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

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

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

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

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

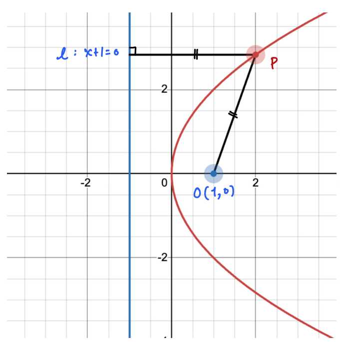


Figure 1.2: The synthetic construction of the parabola. Picture made with Desmos.

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.

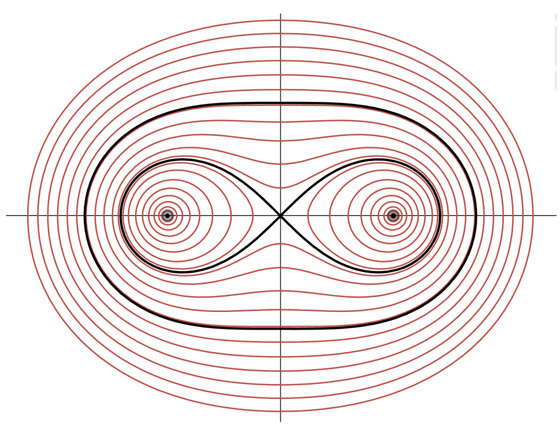


Figure 1.3: The Cassini ovals. Picture made with Desmos.

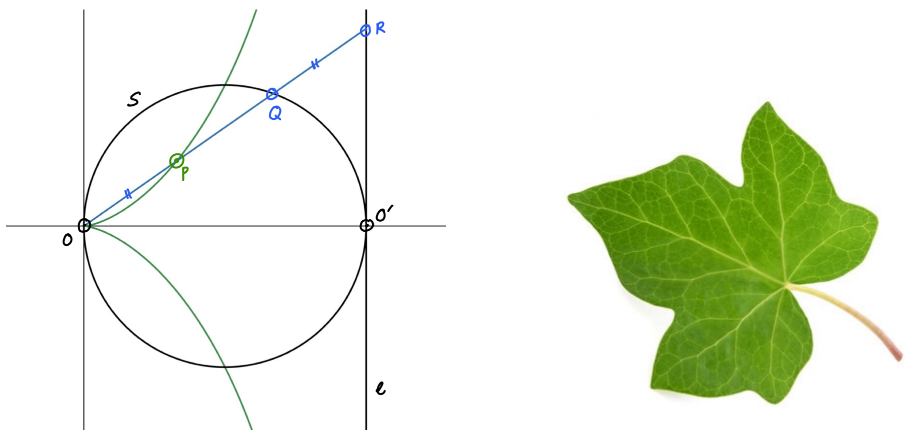
Example 1.2.13 (Cissoid of Diocles)**.** This curve is named after the ancient Greek mathematician Diocles. To construct it, start with a circle $S \subset \mathbb{A}^2_{\mathbb{R}}$ and a point $O \in S$. Construct the diameter OO' to S through O as well as the tangent line ℓ to S through O'. Now for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

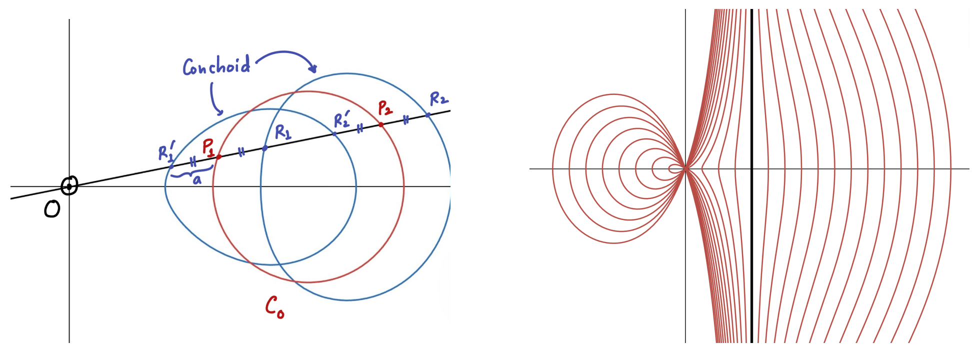
There are many other constructions of this curve: for instance, it is the curve obtained by inverting a parabola in a circle centered at its vertex, and also, if two congruent parabolae are set vertex-to-vertex, and one rolls on the other, then the vertex of the rolling parabola traces out the cissoid.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

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

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].