Some basic facts about algebraic curves

Vishwambhar Pati

In this article we review some basic facts of algebraic curves which will, hopefully, make the article by Arnol'd and Vasil'ev[1] (this issue, page 89) more accessible. The theory of algebraic curves, which received its first impetus from the theory of algebraic equations, dates back to antiquity. Fortunately, however, workers in various disciplines came to contribute to this theory with a variety of motivations and perspectives. For instance, complex analysis and potential theory (Weierstrass, Riemann, Gauss, Hurwitz, Ahlfors), differential equations and Abelian integrals (Riemann, Hilbert, Picard, Fuchs, Abel, Jacobi), algebra and number theory (Noether, Abel, Mordell), and geometry and topology (Gauss, Euler, Klein, Poincare, Riemann, Weyl) are some of the areas where this rich theory had ramifications, and in turn drew on. Indeed, as the article[1] mentions, problems in mechanics and optics led Newton, Fresnel and Huyghens to look at algebraic curves. By the turn of this century, most of the fundamental facts about Riemann surfaces (smooth projective curves), such as topological classification, genus, the Riemann mapping theorem, moduli, the Riemann–Roch theorem, had been established. However, to this day, the fascinating story is far from over. From Faltings' proof of Mordell's conjecture (see for example ref. 2, for a readable account) to string theory in physics, the theory of algebraic curves continues to reverberate through all the mathematical sciences.

In this note, however, we shall touch upon only a few elementary aspects of this vast domain. For the reader who wishes to pursue the story, we recommend refs. 3 and 4 as being excellent ones at an elementary level, and refs. 5–8 for the more mathematically mature.

Preliminaries

We shall assume some familiarity with basic algebra and complex analysis. To simplify our discussion, we will deal only with *plane* curves, over the field **C** of *complex numbers*. The main reason for sticking with the complex numbers is one of *the* most crucial facts of mathematics, to wit:

The fundamental theorem of algebra. Every non-constant polynomial

$$p(z)=\sum_{i=0}^{n} a_i z^i$$

of degree n, with complex coefficients a_i has n complex roots (or 'zeros') $\lambda_1, \lambda_2, \ldots, \lambda_n$, (possibly repeated). Therefore

$$p(z)=a\prod_{i=1}^{n}(z-\lambda_i),$$

where a is a complex scalar (in fact, it is the coefficient a_n of z^n in $p(z)$).

We remark here that there are several other interesting fields other than **C** which have the above property of being algebraically closed. Number theorists, for example, are interested in 'finite characteristic' algebraically closed fields. **C**, however, is the most geometrically (and analytically) accessible one with the property above. There are several proofs of the fundamental theorem of algebra. A proof using Liouville's theorem can be seen in any elementary text on complex analysis, and is a good illustration of how the analytical properties of **C** are brought to bear, to prove an algebraic fact. A pure algebra proof can be found in ref. 9.

Affine algebraic curves

Let us now define what we mean by a plane algebraic complex curve. The affine plane here is complex 2-space, viz. $\mathbf{C}^2$, the space of ordered pairs $\{(x,y): x, y \in \mathbf{C}\}$ of complex numbers. A *plane algebraic complex curve* is defined to be the subset of this plane given by

$$\{(x,y)\in\mathbf{C}^2 : f(x,y)=0\},$$

where

$$f(x,y)=\sum_{m,n,m+n\le d} a_{mn}x^m y^n$$

is a polynomial of degree d (assumed >0) in x and y. For example, when $d=1$, we get *lines*, if $d=2$, we get *conics*, and if $d=3$, we get cubics.

How does one picture these things? There is really no canonical way, since $\mathbf{C}^2$ is real Euclidean 4-space $\mathbf{R}^4$, and $f(x,y)=0$, by equating the real and imaginary parts to 0, gives two real equations, so a complex plane algebraic curve is some (real) two-dimensional object in $\mathbf{R}^4$—not the easiest thing to visualise! However, one way, in the case when $f(x,y)$ has *real* coefficients and some locus of real zeros, one can draw the curve on the real plane $\mathbf{R}^2$. For example, the ones drawn in Figure 1.

However, there are some equations, such as x^2+2y^2

Vishwambhar Pati is in the Statistics–Mathematics Unit, Indian Statistical Institute, Bangalore 560 059.

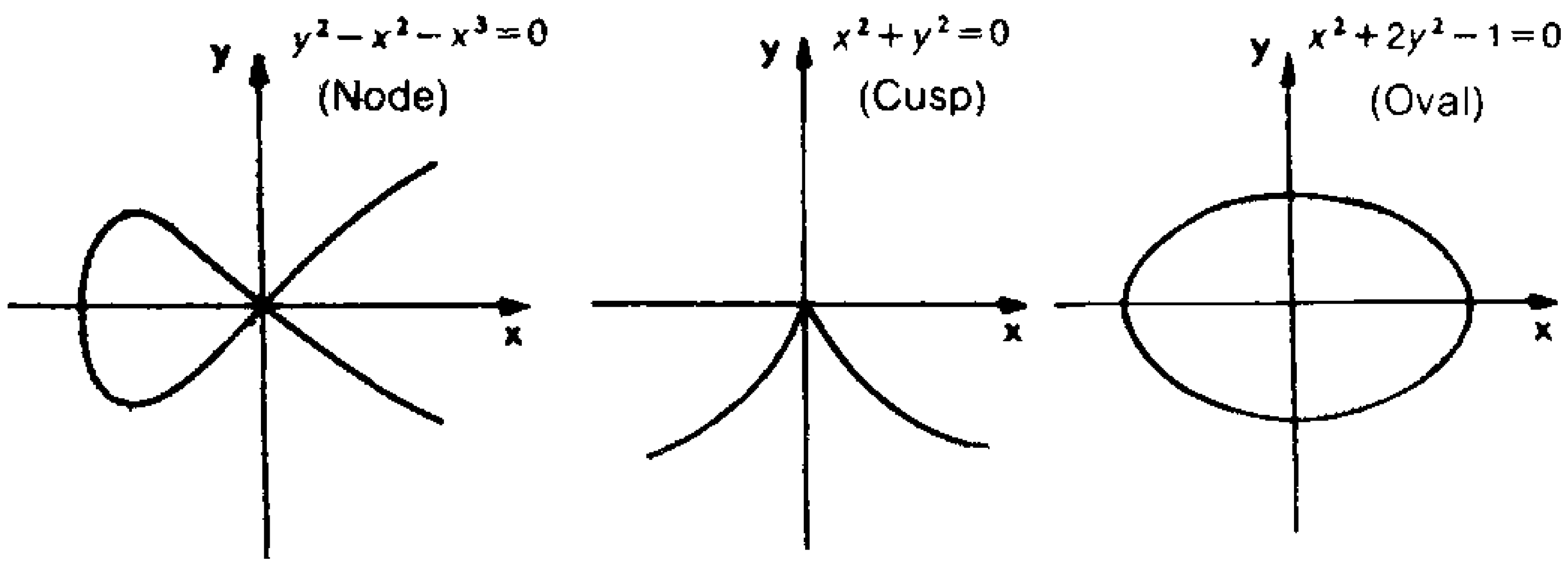

Figure 1.

$+1=0$ which have real coefficients, but no real zeros. The trick is to consider $x^2+2y^2-1=0$, which by the complex change of coordinates $x \to ix$, $y \to iy$, defines the 'same' complex curve, but now has real zeros (an ellipse). Such a curve, whose real locus is homeomorphic to a circle, is called an *oval*.

Projectivization of affine curves

A simple construction on curves (which is crucial to much more than just visualization) is called *projectivization*. There is a way to 'close up' affine *n-space* (by throwing in points at infinity) to get a compact space called *complex projective n-space* denoted $\mathbf{CP}^n$. More precisely, consider the space of complex lines in $\mathbf{C}^{n+1}$ through the origin. Any point $(z_0, z_1, \ldots, z_n) \neq (0,0\ldots0)$ determines a unique such line, and any such line determines a non-zero vector $(z_0, \ldots, z_n)$ up to scaling of *all* the coordinates by a single non-zero complex scalar λ. Thus $\mathbf{CP}^n$ is just the space of points denoted as a ratio $[z_0 : z_1 : \cdots : z_n]$, which is governed by the rules:

(i) All the z_i are not 0, and
(ii) For $\lambda \neq 0$ in $\mathbf{C}$, $[z_0 : z_1 : \cdots : z_n] = [\lambda z_0 : \lambda z_1 : \cdots : \lambda z_n]$.

These coordinates z_i of a point in $\mathbf{CP}^n$, indeterminate up to a common scaling, are called the *homogeneous coordinates* of the point. If one now looks at the (open) subset of $\mathbf{CP}^n$ defined by

$$U_0 = \{[z_0 : z_1 : \cdots : z_n] : z_0 \neq 0\},$$

it is easily seen that this U_0 is none other than affine complex *n*-space $\mathbf{C}^n$. For the point $[z_0 : z_1 : \cdots : z_n]$ in U_0 can be sent to $((z_1/z_0), (z_2/z_0), \ldots (z_n/z_0))$ in $\mathbf{C}^n$, whereas $(w_1, w_2, \ldots, w_n)$ can be sent back to $[1 : w_1 : w_2 : \cdots : w_n]$ in U_0. These two maps, by (i) and (ii) above, are clearly inverses of each other. Similarly, other U_i for $i = 1, 2, \ldots, n$ may be defined, which are all identifiable as affine $\mathbf{C}^n$'s, and $\mathbf{CP}^n$ is covered by these $(n+1)$ affine U_i's.

By all this $\mathbf{CP}^1$ is $U_0 = \mathbf{C}$ with the *single point* $[0 : z_1]$ (all values of $z_1 \neq 0$ lead to *one* point, by scaling) thrown in, so that $\mathbf{CP}^1$ is nothing but $\mathbf{C} \cup \infty$, viz. the Riemann sphere. Similarly, $\mathbf{CP}^2$ is $\mathbf{C}^2 \cup \mathbf{CP}^1$, by the following process: $\mathbf{CP}^1$ is just the space of lines in $\mathbf{C}^2$, so separately close up at ∞ each complex line in $\mathbf{C}^2$ by the point in $\mathbf{CP}^1$ which is defined by that line.

A *complex projective plane curve* is the subset

$$C = \{[z_0 : z_1 : z_2] \in \mathbf{CP}^2 : F(z_0, z_1, z_2) = 0\},$$

where $F(z_0, z_1, z_2)$ is a *homogeneous* polynomial of degree *d*. (Homogeneous of degree *d* means all its terms have total degree *d*, which is equivalent to saying

$$F(\lambda z_0, \lambda z_1, \lambda z_2) = \lambda^d F(z_0, z_1, z_2) \;\; \lambda \in \mathbf{C}.)$$

We need homogeneity to make sense of the definition because of the non-zero scaling indeterminacy of homogeneous coordinates described above in (ii). By putting z_0 (resp. z_1, resp. z_2) equal to 1 and calling the remaining two z-coordinates x and y, we will get the three polynomials $F(1, x, y)$ resp. $F(x, 1, y)$ resp. $F(x, y, 1)$ which will give the three affine pictures of the given projective curve C in the affine open sets U_0 (resp. U_1, resp. U_2). Conversely to projectivize an affine plane curve C given by a polynomial $f(x, y)$ of degree d, just plug $x = (z_1/z_0)$, $y = (z_2/z_0)$ into f, multiply the whole resulting expression by z_0^d to clear denominators, and let the resulting homogeneous degree d polynomial $F(z_0, z_1, z_2)$ be the defining polynomial for the projectivization $\hat{C}$ of C. For example, if $C = (x^2 + y^2 + 1 = 0)$, then

$$\hat{C} = \{[z_0 : z_1 : z_2] : z_0^2 + z_1^2 + z_2^2 = 0\}.$$

Note that C will of course be the affine model of $\hat{C}$ in the affine subset U_0. What do the affine models of $\hat{C}$ in the other affine pieces U_1 and U_2 mean? They describe the behaviour of the original affine curve C as we 'go towards infinity'. In this sense, projectivization brings the limiting behaviour of the affine curve at ∞ on an equal footing with its behaviour at any point of the affine plane.

Pictures of some affine curves and their projectivizations are shown in Figure 2.

Indeed, all *smooth* projective curves are represented (up to topological type) in Figure 3. For a beautiful explanation of Figures 2 and 3, see ref. 3, Ch. 1.

We say a curve (affine or projective) is *irreducible* if its defining-polynomial is irreducible, namely, cannot be

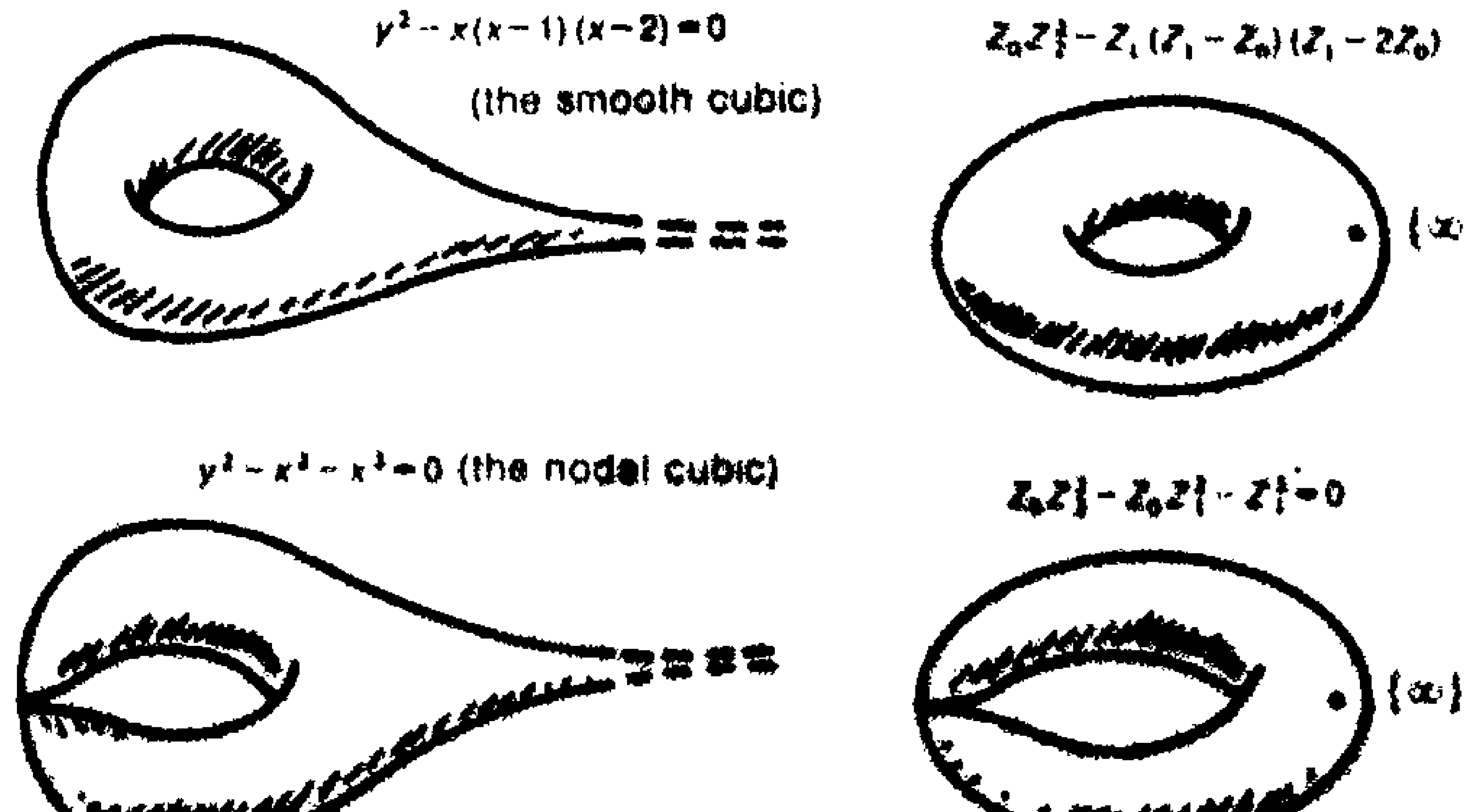

Figure 2.

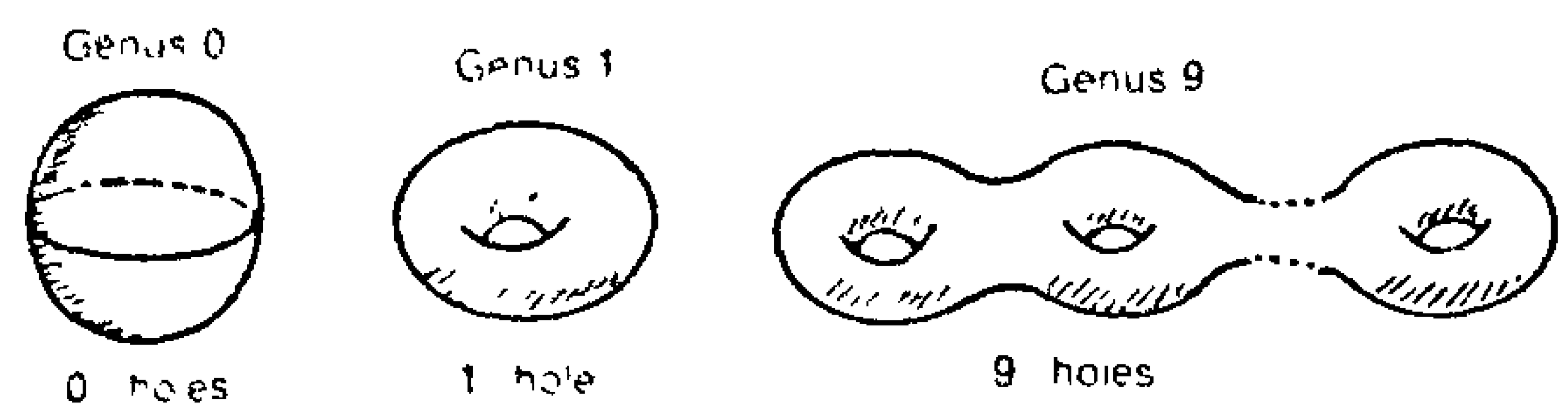

Figure 3.

factorized as a product of polynomials all of strictly smaller degree than the original polynomial. Irreducibility is the natural generalization of the notion of 'primeness' in natural numbers, to polynomials. For example, the curve ($x^2+y^3=0$) is irreducible, whereas the curve ($x^4+y^4=0$) is not, being the union of the four lines $x+\exp[\pi(2k+1)i/4]\,y=0$, where $k=0,1,2,3$. The fact that the ring of polynomials (in any fixed number r of variables) with **C** coefficients is a unique factorization domain (ref. 9, ch. v, section 6) implies that every polynomial has an essentially unique factorization as a product of irreducible polynomials, which in turn means that every algebraic curve (affine, or projective,) has a unique decomposition as a union of irreducible curves. Therefore, in some sense, it is enough to study irreducible curves, and how they meet in general, to understand all curves.

Intersections of curves, Bezout's theorem

The fundamental fact about the intersection of curves is Bezout's theorem. To state it, we need the notion of the *degree* of an irreducible curve. This we define as the degree d of the irreducible defining polynomial f (resp. F) of the irreducible affine (resp. projective) plane curve C. What is the geometric significance of this number d? Let us stick with the affine case for the moment. Consider a general complex line $\lambda x+\mu y=v$, which is nothing but a curve of degree 1 in affine 2-space $\mathbf{C}^2$. This yields

$$y=\rho x+\beta \text{ and/or } x=\tau y+\delta$$

for some constants $\rho,\beta,\tau,\delta\in\mathbf{C}$ depending on whether $\mu\neq 0$ and/or $v\neq 0$. If we plug this expression for, say, y into $f(x,y)$, two possibilities arise: (i) either the resulting polynomial in x vanishes identically, in which case one can show that $(\lambda x+\mu y-v)$ divides f, and the line is thus an irreducible component of the original curve C defined by f. But since we assumed that the curve was irreducible, the curve must be this line or, (ii) it becomes a non-zero polynomial of degree d in x. In the latter case, it will have d roots, say $x=\alpha_1,\ldots,\alpha_d$, (which may not be distinct). In this case the points

$$P_i=(\alpha_i,\rho\alpha_i+\beta),\ i=1,2,\ldots,d$$

are common points of intersection of C and the line. Indeed, by choosing a 'generic enough' line, we can ensure that all the roots α_i are distinct. (For example, the lines $y=x$ and $y=-x$ intersect the curve $x^2+2y+1=0$ in a single point $(-1,-1)$, and $(1,-1)$ respectively, whereas any other line $y=mx$ with $m\neq 1,-1$, intersects it in two points, viz.

$$(-m\pm(m^2-1)^{\frac12},\ -m\pm m(m^2-1)^{\frac12})$$

in $\mathbf{C}^2$.

Thus the degree of an irreducible curve is the number of points in which a generic line intersects it. Further, *any* line always intersects it in *at most d* points, unless the irreducible curve itself happens to be this line. This fact generalizes to all (not necessarily irreducible) curves as: a line which is not a component of the curve intersects it in at most d points. Further, a generic enough line intersects it with 'total intersection multiplicity' d. (Each point of intersection has an order, which is just the multiplicity of the corresponding root α_i in the x (or y) polynomial we get on substituting the linear equation into the defining polynomial $f(x,y)$ of the curve, and this order ≥ 1 in general. We add these up for all the points of intersection to get the total intersection multiplicity.)

A line is just a curve of degree one, so we have just seen that a curve of degree d intersects a curve of degree 1 in at most $d\cdot 1=d$ points. What about the number of intersections of a degree m and a degree n curve? This is answered by

Bezout's Theorem. *Two affine plane irreducible curves of degrees m and n respectively, (which are not identical) intersect in at most mn points. Two irreducible projective curves of degrees m and n respectively (which are not identical) intersect with total intersection multiplicity exactly mn.*

Let us first remark that, in the affine case, the number of intersection points *could be zero*. For example, the parallel lines ($x=0$) and ($x+1=0$) have no points of intersection in $\mathbf{C}^2$. However, they *do* meet at infinity, and so in $\mathbf{CP}^2$, they meet at a single point, which agrees with Bezout's theorem for projective curves. (More precisely, their projectivizations are respectively $z_1=0$ and $z_1+z_0=0$. In the affine set U_2, where $x=(z_0/z_2)$ and $y=(z_1/z_2)$, the affine models of these lines become $y=0$ and $x+y=0$, which meet at ($x=0$, $y=0$), the origin of the affine set U_2. They have no intersection in U_1, since the first line ($z_1=0$) does not meet $U_1=\{z_1\neq 0\}$ at all. This hopefully clarifies the earlier remarks about projectivization and behaviour at ∞.)

The proof of Bezout's theorem depends on an algebraic gadget called the *resultant* of two polynomials. Suppose

$$p(z)=\sum_{i=1}^{n} a_i z^i \text{ and } q(z)=\sum_{i=1}^{m} b_i z^i$$

are two polynomials of degrees n and m respectively.

We want to get a criterion in terms of the coefficients a_i, b_i for these polynomials to have a common root. Suppose λ is a common root. Then, by factorizing,

$$p(z)=(z-\lambda)p_1(z) \text{ and } q(z)=(z-\lambda)q_1(z) \tag{1}$$

so that by crossmultiplying

$$p_1(z)q(z)=q_1(z)p(z), \tag{2}$$

where $\deg(p_1)=n-1$ and $\deg(q_1)=m-1$.

Conversely, if there exist polynomials p_1, q_1 of degrees $(n-1)$ and $(m-1)$ respectively, satisfying (2) above, then we must have

$$\frac{p(z)}{p_1(z)}=\frac{q(z)}{q_1(z)}$$
$$= \text{a polynomial of deg. } 1=az+b.$$

Since $a\neq 0$ and $p(z)=(az+b)p_1(z); q(z)=(az+b)q_1(z)$ have the common root $\lambda=(-b/a)$. So the eq. (2) is equivalent to p, q having a common root. Suppose

$$p_1(z)=\sum_{i=1}^{n-1} c_i z^i; q_1(z)=\sum_{i=i}^{m-1} d_i z^i \tag{3}$$

then, substituting these into (eq. 2) and equating coefficients of z^i we get

$$c_0b_0=d_0a_0 \Rightarrow a_0(-d_0)+b_0(c_0)=0$$
$$c_0b_1+c_1b_0=d_0a_1+d_1a_0$$
$$\Rightarrow a_0(-d_1)+a_1(-d_0)+b_0(c_1)+b_1(c_0)=0,$$

and so on, to get $(n+m)$ linear equations, which can be written in the matrix form

$$(-d_0, \cdots, -d_{m-1}, c_0, \cdots, c_{n-1}) \begin{bmatrix} a_0 & a_1 & \cdots & a_n & 0 & 0 & \cdots & 0 \\ 0 & a_0 & a_1 & \cdots & a_n & 0 & \cdots & 0 \\ 0 & 0 & a_0 & a_1 & \cdots & a_n & \cdots & 0 \\ 0 & 0 & 0 & \cdots & a_0 & a_1 & \cdots & a_n \\ b_0 & b_1 & \cdots & b_m & 0 & 0 & \cdots & 0 \\ 0 & b_0 & b_1 & \cdots & b_m & 0 & \cdots & 0 \\ 0 & 0 & b_0 & b_1 & \cdots & b_m & 0 & \cdot\cdot 0 \\ 0 & 0 & 0 & \cdots & b_0 & b_1 & \cdots & b_m \end{bmatrix} = 0. \tag{4}$$

The determinant of the $(m+n)\times(m+n)$ (with m-rows of a's and n-rows of b's) matrix in the above is called the *resultant* $R(p,q)$ of the two polynomials p and q. For this to have a nontrivial solution c_i, d_j, it is necessary and sufficient that the determinant $R(p,q)=0$, as one knows from linear algebra.

Remark. **Incidentally, this result leads to a nice criterion for a root λ of a polynomial $p(z)$ to be a *repeated* root. If $p(z)=(z-\lambda)^k h(z)$ with $k\geq 2$, then it is easy to differentiate and check that λ is also a root of the differentiated polynomial $p'(z)$, and thus must be a common root of $p(z)$ and $p'(z)$. Thus the resultant** $R(p,p')=0$. This resultant is called the *discriminant* of $p(z)$. For example, the reader may check that for the quadratic az^2+bz+c, the discriminant is $a(b^2-4ac)$, and if $a\neq 0$, this is a genuine quadratic, and as expected from school algebra, has equal roots if and only if $b^2-4ac=0$.

Now let us resume the proof of Bezout's theorem. We remark that if

$$f(x,y)=0 \quad g(x,y)=0$$

are the equations of two curves of degrees n and m respectively, by writing

$$f(x,y)=\sum_{i=1}^{n} a_i(x)y^i$$
$$g(x,y)=\sum_{i=1}^{m} b_i(x)y^i \tag{5}$$

we see that if (x,y) is a point of intersection of the two curves, then y is a common root of the two polynomials of eq. (5), and so

$$R\left(\sum a_i(x)y^i, \sum b_i(x)y^i\right)=0.$$

But this determinant is a polynomial in the coefficients $a_i(x)$, $b_i(x)$, and thus a polynomial $R(x)$ in x. If we could find out its degree d, we could say that the number of solutions for x would be at most d, and thus this d would be an upper bound for the number of intersection points. Since this degree is the highest power of x in $R(x)$, we may as well assume that the a_i's are *homogeneous* of degree $n-i$ and b_j is homogeneous of degree $m-j$. It can then be seen that $R(x)$ (from the matrix in eq. (4)) is homogeneous of degree nm in x. To see this, just replace x by tx, so that $R(x)$ is multiplied by t^d. We need to find d. Now, $x\to tx$ multiplies (by our homogeneity assumption on them) $a_i(x)$ by t^{n-i}, and $b_i(x)$ by t^{m-i}, in the determinant expression for $R(tx)$ from eq. (4). Now multiply the ith row of a's by t^{m-i+1} and the jth row of b's by $n-j+1$. This multiplies $R(tx)$ by a total of $t^{1+2+\dots m+1+2+\dots n}$, viz. $t^{(m^2+n^2+m+n\ 2)}$. But now the first column has the common factor t^{m+n}, the second t^{m+n-1}, and so on, so that pulling these factors out of the determinant column by column, we get $t^{[(m+n)(m+n+1)/2]}R(x)$. So $t^{(m^2+n^2+m+n\ 2)}R(tx)=t^{[(m+n)(m+n+1)/2]}R(x)$ and thus $R(tx)=t^{nm}R(x)$, as claimed. This is what we wanted.

Parametrizations, Newton-Puiseux series

There are always local parametrizations available for affine curves in a small neighbourhood of an arbitrary point (x_0, y_0). That is, if $C=(f(x,y)=0)$, we would like to have functions $x(t)$, $y(t)$ of a single complex parameter t, well defined for small $|t|$ such that

(i) $f(x(t), y(t)) \equiv 0$, and
(ii) $x(0)=x_0$, $y(0)=y_0$.

This is called a *local parametrization* of C at (x_0, y_0). Before one talks about the existence of such, we need to define *smooth* and *singular* points on an affine curve. A point $P=(x_0, y_0)$ on $C=(f(x,y)=0)$ is a *smooth* point if

(i) $f(x_0, y_0)=0$, namely P lies on C,
(ii) $((\partial f/\partial x)(x_0, y_0), (\partial f/\partial y)(x_0, y_0)) \neq 0$;

(ii) above means that there is a non-zero normal direction defined at P and hence a well-defined tangent direction at P. A point which is not smooth is called a *singular* point. For example, on the curve $y^2=x^2+x^3$ (called *the node*, Figures 1 and 2) it is easily checked that only $(0,0)$ is a singular point, and all other points are smooth. Similarly, the *cusp* (Figure 1) defined by $(x^2+y^3=0)$ also has $(0,0)$ as the only singularity. Similarly, the curve $x^2+y^2=0$ has a singular point at the origin. Intuitively, if P is a point on C where C has a sharp corner (see Figure 4,*a*), or through which it is multiple (Figure 4,*b*), then P is a singular point. Note that since singular points are common points of intersection of the curves $f=0$, $(\partial f/\partial x)=0$ and $(\partial f/\partial y)=0$, by the last section, they are *finitely many* in number. A curve with no singular points is called a *smooth* curve.

Similar definitions can be made for projective curves. A singular point on any of the three affine models of a projective curve is called a singular point of the projective curve, and a projective curve that has no singular points is called *smooth*. However, a smooth affine curve can acquire a singularity at ∞ and thus have a singular projectivization. (For example, the smooth affine curve $y+x^3=0$ projectivizes to $z_0^2z_2+z_1^3=0$, whose affine model in U_2 (putting $z_2=1$, $z_0=x$, $z_1=y$) is $x^2+y^3=0$, a cusp singularity at $(0,0)\in U_2$.)

Local parametrization at a smooth point

Suppose (x_0, y_0) is a smooth point on the affine algebraic curve defined by $f(x,y)=0$. Then, by the definition above, at least one of the partials $(\partial f/\partial x)(x_0, y_0)$, $(\partial f/\partial y)(x_0, y_0)$ is non-zero. Say, the latter one is non-zero. Then, since f is a *holomorphic* function of (x,y), one can apply the *holomorphic implicit function theorem* (cf. ref. 10, ch. 1, theorem B.4, e.g.) to conclude that *in a small neighbourhood*, $|x-x_0|<\varepsilon$, the variable y can actually be solved as a holomorphic function $y(x)$ of x. More precisely,

$$y-y_0=\sum_{1\leq i\leq\infty} a_i(x-x_0)^i, \tag{6}$$

which converges absolutely for $|x-x_0|<\varepsilon$. If the other partial is non-zero at $P=(x_0, y_0)$, the roles of x and y can be changed in the foregoing to yield x as a function of y. If both partials are non-zero, each of x, y is solvable locally as a holomorphic function of the other. Note that this solving is always possible locally, and seldom globally.

Example. Consider $x^2+y^2+1=0$. The curve is a smooth curve, and $(0,i)$ is a smooth point. In the neighbourhood $|x|<1$, there is the holomorphic power series expansion of y

$$y-i=i\left(\frac{1}{2}x^2-\frac{1}{8}x^4+\cdots\right),$$

which just arises by writing $y=i(1+x^2)^{\frac{1}{2}}$ and expanding. Note that this expansion is only local, for $|x|<1$, and cannot be extended holomorphically to a disk of radius>1 around $x=0$.

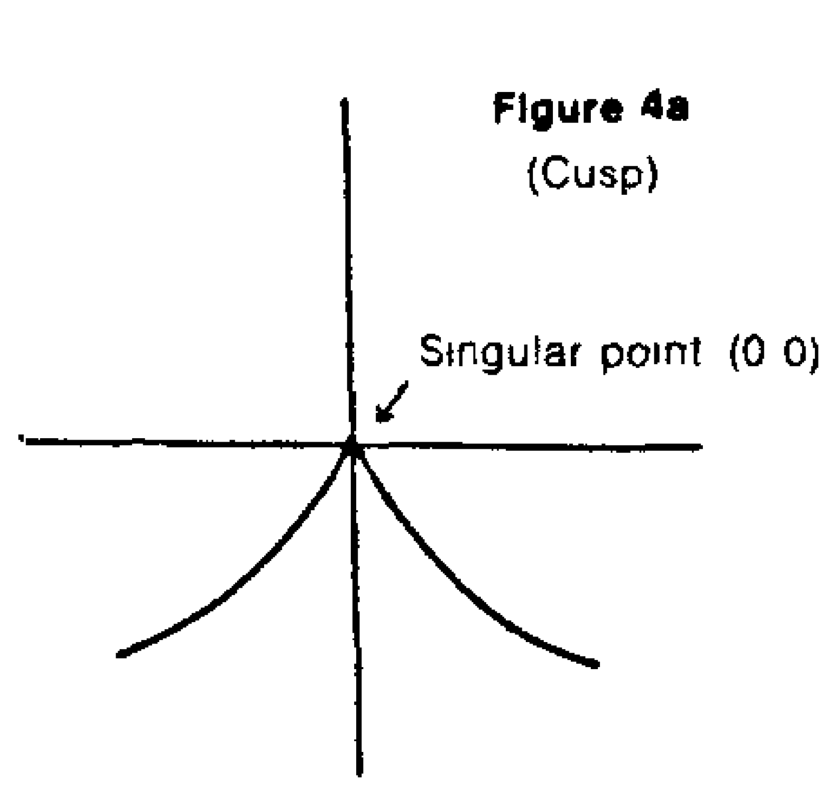

Figure 4a (Cusp)

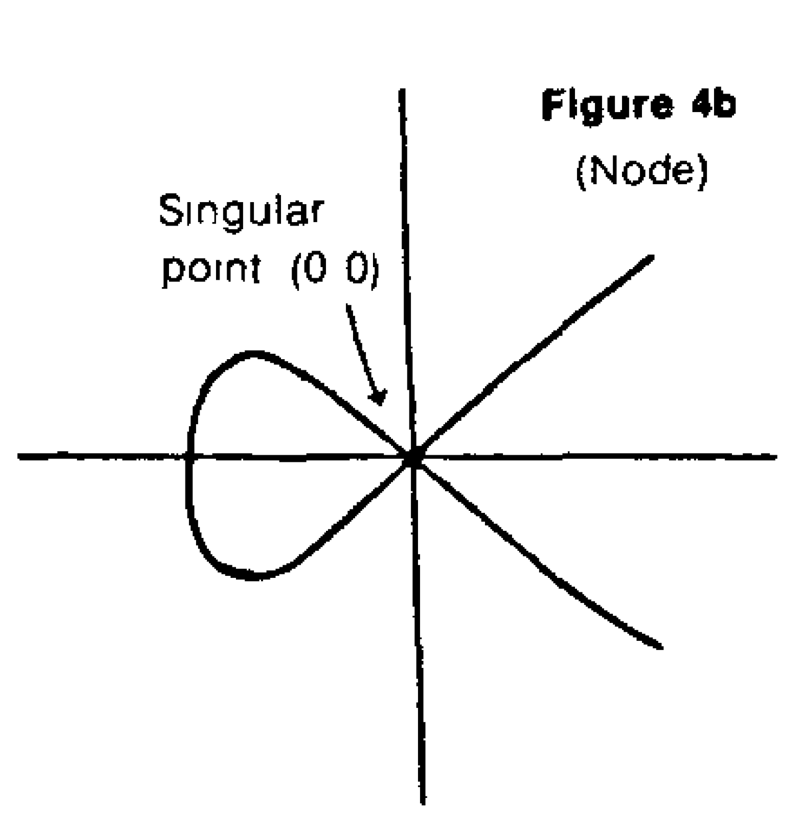

Figure 4b (Node)

Local parametrization at a singular point

Local parametrization in the neighbourhoods of singular points involves the notion of a fractional power series, or Puiseux series. Indeed, holomorphic parametrizations are not possible. For example, look at the singularity $(0,0)$ of the cusp $x^2+y^3=0$. If x were expressible as a holomorphic power series $x=\sum_{i=1}^{\infty} a_i y^i$ in some small y-neighbourhood, then if $a_1\neq 0$, the lowest order term in y in x^2+y^3 would be $a_1^2y^2$ which is non-zero. On the other hand, if a_1 were 0, then y^3 will be the lowest order y term in x^2+y^3, which is again non-zero. In neither case will x^2+y^3 vanish identically therefore. Similarly, y as holomorphic power series of x in a small x neighbourhood can also be ruled out. The only thing one can say in a neighbourhood of the origin is that $x=iy^{\frac{3}{2}}$ or $y=-x^{\frac{2}{3}}$, which are the simplest possible examples of Puiseux series. A *Puiseux series* in $x^{1/d}$ is defined to be a power series in $x^{1/d}$ for some fixed positive integer d. Clearly, a Puiseux series in $x^{1/d}$ is a Puiseux series in $x^{1/kd}$, for any *fixed* positive integer $k\geq 1$, and one can add (or multiply) a Puiseux series in $x^{1/d}$ and another one in $x^{1/d'}$ by regarding *both of them* as formal power series in $x^{1/dd'}$, and adding (or multiplying) them as one does power series. At any rate, we have the basic theorem:

Newton–Puiseux Theorem. *Let us (without loss of generality) assume that (0,0) is a singular point of the* *plane irreducible curve* $f(x,y)=0$. *Then for* $|x|<\varepsilon$, *and* ε *small enough, there is an absolutely convergent Puiseux series expansion*

$$y=\phi(x^{1/d})=\sum_{i=0}^{\infty} a_i x^{i/d} \tag{7}$$

such that $f(x,\phi(x^{1/d}))=0$.

A few words about $x^{1/d}$ are in order before we proceed. A complex number $a\neq 0$ has d dth roots, which are all scalar multiples of each other by the complex numbers $\{e^{2\pi ik/d}: k=0, 1, 2,\dots, d-1\}$ of unit modulus. Of course, 0 has a single dth root 0, and so the quantity $z^{1/d}$ is a 'multivalued function' with branch point 0 (and also ∞, if we projectivize **C** to the Riemann sphere $\mathbf{CP}^1$). Thus a Puiseux series of y in the theorem above is also multivalued in x. This is an essential feature of singular points. Indeed, the number d in the theorem is 1, if and only if y is a smooth function of x.

Let us now actually construct a Puiseux series by the Newton polygon method. (Another proof using resolution of curve singularities can be seen in ref. 11, ch. 4, section 4.3). First let us define the *order* of an infinite series

$$h(x)=a_1x^{(m_1/n_1)}+a_2x^{(m_2/n_2)}+\cdots a_ix^{(m_i/n_i)}+\cdots \tag{8}$$

with positive fractional exponents $(m_1/n_1)<(m_2/n_2)\cdots<(m_i/n_i)<\cdots$ to be the *least* exponent (m_1/n_1), and denote it by $O(h)$.

Now, let $f(x,y)$ be the equation of our curve. First rewrite f as a polynomial in y with x-polynomial coefficients, viz. as

$$f(x,y)=\sum_{i=0}^{n} a_i(x)y^i. \tag{9}$$

For instance, let us take as our prototype example the *rhamphoid cusp* which, written in this form, is

$$f(x,y)=x^4-2x^2y+(1-x+x^2)y^2, \tag{10}$$

a curve of degree 4. Let us denote, in (9) above, $O(a_i)=\alpha_i$ and plot the graph of points $P_i=(i,\alpha_i)$ for $0\le i\le n$. Join successive P_i by line segments. The resulting figure (Figure 5,*a*) is called the *Newton polygon of f*. For instance, in our example (10) above, $\alpha_0=O(a_0)=4$, $\alpha_1=O(a_1)=2$ and $\alpha_2=O(a_2)=0$. The Newton polygon is Figure 5,*b*. We remark here that in eq. (9), the coefficients could even be infinite series of the type eq. (8) for the α_i and Newton polygon to make sense.

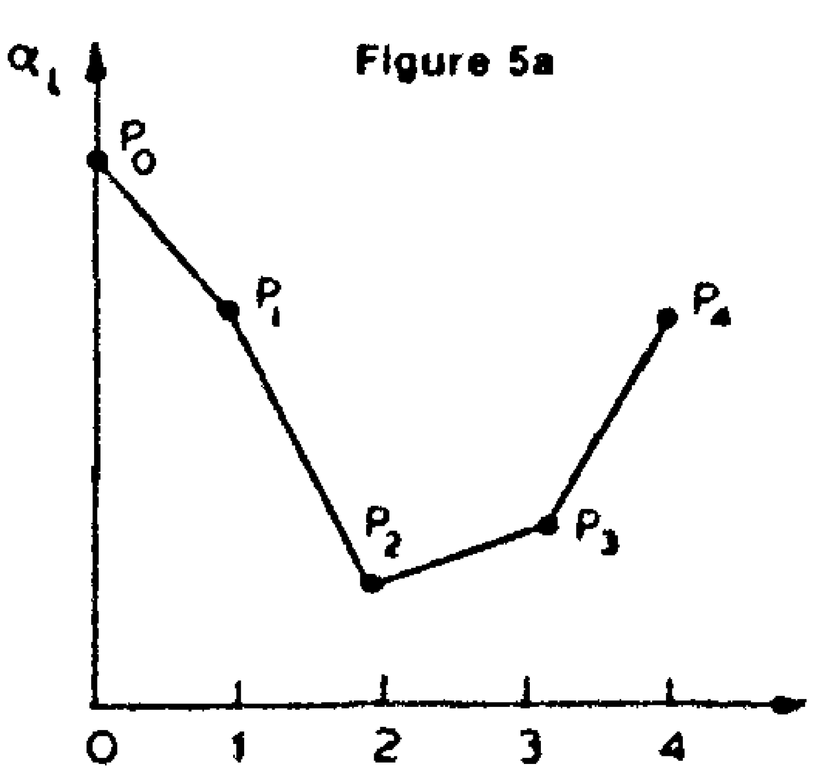

Figure 5a

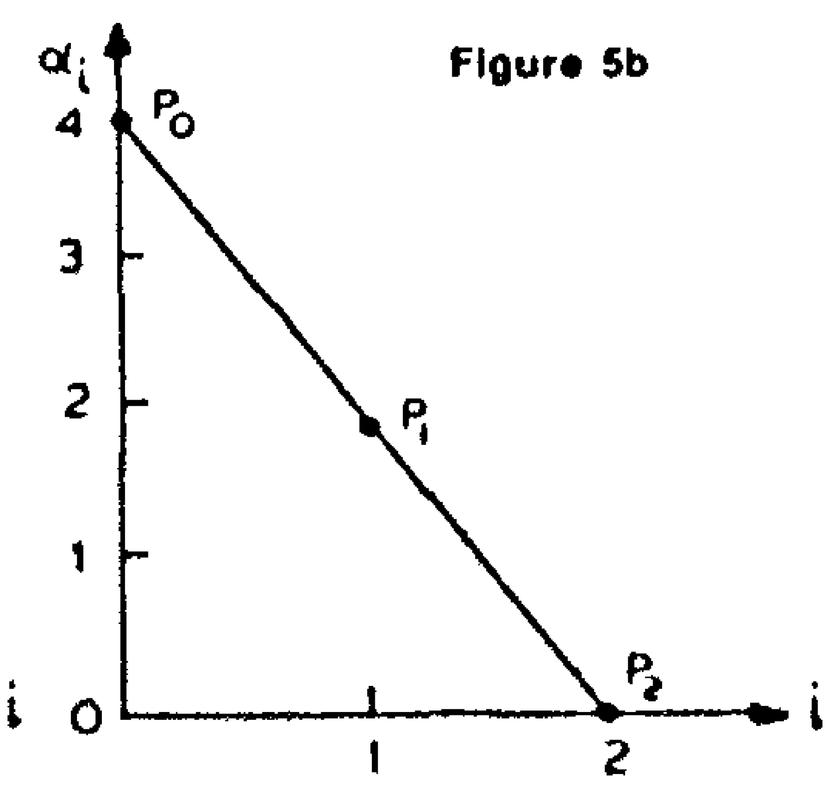

Figure 5b

The next step is to pick a line segment P_iP_j $(i\neq j)$ such that all the points P_k for all $k=0, 1,\dots,n$ lie *above* this segment. For example, in Figure 5,*a*, we could choose the segments P_2P_3 or P_3P_4, but not P_0P_1 or P_1P_3. This means that the number defined by $\alpha_i+i\gamma_1=\alpha_j+j\gamma_1=\beta$, in other words, by

$$\gamma_1=\frac{\alpha_i-\alpha_j}{j-i}$$

satisfies the relation

$$\alpha_k+k\gamma_1\geq\beta=\alpha_i+i\gamma_1=\alpha_j+j\gamma_1,$$
$$\text{for all } k=0, 1,\dots,n. \tag{11}$$

In our example (10), of course, Figure 5,*b* shows that

$$\gamma_1=2 \quad \beta=4=\alpha_k+k\gamma_1 \text{ for } k=0,1,2.$$

To get hold of the Puiseux series of y, let us assume that

$$y=c_1x^{\gamma_1}+c_2x^{\gamma_1+\gamma_2}+c_3x^{\gamma_1+\gamma_2+\gamma_3}+\dots, \tag{12}$$

where $\gamma_1\geq 0$, $\gamma_1>0$ for $i\geq 2$ are rational exponents. This does not make it a Puiseux series, however, and what one needs to do for a Puiseux series is

(i) determine c_i and γ_i by a well-defined procedure
(ii) show that the γ_i for all $i\geq N$ (and some N) have a *fixed* denominator d_N. Then, if $d_1,d_2,\dots,d_{N-1}$ are the denominators of $\gamma_1,\gamma_2,\dots,\gamma_{N-1}$, the whole expression (12) will then clearly be a Puiseux series in $x^{(1/d)}$, where $d=d_1d_2\dots d_N$
(iii) show that the resulting series converges for small $|x|$.

We shall only dwell on (i) and (ii), referring to (ref. 12, ch. 13.) for the proof of (iii). We rewrite the expression (12) as

$$y=x^{\gamma_1}(c_1+y_1), \tag{13}$$

where γ_1 *is the number determined in* (11), from the Newton polygon. Plug this expression (13) into (9) to get

$$f(x,y)=a_0(x)+a_1(x)x^{\gamma_1}(c_1+y_1)+\dots a_n(x)x^{n\gamma_1}(c_1+y_1)^n. \tag{14}$$

Now, by the definition of α_i as $O(a_i)$, each

$$a_i(x)=K_ix^{\alpha_i}+\dots\text{(higher order terms in } x)$$

so that the lowest power of x occurring in the kth term of eq. (14) is $\alpha_k+k\gamma_1$ which, by the choice of γ_1 above, is $\geq\alpha_i+i\gamma_1=\alpha_j+j\gamma_1=\beta$. So all powers of x in eq. (14) are greater or equal to β. Thus we may rewrite eq. (14) as

$$f(x,y)=f(x,x^{\gamma_1}(c_1+y_1))$$
$$=\Bigl(\sum_{\alpha_k+k\gamma_1=\beta}K_kc_1^k\Bigr)x^\beta+\Bigl(x^\beta f_1(x,y_1)\Bigr). \tag{15}$$

where we have collected up the coefficient of the lowest power x^β of x in the first parenthesis of eq. (15), and the rest in the second. For eq. (15) to identically vanish, the coefficient of the lowest power x^β in the first parenthesis above has to equal 0. This will yield a polynomial equation for c_1 (which is nontrivial, since it has at least two terms, contributed by the ith and jth terms in eq. (15), for $i \neq j$, so two different degrees c_1^i and c_1^j definitely appear) which, by the fundamental theorem of algebra, will have at least one complex solution. In our example (10), e.g. we get on substituting $y = x^2(c_1 + y_1)$ that the coefficient of the lowest power x^4 of x on being set equal to 0 gives

$$K_0 + K_1 c_1 + K_2 c_1^2 = 1 - 2c_1 + c_1^2 = 0$$

which means $c_1 = 1$ is the unique solution. (Note that the solution for c_1 will not be unique in general, but just choose one.) Having now found c_1 in our example, we plug $y = x^2(1 + y_1)$ into eq. (10) to get the expression

$$x^4[(x^2 - x) + 2(x^2 - x)y_1 + y_1^2(x^2 - x + 1)] = 0. \quad (16)$$

One sees from eq. (15) that

$$f_1(x, y_1) = x^{-\beta} f(x, x^{\gamma_1}(c_1 + y_1)),$$

because the first parenthesis of eq. (15) vanishes by definition of c_1. For our example (10) of the rhamphoid cusp, since $\beta = 4$, this becomes:

$$f_1(x, y_1) = (x^2 - x) + 2(x^2 - x)y_1 + (x^2 - x + 1)y_1^2. \quad (17)$$

Now one repeats the whole procedure, finding $\gamma_2 > 0$ from the Newton polygon for $f_1(x, y_1)$ and then setting $y_1 = x^{\gamma_2}(c_2 + y_2)$, plugging into expression for f_1, solving a polynomial for c_2, etc. Again, in our example (10), we have the Newton polygon of f_1 to be the Figure 6.

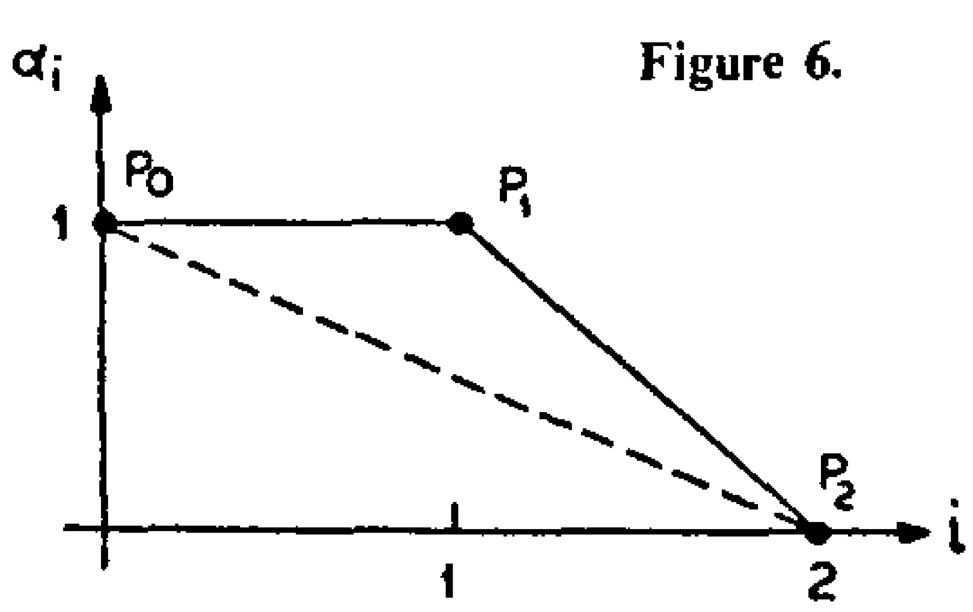

Figure 6.

The segment P_0P_2 of slope $(-\gamma_2) = -1/2$ has P_0, P_1, P_2 above it, and thus

$$y_1 = x^{\gamma_2}(c_2 + y_2) = x^{\frac{1}{2}}(c_2 + y_2). \quad (18)$$

The lowest power of x is x, and equating its coefficient to 0 after substituting for y_1 from eq. (18) into $f_1(x, y_1) = 0$ in eq. (17) leads to $c_2 = 1$.

Now one could go on in this manner to find f_2. However, we end up being lucky, and the expression (17) $f_1(x, y_1) = 0$ leads to

$$(x - x^2)(1 + y_1)^2 = y_1^2$$

$$\frac{1 + y_1}{y_1} = x^{-\frac{1}{2}}(1 - x)^{-\frac{1}{2}}$$

$$\Rightarrow y_1 = \frac{x^{\frac{1}{2}}(1 - x)^{\frac{1}{2}}}{1 - x^{\frac{1}{2}}(1 - x)^{\frac{1}{2}}}$$

$$\Rightarrow y = x^2(1 + y_1) = x^2[1 - x^{\frac{1}{2}}(1 - x)^{\frac{1}{2}}]^{-1}. \quad (19)$$

However,

$$x^{\frac{1}{2}}(1 - x)^{\frac{1}{2}} = x^{\frac{1}{2}} - \frac{1}{2}x^{\frac{3}{2}} + \frac{1}{8}x^{\frac{5}{2}} + \cdots$$

is a Puiseux series in $x^{\frac{1}{2}}$, and after plugging this Puiseux series into eq. (19) above, we easily get the Puiseux series for y.

In general, a more detailed look at the successive Newton polygons for $f_i(x, y_i)$ (see ref. 4, ch. IV, section 3), one shows that the denominators of the γ_i stabilize after a finite stage, so as to yield (ii) above and the Newton–Puiseux theorem.

1. Arnol'd, V. I. and Vasil'ev, V. A., *Am. Math. Soc.*, 1989, **36**, 1149, also this issue.
2. Srinivas, V., *Curr. Sci.*, 1990, **59**, 12.
3. Springer, G., *Riemann Surfaces*, Chelsea, 1957.
4. Walker, R., *Algebraic Curves*, Dover, 1950.
5. Fulton, W., *Algebraic Curves*, W. Benjamin, 1969.
6. Mumford, D., *Curves and their Jacobians*, Univ. Mich. Press, 1975
7. Clemens, H., *A Scrapbook of Complex Curve Theory*, Plenum, New York, 1980.
8. Forster, O., *Riemann Surfaces*, Springer-Verlag, Amsterdam,
9. Lang, S., *Algebra*, Addison-Wesley, 1971.
10. Gunning, R. and Rossi, B., *Analytic Functions of Several Complex Variables*, Prentice-Hall, 1965.
11. Arnol'd, V. I., Gusein-Zade, S. M., Varchenko, A. N., *Singularities of Differentiable Mappings*, Vol. II, Birkhauser, 1988.
12. Picard, J., *Traité d'Analyse*, Vol II, Gauthier-Villars, Paris.