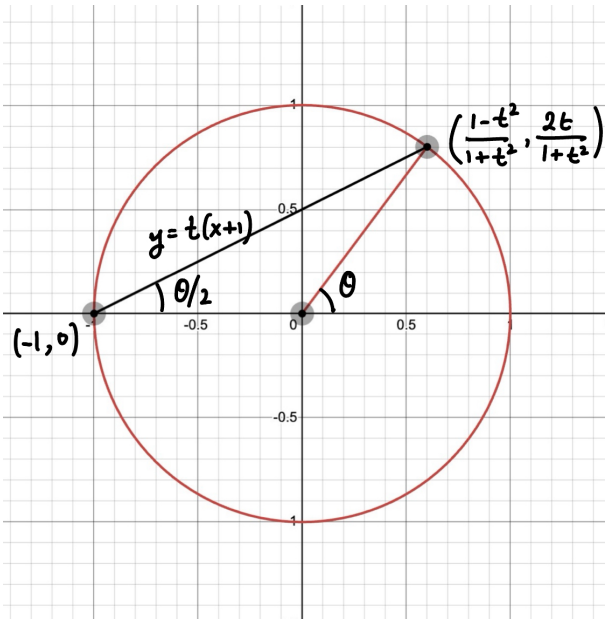


Figure 1.6: Parametrizing the circle $x^2 + y^2 = 1$.

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

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

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

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

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

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

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

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

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

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

from which we conclude that

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

In this case, we can take

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

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

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

■

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

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

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

If we let

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

then it is clear that

$$C \subset C_f.$$

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

$$C = C_f.$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

where

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

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

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

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

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

where

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$C \subset C_f.$$

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

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

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

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

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

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

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

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

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

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

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

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

1.4.1 Affine Changes of Coordinates

Definition 1.4.1. An affine change of coordinates is a transformation

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

of the form

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

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

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

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

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

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

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

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

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

1.4.2 Algebraically Closed Fields

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[11] What would that mean?

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

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

Chapter 2

Exercise Sheets

2.1 Exercise Sheet 1

2.1.1 Numerical and Exploration

Exercise 2.1.1. For an ordered pair (a, b) of rational numbers, consider the polynomial

$$f_{a,b}(x, y) := ax^2 + by^2 - 1 \in \mathbb{Q}[x, y].$$

Let $C(a, b) = C_{f_{a,b}} \subset \mathbb{A}^2_{\mathbb{Q}}$ be the rational affine plane algebraic curve defined by $f_{a,b}$.

(a) Show that $C(2/5, 1/5) = \emptyset$.
(b) Characterize all primes p such that $C(1/p, 1/p) = \emptyset$.
(c) Characterize all pairs (a, b) such that $C(a, b) = \emptyset$.

Exercise 2.1.2.

(a) Play around with graphs of real affine plane algebraic curves (RAPACs) on, say, Desmos or WolframAlpha. What is the coolest thing you can get a graph to do (cross itself thrice, look like a heart, etc.)?
(b) How many pieces (i.e. connected components) can a RAPAC of degree $d = 2$ have? How about $d = 3$? What about $d \in \{4, 5, 6, 7\}$?
(c) What can you say in general? Can you come up with upper or lower bounds for the number of pieces?
(d) Does the number of pieces depend on the nesting relations[1] between them? Does it depend on (or dictate) their shapes (e.g. convexity)?[2]

Exercise 2.1.3.

(a) Let $P \subset \mathbb{A}^2_{\mathbb{R}}$ be the polar curve implicitly defined by the equation

$$r^3 + r\cos\theta - \sin 4\theta = 0.$$

Find a nonconstant polynomial $f(x, y) \in \mathbb{R}[x, y]$ such that the curve $C_f \subset \mathbb{A}^2_{\mathbb{R}}$ defined by f contains P, i.e. satisfies $P \subset C_f$.[3]
(b) What is the degree of your f? What is the smallest possible degree of such an f?
(c) By your choice of f, we have the containment $P \subset C_f$. Is P all of C_f? If so, can you explain why (perhaps by retracing steps)? If not, how would you describe the extraneous components of $C_f \smallsetminus P$? Could you have predicted them? Can you pick an f that provably minimizes the number of extraneous components?
(d) Repeat the same analysis as in (a) through (c) for other such implicitly defined polar curves of your own devising.
(e) Can you perform the same analysis as above for the Archimedean spiral, which is the polar curve implicitly defined by the equation $r = \theta$?

Draw pictures, or get a computer to draw them for you, but beware–is your software doing exactly what you think it is?

[1] What does that mean? What are those?

[2] Here's a harder result to whet your appetite: if $d = 4$ and there is a nested pair of closed ovals, then the inner oval must be convex and there cannot be more components, although there may be up to 4 non-convex components in general. You may not be able to prove this now, but you should be able to solve this problem by the end of the course.

[3] I like to use the symbol $\subset$ to mean "is contained in or equal to". Others prefer the symbol $\subseteq$ to denote the same thing. I will use the symbol $\subsetneq$ when I want to exclude the possibility of equality.

Exercise 2.1.4. Consider the surface defined by the equation $z^3 + xz - y = 0$, pictured in Figure 2.1. The orthogonal projection of this surface to the xy-plane outlines a cuspidal curve.

(a) Find the equation describing this cuspidal curve, and prove the assertion made above.
(b) How does all of this relate to the Cardano formula for the solution to the cubic equation?

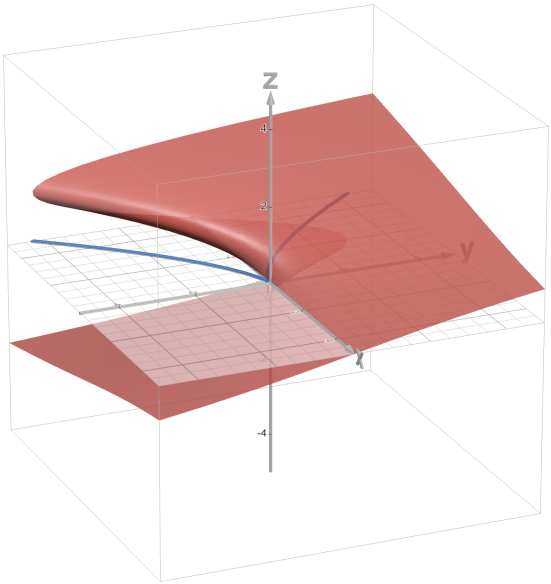


Figure 2.1: The surface $z^3 + xz - y = 0$ when orthogonally projected onto the xy-plane outlines a cuspidal curve. Picture made with Desmos 3D.

Exercise 2.1.5. Can you find a way to use the conchoid of Nichomedes (Example 1.2.14) to trisect a given angle? You may suppose that you know how to construct a conchoid with any given parameters. (Hint: see Figure 2.2.) Once you've done that, use the cissoid of Diocles to give a compass and ruler (and cissoid) construction of $\sqrt[3]{2}$, or of $\sqrt[3]{a}$ for any given $a > 0$. How far can you take this–what else can you do with the cissoid and conchoids of different parameters? Why do these constructions not contradict results from Galois theory?

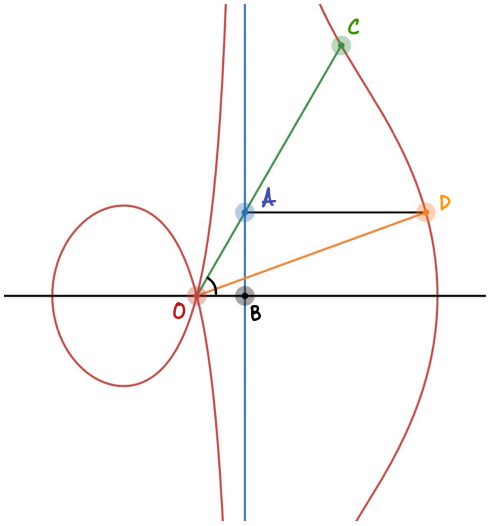


Figure 2.2: The Conchoid of Nichomedes and Angle Trisection. Picture made with Desmos and edited in Notability.

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 [2]. 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

$$\mathrm{Res}(f,g) := \det \mathrm{Syl}(f,g) \in \mathbb{Z}[a_0,\dots,a_m,b_0,\dots,b_n] \subset K.$$

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 $a_0 b_0 \neq 0$, then f and g have a common root $t = t_0 \in K$ iff

$$\mathrm{Res}(f,g) = 0.$$

(What happens if $a_0 b_0 = 0$?) Use this to show that, even if K is not algebraically closed, and $\alpha_1,\dots,\alpha_m$ and $\beta_1,\dots,\beta_n$ 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

$$f(t) = a_1 t^2 + b_1 t + c_1 \text{ and}$$
$$g(t) = a_2 t^2 + b_2 t + c_2,$$

then

$$\mathrm{Res}(f,g) = (a_1 c_2 - a_2 c_1)^2 - (a_1 b_2 - a_2 b_1)(b_1 c_2 - b_2 c_1).$$

In particular, these quadratic equations have a common root (in K, or if necessary, a quadratic extension of K) iff this polynomial of degree 4 in the coefficients vanishes.

(f) (Finishing Example 1.3.10.) Show that if $u(t), v(t) \in k[t]$ are any nonconstant polynomials which define the parametric curve

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

and if

$$f(x,y) := \mathrm{Res}_t(u(t) - x, v(t) - y) \in k[x,y],$$

then $C \subset C_f$ with equality if k is algebraically closed.

Exercise 2.2.5 (Discriminants). Given a field K and a polynomial $f(t) \in K[t]$, the discriminant of f, written $\mathrm{disc}(f)$, is the resultant of f and its (formal) derivative f' with respect to t, up to scalar factors. More precisely, if $f(t) = a_0 t^m + \cdots + a_m$ 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

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

(d) Show that the polynomial $f(t)$ has a repeated root over an algebraic closure of K iff $\operatorname{disc}(f) = 0$. In other words, if there is an α some extension field $K' \supset K$ and a polynomial $q(t) \in K'[t]$ such that

$$f(t) = (x - \alpha)^2 q(t),$$

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.

Bibliography

[1] E. Brieskorn and H. Knörrer, *Plane Algebraic Curves*, vol. 38. Springer Basel AG, 1986.

[2] F. Lemmermeyer, "Parametrizing Algebraic Curves." `https://doi.org/10.48550/arXiv.1108.6219`, 2011.