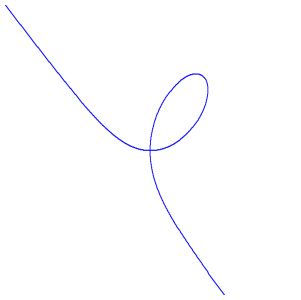
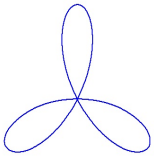
Algebraic Curves

What are curves? Generally speaking, a **curve** is the **zero set** or variety $V(f)$ of a function $f : \mathbb{R}^2 \to \mathbb{R}$, i.e., the set of points $(x, y) \in \mathbb{R}^2$ such that $f(x, y) = 0$. Figure 1 illustrates some famous curves.

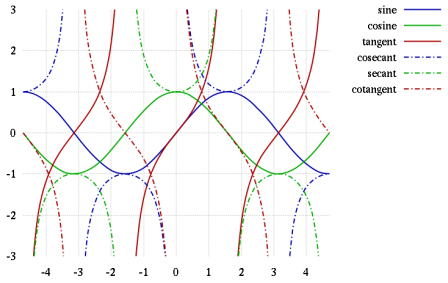


Folium of Descartes: $x^3 + y^3 = 3xy$

Cissoid of Diocles: $(x^2 + y^2)x = 2y^2$



Trifolium $y^4 - y^3 + 2x^2y^2 + 3x^2y + x^4$



Graphs of trigonometric functions

Figure 1: Some famous curves:

Even though our general description of a curve encompasses both a circle $x^2 + y^2 = 1$ and the graphs of the trigonometric functions, our intuition says that the two are fundamentally distinct. One distinction that is apparent is that any line in the plane intersects the circle in only finite points (we will see the proof later), whereas the line $y = 0$ intersects the graph of the trigonometric functions infinitely often. In this lecture, we will focus on curves of the first kind. These curves are called **algebraic curves**, since the defining function is usually a bivariate polynomial, which is an element of the ring $\mathbb{Z}[x, y]$.

¶1. Two choices for representation Consider a line $ax + by = c$ in the plane. Clearly, it defines a curve. The same curve can also be represented as the pair $(x(t), y(t))$, where $x(t) := t$ and $y(t) := (a/b)t - (c/b)$; that is, we can directly give the points on the line in terms of a parameter t. Thus there appear to be two ways to represent the same curve: the **implicit form** is to define the curve as the zero set of a bivariate polynomial; the **parametric form** describes almost all, except finitely many, points on the curve in the form $(x(t), y(t))$, where t is a parameter. The implicit form of the unit circle is the familiar equation $x^2 + y^2 = 1$. But what is its parametric form? We know that all points on the unit circle can be described as $(\cos\theta, \sin\theta)$, $\theta \in [0, 2\pi)$. So this is a valid parametrization. However, let us be restrictive and aim for a parametric form that doesn't use trigonometric functions. By making the substitution $t := \tan\theta/2$, it follows that the desired parametrization is $\left(\frac{1-t^2}{1+t^2}, \frac{2t}{1+t^2}\right)$, where $t = (-\infty, \infty)$. A

geometric interpretation of this parametrization is to fix the point $(-1, 0)$ on the circle and consider all the lines $\ell(t)$ through it with slope t; each $\ell(t)$ intersects the circle at precisely one point, and the coordinates of that point can be obtained by plugging the equation of $\ell(t) = t(x + 1)$ into the equation of the circle to get $x^2 + t^2(x + 1)^2 = 1$ and solve for x in terms of t. Note that we cannot expect $x(t), y(t)$ to be polynomials (Why?).

Though the parametric form has its merits, it has its drawbacks as well: all curves are not "nicely" parametrizable, where nicely can be interpreted as a rational function of univariate polynomials. The study of parameterized curves forms the rich field of differential geometry. In this lecture, we restrict ourselves to study curves in the implicit form, which is the foundation of the field of algebraic geometry.

1 Affine Transforms – Equivalence of Geometry

It is natural to think of two curves as "equivalent" if one can be obtained from the other by a translation of the coordinates, a rotation of the coordinates, or a scaling. Can we generalize this notion? What if we consider the set of invertible linear transformations, in addition to translation? We show that certain concepts are invariant under such transformations, called **affine transformations**. More precisely, an affine map $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ is of the form $\phi(\mathbf{x}) = A\mathbf{x} + \mathbf{t}$, $\mathbf{x} \in \mathbb{R}^2$, where A is an invertible linear map and $\mathbf{t} \in \mathbb{R}^2$ is a translation vector. Note that the set of affine transformations is a group *under composition*; the inverse of $A\mathbf{x} + \mathbf{t}$ is the map $A^{-1}\mathbf{x} - A^{-1}\mathbf{t}$.

We say **two curves f, g are affinely equivalent** if there exists an affine map and a scalar $\lambda \neq 0$ such that $g = \lambda f(\phi(\mathbf{x}))$; note that scaling by a constant cannot be obtained by an affine transformation.

1. What curves are equivalent to a line $y = mx + b$? By the map $x \to (x - b)/m$, we get that the line is equivalent to $y = x$; by mapping $y \to y + x$, we get that the line is equivalent to $y = 0$. Thus all lines are affinely equivalent to $y = 0$, and since affine transformations are invertible under composition, it follows that the set of lines are affinely equivalent.

2. What curves are affinely equivalent to an ellipse $(x/a)^2 + (y/b)^2 = 1$? By the scaling $x \to ax$ and $y \to by$, it follows that all ellipses are affinely equivalent to the circle $x^2 + y^2 = 1$.

3. What curves are affinely equivalent to a parabola $y^2 = 4ax$? By scaling $x \to ax$ and $y \to 2ay$, it follows that all parabolas are affinely equivalent to the standard parabola $y^2 = x$.

4. We can similarly show that a general hyperbola $(x/a)^2 - (y/b)^2 = 1$ is equivalent to a standard rectangular hyperbola $x^2 - y^2 = 1$.

Thus the notion of affine equivalence helps us reduce the number of different types of curves to look at. But what properties of the curve remain affinely invariant? Note that the affine map in each of the cases above somehow maintained the degree of the curve. The **degree** of a curve $f(x, y) = \sum_{i,j} a_{i,j} x^i y^j$ is the degree of the largest monomial appearing in the polynomial; from now on, we use d to represent the degree of f. Based upon what we observed in the examples above, we claim the following:

LEMMA 1. *The degree of a curve is affinely invariant.*

Proof. Let $\phi(\mathbf{x}) = ((px + qy + a), (rx + sy + b))$, where $ps - rq \neq 0$, and $f(x, y) = \sum_{i,j} a_{i,j} x^i y^j$. Suppose

$$g = \lambda f(\phi(\mathbf{x})) = \lambda \sum_{i,j} a_{i,j} (px + qy + a)^i (rx + sy + b)^j.$$

Since $ps - rq \neq 0$, both the linear polynomials $(px + qy + a)$ and $(rx + sy + b)$ are non-zero, and hence $\deg(g) \leq \deg(f)$. Similarly, $f = \lambda^{-1} g(\phi^{-1}(\mathbf{x})$ implies that $\deg(g) \geq \deg(f)$. Thus $\deg(f) = \deg(g)$. **Q.E.D.**

Besides affine equivalence, there's a nice geometric interpretation of the degree. Note that any line in the plane meets another line in at most one point, meets a circle in at most two points, meets a hyperbola in at most two points. It appears that the degree of a curve is an upper bound on the number of intersections of a line with a curve, and indeed it is almost the case.

LEMMA 2. *A line ℓ intersects a degree d curve f in at most d places, unless it's a component of the f.*

Proof. Let $(x(t), y(t))$ be a parametrization of ℓ. Then the points of intersections of ℓ with f are the real roots of the univariate polynomial $g(t) := f(x(t), y(t))$ of degree $\leq d$, unless g is identically zero. By an affine transformation, we can assume that ℓ is the line $y = 0$. Let $f(x, y) = yf_0(x, y) + f_1(x)$. Then $g = f(x, 0) = f_1(x) \equiv 0$. Thus $f(x, y) = yf_0(x, y)$, and hence the line y is a component of f. **Q.E.D.**

Remark: As a consequence of the lemma above, we see that curves such as the graphs of trigonometric functions cannot be algebraic curves.

In the proof above, we have implicitly assumed that affine maps do not change the intersection pattern of a line with a curve, and also that the intersection pattern doesn't depend upon the choice of parameterization of line ℓ . We now make this concept more precise.

Let ℓ be a line which is not a component of $f(x, y)$ and $\mathbf{p} \in \mathbb{R}^2$ be a point common to ℓ and f. Consider a parametrization $(1 - t)\mathbf{a} + t\mathbf{b}$ of ℓ, where $\mathbf{a}, \mathbf{b} \in \ell$; suppose $\mathbf{p} = (1 - t_0)\mathbf{a} + t_0\mathbf{b}$. Then the **intersection number**, $I(\mathbf{p}, f, \ell)$, of the point $\mathbf{p}$ is the multiplicity of t_0 as a root of the univariate polynomial $f((1 - t)\mathbf{a} + t\mathbf{b})$, called the **intersection polynomial**, which is the same as the largest power of $(t - t_0)$ dividing the intersection polynomial.

From the definition it appears that $I(\mathbf{p}, f, \ell)$ depends upon the choice of the parameterization. We claim that this is not the case. Suppose $\ell = (1 - u)\mathbf{c} + u\mathbf{d}$ is another paramaterization of ℓ and let $\mathbf{a}, \mathbf{b}$ correspond to α, β, resp. in this parametrization. Then

$$\begin{aligned}(1 - t)\mathbf{a} + t\mathbf{b} &= (1 - t)((1 - \alpha)\mathbf{c} + \alpha\mathbf{d}) + t(1 - \beta)\mathbf{c} + t\beta\mathbf{d} \\ &= (1 - u(t))\mathbf{c} + u(t)\mathbf{d}\end{aligned}$$

where $u(t) := \alpha - t(\alpha - \beta)$. Thus if $\mathbf{p}$ corresponds to t_0 in the first parametrization, then it corresponds to $u(t_0)$ in the second. Let $\eta(t) := f((1-t)\mathbf{a}+t\mathbf{b})$ and $\psi(u) := f((1-u)\mathbf{c}+u\mathbf{d})$ be the intersection polynomials w.r.t. the two parameterizations. Then we know that $\eta(t) = \psi(u(t))$. Thus, t_0 occurs with multiplicity m in $\eta(t)$ iff $u(t_0)$ occurs with the same multiplicity in $\psi(u)$. Therefore, the multiplicity of a point $\mathbf{p}$ common to ℓ and f is independent of the choice of parametrization of ℓ.

We further claim that $I(\mathbf{p}, f, \ell)$ is invariant under affine maps. Let $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ be an affine map. The map ϕ takes $\mathbf{p} \to \phi(\mathbf{p})$, and the line $\ell = (1 - t)\mathbf{a} + t\mathbf{b}$ to $\ell' := (1 - t)\phi(\mathbf{a}) + t\phi(\mathbf{b})$. However, the curve f is transformed to $f' := f(\phi^{-1})$, because $\mathbf{p}$ is on f iff $\phi(\mathbf{p})$ is on f', which means the domain of f must be transformed by ϕ^{-1} (and not ϕ). Thus the intersection polynomial of ℓ' and f' is

$$f'((1 - t)\phi(\mathbf{a}) + t\phi(\mathbf{b})) = f \circ \phi^{-1}((1 - t)\phi(\mathbf{a}) + t\phi(\mathbf{b})) = f((1 - t)\mathbf{a} + t\mathbf{b})$$

which is the same as f intersecting with ℓ. Thus $I(\mathbf{p}, f, \ell) = I(\phi(\mathbf{p}), f', \ell')$.

¶2. Singular Points: Consider the origin, the curve $y^2 = x^3$, and any line $(t, \lambda t)$ through the origin. The intersection polynomial is $t^2(\lambda - t)$, and thus any line through origin intersects the curve at least twice at the origin; the line $y = 0$ intersects three times. Such points are of considerable interest in our study. The **multiplicity of a point $\mathbf{p}$** on the curve f is the smallest value of $I(\mathbf{p}, f, \ell)$ over all lines ℓ through $\mathbf{p}$. Since $\mathbf{p}$ is on f, $I(\mathbf{p}, f, \ell) \geq 1$. However, there are points, such as the origin in our example $y^2 = x^3$, for which the multiplicity m is ≥ 2, such points are called **singular points**; intuitively, the curve has m branches at $\mathbf{p}$ (though this is not always the case). Our definition, however, is not constructive in nature. Another way to define singular points is as follows: a point $\mathbf{p} = (a, b)$ is a singular point if the smallest non-vanishing monomial in the shifted polynomial $f(x + a, y + b)$ has degree at least two. Let's see why this is equivalent to our earlier definition. Let $f(x + a, y + b) = F_m + F_{m+1} + \cdots + F_d$, where each F_i, $i \geq m \geq 2$, is a homogeneous polynomial of degree i, and $F_m \not\equiv 0$. Any line through (a, b) can be expressed in the parametric form $(a + \alpha t, b + \beta t)$, for different choices of $(\alpha, \beta) \in \mathbb{R}^2$. Thus the intersection polynomial is

$$f(a + \alpha t, b + \beta t) = F_m(\alpha t, \beta t) + F_{m+1}(\alpha t, \beta t) + \cdots = t^m(F_m(\alpha, \beta) + tF_{m+1}(\alpha, \beta) + \cdots).$$

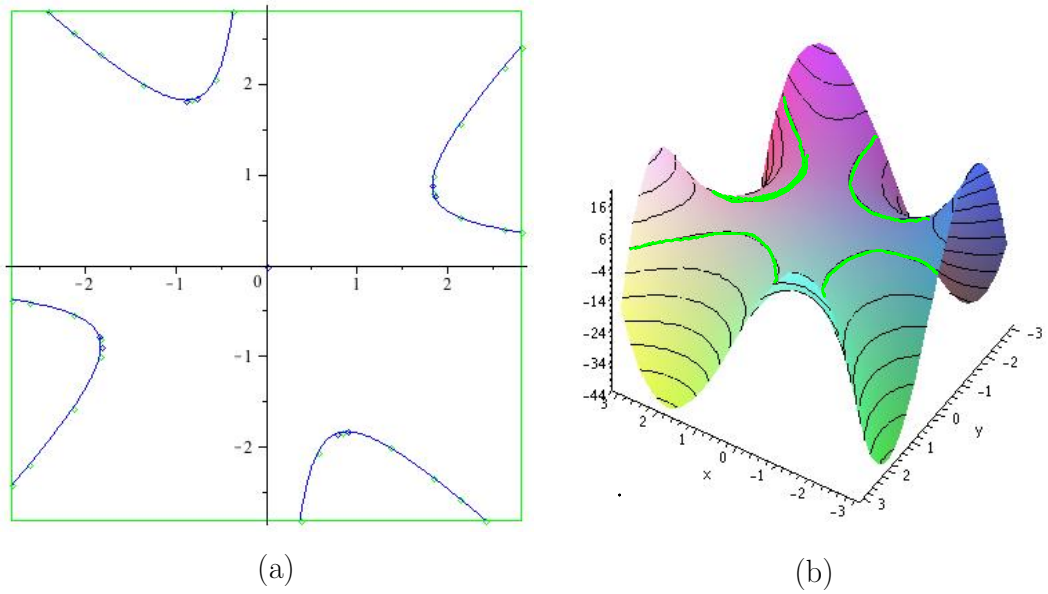
Thus $I(\mathbf{p}, f, \ell) \geq 2$ for all lines ℓ through $\mathbf{p}$. This definition gives us more insight into singular points, however, computationally it is still not good since it involves doing Taylor shifts in both variables. An easier approach is to look at Taylor expansion of f at (a, b):

$$f(x + a, y + b) = f(a, b) + f_x(a, b)x + f_y(a, b)y + \sum_{k\geq 2} \frac{1}{k!} \sum_{i=0}^{k} \binom{k}{i} \frac{\partial^k f}{\partial x^i \partial y^{k-i}} x^i y^{k-i}. \tag{1}$$

Substitute $x = \alpha t, y = \beta t$. For t^2 to divide the resulting intersection polynomial, we want the first three terms to vanish. *Thus* (a, b) *is a singular point of* f *iff*

$$f(a, b) = f_x(a, b) = f_y(a, b) = 0. \tag{2}$$

In fact, (a, b) has multiplicity m iff all the partial derivatives of f of order $< m$ vanish at (a, b). Points with multiplicity one are called simple points, multiplicity two are called double points, and so on. Our intuition that a point with multiplicity m has m branches is, however, always not correct. The Maltese cross is a quartic $f = xy(x^2 - y^2) - (x^2 + y^2)$. The partial derivatives are $f_x = 3x^2y - y^3 - 2x$ and $f_y = x^3 - 3xy^2 - 2y$. Clearly, origin is a singularity with multiplicity two, but are there two branches of f at origin? If we see the plot of f then we observe that origin is in fact an isolated singularity; see Figure 2.



(a) (b)

Figure 2: The Maltese cross: (a) The plot in $\mathbb{R}^2$; (b) The surface in 3-d, with the contour lines corresponding to the variety highlighted.

¶3. Tangents: Given a point $\mathbf{p} = (a, b)$ on a curve f, a tangent to f at $\mathbf{p}$, roughly speaking, is a first order or linear approximation to f at $\mathbf{p}$. More precisely, a **tangent** to f a point $\mathbf{p}$ of multiplicity m is a line ℓ such that $I(\mathbf{p}, f, \ell) > m$. How do tangents look like at simple points, double points etc.? Consider a simple point first. From the Taylor expansion at $\mathbf{p}$ (see (1)) it follows that the parametrized line $(a + \alpha t, b + \beta t)$ is a tangent iff $f_x(\mathbf{p})\alpha + f_y(\mathbf{p})\beta = 0$. Thus the equation of tangent at a simple point (a, b) is given as

$$(x - a)f_x(\mathbf{p}) + (y - b)f_y(\mathbf{p}) = 0.$$

Another way to interpret this is that the tangent is orthogonal to the gradient vector $\nabla f := (f_x, f_y)$ at a simple point. Let's consider the case when $\mathbf{p}$ has multiplicity $m \geq 2$. We know that the expansion of $f(x + a, y + b)$ is of the form

$$F_m + F_{m+1} + \cdots$$

where F_i is a homogeneous polynomial in x, y of degree i. For a line $(a + \alpha t, b + \beta t)$ to have a multiplicity of intersection greater than m at (a, b) we want that $F_m(\alpha, \beta) = 0$. Since F_m is a degree m homogeneous polynomial, we know that it can be factored into linear forms

$$F_m(x, y) = \prod_{i=1}^{m} (a_i x + b_i y).$$

where $a_i, b_i \in \mathbb{C}$. Thus $F_m(\alpha, \beta) = 0$ iff $\alpha/\beta = -b_i/a_i$ for some i. The equations of the m tangents at (a, b) are thus given as

$$(x - a)a_i + (y - b)b_i = 0.$$

Note, however, only the tangents with $a_i, b_i \in \mathbb{R}$ are in the plane $\mathbb{R}^2$. Also, each tangent occurs with a multiplicity, namely the multiplicity of (a_i, b_i) in the factorization of F_m. Points with all m tangents distinct are called ordinary points. Let us look at the tangents to the trifolium $y^4 - y^3 + 2x^2y^2 + 3x^2y + x^4$ at origin. The lower order terms are $3x^2y - y^3 = y(3x^2 - y^2) = y(\sqrt{3}x + y)(\sqrt{3}x - y)$. Thus we have three distinct tangents at origin; see Figure 3. For the Maltese cross, the tangents at origin have imaginary coefficients, $(x \pm iy)$, and hence we do not see any branch of f at the origin in the real plane. This example highlights the fact that we should not always expect branches of the curve to meet at a singularity; this unintuitive phenomenon occurs because we're looking at the variety in $\mathbb{R}^2$.

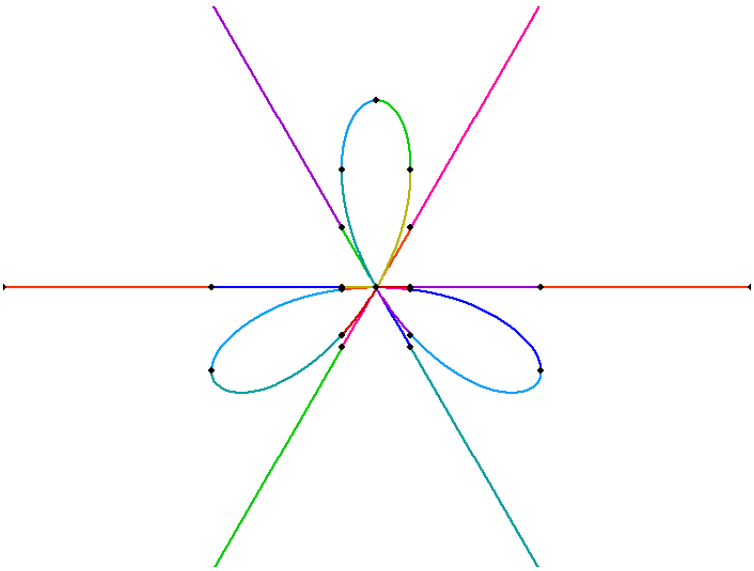


Figure 3: The Trifolium and its three tangents at the origin.

2 Topology of Curves

Our overall aim in these lectures is the following: given an algebraic curve in implicit form $f(x, y) = 0$, determine a "good approximation" of how the curve looks in the plane $\mathbb{R}^2$. What does it mean to have a good approximation? What does it mean to have an approximation to $V(f)$? We first describe what we mean by an approximation, and then what we mean by a good approximation.

Our aim can be described in general as follows: given a set $X \subseteq \mathbb{R}^2$ we want to compute a set $Y \subseteq \mathbb{R}^2$ such that the two are related by some "nice" function $\gamma : X \to Y$. Our intuitive notion of ϵ-approximation is the following: for every point $x \in X$ there is a point $y \in Y$ such that $\|x - y\| \leq \epsilon$. But is this sufficient? Not really, because there may be points in Y that are not ϵ-close to any point in X. Thus we also desire that every point in Y is ϵ-close to some point in X. But this notion is precisely captured by the notion of **Haussdorf distance** between two sets:

$$d_H(X, Y) := \max \left\{ \sup_{x \in X} \inf_{y \in Y} \|x - y\|, \sup_{y \in Y} \inf_{x \in X} \|x - y\| \right\}. \tag{3}$$

Our notion of ϵ-approximation then means that $d_H(X, Y) \leq \epsilon$. It may appear that this notion is sufficient for a good approximation. But ϵ-close does not guarantee that the shapes of the two sets is the same; see Figure **??**. We next describe what properties should the mapping $\gamma : X \to Y$ possess. The proper concept to address these questions is the notion of topology, since topology is the study of properties of objects under transformations independent of the underlying metric.

2.1 Which notion of Topology?

Before we proceed, we need some basic notions from Topology. An **open set** $S \subseteq \mathbb{R}^2$ is a set such that for all points $\mathbf{p} \in S$, there is an $\epsilon > 0$ small enough such that the disc centered at $\mathbf{p}$ with radius ϵ is contained in S. Given a set $X \subseteq \mathbb{R}^2$, a subset $Z \subseteq X$ is said to be open w.r.t. X, if there exists an open set $S \subseteq \mathbb{R}^2$ such that $Z = S \cap X$. For example, if $X = [0, 1)$ then $[0, 1/2)$ is open, $(0, 1/2)$ is open, but $(0, 1/2]$ is not open. A map $\gamma : X \to Y$ is **continuous** if for every set $B \subset Y$ open w.r.t. Y there is a set $A \subset X$ open w.r.t. X such that $\gamma(A) = B$.[1]

Coming back to our **Homeomorphism**

Isotopic

2.2 Locating Critical Points

To detect the correct topology of f we roughly do the following: Imagine a vertical line sweeping the input box from left to right. This line intersects the curve at some points. As we sweep continuously, these points trace the arcs of our curve. Most of the time during the sweep, arcs continue in the direction of the sweep without meeting, except in two cases:

1. the sweep-line becomes a vertical asymptote to the curve (in which case two arcs join); and
2. the sweep-line crosses a singularity, in which case the arcs change their vertical ordering.

This algorithm is essentially the Bentley-Ottmann sweep algorithm for planar line segments. One essential ingredient is to find the two types of points: namely, where the curve has vertical asymptote, and its singularities.

Let us start with how to find **critical points**, that is, points $(\alpha, \beta) \in \mathbb{R}^2$ such that $f(\alpha, \beta) = f_y(\alpha, \beta) = 0$. Another way to think about this is that we want to find a common root to the two *univariate polynomials* $f(\alpha, y)$ and $f_y(\alpha, y)$. We know from the theory of resultants that the two polynomials have a common root if $\texttt{res}(f(\alpha, y), f_y(\alpha, y)) = 0$. We could have performed this test, but the problem is that we do not know α. What if we compute $\texttt{res}_y(f, f_y) \in \mathbb{Z}[x]$, i.e., the resultant of f and f_y treating them as polynomials in y with coefficients in $\mathbb{Z}[x]$, and substitute $x = \alpha$? Is $\texttt{res}_y(f, f_y)(\alpha) = 0$? It is almost true that α is a root of the univariate polynomial $\texttt{res}_y(f, f_y)$ modulo some caveats.

¶4. Resultants of Bivariate Polynomials Let $f, g \in \mathbb{Z}[x, y]$ be such that $\deg_y(f) = m$ and $\deg_y(g) = n$. Treat f, g as polynomials in y with coefficients in $\mathbb{Z}[x]$. The gcd of f, g is defined as

$$\texttt{GCD}(f, g) = \texttt{GCD}(\texttt{cont}(f), \texttt{cont}(g))\texttt{GCD}(\texttt{prim}(f), \texttt{prim}(g)). \tag{4}$$

Note that the content is in $\mathbb{Z}[x]$[2]. From the univariate setting, we know that the resultant of two univariate polynomials is zero iff they have a non-trivial gcd. Can we say something similar for bivariate polynomials? For instance, can we say $\texttt{res}_y(f, g) \equiv zero$ iff $\deg(\texttt{GCD}(f, g)) > 0$? Consider the two polynomials $f = (x^3 + y^3) * (x - 1)$ and $g = (x+y+1)*(x-1)$. Clearly, they have a non-trivial gcd, but their resultant w.r.t. y is ..... Since we are treating f, g as polynomials in y with coefficients in $\mathbb{Z}[x]$, what we can say is that $\texttt{res}_y(f, g) \equiv zero$ iff $\deg_y(\texttt{GCD}(f, g)) > 0$. The argument is essentially the same as in the univariate case: $\deg_y(\texttt{GCD}(f, g)) \geq k$ iff there exists two polynomials P, Q, $\deg_y(P) \leq n - k$ and $\deg_y(Q) \leq m - k$ such that $fP + gQ = 0$. Thus, $\texttt{res}_y(f, g) \equiv 0$ iff $\deg_y(\texttt{GCD}(f, g)) > 0$. [3] Thus to get a non-trivial resultant, we should assume that f, g are **relatively prime**, i.e., their gcd is an absolute constant.

Given two relatively prime polynomials f, g and an $\alpha \in \mathbb{C}$, we want to know when is

$$\texttt{res}_y(f, g)(\alpha) = \texttt{res}(f(\alpha), g(\alpha)). \tag{5}$$

Note the difference between the two sides: on the LHS, we first compute the resultant as a polynomial and then substitute $x = \alpha$; on the RHS, we first compute the polynomials $f(\alpha, y)$ and $g(\alpha, y)$ and then compute their resultant. E.g., suppose $f = (x + 1)y + 2x^2$ and $g = (x + 1)^2y + 3x$. Then $\texttt{res}_y(f, g) = 3x(x + 1) - 2x^2(x + 1)^2$. Suppose in (5) $\alpha = 1$. Then LHS is -2, whereas RHS is $\texttt{res}(2y + 2, 4y + 3) = -2$. So it appears that (5) is true. What if we substitute $\alpha = -1$ in (5)? Then LHS is clearly zero, whereas RHS is a constant. What went wrong? Note that $\alpha = -1$

[1] All these definitions hold in any euclidean space $\mathbb{R}^n$. The definition of continuity generalizes the standard ϵ-δ definition for functions over $\mathbb{R}$.

[2] It is easy to see that each real root α of the content corresponds to a vertical component $x - \alpha$ of the curve

[3] The argument, in fact, works for polynomials $f, g \in D[y]$ for some UFD D. In our case, $D = \mathbb{Z}[x]$.

is a root of *both the leading coefficients* w.r.t. y. Thus the two resultants are different because $f(\alpha), g(\alpha)$ do not have the same degree in y as f, g. What we can show is that (5) almost holds, as long as α is not a root of both $\texttt{lead}_y(f)$ and $\texttt{lead}_y(g)$. The following theorem gives a more precise statement.

THEOREM 3. *Let $f, g \in \mathbb{Z}[x, y]$ be such that $\deg_y(f) = m$ and $\deg_y(g) = n$.*

1. $\texttt{res}_y(f, g) \equiv 0$ *iff* $\deg_y(\texttt{GCD}(f, g)) > 0$.
2. *If f, g are relatively prime then for any $\alpha \in \mathbb{C}$ which is not a root of both $\texttt{lead}_y(f)$ and $\texttt{lead}_y(g)$*

$$\texttt{res}_y(f, g)(\alpha) = C \cdot \texttt{res}(f(\alpha), g(\alpha)) \tag{6}$$

for some non-zero constant C.

Proof. For the proof, we have to show that when α is a root of one of the leading coefficients, say $\texttt{lead}_y(g)$, then (6) holds for an appropriate choice of C, namely $C = \texttt{lead}_y(f)(\alpha)^{n-k}$. **Q.E.D.**

As a corollary we have the following:

COROLLARY 4. *Suppose f, g are relatively prime and $\alpha \in \mathbb{C}$ is not a root of both $\texttt{lead}_y(f)$ and $\texttt{lead}_y(g)$. Then α is a root of $\texttt{res}_y(f, g)$ iff there exists $\beta \in \mathbb{C}$ such that $f(\alpha, \beta) = g(\alpha, \beta) = 0$.*

A curve f is said to be **weakly generic** if its leading coefficient w.r.t. y is a constant.A pair of curves f, g are weakly generic if both the polynomials are weakly generic. Thus the corollary above states that *for a pair of curves f, g in weakly generic position the roots of the resultant correspond to the x-coordinates of some common root (α, β) of f, g.* Now that we have the x-coordinate α of the common root, how do we get hold of the y-coordinate?

Suppose $A, B \in \mathbb{Z}[t]$ and we know that they have exactly one common root β, how can we find this root? Since there is only one common root, the gcd must look like $a(t - \beta)^k$, for $k \geq 1$ and some $a \in \mathbb{Z}$. Thus we know that the coefficient of t^{k-1} is $-ka\beta$. Therefore, if we can get hold of the leading coefficient of the gcd, the second leading coefficient, and its degree then we can express β as a rational function of these quantities. But in the univariate setting, it is trivial to compute the gcd. Another, seemingly more complicated, approach is to use the following property of subresultant sequence $\texttt{sres}_i(A, B)$:

$$\texttt{sres}_0(A, B) = \texttt{sres}_1(A, B) = \cdots = \texttt{sres}_{k-1}(A, B) = 0 \text{ iff } \deg(\texttt{GCD}(A, B)) = k.$$

Moreover, $\texttt{sres}_k(A, B) = \texttt{GCD}(A, B)$. Thus in our case

$$\beta = \frac{\text{coeff}_{k-1}(\texttt{sres}_k(A, B))}{-k \cdot \texttt{lead}(\texttt{sres}_k(A, B))}. \tag{7}$$

The relevance of using the subsresultants becomes relevant in our setting, because for us $A := f(\alpha, y)$ and $B := g(\alpha, y)$, for a real algebraic number α. Given the nature of α it is hard to do symbolic computation to compute the gcd. However, if a relation similar to (6) holds for subresultants then we can compute subresultants for f, g w.r.t. y, which gives us polynomials in $\mathbb{Z}[x]$ and substitute $x = \alpha$. It is not hard to see that if f, g are weakly generic then the following generalization of (6) holds:

$$\texttt{sres}_i(f, g, y)(\alpha) = C \cdot \texttt{sres}_i(f(\alpha), g(\alpha)) \tag{8}$$

for some non-zero constant C. Thus given the x-coordinate α of a common root of f, g, from (7) it follows that the y-coordinate β can be described as

$$\beta = \frac{\text{coeff}_{k-1}(\texttt{sres}_k(f, g, y)(\alpha))}{-k \cdot \texttt{lead}(\texttt{sres}_k(f, g, y)(\alpha))}. \tag{9}$$

(9) describes the corresponding y-coordinate as a rational function of α. However, (9) holds if A and B have only β as common root, or in terms of f, g there is only one common root (α, β) with x-coordinate α. We will call **a pair of curves f, g in generic position** (or generic) if they are weakly generic and distinct common roots of f, g have different x-coordinates. **A curve f is said to be generic** if the pair f, f_y is in generic position. We are now in a position to find the common roots of two curves in generic position.

2.3 Handling a Generic Curve

We now return to our setting when the pair of curves is f, f_y. For $\texttt{res}_y(f, f_y)$ to be well-defined we want that f, f_y are relatively prime, which is the same as saying f is square-free (i.e., f does not factor as $g^2 h$, for some $g \in \mathbb{Z}[x, y]$); note the square-free part of a non-square-free polynomial f is the same as the square-free part of $\texttt{prim}(f)$, since the content is common to both f and f_y. As a consequence, the x-coordinates of critical points of f are real roots of $\texttt{res}_y(f, f_y)$ (see Corollary 4); Moreover, assume f is in generic position, i.e., it does not have vertical asymptotes, vertical components, and no two critical points have the same x-coordinate. As a consequence, we can express the y-coordinate of a common root in terms of its x-coordinate using (9). The following algorithm gives the geometric details for computing the correct topology of a curve in generic position.

INPUT: A square-free polynomial $f \in \mathbb{Z}[x, y]$ in generic position, and a box $I \times J \subseteq \mathbb{R}^2$.
OUTPUT: A piecewise linear graph that is isotopic to f in $I \times J$.

1. Compute the subresultant sequence $\texttt{sres}_i(f, f_y, y)$, $i = 0, \ldots, \deg_y(f)$.
   ◁ *For the most part, only the leading coefficients are needed*
2. Isolate the real roots $\alpha_1, \ldots, \alpha_n$ of $\texttt{sres}_{0,y}(f, f_y)$ in I; $\alpha_0 := I_0, \alpha_{n+1} := I_1$
3. For each root α_i do:
   Isolate the real roots of $f(\alpha_i, y)$ in J and order them vertically in a stack S_i.
   Compute β_i as given by (9).
4. Let $\gamma_i \in (\alpha_i, \alpha_{i+1})$, $i = 0, \ldots, n$ be n rational points.
5. For each γ_i do:
   Isolate the real roots of $f(\gamma_i, y)$ in J and order them vertically in a stack T_i.
6. Now we connect the roots in two consecutive stacks: S_i, T_i ◁ *T_i does not contain critical points* .
   Let $\mathbf{p} := (\alpha, \beta)$ be the critical point in S_i.
   Connect the branches, if any, above $\mathbf{p}$ with one branch each from top of T_i.
   Connect the branches, if any, below $\mathbf{p}$ with one branch each from bottom of T_i.
   Connect the remaining branches, if any, of T_i to $\mathbf{p}$.
7. Connect the stacks S_{n+1}, T_n as above.

Remark: Note that in step 2 we are isolating real roots of a polynomial with algebraic number as coefficients. This requires modifying our exact algorithms for real root isolation to handle input coefficients given as bitstreams, i.e., black boxes that can be queried for any desired *absolute approximation* to the coefficients.

The procedure is illustrated in Figure 4.

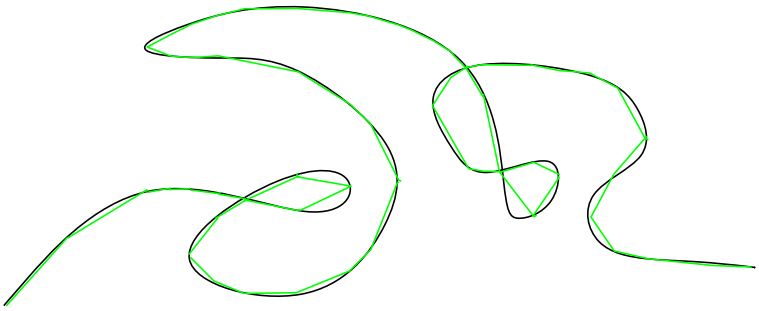


Figure 4: Computing the topology of an algebraic curve using projection

2.4 Handling a Non-Generic Curve

A requirement for the procedure described above is that the curve should be in generic position, that is no two critical points have the same x-coordinate, and the leading coefficient w.r.t. y do not have real roots; in fact, we will ensure

that the latter condition is replaced with the stronger constraint that the leading coefficient is a constant. The idea is to project the curve along a different line. Which lines can we choose? Consider the lines connecting any pair of critical points. Our aim is to find a line ℓ that does not coincide with any of these lines and make that our y-axis. Since our polynomial is square-free, there are only finitely many critical points, and hence most choices of ℓ will suffice. In fact, let $\ell := y = sx$, for a random choice of integer s. What affine map ϕ will take ℓ to the y-axis? Clearly $\phi := (x, y) \to (y - sx, y)$ takes ℓ to the y-axis. However, this map does not preserve the x-axis; it flips it around the origin and scales it by s. To avoid these unwanted outcomes, let us consider the line with slope $1/s$, i.e., $\ell := x - sy$. Then our map $\phi(x, y) = (x - sy, y)$ and, clearly, x-axis remains unchanged. We should also transform the variety, $V(f)$, accordingly. Thus, instead of looking at $V(f)$, we should look at the variety of $f \circ \phi^{-1} = f((x + sy, y))$. The transformation $\mathrm{Sh}_s : f(x, y) \to f(x + sy, y)$ is called a **shear transformation**. For example, let us consider the curve $f = x^4 + y^4 - 2x^2y^2 + 3x^2y - y^3$. Suppose $s = 1$ then the affinely transformed polynomial is $\mathrm{Sh}_1(f) =$. Note the degree of y in the sheared polynomial has decreased from 4 to 3, and the graph of the sheared polynomial shows a *vertical asymptote* at $x = 1$, which was not there initially. Though vertical asymptotes are not a severe issue, they do cause some annoyance in the sense that the resulting leading coefficient is not guaranteed to be a constant. To understand the issue better, let us shear by an unknown factor s. The resulting polynomial $\mathrm{Sh}_s(f) =$. Note that the leading coefficient of $\mathrm{Sh}_s(f)$ in y is a univariate polynomial in s. Moreover, the polynomial vanishes at $s = 1$. We had seen earlier that this implies that both the sheared polynomial and its partial derivative w.r.t. y share a vertical asymptote. To avoid this problem, our choice of s should not be a root of this univariate polynomial. This argument holds in general, because for an arbitrary curve f of degree d, we know $f = F_0 + F_1 + \cdots + F_d$ where F_i is a homogeneous polynomial of degree i. Suppose $F_d = \sum_{i=0}^{d} a_i x^i y^{d-i}$. Then the corresponding leading term of the sheared polynomial is $\mathrm{Sh}_s(F_d) = \sum_{i=0}^{d} a_i (x+sy)^i y^{d-i}$. The leading term in y is the polynomial $\sum_{i=0}^{d} a_i s^{d-i} \in \mathbb{Z}[s]$. Thus our choice of s has to further avoid these d degenerate choices. The argument above, in general, works in the case of two curves f, g and not just f, f_y. To summarize, we have shown the following.

LEMMA 5. *Let $V(f)$ and $V(g)$ be two curves without common components. Then, there exists an $s \in \mathbb{Z}$ such that $V(\mathrm{Sh}_s(f))$ and $V(\mathrm{Sh}_s(g))$ do not have vertically asymptotic arcs, and all common roots of the sheared curves have different x-coordinates.*

Whereas the above lemma is correct, it still misses our aim to make f generic. What we have guaranteed is that the x-coordinates of different singularities are different. But what is the guarantee that the x-coordinates of the *non-singular critical points* are different? In other words, Why can't there be a vertical tangent to the sheared curve that touches it at two different points? Example??? So just looking at pairs of singularities is not sufficient.

View the sheared polynomial $F(S, x, y) := \mathrm{Sh}_S(f) = f(x + Sy, y)$ as a polynomial in S, x, y. Define $D(S, x) := \mathtt{res}_y(F, F_y)$; note that $D(S, x)$ is an algebraic curve. For a given value of $S = s^*$, the roots of the univariate polynomial $D(s^*, x)$ correspond to the x-coordinates of the critical points of the sheared curve $F(s^*, x, y)$. By definition, the multiplicity of some of these roots is greater than one. A choice s^* is bad for us if the multiplicity of distinct roots of $D(s^*, x)$ is more than the multiplicity of distinct roots of $D(s, x)$, where s is in a small neighbourhood of s^*. A stronger way to capture this increase of multiplicity is that the $\deg(\mathtt{GCD}(D, D_x))$ increases at s^*; it is strong, because the degree of gcd may increase but that increment may be because of addition of a new distinct root. From the theory of subresultants, we know that this is equivalent to the vanishing of the smallest non-zero principal subresultant coefficient of D, D_x at s^*. Thus we have the following theorem.

THEOREM 6. *Let f be a square-free polynomial. Define $F(S, x, y) := \mathrm{Sh}_S(f) = f(x + Sy, y)$, $D(S, x) := res_y(F, F_y)$ and*

$$\Delta(S) := \min_k \{\mathtt{psc}_k(D, D_x, x) \not\equiv 0\}.$$

If the curve F is not in generic position, then s is a root of either $\mathtt{lead}_y(F)$ or $\Delta(S)$.

¶5. Bound on bad shear values? The theorem above gives us an upper bound on the number of bad choices of shear values. The degree of $\mathtt{lead}_y(F)$ is at most d, so it can have at most d real roots. What is the degree of $\Delta(S)$? The degree of $D(S, x)$ is at most d^2. Thus the degree of any subresultant coefficient of $D(S, x), D_x(S, x)$, when viewed as a polynomial in x, is at most n^4. Thus the total number of bad choices of shear values is bounded by $n^4 + n$. This suggests that picking a number from $1, \ldots, 2(n^4 + n)$ has probability half of being a successful shear, and so the expected number of trials is 2.

¶6. How to know whether s is a good shear factor? Though we know that most choices of s will satisfy the constraints in Theorem 6, in practice, we need to check whether the sheared polynomial is generic or not, i.e., whether a given s was a valid choice. One way to test this is to compute the x-coordinates α_i's of the critical points, and check whether $\mathtt{sres}_k(f, f_y)(\alpha_i)$, which is a polynomial in x, has one real root. This can be done by constructing a Sturm sequence for $\mathtt{sres}_k(f, f_y)$ and doing a Sturm query for this sequence.

¶7. From the sheared curve to the original curve As long as we want an output that is topologically correct, we can work with the sheared curves. Occasionally, especially when plotting curves, it is good to have the output w.r.t. a specified coordinate system (e.g., the standard coordinate system), since it helps in visualizing the curves. In this case, we have to switch back our coordinate system to undo the shearing. The challenge is to get hold of non-singular critical points w.r.t. the original coordinate system, from the sheared curve. *A shear always maps a non-singular critical point to a regular point*.