Lecture 20
CURVES

This chapter study algebraic curves.

§1. Plane Algebraic Curves

Curves are fairly complex mathematical objects. There are two main mathematical approaches to curves: from the viewpoint of differential geometry [3] or from the viewpoint of algebraic geometry [7]. The former view is more general but less constructive than the latter. We basically take the algebraic viewpoint, but we shall emphasize the computational aspects of curves. We focus mainly on plane algebraic curves. Bruce-Giblin [3] and Fulton [7] are excellent introduction to the differential and algebraic geometric viewpoints, respectively. Other references include [6, 2, 1].

What is a Curve? We all have intuitive ideas about curves because of their striking visual nature. Naively, a curve C is just a set $L(C)$ of points in some topological space S. We call $L(C)$ the **locus** of C. We are most interested in the case where S is the Euclidean plane, $L(C) \subseteq \mathbb{R}^2$. To say that the curve C is "algebraic" means $L(C)$ is the zero set of some non-zero polynomial

$$L(C) = \texttt{Zero}(A), \qquad A(X,Y) \in \mathbb{Z}[X,Y].$$

We also refer to C as "the curve $A(X,Y) = 0$" or simply, $C : A(X,Y) = 0$. We call $A(X,Y)$ the **defining polynomial** of the curve. Moreover, defining polynomials are considered equivalent up to multiplication by a non-zero scalar: $A(X,Y)$ and $cA(X,Y)$ are considered the same defining polynomial for $c \neq 0$. The equation $A(X,Y)$ contains more information than its locus $L(C)$. For instance, if $B(X,Y) = A(X,Y)^2$, then the curve $C' : B(X,Y) = 0$ and the curve $C : A(X,Y) = 0$ have the same locus. But algebraically, they are different: the curve C' is equal to two copies of the curve C. Equivalently, each point p in the locus $L(C')$ as multiplicity 2. These concepts will be made precise using algebraic tools. The choice of the space S is also important: thus equation $X^2 + Y^2 = 0$ gives a "curve" in $S = \mathbb{R}^2$ is just a single point at the origin, but in $S = \mathbb{C}^2$ consists of two lines ($Y = \pm \mathbf{i} X$). We prefer $\mathbb{R}^2$ whenever possible, but it is often convenient to consider $\mathbb{C}^2$. There is a natural notion of points in S satisfying a polynomial equation $A(X,Y) = 0$; the **zero set** $\texttt{Zero}_S(A(X,Y))$ is just the set of points in S which satisfy $A(X,Y) = 0$.

To summarize, a **plane algebraic curve** C is given by its defining polynomial $A(X,Y)$ and the underlying space S. Its locus $L(C)$ is the zero set $\texttt{Zero}_S(A(X,Y)) \subseteq S$. We often refer to C as "the curve $A(X,Y)$" when S is understood.

We see the first divergence between the algebraic viewpoint and the naive view of curves: most applications of curves are only interested in the locus $L(C)$,

and do not care about multiplicity of points in the locus. Fortunately, there is an easy way to remove multiplicities: if the polynomial $A(X,Y)$ is square-free, then all but finitely many points in $L(C)$ has multiplicity 1. Such curves are said to be **reduced**. Note that if the curve C intersects itself, then at the point p of self-intersection, p has multiplicity greater than 1. To admit such curves, we do not insist that *every* point in $L(C)$ have multiplicity 1.

We can simplify our study of curves by observing that if

$$A(X,Y) = A_1(X,Y)A_2(X,Y)$$

then the curve $A = 0$ is the union of two curves $A_1 = 0$ and $A_2 = 0$. We say the curve $A = 0$ is **reducible** if it is the union of two curves $A_1 = 0$ and $A_2 = 0$ where $\deg(A_1)\deg(A_2) > 0$; otherwise, the curve is **irreducible**. Irreducible curves are reduced, but not conversely. It is mathematically appealing to focus on irreducible curves, but this may not always be desirable. One problem is that factorization is quite non-trivial (even though it is in polynomial time). Second, factorization depends on the coefficient domain: even though $A(X,Y) \in \mathbb{Z}[X,Y]$, but we can consider its factors in $D[X,Y]$ for various D ($\mathbb{Z} \subseteq D \subseteq \mathbb{C}$). In contrast to factorization, the concept of square-freeness (i.e., of reduced curves) is intrinsic to $A(X,Y)$: A is square-free iff $\texttt{GCD}(A, A') = 1$, and the concept of `GCD` is defined in $\mathbb{Z}[X,Y]$.

Another area where the algebraic notion of curves diverge from practical applications is that in practice, we are only interested in some subset of $L(C)$ (usually a connected finite part of $L(C)$). The algebraic theory does not handle this refinement. We need semi-algebraic techniques or other methods which are common in geometric modeling. To see some issues arising from handling curve segments, consider the nodal cubic curve $Y^2 - X^2 - X^3 = 0$ illustrated in Figure 1. Consider the portion of the locus indicated by the thick curve: it is a connected set, but in the topology of the curve, the set is actually disconnected. This is related to the issue of curve tracing: we need to be able to "trace" a curve through singularities (such as self-intersection points) in order to properly order points along the curve. But how do we distinguish this computationally? One way is to introduce semi-algebraic geometry, namely geometry that are defined by polynomial inequalities as well as equalities. For example, in Figure 1, we can specify the desired portion by restricting the curve to the rectangular box indicated by dashed lines: i.e., (x,y) such that $a \le x \le b$ and $0 \le y \le c$. This introduces a new level of sophistication into the computational aspects of curves. In short, we need to be aware of gaps between the purely mathematical investigation of curves, and their computational feasibility.

Projective Space. We discuss the topological space S where $L(C)$ lives. We had defined $L(C)$ to be the set $\{(x,y) \in \mathbb{R}^2 : A(x,y) = 0\}$, i.e., we assumed $S = \mathbb{R}^2$. This is the **real affine locus** of the curve $C : A(X,Y) = 0$, and $\mathbb{R}^2$ is the **real affine plane**. If we consider the **complex affine space** $S = \mathbb{C}^2$, then the corresponding **complex affine locus** is given by $\{(x,y) \in \mathbb{C}^2 : A(x,y) = 0\}$. Unless otherwise stated, let $L(C)$ refer to the complex affine locus; the real affine locus is just $L(C) \cap \mathbb{R}^2$.

In general, for any field K, we can denote the affine field K^n by the more fancy notation, $\mathbb{A}^n(K)$. To understand the behaviour of curves at infinity, we must go to projective spaces $\mathbb{P}^n(K)$. In general, the points of $\mathbb{P}^n(K)$ are the equivalence classes over the set $K^{n+1} \setminus \{\mathbf{0}\}$ where $\mathbf{0} = (0,0,\ldots,0)$ and the equivalence relations have the form $(x_0,x_1,\ldots,x_n) \equiv (cx_0,cx_1,\ldots,cx_n)$, for all $c \in K \setminus \{0\}$. Thus a **projective point** is the equivalence class of $(x_0,x_1,\ldots,x_n) \in K^{n+1} \setminus \{0\}$, and we denote this point by $(x_0 : x_1 : \cdots : x_n)$. The tuple $(x_0,x_1,\ldots,x_n)$ is called a **set of homogenous coordinates** for

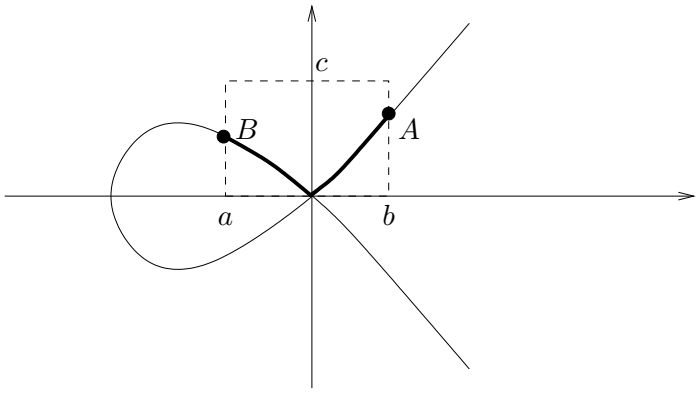

Figure 1: Nodal Cubic curve $Y^2 - X^2 - X^3 = 0$.

$(x_0 : x_1 : \cdots : x_n)$. In case $n = 2$, we have the **projective plane** $\mathbb{P}^2(K)$ over K. We use $K = \mathbb{R}$ and $K = \mathbb{C}$.

Given $A(X, Y) \in \mathbb{Z}[X, Y]$, there is a standard way to define a homogeneous polynomial $\widehat{A}(W, X, Y) \in \mathbb{Z}[W, X, Y]$ of the same degree. The homogeneous polynomial $\widehat{A}(W, X, Y)$ is characterized by the equation

$$\widehat{A}(W, X, Y) = W^d A(X/W, Y/W)$$

where $d = \deg(A)$. For instance, if $A(X, Y) = X^2Y^2 + X^2 + XY - 3Y + 5$ then $\widehat{A}(W, X, Y) = X^2Y^2 + X^2W^2 + XYW^2 - 3YW^3 + 5W^4$. Thus W is the **homogenization variable**. It follows that

$$\widehat{A}(1, X, Y) = A(X, Y). \tag{1}$$

If $H(W, X, Y)$ is any homogeneous polynomial, we can define the polynomial $H^v(X, Y)$ to be $H(1, X, Y)$. Thus, $(\widehat{A}(X, Y))^v = A(X, Y)$ for all $A(X, Y)$.

We define the **projective locus** of $A(X, Y)$ in $\mathbb{P}^2(K)$ to be

$$\{(w : x : y) \in \mathbb{P}^2(K) : \widehat{A}(w, x, y) = 0\}.$$

This definition requires justification since we have used an arbitrary set of homogeneous coordinates (w, x, y) from projective point $(w : x : y)$. However, the justification is easy because a polynomial $H(W, X, Y)$ is homogeneous iff $H(w, x, y) = 0$ iff $H(cx, cy, cw) = 0$ for all $c \neq 0$. Hence, the locus of $A(X, Y)$ in $\mathbb{P}^2(K)$ is well-defined.

The Four Loci for a Curve. We are interested in the projective locus of the curve $A(X, Y) = 0$ in two cases: $K = \mathbb{R}$ and $K = \mathbb{C}$ (real and complex projective loci, respectively). Thus we have define four loci for the curve $A(X, Y)$. How are they related? The relationship between

$$\mathbb{A}^n(\mathbb{R}) \subseteq \mathbb{A}^n(\mathbb{C})$$

is immediate since $\mathbb{R} \subseteq \mathbb{C}$. We also have the relationship between

$$\mathbb{A}^n(K) \subseteq \mathbb{P}^n(K) \tag{2}$$

is achieved via the identification $(x_1, \ldots, x_n) \in \mathbb{A}^n(K)$ iff $(1 : x_1 : \cdots : x_n) \in \mathbb{P}^n(K)$. We call x_0 the homogenization coordinate. We call points of $\mathbb{A}^n(K)$ the

finite points of $\mathbb{P}^n(K)$, and the rest are the **infinite points**, having the form $(0 : x_1 : \cdots : x_n)$. The set of these infinite points corresponds to the hyperplane defined by the equation $x_0 = 0$. Thus $\mathbb{P}^n(K)$ is simply $\mathbb{A}^n(K)$ adjoined with the hyperplane $x_0 = 0$. Summarizing, we have the following of lattice of inclusions:

$$\begin{array}{ccc} \mathbb{A}^n(\mathbb{R}) & \subset & \mathbb{A}^n(\mathbb{C}) \\ \cap & & \cap \\ \mathbb{P}^n(\mathbb{R}) & \subset & \mathbb{P}^n(\mathbb{C}). \end{array} \tag{3}$$

Focusing on $n = 2$ again, consider the affine and projective loci of a curve $C : A(X, Y) = 0$. Let $L(C) \subseteq K^2$ and $\widehat{L}(C) \subseteq \mathbb{P}^2(K)$. We characterize the extra points in $\widehat{L}(C) \setminus L(C)$ as follows: for any polynomial $A(X, Y)$, let $A_i(X, Y)$ denote the homogeneous part of $A(X, Y)$ of degree i. Then, if $d = \deg(A)$, we have

$$A = A_d + A_{d-1} + \cdots + A_0. \tag{4}$$

For instance, if $A(X, Y) = X^2Y^2 + X^2 + XY - 3Y + 5$ then $A_0 = 5$, $A_1 = -3Y$, $A_2 = X^2 + XY$, $A_3 = 0$, $A_4 = X^2Y^2$ and $A_i = 0$ for $i \geq 5$.

Lemma 1 *If $d = \deg(A(X, Y) \geq 1$ then*

$$\widehat{L}(C) \setminus L(C) = \{(x : y : 0) : A_d(x, y) = 0\}.$$

If case $\widehat{L}(C) \subseteq \mathbb{P}^2(\mathbb{C})$, it has exactly d infinite points.

Proof. In proof, from (4) we have

$$\widehat{A} = A_d + WA_{d-1} + \cdots + W^dA_0. \tag{5}$$

But $(x : y : 0)$ is a point of $\widehat{L}(C)$ iff $\widehat{A}(x, y, 0) = 0$. But (5) shows that $\widehat{A}(x, y, 0) = 0$ iff $A_d(x, y) = 0$, as claimed. In case we consider the complex projective plane, $A_d(X, Y)$ factors into exactly d factors,

$$A_d(X, Y) = c\prod_{i=1}^{d}(a_iX - b_iY)$$

and so $(b_i : a_i : 0)$ (for $i = 1, \ldots, d$) are the points at infinity in this locus. Of course, this means that there are at most d such points in $\mathbb{P}^2(\mathbb{R})$. **Q.E.D.**

Ultimately, we are interested in real affine locus but the other loci often yield additional information and permits more elegant mathematical treatment. We shall freely move among the four spaces of (3) as convenient.

Although we focus on plane curves, in general, we can study 1-dimensional varieties in n-dimensional space. These might be called "space curves", to emphasize that they are not necessarily in the plane. It is known that space curves are birationally equivalent to a plane curve – so, up to birational equivalence, the restriction to plane curves is justified.

Exercises

Exercise 1.1: What are the points at infinity of the curve $F(X, Y) = 0$?
(i) $F(X, Y) = X^2 + Y^2 - 1$.
(ii) $F(X, Y) = X^{k+1} - Y^k$ $(k \geq 1)$.
(iii) $F(X, Y) = XY - 1$.
(iv) $F(X, Y) = X^k + bY^k - 1$ ($b > 0$ and k even) ♢

End Exercises

§2. Singular Points and Multiplicity

Consider the curve $C : A(X, Y) = 0$ in some affine space S. For any point $p = (a, b) \in S$, consider the Taylor expansion of $A(X, Y)$ at p:

$$\begin{aligned} A(X,Y) &= \sum_{i\geq 0}\sum_{j\geq 0}\binom{i+j}{i}A_{i,j}(a,b)(X-a)^i(Y-b)^j \\ &= A(a,b) + A_{1,0}(a,b) \end{aligned}$$

where $A_{i,j} = \frac{\partial^{i+j}A}{\partial X^i \partial Y^j}$. We say p has **multiplicity** m if the partial derivatives $A_{i+j}(a, b)$ vanish for all $i + j \leq m - 1$ and some $A_{i+j}(a, b)$ does not vanish for some $i + j = m$. For instance, if $m = 0$ it means that $A(a, b) \neq 0$ and so p is not in the locus of C. If $m = 1$, then p is a **simple point** of C. If $m \geq 2$, then p is a **singular point**. Alternatively, p is a singular point iff $A_{1,0}(p) = A_{0,1}(p) = 0$. Alternatively,

$$\frac{\partial A}{\partial X} = \frac{\partial A}{\partial Y} = 0.$$

Write $\text{mult}_p(C)$ for the multiplicity of p in the curve C. Suppose $\text{mult}_p(C) = m$. Then consider the homogeneous function,

$$\sum_{i\geq 0}^{r}\binom{r}{i}A_{i,r-i}(a,b)(X-a)^i(Y-b)^{r-i}.$$

This equation, in $S = \mathbb{C}^2$, factors into m linear factors. Clearly, the lines passes p, and are called the **tangent lines** at p. If there are m distinct lines, we call this an **ordinary point**, otherwise an **non-ordinary point**.

Example: Consider $A(X, Y) = Y^2 - X^2 - X^3 = 0$ (see Figure 1). We have $A_{1,0}(X, Y) = -2X - 3X^2$ and $A_{0,1}(X, Y) = 2Y$. The origin $\mathbf{0}$ is a singular point since $A_{1,0}(0, 0) = A_{0,1}(0, 0) = 0$. Since $A_{0,2}(X, Y) = 2$, the origin have multiplicity 2. The tangent lines are the linear factors of

$$A_{2,0}(0,0)X^2 + 2A_{1,1}(0,0)XY + A_{0,2}(0,0)Y^2 = -2X^2 + 2Y^2 = 2(Y-X)(Y+X).$$

These are the lines $Y = X$ and $Y = -X$. Hence the origin is an ordinary singularity.

A basic inequality on multiplicity is this: for a projective irreducible curve C,

$$\sum_p mult_p(C)(mult_p(C) - 1) \leq (d-1)(d-2). \tag{6}$$

or if C is a reduced curve, then

$$\sum_p mult_p(C)(mult_p(C) - 1) \leq d(d-1). \tag{7}$$

The proof goes as follows...

§3. Curve Transformation

A transformation is an invertible function between two spaces, S and S'. In the following, assume $S = S' = K^2$ (e.g., $K = \mathbb{R}$ or $K = \mathbb{C}$). Call S the (X, Y)-space and S' the (U, V)-space, where X, Y, U, V are indeterminates. This naming convention allows us to distinguish points in S from S': a point $(a, b) \in K^2$ will be called an (X, Y)-point when viewed as a member of S, and a (U, V)-point when viewed as a member of S'.

A transformation takes each point $(a, b) \in S$ to some point $(a', b') \in S' \cup \{\infty\}$ using some rule. We need ∞ in the range in case the transformation is undefined at (a, b). The simplest transformations are linear transformations. More generally, consider transformations specified by a pair of rational functions, $U(X, Y), V(X, Y) \in K(X, Y)$. Thus,

$$T : (a, b) \mapsto (a', b') = (U(a, b), V(a, b)).$$

Such transformation T is called a **birational map** if it is invertible and the inverse T^{-1} is also given by a pair of rational functions, $X(U, V), Y(U, V) \in K(U, V)$. Thus,

$$T^{-1} : (a', b') \mapsto (a, b) = (X(a', b'), Y(a', b'))$$

where

$$X(U(a, b), V(a, b)) = a, \qquad Y(U(a, b), V(a, b)) = b \tag{8}$$

provided each function is defined at their arguments.

Suppose $F(X, Y) = 0$ defines a curve C in S. Applying the birational transformation to each point (a, b) on $F(X, Y) = 0$, we obtain the collection of points $(a', b') \subseteq S'$. Do these points define the locus of some curve in S'? The answer is yes: (a', b') are points on the curve $C' : F'(U, V) = 0$ where

$$F'(U, V) = F(X(U, V), Y(U, V)).$$

To see this, if $(a', b') \in T(\texttt{Zero}(F))$, then

$$\begin{aligned} F'(a', b') &= F(X(a', b'), Y(a', b')) \\ &= F(X(U(a, b), V(a, b)), Y(U(a, b), V(a, b))) \\ &= F(a, b) \\ &= 0 \end{aligned}$$

i.e., T maps each point in $\texttt{Zero}(F)$ into a point in $\texttt{Zero}(F')$. Conversely, Moreover, every point (a', b') on $F' = 0$ is the image of a point $(X(a', b'), Y(a', b'))$ on $F = 0$. We say that the map $(U(X, Y), V(X, Y))$ **transforms** the curve $F = 0$ into the $F' = 0$. Two curves are **birationally equivalent** if there is a birational map that transforms one to the other.

Example: consider the nodal cubic $F(X, Y) = Y^2 - X^2 - X^3$. Let $U(X, Y) = X, V(X, Y) = Y/X$. Thus, $V(a, b)$ is undefined iff $a = 0$. This specifies a birational map since it is invertible and its inverse transform is $X(U, V) = U$ and $Y(U, V) = UV$. Note that

$$\begin{aligned} F'(U, V) &= F(X(U, V), Y(U, V)) \\ &= F(U, UV) \\ &= (UV)^2 - U^2 - U^3 \\ &= U^2(V^2 - 1 - U). \end{aligned}$$

The curve $F' = 0$ is the union of the line $U = 0$ and the parabola $U = V^2 - 1$. Thus, by a suitable transformation, we "reveal" the true nature of the original curve; it is a revelation in the sense that we have transformed it into familiar or "canonical" curves we already know. This idea is expressed in the theory of canonical forms for curves.

§4. Curve Parametrization

One extremely useful property for curves is parametrization. Sendra [5] is a reference for this.

A real affine curve $F(X,Y)=0$ is said to be **parametrized** by the functions $X=X(t)$ and $Y=Y(t)$ if (1) for almost all $t\in\mathbb{R}$, $F(X(t),Y(t))=0$. and (2) for almost all points (x,y) in the locus, there exists some $t\in\mathbb{R}$ such that $(x,y)=(X(t),Y(t))$. Here "almost all" means with finitely many exceptions. In case $X(T),Y(T)\in\mathbb{R}(T)$ for some indeterminate T, we say $(X(T),Y(T))$ is a **rational parameterization** of F. Only irreducible curves are rationally parametrizable.

Equivalent formulations (see Sendra).

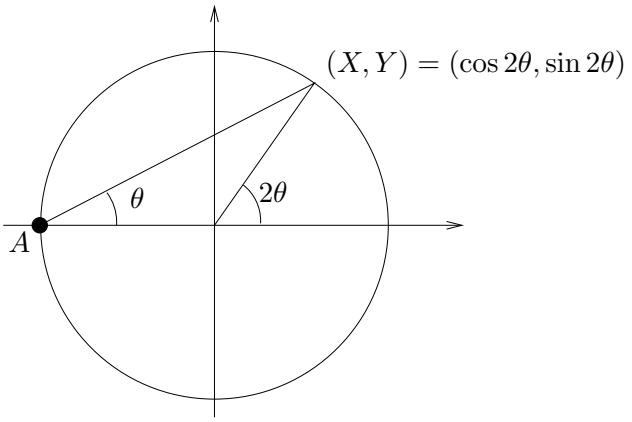

Figure 2: Parametrization of the unit circle

Example: the parabola $F(X,Y)=Y-X^2$ is rationally parametrizable because if $X(t)=t$ and $Y(t)=t^2$ then $F(X(t),Y(t))=0$ holds for all t. In fact, we see that the parabola is even **polynomially parametrizable**. On the other hand, the circle $X^2+Y^2=1$ cannot be polynomially parametrized. It has the rational parametrization

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

Referring to Figure 2, this parametrization is easily understood as followings. Originating from the point $A=(-1,0)$, we have rays $R(\theta)$ making angle θ with the positive X-axis. For $\theta\in(-\pi/2,\pi/2]$, the ray $R(\theta)$ intersects the circle at a unique point $(\cos 2\theta,\sin 2\theta)$. Recall the half-angle formulas, we have $\cos 2\theta=\frac{1-2t^2}{1+t^2}$ and $\sin 2\theta=\frac{2t}{1+t^2}$ where $t=\tan\theta$.

Not all curves are rational. E.g., for instance the cubic $Y^2-X(X^2-1)=0$ has no singular points, and hence is not rational.

On the other hand, at any non-singular point (x,y) of F, the curve $F=0$ can be locally parameterized in the following sense:

Application. Parametrizable curves simplify many computational tasks. For instance, consider the problem of **curve display**: given $F(X,Y)$, a rectangle $R\subseteq\mathbb{R}^2$, and $\varepsilon>0$, to produce a finite set $T\subseteq\mathbb{R}^2$ that is an **ε-cover** for the curve $F(X,Y)=0$ restricted to R. I.e., for every point $p\in R\cap\texttt{Zero}(F)$, every point of the curve either lies outside R or else lies within distance ε from some point of T. To solve this problem, we proceed as follows: assume that we can determine an interval $[t_0,t_1]$ such that for t outside this interval, $(X(t),Y(t))$ lies outside R. Write $p(t)$ for the point $(X(t_0),Y(t_0))$. Initially, assume that

the points $p(t_0)$ and $p(t_1)$ are already added to our set T. In general, we have an interval $[t', t'']$ and $p(t'), p(t'') \in T$. We perform a **distance test** on an interval $[t', t'']$: that is, we "certify" whether for all t in $[t', t'']$, $p(t)$ is within ε distance from $p(t')$ or $p(t'')$. If this test succeeds, we are done. If not, we add into T the point $p(t''')$ where $t''' = (t' + t'')/2$, and recursively perform distance tests for the intervals $[t', t''']$ and $[t''', t'']$. This test is called a **certification**. Alternatively, it is called a "conservative predicate" or a "one-sided predicate". With reasonable implementation, this procedure should terminate.

1. So we would like to know when a curve is rationally parametrizable. We say $F(X, Y)$ is said to be **X-regular** if X^d appears in $F(X, Y)$ and $d = \deg F$. Y-regular is similarly defined. Note that the equation $F(X, Y)$ is not Y-regular. For any degree 2 curve $F = 0$ that is not regular in one of its variables, we can apply the following general parametrization: without loss of generality, say F is a second degree equation that is not Y-regular. Then $F(X, Y) = Y \cdot (aX + b) + (cX^2 + dX + e)$ for some $a, b, c, d, e \in \mathbb{Z}$. Then we choose the rational parametrization

$$X = t, \qquad Y = (ct^2 + dt + e)/(at + b).$$

2. What if $F(X, Y)$ is regular in both X and Y? In this case, we would like to use a "shear transformation", that is, replace X by $X + \alpha Y$ for some $\alpha \in \mathbb{R}$. We just have to see what happens to the homogeneous part $F_d(X, Y)$ where $d = \deg F$. $F_d(X, Y) = aX^2 + bXY + cY^2$. Then $F_d(X + \alpha Y, Y) = Y^2(a\alpha^2 + b\alpha + c) + lot(Y)$. So we only need to choose α so that $a\alpha^2 + b\alpha + c = 0$. This amounts to solving a quadratic equation.

3. What if α turns out to be complex? This happens when $b^2 < 4ac$. For instance, for the circle $F(X, Y) = X^2 + Y^2 - 1$, we get $\alpha^2 + 1 = 0$ or $\alpha = \pm\sqrt{-1}$. This forces us to consider a more general transformation than shears: linear fractional transformations.

UNCLEAR FROM ACCOUNT OF ABHYANKAR...

Remark: shears and linear fractional transformations are special cases of the transformation

$$(X', Y', Z') = (X, Y, Z) \cdot A + (a, b, c)$$

where A is an invertible 3×3 matrix and $(a, b, c) \in \mathbb{R}^3$.

The above outline can be turned into an algorithm for computing a rational parametrization of a conic or a cubic: we leave this

§5. Linear Systems of Curves

We consider the set of all curves of a fixed degree $n \geq 1$. A bivariate polynomial $A(X, Y) \in K[X, Y]$ of degree n has up to $\binom{n+2}{2}$ coefficients in some field $K \subseteq \mathbb{C}$. As these coefficients are distinguished up to multiplication by a non-zero constant, we can view such a curve C as a point in N-dimensional projective space $\mathbb{P}^N(K)$ where

$$N = N(n) = \binom{n+2}{2} - 1 = n(n+3)/2.$$

For instance, $N(2) = 5$ and $N(3) = 9$. Thus the space of conics and cubics are 5- and 9-dimensional, respectively.

Corresponding to a curve $C \in \mathbb{P}^N(K)$, let $A(X, Y) \in K[X, Y]$ be its defining polynomial. There is also the homogeneous polynomial $\widehat{A}(X, Y, W] \in K[X, Y, W]$. For $q \in \mathbb{P}^2(\mathbb{C})$, we will write "$C(q)$" to denote the evaluation of the defining polynomial $\widehat{A}(X, Y, W)$ at the point q. The locus $L(C)$ of C is then

$$L(C) = \{q \in \mathbb{P}^2(\mathbb{C}) : C(q) = 0\}.$$

If λ is a variable λ then the linear polynomial

$$P(\lambda) = \lambda C + (1-\lambda)C'$$

is called the **pencil** generated by C, C'. For each $a \in K$, $P(a)$ represents a curve. We usually prefer to write P_a instead of $P(a)$. The family is $\{P_a : a \in K\}$ is called a **curve pencil**. The pencil is **trivial** when[1] $C = C'$, viewed as projective points in $\mathbb{P}^N(K)$. Alternatively, triviality happens exactly when $P(\lambda)$ can be factored as $(a + b\lambda)C$ for some $a, b \in K$. Pencils are useful for studying the intersection properties of the curves. Call $L(C) \cap L(C')$ the **base** of this pencil. The reason is the following observation:

Lemma 2 *Let $P(\lambda)$ be a non-trivial pencil. Let $a, b \in K$, $a \neq b$.*
(i) We have $P_a \neq P_b$. Thus, there are infinitely many curves in the curve pencil.
(ii) Also $L(P_a) \cap L(P_b)$ is the base of the pencil $P(\lambda)$.

Proof.
(i) If $P_a = P_b$ this means $aC + (1-a)C' = bC + (1-b)C'$. This means $C = C'$, contradiction.
(ii) If $q \in L(P_a) \cap L(P_b)$ then $aC(q) + (1-a)C'(q) = 0 = bC(q) + (1-b)C'(q)$. This easily implies $C(q) = C'(q) = 0$. **Q.E.D.**

Here is an interesting theorem which can be proved using pencils:

Theorem 3 (Generalized Pascal Theorem) *et C, C' be projective plane curves of degree n. If C, C' intersect in n^2 distinct points and there is another curve D of degree $m \leq n$ that passes through nm of these points, then there is third curve D' of degree $n - m$ that is disjoint from D and passes through the remaining $n(n-m)$ of these points.*

Proof. Let I be the set of mn points in $L(C) \cap L(C')$. Choose any point q on $L(D) \setminus I$. We claim that q lies in some curve $P_a = aC + (1-a)C'$ of the pencil $P(\lambda) = \lambda C + (1-\lambda)C'$. This amounts to choosing a so that $aC(q) + (1-a)C'(q) = 0$. By the previous lemma $L(P_a)$ contains the I. Thus $L(P_a) \cap L(D) \geq mn + 1$ since we also have $q \in L(P_a) \cap D$. By Bezout's theorem, P_a and D must share a common component (otherwise their intersection must have at most mn points). Since D is assumed to be irreducible, we must have $L(D) \subseteq L(P_a)$, or viewing D and P_a as polynomials,

$$D | P_a.$$

Alternatively, there is another curve E of degree $n - m$ such that $P_a = DE$. Clearly, $L(E)$ contains all the remaining $n(n-m)$ points of I. **Q.E.D.**

A special case of this theorem is the following: let $(p_0, q_0, p_1, q_1, p_2, q_2)$ be any sequence of 6 distinct points on an ellipse, and consider the three lines $L_i : \overline{p_i q_i}$, $i = 1, 2, 3$. **Pascal's theorem** says that the three points

$$r_0 = L_1 \cap L_2, r_1 = L_2 \cap L_0, r_2 = L_0 \cap L_1$$

are collinear. To deduce this from the generalized Pascal's theorem, we consider the cubic curve C comprising the three lines $\overline{p_i q_i}$ for $i = 0, 1, 2$ and the cubic curve C' comprising the three lines $\overline{q_i p_{i+1}}$. They intersect in the 9 points $\{p_i, q_i, r_i : i = 0, 1, 2\}$. Now, 6 of these points lies on a conic. Hence the remaining three lies on a line.

Exercises

[1] When C, C' are represented by two sets of homogeneous coordinates, then the equality "$C = C'$" amounts to $aC = C'$ for some non-zero $a \in K$.

Exercise 5.1: Any second degree curve is birationally equivalent to a line. Conclude that any two second degree curve is birationally equivalent. ♢

Exercise 5.2: Give another concrete application of the Generalized Pascal theorem. ♢

Exercise 5.3: (Abhyankar) Let C be a rationally parametrizable curve. Then C is polynomially parametrizable iff it has one place at infinity. NOTE: "places" are generalizations of points. ♢

End Exercises

§6. Intersection of Curves

We have already defined the multiplicity of a point P on a curve $f = 0$, $\text{mult}_P(f)$.

We now define the multiplicity $\text{mult}_P(f, g)$ of a point $P = (P_x, P_y)$ in the intersection of two curves $f = 0, g = 0$. There are several ways to do this, but computationally, the best approach is to use the resultant formulation: the naive idea is to simply define $\text{mult}_P(f, g)$ as the multiplicity of P_x as a root of $R(X) = \texttt{res}_Y(f, g)$. The main problem is that we have not used the other component P_y in this definition; this definition may give a larger multiplicity than we want when there is another point Q with the same X-coordinate as P. To get around this, we consider the (horizontal) shear transform,

$$(X, Y) \mapsto (X + tY, Y).$$

Let $R_t(X) = \texttt{res}_Y(f(X + tY, Y), g(X + tY, Y))$ and let $m(t)$ is the multiplicity of 0 as a root of $R_t(X)$. Let $P = (0, 0)$ be the origin. If $f(0, 0) = g(0, 0) = 0$, then the multiplicity of intersection $\text{mult}_P(f, g)$ is defined to be $\min_t m(t)$.

Theorem 4 *he definition of* $\text{mult}_P(f, g)$ *is invariant.*

Theorem 5 *(Bezout) Two projectives curves* f, g *of degrees* m *and* n *intersect in exactly* mn *points, counting multiplicities. More precisely,* $\sum_P \text{mult}_P(f, g) = mn$.

§7. Rational Curves

Do the basic idea about base points of curves and parametrizability of curves...

§8. Properties of Curves in Parametric Form

We now describe the basic properties of curves which are given in the parametric form.

REFERENCES: The following is from Myung-Soo's paper with In-Kwon Lee [4], and also Notes of Jan Stevens at Chalmers (see /prob/curves).

1. Difference between parametric representation and implicit representation:

E.g. The tangent to $A(x, y) = 0$ at point $p(x, y)$ is ... But the tangent to $C(t)$ at point $p = C(t_0)$ is...

The following is from Myung-Soo's paper with In-Kwong Lee (from postech):

Lemma 6 *Let $C(t) = (c_1(t), c_2(t))$ be a parametric curve and $p = C(t)$.*
(i) p is a non-self-intersecting singular point iff $c_1'(t) = c_2'(t) = 0$
(ii) p is a non-singular x-extreme point iff $c_2'(t) = 0$ and $c_1'(t) \neq 0$.
(iii) p is a non-singular y-extreme point iff $c_2'(t) \neq 0$ and $c_1'(t) = 0$.
(iv) p is an inflexion point iff $c_1'(t) \cdot c_2''(t) - c_2'(t) \cdot c_1''(t) = 0$
(v) convex...
(vi) concave...

Lemma 7 *Let C be the implicit curve $f(x, y) = 0$ and $f(p) = 0$.*
(i) p is a singular point iff $f = f_x = f_y = 0$
(ii) p is a non-singular x-extreme point iff $f = f_x = 0$ and $f_y \neq 0$
(iii) p is a non-singular y-extreme point iff $f = f_y = 0$ and $f_x \neq 0$
(iv) p is an inflexion point iff $f = f_{xx} \cdot f_y^2 - 2f_{xy} \cdot f_x f_y + f_{yy} \cdot f_x^2 = 0$.
(v) If C has no inflexion point then C is convex iff $f_{xx} \cdot f_y^2 - 2f_{xy} \cdot f_x f_y + f_{yy} \cdot f_x^2 > 0$.
(vi) If C has no inflexion point then C is concave iff $f_{xx} \cdot f_y^2 - 2f_{xy} \cdot f_x f_y + f_{yy} \cdot f_x^2 < 0$.

Tangent, Hessian of a curve Intersection Multiplicity: we define this so that Bezout's theorem is a trivial consequence of the definition!

Following Jan Stevens:

Assume $F(X, Y, Z), G(X, Y, Z)$ have no common component. Let $R(Y, Z)$ be their resultant w.r.t. X. This is non-zero and homogeneous of degree mn.

Choose a point P not on $F = 0$ or $G = 0$, and outside all lines connecting two common zeros of F and G. Taking new coords, $P = (1 : 0 : 0)$.

Let $Q = (\alpha : \beta : \gamma)$ be a common zero of F, G. By construction, this is the only point on the line $\gamma Y - \beta Z = 0$. Define the **intersection multiplicity** $\text{mult}_Q(F, G)$ of F and G at Q to be the multiplicity of $(\beta : \gamma)$ as a zero of $R(Y, Z)$.

THEOREM: Bezout is true, trivially (?)

Clean up operation:

Tangent, Hessian of a curve Let $F_d(X_0, X_1, X_2)$ be the equation of a curve C and $P = (p_0 : p_1 : p_2)$ be a point on C. The **tangent line** at P is given by the equation

$$\frac{\partial F}{\partial X_0}(P)X_0 + \frac{\partial F}{\partial X_1}(P)X_1 + \frac{\partial F}{\partial X_2}(P)X_2 = 0, \tag{9}$$

provided $\frac{\partial F}{\partial X_i}(P) \neq 0$ for some $i = 0, 1, 2$. I.e., provided P is not a singular point.

There are several ways to understand (9):

- In affine coordinates, $f(x_1, x_2) = F(1, X_1, X_2)$, and $p = (p_1, p_2)$ this equation becomes $\frac{\partial f}{\partial x_1}(p)(x_1 - p_1) + \frac{\partial f}{\partial x_2}(p)(x_2 - p_2) = 0$.

  By Euler's formula, etc (see Jan Steven Notes)

Let L be a tangent line L to a curve C at the point P. We call L an **inflexional tangent** (or a **flex** for short), and P an **inflexion point**, if the intersection multiplicity of L and C at P is at least 3.

Definition: The **Hessian** H_F of F is the curve with equation

$$\det\left(\frac{\partial^2 F}{\partial X_i \partial X_j}\right) = 0.$$

§9. Jacobi Curves and Generalizations

The following is adapted from Nicola Wolpert's thesis.

In this section, assume $f, g \in \mathbb{Z}[X, Y]$ are both normal with respect to X and Y. Suppose $p = (a, b) \in \texttt{ZERO}(f, g)$ where p is not a singularity of f or g. Let $B = [x_{\min}, x_{max}] \times [y_{\min}, y_{\max}]$ be an "isolating box" meaning that $[x_{\min}, x_{\max}]$ is an isolating interval for $\texttt{res}_Y(f, g)$ and $[y_{\min}, y_{\max}]$ is an isolating interval for $\texttt{res}_X(f, g)$. We are interested in methods for detecting (f, g) intersection inside B. Note that such a procedure will output either YES or NO, since there can be at most one intersection inside B.

[FIGURE: isolating box]

In case p is a transversal intersection, there is a simple "box hitting" method to detect p. So assume $p = (a, b)$ is a tangential intersection, i.e., the Jacobi curve

$$h = g_X f_Y - g_Y f_X \tag{10}$$

passes through p. Now, if h intersects f (and hence g) transversally, we can reduce the problem to the detection of (f, h) intersection inside B, and also of (g, h) intersection inside B.

Exercise: device a detection method in case where, if there is an intersection point p, then either the (f, g) or (f, h) intersection is transversal.

[EXPAND: Simple Hitting Box Method]

But if h and f intersect at p tangentially, we can try to see if the "next" Jacobi curve $h_X f_Y - h_Y f_X$ passes through p transversally. This process can be repeated several time. Our goal is to how to do this decisively.

From now on, assume that

$$f_Y(p) \neq 0. \tag{11}$$

If this is false, we can carry out the analogous argument with $f_X(p) \neq 0$ (recall that p is non-singular point of $f = 0$.) We also assume

$$h(p) = 0, \tag{12}$$

i.e., f and g intersect tangentially at p. Then we claim that

$$g_Y(p) \neq 0.$$

To see this, suppose $g_Y(p) = 0$. Then $g_X(p) \neq 0$. But $0 = h(p) = g_X(p) f_Y(p)$ implies $f_Y(p) = 0$, contradiction. Wow inductively define $h_0 = g$ and for $i \geq 0$,

$$h_{i+1} = (h_i)_X - (h_i)_Y \frac{f_X}{f_Y}. \tag{13}$$

Note that $h_i(p)$ is well-defined since $f_Y(p) \neq 0$. In fact, $f_Y h_1$ is the Jacobi curve h, and we also have $h_0(p) = 0$ and $h_1(p) = 0$.

Since $f_Y(p) g_Y(p) \neq 0$, by the Implicit Function Theorem, there are analytic functions $F(X), G(X)$ such that in a neighborhood of $X = a$, we have

$$f(X, F(X)) = 0, \qquad g(X, G(X)) = 0.$$

By differentiating $f(X, F(X))$, we get

$$f_X(X, F(X)) + f_Y(X, F(X)) F'(X) = 0$$

and hence

$$F'(X) = -\frac{f_X(X, F(X))}{f_Y(X, F(X))}$$

Similarly, $G'(X) = -\frac{g_X(X,G(X))}{g_Y(X,G(X))}$.

To understand the behavior of the curves $h_i(X,Y)$ at the point (a,b), we also define for $i \geq 0$,

$$H_i(X) = h_i(X, F(X)). \tag{14}$$

We already know that $H_0(a) = 0$ and $H_1(a) = 0$. Differentiating with respect to X,

$$\begin{aligned}
(H_i(X))' &= (h_i)_X(X,F(X)) + (h_i)_Y(X,F(X))F'(X) \\
&= (h_i)_X(X,F(X)) - (h_i)_Y(X,F(X))\frac{f_X(X,F(X))}{f_Y(X,F(X))} \\
&= h_{i+1}(X,F(X)) \\
&= H_{i+1}(X).
\end{aligned}$$

References

[1] S. S. Abhyankar. *Algebraic Geometry for Scientists and Engineers*. Americal Mathematical Society, Providence, Rhode Island, 1990.

[2] E. Brieskorn and H. Knörrer. *Plane Algebraic Curves*. Birkhäuser Verlag, Berlin, 1986.

[3] J. Bruce and P. Giblin. *Curves and Singularities*. Cambridge University Press, Cambridge, second edition, 1992.

[4] M.-S. Kim and I.-K. Lee. Gaussian approximations of objects bounded by algebraic curves. In *Proc. 1990 IEEE Int'l, Conf. on Robotics and Automation*, pages 322–326, 1990. May 13–18, Cincinnati ,U.S.A.

[5] J. R. Sendra. Rational curves and surfaces: Algorithms and some applications. In F. Chen and D. Wang, editors, *Geometric Computation*, chapter 3. World Scientific Publishing Co., Singapore, 2004. To appear.

[6] R. J. Walker. *Algebraic Curves*. Springer Verlag, Berlin-New York, 1978.

[7] W. Walker. *Algebraic Curves*. The Benjamin/Cummings Pub.Co., Inc, Reading, Massachusetts, 1969.